UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR MOLECULAR

DESCOBERTAS ASSOCIADAS AOS MECANISMOS DE DEFESA ANTIVIRAL BASEADOS EM SILENCIAMENTO POR RNA EM *GLYCINE MAX* E *NICOTIANA BENTHAMIANA*

Guilherme Cordenonsi da Fonseca

Porto Alegre, RS, Brasil
Junho de 2015

Universidade Federal do Rio Grande do Sul
Centro de Biotecnologia
Programa de Pós-Graduação em Biologia Celular e Molecular

DESCOBERTAS ASSOCIADAS AOS MECANISMOS DE DEFESA ANTIVIRAL BASEADOS EM SILENCIAMENTO POR RNA EM *GLYCINE MAX* E *NICOTIANA BENTHAMIANA*

Guilherme Cordenonsi da Fonseca
Orientador: Dr. Rogerio Margis

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para obtenção do título de Doutor

Porto Alegre, RS, Brasil
Junho de 2015

INSTITUIÇÃO E FONTES FINANCIADORAS

Este trabalho foi desenvolvido no Laboratório de Genomas e Populações de Plantas do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.

Este trabalho utilizou recursos financeiros do projeto Genosoja do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e também da Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES).

AGRADECIMENTOS

Agradeço, primeiramente ao meu orientador Dr. Rogerio Margis pela extraordinária orientação que recebi desde a minha iniciação científica, pelo incentivo e pela liberdade de propor idéias de pesquisa novas ao longo de todos esses anos.

Também agradeço aos colegas do Laboratório de Genomas e Populações de Plantas (LGPP), tanto aos atuais, Érika, Franceli, Frank, Henrique, Maria, Nureyev, Pabulo e Priscila, quanto aos que já passaram pelo laboratório, Andréia, Júlio, Maurício e Lorrayne pela amizade e por tornar o LGPP um ambiente espetacular para o trabalho.

Em especial, agradeço aos colegas Guilherme Loss e Luiz Felipe pela camaradagem e apoio, sem as quais esse trabalho jamais seria possível, além de me introduzirem ao mundo da bioinformática.

Também agradeço ao grupo do Núcleo de Genômica Funcional de Plantas (NGFP) liderado pela Dra. Márcia Margis pela troca de idéias e pelo apoio constante ao longo desse doutorado.

Agradeço também os membros da minha comissão de acompanhamento Dr. Arthur Germano Fett Neto e Dr. Guido Lenz, pela prestatividade e ajuda ao longo do desenvolvimento do trabalho dessa tese, e ao Dr. Diego Bonatto por ser o revisor da mesma.

Agradeço também ao Dr. Peter Waterhouse e a seu grupo que me acolheram muito bem em minha estádia na Austrália, aonde aprimorei bastante o meu conhecimento sobre técnicas de biologia molecular.

E especialmente agradeço a minha família, aos meus irmãos Gustavo e Gabriela, e também aos meus pais, por não medirem esforços no auxílio da minha educação.

ÍNDICE

LISTA DE ABREVIATURAS

+ssRNA - “positive sense single-stranded RNA” (RNA fita simples de orientação positiva)
AGO – ARGONAUTE (proteina envolvida na RNAi, presente no complexo RISC, com atividade de RNAse e/ou ligação de RNAs)
CaMV - *Cauliflower mosaic virus* (vírus do mosaico da couve-flor)
CDS – “coding sequence” (sequência codificadora)
CMV - *Cucumber mosaic virus* (vírus do mosaico do pepino)
CP - “coat protein” (proteína da capa viral)
DCL – DICER-LIKE (proteina envolvida na RNAi, com atividade de RNAse e/ou ligação de dsRNAs)
DRB – DOUBLE-STRANDED BINDING PROTEIN
DRD - DEFECTIVE IN RNA-DIRECTED DNA METHYLATION
DRM - DOMAINS REARRANGED METHYLTRANSFERASE
dsRNA – “double strand RNA” (RNA de dupla fita)
EBLN - “endogenous Borna-like N” (Elemento do tipo Borna N endógeno)
ENV - ENVELOPE
GUS – gene da β-glucoronidase
hc-siRNA - “heterochromatic siRNA” (siRNA heterocromático)
HEN - HUA ENHANCER
LCMV - *Lymphocytic choriomeningitis virus* (vírus da coriomeningite linfocitica)
LINE - “long interspersed nuclear elements” (elementos nucleares interespaçados longos)
LTR - “long terminal repeat” (repetições terminais longas)
miRNA – micro RNA
MITE - “miniature inverted-repeat transposable element” (Elemento transponível de repetições invertidas em miniatura)
MP - “movement protein” (proteína do movimento viral)
NAT - “natural antisense” (anti-senso natural)
nat-siRNA - “natural siRNA” (siRNA natural)
ncRNA - “non-coding RNA” (RNA não codificador)
NIRV – “non-retroviral integrated RNA virus” (vírus de RNA não retroviral integrado)
RNA Pol – RNA polimerase

pre-miRNAs - “miRNA precursor” (precursor do miRNA)

pri-miRNAs - “primary miRNA” (miRNA primário)

pri-ta-siRNA - “primary ta-siRNA” (ta-siRNA primário)

PTGS - “Post-transcriptional gene silencing” (silenciamento gênico pós-transcricional)

RdDM - “RNA dependent DNA methylation” (metilação de DNA dependente de RNA)

PVX - “Potato virus X” (vírus X da batata)

RdRp – RNA-dependent RNA polymerase (RNA polimerase RNA dependente)

RDR - RNA POLYMERASE RNA-DEPENDENT

RISC - “RNA-Induced Silencing Complex” (complexo indutor do silenciamento por RNA)

RNAi – RNA de interferência

satRNA – “satellite RNA” (RNA satélite)

siRNA - “short interference RNA” (pequeno RNA interferente)

sRNA - “small RNA” (pequeno RNA)

ta-siRNA - “*trans*-acting siRNA”

TMV – *Tobacco mosaic virus* (vírus do mosaico do fumo)

TGB – TRIPLE GENE BLOCK

TRV - *Tobacco rattle virus* (vírus do chocoalho do fumo)

unc-22 - *Uncordinated*-22

VIGS - “virus-induced gene silencing” (silenciamento gênico induzido por vírus)

vsiRNA - “viral siRNA” (siRNA viral)

VRC - “viral replication complex” (complexo de replicação viral)

XRN – EXORIBONUCLEASE

DESCOBERTAS ASSOCIADAS AOS MECANISMOS DE DEFESA ANTIVIRAL BASEADOS EM SILENCIAMENTO POR RNA EM *GLYCINE MAX* E *NICOTIANA BENTHAMIANA*

Autor: Guilherme Cordenonsi da Fonseca

Orientador: Rogerio Margis

Resumo

O principal mecanismo de defesa das plantas frente a uma infecção viral é baseado no fenômeno chamado interferência por RNA (RNAi). Por meio da ação coordenada de proteínas como Argonautas, Dicers, RNA polimerases dependentes de RNA e proteínas de ligação a RNA de dupla fita (DRBs), o RNA viral é reconhecido e clivado a pequenos RNAs de interferência derivados de vírus (vsiRNAs). Os vsiRNAs, acoplados ao complexo proteíco de indução ao silenciamento, atuam sobre sequências de RNA ou DNA virais, podendo promover a clivagem, inibição da tradução ou metilação de seus alvos (para vírus de DNA). Neste trabalho foram realizados dois estudos que abordam mecanismos de defesa baseados em RNAi em plantas. No primeiro capítulo é descrito a integração do RNA1 do *Cumcumber mosaic virus* (CMV) no genoma de *Glycine max*. Através da análise de bibliotecas de sequenciamento de alta eficiência de RNA mensageiros (mRNAs), pequenos RNAs e DNAs de diferentes cultivares e diferentes tecidos de soja foi possível identificar que o evento de integração envolveu duas moléculas do RNA1 do CMV, o RNA de um retrotransposon e um mRNA de um gene endógeno. No *locus* aonde ocorreu esta integração as duas sequências do RNA1 estão em sentidos opostos. Os pequenos RNAs (sRNAs) das nossas bibliotecas alinham majoritariamente na região do RNA1 do CMV e são praticamente ausentes nas outras regiões da sequência integrada, sugerindo fortemente a formação de um grampo aonde hibridizam ambas as sequências do CMV. A presença desses sRNAs derivados do CMV em todos os tecidos estudados sugere uma provável função antiviral dessa sequencia que foi integrada em soja. No segundo capítulo, por microscopia confocal, foi estudada a interação entre as proteínas DRBs e o *Potato virus X* (PVX) durante a infecção viral em *Nicotiana benthamiana*. É demonstrado que as DRBs 2, 3 e 5 se realocam de sua posição original e se concentram em estruturas chamadas complexos de replicação viral durante a infecção por PVX. Esse fenômeno é um indicativo que essas proteínas podem estar atuando nos primeiros estágios de defesa da planta frente ao vírus.

Tese de Doutorado, Programa de Pós-graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. Julho, 2015.

FINDINGS RELATED TO THE ANTIVIRAL DEFENSE MECHANISM BASED ON RNA SILENCING IN *GLYCINE MAX* AND *NICOTIANA BENTHAMIANA*

Author: Guilherme Cordenonsi da Fonseca

Advisor: Rogerio Margis

Abstract

The main defense mechanism of plants facing a viral infection is based on the phenomenon called RNA interference (RNAi). Through the coordinated action of proteins such as Argonaut, Dicers, RNA dependent RNA polymerases and double-stranded RNA-binding proteins (DRBs), the viral RNA is recognized and cleaved to virus-derived interfering small RNAs (vsiRNAs). The vsiRNAs, coupled to a protein complex that induce the silencing, act on DNA or RNA viral sequences promoting cleavage, translational inhibition or methylation of their targets (for DNA viruses). This work described two studies that address new defense mechanisms based on RNAi in plants. In the first chapter of this thesis is described the integration of the cucumber mosaic virus (CMV) RNA1 in the genome of *Glycine max*. Through the analysis of deep sequencing libraries of messenger RNA, small RNAs and DNA from different cultivars and different soybean tissues it was possible to identified that the integration event involved two molecules of CMV RNA1, the RNA of a retrotransposon and the mRNA of an endogenous gene. In the locus where the integration occurred the two RNA1 sequences are in opposite directions. Small RNAs (sRNAs) from our libraries mostly aligned in the region of CMV RNA1 and are practically absent in other regions of the integrated sequence, strongly suggesting the formation of a hairpin where both CMV sequences hybridize. The presence of these CMV-derived sRNAs in all surveyed tissues suggests a probable antiviral function for the sequence that was integrated into soybeans. In the second chapter, the interaction between the DRB proteins and the Potato virus X (PVX) during viral infection in Nicotiana benthamiana is assessed by confocal microscopy. It is shown that the DRBs 2, 3 and 5 relocate from its original position and concentrated in structures called viral replication complexes during infection by PVX. This is an indication that these proteins can act in the early stages of plant defense against the virus.

Ph.D. Thesis, Graduate Program in Cell and Molecular Biology, Center for Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil. July, 2015.

1.0 INTRODUÇÃO

1.1 Mecanismos moleculares de interação entre vírus e plantas

Vírus são agentes infecciosos intracelulares obrigatórios que infectam todos os tipos de seres vivos se multiplicando exclusivamente dentro das células de seus hospedeiros. A maioria dos vírus de plantas possuem genoma de RNA de polaridade positiva e fita simples (+ss, do inglês "positive-sense single-stranded"), e que codificam para não mais que uma dúzia de proteínas (WANG, 2015). O genoma relativamente pequeno e a dependência exclusiva do metabolismo celular de seu hospedeiro tornam esses sistemas virais muito importantes no estudo das interações patógeno-hospedeiro (MANDADI & SCHOLTHOF, 2013). O processo de infecção viral em plantas pode ser dividido em algumas etapas principais: desmontagem das partículas virais, tradução do genoma viral (para vírus de +ssRNA), modificações da membrana celular seguido da formação de complexos de replicação viral (VRCs, do inglês "viral replication complexes"), encapsidação do genoma viral, movimentação célula-a-célula e deslocamento à longa distância (WANG, 2015).

Durante a infecção viral, muitos vírus de planta induzem a formação de corpos de inclusão característicos que inicialmente foram nomeados "corpos X", devido a sua função desconhecida (GOLDSTEIN, 1924). Essas estruturas foram valiosas para o diagnóstico de doenças em plantas (MARTELLI & RUSSO, 1977), mas somente foi possível descobrir mais sobre sua estrutura e função através do desenvolvimento das técnicas de biologia molecular (LINNIK et al., 2013). Vírus de +ssRNA se replicam nas superfícies citoplasmáticas de membranas celulares modificadas, aonde se encontram os corpos de inclusão, posteriormente, nomeados de complexos de replicação viral (ASURMENDI et al., 2004). As RNA polimerases dependentes de RNA (RdRp, do inglês "RNA-dependent RNA polymerase") de vírus são usualmente ativas como arranjos oligoméricos, e as membranas celulares que elas ocupam servem como estruturas para a montagem desses complexos (NISHIKIORI et al., 2006). Ao rearranjarem as membranas celulares, os VRCs formam um ambiente protegido para o genoma viral (LALIBERTÉ & SANFAÇON, 2010). Além disso, essas estruturas facilitam o acesso do vírus a recursos essenciais na célula como ribossomos, enzimas e nucleotídeos, podendo ser um local de convergência de

complexos ribonucleoproteicos para o transporte intracelular via plasmodesma (SCHOELZ et al., 2011).

Por muitos anos os mecanismos envolvidos na replicação e movimento viral foram estudados separadamente, como se fossem eventos separados, cada um necessitando os seus próprios mecanismos especializados. Entretanto, estudos recentes demonstram que vírus desenvolvem VRCs tanto para replicação quanto para o movimento (HEINLEIN, 2015). Os vírus de planta movem o seu genoma entre as células do hospedeiro via plasmodesma, que são nanoporos intercelulares que conectam a membrana plasmática, citoplasma e retículo endoplasmático através da parede celular (TILSNER et al., 2011). O transporte viral é mediado por uma ou mais proteínas do movimento (MP, do inglês "movement protein") que ligam a ácidos nucleicos de maneira inespecífica, se movem pelo plasmodesma e dilatam os nanoporos que conectam as células (LUCAS, 2006). Já que as MPs são produzidas por tradução de genomas virais, estas provavelmente se ligam co-traducionalmente as moléculas de RNA viral em sua vizinhança nos VRCs associados a membrana do retículo endoplasmático (HEINLEIN, 2015). Estudos recentes demonstraram que VRCs se desenvolvem em sítios específicos de ancoragem no retículo endoplasmático, aonde eles podem se deslocar intra ou intercelularmente via plasmodesma ou permanecer no local como "fábricas" produtoras de moléculas virais, enquanto a infecção se espalha pela planta (KAWAKAMI et al., 2004; TILSNER et al., 2013). Esses mecanismos de formação e movimento de VRCs associados à membranas são fortemente associados ao citoesqueleto, refletindo a exploração pelo vírus de funções dependentes de actina e microtúbulos na organização endomembranar de plantas (BRANDIZZI & WASTENEYS, 2013).

1.2 O silenciamento por RNA como sistema de defesa antiviral

Silenciamento por RNA é o termo genérico dado aos mecanismos regulatórios guiados por RNA que ocorrem em todos os seres eucariotos. Foi descoberto primeiramente em plantas, quando petúnias foram geneticamente modificadas para possuir cópias adicionais do gene de pigmentação floral, chalcona sintase, com o objetivo de intensificar a coloração púrpura das flores. Entretanto, algumas linhagens apresentaram uma variação grande no padrão de coloração das flores, indo do branco ao púrpura intenso. A análise

molecular dessas linhagens revelou que tanto o gene endógeno quanto os transgenes estavam “silenciados” em diferentes graus, um fenômeno que os autores chamaram de “cossupressão” (NAPOLI et al., 1990). Contudo, somente 8 anos depois, evidências que este processo era acionado por RNAs de dupla fita (dsRNAs) foram identificadas em *Caenorhabditis elegans*. Os nematódeos que foram injetados com moléculas de dsRNAs homólogas em sequência de nucleotídeos à região codificadora do gene *unc*-22 (que codifica uma proteína do miofilamento) apresentaram fenótipo idênticos ao de mutantes com perda de função para esse gene, assim como uma redução drástica de seu RNA mensageiro (FIRE et al., 1998). Atualmente é sabido que todos estes sistemas de silenciamento por RNA caracterizados dividem características em comum. Primeiramente, moléculas de dsRNA endógenas ou exógenas são reconhecidas por enzimas RNase do tipo III, conhecidas por DICER, ou DICER-LIKE (DCL) em plantas, que clivam a mólecula em pequenos RNAs interferentes de dupla fita (siRNAs) de 21 a 24 nt (MILLAR & WATERHOUSE, 2005). Proteínas de ligação a dsRNA (DRB, do inglês “Double-stranded RNA-binding protein) podem interagir com a DICER para facilitar a biogênese do siRNA ou o seu acoplamento ao complexo indutor do silenciamento por RNA (RISC). Ao serem produzidos os siRNAs são recrutados pela proteína ARGONAUTA (AGO), componente catalítico do RISC que, ao reconhecer o seu RNA ou DNA alvo, poderá acionar o silenciamento gênico pós-transcricional (PTGS) através de clivagem de RNA ou inibição da tradução, ou o silenciamento gênico transcricional (TGS) que envolve metilação do DNA ou modificações na cromatina (JAMALKANDI & MASOUDI-NEJAD, 2009).

Em plantas, os componentes da maquinaria de silenciamento por RNA se diversificaram para realizar funções especializadas porém algumas vezes redundantes. A planta modelo *Arabidopsis thaliana* pode codificar 4 DCLs, 10 AGOs, 5 DRBs e 6 RDRs. As RDRs polimerizam uma fita de RNA complementar a partir de um RNA de fita simples de uma maneira independente de “primer” (QI et al., 2009). Desse modo, novos siRNAs poderão ser gerados a partir dessa nova molécula de dsRNA, tendo como alvo outros mRNAs que possuem homologia a sua sequência, amplificando assim o sinal do siRNA.

O papel do RNA de interferência (RNAi) na defesa antiviral em plantas tem sido elucidado ao longo dos anos em diferentes trabalhos (LLAVE, 2010). Há basicamente duas linhas de evidências que comprovam esse mecanismo. A primeira evidência é a de que muitos vírus codificam proteínas que suprimem a atividade de componentes envolvidos na

maquinaria do RNAi, como a proteína 2B do CMV que suprime a atividade de clivagem da AGO1 (ZHANG et al., 2006). A segunda evidência é a de que são encontradas nas células das plantas infectadas por vírus pequenos RNAs de interferência virais (vsiRNAs), evidenciando uma estratégia de defesa da planta contra essas infecções (LLAVE, 2010).

Aparentemente, todas as 4 DCL de *Arabidopsis* estão envolvidas na formação dos vsiRNAs (DELERIS et al., 2006). A DCL4 seria a principal envolvida na clivagem dos dsRNAs virais produzindo siRNAs de 21 nt, a classe mais abundante de vsiRNAs encontradas nos tecidos infectados. Na ausência da DCL4, a DCL2 passa a ser a principal formadora de vsiRNAs, produzindo vsiRNAs de 22 nt. Na ausência das duas, a DCL3 produz vsiRNAs de 24 nt (DELERIS et al., 2006). A contribuição da DCL1 é a menor de todas para a formação de vsiRNAs. São encontrados em mutantes para as outras 3 DCL vsiRNAs, porém, em uma quantidade muito menor (GARCIA-RUIZ et al., 2010). AGO1, AGO2, AGO4 e AGO5, acopladas aos vsiRNAs, são as prováveis proteínas envolvidas na mediação de clivagem dos RNA virais ou metilação do DNA, no caso de vírus de DNA (LLAVE, 2010). RDR1, RDR2 e RDR6 estão envolvidas na resistência viral promovida por RNAi em plantas, entretanto elas cumprem papeis distintos em infecções promovidas por diferentes vírus (LLAVE, 2010).

Recentemente foi elucidado o papel das DRBs no silenciamento por RNA. DCL1 interage com DRB1 para o processamento preciso e eficiente dos microRNAs (miRNA) primários em miRNAs maduros (KURIHARA et al., 2006). De maneira similar, DCL4 interage com DRB4 para a correta biogênese dos siRNAs *trans*-atuantes (NAKAZAWA et al., 2007). A DRB1 possui localização nuclear e estudos com construções que possuem a região promotora do gene da DRB1 fusionado ao gene da β-glucoronidase (GUS) indicam que esta proteína é constitutivamente expressa em todos os tecidos de *Arabidopsis,* corroborando com a sua importância na biogênese dos miRNAs (CURTIN et al., 2008). A DRB4 é expressa no tecido vascular, nas raízes, no meristema apical e em anteras em desenvolvimento (EAMENS et al., 2011). Além do processamento de ta-siRNAs, a DRB4, em conjunto com a DCL4, é necessária para o processamento dos RNAs de dupla fita virais durante a resposta antiviral e, além disso, parece inibir o acúmulo da proteína da capa viral através de uma diminuição da sua tradução ou da sua estabilidade (JAKUBIEC et al., 2012).

A DRB2 se localiza no núcleo e estudos utilizando construções que possuem seu

promotor fusionado à GUS demonstraram a sua expressão em pólen, semente madura, semente em germinação, em flores logo após a fertilização e no meristema apical (CURTIN et al., 2008). Recentemente foi demonstrado que DRB2 atua tanto sinergisticamente quanto antagonisticamente à DRB1 na biogênese de diferentes miRNAs no meristema apical (EAMENS et al., 2012).

Ao contrário da DRB1, DRB2 e DRB4, as proteínas DRB3 e DRB5 se localizam no citoplasma. Estudos utilizando construções do promotor do gene DRB3 fusionado ao gene GUS, mostraram que essa proteína é expressa no meristema apical e tecidos adjacentes, estendendo-se até o pecíolo das folhas em plântulas, mas em plantas adultas a expressão ocorre na maioria dos tecidos, porém em níveis mais baixos (CURTIN et al., 2008). Estudos similares demonstraram que a DRB5 também é expressa no meristema apical e no pecíolo das folhas em plântulas, mas em plantas adultas a expressão não é verificada em outros tecidos. Adicionalmente, a DRB5 também é expressa em botões de flores imaturas (CURTIN et al., 2008).

Um estudo recente apontou que DRB3 atua na defesa antiviral contra geminivírus pela via de metilação de DNA direcionada por RNA (RdDM, do inglês "RNA-directed DNA methylation", RAJA et al., 2014). Além disso, as DRB3 e DRB5 parecem atuar na mesma rota de silenciamento via miRNAs de DRB2 (EAMENS et al., 2012). Curiosamente, essa rota alternativa de silenciamento via miRNAs parece ser independente da clivagem do mRNA alvo, tornando esses genes fortes candidatos na mediação do silenciamento via bloqueio da tradução, um mecanismo ainda não muito bem compreendido em plantas (EAMENS et al., 2012).

Apesar dos vários avanços no entendimento do mecanismo de RNAi mediado por vírus, a biogênese dos vsiRNAs ainda não está bem compreendida. Moléculas de dupla fita de RNA são conhecidas por serem intermediários na replicação de vírus de RNA e, então, foram propostas como a principal fonte de geração dos vsiRNAs primários. Além disso, transcritos aberrantes de RNA são reconhecidos por RDRs e transformados em dsRNAs. Entretanto estudos demonstraram que a maioria dos vsiRNAs encontrados em infecções virais possuem orientação positiva (QI et al., 2009), o que não é o esperado para siRNAs originários de dsRNAs e clivados pela proteína DCL, que gera siRNAs nos dois sentidos. Isso levou a proposta de vsiRNAs serem gerados também por meio de estruturas secundárias do RNA viral que formariam regiões de RNA de fita dupla (MOLNÁR et al.,

2005). Entretanto, estruturas secundárias em dupla fita dos RNAs virais não parecem coincidir com a geração de vsiRNAs (QI et al., 2009).

Uma característica importante dos vsiRNAs que vem sendo elucidada é a de serem capazes de inibir a expressão de genes endógenos da planta. Essa capacidade já foi demonstrada em infecções do vírus do mosaico da couve-flor (CaMV) e do fumo (TMV) em *Arabidopsis thaliana* (MOISSIARD & VOINNET, 2006; QI et al., 2009).

1.3 Transferência horizontal de material genético

A transferência horizontal de material genético se caracteriza pela incorporação de moléculas de DNA ou até mesmo de genomas inteiros (DUNNING HOTOPP et al., 2007) entre espécies diferentes e de maneira não relacionada à reprodução (vegetativa ou sexual). Esse evento por muito tempo foi tido como raro e restrito a alguns tipos de organismos (KONDO et al., 2002). No entanto, a transferência genética horizontal tem sido descrita como uma das principais forças evolutivas em organismos procarióticos (HEUER; SMALLA, 2007). Por meio desse mecanismo, bactérias podem reorganizar o seu genoma e adquirir diversas características como resistência a antibióticos, patogenicidade e capacidade de metabolizar novas substâncias (HEUER & SMALLA, 2007).

A transferência de material genético também pode ocorrer entre procariotos e eucariotos, sendo este um evento mais raro. Esse fenômeno foi evidenciado em eucariotos unicelulares que se alimentam de bactérias, tais como *Entamoeba* e *Trichomonas* (ANDERSSON et al., 2001; KOONIN et al., 2000). A bactéria *Wolbachia* é um endossimbionte de uma ampla gama de insetos e é associada intimamente à linhagem celular germinativa de seus hospedeiros, facilitando que processos de transferência horizontal de material genético possam ser levados a geração seguinte. De fato, genes dessa bactéria e até o seu genoma inteiro foram encontrados em mais de uma espécie de inseto (DUNNING HOTOPP et al., 2007; KONDO et al., 2002). A transferência horizontal de genes também pode ocorrer entre eucariotos de diferentes espécies através de vetores virais ou elementos transponíveis como já foi descrito em *Drosophila* (DANIELS et al., 1990).

A aquisição de material genético proveniente de vírus em genomas eucarióticos também é um exemplo bem estudado de transferência horizontal. Já é bem conhecida a

capacidade de retrovírus em se inserir no genoma de seus hospedeiros e no caso de invadirem as células germinativas poderem ser transmitidos para as gerações futuras (STOYE, 2001). Com o tempo esses vírus perdem a capacidade replicativa e se tornam retrovírus endógenos. Esse fenômeno parece ter ocorrido em todos os mamíferos e explica a constatação de 8% do genoma humano ter uma provável origem retroviral (BOCK; STOYE, 2000). Além disso, em vertebrados, vírus de DNA também podem se integrar no genoma e potencialmente têm a capacidade de se manter na prole do hospedeiro (MORISSETTE & FLAMAND, 2010).

Por muito tempo, virologistas acreditavam que vírus de plantas não tinham a capacidade de se integrar no genoma de seus hospedeiros como suas contrapartes animais e bacterianas, contudo, estudos recentes evidenciaram a presença de sequencias virais no DNA vegetal (HARPER et al., 2002). Ao menos três vírus de DNA da família *Caulimoviridae*, que contém todos os pararetrovírus conhecidos de plantas, foram encontrados integrados ao genoma de seus hospedeiros e parecem ser ativos e capazes de gerar infecções episssomais (infecções mediadas por moléculas de DNA não integradas no genoma do hospedeiro) (HARPER et al., 2002). Pararetrovírus são vírus de DNA que transcrevem RNAs com a capacidade de codificar as proteínas virais e polimerizar uma nova molécula de DNA de dupla fita através da ação de uma transcriptase reversa (HARPER et al., 2002). Em tabaco foram encontradas sequências de uma espécie de vírus de DNA pertencente a família *Geminiviridae* mas, ao contrário dos outros exemplos, essas integrações não possuem capacidade de infecção (BEJARANO et al., 1996). Apesar de ainda não ter sido identificado nenhuma espécie de retrovírus em plantas existem, ao menos, duas classes de retrotransposons (elementos genéticos móveis que transcrevem RNA e que codificam as enzimas integrase e transcriptase reversa, sendo assim capazes de se integrar ao genoma de seus hospedeiros) que são capazes de codificar uma proteína similar a ENVELOPE (ENV) de vírus que sintetiza o envelope viral (ZAKI, 2003). Esses retrotransposons possuem potencial para se replicar e gerar uma infecção de maneira similar aos retrovírus, contudo isso ainda não foi verificado (MARCO & MARÍN, 2005).

Apesar de vírus de RNA não retrovirais serem incapazes de codificar a proteína integrase, trabalhos recentes identificaram vírus de RNA integrados no genoma de mamíferos, fungos, insetos e plantas (CHIBA et al., 2011; KOONIN, 2010). Aparentemente a integração nesses casos só foi possível devido a uma recombinação do

RNA viral com o RNA de retrotransposons (GEUKING et al., 2009; HORIE et al., 2010; TAYLOR et al., 2010).

1.4 *Nicotiana benthamiana*, *Potato virus X* e *Cucumber mosaic virus* como organismos modelo para o estudo da resposta antiviral em plantas

Nicotiana benthamiana é a planta mais amplamente utilizada como hospedeira para estudos de virologia vegetal, devido principalmente ao grande número de vírus que podem infectá-la com sucesso (GOODIN et al., 2008). Pertencente a família *Solanaceae*, é uma planta herbácea endêmica da Austrália (GOODIN et al., 2008). Porém, *N. benthamiana* se tornou mais amplamente utilizada em virologia de plantas somente após o desenvolvimento de três grandes avanços na biologia molecular.

O primeiro foi o desenvolvimento de tecnologia capaz de expressar transgenes através de vetores virais. Essa técnica possibilita monitorar o movimento viral *in vivo* revelando aspectos fundamentais da biologia vegetal como o movimento através do plasmodesma e o deslocamento macromolecular intercelular, permitindo também revelar quais as proteínas responsáveis por esses processos (CHAPMAN et al., 1992; ESCOBAR et al., 2003).

O desenvolvimento de vetores virais levou ao desenvolvimento da tecnologia conhecida como silenciamento gênico induzido por vírus (VIGS, do inglês “virus-induced gene silencing”) (KUMAGAI et al., 1995). Essa técnica permite a regulação negativa de praticamente qualquer gene de interesse em plantas (DING & VOINNET, 2007). Uma das grandes vantagens dessa técnica é a capacidade de reduzir o efeito de redundância gênica se o cDNA clonado no vetor viral tiver homologia com mais de um membro de uma família multigênica. Ao mesmo tempo, ao selecionar cuidadosamente a sequência de nucleotídeos utilizada na construção do vetor, um determinado gene pode ser silenciado individualmente (BURCH-SMITH et al., 2004).

Por fim, o desenvolvimento da técnica de agroinfiltração permitiu que genes de interesse, geralmente fusionados a proteínas fluorescentes, possam ser expressos transientemente em células vegetais de uma maneira simples e direta, sendo particularmente útil a experimentos em grande escala (KUNZ & SCHÖB, 1997). Curiosamente, tanto VIGS quanto agroinfiltração funcionam muito bem em *N. benthamiana*, mas insuficientemente em outras plantas como *A. thaliana* (GOODIN et al.,

2008). Recentemente, o genoma e transcriptoma de *N. benthamiana* foram divulgados, tornando o uso dessa planta para estudos de biologia molecular vegetal ainda mais atraente (NAIM et al., 2012; NAKASUGI et al., 2014).

O *Potato virus X* (PVX) é um vírus de planta de importância econômica amplamente utilizado como vetor de expressão e silenciamento servindo como sistema modelo para o estudo de silenciamento gênico, resposta imune e movimentação viral (TILSNER et al., 2013; VERCHOT-LUBICZ et al., 2007). O PVX que pertence à família *Alphaflexiviridae* e ao genêro *Potexvirus* possui um único genoma de +ssRNA que codifica cinco "open reading frames" (ORFs). A primeira ORF codifica a replicase viral. A região central possui três ORFs sobrepostas conhecida como bloco triplo de genes (TGB, do inglês "triple-gene block") que codificam proteínas envolvidas na movimentação viral. E a última ORF codifica a proteína da capa (CP, do inglês "coat protein") necessária tanto para a montagem do virion quanto para o movimento intercelular (HUISMAN et al., 1988).

O PVX é o exemplo típico de vírus de plantas que codificam o TGB, portanto sendo importante para o estudo dos mecanismos de replicação viral e movimento intercelular (VERCHOT-LUBICZ et al., 2010). A proteína TGB1, codificada pela primeira das três ORFs centrais, tem funções de RNA helicase, supressão do silenciamento gênico e ativação da tradução (ATABEKOV et al., 2000; VOINNET et al., 2000). Interações específicas entre subunidades de CP e TGB1 na extremidade 5' de virions desestabiliza a capa viral, permitindo o acesso do genoma viral ao ribossomo, além de dilatar os poros do plasmodesma (ATABEKOV et al., 2000). Isso permite a formação dos VRCs no retículo endoplasmático. TGB2 então induz a formação de grânulos derivados do reticulo endoplasmático, onde se encontram os VRCs (BAMUNUSINGHE et al., 2009). Os VRCs irão encontrar os nanoporos do plasmodesma porque a superfície do retículo endoplasmático é altamente móvel ao longo do plasmodesma. Desta forma, TGB2 e TGB3 recrutam TGB1 para os nanoporos onde é inserida a proteína da capa em forma de complexo proteico junto com o genoma viral, finalizando o movimento intercelular do PVX (TILSNER et al., 2013).

O *Cucumber mosaic virus* (CMV) que pertence à família *Bromoviridae* é a espécie tipo do gênero *Cucumovirus*. É um dos mais comuns dentre os vírus de plantas com substancial importância econômica, podendo infectar mais de mil espécies diferentes de plantas distribuídas em mais de 100 famílias (VAN REGENMORTEL et al., 2000). O

genoma do CMV é composto basicamente por três RNAs de orientação positiva, designados RNA 1 (~3,350 nt), RNA 2 (~3,050 nt), RNA 3 (~2,200 nt) e dois RNAs subgenômicos (RNA 4 e RNA 4A) podendo conter ainda RNAs satélites (SIMON et al., 2004, ZITTER et al., 2009). O RNA 1 codifica a proteína 1A que é um componente do complexo da replicase viral (ZITTER et al., 2009). Além de ser essencial para a replicação esta proteína também está envolvida no movimento viral (EIRAS et al., 2004). O RNA 2 codifica duas proteínas: a proteína 2A, que faz parte do complexo da replicase viral e tem atividade de RNA polimerase RNA-dependente (utiliza um molde de RNA fita simples para sintetizar uma fita complementar) e a proteína 2B que é sintetizada a partir do RNA subgenômico 4A e está envolvida na propagação sistêmica do vírus e também em sua virulência, inibindo o silenciamento gênico pós-transcricional promovido pela planta hospedeira (DIAZ-PENDON et al., 2007). O RNA 3 também codifica duas proteínas: a proteína 3A e a proteína do capsídeo. A proteína 3A é a proteína responsável pelo movimento do vírus entre as células do hospedeiro já que ela aumenta o limite de exclusão por tamanho do plasmodesma (SU *et al*, 2010). Mutações nesta proteína revelam diferenças hospedeiro-específicas na eficiência do movimento viral (ZITTER et al., 2009). A proteína do capsídeo expressa através do RNA subgenômico 4 é a responsável pela encapsidação viral (ZITTER et al., 2009). Os RNAs satélites (satRNAs) são completamente dependentes do CMV para a sua replicação. Eles são RNAs de fita simples que variam de 330 a 405 nt e aparentemente não codificam proteínas (SIMON et al., 2004).

O CMV apresenta uma distribuição cosmopolita e é considerado um importante patógeno tanto nas regiões tropicais e subtropicais, quanto nas temperadas. Por causa de seu amplo espectro de hospedeiros, numerosas espécies de ervas daninhas contribuem para essa distribuição ampla do vírus ao servirem de reservatórios aonde o CMV se mantém até o início da temporada de plantação das culturas de importância econômica (ZITTER et al., 2009).

Os isolados de CMV são divididos em dois grupos (subgrupo I e subgrupo II) baseado em diferenças nos dados sorológicos, mapeamento de peptídeos, hibridização de ácidos nucleicos e similaridade na sequência de nucleotídeos (PALUKAITIS et al., 1992). O subgrupo I ainda pode ser dividido em IA e IB (ROOSSINCK et al., 1999).

Linhagens de CMV capazes de infectar a soja foram reportadas pela primeira vez no Japão (KOSHIMIZU & IIZUKA, 1958). A resposta da soja a infecções virais pelo

CMV é determinada pela combinação de isolados do vírus e cultivares da soja, sugerindo a existência de genes de resistência cultivares-específicos (TAKAHASHI et al., 1980). A infecção do CMV em soja geralmente provoca danos ao crescimento vegetativo da planta e a produção de vagens, mas também pode resultar em nenhum sintoma (TAKAHASHI et al.,1987). Apesar de no Brasil já ter sido reportada a capacidade do CMV em infectar várias plantas como bananeira, maracujá, abóbora, melão, pimentão, pimenta-do-reino, tomate, ervilha e agrião (BOARI et al., 2000 EIRAS et al., 2004), ainda não há indícios de infecção na soja por esse vírus.

2 OBJETIVO GERAL

Analisar possíveis novos mecanismos de defesa por silenciamento de RNA em plantas.

2.1 OBJETIVOS ESPECÍFICOS

- Realizar a montagem *de novo* do RNA 1 do CMV utilizando os dados do sequenciamento em larga escala de sRNAs e ferramentas de bioinformática para a geração dos “contigs”.

- Estudar o perfil de expressão dos sRNAs provenientes desse vírus.

- Demonstrar o mecanismo de integração do RNA 1 do CMV no genoma da soja.

- Buscar a presença dessa integração em outras cultivares.

- Analisar a localização celular das DRBs em folhas de *Nicotiana benthamiana* por microscopia confocal.

- Analisar a localização dessas proteínas frente a uma infecção viral por PVX e a sua interação com as proteínas virais.

CAPÍTULO I - UNUSUAL RNA PLANT VIRUS INTEGRATION IN THE SOYBEAN GENOME LEADS TO THE PRODUCTION OF SMALL RNAS

Autores: Guilherme Cordenonsi da Fonseca[1], Luiz Felipe Valter de Oliveira [2], Guilherme Loss de Morais [12], Ricardo Vilela Abdelnor[4], Alexandre Lima Nepomuceno[4], Peter M. Waterhouse[5], Laurent Farinelli[6], Rogerio Margis[12*]

[1] Centro de Biotecnologia, Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil.

[2] Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil.

[3] Laboratório Nacional de Computação Científica, LNCC, Petropolis, RJ, Brazil

[4] Empresa Brasileira de Pesquisa Agropecuária, EMBRAPA, Londrina, PR, Brazil.

[5] Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Queensland, Australia.

[6] Fasteris SA, Plan-les-Ouates, Switzerland.

Artigo submetido ao periódico BMC Genomics

Abstract:

Horizontal gene transfer (HGT) is known to be a major force in genome evolution. The acquisition of genes from viruses by eukaryotic genomes is a well-studied example of HGT, including rare cases of non-retroviral RNA virus integration. However, the function of these events remains elusive. This study describes the integration of an RNA of the *Cucumber mosaic virus* into the soybean genome. This discovery resulted from an initial metatranscriptomic analysis of small RNAs derived from soybean. *De novo* assembly of small RNAs result in a 3029-nt long contig with homology to the sequence of CMV RNA 1. The presence of this sequence in the genome of *Glycine max*, cultivar Conquista, and *Glycine soja* was confirmed by DNA deep sequencing. The locus where the integration occurred harbors the full CMV RNA 1 sequence followed by the partial sequence of an endogenous mRNA and another sequence of CMV RNA 1 in the opposite sense and orientation. This region was recombined into a retrotransposon located inside an exon of a soybean gene. SmallRNAseq data strongly suggest the formation of a hairpin structure upon hybridization of the complementarity regions of both CMV RNA1 sequences. The nucleotide similarity of the integrated sequence compared to other CMV sequences indicates that the integration event occurred recently. We described a rare event of non-retroviral RNA virus integration in soybean which leads to the production of a double-stranded RNA in a similar fashion to virus resistant transgenic plants.

Keywords: Horizontal gene transfer; RNA plant virus; *Cucumber mosaic virus*; soybean; transposon

Introduction:

Horizontal gene transfer (HGT) is characterized by the exchange of DNA between

unrelated species by means other than sexual reproduction. For many years, HGT events were considered to be rare and restricted to a few types of organisms. More recently, however, the increased availability of genomic data has revealed this phenomenon to be an important force in prokaryotic evolution (DOOLITTLE et al., 2003; HEUER & SMALLA, 2007). Through HGT, different species of bacteria are able to rapidly adapt to new environments by acquiring novel metabolic features, antibiotic resistance and pathogenicity (ARAVIND et al., 1998; HACKER et al., 1997; OCHMAN et al., 2000; WOLF et al., 1999). On the other hand, HGT has long been assumed to play a limited role in the evolution of eukaryotic genomes. One of the reasons for this underestimation was the lack of sequence data from eukaryotes, and a misinterpretation of the number of genes transferred from bacteria to the human genome also helped to undermine the importance of this process in eukaryotic organisms (LANDER et al., 2001; SALZBERG et al., 2001; STANHOPE et al., 2001). Nevertheless, with the increasing number of genomes sequenced in the past few years, many cases of HGT have been reported in eukaryotes (ANDERSSON, 2005; KEELING & PALMER, 2008); even the nearly complete genome of *Wolbachia* was found in the chromosome of its host insect (DUNNING HOTOPP et al., 2007). As with bacteria, HGT events in eukaryotes can provide important adaptive traits such as virulence factors, new metabolic pathways, resistance to abiotic stress and the capability to process new types of nutrients (ACUÑA et al., 2012; ANDERSSON et al., 2007; DANCHIN et al., 2010; EICHINGER et al., 2005; FRIESEN et al., 2006; GRAHAM et al., 2008; LOFTUS et al., 2005; RICARD et al., 2006). Interestingly, in the asexual metazoa bdelloid rotifer, the high ratio of genes acquired from bacteria and fungi may have partially compensated for the lack of genetic variation that sexual reproduction provides (BOSCHETTI et al., 2011; GLADYSHEV; MESELSON; ARKHIPOVA, 2008).

The transfer of genetic material between viruses and eukaryotes has also been documented in past years. Retroviral sequences were found in the chromosomes of vertebrates, and although their functions remain unclear, these endogenous retroviruses comprise a significant fraction of vertebrate genomes (BOCK & STOYE, 2000; STOYE, 2001; STOYE, 2012). For a long time, plant viruses were considered to be unable of integrating into the host genome, as they did not possess integrase activity; however, sequences from pararetroviruses and DNA viruses has been encountered in plant genomes in recent years (BEJARANO et al., 1996; BERTSCH et al., 2009; HARPER et al., 2002). Surprisingly, even the transfer of nonretroviral RNA viruses to plants, mammals and fungi genomes were reported recently (CHIBA et al., 2011; GEUKING et al., 2009; HORIE et al., 2010; TAYLOR & BRUENN, 2009; TAYLOR et al., 2010). The function of these nonretroviral integrated RNA viruses (NIRVs) remains elusive; however, some authors have suggested a possible role in antiviral immunity (KOONIN, 2010; LIU et al., 2010).

Cucumber mosaic virus (CMV) is one of the most common among the RNA viruses of substantial economic importance, as it can infect more than 1,000 different species of plants (PALUKAITIS & GARCÍA-ARENAL, 2003). The CMV genome is composed of three single-stranded, positive-sense RNAs. RNA 1 encodes a protein that is required for viral replication, RNA 2 encodes the viral RNA polymerase and a protein that is able to suppress host post-transcriptional gene silencing (PTGS) mechanisms, and RNA 3 encodes the two proteins required for viral movement (ROOSSINCK, 2001). Recent studies have reported the sequencing and identification of viruses by metatranscriptome analyses of deep sequence small RNA (sRNA) libraries (KREUZE et al., 2009; WU et al., 2010). Using a similar approach, our group identified several sequences from pathogenic, symbiotic and free-living organisms in soybean sRNA libraries (MOLINA et al., 2012).

Interestingly, we also identified sRNAs corresponding to RNA 1 of CMV in all of the libraries surveyed but not to RNA 2 or RNA 3. In this work, we demonstrated that these sRNAs originate from a CMV RNA 1 sequence that previously integrated into the soybean genome.

Material and methods

Plant material and nucleic acid isolation

The small RNA sequences used in this work were obtained from different soybean cultivars used for the generation of deep sequencing libraries related to projects of the GenoSoja consortium (BENKO-ISEPPON et al., 2012).

Leaf and root samples were collected from Conquista cultivar plants grown in a hydroponic system under an aerated 6.6 pH-balanced nutrient solution. First, seeds were pre-germinated on moistened filter paper in a greenhouse with a temperature of 24 ± 2 ºC in the dark for one day and then under light with a 12:12 h photoperiod for 2 days. The plantlets were placed in polystyrene supports, and the roots were immersed in nutrient solution for 15 days until the first trifoliate leaf was fully developed (V2 developmental stage). The leaves and roots were extracted, frozen in liquid nitrogen, and stored at -80 ºC for subsequent RNA extraction.

Seeds of the Conquista cultivar were grown in moistened filter paper for 0, 3, 5 and 7 days in a greenhouse under a temperature regime of 24 ± 2 ºC. At the end of the respective time periods, the samples were frozen in liquid nitrogen and stored at -80 ºC.

Pods of the same cultivar (developmental stage R3 to R5) were collected from field plants grown at the Federal University of Rio Grande do Sul (UFRGS) in Porto Alegre, Brazil.

Flowers samples (developmental stage R2-R3) from the cultivar Urano were collected

from plants grown at the experimental field of the University of Passo Fundo (UPF), Brazil. Whole flowers, stamens, carpels and petals samples were immediately powdered in Trizol (Invitrogen, CA, USA), separately, and stored at -80 ºC for subsequent RNA extraction.

The frozen samples were ground to a fine powder in liquid nitrogen, and total RNA was isolated using Trizol (Invitrogen, CA, USA), following the manufacturer's instructions. RNA quality was evaluated by 1.0% agarose gel electrophoresis, and the amount of RNA was assessed using a Quibit fluorometer and the Quanti-iT RNA assay kit (Invitrogen, CA, USA), following the manufacturer's instructions.

Seeds from 12 different soybean cultivars (*Conquista, Urano, BR 16, Embrapa 48, Williams, Davis, Hills, Paraná, BR 4, IAS 4, Força and Potência*) were incubated in moistened filter paper for a week at 24 ± 2 ºC in a growth chamber. Leaves from the plantlets were collected and ground in liquid nitrogen. DNA was extracted using the PureLink ™ (Invitrogen, CA, USA) kit, according to the manufacturer's instructions. The DNA samples were used in posterior PCR analysis.

DNA, RNA and small RNA sequencing

Approximately 10 μg of total RNA from each sample was sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for processing and sequencing using Solexa technology. Each library was constructed with a pool of 4 to 5 RNA samples from different plants. The “germinating seed” library was constructed from a pool of RNA samples from seeds at 3, 5 and 7 days post incubation in moistened filter paper under 24 ± 2 ºC, as described above. The construction of small RNA libraries from leaf and root samples of the cultivars *BR-16*, *Embrapa-48* and *PI 561356* was described previously (KULCHESKI et

al., 2011). The sequencing of the small RNA samples from seeds, germinating seeds, pods, flowers, carpels, petals and stamens was performed on the Illumina Genome Analyzer GAII, and for the sequencing of the small RNAs, mRNA and DNA from *Conquista* leaves, the Illumina HiSeq 2000 Sequencing System was used.

Briefly, the process for constructing the small RNA libraries consisted of the following: acrylamide gel purification of the RNA bands corresponding to the size range from 20 to 30 nt, the ligation of 3' and 5' adapters to the RNA in two separate subsequent steps, cDNA synthesis followed by acrylamide gel purification and PCR amplification to generate a DNA colony template library for sequencing. After the trimming of the sequences, reads from 19 to 24 nt were used for further analysis.

The mRNAseq library was construct after the Poly(A) purification of the total RNA extract followed by cDNA synthesis using Poli(T) primer. The paired-end 50 bases sequencing was perfomed in inserts of 500 nt.

De novo assembly of small RNAs

The 19- to 24-nt reads of the *Conquista* leaves small RNA library were assembled into contigs using the Velvet 0.7.3.1 *de novo* assembly algorithm (ZERBINO; BIRNEY, 2008) with the following parameters: a harsh length of 17 to 23, a coverage cut-off of 50, an expected coverage of 1,000 and a minimum scaffold length of 100. Only the contigs that matched the CMV RNA 1 sequence were used to assemble the entire sequence in the Molecular Evolutionary Genetics Analysis (MEGA version 5.0) (TAMURA et al., 2011) software using the RNA 1 sequence of CMV isolate CM95 [GenBank: AB188234.1] as a reference.

Mapping analysis

All mapping analyses in this study were performed using the Bowtie 2 tool (LANGMEAD & SALZBERG, 2012). The small RNA reads were mapped against the reference sequence using the standard parameters, except that no mismatches were allowed (-n 0). For the RNA and DNA reads analysis, mapping was performed using the standard paired-end options, adjusting the minimum insert length to 0 (-I 0) and the maximum insert length to 1,000 (-X 1000).

Determination of the entire integrated sequence

To identify the flanking regions of the inserted sequence, the *Conquista* DNA sequencing reads were first mapped, single-end and without allowing mismatches (-n 0), against the CMV reference sequence assemble from the small RNAs contigs described above using the Bowtie 2 tool. Then, pairs of the reads that mapped in the reverse orientation to the 5' end of the sequence and in the sense orientation to the 3' end were used to identify and complement the flanking regions of the sequence. Paired-end Mapping analysis was performed periodically using the standard parameters, to confirm the new sequence. After we identified the integration site, primers were designed to confirm the data by PCR (Figure 1 and Table S1). The sequence was deposited in NCBI GenBank (Acession number: KP137860).

PCR analysis

PCR was performed using ~10 ng of DNA, 10 μM of each oligonucleotide, 5 μM of each dNTP, 2 μl of 10x buffer (100 mM Tris-HCl, pH 8,0), 50 mM $MgCl_2$, 0.5 U of Taq DNA polymerase (Invitrogen, CA, USA) and deionized water up to 20 μl. The PCR conditions were as follows: 5' at 95 ºC; 40 cycles of 30" at 95 ºC, 30" at 60 ºC and 30-180"

(varying according to the amplicon size) at 72 ºC; and a final extension of 5' at 72 ºC. For each reaction, 5 µl was collected and separated on a 1% agarose gel. The samples were stained with SYBR-Gold (Invitrogen, CA, USA), following the manufacturer's instructions.

To further confirm the sequence of the amplicons, each PCR product was sequenced using the Genetic Analyzer ABI Prism 3100 (Applied Biosystems). Approximately 50 ng of each product was add to 5 pmol of the same oligonucleotides used for PCR, along with 2 µl of reaction buffer, 1 µl of BigDye (Applied Biosystems) and deionized water up to 10 µl. Sequencing was performed following the dideoxy chain-termination method using BigDye (Applied Biosystems) according to the manufacturer's instructions.

Phylogenetic analysis

The CMV RNA1 sequences were obtained from the NCBI database (Table S2). The software MEGA (version 5.0) was used to translate the ORFs to protein sequences. The MUSCLE algorithm (EDGAR, 2004) was used to align the sequences. The multiple sequence alignments were manually inspected and edited, and only unambiguously aligned positions were included in the final analysis. The phylogenetic analysis was performed after protein sequence alignments using the Bayesian method, carried out in BEAST1.7 (DRUMMOND & RAMBAUT, 2007). The model of protein evolution used in this analysis was the Dayhoff model for protein matrix substitution. The Yule tree was selected as a tree prior to the Bayesian analysis, and 20,000,000 generations were performed using Markov chain Monte Carlo (MCMC) algorithms. The trees were visualized and edited using FigTree v1.3.1 software (http://tree.bio.ed.ac.uk/software/figtree/).

Results and Discussion

CMV sRNA profile in soybean libraries.

In this study, 11 sRNA libraries were analyzed from different soybean samples (leaf, root, flower, stamen, carpel, petal, pod, seed and germinating seed). The sRNA sequences (ranging from 19 nt to 24 nt) from each library were assembled into contigs using the Velvet program (ZERBINO; BIRNEY, 2008). For all libraries, sequences of 3,029 nt that shared homology to CMV RNA 1 were obtained after the *de novo* assembly of these contigs. This sequence does not cover the entire length of CMV RNA 1, which is usually approximately 3350 nt (ROOSSINCK, 2001). A BLASTn search in the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/) identified that the RNA 1 sequence of CMV isolate CM95 [GenBank: AB188234.1] has the highest similarity to the sequence obtained from the sRNA libraries (nucleotide identity: 93%). The contig sequence covers the entire ORF region of isolate CM95 RNA 1 plus 26 nt of the 5' end and 23 nt of the 3' end. Plants infected by viruses process the viral double-stranded RNAs (dsRNAs) into sRNAs, denominated viral small RNAs (vsRNAs), as a defense mechanism (SHARMA et al., 2012). The biogenesis of vsRNAs occurs by several pathways. The Dicer-like enzyme (DCL in plants) could process the imperfectly base-paired secondary structures of the viral RNAs into primary vsRNAs. These vsRNAs couple with the RNA-induced silence complex (RISC) to promote the cleavage of viral RNAs. Cleaved RNAs are processed into dsRNA by an endogenous RNA-dependent RNA polymerase (RdRp). This dsRNA is cleaved into secondary vsRNAs by the action of a DCL, thus amplifying the RNA silencing signal. Another source for the production of vsRNAs could be the dsRNAs generated during viral replication (SHARMA et al., 2012). Normally the vsRNAs are 20-24 nt length, where the 21 nt is the most abundant class

followed by the 22 nt (DONAIRE et al., 2009). The predominant class of sRNA sequences from CMV RNA 1 was the 22-nt class, which comprised 58% to 90% of all sRNA sequences mapped in each library (Figure 2). sRNAs of 24 nt in length were the second most abundant class, and sRNAs of 21 nt accounted for less than 5% in all libraries surveyed, which is uncommon for viral infections of plants (DONAIRE et al., 2009). In addition, no sRNA sequences were found with homology to CMV RNA 2 or RNA 3. Since, all the libraries originated from samples of soybean cultivars that were growth in different regions, a infection of the same virus seems unlikely. Hence, the CMV sRNAs were assumed to have derived from a CMV RNA 1 sequence that had integrated into the plant genome. To evaluate this hypothesis, primers were designed against the RNA 1 sequence (Figure 1A). PCR reactions confirmed the presence of the CMV RNA 1 sequence in the genomes of all soybean cultivars tested, except for *Williams* (Figure 1B), which is the reference genome of the species (SCHMUTZ et al., 2010). This finding raised two main questions: how was CMV RNA 1 inserted into the soybean genome, and why did this insertion lead the production of sRNAs? For the former question, one possible explanation is that during an infection, the CMV RNA 1 recombined with an endogenous retrotransposon and was then integrated into the plant genome. Indeed, this appears to be the main pathway for the integration of NIRVs (GEUKING et al., 2009; HORIE et al., 2010; LIU et al., 2010). There are some pathways that promote the generation of sRNAs, and these could be used to explain the presence of CMV RNA 1 sRNAs in the soybean cultivars surveyed. Transposable elements (TEs) are inactivated in plants by through RNA-dependent DNA methylation (RdDM), which involves proteins of the RNA interference machinery and leads to the production of sRNAs derived from TEs (PARENT; MARTÍNEZ DE ALBA; VAUCHERET, 2012). If in fact the CMV RNA 1 recombined

with a retrotransposon, it is possible that an RNA containing both sequences triggered the RdDM process. Another possibility is that two CMV RNA 1 sequences were integrated into the plant genome, with one copy being transcribed in the antisense orientation and thus promoting the formation of a dsRNA sequence of CMV RNA 1, thereby leading to the production of sRNAs.

Characterization of the CMV RNA 1 integration into soybean.

To analyze the region where the CMV RNA 1 was integrated, we sequenced the genome of the soybean cultivar *Conquista* using an Illumina Genome Analyzer GAII. Based on these data, the integration event was found to have occurred after a series of recombinations involving two strands of CMV RNA 1, the mRNA of an endogenous gene and the RNA of an endogenous retrotransposon (Figure 3A). The two CMV RNA 1 sequences were in the opposite sense and direction with respect to each other and were flanking a 522-nt sequence that was homologous to the mRNA of a gene encoding the nonclathrin coat protein zeta2-COP or *COPZ2* (Figure 4A). Using the data from Phytozome version 9.0 (http://www.phytozome.net), two homologs to this gene were found in the soybean genome, corresponding to the *loci* Glyma15g13190 and Glyma09g02270; the former showed higher similarity with only 4 mismatches to the sequenced *COPZ2* mRNA sequence (hereafter referred to as lCOPZ2). The CMV RNA 1 sequence upstream of this sequence (upCMV1) contained 3301 nt and showed 93% identity with the CMV RNA 1 of isolate Va (GenBank: JX014246.1, nucleotides 55 to 3360), indicating a deletion of approximately 50 nt at the 5' end of the sequence. The homology breakpoint at the 5' end of the lCOPZ2 was in the fifth exon of the *COPZ2* transcript (Glyma15g13190.1). Interestingly, the remaining nucleotides of this exon shared a short homology of 11 nt to

the 3' end of the upCMV1 sequence, which is in the opposite orientation to lCOPZ2 (Figure 4B). This recombination possibly occurred by the replicase-mediated copy-choice mechanism (SZTUBA-SOLIŃSKA et al., 2011), whereas the hybridization between the viral RNA and the mRNA of *COPZ2* in that region could have facilitated the recombination by forming a heteroduplex with a nascent RNA strand polymerized by the viral RdRp (Figure 4B). As for the 3' end of lCOPZ2 sequence, the homology ended in the 5' untranslated region (UTR) and preceded a 10-nt AT-rich sequence (5'-AAATTTTCCA-3') with no homology to either CMV RNA 1 or the *COPZ* transcript. Thus, this sequence could be a byproduct of the recombination between the 5' UTR of the *COPZ2* transcript and the 3' UTR of the CMV RNA 1 that was localized downstream of the lCOPZ2 (downCMV1), which is in the same orientation as this sequence (Figure 3A). The mechanism behind this recombination remains elusive. The downCMV1 sequence (3127 nt) shares 92% nucleotide similarity with the CMV RNA 1 of isolate Va (GenBank: JX014246.1, nucleotides 25 to 3143), indicating a deletion of less than 30 nt in the 5' end and a deletion of approximately 200 nt in the 3' end of the original viral RNA during recombination.

The 3'UTR of all CMV RNAs contains a region that forms a well-conserved tRNA-like structures in its last 180 nt, which is responsible for the initiation of the minus-strand synthesis (RIZZO & PALUKAITIS, 1989). Secondary structures can favor the template switching of the viral replicase during homologous or nonhomologous RNA recombination, as they can stop the progression of the RdRp replication (FIGLEROWICZ, 2000; NAGY & SIMON, 1997). Thus, this deletion could have occurred due to a non-canonical recombination between the region before the tRNA-like structure of the CMV RNA1 and the 5' UTR of the *COPZ* mRNA. Given that UTRs usually form secondary

structures, we proposed that these two events occurred during the replication of the CMV RNA1 plus strand. First, a template-switching event occurred between the CMV RNA1 minus strand and the *COPZ2* mRNA, allowing the elongation of this nascent RNA using the sequence of the *COPZ2* mRNA as the template (Figure 4B). Second, during the elongation of this chimeric RNA, the viral RdRp stopped in a region of secondary structures of the 5'UTR of the *COPZ2* mRNA and then jumped to a CMV RNA1 plus strand, approximately 200 nt away from the 3' end, after the tRNA-like region (Figure 4C). Thus, the viral replicase continued to elongate the RNA using the CMV RNA1 as template. The whole region (upCMV, loopCOPZ and downCMV) is flanked by the sequences of an endogenous retrotransposon (Figure 3A). Using the data from SoyTEdb, a database of transposable elements in the soybean genome (http://www.soybase.org/soytedb/) (DU et al., 2010), it was possible to determine that this retrotransposon is a Copia LTR (long terminal repeat) named Gmr71. Single copies of Gmr71 are present in the chromosomes 18, 19 and 20 of the *Williams* cultivar genome. The sequence upstream of the upCMV up until the LTR contains 1892 nt, of which 1812 nt share complete similarity with the copies from chromosomes 18 and 20 and contain a single mismatch in the LTR region from the copy of chromosome 19. The downstream region from downCMV to the LTR is 2976 nt long, with 2844 nt sharing similarity with the copies from chromosomes 18 and 20, with 3 mismatches, and with the copy from chromosome 19, with 4 mismatches (one in the LTR region).

The LTR regions are flanked by the sequence of the locus Glyma10g28270, thus indicating that the whole region was integrated into the soybean genome by a retrotransposition event (Figure 3A). We speculate that during CMV replication, recombination occurred between two CMV RNA1 sequences and the mRNA from *COPZ2* and that this event was followed

by a recombination event between this newly formed RNA and the RNA from the LTR Copia retrotransposon, which was integrated into the genome. Seven PCRs were performed to confirm these recombinations (Figure 3D), and the products of these reactions were sequenced. The total length of this insertion, from one LTR to the other, was 11,827 nt (Figure 3A). To assess if this integration is homozygous in the *Conquista* cultivar, a PCR was performed using primers flanking the integration site. An expected product was amplified in the reaction using *Williams* DNA but not in the reaction using *Conquista* DNA (Figure 3E). Because the DNA polymerase used to perform the PCRs cannot synthesize fragments over 10,000 nt, the absence of the same amplicon found in the *Williams* cultivar means that both copies of the Glyma10g28270 *locus* in the *Conquista* genome contain the insertion.

Expression of a hairpin-like RNA in the region of the integrated sequence.

To understand how the production of sRNAs was possible in the integrated sequence, we sequenced samples of sRNAs and mRNAs from leaf tissues of 2-week-old *Conquista* plants using the Illumina HiSeq 2000 system. Small RNAs ranging from 19 to 24 nt were mapped across the integrated sequence. A total of 23,437 reads were mapped, in which the most abundant class was 22-nt sRNAs, comprising 70% of all sRNAs, followed by the 21-nt (13%) and 24-nt classes (8%). Interestingly, the majority of the sRNAs covered two regions (Accession number KP137860, nucleotides 1918 to 4987 and 5733 to 8820) corresponding to sequences that shared homology between upCMV1 and downCMV1 (Figure 3B). This result indicates that the RNA synthesized in this region potentially forms a hairpin-like structure similar to those found in miRNAs, where the lCOPZ sequence acts as a loop and the sequences from upCMV and downCMV hybridize

to form the stem. Inverted repeats that trigger the production of sRNAs are not uncommon in plants. Usually, those sRNAs are of the 24-nt class and are correlated to transposons and other repetitive elements in the genome in a process involving RNA-directed DNA methylation (RdDM) (KASSCHAU et al., 2007; ZHANG; ZHU, 2011). This pathway is important for the stable epigenetic repression of transgenes, specific endogenous genes and repetitive elements across the genome (ZHANG; ZHU, 2011). The biosynthesis of 24-nt sRNAs mainly requires the proteins DICER-LIKE 3 (DCL3) RNA-DEPENDENT, RNA POLYMERASE 2 (RDR2) and RNA POLYMERASE IV (Pol IV) (ZHANG; ZHU, 2011). However, some inverted repeat loci in *Arabidopsis* trigger the production of sRNAs that are not necessarily involved in the RdDM pathway. Some of these regions resemble young miRNAs (FAHLGREN et al., 2007), whereas others are similar to transgenes constructed to induce PTGS in endogenous genes because they have inverted repeats with a fold-back structure (LINDOW et al., 2007). Concerning the latter case, the *IR71 locus* of *Arabidopsis* shares a similar structure to the integrated sequence described in the present work, although it did not match viral sequences (HENDERSON et al., 2006). This locus is a large inverted repeat sequence with repeat arms being ~3.23kb in length. It seems that DCL2 is the major responsible for the accumulation of sRNAs derived from *IR-71*, since *dcl2* mutants showed no production of 22-nt sRNAs derived from this locus (HENDERSON et al., 2006). In *Arabidopsis*, DCL2 and DCL3 are the main proteins responsible for the production of 22- and 24-nt sRNAs, respectively (HENDERSON et al., 2006); thus, the orthologous genes in soybean are likely to act in the synthesis of the CMV-derived sRNAs. Finally, plant transgenic lines producing virus-derived hairpin RNAs can induce RNA silencing of viruses (BONFIM et al., 2007; WANG et al., 2000). In fact, several works demonstrate that transgenic plants expressing hairpins derived from the CMV viral

replicase or from the CMV coat protein can provide strong resistance when challenged with the virus (CHEN et al., 2004; KALANTIDIS et al., 2002; NTUI et al., 2013). Additionally, tobacco plants expressing CMV RNA1 were able to reduce viral accumulation and prevent systemic infection (CANTO; PALUKAITIS, 2001). Thus, soybean cultivars presenting the sequence described in this work are likely to possess some degree of defense against CMV.

To assess if the RNA that triggers the production of the CMV sRNAs is transcribed from the promotor of the locus Glyma10g28270 locus or from the Copia-like retrotransposon, we performed deep sequencing of mRNAs purified from the leaves of the *Conquista* cultivar. Our data shows that paired-end reads that matched both the retrotransposon and the exon of the endogenous gene sequences were found (Figure 3C), which indicates that this RNA is most likely synthesized from the Glyma10g28270 promotor. Curiously, we were able to detect only two paired-end reads matching the CMV sequences (Figure 3C). The rapid degradation of the hairpin-like RNA into sRNAs could possibly explain the lack of sequences mapped to this region.

The integration event was most likely a recent occurrence in Glycine soja.

To evaluate the protein sequence similarity between the sequenced CMV RNA 1 and the actual viral CMV RNA 1 sequences, we performed a phylogenetic analysis using isolates commonly used in similar assays (BALAJI; BHAT; EAPEN, 2008; LIU et al., 2009; ROOSSINCK, 2002). Using the ORF Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), we were able to obtain the whole CMV protein 1a sequence from the upCMV1 nucleotide sequence. This sequence contained 992 aa instead of the 993 aa commonly found in viral isolates but did not contain stop codons

along the protein sequence. Our results showed that the CMV isolate that was integrated into soybean genome had most likely originated from the subgroup IB because it is in the same branch as the IA strain (Figure 5), although Liu and coworkers classified this strain as belonging to the new subgroup IC in their work (LIU et al., 2009). The fact that the *Conquista* CMV sequence is remarkably similar to other actual viral sequences and did not behave as an out-group in the phylogenetic tree indicates that the integration was a recent event. In addition, the comparison between the upCMV1 and the downCMV1 sequences showed only a one-nucleotide difference.

Recently, the genomes of 31 different soybeans accessions were sequenced, including *Glycine soja*, the wild soybean (LAM et al., 2010). Using the available public deep sequence data (ftp://public.genomics.org.cn/BGI/soybean_resequencing/), we mapped the reads from all libraries against the upCMV1 sequence. Curiously, 15 of 17 of the sequenced wild soybeans and 10 of 14 of the sequenced cultivated soybeans presented the CMV integration (Table S3). This could indicate that the integration originally occurred in *Glycine soja* because Lam et al. (2010) showed that the cultivated soybean may have passed for introgression events from the wild bean in a recent past (LAM et al., 2010). One of these events, which occurred after the domestication of the soybean, could be responsible for bringing the CMV insertion from *Glycine soja* to *Glycine max*. However, another possibility could explain the absence of the CMV RNA 1 in some cultivars of *Glycine max*. If an introgression event occurred between a wild soybean that does not have the CMV and a cultivated soybean that does, this could lead to the loss of this locus in this cultivar after several crossings.

Conclusions

In this work, we ifdentified the integration of two CMV RNA 1 sequences into the genomes of both cultivated soybeans and wild soybean, although some cultivars did not present these sequences, such as the reference genome *Williams*. Although the integration of non-retroviral RNA viruses into plants is not a novelty (CHIBA et al., 2011), here we present what is most likely one of the most recent events of viral integration into a plant genome, due the high degree of conservation between the integrated CMV RNA 1 sequences (approximately 95% nucleotide similarity) and the actual viral sequences deposited into GenBank. Moreover, due to RNA recombination with a copia-like retrotransposon, the CMV sequences integrated inside a gene in such a way that the expression of this *locus* leads to the formation of a hairpin-like structure due the hybridization of both RNA 1 sequences. The entire integrated sequence very closely resembles constructs used to make transgenic plants resistant to viral infections (BONFIM et al., 2007), and given that the expression of small RNAs from CMV sequences were found in all plant tissues (Figure 1A), this integrated sequence could similarly confer viral resistance.

References

ACUÑA, R; PADILLA, B. E.; FLÓREZ-RAMOS, C. P.; RUBIO, J.D.; HERRERA, J. C.; BENAVIDES, P.; LEE S-J; YEATS, T. H.; EGAN, A. N.; DOYLE, J. J. & ROSE, J. K. C.; Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proceedings of the National Academy of Sciences of the United States of America*, 109(11): 4197–4202, 2012.

ANDERSSON, J O. Lateral gene transfer in eukaryotes. *Cellular and molecular life sciences*, 62(11): 1182–1197, 2005.

ANDERSON, J.O.; SJÖGREN, A. M.; HORNER, D. S.; MURPHY, C.; DTAL, P. L.; SVÄRD, S. G.; LOGSDON, J. M.; RAGAN, M.; HIRT, R. P. & ROGER, A. J. A genomic survey of the fish parasite Spironucleus salmonicida indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. *BMC genomics*, 8: 51, 2007.

ARAVIND, L.; TATUSOV, R. L.; WOLF, Y.I.; WALKER, D. R. & KOONIN, E. V. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles.

Trends in genetics, 14(11): 442–444, 1998.

BALAJI, S.; BHAT, A. I. & EAPEN, S. J. A phylogenetic reexamination of Cucumber mosaic virus isolates based on 1a , 2a , 3a and 3b proteins. *Indian J. Virol.* 19: 17–25, 2008.

BEJARANO, E. R. KHASHOGGI, A.; WITTY, M. & LICHTENSTEIN, C. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2): 759–764, 1996.

BERTSCH, C.; BEUVE, M.; DOLJA, V. V.; WIRTH, M.; PELSY, F.; HERRBACH, E. & LEMAIRE, O. Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biology direct*, v. 4, p. 21, jan. 2009.

BOCK, M. & STOYE, J. P. Endogenous retroviruses and the human germline. *Current opinion in genetics & development*, v. 10, n. 6, p. 651–5, dez. 2000.

BONFIM, K.; FARIA J. C.; NOGUEIRA, E. O. P. L.; MENDES, E. & ARAGÃO, F. J. L. RNAi-mediated resistance to Bean golden mosaic virus in genetically engineered common bean (Phaseolus vulgaris). *Molecular plant-microbe interactions,* 20(6): 717–726, 2007.

BOSCHETTI, C.; POUHKINA-STANTCHEVA, N.; HOFFMANN, P. & TUNNACLIFFE, A. Foreign genes and novel hydrophilic protein genes participate in the desiccation response of the bdelloid rotifer Adineta ricciae. *The Journal of experimental biology*, 214: 59–68, 2011.

CANTO, T. & PALUKAITIS, P. A cucumber mosaic virus (CMV) RNA 1 transgene mediates suppression of the homologous viral RNA 1 constitutively and prevents CMV entry into the phloem. *Journal of virology*, 75(19): 9114–9120, 2001.

CHEN, Y.-K.; LOHUIS, D.; GOLDBAH, R. & PRINS, M. High frequency induction of RNA-mediated resistance against Cucumber mosaic virus using inverted repeat constructs. *Molecular Breeding*, 14(3): 215–226, 2004.

CHIBA, S.; KONDO, H.; TANI, A.; SAISHO, D.; SAKAMOTO, W.; KANEMATSU, S. & SUZUKI, N. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS pathogens*, 7(7): e1002146, 2011.

DANCHIN, E. G. J.; ROSSO, M-N.; VIEIRA, P.; DE ALMEIDA-ENGLER, J.; COUTINHO, P. M.; HENRISSAT, B. & ABAD, P. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41): 17651–17656, 2010.

DONAIRE, L.; WANG, Y.; GONZALEZ-IBEAS, D.; MAYER, K. F.; ARANDA, M. A. & LLAVE, C. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology*, 392(2): 203–214, 2009.

DOOLITTLE, W. F.; BOUCHER, Y.; NESBØ, C. L.; DOUADY, C. J.; ANDERSSON, J. O. & ROGER, A. J. How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 358(1429): 39–58, 2003.

DRUMMOND, A. J. & RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7: 214, 2007.

DU, J.; GRANT, D.; TIAN, Z.; NELSON, R. T.; ZHU, L.; SHOEMAKER, R. C. & MA, J. SoyTEdb: a comprehensive database of transposable elements in the soybean genome. *BMC genomics*, 11: 113, 2010.

DUNNING HOTOPP, J. C.; CLARK, M. E.; OLIVEIRA, D. C. S. G.; FOSTER, J. M.; FISCHER, P.; MUÑOZ TORRES, M. C.; GIEBEL, J. D.; KUMAR, N.; ISHMAEL, N.; WANG, S.; INGRAM, J.; NENE, R. V.; SHEPARD, J.; TOMKINS, J.; RICHARDS, S.; SPIRO, D. J.; GHEDIN, E.; SLATKO, B. E.; TETTELIN, H. & WERREN, J. H. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science (New York, N.Y.)*, 317(5845): 1753–1756, 2007.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5): 1792–1797, 2004.

EICHINGER, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, 435(7038): 43–57, 2005.

FAHLGREN, N.; HOWELL, M. D.; KASSCHAU, K. D.; CHAPMAN, E. J.; SULLIVAN, C. M.; CUMBIE, J. S.; GIVAN, S. A.; LAW, T. F.; GRANT, S. R.; DANGL, J. L. & CARRINGTON, J. C. High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS one*, 2(2): e219, 2007.

FIGLEROWICZ, M. Role of RNA structure in non-homologous recombination between genomic molecules of brome mosaic virus. *Nucleic acids research*, 28(8): 1714–1723, 2000.

FRIESEN, T. L.; STUKENBROCK, E. H.; LIU, Z.; MEINHARDT, S.; LING, H.; FARIS, J. D.; RASMUSSEN, J. B.; SOLOMON, P. S.; MCDONALD, B. A. & OLIVER, R. P. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature genetics*, 38(8): 953–956, 2006.

GEUKING, M. B.; WEBER, J.; DEWANNIEUX, M.; GORELIK, E.; HEIDMANN, T.; HENGARTNER, H.; ZINKERNAGEL, R. M. & HANGARTNER, L. Recombination of Retrotransposon and Exogenous RNA Virus Results in Nonretroviral cDNA Integration. *Science*, 323(5912): 393–396, 2009.

GLADYSHEV, E. A; MESELSON, M. & ARKHIPOVA, I. R. Massive horizontal gene transfer in bdelloid rotifers. *Science (New York, N.Y.)*, 320(5880): 1210–1213, 2008.

GRAHAM, L. A.; LOUGHEED, S. C.; EWART, K. V. & DAVIES, P. L. Lateral transfer of a lectin-like antifreeze protein gene in fishes. *PloS one*, 3(7): e2616, 2008.

HACKER, J.; BLUM-OEHLER, G.; MÜHLDORFER, I. & TSCHÄPE, H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular microbiology*, 23(6): 1089–1097, 1997.

HARPER, G.; HULL, R.; LOCKHART, B. & OLSZEWSKI, N. Viral sequences integrated into plant genomes. *Annual review of phytopathology*, 40: 119–136, 2002.

HENDERSON, I. R.; ZHANG, X.; LU, C.; JOHNSON, L.; MEYERS, B. C.; GREEN, P. J. & JACOBSEN, S. E. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature genetics*, 38(6): 721–725, 2006.

HEUER, H. & SMALLA, K. Horizontal gene transfer between bacteria. *Environmental biosafety research*, 6 (1): 3–13, 2007.

HORIE, M.; HONDA, T.; SUZUKI, Y.; KOBAYSHI, Y.; DAITO, T.; OSHIDA, T.; IKUTA, K.; JERN, P.; GOJOBORI, T.; COFFIN, J. M. & TOMONOGA, K. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature*, 463(7277): 84–87, 2010.

KALANTIDIS, K.; PSARADAKIS, S.; TABLER, M. & TSAGRIS, M. The occurrence of CMV-specific short Rnas in transgenic tobacco expressing virus-derived double-stranded RNA is indicative of resistance to the virus. *Molecular plant-microbe interactions,* 15(8): 826–833, 2002.

KASSCHAU, K. D.; FAHLGREN, N.; CHAPMAN, E.J.; SULLIVAN, C. M.; CUMBIE, J. S.; GIVAN, S. A. & CARRINGTON, J. C. Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS biology*, 5(3): e57, 2007.

KEELING, P. J. & PALMER, J. D. Horizontal gene transfer in eukaryotic evolution. *Nature reviews. Genetics*, 9(8): 605–618, 2008.

KOONIN, E. V. Taming of the shrewd: novel eukaryotic genes from RNA viruses. *BMC biology*, 8: 2, 2010.

KREUZE, J. F.; PEREZ, A.; UNTIVEROS, M.; QUISPE, D.; FUENTES, S.; BARKER, I. & SIMON, R. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology*, 388(1): 1–7, 2009.

KULCHESKI, F. R.; DE OLIVEIRA, L. F., MOLINA, L. G.; ALMERÃO, M. P.; RODRIGUES, F. A.; MARCOLINO, J.; BARBOSA, J. F.; STOLF-MOREIRA, R.; NEPOMUCENO, A. L.; MARCELINO-GUIMARÃES, F. C.; ABDELNOOR, R. V.; NASCIMENTO, L. C.; CARAZZOLLE, M. F.; PEREIRA, G. A. & MARGIS, R. Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC genomics*, 12(1): 307, 2011.

LAM, H-M.; XU, X.; LIU, X.; CHEN, W.; YANG, G.; WONG, F-L.; LI, M-W.; HE, W.; QIN, N.; WANG, B.; LI, J.; JIAN, M.; WANG, J.; SHAO, G.; WANG, J.; SUN S.S-M. & ZHANG, G. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature genetics*, 42(12): 1053–1059, 2010.

LANDER, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature*, 409(6822): 860–921, 2001.

LANGMEAD, B. & SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4): 357–359, 2012.

LIU, H.; FU, Y.; JIANG, D.; LI, G.; XIE J.; CHENG, J.; PENG, Y.; GHABRIAL, S. A. & YI, X. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *Journal of virology*, 84(22): 11876–11887, 2010.

LIU, Y. Y.; YU, S. L.; LAN, Y. F.; ZHANG, C. L.; HOU S. S.; LI, X. D.; CHEN, X. Z. & ZHU, X. P. Molecular variability of five Cucumber mosaic virus isolates from China. *Acta virologica*, 53(2): 89–97, 2009.

LINDOW, M.; JACOBSEN, A.; NYGAARD, S.; MANG, Y. & KROGH, A. Intragenomic matching reveals a huge potential for miRNA-mediated regulation in plants. *PLoS Comput. Biol.*, 3(11): e238, 2007.

LOFTUS, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature*,

433(7028): 865–868, 2005.

MOLINA, L. G.; DA FONSECA, G. C.; DE MORAIS, G. L.; DE OLIVEIRA, L. F. V.; DE CARVALHO, J. B.; KULCHESKI, F. R. & MARGIS, R. Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries. *Genetics and molecular biology*, 35(1): 292–303, 2012.

NAGY, P. D. & SIMON, A. E. New insights into the mechanisms of RNA recombination. *Virology*, 235(1): 1–9, 1997.

NTUI, V. O. KYNET, K.; AZADI, P.; KHAN, R. S.; CHIN, D. P.; NAKAMURA, I. & MII, M. Transgenic accumulation of a defective cucumber mosaic virus (CMV) replicase derived double stranded RNA modulates plant defence against CMV strains O and Y in potato. *Transgenic research*, 22(6): 1191-1205, 2013.

OCHMAN, H.; LAWRENCE, J. G. & GROISMAN, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784): 299–304, 2000.

PALUKAITIS, P. & GARCÍA-ARENAL, F. Cucumoviruses. *Advances in virus research*, 62: 241–323, 2003.

PARENT, J.-S.; MARTÍNEZ DE ALBA, A. E. & VAUCHERET, H. The origin and effect of small RNA signaling in plants. *Frontiers in plant science*, 3:179, 2012.

RICARD, G. MCEWAN, N. R.; DUTILH, B. E.; JOUANY, J.; MACHEBOEUF, D.; MITSUMORI, M.; MCINTOSH, F. M.; MICHALOWSKI, T.; NAGAMINE, T.; NELSON, N.; NEWBOLD, C. J.; NSABIMANA, E.; TAKENAKA, A.; THOMAS, N. A.; USHIDA, K.; HACKSTEIN, J. H. P. & HUYNEN, M. A. Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC genomics*, 7: 22, 2006.

RIZZO, T. M. & PALUKAITIS, P. Nucleotide Sequence and Evolutionary Relationships of Cucumber Mosaic Virus (CMV) Strains: CMV RNA 1. *Journal of General Virology*, 70(1): 1–11, 1989.

ROOSSINCK, M J. Cucumber mosaic virus, a model for RNA virus evolution. *Molecular plant pathology*, 2(2): 59–63, 2001.

ROOSSINCK, MARILYN J. Evolutionary History of Cucumber Mosaic Virus Deduced by Phylogenetic Analyses. *Journal of Virology*, 76(7), 2002.

SALZBERG, S. L.; WHITE, O.; PETERSON, J. & EISEN, J. A. Microbial genes in the human genome: lateral transfer or gene loss? *Science (New York, N.Y.)*, 292(5523): 1903–1906, 2001.

SCHMUTZ, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature*, 463(7278): 178–183, 2010.

SHARMA, N.; SAHU, P. P.; PURANIK, S. & PRASAD, M. Recent Advances in Plant-Virus Interaction with Emphasis on Small Interfering RNAs (siRNAs). *Molecular biotechnology*, 2012.

STANHOPE, M. J.; LUPAS, A.; ITALIA, M. J.; KORETKE, K. K.; VOLKER, C. & BROWN, J. R. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature*, 411(6840): 940–944, 2001.

STOYE, J P. Endogenous retroviruses: still active after all these years? *Current*

*biology,*11(22): R914–R916, 2001.

STOYE, J. P. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nature reviews. Microbiology*, 10(6): 395–406, 2012.

SZTUBA-SOLIŃSKA, J; URBANOWICZ, A.; FIGLEROWICZ, M. & BUJARSKI, J.J. RNA-RNA recombination in plant virus replication and evolution. *Annual review of phytopathology*, 49: 415–443, 2011.

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. *Molecular biology and evolution*, 28(10): 2731–2739, 2011.

TAYLOR, D. J. & BRUENN, J. The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC biology*, 7: 88, 2009.

TAYLOR, D. J.; LEACH, R. W. & BRUENN, J. Filoviruses are ancient and integrated into mammalian genomes. *BMC evolutionary biology*, 10(1): 193, 2010.

WANG, M. B.; ABBOTT, D. C. & WATERHOUSE, P. M. A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus. *Molecular plant pathology*, 1(6): 347–356, 2000.

WOLF, Y. I.; ARAVIND, L. & KOONIN, E. V. *Rickettsiae* and *Chlamydiae*: evidence of horizontal gene transfer and gene exchange. *Trends in genetics,* 15(5): 173–175, 1999.

WU, Q.; LUO, Y.; LU, R.; LAU, N.; LAI, E. C.; Li, W-X. & DING, S-W. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 107(4): 1606–1611, 2010.

ZERBINO, D. R. & BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5): 821–829, 2008.

ZHANG, C.; GU, H. & GHABRIAL, S. A. Molecular Characterization of Naturally Occurring RNA1 Recombinants of the Comovirus *Bean pod mottle virus*. *Phytopathology*, 97(10): 1255–1262, 2007.

ZHANG, H. & ZHU, J. RNA-directed DNA methylation. *Current Opinion in Plant Biology*, 14(2): 142–147, 2011.

Legend of figures

Figure 1. PCR amplification of the CMV RNA 1 sequence in soybean DNA.

A) A schematic of the CMV RNA 1 sequence. The lines represent the UTRs, and the grey box is the ORF of the CMV 1a protein. The arrows indicate the primers used in the two PCRs. The expected lengths of the two amplicons are also indicated in the figure. B) Two

1% agarose gels that confirmed the amplification of the CMV RNA 1 sequences in all soybean cultivars tested, except in *Williams*. (Cq: *Conquista*; Ur: *Urano*; Br16: *BR 16*; E48: *Embrapa 48*; W: water; Dv: *Davis*; H: *Hill*; Pr: *Paraná*, Br4: *BR 4*; I4: *IAS 4*; For: *Força*; Pot: *Potência*).

Figure 2. Frequency of CMV and BPMV small RNAs

Graphic showing the relative frequency of each class of CMV sRNA in each library. The sRNA libraries were sequenced from RNA isolated from different samples of five soybean cultivars (Root 1- *BR-16* cultivar; Root 2 and Leaf 1 – *Embrapa-48* cultivar; Leaf 2 – *PI 561356* cultivar; Flower, Stamen, Carpel and Petal – *Urano* cultivar; Germinating seed, Seed and Pod – *Conquista* cultivar).

Figure 3. Confirmation of the integration site of the CMV RNA 1 sequence into the genome of *Conquista*.

A) Scheme of the integration site of the two CMV RNA 1 sequences in the third exon of the Glyma10g28270 locus. The CMV sequences (indicated as upCMV1 for the upstream sequence and downCMV1 for the downstream sequence) are flanked by the sequences of a copia-like retrotransposon (indicated in orange and red for the LTR regions) and a sequence of the *COPZ2* transcript (lCOPZ2), as shown in the figure. The orientation of the sequences is indicated by the arrows at their ends. The primers used for PCR are indicated by black arrows. The expected length of each PCR amplicon and the length of all sequences are shown in the figure. B) Raw number of sRNA reads (sequenced from a leaf sample of the soybean cultivar *Conquista*) mapped along the region that integrated into the soybean genome. The distribution pattern of the sRNAs shows that they are derived almost

exclusively from the region of CMV RNA 1. C) Table indicating the raw number of paired-end reads (from an RNAseq library sampled from a leaf of the soybean cultivar *Conquista*) mapped to each region of the integration site. The columns with two regions separated by a bar indicates that the paired-end reads mapped to both regions. D) A 1% agarose gel showing the amplicons from the PCRs performed using DNA from cultivars *Williams* and *Conquista*. The primers used for each PCR are indicated in the scheme above. Primers for the *Cytochrome P450* (*CYP*) gene of soybean were used as positive control. E) A 1% agarose gel showing the results for the PCR performed using primers flanking the integration site of the copia-like retrotransposon into the Glyma10g28270 locus indicated in the scheme above. Only the DNA from *Williams* shows the expected amplicon.

Figure 4. Proposed model for the recombinations between the *COPZ*2 mRNA and the CMV RNAs.

A) Above, a representation of the *COPZ2* transcript sequence. The introns are indicated by lines, blue boxes represent the coding sequences and the green boxes the untranslated regions. The lines connecting the two schemes indicate the regions of the transcript that are found in the sequence integrated into the soybean genome.

B) Proposed model for the first event of the recombination between a CMV RNA1 and the *COPZ2*

mRNA. During viral replication, a template switch occurred between the minus strand of the CMV RNA1 and the *COPZ2* mRNA. A 13-nt region with homology to the end of the CMV RNA1 plus strand facilitated the formation of the heteroduplex. This event allowed the elongation of the nascent RNA using the remaining sequence of the *COPZ2* as the template.

C) Proposed model for the second event of recombination between the nascent RNA containing the CMV RNA1 and *COPZ2* mRNA sequences and a second CMV RNA1. The viral replicase “jumped” from the 5'UTR of the *COPZ2* mRNA to a region approximately 200 nt away from the 3' end of the CMV RNA1 plus strand. Both the UTR and the 3' end of RNA1 harbor secondary structures, which possibly stopped the viral RdRp, thereby allowing the jump to CMV RNA1 and the subsequent elongation of the nascent RNA.

Figure 5. Phylogenetic Analysis.

Phylogenetic analysis of the 1a ORF of different CMV strains and the CMV sequence integrated in soybean (CMV_Cq). The protein sequence of the *Peanut stunt virus* (PSV) was used as the out-group. The Bayesian method was used to construct the tree. The rectangles designate the subgroups of the strains according to different phylogenetic works [63–65]. Posterior probabilities are show above the branches.

Figure 1

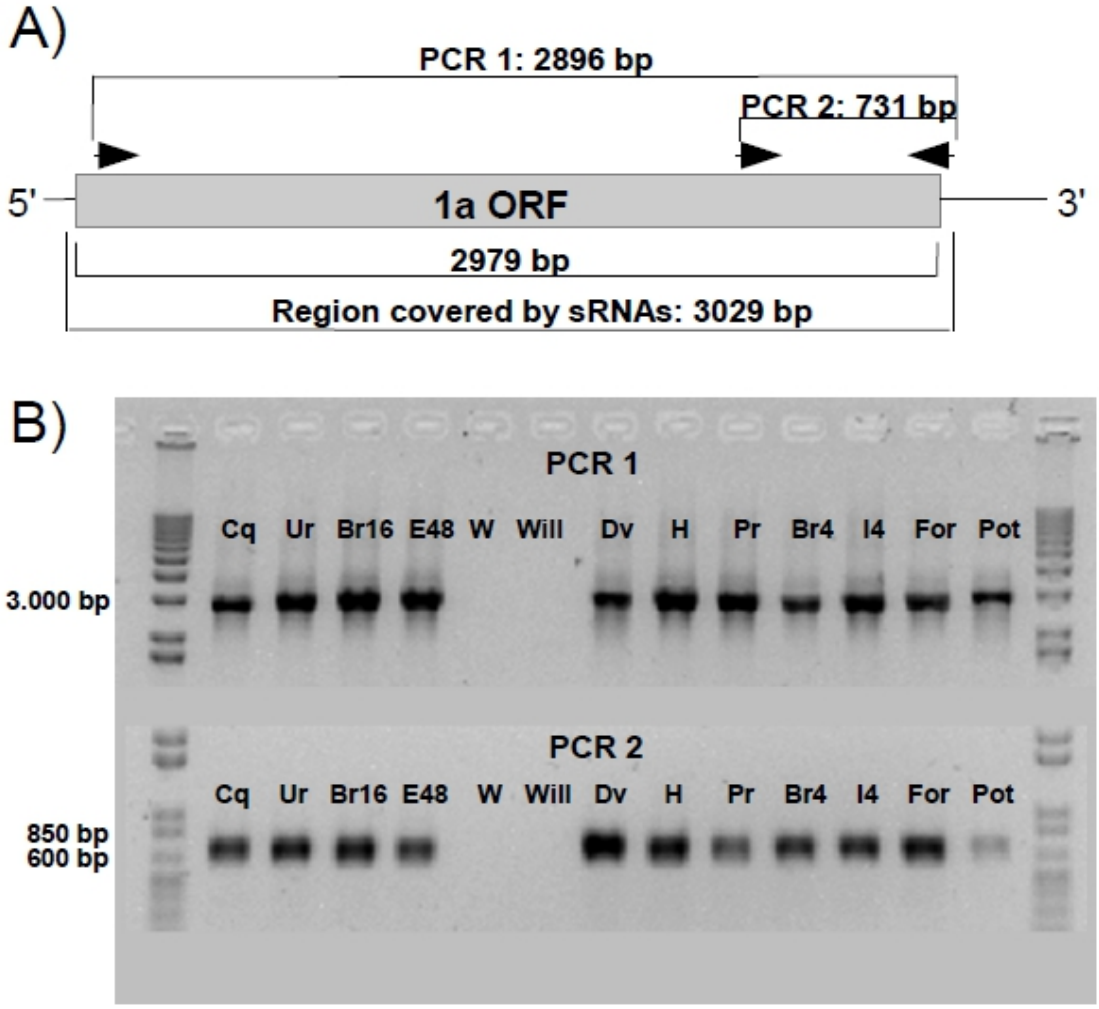
A)
PCR 1: 2896 bp
PCR 2: 731 bp
5'
1a ORF
3'
2979 bp
Region covered by sRNAs: 3029 bp
B)
PCR 1
Cq Ur Br16 E48 W Will Dv H Pr Br4 I4 For Pot
3.000 bp
PCR 2
Cq Ur Br16 E48 W Will Dv H Pr Br4 I4 For Pot
850 bp
600 bp

Figure 2

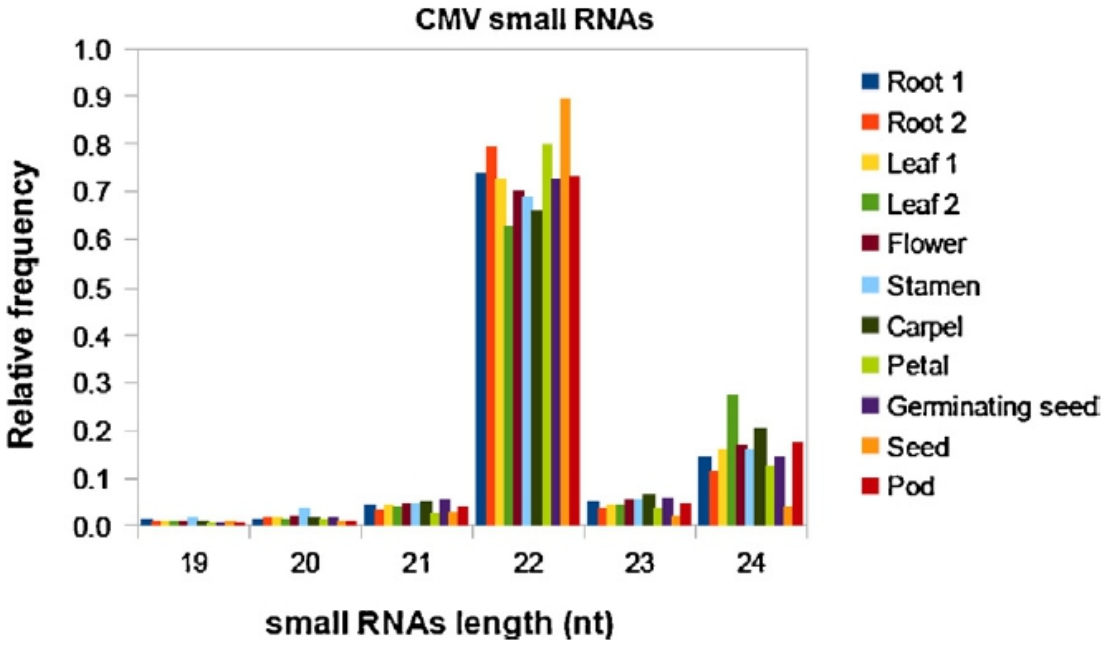
CMV small RNAs
Relative frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
19
20
21
22
23
24
small RNAs length (nt)
Root 1
Root 2
Leaf 1
Leaf 2
Flower
Stamen
Carpel
Petal
Germinating seed
Seed
Pod

Figure 3

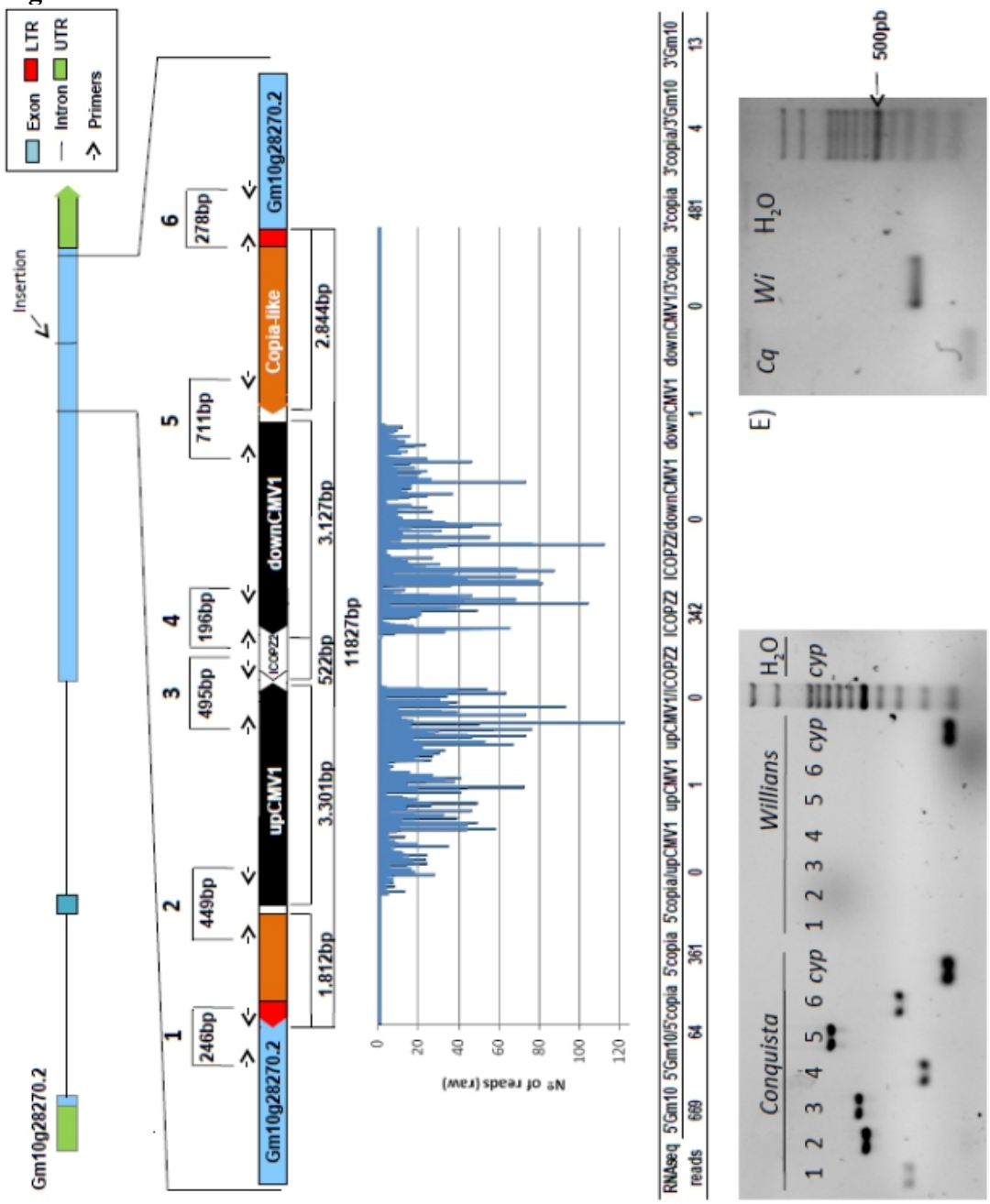
A)
Exon
LTR
Intron
UTR
Primers
Insertion
Gm10g28270.2
Gm10g28270.2
1
246bp
2
449bp
3
495bp
4
196bp
5
711bp
6
278bp
Gm10g28270.2
upCMV1
ICOPZ2
downCMV1
Copia-like
1.812bp
3.301bp
522bp
3.127bp
2.844bp
11827bp
B)
N° of reads (raw)
0
20
40
60
80
100
120
C)
RNAseq reads
5'Gm10 669
5'Gm10/5'copia 64
5'copia 361
5'copia/upCMV1 0
upCMV1 1
upCMV1/ICOPZ2 0
ICOPZ2 342
ICOPZ2/downCMV1 0
downCMV1 1
downCMV1/3'copia 0
3'copia 481
3'copia/3'Gm10 4
3'Gm10 13
D)
Conquista
1 2 3 4 5 6 cyp
Williams
1 2 3 4 5 6 cyp
H2O
cyp
E)
Cq
Wi
H2O
500pb

Figure 4

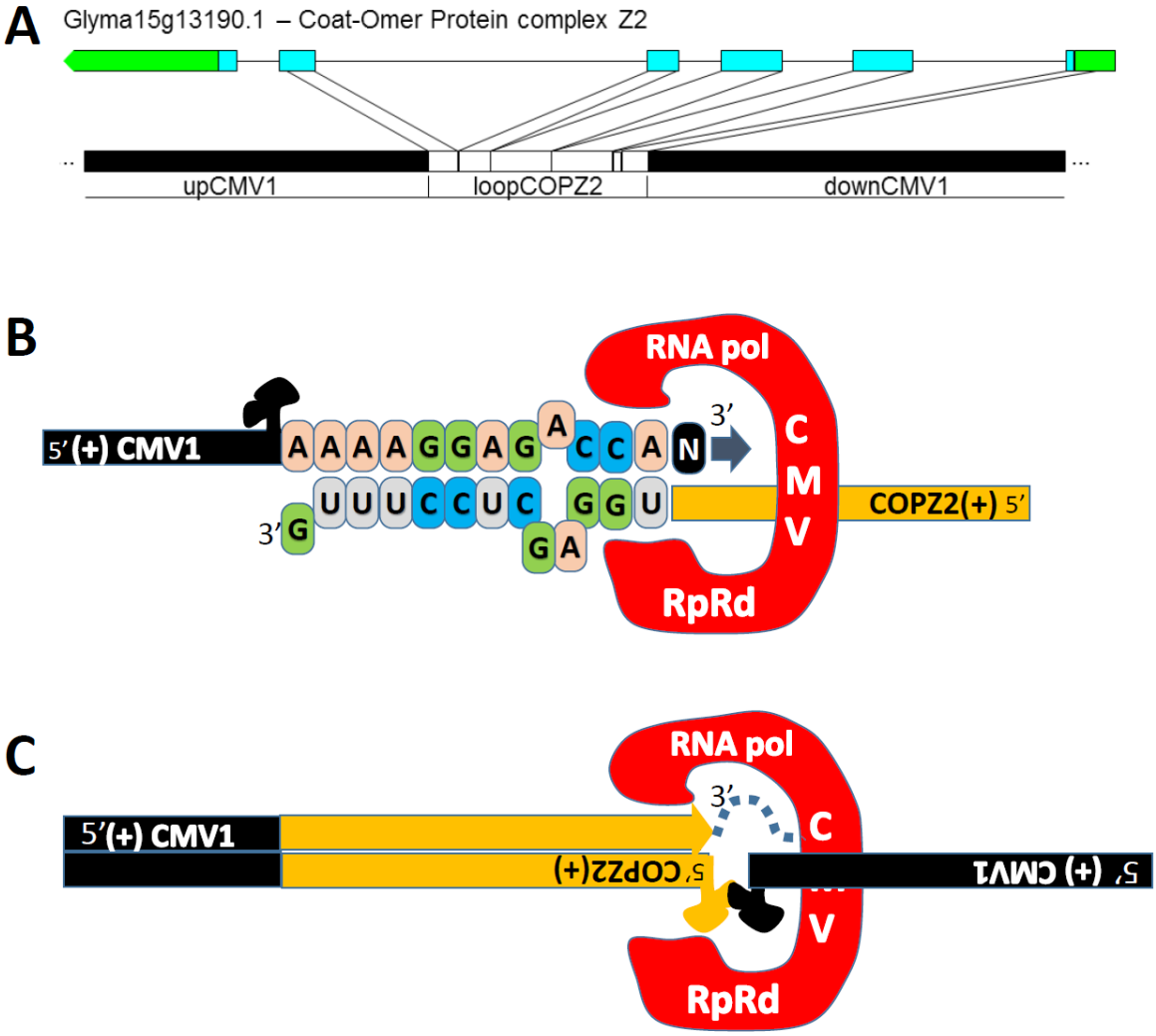
A
Glyma15g13190.1 – Coat-Omer Protein complex Z2
upCMV1
loopCOPZ2
downCMV1
B
RNA pol
5′(+) CMV1
A A A A G G A G
A
C C A
N
3′
C
M
V
U U U C C U C
G G U
3′ G
G A
COPZ2(+) 5′
RpRd
C
RNA pol
5′(+) CMV1
3′
5′ COPZ2(+)
5′ (+) CMV1
V
RpRd

Figure 5

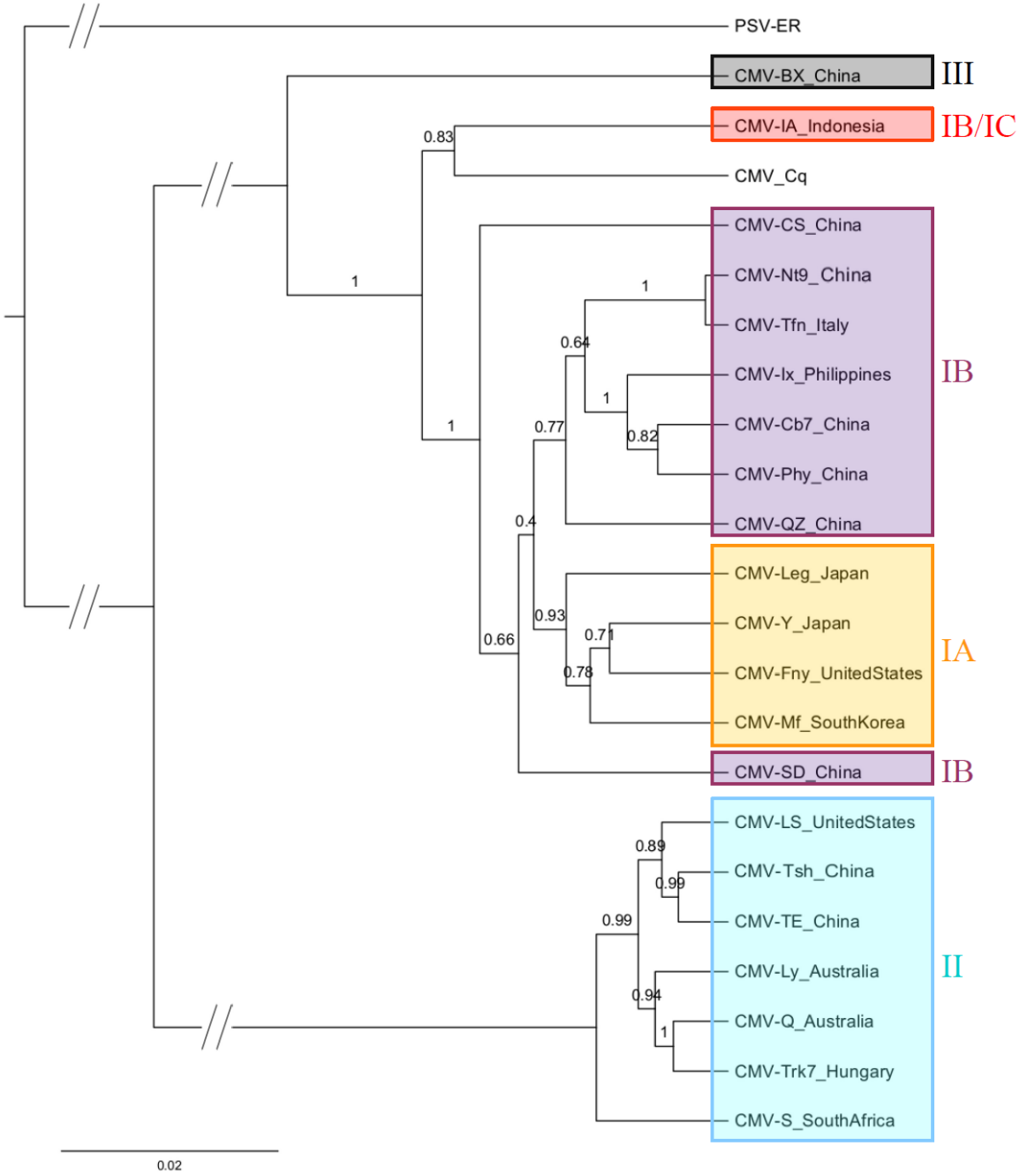
PSV-ER
CMV-BX_China
III
CMV-IA_Indonesia
IB/IC
CMV_Cq
CMV-CS_China
CMV-Nt9_China
CMV-Tfn_Italy
CMV-Ix_Philippines
IB
CMV-Cb7_China
CMV-Phy_China
CMV-QZ_China
CMV-Leg_Japan
CMV-Y_Japan
CMV-Fny_UnitedStates
IA
CMV-Mf_SouthKorea
CMV-SD_China
IB
CMV-LS_UnitedStates
CMV-Tsh_China
CMV-TE_China
CMV-Ly_Australia
II
CMV-Q_Australia
CMV-Trk7_Hungary
CMV-S_SouthAfrica
0.83
1
1
0.64
1
0.77
0.82
0.4
0.66
0.93
0.71
0.78
0.89
0.99
0.99
0.94
1
0.02

Table S1

Primers used to amplify different regions of the sequence integrated into the soybean genome.

Primer name	Nucleotide positions*	Primer sequence
CMV_R1_5'_F	2031-2050/8687-8668	5'-GCTCGAGGAGCAATTACAGC-3'
CMV_R1_5'_R	2159-2140/8559-8578	5'-TGCTGGGTAAGATGGAGGTC-3'
CMV_R1_3'_a	4196-4215/6522-6503	5'-CCAAAGAAGCCTGCACTTTC-3'
CMV_R1_3'_b	4926-4907/5792-5811	5'-AAAGGTGGGAAACCGCTAAG-3'
CMV_R1_3'_c	4738-4757/5980-5961	5'-GGCATTTCTGAGGACCATGT-3'
CMV_R1_3'_d	4923-4942/5795-5776	5'-CTTTGAGCGGGATCTGAGTT-3'
CMV_R1_3'_e	4907-4926/5811-5792	5'-CTTAGCGGTTTCCCACCTTT-3'
Gm10g_F	4796-4816	5'-TCACTGACGTGATTGAAGCAG-3'
Gm10g_R	5041-5060	5'-GCCTTTGCCTCATTCTGTTC-3'
LTR_F	9-33/11666-11690	5'-TGGTTAGGATTCAGTTACAGAGGAG-3'
LTR_R	101-75/11758-11732	5'-TGTAACACACACTTTACAAATCAATCA-3'
Copia_F	1711-1730	5'-GGCTCAGTGCAGTGTGTGAT-3'
Copia_R	9269-9245	5'-AATCTGGTGTGAAAGGAGTTGTACT-3'
lCOPZ_a	5232-5212	5'-TGTTATAGCTGGGAAAGTTGC-3'
lCOPZ_b	5600-5619	5'-CGGAATCCAAAAGAAGGACA-3'

* According to GenBank Accession number KP137860 , except for Gm10g_F and Gm10g_R, which are according to the locus Glyma10g28270.

Table S2

Accession numbers of the CMV RNA1 sequences used in the phylogenetic analysis.

Isolate	Accession number	Origin
CMV_Cq	KP137860	Unknown
PSV-ER	NC_002038	USA
CMV-Leg	D16403.1	Japan
CMV-Y	D12537.1	Japan
CMV-Fny	D00356.1	United States
CMV-Mf	AJ276479.1	South Korea
CMV-CS	AY429435.1	China
CMV-Nt9	D28778.1	China
CMV-Tfn	Y16924.1	Italy
CMV-Ix	U20220.1	Philippines
CMV-Cb7	EF216866.1	China
CMV-Phy	DQ402477.1	China
CMV-QZ	EU414795.1	China
CMV-SD	AF071551.1	China
CMV-IA	AB042292.1	Indonesia
CMV-LS	AF416899.1	United States
CMV-Tsh	EF202595.1	China
CMV-TE*	EU665000.1	China
CMV-Ly	AF198101.1	Australia
CMV-Q	X02733.1	Australia
CMV-Trk7	AJ007933.1	Hungary
CMV-S	Y10884.1	South Africa
CMV-BX	DQ399548.1	China

* Because there is no information about the isolate name in the NCBI database, this isolate was designated according to its host (i.e., *Tagetes erecta*).

Table S3

Presence or absence of the CMV RNA 1 sequence in different soybean accessions.

Accession	N° of CMV RNA 1 paired-end reads	Total reads	Origin*
W 01	355	54,465,171	Beijing area, China
W 02	334	58,190,974	Liaoning, China
W 03	227	59,233,958	Inner Mongolia, China
W 04	462	47,365,616	Henan, China
W 05	279	82,677,560	Henan, China
W 06	42	14,121,208	Heilongjiang, China
W 07	490	58,598,852	Liaoning, China
W 08	2	55,558,012	Heilongjiang, China
W 09	66	16,994,087	Liaoning, China
W 10	0	55,135,521	Heilongjiang, China
W 11	357	53,565,096	Shanxi, China
W 12	434	76,032,002	Anhui, China
W 13	1	74,964,143	Inner Mongolia, China
W 14	0	16,844,015	Inner Mongolia, China
W 15	356	53,036,272	Henan, China
W 16	207	49,792,590	Heilongjiang, China
W 17	293	15,552,743	Liaoning, China
C 01	332	51,131,787	Shandong, China
C 02	0	54,028,270	Liaoning, China
C 08	0	78,640,369	USA
C 12	335	50,670,708	Shanxi, China
C 14	435	55,835,567	Brazil
C 16	0	62,962,367	Japan
C 17	351	56,491,930	Sichuan, China
C 19	380	59,086,921	Jilin, China
C 24	380	57,737,957	Jiangxi, China
C 27	384	62,318,457	Hebei, China
C 30	436	60,615,235	Henan, China
C 33	0	60,522,537	Heilongjiang, China
C 34	424	58,970,721	Guangxi, China
C 35	523	55,237,958	Guangdong, China

- According to the work of Lam et al. (2010).

CAPÍTULO II – INTERACTION OF DRBs AND VIRAL PROTEINS IN *NICOTIANA BENTHAMIANA*

Autores: Guilherme Cordenonsi da Fonseca[1], Peter M. Waterhouse[2], Rogerio Margis[12*]

[1] Centro de Biotecnologia, Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil.

[2]Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Queensland, Australia.

[3]Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil.

Artigo em preparação

Abstract

During viral infection the activation of the RNA interference (RNAi) machinery is one of the main defense pathways of plants. The RNAi system requires the coordinate action of Argonautes (AGO), Dicers-like (DCL), RNA-dependent RNA polymerases (RDR) among other accessory proteins to recognize the replicating double-stranded (ds) viral RNA and process it into viral small RNAs (vsRNAs) that will trigger post-transcriptional silencing of the virus. The dsRNA binding protein (DRB) family is associated with the biogenesis of small RNAs and antiviral defense. Using live cell imaging we were able to show that five of the eight *Nicotiana benthamiana* DRB proteins relocate to viral replication complexes (VRCs) upon Potato virus X (PVX) infection. We propose that these proteins act as viral-invasion sensors in the earliest stages of virus infection as the VRCs migrate toward the nucleus.

Keywords: *Nicotiana benthamiana,* RNAi, dsRNA-binding proteins, Viral replication complexes and Potato Virus X.

Introduction

Plants have elaborated molecular and biochemical responses to cope with pathogen infections (MANDADI & SCHOLTHOF, 2013). Plant antiviral defense relies mostly on the RNA interference (RNAi) mechanism, since the majority of the plant viruses have RNA genomes. RNAi pathways, present in all eukaryotes, are triggered by double-stranded RNAs (dsRNAs) that are processed into small interfering RNAs (siRNAs) by a series of proteins including Dicers, Argonautes (AGO), dsRNA-binding (DRB) and RNA-dependent RNA polymerase (RDR) proteins (EAMENS et al., 2008). The siRNAs can promote transcriptional gene silencing (TGS) in the nuclei by DNA methylation and histone

modification or post-transcriptional gene silencing (PTGS) through RNA cleavage or inhibition of translation (HOFFER et al., 2011). In the model plant *Arabidopsis thaliana* there are 4 DICER-LIKE (DCL), 10 AGO, 5 DRB and 6 RDR proteins that act regulating endogenous gene expression and antiviral defense (EAMENS et al., 2008). During a viral infection, the replicating viral dsRNA is cleaved into viral siRNAs (vsiRNAs), mainly by DCL4 and DRB4, that are loaded into AGO1 to promote the cleavage of single stranded viral RNAs (LLAVE, 2010). The cleavage of the viral RNA can also trigger a secondary response by the polymerization of dsRNA by RDR6 (MOURRAIN et al., 2000). Although the function of DRB1 and DRB4, especially in antiviral resistance, is well-established, the function of the DRBs 2, 3 and 5 is less understood. DRB2 has been show to aid in the biogenesis of specific microRNAs (miRNAs) in the shoot apical meristem, and DRB3 and 5 to assist in silencing the transcripts of the targets of those miRNAs, probably by traslational repression (EAMENS et al., 2012a; EAMENS et al., 2012b). Recently DRB3 together with DCL3 and AGO4 have been shown to participate in antiviral defense against geminiviruses through RNA-directed DNA methylation (RdDM) pathway (RAJA et al., 2014).

Most of the mechanics about RNAi pathways has been elucidated based on RNA (including small RNAs) analysis of different virus-challenged *Arabidopsis* mutant lines (DING; VOINNET, 2007; LLAVE, 2010). However, the intracellular localization of these proteins during a viral infection have been less closely examined. Since all four DCLs appear to be restricted to the nucleus and RNA viruses have citoplasmatic life cycles, it has not been yet clarified how and where the viral dsRNAs first contacts the RNAi proteins. Considering that the double-stranded RNA domains from DRBs can also interact with proteins and DRB2, 3 and 5 are localized in the cytoplasm (HIRAGURI et al., 2005), these

proteins appear to be likely candidates for the first step of viral recognition by the RNAi machinery.

In order to study the interaction between DRBs and viral proteins we fluorescently tagged all the DRBs of *Nicotiana benthamiana* as well as the proteins of Potato virus X (PVX). In this work we show that in response to PVX infection the DRB2, 3 and 5 proteins relocate to the viral replication complexes (VRCs) and might be involved in the earlier stages of antiviral defense.

Material and methods

Plasmid constructs

To generate the binary vector 35S::PVX.CFP-CP a 3.2 kb fragment of the pTXS.GFP-CP plasmid (CRUZ; CHAPMAN, 1996) was obtained by PCR amplification with 5'-AGATTTTCCTAGGCACG-3' (forward) and 5'-CGACCTCGAGTGACAGC-3' (reverse) primers. The product containing the PVX cDNA (flanked by the unique restriction sites AvrII and XhoI) and the GFP-2A-CP cassette was cloned into pGEM-T Easy (Promega). The *GFP* ORF was removed using NheI and HindIII enzymes. The CFP cDNA was amplified using the primers 5'-GCACCAGCTAGCATCGATCATGGTGAGCAAGGGCGA-3' (foward) and 5'-CTCCCGCAAGCTTAAGAAGGTCAAAATTTCTAGATCCGGATTACTTGTACAGCTCGT-3' (reverse) and after digestion with NheI-HindIII was cloned into the binary vector. The modified 3.2 kb fragment was ligated into pART27-35S::PVX.GFP after treatment with AvrII-XhoI enzymes generating the final vector pART27-35S::PVX.CFP-CP. The binary vector pART27-35S::PVX.GFP had been generated by subcloning the 35S-PVX.GFP-NOS fragment from a derivative of pPVX204 (BAULCOMBE; CHAPMAN;

SANTA CRUZ, 1995) into the binary vector pART27 (GLEAVE, 1992) using the unique restriction sites NotI and blunted SpeI. A binary expression vector of PVX TGB1 protein N-terminally fused to YFP was generated by PCR amplification of TGB1 coding sequence from pTXS.GFP-CP previously described with primers 5'-GGATCCCTCGAGATGGATATTCTCATCAGTAG-3'(forward) and 5'-GGATCCGAATTCTGGCCCTGCGCGGACATATGTC-3' (reverse). The PCR fragment was then digested with XhoI and EcoRI, and ligated into XhoI-EcoRI-treated pORE1 binary vector (COUTU et al., 2007) harbouring 35S::YFP to generate pORE1-35S::TGB1-YFP. For expression of free mOrange, the respective cDNA was cloned as NcoI-KpnI fragments into pORE1 to give pORE1-35S::OFP. Gateway binary expression vectors driven by the 35S promoter for production of mOrange tag (pORE2-OFP Gateway) were constructed as follows. To generate pORE2-GFP Gateway vector, a 4.6 kb cassette from pART7-GFP, containing a CaMV 35S promoter, Gateway recombination sites in frame with *GFP* cDNA, and the OCS terminator, was excised as a NotI fragment through partial digestion and ligated into pORE2 binary vector linearized with NotI. To make pORE2-OFP Gateway vector, mOrange cDNA was overlap PCR amplified to generate a SalI-ApaI fragment (containing the attR2 site in frame with mOrange, and the OCS terminator sequence) which was ligated into pORE2-GFP Gateway digested with SalI and ApaI. Coding sequences from *N. benthamiana* DRBs (NAKASUGI et al., 2014) were PCR amplified from leaves cDNA, using Pfu DNA polymerase (Promega), and cloned into pENTR/D-TOPO (Invitrogen) using primers listed in Table S1. The CACC added to forward primers allowed directional cloning into the entry vector and reverse primers eliminated stop codons to allow translational fusion with the fluorescent protein coding sequence. PCR products captured in pENTR/D-TOPO were recombined into pORE2-OFP

Gateway vectors for *Agrobacterium tumefaciens*-mediated plant expression. The *DRB1a* was also cloned in pEarleyGate-101 (EARLEY et al., 2006) to generate the 35S::DRB1a-YFP vector.

Transient expression in *N. benthamiana*

Transient expression in *N. benthamiana* leaves was performed as follows. *Agrobacterium tumefaciens* strain GV3101 transformed with the binary vector of interest were grown overnight at 28ºC in 25 ml of Luria-Bertani (LB) medium with the appropriated antibiotics. After centrifugation the bacterial pellet was ressuspended in infiltration buffer (10 mM of $MgCl_2$, 10 mM of MES and 150 mM of Acetosyringone) to the desired concentration (0.5 OD_{600} to PVX and DRB-OFP constructs and 0.25 to DRB1a-YFP and TGB1-YFP constructs). After 2-3 hours of incubation at room temperature in the dark the solution was infiltrated with a needless syringe into the underside of 4-5 weeks old *N. benthamiana* leaves. The leaves were imaged by confocal microscopy after 3 to 5 days post infiltration.

Confocal microscopy

Leaf tissue transiently expressing fluorescent proteins were analyzed using a Pascal confocal laser scanning microscope (Carl Zeiss, Germany). The underside of the leaves were examined after been mounted onto a slide containing milli-Q water. Cyan fluorescent protein (CFP) was excited with the 458 nm laser line of an argon laser and emitted light captured through a 492-508 nm bandpass filter. Yellow fluorescent protein (YFP) was excited with a 514 nm argon laser line and emitted light captured through a 520-542 nm bandpass filter. Orange (OFP) and mCherry were excited with a green HeNe laser at 543 nm and emitted light was captured through a 557-603 bandpass filter. Cells were imaged

using a 63X water-dipping objective lens with a numerical aperture of 0.95 and all images were collected at a single focal plane using LSM 5 software. Figures were constructed using Photoshop CS5 (Adobe Software).

Phylogenetic analysis

The *A. thaliana* DRBs amino acid sequences were obtained from TAIR database (https://www.arabidopsis.org/), the *N. bentahiama* DRBs amino acid sequences were obtained from *Nicotiana benthamiana* Genome Page (http://sydney.edu.au/science/molecular_bioscience/sites/benthamiana/, (NAKASUGI et al., 2014)) and the CeRDE-4 sequence from the NCBI database (Accession: NC_003281). The MUSCLE algorithm (EDGAR, 2004) was used to align the sequences. The phylogenetic analysis was performed after protein sequence alignments using the Bayesian method, carried out in BEAST1.7 software (DRUMMOND; RAMBAUT, 2007). The model of protein evolution used in this analysis was the Dayhoff model for protein matrix substitution. The Yule tree was selected as a tree prior to the Bayesian analysis, and 20,000,000 generations were performed using Markov chain Monte Carlo (MCMC) algorithms. The trees were visualized and edited using FigTree v1.3.1 software (http://tree.bio.ed.ac.uk/software/figtree/).

Results and Discussion

Viral Replication Complex visualized with fluorescent PVX coat protein.

The majority of plant viruses have single-stranded positive RNA genomes and encode an RNA-dependent RNA polymerase (RdRP) which is used together with host proteins to replicate their genomes (LINNIK et al., 2013). Viral RdRPs are active as oligomeric arrays that usually modify the host cell membrane to form structures referred to

VRCs (ASURMENDI et al., 2004). The process to form the VRCs, as well as the interaction and location of viral RNA and host and viral encoded proteins in these structures has been studied through confocal microscopy (LINNIK et al., 2013; TILSNER et al., 2012, 2013). In order to analyze in greater detail the exact location of interaction between the virus and DRB proteins, an infectious PVX-overcoat construct (CRUZ; CHAPMAN, 1996) was modified to encode cyan fluorescent protein (CFP) fused to the viral coat protein (CP). This construct was used to infect 4-5 week old *Nicotiana benthamiana* leaves by agro-infiltration. Five days post inoculation (dpi) the leaves were examined by confocal microscopy confirming the presence of the VRCs forming ring-like structures adjacent to the nuclei of infected cells (Figure 1) as described in previous reports (TILSNER et al., 2009).

DRB 2, 3 and 5 relocate to VRC during PVX infection.

Plant viruses move through plasmodesmata (PD) to transport their genomes between host cells and spread their infection (TILSNER; AMARI; TORRANCE, 2011b). PVX encodes three proteins that are required for viral cell-to-cell movement in overlapping open reading frames, the “triple gene-block” (TGB; VERCHOT-LUBICZ et al., 2010). TGB1 has been shown to be one of the key proteins involved in the organization of the VRCs through remodeling of host actin (TILSNER et al., 2012). Within the VRC this protein forms aggregates surrounded by a reservoir of recruited host endomembranes where TGB2/3 and the nonencapsidated viral RNA can be localized (TILSNER et al., 2012). Given the importance of this structures in the process of viral infection in plant cells we made a TGB1-YFP construct to use as a marker of the locations and interactions of the DRBs and VRCs. This construct was agro-infiltrated into 4-5 weeks old *N. bentahiama* leaves in combination with the PVX-CFP infectious clone and DRB proteins tagged with

orange fluorescent protein (OFP).

According to a recent report from *N. benthamiana* transcriptome (NAKASUGI et al., 2014), this plant has eight DRB copies in contrast with the five present in *A. thaliana* genome, due to a duplication of the DRB1, 2 and 3 genes (Figure 2). All of them were cloned into OFP tagged constructs. First, the location of these proteins was examined in plants that were not agro-infiltrated with the PVX-CFP construct. DRB1a, 1b and 4 were located exclusively in the nuclei of the cells (Figure 3A) while the others were visible in both the cytoplasm and nuclei (Figure 4A). These results corroborate with previous works describing the localization of this protein family in *A. thaliana* (EAMENS et al., 2012; HIRAGURI et al., 2005). Interesting, in plants co-agro-infiltrated with PVX infectious clone and the OFP constructs, the proteins DRB2a/b, 3a/b and 5 appeared to concentrate in VRCs, diminishing elsewhere in the cell, especially in the nuclei (Figure 4B), indicating a possible relocation of these proteins during a viral infection. In some cells, where DRB4 appear to be overexpressed, it is also possible to see a co-localization of this protein and the virus, but it seems to not concentrate in the ring-like structures around the VRCs (Figure 3B). To further investigate the localization of the DRBs into VRCs the TGB1-YFP construct was added to the agroinfiltration experiments. Examining the cells that simultaneously expressed the three constructs it was possible to identify that TGB1 was localized in the center of the ring-like structures formed by DRB2a/b, 3a/b and 5 during PVX infection (Figure 4C). In previous reports (TILSNER et al., 2012) was demonstrated that TGB1 was located within the center of viral RNA whorls, resembling the structures identified in our experiments. Furthermore, it was demonstrated that DRB2, that share the same dsRNA binding sites with the other DRBs, can bind dsRNA, but not single stranded RNA (HIRAGURI et al., 2005b), suggesting that DRB2a/b, 3a/b and 5 are binding to the

double stranded forms of PVX RNA. In rare cells, where DRB4 were over-expressed, it was possible to see DRB4 outside the nucleus, but it appears to be spread throughout all the cytoplasm and not concentrated aroud TGB1 (Figure 3B). Although it was recently demonstrated that DRB3 together with AGO4 play an important role in the nucleus acting in the defense against geminiviruses through RdRM pathway (RAJA et al., 2014), until now, DRB2, 3 and 5 have not been associated with antiviral responses in the cytoplasm.

A model for DRB2, 3 and 5 antiviral action.

In plants, RNA and DNA virus activate the RNAi silencing through the formation of dsRNA during replication (FUSARO et al., 2006). The viral dsRNA is processed into 21 to 24 nt vsiRNAs due the action of DCL enzymes, mainly by DCL2 and DCL4 with the support of DRB4 (LLAVE, 2010). These vsiRNAs are loaded to AGO and used to guide the degradation of the viral genome. However, DCL enzymes and DRB4 are located exclusively in the nucleus, while RNA viruses have entirely cytoplasmatic life cycles (SCHOELZ; HARRIES; NELSON, 2011). This raises the question of how and where the RNAi silencing machinery contact their target viruses. In *A. thaliana* it was demonstrated that DRB4 is relocated from the nucleus into the cytoplasm upon infection with Turnip yellow mosaic virus (TYMV (JAKUBIEC; YANG; CHUA, 2012b)). Since DRB4 interacts with DCL4 (HIRAGURI et al., 2005b), it is possible that these proteins could be transported to the cytoplasm to generate the vsiRNAs from the replicating virus (JAKUBIEC; YANG; CHUA, 2012b). In our experiments, we could detect a very low level of relocation of *N. benthamiana* DRB4 into VRCs and only in cells that overexpressed this protein (Figure 3B). In contrast DRB2a/b, 3a/b and 5 were substantially relocated to interact with the replicating PVX. This observation, together with previous

reports about roles and cellular locations of PVX encoded proteins, might shed some light in the possible sequence of events that occur in the earlier stages of the antiviral response (LINNIK et al., 2013; TILSNER et al., 2012; TILSNER et al., 2013).

In uninfected cells DRB2a/b, 3a/b and 5 are located into cytoplasm, where DRB4 is in the nucleus (Figure 3A and Figure 4A). In a newly infected cell the cytoplasmatic DRBs could first interact with the replicating virus before it reaches a perinuclear location. In fact, in *A. thaliana* DRB3 and 5 were demonstrated to be strong candidates for mediators of translational repression (EAMENS et al., 2012a), indicating a possible role of these proteins in restraining the virus in the earlier steps of infection through translation inhibition of the viral encoded proteins. When the VRCs interfaces with the nucleus the viral RNA could be accessed and processed into vsiRNAs by DCL2, DCL4 and DRB4 (DELERIS et al., 2006; FUSARO et al., 2006). Once these siRNAs are loaded into AGO1 they could target the single-stranded viral RNA in the VRCs and trigger the production of dsRNA by RDR6, thus generating more vsiRNAs that could move from cell to cell through PD and also long distances in the phloem initiating the latter stage of the RNAi antiviral defense (MOLNAR et al., 2011; SCHWACH et al., 2005).

Considering that DRB3 is expressed throughout the vasculature, whereas DRB2 and DRB5 are expressed mainly in the apical meristem in *A. thaliana* (CURTIN et al., 2008), we proposed that DRB3a/b should play a predominant role in the earlier step of the antiviral defense in *N. benthamiana*. Although DRB2a/b and DRB5 could be providing increased protection to the apical meristem, especially in flowers to prevent the transmission of virus to the seed.

References

ASURMENDI, S.; BERG, R. H.; KOO, J. C. & BEAHY, R. N. Coat protein regulates

formation of replication complexes during tobacco mosaic virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, 101(5): 1415–1420, 2004.

BAULCOMBE, D. C.; CHAPMAN, S. & SANTA CRUZ, S. Jellyfish green fluorescent protein as a reporter for virus infections. *The Plant journal : for cell and molecular biology*, 7(6): 1045–1053, 1995.

CHAPMAN, S.; KAVANAGH, T. & BAULCOMBE, D. Potato virus X as a vector for gene expression in plants. *The Plant journal : for cell and molecular biology*, 2(4): 549–557, 1992.

COUTU, C.; BRANDLE, J.; BROWN, D.; BROWN, K.; MIKI, B.; SIMMONDS, J. & HEGEDUS, D. D. pORE: a modular binary vector series suited for both monocot and dicot plant transformation. *Transgenic research*, 16(6): 771–781, 2007.

CRUZ, S. & CHAPMAN, S. Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. *Proceedings of the National Academy of Sciences of the United States of America*, 93: 6286–6290, 1996.

CURTIN, S. J.; WATSON, J. M.; SMITH, N. A.; EAMENS, A. L.; BLANCHARD, C. L. & WATERHOUSE, P. M. The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS letters*, 582(18): 2753–2760, 2008.

DELERIS, A.; GALLEGO-BARTOLOME, J.; BAO, J.; KASSCHAU, K. D.; CARRINGTON, J. C. & VOINNET, O. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science (New York, N.Y.)*, 313(5783): 68–71, 2006.

DING, S-W. & VOINNET, O. Antiviral immunity directed by small RNAs. *Cell*, 130(3): 413–426, 2007.

DRUMMOND, A. J. & RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7: 214, 2007.

EAMENS, A.; WANG, M-B. SMITH; N. A. & WATERHOUSE, P. M. RNA Silencing in Plants: Yesterday, Today, and Tomorrow. *PLANT PHYSIOLOGY*, 147(2): 456–468, 2008.

EAMENS, A. L.; CURTIN, S. J. & WATERHOUSE, P. M. *Non Coding RNAs in Plants.* Springer Berlin Heidelberg, 2011.

EAMENS, A. L.; KIM, K. W.; CURTIN, S. J. & WATERHOUSE, P. M. DRB2 is required for microRNA biogenesis in Arabidopsis thaliana. *PloS one*, 7(4): p. e35933, 2012a.

EAMENS, A. L.; WOOK KIM, K. & WATERHOUSE, P. M. DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in Arabidopsis thaliana. *Plant signaling & behavior*, 7(10): 1224–1229, 2012b.

EARLEY, K. W.; HAAG, J. R.; PONTES, O.; OPPER, K.; JUEHNE, T.; SONG, K. & PIKAARD, C. S. Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant journal : for cell and molecular biology*, 45(4): 616–629, 2006.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5): 1792–1797, 2004.

FUSARO, A. F. MATTHEW, L.; SMITH, N. A.; CURTIN, S. J.; DEDIC-HAGAN, J.; ELLACOTT, G. A.; WATSON, J. M.; WANG, M-B.; BROSNAN, C.; CARROLL, B. J. & WATERHOUSE, P. M. RNA interference-inducing hairpin RNAs in plants act through the

viral defence pathway. *EMBO reports*, 7(11): 1168–1175, 2006.

GLEAVE, A. P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant molecular biology*, 20(6): 1203–1207, 1992.

HIRAGURI, A.; ITOH, R.; KONDO, N.; NOMURA, Y.; AIZAWA, D.; MURAI, Y.; KOIWA, H.; SEKI, M.; SHINOZAKI, K. & FUKUHARA, T. Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant molecular biology*, 57(2): 173–188, 2005.

HOFFER, P.; IVASHUTA, S.; PONTES, O.; VITINS, A.; PIKAARD, C.; MROCZKA, A.; WAGNER, N. & VOELKER, T. Posttranscriptional gene silencing in nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 108(1): 409–414, 2011.

JAKUBIEC, A.; YANG, S. W. & CHUA, N-H. Arabidopsis DRB4 protein in antiviral defense against Turnip yellow mosaic virus infection. *The Plant journal : for cell and molecular biology*, 69(1): 14–25, jan. 2012.

LINNIK, O.; LIESCHE, J.; TILSNER, J. & OPARKA, K. J. Unraveling the structure of viral replication complexes at super-resolution. *Frontiers in plant science*, 4: 6, 2013.

LLAVE, C. Virus-derived small interfering RNAs at the core of plant-virus interactions. *Trends in plant science*, 15(12): 701–707, 2010.

MANDADI, K. K. & SCHOLTHOF, K-B. G. Plant immune responses against viruses: how does a virus cause disease? *The Plant cell*, v. 25, n. 5, p. 1489–505, maio 2013.

MOLNAR, A.; MELNYK, C. & BAULCOMBE, D. C. Silencing signals in plants: a long journey for small RNAs. *Genome biology*, 12(1): 215, 2011.

MOURRAIN, P.; BÉCLIN, C; ELMAYAN, T.; FEUERBACH, F.; GODON, C.; MOREL, J. B.; JOUETTE, D.; LAOMBE, A. M.; NIKIC, S.; PICAULT, N.; RÉMOUÉ, K.; SANIAL, M.; VO, T. A. & VAUCHERET, H. Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, 101(5): 533–542, 2000.

NAKASUGI, K.; CROWHURST, R.; Bally, J. & WATERHOUSE, P. Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant Nicotiana benthamiana. *PloS one,* 9(3): e91776, 2014.

RAJA, P.; JACKEL, J. N.; LI, S.; HEARD, I. M. & BISARO, D. M. Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. *Journal of virology*, 88(5): 2611–2622, 2014.

SCHOELZ, J. E.; HARRIES, P. A. & NELSON, R. S. Intracellular transport of plant viruses: finding the door out of the cell. *Molecular plant*, 4(5): 813–831, 2011.

SCHWACH, F.; VAISTIJ, F. E.; JONES, L. & BAULCOMBE, D. C. An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for the activity but not the production of a systemic silencing signal. *Plant physiology*, 138(4): 1842–1852, 2005.

TILSNER, J.; LINNIK, O.; CHRISTENSEN, N. M.; BELL, K.; ROBERTS, I. M.; LACOMBE, C. & OPARKA, K. J. Live-cell imaging of viral RNA genomes using a

Pumilio-based reporter. *The Plant journal : for cell and molecular biology*, 57(4): 758–770, 2009.

TILSNER, J.; AMARI, K. & TORRANCE, L. Plasmodesmata viewed as specialised membrane adhesion sites. *Protoplasma*, 248(1): 39–60, 2011.

TILSNER, J.; LINNIK, O. WRIGHT, K. M.; BELL, K.; ROBERTS, A. G.; LACOMBE, C.; SANTA CRUZ, S. & OPARKA, K. J. The TGB1 movement protein of Potato virus X reorganizes actin and endomembranes into the X-body, a viral replication factory. *Plant physiology*, 158(3): 1359–70, 2012.

TILSNER, J.; LINNIK, O.; LOUVEAUX, M.; ROBERTS, I. M.; CHAPMAN, S. N. & OPARKA, K. J. Replication and trafficking of a plant virus are coupled at the entrances of plasmodesmata. *The Journal of cell biology*, 201(7): 981–995, 2013.

VERCHOT-LUBICZ, J.; TORRANCE, L.; SOLOVYEV, A. G.; MOROZOV, S. Y.; JACKSON, A. O. & GILMER, D. Varied movement strategies employed by triple gene block-encoding viruses. *Molecular plant-microbe interactions : MPMI*, 23(10): 1231–1247, 2010.

Legends of figures

Figure 1. Viral replication complexes formation in *N. benthamiana* leaves

CFP tagged to the overcoat protein of PVX appears cyan and it is diffuse throughout nuclei (n) and cytoplasm of *N. benthamiana* epidermal cells. In PVX infected cells occurs the formation of VRCs (v), in some cases close to the nucleus.

Figure 2. Dendrogram of DRB family in *A. thaliana* and *N. benthamiana.*

Phylogenetic analysis of amino acid sequences of genes containing double stranded RNA binding motifs in *A. thaliana* and *N. benthamiana*. CeRDE-4, a DRB from *C. elegans*, was used as the outgroup. The Bayesian method was used to construct the tree, with posterior probabilities showed above the branches.

Figure 3. Location of DRB1a, DRB1b and DRB4 in *N. benthamiana* leaves during PVX infection

A) In uninfected epidermal cells OFP tagged DRB1a, DRB1b and DRB4 localize to nuclei

(n). In some over-expressing cells DRB4-OFP appear also diffuse in cytoplasm. B) In infected cells DRB1a, DRB1b and DRB4 remain in nuclei. DRB4-OFP is diffuse throughout cytoplasm in over-expressing cells.

Figure 4. Cellular localization of DRB2a, DRB2b, DRB3a, DRB3b and DRB5 during PVX infection and formation of ring-like structures around TGB1 within VRCs. A) In uninfected epidermal cells OFP-tagged DRB2a, DRB2b, DRB3a, DRB3b and DRB5 proteins were diffuse throughout cytoplasm and nuclei (n). B) In infected cells, co-agro-infiltrated with PVX-CFP and DRB1a-YFP (as nuclear marker), these proteins are diminished in the nuclei and become more concentrated in cytoplasm, forming ring-like structures as VRCs. C) Co-expression of DRBs-OFP, TGB1-YFP in PVX-CFP infected cells. All DRBs forms ring-like structures around TGB1 within VRCs.

Figure 1

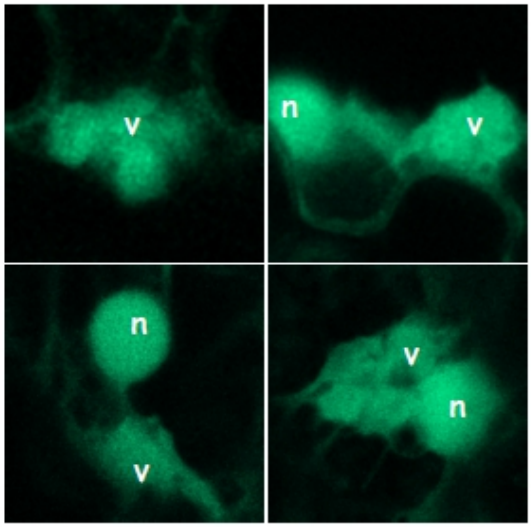

Figure 2

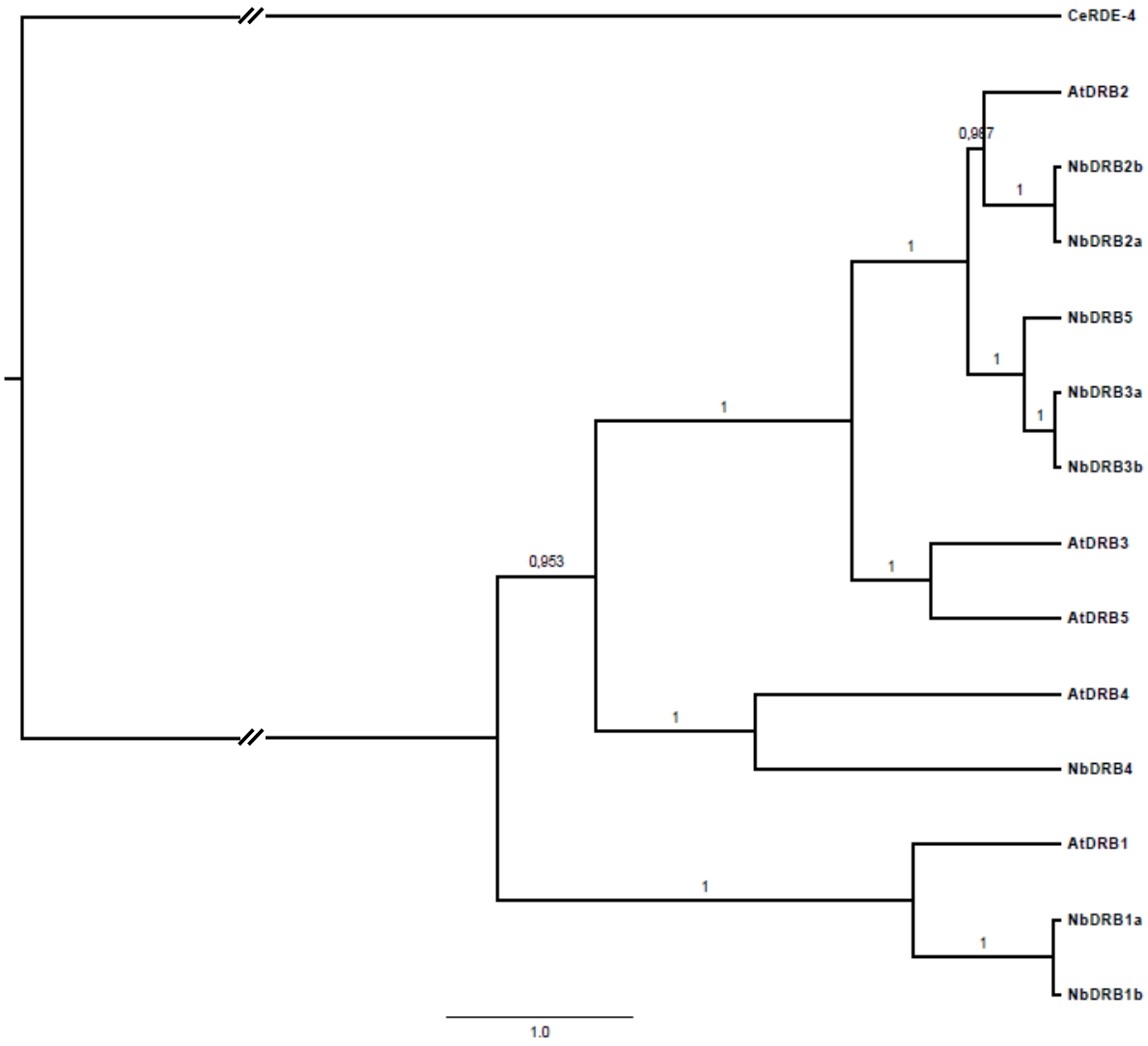
CeRDE-4
AtDRB2
0,987
NbDRB2b
1
NbDRB2a
1
NbDRB5
1
NbDRB3a
1
NbDRB3b
1
0,953
AtDRB3
1
AtDRB5
AtDRB4
1
NbDRB4
AtDRB1
1
NbDRB1a
1
NbDRB1b
1.0

Figure 3

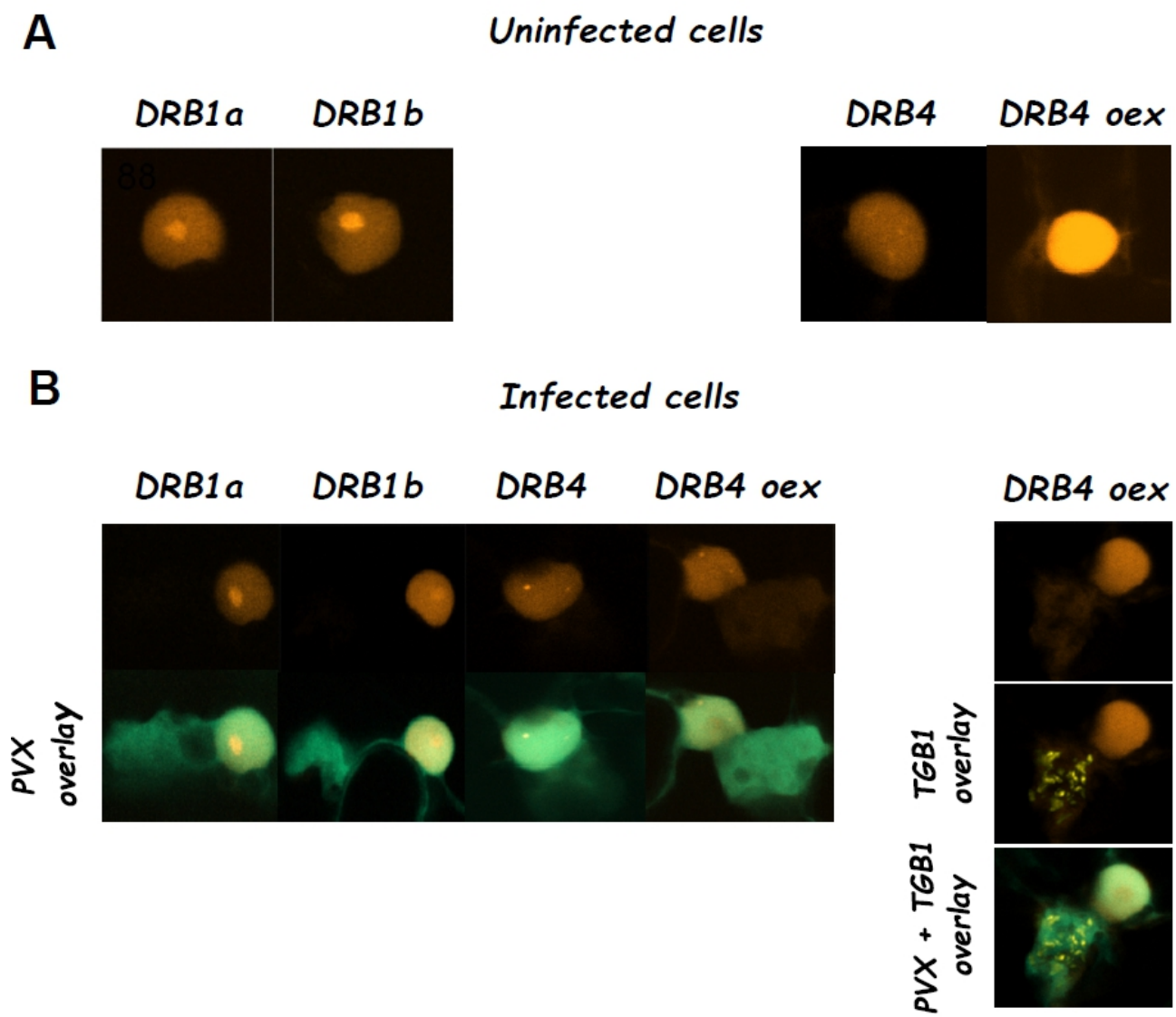
A
Uninfected cells
DRB1a
DRB1b
DRB4
DRB4 oex
B
Infected cells
DRB1a
DRB1b
DRB4
DRB4 oex
PVX overlay
DRB4 oex
TGB1 overlay
PVX + TGB1 overlay

Figure 4

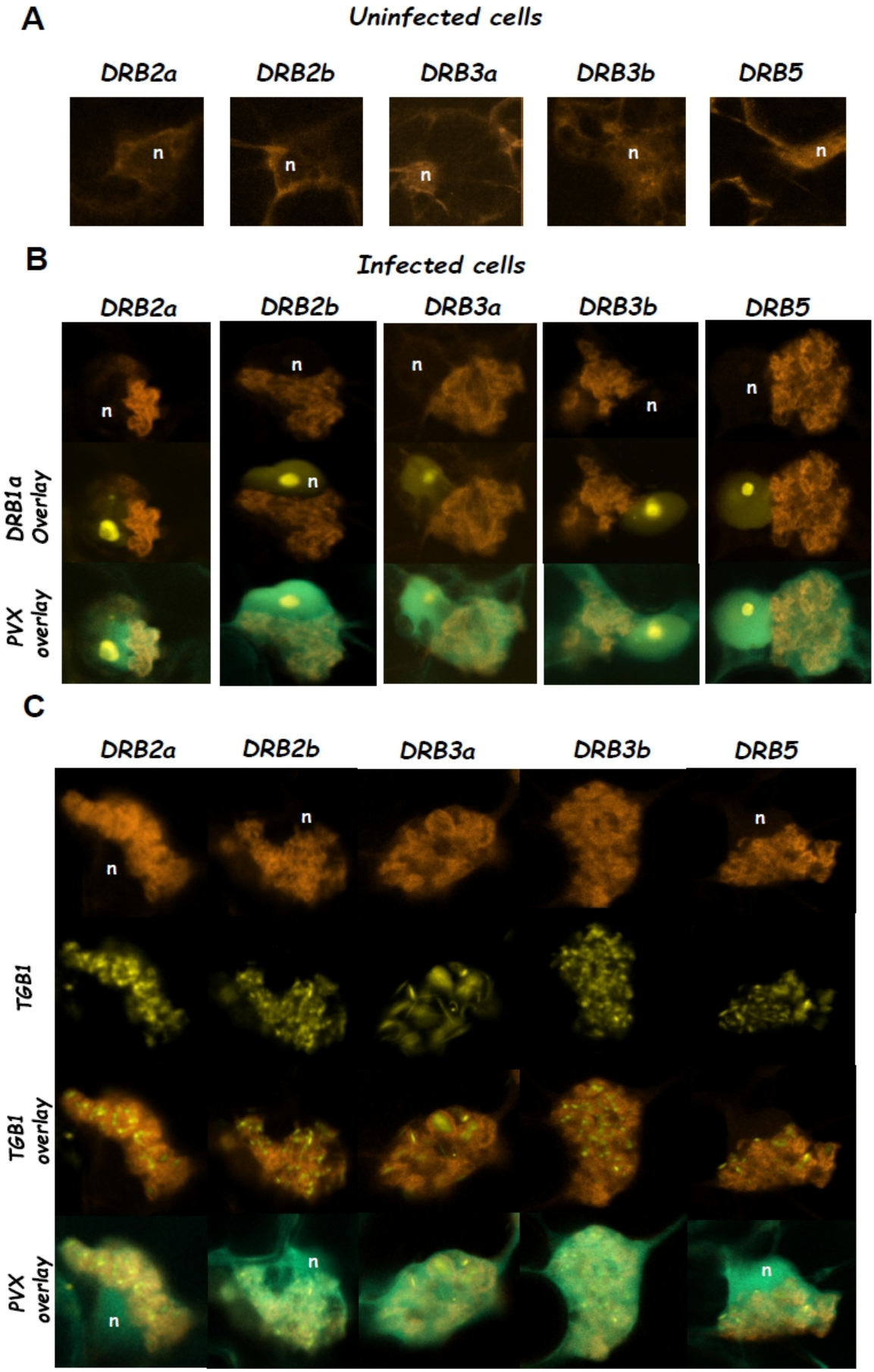
A
Uninfected cells
DRB2a
DRB2b
DRB3a
DRB3b
DRB5
n
B
Infected cells
DRB2a
DRB2b
DRB3a
DRB3b
DRB5
DRB1a Overlay
PVX overlay
C
DRB2a
DRB2b
DRB3a
DRB3b
DRB5
TGB1
TGB1 overlay
PVX overlay

Table S1

Primers used to amplify the *N. benthamiana* DRB genes for cloning.

Primer name	Primer sequence
NbDRB1abFw	5'-CACCATGGCGAAAAATGAGA-3'
NbDRB1aRv	5'-TTAAGAAATTTCTGCCTCTGTGTT-3'
NbDRB1bRv	5'-TTAAGAAACTTCTGCCTCTGTGTT-3'
NbDRB2abFw	5'-CACCATGTACAAGAACCAGTTACA-3'
NbDRB2abRv	5'-TCATAAACTGAGCTCCTCCAAGC-3'
NbDRB3abFw	5'-CACCATGTACAAAAACCAGCT-3'
NbDRB3aRv	5'-TCATATCTTTAGTTGTTTCAAATTCTCA-3'
NbDRB3bRv	5'-TCATATCTTTAGTGGTTTCAAATTCTCA-3'
NbDRB4Fw	5'-CACCATGTTCAAAACGAAGCTGC-3'
NbDRB4Rv	5'-TTAGCGATTTCGATTCAATTC-3'
NbDRB5Fw	5'-CACCATGTATAAGAACCAGCTAC-3'
NbDRB5Rv	5'-TCAGATCTTCAGCTGTTCCAATTT-3'

3. CONSIDERAÇÕES FINAIS E PESPECTIVAS

3.1 CONSIDERAÇÕES FINAIS

Todos os seres vivos podem ser infectados por vírus. Bactérias utilizam endonucleases que clivam o DNA viral. Mamíferos combatem a infecção com anticorpos que atuam em antígenos virais de maneira específica. Em plantas, o sistema de interferência por RNA (RNAi) parece ser a principal defesa antiviral (WATERHOUSE; FUSARO, 2006). Quando um vírus de DNA ou RNA infecta a célula vegetal, estruturas de dupla fita de RNA viral parcialmente complementares ou intermediários de dupla fita polimerizados tanto pela replicase viral (no caso de vírus de RNA) quanto pelas RDRs são utilizados como substratos para a produção dos vsiRNAs (LLAVE, 2010). Os vsiRNAs acoplados ao complexo RISC irão induzir o RNAi via clivagem do RNA viral (LLAVE, 2010). Todo esse processo também conta com a participação de proteínas como AGOs, DCLs e DRBs. Dentre essas proteínas pouco se sabe da função das DRBs com relação ao mecanismo do RNAi em plantas. A DRB1 se localiza exclusivamente no núcleo e é necessária para o correto processamento de miRNAs (EAMENS et al., 2009). De maneira similar, DRB4 é importante para o processamento dos *trans-acting* siRNAs (NAKAZAWA et al., 2007). Adicionalmente a DRB1, em meristema apical de *A. thaliana,* DRB2 atua na biogênese de miRNAs, em um processo em que também participam DRB3 e DRB5, provavelmente em uma rota de silenciamento via inibição da tradução (EAMENS et al., 2012a; EAMENS et al., 2012b). Por muito tempo somente DRB4 tinha sido relacionada a resposta antiviral, participando tanto na biogênese dos vsiRNAs quanto na inibição do acúmulo da CP por meio da diminuição da estabilidade dessa proteína ou o bloqueio de sua tradução (JAKUBIEC; YANG; CHUA, 2012). Recentemente foi demonstrado o papel de

DRB3 na rota de defesa antiviral mediada por metilação frente a uma infecção por Geminivírus (RAJA et al., 2014). Neste trabalho foi demonstrado, pela primeira vez que as proteínas DRB2, DRB3 e DRB5 se localizam nos VRCs durante uma infecção por vírus de RNA. Estudos sugerem que as DRBs atuam na biogênese de pequenos RNAs ao formarem heterodímeros com as DCLs, através dos seus domínios de ligação à RNA de dupla fita, que também promovem a interação proteína-proteína (HIRAGURI et al., 2005b). Além de interagir com DCL3, DRB3 também forma complexos proteícos com AGO4 no núcleo que atuam na metilação do DNA viral (RAJA et al., 2014). Uma vez que DRBs podem se ligar tanto a RNA de dupla fita formado durante a replicação de vírus de RNA, quanto a estruturas de fita dupla do RNA genômico viral (JAKUBIEC; YANG; CHUA, 2012b) e também a proteínas, ainda resta elucidar se DRB2, DRB3 e DRB5 de *N. benthamiana* estão interagindo com o RNA viral ou com as proteínas do PVX.

Eventos de integração de genes virais em eucariotos têm sido demonstrados para uma variedade de organismos como fungos, insetos e mamíferos (BOCK; STOYE, 2000; STOYE, 2001). Esses processos envolvem geralmente a endogenização de genes de retrovírus e vírus de DNA. Na maioria dos casos as sequências virais sofrem mutações ao longo do tempo e se tornam inativas. Contudo, em raras situações, elas podem manter a atividade e inclusive possuir papel importante na biologia celular de seu hospedeiro (BOCK; STOYE, 2000). Entretanto, não há relatos de retrovírus que infectam plantas. Um estudo recente identificou genes de vírus de RNA não retroviral no genoma de diferentes espécies vegetais, similares aos que já haviam sido encontrados em mamíferos, fungos e insetos (KOONIN, 2010; CHIBA et al., 2011). Apesar da função desses genes não estar bem esclarecida, há dados que indicam uma possível atividade antiretroviral. No genoma de fungos foram encontrados genes da CP de vírus da família *Totiviridae* que possuem

mutações no sítio da proteína responsável pela atividade de remoção do Cap 5', porém ainda são capazes de codificar a maior parte das proteínas (TAYLOR; BRUENN, 2010). Isso indica que essas proteínas podem atuar como inibidor negativo da encapsidação viral, como já demonstrado anteriormente em plantas transgênicas superexpressando a CP (REIMANN-PHILIPP, 1998). Em contraste, no fungo *Debaryomyces hansenii* foi encontrada uma RdRp viral que manteve os resíduos catalíticos intactos, sugerindo uma possível contribuição na resposta antiviral ao polimerizar RNAs de dupla fita a partir de transcritos virais, ativando o mecanismo de RNAi (KOONIN, 2010). Esses resultados e outras observações de genes de vírus não retrovirais integrados em genomas de eucariotos levantaram a hipótese de que a razão principal para esses genes se manterem relativamente conservados ao longo do tempo se deve a aquisição de imunidade antiviral para os seus hospedeiros (BERTSCH et al., 2009; TAYLOR & BRUENN, 2010; KOONIN, 2010). Nesse trabalho foi identificado a inserção do RNA1 do CMV no genoma de soja. Ao contrário do demonstrado em outros trabalhos, essa inserção leva a expressão de um grampo que, após processado por DCLs, gera pequenos RNAs homólogos ao RNA1 do CMV, que codifica para a replicase viral. Similarmente ao observado em plantas transgênicas que expressam grampos homólogos em sequência ao genoma viral, as plantas de soja que possuem a inserção do CMV podem apresentar resistência frente a uma infecção por esse vírus (BONFIM et al., 2007).

3.2 PERSPECTIVAS

Para confirmar uma provável resistência antiviral promovida pela inserção do CMV, construções plasmidiais contendo a sequência do CMV fusionada a uma GFP poderão ser expressas de maneira transiente em soja, comparando cultivares que contém a inserção e cultivares que não possuem ela. Desta forma, poderá ser medido tanto a expressão do mRNA da GFP por RT-qPCR tanto quanto pela fluorescência via microscopia confocal entre plantas das diferentes cultivares verificando se há de fato uma redução desse mRNA em plantas contendo a integração do RNA1 do CMV, indicando a ativação do mecanismo de RNAi.

Com relação ao trabalho do Capítulo II sobre a interação entre as proteínas DRBs e PVX, construções plasmidiais foram realizadas contendo outras proteínas do mecanismo de RNAi de *N. benthamiana* fusionadas a OFP (AGO1, AGO2, RDR1, RDR2, RDR5, RDR6 e SGS3). Poderão ser realizados os mesmos experimentos com o objetivo de verificar a interação entre essas proteínas e as proteínas ou RNA viral. Também foram construidos outros vetores plasmidiais contendo as sequências das DRBs fusionados a um peptídeo da hemaglutinina do vírus da influenza humano (HA) e as sequências das outras proteínas do mecanismo de RNAi fusionados ao peptídeo FLAG, para ensaios de imunoprecipitação, com o objetivo de verificar a interação dessas proteínas durante uma infecção por PVX em *N. benthamiana*. Também será realizado, em outros experimentos de imunoprecipitação, a extração do RNA dos imunoprecipitados com o objetivo de verificar a presença do RNA viral via técnica de *Northern blot* ou via PCR com primers específicos do PVX.

4. REFERÊNCIAS BIBLIOGRÁFICAS DA INTRODUÇÃO E CONSIDERAÇÕES FINAIS

ANDERSSON, J. O.; DOOLITTLE, W. F. & NESBØ, C. L. Genomics. Are there bugs in our genome? *Science (New York, N.Y.)*, 292(5523): 1848–1850, 2001.

ASURMENDI, S.; BERG, R. H.; KOO, J. C. & BEAHY, R. N. Coat protein regulates formation of replication complexes during tobacco mosaic virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, 101(5): 1415–1420, 2004.

ATABEKOV, J. G.; RODIONOVA, N. P.; KARPOVA, O. V.; KOZLOVSKY, S. V. & POLJAKOV, V. Y. The movement protein-triggered in situ conversion of potato virus X virion RNA from a nontranslatable into a translatable form. *Virology*, 271(2): 259–263, 2000.

BAMUNUSINGHE, D.; HEMENWAY, C. L.; NELSON, R. S.; SANDERFOOT, A. A. YE, C. M.; SILVA, M. A. T.; PAYTON, M. & VERCHOT-LUBICZ, J. Analysis of potato virus X replicase and TGBp3 subcellular locations. *Virology*, 393(2): 272–285, 2009.

BAULCOMBE, D. C.; CHAPMAN, S. & SANTA CRUZ, S. Jellyfish green fluorescent protein as a reporter for virus infections. *The Plant journal : for cell and molecular biology*, 7(6): 1045–1053, 1995.

BEJARANO, E. R. KHASHOGGI, A.; WITTY, M. & LICHTENSTEIN, C. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 93(2): 759–764, 1996.

BERTSCH, C.; BEUVE, M.; DOLJA, V. V.; WIRTH, M.; PELSY, F.; HERRBACH, E. & LEMAIRE, O. Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biology direct*, v. 4, p. 21, jan. 2009.

BOARI, A. J., MACIEL-ZAMBOLIM, E., CARVALHO, M. G. & ZERBINI, F. M. Caracterização biológica e molecular de isolados do *Cucumber mosaic virus* provenientes de oito espécies vegetais. *Fitopatologia Brasileira,* 25: 49-58, 2000.

BOCK, M. & STOYE, J. P. Endogenous retroviruses and the human germline. *Current opinion in genetics & development*, v. 10, n. 6, p. 651–5, dez. 2000.

BONFIM, K.; FARIA J. C.; NOGUEIRA, E. O. P. L.; MENDES, E. & ARAGÃO, F. J. L. RNAi-mediated resistance to *Bean golden mosaic virus* in genetically engineered common bean (Phaseolus vulgaris). *Molecular plant-microbe interactions,* 20(6): 717–726, 2007.

BRANDIZZI, F. & WASTENEYS, G. O. Cytoskeleton-dependent endomembrane organization in plant cells: an emerging role for microtubules. *The Plant journal : for cell and molecular biology*, 75(2): 339–349, 2013.

BURCH-SMITH, T. M.; ANDERSON, J. C.; MARTIN, G. B. & DINESH-KUMAR, S. P. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant journal : for cell and molecular biology*, 39(5): 734–746, 2004.

CHAPMAN, S.; KAVANAGH, T. & BAULCOMBE, D. Potato virus X as a vector for

gene expression in plants. *The Plant journal : for cell and molecular biology*, 2(4): 549–557, 1992.

CHIBA, S.; KONDO, H.; TANI, A.; SAISHO, D.; SAKAMOTO, W.; KANEMATSU, S. & SUZUKI, N. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS pathogens*, 7(7): e1002146, 2011.

CURTIN, S. J.; WATSON, J. M.; SMITH, N. A.; EAMENS, A. L.; BLANCHARD, C. L. & WATERHOUSE, P. M. The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS letters*, 582(18): 2753–2760, 2008.

DELERIS, A.; GALLEGO-BARTOLOME, J.; BAO, J.; KASSCHAU, K. D.; CARRINGTON, J. C. & VOINNET, O. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science (New York, N.Y.)*, 313(5783): 68–71, 2006.

DIAZ-PENDON, J. A., LI, F., LI, W. X. & DING, S. W. Suppression of antiviral silencing by *Cucumber mosaic virus* 2b protein in *Arabidopsis* is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *The Plant Cell,* 19: 2053–2063, 2007.

DING, S-W. & VOINNET, O. Antiviral immunity directed by small RNAs. *Cell*, 130(3): 413–426, 2007.

DUNNING HOTOPP, J. C.; CLARK, M. E.; OLIVEIRA, D. C. S. G.; FOSTER, J. M.; FISCHER, P.; MUÑOZ TORRES, M. C.; GIEBEL, J. D.; KUMAR, N.; ISHMAEL, N.; WANG, S.; INGRAM, J.; NENE, R. V.; SHEPARD, J.; TOMKINS, J.; RICHARDS, S.; SPIRO, D. J.; GHEDIN, E.; SLATKO, B. E.; TETTELIN, H. & WERREN, J. H. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science (New York, N.Y.)*, 317(5845): 1753–1756, 2007.

EAMENS, A. L.; SMITH, N.; CURTIN, S. J.; WANG, M-B. & WATERHOUSE, P. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA (New York, N.Y.)*, 15(12): 2219-2235, 2009

EAMENS, A. L.; CURTIN, S. J. & WATERHOUSE, P. M. *Non Coding RNAs in Plants.* Springer Berlin Heidelberg, 2011.

EAMENS, A. L.; WOOK KIM, K. & WATERHOUSE, P. M. DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in Arabidopsis thaliana. *Plant signaling & behavior*, 7(10): 1224–1229, 2012a.

EAMENS, A. L.; KIM, K. W.; CURTIN, S. J. & WATERHOUSE, P. M. DRB2 is required for microRNA biogenesis in Arabidopsis thaliana. *PloS one*, 7(4): p. e35933, 2012b.

EIRAS, M., BOARI, A. J., COLARICCIO, A., CHAVES, A. L. R., BRIONES, M. R. S., FIGUEIRA, A. R. & HARAKAVA, R. Characterization of isolates of the cucomovirus *Cocumber mosaic virus* present in Brazil. Journal of Plant Pathology, 86 (1): 61-69, 2004.

ESCOBAR, N. M.; HAUPT, S.; THOW, G.; BOEVINK, P.; CHAPMAN, S. & OPARKA, K. High-throughput viral expression of cDNA-green fluorescent protein fusions reveals novel subcellular addresses and identifies unique proteins that interact with plasmodesmata. *The Plant cell*, 15(7): 1507–1523, 2003.

FIRE, A.; XU, S.; MONTGOMERY, M. K.; KOSTAS, S. A.; DRIVER, S. E & MELLO, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669): 806–811, 1998.

FUSARO, A. F. MATTHEW, L.; SMITH, N. A.; CURTIN, S. J.; DEDIC-HAGAN, J.; ELLACOTT, G. A.; WATSON, J. M.; WANG, M-B.; BROSNAN, C.; CARROLL, B. J. & WATERHOUSE, P. M. RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO reports*, 7(11): 1168–1175, 2006.

GARCIA-RUIZ, H.; TAKEDA, A.; CHAPMAN, E. J.; SULLIVAN, C. M.; FAHLGREN, N.; BREMPELIS, K. J. & CARRINGTON, J. C. *Arabidopsis* RNA-Dependent RNA Polymerases and Dicer-Like Proteins in Antiviral Defense and Small Interfering RNA Biogenesis during *Turnip Mosaic Virus* Infection. *Plant Cell*, 22(2): 481–496, 2010.

GEUKING, M. B.; WEBER, J.; DEWANNIEUX, M.; GORELIK, E.; HEIDMANN, T.; HENGARTNER, H.; ZINKERNAGEL, R. M. & HANGARTNER, L. Recombination of Retrotransposon and Exogenous RNA Virus Results in Nonretroviral cDNA Integration. *Science*, 323(5912): 393–396, 2009.

GOLDSTEIN, B. Cytological Study of Living Cells of Tobacco Plants Affected with Mosaic Disease. *Bulletin of the Torrey Botanical Club*, 51(6): 261, 1924.

GOODIN, M. M.; ZAITLIN, D.; NAIDU, R. A. & LOMMEL, S. A. *Nicotiana benthamiana*: Its History and Future as a Model for Plant – Pathogen Interactions. *Molecular plant-microbe interactions*, 21(8): 1015–1026, 2008.

HARPER, G.; HULL, R.; LOCKHART, B. & OLSZEWSKI, N. Viral sequences integrated into plant genomes. *Annual review of phytopathology*, 40: 119–136, 2002.

HEINLEIN, M. Plant virus replication and movement. *Virology*, 479-480: 657–671, 2015.

HEUER, H. & SMALLA, K. Horizontal gene transfer between bacteria. *Environmental biosafety research*, 6 (1): 3–13, 2007.

HIRAGURI, A.; ITOH, R.; KONDO, N.; NOMURA, Y.; AIZAWA, D.; MURAI, Y.; KOIWA, H.; SEKI, M.; SHINOZAKI, K. & FUKUHARA, T. Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant molecular biology*, 57(2): 173–188, 2005.

HORIE, M.; HONDA, T.; SUZUKI, Y.; KOBAYSHI, Y.; DAITO, T.; OSHIDA, T.; IKUTA, K.; JERN, P.; GOJOBORI, T.; COFFIN, J. M. & TOMONOGA, K. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature*, 463(7277): 84–87, 2010.

HUISMAN, M. J.; LINTHORST, H. J.; BOL, J. F. & CORNELISSEN, J. C. The complete nucleotide sequence of *potato virus X* and its homologies at the amino acid level with various plus-stranded RNA viruses. *The Journal of general virology*, 69: 1789–1798, 1988.

JAKUBIEC, A.; YANG, S. W. & CHUA, N-H. Arabidopsis DRB4 protein in antiviral defense against Turnip yellow mosaic virus infection. *The Plant journal : for cell and molecular biology*, 69(1): 14–25, jan. 2012.

JAMALKANDI, S. A. & MASOUDI-NEJAD, A. Reconstruction of *Arabidopsis thaliana* fully integrated small RNA pathway. *Functional & integrative genomics*, 9(4): 419–432, 2009.

KAWAKAMI, S.; WATANABE, Y. & BEACHY, R. N. Tobacco mosaic virus infection spreads cell to cell as intact replication complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16): 6291–6296, 2004.

KONDO, N.; NIKOH, N. IJICHI, N.; SHIMADA, M. & FUKATSU, T. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proceedings of the National Academy of Sciences of the United States of America*, 99(22): 14280–14285, 2002.

KOONIN, E. V; ARAVIND, L. & KONDRASHOV, A S. The impact of comparative genomics on our understanding of evolution. *Cell*, 101(6): 573–576, 2000.

KOONIN, E. V. Taming of the shrewd: novel eukaryotic genes from RNA viruses. *BMC biology*, 8: 2, 2010.

KOSHIMIZU, Y. & IIZUKA, N. Soybean stunt disease. *Ann. Phytopathol. Soc. Japan,* 23: 27, 1958.

KUMAGAI, M. H.; DONSON, J.; DELLA-CIOPPA, G.; HARVEY, D.; HANLEY, K. & GRILL, L. K. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proceedings of the National Academy of Sciences*, 92(5): 1679–1683, 1995.

KUNZ, H. & SCHÖB, C. Silencing of transgenes introduced into leaves by agroinfiltration: a simple, rapid method for investigating sequence requirements for gene silencing. *Molecular and General Genetics,* 256(5): 581–585, 1997.

KURIHARA, Y.; TAKASHI, Y.; WATANABE, Y. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA (New York, N.Y.)*, 12(2): 206–212, 2006.

LALIBERTÉ, J.-F. & SANFAÇON, H. Cellular remodeling during plant virus infection. *Annual review of phytopathology*, 48: 69–91, 2010.

LINNIK, O.; LIESCHE, J.; TILSNER, J. & OPARKA, K. J. Unraveling the structure of viral replication complexes at super-resolution. *Frontiers in plant science*, 4: 6, 2013.

LLAVE, C. Virus-derived small interfering RNAs at the core of plant-virus interactions. *Trends in plant science*, 15(12): 701–707, 2010.

LUCAS, W. J. Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. *Virology*, 344(1): 169–184, 2006.

MANDADI, K. K. & SCHOLTHOF, K-B. G. Plant immune responses against viruses: how does a virus cause disease? *The Plant cell*, v. 25, n. 5, p. 1489–505, maio 2013.

MARCO, A. & MARÍN, I. Retrovirus-like elements in plants. *Recent Res. Devel. Plant Sci.,* 661(2), 2005.

MARTELLI, G. P. & RUSSO, M. Plant virus inclusion bodies. *Advances in virus research*, 21: 175–266, 1977.

MILLAR, A. A. & WATERHOUSE, P. M. Plant and animal microRNAs: similarities and differences. *Functional & integrative genomics*, 5(3): 129–135, 2005.

MOISSIARD, G. & VOINNET, O. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51): 19593–19598, 2006.

MOLNÁR, A.; CSORBA, T.; LAKATOS, L.; VÁRALLYAY, É.; LACOMME, C. & BURGYÁN, J. Plant virus-derived small interfering RNAs originate predominantly from

highly structured single-stranded viral RNAs. *Journal of Virology*, 79(12): 7812–7818, 2005.

MOLNAR, A.; MELNYK, C. & BAULCOMBE, D. C. Silencing signals in plants: a long journey for small RNAs. *Genome biology*, 12(1): 215, 2011.

MORISSETTE, G. & FLAMAND, L. Herpesviruses and chromosomal integration. *Journal of virology*, 84(23): 12100–12109, 2010.

NAIM, F.; NAKASUGI, K. CROWHURST, R. N.; HILARIO, E.; ZWART, A. B.; HELLENS, R. P.; TAYLOR, J. M.; WATERHOUSE, P. M. & WOOD, C. C. Advanced engineering of lipid metabolism in Nicotiana benthamiana using a draft genome and the V2 viral silencing-suppressor protein. *PloS one*, 7(12): e52717, 2012.

NAKASUGI, K.; CROWHURST, R.; Bally, J. & WATERHOUSE, P. Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant Nicotiana benthamiana. *PloS one,* 9(3): e91776, 2014.

NAKAZAWA, Y.; HIRAGURI, A.; MORIYAMA, H. & FUKUHARA, T. The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. *Plant molecular biology*, 63(6): 777–785, 2007.

NAPOLI, C.; LEMIEUX, C. & JORGENSEN, R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell*, 2(4): 279–289, 1990.

NISHIKIORI, M.; DOHI, K.; MORI, M.; MESHI, T.; NAITO, S. & ISHIKAWA, M. Membrane-bound tomato mosaic virus replication proteins participate in RNA synthesis and are associated with host proteins in a pattern distinct from those that are not membrane bound. *Journal of virology*, 80(17): 8459–8468, 2006.

PALUKAITIS, P.; ROOSSINCK, M. J.; DIETZGEN, R. G. & FRANCKI, R. I. B. Cucumber mosaic virus. In: *Advances in Virus Research*, K. Maramorosch, F.A. Murphy, and A.J. Shatkin, ed. San Diego, CA: Academic Press, pp. 281–348, 1992.

QI, X.; BAO, F. S. & XIE, Z. Small RNA deep sequencing reveals role for Arabidopsis thaliana RNA-dependent RNA polymerases in viral siRNA biogenesis. *PloS one*, 4(3): e4971, 2009.

RAJA, P.; JACKEL, J. N.; LI, S.; HEARD, I. M. & BISARO, D. M. Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. *Journal of virology*, 88(5): 2611–2622, 2014.

ROOSSINCK, M. J., ZHANG, L. & HELLWALD, K. H. Rearrangements in the 5' nontranslated region and phylogenetic analyses of *Cucumber mosaic virus* RNA 3 indicate radial evolution of three subgroups. *Journal of Virology* 73: 6752-6758, 1999.

SCHOELZ, J. E.; HARRIES, P. A. & NELSON, R. S. Intracellular transport of plant viruses: finding the door out of the cell. *Molecular plant*, 4(5): 813–831, 2011.

SIMON, A. E., ROOSINCK, M. J. & HAVELDA, Z. Plant virus satellite and defective interfering RNAs: New Paradigms for a New Century. *Annu. Rev. Phytopathol.,* 42: 415–437, 2004.

STOYE, J. P. Endogenous retroviruses: still active after all these years? *Current biology*, 11(22): R914–R916, 2001.

SU, S., LIU, Z., CHEN, C., ZHANG, Y., WANG, X., ZHU, L., MIAO, L., WANG, X. C. & YUAN, M. *Cucumber mosaic virus* movement protein severs actin filaments to increase the plasmodesmal size exclusion limit in tobacco. *The Plant Cell*, 22: 1373–1387, 2010.

TAKAHASHI, K., TANAKA, T., IIDA, W. & TSUDA, Y. Studies on virus diseases and causal viruses of soybean in Japan. *Bull. Tohoku Natl. Agric. Exp. Stn.,* 62: 1–130, 1980.

TAKAHASHI, K., NAGASAWA, J., TANAKA, T. & IIZUKA, N. Reaction of soybean varieties to the strains of soybean mosaic virus and soybean stunt virus in Japan. *Misc. Publ. Tohoku Natl. Agric. Exp. Stn.*, 7: 1–35, 1987.

TAYLOR, D. J.; LEACH, R. W. & BRUENN, J. Filoviruses are ancient and integrated into mammalian genomes. *BMC evolutionary biology*, 10(1): 193, 2010.

TILSNER, J.; AMARI, K. & TORRANCE, L. Plasmodesmata viewed as specialised membrane adhesion sites. *Protoplasma*, 248(1): 39–60, 2011.

TILSNER, J.; LINNIK, O.; LOUVEAUX, M.; ROBERTS, I. M.; CHAPMAN, S. N. & OPARKA, K. J. Replication and trafficking of a plant virus are coupled at the entrances of plasmodesmata. *The Journal of cell biology*, 201(7): 981–995, 2013.

VERCHOT-LUBICZ, J.; YE, C-M. & BAMUNUSINGHE, D. Molecular biology of potexviruses: recent advances. *The Journal of general virology*, 88: 1643–1655, 2007.

VERCHOT-LUBICZ, J.; TORRANCE, L.; SOLOVYEV, A. G.; MOROZOV, S. Y.; JACKSON, A. O. & GILMER, D. Varied movement strategies employed by triple gene block-encoding viruses. *Molecular plant-microbe interactions : MPMI*, 23(10): 1231–1247, 2010.

VAN REGENMORTEL, M. H. V., FAUQUET, C. M., BISHOP, D. H. L., CARSTENS, E. B., ESTES, M. K., LEMON, S. M., MANILOFF, J., MAYO, M. A., MCGEOCH, D. J., PRINGLE, C. R. & WICKNER, R. B. Virus Taxonomy Classification and Nomenclature of Viruses. *Seventh Report of the International Committee on Taxonomy of Viruses*. Academic Press, California, USA. 2000

VOINNET, O.; LEDERER, C. & BAULCOMBE, D. C. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell*, 103(1): 157–167, 2000.

WANG, A. Dissecting the Molecular Network of Virus-Plant Interactions : The Complex Roles of Host Factors. *Annual Review of Phytopathology*, 53(3): 1–22, 2015.

WANG, M. B.; ABBOTT, D. C. & WATERHOUSE, P. M. A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to *barley yellow dwarf virus*. *Molecular plant pathology*, 1(6): 347–356, 2000.

ZAKI, E. A. Plant retroviruses: structure, evolution and future applications. *African Journal of Biotechnology*, 2(6): 136–139, 2003.

ZHANG, X.; YUAN, Y-R.; PEI, Y.; LIN, S-S.; TUSCHL, T.; PATEL, D. J. & CHUA, N-H. *Cucumber mosaic virus*-encoded 2b suppressor inhibits *Arabidopsis* Argonaute1 cleavage activity to counter plant defense. *Genes & development*, 20(23): 3255–3268, 2006.

ZITTER, T. A. & MURPHY, J. F. Cucumber mosaic. Disponível em: <http://www.apsnet.org/edcenter/intropp/lessons/viruses/Pages/Cucumbermosaic.aspx> Acesso em: 1 de Julho de 2015.

5.0 APÊNDICE

5.1 TRABALHOS CIENTÍFICOS PUBLICADOS.

5.1.1 APÊNDICE I

Genetics and Molecular Biology, 35, 1 (suppl), 292-303 (2012)

www.sbg.org.br

Research Article

Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries

Lorrayne Gomes Molina[1,2], Guilherme Cordenonsi da Fonseca[1], Guilherme Loss de Morais[1], Luiz Felipe Valter de Oliveira[1,2], Joseane Biso de Carvalho[1], Franceli Rodrigues Kulcheski[1] and Rogerio Margis[1,2,3]

[1]*Centro de Biotecnologia e PPGBCM, Laboratório de Genomas e Populações de Plantas, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*
[2]*Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*
[3]*Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

Abstract

A large number of small RNAs unrelated to the soybean genome were identified after deep sequencing of soybean small RNA libraries. A metatranscriptomic analysis was carried out to identify the origin of these sequences. Comparative analyses of small interference RNAs (siRNAs) present in samples collected in open areas corresponding to soybean field plantations and samples from soybean cultivated in greenhouses under a controlled environment were made. Different pathogenic, symbiotic and free-living organisms were identified from samples of both growth systems. They included viruses, bacteria and different groups of fungi. This approach can be useful not only to identify potentially unknown pathogens and pests, but also to understand the relations that soybean plants establish with microorganisms that may affect, directly or indirectly, plant health and crop production.

Key words: next generation sequencing, small RNA, siRNA, molecular markers.

Introduction

Until recently, analysis of the microbial diversity in environmental samples was conducted only after isolation, culture and identification of microorganisms and subsequent sequencing of cloned libraries (Cardenas and Tiedje, 2008). However, these conventional methods are limited to the minority of species that can be cultured (Chistoserdova, 2010). New culture-independent methods, such as metagenomics and metatranscriptomics, have been developed (Xu, 2006; Cardenas and Tiedje, 2008; Adams *et al.*, 2009; Warnecke and Hess, 2009; Chistoserdova, 2010). These methods refer to studies of the collective set of genomes and transcriptomes of mixed microbial communities and may be applied to the exploration of all microorganisms that reside in marine environments, soils, human and animal clinical samples, sludge, polluted environment, and plants (Kent *et al.*, 2007; Zoetendal *et al.*, 2008; Adams *et al.*, 2009; Al Rwahnih *et al.*, 2009; Poretsky *et al.*, 2009; Shi *et al.*, 2009; Desai *et al.*, 2010; Gifford *et al.*, 2010; Roossinck *et al.*, 2010). The metagenomic approaches, including metatranscriptomics, involve the sequencing of random DNA or RNA-derived complementary DNA (cDNA) profiles and subsequent determination of taxonomic diversity and prospective genes related to response to environmental conditions (Rosen *et al.*, 2009). With the advent of new sequencing technologies (high-throughput sequencing), more data can be generated in a relatively short time, in a practical and cost-effective way (Creer *et al.*, 2010). Moreover, this approach allows the direct sequencing of DNA or cDNA, avoiding any cloning bias and leading to large-scale studies (Adams *et al.*, 2009).

Metagenomic analysis of RNA deep-sequencing data has been used in plant disease diagnostics, such as in grapevine (Al Rwahnih *et al.*, 2009; Coetzee *et al.*, 2010), sweet potato (Kreuze *et al.*, 2009), tomato and *Liatris spicata* (Adams *et al.*, 2009). Coupling of metagenomics with pyrosequencing in these studies has also allowed the detection of bacterial and fungal RNA, suggesting that this approach can contribute to massive identification of crop-associated microbiota, including pathogenic, symbiotic, and free-living organisms.

Send correspondence to Rogerio Margis. Centro de Biotecnologia e PPGBCM, Laboratório de Genomas e Populações de Plantas, Prédio 43431, Universidade Federal do Rio Grande do Sul, Caixa Postal 15005, 91501-970 Porto Alegre, RS, Brazil. E-mail: rogerio.margis@ufrgs.br.

5.1.2 APÊNDICE II

OPEN ACCESS Freely available online

PLOS ONE

RFMirTarget: Predicting Human MicroRNA Target Genes with a Random Forest Classifier

Mariana R. Mendoza[1], Guilherme C. da Fonseca[2], Guilherme Loss-Morais[2], Ronnie Alves[3], Rogerio Margis[2], Ana L. C. Bazzan[1]*

1 Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **2** Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **3** Instituto Tecnológico Vale Desenvolvimento Sustentável, Belém, Pará, Brazil

Abstract

MicroRNAs are key regulators of eukaryotic gene expression whose fundamental role has already been identified in many cell pathways. The correct identification of miRNAs targets is still a major challenge in bioinformatics and has motivated the development of several computational methods to overcome inherent limitations of experimental analysis. Indeed, the best results reported so far in terms of specificity and sensitivity are associated to machine learning-based methods for microRNA-target prediction. Following this trend, in the current paper we discuss and explore a microRNA-target prediction method based on a random forest classifier, namely RFMirTarget. Despite its well-known robustness regarding general classifying tasks, to the best of our knowledge, random forest have not been deeply explored for the specific context of predicting microRNAs targets. Our framework first analyzes alignments between candidate microRNA-target pairs and extracts a set of structural, thermodynamics, alignment, seed and position-based features, upon which classification is performed. Experiments have shown that RFMirTarget outperforms several well-known classifiers with statistical significance, and that its performance is not impaired by the class imbalance problem or features correlation. Moreover, comparing it against other algorithms for microRNA target prediction using independent test data sets from TarBase and starBase, we observe a very promising performance, with higher sensitivity in relation to other methods. Finally, tests performed with RFMirTarget show the benefits of feature selection even for a classifier with embedded feature importance analysis, and the consistency between relevant features identified and important biological properties for effective microRNA-target gene alignment.

Citation: Mendoza MR, da Fonseca GC, Loss-Morais G, Alves R, Margis R, et al. (2013) RFMirTarget: Predicting Human MicroRNA Target Genes with a Random Forest Classifier. PLoS ONE 8(7): e70153. doi:10.1371/journal.pone.0070153

Editor: Mikael Boden, The University of Queensland, Australia

Received March 10, 2013; **Accepted** June 16, 2013; **Published** July 26, 2013

Funding: This research was partially supported by the Brazilian funding agencies Coordination for the Improvement of Higher Education Personnel (CAPES) and National Council for Scientific and Technological Development (CNPq). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bazzan@inf.ufrgs.br

Introduction

MicroRNAs (miRNAs) are non-coding RNAs of approximately 22 nucleotides (nt) in length that act as an important post-transcriptional mechanism of gene expression regulation via translational repression or degradation of target mRNAs [1,2]. In both animals and plants, miRNAs are formed after a longer primary transcript (pri-miRNA) by two sequential cleavages, mediated, respectively, by a nuclear and a cytoplasmic RNase III. These processing steps yield a 60–70 nt stem-loop miRNA precursor (pre-miRNA) and next, after the latter is exported to the cytoplasm, a structure of two single RNA strands that corresponds to the mature miRNA, namely the miRNA:miRNA* duplex.

Due to miRNAs participation in important metabolic processes, such as developmental timing, growth, apoptosis, cell proliferation, defense against viruses [3–5], and more recently in tumorigenesis, either as tumor suppressors or oncogenes [6], great efforts have been devoted for the identification of novel miRNAs and targets. Despite the advances in deep sequencing approaches, the use of computational tools is still important for analysis and interpretation of data, among which machine learning (ML) algorithms have been prominent. This approach consists in using known positive and negative examples of miRNA-mRNA associations to train a classifier to distinguish, for instance, real pre-miRNAs from pseudo pre-miRNAs, based on a set of descriptive features extracted from the examples. Among the most commonly applied ML algorithms, one may highlight the use of support vector machine (SVM) [7,8], random forest [9] and naïve Bayes [10] classifiers.

Following this direction, ML-based methods can help in the prediction of miRNA target genes, generating hypotheses regarding miRNA function and potential miRNA:target interactions. However, this is considered to be a more difficult problem, mostly because i) it is hard to distinguish true miRNA-mRNAs hybrids given the millions of possible miRNA-gene combinations and ii) there is still very limited knowledge about the basic mechanisms of microRNA target recognition [11]. Primarily, the interaction of a miRNA and its target occurs by complementarity of their nucleotide sequences, as shown in Fig. 1. Nonetheless, while in plants miRNAs bind their targets with (near) perfect complementarity and mostly in their open read frames [12], in animals, miRNAs sequences have a partial complementarity to their targets and the hybridization may occur in either 3′ untranslated region

6.0 CURRICULUM VITÆ

DA FONSECA, G. C.

1. DADOS PESSOAIS
Nome: Guilherme Cordenonsi da Fonseca
Local e data de nascimento: Porto Alegre, RS, Brasil, 21/09/1985
E-mail: guicf13@gmail.com

2. FORMAÇÃO
Mestrado em Biologia Celular e Molecular (Universidade Federal do Rio Grande do Sul - UFRGS) - 2009 a 2011.
Bacharelado em Ciencias Biologicas com Ênfase Molecular, Celular e Funcional (Universidade Federal do Rio Grande do Sul - UFRGS) - 2003 a 2008.

3. ESTÁGIOS
Bolsa de Iniciacao Cientifica – Orientador: Rogerio Margis – Resumo: Uso da metodologia RNAi para a producao de plantas resistentes a insetos. - BIC/PROPESQ – 01/04/2006 a 31/12/2006.

Bolsa de Iniciacao Cientifica – Orientador: Rogerio Margis – Resumo: Uso da metodologia RNAi como uma alternativa de tratamento para o controle de *Rhipicephalus (Boophilus) microplus*. - PIBIC/CNPq – 01/03/2007 a 31/07/2008

Monitor da disciplina BIO 07001 (Genetica I) – Professora: Marion Schiengold - 2007/1 e 2007/2

4. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS
DA FONSECA, G. C.; SEIXAS, A.; VALLE, D.; LINSS, J. G. B.; TERMIGNONI, C & MARGIS, R. Análise do efeito do RNAi como silenciador do gene da quitina sintase em *Boophilus microplus*. In: XVIII Salão de Iniciação Científica – UFRGS, 2006.

DA FONSECA, G. C.; SEIXAS, A.; VALLE, D.; LINSS, J. G. B.; TERMIGNONI, C. & MARGIS, R. Análise do efeito do RNAi como silenciador do gene da quitina sintase em *Rhipicephalus* (*Boophilus) microplus*. In: XIX Salão de Iniciação Científica – UFRGS, 2007.

DA FONSECA, G. C. & MARGIS, R. Clonagem de dois genes de quitina sintase de *Magnaporthe grisea* em vetor para transformação de arroz. In: XX Salão de Iniciação Científica – UFRGS, 2008.

DA FONSECA, G. C.; DE OLIVEIRA, L. F. V.; LOSS-MORAIS, G.; ABDELNOR, R. V.; NEPOMUCENO, A. L.; FARINELLI, L.; MARGIS, R. Unusual RNA plant virus integration in the soybean genome leads to the production of small RNAs. In: IV Simpósio Brasileiro de Genétia Molecular de Plantas, Bento Gonçalves, 2013.

5. ARTIGOS COMPLETOS PUBLICADOS EM PERIÓDICOS

BALDISSERA, R.; BACH, E.; LIMA, R. P.; MENEGASSI, A.; PIOVESAN, A. R. & FONSECA, G. C. Distribution of Understorey Web Building Spiders Along an Interface Area of Araucaria Forest and Pinus Plantation in Southern Brazil. *Neotropical Biology and Conservation*, 3: 3-8, 2008.

MOLINA, L. G.; DA FONSECA, G. C.; LOSS-MORAIS, G; DE OLIVEIRA, L. F. V.; CARVALHO, J. B.; KULCHESKI, F. R.; MARGIS, R. Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries. *Genetics and Molecular Biology.* 35: 292-303, 2012.

MENDOZA, M. R.; DA FONSECA, G. C.; LOSS-MORAIS, G; ALVES, R.; MARGIS, R.; BAZZAN, A. L. C. RFMirTarget: Predicting Human MicroRNA Target Genes with a Random Forest Classifier. *Plos One*, 8: e70153, 2013.