

Pipeline para a investigação de *Target Mimics* em regiões intergênicas em plantas

Pipeline to investigate *Target Mimics* in intergenic regions in plants

Tharcísio Soares de Amorim
amorim@alunos.utfpr.edu.br
Universidade Tecnológica
Federal do Paraná, Cornélio
Procópio, Paraná, Brasil

Alexandre Rossi Paschoal
paschoal@utfpr.edu.br
Universidade Tecnológica
Federal do Paraná, Cornélio
Procópio, Paraná, Brasil

RESUMO

Os MicroRNAs ou miRNAs compõem o tipo de pequeno RNA (*small RNAs* - sRNA) mais estudados. O miRNA maduro interage com o RNA mensageiro (mRNA) desempenhando um papel regulatório essencial na expressão de genes nas plantas e animais. Os RNAs competidores endógenos (ceRNAs) são transcrições que possuem uma região similar ou igual aos de ligação do miRNAs:mRNA, também conhecida como MRE. Deste modo, os ceRNAs podem regular os miRNA em nível de pós-transcrição ao competir pelo mesmo, liberando os mRNAs para tradução. Em 2014, Ye e co-autores identificaram 42.965 ceRNAs ou *Target Mimics* (TM em plantas) para o genoma de soja, a partir de predição computacional. Entretanto, existe uma carência de informação funcional documentada para estes transcritos. Neste trabalho, nós expandimos essa análise para 28 genomas disponíveis no Ensembl Plants. Para isso, desenvolvemos um pipeline para realizar a análise em larga escala dos TMs. Em seguida, usamos os potenciais *Target Mimics* encontrados e os comparamos com anotações de ncRNA do Ensembl Plants. Nossos resultados mostram que uma grande quantidade de potenciais *Target Mimics* pode ser encontrada em regiões intergenicas e alguns deles são relacionados à ncRNAs conhecidos. Os resultados irão contribuir para a pesquisa de TMs em plantas

PALAVRAS-CHAVE: Non-coding RNA, pipeline, bioinformática, ceRNA, miRNA, *Target Mimics*.

ABSTRACT

The MicroRNAs (miRNAs) are the most studied non-coding RNA (ncRNA). miRNAs interaction with Messenger RNAs (mRNAs) can regulate its expression by translational repression or degradation. However, miRNAs can interact with other RNA type and increase the complex gene regulation network. Competing Endogenous RNAs (ceRNAs), also known as *Target Mimics* (TMs), are transcripts which have a similar or same miRNA binding site. In this case, the TM regulate the miRNAs in a post-transcriptional level, leaving these mRNAs free to be translated. Ye and co-authors identified 42,965 potential *Target Mimics* in intergenic regions in *Glycine max*, by *in silico* analysis. However, there is a lack of functional information for these transcripts. In this work, we expanded their analysis to 26 genomes available at Ensembl Plants. For that, we first developed a pipeline to perform the large-scale TM analysis. Secondly, we used the potential *Target Mimics* found and compared it to the ncRNA annotation from Ensembl Plants. Our results show that a considerable amount of potential *Target Mimics* can be found in intergenic regions and some of them related to known ncRNAs. The results will contribute to the TM research in plants.

Recebido: 04 out. 2018.

Aprovado: 04 fev. 2019.

Direito autorial:

Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.





KEYWORDS: Non-coding RNA, pipeline, bioinformatics, ceRNA, miRNA, Target Mimics.

INTRODUÇÃO

A bioinformática é um campo de estudo interdisciplinar que envolve, principalmente, biologia molecular e genética, ciência da computação, matemática e estatística (CAN *et al.*, 2014). Nela, problemas biológicos com dados de larga escala são resolvidos com abordagem computacional. Dentre diversas áreas de estudo na bioinformática, existem ramos de pesquisa sobre os RNAs não-codificantes (ncRNAs). Os ncRNAs são ácidos ribonucleicos (RNAs) que são transcritos, mas não são traduzidos. De todos ncRNAs, os miRNAs, que compõem um tipo de RNA pequeno (smRNA), é o mais estudado. O miRNA maduro é capaz de interagir com o RNA mensageiro (mRNA) e desempenha um papel regulatório importante na expressão de genes de plantas e animais (CHITWOOD *et al.*, 2007). Os RNAs competidores endógenos (ceRNAs), também conhecidos como Target Mimics (TMs) em plantas, são transcrições que possuem uma região similar ou igual aos de ligação do miRNAs:mRNA, também conhecida como MRE. Deste modo, os ceRNAs podem regular os miRNAs em nível de pós-transcrição ao competir pelo mesmo, liberando os mRNAs para tradução (PASCHOAL *et al.*, 2017).

Considerando este cenário, este projeto foi desenvolvido baseado nas pesquisas realizadas por Ye e coautores (2014), que analisaram as regiões intergênicas do genoma da soja e encontraram 42.965 Target Mimics em potenciais. Entretanto, os autores não analisaram que tipo de transcritos eram esses TMs. Com a mesma ideia, foi desenvolvido um pipeline de modo a expandir as abordagens usadas por Ye e coautores. Nesse sentido, 26 genomas de plantas, disponíveis no site Ensembl Plants (BOLSER *et al.*, 2016), foram analisados. Em um segundo momento, os TMs potenciais foram classificados funcionalmente, através de comparação com anotações de RNAs não codificantes (ncRNAs), também disponíveis no Ensembl Plants.

Espera-se com esse trabalho contribuir para a informação funcional de RNAs reguladores endógenos de miRNA em planta, o que pode ajudar em potenciais terapias gênicas na manipulação biológica de plantas.

METODOLOGIA

EXTRAÇÃO DE REGIÕES INTERGÊNICAS

Para extração das regiões intergênicas, foram utilizados o software Bedtools (QUINLAN *et al.*, 2010) e comandos Unix. Primeiramente, no arquivo de indexação do genoma (GFF3), são selecionadas todas as tuplas cuja terceira coluna diz "gene", o que significa que esta representa a coordenada de uma região gênica. Em seguida, o arquivo extraído do primeiro passo é ordenado de acordo com o tamanho do cromossomo, retirado do arquivo de anotação do genoma (FASTA). Por último, através do comando *complement* do software Bedtools, é retirado o complemento da região gênica, que é a região intergênica.

encontrado e do ncRNA, e verificada a existência de sobreposição de coordenadas de localização no genoma.

RESULTADOS E DISCUSSÕES

As plantas analisadas, quantidade de Target Mimics em potencial encontrados e a quantidade de TMs classificados por tipo estão registrados na Tabela 1.

Tabela 1 – Plantas, TMs potenciais e TMs classificados por tipo

Plantas	TMs	tRNA	snoRNA	rRNA	snRN	pre_miRN
					A	A
<i>Aegilops Tauschii</i>	16989	0	0	1	0	0
<i>Amborella Trichopoda</i>	3777	0	0	0	0	0
<i>Arabidopsis lyrata</i>	2468	0	0	0	0	0
<i>Arabidopsis thaliana</i>	2753	1	0	0	0	3
<i>Brachypodium distachyon</i>	2859	0	0	1	0	1
<i>Brassica Napus</i>	608	0	0	0	0	0
<i>Brassica oleracea</i>	75	0	0	1	0	0
<i>Brassica rapa</i>	1086	1	0	0	0	0
<i>Chlamydomonas reinhardtii</i>	572	0	0	0	0	0
<i>Glycine max</i>	53909	0	0	0	0	0
<i>Gossypium raimondii</i>	5017	1	0	0	0	0
<i>Helianthus Annuus</i>	410	0	0	0	0	0
<i>Hordeum vulgare</i>	22992	0	0	0	0	0
<i>Manihot esculenta</i>	1028	0	0	0	0	0
<i>Medicago truncatula</i>	6223	0	0	1	0	1
<i>Oryza sativa japonica</i>	9599					
<i>Phaseolus Vulgaris</i>	95	0	0	0	0	0
<i>Populus trichocarpa</i>	3204	0	0	2	0	0
<i>Prunus persica</i>	1321	0	0	0	0	0
<i>Selaginella moellendorffii</i>	361	0	0	0	0	0
<i>Solanum lycopersicum</i>	2384	0	0	0	0	0
<i>Solanum tuberosum</i>	5312	0	0	0	1	0
<i>Sorghum bicholor</i>	8155	0	0	14	0	1
<i>Theobroma cacao</i>	316	0	0	1	0	0
<i>Vitis vinifera</i>	1257	0	0	0	0	0
<i>Zea mays</i>	17019	0	0	0	0	0

Fonte: Autor

Na tabela 1, é possível observar que um grande número de Target Mimics em potencial foi encontrado. Já a classificação funcional mostra que apenas um pequeno percentual desse número possui evidência de ncRNAs documentados. Isso se deve ao fato que existe uma carência de anotação de ncRNAs, sobretudo



em regiões intergênicas. Existem outras abordagens capazes de buscar novos ncRNAs, bem como suas funções, em anotações de transcritos disponíveis em bancos de dados públicos, que poderão ser utilizadas em trabalhos futuros. Página | 5 A principal contribuição desse trabalho, entretanto, é a automatização do processo de busca de TMs em regiões intergênicas, estabelecida pelo pipeline. O processo de busca pode ser expandido para toda planta com anotação genética disponível e possibilita a criação de bancos de dados específicos para TMs e outros transcritos.

CONCLUSÕES

O projeto de pesquisa foi executado conforme a proposta estabelecida. Foram analisadas 26 plantas com anotação genômica disponível no site Ensembl Plants e seus respectivos miRNAs anotados no site miRBase. Os resultados mostram que uma grande quantidade de *Target Mimics* em potencial pode ser encontrada nas regiões intergênicas dessas plantas. Entretanto, ao realizar a classificação funcional, é notável que apenas uma pequena quantidade é encontrada, que são 30 classificações ao todo. Isso pode ser atribuído ao fato que existe uma carência de anotação de ncRNAs dos genomas.

REFERÊNCIAS

CAN, Tolga. **Introduction to bioinformatics**. In: miRNomics: MicroRNA Biology and Computational Analysis. Humana Press, Totowa, NJ, 2014. p. 51-71.

CHITWOOD, Daniel H.; TIMMERMANS, Marja CP. Target mimics modulate miRNAs. *Nature genetics*, v. 39, n. 8, p. 935, 2007.

PASCHOAL, Alexandre Rossi et al. **ceRNAs in plants: computational approaches and associated challenges for target mimic research**. Briefings in bioinformatics, 2017.

YE, Chu-Yu et al. **Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean**. *Frontiers in plant science*, v. 5, p. 743, 2014.

QUINLAN, Aaron R.; HALL, Ira M. **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics*, v. 26, n. 6, p. 841-842, 2010.

GRIFFITHS-JONES, Sam et al. **miRBase: microRNA sequences, targets and gene nomenclature**. *Nucleic acids research*, v. 34, n. suppl_1, p. D140-D144, 2006.

FRANCO-ZORRILLA, José Manuel et al. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature genetics*, v. 39, n. 8, p. 1033, 2007.

BOLSER, Dan et al. **Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data**. In: *Plant Bioinformatics*. Humana Press, New York, NY, 2016. p. 115-140.