

**Análise funcional do papel da enzima DNA metiltransferase 2  
(DNMT2) no desenvolvimento e resposta à estresses e  
identificação e caracterização de fragmentos derivados de  
tRNA (tRFs) em *Arabidopsis thaliana*.**

**Cristiane S Alves**

UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"

INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

**Análise funcional do papel da enzima DNA metiltransferase  
2 (DNMT2) no desenvolvimento e resposta à estresses e  
identificação e caracterização de fragmentos derivados de  
tRNA (tRFs) em *Arabidopsis thaliana*.**

**Cristiane de Santis Alves Rosa**

**Orientador: Prof. Dr. Fabio Tebaldi Silveira Nogueira**

Tese apresentada ao Instituto de Biociências, *Campus* de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Rosa, Cristiane de Santis Alves.

Análise funcional do papel da enzima DNA metiltransferase2 (DNMT2) no desenvolvimento e resposta à estresses e identificação e caracterização de fragmentos derivados de tRNA (tRFs) em *Arabidopsis thaliana*. / Cristiane de Santis Alves Rosa. - Botucatu, 2015

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu  
Orientador: Fabio Tebaldi Silveira Nogueira  
Capes: 20203004

1. Plantas - Desenvolvimento. 2. Metilação de DNA. 3. Enzimas - Análise. 4. Acido ribonucleico. 5. Genética vegetal.

Palavras-chave: *Arabidopsis thaliana*; DNMT2; tRF.

*Dedico este trabalho às minhas sobrinhas...  
Que um dia elas possam entender minhas ausências...*

...“Citei, com efeito, noutra obra, uma extensa lista de plantas que os jardineiros chamam plantas loucas, isto é, plantas nas quais se vê surgir de repente um rebento apresentando qualquer caráter novo, e por vezes diferente por completo dos outros rebentos da mesma planta. Estas variações dos gomos, se pode empregar-se esta expressão, podem propagar-se por seu turno por enxerto ou mergulhia, etc, ou algumas vezes mesmo por sementeira. Tais variações se reproduzem raras vezes no estado selvagem; são, porém, bastante frequentes nas plantas cultivadas. Podemos concluir, pois, que a natureza do organismo desempenha o papel principal na produção da forma particular de cada variação, e que a natureza das condições lhe está subordinada; com efeito, vemos muitas vezes na mesma árvore, submetida a condições uniformes, um só gomo, entre milhares de outros produzidos anualmente, apresentar de improviso caracteres novos; vemos, demais, renovos pertencendo a árvores distintas, colocadas em condições diferentes, produzirem quase a mesma variedade - rebentos de pessegueiros, por exemplo, produzirem pêssegos vermelhos, e rebentos de roseira comum produzirem rosas de musgo.

A natureza das condições não tem, pois, talvez mais importância neste caso do que a natureza da faísca, comunicando o fogo a uma massa de combustível, para determinar a natureza da chama.”...

Charles Darwin - A Origem das Espécies

## *Agradecimentos*

*À minha família, por todo o apoio, dedicação e amor.*

*Ao meu marido, pela paciência, apoio e amor.*

*Aos colegas e funcionários do IBB, sempre dispostos à ajudar.*

*Aos colegas e amigos do Departamento de Genética, sempre companheiros e prestativos.*

*Ao meu orientador, pela confiança depositada em mim e por todo apoio durante a realização deste trabalho.*

*Aos colegas e amigos do laboratório, por todo o companheirismo, cumplicidade e discussões científicas e técnicas. Sem vocês este trabalho não teria chegado ao fim!*

*Aos colegas da Unicamp, por toda ajuda prestada.*

*Aos funcionários do Huntsman Cancer Institute e aos colegas e amigos do 'Cairns Lab', que me receberam com tanto carinho e tanto me ensinaram.*

*Ao funcionários do Cold Spring Harbor Laboratory e do Delbrück, e aos colegas e amigos do 'Timmermans Lab', que abriram as portas do laboratório e de um novo mundo científico. Foram tantos aprendizados...*

*À FAPESP (Processos #2011/19512-6 e #2013/21108-4) e CNPq (Processo # 480628/2012-2), pelo apoio financeiro.*

## Sumário

RESUMO .....	3
ABSTRACT .....	5
INTRODUÇÃO.....	7
REFERÊNCIAS BIBLIOGRÁFICAS.....	10
<b>CAPÍTULO 1.....</b>	<b>12</b>
1-REVISÃO BIBLIOGRÁFICA .....	13
2- OBJETIVOS .....	24
<i>Objetivos Gerais</i> .....	24
<i>Objetivos Específicos</i> .....	24
3. MATERIAL E MÉTODOS .....	25
3.1. <i>Material vegetal e condições de crescimento</i> .....	25
3.2. <i>Extração de DNA, RNA e preparo das amostras</i> .....	25
3.3. <i>Análises de sequências para filogenia</i> .....	25
3.4. <i>Obtenção das plantas transgênicas</i> .....	26
3.5. <i>Análise da funcionalidade da enzima AtDNMT2 via metilação da C38 do tRNA Asp<sup>GTC</sup></i> .....	27
3.6. <i>Localização celular da AtDNMT2 em folhas de tabaco</i> .....	28
3.7. <i>Preparação das bibliotecas de RNA tratadas com bissulfito de sódio</i> .....	28
3.8. <i>Análises de Bioinformática para as bibliotecas de RNA tratadas com bissulfito de sódio</i> .....	29
3.9. <i>Análise fenotípica do embrião</i> .....	30
3.10. <i>Análise de expressão via transcrição reversa (RT) PCR quantitativo</i> .....	30
3.11. <i>Ensaio de Northern Blot não radioativo</i> .....	30
3.12. <i>Ensaio de estresse oxidativo</i> .....	30
3.13. <i>Teste para indução de estresse</i> .....	31
3.14. <i>Cruzamentos entre arabidopsis mutantes</i> .....	31
3.15. <i>Análise de germinação</i> .....	32
3.16. <i>Preparação das bibliotecas de mRNA (RNA-seq)</i> .....	32
3.17. <i>Análises de bioinformática das bibliotecas de mRNA (RNA-seq)</i> .....	33
4. RESULTADOS E DISCUSSÃO.....	34
4.1. <i>Análises in silico do gene AtDNMT2</i> .....	34
4.2. <i>Localização subcelular da enzima AtDNMT2</i> .....	38
4.3. <i>Caracterização molecular da planta mutante Atdnmt2</i> .....	40
4.4. <i>Identificação de novos alvos para AtDNMT2</i> .....	42

4.5. Análise fenotípica de plântulas e embrião de arabidopsis Col-0 e <i>Atdnmt2</i> .....	43
4.6. Avaliação da expressão do gene <i>AtDNMT2</i> e sua possível função sob estresse oxidativo .....	44
4.7. Alteração da expressão do gene <i>AtDNMT2</i> em resposta a outros estresses abióticos .....	48
4.8. <i>DNMT2</i> e a família <i>HD2</i> .....	52
4.9. Análises das taxas de germinação nas plantas Col-0, <i>dnmt2</i> e dos duplo mutantes <i>HDA</i> e <i>DHA</i> em diferentes concentrações de sacarose .....	53
4.10. Avaliação de alterações transcricionais na planta mutante <i>Atdnmt2</i> .....	63
5. CONCLUSÕES.....	69
6. REFERÊNCIAS BIBLIOGRÁFICAS .....	70
<b>CAPÍTULO 2.....</b>	<b>81</b>
INTRODUCTION.....	83
RESULTS AND DISCUSSION .....	85
<i>Plant tRFs do not correlated with their parental tRNAs in terms of copies and abundance and are differentially expressed in tissues/organs</i> .....	85
<i>Plant tRFs accumulate differentially under non-physiological conditions</i> .....	87
<i>Biogenesis of most plant tRFs seems to be Dicer-like independent</i> .....	88
<i>AGOs can incorporate tRFs but do not have targets</i> .....	89
CONCLUSIONS .....	91
MATERIAL AND METHODS .....	92
FIGURE LEGENDS.....	93
FIGURES.....	98
REFERENCES.....	105
<b>APÊNDICES.....</b>	<b>109</b>
<i>APÊNDICE 1 - GENOTIPAGEM DOS MUTANTES ATDNMT2 E ATHD2A</i> .....	110
<i>APÊNDICE 2- TABELA DE OLIGONUCLEOTÍDEOS UTILIZADOS:</i> .....	111
<i>APÊNDICE 3 - GENOTIPAGEM DOS DUPLOS MUTANTES ATDNMT2;ATHD2A</i> .....	112
<i>APÊNDICE 4 - DENDOGRAMA DE SIMILARIDADE DA ENZIMA NSUN2</i> .....	113
<i>APÊNDICE 5 – ARTIGO PUBLICADO COMO COLABORADORA</i> .....	114
<i>APÊNDICE 6 – CAPÍTULO DE LIVRO PUBLICADO COMO COLABORADORA</i> .....	124



## Resumo

A metilação do DNA está relacionada à regulação gênica, memória celular, silenciamento de elementos transponíveis, *imprinting* genômico e repressão de pseudo-elementos provenientes de sequências duplicadas. Os padrões de metilação são estabelecidos, mantidos e traduzidos em estados funcionais apropriados da maquinaria de metilação do DNA, a qual inclui uma família classificada em três grupos de enzimas do tipo metiltransferases: DNMT1, DNMT3 e DNMT2. A DNA metiltransferase 2 (DNMT2) foi identificada na busca de novos candidatos à uma segunda DNA metiltransferase. Esta enzima não possui função biológica definida, porém, é capaz de metilar tanto DNA quanto RNA, em especial RNA transportadores (tRNAs). A DNMT2 está localizada tanto no núcleo quanto no citoplasma em células humanas, sendo capaz de migrar do núcleo para o citoplasma em resposta a estresses celulares. É provável que a enzima metile o tRNA no citoplasma, possivelmente para protegê-lo contra clivagens em situações de estresse. Quando estas clivagens ocorrem de forma específica, pequenos fragmentos de RNA são gerados (denominados tRFs), fato observado em diversas espécies, incluindo *Arabidopsis thaliana*. Aparentemente, estes fragmentos de RNA fazem parte de uma nova via de interferência por RNA (RNAi). Contudo, seu papel biológico ainda não foi definido. O objetivo deste trabalho foi determinar a função da enzima DNMT2 de plantas durante o desenvolvimento e em resposta a estresses, além de estabelecer seu possível papel na proteção de tRNAs. Até o momento, foi demonstrado que a enzima AtDNMT2 possui localização, tanto nuclear quanto citoplasmática e também pode ser visualizada em estruturas que aparentam ser citoesqueletos. Foi possível determinar que AtDNMT2 não atua na proteção do tRNA Asp<sup>GTC</sup> durante estresse oxidativo, porém é positivamente regulada durante diferentes tipos de estresse. A planta mutante *dnmt2* não possui fenótipo característico, fato que dificulta o entendimento da função biológica do gene, levando à aprofundar os estudos moleculares a fim de revelar outros possíveis alvos e interações moleculares. Neste contexto, o presente trabalho revelou dois novos alvos (tRNA Gly<sup>GCC</sup> e tRNA Gly<sup>CCC</sup>) para a enzima AtDNMT2 e possíveis vias de atuação gênica para *AtDNMT2*, assim como revelou sua interação com o gene *AtHD2A* e possível papel desses genes em vias epigenéticas de controle da germinação e/ou crescimento inicial de plântulas de *Arabidopsis thaliana*. Dados gerados a partir da confecção de bibliotecas de RNA mensageiro, corroboram os dados gerados previamente para a localização nuclear, incluindo o citoesqueleto, os dados de estresse e condições não

fisiológicas, e corrobora a suspeita de que, de fato, *AtDNMT2* esteja envolvido com o atraso na germinação e/ou crescimento das sementes provenientes do cruzamento dos duplos mutantes. Em relação aos tRFs, foi possível determinar que sua abundância não correlaciona com o número de cópias dos tRNAs parentais e que, em espécies vegetais evolutivamente distintas, o fragmento correspondente à extremidade 5' do tRNA (5'tRFs) é o mais abundante. Análises *in silico* revelaram que os tRFs, em sua maioria 5'tRFs, acumulam-se de maneira distinta nos diferentes tecidos de *Arabidopsis*, dado confirmado por experimentos *in vivo*. Posteriormente, verificou-se que a biogênese dos tRFs avaliados no decorrer deste trabalho, tanto 3' quanto 5', é independente de enzimas Dicer-like, apesar de haver incorporação de tRFs nas diferentes Argonautas de *Arabidopsis*, como observado em análises *in silico*.

Palavras-chave: DNMT2, tRNA, tRF, *Arabidopsis thaliana*.

## Abstract

DNA methylation is associated with genetic regulation, cell memory, silencing of transposable elements, genomic imprinting and repression of pseudo-elements coming from duplicate sequences. Methylation patterns are established, kept and translated via an appropriate functional DNA methylation machinery, which includes a family of proteins classified into three methyltransferase enzyme groups: DNMT1, DNMT3 e DNMT2. DNA methyltransferase 2 (DNMT2) was first identified by searching for novel DNA methyltransferase candidates. DNMT2 is highly conserved in different kingdoms and does not have a biological function well defined so far; however, it has been shown that DNMT2 can methylate both DNA and RNA in animal cells, most specifically transfer RNA (tRNA). In human cells, DNMT2 is localized both in the nucleus and in the cytoplasm, being capable to migrate from nucleus to cytoplasm under stress conditions. In the cytoplasm, DNMT2 methylates tRNAs, possibly to protect against cleavage events that occur under stress conditions. When these cleavages occur in a specific pattern, small RNA fragment emerges (tRFs). tRFs are found in several species, including *Arabidopsis thaliana*. It seems that these tRNA fragments are part of a new RNAi pathway. However, its biological role has not been reveal yet. The aim of this work is to evaluate the possible role(s) of DNMT2 in plant development and stress response and also establish its possible role in tRNA protection. So far we demonstrated that AtDNMT2 has both nuclear and cytoplasmic cellular localization and can also be visualized in what seen to be the cytoskeleton. We determined that AtDNMT2 does not play role in tRNA Asp<sup>GTC</sup> protection under oxidative stress, though *AtDNMT2* is up regulated in different stresses. The mutant plant *Atdnmt2* does not have obvious phenotype, what makes harder to understand its biological role, leading us to deeper molecular studies. In this context, the present work reveals two new AtDNMT2 targets (tRNA Gly<sup>GCC</sup> and tRNA Gly<sup>CCC</sup>) and the *AtDNMT2* interaction with *AtHD2A* and both genes could be playing a possible role in controlling germination and/or growing at epigenetics pathways from *Arabidopsis thaliana* seeds in early steps from development. Data from RNA-seq reinforce our previous data regarding cell localization including cytoskeleton, data from stress and non-physiological conditions and also agree with our hypothesis that indeed AtDNMT2 is involved with the delay in germination/growing of the seeds. Regarding tRFs, it was possible to determine that tRFs abundance did not correlate whit tRNA parent copy number and in evolutionary distinct plant species, 5'tRF is the most

abundant small RNA. *In silico* analyses reveal that tRFs, mostly 5'tRF, accumulate differently in arabidopsis tissues, confirmed by *in vivo* experiments. After that, we verified that both 3'and 5' tRFs biogenesis are Dicer-like independent, despite the fact that tRFs are incorporate in different arabidopsis Argonauts, as we observed *in silico*.

Key words: DNMT2, tRNA, tRF, *Arabidopsis thaliana*.

## Introdução

A metilação de DNA está envolvida com regulação gênica, memória celular, silenciamento de elementos transponíveis, *imprinting* genômico e repressão de pseudo-elementos provenientes de duplicação de sequências. Padrões de metilação são estabelecidos, mantidos e traduzidos via maquinaria de metilação de DNA funcional e apropriada, o que inclui, em quase todos os eucariotos, a família de proteínas classificada em três grupos de enzimas DNA metiltransferases (DNMTs): DNMT1, DNMT3 e DNMT2.

De forma geral, plantas possuem mais genes codificantes para enzimas DNA metiltransferases, dentre elas podemos destacar a planta *Arabidopsis thaliana*, com dez possíveis genes distribuídos em quatro famílias: 1) DNA metiltransferase (MET): uma CpG DNA metiltransferase de manutenção similar a DNMT1; 2) Cromometiltransferase (CMT): uma CpNpG cromometilase específica de plantas, na qual falta a extensão N-terminal; 3) Metiltransferase com Domínio Rearranjado (DRM): *de novo* DNA metiltransferase específica de plantas, que possui um arranjo diferente dos motivos catalíticos conservados; 4) DNMT2: uma enigmática metiltransferase, porém, altamente conservada (Vanyushin e Ashapkin, 2009).

A DNA metiltransferase 2 (DNMT2) foi identificada na busca de novas candidatas à DNA metiltransferases e sua classificação como DNA metiltransferase se deu graças à um motivo catalítico DNA metiltransferase, altamente conservado, presente na molécula (Yoder e Bestor, 1998) e, posteriormente, foi demonstrado que DNMT2 é capaz de se ligar ao DNA, *in vivo*, e essa ligação é resistente à denaturação (Dong et al., 2001). A enzima DNMT2 deve desempenhar papel biológico relevante, pois é uma molécula antiga que está presente desde LECA (último ancestral comum eucarioto) (Jurkowski e Jeltsch, 2011) e, além disso, é conservado em diversas espécies, tais como humano, camundongo, mosca, peixe, lagarta, plantas, levedura, entre outros.

O papel de DNMT2 na regulação de elementos transponíveis ainda permanece em debate, entretanto parece desempenhar importante papel no silenciamento de retrotransposons de *Dictyotellium* e *Drosophila* (Kuhlmann et al., 2005; Phalke et al., 2009). Contudo, existem evidências da ausência de padrões de metilação de DNA em *Drosophila* (Raddatz et al., 2013), o que acende o debate acerca padrões de metilação e técnicas para sua detecção.

Outra informação interessante sobre a enigmática DNMT2 é que cerca de 6%

dos genes de *Aedes aegypti* são metilados e 75% desses *loci* possuem pelo menos um transposon, e, a DNMT2 é a única DNA metiltransferase presente nesse organismo (Ye et al., 2013). Entretanto, é possível que DNMT2 tenha duas funções: metilação de DNA e tRNA (Jeltsch et al., 2006). Enquanto que a metilação de DNA não foi demonstrada em todos os organismos, a metilação da C38 de alguns tRNAs já foi extensivamente demonstrada em diversos organismos, como rato, *Drosophila*, *Arabidopsis thaliana*, bactérias entre outros (Goll et al., 2006; Shanmugam et al., 2014).

Até o momento, apenas alguns trabalhos são dedicados à desvendar o papel da DNMT2 em plantas, dentre eles, destaca-se a publicação de Song e colaboradores (2010), a qual evidencia a interação entre as enzimas AtDNMT2 e AtHD2A no núcleo da célula, indicando que AtDNMT2 pode desempenhar papel na metilação do DNA em plantas. A família histona deacetilase 2 (HD2) é encontrada apenas no reino vegetal, e entre seus membros, a enzima AtHD2A é capaz de mediar repressão transcricional, além de já ter sido correlacionada com o controle da via de dorsoventralidade foliar, provavelmente atuando juntamente com os fatores de transcrição *AS1* e *AS2* (*Asymmetric leaves 1* e *2*) (Ueno et al., 2007), podendo também estar correlacionada aos estádios iniciais do desenvolvimento somático embrionário (Zhou et al., 2004), início do crescimento de plântulas e ainda, na regulação da germinação de sementes (Colville et al., 2011).

A DNMT2 também parece desempenhar papel de proteção de tRNAs em condições de estresse, como demonstrado em *Drosophila* durante estresse oxidativo e altas temperaturas e a clivagem das moléculas de tRNA, é protegida pela metilação da C38 dos tRNAs Asp<sup>GTC</sup>, Val<sup>AAC</sup> e Gly<sup>GCC</sup> (Schaefer et al., 2010). DNMT2 também é capaz de ser transportada do núcleo para o citoplasma em situação de estresse oxidativo em células *HeLa* (Thiagarajan et al., 2011) e em *S. pombe*, a metilação de tRNA pela homóloga de DNMT2 é responsável pela sinalização de nutrientes (Becker et al., 2012). Além disso, condições de estresse podem induzir a liberação de fragmentos que correspondem à metade do tRNA. Tal clivagem ocorre no *loop* do anticódon, tanto em animais como em *arabidopsis* (Thompson et al., 2008). Em células de *Drosophila*, Schaefer e colaboradores (2010) demonstraram que DmDNMT2 é capaz de proteger o tRNA contra clivagens induzidas por estresse oxidativo e térmico.

Os avanços na área de sequenciamento em larga escala dos últimos anos

tornaram possível a identificação de pequenos RNAs (sRNAs) provenientes de diversos RNAs não codantes, incluindo tRNAs. Buscando identificar novas populações de pequenos RNAs, dois grupos realizaram pirosequenciamento em larga escala utilizando células *HeLa* e células de carcinoma de próstata humana. Os *reads* avaliados revelaram sequências provenientes de tRNA que variaram entre 18 e 22 nucleotídeos. Tais RNAs foram chamados de fragmentos derivados de tRNA (tRFs) e análises detalhadas sugerem que esses fragmentos são resultados de processamento específico, e não de degradação aleatória (Cole et al., 2009; Lee et al., 2009).

A linhagem de tRF que é derivada da extremidade 5' do tRNA maduro é denominado 5'tRF e a extremidade 3' proveniente do *loop*-D da estrutura do tRNA, 3'tRF. Em humanos, o processo de clivagem ocorre na base imediatamente após o motivo GG posicionado nas bases 17 e 18 (Cole et al., 2009). O envolvimento da enzima Dicer no processamento de alguns tRFs foi evidenciado em células *HeLa*, favorecendo a teoria de que estes fragmentos de RNAs realmente são frutos de um processamento específico e fazem parte de uma nova classe de pequenos RNAs (Cole et al., 2009).

Em arábido, estes tRFs também foram encontrados em condições de estresse nutricional. As sequências de 19 nt clivadas da extremidade 5' do Gly-tRNA<sup>TCC</sup> e do Asp-tRNA<sup>GTC</sup> representam mais de 80% dos pequenos RNAs derivados de tRNA provenientes de tecidos radiculares. Estes dados revelam a existência de um padrão de expressão temporal e espacial de pequenos RNAs derivados de tRNAs com clivagem específica advindos de moléculas de tRNA, corroborando com as observações realizadas em células *HeLa*, descartando a hipótese de degradação aleatória. Estes dados podem revelar a existência de uma nova via de processamento de pequenos RNAs (Hsieh et al., 2009).

Existem alguns fatos e muitas incertezas a respeito da enzima DNMT2 e dos RNAs não codantes denominados tRFs. Pouco se sabe sobre suas funções biológicas e suas interações com outras moléculas, principalmente no reino vegetal, tanto acerca da DNMT2 quanto dos tRFs. O presente trabalho teve como objetivos investigar possível papel da DNMT2 no desenvolvimento e na resposta à estresses, e também o processamento, comportamento e resposta à estresses dos tRFs, fazendo uso da planta modelo *Arabidopsis thaliana*. Com isso, o presente documento foi dividido em dois capítulos: o primeiro trata de um possível papel da enzima DNMT2 no desenvolvimento inicial de arábido e resposta à estresses; o segundo discorre sobre

possíveis vias de processamento de alguns tRFs e sua resposta à estresse.

## Referências Bibliográficas

- ALBALAT, R. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus *Branchiostoma floridae*. *Development Genes and Evolution*, v. 218, n. 11-12, p. 691-701, 2008.
- ALONSO, J. M. et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, v. 301, n. 5633, p. 653-657, 2003.
- BECKER, M. et al. Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Research*, v. 40, n. 22, p. 11648-11658, 2012.
- COLE, C. et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA*, v. 15, n. 12, p. 2147-2160, 2009.
- COLVILLE, A. et al. Role of HD2 genes in seed germination and early seedling growth in *Arabidopsis*. *Plant Cell Reports*, v. 30, n. 10, p. 1969-1979, 2011.
- DONG, A. et al. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Research*, v. 29, n. 2, p. 439-448, 2001.
- GOLL, M. G. et al. Methylation of tRNA Asp by the DNA methyltransferase homolog Dnmt2. *Science*, v. 311, n. 5759, p. 395-398, 2006.
- HSIEH, L.-C. et al. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiology*, v. 151, n. 4, p. 2120-2132, 2009.
- JELTSCH, A.; NELLEN, W.; LYKO, F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends in Biochemical Sciences*, v. 31, n. 6, p. 306-308, 2006.
- JURKOWSKI, T. P.; JELTSCH, A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS One*, v. 6, n. 11, p. e28104, 2011.
- KUHLMANN, M. et al. Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Research*, v. 33, n. 19, p. 6405-6417, 2005.
- LEE, Y. S. et al. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes and Development*, v. 23, n. 22, p. 2639-2649, 2009.



- PHALKE, S. et al. Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics*, v. 41, n. 6, p. 696-702, 2009.
- RADDATZ, G. et al. Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, n. 21, p. 8627-8631, 2013.
- SCHAEFER, M. et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes and Development*, v. 24, n. 15, p. 1590-1595, 2010.
- SHANMUGAM, R. et al. The Dnmt2 RNA methyltransferase homolog of *Geobacter sulfurreducens* specifically methylates tRNA-Glu. *Nucleic Acids Research*, v. 42, n. 10, p. 6487-6496, 2014.
- THIAGARAJAN, D.; DEV, R. R.; KHOSLA, S. The DNA methyltransferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics*, v. 6, n. 1, p. 103-113, 2011.
- THOMPSON, D. M. et al. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA*, v. 14, p. 2095-2103, 2008.
- UENO, Y. et al. Histone deacetylases and ASYMMETRIC LEAVES2 are involved in the establishment of polarity in leaves of Arabidopsis. *The Plant Cell*, v. 19, n. 2, p. 445-457, 2007.
- VANYUSHIN, B. F.; ASHAPKIN, V. V. *DNA methylation in plants*. New York: Nova Science Publishers, 2009.
- YE, Y. H. et al. Infection with a Virulent Strain of Wolbachia Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito *Aedes aegypti*. *PloS One*, v. 8, n. 6, p. e66482, 2013.
- YODER, J. A.; BESTOR, T. H. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Human Molecular Genetics*, v. 7, n. 2, p. 279-284, 1998.
- ZHOU, C. et al. Expression and function of HD2-type histone deacetylases in Arabidopsis development. *The Plant Journal: for Cell and Molecular Biology*, v. 38, n. 5, p. 715-724, 2004.

## *Capítulo 1*

*DNA metiltransferase 2 (DNMT2) em plantas: possível papel  
no desenvolvimento e na resposta à estresses.*

## **1-Revisão bibliográfica**

### ***Epigenética: o maestro da regulação gênica***

O termo “Epigenética” é atribuído à Conrad Waddington (1939), quem no final da década de 30 utilizou o termo na tentativa de explicar as diferenças que ocorrem durante o desenvolvimento, após a fertilização, as quais dão origem à diferentes órgãos e tecidos, e que não estavam presentes no ovo fertilizado originalmente. O termo sofreu mudanças em seu significado, a fim de refletir um significado mais amplo do que apenas ‘desenvolvimento’, ou seja, mais adequado ao conhecimento atual, tal como: ‘mudanças na função gênica que são herdáveis via mitose e/ou meiose e não envolvem mudanças na sequência do DNA’ (Wu e Morris, 2001). Mais recentemente, Bird (2007) definiu o termo como: ‘adaptação estrutural de regiões cromossômicas para registrar, sinalizar ou perpetuar estado de atividade alterado’ e Reinberg: ‘aqueles processos que asseguram a hereditariedade da variação (-genética) além (-epi) de modificações na sequência de DNA’ (Bonasio et al., 2010). Em suma, eventos epigenéticos são processo normais, reversíveis e herdáveis, podendo ser transmitidos após vários ciclos de replicação celular e englobam os mecanismos: metilação de DNA, modificação de histonas e interferência por RNA (Razin e Riggs, 1980; Bird, 2007; Goldberg et al., 2007; Berger et al., 2009).

Nos dias atuais, já se sabe que a regulação gênica depende de complexas interações entre fatores de transcrição, remodeladores de cromatina, modificações de histonas, histonas variantes, metilação de DNA e moléculas de RNA não codantes. Dentro deste contexto, surge a cromatina: estrutura que contém toda informação biológica e, ao mesmo tempo, é responsável tanto pelo acesso ao genoma quanto pela correta compactação e armazenamento de todo o DNA. A compactação do DNA é necessária para que a informação genética caiba dentro do núcleo da célula, fato que afeta todos os processos baseados no DNA, tais como transcrição, recombinação, replicação e reparo. Contudo, o acesso ao DNA é necessário para que estes mesmos processos sejam iniciados, logo, a cromatina deve ser uma estrutura dinâmica para permitir ou bloquear o acesso ao DNA. Os processos regulatórios do genoma só são possíveis graças à inerente plasticidade da cromatina. Esta plasticidade é caracterizada pela capacidade de afrouxamento e compactação da cromatina que ocorre graças à unidade básica da cromatina: o nucleossomo. O nucleossomo é composto por um octâmero de histonas – duas moléculas H2A, duas H2B, H3 e H4 - que é cercado por

uma região de 147 pares de base (pb) da dupla fita de DNA (Luger et al., 1997).

### ***Modificação de Histonas***

As histonas possuem caudas N-terminais que podem variar de tamanho contendo de 15 a 40 aminoácidos, podendo ser alvos de inúmeras modificações pós-traducionais. As modificações de histonas são realizadas por enzimas específicas capazes de adicionar ou remover tais modificações nos aminoácidos de suas caudas (Smith e Peterson, 2005). As principais modificações são: acetilação, metilação, fosforilação e ubiquitinação (Khorasanizadeh, 2004; Petruk et al., 2012; Abmayr e Worman, 2012). As modificações são capazes de alterar a estrutura e a função da cromatina, sinalizando eventos como ativação da transcrição, silenciamento ou deposição de histonas (Hake et al., 2004; Khorasanizadeh, 2004; Luger et al., 2012).

Modelos sugerem que modificações de histona alteram a estrutura da cromatina, alterando as interações DNA-histona e histona-histona (Hansen e Wolffe, 1998; Wolff e Hayes, 1999). Dentre as modificações amino-terminais de histonas destacam-se a acetilação e a deacetilação. Normalmente o resíduo de lisina (K) é alvo dessas reações que, por serem antagônicas, asseguram a reversibilidade da marca epigenética. A acetilação remove cargas positivas da cauda amino-terminal, o que diminui a afinidade com a fita de DNA e, por conseguinte, facilita a ação de fatores de transcrição e polimerases de RNA, aumentando assim, os níveis transcricionais (Kouzarides, 2000; Yang, 2004; Choudhary et al., 2009).

*Arabidopsis thaliana* possui 12 enzimas tipo Histona Acetilases (HAT) e 18 enzimas tipo Histona Deacetilases (HDAC) (Pandey et al., 2002), número aproximado ao encontrado em mamíferos e maior do que o encontrado em outros eucariotos. As HDACs são conservadas através dos diversos reinos eucariotos; entretanto, existe uma família planta-específica, a HD2, cuja função ainda não está bem elucidada.

A família HD2 foi primeiramente caracterizada em embriões de milho (Brosch et al., 1996). HD2 localiza-se no núcleo, sugerindo que uma das possíveis funções seria a regulação da função e da estrutura da cromatina ribossomal e, provavelmente, ela também desempenha a função de deacetilação de histonas e, até o momento, não é encontrada em nenhum outro reino além do vegetal (Lusser et al., 1997; Pandey et al., 2002; Lawrence et al., 2004; Kim et al., 2014).

A família HD2 também foi identificada em outras espécies vegetais, dentre elas

*Arabidopsis thaliana*, na qual foram encontrados quatro membros – *AtHD2A/HDT1*, *AtHD2B/HDT2*, *AtHD2C/HDT3* e *AtHD2D/HDT4* (Pandey et al., 2002). Existem evidências que sugerem que a família HD2 pode desempenhar um papel crítico no desenvolvimento reprodutivo, durante a embriogênese, desenvolvimento embrionário e desenvolvimento da semente (Wu et al., 2000; Zhou et al., 2004; Demetriou et al., 2009; Colville et al., 2011).

### **Metilação de DNA**

A metilação de DNA foi detectada pela primeira vez no final da década de 40 (Hotchkiss, 1948) e desde então vem sendo amplamente estudada e atualmente, sendo uma das modificações epigenéticas mais conhecidas (Goldberg et al., 2007). A metilação de DNA é associada ao silenciamento gênico em eucariotos; consiste na adição do radical metil ao carbono 5 da base citosina ( $m^5C$ ). Também é uma marca associada à memória celular, silenciamento de elementos transponíveis, *imprinting* genômico e repressão de pseudo-elementos provenientes de sequências duplicadas (Albalat, 2008).

Em plantas, a metilação de DNA ocorre predominantemente em transposons e outros elementos repetitivos (Zhang et al., 2006), no entanto, durante a evolução animal, esses papéis podem variar amplamente, e as funções irão depender de padrões de metilação específicos da espécie considerada (Albalat, 2008).

Em mamíferos, normalmente a metilação de DNA ocorre no dinucleotídeo CpG, dentro do contexto de repressão transcricional (Goll e Bestor, 2005). Estima-se que aproximadamente 70% a 80% dos dinucleotídeos CpGs são metilados no genoma humano (Ehrlich et al., 1982), porém algumas metilações não-CpG podem ser encontradas em células tronco embrionárias (Ramsahoye et al., 2000; Bird, 2002). Em *Arabidopsis thaliana*, aproximadamente 20% de seu genoma é metilado (Zhang et al., 2006), sendo que 24% da metilação é observada no contexto CpG, 6,7% CpHpG e 1,7% CpHpH (onde H pode ser A, T ou C) (Cokus et al., 2008). Em *arabidopsis*, evidências conectam importantes metilações não-CpG que ocorrem durante o desenvolvimento à modificação de histonas e maquinaria de RNAi (Chan et al., 2006).

Os padrões de metilação são estabelecidos, mantidos e traduzidos em estados funcionais apropriados da maquinaria de metilação do DNA, o que, em animais inclui uma família dividida em três classes de enzimas metiltransferases DNA metiltransferases (DNMTs): DNMT1, DNMT2 e DNMT3.

A enzima DNMT1 é responsável pela manutenção do padrão de metilação já estabelecidos (Bestor et al., 1988; Leonhardt et al., 1992; Goll e Bestor 2005; Cheng e Blumenthal, 2008; Chédin et al., 2011); a enzima DNMT3 estabelece os padrões de metilação *de novo* em células germinativas e jovens embriões (Okano et al., 1998); e, por fim, a enigmática enzima DNMT2 que, apesar de demonstrar baixa atividade de metilação de DNA, é capaz de metilar RNAs transportadores (tRNAs) (Goll et al., 2006).

No reino vegetal, a variabilidade de genes referentes à metilação de DNA costuma ser maior do que em animais. Em milho foram encontrados oito genes (Qian et al, 2014), sete em tomateiro (Cao et al., 2014), 10 em arroz (Ahmad et al., 2014) e *Arabidopsis thaliana* conta com 10 possíveis genes, os quais codificam quatro classes: MET, CMT, DRM e DNMT2 (Vanyushin e Ashapkin, 2009).

Em arabidopsis, a classe DNA metiltransferase (MET) é composta por quatro genes, *MET1*, *MET2a*, *MET2b* e *MET3*. A enzima MET1 desempenha a mesma função que a DNMT1: mantém os padrões de metilação já estabelecidos dentro do contexto CpG (Finnegan e Dennis, 1993; Ronemus et al., 1996; Finnegan et al., 2000; Law e Jacobsen, 2010).

A classe CMT (Cromometilase) possui um domínio cromo e não possui a extensão N-terminal longa, característica das outras DNA metiltransferases dos eucariotos superiores. Esta classe parece existir somente no reino vegetal e é composta por três genes *CMT1*, *CMT2* e *CMT3* (Henikoff e Comai, 1998). A enzima CMT3 é específica para a metilação do tipo CpHpG (Bartee et al., 2001; Lindroth et al., 2001). A proteína CMT3 se liga a cauda metilada da histona H3, indicativo de que enzimas de metilação de histonas também possam atuar no recrutamento da CMT3 (Lindroth et al., 2004).

A classe DRM (DRM1 e DRM2) ou Metiltransferases com Domínios Rearranjados é responsável pela metilação *de novo*, principalmente pela enzima DRM2 dentro do contexto CpHpH (Cao et al., 2002; Cao et al., 2003, Chan et al., 2004). Essa classe faz parte da via RdDM (Metilação de DNA direcionada por RNA), em que os transcritos de RNA dependentes de RNA polimerase IV são convertidos em RNA dupla-fita pela enzima RDR2 (RNA dependente de RNA polimerase 2) e processados pela DCL3 (Dicer Like 3) em siRNAs de 24 nucleotídeos (Wassenegger et al., 1994; Jones et al., 2001; Greenberg et al., 2011; Zhang et al., 2013; Zhong et al., 2014).

A classe DNMT2 em plantas, assim como em animais, permanece obscura, mas também é capaz de metilar o tRNA Asp<sup>GTC</sup> (Goll et al., 2006). DNMT2 em *Drosophila*, é responsável pela metilação dos tRNAs Asp<sup>GTC</sup>, Gly<sup>GCC</sup> e Val<sup>AAC</sup> bem como pela proteção de tRNA em situações de estresse (Schaefer et al., 2010).

### ***DNMT2: uma enigmática metiltransferase***

A enzima DNMT2 foi identificada em 1998 por Yoder e Bestor, na busca de novos candidatos a uma segunda DNMT ou METase em humano e camundongo. O domínio regulador comum a outras METases, correspondente à região N-terminal, não é encontrado nesta nova molécula. Proteínas homólogas à DNMT2, presentes em organismos conhecidos por não metilar seus genomas, possuem uma sequência-alvo de reconhecimento específico, incluindo o tripeptídeo invariável Cys-Phe-Thr (CFT). Além disso, experimentos *in vitro* demonstraram que a DNMT2 se liga ao DNA, formando um complexo resistente à denaturação (Dong et al., 2001). A DNMT2 possui todos os resíduos que são utilizados para a catálise da DNA m<sup>5</sup>C conservados, porém, também possui um resíduo de cisteína o qual é fortemente conservado no grupo da DNMT2, mas ausente nos grupos das DNA m<sup>5</sup>C metiltransferases. O motivo CFT (conservado e existente apenas entre as proteínas DNMT2) pode ser responsável pela atividade catalítica da DNMT2 (Jurkowski et al., 2008).

Trabalhos demonstram que o *locus DNMT2* em *Drosophila* gera dois transcritos que se diferenciam na porção N-terminal. Em um dos transcritos, o motivo I, responsável pelo sítio catalítico DNA metiltransferase é eliminado, enquanto o outro transcrito mantém o motivo I. Ambos os transcritos foram detectados em tanto nos estádios de desenvolvimento como em moscas adultas, sugerindo uma função enzimática global através das fases do desenvolvimento (Schaefer *et al.*, 2010).

Em 2003, foi comprovado que a DNMT2 é um subgrupo da família gênica m<sup>5</sup>C (metilação do carbono 5 na base citosina). A proteína foi superexpressa em *Drosophila*, gerando uma hipermetilação anômala no genoma da mosca (Tang *et al.*, 2003). A baixa atividade da enzima DNMT2 *in vitro* pode explicar a falha em detectar sua atividade em estudos anteriores e também pode explicar, em parte, o fenótipo fraco observado após deleção do gene *dnmt2* em camundongos. A DNMT2 também é capaz de modificar o DNA em resíduos CG, semelhante às outras METases identificadas até o momento. É possível que outras METases possam compensar a função do gene *dnmt2* mutado em animais, o que também ajudaria a explicar o fenótipo fraco em

mutantes nocautes. As sequências *consenso* derivadas das reações de metilação *in vitro* sobrepõem o local do alvo de fatores de transcrição que possuem papéis importantes na regulação gênica de diferentes vias de sinalização e interagem com diversas outras proteínas envolvidas na regulação transcricional. Desde que foi demonstrado que a metilação CG inibe a ligação por diferentes fatores de transcrição, a sequência *consenso* sugere que a DNMT2 pode estar envolvida na regulação da ligação do DNA e na transativação por um fator de transcrição. Assim, a DNMT2 possivelmente tem um papel mais especializado *in vivo* do que outras METases conhecidas (Hermann et al., 2003).

A falta de fenótipo óbvio em mutantes para esta enzima intriga ainda mais o mundo científico, pois ainda não foi possível atribuir uma função biológica precisa para a DNMT2. A enzima é conservada em diversas espécies e, segundo análises de bioinformática, é a mais antiga entre as METases (Goll et al., 2006; Jeltsch et al., 2006).

A DNMT2 também pode estar relacionada à longevidade. Em *Drosophila*, foi realizado experimento no qual as moscas foram submetidas ao estresse oxidativo, que está relacionado a diversas doenças degenerativas, ao acúmulo de danos oxidativos e ao envelhecimento. As moscas transgênicas que superexpressam a enzima demonstram maior resistência ao estresse do que o controle (Mund et al., 2004). Em 2006, outro trabalho também demonstrou o aumento da tolerância ao estresse em amebas (Fisher et al., 2006). A DNMT2 também pode estar envolvida com a integridade do genoma, protegendo-o contra elementos transponíveis e protegendo a integridade dos telômeros (Schaefer e Lyko, 2010).

Em organismos como a *Dictyostelium*, acreditava-se não haver metilação de DNA. Contudo, descobriu-se que a DNMT2 é capaz de metilar seu genoma em baixos níveis. No trabalho de Katoh e colaboradores (2006), os autores observaram fenótipos sutis em mutante nulo para o gene *DNMT2*, porém reproduzíveis, na esporulação e na fragmentação do *sorus*, dando indícios de que a enzima é fundamental para a finalização de seu correto desenvolvimento. Estes resultados demonstram consistência com o aumento temporal da metilação durante o desenvolvimento (Katoh et al., 2006). Estudos em peixes *zebrafish* revelaram que a DNMT2 atua em diversas fases do desenvolvimento. Peixes mutantes com expressão nula do gene apresentaram defeitos



no desenvolvimento de órgãos, tais como a retina, o fígado e o cérebro. (Rai et al., 2007).

As DNMTs 1 e 3 estão localizadas no núcleo celular (Leonhardt et al. 1992), enquanto que a DNMT2 está preferencialmente no citoplasma (Goll et al., 2006). Recentemente, foi demonstrado a realocação da DNMT2 do núcleo para o citoplasma, em resposta ao estresse do ambiente celular em células *HeLa* (Thiagarajan et al., 2011). De acordo com estes dados, pode-se dizer que a DNMT2 é tanto uma enzima citoplasmática quanto nuclear. Ela também aparece em abundância durante a divisão celular, sendo que sua localização é altamente dinâmica durante o ciclo celular (Schaefer et al., 2008).

Em outro trabalho recente, foi demonstrado que a DNMT2 de *Arabidopsis* (*AtDNMT2*) também é encontrada no núcleo e interage com o tipo 2 da histona deacetilase (*AtHD2*), família de histona deacetilase única em plantas. A *AtDNMT2* tem a habilidade de reprimir a expressão gênica à nível transcricional, sendo possível o envolvimento da DNMT2 na atividade de deacetilação de histonas e na rede regulatória epigenética de plantas (Song et al., 2010).

Thiagarajan e colaboradores (2011), demonstraram em células *HeLa* transgênicas superexpressando o gene *DNMT2*, que a enzima também está relacionada com letalidade celular. Contudo, até o momento a grande maioria dos trabalhos indica que indivíduos mutantes são férteis e viáveis, não apresentando nenhum fenótipo óbvio (Goll et al., 2006; Jeltsch et al., 2006). Como resultado do experimento de Thiagarajan *et al.* (2011), diversos genes envolvidos na infecção viral foram desregulados sob o aumento da expressão da DNMT2, sugerindo um possível papel da enzima como parte do mecanismo do sistema imune.

Para surpresa do mundo científico, foi averiguado que a enzima DNMT2 tem como substrato tanto DNA quanto RNA transportadores (tRNAs) (Krauss e Reuter, 2011). A metilação em tRNA se dá na posição da citosina 38, localizada logo após o *loop* do anticódon (Goll et al., 2006). Os tRNAs possuem um papel biológico central, atuando como um adaptador entre o RNA mensageiro (mRNA) e os aminoácidos, no qual o anticódon interage com o códon do mRNA, permitindo que um aminoácido anexado na outra extremidade do tRNA se condense com a cadeia de peptídeo em um segundo tRNA, através da ação do ribossomo (Hopper e Phizicky, 2003).

A primeira vez que se observou a degradação de tRNA foi durante estresse nutricional de protozoário, fato que sugeria uma adaptação a esse tipo de estresse. As

regiões clivadas dos tRNAs foram observadas na fase inicial do estresse, e também notou-se que nenhum tRNA degradado aparentava ser produto do mecanismo de controle de qualidade do processamento do tRNA, já que em mutantes para defeitos deste processamento, a quantidade de fragmentos não era alterada, sugerindo uma via distinta da via de degradação de tRNAs aberrantes (Lee e Collins, 2005).

O estresse oxidativo é responsável pela clivagem do tRNA em células de leveduras, plantas e mamíferos, e é uma característica conservada da resposta celular ao estresse oxidativo (Thompson et al., 2008). Outros trabalhos, nos quais foram utilizados *arabidopsis* e linhagens de células humanas, também foram observadas pequenas quantidades de fragmentos de tRNA na ausência de qualquer estresse, fato que sugere que a clivagem do tRNA é ativo constitutivamente em níveis basais. Porém, sua atividade é aumentada na presença de algum tipo de estresse. Este mecanismo pode ser desencadeado a fim de diminuir os transcritos durante estresse celular; entretanto, resultados em levedura e células humanas demonstram que a quantidade de tRNAs maduros não é alterada, mesmo durante o pico de fragmentos de tRNAs. Durante o estresse celular, a ativação de outras nucleases pode ser desencadeada, utilizando outros RNAs - que ainda não foram identificados - como alvos. Estes dados sugerem que a clivagem do tRNA pode ter um papel de transdutor de sinal durante estresse, ou os fragmentos de tRNA podem, potencialmente, interagir com a maquinaria de tradução de uma forma inibitória (Thompson et al., 2008).

Em experimento com peixes *zebrafish*, foi observada a metilação *in vitro* de RNAs de aproximadamente 80 bases pela DNMT2, tamanho consistente com o tRNA<sup>Asp</sup> destes peixes. Estes dados sugerem que a função biológica da DNMT2 provavelmente envolve a metilação citoplasmática do RNA (Rai et al., 2007). Em outro trabalho foi observado o aparecimento de fragmentos de tRNA Asp, tRNA Val e tRNA Gly em *Drosophila* com o alelo mutante *dnmt2* – responsável pela perda dos dois transcritos do *DNMT2* -, após tratamento com bissulfito e posterior sequenciamento, indicando que a DNMT2 protege alguns tRNAs da clivagem em situações de estresse (Schaefer et al., 2010). Em experimentos com células *HeLa*, foi demonstrado que a DNMT2 interage com diversas proteínas envolvidas no processamento de RNA (Thiagarajan et al., 2011).

Em *Entamoeba histolytica*, a falta de glicose leva ao acúmulo de enolase (enzima envolvida na glicólise) no núcleo, levando à formação do complexo DNMT2-enolase,

havendo uma significativa redução na metilação do tRNA Asp no parasita. Estes dados estabelecem a enolase como a primeira proteína que interage com a DNMT2, e traz à tona um papel inesperado de uma enzima glicolítica na modulação da atividade da DNMT2 (Tovy et al., 2010).

Novas descobertas sugerem que a DNMT2 participa de processos epigenéticos hereditários mediados por RNA em camundongos. Ao fazer uso da paramutação *Kit*, a qual apresenta fenótipos de cauda e pés brancos, Kiani e coautores (2013) realizaram uma série de cruzamentos entre camundongos portadores da paramutação com camundongos *dnmt2*, e observaram que a progênie do cruzamento *dnmt2* *-/-* e *Kit* *+/+* não apresenta o fenótipo da cauda e pés brancos. Após uma série de cruzamentos, onde o animal macho não apresenta a enzima DNMT2, foi observado que a progênie não apresentava o fenótipo, levando à conclusão de que DNMT2 é requerida apenas durante a gametogênese paterna ou em estádios iniciais do desenvolvimento embrionário em camundongos (Kiani et al., 2013).

Até o momento, poucos trabalhos remetem a um possível papel da DNMT2 em plantas. Já foi demonstrado que DNMT2 em *Arabidopsis thaliana* (AtDNMT2) é capaz de metilar a C38 do tRNA Asp<sup>GTC</sup> (Goll et al., 2006). Existem evidências de que AtDNMT2 pode participar da metilação de DNA em plantas, já que a interação entre AtDNMT2 e AtHD2A foi demonstrada e DNMTs parecem agir como repressores transcricionais (Song et al., 2010). Como descrito anteriormente, HD2s são histona deacetilases específicas de plantas. AtHD2A é capaz de intermediar a repressão transcricional e já foi demonstrado que é associada ao controle da dorsoventralidade foliar, provavelmente atuando juntamente com os fatores de repressão gênica AS1 e AS2 (Asymmetric leaves1 e 2) (Ueno et al., 2007) e parece estar correlacionada com os estádios iniciais do desenvolvimento embrionário somático (Zhou et al., 2004). Recentemente, Colville e colaboradores (2011) descreveram possível papel para os genes da família HD2 durante o início do crescimento de plântulas, regulando a germinação de sementes.

Baseado nos dados existentes em DNMT2 de diferentes espécies, é possível argumentar que haja uma correlação entre a DNMT2 e a geração de metades de tRNAs, porém a função biológica tanto da enzima quanto das metades de tRNAs permanece desconhecida, fato este que está aguçando a curiosidade de pesquisadores de diversas áreas.

## ***Metilação de RNA***

Já se conhece mais de 100 modificações químicas que podem ocorrer no RNA, incluindo a m<sup>5</sup>C. A metilação de RNA se dá em diversos tipos de RNA, como: RNA mensageiro (mRNA) (Banerjee, 1980; Cowling, 2010), RNA ribossomal (rRNA) (Tscherne et al., 1999; Foster et al., 2003; Andersen et al., 2006), RNA transportador (tRNA) (Goll e Bestor, 2005; Goll et al., 2006; Auxilien et al. 2007; Rai et al., 2007; Grosjean et al. 2008; Jurkowski et al., 2008), RNAs virais (Pichlmair et al., 2006; Daffis et al., 2010), além de RNAs regulatórios em animais e plantas – siRNAs, piRNAs, microRNAs e seus precursores (Luciano et al., 2004; Yu et al., 2005; Yang et al., 2006; Horwich et al., 2007; Saito et al., 2007).

RNA metiltransferases ou RNA-MTases incluem mais de 60 membros, com centenas de homólogos divididos em quatro superfamílias: RFM (Rossmann Fold MTase), SPOUT, SAM-dependent MTases e FAD/NAD-binding protein (Czerwoniec et al., 2009; Motorin e Helm, 2011).

A metilação do tipo m<sup>5</sup>C em tRNA possui funções biológicas de grande importância como, por exemplo, garantir a estabilidade da molécula (Helm, 2006), a regulação fina da tradução – manutenção do frame de leitura (Motorin e Helm, 2011) e possível papel de proteção contra estresses (Schaefer et al., 2010). Especula-se que uma possível função da regulação de padrões epigenéticos herdáveis e em tRNAs a m<sup>5</sup>C mais comum se dá nas posições 48 e 49 (Sprinzl et al., 2005; Motorin et al., 2010).

A diferença no comportamento químico dos nucleotídeos citosina e m<sup>5</sup>C quando expostos ao bissulfito de sódio em pH ácido, forneceu uma valiosa ferramenta para o sequenciamento baseado no perfil de metilação do DNA genômico (Frommer et al., 1992). Todas as citosinas não metiladas, quando tratadas com bissulfito de sódio, são deaminadas e convertidas em uracilas (Us), aparecendo como timinas (Ts) após reação de sequenciamento. As m<sup>5</sup>C são refratárias à deaminação, se mantendo como citosinas (Cs) após a reação de sequenciamento, um indicativo da metilação da citosina. O método de sequenciamento por bissulfito acoplado à técnicas de sequenciamento em larga escala, pode definir padrões de metilação de fitas individuais de DNA, com resolução de base única e tem sido escalonada a fim de definir metilomas de genomas eucariotos inteiros (Meissner et al., 2008; Cokus et al., 2008; Lister et al., 2009).

O uso do bissulfito de sódio em moléculas de RNA foi reportado pela primeira

vez em há mais de quatro décadas, e também foi demonstrado que a deaminação de bases Cs em moléculas de RNA requer a exposição da base, na forma de ácido nucléico de fita simples (Shapiro et al., 1970). A metodologia de bissulfito foi utilizada a fim de determinar a estrutura secundária do RNA ribossomal bacteriano, aproveitando o fato da conversão da C para U apenas ocorrer nas bases expostas e não nas bases C-G ligadas por pontes de hidrogênio, em regiões dupla-fita (Woese et al., 1980). O sequenciamento em larga escala de RNA tratado com bissulfito de sódio também tem sido utilizado para distinguir transcritos senso e anti-senso, pelo fato da conversão por bissulfito alterar o contexto da sequência das fitas senso e anti-senso de modo que elas percam a complementariedade completa, se tornando distinguíveis umas das outras (He et al., 2009).

A técnica de sequenciamento de RNA tratado com bissulfito de sódio foi estabelecida com o intuito de mapear tanto as m<sup>5</sup>Cs conhecidas como as desconhecidas em moléculas de tRNAs e rRNAs (Schaefer et al., 2009) e utilizado com sucesso para determinar dois novos alvos para DNMT2: tRNA Gly e Val (Schaefer et al., 2010).

Com isso, as hipóteses para esse trabalho incluíram: outros tRNAs, além do tRNA Asp<sup>GTC</sup> como alvos para AtDNMT2; proteção de tRNAs pela enzima AtDNMT2; participação de AtDNMT2 em vias epigenéticas do desenvolvimento vegetal e resposta à estresse. Para tal, foram propostos os seguintes objetivos gerais e específicos:

## 2- Objetivos

### Objetivos Gerais

Avaliar o papel da enzima DNMT2 no desenvolvimento vegetal e em resposta a estresses, bem como seu possível papel na proteção de tRNAs na planta modelo *Arabidopsis thaliana*.

### Objetivos Específicos

- Avaliar a localização intracelular da enzima DNMT2.
- Avaliar a função da DNMT2 durante o desenvolvimento vegetal de *Arabidopsis thaliana*
- Avaliar a alteração da expressão do gene *DNMT2* e das metades de tRNA em plantas de *Arabidopsis thaliana* sob estresse oxidativo;
- Avaliar a expressão do gene *DNMT2* em resposta a diferentes tipos de estresse;
- Avaliar a interação genética entre as enzimas DNMT2 e HD2A e seu papel no processo de germinação de sementes de *Arabidopsis thaliana*.
- Avaliar alterações transcricionais na planta mutante *Atdnmt2* em estádios iniciais do desenvolvimento de *Arabidopsis thaliana*.

### **3. Material e Métodos**

#### **3.1. Material vegetal e condições de crescimento**

A espécie vegetal utilizada como modelo para a realização deste trabalho foi *Arabidopsis thaliana* (ecótipo Columbia-0 - Col-0). Para a desinfecção das sementes, as mesmas foram lavadas por 10 minutos em água sanitária 50% e 2% de Tween® e posteriormente lavadas com água destilada autoclavada de três a cinco vezes. As sementes desinfetadas foram semeadas em placa de petri contendo meio MS (Murashige e Skoog, 1962) modificado sólido, acrescido de 1% de sacarose (meio MS comercial sem vitaminas - PhytoTechnology). As placas foram acondicionadas no escuro a 4°C por 48 horas para quebra da dormência e transferidas para crescimento sob iluminação artificial em fotoperíodo de 16 horas de claro e 8 horas de escuro à aproximadamente 22°C, até o momento da coleta ou transferência para o substrato.

Sementes mutantes de arábidoopsis com inserção de T-DNA adjacente ao exon sete do gene *DNMT2* (mutante *dnmt2*) foram obtidas da coleção Salk (136635C) (Alonso et al., 2003; Goll et al., 2006). Sementes mutantes de arábidoopsis com inserção T-DNA dentro do exon três do gene *HD2A* (mutante *hd2a*) foram obtidas da coleção Gabi (CS348583) (Alonso et al., 2003; Colville et al., 2011). Os dois mutantes foram confirmados por genotipagem de DNA genômico (Apêndice 1). Os oligonucleotídeos utilizados estão descritos no (Apêndice 2).

#### **3.2. Extração de DNA, RNA e preparo das amostras**

A extração de DNA foi realizada segundo protocolo adaptado descrito por Doyle e Doyle (1987). A extração do RNA total foi realizada utilizando Trizol Reagent® (Invitrogen) seguindo as recomendações do fabricante. O RNA foi tratado utilizando o kit comercial Turbo DNA-free™ (Ambion), seguindo as recomendações da fabricante. A síntese do cDNA foi realizada utilizando o kit comercial ImpromII Reverse Transcriptase® (Promega).

#### **3.3. Análises de sequências para filogenia**

As sequências no formato FASTA das proteínas utilizadas para geração de árvores filogenéticas foram obtidas no banco de dados Tair (*The Arabidopsis*

*Information Resource* - <https://www.arabidopsis.org> ) e inseridas no *software* CLC Sequence Viewer 7.6 da empresa Qiagen ( <http://www.clcbio.com/products/clc-sequence-viewer/> ). Primeiro, foi realizado o alinhamento das proteínas selecionadas e posteriormente, a ferramenta para construção de árvores filogenéticas foi empregada, utilizando os parâmetros estabelecidos para o método de máxima verossimilhança utilizando 1000 replicatas *Bootstrap*.

### 3.4. Obtenção das plantas transgênicas

Para a expressão do gene *AtDNMT2* fusionado com *mRFP* sob o controle do promotor CaMV35S, primeiramente, foram realizadas reações de PCR, com o kit comercial Phusion®High-Fidelity DNA Polymerase (NEB), para amplificação do fragmento correspondente ao cDNA da enzima *AtDNMT2*. Os oligonucleotídeos (Apêndice 2) foram desenhados a partir da sequência disponível em bancos de dados (TAIR e GenBank). Os amplicons de 1370pb foram verificados através de eletroforese em gel de agarose 1,5%.

O cDNA do gene *AtDNMT2* foi inserido no vetor pENTR, utilizando o pENTR™/D-TOPO® Cloning Kit (Invitrogen), seguindo as indicações do fabricante. Foram utilizadas células competentes da cepa DH5α de *Escherichia coli* para transformação via choque térmico. Após, as bactérias foram plaqueadas em placa de Pertri contendo meio LB (Luria-Bertani) sólido contendo antibiótico canamicina 50µg/ml. As colônias foram selecionadas através de PCR com os oligonucleotídeos M13.

Posteriormente, foi feita a minipreparação do DNA plasmidial de algumas colônias contendo o fragmento esperado, utilizando o kit comercial Zyppy™ Plasmid Miniprep Kit (Zymo Research), seguindo manual do fabricante. Após a minipreparação, foi realizado outra PCR para confirmação do fragmento e, posteriormente, enviadas para sequenciamento Sanger. As sequências foram avaliadas no *software* CLC Sequence Viewer 7.6 e submetidas à pesquisa junto ao banco de dados do NCBI, website (<http://www.ncbi.nlm.nih.gov/blast>) utilizando a ferramenta BLAST.

Após, foram realizadas duas reações do tipo LR, com o kit comercial Gateway® LR Clonase® enzyme mix (Invitrogen) a fim de recombinar o cDNA do gene *AtDNMT2* nos vetores binários pGWB554 -P35S:DNMT2:mRFP- e pGWB555 -



P35S:mRFP:DNMT2. Os vetores utilizados foram desenvolvidos e gentilmente cedidos por Nakagawa et al., 2007. Após, foi realizada a minipreparação do DNA plasmidial de colônias contendo o fragmente esperado, utilizando o kit comercial Zyppy™ Plasmid Miniprep Kit (Zymo Research), seguindo instruções do fabricante. Após a minipreparação, o DNA plasmidial foi enviado para sequenciamento Sanger. As sequências foram avaliadas no *software* CLC Sequence Viewer 7.6 para verificação da orientação e correta sequência do cDNA do gene *AtDNMT2*. O DNA plasmidial de ambas as construções foram transformados em células competentes de *Agrobacterium tumefaciens* cepa GV3101. As colônias contendo os vetor binário foram selecionadas através de antibióticos e reações de PCR e, posteriormente, transformadas em *Arabidopsis thaliana*.

As plantas transgênicas foram obtidas segundo o protocolo de transformação de arábido *Floral Dip* adaptado por Weigel e Glazebrook, 2002. A seleção das plantas transformadas foi feita em placa de petri contendo meio MS (comercial) sólido e antibióticos de seleção.

### **3.5. Análise da funcionalidade da enzima AtDNMT2 via metilação da C38 do tRNA Asp<sup>GTC</sup>**

Plântulas Col-0 e *dnmt2* com dois dias de idade foram utilizadas a fim de testar a funcionalidade da enzima DNMT2. Após a extração de RNA, o DNA complementar foi obtido através do kit comercial SuperScript® III Reverse Transcriptase (Invitrogen), seguindo as recomendações do fabricante e um oligonucleotídeo reverso complementar ao tRNA Asp<sup>GTC</sup> tratado com bissulfito de sódio foi utilizado na reação (Apêndice 1). Para a reação de PCR e logo após, Nested-PCR, foram utilizados três oligonucleotídeos, sendo que o reverso foi o mesmo utilizado previamente para a confecção da fita complementar e dois oligos *forward* foram utilizados (Apêndice 2), um para cada reação.

Após a confirmação da correta amplificação do tRNA em gel de agarose 2%, o produto de PCR foi clonado no vetor pCR 2.1, utilizando o TOPO®-TA Cloning® Kit for sequencing (Invitrogen), seguindo as instruções do fabricante e transformadas em células competentes da cepa DH5α de *Escherichia coli* via choque térmico. Para cada genótipo – Col-0 e *dnmt2* – pelo menos 20 colônias das bactérias contendo o vetor clonado foram selecionadas da placa contendo meio LB (Luria-Bertani) contendo 100

µg/ml ampicilina e colocadas para crescer em meio LB líquido contendo 100 µg/ml ampicilina a fim de isolar o vetor contendo o tRNA Asp<sup>GTC</sup>.

Para confirmar a presença dos plasmídeos clonados, a enzima EcoRI (NEB) foi utilizada, seguindo o protocolo padrão, após isolamento do DNA plasmidial. Vinte plasmídeos positivos de cada genótipo foram enviados para sequenciamento Sanger e os resultados foram comparados com a sequência do tRNA Asp<sup>GTC</sup> maduro através do *software* CLC Sequence Viewer 7.6 da empresa Qiagen (<http://www.clcbio.com/products/clc-sequence-viewer/>).

### 3.6. Localização celular da AtDNMT2 em folhas de tabaco

Folhas de tabaco (*Nicotiana tabacum*) de plantas de três e/ou quatro semanas de idade foram infiltradas com duas soluções feitas com MES hidratado, MgCl<sub>2</sub> e Acetosiringona, contendo *Agrobacterium tumefaciens* cepa GV3101. Cada uma das agrobactérias continham um cassete de expressão: p35S:mRFP:DNMT2 e p35S:DNMT2:mRFP. As folhas de tabaco infiltradas foram visualizadas sob microscópio de fluorescência após 26 e 48 horas do procedimento. Folhas foram coletadas e armazenadas à -80°C para análises transcricionais posteriores.

### 3.7. Preparação das bibliotecas de RNA tratadas com bissulfito de sódio

As plântulas dos genótipos Col-0 e *dnmt2* foram coletadas com 48 horas de germinação e tratadas de acordo com protocolo publicado por Khoddami e Cairns (2014). Após a extração do RNA total, o RNA ribossomal foi removido com o auxílio do kit comercial RiboZero – Plant seed/roots (Epicentre). As bibliotecas de RNA foram preparadas utilizando o kit comercial *Illumina TruSeq Small RNA Sample Prep Kit*, de acordo com instruções do fabricante. O sequenciamento foi realizado no *High Throughput Genomics Facility*, localizado no *Huntsman Cancer Institute* da Universidade de Utah (EUA).

O RNA total foi denaturado e adaptadores 3' e 5' foram ligados às moléculas de RNA. O RNA ligado ao adaptador foi retrotranscrito utilizando oligonucleotídeos complementares à sequência do adaptador 3'. Para tanto, utilizou-se o kit comercial Superscript II Reverse Transcriptase (Invitrogen). Os adaptadores foram estendidos e os *index tags* foram adicionados ao cDNA via PCR (11 ciclos). A concentração da

biblioteca foi estimada utilizando-se o kit comercial Invitrogen *Qubit dsDNA HS Assay* e uma alíquota da biblioteca foi utilizada a fim de verificar a distribuição do tamanho da mesma. Nesta fase, foi empregado o sistema Agilent 2200 *Tape Station*, o qual utiliza o *High Sensitivity DIK Assay*. A concentração das bibliotecas foi ajustada para aproximadamente 5 nM e PCR quantitativa foi realizada a fim de calcular a molaridade dos adaptadores ligados às moléculas das bibliotecas, utilizando-se o kit comercial Kapa Biosystems (Kapa Library Quant Kit), seguindo as instruções do fabricante. A concentração das bibliotecas foi então ajustada para sua preparação para o sequenciamento Illumina.

### **3.8. Análises de Bioinformática para as bibliotecas de RNA tratadas com bissulfito de sódio**

O pipeline e os pacotes utilizados neste trabalho foram desenvolvidos pelo Dr. Vahid Khoddami e pela equipe de Bioinformática do Huntsman Cancer Institute (dados não publicados).

O *index* do transcriptoma foi feito utilizando-se o *software* comercial *Novoalign* (<http://www.novocraft.com/products/novoalign/>) e incluiu eventos de *splicing* e foi utilizada uma opção disponível no *software* capaz de gerar um *index* do transcriptoma virtualmente tratado com bissulfito de sódio, convertendo as Cs em Ts. Esse *index* foi utilizado para realizar o alinhamento dos arquivos Fastq. Todos os *reads* iguais ou maiores que 18 pb foram utilizados para as análises. Para processar os *outputs* dos alinhamentos e lidar com as junções dos *splicings*, foi utilizado o pacote USeq da Universidade de Utah (<http://sourceforge.net/projects/useq/>).

Para as análises de metilação, *software* customizado desenvolvido pelo grupo de Bioinformática da Universidade de Utah foi utilizado. Tal *software* (ainda não publicado) conta o número de Cs em cada posição apenas se pelo menos uma C for observada naquele ponto em pelo menos um *read* mapeado naquela região. O *index* de todas as Cs mantidas (Cs potencialmente metiladas) foi utilizado para montar a tabela final de trabalho. Outro programa customizado (ainda não publicado) também foi utilizado para calcular o *p-value* de cada local em particular. A tabela final de trabalho foi anotada a fim de identificar os genes candidatos com diferentes padrões de metilação de RNAs.

### **3.9. Análise fenotípica do embrião**

Silíquas de plantas de arábido *arabidopsis* contendo os embriões foram coletadas em diferentes estádios de desenvolvimento. Os embriões foram visualizados e fotografados sob microscópio óptico. Os embriões foram preparados de acordo com protocolo de Berleth e Jurgens (1993).

### **3.10. Análise de expressão via transcrição reversa (RT) PCR quantitativo**

O kit comercial GoTaq qPCR Master Mix (Promega) foi utilizado seguindo o protocolo padrão, os oligonucleotídeos utilizados estão descritos no Apêndice 1. Todas as reações foram realizadas em triplicatas biológicas e duplicatas técnicas. O procedimento foi realizado no equipamento Step One Plus Real Time® (Applied Biosystems). As análises foram realizadas seguindo o método  $2^{-\Delta\Delta CT}$  (Livak e Schmittgen, 2001).

### **3.11. Ensaio de Northern Blot não radioativo**

O ensaio de Northern Blot não radioativo foi realizado segundo protocolo adaptado de Kim et al. (2010). Foi utilizada sonda LNA (Locked Nucleic Acid), descrita no Apêndice 1 e duas replicatas biológicas foram utilizadas.

Para a adequação do protocolo, ao invés da utilização de EDC (1-etil-3-(3-dimetilaminopropil) carbodiimida) para a realização do *cross-linking*, a membrana contendo o RNA foi submetida à radiação ultra violeta à 1200J por 2 minutos. O substrato utilizado para visualização das bandas também foi modificado. O CSPD, substrato quimioluminescente, foi substituído pelo NBT/BCIP, substrato que gera um precipitado de cor púrpura comumente utilizado na metodologia de Hibridização *in situ* (Javelle et al, 2011).

### **3.12. Ensaio de estresse oxidativo**

Para a realização do experimento, 10 mg de sementes de *A. thaliana* Col-0 foram desinfetadas e colocadas para germinar em erlenmeyers contendo meio de cultura MS/2 líquido e 1,5% de sacarose foi acrescido ao meio nutritivo (meio MS comercial sem vitaminas - PhytoTechnology) (Gauer, 2004). Após a quebra da dormência das sementes, os erlenmeyers foram transferidos para crescimento sob agitação contínua

suave (60 RPM) e temperatura de 22°C por 10 dias. Após 10 dias, foram adicionados 10 mM de peróxido de hidrogênio (Thompson *et al.*, 2008) em cada erlenmeyer e, como controle do experimento, foi adicionado o mesmo volume de meio MS/2. As plântulas foram coletadas após quatro, 16, 24, 36 e 48 horas de exposição das plantas ao estresse oxidativo e armazenadas à -80°C para posterior extração de RNA. As plântulas utilizadas como controle foram coletadas após quatro e 16, 24, 36 e 48 horas do início do tratamento.

### **3.13. Teste para indução de estresse**

Para todos os experimentos, meio MS líquido foi utilizado e o respectivo tratamento foi adicionado na placa contendo papel filtro estéril. Para o controle, meio MS (meio MS comercial sem vitaminas - PhytoTechnology) acrescido de 1% de sacarose foi utilizado. As plântulas foram coletadas com 48 e 72 horas e imediatamente armazenadas à -80°C até o momento do uso. Para o experimento de sacarose, foram adicionados 2%, 3% e 4% de sacarose ao meio de cultura (Colville *et al.*, 2011).

Para o experimento no qual foi empregado ácido abscísico (ABA), as seguintes concentrações foram utilizadas: 0,3 nM, 3 µM e 30 µM (adaptado de Weigel e Glazebrook, 2002).

### **3.14. Cruzamentos entre arabis mutantes**

As flores doadoras de pólen de arabis foram emasculadas, suas anteras foram retiradas com o auxílio de pinças. A seguir, o pistilo foi isolado das demais flores, as quais foram retiradas do ramo a fim de evitar a polinização. Dois dias após a emasculação, o pistilo recebeu o pólen contido nas anteras das flores doadoras, manualmente. Cada cruzamento foi devidamente identificado com fita adesiva.

Foram realizados cruzamentos recíprocos (Tabela 1) e as progênes denominadas: 1) duplo mutante HAD: onde a planta mutante doadora de pólen é *hd2a* e a planta receptora é *dnmt2*; 2) duplo mutante DHA: onde a planta mutante doadora de pólen é *dnmt2* e a planta receptora é *hd2a*.

Após o amadurecimento das silíquas provenientes dos cruzamentos, as sementes foram coletadas e armazenadas em tubos tipo *ependorf*, acondicionadas em estufa a 37°C para secagem por aproximadamente 24 horas e armazenadas a 4°C até o momento do uso.

Para seleção das plantas duplo-mutantes, as sementes foram semeadas em placa de Petri, como descrito anteriormente. Após 15 dias, as plântulas foram transferidas para o substrato. Após a aclimação das mesmas, foi coletado tecido foliar e extraído DNA para genotipagem.

**Tabela 1:** Genótipos utilizados para os cruzamento controles entre as arabisopsis.

<b>Doador de pólen</b>	<b>Receptor de pólen</b>
<i>dnmt2</i>	<i>hd2a</i>
<i>hd2a</i>	<i>dnmt2</i>

### 3.15. Análise de germinação

Para avaliar o efeito da sacarose na germinação, as sementes de arabisopsis foram germinadas de acordo com o item 3.1, em meio MS comercial foram adicionados 2%, 3% e 4% de sacarose (Colville et al., 2011). A presença dos cotilédones foi avaliada e a semente com cotilédones aparentes foi considerada germinada em 24, 48, 72 e 96 horas após terem sido colocadas na câmara de crescimento.

Foram utilizadas as sementes selecionadas por genotipagem (Apêndice 3) provenientes da progênie do cruzamento entre *dnmt2* e *hd2a*, descritas no item 3.14 (os duplos mutantes HAD e DHA), além das sementes *s4.dnmt2*, *s4.hd2a* e *s4.Col-0*, onde *s* (*siblings*) é utilizado para diferenciar as linhagens provenientes do cruzamento dos demais genótipos e "4" para indicar a geração F4. Os genótipos Col-0 (planta selvagem ou WT) e *dnmt2* também foram utilizados nas análises.

### 3.16. Preparação das bibliotecas de mRNA (RNA-seq)

As plântulas dos genótipos Col-0 e *dnmt2* foram coletadas com 48 horas de crescimento. Quatro réplicas biológicas de cada genótipo foram preparadas e avaliadas.

A construção das bibliotecas *single-end* de mRNA foi realizada utilizando o kit comercial NEBNext Ultra RNA Library Prep Kit for Illumina® (NEB), seguindo instruções do fabricante. Para definir a distribuição do tamanho e qualidade das

bibliotecas, uma alíquota de cada biblioteca foi empregada no sistema de análises biológicas Agilent 2200, utilizando-se o Agilent dsDNA High Sensitivity DNA Assay, seguindo instruções do fabricante. As bibliotecas foram ajustadas para uma concentração de aproximadamente 5 nM e a PCR quantitativa foi realizada utilizando o kit comercial Kapa Library Quantification Kit for Illumina® (Kapa Biosystems). A concentração das bibliotecas foi então ajustada para sua preparação para o sequenciamento Illumina.

### **3.17. Análises de bioinformática das bibliotecas de mRNA (RNA-seq)**

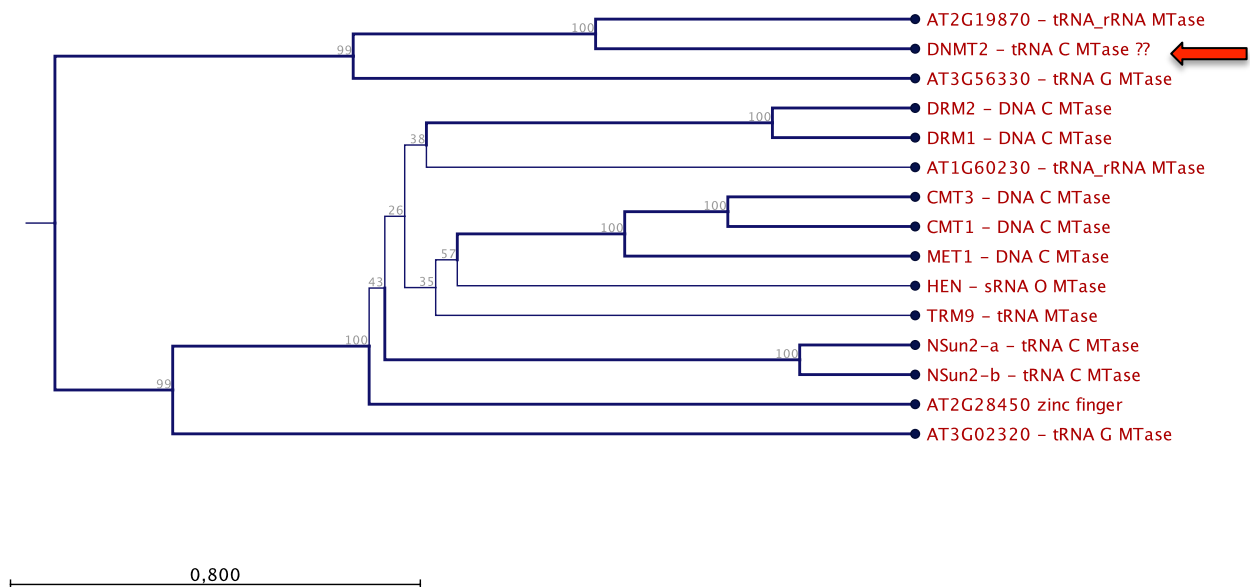
As análises de bioinformática das bibliotecas de mRNA foram realizadas através da plataforma Galaxy ( <https://galaxyproject.org> ), utilizando o *software* STAR (Dobin et al., 2012) para o mapeamento dos fragmentos com o genoma de referência e o pacote DEseq (Anders e Huber, 2010) para a análise de expressão diferencial. O mapeamento foi realizado sem *mismatches* e apenas fragmentos que mapearam em um único *locus* foram permitidos.

## 4. Resultados e Discussão

### 4.1. Análises *in silico* do gene *AtDNMT2*

As vias epigenéticas em vegetais são mais complexas do que em animais, e a metilação de DNA não é restrita ao contexto CpG. Apenas em relação à DNA metiltransferases, plantas possuem dez prováveis genes.

Com a finalidade de estabelecer relação de similaridade entre a enzima DNMT2 e demais metiltransferases de *A. thaliana*, algumas enzimas DNA metiltransferases e algumas RNA m<sup>5</sup>C (metiltransferases que metilam o carbono 5 na base citosina) foram selecionadas a partir de bancos de dados públicos e árvore filogenética foi gerada. Como demonstrado na Figura 1, *AtDNMT2* aparenta ser mais relacionada com RNA metiltransferases do que com DNA metiltransferases, dado que corrobora a hipótese sugerida por Jurkowski e Jeltsch (2011), na qual relatam que durante a evolução, DNMT2 perdeu o domínio N-terminal, indicando que houve uma troca de substrato, de DNA para RNA transportador. As mesmas análises filogenéticas realizadas pela dupla de pesquisadores em 2011 também descarta a possibilidade sugerida por Goll e colaboradores (2006), na qual DNMT2 teria surgido primeiro durante a evolução das proteínas C5 metiltransferases e a partir dela, surgiriam as DNA metiltransferases.



**Figura 1:** Árvore filogenética de similaridade construída utilizando o método de máxima verossimilhança (ML), com 1000 replicatas *Bootstrap*, demonstrando a relação entre algumas das metiltransferases encontradas em *A. thaliana*. O valor de *bootstrap* está representado nos nós dos ramos. A seta indica a enzima *AtDNMT2*.



A enzima de metilação DNMT2 é uma molécula conservada entre os eucariotos, abrangendo organismos tão diferentes quanto *S. pombe* e humanos (Jeltsch et al., 2006), e já foi demonstrado, *in vitro*, que em células *HeLa* e *Drosophila*, DNMT2 é capaz de metilar a citosina 38 (C38), nucleotídeo localizado logo após o *loop* do anticódon dos tRNAs Asp, Gly, e Val. Em plantas, até o momento, só se tem conhecimento do tRNA Asp<sup>GTC</sup> como alvo direto da DNMT2 (Goll et al., 2006).

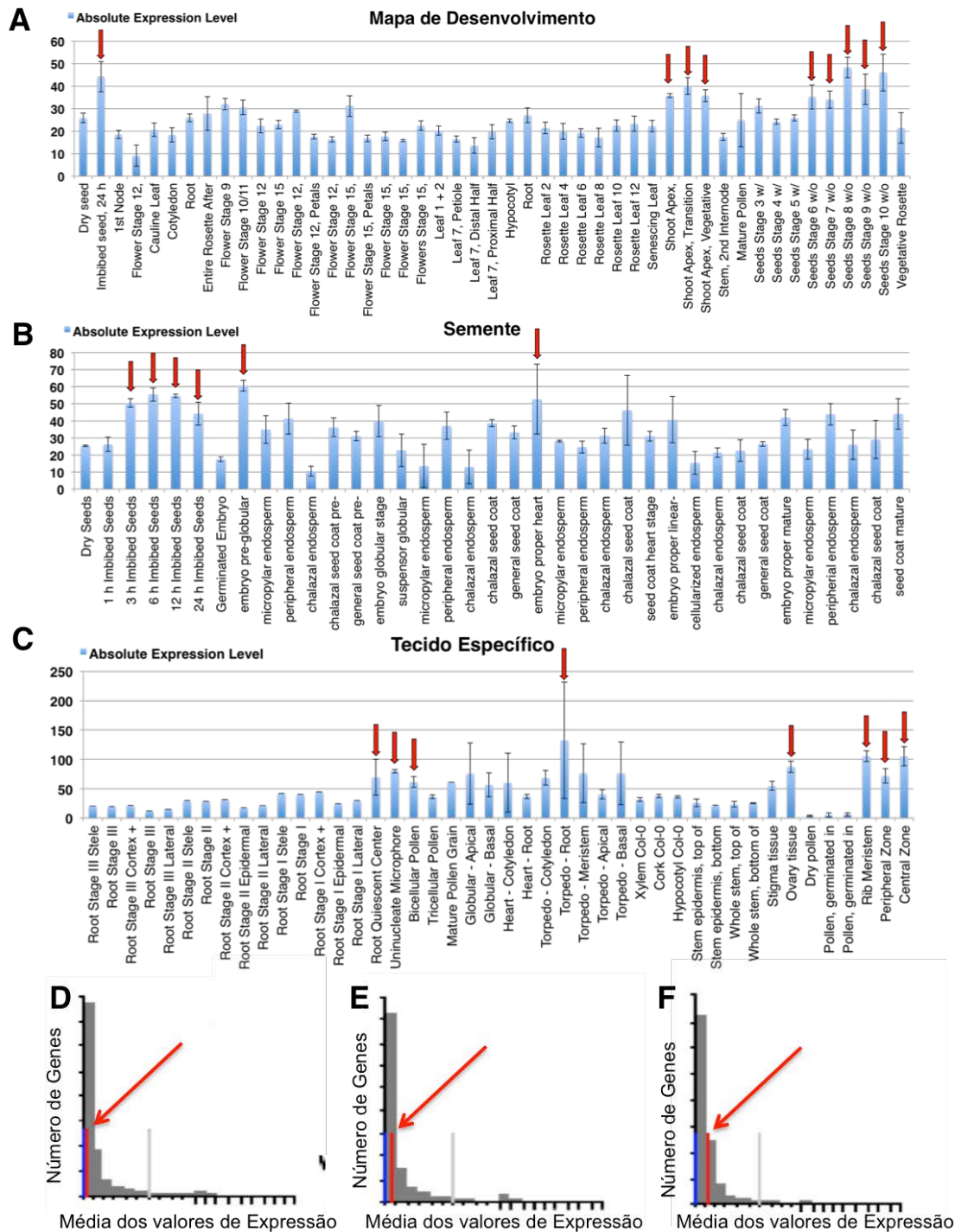
A enzima DNMT2 foi intitulada como DNA metiltransferase graças à presença de todos os dez motivos conservados, compartilhados e característicos da classe de enzimas DNA m<sup>5</sup>C. Porém, o domínio de reconhecimento de alvos (TRD) é altamente conservado e único em todas as proteínas DNMT2, assim como a falta da porção N-terminal, a qual é funcionalmente importante nas demais DNA metiltransferases (Goll e Bestor, 2005; Dong et al., 2001). Tanto a m<sup>5</sup>C de DNA, quanto de RNA compartilham o mesmo mecanismo de ação, porém fazem uso de diferentes resíduos catalíticos (Liu e Santi, 2000). *S. cerevisiae*, *C. elegans* e *O. dioica* não possuem nenhuma enzima DNA metiltransferase, tão pouco são detectadas metilações em seus genomas, porém, a presença da proteína DNMT2 em espécies relacionadas indicam que estes três organismos perderam o gene *DNMT2* durante a evolução (Jurkowski e Jeltsch, 2011), essa informação pode sugerir que DNMT2 não desempenhe função celular essencial ou outra(s) enzima(s) exerçam papel redundante.

Contudo, aproximadamente 8% da metilação no contexto CpHpH (H: A, T ou C) em *Arabidopsis* não é metilado pelas enzimas DRM1, DRM2 ou CMT2 (Stroud et al., 2012), sugerindo que outras DNA METases sejam responsáveis pela metilação desses nucleotídeos, não descartando a possibilidade desta outra enzima ser a DNMT2, já que estudos *in vitro* demonstraram que DNMT2 é capaz de se ligar ao DNA (Dong et al., 2001).

Também é interessante ressaltar que enzimas DNA m<sup>5</sup>C são capazes de realizar outras reações enzimáticas como a deaminação da C4 (citosina na posição 5), adição de aldeídos na C5 (citosina na posição 5) (Shen et al., 1992; Liutkeviciutė et al., 2009), modificações estas que podem ser de grande valia para a atuação do sistema imune contra vírus e para a maturação de células B em animais (Holmes et al., 2007; Dong et al., 2001).

A fim de começar a entender o funcionamento da enzima DNMT2 em plantas, dados relacionados à expressão do gene *DNMT2* em diferentes órgãos e tecidos de *Arabidopsis* foram obtidos de banco de dados públicos ( <http://bar.utoronto.ca/efp/cgi->

[bin/efpWeb.cgi](#) ) e dispostos em gráficos (Figuras 2A, B e C). Tais dados demonstram que a expressão de *DNMT2* é ubíqua e ocorre em níveis basais, com um ligeiro aumento de expressão na semente embebida por 24 horas, na região meristemática apical e nas sementes com 6 a 10 semanas de desenvolvimento (Fig. 2A), e ainda, a expressão de *DNMT2* é menor do que a maior parte dos genes expressos nessa categoria (Fig. 2D). Durante a embriogênese, também observa-se pequeno aumento na expressão da *DNMT2*, de forma quase constante no decorrer dos diferentes estádios de desenvolvimento do embrião, mas o aumento é ligeiramente maior nos estádios pré-globular, coração e na semente embebida por 3, 6, 12 e 24 horas (Fig. 2B), a expressão de *DNMT2* também é menor do que a maior parte dos genes expressos nessa categoria (Fig. 2E). Ainda, pode-se destacar o aumento de expressão da *DNMT2* no centro quiescente da raiz, nos estádios uni e bicelular de desenvolvimento do grão de pólen, na região onde a raiz se desenvolverá no estágio torpedo do desenvolvimento embrionário, no ovário e, finalmente, no meristema apical caulinar (Fig. 2C), aqui, observa-se que a expressão de *DNMT2* é ligeiramente maior do que a maior parte dos genes expressos nessa categoria (Fig. 2F).



**Figura 2:** Mapa de expressão absoluta de *DNMT2* em *Arabidopsis thaliana* (At5G25480) nos diferentes estádios do desenvolvimento **A**, **B** e **C**; as setas vermelhas representam os estádios com maior expressão em relação aos demais estádios. **(A)** Diferentes órgão/tecidos. **(B)** Desenvolvimento da semente. **(C)** Tecidos específicos. Os gráficos **D**, **E** e **F** indicam o valor de expressão máxima do gene *DNMT2* contra a média de expressão de todos os genes; as setas e as colunas vermelhas indicam a posição de *DNMT2*: **(D)** Diferentes órgão/tecidos. **(E)** Desenvolvimento da semente. **(F)** Tecidos específicos. Todos os dados foram obtidos através da ferramenta *Arabidopsis eFP Browser* disponível na página *BAR - The Bio-Analytic Resource for Plant Biology* – (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>) (Winter et al., 2007; Schmid et al., 2005).

As variações nos níveis de expressão absoluta do gene *AtDNMT2* durante a formação do embrião, início da ativação da semente, regiões de alto índice de atividade celular e em células germinativas, sugerem que a enzima *AtDNMT2* tenha função nesses órgãos/tecidos, embora sua expressão em condições normais seja baixa.

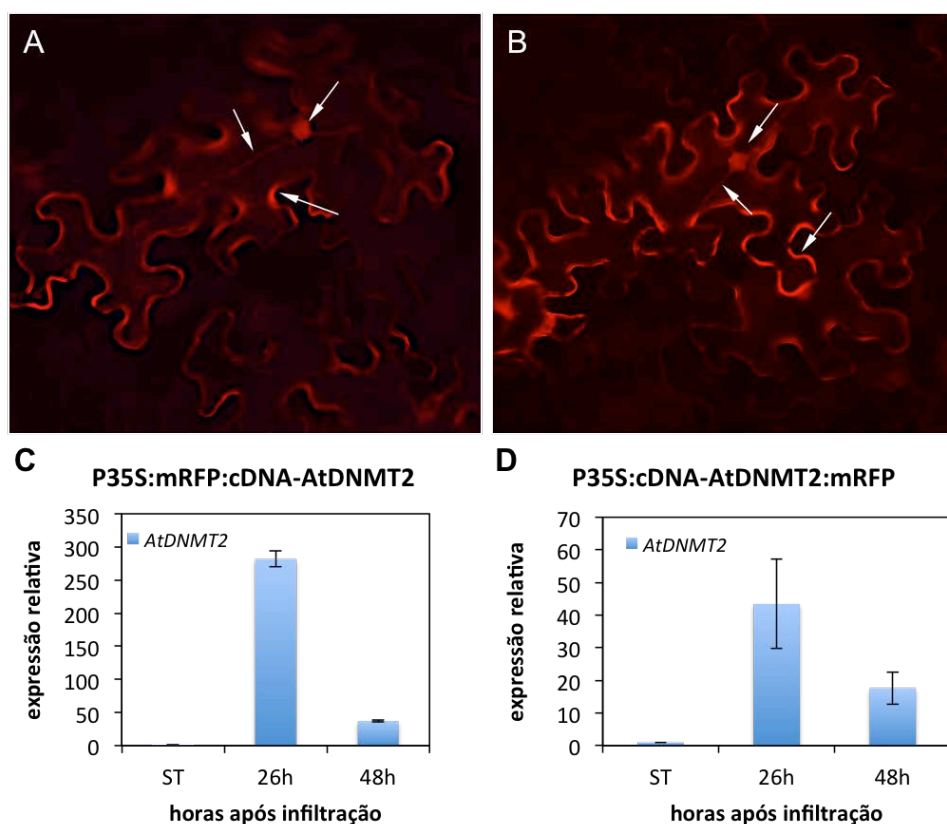
Portanto, análises fenotípicas e moleculares mais detalhadas são necessárias para tentar compreender a função da *AtDNMT2*, uma proteína evolutivamente antiga e conservada em diversos e distintos organismos. E ainda, pode-se ponderar que por ser produzida em quase todos os tipos celulares, *AtDNMT2* deva possuir função biológica relevante, fato que justificaria o gasto de energia das células para sua produção, talvez desempenhando papel no metabolismo básico da planta, ou ainda, exercendo papel na manutenção da estrutura do tRNA.

#### **4.2. Localização subcelular da enzima *AtDNMT2***

A fim de obter informações à respeito da localização subcelular da enzima *DNMT2* em *Arabidopsis thaliana*, e assim, obter informações adicionais sobre seu possível envolvimento com ambos os ácido nucléicos, dois cassetes de expressão foram construídos contendo o cDNA do gene *AtDNMT2* fusionado com a proteína fluorescente mRFP tanto no C-, quanto N-terminal. Após a clonagem em *Agrobacterium tumefaciens*, a mesma foi infiltrada em folhas jovens de tabaco e após 26 e 48 horas, levadas ao microscópio de fluorescência para visualização da proteína quimérica. Além disso, as amostras de tecido foliar foram coletados para análise de expressão do gene *AtDNMT2*.

Vinte e seis horas após agroinfiltração das folhas de tabaco, não foi possível visualizar fluorescência, porém, detectou-se alta expressão do transgene *AtDNMT2* de ambas as construções (Figs. 3C e 3D), e 48 horas após agroinfiltração, observou-se a fluorescência proveniente da atividade da mRFP tanto no núcleo quanto no citoplasma das células de tabaco e, surpreendentemente, observou-se a expressão da proteína vermelha no que aparenta ser o citoesqueleto das células (Figs. 3A e 3B) e, nesse intervalo de tempo, a expressão do transgene diminui drasticamente (Figs. 3C e 3D). O decaimento no acúmulo dos transcritos de *DNMT2* após 48 horas era esperado, já que mRNAs de *Arabidopsis* possuem meia vida variando entre 0,2 e 24 horas (Narsai et al., 2007). Como a visualização da proteína vermelha se dá apenas 48 horas após infiltração, pode-se especular que a transcrição do gene *AtDNMT2* é lenta,

necessitando de mais tempo para acumular a quantidade de moléculas necessárias para visualização da fluorescência.



**Figura 3:** Agroinfiltração em folhas de tabaco. Imagem de microscopia de fluorescência de folhas de tabaco após 48 horas de infiltração com *Agrobacterium tumefaciens* GV3101 com um aumento de 40x. Observa-se em ambas as imagens, fluorescência vermelha (mRFP) no núcleo, citoplasma e citoesqueleto, indicado pelas setas brancas. (A) Imagem da expressão do cassete *p35S:mRFP:cDNA-AtDNMT2*. (B) Imagem da expressão do cassete *p35S:cDNA-AtDNMT2:mRFP*. Análise de expressão do gene *AtDNMT2*, via RT-qPCR de folhas de tabaco com 26 e 48 horas após agroinfiltração em relação à folhas de tabaco não infiltradas (ST). A média e o desvio padrão de três amostras biológicas foram utilizadas para a confecção do gráfico. O gene  $\beta$ -tubulina de tabaco foi utilizado como controle endógeno. (C) Análise de expressão do gene *AtDNMT2* proveniente do cassete *p35S:mRFP:cDNA-AtDNMT2*. (D) Análise de expressão do gene *AtDNMT2* proveniente do cassete *p35S:cDNA-AtDNMT2:mRFP*.

Inicialmente, a localização da DNMT2 humana foi descrita como sendo citoplasmática (Goll et al., 2006), posteriormente foi demonstrado que Ehmeth, homóloga da DNMT2 em *E. histolytica*, e DmDNMT2 localizam-se tanto no núcleo quanto no citoplasma (Tovy et al., 2010; Schaefer et al., 2008), e em *Drosophila*, a DNMT2 é relocada para o núcleo durante fases específicas do ciclo celular, interagindo com a matriz nuclear (Schaefer et al., 2008). Embora organismos que possuem apenas a enzima DNMT2 como DNA m<sup>5</sup>C, como *Drosophila* e *S. mansoni*, não apresentem um padrão definido para a metilação de DNA em seus genomas (Raddatz et al., 2013), Ehmeth é capaz de interagir, *in vitro*, com a matriz nuclear, via

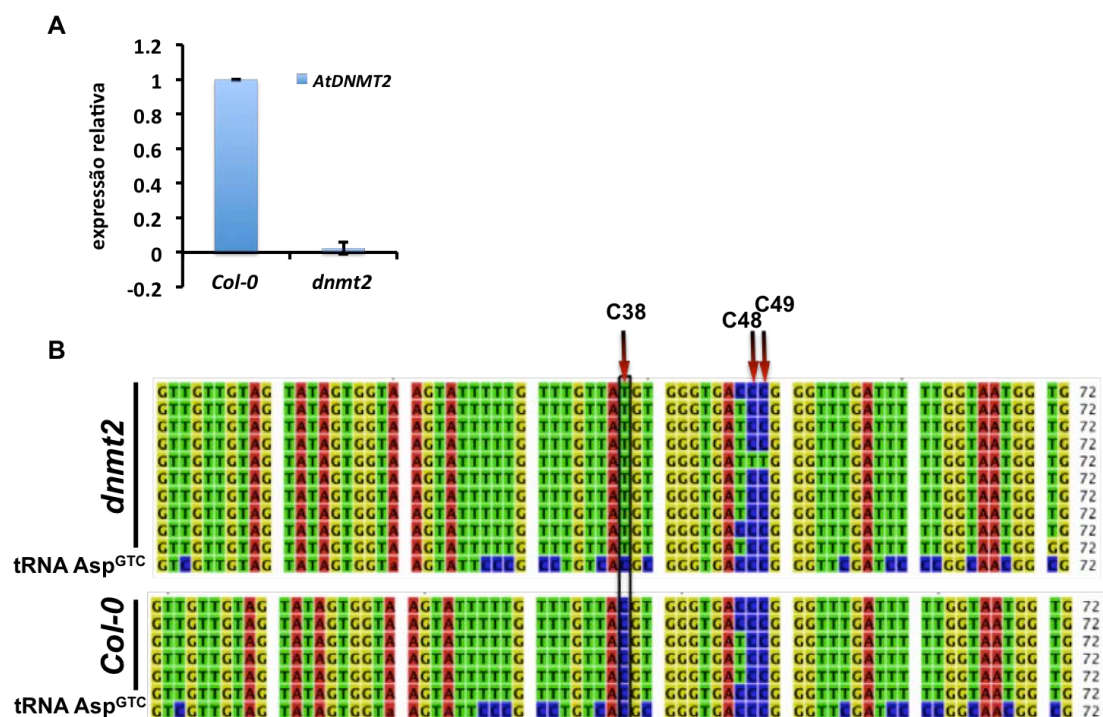
EhMRS2, um elemento de DNA que contém regiões de anexação *consensus scaffold*/matriz (S/MAR), que por sua vez são elementos regulatórios em *cis* (Banerjee et al., 2005).

Com isso, a visualização da AtDNMT2 no núcleo de células de tabaco indica a possibilidade de que a enzima AtDNMT2 também possa estar interagindo indiretamente com o DNA, intermediada por outras proteínas nucleares, já que, até o momento, não há informações de atividade de metilação de DNA pela enzima DNMT2 em nenhum organismo. Embora experimentos adicionais com controles internos precisem ser realizados, os dados preliminares sugerem que a AtDNMT2 pode associar-se com o citoesqueleto das células de tabaco, fato que também abre espaço para a exploração de novas funções para a enzima, como sua associação à proteínas de transporte ou sinalização celular, além da já sugerida função durante a mitose de células de *Drosophila*, quando a DNMT2 pode ser encontrada em microtúbulos, onde a enzima pode acessar o DNA (Schaefer et al., 2008). A visualização da fluorescência acumulada nas bordas das células pode ter sido causada pelo entumescimento dos tonoplastos (Reisen et al., 2005), que estaria “empurrando” o conteúdo citoplasmático para as bordas das células.

Novos experimentos com a adição dos controles das organelas celulares e citoesqueleto serão realizados em momento oportuno a fim de confirmar os resultados demonstrados no presente trabalho.

### **4.3. Caracterização molecular da planta mutante *Atdnmt2***

Para determinar a função biológica da enzima AtDNMT2, planta mutante contendo inserção de T-DNA no éxon 7 do gene *DNMT2* – *Atdnmt2* ou *dnmt2* (Goll et al., 2006) foi obtida e caracterizada. Para tanto, inicialmente foi realizada a confirmação da presença do T-DNA (Apêndice 1, Fig 1). Em seguida, a expressão do gene *AtDNMT2* na planta mutante *Atdnmt2* foi analisada a fim de confirmar o *knockdown* do gene (Fig 4A), assim como uma análise funcional da enzima AtDNMT2, via tratamento com bissulfito de sódio e sequenciamento Sanger do único alvo conhecido até o momento – tRNA Asp<sup>GTC</sup> (Fig 4B).



**Figura 4:** Caracterização molecular da planta mutante *Atdnmt2*. **(A)** Análise de expressão do gene *AtDNMT2* de plântulas *Atdnmt2* em relação à plântulas *Col-0* (wt) crescidas em meio MS. O gene *AtPP2A* foi utilizado como controle endógeno. **(B)** Sequências das amostras de plântulas *dnmt2* e *Col-0* tratadas com bissulfito de sódio alinhadas com o tRNA Asp<sup>GTC</sup>. O painel superior representa a linhagem da planta mutante *Atdnmt2* e o painel inferior, a linhagem *Col-0*. As setas indicam as posições em que as citosinas são sabidamente metiladas no tRNA maduro: C38 pela enzima DNMT2, C48 e C49 pela enzima NSun2.

A análise de expressão do gene *AtDNMT2* da planta mutante *Atdnmt2*, comparada com a planta *Col-0* (tipo selvagem - wt) demonstra claramente uma diminuição abrupta da detecção dos transcritos, indicando que a planta mutante não é funcional, já que quase não produz mRNA (Fig 4A).

Posteriormente, uma análise funcional da proteína *AtDNMT2* foi realizada para confirmar sua falta de atividade. Nos painéis demonstrados pela Figura 4B, pode-se visualizar que na posição C38 ocorreu a conversão das Cs da linhagem *Atdnmt2*, demonstrando que a *AtDNMT2* não está desempenhando seu papel de metilação na C38, já que, se a metilação tivesse ocorrido, as Cs estariam protegidas da conversão por bissulfito de sódio e continuariam como C após a reação de sequenciamento. Já na linhagem *Col-0*, a conversão não ocorre, indicando que a C38 foi metilada e portanto protegida da ação do bissulfito de sódio. Como controle para a verificação da eficiência do tratamento, foram utilizadas as C48 e C49, que em tRNAs, incluindo o tRNA Asp<sup>GTC</sup>, são metiladas pela enzima NSun2, uma enzima de m<sup>5</sup>C de tRNAs (Tuorto et al., 2012; Militello et al., 2014).

Como já descrito por Goll e colaboradores (2006), a planta mutante *Atdnm2* não possui nenhum fenótipo óbvio na planta adulta e produz sementes viáveis, entretanto, os autores não avaliaram a possibilidade de alterações fenotípicas nas fases iniciais do desenvolvimento e no desenvolvimento embrionário. Além disso, até o momento não foram descritos novos alvos para a enzima DNMT2 em *Arabidopsis thaliana*.

#### 4.4. Identificação de novos alvos para AtDNMT2

Até o momento, o tRNA Asp<sup>GTC</sup> é o único alvo conhecido para a DNMT2 em *Arabidopsis thaliana* (Goll et al., 2006). Porém, em outros organismos, diferentes alvos já foram encontrados, como em *Drosophila*, onde a enzima é capaz de metilar outros dois tRNAs além do Asp<sup>GTC</sup>, o tRNA Gly<sup>GCC</sup> e o tRNA Val<sup>ACC</sup> (Schaeffer et al., 2010) e na bactéria *Geobacter sulfurreducens*, na qual enzima homóloga a DNMT2 metila especificamente o tRNA Glu (Shanmugam et al., 2014).

A fim de encontrar outros possíveis alvos para AtDNMT2, a técnica SB-RNAseq (*Sodium Bisulfite RNA sequencing*) foi empregada (Koddami e Cairns, 2014). A escolha foi feita baseada na capacidade da técnica em distinguir todas as citosinas (Cs) metiladas das não metiladas (Frommer et al., 1992). Desta forma, o experimento foi realizado utilizando plântulas mutantes *dnmt2* e Col-0 (plantas selvagens ou WT).

Os resultados do experimento de SB-RNAseq revelaram pelo menos dois novos alvos para a RNA metiltransferase AtDNMT2: tRNA Gly<sup>GCC</sup> e tRNA Gly<sup>CCC</sup> (Tabela 2). As análises também revelaram outros RNAs, que não tRNAs, como possíveis candidatos, tais como alguns transcritos de mRNA e rRNA que, de acordo com os dados, são diferentemente metilados em ambos os genótipos, porém análises mais detalhadas serão realizadas posteriormente a fim de confirmar os dados obtidos.

**Tabela 2:** Porcentagem de metilação na base C dos tRNAs candidatos diferencialmente metilados das bibliotecas de RNA bissulfito de plântulas, Col-0 e *Atdnmt2* com 48 horas de crescimento em meio MS. O Teste de Fisher foi aplicado e considerado significativo para *p-values* ≤ 0,05.

	Col-0	<i>Atdnmt2</i>	<i>p-value</i>
tRNA Asp <sup>GTC</sup>	90,6%	0,4%	0
tRNA Gly <sup>GCC</sup>	92,3%	4,35%	0
tRNA Gly <sup>CCC</sup>	95,5%	1,12%	0.02

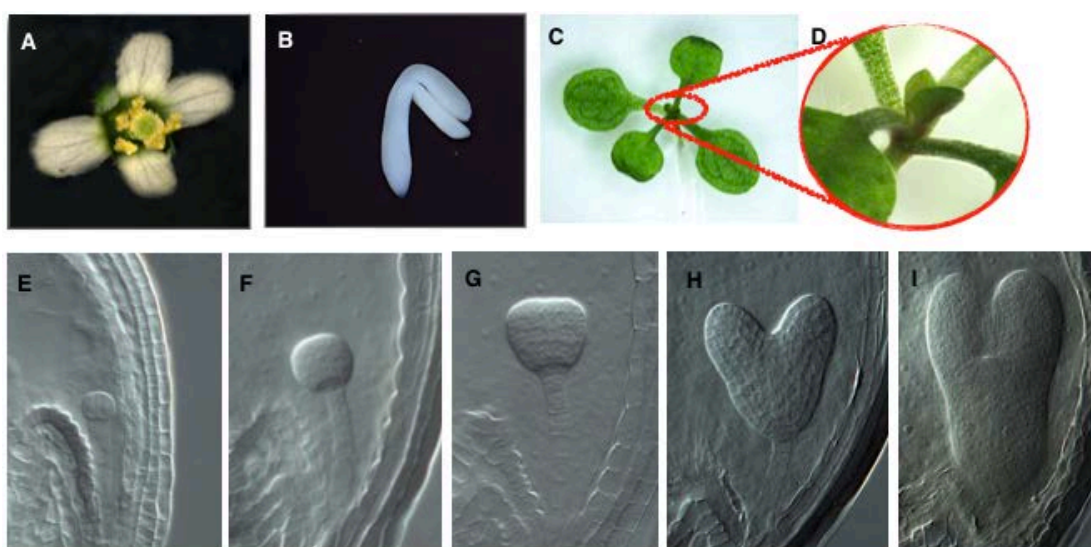
Como esperado, as análises em plântulas germinadas e crescidas por 48 horas em meio MS comercial acrescidas de 1% de sacarose, demonstraram que AtDNMT2 é



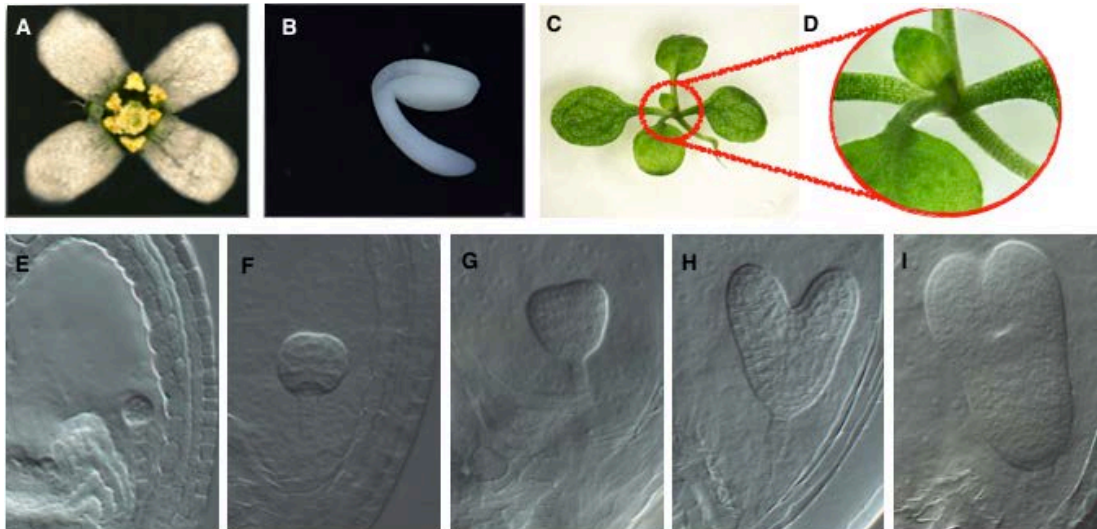
responsável pela metilação da C38 no tRNA Asp<sup>GTC</sup>, assim como demonstrado por Goll e colaboradores (2006), e aqui foi utilizado como controle positivo para o experimento (Tabela 2). Era esperado encontrar pelo menos mais dois alvos: tRNA Gly<sup>GCC</sup> e tRNA Val<sup>ACC</sup>, descrito previamente em *Drosophila* por Schaefer e colaboradores (2010). Além desses dois tRNAs, DNMT2 de outros organismos também possuem a habilidade de metilar outros tRNAs na mesma posição. Por exemplo, a enzima Pmt1 (*Schizosaccharomyces pombe*) e a DNMT2 da bactéria *Geobacter sulfurreducens*, as quais metilam o tRNA Glu na posição C38 (Becker et al., 2012; Shanmugam et al., 2014). Deste modo, é plausível supor que existam outros alvos para AtDNMT2 além do previamente descrito em plantas. Outros dois tRNAs-candidatos foram observados neste experimento: tRNA Gly<sup>GCC</sup> e tRNA Gly<sup>CCC</sup>, os quais serão validados em experimentos futuros.

#### 4.5. Análise fenotípica de plântulas e embrião de arabidopsis Col-0 e *Atdnmt2*

Embriões em diferentes estádios de desenvolvimento, assim como sementes embebidas por 24 horas e plântulas com 10 dias de idade do genótipo Col-0 (planta selvagem ou WT) e mutante *Atdnmt2* foram coletadas e preparadas a fim de verificar eventuais alterações fenotípicas que indicassem função biológica da enzima DNMT2. Entretanto, nenhuma diferença óbvia foi observada em plântulas e embriões do mutante *dnmt2* quando comparados ao Col-0 (Figs. 5 e 6).



**Figura 5:** Imagens de diferentes estádios de desenvolvimento em Col-0. (A) Flores; (B) Semente embebida em água por 24; (C) Plântula com 10 dias; (D) Inserção das folhas em plântulas com 10 dias; (E) estágio Octante; (F) estágio globular; (G) estágio triangular; (H) estágio coração; (I) estágio torpedo.



**Figura 6:** Imagens de diferentes estádios de desenvolvimento em *dnmt2*. (A) Flores; (B) Semente embebida em água por 24; (C) Plântula com 10 dias; (D) Inserção das folhas em plântulas com 10 dias; (E) estágio Octante; (F) estágio globular; (G) estágio triangular; (H) estágio coração; (I) estágio torpedo.

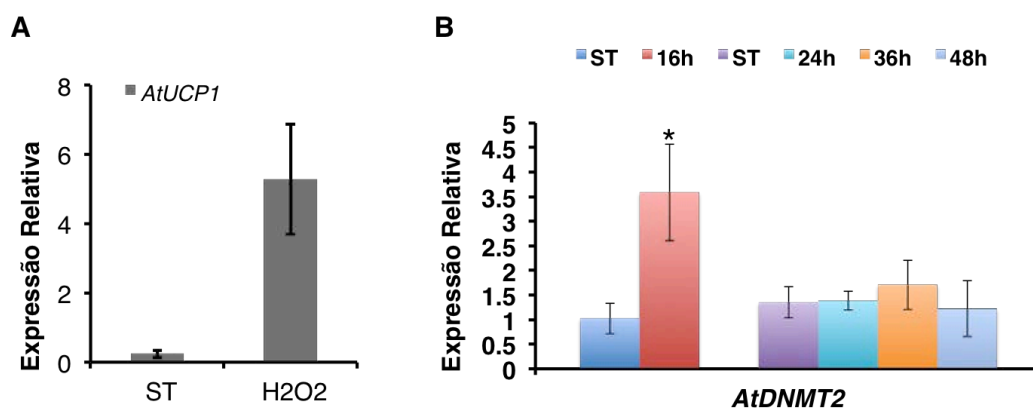
Anteriormente, nenhum fenótipo havia sido observado em *arabidopsis*, camundongos ou *Drosophila* na ausência da proteína DNMT2, embora análises detalhadas de desenvolvimento embrionário não haviam sido realizadas (Goll et al., 2006). Porém, estudos realizados em *zebrafish* indicaram que o *knockdown* do gene *DNMT2* resulta em falhas no desenvolvimento de órgãos como fígado, cérebro e retina de embriões com 80 horas após a fecundação. Baseados nestas observações, nossa hipótese era que possíveis alterações ocorressem em alguma fase do desenvolvimento embrionário ou estádios iniciais de desenvolvimento da plântula de *arabidopsis*. Porém, a ausência de diferenças fenotípicas entre o mutante *dnmt2* e Col-0, em condições normais de crescimento, sugere que a função biológica de *DNMT2* pode estar associada à respostas à condições de estresse. Portanto, fez-se necessária análises moleculares e fenotípicas em condições de estresses para possivelmente revelar vias de atuação da enzima AtDNMT2.

#### 4.6. Avaliação da expressão do gene *AtDNMT2* e sua possível função sob estresse oxidativo

Condições de estresse oxidativo podem induzir a clivagem de tRNAs, liberando as metades de tRNAs em plantas e animais (Thompson et al., 2008). Sabe-se que a enzima DNMT2 é capaz de metilar o tRNA Asp<sup>GCT</sup> em *arabidopsis* (Goll et al., 2006; este trabalho) e, ainda, que em células de *Drosophila* sob condições de estresse

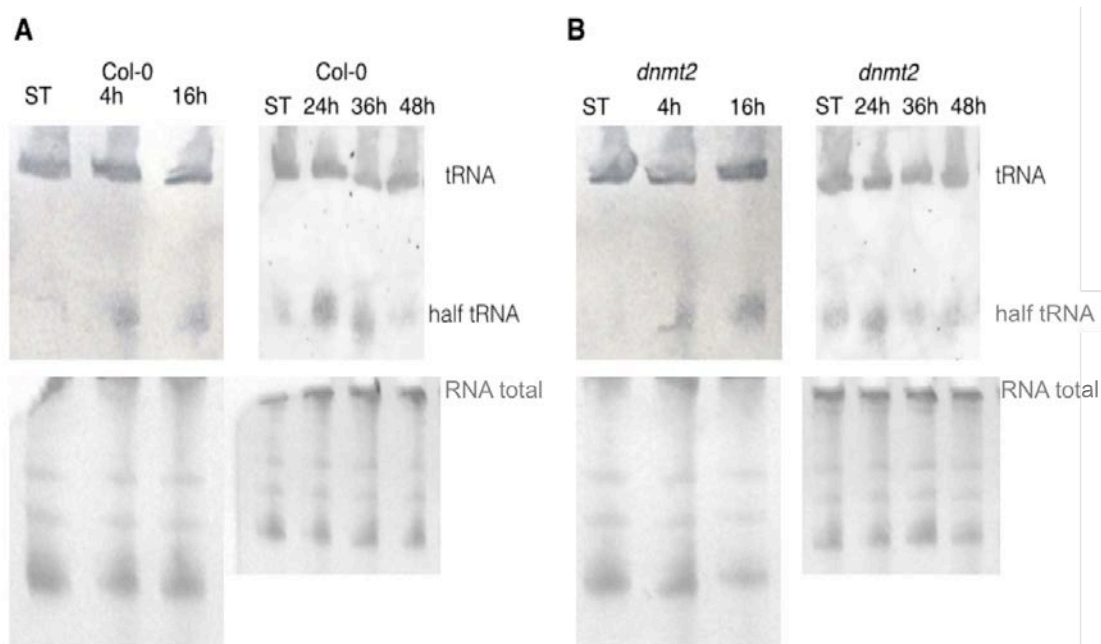
oxidativo e de altas temperaturas (*heat shock*), DmDNMT2 é capaz de proteger o tRNA contra essas clivagens (Schaefer et al., 2010).

Para avaliar a hipótese de que AtDNMT2 exerce papel similar na proteção do tRNA Asp<sup>GTC</sup> em arábido, plântulas Col-0 e mutante *dnmt2* foram submetidas a condições de estresse oxidativo, com aplicação de peróxido de hidrogênio (H<sub>2</sub>O<sub>2</sub>). Inicialmente, foram avaliados os níveis dos transcritos do gene *AtDNMT2* em plântulas Col-0 submetidas ao estresse oxidativo. Conforme os resultados de RT-qPCR indicam, os níveis transcricionais do gene *AtUCP1* (utilizado como controle para o estresse oxidativo; Begcy et al., 2011) aumentam (Figura 7A), corroborando o efeito molecular do estresse oxidativo (Schaeffer et al., 2010). Os níveis de transcritos do gene *AtDNMT2* são elevados quando plântulas de arábido são submetidas a 16 horas de estresse oxidativo (Figura 7B). Porém, após 24, 36 e 48 horas de tratamento os níveis transcricionais do gene *AtDNMT2* não apresentaram mudança significativa (Figura 7B).



**Figura 7:** RT-qPCR de plântulas Col-0 crescidas em meio MS/2 líquido, submetidas à estresse oxidativo. O estresse foi gerado adicionando-se 10mM de H<sub>2</sub>O<sub>2</sub> ao meio por 16, 24, 36 e 48 horas. O gene *AtPP2A* foi utilizado como controle endógeno. (A) Análise de expressão do *AtUCP1* utilizado como controle para o experimento, após 16 horas de estresse (Begcy et al., 2011). (B) Análise de expressão de *AtDNMT2*.

Após realização das análises de expressão, a fim de examinar a hipótese da enzima AtDNMT2 proteger o tRNA Asp<sup>GTC</sup> sob condições de estresse oxidativo, foi avaliado o acúmulo de fragmentos correspondentes à metade do tRNA Asp<sup>GTC</sup> em plântulas Col-0 e *dnmt2* submetidas ao estresse oxidativo, via técnica de *Northern Blot* não radioativo. Observou-se aumento da intensidade nas bandas correspondentes à metade do tRNA Asp<sup>GTC</sup> após quatro e 16, 24 e 36 horas de aplicação de estresse, e uma diminuição da intensidade das bandas após 48 horas da aplicação do estresse, tanto na planta selvagem Col-0 quanto no mutante *Atdnmt2* (Figura 8).



**Figura 8:** Northern Blot de plântulas arabidopsis crescidas em meio MS/2 líquido, submetidas à estresse oxidativo. O estresse foi gerado adicionando-se 10mM de H<sub>2</sub>O<sub>2</sub> ao meio MS comercial contendo 1% de sacarose. As bandas mais robustas e de igual intensidade representam o tRNA Asp<sup>GTC</sup> íntegro. Abaixo, estão as imagens do RNA utilizado para o experimento, para cada amostra biológica, 20 µg de RNA total foram utilizados. (A) Col-0. (B) *Atdnmt2*.

Apesar da detecção do aumento nas metades do tRNA Asp<sup>GTC</sup> em ambos os genótipos avaliados sob condições de estresse oxidativo, foi observada aumento de igual intensidade entre as plântulas Col-0 e *Atdnmt2* nos tempos de estresse em que foram submetidas. Era esperado que, com a falta da atividade da enzima AtDNMT2 nas plântulas *dnmt2*, houvesse maior acúmulo dos fragmentos correspondentes às metades do tRNA avaliado. Conforme observado em *Drosophila*, células *HeLa*, *E. histolytica* e *S. pombe*, a ação da DNMT2 é correlacionada com estresse de calor, oxidativo e nutricional (Fisher et al., 2006; Tovy et al., 2010; Schaefer 2010; Thiagarajan et al., 2011; Becker et al., 2012), e ainda, Mytych e colaboradores (2014), relataram que o gene *DNMT2* tem sua expressão aumentada em células *HeLa* durante estresse oxidativo. Além disto, já se sabe que condições de estresse podem induzir a liberação de fragmentos que correspondem à metade do tRNA. Essa clivagem se dá no *loop* do anticódon, tanto em animais como em arabidopsis (Thompson et al., 2008). Goll e colaboradores (2006) demonstraram que a DNMT2 humana é capaz de metilar a C38 do tRNA Asp<sup>GTC</sup> *in vitro*. Em células de *Drosophila*, Schaefer e colaboradores (2010) demonstraram que DmDNMT2 é capaz de proteger o tRNA – incluindo o

tRNA Asp<sup>GTC</sup> - contra clivagens induzidas por estresse oxidativo e térmico. Os resultados encontrados no presente trabalho sugerem que a metilação da C38 no tRNA Asp<sup>GTC</sup> em arábido pode não ter relação com a proteção do tRNA em condições de estresse oxidativo, podendo ser uma metilação estrutural, que somada as demais modificações presentes no tRNA, mantém a estabilidade da molécula.

A metilação da citosina (m<sup>5</sup>C) está presente em tRNAs, em diferentes posições e é apenas uma dentre as diversas modificações dos tRNAs. Apesar da m<sup>5</sup>C em tRNA ser bem conhecida, sua função ainda permanece incompreendida. Recentemente, Tuorto et al. (2012), com o intuito de caracterizar funcionalmente a m<sup>5</sup>C em tRNA, geraram camundongos com falta das duas enzimas de metilação de tRNA (DNMT2 e NSun2) e observaram que a DNMT2 pode atuar conjuntamente com a NSun2, provocando defeitos no desenvolvimento e até letalidade, concluindo que a metilação de tRNA é um regulador da síntese proteica.

Em animais, segundo modelo proposto por Tuorto e colaboradores (2012), a estabilidade do tRNA se dá pela atuação de duas tRNA metiltransferases, DNMT2 e a NSun2. Quando um organismo é submetido à algum tipo de estresse, estas enzimas são inibidas e o tRNA se torna alvo de enzimas de degradação, o que geraria as metades de tRNA (Durdevic et al., 2013). Os resultados apresentados no presente trabalho, indicam aumento na expressão do gene *AtDNMT2*, porém as possíveis diferenças no nível proteico não foram avaliadas. Pode-se ponderar que devido à presença de homólogos a NSun2 em arábido (Fig. 1 e Apêndice 4), o tRNA se manteve estável durante este período de exposição à H<sub>2</sub>O<sub>2</sub>, o que anularia o efeito da falta de enzima DNMT2 no mutante *dnmt2*, pelo menos nas condições experimentais avaliadas neste estudo.

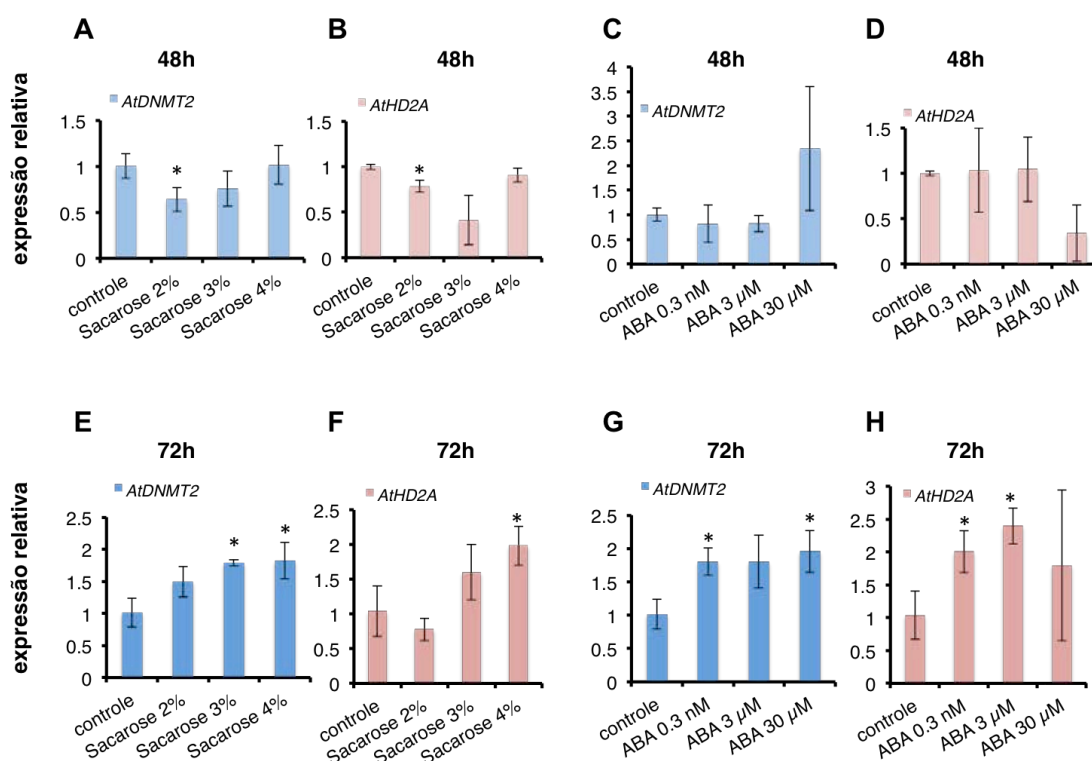
Portanto, pode-se concluir que houve aumento da liberação dos fragmentos de tRNA em resposta ao estresse oxidativo, corroborando os dados de Thompson e colaboradores (2008); porém, não há nenhum indicativo de que *AtDNMT2* possua função na proteção do tRNA Asp<sup>GTC</sup> contra clivagens durante estresse oxidativo, ao contrário do que parece ocorrer em *Drosophila* (Schaefer et al., 2010). Talvez o aumento de expressão do gene *AtDNMT2* tenha outra função biológica que não a proteção de tRNA, podendo agir como cofator em outras vias de resposta à estresse.

#### 4.7. Alteração da expressão do gene *AtDNMT2* em resposta a outros estresses abióticos

Com o intuito de continuar estudando a resposta dos genes *AtDNMT2* em condições de estresse e seu possível papel durante a germinação e/ou crescimento, a expressão dos genes *AtDNMT2* e *AtHD2A* foram avaliadas em plântulas Col-0 submetidas a três diferentes tipos de estresse: concentrações crescentes de sacarose (2%, 3% e 4 %) (Colville et al., 2011) e a presença do fitormônio ácido abscísico ou ABA (0.3nM, 3µM e 30µM) (adaptado de Weigel e Glazebrook, 2002) e em dois estádios de desenvolvimento, 48 e 72 horas após estratificação de dois dias a 4°C no escuro. Em condições de laboratório, com controle de fotoperíodo, temperatura e nutrição, de 24 a 72 horas após estratificação das sementes, as mesmas devem completar o processo germinativo (Boyes et al., 2001; Zanten et al., 2011). Os diferentes estádios de desenvolvimento escolhidos devem representar o conjunto de genes ativos durante a emergência da raiz, 48 horas, e durante o surgimento do hipocótilo e cotilédones, 72 horas, a fim de averiguar possíveis diferenças na resposta de *AtDNMT2* em condições de estresse nas diferentes fases iniciais do desenvolvimento. O gene *AtHD2A* foi avaliado neste experimento pois a interação da proteína DNMT2 e HD2A já foi demonstrada (Song et al., 2010), assim como sua relação com a germinação de sementes de arábido (Colville et al., 2011).

O gene *AtDNMT2* sofre uma pequena, porém significativa repressão nas plântulas germinadas e mantidas à 2% de sacarose quando comparadas com as plântulas controle, após 48 horas de crescimento (Figura 9A). No tratamento com ABA, durante mesmo período de exposição da semente, não houve alteração estatisticamente significativa do acúmulo dos transcritos de *AtDNMT2* (Figura 9C). Setenta e duas horas de exposição da semente à 2% de sacarose induziram o acúmulo de transcritos do gene *AtDNMT2*, quando comparados ao experimento controle, tanto em altas concentrações de sacarose (3% e 4%) quanto nas concentrações de 0,3 nM e 30 µM de ABA (Figuras 9E e G). No período de 48 horas, os transcritos de *AtHD2A* parecem acompanhar os níveis de *AtDNMT2* durante a germinação em diferentes concentrações de sacarose (Figura 9B). Neste período de tratamento, *AtHD2A* manteve os mesmos níveis transcricionais durante tratamento com ABA, em relação às plântulas controles não tratadas (Figura 9D). Após período de 72 horas, os transcritos

do gene *AtHD2A* são induzidos nos mesmos tratamentos que o gene *AtDNMT2* (Figuras 9F e H).



**Figura 9:** Expressão relativa via RT-qPCR de plântulas de *Arabidopsis thaliana* em diferentes tratamentos. Todas as plântulas cresceram em meio MS líquido, condições de dia longo e temperatura controlada entre 20-22°C. Além disso, 1% de sacarose foi adicionado ao meio das plântulas utilizadas como controle. *AtPP2A* foi utilizado como controle endógeno e a expressão de cada tratamento foi calculada em relação à amostra controle do experimento. O Teste t de Student foi aplicado e considerado significativo para  $p\text{-value} \leq 0,05$ . Aproximadamente 20 mg de plântulas foram coletadas com 48 horas após colocadas para germinar. **(A)** Análises de expressão de *AtDNMT2* a 2%, 3% e 4% de sacarose; **(B)** Análises de expressão de *AtHD2A* a 2%, 3% e 4% de sacarose; **(C)** Análises de expressão de *AtDNMT2* a 0.3nM, 3μM e 30μM de ABA; **(D)** Análises de expressão de *AtHD2A* a 0.3nM, 3μM e 30μM de ABA. As plântulas foram coletadas com 72 horas após colocadas para germinar. **(E)** Análises de expressão de *AtDNMT2* a 2%, 3% e 4% de sacarose; **(F)** Análises de expressão de *AtHD2A* a 2%, 3% e 4% de sacarose; **(G)** Análises de expressão de *AtDNMT2* a 0.3nM, 3μM e 30μM de ABA; **(H)** Análises de expressão de *AtHD2A* a 0.3nM, 3μM e 30μM de ABA.

Em *S. pombe*, Pmt1, enzima homóloga à DNMT2, é capaz de metilar tRNA em condições de privação nutricional (Becker et al., 2012); por outro lado, Ehmet (homóloga da DNMT2 do organismo *E. histolytica*) parece ser inibida sob privação de glicose e a enzima enolase se acumula no núcleo, formando um complexo com Ehmet (Tovy et al., 2010). Independentemente de como o estresse regula a expressão do gene *DNMT2*, a energia metabólica parece desempenhar papel importante no controle da maquinaria enzimática epigenética (Tovy et al., 2010).

As plantas fazem uso de diferentes tipos de açúcares em seu metabolismo básico para um correto desenvolvimento e crescimento. As plantas necessitam de açúcar, já que são a fonte primária de carbono e de energia para quase todos os tipos celulares, além de controlar o metabolismo e resistência ao estresse. Na última década, o papel sinalizador de açúcares e seus dramáticos efeitos no crescimento e desenvolvimento das plantas se tornou evidente (Duarte et al., 2013, Gupta et al., 2015; Chen 2015).

O metabolismo de açúcares em plantas é extremamente complexo e suas concentrações variam em resposta tanto ao desenvolvimento quanto à fatores ambientais. Durante os primeiros estádios de desenvolvimento de *arabidopsis*, altos níveis de açúcares exógenos são capazes de reprimir o alongamento do hipocótilo, a expansão e esverdeamento do cotilédone e, conseqüente, o desenvolvimento da parte aérea (Rolland et al., 2006). O perfil transcricional das plântulas de *arabidopsis* também revela o importante papel da glicose, que é responsável por uma intrincada rede de genes que se comunicam com hormônios e vias de respostas à estresses (Duarte et al., 2013). Estas cascatas de eventos desencadeiam a repressão de genes, seguidas pela ativação de outros genes, o que sugere uma reprogramação ao nível de remodelamento da cromatina (Price et al., 2004). Já se sabe que a família de histonas deacetilase HD2 - exclusiva de plantas - responde à diferentes açúcares e a HD2A é um dos membros mais responsivos, além de já ter sido descrita a interação proteica entre a AtDNMT2 e as AtHD2 (Song et al., 2010). No mutante *Athd2a*, a taxa de germinação é maior em relação à planta selvagem na ausência de glicose. Além disso, o mutante possui redução da inibição de germinação em concentrações crescentes de glicose (Colville et al., 2011). Esses dados, somados ao fato de que DNMT2 é requerida nos eventos iniciais da embriogênese animal (Kiani et al., 2013), sugerem que possivelmente a falta de ambas as proteínas pode alterar a eficiência de germinação e/ou crescimento da *arabidopsis*.

Além dos açúcares, hormônios vegetais são moléculas sinalizadoras que também possuem papel crucial no desenvolvimento vegetal e são responsáveis pela comunicação entre células, tecidos e órgãos. Além disso, alguns hormônios vegetais são essenciais para a regulação da dormência das sementes e germinação. Quando o desenvolvimento da semente está finalizado – a fase de maturação terminou e a fase de dessecação se inicia – o ABA acumula dando início à dormência. ABA está envolvido na indução e na manutenção da dormência da semente (Maia et al., 2014). A



expressão gênica durante o desenvolvimento embrionário e inibição da germinação são regulados por ABA por meio de fatores de transcrição (Finkelstein e Gibson, 2002; Lee et al., 2015).

O papel da metilação/demetilação do DNA não é completamente entendido no processo de germinação e desenvolvimento inicial das plantas. Em arábida, a acetilação/deacetilação de histonas também parece regular o desenvolvimento da semente, e membros da família HD2, como por exemplo a *AtHD2A*, são expressas nos óvulos, embriões, meristema apical, folhas primárias e são fortemente induzidas durante a embriogênese somática (Zhou et al., 2004).

O aumento dos níveis dos transcritos *AtDNMT2* e *AtHD2A* após 72 horas de crescimento e germinação em altas concentrações de sacarose e ABA sugerem que ambos os genes possam atuar em vias similares em resposta à estresses abióticos, corroborando dados que demonstram resposta da família HD2 à estresses bióticos e abióticos (Sridha e Wu, 2006), assim, os membros da família devem desempenhar papéis na resposta à estresse e germinação da semente. *AtDNMT2* é regulado por estresse, de acordo com dados apresentados no presente trabalho e com outros modelos biológicos. Em células de *Drosophila*, Schaefer e colaboradores (2010) sugerem que DNMT2 desempenha um papel de protetor de tRNAs. No parasita *E. histolytica*, *Ehmeth* (homólogo à DNMT2), parece exercer papel essencial na divisão celular e resposta à estresse oxidativo, porém o mecanismo com que isso acontece permanece desconhecido (Fisher et al., 2006). DNMT2 também pode participar do processamento de RNAs durante estresse celular, demonstrado por Thiagarajan e colaboradores (2011) em experimentos nos quais a enzima DNMT2, em células *HeLa* sob condições de estresse, realocaliza do núcleo para grânulos de estresse no citoplasma e para corpos de processamento de RNA. Em *S. pombe*, *Pmt1* (homóloga à DNMT2) também possui atividade de metilação de tRNA, entretanto, a metilação de tRNA parece ser dependente do *status* nutricional (Becker et al., 2004).

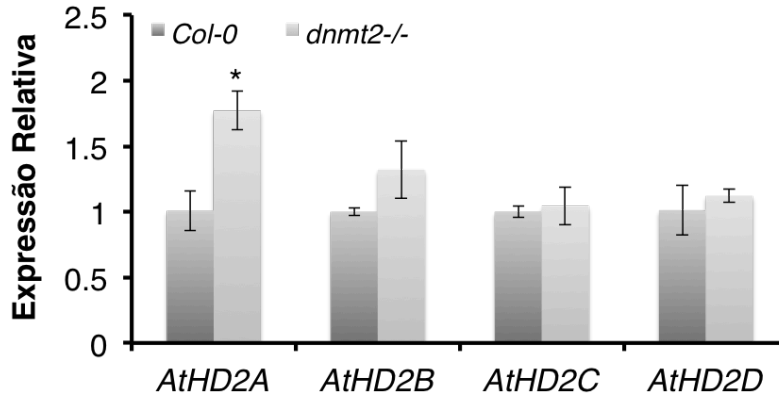
Baseado nas informações disponíveis na literatura em relação à DNMT2, essa enzima parece desempenhar diferentes papéis em diferentes organismos. DNMT2 é uma proteína conservada entre os diversos reinos e parece ser a DNA metiltransferase mais antiga que, provavelmente mudou de função e substrato em algum momento durante a evolução, de acordo com as necessidades de cada espécie. A relação entre DNMT2 e estresse parece ser clara, enquanto que sua exata função biológica, seja ela proteção de tRNA, metilação de tRNA ou processamento de RNA, parece mudar de

acordo com a espécie observada. Para uma visão mais abrangente acerca da relação do gene *AtDNMT2* e as possíveis vias em que possa estar atuando, seja pela regulação traducional via metilação de tRNA, metilação de outros RNAs, metilação de DNA, outros tipos de reações enzimáticas em DNA e/ou RNA, ou ainda, atuando como co-fator, seria de grande valia a obtenção do transcriptoma da planta mutante *Atdnmt2*.

#### **4.8. DNMT2 e a família HD2**

As proteínas da família HD2 e DNMT2 interagem no núcleo e, uma das hipóteses é que elas estejam atuando conjuntamente em vias epigenéticas, desempenhando papéis no desenvolvimento de plantas e na resposta a estresses (Song et al., 2010). A fim de averiguar uma possível interação molecular entre *AtDNMT2* e *AtHD2*, os níveis transcricionais dos quatro membros da família *AtHD2* foram avaliados nas plântulas do mutante *Atdnmt2* e Col-0, 48 horas após período de estratificação. Em condições de laboratório, com controle de fotoperíodo, temperatura e nutrição, de 24 a 72 horas após estratificação das sementes, as mesmas devem completar o processo germinativo (Boyes et al., 2001 Zanten et al., 2011). Os diferentes estádios de desenvolvimentos escolhidos devem representar o conjunto de genes ativos durante a emergência da raiz, 48 horas, e durante o surgimento do hipocótilo e cotilédones, 72 horas, a fim de averiguar possíveis diferenças na resposta de *AtDNMT2* nas diferentes fases iniciais do desenvolvimento.

Os dados demonstram que, entre os quatro membros da família HD2, apenas o gene *AtHD2A* é induzido no mutante *dnmt2*, quando comparado com Col-0. Os outros três membros da família não apresentaram alteração transcricional significativa nestas condições (Figura 10). Esse dado sugere que *AtDNMT2*, direta ou indiretamente, afeta a expressão especificamente do gene *AtHD2A*.



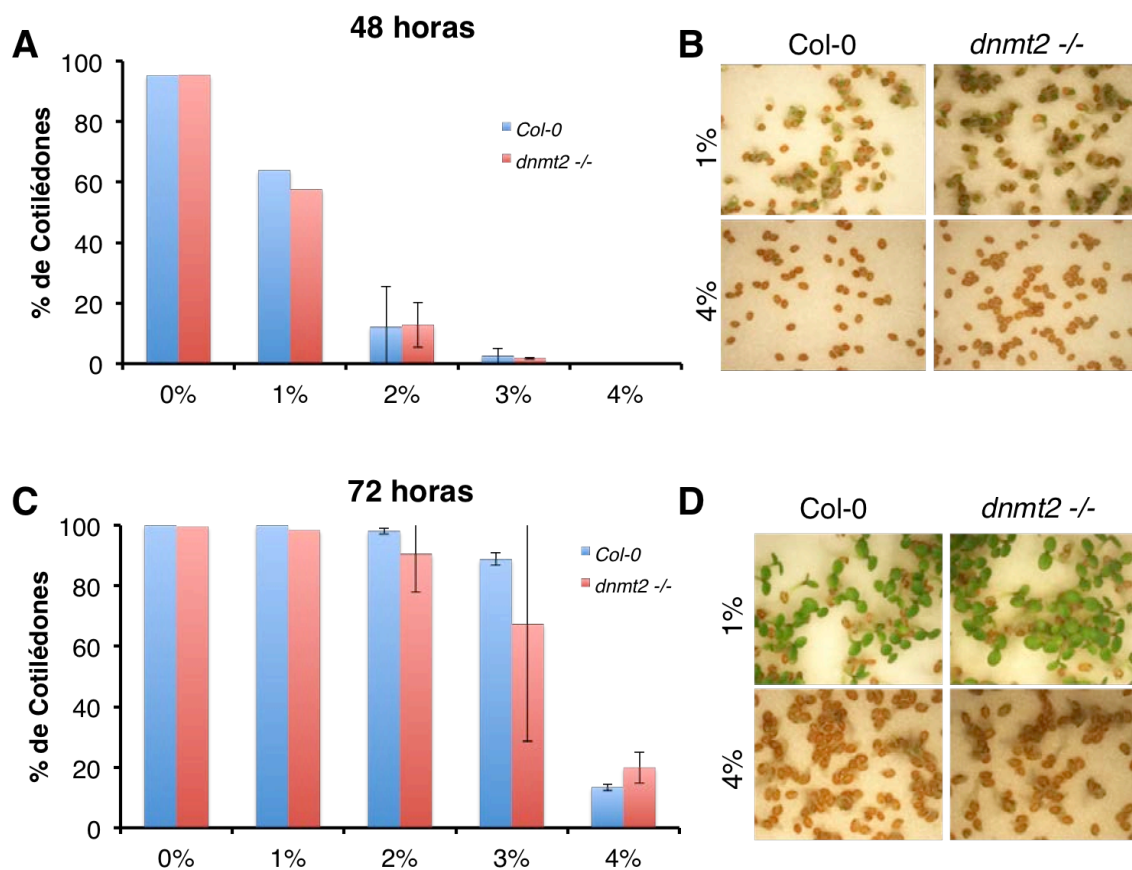
**Figura 10:** Análise de expressão dos genes *AtHD2A*, *AtHD2B*, *AtHD2C*, *AtHD2D* em plântulas *dnmt2* e Col-0, crescidas em meio MS sólido, acrescido de 1% de sacarose, aproximadamente 20 mg de plântulas foram coletadas após 48 horas de crescimento. *AtPP2A* foi utilizado como controle endógeno. O Teste t de Student foi aplicado e considerado significativo (\*) para p-values  $\leq 0,05$ .

Embora ainda seja necessário avaliar a expressão do gene *AtDNMT2* nas plantas mutantes *Athd2a*, nossos resultados e de outros (Song et al., 2010) indicam que as proteínas AtDNMT2 e AtHD2A interagem molecularmente.

#### 4.9. Análises das taxas de germinação nas plantas Col-0, *dnmt2* e dos duplo mutantes HDA e DHA em diferentes concentrações de sacarose

Na busca de um possível envolvimento do gene *AtDNMT2* nas fases iniciais da germinação e/ou crescimento da arabidopsis, foram realizados experimentos em que os genótipos Col-0 e *dnmt2* foram germinados em diferentes concentrações de sacarose (0%, 1%, 2%, 3%, 4%) e aqui, apenas as plântulas com cotilédones aparentes foram consideradas germinadas (Weigel e Glazebrook 2002). No primeiro experimento, foram utilizadas entre 250 e 300 sementes de cada genótipo, todas as concentrações de sacarose foram utilizadas, enquanto que no segundo, foram utilizadas entre 100 e 200 sementes de cada genótipo e apenas as concentrações de 2%, 3% e 4% foram empregadas e as concentrações de 0% e 1% não foram utilizadas porque as sementes de ambos os genótipos apresentaram germinação de aproximadamente 100% das sementes tanto após 24 horas, quanto após 48 horas. Os cotilédones foram contados 24, 48, 72 e 96 horas após colocadas para germinar, conforme descrito no item 3.1. Observou-se que em todas as concentrações de sacarose, em ambos os experimentos, as sementes iniciaram o processo de germinação somente após 24 horas. Nas placas sem sacarose (0%), após 48 horas, aproximadamente 95% das sementes de ambos os

genótipos haviam germinado e com 72 horas, 100% das sementes haviam germinado. Nas placas contendo 1% de sacarose, após 24 horas, aproximadamente 60% das sementes de ambos os genótipos haviam germinado e com 72 horas, 100% das sementes Col-0 e 98% das sementes *dnmt2* haviam germinado. Nos experimentos contendo 2% de sacarose, após 48 horas, ambos os genótipos tiveram em média 12% de sementes germinadas, enquanto que após 72 horas, em média, mais de 90% das sementes de ambos os genótipos estavam germinadas (Fig. 11). Nos experimentos contendo 3% de sacarose, aproximadamente 2% das sementes de ambos os genótipos germinaram após 48 horas. Após 72 horas, cerca de 88% das sementes Col-0 e 67% das sementes *Atdnmt2* germinaram. Finalmente, no experimento contendo 4% de sacarose, nenhuma semente de ambos os genótipos germinou após 48 horas e 13% das sementes Col-0 e 19% das *Atdnmt2* germinaram após 72 horas (Fig. 11). Em geral, os dados demonstram taxas similares de germinação das sementes Col-0 e do mutante *Atdnmt2*. Tais observações indicam que a não atividade da enzima DNMT2 não afeta a taxa de germinação de sementes de arábido. Entretanto, é possível que a enzima AtDNMT2 atue como cofator com outros fatores epigenéticos que afetam a germinação e crescimento. Um desses fatores pode ser a enzima AtHD2A (Zhou et al., 2004).

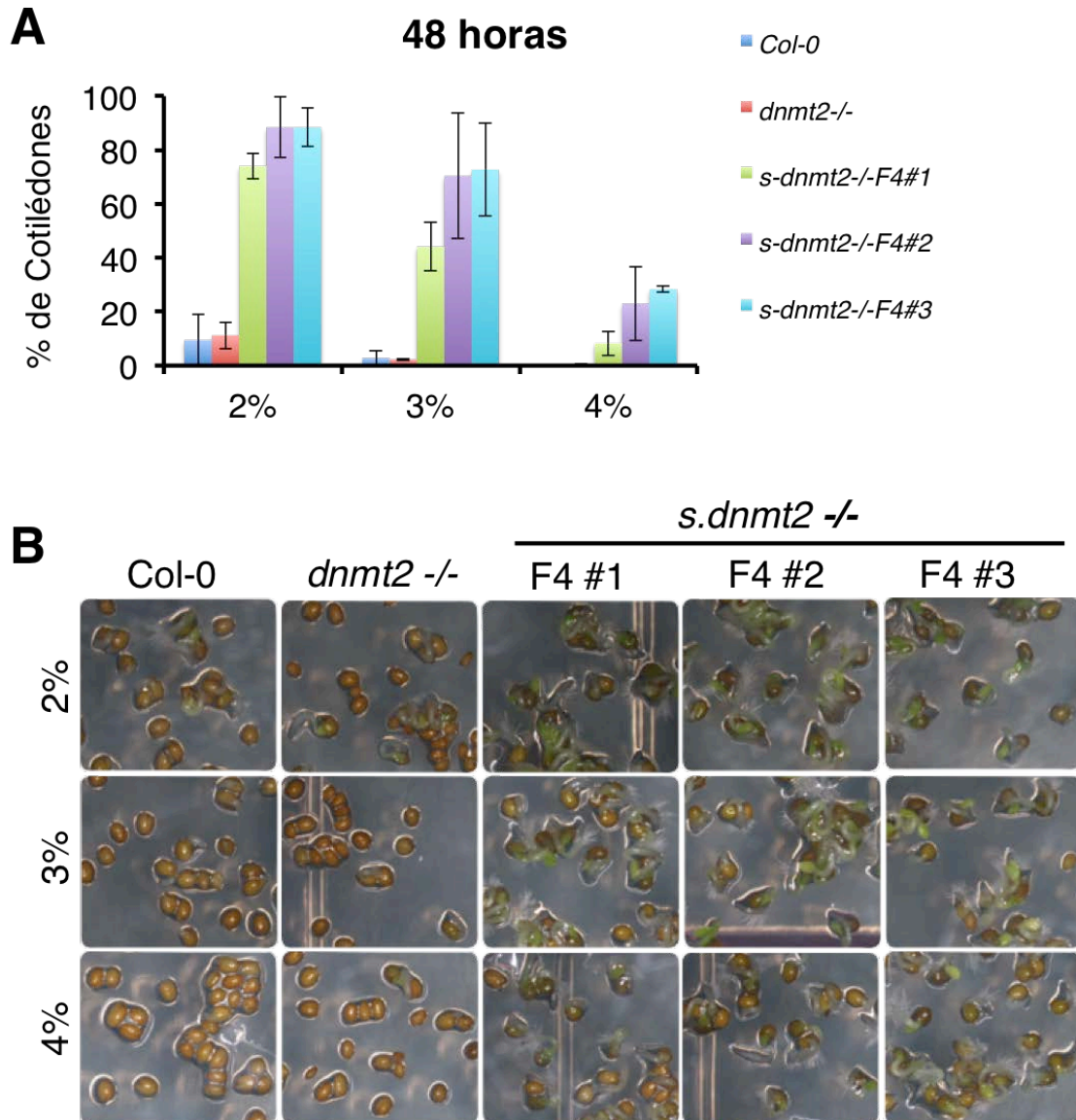


**Figura 11:** Taxa de germinação de sementes Col-0 e plantas mutantes *Atdnmt2* em concentrações de sacarose de 0%, 1%, 2%, 3% e 4%. **(A)** Porcentagem de cotilédones em ambos os genótipos em plântulas com 48 horas de crescimento. **(B)** Imagem de ambos os genótipos crescidos em 1% e 4% de sacarose, com 48 horas de crescimento. **(C)** Porcentagem de cotilédones em ambos os genótipos em plântulas com 72 horas de crescimento. **(D)** Imagem de ambos os genótipos crescidos em 1% e 4% de sacarose, com 72 horas de crescimento.

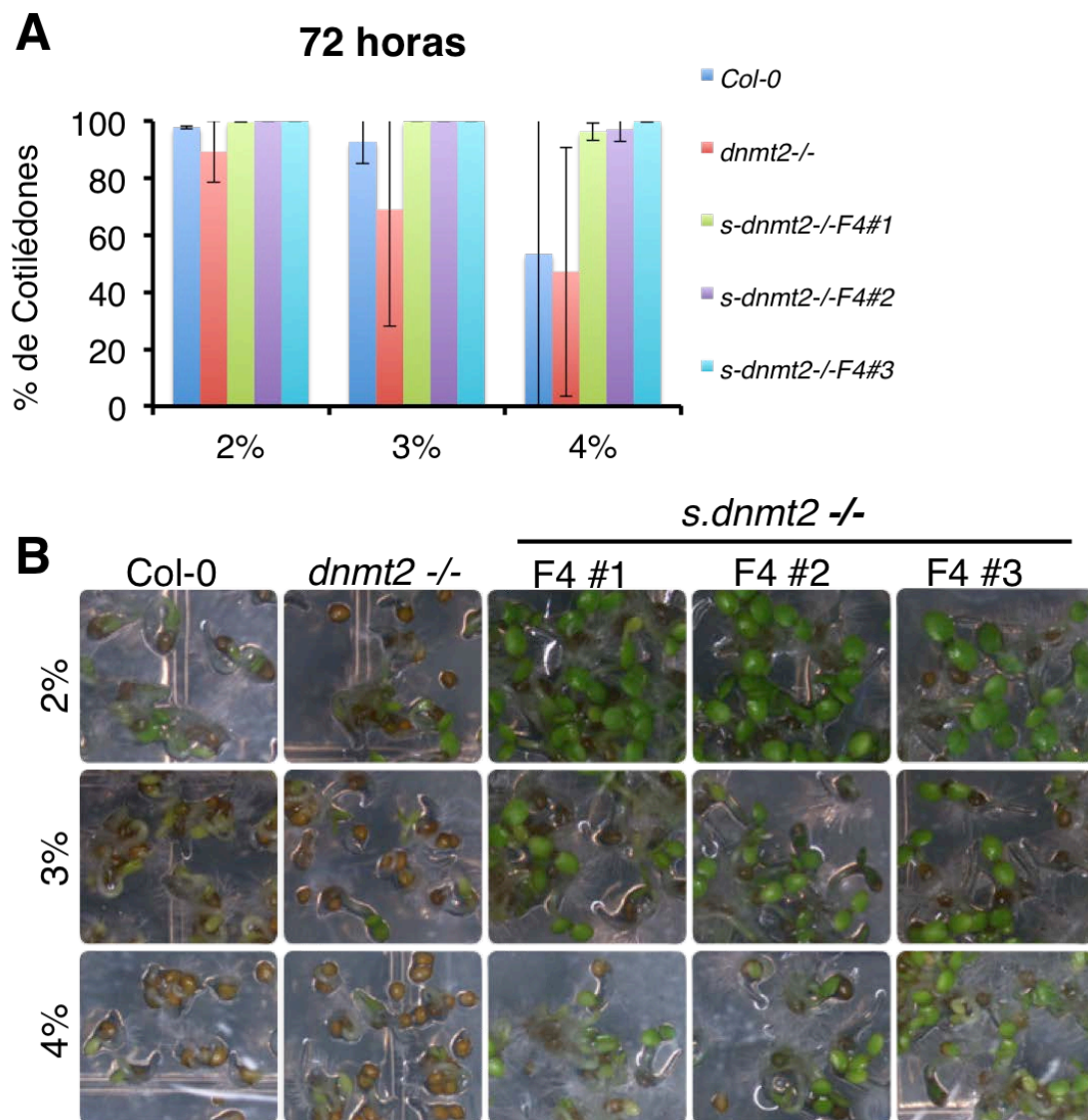
Recentemente foi demonstrado que *AtDNMT2* interage diretamente com enzimas *AtHD2*, incluindo a *AtHD2A*. Tal interação parece ser importante para o controle transcricional gênico (Song et al., 2010). Com o intuito de avaliar possível interação genética entre *AtDNMT2* e *AtHD2A* durante a germinação de sementes, plantas duplo-mutantes recíprocos *Atdnmt2;Ahd2a* foram gerados (DHA e HDA). Linhagens de duplo-mutantes, parentais e mutantes simples, provenientes do mesmo cruzamento, foram selecionados e avançados até a quarta geração (F4).

Inicialmente, sementes de três linhagens do mutante simples *s4.dnmt2* (vide Metodologia) da quarta geração (F4), oriundas do cruzamento, assim como as sementes Col-0 e *Atdnmt2*, foram germinadas em diferentes concentrações de sacarose a fim de avaliar possíveis alterações na germinação e/ou crescimento entre os genótipos. Entre 150 e 250 sementes de cada genótipo e de cada linhagem foram utilizadas em duas replicatas técnicas do experimento. As concentrações de 2%, 3% e

4% de sacarose foram empregadas e as sementes germinadas foram contadas 48 e 72 horas após colocadas para germinar, seguindo o mesmo critério utilizado anteriormente. Após 48 horas, foi observado que nos experimentos contendo 2% de sacarose, os genótipos Col-0 e *Atdnmt2* tiveram uma média de germinação de aproximadamente 10%, enquanto que a linhagem s4.dnmt2 #1 apresentou uma média de 74% de sementes germinadas e as linhagens s4.dnmt2 #2 e s4.dnmt2 #3 apresentaram uma média de germinação de aproximadamente 88% (Fig. 11). Adicionando-se 3% de sacarose ao meio MS, observou-se média de germinação de aproximadamente 2,5% nos genótipos Col-0 e *dnmt2*, 44% na linhagem s4.dnmt2 #1 e 71% nas linhagens s4.dnmt2 #2 e s4.dnmt2 #3. Já adicionando-se 4% de sacarose ao meio de cultura, nenhuma semente Col-0 germinou e apenas 0,3% das sementes *dnmt2* germinaram, enquanto que a linhagem s4.dnmt2 #1 apresentou 8% de germinação e as linhagens s4.dnmt2 #2 e s4.dnmt2 #3 de 15% (Fig. 12). Observou-se que após 72 horas nos experimentos contendo 2% de sacarose, as sementes Col-0 apresentaram média de 98% de germinação e de sementes da *Atdnmt2* de aproximadamente 90%, enquanto que as linhagens s4.dnmt2 #1, s4.dnmt2 #2 e s4.dnmt2 #3 apresentaram média de germinação de aproximadamente 100%. Após adicionar 3% de sacarose ao meio MS, observou-se média de germinação de aproximadamente 92% no genótipo Col-0, 70% no genótipo *dnmt2* e de 100% nas linhagens s4.dnmt2 #1, s4.dnmt2 #2 e s4.dnmt2 #3. Finalmente, em meio de cultura MS contendo 4% de sacarose, média de aproximadamente 50% de sementes Col-0 e *Atdnmt2* germinaram, enquanto que a linhagem s4.dnmt2 #1 apresentou uma média de 96% de germinação, a linhagem s4.dnmt2 #2, 97% e s4.dnmt2 #3 aproximadamente de 100% (Fig.13).



**Figura 12:** Taxa de germinação de sementes Col-0, *Atdnmt2* e sementes F4 *s4.dnmt2* #1, *s4.dnmt2* #2 e *s4.dnmt2* #3 em concentrações de sacarose de 2%, 3% e 4%, com 48 horas de crescimento. **(A)** Porcentagem de cotilédones em todos os genótipos. **(B)** Imagem dos genótipos crescidos em 2%, 3% e 4% de sacarose.



**Figura 13:** Taxa de germinação de sementes Col-0, *Atdnmt2* e sementes F4 *s4.dnmt2* #1, *s4.dnmt2* #2 e *s4.dnmt2* #3 em concentrações de sacarose de 2%, 3% e 4%, com 72 horas de crescimento. (A) Porcentagem de cotilédones em todos os genótipos. (B) Imagem dos genótipos crescidos em 2%, 3% e 4% de sacarose.

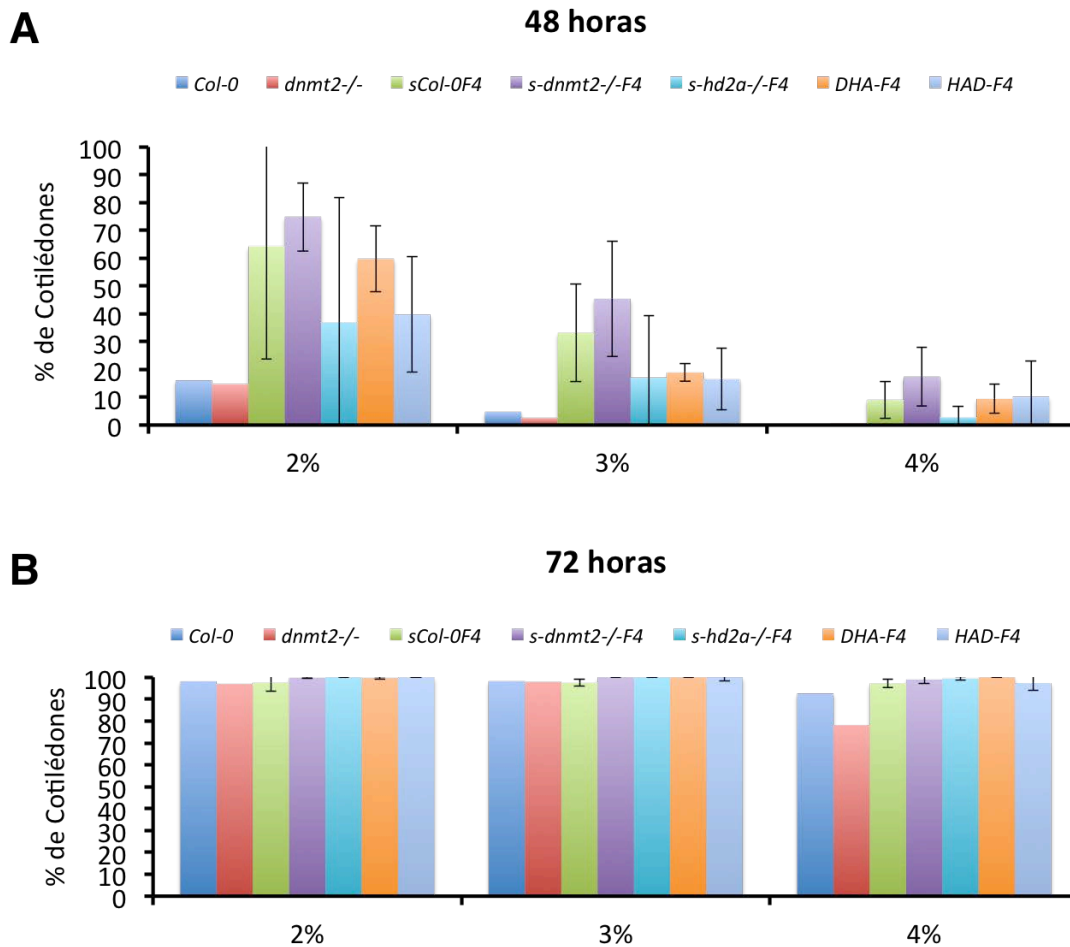
Posteriormente, sementes na quarta geração (F4) de cinco genótipos, oriundas do mesmo cruzamento (três linhagens *s4.Col-0*, três linhagens *s4.dnmt2*, duas linhagens *s4.hd2a*, três linhagens HAD e três linhagens DHA), além de sementes Col-0 e *Atdnmt2*, foram germinadas em 2%, 3% e 4% de sacarose a fim de avaliar possíveis alterações na germinação e/ou crescimento entre os genótipos. As sementes germinadas foram contadas 48 e 72 horas após colocadas para germinar seguindo mesmo critério utilizado anteriormente. Até o momento foi realizado apenas uma



replicata técnica e as linhagens foram contadas separadamente, porém o resultado apresenta as linhagens de cada genótipo agrupadas.

Após 48 horas, no experimento contendo 2% de sacarose, foi observada germinação de aproximadamente 15% para as sementes Col-0 e *Atdnmt2*. Para os genótipos s4.Col-0, s4.dnmt2, s4.hd2a, DHA e HAD, a média da porcentagem de germinação foi de aproximadamente 65%, 74%, 37%, 60% e 40%, respectivamente. Na concentração de 3% de sacarose, a germinação foi em média 5% para as sementes Col-0 e 3% para as sementes *Atdnmt2*. Para os genótipos s4.Col-0, s4.dnmt2, s4.hd2a, DHA e HAD, a média da porcentagem de germinação foi de 33%, 45%, 17%, 18% e 16%, respectivamente (Fig. 14A). Finalmente, quando as sementes foram incubadas em meio MS contendo 4% de sacarose, não houve germinação do genótipo Col-0, aproximadamente 0.66% para o genótipo *Atdnmt2*. Já para os genótipos s4.Col-0, s4.dnmt2, s4.hd2a, DHA e HAD, a média da porcentagem de germinação foi de 9%, 17%, 3%, 9% e 10%, respectivamente (Fig. 14A).

Finalmente, após 72 horas, com exceção das sementes Col-0 e *Atdnmt2* na placa contendo 4% de sacarose, a qual foi notada germinação de 92% e 78% respectivamente, todos os demais experimentos e genótipos tiveram taxa de germinação que variou entre 97 e 100% (Figura 14B).



**Figura 14:** Taxa de germinação de sementes Col-0, *Atdnmt2* e sementes F4 s4.Col-0 (3 linhagens diferentes), s4.dnmt2 (3 linhagens diferentes), s4.hd2a (2 linhagens diferentes), DHA (3 linhagens diferentes), HAD (3 linhagens diferentes) em concentrações de sacarose de 2%, 3% e 4%, com 72 horas de crescimento. **(A)** Porcentagem de cotilédones em todos os genótipos após 48 horas de crescimento. **(B)** Porcentagem de cotilédones em todos os genótipos após 72 horas de crescimento.

A transcrição dos genes em muitos eucariotos é regulado pela metilação de DNA e modificação de histonas. A reprogramação epigenética se dá ao apagar as marcas epigenéticas entre as gerações, entretanto, existem diferenças nesse processo entre animais e plantas, provavelmente devido as suas diferenças na biologia reprodutiva (Kawashima e Berger, 2014).

Em humanos, RNA mensageiros paternos podem ser encontrados no zigoto durante a fertilização. Acreditava-se que estes mRNAs fossem sobras da espermatogênese, mas já se sabe que essas moléculas são importantes para as fases

iniciais do desenvolvimento (Ostermeir et al., 2004). E ainda, novas descobertas sugerem que DNMT2 é requerida apenas durante a gametogênese paterna ou nas fases iniciais do desenvolvimento de camundongos (Kiani et al., 2013).

Em mamíferos, as linhagens germinativas são definidas antes da meiose, em um estágio inicial da embriogênese, enquanto que em plantas, a diferenciação das linhagens germinativas ocorre depois da meiose. Durante o início da embriogênese das linhagens germinativas dos mamíferos, ocorre a primeira demetilação global do genoma e depois, a segunda demetilação global ocorre no zigoto, após a fecundação. Contudo, o mecanismo responsável pela reprogramação epigenética continua desconhecido (Kawashima e Berger, 2014).

Apesar de no reino vegetal alguns genes, transgenes e elementos como TEs (elementos de transposição) permanecerem metilados entre as gerações e, aparentemente não ocorrer a reprogramação epigenética, análises globais e moleculares detalhadas do epigenoma, demonstraram que a reprogramação epigenética ocorre durante a reprodução sexuada de plantas (Kawashima e Berger, 2014).

Os resultados apresentados nesse trabalho sugerem taxa de germinação similar de sementes oriundas de planta selvagem (Col-0) e do mutante *Atdnmt2* (Fig. 11). Entretanto, os resultados evidenciam diferenças entre as taxas de germinação das sementes das linhagens provenientes do cruzamento *Atdnmt2 X Athd2a*, quando comparadas as taxas de germinação de sementes Col-0 e do mutante *dnmt2*. *AtHD2A* interfere em processos germinativos (Colville et al., 2011) e a falta da *AtDNMT2* parece aumentar a taxa de germinação após quatro gerações de autofecundação. Visto que os genomas maternos e paternos contribuem igualmente para o transcriptoma dos embriões de plantas (Nodine e Bartel, 2012), é razoável concluir que o aumento na taxa de germinação se dá por efeitos epigenéticos. Ambas as enzimas *AtDNMT2* e *AtHD2A* devem afetar a gametogênese masculina ou os estádios iniciais da embriogênese. Essa observação está de acordo com Kiani e colaboradores (2013), os quais sugerem o envolvimento da enzima DNMT2 em eventos iniciais da embriogênese de camundongos e com relatos que indicam que a enzima HD2A de arabis é um regulador fino de germinação de sementes (Zhou, et al., 2004; Colville et al., 2011).

Uma possível explicação para a germinação e/ou crescimento acelerado do embrião em todos os genótipos das plântulas provenientes dos mesmos cruzamentos entre os mutantes *Atdnmt2* e *Athd2a* pode ser o fato de ambas as enzimas, DNMT2 e

HD2A, estarem atuando conjuntamente a fim de regular negativamente alguns *loci*, os quais podem ser genes ou ainda elementos de transposição inativos nas fases iniciais do desenvolvimento. Uma vez que a característica aparece em todos os descendentes do mesmo cruzamento, é possível que a mesma seja mantida em diferentes gerações por consequência da manutenção da acetilação de histonas e da deficiência na função da AtDNMT2 (metilação de RNA e/ou cofator da enzima AtHD2A).

Modificações de histonas são parcialmente herdadas através de divisões celulares e a reprogramação epigenética deve ocorrer na fertilização (Kawashima e Berger, 2014). É possível que a enzima AtDNMT2 possua outros alvos além de tRNAs ou atue na regulação do gene *AtHD2A*, ou ainda desempenhe papel de cofator juntamente com a AtHD2A, a fim de remover acetilação de histonas. Como resultado, alguns *loci* que normalmente estão desativados podem ser ativados ou parcialmente ativados.

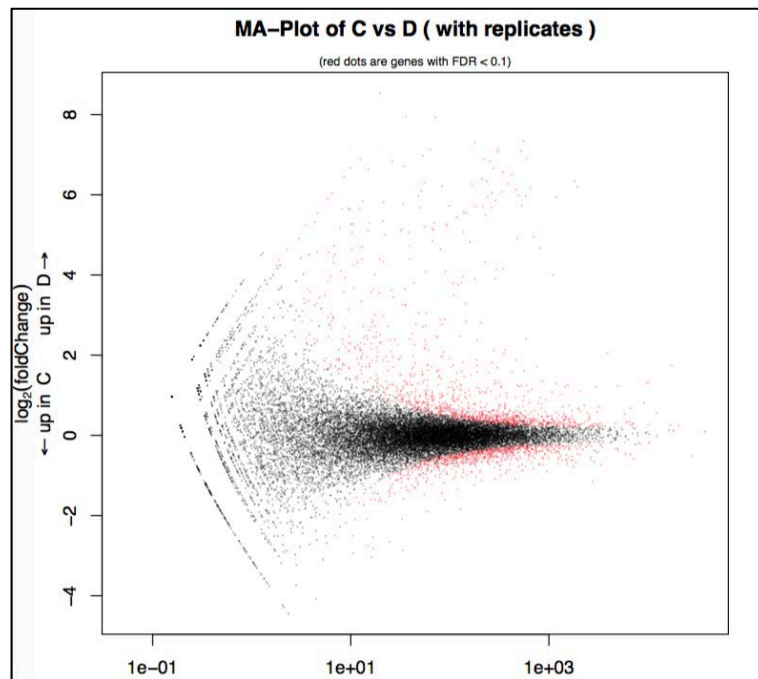
Além disso, o perfil de metilação de pólen indica mudança dinâmica na metilação do DNA durante a gametogênese masculina, porém, ainda são desconhecidas as enzimas responsáveis pela metilação do DNA nesse fase. A metilação não CpG aumenta nas células vegetativas do grão de pólen em regiões ricas em elementos de transposição, provavelmente com a finalidade de regular tais elementos (Borges et al., 2012). Normalmente, retrotransposons LTR (repetições terminais longas) são regulados pela via de RdDM – Metilação de DNA direcionada por RNA, na qual os transcritos dependentes de RNA polimerase IV são convertidos em RNA fita-dupla pela RDR2 (RNA dependente de RNA polimerase II) e processadas pela DCL3 (Dicer like 3) em siRNAs de 24 nucleotídeos (Zhong et al, 2014). Baixos níveis de siRNAs correspondentes aos retrotransposons LTR são encontrados em células vegetativas do grão de pólen. Entretanto, siRNAs de 21 e não 24 nucleotídeos são encontrados nessa fase. O modelo corrente é de que vias não canônicas RdDM assumem o papel nesse estágio a fim de controlar esses elementos. Como a DNMT2 aumenta sua expressão durante a gametogênese (Fig. 2), é possível especular papel da DNMT2 no controle destes elementos.

#### 4.10. Avaliação de alterações transcricionais na planta mutante *Atdnmt2*

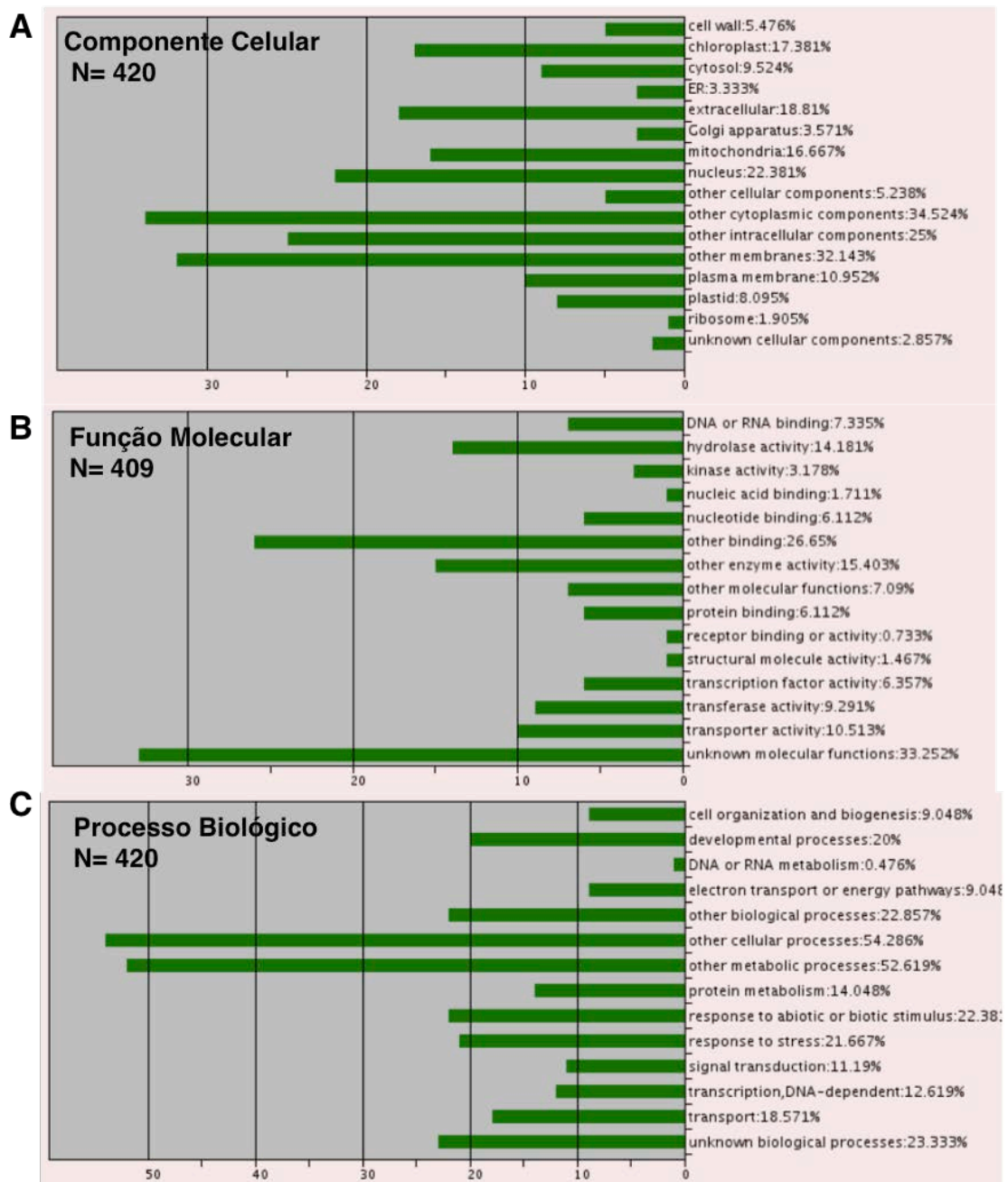
A fim de avaliar quais vias são desreguladas na ausência da enzima funcional AtDNMT2, bibliotecas de mRNA (RNA-seq) foram preparadas a partir de tecidos de plântulas de arábido após 48 horas de crescimento e avaliadas comparando-se os genótipos Col-0 e *Atdnmt2*.

Os dados gerados pelas bibliotecas mostram poucos genes desregulados, sendo que, aproximadamente 1,35% dos genes da planta mutante *Atdnmt2* estão positivamente regulados em relação à planta Col-0 e, apenas 0,26% dos genes da planta Col-0 estão positivamente regulados em relação à planta mutante *Atdnmt2*. Os dados podem ser visualizados no gráfico da Figura 15.

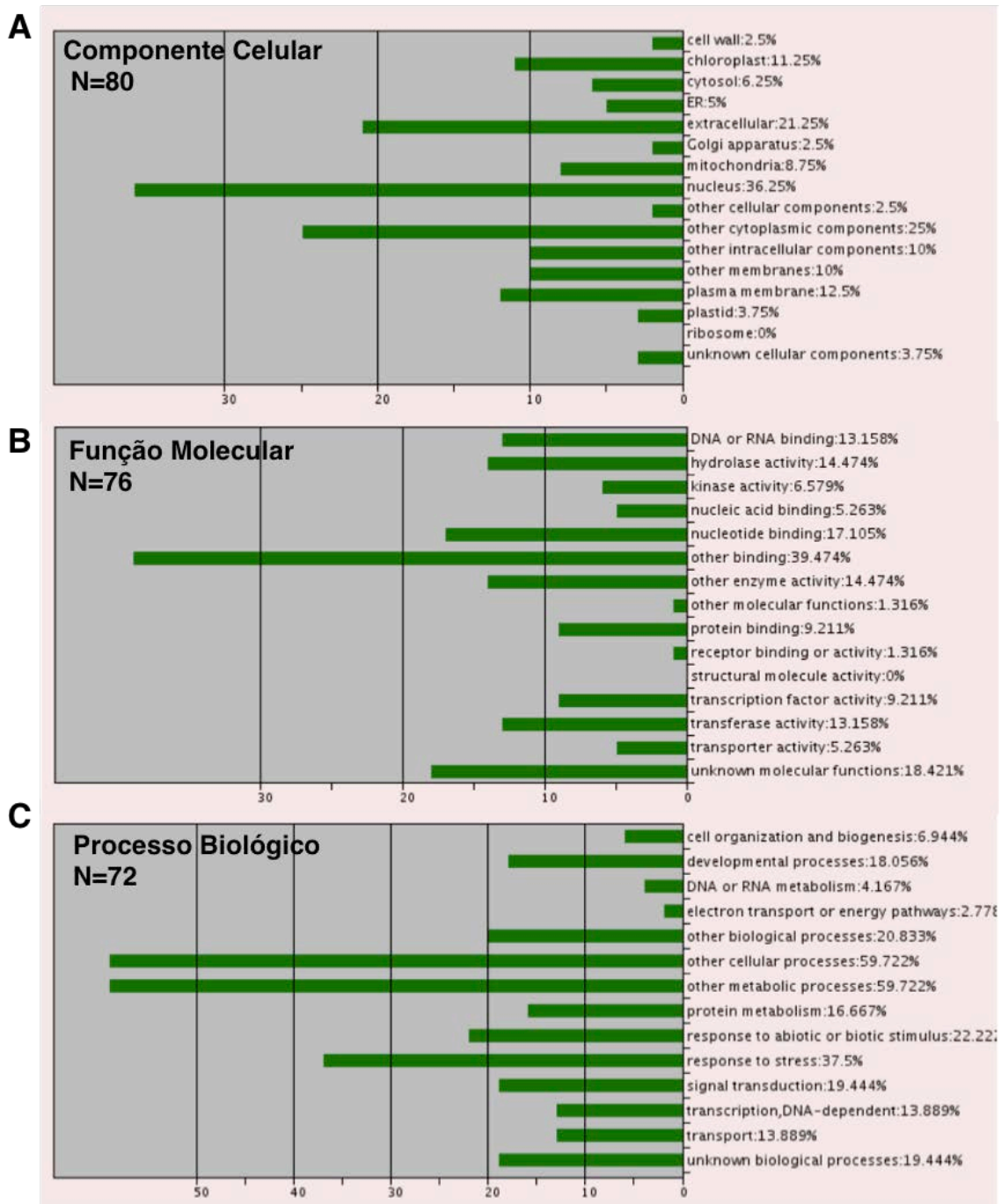
Os genes desregulados indicam um papel mais abrangente e, em sua maioria, ligados às funções básicas da planta (Figs. 16 e 17), como previamente hipotetizado no presente trabalho.



**Figura 15:** Col-0 (C) x *Atdnmt2* (D) MA-Plot, com quatro replicatas biológicas cada, os pontos vermelhos são genes com FDR < 0.1 (taxa de falso descobrimento), os dados foram gerados pela plataforma Galaxy.



**Figura 16:** Genes positivamente regulados na planta mutante *Atdnmt2*. Os gráficos foram gerados pela ferramenta *Bulk Data Retrieval* (<https://www.arabidopsis.org/tools/bulk/go/index.jsp>), onde cada gene pode ser classificado em mais de uma categoria, de acordo com sua anotação. **(A)** Categoria de Componentes celulares, contendo 420 genes. **(B)** Categoria de Função Molecular, com 409 genes. **(C)** Categoria de Processos Biológicos, com 420 genes.



**Figura 17:** Genes positivamente regulados na planta Col-0. Os gráficos foram gerados pela ferramenta *Bulk Data Retrieval* (<https://www.arabidopsis.org/tools/bulk/go/index.jsp>), onde cada gene pode ser classificado em mais de uma categoria, de acordo com sua anotação. **(A)** Categoria de Componentes celulares, contendo 80 genes. **(B)** Categoria de Função Molecular, com 76 genes. **(C)** Categoria de Processos Biológicos, com 72 genes.

A grande maioria dos genes desregulados no mutante *Atdnmt2* ainda são desconhecidos. Entretanto, algumas informações se destacam, como por exemplo: aproximadamente 17% e 16% dos genes positivamente regulados na planta mutante *Atdnmt2* pertencem ao cloroplasto e à mitocôndria, respectivamente, e aproximadamente 9% dos genes pertencem ao transporte de elétrons ou vias energéticas, enquanto que 53% dos genes estão envolvidos com processos metabólicos. Aproximadamente 22% desses genes são componentes do núcleo, 34,5% do citoplasma e 32% de outras membranas ainda não identificadas; 14% dos genes fazem parte do metabolismo de proteínas e apenas 0,5% no metabolismo de DNA ou RNA. Cerca de 22% e 21% respondem à estímulos bióticos e abióticos e à estresse, respectivamente, 11% na transdução de sinal, 18,5% em transporte, 12,6% na transcrição dependente de DNA e 20% em processos do desenvolvimento. Algumas das funções moleculares em que os genes na planta mutante estão positivamente regulados são: 10% atividade de transporte, 9% atividade de transferase, 6% atividade de fator de transcrição, 14% atividade de hidrólise e 3% atividade quinase, mas a grande maioria das funções moleculares envolvem genes ainda não caracterizados.

Os genes positivamente regulados na planta Col-0 que se destacam, são componentes do núcleo (36%), cloroplasto (11%), mitocôndria (9%), citoplasma (25%) e membrana plasmática (12,5%); 18% estão envolvidos com processos do desenvolvimento, 22% resposta à estímulos bióticos e abióticos, 37,5% resposta à estresse, 20% na transdução de sinal, 60% processos metabólicos, 16,6% e 4% metabolismo de proteínas e DNA ou RNA, respectivamente, 14% transporte e 14% transcrição dependente de DNA. Algumas das funções moleculares em que os genes na planta selvagem (Col-0) estão positivamente regulados são: 5% atividade de transporte, 13% atividade de transferase, 9% atividade de fator de transcrição, 15% atividade de hidrólise e 6% atividade quinase, porém a grande maioria das funções moleculares também envolvem genes ainda não caracterizados.

Até o momento, não há informações na literatura sobre bibliotecas de RNA mensageiro de mutantes *DNMT2* em outros organismos. O fato de encontrarmos genes desregulados que fazem parte do núcleo e citoplasma, corroboram o resultado da localização celular, apresentado no início do presente trabalho, no qual pode-se visualizar a mRFP fusionada com a AtDNMT2 tanto no núcleo das células de tabaco quanto no citoplasma.



A hipótese apresentada previamente no presente trabalho, de que *AtDNMT2* estaria envolvido com metabolismo básico celular, por estar presente em praticamente todos os tecidos de forma ubíqua (Fig. 2), também se confirma, pois grande parte dos genes desregulados estão de fato envolvidos com algum tipo de metabolismo, entre eles, metabolismo de proteínas ou de ácidos nucleicos.

A exata função da DNMT2 em situações de estresse permanece incompreendida, porém está cada vez mais clara sua relação, como já concluído por outros autores (Fisher et al., 2006; Tovy et al., 2006; Schaefer et al., 2010; Thiagarajan et al., 2011; Becker et al., 2012; Mytych et al., 2014). Aqui, diversos genes envolvidos, principalmente com a resposta estresse de calor são desregulados na planta mutante *Atdnmt2*, assim como resposta à estresse salino, injúria, oxidativo, luz, patógenos, resposta à ABA entre outros. O fato desses genes serem desregulados na planta mutante corrobora os dados descritos previamente (Figs. 7 e 9), em que observa-se um aumento na expressão de *AtDNMT2* em condições de estresse oxidativo e condições não fisiológicas. Futuros trabalhos deverão ser desenvolvidos a fim de desvendar a atividade da DNMT2 nessas vias, seja por metilação de RNA ou DNA, sinalização, cofator ou outra atividade enzimática, como por exemplo a deaminase de ácido nucleicos (Shen et al., 1992; Liutkeviciutė et al., 2009).

Uma informação extremamente interessante e que pode vir a confirmar hipóteses levantadas por Kiani e colaboradores (2013) é que alguns genes positivamente regulados na planta mutante envolvidos com desenvolvimento embrionário, caracterizados ou não, são expressos apenas em estruturas masculinas (diferentes estádios), como os genes AT4G34600 (envolvido no crescimento do tubo polínico), AT5G12950 (envolvido em diversos estádios de desenvolvimento do pólen), AT5G54740 (envolvido no desenvolvimento do grão de pólen) e apenas um gene positivamente regulado na planta Col-0, é encontrada em ovário, o AT3G50330 (envolvido na formação do carpelo e desenvolvimento do septo do ovário). Será interessante avaliar a expressão destes genes nas plantas provenientes do cruzamento realizado previamente neste trabalho (HAD, DHA, *s.dnmt2* e *s.hd2a*) a fim de verificar se estes genes estariam de fato interferindo no tempo de germinação e/ou crescimento das sementes avaliadas (Fig. 14). Não foram encontrados *loci* correspondentes à elementos de transposição, uma das hipóteses sugeridas previamente neste trabalho, para este atraso na germinação e/ou crescimento das sementes vindas do cruzamento entre a planta mutante *Atdnmt2* e *Athd2a*. Contudo, diversos outros genes encontrados

nas bibliotecas poderiam justificar um atraso na germinação e/ou crescimento das sementes geradas previamente, e, a realização de futuros experimentos, levando em consideração os dados gerados até o momento devem ajudar responder à estas questões.

## 5. Conclusões

- A enzima AtDNMT2 está localizada tanto no núcleo como no citoplasma.
- Não há diferenças fenotípicas entre *Atdnmt2* e Col-0 nas fases embrionárias, após 24 horas embebidas e plântulas com 10 dias de crescimento;
- O gene *AtDNMT2* é positivamente regulado durante diferentes tipos de estresses.
- A enzima AtDNMT2 não atua na proteção do tRNA Asp<sup>GTC</sup> durante estresse oxidativo nas condições experimentais do presente trabalho;
- O gene *AtHD2A* parece ser regulado negativamente pela enzima AtDNMT2.
- As enzimas AtDNMT2 e AtHD2A parecem controlar conjuntamente gene(s) envolvido(s) na germinação e/ou crescimento inicial de *Arabidopsis thaliana*.
- O gene *AtDNMT2* pode desempenhar papel indireto na gametogênese masculina.

## 6. Referências Bibliográficas

- ABMAYR, S. M.; WORKMAN, J. L. Holding on through DNA replication: Histone modification or modifier? *Cell*, v. 150, p. 875-877, 2012.
- AHMAD, F. et al. Comprehensive gene expression analysis of the DNA (cytosine-5) methyltransferase family in rice (*Oryza sativa L.*). *Genetics and Molecular Research*, v. 13, n. 3, p. 5159-5172, 2014.
- ALBALAT, R. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus *Branchiostoma floridae*. *Development Genes and Evolution*, v. 218, n. 11-12, p. 691-701, 2008.
- ALONSO, J. M. et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, v. 301, n. 5633, p. 653-657, 2003.
- ANDERSEN, N. M.; DOUTHWAITE, S. YebU is a m<sup>5</sup>C methyltransferase specific for 16 S rRNA nucleotide 1407. *Journal of Molecular Biology*, v. 359, p. 777-786, 2006.
- AUXILIEN, S. et al. Archease from *Pyrococcus abyssi* improves substrate specificity and solubility of a tRNA m<sup>5</sup>C methyltransferase. *Journal of Biological Chemistry*, v. 282, n. 26, p. 18711-18721, 2007.
- BANERJEE, A. K. 5-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiology Review*, v. 44, p. 175-205, 1980.
- BANERJEE, S. et al. Entamoebahistolytica DNA methyltransferase (EhMeth) is a nuclear matrix protein that binds EhMRS2, a DNA that includes a scaffold/matrix attachment region (S/MAR). *Molecular and Biochemical Parasitology*, v. 139, n. 1, p. 91-97, 2005.
- BARTEE, L.; MALAGNAC, F.; BENDER, J. Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes and Development*, v. 15, n. 410, p. 1753-1758, 2001.
- BECKER, M. et al. Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Research*, v. 40, n. 22, p. 11648-11658, 2012.
- BERGER, S. L. et al. An operational definition of epigenetics An operational definition of epigenetics. *Genes & Development*, v. 23, p. 781-783, 2009.
- BERLETH, T.; JÜRGENS G. The role of the monopteros gene in organising the basal body region of the Arabidopsis embryo. *Development*, v. 118, p. 575-587, 1993.
- BESTOR, T. et al. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal*

- of Molecular Biology*, v. 203, p. 971-983, 1988.
- BIRD, A. DNA methylation patterns and epigenetic memory DNA methylation patterns and epigenetic memory. *Genes & Development*, v. 16, p. 6-21, 2002.
- BIRD, A. Perceptions of epigenetics. *Nature*, v. 447, p. 396-398, 2007.
- BONASIO, R.; TU, S.; REINBERG, D. Molecular signals of epigenetic states. *Science*, v. 330, p. 612-616, 2010.
- BORGES, F.; CALARCO, J. P.; MARTIENSSSEN, R. A. Reprogramming the epigenome in Arabidopsis pollen. *Cold Spring Harbor Symposia on Quantitative Biology*, v. 77, p. 1-5, 2012.
- BOYES, D. C. et al. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *The Plant Cell*, v. 13, n. 7, p. 1499-1510, 2001.
- BROSCH, G. et al. Purification and Characterization of a High Molecular Weight Histone Deacetylase. *Biochemistry*, v. 2960, n. 96, p. 15907-15914, 1996.
- CAO, D. et al. Genome-wide identification of cytosine-5 DNA methyltransferases and demethylases in *Solanumlycopersicum*. *Gene*, v. 550, n. 2, p. 230-237, 2014.
- CAO, X. et al. Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. *Current Biology*, v. 13, p. 2212-2217, 2003.
- CAO, X.; JACOBSEN, S. E. Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Current Biology*, v. 12, n. 2, p. 1138-1144, 2002.
- CHAN, S. W. L. et al. RNAi, DRD1, and histone methylation actively target developmentally important Non-CG DNA methylation in Arabidopsis. *PLoS Genetics*, v. 2, n. 6, p. 791-797, 2006.
- CHAN, S. W.-L. et al. RNA silencing genes control de novo DNA methylation. *Science*, v. 303, p. 1336, 2004.
- CHEN, L.-Q. et al. A Cascade of Sequentially Expressed Sucrose Transporters in the Seed Coat and Endosperm Provides Nutrition for the Arabidopsis Embryo. *The Plant Cell Online*, p. 1-13, 2015.
- CHÉDIN, F. The DNMT3 family of mammalian de novo DNA methyltransferases. *Progress in Molecular Biology and Translational Science*, v. 101, p. 255-285, 2011.
- CHENG, X.; BLUMENTHAL, R. M. Mammalian DNA Methyltransferases: A Structural Perspective. *Cell Structure*, v. 16, p. 341-350, 2008.

- CHOUHARY, C. et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, v. 325, p. 834-840, 2009.
- COKUS, S. J. et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, v. 452, p. 215-219, 2008.
- COLVILLE, A. et al. Role of HD2 genes in seed germination and early seedling growth in Arabidopsis. *Plant Cell Reports*, v. 30, n. 10, p. 1969-1979, 2011.
- COWLING, V. H. Regulation of mRNA cap methylation. *Biochemistry Journal*, v. 425, p. 295-302, 2010.
- CZERWONIEC, A. et al. MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Research*, v. 37, p. D118-121, 2009.
- CZERWONIEC, A. et al. Folds and functions of domains in RNA modification enzymes. In: GROSJEAN, H. (Ed.). *DNA and RNA modification enzymes: comparative structure, mechanism, functions, cellular interactions and evolution*. Austin, TX: Landes Bioscience, 2009. p. 289-302.
- DAFFIS, S. et al. 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*, v. 468, p. 452-456, 2010.
- DEMETRIOU, K. et al. Epigenetic chromatin modifiers in barley: I. Cloning, mapping and expression analysis of the plant specific HD2 family of histone deacetylases from barley, during seed development and after hormonal treatment. *Physiologia Plantarum*, v. 136, p. 358-368, 2009.
- DONG, A. et al. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Research*, v. 29, n. 2, p. 439-448, 2001.
- DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, v. 19, p. 11-15, 1987.
- DUARTE, G. T. et al. Involvement of microRNA-related regulatory pathways in the glucose-mediated control of Arabidopsis early seedling development. *Journal of Experimental Botany*, v. 64, n. 14, p. 4301-4312, 2013.
- DURDEVIC, Z. et al. The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. *Cell Reports*, v. 4, n. 5, p. 931-937, 2013.
- EHRlich, M. et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, v. 10, n. 8, p. 2709-2721, 1982.
- FINKELSTEIN, R. R.; GIBSON, S. I. ABA and sugar interactions regulating development: Cross-talk or voices in a crowd? *Current Opinion in Plant Biology*, v. 5, p. 26-32, 2002.

- FINNEGAN, E. J.; DENNIS, E. S. Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Research*, v. 21, n. 10, p. 2383-2388, 1993.
- FINNEGAN, E. J.; PEACOCK, W. J.; DENNIS, E. S. DNA methylation, a key regulator of plant development and other processes. *Current Opinion in Genetics & Development*, v. 10, n. 2, p. 217-223, 2000.
- FISHER, O.; SIMAN-TOV, R.; ANKRI, S. Pleiotropic phenotype in *Entamoebahistolytica* overexpressing DNA methyltransferase (Ehmet). *Molecular and Biochemical Parasitology*, v. 147, n. 1, p. 48-54, 2006.
- FOSTER, P. G. et al. The first structure of an RNA m<sup>5</sup>C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate. *Structure*, v. 11, p. 1609-1620, 2003.
- FROMMER, M. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, v. 89, p. 1827-1831, 1992.
- GAUER, L. Caracterização funcional do gene *Bzo2h2* de *Arabidopsis thaliana*: um regulador da transcrição homólogo ao locus *Opaco2* de milho. 2004. Tese (Doutorado em Genética e Biologia Molecular) - Instituto de Biologia, Universidade Estadual de Campinas, 2004.
- GOLDBERG, A. D.; ALLIS, C. D.; BERNSTEIN, E. Epigenetics: a Landscape Takes Shape. *Cell*, v. 128, p. 635-638, 2007.
- GOLL, M. G. et al. Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science*, v. 311, n. 5759, p. 395-398, 2006.
- GOLL, M. G.; BESTOR, T. H. Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry*, v. 74, p. 481-514, 2005.
- GREENBERG, M. V. C. et al. Identification of genes required for de novo DNA methylation in arabidopsis. *Epigenetics*, v. 6, p. 344-354, 2011.
- GROSJEAN, H. et al. RNomics and Modomics in the halophilic archaea *Haloferaxvolcanii*: identification of RNA modification genes. *BMC Genomics*, v. 9, p. 470, 2008.
- GUPTA, A.; SINGH, M.; LAXMI, A. Interaction between Glucose and Brassinosteroid during regulation of lateral root development in *Arabidopsis thaliana*. *Plant Physiology*, v. 168, n. 1, p. 307-320, 2015.
- HAKE, S. B.; XIAO, A.; ALLIS, C. D. Linking the epigenetic “language” of covalent histone modifications to cancer. *British Journal of Cancer*, v. 90, p. 761-769, 2004.

- HANSEN, J. C.; TSE, C.; WOLFFE, A. P. Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry*, v. 37, p. 17637-17641, 1998.
- HELM, M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Research*, v. 34, p. 721-733, 2006.
- HE, X.-J. et al. An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell*, v. 137, n. 3, p. 498-508, 2009.
- HENIKOFF, S.; COMAI, L. A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. *Genetics*, v. 149, p. 307-318, 1998.
- HERMANN, A.; SCHMITT, S.; JELTSCH, A. The Human Dnmt2 Has Residual DNA-(Cytosine-C5). *Biochemistry*, v. 278, n. 34, p. 31717-31721, 2003.
- HOTCHKISS, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *Journal of Biology and Chemistry*, n. 175, p. 315-332, 1948.
- HOPPER, A. K.; PHIZICKY, E. M. tRNA transfers to the limelight. *Genes & Development*, v. 17, p. 162-180, 2003.
- HORWICH, M. D. et al. The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Current Biology*, v. 17, p. 1265-1272, 2007.
- JAVELLE, M.; MARCO, C. F.; TIMMERMANS, M. In Situ Hybridization for Precise Localization of Transcripts in Plants. *Journal of Visualized Experiments*, v. 57, p. e3328, 2011.
- JELTSCH, A.; NELLEN, W.; LYKO, F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends in Biochemical Sciences*, v. 31, n. 6, p. 306-308, 2006.
- JONES, L.; RATCLIFF, F.; BAULCOMBE, D. C. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Current Biology*, v. 11, p. 747-757, 2001.
- JURKOWSKI, T. P.; JELTSCH, A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS One*, v. 6, n. 11, p. e28104, 2011.
- JURKOWSKI, T. P. et al. Mapping the tRNA binding site on the surface of human DNMT2 methyltransferase. *Biochemistry*, v. 51, n. 22, p. 4438-4444, 2012.
- JURKOWSKI, T. P. et al. Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA*, v. 14, n. 8, p. 1663-1670, 2008.



- KAWASHIMA, T.; BERGER, F. Epigenetic reprogramming in plant sexual reproduction. *Nature Reviews Genetics*, v. 15, p. 613-624, 2014.
- KIANI, J.; RASSOULZADEGAN, M. A load of small RNAs in the sperm - how many bits of hereditary information? *Cell Research*, v. 23, n. 1, p. 18-19, 2013.
- KATOH, M. et al. Developmentally Regulated DNA Methylation in *Dictyosteliumdiscoideum*. *Eukaryotic Cell*, v. 5, n. 1, p. 18-25, 2006.
- KHODDAMI, V.; CAIRNS, B. R. Transcriptome-wide target profiling of RNA cytosine methyltransferases using the mechanism-based enrichment procedure Aza-IP. *Nature Protocols*, v. 9, n. 2, p. 337-361, 2014.
- KHORASANIZADEH, S. The Nucleosome: From Genomic Organization to Genomic Regulation. *Cell*, v. 116, p. 259-272, 2004.
- KIM, Y. K. et al. Ribosomal protein s6, a target of rapamycin, is involved in the regulation of rRNA genes by possible epigenetic changes in arabidopsis. *Journal of Biological Chemistry*, v. 289, p. 3901-3912, 2014.
- KIM, S. W. et al. A sensitive non-radioactive northern blot method to detect small RNAs. *Nucleic Acids Research*, v. 38, n. 7, p. e98, 2010.
- KOUZARIDES, T. Acetylation: a regulatory modification to rival phosphorylation? *The EMBO Journal*, v. 19, n. 6, p. 1176-1179, 2000.
- KRAUSS, V.; REUTER, G. DNA methylation in *Drosophila*--a critical evaluation. *Progress in Molecular Biology and Translational Science*, v. 101, p. 177-191, 2011.
- LAW, J. A.; JACOBSEN, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, v. 11, p. 204-220, 2010.
- LAWRENCE, R. J. et al. A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Molecular Cell*, v. 13, p. 599-609, 2004.
- LEE, S. R.; COLLINS, K. Starvation-induced cleavage of the tRNA anticodon loop in *Tetrahymenathermophila*. *The Journal of Biological Chemistry*, v. 280, n. 52, p. 42744-42749, 2005.
- LEE, H. G.; LEE, K.; SEO, P. J. The Arabidopsis MYB96 transcription factor plays a role in seed dormancy. *Plant Molecular Biology*, v. 87, p. 371-381, 2015.
- LEONHARDT, H. et al. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, v. 71, n. 5, p. 865-873, 1992.
- LISTER, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, v. 462, n. 7271, p. 315-322, 2009.

- LINDROTH, A. et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science*, v. 292, p. 2077-2080, 2001.
- LINDROTH, A. M. et al. Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *The EMBO Journal*, v. 23, n. 21, p. 4286-4296, 2004.
- LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*, v. 25, p. 402-408, 2001.
- LIUTKEVIČIŪTĖ, K. et al. Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chemical Biology*, v. 5, p. 400-402, 2009.
- LUCIANO, D. J. et al. RNA editing of a miRNA precursor RNA editing of a miRNA precursor. *RNA*, v. 10, p. 1174-1177, 2004.
- LUGER, K. et al. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, v. 389, p. 251-260, 1997.
- LUGER, K.; DECHASSA, M. L.; TREMETHICK, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology*, v. 13, n. 7, p. 436-447, 2012.
- LUSSER, A et al. Identification of maize histone deacetylase HD2 as an acidic nucleolarphosphoprotein. *Science*, v. 277, n. 1992, p. 88-91, 1997.
- MAIA, J. et al. Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated Arabidopsis seeds. *New Phytologist*, v. 203, n. 1, p. 81-93, 2014.
- MEISSNER, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, v. 454, p. 766-770, 2008.
- MILITELLO, K. T. et al. A map of 5-methylcytosine residues in *Trypanosoma brucei* tRNA revealed by sodium bisulfite sequencing. *Molecular and Biochemical Parasitology*, v. 193, n. 2, p. 122-126, 2014.
- MOTORIN, Y.; HELM, M. RNA nucleotide methylation. *WIREs RNA*, v. 2, n. 5, p. 611-631, 2011.
- MOTORIN, Y.; LYKO, F.; HELM, M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Research*, v. 38, n. 5, p. 1415-1430, 2010.
- MUND, C. et al. Comparative analysis of DNA methylation patterns in transgenic *Drosophila* overexpressing mouse DNA methyltransferases. *The Biochemical Journal*, v. 378, pt. 3, p. 763-768, 2004.
- MYTYCH, J. et al. Nanodiamond-mediated impairment of nucleolar activity is accompanied by oxidative stress and DNMT2 upregulation in human cervical

- carcinoma cells. *Chemico-Biological Interactions*, v. 220, p. 51-63, 2014.
- NARSAI, R. et al. Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *The Plant Cell*, v. 19, n. 11, p. 3418–3436, 2007.
- NODINE, M. D.; BARTEL, D. P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes & Development*, v. 24, n. 23, p. 2678-2692, 2010.
- OKANO, M.; XIE, S.; LI, E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Research*, v. 26, n. 11, p. 2536-2540, 1998.
- OSTERMEIER, G. C. et al. Delivering spermatozoan RNA to the oocyte. *Nature*, v. 429, n. 6988, p. 154, 2004.
- PANDEY, R. et al. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Research*, v. 30, n. 23, p. 5036-5055, 2002.
- PETRUK, S. et al. TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell*, v. 150, p. 922-933, 2012.
- PICHLMAIR, A.etal. RIG-I-mediated antiviral responses to single-stranded RNA bearing 59-phosphates. *Science*,v. 314, p. 997-1001,2006.
- PRICE, J. et al. Mechanisms of glucose signaling during germination of *Arabidopsis*. *Plant Physiology*, v. 132, p. 1424-1438, 2003.
- QIAN, Y. et al. Genome-wide identification and expression profiling of DNA methyltransferase gene family in maize. *Plant Cell Report*, v. 33, p. 1661-1672, 2014.
- RADDATZ, G. etal. Dnmt2-dependent methylomeslackdefined DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, n. 21, p. 8627-8631, 2013.
- RAI, K. etal. Dnmt2 functionsinthecytoplasm to promoteliver, brain, and retina developmentinzebrafish. *Genes and Development*, v. 21, n. 3, p. 261-266, 2007.
- RAMSAHOYE, B. H. et al. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, v. 97, n. 10, p. 5237-5242, 2000.
- RAZIN A.; RIGGS, A. D. DNA methylation and gene function. *Science*, v. 210, n. 4470, p. 604-610, 1980.

- REISEN, D.; MARTY, F.; LEBORGNE-CASTEL, N. New insights into the tonoplast architecture of plant vacuoles and vacuolar dynamics during osmotic stress. *BMC Plant Biology*, v. 5, p. 13, 2005.
- ROLLAND, F.; BAENA-GONZALEZ, E.; SHEEN, J. Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annual Review of Plant Biology*, v. 57, p. 675-709, 2006.
- RONEMUS, M. J. et al. Demethylation-Induced Developmental Pleiotropy in Arabidopsis. *Science*, v. 273, p. 654-657, 1996.
- SAITO, K. et al. Pimet, the Drosophila homolog of HEN1, mediates 2' - O -methylation of Piwi- interacting RNAs at their 3' ends. *Gene & Development*, v. 21, p. 1603-1608, 2007.
- SCHAEFER, M. et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes and Development*, v. 24, n. 15, p. 1590-1595, 2010.
- SCHAEFER, M.; LYKO, F. Solving the Dnmt2 enigma. *Chromosoma*, v. 119, n. 1, p. 35-40, 2010.
- SCHAEFER, M.; STERINGER, J. P.; LYKO, F. The Drosophila cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS ONE*, v. 3, n. 1, p. e1414, 2008.
- SCHAEFER, M. et al. Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines. *Cancer Research*, v. 69, n. 20, p. 8127-8132, 2009.
- SCHMID, M. et al. A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, v. 37, n. 5, p. 501-506, 2005.
- SHEN, J. C.; RIDEOUT, W. M.; JONES, P. A. High frequency mutagenesis by a DNA methyltransferase. *Cell*, v. 71, n. 7, p. 1073-1080, 1992.
- SHAPIRO, R.; COHEN, B. I.; SERVIS, R. E. Specific deamination of RNA by sodium bisulphite. *Nature*, v. 227, p. 1047-1048, 1970.
- SHANMUGAM, R. et al. The Dnmt2 RNA methyltransferase homolog of Geobacter sulfurreducens specifically methylates tRNA-Glu. *Nucleic Acids Research*, v. 42, n. 10, p. 6487-6496, 2014.
- SMITH, C. L.; PETERSON, C. L. ATP-Dependent Chromatin Remodeling. *Current Topics in Developmental Biology*, v. 65, p. 115-148, 2005.
- SPRINZL, M.; VASSILENKO, K. S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, v. 33, p. D139-D140, 2005.
- SONG, Y. et al. Arabidopsis DNA methyltransferase AtDNMT2 associates with histone deacetylase AtHD2s activity. *Biochemical and Biophysical Research Communications*, v. 396, p. 187-192, 2010.

- STROUD, H. et al. DNA methyltransferases are required to induce heterochromatic replication in Arabidopsis. *PLoS Genetics*, v. 8, n. 7, p. e1002808, 2012.
- SRIDHA, S.; WU, K. Identification of AtHD2C as a novel regulator of abscisic acid responses in Arabidopsis. *The Plant Journal*, v. 46, p. 124-133, 2006.
- TANG, L. Y. et al. The Eukaryotic DNMT2 Genes Encode a New Class of Cytosine-5 DNA Methyltransferases. *Journal Biology Chemistry*, v. 278, n. 36, p. 33613-33616, 2003.
- THIAGARAJAN, D.; DEV, R. R.; KHOSLA, S. The DNA methyltransferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics*, v. 6, n. 1, p. 103-113, 2011.
- THOMPSON, D. M. et al. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA*, v. 14, n. 10, p. 2095-2103, 2008.
- TOVY, A. et al. A New Nuclear Function of the *Entamoebahistolytica* Glycolytic Enzyme Enolase: The Metabolic Regulation of Cytosine-5 Methyltransferase 2 (Dnmt2) Activity. *PLoS Pathogens*, v. 6, n. 2, p. e1000775, 2010.
- TSCHERNE, J. S. et al. Purification, cloning, and characterization of the 16S RNA m2G1207 methyltransferase from Escherichia coli. *Journal of Biological Chemistry*, v. 274, n. 2, p. 924-929, 1999.
- TUORTO, F. et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nature Structural & Molecular Biology*, v. 9, p. 900-905, 2012.
- UENO, Y. et al. Histone deacetylases and ASYMMETRIC LEAVES2 are involved in the establishment of polarity in leaves of Arabidopsis. *The Plant Cell*, v. 19, n. 2, p. 445-457, 2007.
- VANYUSHIN, B. F.; ASHAPKIN, V. V. *DNA methylation in plants*. New York: Nova Science Publishers, 2009.
- WADDINGTON, C. H. *Introduction to modern genetics*. London: Allen and Unwin, 1939.
- WASSENEGGER, M. et al. RNA-directed de novo methylation of genomic sequences in plants. *Cell*, v. 76, n. 3, p. 567-576, 1994.
- WEIGEL, D.; GLAZEBROOK, J. *Arabidopsis: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2002.
- WINTER, D. et al. An “electronic fluorescent pictograph” Browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, v. 2, n. 8, p. 1–12, 2007.

- WOLFFE, A. P.; HAYES, J. J. Chromatin disruption and modification. *Nucleic Acids Research*, v. 27, n. 3, p. 711-720, 1999.
- WU, C. T.; MORRIS, J. R. Genes, genetics, and epigenetics: a correspondence. *Science*, n. 293, p. 1103-1105, 2001.
- WOESE, C. R. et al. Secondary structure model bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Research*, v. 8, p. 2275-2293, 1980.
- WU, K. et al. Functional analysis of HD2 histone deacetylase homologues in *Arabidopsis thaliana*. *The Plant Journal*, v. 22, n. 1, p. 19-27, 2000.
- YANG, X. J. Lysine acetylation and the bromodomain: A new partnership for signaling. *BioEssays*, v. 26, p. 1076-1087, 2004.
- YANG, Z. et al. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Research*, v. 34, n. 2, p. 667-675, 2006.
- YODER, J. A.; BESTOR, T. H. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Human Molecular Genetics*, v. 7, n. 2, p. 279-284, 1998.
- YU, B. et al. Methylation as a crucial step in plant microRNA biogenesis. *Science*, v. 307, p. 932-935, 2005.
- ZANTEN, V. et al. Seed Maturation in *Arabidopsis thaliana* Is Characterized by Nuclear Size Reduction and Increased Chromatin Condensation. *Proceedings of the National Academy of Sciences of the United States of America*, v. 108, n. 50, p. 20219-20224, 2011.
- ZHANG, H.; HE, X.; ZHU, J.-K. RNA-directed DNA methylation in plants: Where to start? *RNA Biology*, v. 10, p. 1593-1596, 2013.
- ZHANG, X. et al. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*. *Cell*, v. 126, p. 1189-1201, 2006.
- ZHONG, X. et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell*, v. 157, n. 5, p. 1050-1060, 2014.
- ZHOU, C. et al. Expression and function of HD2-type histone deacetylases in *Arabidopsis* development. *The Plant Journal*, v. 38, n. 5, p. 715-724, 2004.

## *Capítulo 2*

### *Fragmentos derivados de tRNA: identificação e caracterização em plantas*

## Genome-wide identification and characterization of tRNA-derived fragments (tRFs) in plants

Cristiane S Alves<sup>1</sup>, Renato Vicentini<sup>2</sup>, Gustavo Turqueto Duarte<sup>2</sup>, Fabio TS Nogueira<sup>1,3</sup>

<sup>1</sup>Departamento de Genetica, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Distrito de Rubião Jr., s/n. CEP 18618-970 Botucatu, SP, Brazil.

<sup>2</sup>Centro de Biologia Molecular e Engenharia Genetica (CBMEG), Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brazil.

<sup>3</sup>Laboratory of Molecular Genetics of Plant Development, Departamento de Ciencias Biologicas, ESALQ/USP, Av. Pdua Dias s/n. CEP 13418-900 Piracicaba, SP. Brazil.

[cris\\_desantis@icloud.com](mailto:cris_desantis@icloud.com)

[shinapes@unicamp.br](mailto:shinapes@unicamp.br)

[gtduarte@unicamp.br](mailto:gtduarte@unicamp.br)

[ftsnogue@usp.br](mailto:ftsnogue@usp.br)

### Abstract

Small RNAs (sRNAs) derived from tRNAs, named tRFs (3'-loop, 5'-loop, anti-codon loop) have been reported in several organisms, including human cells and plants and they can interfere with gene expression and their biogenesis and biological functions remains unknown. Here, we capitalized on small RNA sequencing data from distinct species such as *Arabidopsis*, rice and *Physcomitrella patens* to examine the diversity of tRFs in plants. We analyzed 19 to 20-nt tRFs derived from 5' (5'tRF) and 3' (3'CCA tRF) tRNA loops. *In silico* analyzes of tRFs from these three evolutionary distant species showed that their abundance did not correlated with the number of genomic copies of parental tRNAs and that 19-20 nt 5'tRF is more abundant in all three species. *In silico* and *in vivo* analyzes unraveled differential accumulation of tRFs in distinct *Arabidopsis* tissues, suggesting that they are not subproducts of tRNA degradation. We also verified that the biogenesis of most *Arabidopsis* 19-20 nt 5'tRFs and 3'CCA tRFs is Dicer-like independent, though they seem to be associated with Argonaut proteins. Moreover, both types of tRFs accumulated in response to oxidative stress.

Keywords: small RNA, tRNA, tRF, *Arabidopsis thaliana*, high throughput data.



## Introduction

Transfer RNAs (tRNAs) are an ancient form of non-coding RNA (ncRNA) that provides a connection between transcription and translation in most organisms. Each eukaryotic genome has at least 32 distinct families of tRNAs to decode the 20 amino acids in a proper way. tRNA biogenesis is a complex process that involves several enzymes, distinct cell localization for each step and usually differs among yeast, plants and vertebrates. Apparently, the biogenesis pathway “chosen” in a particular species relies on the evolution of the processing enzymes found in each species (for review, see Hopper and Shaheen, 2008). In general, tRNA genes are transcribed by RNA polymerase III (RNA Pol III) and their transcription is terminated by a series of Us (Hopper et al., 2010; Phizicky and Hopper, 2010). The 5' leader sequence is removed from precursor tRNAs (pre-tRNAs) by the RNase P (Walker and Engelke, 2006). RNase Z then removes the 3' U trailer sequence and CCA nucleotides are added to the trimmed 3' end by tRNA nucleotidyl transferase. Finally, pre-tRNA undergoes further modification steps to mature as functional aminoacylated tRNAs. At least in *S cerevisiae*, and probably in plants, the primary termini processing of tRNAs occurs in the nucleolus. Afterwards, various modification steps occur in the nucleoplasm and at the inner nucleoplasm membrane, from where the tRNA is exported to the cytoplasm to mature. After maturation, it can be imported to the nucleus where aminoacylation takes place. In this step of the biogenesis, tRNAs can either stay in the nucleus or be exported to the cytoplasm depending on cellular needs. (Czerwoniec et al., 2009; Hopper et al., 2008; Hopper et al., 2010). Interestingly, tRNAs can assume additional biological and cellular tasks such as an unexpected role in inhibiting apoptosis by binding cytochrome *c* and consequently blocking the caspase activation in *HeLa* cells (Mei et al., 2010). Moreover, they can be used as primers for viral reverse transcription; uncharged tRNAs can act as stress sensors and as regulators of gene expression (for review see Phizicky and Hopper 2010). Moreover, recently discovered new tRNAs may perform non-canonical, extra-translational functions that may be linked to human health and disease (Parisien et al., 2013).

Recent advances in high-throughput sequencing techniques made possible the identification of small RNAs (sRNAs) derived from various cellular non-coding RNAs, including tRNAs. Most known and well-studied sRNAs are typically 19–35 nt in length and are classified into microRNAs (miRNAs), endogenous small interfering

RNAs (endo-siRNAs) and Piwi-interacting small RNAs (piRNAs) (Farazi et al., 2008). Small RNAs from tRNAs, collectively termed tRNA-derived fragments (tRFs), have been identified in different species and cell types (Li et al., 2012; Haussecker et al., 2010; Hsieh et al., 2009; Lee et al., 2009; Cole et al., 2009; Cai et al., 2013; Sobala et al., 2013). tRFs are in the size range of 13–30 nts and display distinct expression patterns, abundance, cellular localizations, or biological roles compared with their parental tRNA molecules. Cleavage does not occur randomly and their proposed nomenclature and classification are based on the regions of parental tRNA cleavage, including 3' U tRFs that are processed from precursor tRNAs (pre-tRNAs) by RNase Z and unknown tRNases and consist of the sequence between the cleavage site and the RNA PolIII run-off poly(U) tract (Gebetsberger and Polacek, 2013). Mature tRNA can generate two main types of tRFs: one processed from the 5' end (5' tRFs) and another from the 3' end, harboring the added CCA sequence (3' CCA tRFs) (Gebetsberger and Polacek, 2013). In addition to tRFs, tRNA halves (tRNA-derived, stress-induced small RNAs or short tiRNAs), deriving from both the 5' and 3' part of full-length tRNA molecule, are produced by cleavage in the anticodon loop (Emara et al., 2010).

tRFs were firstly discovered in *HeLa* cells (Lee et al., 2009) and since then, several works have been undertaken to comprehend how these small non-coding RNAs integrate the small RNA world (Cole et al., 2009; Haussecker et al., 2010; Li et al., 2012; Sobala and Hutvagner, 2013). In a few examples, 5'- and 3' CCA tRFs are generated by the action of Dicer in a similar manner to siRNAs and miRNAs, suggesting they may have similar biogenesis and regulatory functions. Nevertheless, most tRFs are essentially restricted to the cytoplasm and some has been shown to be produced by the action of other endonucleases, such as angiogenin (Yamasaki et al., 2009). Therefore, it is possible that a combination of ribonucleases is necessary in generating different species of these tRNA-derived sRNAs (Sobala and Hutvagner, 2013).

While cellular functions have been assigned for tRNA halves (Yamasaki et al., 2009; Emara et al., 2010), functions of the tRNA-derived fragments proposed so far do not yet allow drawing any clear picture of their ncRNA biology. Cellular and molecular functions of tRFs may be mediated by their association with AGO proteins (Cole et al., 2009; Haussecker et al., 2010; Li et al., 2012) and, at least in some

instances, may involve the inhibition of protein translation and down regulation of target genes by transcript cleavage, resembling miRNA/siRNA biological functions (Li et al., 2012; Sobala and Hutvagner, 2013). Another interesting hypothesis is that tRFs may compete for AGO association with other cellular sRNAs and thus they might act as global regulators of RNA-mediated silencing (Haussecker et al., 2010). These findings suggest that tRNA-derived RNA molecules are, in some cell types and biological contexts, used for broad control of gene expression.

As discussed above, several works evaluating the expression, biogenesis and possible biological function(s) of tRFs have been done mostly in animal cells. Conversely, few works have been done in plants to show the presence of tRFs. In *Arabidopsis thaliana*, the 5' tRF of Asp<sup>GTC</sup> and the 5' and 3' CCA tRFs of Gly<sup>TCC</sup> tRNAs were found to be highly expressed in root tissues treated with phosphate deprivation (Hsieh et al., 2009). In barley, the His<sup>GTG</sup> tRF was the most abundant sRNA molecule identified in small RNA populations under phosphorous-deficient conditions (Hackenberg et al., 2012), while in rice the 5' Ala<sup>AGC</sup> and Pro<sup>CGG</sup> tRFs demonstrated differential expression in the callus and leaves (Chen et al., 2011). Additionally, some 5' tRFs were shown to be associated with *Arabidopsis* AGO proteins (Loss-Morais et al., 2013). In this work, we capitalized on the publicly small RNA sequencing data from several species such as *Arabidopsis*, rice, and *Physcomitrella patens* to comprehensively examine the diversity, expression and possible biogenesis of tRFs in plants. We focused on 5'- and 3' CCA tRFs and showed that both are conserved across plant kingdom and their accumulation is not correlated with parental tRNA loci number. Most identified tRFs are not DCL-dependent and some indeed accumulated at higher levels in *dcl* mutant tissues. Moreover, 5' tRFs and 3' CCA tRFs accumulated in response to oxidative stress.

## **Results and Discussion**

### **Plant tRFs do not correlated with their parental tRNAs in terms of copies and abundance and are differentially expressed in tissues/organs**

We decided to use publicly available databases from small RNA libraries from distinct evolutionary plant species (Supplemental Table 1). We select reads from 19 to 20 nts, previously reported to be the most abundant tRFs (Lee et al., 2009; Cai et al,

2013) and mapped them to their respective parental tRNAs. We only analyzed reads that fully mapped with mature tRNAs from *A. thaliana* (for Arabidopsis and rice tRFs) and *P. patens*. Firstly, we asked whether there was a correlation between parental tRNA *loci* number and tRF abundance by estimating the total number of tRNA *loci* from each anticodon and the number of tRFs that mapped to each particular tRNA. At this point, we did not distinguish 19-20 nt 5' or 3' CCA tRFs. Apparently, there was low or no correlation between parental tRNA *loci* number and tRF abundance in *A. thaliana*, *O. sativa* or *P. patens* (Fig. 1A). In several animal species, tRFs did not derived from abundant cellular tRNAs (Kumar et al., 2014). In agreement with this data, we did not observed correlation between tRF abundance and tRNA copy number (Fig. 1A). For instance, *Arabidopsis thaliana* tRNA Arg<sup>CCG</sup> has four *loci* that mapped to 260 tRF reads, while tRNA Pro<sup>TGG</sup>, which has 45 copies in the genome, mapped only to 12 reads (data not shown). Our data is in agreement with other species (Kumar et al., 2014), except for *Tetrahymena* (Couvillion et al., 2010), and suggest that plant tRFs are not subproducts of random cleavage. (Li et al., 2012; Lee et al., 2009; Cole et al., 2009; Hurto et al, 2011). In fact, each type of tRF seems to be generated from a specific parental tRNA and different kinds of enzymes could be acting on their cleavage (Hurto, 2011). Moreover, tRF cleavage site is highly conserved among evolutionary distant plant species (Supplemental Fig. 1).

Next, we aligned tRFs from different sRNA libraries with parental tRNAs in order to verify if tRFs have a bias regarding cleavage position. Similarly to other species (Lee et al., 2009; Li et al., 2012; Cai et al, 2013), we observed a peak of tRFs into 5' tRNA termini in all three species (Fig 1B). The precision with which individual tRFs are generated suggests that tRFs are not generated by random exonucleolytic digestion of tRNA precursors. In addition, because of the method of sRNA sequencing in these databases requires reverse transcriptase to read through the tRF into the adaptor sequence, tRNA modifications would, if anything, lower the number of reads that we observed, not create artificial short sequences.

To evaluate whether tRF count varies among tissues/organs, we compared total 5' and 3'CCA tRF accumulation in Arabidopsis tissues/organs by estimating TPM (transcripts per million) from publicly available databases (Supplemental Table 1). We observed differential accumulation among tissues/organs, in which the highest TPM was observed in seedlings and the lowest in rosette leaves. Independently of the

organ/tissues evaluated, 5' tRFs were the most abundant sRNAs observed in the libraries (Fig. 2A). This data agrees with recent work from Kumar and co-authors (2014), in which the authors did a meta-analysis of several tRFs and showed that a number of animal tRFs accumulated differentially in tissues/organs from distinct species, including humans.

To evaluate the abundance of individual tRFs, we employed stem-loop qRT-PCR using total RNA from different *Arabidopsis* organs. Stem-loop qRT-PCR is a standard technique to accurately quantify miRNAs (Varkonyi-Gasic et al., 2007; Ortiz-Morea et al, 2013). We selected two *Arabidopsis* 5' tRFs and two 3' CCA tRFs (Fig. 2B). tRF transcripts were readily detectable, indicating that they were authentic sRNAs. We observed differential accumulation mainly in siliques as compared with 15-d-old seedling tissues (Fig. 2C). The differential accumulation of both 5' tRFs and 3' CCA tRFs in *Arabidopsis* organs reinforced the idea that these small RNAs are not subproducts of random tRNA degradation in plants, similar what have been observed in animal cells (Kumar et al, 2014).

### **Plant tRFs accumulate differentially under non-physiological conditions**

In order to verify how tRFs behave under non-physiological conditions, we decided to evaluate their accumulation patterns in *Arabidopsis* plants growing under oxidative stress conditions, as well as in the presence or absence of glucose and the phytohormone abscisic acid (ABA). We employed stem-loop qRT-PCR to detect transcripts of four tRFs (Fig. 2B). Under oxidative stress conditions (presence of hydrogen peroxide), we observed a general tendency for all tRFs to accumulate at higher levels than control conditions (Fig.3A). Under some stress conditions, aminoacylation levels of tRNAs change and uncharged tRNAs could participate of different biological processes to regulate global gene expression, helping the cell to survive under stressing conditions. tRNAs could act as stress signals, regulating cell death or being a substrate for tRF formation (Mei et al., 2010; Phizicky and Hopper, 2010). Previous reports showed that eukaryotic tRNA is cleaved to tRNA-halves under some stress conditions especially oxidative stress and selected tRNA-halves are able to inhibit protein synthesis (Thompson et al., 2008; Ivanov et al., 2011; Saikia et al., 2012). Likewise, archeal model *Haloferax volcanii* generates different tRFs under stress conditions and one of them is capable of inhibiting protein synthesis

(Gebetsberger et al., 2012). Interesting, Arabidopsis root tissues accumulate high levels of 19- nt 5'tRFs under Pi starvation (Hsieng et al., 2009). Our data showed that four analyzed tRFs accumulated under oxidative stress condition (Fig. 3A). It is possible they accumulate under different stress conditions as well, as it has been shown for some Arabidopsis tRFs through *in silico* analyzes (Loss-Morais et al., 2013).

Besides energy supply and carbon storage, glucose is a key regulator for diverse processes in plants such as germination and development and also is a signaling for internal modulation in response to environmental changes (Moore et al., 2003; Tsutsui et al., 2015) and the phytohormone abscisic acid (ABA) is also a signaling molecule that participate in different regulatory pathways (for review see: Rook et al., 2006; Mittler et al., 2015). In the presence of glucose, only transcripts of 5'tRF-Arg<sup>TCG</sup> seemed to accumulate at higher levels (Fig. 3B). Conversely, all four tRFs accumulated similarly in the presence or absence of ABA (Fig. 3C).

### **Biogenesis of most plant tRFs seems to be Dicer-like independent**

Next, we asked whether similar biogenesis pathways as mi/siRNAs could generate plant tRFs. Thus, we analyzed arabidopsis libraries from mutants and double mutants associated with small RNA machinery and RNA processing and compared them with wild type (WT) libraries (see Material and Methods). We initially estimated tRF counts from *hyl1-2*, *hyl1-2;dcl1-20* and *hyl1-2;dcl1-21* rosette leaf libraries and compared with WT library from the same experiment. We observed that three 19-20 nt 3' CCA tRF and one 5' tRF differentially accumulated in mutant libraries as comparing with WT library. However, most 5' and 3'-CCA tRF counts did not significantly differ in mutants as compared with WT (Table 1). In general, the accumulation of only few tRFs seemed to change in sRNA biogenesis pathway-associated mutants. This observation supports the idea that the biogenesis of most plant tRFs is independent of these factors.

There are controversies regarding what enzymes are important for tRF biogenesis in different organisms. For example, RNase Z is responsible for 3'pre-tRNA cleavage tRNA-halves could be processed by Angiogenin in mammals, Rny1 in yeast and PrrC, colicin D and colicin E5 in bacteria (Hackenberg et al., 2012). For some 5' tRFs, Dicer was shown to be responsible for its processing in *HeLa* cells.

Dicer is able to process 3'CCA tRFs in *HeLa* as well, and also Angiogenin and others RNases A in a Dicer-independent pathway (For review, see Raina and Ibba, 2014). To address individual processing of 5' tRF and 3'CCA tRF in arabidopsis, we decided to evaluate the accumulation patterns of four tRFs (Fig. 2B) by comparing their expression patterns in tissues of 15-d-old seedlings of *dcl1*, -2, -3, -4 and *hyl1* mutants with WT tissues (Fig. 4). Most tRFs accumulated similarly in WT and mutant tissues, excepted for 3' CCA tRF-Ser<sup>GGA</sup>. The accumulation of this tRF was significantly lower in *hyl1* when compared with WT (Fig. 4). Interestingly, we observed higher accumulation of transcripts, specially from 5' tRFs and 3'CCA tRF-Ser<sup>GGA</sup> in *dcl2*, -3, -4 mutants (Supplemental Fig. 2). DCL1 and HYL1 are enzymes associated with microRNA biogenesis, while DCL2 and DCL3 are associated with siRNA and DCL4 with trans-acting siRNA (ta-siRNA) biogenesis (Kim et al., 2008). Given that the processing of several sRNA classes is compromised in *dcl2*, -3, -4 mutants, it is possible that the high accumulation of tRFs in these mutants (Fig. 4 and Supplemental Fig. 2) is due to the lack of competition with other sRNAs. In other words, as *dcl2*, -3, -4 mutants have much less production of siRNAs/ta-siRNAs, it is likely easier to detect and amplify tRFs in these mutants than in WT tissues. However, we cannot roll out the possibility that Dicer-likes are important enzymes for the processing of tRFs in particular developmental stages, organs and/or environmental conditions.

### **AGOs can incorporate tRFs but do not have targets**

To evaluate if tRFs are associated with Argonautes (AGOs), we took advantage of publicly available high throughput sequencing data from different Arabidopsis AGO immunoprecipitation experiments (Supplemental Table 1). Firstly, we looked for a bias in the 5' nucleotide of tRFs that incorporate into AGO1, AGO2, AGO4 and AGO5. As a result of our analyzes, guanine (G) is the most abundant nucleotide in tRFs incorporate in the Argonautes (Fig. 5A). This is agreement with our previous analyses that showed a conserved bias for G in the 5'end of 19-20 nt 5'tRF and 20 nt 3' CCA tRF from Arabidopsis, rice, and moss (Supplemental Fig. 1). After, we identified 19 to 25 nt tRFs that were associated with AtAGOs 1, 2, 4 and 5. For AtAGO1, the most abundant tRF size was 20 nts. Conversely, only 19 nt tRFs were found to be associated with AtAGO2. For AGO4, 38% of 24nt and 31% of 19 nt long. Finally for AGO 5, 83% of 19nt and 17% of 24nt long (Fig. 5B, C, D and E). Taken together, these data

reveals a bias for the 19 nt long tRFs by the Argonautes. We also find that a few tRFs incorporate in Argonaute7 (Fold Change  $\geq 1$ ) (Fig. 5F), and AGO7 incorporated tRFs also have a bias for 19nt long, against 14% for 20nt and 29% for 21nt (Fig. 5G).

A few biological functions has been linked to tRFs, for instance HIV (human immunodeficiency virus) uses the 3'end of human tRNA Lys as a primer for its retrotranscriptase and DNA synthesis (Yeung et al., 2009). The 3'tRF was shown to associate with Dicer and Argonaute2 (AGO2), causing the RNA cleavage, demonstrating a role for tRFs in viral gene silencing and others tRFs has been shown to associate with AGOs, potentially acting as a siRNA or miRNA (Haussecker et al., 2010).

It is known that AGO1 has preference for 5'U 21 nt microRNAs, AGO2 5'A 21 nt tasiRNAs, AGO4 5'A 24 nt siRNAs, AGO5 5'C 21, 22, 24 nt siRNAs and AGO7 for 21 nt and no selective 5'nucleotide (reviewed by Kim, 2008; Bologna and Voinnet, 2014) According to available literature, there is no evidence that plant Argonautes proteins would incorporate tRFs with a biological purpose. The nucleotide bias that we found on the analyzed data did not agree with AGOs nucleotide preferences. Moreover, we did not find any evidence or potential target for the tRFs found in our study via psRNATarget (data not show) that justifies the belief that plants could have similar micro/siRNA machinery or biogenesis pathway, in disagreement with previously reported putative targets found in arabidopsis (Loss-Morais et al., 2013).

Other reports showed that 5'tRFs can inhibit protein synthesis, like archeon *Haloferax volcanii* (Gebetsberger et al., 2012), it has been shown that 26 nt long 19-20 nt 5'tRF Val can inhibit peptide bound formation and 19-20 nt 5'tRF Ile is not. Other study found that four different 19 nt long 5'tRFs from human cells also can inhibit protein synthesis *in vivo* and *in vitro* without a complementary sequence on the mRNA, but a GG motive at 3'end is required. These data shed light on the possibility that tRFs, indeed, have a distinct mechanism of action from microRNA and siRNA, and probably is an ancient mechanism to control gene expression (Sobala and Hutvagner, 2013). Our findings support a DCL/AGO independent pathway. It is more plausible to search new pathways and mechanisms of action for tRFs, makes more sense that these newly find molecules have and old and maybe simpler mechanisms, after all tRFs are formed by tRNA, the essential molecule for protein synthesis, one of the oldest molecules that we know of and we are just beginning to understand its



complexity, biogenesis and non-canonical functions. Within the next generation sequencing, all these new small and long RNAs are emerging together with new functions and mechanisms and we are just entering in an exciting and brand new RNA world.

## Conclusions

tRNA is one of the oldest RNA molecules and has been shown to play others biological roles rather than the canonical one into protein synthesis pathway. Both charged and uncharged tRNAs can play different roles like being substrate for non-ribosomal peptide bond formation on the cell wall formation and protein labeling for degradation, in the first case and, regulate global gene expression in response to changes in the amino acid pool in the cell, furthermore, tRNAs are one of the most stable RNAs (Gebetsberger and Polacek, 2013; Raina and Ibba 2014). Small non-coding RNAs plays important role on gene regulation and they are being extensively studied in the past few years. Among them, tRFs have emerged as a potential translational regulator as well as gene expression regulator. tRFs are present in all three domains of life and their abundance did not seem to be correlated to tRNA *loci* number, at least, in human cells (Kumar et al., 2014). tRNA is the bridge between RNA world and protein word, gene duplication and recombination are driving forces on gene evolution, changing gene structure and function (Kanai et al., 2015). Recently studies demonstrated that plants have a slight higher mutation frequency than mammals, and among plant species, they also have a divergence mutation's rate (Akkuratov, 2014). tRNA can be processed to generate small tRNA fragments. Apparently, tRFs have biological function based on a few evidences, for instance the fragmentation pattern can change according to the developmental stage and some tRFs can bound to Argonaute/Piwi proteins (Raina et al., 2014; Kumar et al., 2014).

In this work, we conclude that tRFs from three distinct evolutionary plant species did not correlated with parental tRNAs and 19-20 nt 5'tRF is much more abundant than 3'tRF in all three analyzed species, also tRFs accumulate differently in distinct arabidopsis organs. Arabidopsis 5'tRFs and 3'CCA tRFs biogenesis are Dicer-like independent, even incorporating in Argonauts proteins. More work should be done in order to unravel tRFs biogenesis and mechanisms pathways in plants.

## Material and Methods

**Bioinformatics analyses:** Small RNA libraries (Sup. Table 1) were searched for identification of tRFs by performing a BLAT search (minimal identity = 100, tile size = 7, and minimal score = 0) of sRNAs with perfect match against the whole set of tRNAs of each plant species (Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: A database of transfer RNA genes detected in genomic sequence). tRFs expression were estimated normalizing the amount of reads of each sRNA by the total sRNA library sequence size, with a minimal of 10 per 1 million of reads. Statistical analysis was calculated by Fisher's exact test.

**Plant material, growth conditions and tissue collection:** In this study the *Arabidopsis thaliana* lines background used were Columbia (Col0) wild-type (WT), *hyl1-2* (Vazquez et al., 2004), *dcl1-11* (Zhang et al., 2008), *dcl2-1* and *dcl3-1* (Xie et al., 2004), *dcl4-2* (Xie et al., 2005). All the seeds were germinated on commercial solid Murashige and Skoog (MS) medium with 3% sucrose. All plants were growth under long days conditions (16 h light, 8 h dark) at 19-22°C. The seedlings were collected with 10 (root and shoot from *dcl1*, *hyl1* and WT) and 15 (shoot) days after germination (*dcl2*, *dcl3*, *dcl4*, WT) and for each biological sample a pool with 10 to 15 plants were used. For the tissues analysis, apical meristem and leaves were collected 15 days after germination; the first four siliques from different *Arabidopsis* was harvest and pooled together; a pool containing buds from different plants was formed.

**Oxidative stress treatment:** *Arabidopsis thaliana* ecotype Columbia 0 was used. Seeds were sterilized and stratified for 48 h at 4°C then transferred to a growth chamber under 60 rpm rotation. Plants were growth on erlenmeyers containing half strength modified Murashige and Skoog (MS) medium with vitamins (PhytoTechnology Laboratories® - Prod no: M519) with 1.5% sucrose. All plants were growth under constantly light at 19-22°C. After nine days, 10 mM H<sub>2</sub>O<sub>2</sub> was added to MS medium (Thompson et al., 2008). For control plants, ½ MS was used instead H<sub>2</sub>O<sub>2</sub>. After 16h, both control and treated seedlings were harvested and quickly frozen in liquid nitrogen.

**ABA and glucose treatment:** Arabidopsis (Col-0) seeds were sown in 10 mL liquid half-strength Murashige and Skoog media supplied with 0.3% glucose and were stratified for two days in the dark at 4°C, following five days of growth under continuous light at 24°C. The media were then replaced by 10 mL glucose-free MS/2 for a 24h growth, following treatments with 2% glucose, 2% mannitol (osmotic control) or 100 µM ABA (all Sigma-Aldrich reagents). Evaluations of RNA profile changes by qRT-PCR and high-throughput sequencing were performed as previously described (Duarte et al., 2013).

**RNA isolation and quantitative analysis:** The plant material was ground under liquid nitrogen and total RNA extracted with TRIzol® Reagent (Ambion) according to manufacturer instructions. The RNA was treated with Turbo DNA-free™ (Ambion). A PCR reaction was performed to guarantee no DNA contamination. In order to detect and quantify tRFs, stem-loop RT-PCR was performed according to Chen<sup>X</sup>. Oligos for tRFs class 1 Arg-TCG (chr1.trna186-ArgTCG) and Arg-CCT (chr1.trna74-ArgCCT) and for tRFs class 2 Tyr-GTA (chr3.trna24-TyrGTA) and Ser-GGA (chr2.trna11-SerGGA) was designed with no mismatches in the six nucleotides from 5' end and 3' end (including CCA) respectively, of the specific tRNA (primers Supplementary Table 2). The reactions were carried out using Go-Taq® qPCR Master Mix (Promega), following the manufacturer instructions and Applied StepOne™ Real-Time PCR System was used to perform the reactions. For all analysis, *AtActin* (AT3G18780) was used as an endogenous gene. All the reactions were run in triplicate. Gene expression analysis by qRT-PCR was conducted as described by Udvardi and co-authors (2008), evaluating gene expression with the Delta-Delta Ct method (Livak and Schmittgen, 2001). Differences in gene expression were considered significant for  $P < 0.05$  according to two-tailed Student's t-test.

### Figure legends

**Figure 1:** Small RNAs (sRNAs) matching to 5' and 3' termini of tRNAs and their respective parental tRNAs are conserved across plant species. (A) Negative correlation between the abundance of conserved tRNA-derived fragments (tRFs) retrieved from sRNA libraries and the number of loci of their respective parental tRNAs in distinct

plant species. The values were  $\log_{10}$ -transformed to facilitate the comparisons. **(B)** 5' terminal of conserved plant tRNAs are highly and specifically processed to yield tRFs (5' terminal peak). The tRNA position (with respect to its 5' end) at which the 5' end of the sRNA reads are matched are used, as depicted by the bottom illustration. At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Pp, *Physcomitrella patens*.

**Figure 2:** Abundance of tRFs. **(A)** Graph indicating the total number of reads from *Arabidopsis thaliana* from publicly available data (see Supplemental Table 1). Blue bars represent total reads with 19 and 20 nucleotides; Red bars indicate total reads mapping in tRNA 5' termini position; Green bars indicate total reads that mapped tRNA 3'CCA termini position. **(B)** Representation of mature tRNAs used in this study and its respective tRFs: 5'tRF-Arg<sup>CCT</sup>, 5'tRF Arg<sup>TCG</sup>, 3'tRF-Ser<sup>GGA</sup> and 3'tRF Tyr<sup>GTA</sup> **(C)** tRFs in Col-0 bud, meristem and silique. All expression values are in comparison to Col-0 leaves. Values are mean  $\pm$  SD of two biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with leaves. Each bar represents the mean value and standard deviation of at least two independent biological replicates. *Actin* was used as an endogenous control. Two tRF of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were used.

**Figure 3:** Abundance of tRFs in Col-0 seedlings grown in different treatments. All expression values are in comparison with no treated -standard (ST)- samples. Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed). *Actin* was used as an endogenous control. Two tRFs of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were analyzed. **(A)** Oxidative stress. **(B)** Glucose. **(C)** ABA.

**Figure 4:** Abundance of tRFs in Col-0, *dcl1-11* and *hyll-2* 10 days old seedlings. All expression values are in comparison to Col-0. Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with Col-0. Each bar represents the mean value and standard deviation of three independent biological

replicates. *Actin* was used as an endogenous control. Two tRFs of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were used.

**Figure 5:** Argonaut incorporation: **(A)** Percentage of each 5' terminal nucleotide in total tRFs in *Arabidopsis thaliana* and those enriched in AGO1, AGO2, AGO4 and AGO5 complexes. **(B)** Percentage of each tRF length in AGO1 complex in immunoprecipitation experiments. **(C)** Percentage of each tRF length in AGO2 complex in immunoprecipitation experiments. **(D)** Percentage of each tRF length in AGO4 complex in immunoprecipitation experiments. **(E)** Percentage of each tRF length in AGO5 complex in immunoprecipitation experiments. **(F)** Percentage of each tRF length incorporated in AGO7 complex. **(G)** Enrichment of each tRF in AGO7 complex compared to total input from immunoprecipitation experiments.

**Supplemental Figure 1:** Relative nucleotide bias at each position of the tRFs in total reads from *Arabidopsis thaliana* (At), *Physcomitrella patens* (Pp) and *Oryza sativa* (Os). **(A)** 19 nucleotides 5'tRFs. **(B)** 20 nucleotides 5'tRFs. **(C)** 19 nucleotides 3'tRFs. **(D)** 20 nucleotides 3'tRFs. *Arabidopsis thaliana*.

**Supplemental Figure 2:** Abundance of tRFs in *Arabidopsis thaliana* 15 days old seedlings. Expression values are in comparison to Col-0 (WT). Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with leaves. Each bar represents the mean value and standard deviation of three independent biological replicates. *Actin* was used as an endogenous control. Two tRF of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'CCA tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were used. **(A)** 5'tRF-Arg<sup>CCT</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*. **(B)** 3'CCA tRF-Ser<sup>GGA</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*. **(C)** 5'tRF- Arg<sup>TCG</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*. **(D)** 3'tRF- Tyr<sup>GTA</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*.

**Table 1:** Log 2 Fold Change in 5 weeks old roseleaves of double mutants *dcl1-20;hyl1-2*, *dcl1-21;hyl1-2* and single mutant *hyl1-2* compared with wild type (GEO accession: GSM738729-8696, GSM738730-8704, GSM738728-8685 and GSM738727-8676 respectively):

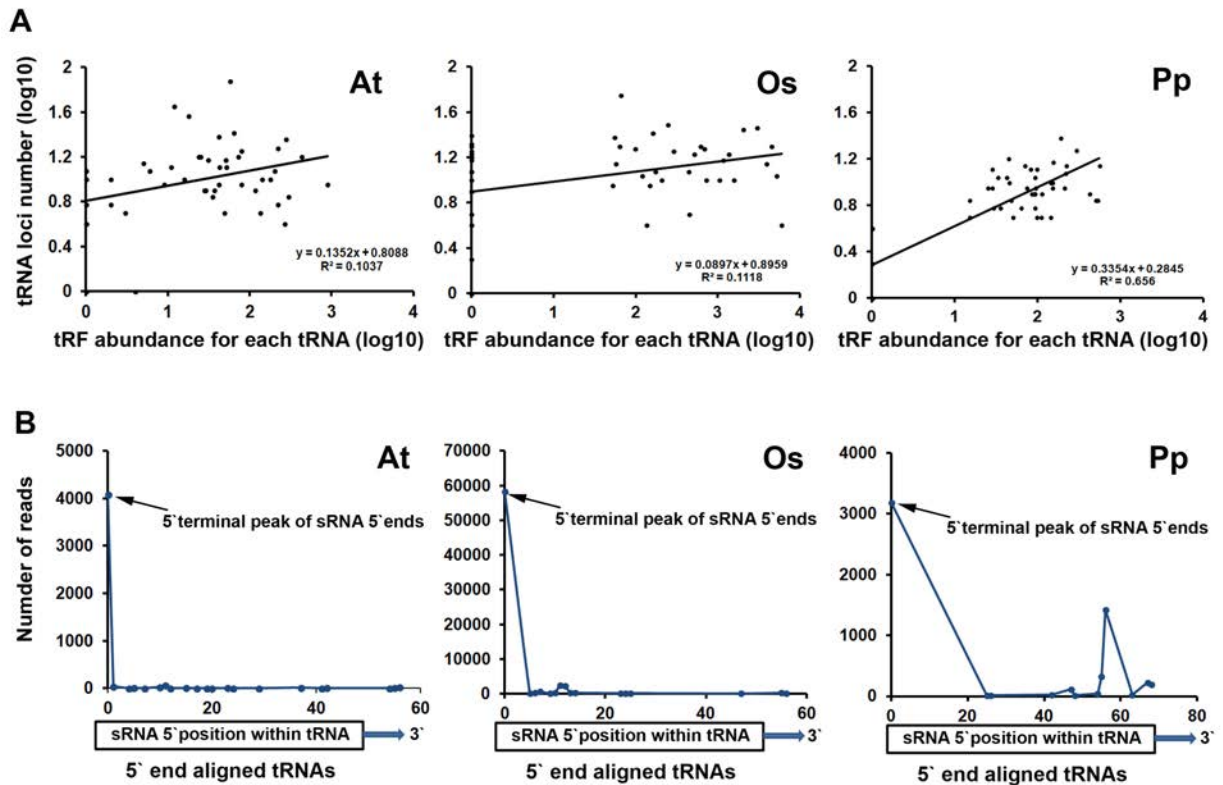
	19-20nt				tRF position
	WT	<i>hyl1-2</i>	<i>hyl1-2/dcl1-20</i>	<i>hyl1-2/dcl1-21</i>	
<b>lib size</b>	33540825	42295423	24611317	56751521	
<b>chr3.trna24-TyrGTA</b>	5990	3393 <sup>a</sup>	4566	5827 <sup>a</sup>	3'
<b>chr2.trna11-SerGGA</b>	3504	2034 <sup>a</sup>	2252 <sup>a</sup>	2331 <sup>a</sup>	3'
<b>chr1.trna186-ArgTCG</b>	7397	6653 <sup>a</sup>	1518 <sup>a</sup>	13692 <sup>a</sup>	5'
<b>chr4.trna74-LeuTAG</b>	2020	1416 <sup>a</sup>	468 <sup>a</sup>	3393	3'

**Supplemental Table 1:** GEO accessions:

ORGANISM	GEO	Description
Pp	GSM313212	WT- protonema 10 days
Pp	GSM313214	<i>dcl3-5</i> protonema 10 days
Pp	GSM313215	<i>dcl3-10</i> protonema 10 days
Pp	GSM313216	<i>rdr6-19</i> protonema 10 days
Os	GSM520637-18674	<i>dcl1r-2</i> seedlings
Os	GSM520638-18676	<i>dcl3a-17</i> seedlings
Os	GSM520640-18679	WT seedlings
Os	GSM520639-18678	<i>rdr2-2</i> seedlings
Os	GSM520636-18672	IP AGO16 seedlings
Os	GSM455965-11464	total extract
Os	GSM455964	IP AGO1c seedlings
Os	GSM455963	IP AGO1b seedlings
Os	GSM455962	IP AGO1a seedlings
Os	GSM520634-18667	IP AGO4a seedlings
At		wt_processed
At		met1_processed
At		rdd_processed
At		ddc_processed
At		abh11_processed
At		ein56abh11_processed
At		ein56_processed
At	GSM284747	WT floral bud
At	GSM284748	<i>abh1-1</i> floral bud
At	GSM284750	<i>abh1-1 ein5-6</i> floral bud
At	GSM284749	<i>ein5-6</i> floral bud
At	GSM121455	Wt inflorescence bud 5 weeks
At	GSM121454	<i>rdr6</i> inflorescence bud 5 weeks

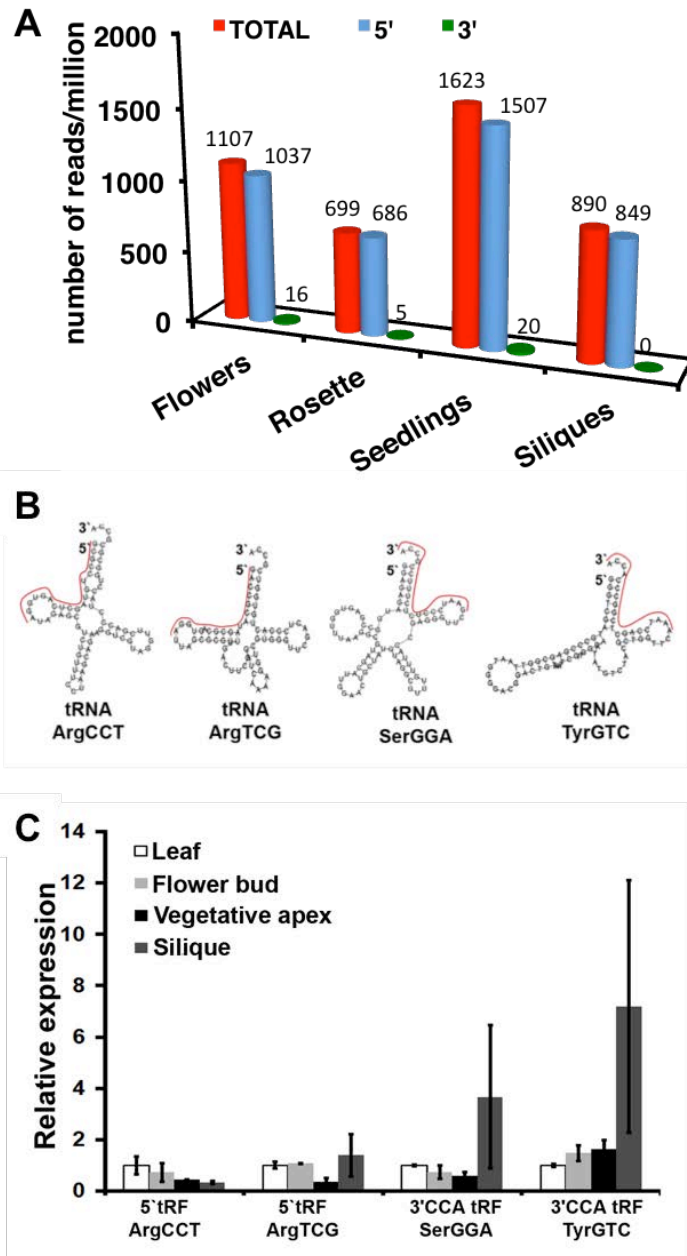
At	GSM121453	<i>rdr2</i> inflorescence bud 5 weeks
At	GSM121457	<i>dcl234</i> inflorescence bud 5 weeks
At	GSM154336	WT inflorescence stages 1-12
At	GSM154368	<i>rdr6-15</i> inflorescence stages 1-12
At	GSM154361	<i>dcl1-7</i> inflorescence stages 1-12
At	GSM154367	<i>rdr2-1</i> inflorescence stages 1-12
At	GSM154365	<i>rdr1-1</i> inflorescence stages 1-12
At	GSM154363	<i>dcl3-1</i> inflorescence stages 1-12
At	GSM154364	<i>dcl4-2</i> inflorescence stages 1-12
At	GSM154375	WT seedlings 6 days
At	GSM154376	<i>rdr6-15</i> seedlings 6 days
At	GSM149079	AGO4 complex extract
At	GSM343001-14391	WT inflorescence bud stages 1-12
At	GSM343005-14394	<i>ago1-25</i> inflorescence bud stages 1-12
At	GSM893113	WT immature inflorescence bud
At	GSM893117	<i>pol V-3</i> immature inflorescence bud
At	GSM893116	<i>pol V-2</i> immature inflorescence bud
At	GSM893118	WT immature inflorescence bud
At	GSM893124	<i>rdr2-1</i> immature inflorescence bud
At	GSM893122	<i>rdm1-4</i> immature inflorescence bud
At	GSM893121	<i>dms3-1</i> immature inflorescence bud
At	GSM893120	<i>drd1-1</i> immature inflorescence bud
At	GSM893119	<i>dms4-1</i> immature inflorescence bud
At	GSM738729-8696	<i>hyll-2 dcl1-20</i> roseleaves 5 weeks
At	GSM738727-8676	WT roseleaves 5 weeks
At	GSM738730-8704	<i>hyll-2 dcl1-21</i> roseleaves 5 weeks
At	GSM738728-8685	<i>hyll-2</i> roseleaves 5 weeks
At	GSM738731-8720	WT flowers and floral buds 8 weeks
At	GSM738734-8773	<i>hyll-2 dcl1-21</i> flowers and floral buds 8 weeks
At	GSM738733-8749	<i>hyll-2 dcl1-20</i> flowers and floral buds 8 weeks
At	GSM738732-8732	<i>hyll-2</i> flowers and floral buds 8 weeks
At	GSM118374	<i>in vitro</i> seedlings 6 days
At	GSM118375	siliques 2 months
At	GSM118373	roseleaves 6 weeks
At	GSM118372	flowers 4 weeks
At	GSM253622	IP AGO1 complex
At	GSM253623	IP AGO2 complex
At	GSM253624	IP AGO4 complex
At	GSM253625	IP AGO5 complex
At	GSM304284	inputAGO7
At	GSM304285	IP AGO7

## Figures

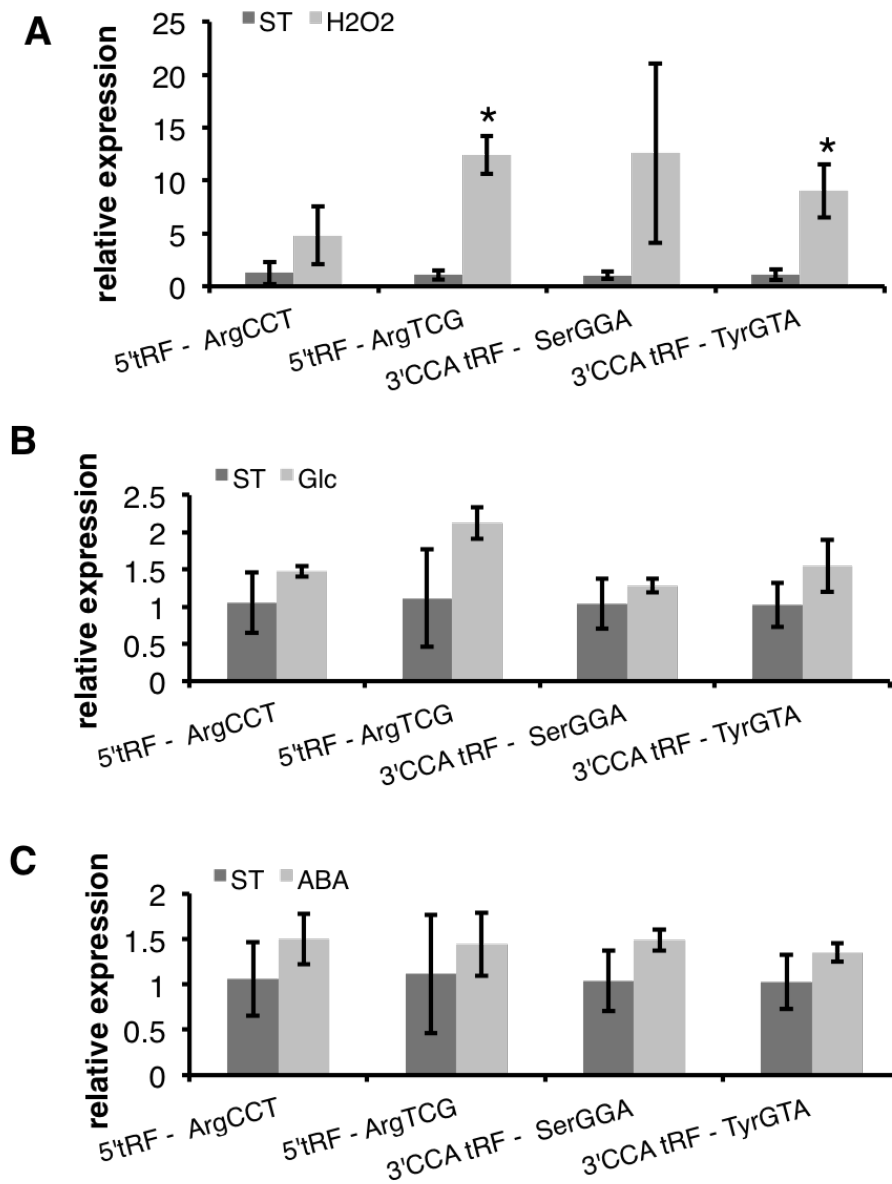


**Figure 1:** Small RNAs (sRNAs) matching to 5' and 3' termini of tRNAs and their respective parental tRNAs are conserved across plant species. **(A)** Negative correlation between the abundance of conserved tRNA-derived fragments (tRFs) retrieved from sRNA libraries and the number of loci of their respective parental tRNAs in distinct plant species. The values were log<sub>10</sub>-transformed to facilitate the comparisons. **(B)** 5' terminal of conserved plant tRNAs are highly and specifically processed to yield tRFs (5' terminal peak). The tRNA position (with respect to its 5' end) at which the 5' end of the sRNA reads are matched are used, as depicted by the bottom illustration. *At*, *Arabidopsis thaliana*; *Os*, *Oryza sativa*; *Pp*, *Physcomitrella patens*.

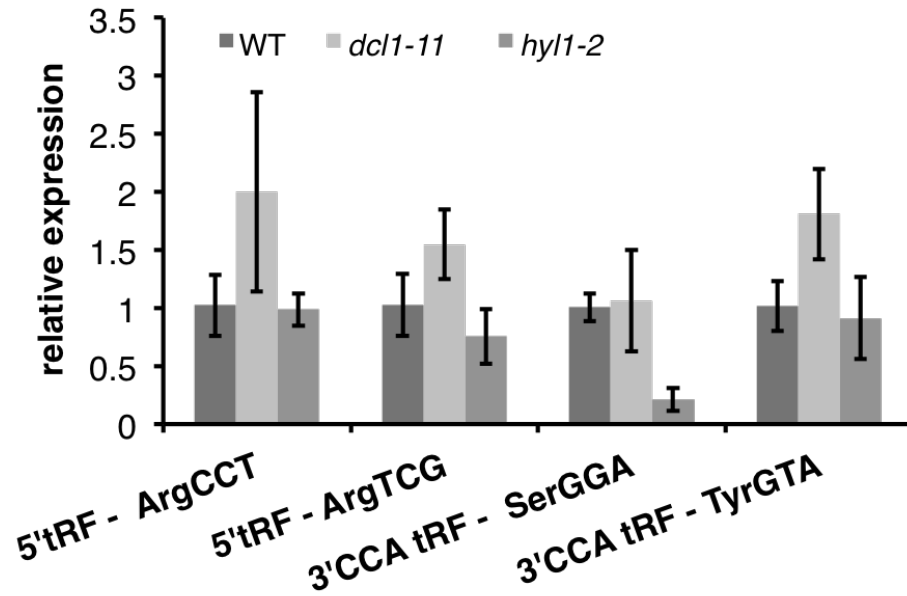




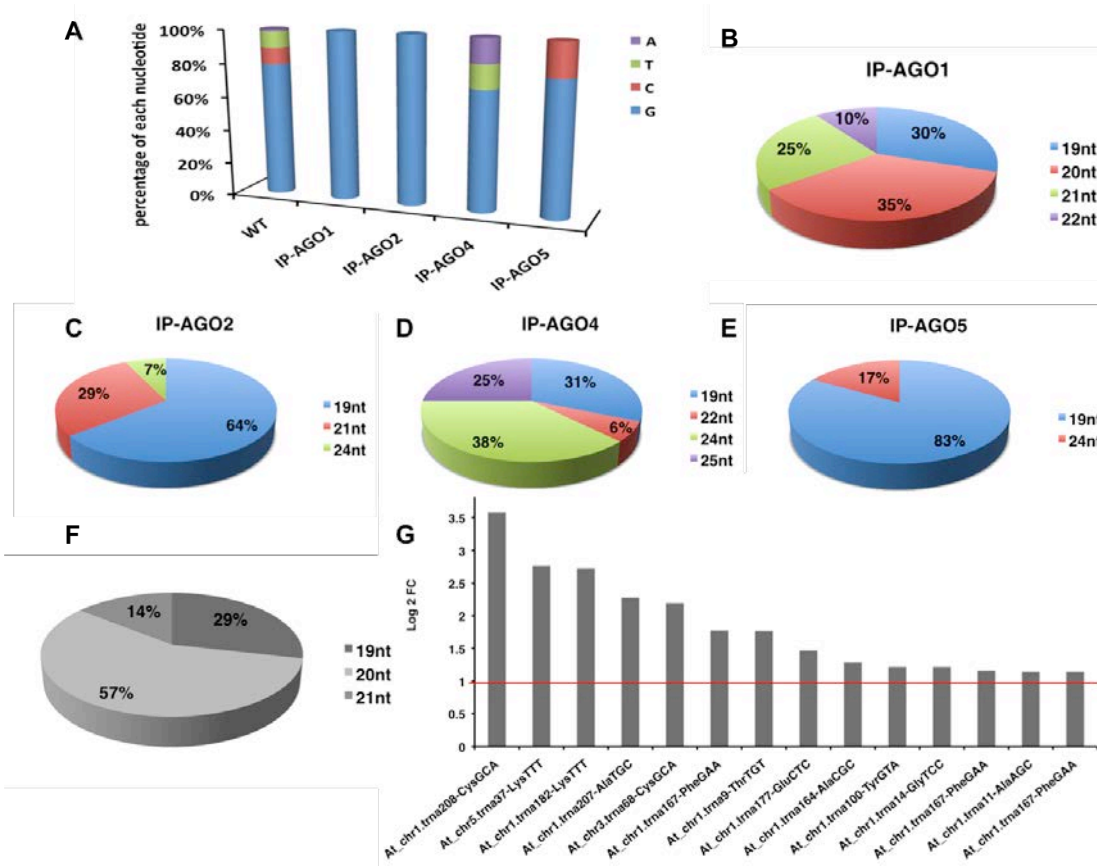
**Figure 2:** Abundance of tRFs. (A) Graph indicating the total number of reads from *Arabidopsis thaliana* from publicly available data (see Supplemental Table 1). Blue bars represent total reads with 19 and 20 nucleotides; Red bars indicate total reads mapping in tRNA 5' termini position; Green bars indicate total reads that mapped tRNA 3'CCA termini position. (B) Representation of mature tRNAs used in this study and its respective tRFs: 5'tRF-Arg<sup>CCT</sup>, 5'tRF Arg<sup>TCG</sup>, 3'tRF-Ser<sup>GGA</sup> and 3'tRF Tyr<sup>GTA</sup> (C) tRFs in Col-0 bud, meristem and silique. All expression values are in comparison to Col-0 leaves. Values are mean  $\pm$  SD of two biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with leaves. Each bar represents the mean value and standard deviation of at least two independent biological replicates. *Actin* was used as an endogenous control. Two tRF of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were used.



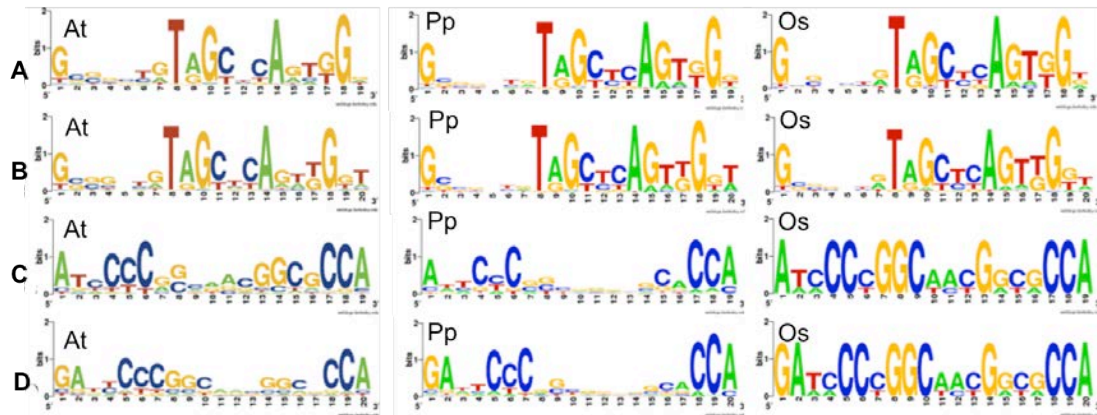
**Figure 3:** Abundance of tRFs in Col-0 seedlings grown in different treatments. All expression values are in comparison with no treated -standard (ST)- samples. Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed). Actin was used as an endogenous control. Two tRFs of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were analyzed. (A) Oxidative stress. (B) Glucose. (C) ABA.



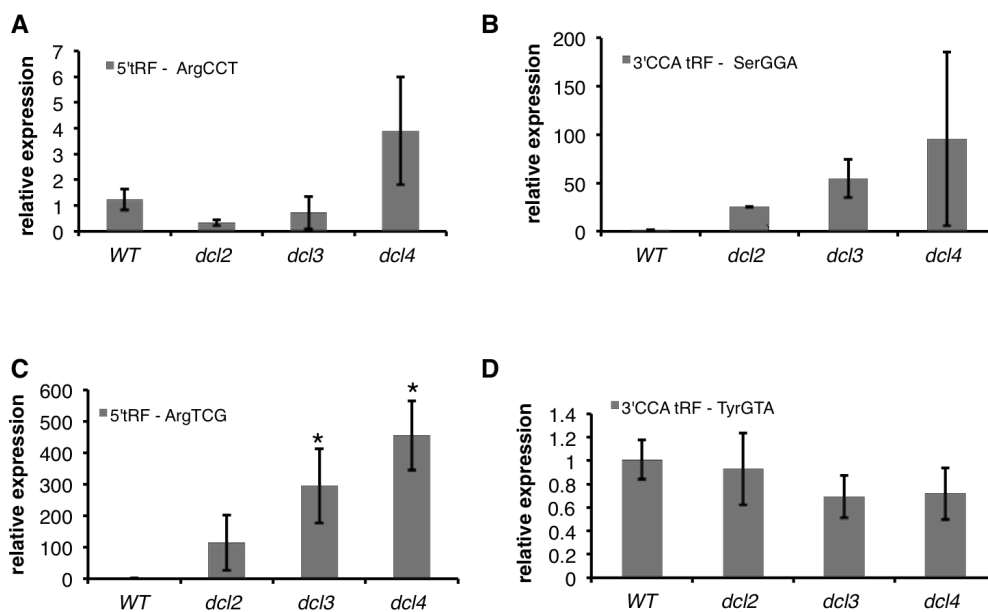
**Figure 4:** Abundance of tRFs in Col-0, *dcl1-11* and *hyl1-2* 10 days old seedlings. All expression values are in comparison to Col-0. Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with Col-0. Each bar represents the mean value and standard deviation of three independent biological replicates. *Actin* was used as an endogenous control. Two tRFs of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were used.



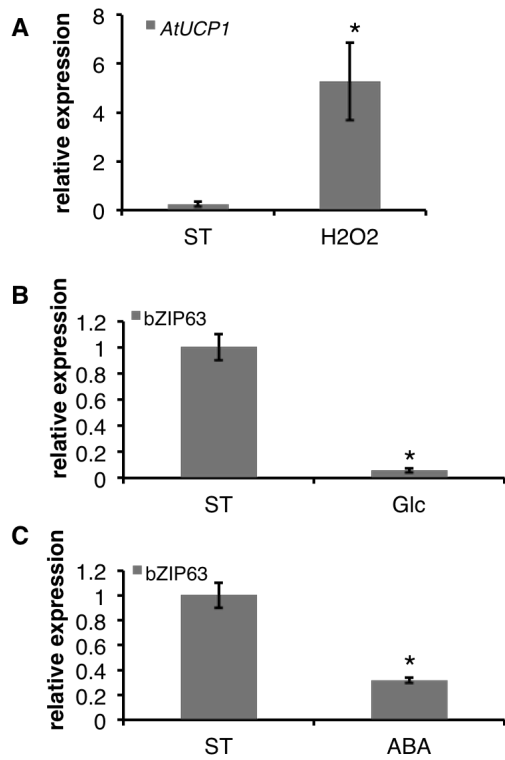
**Figure 5:** Argonaute incorporation: **(A)** Percentage of each 5' terminal nucleotide in total tRFs in *Arabidopsis thaliana* and those enriched in AGO1, AGO2, AGO4 and AGO5 complexes. **(B)** Percentage of each tRF length in AGO1 complex in immunoprecipitation experiments. **(C)** Percentage of each tRF length in AGO2 complex in immunoprecipitation experiments. **(D)** Percentage of each tRF length in AGO4 complex in immunoprecipitation experiments. **(E)** Percentage of each tRF length in AGO5 complex in immunoprecipitation experiments. **(F)** Percentage of each tRF length incorporated in AGO7 complex. **(G)** Enrichment of each tRF in AGO7 complex compared to total input from immunoprecipitation experiments.



**Supplemental Figure 1:** Relative nucleotide bias at each position of the tRFs in total reads from *Arabidopsis thaliana* (At), *Physcomitrella patens* (Pp) and *Oryza sativa* (Os). (A) 19 nucleotides 5'tRFs. (B) 20 nucleotides 5'tRFs. (C) 19 nucleotides 3'tRFs. (D) 20 nucleotides 3'tRFs. *Arabidopsis thaliana*.



**Supplemental Figure 2:** Abundance of tRFs in *Arabidopsis thaliana* 15 days old seedlings. Expression values are in comparison to Col-0 (WT). Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with leaves. Each bar represents the mean value and standard deviation of three independent biological replicates. *Actin* was used as an endogenous control. Two tRF of each class (5'tRF-Arg<sup>CCT</sup> and Arg<sup>TCG</sup> and for 3'CCA tRF-Ser<sup>GGA</sup> and Tyr<sup>GTA</sup>) were used. (A) 5'tRF-Arg<sup>CCT</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*. (B) 3'CCA tRF-Ser<sup>GGA</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*. (C) 5'tRF- Arg<sup>TCG</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*. (D) 3'tRF- Tyr<sup>GTA</sup> abundance in *Atdcl2*, *Atdcl3* and *Atdcl4*.



**Supplementary Figure 4:** Controls genes used for stress treatments. All expression values are in comparison to non-treated samples. Values are mean  $\pm$  SD of three biological replicates. Changes in transcript accumulation were considered significant with  $P \leq 0.05$  according to Student's t-test (two tailed) comparing with leaves. Each bar represents the mean value and standard deviation of three independent biological replicates. Actin was used as an endogenous control. (A) Oxidative stress. (B) Glucose stress. (C) ABA stress.

## References

- [1] Akkuratov EE, Walters L, Saha-Mandal A, Khandekar S, Crawford E, Zirbel CL, et al. Bioinformatics analysis of plant orthologous introns: identification of an intronic tRNA-like sequence. *Gene* 2014;548:81–90. doi:10.1016/j.gene.2014.07.012.
- [2] Bariola PA, Howard CJ, Taylor CB, Verburg MT, Jaglan VD, Green PJ. The Arabidopsis ribonuclease gene RNS1 is tightly controlled in response to phosphate limitation. *Plant J* 1994;6:673–85. doi:10.1046/j.1365-313X.1994.6050673.x.
- [3] Bologna NG, Voinnet O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annu Rev Plant Biol* 2014;65:473–503. doi:10.1146/annurev-arplant-050213-035728.
- [4] Cai P, Piao X, Hao L, Liu S, Hou N, Wang H, et al. A Deep Analysis of the Small Non-Coding RNA Population in Schistosoma japonicum Eggs. *PLoS One* 2013;8. doi:10.1371/journal.pone.0064003.
- [5] Chan PP, Lowe TM. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 2009;37:93–7. doi:10.1093/nar/gkn787.
- [6] Chen D, Meng Y, Yuan C, Bai L, Huang D, Lv S, et al. Plant siRNAs from introns mediate DNA methylation of host genes. *RNA* 2011;17:1012–24. doi:10.1261/rna.2589011.
- [7] Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JWS, et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 2009;15:2147–60. doi:10.1261/rna.1738409.
- [8] Couvillion MT, Bounova G, Purdom E, Speed TP, Collins K. A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. *Mol Cell* 2012;48:509–20. doi:10.1016/j.molcel.2012.09.010.
- [9] Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, et al. MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res* 2009;37:D118–21. doi:10.1093/nar/gkn710.
- [10] Duarte GT, Mantioli CC, Pant BD, Schlereth A, Scheible WR, Stitt M, et al. Involvement of microRNA-related regulatory pathways in the glucose-mediated control of Arabidopsis early seedling development. *J Exp Bot* 2013;64:4301–12. doi:10.1093/jxb/ert239.
- [11] Emara MM, Ivanov P, Hickman T, Dawra N, Tisdale S, Kedersha N, et al. Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J Biol Chem* 2010;285:10959–68. doi:10.1074/jbc.M109.077560.
- [12] Farazi T a, Juranek S a, Tuschl T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 2008;135:1201–14. doi:10.1242/dev.005629.
- [13] Gebetsberger J, Polacek N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol* 2013;10:1798–806. doi:10.4161/rna.27177.

- [14] Gebetsberger J, Zywicki M, Künzi A, Polacek N. tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*. *Archaea* 2012;2012:1–11. doi:10.1155/2012/260909.
- [15] Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-L, Zhang X, et al. Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science* (80- ) 2006;311:395–8. doi:10.1126/science.1120976.
- [16] Hackenberg M, Huang P-J, Huang C-Y, Shi B-J, Gustafson P, Langridge P. A Comprehensive Expression Profile of MicroRNAs and Other Classes of Non-Coding Small RNAs in Barley Under Phosphorous-Deficient and -Sufficient Conditions. *DNA Res* 2012:1–17. doi:10.1093/dnares/dss037.
- [17] Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay M a. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 2010;16:673–95. doi:10.1261/rna.2000810.
- [18] Hillwig MS, LeBrasseur ND, Green PJ, MacIntosh GC. Impact of transcriptional, ABA-dependent, and ABA-independent pathways on wounding regulation of RNS1 expression. *Mol Genet Genomics* 2008;280:249–61. doi:10.1007/s00438-008-0360-3.
- [19] Hopper AK, Pai D a, Engelke DR. Cellular dynamics of tRNAs and their genes. *FEBS Lett* 2010;584:310–7. doi:10.1016/j.febslet.2009.11.053.
- [20] Hopper AK, Shaheen HH. A decade of surprises for tRNA nuclear-cytoplasmic dynamics. *Trends Cell Biol* 2008;18:98–104. doi:10.1016/j.tcb.2008.01.001.
- [21] Hsieh L-C, Lin S-I, Shih AC-C, Chen J-W, Lin W-Y, Tseng C-Y, et al. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiol* 2009;151:2120–32. doi:10.1104/pp.109.147280.
- [22] Hurto RL. Unexpected functions of tRNA and tRNA processing enzymes. In: Collins LJ, editor. *RNA Infrastruct. Networks*, Landes Bioscience and Springer Science; 2011, p. 137–55. doi:10.1007/978-1-4614-0332-6\_9.
- [23] Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell* 2011;43:613–23. doi:10.1016/j.molcel.2011.06.022.
- [24] Kanai A. Disrupted tRNA Genes and tRNA Fragments: A Perspective on tRNA Gene Evolution. *Life* 2015;5:321–31. doi:10.3390/life5010321.
- [25] Kim VN. Sorting out small RNAs. *Cell* 2008;133:25–6. doi:10.1016/j.cell.2008.03.015.
- [26] Kumar P, Anaya J, Mudunuri SB, Dutta A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol* 2014;12:1–14.
- [27] Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* 2009;23:2639–49. doi:10.1101/gad.1837609.

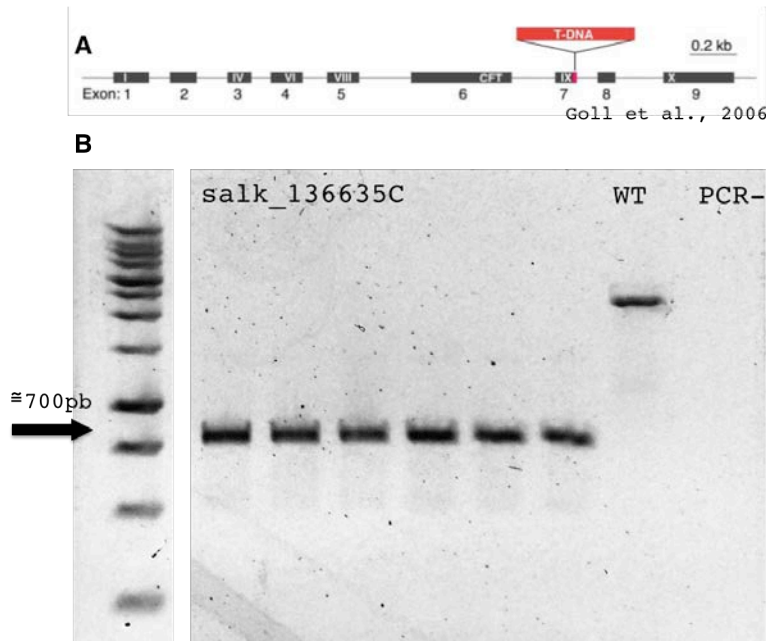


- [28] Li Z, Ender C, Meister G, Moore PS, Chang Y, John B. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res* 2012;40:6787–99. doi:10.1093/nar/gks307.
- [29] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method. *Methods* 2001;25:402–8. doi:10.1006/meth.2001.1262.
- [30] Loss-Morais G, Waterhouse PM, Margis R. Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. *Biol Direct* 2013;8:6. doi:10.1186/1745-6150-8-6.
- [31] MacIntosh GC, Hillwig MS, Meyer A, Fligel L. RNase T2 genes from rice and the evolution of secretory ribonucleases in plants. *Mol Genet Genomics* 2010;283:381–96. doi:10.1007/s00438-010-0524-9.
- [32] Mei Y, Yong J, Liu H, Shi Y, Meinkoth J, Dreyfuss G, et al. tRNA binds to cytochrome c and inhibits caspase activation. *Mol Cell* 2010;37:668–78. doi:10.1016/j.molcel.2010.01.023.
- [33] Mittler R, Blumwald E. The Roles of ROS and ABA in Systemic Acquired Acclimation. *Plant Cell Online* 2015;27:64–70. doi:10.1105/tpc.114.133090.
- [34] Moore B, Zhou L, Rolland F, Hall Q, Cheng W-H, Liu Y-X, et al. Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* (80- ) 2003;300:332–6. doi:10.1126/science.1080585.
- [35] Müller S, Windhof IM, Maximov V, Jurkowski T, Jeltsch A, Förstner KU, et al. Target recognition, RNA methylation activity and transcriptional regulation of the *Dictyostelium discoideum* Dnmt2-homologue (DnmA). *Nucleic Acids Res* 2013;41:8615–27. doi:10.1093/nar/gkt634.
- [36] Ortiz-Morea FA, Vicentini R, Silva GFF, Silva EM, Carrer H, Rodrigues AP, et al. Global analysis of the sugarcane microtranscriptome reveals a unique composition of small RNAs associated with axillary bud outgrowth. *J Exp Bot* 2013;64:2307–20. doi:10.1093/jxb/ert089.
- [37] Parisien M, Wang X, Pan T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol* 2013;10:1853–67. doi:10.4161/rna.27361.
- [38] Phizicky EM, Hopper AK. tRNA biology charges to the front. *Genes Dev* 2010;24:1832–60.
- [39] Raina M, Ibba M. tRNAs as regulators of biological processes. *Front Genet* 2014;5:171. doi:10.3389/fgene.2014.00171.
- [40] Rook F, Hadingham S a., Li Y, Bevan MW. Sugar and ABA response pathways and the control of gene expression. *Plant, Cell Environ* 2006;29:426–34. doi:10.1111/j.1365-3040.2005.01477.x.
- [41] Saikia M, Krokowski D, Guan BJ, Ivanov P, Parisien M, Hu GF, et al. Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *J Biol Chem* 2012;287:42708–25. doi:10.1074/jbc.M112.371799.

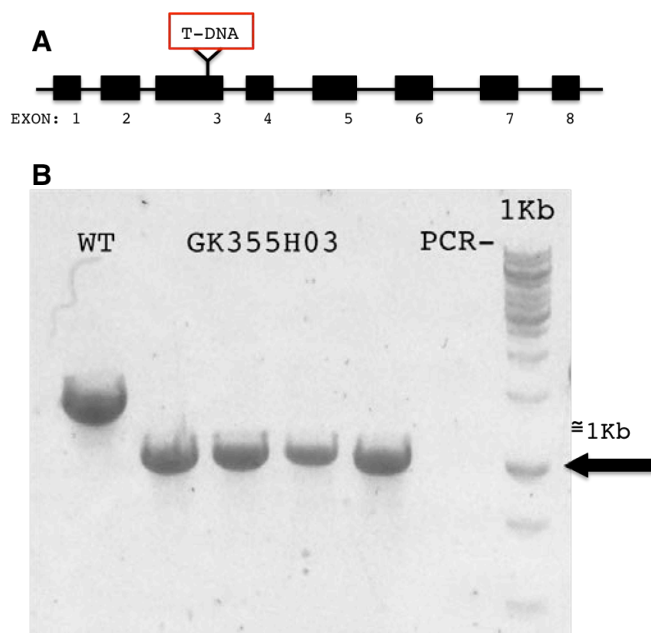
- [42] Sobala A, Hutvagner G. Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. *RNA Biol* 2013;10:553–63. doi:10.4161/rna.24285.
- [43] Sobala A, Hutvagner G. Transfer RNA-derived fragments: origins, processing, and functions. *Wiley Interdiscip Rev RNA* 2011;2:853–62. doi:10.1002/wrna.96.
- [44] Thompson DM, Lu C, Green PJ, Parker R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* 2008;14:2095–103. doi:10.1261/rna.1232808.
- [45] Tsutsui T, Nakano a., Ueda T. The Plant-Specific RAB5 GTPase ARA6 is Required for Starch and Sugar Homeostasis in *Arabidopsis thaliana*. *Plant Cell Physiol* 2015;0:1–11. doi:10.1093/pcp/pcv029.
- [46] Udvardi MK, Czechowski T, Scheible W-R. Eleven golden rules of quantitative RT-PCR. *Plant Cell* 2008;20:1736–7. doi:10.1105/tpc.108.061143.
- [47] Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* 2007;3:12. doi:10.1186/1746-4811-3-12.
- [48] Vazquez F, Gascioli V, Cr  t   P, Vaucheret H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr Biol* 2004;14:346–51. doi:10.1016/j.cub.2004.01.035.
- [49] Walker SC, Engelke DR. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit Rev Biochem Mol Biol* 2006;41:77–102. doi:10.1080/10409230600602634.
- [50] Xie Z, Allen E, Wilken A, Carrington JC. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 2005;102:12984–9. doi:10.1073/pnas.0506426102.
- [51] Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, et al. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* 2004;2:E104. doi:10.1371/journal.pbio.0020104.
- [52] Yamasaki S, Ivanov P, Hu G-F, Anderson P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J Cell Biol* 2009;185:35–42. doi:10.1083/jcb.200811106.
- [53] Yamasaki S, Ivanov P, Hu G-F, Anderson P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J Cell Biol* 2009;185:35–42. doi:10.1083/jcb.200811106.
- [54] Yeung ML, Bennasser Y, Watashi K, Le SY, Houzet L, Jeang KT. Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: Evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic Acids Res* 2009;37:6575–86. doi:10.1093/nar/gkp707.
- [55] Zhang J-F, Yuan L-J, Shao Y, Du W, Yan D-W, Lu Y-T. The disturbance of small RNA pathways enhanced abscisic acid response and multiple stress responses in *Arabidopsis*. *Plant Cell Environ* 2008;31:562–74. doi:10.1111/j.1365-3040.2008.01786.x.

## *Apêndices*

## Apêndice 1 - Genotipagem dos mutantes *Atdnmt2* e *Athd2a*



**Figura 1:** Genotipagem das plantas mutantes de arábido, que não produzem a enzima DNMT2, obtidas através do TAIR (<http://www.arabidopsis.org>). **(A)** Esquema retirado de Goll et al., 2006 demonstrando a região de inserção do T-DNA. **(B)** PCR para genotipagem das plantas mutantes, obtido com iniciadores *forward* e reverso para o *locus* correspondente à região codificadora da enzima DNMT2 e um iniciador *forward* para o T-DNA. As bandas correspondentes às amostras de salk\_136635C são únicas e de aproximadamente 700pb, tamanho esperado para mutantes homocigotos contendo a inserção do T-DNA. WT: controle positivo da reação, contendo DNA de arábido selvagem. PCR-: controle negativo da reação.

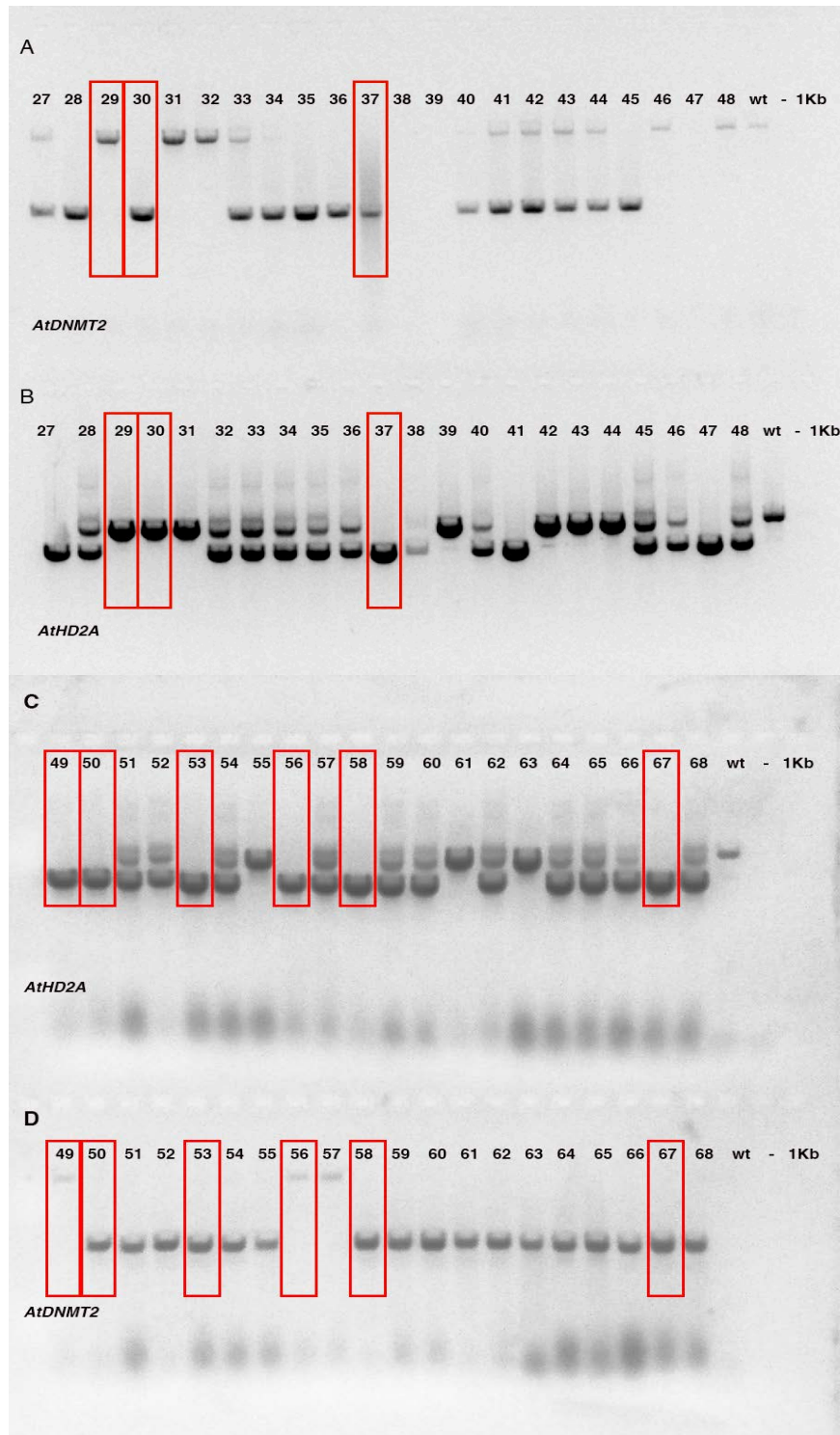


**Figura 2:** Genotipagem das plantas mutantes de arábido, que não produzem a enzima HD2A, obtidas através do TAIR (<http://www.arabidopsis.org>). **(A)** Esquema demonstrando a região de inserção do T-DNA. **(B)** PCR para genotipagem das plantas mutantes, obtido com iniciadores *forward* e reverso para o *locus* correspondente à região codificadora da enzima HD2A e um iniciador *forward* para o T-DNA. As bandas correspondentes às amostras de GK355H03 são únicas e de aproximadamente 1Kb, tamanho esperado para mutantes homocigotos contendo a inserção do T-DNA. WT: controle positivo da reação, contendo DNA de arábido selvagem. PCR-: controle negativo da reação

## Apêndice 2- Tabela de oligonucleotídeos utilizados:

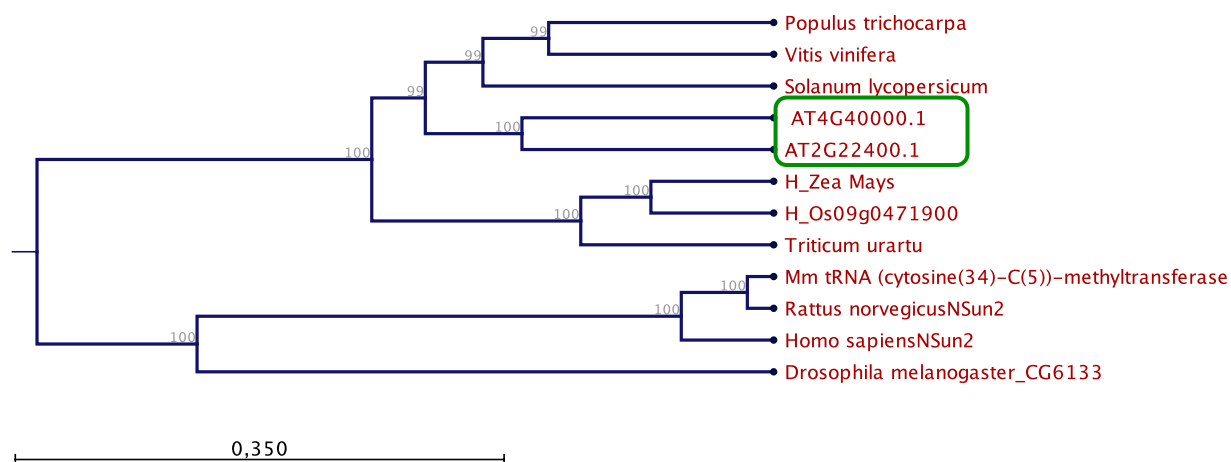
OLIGO	SEQUÊNCIA	IDENTIDADE
Gabi LB 08409	ATATTGACCATCATACTCATTGC	
LBbl.3_foward	ATTTTGCCGATTTTCGGAAC	
qAt_PP2A - forward	TGATGCAATCTCTCATTCGGATAG	AT1G13320
qAt_PP2A - reverse	AGCATGGCCGTATCATGTTCT	AT1G13320
At_dnmt2 - forward	CACCGGATCCATTAATTTTGTTAAGCAGCAGC	AT5G25480
At_dnmt2 - reverse	GCAGAGCTCAATTATGATTAACGGGCCGA	AT5G25480
qAt_dnmt2 - forward	GGCTTTTTTCGTTTCTCAGGATTCTT	AT5G25480
qAt_dnmt2 - reverse	CCGAAGTCTCAAATCCAACAACA	AT5G25480
At_hd2a - forward	CTTAGCCATGGAGTTCTGGG	AT3G44750
At_hd2a - reverse	CAAGTGCATTCCCTGAGTTG	AT3G44750
qAt_hd2a - forward	GGGGAAAGGCTGCAAACCAG	AT3G44750
qAt_hd2a - reverse	GCTTGGCCTTGTTGTGAGACTC	AT3G44750
qAt_hd2b - forward	GAATGCTAACCAGAGCCCCAAG	AT5G22650
qAt_hd2b - reverse	CTCTACCCTTCCCTTGCCC	AT5G22650
qAt_hd2c - forward	GCCAAGCAAGATGATGACGC	AT5G03740
qAt_hd2c - reverse	CTCTTCCTCCTCGTCTCCAGAG	AT5G03740
qAt_hd2d - forward	GGTGGCTCCTCCGAGCAAAAAG	AT2G27840
qAt_hd2d - reverse	GGCATCTCTTCTTCCCTCCTG	AT2G27840
AtUCP1- forward	GATGGTGGCGGCTGGTA	AT3G54110
AtUCP1 - reverse	CGCCGACGCAAGCAG	AT3G54110
At_csd2- forward	GTGGACAATCAGATTCCTCTG	AT2G28190
At_csd2 - reverse	TTCCAGTGGTCAGACTAAGC	AT2G28190
At_actin - forward	GACCTTGCTGGACGTGACCTTAC	AT3G18780
At_actin - reverse	GTAGTCAACAGCAACAAGGAGAGC	AT3G18780
At_arf3- forward	ATCACCTCTTCCGTCTTGCTT	AT2G24765
At_arf3 - reverse	GGAAACCAAACAGCCTGCATCT	AT2G24765
At_rns1 - forward	CGCTCAAACGAAGTAATTCTGAGG	AT2G02990
At_rns1 - reverse	GTCACAGTATGATCCTGGCC	AT2G02990
Universal Reverse	GTGCAGGGTCCGAGGT	
At_trf1-ArgTCG-F	CTGACCGCATAGCGCAGTGGAT	At_chr1.trna186-ArgTCG
At_trf1-ArgTCG-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACATCCAC	At_chr1.trna186-ArgTCG
At_trf1-ArgCCT-F	CTGCGCCTGTAGCTCAGTGGGA	At_chr1.trna74-ArgCCT
At_trf1-ArgCCT-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCCACT	At_chr1.trna74-ArgCCT
At_trf2-TyrGTA-F	CTAAATCCAGCTCGGCCACCA	At_chr3.trna24-TyrGTA
At_trf2-TyrGTA-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGGTGG	At_chr3.trna24-TyrGTA
At_trf2-SerGGA-F	CTGAATCCCTCTCTTTCCCCCA	At_chr2.trna11-SerGGA
At_trf2-SerGGA-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGGGGG	At_chr2.trna11-SerGGA
At-Asp <sup>GTC</sup> - R	CATCACCATTACCAAAAATCAAACCC	AT5G40545
At-Asp <sup>GTC</sup> - F	CATCGGTTGTGTAGTATAGTGGTAAG	AT5G40545
At-Asp <sup>GTC</sup> - NF	CATGTTGTTGTAGTATAGTGGTAAGTATTTTGTG	AT5G40545
LNA - At_tRNA <sup>asp</sup> (GTC)	CT+TAC+CAC+TAT+ACT+ACA+ACG	AT5G40545(Goll et al., 2006)

### Apêndice 3 - Genotipagem dos duplos mutantes *Atdnmt2;Athd2a*



**Figura 1:** Produtos de PCR das genotipagens dos duplos mutantes. Os números representam os indivíduos provenientes dos cruzamentos. Os retângulos vermelhos indicam alguns dos genótipos selecionados, onde: 29- s.WT, 30- s.dnmt2, 37- HAD, 49- s.hd2a, 50- HAD, 53-HAD, 55- s.dnmt2, 56- s.hd2a, 58- HAD, 61- s.dnmt2, 63, s.dnmt2, 67-HAD. **(A)** Detecção do T-DNA do mutante *dnmt2* da F2 dos cruzamentos HAD. **(B)** Detecção do T-DNA do mutante *hd2a* da F2 dos cruzamentos HAD. **(C)** Detecção do T-DNA dos mutantes *hd2a* **(D)** Detecção do T-DNA dos mutantes *dnmt2* da F2 do cruzamento HAD.

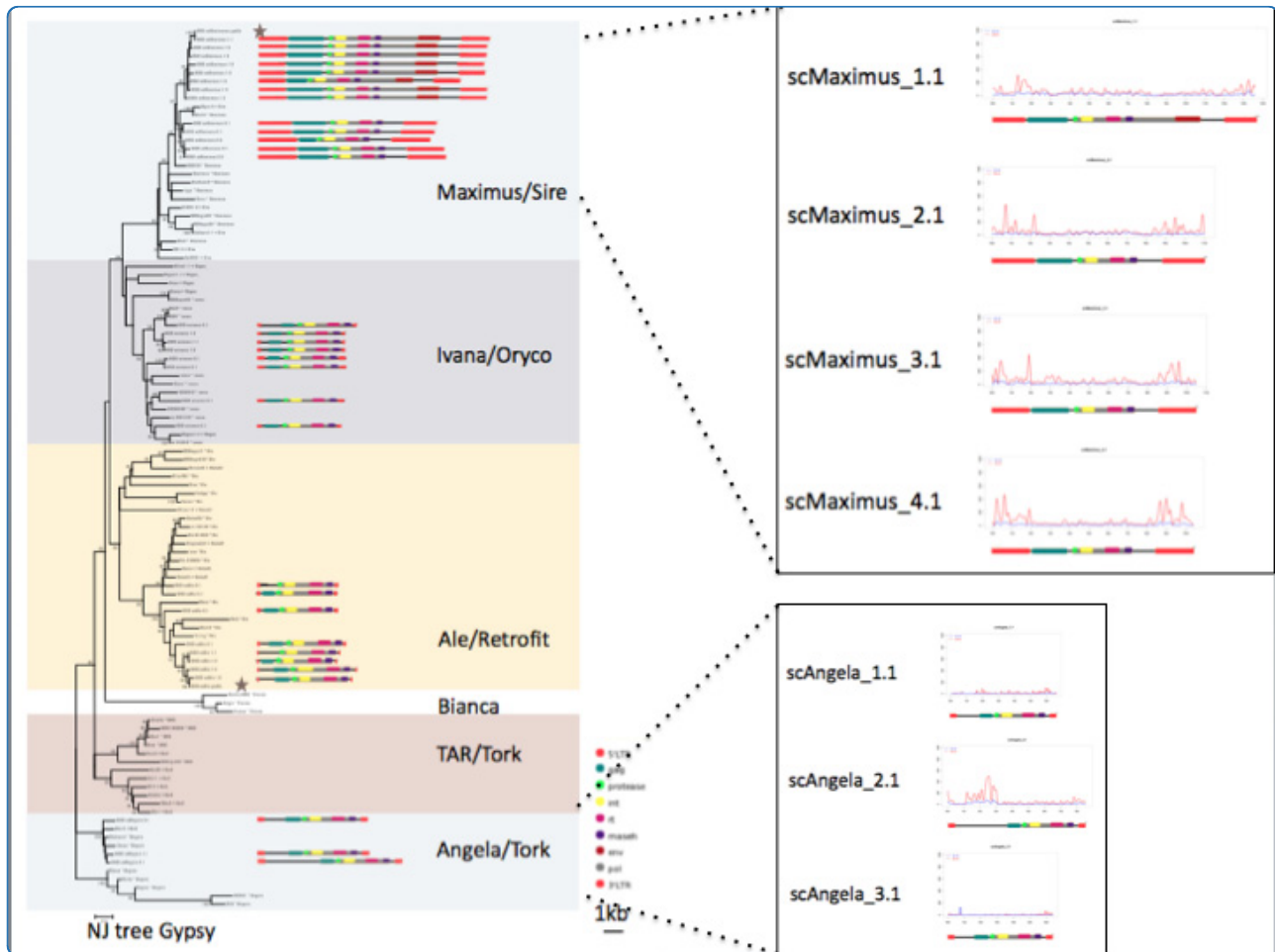
## Apêndice 4 - Dendograma de similaridade da enzima NSun2



**Figura 1:** Árvore filogenética de similaridade construída utilizando o método de máxima verossimilhança (ML), com 1000 replicatas *Bootstrap*, demonstrando a relação entre as enzimas NSun2 de diferentes organismos, tanto animais, quanto vegetais, o retângulo destaca as duas proteínas homólogas à NSun2 em *A. thaliana*. O valor de *bootstrap* está representado nos nós dos ramos.

*Apêndice 5* – Artigo publicado como colaboradora





## Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns

Domingues *et al.*

RESEARCH ARTICLE

Open Access

# Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns

Douglas S Domingues<sup>1,4†</sup>, Guilherme MQ Cruz<sup>1†</sup>, Cushla J Metcalfe<sup>1</sup>, Fabio TS Nogueira<sup>2</sup>, Renato Vicentini<sup>3</sup>, Cristiane de S Alves<sup>2</sup> and Marie-Anne Van Sluys<sup>1\*</sup>

## Abstract

**Background:** Sugarcane is an important crop worldwide for sugar production and increasingly, as a renewable energy source. Modern cultivars have polyploid, large complex genomes, with highly unequal contributions from ancestral genomes. Long Terminal Repeat retrotransposons (LTR-RTs) are the single largest components of most plant genomes and can substantially impact the genome in many ways. It is therefore crucial to understand their contribution to the genome and transcriptome, however a detailed study of LTR-RTs in sugarcane has not been previously carried out.

**Results:** Sixty complete LTR-RT elements were classified into 35 families within four *Copia* and three *Gypsy* lineages. Structurally, within lineages elements were similar, between lineages there were large size differences. FISH analysis resulted in the expected pattern of *Gypsy*/heterochromatin, *Copia*/euchromatin, but in two lineages there was localized clustering on some chromosomes. Analysis of related ESTs and RT-PCR showed transcriptional variation between tissues and families. Four distinct patterns were observed in sRNA mapping, the most unusual of which was that of *Ale1*, with very large numbers of 24nt sRNAs in the coding region. The results presented support the conclusion that distinct small RNA-regulated pathways in sugarcane target the lineages of LTR-RT elements.

**Conclusions:** Individual LTR-RT sugarcane families have distinct structures, and transcriptional and regulatory signatures. Our results indicate that in sugarcane individual LTR-RT families have distinct behaviors and can potentially impact the genome in diverse ways. For instance, these transposable elements may affect nearby genes by generating a diverse set of small RNA's that trigger gene silencing mechanisms. There is also some evidence that ancestral genomes contribute significantly different element numbers from particular LTR-RT lineages to the modern sugarcane cultivar genome.

**Keywords:** LTR Retrotransposons, Sugarcane, Genome, FISH, Epigenetic, Small RNA

## Background

Plant genome evolution is closely associated with polyploidy and gene amplification, the most highly amplified genes being Long Terminal Repeat retrotransposons (LTR-RTs) [1,2]. LTR-RT proliferation is regulated by the genome at both the transcriptional and post-transcriptional level [3]. LTR-RTs, like other Transposable Elements (TEs), can not only affect a genome by

expansion by proliferation, but also by, for example, providing a template for recombination, inserting into coding regions and disrupting gene expression, or affecting the transcription of neighboring genes (reviewed by [4]).

Based on the massive amounts of genomic sequence data released in the last 15 years, phylogenetic analysis of plant LTR-RTs has identified distinct evolutionary lineages within *Gypsy* and *Copia*, the two main plant LTR-RT superfamilies [5]. These lineages are widespread in both monocot and eudicot genomes. Uncovering of this fine structuring within *Gypsy* and *Copia*

\* Correspondence: mavsluys@usp.br

† Contributed equally

<sup>1</sup>GaTE Lab, Depto. de Botânica, Inst. de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-090 São Paulo, Brazil

Full list of author information is available at the end of the article

plant LTR-RTs has lead to a better understanding of the diversification of LTR-RTs and shed light on their structure and genomic distribution [6-8].

Sugarcane is an important crop worldwide, being a major source of sugar, and is also increasingly being used for the production of renewable energy sources such as ethanol. Despite its economic importance, the sequencing of the sugarcane genome is at the pilot stage [9]. Modern cultivars are highly polyploid and have one of the largest (10Gb) and most complex crop genomes, and chromosome numbers varying between 100-130 [10].

There are a few reports on transposable elements in sugarcane sequences [11-14] but there is no detailed study on sugarcane LTR-RTs, their phylogenetic classification, and potential impact on genomic distribution and transcriptional activity. In addition, the importance of accurate and complete TE annotation is increasingly recognized as a priority in plant genome sequencing projects to minimize the inaccuracy of gene annotation and facilitate functional gene studies [15].

To set the basis for future genome interpretation, LTR-RTs in the sugarcane genome were characterized using available public resources and BAC sequences from the BIOEN project [16]. We examined the structure, genomic distribution, phylogenetic diversity, transcriptional activity and regulation of sugarcane LTR-RTs. This study reveals that within well-defined phylogenetic lineages, that while TEs within LTR-RTs families are structurally similar, they have distinct transcriptional and regulatory signatures. Taken together, these results support the growing evidence that LTR-RTs contribute to genomic diversity, but with a wide range of potential outcomes.

## Results

### Characterization of LTR retrotransposons in sugarcane: phylogeny and structural features

Plant LTR-RT evolutionary lineage names are not used consistently within the literature [6-8], we therefore chose to include sequences from more than one source, so that we could directly compare our results with published data. Our inferred evolutionary histories suggest that, at least for the sequences we analyzed within *Gypsy* superfamily, the *DEL* lineage is equivalent to *Tekay*, and within the *Copia* superfamily, the *Maximus* lineage is equivalent to *Sire*, *Ale* to *Retrofit* and *Ivana* to *Oryco* (Figures 1 and 2; Additional file 1). The relationship between the *TAR*, *Angela*, *Tork* and *Bianca* lineages is more problematic. While the *Bianca* lineage is not included in the GyDb [8] *Tork* appears to be the *Angela* and *TAR* lineages combined (Figure 2; Additional file 1: Figure S2).

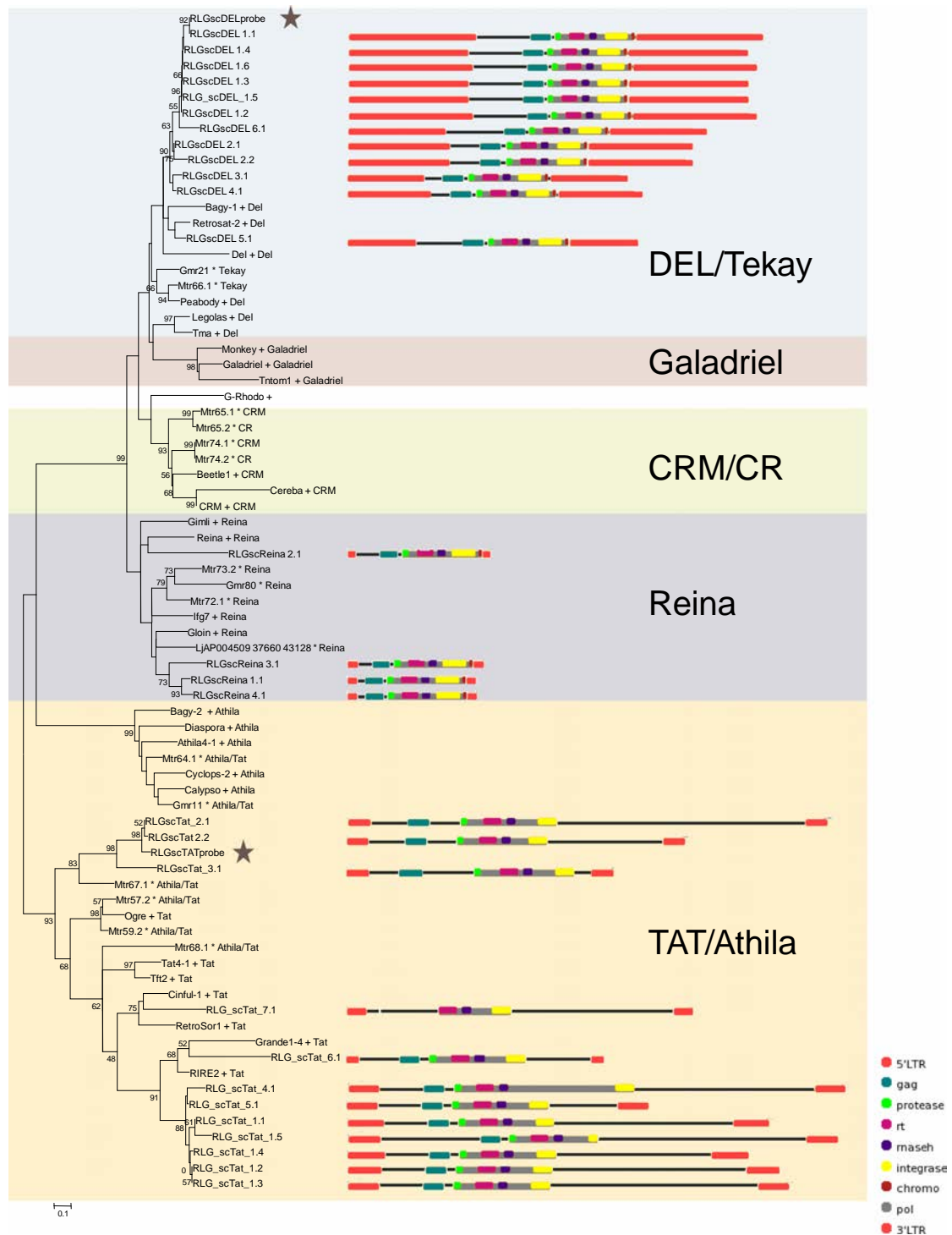
Of the sixty full-length sequences extracted from sugarcane BAC sequences, thirty-two sequences were assigned to the *Copia* superfamily and twenty-eight to the *Gypsy* superfamily (Table 1; Figures 1 and 2). We identified four of the six major plant evolutionary *Copia* lineages described by Wicker and Keller [6] in the sugarcane genome (Figure 2; Additional file 1: Figure S2), and three of the six *Gypsy* lineages described by Du et al. [7] and the GyDB [8] (Figure 1; Additional file 1: Figure S1). In terms of sequence numbers, the *Maximus* lineage was the most highly represented within the *Copia* superfamily, and the *Del* and *Tat* lineages were equally represented within the *Gypsy* superfamily. Based on sequence identity within the LTRs, the elements were classified into 35 families (Table 1) [5]. In order to simplify reading of the text we have referred to these families as simply '*Ale1*' for example, rather than the by the full name '*RLC\_scAle1*'.

We were able to identify all the internal coding domains (gag, aspartic protease, reverse transcriptase, integrase and RNaseH) in all the sugarcane sequences, apart from the gag domain in *RLG\_scTat\_7.1* (Figure 1). In addition, a putative envelope domain was identified in elements from the *Maximus1* family (Figure 2) and a chromodomain in all *DEL* and *Reina* elements (Figure 1) [8].

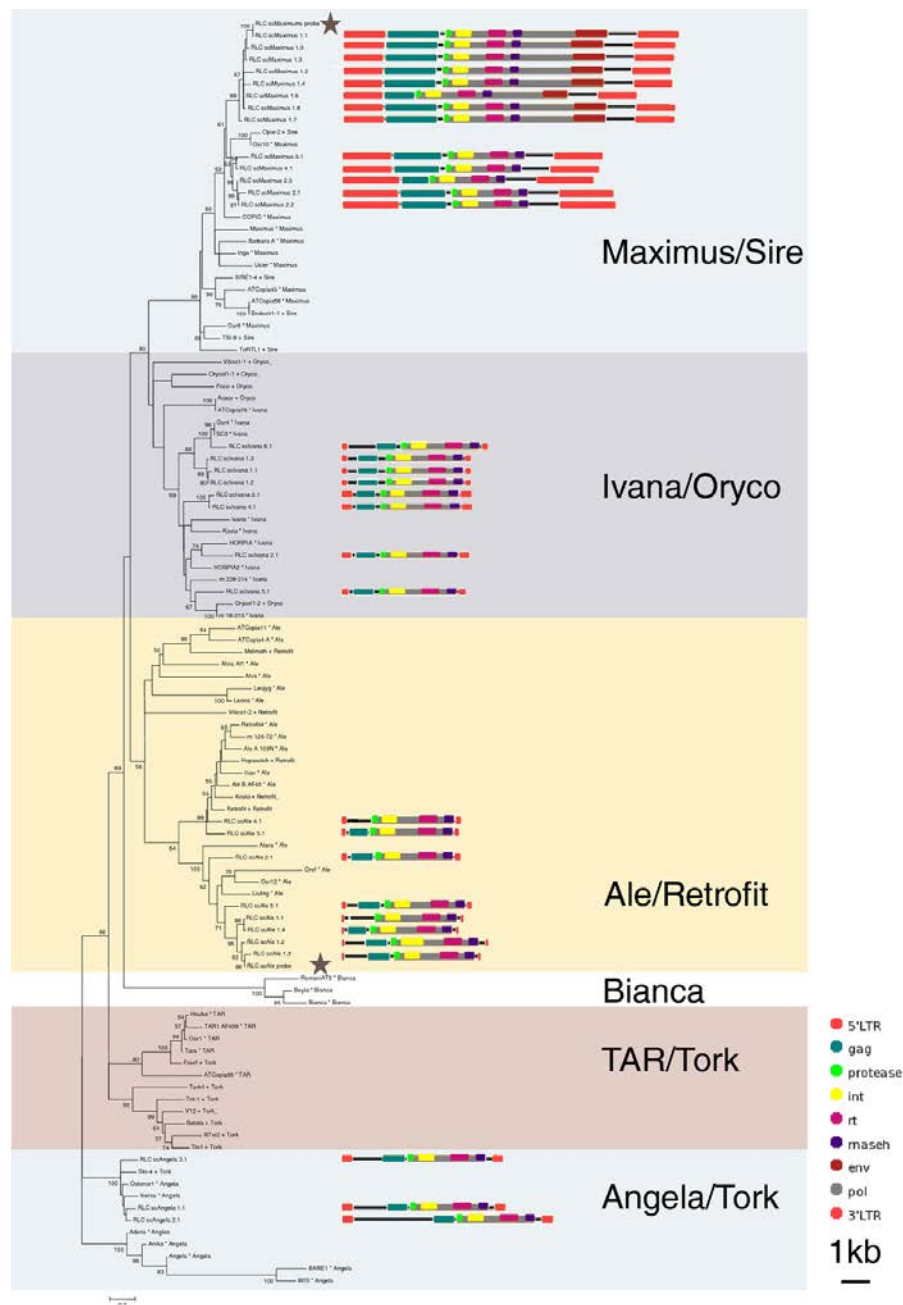
There was a general pattern in overall size of elements from lineages in each of the two superfamilies. In both superfamilies, there was one lineage that is very large (*Maximus* and *Del*, 10.2-16.7 kb), and at least one lineage that is comparatively smaller (*Ivana*, *Ale* and *Reina*, 4.7-5.9 kb) (Figures 1 and 2). Within the *Copia* superfamily, there is a lineage that is an in-between size (*Angela* at 6.5-8.5 kb), while within the *Gypsy* superfamily, the sequences of the *Tat* lineage (9.2-17.7 kb) are more comparable in size to those of the *DEL* lineage (11.3-16.7 kb) (Figures 1 and 2). Differences in total length were chiefly due to differences in LTR size, and the presence and size of spacer regions between the internal coding domain and the LTRs, rather than insertions within the gag/pol coding regions (Figures 1 and 2). The exception is the *Tat* elements, which contain insertions within the pol region, but in most cases between actual coding domains (Figure 1).

### Distribution of LTR retrotransposons in metaphases

Given that transposable elements can be distributed throughout a genome, it is difficult to distinguish between random and real signals from a probe derived from a transposable element. The probe for the *Del* lineage was therefore prepared twice, one probe labeled with DIG and the other with biotin, and hybridized separately to the same slide before and after stripping.



**Figure 1 Neighbour-joining (NJ) phylogenies of Copia and Gypsy families.** NJ phylogeny of Gypsy families based on reverse transcriptase, with schematic representations of sugarcane elements. Sequences from the Gypsy database [8] are denoted with a plus symbol, those from Du *et al.* [7] with a star. Robustness of the nodes was estimated by 500 bootstrap replications. Bootstrap values below 50% are not shown. Lineages are indicated by names and colored blocks. A star indicates the sequences used as probes for fluorescent *in situ* hybridization (Figure 3). Schematic representations were created using domain draw [17]. A scale and a key for the domains represented in the schematic representations are shown in the bottom right hand corner. Abbreviations and color coding of domains: LTR = long terminal repeat (orange); gag = Gag (dark green); protease = Protease (light green); rt = Reverse Transcriptase (pink); maseh = Ribonuclease H (purple); integrase = Integrase (yellow); chromo = Chromodomain (brown); env = Envelope (brown); pol = Polyprotein (grey).



**Figure 2 Neighbour-joining (NJ) phylogenies of Copia and Gypsy families.** NJ phylogeny of Copia families based on reverse transcriptase, with schematic representations of sugarcane sequences. Sequences from the Gypsy database [8] are denoted with a plus symbol, those from Wicker and Keller [6] with a star. Robustness of the nodes was estimated by 500 bootstrap replications. Bootstrap values below 50% are not shown. Lineages are indicated by names and colored blocks. A star indicates the sequences used as probes for fluorescent *in situ* hybridization (Figure 3). Schematic representations were created using domain draw [17]. A scale and a key for the domains represented in the schematic representations are shown in the bottom right hand corner. Abbreviations and color coding of domains: LTR = long terminal repeat (orange); gag = Gag (dark green); protease = Protease (light green); rt = Reverse Transcriptase (pink); maseh = Ribonuclease H (purple); integrase = Integrase (yellow); chromo = Chromodomain (brown); env = Envelope (brown); pol = Polyprotein (grey).

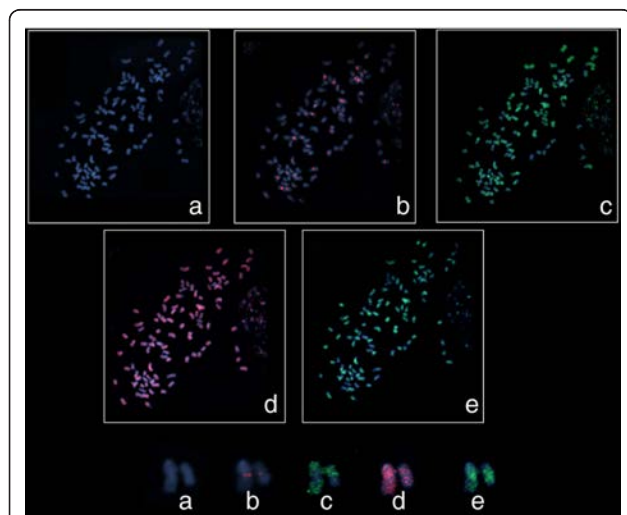
The distribution of the *Del* probe was compared on the same 10 metaphases visualized with anti-digoxigenin-rhodamine (Red) or with NeutrAvidin-Oregon Green-488 (Green) (Additional file 2). The distribution of the

probe signal was similar in all 10 metaphases, and therefore the signal was considered to be valid. Similar *in situ* hybridization conditions were used for all subsequent experiments. A BAC clone (SCHRBA\_239\_N21),

**Table 1 General features of sugarcane LTR-RT lineages**

Superfamily/Lineage	Size (kb)	LTR (bp)	Families	Sequences
<i>Copia</i>				
Ale	4.7-5.9	116-238	5	8
Angela	6.5-8.5	434-461	3	3
Ivana	5-5.9	238-454	6	8
Maximus	10.2-13.6	1607-2004	4	13
Total <i>Copia</i>			18	32
<i>Gypsy</i>				
DEL	11.3-16.7	2762-5139	6	12
Reina	5.1-5.7	315-416	4	4
TAT	9.2-17.7	458-1345	7	12
Total <i>Gypsy</i>			17	28
Overall Total			35	60

identified by our group [16], contains known sugarcane centromeric repeats [18,19]. This BAC clone hybridized to the middle region of the chromosome, which consists mainly of centromeric specific sequences (Figure 3b). Seven LTR-RT probes were used, three of which are from *Gypsy* superfamily members and four from *Copia* superfamily members, representing all the major lineages already described. No convincing signal was obtained for the *Reina1*, *Ivana1* and *Angela1* probes,



**Figure 3 Fluorescent *in situ* hybridization patterns observed for sugarcane LTR-RTs.** A pair of representative chromosomes is shown underneath the metaphase spreads. Chromosomes are stained with DAPI, probes were labelled with Digoxigenin (DIG) and detected with Anti-DIG-Rhodamine (red signal) or with Biotin and detected with NeutrAvidin-Oregon Green (green signal) (a) DAPI counterstaining only. (b) BAC SCHRba\_239\_N21 which contains known sugarcane centromeric repeats (c) scAle probe (euchromatic pattern). (d) scMaximus probe (dispersed pattern) (e) scDELprobe (centromeric/pericentromeric pattern). The LTR-RT probes were 1.9-2.9 kb and included the reverse transcriptase domain.

using the same conditions that were used for the other LTR-RT probes.

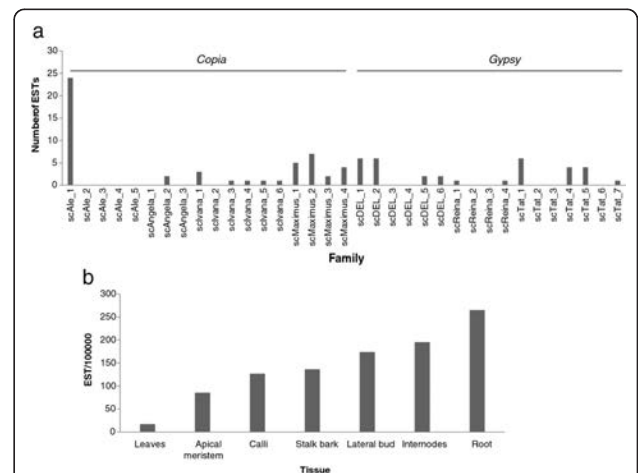
The probe from one of the two *Gypsy* lineages, *Del1*, localized almost exclusively in a broad band around the centromeres. This observation suggests that it is a heterochromatic or pericentromeric specific element (Figure 3e). The second *Gypsy* element examined, *Tat2*, displays a generally strong broad dispersed pattern, but with some concentration along some chromosome arms (data not shown).

Two *Copia* probes were found dispersed along the chromosome arms and none localized exclusively at or near the centre of the chromosomes. The *Ale1* lineage probe was found in high concentrations along particular chromosome arms (Figure 3c). Signals from the *Maximus* lineage probe were generally dispersed (Figure 3d).

**Transcriptional activity of sugarcane LTR retrotransposons**

We associated 84 ESTs from the sugarcane cultivar SP80-3280 related to full-length LTR retrotransposons. The largest number of transcripts was identified from the root libraries (Figure 4b), followed by the internode, lateral bud and calli libraries.

The *Copia* superfamily was more highly represented in the EST database than the *Gypsy* superfamily, with 51 and 33 sequences, respectively (Figure 4a). Previously described full length cDNA sequences [20] were assigned to 8 of the 35 families we identified from BAC sequences (Additional file 3). In this study, no ESTs were identified for 14 families (Figure 4a). We assigned ESTs to another 14 families not previously described as transcriptionally active [11,20] (Figure 4a), however for



**Figure 4 Transcriptional activity of sugarcane LTR-RTs.** ESTs from the cultivar SP80-3280 available from NCBI were assigned to an LTR-RT family, using the criteria described by Wicker et al. (2007) [5]. (a) Total normalized number of ESTs assigned to each LTR-RT family and (b) Total normalized number of ESTs by tissue type.

one family, *Ale3*, for which a previously described cDNA sequence was assigned, no ESTs were identified.

RT-PCR analysis confirmed the transcriptional activity of sugarcane LTR-RTs in leaves and lateral buds. We also experimentally confirmed transcriptional activity for five new families: *Ale2*, *Ivana6*, *Del2*, *Reina1* and *Reina3*.

The most represented family in the EST data, *Ale1*, had an intense band, confirming its transcriptional activity, and had the same intensity of signal in both leaves and lateral buds. Transcriptional patterns differed between tissues and LTR-RT families (Figure 5a). The *Ivana* and *Reina* lineages illustrate that individual families have distinct transcriptional patterns: for instance, the *Ivana6* signal is more intense than that of *Ivana1*; moreover, *Ivana1* has higher transcriptional activity in lateral buds, while *Ivana6* is more active in leaves. *Reina1* and *Reina3* are both transcriptionally active in lateral buds, but clear differences in band intensity can be observed (Figure 5a).

#### Distribution of small RNAs among LTR retrotransposon lineages

We divided the sRNA population into two major classes: 20-22nt and 23-25nt sRNA sequences. We refer hereafter to these classes as 21-nt class (20-22nt) and 24-nt class (23-25nt). In this analysis we allowed a 0-2 nt mismatch between the sRNA and LTR-RT reference sequences, in which 25% and 22% of the 21 and 24 ntRNAs class, respectively, showed a perfect match with the LTR-RTs sequences. All identified lineages had at least one family with sRNA from both classes mapped to LTR-RTs. In the *Copia* superfamily, all families within the *Maximus* lineage had more than 8000 sRNAs mapped, mostly belonging to the 24-nt class (Additional file 4: Figure S5 and Additional file 5). These elements had the highest sRNA counts, about 1.5 to 5.5 fold higher than other lineages and these counts mapped chiefly to the LTRs (Additional file 4: Figure S5). *Del1* is the only representative of the *Gypsy* superfamily that has comparable sRNA counts. The *Ivana* and *Reina* lineages were the lineages with the least number of sRNA counts (Additional file 5). Within the *Ivana* lineage, the element *RLC\_scIvana6.1* had the highest number of sRNA matches, most of which belong to the 21-nt class (Additional file 5). The same pattern was observed in *RLG\_Reina3.1*. The remaining *Ivana* and *Reina* families had lower sRNA counts.

To evaluate whether sRNAs preferentially matched specific regions of each LTR-RT, we compared the sRNA distribution within LTRs and coding domains (Additional file 4: Figure S1-S7). Various distribution patterns, both within the LTRs and the coding domains were observed. Families of the *Maximus* lineage and *Del1* had the high 24-nt sRNA counts within the LTRs.

The 5' region of both LTRs from *Del2* and *Del3* families had a peak of 21-nt sRNAs. Most LTR-RT families did not have high sRNA counts within the coding domains, but some cases of high counts of sRNAs were noted. The *Ale1* family had > 1000 counts of the 24-nt class sRNAs within the integrase and protease domains and the *Tat2* family had > 6000 counts of the 24-nt class sRNAs within the *RnaseH* domain (Additional file 4: Figure S4, Figure S3). All other cases of high counts in coding domains are related to 21-nt class sRNAs, as seen in the *Ivana6* and *Reina3* families, both with around 300 counts. Interestingly, *Angela2* had peaks of > 200 24-nt sRNA counts in a 3.5 kb non-coding region between the 5'LTR and the *gag* domain (Additional file 4: Figure S5).

#### Discussion

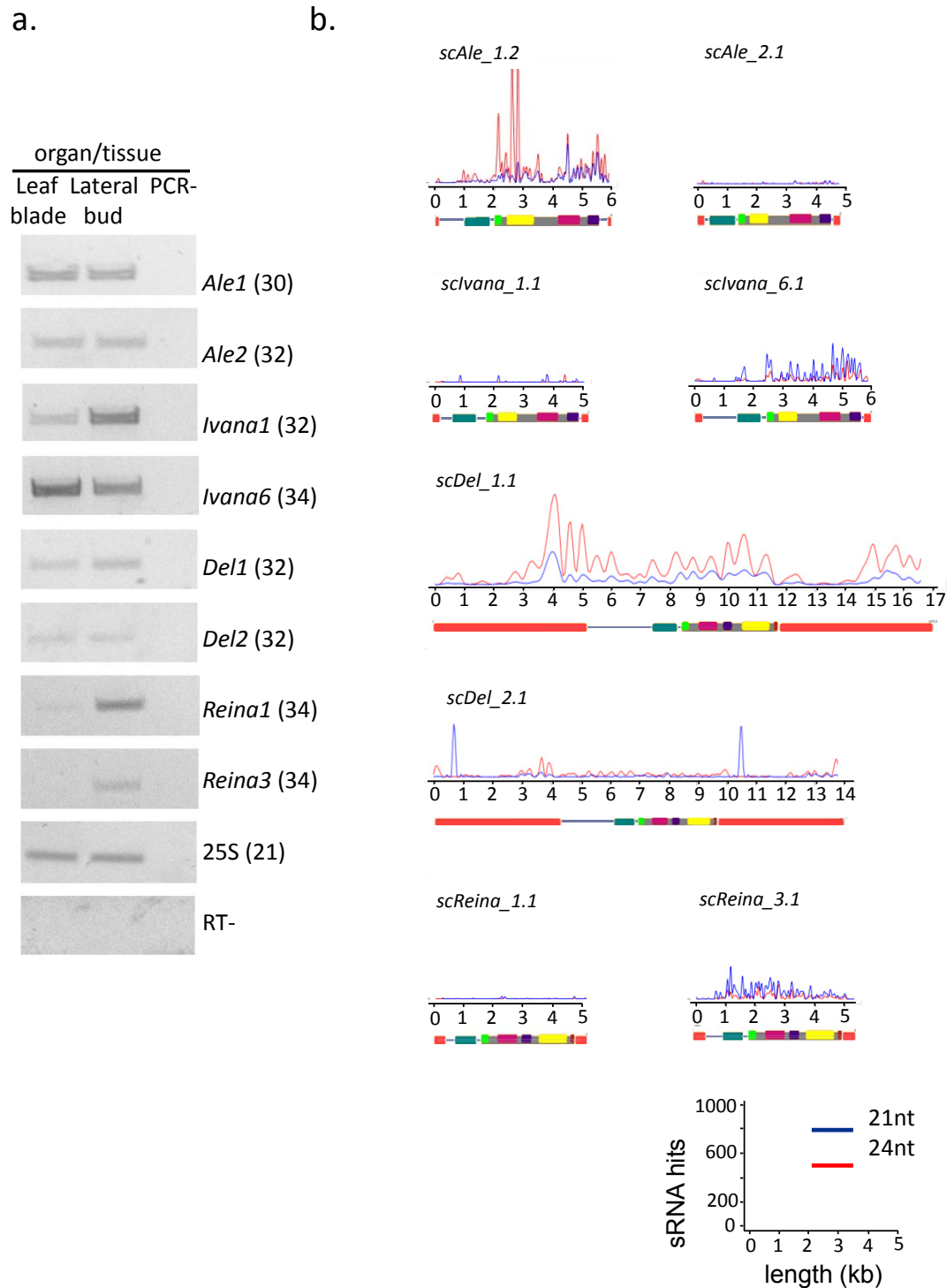
We characterized LTR-RTs in the sugarcane genome using BAC sequences available from the BIOEN project and publicly available genomic and EST resources, which allowed us to provide the most complete overview of the structural and phylogenetic diversity of these elements to date. The 35 LTR-RT families identified belong to four lineages from the *Copia* superfamily and to three from the *Gypsy* superfamily. Among all these families we observed distinct patterns of differences in structural features, chromosomal distribution, transcriptional activity, and sRNAs targeting.

#### Sugarcane LTR-RTs are structurally diverse and belong to evolutionary lineages shared by monocots and eudicots

Previous surveys of LTR-RTs in plants defined six major common evolutionary *Copia* lineages [6] and six major common evolutionary *Gypsy* lineages [8]. *Bianca* was the only *Copia* superfamily lineage in which sugarcane LTR-RTs were not found. Absence of *Bianca* in the sugarcane genome dataset analyzed may be explained by its relatively low representation in other plants. Indeed, rice has few *Bianca* families [6] and soybean has none [7].

We also did not find sugarcane LTR-RTs elements from the *Galadriel* and *CRM* lineages in *Gypsy* superfamily. The *Galadriel* lineage seems to be an under-represented lineage among monocots, since there are only reports of *Galadriel* elements in banana [21]. *CRM* elements are also minor components of *Gypsy* retrotransposons in the model grass rice [7], which may explain its low representation in sugarcane.

Although very little sequence of the sugarcane genome is available, the number of LTR-RT families we identified was comparable to previous large scale LTR-RTs analyses in BAC clones from *Medicago truncatula* (232 Mb) [22] and *Capsicum annuum* (35.6 Mb) [23]. This finding suggests that we have a good overview of the diversity of sugarcane LTR-RTs.



**Figure 5 Mapping of small RNAs in confirmed transcriptionally active LTR-RTs families. (a)** Sugarcane TEs are transcriptionally active in leaf and bud tissues. Expression profiles of selected TEs in leaf blade and developing lateral buds of the sugarcane hybrid SP80-3280. Ribosomal 25S gene (Sc25S) was used as a loading control. RT- (reaction without RT) and PCR- (reaction without cDNA) are shown as negative controls. Numbers in parentheses represent PCR cycles for each amplicon. **(b)** Mapping of sRNA within 8 LTR-RT families. 20-22nt sRNAs are represented as blue lines, 23-25nt sRNAs as red lines. A scaled schematic (also shown in Figures 1 and 2), is shown below each sRNA map.



### Sugarcane LTR-RT families have distinct chromosomal distributions

Transposable elements are not distributed randomly along eukaryotic chromosomes [24]. In particular, *Gypsy* elements tend to be found in heterochromatic regions, *Copia* elements are more dispersed throughout the genome [25,26]. In addition, heterochromatic and pericentromeric regions are enriched in TEs, such as those of the *CR/CRM* and *Tat* lineages, compared with euchromatic regions [27-29]. Using fluorescence in situ hybridization (FISH), we examined the distribution of representatives from seven lineages identified within the *Gypsy* and *Copia* superfamilies in sugarcane chromosomes. No detectable FISH signal was obtained for the *Angela*, *Ivana* and *Reina* lineages; in the case of *Ivana* preliminary estimates from the R570 BAC library suggests that there is approximately 50 copies of *Ivana1* per haploid genome in sugarcane (data not shown), which suggests that the lack of signal is due to low copy number. We suspect that this observation may be also valid for *Angela* and *Reina*.

The two *Copia* probes, *Ale1* and *Maximus1*, had the expected euchromatic patterns, but they were distinctly different. The signals from the *Ale* probe had localized clusters along particular chromosome arms, while those of the *Maximus1* probe were widespread and dispersed, without any signs of concentration in particularly regions (Figure 3d). It has been suggested that as TEs accumulate in a genome, even if they insert randomly, they are more likely to insert within other TEs, forming clusters of TEs. Because these clusters of TEs are most likely selectively neutral, they will be free to expand [24]. The *Ale* distribution pattern observed in sugarcane is reminiscent of such clusters of TEs.

The sugarcane BAC SCHRBA 239\_N21 used as a centromeric-specific probe in this study contains the SCEN repeat interspersed with LTR-RTs from the *Tat* lineage [16]. Previous studies in sugarcane [18,19] identified a centromere-specific repeat (SCEN) and centromere associated LTR-RTs, classified as *CR/CRM*-like. We did not identify any full length LTR-RTs from the *CR/CRM* lineage in this study, but sequence from a related *Gypsy* lineage, *Del1*, did hybridize to a broad region around and within the centromeric region (Figure 3b and 3e). *Del* lineage sequences have been identified at other plant centromeres [28], however our results suggest that the *Del* sequence we used as a probe is not strictly centromeric specific, but is rather preferentially found in and around heterochromatic regions of the centromeric repeats. The *Tat* sequences identified in the BAC SCHRBA 239\_N21 clustered with the *Tat1*, *Tat4*, *Tat5* and *Tat6* families, (data not shown) which is quite distinct from the *Tat2* family (Figure 1). The *Tat* FISH probe used is most closely related to the *Tat2* sequences

(Figure 1) and had a clusters of localization along some chromosomes, like those seen with the *Ale1* probe, but with much more intense signals.

Modern sugarcane cultivars are interspecific hybrids between *Saccharum officinarum* and *S. spontaneum* and backcrossed with *S. officinarum* [30]. The resulting genomes are highly polyploid, 70-80% of which is from *S. officinarum*, 10-23% from *S. spontaneum*, the rest being recombinants [10,31]. The derivation of the modern sugarcane genome suggests that, rather than being the result of TEs inserting into clusters of TEs, the distribution of *Tat2* and *Ale1* may be the result of differential transmission of TEs from the parental genomes. Future work using genomic in situ hybridization [31] combined with FISH should enable us to distinguish between these two hypotheses.

### Transcriptional activity of LTR-RTs and their associated sRNAs

Transposable elements make up a substantial proportion of plant genomes, and are often transcriptionally active. A very stringent search for sugarcane ESTs associated with TEs confirmed that transcriptional activity appears to be a general feature of some sugarcane LTR-RTs, as reported for other monocots [32,33]. Since we do not have detailed information about cDNA library preparation of EST data, we cannot identify if transcribed sequences were in sense or antisense orientation. Interestingly, within each lineage, the number of transcripts mapped to each family was variable (Figure 4a). The most striking example is the *Ale* lineage, which has the highest number of transcripts, all of which mapped to a single family, *Ale1*. There is a similar, but not as definitive pattern in all other lineages, except for *Maximus*, where all families have similar numbers of transcripts.

In terms of tissue specificity (Figure 4b), the highest number of LTR-RT associated transcripts was identified from the root libraries, followed by the internode, lateral bud and calli libraries. A previous macroarray analysis identified calli as the tissue with the highest number of TEs being expressed in sugarcane [20], however in this study only calli, apical meristem, leaf roll and flower were analyzed. These differences in expression according to tissue, in particular that leaf is the tissue with least number of ESTs identified, may help guide future studies of transposable elements in sugarcane.

As they are potentially highly mutagenic, the activity of LTR-RTs is usually controlled by the host genome through the siRNA machinery. The specificity of this response is achieved by a surveillance system that detects aberrant RNA. The proliferative nature of TEs makes them prone to insert in the genome in such way that both sense and anti-sense transcripts are produced, generating dsRNA, and activating the siRNA system.

Two main classes of siRNAs are generated, the 21-nt class regulates post-transcriptionally related mRNAs while the 24-nt class is involved in RNA-dependent DNA methylation (RdDM) and heterochromatin maintenance and therefore suppresses gene expression at transcriptional level [34].

Previous studies mapping sRNAs to LTR-RTs in wheat and maize genomes [35,36], presented an overall study of TE superfamilies and described a pattern of concentration of 24nt sRNA in the LTRs. Our work, which focused on mapping to individual LTR-RT elements, identified distinct patterns of sRNA targeting within sugarcane LTR-RT lineages and families.

The previously described '24nt LTR' pattern was observed for all reference sequences in the *Maximus* families, and for *Del1* and *Tat3* (Additional file 5). For all other reference sequences a different type of pattern was observed. For almost half of the families (18 out of 33) very few sRNAs (< 2000 counts) were mapped to the reference copy (Additional file 5). Two other patterns were observed, one in which high numbers of 21nt sRNAs mapped along the coding region, represented by *Ivana6* and *Reina3* (Additional file 5), and one in which a very large number of 24nt sRNAs mapped within the coding region, seen only in *Ale1*.

Very few mapped sRNAs indicates that elements from those families either are not transcriptionally active, or they are very recently activate and have not yet triggered the host small RNA-dependent silencing machinery. If they are not transcriptionally active, silencing may be being maintained by ancient methylation. A previous study has shown that ~63% of methylated regions were maintained without persistent targeting by sRNAs [37], explaining the absence of both 21 and 24nt sRNAs.

The high numbers of 21nt sRNAs mapped to the coding region of some families suggests that they are being regulated primarily post-transcriptionally. This has been previously reported for MITEs [36] and may indicate recent activation of transcription, prior to triggering of the RdDM machinery.

The unique pattern observed in the *Ale1* family suggests that the RdDM machinery is guiding methylation to the coding region of the element, not the promoter region. The "body-methylated gene" concept was first described in plants in 2006, when Zhang and collaborators showed that over one third of *Arabidopsis* expressed genes were methylated in the coding region, but not in the promoter region [37]. The *Ale1* family has a distinct profile, it has a 'body-gene'-like sRNA pattern, it is the most transcriptionally active LTR-RT in sugarcane and is concentrated in euchromatic regions and gene-rich BAC sequences ([12] and Figure 3c). At this point it is difficult to make conclusions from these

observations, however these results are intriguing and warrant further investigation.

## Conclusions

This is the first study to perform a concomitant survey of phylogenetic diversity, chromosomal distribution, structure, transcriptional activity and interaction with sRNAs of LTR-RTs in a plant genome. We assigned 60 LTR-RT elements to 35 families within four *Copia* and three *Gypsy* lineages. Two lineages, one *Copia* and one *Gypsy* lineage, showed distinct patterns of signal clustering along some chromosomes in the FISH analysis. Given that the modern sugarcane cultivars are hybrids with highly unequal contributions from the ancestral genomes, the FISH patterns suggest that for these lineages there has been higher numbers of elements from one ancestral genome than the other. For the transcriptional and sRNA mapping analyses we chose to analyze at the family level. Individual families had distinct transcript and sRNA mapping profiles, suggesting that they are differentially expressed and regulated. The *Ale1* family was particularly unusual in that it had 'body-gene'-like sRNA pattern, it is the most transcriptionally active LTR-RT in sugarcane and is concentrated in euchromatic regions. Overall, our results indicate that LTR-RTs could impact the genome in different ways at the family levels.

## Methods

### Identification and retrieval of sugarcane LTR retrotransposon sequences

All BACs used are from the R570 sugarcane cultivar library [38]. BACs sequenced for the BIOEN Project [16] and public sugarcane BAC sequences available at the National Center for Biotechnology Information (NCBI) website as at 01/02/2011 were screened for full-length LTR elements using LTR\_STRUC [39] with the most thorough stringency (1). Sixty sequences were retrieved and provisionally assigned to the *Gypsy* or *Copia* superfamily by submission against cores in the Gypsy Database (GyDB) [8] using BLASTX. To determine whether the sequences were complete elements, we identified target site duplications (TSDs) by submitting the full length sequences as a query and subject to a blast2seq [40] on the NCBI website.

### Phylogenetic analysis

Sugarcane LTR-RTs, including the probes used for fluorescence *in situ* hybridization, were assigned to previously described plant LTR lineages [6-8] by phylogenetic analysis using the translated reverse transcriptase (RT) domain excised from all the sugarcane LTR-RTs and published RT sequences.

For both phylogenies we downloaded RT alignments from the Gypsy database (GyDB) [8], and removed non-plant sequences. *Gypsy* sequences were also taken from Du et al. (2010) [7], *Copia* sequences were taken from Wicker and Keller (2007) [6]. All sequences were renamed to reflect published lineage names. Sequences were aligned using MUSCLE with default settings [41] and manually adjusted by eye. The optimal model of amino acid substitution was estimated using MEGA5 [42] with default settings. Neighbor-joining and maximum-likelihood phylogenies were estimated with MEGA5 [42] using the highest-ranked substitution model available and a bootstrap of 500 replicates.

#### Assignment to Families within Lineages and naming of sequences

Sugarcane LTR-RTs were assigned to families within lineages on the basis of 80% sequence identity in at least 80% of their LTRs [5]. Although previous reports assign names to some sugarcane LTR-RT families [11,14,20], we opted to standardize the name of sugarcane LTR-RT sequences, using a more straightforward strategy, based in the proposed universal classification of TEs by Wicker et al. (2007) [5]. Sequences were named 'RLC' (*Copia*) or 'RLG' (*Gypsy*), 'sc' for 'sugarcane', the lineage name e.g. 'Ale', the family number e.g. '1', then each sequence within a family was numbered sequentially. For example 'RLC\_scAle\_1.1' is the first sequence named within the *Ale* lineage, family 1, superfamily *Copia*.

#### Analysis of the structure of Sugarcane LTR-RTs

Coding domains were identified using Pfam, or by alignment with MUSCLE [41] against the domain alignment from the GyDb [8]. Full-length sequences were aligned and analyzed using BioEdit [43], using the toggle translate option so that we could align the coding domains as well as the LTRs, TSDs, and the regions between the LTRs and the coding domains. LTRs were identified by submitting the sequence of the entire sugarcane LTR-RT as both a query and subject to a MEGABLAST [40] analysis. The beginning of the LTRs, regions between the LTRs and the coding domains, and the TSDs were manually aligned in BioEdit [43]. Co-ordinates of the beginning of all features of each element were recorded in an Excel table and the information submitted to domain draw [17] to create a schematic representation of each sugarcane LTR-RT.

General features of each sequence, as element size, LTR size, Target Side Duplications (TSD) and GenBank accession numbers are presented in Additional file 6.

#### Sugarcane EST database screening

All full-length LTR-RTs were used as queries in a BLASTN search against EST sequences from the

sugarcane cultivar SP80-3280. The ESTs were obtained using ENTREZ at NCBI <http://www.ncbi.nlm.nih.gov/Entrez/>. A total of 155,354 sugarcane ESTs were analyzed, all of them from the SUCEST (Sugarcane EST) project [44].

ESTs similar to LTR-RTs were assigned to a family according to the criteria based on Wicker et al. [5]: 80% coverage with 80% nucleotide identity.

The number of hits for each library was normalized by dividing the raw number of hits by the total number of valid reads. The normalized numbers of hits per library were then combined according to tissue type. The final number was multiplied by 100,000, so that in Figure 4 the X axis represents the number of ESTs per 100,000 transcripts from each tissue.

#### Association of cDNAs to full-length LTR-RTs

Thirty manually curated sugarcane cDNAs related to LTR-RTs [20], described using an older nomenclature, were assigned to a family according to the same criteria used for the ESTs.

#### RNA extraction and Reverse Transcriptase (RT) PCR Analysis

Leaf blade tissues were collected from one-month-old sugarcane plants (cultivar SP 80-3280) grown under greenhouse conditions. Mature eight-month-old plants of the same cultivar were used to obtain lateral buds. Stalk pieces with one bud (single eye sets) were planted in plastic trays containing a commercial planting mix (Plantmax, Eucatex, Brazil). After five days, developing buds were collected for RNA extraction. Two independent biological replicates were collected for leaf blade and lateral bud tissues. Total RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions.

Primers were designed within the reverse transcriptase domain using Primer3Plus [45] to amplify all known elements from a family. Total RNA was treated with DNase I Amp Grade (Invitrogen) to remove any residual genomic DNA. Three micrograms of DNase-treated RNA was used to generate the first strand cDNA using ImProm II Reverse Transcriptase (Promega) according to the manufacturer's instructions. The reaction mixture was placed in a GeneAmp9700 thermocycler (Applied Biosystems) and incubated at 16°C for 30 minutes, followed by 60 cycles of pulsed reverse transcription at 30°C for 30 seconds, 42°C for 30 seconds, and 50°C for one second. cDNA dilutions were used in PCR reactions as following: 1.0 µL of cDNA, 10 pmol of each primer, GoTaqmastermix, and 1 U of GoTaq DNA Polymerase (Promega) in a total volume of 25 µL. The reactions were placed in the thermocycler with the following conditions: 94°C for three minutes and

appropriate cycle numbers of 94°C for 30 seconds, 55°C or 60°C for 30 seconds, and 72°C for 45 seconds. All reactions were repeated at least twice.

#### Small RNA library construction and bioinformatic analysis

To evaluate the small RNA “landscape” of sugarcane LTR-RTs, we prepared a sRNA library from leaves of one-month old SP80-3280 sugarcane cultivar plants, grown under greenhouse conditions. Ten micrograms of total RNA, prepared using TRIzol reagent (Invitrogen) according to the manufacturer’s instructions, were used to generate sRNA library following Illumina’s modified protocol. The sRNA fraction of 19-28 nt was purified by size fractionation on a 15% TBE-Urea polyacrylamide gel. A 5’-adenylated single-stranded adapter was first ligated to the 3’-end of the RNA using T4 RNA ligase without ATP, followed by a second single-stranded adapter ligation at the 5’-end of the RNA using T4 RNA ligase in the presence of ATP. The resulting products were fractionated on a 10% TBE-Urea polyacrylamide gel and then used for cDNA synthesis and PCR amplification. The resulting library was sequenced on an Illumina Genome Analyzer (GA-IIx) following the manufacturer’s protocol available at <http://www.fasteris.com>.

A total of 4,388,665 20-25nt raw sequences were retrieved in a FASTQ formatted file and the adapter sequences were removed using Perl Scripts. After trimming of adapter sequences, the inserts were sorted into separate files according to their lengths. We used the program MAQ [46] to map 20-25 ntsRNA reads against sugarcane LTR-RT reference sequences (sequence 1 from each family). MAQ is a program that rapidly aligns short reads to reference genome sequences, and in this study we allowed 0-2 nt mismatches between the sRNA and LTR-RTs sequences. Three percent of the total library, that is, 131,641 high quality raw 20-25nt sequences matched against the sugarcane LTR-RT sequences. These sRNAs sequences have been submitted to the NCBI Gene Expression Omnibus database <http://www.ncbi.nlm.nih.gov/geo> under accession number GSE35143.

#### Fluorescence in situ hybridization (FISH)

The distribution of the sugarcane LTR-RTs was analyzed by fluorescence *in situ* hybridization (FISH) on metaphase chromosomes. In order to compare the distribution of the LTR-RT relative to the centromere, a centromeric BAC [16] was also used as a probe. A single representative probe was used for each evolutionary lineage (Figure 3). The sequence of each probe was submitted as a query to a BLASTN analysis against a database of sugarcane cDNAs related to TEs identified in our lab [11,20] to check that, at 85% stringency, it would not hybridize against other elements.

All LTR-RT probes were 1.9 to 2.9 kb and covered the reverse transcriptase domain. For the *Ale1* and *Ivana1* families, probes were selected from previously reported cDNA sequences [11,20]. For *Ale1*, we used cDNA TE137 (GenBank accession [GenBank:JN786875]) and TE049 for *scIvana1* (GenBank accession [GenBank:DQ115032]) on the basis of size (> 1.9 kb) and the presence of the reverse transcriptase domain. For all other lineages primers were designed from alignments of the RT domain using Primer3Plus [45]. All kits were used according to the manufacturer’s instructions. The probe sequences were PCR amplified from R570 cultivar genomic DNA using Elongase (Invitrogen) or GoTaq (Promega) with 2 mM MgCl<sub>2</sub>, 0.2 mM dNTPs, 0.2 μM primers, 1 ng/μL genomic DNA and 0.025 units/uL of Enzyme. Cycling conditions were as described in the Expand Long Template PCR System (Roche). The resulting amplicons were separated on 1% agarose, gel-purified using the NucleoSpin Extract II kit (Macherey Nagel), ligated into the pGEM T-Easy Vector (Promega), and cloned into DH10B electrocompetent cells according to standard procedures [47]. Minipreps from three clones from each PCR reaction and from the cDNA clones were prepared using standard alkaline precipitation methods [47], and sequenced using the vector primers M13F/R. In order to obtain a probe that consisted of just the probe, one miniprep for each lineage was diluted 1:1000 and used as template in 100 μL PCR reaction with M13F/R primers to amplify the insert only, using GoTaq (Promega) in same reaction conditions as above, but with the following cycling conditions, initial denaturation 95°C 3 min, 35 cycles of 95°C 1 min, 55°C 30 sec, 72°C 2 min, followed by a final extension of 72°C for 3 min. The resulting amplicons were separated on 1% agarose, gel purified using the NucleoSpin Extract II kit (Macherey Nagel) and quantified using a NanoDrop Spectrometer (ThermoScientific). For the centromeric BAC probe, BAC DNA was extracted using the Large-Construct Kit (Qiagen).

Between 350-700 ng of probe DNA was used in a 20 μL nick translation reaction with Digoxigenin (DIG)-11-dUTP (Invitrogen) or Biotin-16-dUTP (Invitrogen) and the NT mix (Roche). Labeling efficiency was tested according to Heslop (2000) [48] (protocol 4.7). The probe was only used if the 1:1000 dilution was clearly visible.

Sections of sugarcane stalk from the cultivar SP80-3280 were planted in a mixture of 1/2 soil 1/2 vermiculite, root tips harvested within 1-3 days and placed directly into 2 mM 8-hydroxyquinoline for 6 hours at 18°C. Next, they were transferred to 3:1 ethanol:acetic acid fixative and stored at -20°C. Root samples were prepared according to Heslop(2000) [48], protocol 5.3, except that they were digested in either 2% cellulase/

0.2% macerozyme/20% pectinase or 1% cellulase/0.2% macerozyme for 2 1/2 to 3 hours (depending on how large the root tip was) at 37°C.

Hybridization and detection was performed according to Heslop (2000) [48] using protocols 8.1, 8.4, 9.1 and 9.2, with the following conditions: the slide was dried for 30 min at 50-60°C and pretreated with both RNase A and pepsin (20 min at 37°C); 1 µL of each labeled probe was added to a 20 µL hybridization mix of 50% formamide/2xSSC/10% dextran sulphate/1% SDS; the slide was denatured in 50 mL of 70% formamide/2xSSC at 70°C for 2 min and then dehydrated through an ice-cold ethanol series (70%, 85%, 100% ethanol); washes were 80-82% stringent, 20% formamide with 0.1 or 0.2 xSSC at 42°C; DIG-labeled probes were detected with anti-digoxigenin-rhodamine (Roche), biotin-labeled with NeutrAvidin-Oregon Green-488 (Molecular Probes).

The slide was stained with DAPI, observed and photographed with an Zeiss AxioPlan2 microscope and captured using an AxioCam MR camera and the Isis Fluorescence Imaging System (MetaSystems). Nine to 25 metaphases were photographed for each probe. Slides were stripped by carefully removing the immersion oil, soaked in 4xSSC/0.1% Tween 20 at 37°C until the coverslip floated off, transferred to fresh 4xSSC/0.1% Tween 20 for 3 hours with gentle shaking, transferred to 3:1 ethanol:acetic acid fixative for 30 min and then dehydrated through an ethanol series (70%, 85%, 100% ethanol) for 5 min each at room temperature and air dried for 1 hour.

## Additional material

**Additional file 1: Gypsy and Copia Maximum Likelihood phylogenies.** Maximum-likelihood phylogeny of *Gypsy* families (Figure 1) based on reverse transcriptase. Sequences from the Gypsy database [8] are denoted with a plus symbol, those from Du et al. [7] with a star. Maximum-likelihood phylogeny of *Copia* families (Figure 2) based on reverse transcriptase. Sequences from the Gypsy database [8] are denoted with a plus symbol, those from Wicker and Keller [6] with a star. Robustness of the nodes was estimated by 500 bootstrap replications. Bootstrap values below 50 are not shown.

**Additional file 2: Fluorescence in situ hybridization with the DEL probe.** Figure of fluorescence *in situ* hybridization using the *Del* probe, prepared and labelled twice, once with Digoxigenin (DIG) and detected with Anti-DIG-Rhodamine (red signal) and once with Biotin and detected with NeutrAvidin-Oregon Green (green signal). The probes were hybridized to the same slide in consecutive FISH experiments under the same conditions. The same pattern was observed for both probes, suggesting that the signal was real, and the same FISH conditions was used for all LTR-RT probe.

**Additional file 3: Assignment of previously described cDNA sequences to LTR-RT families.** Thirty manually curated sugarcane cDNAs related to LTR-RTs [20] were assigned to a family according Wicker et al. [5]: 80% coverage with 80% nucleotide identity.

**Additional file 4: sRNA mapping to individual LTR-RT elements.** Mapping of sRNAs within each LTR-RT family (Figures 1 to 7). 20-22nt sRNAs are represented as blue lines, 23-25ntsRNAs as red lines. Each

figure shows a different lineage, and includes all the families of that lineage. A scaled schematic (also shown in Figures 1 and 2), is shown below each sRNA map.

**Additional file 5: Total 20-22nt and 23-25nt sRNA counts for each LTR-RT family.** Total 20-22nt (black) and 23-25nt (grey) sRNA counts for each LTR-RT family, with a mismatch of 2nt allowed.

**Additional file 6: LTR-RT information.** Name assigned in this paper, pre-existing name from [14,20], GenBank accession number, size of full-length element, length of 5' and 3' LTRs, and sequence of 5 and 3' TSDs, for individual LTR-RT sequences.

## Acknowledgements

The authors thank Dr Marguerite Rodier-Goud (CIRAD, Montpellier, France), Dr Angélique D'Hont (CIRAD, Montpellier, France), Dr Reginaldo Carvalho (Universidade Federal Rural de Pernambuco, Recife, Brazil) and Dr. Ana Christina Brasileiro Vidal (Universidade Federal de Pernambuco, Recife, Brazil) for help with the preparation of metaphases from sugarcane and fluorescence *in situ* hybridization. We also thank GaTE Lab members Dr Nathalia de Setta, Andréia Prata, Edgar Andres Ochoa-Cruz and Kleber Alves Gomes for their help with sugarcane BAC sequencing under the BIOEN project. This work was partly funded by FAPESP-BIOEN 2008/52074-0 and CNPq grants to MAVS. DSD, GMQC and CJM are supported by fellowships from FAPESP. DSD was also supported by a fellowship from CAPES.

## Author details

<sup>1</sup>GaTE Lab, Depto. de Botânica, Inst. de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-090 São Paulo, Brazil. <sup>2</sup>Depto. de Genética, Inst. de Biociências, Universidade Estadual Paulista, campus de Botucatu, Distrito de Rubião Jr, s/n, 18618-000 Botucatu, Brazil. <sup>3</sup>Systems Biology Laboratory, Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, 13083-875 Campinas, Brazil. <sup>4</sup>Plant Biotechnology Laboratory, Instituto Agrônomo do Paraná, Rod. Celso Garcia Cid (PR-445), km375, 86047-902 Londrina, Brazil.

## Authors' contributions

The project was designed by MAVS, DSD and GMQC, and co-ordinated by MAVS. DSD, GMQC and CJM did the bio-informatic analyses. CJM carried out the FISH assays. FTSN and RV made the sRNA library and did the sRNA analyses. CSA did the RT-PCRs. DSD, GMQC, CJM, FTSN, RV and MAVS wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests

Received: 27 October 2011 Accepted: 16 April 2012

Published: 16 April 2012

## References

1. Ainouche ML, Jenczewski E: **Focus on polyploidy.** *New Phytol* 2010, **186**:1-4.
2. Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhou B, Grandbastien M: **Impact of transposable elements on the organization and function of allopolyploid genomes.** *New phytol* 2010, **186**:37-45.
3. Hirochika H, Okamoto H, Kakutani T: **Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation.** *Plant cell* 2000, **12**:357-368.
4. Feschotte C, Jiang N, Wessler SR: **Plant transposable elements: where genetics meets genomics.** *Nature reviews Genetics* 2002, **3**:329-341.
5. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhou B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH: **A unified classification system for eukaryotic transposable elements.** *Nat Rev Genet* 2007, **8**:973-982.
6. Wicker T, Keller B: **Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families.** *Genome Res* 2007, **17**:1072-1081.
7. Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC, Ma J: **Evolutionary conservation, diversity and specificity of LTR-**

- retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J* 2010, **63**:584-598.
8. Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Muñoz-Pomer A, Sempere JM, Latorre A, Moya A: **The Gypsy Database (GyDB) of mobile genetic elements: release 2.0.** *Nucleic Acids Res* 2011, **39**:D70-D74.
  9. Souza GM, Berges H, Bocs S, Casu R, D'Hont A, Ferreira JE, Henry R, Ming R, Potier B, Van Sluys M-A, Vincenz M, Paterson AH: **The Sugarcane Genome Challenge: Strategies for Sequencing a Highly Complex Genome.** *Tropical Plant Biology* 2011, **4**:145-156.
  10. Piperidis G, Piperidis N, D'Hont A: **Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane.** *Molecular Genet Genomics* 2010, **284**:65-73.
  11. Rossi M, Araujo PG, Van Sluys M-A: **Survey of transposable elements in sugarcane expressed sequence tags (ESTs).** *Genet Mol Bio* 2001, **24**:147-154.
  12. Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A: **Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome.** *Plant J* 2007, **50**:574-585.
  13. Wang J, Roe B, Macmill S, Yu Q, Murray JE, Tang H, Chen C, Najjar F, Wiley G, Bowers J, Van Sluys M-A, Rokhsar DS, Hudson ME, Moose SP, Paterson AH, Ming R: **Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes.** *BMC Genomics* 2010, **11**:261.
  14. Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, Droc G, Zini C, Glaszmann J-C, Van Sluys M-A, D'Hont A: **High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane.** *New Phytol* 2010, **189**:629-642.
  15. Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W: **Consistent over-estimation of gene number in complex plant genomes.** *Current Opin Plant Biol* 2004, **7**:732-736.
  16. Setta N, Cruz G, Cruz E, Gomes K, Campos R, Hotta C, Vilela M, Vincenz M, Vautrin S, Souza G, Bérge H, Gaiarsa J, Kitajima J, Van Sluys M: **Sugarcane genome: a snapshot from 100 sequenced BACs.** *Plant and Animal Genomes XIX Conference San Diego* 2011, USA.
  17. Fink J, Hamilton N: **DomainDraw: A Macromolecular Schematic Drawing Program.** *In Silico Biol* 2007, **7**:14.
  18. Nagaki K, Tsujimoto H, Sasakuma T: **A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions.** *Chromosome Res* 1998, **6**:295-302.
  19. Nagaki K, Murata M: **Characterization of CENH3 and centromere-associated DNA sequences in sugarcane.** *Chromosome research* 2005, **13**:195-203.
  20. de Araujo PG, Rossi M, de Jesus EM, Saccaro NL, Kajihara D, Massa R, de Felix JM, Drummond RD, Falco MC, Chabregas SM, Ulian EC, Menossi M, Van Sluys M-A: **Transcriptionally active transposable elements in recent hybrid sugarcane.** *The Plant Journal* 2005, **44**:707-717.
  21. Hribová E, Neumann P, Matsumoto T, Roux N, Macas J, Dolezel J: **Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing.** *BMC Plant Biol* 2010, **10**:204.
  22. Wang H, Liu J-S: **LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice.** *BMC Genomics* 2008, **9**:382.
  23. Park M, Jo S, Kwon J-K, Park J, Ahn JH, Kim S, Lee Y-H, Yang T-J, Hur C-G, Kang B-C, Kim B-D, Choi D: **Comparative analysis of pepper and tomato reveals euchromatin expansion of pepper genome caused by differential accumulation of Ty3/Gypsy-like elements.** *BMC Genomics* 2011, **12**:85.
  24. Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P: **The struggle for life of the genome's selfish architects.** *Biology Direct* 2011, **6**:19.
  25. Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin A, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell A, Harrison GE: **The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution.** *Genetica* 1997, **100**:197-204.
  26. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Ottillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, et al: **The *Sorghum bicolor* genome and the diversification of grasses.** *Nature* 2009, **457**:551-556.
  27. Mizuno H, Ito K, Wu J, Tanaka T, Kanamori H, Katayose Y, Sasaki T, Matsumoto T: **Identification and mapping of expressed genes, simple sequence repeats and transposable elements in centromeric regions of rice chromosomes.** *DNA Res* 2006, **13**:267-274.
  28. Theuri J, Phelps-Durr T, Mathews S, Birchler J: **A comparative study of retrotransposons in the centromeric regions of A and B chromosomes of maize.** *Cytogenetic Genome Res* 2005, **110**:203-208.
  29. Weber B, Schmidt T: **Nested Ty3-gypsy retrotransposons of a single *Beta procumbens* centromere contain a putative chromodomain.** *Chromosome Res* 2009, **17**:379-396.
  30. Grivet L, Arruda P: **Sugarcane genomics: depicting the complex genome of an important tropical crop.** *Current Opin Plant Biol* 2001, **5**:122-127.
  31. D'Hont A: **Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana.** *Cytogenetic Genome Res* 2005, **109**:27-33.
  32. Vicent CM: **Transcriptional activity of transposable elements in maize.** *BMC Genomics* 2010, **11**:601.
  33. Echenique V, Stamova B, Wolters P, Lazo G, Carollo L, Dubcovsky J: **Frequencies of Ty1-copia and Ty3-gypsy retroelements within the Triticaceae EST databases.** *Theoretical Appl Genet* 2002, **104**:840-844.
  34. Baulcombe D: **RNA silencing in plants.** *Nature* 2004, **431**:356-363.
  35. Nobuta K, Lu C, Shrivastava R, Pillay M, De Paoli E, Accerbi M, Arteaga-Vazquez M, Sidorenko L, Jeong DH, Yen Y, others: **Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant.** *Proc Natl Acad Sci USA* 2008, **105**:14958-14963.
  36. Cantu D, Vanzetti LS, Sumner A, Dubcovsky M, Matvienko M, Distelfeld A, Michelmore RW, Dubcovsky J: **Small RNAs, DNA methylation and transposable elements in wheat.** *BMC Genomics* 2010, **11**:408.
  37. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*.** *Cell* 2006, **126**:1189-1201.
  38. Tomkins J, Yu Y, Miller-Smith H, Frisch D, Woo S, Wing R: **A bacterial artificial chromosome library for sugarcane.** *Theoretical Appl Genet* 1999, **99**:419-424.
  39. McCarthy EM, McDonald JF: **LTR\_STRUC: a novel search and identification program for LTR retrotransposons.** *Bioinformatics* 2003, **19**:362-367.
  40. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmsberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrahi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, et al: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2011, **39**:D38-D51.
  41. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.
  42. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Mol Biol Evol* 2011, **28**:2731-2739.
  43. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95-98.
  44. Vettore AL, Silva FRD, Kemper EL, Arruda P: **The libraries that made SUCEST.** *Genet Mol Biol* 2001, **24**:1-7.
  45. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA: **Primer3Plus, an enhanced web interface to Primer3.** *Nucleic Acids Res* 2007, **35**:W71-W74.
  46. **Mapping and Assembly with Qualities.** [http://maq.sourceforge.net].
  47. Sambrook J, Fritsch EF, Maniatis T: *In Molecular Cloning: A Laboratory Manual* Edited by: Irwin N, Ford N, Nolan C, Ferguson M Cold Spring Harbor Laboratory Press, 2 1989.
  48. Heslop-Harrison P, Schwarzacher T: *Practical In situ Hybridization* Oxford, UK: BIOS Scientific Publishers Ltd; 2000.

doi:10.1186/1471-2164-13-137

**Cite this article as:** Domingues et al.: Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics* 2012 **13**:137.

*Apêndice 6* – Capítulo de livro publicado como colaboradora

# RNA Activation e Outros Efeitos Mediados por dsRNAs

---

---

## Capítulo 13

---

---

Prof. Dr. Fabio TS Nogueira<sup>1</sup>, Ms. Cristiane S Alves<sup>1</sup>, Prof. Dr. Paulo C Ferreira<sup>2</sup>

<sup>1</sup>Depto. de Genética, Instituto de Biociências de Botucatu, UNESP

<sup>2</sup>Instituto de Bioquímica Médica, UFRJ

### Um Surpreendente Papel dos Pequenos RNAs de Fita Dupla na Ativação Gênica

Como foi visto nos capítulos anteriores, RNAi é um mecanismo de silenciamento gênico pós-transcricional, iniciado pela produção de pequenos duplexes de RNAs (siRNAs) de 21-28 nucleotídeos (nt) a partir de longas moléculas de RNA de dupla fita (dsRNAs). Tais pequenos RNAs são incorporados no complexo proteico RISC (*RNA-induced silencing complex*) e guiam a clivagem de moléculas de RNA complementares ou RNAs alvos (Hannon, 2002). Além da regulação pós-transcricional, os RNAs de interferência estão também envolvidos em processo de silenciamento transcricional via metilação do DNA nuclear em plantas (Hamilton et al., 2002), *Schizosaccharomyces pombe* (Volpe et al., 2002), *Drosophila melanogaster* (Pal-Bhadra et al., 2004) e em mamíferos (Morris et al., 2004). Ademais, em protozoários ciliados, como o *Tetrahymena thermophila*, dsRNAs são capazes de atuar como guias na deleção de sequências genômicas específicas, que podem variar de centenas até 20.000 pares de bases, durante a conjugação destes organismos. O mecanismo como estes dsRNAs atuam permanece desconhecido. Estudos com estes organismos demonstraram que dsRNAs também são capazes de deletar DNA exógeno, funcionando como um “serviço de vigilância” celular (Yao et al., 2003).

Existem, entretanto, alguns casos em que RNAs pequenos atuam positivamente na regulação da expressão de genes alvos. Há pouco mais de quatro décadas, R.J. Britten e E.H. Davidson (1969) propuseram a teoria de que os então denominados “RNAs ativadores” - transcritos de regiões redundantes e não codificadoras do genoma - seriam capazes de ativar a expressão de vários genes codificadores de proteínas. Segundo os autores, os “RNAs ativadores” seriam complementares a regiões regulatórias de alguns genes, de modo que RNAs de fita única formariam



um complexo com a fita dupla de DNA. Esse complexo seria responsável pela regulação transcricional de genes alvos. Embora na época tal teoria não tenha sido formalmente testada, recentemente, alguns trabalhos apontam para a intrigante possibilidade de dsRNAs estarem envolvidos em mecanismos de ativação da expressão gênica. Tal mecanismo é denominado *double-strand RNA (dsRNA)-induced gene activation, RNA activation* ou simplesmente RNAa (Rossi, 2007). Esse mecanismo foi demonstrado pelo uso de RNAs sintéticos de 19-21 nt que formam duplexes e são complementares a regiões específicas de promotores de genes humanos. Tais duplexes sintéticos são coletivamente denominados *small activating RNAs* (saRNAs) ou *antigene RNAs* (agRNAs), para diferenciar de dsRNAs pequenos que atuam no mecanismo de RNAi (siRNAs). Para padronização de nomenclatura, esse capítulo tratará esses duplexes de RNA coletivamente como saRNAs. O reconhecimento de sequências de promotores pelos saRNAs é capaz de aumentar os níveis de expressão de genes específicos em até vinte vezes em células de mamíferos (Li et al., 2006; Janowski et al., 2007; Schwartz et al., 2008).

A descoberta de que pequenos duplexes, similares a siRNAs, são capazes de ativar a expressão gênica indica que devem existir diferenças nos mecanismos de silenciamento e ativação gênica. Ambos os grupos de pesquisa, que identificaram o fenômeno de RNAa, estavam inicialmente estudando mecanismos de silenciamento transcricional mediado por siRNAs. Li et al. (2006) delinearam experimentos nos quais duplexes de RNA eram complementares a regiões ricas em CpG de promotores específicos. Trabalhos anteriores demonstraram, de maneira convincente, que duplexes induzem o silenciamento transcricional quando complementares a regiões de promotores ricas em CpG via indução de modificações pós-traducionais em histonas ou no nível de metilação do DNA cromossômico (Morris et al., 2004; Ting et al., 2005). De fato, Li et al. (2006) obtiveram níveis de silenciamento transcricional dos genes *E-cadherin*, *p21* e *VEGF* semelhantes aos encontrados por outros grupos de pesquisa. Entretanto, quando os autores transfectaram células humanas com saRNAs complementares a regiões distintas nos mesmos promotores, isto é, regiões possuindo baixo conteúdo de GC e ausência de sequências invertidas e/ou repetidas, eles observaram resposta completamente diferente: ativação da expressão gênica. O aumento no acúmulo de transcritos e de proteína dos genes alvos foi associado à presença desses saRNAs. Além disso, as células transfectadas apresentaram algumas alterações fenotípicas, provavelmente devido ao aumento de expressão dos genes alvos. Baseados nessas observações, os autores concluíram que saRNAs são capazes de induzir a expressão gênica via sua complementaridade com sequências de promotores.

Janowski et al. (2007) chegaram à mesma conclusão, avaliando a ação de duplexes de RNAs complementares a outros promotores. Neste caso, a sequência alvo foi a do promotor do gene *progesterone receptor (PR)*. Os autores observaram alterações na resposta fisiológica de células transfectadas com os duplexes devido ao aumento na expressão do gene *PR*.

Uma observação importante feita pelos dois grupos foi que a ativação gênica via saRNAs não ocorre ao acaso, mas sim é determinada por tipos celulares específicos e pela regulação epigenética do próprio gene alvo. Li et al. (2006), por exemplo, detectaram aumento nos níveis de transcritos e de proteína do gene *E-cadherin* em células de câncer de próstata (PC-3 e DU-145) após transfecção com saRNAs. Tal efeito não foi observado em células HeLa, nas quais o promotor do gene em questão é altamente metilado em sequências CpG. Observações desse trabalho indicam que o nível de metilação do promotor do gene alvo determina a eficiência de ativação da expressão via saRNAs. Janowski et al. (2007) relatam que a variação nos níveis basais de expressão do gene *PR* em distintas células de câncer de mama, determina a eficiência de ativação via saRNAs. Em células expressando

altos níveis de *PR*, a ativação foi ineficiente, enquanto que células expressando níveis basais do gene *PR* apresentaram aumentos expressivos no acúmulo de transcritos e de proteína *PR*. Portanto, os dados desses dois trabalhos indicam que RNAa pode ser um mecanismo geral para controlar a expressão de genes específicos em determinados tipos e condições celulares, gerando respostas metabólicas e fenotípicas que podem estar relacionadas ao desenvolvimento de cânceres.

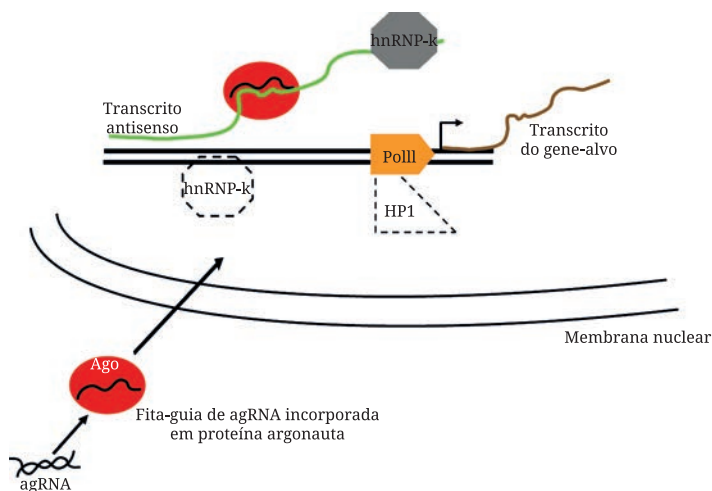
A partir de sua descoberta, várias perguntas emergiram em relação ao fenômeno de regulação gênica via RNAa: a primeira pergunta que surge é referente ao mecanismo molecular implicado no processo de ativação gênica, visto que saRNAs modificados são também capazes de induzir o silenciamento de genes alvos (Janowski et al., 2007). A ativação ou silenciamento gênico são determinados pela especificidade de sequência entre os saRNAs e seus promotores alvos (Rossi, 2007). Uma única alteração de base nucleotídica é suficiente para ativação ou repressão transcricional (Janowski et al., 2007). Concomitantemente, os mesmos autores avaliaram várias regiões alvos na sequência do promotor do gene *PR* e testaram diferentes tipos de saRNAs para ativação ou silenciamento gênico. Os resultados de tais experimentos falharam em apontar um consenso de sequência de DNA que diferencie a ativação de silenciamento. Além disso, a mesma análise realizada em outros promotores demonstrou que não há um aparente consenso na região do promotor em que a ativação ou silenciamento é mais eficiente. Por exemplo, o saRNA mais eficiente para a ativação do gene *PR* é complementar à região -11 a +8, enquanto que, para o gene *major vault protein*, o saRNA mais eficiente na ativação é complementar à região -54 a -36 de seu promotor. Tais resultados indicam que são necessários dados experimentais adicionais para esclarecer qual(is) mecanismo(s) molecular(es) é(são) responsável(is) pela distinção da ação de saRNAs como ativadores ou repressores da expressão gênica. Entretanto, uma pista inicial é que saRNAs direcionados a regiões do genoma propensas a metilação (por exemplo, regiões ricas em sequências CpG) irão provavelmente atuar como repressores ao invés de ativadores da expressão de genes alvos, pelo menos no caso do gene *E-cadherin* (Li et al., 2006).

Além de induzir o silenciamento, saRNAs que diferem em poucas bases podem ser mais ou menos eficientes ativadores da expressão de genes alvos, resultado semelhante ao observado para siRNAs (Janowski et al., 2007). Tanto os saRNAs perfeitamente complementares à sequência alvo quanto os com poucos “*mismatches*” são capazes de reconhecer a mesma sequência em questão. Portanto, eles competem entre si por sítios de reconhecimento nos promotores, e essa competição afeta o grau de ativação do gene alvo. Esta observação sugere que a geometria de reconhecimento da sequência alvo é um fator crucial para uma eficiente ativação da expressão gênica.

Outra pergunta importante em relação ao mecanismo de RNAa *versus* RNAi é qual ou quais proteínas participam da ativação gênica utilizando duplexes de RNA. Estudo prévio de silenciamento gênico em mamíferos utilizando duplexes demonstrou o envolvimento das proteínas da família Argonauta, Ago1 e Ago2, as quais são componentes essenciais do mecanismo de RNAi (Janowski et al., 2006). Até o momento, não há um consenso de qual proteína argonauta participa no mecanismo de RNAa, embora células com baixa produção da proteína Ago2 pareçam ser incapazes de induzir a ativação gênica via saRNAs (Li et al., 2006). As proteínas Argonautas utilizam pequenos RNAs regulatórios, tais como siRNAs e microRNAs, como guias para o reconhecimento de RNAs mensageiros (mRNAs) alvos. Tal reconhecimento induz a clivagem do mRNA alvo ou repressão da tradução durante o silenciamento gênico via RNAi (Meister et al., 2004). O envolvimento de proteínas argonautas

na atividade de saRNAs sugere que esses dsRNAs têm o potencial de ligar-se a moléculas de RNA, ou seja, os saRNAs guiam o complexo *RISC* para o RNA alvo.

De fato, um trabalho recente demonstra que transcritos antissenso podem ser os alvos primários de saRNAs (Schwartz et al., 2008), pelo menos no caso do gene *PR*. Os autores identificaram transcritos de RNA antissenso que são complementares à região promotora do gene *PR*. Tais RNAs são transcritos na orientação antissenso a partir de exons do gene *PR*, incluindo parte de sua região promotora na qual está contida sequência de reconhecimento dos saRNAs. Os transcritos acumulam-se em níveis elevados em células humanas específicas e não são alvos de clivagem via saRNAs. Entretanto, células contendo níveis baixos desses transcritos antissenso são menos eficientes ou mesmo incapazes de ativar a expressão do gene *PR* via saRNAs, sugerindo que a atividade dos saRNAs na ativação gênica é dependente de RNAs antissenso. Schwartz e colaboradores (2008) demonstraram que saRNAs ligam-se aos transcritos antissenso e recrutam proteínas argonautas para formar o complexo saRNA-RNA antissenso proteína argonauta. Tal complexo atua como um *scaffold*, recrutando ou redirecionando outras proteínas, tais como a *heterochromatin protein1* (HP1) e a *heterogeneous nuclear ribonucleoprotein-k* (hnRNP-k) (Yahi et al., 2008; Venables et al., 2008). Tanto HP1 quanto hnRNP-k são fatores essenciais para a interação DNA e RNA. Um possível modelo do mecanismo de RNAa (proposto por Schwartz et al., 2008) é ilustrado na Figura 1.



**Figura 1. Modelo esquemático do possível mecanismo de ativação gênica via saRNAs.** Linhas cheias e tracejadas representam fatores que se associam e desassociam, respectivamente, com a sequência do promotor do gene alvo após transfecção com saRNAs. Os saRNAs recrutam proteína Argonata para o transcrito antissenso e isto induz a relocação da proteína heterogeneous nuclear ribonucleoprotein-k (hnRNP-k) do promotor para o transcrito antissenso. O complexo saRNA-transcrito antissenso-proteína forma-se nas proximidades do promotor do gene alvo e afeta o balanço da regulação gênica. Ago, Argonata; HP1, heterochromatin protein1; PolII, RNA polimerase II.

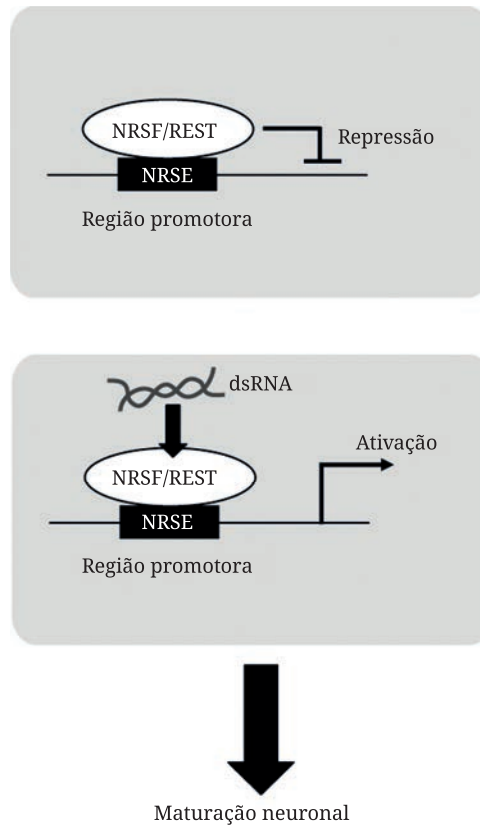
Embora o modelo descrito anteriormente seja o mais recentemente proposto, o exato mecanismo molecular que utiliza saRNAs para ativação gênica ainda necessita de mais

esclarecimentos. É possível que tal mecanismo seja específico de determinados promotores ou de determinados tipos celulares. Além disso, fatores adicionais podem desempenhar importante papel na ativação de genes alvos via o mecanismo de RNAa. Li et al. (2006) demonstraram que a ativação do gene *E-cadherin* em células de câncer de próstata (PC-3), utilizando saRNAs, ocorre devido a uma redução no padrão de metilação da proteína histona 3 (H3). Altos níveis de metilação de H3 reprimem a expressão de vários genes nesses tipos celulares, incluindo o *E-cadherin* (Shi et al., 2003). Novamente, experimentos adicionais, avaliando a ativação de outros genes com saRNAs distintos, são necessários para um melhor entendimento do mecanismo de RNAa.

Embora comprovado experimentalmente, o fenômeno de RNAa, tal como discutido anteriormente, não foi ainda identificado para duplexes de RNA endógenos. Contudo, Kuwabara et al. (2004) relataram a identificação de um pequeno duplex (21 nt) que possui sequência antissenso complementar ao elemento em *cis* NRSE, o qual está presente em promotores de genes especificamente ativos em neurônios. Esse elemento em *cis* é reconhecido pela proteína NRSF/REST que mantém tais genes desligados durante a manutenção do estado não diferenciado de células pré-neuronais. Esse desligamento depende da capacidade de NRSF/REST ligar-se ao elemento NRSE (Lunyak e Rosenfeld, 2005). O dsRNA antissenso ao elemento NRSE, produzido a partir de molécula precursora de RNA não codificante (~20-40 nt), interage com a proteína NRSF/REST sem afetar sua ligação com o elemento em *cis*, ou seja, a proteína repressora mantém-se associada aos promotores dos genes alvos. Entretanto, a formação do complexo dsRNA-proteína NRSF/REST funciona como ativador da expressão gênica (Figura 2). Em outras palavras, a presença do duplex libera genes específicos de seu estado de repressão gênica, gerado, inicialmente, pela ligação do repressor transcricional NRSF/REST na sequência NRSE de regiões promotoras. Embora o mecanismo ainda seja desconhecido, os autores sugerem que esse *switch* na regulação da expressão gênica em células pré-neuronais induz sua maturação final em neurônios (Figura 2). Portanto, RNAs de dupla fita endógenos podem funcionar como ativadores da expressão gênica, desempenhando papéis-chave na diferenciação celular, o que pode implicá-los também na formação de células tumorais.

De forma similar ao uso de siRNAs no mecanismo de RNAi, células utilizam microRNAs (miRNAs) para regular a expressão gênica endógena via repressão da tradução e/ou clivagem de mRNA alvo complementar (Flynt e Lai, 2008). Recentemente, Place e colaboradores (2008) relataram um caso em que miRNAs podem funcionar também como ativadores da expressão gênica, tal como os saRNAs. Buscas computacionais de sequências complementares a miRNAs no genoma humano relevaram a complementaridade do microRNA miR-373 à sequência dos promotores dos genes *E-cadherin* e *CSDC2*. O mesmo grupo já havia demonstrado que saRNAs complementares ao promotor do gene *E-cadherin* são capazes de induzir sua expressão em células tumorais (Li et al., 2006).

A biossíntese de miRNAs é iniciada pela transcrição dos genes *MIR* via RNA PolII. O transcrito primário, ou pri-miRNA, é um RNA longo de fita simples e poliadenilado que forma estrutura secundária do tipo *hairpin*. O pri-miRNA é processado via a ação de diferentes enzimas, produzindo o precursor pre-miRNA. O pre-miRNA é processado novamente, gerando um “duplex” imperfeito de RNA (20-22 nt) que contém tanto o miRNA maduro quanto sua fita complementar (denominado miRNA\*). O miRNA maduro é posteriormente incorporado ao complexo proteico RISC (tal como os siRNAs) e guia a clivagem de mRNAs complementares ou promove a repressão da tradução de genes alvos (Flynt e Lai, 2008).



**Figura 2. Representação esquemática de ativação da expressão gênica via dsRNAs complementares ao elemento em cis NRSE.** Complexo proteico contendo proteína NRSF/REST liga-se ao elemento em cis NRSE, reprimindo a expressão de genes específicos de neurônios. Em determinadas condições fisiológicas e estágio de desenvolvimento, células pré-neuronais produzem dsRNAs que ligam-se ao complexo proteico NRSF/REST, ativando os mesmos genes e induzindo a maturação neuronal.

A possibilidade de que miRNAs também funcionem como ativadores transcricionais foi primeiramente cogitada com base em dados de complementaridade entre os saRNAs e suas seqüências alvos. Os saRNAs com poucos *mismatches* são ainda capazes de ativar a expressão gênica (Li et al., 2006). O não pareamento de algumas bases entre a seqüência do miRNA maduro e seu alvo é comum e tolerável durante o silenciamento gênico mediado por esses RNAs regulatórios (Faller e Guo, 2008), o que sugere que miRNAs podem apresentar *mismatches* em relação à seqüência promotora alvo e ainda assim serem funcionais. Desta forma, Place et al. (2008) transfectaram células de câncer de próstata (PC-3) tanto com o miR-373 maduro quanto com o pre-miR-373, os quais possuem região de complementaridade com alguns *mismatches* em relação à seqüência dos promotores dos genes *E-cadherin* e *CSDC2*. Os autores demonstraram que a expressão desses genes foi induzida por transfecção de células com microRNA miR-373. Tal indução foi específica e dependente do nível de complementaridade entre o microRNA e a seqüência alvo dos promotores em questão.

Além disso, tal como observado para os saRNAs, a eficiência de indução gênica via miRNAs também é dependente de fatores epigenéticos (metilação, etc.) que interferem na atividade do promotor.

Apesar de demonstrarem incontestavelmente que células transfectadas com o miR-373 induzem a expressão gênica, os autores não conseguiram demonstrar que o miR-373 endógeno é capaz de funcionar da mesma maneira. É possível que a atividade endógena de ativação gênica via miRNAs seja específica de tipos celulares ou estágios de desenvolvimento celular não testados pelos autores. Tal hipótese é embasada pela recente descoberta de que miRNAs podem ativar não somente a expressão gênica, mas também a tradução de proteínas (Vasudevan et al., 2007). Sequências em tandem AUUUA, denominadas elementos AREs, estão presentes na região 3' não traduzida de alguns mRNAs. Os elementos AREs são sítios de ligação para as proteínas Ago2 e *fragile X mental retardation-related protein 1* (FXR1). Tal associação leva ao aumento da tradução de proteínas de genes específicos. Como Ago2 está associada a microRNAs, Vasudevan et al. (2007) investigaram a possibilidade de o aumento na tradução ser devido à presença de miRNAs. De fato, os autores identificaram miRNAs complementares aos elementos AREs. Interessantemente, esses miRNAs podem mediar a ativação ou repressão da tradução, dependendo do estágio de divisão em que as células se encontram durante o ciclo celular. Em células no estágio de proliferação, complexo contendo Ago2 e microrribonucleoproteínas é recrutado por miRNAs para o mRNA alvo, reprimindo a tradução de proteínas. Contudo, após o bloqueio do ciclo celular, esse complexo, juntamente com a proteína FXR1, associa-se ao mRNA alvo promovendo a tradução de proteínas e, portanto, induzindo a ativação de genes alvos. Tal associação do complexo proteico ao mRNA alvo é também mediada por miRNAs, indicando um papel duplo dos miRNAs no controle da tradução de proteínas. Interessantemente, o papel de ativador gênico parece ser desempenhado por miRNAs bem documentados na literatura, embora essa faceta de sua ação tenha sido notada somente recentemente. Por exemplo, o microRNA Let-7, conhecido por reprimir a tradução de seus alvos em células animais (Roush e Slack, 2008), pode também ativá-la, dependendo do estágio de desenvolvimento celular (Vasudevan et al., 2007). Essa descoberta, juntamente com a indução da expressão gênica, imediatamente sugere que fatores específicos do desenvolvimento celular regulam a ação de miRNAs como ativadores ou repressores da atividade gênica. Quais seriam e como agem esses reguladores estão entre as intrigantes perguntas biológicas a serem respondidas pela ciência em futuro próximo. A ativação de certos genes via microRNAs, tanto transcricional como traducionalmente, revela uma complexa e fundamental importância desses pequenos RNAs como reguladores endógenos da expressão gênica.

Os resultados discutidos nos parágrafos anteriores indicam claramente uma nova via de ação de pequenos duplexes de RNA na regulação gênica de eucariotos. Não apenas repressores da expressão gênica, os duplexes de RNA (incluindo miRNAs) possuem também papel relevante na ativação de genes em determinados tipos celulares e condições fisiológicas. Essa descoberta possui implicações claras no uso de RNAi e RNAa como ferramentas da regulação da expressão gênica. Ambas as técnicas vêm sendo propostas como possíveis modalidades terapêuticas para o tratamento de doenças específicas, incluindo alguns tipos de cânceres (Janowski et al., 2007). Um passo inicial nessa direção foi iniciado com os trabalhos de Mao et al. (2008) e Yang et al. (2008). Os dois trabalhos demonstraram que a ativação de genes associados ao câncer de bexiga, via saRNAs, reduziu a proliferação e viabilidade das células tumorais. Tais genes, quando ativados, induzem uma cascata de eventos que culmina na repressão de fatores importantes para o desenvolvimento de células tumorais.

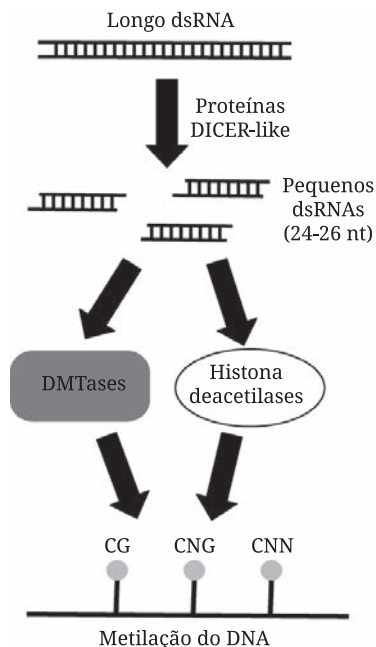
Contudo, todos os experimentos utilizando RNAa, até o momento, foram realizados em culturas celulares. Portanto, os eventos de ativação gênica descritos foram essencialmente transientes (Li et al., 2006; Janowski et al., 2007; Schwartz et al., 2008; Place et al., 2008; Yang et al., 2008; Mao et al., 2008). Com o intuito de continuar as pesquisas nesta intrigante área, além de células humanas, também já foram realizados testes em células de chimpanzé (WES) e macaco verde africano (COS1). Estes animais foram escolhidos por possuírem um genoma quase idêntico ao nosso e já serem utilizados para testes no desenvolvimento de novos fármacos. A ativação de genes como *E-cadherin*, *p21*, *VEGF*, *p53*, entre outros, potencialmente envolvidos no tratamento de cânceres, foi desencadeada nas células WES e COS1, pela introdução de moléculas de saRNA humanas. Este resultado abre espaço para a validação de novas drogas baseadas em RNAa em outros organismos que não humanos (Huang et al., 2010). O próximo passo é analisar esse mecanismo de regulação gênica do organismo e avaliar se a indução de expressão de genes específicos também ocorre nessas condições. Somente após este importante passo, ampliaremos nosso conhecimento de modo a utilizar a ferramenta de RNAa da maneira mais eficiente e segura possível. Entretanto, podemos já vislumbrar um futuro (possivelmente próximo) no qual será possível combinar RNAi e RNAa de modo a silenciar e ativar genes simultaneamente com propósitos terapêuticos. De fato, a empresa *Alnylam Pharmaceuticals* nos EUA, uma das líderes mundiais no uso de RNAi como modalidade terapêutica, anunciou, em agosto de 2008, que irá investir no uso de RNAa em terapias genéticas em futuro próximo. Por exemplo, o mecanismo de RNAa pode ser utilizado para reativar genes supressores de tumores, os quais estão normalmente silenciados em células tumorais (Zhao and Epstein, 2008). Os estudos com células tumorais já tiveram algum avanço em testes utilizando a tecnologia de ativação por RNA. Na China, pesquisadores conseguiram inibir o crescimento de células cancerígenas com a ativação do *p21*, que modula negativamente a progressão do ciclo celular, via saRNA, e ainda aumentar a sensibilidade das células tumorais a drogas utilizadas em tratamentos quimioterápicos (Wei et al., 2010). Não apenas o tratamento de alguns tipos de câncer pode ser beneficiado pelos avanços tecnológicos desta crescente área, como também outros problemas de saúde. Um exemplo envolve o gene *VEGF* (fator de crescimento vascular), cuja ativação em células primárias musculares esponjosas do corpo cavernoso humano foi demonstrada recentemente. Esta resposta positiva abre portas para um possível medicamento para casos de disfunção erétil (Chen et al., 2011).

A manipulação de células para indução de pluripotência sempre foi um enorme desafio para a ciência moderna. Agora, o sonho de manipular o destino celular está cada vez mais próximo de se tornar realidade. Com a recente descoberta de que um importante fator de transcrição – *NANOG*, o qual tem participação crítica na regulação do destino da célula – é superexpresso na presença de saRNA, ilumina-se o caminho para a ativação de genes relacionados com o desenvolvimento e a manipulação do destino celular (Wang et al., 2012).

## **Silenciamento Transcricional: dsRNAs Atuando no DNA**

Na seção anterior, foi discutido sobre o papel de pequenos dsRNAs na ativação e no silenciamento pós-transcricional de genes específicos, formando o “Yin e Yang” da regulação da expressão gênica em eucariotos. Como a ativação gênica via dsRNAs (RNAa) foi abordada anteriormente, esta seção será dedicada a discutir mais detalhadamente os mecanismos envolvidos no silenciamento transcricional mediado por dsRNAs.

O silenciamento transcricional dependente de dsRNAs foi primeiramente documentado em plantas infectadas por viroides (Wassenegger et al., 1994) e pode envolver tanto a metilação de sequências específicas de DNA (*RNA-dependent DNA methylation* ou RdDM) quanto modificações na cromatina. Em plantas, pequenos dsRNAs ou siRNAs (24-26 nt) recrutam DNA metiltransferases (DMTases) e proteínas modificadoras de cromatina (*e.g.*, histona deacetilases) para a metilação *de novo* de regiões do genoma, principalmente regiões ricas em sequências CG, CNG e CNN (Figura 3; Kawasaki et al., 2005).



**Figura 3. Metilação de DNA mediada por dsRNAs em plantas.** Longos dsRNAs são processados por proteínas DICER-like em dsRNAs de 24-26 nt. Esses dsRNAs são incorporados no complexo contendo proteína Argonata, recrutando DNA metiltransferases (DMTases) e proteínas modificadoras de cromatina (histona deacetilases) para a metilação *de novo* de regiões do genoma ricas em sequências CG, CNG e CNN.

Os dsRNAs são utilizados por complexos proteicos como guias para identificar genes alvos exógenos ou endógenos e induzir o silenciamento transcricional. No caso de plantas, tanto viroides quanto transgenes são utilizados como substratos para gerar siRNAs, que guiam a metilação da base citosina em sequências homólogas no genoma vegetal (Henderson e Jacobsen, 2007), funcionando como um mecanismo de defesa contra agentes exógenos. Os principais alvos endógenos para o silenciamento transcricional dependente de dsRNAs são os transposons (Girard e Hannon, 2008). Transposons são ácidos nucleicos parasitas



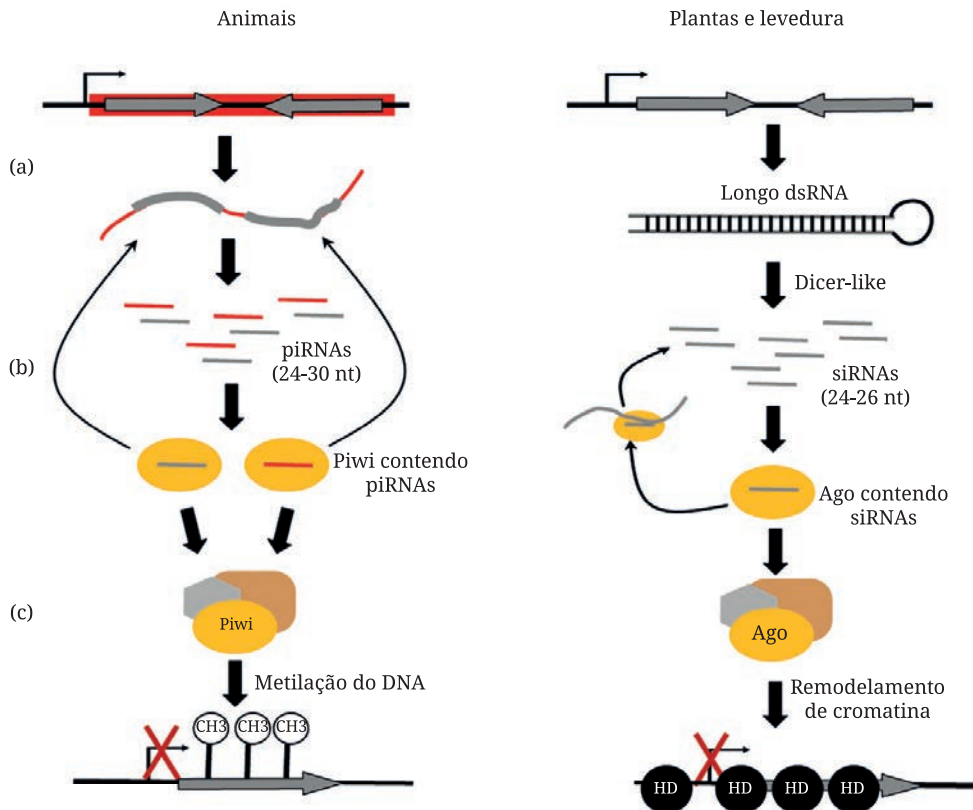
capazes de mover-se e propagar-se no genoma hospedeiro, alterando a expressão gênica (Kazazian Junior, 2004). A movimentação e propagação desordenadas de transposons podem levar à disrupção de genes importantes, além de causar rearranjos cromossômicos, gerando, muitas vezes, mutações deletérias no organismo hospedeiro. Por isso, o genoma de vários organismos eucarióticos utiliza siRNAs como defesa contra transposons, ou seja, esses pequenos dsRNAs participam de mecanismos de supressão de transposons. Tais mecanismos variam nos detalhes de organismo para organismo, mas, em geral, as etapas são conservadas.

No caso de silenciamento transcricional de transposons, três etapas são essencialmente constantes em todos os organismos estudados: (1) detecção de *loci* de transposons via produção de transcritos que são reconhecidos como substratos pela maquinaria de RNAi; (2) produção de pequenos RNAs (amplificação); e (3) incorporação de pequenos RNAs em complexo proteico e posterior recrutamento de DMTases e proteínas modificadoras de cromatina para metilação de DNA e formação de heterocromatina (Girard e Hannon, 2008).

Em plantas e leveduras, a etapa de reconhecimento pode iniciar-se pela transcrição de sequências invertidas de transposons, o que leva à formação de fitas duplas de RNA. Tais transcritos tornam-se substratos para a geração de siRNAs por meio da ação de proteínas DICER-like (Figura 3; Slotkin e Martienssen, 2007). Em animais, a etapa de reconhecimento é menos conhecida, visto que envolve uma classe de pequenos RNAs identificada somente recentemente, os *Piwi-interacting RNAs* (piRNAs; Girard e Hannon, 2008). Os piRNAs são pequenos RNAs de 24-30 nucleotídeos que associam-se especificamente com proteínas denominadas *Piwi*, as quais compartilham domínios conservados com as proteínas Argonautas (Girard et al., 2006). Diferentemente de siRNAs, piRNAs são produzidos a partir de precursores de RNA de fita simples policistrônicos, os quais são transcritos de *loci* formados por *clusters* desses pequenos RNAs. Tais precursores são posteriormente processados em piRNAs por meio de mecanismo ainda não elucidado (Girard e Hannon, 2008). Interessantemente, alguns transposons estão contidos nesses *clusters* como sequências invertidas (Figura 4a). Portanto, após seu processamento, os transcritos policistrônicos geram piRNAs complementares à sequência do *locus* de origem e geram também pequenos RNAs complementares às sequências invertidas dos transposons (Figura 4b; Brennecke et al., 2007).

Embora o mecanismo de geração de piRNAs primários ainda não esteja esclarecido, estes são utilizados para clivar mais precursores de piRNAs, via *Piwi*, aumentando conseqüentemente o número final de piRNAs (piRNAs secundários). No caso de plantas e leveduras, os siRNAs primários incorporados na proteína Argonauta clivam outros transcritos de transposons com sequências conservadas, gerando os siRNAs secundários (Figura 4b; Sugiyama et al., 2005). A etapa de amplificação mantém o sistema constantemente alimentado de piRNAs e siRNAs, garantindo a eficiência do silenciamento.

Na etapa de repressão transcricional, o complexo proteico contendo *Piwi* e piRNAs recruta DMTases para metilar a sequência de DNA correspondente aos transposons alvos (Figura 4c). Em plantas e leveduras, o complexo proteico contendo Argonauta e siRNAs recruta tanto DMTases quanto histona deacetilases para metilação de DNA e formação de heterocromatina, respectivamente (Figura 3 e 4c). Tais modificações epigenéticas impedem a transcrição dos transposons, inibindo sua atividade no genoma. Entretanto, é importante ressaltar que nem todos os transposons existentes no genoma são regulados pelo mecanismo de silenciamento transcricional mediado por pequenos dsRNAs ou piRNAs. Além disso, em alguns casos, o silenciamento transcricional e pós-transcricional de transposons ocorre concomitantemente (Kawasaki et al., 2005).



**Figura 4. Silenciamento transcricional de transposons em eucariotos.** (a) Em animais, piRNAs (24-30 nt) complementares a transposons são gerados a partir de precursor de RNA policistrônico de piRNAs (vermelho) que contém sequências invertidas de transposons (cinza) embebedas no *locus*. Já em plantas e levedura, *locus* de transposon (cinza) com sequências invertidas produz um dsRNA longo com estrutura do tipo *harpin*. (b) Embora ainda não elucidado, mecanismo inicial de clivagem gera piRNAs primários a partir de seu precursor. Tais piRNAs são incorporados na proteína *Piwi* e induzem a clivagem *de novo* de precursores, gerando os piRNAs secundários. O dsRNA longo é clivado pela proteína *DICER-like*, gerando siRNAs (24-26 nt), que, por sua vez, são incorporados em proteína Argonauta (*Ago*). O complexo siRNA-proteína Argonauta cliva transcritos de transposons alvos, gerando os siRNAs secundários. A produção de RNAs secundários é denominada de etapa de amplificação. (c) O complexo piRNA-*Piwi* recruta DMTases para inserir grupos metil (CH3) nas sequências alvos de transposons de animais. Já em plantas, o complexo Argonauta-siRNA recruta tanto DMTases (como mostrado na Figura 3) como proteínas remodeladoras de cromatina, tais como as histonas deacetilases (HD). Tanto a metilação como a formação de heterocromatina eficientemente promovem o silenciamento transcricional de transposons.

Outras sequências de DNA também produzem siRNAs e tornam-se alvos de silenciamento transcricional. Tais regiões no genoma incluem sequências de DNA repetitivas e invertidas, geralmente localizadas na heterocromatina. Na levedura *Schizosaccharomyces pombe*, sequências repetitivas de DNA estão presentes na região pericentromérica dos cromossomos, formando a heterocromatina. Essas sequências são reguladas transcricionalmente por siRNAs gerados a partir de RNAs transcritos naquela região (Volpe et al., 2002). O silenciamento transcricional mediado por siRNAs em *S. pombe* ocorre exclusivamente via modificação da heterocromatina, enquanto que, em animais e plantas, o mecanismo predominante é o RdDM.

Em plantas, regiões repetitivas de DNA também podem estar presentes em genes codificadores de proteínas, tornando-os alvos de silenciamento por RdDM. O melhor exemplo descrito até o momento é o do gene *FWA* que codifica um fator de transcrição envolvido no controle do período de florescimento de *Arabidopsis thaliana* (Soppe et al., 2000). *FWA* é expresso no endosperma de sementes de *A. thaliana*, enquanto que sua expressão é reprimida em outros tecidos jovens da planta. Tal repressão é causada por metilação de sequências simétricas do tipo CpG e assimétricas do tipo CpHpH, presentes em sequências repetitivas de DNA no promotor e na região 5' não traduzida do gene *FWA*.

A manutenção do silenciamento transcricional do gene *FWA* é um processo extremamente complexo e envolve mecanismos dependentes e independentes de siRNAs (Lavrov e Kibanov, 2007). Neste capítulo, será discutido somente o mecanismo envolvendo RdDM. As regiões repetitivas presentes no promotor, próximas à região 5' não traduzida do gene *FWA*, são transcritas por uma RNA polimerase IV (RNAPolIV), gerando um transcrito de RNA de fita simples. Esse transcrito torna-se substrato para a proteína denominada *RNA-dependent RNA polimerase* (RdRP) que gera um dsRNA longo. O dsRNA longo, por sua vez, é processado pela maquinaria de RNAi, gerando siRNAs (24-26 nt). Finalmente, os siRNAs são incorporados na proteína Argonauta, que, neste caso, forma um complexo com a RNAPolIV. Tal complexo liga-se às regiões regulatórias do gene *FWA* e recruta DMTases para a metilação *de novo* (Lavrov e Kibanov, 2007). A regulação da expressão do gene *FWA* via RdDM é extremamente importante para a manutenção do correto período de florescimento em *A. thaliana*. Em plantas transgênicas ou mutantes, nas quais genes envolvidos no mecanismo de RdDM são inativos, ou seja, não há repressão de expressão gênica via RdDM, *FWA* é expresso em vários tecidos na planta jovem e adulta, resultando em um significativo atraso no florescimento. Este exemplo em plantas ilustra a complexidade da regulação transcricional mediada por dsRNAs, a qual não depende somente de pequenos RNAs, mas também das várias proteínas envolvidas na biogênese e ação de siRNAs, além daquelas envolvidas no processo de metilação e remodelamento da cromatina.

Como foi visto até o momento, a função de pequenos dsRNAs ou siRNAs na repressão da transcrição é funcionar como guias para revelar áreas no genoma, nas quais os complexos proteicos de metilação e remodelamento de cromatina devem ligar-se e atuar. A identificação de tais áreas ocorre devido ao pareamento entre os siRNAs e a sequência alvo de DNA ou RNA. Depois desse reconhecimento, DNA metiltransferases e proteínas modificadoras de cromatina, além de outras, são recrutadas, metilando o DNA e compactando a cromatina. Entretanto, se dsRNAs participam ativamente da modificação do DNA, e estes reconhecem primariamente sequências alvos de RNA, a etapa de transcrição inicial é necessária. Em outras palavras, o precursor de RNA necessita ser produzido para servir de substrato para a maquinaria de RNAi e, portanto, gerar os pequenos RNAs primários. Pelo menos em plantas, uma solução para essa contradição parece existir: diferente de animais e levedura, plantas possuem um quarto tipo de RNA polimerase, a RNAPolIV. A atividade desta RNA polimerase parece não ser afetada pelo estado de heterocromatina do genoma, permitindo assim a transcrição de sequências repetitivas de DNA. Embora este modelo seja ainda especulativo, a presença da RNAPolIV parece contribuir com a eficiência do silenciamento transcricional dependente de siRNAs. Em animais e leveduras, a proteína responsável pela transcrição inicial de sequências repetitivas não foi ainda identificada.

A repressão no nível transcricional, mediada por dsRNAs, também ocorre em genes humanos. A transfecção de células humanas com dsRNAs complementares a sequências de

promotores específicos leva ao silenciamento transcricional via formação de heterocromatina. Tal silenciamento requer a presença de uma molécula de RNA complementar à região promotora do gene alvo, que se liga à fita antissenso dos dsRNAs exógenos. A hibridação entre a fita antissenso do dsRNA e o RNA complementar ao promotor (*promoter-associated RNA* ou pRNA) funciona como um *tag* para que as proteínas remodeladoras de cromatina associem-se com o gene alvo e promovam o seu silenciamento (Han et al., 2007). Este e outros trabalhos sugerem a existência de dsRNAs endógenos também capazes de promover silenciamento transcricional de genes humanos. Embora nenhum siRNA endógeno associado à repressão no nível de transcrição de genes codificadores de proteínas tenha sido identificado até o momento, há trabalhos recentes que indicam que miRNAs participam do silenciamento transcricional, além do silenciamento pós-transcricional e repressão da tradução de proteínas (Kim et al., 2008; Younger e Corey, 2011). Com certeza, novos estudos irão surgir demonstrando o papel de dsRNAs endógenos (incluindo miRNAs) na regulação transcricional de genes, aumentando ainda mais a complexidade existente nas diferentes “camadas” de regulação gênica em eucariotos.

## Referências Bibliográficas

- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam R. and Hannon, G. J. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128:1089-1103. PMID:17346786. <http://dx.doi.org/10.1016/j.cell.2007.01.043>
- Britten, R. J. and Davidson, E. H. 1969. Gene regulation for higher cells: a theory. *Science* 165:349-357. PMID:5789433. <http://dx.doi.org/10.1126/science.165.3891.349>
- Chen, R., Wang, T., Rao, K., Yang, J., Zhang, S., Liu, J. and Ye, Z. 2011. Up-regulation of VEGF by small activator RNA in human corpus cavernosum smooth muscle cells. *J Sex Med* 10:2773-2780. PMID:21819543. <http://dx.doi.org/10.1111/j.1743-6109.2011.02412.x>
- Faller, M. and Guo, F. 2008. MicroRNA biogenesis: there's more than one way to skin a cat. *Biochim Biophys Acta* 1779:663-667.
- Flynt, A. S. and Lai, E. C. 2008. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Gen* 9:831-842.
- Girard, A., Sachidanandam, R., Hannon, G. J. and Carmell, M. A. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199-202. PMID:16751776.
- Girard, A. and Hannon, G. J. 2008. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol* 18:136-148. PMID:18282709 PMCid:PMC2995447. <http://dx.doi.org/10.1016/j.tcb.2008.01.004>
- Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. 2002. Two classes of short interfering RNA in RNA silencing. *Embo J* 21:4671-4679. PMID:12198169 PMCid:PMC125409. <http://dx.doi.org/10.1093/emboj/cdf464>
- Han, J., Kim, D. and Morris, K. V. 2007. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci USA* 104:12422-12427. PMID:17640892 PMCid:PMC1924466. <http://dx.doi.org/10.1073/pnas.0701635104>
- Hannon, G. J. 2002. RNA interference. *Nature* 418:244-251. PMID:12110901. <http://dx.doi.org/10.1038/418244a>
- Henderson, I. R. and Jacobsen, S. E. 2007. Epigenetic inheritance in plants. *Nature* 447:418-424. PMID:17522675. <http://dx.doi.org/10.1038/nature05917>
- Huang, V., Qin, Y., Wang, J., Wang, X., Place, R. F., Lin, G., Lue, T. F. and Li, L. C. 2010. RNAa is conserved in mammalian cells. *Plos One* 5:e8848. PMID:20107511 PMCid:PMC2809750. <http://dx.doi.org/10.1371/journal.pone.0008848>

- Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., Minna, J. D. and Corey, D. R. 2006. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* 13:787-792. PMID:16936728. <http://dx.doi.org/10.1038/nsmb1140>
- Janowski, B. A., Younger, S. T., Hardy, D. B., Ram, R., Huffman, K. E. and Corey, D. R. 2007. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs, *Nat Chem Biol* 3:166-173. PMID:17259978. <http://dx.doi.org/10.1038/nchembio860>
- Kawasaki, H., Taira, K. and Morris, K. V. 2005. siRNA induced transcriptional gene silencing in mammalian cells. *Cell Cycle* 4:442-448. PMID:15684610. <http://dx.doi.org/10.4161/cc.4.3.1520>
- Kazazian Junior, H. H. 2004. Mobile elements: drivers of genome evolution. *Science* 303:1626-1632. PMID:15016989. <http://dx.doi.org/10.1126/science.1089670>
- Kim, D. H., Sætrom, P., Snøve Junior, O. and Rossi, J. J. 2008. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci USA* 105:16230-16235. PMID:18852463 PMID:PMC2571020. <http://dx.doi.org/10.1073/pnas.0808830105>
- Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K. and Gage, F. H. 2004. A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* 116:779-793. [http://dx.doi.org/10.1016/S0092-8674\(04\)00248-X](http://dx.doi.org/10.1016/S0092-8674(04)00248-X)
- Lavrov, S. A. and Kibanov, M. V. 2007. Noncoding RNAs and chromatin structure. *Biochemistry (Mosc)* 72:1422-1438.
- Li, L.-C., Okino, S. T., Zhao, H., Pookot, D., Place, R. F., Urakami, S., Enokida, H. and Dahiya, R. 2006. Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci USA* 103:17337-17342. PMID:17085592 PMID:PMC1859931. <http://dx.doi.org/10.1073/pnas.0607015103>
- Lunyak, V. V. and Rosenfeld, M. G. 2005. No rest for REST: REST/NRSF regulation of neurogenesis. *Cell* 121:499-501. PMID:15907461. <http://dx.doi.org/10.1016/j.cell.2005.05.003>
- Mao, Q., Li, Y., Zheng, X., Yang, K., Shen, H., Qin, J., Bai, Y., Kong, D., Jia, X. and Xie, L. 2008. Up-regulation of *E-cadherin* by small activating RNA inhibits cell invasion and migration in 5637 human bladder cancer cells. *Biochem. Biophys Res Commun* 375:466-470. PMID:18725195. <http://dx.doi.org/10.1016/j.bbrc.2008.08.059>
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* 15:185-197. PMID:15260970. <http://dx.doi.org/10.1016/j.molcel.2004.07.007>
- Morris, K. V., Chan, S. W., Jacobsen, S. E. and Looney, D. J. 2004. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* 305:1289-1292. PMID:15297624. <http://dx.doi.org/10.1126/science.1101372>
- Pal-Bhadra, M., Bhadra, U. and Birchler, J. A. 2004. Interrelationship of RNA interference and transcriptional gene silencing in *Drosophila*. *Cold Spring Harb Symp Quant Biol* 69:433-438. PMID:16117678. <http://dx.doi.org/10.1101/sqb.2004.69.433>
- Place, R.F., Li, L., Pookot, D., Noonan, E. J., and Dahiya, R. 2008. MicroRNA-373 Induces Expression of Genes with Complementary Promoter Sequences. *Proceedings of the National Academy of Sciences of the United States of America* 105 (5) (February 5): 1608–13. <http://dx.doi.org/10.1073/pnas.0707594105>
- Roush, S. and Slack, F. J. 2008. The let-7 family of microRNAs. *Trends Cell Biol* 18:505-516 PMID:18774294. <http://dx.doi.org/10.1016/j.tcb.2008.07.007>
- Rossi, J. J. 2007. Transcriptional activation by small RNA duplexes. *Nat Chem Biol* 3:136-137. PMID:17301798. <http://dx.doi.org/10.1038/nchembio0307-136>
- Schwartz, J. C., Younger, S. T., Nguyen, N., Hardy, D. B., Monia, B. P., Corey, D. R. and Janowski, B. A. 2008. Antisense Transcripts Are Targets for Activating Small RNAs. *Nature Structural & Molecular Biology* 15 (8) (August): 842-8. <http://dx.doi.org/10.1038/nsmb.1444>
- Shi, Y., Sawada, J., Sui, G., Affar, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P., Nakatani, Y. and Shi, Y. 2003. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422:735-738. PMID:12700765. <http://dx.doi.org/10.1038/nature01550>
- Slotkin, R. K. and Martienssen, R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8:272-285. PMID:17363976. <http://dx.doi.org/10.1038/nrg2072>

- Sugiyama, T., Cam, H., Verdel, A., Moazed, D. and Grewal, S. I. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci USA* 102:152-157. PMID:15615848 PMCID:PMC544066. <http://dx.doi.org/10.1073/pnas.0407641102>
- Ting, A. H., Schuebel, K. E., Herman, J. G. and Baylin, S. B. 2005. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet* 37:906-910. PMID:16025112 PMCID:PMC2659476. <http://dx.doi.org/10.1038/ng1611>
- Yahi, H., Fritsch, L., Philipot, O., Guasconi, V., Souidi, M., Robin, P., Poleskaya, A., Losson, R., Harel-Bellan, A. and Ait-Si-Ali, S. 2008. Differential cooperation between heterochromatin protein HP1 isoforms and MyoD in myoblasts. *J Biol Chem* 283:23692-23700. PMID:18599480 PMCID:PMC3259757. <http://dx.doi.org/10.1074/jbc.M802647200>
- Yao, M. C., Fuller, P. and Xi, X. 2003. Programmed DNA deletion as an RNA-guided system of genome defense. *Science* 300:1581-1584. PMID:12791996. <http://dx.doi.org/10.1126/science.1084737>
- Vasudevan, S., Tong, Y. and Steitz, J. A. 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318:1931-1934. PMID:18048652. <http://dx.doi.org/10.1126/science.1149460>
- Venables, J. P., Koh, C. S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E. R., Watier, V., Durand, M., Lucier, J. F., Gervais-Bird, J., Tremblay, K., Prinos, P., Klinck, R., Elela, S. A. and Chabot, B. 2008. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol* 28:6033-6043. PMID:18644864 PMCID:PMC2547008. <http://dx.doi.org/10.1128/MCB.00726-08>
- Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. and Martienssen, R. A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833-1837. PMID:12193640. <http://dx.doi.org/10.1126/science.1074973>
- Yang, K., Zheng, X.-Y., Qin, J., Wang, Y.-B., Bai, Y., Mao, Q.-Q., Wan, Q., Wu, Z.-M. and Xie, L.-P. 2008. Up-regulation of p21WAF1/Cip1 by saRNA induces G1-phase arrest and apoptosis in T24 human bladder cancer cells. *Cancer Let* 265:206-214.
- Zhaoand, Y. and Epstein, R. J. 2008. Programmed Genetic Instability: A Tumor-Permissive Mechanism for Maintaining the Evolvability of Higher Species through Methylation-Dependent Mutation of DNA Repair Genes in the Male Germ Line. *Mol Biol Evol* 25:1737-1749. PMID:18535014 PMCID:PMC2464741. <http://dx.doi.org/10.1093/molbev/msn126>
- Wang, X., Wang, J., Huang, V., Place, R. F. and Li, L. C. 2012. Inducing of NANOG expression by targeting promoter sequence with small activating RNA antagonizes retinoic acid-induced differentiation. *Biochem J* 443:821-828. PMID:22339500 PMCID:PMC3327998. <http://dx.doi.org/10.1042/BJ20111491>
- Wassenegger, M., Heimes, S., Riedel, L. and Sänger, H. L. 1994. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* 76:567-576. [http://dx.doi.org/10.1016/0092-8674\(94\)90119-8](http://dx.doi.org/10.1016/0092-8674(94)90119-8)
- Wei, J., Zhao, J., Long, M., Han, Y., Wang, X., Lin, F., Ren, J., He, T. and Zhang, H. 2010. p21WAF1/CIP1 gene transcriptional activation exerts cell growth inhibition and enhances chemosensitivity to cisplatin in lung carcinoma cell. *BMC Cancer* 10:632. PMID:21087528 PMCID:PMC2995802. <http://dx.doi.org/10.1186/1471-2407-10-632>
- Younger, S. T. and Corey, D. R. 2011. Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Res* 39(13):5682-91 PMID:21427083 PMCID:PMC3141263. <http://dx.doi.org/10.1093/nar/gkr155>