

Caracterização *in silico* das sHSPs de soja

LOPES, VALÉRIA S.¹; DE CARVALHO, MAYRA C. C. G.²; DIAS, WALDIR P.²; MARCELINO-GUIMARÃES, FRANCISMAR C.². ¹Universidade Estadual de Londrina, ²Empresa Brasileira de Pesquisa Agropecuária. Embrapa Soja, Caixa Postal, 231, 86001-970, Londrina, Paraná.
e-mail: valopes@cnpso.embrapa.br

Introdução

As proteínas de choque térmico são divididas em proteínas de alta massa molecular (60-100D) e de baixa massa molecular (16-42 kDa) ou *small heat shock proteins* (sHSPs) (SARKAR et al., 2009). As sHSPs são chaperonas moleculares capazes de impedir a ação de desnaturases nas células. Desempenham ainda função no dobramento de proteínas nascentes e no redobramento das desnaturadas. As sHSPs vegetais são codificadas por uma família multigênica nuclear sendo sua principal característica a presença do conservado de 80-100 aminoácidos na região C-terminal denominado domínio α -cristalino (α -cristalina domain – ACD) (SARKAR et al., 2009). Além disso, a região *upstream* à sequência codante das sHSP apresenta comumente várias repetições do domínio 5'nGAAn3' (HSE, *heat shock element*) que são reconhecidas e ativadas pelos *heat shock factors* ou HSF (SARKAR et al., 2009).

As sHSPs são frequentemente associadas à resposta ao estresse por fatores abióticos e, mais recentemente, tem sido também associadas a resposta a nematóides (FUGANTI et al., 2010; KANDOTH et al., 2011). Ainda não se conhece os fatores específicos que desencadeiam a expressão dessas proteínas durante o estresse biótico, mas as próprias alterações metabólicas decorrentes do ataque de patógenos podem gerar estímulos indutores muito parecidos aos encontrados nas vias de estresse abiótico (BARCALA et al., 2008).

Dada a diversificação funcional das sHSPs, essas proteínas são tidas como alvo ideal para a geração de variedades agrícolas tolerantes a condições de estresses abióticos e bióticos. Isso é especialmente importante na cultura da soja se considerarmos sua ampla distribuição. Até o momento, uma única sHSP de soja cuja expressão varia entre plantas resistentes e suscetíveis a *M. javanica* foi identificada (FUGANTI et al., 2010) mas, certamente, a caracterização da família de GmHSPs deverá auxiliar na identificação de outros candidatos potenciais para o desenvolvimento de cultivares mais produtivas em condições desfavoráveis. Neste sentido, o presente trabalho pretendeu fazer uma caracterização *in silico* das regiões codificadoras e reguladoras das sHSPs de soja e identificar padrões de *cis* elementos potencialmente envolvidos na resposta a nematóides.

Material e Métodos

As small HSPs de soja foram identificadas no banco de dados Superfamily e através das ferramentas blastn e tblastn contra o genoma da soja. O software Mega 4.0 foi utilizado na construção da árvore filogenética pelo método *neighbor-joining*. Para as predições de domínios e localização subcelular foram utilizados os softwares MEME, PROSITE, PSORT, PREDOTAR, TargetP.

Foram consideradas até 1000pb *upstream* ao sítio de início da transcrição na busca por cis elementos no PlantCare e PLACE. Os elementos HSE foram encontrados apenas em dois

modelos gênicos nas buscas feitas no PLACE e em todos os glymas quando a busca foi realizada no PlantCare. Assim, sua caracterização foi feita pela análise dos 500pb *upstream* no software MEME. O software FrameWorker foi utilizado para buscar por modelos combinatórios de *cis* elementos comuns entre as sHSP de soja e por modelos possivelmente relacionados a indução por nematóides. Nesta última análise foram utilizados genes previamente relatados como importantes em soja em resposta a nematóides.

Resultados e Discussão

As buscas por GmsHSP resultaram em 26 modelos gênicos não redundantes que foram utilizados nas análises da região peptídica (Tabela 1). As sHSP de soja estão organizadas em 14 dos 20 cromossomos sendo os cromossomos 7, 13 e 8 os que apresentam maior número de genes sHSP (Tabela 1). Todas as sHSP dos cromossomos 7 e 13 estão filogeneticamente reunidas em um clado o que pode evidenciar a ocorrência de eventos de duplicação e/ou diploidização (Figura 1). De forma semelhante as sHSP do cromossomo 8 formaram um clado específico na árvore filogenética. A análise filogenética sugere também que os genes para sHSP nos cromossomos 10 e 20 tenham tido como ancestral comum o cromossomo 16.

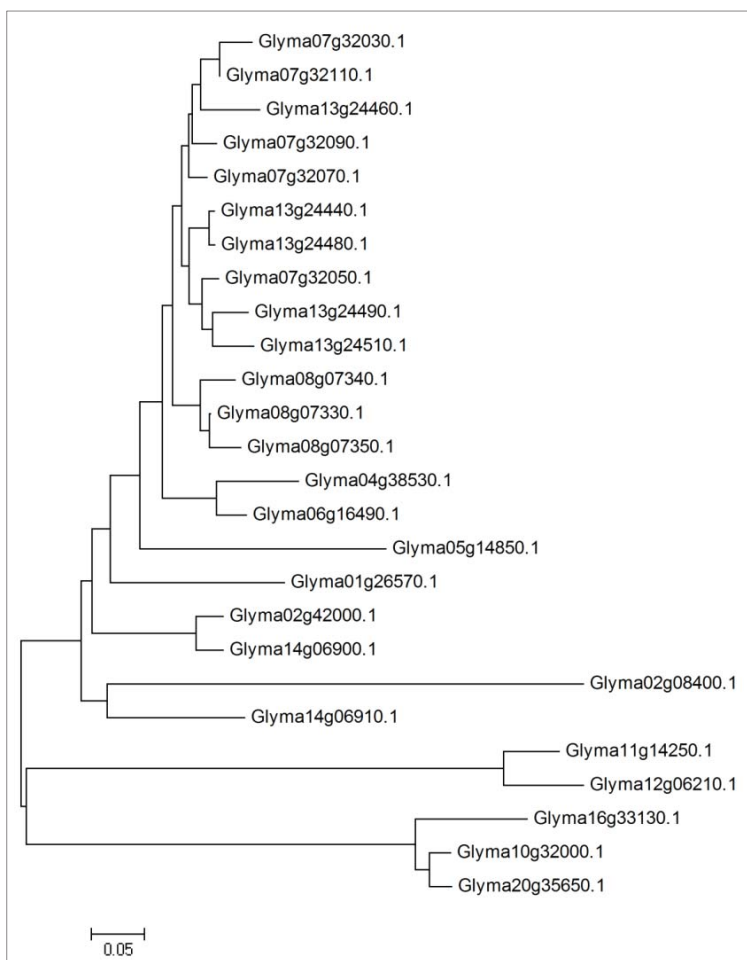


Figura 1. Análise filogenética dos glymas a partir da sequência peptídica pelo programa MEGA 4.0.

sHSP de soja que respondem a estresse biótico (Tab. 1) estão os Wbox/WRKY conhecidamente importantes na indução de genes de defesa via elicitor e os CAAT box recentemente relacionados ao aumento da expressão de sHSPs em raízes infectadas com nematóides. Os domínios HSE foram identificados pelo MEME/MAST (e value ≤ 0.001) em repetições de 1 a

Quando alinhadas as regiões codificadoras dos 26 glymas para sHSPs foi encontrada uma homologia de 40%, principalmente evidenciada na região C terminal das sequências. Nesta região foi identificada a ocorrência de uma sequência conservada de 100 aminoácidos definida como domínio ACD (MEME/MAST) com *p-value* mínimo de $7,62e-42$. Os demais softwares utilizados para predição de domínios e localização celular identificaram 21 glymas como prováveis proteínas citoplasmáticas e 4 como proteínas organelares (Tabela 1).

Dois glymas (Glyma04g14850 e Glyma07g32090) foram excluídos das análises dos *cis* elementos presentes nas sHSP de soja por não apresentarem as regiões 5'UTR e promotora no phytozome. Foi identificada a ocorrência do promotor TATAbox presente até -30 bases do sítio de início da transcrição em todas as sHSP. Entre os *cis* elementos identificados nas

3 por sequência em 16 glymas apenas com e value ≤ 0.001 (Tab. 1). Alguns glymas possuem um domínio HSE distante até 80 bases do sítio de início da transcrição, o que pode sugerir função para esses genes na resposta a infecção com nematóides já que o mesmo padrão foi evidenciado para sHSP de girassol responsivas a este estresse.

Tabela 1. Caracterização do Glymas quanto à posição cromossômica, localização na célula, presença e posição de domínios.

Glymas	Posição cromossômica	Domínio ACD	Localização	Núcleos HSEs	Posição	Domínio TA-rich	CAAT/CCAAT	Wbox/WRKY
01g26570	35.047.848 a 35.048.596	Sim	Citoplasma	1A	(-70 a -41)	—	(-379/-335/-282/-212/-112/-16)	—
02g08400	6.593.674 a 6.595.430	Sim	Citoplasma	—	—	—	(-491/-302/-122)	—
02g42000	47.097.366 a 47.098.060	Sim	Citoplasma	2B	(-65 a -51) e (-125 a -111)	—	(-475/-470/-450/-302/-293/-161/-88)	(-266)
04g38530	44.921.535 a 44.922.271	Sim	Citoplasma	2A	(-67 a -38) e (-101 a -72)	—	(-491/-302/-122)	—
06g14850	16.152.742 a 16.153.776	Sim	plastídeos/mitocôndrias	—	—	—	—	—
06g16490	12.926.962 a 12.927.793	Sim	Citoplasma	1A	(-62 a -33)	—	(-359/-231/-156)	(-398/-174/-173/-44/-172/-43)
07g32030	36.962.519 a 36.963.220	Sim	Citoplasma	1A	(-116 a -87)	—	(-303/-180)	—
07g32050	36.977.969 a 36.978.814	Sim	Citoplasma	1A e 1B	(-62 a -33) e (-120 a -106)	—	(-444)	—
07g32070	36.987.613 a 36.988.240	Sim	Citoplasma	2A	(-29 a 0) e (-61 a -32)	—	(-403/-204)	—
07g32090	36.995.756 a 36.996.604	Sim	Citoplasma	—	—	—	—	—
07g32110	37.000.908 a 37.001.669	Sim	Citoplasma	1A	(-47 a -18)	—	—	(-218)
08g07330	5.260.106 a 5.260.914	Sim	Citoplasma	1A e 1B	(-65 a -36) e (-147 a -133)	—	—	(-309/-308)
08g07340	5.263.766 a 5.264.591	Sim	Citoplasma	1A e 2B	(-67 a -38) e (-136 a -122) e (-187 a -173)	—	(-314/-277/-240/-221/-190)	—
08g07350	5.268.144 a 5.269.164	Sim	Citoplasma	2B	(-33 a -19) e (-84 a -70)	—	(-349/-137/-118/-108/-87)	—
10g32000	40.472.306 a 40.473.244	Sim	cloroplastos/mitocôndrias	2A	(-77 a -48) e (-227 a -198)	-31 a -5	(-372/-255)	(-409/-248/-408/-259)
11g14250	10.224.962 a 10.225.801	Sim	plastídeos/mitocôndrias	1A	(-49 a -20)	—	(-442/-280/-249/-44)	—
12g06210	4.235.504 a 4.236.356	Sim	Citoplasma	—	—	—	(-161/-132/-30)	—
13g24440	27.839.030 a 27.840.066	Sim	Citoplasma	—	—	—	(-496/-306/-217)	—
13g24460	27.847.751 a 27.848.553	Sim	Citoplasma	2A e 1B	(-32 a -3) e (-64 a -35) e (-441 a -427)	-199 a 183	(-422/-71)	(-7)
13g24480	27.853.841 a 27.854.840	Sim	Citoplasma	—	—	—	(-485/-484/-463/-335/-248)	(-484/-379)
13g24490	27.862.859 a 27.863.745	Sim	Citoplasma	1A e 1B	(-61 a -32) e (-119 a -105)	—	(-263)	(-466/-439/-333)
13g24510	27.869.587 a 27.870.386	Sim	Citoplasma	1A	(-66 a -37)	—	(-479/-158)	(-401/-345)
14g06900	5.086.252 a 5.086.956	Sim	Citoplasma	3B	(-61 a -47) e (-125 a -111) e (-187 a -173)	—	(-495/-379/-220/-178)	(-291)
14g06910	5.087.964 a 5.088.623	Sim	Citoplasma	3A	(-100 a -71) e (-165 a -136) e (-224 a -195)	—	(-240/-76/-58)	(-73)
16g33130	36.161.134 a 36.161.989	Sim	plastídeos/mitocôndrias	2A e 1B	(-152 a -123) e (-201 a -172) e (-248 a -234)	—	(-398/-273)	—
20g35650	43.885.724 a 43.886.516	Sim	plastídeos/mitocôndrias	1A	(-215 a -186)	—	—	(-23)

Núcleos HSEs: (A)=HSE I e (B)=HSE II.

Entre as sHSP de soja foram encontradas regiões ricas em TA nas posições -31 e -199 dos glymas Glyma10g32000, (TA)₁₃, e Glyma13g24460, (TA)₉ respectivamente; os quais são induzidos em raízes de plantas de soja infectadas com nematóides (FUGANTI et al., 2010; KANDOTH et al., 2011). Apenas 13 sequências de 24 avaliadas foram incluídas no modelo com um único elemento HSE distante até -300pb (p -value 1,67e-8). Este resultado corrobora com a análise realizada pelo MEME, demonstrando haver variação no posicionamento dos elementos HSE entre as GmsHSP. O modelo formado sugere a ocorrência de 3 padrões distintos entre as sHSP, um formado pelo glyma 01g26570, outro formado pelos glymas 13g24510, 08g07350, 08g07330, e o último formado pelos glymas 06g16490, 07g32070, 07g32110, 16g33130, 20g35650, 08g07340, 10g32000 e 13g24460 (para os glymas sublinhados há evidência de indução por estresse com nematóides).

Conclusões

As análises *in silico* sugerem que as sHSP de soja sejam representadas por 26 genes localizados em 14 cromossomos. Todas as sequências avaliadas são anotadas como sHSPs, mas os resultados obtidos sugerem que possivelmente alguns desses genes sejam proteínas com domínio ACD que não fazem parte da família de sHSPs. Apenas 16 dos 26 genes analisados possuem ao menos um elemento HSE, mas ensaios experimentais são fundamentais para a avaliação da resposta desses genes ao "heat shock". As sHSP de soja responsivas ao estresse de infecção por nematóides apresentam o padrão específico de disposição de elementos HSE, CAATbox e Wbox previamente reportado para genes de sHSP de outras espécies.

Referências

BARCALA, M.; GARCIA, A.; CUBAS, P.; ALMOGUERA, C.; JORDANO, J.; FENOLL, C.; ESCOBAR, C. Distinct heat-shock element arrangements that mediate the heat shock, but not the late-embryogenesis induction of small heat-shock proteins, correlate with promoter activation in root-knot nematode feeding cells. **Plant Molecular Biology**. v.66, p.151–164, 2008.

FUGANTI, R.; MACHADO, M. F. P. S.; LOPES, V. S.; SILVA, J. F. V.; ARIAS, C. A. A.; ROCKENBACH-MARIN, S. R.; BINNECK, E.; ABDELNOOR, R. V.; MARCELINO, F. C.; NEPOMUCENO, A. L. Size of AT(n) insertions in promoter region modulates Gmhsp17.6-L mRNA transcript levels. **Journal of Biomedicine and Biotechnology**. v.10, 2010.

KANDOTH, P. K.; ITHAL, N.; RECKNOR, J.; MAIER, T.; NETTLETON, D.; BAUM, T. J.; MITCHUM, M. G. The soybean Rhg1 locus for resistance to the soybean cyst nematode *Heterodera glycines* regulates the expression of a large number of stress- and defense-related genes in degenerating feeding cells. **Plant Physiology**. v. 155, p. 1960–1975, 2011.

SARKAR, N. K.; KIM, Y-K.; GROVER, A. Rice sHsp genes: genomic organization and expression profiling under stress and development. **BMC Genomics**. v.10, n.393, p.1471-2164, 2009.