Universidade Federal de São João del-Rei
Programa de Pós-Graduação em Bioengenharia

VANDER FILLIPE DE SOUZA

MAPEAMENTO GENÉTICO E ANÁLISE FUNCIONAL PARA CARACTERES AGROINDUSTRIAIS EM SORGO PARA PRODUÇÃO DE BIOENERGIA

SÃO JOÃO DEL REI
MINAS GERAIS – BRASIL
FEVEREIRO DE 2016

VANDER FILLIPE DE SOUZA

MAPEAMENTO GENÉTICO E ANÁLISE FUNCIONAL PARA CARACTERES AGROINDUSTRIAIS EM SORGO PARA PRODUÇÃO DE BIOENERGIA

Tese submetida ao Programa de Pós-Graduação em Bioengenharia da Universidade Federal de São João del-Rei como parte dos requisitos necessários para a obtenção do título de *Doctor Scientiae*.

SÃO JOÃO DEL REI
MINAS GERAIS – BRASIL
FEVEREIRO DE 2016

Ficha catalográfica elabora pelo Setor de Processamento Técnico da Divisão de Biblioteca da UFSJ[i]

Souza, Vander Fillipe de
S729m Mapeamento genético e análise funcional para caracteres agroindustriais em sorgo para produção de bioenergia [manuscrito] / Vander Fillipe de Souza. – 2016.
104f. ; il.

Orientador: Cynthia Maria Borges Damasceno.

Tese (doutorado) – Universidade Federal de São João del-Rei. Departamento de Engenharia de Biossistemas.
Inclui referências.

1. Biocombustível 2. Genotyping-by-Sequencing 3. GWAS 4. QTL 5. RT-qPCR 6. Sorghum bicolor I. Damasceno, Cynthia Maria Borges (orientador) II. Universidade Federal de São João del- Rei. Departamento de Engenharia de Biossistemas III. Título

CDU 57.08:664.788

AGRADECIMENTOS

Agradeço a Deus, pela força de vontade para concluir mais essa etapa.

À Universidade Federal de São João del-Rei e ao Programa de Pós-Graduação em Bioengenharia, pela oportunidade e pela concessão da bolsa de estudos.

À Embrapa Milho e Sorgo, por oferecer toda a estrutura para a realização desse trabalho.

À Dra. Cynthia Damasceno, por aceitar me orientar e pelo apoio durante o doutorado.

À Dra. Maria Marta Pastina, pela coorientação e por todos os auxílios e ensinamentos sobre análises estatísticas.

Ao Dr. Luciano da Costa e Silva e ao Prof. Dr. Antonio Augusto Franco Garcia, pela participação na banca e pelas contribuições na tese.

Ao Dr. Robert Schaffert e ao Dr. Rafael Parrella, por toda ajuda desde o meu mestrado e por todos os ensinamentos sobre a cultura do sorgo.

Ao Dr. Jurandir Magalhães, pelas sugestões e contribuições, principalmente nas análises de mapeamento associativo.

À Dra. Maria Lucia Simeone, pelo apoio nas análises de caracterização da biomassa.

Ao Dr. Roberto Noda, pelo auxílio no processamento dos dados genotípicos.

Ao Dr. Guilherme da Silva Pereira, por compartilhar seus conhecimentos sobre mapeamento de QTLs e programação em R.

À Dra. Beatriz de Almeida Barros, pelas instruções e auxílios nas análises realizadas no laboratório de biologia molecular.

A todos do galpão de melhoramento de sorgo, do núcleo de biologia aplicada e dos laboratórios de composição centesimal e análise do caldo de sorgo da Embrapa Milho e Sorgo.

Aos meus familiares, principalmente à minha mãe Maria Augusta, à Carla de Fátima e aos familiares dela, pelo estímulo e pelo apoio.

A todos os colegas e amigos que fiz na Embrapa e na UFSJ, pelas ideias compartilhadas.

“Toda a nossa ciência, comparada com a realidade, é primitiva e infantil
e, no entanto, é a coisa mais preciosa que temos”
(*Albert Einstein*)

RESUMO

SOUZA, Vander Fillipe de (DS). Universidade Federal de São João del Rei, Fevereiro de 2016. **MAPEAMENTO GENÉTICO E ANÁLISE FUNCIONAL PARA CARACTERES AGROINDUSTRIAIS EM SORGO PARA PRODUÇÃO DE BIOENERGIA**. Orientadora: Cynthia Maria Borges Damasceno. Coorientadora: Maria Marta Pastina.

Devido à demanda crescente por recursos energéticos renováveis, o interesse por culturas agrícolas com versatilidade para geração de biocombustíveis e cogeração de eletricidade é crescente. O sorgo possui características que o torna uma promissora matéria-prima para produção de bioenergia, que incluem desde a produção de açúcares fermentescíveis no colmo até a produção de biomassa com composições diferenciadas para a combustão em caldeiras ou para a produção de etanol de segunda geração. O objetivo do presente trabalho foi estudar o controle genético dos caracteres relacionados à produção de bioenergia em genótipos de sorgo. Para isso, foram utilizadas estratégias de mapeamento genético para caracteres de interesse agroindustrial e a expressão de genes relacionados à biossíntese de lignina via PCR em tempo real. Uma população de RILs de sorgo sacarino composta por 223 indivíduos e um painel de diversidade genética de sorgo contendo 100 linhagens foram genotipados por sequenciamento e utilizados, respectivamente, para o mapeamento de QTLs e o mapeamento associativo. Sessenta genótipos do painel também foram utilizados para o ensaio de expressão gênica. Para o mapeamento de QTLs na população de RILs de sorgo sacarino, cujos genitores, Brandes e Wray, são contrastantes para o teor de açúcares, foi utilizada a abordagem de mapeamento por múltiplos intervalos para múltiplos caracteres em três safras. Ao todo foram identificados 65 QTLs para os oito caracteres analisados, alguns foram comuns entre os caracteres, o que sugere ligação ou pleiotropia, e todos os caracteres apresentaram ao menos um efeito epistático significativo. Para o mapeamento associativo foi utilizada uma abordagem de modelos mistos com base em múltiplos locos, incluindo as matrizes K e Q. Ao todo foram mapeados treze SNPs significativos, considerando a correção de Bonferroni (alpha = 0,05) para múltiplos testes, para os cinco caracteres avaliados. A fim de melhor entender a síntese de lignina em sorgo e

identificar possíveis alvos para o programa de melhoramento, um estudo de expressão gênica foi realizado para vários genes da via. Dos vários genes testados, apenas um gene da família que codifica enzimas hidroxicianomil transferase, *HCT1*, apresentou correlação significativa (*p*-valor < 0,01; r = 0,76) com o teor de lignina. No todo, os resultados do presente trabalho contribuem para o melhor entendimento da arquitetura genética de caracteres relacionados à produção de bioenergia, e, também, para a futura aplicação da seleção assistida por marcadores moleculares no programa de melhoramento de sorgo.

Palavras-chave: Biocombustível; *Genotyping-by-Sequencing*; GWAS; QTL; RT-qPCR; *Sorghum bicolor*.

ABSTRACT

SOUZA, Vander Fillipe de (DS). Federal University of São João del Rei, February 2016. **GENETIC MAPPING AND FUNCTIONAL ANALYSIS FOR AGRO-INDUSTRIAL TRAITS IN SORGHUM FOR BIOENERGY PRODUCTION**. Advisor: Cynthia Maria Borges Damasceno. Co-advisor: Maria Marta Pastina.

Due to growing demand for renewable energy sources, interest in crops dedicated to biofuel production and cogeneration is expanding. Sorghum has several traits that make it a promising feedstock for bioenergy production, ranging from presence of fermentable sugars in the stalk for straightforward ethanol production, to high biomass with differential compositions for biomass combustion in furnaces or for second-generation ethanol production. The present work aimed to study the genetic control of traits related to the production of bioenergy in sorghum genotypes. To conduct this study, different genetic mapping strategies were used for studying traits associated to biomass production and quality, as well as gene expression analysis related to lignin biosynthetic pathway. A RIL population of 223 lines and a diversity panel of 100 individuals were genotyped by sequencing (GBS) and used for QTL mapping and association mapping, respectively, while a panel subgroup of sixty contrasting genotypes for lignin content were used for gene expression assays. A multivariate approach using mixed models was applied for QTL mapping of bioenergy related traits using the sweet sorghum RIL population, whose genitors Brandes and Wray contrast for juice sugar content. In total, 65 QTLs were identified for the eight traits analyzed, some QTLs collocated and all traits had at least one significant epistatic effect. Kinship matrices and population structure were used in the association mapping analysis performed on the diversity panel, as well as SNPs selected according to a model for multiple loci. In total, thirteen significant SNPs were mapped, according to the Bonferroni correction ($\alpha = 0.05$), for the five traits evaluated. To identify potential breeding targets and also better understand lignin synthesis in sorghum, a lignin gene expression analysis was performed. From the several genes tested, only one gene belonging to hydroxycyanomil transferase family, *HCT1*, showed significant correlation ($p < 0.01$, $r = 0.76$) with lignin content. These findings contribute to a better understanding of the genetic architecture of traits related to bioenergy production, and

to the future implementation of marker-assisted selection in sorghum breeding programs.

Keywords: Biofuel; *Genotyping-by-Sequencing*; GWAS; QTL; RT-qPCR; *Sorghum bicolor*.

SUMÁRIO

1. INTRODUÇÃO

1.1. A IMPORTÂNCIA DO SORGO NO CONTEXTO DA PRODUÇÃO DE BIOENERGIA

A cultura do sorgo (*Sorghum bicolor)* apresenta ampla diversidade genética, utilizada para geração de genótipos específicos para diversos produtos. O sorgo granífero, por exemplo, possui baixo porte e elevada produção de grãos, com foco na produção de ração para animais monogástricos. O sorgo forrageiro apresenta porte superior, com produção de massa verde e de grãos proporcionais aos valores necessários para a nutrição de ruminantes. O sorgo pastejo, cruzamento entre *Sorghum bicolor* x *Sorghum sudanense,* possui alta capacidade de perfilhamento e rebrota, destinados à formação de pastagens (Sawazaki 1998). Além das cultivares desenvolvidas para alimentação animal, genótipos de sorgo foram melhorados especificamente para a produção de biocombustíveis e cogeração de eletricidade (Naik *et al.* 2010).

As cultivares de sorgo sacarino são destinadas à produção de bioetanol a partir dos açúcares presentes no caldo do colmo. Em média, os genótipos de sorgo sacarino apresentam entre 13 e 24 ºBrix de sólidos solúveis no caldo, com teor de açúcares entre 7 e 15% (Reddy *et al.* 2005; Almodares & Hadi 2009). Estes valores são similares aos observados na cana-de-açúcar, por exemplo. Além disso, o processamento agroindustrial do sorgo é o mesmo da cana, o que permite a utilização da mesma infraestrutura agroindustrial montada, com a possibilidade de processamento misto. O que possibilita a redução no período de ociosidade das usinas e as consequentes oscilações no preço do etanol (Yu *et al.* 2012).

De forma geral, apesar das possibilidades de retração do setor sucroalcoleiro, a demanda por etanol é crescente e os custos de produção tendem a decrescer ao longo do tempo, como demonstra projeções de Jonker *et al.* (2015). O sorgo sacarino pode fornecer matéria-prima complementar, em sistemas de produção que incluem desde o plantio na entressafra ou reforma de canaviais, ou em áreas sem aptidão para o cultivo da cana-de-açúcar.

De acordo com o Ministério da Agricultura, Pecuária e Abastecimento (Mapa), a produção de cana-de-açúcar no ano civil 2014 alcançou 631,8 milhões de toneladas. Este montante foi 2,5% inferior ao registrado no ano anterior, 648,1 milhões de toneladas. No entanto, a fabricação de etanol cresceu 3,3% e foi superior à dos anos anteriores, atingindo um montante de 28.526 mil m³, enquanto a produção de açúcar teve queda de 5%. Além disso, a quantidade de ATR (Açúcar Total Recuperável) na cana-de-açúcar, que é uma medida da qualidade da matéria-prima, teve queda de 2,71% (BEN 2015).

Desta forma, o sorgo sacarino é uma alternativa viável, tanto para o incremento de safras como para atender a uma demanda emergencial. Visto que seu rápido ciclo de cultivo, aproximadamente 120 dias, recompensa a taxa de produção de biomassa relativa à cana-de-açúcar (Zegada-Lizarazu & Monti 2012). Além disso, a cultura do sorgo é relativamente tolerante ao estresse hídrico, condição associada à redução da produtividade de cana (Conab 2015), e eficiente no uso da água e de nutrientes, o que pode tornar seu balanço energético mais eficiente do que o das demais culturas energéticas (Regassa & Wortmann 2014).

O bagaço do sorgo sacarino também pode ser utilizado na queima direta em caldeiras e cogeração de eletricidade. Estratégia atrativa atualmente devido ao baixo nível dos reservatórios das hidroelétricas, mas que permanecerá pertinente, uma vez que utiliza matéria-prima renovável para combustão (Khatiwada *et al.* 2016). Projeções realizadas por Jonker *et al.* (2015) e Khatiwada *et al.* (2016) também incluem a viabilização da produção de etanol de segunda geração neste cenário, que requer o pré-tratamento físico-químico da biomassa.

O etanol de segunda geração (2G) é obtido a partir da despolimerização dos componentes da parede celular, com a remoção da lignina, e a hidrólise ácida ou enzimática da celulose e hemicelulose para subsequente fermentação dos carboidratos simples resultantes (Vermerris *et al.* 2007). Com o objetivo de viabilizar esta tecnologia mais rapidamente, genótipos de sorgo estão sendo melhorados para essa finalidade. O sorgo biomassa, ou lignocelulósico, apresenta alta produtividade de biomassa sem caldo e sem açúcares, sendo destinado especificamente para produção de etanol 2G ou cogeração de eletricidade.

As cultivares de sorgo biomassa apresentam composição diferenciada da parede celular, com variação no teor lignina, para facilitar a hidrólise dos materiais ou para aumentar seu poder calorífico (Naik *et al.* 2010). Em geral, o sorgo biomassa é

sensível ao fotoperiodismo, ou seja, floresce apenas em condições de dias curtos (< 12 horas e 20 minutos). Esta característica possibilita a ampliação do estágio de desenvolvimento vegetativo conforme a época e o local de implantação dessa cultura no campo, o que por sua vez proporciona um significativo aumento no acúmulo total de biomassa (Rooney & Aydin 1999).

Neste contexto, o desenvolvimento de cultivares melhoradas é uma das etapas mais importantes para aumentar a produtividade e possibilitar a expansão da cultura do sorgo para produção de bioenergia. Como os principais caracteres de interesse agroindustrial são controlados por múltiplos genes, o entendimento do controle genético destes caracteres é fundamental para avanços nos processos de melhoramento genético.

1.2. ESTRATÉGIAS DE MAPEAMENTO GENÉTICO PARA CARACTERÍSTICAS QUANTITATIVAS

Com o advento dos marcadores moleculares, tornou-se possível o mapeamento genético de locos que controlam a expressão de caracteres quantitativos (do inglês, *Quantitative Trait Loci* - QTLs), assim como estimar suas posições no genoma, seus efeitos e os tipos de interações alélicas (Falconer & Mackay 1996). Essa estratégia também permitiu um melhor entendimento das bases genéticas das correlações entre caracteres e o estudo das interações entre QTLs e ambientes (Malosetti *et al*. 2008).

Os estudos de mapeamento genético foram iniciados a partir de marcadores morfológicos. Porém, os marcadores morfológicos apresentam ocorrência limitada e podem ser influenciados pelo ambiente. Os marcadores bioquímicos, embora sejam mais numerosos, também são influenciados pelo ambiente, o que pode, consequentemente, levar à identificação de falsos positivos e negativos em estudos de mapeamento genético. Já a utilização de marcadores moleculares permite detectar os polimorfismos no próprio material genético (Zargar *et al.* 2015).

Com o desenvolvimento da segunda geração de sequenciamento (do inglês, *Next-Generation Sequencing* - NGS), o custo do sequenciamento caiu intensamente, de modo que ele está sendo rotineiramente utilizado para genotipagem em larga

escala e desenvolvimento de marcadores moleculares (Wetterstrand 2014). A plataforma de genotipagem por sequenciamento (do inglês, *Genotyping-by-Sequencing* - GBS) é uma tecnologia NGS que pode ser utilizada para identificar polimorfismos em uma população (Elshire *et al.* 2011). Estes polimorfismos ocorrem devido a alterações de nucleotídeos, tais como transição e transversão, e indels, inserções e deleções. O polimorfismo de nucleotídeo único (do inglês, *Single Nucleotide Polymorphism* - SNP) tem ampla aplicação em pesquisas de melhoramento de plantas, incluindo o mapeamento de QTLs e os estudos de associação genômica ampla (do inglês, *Genome-Wide Association Study* - GWAS) (Zargar *et al.* 2015).

O mapeamento de QTLs é baseado em análises de ligação ou em análises de associação, sendo que o princípio de ambas é a existência de desequilíbrio de ligação (do inglês, *linkage disequilibrium* - LD) no genoma, que representa a associação não aleatória entre alelos de diferentes locos em uma população (Gupta *et al.* 2005). As análises de ligação são comumente conduzidas em populações experimentais obtidas a partir de cruzamentos controlados, como as populações biparentais. A precisão com a qual um QTL pode ser localizado em relação a um marcador é diretamente proporcional ao número de meioses ocorridas na população de mapeamento, ou seja, ao número de oportunidades de recombinação entre o marcador e o loco que controla o caractere. E quanto maior a população, maior será a probabilidade de ocorrer uma recombinação na região-alvo (Flint-Garcia *et al.* 2003).

Inicialmente, os modelos de mapeamento de QTLs via análise de ligação testavam apenas as diferenças entre as médias fenotípicas associadas ao marcador. Jansen (1993) e Zeng (1993, 1994) propuseram, independentemente, um método baseado em regressão múltipla que permitiu incluir tanto o efeito de um QTL como os efeitos de covariáveis, que são um subconjunto de marcadores selecionados, no modelo. Este método é descrito como mapeamento por intervalo composto (do inglês, *Composite Interval Mapping* - CIM). Posteriormente foi proposto o mapeamento por múltiplos intervalos (do inglês, *Multiple Interval Mapping* - MIM) que considera os próprios efeitos de QTLs mapeados como cofatores no modelo (Kao *et al.* 1999).

Em contraste com o método de mapeamento em populações biparentais, o mapeamento associativo, ou GWAS, explora a diversidade alélica e a recombinação histórica e evolutiva em um conjunto de linhagens diversas. Logo, o mapeamento associativo é capaz de proporcionar maior resolução que o mapeamento de ligação.

Entretanto, as duas abordagens são complementares uma vez que o mapeamento em uma população biparental oferece maior poder estatístico para a detecção de QTLs (Yu & Buckler 2006).

Enquanto uma população biparental apresenta em média o mesmo grau de relacionamento entre os indivíduos, em um estudo de associação, as populações apresentam um complexo padrão de subdivisões (estrutura populacional) e de parentesco que podem ser estimados a partir dos marcadores. Esta é uma questão fundamental para a análise de associação, uma vez que um determinado marcador pode ser afetado por um conjunto de efeitos correlatos do restante do genoma e resultar na identificação de falsos positivos no mapeamento (Hoffman 2013).

O controle de associações espúrias é melhorado quando a estrutura populacional e o grau de parentesco entre os indivíduos são levados em consideração no modelo GWAS, ajustando uma matriz de efeitos fixos e aleatórios, respectivamente. Como o LD não necessariamente resulta de ligação física, já que diversos fatores podem causá-lo, como epistasia, seleção, deriva genética, e a própria estrutura populacional, entre outros, o controle dos falsos positivos permite uma maior precisão na identificação de regiões potenciais relacionadas com os caracteres de interesse (Yu *et al.* 2006).

Assim, tanto o mapeamento de QTLs como o associativo podem ser empregados como uma importante ferramenta para o melhoramento de plantas, uma vez que permitem a identificação de marcadores ou genes relacionados a características fenotípicas de interesse. As informações geradas podem auxiliar no entendimento do controle genético dos caracteres e os SNPs mapeados podem ser empregados na seleção assistida por marcadores, contribuindo significantemente para acelerar o desenvolvimento de novas cultivares com alta biomassa e composição adequada aos processos de produção de biocombustíveis em programas de melhoramento de sorgo.

1.3. ANÁLISE DE EXPRESSÃO GÊNICA VIA PCR QUANTITATIVO EM TEMPO REAL

Além da estratégia de mapeamento de locos de características quantitativas em populações biparentais e painéis de diversidade, outras técnicas podem ser utilizadas para a identificação de genes candidatos relacionados aos caracteres de interesse. Uma das possibilidades é o uso de homologia de sequências de genes previamente estudados em outras culturas. Essa é uma estratégia interessante quando há indícios de que o controle genético permanece conservado entre as diferentes espécies (Li *et al.* 2008).

Com o uso do algoritmo *Basic Local Alignment Search Tool* (BLAST) é possível a comparação de sequências de diversos genomas e obter a relação de compatibilidade entre as sequências (Altschu *et al.* 1990). As sequências-alvo podem ser estudadas por técnicas de reação em cadeia da polimerase (do inglês, *Polymerase Chain Reaction* - PCR) com primers específicos. Entretanto, a funcionalidade, ou a expressão do gene de interesse, deve ser verificada a partir dos níveis de mRNA. Este é um ponto importante, uma vez que os genes podem apresentar expressão diferencial em função dos diferentes tecidos, estágios e estímulos ambientais (Zhao & Dixon 2011).

A reação de transcriptase reversa do mRNA seguida da reação em cadeia da polimerase quantitativa em tempo real (do inglês, *real-time quantitative reverse transcription* PCR, ou RT-qPCR) é uma importante técnica que permite quantificar a expressão gênica de alvos específicos (Heid *et al.* 1996). Embora existam outros métodos utilizados para medir a expressão do mRNA, o RT-qPCR é um dos mais sensíveis graças ao processo de amplificação exponencial. A quantificação dos produtos de RT-qPCR é possível através da detecção de moléculas fluorescentes que intensificam a radiação em função da amplificação do material genético, como os corantes SYBR Green e as sonda de hibridação TaqMan (Peinnequin *et al.* 2004).

Genes que codificam enzimas envolvidas na biossíntese de lignina foram analisados em diversas espécies de interesse comercial como tabaco, milho e alfafa (Li *et al.* 2008). Desta forma, a técnica de RT-qPCR pode ser utilizada para o estudo da expressão destes genes na cultura do sorgo, assim como para quantificar as diferenças de expressão relativas aos diferentes genótipos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMODARES, A.; MOSTAFAFI DARANY, S.M. (2006). Effects of planting date and time of nitrogen application on yield and sugar contente of sweet sorghum. *J. Environmental Biology* 27:601-5.

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215:403-410.

BEN (2015). Balanço Energético Nacional 2015: Ano base 2014. Empresa de Pesquisa Energética - EPE.

CONAB (2015). Acompanhamento da safra brasileira de cana-de-açúcar: Safra 2015/16, Terceiro levantamento. Companhia Nacional de Abastecimento.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6(5).

FALCONER, D.S.; MACKAY, T.F.C. (1996). Introduction to Quantitative Genetics. Ed. 4. Longman, New York, USA, 464 p.

FLINT-GARCIA, S.A.; THORNSBERRY, J.M.; BUCKLER, E.S. (2003). Structure of linkage disequilibrium in plants. *Annuals Reviews in Plant Biology* 54:357-374.

GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L. (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant. Mol. Biol.* 57(4):461-485.

HEID, C.A.; STEVENS, J.; LIVAK, K.J.; WILLIAMS, P.M. (1996). Real Time Quantitative PCR. *Genome Res.*6:986–994.

HOFFMAN, G.E. (2013). Correcting for Population Structure and Kinship Using the Linear Mixed Model: Theory and Extensions. *Plos One* 8(10).

JANSEN, R.C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205–211.

JONKER, J.G.G.; HILST, F.V.D.; JUNGINGER, H.M.; CAVALETT, O.; CHAGAS, M.F.; FAAIJ, A.P.C. (2015). Outlook for ethanol production costs in Brazil up to 2030, for different biomass crops and industrial technologies. *Applied Energy* 147:593–610.

KAO, C.H.; ZENG, Z.B.; TEASDALE, R.D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* 152(3):1203-1216.

KHATIWADA, D.; LEDUC, S.; SILVEIRA, S.; MCCALLUM, I. (2016). Optimizing ethanol and bioelectricity production in sugarcane biorefineries in Brazil. *Renewable Energy* 85:371-386.

Li, X.; Weng, J.; Chapple C. (2008). Improvement of biomass through lignin modification. *The Plant Journal* 54:569–581.

MALOSETTI, M., RIBAUT, J.M., VARGAS, M., CROSSA, J., VAN EEUWIJK, F. (2008). A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (Zea mays L.). *Euphytica* 161:241-257.

NAIK, S.N., GOUD, V.V., ROUT, P.K., DALAI, A.K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* 14(2):578-597.

PEINNEQUIN, A.; MOURET, C.; BIROT, O.; ALONSO, A.; MATHIEU, J.; CLARENÇON, D.; AGAY, D.; CHANCERELLE, Y.; MULTON, E. (2004). Rat pro-inflammatory cytokine and cytokine related mRNA quantification by real-time polymerase chain reaction using SYBR green. *BMC Immunology* 5:3.

REDDY, B.V.S.; RAMESH, S.; REDDY, P.S.; RAMAIH, B.; SALIMATH, P.M.; KACHAPUR, R. (2005). Sweet sorghumea potential alternative raw material for bio-ethanol and bio-energy. Int Sorghum Millets News 46:79-86.

REGASSA, T.H.; WORTMANN, C.S. (2014). Sweet sorghum as a bioenergy crop: Literature review. *Biomass and Bioenergy* 64:348-355.

ROONEY, W.L.; AYDIN, S. (1999). Genetic control of a photoperiod-sensitive response in Sorghum bicolor (L.) Moench. *Crop Science* 39:397-400.

SAWAZAKI, E. (1998). Sorgo forrageiro ou misto, sorgo granífero, sorgo vassoura Sorghum bicolor L. Moench. In: FALH, J. L. Instruções agrícolas para as principais culturas econômicas 6:44-49.

VERMERRIS, W.; SABALLOS, A.; EJETA, G.; MOSIER, N.S.; LADISCH, M.R.; CARPITA, N.C. (2007). Molecular breeding to enhance ethanol production from corn and sorghum stover. *Crop Science* 47(S3):S142-S153.

WETTERSTRAND, K.A. (2014). DNA sequencing costs: data from the NHGRI large-scale genome sequencing program. 2011 Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcosts. Accessed [17 january 2016].

YU, J. et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203–208.

YU, J.; BUCKLER, E.S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17:155–160.

YU, J.; ZHANG, T.; ZHONG, J.; ZHANG, X.; TAN, T. (2012). Biorefinery of sweet sorghum stem. *Biotechnology Advances* 30:811–816.

ZARGAR, S.M.; RAATZ, B.; SONAH, H.; MUSLIMANAZIR; BHAT, J.A.; DAR Z.A.; AGRAWAL, G.K.; RAKWAL, R. (2015). Recent Advances in Molecular Marker

Techniques: Insight into QTL Mapping, GWAS and Genomic Selection in Plants. *J. Crop Sci. Biotech.* 18(5):293-308.

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

ZENG, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468.

ZHAO, Q.; DIXON, R.A. (2011). Transcriptional networks for lignina biosynthesis: more complex than we thought? *Trends in Plant Science* 16(4).

2. CAPÍTULO 1

MAPEAMENTO DE QTLs EM SORGO SACARINO PARA CARACTERES RELACIONADOS À PRODUÇÃO DE BIOENERGIA

RESUMO

O sorgo é uma importante cultura que tem ganhado destaque como potencial matéria-prima para produção de bioenergia. Além disso, o sorgo é um excelente modelo para estudos genéticos em gramíneas, por ser uma espécie diploide com genoma compacto e sequenciado. O objetivo do presente estudo foi mapear QTLs relacionados à produção de bioenergia em sorgo. Uma população de RILs de sorgo sacarino, constituída por 223 linhagens endogâmicas recombinantes, derivadas do cruzamento entre os genitores Brandes e Wray, ambos sacarinos, contrastantes para qualidade e quantidade de açúcares presentes no caldo do colmo, foi utilizada para o mapeamento. Os experimentos foram conduzidos em três safras, em Sete Lagoas, MG, Brasil. Os caracteres avaliados foram florescimento, altura de plantas, produção de massa verde (PMV), sólidos solúveis totais (Brix), sacarose (Pol), açúcares redutores (AR) e fibras. A população foi genotipada por sequenciamento e a detecção de QTLs foi realizada pelo método de mapeamento por múltiplos intervalos para múltiplos caracteres (*Multiple Trait Multiple Interval Mapping* – MTMIM) no pacote *OneQTL* do R. Foram identificados 65 QTLs pela abordagem MTMIM e todas as variáveis analisadas apresentaram efeitos epistáticos significativos entre QTLs. Novos QTLs, ainda não relatados, foram identificados no presente estudo, o que pode ser em consequência da alta densidade de marcadores utilizados, do método de mapeamento empregado ou pelo fato de ambos os genitores serem sacarinos. Estes resultados contribuem para o melhor entendimento da arquitetura genética de caracteres relacionados à produção de bioenergia, e, também, para uma futura aplicação na seleção assistida por marcadores moleculares para o melhoramento de sorgo sacarino.

Palavras-chave: *Sorghum bicolor*; linhagens endogâmicas recombinantes; bioetanol, *Genotyping-by-Sequencing*; *Quantitative Trait Loci.*

ABSTRACT

Sorghum is an important crop that has gained attention as a potential feedstock for bioenergy production. It is an excellent model for genetic studies in grasses, due to its compact sequenced genome and diploid status. The aim of this study was to map QTLs related to bioenergy production. A RIL population of sweet sorghum, consisting of 223 lines derived from a cross between the genitors Brandes and Wray, both sweet and contrasting regarding the quality and quantity of sugars present in the stem juice, was used for genetic mapping studies. The experiment was carried out in Sete Lagoas, Minas Gerais, Brazil, for three harvests. The evaluated traits were flowering, plant height, fresh mass production (FMP), total soluble solids (Brix), sucrose (Pol), reducing sugars (RS), and fiber. The population was genotyped by sequencing (GBS) and the detection of QTLs was performed by the method of Multiple Trait Multiple Interval Mapping (MTMIM), using the OneQTL package of the R software. Sixty five QTLs were identified by the MTMIM approach and all analysed variables had significant epistatic effects between QTLs. Novel QTLs were identified in this study, which may be due to the high density of markers used, the mapping method employed, or that both genitors are sweet. These findings contribute to a better understanding of the genetic architecture of traits related to bioenergy production, and to the future implementation of marker-assisted selection in sweet sorghum breeding programs.

Keywords: *Sorghum bicolor,* recombinant inbred lines, bioethanol, Genotyping-by-Sequencing; Quantitative Trait Loci.

2.1. INTRODUÇÃO

O sorgo (*Sorghum bicolor*) apresenta ampla diversidade genética, o que possibilita o desenvolvimento de cultivares para diferentes finalidades e produtos, como forragens, grãos e biocombustíveis. As cultivares de sorgo sacarino apresentam colmos suculentos ricos em açúcares, semelhantes a cana-de-açúcar. Além disso, é uma cultura versátil e adaptada a diversas condições de solo e clima, o que a torna uma promissora alternativa para a produção de energia ao redor do mundo (Zegada-Lizarazu & Monti 2012). No Brasil, a utilização do sorgo sacarino como matéria-prima complementar pode aumentar não só a eficiência da produção de etanol, com a redução de ociosidade das usinas durante a entressafra da cana-de-açúcar, como também a de bioeletricidade, a partir da queima da biomassa residual (Jonker *et al.* 2015).

Os caracteres mais importantes na seleção de genótipos de sorgo sacarino para a produção de bioenergia são basicamente a duração do ciclo de cultivo, o porte da planta, e a produção total de açúcares e fibras da biomassa (Regassa & Wortmann 2014). Tais caracteres são de natureza quantitativa, controlados por muitos genes, recebendo forte influência do ambiente. O desenvolvimento de tecnologias de genotipagem com base no sequenciamento do genoma vem permitindo maior precisão e detalhamento no mapeamento de locos que controlam a expressão de características quantitativas (do inglês, *Quantitative Trait Loci* - QTL), devido à maior disponibilidade de marcadores ao longo do genoma (Poland & Rife 2012).

A identificação de regiões que controlam a expressão de características fenotípicas, e a compreensão da relação entre polimorfismos nas sequências de DNA e a variabilidade observada nos fenótipos dos indivíduos, podem proporcionar aplicações relevantes para os programas de melhoramento genético do sorgo sacarino por meio de seleção assistida por marcadores moleculares (He *et al.* 2014). Além disso, por ser uma espécie diploide (2n=2x=20) e com genoma relativamente compacto, aproximadamente 726.616.606 pares de bases (Patterson *et al.* 2009), o sorgo tornou-se uma excelente plataforma para estudos genéticos, além de modelo para o estudo de culturas energéticas com genoma mais complexo, como a cana-de-açúcar (Mullet *et al.* 2014).

Vários estudos de mapeamento de QTLs relacionados à bioenergia em sorgo são descritos na literatura (Murray *et al.* 2008; Guan *et al*. 2010; Shiringani *et al*. 2011; Mace & Jordan 2011; Felderhoff *et al.* 2012). Entretanto, os estudos anteriores geralmente utilizaram populações de RILs geradas a partir de cruzamento entre um genitor sacarino e outro não sacarino, além de um baixo número de marcadores quando comparado ao obtido a partir da técnica de genotipagem por sequenciamento (do inglês, *Genotyping By Sequencing* - GBS).

Desta forma, o presente trabalho teve como objetivo estudar a arquitetura genética de caracteres agroindustriais em sorgo sacarino, considerando dados fenotípicos de três safras. Os dados foram obtidos de uma população de linhagens endogâmicas recombinantes (do inglês, *Recombinant Inbred Lines* - RIL) derivada de genitores sacarinos contrastantes para caracteres agroindustriais, como teor de sacarose, açúcares redutores e fibras. A população foi genotipada por sequenciamento e foi empregado o método de mapeamento por múltiplos intervalos para múltiplos caracteres (do inglês, *Multiple Trait Multiple Interval Mapping* - MTMIM), que permite o mapeamento para múltiplos ambientes com a incorporação das correlações genéticas entre safras (Silva *et al.* 2012).

2.2. MATERIAL E MÉTODOS

2.2.1. Material Vegetal, Delineamento e Descrição da Área Experimental

Foram avaliadas 223 RILs, obtidas a partir do cruzamento entre os genitores Brandes (BR501) e Wray (BR505) pelo método “descendente de uma única semente” (do inglês, *Single Seed Descent* - SSD) (Brim 1966). Ambos os genitores são sacarinos, com alto teor de açúcares totais, porém contrastantes quanto a quantidade e a qualidade dos açúcares. O genótipo Wray apresenta elevado teor de sacarose e baixo teor de açúcares redutores. Já o genótipo Brandes foi tradicionalmente utilizado para produção de xarope na América do Norte e apresenta baixo teor de sacarose, indesejável neste caso devido a cristalização dos açúcares. Entretanto, altos teores de sacarose podem ampliar o período de utilização industrial do sorgo sacarino,

devido a estabilidade dessa molécula, sendo uma característica desejável na produção de bioenergia.

As RILs foram avaliadas a partir da geração $F_{2:6}$ e os genitores foram utilizados como testemunhas no delineamento experimental, que foi em látice 15x15, com três repetições, totalizando 675 parcelas. As parcelas foram compostas por duas linhas de 5 m de comprimento, espaçadas por 0,70 m. Os experimentos foram conduzidos nas safras 2010/11, 2011/12 e 2013/14 em áreas experimentais pertencentes à Embrapa Milho e Sorgo, em Sete Lagoas, MG, Brasil.

As safras 2010/11 e 2011/12 foram semeadas nos dias 3 de fevereiro de 2011 e 13 de dezembro de 2012, respectivamente, na mesma área experimental (coordenadas geográficas -19.449760, -44.176479). A safra 2013/14 foi semeada em 17 de outubro de 2013 em área experimental diferente (coordenadas geográficas -19.473939, -44,174442). Os dados climáticos referentes às safras são apresentados na Figura 1. Os dados das análises de solo das duas áreas experimentais são apresentados na Tabela 1. As semeaduras foram realizadas utilizando o sistema de plantio direto e as safras receberam irrigação suplementar à precipitação local, por aspersão convencional, durante todo o ciclo. Na adubação de plantio foram aplicados 400 $Kg.ha^{-1}$ do formulado 8-28-16 (NPK), e 200 $Kg.ha^{-1}$ de ureia foram utilizados na adubação de cobertura 20 dias pós-semeadura.

2.2.2. Dados Fenotípicos

Os seguintes caracteres foram avaliados: época de florescimento (Floresc.), em dias após semeadura; altura média da parcela (Altura), em cm; produção de massa verde (PMV), em $t.ha^{-1}$; extração de caldo (Extração), em % da biomassa; sólidos solúveis totais (Brix), em ºBrix; teor de sacarose (Pol), em % do caldo; açúcares redutores (AR), em % do caldo; e fibras do colmo (Fibras), em % da biomassa.

A produção de biomassa foi calculada em quilos por parcela e transformada em tonelada por hectare. A extração de caldo foi realizada em prensa hidráulica com pressão mínima e constante de 250 $kgf.cm^{-2}$ durante 1 minuto, sobre uma amostra de 500 g de biomassa fresca desintegrada e homogeneizada. O teor de sólidos solúveis

totais foi aferido em refratômetro digital de leitura automática em ºBrix e o conteúdo de fibras foi estimado pelo método de Tanimoto (1964).

Na primeira safra, 2010/11, os teores de sacarose e açúcares redutores foram avaliados, respectivamente, em polarímetro, após clarificação do caldo com mistura à base de alumínio, e destilação com Fehling A e B, conforme método de Lane & Eynon (1934). Os resultados alimentaram uma curva de calibração em espectroscopia no infravermelho próximo (do inglês, *Near-Infrared Spectroscopy* - NIRS) e as safras subsequentes foram analisadas utilizando o modelo validado para Pol e AR via NIRS semelhante ao desenvolvido por Guimarães *et al.* (2014).

2.2.3. Análises Fenotípicas

As análises fenotípicas das múltiplas safras para cada caractere avaliado foram realizadas via modelos mistos e os componentes de variância foram estimados via máxima verossimilhança restrita (REML) considerando o seguinte modelo:

$$y_{ijkl} = \mu + s_l + r_{k(l)} + b_{j(kl)} + t_i + ts_{il} + \varepsilon_{ijkl}$$

em que: y_{ijkl} é o valor fenotípico observado para o indivíduo i no bloco j, repetição k e safra l; μ é a média geral; s_l é o efeito fixo da l-ésima safra (l = 1,..., L; L = 3); $r_{k(l)}$ é o efeito fixo da k-ésima repetição (k = 1, ..., K; K = 3) na safra l; $b_{j(kl)}$ é o efeito aleatório do j-ésimo bloco (j = 1,..., J; J = 15) na repetição k e na safra l; t_i é o efeito ora fixo ora aleatório do i-ésimo tratamento (i = 1, ..., I; I = 225); ts_{il} é o efeito ora fixo ora aleatório da interação do i-ésimo tratamento com a l-ésima safra; e ε_{ijkl} é o efeito aleatório residual. O termo t_i foi separado em dois grupos, sendo g_i o efeito aleatório dos genótipos da população (i = 1, ..., I_g; I_g = 223) e p_i o efeito fixo dos genitores da população (i = I_g + 1, ..., I_g + I_c; I_c = 2) que foram incluídos como testemunha no delineamento experimental. Da mesma forma, o termo ts_{il} foi separado em dois, sendo gs_{il} o efeito aleatório da interação dos genótipos da população (i = 1, ..., I_g; I_g = 223) com a l-ésima safra e ps_{il} o efeito fixo da interação dos genitores da população (i = I_g + 1, ..., I_g + I_c; I_c = 2) com a l-ésima safra.

Para as análises dos efeitos aleatórios do modelo, foi utilizado o teste da razão de verossimilhança (do inglês, *Likelihood Ratio Test* - LRT), realizado a partir da diferença entre as deviances para os modelos com e sem o efeito a ser testado, o qual

apresenta distribuição qui-quadrado com 1 grau de liberdade. Os efeitos fixos foram testados, utilizando a estatística de Wald, e mantidos no modelo quando significativos (*p*-valor < 0,05). O coeficiente de variação (CV) e a herdabilidade com base nas médias (h_m^2) foram estimados por:

$$CV = \frac{\sqrt{\sigma_e^2}}{\bar{x}} \times 100 \qquad h_m^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gs}^2}{l} + \frac{\sigma_e^2}{k.l}}$$

sendo: σ_g^2 , σ_{gs}^2 e σ_e^2, respectivamente, a variância genética, a variância da interação entre genótipos e safras, e a variância residual; $\bar{x}$ é a média geral; k é o número de repetições e l é o número de safras.

Para o ajuste das médias, ou seja, para estimar as médias dos genitores via BLUE (*Best Linear Unbiased Estimator*) e para predizer as médias dos genótipos via BLUP (*Best Linear Unbiased Predictions*) foi utilizado o seguinte modelo:

$$y_{ijkl} = \mu + s_l + r_{k(l)} + b_{j(kl)} + t_{il} + \varepsilon_{ijkl}$$

em que: y_{ijkl} é o valor fenotípico observado para o indivíduo *i* no bloco *j*, repetição *k* e safra *l*; μ é a média geral; s_l é o efeito fixo da *l*-ésima safra (l = 1,..., L; L = 3); $r_{k(l)}$ é o efeito fixo da *k*-ésima repetição (k = 1, ..., K; K = 3) na safra *l*; $b_{j(kl)}$ é o efeito aleatório do *j*-ésimo bloco (j = 1,..., J; J = 15) na repetição *k* e na safra *l*; t_{il} é o efeito ora fixo ora aleatório do *i*-ésimo tratamento (i = 1, ..., I; I = 225) na safra *l*; e ε_{ijkl} é o efeito aleatório residual. O termo t_{il} foi separado em dois grupos, sendo g_{il} o efeito aleatório dos genótipos da população (i = 1, ..., I_g; I_g = 223) e p_{il} o efeito fixo dos genitores da população ($i = I_g + 1$, ..., $I_g + I_c$; I_c = 2). Os vetores dos efeitos de blocos, residuais e genéticos apresentaram distribuição normal multivariada com média zero e matriz de variância-covariância (VCOV) $\mathbf{B}_m$, $\mathbf{R}_m$ e $\mathbf{G}_m \otimes \mathbf{I}_n$, respectivamente. Sendo que: $\otimes$ é o produto de Kronecker e $\mathbf{I}_n$ é uma matriz identidade com dimensão n x n.

Diferentes estruturas para as matrizes de variância e covariância (VCOV) dos efeitos genéticos (**G**), de blocos (**B**) e residuais (**R**) foram comparadas utilizando AIC (Akaike Information Criterion) (Akaike 1974) e BIC (Bayesian Information Criterion) (Schwarz 1978). Em resumo, foram comparadas estruturas de VCOV que assumiam ausência de correlação e homogeneidade (Identidade - ID) ou heterogeneidade de variâncias (Diagonal - DIAG), variâncias e covariâncias homogêneas (uniforme - UNIF), e heterogeneidade de variâncias e existência de correlações específicas (Não estruturada - UNST).

O ajuste foi realizado inicialmente para as matrizes (**G**), de acordo com os menores valores de AIC e BIC. A estrutura de VCOV selecionada foi fixada para os efeitos genéticos e os ajustes foram realizados para os efeitos residuais. Por fim, com as matrizes de VCOV fixadas para os efeitos genéticos e residuais, foram realizados os ajustes dos efeitos de blocos.

As análises de modelos mistos foram realizadas utilizando o software GenStat (v. 16.1) (VSN International 2014). Correlações genéticas entre as médias ajustadas dos genótipos para cada par dos caracteres foram calculadas pelo método de Pearson e testadas assumindo nível global de significância de 0,05 utilizando o pacote psych (Revelle 2014) disponível no software R (R Core Team 2014).

2.2.4. Mapeamento de QTLs

O procedimento de genotipagem por sequenciamento (do inglês, *Genotyping-by-Sequencing* - GBS) foi realizado pelo Institute for Genomic Diversity (IGD, Cornell University, Ithaca, NY, EUA) de acordo com o protocolo descrito em detalhes por Elshire et al. (2011) em plataforma HiSeqTM 2000 (Illumina, Inc.). Bibliotecas com controle negativo foram construídas a partir da digestão de DNA genômico dos indivíduos com a enzima ApeKI, que reconhece sítio de cinco bases (sendo uma degenerada) e é sensível à metilação.

O pipeline Tassel-GBS (Glaubitz *et al*. 2014) implementado no software Tassel (v. 4.3.8) foi utilizado para descoberta de SNPs e indels. Para tanto, tags de GBS de 64 pb foram alinhadas contra os dez cromossomos do genoma do sorgo (v. 2.1) (Paterson *et al*. 2009). As *tags* utilizadas no mapeamento estavam presentes, no mínimo, três vezes no conjunto de três bibliotecas.

Dados perdidos foram imputados utilizando o software Npute (Roberts *et al*. 2007). Este programa utiliza informações dos marcadores próximos aos dados perdidos em um espaço amostral, ou tamanho de janela, determinado por teste preliminar de acurácia da imputação com dados não perdidos. Foram testados tamanhos de janela variando de 5 a 150 SNPs sequenciais para cada cromossomo, as janelas que apresentaram maior acurácia foram selecionadas para imputação. Os dados imputados foram filtrados para uma frequência alélica mínima (do inglês, *minor*

allele frequency – MAF) de 40%, considerando uma possível distorção na segregação 1:1 da população RILs.

A busca por QTLs baseou-se em modelos de mapeamento por múltiplos intervalos (do inglês, *Multiple Interval Mapping* - MIM) (Kao *et al.* 1999) e MIM para múltiplos caracteres (MTMIM) (Silva *et al.* 2012), implementados no pacote desenvolvido para o software R denominando OneQTL (L. da Costa e Silva, comunicação pessoal). Foi utilizado o seguinte modelo linear de múltiplos QTLs para múltiplos ambientes (safras):

$$y_{ei} = \mu_e + \sum_{r=1}^{R} (a_{er} x_{air}) + \varepsilon_{ei}$$

em que: y_{ei} é a média ajustada para o indivíduo *i* (*i* = 1, 2,..., I_g) no ambiente *e* (*e* = 1,..., *E; E* = 3), se E = 1, o modelo de mapeamento torna-se univariado; μ_e é o intercepto para cada ambiente *e;* a_{er} é o efeito genético aditivo do QTL *r* (*r* = 1, 2,..., *R* QTLs) no ambiente *e;* $\varepsilon i \sim N(0,\Sigma_E)$ é o resíduo; x_{air} é uma variável para o genótipo do SNP *r* no indivíduo *i*, e assume valores -1 e 1 para os genótipos referentes aos genitores Brandes e Wray, respectivamente.

Devido à grande densidade de marcadores, não foi necessário estimar probabilidades condicionais para o genótipo do QTL dentro do intervalo entre marcadores. As médias ajustadas marginalmente para cada safra foram utilizadas no modelo MIM univariado (E = 1). Posteriormente, as médias marginais foram utilizadas no modelo MTMIM (E = 3), permitindo a detecção de possíveis QTLs com base na informação de múltiplos ambientes.

A construção do modelo de múltiplos QTLs baseou-se em busca *forward* de possíveis QTLs ao se testar a significância de seus efeitos principais em cada posição do genoma, que corresponderam aos próprios marcadores SNP. As posições dos QTLs no modelo foram refinadas a cada três ciclos de inclusão de QTLs. Finalmente, as permanências dos efeitos principais foram testadas via eliminação *backward*. Nos modelos multivariados, coeficientes de regressão aparentemente não relacionados (do inglês, *Seemingly Unrelated Regressions*) (Zellner 1962) resultantes da eliminação *backward* tiveram seus efeitos estimados e foram marcados como não significativos, ao invés de mantê-los com efeitos nulos. Foram utilizados os níveis de significância baseados na estatística *Score* (Zou & Zeng 2008) de 0,10 e 0,05 nas respectivas etapas de entrada (busca *forward*) e saída (eliminação *backward*) de QTLs em cada modelo. O método da máxima verossimilhança foi adotado para a

estimação dos efeitos genéticos e do componente de variância residual. Também foram obtidos os valores de LOD score associados a todas as posições genômicas avaliadas.

2.3. RESULTADOS

2.3.1. Análises Fenotípicas

As informações climáticas do decorrer do ciclo de cultivo de cada safra, como temperatura, pluviosidade e fotoperíodo, podem ser observadas na Figura 1. O fotoperíodo foi menor no decorrer do ciclo da primeira safra, com semeadura em fevereiro. Isso pode explicar a redução do estádio vegetativo, que ocorre até o florescimento. Na primeira safra, o florescimento ocorreu em média 73 dias após a semeadura. Nas demais safras, o florescimento ocorreu em média 90 e 99 dias após a semeadura.

Como o ciclo vegetativo é reduzido em função da época de semeadura, a altura e a produção de massa verde também seguem essa tendência. A terceira safra apresentou a maior amplitude de variação para florescimento e maiores valores de médias ajustadas para altura, PMV e extração (Figura 2). Além do fotoperíodo ter sido maior na terceira safra, a área experimental também apresentava maior saturação de bases, além de menor teor de alumínio e de acidez no solo, quando comparada à área experimental utilizada nas primeiras safras (Tabela 1). Assim, as diferenças entre as médias fenotípicas das safras podem ser atribuídas principalmente às diferentes épocas de semeadura, e no caso da terceira safra à diferente área experimental.

Quanto ao acúmulo de açúcares no colmo, no sorgo este processo ocorre após o florescimento, diferente do que ocorre com a cana-de-açúcar (Tarpley & Vietor 2007). Desta forma, os caracteres sólidos solúveis totais (Brix) e sacarose (Pol) não apresentaram variação entre safras tão visíveis como as observadas para altura, florescimento e PMV. No entanto, o efeito de safra foi significativo para todos os caracteres avaliados, assim como o efeito de genótipos e da interação entre genótipos e safras. Em geral, todos os efeitos foram significativos, exceto o efeito de

testemunhas para florescimento, ou seja, para este caractere não foi necessária a inclusão do efeito de testemunhas no modelo para a obtenção das médias ajustadas para as RILs.

Para a maioria dos caracteres, as variâncias da interação entre genótipos e safras foram de menor magnitude do que as variâncias genéticas. Entretanto, a variância residual apresentou menor magnitude quando comparada à variância genética apenas para florescimento e altura. Os coeficientes de variação residual permitem verificar a qualidade experimental, cujos valores podem ser considerados dentro do padrão, e as herdabilidades médias reforçam que a maior proporção da variabilidade total é de natureza genética (Tabela 2).

Para a comparação dos diferentes modelos para as matrizes de variância-covariância (VCOV), foram utilizados principalmente os valores de AIC. Para os caracteres florescimento, altura, PMV, extração e fibras, modelos não estruturados (UNST) foram selecionados para os efeitos genéticos. No entanto, para Brix e Pol, matrizes com variâncias homogêneas e mesma covariância (UNIF) foram selecionadas (Tabela 3).

As médias ajustadas das RILs em relação às médias dos genitores (Tabela 2) demostraram a ocorrência de segregação transgressiva, o que indica que esta população pode ser utilizada para a identificação de genótipos com grande potencial de utilização em programas de melhoramento. Dentre os caracteres avaliados, Brix e POL foram os mais contrastantes fenotipicamente entre os genitores, gerando, por consequência, ampla variabilidade de fenótipos na população de RILs, ou seja, a população estudada é bastante adequada para mapear QTLs para esses caracteres.

As correlações genéticas entre os caracteres avaliados, obtidos com base nas médias BLUP para as RILs, são apresentadas na Figura 3. Extração e AR não apresentaram correlações genéticas significativas com Florescimento e PMV, e o teor de Fibras não apresentou correlações significativas com Brix, Pol e AR. A inexistência de correlações negativas entre os teores de açúcares e o teor de fibras indica que podem ser conduzidas, simultaneamente, estratégias de melhoramento para a produção de etanol de primeira e de segunda geração, ou a cogeração de eletricidade.

2.3.2. Mapeamento de QTLs

Com relação aos dados genotípicos, inicialmente, foi obtido um total de 461.241 marcadores SNP, variando de 31.969 (cromossomo 8) a 71.878 (cromossomo 1). O tamanho das janelas de imputação do software Nputе variou entre 12 SNPs para o cromossomo 7 a 18 SNPs para o cromossomo 4, com acurácia média de 98,4%. O número final de marcadores passou a ser 100.291, após imputação e filtragem para MAF igual a 40%, variando de 4.086 (cromossomo 8) a 14.326 (cromossomo 2).

Utilizando a abordagem MTMIM, com base nas médias ajustadas obtidas para cada safra, foi possível identificar um total de 65 possíveis QTLs para os caracteres avaliados na população de mapeamento de sorgo sacarino. Nas Tabelas 5.1 e 5.2 é possível visualizar as posições físicas, as estimativas dos efeitos e o valor de LOD para cada possível QTL identificado. Nas Figuras 3 e 4, são apresentados os QTLs identificados via MIM, com base nas médias marginais de cada safra, e via MTMIM, utilizando-se as médias das três safras.

O genitor Wray apresentou a maior proporção de alelos favoráveis para os caracteres florescimento, altura, PMV, Brix e Pol, enquanto o Brandes apresentou para fibras, AR e extração. Entretanto, os dois genitores contribuíram com alelos favoráveis para quase todos os caracteres avaliados, exceto AR, cujos alelos favoráveis foram herdados exclusivamento do genitor Brandes (Tabelas 4.1 e 4.2). A combinação dos alelos favoráveis de ambos os genitores pode explicar a presença de linhagens com segregação transgressiva à média dos pais na população (Tabela 2).

Quanto à interação entre QTLs, ocorreram ao todo 22 efeitos epistáticos, dos quais 14 ocorreram na primeira safra, 9 na segunda e apenas 6 na terceira. Sendo que 6 foram comuns para a primeira e a segunda safra, e apenas 1 foi comum para a segunda e a terceira safra. Não ocorreu efeito epistático comum entre a primeira e a terceira safra. A maioria dos efeitos epistáticos foi positiva, ou seja, ocorreram mais incrementos do que reduções nos valores fenotípicos em decorrência das interações entre os locos. Entretanto, para Pol todos os efeitos epistáticos foram negativos na primeira safra, por exemplo (Tabela 5).

2.4. DISCUSSÃO

Conforme Paterson *et al.* (2009), o genoma do sorgo deve codificar aproximadamente 30.000 genes distribuídos nos seus dez cromossomos, tornando assim o estudo do controle genético dos caracteres de interesse uma tarefa extremamente difícil. A qualidade e a precisão do mapeamento de QTLs podem ser verificadas a partir dos resultados de estudos anteriores em que as posições dos genes próximos aos QTLs são conhecidas, como para florescimento, ou maturação, e para altura de plantas em sorgo (Murphy *et al*. 2011). O controle genético destas variáveis já está bem elucidado. A altura de plantas é controlada por quatro locos (*Dw1, Dw2, Dw3 e Dw4*) e o florescimento por seis locos, *Ma1* a *Ma4* relatados por Quinby (1974) e *Ma5* e *Ma6* por Rooney & Aydin (1999).

Murphy *et al.* (2011) atribuíram ao *locus Ma1* uma proteína reguladora (SbPRR37) situada entre 40,27 e 40,28 Mpb no cromossomo 6. No presente estudo, foi identificado um QTL para florescimento, extremamente significativo, com LOD igual a 56,45, na posição 40.740.202 pb. Relativamente próximo à região relatada quando comparado, por exemplo, ao trabalho de Higgins *et al.* (2014), que identificaram um SNP significativo cerca de 2 Mpb de distância do *Ma1*, na posição 42,07 Mpb, utilizando modelos de análise GWAS.

Trabalhos iniciais com sorgo sugeriam que o *Ma1* estaria ligado ao *locus* de nanismo *Dw2* (Quinby 1974), porém diversos estudos independentes estimam a posição do *Dw2* a aproximadamente 43 Mpb no cromossomo 6 (Klein *et al*. 2008; Morris *et al*. 2013; Thurber *et al*. 2013). No presente trabalho, foi identificado um QTL para altura na posição 42.204.434 pb no cromossomo 6, enquanto que, Higgins *et al*. (2014) sugeriram localização para *Dw2* entre 44,30 e 44,45 Mpb no mesmo cromossomo.

Não foram localizados QTLs para altura no cromossomo 7, na região do *Dw3*, em contraposição ao trabalho de Zou *et al*. (2012), que utilizou uma população RILs oriunda do cruzamento entre sorgo granífero (insensível ao fotoperíodo) x sacarino (sensível ao fotoperíodo). A ausência deste QTL pode ser atribuída ao fato de ambos os genitores serem insensíveis ao fotoperíodo, o que pode ter gerado uma população de mapeamento não polimórfica para esta região.

Upadhyaya *et al*. (2013) realizaram o mapeamento associativo de uma minicoleção nuclear de sorgo em ambientes tropicais utilizando 14.739 SNPs e identificaram marcadores significativos para altura e maturação próximos a QTLs previamente mapeados em sorgo. Quatro SNPs, localizados entre 554.233 e 554.279

pb no cromossomo 6, foram associados a maturação, com p-valores de mesma magnitude, e quatro dos cinco genes próximos a este local são de transportadores, sendo o mais próximo um transportador de açúcar (Sb06g000520). No presente trabalho, foi identificado um QTL na posição 575.104 pb, aproximadamente 20 kb desse gene, relacionado ao florescimento (Floresc.3).

Ainda no cromossomo 6, os autores identificaram um SNP na posição 44.980.895 pb, que está a 28,45 kb de um gene de resposta ao fotoperíodo (Sb06g016300). No presente trabalho, foi detectado um QTL na posição 40.740.202 pb. Zhang *et al.* (2015) relataram que a região pericentromérica do cromossomo 6 tem mostrado repetidamente evidências do controle genético da altura e da maturação, e a baixa recombinação nesta região permite intervalos de confiança de QTLs que atravessam o centrômero e cobrem amplas áreas no genoma do sorgo, como pode ser observado na Figura 4.

Alguns QTLs foram colocalizados entre os diferentes caracteres utilizados para o mapeamento, ou em regiões extremamente próximas. No cromossomo 1, entre 6.397.372 e 6.917.852 pb, foram mapeados QTLs para florescimento, extração de caldo, Brix e Pol. Estes QTLs podem estar associados a curva de maturação dos materiais, pois o acúmulo de açúcares está fisiologicamente relacionado à época de florescimento (Fernandes *et al.* 2014).

Para altura e PMV, dois QTLs foram colocalizados entre 42.132.655 e 42.204.434 pb no cromossomo 6, e 12.890.638 e 13.610.290 pb no cromossomo 10. Este resultado é justificável uma vez que o PMV está diretamente relacionado à altura das plantas. Além disso, altura de plantas também apresentou um QTL colocalizado com Fibras, Brix e Pol, no cromossomo 9 entre 10.110.884 e 10.163.280 pb, o que aponta uma base genética comum entre os carboidratos estruturais e os açúcares de reserva energética.

Da mesma forma, no cromossomo 3, entre 66.431.004 e 68.294.572, foram identificados QTLs para altura, fibras, Brix e Pol, porém apresentaram LOD mais elevado para Brix e Pol. Murray *et al.* (2008a) identificaram um QTL no cromossomo 3 para porcentagem de celulose estrutural no colmo, e sugeriram que o controle para teor de fibras é pleiotrópico com a concentração de açúcares no colmo.

No cromossomo 6, em posições entre 46.121.291 e 46.635.675 pb, foram identificados QTLs para extração, fibras, Brix e Pol. Além de dois QTLs específicos para Brix e Pol, em posições entre 10.940.929 e 12.196.799 pb, e entre 52.508.130 e

53.554.118 pb, também no cromossomo 6. Shiringani *et al.* (2010), utilizando população de RILs com 188 linhagens de um cruzamento entre granífero (M71) e sacarino (SS79), genotipada com 157 marcadores AFLP, SSR, e EST-SSR e utilizando mapeamento por intervalo composto (CIM), identificaram QTLs no cromossomo 6 que mostraram múltiplos efeitos para florescimento, altura da planta, Brix e Pol. Além disso, Burks *et al*. (2015) identificaram um SNP significativo a 51.801.476 pb no cromossomo 6 para teor de açúcares, rendimento de açúcares por área, volume de caldo, além de cor de nervura central e umidade do colmo, em mapeamento associativo de um painel de diversidade com 252 genótipos de sorgo.

No presente trabalho foram identificados QTLs para Brix e Pol em 8 cromossomos (exceto cromossomos 7 e 10). Murray *et al*. (2008a) em estudo com 176 linhagens endogâmicas recombinantes $F_{4:5}$, oriundas do cruzamento entre sorgo sacarino (Rio) e granífero (BTx623), também identificaram QTLs para Brix e açúcares totais em sorgo nos cromossomos 3 e 6. Natoli *et al*. (2002) identificaram QTLs de efeito maior para Brix no cromossomo 3 e Ritter *et al*. (2008) localizaram QTLs próximo ao telômero do braço longo do cromossomo 9. Os dois últimos trabalhos também identificaram QTLs para rendimento de açúcares no cromossomo 5.

Ritter *et al*. (2008) utilizaram 228 marcadores SSR e AFLP para genotipar 184 RILs $F_{2:6}$ obtidas a partir do cruzamento entre o genitor Rio e uma linhagem R elite do tipo granífero, e identificaram QTLs nos cromossomos 1, 3, 5 e 6 para teor de açúcares. Guan *et al*. (2011) também identificaram QTLs para Brix nos cromossomos 1, 2 e 3, utilizando 636 marcadores SSR, em 186 plantas F_2 e 186 $F_{2:3}$ também derivadas de cruzamento entre sorgo granífero e sacarino.

No presente trabalho, a maioria dos QTLs identificados para Pol foram colocalizados com os de Brix, exceto três QTLs no cromossomo 4. Este resultado é condizente, uma vez que a sacarose presente no caldo é também aferida como um sólido solúvel. Porém, além da sacarose, o teor de outros componentes orgânicos e inorgânicos presentes no caldo, como ácidos orgânicos e minerais, também é estimado pelo refratômetro, o que talvez esteja relacionado aos demais QTLs mapeados para Brix.

Apesar de já existirem relatos de QTLs nos mesmos cromossomos, é difícil determinar se os locos identificados para características quantitativas de outros estudos estão colocalizados com os detectados no presente trabalho, ou se estão associados aos mesmos genes, devido à falta de relação entre o mapa físico e os

mapas genéticos desses estudos. No entanto, Mace & Jordan (2011) realizaram uma meta-análise abrangente utilizando 48 estudos de QTLs de sorgo publicados entre 1995 e 2010, e ao todo, 771 QTL relativos a 161 caracteres foram projetados em um mapa consenso de sorgo. Além disso, os autores também identificaram cinco regiões principais com alta densidade de QTLs, que apresentaram mais de 20 QTLs por 0,5 cM, localizadas nos cromossomos 1 (~70 cM), 6 (~84 cM), 10 (~58 cM), e duas no cromossomo 7 (~108 cM e ~125 cM). No cromossomo 6, em uma região de apenas 5 cM, onde estão localizados *Dw2* e *Ma1*, são descritos 36 QTLs, incluindo oito QTLs para altura, três para maturidade, dois relacionados a biomassa do colmo e cinco relacionados a açúcares.

Os mesmos autores descreveram nove regiões com alta densidade de genes, que podem apresentar mais de 80 genes por 0,5 Mbp, identificadas em 5 cromossomos. Entre essas estão o cromossomo 1, entre 1,5 e 2,0 Mbp com 99 genes; cromossomo 3, entre 61,5 e 62,0 Mbp com 82 genes; e o cromossomo 6, entre 56,0 e 56,5 Mbp com 84 genes, entre 58,0 e 58,5 Mbp com 88 genes, e entre 60,0 e 60,5 Mbp com 82 genes. Assim, é possível observar que a maior parte dos QTLs mapeados no presente estudo está próxima às regiões então relatadas. Este fato também pode explicar o maior intervalo de confiança observado para os QTLs identificados nestes cromossomos.

Para identificar genes candidatos para Brix em QTLs de grande efeito, Murray *et al*. (2009) desenharam marcadores baseados em sequências de BLAST de mais de 100 enzimas metabólicas de amido e sacarose e transportadores de açúcar de diferentes espécies (Kanehisa *et al*. 2006). A única associação significativa para Brix ocorreu no cromossomo 1, aproximadamente 12 kb de distância de um homólogo de glicose-6-fosfato-isomerase (SB00166.1). Lv et al. (2013) também utilizaram a estratégia de mapeamento associativo em 125 variedades de sorgo sacarino, porém, só detectaram associação entre um marcador (xtxp340), também no cromossomo 1, com Brix em apenas uma das safras. A posição física deste marcador foi obtida no banco de dados CSGRqtl (Zhang *et al*. 2013), aproximadamente 69,74 Mpb, e corresponde a um transportador de açúcar UDP.

No presente estudo, foi identificado um QTL para Brix e Pol no início do cromossomo 1, 6.554.131 pb, entre dois genes UDP-Glicosil Transferase (Sb01g007620 e Sb01g007630). É importante destacar que UDP-Glicose está diretamente envolvida com a síntese de sacarose, e combina-se com frutose-6-P para

gerar sacarose-P por ação da enzima sacarose fosfato sintase (SPS). Posteriormente sacarose-P é convertida em sacarose pela enzima sacarose fosfato fosfatase (SPP). A sacarose também pode ser convertida em frutose e UDP-Glicose para a respiração celular e biossíntese de biopolímeros, como amido e componentes da parede celular (Wang *et al*. 2013).

Para AR, dois QTLs também foram mapeados próximos a genes relacionados a sua função fisiológica. O primeiro, no cromossomo 1, 44.337.325 Mpb, está próximo a um gene da família Glicosil hidrolase (Sb01g026390), que se encontra entre 44.323.674 e 44.334.289 Mpb. Enzimas desta família atuam na hidrólise de açúcares complexos, dentre outras funções (Davies & Henrissat 1995). O segundo QTL, no cromossomo 4, 434.967 Mpb, está a aproximadamente 4 Kpb de uma proteína similar a uma invertase vacuolar (Sb04g000615). Estes resultados são condizentes, uma vez que a variável AR representa especificamente os teores de glicose e frutose oriundos da inversão da sacarose.

Como pode ser observado na literatura, as populações de mapeamento utilizadas anteriormente para sorgo sacarino, geralmente, são oriundas do cruzamento entre linhagens sacarinas e não sacarinas (sorgo granífero, em geral), e utilizaram mapas de ligação de baixa densidade. Além disso, utilizaram técnicas de mapeamento com menor poder de detecção de QTLs. Desta forma, a partir do presente estudo, foi possível validar QTLs já identificados em outros trabalhos, e, também, identificar novos QTLs, e possíveis interações epistáticas. Os resultados são apresentados para regiões do genoma com posições físicas detalhadas, o que permite que novos estudos mais aprofundados sejam realizados para o entendimento do controle genético dessas variáveis de interesse agroindustrial.

2.5. REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6):716-723.

BRIM, C.A. (1966). A modified pedigree method of selection in soybeans. *Crop Science* 6:220.

BURKS, P.S.; KAISER, C.M.; HAWKINS, E.M.; BROWN P.J. (2015). Genomewide Association for Sugar Yield in Sweet Sorghum. *Crop Science* 55.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKLER, E.S.; MITCHELL, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):1-9.

FELDERHOFF, T.J.; MURRAY, S.C.; KLEIN, P.E.; SHARMA, A.; HAMBLIN, M.T.; KRESOVICH, S.; VERMERRIS, W.; ROONEY, W. L. (2012). QTLs for Energy-related Traits in a Sweet × Grain Sorghum [*Sorghum bicolor* (L.) Moench] Mapping Population. *Crop Science* 52:2040–2049.

FERNANDES, G.; BRAGA, T.G.; FISCHER, J.; PARRELLA, R.A.C.; RESENDE, M.M.; CARDOSO, V.L. (2014). Evaluation of potential ethanol production and nutrients for four varieties of sweet sorghum during maturation. *Renewable Energy* 71:518-524.GLAUBITZ, J.C.; CASSTEVENS, T.M.; LU, F.; HARRIMAN, J.; ELSHIRE, R.J.; SUN, Q.; BUCKLER, E.S. (2014). Tassel-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9(2):1-11.

GUAN, Y.A.; WANG, H.L.; QIN, L.; ZHANG, H.W.; YANG, Y.B.; GAO, F.J.; LI, R.Y.; WANG, H.G. (2011). QTL mapping of bio-energy related traits in Sorghum. *Euphytica* 182(3):431–440.

GUIMARÃES, C.C.; SIMEONE, M.L.F.; PARRELLA, R.A.C.; SENA, M.M. (2014). Use of NIRS to predict composition and bioethanol yield from cell wall structural components of sweet sorghum biomass. *Microchemical Journal* 117:194–201.

HALEY, C.S.; KNOTT, S.A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69:315–324.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z.; LIU, H.; LI, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* 5:484.

HIGGINS, R.H.; THURBER, C.S.; ASSARANURAK, I.; BROWN, P.J. (2014). Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families. *G3*, 4(9):1593-602.

JONKER, J.G.G.; HILST, F.V.D.; JUNGINGER, H.M.; CAVALETT, O.; CHAGAS, M.F.; FAAIJ, A.P.C. (2015). Outlook for ethanol production costs in Brazil up to 2030, for different biomass crops and industrial technologies. *Applied Energy* 147:593-610.

KANEHISA, M.; GOTO, S.; HATTORI, M.; AOKI-KINOSHITA, K. F.; ITOH, M., KAWASHIMA, S.; KATAYAMA, T.; ARAKI, M.; HIRAKAWA, M. (2006). From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res* 34:D354–D357.

KAO, C.H.; ZENG, Z.B.; TEASDALE, R.D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* 152(3):1203-1216.

KLEIN, R.R.; MULLET, J.E.; JORDAN, D.R.; MILLER, F.R.; ROONEYET, W.L. (2008). The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. *Crop Sci.* 48:S12–S26.

LANE, J.H.; EYNON, L. (1934). Determination of reducing sugars by Fehling solution with methylene blue indicator. *Norman Rodge*.

LV, P.; JI, G.; HAN, Y.; HOU, S.; LI, S.; MA, X.; DU, R.; LIU, G. (2013). Association analysis of sugar yield-related traits in sorghum [*Sorghum bicolor* (L.)]. *Euphytica* 193:419–431.

MACE, E.S.; JORDAN, D.R. (2011). Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. *Theor Appl Genet* 123:169-191.

MORRIS, G.P.; RAMU, P.; DESHPANDE, S.P.; HASH, C.T.; SHAHET, T. (2012). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci*. 110:453–458.

MULLET, J.; MORISHIGE, D.; MCCORMICK, R.; TRUONG, S.; HILLEY, J.; MCKINLEY, B.; ANDERSON, R.; OLSON, S. N.; ROONEY, W. (2014). Energy Sorghum - a genetic model for the design of C4 grass bioenergy crops. *Journal of Experimental Botany* 65(13):3479–3489.

MURPHY, R.L.; KLEIN, R.R.; MORISHIGE, D.T.; BRADY, J. A.; ROONEY, W.L.; MILLER, F.R.; DUGAS, D.V.; KLEIN, P.E.; MULLET, J.E. (2011). Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodicflowering in sorghum. *PNAS* 108(39):16469-16474.

MURRAY, S.C.; ROONEY, W.L.; HAMBLIN, M.T.; MITCHELL, S.E.; KRESOVICH, S. (2009). Sweet sorghum genetic diversity and association mapping for brix and height. *The Plant Genome* 2(1):48.

MURRAY S.C.; ROONEY, W.L.; MITCHELL, S.E.; SHARMA, A.; KLEIN, P.E.; MULLET, J.E.; KRESOVICH, S. (2008a). Genetic Improvement of Sorghum as a Biofuel Feedstock: I. QTL for Stem Sugar and Grain Nonstructural Carbohydrates. *Crop Science* 48.

MURRAY S.C.; ROONEY, W.L.; MITCHELL, S.E.; SHARMA, A.; KLEIN, P.E.; MULLET, J.E.; KRESOVICH, S. (2008b) Genetic Improvement of Sorghum as a

Biofuel Feedstock: II. QTL for Stem and Leaf Structural Carbohydrates. *Crop Science* 48.

NATOLI, A.; GORNI, C.; CHEGDANI, F.; MARSAN, P.A.; COLOMBI, C.; LORENZONI, C. (2002). Identification of QTLs associated with sweet sorghum quality. *Maydica* 47(3/4):311–22.

PATERSON, A.H.; BOWERS, J.E.; BRUGGMANN, R.; DUBCHAK, I.; GRIMWOOD, J.; GUNDLACH, H.; HABERER, G.; HELLSTEN, U.; MITROS, T.; POLIAKOV, A.; SCHMUTZ, J.; SPANNAGL, M.; TANG, H.; WANG, X.; WICKER, T.; BHARTI, A.K.; CHAPMAN, J.; FELTUS, F.A.; GOWIK, U.; GRIGORIEV, I.V.; LYONS, E.; MAHER, C. A.; MARTIS, M.; NARECHANIA, A.; OTILLAR, R.P.; PENNING, B.W.; SALAMOV, A.A.; WANG, Y.; ZHANG, L.; CARPITA, N.C.; FREELING, M.; GINGLE, A.R.; HASH, C.T.; KELLER, B.; KLEIN, P.; KRESOVICH, S.; MCCANN, M.C.; MING, R.; PETERSON, D.G.; RAHMAN, M.; WARE, D.; WESTHOFF, P.; MAYER, K.F.X.; MESSING, J.; ROKHSAR, D.S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551–556.

POLAND, J.A.; RIFE, T.W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome* 5:3.

QUINBY, J.R. (1974). Sorghum improvement and the genetics of growth, College Station: Texas Agricultural Experiment Station.

R CORE TEAM. (2014). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

REGASSA, T.H.; WORTMANN, C.S. (2014). Sweet sorghum as a bioenergy crop: Literature Review. *Biomass and Bioenergy* 64:348-355.

REVELLE, W. (2014) psych: procedures for psychological, psychometric, and personality research. Evanston. Disponível em: <http://cran.r-project.org/package=psych>.

RITTER, K.B.; JORDAN, D.R.; CHAPMAN, S.C.; GODWIN, I.D.; MACE, E.S.; MCINTYRE, C.L. (2008). Identification of QTL for sugar-related traits in a sweet x grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. *Mol Breeding* 22:367–384.

ROBERTS, A.; MCMILLAN, L.; WANG, W.; PARKER, J.; RUSYN, I.; THREADGILL, D. (2007). Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics* 23(13):i401–i407.

ROONEY, W.L.; AYDIN, S. (1999). Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. *Crop Sci.* 39:397–400.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6:461-464.

SHIRINGANI, A. L.; FRISCH, M.; FRIEDT, W. (2010). Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench. *Theoretical and Applied Genetics* 121(2):323–336.

SILVA, L.D.C.E.; WANG, S.; ZENG, Z.B. (2012). Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genetics* 13(67):1–24.

SILVA, G.A.P.; GEZAN, S.A.; CARVALHO, M.P.; GOUVÊA, L.R.L.; VERARDI, C.K.; OLIVEIRA, A.L.B.; GONÇALVES, P.S. (2014). Genetic parameters in a rubber tree population: heritabilities, genotype-by-environment interactions and multi-trait correlations. Tree Genetics & Genomes 10(6):1511:1518.

TANIMOTO, T. (1964). The press method of cane analysis. *Hawaiian Planter's Record* 57(2):133-150.

TARPLEY, L.; VIETOR, D.M. (2007). Compartmentation of sucrose during radial transfer in mature sorghum culm. *BMC Plant Biology* 7:33.

THURBER, C.S.; MA, J.M.; HIGGINS, R.H.; BROWN, P.J. (2013). Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biol.* 14: R68.

UPADHYAYA, H.D.; WANG, Y.H.; GOWDA, C.L.L.; SHARMA, S. (2013). Association mapping of maturity and plant height using SNP markers with the *sorghum* mini core collection. *Theor Appl Genet* 126:2003–2015.

VSN INTERNATIONAL. (2014). GenStat for Windows 16th Edition. Hemel Hempstead: VSN International.

WANG, J.; NAYAK, S.; KOCH, K.; MING, R. (2013). Carbon partitioning in sugarcane (*Saccharum species*). *Front. Plant Sci.* 4.

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

ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57(298):348–368.

ZHANG, D.; GUO, H.; KIM, C.; LEE, T.H. (2013). CSGRqtl, a comparative quantitative trait locus database for *Saccharinae* grasses. *Plant physiology* 161:594–9.

ZHANG, D.; KONG, W.; ROBERTSON, J.; GOFF, V.H.; EPPS, E.; KERR, A.; MILLS, G.; CROMWELL, J.; LUGIN, Y.; PHILLIPS, C.; PATERSON, A.H. (2015). Genetic analysis of inflorescence and plant height components in sorghum (*Panicoidae*) and comparative genetics with rice (*Oryzoidae*). *BMC Plant Biology*, 15:107.

ZOU, W.; ZENG, Z.-B. (2008). Statistical methods for mapping multiple QTL. *International Journal of Plant Genomics* 2008:1–8.

ZOU, G.; ZHAI, G.; FENG, Q.; YAN, S.; WANG, A.; ZHAO, Q.; SHAO, J.; ZHANG, Z.; ZOU, J.; HAN, B.; TAO, Y. (2012). Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *Journal of Experimental Botany* 63(15):5451–5462.

TABELAS

Tabela 1 - Análise de solo das áreas experimentais, localizadas na Embrapa Milho e Sorgo, Sete Lagoas, MG, Brasil. Sendo 1 a área experimental utilizada para a avaliação da população de RILs de sorgo sacarino na primeira e segunda safra, e 2 para a terceira safra.

Área	Profundidade (cm)	pH (H_20)	H+Al ($cmol_c.dm^{-3}$)	M.O. ($g.kg^{-1}$)	C (Total)	SB ($cmol_c.dm^{-3}$)	CTC ($cmol_c.dm^{-3}$)	V (%)	Sat Al (%)
1	0-20	4,80	10,99	4,32	2,51	2,34	12,62	18,55	32,36
1	20-40	4,90	8,55	3,84	2,23	1,70	10,25	16,55	32,06
2	0-20	6,00	7,39	5,28	3,07	10,58	17,97	58,88	0,38
2	20-40	6,10	7,36	4,63	2,69	10,30	17,66	58,32	0,48

H+Al: acidez potencial; M.O.: matéria orgânica; C: carbono orgânico; SB: soma de bases; CTC: capacidade de troca catiônica; V: saturação por bases; Sat Al: saturação por alumínio.

Tabela 2 – Análise dos efeitos fixos do modelo via Teste de Wald e dos efeitos aleatórios via teste da razão de verossimilhança (LRT), estimativas dos componentes de variância σ^2, coeficientes de variação (CV) e herdabilidades com base nas médias (h^2_m) para caracteres agroindustriais avaliados ao longo de três safras em uma população de RILs de sorgo sacarino, oriunda do cruzamento entre as cultivares Brandes e Wray. Médias ajustadas via BLUE para os genitores e via BLUP, com os valores mínimos (Mín), médios (Méd) e máximos (Máx), para os indivíduos da população de RILs.

Caractere	Efeitos Fixos			Efeitos Aleatórios			Componentes de Variância						Médias Ajustadas		Médias Ajustadas (RILs)		
	s_l	$r_{k(l)}$	p_{il}	g_{il}	gs_{il}	$b_{j(kl)}$	σ^2_b	σ^2_g	σ^2_{gs}	σ^2_e	CV	h^2_m	Brandes	Wray	Mín	Méd	Máx
Floresc. (DAS)	**	**	NS	**	**	*	0,64	27,37	16,23	24,76	5,67	0,77	88	88	75	88	99
Altura (cm)	**	**	**	**	**	**	0,00	0,04	0,02	0,03	6,33	0,81	247	289	206	268	324
PMV ($t.ha^{-1}$)	**	**	**	**	**	**	12,80	33,30	32,20	108,40	26,58	0,59	32,79	45,55	23,17	39,17	51,94
Extração (%)	**	**	**	**	**	**	1,60	2,44	3,07	8,50	4,76	0,55	61,80	61,05	54,64	61,42	64,99
Brix (%)	**	**	**	**	**	**	0,40	1,77	1,03	2,11	12,04	0,75	8,94	15,20	8,08	12,07	15,50
Pol (%)	**	**	**	**	**	**	0,36	1,72	0,90	2,11	20,26	0,76	3,94	10,41	2,84	7,17	10,81
AR (%)	**	**	**	**	**	**	0,01	0,02	0,02	0,12	17,89	0,52	2,27	1,57	1,57	1,92	2,49
Fibras (%)	**	**	*	**	**	**	0,39	0,53	0,63	1,54	9,78	0,58	13,01	12,41	10,67	12,71	15,50

** e *: significativo a 1% e 5%, respectivamente. NS: não significativo. s_l: efeito da *l*-ésima safra; $r_{k(l)}$: efeito da *k*-ésima repetição na safra *l*; p_{il}: efeito dos genitores da população, incluídos como testemunha no delineamento experimental; g_{il}: efeito dos genótipos da população; gs_{il}: efeito da interação entre genótipos e safras; e $b_{j(kl)}$: efeito do *j*-ésimo bloco na repetição *k* e na safra *l*.

Tabela 3 - Descrição dos modelos examinados para as matrizes de variância-covariância genética (**G**), residual (**R**), de blocos (**B**) para a análise fenotípica conduzida com os dados das três safras de avaliação da população de RILs. As estruturas testadas foram: identidade (ID), diagonal (DIAG), uniforme (UNIF), e não estruturada (UNST). Os valores em negrito representam as estruturas selecionadas conforme o menor valor de AIC e BIC para cada caractere.

Caractere	Estrutura	Matriz **G**		Matriz **R**		Matriz **B**	
		AIC	BIC	AIC	BIC	AIC	BIC
Floresc.	ID	13397	13414	12903	12948	**12068**	**12124**
	DIAG	13172	13200	**12068**	**12124**	12070	12137
	UNIF	13224	13246	12903	12954	12070	12131
	UNST	**12903**	**12948**	12068	12140	12075	12159
Altura	ID	18444	18461	18217	18260	**18170**	**18204**
	DIAG	18434	18462	**18157**	**18204**	18170	18215
	UNIF	18238	**18260**	18218	18267	18171	18210
	UNST	**18217**	18262	18162	18226	18175	18236
PMV	ID	15780	15797	15624	15669	**15442**	**15498**
	DIAG	15698	15726	**15442**	**15498**	15446	15513
	UNIF	15723	15746	15623	15673	15444	15506
	UNST	**15624**	**15669**	15444	15516	15452	15536
Extra	ID	10813	10830	10728	10773	10666	**10722**
	DIAG	10778	10806	**10666**	**10722**	**10664**	10731
	UNIF	10772	10794	10729	10779	10668	10730
	UNST	**10728**	**10773**	10669	10741	10669	10753
Brix	ID	8396	8413	8247	**8269**	**8242**	**8269**
	DIAG	8397	8425	**8242**	8276	8244	8281
	UNIF	**8247**	**8269**	8248	8276	8244	8276
	UNST	8248	8292	8247	8298	8248	8302
POL	ID	8345	8361	8187	**8209**	**8187**	**8209**
	DIAG	8347	8375	**8187**	8220	8190	8224
	UNIF	**8187**	**8209**	8189	8217	8188	8217
	UNST	8192	8237	8193	8243	8195	8246
AR	ID	1998	2015	**1866**	**1911**	1866	**1911**
	DIAG	1910	1938	NC	NC	**1857**	1913
	UNIF	1960	1982	1867	1917	1867	1917
	UNST	**1866**	**1911**	NC	NC	1858	1931
Fibra	ID	7518	7535	7338	7383	7139	**7198**
	DIAG	7414	7442	7142	**7198**	**7134**	7204
	UNIF	7462	7484	7340	7391	7141	7205
	UNST	**7338**	**7383**	**7139**	7211	7138	7225

NC: não convergiu

Tabela 4.1 - Estimativas dos efeitos dos QTLs em cada safra e seus respectivos erros padrão (EP), cromossomos (Cr.), posições físicas (em megabases - Mb) e *LOD Score* (LOD) resultantes da análise MTMIM realizada com base nas médias marginais dos três anos de avaliação de uma população de RILs de sorgo sacarino. Os efeitos dos QTLs encontram-se na mesma escala do caractere avaliado, sendo: Florescimento em DAS; Altura em cm; PMV em $t.ha^{-1}$ e Fibras em %. Efeitos positivos representam alelos herdados do genitor Wray e efeitos negativos representam alelos do genitor Brandes.

QTL	Cr	Posição (Mb)	LOD	Efeito					
				Safra1	(EP)	Safra2	(EP)	Safra3	(EP)
Floresc1	1	6,92	7,28	0,75	(0,14)	1,30	(0,28)	1,38	(0,30)
Floresc2	1	66,08	11,88	-0,93	(0,13)	-1,34	(0,27)	-0,37	(0,29)
Floresc3	6	0,58	16,97	0,61	(0,13)	1,86	(0,28)	1,96	(0,30)
Floresc4	6	40,74	56,45	-0,94	(0,14)	-4,60	(0,29)	-5,28	(0,31)
Floresc5	9	2,86	10,05	0,86	(0,13)	1,84	(0,27)	1,24	(0,30)
Floresc6	10	39,91	14,45	0,42	(0,13)	1,73	(0,28)	1,39	(0,30)
Altura1	1	69,11	12,39	6,43	(1,05)	7,29	(1,24)	2,79	(1,10)
Altura2	3	68,29	9,36	3,41	(1,08)	4,47	(1,28)	7,28	(1,12)
Altura3	6	42,20	28,93	-1,96	(1,08)	-10,15	(1,28)	-6,27	(1,12)
Altura4	9	10,12	13,13	7,83	(1,06)	7,32	(1,25)	7,34	(1,11)
Altura5	10	13,61	16,90	5,55	(1,07)	9,84	(1,27)	3,90	(1,10)
PMV1	3	21,33	9,94	0,37	(0,27)	1,30	(0,42)	2,79	(0,41)
PMV2	6	42,13	18,34	-0,43	(0,27)	-2,76	(0,41)	-3,32	(0,41)
PMV3	7	59,96	7,78	0,83	(0,27)	1,46	(0,42)	2,40	(0,41)
PMV4	10	12,89	11,10	1,41	(0,27)	2,48	(0,41)	2,39	(0,40)
Fibras1	1	58,99	7,36	-0,08	(0,07)	-0,13	(0,03)	-0,19	(0,03)
Fibras2	3	66,43	12,63	-0,33	(0,07)	-0,21	(0,03)	-0,27	(0,03)
Fibras3	4	64,60	7,85	0,37	(0,07)	0,16	(0,03)	0,10	(0,03)
Fibras4	6	34,85	11,19	0,02	(0,02)	0,07	(0,03)	-0,11	(0,04)
Fibras5	6	46,12	8,85	0,24	(0,07)	-0,01	(0,03)	-0,11	(0,04)
Fibras6	8	37,46	14,31	-0,29	(0,07)	-0,21	(0,03)	-0,25	(0,03)
Fibras7	9	10,11	9,13	-0,16	(0,07)	0,03	(0,03)	0,16	(0,03)
Fibras8	10	2,46	10,42	0,20	(0,07)	0,13	(0,03)	0,22	(0,04)
Fibras9	10	56,60	8,18	-0,36	(0,07)	-0,16	(0,03)	-0,12	(0,03)

Tabela 4.2 - Estimativas dos efeitos dos QTLs em cada safra e seus respectivos erros padrão (EP), cromossomos (Cr), posições físicas (em megabases - Mb) e *LOD Score* (LOD) resultantes da análise MTMIM realizada com base nas médias marginais dos três anos de avaliação de uma população de RILs de sorgo sacarino. Os efeitos dos QTLs encontram-se na mesma escala do caractere avaliado, sendo: Brix em ºBrix; e Pol, AR e Extração em %. Efeitos positivos representam alelos herdados do genitor Wray e efeitos negativos representam alelos do genitor Brandes.

QTL	Cr	Posição (Mb)	LOD	Efeito					
				Safra1	(EP)	Safra2	(EP)	Safra3	(EP)
Brix1	1	6,55	21,50	0,43	(0,06)	0,35	(0,06)	0,58	(0,06)
Brix2	2	30,98	12,40	0,15	(0,06)	-0,13	(0,06)	-0,18	(0,06)
Brix3	3	15,97	20,40	0,41	(0,06)	0,21	(0,06)	0,48	(0,06)
Brix4	3	66,81	25,00	0,70	(0,06)	0,44	(0,06)	0,31	(0,06)
Brix5	4	0,98	7,41	-0,36	(0,06)	-0,22	(0,06)	-0,17	(0,06)
Brix6	4	20,86	9,20	0,30	(0,06)	0,23	(0,07)	0,29	(0,06)
Brix7	4	52,44	6,89	0,01	(0,06)	-0,25	(0,07)	-0,21	(0,06)
Brix8	4	63,11	13,60	-0,01	(0,06)	0,17	(0,06)	0,41	(0,06)
Brix9	5	4,78	11,90	0,37	(0,06)	0,47	(0,06)	0,31	(0,06)
Brix10	6	12,20	20,30	0,10	(0,06)	-0,30	(0,06)	0,17	(0,06)
Brix11	6	46,64	5,76	-0,07	(0,07)	-0,16	(0,07)	-0,32	(0,07)
Brix12	6	52,51	12,40	-0,40	(0,06)	-0,21	(0,07)	-0,44	(0,06)
Brix13	8	43,62	15,80	0,25	(0,06)	-0,02	(0,06)	-0,15	(0,06)
Brix14	9	2,55	9,36	0,09	(0,06)	0,37	(0,06)	0,04	(0,06)
Brix15	9	10,16	13,70	0,36	(0,06)	0,33	(0,06)	0,46	(0,06)
Pol1	1	6,55	17,80	0,45	(0,06)	0,38	(0,06)	0,60	(0,06)
Pol2	2	30,68	7,71	-0,01	(0,06)	-0,19	(0,06)	-0,24	(0,06)
Pol3	3	15,90	11,70	0,41	(0,06)	0,25	(0,06)	0,41	(0,06)
Pol4	3	68,07	24,00	0,66	(0,06)	0,42	(0,06)	0,39	(0,06)
Pol5	4	63,11	11,70	0,05	(0,06)	0,23	(0,06)	0,41	(0,06)
Pol6	5	4,78	14,00	0,36	(0,06)	0,49	(0,06)	0,33	(0,06)
Pol7	6	10,94	12,90	0,04	(0,06)	-0,21	(0,06)	0,18	(0,06)
Pol8	6	46,64	13,60	-0,11	(0,07)	-0,22	(0,07)	-0,38	(0,07)
Pol9	6	53,55	10,40	-0,41	(0,06)	-0,19	(0,06)	-0,29	(0,06)
Pol10	8	45,39	7,29	0,09	(0,06)	-0,03	(0,06)	-0,24	(0,06)
Pol11	9	2,77	8,19	0,18	(0,06)	0,40	(0,06)	0,21	(0,06)
Pol12	9	10,14	9,43	0,30	(0,06)	0,28	(0,06)	0,38	(0,06)
AR1	1	44,34	8,96	-0,05	(0,01)	-0,03	(0,01)	-0,01	(0,01)
AR2	3	57,50	12,90	-0,07	(0,01)	-0,03	(0,01)	-0,01	(0,01)
AR3	3	67,59	12,00	-0,09	(0,01)	-0,04	(0,01)	-0,02	(0,01)
AR4	4	0,43	17,70	-0,11	(0,01)	-0,04	(0,01)	-0,03	(0,01)
AR5	6	42,00	8,28	-0,07	(0,01)	-0,01	(0,01)	-0,02	(0,01)
Extração1	1	6,40	15,10	-0,11	(0,13)	-0,54	(0,08)	-0,54	(0,11)
Extração2	1	51,68	7,29	-0,14	(0,13)	-0,34	(0,08)	-0,56	(0,11)
Extração3	3	47,72	9,13	-0,23	(0,13)	-0,36	(0,08)	-0,66	(0,11)
Extração4	5	4,78	11,10	-0,35	(0,13)	-0,48	(0,08)	-0,34	(0,11)
Extração5	5	58,13	8,34	-0,12	(0,13)	0,35	(0,08)	0,34	(0,11)
Extração6	6	46,18	9,50	-0,16	(0,13)	0,17	(0,08)	0,47	(0,11)
Extração7	7	57,48	8,82	0,12	(0,13)	0,24	(0,08)	0,01	(0,11)
Extração8	8	0,02	10,40	-0,24	(0,13)	-0,29	(0,08)	0,02	(0,11)
Extração9	10	20,15	8,59	0,49	(0,13)	-0,13	(0,08)	0,06	(0,11)

Tabela 5 - Estimativas dos efeitos epistáticos entre QTLs em cada safra e os respectivos erros padrão (EP) resultantes da análise MTMIM com base nas médias marginais dos três anos de avaliação de uma população de RILs de sorgo sacarino. Efeitos positivos e negativos podem ser interpretados como aumento e redução, respectivamente, no fenótipo em consequência da interação epistática entre os locos.

Caractere	QTL x QTL	Safra1	(EP)	Safra2	(EP)	Safra3	(EP)
Florescimento (DAS)	3 x 4	-	-	-0,67	(0,17)	-1,35	(0,26)
	4 x 6	0,42	(0,11)	0,88	(0,16)	-	-
Altura (cm)	1 x 3	-	-	-	-	-2,78	(0,71)
	3 x 4	-	-	-	-	1,97	(0,71)
	3 x 5	3,12	(0,82)	4,69	(0,82)	-	-
PMV ($t.ha^{-1}$)	2 x 4	0,79	(0,25)	1,06	(0,30)	-	-
Extração (%)	1 x 7	0,26	(0,10)	0,16	(0,05)	-	-
	4 x 8	0,32	(0,10)	-	-	-	-
	5 x 9	-	-	-0,13	(0,05)	-	-
Brix (ºBrix)	1 x 10	0,13	(0,04)	-	-	-	-
	2 x 6	-	-	-	-	-0,17	(0,05)
	3 x 10	-	-	-	-	0,12	(0,05)
	3 x 13	0,13	(0,04)	-	-	-	-
	10 x 13	-	-	-	-	0,14	(0,05)
	10 x 14	-	-	0,14	(0,05)	-	-
Pol (%)	2 x 8	-0,12	(0,05)	-0,13	(0,05)	-	-
	4 x 9	-0,12	(0,04)	-	-	-	-
	6 x 12	-0,12	(0,04)	-	-	-	-
	7 x 8	-0,20	(0,05)	-	-	-	-
AR (%)	1 x 2	0,03	(0,01)	-	-	-	-
Fibras (%)	5 x 9	-0,15	(0,06)	-0,05	(0,02)	-	-
	6 x 8	0,18	(0,05)	-	-	-	-

FIGURAS

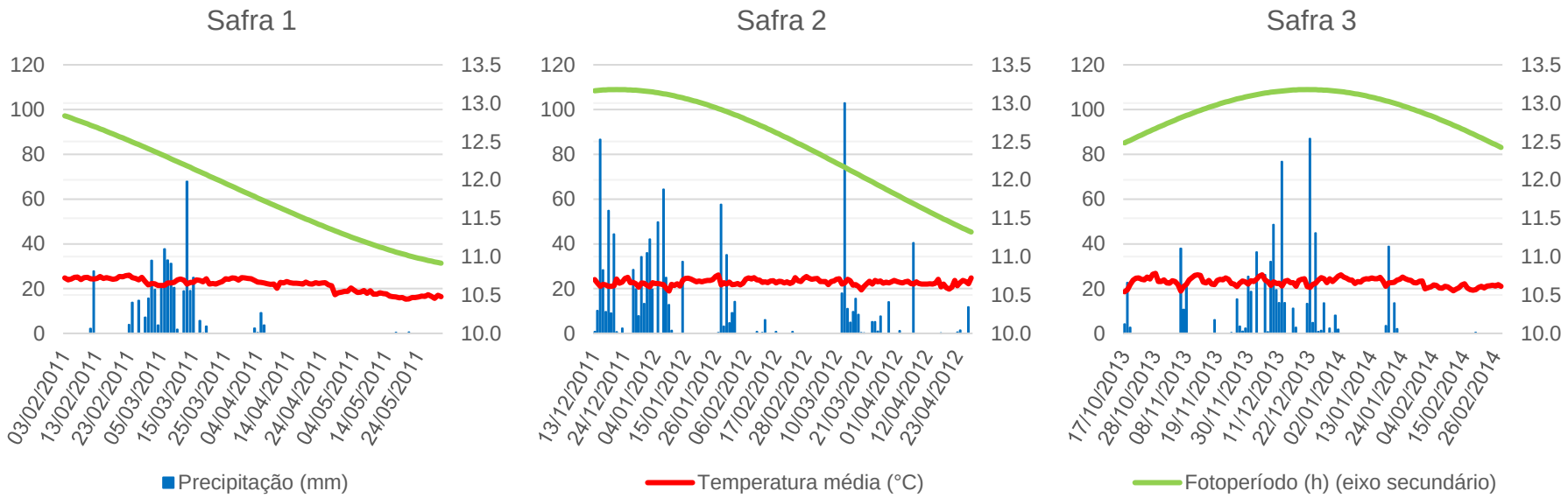

Figura 1 - Precipitação e temperatura média diária e fotoperíodo ao longo dos ciclos de cultivo das três safras da população RILs de sorgo sacarino. As informações climáticas foram obtidas pela estação meterológica da Embrapa Milho e Sorgo, Sete Lagoas, MG, Brasil.

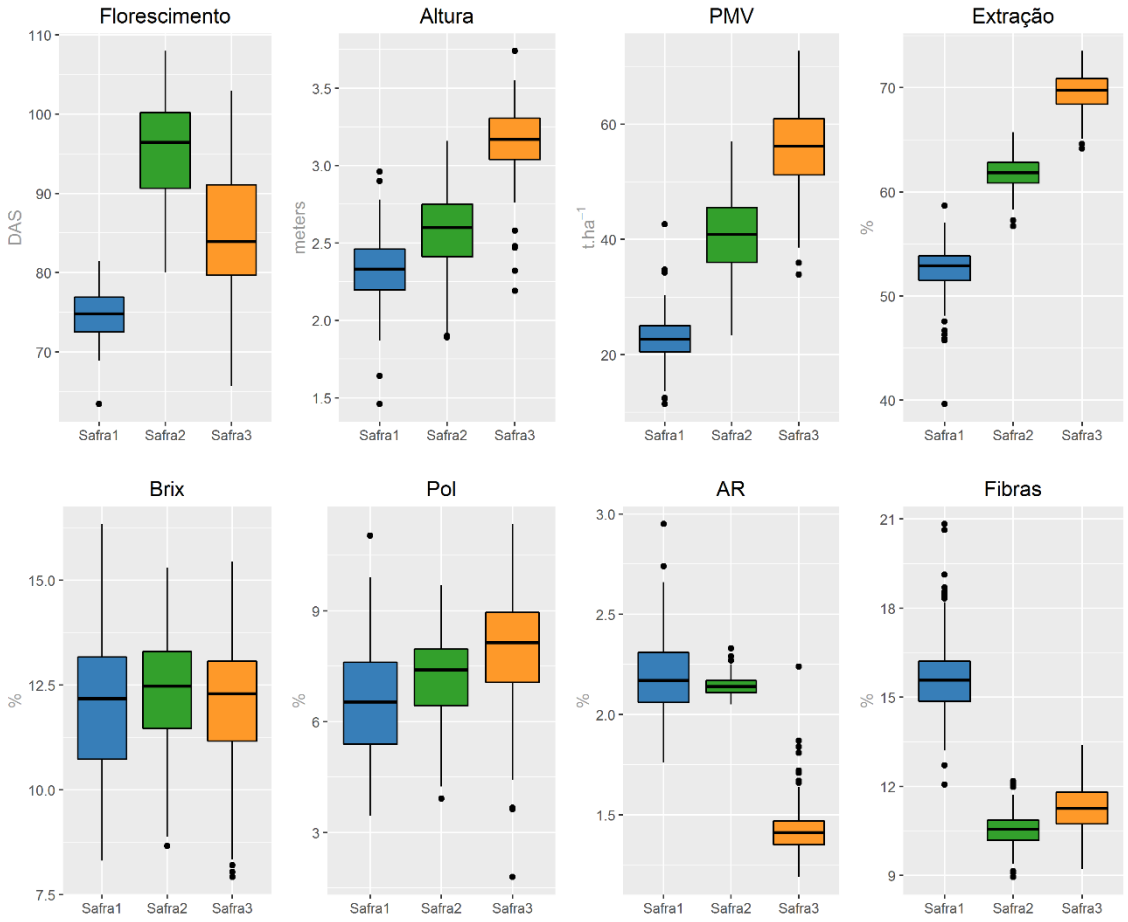

Figura 2 - Distribuição das médias ajustadas, em formato boxplot, para os caracteres agroindustriais avaliados em cada uma das safras, 2010/11 (azul), 2011/12 (verde) e 2013/14 (amarelo). Sendo época de florescimento (Florescimento), em dias após semeadura; altura média da parcela (Altura), em cm; produção de massa verde (PMV), em $t.ha^{-1}$; extração de caldo (Extração), em % da biomassa; sólidos solúveis totais (Brix), em ºBrix; teor de sacarose (Pol), em % do caldo; açúcares redutores (AR), em % do caldo; e fibras do colmo (Fibras), em % da biomassa.

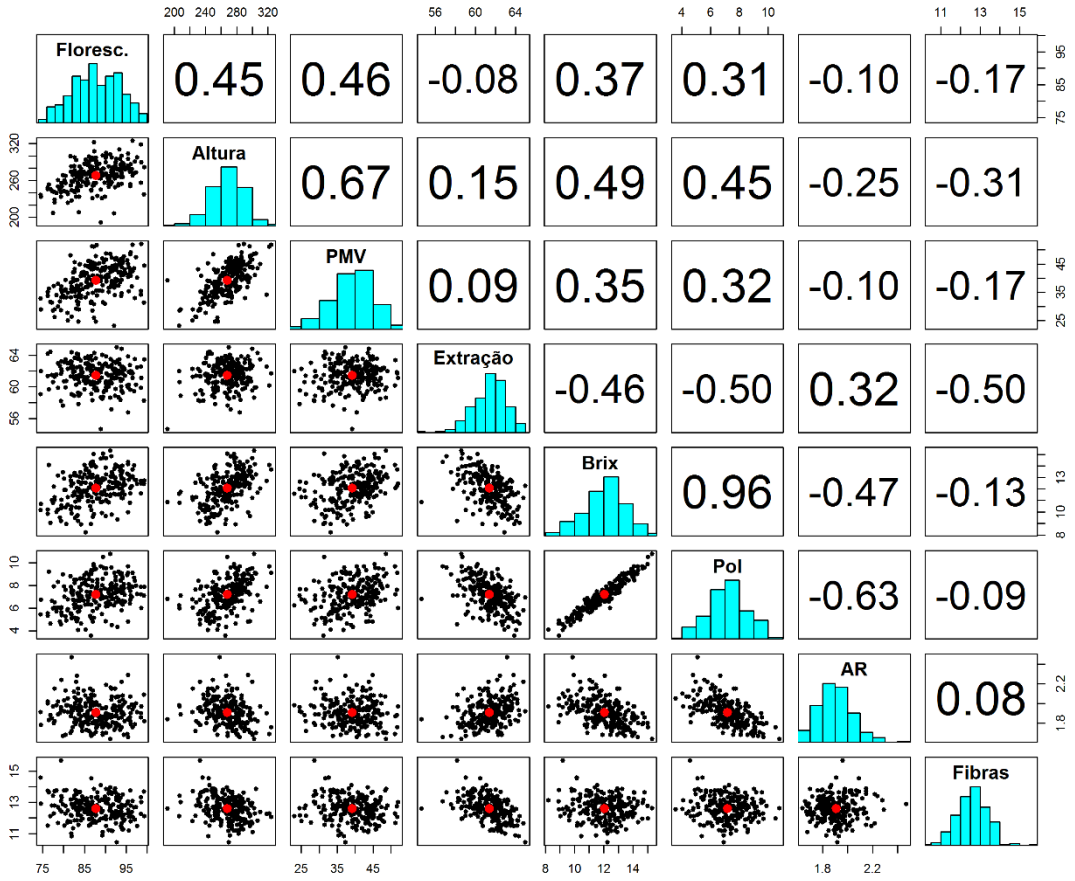

Figura 3 - Histogramas das médias ajustadas de cada caractere (na diagonal), gráficos de dispersão com os valores médios representados em vermelho (abaixo da diagonal) e valores de correlações genéticas (acima da diagonal) entre pares de caracteres avaliados na população de RILs de sorgo sacarino. Sendo época de florescimento (Floresc.), em dias após semeadura; altura média da parcela (Altura), em cm; produção de massa verde (PMV), em $t.ha^{-1}$; extração de caldo (Extração), em % da biomassa; sólidos solúveis totais (Brix), em ºBrix; teor de sacarose (Pol), em % do caldo; açúcares redutores (AR), em % do caldo; e fibras do colmo (Fibras), em % da biomassa.

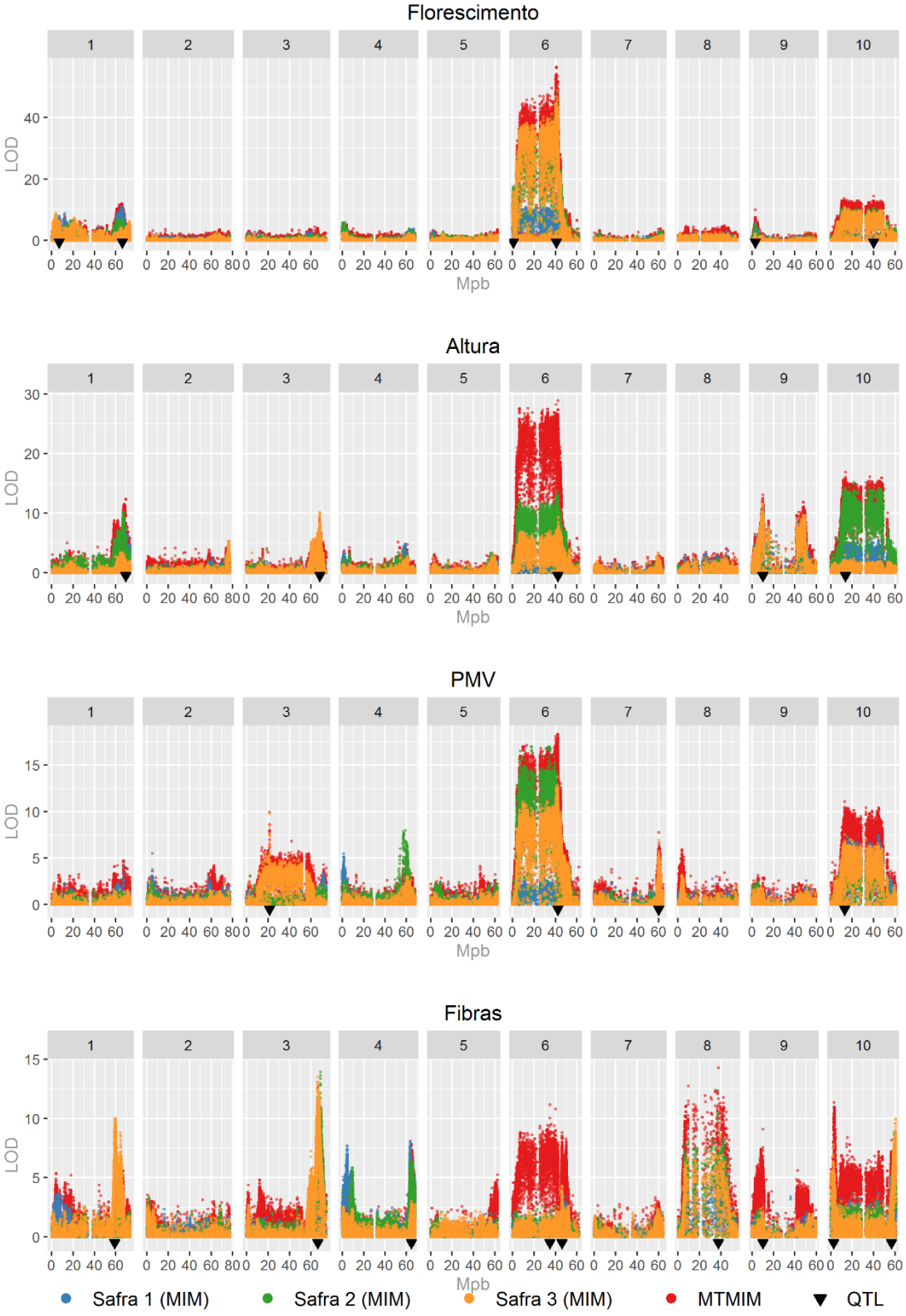

Figura 4 - Perfil obtido, para os 10 cromossomos, a partir dos valores de *LOD Score* resultantes do mapeamento por múltiplos intervalos (MIM) para florescimento, altura, produção de massa verde (PMV) e fibras em cada safra da população de RILs de sorgo sacarino, e do mapeamento por múltiplos intervalos multivariado (MTMIM), com os QTLs mapeados nas três safras conjuntamente.

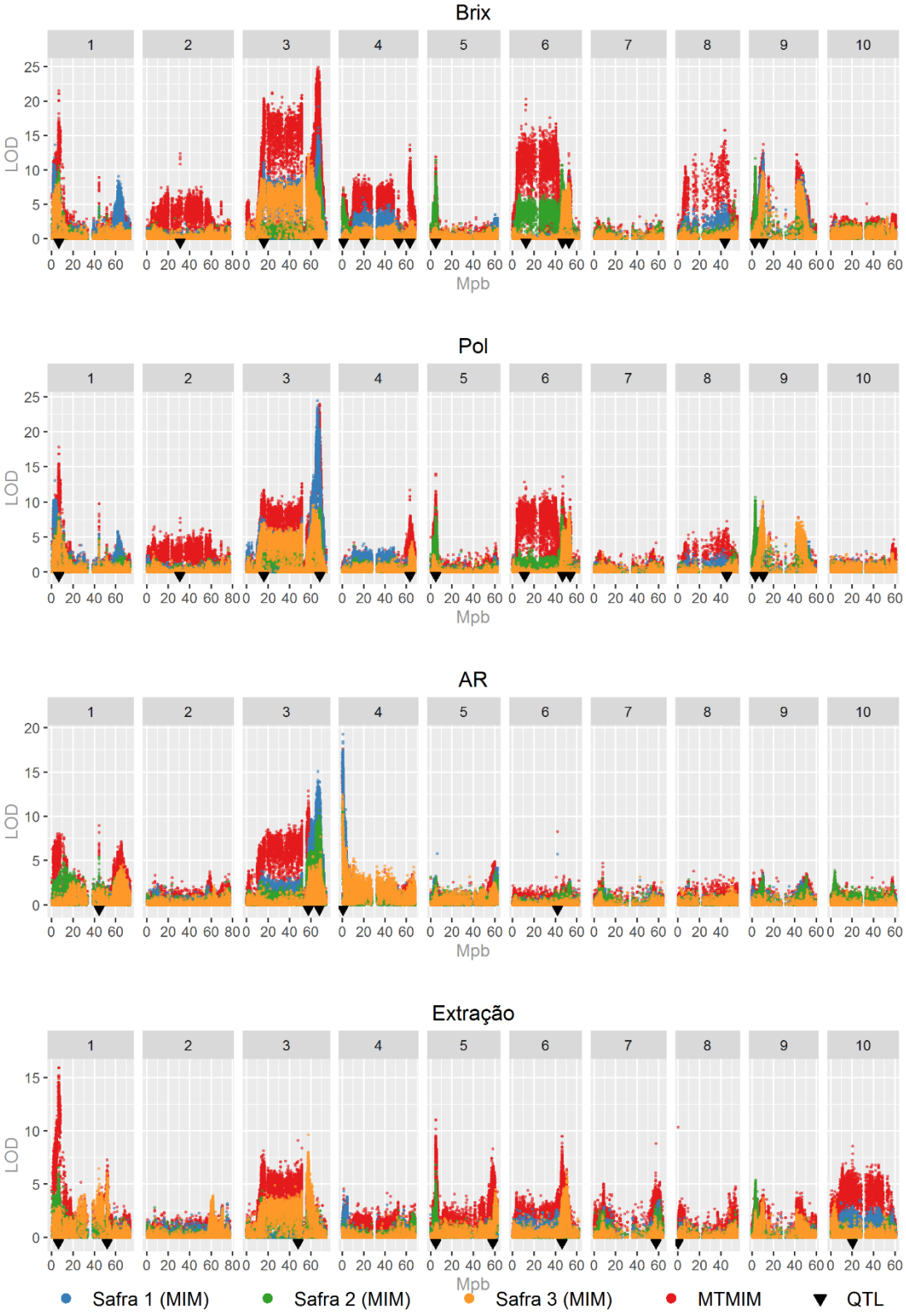

Figura 5 - Perfil obtido, para os 10 cromossomos, a partir dos valores de *LOD Score* resultantes do mapeamento por múltiplos intervalos (MIM) para sólidos solúveis totais (Brix), sacarose (Pol), açúcares redutores (AR) e extração de caldo em cada safra da população de RILs de sorgo sacarino, e do mapeamento por múltiplos intervalos multivariado (MTMIM), com os QTLs mapeados nas três safras conjuntamente.

3. CAPÍTULO 2

MAPEAMENTO ASSOCIATIVO, A PARTIR DE UMA ABORDAGEM PARA MÚLTIPLOS LOCOS, PARA CARACTERES RELACIONADOS À PRODUÇÃO E QUALIDADE DA BIOMASSA EM UM PAINEL DE SORGO

RESUMO

O *Sorghum bicolor* representa uma potencial matéria-prima para a produção de bioenergia, e o teor de lignina da biomassa é um componente importante para direcionar a seleção e o desenvolvimento de genótipos, tanto para a combustão quanto para a produção de etanol de segunda geração. O objetivo do presente trabalho foi identificar regiões genômicas e/ou genes candidatos associados à qualidade e produção de biomassa através do estudo de associação genômica ampla, utilizando-se um painel diverso de sorgo com 100 genótipos. Os experimentos foram conduzidos em duas safras, 2011 e 2012, em Sete Lagoas, MG, Brasil. Os caracteres avaliados foram: época de florescimento, altura de plantas, produção de matéria verde (PMV), produção de matéria seca (PMS) e teor de lignina em detergente ácido. Para o mapeamento associativo foram utilizados 260.408 marcadores do tipo SNP, obtidos via genotipagem por sequenciamento. Diferentes modelos mistos foram examinados, com e sem a incorporação do grau de parentesco entre os indivíduos via matriz de VanRaden (K), bem como com e sem as informações de estrutura populacional, por Monte Carlo via cadeias de Markov (Q) ou via componentes principais (*Principal Components* - PC). Com base nos critérios AIC e BIC, dentre os modelos Naïve, PC, Q, K, K + PC e K+Q, o modelo K + Q foi selecionado para ser utilizado na abordagem de múltiplos locos para a análise de associação. Ao todo, foram mapeados cinco SNPs para altura; três para florescimento; dois para lignina e para PMV; e um SNP para PMS, considerando a correção de Bonferroni (α = 0,05). Os SNPs mapeados para florescimento e altura corroboram resultados anteriores já publicados em outros trabalhos e um mesmo SNP foi significativo para florescimento, PMV e PMS. O SNP mais significativo para lignina foi mapeado dentro de uma região da família gênica *Small Auxin Upregulated RNA* (SAUR), de proteínas responsivas à auxina. Esses resultados podem auxiliar no desenvolvimento de estratégias de melhoramento do sorgo biomassa para a produção de bioenergia.

Palavras-chave: *Sorghum bicolor*; *Genotyping-by-Sequencing*; *Quantitative Trait Loci*; GWAS; lignina.

ABSTRACT

Sorghum bicolor is presented as a potential feedstock for bioenergy production. In this context, lignin content is an important biomass component for genotype selection and development for cogeneration or second generation ethanol production. This work aimed to identify genomic regions and/or candidate genes associated with biomass production and quality applying a GWAS approach using a genetic diverse sorghum panel (100 genotypes). Experiments were carried out in Sete Lagoas, Minas Gerais, Brazil, during the 2011 and 2012 cropping seasons. Characters evaluated in sorghum plants during harvest time were: flowering date, plant height, fresh and dry mass production (FMP and DMP), and lignin content. For GWAS mapping, 260.408 SNP markers were obtained by GBS analysis. The degree of kinship between the genotypes (K) was calculated via VanRaden matrix and the population structure was determined by principal component analysis (PC) and Markov Chain Monte Carlo (Q). AIC and BIC criteria were used to select the best model (Naive, PC, Q, K, K + PC and K + Q) and the K + Q model was selected to be applied in the multi-locus mixed model (MLMM) approach used for GWAS analysis. Five SNPs were mapped for plant height; three for flowering; two for lignin and FMP; and one SNP for DMP, according to Bonferroni correction (p-value < 0.05). The SNPs mapped for flowering and plant height corroborate previous results. A same SNP was significant for flowering, FMP and DMP. The most significant SNP for lignin is within a region containing a gene belonging to the Small Auxin Upregulated RNA (*SAUR*) gene family, which are proteins responsive to auxin. These results and future studies will contribute to developing new strategies for biomass sorghum breeding form bioenergy production.

Keywords: *Sorghum bicolor*; *Genotyping-by-Sequencing*; *Quantitative Trait Loci*; GWAS; lignin

3.1. INTRODUÇÃO

A biomassa do sorgo vem se tornando uma importante matéria-prima para produção de bioenergia (Jonker *et al.* 2015). Atualmente, os programas de melhoramento genético do sorgo trabalham no desenvolvimento de variedades específicas para esta finalidade, com alta produtividade e composição diferenciada da biomassa. Dentre as principais estratégias para o melhoramento do sorgo biomassa está a alteração do teor de lignina (Yan *et al.* 2015). A lignina é um composto polifenólico que está associado à celulose na parede celular, e confere rigidez e resistência aos tecidos vegetais (Weng & Chapple 2010).

A lignificação é um importante mecanismo de proteção, do ponto de vista evolutivo. Entretanto, altos teores de lignina tornam os materiais genéticos recalcitrantes para uso em biorrefinarias. Isto é um problema, por exemplo, para produção de etanol de segunda geração, que precisa desconstruir a parede celular para fermentar os açúcares oriundos da celulose. Por outro lado, a lignina possui elevado poder calorífico e pode ser utilizada na queima direta da biomassa para produção de calor e energia (Naik *et al.* 2010). Entretanto, a determinação do teor de lignina nos materiais vegetais baseia-se em métodos demorados, laboriosos ou de alto custo (Lupoi *et al.* 2015).

A fim de acelerar as etapas no processo de melhoramento genético das cultivares de sorgo para produção de bioenergia, faz-se necessário o estudo do controle genético da lignina assim como dos caracteres relacionados à produção de biomassa em campo, como a altura ou porte das plantas, a sensibilidade ao fotoperíodo e a produtividade total de biomassa fresca e seca (Mullet *et al.* 2014). O mapeamento associativo é uma das técnicas mais promissoras para o mapeamento genético, ou estudo de associação genômica ampla (do inglês, *Genome Wide Association Studies* - GWAS) (Mackay *et al*. 2009). O GWAS, inicialmente desenvolvido para estudo de doenças em populações humanas, tem sido amplamente utilizado para identificação de polimorfismos de características complexas em plantas cultivadas (Flint-Garcia 2013).

Vários modelos genético-estatísticos para GWAS foram propostos. Atualmente, a estrutura de população e o grau de relacionamento genético entre os indivíduos são considerados as principais causas de falsas associações entre

fenótipos e genótipos e, portanto, devem ser incorporados aos modelos para um maior controle de falsos-positivos (Yu *et al*. 2006). Nas análises de associação, a correção para estrutura populacional é feita a partir da incorporação de cofatores, incluídos como efeitos fixos no modelo, que representam o agrupamento genealógico ou as subpopulações presentes no painel. Além disso, a utilização de modelos lineares mistos permite também a inclusão do grau de parentesco entre os indivíduos no modelo associativo, a partir de uma matriz de relacionamento genético (Kang *et al.* 2010).

Inicialmente, os modelos de GWAS baseavam-se em testes individuais realizados para cada loco. No entanto, Segura *et al.* (2012) propuseram uma abordagem de modelos mistos para múltiplos locos (do inglês, *Multi-loci Mixed Models* - MLMM), que considera SNPs como cofatores no modelo. A proposta de múltiplos locos no modelo GWAS é similar à abordagem utilizada para o mapeamento de QTLs por intervalo composto (Jansen 1993) e por múltiplos intervalos (Zeng 1994). No entanto, no caso do GWAS, o uso de uma abordagem para múltiplos locos é ainda mais relevante, pois um efeito confundido pode estar presente ao longo do genoma, devido a existência de desequilíbrio de ligação, e não apenas localmente, devido à ligação física entre locos.

Desta forma, o MLMM aumenta o poder do teste e diminui a taxa de falsos positivos. Nesse contexto, o objetivo do presente trabalho foi identificar regiões genômicas associadas à produção e qualidade de biomassa em um painel com ampla diversidade genética de sorgo, utilizando a abordagem MLMM, e assim identificar possíveis genes candidatos que serão investigados e validados em estudos futuros quanto ao potencial de utilização na seleção assistida por marcadores moleculares.

3.2. MATERIAL E MÉTODOS

3.2.1. Material Genético e Delineamento Experimental

Foram avaliados 100 genótipos que compõem um painel com ampla diversidade genética de sorgo para produção de bioenergia, oriundos da coleção núcleo do CIRAD, ICRISAT e do Banco de Germoplasma da Embrapa Milho e Sorgo. Esse painel é composto por acessos de diferentes raças e sub-raças do sorgo (*caudatum, bicolor, guinea, kafir e durra*) de pelo menos 19 países e de diferentes regiões, como Brasil, Estados Unidos da América, Índia, África e Norte da Austrália. Os genótipos foram escolhidos por serem bastante distintos quanto à composição da biomassa e ao teor de lignina. No entanto, também apresentam diversidade de porte e diferentes graus de sensibilidade ao fotoperíodo.

Os experimentos foram conduzidos em duas safras, a primeira semeada em fevereiro de 2011 e a segunda em janeiro de 2012, em área experimental pertencente à Embrapa Milho e Sorgo, em Sete Lagoas, MG, Brasil. As parcelas experimentais foram compostas por duas linhas de 5 m de comprimento, espaçadas 0,70 m, com densidade de 9 plantas por metro linear. O delineamento foi em látice 10x10 com três repetições.

Os caracteres avaliados foram: época de florescimento (Floresc.), em dias decorridos da semeadura à antese; altura de plantas (Altura), média da parcela em cm; produção de massa verde (PMV), a partir do peso total da parcela convertido em tonelada por hectares; produção de massa seca (PMS), aferido a partir de subamostras secas em estufa a 65 ºC até apresentar peso constante; e teor de lignina em detergente ácido (Lignina), conforme Van Soest (1991).

3.2.2. Análises Fenotípicas

Os caracteres fenotípicos foram analisados com base na abordagem de modelos mistos e os componentes de variância foram estimados via máxima verossimilhança restrita (do inglês, *Restricted Maximum Likelihood* – REML) a partir do seguinte modelo:

$$y_{ijkl} = \mu + s_l + r_{k(l)} + b_{j(kl)} + g_i + gs_{il} + \varepsilon_{ijkl}$$

em que: y_{ijkm} é o valor fenotípico observado para o indivíduo *i* no bloco *j*, repetição *k* e safra *l*; μ é a média geral; s_l é o efeito fixo da *l*-ésima safra (l = 1, ..., L; L = 2); $r_{k(l)}$ é o efeito fixo da *k*-ésima repetição (k = 1, . . . , K; K = 3) na safra *l*; $b_{j(kl)}$ é o efeito aleatório do *j*-ésimo bloco (j = 1, . . . , J ; J = 10) na repetição *k* e na safra *l*; g_i é o efeito aleatório do *i*-ésimo genótipo (i = 1, . . . , I ; I = 100); gs_{il} é o efeito aleatório da interação do *i*-ésimo genótipo com a safra *l*; e ε_{ijkl} é o resíduo.

Além dos efeitos fixos de safras e repetições, as covariáveis estande de plantas e época de florescimento foram testadas com base no teste de Wald e mantidas no modelo quando significativas (*p*-valor < 0,05). Para as análises dos efeitos aleatórios, blocos, genótipos e interação genótipo x safra (GxS), foi utilizado o teste da razão de verossimilhança (do inglês, *Likelihood Ratio Test* - LRT), realizado a partir da diferença entre as deviances para os modelos com e sem o efeito a ser testado, o qual apresenta distribuição qui-quadrado com 1 grau de liberdade.

Para obtenção dos coeficientes de variação e das herdabilidades, as seguintes equações foram utilizadas, respectivamente:

$$CV = \frac{\sqrt{\sigma_e^2}}{\bar{x}} \times 100 \qquad h_m^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gs}^2}{l} + \frac{\sigma_e^2}{k.l}}$$

sendo: σ_g^2 , σ_{gs}^2 e σ_e^2: a variância genética, a variância da interação entre genótipos e safras, e a variância residual, respectivamente. $\bar{x}$: a média geral, k: o número de repetições e l: o número de ambientes (safras).

Para o ajuste das médias, ou seja, para predizer as médias dos genótipos via BLUP (*Best Linear Unbiased Predictions*) foi utilizado o seguinte modelo:

$$y_{ijkl} = \mu + s_l + r_{k(l)} + b_{j(kl)} + g_{il} + \varepsilon_{ijkl}$$

em que: y_{ijkm} é o valor fenotípico observado para o indivíduo *i* no bloco *j*, repetição *k* e safra *l*; μ é a média geral; s_l é o efeito fixo da *l*-ésima safra (l = 1, ..., L; L = 2); $r_{k(l)}$ é o efeito fixo da *k*-ésima repetição (k = 1, . . . , K; K = 3) na safra *l*; $b_{j(kl)}$ é o

efeito aleatório do *j*-ésimo bloco (*j* = 1, . . . , *J* ; *J* = 10) na repetição *k* e na safra *l*; g_{il} é o efeito aleatório do *i*-ésimo genótipo (*i* = 1, . . . , *I* ; *I* = 100) na safra *l*; e $\boldsymbol{\varepsilon}_{ijkl}$ é o resíduo. Os vetores dos efeitos genéticos e residuais apresentaram distribuição normal multivariada com média zero e matriz de variância-covariância (VCOV) $\mathbf{G}_m$ e $\mathbf{R}_m$, respectivamente.

Diferentes estruturas para as matrizes de variância e covariância (VCOV) genética (G) foram comparadas utilizando o critério AIC (*Akaike Information Criterion*) (Akaike 1974) e BIC (*Bayesian Information Criterion*) (Schwarz 1978). Em resumo, foram comparadas estruturas de VCOV que assumiam ausência de correlação entre ambientes e homogeneidade (Identidade) ou heterogeneidade de variâncias (Diagonal), e estruturas com heterogeneidade de variâncias e existência de correlações genética específicas para cada par de ambientes, ou seja, cada par de safras (Não-estruturada) (Malosetti *et al.* 2013). Para as matrizes de VCOV residuais foram utilizados modelos diagonais.

As análises de modelos mistos foram realizadas utilizando o software GenStat (v. 16.1) (VSN International 2014). Correlações genéticas entre as médias ajustadas dos genótipos para cada par de caracteres foram calculadas pelo método de Pearson e testadas assumindo nível global de significância de 0,05 utilizando o pacote psych (Revelle 2014) disponível no software R (R Core Team 2014).

3.2.3. Extração de DNA Genômico e Genotipagem Via GBS

Amostras de folhas foram coletadas para extração de DNA, sendo que cada amostra de DNA genômico foi extraída de uma única planta do painel por meio do *DNeasy Plant Mini Kit* da QIAGEN. A qualidade e quantidade do DNA extraído foram verificadas por Nanodrop e gel de agarose. As amostras foram enviadas ao *Institute for Genomic Diversity* (IGD, Cornell University, Ithaca, NY, EUA) para genotipagem por sequenciamento (do inglês, *Genotyping-by-Sequencing* – GBS), de acordo com o procedimento desenvolvido por Elshire *et al.* (2011). A enzima ApeKI foi utilizada para a digestão, e as bibliotecas de GBS foram sequenciadas na plataforma HiSeq2000 (Illumina, Inc.).

O pipeline do programa TASSEL Version 4 (Bradbury et al., 2007) foi utilizado para o *SNP call* dos resultados do sequenciamento. As *tags* sequenciadas foram alinhadas com a versão do genoma *Sorghum bicolor* v2.1 (Paterson *et al*. 2009; Goodstein *et al*. 2012). Os sítios ambíguos ou heterozigotos foram definidos como dados perdidos, que foram imputados em cada cromossomo separadamente usando o software NPUTE (Roberts *et al.* 2007), com janela para imputação que variou entre 50 e 70 SNPs. Os dados foram posteriormente filtrados para a frequência do alelo menos comum (do inglês, *Minor Allele Frequency* – MAF) de 5%, para eliminar possíveis erros de sequenciamento (Glaubtiz *et al*. 2014).

3.2.4. Análises Genotípicas

A matriz de parentesco (K) entre os genótipos foi calculada utilizando o método proposto por VanRaden (2008) e implementado no pacote GAPIT (Lipka *et al.* 2012) disponível no software R. Para estimar a estrutura populacional, foi utilizada a análise de componentes principais (do inglês, *Principal Componente Analysis* - PCA), a partir do pacote pcaMethods do Bioconductor (Stacklies *et al*. 2007), disponível no software R, e também, a análise Bayesiana com base no processo iterativo de Monte Carlo via Cadeias de Markov (MCMC), implementada no software Structure (Pritchard *et al.* 2000).

Para viabilizar a análise no software Structure, em relação ao tempo computacional, foi realizada uma amostragem aleatória de 5.000 SNPs, a partir do arquivo final de dados genotípicos já imputados e filtrados, utilizando a função *sample* do software R. Foram considerados os seguintes parâmetros no Structure: modelo *admixture* com frequências alélicas correlacionadas, *burn-in* igual a 10.000 e número de interação MCMC igual a 100.000, com cinco repetições independentes para cada número de subpopulações (*k*) analisado, que variou de 1 a 10. O número de subpopulações mais apropriado foi definido com base na probabilidade log e pela taxa de mudança no delta *k* entre os sucessivos valores de *k*, conforme o método de Evanno *et al.* (2005).

3.2.5. Mapeamento Associativo

As médias ajustadas, obtidas a partir da análise de modelos mistos conjunta para as duas safras, das variáveis altura de plantas, florecimento, PMV, PMS e lignina foram utilizadas no estudo de associação genômica ampla. Inicialmente, diferentes modelos foram examinados (Naíve, Q, PC, K, Q+K e PC+K). Em seguida, o melhor modelo foi selecionado para cada variável com base nos valores de AIC e BIC. O mapeamento associativo foi realizado no pacote MLMM (Segura *et al*. 2012) disponível no software R. Este método é baseado em sucessivas etapas de inclusão, e posteriormente, etapas de exclusão de SNPs como cofatores no modelo. Nas etapas de inclusão, o cofator é o SNP mais significativo da etapa anterior, de forma acumulativa. As inclusões são descontinuadas quando a relação entre a variância genética e a variância fenotípica é próxima de zero. De forma inversa, nas etapas de exclusão, os SNPs menos significativos são eliminados do modelo. A correção de Bonferroni (α = 0,05) foi considerada para determinar o nível de significância individual de cada teste (α / n° de SNPs) e o modelo selecionado foi aquele que apresentou o maior número de locus com *p*-valor abaixo do valor da correção de Bonferroni (mBonf). Os gráficos quantil-quantil (QQ Plots) dos *p*-valores esperados *versus p*-valores observados para os diferentes caracteres também foram utilizados para comparar os diferentes modelos multi-locos. Por fim, as informações de anotações gênicas das regiões onde os SNPs significativos estavam posicionados foram verificadas na versão 2.1 do genoma de *S. bicolor* no banco de dados genômicos Phytozome (Goodstein *et al*. 2012).

3.3. RESULTADOS

3.3.1. Análises Fenotípicas

Na análise dos efeitos fixos do modelo, foi possível verificar que o efeito de safra foi significativo para todos os caracteres. A covariável estande de plantas não foi significativa para lignina, e a época de florescimento só foi significativa para PMV, conforme teste de Wald. Para os efeitos aleatórios, o efeito de blocos não foi significativo para florescimento. Entretanto, os demais efeitos foram significativos para todos os caracteres de acordo com o teste LRT (Tabela 1).

As variâncias genéticas dos caracteres avaliados (florescimento, altura, PMV, PMS e lignina) apresentaram elevada magnitude quando comparadas aos demais componentes de variância. As herdabilidades com base em médias (h^2_m) também reforçam que a maior proporção da variabilidade total do painel é de natureza genética, variando entre 0,75 para PMS e 0,96 para altura. Os coeficientes de variação residual (CV) estão adequados e variaram entre 6,79% para florescimento e 22,47% para PMS (Tabela 2).

Das diferentes estruturas de matrizes de variância e covariância testadas para o efeito de genótipos (G) a matriz não-estruturada apresentou os menores valores de AIC e BIC para todos os caracteres avaliados. Isto indica a existência de diferentes variâncias genéticas para cada safra e de correlações específicas para cada par de safras (Tabela 3).

Após o ajuste final do modelo para cada caractere, as médias ajustadas obtidas a partir do melhor preditor linear não viesado (do inglês, *Best Linear Unbiased Predictor* – BLUP) foram correlacionadas. De maneira geral, os caracteres agronômicos (florescimento, altura, PMV e PMS) apresentaram correlações genéticas positivas e de alta magnitude entre si, enquanto o teor de lignina apresentou correlações genéticas negativas, porém não significativas pelo teste de Pearson (*p*-valor > 0,05) com os demais caracteres avaliados (Figura 1).

3.3.2. Análises Genotípicas

Após imputação dos dados genotípicos, 616.901 marcadores polimórficos foram obtidos. A acurácia média da imputação foi maior que 97%. Posteriormente, marcadores com frequência alélica muito baixa (MAF < 5%) foram eliminados e foram obtidos 260.408 SNPs após filtragem. Considerando que o genoma do sorgo tem 726.616.606 pares de bases, em média foi obtido um SNP a cada 2.790 pares de bases, aproximadamente. Entretanto, a distribuição dos SNPs não é uniforme no genoma, as extremidades dos cromossomos geralmente apresentam maior densidade de SNPs, enquanto as regiões dos centrômeros apresentam baixa cobertura, como pode ser observado na Figura 2.

O grau de relacionamento entre indivíduos pode ser visualizado por meio da matriz de parentesco de VanHaden (Figura 3). A maior parte das estimativas de parentesco situaram-se entre 0 e 0,5, na escala entre 0 e 2, sendo que 2 representa o valor máximo de semelhança genética. Assim, a diversidade genética existente no painel oferece base para o uso do mapeamento associativo como estratégia para a busca de regiões genômicas associadas aos caracteres relacionados à produção e à qualidade da biomassa em sorgo.

Conforme o resultado do Structure (Δk 2 = 1249,92), o painel pode ser subdividido basicamente em duas subpopulações. Para visualizar a relação entre os métodos empregados para a determinação da estrutura populacional, os eixos x e y da Figura 4 correspondem aos dois primeiros componentes da análise de PCA, cujos pontos (linhagens) apresentam formato e cor de acordo com os resultados obtidos a partir do software Structure. Assim, com base nessa Figura, é possível observar uma concordância entre os resultados dos dois métodos.

3.3.3. Mapeamento Associativo

As informações de estrutura populacional (PC e Q) e de parentesco entre os genótipos (K) foram utilizadas para comparar e definir o melhor modelo de análise de mapeamento associativo. Os resultados das comparações via AIC e BIC estão presentes na Figura 5. É possível observar que a estrutura populacional estimada via Structure (Q) apresentou melhor ajuste quando comparada à utilização dos componentes principais (PC). Além disso, os modelos que utilizam ambas as correções (K+Q) apresentaram os menores valores de AIC e BIC para todos os caracteres avaliados e, portanto, foram os modelos selecionados para a análise de mapeamento associativo. Na comparação das estimativas lambda dos modelos MLMM por meio dos gráficos quantil-quantil (QQ Plots) é possível observar o melhor ajuste na correção para falsos positivos a partir da inclusão de SNPs como cofatores (Figura 5).

Os resultados dos estudos de associação genômica ampla (GWAS) para os caracteres avaliados pelo critério mBonf ($\alpha = 0,05$) podem ser visualizados na Figura 6. Foram mapeados cinco SNPs para altura de plantas; três para época de florescimento; dois para teor de lignina e para PMV; e um SNP para PMS. Na Tabela 4, estão apresentados o número do cromossomo, posição física, p-valor e MAF de cada SNP significativo.

3.4. DISCUSSÃO

Assim como os componentes de variância demostraram ampla variabilidade genética e alta herdabilidade para os caracteres avaliados, os dados genotípicos evidenciaram baixo grau de relacionamento entre os indivíduos. Além disso, a presença de SNPs significativos nas regiões previamente relatadas para altura e florescimento fundamentam os resultados obtidos no presente trabalho e corroboram estudos anteriores (Brown *et al.* 2008; Morris *et al.* 2013; Higgins *et al.* 2014).

A utilização do Structure para inferir a ascendência genética e permitir a correção para estrutura de população foi amplamente empregada para o sorgo em outros trabalhos (Casa *et al.* 2008; Brown *et al.* 2008; Caniato *et al.* 2011; Caniato *et al.* 2014). Contudo, a utilização da PCA também foi testada no presente estudo. Nesse caso, os dois primeiros componentes principais obtidos com base nos dados genotípicos dos marcadores SNP explicaram apenas 14,7% da variação dos dados observados, e os dez primeiros PCs, cerca de 35,5%. Apesar de a PCA ser amplamente utilizada e sua aplicação ser fundamentada na literatura (Price *et al.* 2006), seu resultado pode refletir mais especificamente o relacionamento familiar, conforme relatado por Price *et al.* (2010).

Apesar dos dois primeiros componentes da PCA apresentarem certa concordância com a estrutura populacional obtida a partir do software Structure (Figura 4), o modelo que incorpora a matriz Q (Structure) apresentou melhor ajuste, com menores valores de AIC e BIC. Além disso, o modelo que utilizou os PCs como cofatores não apresentou SNPs significativos para lignina, por exemplo, pela correção de Bonferrroni. Assim, no presente caso, a simplificação dos dados de marcadores pela PCA pode ter provocado uma perda de informações relevantes, quando comparado aos resultados do software Structure.

Com relação aos resultados do GWAS, o SNP mais significativo para florescimento foi mapeado na posição 41.234.461 pb do cromossomo 6 (Tabela 3). De acordo com Murphy *et al.* (2011), o loco *Ma1* (SbPRR37) está situado entre 40,27 e 40,28 Mpb no cromossomo 6. Entretanto, Higgins *et al.* (2014) também identificaram um SNP a 2 Mb do *Ma1* em 42,07 Mpb. Além de ser significativo para florescimento o SNP S6_41.234.461 também foi significativo para PMV e PMS. Mesmo utilizando época de florescimento como covariável para o ajuste das médias de PMV e PMS, o SNP permaneceu significativo para estes caracteres. O alelo G do S6_41.234.461 conferiu em média um aumento de 12,5 dias no ciclo até o florecimento; 27,25 $t.ha^{-1}$ na PMV e 10,3 $t.ha^{-1}$ na PMS em relação ao alelo A, como pode ser observado na Figura 7.

Inicialmente pensava-se que o *Ma1* estaria ligado ao *locus* de nanismo *Dw2* (Quinby 1974), porém diversos estudos independentes estimam a posição do *Dw2* na posição 42,2 Mpb do cromossomo 6 (Klein *et al*. 2008; Morris *et al*.

2013; Thurber *et al*. 2013). Para o caractere altura, um dos SNPs significativos foi mapeado na posição 41.309.405 pb do cromossomo 6, localizado a 74.944 pb do SNP mais significativo para florescimento.

O SNP mais significativo para altura (S9_58.760.024) foi localizado entre a região relatada para o locus *Dw1*, mapeado aproximadamente a 57 Mpb no cromossomo 9 (Morris *et al*. 2013; Thurber *et al*. 2013), e um outro QTL para altura (SbFL9.1), também localizado no chr9 a aproximadamente 59 Mpb (Thurber *et al*. 2013). Dos cinco SNPs significativos para o caractere florescimento, quatro encontram-se dentro de genes, porém ainda sem anotação no Phytozome (v.2.1).

A existência de um SNP comum para florescimento, PMV e PMS (S6_41.234.461) sugere efeito pleiotrópico. Este SNP representa a relação intrínsica da maturação com a produção final de biomassa, uma vez que os incrementos vegetativos tendem a cessar com o início do estágio reprodutivo. As médias ajustadas de altura, florescimento, PMV e PMS já apresentavam forte correlação positiva e significativa, enquanto lignina apresentava correlação negativa, porém, não significativa com os demais caracteres. Contudo, o SNP S6_41.234.461 representa um alvo interessante para a manipulação do ciclo e incremento da produção de biomassa.

Novaes *et al.* (2010) associaram a relação negativa entre crescimento vegetativo em espécies arbóreas e teor de lignina a uma possível competição entre o carbono alocado pela lignina e o alocado pela celulose e hemicelulose. Entretanto, a ausência de correlação entre PMV e lignina no painel possibilita a identificação de materiais com elevada produtividade de biomassa e diversificado teor de lignina, que podem servir tanto para produção de etanol de segunda geração, como para cogeração de eletricidade. A produção de etanol de segunda geração necessita de materiais com baixo teor, devido aos efeitos de bloqueio e adsorção de enzimas hidrolíticas realizados pela lignina; enquanto a cogeração de eletricidade visa altos teores, devido ao alto poder calorífico conferido pela lignina (Whitfield *et al.* 2012; Frei 2013).

No cromossomo 6, além dos SNPs mapeados para florescimento e altura de plantas nas regiões relacionadas ao *Ma1* e *Dw2*, outro SNP para PMV e um SNP para lignina também foram localizados neste cromossomo em posições distintas. Thurber *et al.* (2013) reportaram alta frequência de introgressão ao

longo do cromossomo 6 em uma população de cruzamento controlado e isto poderia ser atribuído a outros locos-alvo neste cromossomo. Já Morris *et al.* (2013) relataram uma região com baixa heterozigosidade que se estende desde aproximadamente 6,6 Mb a 42 Mb, também no cromossomo 6, e sugeriram que outros locos para altura e florescimento poderiam ser localizados nesta região.

Para o caractere teor de lignina, foram localizados dois SNPs significativos, um na posição 57.358.040 pb do cromossomo 6, como mencionado acima, e outro na posição 11.638.864 do cromossomo 8. O SNP S8_11.638.864 apresentou MAF de 0,06 e encontra-se em uma região com poucos genes descritos. No entanto, o SNP S6_57.358.040 apresentou MAF de 0,27 e foi localizado próximo a três genes da família SAUR (*Small Auxin Upregulated RNA*), Sobic.006G216400.1 a aproximadamente 2.203 pb, Sobic.006G216500.1 a 9.153 pb e Sobic.006G216700.1 a 20.622 pb do SNP. O alelo A do SNP S6_57.358.040 e, também, o alelo A do SNP S8_11.638.864 proporcionaram redução de 1,26% e 2,34%, respectivamente, no teor de lignina, em relação aos alelos C (Figura 7), o que representa uma variação significativa para esse caractere.

A região que circunda o SNP mais significativo para lignina no cromossomo 6, onde encontram-se os três genes da família SAUR, possui aproximadamente 24 kb. No sorgo, um total de 71 genes SAUR foram descritos por Wang *et al.* (2010) e os genes mapeados no presente estudo podem corresponder aos genes SbSAUR45, SbSAUR46 e SbSAUR47, conforme as posições físicas relatadas. SAURs são genes de resposta primária na via de sinalização de auxina (Chen *et al.* 2014) e a possibilidade do SNP mais significativo para lignina (S6_57.358.040) estar associado com genes responsivos a auxina pode estar fundamentada na literatura.

As moléculas de lignina são compostas por diferentes subunidades, ou monolignóis, e estudos anteriores demonstraram que a intervenção na sua via de biossíntese, com o silenciamento de determinados genes, resulta na hiperacumulação de moléculas intermediárias que podem interferir na sinalização celular e no transporte de auxina (Brown *et al.* 2001; Besseau *et al.* 2007; Li *et al*. 2010). Apesar de não ser a situação do presente estudo, a correlação negativa entre o teor de lignina e o desenvolvimento vegetativo pode

também estar relacionada à intervenções na sinalização e no transporte de auxina.

Desta forma, os resultados podem corroborar os estudos que relacionam a auxina à via fenilpropanoide e suas possíveis interferências e interações. O SNP mapeado pode não estar diretamente associado à via de síntese da lignina, mas fornece pistas sobre interações entre biomoléculas, que demandam estudos mais específicos sobre redes moleculares da interação entre genes, proteínas e metabólitos. Os resultados apresentam SNPs e genes candidatos que poderão ser utilizados em estratégias de mapeamento fino, para a identificação de SNPs causativos. Após a validação, os SNPs mapeados poderão ser utilizados em estratégias de clonagem e seleção assistida por marcadores moleculares para auxiliar o melhoramento de sorgo biomassa para a produção de bienergia.

3.5. REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6):716-723.

BESSEAU, S.; HOFFMANN, L.; GEOFFROY, P.; LAPIERRE, C.; POLLET, B.; LEGRAND, M. (2007). Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth. *The Plant Cell* 19:148–162.

BRADBURY, P.J.; ZHANG, Z.; KROON, D.E.; CASSTEVENS, T.M.; RAMDOSS, Y.; BUCKLER, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635.

BROWN, P.J.; ROONEY, W.L.; FRANKS, C.; KRESOVICH, S. (2008). Efficient Mapping of Plant Height Quantitative Trait Loci in a Sorghum Association Population With Introgressed Dwarfing Genes. *Genetics* 180:629–637.

BROWN, D.E.; RASHOTTE, A.M.; MURPHY, A.S.; NORMANLY, J.; TAGUE, B.W.; PEER, W.A.; TAIZ, L.; MUDAY, G.K. (2001). Flavonoids act as negative regulators of auxin transport in vivo in Arabidopsis. *Plant Physiol* 126:524-535.

CANIATO, F.F.; GUIMARÃES, C.T.; HAMBLIN, M.; BILLOT, C.; RAMI, J.F.; HUFNAGEL, B.; KOCHIAN, L.V.; LIU, J.; GARCIA, A.A.F.; HASH, C.T.; RAMU, P.; MITCHELL, S.; KRESIVICH, S.; OLIVEIRA, A.C.; AVELLAR, G.; BORÉM, A.; GLASZMANN, J.C.; SCHAFFERT, R.E.; MAGALHÃES, J.V. (2011). The relationship between population structure and aluminum tolerance in cultivated sorghum. *PLoS ONE* 6(6).

CANIATO, F.F.; HAMBLIN, M.T.; GUIMARAES, C.T.; ZHANG, Z.; SCHAFFERT, R.E.; KOCHIAN, L.V.; MAGALHAES, J.V. (2014). Association mapping

provides insights into the origin and the fine structure of the sorghum aluminum tolerance locus, AltSB. *PLoS ONE* 9(1):e87438.

CASA, A.M.; PRESSOIR, G.; BROWN, P.J.; MITCHELL, S.E.; ROONEY, W.L.; TUINSTRA, M.R.; FRANKS, C.D.; KRESOVICH, S. (2008). Community Resources and Strategies for Association Mapping in Sorghum. *CROP SCIENCE* 48:30–40.

CHEN, Y.; HAO, X.; CAO, J. (2014). Small auxin upregulated RNA (SAUR) gene family in maize: Identification, evolution, and its phylogenetic comparison with Arabidopsis, rice, and sorghum. *J Integr Plant Biol* 56:133–150.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6(5).

EVANNO, G.; REGNAUT, S.; GOUDET, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611–2620.

FLINT-GARCIA, S.A. (2013). Genetics and Consequences of Crop Domestication. *Journal of Agricultural and Food Chemistry*. 61(35): 8267-8276.

FREI, M. (2013). Lignin: Characterization of a Multifaceted Crop Component. *The Scientific World Journal.*

GLAUBITZ, J.C.; CASSTEVENS, T.M.; LU, F.; HARRIMAN, J.; ELSHIRE, R.J.; SUN, Q. et al. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* 9(2): e90346.

GOODSTEIN, D.M.; SHU, S.; HOWSON, R. et al (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40:D1178–D1186.

HIGGINS, R.H.; THURBER, C.S.; ASSARANURAK, I.; BROWN, P.J. (2014). Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families. *G3journal* 4:1593-1602.

JANSEN, R.C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205–211.

JONKER, J.G.G.; HILST, F.V.D.; JUNGINGER, H.M.; CAVALETT, O.; CHAGAS, M.F.; FAAIJ, A.P.C. (2015). Outlook for ethanol production costs in Brazil up to 2030, for different biomass crops and industrial technologies. *Applied Energy*147:593–610.

KANG, H.M.; SUL, J.H.; SERVICE, S.K.; ZAITLEN, N.A.; KONG, S.; FREIMER, N.B.; SABATTI, C.; ESKIN, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42(4).

KLEIN, R.R.; MULLET, J.E.; JORDAN, D.R.; MILLER, F.R.; ROONEY, W.L.; MENZ, M.A.; FRANKS, C.D.; KLEIN, P.E. (2008). The Effect of Tropical Sorghum Conversion and Inbred Development on Genome Diversity as Revealed by High-Resolution Genotyping. *Crop Sci.* 48(S1):S12–S26.

LI, X.; BONAWITZ, N.D.; WENG, J.K.; CHAPPLE, C. (2010). The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thalianais* independent of flavonoids. *Plant Cell* 22:1620–1632.

LIPKA, A.E.; TIAN, F.; WANG, Q.; PEIFFER, J.; LI, M.; BRADBURY, P.J.; GORE, M.A.; BUCKLER, E.S.; ZHANG, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28(18):2397-2399.

LUPOI, J.S.; SINGH, S.; PARTHASARATHI, R.; SIMMONS, B.A.; HENRY, R.J. (2015). Recent innovations in analytical methods for the qualitative and

quantitative assessment of lignina. *Renewable and Sustainable Energy Reviews* 49:871–906.

MACKAY, T.F.C.; STONE, E.A.; AYROLES, J.F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10.

MALOSETTI, M.; RIBAUT, J.M.; VAN EEUWIJK, F.A. (2013). The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Plant Physiology, Frontiers in Physiology* 44:4.

MORRIS, G.P.; RAMU, P.; DESHPANDE, S.P.; HASH, C.T.; SHAH, T.; UPADHYAYA, H.D.; RIERA-LIZARAZU, O.; BROWN, P.J.; ACHARYA, C.B.; MITCHELL, S.E.; HARRIMAN, J.; GLAUBITZ, J.C.; BUCKLER, E.S.; KRESOVICH, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *PNAS* 8(2):453-458.

MULLET, J.; MORISHIGE, D.; MCCORMICK, R.; TRUONG, S.; HILLEY, J.; MCKINLEY, B.; ANDERSON, R.; OLSON, S.N.; ROONEY, W. (2014). Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. Journal of Experimental Botany, 65(13):3479-3489.

MURPHY, R.L.; KLEIN, R.R.; MORISHIGE, D.T.; BRADY, J.A.; ROONEY, W.L.; MILLER, F.R.; DUGAS, D.V.; KLEIN, P.E.; MULLET, J.E. (2011). Coincident light and clock regulation of pseudoresponse regulator protein 37(PRR37) controls photoperiodicflowering in sorghum. *PNAS* 108(39):16469-16474.

NAIK, S.N.; GOUD, V.V.; ROUT, P.K.; DALAI, A.K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* 14:578–597.

NOVAES, E.; KIRST, M.; CHIANG, V.; WINTER-SEDEROFF, H.; SEDEROFF, R. (2010). Lignin and Biomass: A Negative Correlation for Wood Formation and Lignin Content in Trees. *Plant Physiology* 154:555–561.

PATERSON, A.H.; BOWERS, J.E.; BRUGGMANN, R.; DUBCHAK, I.; GRIMWOOD, J. et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 29:551-6.

PRICE, A.L.; PATTERSON, N.J.; PLENGE, R.M.; WEINBLATT, M.E.; SHADICK, N.A.; REICH, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, *38*(8), 904-909.

PRICE, A.L.; ZAITLEN, N.A.; REICH, D.; PATTERSON, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Genetics* 11:459:463.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

QUINBY, J.R. (1974). Sorghum improvement and the genetics of growth College Station, TX: Texas A&M University Press.

R CORE TEAM (2014). The Comprehensive R Archive Network.[http://cran.r-project.org/].

REVELLE, W. (2014) psych: procedures for psychological, psychometric, and personality research. Evanston. Disponível em: <http://cran.r-project.org/package=psych>.

ROBERTS, A.; MCMILLAN, L.; WANG, W.; PARKER, J.; RUSYN, I.; THREADGILL, D. (2007). Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics* 23(13):i401–i407.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6:461-464.

SEGURA, V.; VILHJÁLMSSON, B.J.; PLATT, A.; KORTE, A.; SEREN, U.; LONG, Q.; NORDBORG, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* 44:7.

SILVA, G.A.P.; GEZAN, S.A.; CARVALHO, M.P.; GOUVÊA, L.R.L.; VERARDI, C.K.; OLIVEIRA, A.L.B.; GONÇALVES, P.S. (2014). Genetic parameters in a rubber tree population: heritabilities, genotype-by-environment interactions and multi-trait correlations. Tree Genetics & Genomes 10(6):1511:1518.

STACKLIES, W.; REDESTIG, H.; SCHOLZ, M.; WALTHER, D.; SELBIG, J. (2007). pcaMethods – a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23:1164–1167.

THURBER, C.S.; MA, J.M.; HIGGINS, R.H.; BROWN, P.J. (2013). Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biology* 14:R68.

VAN SOEST, P. V., ROBERTSON, J. B., & LEWIS, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of dairy science*, *74*(10), 3583-3597.

VANRADEN, P.M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci*. 91:4414-4423.

VSN INTERNATIONAL. (2014). GenStat for Windows 16th Edition. Hemel Hempstead: VSN International.

WANG, S.; BAI, Y.; SHEN, C.; WU, Y.; ZHANG, S.; JIANG, D.; GUILFOYLE, T.J.; CHEN, M.; QI, Y. (2010). Auxin-related gene families in abiotic stress response in Sorghum bicolor. *Funct Integr Genomics* 10: 533–546.

WENG, J.; CHAPPLE, C. (2010). The origin and evolution of lignin Biosynthesis. New Phytologist, 187: 273–285.

WHITFIELD, M.B.; CHINN, M.S.; VEAL, M.W. (2012). Processing of materials derived from sweet sorghum for biobased products. *Industrial Crops and Products* 37:362-375.

YAN, Z.; LI, J.; LI, S.; CHANG, S.; CUI, T.; JIANG, Y.; CONG, G.; YU, M.; ZHANG, L. (2015). Impact of lignin removal on the enzymatic hydrolysis of fermented sweet sorghum bagasse. *Applied Energy* 160(15):641-647.

YU, J.; PRESSOIR, G.; BRIGGS, W.H.; BI, I.V.; YAMASAKI, M.; DOEBLEY, J.F.; MCMULLEN, M.D.; GAUT, D.M.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. (2006). A unified mixed-model method for association mapping. *Nature genetics* 38:2.

ZENG, Z.B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468.

TABELAS

Tabela 1 - Análise dos efeitos fixos (Safra (S), Repetição, Estande e Florescimento) via teste de Wald e dos efeitos aleatórios (Blocos, Genótipos (G) e interação entre Genótipos e Safras (GxS)) via teste da razão de verossimilhança (LRT) de cinco caracteres agroindustriais do painel de diversidade genética de sorgo para caracteres relacionados à produção de bioenergia, nas safras 2010/11 e 2011/12.

Caractere	Wald				LRT		
	Safra (S)	Repetição	Estande	Floresc.	Blocos	Genótipos (G)	GxS
Floresc.	74,25**	0,83^{NS}	16,27**	--	0,17^{NS}	296,75**	45,85**
Altura	25,62**	2,03^{NS}	17,72**	1,87^{NS}	78,47**	898,80**	30,46**
PMV	120,63**	2,85*	13,21**	1,50^{NS}	27,89**	394,56**	89,28**
PMS	201,12**	7,55**	13,83**	11,22**	8,83**	291,59**	83,07**
Lignina	22,59**	6,70**	0,00^{NS}	0,02^{NS}	18,69**	244,28**	11,27**

** e *: significativo a 1% e 5%, respectivamente. NS: não significativo.

Tabela 2 - Estimativas da variância genética (σ^2_g), variância de blocos (σ^2_b), variância da interação entre genótipos e safras (σ^2_{gs}), variância residual (σ^2_e), coeficiente de variação residual (CV) e herdabilidade ao nível de média (h^2_m) do painel de diversidade genética de sorgo para caracteres relacionados à produção de bioenergia, nas safras 2010/11 e 2011/12.

Caractere	σ^2_g	σ^2_b	σ^2_{gs}	σ^2_e	CV	h^2_m
Floresc.	42,41	0,29	14,56	26,25	6,79	0,78
Altura	0,65	0,03	0,03	0,06	9,97	0,96
PMV	205,95	14,99	68,38	66,17	14,76	0,82
PMS	20,83	1,17	10,37	10,80	22,47	0,75
Lignina	1,23	0,15	0,23	0,96	16,75	0,82

Tabela 3 - Diferentes modelos examinados para as estruturas das matrizes de variância-covariância para os efeitos genéticos (G) do painel de diversidade genética de sorgo para a produção de bioenergia, nas safras 2010/11 e 2011/12. Os valores em negrito representam as estruturas selecionadas conforme o menor valor de AIC e BIC para cada caractere.

Caractere	Estrutura	Matriz **G**	
		AIC	BIC
Floresc.	Identidade	3768	3781
	Diagonal	3749	3766
	Não-Estruturada	**3667**	**3689**
Altura	Identidade	808	821
	Diagonal	804	822
	Não-Estruturada	**573**	**595**
PMV	Identidade	4747	4760
	Diagonal	4727	4745
	Não-Estruturada	**4636**	**4658**
PMS	Identidade	3593	3607
	Diagonal	3571	3589
	Não-Estruturada	**3507**	**3529**
Lignina	Identidade	2077	2090
	Diagonal	2076	2093
	Não-Estruturada	**2013**	**2035**

Tabela 4 - SNPs significativos identificados no mapeamento associativo para caracteres relacionados à produção de bioenergia do painel de diversidade genética de sorgo pelo método MLMM utilizando correção de Bonferroni (α = 0,05).

QTL	SNP	Cromossomo	Posição (pb)	p-valor	MAF
Floresc.1	S6_41.234.461	6	41.234.461	4,26E-14	0,13
Floresc.2	S3_46.171.140	3	46.171.140	4,90E-10	0,06
Floresc.3	S7_61.157.767	7	61.157.767	2,61E-07	0,13
Altura1	S9_58.760.024	9	58.760.024	2,69E-17	0,22
Altura2	S8_4.868.512	8	4.868.512	9,77E-10	0,45
Altura3	S3_45.240.043	3	45.240.043	1,08E-08	0,30
Altura4	S6_41.309.405	6	41.309.405	1,86E-08	0,27
Altura5	S5_58.728.867	5	58.728.867	5,43E-08	0,20
PMV1	S6_41.234.461	6	41.234.461	2,11E-10	0,13
PMV2	S6_3.285.943	6	3.285.943	1,42E-07	0,41
PMS1	S6_41.234.461	6	41.234.461	2,48E-10	0,13
Lignina1	S6_57.358.040	6	57.358.040	3,93E-08	0,27
Lignina2	S8_11.638.864	8	11.638.864	1,12E-07	0,06

FIGURAS

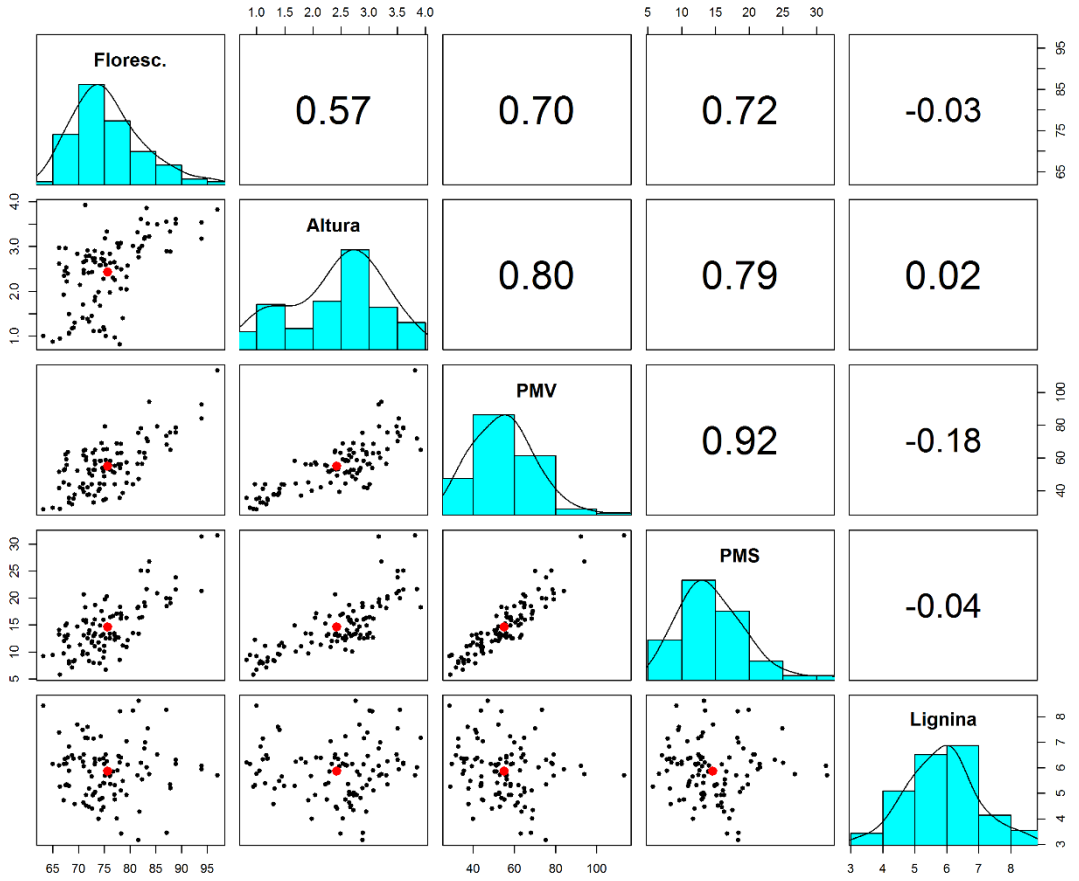

Figura 1 - Histogramas das médias ajustadas para cada caractere (na diagonal), gráficos de dispersão com os valores médios representados em vermelho (abaixo da diagonal) e valores de correlações genéticas (acima da diagonal) entre os pares de caracteres avaliados. Sendo época de florescimento (Floresc.), em dias após semeadura; altura média da parcela (Altura), em m; produção de massa verde (PMV), em $t.ha^{-1}$; produção de massa seca (PMS), em $t.ha^{-1}$; e teor de lignina (Lignina), em %.

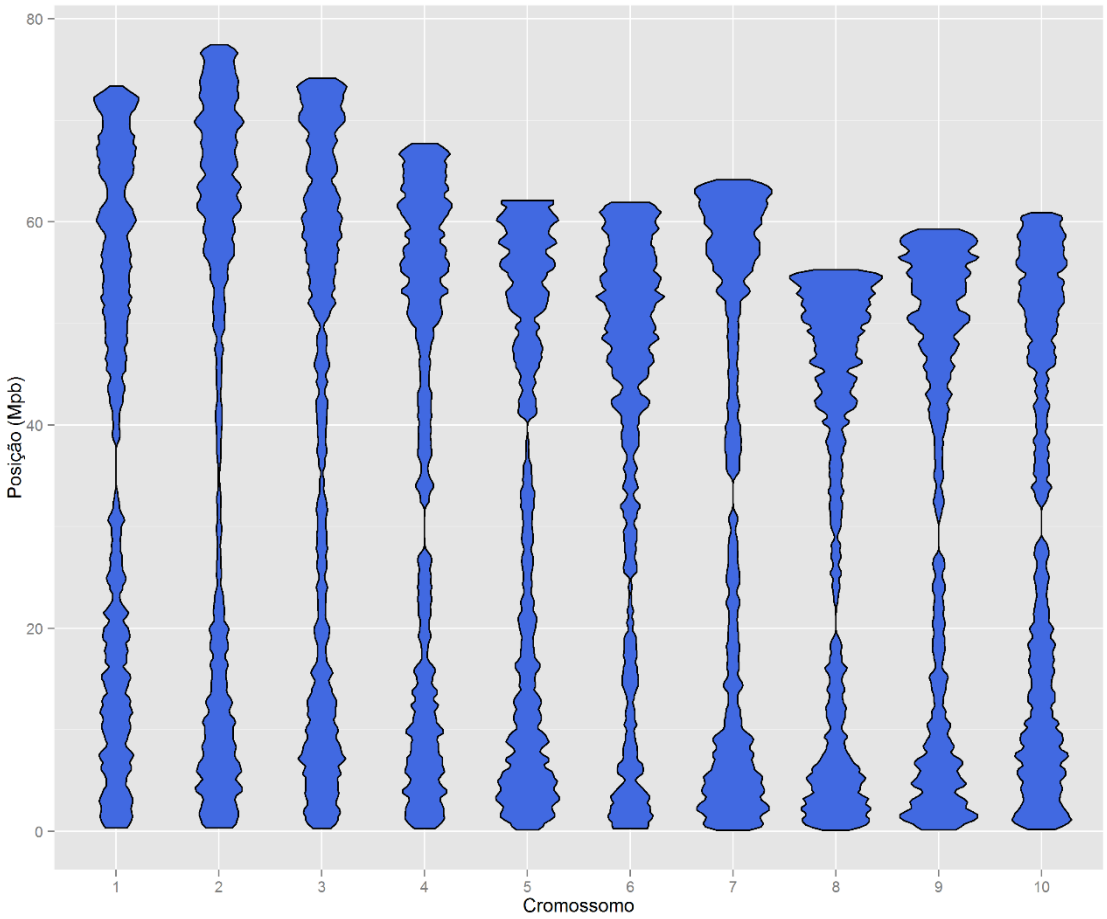

Figura 2 - Densidade de marcadores SNP por cromossomo no painel de diversidade genética de sorgo para produção de bioenergia.

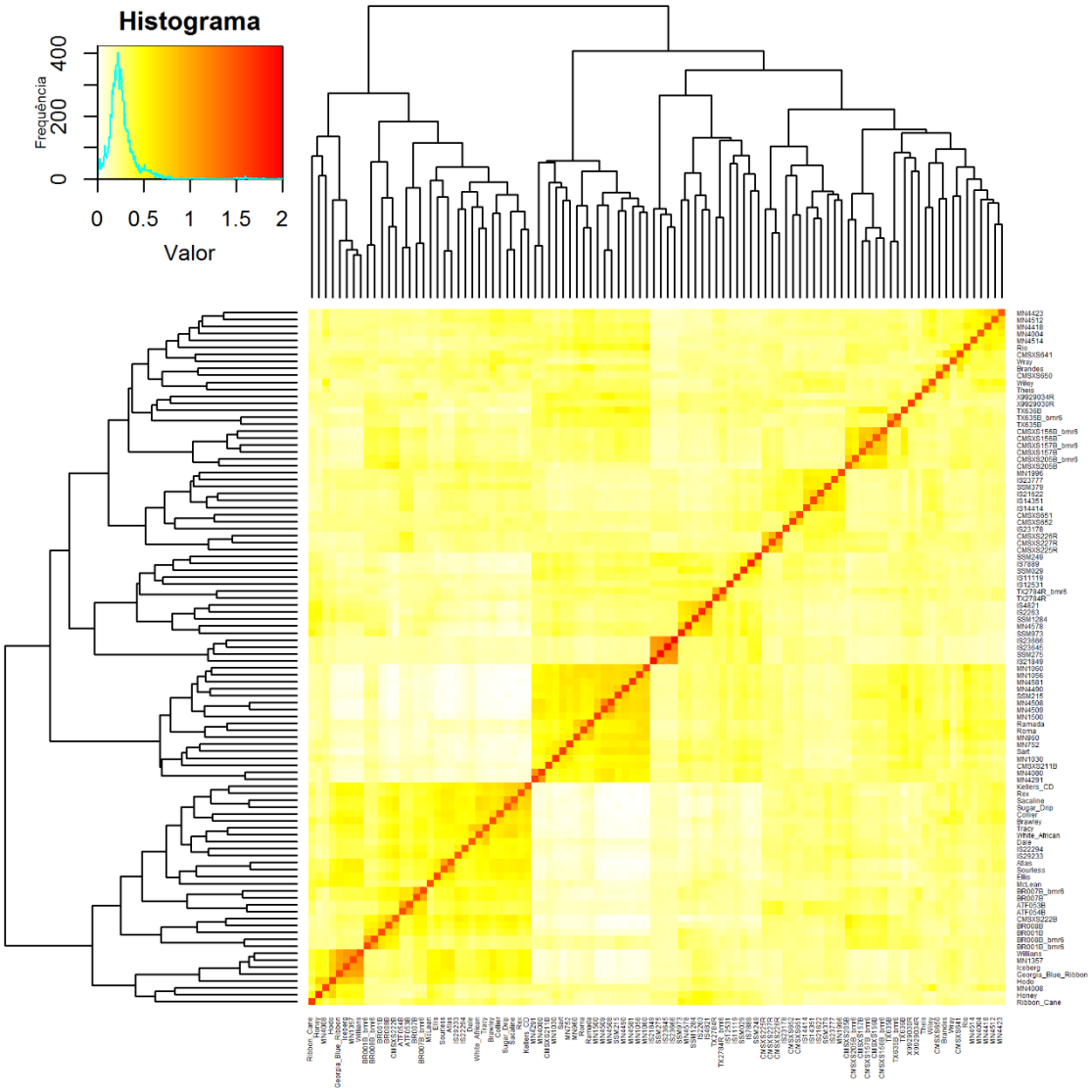

Figura 3 - Heatmap e dendrograma obtidos com base na matriz de parentesco entre os 100 genótipos que compõem o painel de diversidade genética de sorgo para a produção de bioenergia. O histograma (canto superior esquerdo) mostra a distribuição dos valores do coeficiente de coancestria em toda a matriz de parentesco.

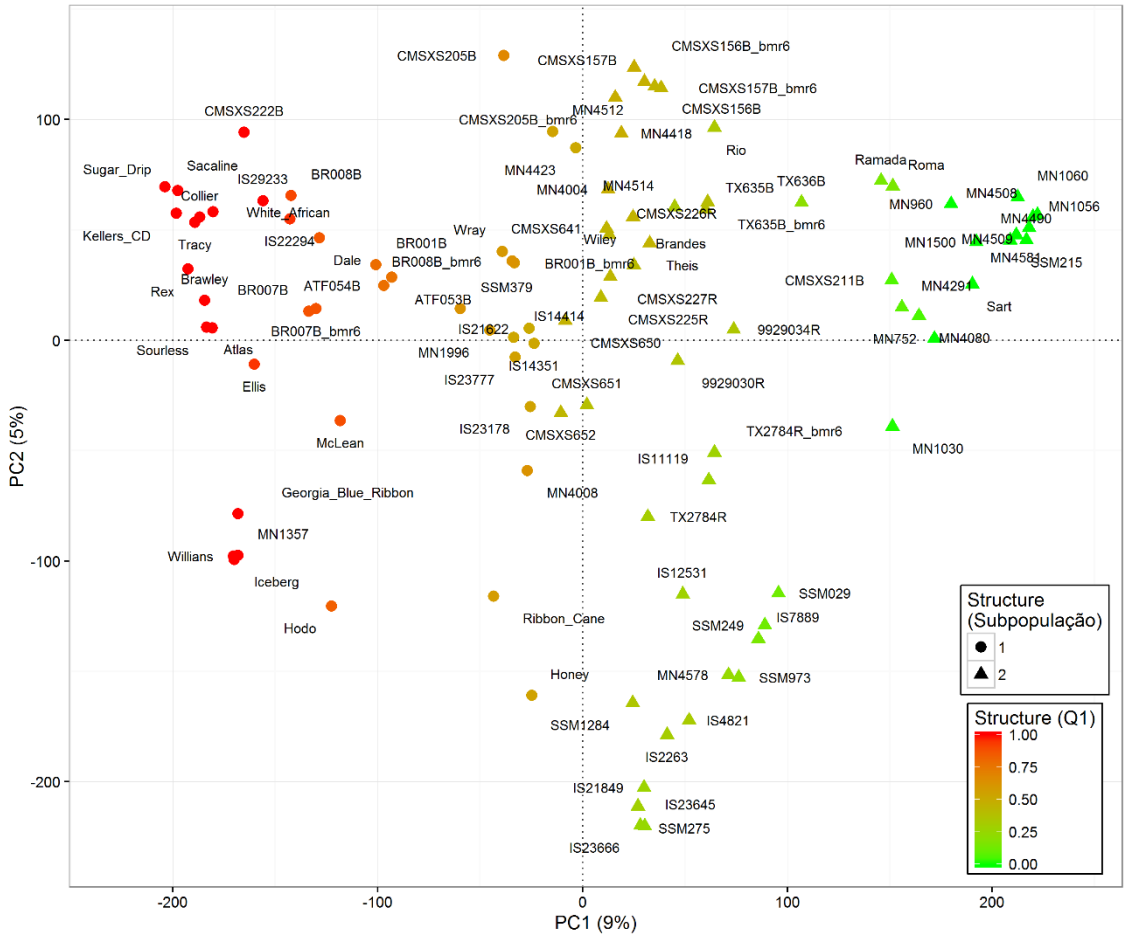

Figura 4 - Estrutura populacional de acordo com o software Structure (os círculos representam a subpopulação 1 e os triângulos a subpopulação 2) e respectivas proporções (gradiente de cores) de pertencer a subpopulação 1 (Q1), plotados sobre os dois primeiros eixos da análise de componentes principais (PCA) dos dados genotípicos do painel de diversidade genética de sorgo para produção de bioenergia.

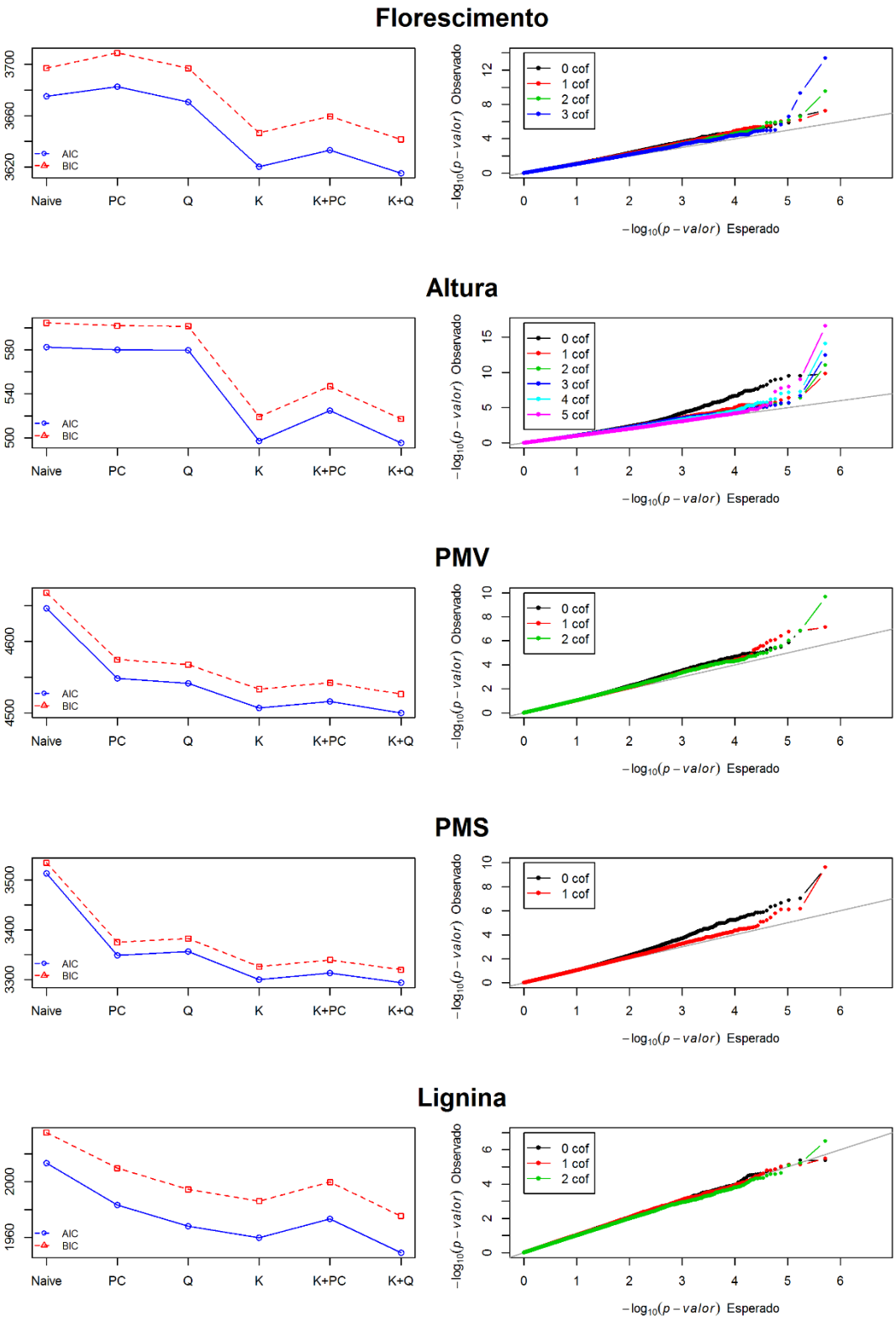

Figura 5 - Seleção de modelos (Naive, PC, Q, K, K+PC e K+Q) de acordo com os critérios AIC e BIC (menor valor, melhor modelo), na coluna da esquerda, e ajuste dos modelos selecionados (K+Q) considerando a adição de cofatores (SNPs) até alcançar o modelo otimizado, conforme o critério mBonf do MLMM, na coluna da direita, para cada um dos caracteres avaliados no painel de diversidade genética de sorgo para produção de bioenergia.

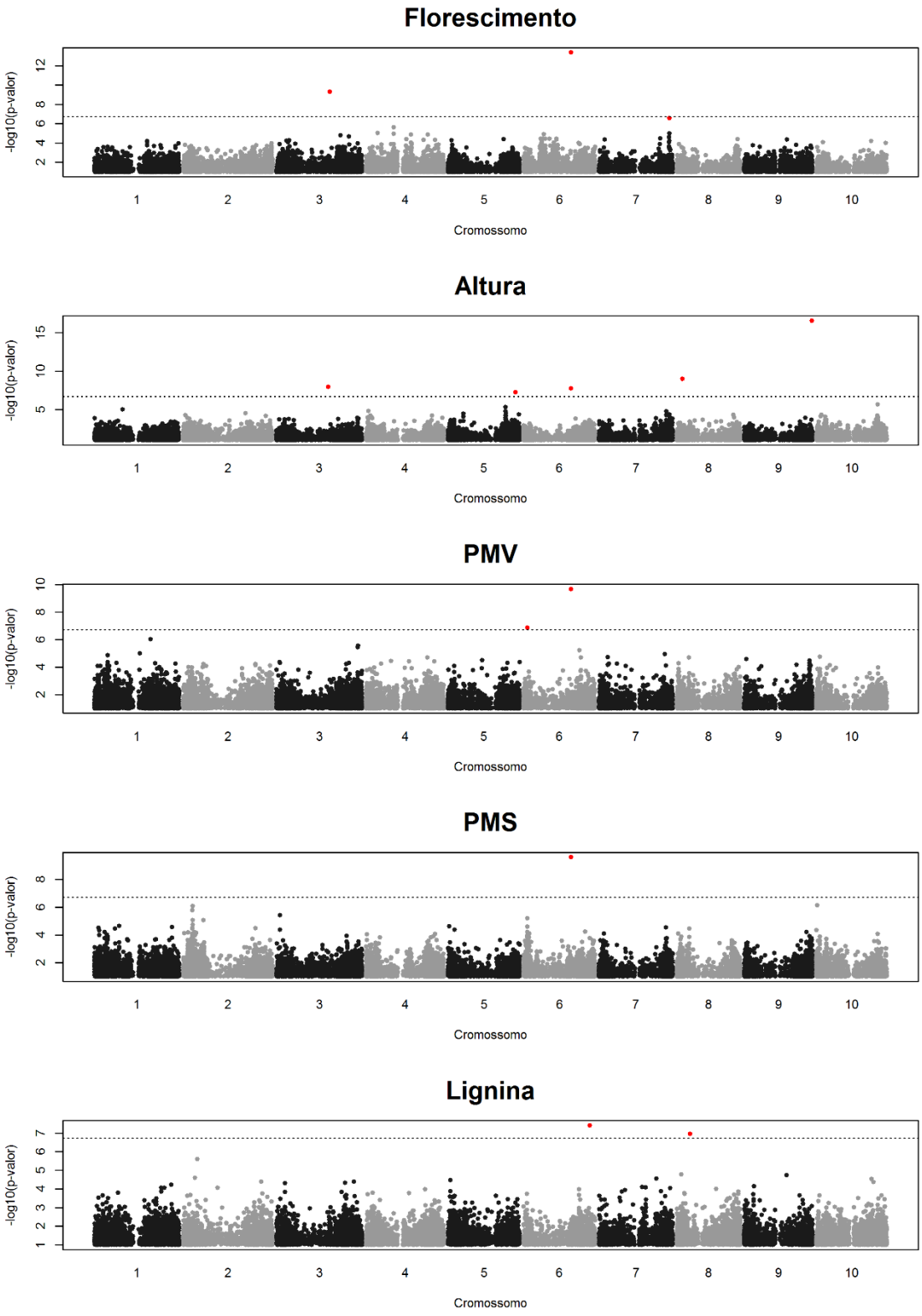

Figura 6 – Marcadores SNP significativos (pontos em vermelho) identificados por associação genômica ampla (GWAS) para florescimento, altura, PMV, PMS e lignina no modelo otimizado via MLMM, com valor de significância determinado pela correção de Bonferroni a 5% de probabilidade [$-\log_{10}(p\text{-valor})$ maior ou igual a 6,72], no painel de diversidade genética de sorgo para produção de bioenergia.

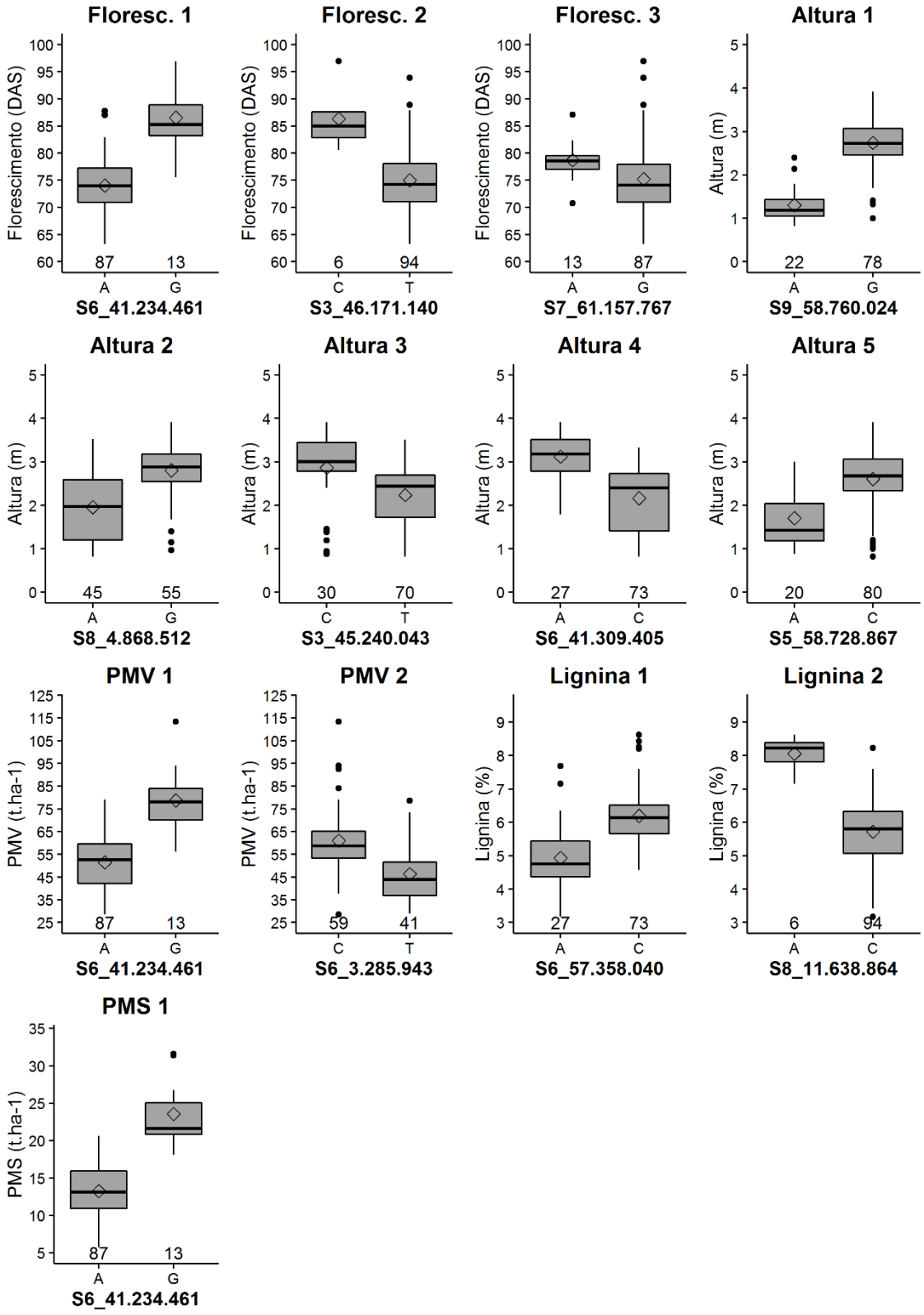

Figura 7 - Valores ajustados dos caracteres agroindustriais para cada um dos alelos dos SNPs significativos, conforme correção de Bonferroni, nos estudos de associação genômica ampla (GWAS) via MLMM do painel de diversidade de sorgo para produção de bioenergia. No eixo x são apresentadas as duas versões alélicas de cada SNP e suas respectivas frequências. Os losangos nos diagramas de caixa representam a média do caractere para cada alelo.

4. CAPÍTULO 3

ANÁLISE DE EXPRESSÃO DOS GENES ASSOCIADOS À SÍNTESE DE LIGNINA EM UM PAINEL DIVERSO DE SORGO

RESUMO

A composição bioquímica da biomassa vegetal é um dos fatores mais importantes para a produção de bioenergia. A lignina é um dos principais componentes da parede celular vegetal e interfere no processo de conversão da biomassa em combustíveis de segunda geração. No entanto, a lignina aumenta o poder calorífico da biomassa, o que é desejável para a cogeração de eletricidade. Genótipos de sorgo apresentam ampla variação no teor de lignina e podem fornecer matéria-prima de qualidade para ambas as finalidades. O objetivo deste trabalho foi avaliar a expressão de genes envolvidos na biossíntese de lignina em sorgo para a identificação de genes-alvo para o programa de melhoramento. Um painel diverso de sorgo composto por 100 genótipos foi avaliado para o teor de lignina para a identificação dos acessos com alta produtividade de biomassa e variados teores de lignina. Sessenta genótipos, trinta com maior e trinta com menor teor de lignina, foram selecionados para o ensaio de expressão gênica relacionado à síntese de lignina. A análise de expressão gênica foi feita por RT-qPCR pelo método de SybrGreen em um grupo de dezesseis genes de sorgo homólogos a genes de Arabidopsis previamente identificados como associados à biossíntese de lignina. Inicialmente, os dezesseis genes foram testados em oito genótipos contrastantes do painel, e uma análise de correlação entre a expressão e o teor de lignina foi utilizada para a identificação de relações significativas. Apenas o gene *HCT1*, que codifica para a enzima hidroxicinamoil transferase, apresentou correlação significativa ($p < 0,01$; $r = 0,87$) e, subsequentemente, a análise de expressão para esse gene foi investigada para os sessenta indivíduos previamente selecionados pelo teor de lignina. A expressão do gene *HCT1* apresentou relação linear significativa ($p < 0,01$; $r^2 = 0,58$) com o teor de lignina dos sessenta genótipos. O gene *HCT1* atua no início da síntese de lignina e de outros fenilpropanoides, como relatam trabalhos anteriores. O presente trabalho reforça a relação direta entre a expressão do gene *HCT* em plantas jovens e o teor de lignina no ponto de corte do sorgo. Desta forma, este gene pode ser utilizado para estudos mais aprofundados da biossíntese de lignina em sorgo, além de ser um alvo para seleção assistida por marcadores moleculares.

Palavras-chave: *Sorghum bicolor*; RT-qPCR; *HCT*, *hidroxicinamoil transferase*.

ABSTRACT

Plant biomass biochemical composition is one of the most important factors regarding bioenergy production. Lignin is a major component of plant cell walls and it can interfere with the process of biomass conversion to second generation ethanol. Nevertheless, lignin increases the calorific value of biomass, which is desirable for electricity cogeneration. Sorghum genotypes have a wide variation regarding lignin content and they can provide a quality feedstock for both end purposes. The aim of this work was to evaluate the expression of genes involved in lignin biosynthesis to identify potential gene targets for sorghum breeding. A diversity panel of 100 sorghum genotypes was evaluated for lignin content to identify genotypes showing high biomass production and varying levels of lignin. Sixty genotypes, thirty with higher and thirty with lower lignin content were selected for gene expression analysis related to lignin biosynthesis. Lignin gene expression was evaluated by RT-qPCR using SybrGreen in a set of sixteen sorghum genes homologous to previously identified lignin genes in Arabidopsis. Initially, these sixteen genes were analyzed in eight contrasting genotypes, and to identify significant relationships, correlation analysis was applied to the expression and the lignin content data. Only the *HCT1* gene, which codes for a hydroxycinnamoyl transferase, showed a significant correlation ($p < 0.01$; $r = 0.87$) to lignin content. Subsequently, *HCT1* expression analysis was investigated in the 60 previously selected contrasting genotypes. The results determined that *HCT1* showed significant linear relationship ($p < 0.01$, $r^2 = 0.58$) with the lignin content. The enzyme hydroxycinnamoyl transferase (*HCT*) acts at the beginning of lignin and other phenylpropanoid pathways, as reported in previous studies. The present work determines a relationship between *HCT1* gene expression in young plants and high lignin levels at harvest time. This gene may be used for future studies on lignin biosynthesis in sorghum, in addition to become a target for marker assisted selection.

Keywords: *Sorghum bicolor*; RT-qPCR; *HCT*, hidroxicinamoil transferase.

4.1. INTRODUÇÃO

A lignina é um polímero aromático complexo que é depositado nas paredes celulares secundárias de todas as plantas vasculares, e proporciona rigidez e resistência aos tecidos vegetais. A lignificação foi uma das adaptações críticas que permitiram a colonização dos ambientes terrestres pelas plantas vasculares há cerca de 500 milhões de anos (Weng & Chapple 2010). As propriedades físicas e químicas da lignina conferem à planta estrutura de defesa contra pragas e doenças e contra o tombamento. Além de conceder aos vasos do xilema superfície hidrofóbica que permite o transporte de água e nutrientes nos vegetais superiores (Koch *et al*. 2004).

No contexto bioenergético a lignina é um importante componente da biomassa. Do ponto de vista termoquímico, especificamente na combustão direta, o alto teor de lignina aumenta o valor energético dos materiais vegetais, uma vez que a lignina contém menos oxigênio do que a celulose e hemicelulose e tem um poder calorífico de 22 a 24 kJ g^{-1}, que é 30 a 50 por cento maior do que a de outros componentes da parede celular. Em contraste, a lignina é inibidora de processos de conversão biológica, tais como a fermentação para produção de etanol segunda geração (Frei 2013).

As matérias-primas para a produção de etanol de segunda geração são os polissacáridos da parede celular, mais especificamente a celulose e a hemicelulose, que são convertidos em açúcares simples, por hidrólise da biomassa. A lignina dificulta a extração destes componentes, sendo necessário métodos de pré-tratamento para neutralizá-la ou removê-la (Naik *et al.* 2010). Além do alto custo energético, a degradação da lignina produz moléculas que podem interferir posteriormente nas etapas de sacarificação e fermentação (Sun *et al.* 2014). Assim, existe um grande interesse na possibilidade da manipulação genética da biossíntese de lignina para melhorar a extração e processamento dos polissacarídeos, ou aumentar o poder calorífico da biomassa.

A lignina é composta pela combinação dos monolignóis p-cumaril álcool (monolignol H), coniferílico álcool (monolignol G), e álcool sinapyl (monolignol S), e já existe um consenso sobre enzimas envolvidas na biossíntese dos monolignóis (Boerjan, *et al.* 2003; Vanholme *et al.* 2010). A síntese das

subunidades da lignina, os monolignóis, é feita a partir da fenilalanina, e requer desaminação, hidroxilação em uma, duas, ou três posições do anel aromático e a metilação de um ou dois destes grupos hidroxila. Assim como duas reduções sucessivas do ácido carboxílico, primeiramente em um aldeído e, em seguida, em um álcool.

O processo de lignificação é iniciado pela enzima Fenilalanina amônia liase (PAL). As enzimas cinamato 4-hidroxilase (C4H), 4-hidroxicinamato 3-hidroxilase (C3H) e ferulato 5-hidroxilase (F5H) são três monoxigenases citocromo P450-dependente diferentes. Cafeoil CoA O-metiltransferase (CCoAOMT) e cafeato/5-hidroxiferulato O-metiltransferase (COMT) são metiltransferases, e hidroxicinamoil CoA redutase (CCR) e cinamil álcool desidrogenase (CAD) são oxidoredutases. A 4-cumarato CoA ligase (4CL) é uma enzima ATP-dependente que catalisa a síntese de p-coumaroil-CoA, que por sua vez pode ser catalizada em p-coumaroil shikimato pela enzima hidroxicinamoil transferase (HCT) (Bonawitz & Chapple, 2010). A via de síntese de p-cumaril álcool (monolignol H) requer um subconjunto de apenas cinco destas enzimas (PAL, C4H, 4CL, CCR e CAD), enquanto para a síntese de coniferílico álcool (monolignol G) mais três enzimas são necessárias (HCT, C3H e CCoAOMT). Já o álcool sinapyl (monolignol S) exige todas as dez enzimas, incluindo F5H e COMT.

A biossíntese de monolignóis está bem elucidada em relação ao transporte e a sua ativação na parede celular. Desta forma, tem sido o alvo mais frequente nas tentativas de redução ou alteração da lignina. Os genes e enzimas necessárias para a biosíntese de monolignóis já foram identificados em várias espécies de plantas, incluindo *Arabidopsis*, alfafa, álamo, arroz, tabaco e eucalipto (Dixon *et al*. 2001), e as características mais relevantes da via parecem ser conservadas entre as espécies (Bonawitz & Chapple 2013).

Além de variar entre espécies de plantas, o teor de lignina varia nos tecidos e nos estágios de desenvolvimento vegetal. O processo de lignificação é iniciado para compor as paredes celulares primárias dos elementos do xilema no início da formação das paredes celulares secundárias. Em outros tecidos, é depositada posteriormente durante a fase vegetativa e esse período pode variar ao longo de meses (Wang *et al.* 2013). Em sorgo, Saballos *et al.* (2012) consideraram plantas coletadas entre 35 e 50 dias após a semeadura para

ensaios de expressão de genes envolvidos nesse processo, uma vez que os tecidos ainda se mostravam pouco lignificados.

Estudos da expressão diferencial de genes associados a lignina foram realizados em sorgo, principalmente em genótipos mutantes de nervura marrom (do inglês, *brown midrib* - bmr) (Bout & Vermerris 2003; Sattler *et al.* 2010; Saballos *et al.* 2012). Entretanto, se faz necessário o estudo da expressão diferencial dos diversos genes envolvidos na biossíntese de lignina que ocorre naturalmente devido a ampla diversidade genética do sorgo. Nesse contexto, o objetivo desse trabalho foi identificar genes candidatos relacionados ao processo de lignificação, a partir da quantificação via RT-qPCR, e da correlação entre a expressão desses genes em plantas jovens de sorgo com o teor de lignina predito na planta em ponto de corte.

4.2. MATERIAL E MÉTODOS

4.2.1. Material Genético e Delineamento Experimental

Um painel de diversidade genética de sorgo, composto por 100 linhagens, foi utilizado para quantificar o teor de lignina visando a produção de bioenergia. Os materiais do painel são oriundos da coleção núcleo do CIRAD, ICRISAT e do Banco de Germoplasma da Embrapa Milho e Sorgo. Os genótipos foram avaliados em duas safras em campo, a primeira semeada em fevereiro de 2011 e a segunda em janeiro de 2012, em área experimental pertencente à Embrapa Milho e Sorgo, em Sete Lagoas, MG, Brasil. As parcelas experimentais foram compostas por duas linhas de 5 m de comprimento, espaçadas 0,70 m, com densidade de nove plantas por metro linear. O delineamento foi em látice 10x10 com três repetições. Amostras de colmos de cada parcela foram utilizadas para determinar o teor de lignina dos materiais.

4.2.2. Determinação do Teor de Lignina em Detergente Ácido (LDA) e Análise Fenotípica

Após maturidade fisiológica dos grãos, aproximadamente 120 dias após a semeadura, amostras de cinco colmos sem panículas foram colhidas de cada parcela, trituradas e secas em estufa a 65 ºC. O método de quantificação da lignina dos colmos utilizado foi o de lignina em detergente ácido (LDA), conforme Van Soest (1991). Este método consiste na extração sequencial dos diferentes componentes da biomassa por etapas sucessivas de digestão química até que a lignina residual seja isolada.

Assim, aproximadamente 500 mg de cada amostra homogeneizada foram acondicionados em bolsas filtro e aquecidos em solução de detergente neutro, de detergente ácido e de ácido sulfúrico 72% m/v, com separação dos resíduos por filtração e secagem a 105 ºC entre as etapas. Os produtos finais foram pesados para a determinação da porcentagem de lignina em detergente ácido.

Os dados fenotípicos do teor de lignina foram analisados considerado o seguinte modelo:

$$y_{ijkl} = \mu + a_l + r_{k(l)} + b_{j(kl)} + g_{im} + \underline{\varepsilon}_{ijkl}$$

em que: y_{ijkm} é o valor fenotípico observado para o indivíduo i no bloco j, repetição k e safra $_l$; μ é a média geral; a_l é o efeito fixo da l-ésima safra (l = 1, ..., L; L = 2); $r_{k(l)}$ é o efeito fixo da k-ésima repetição (k = 1, . . . , K; K = 3) na safra l; $b_{j(kl)}$ é o efeito fixo do j-ésimo bloco (j = 1, . . . , J ; J = 10) na repetição k e na safra l; g_{il} é o efeito fixo do i-ésimo genótipo (i = 1, . . . , I ; I =100) na safra l; e ε_{ijkl} é o resíduo.

4.2.3. Seleção de Materiais Contrastantes para a Análise de Expressão Gênica

Para a análise de expressão dos genes relacionados à síntese de lignina, 60 genótipos foram selecionados após a análise fenotípica da lignina das 100 linhagens avaliadas. Foram selecionados os 30 genótipos com maior e os 30 com menor teor de lignina. Os 60 genótipos foram semeados em casa de vegetação, com controle de temperatura e umidade. Cada parcela experimental foi composta por um vaso com cinco plantas e o delineamento utilizado foi em blocos casualizados, com três repetições. Os colmos das cinco plantas da parcela foram utilizados para extração de RNA. A coleta foi realizada 38 dias após semeadura, com os tecidos ainda jovens e pouco lignificados (quando os genes associados à síntese de lignina estão mais ativos). As amostras foram acondicionadas imediatamente após a coleta, e durante o processo de maceração, em nitrogênio líquido. Após a maceração de toda a amostra, uma subamostra foi coletada e mantida em ultrafreezer a -80 ºC até a extração do RNA.

4.2.4. Identificação de Genes Relacionados à Via de Biossíntese da Lignina em Sorgo e Desenho de Primers Específicos

Os genes envolvidos na via biossintética da lignina foram identificados em sorgo utilizando-se sequências homólogas dos genes correspondentes às proteínas relatadas na literatura para *Arabidopsis* (Ehlting *et al.* 2005). Utilizando-se o programa BLAST (Altschul *et al.* 1997), estas sequências homólogas foram comparadas à sequência genômica de sorgo no Phytozome (Goodstein *et al*. 2012), e também aos bancos contendo sequências ESTs de sorgo, como NCBI (Geer *et al*. 2010) e GRAMENE (Youens-Clark *et al*. 2010), para a correta predição de intros/exons dos genes de interesse.

Desta forma, 16 pares de primers específicos para genes que codificam enzimas da via de biossíntese de lignina (Tabela 1) foram desenhados com o programa ‘Primer3 v.0.4.0’ desenvolvido por Rozen & Skaletsky (2000), selecionando-se primers com Tm em torno de 60 ºC. Para diferenciar entre a amplificação resultante do cDNA sintetizado ou DNA genômico residual, os pares de primers gene-específicos foram desenhados de forma que

flanqueassem introns, sempre que possível. Também foi realizada busca no genoma do sorgo com cada primer a fim de se determinar a especificidade do mesmo para o respectivo gene.

4.2.5. Extração de RNA e Análise de Expressão Gênica

O RNA total de cada genótipo foi extraído por meio do RNeasy Mini Kit (Qiagen). Após extração e tratamento com DNAse I, o RNA foi quantificado em espectrofotômetro NanoDrop ND-1000 (Thermo Scientific), por leitura em 260 nm, e qualificado em eletroforese em gel de agarose, pela integridade das bandas de RNA ribossomal.

A primeira fita de cDNA foi sintetizada utilizando-se o kit '*High Capacity cDNA Reverse Transcription*' (Applied Biosystems). Cada amostra de cDNA foi diluída 10 vezes antes de amplificação com primers específicos. A expressão gênica foi avaliada por meio de RT-PCR quantitativo em tempo real no equipamento ABI Prism 7500 (Applied Biosystems).

As condições de amplificação foram testadas para cada par de primer, realizando-se o teste de eficiência, no qual apenas primers com eficiência maior que 90% foram selecionados para etapas posteriores. Para efeitos de normalização, a expressão endógena foi quantificada com o controle glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) detalhado na Tabela 1.

A quantificação dos cDNAs sintetizados foi feita pelo método SYBR Green de detecção (Applied Biosystems). As mudanças em fluorescência do *SYBR Green I dye* foram monitoradas em cada ciclo pelo programa do sistema ABI 5700 e o ciclo '*threshold*' (CT) para cada reação foi calculado. Os tratamentos foram analisados em triplicatas. Todas as reações foram submetidas às mesmas condições de análise e normalizadas pelo sinal do corante de referência passiva ROX para correção de flutuações na leitura ao longo da reação.

Os valores do CT dos genes-alvo foram subtraídos do valor do CT do normalizador, o que resultou no valor de ΔCT. Os valores de ΔCT dos genes-alvo foram subtraídos do valor do ΔCT do calibrador (controle), e então, calculado o valor de $\Delta\Delta$CT. A determinação dos níveis de expressão dos genes-alvo foi realizada pela quantificação relativa (*relative quantification* - RQ),

utilizando-se a equação $RQ=2^{-\Delta\Delta CT}$ (Livak & Schmittgen 2001). Por fim, foi calculada a média de RQ das triplicatas.

4.2.6. Análises de Correlação e Regressão entre a Expressão Gênica e o Teor de Lignina

Inicialmente oito genótipos contrastantes, quatro com baixo teor e quatro com alto teor de lignina, foram selecionados para o teste de correlação entre o teor de lignina predito dos genótipos na fase de maturação da planta e dos valores de expressão relativa dos 16 genes da Tabela 1. O objetivo foi simplificar o teste da relação direta entre a expressão gênica e o teor de lignina utilizando apenas os valores extremos dos materiais, reduzindo-se o número de análises iniciais. Após a análise de expressão utilizando-se os oito genótipos e os 16 genes-alvo, apenas os primers que apresentaram correlação significativa com o teor de lignina foram utilizados para análise de expressão nos sessenta genótipos contrastantes. As análises de correlação e regressão entre os valores de RQ e as médias do teor de lignina desses 60 genótipos avaliados foram realizadas no programa R, por meio dos pacotes agricolae e ggplot2 (R Core Team 2015).

4.3. RESULTADOS

Baseando-se nas médias ajustadas dos 100 materiais que compõem o painel diverso de sorgo, 30 genótipos como maior e 30 com menor teor de lignina em detergente ácido (LDA) foram selecionados. As médias ajustadas para o teor

de lignina dos 60 materiais contrastantes selecionados para as análises de expressão gênica podem ser visualizadas na Figura 1. As médias ajustadas para o teor de lignina variaram entre 3,38% (BR506) a 8,99% (IS23666).

Os resultados da correlação entre os valores de RQ dos 16 genes testados e o teor de lignina predito dos oito genótipos contratantes, quatro com alto e quatro com baixo teor de lignina, estão apresentados na figura 2. Nessa análise inicial, apenas o gene *HCT1* apresentou correlação significativa (*p*-valor < 0,01) e de alta magnitude (r = 0,87) com o teor de lignina. Desta forma, o *HCT1* foi selecionado para a análise de expressão utilizando-se os 30 genótipos com maior e os 30 com menor teor de lignina, para que todos os 60 materiais fossem considerados na análise de regressão entre o teor de lignina e o RQ (expressão gênica do *HCT1*).

A análise de correlação e a regressão linear entre níveis de expressão do gene *HCT1* e o teor de lignina para os 60 genótipos avaliados manteve-se significativa ($p < 1{,}33 \times 10^{-12}$) com coeficiente de correlação (r) igual a 0,76 e com coeficiente de determinação (r^2) igual a 0,58, ou seja, a variável expressão do gene *HCT1* explicou cerca de 58% da variabilidade observada para o teor de lignina nesses 60 genótipos (Figura 3). Como o teor de lignina é uma característica quantitativa e diversos outros fatores estão envolvidos no teor final, incluindo fatores ambientais, é normal que a equação não explique perfeitamente a relação entre os dados. Entretanto, a relação linear positiva é satisfatória e reforça a importância deste gene na via de biossíntese de lignina, bem como a presença de variação de sua expressão no painel diverso de sorgo utilizado.

4.4. DISCUSSÃO

A enzima HCT constitui um ponto chave no estudo da lignina pois separa as vias de síntese dos monolignóis G e S, do monolignol H. Paralelo a isso, serve como substrato para síntase de outros produtos da via fenilpropanoide, que incluem flavonoides, antocianinas, taninos, hidroxicinnamatos e os componentes fenólicos dos biopolímeros esporopolenina e suberina. (Bonawitz & Chapple 2013). A HCT foi recentemente caracterizada para o sorgo (Walker *et al*. 2013)

e catalisa a síntese do quinato-chiquimato e ésteres do ácido p-cumárico, que são os substratos do citocromo P450 3-hidroxilase (CYP98A3) (Schoch *et al.* 2001; Franke *et al.* 2002).

Como a simples redução da concentração total de lignina pode comprometer a estrutura da planta, o controle genético da HCT pode possibilitar a composição diferenciada da lignina, e reduzir a recalcitrância das fibras sem prejudicar o desenvolvimento vegetal. Segundo o trabalho de Hoffmann *et al.* (2004), o silenciamento do *HCT* em *Nicotiana benthamiana* além de reduzir 15% no teor de lignina, aumentou relativamente o monômero H (de 0,2% para 8%) sem afetar a proporção de monômeros G (aproximadamente 30% em todos os casos). Desta forma, os monômeros S foram reduzidos de, aproximadamente, 70% para 63%, em plantas *N. benthamiana HCT*-silenciadas. Além disso, os materiais foram testados quanto a suscetibilidade dos tecidos à celulase, e as plantas *HCT*-silenciadas apresentaram biomassa residual 70% menor que as plantas controle. Este fato evidencia uma maior eficiência da hidrólise em tecidos com concentração diferenciada de lignina.

No entanto, a intervenção ou bloqueio na síntese de monolignóis, com o silenciamento do *HCT*, pode resultar na hiperacumulação de intermediários da via ou seus derivados. O acúmulo destes substratos pode ser amplamente citotóxico ou interferir na sinalização celular, como demonstram estudos com genótipos mutantes. A própria coloração púrpura de folhas de plantas silenciadas sugeriru o acúmulo de flavonoides e antocianinas (Hoffmann *et al.* 2004). Li *et al.* (2010) relataram que o silenciamento do *HCT* em *Arabidopsis thaliana* provocou uma acumulação de resíduos de p-hidroxifenil e o decréscimo do conteúdo de guaiacil e siringil, porém gerando plantas raquíticas.

Assim, a identificação de variações alélicas do gene *HCT1* no painel de diversidade, que confiram modificações no teor e/ou composição de lignina sem os efeitos colaterais acarretados pelo silenciamento do gene mencionados acima, tem grande potencial para uso no programa de melhoramento de sorgo biomassa. A partir da identificação futura de polimorfismos do tipo SNP (*Single Nucleotide Polimorfismo*) no gene *HCT1*, em materiais contrastantes para lignina, será possível desenvolver e validar marcadores nos demais genótipos do painel, a fim de se desenvolver métodos de seleção assistida para introgressão do alelo superior em linhagens elite de sorgo biomassa.

4.5. REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17):389-402.

BOERJAN, W.; RALPH, J.; BAUCHER, M. (2003). Lignin biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 54:519-546.

BONAWITZ, N.D.; CHAPPLE, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. In: *Annual Review of Genetics*. Campbell, A.; Lichten, M.; Schupbach, G. Eds., 44:337–363.

BONAWITZ, N.D.; CHAPPLE, C. (2013). Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty? *Current Opinion in Biotechnology* 24:336–343.

BOUT, S.; VERMERRIS, W. (2003). A candidate-gene approach to clone the sorghum Brown midrib gene encoding caffeic acid O-methyltransferase. Mol Genet Genom269: 205–214.

DIXON, R.A.; CHEN, F.; GUO, D.; PARVATHI, K. (2001). The biosynthesis of monolignols: a “metabolic grid” or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 57:1069–84.

EHLTING, J.; MATTHEUS, N.; AESCHLIMAN, D.S.; LI, E.Y.; HAMBERGER, B.; CULLIS, I.F.; ZHUANG, J.; KANEDA, M.; MANSFIELD, S.D.; SAMUELS, L.; RITLAND, K.; ELLIS, B.E.; BOHLMANN, J.; DOUGLAS, C.J. (2005). Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. Plant Journal 42(5):618-640.

FRANKE, R.; HEMM, M.R.; DENAULT, J.W.; RUEGGER, M.O.; HUMPHREYS, J.M.; CHAPPLE, C. (2002). Changes in secondary metabolism and deposition of an unusual lignin in the ref8mutant of Arabidopsis. *Plant Journal* 30:47–59.

FREI, M. (2013). Lignin: Characterization of a Multifaceted Crop Component. *The Scientific World Journal*.

GEER, L.Y.; MARCHLER-BAUER, A.; GEER, R.C.; HAN, L.; HE, J.; HE, S.; LIU, C.; SHI, W.; BRYANT, S.H. (2010). The NCBI BioSystems database. *Nucleic Acids Res*. 38.

GOODSTEIN, D.M.; SHU, S.; HOWSON, R. et al (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40:D1178–D1186.

HOFFMANN, L.; BESSEAU, S.; MARTZ, F.; GEOFFROY, P.; RITZENTHALER, C.; MEYER, D.; LAPIERRE, C.; POLLET, B.; LEGRAND, M. (2004). Silencing of Hydroxycinnamoyl-Coenzyme A Shikimate/Quinate Hydroxycinnamoyltransferase Affects Phenylpropanoid Biosynthesis. *The Plant Cell* 16:1446–1465.

KOCH, G.W.; SILLETT, S.C.; JENNINGS, G.M.; DAVIS, S.D. (2004). The limits to tree height. *Nat. Biotechnol*. 428:851-854.

LI, X.; BONAWITZ, N.D.; WENG, J.K.; CHAPPLE, C. (2010). The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thalianais* independent of flavonoids. *Plant Cell* 22:1620–1632.

LIVAK, K.J.; SCHMITTGEN, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *methods*,*25*(4), 402-408.

NAIK, S.N.; GOUD, V.V.; ROUT, P.K.; DALAI, A.K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* 14:578–597.

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

ROZEN, S.; SKALETSKY, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In: KRAWETZ, S.; MISENER, S. (Ed.). Bioinformatics methods and protocols: methods in molecular biology. Totowa: Humana Press 365-386.

SABALLOS, A.; SATTLER, S.E.; SANCHEZ, E.; FOSTER, T.P.; XIN, Z.; KANG, C.H.; PEDERSEN, J.F.; VERMERRIS, W. (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate: Coenzyme A ligase involved in lignin biosynthesis in sorghum. *Plant J.* 70: 813–830.

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

SCHOCH, G.; GOEPFERT, S.; MORANT, M.; HEHN, A.; MEYER, D.; ULLMANN, P.; WERCK-REICHHART, D. (2001). CYP98A3 from Arabidopsis thaliana is a 39-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J. Biol. Chem.* 276:36566–36574.

SUN, S.; WEN, J.; MA, M.; SUN, R. (2014). Structural Elucidation of Sorghum Lignins from an Integrated Biorefinery Process Based on Hydrothermal and Alkaline Treatments. *J. Agric. Food Chem*. 62:8120−8128.

VAN OVERBEEK, J.; BLONDEAU, R.; HORNE, V. (1951). Trans-cinnamic acid as an anti-auxin. *Am J Bot* 38:589-595.

VAN SOEST, P. V., ROBERTSON, J. B., & LEWIS, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of dairy science*, *74*(10), 3583-3597.VANHOLME, R.; DEMEDTS, B.; MORREEL, K.; RALPH, J.; BOERJAN, W. (2010). Lignin biosynthesis and structure. *Plant Physiology* 153(3):895–905.

WALKER, A.M.; HAYES, R.P.; YOUN, B.; VERMERRIS, W.; SATTLER, S.E.; KANG, C. (2013). Elucidation of the Structure and Reaction Mechanism of Sorghum Hydroxycinnamoyltransferase and Its Structural Relationship to Other Coenzyme A-Dependent Transferases and Synthases. *Plant Physiology* 162:640–651.

WANG, Y.; CHANTREAU, M.; SIBOUT, R.; HOWKINS, S. (2013) Plant cell wall lignification and monolignol metabolism.Front Plant Sci4: 220.

WENG, J.; CHAPPLE, C. (2010). The origin and evolution of lignin Biosynthesis. *New Phytologist* 187: 273–285.

YOUENS-CLARK, K.; BUCKLER, E.; CASSTEVENS, T.; CHEN, C.; DECLERCK, G.; DERWENT, P.; DHARMAWARDHANA, P.; JAISWAL, P.; KERSEY, P.; KARTHIKEYAN, A.S.; LU, J.; MCCOUCH, S.R.; REN, L.; SPOONER, W.; STEIN, J.C.; THOMASON, J.; WEI, S.; WARE, D. (2011). Gramene database in 2010: updates and extensions. *Nucleic Acids Research* 39: D1085-94.

TABELAS

Tabela1 - Genes relacionados à síntese de lignina identificados no genoma de sorgo por homologia de sequência e seus respectivos primers, desenhados para utilização no ensaio de expressão gênica de 60 genótipos contrastantes quanto ao teor de lignina do painel de diversidade genética de sorgo para produção de bioenergia.

Enzima	Gene em *Arabidopsis*	Homólogo no sorgo	Fragmento genômico (pb)	Amplicon (pb)	Primer-F	Primer-R
4CL1	At1g51680	Sobic.004G272700	88	88	GTGTTCTACAAGAGGCTACACAAGGT	CGCGTAGCTCTCTCCTCAGAAT
4CL2	At1g51680	Sobic.004G062500	70	70	GTTCCAAGCTTCCCGACATC	CTCATCTTCCCGAAGCAGTAGGT
C3H1	At2g40890	Sobic.009G181800	626	187	CGCTTGACAATGAAGATCATT	CCCACTCAACGGAGATGACT
C3H2	At2g40890	Sobic.003G327800	1478	104	TGGTCGAGTCCGTCTACAAG	CCGCGTGATGTTGTTGAA
C4H1	At2g30490	Sobic.003G337400	70	70	CGACCACTGGCGCAAGA	GTACTGCTGCACCACCTTGTTG
C4H2	At2g30490	Sobic.002G126600	1637	102	GGACAACTTCGTCCAGGAAC	TTGATCTCGCCTTTCCTCTC
CAD1	At4g34230	Sobic.004G071000	81	81	GAGGTGCTCCAGTTCTGCG	CAGCGCCTCGTTCACGTAC
CAD2	At4g34230	Sobic.004G071000	736	91	ACCTCGGGGCTTCAAAGTA	ACGCCGTACTTGCTCACCT
CCoAOMT1	At4g34050	Sobic.010G052200	237	110	GCCTGCTCAAGAGCGACGACC	CCATGGGTGCTTGGCGGTGA
CCoAOMT2	At4g34050	Sobic.007G218800	55	55	CTCGGCGGCGATCAAG	CTCGACGCGCTCATCCTTA
CCR1	At1g15950	Sobic.007G141200	2971	88	CATCCTCGCCAAGCTCTTC	CGAGAACTTGTACGGCTGCT
COMT2	At5g54160	Sobic.007G047300	80	80	AACAAGGCGTACGGGATGAC	TTCATGCCCTCGTTGAACAC
F5H1	At4g36220	Sobic.001G196300	2806	94	CGACAACATCAAGGCCATC	GTGCATCATCTCCGCCATC
HCT1	At2g40890	Sobic.006G136800	3224	102	ATCTCGGCCTTCCCTCTACT	AGACATGCCATCCGCTACAT
HCT2	At2g40890	Sobic.004G212300	4049	104	TCGACTTCTCCGACGACAC	GGCTACGTGGTGCTGCAT
PAL1	At2g37040	Sobic.004G220700	104	104	TCATGTTTGCGCAGTTCTCT	CTTGAAGCCGTAGTCCAAGC
GAPDH	-	Sobic.004G205100	151	66	TTGAGGGTCTGATGACCACTGT	CCGAGGGCCCATCAACA

FIGURAS

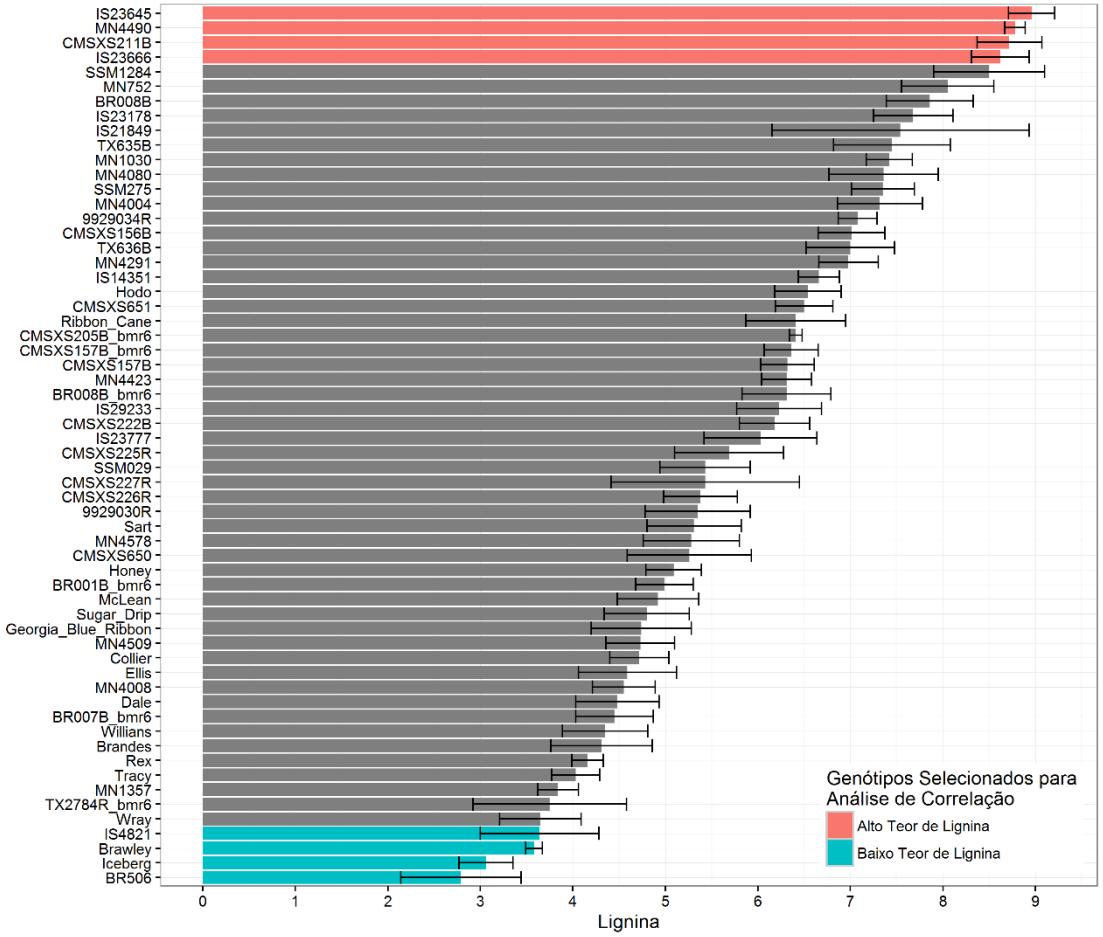

Figura 1 – Médias ajustadas para o teor de lignina de 60 materiais do painel de diversidade genética de sorgo para a produção de bioenergia selecionados para a análise de expressão gênica. As barras em preto representam o erro padrão da média. Quatro genótipos com maior teor (em vermelho) e quatro genótipos com menor teor (em azul) de lignina foram previamente selecionados para a análise de correlação com a expressão gênica. Posteriormente, os genes significativos na análise de correlação foram utilizados para a análise de regressão entre a expressão gênica e o teor de lignina dos 60 materiais.

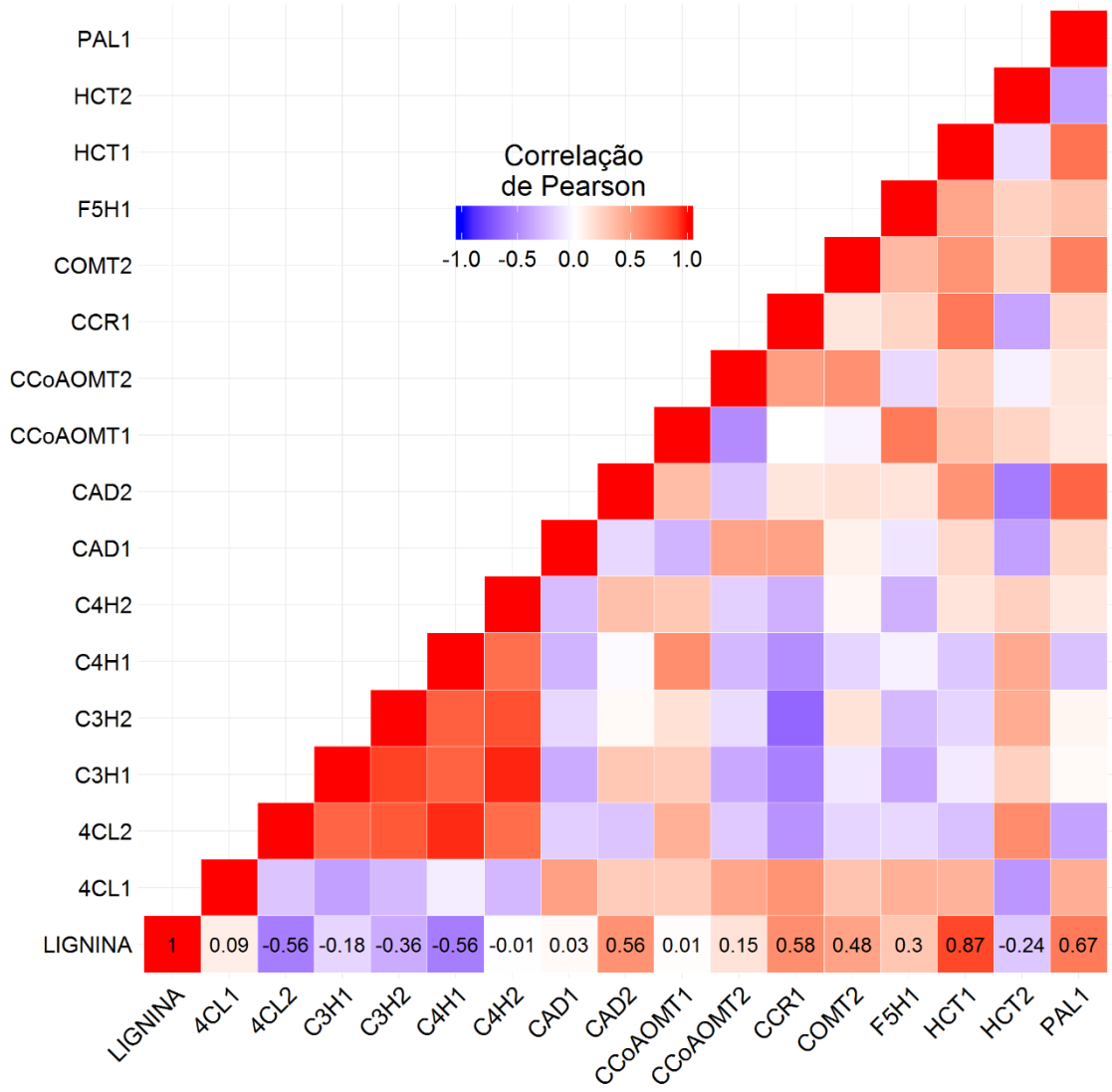

Figura 2 - Matriz de correlação entre a expressão gênica (RQ) de 16 genes da via de biossíntese de lignina, avaliados em sorgo 38 dias após semeadura, e o teor de lignina predito dos 8 genótipos contrastantes, avaliados no ponto de maturação fisiológica dos grãos.

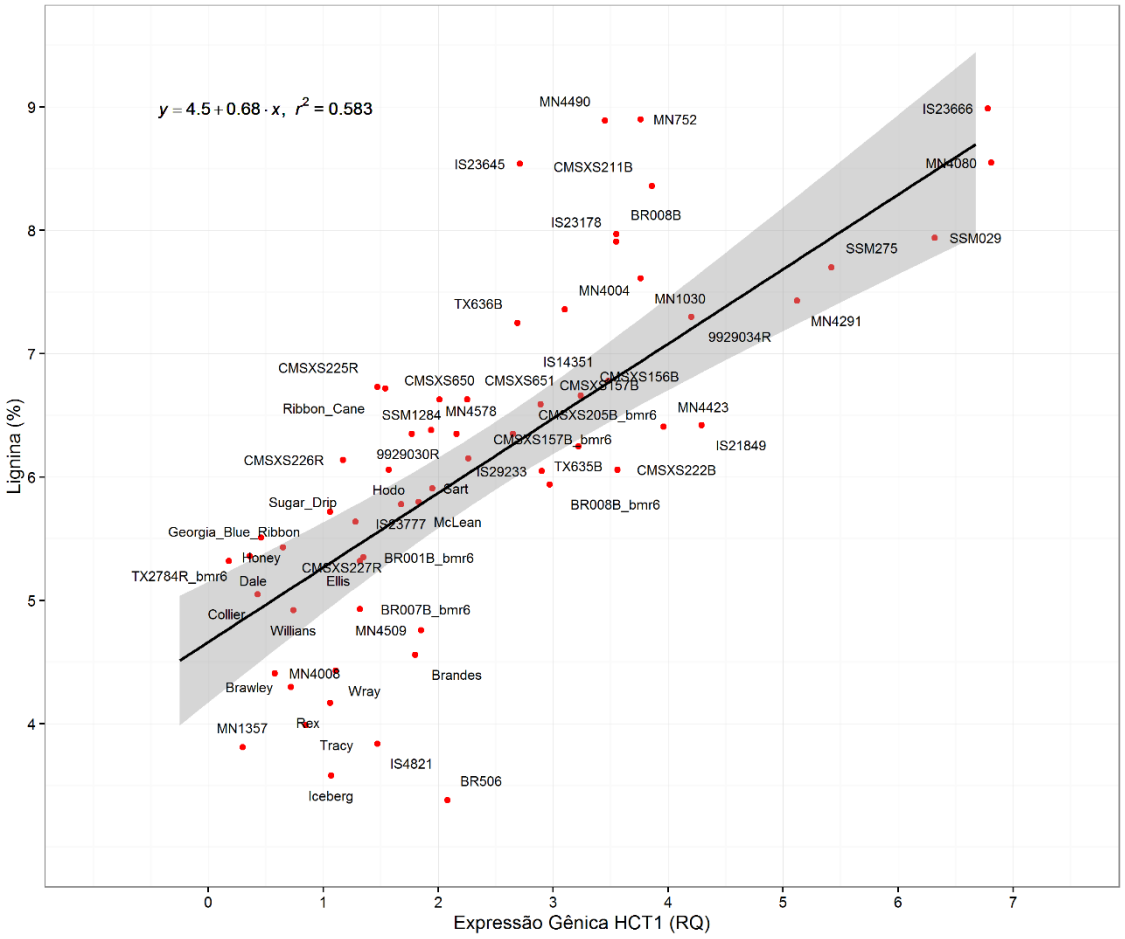

Figura 3 - Regressão entre a expressão gênica (RQ) do *HCT1* avaliado em sorgo com 38 dias após semeadura e o teor de lignina predito dos 60 genótipos contrastantes, 30 com maior e 30 com menor teor de lignina, avaliados no ponto de maturação fisiológica dos grãos. A relação entre os fatores é dada por: $y = 4{,}5 + 0{,}68 \times x$. A variável expressão do gene *HCT1* explicou cerca de 58% (r^2=0,58) da variabilidade observada para o teor de lignina. O gradiente de coloração cinza representa o intervalo de confiança de 95%.

5. CONCLUSÃO GERAL

As técnicas de mapeamento de QTLs utilizadas no presente estudo permitiram a identificação de diversas regiões relacionadas ao controle genético dos caracteres agroindustriais avaliados no sorgo. Como altura e florescimento já foram amplamente estudados na cultura do sorgo, os resultados de literatura relativos a esses caracteres serviram como referência para comparações. Foi possível concluir que os resultados do presente trabalho corroboram os resultados anteriores, o que reforça os métodos utilizados e os resultados das demais variáveis analisadas.

No mapeamento de QTLs utilizando-se a população RILs, além de ambos os genitores serem sacarinos, a utilização da técnica GBS e do método de mapeamento MTMIM possibilitou a identificação de novos QTLs, principalmente para Brix e Pol, afora os já relatados na literatura. No caso do mapeamento associativo, além da inclusão das matrizes Q e K, o uso do algoritmo MLMM com a inclusão de cofatores otimizou os ajustes dos modelos de análise GWAS e possibilitou a identificação de SNPs significativos para florescimento, altura, produção de massa verde e seca, e lignina, considerando a correção de Bonferroni. Futuras aplicações de estratégias de mapeamento fino nas regiões genômicas identificadas poderão permitir a descoberta de SNPs causativos.

A análise de expressão dos genes envolvidos na via de biossíntese de lignina também possibilitou a identificação de um alvo gênico específico para estudos mais aprofundados. Devido à assimetria temporal entre a fase de expressão das enzimas da via de biossíntese da lignina e a fase em que o tecido se apresenta lignificado, outros genes ainda podem apresentar correlação com o teor de lignina em determinado momento. Assim, ainda é necessário o estudo de expressão ao longo dos diversos estágios de desenvolvimento do sorgo. Entretanto, devido à significativa relação com teor final de lignina, o gene *HCT1* é um forte candidato para a aplicação na seleção assistida por marcadores moleculares, por exemplo.

Interessantemente, existem relatos na literatura da influência da enzima HCT na via de fenilpropanoide e de sua relação com o transporte de auxina na

planta. O que relaciona indiretamente os resultados dos Capítulos 2 e 3, uma vez que um dos SNPs associados ao teor de lignina no painel de sorgo encontra-se entre dois genes da família *SAUR*, de proteínas responsivas à auxina, e o gene com expressão correlacionada com o teor de lignina, o *HCT1*, atua na bifurcação da via fenilpropanoide entre a síntese de lignina e de flavonoides. Hipóteses anteriores relacionam o acúmulo de flavonoides, devido ao silenciamento do *HCT*, à interferência na sinalização e ao transporte de auxina.

Assim, o presente estudo contribuiu com novas informações sobre o controle genético de caracteres agroindustriais associados à produtividade de biomassa em sorgo, além de oferecer SNPs e genes candidatos para futuras aplicações no melhoramento de cultivares de sorgo, com a finalidade de aumentar a oferta de matéria-prima sustentável e de qualidade para a produção de bioenergia no Brasil.