



## Análise de enriquecimento funcional de redes gênicas em cana-de-açúcar

Luiza Oliveira Romão<sup>1</sup> e Renato Vicentini<sup>1</sup>

<sup>1</sup>Laboratório de Bioinformática e Biologia de Sistemas, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brasil

### INTRODUÇÃO

A cana-de-açúcar (*Saccharum* spp.) é caracterizada, principalmente, por sua capacidade de acumular altas concentrações de sacarose (em torno de 0,7 M) em seus colmos maduros (MOORE, 1995), tornando-a uma das mais importantes plantas cultiváveis do planeta, principalmente para os países das regiões tropical e subtropical.

A Biologia de Sistemas é o estudo de como diferentes elementos fundamentais como DNA, RNA, proteínas, e organismos interagem entre si, visando assim a integração e modelagem de extensos conjuntos de dados. O aumento do conhecimento em nível sistêmico da relação entre genes, proteínas e organismos tem permitido a criação de uma nova geração de modelos para a predição de comportamentos a partir de dados moleculares (YIN, et al; 2004). Desta forma, o principal desafio científico desta proposta será o desenvolvimento de modelos e ferramentas que permitam realizar a ligação entre planta e mudanças ambientais.

A análise de enriquecimento visa revelar o aumento ou diminuição de vias funcionais que possam estar associadas às respostas fisiológicas do organismo. Para se analisar as respostas funcionais do transcriptoma de um organismo, os dados de expressão gênica são analisados sobre a perspectiva de processos biológicos possibilitada pelo enriquecimento de termos GO (Gene Ontology, ASHBURNER, et al; 2000). Essa análise busca comparar determinados subconjuntos de transcritos em relação a todo o transcriptoma, verificando-se processos biológicos estatisticamente enriquecidos nos subconjuntos de transcritos testados.

No presente estudo, análises de enriquecimento funcional em redes regulatórias de cana-de-açúcar estão sendo conduzidas com intuito de integrar dados genéticos e moleculares relacionados ao papel da cana-de-açúcar em resposta a estresse hídrico, mudanças na relação fonte-dreno e no foto período, assim como durante a indução do desenvolvimento floral.

### OBJETIVOS

Avaliar e determinar o enriquecimento funcional de redes gênicas de cana-de-açúcar quando em situação de estresse abiótico e mudanças no desenvolvimento. Especificamente os objetivos são: 1) Analisar o enriquecimento funcional em genes de *Arabidopsis thaliana* utilizando os genes ortólogos aos genes de cana-de-açúcar; 2) Anotar funcionalmente os genes de cana-de-açúcar; 3) Montar um banco de dados de genes de cana-de-açúcar para análise de enriquecimento; 4) Analisar o enriquecimento funcional em cana-de-açúcar quando esta está submetida a diferentes situações de estresse abiótico e a mudanças no desenvolvimento vegetal; 5) Aplicar métricas estatísticas para avaliar os resultados.

## MÉTODOS

### *Determinação de redes gênicas de cana-de-açúcar*

A rede de regulação gênica de cana-de-açúcar em diferentes condições foi previamente determinada pelo grupo de pesquisa do Laboratório de Bioinformática e Biologia de Sistemas do CBMEG – UNICAMP utilizando o algoritmo de cinzelamento heurístico de agrupamentos (ou “HCCA”, de Heuristic Cluster Chiseling Algorithm), descrito por Mutwil *et al.* (2010).

### *Genes ortólogos em Arabidopsis thaliana*

A análise de filogenia para a prospecção dos reais ortólogos de genes conhecidos em outras espécies na cana-de-açúcar está sendo realizada através do software Phylxpress (VICENTINI, *et al.*; 2012). Nestas, é possível encontrar os transcritos de cana-de-açúcar e como eles se agrupam com os genes de *Arabidopsis thaliana*, fornecendo meios de encontrar genes ortólogos pelo agrupamento destas sequências.

### *Anotação funcional e construção do banco de dados de genes de cana-de-açúcar*

Para a anotação dos transcritos, está sendo utilizada a versão standalone do programa BLAST v 2.7.1+ (ALTSCHUL, *et al.*; 1990). Inicialmente, as sequências são alinhadas com proteínas do banco de dados Phytozome pelo programa BLASTx (capaz de alinhar nucleotídeos com proteínas) com os seguintes parâmetros: máximo de 20 hits por transcrito, “e-value”  $\leq 1e-05$ . Para entender a resposta da cana-de-açúcar nas diferentes condições avaliadas, os transcritos anotados foram classificados por ontologias gênicas utilizando o Blast2GO (GOTZ, *et al.*; 2008).

### *Análises de enriquecimento funcional e avaliação estatística*

Os dados anotados do transcriptoma da cana-de-açúcar são submetidos a análise de enriquecimentos de termos e categorias do Gene Ontology (ASHBURNER, *et al.*; 2000) utilizando programas BINGO (MAERE, *et al.*; 2005) e ClueGO (BINDEA, *et al.*; 2009), com teste hipergeométrico e correção de taxa de falsa descoberta de Benjamini e Hochberg ( $FDR < 0,05$ ), ambos plugins do software Cytoscape (SHANNON, *et al.*; 2003).

## RESULTADOS PRELIMINARES

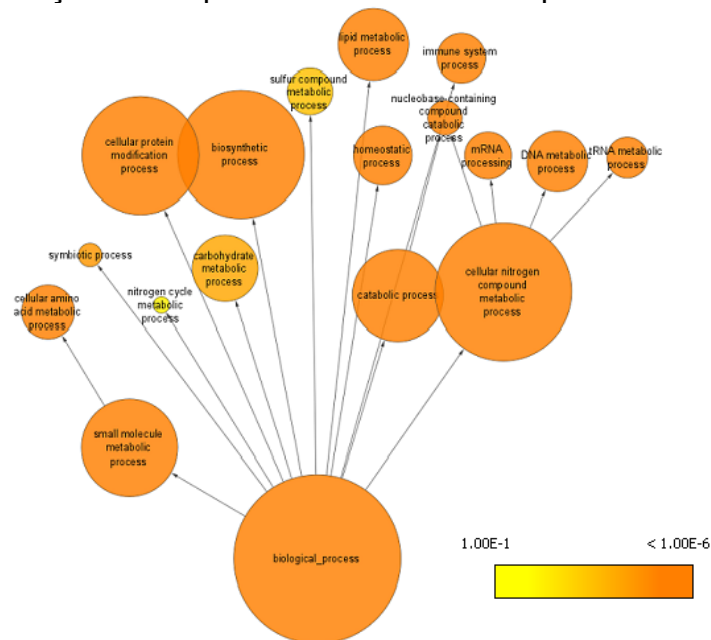
Foram submetidos a análise de enriquecimento funcional no programa BiNGO os três clusters contendo as maiores quantidades de genes, revelando quantas e quais categorias estavam enriquecidas (Tabela 1). Além disso, toda a rede contendo dados anotados do transcriptoma da cana-de-açúcar foi submetida a análise, sendo representada na Figura 1. Esta análise resultou com destaque para o enriquecimento da categoria associada a fotossíntese. Nas Figuras 1 e 2 o tamanho de cada círculo representa o número de genes que estão incluídos em cada termo da ontologia genética; o gradiente de cor indica o enriquecimento de acordo com o p-value.

Sub-rede	Nº de genes	Categorias enriquecidas
Cluster 6	220	859
Cluster 48	223	120
Cluster 64	219	91

**Tabela 1.** Principais sub-redes analisadas via BiNGO.



**Figura 1.** Organização hierárquica da análise de enriquecimento via BiNGO da rede total.



**Figura 2.** Organização hierárquica da análise de enriquecimento via BiNGO do Cluster 48.

Foram realizadas análises de enriquecimento também pelo ClueGO dos três principais clusters e de toda rede. Abaixo (Figuras 3, 4, 5 e 6) estão representadas as análises da sub-rede 6 e também da rede global. As análises foram feitas com  $kappa\ score = 0,4$  e com a especificidade dos termos da ontologia gênica reduzida conforme a quantidade de genes, afim de retornar redes que tenham menor quantidade de nós.

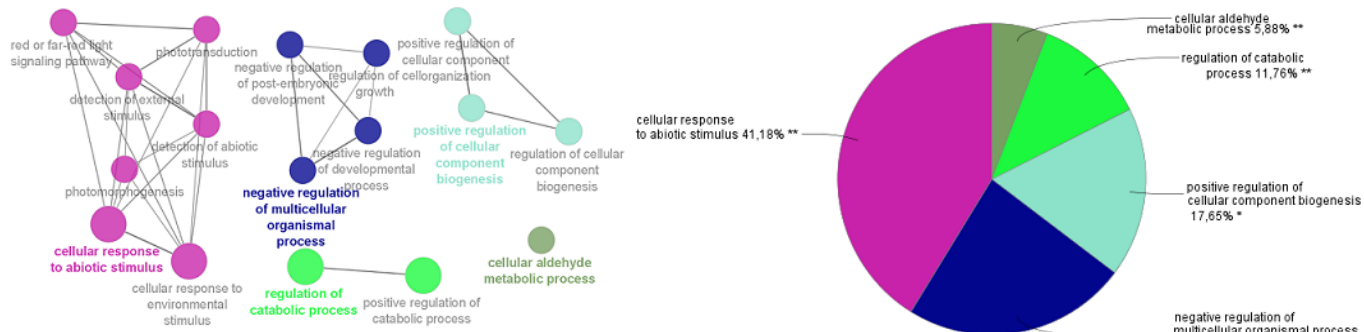


Figura 3. Análise de enriquecimento do Cluster 6 realizada pelo programa ClueGO.

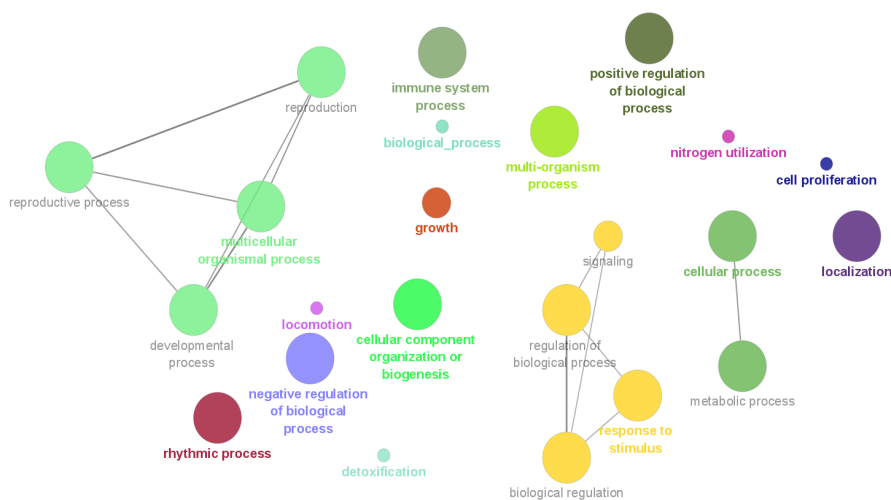


Figura 4. Análise de enriquecimento da rede total realizada pelo programa ClueGO.

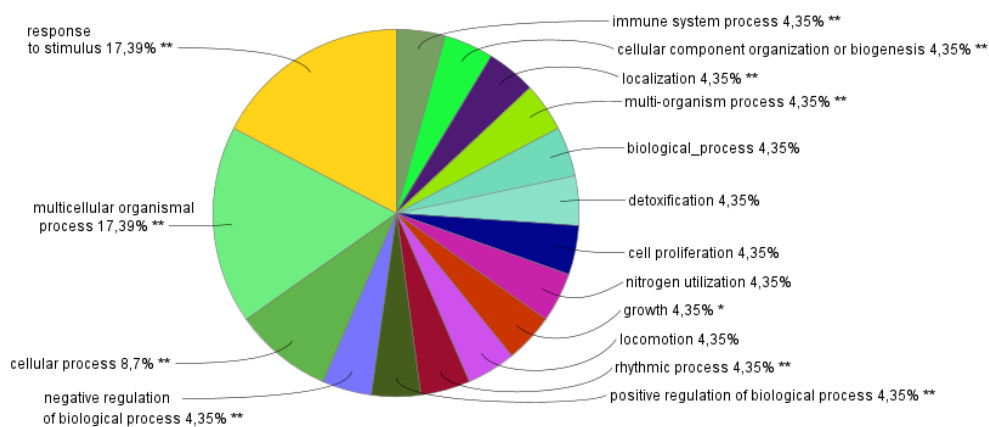


Figura 5. Análise de enriquecimento da rede total realizada pelo programa ClueGO

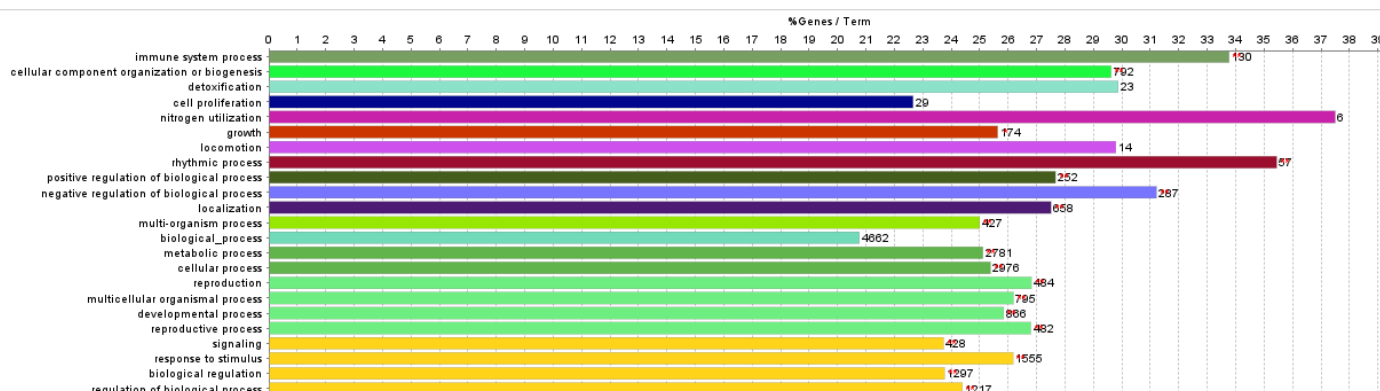


Figura 6. Análise de enriquecimento da rede total realizada pelo programa ClueGO.

Estes resultados mostram que a cana-de-açúcar está ativamente respondendo a estímulos ambientais (Figuras 3 e 5), mas principalmente foi possível identificar redes de co-expressão gênica envolvidas na resposta fotossintética frente às mudanças de fotoperíodo.

## REFERÊNCIA BIBLIOGRÁFICAS

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J Mol Biol.* 215(3):403-10.
- Anisimova M, Gascuel O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 55(4):539-52.
- Ashburner et al. Gene ontology: tool for the unification of biology. *Nat Genet.* May 2000;25(1):25-9.
- Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 25(8):1091-3.
- Götz S. et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite, *Nucleic Acids Research*, 36: 3420-3435.
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics* 21, 3448-3449.
- Moore PH (1995). Temporal and Spatial Regulation of Sucrose Accumulation in the Sugarcane Stem. *Functional Plant Biol* 22:661-679.
- Mutwil M., Usadel B., Schütte M., Loraine A, Ebenhöf O, Persson S. (2010). Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol.* 52 (1) 29-43.
- Nakamura, Yamada, Tomii, Katoh (2018) Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34:2490–2492.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*; 13(11):2498-504
- Souza AP, Gaspar M, Silva EA, Ulian EC, Waclawovsky AJ, Nishiyama MY et al. (2008) Elevated CO<sub>2</sub> increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane. *Plant Cell Environ* 31: 1116– 1127.
- Tyson, J. J. (2007). Bringing cartoons to life. *Nature*, 445(7130), 823.
- Vicentini R, Del Bem LEV, Van Sluys MA, Nogueira FTS, Vincenz M (2012) Gene Content Analysis of Sugarcane Public ESTs Reveals Thousands of Missing Coding-Genes and an Unexpected Pool of Grasses Conserved ncRNAs. *Tropical Plant Biology.* 5:2, 199-205.
- Yin X, Struik PC, Kropff MJ (2004). Role of crop physiology in predicting gene-to-phenotype relationships. *Trends Plant Sci* 9:426-432.