



**XII** Congresso  
Fluminense  
de Iniciação Científica  
e Tecnológica

**V** Congresso  
Fluminense  
de Pós-Graduação

Ciência para o Desenvolvimento Sustentável

## Identificação de genes de resistência a estresse biótico em soja (*Glycine max*) a partir da integração de GWAS e redes de coexpressão gênica

*Fabrício de Almeida Silva, Dayana Kelly Turquetti Moraes, Thiago Motta Venancio*

A soja (*Glycine max* (L.) Merr.) é a principal leguminosa produzida no mundo, usada para o consumo humano e animal, além de aplicações industriais diversas. As doenças e pragas geram perdas consideráveis para produtores de soja em todo o mundo. Diversos estudos associativos identificaram marcadores moleculares relacionados a resistência contra o estresse biótico no genoma da soja, mas a identificação de genes causais a partir dos marcadores ainda é um desafio. Nesse estudo, integramos dados públicos de marcadores moleculares e transcriptoma de soja para reconstruir redes de coexpressão gênica e identificar genes candidatos de alta confiança envolvidos na resistência a estresse biótico. Através de uma extensiva curadoria na literatura, identificamos 685 marcadores associados à resistência a pragas e patógenos, dentre os quais 293 são associados a fungos, 154 a oomicetos, 110 a insetos, 102 a nematódeos e 26 a vírus. Usando os pacotes GenomicRanges e IRanges em ambiente R, selecionamos os genes localizados em intervalos de 1 Mb de distância de cada marcador, obtendo um total de 17375, 9299, 8994, 5644 e 931 genes candidatos para a resistência a fungos, nematódeos, oomicetos, insetos e vírus, respectivamente. Os dados de transcriptoma de soja sob estresse causado por fungos, insetos e oomicetos foram baixados do banco de dados *Soybean Expression Atlas*. Genes codificadores de proteínas relacionadas à patogênese (PR) foram obtidos do banco de dados UniProt e usados como genes-guia. Para cada conjunto de dados de transcriptoma, removemos genes cujos valores medianos de expressão (TPM) eram inferiores a 1 e, em seguida, selecionamos 80% dos genes com maiores valores de variância. O pacote BioNERO foi usado para ajuste dos dados, normalização quantil e reconstrução de redes de coexpressão gênica. Os genes candidatos de alta confiança foram aqueles que tiveram indução ou inibição significativa da expressão durante o estresse e que foram coexpressos com os genes-guia específicos para a resposta a fungos, insetos e oomicetos. A filtragem sistemática reduziu de modo considerável o número inicial de genes candidatos. Análises de enriquecimento funcional revelaram que os genes candidatos de alta confiança estão envolvidos em processos biológicos e vias metabólicas importantes para a defesa contra patógenos e pragas.

Palavras-chave: genômica populacional, QTL, bioinformática.

Instituições de fomento: CAPES, CNPq, UENF.