

sbRNAs DE INSETOS: VALIDAÇÃO ESTRUTURAL

Beatriz Benedette Pessoa (PIBIC/CNPq/FA/UEM), Katharina Tanaka (PIBIC/CNPq/FA/UEM), Flávio Augusto Vicente Seixas (Co-orientador), Daniel Caligari (Co-orientador), Maria Aparecida Fernandez (Orientador) E-mail: ra119904@uem.br

Universidade Estadual de Maringá, Centro de Ciências Biológicas e da Saúde, Maringá, PR.

Ciências Biológicas, Bioquímica, Biologia Molecular

Palavras-chave: RNAs não codificantes, sbRNA, bioinformática

RESUMO

Os ncRNAs (RNAs não codificantes) são caracterizados como segmentos transcritos funcionais, porém não traduzidos em proteínas. Os *stem-bulge* RNAs (sbRNAs) de invertebrados são homólogos aos Y RNAs de vertebrados, ambos possuindo estrutura tipo haste-loop e parte deles já descritos com funcionalidade no licenciamento da replicação do DNA. Em insetos já foram descritos um gene de sbRNAs em *Bombx mori*, *BmsbRNA*, dois em *Drosophila melanogaster*, Dm1 e Dm2, além de genes Y RNA-like, identificados somente *in silico*, em *Anopheles gambiae* (Y-RNA-like AgYRNA). Dm1 de *D. melanogaster* apresentou funcionalidade no início da replicação do DNA, como verificado nos Y RNA de vertebrados e sbRNAs do nematóide *Caenorhabditis elegans*. Entretanto, a predição da classe que estes sbRNAs de insetos pelo novo software NYCPred ainda não foi avaliado, por isso estas avaliações foram propostas para o presente projeto. Os resultados mostraram que os genes Dm1 e o putativo Y-RNA-like AgYRNA.1 são estruturalmente semelhantes aos modelos descritos para sbRNAs, porém entre os genes avaliados somente Dm1 foi anotado como sbRNAs pelo software NYCPred.

INTRODUÇÃO

As regiões não codificantes representam a maior parte dos genomas eucarióticos, dentre as quais encontram-se genes que são transcritos, porém não codificam proteínas. Estes genes são denominados RNAs não codificantes (non coding RNAs, ncRNAs), sendo seus produtos gênicos tais como os tRNA (RNAs transportadores), rRNA (RNA ribossomais), miRNA (micro RNAs que atuam como silenciadores, regulando a expressão gênica); os snRNA (pequenos RNAs não-codificantes que atuam em diversos processos moleculares) e lncRNA (longos RNAs não-codificantes que estão envolvidos em vários processos celulares como transcrição, organização da cromatina e processamento de RNA), dentre outros. Os ncRNAs são genes antes considerados sem atividade, mas hoje apontados como essenciais para todos os organismos vivos, sendo que a função destes genes ainda é muito estudada, porém

de difícil compreensão integrada (Kowalski e Krude, 2015). Dentro os ncRNAs estão os Y RNAs, sendo que os detectados em mamíferos se apresentam essenciais para o processo de iniciação da replicação do DNA. Estudos funcionais de sbRNAs descritos em *Caenorhabditis elegans* demonstraram que estes RNAs são essenciais na progressão da fase S, viabilidade da embriogênese e atuação no início da replicação. Desse modo, a identificação destes genes e pseudogenes em insetos, demonstraram um grande avanço na área do estudo de RNA não codificantes (Kowalski e Krude, 2015). A obtenção de um novo software (NYCPred: Non-Coding/Y-RNA Prediction; Lima et al., 2021) capaz de analisar sequências de genes conhecidos como RNAs não-codificantes, foi uma conquista importante nessa área de pesquisa. O programa possui capacidade de classificar rapidamente milhares de sequências entre 13 tipos diferentes de RNAs não-codificantes, com precisão superior a alguns dos métodos mais avançados. Os 4 genes putativos de YRNA-like para *Anopheles gambiae*, nunca foram validados quanto à probabilidade de que realmente possam ser sbRNAs. As descrições do gene BmsbRNA em *B. melanogaster*, nunca foram validados em um programa como o NYCPre (Duarte Junior et al., 2015; 2019). Portanto, neste trabalho, os genes descritos como sbRNA e/ou Y RNA-like de insetos foram avaliados no software NYCPre.

MATERIAIS E MÉTODOS

Para a realização inicial do projeto, foi executada uma busca das sequências de cada sbRNA-alvo no banco de dados NCBI (<https://www.ncbi.nlm.nih.gov/>), utilizando o software BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), com o objetivo de identificar com precisão as sequências dos genes que constituem o grupo de Y RNA-like de *Anopheles gambiae*; *BmsbRNAs* de *B. mori* (MN_654364.1) (Duarte Junior et al., 2015) e Dm1 (NR_167817.1) e Dm2 (MN_661249.1) de *D. melanogaster* (Duarte Junior et al., 2019). Estas sequências foram armazenadas em formato “fasta” e utilizadas como *input* no programa NYCPre (<https://www.gpea.uem.br/ncypred/submit>) para verificar a classe dos respectivos ncRNAs. Finalmente, as sequências foram submetidas na plataforma RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) para a predição da estrutura secundária.

RESULTADOS E DISCUSSÃO

Os resultados das previsões da estrutura secundária analisados pelo software RNAfold (Figura 1) mostram que os genes *DmsbRNA1* (Dm1) de *D. melanogaster* e o putativo Y-RNA-like AgYRNA.1 de *A. gambiae* são estruturalmente semelhantes aos modelos descritos para sbRNAs de *C. elegans* (revisão em Kowalski e Krude, 2015), que inclui, por exemplo, hastes inferiores e superiores separadas por uma protuberância de citosina, uma cauda média e central, além de uma cauda de poliuridina e loop central. Para outras sequências, foram encontrados modelos relativamente diferentes com a estrutura de consenso descrita, o que pode ser

explicado pelos parâmetros do algoritmo de predição analisados em "modo padrão", ou que existem variações toleradas para a atuação funcional dos genes em questão.

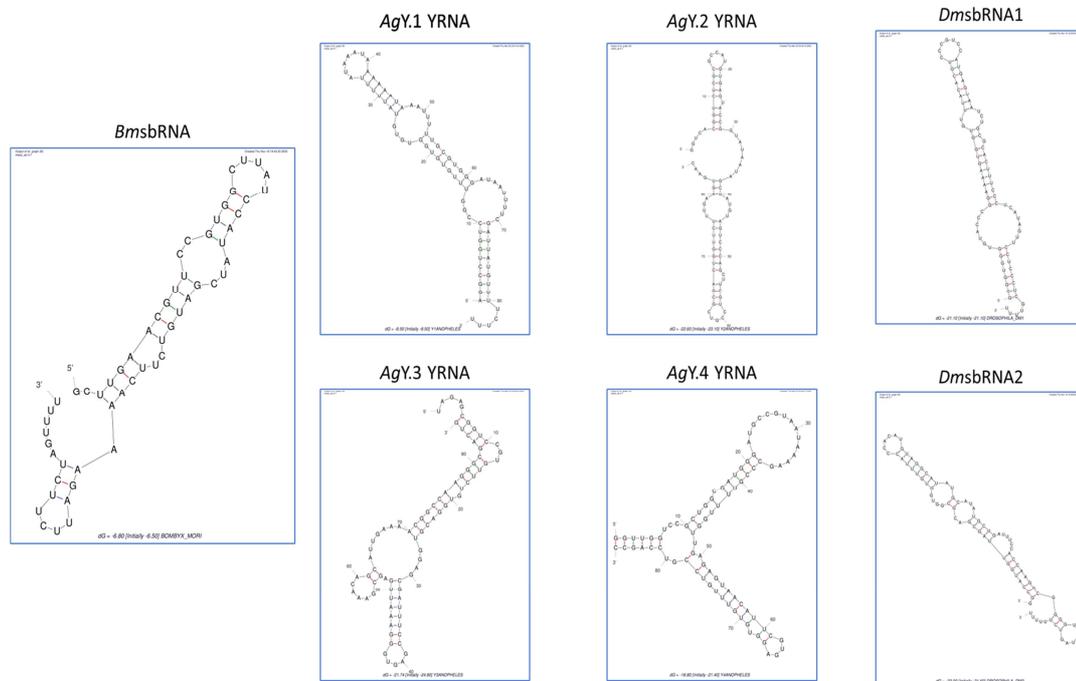


Figura 1. Predição da estrutura secundária mais estável dos sbRNAs e YRNAs-like analisados pelo software RNAfold. *BmsbRNA* de *Bombyx mori*; *AgY.1*, *AgY.2*, *AgY.3* e *AgY.4* de *Anopheles gambiae*; *DmsbRNA1* (Dm1) e *DmsbRNA2* de *Drosophila melanogaster*.

Os resultados da submissão no software NYCPrede mostraram que todas as sequências foram anotadas como microRNAs (miRNAs), exceto as sequências *DmsbRNA1*, *DmsbRNA2* e *AgYRNA3* (Tabela 1). O resultado de Y RNA-like para *DmsbRNA1* (Dm1) de *D. melanogaster* confirmou o resultado de funcionalidade para esse gene descrita anteriormente, isto é, como um sbRNA que atua na iniciação da replicação de DNA (Duarte Junior *et al.*, 2019). O diferencial descrito no software NYCPrede para os genes *DmsbRNA2* (Dm2) de *D. melanogaster* como tRNA e de intron-gp II para o gene *AgY.3* de *A. gambiae* leva a discussão sobre a classe exata que esses genes pertencem. Este software realiza verificação de homologia entre sequências de consenso previamente armazenadas no banco de dados do programa, o que pode demonstrar uma limitação em relação aos sbRNAs. A identidade com microRNA e outros ncRNAs indica que as regiões aqui analisadas podem conter regiões derivadas da mesma estrutura.

Tabela das classes de ncRNAs

id	seq	prediction	5.8S-rRNA	5S-rRNA	CD-box	HACA-box	Intron-gp	Intron-gp-I	Leader	Riboswitch	Ribozyme	Y-RNA	Y-RNA-like	miRNA	tRNA
DmsbRNA1	GGGGTGGGTGTACCCGAAAGTGGTY-RNA		0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.513	0.000	0.430	0.046
DmsbRNA2	GGCCATGGTTAGCGACGCGGTGGT tRNA		0.000	0.000	0.011	0.007	0.004	0.000	0.073	0.002	0.000	0.000	0.043	0.007	0.850
BmsbRNA	GCTTGAACGTTCCGTGGCTTATCCATA miRNA		0.000	0.000	0.006	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.963	0.003
AgY1.1	AGGCCTGGTCCGGTTTGTGGTGTG miRNA		0.000	0.000	0.024	0.090	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.883	0.000
AgY1.2	GGTCACGGTCTCGGCCATGTGAGTA miRNA		0.000	0.000	0.013	0.052	0.000	0.004	0.000	0.001	0.002	0.000	0.000	0.928	0.000
AgY1.3	TAGAGCGGTCGGTGTCTGTGGACGT Intron-gp-II		0.000	0.000	0.000	0.001	0.001	0.985	0.001	0.001	0.001	0.000	0.006	0.001	0.003
AgY1.4	GGTTGGTCCGTCGGCATGGGATGCC miRNA		0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.998	0.000

Tabela 1. Anotação das classes de ncRNAs previstas no software NYCPre. *BmsbRNA* de *Bombyx mori*; *AgY1.1*, *AgY1.2*, *AgY1.3* e *AgY1.4* de *Anopheles gambiae*; *DmsbRNA1* (Dm1) e *DmsbRNA2* de *Drosophila melanogaster*.

CONCLUSÕES

Em suma, o estudo investigou sbRNAs e Y RNAs-like em insetos, usando NYCPre para classificá-los. Resultados indicaram similaridades estruturais com sbRNAs, apesar da classificação como microRNAs pelo NYCPre, enfatizando desafios na classificação de ncRNAs e a relevância contínua da pesquisa para entender a biologia molecular.

AGRADECIMENTOS

Agradeço ao CNPq pelo auxílio financeiro para a realização do presente trabalho.

REFERÊNCIAS

- DUARTE JUNIOR F. F; LIMA NETO Q. A, et al. Identification and molecular structure analysis of a new noncoding RNA, a sbRNA homolog, in the silkworm *Bombyx mori* genome. **Mol BioSyst**, v. 11, p. 801-808, 2015.
- DUARTE JUNIOR F. F; BUENO P. S. A; PEDERSEN S. L; RANDO F. S; PATTARO JÚNIOR J. R; CALIGARI D, et al. Identification and characterization of stem-bulge RNAs in *Drosophila melanogaster*. **RNA Biology**, v. 16, p. 330-339, 2019.
- KOWALSKI M. P AND KRUDE T. Functional roles of non-coding Y RNAs. **The Int J of Bioch & Cell Biol**, v.66, p. 20–29, 2015.
- LIMA D.S.; AMICHI L.J.A.; FERNANDEZ M.A.; CONSTANTINO A.A.; SEIXAS F. A. V. NCYPred: A Bidirectional LSTM Network With Attention for Y RNA and Short Non-Coding RNA Classification. **IEEE/ACM Trans Comput Biol Bioinform**, v. 20, p. 557-565, 2021.