

MAPEAMENTO ASSOCIATIVO, IDENTIFICAÇÃO DE REGIÕES GENÔMICAS E SELEÇÃO DE GENES CANDIDATOS PARA COMPOSTOS BIOATIVOS DE SORGO PARA ALIMENTAÇÃO HUMANA

Wellingson Assunção Araujo ¹

¹ Departamento de Genética - Escola Superior de Agricultura Luiz de Queiroz - Universidade de São Paulo

RESUMO

Os compostos bioativos são constituintes extranutricionais que normalmente ocorrem em pequenas quantidades nos alimentos. Muitos têm efeitos antioxidantes que podem contribuir para a prevenção de doenças crônicas, como diabetes, doenças cardiovasculares e obesidade. O sorgo é um cereal considerado uma boa fonte de compostos bioativos e não possui glúten, portanto é totalmente seguro para a produção de alimentos para indivíduos com doença celíaca. Por esses motivos, foi considerado um cereal promissor para uso em alimentos humanos e para melhoria da saúde. Esses compostos variam entre os genótipos de sorgo, portanto, são necessários mais estudos para avaliar seus níveis. No entanto, suas análises químicas são, em geral, caras e demoradas. Portanto, o objetivo deste trabalho foi gerar informações sobre a avaliação desses compostos no sorgo e identificar genes candidatos que possam permitir a seleção de genótipos para o desenvolvimento de cultivares superiores de sorgo. Um painel de sorgo com 240 membros, composto por linhas com ampla variabilidade genética, foi fenotipado para compostos fenólicos totais para análises de mapeamento associativo. Entende-se que dentro destes fenólicos podem estar sendo abrangidos antocianinas, taninos e demais compostos por isso em trabalhos subsequentes serão feitas análises fenotípicas para os compostos mais refinados. Através de metodologias de bioinformática foram encontrados alguns genes que conferem a genótipos um alto conteúdo de compostos bioativos, poderão ser utilizados para o melhoramento genético e tornar o sorgo promissor para uso em alimentos humanos e para promoção da saúde.

INTRODUÇÃO

Os cereais são um importante grupo de espécies cultivadas destinadas à alimentação de grande parte da população mundial, especialmente em países de baixa e média renda, onde a dieta geralmente consiste em grandes quantidades de alimentos básicos com poucos alimentos ricos em micronutrientes e compostos bioativos, como frutas, vegetais, produtos de origem animal, peixes e grãos integrais (HarvestPlus, <http://www.harvestplus.org>). Compostos bioativos são constituintes extranutricionais que geralmente ocorrem em pequenas quantidades nos alimentos,

como por exemplo, os flavonoides e os ácidos fenólicos (KRIS-ETHERTON *et al.*, 2002; DYKES & RONNEY, 2007; CROZIER *et al.*, 2009).

Muitos desses compostos possuem efeitos antioxidantes que podem contribuir para a prevenção de doenças crônicas, tais como diabetes, doenças cardiovasculares e obesidade (KRIS-ETHERTON *et al.*, 2002; DEL RIO *et al.*, 2013). Dentre os cereais mais cultivados, o sorgo [*Sorghum bicolor* (L.) Moench] é considerado uma boa fonte de compostos bioativos apresentando uma variedade de genótipos com diferentes níveis de ácidos fenólicos, flavonóides (por exemplo, antocianinas e tanino) e amido resistente (DYKES & RONNEY, 2007; WU *et al.*, 2012). Entretanto, a maioria do sorgo cultivado é destinada basicamente à alimentação animal e, somente em poucos países da África e da Ásia, o sorgo é utilizado na alimentação humana (DICKO *et al.*, 2006).

O sorgo tem um valor nutricional semelhante ao do milho e tem sido considerado uma boa fonte de compostos bioativos. Além disso, ele não possui glúten, sendo totalmente seguro para pessoas celíacas. Outra importante característica do sorgo é a sua alta tolerância à seca e à acidez do solo, o que torna mais acessível o seu cultivo em regiões áridas e semiáridas do mundo. Ademais, assim como outros cereais, o sorgo tem características que o tornam mais atrativo quando comparado a frutas e vegetais. Por exemplo, seus grãos são secos, o que permite a estocagem por um período mais longo e possui boa estabilidade de seus extratos concentrados, fornecendo várias alternativas para ser utilizado como fonte de alimento, como para a produção de farelo, farinha, pães, bolo, dentre outros (DYKES & ROONEY, 2007). Devido a essas razões, o sorgo vem sendo apontado como um cereal promissor para a alimentação humana e para a promoção da saúde.

O sorgo apresenta uma valiosa fonte e uma grande diversidade de compostos bioativos devido à sua ampla variabilidade genética (WU *et al.*, 2012). Essa diversidade genética e fenotípica pode ser utilizada no melhoramento genético para a obtenção de linhagens elite. No entanto, as análises químicas tradicionalmente utilizadas para a identificação dos compostos funcionais são, em geral, muito onerosas, demandam muito tempo e geram grande quantidade de resíduos.

Dessa forma, a pesquisa genética do sorgo e a compreensão dos mecanismos moleculares responsáveis pelas variações nos níveis desses compostos, tornam-se importantes ferramentas para o melhoramento assistido por marcadores moleculares e o desenvolvimento de variedades com potencial para uma maior quantidade e maior biodisponibilidade de compostos bioativos. Diante disso, a proposta desse trabalho foi identificar regiões genômicas do sorgo associadas às características funcionais visando a seleção de genes candidatos envolvidos nas vias de síntese de compostos bioativos. Tais informações podem ser utilizadas no programa de melhoramento genético de sorgo para, de maneira mais rápida e mais

econômica, auxiliar na identificação de genótipos superiores que apresentem grãos com maior qualidade nutricional e funcional para uso em biofortificação.

MATERIAL E MÉTODOS

Material genético

Neste estudo, foi utilizado um painel associativo contendo 240 linhagens de sorgo com ampla variabilidade genética. Esse painel é composto por uma coleção de linhagens representativas do sorgo cultivado e linhagens usadas em programas de melhoramento nos Estados Unidos e no Brasil (CASA *et al.*, 2008), que tem sido utilizado com sucesso para o estudo de outras características, como eficiência na utilização de fósforo (HUFNAGEL *et al.*, 2014).

Mapeamento associativo

O mapeamento associativo foi utilizado com o objetivo de identificar regiões genômicas associadas aos compostos fenólicos totais que possuem potencial para a promoção da saúde humana. Para esta finalidade, serão usados dados de marcadores SNP obtidos para o painel de linhagens de sorgo via GBS (*Genotyping-by-Sequencing*, ELSHIRE *et al.*, 2011). Essa genotipagem gerou aproximadamente 300.000 marcadores do tipo SNP. Para a realização das análises, também foram utilizados os dados fenotípicos correspondentes à quantificação das características: fenólicos totais. As análises dessas características foram realizadas em sistema HPLC (*High performance liquid chromatography*) (YANG *et al.* 2012; CARDOSO *et al.*, 2014).

Os dados genotípicos foram imputados para redução no número de genótipos perdidos, utilizando o programa Npute (ROBERTS *et al.*, 2007). Em seguida, os SNPs monomórficos foram filtrados usando como parâmetro um máximo de 0,01 para a frequência do alelo menos comum (MAF, do inglês *minor allele frequency*) (BRADBURY *et al.*, 2007). A análise de mapeamento associativo foi realizada no programa GAPIT (*Genomic Association and Prediction Integrated Tool* – version 3). a partir da abordagem de modelos lineares mistos (MLM, do inglês *mixed linear models*) proposta por ZHANG *et al.* (2010), utilizando modelos que consideram os efeitos da estrutura populacional (**Q**), da matriz de relacionamento genético entre indivíduos (**K**), ou ambos os efeitos (**Q+K**), para o controle de falsos positivos (YU *et al.*, 2006), i.e. falsas associações entre genótipo e fenótipo.

Identificação de regiões genômicas do sorgo e seleção de genes candidatos

As regiões genômicas localizadas próximas dos SNPs significativos nas análises associativas e que estejam dentro de um intervalo genômico definido com base no desequilíbrio de ligação local calculado no painel de linhagens de sorgo serão utilizadas para a busca por genes candidatos com funções relacionadas à biossíntese dos compostos bioativos em sorgo. Essas análises foram realizadas

utilizando diferentes ferramentas de bioinformática para o alinhamento de sequências de DNA e busca por similaridades, utilizando o banco de dados do genoma do sorgo, Phytozome (<https://phytozome.jgi.doe.gov/>), e o GenBank NCBI (<http://www.ncbi.nlm.nih.gov>).

Após a análise de GWAS os SNPs que a partir do teste F destacaram-se nos picos do gráfico de Manhattan ("arranha-céus") os quais sugere que a região genômica circundante tem uma forte associação com a característica foram analisados, para esse estudo foram selecionados os primeiros 11 SNPs.

No Phytozome foi utilizada a ferramenta JBrowse onde foi selecionado o genoma de *Sorghum bicolor* v3.1.1. foram aplicadas as seguintes metodologias de busca: para o alinhamento utilizou-se o BLASTX Arabi/Chlamy/Rise, para o mapeamento de transcritos foi selecionado a opção mapped prior version genes e para a visualização de transcritos foram utilizadas as formas transcrit e alternative trascript.

A análises de possíveis genes nas regiões dos SNPs selecionados foi feita seguindo os seguintes passos:

- A posição do SNP assim como o cromossomo foram indicados no campo de busca do JBrowse e a região flanqueadora deste foi estudada com base na similaridade de sequências com os genomas de arabidopsis e arroz.
- Foi observada a notação funcional no genoma do Sorgo para os transcritos próximos aos SNPs.
- Investigou-se a homologia das proteínas dos genes com outras espécies para aumentar o conhecimento funcional da proteína.
- Além disso, com base em revisões bibliográficas, foram identificados e selecionados genes candidatos que foram previamente caracterizados em estudos já realizados em sorgo ou em outras espécies, e que estejam relacionados com a biossíntese de compostos bioativos.

Um exemplo da abordagem é representado nas figuras abaixo

Passo 1 – No banco de dados Phytozome foi utilizada a ferramenta JBrowse. Em laranja está o genoma de sorgo, em vermelho é mostrada a região próxima aos SNPs S4_60804216, S4_60793207, S4_60804649, em azul são todos os parâmetros utilizados na ferramenta, em rosa é representado um transcrito de sorgo e em amarelo uma notação gênica para arábido.

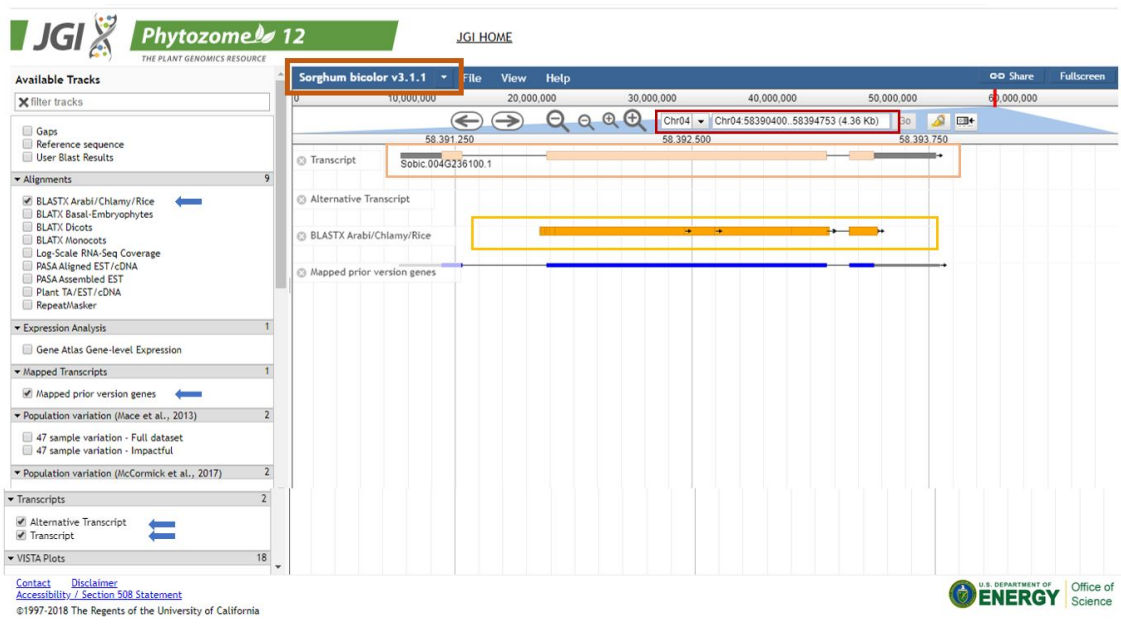


Figura 1 – Representação da plataforma do banco de dados Phytozome. Em laranja está o genoma de sorgo, em vermelho é mostrada a região próxima aos SNPs S4_60804216, S4_60793207, S4_60804649, em azul são todos os parâmetros utilizados na ferramenta, em rosa é representado um transcrito de Sorgo e em amarelo uma notação gênica para arábido.

Passo 2 – Clicando no transcrito de sorgo é possível estudar um pouco da sua notação funcional tanto com informações do Phytozone como do PATHWAY, PFAM e outros bancos de dados. Além disso quando selecionado a sequência resultante do BLAST X é possível ter acesso as informações da notação funcional em espécies modelo e os dados estatísticos da análise. Esses campos são representados em verde.

The screenshot displays the Phytozone 12 web interface. The main content area shows gene information for *Sorghum bicolor* (Gene Sobic.0040236100). Key details include the transcript name, location on chromosome 04, and a description: "(1 of 8) PTHR11709/PTHR11709.SF118 - MULTI-COPPER OXIDASE // SUBFAMILY NOT NAMED". A protein domain view shows a single domain from position 1 to 573. Below this, functional annotations for this locus are listed, including two entries from the PATHWAY database: "matrineol biosynthesis" (PWW-5865) and "jucoidin B biosynthesis" (PWW-6824). On the right, a comparison with *Arabidopsis thaliana* (match Arath|AT5G48100.1_Gx8P) is shown. This section includes primary data such as Name, Type (match), Score (1e-155), and Description: "Laccase/Diphenol oxidase family protein AT5G48100.1 AT5G48100.1". It also lists attributes like Alias (TT10, LAC15, etc.), Description, ID, Seq_id, Source, Species, and a link to the Region sequence.

Figura 1 – Representação da plataforma do banco de dados Phytozone (Continuação) Em verde podem ser vistas características de anotações funcionais em Sorgo e espécies modelo.

RESULTADOS

Mapeamento Associativo

Análise gráfica do quantil-quantil (QQ)

O gráfico quantil-quantil (QQ) foi útil para avaliar quão bem o modelo usado no GWAS explica a estrutura da população e a relação de parentesco. No gráfico abaixo, os logaritmos negativos dos valores P dos modelos ajustados no GWAS foram plotados contra seu valor esperado na hipótese nula, a que indica nenhuma associação com a característica. Como pode ser percebido a maioria dos SNPs testados provavelmente não está associada à característica, isso pode ser visto devido a maioria dos pontos no gráfico QQ deve estar na linha diagonal. Desvios desta linha sugerem presença de associações. É possível observar que os SNPs na seção superior direita do gráfico, desvie-se da diagonal. Esses SNPs provavelmente estão associados com a característica em estudo

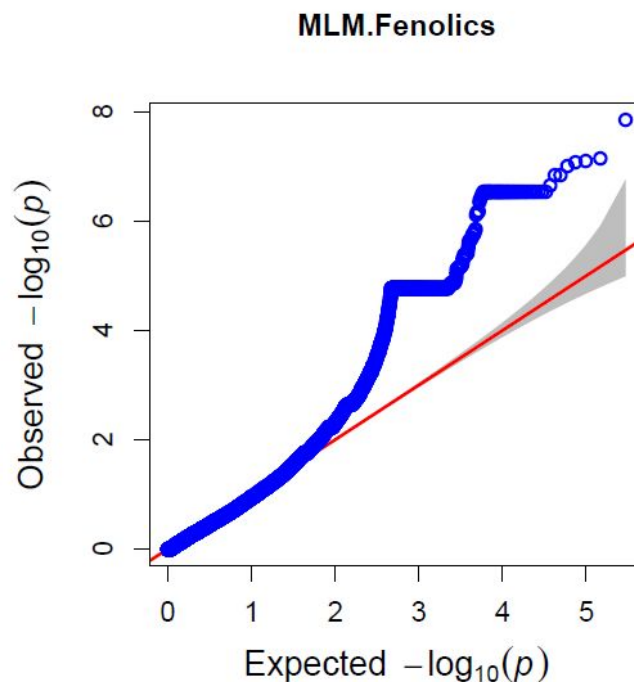


Figura 3 - Quantil-quantil (QQ) - gráfico de valores-P. O eixo Y é a base negativa observada 10 logaritmo dos valores P, e o eixo X é o logaritmo negativo esperado da base 10 dos valores P, sob a suposição de que os valores P seguem uma distribuição uniforme de [0,1]. As linhas pontilhadas mostram o intervalo de confiança de 95% para o gráfico QQ sob a hipótese nula de não associação entre o SNP e a característica.

Gráfico de Manhattan

Com o gráfico de Manhattan foi possível observar a dispersão que resume os resultados do GWAS. No eixo X é representada a posição genômica de cada SNP, e o eixo Y é o logaritmo negativo do valor P obtido do modelo GWAS especificamente a partir do teste F para testar H0 a qual testa nenhuma associação entre o SNP e a característica.

Grandes picos no gráfico Manhattan sugerem que a região genômica circundante tem uma forte associação com a característica. Como podemos perceber ocorreram picos que se sobressaem em praticamente todos os cromossomos, o que já era esperado devido a características de fenólicos totais abrangerem diferentes compostos bioativos e cada um deles possuir uma característica poligênica. Entretanto foram selecionados os SNPs que apresentaram uma maior relevância na análise, os quais estão presentes nos cromossomos 1,4,5 e 9.

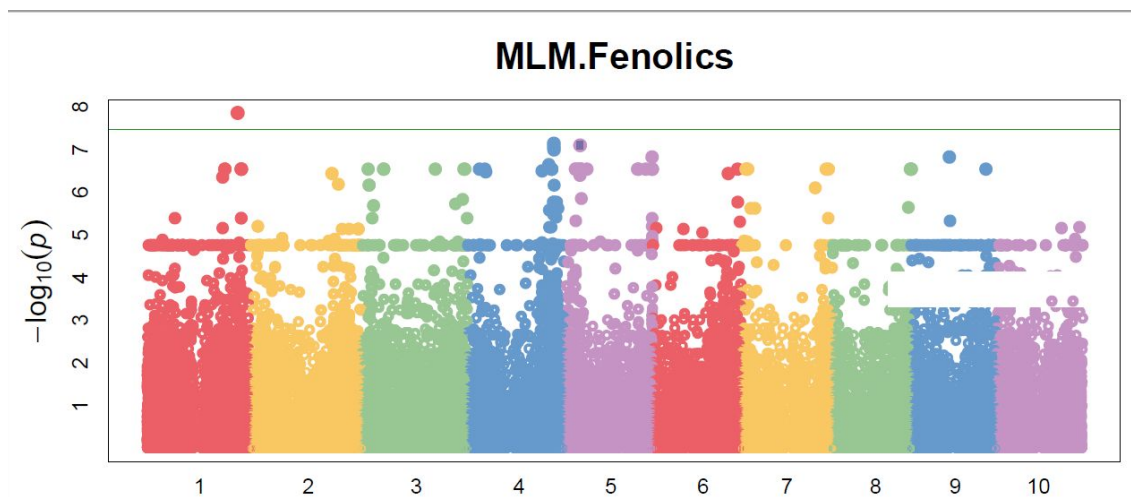


Figura 4 – Gráfico de Manhattan. O eixo X é a posição genômica dos SNPs no genoma e o eixo Y é a base logarítmica 10 negativa dos valores P. Cada cromossomo é colorido de forma diferente.

Tabela de Associação

Na tabela abaixo são mostrados os resultados do GWAS. Ela fornece um resumo detalhado dos resultados advindos do GWAS. É possível ver os resultados para cada SNP acima do limite do alelo menor frequência. Os SNPs estão classificados por seus valores de P (do menor para o maior). Para esse trabalho foram selecionados os 11 primeiros SNPs desta tabela, pois os mesmos se relacionam com os resultados do gráfico de Manhattan, percebeu-se uma predominância de regiões nos cromossomos 1, 4 e 5. Além disso as regiões mostraram-se próximas, diante disso focou-se no estudo dessas regiões.

SNP	Chromosome	Position	P.value	maf
S1_64220969	1	64220969	1,40E+06	0,0347
S4_60804216	4	60804216	7,24E+06	0,052
S5_11221818	5	11221818	7,83E+06	0,0173
S4_60793207	4	60793207	8,46E+06	0,0289
S4_60804649	4	60804649	9,86E+06	0,0636
S5_61846894	5	61846894	1,45E+07	0,0173
S4_57348059	4	57348059	2,25E+07	0,237
S1_66776598	1	66776598	2,90E+07	0,0116
S1_66776617	1	66776617	2,90E+07	0,0116
S1_66790647	1	66790647	2,90E+07	0,0116
S1_66790650	1	66790650	2,90E+07	0,0116

Tabela 1 - Resultados do GWAS para os SNPs analisados. Esta tabela fornece o ID SNP, cromossomo, posição bp, valor P e a frequência alélica menor (maf)

Busca de genes relacionados a característica de compostos fenólicos totais

A tabela a seguir resume todas as análises feitas pela busca de genes candidatos a características compostos bioativos . Nela consta o código do transcrito de Sorgo no banco de dados Phytozone, o cromossomo e a posição inicial, as informações de anotação foram obtidas através do BLASTX contra genoma de espécies modelo como Arabidopsis e Arroz Foi levado em consideração as características da notação funcional do Genoma de sorgo e também as informações dos bancos de dados PATHWAY e PFAM. Como resultados foram encontrados genes que tem fortes indícios de associação com o processo de produção de compostos bioativos.

Sorgo Código no Phytozone	Cromos	Posição inicial do loco	Anotação	ID / Gene Arabidopsis
Sb04g031730	4	61663774	WD40 Tanino - Tannin1	AT3G26640 / LWD2, TTG1.
Sobic.004G2361 00.1	4	58390963	Polifenol oxidase (LAC)	AT5G48100 / TT10
Sb04g037630	4	67224600	Mate efflux family protein	AT3G59030 / TT12
Sobic.004G3101 00.1	4	64699665	Flavonol biosynthesis Rutin biosynthesis	AT5G08640 / FLS1
Sb01g032770	1	62819065	R2R3 MYB	AT5G35550 / TT2
Sb05g008850	5	17709528	Encodes chalcone synthase (CHS)	AT5G13930 / TT4

Tabela 2 - Resultado da busca por genes candidatos a características compostos bioativos.

DISCUSSÃO

Genes candidatos

Genes TTG1, LWD2, Tan1 e os fatores de transcrição WD40

O gene TTG1 codifica um fator de transcrição WD. Busca de Genes relacionados a característica de Compostos fenólicos Totais. Para a cultura do Feijoeiro fava esse gene tem uma grande importância por estar ligado a produção de antocianinas e taninos, para a cultura é importante que os genótipos tenham as formas recessivas dos genes expressando os fenótipos do tipo zt-1 e zt-2, as quais possuem uma interrupção na biossíntese desses compostos, pois assim eles são melhores aproveitados quando utilizados como alimento (Gutierrez et al 2019). O objetivo do melhoramento de Feijoeiro é contrário ao que se pretende com o melhoramento de Sorgo para a alimentação humana, no Sorgo se busca o acúmulo de desses compostos, pois estes estão ligados a várias vantagens para a saúde entretanto para a cultura do feijão fava, o qual ainda é utilizado também para a alimentação animal, o aumento na energia necessária para a digestibilidade desses compostos não é interessante (Bond, 1976).

As proteínas WD-repeat são muito conhecidas e possuem diversos estudos os quais indicam que estas são encontradas nos eucariotos e desempenham um papel importante na regulação de uma ampla variedade de funções celulares, como transdução, transcrição e proliferação de sinais. Em um estudo desenvolvido por Sompornpailin et al, 2002 foi isolado um cDNA que codifica uma proteína de repetição de WD chamada de PFWD, das folhas pigmentadas com antocianina de *Perilla frutescens*. Quando esse gene foi superexpressado em *Arabidopsis* as plantas exibiram alterações fenotípicas, incluindo aumento da produção de antocianina. Dado os resultados desse trabalho é possível indicar que esse gene tem um papel importante na produção de compostos bioativos como as antocianinas.

O gene Tannin1 (Tan1) possui estudos bem conclusivos na cultura do Sorgo, na cultura o gene codifica o fator de transcrição de repetição WD40 (Sb04g031730), controlando a biossíntese de taninos no sorgo, esse resultado provém de um experimento no qual foi feita uma transformação estrutura do gene Tan1 em um mutante *Arabidopsis* o qual não acumularam tanino nas sementes. Tan1 é semelhante ao PAC1, que codifica um regulador da via das antocianinas do milho (Mizuno et al, 2014)

Wu,(2012) relatou desenvolveu um grande estudo sobre QTL relacionados com taninos em Sorgo, no qual após estudos preliminares focou no cromossomo 4 no trabalho 20 genes candidatos envolvidos na via biossintética de tanino de *Arabidopsis* e outras espécies foram investigados. O estudo resultou em quatro modelos de genes de sorgo: Sb04g037630.1, Sb04g030570.1, Sb04g031710.1 e Sb04g031710.1 e Sb04g031730 os quais estavam no cromossomo 4 segundo os autores esses genes são homólogos dos genes TTG1, TT12, LEC2 e FUS3 em

Arabidopsis e LAR1 Vitis vinifera. Esses resultados confirmam a análise feita pelo mapeamento associativo que indicou diversos SNPs associados à característica na região do cromossomo 4, assim como a presença do transcrito Sb04g031730 resultante da análise feita no JWBrowse.

TT10

O gene TT10 em arabidopsis codifica uma proteína com função polifenol oxidases do tipo lacase. tem sua expressão na testa das sementes, tendo como produtos os flavonóides, proantocianidinas e flavonóis. quando o gene se apresenta na forma recessiva (tt10) ocorre uma maior quantidade de epicatequina e proantocianidinas mais solúveis do que as sementes do tipo selvagem. Indicando que o gene TT10 está envolvido na polimerização oxidativa de flavonóides e funciona como uma oxidase flavonóide do tipo lacase (Pourcel et al., 2005).

TT12 MATE efflux family protein

O trabalho envolvendo o gene TT12 o qual tem função de acumular no vacúolo precursores de proantocianidina na semente. Um experimento no qual um mutante de Arabidopsis foi transformado para expressar o fenótipo recessivo para esse gene mostraram ausência de flavan-3-ols e proantocianidinas e redução flavonol quercetin-3-O-ramnídeo. este estudo também indica que o TT12 transporte flavan-3-ols glicosilado in vivo e que faz mediação no transporte de antocianinas in vitro (Marinova et al., 2007). Para a cultura do sorgo não foram encontrados estudos mais conclusivos sobre esse gene. entretanto na análise ele possui grande semelhança estrutural com o transcrito Sb04g037630.

TT2

TT2 indicam que este codifica um fator de transcrição putativo do domínio R2R3 MYB essa proteína possui uma alta similaridade à proteína OsMYB3 do arroz e ao fator COLORLESS1 do milho .entende-se também que que um complexo ternário composto de TT2, TT8 e TTG1 seja necessário para a expressão correta da BAN (BANYULS) no endotélio da semente e que esses genes também estejam diretamente ligados a resistência a estresses abióticos (Nesi et al., 200; Chen et al., 2012)

FLS1 Flavonol synthase 1

O gene FLS1 possui grandes estudos nas vias de biossíntese e características bioquímicas dos flavonoides, esse composto bioativo possui uma grande importância para a saúde humana, quando o FLS1 é superexpressando é possível perceber nas plantas um acúmulo de flavonóides comparada com o tipo selvagem, mostrando também mudança fenotípica nas sementes, todavia quando o gene se encontra na forma recessiva fls-3 são obtidas plantas com maiores teores de antocianinas e menores de flavonoides (Nguyen et al., 2016) Diante desses resultados é possível sugerir que esse gene possui grande relevância na produção de compostos bioativos e deve ser estudado mais a fundo na cultura do Sorgo.

TT4

TT4 codifica a chalcona sintase (CHS), uma enzima essencial na etapa inicial da biossíntese dos flavonóides, e a perda da função do TT4 causa uma falta de flavonóides. A deficiência de desse composto leva ao aumento do teor de ácidos graxos em sementes oleaginosas. Com dados de cruzamento Xuan et al, (2018) mostrou que com a perda da função TT4 ocorreu uma ativação da glicólise e aumento da produção de auxinas em embriões de sementes, aumentando assim a biossíntese de ácidos graxos, todavia a mucilagem das sementes não foi alterada, mostrando assim que o teor de flavonoides é importante nesses dois processos.

CONCLUSÃO

Como parte de estudos iniciais relacionados a mapeamento associativo para compostos Bioativos em Sorgo, como os fenólicos totais, puderam ser encontrados diferentes genes candidatos baseando-se em genes de Arabidopsis os quais possuem uma gama maior de estudos, estudos estes que podem ser aplicados na cultura do Sorgo a fim de validar todos esses genes na espécie.

REFERÊNCIAS

- BOND, D. A. In vitro digestibility of the testa in tannin-free field beans (*Vicia faba* L.). J. Agric. Sci. 86, 561–566 (1976).
- RADBURY, P.J. et al. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, v. 23, p. 2633–2635, 2007.
- CARDOSO, L. M et al. Effects of processing with dry heat and wet heat on the antioxidante profile of sorghum. Food Chemistry, v. 152, p. 210-217, 2014.
- CASA, M. et al. Community resources and strategies for association mapping in sorghum. Crop Sci. v. 48, p. 30–40, 2008.
- CHEN M, WANG Z, ZHU Y, et al. The effect of transparent TESTA2 on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in Arabidopsis. Plant Physiol. 2012;160(2):1023–1036. doi:10.1104/pp.112.202945
- CROZIER, A.; JAGANATH, I.B.; CLIFFORD, M.N. Dietary phenolics: chemistry, bioavailability and effects on health. Natural product reports, v. 26, n. 8, p. 1001–1043, 2009.
- DEL RIO, D. et al. Dietary (Poly)phenolics in Human Health: Structures, Bioavailability, and Evidence of Protective Effects Against Chronic Diseases. Antioxidants & redox signaling, v. 18, n. 14, p. 1818–92, 2013.
- DICKO, M.H. et al. Sorghum grain as human food in Africa: relevance of content of starch and amylase activities. African Journal of Biotechnology, v. 5, n. 5, p. 384–395, 2006.
- DYKES, L; ROONEY, L. Phenolic Compounds in Cereal Grains and Their Health Benefits. Cereal Foods World, v. 52, n. 3, p. 105–111, 20

ELSHIRE, R J et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, v. 6, p. 1-10, 2011.

GUTIERREZ, N., TORRES, A.M. Characterization and diagnostic marker for TTG1 regulating tannin and anthocyanin biosynthesis in faba bean. Sci Rep 9, 16174 (2019) doi:10.1038/s41598-019-52575-x

HUFNAGEL, B. et al., Duplicate and Conquer: multiple homologs of Phosphorus-Starvation Tolerance-1 enhance phosphorus acquisition and sorghum performance on low-phosphorus soils. Plant Physiology, v. 166, p. 659-677, 2014.

KRIS-ETHERTON, P.M. et al. Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. The American Journal of Medicine, v. 113 Suppl, n. 1, p. 71S–88S, 2002.

MARINOVA K, POURCEL L, WEDER B, et al. The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H⁺ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell. 2007;19(6):2023–2038. doi:10.1105/tpc.106.046029

MIZUNO, H., YAZAWA, T., KASUGA, S. et al. Expression level of a flavonoid 3'-hydroxylase gene determines pathogen-induced color variation in sorghum. BMC Res Notes 7, 761 (2014) doi:10.1186/1756-0500-7-761

NESI N, JOND C, DEBEAUJON I, CABOCHE M, LEPINIEC L. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell. 2001;13(9):2099–2114. doi:10.1105/tpc.010098

NGUYEN, NH., KIM JH., KWON J., JEONG CY., LEE W., LEE D., HONG SW., LEE H. Characterization of Arabidopsis thaliana FLAVONOL SYNTHASE 1 (FLS1) -overexpression plants in response to abiotic stress in Plant Physiol Biochem. 2016 Jun;103:133-42. doi: 10.1016/j.plaphy.2016.03.010..

POURCEL L, ROUTABOUL JM, KERHOAS L, CABOCHE M, LEPINIEC L, DEBEAUJON I. TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. Plant Cell. 2005;17(11):2966–2980. doi:10.1105/tpc.105.035154

SOMPORNPAILIN, K., MAKITA, Y., YAMAZAKI, M. et al. A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in *Perilla frutescens* . Plant Mol Biol 50, 485–495 (2002) doi:10.1023/A:101985092162

YANG, L. et al. Sorghum phenolics demonstrate estrogenic action and Induce apoptosis in nonmalignant colonocytes. Nutrition and Cancer, p. 64, v. 3, p. 419–427, 2012.

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

WU Y, LI X, XIANG W, et al. Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1. Proc Natl Acad Sci U S A. 2012;109(26):10281–10286. doi:10.1073/pnas.1201700109

XUAN L., ZHANG C., YAN T., WU D., HUSSAIN N., LI .1, CHEN M., PAN J., JIANG L..TRANSPARENT TESTA 4-mediated flavonoids negatively affect embryonic fatty acid biosynthesis in Arabidopsis in Plant Cell Environ. 2018 Dec;41(12):2773-2790. doi: 10.1111/pce.13402. Epub 2018 Aug 7.

ZHANG, Z. et al. Mixed linear model approach adapted for genome-wide association studies. Nature genetics, v. 42, n. 4, p. 355–60, abr. 2010