



## Investigação dos mecanismos moleculares associados à resistência às cigarrinhas-das-pastagens em *Urochloa Decumbens* utilizando redes de co-expressão

Palavras-Chave: REDES COMPLEXAS, FORRAGEIRAS TROPICAIS, ESTRESSE BIÓTICO

Autores(as):

Jéssica Faversoni Diniz (autora), IB – UNICAMP

Alexandre Hild Aono (coautor), IB – UNICAMP

Guilherme Francio Niederauer (coautor), IB – UNICAMP

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Anete Pereira de Souza (orientadora), IB - UNICAMP

---

### INTRODUÇÃO:

No Brasil a agropecuária se configura como uma das atividades mais relevantes para manutenção do produto interno bruto (PIB) nacional (CEPEA, 2022), de modo que a busca por formas de melhorias no setor se mostra de extrema importância. Haja vista que a pecuária brasileira é baseada em sistemas de produção extensivos e dependentes da utilização de gramíneas, o melhoramento destas forrageiras tem grande potencial de aumentar a produtividade da pecuária nacional (SANTOS, 2022).

A maioria das áreas de pastagens do país são compostas por espécies do gênero *Urochloa*, sendo *U. decumbens*, *U. humidicola*, *U. brizantha* e *U. ruziziensis* as mais empregadas (DOVALLE, 2009; SANTOS, 2022). Dentre estas, *U. decumbens* é uma variedade que recebe grande atenção dos programas de melhoramento em virtude de suas características agrônômicas, tais como: boa adaptação a solos ácidos e pouco férteis, resistência às secas sazonais, capacidade de suportar a pressão exercida pelo pastejo, eficiência no uso de nitrogênio e produção de uma forragem palatável e nutritiva ao gado (LOW, 2015; MATEUS *et al.*, 2015; SANTOS, 2022). Entretanto, *U. decumbens* se mostra suscetível a diferentes espécies de cigarrinha-das-pastagens, como as dos gêneros *Notozulia*, *Mahanarva* e *Deois*. (MATIAS *et al.*, 2021).

A relação específica cigarrinha-*Urochloa* é pouco descrita na literatura, sendo as vias bioquímicas relacionadas ao processo de resistência observadas apenas para algumas espécies do gênero (BARROS, 2021). Nesse sentido, considerando as lacunas existentes quanto aos fatores genéticos que determinam a resistência de *Urochloa* às cigarrinhas, bem como os processos biológicos envolvidos, uma abordagem baseada em redes de co-expressão se mostra promissora. Com o uso de grandes conjuntos de dados, é possível obter uma visão mais ampla de como os processos biológicos observados são regulados por grupos de genes funcionalmente relacionados (SERIN *et al.*, 2016). Assim, essa abordagem se torna ainda mais relevante quando se trata da compreensão dos mecanismos de resistência, visto que estes, em geral, envolvem o *cross-talk* de diferentes produtos gênicos relacionados a vias metabólicas intrincadas (SHIGENAGA, 2017).

Considerando o exposto acima, o presente projeto propõe a utilização de redes de co-expressão gênica de *Urochloa* para investigar o processo de resistência às cigarrinhas. Com a realização de análises de redes complexas em grupos gênicos associados à resistência, espera-se propiciar uma compreensão mais ampla de quais fatores genéticos e moleculares estão envolvidos na resistência.

#### **METODOLOGIA:**

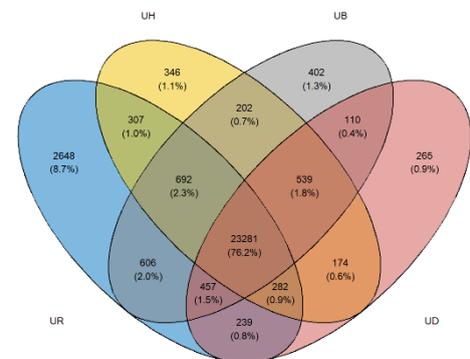
Como material vegetal utilizado, foram reunidos dois experimentos de sequenciamento de RNA do gênero *Urochloa*, contemplando: (i) dados de folha de 104 acessos, sendo 11 de *U. ruziziensis*, 25 de *U. decumbens*, 27 de *U. brizantha* e 29 de *U. humidicola* (HANLEY *et al.*, 2021) e (ii) dados de raízes de 90 mudas de *U. decumbens* com exposição ao alumínio (SALGADO *et al.*, 2017). A qualidade das leituras de sequenciamento foi avaliada através do FastQC e as sequências de baixa qualidade foram filtradas com o Trimmomatic v0.39 (BOLGER *et al.*, 2014). Com as leituras de sequenciamento filtradas, foi realizado, através do software STAR v2.7.3a (DOBIN *et al.*, 2013), o mapeamento das reads contra o genoma de referência de *U. ruziziensis*. A montagem do transcriptoma, por sua vez, foi executada com o StringTie v2.1.6 (PERTEA, *et al.*, 2015) e a qualidade dos transcritos montados foi avaliada a partir da comparação com sequências ortólogas conservadas, através da ferramenta BUSCO v5.4.7 (SIMÃO *et al.*, 2015), e quanto aos índices de comprimento dos scaffolds, com a ferramenta BMap v39.01 (BUSHNELL, 2014). Para inferência da quantificação de expressão gênica foi empregado o software Salmon obtendo-se uma matriz de contagem gênica que contempla o gênero *Urochloa*.

A partir do processo de análise de dados de RNA-Seq, foi construída uma rede de co-expressão gênica global. Para tal, os valores de contagem gênica foram normalizados e utilizados para cálculo de uma matriz de correlação de Pearson e o pacote WGCNA (LANGFELDER, 2008) foi empregado para construção de uma rede livre de escala e a matriz de dissimilaridade associada. Com esta matriz, através de um método de agrupamento hierárquico, foram definidos os grupos gênicos co-expressos. A partir desses grupos foram selecionados aqueles potencialmente associados com processos biológicos envolvidos na resistência. Sendo essa seleção realizada com base em regiões de QTL previamente descritas (FERREIRA, 2019), marcadores moleculares associados à resistência identificados por meio de metodologias de aprendizado de máquina (trabalho em desenvolvimento pelo grupo de pesquisa seguindo o método proposto por AONO *et al.*, 2020, 2022) e o que se tem na literatura acerca do processo de defesa a estresses bióticos. Considerando o conjunto de genes selecionado, foi realizada a análise com o pacote topGO (ALEXA, 2022) para definição de categorias de ontologia gênica enriquecidas nos módulos. Ainda, o banco de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (KANEHISA, 2000) serviu como referência para definição de vias metabólicas enriquecidas associadas à resistência. Adicionalmente, será construída uma rede de co-expressão específica com os genes identificados como presentes em módulos de interesse. Essa rede será construída considerando parâmetros mais restritivos e suas avaliações e análises serão mais rigorosas de modo que as inferências funcionais finais sejam mais finas.

#### **RESULTADOS E DISCUSSÃO:**

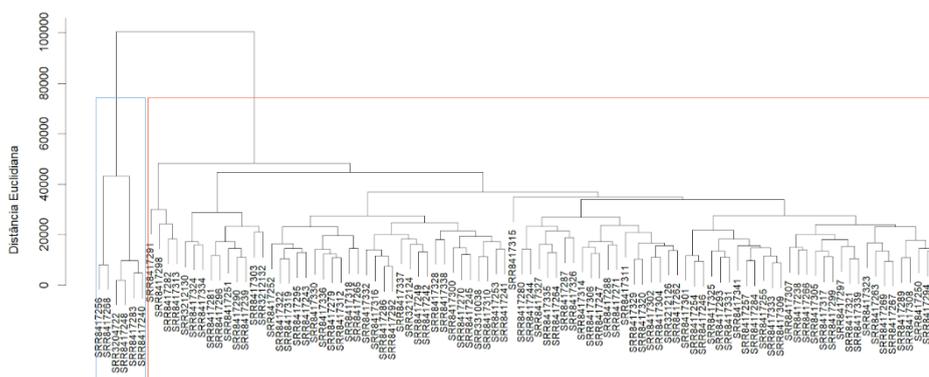
- **Análise de Dados de RNA-Seq**

Inicialmente, o conjunto de dados utilizado contemplava 36.719.525 reads, porém após a filtragem executada através do Trimmomatic foram mantidas 33.817.671 reads. A etapa de montagem do transcriptoma realizada através do StringTie, por sua vez, resultou em 99.633 genes montados representativos do transcriptoma total avaliado. Quanto a quantificação da expressão gênica estimada com o software Salmon, foi realizada para cada transcrito nas diferentes amostras, gerando-se uma matriz de contagem gênica. Essa matriz, por sua vez, passou por algumas etapas de filtragem a fim de que genes com níveis nulos de expressão fossem removidos. Para realização das etapas de filtragem a matriz gênica foi dividida de modo que se obtivesse uma matriz específica para cada uma das quatro espécies (*U. decumbens*, *U. humidicola*, *U. ruzizensis*, *U. brizantha*) contidas no conjunto de dados reunido. Essa separação foi realizada para que, após o filtro, ainda se mantivessem amostras representativas das espécies em questão. Assim, para cada uma das matrizes individuais foram removidos os genes com valor de transcritos por milhão (TPM) menor que 0,5 em até 25% das amostras. Após a filtragem da matriz de cada uma das espécies, foi analisado o conjunto de genes que foram mantidos simultaneamente nos quatro subconjuntos de dados (**Figura 1**). Os 23.281 genes contidos na intersecção dos filtros aplicados a cada um dos subconjuntos foram mantidos para a execução das etapas posteriores do projeto.



**Figura 1. Diagrama de Venn:** representa a intersecção dos genes mantidos para cada um dos subconjuntos após as etapas de filtro aplicadas. UH: *U. humidicola*; UB: *U. brizantha*; UD: *U. decumbens*; UR: *U. ruzizensis*

A fim de se avaliar o nível de agrupamento entre as amostras utilizadas, foi realizado um dendrograma com o conjunto de dados. Ao analisar o dendrograma das amostras foram identificados dois grupos distintos, um menor contendo apenas 6 amostras e outro contendo todas as demais. Essas seis amostras correspondem a diferentes espécies e são oriundas de ambos os experimentos de RNA-Seq utilizados, portanto é possível que correspondam a outliers na coleta de dados. Ainda, foi realizada uma análise de componentes principais (PCA) buscando-se por padrões gerados pelas duas primeiras PCs que poderiam auxiliar na identificação de amostras anômalas. O gráfico de dispersão de PCA do conjunto de dados completo evidencia as mesmas 6 amostras observadas no dendrograma com um padrão distinto de agrupamento. Porém quando a análise é realizada com a extração dessas amostras é possível observar uma dispersão mais homogênea dos dados. Posto isso, optou-se pela retirada dessas amostras visto que seus dados podem representar outliers que distorceriam as análises posteriores. A remoção dessas amostras problemáticas pode ajudar a melhorar a qualidade geral e coerência do conjunto de dados.



**Figura 2. Dendrograma para agrupamento hierárquico das amostras:** no eixo x são representadas as amostras e no eixo y a medida de distância entre as amostras. Evidenciado em vermelho se encontra o cluster que contém as amostras mantidas no dataset e em azul as seis amostras retiradas

Para inferência funcional dos genes diferencialmente expressos, suas sequências foram comparadas com os bancos de dados contidos na plataforma Trinotate (BRYANT *et al.*, 2017). As sequências presentes nos respectivos bancos de dados estão acompanhadas de suas respectivas anotações funcionais e termos GO associados. Assim, ao realizar o alinhamento dos genes de interesse contra essas sequências, foram identificadas as correspondências entre eles. Com base nessas inferências funcionais, os genes estão sendo analisados em relação ao seu potencial papel na expressão das respostas de defesa da planta frente ao ataque de cigarrinhas.

- **Modelagem da Rede de Co-expressão**

Com as quantificações de leituras de sequenciamento estimadas para os genes montados, está sendo construída uma rede de co-expressão gênica global. O pacote WGCNA do R foi empregado para definir uma estrutura de rede livre de escala a partir da matriz de correlação, considerando a estimativa de uma potência  $\beta$  para ajustar a distribuição de conectividade. Foram executados alguns testes a fim de se observar qual valor  $\beta$  está associado ao melhor ajuste da distribuição e a partir disso foi definida a potência de valor 8 visto que com esse valor o  $R^2$  obtido é de 0,8. Isto é, resulta em uma matriz associada a uma maior similaridade com a estrutura de uma rede livre de escala e se mantém um bom nível de conectividade média entre os componentes da rede (BARABÁSI, 2000). Ainda, foi realizada a criação de uma matriz de sobreposição topológica e o cálculo de uma matriz de dissimilaridade. Com esta matriz final, utilizando-se um método de agrupamento hierárquico, foram definidos 110 grupos gênicos co-expressos (**Tabela 1**). Esses grupos correspondem a conjuntos de genes que apresentam perfis de expressão semelhantes em diferentes amostras.

<b>Total de genes contidos nos grupos</b>	23.281
<b>Máximo de genes contidos em um único grupo</b>	6.482
<b>Mínimo de genes contidos em um único grupo</b>	50
<b>Mediana do conjunto de dados</b>	126,5

**Tabela 1.** Caracterização dos grupos gênicos co-expressos

- **Identificação de Módulos Associados à Resistência**

Para cada grupo gênico definido através do WGCNA foi realizada uma análise de enriquecimento de termos e categorias do Gene Ontology (GO), de modo a caracterizar quais processos biológicos são específicos de cada grupo (YOUNG *et al.*, 2010). Grupos associados com termos GO referentes a processos de transporte de íons, atividades quinases, homeostase celular, resposta de defesa, regulação de resposta a estímulos, são alguns exemplos de funções biológicas de interesse observadas. Ainda, foi realizada uma análise de enriquecimento para as vias metabólicas inferidas através do banco de dados KEGG. A análise dessas funções enriquecidas em conjunto com o que se encontrou na literatura e as inferências de genes associados aos QTLs descritos, tem sido empregada para a escolha de grupos relevantes no que tange ao processo de resistência.

## **CONCLUSÕES:**

As características genéticas de *Urochloa* spp. são escassas, portanto, a identificação de regiões do genoma do gênero associadas à resistência às cigarrinhas representa uma grande contribuição para os

programas de melhoramento que visam a fixação dessa característica nas diferentes cultivares. Nesse sentido, a partir de dados de experimentos de RNA-Seq, o projeto engloba um conjunto de avaliações a fim de se compreender as funções biológicas e mecanismos de interação molecular associados à expressão da resistência em plantas do gênero *Urochloa* frente ao ataque das cigarrinhas-das-pastagens. Espera-se que a modelagem de redes com um grande conjunto de dados possibilite o estabelecimento de uma visão sistêmica dos fatores envolvidos na resistência, além de direcionar os programas de melhoramento para diminuir o impacto produtivo causado pela praga.

---

## BIBLIOGRAFIA

- ALEXA, A., RAHNENFUHRER J.. *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.50.0, 2022.
- AONO, Alexandre H., et al. "A joint learning approach for genomic prediction in polyploid grasses." *bioRxiv*, 2022.
- AONO, Alexandre H., et al. "Machine learning approaches reveal genomic regions associated with sugarcane brown rust resistance." *Scientific reports* 10.1, p. 1-16, 2016.
- AUAD, A. M., and S. E. B. da Silva. "30.2 Pasture Pest Insects." *Natural Enemies of Insect Pests in Neotropical Agroecosystems*, p: 370, 2019.
- BARABÁSI, A.L. et al. "Scale-free characteristics of random networks: the topology of the world-wide web." *Physica A: statistical mechanics and its applications*, v. 281, n. 1-4, p. 69-77, 2000.
- BARROS, Rafael De A., et al. "Differential defense responses of tropical grasses to *Mahanarva spectabilis* (Hemiptera: Cercopidae) infestation." *Anais da Academia Brasileira de Ciências* 93, 2021.
- BOLGER, Anthony M., LOHSE, Marc, and USADEL, Bjoern. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30.15, p. 2114-2120, 2014.
- BRYANT, Donald M., et al. "A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors." *Cell reports* 18.3, p. 762-776, 2017.
- BUSHNELL, B. "BBMap: A Fast, Accurate, Splice-Aware Aligner." Disponível em: <https://www.osti.gov/servlets/purl/1241166>. Acesso em: 2023.
- CEPEA - Centro de Estudos Avançados em Economia Aplicada. PIB agronegócio. (2022) Disponível em: <http://cepea.esalq.usp.br/pib/>. Acesso em: 18 out. 2022.
- DOBIN, A. et al. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics*, v. 29, n. 1, p. 15-21, 2013.
- DOVALLE, Cacilda B., JANK, Liana, e RESENDE, Rosângela M. S.. "O melhoramento de forrageiras tropicais no Brasil." *Revista Ceres* 56.4, p. 460-472, 2009.
- FERREIRA, Rebecca C. U., et al. "Genetic mapping with allele dosage information in tetraploid *Urochloa decumbens* (Stapf) RD Webster reveals insights into spittlebug (*Notozulia entreriana* Berg) resistance." *Frontiers in Plant Science* 10, p. 92, 2019.
- HANLEY, Steve J., et al. "Allele mining in diverse accessions of tropical grasses to improve forage quality and reduce environmental impact." *Annals of Botany* 128.5, p. 627-637, 2021.
- KANEHISA, M., and GOTO, S.. "KEGG: kyoto encyclopedia of genes and genomes." *Nucleic acids research* 28.1 (2000): 27-30.
- LANGFELDER, P.; HORVATH, S. "WGCNA: an R package for weighted correlation network analysis." *BMC Bioinformatics*, v. 9, n. 1, p. 1-13, 2008.
- LOW, Susan G. "Signal grass (*Brachiaria decumbens*) toxicity in grazing ruminants." *Agriculture* 5.4, p. 971-990, 2015.
- MATEUS, Rogério G., et al. "Genetic parameters and selection of *Brachiaria decumbens* hybrids for agronomic traits and resistance to spittlebugs." *Crop Breeding and Applied Biotechnology* 15, p. 227-234, 2015.
- MATIAS, Filipe Inácio, et al. "*Urochloa* spp. multivariate performance: Similarities and divergences among intra-and interspecific populations." *Crop Science* 61.2 (2021): 1104-1116.
- PERTEA, M. et al. "StringTie enables improved reconstruction of a transcriptome from RNA-seq reads." *Nature biotechnology*, v. 33, n. 3, p. 290-295, 2015.
- SALGADO, Leonardo R., et al. "De novo RNA sequencing and analysis of the transcriptome of signalgrass (*Urochloa decumbens*) roots exposed to aluminum." *Plant Growth Regulation* 83.1 (2017): 157-170.
- SANTOS, PM, et al. "Demandas para pesquisas e desenvolvimento para as pastagens no Brasil.", 2022.
- SERIN, Elise AR, et al. "Learning from co-expression networks: possibilities and challenges." *Frontiers in plant science* 7 (2016): 444.
- SHIGENAGA, Alexandra M., et al. "Towards engineering of hormonal crosstalk in plant immunity." *Current opinion in plant biology* 38, p. 164-172, 2017.
- SIMÃO, F. A., et al. "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." *Bioinformatics*, v. 31, n. 19, p. 3210-3212, 2015.
- YOUNG, M. D. et al. "Gene ontology analysis for RNA-seq: accounting for selection bias." *Genome Biology*, v. 11, n. 2, p. R14, 2010.