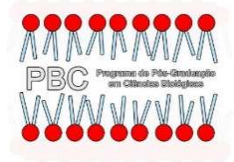

UNIVERSIDADE ESTADUAL DE MARINGÁ

CENTRO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

ÁREA DE CONCENTRAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Francisco Ferreira Duarte Junior

sbRNAs DE INSETOS: DESCRIÇÃO E ANÁLISE DA SUA INTERAÇÃO COM PROTEÍNAS DA CROMATINA

Maringá

Janeiro – 2020

Francisco Ferreira Duarte Junior

sbRNAs DE INSETOS: DESCRIÇÃO E ANÁLISE DA SUA INTERAÇÃO COM PROTEÍNAS DA CROMATINA

Tese apresentada ao curso de Pós-graduação em Ciências Biológicas (Área de concentração: Biologia Celular e Molecular) da Universidade Estadual de Maringá, para obtenção do grau de Doutor em Ciências Biológicas.

Orientadora: Dra. Maria Aparecida Fernandez

Co-orientador: Dr. Quirino Alves de Lima Neto

Maringá

2020

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

Duarte Junior, Francisco Ferreira
D812s sbRNAs de insetos : descrição e análise da sua interação com proteínas da cromatina / Francisco Ferreira Duarte Junior. -- Maringá, 2020.
69 f. : il. color., figs., tabs.

Orientadora: Profª. Drª. Maria Aparecida Fernandez.
Coorientador: Prof. Dr. Quirino Alves de Lima Neto.
Tese (Doutorado) - Universidade Estadual de Maringá, Centro de Ciências Biológicas, Programa de Programa de Pós-Graduação em Ciências Biológicas (Biologia Celular), 2020.

1. RNA não-codificante. 2. *Drosophila melanogaster*. 3. *stem-bulge* RNAs (sbRNAs). 4. Y RNA. I. Fernandez, Maria Aparecida , orient. II. Lima Neto, Quirino Alves de , coorient. III. Universidade Estadual de Maringá. Centro de Ciências Biológicas. Programa de Pós-Graduação em Ciências Biológicas (Biologia Celular). IV. Título.

CDD 23.ed. 572.88

Síntique Raquel Eleuterio – CRB 9/1641

Francisco Ferreira Duarte Junior

sbRNAs DE INSETOS: DESCRIÇÃO E ANÁLISE DA SUA INTERAÇÃO COM PROTEÍNAS DA CROMATINA

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas do Centro de Ciências Biológicas da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre/Doutor em Ciências Biológicas, sob apreciação da seguinte banca examinadora:

Aprovado em: 17 / 02 / 2020

Profa Dra Maria Aparecida Fernandez
(Presidente)

Profa Dra Nádia Monesi
(Membro examinador externo- USP)

Prof. Dr. Marco Aurélio Schüler de Oliveira
(Membro examinador- UEM)

Profa Dra Emy Luiza Ishii Iwamoto
(Membro examinador- UEM)

Prof. Dr. Wanderley Dantas dos Santos
(Membro examinador- UEM)

BIOGRAFIA

Francisco Ferreira Duarte Junior, filho de Francisco Ferreira Duarte e Anna Maria de Jesus, nasceu na cidade de Maringá, no Estado do Paraná, em 31 de março de 1990. No ano de 2004, concluiu o Ensino Fundamental no Colégio Adventista de Maringá. Concluiu o Ensino Médio na mesma instituição, em 2007. No ano de 2009, iniciou a graduação em Ciências Biológicas pela Universidade Estadual de Maringá - UEM, concluindo em 2013. No ano seguinte, ingressou no Mestrado do Programa de Pós-Graduação em Ciências Biológicas (Biologia Celular e Molecular) da UEM, tendo defendido sua Dissertação no ano de 2016. Atualmente é aluno de Doutorado no mesmo programa. Em 2017, concluiu o curso de Tecnólogo em Análise e Desenvolvimento de Sistemas pelo Centro de Ensino Superior de Maringá - Unicesumar. Desde seu trabalho de conclusão de curso, vem trabalhando com a descoberta, expressão e interação de genes de *stem-bulge* RNAs (sbRNAs), pertencentes à família dos *non-coding* RNAs, sendo integrante do grupo pioneiro na detecção destas moléculas em insetos, como o bicho-da-seda (*Bombyx mori*) e a mosca da fruta (*Drosophila melanogaster*). No ano de 2019, iniciou uma sociedade na startup Helix Biotecnologia, selecionada no edital Sinapse da Inovação da Fundação Araucária (Chamada Pública 022/2018) para prestação de serviços com sequenciamento de DNA.

AGRADECIMENTOS

A Deus, por me dar o presente da vida e a força para superar tudo.

À Profa. Dra. Maria Aparecida Fernandez, por sua orientação, paciência, confiança e por me dar a oportunidade de realizar esta etapa do meu sonho. Por me dar o direcionamento, conhecimento e compartilhar suas experiências comigo.

Ao meu amigo e co-orientador, Prof. Dr. Quirino Alves de Lima Neto, por todos estes anos de ensinamentos constantes, pensamentos críticos e inspiração, além de sua amizade. O senhor vai longe!!

Aos meus familiares, em especial à minha mãe e meu pai, meus grandes exemplos de dedicação, garra e trabalho. Obrigado por todo seu amor, apoio e compreensão, por todo investimento e dedicação para me tornarem uma pessoa melhor. Espero que tenham tanto orgulho de mim, quanto eu tenho de vocês.

A todos os meus colegas do Laboratório de Organização Funcional do Núcleo (LORF), por me auxiliarem e dedicarem uma porção de seu tempo comigo. Em especial aos meus amigos José Renato Pattaro Júnior, Fabiana dos Santos Rando, Paulo Bueno e Daniel Caligari, que tiveram paciência para me aconselhar, disposição em me ajudar e pelo conhecimento compartilhado. Não posso deixar de agradecer à auxiliar de laboratório e colega, Marli Licero Schuete Silva, pelas diversas atividades que desempenha no laboratório e por sua disposição em nos ajudar na rotina laboratorial.

A todos os meus amigos, em especial, Denison Martins, Lucas Vieira e Ricardo Drozino, pelas conversas, conselhos, apoio e incentivo.

Ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Estadual de Maringá.

À Universidade Estadual de Maringá – UEM e ao Complexo de Centrais de Apoio à Pesquisa – COMCAP/UEM.

À Fundação Araucária, CAPES e CNPq, pelo apoio financeiro.

A todos os membros da banca examinadora, pelo tempo disponibilizado, conhecimento e conselhos, críticas e elogios.

A todas as outras pessoas, não mencionadas aqui, que contribuíram de alguma forma para a realização deste trabalho.

Obrigado!!!

APRESENTAÇÃO

Em consonância com as regras do Programa de Pós-Graduação em Ciências Biológicas, esta tese é composta por um resumo geral do trabalho em português e em inglês, e dois artigos de pesquisa científica, originais, contemplando os resultados obtidos ao longo do desenvolvimento do doutorado. O primeiro artigo, que foi redigido e publicado na revista *RNA Biology* (ISSN: 1555-8584/ *Impact Factor* JCR: 5,28), compreende resultados referentes ao período de 2016 a 2018. Nele, são descritos os primeiros genes de *stem-bulge* RNAs (sbRNAs) em *Drosophila melanogaster*, batizados de Dm1 e Dm2. O segundo artigo foi redigido e será submetido para a mesma revista, apresentando os resultados referentes ao período de 2018 a 2019. Por meio de análises de espectrometria de massas, são apresentados os resultados da interação entre os stem-bulge RNAs de *Drosophila melanogaster* com proteínas da cromatina, juntamente com a comparação da expressão dos genes de sbRNAs em indivíduos adultos, machos e fêmeas, de *D. melanogaster*.

RESUMO

Os RNAs não-codificantes de proteínas (ncRNAs) são descritos como participantes de diversos processos celulares. Dentro deste grupo de ncRNAs estão os Y RNAs, sendo que os detectados em mamíferos são descritos como essenciais para o processo de iniciação da replicação do DNA. Recentemente, foi mostrado que os Y RNAs descritos em *Xenopus laevis* e *Danio rerio* também são essenciais para a iniciação da replicação do DNA, mas somente após a segunda parte da fase de blástula, na embriogênese. O termo "Y RNA" se deve ao fato destes terem sido encontrados no citoplasma de células de mamíferos (cYtoplasmatic RNAs) e assim, serem diferenciados dos RNAs encontrados no núcleo (nUclear RNAs). Os Y RNAs foram primeiramente descritos em 1981 como componentes das partículas de ribonucleoproteína, RNP, complexadas com as proteínas Ro60 e La, detectados no soro de pacientes com Lúpus eritematoso sistêmico ou com síndrome de Sjögren.

Os Y RNAs são transcritos pela RNA polimerase III e apresentam uma estrutura secundária característica, na forma de haste-loop. Essa estrutura particular é formada devido a uma complementaridade parcial nas extremidades 5´ e 3´ dos Y RNAs, o que leva à formação de uma estrutura de dupla fita, ligada por um loop de fita simples. Os tamanhos dos Y RNAs podem variar de 70 a 115 nucleotídeos. Além disso, podem ser encontrados até quatro genes altamente conservados ao longo da evolução dos vertebrados. Em humanos, são expressos quatro genes: RNY1 (NR_004391), RNY3 (NR_004392), RNY4 (NR_004393) e RNY5 (NR_001571.2). Os quatro genes estão localizados no cromossomo 7, posição 7q36, não apresentam sequências intervenientes e possuem 112, 101, 93 e 83 nucleotídeos, respectivamente. Embora tenha sido descrito juntamente com os demais, o gene RNY2 foi removido da lista pois foi constatado que se tratava de um subproduto da degradação do hY1 RNA. É descrito que em roedores também foram detectados os quatro genes, entretanto ocorre expressão de somente dois destes genes, sendo eles Y1 e Y3. Em camundongos, somente os genes mY1 (NR_004419.1) de 111 nucleotídeos e mY3 (NR_024202.2) de 101 nucleotídeos são descritos e expressos.

Os Y RNAs já foram descritos em diversos organismos, desde procariotos até eucariotos. Em cultura de células de fibroblastos de pulmão de hamster chinês, da linhagem GMA32, foram detectados quatro genes de Y RNAs: chY1, chY3, chY4 e

chY5. Entretanto, apenas os chY1 e chY3 são expressos. Independentemente do fato de que os genes chY4 e chY5 não são expressos, os mesmos ainda são capazes de promover a iniciação da replicação do DNA do sistema *in vitro* de células humanas. Isso provavelmente se deve ao fato de que esses chY RNAs possuem em sua região denominada *upper stem*, os nucleotídeos que foram descritos como indispensáveis para a ativação da replicação do DNA *in vitro*.

Os padrões estruturais, de sequência e motivos, descritos para os Y RNAs, foram utilizados para descrever genes homólogos em invertebrados, que levou ao estabelecimento de uma outra família de ncRNA: os *Stem-bulge* RNAs (sbRNAs). Estes ncRNAs, compõem uma família de pequenos ncRNAs encontrados em invertebrados, primeiramente descritos em *Caenorhabditis elegans*. Tanto os Y RNAs quanto os sbRNAs estão fortemente relacionados à replicação do DNA cromossomal, atuando como fatores essenciais para a iniciação da replicação em humanos e outros vertebrados, além de serem essenciais para a viabilidade de embriões em nematoides. Estudos comprovaram que fragmentos de Y RNAs estão altamente presentes em células, tecidos e fluídos corporais de mamíferos, incluindo seres humanos, assim como em tumores.

Os Y RNAs de *C. elegans* (CeY RNA), de *Branchiostoma floridae* (BfY RNA) e de *Deinococcus radiodurans* (DrY RNA) não possuem sequencias relativas ao domínio mínimo da haste superior dos Y RNA de vertebrados, e desse modo não são capazes de apresentar a atividade essencial para o licenciamento da iniciação da replicação em sistemas *in vitro* de células humanas. Entretanto, em *C. elegans*, foram descritos os *stem-bulge* RNAs (sbRNAs), moléculas homólogas em estrutura e função aos Y RNAs de vertebrados.

Os sbRNAs de *C. elegans* são caracterizados por possuírem dois motivos distintos, localizados no final das extremidades 5' e 3'. Estes motivos conservados têm o potencial de formar uma cadeia dupla hélice apresentando a sequência conservada (UUAUC), separada por uma única cadeia de RNA em *loop* que pode ter diferentes tamanhos. Além desta sequência, constatou-se que as moléculas de sbRNA só atuam na iniciação da replicação se apresentarem um trecho altamente conservado, denominado domínio mínimo, composto pela sequência GUG – CAC. Este domínio mínimo está presente tanto nos Y RNAs quanto nos sbRNAs.

Até o ano de 2015 não haviam indícios tanto de Y RNAs quanto sbRNAs em insetos. A única exceção era a sugestão de um gene putativo de Y RNA, descrito *in silico*,

para *Anopheles gambiae*. Então, surgiu a descrição do primeiro gene de sbRNA, identificado e clonado a partir do genoma de *Bombyx mori*. Esse ncRNA é homólogo aos sbRNAs de *C. elegans* e foi denominado de *Bm*sb*RNA*.

Uma vez que não haviam registros de nenhum outro possível sbRNA em insetos, realizou-se a busca por estas moléculas em *Drosophila melanogaster*. Por meio de ferramentas de bioinformática e dinâmica molecular, dois candidatos a sbRNAs foram identificados no genoma deste inseto e suas estruturas avaliadas quanto à estabilidade. Estes foram denominados Dm1 e Dm2, respectivamente. Em seguida, foram construídos primers para experimentos de expressão relativa, mostrando que os sbRNAs de *D. melanogaster* são expressos tanto em linhagens celulares quanto em moscas adultas. Além disso, foi constatado que estes genes são mais expressos em moscas adultas do sexo masculino. Experimentos funcionais foram realizados em um sistema de núcleos isolados, a fim de verificar se os genes candidatos eram capazes de substituir Y RNAs endógenos, na função de iniciação da replicação do DNA. Foi constatado que apenas Dm1 era funcional, tornando este o primeiro registro funcional de um sbRNA de insetos. Adicionalmente, análises de espectrometria de massas indicam que Dm1 está associado com proteínas relacionadas com a replicação do DNA, o que corrobora com dados obtidos no teste funcional, além da detecção de peptídeos relacionados à resposta imune. Estes peptídeos são as cecropinas A1, que também estão associadas com a infecção do *Bm*NPV ao bicho-da-seda *Bombyx mori*. Portanto, acredita-se que estas cecropinas também devam interagir com o *Bm*sbRNA, hipótese que será investigada em experimentos futuros. Por fim, os resultados para Dm2 indicam que esta molécula está relacionada com proteínas do citoesqueleto, em outras palavras, proteínas de caráter estrutural e por isso não foi detectada sua atividade de licenciamento da iniciação da replicação.

Palavras-chave: *Stem-bulge* RNAs, *non-coding* RNAs, *Drosophila melanogaster*, Dm1, Dm2, expressão relativa, replicação do DNA.

ABSTRACT

Non-coding RNAs (ncRNAs) are described as participants in several cellular processes. Within this group, there are Y RNAs, detected in mammals and described as essentials for the initiation of the DNA replication process. Recently, it was shown that the Y RNAs described in *Xenopus laevis* and *Danio rerio* are essential for the initiation of DNA replication, but only after the second part of the blastula phase, in embryogenesis. The term “Y RNA” is because they were found in the cytoplasm of mammalian cells (cYtoplasmatic RNAs) and thus, differentiated from the RNAs found in the nucleus (nUclear RNAs). Y RNAs were first described in 1981, as components of the ribonucleoprotein particles (RNPs), complexed with Ro60 and La proteins, detected in the serum of patients with systemic lupus erythematosus or Sjögren's syndrome.

Y RNAs are transcribed by the RNA polymerase III and have a characteristic secondary structure in a stem-loop form. This particular structure is formed due to a partial complementarity at the 5 'and 3' ends of the Y RNAs, which leads to the formation of a double-stranded structure, connected by a single strand loop. The sizes of Y RNAs can vary from 70 to 115 nucleotides. Also, up to four highly conserved genes can be found throughout the evolution of vertebrates. In humans, four genes are expressed: RNY1 (NR_004391), RNY3 (NR_004392), RNY4 (NR_004393) and RNY5 (NR_001571.2). The four genes are located on chromosome 7, position 7q36, have no intervening sequences and have 112, 101, 93 and 83 nucleotides, respectively. Although it was described together with the others, the RNY2 gene was removed from the list because it was found to be a by-product of hY1 RNA degradation. It is described that the four genes were also detected in rodents, however, only two of these genes are expressed, Y1 and Y3. In mice, only the 111 nucleotide mY1 (NR_004419.1) and 101 nucleotide mY3 (NR_024202.2) genes are described and expressed.

Y RNAs have been described in several organisms, from prokaryotes to eukaryotes. In GMA32 lineage, a culture of cells from Chinese hamsters, four Y RNA genes were detected: chY1, chY3, chY4, and chY5. However, only chY1 and chY3 are expressed. Regardless of the fact that the chY4 and chY5 genes are not expressed, they are still capable of promoting the initiation of DNA replication in the human cell system in vitro. This is probably due to the fact that these chY RNAs have in their region called

the upper stem, the nucleotides that have been described as indispensable for the activation of DNA replication *in vitro*.

The structural patterns, sequence, and motifs, previously described for Y RNAs, were used to search for homologous genes in invertebrates, which led to the establishment of another family of ncRNA: The stem-bulge RNAs (sbRNAs). These ncRNAs consist of a family of small ncRNAs found in invertebrates, first described in *Caenorhabditis elegans*. Both Y RNAs and sbRNAs are strongly related to chromosomal DNA replication, acting as essential factors for the initiation of replication in humans and other vertebrates. They also are described as being essential for the viability of embryos in nematodes. Studies have shown that fragments of Y RNAs are highly present in cells, tissues and body fluids of mammals, including humans, as well as tumors.

The Y RNAs of *C. elegans* (CeY RNA), *Branchiostoma floridae* (BfY RNA) and *Deinococcus radiodurans* (DrY RNA) do not have sequences related to the minimal domain of the upper stem of the vertebrate Y RNA and thus are not capable to present the essential activity for licensing the initiation of replication in an *in vitro* human cell system. However, in *C. elegans*, stem-bulge RNAs (sbRNAs), molecules homologous in structure and function to vertebrate Y RNAs, have been described.

The sbRNAs from *C. elegans* are characterized by having two distinct motifs, located at the end of the 5 'and 3' ends. These conserved motifs have the potential to form a double helix strand presenting the conserved sequence (UUAUC), separated by a single looped RNA strand that can be variable in size. In addition to this sequence, it was found that sbRNA molecules only act in the initiation of DNA replication if they present a highly conserved stretch, called the minimal domain, composed of the double-stranded sequence GUG - CAC. This minimal domain is present in both Y RNAs and sbRNAs.

Until 2015, there was no evidence of either Y RNAs or sbRNAs in insects. The only exception was the suggestion of a putative Y RNA gene, described *in silico*, for *Anopheles gambiae*. Then, the first sbRNA gene was identified and cloned from the *Bombyx mori* genome. This ncRNA was shown to be homologous to the *C. elegans* sbRNAs and was called *Bm*sbRNA.

Since there were no records of any other possible sbRNA in insects, the search for these molecules was carried out in *Drosophila melanogaster*. Using bioinformatics and molecular dynamic tools, two candidates for sbRNAs were identified in the genome of this insect and its structures were evaluated for stability. These were called Dm1 and

Dm2, respectively. Then, primers were constructed for relative expression experiments, showing that *D. melanogaster* sbRNAs are expressed in both cell lines and adult flies. In addition, it was found that these genes are more expressed in adult male flies. Functional experiments were carried out in a system of isolated nuclei, in order to verify whether the candidate genes were capable of replacing endogenous Y RNAs, in the function of initiating DNA replication. It was found that only Dm1 was functional, making this the first functional record of an insect sbRNA. Also, mass spectrometry analyzes indicate that Dm1 is associated with proteins related to DNA replication, which corroborates with data obtained in the functional test. Moreover, it was detected peptides related to the immune response. These peptides are cecropins A1, which are also associated with the infection of *Bm*NPV to the silkworm *Bombyx mori*. Therefore, it is believed that these cecropins should also interact with *Bm*sbRNA, a hypothesis that will be investigated in future experiments. Finally, the results for Dm2 indicate that this molecule is related to proteins of the cytoskeleton, in other words, proteins with structural characteristics. This result might explain why its licensing activity in the initiation of replication was not detected.

Keywords: Stem-bulge RNAs, non-coding RNAs, *Drosophila melanogaster*, Dm1, Dm2, relative expression, DNA replication.

Artigo 1: Identification and characterization of stem-bulge RNAs in *Drosophila melanogaster*.

(Publicado na revista *RNA Biology* - ISSN: 1555-8584/ *Impact Factor* JCR: 5,28)

Identification and characterization of stem-bulge RNAs in Drosophila melanogaster

Francisco Ferreira Duarte Junior[a], Paulo Sérgio Alves Bueno[b], Sofia L. Pedersen[c], Fabiana dos Santos Rando[d], José Renato Pattaro Júnior[b], Daniel Caligari[a], Anelise Cardoso Ramos[a], Lorena Gomes Polizelli[a], Ailson Francisco dos Santos Lima[a], Quirino Alves de Lima Neto[a], Torsten Krude[c], Flavio Augusto Vicente Seixas[b], and Maria Aparecida Fernandez[a]*

[a] Departamento de Biotecnologia, Genética e Biologia Celular, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil; [b] Departamento de Tecnologia, Universidade Estadual de Maringá, campus Umuarama, Av. Ângelo Moreira da Fonseca, 1800, Umuarama, 87506-370, Paraná, Brazil; [c] Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, United Kingdom. [d] Center for Molecular, Structural and Functional Biology - CBM/COMCAP, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil.

***Corresponding author:** Maria Aparecida Fernandez; mafernandez@uem.br

Identification and characterization of stem-bulge RNAs in *Drosophila melanogaster*

Abstract

Non-coding Y RNAs and stem-bulge RNAs are homologous small RNAs in vertebrates and nematodes, respectively. They share a conserved function in the replication of chromosomal DNA in these two groups of organisms. However, functional homologues have not been found in insects, despite their common early evolutionary history. Here, we describe the identification and functional characterization of two sbRNAs in *Drosophila melanogaster*, termed Dm1 and Dm2. The genes coding for these two RNAs were identified by a computational search in the genome of *D. melanogaster* for conserved sequence motifs present in nematode sbRNAs. The predicted secondary structures of Dm1 and Dm2 partially resemble nematode sbRNAs and show stability in molecular dynamics simulations. Both RNAs are phylogenetically closer related to nematode sbRNAs than to vertebrate Y RNAs. Dm1, but not Dm2 sbRNA is abundantly expressed in *D. melanogaster* S2 cells and adult flies. Only Dm1, but not Dm2 can functionally replace Y RNAs in a human cell-free DNA replication initiation system. Therefore, Dm1 is the first functional sbRNA described in insects, allowing future investigations into the physiological roles of sbRNAs in the genetically tractable model organism *D. melanogaster*.

Keywords: DNA replication; sbRNA; non-coding RNAs; Y RNA.

Subject classification codes: Bioinformatics; Cell Biology, RNomics; Small and large Non-coding RNAs; Trancriptome.

Introduction

Non-coding RNAs (ncRNAs) regulate many fundamental pathways in eukaryotic organisms. Two families of ncRNAs play essential functional roles during chromosomal DNA replication: Y RNAs in vertebrates and stem-bulge RNAs (sbRNAs) in nematodes [1, 2]. Y RNAs have been shown to be essential for the initiation of chromosomal DNA replication in a cell-free system, for DNA replication and cell proliferation in cultured vertebrate cells and for early development and viability of *Xenopus laevis* and the zebrafish, *Danio rerio* [3, 4, 5]. sbRNAs are able to functionally replace endogenous Y RNAs in a human DNA replication initiation system. They are also essential for DNA replication and cell proliferation in the nematode *Caenorhabditis elegans*, as well as its development and viability [6].

Both Y and sbRNAs show homology in function and structure [1, 2, 7, 8]. Structurally, they comprise short stem-loop RNAs of around a hundred nucleotides in length. The partially complementary 5' and 3' ends hybridize to form a double-stranded stem structure with a central single-stranded loop. Common to both Y and sbRNAs are conserved nucleotide sequence elements in the stem structure. They comprise a short helix of 7-10 base pairs flanked by G-C base pairs either end, and a highly conserved double-stranded GUG-CAC tri-nucleotide motif near the center of this domain. This domain is located in the upper stem of both ncRNA families, which opens up into the central loop domain. Importantly, these conserved motifs are essential for the function of these RNAs during the initiation of chromosomal DNA replication in a cell-free system because mutations in these elements abrogate their function [3, 6, 9, 10].

Despite the similarities, there are a few differences between Y and sbRNAs. On the one hand, vertebrate Y RNAs contain a second helical motif with a bulged C residue towards the terminus of the stem-loop, which bind orthologues of the Ro60 protein. In

fact, Y RNAs were originally described as the RNA component of Ro-ribonucleotide particles (Ro-RNPs), based on their association with Ro60 [11]. However, neither the Ro-binding domain nor the Ro60 protein are essential for Y RNA function during DNA replication [3, 9, 12], and they are not found in sbRNAs [1, 2, 6]. On the other hand, sbRNAs contain a highly conserved UUAUC penta-nucleotide motif at the 5' end of the central single-stranded domain, which appears to play an additional stimulatory role for DNA replication [1, 2, 6].

In evolutionary terms, functional Y RNAs are found in all vertebrates investigated so far, and sbRNAs in several nematodes including *C. elegans* and *Caenorhabditis briggsae* [1, 6, 13, 14]. Yet there is a clear lack of information about the presence of these RNAs in the large group of insects, which diverged from a common ancestor after the nematodes. Computational searches have not provided evidence for a wide-scale conservation of either RNA family in insects, including the major model organism *Drosophila melanogaster* [8]. However, there are isolated descriptions of a candidate Y RNA in *Anopheles gambiae* [8] and an sbRNA in the silkworm *Bombyx mori* [15], suggesting that Y or sbRNAs might be conserved to some extent in insects. It is unknown if these isolated examples play a functional role in DNA replication.

In this study, we conducted a computational search for sbRNA genes in *D. melanogaster* based on homology motifs present in nematode sbRNAs. We present the first two candidate sbRNA genes in *D. melanogaster*, coding for two sbRNAs named Dm1 and Dm2. Both their predicted secondary structures resemble sbRNAs and show stability in molecular dynamics simulations. Dm1 is more abundantly expressed than DM2 in *D. melanogaster* S2 cells and in adult flies. Only Dm1 can functionally replace Y RNAs in a human cell-free DNA replication initiation system. Therefore, Dm1 is the first functional sbRNA described in insects, allowing future investigations into the

physiological roles of sbRNAs in the genetically tractable model organism *D. melanogaster*.

Results

Identification of sbRNA candidate genes in Drosophila melanogaster

To identify candidate sbRNA genes in the genome of fruit fly *Drosophila melanogaster*, we searched for sequences with predicted transcription start and stop sites containing conserved nucleotide motifs present in nematode sbRNAs (see Materials and Methods). The search resulted in two candidate sbRNA genes, named Dm1 and Dm2, which transcribe into RNA molecules of 85 and 89 nucleotides, respectively (Figure S1 and S2).

We then assessed the predicted secondary structures of these two candidate sbRNAs using the mFold algorithm (Fig. 1). Dm1 provided a single structure, while Dm2 provided seven alternative structures. The difference of free energy between the predicted structures of Dm2 was approximately 0.1 Kcal, suggesting that this molecule alternates between the seven related conformations. The most stable predicted secondary structures of Dm1 and Dm2 show an overall stem-bulge shape (Fig. 1) that is highly similar to vertebrate Y RNAs and nematode sbRNAs [2]. Both Dm1 and Dm2 contain a double-stranded stem, subdivided into small domains by internal loops and bulges, and a central single-stranded loop (Fig. 1). They lack a terminal helix with a bulged C forming a binding site for Ro60 protein, which is typically found in vertebrate Y RNAs. Both candidates also contain a conserved double-stranded GUG-CAC motif present in both nematode sbRNAs and vertebrate Y RNAs, which is essential for their function in chromosomal replication [2]. Both Dm1 and Dm2 contain a UUUAC penta-nucleotide

motif downstream of the GUG motif, similar to the UUAUC motif found in nematode sbRNAs at this position. However, unlike nematode sbRNAs, Dm1 has a partially base-paired UUUAC penta-nucleotide motif and a non-conserved bulged G next to the trinucleotide motif (Fig. 1A). Dm2 has an unusually short double-stranded stem around the trinucleotide motif and lacks flanking G-C base pairs (Fig. 1B), which are conserved in both vertebrate Y RNAs and nematode sbRNAs.

We conclude that we have identified two candidate sbRNA genes in the genome of *D. melanogaster*, which share several key elements with vertebrate Y RNAs and nematode sbRNAs.

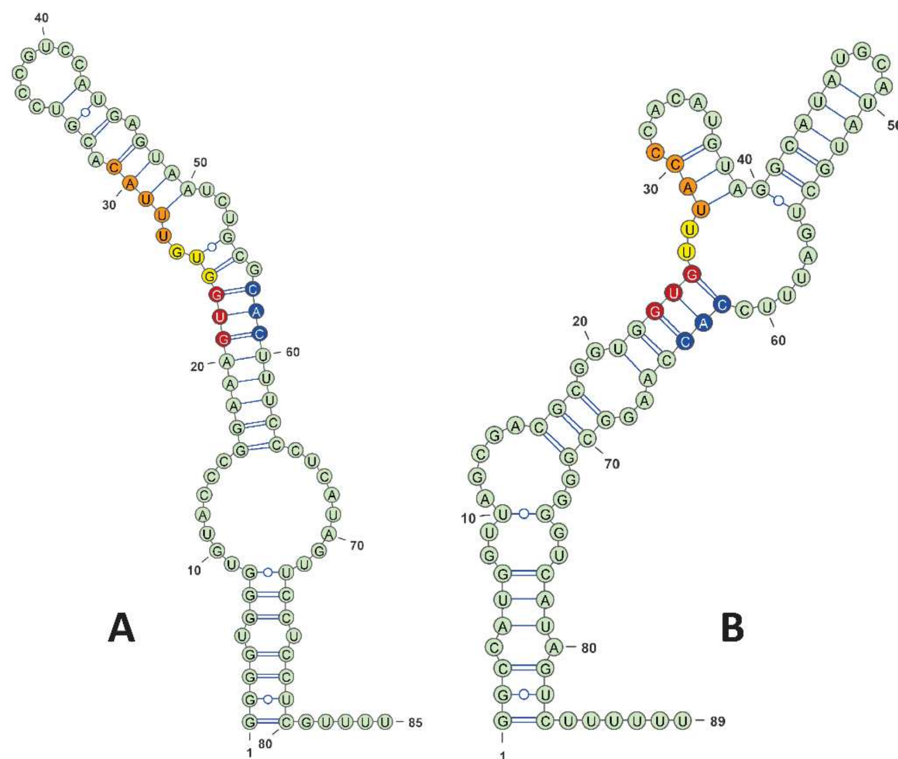

Figure 1. Predicted secondary structures of two putative *Drosophila melanogaster* sbRNAs. An evolutionarily conserved trinucleotide domain present stem-bulge and Y RNAs that is functionally essential for DNA replication activity is highlighted in red/blue, a variable region in yellow, and a second conserved sequence present in sbRNAs in orange. Predicted secondary structure for Dm1(A) and Dm2 (B).

Molecular Dynamics

Computational simulations of the molecular dynamics of small sbRNAs and Y RNAs allow predictions about 3D structure stability and rigidity (Fig. S3-S8). We therefore performed these computational simulations on the *D. melanogaster* candidate sbRNAs Dm1 and Dm2 in comparison to small reference vertebrate Y RNAs (Fig. 2, 3, S1 and S2). The 3D structures of Dm1 and Dm2 reached thermodynamic equilibrium after 15 ns and 30 ns, respectively (Figure 2). However, both Dm1 and Dm2 were much more stable than the reference Y RNAs. The stability is evaluated as a function of the oscillation. The smaller the oscillation, the greater the stability. The oscillation of Dm1 and Dm2 are the ones closest to a straight line after 30ns of simulation (Fig 2b). The radius of gyration of all RNAs was constant over time for all RNAs analyzed, indicating that these structures maintained their folded configuration over the course of the simulation (Fig 2a). Together, these results suggest that the proposed three-dimensional structures of both Dm1 and Dm2 are stable.

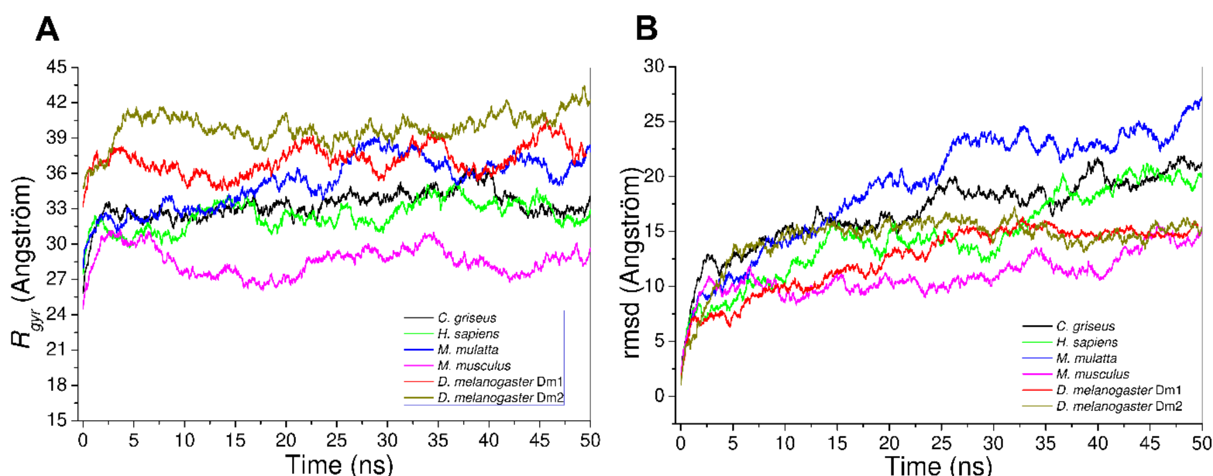

Figure 2. Parameters of computational molecular dynamics simulations. Root mean square deviation (A) and Radius of gyration (B) obtained from the MD simulation from different Y RNAs and sbRNAs. 3D representations of *D. melanogaster* Dm1 and Dm2 RNAs, Y3 RNA from *C. griseus*, *H. sapiens* and *M. musculus* and Y5 RNA from *M. mulatta* were used for this analysis (see Fig. S3-S8, corresponding .pdb files).

We next performed an analysis of individual nucleotide dynamics of Dm1 and Dm2, focusing on the conserved GUG-CAC trinucleotide motifs. The analysis of the root mean square fluctuation (rmsf) for C_1’ atoms of each nucleotide suggests that the stem region containing the conserved GUG-CAC motif is a region of high mobility in Dm1 (Fig 3A). This region is much more stable in Dm2 (Fig. 3B), whose behavior was similar to the reference Y RNAs (Fig. S3). We expected the pairing of this functional triplet in Dm1 to have up to eight hydrogen bonds. However, the average number of hydrogen bonds in this region was 2.4 ± 1.41 during the simulation, while Dm2 had 1.7 ± 1.22, less than half of that expected (Fig. S4). Although Dm1 presents a higher average number of H-bonds than Dm2, their number of H-bonds equalized at the end of the simulation (equilibrium region). The reduced numbers of hydrogen bonds suggest that this region has high mobility, in contrast to known sbRNAs [15].

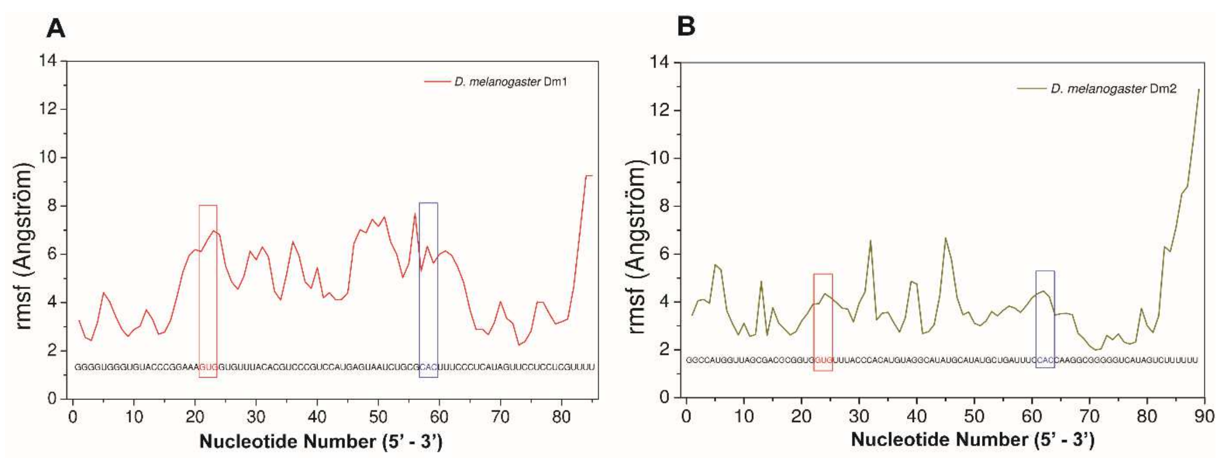

Figure 3. Parameters of computational molecular dynamics simulations. Root mean square fluctuation (rmsf) of C1' from *D. melanogaster* Dm1 (A) and Dm2 (B) RNAs. Red boxes highlight the rmsf of the GUG triplet and blue boxes of the CAC triplet.

The GUG-CAC motif of Dm1 has a greater flexibility compared to other Y RNAs. The most likely explanations for this behavior are: (1) the loop proximity may have promoted instability in the region; or (2) the long period of simulation (50 ns) may have expanded the imperfections in the force field parameters throughout the simulation.

Previous analyses of the conserved upper stem domain of human Y1 RNA by circular dichroism and solution state NMR indicated that this domain assumes an overall A-form RNA helix, but central bases including the GUG-CAC motif are destabilized and may actually dynamically flip out of the helix [10]. Our molecular dynamics simulations suggest that Dm1, but not Dm2, might show a similar feature of localized high flexibility. Importantly, mutations within the GUG-CAC motif in Y1 RNA led to both structural disturbances and loss of function during the initiation step of chromosomal DNA replication in human cell nuclei [3, 9, 10]. It is therefore possible that Dm1 RNA might be a functional homologue of vertebrate Y RNAs as it contains this essential motif and associated structural features.

In the next set of experiments, we turned to phylogenetic and functional analyses of the Dm1 and Dm2 candidate sbRNAs in relation to vertebrate Y and nematode sbRNAs.

Phylogenetic analysis

Vertebrate Y RNAs and nematode sbRNAs share structural and functional homologies [1, 2]. Therefore, we performed a phylogenetic analysis of Dm1 and Dm2 in comparison to Y and sbRNAs (Fig. 4). We included a group of 5S rRNA sequences from insect,

nematode and vertebrate organisms into the analysis and calculated the phylogenetic tree with these 5S rRNAs as the root. As expected, the 5S rRNAs form a separate clade from the Y and sbRNAs (Fig. 4). Vertebrate Y RNAs separated into the four distinct major clades of Y1, Y3, Y4 and Y5, and a heterogenous outgroup of teleost and some amphibian Y RNAs, consistent with earlier reports [7, 8]. Nematode sbRNAs form a distinct clade from the vertebrate Y RNAs (Fig. 4). This clade is more closely related to Y5 (and Chinese hamster Y4 as an outlier), than to the three other Y RNA clades. Interestingly, Dm1 RNA forms an outgroup of this nematode sbRNA clade together with an sbRNA candidate of the silkworm *Bombyx mori* described earlier [15] and *C. elegans* CeN72 sbRNA. Dm2 is an isolated outlier between sbRNAs and 5S rRNA. Taken together, these data strongly suggest that these small insect candidate sbRNAs are more closely related to nematode sbRNAs than to vertebrate Y RNAs.

Figure 4. Global Phylogenetic Tree of Y RNAs and sbRNAs. 5S Ribosomal RNA family (Black). sbRNA family (Blue). Y5 RNA family (Red). Y4 RNA family (Purple). Y3 RNA family (Orange). Y1 RNA family (Green). The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [30]. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed [31]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the

Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The analysis involved 77 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 18 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [23].

These phylogenetic data suggest that insect sbRNAs and nematode sbRNAs may have evolved in their respective lineages from a common ancestor and may therefore share some common functionality. From these data, however, it is unclear whether vertebrate Y RNAs, nematode or insect sbRNAs have also evolved from a common ancestor, or whether they are the product of a convergent evolution into a partially shared functionality [2]. To address this issue, phylogenetic work would need to be conducted with a substantially larger number of functionally and structurally related sbRNA homologues from other insects and arthropods, and from crustacean, mollusk and even more distantly related groups of organisms. As these data are currently unavailable, future work aimed at identifying such additional distant homologues would be an important goal in the field.

Expression of Dm1 and Dm2 sbRNAs

Y RNAs and sbRNAs are abundantly expressed in vertebrates and nematodes, respectively, at levels approximately ten times below those of ribosomal 5S RNAs [3]. However, individual expression levels vary substantially between different types of Y and sbRNAs within a cell type or tissue, and between different cell types and different tissues of the same organism [2, 3, 6, 16, 17]. Next, we determined the relative expression levels of Dm1 and Dm2 sbRNAs in *D. melanogaster* S2 cells and adult flies (Fig. 5). In S2 cells, the expression of the Dm1 and Dm2 genes was approximately 250-fold and 1,000-fold lower than the 5S gene, respectively. In adult flies, Dm1 and Dm2 sbRNA

expression was 66-fold and 900-fold lower than the 5S gene, respectively.

We conclude that the relative expression level of Dm1 sbRNA in *D. melanogaster* is similar to sbRNAs in nematodes and Y RNAs in vertebrates, whereas Dm2 sbRNA is expressed at one to two orders of magnitude below that, if at all.

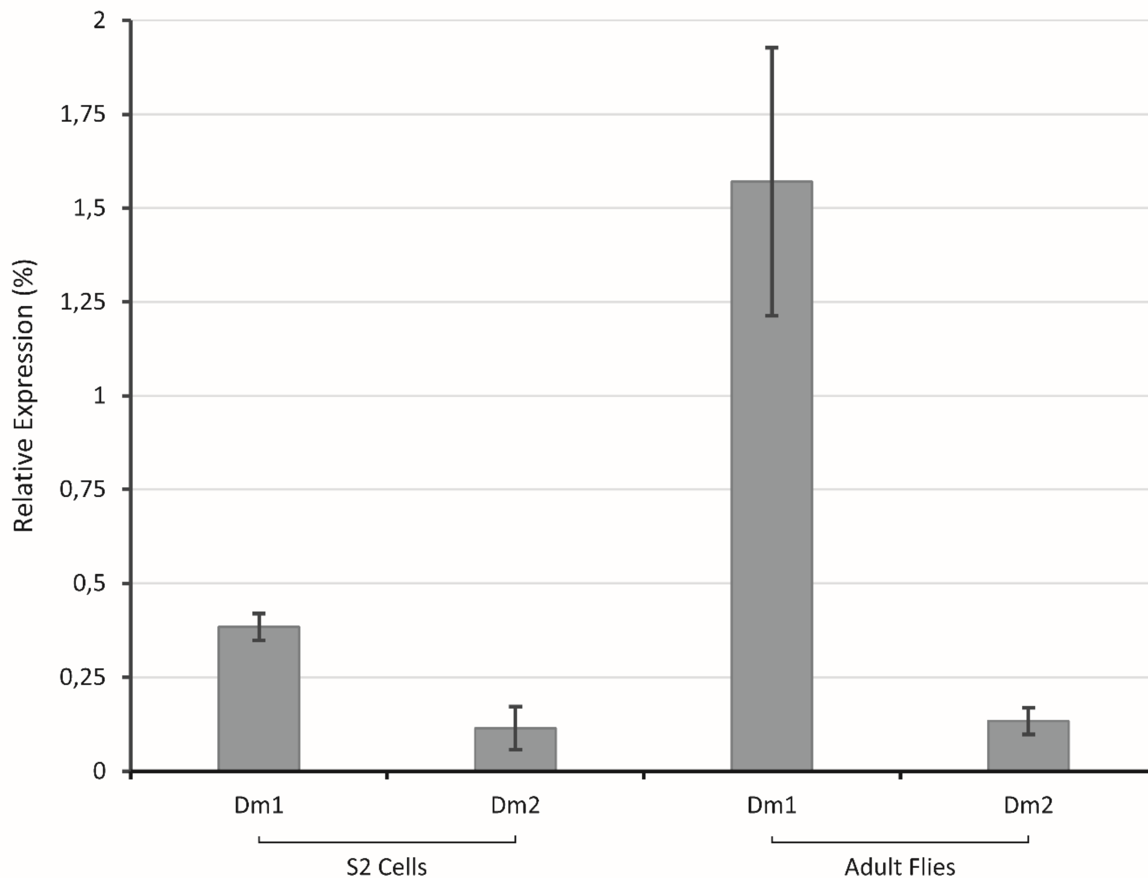

Figure 5. Expression of Dm1 and Dm2 sbRNAs, relative to 5S rRNA. Total RNA was prepared from *Drosophila melanogaster* S2 cells and adult flies, and quantified by qRT-PCR. Percentages of relative expression are shown as mean values ± standard deviations of n = 4 independent experiments.

Functional homologies to human Y RNA

Vertebrate Y RNAs and nematode sbRNAs have essential functions in chromosomal DNA replication [2]. Depletion of Y RNAs and sbRNAs *in vivo* by RNA interference or

antisense morpholino oligonucleotides leads to an inhibition of DNA replication and cell proliferation, and to death during early development [3, 4, 6]. In a cell-free system derived from human cells, Y RNAs are required specifically in late G1 phase nuclei for the initiation of chromosomal DNA replication, but not for subsequent elongation synthesis [3, 5]. Importantly, non-human vertebrate Y RNAs and several nematode sbRNAs can functionally substitute for human Y RNAs in this system [6, 9], indicating that they are functionally redundant and homologous. To address whether *Drosophila* Dm1 and Dm2 sbRNAs were functionally homologous to these RNAs, we tested whether they could functionally substitute for vertebrate Y RNAs in this system (Fig. 6).

We synthesised Dm1 and Dm2 sbRNAs *in vitro* from synthetic templates and purified the full-length RNAs (Fig. 6A). We then tested these RNAs in the human cell-free DNA replication initiation system (Fig. 6B). Incubation of late G1 phase template nuclei in a control extract from proliferating HeLa cells resulted in a significant initiation of chromosomal DNA replication in about 35% of these nuclei. This is above the background of about 5% of contaminating S phase nuclei replicating in the absence of extract, as described previously [3, 18]. Degradation of the endogenous human Y RNAs resulted in an inhibition of replication initiation, reducing the percentages of replicating nuclei to near-background values (Fig. 6B), as described previously [3]. Significantly, addition of Dm1 restored the initiation function of the Y RNA-depleted extract, whereas addition of Dm2 did not (Fig. 6B). We therefore conclude that Dm1 sbRNA, but not Dm2 sbRNA, is a functional homologue of human Y RNAs in this DNA replication initiation system.

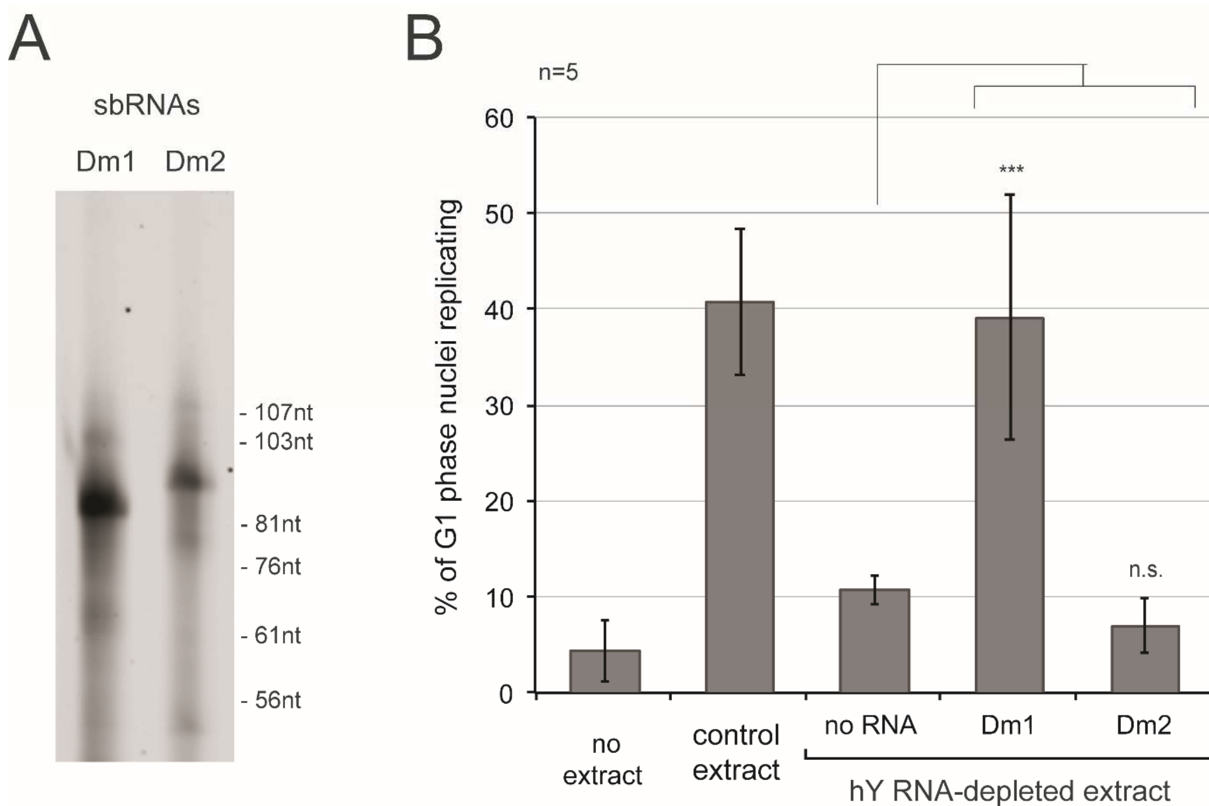

Figure 6. DNA replication initiation function. (A) Synthesis of Dm1 and Dm2 sbRNAs. The indicated RNAs were synthesised by *in vitro* transcription and visualised after denaturing polyacrylamide gel electrophoresis. Molecular sizes of single-stranded DNA oligonucleotide markers are indicated. (B) RNA-dependent initiation of chromosomal DNA replication in a human cell-free system. Human late G1 phase template nuclei were incubated in the absence or presence of control or Y RNA-depleted human cytosolic extracts supplemented with 100 ng of purified Dm1 or Dm2 sbRNAs, as indicated. Percentages of replicating template nuclei are shown for each reaction as mean values ± standard deviations of n = 5 independent experiments. Brackets indicate results of t-tests (unpaired, two-tailed with unequal variance) of control against the experimental samples (** p = 0.007; n.s., not significant).

Discussion

In this study, we identified by a computational search for conserved sequence motifs present in nematode sbRNAs two short candidate RNAs in the genome of the fruit fly *Drosophila melanogaster*, which we termed Dm1 and Dm2. Both RNAs have predicted secondary structures partially resembling nematode sbRNAs and both show stability in

molecular dynamics simulations. These candidate RNAs are phylogenetically closer to nematode sbRNAs than to the homologous vertebrate Y RNAs. However, we found that only Dm1, but not Dm2, is abundantly expressed and can functionally replace Y RNAs in a human cell-free DNA replication initiation system.

Taken together, our results support the conclusion that Dm1 is a genuine stem-bulge RNA expressed in *D. melanogaster*, which is functionally related to nematode sbRNAs and vertebrate Y RNAs. A different short sbRNA candidate gene has previously been described in another insect, the silkworm *Bombyx mori* [15]. However, it is yet unknown whether this BmsbRNA has an activity during DNA replication and would thus be functionally homologous to Dm1 sbRNA, vertebrate Y RNAs, or other nematode sbRNAs. Therefore, Dm1 sbRNA would be the first functional sbRNA described in insects. This discovery should stimulate future work into identifying and characterizing additional functional sbRNA or Y RNA homologues in insects, and to study their physiological roles during development and tissue maintenance in whole organisms.

In contrast, the Dm2 sbRNA gene is only very weakly expressed in *D. melanogaster* (if at all). Dm2 shares fewer conserved elements with vertebrate Y RNAs and nematode sbRNAs than Dm1. In the evolutionary tree, Dm2 remains isolated between other sbRNAs and the 5S rRNA outgroup. This correlates with the observation that Dm2 did not show any significant activity in initiating DNA replication *in vitro*. These observations suggest that Dm2 may not be a genuine sbRNA or Y RNA homologue and may have a different physiological role. Interestingly, the overall structure of Dm2 is not inconsistent with the possibility that it could be a miRNA precursor [6, 19, 20]. We will address this possibility separately in a future communication.

In conclusion, we have shown that small non-coding stem-loop RNAs with conserved structural elements and a homologous function in DNA replication are not only

present in nematodes and vertebrates, but also in insects. This finding should stimulate further work into the evolution of these RNAs and into elucidating molecular mechanisms underlying their physiological function using the powerful genetic tools available in *D. melanogaster*.

Materials and Methods

Search for sbRNA gene sequences

The search for sequences similar to sbRNAs was carried out in the genome of *Drosophila melanogaster* (genome id: 47), using BLASTn [21], starting from fragments of conserved sequences for sbRNAs containing potential start and stop points of transcription. Sequences found were submitted to mFold web server [22], using default parameters, in order to obtain their secondary structure. This secondary structure was compared with the standard model for sbRNAs, described for *Caernohabditis elegans* [1].

Phylogenetic analysis

The phylogenetic tree was constructed using sequences for Y RNAs [8] and sbRNAs related to DNA replication [1, 6, 7]. This analysis was carried out with MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets [23], using a bootstrap-neighborhood analysis with 10000 steps and 5S genes as outgroup.

Molecular Dynamics

RNA secondary structures were constructed with Varna Applet [24] and used for graphical representation. The 3D representations were generated by RNA Composer server [25], and used for molecular dynamics simulations (MD). These 3D

representations are supplied as .pdb files (supplementary figures S3-S8, Supplementary Material online).

The MD simulations were carried out and analyzed in the programs NAMD2 [26] and VMD [27] using Charmm C35b2/C36a2 force field [28]. For this procedure, the spatial coordinates of the RNAs were virtually immersed in a periodic box containing TIP3 water and sufficient amount of sodium counter ions to neutralize the system charges. The box size was at least 15 Å away from the outer surface of the RNA. The simulations were carried out in steps. In the first one, all atoms of the system were minimized by 20,000 steps of Conjugated Gradient (CG). In the second step, water and ions were equilibrated by 60 ps. In the third step, all atoms of the system were minimized again by another 20,000 steps of CG. In the fourth and final step, all atoms of the system were equilibrated during 50 ns using 1 atm pressure, 300 K temperature, and a constant number of atoms (*NPT* ensemble). The other simulation parameters were adjusted according to the protocol established by [29]. The simulations took place in 20 nodes of an Intel Xeon E5-2695v2 Ivy Bridge, 2.4 GHZ (480 cores) processors of the St Dumont cluster at LNCC, Brazil.

Relative quantification of Drosophila RNAs by qRT-PCR

Total RNA was extracted from *D. melanogaster* cell cultures and adult wild-type flies. *D. melanogaster* S2 cells (provided by Dr. Alain Debec, Institut Jacques Monod, Paris, France), derived from the embryo of *Drosophila melanogaster*, were maintained in Shields and Sang M3 insect culture media (Sigma-Aldrich), supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco) and antibiotics at a temperature of 26°C. For RNA extraction, the media was discarded, followed by extraction with TRIzol LS (Invitrogen). *D. melanogaster* flies were acquired from the State University of

Maringá, Maringá - Paraná, Brazil. For whole organism RNA extraction, adult flies were macerated in liquid nitrogen, followed by extraction with TRIzol LS (Invitrogen).

Total RNA sample concentrations were quantified by nanodrop spectroscopy using NanoDrop2000 (Thermo-Fisher) and standardized for a total of 2000 ng of RNA for the treatment with DNAse I (Biolabs). The reverse transcriptase kit used to synthesize the first cDNA was iScript (Bio-Rad). The qPCR reactions were carried out by the following kit: Power SYBR Green PCR Master Mix (Thermo Fisher), with 40 cycles and annealing temperature of 58°C, in the LightCycler® 96 System (Roche) equipment. The primer sequences used were:

- Dm1: 5' GGGGTGGGTGTACCCGGAAA-3' and 5'AAAACGAGGAGGAACTATGAGGG-3'
- Dm2: GGCCATGGTTAGCGACGCG and 5'AAAAAAGACTATGACCCCCGCC-3'
- 5S rRNA: 5' GCCAACGACCATACCACGC-3' and 5'AAAAAGTTGTGGACGAGGCCAA-3'

The relative amount of expression (Ar) was calculated (Equation 1). Each sample was carried out in duplicates and calculations were based on 4 independent experiments.

$$Ar = 2^{(-\Delta CT)} = 2^{-(Ct\ of\ the\ sample\ -\ Ct\ of\ 5S\ gene)} \quad (1)$$

RNA synthesis in vitro

Dm1 and Dm2 sbRNAs were synthesized in vitro by bacteriophage SP6 RNA polymerase from two annealed complementary synthetic DNA oligonucleotides (Sigma-Aldrich Co.

LLC), as described previously [10]. Sequences of the four DNA oligonucleotides with added SP6 promoter sequences (underlined) are as follows:

- Dm1 forward 5'ATTTAGGTGACACTATAGGGGGTGGGTGTACCCGGAAAGTGGTGTT TACACGTCCCGTCCATGAGTAATCTGCGCACTTTCCCTCATAGTTCCT CCTCGTTTT-3'
- Dm1 reverse 5'AAAACGAGGAGGAACTATGAGGGAAAGTGCGCAGATTACTCATGG ACGGGACGTGTAAACACCACTTTCCGGGTACACCCACCCCCTATAG TGTCACCTAAAT-3'
- Dm2 forward 5'ATTTAGGTGACACTATAGGGCCATGGTTAGCGACGCGGTGGTGTTT ACCCACATGTAGGCATATGCATATGCTGATTTCCACCAAGGCGGGGG TCATAGTCTTTTTT-3'
- Dm2 reverse 5'AAAAAAGACTATGACCCCCGCCTTGGTGGAAATCAGCATATGCAT ATGCCTACATGTGGGTAAACACCACCGCGTCGCTAACCATGGCCCCT ATAGTGTCACCTAAAT-3'

Transcription produced the following synthetic RNAs:

- Dm1 sbRNA 5'GGGGUGGGUGUACCCGGAAAGUGGUGUUUACACGUCCCGUCCAU GAGUAAUCUGCGCACUUUCCCUCAUAGUUCCUCCUCGUUUU-3'
- Dm2 sbRNA 5'GGCCAUGGUUAGCGACGCGGUGGUGUUUACCCACAUGUAGGCAU

AUGCAUAUGCUGAUUUCCACCAAGGCGGGGGUCAUAGUCUUUUUU-3’

Following in vitro synthesis, template DNA was degraded by DNAse I treatment and the newly transcribed RNA was purified by phenol/chloroform extraction and ethanol precipitation. The RNA was then visualised by electrophoresis on denaturing polyacrylamide gels containing 8M urea, as described previously [3, 4].

DNA replication reactions

Human HeLaS3 and EJ30 cells were grown as proliferating monolayers, and template nuclei were prepared from mimosine-arrested late G1 phase EJ30 cells as described previously [18].

DNA replication initiation reactions were performed as detailed previously [3, 18]. These contained late G1 phase template nuclei prepared from mimosine-arrested human EJ30 cells, a cytosolic extract from proliferating HeLaS3 cells, *in vitro*-synthesised RNAs, and a buffered mix of ribo- and deoxyribonucleoside triphosphates. The mix included digoxigenin-dUTP as a tracer for newly synthesised DNA. Inclusion of creatine phosphate/phosphocreatine kinase provided an energy regenerating system. Endogenous human Y RNAs were depleted from the HeLa cell extract by endogenous RNAseH activity targeted by antisense DNA oligonucleotides, as detailed previously [3].

After a 3-hour reaction time, nuclei were fixed and sedimented on polylysine-coated glass coverslips. Digoxigenin-labelled DNA, the product of the replication reactions, was detected by anti-digoxigenin fluorescein-conjugated F_{ab} fragments (Roche), and total DNA was stained with propidium iodide, as described previously [3, 18]. Confocal fluorescence microscopy was performed on a Leica SP1 microscope using 40x lens magnification; individual channels were recorded simultaneously. The

percentages of replicating nuclei were determined from randomly chosen microscopic fields. At least 200 nuclei were scored per reaction.

Acknowledgements

This work was supported by Fundação Araucária (147/14 and 40/16), Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) and National Council for Scientific and Technological Development – Brazil - CNPq (grant number 305960/2015-6). SLP was supported by the Department of Zoology, University of Cambridge. FFDJ, JRPJ and LGP received graduated CAPES fellowships; PSAB, FSR and ACR received graduated fellowships from CNPq; DC and AFSL received undergraduate fellowships from CNPq and Fundação Araucária, respectively. The authors would like to thank Dr. Alain Debec for providing the *Drosophila melanogaster* S2 strain cells, LNCC for computational facilities; COMCAP/UEM for equipment facilities; and Dr. Omar Pereira and Milena Pavlickova for critical discussions.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Boria I, Gruber AR, Tanzer A, et al. Nematode sbRNAs: homologs of vertebrate Y RNAs. Journal of molecular evolution. 2010 Apr;70(4):346-58. doi: 10.1007/s00239-010-9332-4. PubMed PMID: 20349053.
2. Kowalski MP, Krude T. Functional roles of non-coding Y RNAs. The international journal of biochemistry & cell biology. 2015 Sep;66:20-9. doi: 10.1016/j.biocel.2015.07.003. PubMed PMID: 26159929; PubMed Central PMCID: PMC4726728.
3. Christov CP, Gardiner TJ, Szuts D, et al. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. Molecular and cellular biology. 2006 Sep;26(18):6993-7004. doi: 10.1128/MCB.01060-06. PubMed PMID: 16943439; PubMed Central PMCID: PMC1592862.
4. Collart C, Christov CP, Smith JC, et al. The midblastula transition defines the onset of Y RNA-dependent DNA replication in Xenopus laevis. Molecular and cellular biology. 2011 Sep;31(18):3857-70. doi: 10.1128/MCB.05411-11. PubMed PMID: 21791613; PubMed Central PMCID: PMC3165727.
5. Krude T, Christov CP, Hyrien O, et al. Y RNA functions at the initiation step of mammalian chromosomal DNA replication. Journal of cell science. 2009 Aug 15;122(Pt 16):2836-45. doi: 10.1242/jcs.047563. PubMed PMID: 19657016.
6. Kowalski MP, Baylis HA, Krude T. Non-coding stem-bulge RNAs are required for cell proliferation and embryonic development in C. elegans. Journal of cell science. 2015 Jun 1;128(11):2118-29. doi: 10.1242/jcs.166744. PubMed PMID: 25908866; PubMed Central PMCID: PMC4450293.
7. Mosig A, Guofeng M, Stadler BM, et al. Evolution of the vertebrate Y RNA cluster. Theory in biosciences = Theorie in den Biowissenschaften. 2007 Aug;126(1):9-14. doi: 10.1007/s12064-007-0003-y. PubMed PMID: 18087752.
8. Perreault J, Perreault JP, Boire G. Ro-associated Y RNAs in metazoans: evolution and diversification. Molecular biology and evolution. 2007 Aug;24(8):1678-89. doi: 10.1093/molbev/msm084. PubMed PMID: 17470436.
9. Gardiner TJ, Christov CP, Langley AR, et al. A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. Rna. 2009 Jul;15(7):1375-85. doi: 10.1261/rna.1472009. PubMed PMID: 19474146; PubMed Central PMCID: PMC2704080.
10. Wang I, Kowalski MP, Langley AR, et al. Nucleotide contributions to the structural integrity and DNA replication initiation activity of noncoding y RNA. Biochemistry. 2014 Sep 23;53(37):5848-63. doi: 10.1021/bi500470b. PubMed PMID: 25151917.
11. Lerner MR, Boyle JA, Hardin JA, et al. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science. 1981 Jan 23;211(4480):400-2. PubMed PMID: 6164096.
12. Langley AR, Chambers H, Christov CP, et al. Ribonucleoprotein particles containing non-coding Y RNAs, Ro60, La and nucleolin are not required for Y RNA function in DNA replication. PloS one. 2010 Oct 27;5(10):e13673. doi: 10.1371/journal.pone.0013673. PubMed PMID: 21060685; PubMed Central PMCID: PMC2965120.
13. Aftab MN, He H, Skogerbo G, et al. Microarray analysis of ncRNA expression patterns in Caenorhabditis elegans after RNAi against snoRNA associated

proteins. BMC genomics. 2008 Jun 11;9:278. doi: 10.1186/1471-2164-9-278. PubMed PMID: 18547420; PubMed Central PMCID: PMC2442092.
14. Deng W, Zhu X, Skogerbo G, et al. Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, and expression. Genome research. 2006 Jan;16(1):20-9. doi: 10.1101/gr.4139206. PubMed PMID: 16344563; PubMed Central PMCID: PMC1356125.
15. Duarte Junior FF, de Lima Neto QA, Rando Fdos S, et al. Identification and molecular structure analysis of a new noncoding RNA, a sbRNA homolog, in the silkworm Bombyx mori genome. Molecular bioSystems. 2015 Mar;11(3):801-8. doi: 10.1039/c4mb00595c. PubMed PMID: 25521575.
16. Christov CP, Trivier E, Krude T. Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. British journal of cancer. 2008 Mar 11;98(5):981-8. doi: 10.1038/sj.bjc.6604254. PubMed PMID: 18283318; PubMed Central PMCID: PMC2266855.
17. Kheir E, Krude T. Non-coding Y RNAs associate with early replicating euchromatin in concordance with the origin recognition complex. Journal of cell science. 2017 Apr 1;130(7):1239-1250. doi: 10.1242/jcs.197566. PubMed PMID: 28235841.
18. Krude T. Initiation of human DNA replication in vitro using nuclei from cells arrested at an initiation-competent state. The Journal of biological chemistry. 2000 May 5;275(18):13699-707. PubMed PMID: 10788489.
19. Chandra S, Vimal D, Sharma D, et al. Role of miRNAs in development and disease: Lessons learnt from small organisms. Life sciences. 2017 Sep 15;185:8-14. doi: 10.1016/j.lfs.2017.07.017. PubMed PMID: 28728902.
20. Hall AE, Turnbull C, Dalmay T. Y RNAs: recent developments. Biomolecular concepts. 2013 Apr;4(2):103-10. doi: 10.1515/bmc-2012-0050. PubMed PMID: 25436569.
21. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. Journal of molecular biology. 1990 Oct 5;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2. PubMed PMID: 2231712.
22. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research. 2003 Jul 1;31(13):3406-15. PubMed PMID: 12824337; PubMed Central PMCID: PMC169194.
23. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016 Jul;33(7):1870-4. doi: 10.1093/molbev/msw054. PubMed PMID: 27004904.
24. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 2009 Aug 1;25(15):1974-5. doi: 10.1093/bioinformatics/btp250. PubMed PMID: 19398448; PubMed Central PMCID: PMC2712331.
25. Popenda M, Szachniuk M, Antczak M, et al. Automated 3D structure composition for large RNAs. Nucleic acids research. 2012 Aug;40(14):e112. doi: 10.1093/nar/gks339. PubMed PMID: 22539264; PubMed Central PMCID: PMC3413140.
26. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. Journal of computational chemistry. 2005 Dec;26(16):1781-802. doi: 10.1002/jcc.20289. PubMed PMID: 16222654; PubMed Central PMCID: PMC2486339.
27. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996 Feb;14(1):33-8, 27-8. PubMed PMID: 8744570.

28. Mackerell AD, Jr., Feig M, Brooks CL, 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of computational chemistry. 2004 Aug;25(11):1400-15. doi: 10.1002/jcc.20065. PubMed PMID: 15185334.
29. Pacheco Homem D, Flores R, Jr., Tosqui P, et al. Homology modeling of dihydrofolate reductase from T. gondii bonded to antagonists: molecular docking and molecular dynamics simulations. Molecular bioSystems. 2013 Jun;9(6):1308-15. doi: 10.1039/c3mb25530a. PubMed PMID: 23450239.
30. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution. 1993 May;10(3):512-26. doi: 10.1093/oxfordjournals.molbev.a040023. PubMed PMID: 8336541.
31. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution; international journal of organic evolution. 1985 Jul;39(4):783-791. doi: 10.1111/j.1558-5646.1985.tb00420.x. PubMed PMID: 28561359.

Supplementary materials

Identification and characterization of stem-bulge RNAs in *Drosophila melanogaster*

Francisco Ferreira Duarte Junior[a], Paulo Sérgio Alves Bueno[b], Sofia L. Pedersen[c], Fabiana dos Santos Rando[d], José Renato Pattaro Júnior[b], Daniel Caligari[a], Anelise Cardoso Ramos[a], Lorena Gomes Polizelli[a], Ailson Francisco dos Santos Lima[a], Quirino Alves de Lima Neto[a], Torsten Krude[c], Flavio Augusto Vicente Seixas[b], and Maria Aparecida Fernandez[a]*

[a] Departamento de Biotecnologia, Genética e Biologia Celular, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil; [b] Departamento de Tecnologia, Universidade Estadual de Maringá, campus Umuarama, Av. Ângelo Moreira da Fonseca, 1800, Umuarama, 87506-370, Paraná, Brazil; [c] Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, United Kingdom. [d] Center for Molecular, Structural and Functional Biology - CBM/COMCAP, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil.

Keywords: DNA replication; sbRNA; non-coding RNAs; Y RNA

* **Corresponding author:** Maria Aparecida Fernandez; mafernandez@uem.br

Dm1

>AE014298.5 Drosophila melanogaster chromosome X
AATAGGAAATCTTAAATTTAGAAACGATGTGCGAACCTATTAATGCCATTACTTGGACCGCAGCTGTTTG
CCTGGGTGAT**GGGGTGGGTGTACCCGGAAAGTGGTGTTTACACGTCCCGTCCATGAGTAATCTGCGCACT**
TTCCCTCATAGTTCCTCCTCGTTTTCCCAGTTTCATTTTATTTGATTTGTTCGTTATTTGTTTGTTTGTT

Transcript:
GGGGUGGGUGUACCCGGAAAGUGGUGUUUACACGUCCCGUCCAUGAGUAAUCUGCGCACUUUCCCUCAUA
GUUCCUCCUCGUUUU

Figure S1. The *Drosophila melanogaster* stem-bulge RNA 1 gene and its regulatory sequences in the genome. A putative TATA box-like sequence in the predicted upstream RNA Polymerase III promoter region is shown in green. The nucleotide sequence predicted to be transcribed is represented in bold. The sequence motif conserved with nematode sbRNAs is highlighted in yellow. The poly-T tail, were the transcription ends, is highlighted in aqua-blue. The Dm1 gene is located between the position 16633124 – 16633208 of the *D. melanogaster* chromosome X.

Dm2

>AE014296.5 Drosophila melanogaster chromosome 3L
TTTGTCCGTATGTCCGTTGTCCGGCTGAAAATATTCCAAACAATGTAGACAAGAAGTGCTCAGGCCGGGC
GGAGTTTTTGT**GGCCATGGTTAGCGACGCGGTGGTGTTTACCCACATGTAGGCATATGCATATGCTGATT**
TCCACCAAGGCGGGGGTCATAGTCTTTTTTCTTTTTTTGAGGGGGGTTTGTGGTGGAGGAGTGGGCCGTC

Transcript:
GGCCAUGGUUAGCGACGCGGUGGUGUUUACCCACAUGUAGGCAUAUGCAUAUGCUGAUUUCCACCAAGGC
GGGGGUCAUAGUCUUUUUU

Figure S2. The *Drosophila melanogaster* stem-bulge RNA 2 gene and its regulatory sequences in the genome. A putative TATA box-like sequence in the predicted upstream RNA Polymerase III promoter region is shown in green. The nucleotide sequence predicted to be transcribed is represented in bold. The sequence motif conserved with nematode sbRNAs is highlighted in yellow. The poly-T tail, were the transcription ends, is highlighted in aqua-blue. The Dm2 gene is located between the position 12435720 – 12435808 of the *D. melanogaster* chromosome 3L.

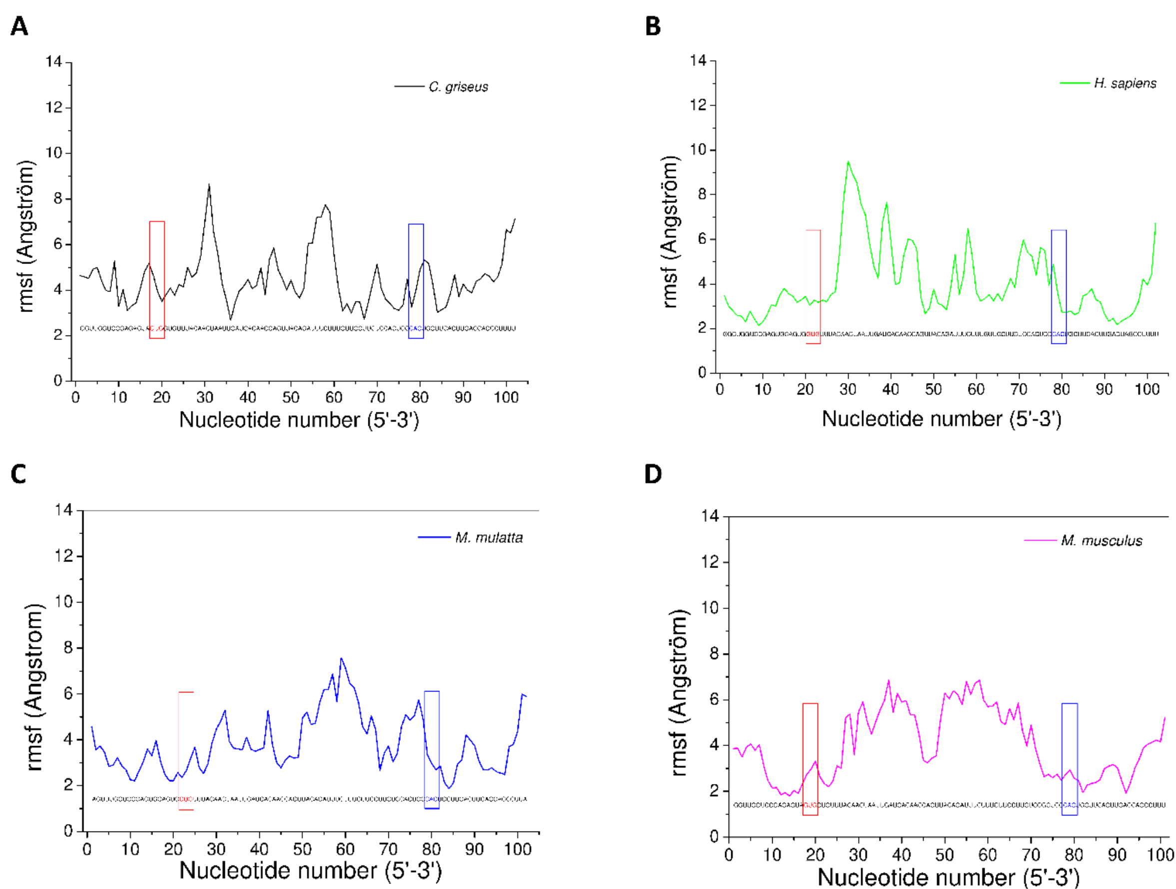

Figure S3. Molecular dynamics simulations. Root mean square fluctuation (rmsf) of C1' from Y RNAs used as reference. 3D representations of Y3 RNA from *Cricetulus griseus* (A), *Homo sapiens* (B) and *Mus musculus* (D) and Y5 RNA from *Macaca mulatta* (C) were used for this analysis (see supplementary figures S7-S10 for the corresponding pdb files). Red boxes highlight the rmsf of the paired GUG triplet and blue boxes of the paired CAC triplet.

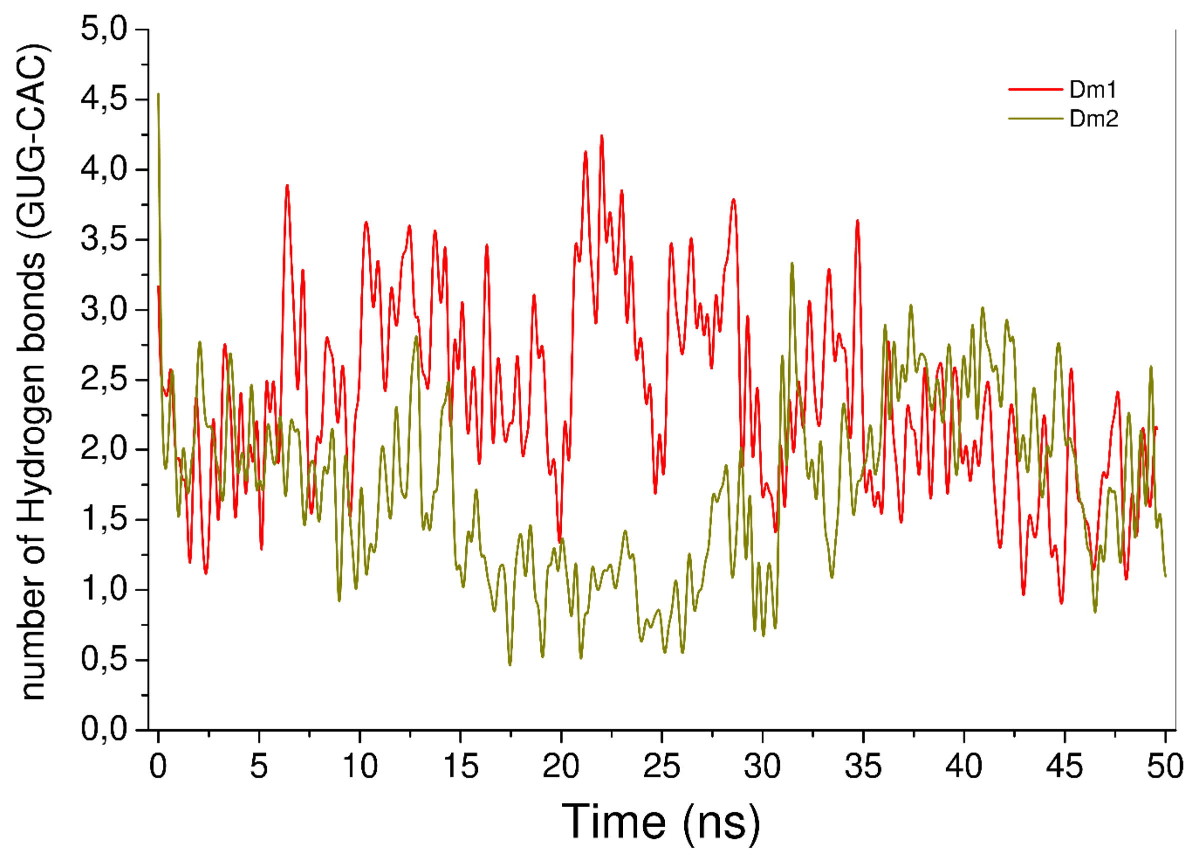

Figure S4. Molecular dynamics simulations. Number of hydrogen bonds within the GUG-CAC triplet of the *Drosophila melanogaster* Dm1 and Dm2 RNAs.

Legend to Figure S5.

3D representation of Dm1 RNA from *Drosophila melanogaster* in pdb file format.

Legend to Figure S6.

3D representation of Dm2 RNA from *Drosophila melanogaster* in pdb file format.

Legend to Figure S7.

3D representation of Y3 RNA from *Cricetulus griseus* in pdb file format.

Legend to Figure S8.

3D representation of Y3 RNA from *Homo sapiens* in pdb file format.

Legend to Figure S9.

3D representation of Y5 RNA from *Macaca mulatta* in pdb file format.

Legend to Figure S10.

3D representation of Y3 RNA from *Mus musculus* in pdb file format.

```
CLUSTAL O(1.2.4) multiple sequence alignment


CeN134_Caenorhabditis_elegans       -------------------------GGGTTATTTATTCTCTCG------ATTC-------  22
CeN72_Caenorhabditis_elegans        ----CATCATCGGTCCGGTGTTGATGGGTTATTATCCTGTGG------------------  38
CeN71_Caenorhabditis_elegans        -GAATTCCTGCGGTCCGGATCGTATGGGTTATCAATTCTC--------------------  39
CeN133_Caenorhabditis_elegans       --------ATCGGTCCGAAGTTGATGGGTTACCAATTGAATTCTCTCT-GTTCTTTTAAA  51
CeN73-1_Caenorhabditis_elegans      -----CACATCGGTCCGGAGTTGATGGGTTACCAGATTAATTCTTCTG-CTTGCAGGAGA  54
CeN73-2_Caenorhabditis_elegans      -----CACATCGGTCCGGAGTTGATGGGTTA-CCCAGTCATTCTTCTG-CTTGCAGGAGA  53
CeN74-1_Caenorhabditis_elegans      ------GTCTCGGTCCGGCGTCAGTGGGTTATCGTATTTCTCT-----------------  37
CeN74-2_Caenorhabditis_elegans      ------GTATCGGTCCGGCGTCAGTGGGTTATCAAGTTTGCCT-----------------  37
CeN135_Caenorhabditis_elegans       -ATCAGGCATAGGTCCGGAGTCGGTGGGTTATCTGAAGAAACC-----------------  42
CeN76_Caenorhabditis_elegans        CAGACAGGCGTGGTCCGGAGTCGGTGGGTTACCTTTGAAACCC-----------------  43
CeN75_Caenorhabditis_elegans        -----AAATACGGTCCGGAGTCGGTGGGTTATCTGAGAAGC-------------------  36
CeN77_Caenorhabditis_elegans        -----CAATTCGGTCCGGAGTCAATGGGTTATCTTTCAAAAAA-----------------  38
Dm2_Drosophila_melanogaster         -----GGCCATGGTTAGCGACGCGGTGGTGTTTACCCACATGTAGGCATATGC-------  48
Dm1_Drosophila_melanogaster         -----GGGGTGGGTGTACCCGGAAAGTGGTGTTTACACGTCCCGTCCA-----------T  44
BmsbRNA_Bombyx_mori                 --------GCTT--GAACGTCTCGTGGCTTATCCATAT----------------------  28
xY5_Xenopus_laevis                  -------AGTTGGTCCGATAATGGTGGGTTACCGTTTGTTTACGAAA-------------  40
XtY5_Xenopus_tropicalis             -------AGTTGGTCCGATAACAGTGGGTTATC-GTTGTTTACGAAA-------------  39
CpY5_Cavia_porcellus                -------AGTTGGTCCGAAGGCTGTGGGTTATTGTTATACATGGTTAA-CAT--------  44
PtY5_Pan_troglodytes                -------AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAACATT--------  45
hY5_Homo_sapiens                    -------AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAACAT---------  44
MmY5_Macaca_mulatta                 -------AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAACAT---------  44
chY5_Cricetulus_griseus             -------AGTTGGTCCGAAGGCTGTGGGTTATTGTTAAGTGTTTAGCATTTG--------  45
CfY5_Canis_familiaris               -------AGTTGGTCCGAGGGCTGTGGGTTATTGTTAAGCTGTTTAACATTG--------  45
LaY5_Loxodonta_africana             -------AGTTGGTCCGAGGGCTGTGGGTTATTGTTAAACTGTTTAGCATTG--------  45
MdY1_Monodelphis_domestica          -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAAATGATTGTTCACAGTCAGTTACA  53
chY1_Cricetulus_griseus             -------GGCTGGTCCGATGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
OcY1_Oryctolagus_cuniculus          -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
mY1_Mus_musculus                    -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
RnY1_Rattus_norvegicus              -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
BtY1_Bos_taurus                     -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
CpY1_Cavia_porcellus                -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGT-ACA  52
hY1_Homo_sapiens                    -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
PtY1_Pan_troglodytes                -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
MmY1_Macaca_mulatta                 -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
CfY1_Canis_familiaris               -------GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACA  53
zY1_Danio_rerio                     -------GGCTGGTCCGAAGGCGGTGGGTTAGTCACAATTGATTGCTACAGTCAGTTACA  53
zY3_Danio_rerio                     -------GGCTGGTCCGAAGGCGGTGG---------------------------------  20
xY3_Xenopus_laevis                  -------GGCTGGTCCGAAGG-CAGTGGTTGCCACCATTAATTGATTACAGACAGTTACA  52
XtY3_Xenopus_tropicalis             -------GGCTGGTCCGAAGG-CAGTGGCTGCTACCATTAATTGATTACAGACAGTTACA  52
MdY3_Monodelphis_domestica          -------GGCTGGTCCGATTG-CAGTGGTAACTCTAATTAATTGATTACAGTCAGTTACA  52
LaY3_Loxodonta_africana             -----GGCTGGTCCGAGTGCA-GTGTGAGGCTTACAACTAATTGATCACAACCAGTTACA  54
RnY3_Rattus_norvegicus              -------GGTTGGTCCGCGAG-TAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
chY3_Cricetulus_griseus             -------GGTTGGTCCGAGAG-TAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
mY3_Mus_musculus                    -------GGTTGGTCCGAGAG-TAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
CfY3P_Canis_familiaris              -------GGCTGGTCCGAGTG-CAGTGGTGCTTACAATTAATTGATCACAGCCAGTTACA  52
BtY3_Bos_taurus                     -------GGCTGGTCCGAGTG-CAGTGGTGTTTACAATTAATTGATCACAGCCAGTTACA  52
hY3_Homo_sapiens                    -------GGCTGGTCCGAGTG-CAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
PtY3_Pan_troglodytes                -------GGCTGGTCCGAGTG-CAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
CpY3_Cavia_porcellus                -------GGCTGGTCCGAGTG-CAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
MmY3_Macaca_mulatta                 -------GGTTGGTCCGAGTG-CAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
OcY3_Oryctolagus_cuniculus          -------GGCTGGTCCGAGTG-CAGTGGTGTTTACAACTAATTGATCACAACCAGTTACA  52
xY4_Xenopus_laevis                  -------GGTTGGTCCGAAAGTTGTGGGTTATCCAAATCATTC--AGTTAGTAT----CA  47
chY4_Cricetulus_griseus             -------GGTTGGTCTGATGTTAACGTGTTATTGGTAT----TAACTCTAG-------TG  42
MdY4_Monodelphis_domestica          -------GGCTGGTCCGATGGCAGTGGTTTACCAGAACTTATTGATATTAGTTT----CA  49
BtY4_Bos_taurus                     -------GGCTGGTCCGATGGTAGTGGGTTACCAGAACTTATTAACGTTAGTGT----CA  49
hY4_Homo_sapiens                    -------GGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGTGT----CA  49
PtY4_Pan_troglodytes                -------GGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGTGT----CA  49
MmY4_Macaca_mulata                  -------GGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGTGT----CA  49
OcY4_Oryctolagus_cuniculus          -------GGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGTGT----CA  49
CfY4_Canis_familiaris               -------GGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGTGT----CA  49
CpY4_Cavia_porcellus                -------GGCTGGTCCGATGGCAGTGGGTTATCAGAACTTATTAACATTAGTGT----CA  49
LaY4_Loxodonta_africana             -------GGCTGGTCCGATGGCAGTGGGTTATCAGAACTTATTAACGTTAGTGT----CA  49
```

```
CeN134_Caenorhabditis_elegans      ---AAATGCTCTTCACAATGCTTCACAGTGGT----TAATAATGCTTGTTTCTCCCATCA  75
CeN72_Caenorhabditis_elegans       --------TGCTTGTCGTCGC----------TGATCACATTCACCGTCCTCTACACATCA  80
CeN71_Caenorhabditis_elegans       ------------------------------------------------AACCACCCATAC  52
CeN133_Caenorhabditis_elegans      G------------------------------------------AACTACCCATCCCATCA  69
CeN73-1_Caenorhabditis_elegans     GCCCGGTGTCCTTGTGATGCCAAACCCGTGTTCCTAACAGAATACAACCCCTTCCCATCG  114
CeN73-2_Caenorhabditis_elegans     GCCCGGTGTCCTTGTGATGCCAAACCCGTGTTCCTAACAGAATACAAACCCTTCCCATCG  113
CeN74-1_Caenorhabditis_elegans     -------------------------------CCC------TT-CGGGGAATTTCCCATCG  59
CeN74-2_Caenorhabditis_elegans     -------------------------------CCC------TT-CGG-GAATTTCTCATCG  58
CeN135_Caenorhabditis_elegans      -------------------------------A--------------------CCCCATCG  51
CeN76_Caenorhabditis_elegans       -------------------------------CCC-----------------CTCCCATCG  55
CeN75_Caenorhabditis_elegans       ---------------------------------------------------CCCCCATCG  45
CeN77_Caenorhabditis_elegans       -------------------------------AAA------AAAAAAAAAACCCCCCATTG  61
Dm2_Drosophila_melanogaster        --------------------------------------------ATATGCTGATTTCCAC  64
Dm1_Drosophila_melanogaster        GAGT---------------------------------A------ATCTGCGCACTTTCCC  65
BmsbRNA_Bombyx_mori                --------------------------------------------------------CGAT  32
xY5_Xenopus_laevis                 --------------------------------------------TTCCCCC-CACCGTTG  55
XtY5_Xenopus_tropicalis            --------------------------------------------TTCTCCC-CACCGTTG  54
CpY5_Cavia_porcellus               -------------------------------------T----GTCTCCCCCCACA-CCCG  62
PtY5_Pan_troglodytes               -------------------------------------T----GTCTCCCCCCACA-ACCG  63
hY5_Homo_sapiens                   -------------------------------------T----GTCTCCCCCCACA-ACCG  62
MmY5_Macaca_mulatta                -------------------------------------T----GTCTCCCCCCACA-ACCG  62
chY5_Cricetulus_griseus            -------------------------------------T------CTCCCCCCACACCCTA  62
CfY5_Canis_familiaris              -------------------------------------T------CTCCCCCCACAACCAT  62
LaY5_Loxodonta_africana            -------------------------------------T------CTCCCACCCACACCCG  62
MdY1_Monodelphis_domestica         GATCGATCTCCTTGT----------------------TCTCTCTTTCCCTCCCTTCTCAC  91
chY1_Cricetulus_griseus            GATTGAACTCCTGTT----------------------CTAC--ACTTCCCCCCTTCTCAC  89
OcY1_Oryctolagus_cuniculus         GACCGATCTCCTGA-----------------------TCTACTCTTTCCCCCCTTGTCAC  90
mY1_Mus_musculus                   GATTGAACTCCTGT-----------------------TCTACACTTTCCCCCCTTCTCAC  90
RnY1_Rattus_norvegicus             GATTGAACTCCTGT-----------------------TCTACACTTTCCCCCCTTCTCAC  90
BtY1_Bos_taurus                    GATCGAACTCCTCGT----------------------TCTACTCTTCCC-CCCTTCTCAC  90
CpY1_Cavia_porcellus               GATTGAACTCCTTGT----------------------TCTACTCTTTCCCCCCTTCTCAC  90
hY1_Homo_sapiens                   GATCGAACTCCTTGT----------------------TCTACTCTTTCCCCCCTTCTCAC  91
PtY1_Pan_troglodytes               GATCGAACTCCTTGT----------------------TCTACTCTTTCCCCCCTTCTCAC  91
MmY1_Macaca_mulatta                GATCGAACTCCTTGT----------------------TCTACTCTTTCCCCCCTTCTCAC  91
CfY1_Canis_familiaris              GATTGAACTCCTTGT----------------------TCTACTCTTTCCCCCCTTCTCAC  91
zY1_Danio_rerio                    GAACTCCTTGTTTCT----------------------T------CTCTCCCCTCTCCCAC  85
zY3_Danio_rerio                    -------GTTAGTCT----------------------T------CTCTCCCCTCTCCCAC  45
xY3_Xenopus_laevis                 GACTTCTT----TGT----------------------T------CTTC-TCCCCTCCCAC  79
XtY3_Xenopus_tropicalis            GACTTCTT----TGT----------------------T------CTTC-TCCCCTCCCAC  79
MdY3_Monodelphis_domestica         GATTTCTT----TGT----------------------T------CTTTCTCCGCTCCCAC  80
LaY3_Loxodonta_africana            GATTCCTT----TGT----------------------T------CCTTCTCCACTCCCAC  82
RnY3_Rattus_norvegicus             GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
chY3_Cricetulus_griseus            GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
mY3_Mus_musculus                   GATTTCTT----TGT----------------------T------CCTTCTCCGCTCCCAC  80
CfY3P_Canis_familiaris             GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
BtY3_Bos_taurus                    GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
hY3_Homo_sapiens                   GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
PtY3_Pan_troglodytes               GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
CpY3_Cavia_porcellus               GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
MmY3_Macaca_mulatta                GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
OcY3_Oryctolagus_cuniculus         GATTTCTT----TGT----------------------T------CCTTCTCCACTCCCAC  80
xY4_Xenopus_laevis                 CTAA--------------------------------------CCTTCTATTTCACCCCAC  69
chY4_Cricetulus_griseus            CCGC---------------------------------TATAGTACATATAGGCCCCTCAC  69
MdY4_Monodelphis_domestica         CAAC---------------------------------AAGTTAATATATTCCACCCCCAC  76
BtY4_Bos_taurus                    CTAA---------------------------------AG-TTGGTATACAACCCCCCA-C  74
hY4_Homo_sapiens                   CTAA---------------------------------AG-TTGGTATACAACCCCCCA-C  74
PtY4_Pan_troglodytes               CTAA---------------------------------AG-TTGGTATACAACCCCCCA-C  74
MmY4_Macaca_mulata                 CTAA---------------------------------AG-TTGGTATACAACCCCCCA-C  74
OcY4_Oryctolagus_cuniculus         CTAA---------------------------------AG-TTGGTATACAACCCCCCA-C  74
CfY4_Canis_familiaris              CTAA---------------------------------AG-TTGGTATACAACCCCCCA-C  74
CpY4_Cavia_porcellus               CTAA---------------------------------AG-TTGGTATACCACCCCCCCAC  75
LaY4_Loxodonta_africana            CTAA---------------------------------AG-TTGGTATACAACCCCCCCAC  75
```

```
CeN134_Caenorhabditis_elegans       ACAA-CAACTTGACCGAATT--------  94
CeN72_Caenorhabditis_elegans        TCAC-AAATTTGACCGATGTT-------  100
CeN71_Caenorhabditis_elegans        GAAC-TAACTTGACTACCGGAATT----  75
CeN133_Caenorhabditis_elegans       ACAT-CAACTTGATCGAAAAACATTTTT  96
CeN73-1_Caenorhabditis_elegans      ACAC-CAACTTGACCGTTGTT-------  134
CeN73-2_Caenorhabditis_elegans      ACAC-CAACTTGACCGTTGT--------  132
CeN74-1_Caenorhabditis_elegans      GCAT-CAACTTGACCGTTGCGT------  80
CeN74-2_Caenorhabditis_elegans      GCAC-CAACTTGACCGTTGCT-------  78
CeN135_Caenorhabditis_elegans       GCAA-CAACTTGACTTCC----------  68
CeN76_Caenorhabditis_elegans        GCAC-CAACTTGACCGTTCCTGTT----  78
CeN75_Caenorhabditis_elegans        ACAC-CAACTTGACCGATGA--------  64
CeN77_Caenorhabditis_elegans        ACAA-CAACTTGACCGGCGT--------  80
Dm2_Drosophila_melanogaster         CAAGGCGGGGGTCATAGTCTTTTTT---  89
Dm1_Drosophila_melanogaster         TCAT-AGTT---CCTCCTCGTTTT----  85
BmsbRNA_Bombyx_mori                 GTCT-TCAAAAGATTCTTCTAGTTTT--  57
xY5_Xenopus_laevis                  CCAT-TGACTAACGA-------------  69
XtY5_Xenopus_tropicalis             CCAT-TGACTAACG--------------  67
CpY5_Cavia_porcellus                TGCT-TGACTAACTGAGTGTTTT-----  84
PtY5_Pan_troglodytes                CGCT-TGACTAGCTTGCTGTTTT-----  85
hY5_Homo_sapiens                    CGCT-TGACTAGCTTGCTGTTTT-----  84
MmY5_Macaca_mulatta                 CGCT-TGACTAGCTTGACGTTTT-----  84
chY5_Cricetulus_griseus             -ACT-TGACTAGCTTGAGCTTTTT----  84
CfY5_Canis_familiaris               C-CT-TGACTGGCTTCGCGTTTT-----  83
LaY5_Loxodonta_africana             ACCT-TGACTAGCTGTGCCTTTT-----  84
MdY1_Monodelphis_domestica          TACT-GCACTCGACTAGTCTTT------  112
chY1_Cricetulus_griseus             TACT-GCACTTGACTAGTCTTTT-----  111
OcY1_Oryctolagus_cuniculus          TACT-GCACTAGACTAGTCTTT------  111
mY1_Mus_musculus                    TACT-GCACTTGACTAGTCTTT------  111
RnY1_Rattus_norvegicus              TACT-GCACTTGACTAGTCTT-------  110
BtY1_Bos_taurus                     TACT-GCACTTGACTAGTCTTT------  111
CpY1_Cavia_porcellus                TACT-GCACTTGACTAGTCTTT------  111
hY1_Homo_sapiens                    TACT-GCACTTGACTAGTCTTT------  112
PtY1_Pan_troglodytes                TACT-GCACTTGACTAGTCTTT------  112
MmY1_Macaca_mulatta                 TACT-GCACTTGACTAGTCTTT------  112
CfY1_Canis_familiaris               TACT-GCACTTGACTAGTCTTT------  112
zY1_Danio_rerio                     CGCT-GAACTTGACCAGCCTTT------  106
zY3_Danio_rerio                     CGCT-AAACTTGACCAGTCTTT------  66
xY3_Xenopus_laevis                  TGCT-TCCCTTGACTAGCCT--------  98
XtY3_Xenopus_tropicalis             TGCT-GCCCTTGACTAGTCT--------  98
MdY3_Monodelphis_domestica          TGCT-TCACTTGACTAGTCTTT------  101
LaY3_Loxodonta_africana             TGCT-TCACTAGACCGGTCTTT------  103
RnY3_Rattus_norvegicus              TGCT-TCACTTGACCAGCCTTT------  101
chY3_Cricetulus_griseus             TGCT-TCACTTGACCAGCCTTTT-----  102
mY3_Mus_musculus                    TGCT-TCACTTGACCAGCCTTT------  101
CfY3P_Canis_familiaris              TGCT-TCACTTGACCGGCCTTT------  101
BtY3_Bos_taurus                     TGCT-TCACTTGACTAGCCTTT------  101
hY3_Homo_sapiens                    TGCT-TCACTTGACTAGCCTTT------  101
PtY3_Pan_troglodytes                TGCT-TCACTTGACTAGCCTTT------  101
CpY3_Cavia_porcellus                TGCT-TCACTTGACTAGCCTTT------  101
MmY3_Macaca_mulatta                 TGCT-TCACTTGACCAGCCTTT------  101
OcY3_Oryctolagus_cuniculus          TGCT-TCACTTGACCAGTCTTT------  101
xY4_Xenopus_laevis                  TGCT-GACCTTGACTGGCCA--------  88
chY4_Cricetulus_griseus             TGAT-AAGTTTGACTGGCTTTTT-----  91
MdY4_Monodelphis_domestica          TGCT-AAATTTGACTGGCCTT-------  96
BtY4_Bos_taurus                     TGCT-AAATTTGACTGGCTTT-------  94
hY4_Homo_sapiens                    TGCT-AAATTTGACTGGCTTT-------  94
PtY4_Pan_troglodytes                TGCT-AAATTTGACTGGCTTT-------  94
MmY4_Macaca_mulata                  TGCT-AAATTTGACTGGCTTT-------  94
OcY4_Oryctolagus_cuniculus          TGCT-AAATTTGACTGGCTTT-------  94
CfY4_Canis_familiaris               TGCT-AGATTTGACTGGCTTT-------  94
CpY4_Cavia_porcellus                TGCT-AAATTTGACTGGCTTT-------  95
LaY4_Loxodonta_africana             TGCT-GAATTTGACTGGCTTT-------  95
```

Figure S11. Sequence alignment of 62 stem-bulge RNA and Y RNA genes using Clustal Omega v. 1.2.4.

Artigo 2: Stem-bulge RNAs from *Drosophila melanogaster* and their interaction with chromatin proteins.

(Será submetido à revista *RNA Biology* - ISSN: 1555-8584/ *Impact Factor* JCR: 5,28)

Stem-bulge RNAs from *Drosophila melanogaster* and their interaction with chromatin proteins

Francisco Ferreira Duarte Junior[a*], José Renato Pattaro Júnior[b], Eduardo Jorge Pilau[c], Quirino Alves de Lima Neto[a], Letícia Aparecida de Oliveira[a], Fabiana dos Santos Rando[d], Daniel Caligari[a], Daniele Maria Zanzarin[a], Flavio Augusto Vicente Seixas[b] and Maria Aparecida Fernandez[a*]

[a] Departamento de Biotecnologia, Genética e Biologia Celular, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil; [b] Departamento de Tecnologia, Universidade Estadual de Maringá, campus Umuarama, Av. Ângelo Moreira da Fonseca, 1800, Umuarama, 87506-370, Paraná, Brazil; [c] Departamento de Química, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil; [d] Center for Molecular, Structural and Functional Biology - CBM/COMCAP, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil.

***Corresponding author:** Maria Aparecida Fernandez, mafernandez@uem.br;
Francisco Ferreira Duarte Junior, juniorf.duarte@hotmail.com;

Stem-bulge RNAs from *Drosophila melanogaster* and their interaction with chromatin proteins

Abstract

Non-coding RNAs are molecules responsible for several metabolic activities, except protein-coding. Among them, the Y RNAs were first detected in patients with autoimmune diseases, such as Erythematous Lupus and Sjögren syndrome. The group of Y RNA is present in both prokaryotes and eukaryotes. They are related to the initiation of DNA replication, in eukaryotes, more specifically in vertebrates. However, prokaryote Y RNAs do not have this function, performing the degradation of mature RNAs. In parallel, another group performs the initiation of DNA replication, in invertebrates, the stem-bulge RNAs. They are homologous to Y RNAs in structure and function. Stem-bulge RNAs present a "UUAUC" sequence, evolutionarily conserved. Until 2015, there was no register of any sbRNA gene in insects. The first expressed gene was described for the silkworm, *Bombyx mori*. Furthermore, two genes, Dm1 and Dm2, were found in the fruit fly *Drosophila melanogaster*. While Dm1 presented functional activity in the initiation of DNA replication, this function could not be observed for Dm2. With this in mind, the literature does not contain further information about sbRNAs in insects. So, we aimed to observe if there was any differential expression of Dm1 and Dm2 in male and female adult flies. Additionally, we used mass spectrometry to investigate which cytoplasmic proteins were interacting with the three insect sbRNAs, previously synthesized and bound to agarose beads. This investigation could Our data shows that male flies express Dm1 and Dm2 at a higher rate than females. Moreover, we found out that Dm1 bound with proteins related with replicative and immune activities, such as cecropins A1. Dm2 bound with proteins related to structure maintenance, but no

replicative proteins. This is supported by results previously described for these two sbRNAs.

Keywords: DNA replication; sbRNA; non-coding RNAs; Y RNA, *Drosophila melanogaster*, *Bombyx mori*, Mass Spectrometry.

Introduction

Non-coding RNAs (ncRNA) are described as molecules with metabolic functions different from protein codification (Kowalski & Krude., 2015). Among then, there are mid-sized molecules, ranging between 70 and 115 nucleotides, with none to four expressing genes, named Y RNAs (Mosig et al., 2007; Perreault et al., 2007). The "Y RNA" name derives from cYtoplasmic RNAs, local of their first detection in mammalian cells (Lerner et al., 1981). The presence of Y RNAs was first observed in humans, they were detected as ribonucleoproteins (RNPs), complexed with Ro60 and La proteins, in blood serum from patients with Erythematous Lupus and Sjögren syndrome (Hendrick et al., 1981; Lerner et al., 1981). Later, they were described as participants in the initiation of DNA replication, with all four genes replacing each other in this function (Christov et al., 2006; Krude et al., 2009; Gardiner et al., 2009).

Recently, homologous genes were described for Chinese hamster, named chY1, chY3, chY4 and chY5 and presenting great identity to the four human genes (hY1, hY3, hY4, and hY5). However, only chY1 and chY3 genes were shown to be expressed, even with all four being able to replace endogenous RNAs in the initiation of replication (Lima Neto et al., 2016). The Y RNAs have been investigated and detected in several organisms, from prokaryotes (Chen et al., 2014) to eukaryotes (Perreault et al., 2007). Y RNAs are transcribed by RNA Polymerase III and present a characteristic secondary structure, with a stem-loop shape (Maraia et al., 1996; Teunissen et al., 2000; Christov et al., 2006; Kowalski & Krude, 2015). Additionally, Y RNAs have an unpaired C residue in the Y RNAs lower-stem, a binding site for the Ro60 protein (Lerner et al., 1981). While the literature suggests that vertebrate Y RNAs can participate in the initiation of DNA replication. This information is not valid for invertebrate Y RNAs, such as the ones found in *Caenorhabditis elegans* (CeY RNA), *Branchiostoma floridae* (BfY RNA) and

Deinococcus radiodurans (DrY RNA) (Gardiner et al., 2009). With this in mind, some authors claim that another group of non-coding RNAs plays this role in invertebrates, this group is known as stem-bulge RNAs (sbRNAs) (Boria et al., 2010; Perreault et al., 2007; Kowalski et al., 2015; Duarte Junior et al., 2015; 2019). Both sbRNAs and Y RNAs have similarities, such as the RNA Polymerase III transcription motif, the stem-loop secondary structure and a conserved minimal domain of double-stranded nucleotides (GUG-CAC), which is minimal for their functioning as essentials to licensing of the initiation of DNA replication (Christov et al., 2006; Gardiner et al., 2009; Wang et al., 2014; Kowalski et al., 2015). However, there are differences between them. The sbRNAs contain a conserved sequence (UUAUC) and do not show the unpaired C residue, present in the Y RNAs (Boria et al., 2010). The sbRNAs are still neglected, compared with other ncRNAs, and were first detected in nematodes (Deng et al., 2006). Recently, some sbRNA genes were shown to be expressed in insects, such as the *Bm*sbRNA from the silkworm *Bombyx mori* (Duarte Junior et al., 2015) and the Dm1 and Dm2 from *Drosophila melanogaster* (Duarte Junior et al., 2019). However, Dm1 is the only sbRNA, from insects, capable of replacing endogenous Y RNAs in the initiation of DNA replication in a human cell-free system (Duarte Junior et al., 2019). The sbRNAs lack further information regarding their functional roles and location, while Y RNAs have been observed in the bloodstream (Dhahbi et al., 2014), in vesicles from immune cells and in several other extracellular vesicles, from humans and murines to retroviruses (Nolte-'t Hoen et al., 2012; Driedonks & Nolte-'t Hoen, 2019).

Therefore, to unravel the functional roles of the insect sbRNAs, we intended to observe the interaction between these molecules and cytoplasmic proteins. Using mass spectrometry, we searched for possible functions base on the proteins bound, inferring the roles played by these sbRNAs. We bound and co-precipitated two synthetic sbRNAs

from *D. melanogaster* (Dm1 and Dm2) to agarose beads and total protein extracts from S2 embryonary cell cultures. The proteins that bound to these sbRNAs show unprecedented results, suggesting that the Dm1 sbRNA binds with proteins related to immune responses and replicative proteins and Dm2 binds with structure maintenance proteins. Furthermore, we checked the differential expression of both Dm1 and Dm2 between male and female flies of *D. melanogaster*. We found out that male flies have the highest expression of both sbRNAs. We also observed that the sum of Dm1 and Dm2 expressions is higher in males than females, suggesting that their expression might be related with proliferative events, such as spermatogenesis, that occurs in testis.

Results

Relative expression of DmsbRNA in males and females

The expression of *Drosophila* sbRNAs (Dm1 and Dm2) has already been observed in cell cultures and adult insects. The literature suggests that Dm1 is more expressed than Dm2, with that one being more frequent in flies than S2 cells (Duarte Junior et al., 2019). However, there is no further information if there are any differences in their expression between males and females (Figure 1). In order to access this question, we used RT-qPCR to compare the expression rates in male and female adult flies. Experiments showed that both Dm1 and Dm2 were more expressed in males. However, only Dm2 expression was statistically different, between males and females. When compared with 5S expression, Dm1 gene was 204-fold lower (0.49 %), while Dm2 expressed 192-fold lower (0.52 %), approximately. In females, expression rates were 294-fold lower (0,34%) and 357-fold lower, for Dm1 and Dm2, respectively.

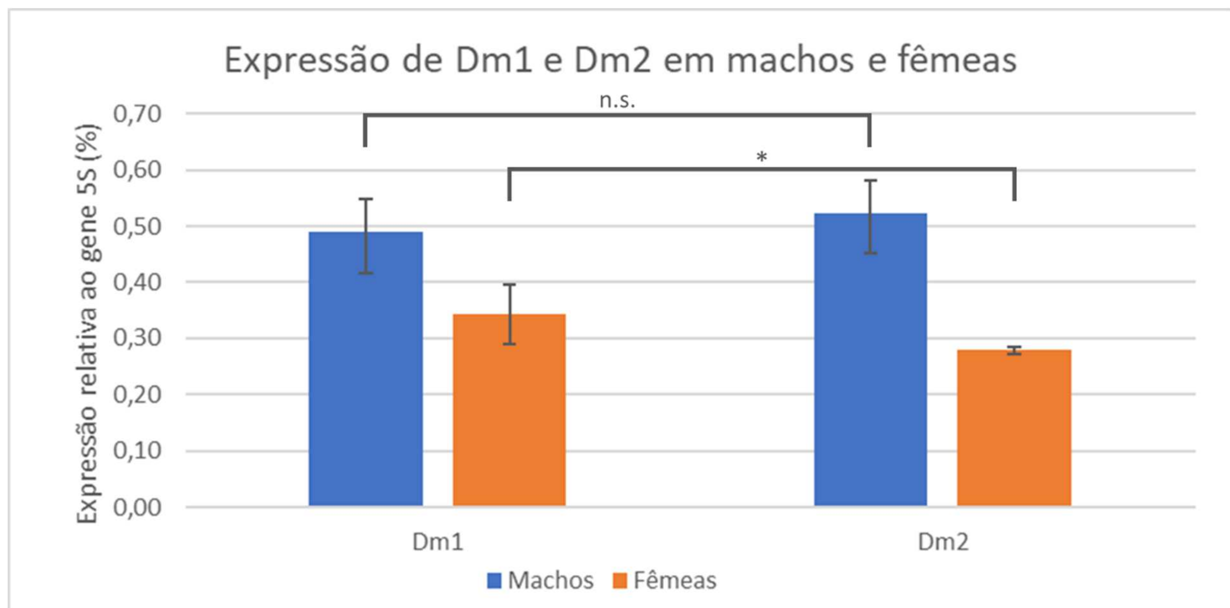

Figure 1. Relative expression of Dm1 and Dm2 sbRNAs in male and female adult flies. The expression rates are relative to 5S rRNA expression, quantified by RT-qPCR. Brackets show results of one-way ANOVA with Tukey's Test (* = P<0,05; n.s. = non-significant). Samples were analyzed in triplicates and data are shown as mean values ± standard deviations of three independent experiments.

Protein-RNA interaction assays

Insect stem-bulge RNAs (Dm1, Dm2, and *Bm*sbRNA) were successfully synthesized by SP6 polymerase. After DNAse I treatment, synthetic insect sbRNAs were coupled with pre-treated beads, forming an RNA-bead complex (Figure 2A). The formation of RNA-bead complexes was verified by agarose gel electrophoresis (Figure 2B), where uncoupled RNA migrates through the gel and the RNA-bead complex is restrained to their respective wells (Figure 2B, odd lanes).

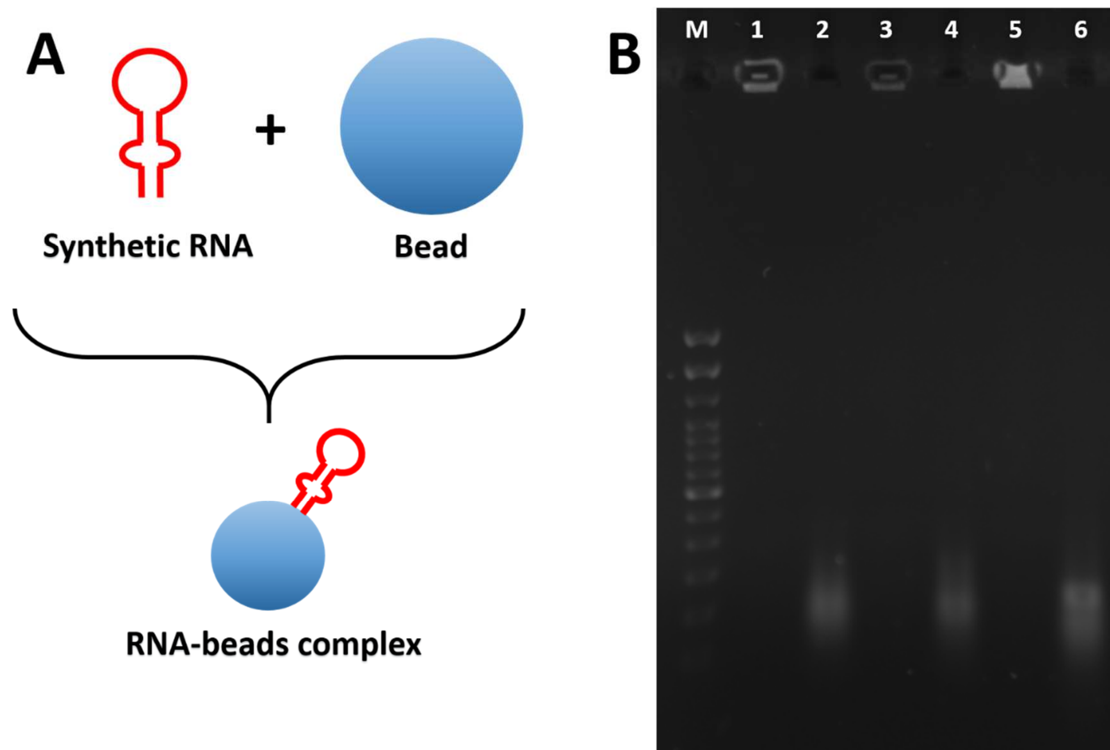

Figure 2. Schematic interpretation of synthetic stem-bulge RNAs from *D. melanogaster* complex. (A) RNA-beads complex, composed of synthetic sbRNAs from *Drosophila* melanogaster (Dm1 and Dm2) and adipic acid dihydrazide agarose beads. (B) Agarose electrophoresis showing the complexes (odd lanes) and free synthetic RNAs (even lanes). The RNAs are Dm1 (lanes 1 and 2), Dm2 (lanes 3 and 4), and BmsbRNA (lanes 5 and 6). The "M" lane represents the 100 bp ladder (Invitrogen).

Following that confirmation, pre-depleted (with RNA-free beads) samples of total proteins, from the respective organisms, were incubated with the respective RNA-beads complexes. The proteins, bound with the RNA-beads complex, were washed, separated and digested with trypsin. The digested samples were treated and loaded in mass spectrometer. The results were analysed for each sample, with no common proteins among then (Table S1).

The experiments with Dm1 sbRNA suggest that it bound with three proteins: two cecropins A1 (Uniprot IDs: P81688 and P81685) and an uncharacterized protein (Uniprot

ID: B4P7W4). We detected cecropins A1 from *D. sechellia* and *D. mauritana*. However, both species belongs to the *melanogaster* group (Ramos-Onsins & Aguadé, 1998; Clark et al., 2007). Besides, cecropin A1 from *D. sechellia* has a valine residue in the 50th position of its chain, compared to isoleucine in cecropin A1 from *D. melanogaster* (Uniprot ID: C0HKQ7). Binding experiments with Dm2 resulted in four proteins: IP05681p (Uniprot ID: Q4QPY5), GEO12070p1 (Uniprot ID: Q9VKK9), Trichohyalin (Uniprot ID: A0A1W4VAI5), Protein hu-li tai shao (Uniprot ID: Q02645).

Discussion

In this study, we investigated if the expression rates vary between male and female *Drosophila* flies. RT-qPCR assays indicate that both Dm1 and Dm2 are more expressed in male flies, with expression rates of 0.49 % and 0.52%, respectively. We also checked that the total expression (sum of Dm1 and Dm2) is also higher in males, approximately 1.01 % for Dm1, and 0.62 % for Dm2. This might be a reflex of the elevated cell proliferation that occurs in gonads, especially in testis with ongoing spermatogenesis. This event is very present in male insects, after the metamorphosis, since it precedes mating (Hoy, 2013).

In the meantime, RNA/protein binding experiments identified some proteins interacting with our synthetic sbRNAs from insects. Dm1 was observed binding with immune-related peptides, named cecropins A1, and an uncharacterized protein, described as a participant in several replicative events such as heterochromatin processing and regulation of transcription. Cecropins consist of a group of antimicrobial peptides acts in the innate immune systems, especially against bacterial and fungal infections. This function is related to their structures, composed of two amphipathic α-helices, responsible for penetrating cellular membranes of bacteria (Samakovlis et al., 1990; Chen et al.,

2018). It was verified that cecropins were effective against most fungi, present in *Drosophila's* environment (Ekengren & Hultmark, 1999), which supports the theory that, maybe, S2 cells were secreting cecropins as a reflex to some undetected latent contamination of the cell culture. The uncharacterized protein (B4P7W4) is described acting in negative regulation of transcription, germ-line cyst formation, pole cell development and migration and regulation of heterochromatin (Clark et al., 2007). This protein is detected in polar granules, small RNA/protein complexes with no membranes, that are aggregated in the primordial germ cells of many higher eukaryotes (Schisa et al., 2001). Although Dm1 sbRNA is the only insect gene that has been observed acting licensing of the initiation of DNA replication (Duarte Junior et al., 2019), our results support that previous affirmation (associated with this uncharacterized protein) and suggests that this molecule can perform even more metabolic functions, depending what protein it complexes with. Therefore, we might assume that Dm1 might play a role in immune responses (when associated with cecropins) and in DNA replication (when associated with the uncharacterized proteins and data published by Duarte Junior et al., 2019).

Dm2 sbRNA was found interacting with four proteins, which biological function was structural maintenance of cuticles, organization of sarcomere, ovarian and testicular fusomes, and adult somatic muscle development. The proteins IP05681p and GEO12070p1 are described as inhibitors for protein phosphatases (Leach et al., 2002) and regulator of transduction signals, an event that occurs in olfaction (Sato & Touhara, 2008). The third protein, Trichohyalin, is reported as a structural constituent of the cuticle (Clark et al., 2007). The last one, Protein hu-li tai shao, is the most intriguing for Dm2 results. This protein, whose name means “Too little nursing”, is involved in several biological activities. It is related to actin filament binding, required in developing egg

chambers, interacting with other developmental proteins that transport nurse cells/oocytes (Yue & Sprandling, 1992). Also, this protein acts in oogenesis, ovarian and testicular fusome organization (assembly, arrangement or disassembly of constituent parts of the fusome, organelle derived from the spectrosome), sarcomere organization and adult somatic muscle development (Matthews et al., 2015). The results shown for Dm2 might suggest the reasons why it could not substitute endogenous Y RNAs in the initiation of DNA replication in a cell-free system (Duarte Junior et al., 2019) since the proteins bound to Dm2 show functions and profiles different than the proteins bound to Dm1.

Y RNAs were initially discovered in association with Ro60 and La proteins, forming ribonucleoprotein complexes (RNPs) (Hendrick et al., 1981; Lerner et al., 1981). Also, Y RNAs are already known for their association with diverse proteins (Zhang et al., 2011; Kowalski et al., 2015), but the information obtained from our results is unknown for sbRNAs. Here, we show the first data of interaction between sbRNAs from *D. melanogaster* with proteins from cytoplasm of S2 cell cultures. Furthermore, our data support the idea that Dm1 is related to the initiation of DNA replication, once it was found bound with proteins with the same functional profile. Regarding Dm2, we might have found a clue that it might perform structural maintenance activity and why it was not capable of playing the same function as the Dm1, in a cell-free system (Duarte Junior et al., 2019).

Materials and Methods

Relative expression of DmsbRNA in males and females

Adult flies, of the *Drosophila melanogaster* species, were acquired from the State University of Maringá, Maringá - Paraná, Brazil. They were separated by gender and macerated in liquid nitrogen, followed by extraction with TRIzol LS (Invitrogen).

Total RNA sample concentrations were quantified with NanoDrop2000 (Thermo-Fisher) equipment and standardized for a total of 5000 ng of RNA to be treated with DNAse I (Biolabs). The reverse transcriptase kit used to synthesize the first cDNA was the Quantinova Reverse Transcription kit (Qiagen). For the RT-qPCR reactions, we used the Fast SYBR Green PCR Master Mix (Thermo Fisher), with 40 cycles and annealing temperature of 58°C, in the Rotor-Gene Q (Qiagen) equipment. The primer sequences and methodology were according to described in the literature (Duarte Junior et al., 2019). Our experiments were conducted in triplicates, based in 3 independent experiments (n=3).

Cell cultures and protein extraction

Protein samples were extracted from different cellular cultures. The cytoplasmic proteins were extracted from S2 embryonic cells (fruit fly - *Drosophila melanogaster*). S2 cells were maintained with Shields and Sang M3 insect culture media (Sigma-Aldrich), supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco) and antibiotics at a temperature of 26°C. This cellular strain was kindly provided by Dr. Alain Debec (Institut Jacques Monod, Paris, France).

Protein-RNA interaction assays

Stem-bulge RNA genes from *Drosophila melanogaster* Dm1 (NCBI ID: MN654365) and Dm2 (NCBI ID: MN661249) were cloned in TOPO-TA (Duarte Junior et al., 2019) and used as templates in conventional PCR amplification. PCR amplicons were used as templates in the *in vitro* reaction, carried by bacteriophage SP6 RNA polymerase, as previously described (Wang et al., 2014). RNA samples, synthesized by SP6 Polymerase, were treated with DNAse I (Biolabs) and purified by phenol/chloroform extraction and ethanol precipitation. To verify the quality of the treated RNAs, samples

were analysed in agarose gel electrophoresis. The RNAs were oxidized and covalently coupled to adipic acid dihydrazide agarose beads (Invitrogen). The coupling reaction was quenched with sodium cyanoborohydride and, then, the RNA-coupled beads were washed with RNA Binding Buffer (20 mM K-Hepes pH 7,8; 100 mM K-Ac; 2 mM $MgCl_2$; 1 mM DTT). After the protein-RNA-bead complex was obtained, samples were washed with RNA Binding Buffer 3 times and heated, without SDS, for 5 minutes at 95ºC. The supernatant was reserved for the Mass Spectrometry protocol. All the protocol was carried out using RNAse free water (Sigma) for the solutions and was followed as previously described (Zhang et al., 2011), with the adjustments indicated above.

Mass Spectrometry experiments

The proteins previously eluted, were submitted to treatment and digestion with trypsin (Promega), according to the protocol established by Villén & Gygi (2008), with modifications by Laboratório de Espectrometria de Massas LNBio (LNBio: https://lnbio.cnpem.br/facilities/mass-spectrometry/sample-preparation/).

LC-MS/MS analyses with tryptic peptides solutions, derived from proteins bound to RNA-bead complex, were performed on an M-Class Waters nano-LC coupled to a Waters Xexo G2-si mass spectrometer (TOF geometry) using MS^e acquisition mode. The UPLC system was equipped with a desalination column (trap column Waters Symmetry 38 C18; 20 mm x 180 id µm; 5 µm particle size) followed by an analytical column (C18 Waters BEH130; 100 mm x 100 id µm; 1.7 µm particle size). Injected samples were desalted on the trap column for 3 min, with a 5 µL min-1 flow of 97:3 H_2O/MeCN with 0.1% (v/v) formic acid. The following step was the elution, performed by directing the flow into the analytical column, using 1.0 µL min-1 flow with a gradient of 97:3 to 30:70% H_2O/MeCN with 0.1% (v/v) formic acid, for a total analysis time of 15 min.

Then, the RAW files were processed using ProteinLynx Global Server version 3.0.3 (Waters). Each organism had its protein database downloaded in Uniprot (https://www.uniprot.org/). The samples were analysed in the mass spectrometer (Waters) from the Complexo de Centrais de Apoio à Pesquisa from the Universidade Estadual de Maringá – COMCAP/UEM.

Acknowledgments

This work was supported by Fundação Araucária (147/14 and 40/16), Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) and National Council for Scientific and Technological Development – Brazil - CNPq (grant number 305960/2015-6). The authors would like to thank the COMCAP/UEM, for equipment facilities.

Conflict of Interest

The authors declare no conflict of interest.

References

Boria I, Gruber AR, Tanzer A, et al. Nematode sbRNAs: homologs of vertebrate Y RNAs. Journal of molecular evolution. 2010 Apr;70(4):346-58. doi: 10.1007/s00239-010-9332-4. PubMed PMID: 20349053.

Chen, X., Sim, S., Wurtmann, E.J., Feke, A., Wolin, S.L. Bacterial noncoding YRNAs are widespread and mimic tRNAs. *RNA*, 2014, 20, 1715–1724.

Chen, Y.-W., Wu, C.-P., Wu, T.-C., & Wu, Y.-L. Analyses of the transcriptome of Bombyx mori cells infected with either BmNPV or AcMNPV. 2018, Journal of Asia-Pacific Entomology, 21(1), 37–45. doi:10.1016/j.aspen.2017.10.009

Christov CP, Gardiner TJ, Szuts D, et al. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. Molecular and cellular biology. 2006 Sep;26(18):6993-7004. doi: 10.1128/MCB.01060-06. PubMed PMID: 16943439; PubMed Central PMCID: PMC1592862.

Clark, A., Eisen, M., Smith, D., et al. Evolution of genes and genomes on the Drosophila phylogeny. 2007, Nature 450, 203–218.

Deng, W.; Zhu, X.; Skogerbo, G.; et al. Organization of the *Caenorhabditis elegans* small non-coding transcriptome: genomic features, biogenesis, and expression. Genome research, 2006, Jan, 16, (1):20-9.

Dhahbi, J. M., Spindler, S. R., Atamna, H., Boffelli, D., Martin, D. I. Deep Sequencing of Serum Small RNAs Identifies Patterns of 5′ tRNA Half and YRNA Fragment Expression Associated with Breast Cancer, *Biomarkers in Cancer*, 2014, 6, 37–47.

Driedonks, T.A.P., and Nolte-'t Hoen, E.N.M. Circulating Y-RNAs in Extracellular Vesicles and Ribonucleoprotein Complexes; Implications for the Immune System. 2019, *Front. Immunol.* 9:3164. doi: 10.3389/fimmu.2018.03164

Duarte Junior FF, de Lima Neto QA, Rando Fdos S, et al. Identification and molecular structure analysis of a new noncoding RNA, a sbRNA homolog, in the silkworm Bombyx mori genome. Molecular bioSystems. 2015 Mar;11(3):801-8. doi: 10.1039/c4mb00595c. PubMed PMID: 25521575.

Duarte Junior, F. F; Bueno, P. S. A; Pedersen, S. L; Rando, F. S; Pattaro Junior, J. R; Caligari, D; Ramos, A. C; Polizelli, L. G; Lima, A. F. S; Lima Neto, Q. A; Krude, T; Seixas, F. A. V; Fernandez, M. A. Identification and characterization of stem-bulge RNAs in Drosophila melanogaster. RNA Biology, 2019, DOI: 10.1080/15476286.2019.1572439.

Ekengren, S. & Hultmark, D. *Drosophila* cecropin as an antifungal agent. 1999, Insect Biochemistry and Molecular Biology, 29, 965–972.

Gardiner TJ, Christov CP, Langley AR, et al. A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. Rna. 2009 Jul;15(7):1375-85. doi: 10.1261/rna.1472009. PubMed PMID: 19474146; PubMed Central PMCID: PMC2704080.

Hendrick J. P.; Wolin S. L.; Rinke J.; Lerner M. R.; Steitz J. A. Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. Mol. Cell. Biol. 1981, 1:1138–1149.

Hoy, M. A. Insect Molecular Genetics. Academic Press. 30th April, 30th, 2013.

Kowalski MP, Baylis HA, Krude T. Non-coding stem-bulge RNAs are required for cell proliferation and embryonic development in C. elegans. Journal of cell science. 2015 Jun 1;128(11):2118-29. doi: 10.1242/jcs.166744. PubMed PMID: 25908866; PubMed Central PMCID: PMC4450293.

Kowalski MP, Krude T. Functional roles of non-coding Y RNAs. The international journal of biochemistry & cell biology. 2015 Sep;66:20-9. doi: 10.1016/j.biocel.2015.07.003. PubMed PMID: 26159929; PubMed Central PMCID: PMC4726728.

Krude T, Christov CP, Hyrien O, et al. Y RNA functions at the initiation step of mammalian chromosomal DNA replication. Journal of cell science. 2009 Aug 15;122(Pt 16):2836-45. doi: 10.1242/jcs.047563. PubMed PMID: 19657016.

Leach, C. Domains of type 1 protein phosphatase inhibitor-2 required for nuclear and cytoplasmic localization in response to cell-cell contact. 2002, Journal of Cell Science, 115(19), 3739–3745. doi:10.1242/jcs.00052

Lerner MR, Boyle JA, Hardin JA, et al. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science. 1981 Jan 23;211(4480):400-2. PubMed PMID: 6164096.

Lima Neto, Q. A., Duarte Junior, F. F., Bueno, P. S. A., Seixas, F. A. V., Kowalski, M. P., Kheir, E. Krude, T. & Fernandez, M. A. Structural and functional analysis of four non-coding Y RNAs from Chinese hamster cells: identification, molecular dynamics simulations, and DNA replication initiation assays. *BMC Molecular Biol*, 2016, 17:1-9

Maraia, R., Sakulich, A. L., Brinkmann, E. & Green, E. D. Gene encoding human Ro-associated autoantigen Y5 RNA. *Nucleic Acids Res*, 1996, 24, 3552-9.

Matthews, B. B., dos Santos, G., Crosby, M. A., Emmert, D. B., St. Pierre, S. E., … Gramates, L. S. 2015, Gene Model Annotations for Drosophila melanogaster: Impact of High-Throughput Data G3: Genes Genomes Genetics, 5(8), 1721–1736. doi:10.1534/g3.115.018929

Mosig A, Guofeng M, Stadler BM, et al. Evolution of the vertebrate Y RNA cluster. Theory in biosciences = Theorie in den Biowissenschaften. 2007 Aug;126(1):9-14. doi: 10.1007/s12064-007-0003-y. PubMed PMID: 18087752.

Nolte-'t Hoen, E.N., Buermans, H.P., Waasdorp, M., Stoorvogel, W., Wauben, M.H., 'tHoen, P.A., Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. 2012, Nucleic Acids Res. 40, 9272–9285.

Perreault J, Perreault JP, Boire G. Ro-associated Y RNAs in metazoans: evolution and diversification. Molecular biology and evolution. 2007 Aug;24(8):1678-89. doi: 10.1093/molbev/msm084. PubMed PMID: 17470436.

Ramos-Onsins, S. and Aguadé, M. Molecular evolution of the Cecropin multigene family in Drosophila. functional genes vs. pseudogenes. 1998, Genetics, Sep, 150(1): 157–171.

Samakovlis, C., Kimbrell, D.A., Kylsten, P., Engstrom, A. and Hultmark, D. The immune response in Drosophila: pattern of cecropin expression and biological activity. 1990, The EMBO Journal vol.9 no.9 pp.2969 - 2976.

Sato, K., & Touhara, K. Insect Olfaction: Receptors, Signal Transduction, and Behavior. 2008, Chemosensory Systems in Mammals, Fishes, and Insects, 203–220. doi:10.1007/400_2008_10

Schisa JA, Pitt JN, Priess JR. Analysis of RNA associated with P granules in germ cells of C. elegans adults. Development (Cambridge, England). 2001 Apr;128(8):1287-1298.

Teunissen, S. W., Kruithof, M. J., Farris, A. D., Harley, J. B., Venrooij, W. J. & Pruijn, G. J. Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. *Nucleic Acids Res*, 2000, 28, 610-9.

Villén, J., Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nature Protocols, 2008, 3, 10, 1630-38.

Wang I, Kowalski MP, Langley AR, et al. Nucleotide contributions to the structural integrity and DNA replication initiation activity of noncoding y RNA. Biochemistry. 2014 Sep 23;53(37):5848-63. doi: 10.1021/bi500470b. PubMed PMID: 25151917.
Yue, L., Spradling, A.C. hu-li tai shao, a gene required for ring canal formation during Drosophila oogenesis, encodes a homolog of adducin. 1992. Genes Dev. 6: 2443-2454.
Zhang, A. T., Langley, A. R., Christov, C. P., Kheir, E., Shafee, T., Gardiner, T. J. & Krude, T. Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *J Cell Sci*, 2011, 124, 2058-69.

Supplementary materials

Stem-bulge RNAs from insects and their interaction with chromatin proteins

Francisco Ferreira Duarte Junior[a]*, José Renato Pattaro Júnior[b], Eduardo Jorge Pilau[c], Quirino Alves de Lima Neto[a], Letícia Aparecida de Oliveira[a], Fabiana dos Santos Rando[d], Daniel Caligari[a], Daniele Zanzarin[a], Flavio Augusto Vicente Seixas[b] and Maria Aparecida Fernandez[a]*

[a] Departamento de Biotecnologia, Genética e Biologia Celular, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil; [b] Departamento de Tecnologia, Universidade Estadual de Maringá, campus Umuarama, Av. Ângelo Moreira da Fonseca, 1800, Umuarama, 87506-370, Paraná, Brazil; [c] Departamento de Química, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil; [d] Center for Molecular, Structural and Functional Biology - CBM/COMCAP, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, 87020-900, Paraná, Brazil.

***Corresponding author:** Maria Aparecida Fernandez, mafernandez@uem.br;
Francisco Ferreira Duarte Junior, juniorf.duarte@hotmail.com;

Table S1: Proteins bound to non-coding RNAs and identified by Mass Spectroscopy.					
sbRNA Bound	**Protein ID**	**Protein Name**	**Protein Average Mass**	**Protein Cover (%)**	**Main Functions**
Dm1	P81688	Cecropin-A1 *D. sechellia*	67.568.872	87.30	- Antibacterial humoral response; - Innate immune response.
Dm1	P81685	Cecropin-A1 *D. mauritiana*	67.709.575	42.86	- Antibacterial humoral response; - Innate immune response.
Dm1	B4P7W4	Uncharacterized protein	87.125.052	54.93	- Germ-line cyst formation; - Negative regulation of transcription (DNA-templated); - Pole cell development; - Pole cell migration; - Positive regulation of heterochromatin assembly.
Dm2	Q4QPY5	IP05681p	148.490.155	63.28	- Protein phosphatase inhibitor; - Regulation of signal transduction.
Dm2	Q9VKK9	GEO12070p1	147.349.116	63.78	- Protein phosphatase inhibitor; - Regulation of signal transduction.
Dm2	A0A1W4VAI5	Trichohyalin	572.175.352	21.40	- Structural constituent of cuticle.
Dm2	Q02645	Protein hu-li tai shao "too little nursing"	1.288.518.459	7.09	- Actin filament binding; - adult somatic muscle development; - oogenesis; - Ovarian and testicular fusome organization; - Sarcomere organization.