



ANÁLISE INTEGRATIVA E COMPARATIVA DE DADOS TRANSCRIPTÔMICOS DE SOJA (*Glycine max*)

*Fabricio Brum-Machado, Kanhu Charan Moharana,
Rajesh Kumar Gazara, Thiago Motta Venancio*

A soja (*Glycine max*) é uma das mais importantes culturas na agricultura mundial, constituindo uma fonte crucial de proteína e óleo na alimentação humana e animal. Desde sua introdução, a cultura da soja apresentou uma evolução significativa e tem sido cada vez mais intensificada pelas tecnologias geradas por Universidades e Institutos de pesquisas, com o desenvolvimento de cultivares melhorados. Contudo, ainda são observadas e esperadas perdas relevantes em decorrência de condições ambientais adversas e estresses que comprometem o crescimento, desenvolvimento e produtividade dessa cultura. Com a publicação do genoma da soja em 2010, surgiram diversas frentes em estudo molecular e genômico desta espécie, incluindo análises transcriptômicas em diversas condições e tecidos. No entanto, análises integrativas destes dados visando identificar padrões transcricionais globais subjacentes às transições morfológicas e fisiológicas ainda são escassos. O objetivo deste trabalho é integrar dados transcriptômicos acumulados ao longo dos últimos 7 anos, buscando semelhanças, diferenças e especificidades entre esses perfis transcricionais. Os dados foram baixados do banco de dados do SRA (Short Read Archive do NCBI), analisados e filtrados por qualidade com os programas FASTQC e FASTX, mapeados contra o genoma de referência *Wm82.a2.v1* com o software STAR. A normalização dos valores de FPKM e RPKM foi realizada com o software StringTie. Foram encontradas e selecionadas 1.518 amostras divididas em 108 experimentos e 2.107 arquivos FASTQ que, mesmo compactados totalizam aproximadamente 5.5 Tb. Ao todo os dados compreendem 124 cultivares, tendo sido a Williams 82 usada em 40% dos estudos. Dentre os tecidos, encontramos 7 principais grupos: semente, folha, flor, raiz, caule, vagem e calo. A plataforma Illumina HiSeq 2000 foi utilizada em mais de 60% dos experimentos. Realizamos o agrupamento das amostras usando a análise de componentes principais (*Principal component analysis*, PCA). Esperamos que a coleção de dados analisados de forma sistemática seja útil para não somente revelar aspectos importantes da fisiologia de soja, como também forneça uma rica fonte de dados para a comunidade científica.

Palavras-chave: Transcriptoma, Soja, *Glycine max*, Expressão diferencial, Biotecnologia, RNA-Seq.

Instituição de fomento: CAPES, FAPERJ.