



## Caracterização de RNAs longos não codificantes (lncRNAs) em milho submetido a colonização pela bactéria endofítica diazotrófica *Herbaspirillum seropedicae* e adição de ácido húmico

Leandro de Oliveira Silva, Lucas Maciel Vieira, Maria Emilia M. T. Walter, Thiago M. Venâncio, Fábio Lopes Olivares, Clicia Grativol

O milho (*Zea mays*) é uma das mais importantes culturas de cereais, empregado na alimentação humana, produção de ração animal e de bioenergia. Para otimizar seu cultivo é possível utilizar bactérias diazotróficas e promotoras de crescimento. Contudo, a complexa relação planta-bactéria benéfica ainda não foi completamente elucidada, o que inclui o papel dos longos RNAs não codificantes (lncRNAs). Até o presente momento, não foram publicados trabalhos que mostram a participação de lncRNAs de plantas em relação simbiótica. O presente trabalho teve como objetivo caracterizar, por meio de análises de bioinformática, lncRNAs expressos em milho submetido a colonização pela bactéria diazotrófica *Herbaspirillum seropedicae* com adição de ácido húmico. Seis bibliotecas de RNA-seq de milho foram utilizadas para a identificação e caracterização dos lncRNAs novos e dos já conhecidos. Reads foram alinhados ao genoma de *Z. mays* (B73 RefGen\_v4) com o uso do HISAT2 e StringTie foi utilizado para montar os transcritos. Um transcriptoma único foi obtido, com 192.521 transcritos. A expressão diferencial dos lncRNAs foi calculada por meio de DESeq2. Um modelo baseado em *machine learning* foi criado para aumentar a acurácia da predição de novos lncRNAs em milho. Buscas na literatura e em bancos de dados permitiram obter 124.739 diferentes lncRNAs, empregados na construção de um banco de sequências para o cálculo da expressão dos lncRNAs conhecidos. Por meio de BLASTN com o transcriptoma, 32.645 lncRNAs conhecidos foram identificados. Destes, 3.263 lncRNAs foram diferencialmente expressos (1.502 up-regulados e 1.761 down-regulados). Os transcritos montados foram submetidos a diferentes ferramentas, como forma de identificar os novos lncRNAs e excluir sequências codificantes: filtro de tamanho; cálculo do potencial de codificação de proteína (CPC); cálculo do tamanho da ORF por meio de TransDecoder; alinhamento contra um banco de ncRNAs (Rfam) e, por fim, análise com o modelo. O total de 6.736 novos lncRNAs de alta confiabilidade foram identificados, sendo 1.050 deles diferencialmente expressos (480 up-regulados e 570 down-regulados). Os próximos passos incluem: predição dos genes-alvo dos lncRNAs e caracterização por meio de Gene Ontology (GO) e Kyoto Encyclopedia of Genes and Genomes (KEGG); verificação de mecanismos de splicing alternativo; mapeamento dos lncRNAs em bibliotecas públicas de miRNAs e validação de seus alvos por meio de bibliotecas de degradoma; construção de uma rede da interação lncRNA-miRNA de milho. Espera-se, com este trabalho, contribuir para uma maior compreensão dos papéis desempenhados pelos lncRNAs de milho.

Instituição do Programa de IC, IT ou PG: UENF  
Fomento da bolsa: FAPERJ-UENF