

MINERAÇÃO DE MOTIVOS SSR EM SEQUÊNCIAS EXPRESSAS DE FEIJÃO-CAUPI CONTENDO TAGS SUPERSAGE DIFERENCIALMENTE EXPRESSAS SOB DESIDRATAÇÃO RADICULAR

Renato Santos¹; Carolina Basante²; José R. C. Ferreira Neto³; Valesca Pandolfi³; Maurisrael de M. Rocha⁴; Kaesel J. Damasceno e Silva⁴; Alexandre Nepomuceno⁵; Rosana P. V. Brondani⁶; Ana M. Benko-Iseppon³; Ederson A. Kido³

¹Biólogo, graduando, Universidade Federal de Pernambuco, Av. Prof. Moraes Rêgo S/N, Recife, PE. E-mail: scalk2004@hotmail.com.

²Biólogo, graduando, Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros S/N, Recife, PE.

³Pesquisador, Universidade Federal de Pernambuco, Av. Prof. Moraes Rêgo S/N, Recife, PE.

⁴ Pesquisador, Embrapa Meio-Norte, Av. Duque de Caxias, 5650, Teresina, PI.

⁵ Pesquisador, Embrapa Soja, Rodovia Carlos João Strass, Londrina, PR.

⁶ Pesquisador, Embrapa Arroz e feijão, Rodovia GO-462, km 12, Sto Antônio de Goiás, GO.

Resumo – O feijão-caupi é uma planta importante na nutrição e na economia de pequenos produtores das regiões Norte e Nordeste do Brasil. Entretanto, sua produtividade ainda é pequena, quando comparada com seu potencial e o de outras leguminosas. Uma das alternativas para auxiliar o melhoramento genético da cultura e associar tecnologias modernas, como o desenvolvimento de marcadores moleculares do tipo microssatélites (SSR) e a identificação de genes importantes em resposta a estresses, via estudos de transcriptômica, proporcionados pela técnica SuperSAGE. Nesse sentido, *tags* SuperSAGE de 26 pb, de genótipo tolerante a desidratação radicular, que foram observadas sendo induzidas após aplicação do estresse (até 150 min), em relação ao controle sem estresse, foram ancoradas em ESTs de feijão-caupi. Essas ESTs contendo as *tags* induzidas foram mineradas quanto à presença de motivos SSR. Os motivos identificados em 135 ESTs permitiram desenhar pares de *primers* para desenvolvimento de marcadores moleculares SSR, agora baseados em DNA, e não mais anônimos, visando aproveitar esse tipo de polimorfismos para genotipagem de materiais do melhoramento. Sua associação com resposta a estresses está sendo investigada.

Palavras-chave: *Vigna unguiculata*, estresse abiótico, biblioteca genômica, bioinformática

Introdução

O feijão caupi [*Vigna unguiculata* (L.) Walp] é uma leguminosa importante na alimentação humana por ser uma excelente fonte de proteínas e aminoácidos essenciais. Dada à sua ampla utilização, possui importância econômica e agrônômica no Brasil, principalmente nas regiões Norte e Nordeste. A espécie apresenta ainda um pool gênico diverso, com ampla plasticidade, capaz de tolerar uma gama de condições adversas. Apesar disso, sua produtividade está ainda longe de ser alcançada em relação ao potencial existente.

A Rede NordEST representa um esforço conjunto e pioneiro de diversas instituições em realizar a genômica funcional do feijão-caupi. Para tanto, estudos abrangendo estresses bióticos (CABMV e CMV) e abióticos (desidratação radicular e exposição ao sal NaCl), com sequenciamento de ESTs (*Expressed Sequence Tags*) e de *tags* SuperSAGE (*SuperSerial Analysis of Gene Expression*), permitiram análises globais dos transcriptomas. Dentre as metodologias de transcriptômica, a técnica de SuperSAGE representa uma das mais eficientes, sendo de arquitetura aberta, com altos rendimentos, especificidades e sensibilidade (TERAUCHI *et*

al., 2008). O emprego dessa técnica, na identificação de genes que confirmam melhor tolerância aos estresses estudados, poderá proporcionar melhor entendimento dos mecanismos envolvidos, com possibilidades de aplicações no melhoramento da espécie. Marcadores moleculares de DNA também é outra ferramenta importante para uso em programas de melhoramento. Dentre esses, aqueles do tipo microssatélites (SSR; *Simple Sequence Repeats*), constituídas por repetições de nucleotídeos (motivos) em *tandem* (ex: ATATAT), são um dos mais informativos, amostrando regiões altamente polimórficas dos genomas. Em ESTs, a presença de motivos SSR é significativa, estando presentes em transcritos de enzimas metabólicas, proteínas estruturais e de armazenamento, fatores de transcrição, dentre outros. Essa presença sugere que algumas dessas repetições podem estar envolvidas em processos de regulação metabólica e de evolução gênica (LI *et al.*, 2004).

O presente trabalho teve como objetivo a busca de motivos SSR em ESTs contendo *tags* SuperSAGE diferencialmente expressas em feijão caupi após estresse de desidratação radicular por até 150 minutos.

Material e Métodos

As *tags* disponíveis para o trabalho foram geradas de bibliotecas SuperSAGE oriundas de genótipos tolerante e sensível, selecionados e fornecidos pela Embrapa Meio-Norte (Teresina, PI), em experimento de desidratação radicular, com coleta de amostras a cada 25 minutos (até 150 minutos), após suspensão de cultivo hidropônico em cada de vegetação, com plantas de 3 meses, conduzido pela Embrapa Soja (Londrina, PR). As bibliotecas foram geradas conforme MOLINA *et al.* (2008) e sequenciadas por intermédio da empresa GenXPro GmHb (Frankfurt, Alemanha). *Tags* foram analisadas com auxílio do software DiscoverySpace 4.0 (ref) e classificadas em induzidas ou reprimidas, ao nível de 5% de probabilidade ($p < 0,05$). Valores referentes à modulação da expressão (FC, *fold change*) foram obtidos através da razão entre frequências observadas para uma mesma *tag*, nas bibliotecas com estresse em relação ao controle sem estresse. *Unitags* (*tags* diferentes) foram ancoradas (BlastN) em ESTs de feijão-caupi clusterizadas pela rede NordEST e anotadas a partir de transcritos de *Phaseolus vulgaris* da base de dados Phytozome (<http://www.phytozome.net/>). Foram considerados somente alinhamentos BlastN (*tag* – EST) perfeitos ou com somente um erro em 26 pb, não podendo este ocorrer nas quatro bases iniciais que caracteriza a *tag* (CATG). ESTs contendo *unitags* diferencialmente expressas, que foram induzidas no genótipo tolerante (quando sob estresse, em comparação ao respectivo controle negativo) e também reprimida ou não significativa ao nível de 5%, ou mesmo ausente na comparação do genótipo sensível (sob estresse em relação ao controle sem estresse), foram mineradas quanto à presença de SSR. Este foi definido com tendo um número mínimo de 9, 6, 5, 5, 4 repetições, respectivamente, para motivos de di, tri, tetra, penta e hexanucleotídeos, identificados a partir da ferramenta *on line SSRfinder* (<http://www.fresnostate.edu/ssrfinder/>). Nesta detecção, as 30 primeiras bases (extremidade 5') e as 30 últimas bases (3') foram desconsideradas, para facilitar o posterior desenho de *primers*, o qual foi feito usando-se a ferramenta *Primer3* (<http://frodo.wi.mit.edu/>), com vistas ao futuro desenvolvimento de marcadores moleculares funcionais, do tipo SSR.

Resultados e Discussão

De 5820 ESTs (*contigs* e *singlets*) que alinharam com *tags* SuperSAGE induzidas no genótipo tolerante sob estresse (em comparação ao controle sem estresse), um total de 135 mostraram motivos SSR detectados (Tabela 1). Os motivos mais abundantes foram de trinucleotídeos (56,3 %), seguidos por aqueles de dinucleotídeos (26,7 %), enquanto que os motivos de tetranucleotídeos, pentanucleotídeos e hexanucleotídeos foram os menos frequentes nas sequências analisadas. A predominância dessas repetições em seqüências

expressas de vegetais tem sido relatada (MUN *et al.*, 2006). MORGANTE *et al.* (2002), investigando SSR em diferentes frações genômicas de *Arabidopsis*, observaram que a frequência relativa de trinucleotídeos duplicou em regiões codificantes, quando comparada com frações não-codificantes, possivelmente como resultado de pressão de seleção (METZGAR *et al.*, 2000). Motivos de dinucleotídeos foram os que apresentaram a maior quantidade de repetições (até 23) e maior extensão média, sendo os do tipo CT/TC os mais abundantes (Tabela 1); motivos trinucleotídeos, por sua vez, apresentaram até nove repetições e extensão média de 6,6 repetições (cerca de 20 pb), sendo que o motivo AGA/GAA/AAG, foi considerado o mais frequente (Tabela 1). A extensão de SSRs é uma das informações mais importantes sobre esses marcadores, pois quanto maior essa variável maior seu potencial polimórfico, aumentando assim a informatividade do marcador.

Tabela 1. Caracterização dos SSRs presentes em ESTs contendo *unitags* SuperSAGE diferencialmente expressas em feijão-caupi submetido à desidratação por até 150 min., em sistema hidropônico.

Motivo SSR	Número de ESTs (%)	Número de repetições de motivos SSR (mínimo - máximo)	Média de repetições de motivos SSR	Motivo SSR mais abundante (frequência)
Dinucleotídeo	36 (26,7)	9 – 23	11,9	*CT/TC (17)
Trinucleotídeo	76 (56,3)	6 – 9	6,6	*AGA/GAA/AAG (13)
Tetranucleotídeo	10 (7,4)	5	5,0	*CTTT/TTTC (3)
Pentanucleotídeo	3 (2,2)	5 – 6	5,3	uniforme
Hexanucleotídeo	10 (7,4)	4 – 6	4,6	uniforme
Total	135 (100%)	-	-	-

*Combinações de nucleotídeos para diferentes motivos SSR foram agrupadas em classes únicas, considerando-se a complementaridade de bases do DNA. Motivos dinucleotídeos, por exemplo, foram agrupados em 4 classes únicas de repetição: (1) AT/TA; (2) AG/GA / TC/CT; (3) AC/CA /TG/GT; e (4) GC/CG.

Para essas ESTs contendo motivos SSR o desenho de *primers* possibilita amplificação via PCR a partir de DNA extraído. Algumas das ESTs com *primers* disponíveis, que alinharam com *unitags* SuperSAGE, as quais foram observadas induzidas no genótipo tolerante (quando sob estresse, em comparação ao respectivo controle negativo) e reprimidas, não significativas (ao nível de 5%) ou mesmo ausentes no genótipo sensível (quando comparado o contraste sob estresse em relação ao controle sem estresse) encontram-se na Tabela 2. Para vários desses ESTs, a importância deles em respostas a estresses pode ser inferida pelas descrições de gene e/ou função disponíveis em suas anotações (ex: fatores de transcrição do tipo *WRKY*, proteínas atuantes em vias de sinalização por hormônios, do tipo Etileno, ácido indol-acético, etc) e pelos valores de FC (*fold change*; modulação da expressão) positivos, que foram estimados (Tabela 2).

A presença de SSR em ESTs associadas às *unitags* diferencialmente expressas nos genótipos contrastantes estudados representa uma rica fonte de informação, com potencial para aplicação no melhoramento vegetal de feijão-caupi, pois possibilita o desenvolvimento de dois tipos de marcadores: aqueles baseados em RNA, devendo ser observadas as variações de expressão a partir das *unitags* observadas, em reações de RTqPCR, estas nem sempre acessíveis ou disponíveis na maioria dos laboratórios, mesmo aqueles de biologia molecular, e aqueles marcadores baseados em DNA, mais comumente encontrados. Se constatada a importância desses SSRs na genotipagem de materiais do melhoramento, isso pode vir a ser útil para os programas de

melhoramento, uma vez que sua importância em nível de transcrição, em resposta ao estresse de desidratação radicular já fora levantada pela constatação das *tags* SuperSAGE. Deve ser ressaltado que repetições de motivos SSR podem estar envolvidos em processos de regulação gênica, além de que, variação na extensão de SSRs no interior de regiões codificantes deve ser crítica para a atividade do gene, pois expansões e contrações nessas estruturas afetam, diretamente, o produto gênico correspondente, podendo causar mudanças fenotípicas (LI *et al.*, 2004). A presença desses motivos e a associação com as respostas a estresses estão sendo investigadas, bem como sua aplicabilidade na genotipagem de materiais oriundos do melhoramento.

Tabela 2. Resultados parciais mostrando ESTs de feijão-caupi ancorando *unitags* SuperSAGE diferencialmente expressas em raízes de genótipo contrastantes (tolerante T e sensível S) de feijão-caupi submetidos à desidratação (até 150 min.), bem como anotação e motivo SSR presente, com número de repetições (Rep).

EST Contig	SuperSAGE Unitag	T		S		Anotação	(Motivo) Rep.
		FC	Reg	FC	Reg		
11347	CpD102841	51,0	UR	-1,1	ns	<i>SEUSS-like 2</i>	(CAG)7
1851	CpD72243	42,6	UR	-1,0	ns	<i>WRKY DNA-binding prot. 11</i>	(ACCGCA)4
7268	CpD142239	9,9	UR	1,1	ns	n.a	(AAC)6
10896	CpD83800	9,6	UR	1,1	ns	<i>indoleacetic acid-induced prot. 16</i>	(AGTAAC)4
12546	CpD15182	8,6	UR	1,6	ns	<i>S-ribonuclease binding prot.</i>	(ATT)6
110	CpD66444	8,0	UR	-1,0	ns	<i>cell division control 2</i>	(GA)12
6090	CpD88962	6,1	UR	-1,0	ns	<i>Ethylene insensitive 3 fam. prot.</i>	(AT)9
6517	CpD12763	5,7	UR	-1,2	ns	n.a	(AGAAG)5
9491	CpD102796	5,1	UR	-1,2	ns	<i>Ethylene insensitive 3 fam. prot.</i>	(ACA)7
3666	CpD79822	49,0	UR	-1,7	DR	<i>UDP-D-glucuronate 4-epimerase 6</i>	(TC)9
2958	CpD121944	24,9	UR	-5,1	DR	<i>20S proteasome alpha subunit C1</i>	(CT)18
3631	CpD1568	17,8	UR	-3,7	DR	<i>vesicle-associated membrane prot. 724</i>	(CT)23
240	CpD134490	13,9	UR	-2,6	DR	<i>plant UBX domain containing prot. 4</i>	(CT)15
4961	CpD69643	13,7	UR	-2,7	DR	n.a	(CT)9
517	CpD109360	12,4	UR	-1,5	DR	<i>SAM domain-containing prot.</i>	(CT)9
12017	CpD136091	12,0	UR	-1,7	DR	<i>Survival prot. SurE-like phosphatase</i>	(TTC)6
11606	CpD81629	11,5	UR	-1,9	DR	<i>glycosyl hydrolase 9C2</i>	(TCT)7

FC: razão entre frequências de uma mesma *tag* nas bibliotecas com estresse sem estresse; Reg: regulação; UR: *tags* induzidas em relação ao controle sem estresse; DR: *tags* reprimidas; n.s: não significativa ($p < 0,05$)

Conclusões

A mineração de dados detectou motivos SSR em sequências ESTs de feijão-caupi que ancoravam *tags* SuperSAGE consideradas induzidas no genótipo tolerante, quando sob estresse de desidratação radicular de até 150 min., em relação ao controle sem estresse. Para estes alvos moleculares, além da possibilidade de desenvolvimento de marcadores moleculares com base em RNAs e nas *tags* detectadas, os motivos SSR detectados possibilitam o desenvolvimento de marcadores moleculares do tipo SSR, baseados em DNA, e funcionais, não sendo mais anônimos, com potencial para explorar polimorfismos decorrentes dessa classe de marcadores, agora associada com a resposta diferencial ao estresse em questão.

Agradecimentos

Os autores agradecem à Embrapa e ao CNPq, pelo financiamento e bolsas, e as unidades Embrapa Arroz e feijão, Embrapa Meio-Norte e Embrapa Soja pelos apoios diversos de seus pesquisadores.

Referências

- TERAUCHI, R.; MATSUMURA, H.; KRÜGER, D. H.; KAHL, G. SuperSAGE: The most advanced transcriptoma technology for functional genomics. In: KAHL, G.; MEKSEM, K. (eds.). **The handbook of plant functional genomics**. 1^a ed. Weinheim: Wiley-VCH Verlag GmbH & Co, cap. 3, p. 37-54, 2008.
- LI, Y-C.; ABRAHAM, B.K.; FAHIMA, T.; NEVO, E. Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution* v.21, p. 991-1007, 2004.
- MOLINA, C.; ROTTER, B.; HORRES, R.; UDUPA, S.M.; BESSER, B.; BELLARMINO, L.; BAUM, M.; MATSUMURA, H.; TERAUCHI, R.; KAHL, G.; WINTER, P. SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. *BMC Genomics*, v.9, n.1, p.553-580, 2008.
- MUN, J-H.; KIM, D-J.; CHOI, H-K.; GISH, J.; DEBELLÉ, F.; MUDGE, J.; DENNY, R.; ENDRE, G.; SAURAT, O.; DUDEZ, A-M.; KISS, G.B.; ROE, B.; YOUNG, N.D.E.; COOK D.R. Distribution of microsatellites in genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics* , v.172, p.2541-2555, 2006.
- MORGANTE, M.; HANAFEY, M.E.; POWELL, W. Microsatellite are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, v.30, n.2, p.194-200, 2002.
- METZGAR, D.; BYTOF, J.; WILLS, C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research*, v. 10, n.2, p.72-80, 2000.