



GABRIELLE MARIA ROMEIRO LOMBARDI

**ASSOCIAÇÃO ENTRE CARACTERES E SELEÇÃO DE
HÍBRIDOS DE SORGO BIOMASSA PARA PRODUÇÃO DE
BIOENERGIA**

**LAVRAS – MG
2021**

GABRIELLE MARIA ROMEIRO LOMBARDI

**ASSOCIAÇÃO ENTRE CARACTERES E SELEÇÃO DE HÍBRIDOS DE SORGO
BIOMASSA PARA PRODUÇÃO DE BIOENERGIA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração Genética e Melhoramento de Plantas para obtenção do título de Doutora.

Dr. José Airton Rodrigues Nunes

Orientador

Dr. Rafael Augusto da Costa Parrella

Coorientador

**LAVRAS – MG
2021**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Lombardi, Gabrielle Maria Romeiro.

Associação entre Caracteres e Seleção de Híbridos de Sorgo
Biomassa para Produção de Bioenergia / Gabrielle Maria Romeiro
Lombardi. - 2021.

89 p.

Orientador(a): José Airton Rodrigues Nunes.

Coorientador(a): Rafael Augusto da Costa Parrella.

Tese (doutorado) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Sorghum bicolor L. 2. Análise de Trilha. 3. FAI-BLUP. I.
Nunes, José Airton Rodrigues. II. Parrella, Rafael Augusto da
Costa. III. Título.

GABRIELLE MARIA ROMEIRO LOMBARDI

**ASSOCIAÇÃO ENTRE CARACTERES E SELEÇÃO DE HÍBRIDOS DE SORGO
BIOMASSA PARA PRODUÇÃO DE BIOENERGIA**

**ASSOCIATION BETWEEN TRAITS AND SELECTION OF SORGHUM BIOMASS
GENOTYPES FOR BIOENERGY**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração Genética e Melhoramento de Plantas para obtenção do título de Doutora.

APROVADA em 15 de dezembro de 2021

Dr. José Airton Rodrigues Nunes	UFLA
Dr. Rafael Augusto da Costa Parrella	Embrapa Milho e Sorgo
Profª. Drª. Flávia Maria Avelar Gonçalves	UFLA
Prof. Dr. Tiago Olivoto	UFMS
Prof. Dr. Pedro Crescêncio Souza Carneiro	UFV


Prof. Dr. José Airton Rodrigues Nunes
Orientador

**LAVRAS – MG
2021**

A minha linda filha Laura,
Dedico.

AGRADECIMENTOS

Agradeço à Deus por estar sempre presente e me guiando.

Ao meu marido Ramon pelo apoio, incentivo, aconselhamentos, companheirismo e cumplicidade, amo você.

A minha filha Laura, pela alegria, bondade, generosidade e compreensão. Graças a você continuo caminhando. Você é incrível, te amo.

A minha irmã Nicolay, meu cunhado Rogério e meus sobrinhos, Sophia e Gabriel, pela leveza, alegria, suporte e amizade, vocês são demais.

A minha tia Marilena e ao meu tio Chico, pelo acolhimento, suporte, incentivo e companheirismo, sem vocês essa caminhada seria bem difícil. Obrigada.

A minha avó Liliana por me incentivar no estudo e sempre querer que eu cresça profissionalmente.

Aos meus pais, Carlos e Isabel, que me apoiaram e me incentivaram. Especialmente minha mãe. Muito Obrigada.

A todos os professores do Programa de Pós-graduação em Genética e Melhoramento de Plantas da Universidade Federal de Lavras, pelos conhecimentos transmitidos e principalmente oportunidades recebidas, em especial à professora Flávia Maria e ao professor José Airton o qual além de referência intelectual, tenho como espelho em minha vida, tanto profissional como pessoal. Obrigada professor pelos ensinamentos e pela orientação.

As amigas Fernanda e Risely pelos conselhos, amizade e entretenimento.

Ao pesquisador Rafael Parrella, por gentilmente ter cedido os dados para confecção desta tese, pelas oportunidades oferecidas e pela coorientação. Obrigada Rafael por ser este excelente pesquisador e espelhar o melhor dos que estão a sua volta.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas pela oportunidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pela concessão de bolsa de estudos.

À todos os colegas do Programa de Pós-Graduação em Genética e Melhoramento de Plantas e funcionários do Departamento de Biologia, que contribuíram de forma imprescindível para a minha vida acadêmica.

RESUMO GERAL

O sorgo biomassa, cultura energética promissora para produção de bioenergia, é eficiente tanto para geração de energia elétrica quanto para produção de etanol de segunda geração (2G). A eficiência da seleção de genótipos superiores desta cultura depende da finalidade industrial e do conhecimento do inter-relacionamento dos vários caracteres que de forma direta ou indireta influenciam na expressão fenotípica. Além disso, a seleção de genótipos envolvendo múltiplos caracteres mediante o emprego de índices adequados como o FAI-BLUP pode contribuir para o melhor direcionamento em programas de melhoramento da cultura para a seleção do genótipo ideal. Dessa forma, o objetivo com este trabalho foi, no primeiro capítulo, estimar as associações entre caracteres do sorgo biomassa; estimar os efeitos diretos e indiretos e verificar a existência de relação causa e efeito dos demais caracteres com o poder calorífero teórico (PCT) via análise de trilha; e fornecer uma técnica complementar para análise dos caracteres e direcionamento dos programas de melhoramento de sorgo biomassa por meio da Mineração de Dados. O PCT apresentou tanto associações nulas quanto positivas com os caracteres avaliados. Destacando-se as associações positivas com florescimento (FLOR); altura de planta (AP); produtividade de biomassa verde do colmo (PMVc); massa seca do colmo (MSc) e foliar (MSf); fibra em detergente neutro e ácido foliar (FDNf e FDAf); lignina do colmo (LIGc) e foliar (LIGf); e hemicelulose foliar (HEMf). Os caracteres FLOR, AP, PMVc, MSf e LIGc foram considerados os principais determinantes das alterações do PCT, podendo ser utilizados na seleção indireta deste caráter. Além disso, os caracteres FLOR, AP e PMVc influenciaram direta ou indiretamente a expressão dos caracteres MSf e LIGc, podendo ser utilizados na seleção indireta desses caracteres. A técnica de mineração de dados, árvore de regressão, permitiu melhor visualização das relações entre os caracteres, auxiliando em algumas explicações sobre a influência dos caracteres sobre o PCT. No segundo capítulo, o objetivo foi identificar e selecionar genótipos superiores de sorgo biomassa para a produção de etanol de segunda geração considerando a seleção para múltiplos caracteres, bem como, verificar qual estratégia seria a mais adequada para seleção desses genótipos pelo índice FAI-BLUP considerando experimentos em múltiplos locais. As estratégias empregadas para calcular o índice FAI-BLUP foram: primeira estratégia (E1) uso dos valores genéticos preditos em cada local; segunda estratégia (E2) uso dos valores genéticos preditos livres de interação considerando todos os locais; terceira estratégia (E3) uso dos valores genéticos preditos capitalizando a interação GxE; quarta e quinta estratégia (E4;E5) houve o uso dos valores genéticos preditos na E1 e na E3, respectivamente e criação de novos caracteres por local. Os genótipos 4, 13, 17 e 18 destacaram-se quando a superioridade, previsibilidade e adaptabilidade genotípica, sendo interessante o uso destes visando à produção de etanol de segunda geração. O uso da E2 e E3 foram mais adequadas para seleção de genótipos superiores pelo índice FAI-BLUP considerando experimentos em múltiplos locais e visando a recomendação, respectivamente, para locais que apresentam e/ou não o mesmo padrão de interação GxE da rede experimental avaliada.

Palavras-chave: *Sorghum bicolor* L., análise de trilha, correlação, bioenergia, árvore de decisão, FAI/BLUP.

GENERAL ABSTRACT

Biomass sorghum, a promising energy crop for bioenergy production, is efficient for both electricity generation and second generation (2G) ethanol production. The efficiency of the selection of superior sorghum biomass genotypes depends on the industrial purpose and the knowledge of the interrelationship of the several traits that directly or indirectly influence the phenotypic expression. In addition, the selection of genotypes involving multiple traits using suitable indices such as FAI-BLUP can contribute to better targeting in crop improvement programs for genotype selection based on an ideal genotype. Thus, the objective of this study was, in the first chapter, to estimate the associations between sorghum biomass traits; estimate the direct and indirect effects and verify the existence of cause-and-effect relationship of the other traits with the theoretical heating power (THP) via path analysis; and provide a complementary technique for multi-trait analysis and targeting sorghum biomass improvement programs through Data Mining. The THP showed both null and positive associations with the evaluated traits. Highlighting the positive associations with flowering (FLOW); plant height (PH); green stem biomass productivity (GSBP); stem (SDM) and leaf (LDM) dry mass; leaf neutral detergent fiber (LNDF); leaf acid detergent fiber (LADF); stem lignin (SL); leaf lignin (LL); and leaf hemicellulose (LH). The traits FLOW, PH, GSBP, LDM and SL were considered the main determinants of THP variation, and can be used in the indirect selection of this target trait. In addition, the traits FLOW, PH and GSBP directly or indirectly influenced the expression of the traits LDM and SL, and can be used in the indirect selection of these traits. The data mining technique, regression tree, allowed a better visualization of the relationships between the traits, helping in some explanations about the influence of the traits on the THP. In the second chapter, the objective was to identify and select superior sorghum biomass genotypes to produce second-generation ethanol considering the selection for multiple traits, as well as to verify which strategy would be the most appropriate for the selection of these genotypes by the FAI-BLUP index considering experiments in multiple locations. The strategies used to calculate the FAI-BLUP index were: first strategy (E1) use of the genetic values predicted in each location; second strategy (E2) use of predicted genetic values free of genotype by environment (GxE) interaction considering all locations; third strategy (E3) use of predicted genetic values capitalizing on the GxE interaction; fourth strategy (E4) use of genetic values predicted in E1 and creating new traits by location; fifth strategy (E5) using predicted values in E3 and creating new traits per location. Genotypes 4, 13, 17 and 18 stood out in terms of superiority, predictability and genotypic adaptability, and their use is interesting for second-generation ethanol production. The use of E2 and E3 were the most suitable for the selection of superior genotypes by the FAI-BLUP index considering experiments in multiple locations and aiming at the recommendation, respectively, for locations that present and/or not the same pattern of GxE interaction of the evaluated experimental network.

Keywords: Sorghum bicolor L., path analysis, correlation, bioenergy, decision tree, FAI/BLUP.

Sumário

	CAPÍTULO 1 – INTRODUÇÃO GERAL.....	10
1	INTRODUÇÃO.....	10
2	REFERENCIAL TEÓRICO.....	11
2.1	Sorgo Biomassa.....	11
2.2	Correlação entre caracteres e análise de trilha em sorgo biomassa.....	16
2.3	Mineração de Dados.....	20
2.4	Seleção para múltiplos caracteres.....	23
	REFERÊNCIAS.....	25
	CAPÍTULO 2 – ASSOCIAÇÃO ENTRE CARACTERES DE SORGO BIOMASSA USANDO TÉCNICAS DE ANÁLISE DE TRILHA E ÁRVORE DE REGRESSÃO COM APRENDIZADO DE MÁQUINA.....	34
1	INTRODUÇÃO.....	37
2	MATERIAL E MÉTODOS.....	40
2.1	Local, Plano Experimental e Manejo.....	40
2.2	Caracteres mensurados.....	40
2.3	Análise Estatística.....	41
3	RESULTADOS.....	43
4	DISCUSSÃO.....	49
5	CONCLUSÃO.....	55
	REFERÊNCIAS.....	57
	CAPÍTULO 3 – ESTRATÉGIAS DE SELEÇÃO DE GENÓTIPOS DE SORGO BIOMASSA EM ENSAIOS MULTI-AMBIENTAIS COM BASE EM MÚLTIPLOS CARACTERES.....	63
1	INTRODUÇÃO.....	66
2	MATERIAL E MÉTODOS.....	67
2.1	Local experimental, plano experimental e manejo.....	67
2.2	Caracteres mensurados.....	68
2.3	Análise estatística.....	69
3	RESULTADOS.....	71
4	DISCUSSÃO.....	76
5	CONCLUSÃO.....	82
	REFERÊNCIAS.....	83

CAPÍTULO 1 – INTRODUÇÃO GERAL

1 INTRODUÇÃO

As crescentes preocupações com a segurança energética e os impactos ambientais com o uso de combustíveis fósseis, aliado à necessidade de diversificar a matriz energética mundial, expandiram a demanda por fontes renováveis de energia (SCOTT et al., 2015). O sorgo biomassa [*Sorghum bicolor* (L). Moench], cultura energética promissora para produção de bioenergia (ABBAS et al., 2020; SIMEONE et al., 2020), apresenta ciclo curto de, aproximadamente, 180 dias; altura elevada, sendo superior a 4 metros; alta produtividade de biomassa verde e seca, em torno de 150 t ha⁻¹ e 50 t ha⁻¹, respectivamente (HEITMAN et al., 2017; SIMEONE et al., 2020); ampla adaptabilidade para regiões tropicais e temperadas (STEPHENSON et al., 2021; LIANG et al., 2021); e alto poder calorífico em caldeiras (CASTRO et al., 2015; DA SILVA et al., 2018; OLIVEIRA et al., 2019). Esta cultura é considerada eficiente tanto para geração de energia elétrica quanto para produção de etanol de segunda geração (2G) (CASTRO et al., 2017; SILVA et al., 2020; SIMEONE et al., 2020).

Os caracteres morfoagronômicos e os que compõem a biomassa vegetal do sorgo biomassa, mais especificamente os compostos lignocelulósicos (celulose, hemicelulose e lignina), podem influenciar positivamente ou negativamente na obtenção do genótipo ideal para cada finalidade (SIMS et al., 2010). Quanto ao teor de lignina, por exemplo, busca-se genótipos com maiores teores para incrementar o poder calorífico da cultura visando a geração de energia elétrica (XIAO et al., 2001; MAY et al., 2015; CASTRO et al., 2015; OLIVEIRA et al., 2019) e genótipos com menores teores para melhorar a hidrólise enzimática e o processo fermentativo visando a produção de etanol 2G (DAMASCENO et al., 2010; VAN DER WEIJDE et al., 2016). Assim, a obtenção do ideótipo de sorgo biomassa depende da finalidade industrial (CASTRO et al., 2017; SIMEONE et al., 2020) e envolve vários caracteres simultaneamente (VENDRUSCOLO et al., 2016; HABYARIMANA et al., 2020). Esses caracteres apresentam grande variação quanto a direção e a magnitude das associações (CASTRO et al., 2015; VENDRUSCOLO et al., 2016; ALMEIDA et al., 2019; DOS SANTOS et al., 2020; HABYARIMANA et al., 2020).

Associações positivas interessantes para fins de melhoramento da cultura foram observadas entre a altura de planta e a produtividade de biomassa seca (FERNANDES et al., 2018; DOS SANTOS et al., 2020) e entre a altura de planta e o teor de lignina (OLIVEIRA et al., 2019). Este conhecimento, permite a utilização do caráter altura para selecionar genótipos

de sorgo biomassa com maior produtividade de biomassa seca e com maior teor de lignina de maneira menos trabalhosa e onerosa, sendo extremamente interessante para obtenção de genótipos de sorgo biomassa visando a geração de energia elétrica. No entanto, até o presente momento não foi observado relação de causa e efeito das associações desses caracteres, indicando que a altura, possivelmente, não é o principal determinante das alterações desses caracteres (VENDRUSCOLO et al., 2016).

O envolvimento de vários caracteres simultaneamente na seleção do genótipo ideal e a ausência de conhecimento da relação entre causa e efeito das associações entre esses caracteres impactam no direcionamento dos programas de melhoramento da cultura para geração de energia elétrica e para produção de etanol 2G (VENDRUSCOLO et al., 2016; OLIVOTO et al., 2017). Além disso, considerando a obtenção do genótipo ideal para produção de etanol 2G, as associações positivas observadas entre a altura e o teor de lignina, por exemplo, apresentam causas indesejáveis.

Dessa forma, o objetivo com este trabalho foi estimar as associações entre caracteres do sorgo biomassa; estimar os efeitos diretos e indiretos e verificar a existência de relação causa e efeito dos caracteres com o poder calorífero teórico (PCT) via análise de trilha; e fornecer uma técnica complementar para análise dos caracteres e direcionamento dos programas de melhoramento em sorgo biomassa para geração de energia elétrica por meio da Mineração de Dados e, identificar e selecionar genótipos superiores de sorgo biomassa para a produção de etanol de segunda geração considerando a seleção para múltiplos caracteres, bem como, verificar qual estratégia seria a mais adequada para seleção desses genótipos pelo índice FAI-BLUP considerando experimentos em múltiplos locais.

2 REFERENCIAL TEÓRICO

2.1 Sorgo Biomassa

O sorgo [*Sorghum bicolor* (L). Moench] é uma espécie pertencente ao Reino Plantae; Divisão Magnoliophyta (Angiospermas); Classe Liliopsida (Monocotiledonea); Ordem Poales; Família Poaceae (Gramíneas), gênero *Sorghum* (SANTOS et al., 2005). A África Oriental é considerada como centro de origem desta cultura, embora algumas evidências indiquem que possa ter havido duas regiões de dispersão independentes: África e Índia, onde se encontram atualmente, a maior variabilidade em espécies silvestres e cultivadas da cultura (RIBAS, 2008; REGASSA et al., 2014; BORÉM et al., 2014).

O sorgo é uma espécie diploide com 10 pares de cromossomos ($2n = 2x = 20$) (SANTOS et al., 2005). Apresenta flores andróginas, sendo basicamente uma espécie autógama, com pequena taxa de fecundação cruzada a qual varia conforme o genótipo e o ambiente (em torno de 2 a 10%) e, quando ocorre, é realizada pelo vento (SMITH; FREDERIKSEN, 2005). Devido a isso, a autofecundação foi à base nos programas de desenvolvimento de variedades, que possibilitou o estabelecimento da cultura, antes da era do híbrido. Atualmente, essas linhagens atuam como base nos programas de desenvolvimento de híbridos (SMITH; FREDERIKSEN, 2005), visto que os cruzamentos entre as linhagens, frequentemente apresentam vigor híbrido (SMITH; FREDERIKSEN, 2005; PFEIFFER et al., 2010; BUNPHAN et al., 2015; LOMBARDI et al., 2018, HE et al., 2020). Os híbridos de sorgo são produzidos pelo cruzamento entre uma linhagem macho estéril (linhagem A) e uma linhagem fértil polinizadora (linhagem R). A macho esterilidade do sorgo é genético-citoplasmática, a qual envolve genes mitocondriais, herdados maternalmente, e genes restauradores de fertilidade de natureza nuclear, constituindo um sistema binário.

O sorgo pode ser melhorado para diversas finalidades (ROONEY, 2004) devido sua ampla diversidade genética. Atualmente, o sorgo é classificado em cinco grupos: i) sorgo granífero, cujo produto principal é o grão que é utilizado na alimentação humana e animal; ii) sorgo forrageiro, cujo produto principal é a biomassa a qual é utilizada para produção de silagem e/ou etanol; iii) sorgo vassoura, cujo produto principal é a panícula que é utilizada para confecção da “vassoura de sorgo” ou vassoura caipira; iv) sorgo sacarino, cujo produto principal é o teor de açúcares fermentáveis que é utilizado para produção de etanol; e, por fim, o v) sorgo biomassa, cujo produto principal é a biomassa a qual é utilizada para produção de etanol de segunda geração (etanol 2G ou etanol lignocelulósico), geração de energia elétrica, produção de gás combustível, entre outros (EMBRAPA, 2014; BORÉM et al., 2014).

O sorgo biomassa é uma planta anual com alto potencial para a produção de bioenergia (SILVA et al., 2020) por possuir ciclo curto (150-180 dias), alta produção de biomassa verde (até 150 t ha^{-1}) e seca (até 50 t ha^{-1}) (PARRELA et al., 2011; PACKER; ROONEY, 2014), alto poder calorífico em caldeiras (até $4.500 \text{ Kcal kg}^{-1}$ de matéria seca) (MAY et al., 2014), ampla adaptabilidade para regiões tropicais e temperadas (VERMERRIS; SABALLOS, 2013; SILVA et al., 2018; OLIVEIRA et al., 2019) e grande variabilidade quanto ao teor de lignina na planta (entre 1 e 25%), podendo ser utilizado, dependendo do teor de lignina, tanto para geração de energia elétrica quanto para produção de etanol 2G (ZHAO et al., 2009; SKONIESKI et al., 2010; PARRELLA et al., 2011; DAMASCENO et al., 2013).

A alta produção de biomassa deve-se, principalmente, ao cultivo de genótipos de sorgo biomassa sensíveis ao fotoperíodo. A sensibilidade ao fotoperíodo possibilita a ampliação do ciclo vegetativo e, conseqüentemente, incremento da produção de biomassa por hectare/ciclo (PEREIRA et al., 2012). O fotoperiodismo, resposta do desenvolvimento da planta à duração dos períodos de luz e de escuro, afeta a indução floral e o crescimento final da cultura. Os genótipos de sorgo biomassa respondem de forma diferente à duração do período luminoso diário, sendo classificados como sensíveis ou insensíveis ao fotoperíodo. O controle da sensibilidade ao fotoperiodismo e da maturação (indução de floração) em sorgo está associado ao efeito de seis genes com dois alelos: Ma1,ma1; Ma2,ma2; Ma3,ma3; Ma4,ma4; Ma5,ma5; e Ma6,ma6 (ROONEY; AYDIN, 1999). Os genes Ma1 a Ma4 controlam o ciclo, enquanto os genes Ma5 e Ma6 são responsáveis pela sensibilidade ao fotoperiodismo. Os genótipos Ma5Ma5ma6ma6 e ma5ma5Ma6Ma6 são insensíveis ao fotoperíodo e florescem entre 60 e 70 dias após a germinação, enquanto genótipos derivados de cruzamentos entre estas linhagens são sensíveis (Ma5ma5Ma6ma6), florescendo entre 150 e 180 dias. Contudo, genótipos com constituição genética Ma1__ são sensíveis ao fotoperíodo, independentemente dos demais genes e, genótipos com constituição genética ma1ma1 são insensíveis ao fotoperíodo (QUINBY; SCHERTZ, 1970). Portanto, para a obtenção de híbridos sensíveis, busca-se selecionar genótipos com alelo Ma1 (PARRELLA et al., 2011).

O sorgo sensível é uma planta de dias curtos, ou seja, floresce apenas quando o fotoperíodo é menor que um fotoperíodo crítico. Assim sendo, se o comprimento do dia aumenta, a gema apical permanece vegetativa (a planta não floresce), ao passo que se o comprimento do dia decresce, tornando-se menor que 12 horas e 20 minutos, haverá diferenciação da gema apical em gema floral (a planta floresce). Desta forma, genótipos de sorgo biomassa, sensíveis ao fotoperíodo, semeados nos meses de setembro ou outubro em regiões com fotoperíodo maior que 12 horas e 20 minutos, irão iniciar o desenvolvimento da gema floral a partir de março do ano seguinte, ampliando o ciclo vegetativo e, concomitantemente, possibilitando maior produção de biomassa por hectare/ciclo em comparação a cultivares insensíveis ao fotoperíodo.

Além da elevada produção de biomassa, a qualidade da biomassa é extremamente importante, para obtenção de genótipos de sorgo biomassa com alto potencial para a geração de bioenergia (DAMASCENO et al., 2010; SILVA et al., 2020). Como comentado anteriormente, a biomassa dessa cultura pode ser utilizada tanto para produção de etanol 2G (AMADUCCI et al., 2000; DAVILA-GOMEZ et al., 2011; SILVA et al., 2020) quanto para geração de energia elétrica (BENNETT; ANEX, 2009; ZEGADA-LIZARAZU; MONTI,

2012). A umidade e composição da biomassa vegetal, mais especificamente dos compostos lignocelulósicos (celulose, hemicelulose e lignina), podem influenciar positivamente ou negativamente na obtenção do genótipo ideal para cada finalidade (SIMS et al., 2010).

A lignina, por exemplo, fornece aos tecidos vegetais, impermeabilização, reforço estrutural e resistência ao ataque tanto biológico como físico e químico (XIAO; SUN; SUN, 2001). O teor ideal deste componente dependerá da finalidade do uso do genótipo, sendo desejável genótipos com maiores teores de lignina visando a geração de energia elétrica (MAY et al., 2015; CASTRO et al., 2015) e genótipos com menores teores de lignina visando a produção de etanol 2G (DAMASCENO et al., 2010; VAN DER WEIJDE et al., 2016).

Na geração de energia elétrica que é realizada por meio da combustão direta da biomassa, transformando energia térmica em elétrica (ZEGADA-LIZARAZU; MONTI, 2012), quanto maior a energia liberada por massa do sólido quando queimado totalmente na presença de oxigênio maior será o poder calorífero. Como a lignina apresenta elevado poder calorífico (6100 Kcal kg⁻¹) genótipos com elevados teores de lignina são desejáveis. Isto ocorre em função da lignina possuir alta resistência a degradação térmica, devido ao elevado número de carbonos presentes em sua estrutura e baixo teor de oxigênio, se comparado a celulose e hemicelulose (PETROFF; DOAT, 1978; PEREIRA et al., 2012; MAY et al., 2015; CASTRO et al., 2015).

O sorgo biomassa possui elevado poder calorífico, em torno de 4500 Kcal kg⁻¹ de biomassa seca (CASTRO et al., 2015; DA SILVA et al., 2018), se comparado ao eucalipto (4501 Kcal kg⁻¹), cana-de-açúcar (4274 Kcal kg⁻¹) (PAULA et al., 2011), switchgrass (*Panicum virgatum*, L.) (4204 Kcal kg⁻¹) (SAINI et al., 2017) e capim elefante (*Pennisetum purpureum*, S.) (4200 Kcal kg⁻¹) (BORGES et al., 2016). Além disso, o sorgo biomassa pode atingir 50% de umidade em campo (PARRELLA et al., 2013), sendo compatível com as necessidades da maioria das caldeiras. A umidade do material interfere diretamente na temperatura de combustão, podendo, quando em elevada umidade, dificultar a queima da biomassa (WU et al., 2011).

Na produção de etanol de 2G, antes da conversão de açúcares simples em biocombustíveis pela fermentação, é realizada a despolimerização dos polissacarídeos da parede celular em açúcares simples e fermentáveis, pelos processos de pré-tratamento e sacarificação (LI et al., 2007). O pré-tratamento, que pode ser químico, físico ou enzimático, visa aumentar o acesso das enzimas aos polissacarídeos que posteriormente serão fragmentados em monômeros fermentescíveis (LIU et al., 2017) e reduzir a interação entre os carboidratos da parede celular e a lignina. É desejável menor interação entre os carboidratos

da parede celular e da lignina, pois a lignina atua negativamente na hidrólise enzimática por três principais fatores: a adsorção competitiva, a inibição química e o impedimento estérico (LI et al., 2007; KE; CHEN, 2013), podendo inibir o processo fermentativo subsequente (DAMASCENO et al., 2010; VAN DER WEIJDE et al., 2016) e encarecer a produção de etanol (SIMS et al., 2010). Com isso, genótipos de sorgo biomassa com menores teores de lignina são desejáveis visando a produção de etanol 2G.

A mutação *bmr* (brown midrib) é uma das formas de se alterar a qualidade e a quantidade de lignina nos genótipos de sorgo biomassa de forma a favorecer o processo de produção de etanol (CHERNEY et al., 1991). A mutação *bmr* foi primeiro observada em 1924 em Saint Paul, Minnesota, Estados Unidos e ocorreu de forma espontânea em uma planta de milho (*Zea mays* L.) (JORGENSEN, 1931). Os genes responsáveis pela mutação *bmr* reduzem a atividade de enzimas envolvidas no processo de biossíntese de lignina (HALPIN et al., 1998; SATTLER et al., 2009) acarretando menor teor de lignina nas plantas mutantes em relação a seus pares isogênicos, além da formação de tecido vascular marrom-avermelhado nas folhas e nos colmos (PORTER et al., 1978; SATTLER et al., 2010). A obtenção de genótipos de sorgo *bmr* ocorreu, segundo a literatura, por meio da indução química de sementes com dietil sulfato, gerando dezenove mutantes *bmr* de ocorrência independente, identificados em progênies segregadas. Dentre os dezenove mutantes, foram selecionados três de melhores características agrônômicas o *bmr6*, *bmr12* e *bmr18* (BOUT; VERMERRIS, 2003). O *bmr6* provoca redução da atividade da enzima cinamil álcool desidrogenase (CAD), enquanto os *bmr12* e *bmr18* diminuem a atividade da enzima cafeato O-metiltransferase (COMT) na síntese de lignina da planta de sorgo (OLIVER et al., 2005).

Genótipos de sorgo biomassa *bmr* (*bmr6* e *bmr12*) apresentaram eficiência na conversão de biomassa, promovendo maior liberação de açúcares fermentáveis após sacarificação enzimática da biomassa quando comparadas com os não mutantes correspondentes (SABALLOS et al., 2008; DIEN et al., 2009)

Dessa forma, a obtenção do genótipo ideal de sorgo biomassa depende da finalidade industrial, geração de energia elétrica ou produção de etanol 2G, e envolve vários caracteres simultaneamente. Logo, o estudo desses caracteres, suas associações e a seleção de genótipos considerando-os simultaneamente são de extrema importância, visto que podem ter efeito positivo ou negativo na seleção.

2.2 Correlação entre caracteres e análise de trilha em sorgo biomassa

A correlação é um parâmetro estatístico que mede o grau de associação linear entre duas variáveis ou dois caracteres. O grau de associação é medido por meio do coeficiente de correlação, o qual é adimensional e seu valor absoluto não supera a unidade podendo assumir valores no intervalo $[-1, 1]$. Quando a correlação é positiva, os caracteres variam na mesma direção; quando é negativa, os caracteres variam em direção oposta e, quando a correlação é nula os caracteres não possuem associação linear (RAMALHO et al., 2012).

A correlação que pode ser diretamente mensurada a partir de medidas de dois caracteres, é a fenotípica. Essa correlação tem causas genéticas e ambientais, porém só as genéticas envolvem uma associação de natureza herdável, podendo ser utilizada na orientação de programas de melhoramento (CRUZ et al., 2012). Tem-se que a ocorrência da correlação genética é devido a pleiotropia e a ligações gênicas, contudo, a causa da correlação genética é, principalmente, devido a pleiotropia. As ligações gênicas são causas transitórias, especialmente em populações derivadas de cruzamento entre linhagens divergentes (FALCONER; MACKAY, 1996).

Nos programas de melhoramento, se dois caracteres apresentam correlação genética positiva, é possível obter ganhos para um deles por meio da seleção indireta no outro caráter associado. Em alguns casos, a seleção indireta, com base na resposta correlacionada, pode levar a progressos mais rápidos do que a seleção direta do caráter desejado, principalmente, quando este sofre pronunciado efeito do ambiente e, assim, apresenta baixa herdabilidade (RAMALHO et al., 1993; CRUZ et al., 2012). Entretanto, se um caráter correlacionasse negativamente com alguns e positivamente com outros, deve-se tomar o cuidado de, ao selecioná-lo, não provocar mudanças indesejáveis em outros.

Em sorgo biomassa, observa-se alta correlação entre a produção de biomassa verde e seca, altura da planta, número de colmos (VENDRUSCOLO et al., 2016; CASTRO et al., 2015; PERAZZO et al., 2014) e dias para o florescimento (VENDRUSCOLO et al., 2016; CASTRO et al., 2015). A produção de biomassa no sorgo apresenta baixa herdabilidade (DOS SANTOS et al., 2020), sendo mais oneroso e trabalhoso a obtenção de genótipos superiores considerando este caráter quando comparado com a seleção de genótipos visando elevada produção de biomassa baseados em caracteres de rápida mensuração, como a altura. Isto acontece pois a altura de planta é um caráter que apresenta elevada herdabilidade (CASTRO et al., 2015; BURKS et al., 2015; PATEL et al., 2019) e correlação positiva e de magnitude

elevada com a produção de biomassa desde os trinta dias após o plantio (FERNANDES et al., 2018).

Contudo, apesar de muito útil no melhoramento de plantas, os coeficientes de correlação não fornecem informações sobre o efeito direto e indireto dos caracteres, isto é, as informações de correlações são restritas a um par de caracteres, podendo não representar a verdadeira associação entre dois caracteres, uma vez que pode ocorrer influência de um terceiro caráter, ou um conjunto de caracteres, sobre uma estimativa de correlação (WRIGHT, 1921). Este fato, foi observado para o caráter altura de planta, no sorgo biomassa, ao realizar o estudo da análise de trilha, visto que apesar da altura de planta apresentar alta correlação com a produção de biomassa, ela não influenciou direta ou indiretamente este caráter (VENDRUSCOLO et al., 2016). Possivelmente, a altura de planta sofreu influência de um terceiro caráter ou grupo de caracteres.

Nesse sentido, o estudo da análise de trilha, proposta por Wright (1921), surge como alternativa interessante à correlação, pois esta realiza o desdobramento do coeficiente de correlação em efeitos diretos e indiretos de um conjunto de caracteres sobre um caráter básico ou principal permitindo assim o estudo mais detalhado das associações num conjunto de caracteres. As estimativas da análise de trilha são obtidas por meio de equações de regressão, em que os caracteres são previamente padronizados. Apesar da correlação ser uma característica intrínseca a dois caracteres em dada condição experimental, sua decomposição é dependente do conjunto de caracteres estudados, que normalmente são avaliados pelo conhecimento do melhorista de suas importâncias e de possíveis inter-relações expressas em um modelo de relacionamento de causa e efeito, conhecido como “diagrama causal” (ou “diagrama de trilha”) (CRUZ et al., 2012).

O diagrama causal pode considerar um único modelo causal (diagrama de uma cadeia) ou mais de um modelo causal, como o diagrama de duas cadeias. A quantidade de cadeias será determinada pelo melhorista de acordo com o conhecimento do mesmo em relação a cultura e caracteres que estão sendo avaliados. A análise de trilha, nos casos em que se considera um único modelo causal, é simplesmente uma análise de regressão parcial padronizada, sendo útil no desdobramento dos coeficientes de correlação em efeitos diretos e indiretos. Contudo, nos casos em que se considera mais de um modelo causal, como os diagramas de duas cadeias, a análise de trilha constitui-se em uma expansão da regressão múltipla, o que é comumente encontrado em estudos de melhoramento envolvendo a produtividade da cultura, os componentes primários da produtividade e também os componentes secundários (CRUZ et al., 2012).

Após estabelecido do diagrama causal, as correlações observadas são decompostas em conjunto de termos denominados de coeficientes de trilha, os quais apresentam os caminhamentos simples e complexos (JOHNSON; WICHERN, 2007). O coeficiente de trilha pode ser estimado pelo método dos quadrados mínimos e tem as seguintes propriedades associadas a ele: (1) sendo um coeficiente de regressão, ele tem direção, podendo ser negativo ou positivo e maior ou menor quantidade; (2) sendo um coeficiente de regressão padronizado, ele pode ser utilizado para comparar efeito de caracteres mensuráveis em diferentes escalas; e (3) não tendo unidade física, ele se assemelha a um coeficiente de correlação (CRUZ et al., 2012).

Os resultados dos coeficientes de trilha podem auxiliar na tomada de decisão da estratégia de seleção mais adequada quando: (i) o efeito direto do carácter causal (X) sobre o carácter principal (Y) for semelhante, em grandeza e sinal, à correlação observada entre X e Y, tem-se que a resposta correlacionada por meio da seleção indireta do carácter principal é eficiente, logo o melhorista pode praticar a seleção indireta; (ii) a correlação entre X e Y for positiva e significativa, mas X tiver pequeno efeito direto positivo ou mesmo negativo sobre o carácter principal, o melhorista deve dar maior atenção aos efeitos indiretos dos demais caracteres, visto que este fato indica ausência de causa e efeito, isto é, o carácter X não é o principal determinante das alterações no carácter principal, existindo outros caracteres que poderão proporcionar maior impacto em termos de ganho de seleção; (iii) a correlação entre X e Y for negligenciável, mas o efeito direto for positivo e elevado, os efeitos indiretos é que devem ser considerados no processo seletivo; (iv) a correlação entre X e Y for negativa, porém com efeito direto positivo e elevado, é recomendável utilizar um esquema seletivo restrito a fim de eliminar os efeitos indiretos indesejáveis para aproveitar o efeito direto existente (VENCOVSKY; BARRIGA, 1992; CRUZ et al., 2012).

A análise de trilha apresenta algumas limitações. A confiabilidade dos coeficientes de trilha pode ser adversamente afetada pelos efeitos de multicolinearidade (ou intercorrelação) existente entre os caracteres que compõem o diagrama causal, isto é, o coeficiente de trilha pode ser enviesado caso os caracteres causais sejam correlacionados entre si no modelo de regressão (CRUZ et al., 2014). Além disso, a ausência de altas correlações não indica ausência de multicolinearidade (CARVALHO; CRUZ, 1996).

A multicolinearidade aumenta os erros padrão dos coeficientes e pode acarretar: (i) grandes alterações nas estimativas dos coeficientes de regressão quando se exclui ou inclui um ou poucos caracteres, (ii) grandes amplitudes para intervalos de confiança correspondentes aos coeficientes de regressão que representam importantes caracteres causais,

(iii) coeficientes de regressão que não apresentam o sinal algébrico esperado, com base em resultados já obtidos ou com base no conhecimento prévio que o pesquisador tem a respeito do assunto estudado (CRUZ et al., 2014).

Há diversos métodos para diagnosticar e contornar a presença de multicolinearidade, entre eles tem-se a análise dos autovalores associados à matriz de correlação, ou teste do número de condições da matriz (NC), proposto por Montgomery e Peck (1981), que consiste na razão entre o maior e menor autovalor da matriz de correlação. A multicolinearidade é considerada fraca quando NC é menor que 100, moderada a forte se o NC estiver no intervalos de $[100 < NC < 1000]$ e severa se $NC > 1000$. Se a multicolinearidade for considerada fraca, ela não é um problema sério, não necessitando assim ser contornada. Contudo, se for considerada moderada a forte ou severa é necessário contorná-la. A multicolinearidade pode ser contornada por meio da eliminação de caracteres causais do modelo ou pelo emprego de uma metodologia alternativa para a estimação dos parâmetros, denominada análise de trilha em crista (CARVALHO, 1995; CRUZ et al., 2014).

O método de análise de trilha em crista adiciona uma constante positiva k ($0 < k < 1$) na diagonal principal da matriz de correlações dos caracteres explicativos, visando obter soluções para o vetor de parâmetros B . No entanto, na presença de multicolinearidade severa, as estimativas dos coeficientes de análise de trilha em crista mostram-se instáveis para os valores iniciais de k . À medida que k aumenta, essas estimativas tendem a se estabilizar, proporcionando estimativas confiáveis do parâmetro B . Boa estabilização ocorre com maiores valores de k , entretanto, quanto maior o seu valor, maior será o viés da análise de trilha (CRUZ et al., 2014).

A análise de trilha, apesar das limitações, tem sido bastante utilizada no melhoramento de diversas culturas, inclusive com sorgo, como sorgo sacarino (LOMBARDI et al., 2015), sorgo granífero (MENGISTU et al., 2020), sorgo forrageiro (GOSWAMI et al., 2020) e sorgo biomassa (VENDRUSCOLO et al., 2016), principalmente, por possibilitar o conhecimento da influência de um carácter qualquer sobre o carácter principal e, assim, estabelecer a melhor estratégia de seleção para incrementar o melhoramento genético.

As análises de correlação e de trilha são comumente utilizadas pelos programas de melhoramento genético, principalmente, para entender a associação entre os caracteres e a relação entre causa e efeito deles. Técnicas que podem complementar essas análises surgiram na área de ciência da computação como, por exemplo, a técnica de árvore de decisão da etapa de mineração de dados, sendo por vezes mais robustas e de fácil interpretação.

2.3 Mineração de Dados

A Mineração de Dados (MD) é uma das etapas do processo de Descoberta de Conhecimento em Base de Dados (Knowledge Discovery in Database (KDD)) (FAYYAD et al., 1996). O processo de KDD é dividido em cinco grandes etapas, sendo: (i) seleção de um conjunto de dados ou amostra dos dados com os quais o processo de descoberta será realizado. Para realizar a seleção dos dados é preciso entender o objetivo do estudo, como por exemplo: identificar padrões de genótipos altamente produtivos, e, em seguida, estudar os dados que serão utilizados; (ii) pré-processamento dos dados selecionados por meio da limpeza dos dados. Esta etapa corrige inconsistência nos dados, tais como, valores faltantes, outliers e suavização de dados ruidosos; (iii) transformação dos dados selecionados. Nesta etapa poderão ser empregadas operações de projeção e redução, visando preparar os dados para serem processados pelos algoritmos de mineração; (iv) mineração dos dados, a qual visa extrair conhecimentos dos dados. Esta etapa é realizada por meio da análise de dados e aplicação de algoritmos, muitas vezes de forma repetitiva; e, por último tem-se (v) interpretação e avaliação dos resultados encontrados e, conseqüentemente, apresentação dos conhecimentos úteis adquiridos (FREITAS; LAVINGTON, 1998; CASTRO; FERRARI, 2016).

As principais técnicas de mineração de dados utilizadas para extrair conhecimento dos dados encontram-se classificadas em duas categorias: Tarefas Descritivas e Tarefas Preditivas (CASTRO; FERRARI, 2016). As tarefas descritivas visam entender ou descrever um conjunto de dados que não apresentam saídas associadas (“caráter alvo”) descobrindo propriedades, relações e padrões entre os caracteres ou certo subconjunto de dados, sendo, por vezes, tarefas de aprendizagem não supervisionada, como exemplo tem-se as técnicas de: Agrupamento (clustering), Padrões Sequenciais e Regras de Associação. Enquanto as tarefas preditivas, o processo de mineração é baseado no tratamento de um determinado caráter existente (“caráter alvo”) o qual se deseja descrever ou prever a partir dos demais caracteres, ou seja, o conjunto de dados contém o caráter que se deseja prever, sendo consideradas tarefas de aprendizagem supervisionadas, tendo as técnicas de Classificação e Regressão como as mais utilizadas nesta tarefa (HAN et al., 2012; CASTRO; FERRARI, 2016).

A escolha da técnica ou da tarefa de mineração que será utilizada dependerá de qual tipo de conhecimento se busca extrair dos dados. Cada tarefa possui um conjunto de algoritmos que permite extrair padrões consistentes. A escolha do algoritmo está relacionada com a capacidade de manipular certos tipos de dados, o formato dos dados de entrada do

algoritmo, compreensibilidade do modelo de saída, nível de familiaridade do usuário, escalabilidade, requisitos de software e hardware (CABENA et al., 1997).

A árvore de decisão, técnica de mineração de dados de aprendizagem supervisionada e não paramétrica, pode descrever as relações entre os caracteres independentes (variáveis de entrada ou atributos) e o caráter alvo (variável de saída ou classe) na forma de uma estrutura de árvore só que em sentido invertido, sem a necessidade de assumir hipóteses a priori (BLOCKEEL; RAEDT, 1998). Somado a isto, a árvore de decisão pode selecionar e determinar a importância dos caracteres pelo nível de associação, força da influência e o número de casos influenciados destes.

Estruturalmente, a árvore de decisão representa uma segmentação hierárquica dos dados sendo composta pelo nó raiz, nó de decisão e nó terminal (nó folha). O nó raiz representa a população inteira ou amostra e é dividido em dois ou mais conjuntos, formando sub-nós que podem ser nós de decisão ou nós terminais; nó de decisão representa um sub-nó que é dividido em sub-nós adicionais; e nó terminal ou folha representa um sub-nó que não é dividido.

A descrição do caráter alvo é realizada atravessando-o ao longo da árvore de decisão, partindo-se do nó raiz até chegar em uma folha. Ao alcançar um nó de decisão o caráter é confrontado com um teste associado aquele nó e o resultado determina o próximo nó de travessia, formando os ramos da árvore. Esses testes, na maioria dos casos, consistem na comparação do valor de um caráter independente com um valor constante. Se o caráter em teste num determinado nó é do tipo categórico, o número de ramos a partir do nó de decisão poderá ser igual ou menor ao número de categorias que este caráter possa assumir. Já para caracteres em testes do tipo numérico o nó de decisão se ramifica em dois fazendo a comparação do tipo maior ou menor que uma dada constante. Finalmente, ao chegar em uma folha, o caráter alvo recebe um valor de acordo com uma função associada ao mesmo, sendo um valor aproximado quando o mesmo for do tipo numérico (VAN DIEPEN; FRANCES, 2006).

De maneira geral, a árvore de decisão é criada a partir de um processo de indução (FONSECA, 1994). O processo de indução visa escolher o caráter independente que será utilizado em determinado nó de decisão. Existem diferentes métodos utilizados no processo de indução para escolha do caráter em teste em cada nó de decisão, como índice de Gini ou a entropia (medida de desorganização de um sistema) (TAN et al., 2009). O método mais utilizado é o ganho de informação, também interpretado como redução de entropia (MEIRA, 2008). O atributo com maior redução de entropia é o que divide os dados em grupos mais

homogêneos. Isso permite reduzir o número de testes para classificação de um exemplo, e consequentemente o tamanho da árvore (HAN et al., 2012).

Após a construção da árvore de decisão, esta pode estar super ajustada aos dados utilizados e, via de regra, com alta complexidade. Esta condição é chamada de overfitting e pode ser evitada com o uso de podas da árvore, que a tornam mais generalista e menos complexa. Existem duas categorias principais de poda, a pré-poda e a pós-poda. Na pré-poda, a árvore de decisão interrompe o crescimento em função de determinadas restrições disponíveis nos algoritmos ou determinadas pelo pesquisador, de forma que um ramo de decisão, para ser criado, deve atender a estas restrições, caso contrário, se torna uma folha. Como tipo de restrições tem-se: nível de profundidade da árvore, número mínimo de registros cobertos por folha, significância estatística e ganho de informação (TAN et al., 2009).

No método de pós-poda, a árvore é construída por completo e em seguida busca-se ramos que podem ser podados, seja pela substituição de ramos (sub-árvore) por uma folha, ou pela elevação de uma sub-árvore na estrutura da árvore principal, podando-se assim algum ramo intermediário. A árvore de decisão, depois de pronta, pode ser utilizada para descobrir a estrutura preditiva do problema, permitindo a compreensão de quais caracteres e associações desses caracteres conduzem ao carácter alvo. Regras podem ser extraídas de árvores de decisão, no formato “SE... → ENTÃO”, seguindo o caminho do nó raiz até uma das folhas, facilitando assim sua compreensão, uma vez que permitem a análise dos padrões encontrados de forma individual (TAN et al., 2009).

Vale destacar que as árvores de decisão podem trabalhar com dados do tipo numérico, categórico ou ambos. Quando o carácter alvo for do tipo categórico (classificatório) utiliza-se árvore de classificação e quando o carácter alvo for do tipo numérico (quantitativo) utiliza-se árvore de regressão (CRIMINISI; SHOTTON, 2013; WU et al., 2008). As árvores de regressão preservam a interpretabilidade das árvores de decisão podendo facilmente interpretar as relações entre os caracteres dependentes e o carácter alvo por meio da estrutura da árvore. Logo, o processo de regressão visa prever o carácter alvo a partir dos demais caracteres observados e explicar a relação existente entre os caracteres independentes e o carácter alvo (BREIMAN et al., 1984).

O sucesso da árvore de decisão deve-se ao fato de ser uma técnica extremamente simples que não necessita de parâmetros de configuração e geralmente tem um bom grau de assertividade (LOUPPE, 2014). Todavia, a importância ou associação de um carácter independente não necessariamente está ligada com a causa de variação do carácter alvo (OBERMEYER; EMANUEL, 2016), sendo interessante nos programas de melhoria

genético utilizar essa técnica de forma complementar a técnica de correlação e análise de trilha dos caracteres ou vice-versa, para auxiliar no entendimento e na visualização das associações.

2.4 Seleção para múltiplos caracteres

Como comentado anteriormente, a seleção do genótipo ideal de sorgo biomassa depende da finalidade industrial e envolve vários caracteres simultaneamente, os quais podem apresentar associações positivas ou negativas. Essas associações podem provocar, de acordo com a seleção realizada, mudanças desejáveis em alguns caracteres e indesejáveis em outros. Logo, o uso de procedimentos genético-estatísticos, que permitam minimizar os efeitos indesejados das respostas correlacionadas e que considerem a seleção para múltiplos caracteres, é extremamente recomendada para a obtenção de ideótipos e para o aumento da eficiência dos programas de melhoramento (SANTOS et al., 2007; RAMALHO et al., 2012).

Alternativas de seleção de múltiplos caracteres podem ser agrupadas em três metodologias: (a) método de tandem, que baseia-se na seleção de uma característica por vez, é considerado um procedimento demorado e de eficiência duvidosa, e quase não é praticado; (b) método dos níveis independentes de eliminação, que estabelece níveis mínimos ou máximos para cada característica, podendo haver descarte de indivíduos fora do limite preestabelecido para uma determinada característica, porém, ótimos para as demais; (c) índice de seleção, que combina a informação de todos os caracteres, de modo que a seleção é fundamentada em um único valor, envolvendo todos os demais (CRUZ; REGAZZI; CARNEIRO, 2012; RAMALHO et al., 2012).

O índice de seleção tornou-se uma importante ferramenta no melhoramento de plantas e tem sido amplamente usada nos programas de melhoramento para a seleção de genótipos superiores, pois permite combinar as múltiplas informações contidas na unidade experimental, de modo que seja possível a seleção do genótipo superior com base em vários caracteres de interesse econômico, simultaneamente (CRUZ et al., 2012). Além disso, permite que o desempenho de um genótipo para uma determinada característica, compense alguma outra, podendo ser considerado mais flexível quando comparado aos níveis independentes de eliminação (BERNARDO, 2002).

Existem diversos índices de seleção descritos na literatura que são usados no melhoramento vegetal, como, por exemplo, os índices de: Smith (1936) e Hazel (1943), Willians (1962), Mulamba e Mock (1978), Índice Z (NUNES et al., 2005) e o índice FAI-BLUP (ROCHA et al., 2018). No melhoramento genético, o uso do índice FAI-BLUP

mostrou-se superior ao índice de Smith-Hazel e ao índice de Mulamba e Mock na seleção de genótipos de capim-elefante (*Pennisetum purpureum*) (ROCHA et al., 2018) e de sorgo sacarino (BOTELHO et al., 2021), respectivamente. Sendo seu uso recomendado para a seleção de múltiplos caracteres simultaneamente, independente do sentido que deseja para o carácter em seleção.

O índice FAI-BLUP (Factor analysis and ideotype-design – análise de fatores e design de ideótipos) tem a capacidade de contornar problemas como multicolinearidade e atribuição de pesos econômicos. Neste índice é realizada a análise de fatores exploratória, com o intuito de se extrair as cargas fatoriais da matriz de correlação genética. Posteriormente, são determinados os ideótipos baseando-se na combinação de fatores desejáveis e indesejáveis para o melhoramento genético, sendo que, o número de ideótipos vai depender do número de autovalores maiores ou iguais a unidade.

Depois da determinação dos ideótipos, estima-se as distâncias genótipo-ideótipo, que são convertidas em uma estimativa de similaridade relativa, possibilitando assim, o ranqueamento dos genótipos. Assim, além de resolver o problema da multicolinearidade e da atribuição de pesos econômicos, o índice considera a estrutura das correlações obtidas dos dados e direciona os melhoristas na seleção de genótipos mais próximos ao ideótipo (ROCHA et al., 2018).

REFERÊNCIAS

- ABBAS, T.; ISSA, M.; ILINCA, A. Biomass cogeneration Technologies: A Review. **Scientific Research**, v. 10, n.1, p.1-14. 2020.
- ALMEIDA, L.G.F.; PARRELLA, R.A.C.; SIMEONE, M.L.F.; RIBEIRO, P.C.O.; SANTOS, A.S.; COSTA, A.S.V.; GUIMARÃES, A.G.; SCHAFFERT, R.E. Composition and growth of sorghum biomass genotypes for ethanol production. **Biomass and Bioenergy**, v.122, p.343-348. 2019.
- AMADUCCI, S.; AMADUCCI, M.T.; BENATI, R.; VENTURI, G. Crop yield and quality parameters of four annual fibre crops (hemp, kenaf, maize and sorghum) in the North of Italy. **Ind. Crops Prod**, v.11, p.179-186. 2000.
- BENNET, A.S.; ANEX, R.P. Production, transportation and milling costs of sweet sorghum as a feedstock for centralized bioethanol production in the upper Midwest. **Bioresour. Technol**, v.100, p.1595-1607. 2009.
- BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbury, Minnesota: Stemma Press, 2002. 368 p.
- BLOCKEEL, H.; RAEDT, L.D. “Top-Down Induction of First-Order Logical Decision Trees”, **Artificial Intelligence**, v. 101, n.2, p.1-12. 1998.
- BORÉM, A.; PIMENTEL, L.; PARRELLA, R.A.C. **Sorgo do plantio à colheita**. Viçosa: Editora UFV, 2014. 275 p.
- BORGES, L.S.; AQUINO, F. C.; EVANGELISTA, A. F. Potencial do capim elefante para geração de bioenergia – revisão. **Revista Eletrônica**, v.13, p.4518-4523. 2016.
- BOTELHO, T.T., LEITE, P.S.S., PARRELLA, R.A.C., NUNES, J.A.R. Strategies for multi-trait selection of sweet sorghum progenies. **Crop Breeding and Applied Biotechnology**, v. 21, p.1-9. 2021.
- BOUT, S.; VERMERRIS, W. A candidate-gene approach to clone the sorghum Brown midrib gene encoding caffeic acid O-methyltransferase. **Mol. Genet. Genomics**, v.269, p.205-214. 2003.
- BREIMAN, L.; FRIEDMAN, J.; STONE, C.J.; OLSHEN, R.A. Classification and regression trees. **CRC press**, v.2, p.13-16. 1984.
- BUNPHAN, D.; JAISIL, P.; SANITCHON, J.; JNOLL, J.E.; ANDERSON, W.F. Heterosis and Combining Ability of F1 Hybrid Sweet Sorghum in Thailand. **Crop Science**, v.55, p. 178-187. 2015.
- BURKS, P.S.; KAISER, C.M.; HAWKINS, E.M.; BROWN, P.J. Genomewide association for sugar yield in sweet sorghum. **Crop Science**, v.55, p.2138–2148. 2015.

CABENA, P.; VERHEES, J.; ZANASI, A.; HADJNIAN, P. *Discovering Data Mining: From Concept to Implementation*. **Prentice Hall**, 1 edição, 1997.

CARVALHO, P. C.; CRUZ, C. D. Diagnosis of multicollinearity: assessment of the condition of correlation matrices used in genetic studies. **Brazilian Journal of Genetics**, v. 19, n. 3, p. 479-484. 1996.

CARVALHO, S.P. **Métodos alternativos de estimação de coeficientes de trilha e índices de seleção sob multicolinearidade**. f. Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade Federal de Viçosa, Viçosa, 1995.

CASTRO, E., NIEVES, I.U., RONDÓN, V., SAGUES, W.J., FERNÁNDEZ-SANDOVAL, M.T., YOMANO, L.P. Potential for ethanol production from different sorghum cultivars. **Ind. Crops Prod**, v.109, p.367–373. 2017.

CASTRO, F.M.R.; BRUZI, A.T.; NUNES, J.A.R.; PARRELA, R.A.C.; LOMBARDI, G.M.R.; ALBUQUERQUE, C.J.B.; LOPES, M. Agronomic and Energetic Potential of Biomass Sorghum Genotypes. **American Journal of Plant Science**, v. 6,p1-14. 2015.

CASTRO, L.N.; FERRARI, D.G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016. 231p.

CHERNEY, J.H.; CHERNEY, D.J.R.; AKIN, D.E.; AXTELL, J.D. Potential of Brown-Midrib, Low-Lignin Mutants for Improving Forage Quality. **Advances in Agronomy**, v. 46, p.157-198. 1991.

CRIMINISI, A.; SHOTTON, J. Decision forests for computer vision and medical image analysis. **Springer Science & Business Media**, v.13, p.22-25. 2013.

CRUZ, C.D.; CARNEIRO, P.C.S.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético v2**. Viçosa: UFV, 2014. 688 p.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético v1**. 4.ed. Viçosa: Editora UFV, 2012. 514p.

DA SILVA, M.J., CARNEIRO, P.C.S., CARNEIRO, J.E.S., DAMASCENO, C.M.B., PARRELLA, N.N.L.D., PASTINA, M.M., SIMEONE, M.L.F., SCHAFFERT, R.E., PARRELLA, R.A.C. Evaluation of the potential of lines and hybrids of biomass sorghum. **Ind. Cro. and Prod**, v.125, p.379-385. 2018.

DAMASCENO, C.M.B.; PARRELLA, R.A.C.; SIMEONE, M.L.P.; SCHAFFERT, R.E.; MAGALHÃES, J.V.M. **Caracterização bioquímica de genótipos de sorgo quanto ao teor de lignina e análise molecular de rotas metabólicas visando à produção de etanol de segunda geração**. XXVIII Congresso Nacional de Milho e Sorgo, Goiânia: Associação Brasileira de Milho e Sorgo. 2010.

DAMASCENO, C.M.B.; PARRELLA, R.A.C.; SOUZA, V.F.; SIMEONE, M.L.F.; SCHAFFERT, R.E. **Análise Morfoagronômica e Bioquímica de Um Paineiro de Sorgo Energia para Características Relacionadas à Qualidade da Biomassa**. ISSN 1679-1150 Versão eletrônica. Circular Técnica 190. 2013.

DAVILA-GOMEZ, F.J.; CHUCK-HERNANDEZ, C.; PEREZ-CARRILLO, E.; ROONEY, W.L.; SERNA-SALDIVAR, S.O. Evaluation of bioethanol production from five different varieties of sweet and forage sorghums (*Sorghum bicolor* (L.) Moench). **Industrial Crops and Products**, v.33, p.611-616, 2011.

DIEN, B.S.; SARATH, G.; PEDERSEN, J.F.; SATLER, S.E.; CHEN, H.; FUNNELL-HARRIS, D.L.; NICHOLS, N.N.; COTTA, M.A. Improved sugar conversion and ethanol yield for forage sorghum (*Sorghum bicolor* (L.) Moench) lines with reduced lignin contents. **BioEnergy Research**, v. 2, p. 153-164, 2009.

DOS SANTOS, J.P.R., FERNANDES, S.B., MCCOY, S., LOZANO, R., BROWN, P.J., LEAKEY, A.D.B., BUCKLER, E.S., GARCIA, A.A.F., GORE, M.A. Novel Bayesian Networks for Genomic Prediction of Developmental Traits in Biomass Sorghum. **G3 (Bethesda)**, v.10, n.2, p.769-781. 2020.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária, Milho e Sorgo. **Cultivo do Sorgo**. Brasília, Editora: Embrapa. 2014.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**.4.ed. Edinburgh: Longman Group Limited, 1996. 464p.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. **Communications of the ACM**, v.39, n.11, p.27–34, 1996.

FERNANDES, S.B.; DIAS, K.O.G.; FERREIRA, D.F.; BROWN, P.J. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. **Theoretical and Applied Genetics**, v.131, n.3, p.747–755. 2018.

FONSECA, M.M.R. **Indução de Árvores de decisão** (hist-class - proposta de um algoritmo não paramétrico). Master's thesis, Universidade Nova de Lisboa, 1994.

FREITAS, A. A; LAVINGTON, S. H. **Mining Very Large Databases with Parallel Processing. The Kluwer international series on advances in database systems**. Kluwer Academic Publishers, Boston, 1998. 312p.

GOSWAMI, S.J.; PATEL, P.T.; GAMI, R.A.; PATEL, R.N.; KHATRI, A.B. Correlation and path analysis study of different characters for grain yield and fodder purpose in sorghum [*Sorghum bicolor* (L.) Moench]. **Electronic Journal of Plant Breeding**, v.11, n.4, p.1-9. 2020.

HABYARIMANA, E., LOPEZ-CRUZ, M., BALOCH, F.S. Genomic Selection for Optimum Index with Dry Biomass Yield, Dry Mass Fraction of Fresh Material, and Plant Height in Biomass Sorghum. **Genes**, v.11, n.1, p.61-68. 2020.

HALPIN, C.; HOLT, K.; CHOJECKI, J.; OLIVER, D.; CHABBERT, B.; MONTIES, B.; EDWARDS, K.; BARAKATE, A.; FOXON, G.A. Brown-midrib maize (bmr-1): a mutation affecting the cinnamyl alcohol dehydrogenase gene, **Plant J**. v. 14, p.545–553, 1998.

HAN, J.; KAMBER, M.; PEI, J. **Data mining : concepts and techniques**. Amsterdam; Boston: Elsevier/Morgan Kaufmann, 2012.

HAZEL, L.N. The genetic basis for constructing selection indexes. **Genetics**, v. 28, p. 476-490, 1943.

HE, S.; TANG, C.; WANG, M.L.; LI, S.; DIALLO, B.; XU, Y.; ZHOU, F.; SUN, L.; SHI, W.; XIE, G.H. Combining ability of cytoplasmic male sterility on yield and agronomic traits of sorghum for grain and biomass dual-purpose use. **Industrial Crops and Products**, v. 157, p. 14-27. 2020.

HEITMAN, A.J., M.S. CASTILLO, T.J. SMYTH, C.R. CROZIER, Z. WANG, R.W. HEINIGER, R.J. GEHL. Nitrogen fertilization effects on dry matter yield and nutrient removal of biomass- and sweet-sorghum. **Agron. J.**, v.109, p.1352–1358. 2017.

JOHNSON, R.A., WICHERN, D.W. **Applied Multivariate Statistical Analysis**. 6th Edition, Pearson Prentice Hall, Upper Saddle River. 2007.

JORGENSON, L.R. Brown midrib in maize and its linkage relations. **Journal of the American Society of Agronomy**, v.23, p.549-557, 1931.

KE, J.; CHEN, S. Thermal decomposition of lignin structural modification in termite digested softwood (II). **Fuel**, v. 104, p. 781-787. 2013.

LI, J.; HENRIKSSON, G.; GELLRSTEDT, G. Lignin depolymerization/repolymerization and its critical role for delignification of aspen wood by steam explosion. **Bioresource Technology**, v. 98, p.3061-3068, 2007.

LIANG, X., HOOGENBOOM, G., VOULGARAKI, S., BOOTE, K.J., VELLIDIS, G. Deriving genetic coefficients from variety trials to determine sorghum hybrid performance using the CSM–CERES–Sorghum model. **Agron. J.** Online Version. 2021.

LIU, G.; ZHANG, Q.; LI, H.; QURESHI, A.S.; ZHANG, J.; BAO, X; BAO, J. Dry biorefining maximizes the potentials of simultaneous saccharification and co-fermentation for cellulosic ethanol production. **Biotechnology and Bioengineering**, v.115, p.60-69, 2017.

LOMBARDI, G.M.R.; NUNES, J.A.R.; PARRELLA, R.A.C.; TEIXEIRA, D.H.L.; BRUZI, A.T.; DURÃES, N.N.L.; FAGUNDES, T.G.; Path analysis of agro-industrial traits in sweet sorghum. **Genetic and Molecular Research**, v.14, p.16392-16402. 2015.

LOMBARDI, G.M.R.; NAVEGANTES, P.C.A.; PEREIRA, C.H.; OLIVEIRA, J.O.; PARRELLA, R.A.C.; CASTRO, F.M.R.; ROCHA, M.J.; ORNELAS, D.O.; BRUZI, A.T.; NUNES, J.A.R. Heterosis in sweet sorghum. **Pesquisa Agropecuária Brasileira**, v.53, n.5, p.593-601. 2018.

LOUPPE G. **Understanding random forests: From theory to practice**. Tese de Doutorado, Universidade de Lieja. 2014.

MAY, A. et al. **Sorgo biomassa para a cogeração de energia**. Sete Lagoas: Embrapa Milho e Sorgo, 2015. 7 p. (Circular Técnica, 211)

- MAY, A.; PARRELLA, R. A. da C.; DAMASCENO, C. M. B.; SIMEONE, M. L. F. Sorgo como matéria-prima para produção de bioenergia: etanol e cogeração. **Informe Agropecuário**, Belo Horizonte, v. 35, n. 278, p. 14-20, 2014.
- MEIRA, C.A.A.; RODRIGUES, L.H.A.; MORAES, S. A. Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. **Tropical Plant Pathology**, v.33, n.2, p.114–124. 2008.
- MENGISTU, G.; SHIMELIS, H.; LAING, M. LULE, D.; MATHEW, I. Genetic variability among Ethiopian sorghum landrace accessions for major agro-morphological traits and anthracnose resistance. **Euphytica**, v. 216, p.113-124. 2020.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York : J. Wiley, 1981.
- MULAMBA, N.N.; MOCK, J.J. Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. **Egypt Journal of Genetics and Cytology**, Alexandria, v. 7, p. 40- 51, 1978.
- NUNES, J.A.R.; RAMALHO, M.A.P.; ABREU, A.F.B. Graphical method in studies of adaptability and stability of cultivars. **Annual Report of the Bean Improvement Cooperative**, v. 48, p. 182-183, 2005.
- OBERMEYER, Z.; EMANUEL, E.J. Predicting the future—big data, machine learning, and clinical medicine. **The New England Journal of Medicine**, NIH Public Access, v.13, p.375-391. 2016.
- OLIVEIRA, I.C.M.; MARÇAL, T. de S.; BERNARDINO, K. da C.; RIBEIRO, P.C. de O.; PARRELLA, R.A. da C.; CARNEIRO, P.C.S.; SCHAFFERT, R.E.; CARNEIRO, J.E. de S. Combining ability of biomass sorghum lines for agroindustrial characters and multitrait selection of photosensitive hybrids for energy cogeneration. **Crop Science**, v.59, p.1554-1566, 2019.
- OLIVER, A.L.; PEDERSEN, J.P.; GRANT, W.; KLOPFENSTEIN, T.J. Comparative effects of the sorghum bmr6 and bmr12 genes: I. Forage sorghum yield and quality. **Crop Sci**, v.45, p.2234-2239. 2005.
- OLIVOTO, T., SOUZA, V.Q., NARDINO, M., CARVALHO, I.R., FERRARI, M., PELEGRIN, A.J., SZARESKEI, V.J., SCHMIDT, D. Multicollinearity in Path Analysis: A Simple Method to Reduce Its Effects. **Agron. J**, v.109, n.1, p.131-142. 2017.
- PACKER, D.J.; ROONET, W.L. High-parent heterosis for biomass yield in photoperiod-sensitive sorghum hybrids. **Field Crops Research**, v.167, p.153–158. 2014.
- PARRELLA, R.A.; SCHAFFERT, R.E.; MAY, A.; EMYGDIO, B.; PORTUGAL, A.F.; DAMASCENO, C.M.B. **Desempenho agrônômico de híbridos de sorgo biomassa**. Boletim de Pesquisa e Desenvolvimento / Embrapa Milho e Sorgo, Sete Lagoas, ISSN 1679-0154; 41. 2011. 19 p.
- PARRELLA, R.A.C.; RESENDE, A.V.; MAY, A.; DA SILVA, D.D.; DOS SANTOS, F.C.; COTA, L.C.; DE OLIVEIRA, P.A.; MENDES, S.M.; DA COSTA, R.V.; DE MIRANDA,

R.A. **Cultivo de sorgo biomassa para cogeração de energia**. Sete Lagoas: Embrapa Milho e Sorgo, 2013. 69p. (Embrapa Milho e Sorgo. Boletim de Pesquisa e Desenvolvimento, 152).

PATEL, B.; PATEL, A.; GAMI, B.; PATEL, P. Energy balance, GHG emission and Economy for cultivation of high biomass varieties of Bamboo, Sorghum and Pearl millet as energy crops at marginal ecologies of Gujarat State in India. **Renewable Energy**, v.148, p.816-823. 2019.

PAULA, L.E.R.; TRUGILHO, P.F.; NAPOLI, A.; BIANCHI, M.L. Characterization of residues from plant biomass for use in energy generation. **Cerne**, v.17, n.2, p.237-246. 2011.

PERAZZO, A.F.; CARVALHO, G.G.P.; SANTOS, E.M.; PINHO, R.M.A.; CAMPOS, F.S.; MACEDO, C.H.O.; AZEVEDO, J.A.G.; TABOSA, J.N. Agronomic evaluation of 32 sorghum cultivars in the Brazilian semi-arid region. **Revista Brasileira de Zootecnia**, v.43, n.5, p.232-237. 2014.

PEREIRA, L.S.; CORDERY, I.; IACOVIDES, I. Improved indicators of water use performance and productivity for sustainable water conservation and saving, **Agricultural Water Management**, Amsterdam, v.108, p.39–51, 2012.

PETROFF, G.; DOAT, J. Pirolise des bois tropicaux: influence de la composition chimique des bois sur les produits de distillation. **Bois et forêts de tropiques**, v.117, p.51-64, 1978.

PFEIFFER, T.W.; BITZER, M.J.; TOY, J.; PETERSEN, J.F. Heterosis in sweet sorghum and selection of a new sweet sorghum hybrid for use in syrup production in Appalachia. **Crop Science**, v.50, p.1788-1794. 2010.

PORTER, K.S.; AXTELL, J.D.; LICHTENBERG, V.L.; COLENBRANDER, V.F. Phenotype, fiber composition and in vitro organic matter disappearance of chemically induced born midrib mutants of sorghum. **Crop Science**, v.18, p.205-212, 1978.

QUINBY, J.R.; SCHERTZ, K.F. **Sorghum genetics, breeding, and hybrid seed production**. In: Wall JS, Ross WM (eds) **Sorghum production and utilization**. AVI Publishing, Westport, Connecticut, 1970. 171p.

RAMALHO, M.A.P.; ABREU, A.F.B.; SANTOS, J.B.; NUNES, J.A.R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. 1ª Edição. Lavras- MG. Editora: UFLA, 2012. 522p.

RAMALHO, M.A.P.; SANTOS, J.B.; ZIMMERMANN, M.J. **Genética quantitativa em plantas autógamas: aplicações ao melhoramento do feijoeiro**. Goiânia: UFG, 1993. 271 p.

REGASSA, T.H; WORTMANN, C.S. Sweet Sorghum as a bioenergy crop: Literature review. **Biomass and Bioenergy**, v.6, p. 1 – 8. 2014.

RIBAS, P.M. **Cultivo do sorgo**. 2008. Disponível em: http://www.cnpms.embrapa.br/publicacoes/sorgo_4_ed/plantio-plantio.html Acesso em: 10 de abril de 2021.

ROONEY, W. L. Sorghum Improvement - Integrating Traditional and New Technology to Produce Improved Genotypes. **Advances in Agronomy**, College Station, v.83, p.37-109. 2004.

ROONEY, W.L.; AYDIN, S. Genetic control of a photoperiod-sensitive response in Sorghum bicolor (L.) Moench. **Crop Science**, v. 9, p. 397-400. 1999.

ROCHA, J.R.A.S.C.; MACHADO, J.C.; CARNEIRO, P.C.; Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy**, v.10, p. 52-60, 2018.

SABALLOS, A. Development and utilization of sorghum as a bioenergy crop. In: VERMERIS, W. (Ed.). Genetic improvement of bioenergy crops. **Springer**, v.5, p.211-248. 2008.

SAINI, P., KOFF, J.P., ALLISON, A., HAMILTON, C. Changes in Lignocellulosic Polymers, Carbon, and Energy in Switchgrass for Bioenergy Production. **Agron. J.**, v.109, n.5, p.1935-1943. 2017.

SANTOS, F. G.; CASELA, C. R.; WAQUIL, J. M. **Melhoramento de sorgo**. In: BORÉM, A. Melhoramento de espécies cultivadas, 2ª Ed., 2005.

SANTOS, F. S.; AMARAL JUNIOR, A.T.; FREITAS JUNIOR, S.P.; RANGEL, R.M.; PEREIRA, M.G. Predição de ganhos genéticos na população de milho pipoca UNB-2U sob seleção recorrente. **Bragantia**, v. 66, n. 3, p. 389–396, 2007.

SATTLER, S.E.; FUNNELL-HARRIS, D.L.; PEDERSEN, J.F. Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. **Plant Science**, v.178, p.229-238, 2010.

SATTLER, S.E.; SAATHOFF, A.J.; HAAS, E.J.; PALMER, N.A.; FUNNELL-HARRIS, D.L.; SARATH, G.; PEDERSEN, J.F. A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the Sorghum brown midrib 6 phenotype, **Plant Physiol**, v.150, p.584–595, 2009.

SCOTT, V., HASZELDINE, R.S., TETT, S.F.B., OSCHLIES, A. Fossil fuels in a trillion tonne world. **Nature Clim. Change**, v.5, p.419–423. 2015.

SILVA, M.J. da; CARNEIRO, P.C.S.; CARNEIRO, J.E. de S.; DAMASCENO, C.M.B.; PARRELLA, N.N.L.D.; PA ST I NA, M.M.; SIMEONE, M.L.F.; SCHAFFERT, R.E.; PARRELLA, R.A. da C. Evaluation of the potential of lines and hybrids of biomass sorghum. **Industrial Crops and Products**, v.125, p.379-385, 2018.

SILVA, M.J. da; DAMASCENO, C.M.B.; CARNEIRO, J.E. de S.; PEREIRA, H.D.; CARNEIRO, P.C.S.; SCHAFFERT, R.E.; PARRELLA, R.A. da C. Combining ability of biomass sorghum in different crop years and sites for bioenergy generation. **Agron. J.**, v.112, p.1549-1563, 2020.

SIMEONE, M.L.F., PARRELLA, R.A.C., DAMASCENO, C.M.B., SCHAFFERT, R.E. Prediction of high-biomass sorghum quality using near infrared spectroscopy to monitoring

calorific value, moisture, and ash content. **International Journal of Development Research**, v.10, n.9, p.40916-40920. 2020.

SMITH, H. F. A discriminant function for plant selection. **Annual Eugenics**, v. 7, p. 240-250, 1936.

SMITH, C. W.; FREDERIKSEN, R. A. **Sorghum: origin, history, technology, and production**. Wile Series in Crop Science, Series Editor Texas A & M University. 2005. 824p.

SIMS, R.E.H.; MABEE, W.E.; SADDLER, J.; TAYLOR, M. An overview of second generation biofuel technologies. **Bioresource Technology**, v.101, n.6, p.1570-1580. 2010.

SKONIESKI, F.R.; NORBERG, J.L.; AZEVEDO, E.B. de; DAVID, D.B.; KESSLER, J.D.; MENEGAZ, A.L. Produção, caracterização nutricional e fermentativa de silagens de sorgo forrageiro e sorgo duplo propósito. **Acta Scientiarum**. Animal Science, v.32, n.1, p.27-32, 2010.

STEPHENSON, T.D., CARVALHO, H.D.R., CASTILLO, M.S., CROZIER, C.R., SMYTH, T.J., HEITMAN, J.L. Water use and biomass yield of bioenergy crops in the North Carolina Piedmont. **Agron. J.** Online Version. 2021.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining**. Rio de Janeiro: Ciência Moderna, 2009.

VAN DER WEIJDE, T.; TORRES, A.F.; DOLSTRA, O.; DECHESNE, A.; VISSER, R.G.F.; TRINDADE, L.M. Impact of Different Lignin Fractions on Saccharification Efficiency in Diverse Species of the Bioenergy Crop Miscanthus. **Bioenergy Research**, v.9, p.146-156. 2016.

VAN DIEPEN, M.; FRANCES, P.H. Evaluating chi-squared automatic interaction detection. **Information Systems**, v.31, n.8, p.814-831. 2006

VENCOVSKY, R.; BARRIGA, P. Genética biométrica no fitomelhoramento. Ribeirão Preto: **Revista Brasileira de Genética**. 1992. 496 p.

VENDRUSCOLO, T.P.S.; BARELLI, M.A.A.; CASTRILLON, M.A.S.; SILVA, R.S.; OLIVEIRA, F.T.; CORREA, C.L.; ZAGO, B.W.; TARDIM, F.D. Correlation and path analysis of biomass sorghum production. **Genetics and Molecular Research**, v.15, n.4, p.1-16, 2016.

VERMERRIS W.; SABALLOS, A. Genetic enhancement of sorghum for biomass utilization. **Genomics of the Saccharinae**, v.11, p.391-425. 2013.

WILLIAMS, J.S. The evaluation of a selection index. **Biometrics**, North Carolina, v.18, p. 375-393, 1962.

WRIGHT, S. Correlation and causation. **Journal of Agricultural Research**, v.20, p.557-585. 1921.

WU, M.R., SCHOTT, D.L., LODEWIJKS, G. Physical properties of solid biomass. **Biomass and Bioenergy**, v.35, p.2093-2105. 2011.

WU, X.; KUMAR V.; QUINLAN, J.R.; GHOSH, J.; YANG, Q.; MOTODA, H.; MCLACHLAN, G.J.; LIU, B.; PHILIP, S.Y. Top 10 algorithms in data mining. **Knowledge and information systems**, v.14, n.1, p.1–37. 2008.

XIAO, B.; SUN, F.; SUN, R.C. The chemical modification of lignins with succinic anhydride in aqueous systems, **Polymer Degradation and Stability**, v.71, n.2, p.223-231, 2001.

ZEGADA-LIZARAZU, W.; MONTI, A. Are we ready to cultivate sweet sorghum as a bioenergy feedstock? A review on field management practices. **Biomass and Bioenergy**, v.40, p.1-12,. 2012.

ZHAO, Y.L.; A. DOLAT, Y. STEINBERGER, X.; WANG, A.; OSMAN; XIE, G.H. Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. **Field Crops Research**, v.111, n.1, p.55–64. 2009.

**CAPÍTULO 2 – ASSOCIAÇÃO ENTRE CARACTERES DE SORGO BIOMASSA
USANDO TÉCNICAS DE ANÁLISE DE TRILHA E ÁRVORE DE REGRESSÃO
COM APRENDIZADO DE MÁQUINA**

RESUMO

O objetivo com este estudo foi estimar as associações entre caracteres do sorgo biomassa; estimar os efeitos diretos e indiretos e verificar a existência de relação causa e efeito dos caracteres com o poder calorífero teórico via análise de trilha; e fornecer uma técnica complementar para análise dos caracteres e direcionamento dos programas de melhoramento de sorgo biomassa por meio da Mineração de Dados. Foram avaliados vinte e cinco híbridos de sorgo biomassa sob delineamento látice quadrado triplo 5x5. Avaliaram-se os caracteres: florescimento (FLOR); altura de planta (AP); produtividade de biomassa verde do colmo (PMVc), das folhas (PMVf) e das panículas (PMVp); massa seca do colmo (MSc) e foliar (MSf); fibra em detergente neutro do colmo (FDNc) e foliar (FDNf); fibra em detergente ácido do colmo (FDAc) e foliar (FDAf); lignina do colmo (LIGc) e foliar (LIGf); hemicelulose do colmo (HEMc) e foliar (HEMf); a concentração de celulose do colmo (CCc) e foliar (CCf); e o poder calorífero teórico (PCT). Os dados foram submetidos a análise exploratória, a análise univariada, correlação genotípica, análise de trilha e a técnica de árvore de regressão. A análise de trilha e a árvore de regressão consideraram o PCT como carácter alvo. O PCT apresentou tanto associações nulas quanto positivas com os caracteres avaliados. Destacando-se as associações positivas com FLOR, AP, PMVc, MSf, MSc, HEMc, LIGc, FDNf, FDAf, HEMf, LIGf. Os caracteres FLOR, AP, PMVc, MSf e LIG foram consideradas os principais determinantes das alterações do PCT podendo ser utilizados na seleção indireta deste carácter. Além disso, os caracteres FLOR, AP e PMVc influenciaram direta ou indiretamente a expressão dos caracteres MSf e LIGc, podendo ser utilizados na seleção indireta dos caracteres MSf e LIGc. A técnica de Mineração de Dados empregada permitiu melhor visualização das relações entre os caracteres, auxiliando em algumas explicações sobre a influência dos caracteres sobre o PCT.

Palavras-chave: *Sorghum bicolor* L., análise de trilha, correlação, bioenergia, árvore de decisão.

ABSTRACT

The aim of this study was to estimate the associations between sorghum characters biomass; to estimate the direct and indirect effects and verify the existence of a cause-and-effect relationship of the traits with the theoretical heating power via trail analysis; and provide a complementary technique for character analysis and targeting sorghum biomass improvement programs through Data Mining. Twenty-five sorghum biomass hybrids were evaluated under a 5x5 triple square lattice design. The following characters were evaluated: flowering (FLOW); plant height (PH); production of total green mass, stem (GSBS), leaf (GSBL) and panicles (GSBP); dry mass, stem (SDM) and leaf (LDM); neutral detergent fiber, stem (SNDF) and leaf (LNDF); acid detergent fiber, stem (SADF) and leaf (LADF); lignin, stem (SL) and leaf (LL); hemicellulose, stem (SH) and leaf (LH); the concentration of cellulose in the stem (SC) and in the leaf (LC); and the theoretical calorific value (THP). The data were submitted to exploratory analysis, univariate analysis, genotypic correlation, trail analysis and the regression tree technique. The path analysis and the regression tree considered THP as target character. THP also showed both null and positive associations with the evaluated characters. Highlighting the positive associations with FLOW, PH, GSBS, MSf, SDM, SH, SL, LNDF, LADF, LH, LL. The characters FLOW, PH, GSBS, LDM and lignin were considered the main determinants of THP alterations and can be used in the indirect selection of this character. In addition, the characters FLOW, PH and GSBS directly or indirectly influence the expression of the characters LDM and SL and can be used in the indirect selection of these characters. The Data Mining technique used allowed a better visualization of the relationships between the characters, helping to explain the influence of the characters on the THP.

Keywords: Sorghum bicolor L., path analysis, correlation, bioenergy, decision tree.

1 INTRODUÇÃO

A demanda mundial por fontes renováveis de energia elétrica se expandiu em decorrência do alto consumo de combustível fóssil e seus impactos ambientais (MOOD et al., 2013). A biomassa, maior fonte de bioenergia do mundo cuja combustão causa menor emissão de substâncias tóxicas e nocivas (KAMBO; DUTTA, 2014), pode ser produzida por qualquer material vegetal, animal ou residual (humano e industrial) (DERMIBAS, 2001). A cultura da cana-de-açúcar (*Saccharum officinarum* L.) e as culturas florestais (ex. eucalipto (*Eucalyptus* spp.)) são as principais matérias-primas utilizadas na produção mundial de bioenergia via combustão de biomassa (VENDRUSCOLO et al., 2016; ABBAS et al., 2020). No entanto, a busca de novas fontes de biomassa, principalmente, oriundas de culturas anuais, visando suprir a demanda mundial de bioenergia, é crescente em diversos países, como no Brasil (DA SILVA et al., 2020a), nos Estados Unidos da América (PARKHURST et al., 2016; STEPHENSON et al., 2021), na Itália (SERRA et al., 2017), na Austrália (BLAKERS et al., 2017), na Índia (PATEL et al., 2019) e em Portugal (FERREIRA et al., 2017).

O sorgo biomassa, cultura energética promissora para produção de bioenergia (ABBAS et al., 2020; SIMEONE et al., 2020), apresenta ciclo curto de, aproximadamente, 180 dias; altura elevada, sendo superior a 4 metros; alta produtividade de biomassa verde e seca, em torno de 150 t ha⁻¹ e 50 t ha⁻¹, respectivamente (HEITMAN et al., 2017; SIMEONE et al., 2020); ampla adaptabilidade para regiões tropicais e temperadas (STEPHENSON et al., 2021; LIANG et al., 2021); e alto poder calorífico em caldeiras (CASTRO et al., 2015; DA SILVA et al., 2018; OLIVEIRA et al., 2019).

O poder calorífico pode ser definido como a quantidade de energia liberada por massa do sólido quando queimado totalmente na presença de oxigênio (SIMEONE et al., 2020). Logo, quanto maior o poder calorífico de uma cultura maior será a geração de energia elétrica pela queima da biomassa. O poder calorífico do sorgo biomassa é considerado elevado, em torno de 4500 Kcal Kg⁻¹ de biomassa seca (CASTRO et al., 2015; DA SILVA et al., 2018), principalmente, quando comparado com as principais culturas utilizadas como matéria-prima na produção mundial de bioenergia via combustão de biomassa, como eucalipto (4501 Kcal Kg⁻¹), cana-de-açúcar (4274 Kcal kg⁻¹) (PAULA et al., 2011), switchgrass (*Panicum virgatum*, L.) (4204 Kcal kg⁻¹) (SAINI et al., 2017) e capim elefante (*Pennisetum purpureum*, S.) (4200 Kcal kg⁻¹) (BORGES et al., 2016).

O elevado poder calorífero do sorgo biomassa deve-se, entres outros caracteres, a umidade e ao teor de lignina (ABBAS et al., 2020; SIMEONE et al., 2020). A umidade da

biomassa interfere diretamente na temperatura de combustão, podendo, quando elevada, dificultar a queima da biomassa (WU et al., 2011; CASTRO et al., 2015), aumentando os custos para geração de energia elétrica; sendo ideal, teor de umidade das plantas em torno de 50% no momento da colheita, o que é facilmente tangível para a cultura (CASTRO et al., 2015). A lignina, composto lignocelulósico presente nos vegetais possui alta resistência a degradação térmica, devido ao elevado número de carbonos presentes em sua estrutura e baixo teor de oxigênio, apresentando, assim, elevado poder calorífico em torno de 6100 Kcal kg⁻¹ (XIAO et al., 2001). Logo, o cultivo de genótipos de sorgo biomassa com maiores teores de lignina é ideal para incrementar o poder calorífico da cultura e para maior geração de energia elétrica.

Em conjunto com esses caracteres (umidade e teor de lignina), a obtenção de genótipos com alta produtividade de biomassa (verde e seca) é extremamente importante para a geração de energia elétrica (NHUCHHE; AFZAL, 2017; DA SILVA et al., 2020a), pois quanto maior a quantidade de massa para se queimar maior será a geração de energia elétrica. A produtividade foi fortemente incrementada com a ampliação do ciclo vegetativo da cultura, isto é, com o cultivo de genótipos sensíveis ao fotoperíodo (DA SILVA et al., 2020b). Contudo, a produtividade apresenta baixa herdabilidade (DOS SANTOS et al., 2020), sendo resultante da expressão e associação de diversos caracteres (OLIVEIRA et al., 2021). Assim sendo, é caro e trabalhoso a obtenção de genótipos superiores para produtividade por meio do fenótipo (FERNANDES et al., 2018) o mesmo ocorre para o teor de lignina, apesar deste apresentar elevada herdabilidade (OLIVEIRA et al., 2018).

A associação entre os caracteres, que pode ser positiva ou negativa, é um dos recursos utilizados pelos programas de melhoramento genético para auxiliar na seleção e obtenção de genótipos superiores (VENDRUSCOLO et al., 2016; GOSWAMI et al., 2020). Visto que, pode-se ter rápidos progressos com a seleção indireta de genótipos superiores para o carácter de interesse, por meio da seleção pelo carácter associado, principalmente, quando o carácter de interesse apresenta baixa herdabilidade e dificuldade na mensuração e o carácter associado não (WRIGTH, 1921). Por outro lado, a seleção de um carácter pode ocasionar mudanças indesejáveis em outros, pois este pode associar-se tanto positivamente quanto negativamente com os outros caracteres (WRIGTH, 1921, LOMBARDI et al., 2015; OLIVEIRA et al., 2021). Outro fator limitante da associação entre dois caracteres é a possível influência de um terceiro carácter, ou conjunto de caracteres, sobre a estimativa de associação (WRIGTH, 1921). Dessa forma, é importante estudar a associação entre dois caracteres (correlação), bem como, as associações de um conjunto de caracteres sobre o carácter de interesse (análise de

trilha) para verificar a existência de relação causa e efeito entre os caracteres; fornecendo assim maior poder na tomada de decisão aos melhoristas (LOMBARDI et al., 2015; VENDRUSCOLO et al., 2016; GOSWAMI et al., 2020; OLIVEIRA et al., 2021).

A análise de trilha é uma técnica extremamente útil no melhoramento de diversas culturas, principalmente, por possibilitar o conhecimento da relação entre causa e efeito entre os caracteres (WRIGHT, 1921). Contudo, a análise de trilha apresenta algumas limitações como a multicolinearidade, a dimensionalidade dos dados e o fato de não apresentar resultado visual e de fácil interpretação (OLIVOTO et al., 2017). Algumas dessas limitações podem ser contornadas por meio do uso complementar de outras técnicas de mineração de dados, como a árvore de decisão. A árvore de decisão é uma técnica extremamente simples, visual e de fácil interpretação, que não necessita de parâmetros de configuração e geralmente tem um bom grau de assertividade (LOUPPE, 2014), podendo ser utilizadas para descrever as associações entre os caracteres avaliados (BREIMAN et al., 1984), porém sem verificar a causa de variação do carácter alvo (OBERMEYER; EMANUEL, 2016).

No sorgo biomassa os caracteres apresentam grande variação quanto a direção e a magnitude das associações (CASTRO et al., 2015; VENDRUSCOLO et al., 2016; ALMEIDA et al., 2019; DOS SANTOS et al., 2020; HABYARIMANA et al., 2020). Uma associação interessante e benéfica ao melhoramento da cultura foi observada, desde os 30 dias após a semeadura da cultura, entre a altura de planta e a produtividade de biomassa seca (FERNANDES et al., 2018; DOS SANTOS et al., 2020). Este conhecimento, permite a utilização do carácter altura para selecionar genótipos de sorgo biomassa com maior produtividade de biomassa seca de maneira menos trabalhosa e onerosa. No entanto, até o presente momento não foi observado relação entre causa e efeito da associação desses caracteres, indicando que a altura, possivelmente, não é o principal determinante das alterações na produtividade (VENDRUSCOLO et al., 2016).

A ausência de conhecimento da relação entre causa e efeito das associações entre os caracteres de sorgo biomassa e a dificuldade de interpretação das análises de associação impactam no entendimento da associação dos caracteres do sorgo biomassa e no direcionamento dos programas de melhoramento da cultura (VENDRUSCOLO et al., 2016; OLIVOTO et al., 2017).

Dessa forma, o objetivo com este trabalho foi estimar as associações entre caracteres do sorgo biomassa; estimar os efeitos diretos e indiretos e verificar a existência de relação causa e efeito dos caracteres com o poder calorífero teórico (PCT) via análise de trilha; e

fornecer uma técnica complementar para análise dos caracteres e direcionamento dos programas de melhoramento de sorgo biomassa por meio da Mineração de Dados.

2 MATERIAL E MÉTODOS

2.1 Local, Plano Experimental e Manejo

O experimento foi conduzido no ano agrícola de 2016/2017 no Centro de Desenvolvimento Científico e Tecnológico em Agropecuária da Universidade Federal de Lavras (UFLA), no município de Lavras, Estado de Minas Gerais, Brasil (21°14' S; 45°00' O; 932 m). O clima de Lavras, foi classificado, segundo Koppen, como Cwa (ALVARES et al., 2013) e, apresentou durante a condução do experimento, precipitação total de 361,3 mm e temperatura média de 22,1°C. O solo da área experimental foi classificado como Latossolo Vermelho-amarelo, de acordo o Sistema Brasileiro de Classificação de Solos, correspondente ao Oxisol (SOIL SURVERY STAFF, 2014).

O experimento foi instalado, em 26 de novembro de 2016, segundo o delineamento látice quadrado triplo 5 x 5. Foram avaliados vinte e cinco genótipos sensíveis ao fotoperíodo, sendo vinte e um híbridos experimentais de sorgo biomassa e quatro híbridos comerciais - HC (HC1: BRS716; HC2: BRS656, HC3: TESTSB; HC4: Volumax). As parcelas experimentais, constituídas por duas linhas de 5,0 m de comprimento e espaçadas de 0,70 m entre fileiras, foram aradas usando um conjunto de discos tandem a uma profundidade de 20 cm seguido por um cultivador de campo para a preparação do canteiro de sementes. A densidade de plantas adotada foi de 140.000 plantas ha⁻¹. Na semeadura, realizada de forma direta, foi aplicado 450 kg ha⁻¹ do formulado NPK 08-28-16 nos sulcos de plantio e quando as plantas estavam no estágio V4-V5 foi aplicado, em cobertura, 200 kg de ureia ha⁻¹. Além disso, durante a condução do experimento, foi realizado o controle de plantas daninhas por meio da aplicação de atrazina (3 kg i.a. ha⁻¹) complementado por capina mecânica, quando necessário. A colheita foi realizada manualmente, em média, aos 180 dias após a semeadura.

2.2 Caracteres mensurados

Os caracteres agrônômicos mensurados foram: florescimento (FLOR; dias) de pelo menos 50% das plantas da parcela, em dias após o semeadura; altura da planta (AP; m) na média de oito plantas aleatórias, medidas da superfície do solo ao ápice da panícula, no momento da colheita; produtividade de massa verde (PMV; t ha⁻¹) de oito plantas aleatórias por parcela, com panícula e folhas, produtividade de massa verde do colmo (PMVc; t ha⁻¹) de oito plantas por parcela, sem panícula e folhas, produtividade de massa verde das folhas

(PMVf; t ha⁻¹) de oito plantas por parcela, sem colmo e panícula e, produtividade de massa verde das panículas (PMVp; t ha⁻¹) de oito plantas por parcela, sem colmo e folhas, ambos pesados por meio de uma balança de suspensão digital, em Kg, convertida em t ha⁻¹; massa seca (MS, %) determinada pela soma entre a massa seca foliar e do colmo de oito plantas sobre a massa verde dos mesmos; massa seca foliar (MSf, %) determinada pela razão entre a massa seca e a massa verde foliar de oito plantas; massa seca do colmo (MSc, %) determinada pela razão entre a massa seca e a massa verde do colmo de oito plantas, a massa seca foi obtida pela secagem do material em estufa e dessecador até que as amostras apresentassem peso constante, seguindo a metodologia de Zhao et al. (2009).

Com base na MS, determinou-se os caracteres de composição da biomassa: fibra em detergente neutro do colmo e foliar (FDNc e FDNl, %), determinado pela razão entre o peso da amostra após a digestão em 35 ml de solução de detergente neutro e 0,5 ml de amilase, filtragem e secagem em estufa a 105°C e o peso da amostra antes da digestão (0,35g) de cada componente (colmo e folha); fibra em detergente ácido (FDA, %), determinado pela razão entre o peso da amostra após a digestão em 35 ml de solução de detergente ácido, filtragem e secagem em estufa a 105°C e o peso da amostra antes da digestão (0,35g) de cada componente (colmo e folha); lignina do colmo e foliar (LIGc e LIGf, %), determinado segundo a metodologia de Soest (1963) para cada componente (colmo e folha); hemicelulose do colmo e foliar (HEMc e HEMf, %), obtido pela diferença entre FDN e FDA; e, a concentração de celulose do colmo e foliar (CCc e CCf, %) obtido pela diferença entre FDA e LIG.

Após a mensuração desses caracteres, obteve-se a variável básica PCT (poder calorífico teórico) por meio da expressão: , em que, PMV é a produção de massa verde; MSm é a massa seca média e, FDNm é o teor médio de fibra em detergente neutro na planta.

2.3 Análise Estatística

Os dados coletados dos caracteres foram submetidos a análise exploratória. Na análise exploratória, foi realizada análise descritiva dos caracteres e exame gráfico das relações entre os caracteres, visando, principalmente, identificar possíveis valores atípicos (outliers), presença de dados faltantes (missing) e avaliar algumas suposições básicas, como normalidade e homocedasticidade. A análise exploratória foi realizada utilizando o pacote Scikit Learn versão 0.24.2 (PEDREGOSA et al., 2011) em ambiente Python 3 (LUTZ, 2014).

Os pressupostos da análise de variância, após avaliação em ambiente Python, foram verificados por meio do teste de Shapiro-Wilk e Levene quanto a normalidade dos erros e

homocedasticidade das variâncias, respectivamente. Constatado que os dados atendiam os pressupostos, realizou-se análise univariada dos caracteres dos genótipos de sorgo biomassa de acordo com o modelo estatístico: $y^c = X^c B^c + Z^c b^c + W^c g^c + e^c$, em que: y^c é o vetor de dados fenotípicos do c-ésimo carácter; X^c, Z^c, W^c são matrizes de incidência dos efeitos fixos e aleatórios; B^c é o vetor de efeitos fixos das repetições; b^c é o vetor de efeito de blocos dentro de repetições, sendo $b^c \sim N(0, I\sigma^2b)$; g^c é o vetor de efeitos dos genótipos, sendo $g^c \sim N(0, I\sigma^2g)$; e^c é o vetor de efeitos do erro experimental, sendo $e^c \sim N(0, I\sigma^2e)$. A qualidade experimental foi aferida pela estimativa da acurácia seletiva ($AS = \sqrt{1-1/F}$), a qual reflete a confiabilidade na seleção de genótipos a partir dos valores fenotípicos (RESENDE; DUARTE, 2007). A análise univariada e a verificação dos pressupostos da análise de variância foram realizadas, respectivamente, por meio do pacote `lmtest` (ZEILEIS; HOTHORN, 2002) e por meio da função `gamem()` do pacote `metan` (OLIVOTO; LÚCIO, 2020) em ambiente R (R CORE TEAM, 2018).

Em seguida, estimou-se a associação entre os todos caracteres avaliados e os efeitos diretos e indiretos destes por meio da correlação genotípica e da análise de trilha, respectivamente. As significâncias das correlações genotípicas foram verificadas via bootstrap. Para a análise de trilha, a matriz de correlação genotípica das variáveis explicativas foi previamente diagnosticada quanto à multicolinearidade. O teste utilizado foi do número de condição da matriz (NC), que consiste na razão entre o maior e menor autovalor da matriz de correlação (MONTGOMERY; PECK, 1981; OLIVOTO et al., 2017). A multicolinearidade é considerada fraca quando NC é menor que 100, moderada a forte se $100 < NC < 1000$ e severa se $NC > 1000$. Constatada multicolinearidade severa, adicionou-se uma constante k de 0,04684 na diagonal principal da matriz de correlação das variáveis explicativas à semelhança do método de regressão em crista.

A análise de trilha foi realizada em diagrama causal de duas cadeias. A primeira cadeia constou do carácter PCT como variável básica e pelos caracteres MSc, FDNc, FDAc, LIGc, HEMc, CCc, MSf, FDNf, FDAf, LIGf, HEMf e CCf, como variáveis explicativas primárias. A segunda cadeia constou das variáveis explicativas secundárias FLOR, AP, PMVf, PMVc e PMVp sobre as variáveis primárias. A análise de correlação, o teste de multicolinearidade e a análise de trilha foram realizadas por meio da função `path_analysis_seq()` do pacote `metan` (OLIVOTO; LÚCIO, 2020), as figuras foram realizadas utilizando o pacote `ggplot2` (WICKHAM, 2016) em ambiente R (R CORE TEAM, 2018).

A árvore de decisão foi construída considerando o carácter PCT como alvo e os demais caracteres como independentes. Como o carácter alvo PCT é do tipo numérico utilizou-se a árvore de regressão. Construiu-se a árvore de regressão por meio da classe `sklearn.tree.DecisionTreeRegressor` (SCIKIT LEARN, 2020) do pacote Scikit Learn versão 0.24.2 (PEDREGOSA et al., 2011) em ambiente Python 3 (LUTZ, 2014). Para a construção da árvore foi considerado todas as instâncias avaliadas (genótipos), sendo 70% dos dados para treino e 30% para teste (`test_size=0,30`), parâmetro `RandomState` equivalente a 1 (`random_state=1`) para que o algoritmo construa sempre a mesma árvore e a profundidade da árvore equivalente a 4 (`max.depth=4`). A profundidade da árvore foi definida com base no melhor ajuste do modelo, sendo adotado a profundidade 4, pois na mesma o coeficiente de determinação (R^2) foi considerado o mais alto. Após a construção da árvore foi observado a importância dos caracteres independentes para o modelo por meio do atributo `feature_importances` da classe `sklearn.tree.DecisionTreeRegressor` (SCIKIT LEARN, 2020).

3 RESULTADOS

A estimação da associação genotípica dos caracteres e a seleção, direta ou indireta, de genótipos superiores pelos programas de melhoramento genético dependem da existência de variabilidade genética (CRUZ et al., 2012; SINHA; KUMARAVADIVEL, 2016). A variabilidade genética ($p < 0,05$) foi observada para todos os caracteres mensurados (Figura 1) e a associação entre eles variou desde fortemente positiva até moderadamente negativa (Figura 2).

Os caracteres agronômicos, FLOR, AP, PMVc, MSf e MSc, apresentaram elevadas associações positivas entre si ($0,54 < r_g < 0,90$) e associações nulas com PMVf ($0,13 < r_g < 0,32$) e com PMVp ($-0,26 < r_g < 0,18$) (Figura 2). Os caracteres que compõem a biomassa do colmo, FDNc, FDAc, LIGc, HEMc e CCc, apresentaram elevadas associações positivas entre si ($0,48 < r_g < 0,98$) (Figure 2). Associações similares foram observadas entre os caracteres que compõem a biomassa da folha, sendo elevadas e positivas entre FDNf, FDAf, HEMf e CCf ($0,79 < r_g < 0,99$) e nulas com LIGf ($-0,07 < r_g < 0,28$) (Figura 2). As associações entre os caracteres que compõem a biomassa do colmo e da folha foram consideradas nulas ($-0,23 < r_g < 0,26$), com exceção para as associações entre LIGf x LIGc a qual apresentou associação elevada e positiva ($r_g = 0,65$) (Figura 2).

As associações entre os caracteres agronômicos e os caracteres que compõem a biomassa apresentaram grandes variações ($-0,29 < r_g < 0,74$) (Figura 2). Considerando os caracteres que compõem a biomassa do colmo, tem-se associações positivas entre a LIGc com

o FLOR, AP, PMVc, MSc e MSf ($rg > 0,57$), entre a HEMc e a PMVf ($rg > 0,51$) e entre FDAc, FDNc e CCc com PMVp ($rg > 0,50$) (Figura 2). Considerando os caracteres que compõem a biomassa foliar, têm-se associações positivas entre os caracteres FDAf, FDNf, LIGf e HEMf com FLOR, AP, PMVc e MSf ($rg > 0,44$) e associações nulas entre os caracteres PMVf e PMVp com todos os caracteres que compõem a biomassa foliar ($-0,28 < rg < 0,20$) (Figura 2). O PCT apresentou associações fortemente positivas com os caracteres FLOR, AP, PMVc, MSf, MSc, HEMc, LIGc, FDNf, FDAf, HEMf, LIGf ($0,49 < rg < 0,95$) e associações fracas ou nulas com os caracteres PMVf, PMVp, FDNc, FDAc, CCc e CCf ($rg < 0,41$) (Figura 2).

O coeficiente de determinação (R^2) da análise de trilha foi de 90,66%, indicando que o PCT pode ser explicado pelos efeitos das características analisadas. O efeito residual foi de 30,56% e reflete a contribuição das variáveis do modelo para a variável básica. Portanto, o modelo explicativo utilizado expressou a relação de causa e efeito entre as variáveis primárias e o PCT.

As variáveis explicativas primárias MSc, FDNc, FDAc, HEMc, CCc, FDNf, FDAf, LIGf, HEMf e CCf influenciaram indiretamente o PCT pois, apesar da elevada associação positiva desses caracteres com o PCT, o efeito direto desses caracteres foi menor que o efeito residual (Figura 3 (A)). Logo, esses caracteres apresentaram pequeno efeito direto sobre o PCT, não sendo os principais determinantes nas alterações deste carácter, existindo outros caracteres que podem proporcionar maior impacto na seleção indireta, como as variáveis explicativas primárias MSf e LIGc. As variáveis explicativas primárias MSf e LIGc influenciaram diretamente o PCT, visto que o efeito direto foi superior ao efeito residual (Figura 3 (A)). As variáveis explicativas secundárias FLOR, AP e PMVc apresentaram elevada associação e efeito direto positivo com PCT, além de efeitos indiretos via outros caracteres (Figura 3 (B)). As demais variáveis secundárias, PMVp e PMVf, apresentaram influência sobre o PCT (Figura 3 (B))

Ao verificar os efeitos das variáveis explicativas secundárias sobre as primárias, observou-se, em média, coeficiente de determinação (R^2) de moderada magnitude, entre 27% e 82%, e elevado efeito residual, entre 42 e 86%, indicando que os variáveis explicativas secundárias não explicaram a contento a variação nos caracteres primários. Exceto, para os caracteres MSf e LIGc, os quais apresentaram elevado coeficiente de determinação ($>66\%$) e moderado efeito residual ($<58\%$). Para o MSf, as variáveis explicativas secundárias FLOR e AP apresentaram elevada associação e efeito direto positivo, podendo ser utilizadas na seleção indireta deste carácter e o carácter PMVc apresentou influência indireta, principalmente

devido ao efeito indireto via FLOR e AP (Figuras 2 e 4). Para a LIGc foi observado, apenas, influência indireta dos caracteres FLOR, AP e PMVc (Figura 4).

A árvore de regressão apresentou elevado coeficiente de determinação ($R^2 = 80\%$) (Figura 5). A estrutura da árvore foi composta pelo nó raiz o qual recebeu o carácter FLOR, doze nós de decisão, os quais foram compostos pelos caracteres AP (2), PMVc (4), MSc (2), MSf (1), LIGc (1), LIGf (1) e FDAf (1) e 14 nós folhas (Figura 5). Os caracteres considerados de maior importância para a construção da árvore foram FLOR, PMVc, LIGc, LIGf, MSc e MSf (Figura 6). Pode-se observar que o PCT foi sendo incrementado a medida em que o período vegetativo da cultura ($FLOR > 110$ dias), a produtividade e o teor de lignina do colmo apresentaram elevada expressão ($PMVc > 59,12 \text{ t ha}^{-1}$; $LIGc > 9,35\%$), obtendo quando seguido essas regras valores de PCT superior a 35 (Figura 5). Considerando o caminho apresentado anteriormente quando o LIGc apresentou valores inferiores a 9,35% o carácter PMVc apresentou grande influência na expressão do PCT, o qual apresentou maiores valores quanto o valor do PMVc foi superior a $65,71 \text{ t ha}^{-1}$.

Quando o período vegetativo foi superior a 110 dias porém a produtividade do colmo foi inferior a 59 t ha^{-1} observou-se menores valores para o PCT. Esses valores foram incrementados quando a LIGf foi inferior a 8,13% e MSc foi superior a 38,35%. Quando o período vegetativo foi inferior a 110 dias maiores valores de PMVc e de AP ajudaram a incrementar o valor do PCT (Figura 5).

Figura 1 - Herdabilidade (%), média (\bar{x}) e significância da variância genotípica pelo teste F, dos caracteres de sorgo biomassa.

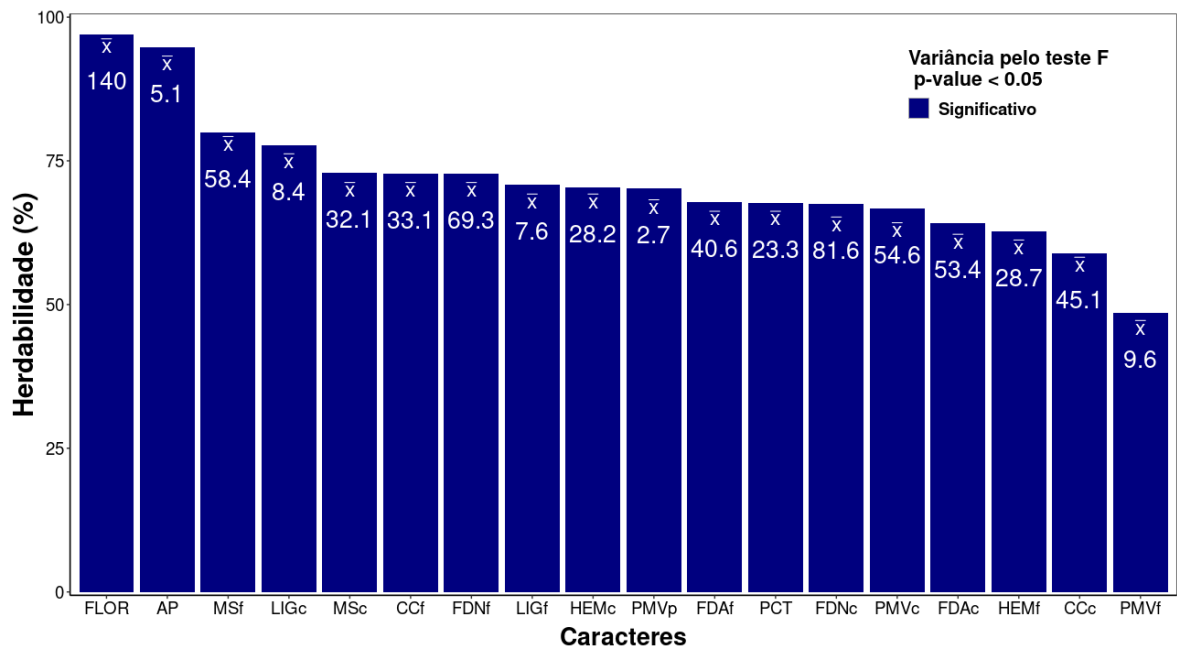
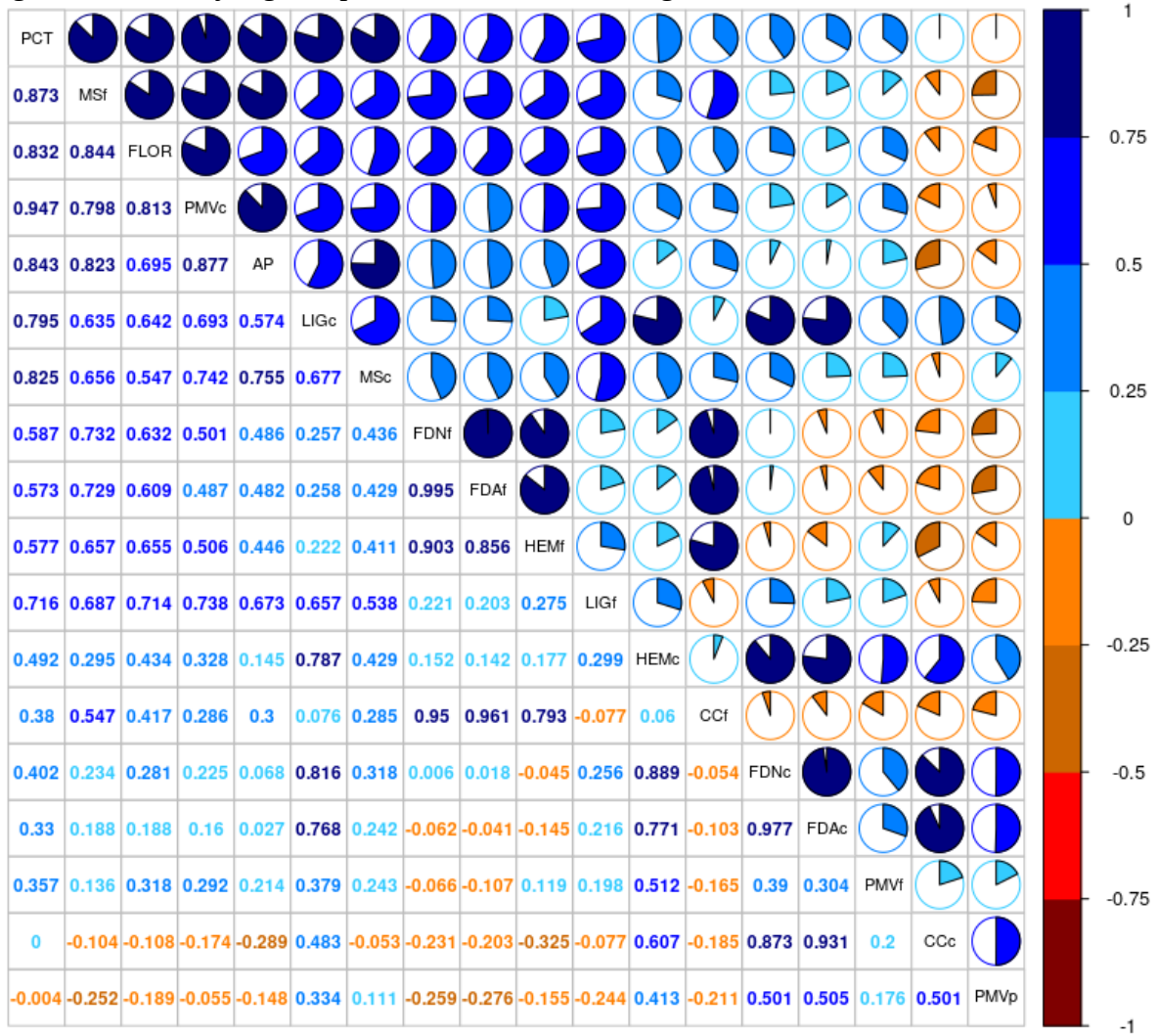
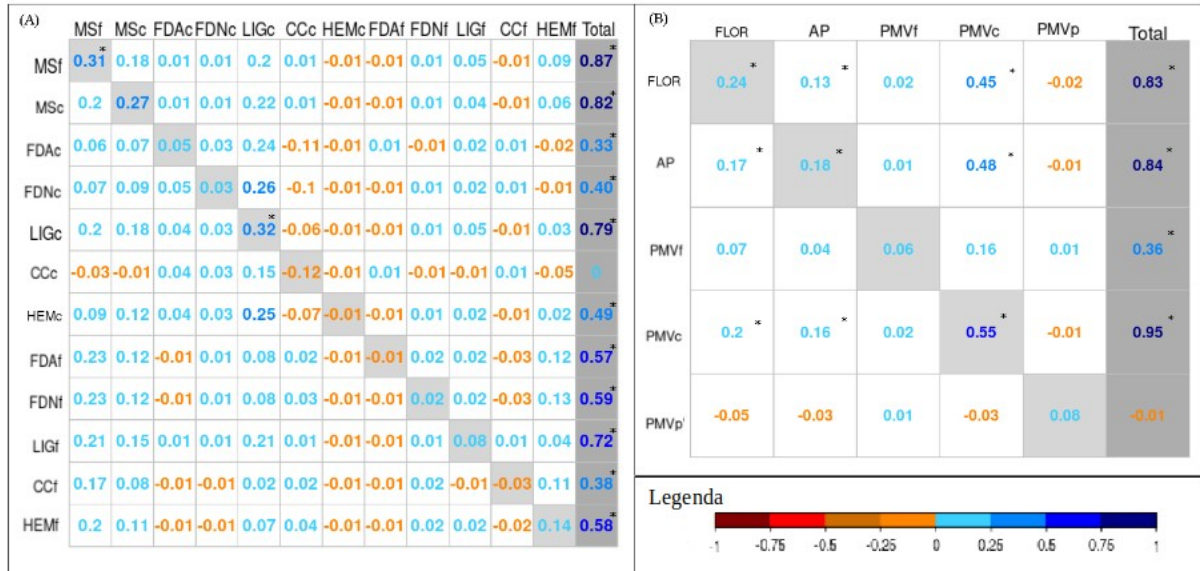


Figura 2 - Correlação genotípica dos caracteres de sorgo biomassa.



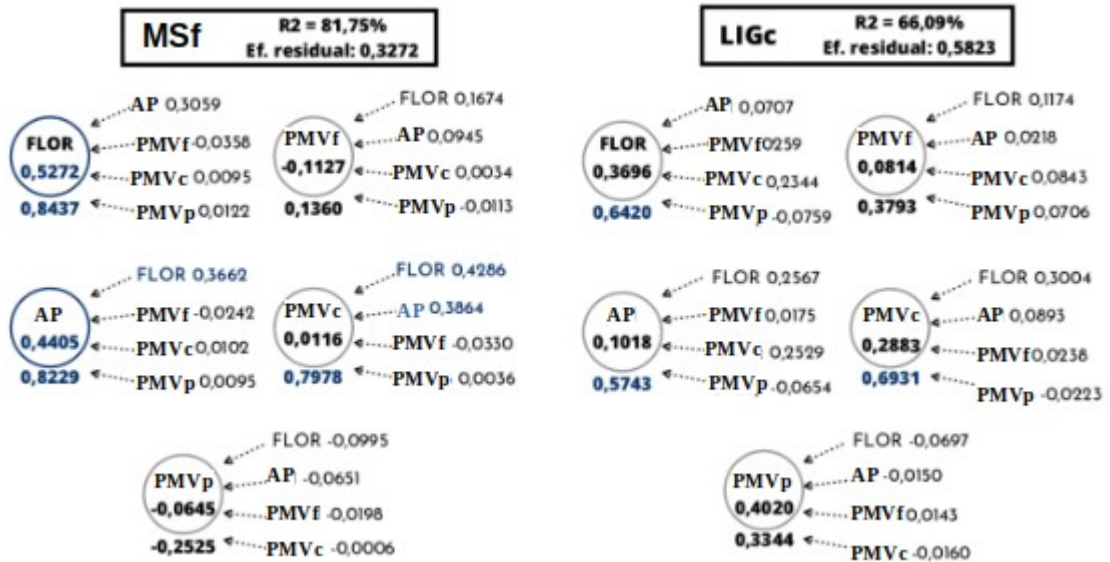
Fonte: Do autor (2021).

Figura 3 - Estimativas dos efeitos diretos (diagonal principal, cinza claro), efeitos indiretos (linha, branco) e efeito total (última coluna, cinza escuro) das variáveis primárias (A) e das variáveis secundárias (B) sobre a variável principal PCT de sorgo biomassa.



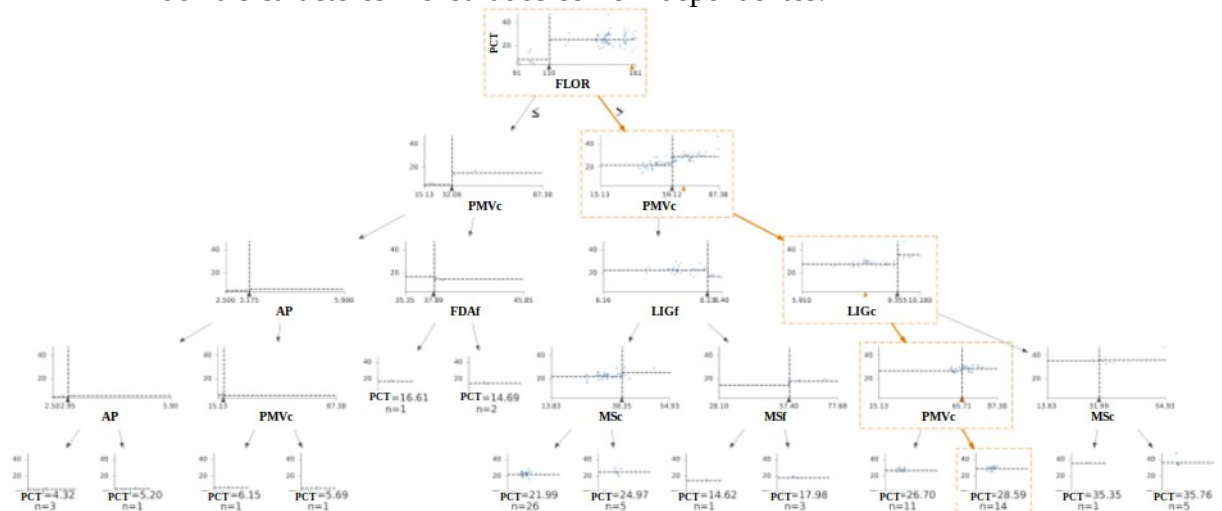
Fonte: Do autor (2021).

Figura 4 - Estimativas dos efeitos das variáveis explicativas secundárias sobre as variáveis primárias MSf e LIGc de sorgo biomassa.



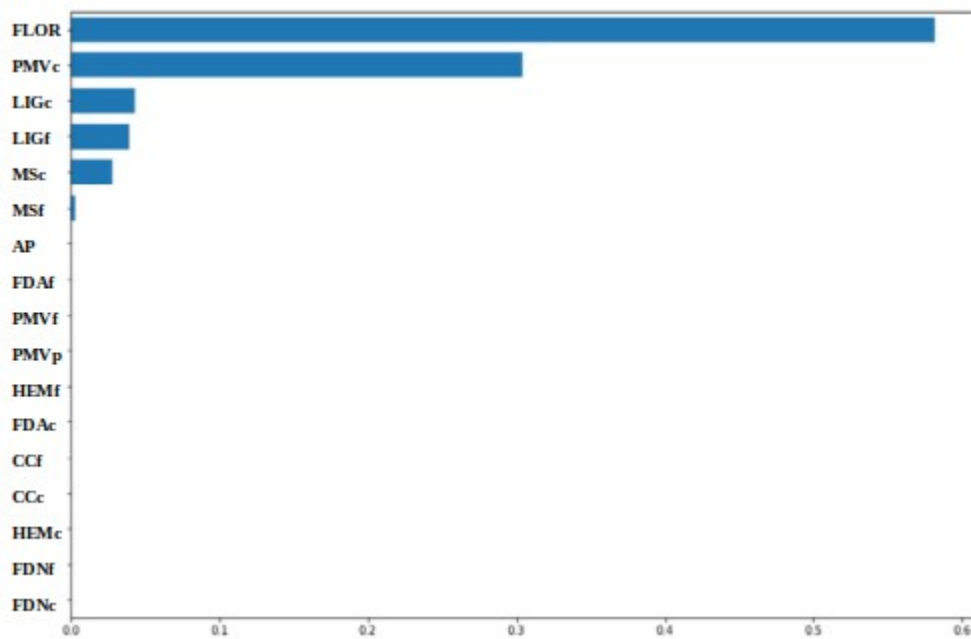
Fonte: Do autor (2021).

Figura 5 - Árvore de regressão ($R^2 = 80\%$) considerando o carácter PCT como alvo e os demais caracteres mensurados como independentes.



Fonte: Do autor (2021).

Figura 6 - Importância dos caracteres na árvore de regressão considerando o carácter PCT como carácter alvo.



Fonte: Do autor (2021).

4 DISCUSSÃO

Associações positivas entre os caracteres agronômicos FLOR, AP, PMV e MS já haviam sido observadas em sorgo biomassa (CASTRO et al., 2015; VENDRUSCOLO et al., 2016; FERNANDES et al., 2018; DOS SANTOS et al., 2020; HABYARIMANA et al., 2020) e podem ser explicadas pelo período vegetativo da cultura (sensibilidade ao fotoperíodo), índice de área foliar e interceptação e uso eficiente da radiação (OLSON et al., 2012;

CASTRO et al., 2015; CHUNG et al., 2017; OLIVEIRA et al., 2019). A sensibilidade ao fotoperíodo e a maturação (indução ao florescimento) da cultura são controlados pelos mesmos seis genes e dois alelos (Ma1,ma1; Ma2,ma2; Ma3,ma3; Ma4,ma4; Ma5,ma5; e Ma6,ma6), os quais possuem interações alélicas em que genótipos com constituição genética Ma1__ são sensíveis ao fotoperíodo, independentemente dos demais genes e, genótipos com constituição genética ma1ma1 são insensíveis ao fotoperíodo (ROONEY; AYDIN, 1999). O gene Ma1 está ligado com os genes que controlam a altura da planta (Dw1,dw1; Dw2,dw2; Dw3,dw3; e Dw4,dw4), principalmente, com o Dw2 (HILLEY et al., 2017). A altura de planta é determinada pelo número e comprimento de entrenós e influencia a espessura do diâmetro dos mesmos e, conseqüentemente, no maior acúmulo de biomassa, tendo maiores expressões no colmo e menores nos tecidos foliares e panícula (KEBRON et al., 2017).

As associações entre os caracteres FLOR, AP e PMV ocorrem devido a expressão genética e a ligação entre os genes que controlam a sensibilidade do fotoperíodo e a maturação com os genes que controlam a altura da planta. Em relação aos caracteres MSf e MSc, alguns estudos indicam a influência dos genes que controlam a expressão genética do FLOR e AP e dos genes relacionados a formação da parede celular na expressão desses caracteres (MCKINLEY et al., 2016; HILLEY et al., 2017), além da influência de fatores ambientais, como a incidência de luz (CHUNG et al., 2017; PERRIER et al., 2017). Maiores acúmulos de MSf e MSc ocorrem após a fase vegetativa da cultura, visto que, após a floração os genes envolvidos na formação da parede secundária e da lignificação (genes da família dos fatores de transcrição NAC (genes NAC)), que atuam, também, no alongamento da parede celular são, em partes, retidos no colmo do sorgo e contribuem para o aumento no diâmetro do mesmo e dos componentes estruturais, devido ao aumento da espessura das paredes secundárias (MCKINLEY et al., 2016). Enquanto, outros genes envolvidos no alongamento da parede celular influenciam na área e no ângulo da folha (HILLEY et al., 2017), permitindo maior interceptação e uso eficiente da radiação (CHUNG et al., 2017). O presente estudo confirmou a influência direta dos caracteres FLOR e AP e indireta do carácter PMVc na expressão genotípica do MSf (Figura 4), podendo os caracteres FLOR e AP e o carácter PMVc juntamente com os dois anteriores serem utilizados na seleção indireta do carácter MSf. A ausência do efeito direto dos caracteres agrônomicos para o MSc deve ser melhor avaliada, mas acredita-se que o resultado observado neste estudo, para este carácter, deve-se aos fatores genéticos, as interações entre esses fatores e seus efeitos nos processos fisiológicos e nas características morfológicas da planta (KEBROM et al., 2017; PERRIER et al., 2017).

Assim, a expressão genotípica dos genes que controlam os caracteres agronômicos, são os principais responsáveis pelas associações positivas observadas entre eles e pelos resultados observados em estudos relacionados ao desenvolvimento do sorgo biomassa, os quais apresentaram elevado incremento da expressão dos caracteres agronômicos ao longo de todo o desenvolvimento da cultura e menores incrementos após a floração da mesma, exceto para os caracteres PMVf e PMVp (OLSON et al., 2012; DIALLO et al., 2019; CASTRO et al., 2021). Os caracteres PMVf e PMVp, aparentemente, sofrem maior influência dos níveis de expressão dos genes que estão envolvidos na resposta dos hormônios vegetais, ou seja, daqueles que influenciam os aspectos fenológicos e fisiológicos da cultura (OLSON et al., 2012; TRUONG et al., 2015; SERRA et al., 2017; KEBROM et al., 2017) e, possivelmente, devido a isso, não apresentaram associações lineares com os demais caracteres agronômicos (Figura 2). A PMVf apresenta incremento moderado durante o período vegetativo e constante ou decréscimo após a floração (OLSON et al., 2012), período este em que a taxa de senescência foliar se iguala a taxa de crescimento, não apresentando incremento equivalente aos demais caracteres agronômicos avaliados. Enquanto a PMVp apresenta incremento acentuado após a floração, período em que os genótipos cessam o desenvolvimento vegetativo para priorizar a formação de grãos (SERRA et al., 2017).

Dentre o observado neste estudo quanto a associações dos caracteres agronômicos tem-se que as associações entre o FLOR, AP e MS com PMV verificados em outros estudos, devem-se, praticamente, as associações desses caracteres com a PMVc (Figura 2). A princípio, o conhecimento dessa associação não ocasionaria grandes problemas com o uso desses caracteres agronômicos para a seleção indireta da PMV, pois a PMVc representou, em média, 80% da PMV (Figura 1) podendo, inclusive, orientar os melhoristas na eliminação de alguns caracteres durante a seleção de genótipos superiores. Contudo, alguns estudos relataram que apesar da altura apresentar elevada associação com a produtividade de biomassa, a mesma não influenciou direta ou indiretamente este carácter (VENDRUSCOLO et al., 2016).

Considerando os caracteres que compõem a biomassa, associações positivas foram observadas entre os caracteres FDN, FDA, LIG, HEM e CC as quais, assim como as associações entre os caracteres agronômicos, já haviam sido observadas em sorgo biomassa (CASTRO et al., 2015; LI et al., 2017; ALMEIDA et al., 2019). Associações positivas entre FDN e FDA com LIG, HEM e CC eram esperadas visto que o FDN estima os componentes estruturais da parede vegetal que são insolúveis em detergente neutro como a LIG, HEM e CC e o FDA estima os componentes estruturais da parede vegetal insolúveis em detergente ácido

como a LIG e a CC (SOEST, 1963), sendo interessante entender as associações entre LIG, HEM e CC, os quais são os principais componentes da parede celular. As associações entre LIG, HEM e CC foram positivas e podem ser explicadas pela expressão dos genes NAC (XIA et al., 2018) e pela estrutura e ligações entre esses caracteres durante a constituição da parede celular as quais foram elucidadas em estudos anteriores (ATALLA et al., 1993; GRABBER et al., 2004; JUNG et al., 2015; RAVINDRAN; JAISWAL, 2016). Resumidamente, os genes NAC induzem o espessamento da parede celular secundária e a deposição de CC, HEM e LIG no colmo (ZHONG et al., 2010; XIA et al., 2018). Estruturalmente, tem-se que a LIG reticula a CC e HEM, formando uma matriz altamente impermeável e complexa, que protege os polissacarídeos da parede celular, no caso CC e HEM, da degradação (GRABBER et al., 2004) e somando a isto, tem-se que os polímeros de HEM se ligam a CC e a LIG por meio de ligações de hidrogênio e ligações covalentes, respectivamente (RAVINDRAN; JAISWAL, 2016). Assim sendo, a expressão dos caracteres LIG, HEM e CC variam no mesmo sentido, principalmente, no colmo. As associações nulas observadas entre HEM_f e CC_f com a LIG_f pode ser decorrente da menor lignificação foliar a qual apresenta menor relação LIG:CC e LIG:HEM quando comparado com os constituintes da parte do colmo (Figura 1) (SARATH et al., 2008; MCKINLEY et al., 2018).

Apesar da associação observada entre os caracteres que compõem a biomassa, as associações entre os caracteres que compõem a biomassa do colmo com aqueles que compõem a biomassa da folha foram consideradas nula (Figura 2). Isto ocorre, possivelmente, devido aos efeitos dos fatores genéticos e ambientais na fisiologia e morfologia dos tecidos vegetais do colmo e da folha (FIRDOUS; GILANI, 2001; SARATH et al., 2008; PERRIER et al., 2017). As folhas do sorgo biomassa diferenciam do colmo quanto a expressão dos compostos lignocelulósicos ao longo do desenvolvimento da cultura; apresentando, no final do ciclo, maiores teores de água e proteínas e menores teores de CC e LIG em relação ao colmo (FIRDOUS; GILANI, 2001; SARATH et al., 2008; MCKINLEY et al., 2018). Uma possível explicação genética seria os níveis de expressão dos genes que estão envolvidos na resposta dos hormônios vegetais (auxina, citocinina, giberilina e etileno, entre outros) (TRUONG et al., 2015; KEBROM et al., 2017), tendo, por exemplo, maior expressão do etileno no colmo e da auxina na folha, porém não foi encontrado informações na literatura sobre as causas genéticas que explicassem de forma mais clara a ausência de associação linear entre os caracteres que compõem a biomassa do colmo e da folha no sorgo biomassa.

Vale destacar, também, uma outra possível explicação para a associação nula entre os caracteres que compõem a biomassa do colmo com os que compõem a biomassa da folha que

é a associação negativa entre os componentes estruturais (CC, HEM e LIG) e os não estruturais (proteínas, minerais, entre outros) na composição da biomassa da cultura (PERRIER et al., 2017; MCKINLEY et al., 2018). Assim, como as folhas necessitam de maiores concentrações de componentes não estruturais para realização dos seus processos fisiológicos, há uma menor concentração dos componentes estruturais na mesma, diferentemente do colmo (FIRDOUS; GILANI, 2001; SARATH et al., 2008; MCKINLEY et al., 2018).

É importante lembrar que as associações nulas, observadas neste estudo, podem ser decorrentes da colheita uniforme dos genótipos (~ 180 dias após a semeadura), tendo assim diferenças entre os estádios fenológicos de maturação dos genótipos avaliados, havendo genótipos no início do período de maturação e outros no final do período de maturação. Esse fato, provavelmente, influenciou o nível de expressão genotípica de alguns genótipos para alguns caracteres. Além disso, ressalta-se que as associações nulas, como as observadas entre os componentes da biomassa do colmo com os da folha e entre os caracteres agronômicos com a PMVf e PMVp, não necessariamente significam ausência de associação entre os caracteres, mas apenas que esses caracteres não apresentam associação linear entre si (Figura 2).

A ausência da associação linear entre a maioria dos caracteres agronômicos com os caracteres que compõem a biomassa contrapõem as observações sobre a expressão desses caracteres ao longo do desenvolvimento da cultura, porém pode ser explicada pela expressão genética de alguns genes que atuam nesses caracteres, como os genes NAC e os genes que controlam a resposta dos hormônios (FIRDOUS; GILANI, 2001; OLSON et al., 2012; HOFFMANN; ROONEY, 2014; TRUONG et al., 2015; KEBROM et al., 2017; MCKINLEY et al., 2018; DIALLO et al., 2019; CASTRO et al., 2021). Os genes NAC, que codificam cerca de 180 proteínas no sorgo, estão envolvidos em vários processos de desenvolvimento da cultura, como o desenvolvimento dos meristemas apicais, o transporte de auxinas, a divisão celular, o espessamento da parede secundária e a resposta da planta a estresses bióticos e abióticos (XIA et al., 2018; SANJARI et al., 2019) e apresentam diferentes níveis de expressão ao longo do desenvolvimento da cultura, inclusive entre os caracteres agronômicos e os caracteres que compõem a biomassa (MCKINLEY et al., 2016; PIERRER et al., 2017). Por isso que a maioria dos caracteres agronômicos não foram os principais determinantes das variações nos caracteres que compõem a biomassa, apesar desses caracteres apresentarem comportamento similar ao longo do desenvolvimento da cultura (Figura 4), exceto para o carácter LIGc (FIRDOUS; GILANI, 2001; OLSON et al., 2012; HOFFMANN; ROONEY,

2014; CASTRO et al., 2021). Esses conhecimentos implicam que os caracteres que compõem a biomassa podem ser melhorados independente dos caracteres agronômicos. Isso tem implicações importantes para os programas de melhoramento de sorgo biomassa os quais podem obter genótipos com maiores teores de compostos lignocelulósicos independente da expressão dos caracteres agronômicos.

O carácter LIGc diferente da maioria dos caracteres que compõem a biomassa apresentou associação positiva com os caracteres FLOR, AP, PMVc, MSc e MSf (Figura 2) sendo influenciado indiretamente pelos caracteres FLOR, AP e PMVc (Figura 4). Este resultado demonstrou que realizar a seleção para este carácter com base apenas nas associações positivas observadas com os caracteres agronômicos ($0,33 < r_g < 0,70$) (Figura 2), poderá não resultar ganhos satisfatórios com a seleção. Neste caso, a melhor estratégia seria a seleção simultânea de caracteres, principalmente, os caracteres FLOR, AP e PMVc, cujos efeitos indiretos são significativos (CRUZ et al., 2012).

O PCT, principal carácter nos estudos com sorgo biomassa para geração de energia elétrica, apresentou associações positivas com a maioria dos caracteres avaliados neste estudo porém, somente as variáveis explicativas primárias MSf e LIGc e as variáveis explicativas secundárias FLOR, AP e PMVc foram consideradas as principais determinantes das alterações do PCT podendo ser utilizadas na seleção indireta deste carácter (Figura 2 e 3). Resultado similar foi observado por meio da árvore de regressão a qual, apesar de não indicar a causa de variação do PCT, atribuiu elevada importância aos caracteres FLOR, PMVc, LIGc e MSf em que genótipos com maiores expressões para estes caracteres apresentaram maiores valores de PCT (Figura 6). Esse resultado, em partes, era esperado visto que quanto maior a energia liberada pela biomassa da cultura quando queimada totalmente na presença de oxigênio maior será o poder calorífico (SIMEONE et al., 2020) e que alguns estudos já demonstraram que a energia liberada será maior na presença de menor umidade e de maior teor de lignina na biomassa (ABBAS et al., 2020; SIMEONE et al., 2020).

A umidade da biomassa interfere diretamente na temperatura de combustão podendo, quando elevada, influenciar negativamente na otimização da combustão e no valor calorífico, pois parte do calor gerado é consumido na evaporação da água e também no aquecimento dos vapores para atingir a temperatura dos demais gases (WU et al., 2011; CASTRO et al., 2015; SIMEONE et al., 2020). Então quanto menor a umidade da biomassa maior será a porcentagem de massa seca (MS) ($UM\% = 100 - MS\%$) e maior será o PCT (CASTRO et al., 2015; ABBAS et al., 2020; SIMEONE et al., 2020). Dessa forma, esperava-se que o MS do colmo e folha influenciassem diretamente o PCT. Inclusive, pela árvore de regressão pode-se

observar que maiores valores de MSc influenciaram positivamente os valores de PCT, principalmente, quando houve menores valores de PMVc ($PMVc < 59,12 \text{ t ha}^{-1}$). Logo o MSc compensou, em partes, a variação do PCT quando os genótipos apresentaram menores PMVc (Figura 5). Acredita-se que a ausência do efeito direto do MSc sobre PCT deve-se aos diferentes níveis de maturação dos genótipos avaliados e aos fatores genéticos, ambientais e as interações entre esses fatores na expressão desse carácter (KEBROM et al., 2017; PERRIER et al., 2017).

O carácter LIG apresenta elevada resistência a degradação térmica, devido ao elevado número de carbonos presentes em sua estrutura e baixo teor de oxigênio, se comparado a CC e HEM (XIAO et al., 2001; CASTRO et al., 2015; SIMEONE et al., 2020). O sorgo biomassa apresenta poder calorífico em torno de $4500 \text{ Kcal Kg}^{-1}$ de biomassa seca enquanto a LIG apresenta poder calorífico em torno de $6100 \text{ Kcal kg}^{-1}$ de biomassa seca (PETROFF; DOAT, 1978; CASTRO et al., 2015), o que justifica a influência deste carácter na variação do PCT. As maiores concentrações de LIG no colmo, principalmente, devido as diferenças nos níveis de lignificação dessas estruturas e da atuação dos genes que controlam o enlogamento da parede celular e os genes que estão envolvidos na resposta dos hormônios vegetais podem justificar maior influência da LIGc do que da LIGf na variação do PCT.

E por fim, em relação as variáveis explicativas secundárias, não foi observado em estudos anteriores como esses caracteres (FLOR, AP e PMV) influenciam a quantidade de energia liberada durante a combustão sendo, neste estudo, considerados determinantes nas variações do PCT (Figuras 3 e 4). Acredita-se que esses caracteres agronômicos influenciam diretamente o PCT devido a expressão dos mesmos, principalmente, pelo fato dos caracteres FLOR e AP estarem ligados geneticamente (genes Ma1 e genes Dw) e influenciarem a expressão do PMVc e esses caracteres agronômicos influenciarem direta ou indiretamente a expressão de alguns caracteres que influenciam diretamente a quantidade de energia liberada durante a combustão, como os caracteres MSf e LIGc (Figuras 3 e 4). A relação entre os caracteres PMVc e LIGc pode ser observado pela árvore de regressão, em que genótipos que apresentaram PMVc superior a 59 t ha^{-1} porém baixo teor de LIGc ($LIGc < 9,35\%$) tiveram o valor de PCT incrementado quando considerado maiores PMVc ($PMVc > 65 \text{ t ha}^{-1}$) para a cultura, demonstrando que menores LIGc são compensados pela maior PMVc (Figura 5).

5 CONCLUSÃO

Os caracteres agronômicos apresentaram associações positivas entre si e associações nulas com PMVf e PMVp. Os caracteres que compõem a biomassa do colmo, assim como os

caracteres que compõem a biomassa da folha, FDN, FDA, LIG, HEM e CC, apresentaram associações positivas entre si, porém as associações entre os caracteres que compõem a biomassa do colmo com os que compõem a biomassa da folha foram consideradas nulas.

As associações entre os caracteres agrônômicos e os caracteres que compõem a biomassa apresentaram tanto associações nulas quanto positivas. Destacam-se as associações positivas entre a LIGc com o FLOR, AP, PMVc, MSc e MSf. O PCT também apresentou tanto associações nulas quanto positivas com os caracteres avaliados. Destacando-se as associações positivas com FLOR, AP, PMVc, MSf, MSc, HEMc, LIGc, FDNf, FDAf, HEMf, LIGf.

Os caracteres FLOR, AP, PMVc, MSf e LIG foram consideradas os principais determinantes das alterações do PCT podendo ser utilizados na seleção indireta deste carácter. Além disso, os caracteres FLOR, AP e PMVc influenciarem direta ou indiretamente a expressão dos caracteres MSf e LIGc, podendo para a seleção indireta dos caracteres MSf e LIGc, levar em consideração a seleção indireta ou simultânea, dependendo do carácter alvo, dos caracteres FLOR, AP e PMVc.

A técnica de Mineração de Dados empregada permitiu melhor visualização das relações entre os caracteres, auxiliando em algumas explicações sobre a influência dos caracteres sobre o PCT e somando a isto diagnosticou como caracteres importantes para a construção da árvore todos aqueles que foram determinantes das alterações do PCT, exceto o AP. Sendo assim, interessante seu uso combinado com as análises de correlação e de trilha para melhor entendimento das relações entre os caracteres.

REFERÊNCIAS

- ABBAS, T.; ISSA, M.; ILINCA, A. Biomass cogeneration Technologies: A Review. **Scientific Research**, v. 10, n.1, p.1-14. 2020.
- ALMEIDA, L.G.F.; PARRELLA, R.A.C.; SIMEONE, M.L.F.; RIBEIRO, P.C.O.; SANTOS, A.S.; COSTA, A.S.V.; GUIMARÃES, A.G.; SCHAFFERT, R.E. Composition and growth of sorghum biomass genotypes for ethanol production. **Biomass and Bioenergy**, v.122, p.343-348. 2019.
- ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; GONÇALVES, J.L.M.; SPAROVEK, G. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v.22, p.711-728. 2013.
- ATALLA, R.H.; HACKNEY, J.M.; UHLIN, I.; THOMPSON, N.S. Hemicelluloses as structure regulators in the aggregation of native cellulose. **International Journal of Biological Macromolecules**, v.15,n.2, p.109–112. 1993.
- BLAKERS, A.; LU, B.; STOCKS, M. 100% renewable electricity in Australia. **Energy**, v.133, p.471-482. 2017.
- BORGES, L.S.; AQUINO, F. C.; EVANGELISTA, A. F. Potencial do capim elefante para geração de bioenergia – revisão. **Revista Eletrônica**, v.13, p.4518-4523. 2016.
- BREIMAN, L.; FRIEDMAN, J.; STONE, C.J.; OLSHEN, R.A. Classification and regression trees. **CRC press**, v.2, p.13-16. 1984.
- CASTRO, F.M.R., LOMBARDI, G.M.R., NUNES, J.A.R., PARRELLA, R.A.C., BRUZI, A.T. Accumulation of biomass and lignocellulosic compounds in photoperiod-sensitive biomass sorghum genotypes. **Biomass and Bioenergy**. 2021 (prelo)
- CASTRO, F.M.R.; BRUZI, A.T.; NUNES, J.A.R.; PARRELLA, R.A.C.; LOMBARDI, G.M.R.; ALBUQUERQUE, C.J.B.; LOPES, M. Agronomic and Energetic Potential of Biomass Sorghum Genotypes. **American Journal of Plant Science**, v. 6,p1-14. 2015.
- CHUNG, Y. S.; CHOI, S. C.; SILVA, R. R.; KANG, J. W.; EOM, J. H.; KIM, C. Case study: Estimation of sorghum biomass using digital image analysis with Canopeo. **Biomass and Bioenergy**, v.105, p.207–210. 2017.
- CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético v1..** 4.ed. Viçosa: Editora UFV, 2012. 514p.
- DA SILVA, M.J., CARNEIRO, P.C.S., CARNEIRO, J.E.S., DAMASCENO, C.M.B., PARRELLA, N.N.L.D., PASTINA, M.M., SIMEONE, M.L.F., SCHAFFERT, R.E., PARRELLA, R.A.C. Evaluation of the potential of lines and hybrids of biomass sorghum. **Ind. Cro. and Prod**, v.125, p.379-385. 2018.
- DA SILVA, M.J., DAMASCENO, C.M.B., CARNEIRO, J.E.S., PEREIRA, H.D., CARNEIRO, C.S., SCHAFFERT, R.E., PARRELLA, R.A.C. Combining ability of biomass

sorghum in different crop years and sites for bioenergy generation. **Agron. J.** 12(3):1549-1563. 2020a.

DA SILVA, M.J., DAMASCENO, C.M.B., GUIMARÃES, C.T., PINTO, M.O., BARROS, B.A., CARNEIRO, J.E.S., SCHAFFERT, R.E., PARRELLA, R.A.C. Introgression of the bmr6 allele in biomass sorghum lines for bioenergy production. **Euphytica**, v.216, n.95, p.1-12. 2020b.

DERMIBAS, A. Biomass resource facilities and biomass conversion processing for fuels and chemicals. **Energy Conversion and Management**, v.42, n.11, p.1357–1378. 2001.

DIALLO, B., LI, M., TANG, C., AMEEN, A., ZHANG, W., XIE, G. H. Biomass yield, chemical composition and theoretical ethanol yield for different genotypes of energy sorghum cultivated on marginal land in China. **Industrial Crops and Products**, v.137, p.221–230. 2019.

DOS SANTOS, J.P.R., FERNANDES, S.B., MCCOY, S., LOZANO, R., BROWN, P.J., LEAKEY, A.D.B., BUCKLER, E.S., GARCIA, A.A.F., GORE, M.A. Novel Bayesian Networks for Genomic Prediction of Developmental Traits in Biomass Sorghum. **G3 (Bethesda)**, v.10, n.2, p.769-781. 2020.

FERNANDES, S.B.; DIAS, K.O.G.; FERREIRA, D.F.; BROWN, P.J. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. **Theoretical and Applied Genetics**, v.131, n.3, p.747–755. 2018.

FERREIRA, S., MONTEIRO, E., BRITO, P., VILARINHO, C. Biomass resources in Portugal: Current status and prospects. **Renewable and Sustainable Energy Reviews**, v.78, p.1221–1235. 2017.

FIRDOUS, R.; GILANI, A. H. “Changes in Chemical Composition of Sorghum as Influenced by Growth Stage and Cultivar” Asian-Australasian Journal of Animal Sciences. **Asian Australasian Association of Animal Production Societies**, v.14, n.7, p. 935–940. 2001.

GOSWAMI, S.J.; PATEL, P.T.; GAMI, R.A.; PATEL, R.N.; KHATRI, A.B. Correlation and path analysis study of different characters for grain yield and fodder purpose in sorghum [Sorghum bicolor (L.) Moench]. **Electronic Journal of Plant Breeding**, v.11, n.4, p.1-9. 2020.

GRABBER, J.H., RALPH, J., LAPIERRE, C., BARRIÈRE, Y. Genetic and molecular basis of grass cell-wall degradability. I. Lignin–cell wall matrix interactions. *C. R. Biol*, v.327, p.455–465. 2004.

HABYARIMANA, E., LOPEZ-CRUZ, M., BALOCH, F.S. Genomic Selection for Optimum Index with Dry Biomass Yield, Dry Mass Fraction of Fresh Material, and Plant Height in Biomass Sorghum. **Genes**, v.11, n.1, p.61-68. 2020.

HEITMAN, A.J., M.S. CASTILLO, T.J. SMYTH, C.R. CROZIER, Z. WANG, R.W. HEINIGER, R.J. GEHL. Nitrogen fertilization effects on dry matter yield and nutrient removal of biomass- and sweet-sorghum. **Agron. J.**, v.109, p.1352–1358. 2017.

HILLEY, J.L., WEERS, B.D., TRUONG, S.K. ET AL. Sorghum Dw2 Encodes a Protein Kinase Regulator of Stem Internode Length. **Sci Rep**, v.7, p.4616 -4629. 2017.

HOFFMANN, L., ROONEY, W.L. Accumulation of Biomass and Compositional Change Over the Growth Season for Six Photoperiod Sorghum Lines. **Bioenerg. Res**, v.7, p.811–815. 2014.

JUNG, S.J., KIM, S.-H., CHUNG, I.-M. Comparison of lignin, cellulose, and hemicellulose contents for biofuels utilization among 4 types of lignocellulosic crops. **Biomass and Bioenergy**, v.83, p.322–327. 2015.

KAMBO, H.S., DUTTA, A. Strength, storage, and combustion characteristics of densified lignocellulosic biomass produced via torrefaction and hydrothermal carbonization. **Applied Energy**, v.135, p.182-191. 2014.

KEBROM, T.H., MCKINLEY, B., MULLET, J.E. Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. **Biotechnol Biofuels**, v.10, p.159-167. 2017.

LI, J.; HENRIKSSON, G.; GELLRSTEDT, G. Lignin depolymerization/repolymerization and its critical role for delignification of aspen wood by steam explosion. **Bioresource Technology**, v. 98, p.3061-3068, 2007.

LI, M., WANG, J., DU, F., DIALLO, B., XIE, G.H. High-throughput analysis of chemical components and theoretical ethanol yield of dedicated bioenergy sorghum using dual-optimized partial least squares calibration models. **Biotechnol Biofuels**, v.10, 206-218. 2017.

LIANG, X., HOOGENBOOM, G., VOULGARAKI, S., BOOTE, K.J., VELLIDIS, G. Deriving genetic coefficients from variety trials to determine sorghum hybrid performance using the CSM–CERES–Sorghum model. **Agron. J.** Online Version. 2021

LOMBARDI, G.M.R.; NUNES, J.A.R.; PARRELLA, R.A.C.; TEIXEIRA, D.H.L.; BRUZI, A.T.; DURÃES, N.N.L.; FAGUNDES, T.G.; Path analysis of agro-industrial traits in sweet sorghum. **Genetic and Molecular Research**, v.14, p.16392-16402. 2015.

LOUPPE G. **Understanding random forests: From theory to practice**. Tese de Doutorado, Universidade de Lieja. 2014.

LUTZ, M. **Python Pocket Reference: Python in your pocket**, 5th Edition. O'Reilly Media, Inc. 2014.

MCKINLEY, B. A., OLSON, S. N., RITTER, K. B., HERB, D. W., KARLEN, S. D., LU, F., ... MULLET, J. E. Variation in energy sorghum hybrid TX08001 biomass composition and lignin chemistry during development under irrigated and non-irrigated field conditions. **PLOS ONE**, 13(4). 2018.

MCKINLEY, B., ROONEY, W., WILKERSON, C., MULLET, J. Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of Sorghum bicolor. **The Plant Journal**, v.88, n.4, p.662–680. 2016.

- MENDIBURU, F. **agricolae: Statistical Procedures for Agricultural Research**. R package version 1.3-3. <https://CRAN.R-project.org/package=agricolae>. 2020.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York : J. Wiley, 1981.
- MOOD, S.H., GOLFESHAN, A.H., TABATABAEI, M., JOUZANI, G.S., NAJAFI, G.H., GHOLAMI, M., GHOLAMI, M., ARDJMAND, M. Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. **Renewable and Sustainable Energy Reviews**, v.27, p.77–93. 2013.
- NHUCHHEN, D.R., AFZAL, M.T. HHV Predicting Correlations for Torrefied Biomass Using Proximate and Ultimate Analyses. **Bioengineering**, v.4, n.7, p.1-15. 2017.
- OBERMEYER, Z.; EMANUEL, E.J. Predicting the future—big data, machine learning, and clinical medicine. The New England Journal of Medicine, **NIH Public Access**, v.13, p.375-391. 2016.
- OLIVEIRA, A.A., PASTINA, M.M., DE SOUZA, V.F., PARRELLA, R.A.C., NODA, R.W., SIMEONE, M.L.F., SCHAFFERT, R.E., MAGALHÃES, J.V., DAMASCENO, C.M.B., MARGARIDO, G.R.A. Genomic prediction applied to high-biomass sorghum for bioenergy production. **Mol Breeding**, v.2, p.38:49. 2018.
- OLIVEIRA, I.C.M.; MARÇAL, T. de S.; BERNARDINO, K. da C.; RIBEIRO, P.C. de O.; PARRELLA, R.A. da C.; CARNEIRO, P.C.S.; SCHAFFERT, R.E.; CARNEIRO, J.E. de S. Combining ability of biomass sorghum lines for agroindustrial characters and multitrait selection of photosensitive hybrids for energy cogeneration. **Crop Science**, v.59, p.1554-1566, 2019.
- OLIVEIRA, T.C., BARELLI, M.A.A., SILVA, V.P., FELIPIN-AZEVEDO, R., GONÇALVES, D.L., SANTODS, P.R.J., POLETINE, J.P., GALBIATI, C., TARDIN, F.D. Correlations between characters and path analysis in sweet sorghum (*Sorghum bicolor* (L.) Moench) genotypes for juice production. **AJCS**, v.15, n.02, p.290-296. 2021.
- OLIVOTO, T., LÚCIO, A.D. metan: an R package for multi-environment trial analysis. **Methods Ecol Evol**, v.11, p.783-789. 2020.
- OLIVOTO, T., SOUZA, V.Q., NARDINO, M., CARVALHO, I.R., FERRARI, M., PELEGRIN, A.J., SZARESKEI, V.J., SCHMIDT, D. Multicollinearity in Path Analysis: A Simple Method to Reduce Its Effects. **Agron. J**, v.109, n.1, p.131-142. 2017.
- OLSON, S.N., RITTER, K., ROONEY, W., KEMANIAN, A., MCCARL, B. A., ZHANG, Y., HALL, S., PACKER, D., MULLET, J. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. **Biofuels, Bioproducts and Biorefining**, v.6, n.6, p.640–655. 2012.
- PARKHURST, K.M., SAFFRON, C.M., MILLER, R.O. An energy analysis comparing biomass torrefaction in depots to wind with natural gas combustion for electricity generation. **Applied Energy**, v.179, p.171–181. 2016.

PATEL, B.; PATEL, A.; GAMI, B.; PATEL, P. Energy balance, GHG emission and Economy for cultivation of high biomass varieties of Bamboo, Sorghum and Pearl millet as energy crops at marginal ecologies of Gujarat State in India. **Renewable Energy**, v.148, p.816-823. 2019.

PAULA, L.E.R.; TRUGILHO, P.F.; NAPOLI, A.; BIANCHI, M.L. Characterization of residues from plant biomass for use in energy generation. **Cerne**, v.17, n.2, p.237-246. 2011.

PEDREGOSA, F., VAROQUAZ, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **JMLR**, v.12, p.2825-2830. 2011.

PERRIER, L., ROUAN, L., JAFFUEL, S., CLÉMENT-VIDAL, A., ROQUES, S., SOUTIRAS, A., ... LUQUET, D. Plasticity of Sorghum Stem Biomass Accumulation in Response to Water Deficit: A Multiscale Analysis from Internode Tissue to Plant Level. **Frontiers in Plant Science**, v.8, p.1-8. 2017.

PETROFF, G.; DOAT, J. Pirolise des bois tropicaux: influence de la composition chimique des bois sur les produits de distillation. **Bois et forêts de tropiques**, v.117, p.51-64,1978.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria. 2018.

RAVINDRAN, R., JAISWAL, A. K. A comprehensive review on pre-treatment strategy for lignocellulosic food industry waste: Challenges and opportunities. **Bioresource Technology**, v.199, p.92–102. 2016.

RESENDE, M.D.V., DUARTE, J.B. Precision and quality control in variety trials. **Pesqui. Agropec. Trop**, v.37, p.182–194. 2007.

ROONEY, W.L.; AYDIN, S. Genetic control of a photoperiod-sensitive response in Sorghum bicolor (L.) Moench. **Crop Science**, v. 9, p. 397-400. 1999.

SAINI, P., KOFF, J.P., ALLISON, A., HAMILTON, C. Changes in Lignocellulosic Polymers, Carbon, and Energy in Switchgrass for Bioenergy Production. **Agron. J.**, v.109, n.5, p.1935-1943. 2017.

SANJARI, S., SHIRZADIAN-KHORRAMABAD, R., SHOBBAR, ZS. ET AL. Systematic analysis of NAC transcription factors' gene family and identification of post-flowering drought stress responsive members in sorghum. **Plant Cell Rep**, v.38, p.361–376. 2019.

SARATH G, MITCHELL RB, SATTLER SE, FUNNELL D, PEDERSEN JF, GRAYBOSCH RA ET AL. Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. **J Ind Microbiol Biotechnol**, v.35, n.5, p.343–354.2008.

SCIKIT-LEARN. **User Guide: Ensemble methods**. 2020. Disponível em: <https://scikit-learn.org/stable/modules/ensemble.html>.

- SERRA, P., COLAUZZI, M., AMADUCCI, S. Biomass sorghum production risk assessment analysis: A case study on electricity production in the Po Valley. **Biomass and Bioenergy**, v.96, p.75–86. 2017.
- SIMEONE, M.L.F., PARRELLA, R.A.C., DAMASCENO, C.M.B., SCHAFFERT, R.E. Prediction of high-biomass sorghum quality using near infrared spectroscopy to monitoring calorific value, moisture, and ash content. **International Journal of Development Research**, v.10, n.9, p.40916-40920. 2020.
- SINHA, S., KUMARAVADIVEL, N. "Understanding Genetic Diversity of Sorghum Using Quantitative Traits", **Scientifica**, v.2016, p.1-8. 2016.
- SOEST, P.J.V. Use of detergents in the analysis of fibrous feeds. II. A rapid method for the determination of fiber and lignin. **Journal of the A.O.A.C.**, v.46. n.5, p.1-14. 1963.
- SOIL SURVEY STAFF. **Keys to Soil Taxonomy**. United States Department of Agriculture. 2014.
- STEPHENSON, T.D., CARVALHO, H.D.R., CASTILLO, M.S., CROZIER, C.R., SMYTH, T.J., HEITMAN, J.L. Water use and biomass yield of bioenergy crops in the North Carolina Piedmont. **Agron. J.** Online Version. 2021.
- TRUONG, S. K., MCCORMICK, R. F., ROONEY, W. L., MULLET, J. E. Harnessing Genetic Variation in Leaf Angle to Increase Productivity of Sorghum bicolor. **Genetics**, v.201, n.3, p.1229–1238. 2015.
- VENDRUSCOLO, T.P.S.; BARELLI, M.A.A.; CASTRILLON, M.A.S.; SILVA, R.S.; OLIVEIRA, F.T.; CORREA, C.L.; ZAGO, B.W.; TARDIM, F.D. Correlation and path analysis of biomass sorghum production. **Genetics and Molecular Research**, v.15, n.4, p.1-16, 2016.
- WRIGHT, S. Correlation and causation. **Journal of Agricultural Research**, v.20, p.557-585. 1921.
- WU, M.R., SCHOTT, D.L., LODEWIJKS, G. Physical properties of solid biomass. **Biomass and Bioenergy**, v.35, p.2093-2105. 2011.
- XIA, J., ZHAO, Y., BURKS, P., PAULY, M., BROWN, P. J. A sorghum NAC gene is associated with variation in biomass properties and yield potential. **Plant Direct**, v.2, n.7, p.70-78. 2018.
- XIAO, B.; SUN, F.; SUN, R.C. The chemical modification of lignins with succinic anhydride in aqueous systems, **Polymer Degradation and Stability**, v.71, n.2, p.223-231, 2001.
- ZEILEIS, A., HOTHORN, T. Diagnostic Checking in Regression Relationships. **R News**, v.2, n.3, p.7-10. 2002.
- ZHAO, Y.L.; A. DOLAT, Y. STEINBERGER, X.; WANG, A.; OSMAN; XIE, G.H. Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. **Field Crops Research**, v.111, n.1, p.55–64. 2009.

**CAPÍTULO 3 – ESTRATÉGIAS DE SELEÇÃO DE GENÓTIPOS DE SORGO
BIOMASSA EM ENSAIOS MULTI-AMBIENTAIS COM BASE
EM MÚLTIPLOS CARACTERES**

RESUMO

O sorgo biomassa apresenta-se como matéria-prima interessante para a produção de etanol de segunda geração. Dessa forma, o objetivo com este trabalho foi identificar e selecionar genótipos superiores de sorgo biomassa para a produção de etanol de segunda geração considerando a seleção para múltiplos caracteres, bem como, verificar a estratégia mais adequada para seleção desses genótipos pelo índice FAI-BLUP considerando experimentos em múltiplos locais. Foram avaliados vinte e cinco genótipos de sorgo biomassa sob delineamento alfa-látice triplo 5x5 em experimento conduzido em Janaúba e Sete Lagoas, Minas Gerais e Sinop, Mato Grosso, no ano agrícola de 2018/2019. Avaliou-se os caracteres: florescimento (FLOR); altura de planta (AP); produtividade de massa verde (PMV); massa seca (MS); umidade (UM); produtividade de massa seca (PMS); fibra em detergente neutro (FDN); fibra em detergente ácido (FDA); lignina (LIG); hemicelulose (HEM); e a concentração de celulose (CEL). Os dados foram submetidos a análise exploratória, a análise univariada por local e conjunta e a cinco estratégias para seleção de genótipos superiores pelo índice FAI-BLUP. As estratégias empregadas foram: primeira estratégia (E1) uso dos valores genéticos preditos em cada local (BLUPs dos genótipos); segunda estratégia (E2) uso dos valores genéticos preditos livres de interação considerando todos os locais (BLUPs dos genótipos); terceira estratégia (E3) uso dos valores genéticos preditos capitalizando a interação GxE (BLUPs da interação); quarta estratégia (E4) uso dos valores genéticos preditos na E1 e criação de novos caracteres por local; quinta estratégia (E5) uso dos valores preditos na E3 e criação de novos caracteres por local. Ao aplicar o índice FAI-BLUP, independente da estratégia empregada, o ideótipo favorável obtido foi no sentido de aumentar os caracteres FLOR, AP, PMV, MS, PMS, FDN, FDA, HEM e CEL e diminuir os caracteres UM e LIG, enquanto o ideótipo desfavorável foi no sentido contrário. Os genótipos 4, 13, 17 e 18 destacaram-se quando a superioridade, previsibilidade e adaptabilidade genotípica sendo interessante o uso destes visando a produção de etanol de segunda geração. O uso das estratégias E2 e E3 foram as mais adequadas para seleção de genótipos superiores pelo índice FAI-BLUP considerando experimentos em múltiplos locais e visando a recomendação, respectivamente, para locais que não apresentam e/ou apresentam o mesmo padrão de interação GxE da rede experimental avaliada.

Palavras-chave: *Sorghum bicolor* L., FAI-BLUP, bioenergia, seleção genotípica.

ABSTRACT

Biomass sorghum is an interesting raw material to produce second-generation ethanol. Thus, the aim of this study was to identify and select superior sorghum biomass genotypes for the second-generation ethanol production considering the selection for multiple traits, as well as to verify the most adequate strategy for selection of these genotypes by FAI-BLUP index considering experiments at multiple locations. Twenty-five sorghum biomass genotypes were evaluated under a 5x5 triple alpha-lattice design in an experiment conducted in Janaúba and Sete Lagoas, Minas Gerais and Sinop, Mato Grosso, in the 2018/2019 harvest. Characters were evaluated: flowering (FLOW); plant height (PH); green mass productivity (GMP); dry mass (DM); humidity (HU); dry mass productivity (DMP); neutral detergent fiber (NDF); acid detergent fiber (ADF); lignin (LIG); hemicellulose (HEM); and the cellulose concentration (CEL). The data were submitted to exploratory analysis, individual and joint univariate analysis and five strategies for selection of superior genotypes by the FAI-BLUP index. The strategies employed were: first strategy (E1) use of predicted breeding values in each location (genotype BLUPs); second strategy (E2) use of interaction-free predicted genetic values considering all locations (BLUPs of genotypes); third strategy (E3) use of predicted breeding values capitalizing on the GxE interaction (BLUPs of the interaction); fourth strategy (E4) use of genetic values predicted in E1 and creation of new characters per location; fifth strategy (E5) use of predicted values in E3 and creation of new characters per location. When applying the FAI-BLUP index, for all strategy used, the favorable ideotype obtained was to increase the characters FLOW, PH, GMP, DM, DMP, NDF, ADF, HEM and CEL and decrease the characters HU and LIG, while that the unfavorable ideotype was in the opposite direction. Genotypes 4, 13, 17 and 18 stood out in terms of superiority, predictability and genotypic adaptability, being interesting to use them for the production of second-generation ethanol. The use of the E2 and E3 strategy were the most indicate for selecting superior genotypes by the FAI-BLUP index considering experiments in multiple locations and aiming at the recommendation, respectively, for locations that do not present and/or present the same pattern of GxE interaction in the evaluated experimental network.

Keywords: Sorghum bicolor L., FAI-BLUP, bioenergy, genotypic selection.

1 INTRODUÇÃO

A seleção de genótipos de sorgo biomassa (*Sorghum bicolor* (L.) Moench), importante matéria-prima para suprir a demanda por fontes renováveis de energia diminuindo, assim, os impactos ambientais e incrementando a matriz energética global (SCOTT et al., 2015), depende da finalidade industrial da cultura e envolve vários caracteres simultaneamente (VENDRUSCOLO et al., 2016; HABYARIMANA et al., 2020). Esses caracteres apresentam grande variação quanto a direção e a magnitude das associações (CASTRO et al., 2015; VENDRUSCOLO et al., 2016; ALMEIDA et al., 2019; DOS SANTOS et al., 2020; HABYARIMANA et al., 2020), podendo ocasionar de acordo com a seleção realizada, mudanças desejáveis em alguns caracteres e indesejáveis em outros.

O sorgo biomassa é eficiente tanto para a geração de energia elétrica quanto para a produção de etanol de segunda geração (2G) (CASTRO et al., 2017; SIMEONE et al., 2020). O genótipo ideal de sorgo biomassa, visando a produção de etanol 2G, deve apresentar, além de alto rendimento e baixo teor de umidade (<60%), baixo teor de lignina e altos teores de celulose e hemicelulose (OLSON et al., 2012; GURAGAINA et al., 2014; ALMEIDA et al., 2019), isto pois a lignina atua negativamente na hidrólise enzimática, podendo, inclusive, inibir o processo fermentativo subsequente e conseqüentemente a produção de etanol 2G (VAN DER WEIJDE et al., 2016). Assim, considerando as associações observadas entre esses caracteres, a seleção indireta dos genótipos por meio da altura de planta pode influenciar positivamente no incremento de biomassa seca (FERNANDES et al., 2018; DOS SANTOS et al., 2020) e negativamente no decréscimo do teor de lignina (OLIVEIRA et al., 2019). Logo, o uso de procedimentos genético-estatísticos, que permitam minimizar os efeitos indesejados das respostas correlacionadas e que considerem a seleção para múltiplos caracteres, é extremamente recomendada para a obtenção do genótipo ideal e para o aumento da eficiência dos programas de melhoramento (RAMALHO et al., 2012).

Vários índices de seleção foram propostos para a seleção de genótipos superiores considerando múltiplos caracteres, como, por exemplo, os índices de: Smith (1936) e Hazel (1943), Willians (1962) e Mulamba e Mock (1978). Apesar desses índices considerarem vários caracteres simultaneamente para a seleção dos genótipos, estes apresentam algumas limitações que reduzem a precisão da seleção e podem ocasionar conclusões errôneas. Entre as principais limitações desses índices estão a dificuldade em atribuir pesos aos caracteres de interesse e a presença de multicolinearidade resultante da correlação entre esses caracteres. O índice FAI-BLUP (Factor analysis and ideotype-design – análise de fatores e design de

ideótipos) tem a capacidade de contornar os problemas de multicolinearidade e atribuição de pesos econômicos, além de considerar a estrutura das correlações dos caracteres e de direcionar os melhoristas na seleção de genótipos mais próximos ao ideótipo (ROCHA et al., 2018). O índice FAI-BLUP apresentou maiores ganhos na seleção de genótipos quando comparado aos índices de Smith-Hazel e de Mulamba e Mock (ROCHA et al., 2018; BOTELHO et al., 2021), sendo considerado eficiente para a seleção de genótipos superiores no melhoramento genético (DA SILVA et al., 2018; OLIVEIRA et al., 2019; ROCHA et al., 2019; WOYANN et al., 2020) e uma boa estratégia para a seleção de genótipos visando a produção de biocombustíveis (ALVES et al., 2019; WOYANN et al., 2020).

As vantagens apresentadas pelo índice FAI-BLUP possibilitaram que seu uso fosse rapidamente difundido entre os programas de melhoramento (DA SILVA et al., 2018; SALVADOR et al., 2018; ALVES et al., 2019; ROCHA et al., 2019; BOTEGA, 2019; WOYANN et al., 2020; ALMEIDA et al., 2021; BOTELHO et al., 2021; MEIER et al., 2021; PEIXOTO et al., 2021). Os BLUPs utilizados nesses estudos diferem quanto as estratégias empregadas, principalmente, quando se avalia a interação com outra fonte de variação, como, por exemplo, local, havendo estudos cuja estratégia empregada considera os valores genéticos preditos capitalizando a interação (BLUPS da interação obtidos na análise conjunta) (ROCHA et al., 2018; WOYANN et al., 2020; BOTELHO et al., 2021; PEIXOTO et al., 2021) e estudos que consideram os valores genéticos preditos dos genótipos nas análises individuais (BLUPS dos genótipos) e obtém novos caracteres de acordo com a fonte de variação (local, ano, safra, geração, entre outros) (DA SILVA et al., 2018; SALVADOR et al., 2018; OLIVEIRA et al., 2019; ROCHA et al., 2019) visando capitalizar o efeito da interação.

Dessa forma, o objetivo com este trabalho foi identificar e selecionar genótipos superiores de sorgo biomassa para a produção de etanol de segunda geração considerando a seleção para múltiplos caracteres, bem como, verificar qual estratégia seria a mais adequada para seleção desses genótipos pelo índice FAI-BLUP considerando experimentos em múltiplos locais.

2 MATERIAL E MÉTODOS

2.1 Local experimental, plano experimental e manejo

Os experimentos foram conduzidos no ano agrícola de 2018/2019 nos municípios de Janaúba (15°48' S; 43°18' O; 516 m) e Sete Lagoas (19°27' S; 44°14' O; 767 m), Estado de Minas Gerais, e Sinop (11°50' S; 55°38' O; 384 m), Estado de Mato Grosso, Brasil. Janaúba,

Sete Lagoas e Sinop, cujo os climas foram classificados, segundo Koppen, como Aw (clima tropical de savana) (ALVARES et al., 2013) apresentaram, respectivamente, durante a condução do experimento precipitação total de 684; 165; 1048 mm e temperatura média de 26,6; 23,3; 26,4 °C.

Os experimentos foram instalados, em novembro de 2018, segundo o delineamento alfa-látice triplo 5 x 5. Foram avaliados vinte e cinco genótipos sensíveis ao fotoperíodo, sendo vinte e dois híbridos experimentais bmr de sorgo biomassa e três híbridos comerciais - HC (HC1: BRS716; HC2: BRS655, HC3: Volumax). As parcelas experimentais, constituídas por duas linhas de 5,0 m de comprimento e espaçadas de 0,70 m entre fileiras, foram aradas usando um conjunto de discos tandem a uma profundidade de 20 cm seguido por um cultivador de campo para a preparação do canteiro de sementes. A densidade de plantas adotada foi de 140.000 plantas ha⁻¹. Na semeadura, que foi realizada de forma direta, foi aplicado 450 kg ha⁻¹ do formulado NPK 08-28-16 nos sulcos de plantio e quando as plantas estavam no estágio V4-V5 foi aplicado, em cobertura, 200 kg de ureia ha⁻¹. Além disso, durante a condução do experimento, foi realizado o controle de plantas daninhas por meio da aplicação de atrazina (3 kg i.a. ha⁻¹) complementado por capina mecânica, quando necessário. A colheita foi realizada manualmente, em média, aos 180 dias após a semeadura.

2.2 Caracteres mensurados

Os caracteres agronômicos mensurados foram: florescimento (FLOR; dias) de pelo menos 50% das plantas da parcela, em dias após a semeadura; altura da planta (AP; m) na média de oito plantas aleatórias, medidas da superfície do solo ao ápice da panícula, no momento da colheita; produtividade de massa verde (PMV; t ha⁻¹) de oito plantas aleatórias por parcela, com panícula e folhas, pesados por meio de uma balança de suspensão digital, em Kg, convertida em t ha⁻¹; massa seca (MS, %) determinada pela secagem do material utilizado para mensurar o PMV em estufa e dessecador até que as amostras apresentassem peso constante, seguindo a metodologia de Zhao et al. (2009), umidade (UM; %) obtida pela diferença entre 100% e a MS, produtividade de massa seca (PMS; t ha⁻¹) de oito plantas aleatórias por parcela, com panícula e folhas, pesados após a secagem do material por meio de uma balança de suspensão digital, em Kg, convertida em t ha⁻¹.

Com base na MS, determinou-se os caracteres de composição da biomassa: fibra em detergente neutro (FDN, %), determinado pela razão entre o peso da amostra após a digestão em 35 ml de solução de detergente neutro e 0,5 ml de amilase, filtragem e secagem em estufa a 105°C e o peso da amostra antes da digestão (0,35g); fibra em detergente ácido (FDA, %),

determinado pela razão entre o peso da amostra após a digestão em 35 ml de solução de detergente ácido, filtragem e secagem em estufa a 105°C e o peso da amostra antes da digestão (0,35g); lignina (LIG, %), determinado segundo a metodologia de Soest (1963); hemicelulose (HEM, %), obtido pela diferença entre FDN e FDA; concentração de celulose (CEL, %) obtido pela diferença entre FDA e lignina.

2.3 Análise estatística

Os dados coletados dos caracteres foram submetidos a análise exploratória. Na análise exploratória, foi realizada análise descritiva dos caracteres e exame gráfico das relações entre os caracteres, visando, principalmente, identificar possíveis valores atípicos (outliers), presença de dados faltantes (missing) e avaliar algumas suposições básicas, como normalidade, linearidade e homocedasticidade. A análise exploratória foi realizada utilizando o pacote Scikit Learn versão 0.24.2 (PEDREGOSA et al., 2011) em ambiente Python 3 (LUTZ, 2014).

Os pressupostos de normalidade dos erros e homocedasticidade das variâncias foram verificados por meio do teste de Shapiro-Wilk e Levene, respectivamente. Constatado que os dados atendiam os pressupostos, foram realizadas, utilizando-se a abordagem de modelos mistos, as análises univariada por local e conjunta dos dados. As análises por local (EQUAÇÃO 1) e conjunta (EQUAÇÃO 2) para cada carácter foram realizadas de acordo com os seguintes modelos estatísticos:

$$y^c = X^c B^c + Z^c b^c + W^c g^c + e^c \quad (1)$$

$$y^c = X_1^c B_1^c + X_2^c B_2^c + Z_1^c b_1^c + W^c g^c + T^c i^c + e^c \quad (2)$$

em que: y^c é o vetor de dados fenotípicos da c -ésima característica; $X^c, X_1^c, X_2^c, Z^c, Z_1^c, W^c, T^c$ são matrizes de incidência dos efeitos fixos e aleatórios; B^c, B_1^c, B_2^c são vetores de efeitos fixos das repetições, dos locais e das repetições dentro dos locais acrescidas da média geral, respectivamente; b^c, b_1^c são os vetores de efeito de blocos dentro de repetições, sendo $b^c \sim N(0, I\sigma^2b)$, e de blocos dentro de repetições nos locais, sendo $b_1^c \sim N(0, I\sigma^2b1)$; g^c é o vetor de efeitos dos genótipos, sendo $g^c \sim N(0, I\sigma^2g)$; i^c é o vetor de efeitos das interações genótipos por locais, sendo $i^c \sim N(0, I\sigma^2i)$; e^c é o vetor de efeitos do erro experimental, sendo $e^c \sim N(0, I\sigma^2e)$; $\sigma^2b, \sigma^2b1, \sigma^2g, \sigma^2i, \sigma^2e$ são variâncias de blocos dentro de repetições, de blocos dentro de repetições nos locais, de genótipos, da interação genótipos por locais e do erro experimental, respectivamente.

Após obter os BLUPS nas análises univariadas foi aplicado o índice FAI-BLUP (factor analysis and genotype-ideotype distance), visando a seleção de genótipos superiores (ROCHA et al., 2018). O índice FAI-BLUP foi aplicado considerando cinco estratégias, de forma a identificar qual seria a mais adequada para seleção de genótipos considerando análises em múltiplos locais.

Na primeira estratégia (E1) o índice FAI-BLUP foi aplicado considerando cada local individualmente (análise individual por local) e utilizou-se os valores genéticos preditos em cada local (BLUPs dos genótipos). Na segunda (E2) e terceira estratégia (E3) o índice foi aplicado considerando todos os locais de forma conjunta (análise conjunta) e utilizou-se, respectivamente, os valores genéticos preditos livres de interação (BLUPs dos genótipos) e os valores genéticos preditos capitalizando a interação GxE (BLUPs da interação). Na quarta (E4) e quinta estratégia (E5) o índice foi aplicado considerando a criação de novos caracteres de acordo com o local avaliado, tendo assim 33 caracteres avaliados ao invés de 11, visto que o carácter FLOR, por exemplo, desdobrou-se em FLOR_J (florescimento Janaúba), FLOR_SL (florescimento Sete Lagoas) e FLOR_S (florescimento Sinop) e utilizou-se, respectivamente, os valores genéticos preditos obtidos nas análises individuais (BLUPs dos genótipos – equivalente aos BLUPs da E1) e os valores preditos capitalizado a interação GxE obtidos na análise conjunta (BLUPs da interação – equivalente aos BLUPs da E3).

Os BLUPs de cada estratégia foram utilizados na análise de fatores. O número de fatores (n) deve ser igual ao número de componentes principais cujo autovalores apresentaram valores maiores ou iguais a unidade, ou seja, um. Após a determinação do número de fatores determinou-se o número de ideótipos, onde o estimador utilizado foi (EQUAÇÃO 3):

$$NI = 2^n \quad (3)$$

em que: NI: número de ideótipos; n: número de fatores.

O ideótipo favorável obtido foi no sentido de aumentar os caracteres FLOR, AP, PMV, MS, PMS, FDN, FDA, HEM e CEL e diminuir os caracteres UM e LIG. Por outro lado, o ideótipo desfavorável foi no sentido contrário, ou seja, diminuir os caracteres FLOR, AP, MS, PMV, PMS, HEM, FDN, FDA e CEL e aumentar os caracteres UM e LIG em todas as estratégias empregadas. Inclusive, nas estratégias 4 e 5 o ideótipo ideal, por exemplo, foi no sentido de aumentar os caracteres FLOR_J, FLOR_SL e FLOR_S e diminuir os caracteres UM_J, UM_SL e UM_S.

Depois do estabelecimento dos ideótipos, as distâncias entre os genótipos avaliados e os ideótipos favoráveis foram calculadas e transformadas em uma medida de similaridade relativa, que possibilitou o ranqueamento dos genótipos (EQUAÇÃO 4):

$$P_{jk} = \frac{\frac{1}{d_{jk}}}{\sum_{j,k=1}^{n,m} \frac{1}{d_{jk}}} \quad (4)$$

em que: P_{jk} é a similaridade relativa da progênie j ao ideótipo k; d_{jk} é a distância da progênie j ao ideótipo k, baseada na distância euclidiana média padronizada.

Após o ranqueamento dos genótipos foi realizada a seleção dos melhores genótipos, com base no ideótipo, de acordo com a intensidade de seleção aplicada (15%). O ganho esperado (GS, %) com a seleção foi estimado para cada carácter em cada estratégia (EQUAÇÃO 5):

$$GS(\%) = \frac{X_{mc} - X_{0c}}{X_{0c}} * 100 \quad (5)$$

em que: X_{mc} é a média dos BLUPs dos genótipos selecionados para o caractere c; X_{0c} é média geral dos BLUPs dos genótipos para o carácter c em avaliação.

As análises estatísticas pela abordagem de modelos mistos foram realizadas utilizando o pacote metan (OLIVOTO e LÚCIO, 2020) e as figuras foram realizadas utilizando o pacote ggplot2 (WICKHAM, 2016) em ambiente R (R CORE TEAM, 2018). As rotinas utilizadas para realizar as análises e as figuras estão disponíveis em https://github.com/GabrielleLombardi/FAI_BLUP_Strategies.

3 RESULTADOS

A variabilidade genética, essencial para o progresso genético (ROCHA et al., 2018), foi observada ($p < 0,05$) em todas as estratégias empregadas para todos os caracteres mensurados. Também, foi observado interação genótipo por local na análise conjunta para todos os caracteres mensurados. O índice de seleção FAI-BLUP apresentou, de acordo com a estratégia empregada, diferenças nos números de fatores, na proporção da variância explicada por eles, nas cargas fatoriais e comunalidades, na classificação, na seleção dos genótipos e no ganho esperado com a seleção dos genótipos de sorgo de biomassa (Figuras 1, 2, 3, 4, 5 e 6).

Na E1, nos experimentos de Sete Lagoas e Sinop os três primeiros componentes principais apresentaram autovalores maiores que um e no experimento de Janaúba os quatro primeiros (Figura 1 (A)). Na E2 e E3 os quatro primeiros componentes apresentaram autovalores maior que um e na E4 e E5 os sete primeiros componentes (Figura 3 (A, B, C e D)). Os dados podem ser reduzidos nos números de componentes principais que apresentaram

autovalores maiores que um (KAISER, 1958), portanto, considerando a E1, os dados foram reduzidos em três fatores nos experimentos de Sete Lagoas e Sinop e quatro fatores no experimento de Janaúba (Figuras 1 (B)). Na E2 e E3 os dados foram reduzidos em quatro fatores e na E4 e E5 em sete fatores (Figura 4 (A, B, C e D)). Esses fatores foram suficientes para representar mais de 80% de toda a variabilidade nas estratégias avaliadas, visto que a proporção da variância explicada por esses fatores na E1 foi de 89,9% em Janaúba, 84,9% em Sete Lagoas e 87,7% em Sinop; na E2 foi de 90,4%; na E3 foi de 90,5%; na E4 foi de 90,2%; e na E5 de 92,1% (Figuras 1 (B) e 4 (A, B, C e D)).

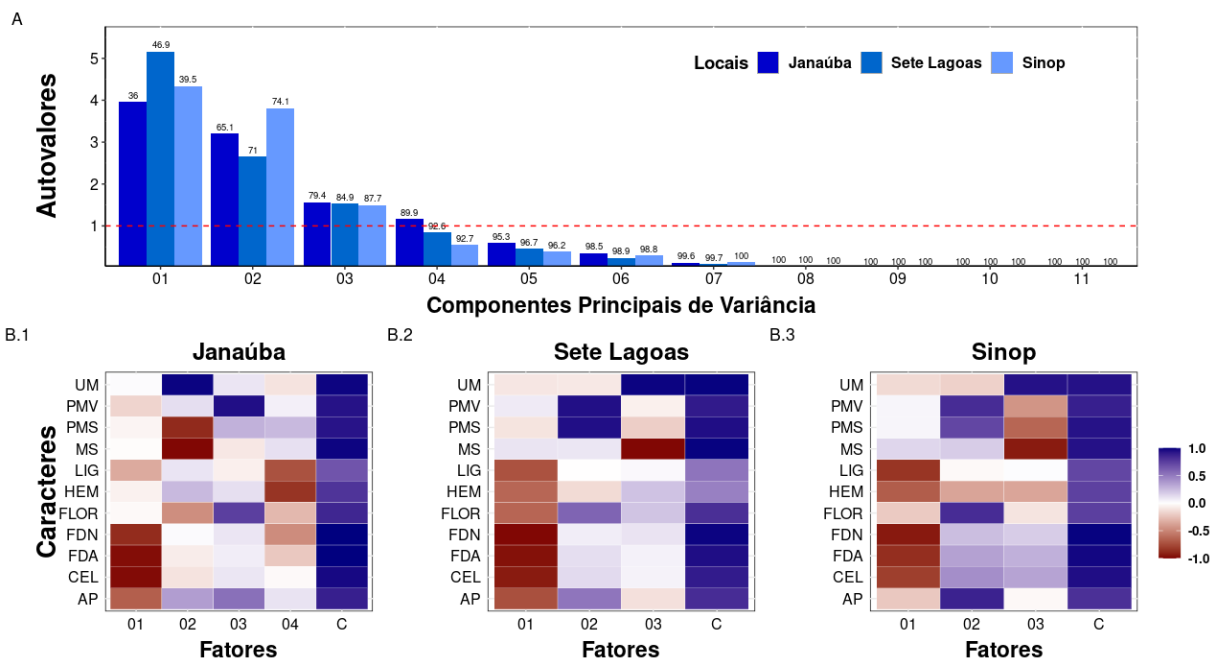
Como os fatores são caracteres abstratos que são obtidas a partir da redução da dimensionalidade dos caracteres originais mensurados, tem-se em cada fator um grupo de caracteres que apresentam alta correlação entre si, independente da direção, e baixa correlação com os caracteres contidos nos outros fatores (CRUZ e CARNEIRO, 2003). Na E1, o experimento de Janaúba apresentou alta correlação genética no primeiro fator (FA-01) entre AP, FDN, CEL e FDA; no segundo fator (FA-02) entre UM, PMS e MS, no terceiro fator (FA-03) entre PMV e FLOR e no quarto fator (FA-04) entre LIG e HEM (Figura 1 (B.1)). O experimento de Sete Lagoas, apresentou alta correlação genética no FA-01 entre AP, FDN, LIG, HEM, FLOR, CEL e FDA; no FA-02 entre PMV e PMS e no FA-03 entre UM e MS (Figura 1 (B.2)). O experimento de Sinop apresentou alta correlação genética no FA-01 entre FDN, LIG, HEM, CEL e FDA; no FA-02 entre AP, PMV, FLOR e PMS e no FA-03 entre UM e MS (Figura 1 (B.3)).

Na E2 e E3 alta correlação genética foi observada no FA-01 entre AP, FDN, CEL e FDA; no FA-02 entre UM e MS, no FA-03 entre PMV, FLOR e PMS e no FA-04 entre LIG e HEM (Figura 4 (A)). Na E4 alta correlação genética foi observada no FA-01 entre SL_FLOR, S_AP, SL_AP, J_AP, SL_FDN, SL_FDA, SL_CEL e J_CEL; no FA-02 entre J_FLOR, S_LIG, S_FDN, S_FDA e S_CEL; no FA-03 entre S_FLOR, S_PMV, SL_PMV, J_PMV, S_PMS, SL_PMS; no FA-04 (quarto fator) entre S_MS, S_UM, S_HEM, SL_HEM e J_HEM; no FA-05 (quinto fator) entre SL_LIG, J_LIG, J_FDN, J_FDA; no FA-06 (sexto fator) entre SL_MS e SL_UM; no FA-07 (sétimo fator) entre J_MS, J_UM e J_PMS (Figura 4 (B)). Na E5 alta correlação genética foi observada no FA-01 entre S_AP, SL_AP, J_AP, SL_FDN, J_FDN, SL_FDA, J_FDA, SL_CEL e J_CEL; no FA-02 entre J_MS, J_UM, J_PMS; no FA-03 entre S_HEM, SL_HEM e J_HEM; no FA-04 entre S_FLOR, SL_FLOR, S_PMV, SL_PMV, J_PMV, S_PMS e SL_PMS; no FA-05 entre SL_LIG e J_LIG; no FA-06 entre SL_MS e SL_UM; no FA-07 entre J_FLOR, S_MS, S_UM, S_LIG, S_FDN, S_FDA e S_CEL (Figura 4 (C)).

Em relação ao ideótipo de sorgo biomassa para a produção de etanol de segunda geração, quatro genótipos com maior potencial foram selecionados de acordo com a estratégia empregada. Os genótipos selecionados foram: 4, 10, 13 e 20 em Janaúba (Figura 2 (A.1)); 3, 4, 16 e 17 em Sete Lagoas (Figura 2 (A.2)); e 1, 6, 8 e 23 em Sinop (Figura 2 (A.3)), considerando a E1; 4, 13, 17 e 18, considerando a E2 e E3 (Figura 5 (A)); 3, 4, 15 e 18 considerando E4 e 3, 4, 7 e 15 considerando a E5 (Figura 5 (B e C)).

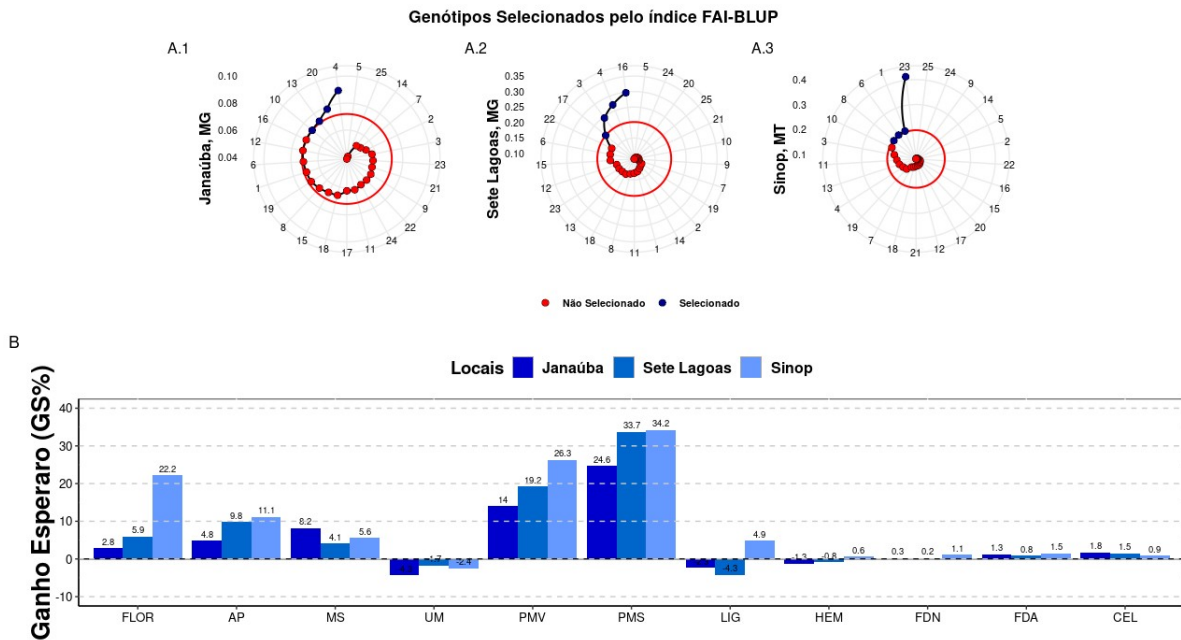
Os ganhos esperados (GS) com a seleção destes genótipos foram previstos no sentido desejável para todos os caracteres mensurados, exceto para: LIG em Sinop e HEM em Janaúba e Sete Lagoas, considerando a E1 (Figura 2 (B)); e, em média, para FDN e HEM considerando a E2 e E3 (Figura 6 (A)); para HEM, FDN e FDA considerando a E4 (Figura 6 (B)); e para FLOR e HEM considerando a E5 (Figura 6 (C)). De maneira geral o GS foi, em média, superior para os caracteres AP, MS, UM, PMV, PMS e FDA utilizando a E3; para os caracteres FLOR e LIG utilizando a E4 e para os caracteres FDN, CEL e HEM utilizando a E5.

Figura 1 - Estimativas de autovalores pela análise de componentes principais e a proporção da variância explicada por eles em cada local (A), cargas fatoriais e comunalidades (c) dos fatores, para Janaúba (B.1), Sete Lagoas (B.2) e Sinop (B.3), cujo número de autovalores foram maiores ou iguais a um com base no índice FAI-BLUP.



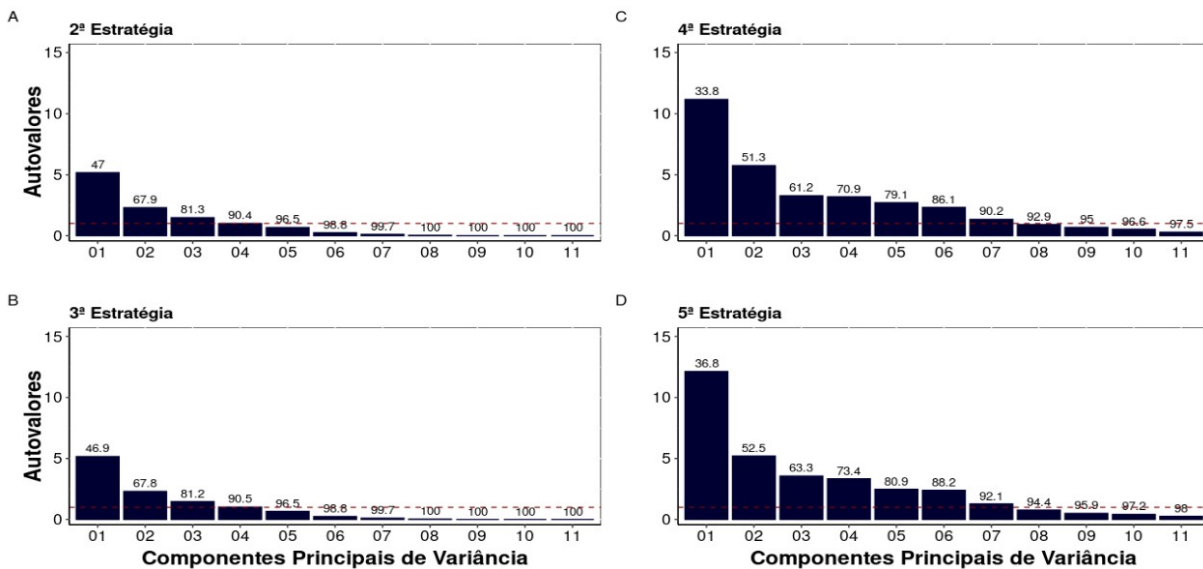
Fonte: Do autor (2021).

Figura 2 - Classificação e seleção dos genótipos de sorgo de biomassa para Janaúba (A.1), Sinop (A.2), Sete Lagoas (A.3) e Ganho esperado (GS %) com a seleção dos genótipos em cada local pelo índice FAI-BLUP (B).



Fonte: Do autor (2021).

Figura 3 - Estimativas de autovalores pela análise de componentes principais e a proporção da variância explicada por eles considerando as estratégias E2 (A), E3 (B), E4 (C) e E5 (D) com base no índice FAI-BLUP.



Fonte: Do autor (2021).

Figura 4 - Cargas fatoriais e comunalidades (C) dos fatores considerando as estratégias E2 (A), E3 (B), E4 (C) e E5 (D) em que o número de autovalores foram maiores ou iguais a um com base no índice FAI-BLUP.

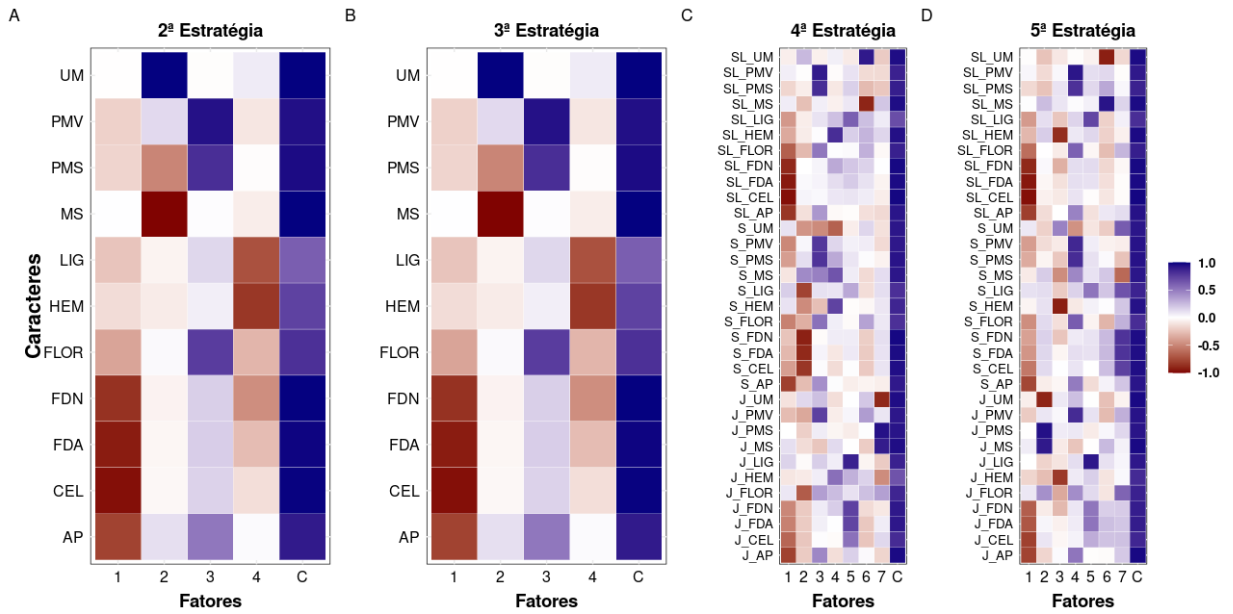


Figura 5 - Classificação e seleção dos genótipos de sorgo de biomassa considerando as estratégias E2 (A), E3 (B), E4 (C) e E5 (D) pelo índice FAI-BLUP.

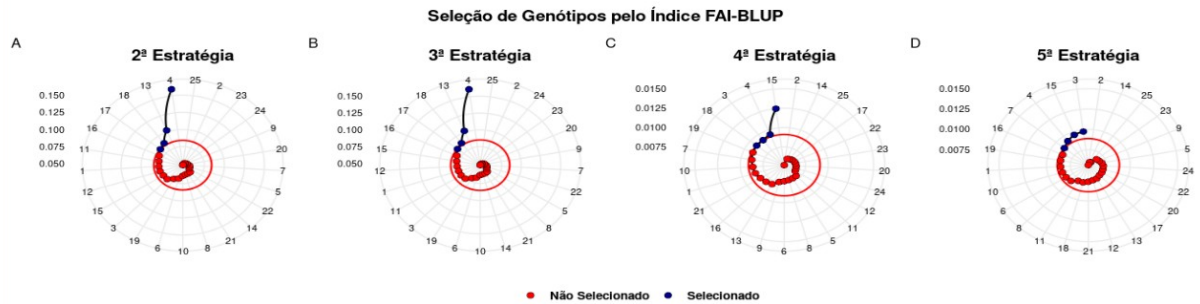
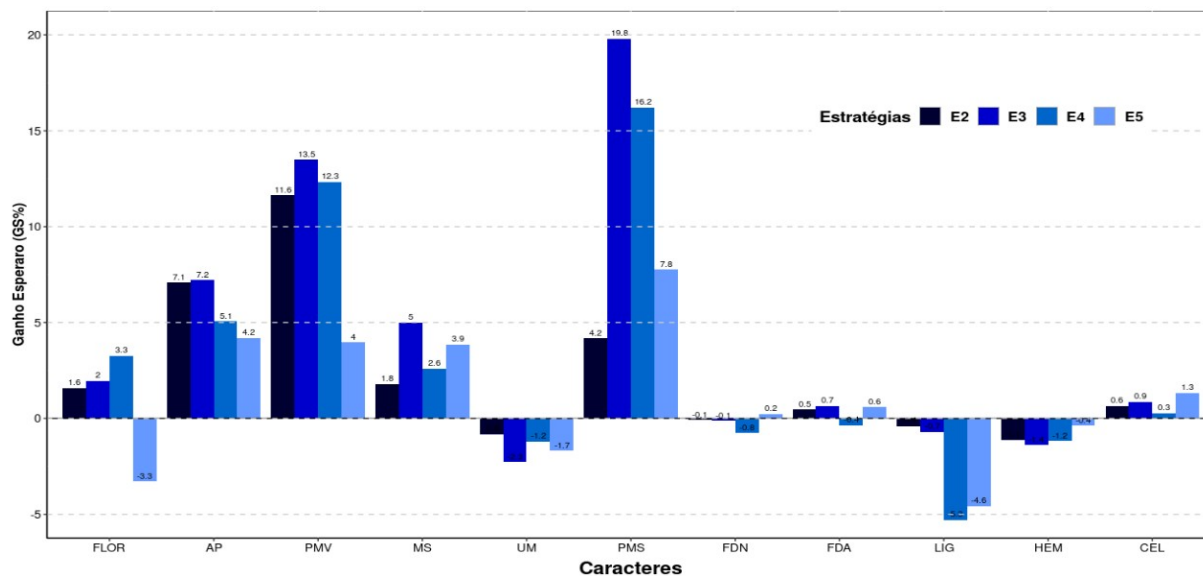


Figura 6 - Ganho esperado (GS %) com a seleção dos genótipos considerando as estratégias E2, E3, E4 e E5 pelo índice FAI-BLUP.



Fonte: Do autor (2021).

4 DISCUSSÃO

A análise de fatores é amplamente utilizada para auxiliar no melhoramento genético, principalmente, na seleção de genótipos superiores (SOUZA et al., 2020). Isto ocorre pois, a análise de fatores, ao reduzir a dimensionalidade dos caracteres em fatores que os agrupam de acordo com intensidade de associação dos mesmos, possibilita a descrição e interpretação dos fatores na área biológica, bem como do comportamento dos caracteres (CRUZ et al., 2014). Logo, ao utilizar-se da análise de fatores, o índice FAI-BLUP busca compreender o comportamento dos caracteres e fornecer uma interpretação biológica dos fatores e dos ideótipos (ROCHA et al., 2018). O comportamento dos caracteres variou tanto dentro da mesma estratégia (E1) quanto entre as estratégias aplicadas (estratégias E1, E2, E3, E4 e E5), o que dificultou a generalização da interpretação biológica dos fatores e dos ideótipos (Figuras 1 e 4). Esse fato, era esperado visto a variação dos valores genotípicos preditos e dos números de caracteres avaliados entre as estratégias.

Na E1, no experimento conduzido em Sinop tem-se o agrupamento dos caracteres lignocelulótico no FA-01 (FDN, FDA, LIG, HÉM e CEL), os caracteres agrônômicos no FA-02 (FLOR, AP, PMV e PMS) e os caracteres relacionados a umidade da planta no FA-03 (MS e UM) (Figura 1). O comportamento desses caracteres em sorgo biomassa, bem como, a interpretação biológica desses fatores, corresponderam com as associações positivas observadas entre estes na literatura (CASTRO et al., 2015; VENDRUSCOLO et al., 2016; LI

et al., 2017; FERNANDES et al., 2018; ALMEIDA et al., 2019; DOS SANTOS et al., 2020; HABYARIMANA et al., 2020).

No experimento de Janaúba e Sete Lagoas entre os caracteres considerados agronômicos, os caracteres AP em Janaúba e AP e FLOR em Sete Lagoas, se agruparam com os caracteres considerados lignocelulóticos. Alguns estudos observaram associações positivas entre esses caracteres (ALMEIDA et al., 2019), inclusive aplicando o índice FAI-BLUP (OLIVEIRA et al., 2019) enquanto outros não observaram (CASTRO et al., 2015). Esse comportamento pode estar atrelado a performance similar desses caracteres ao longo do desenvolvimento da cultura (CASTRO et al., 2021) e a influência dos fatores ambientais (JOSHI et al., 2018; CASTRO et al., 2021) e da expressão dos genes envolvidos no desenvolvimento da cultura, como, por exemplo, os genes NAC (genes da família dos fatores de transcrição NAC) que atuam tanto no desenvolvimento do meristema apical quanto na constituição da parede celular (KEBRON et al., 2017; XIA et al., 2018; SANJARI et al., 2019). De maneira geral, é possível observar que em Janaúba os caracteres lignocelulóticos, os agronômicos e os caracteres relacionados a umidade na planta foram agrupados, respectivamente, nos fatores 1 e 4; 3; e 2; e, em Sete Lagoas eles foram agrupados, respectivamente, nos fatores 1, 2 e 3.

Após a alocação dos caracteres nos fatores e a compreensão do seu significado biológico, outro ponto que deve ser observado é a obtenção dos ideótipos. Os ideótipos são definidos a partir da combinação de fatores desejáveis e indesejáveis para o objetivo da seleção (ROCHA et al., 2018). Um fator desejável deve ter valores genéticos desejáveis enquanto um fator indesejável deve ter valores genéticos indesejáveis para todos os caracteres sob seleção agrupados nele (ROCHA et al., 2018). Os valores genéticos podem ser de máximo, mínimo, médio ou um valor genético específico (ROCHA et al., 2018). Assim, caso um fator agrupe positivamente caracteres cujo valor genético hipotetizado é máximo para um e mínimo para o outro, o índice FAI-BLUP, durante a análise de distância genótipo-ideótipo considerará o ideótipo ou no sentido desejável ou no sentido indesejável acarretando a seleção no sentido desejável para um dos caracteres, e, conseqüentemente, indesejado para o outro. Logo, o ganho desejado com a seleção somente será observado para um dos caracteres presente naquele fator (ALMEIDA et al., 2019).

O agrupamento positivo, em um mesmo fator, entre caracteres com valores genéticos opostos foi observado neste estudo (Figura 1 e 4). A LIG, cujo valor genético hipotetizado era mínimo, se agrupou no experimento de Janaúba com a HEM (FA-04), no experimento de Sete Lagoas com AP, FLOR e com os demais caracteres lignocelulóticos (FA-01) e no experimento

de Sinop com os demais caracteres lignocelulóticos (FA-01) cujos valores genéticos hipotetizados eram máximo (Figura 1). Esse agrupamento, após a seleção dos genótipos ideais para cada experimento, ocasionou ganhos negativos ou pouco expressivos para alguns caracteres (Figura 2). Por exemplo, tem-se com a seleção dos genótipos, em Janaúba, ganhos no sentido desejado para a LIG (-4,3%) e indesejado para a HEM (-1,3%) (Figura 2). Os ganhos observados em Janaúba para LIG e HEM indicam que a obtenção do ideótipo ocorreu no sentido desejado para o carácter LIG no FA-04. Essa análise pode ser extrapolada para os demais experimentos, tendo assim em Sete Lagoas a obtenção dos genótipos ocorreu no sentido desejado e em Sinop no sentido indesejado para o carácter LIG no FA-01 (Figura 2).

O uso da E1 resultou, para a maioria dos caracteres avaliados, elevados ganhos com a seleção sendo, assim, indicado seu uso para a seleção e recomendação de genótipos superiores considerando, exclusivamente, o local avaliado (Figura 2). A avaliação de genótipos em um único local não permite que o componente da interação seja isolado, podendo gerar confusão na estimativa da variância genética como também na herdabilidade, comprometendo o ganho esperado com a seleção (TERASAWA JÚNIOR et al., 2008; PEIXOTO et al., 2021). Somado a isso, tem-se os baixos valores de coincidência entre os genótipos selecionados em cada local na E1, indicando que o efeito da interação genótipo por local (GxE) observado na análise conjunta foi do tipo complexa, tornando assim desaconselhável o uso desta estratégia visando a recomendação de genótipos superiores para diferentes locais. A interação GxE foi considerada responsável por essas oscilações em outros estudos aplicando o índice FAI-BLUP em que os genótipos selecionados apresentaram baixo índice de coincidência entre os locais avaliados (PEIXOTO et al., 2021).

A estratégia apropriada para seleção de genótipos superiores visando maximizar os ganhos para uma série de locais deve estimar o ganho utilizando-se dos valores genéticos preditos na análise conjunta (CRUZ et al., 2014). Essa estratégia foi empregada em alguns estudos com o índice FAI-BLUP e os valores genéticos preditos nesses estudos foram livres de interação, ou seja, os blups dos genótipos obtidos na análise conjunta (BOTEGA, 2019; WOYANN et al., 2020; BOTELHO et al., 2021; PEIXOTO et al., 2021). O uso desses valores genéticos ocorre devido à metodologia de modelos mistos penalizar os valores genéticos preditos (MAIA et al., 2009), assim é esperado comportamento similar dos genótipos selecionados para os caracteres mensurados em outros locais desconhecidos ou com padrão de interação diferente daquele da rede experimental ou com alta heterogeneidade ambiental dentro de local (BASTOS et al., 2007; OLIVEIRA et al., 2021). Os valores genéticos preditos sem capitalizar o efeito da interação, também, podem ser utilizados para selecionar e

recomendar genótipos para locais que apresentem o mesmo padrão de interação dos ambientes avaliados (TORRES et al., 2015), contudo, neste caso, deve-se considerar o uso de valores genéticos preditos que capitalizem o efeito da interação GxE (BASTOS et al., 2007).

Ao empregar, na análise conjunta, o índice FAI-BLUP considerando o uso dos valores genéticos preditos livres de interação (E2) e o uso dos valores genéticos preditos que capitalizem o efeito da interação GxE (E3), pode-se observar resultados similares nos números de fatores, no agrupamento dos caracteres nos fatores, na interpretação biológica dos fatores, nos genótipos selecionados e no sentido do ganho com a seleção (Figuras 3, 4, 5 e 6). Resultados divergentes foram observados apenas na magnitude do ganho com a seleção (Figura 6). Em ambas as estratégias, os caracteres lignocelulóticos, os agrônômicos e os caracteres relacionados a umidade na planta foram agrupados, respectivamente, nos fatores 1 e 4; 3; e 2 (Figura 4) e os genótipos selecionados foram 4, 13, 17 e 18 (Figura 5). Isto indica que estes genótipos possuem comportamento similar em locais que apresentam ou não o mesmo padrão de interação dos locais avaliados e, somado a isto, pode indicar maior expressão genotípica dos caracteres sob seleção e maior adaptabilidade e estabilidade genotípica (MAIA et al., 2009; TORRES et al., 2015).

Os ganhos com a seleção, nas estratégias E2 e E3, foram no sentido desejado para todos os caracteres mensurados, com exceção FDN e HEM (Figura 6). A aplicar a E3 observou maiores ganhos com a seleção do que quando aplicou a E2. Os maiores ganhos com a seleção observados quando se capitalizou a interação, deve-se ao incremento do valor genotípico predito dos caracteres. Assim, os genótipos selecionados para locais que apresentem o mesmo padrão de interação da rede experimental avaliada tendem a expressar maiores ganhos do que quando selecionados para locais com padrões diferentes aos avaliados no experimento. Essas duas estratégias foram consideradas eficientes para seleção e recomendação de genótipos para locais que apresentam o mesmo padrão (E2 e E3) ou não (E2) daquele da rede experimental avaliada. E somado a isto, ao analisar ambas as estratégias em conjunto pode-se melhor vislumbrar a adaptabilidade e estabilidade dos genótipos selecionados.

Outra estratégia que está sendo amplamente utilizada, para a identificação de genótipos com maior potencial produtivo e de alta qualidade, independente dos ambientes avaliados, empregando o índice FAI-BLUP considera as médias BLUP dos genótipos estimadas para cada local e carácter como novos caracteres (DA SILVA et al., 2018; SALVADOR et al., 2018; OLIVEIRA et al., 2019; ROCHA et al., 2019), de forma a capitalizar os efeitos da interação GxE (SALVADOR et al., 2018; ROCHA et al., 2019).

Utilizando essa estratégia, tem-se que os caracteres agrupados no mesmo fator apresentam maiores estimativas de correlação genética e interações não significativas entre as fontes de variação (ROCHA et al., 2019), no caso deste estudo, interações não significativas entre genótipos e locais. Além dessa estratégia, denominada E4, este estudo propôs, uma quinta estratégia (E5) a qual utiliza os valores preditos capitalizados pela interação GxE obtidos na análise conjunta e considera novos caracteres por local de forma a verificar a interação GxE e selecionar e recomendar genótipos para locais que apresentem o mesmo padrão de interação dos ambientes avaliados.

Na E4 e na E5, observou-se o agrupamento dos caracteres em sete fatores. Os caracteres lignocelulóticos, os agronômicos e os caracteres relacionados a umidade na planta foram agrupados, respectivamente, nos fatores 1, 2, 4 e 5; 3; 6 e 7 na E4 e nos fatores 1, 3, 5 e 7; 4; 2 e 6 na E5 (Figura 4). Na E4, houve dificuldade na interpretação biológica de alguns fatores, visto que os caracteres com diferentes interpretações biológicas foram agrupados no mesmo fator, como, por exemplo, o fator 4 apresentou caracteres relacionados a umidade da planta e caracteres lignocelulósicos. Essa dificuldade de interpretação biológica dos fatores é comum de acontecer quando se avalia elevado número de caracteres simultaneamente (CRUZ et al., 2014), o qual nesta estratégia foi triplicada em relação às estratégias anteriores. Além disso, elevado número de caracteres pode acarretar no uso de maior número de fatores para realizar o agrupamento dos caracteres, o que, conseqüentemente, ocasiona maior erro nas estimativas e seleção do genótipo superior (CRUZ et al., 2014).

Analisando o agrupamento dos caracteres nos fatores, tem-se, na E4 e na E5, que os caracteres AP, PMV e HEM não foram influenciadas pela interação GxE, visto que foram agrupados no mesmo fator (Figura 4). Esse resultado discorda dos obtidos pela análise conjunta a qual verificou interação GxE para todos os caracteres mensurados, porém suporta as interações GxE observadas para os demais caracteres. Analisando a E5, pode-se entrever que Sinop foi o local que apresentou maior influência no efeito da interação GxE. Essa interpretação é suportada pela ausência de similaridade entre os genótipos selecionados na E1 em Sinop com os selecionados nas demais estratégias (Figuras 2 e 5). As estratégias E4 e E5 apresentaram, apesar do ranqueamento divergente, elevada similaridade entre os genótipos selecionados, os quais foram, respectivamente, 3, 4, 15 e 18 e 3, 4, 7 e 15 (Figura 5). Os ganhos com a seleção na E4 e E5 foram no sentido desejável para todos os caracteres, exceto para HEM, FDN e FDA na E4 e para FLOR e HEM na E5 (Figura 6). Os ganhos no sentido indesejado para o FLOR na E5 deve-se principalmente pela seleção do genótipo 7 o qual possui elevada precocidade frente aos demais genótipos avaliados neste estudo.

Ao empregar as estratégias E4 e E5 foi possível observar que apesar do termo capitalizar ser utilizado para descrever a E4, a interação GxE, não é capitalizada verdadeiramente, apesar da interação continuar “inflada” nos valores genotípicos preditos por local. Assim, acredita-se que a E4 possibilita vislumbrar interações do tipo complexa entre as fontes de variação e entrever o fator (local, safra,...) dentro da fonte de variação que possui maior influência no efeito da interação, e visa minimizar esse efeito para a seleção e recomendação dos genótipos. Dessa forma, na E4, busca-se selecionar e recomendar genótipos para locais que apresentam o mesmo padrão de interação ou diferentes padrões daquele da rede experimental avaliada; enquanto na E5 ao, realmente capitalizar a interação GxE e tentar minimizar ela busca-se selecionar e recomendar os genótipos para os locais que apresentem o mesmo padrão da interação ou pequenas alterações neste padrão. Porém, o uso dessas estratégias não é recomendado quando se tem um número elevado de caracteres e de fatores na fonte de variação – elevado número de locais, safras, entre outros - pois dificulta a interpretação biológica dos fatores e eleva o erro nas estimativas (CRUZ et al., 2014).

De maneira geral, considerando o emprego das cinco estratégias, pode-se observar que houve similaridade dos genótipos selecionados entre elas sendo de 25% entre: E1 Janaúba e E1 Sete Lagoas, E1 Janaúba e E4/E5, E2/E3 e E5; e de 50% entre: E1 Janaúba e E2/E3, E1 Sete Lagoas e E2/E3/E4/E5, E2/E3 e E4; de 75% entre E4 e E5; e de 100% entre E2 e E3. Apesar da similaridade variar entre as estratégias empregadas é possível destacar a superioridade, previsibilidade e adaptabilidade do genótipo 4, o qual foi selecionado em todas as estratégias empregadas, com exceção da E1 em Sinop. Além disso, destaca-se o uso dos genótipos 4, 13, 17 e 18 para locais que apresentem ou não padrões de interações similares aos locais avaliados neste estudo.

O ganho de seleção variou entre as estratégias empregadas, sendo no sentido desejável para a maioria dos caracteres mensurados (Figura 6). A HEM foi a única característica que apresentou ganhos no sentido indesejado em todas as estratégias empregadas. Como as estratégias são empregadas para finalidades diferentes, a comparação entre elas por meio do ganho de seleção não faria sentido. Com isso, tem-se que o uso das estratégias deverão ser empregadas para a seleção e recomendação de genótipos superiores considerando: i. exclusivamente o local avaliado - E1; ii. locais que apresentam o mesmo padrão de interação GxE daquele da rede experimental avaliada - E2 e E3; iii. locais que apresentem padrões de interação diferente daquele da rede experimental avaliada – E2.

As estratégias E4 e E5 são recomendadas com ressalvas para a seleção e recomendação dos genótipos superiores, considerando: i. locais que apresentem ou não

mesmo padrão de interação GxE daquele da rede experimental avaliada – E4 e ii. locais que apresentem o mesmo padrão da interação ou pequenas alterações neste padrão – E5, porém ambas estratégias devem ser empregadas em experimentos que possuem pequeno número de caracteres e de fatores da fonte de variação da interação sob avaliação. Sendo o uso dessas duas estratégias interessante para vislumbrar interações do tipo complexa entre as fontes de variação e entrever o fator (local, safra,...) dentro da fonte de variação que possui maior influência no efeito da interação, bem como minimizar esse efeito para a seleção e recomendação dos genótipos.

5 CONCLUSÃO

Os genótipos 4, 13, 17 e 18 destacaram-se quando a superioridade, previsibilidade e adaptabilidade genotípica sendo interessante o uso destes visando a produção de etanol de segunda geração. O uso da estratégia E2 e E3 foram as mais adequadas para seleção desses genótipos pelo índice FAI-BLUP considerando experimentos em múltiplos locais e visando a recomendação, respectivamente, para locais que não apresentam e/ou apresentam mesmo padrão de interação GxE da rede experimental avaliada.

REFERÊNCIAS

- ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; GONÇALVES, J.L.M.; SPAROVEK, G. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**. v. 22, p.711-728. 2013.
- ALMEIDA, G.Q., ROCHA, J.R.A.S.C., PESSOA, H.P., ALVES, F.M., CHAVES, L.J. Selection of *Hanicornia speciosa* germplasm accessions based on the FAI-BLUP index. **Pesq. Agropec. Trop.**, v.51. 2021.
- ALMEIDA, L.G.F., PARRELLA, R.A.C., SIMEONE, M.L.F., RIBEIRO, P.C.O., DOS SANTOS, A.S., DA COSTA, A.S.V., SCHAFFERT, R. E. Composition and growth of sorghum biomass genotypes for ethanol production. **Biomass and Bioenergy**, v.122, p.343–348. 2019.
- ALVES, R.S., TEODORO, P.E., PEIXOTO, L.A., SILVA, L.A., LAVIOLA, B.G., RESENDE, M.D.V., BHERING, L.L., Multiple-trait BLUP in longitudinal data analysis on *Jatropha curcas* breeding for bioenergy. **Ind. Crops Prod**, v. 130, p. 558–561. 2019.
- BASTOS, I.T., BARBOSA, M.H.P., RESENDE, M.D.V., PETERNELII, L.A., SILVEIRA, L.C.I., DONDA, L.R., FORTUNATO, A.A., COSTA, P.M.A., FIGUEIREDO, I.C.R. Avaliação da interação genótipo x ambiente em cana-de-açúcar via modelos mistos. **Pesquisa Agropecuária Tropical**, 37, 195-203. 2007.
- BOTEGA, G.P. **Eficiência do índice FAI-BLUP na seleção de genótipos Bourbon**. Dissertação (Mestrado em Genética e Melhoramento de plantas) – Universidade Federal de Lavras, Lavras, 2019.
- BOTELHO, T.T., LEITE, P.S.S., PARRELLA, R.A.C., NUNES, J.A.R. Strategies for multi-trait selection of sweet sorghum progenies. **Crop Breeding and Applied Biotechnology**, v. 21, p.1-9. 2021.
- CASTRO, E., NIEVES, I.U., RONDÓN, V., SAGUES, W.J., FERNÁNDEZ-SANDOVAL, M.T., YOMANO, L.P. Potential for ethanol production from different sorghum cultivars. **Ind. Crops Prod**, v. 109, p.367–373. 2017.
- CASTRO, F.M.R., LOMBARDI, G.M.R., NUNES, J.A.R., PARRELLA, R.A.C., BRUZI, A.T. Accumulation of biomass and lignocellulosic compounds in photoperiod-sensitive biomass sorghum genotypes. **Biomass and Bioenergy**. 2021 (prelo)
- CASTRO, F.M.R.; BRUZI, A.T.; NUNES, J.A.R.; PARRELLA, R.A.C.; LOMBARDI, G.M.R.; ALBUQUERQUE, C.J.B.; LOPES, M. Agronomic and Energetic Potential of Biomass Sorghum Genotypes. **American Journal of Plant Science**, v. 6,p1-14. 2015.
- CRUZ, C.D., CARNEIRO, P.C.S. **Modelos Biométricos aplicados ao melhoramento genético**. Viçosa: UFV, v.2, 2003.

CRUZ, C.D., CARNEIRO, P.C.S. REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético (3. ed.)**. Viçosa: Editora UFV. v.3, 2014.

DA SILVA, M.J., CARNEIRO, P.C.S., CARNEIRO, J.E.S., DAMASCENO, C.M.B., PARRELLA, N.N.L.D., PASTINA, M.M., SIMEONE, M.L.F., SCHAFFERT, R.E., PARRELLA, R.A.C. Evaluation of the potential of lines and hybrids of biomass sorghum. **Ind. Cro. and Prod**, v.125, p.379-385. 2018.

DOS SANTOS, J.P.R., FERNANDES, S.B., MCCOY, S., LOZANO, R., BROWN, P.J., LEAKEY, A.D.B., BUCKLER, E.S., GARCIA, A.A.F., GORE, M.A. Novel Bayesian Networks for Genomic Prediction of Developmental Traits in Biomass Sorghum. **G3 (Bethesda)**, v.10, n.2, p.769-781. 2020.

FERNANDES, S.B.; DIAS, K.O.G.; FERREIRA, D.F.; BROWN, P.J. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. **Theoretical and Applied Genetics**, v.131, n.3, p.747–755. 2018.

GURAGAINA, Y.N., GANESH, K.M., BANSAL, S., SATHISH, R.S., RAO, N., VADLANI, P.V. Low-lignin mutant biomass resources: Effect of compositional changes on ethanol yield. **Ind. Crop. Prod.** v.61 p.1–8. 2014.

HABYARIMANA, E., LOPEZ-CRUZ, M., BALOCH, F.S. Genomic Selection for Optimum Index with Dry Biomass Yield, Dry Mass Fraction of Fresh Material, and Plant Height in Biomass Sorghum. **Genes**, v.11, n.1, p.61-68. 2020.

HAZEL, L.N. The genetic basis for constructing selection indexes. **Genetics**, v. 28, p. 476-490, 1943.

JOSHI, R., SINGLA-PAREEK, S.L., PAREEK, A. Engineering abiotic stress response in plants for biomass production. **Journal of Biological Chemistry**, v. 293, n.14, p.5035–5043. 2018.

KAISER H. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, p.187–200. 1958.

KEBROM, T.H., MCKINLEY, B., MULLET, J.E. Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. **Biotechnol Biofuels**, v.10, p.159-167. 2017.

LI, M., WANG, J., DU, F., DIALLO, B., XIE, G.H. High-throughput analysis of chemical components and theoretical ethanol yield of dedicated bioenergy sorghum using dual-optimized partial least squares calibration models. **Biotechnol Biofuels**, v.10, 206-218. 2017.

LUTZ, M. **Python Pocket Reference: Python in your pocket**, 5th Edition. O'Reilly Media, Inc. 2014.

MAIA, M.C.C., RESENDE, M.D.V., PAIVA, J.R., CAVALCANTI, J.J.V., BARROS, L. M. Seleção simultânea para produção, adaptabilidade e estabilidade genotípicas em clones de cajueiro, via modelos misto. **Pesquisa Agropecuária Tropical**, 39, 43-50. 2009.

- MEIER, C., MARCHIORO, V.M., MAIRA, D., OLIVOTO, T., KLEIN, L.A. Genetic parameters and multiple-trait selection in wheat genotypes. **Pesq. Agropec.Trop.** v.51, 2021.
- MULAMBA, N.N.; MOCK, J.J. Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. **Egypt Journal of Genetics and Cytology**, Alexandria, v. 7, p. 40- 51, 1978.
- OLIVEIRA, I.C.M.; MARÇAL, T. de S.; BERNARDINO, K. da C.; RIBEIRO, P.C. de O.; PARRELLA, R.A. da C.; CARNEIRO, P.C.S.; SCHAFFERT, R.E.; CARNEIRO, J.E. de S. Combining ability of biomass sorghum lines for agroindustrial characters and multitrait selection of photosensitive hybrids for energy cogeneration. **Crop Science**, v.59, p.1554-1566, 2019.
- OLIVEIRA, L.G.M., MARQUES, A., LOPES, E.D., GONÇALVES, J.F., MARTINS, N.S., PENA, C.A.A, ARBEX, D.C., LAIA, M.L. Produtividade, adaptabilidade e estabilidade genotípica de clones de *Eucalyptus* spp. e *Corymbia* spp. em diferentes espaçamentos de plantio. **Scientia Forestalis**, v.49, p.131-136. 2021.
- OLIVOTO, T.; LÚCIO, A.D. metan: an R package for multi-environment trial analysis. **Methods Ecol Evol.** v.11, p.783-789. 2020.
- OLSON, S.N., RITTER, K., ROONEY, W., KEMANIAN, A., MCCARL, B.A., ZHANG, Y., HALL, S., PACKER, D., MULLET, J. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. **Biofuels, Bioproducts and Biorefining**, v.6, n.6, p.640–655. 2012.
- PEDREGOSA, F., VAROQUAZ, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **JMLR**, v.12, p.2825-2830. 2011.
- PEIXOTO, M.A., COELHO, I.F., EVANGELISTA, J.S.P.C., SANTOS, S.S.O., ALVES, R.S., PINTO, J.F.N., REIS, E.F., BHERING, L.L. Selection of maize hybrids: an approach with multi-trait, multi-environment, and ideotype-design. **Crop Breed. Appl. Biotechnol.** v.21, n.2, p.1-12. 2021
- R CORE TEAM. **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria. 2018.
- RAMALHO, M.A.P.; ABREU, A.F.B.; SANTOS, J.B.; NUNES, J.A.R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas.** 1ª Edição. Lavras- MG. Editora: UFLA, 2012. 522p.
- ROCHA, J.R.A.S.C.; MACHADO, J.C.; CARNEIRO, P.C.; Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy**, v.10, p. 52-60, 2018.

ROCHA, J.R.A.S.C.; NUNES, K.V.; CARNEIRO, A.L.N.; MARÇAL, T.D.S.; SALVADOR, F. V.; CARNEIRO, P.C.S.; CARNEIRO, J.E.S. Selection of superior inbred progenies toward the common bean ideotype. **Agronomy Journal**, v. 111, n. 1, p. 1181-1189, 2019.

SALVADOR, F.V. **Seleção recorrente no melhoramento do feijoeiro: otimização das etapas de seleção e recombinação de progênies**. Tese (Doutorado em Genética e Melhoramento de plantas) – Universidade Federal de Viçosa, Viçosa, 2018.

SANJARI, S., SHIRZADIAN-KHORRAMABAD, R., SHOBBAR, ZS. ET AL. Systematic analysis of NAC transcription factors' gene family and identification of post-flowering drought stress responsive members in sorghum. **Plant Cell Rep**, v.38, p.361–376. 2019.

SCOTT, V., HASZELDINE, R.S., TETT, S.F.B., OSCHLIES, A. Fossil fuels in a trillion tonne world. **Nature Clim. Change**, v. 5, p. 419-423. 2015.

SIMEONE, M.L.F., PARRELLA, R.A.C., DAMASCENO, C.M.B., SCHAFFERT, R.E. Prediction of high-biomass sorghum quality using near infrared spectroscopy to monitoring calorific value, moisture, and ash content. **International Journal of Development Research**, v.10, n.9, p.40916-40920. 2020.

SMITH, H. F. A discriminant function for plant selection. *Annual Eugenics*, v. 7, p. 240-250, 1936.

SOEST, P.J.V. Use of detergents in the analysis of fibrous feeds. II. A rapid method for the determination of fiber and lignin. **Journal of the A.O.A.C.**, v.46. n.5, p.1-14. 1963.

SOUZA, M.H., PEREIRA JUNIOR, J.D., STECKLING, S.M., MENCALHA, J., DIAS, F.S., ROCHA, J.R.A.S.C., CARNEIRO, P.C.S., CARNEIRO, J.E.S. New genotypic adaptability and stability analyses using Legendre polynomials and genotype-ideotype distances. **BioRxiv**. Online version. 2020.

TERASAWA JUNIOR, F., VENCOSKY, R., KOEHLER, H. Environment and genotype-Environment interaction in maize breeding in Paraná, Brazil. **Crop Breed. Appl. Biotechnol.** v.8, p.17-22, 2008.

TORRES, F.E., TEODORO, P.E., SAGRILO, E., CECCON, G., CORREA, A.M. Interação genótipo x ambiente em genótipos de feijão-caupi semiprostrado via modelos mistos. **Bragantia**. v.10, p.159-167. 2015.

VAN DER WEIJDE, T.; TORRES, A.F.; DOLSTRA, O.; DECHESNE, A.; VISSER, R.G.F.; TRINDADE, L.M. Impact of Different Lignin Fractions on Saccharification Efficiency in Diverse Species of the Bioenergy Crop *Miscanthus*. **Bioenergy Research**, v.9, p.146-156. 2016.

VENDRUSCOLO, T.P.S.; BARELLI, M.A.A.; CASTRILLON, M.A.S.; SILVA, R.S.; OLIVEIRA, F.T.; CORREA, C.L.; ZAGO, B.W.; TARDIM, F.D. Correlation and path analysis of biomass sorghum production. **Genetics and Molecular Research**, v.15, n.4, p.1-16, 2016.

WILLIAMS, J.S. The evaluation of a selection index. **Biometrics**, North Carolina, v.18, p.375-393, 1962.

WOYANN, L.G.; MEIRA, D.; ZDZIARSKI, A.D.; MATEI, G.; MILIOLI, A.S.; ROSA, A.C.; MADELLA, L.A.; BENIN, G. Selection indexes based on linear-bilinear models applied to soybean breeding. **Agronomy Journal**. v.112, p.175-182, 2020

XIA, J., ZHAO, Y., BURKS, P., PAULY, M., BROWN, P. J. A sorghum NAC gene is associated with variation in biomass properties and yield potential. **Plant Direct**, v.2, n.7, p.1-12. 2018.

ZHAO, Y.L.; A. DOLAT, Y. STEINBERGER, X.; WANG, A.; OSMAN; XIE, G.H. Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. **Field Crops Research**, v.111, n.1, p.55–64. 2009.