



Análise Computacional Integrativa de Dados Transcriptômicos de Soja (*Glycine max*)

Fabricio Brum Machado, Kanhu C. Moharana, Fabricio Almeida-Silva, Rajesh K. Gazara, Francisnei Pedrosa-Silva, Fernanda S. Coelho, Clícia Grativol, Thiago M. 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, com o desenvolvimento de cultivares melhorados, estimulados pelas tecnologias geradas por Universidades e Institutos de pesquisas. Contudo, ainda são observadas e esperadas perdas relevantes em sua produção, em decorrência de condições ambientais adversas, que comprometem o crescimento, desenvolvimento e produtividade dessa cultura. Com a publicação do genoma da soja em 2010, surgiram diversas frentes de estudo molecular e genômico desta espécie, incluindo análises transcriptômicas em diversas condições e tecidos. No entanto, análises integrativas desses dados, visando identificar padrões transcricionais globais em tecidos específicos, ainda são escassos. Dessa forma, o objetivo deste trabalho foi integrar os dados transcriptômicos, disponibilizados na base de dados, relacionados aos diferentes tecidos da soja, buscando semelhanças, diferenças e especificidades. Neste estudo, foram coletados dados de 1.298 amostras de transcriptoma de soja disponíveis no banco de dados público do SRA (Short Read Archive do NCBI), que foram filtrados, processados e mapeados sistematicamente contra o genoma de referência de soja (Wm82.a2.v). Dos genes anotados, 94% (52.737 / 56.044) destes, mostraram expressão detectável em pelo menos uma das amostras. Foram revelados três grupos principais através do agrupamento não supervisionado, incluindo amostras de partes subterrâneas, aéreas e relacionadas a sementes. Foram encontrados 452 genes com níveis de expressão uniformes e constantes, o que apoia seu papel como genes de manutenção (*housekeeping genes*). Além destes, 1.349 genes mostraram padrões de expressão tecido específico. Uma análise em nível de transcrição revelou que 95% (70.963 / 74.490) dos transcritos conhecidos se sobrepõem aos relatados neste trabalho, enquanto 3.256 dos transcritos representam potencialmente novos splicings alternativos. A relevância desse trabalho consiste em que todo o conjunto de dados deste trabalho pode ser baixado ou acessado por meio de uma interface da Web em <http://venanciogroup.uenf.br/resources/>. Assim, os dados encontrados nesse trabalho formam um atlas transcriptômico de acesso público, podendo auxiliar pesquisas em áreas como genética, genômica e transcriptômica. Desempenhando, dessa forma, um importante papel em programas de biotecnologia e melhoramento da soja.