

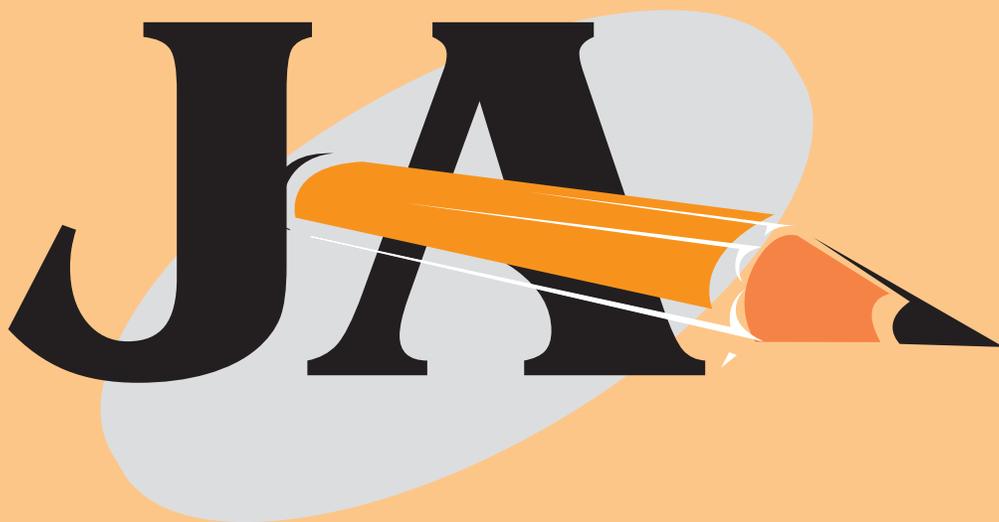
DOCUMENTOS

440

ISSN 2176-2937
Agosto/2021

XVI Jornada Acadêmica da Embrapa Soja

Resumos expandidos



**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Soja
Ministério da Agricultura, Pecuária e Abastecimento**

DOCUMENTOS 440

XVI Jornada Acadêmica da Embrapa Soja Resumos expandidos

*Regina Maria Villas Bôas de Campos Leite
Kelly Catharin*
Editoras Técnicas

Embrapa Soja
Londrina, PR
2021

Exemplares desta publicação podem ser adquiridos na:

Embrapa Soja
Rod. Carlos João Strass, s/n
Acesso Orlando Amaral, Distrito da Warta
CEP 86001-970
Caixa Postal 231
Londrina, PR
Fone: (43) 3371 6000
www.embrapa.br/soja
www.embrapa.br/fale-conosco/sac

**Comitê Local de Publicações
da Embrapa Soja**

Presidente
Alvadi Antonio Balbinot Junior

Secretária-Executiva
Regina Maria Villas Bôas de Campos Leite

Membros
Clara Beatriz Hoffmann-Campo, Claudine Dinali Santos Seixas, Ivani de Oliveira Negrão Lopes, Liliane Márcia Mertz-Henning, Marco Antônio Nogueira, Mariangela Hungria da Cunha, Mônica Juliani Zavaglia Pereira e Norman Neumaier

Supervisão editorial
Vanessa Fuzinato Dall' Agnol

Normalização bibliográfica
Valéria de Fátima Cardoso

Projeto gráfico da coleção
Carlos Eduardo Felice Barbeiro

Edição eletrônica e capa
Vanessa Fuzinato Dall' Agnol

1ª edição
PDF digitalizado (2021).

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Soja

Jornada Acadêmica da Embrapa Soja (16. : 2021: Londrina, PR).

Resumos expandidos [da] XVI Jornada Acadêmica da Embrapa Soja / Regina Maria Villas Boas de Campos Leite, Kelly Catharin, editoras técnicas – Londrina: Embrapa Soja, 2021.

163 p. (Documentos / Embrapa Soja, ISSN 2176-2937 ; n. 440).

1. Soja-Pesquisa. 2. Pesquisa agrícola. I. Série.

CDD: 630.2515 (21. ed.)

Análise *in silico* de genes para ganho em teor e qualidade proteica em semente de soja

MARIN, D. R.¹; SOUZA, E. M. G.¹, MOLINARI, M. D. C.²; MARIN, S. R. R.³; NEPOMUCENO, A. L.⁴; MERTZ-HENNING, L. M.⁴

¹Unifil, Centro Universitário Filadelfia, Bolsista PIBIC/CNPq, Londrina, PR, dani.rockenbachmarin@gmail.com; ²Pós-doutoranda, Embrapa Soja; ³Analista, Embrapa Soja; ⁴Pesquisador, Embrapa Soja.

Introdução

A grande importância da soja está relacionada com sua composição, principalmente em relação ao teor de proteína (38% - 42%) e óleo (18% - 22%) que proporciona uso na alimentação humana e animal, e aplicação na indústria farmacêutica, indústria química e para produção de biocombustível (Patil et al., 2018). Não somente a quantidade, mas também a qualidade da proteína é crítica para o uso na alimentação animal, e a composição de aminoácidos na proteína está entre os principais fatores que determinam a qualidade da soja. A proteína de soja possui uma proporção adequada dos aminoácidos essenciais para alimentação humana e animal, com exceção dos aminoácidos sulfurados metionina e cisteína (Zhang et al., 2018), dessa forma, o aumento destes aminoácidos na proteína pode promover melhor qualidade nutricional.

O melhoramento genético para teor e composição da proteína é desafiador, pois a composição química da semente é influenciada por fatores como ambiente e genótipo (Kim et al., 2016), e por essas características serem controladas por diversos genes (Patil et al., 2017). Numerosos estudos mapearam principalmente no cromossomo 20 diversos QTLs associados a composição e teor de proteína na semente de soja (Van; McHale, 2017), porém poucos fizeram a caracterização desses QTLs. A caracterização *in silico* de genes relacionados com o acúmulo e a composição da proteína possibilitam a prospecção de candidatos mais promissores para uso em estratégias biotecnológicas para melhoria dessas características. Assim, esse trabalho tem como objetivo a análise *in silico* de genes presentes no QTL mPCM20-3, envolvido no aumento de quantidade proteica e de aminoácidos sulfurados (metionina e cisteína) em grãos de soja, visando selecionar alvos para edição genômica.

Materiais e Métodos

A partir do trabalho de Van e McHale (2017), foram selecionados cinco genes (Glyma.20G121200; Glyma.20G121300; Glyma.20G122300; Glyma.20G122400; Glyma.20G122500) presentes no QTL mPCM20-3 (identificado no cromossomo 20 de soja). As sequências proteicas dos genes e seus respectivos transcritos foram recuperados a partir do banco de dados Phytozome¹. Em seguida, foi realizada a busca por genes ortólogos em espécie modelo para dicotiledônea (*Arabidopsis thaliana* – genoma TAIR10) e cópias no genoma da soja (Wm82.a4.v1) a partir da ferramenta BlastN do genome browser Persephone². O alinhamento de sequências proteicas foi realizado no software Muscle³ e Jalview V2 (Waterhouse et al., 2009). A predição do sítio de localização subcelular foi realizado a partir do software TargetP (Emanuelsson et al., 2007) para identificação do local de ação do gene. A árvore filogenética foi obtida a partir do software ITol⁴. A identificação da via metabólica foi obtida a partir do banco KEGG pathways⁵. Os domínios proteicos conservados foram identificados através do software MotifScan⁶ via banco de dados de famílias gênicas pfam, utilizando alinhamentos múltiplos locais e globais HMMs (*Hidden Markov Models*). As expressões e *heatmap* em sementes de soja foram obtidas em transcritos por milhão-TPM do banco de dados transcriptômicos de soja Soybean Expression Atlas⁷ (Machado et al., 2020).

Resultados e Discussão

Todos os genes selecionados apresentam ortólogos na espécie modelo para dicotiledôneas, *Arabidopsis thaliana*. O gene com maior similaridade proteica com *Arabidopsis* foi o Glyma.20G122500 com 87% e o de menor simi-

¹ <https://phytozome.jgi.doe.gov/pz/portal.html>

² <https://persephonesoft.com/>

³ <https://www.ebi.ac.uk/Tools/msa/muscle/>

⁴ <https://itol.embl.de/upload.cgi>

⁵ <https://www.genome.jp/kegg/pathway.html>

⁶ https://myhits.sib.swiss/cgi-bin/motif_scan

⁷ http://venanciogroup.uenf.br/cgi-bin/gmax_atlas/index.cgi

laridade foi o Glyma.20G122300 com 53%. Somente o Glyma.20G122300 está presente em cópia única, os demais Glymas apresentam-se em cópia duplicada no cromossomo 10. A similaridade entre cópias foi $\geq 93\%$. Todos os cinco genes encontram-se em tandem no cromossomo 20 entre as posições 36.300.000 a 36.600.000 cM, bem como suas cópias no cromossomo 10 que também se encontram posicionados em tandem entre as posições 49.100.000 a 49.300.000 cM (Tabela 1).

Os genes Glyma.20G121200 e Glyma.20G121300 codificam para uma enzima Diacilglicerol difosfato fosfatase (E.C.3.1.3.4) que atua no metabolismo de Glicerolípido, Glicerofosfolípido, éter lipídico e esfingolípido. Os genes Glyma.20G122300 e Glyma.20G122400 codificam para uma enzima Procolágeno-lisina 5-dioxigenase (E.C.1.14.11.4) que atua na degradação de lisina. Por fim, o gene Glyma.20G122500 codifica para uma enzima Frutose-bifosfato aldolase (E.C.4.1.2.13) que atua na via de pentose fosfato, glicólise e gluconeogenese, além de atuar também no metabolismo da frutose e manose e na fixação de carbono.

Ao comparar as sequências proteicas entre as espécies foi possível observar que para os genes codificadores da enzima Diacilglicerol difosfato fosfatase e Procolágeno-lisina 5-dioxigenase, não foi detectada a presença de peptídeo sinal, o que indica que as proteínas têm atuação permanente no citosol (Figura 1A e 1 C). Já a enzima Frutose-bifosfato aldolase apresentou 98% de probabilidade de atuar no cloroplasto. O peptídeo de transferência cloroplasmática foi observado na posição 30pb em ambas as espécies (Figura 1B).

Tabela 1. Anotação funcional dos genes em *Arabidopsis thaliana* e soja. A porcentagem de similaridade (%S) refere-se à propriedade química similar entre as proteínas. Legenda: E.C. (*enzyme code*): classificação número enzimática. TT: tamanho do transcrito. *Start* e *Stop*: posição inicial e final no genoma respectivamente. NT: Número de transcritos alternativos. NE: Número de exons. Chr: cromossomo.

Genes presentes no QTL mPCM20-3 -Chr 20 - correlacionado positivamente a quantidade de proteína e aminoácidos sulfurados (cisteína + metionina)													
Gene ID	Cópias	%S	NT	NE	start	stop	Anotação funcional	E.C.	Ortólogo	%S	NT	NE	
Glyma.20G121200	Glyma.10G270100	93	2	8	36.384.081	36.391.030	Diacylglycerol difosfato fosfatase	3.1.3.4	AT2G01180	64	2	2 e 1	
Glyma.20G121300	Glyma.10G270000	98	2	7	36.395.415	36.400.718	Diacylglycerol difosfato fosfatase	3.1.3.4	AT1G15080	76	1	7	
Glyma.20G122300				4	9/8/6.16	36.506.583	36.511.326	Procollagen-lysine 5-dioxygenase	1.14.11.4	AT3G18210	53	2	9
Glyma.20G122400	Glyma.10G268600	97	1	9	36.516.501	36.521.436	Procollagen-lysine 5-dioxygenase	1.14.11.4	AT3G18210	59	2	9	
Glyma.20G122500	Glyma.10G26850	99	1	6	36.524.182	36.526.957	Fructose-bisphosphate aldolase	4.1.2.13	AT2G01140	87	1	4	

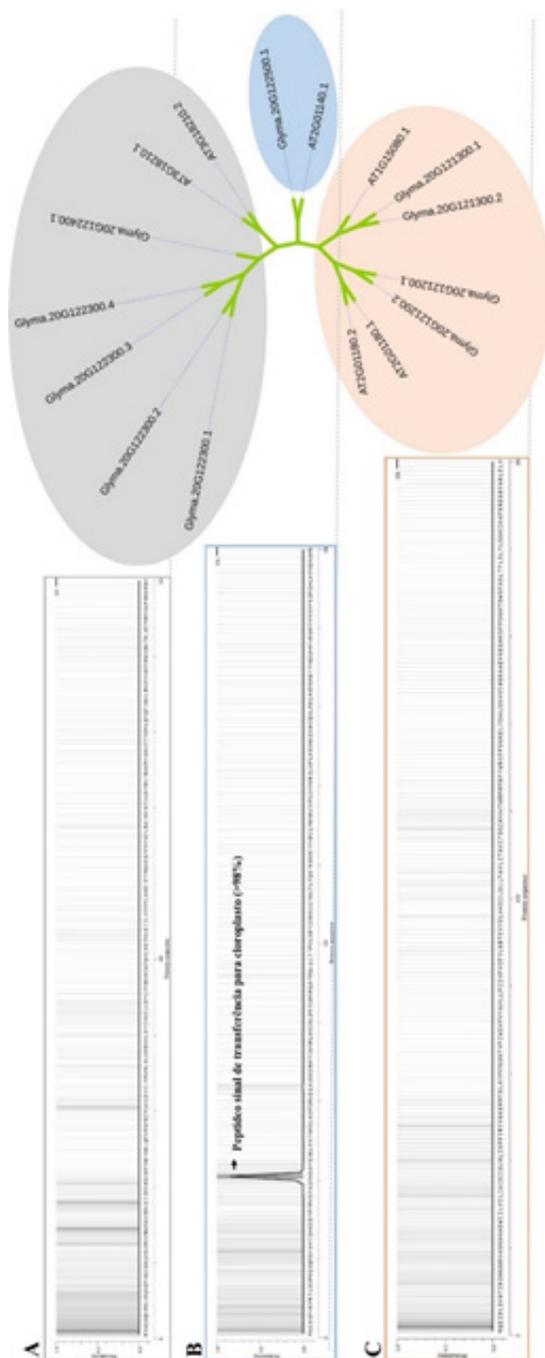


Figura 1. Identificação do sítio de localização subcelular nas proteínas dos cinco genes de soja e seus respectivos ortólogos em *Arabidopsis*. **A.** Genes codificadores para enzima Diacilglicerol difosfato fosfatase (E.C.3.1.3.4). **B.** Gene codificador para enzima Fructose-bisphosphate aldolase (E.C. 4.1.2.13). **C.** Genes codificadores para enzima Procollagen-lysine 5-dioxygenase (E.C. 1.14.11.4).

A análise de domínio proteico (Figura 2) revelou que todos os genes, transcritos e ortólogos que codificam para Diacilglicerol difosfato fosfatase apresentam um domínio PAP2 entre as posições 120 a 280 em soja e 100 a 260 em *Arabidopsis*. Os genes que codificam para Procollagen-lysine 5-dioxygenase apresentaram domínios 2OG-Fell_Oxy apenas no gene Glyma.20G122400 de soja e no ortólogo em *Arabidopsis*. No Glyma.20G122300 não há presença de domínio ativo, pois a região entre as posições 280 a 340 do domínio sofreu fortes alterações físico-químicas e de identidade em todos os transcritos, podendo ser este um indicativo de que este gene não seja funcional. Para os outros o domínio encontra-se entre as posições 240-340 em soja e 260 a 340 em *Arabidopsis*. Já o gene Glyma.20G122500 que codifica para a enzima Fructose-bisphosphate aldolase foi identificado o domínio Glicolítico na posição 40 a 380 em soja e *Arabidopsis*. A similaridade proteica entre as espécies foi de 87%.

O *heatmap* mostra as médias de expressões em semente dentro de cada bioensaio (SAMN) avaliado, altas expressões estão representadas na barra em azul e baixas expressões em amarelo. A média das expressões dos 158 bioensaios estão representados no gráfico de barras. Os genes Glyma.20G122500 e sua cópia Glyma.10G26850 apresentam as maiores expressões em sementes. Estes codificam para a enzima Fructose-bisphosphate aldolase. Entre os genes que codificam para a enzima Diacilglicerol difosfato fosfatase os genes Glyma.20G121300 e sua cópia Glyma.10G270000 são os mais expressos em sementes quando comparados aos genes Glyma.20G121200 e sua cópia Glyma.10G270100 (Figura 3). Por fim, todos os genes que codificam para a enzima Procollagen-lysine 5-dioxygenase apresentam altas expressões em sementes (Glyma.20G122300; Glyma.20G122400; Glyma.10G268600) (Figura 3).

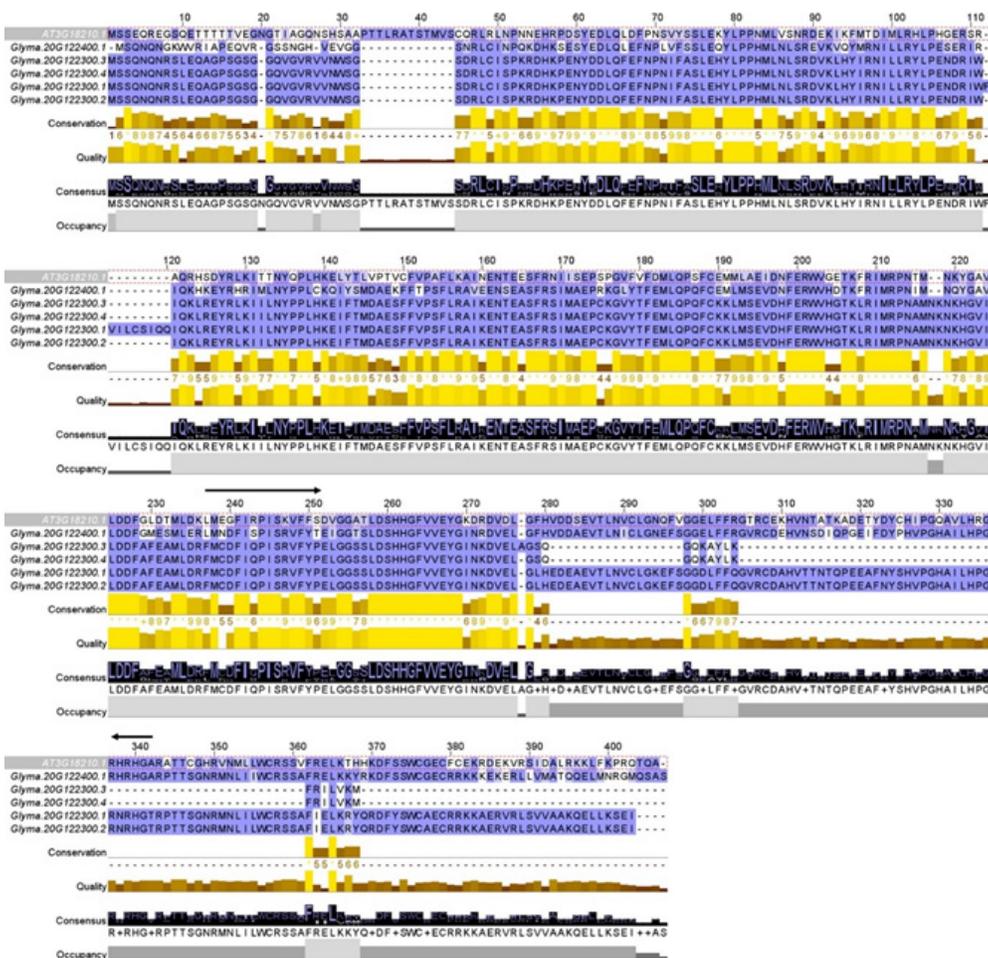


Figura 2. Alinhamento das seqüências proteicas dos genes codificadores da enzima Procollagen-lysine 5-dioxygenase (E.C. 1.14.11.4). Em vermelho estão representadas regiões com mutações não sinônimas. Setas em preto representam a posição do domínio ativo 20G-Fell_Oxy. Os resíduos conservados são destacados (azul:> 80% de concordância, azul médio:> 60% de concordância, azul claro:> 40% de concordância; apenas os resíduos que concordam com o resíduo de consenso para cada coluna são coloridos). O histograma de anotação de conservação abaixo do alinhamento reflete a conservação das propriedades físico-químicas e marca os resíduos absolutamente conservados (pontuação 11) com um asterisco amarelo (*) e colunas onde as propriedades físico-químicas são conservadas (pontuação 10) com um (+) amarelo; posições menos conservadas são mostradas em cores mais escuras com pontuação decrescente.

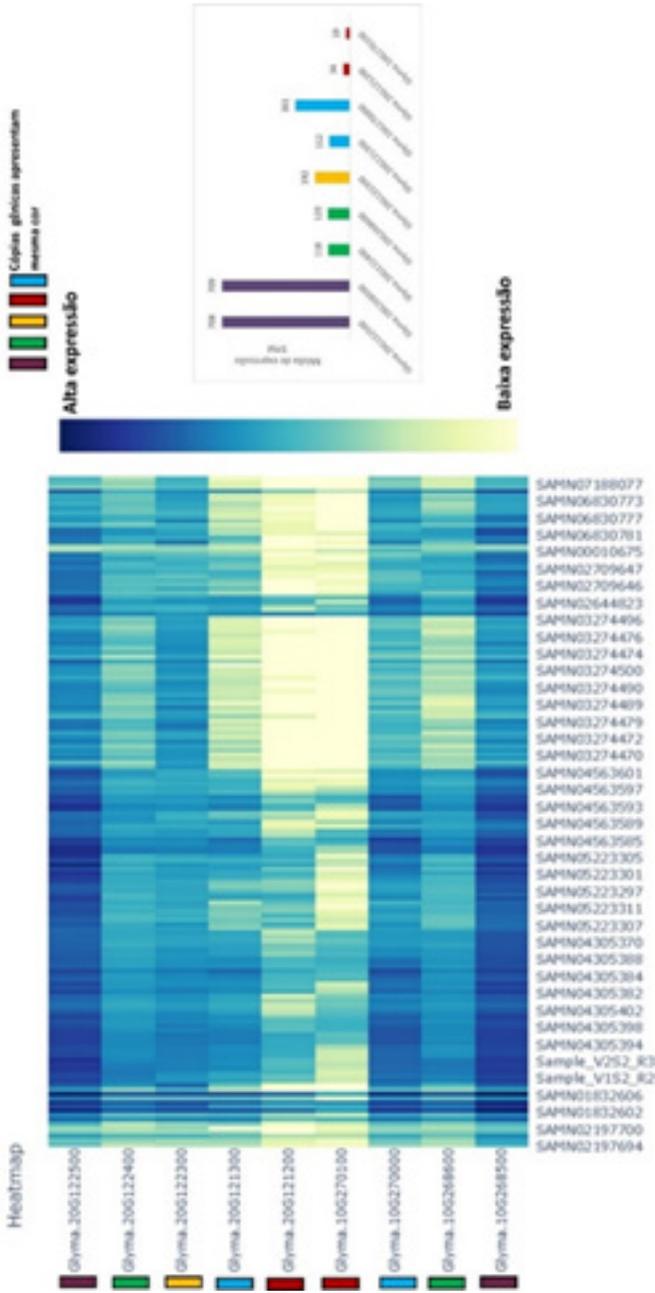


Figura 3. Heatmap contendo as expressões em transcritos por milhão-TPM dos cinco genes em 158 bioensaios de sementes obtidos a partir do banco de dados Soybean Expression Atlas (Machado et al., 2020). O gráfico de barras representa as médias em TPM dos 158 ensaios avaliados.

Conclusão

Os genes que codificam para as enzimas Diacilglicerol difosfato fosfatase e Procollagen-lysine 5-dioxygenase atuam no citoplasma enquanto o gene que codifica para enzima Fructose-bisphosphate aldolase atua no cloroplasto.

Embora o gene Glyma.20G122300 apresente alta expressão, não apresenta domínio ativo; já os Glymas.20G121200 e Glyma.10G270100 apresentam domínios ativos, porém suas expressões são menores.

Entre os genes com alta expressão em sementes para ambas as cópias e domínios ativos estão o Glyma.20G122400 (Procollagen-lysine 5-dioxygenase) e o Glyma.20G122500 Fructose-bisphosphate aldolase, podendo ser alvos interessantes para edição genômica a fim de aumentar quantidade e qualidade da proteína.

Referências

EMANUELSSON, O.; BRUNAK, S.; VON HEIJNE, G.; NIELSEN, H. Locating proteins in the cell using TargetP, SignalP and related tools. **Nature Protocols**, v. 2, n. 4, p. 953-971, 2007.

KIM, M.; SCHULTZ, S.; NELSON, R. L.; DIERS, B. W. Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15. **Crop Science**, v. 56, n. 1, p. 219-225, 2016.

MACHADO, F. B.; MOHARANA, K. C.; ALMEIDA-SILVA, F.; GAZARA, R. K.; PEDROSA-SILVA, F.; COELHO, F. S.; GRATIVOL, C.; VENANCIO, T. M. Systematic analysis of 1298 RNA-Seq samples and construction of a comprehensive soybean (*Glycine max*) expression atlas. **The Plant Journal**, v. 103, n. 5, p. 1894-1909, 2020.

PATIL, G.; MIAN, R.; VUONG, T.; PANTALONE, V.; SONG, Q.; CHEN, P.; SHANNON, G. J.; CARTER, T. C.; NGUYEN, H. T. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. **Theoretical and Applied Genetics**, v. 130, n. 10, p. 1975-1991, 2017.

PATIL, G.; VUONG, T. D.; KALE, S.; VALLIYODAN, B.; DESHMUKH, R.; ZHU, C.; WU, X.; BAI, Y.; YUNGBLUTH, D.; LU, F.; KUMPATLA, S.; SHANNON, J. G.; VARSHNEY, R. K.; NGUYEN, H. T. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. **Plant Biotechnology Journal**, v. 16, n. 11, p. 1939-1953, 2018.

VAN, K.; McHALE, L. K. Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] seed. **International Journal of Molecular Sciences**, v. 18, n. 6, 1180, 2017. 21 p.

WATERHOUSE, A. M.; PROCTER, J. B.; MARTIN, D. M.; CLAMP, M.; BARTON, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. **Bioinformatics**, v. 25, n. 9, p. 1189-1191, 2009.

ZHANG, J.; WANG, X.; LU, Y.; BHUSA, S. J.; SONG, Q.; CREGAN, P. B.; YEN, Y.; BROWN, M.; JIANG, G. L. Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. **Molecular Plant**, v. 11, n. 3, p. 460-472, 2018.