

Identificação por RNAseq e validação por qPCR de genes diferencialmente expressos em sorgo em resposta ao estresse hídrico

Beatriz de Almeida Barros⁽¹⁾; Andrea Almeida Carneiro⁽²⁾; Paloma Alessandra Alves⁽³⁾; Raquel Oliveira Moreira⁽³⁾; Meire de Cassia Alves⁽¹⁾; Roberto Willians Noda⁽²⁾; Newton Portilho Carneiro⁽²⁾

⁽¹⁾ Analista de Pesquisa e Desenvolvimento, Embrapa Milho e Sorgo, Sete Lagoas, MG; ⁽²⁾ Pesquisador, Embrapa Milho e Sorgo, Sete Lagoas, MG; ⁽³⁾ Estagiária, Embrapa Milho e Sorgo, Sete Lagoas, MG. newton.carneiro@embrapa.br

RESUMO: Um dos principais fatores que mais afeta a produção agrícola é o estresse hídrico. Desenvolvimento de cultivares de sorgo mais tolerantes ao estresse hídrico é uma das principais alternativas para um aumento de produtividade sob baixa disponibilidade de água. A tolerância à seca é uma característica complexa e afeta as respostas de genes em muitos processos biológicos em plantas. Nesse trabalho foram comparados os níveis de expressão de genes por RNAseq e qPCR de variedades de sorgo contrastantes para a tolerância ao estresse hídrico e em duas condições hídricas. O RNA total foi extraído de raízes de plantas de sorgo mais tolerantes na presença (50% menos irrigado) e ausência (100% controle irrigado) do estresse. Os cDNAs foram construídos, sequenciados e analisados utilizando a plataforma Galaxy (CNPTIA). Genes com diferenças de expressão superior a dez vezes em condições contrastantes em análise de RNAseq e de baixo número de cópia foram submetidos ao qPCR. Os resultados mostraram consistentes apenas com os genes mais expressos. Houve grande divergência com os genes menos e igualmente expressos entre os tratamentos contrastes com variações chegando superior a 30 vezes entre os dois métodos. Apesar da divergência esses resultados ajudam a consolidar os resultados de RNAseq de expressão de genes relacionados com estresse hídrico e um maior conhecimento das vias metabólicas relacionados a esse processo.

Termos de indexação: sorgo, seca, RNASeq, qPCR.

INTRODUÇÃO

Plantas sujeitas ao estresse na maioria das vezes combinam uma série de respostas fisiológicas e bioquímicas a nível molecular, celular e em toda a planta (Bohnert et al., 1995; Bray, 1993, 1997;

Chaves et al., 2003) que podem levar a alterações na planta que permitam uma maior tolerância a curto e longo prazo ao estresse hídrico. Essas alterações a curto prazo podem variar desde um ajuste osmótico (Morgan, 1984) através do sequestro de íons na planta (Mimura et al, 2003) até um sinal raiz-folha que possa reduzir a transpiração e o crescimento da planta. A longo prazo a planta pode modificar o tamanho e arquitetura do sistema radicular capacitando a planta de absorver maiores quantidades de água durante condições de estresses hídrico. Pouco se conhece sobre as redes regulatórias que traduzem as mudanças ambientais em alterações metabólicas necessárias durante as respostas de adaptação à seca. Um grande número de genes é expresso diferencialmente sob seca em diferentes tecidos, estádios e condições (Reynolds & Tuberosa 2008; Deyholos 2009) e a identificação desses genes responsivos ao déficit hídrico se torna fundamental para o desenvolvimento de linhagens tolerantes, cuja produção possa minimizar os efeitos da seca. Neste trabalho, mostramos a identificação de alguns desses genes diferencialmente expressos em resposta ao estresse hídrico em sorgo por RNAseq e sua validação por qPCR.

MATERIAL E MÉTODOS

Material Genético

Foram utilizados dois híbridos de sorgo contrastantes para tolerância à seca, denominados: genótipo A (tolerante) e genótipo B (sensível). No pré-florescimento, foram impostos dois tratamentos hídricos: irrigado e submetido a restrição hídrica. No primeiro, a reposição hídrica foi realizada diariamente até o solo atingir a umidade próxima a CC (tensão da água no solo de aproximadamente - 18 kPa), enquanto que, no segundo tratamento, a indução do estresse hídrico foi realizada

diariamente aplicando-se no máximo 50% da água total disponível, ou seja, até a tensão da água no solo atingir no mínimo -138 kPa para o solo especificado. Esse estresse foi mantido por 10 dias.

Extração de RNA e síntese de cDNA

Antes da extração, cada amostra foi lavada rapidamente em água para retirada do solo residual e macerada em nitrogênio líquido. O RNA total foi extraído utilizando o RNeasy Plant Mini Kit (Qiagen) de acordo com as recomendações do fabricante. O cDNA foi sintetizado a partir de 1,0 ug de RNA total utilizando High Capacity cDNA Reverse Transcription Kit de acordo com as recomendações do fabricante.

RNASeq

Cerca de 10 ug de RNA total de duas réplicas biológicas de cada tratamento foram enviados para a empresa Eurofins para sequenciamento. As análises de sequências de RNAseq de sorgo utilizaram o programa TopHat (<http://tophat.cbcb.umd.edu/>) que alinha sequências do genoma e determina splicing alternativos. O Cufflinks (<http://cufflinks.cbcb.umd.edu/>) usa o mapa para localizar os transcritos. O Cuffdiff é parte do pacote do Cufflinks que faz o alinhamento de genes em duas ou mais condições determinando genes diferencialmente expressos utilizando análises estatísticas rigorosas. O TopHat e Cufflinks são sistema baseados em UNIX. Para utilizar uma interface gráfica mais acessível utilizamos a plataforma Galaxy disponibilizado pela Embrapa Informática Agropecuária. O CummeRbund (<http://compbio.mit.edu/cummeRbund/>) produz informações de expressão nos formatos gráficos de vulcano, scatter e boxplots utilizando os dados do Cufflinks, em conexão com o ambiente de estatística R e pode ser acessível pelo Bioconductor (<http://www.bioconductor.org/>).

qRT-PCR

Para a escolha dos transcritos diferencialmente expressos de sorgo e sob condições contrastantes para estresse hídricos (100% Reidratado x 50% Reidratado), das bibliotecas de raiz (A-Root-100_x_A-Root-50) foram elencados os transcritos com as maiores expressões diferenciais [\log_2 (Fold Change)] positivas (Up regulation) e negativas (Down regulation), e os transcritos com expressões mais semelhantes (no significant differential expression = NoDiff), excetuando-se os transcritos sem expressão detectada e os transcritos sem teste estatístico. Para as categorias up e down regulated, também foram excluídos os transcritos sem significância estatística da expressão diferencial. Os

transcritos sem splicing alternativo (um gene com um único transcrito) foram utilizados para o qRT-PCR. Na tabela 1 é mostrado aqueles genes que foram comparados pelos dois métodos.

Os primers foram desenhados para todos os genes selecionados utilizando o software PrimerExpress (Applied Biosystems). As reações de PCR em tempo real foram conduzidas em 7500 Fast Real Time PCR System (Applied Biosystems) utilizando Fast Syber Green Master Mix (Applied Biosystems) de acordo com as recomendações do fabricante. Um gene que codifica actina foi utilizado como gene de referência para ambos os materiais e quantificação relativa foi calculada utilizando o método $\Delta\Delta C_t$ (Livak e Schmittgen, 2001).

RESULTADOS E DISCUSSÃO

As análises de sequências de RNAseq de sorgo foram feitas na plataforma Galaxy e apresentaram os números de reads por amostra da Tabela 1.

Tabela 1: Reads por amostra.

Amostras	# Reads	# Reads mapeados (%)	# Reads únicos mapeados (%)
A1-Root-100	116.579.378	102.243.966 (0.88)	90.834.215 (0.80)
A1-Root-50	89.855.346	81.632.879 (0.91)	73.009.796 (0.81)
B1-Root-100	179.914.946	160.446.644 (0.89)	143.547.667 (0.80)
B1-Root-50	122.522.510	105.978.571 (0.87)	92.389.881 (0.75)
A3-Root-100	107.365.688	98.555.039 (0.92)	85.399.469 (0.79)
A3-Root-50	79.233.974	61.298.435 (0.77)	41.601.856 (0.52)
B3-Root-100	123.638.136	114.598.181 (0.93)	98.183.283 (0.79)
B3-Root-50	120.017.974	93.215.627 (0.77)	79.800.483 (0.66)
A1-Leaf-100	100.936.758	95.953.732 (0.95)	87.141.552 (0.86)
A1-Leaf-50	96.462.640	91.790.319 (0.95)	82.848.490 (0.86)
B1-Leaf-100	106.853.764	100.825.043 (0.94)	91.879.058 (0.86)
B1-Leaf-50	185.613.262	176.839.393 (0.95)	160.287.282 (0.86)
A3-Leaf-100	219.889.864	209.038.068 (0.95)	190.179.049 (0.86)
A3-Leaf-50	110.288.924	104.385.604 (0.95)	94.117.866 (0.85)

A quantidade de genes diferencialmente expressos é apresentada na Tabela 2.

Tabela 2: Genes diferencialmente expressos.

Condição	# genes
A leaf 100 X A leaf 50	540
B leaf 100 X B leaf 50	3488
A leaf 100 X B leaf 100	3592
A leaf 50 X B leaf 50	3496

O BiNGO (Maere et al., 2005), um aplicativo para Cytoscape 3.0.2 (Shannon et al., 2003), foi utilizado para determinar quais categorias são diferencialmente representados, com base estatística, nos três domínios (Componente Celular, Função Molecular e Processo Biológico) do *Gene Ontology* (GO). Na Figura 1, observa-se as categorias diferencialmente representadas, destacadas com cores para o domínio Processo Biológico. Resulta da comparação entre as bibliotecas de folhas do genótipo A, com 100% e

50% de reidratação, com 540 genes diferencialmente expressos, sendo 252 superexpressos e 288 subexpressos. A variação de cor (de amarelo a vermelho) é relativa a menor diferença (amarelo) ou a maior diferença (vermelho) de representação estatística entre as bibliotecas, ou seja, quanto maior a diferença estatística entre as quantidades de genes de uma dada categoria GO em cada biblioteca, mais próxima do vermelho será a cor de preenchimento do círculo da categoria no gráfico. As categorias diferencialmente representadas estão relacionadas com respostas a estímulos, transporte e processos metabólicos para o domínio Processo Biológico. Para o domínio de Função Molecular, as categorias mais diferencialmente representadas são atividades catalíticas e algumas categorias de ligação, como ligação a ferro, a íons metálicos, a polissacarídeos e outros. Em Componente Celular, as categorias ligadas a organelas, citoplasma e região intracelular foram as diferencialmente representadas.

Foram elencados 72 transcritos para desenho do primer e reação de qPCR, conforme Tabela 3. Todos os resultados de superexpressão (Up regulated) que apresentaram resultados de qPCR, confirmaram a tendência de aumento de expressão dos transcritos em situação de restrição hídrica. Para os resultados sem diferença de expressão apenas um resultado confirmou os dados de RNAseq.

Tabela 3 – Número de transcritos diferencialmente expressos selecionados com RNAseq, reações que apresentaram resultados de qPCR e resultados confirmados.

Categoria	RNAseq	qPCR	RNAseq=qPCR
50-100_NoDiff	37	31	1
50-100_Up	18	12	12
50-100_Down	17	13	7

A comparação da expressão pelo método de RNAseq e qPCR mostrou grande variação no sorgo, principalmente para a categoria sem diferença de expressão (*NoDiff*) entre as amostras 100% e 50% reidratadas. As figuras 2 e 3 mostram os níveis de expressão do qPCR para alguns transcritos.

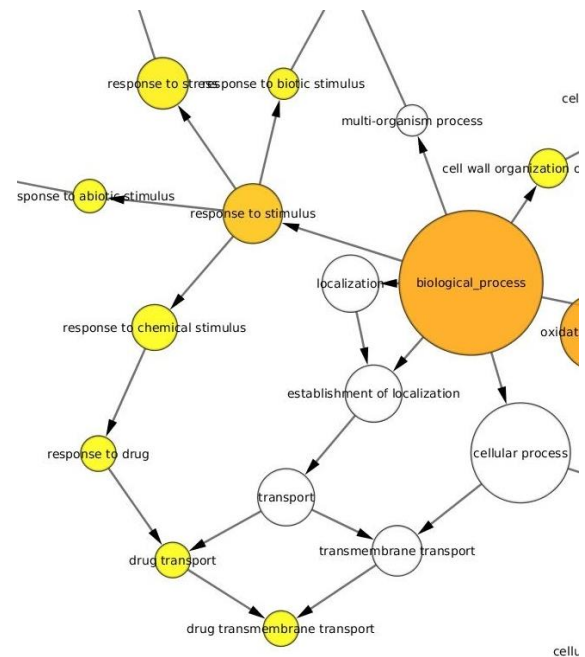


Figura 1: Parte da imagem da análise de categorias (destacadas com cores) do domínio Processo Biológico do *Gene Ontology* diferencialmente representadas entre as bibliotecas de folhas do genótipo A, com 100% e 50% de reidratação.

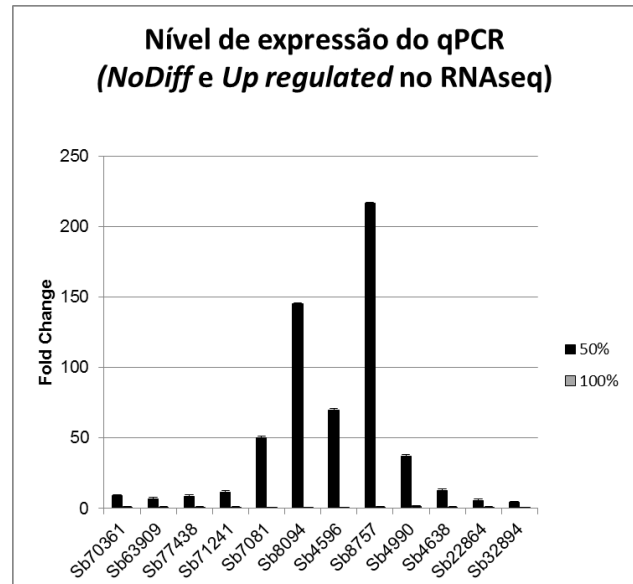


Figura 2: qPCR - Nível de expressão dos transcritos. *NoDiff* em RNAseq: Sb70361, Sb63909, Sb77438, Sb71241. *Up regulated*: Sb7081, Sb8094, Sb4596, Sb8757, Sb4990, Sb4638, Sb22864, Sb32894. Os transcritos *NoDiff* apresentam tendência de aumento de expressão na amostra 50% reidratada, tornando os resultados incompatíveis entre RNAseq e qPCR. Já para os transcritos *Up regulated*, os resultados são coerentes entre as metodologias testadas.

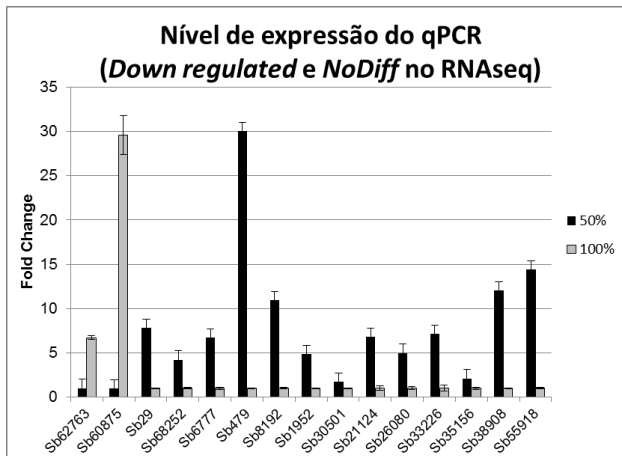


Figura 3: qPCR - Nível de expressão dos transcritos. *Down regulated*: Sb62763, Sb60875, Sb68252. *NoDiff* em RNAseq: Sb29, Sb6777, Sb479, Sb8192, Sb1952, Sb30501, Sb21124, Sb26080, Sb33226, Sb35156, Sb38908, Sb55918. Dos transcritos *NoDiff*, apenas o Sb30501 apresenta um aumento de expressão na amostra 50% reidratada menor que duas vezes. Para os transcritos *Down regulated*, da figura, os resultados são coerentes entre as metodologias testadas para os transcritos Sb62763 e Sb60875 e incoerente para o Sb29.

CONCLUSÕES

A caracterização de genes relacionados a processos biológicos envolvidos no estresse hídrico em sorgo é de grande importância para o desenvolvimento de genótipos mais tolerantes a esse estresse. Isso poderia ser realizado tanto com a obtenção de transgênicos como com a transferência de genes via programa de melhoramento. No entanto, a complexidade do estresse de seca faz com que a identificação de genes específicos para transgenia ou transferência gênica seja difícil. Uma das alternativas é o estudo extensivo da expressão gênica em diferentes genótipos e em diferentes situações com extensiva análise de transcriptomas. Nesse trabalho, foram levantados genes expressos em dois diferentes genótipos e dois regimes hídricos. Alguns desse genes foram analisados por qPCR para verificar se os resultados de diferença de expressão são consistentes. O sequenciamento por RNAseq trouxe um avanço em relação as análises realizadas por microarranjos. Ambos métodos são robustos, mas em alguns casos o microarranjo ainda depende de hidridização cruzada trazendo pouco informação entre membros de família gênica. Por outro lado, o RNAseq por sua vez depende de extensa análise de bioinformática. Apesar disso, é um método mais robusto e informativo. Os resultados apresentados nesse trabalho mostraram consistência dos dados de RNAseq e qPCR principalmente para os

transcritos que foram mais expressos (*Up regulated*) sob a condição de estresses. Para aqueles que têm expressão igual houve incoerência de resultados e para *Down regulated*, os resultados foram medianamente coerentes. Novas análises de qPCR serão realizadas para confirmar os resultados, entretanto os resultados de RNAseq são importantes para o entendimento dos mecanismos celulares envolvidos na tolerância de sorgo ao estresse de seca.

AGRADECIMENTOS

Esse projeto teve a participação da FAPEMIG e Embrapa.

REFERÊNCIAS

- Bohnert, H.J., Nelson, D.E. and Jensen, R.G. (1995) Adaptations to environmental stresses. *Plant Cell* 7, 1099-1111.
- Bray EA (1997) Plant responses to water deficit. *Trends Plant Sci* 2:
- Chaves MM, Maroco JP, Pereira JS. 2003. Understanding plant responses to drought – from genes to the whole plant. *Functional Plant Biology* 30: 239–264.
- Deyholos, M. K. (2010), Making the most of drought and salinity transcriptomics. *Plant, Cell & Environment*, 33: 648–654.
- Livak, J. K., Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method, *Methods* 25 p402-408.
- Maere S¹, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005 Aug 15;21(16):3448-9. Epub 2005 Jun 21.
- Morgan, J.M. 1984. Osmoregulation and water stress in higher plants. *Annu. Rev. Plant Physiol.* 35:299–31.
- Reynolds M., Tuberosa R. (2008). Translational research impacting on crop productivity in drought-prone environments. *Curr. Opin. Plant Biol.* 11, 171–179 10.1016.
- Shannon P¹, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003 Nov;13(11):2498-504.



XXXI CONGRESSO NACIONAL DE MILHO E SORGO

"Milho e Sorgo: inovações,
mercados e segurança alimentar"
