

Investigação de Target Mimics em dados públicos de lncRNAs

Target Mimics investigation in lncRNA public data

RESUMO

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

Os MicroRNAs (miRNAs) compõem o tipo de pequeno RNA (*small RNAs* - sRNA) mais estudados. O miRNA maduro interage com o RNA mensageiro (mRNA) desempenhando um importante papel regulatório na expressão de genes nas plantas e animais. Os RNAs competidores endógenos (ceRNAs) são transcrições que possuem uma região similar ou igual a da 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 2017, Yuan e coautores disponibilizaram o banco de dados PceRBase, onde foram armazenados Target Mimics de 28 espécies de plantas, encontrados através de um pipeline de autoria própria. Neste contexto, este trabalho foi realizado com os objetivos de: atualizar a literatura, encontrando novos transcritos; desenvolver um pipeline próprio para encontrar novos Target Mimics, automatizando o processo para qualquer espécie de planta, e; realizar análise em espécies não analisadas anteriormente. Os resultados irão contribuir para a pesquisa de TMs em plantas

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

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

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



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. Yuan and co-authors made available PceRBase, a Target mimics database in over 25 plant species. In this scenario, the goals of this work are: to update the literature by finding new transcripts; to develop our own pipeline, expanding the process to any plant species, and; to analyse new plant species. The results will contribute to the TM research in plants.

KEYWORDS: Keyword one. Keyword two. Keyword three.

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 os ncRNAs, os miRNAs, que compõem um tipo de RNA pequeno (smRNA), são os mais estudados. 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 Yuan e coautores (2017), que analisaram dados públicos de 28 genomas de plantas, com o intuito de encontrar Target Mimics. Com a mesma ideia, foi desenvolvido um pipeline próprio para investigação de Target Mimics. Nesse sentido, 35 genomas de plantas, disponíveis nos repositórios Ensembl Plants (BOLSER *et al.*, 2016), CANTATADB (SZCZEŚNIAK *et al.*, 2015) e GreenC (PAYUVÍ *et al.*, 2015) foram analisados. Em uma próxima etapa, serão analisados dados de RNA-Seq para as plantas que possuem tal dado disponível publicamente, para verificação dos transcritos encontrados.

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

METODOLOGIA

OBTENÇÃO DE DADOS PÚBLICOS DE lncRNA

Os lncRNAs utilizados para este trabalho foram obtidos em três diferentes repositórios, com a finalidade de aumentar a quantidade e diversidade da análise. São eles: Ensembl Plants, CanTaTaDB e GreenC.

REMOÇÃO DE REDUNDÂNCIAS E PADRONIZAÇÃO DE NOMENCLATURA

Após selecionados os transcritos dos três repositórios, existe a possibilidade de um mesmo transcrito constar em dois ou mais repositórios, o que pode resultar

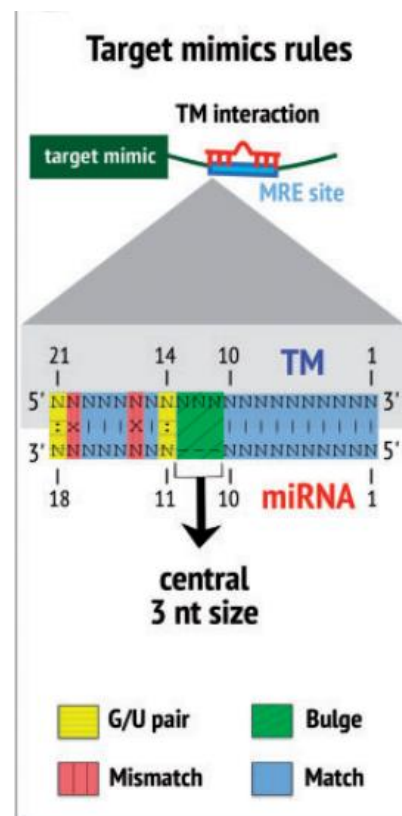
repetições desnecessárias. Para remover essas redundâncias, foi utilizado o software CD-HIT, o qual realiza comparações compara todos os transcritos dos diferentes bancos e forma clusters com as repetições.

Após a remoção das redundâncias, os lncRNAs remanescentes têm seus nomes padronizados; processo que é realizado através de um script desenvolvido em Python, de autoria própria.

IDENTIFICAÇÃO DE TARGET MIMICS EM POTENCIAL

A busca por TMs em potencial foi realizada de três formas: através de scripts em linguagem de programação Perl, disponibilizados por Ye e coautores (2014); através do software Tapir e; através do software TarHunter. A busca compara miRNAs já documentados de cada planta, disponíveis no site miRBase (GRIFFITHS-JONES *et al.*, 2006), com as sequencias de nucleotídeos dos lncRNAs obtidos nos repositórios. As regras para identificação de Target Mimics, de acordo com o trabalho publicado por Franco Zorilla (2007), são mostradas na figura 1

Figura 1 – Regras para identificação de Target Mimics



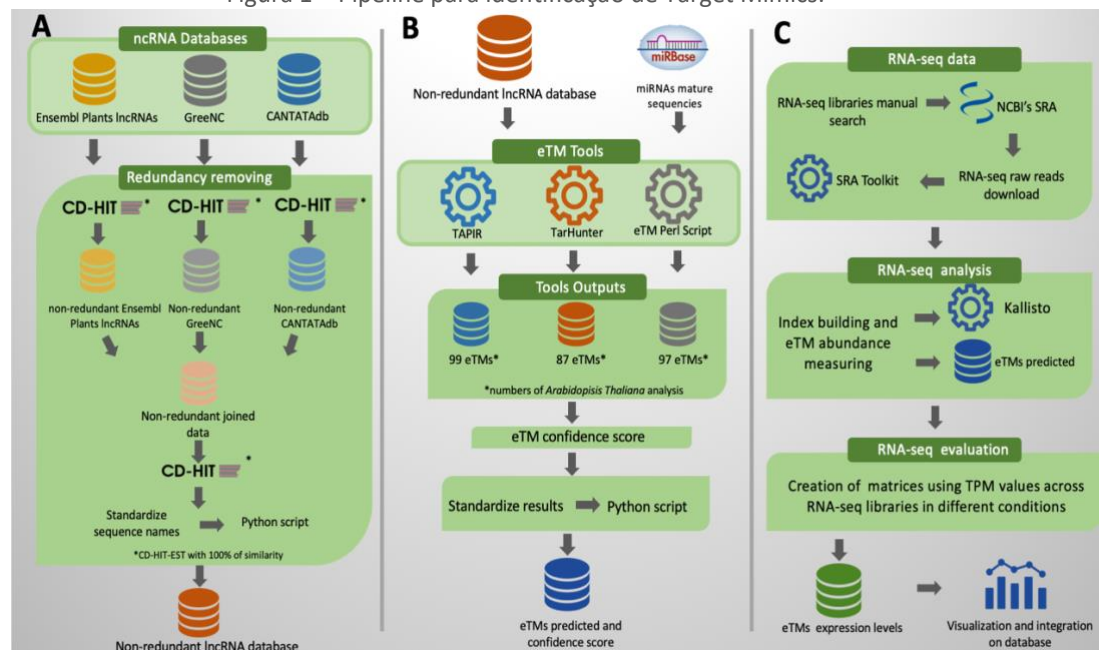
Fonte: PASCHOAL *et al.*, 2017. As regras dizem que (i) deve existir uma protuberância central de 3 nucleotídeos no TM e entre as posições 10 e 11 do miRNA, representado em verde (ii) pares Guanina/Uracila (G/U) só podem aparecer após a posição 10 do miRNA e no final da interação, representado em amarelo (iii) um ou dois *mismatches* (nucleotídeos diferentes) são tolerados após a posição 11 do miRNA, em vermelho.

ANÁLISE RNA-SEQ

Em um próximo momento, será realizada uma análise para dar confiabilidade nos transcritos encontrados na primeira etapa, para as espécies que possuem dados de RNA-seq disponibilizados publicamente. Esta etapa será realizada por pesquisadores parceiros de outros países.

A Figura 2 mostra o pipeline que representa todo o processo de identificação de Target Mimics, desde a obtenção dos dados de lncRNAs das plantas até a identificação de confiabilidade dos TMs em potenciais (análise de RNA-Seq).

Figura 1 – Pipeline para identificação de Target Mimics.



Fonte: Autor. O pipeline foi desenvolvido em três etapas. (A) é onde deve ser feito o download dos dados públicos de lncRNAs, remoção de redundâncias e padronização de nomenclatura. Na etapa (B) devem ser utilizados os dados de lncRNA padronizados e miRNAs nos softwares de busca de TMs, gerando no final um banco de dados com os TMs em potenciais. Por último, na etapa (C) devem ser utilizados os dados de RNA-seq, quando disponíveis, juntamente com os TMs em potenciais, de forma a dar confiabilidade aos dados encontrados na etapa (B).

RESULTADOS E DISCUSSÕES

As plantas analisadas e quantidade de Target Mimics em potencial encontrados estão registrados na Tabela 1.

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

Plantas	Tapir	TarHunter	Script Perl
<i>Amborella trichopoda</i>	34	30	32
<i>Arabidopsis lyrata</i>	125	23	80
<i>Arabidopsis thaliana</i>	98	46	94
<i>Brachypodium distachyon</i>	302	75	69
<i>Brassica napus</i>	68	6	23
<i>Brassica oleracea</i>	0	0	2
<i>Brassica rapa</i>	51	3	28
<i>Carica papaya</i>	9	1	2
<i>Chlamydomonas reinhardtii</i>	33	4	10
<i>Citrus clementina</i>	1	0	1
<i>Citrus sinensis</i>	22	0	7
<i>Cucumis sativus</i>	0	0	0
<i>Fragaria vesca</i>	15	0	4
<i>Glycine max</i>	182	78	160
<i>Gossypium raimondii</i>	12	5	13
<i>Helianthus annuus</i>	0	0	0
<i>Hordeum vulgare</i>	31	14	31
<i>Linum usitatissimum</i>	39	0	1
<i>Malus domestica</i>	81	21	75
<i>Manihot esculenta</i>	110	8	30
<i>Medicago truncatula</i>	172	62	143
<i>Oryza sativa</i>	294	52	110
<i>Phaseolus vulgaris</i>	0	0	1
<i>Physcomitrella patens</i>	159	43	75
<i>Populus trichocarpa</i>	108	18	52
<i>Prunus persica</i>	24	33	65
<i>Ricinus communis</i>	2	0	2
<i>Selaginella moellendorffii</i>	4	0	0
<i>Solanum lycopersicum</i>	26	3	43
<i>Solanum tuberosum</i>	99	58	66
<i>Sorghum bicolor</i>	106	20	52
<i>Theobroma cacao</i>	17	4	9
<i>Triticum aestivum</i>	213	114	156
<i>Vitis vinifera</i>	27	6	16
<i>Zea mays</i>	165	110	226

Fonte: Autor

Na tabela 1, é possível observar que um grande número de Target Mimics em potencial foi encontrado, separados por espécie e ferramenta. É notável que

algumas espécies possuem poucos TMs em potencial encontrados. Isso se deve ao fato que algumas espécies possuem pouca documentação em RNAs não-codificadas (lncRNAs e/ou miRNAs). Isso não significa necessariamente que exista poucos TMs no genoma em questão.

CONCLUSÕES

O projeto de pesquisa foi executado conforme a proposta estabelecida. Foram analisadas 36 plantas com anotação de lncRNAs disponível em pelo menos um dos três repositórios selecionados (metodologia) e seus respectivos miRNAs anotados no site miRBase. Os resultados mostram que uma grande quantidade de *Target Mimics* em potencial pode ser encontrada na maioria das plantas analisadas. A próxima etapa deste projeto consistirá na análise RNA-Seq, identificar o grau de confiabilidade dos dados encontrados. Este é um projeto em parceria com pesquisadores de universidades da Turquia e China, que finalizarão o presente trabalho.

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.

SZCZEŚNIAK, Michał W.; ROSIKIEWICZ, Wojciech; MAKALOWSKA, Izabela. **CANTATadb: a collection of plant long non-coding RNAs**. *Plant and Cell Physiology*, v. 57, n. 1, p. e8-e8, 2015

PAYTUVÍ GALLART, Andreu et al. **GREENC: a Wiki-based database of plant lncRNAs**. *Nucleic acids research*, v. 44, n. D1, p. D1161-D1166, 2015.