Rocha et al. BMC Plant Biology (2019) 19:548
https://doi.org/10.1186/s12870-019-2180-5

BMC Plant Biology

RESEARCH ARTICLE Open Access

Unraveling candidate genes underlying biomass digestibility in elephant grass (*Cenchrus purpureus*)

João Romero do Amaral Santos de Carvalho Rocha[1], Tiago de Souza Marçal[1], Felipe Vicentino Salvador[1], Adriel Carlos da Silva[1], Pedro Crescencio Souza Carneiro[1], Marcos Deon Vilela de Resende[2], Jailton da Costa Carneiro[3], Ana Luisa Sousa Azevedo[3], Jorge Fernando Pereira[3] and Juarez Campolina Machado[3*]

Abstract

Background: Elephant grass [*Cenchrus purpureus* (Schumach.) Morrone] is used for bioenergy and animal feed. In order to identify candidate genes that could be exploited for marker-assisted selection in elephant grass, this study aimed to investigate changes in predictive accuracy using genomic relationship information and simple sequence repeats for eight traits (height, green biomass, dry biomass, acid and neutral detergent fiber, lignin content, biomass digestibility, and dry matter concentration) linked to bioenergetics and animal feeding.

Results: We used single-step, genome-based best linear unbiased prediction and genome association methods to investigate changes in predictive accuracy and find candidate genes using genomic relationship information. Genetic variability ($p < 0.05$) was detected for most of the traits evaluated. In general, the overall means for the traits varied widely over the cuttings, which was corroborated by a significant genotype by cutting interaction. Knowing the genomic relationships increased the predictive accuracy of the biomass quality traits. We found that one marker (M28_161) was significantly associated with high values of biomass digestibility. The marker had moderate linkage disequilibrium with another marker (M35_202) that, in general, was detected in genotypes with low values of biomass digestibility. In silico analysis revealed that both markers have orthologous regions in other C4 grasses such as *Setaria viridis*, *Panicum hallii*, and *Panicum virgatum*, and these regions are located close to candidate genes involved in the biosynthesis of cell wall molecules (xyloglucan and lignin), which support their association with biomass digestibility.

Conclusions: The markers and candidate genes identified here are useful for breeding programs aimed at changing biomass digestibility in elephant grass. These markers can be used in marker-assisted selection to grow elephant grass cultivars for different uses, e.g., bioenergy production, bio-based products, co-products, bioactive compounds, and animal feed.

Keywords: Gene annotation, Napier grass, *Pennisetum purpureum*, SSR marker, Trait-marker association

* Correspondence: juarez.machado@embrapa.br
[3]Embrapa Gado de Leite, Juiz de Fora, MG 36038-330, Brazil
Full list of author information is available at the end of the article

Background

Elephant grass [*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 [1]. It has a range of adaptation to different levels of altitude, precipitation, and soils, and has important agronomic traits [2], particularly its high biomass production. Depending on the environment and cultivar characteristics, green biomass can reach 300 Mg ha^{-1} $year^{-1}$ [3]; consequently, elephant grass is used for multiple purposes, including the production of bio-based compounds [4] and molecules with pharmaceutical and industrial applications [5–7]. It has been targeted by bioenergy programs, because its annual dry matter production is greater than that of sugarcane or eucalyptus, which are the most-used biomass energy sources in Brazil [2, 8]. However, the most common use for elephant grass in Brazil is in animal feed, particularly for dairy cattle.

Elephant grass is a tetraploid species (2n = 4x = 28) with two genomes (A′A′BB) [9]. Its genome is homologous to that of pearl millet [*Pennisetum glaucum* (L.) R. Br. syn. *Cenchrus americanus* (L.) Morrone], and cytogenetic studies have been conducted on these species in order to identify hybrids [9, 10]. Elephant grass has not been significantly studied, so its genomic information is little-known compared to that of more economically important crops. An RNAseq study [11], two studies on genome sequencing [12, 13], and one transcriptomic and metabolomic study [7] have been published for elephant grass. However, much is still needed to be elucidated for the next-generation breeding of elephant grass [14]. For example, few trait-marker association studies have been conducted because of the low availability of molecular markers, but simple sequence repeat (SSR) markers have been transferred from pearl millet to elephant grass [15] and used in genetic diversity studies to increase gene bank diversity [16]. Although SSR markers appear in the genome at a lower frequency than single nucleotide polymorphisms (SNPs), they can be used in genomic association studies. For example, SSR markers have been used to identify loci associated with yield components and fiber quality in cotton [17], resistance to *Sclerotinia sclerotiorum* in *Brassica napus* [18], and quantitative traits contributing to yield in sugarcane [19]. In general, many SSR markers are used in trait-marker association studies, but few may provide satisfactory results in species with limited sequence information, such as creeping bentgrass (*Agrostis stolonifera* L.) [20].

Genomic association studies identify candidate genes and markers linked to important traits. When considering multipurpose traits (i.e., traits with different purposes, such as animal feeding and bioenergy), the development of selection procedures based on molecular markers can radically streamline and accelerate elephant grass improvement [21]. Appropriate targets for breeding elephantgrass for forage and bioenergy use are agronomic traits (height, green biomass, and dry biomass) and quality traits (acid and neutral detergent fiber, lignin content, biomass digestibility, and dry matter concentration). Because elephant grass is vegetatively propagated and crosses are not widely used, the use of germplasm collections for association mapping is based on historical and naturally occurring recombination events.

In this context, the goals of this study were to (*i*) investigate changes in predictive accuracy using genomic relationship information in statistical models, (*ii*) investigate significant associations between SSR markers and eight traits, evaluated in different cuttings, in a germplasm collection of elephant grass genotypes, and (*iii*) identify candidate genes linked to these traits with their respective gene annotations. By achieving these goals, we aimed to increase the speed and accuracy of breeding elephant grass for different purposes.

Results

Genetic variation, genotype by cutting interaction, and predictive accuracy using a single-step, genome-based best linear unbiased prediction (ssGBLUP) model

Initially, a ssGBLUP model was used to fit the full dataset in order to investigate genetic variability and the genotype by cutting interaction and residual genetic variability and the genotype by cutting interaction. Genetic variability ($p < 0.05$) was detected for seven traits (height, green biomass, dry biomass, dry matter concentration, acid detergent fiber, neutral detergent fiber, and lignin content) but not biomass digestibility (Fig. 1). Regarding the genotype by cutting interaction, a 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.

Log-likelihood ratio tests revealed that residual genetic effects were significant ($p < 0.05$) for height, green biomass, and dry biomass but not for biomass quality traits (i.e., dry matter concentration, acid detergent fiber, neutral detergent fiber, biomass digestibility, and lignin content) ($p > 0.05$) (Fig. 1). There was a significant residual genotype by cutting interaction ($p < 0.05$) for green biomass, dry biomass, dry matter concentration, and lignin (Fig. 1).

The predictive accuracy of the ssGBLUP model (considering the H^{-1} matrix) ranged from 0.58 (biomass digestibility) to 0.84 (acid detergent fiber, dry matter concentration, and lignin). When the model did not include the relationship between the genotypes (simple repeatability plus the genotype by cutting interaction

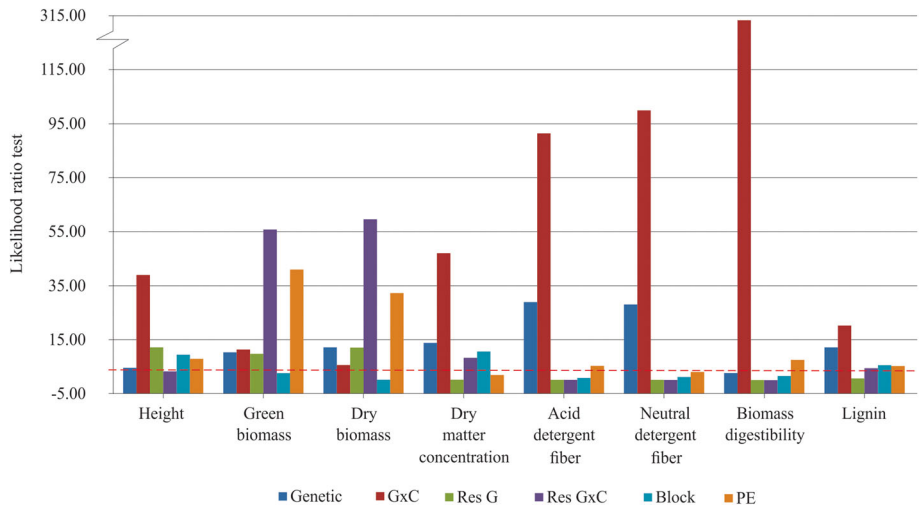

Fig. 1 Likelihood ratio tests for genetic effects (Genetic), genotype by cutting effects (G × C), residual genetic effects (Res G), residual genotype by cutting interaction effects (Res G × C), block effects (Block) and permanent environment effects (PE) considering ssGBLUP model. All bars above the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

model), the accuracy ranged from 0.59 (biomass digestibility) to 0.89 (dry biomass). Inclusion of the relationship matrix in the ssGBLUP model increased the accuracy of the biomass quality traits (Fig. 2), except for biomass digestibility.

Overall means and accuracies for each cutting

Because a significant effect of the genotype by cutting interaction was observed for all traits (ssGBLUP model, Fig. 1), trait-marker associations were investigated for each cutting. The overall means and accuracies of each cutting are presented in Fig. 3. The accuracy values ranged from 0.47 (acid detergent fiber and biomass digestibility for the fifth cutting and biomass digestibility for the first cutting) to 0.88 (dry matter concentration for the second cutting). In general, the overall means for all traits varied widely among the cuttings, corroborating the significant genotype by cutting interaction. For example, the second cutting had the highest values for height, green biomass, dry biomass, acid detergent fiber, neutral detergent fiber, and lignin, but had the lowest values for biomass digestibility.

Genome association study

We analyzed 90 elephant grass genotypes that had phenotypic data and SSR alleles available, and identified one allele of the M28 marker (M28_161) that was significantly associated with biomass digestibility. This association was detected only for the first cutting. A

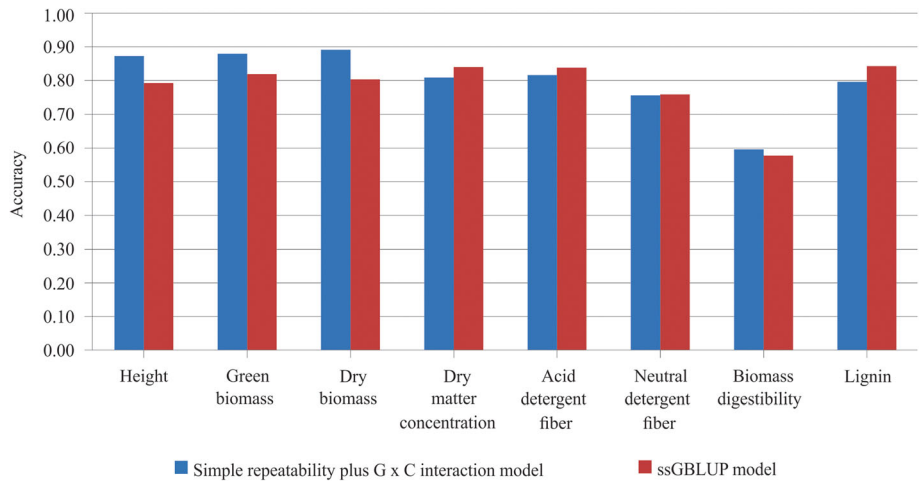

Fig. 2 Accuracy of breeding value for the simple repeatability plus genotype by cutting interaction model and ssGBLUP model

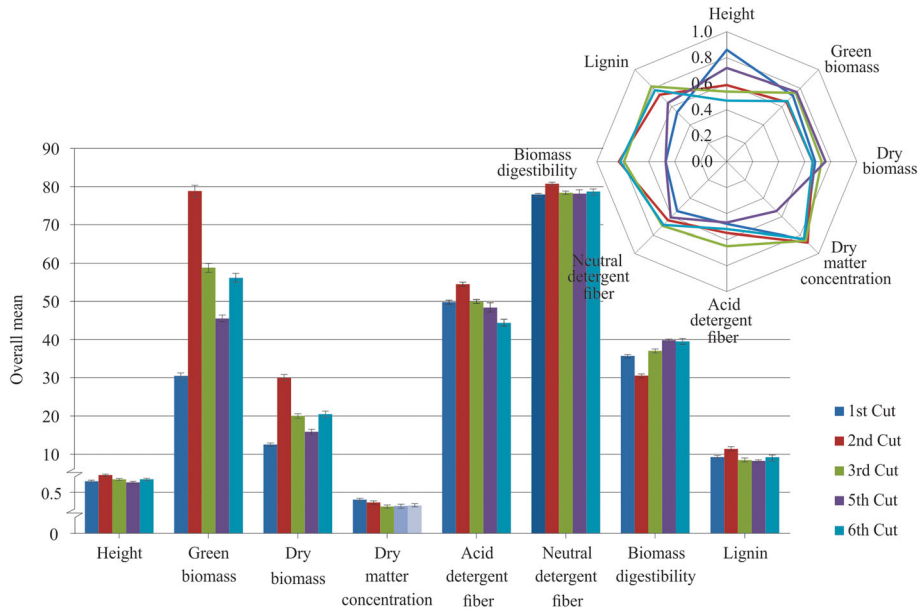

Fig. 3 Overall means (bars plot) with the standard errors and accuracy of breeding values (radar plot) for five cuttings recorded in 90 genotypes of elephant grass

Manhattan plot (Fig. 4, right) shows the $-\log(P-value)$ for all SSRs that affected biomass digestibility, while the quantile-quantile (QQ) plot (Fig. 4, left) displays significant deviations of the observed $-\log(P-value)$ from those expected.

To identify the linkage of all markers, we performed linkage disequilibrium analysis (Fig. 5). In this analysis, we were interested in identifying markers that may have been linked to M28_161, which was significantly linked to high 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 elephant grass genotypes with low values of biomass digestibility. All other r^2-values were lower than 0.11, and varied from 0.00 to 0.11.

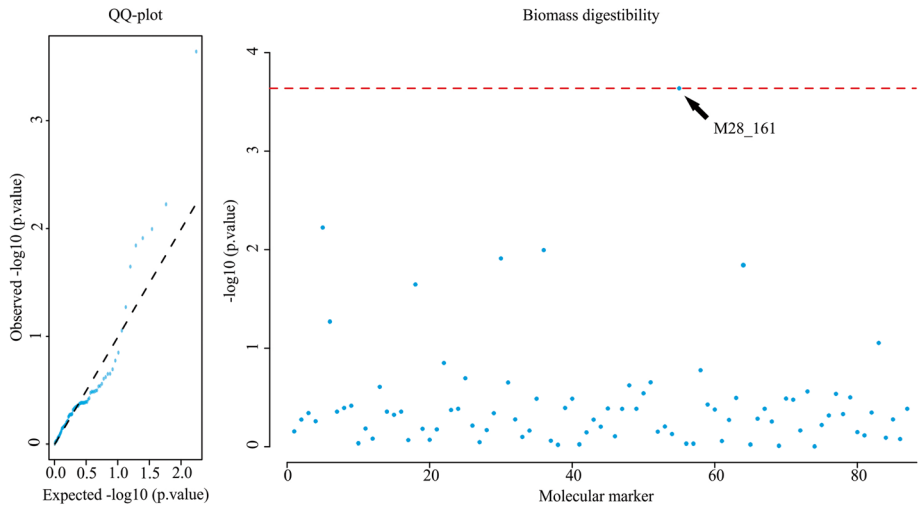

Fig. 4 QQ-plot (on the left) and Manhattan plot (on the right) for genome association of biomass digestibility phenotype with SSR markers on the first cut of elephant grass. Dashed red lines on Manhattan plot (false discovery rate) indicate the minimum threshold to select significant markers. The arrow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

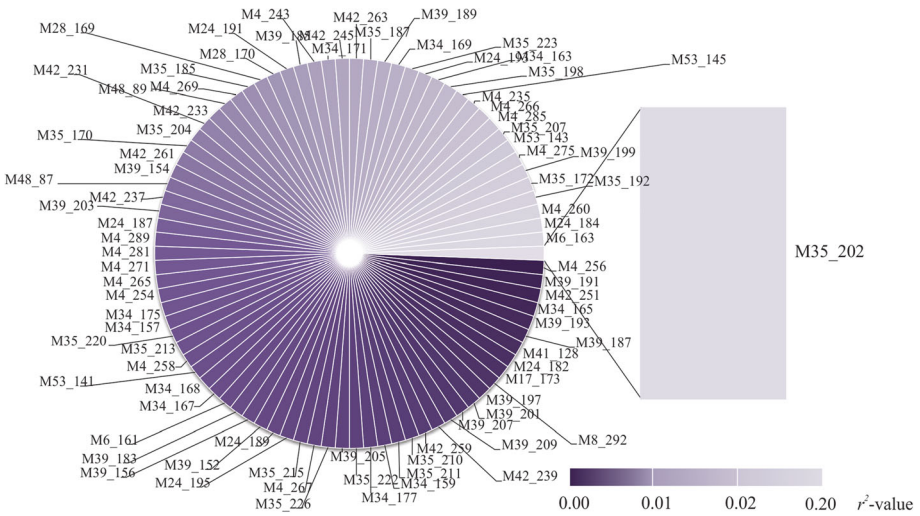

Fig. 5 Linkage disequilibrium (r^2-values) between all markers used in this study with the marker M28_161, which was found as significantly associated with biomass digestibility in elephant grass

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

A 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), M28_161 is close to candidate genes linked to pathways influenced by plant hormones (salicylic acid or abscisic acid). However, in most cases, M28_161 and M35_202 are associated with candidate genes involved in the synthesis of cell wall components. In *P. halli*, M28_161 is close to candidate genes involved in lignin biosynthetic processes and cell wall organization (chromosome 7), and in *Panicum virgatum*, it is close to a candidate gene involved in lignin catabolic processes and oxidation-reduction processes (located on chromosome 8). In *S. viridis*, both M28_161 and M35_202 were found to have orthologous regions. M28_161 is close to a candidate gene on chromosome 3 that plays a role in xyloglucan biosynthesis, while M35_202 is close to a candidate gene on chromosome 7 that functions in lignin biosynthesis. For *Setaria italica*, a successful BLAST search was conducted for the M28_161 sequence, but no candidate genes related to digestibility were found. However, for *C. americanus*, no BLAST results were obtained when using the M28 and M35 sequences.

Discussion

Genetic variation and genotype by cutting interaction

We observed a significant effect of the genotype by cutting interaction for all eight traits analyzed, so the genomic association study was conducted by considering a single cutting at a time. The lack of genetic variability

Table 1 Candidate genes, in the genome of other C4 species, that are near to homologous sequences to the SSR markers M28_161 and M35_202

Marker	Candidate gene	Reference genome	Chr[a]	Biological pathway	Ortholog locus on *A. thaliana*	Locus position[b]	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

[a]Chromosome; [b]Locus that is closest to the SSR marker

(i.e., for biomass digestibility) revealed by the ssGBLUP analysis does not mean that there was no genetic variability, because the interaction effect may have reduced it. Residual terms were added to the ssGBLUP model to capture nonadditive effects. For crops that exhibit clonal propagation and are not lines (such as elephant grass), nonadditive effects represent epistatic and dominance effects [22], and additive effects were not explained by the genomic relationship matrix, so these effects would not have inflated the residual estimates or overlaid the other effects. The nonsignificant residual genetic effect for the biomass quality traits indicates that part of the additive fraction that plays a role in the genetic architecture of these traits could be explained by only a few markers (87 alleles). For the morpho-agronomic traits the residual genetic effect was significant, so the genomic relationship matrix constructed from 87 markers did not explain the additive genetic fraction for height, green biomass, and dry biomass.

Predictive accuracy

Cell component fractions that could affect forage digestibility include cellulose, lignin, hemicellulose, and cell wall proteins [23]. Therefore, biomass digestibility is dependent upon several components which makes it a complex trait that may be explained by the fact of the accuracy has been moderate magnitude considering the genomic information. Recent transcriptomic and genomic studies [7, 11–13] and the development of SNPs for elephant grass [12] have increased accuracy, and studies using molecular markers have been important for the development of SNP markers. New genotyping, sequencing, and bioinformatics tools have increased accuracy and decreased the price per sequenced base or molecular genotype [14], and selection time [24].

Trait-marker association analysis

None of the agronomic traits (height, green biomass, and dry biomass) and four of the quality traits (acid and neutral detergent fiber, dry matter concentration, and lignin content) were associated with SSR markers. This was probably related to the low number of SSR markers used (18 SSR markers that originated in 87 alleles) when compared to the genome size of elephant grass, which has recently been estimated as 2.1 Gb [13]. However, one quality trait (biomass digestibility) was associated with the SSR marker M28_161 when the dataset from the first cutting was analyzed. This marker was in linkage disequilibrium with M35_202, so is associated with biomass digestibility. However, these markers are linked to different values of biomass digestibility, because M28_161 was significantly associated with high values of biomass digestibility while M35_202 was more frequently detected in genotypes with low values of biomass digestibility. This is an interesting result considering that digestibility is an important trait for many plant species, and understanding its impact on plant quality and the genomic regions associated with it has been the focus of many research groups [25–27].

The SSR markers used here were previously developed for pearl millet [28, 29], and were identified by our group to cross-amplify the elephant grass genome [15]. It is unknown to what extent these markers are distributed in the elephant grass genome, but our analysis shows that some markers have moderate linkage disequilibrium. The average similarity coefficient among 107 accessions of the Active Elephant Grass Germplasm Bank maintained by Embrapa Gado de Leite (where the genotypes were obtained) was 0.651, which indicates genetic variability [15]. In addition, Azevedo et al. [15] detected only one group of similarity for the genotypes used in the present study. In this context, neither the small number of markers used here nor the absence of population structure correction were obstacles to detect marker M28_161 as being significantly associated with biomass digestibility. Although trait-marker association analysis can be performed using non-specific markers (such as diversity arrays technology) or by SNP genotyping, SSR markers are important when genome sequencing and bioinformatics are not trivial tasks.

Marker annotation and in silico pathway

We ran BLAST searches using sequences of the markers M28_161 and M35_202, and the genome of the related C4 grasses *S. italica*, *S. viridis*, *P. virgatum*, *P. halli*, and *C. americanus*. These species are closely related to elephant grass [30, 31]. No results were found for the *C. americanus* genome sequence, which was unexpected because the SSR markers used here were developed from pearl millet [15] and the A′A′ genome of elephant grass is homologous to the A genome of pearl millet [9, 10]. For the other species, six candidate genes that are associated with the biosynthesis of cell wall molecules were identified.

One candidate gene annotated on chromosome 3 of *S. viridis* (Sevir. 3G340800) is orthologous to locus AT2G20370 of *Arabidopsis thaliana*, and codes for a xyloglucan galactosyltransferase responsible for different functions, including the synthesis of cell wall materials. Xyloglucan, which is a component of the plant cell wall, is a type of hemicellulose that has the ability to bind to cellulose to form a cellulose-xyloglucan network linked through hydrogen bonds (see review by Pauly and Keegstra [23]). It is unclear whether xyloglucan decreases biomass digestibility, but it is associated with lignin, which is a cell wall component that commonly negatively affects digestibility. Xyloglucan binds to cellulose to form an aggregate that, in some types of cell, can be

embedded in a matrix that contains lignin [23]. However, it is important to note that although xyloglucan is the most abundant hemicellulose in the primary cell walls of dicotyledons, in grasses such as elephant grass, its abundance is lower [23]. The effects of xyloglucan galactosyltransferase on biomass digestibility in elephant grass should be carefully evaluated.

No markers were associated with lignin content. Lignin is the second-most abundant biopolymer on Earth [32], and is a highly condensed phenylpropanoid matrix that is relatively difficult to digest by ruminal microorganisms and intestinal enzymes [33]. In many plant species, there is a negative correlation between lignin content and digestibility [34, 35] that can affect animal performance, because a small increase in dry matter digestibility (1%) can increase beef cattle daily weight gains by 3.2% [36]. In general, the quality of forage grasses decreases as they mature as a consequence of secondary cell wall deposition and the lignification of sclerenchyma cells. This does not mean that decreasing the lignin content is the only way to improve plant digestibility, but it is commonly accepted that lignin is a potential target for that purpose. In this context, the candidate genes Pahal. G00901, Pavir. 8KG357500, and Sevir. 7G164200, annotated in the *P. halli*, *P. virgatum*, and *S. viridis* genomes, respectively, are useful assets. These candidate genes are orthologous to the loci AT5G48930, AT3G09220, and AT5G48930 in *Arabidopsis*. AT5G48930 is a hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase that is involved in the phenylpropanoid pathway, and plays a role in the production of hydroxycinnamyl alcohols (or monolignols) that serve as the building blocks of lignin [37]. The lignin content can also affect bioenergy production [38], and can be decreased or increased depending on how the biomass is treated for energy generation. For example, for the conversion of lignocellulosic biomass to ethanol, polysaccharides from the cell wall need to be hydrolyzed to simple sugars and then fermented to ethanol [39]. In this case, reducing the lignin content can increase ethanol production through conventional biomass fermentation [38]. However, when considering biomass combustion, more lignin is needed for high energy conversion [40]. The molecular basis of lignin content in elephant grass is unknown, but the candidate genes identified here can be used as targets for gene manipulation. Either by transgenesis or gene editing, the manipulation of genes associated with lignin production can result in germplasm that is ideal for animal feeding or bioenergy [35, 38, 41, 42].

Some candidate genes are associated with plant hormones (salicylic acid or abscisic acid), the cross-talk of which plays an important role in the molecular responses of plants to stress [43], and two of them are involved in responses to biotic and abiotic stressors. The candidate gene Pahal. G00889 (annotated on chromosome 7 of *P. halli*) is orthologous to locus AT5G05190 in *Arabidopsis*, in which a hypothetical protein is involved in different functions, including responses to fungi [44]. The candidate gene Pahal. E03247 (also from *P. halli*) has an orthologous locus in *Arabidopsis* (AT3G05880) that codes for a small and highly hydrophobic protein that is involved in the hyperosmotic salinity response and response to cold. Validating these candidate genes in elephant grass would increase our knowledge of how this species responds to stressors such as spittlebug (Cercopidae) attack and drought or flooding, which are important factors for breeding programs in Brazil [14].

Conclusions

This study showed that, even by using a few SSR markers, it is possible to identify candidate genes associated with biomass digestibility. It also showed that it was possible to increase predictive accuracy in elephant grass by incorporating genomic relationship information. Because there is little genomic information for elephant grass available, our findings may improve elephant grass breeding. For example, marker-assisted selection can be applied, and markers associated with biomass digestibility have the potential to drive elephant grass selection for different uses (e.g., bioenergy production and animal feed). Further validation of the candidate genes revealed here may lead to a better understanding of biomass digestibility variation and its genomic basis.

Methods

Plant materials and experimental information

One hundred elephant grass genotypes (Additional file 1: Table S1) from the Active Elephant Grass Germplasm Bank (Embrapa Gado de Leite, Brazil) were planted in 0.20-m-deep furrows with 80 kg ha^{-1} P_2O_5 fertilizer applied at planting on December 23rd, 2011. The red-yellow latosol soil at the Embrapa Gado de Leite experimental station in Coronel Pacheco, MG, Brazil (latitude 21°33′18′′ S, longitude 43°15′51′′ W, 417 m.a.s.l.) 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 4-m rows that were planted side by side, 1.5 m apart. Plots were allocated in a 10 × 10 simple lattice design, with two replications. At 30 days after planting, the plots were cut to 0.30-m stubble height (uniformly cut). The number of days to reach each of the six growing seasons (cuttings) started at this time. Maintenance fertilization was made with 300 kg ha^{-1} of a N-P_2O_5-K_2O formulation (20:05:20 blended granular

fertilizer) after all cuttings. Fertilization was conducted according to the results of a soil analysis.

Six cuttings were conducted: the first cutting (September 28th, 2012) was performed at 250 regrowth days, the second (June 5th, 2013) at 250 regrowth days, the third (April 16th, 2014) at 315 regrowth days, the fourth (January 16th, 2015) at 275 regrowth days, the fifth (November 27th, 2015) at 315 regrowth days, and the sixth (June 24th, 2016) at 210 regrowth days. The fourth cutting collected propagation material, i.e., no phenotypic data were obtained at this cutting. We declare that all plant materials in this study represent voucher specimens that have been deposited in a publicly available herbarium (Active Elephant Grass Germplasm Bank) and that comply with institutional, national, and international guidelines for the collection and cultivation of any plant materials.

Phenotypic traits

The plants were phenotyped at first, second, third, fifth, and sixth cuttings for each of the following traits: (*i*) height (m) was obtained from the arithmetic mean of the height of three randomly selected plants in each plot, measured from ground level to the curve of the last completely expanded leaf; (*ii*) green biomass (Mg ha^{-1}) was obtained from a cutting taken at 7.5 cm stubble height in a 3-m section in the middle of the rows using a gasoline-powered strimmer and that was 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 [45]; (*vii*) biomass digestibility (g Kg^{-1}) was determined by the method described by Tilley and Terry [46]; and (*viii*) dry matter concentration (%) was obtained by sampling three complete plants from each plot, which were dried in a kiln after weighing (fresh weight) until weight stabilization. The samples were weighed (dry weight) again, and 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, lignin content, and biomass digestibility, random samples of three complete plants from each plot were collected before cutting the experimental plots. These samples were dried in a forced-air circulation oven at 56 °C for 72 h. After drying, the samples were ground to small particles (1 mm) in a Wiley type grinder and analyzed as described above. For the third, fifth, and sixth cuttings, acid and neutral detergent fiber, lignin content, biomass digestibility, and dry matter concentration were measured using near-infrared spectroscopy (NIRS). Data generated from the first and second cuttings, and from other experiments (i.e., by traditional methodologies of biomass quality analysis), were used for NIRS calibration. The phenotypic data are shown in Additional file 2: Table S2.

Genotyping, quality control, and imputation

Eighteen SSR markers were used for genotyping, as described by Azevedo et al. [15]. The alleles for each marker are shown in Additional file 3: Table S3. Due to the multiallelic nature of SSR markers associated with the polyploidy of elephant grass, each allele was considered a marker (totaling 111 markers). For each marker, individuals were coded as 0 (absence of allele) or 1 (presence of allele) according to Viana et al. [47]. SSRs with more than 15% missing values (i.e., a call rate of at least 85%) and/or a frequency of minor alleles of above 1% were removed. The following imputation algorithm was used for the missing value data point in a marker matrix (M_i): $M_i = \begin{cases} if: p_i \leq 0.5 \rightarrow M_i = 0 \\ if: 0.5 < p_i \leq 1 \rightarrow M_i = 1 \end{cases}$, where p_i is the allele frequency associated with the presence of a marker at locus *i*. The algorithm was directly implemented as an R function. After quality control, 87 markers were used. According to previous studies [28, 29, 48, 49], the SSR markers used here are on chromosomes 1, 3, 4, 5, 6, and 7 of pearl millet.

ssGBLUP

A mixed model methodology was adopted for statistical analyses using ssGBLUP [50, 51]. The statistical model was denoted by the following expression: $y = Xm + Za + Zg + Wb + Ti + Tr + Qp + \varepsilon$, where y is the vector of responses across the five cuttings (only using phenotypic data); m is the vector of the effects of the measurement-replication combination (assumed as fixed) added to the overall mean; a is the vector of genetic effects (assumed as random); g is the vector of residual genetic effects (assumed as random); b is the vector of block effects (assumed as random); i is the vector of the genotype by cutting interaction; r is the vector of the residual genotype by cutting interaction; p is the vector of permanent environment effects (random); ε is the vector of residues (random); and X, Z, W, T, and Q represent the incidence matrices for these effects.

The following distributions of random effects were considered: $a \sim N(G \otimes \sigma_a^2)$; $g \sim N(I \otimes \sigma_{rg}^2)$; $b \sim N(I \otimes \sigma_b^2)$; $i \sim N(G_i \otimes \sigma_i^2)$; $r \sim N(I \otimes \sigma_{ri}^2)$; $p \sim N(I \otimes \sigma_p^2)$; and $\varepsilon \sim N(I \otimes \sigma_e^2)$, where G is a matrix of genomic additive relationships, I is an identity matrix of appropriate dimensions, G_i is a matrix of genomic interactions (genotype by cutting interaction), 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 residual variance components, respectively. The model above includes residual genetic effects and residual genotype by cutting interactions, according to Oakey et al. [22].

An additive relationship matrix structure (G) was used according to Resende et al. [24], and is denoted by the expression $G = \frac{Z^* Z^{*'}}{\sum_i^n p_i(1-p_i)}$, where $Z^* = Z - P$, in which 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 that were not genotyped, the inverse of the genomic relationship matrix (H^{-1}) was adopted, according to Legarra et al. [50]. The expression states that $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{*-1} - A_{22}^{-1} \end{bmatrix}$, where A^{-1} is the inverse of the pedigree relationship for all elephant grass genotypes and A_{22}^{-1} is the inverse of the pedigree relationship for only genotyped elephant grass genotypes. A and A_{22} was an identity matrix for this study because there was no information about the pedigree. To compute the exact inverse of G to compose the H^{-1} matrix, we used the algorithm $G^* = 0.95G + 0.05A_{22}$ [52].

From the matrix H^{-1}, the ssGBLUP procedure was run according to the specified model. For the random effects of the model, significance of the likelihood ratio test was evaluated using a chi-square test with one degree of freedom. ssGBLUP was performed using ASReml 4.1 software [53].

The accuracy ($r_{\hat{a}a}$) of the additive effect considering the ssGLUP model was estimated as $r_{\hat{a}a} = \sqrt{1 - \frac{PEV_a}{\hat{\sigma}_a^2}}$, where PEV_a is the predictive error variance that is obtained by diagonal elements inverse of the left-hand side of the mixed model equation for the additive effect.

Simple repeatability plus genotype by cutting (G × C) interaction model

The model was run without considering the relationship (simple repeatability plus G x C interaction model) between the genotypes, as follows: $y = Xm + Zg + Wb + Ti + Qp + \varepsilon$, where y is the vector of response across the five cuttings (using only phenotypic data); m is the vector of the effects of the measurement-replication combination (assumed as fixed) added to the overall mean; g is the vector of genetic effects (assumed as random); b is the vector of block effects (assumed as random); i is the vector of the genotype by cutting interaction; p is the vector of the permanent environment effects (random); ε is the vector of residues (random); and X, Z, W, T, and Q represent the incidence matrices for these effects.

The following distributions of random effects were considered: $g \sim N(I \otimes \sigma_g^2)$; $b \sim N(I \otimes \sigma_b^2)$; $i \sim N(G_i \otimes \sigma_i^2)$; $p \sim N(I \otimes \sigma_p^2)$; and $\varepsilon \sim N(I \otimes \sigma_e^2)$, where I is an identity matrix of appropriate dimensions and σ_g^2, σ_b^2, σ_i^2, σ_p^2, and σ_e^2 are the genetic, block, genotype by cutting interaction, permanent environment, and residual variance components, respectively. The simple repeatability plus G × C interaction model was performed in ASReml 4.1 [53].

The accuracy ($r_{\hat{g}g}$) of the genotype effect (genetic effects) considering the simple repeatability plus G x C interaction model was estimated as $r_{\hat{g}g} = \sqrt{1 - \frac{PEV_g}{\hat{\sigma}_g^2}}$, where PEV_g is the predictive error variance that was obtained by the diagonal elements inverse of the left-hand side of the mixed model equation for the genetic effect.

Genome association study

The 90 elephant grass genotypes that had been genotyped and phenotyped were used in the genome association study. The R package "sommer" (GWAS2 function [54]) revealed a significant association between the markers and the phenotypic traits. The genotype by cutting interaction was included in the association study for each cutting using the model $y = Xm + Mu + Za + Zg + Wb + \varepsilon$, where y is the vector of response for each cutting; m is the vector of the effects of the replication (assumed as fixed) added to the overall mean; u is the vector of markers (assumed as fixed); a is the vector of genetic effects (assumed as random); g is the vector of residual genetic effects (assumed as random); b is the vector of block effects (assumed as random); ε is the vector of residue (random); and X, M, Z, and W represent the incidence matrices for these effects.

The following distributions of random effects were considered: $a \sim N(G \otimes \sigma_a^2)$, $g \sim N(I \otimes \sigma_{rg}^2)$, $b \sim N(I \otimes \sigma_b^2)$, and $\varepsilon \sim N(I \otimes \sigma_e^2)$. In these expressions, G is a matrix of genomic additive relationships, I is an identity matrix of appropriate dimensions, and σ_a^2, σ_{rg}^2, $\sigma_{b,}^2$ and σ_ε^2 are the additive, residual genetic, block, and residual variances, respectively.

Markers with $-\log(P - value)$ up to the false discovery rate (FDR) threshold were considered candidate markers. To compute the FDR, a 0.02 threshold level and $p < 0.05$ were set.

We used the LD.Measures function in the R package LDcorSV to compute linkage disequilibrium (r^2), as described by Hill and Robertson [55]: $r^2 = \frac{[p(AB) - p(A)p(B)]^2}{p(A)p(B)[1-p(A)][1-p(B)]}$, where $p(AB)$ is the frequency of the haplotype AB and $p(A)$ and $p(B)$ are the frequencies of alleles A and B, respectively. Therefore, r^2 ranged from 0 (when the two markers were in perfect equilibrium) to 1

(when the two markers provided identical information). Manhattan plots and QQ plots were obtained using the "sommer" package [54] in R [56].

Alignment, candidate genes, and gene annotation

A list of candidate genes was assembled by BLAST, with default parameters set, using the plant comparative genomics portal Phytozome [57] based on related species (reference genomes; e.g., *S. italica*, *S. viridis*, *P. virgatum*, and *P. halli*). A BLAST search of the *C. americanus* genome was performed on the US National Center for Biotechnology Information website (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For these searches, sequences of the primers M28R (CGAATACGTATGGAGAACTGCG CATC) and M35R (ATCCACCCGACGAAGGAAAC GA) were used. For gene annotation, ortholog genes in *A. thaliana* were searched using the Arabidopsis Information Resource database [58].

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12870-019-2180-5.

Additional file 1: Table S1. Genotypes of elephant grass and their respective codes. The genotypes are part of the Active Elephant grass Germplasm Bank (BAGCE) maintained by Embrapa Gado de Leite (Embrapa Dairy Cattle).

Additional file 2: Table S2. Phenotyping data for the 100 elephant grass genotypes obtained on five cuttings (days 250, 500, 815, 1405 and 1615 after the uniformity cut). Abbreviations used: Cut (cutting day), Gen (genotype - see codes in Table S1), Cut.Rep (concatenation of cut and rep columns), Int (genotype by cutting interaction), Rep (replication), Height (in meters), GB (green biomass, in Mg ha-1), DB (dry biomass, in Mg ha-1), DM (dry matter concentration, in %), ADF (acid detergent fiber, in g Kg-1), NDF (neutral detergent fiber, in g Kg-1), DIG (biomass digestibility, in g Kg-1) and LIG (Lignin content, in g Kg-1).

Additional file 3: Table S3. Molecular data for the 90 elephant grass genotypes. "Code" indicates de code of the genotypes (see Additional file 1: Table S1).

Abbreviations

ABA: Abscisic acid; ADF: Acid detergent fiber; BAGCE: Active Elephant Grass Germplasm Bank; BLAST: Basic local alignment search tool; Cut.Rep: Concatenation of cut and rep columns; DArT: Diversity arrays technology; DB: Dry biomass; DIG: Biomass digestibility; DM: Dry matter concentration; FDR: False discovery rate; G × C: Genotype by cutting interaction; GA: Genome association; GB: Green biomass; Gen: Genotype; Int: Genotype by cutting interaction; LIG: Lignin; NDF: Neutral detergent fiber; NIRS: Near-infrared spectroscopy; PE: Permanent environment; QQ plot: Quantile-quantile plot; Rep: Replication; Res G × C: Residual genotype by cutting interaction; Res G: Residual genetic; SA: Salicylic acid; SNP: Single nucleotide polymorphism; ssGBLUP: Single-step, genome-based best linear unbiased prediction; SSR: Simple sequence repeat

Acknowledgments

We acknowledge financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). We are also thankful to the partnership of Associação para o Fomento à Pesquisa de Melhoramento de Forrageiras (UNIPASTO).

Authors' contributions

PCSC and JCM conceived and designed the experiments; JRASCR, ALSA, JCC and JCM performed the experiments; JRASCR, TSM, FVS, ACS, PCSC, MDVR, ALSA, JFP and JCM analyzed the data; JRASCR, PCSC, ALSA, JFP and JCM wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We also received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and Associação para o Fomento à Pesquisa de Melhoramento de Forrageiras (UNIPASTO). JRASCR and TSM received PhD scholarships and FVS received MSc scholarship from CAPES. ACS received MSc scholarship from FAPEMIG. PCSC and MDVR received research fellowships from CNPq. The funding agencies payed the costs associated with the research (experimental costs and publication fees) but played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and in its supplementary information files. All plant materials used in this article belong to the Active Elephant Grass Germplasm Bank maintained by Embrapa Gado de Leite and are available from the corresponding author on reasonable request and signing of material transfer agreement.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

[1]Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil. [2]Embrapa Florestas, Colombo, PR 83411-000, Brazil. [3]Embrapa Gado de Leite, Juiz de Fora, MG 36038-330, Brazil.

Received: 16 May 2019 Accepted: 1 December 2019
Published online: 10 December 2019

References

1. Negawo AT, Teshome A, Kumar A, Hanson J, Jones CS. Opportunities for Napier grass (*Pennisetum purpureum*) improvement using molecular genetics. Agronomy. 2017;7:28.
2. Fontoura CF, Brandão LE, Gomes LL. Elephant grass biorefineries: towards a cleaner Brazilian energy matrix? J Clean Prod. 2015;96:85–93.
3. Pereira AV, Lédo FJS, Morenz MJF, Leite JLB, Santos AMB, Martins CE, Machado JC. BRS Capiaçu: cultivar de capim-elefante de alto rendimento para produção de silagem. Juiz de Fora: Embrapa Gado de Leite; 2016. p. 1–6.
4. 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;46:704–8.
5. Mambe FT, Voukeng IK, Beng VP, Kuete V. Antibacterial activities of methanol extracts of *Alchornea cordifolia* and four other Cameroonian plants against MDR phenotypes. J Taibah Univ Med Sci. 2016;11:121–7.
6. 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 Design. 2016;89:839–47.
7. Zhou S, Chen J, Laib Y, Yinc G, Chena P, Pennermanc KK, Yand H, Wua B, Zhanga H, Yie X. Integrative analysis of metabolome and transcriptome

reveals anthocyanins biosynthesis regulation in grass species *Pennisetum purpureum*. Ind Crop Prod. 2019;138:111470.
8. Rocha JRASC, Machado JC, Carneiro PCS, Carneiro JC, Resende MDV, Lédo FJS, Carneiro JES. Bioenergetic potential and genetic diversity of elephantgrass *via* morpho-agronomic and biomass quality traits. Ind Crop Prod. 2017;95:485–92.
9. Reis GB, Mesquita AT, Torres GA, Andrade-Vieira LF, Pereira AV, Davide LC. Genomic homeology between *Pennisetum purpureum* and *Pennisetum glaucum* (Poaceae). Comp Cytogenet. 2014;8:199–209.
10. Anderson WF, Casler MD, Baldwin BS. Improvement of perennial forage species as feedstock for bioenergy. In: Vermerris W, editor. Genetic improvements in bioenergy crops. New York: Springer; 2008. p. 309–45.
11. Zhou S, Wang C, Frazier TP, Yan H, Chen P, Chen Z, Huang L, Zhang X, Peng Y, Ma X, et al. The first Illumina-based de novo transcriptome analysis and molecular marker development in Napier grass (*Pennisetum purpureum*). Mol Breeding. 2018;38:95.
12. Paudel D, Kannan B, Yang X, Harris-Shultz K, Thudi M, Varshney RK, Altpeter F, Wang J. Surveying the genome and constructing a high-density genetic map of napiergrass (*Cenchrus purpureus* Schumach). Sci Rep. 2018;8:14419.
13. Wang C, Yan H, Li J, Zhou S, Liu T, Zhang X, Huang L. Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum 'Zise') and identification of its SSR markers. Mol Breeding. 2018;38:94.
14. Pereira JF, Azevedo ALS, Filho MP, Romanel EAC, Pereira AV, Vigna BBZ, Sobrinho FS, Benites FRG, Lédo FJS, Brito GG, et al. Research priorities for next-generation breeding of tropical forages in Brazil. Crop Breed Appl Biotechnol. 2018;18:314–9.
15. Azevedo ALS, Costa PP, Machado JC, Machado MA, Pereira AV, Lédo JFS. Cross-species amplification of *Pennisetum glaucum* microsatellite markers in *Pennisetum purpureum* and genetic diversity of Napier grass accessions. Crop Sci. 2012;52:1776–85.
16. Negawo AT, Jorge A, Hanson J, Teshome A, Muktar MS, Azevedo ALS, Lédo FLS, Machado JC, Jone CS. 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.
17. Qin H, Chen M, Yi X, Bie S, Zhang C, Zhang Y, Lan J, Meng Y, Yuan Y, Jiao C. 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.
18. Gyawali S, Harrington M, Durkin J, Horner K, Parkin IAP, Hegedus DD, Bekkaoui D, Buchwaldt L. Microsatellite markers used for genome-wide association mapping of partial resistance to *Sclerotinia sclerotiorum* in a world collection of *Brassica napus*. Mol Breeding. 2016;36:72.
19. Siraree A, Banerjee N, Kumar S, Khan MS, Singh PK, Kumar S, Sharma S, Singh RK, Singh J. Identification of marker-trait associations for morphological descriptors and yield component traits in sugarcane. Physiol Mol Biol Plants. 2017;23:185–96.
20. Jespersen D, Ma X, Bonos SA, Belanger FC, Raymer P, Huang B. Association of SSR and candidate gene markers with genetic variations in summer heat and drought performance for creeping bentgrass. Crop Sci. 2018;58:2644–56.
21. Biazzi E, Nazzicari N, Pecetti L, Brummer EC, Palmonari A, Tava A, Annicchiarico P. Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. PLoS One. 2017;12: e0169234.
22. Oakey H, Cullis B, Thompson R, Comadran J, Halpin C, Waugh R. Genomic selection in multi-environment crop trials. G3 - Genes Genom Genet. 2016;6: 1313–26.
23. Pauly M, Keegstra K. Biosynthesis of the plant cell wall matrix polysaccharide xyloglucan. Annu Rev Plant Biol. 2016;67:235–59.
24. Resende MDV, Silva FF, 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. Viçosa: Editora UFV; 2014. p. 1–881.
25. 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.
26. 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;109:2749–61.
27. Leng P, Ouzunova M, Landbeck M, Wenzel G, Eder J, Darnhofer B, Lübberstedt T. Quantitative trait loci mapping of forage Stover quality traits in six mapping populations derived from European elite maize germplasm. Plant Breed. 2018;137:139–47.
28. Allouis S, Qi X, Lindup S, Gale M, Devos KM. Construction of a BAC library of pearl millet, *Pennisetum glaucum*. Theor Appl Genet. 2001;102:1200–5.
29. 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;43:2284–90.
30. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. Reference genome sequence of the model plant Setaria. Nat Biotechnol. 2012;30:555–61.
31. 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.
32. Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. Annu Rev Plant Biol. 2003;54:519–46.
33. Liu N, Yu P. Molecular clustering, interrelationships and carbohydrate conformation in hull and seeds among barley cultivars. J Cereal Sci. 2011;53:379–83.
34. Wu Z, Zhang M, Wang L, Tu Y, Zhang J, Xie G, Zou W, Li F, Guo K, Li Q, 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.
35. Lam PY, Tobimatsu Y, Takeda Y, Suzuki S, Yamamura M, Umezawa T, Lo C. Disrupting flavone synthase II alters lignin and improves biomass digestibility. Plant Physiol. 2017;174:972–85.
36. Casler MD, Vogel KP. Accomplishments and impact from breeding for increased forage nutritional value. Crop Sci. 1999;39:12–20.
37. Fraser CM, Chapple C. The phenylpropanoid pathway in Arabidopsis. Arabidopsis Book. 2011;9:e0152.
38. Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M Jr, Chen F, Foston M, Ragauskas A, Bouton J, et al. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci. 2011;108:3803–8.
39. Mielenz JR. Ethanol production from biomass: technology and commercialization status. Curr Opin Microbiol. 2001;4:324–9.
40. Dorez G, Ferry L, Sonnier R, Taguet A, Lopez-Cuesta JM. Effect of cellulose, hemicellulose and lignin contents on pyrolysis and combustion of natural fibers. J Anal Appl Pyrol. 2014;107:323–31.
41. 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.
42. Sato H, Sakamoto S, Mitsuda N, Ohme-Takagi M, Takamizo T. Improvement of cell wall digestibility in tall fescue by *Oryza sativa* SECONDARY WALL NAC DOMAIN PROTEIN2 chimeric repressor. Mol Breeding. 2018;38:36.
43. 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;91:727–40.
44. 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;27:857–73.
45. Goering HK. Forage fiber analysis: Apparatus, reagents, procedures and some applications, Agricultural handbook 379. Washington: U.S. Agricultural Research Service; 1970. p. 1–20.
46. Tilley JMA, Terry RA. A two-stage technique for the *in vitro* digestion of forage crops. Grass Forage Sci. 1963;18:104–11.
47. Viana AP, Resende MDV, Riaz S, Walker MA. Genome selection in fruit breeding: application to table grapes. Sci Agric. 2016;73:142–9.
48. Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, Chantereau J, Gerard B, Ndjeunga J, Bezançon G, 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.
49. Rajaram V, Nepolean T, Senthilvel S, Varshney RK, Vadez V, Srivastava RK, Shah TM, Supriya A, Kumar S, Kumari BR, 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.
50. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. J Dairy Sci. 2009;92:4656–63.
51. 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; 93:743–52.
52. Masuda Y, Misztal I, Tsuruta S, Legarra A, Aguilar I, Lourenco DAL, Fragomeri BO, Lawlor TJ. 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;99:1968–74.
53. Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. ASReml User Guide Release 4.1. UK: VSN International Ltd; 2015.
54. Covarrubias-Pazaran G. Genome-assisted prediction of quantitative traits using the R package *sommer*. PLoS One. 2016;11:e0156744.
55. Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theor Appl Genet. 1968;38:226–31.
56. R Development Core Team R. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2015.
57. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2011;40:D1178–86.
58. 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.

Publisher's Note