

SANTOS; Eduardo Pereira da Silva<sup>1</sup>, LIMA; Marcelo de Freitas<sup>2</sup>

## RESUMO

**Projeto:** PIC1804-2020 Variedades comerciais são essenciais para a humanidade por compor a base da alimentação de diferentes culturas. Essas variedades estão submetidas a estresses bióticos e abióticos e a seleção dos genótipos resistentes/tolerantes é essencial para manutenção da produção. Portanto, é necessário fortalecer a criação ou atualização de bancos de dados e ferramentas de visualização de interação proteica que permitam estudar os mecanismos de respostas das espécies de interesse a estresses específicos. Dentre os diversos campos do saber, a área de bioinformática permite realizar estudos de expressão gênica e ortologia/paralogia. Desde 2019, integrantes do Laboratório de Biologia Molecular e Biotecnologia (LBMB/DBQ/IQ) desenvolvem a ODAP, uma plataforma de análise de dados voltada para a área de bioinformática que tem como objetivo a realização de estudos de genômica e transcriptômica em diferentes espécies. No período 2020-21, uma ferramenta de ortologia/paralogia foi adicionado a plataforma com o objetivo geral de filtrar genes que possuem ancestralidade compartilhada através de um evento de especiação e poderão ajudar a desvendar as vias conservadas de resposta a estresses presentes em espécies próximas ou distantes evolutivamente. Para tal, inicialmente, foi gerado um banco de dados genético para cada espécie selecionada em conjunto com a anotação gênica. Posteriormente, foi elaborado um método de identificação das relações de ortologia/paralogia para confecção do banco de dados com intuito de integrá-lo à ODAP. As descrições funcionais disponíveis foram obtidas de bancos dedicados aos projetos genomas ou do *National Center for Biotechnology Information* (NCBI). Optou-se pelo sistema gerenciador de banco de dados *Postgresql* para armazenar dados em tabelas e o *OrthoMCL* para determinação da similaridade de sequências e formação de grupos de genes evolutivamente relacionados. Depois, esses dados foram complementados por execução do alinhamento múltiplo de sequências e construção de árvores filogenéticas. A visualização é realizada via ferramentas *web* que permitem a exibição de grafos e dados customizados. Após a confecção de *scripts* direcionados para o reconhecimento de genes ortólogos/parálogos, o resultado foi a atualização da plataforma ODAP para a versão 2 (V.2). Neste momento, o operador da plataforma conseguirá orientar sua busca de alvos existentes em listas de genes diferencialmente expressos de duas ou mais espécies e reconhecer quais podem assumir funções correlatas ou não em resposta aos tratamentos. Para validar a nova ferramenta, diferentes sequenciamentos de RNA (RNA-seq) de *Arabidopsis* e arroz foram selecionados de acordo com o tipo de estresse abiótico (ex: salino) e submetidos ao *pipeline* completo de análise da plataforma. Os resultados indicaram que a integração da nova ferramenta foi um sucesso e genes ortólogos/parálogos foram observados em listas de genes das duas espécies submetidas a tratamento similar. Através da plataforma ODAP V.2, é possível, a partir dos dados brutos oriundos de um RNA-seq, identificar os genes diferencialmente expressos, genes ortólogos/parálogos e redes de interação proteica em grupos e subgrupos pré-determinados. Ao longo dos anos, a plataforma vem agregando novas ferramentas com a finalidade de prestar diferentes serviços à comunidade científica e grupos empresariais ao dar suporte *in silico* aos estudos desenvolvidos *in vitro* e *in vivo*.

**PALAVRAS-CHAVE:** RNAseq, biomassa, bioinformática, ortólogos

<sup>1</sup> UFRJ, eduardopsantos@ufrj.br

<sup>2</sup> UFRJ, mflima@ufrj.br

