



# ANÁLISE DE ENRIQUECIMENTO FUNCIONAL DE REDES GÊNICAS EM CANA-DE-AÇÚCAR

**Palavras-Chave:** Cana-de-açúcar, Enriquecimento, Redes gênicas

**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 é um recurso essencial relacionado ao desenvolvimento econômico, tendo o Brasil como um dos seus maiores produtores. Deste insumo, desenvolvem-se importantes processos como a produção de açúcar e etanol, fabricação de biocombustíveis e geração de energia sustentável. Em estudos prévios, foi possível a identificação e caracterização de genes da cana-de-açúcar possivelmente envolvidos em aumento de biomassa e adaptações a mudanças ambientais. A análise de enriquecimento funcional de redes gênicas em cana-de-açúcar permite a aquisição de respostas biológicas relevantes que forneçam conhecimentos sobre a planta e que possam ser utilizados nas áreas de transcriptoma, biologia de sistemas, proteômica e mapeamento metabólico. No presente estudo, análises de enriquecimento funcional em redes regulatórias de cana-de-açúcar foram 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 fotoperíodo, assim como durante a indução do desenvolvimento floral.

## METODOLOGIA

### *Enriquecimento funcional em genes de Arabidopsis thaliana*

Inicialmente, foi realizado o treinamento para manuseio das ferramentas de análise de enriquecimento funcional de genes. Foram utilizados os dados previamente anotados pelo grupo de pesquisa do Laboratório de Bioinformática e Biologia de Sistemas do IB, em especial a relação de ortologia entre os genes de cana-de-açúcar e *Arabidopsis thaliana*. Estes dados foram submetidos à análise de enriquecimento de termos e categorias do Gene Ontology (GO) (ASHBURNER, et al; 2000).

Utilizando os 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), foi realizada a análise de enriquecimento dos termos GO para determinar quais categorias funcionais foram super-

representadas nos transcritos que apresentaram expressão nas redes identificadas.

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

Para anotação dos genes e posterior montagem do banco de dados provenientes da cana-de-açúcar, foi realizado o alinhamento das sequências por meio do programa BLAST+ (ALTSCHUL, et al; 1990). Para tal, foram utilizados os parâmetros -p Blastx, -b 20, -v 20, -e 0.00001, -F F. As sequências dos transcritos da cana-de-açúcar foram alinhadas contra todo o banco de dados NR do NCBI (<https://www.ncbi.nlm.nih.gov/>).

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

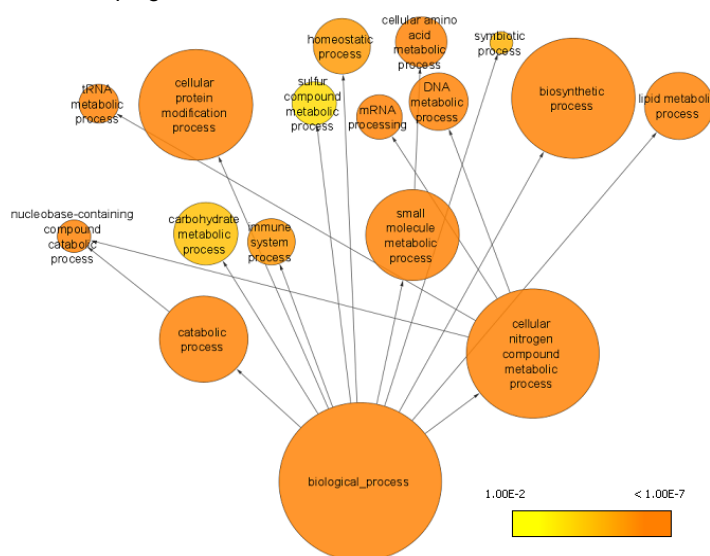
Os dados anotados do transcriptoma da cana-de-açúcar foram submetidos a análise de enriquecimentos de termos e categorias GO pelo plugin BINGO. Pelo plugin MCODE foram selecionados os clusters mais densamente conectados para apresentação dos resultados da análise. Para tal, foram utilizados os parâmetros Degree cut-off= 2, node density cut-off= 0.1, node score cut-off= 0.2, K-score= 2 e máx. depth= 100.

## RESULTADOS

Através do plugin BINGO foi realizada a análise dos três maiores grupos de genes da rede (Tabela 1), revelando quantas e quais categorias estavam mais enriquecidas, assim como a análise de toda a rede (Figura 1), com destaque para o enriquecimento das categorias associadas a processos biossintéticos, modificação de proteínas e processos metabólicos formadores de nitrogênio celular. Na Figura 1 o tamanho de cada nó representa a quantidade de genes que estão incluídos em cada termo da ontologia gênica; o gradiente de cor indica o enriquecimento de acordo com o p-value. Estas análises foram realizadas nos dados da rede de co-expressão do experimento com fotoperíodo de 8h/16h (luz/escuro) de cana-de-açúcar, em comparação ao fotoperíodo normal de 12h/12h (luz/escuro). A rede de co-expressão deste experimento possui 11.335

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

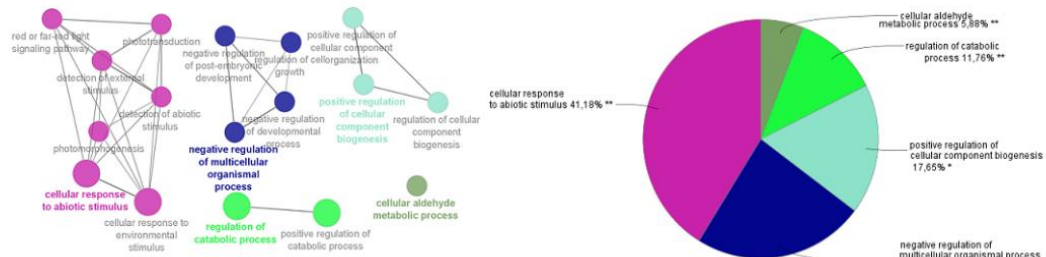
**Tabela 1.** Números das principais sub-redes analisadas através do plugin BINGO.



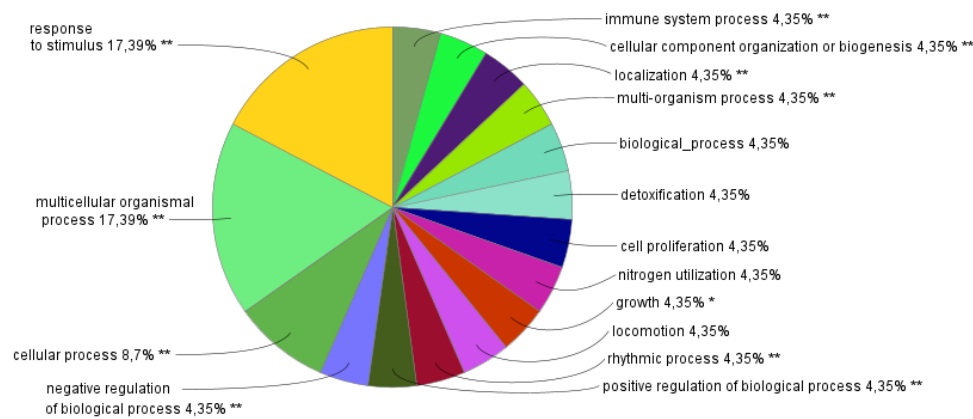
**Figura 1.** Organização hierárquica da análise de enriquecimento via BINGO da rede total.

transcritos organizados em 130 clusters.

Foram realizadas análises de enriquecimento também pelo plugin ClueGO dos três principais clusters e de toda rede. As figuras abaixo apresentam os resultados das análises da sub-rede 6 (Figura 2) e também da rede global (Figura 3). As análises foram feitas com kappa score = 0,4 e com a especificidade dos termos da ontologia gênica reduzida conforme a quantidade de genes, para retornar redes que tenham menor quantidade de nós.



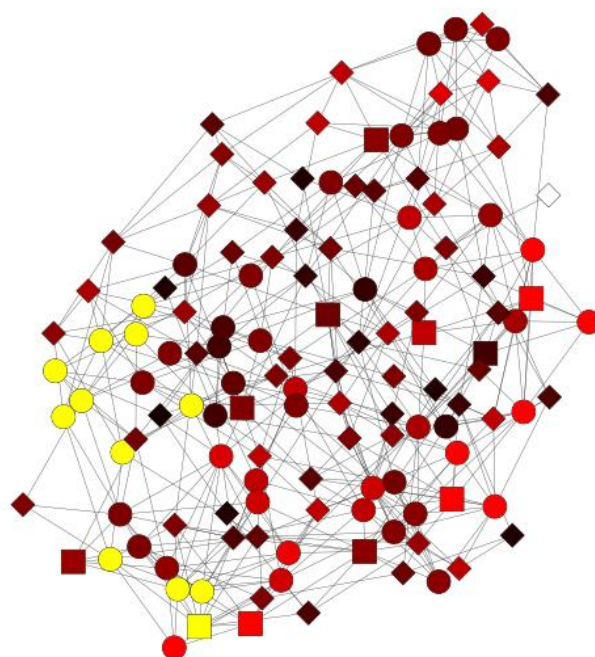
**Figura 2.** Rede de interações e análise gráfica de enriquecimento do Cluster 6 realizado pelo programa ClueGO.



**Figura 3.** Análise gráfica de enriquecimento da rede total realizada pelo programa ClueGO

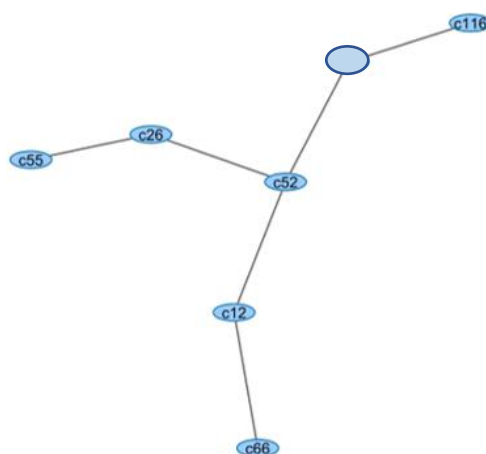
Após a realização da anotação funcional realizada por meio do BLAST dos transcritos da cana-de-açúcar contra todo o banco de dados NR do NCBI, foi obtida anotação para mais de 76.000 transcritos. Estes transcritos foram submetidos à análise de enriquecimento funcional pelo plugin BINGO.

Utilizando o algoritmo do MCODE (Degree cut-off= 2, node density cut-off= 0.1, node score cut-off= 0.2, K-score= 2 e máx. depth= 100) foram analisados os 130 clusters da rede gênica da cana-de-açúcar, sendo encontrados 12 clusters (Figura 4) com grande conectividade. Dentro deste conjunto, temos 6 clusters com enriquecimento funcional relevante (Corrected p-value < 0.05) (Tabela 2 e Figura 5).



**Figura 4.** Rede de alta densidade gerada pelo MCODE, com os 12 clusters densamente conectados destacados em amarelo.

Cluster	Termos GO
116	Protein localization to peroxisome; Establishment of protein localization to peroxisome; Peroxisomal transport;
12	Regulation of secondary metabolic process; Regulation of phenylpropanoid metabolic process
66	RNA processing; Gene expression; RNA metabolic process
26	RNA modification; ncRNA catabolic process; miRNA catabolic process;
55	RNA modification; RNA processing; snRNA metabolic process; RNA metabolic processing;
52	ncRNA procesing; ncRNA metabolic process; RNA processing; rRNA procesing;



- ✓ Regulation of phenylpropanoid and secondary metabolic
- ✓ Peroxisomal transport
- ✓ RNA modification
- ✓ ncRNA and miRNA, snRNA catabolic and metabolic process

**Figura 5.** Rede de clusters que apresentaram enriquecimento funcional com correct p-value < 0.05

**Tabela 2.** Termos GO enriquecidos nos 12 clusters da sub-rede gerada no MCODE.

Neste conjunto de dados é possível observar a grande presença de termos GO enriquecidos associados à diferentes classes de RNA. Em estudos prévios, é definido que a expressão de smallRNAs, como o miRNA, tem papéis importantes na resposta apresentada pelas plantas durante o estresse abiótico e biótico (THIEBAUT, 2017). O miRNA é um importante regulador de expressão gênica em plantas, que é expresso normalmente durante o desenvolvimento ou em resposta à desafios ambientais, tendo como alvos, por exemplo, os genes que estão envolvidos no transporte de nutrientes e produção metabólica (CARNAVALE, 2013). Outros estudos sugerem que long ncRNAs em plantas

podem exercer diversas funções diferentes na resposta ao estresse, como realizar a modificação de histonas, metilar DNA, ser o precursor de sRNAs e/ou bloquear a interação entre miRNAs e seus alvos (WANG, 2017).

## **CONCLUSÕES**

Sendo a cana-de-açúcar um recurso importante para o desenvolvimento econômico nacional, é de suma importância a análise do seu comportamento em resposta às situações de estresse. As análises iniciais mostraram que a cana-de-açúcar está ativamente respondendo a estímulos ambientais, mas principalmente foi possível identificar redes de co-expressão gênica envolvidas na resposta fotossintética frente às mudanças de fotoperíodo. Ademais, os resultados obtidos posteriormente sugerem que a cana-de-açúcar apresenta mecanismos de resposta ao estresse relacionados com diversas classes de RNA, tais como miRNA, siRNA e ncRNA.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

ALTSCHUL, Stephen F. et al. Basic local alignment search tool. *Journal of molecular biology*, v. 215, n. 3, p. 403-410, 1990. ASHBURNER, Michael et al. Gene ontology: tool for the unification of biology. *Nature genetics*, v. 25, n. 1, p. 25-29, 2000.

BINDEA, Gabriela et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, v. 25, n. 8, p. 1091- 1093, 2009.

CARNAVALE BOTTINO, Mariana et al. High-throughput sequencing of small RNA transcriptome reveals salt stress regulated microRNAs in sugarcane. *PLoS one*, v. 8, n. 3, p. e59423, 2013.

CONESA, Ana et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, v. 21, n. 18, p. 3674-3676, 2005.

MAERE, Steven; HEYMANS, Karel; KUIPER, Martin. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, v. 21, n. 16, p. 3448-3449, 2005.

SHANNON, Paul et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, v. 13, n. 11, p. 2498-2504, 2003.

THIEBAUT, Flávia et al. Roles of non-coding RNA in sugarcane-microbe interaction. *Non-coding RNA*, v. 3, n. 4, p. 25, 2017.

WANG, Jingjing et al. Non-coding RNAs and their roles in stress response in plants. *Genomics, proteomics & bioinformatics*, v. 15, n. 5, p. 301-312, 2017.