



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

**Identificação da integração do vírus do mosaico do pepino no
genoma da soja e seu reconhecimento pelo sistema de produção
de pequenos RNAs**

Guilherme Cordenonsi da Fonseca

Porto Alegre, março de 2011

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

**Identificação da integração do vírus do mosaico do pepino no
genoma da soja e seu reconhecimento pelo sistema de produção
de pequenos RNAs**

Guilherme Cordenonsi da Fonseca

Dissertação submetida ao Programa de Pós-graduação
em Biologia Celular e Molecular do Centro de
Biotecnologia da UFRGS como requisito parcial para a
obtenção do título de Mestre

Orientador: Prof. Dr. Rogerio Margis

Porto Alegre
Março de 2011

INSTITUIÇÃO E FONTES FINANCIADORAS

Este trabalho foi desenvolvido no Laboratório de Genomas e Populações de Plantas do Centro de Biotecnologia da UFRGS.

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, Ministério da Ciência e Tecnologia.

AGRADECIMENTOS

Agradeço, primeiramente aos meus colegas Felipe e Guilherme Loss pelo companheirismo e camaradagem, além das ajudas com os testes de bioinformática, sem os quais esse trabalho jamais seria realizado. Aos demais colegas do LGPP, Júlio, Andréia, Cláudia, Fernanda, Maurício, Fran, Josi, Lorryne, Nicole e Bruno pela amizade e pelo ambiente excepcional do laboratório. A professora Dra. Márcia Margis, e aos demais colegas do NGFP, pela disposição, acolhimento e auxílio em muitos momentos ao longo deste Mestrado. Ao professor Dr. Gialcarlo Pasquali pela grande ajuda com as revisões da redação científica e deste trabalho. E ao professor Dr. Rogerio Margis por estar comigo praticamente desde o início da minha carreira científica sendo um extraordinário orientador em todos os momentos.

E um agradecimento especial a minha família aos meus irmãos, Gus e Gabi, e aos meus pais por sempre fazerem o máximo pela minha educação e principalmente por serem os pilares da minha formação pessoal.

ÍNDICE

LISTA DE ABREVIATURAS.....	6
RESUMO.....	10
ABSTRACT.....	11
1. REVISÃO BIBLIOGRÁFICA.....	12
1.1 A Soja.....	12
1.1.1 Histórico da soja.....	12
1.1.2 Importância e usos da soja.....	13
1.2 O Vírus do Mosaíco do Pepino.....	14
1.3 RNA de interferência.....	16
1.3.1 Os diversos tipos de pequenos RNAs de interferência.....	16
1.3.2 Pequenos RNAs de interferência virais em plantas.....	20
1.4 Transferência Horizontal de Material Genético.....	21
1.5 Montagem <i>de novo</i> de sequências de RNA por meio de siRNAs.....	23
2. OBJETIVOS.....	25
2.1 Objetivo Geral.....	25
2.2 Objetivos Específicos.....	25
3. MATERIAL E MÉTODOS.....	26
3.1 Material biológico para as bibliotecas de sRNAs.....	26
3.1.1 Experimento de estresse abiótico.....	26
3.1.2 Experimento de estresse biótico.....	27
3.1.3 Diferentes estádios de desenvolvimento da soja.....	28
3.1.4 Flor.....	28
3.2 Extração de RNA.....	28
3.3 Sequenciamento em larga escala dos sRNAs.....	29
3.4 Montagem <i>de novo</i> do RNA 1 do CMV e dos RNAs do BPMV.....	29
3.5 Desenho de oligonucleotídeos para PCR, RT-qPCR e sequenciamento.....	30
3.6 Extração de DNA.....	31
3.7 PCR para a confirmar a integração do RNA 1 do CMV no genoma da soja.....	31
3.8 Sequenciamento de DNA.....	32
3.9 Síntese de DNA complementar (cDNA).....	32
3.10 RT-qPCR.....	33
3.11 Caracterização dos siRNAs produzidos pelo RNA 1 do CMV e pelos RNAs do BPMV.....	34
3.12 Análise do perfil de expressão dos siRNAs produzidos pelo RNA 1 do CMV.....	34
3.13 Análise dos siRNAs diferencialmente expressos.....	34
3.14 Predição dos alvos dos siRNAs diferencialmente expressos.....	35
3.15 Busca de estruturas em grampo para o RNA 1 do CMV.....	35
4. RESULTADOS.....	37
4.1 Sequenciamento em larga escala das bibliotecas de sRNAs e montagem <i>de novo</i> das sequências do RNA 1 do CMV e dos RNAs 1 e 2 do BPMV.....	37
4.2 Confirmação da integração do RNA 1 do CMV em diferentes cultivares de soja por PCR.....	38

4.3 Confirmação do produto de PCR por sequenciamento.....	39
4.4 Caracterização dos siRNAs do RNA 1 do CMV e do RNA 1 e RNA 2 do BPMV.....	39
4.5 Distribuição dos siRNAs ao longo da sequência do RNA 1 do CMV..	42
4.6 Análise dos siRNAs do CMV diferencialmente expressos.....	49
4.7 Prováveis alvos dos siRNAs diferencialmente expressos.....	52
4.8 Análise da expressão do RNA e dos siRNAs do RNA 1 do CMV.....	53
5. DISCUSSÃO.....	56
5.1 Integração do CMV no genoma da soja (<i>Glycine max</i>).....	56
5.2 Perfil dos siRNAs do RNA 1 do CMV e dos RNAs 1 e 2 do BPMV.....	59
5.3 Expressão do RNA 1 e dos siRNAs derivados do CMV.....	62
5.4 Expressão diferencial dos siRNAs e possíveis alvos.....	63
5.5 Possíveis benefícios da integração do CMV ao genoma da soja.....	65
6. CONCLUSÕES E PERSPECTIVAS.....	69
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	71

LISTA DE ABREVIATURAS

AGO – argonauta, proteína envolvida na RNAi, presente no complexo RISC, com atividade de RNase e/ou ligação de RNAs

APK – proteína quinase da *Arabidopsis thaliana*, do inglês *Arabidopsis thaliana* protein kinase

BDV - vírus da doença de Borna, do inglês, “Borna disease virus”

BLAST – “Basic Local Alignment Search Tool”, ferramenta de busca ou procura de sequências nucleotídicas ou peptídicas por alinhamento básico local

BPMV – vírus da mancha da vagem do feijoeiro, do inglês, “Bean pod mottle virus”

°C – graus Celsius

CaMV - vírus do mosaico da couve-flor, do inglês, “Cauliflower mosaic virus”

cDNA – DNA complementar

CDS – sequência codificante, do inglês, “coding sequence”

CMV - vírus do mosaico do pepino, do inglês, “Cucumber mosaic virus”

DCL – do tipo “dicer”, do inglês, Dicer-Like (proteína envolvida na RNAi, com atividade de RNase e/ou ligação de dsRNAs)

DNA – ácido desoxirribonucleico, do inglês, “desoxiribonucleic acid”

DRM – metiltransferase rearranjada por domínios, do inglês, “Domains Rearranged Methyltransferase”

DRD – defeitivo em metilação de DNA direcionada por RNA, do inglês, “Defective In RNA-Directed DNA Methylation”

DSRC - controle de raiz sensível à seca, do inglês, “drought-sensitive root control”

dsRNA – RNA de dupla fita, do inglês, “double strand RNA”

DSRT - raiz tratada sensível à seca, do inglês, “drought-sensitive root treated”

DTRC - controle de raiz tolerante à seca, do inglês, “drought-tolerant root control”

DTRT - raiz tratada tolerante à seca, do inglês, “drought-tolerant root treated”

EBLN - elemento do tipo Borna N endógeno, do inglês, “endogenous Borna-like N”

ENV - envelope

hai – horas após a inoculação

hc-siRNA – siRNA heterocromáticos, do inglês, “heterochromatic siRNA”

HEN – ativador de HUA (ativador de agamous), do inglês, HUA Enhancer

IAP - partícula-A intracisternal, do inglês, Intracisternal A-particle

LCMV - vírus da coriomeningite linfocítica, do inglês, Lymphocytic choriomeningitis virus

LINE - elementos nucleares interestaçados longos, do inglês, “long interspersed nuclear elements”

LTR – repetições terminais longas, do inglês, “long terminal repeat”

M – molar

µg - micrograma

mg - miligrama

µL – microlitro

M-MLV – vírus da leucemia de murinos Moloney, do inglês, “Moloney Murine Leukemia virus”

miRNA – micro RNA

MITE - Elemento transponível de repetições invertidas em miniatura, do inglês, “miniature inverted-repeat transposable element”

mRNA – RNA mensageiro

NAT – anti-senso natural, do inglês, “natural antisense”

nat-siRNA – siRNA natural, do inglês, “natural siRNA”

NCBI – “National Center for Biotechnology Information”

ncRNA – RNA não codificador, do inglês, “non-coding RNA”

NIRV – vírus de RNA integrado não retroviral, do inglês, “non-retroviral integrated RNA virus”

nt - nucleotídeo

PCR – reação em cadeia da polimerase, do inglês, “polymerase chain reaction”

PDS - fito-eno dessaturase, do inglês, phytoene desaturase

RNA Pol – RNA polimerase

pre-miRNAs – precursor do miRNA, do inglês, “miRNA precursor”

pri-miRNAs - miRNA primário, do inglês, “primary miRNA”

pri-ta-siRNA - ta-siRNA primário, do inglês, “primary ta-siRNA”

PTGS - silenciamento gênico pós-transcricional, do inglês, “Post-transcriptional gene silencing”

RdDM - metilação de DNA dependente de RNA, do inglês, “RNA dependent DNA methylation”

RdRp – RNA polimerase RNA dependente, do inglês, RNA-dependent RNA polymerase

RDR - RNA polimerase RNA dependente, do inglês, RNA-dependent RNA polymerase

RISC - complexo indutor do silenciamento por RNA, do inglês, “RNA-Induced Silencing Complex”

RNA – ácido ribonucleico, do inglês, “ribonucleic acid”

RNAi – RNA de interferência

RRLC - controle de folhas resistentes à ferrugem, do inglês, “rust-resistant leaf control”

RRLI - folhas inoculadas resistentes à ferrugem, do inglês, “rust-resistant leaf inoculated”

rRNA – RNA ribossomal

RSLC - controle de folhas suscetíveis à ferrugem, do inglês, “rust-susceptible leaf control”

RSLI - folhas inoculadas suscetíveis à ferrugem, do inglês, “rust-susceptible leaf inoculated”

RT-qPCR - PCR quantitativa em tempo real, do inglês, “Real-time quantitative PCR”

satRNA – RNA satélite, do inglês, “satellite RNA”

siRNA – pequeno RNA interferente, do inglês, “short interference RNA”

SOAP – programa de alinhamento de pequenos oligonucleotídeos, do inglês, “short oligonucleotide alignment program”

sRNA – pequeno RNA, do inglês, “small RNA”

ta-siRNA – siRNA transatuante, do inglês, “*trans*-acting siRNA”

tBLASTn – algoritmo da ferramenta BLAST que inspeciona um banco de dados de sequências codificantes, utilizando sequências peptídicas como isca.

TMV – vírus do mosaico do fumo, do inglês, “Tobacco mosaic virus”

tRNA – RNA transportador

TRV - vírus do chocoalho do fumo, do inglês, “Tobacco rattle virus”

U - uridina

unc-22 – descordenado – 22, do inglês, “Uncoordinated - 22”

vsiRNA – siRNA viral, do inglês, “viral siRNA”

XRN – exorribunuclease

RESUMO

A soja (*Glycine max*) é uma das culturas mais importantes do mundo, os seus grãos servem tanto para a alimentação quanto para a extração de óleo para a fabricação do biodiesel. O vírus do mosaico do pepino (CMV, do inglês, “Cucumber mosaic virus”) é um vírus de RNA, patogênico a diversas plantas. O RNA de interferência é um sistema de silenciamento de RNA presente na maioria dos eucariotos no qual precursores de RNA de dupla fita (dsRNA) são processados em pequenos RNAs (sRNAs) de 21-24 nucleotídeos (nt), que podem regular a atividade de genes, elementos genéticos e vírus de uma maneira sequência específica. A integração de vírus de DNA e de retrovírus no genoma do hospedeiro já é bem conhecida tanto para sistemas eucarióticos quanto para procarióticos. Mais recentemente, foi observada a integração de vírus de RNA não retrovirais (NIRVs) em mamíferos. O presente trabalho é o primeiro a demonstrar tal evento no genoma de plantas. A partir das sequências dos sRNAs de 19-24 nt de 15 bibliotecas de sRNAs sequenciados de amostras de tecidos de soja, foram montadas sequências contíguas (“contigs”) pelo programa SOAP, algumas das quais apresentaram homologia de sequência ao RNA 1 do CMV. Por montagem *de novo* desses contigs foi obtida uma sequência de 3.092 nt do RNA 1 do CMV, presente em todas as bibliotecas pesquisadas de pelo menos cinco cultivares diferentes de soja. A presença dessa sequência foi confirmada em outras sete cultivares, exceto em “Willians”. Foi observada uma maior presença de sRNAs derivados do CMV senso do que anti-senso nas 15 bibliotecas sequenciadas. Os sRNAs de 22-nt foram os mais abundantes. Para o vírus da mancha da vagem do feijoeiro (BPMV, do inglês “Bean pod mottle virus”) presente em uma das bibliotecas, os sRNAs de 21-nt e 22-nt representaram em torno de 80% do total de sRNAs. Foram encontrados sRNAs que variaram sob estresse biótico (*Phakospora pachyrhizi*) e abiótico (seca) e entre diferentes cultivares. A expressão do RNA 1 aumentou nas plantas sob estresse. Provavelmente o evento de integração ocorreu via recombinação à um retrotransposon.

Palavras-chave: *Glycine max*; *Cucumber mosaic virus*; RNAi, transferência horizontal, NIRV.

ABSTRACT

The Soybean (*Glycine max*) is one of the world's most important crops, its seeds are used both as food and for the extraction of oil to manufacture biodiesel. The *Cucumber mosaic virus* (CMV) is a pathogenic RNA virus of plants. The RNA interference is a system of RNA silencing present in most eukaryotes in which precursors of double-stranded RNA (dsRNA) are processed into small RNAs (sRNAs) of 21-24 nucleotides (nt), which can regulate the activity of genes, genetic elements and virus in a sequence-specific manner. The integration of DNA virus and retrovirus into the host genome is well known both for prokaryotic and eukaryotic systems. The integration of non-retroviral RNA virus (NIRVs) in mammals was previously observed, but the present work is the first to demonstrate such an event in a plant genome. The sequences of the sRNAs ranging from 19 to 24 nt, in 15 libraries of sRNAs sequenced from samples of soybean tissues, were assembled in contigs by the program SOAP, with some preentering sequence homology to the RNA 1 of CMV. By *de novo* assembling of these contigs it was obtained a sequence of 3,092 nt of the CMV RNA 1, present in all libraries surveyed in at least five different varieties of soybeans. The presence of this sequence was confirmed by PCR in seven other cultivars, but in "Williams". We observed a greater presence of sRNAs derived from CMV of sense orientation than antisense in the 15 libraries sequenced. The 22-nt sRNAs were the most abundant. For the *Bean pod mottle virus* (BPMV), present in one of the libraries, the 21 and 22 nt sRNAs were represented by around 80% of all sRNAs. The sRNAs were found varying under biotic stress (*Phakospora pachyrhizi*) and abiotic (drought) and among different cultivars. RNA 1 expression increased in plants under stress. Probably the integration event occurred via recombination of a retrotransposon.

Keywords: *Glycine max*; Cucumber mosaic virus, RNAi, horizontal transfer, NIRV.

1. REVISÃO BIBLIOGRÁFICA

1.1. A Soja

1.1.1. Histórico da soja

A soja é uma planta pertencente à família Fabaceae (leguminosas) do gênero *Glycine*, cuja principal espécie cultivável é a espécie *G. max*. Esta leguminosa é originária da China. Contudo, sua forma cultivada é muito diferente das suas ancestrais, rasteiras, que se desenvolveram ao longo da costa leste chinesa em terras baixas e úmidas, junto a juncos nas proximidades de lagos e rios. As primeiras citações à planta ocorreram no período de 2883 a 2838 a.C., quando a soja era considerada um grão sagrado ao lado do arroz, do trigo, da cevada e do milho, sendo provavelmente domesticada entre os séculos XII e XI a.C. (HYMOWITZ, 1970).

Apesar de ser consumida e conhecida pela civilização oriental por milhares de anos a soja só foi introduzida na Europa no final do século XV nos jardins botânicos da Inglaterra, França e Alemanha. De acordo com os primeiros relatos a produção de soja foi restrita à China até aproximadamente 1894, no término da guerra entre China e Japão, quando os japoneses começaram a importar a soja como fertilizante (EMBRAPA, 2010a). Nos Estados Unidos variedades de soja originárias da China foram introduzidas em 1875 (HYMOWITZ & HARLAN, 1983). No Brasil as mais antigas citações encontradas na literatura sobre a soja se referem a experiências feitas na Bahia por Gustavo Dutra, em 1882, e em Campinas por Daffert, em 1892 (LEAL, 1967). No Rio Grande do Sul a soja teria sido cultivada pela primeira vez no ano de 1900 em Dom Pedrito e em 1901 foi publicado o primeiro trabalho com a soja no Estado pelo professor Guilherme Minssen (COSTA, 1996).

Somente na segunda década do século XX o teor do óleo e as proteínas do grão de soja começaram a despertar o interesse das indústrias mundiais. Entretanto, as tentativas de introdução do cultivo comercial da soja em países da Europa como Inglaterra, Alemanha e Rússia não foram bem sucedidas devido às

condições climáticas desfavoráveis (EMBRAPA, 2010a).

1.1.2. Importância e usos da soja

A soja é uma leguminosa de alto valor nutritivo. O seu grão possui grandes quantidades de princípios ativos e elementos dietéticos, representando uma das melhores fontes vegetais nutritivas por conter cerca de 38% de proteínas, 19% de lipídeos e 23% de carboidratos (EMBRAPA, 2010b). O grão de soja é muito importante como fonte de proteínas para muitos alimentos humanos como massas, carne de soja, cereais, bebidas e alimentos dietéticos além de ser a proteína predominante do farelo para a alimentação animal (GUNSTONE, 2005). O óleo refinado de soja é utilizado para a fabricação de vários produtos alimentícios como óleos de cozinha, margarinas, temperos, pastas e gorduras vegetais e diversos outros tipos de produtos como desinfetantes, inseticidas, plastificadores e tintas (EMBRAPA, 2010c). Mais recentemente o óleo de soja também vem sendo testado por instituições de pesquisa para a fabricação de biodiesel como fonte de energia alternativa ao petróleo (EMBRAPA, 2010d).

Graças à diversidade em qualidade e quantidade de derivados de soja e também pela capacidade do seu óleo substituir o uso de gorduras animais para gorduras vegetais mais saudáveis a soja é uma das mais importantes mercadorias do mundo. A produção mundial de soja mundial na safra 2008/2009 foi de 210,6 milhões de toneladas em uma área plantada de 96,3 milhões de hectares. À América do Sul corresponde aproximadamente metade dessa produção, com 95,3 milhões de toneladas. Os Estados Unidos é o maior produtor mundial desse grão (80,5 milhões de toneladas), seguido pelo Brasil (57,1 milhões de toneladas). O Rio Grande do Sul é o terceiro Estado do Brasil que mais produziu grãos de soja na safra 2008/2009, com um total de 7,7 milhões de toneladas, sendo que o Mato Grosso é o maior produtor nacional seguido do Paraná (EMBRAPA, 2010e).

Portanto, a soja é um dos mais importantes itens da balança comercial brasileira. A exportação do grão rendeu ao Brasil cerca de US\$ 11 bilhões em 2009 o maior valor obtido no mundo (ABIOVE, 2010).

1.2. O Vírus do Mosaico do Pepino

O *Cucumber mosaic virus* (CMV), ou vírus do mosaico do pepino, é uma espécie do gênero *Cucumovirus*. O CMV é um vírus de RNA patogênico de plantas, com partículas virais isométricas de 30 nm de diâmetro e pertencente à família Bromoviridae (EIRAS *et al.*, 2004). É um dos mais comuns dentre os vírus de plantas com substancial importância econômica, podendo infectar 1.200 espécies diferentes de plantas distribuídas em mais de 100 famílias (VAN REGENMORTEL *et al.* 2000). O genoma do CMV é composto por três RNAs de orientação positiva, designados RNA 1 (~3.350 nucleotídeos, ou nt), RNA 2 (~3.050 nt), RNA 3 (~2.200 nt) e dois RNAs subgenômicos denominados RNA 4 e RNA 4A. Este vírus pode conter ainda RNAs satélites (SIMON *et al.*, 2004, ZITTER *et al.*, 2009).

O RNA 1 do CMV codifica a proteína 1A, que é um componente do complexo da replicase viral, contendo dois domínios funcionais: o domínio N-terminal, que possui atividade de metiltransferase, metilando o final 5' dos RNAs virais, e o domínio C-terminal, que possui atividade de helicase, auxiliando o processo de replicação 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 do CMV codifica duas proteínas: a proteína 2A, que faz parte do complexo da replicase viral e tem atividade de RNA polimerase RNA-dependente, isto é, que 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. Esta proteína 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 do CMV também codifica duas proteínas: a proteína 3A e a proteína da capa. A proteína 3A é a proteína responsável pelo movimento do vírus entre as células do hospedeiro já que aumenta o limite de exclusão por tamanho do plasmodesma (SU *et al.*, 2010). Mutações nesta proteína revelaram diferenças

hospedeiro-específicas na eficiência do movimento viral. A proteína da capa é expressa pelo RNA subgenômico 4 e 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 (ssRNAs) que variam de 330 a 405 nt e, aparentemente não codificam proteínas. A maioria dos satRNAs do CMV atenua os sintomas induzidos pelo vírus. Entretanto, já foi demonstrado que, para tomate e tabaco, a presença de satRNAs induz a clorose ou a necrose no hospedeiro (SIMON *et al.*, 2004).

O CMV replica-se no citoplasma das células vegetais, onde a partir da ação da replicase, fitas de RNA de orientação negativa são sintetizadas e servem como molde para a síntese de fitas de orientação positiva que são codificadas pela maquinaria de síntese proteica do hospedeiro. O movimento viral entre as células dos hospedeiros é realizado através do plasmodesma e o vírus também usa o floema para a movimentação de longa distância na planta (ZITTER *et al.*, 2009).

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 plantas invasoras das lavouras contribuem para a distribuição ampla do vírus ao servirem de reservatórios onde o CMV mantém-se até o início da temporada de plantação das culturas de importância econômica (ZITTER *et al.*, 2009). Além disso, mais de 80 espécies de afídeos, também conhecidos como pulgões, são capazes de transmitir o vírus de uma maneira não persistente. O inseto primeiramente alimenta-se de uma planta infectada, de onde adquire o vírus e em seguida alimenta-se de uma outra planta saudável, repassando o CMV a este novo potencial hospedeiro (MAUCK *et al.*, 2010). Foi descrito que abóboras (*Cucurbita pepo*) infectadas por CMV produzem uma quantidade maior de substâncias voláteis, capazes de atrair duas espécies de afídeos (*Myzus persicae* e *Aphis gossypii*), demonstrando que o CMV é capaz de alterar o fenótipo de seus hospedeiros para facilitar a sua dispersão como já foi demonstrado para outros patógenos e parasitas (MAUCK *et al.*, 2010).

As linhagens de CMV estão divididas em dois grupos baseado em

diferenças nos dados sorológicos, mapeamento de peptídeos, hibridização de ácidos nucleicos e similaridade nas sequências de nucleotídeos (PALUKAITIS *et al.*, 1992).

Linhagens de CMV capazes de infectar a soja foram descritas 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 à produção de vagens, mas também pode resultar em nenhum sintoma (TAKAHASHI *et al.*, 1987). Apesar de no Brasil já ter sido relatada 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 este vírus.

1.3. RNA de interferência

1.3.1. Os diversos tipos de pequenos RNAs de interferência

Pequenos RNAs não codificadores de proteínas (ncRNAs, do inglês, “non-protein coding small RNAs”) são componentes regulatórios essenciais identificados na maioria dos eucariotos, e que possuem papéis diversificados e complexos na biologia da célula. As primeiras evidências do fenômeno de interferência por RNA (RNAi, do inglês, “RNA interference”) foi descoberto em petúnia quando, com a intenção de alterar a cor da flor, plantas transgênicas com cópias adicionais do gene codificador da chalcona sintase, essencial para a pigmentação, apresentavam flores com perda de coloração indicando que a atividade desse gene foi suprimida (NAPOLI *et al.*, 1990). Contudo, somente oito anos depois evidências de que este processo era acionado por RNAs de dupla fita (dsRNAs, do inglês, “double-stranded RNAs”) 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, o qual codifica uma proteína do miofilamento apresentaram fenótipo idêntico ao de mutantes com perda-de-função para esse gene, assim como uma redução drástica de seu mRNA (FIRE *et al.*, 1998). Atualmente as principais etapas do processo pelo qual dsRNAs suprimem a atividade de genes foram elucidadas. Em plantas, animais e na maioria dos fungos moléculas de dsRNA endógenas ou exógenas são reconhecidas por enzimas RNase do tipo III, conhecidas por “dicer” ou “dicer-like” (DCL), as quais clivam as moléculas de dsRNAs em pequenos RNAs interferentes de dupla fita (siRNAs, do inglês, “small interfering RNAs”) de 21 a 24 nt (MILLER & WATERHOUSE, 2005). Uma destas fitas é acoplada ao complexo indutor do silenciamento por RNA (RISC, do inglês, RNA Induced Silencing Complex)) que promove geralmente a clivagem ou o bloqueio da tradução de RNAs que possuem homologia pela sequência do siRNA (JAMALKANDI & MASOUDI-NEJAD, 2009).

Uma série de sRNAs de diferentes tamanhos tem sido identificada em plantas nos últimos anos, provenientes de diferentes tipos de moléculas precursoras e rotas de biosíntese. Os micro RNAs (miRNAs) são uma das classes de siRNAs mais estudadas ultimamente e tem importantes funções nas plantas como o controle do desenvolvimento foliar e floral, a formação e a iniciação da raiz lateral e a resposta a estresses bióticos e abióticos como alta salinidade, frio e seca (JAMALKANDI & MASOUDI-NEJAD, 2009).

Os miRNAs são sintetizados por genes que, ao serem transcritos geram RNAs de fita simples os quais possuem, em sua estrutura, regiões complementares capazes de se parear em dupla fita na forma de grampos. Estes transcritos, de aproximadamente 1.000 nt, são chamados de miRNAs primários (pri-miRNAs) e produzidos majoritariamente pela RNA polimerase II (RNA Pol II), sendo poliadenilados (XIE *et al.*, 2005). Os pri-miRNAs são então processados pela DCL1 gerando precursores em forma de grampo de 70 a 100 nt chamados de pre-miRNAs. Estes são novamente processados pela DCL1 resultando em um pequeno dsRNA que possui geralmente 21 a 24 nt não perfeitamente pareados (XIE & QI, 2008). As terminações 2'-hidroxila de ambas as fitas são metiladas pela enzima “HUA enhancer 1” (HEN1) para proteger o miRNA de atividades exonucleotídicas (YU *et al.*, 2005). Já no citoplasma o sRNA de dupla fita é,

então, acoplado à proteína ARGONAUTE1 (AGO1) um dos principais componentes do RISC. Enquanto uma das fitas permanece no complexo a outra é degradada (SCHWARZ *et al.*, 2003).

O RISC conduz então a clivagem ou repressão traducional do mRNA alvo (JAMALKANDI & MASOUDI-NEJAD, 2009). A clivagem do RNA alvo ocorre geralmente entre a décima e décima primeira base do miRNA, gerando duas fitas de RNA. Uma delas é alvo da 5'-3' exorribonuclease 4 (XRN4), enquanto a outra pode ser alvo da RNA polimerase RNA-dependente 6 (RDR6), a qual polimeriza uma fita de RNA complementar a partir de um ssRNA de uma maneira independente de oligonucleotídeo iniciador. Deste modo, novos siRNAs poderão ser gerados a partir da nova molécula de dsRNA tendo como alvo outros mRNAs que possuem homologia à sua sequência, amplificando, assim, o sinal do miRNA (QI *et al.*, 2009).

Uma outra classe de sRNAs presentes em plantas é a dos siRNAs transatuantes (ta-siRNAs, do inglês, “trans-acting siRNAs”). Assim como os miRNAs, os ta-siRNAs são originários de locos gênicos definidos, chamados de TAS (PERAGINE, 2004). A expressão dos ta-siRNAs também depende da RNA Pol II que transcreve, a partir dos locos TAS, longos transcritos de ssRNAs que possuem sítios-alvo para miRNAs, também conhecidos como pri-ta-siRNAs (AXTELL *et al.*, 2006). A clivagem desses transcritos é mediada pelo complexo RISC contendo a proteína AGO1 e o miRNA homólogo ao sítio-alvo já mencionado. Após a clivagem há a geração de duas moléculas de RNA que em sendo alvos da RDR6, resultam em moléculas de dsRNAs, as quais são reconhecidas e clivadas pela DCL4 em pequenos siRNAs de 21 nt, chamados de ta-siRNAs maduros (GASCIOLLI, 2005). Os ta-siRNAs são metilados em suas terminações 2'-hidroxila pela HEN1 e são incorporados ao complexo RISC contendo AGOs que, por sua vez, possuem atividade de clivagem (AGO1 e AGO7), podendo mediar o silenciamento gênico pós-transcricional (PTGS, do inglês, “post-transcriptional gene silencing”) de mRNAs endógenos ou dos próprios pri-ta-siRNAs (VAZQUEZ *et al.*, 2004). O nome e abreviatura ta-siRNAs vem exatamente da capacidade destes siRNAs em atuar sobre alvos diferentes do pri-ta-siRNA seja em suas origens genômicas ou em suas sequências de

nucleotídeos (XIE & QI, 2008).

Os pequenos siRNAs associados a transcritos anti-senso naturais (nat-siRNAs) são originados de dsRNAs formados por homologia de sequência de dois transcritos que podem ser ncRNAs ou mRNAs (JAMALKANDI & MASOUDI-NEJAD, 2009). Estes transcritos anti-senso naturais (NATs, do inglês, “natural antisense transcripts”) podem ser originários do mesmo locus quando esse é transcrito em ambos os sentidos (*cis*-NAT), ou de locus diferentes, sendo chamados de *trans*-NAT (JAMALKANDI & MASOUDI-NEJAD, 2009). Nos casos mais estudados até agora um dos NATs é transcrito constitutivamente e a indução do outro NAT é que promove a formação do dsRNA precursor. Em *Arabidopsis thaliana* essa indução é realizada em condições de estresse salino ou pela infecção bacteriana (BORSANI, 2005, KATIYAR-AGARWAL *et al.*, 2006).

O transcrito induzido por estresses parece ser sintetizado pela RNA Pol IV e não codifica qualquer proteína (BORSANI, 2005). A região complementar aos dois transcritos é então reconhecida pela DCL2, a qual cliva o dsRNA formando os nat-siRNAs primários que possuem 24 nt e são metilados pela HEN1 (BORSANI, 2005). Os nat-siRNAs são acoplados no RISC por uma AGO ainda não revelada, mas que tem a capacidade de clivar o RNA alvo (XIE & QI, 2008). O alvo dos nat-siRNAs primários é o próprio NAT codificante, o qual é expresso constitutivamente. Após a clivagem deste transcrito, a RDR6 produz dsRNAs que são reconhecidos pela DCL1 que, por sua vez, produz os nat-siRNAs secundários de 21 nt, os quais possuem como alvo o próprio NAT constitutivo ou outros mRNAs da planta (BORSANI, 2005, KATIYAR-AGARWAL *et al.*, 2006).

Os siRNAs heterocromatínicos (hc-siRNAs) de plantas são tipicamente constituídos por 24 nt e são associados a sequências repetitivas no genoma como transposons, retroelementos, DNAs ribossomais e regiões centroméricas onde realizam o silenciamento transcricional por uma via conhecida como metilação do DNA dependente de RNA (RdDM, do inglês, “RNA-dependent DNA methylation”; LIPPMAN & MARTIENSSEN, 2004).

Transcritos são produzidos em regiões repetitivas do genoma pela RNA Pol IV e, então, são transformados em dsRNAs pela RDR2 e, subsequentemente clivados em hc-siRNAs pela DCL3 e metilados pela HEN1 (XIE & QI, 2008). Os

hc-siRNAs promovem a metilação do DNA de regiões homólogas à sua sequência em um processo que necessita da AGO4 e da RNA Pol V, possivelmente agrupadas em um mesmo complexo proteico (PONTES *et al.*, 2006). O complexo RISC com hc-siRNA, AGO4 e RNA Pol V recruta a metiltransferase rearranjada por domínios 2 (DRM2, do inglês, “domains rearranged methyltransferase”) e a metiltransferase defectiva em metilação de DNA direcionada por RNA 1 (DRD1, do inglês, “defective in RNA-directed DNA methylation”) as quais promovem a metilação do DNA e das histonas promovendo, assim, o silenciamento gênico transcricional desse locus (PONTES *et al.*, 2006). Foi demonstrado que a AGO6 também pode estar envolvida na rota dos hc-siRNAs, possuindo características redundantes às da AGO4 (ZHENG *et al.*, 2007).

1.3.2. Pequenos RNAs de interferência virais em plantas

O papel da RNAi na defesa antiviral em plantas tem sido elucidado ao longo dos últimos anos em diferentes trabalhos (LLAVE, 2010). Há basicamente duas linhas de evidências que comprovam este mecanismo. A primeira evidência é a de que muitos vírus codificam proteínas que suprimem a atividade de componentes envolvidos na maquinaria da RNAi como a proteína 2B do CMV, a qual suprime a atividade de clivagem da AGO1 (ZHANG, *et al.*, 2006). A segunda evidência é a de que siRNAs virais (vsiRNAs) são encontrados nas células das plantas infectadas por vírus, evidenciando uma estratégia de defesa da planta contra estas infecções (LLAVE, 2010).

Aparentemente todas as quatro DCLs de *A. thaliana* estão envolvidas na formação dos vsiRNAs (DELERIS *et al.*, 2006). A DCL4 seria a principal enzima 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 passaria a ser a principal formadora de vsiRNAs, produzindo vsiRNAs de 22 nt. Na ausência das duas, a DCL3 produziria vsiRNAs de 24 nt (DELERIS *et al.*, 2006). A contribuição da DCL1 é a menor de todas para a formação de vsiRNAs. Em mutantes de *A. thaliana* para as outras três DCLs vsiRNAs são encontrados, mas em uma quantidade muito menor (GARCIA-RUIZ

et al., 2010). AGO1, AGO2, AGO4 e AGO5, acopladas aos vsiRNAs, são as prováveis proteínas envolvidas na clivagem dos RNAs virais ou na metilação do DNA, no caso de vírus de DNA. RDR1, RDR2 e RDR6 estão envolvidas na resistência viral promovida por RNAi em plantas. Entretanto, elas cumprem papéis distintos em infecções promovidas por diferentes vírus (LLAVE, 2010).

Apesar dos vários avanços no entendimento do mecanismo de RNAi mediado por vírus, a biogênese dos vsiRNAs ainda não está muito bem compreendida. Moléculas dsRNAs 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 disto, transcritos aberrantes de RNA são reconhecidos por RDRs e transformados em dsRNAs. Entretanto, estudos demonstraram que a maioria dos vsiRNAs encontrados em plantas após infecções virais possuem orientação positiva (QI *et al.*, 2009), o que não é esperado para siRNAs originários de dsRNAs e clivados pela proteína DCL, a qual gera siRNAs nos dois sentidos. Isto levou à proposta de que vsiRNAs são gerados também por meio de estruturas secundárias do RNA viral, as quais formariam regiões de dsRNA (MOLNAR *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, sendo elucidada recentemente, é a capacidade destes de inibir a expressão de genes endógenos da planta. Esta capacidade já foi demonstrada em plantas após infecções pelo vírus do mosaico da couve-flor (CaMV, do inglês, "Cauliflower Mosaic Virus") e do fumo (TMV, do inglês, "Tobacco Mosaic Virus") em *A. thaliana* (MOISSIARD & VOINNET, 2006, QI *et al.*, 2009).

1.4. 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 (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 é restrito a alguns tipos de organismos (KONDO *et al.*, 2009). Ultimamente, no entanto, a transferência genética horizontal tem sido descrita como uma das principais forças evolutivas em organismos procarióticos (HEUER & SMALLA, 2007). Através 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 um evento mais raro. Esse fenômeno foi evidenciado em Eucariotos unicelulares que se alimentam de bactérias, tais como *Entamoeba* e *Trichomonas* (DE KONING *et al.*, 2000, ANDERSON *et al.*, 2001). A bactéria *Wolbachia* é um patógeno 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, realizados nessas células, 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 (HOTOPP *et al.*, 2007, KONDO *et al.*, 2009). 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 que 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 (MORISSETE & 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 bacteriais, contudo, estudos recentes evidenciaram a presença de sequências 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 episomais (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 codificam as enzimas integrase e transcriptase reversa, assim sendo 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 e insetos (KONIN, 2010). Aparentemente a integração nesses casos só foi possível graças a uma recombinação do RNA viral com o RNA de retrotransposons (GEUKING *et al.*, 2009, HORIE *et al.*, 2010, TAYLOR *et al.*, 2010).

1.5. Montagem de novo de sequências de RNA por meio de siRNAs

Kreuze *et al.* (2009) identificaram novos vírus e sequenciaram genomas virais inteiros através de dados obtidos por sequenciamento em larga escala de

sRNAs isolados de plantas infectadas. Através das rotas de RNAi, moléculas de siRNAs são formadas ao longo de toda a sequência do RNA viral. Assim, através da similaridade de sequência dessas moléculas, “contigs” podem ser formados utilizando ferramentas de bioinformática similares às utilizadas para a montagem de genomas. Esse conjunto de técnicas permite o sequenciamento e a identificação de RNAs que são alvos de rotas de silenciamento gênico via RNAi.

No presente trabalho, utilizando dados de sequenciamento em larga escala de sRNAs do projeto GENOSOJA, ao qual o nosso grupo de pesquisa faz parte, inicialmente realizados com a intenção de identificar novos miRNAs nessa planta, foram identificados em diferentes cultivares de soja sob condições controle e submetidas à estresse biótico e abiótico, assim como de diferentes tecidos e estágios de desenvolvimento, sequências homólogas ao RNA 1 do CMV. Em uma dessas bibliotecas de sequenciamento de sRNAs foi identificada a presença dos dois RNAs genômicos do *Bean pod mottle virus* (BPMV), um vírus de RNA patogênico da soja. Como só foi encontrado um dos RNAs do CMV nas diferentes bibliotecas, ao contrário do identificado em trabalhos de sequenciamento de sRNAs de plantas infectadas por vírus, aonde são detectados vsRNAs para todos os RNAs genômicos (DONAIRE et al., 2009), como no caso do BPMV, foi levantada a hipótese de que a sequência do RNA 1 do CMV estaria integrada no genoma da soja.

2. OBJETIVOS

2.1. Objetivo Geral

Identificar e caracterizar a integração do RNA 1 do CMV no genoma da soja e suas prováveis implicações fisiológicas assim como o perfil dos siRNAs produzidos tanto pelo CMV integrado quanto pelos RNAs do BPMV.

2.2. Objetivos Específicos

- Realizar a montagem *de novo* do RNA 1 do CMV e do RNA 1 e 2 do BPMV utilizando os dados do sequenciamento em larga escala de sRNAs e ferramentas de bioinformática para a geração dos “contigs”.
- Demonstrar a integração do RNA 1 do CMV no genoma da soja em diferentes cultivares.
- Identificar e comparar o perfil de transcrição do RNA 1 do CMV integrado em cultivares de soja sob condições controle e de estresse.
- Comparar a expressão dos siRNAs provenientes do CMV sob condições controle e de estresse.
- Analisar o perfil dos siRNAs provenientes do RNA 1 do CMV e dos RNAs do BPMV em todas as diferentes bibliotecas de sequenciamento.
- Identificar siRNAs provenientes do RNA 1 do CMV que são diferencialmente expressos na soja sob condições controle e de estresse.

3. MATERIAL E MÉTODOS

3.1. Material biológico para as bibliotecas de sRNAs

As sequências de sRNAs analisadas nesse trabalho vieram de plantas de soja de 5 cultivares: “Embrapa 48”, “BR 16”, “PI561356”, “Conquista” e “Urano”.

3.1.1. Experimento de estresse abiótico

As cultivares “Embrapa 48” resistente ao estresse hídrico e a “BR 16” sensível à esse estresse foram cultivadas em uma estufa localizada na Embrapa-Soja em Londrina, Paraná. As plantas foram cultivadas usando um sistema hidropônico utilizando contêineres de plástico de 30 litros contendo uma solução nutritiva de pH 6,6. As sementes foram pré-germinadas em papel umedecido mantidos no escuro à temperatura de $25\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ e $65\% \pm 5\%$ de umidade relativa. As plântulas foram colocadas em suportes de poliestireno de maneira a deixar as raízes da planta completamente submersas na solução nutritiva. Os recipientes contendo as plantas foram mantidos em uma estufa à $25\text{ }^{\circ}\text{C} \pm 2\text{ }^{\circ}\text{C}$ e $60\% \pm 5\%$ de umidade relativa sob luz natural por 12 horas. Após 15 dias, mudas com a primeira folha trifoliada completamente desenvolvida (estádio vegetativo V2) foram submetidas a diferentes estresses hídricos. A solução nutritiva foi removida de cada contêiner plástico aonde as raízes das plantas estavam submersas por 0 minutos (T0 ou controle), 125 minutos (T125) e 150 minutos (T150). No final de cada tratamento as raízes das plantas foram imediatamente congeladas por nitrogênio líquido e armazenadas a $-80\text{ }^{\circ}\text{C}$ para posterior extração de RNA. Cada tratamento foi realizado em triplicata, onde cada uma dessas foi composta por 5 plantas. Quatro bibliotecas de sRNAs foram construídas desses tratamentos: 1) raízes das plantas sensíveis à deficiência hídrica não submetidas ao estresse (controle), também chamadas de Raiz sensível à seca controle (DSRC); 2) raízes das plantas sensíveis à deficiência hídrica submetidas à 125 minutos e 150 minutos de estresse, chamadas por Raiz sensível à seca tratada (DSRT); 3) raízes das plantas tolerantes à deficiência hídrica não submetidas ao

estresse (controle), também chamadas de raiz tolerante à seca controle (DTRC); 4) raízes das plantas tolerantes à deficiência hídrica submetidas a 125 minutos e 150 minutos de estresse, chamados de raiz tolerante à seca tratada (DTRT).

3.1.2. Experimento de estresse biótico

A reação ao fungo da ferrugem asiática foi avaliada em plantas de soja na estufa da Embrapa-Soja em Londrina, Paraná usando uma população de campo de *Phakospora pachyrhizi* coletada de plantações de soja no estado do Mato Grosso e mantidas por mais de 10 gerações na cultivar susceptível BRSMS-Bacuri. A cultivar “Embrapa 48” foi usada como planta suscetível a infecção, já que desenvolve lesões susceptíveis após a infecção por *Phakospora pachyrhizi*, enquanto a cultivar “PI561356” foi usada como planta resistente a infecção já que desenvolve uma lesão castanho avermelhada com poucos ou nenhum esporo. Urediniosporos foram coletados de plantas BRSMS-Bacuri infectadas, em uma estufa separada, através de batidas nas folhas infectadas sobre uma bandeja plástica e então foram diluídos em água destilada com 0,05% de Tween-20 em uma concentração final de 3×10^5 esporos/mL. Essa suspensão de esporos foi aplicada em potes contendo 3 plantas de soja do estágio vegetativo V2. A mesma solução só que sem esporos foi utilizada para o experimento controle. Esse processo foi repetido nas duas cultivares. Após a aplicação das soluções as plantas foram protegidas por sacolas plásticas umedecidas para evitar a infecção das plântulas utilizadas no tratamento controle. A terceira folha trifoliolada das três plantas foi coletada 12 horas após a inoculação (hai) para a extração de RNA. Por intermédio deste experimento, foram construídas quatro bibliotecas de sRNAs de folhas compostas por: 1) folhas das plantas suscetíveis a infecção tratadas com a solução controle, também chamada de folhas suscetíveis a ferrugem controle (RSLC); 2) folhas das plantas suscetíveis a infecção tratadas com a solução de esporos, também chamada de folhas suscetíveis a ferrugem inoculadas (RSLI); 3) folhas das plantas resistentes a infecção tratadas com a solução controle, também chamada de folhas resistentes a ferrugem controle (RRLC); 4) folhas das plantas resistentes a infecção tratadas com a solução de esporos, também

chamada de folhas resistentes a ferrugem inoculadas (RRLI).

3.1.3. Diferentes estádios de desenvolvimento da soja

RNA de sementes maduras, vagens e sementes de 3 estádios de germinação diferentes da cultivar de soja “Conquista” também foram extraídos para serem usados na construção de bibliotecas de sRNAs. Sementes foram incubadas por 3, 5 e 7 dias em papel filtro umedecido à 26 °C em uma câmara de crescimento. Vagens (dos estádios R3-R5) foram coletadas de plantas cultivadas em estufa localizada na Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre. Sementes maduras, vagens e sementes germinadas por 3, 5 e 7 dias foram congeladas em nitrogênio líquido e trituradas para posterior e imediata extração de RNA. Para estas 3 bibliotecas, deu-se o nome de Semente (para RNAs de semente madura), Vagem (para os RNAs de vagem) e Germinação (Para os RNAs das sementes de 3 estádios de germinação).

3.1.4. Flor

Amostras de flor (dos estádios R2-R3) foram coletadas da cultivar de soja “Urano” cultivadas no campo experimental da Universidade de Passo Fundo (UPF), Rio Grande do Sul. A flor inteira e amostras do carpelo, estame e pétala foram pulverizadas para posterior extração de RNA. A partir desses RNAs outras 4 bibliotecas foram construídas chamadas de Flor (RNAs da flor inteira), Pétala (RNAs da pétala), Estame (RNAs do estame) e Carpelo (RNAs do carpelo).

3.2. Extração de RNA

O RNA total das diferentes amostras foi extraído usando o reagente Trizol (Invitrogen, CA, EUA) de acordo com as instruções do fabricante e a qualidade do RNA foi avaliada por eletroforese em gel de agarose 1.0%. A quantidade de RNA foi analisada utilizando o kit Quant-iT RNA assay (Invitrogen, CA, EUA) e o fluorímetro Qubit de acordo com as instruções do fabricante.

3.3. Sequenciamento em larga escala dos sRNAs

Aproximadamente 10 µg do RNA total extraído foi enviado à Fasteris Life Sciences SA (Plan-les-Ouates, Suíça) para processamento e sequenciamento usando a tecnologia Solexa no aparelho Illumina Genome Analyzer GAII e HiSeq2000. As bibliotecas de sRNAs foram construídas a partir das amostras de RNA obtidos nos experimentos previamente descritos (DSRC, DSRT, DTRC, DTRT, RSLC, RSLI, RRLC, RRLI, Semente, Vagem, Germinação, Flor, estame, Carpelo e Pétala) sequenciados em um total de 2 canais. Basicamente o processo de sequenciamento de sRNAs pelo Illumina consiste dos seguintes passos: purificação das bandas de RNA correspondentes ao intervalo de tamanho de 15 a 30 nt em gel de poliacrilamida, após eletroforese das amostras de RNA. Ligação de adaptadores 5' e 3' ao RNA em duas etapas separadas e subsequentes, seguidas de purificação em gel de poliacrilamida, síntese de cDNA seguido de purificação em gel de poliacrilamida e uma etapa final de amplificação por PCR (Reação em Cadeia da Polimerase) para gerar as bibliotecas de colônias de cDNA para o sequenciamento pelo Illumina. Após a remoção dos adaptadores as sequências são separadas de acordo com os seus tamanhos (de 19 à 24 nt) para posterior análise.

3.4. Montagem *de novo* do RNA 1 do CMV e dos RNAs do BPMV

As sequências dos sRNAs de 19 a 24 nt das diferentes bibliotecas foram agrupadas por cultivar (“Embrapa 48”: DTRC, DTRT, RSLC e RSLI; “BR 16”: DSRC e DSRT; “PI561356”: RRLC e RRLI; “Conquista”: Semente, Vagem e Germinação; “Urano”: Flor, Carpelo, Pétala e Estame). Os sRNAs das diferentes cultivares foram agrupados em “contigs” usando o algoritmo para montagem *de novo* do programa Velvet 0.7.3.1 (ZERBINO & BIRNEY, 2008) com os seguintes parâmetros: “hash length”: 23; “coverage cut-off”: 50; “expected coverage”: 1,000 e “minimum scaffold length”: 100. Os “contigs” montados que possuíam correspondência de sequência ao genoma da soja foram eliminados das análises

posteriores por BLAST (“Basic Local Alignment Search Tool”) versão 2.2.20 contra o genoma da soja (*Glycine max*) disponível no Phytozome (ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/Gmax/assembly/sequence/) com os seguintes parâmetros: BLASTn; “Expectation value” (-e): 1e-10, “cost to open a gap (-G): -6; “cost to extend a gap” (-E): -6; “penalty for a nucleotide mismatch”: -5. Os contigs remanescentes foram usados em uma nova busca por BLAST no banco de dados do NCBI (“National Center for Biotechnology Information”) utilizando a ferramenta BLASTn. Os “contigs” que possuíram sequência similar ao CMV ou ao BPMV foram separados para a montagem *de novo* desses RNAs. Utilizando o alinhamento Explorer/CLUSTALW do programa Molecular Evolutionary Genetics Analysis (MEGA versão 4.0) os contigs foram alinhados à sequência de RNA de maior homologia encontrada nas buscas por BLAST no banco de dados do NCBI para cada um dos 3 RNAs encontrados com o objetivo de montar a sequência desses RNAs para cada cultivar.

3.5. Projeção dos oligonucleotídeos para PCR e RT-qPCR e sequenciamento

Utilizando a ferramenta Primer3 (<http://frodo.wi.mit.edu/primer3/>) (ROZEN & SKALETSKY, 2000) foram desenhados oligonucleotídeos senso e anti-senso para a amplificação de regiões da sequência do RNA 1 do CMV para as análises por PCR qualitativa como para as análises de RT-qPCR e sequenciamento (Figura 1). Para as análises de RT-qPCR um par de oligonucleotídeos foi desenhado para gerar um produto de amplificação de 179 pb na região 3' do RNA viral (Figura 1). Todos os oligonucleotídeos utilizados nesse trabalho (Tabela 1) possuem um temperatura ótima de anelamento de 60°C nas condições reacionais utilizadas.

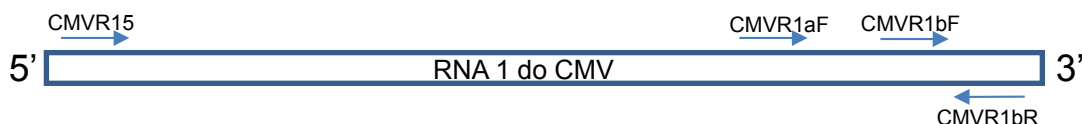


Figura 1. Posição dos oligonucleotídeos projetados em referência a sequência do RNA 1 do CMV. As setas azuis indicam os oligonucleotídeos assim como as suas orientações. Acima ou abaixo das setas esta o nome dos respectivos oligonucleotídeos. A figura não está em escala. Tamanho total do fragmento 3092 nt.

Tabela 1. Lista de oligonucleotídeos utilizados e suas respectivas sequências.

Oligonucleotídeo	Seqüência de nucleotídeos
CMVR15	5'- GCTCGAGGAGCAATTACAGC-3'
CMVR1aF	5'- CCAAAGAAGCCTGCACTTTC-3'
CMVR1bF	5'- GGCATTTCTGAGGACCATGT-3'
CMVR1bR	5'- AAAGGTGGGAAACCGCTAAG -3'
MIR156a	5'-TGACAGAAGAGAGTGAGCAC-3'
MIR156b	5'-TGACAGAAGAGAGAGAGCACACA-3'
MIR1520d	5'-ATCAGAACATGACACGTGACAA-3'
Reverse	5'-GTGCAGGGTCCGAGGT-3'
Poli25TV	5'- TTTTTTTTTTTTTTTTTTTTTTTTTTV-3'

3.6. Extração de DNA

DNA de 12 cultivares de soja (“Conquista”, “Urano”, “BR 16”, “Embrapa 48”, “Willians”, “Davis”, “Hill”, “Paraná”, “BR 4”, “IAS 4”, “Força” e “Potência”), fornecidas pela EMBRAPA-Soja de Londrina-PR, foram extraídos. As sementes maduras dessas cultivares foram incubadas em papel filtro umedecido por uma semana à 26°C em câmara de crescimento. Folhas das plântulas de cada cultivar foram pulverizadas com cadinho e pistilo em nitrogênio líquido. Aproximadamente 100 mg do material pulverizado foi utilizada para a extração de DNA com o auxílio do *kit* PureLink™ (Invitrogen), seguindo as instruções do fabricante. As amostras de DNA foram diluídas na proporção de 1:10 em água destilada deionizada autoclavada e estocadas a -20°C.

3.7. PCR para confirmar a integração do RNA 1 do CMV no genoma da soja

As PCRs foram realizadas utilizando 1 µl do DNA extraído das 12 cultivares de soja diluído 1:10 (~10ng/µl de DNA) foram adicionados 1 µl de cada oligonucleotídeo (10 µM); 1 µl de dNTP (5 µM); 2 µl de tampão de reação 10x (Tris-HCl 100 mM. pH 8,0); 1 µl de MgCl₂ (50 mM); 0,1µl de Taq DNA polimerase (Invitrogen) (5U/µl) e 12,9 µl de água deionizada autoclavada num total de 20 µL de reação. Foram realizadas 3 reações com conjuntos de oligonucleotídeos diferentes. Em uma reação foram utilizados os oligonucleotídeos CMVR15 e CMVR1bR (produto de amplificação esperado: 3000 pb) em outra os

oligonucleotídeos CMVR1aF e CMVR1bR (produto de amplificação esperado: 731 pb) e na última os oligonucleotídeos CMVR1bF e CMVR1bR (produto de amplificação esperado: 179 pb) (Figura 1, Tabela 1).

As reações de PCR foram realizadas em aparelho Viriti (Applied Biosystems), sob as seguintes condições: 95°C por 5 minutos e 40 ciclos de 95 °C por 30 s, 60 °C por 30 s e 72 °C por 3 min. Uma alíquota de 5 µL de cada um dos produtos de PCR foi resolvida por eletroforese em gel de agarose 2%, corado com o fluoróforo “SYBR-Gold” (Invitrogen).

3.8. Sequenciamento de DNA

Foi confirmada a sequência para um dos produtos de PCR resultante da reação que utilizou os oligonucleotídeos CMVR1bF e CMVR1bR (Figura 1, Tabela 1) no sequenciador automático Genetic Analyzer ABI Prism 3100 (Applied Biosystems), pertencente ao Centro de Biotecnologia/UFRGS.

Em aproximadamente 50 ng do produto de PCR foram adicionados 5 pmoles do oligonucleotídeo CMVR1bF, 2 µL de tampão de reação, 1 µL de marcador (“Big Dye”) e água deionizada autoclavada até o volume de 10µL. A reação de termociclagem ocorreu por uma etapa inicial de desnaturação de 96 °C, por 3 minutos, seguida por 25 ciclos de 96 °C por 10 segundos, 55 °C por 5 segundos e 60 °C por 4 minutos. Ao final desta reação, a amostra foi precipitada com isopropanol 75%, e posteriormente lavada com etanol 60%. Depois de seca, foi adicionado formamida à amostra, a qual foi desnaturada (95°C por 5 minutos) e injetada no sequenciador.

3.9. Síntese de DNA complementar (cDNA)

Aproximadamente 2 µg do RNA total extraído dos experimentos de estresse biótico e abiótico e o oligonucleotídeo poli25TV (1µg) foram misturados e tiveram seu volume final ajustado em 10µL com água deionizada livre de Ribonucleases (RNases) para posterior desnaturação a 70 °C por 5 minutos. Após essa etapa, o tubo contendo o RNA purificado e desnaturado foi rapidamente

transferido para banho de gelo. Posteriormente, foram adicionados 2 µL de dNTPs (5mM), 6 µL de tampão de reação 5X, 1 µL transcriptase reversa M-MLV (200U/µL) e água deionizada livre de RNases, totalizando 30 µL. A reação de síntese da fita complementar ocorreu a 40 °C por 1 hora. O cDNA produzido foi diluído na proporção de 1:10 em água deionizada livre de RNases e estocado a -80 °C. A partir desta diluição, uma segunda diluição foi feita, na proporção de 1:100, a qual foi utilizada para os experimentos de RT-qPCR.

Para a síntese de cDNA a partir dos sRNAs, ao invés do primer poli25TV, foi utilizado o primer reverso universal “stem-loop” de acordo com Cheng e colaboradores (2005). A sequência do “stem-loop” consiste em 44 nucleotídeos conservados e seis variáveis que serão específicos ao final 3' dos sRNAs (5' GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACNNNNNN 3').

3.10. RT-qPCR

A técnica de PCR quantitativa em tempo real (RT-qPCR) foi utilizada para analisar a expressão do RNA 1 do CMV integrado ao genoma da soja entre as plantas submetidas a estresses bióticos e abióticos. Todas as reações de RT-qPCR foram realizadas utilizando 2 replicatas técnicas e 3 replicatas biológicas. Os oligonucleotídeos utilizados para a detecção do RNA 1 do CMV foram o CMVR1bF e CMVR1bR e para a normalização dos resultados de expressão foram utilizados os oligonucleotídeos MIR156a, MIR156b e MIR1520d junto com o oligonucleotídeo Reverse que amplificam miRNAs (Tabela 1). Kulcheski *et al.* (2010) demonstraram que genes de miRNAs são melhores do que genes que codificam proteínas para normalizar os dados de RT-qPCR em soja.

Em 10 µL de cDNA diluídos 1:100 ($\geq 0,1\text{ng}/\mu\text{L}$) extraídos das plantas dos experimentos de estresse biótico e abiótico, foram adicionados 0,4 µL de cada oligonucleotídeo (5 µM); 0,4 µL de dNTP (5 mM); 2 µL tampão de reação 10X (Tris-HCl 100mM, pH 8,0); 1,2 µL MgSO₄ (50 mM); 2 µL de SYBR Green (1X), 0,05 µL de Taq DNA Polimerase (Invitrogen) (5U/µL) e 3,55 µL de água deionizada autoclavada, resultando em 20µL de reação.

As RT-qPCRs foram realizadas no equipamento 7500 Real Time System

(Applied Biosystems), tendo uma etapa inicial de 95 °C por 5 minutos, seguida de 40 ciclos de 95 °C por 15 s, 60 °C por 10 s e 72 °C por 15 s. Posteriormente as amostras foram mantidas a 50 °C por 2 min e aquecidas até 99°C (0,4 °C/s), para obtenção da curva de dissociação ou desnaturação.

Os cálculos de expressão relativa foram feitos pelo método $2^{-\Delta\Delta C_t}$ (LIVAK & SCHMITTGEN, 2001). As análises foram feitas pelo teste estatístico de Análise da Variância (ANOVA) seguido pelo teste de Turkey HSD (“Honestly Significant Difference”).

3.11. Caracterização dos siRNAs produzidos pelo RNA 1 do CMV e pelos RNAs do BPMV

Para analisar a quantidade de siRNAs de diferentes tamanhos a sua orientação e a distribuição dos mesmos ao longo das sequências de RNAs virais encontradas nas diferentes bibliotecas de sRNAs foi utilizado o algoritmo SOAPaligner/soap2 do programa SOAP (“Short Oligonucleotide Analysis Program”) versão 2.20 (LI *et al.*, 2009). Esse programa alinha as sequências geradas (“reads”) pelo sequenciamento de larga escala realizados pela tecnologia Solexa/Illumina à uma determinada sequência indicando a sua posição e orientação (senso ou anti-senso). Os siRNAs (de 19-24 nt) de cada uma das diferentes bibliotecas de sRNAs foram alinhados separadamente ao RNA 1 do CMV e aos RNA 1 e 2 do BPMV (encontradas nessas mesmas bibliotecas como já previamente descrito) pelo programa SOAP de modo a não permitir nenhum “mismatch” entre as sequências alinhadas.

3.12. Análise do padrão de expressão dos siRNAs produzidos pelo RNA 1 do CMV

A expressão dos siRNAs de 19-24 nt produzidos pelo RNA 1 do CMV foi verificada nas diferentes bibliotecas de estresse biótico e abiótico (DSRC, DSRT, DTRC, DTRT, RSLC, RSLI, RRLC e RRLI). O total de siRNAs produzidos pelo RNA 1 do CMV foi normalizado pelo total de siRNAs se suas respectivas

bibliotecas. O teste do qui-quadrado foi utilizado para verificar se a diferença entre os valores obtidos para cada biblioteca é estatisticamente significativa.

3.13. Análise dos siRNAs diferencialmente expressos

Para a análise da expressão dos siRNAs entre as diferentes bibliotecas foi usado o pacote do programa estatístico R denominado “DEGseq” (WANG et al., 2010). Esse programa modela a sequência através de um processo de amostragem aleatória seguido de uma distribuição binomial e, então, aplica o teste exato de Fisher para cada gene. Para essa análise foram utilizados os dados normalizados cada sequência única de sRNAs. Em seguida, o “fold-change”, o valor de P e o valor de q foram computados pela “DEGseq” para identificar os siRNAs diferencialmente expressos. Foram feitas comparações duas a duas entre as bibliotecas controle e submetidas aos estresses bióticos e abióticos. Também examinamos as diferenças entre as diferentes cultivares sob o mesmo tratamento. As comparações entre as bibliotecas de sRNAs se deram da seguinte maneira: DSRT-DSRC, DTRT-DTRC, DTRC-DSRC, DTRT-DSRT, RSLI-RSLC, RRLI-RRLC, RRLC-RSLC e RRLI-RSLI.

3.14. Predição dos alvos dos siRNAs diferencialmente expressos.

Os prováveis alvos dos siRNAs diferencialmente expressos do RNA 1 do CMV entre as diferentes bibliotecas de sRNAs foram buscados através do uso da ferramenta de predição de alvos de miRNAs/ta-siRNAs de plantas do sítio “UEA (University of East Anglia) sRNAs toolkit” (<http://srna-tools.cmp.uea.ac.uk/plant/cgi-bin/srna-tools.cgi>) (MOXON et al., 2008). As regras para a predição de alvos são as sugeridas por Allen e colaboradores (2005) e Schwab e colaboradores (2005). Para a identificação das sequências-alvo dos siRNAs foi utilizado o programa “BLASTn” do banco de dados do Phytozome (<http://www.phytozome.net/>).

3.15. Busca de estruturas em grampo para o RNA 1 do CMV.

A busca por estruturas secundárias em forma de grampo no RNA 1 do CMV foi realizada utilizando o programa Mfold (ZUKER, 2003). Foram analisadas as sequências do RNA 1 inteira e cinco outras sequências de 240 nt identificadas como regiões de alta produção de siRNAs.

4. RESULTADOS

4.1. Sequenciamento em larga escala das bibliotecas de sRNAs e montagem *de novo* das sequências do RNA 1 do CMV e dos RNA 1 e 2 do BPMV.

As bibliotecas de sRNAs analisadas nesse estudo tiveram os seguintes números de “reads” (sRNAs sequenciados) totais: DSRC: 1.449.588, DSRT: 4.302.713, DTRC: 4.916.869, DTRT: 1.349.605, RSLC: 3.813.712, RSLI: 3.592.903, RRLC: 3.849.662, RRLI: 3.818.259, Germinação: 4.190.789, Vagem: 3.575.804, Semente: 3.604.462, Carpelo: 3.530.011, Estame: 3.275.049, Pétala: 3.360.601, Flor: 3.392.134. Em todas as bibliotecas os sRNAs de 21 e 24 nt foram os mais abundantes (Figura 2 e Figura 3).

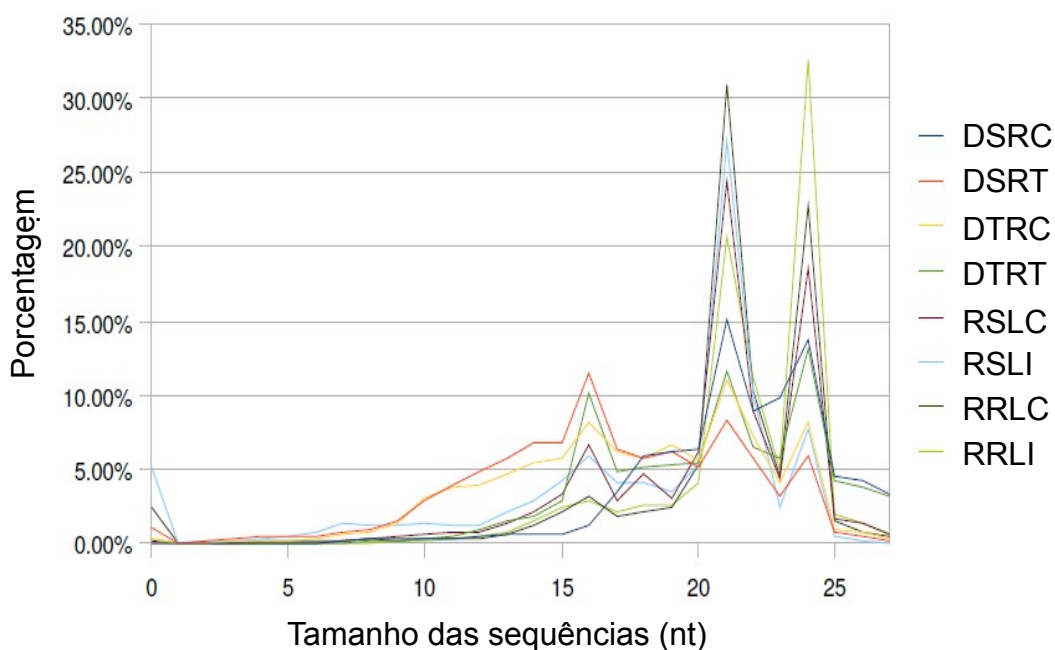


Figura 2. Abundância relativa das sequências de tamanhos diferentes obtidas pelo sequenciamento realizado pela plataforma Solexa/Illumina das bibliotecas de pequenos RNAs de estresse biótico e abiótico. As bibliotecas DSRC e DSRT pertencem à cultivar “BR16”, DTRC, DTRT, RSLC e RSLI à cultivar “Embrapa 48” e RRLC e RRLI à cultivar “PI561356”.

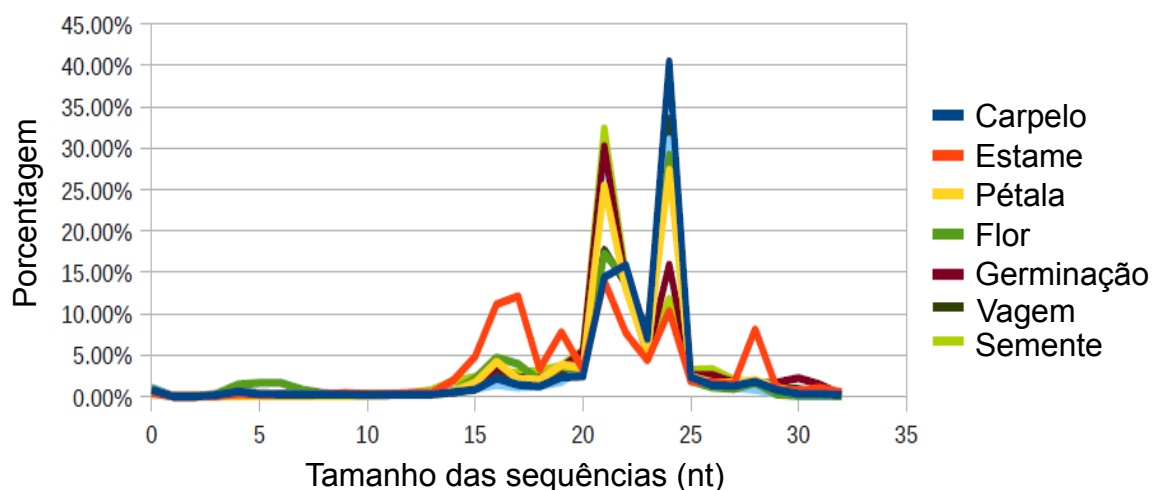


Figura 3. Abundância relativa das sequências de tamanhos diferentes obtidas através do sequenciamento realizado pela plataforma Solexa/Illumina das bibliotecas de pequenos RNAs das cultivares de soja “Conquista” (Germinação, Vagem e Semente) e “Urano” (Flor, Estame, Pétala e Carpelo).

Foram obtidas por montagem *de novo* dos pequenos RNAs sequências de 3.092 nt com homologia ao RNA 1 do CMV em todas as 17 bibliotecas. Todas elas possuem a mesma sequência de nucleotídeos. Uma busca por “BLASTn” no banco de dados “nucleotide collection” do NCBI revelou que a sequência obtida das bibliotecas de sRNAs possui uma maior homologia ao RNA 1 da linhagem “Legume” do CMV (código de acesso: D16403.1) com uma identidade de nucleotídeos de 93% (2861 nt).

Sequências do RNA 1 de 5.991 nt e do RNA 2 de 3.662 nt do BPMV também foram obtidas por montagem *de novo*, mas apenas na biblioteca Semente. O RNA 1 do BPMV apresentou maior homologia ao isolado “Iowa-Prignitz 10” (código de acesso: GQ996948) com uma identidade de nucleotídeos de 97% e o RNA 2 apresentou maior homologia à linhagem “K-Hopkins1” (código de acesso: AF394609) com uma identidade de nucleotídeos de 97%.

4.2. Confirmação da integração do RNA 1 do CMV em diferentes cultivares de soja por PCR.

Utilizando oligonucleotídeos específicos (Figura1 e Tabela 1) para a sequência do RNA 1 do CMV encontrada em nossas bibliotecas de sRNAs foram realizadas reações de PCR utilizando o DNA de 12 cultivares de soja, quais

sejam: “Conquista”, “Urano”, “BR 16”, “Embrapa 48”, “Willians”, “Davis”, “Hill”, “Paraná”, “BR 4”, “IAS 4”, “Força” e “Potência”. Produtos de amplificação com o tamanho esperado de 2896 pb (CMVR15 e CMVR1bR) e 731 pb (CMVR1aF e CMVR1bR) foram obtidos para todas as cultivares exceto a “Willians” (Figura 4).

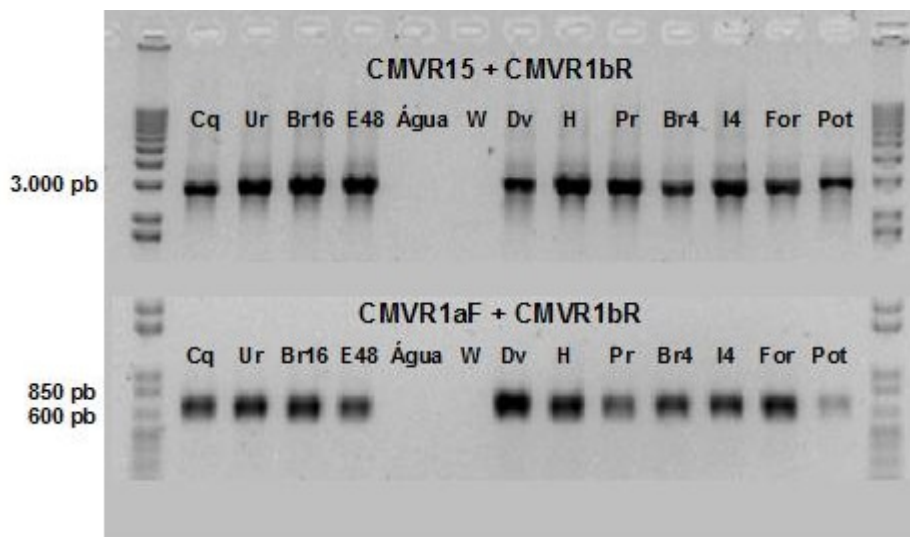


Figura 4. Eletroforese em gel de agarose 2% dos produtos de amplificação de duas reações de PCR utilizando o DNA de 12 cultivares de soja. A reação utilizando os oligonucleotídeos CMVR15 e CMVR1bR está representada acima e a qual foi utilizado os oligonucleotídeos CMVR1aF e CMVR1bR abaixo. No extremo esquerdo e direito de cada gel foi aplicado o marcador de peso molecular. Cq: “Conquista”, Ur: “Urano”, Br16: “BR 16”, E48: “Embrapa 48”, W: “Willians”, Dv: “Davis”, H: “Hill”, Pr: “Paraná”, Br4: “BR 4”, I4: “IAS 4”, For: “Força” e Pot: “Potência”.

4.3. Confirmação do produto de PCR por sequenciamento.

O sequenciamento do produto de amplificação por PCR utilizando os oligonucleotídeos CMVR1bF e CMVR1bR e o DNA da cultivar “Conquista” demonstrou que este possui a mesma sequência de nucleotídeos esperada para a determinada região do RNA 1 do CMV encontrado nas bibliotecas de sRNAs (dado não apresentado).

4.4. Caracterização dos siRNAs do RNA 1 do CMV e do RNA 1 e RNA 2 do BPMV.

As bibliotecas de sRNAs analisadas nesse estudo obtiveram os seguintes números de siRNAs de 19 à 24 nt correspondentes ao RNA 1 do CMV: DSRC: 1061, DSRT: 3080, DTRC: 9585, DTRT: 1613, RSLC: 7759, RSLI: 5304, RRLC: 25915, RRLI: 15803, Flor: 7910, Estame: 3153, Carpelo: 8259, Pétala: 7809, Germinação: 3262, Semente: 1617 e Vagem: 3975.

Os siRNAs mais abundantes em todas as bibliotecas foram os de 22 nt, variando de 90% de todos os siRNAs na biblioteca semente a 58% dos siRNAs da biblioteca RRLI. Os siRNAs de 24 nt foram a segunda classe mais abundante variando de 32% na biblioteca RRLI a 4% na biblioteca Semente (Figura 5).

Na média de todas as bibliotecas os siRNAs de 22 nt foram os mais abundantes com 73% de todos os siRNAs seguido pelos de 24 nt com 16%, 21 nt e 23 nt com 4%, 20 nt com 2% e 19 nt com 1% (Figura 5).

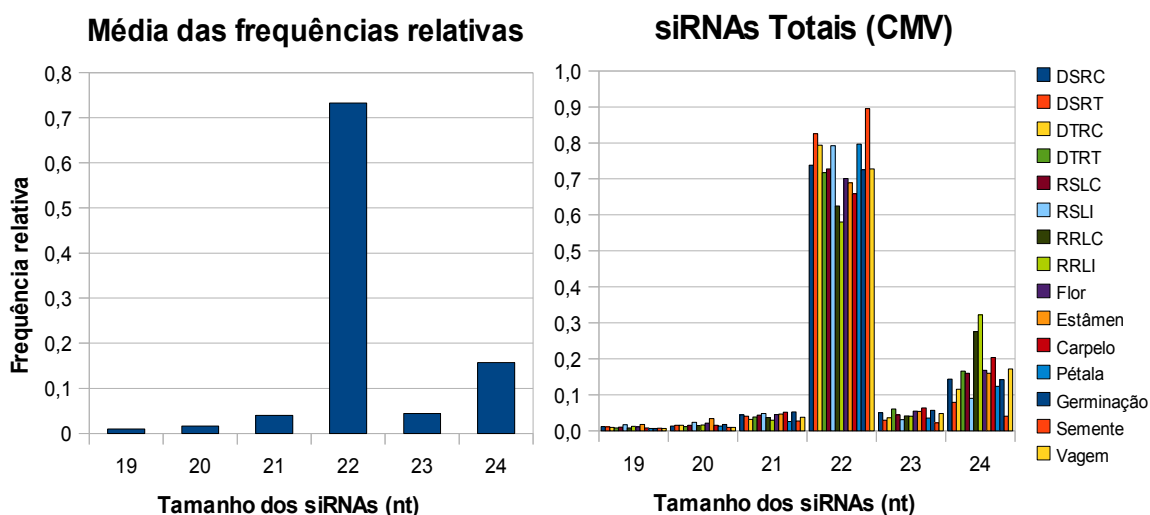


Figura 5. A frequência de cada siRNA de 19 a 24 nt é relativa a abundância de cada uma dessas classes de tamanho ao total de siRNAs referentes ao RNA 1 do CMV de sua biblioteca. O gráfico da esquerda corresponde à média das frequências dos siRNAs de tamanhos diferentes em todas as bibliotecas. O gráfico da direita apresenta a frequência de cada siRNA de 19 a 24 nt dentro de sua própria biblioteca.

Quanto a orientação das sequências, na média de todas as bibliotecas os siRNAs senso corresponderam a 57% e os anti-senso a 43%. As bibliotecas Vagem e Pétala tiveram as menores diferenças entre siRNAs senso e anti-senso com 53% do primeiro e 47% do segundo. A biblioteca DSRC apresentou a maior diferença com 63% para siRNAs senso e 37% para os anti-senso.

Levando em consideração a orientação dos siRNAs, a classe mais

abundante na média de todas as bibliotecas foi a dos de 22 nt senso com 42%. seguida de 22 nt anti-senso com 32%, 24 nt senso com 8% e 24 nt anti-senso com 7% (Figura 6B). Para os siRNAs de 22 nt senso e anti-senso o maior valor foi encontrado na biblioteca Semente com 50% para o primeiro e 40% para o segundo e o menor valor na biblioteca RRLI com 35% e 23%, respectivamente (Figura 6A). Para os siRNAs de 24 nt a situação se inverteu, na biblioteca RRLI o maior valor obtido foi de 19% para o senso e 14% para o anti-senso, enquanto que os menores valores foram encontrados na biblioteca de Semente com 2% para as duas orientações (Figura 6B).

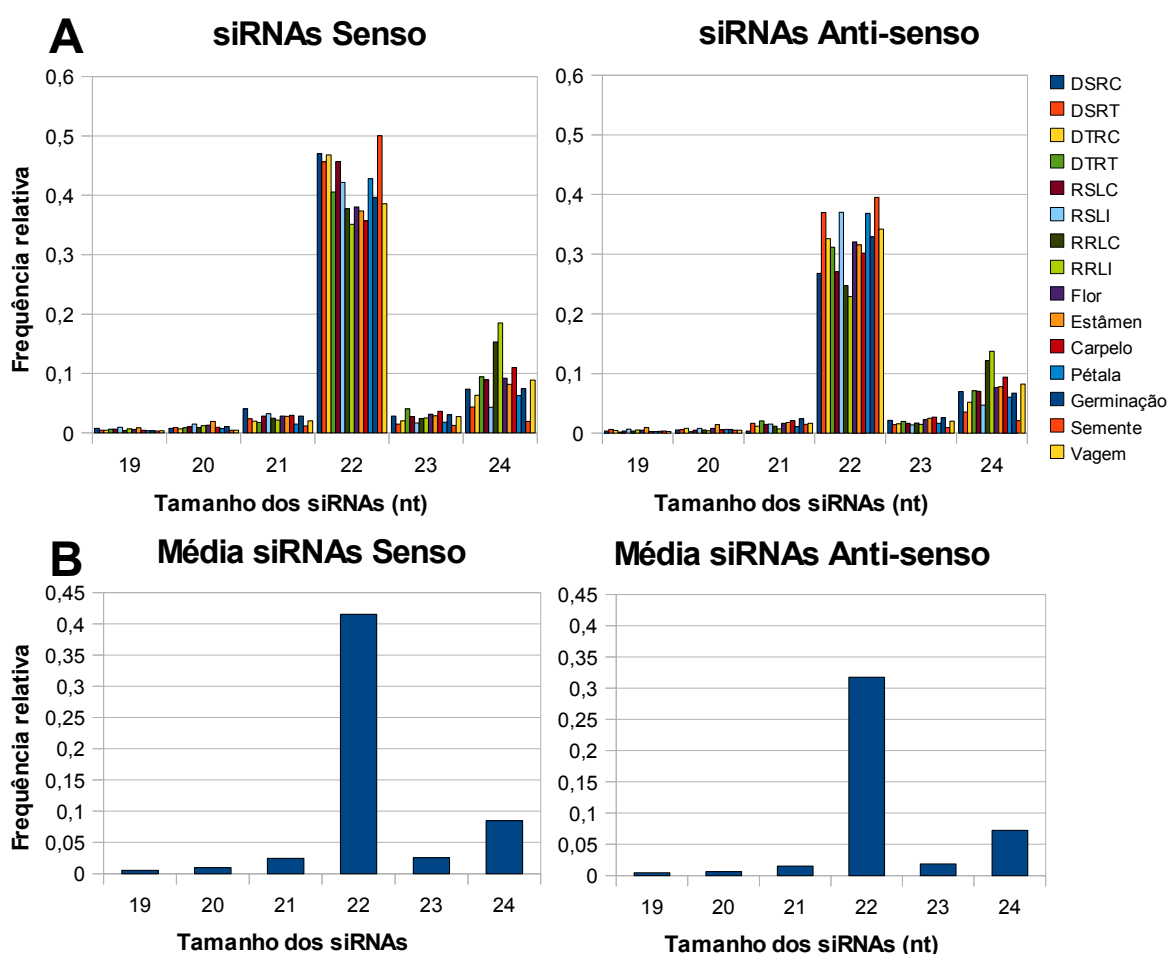


Figura 6. A frequência de cada siRNA de 19 à 24 nt é relativa a abundância de cada uma dessas classes de tamanho ao total de siRNAs referentes ao RNA 1 do CMV de sua biblioteca. **A)** Gráficos da frequência dos siRNAs senso (esquerda) e anti-senso (direita) de 19 à 24 nt relativa ao total de siRNAs do CMV de cada biblioteca. **B)** gráfico correspondente à média das frequências dos siRNAs senso (esquerda) e anti-senso (direita) de tamanhos diferentes em todas as bibliotecas.

Na biblioteca Semente, além do RNA 1 do CMV, foram encontrados por montagem *de novo* o RNA 1 e RNA 2 do BPMV. O total de siRNAs foi de 62.060 para o RNA 1 e 18.874 para o RNA 2. A classe mais abundante de siRNAs provenientes dessas sequências foi a de 22 nt contabilizando 47% do total de siRNAs do RNA 1 e 45% do RNA 2. Contudo, diferente do CMV, os siRNAs de 21 nt foram a segunda classe mais frequente correspondendo à 40% para o RNA 1 e 41% para o RNA 2 (Figura 7A). Quanto a orientação dos siRNAs para o RNA 1, 56% são senso e 44% anti-senso. Para o RNA 2 54% são senso e 46% anti-senso (Figura 7B).

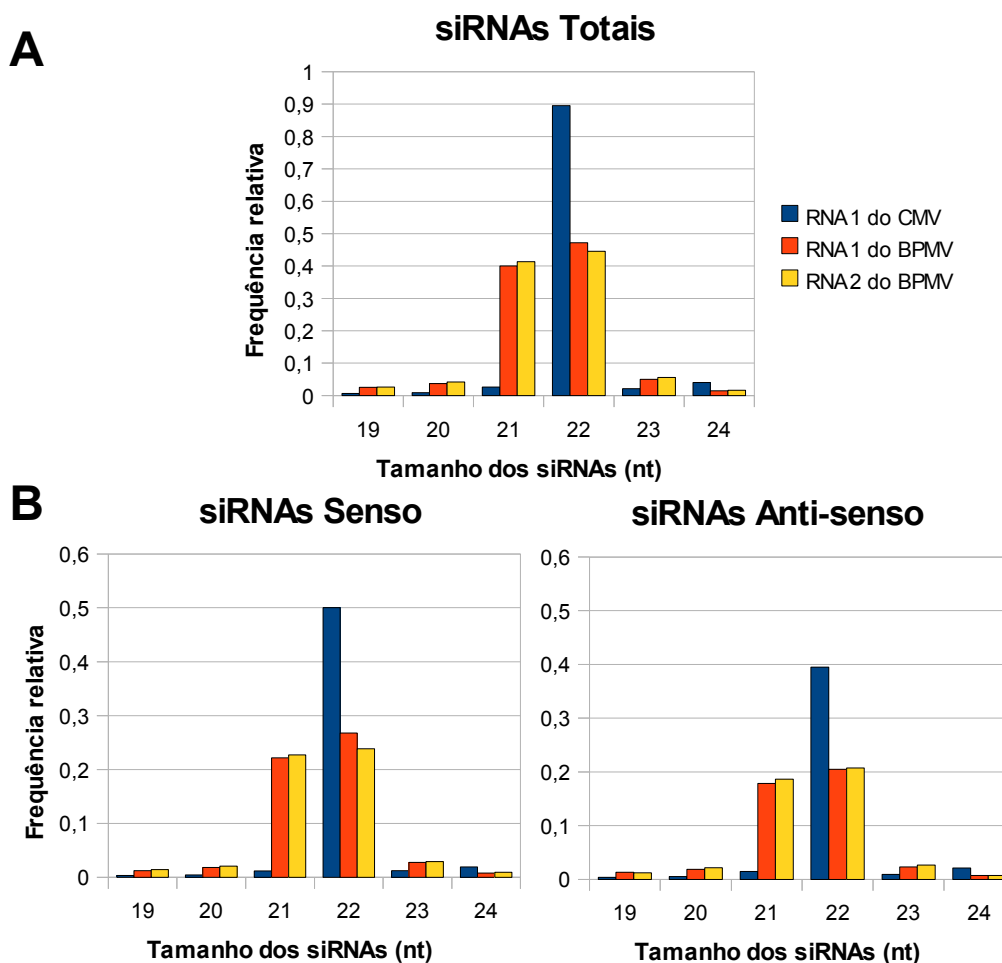


Figura 7. A) O gráfico apresenta a frequência de cada siRNA de 19 à 24 nt relativa ao total dos siRNAs do RNA 1 do CMV, RNA 1 do BPMV ou RNA 2 do BPMV dentro da biblioteca Semente. **B)** Gráficos da frequência dos siRNAs senso (esquerda) e anti-senso (direita) de 19 à 24 nt relativa ao total de siRNAs do RNA 1 do CMV, RNA 1 do BPMV e RNA 2 do BPMV.

4.5. Distribuição dos siRNAs ao longo da sequência do RNA 1 do CMV.

Para examinar a distribuição dos siRNAs foi feito um mapeamento desses ao longo da sequência do RNA 1 do CMV. Foram mapeados os siRNAs de 21, 22, e 24 nt das bibliotecas de estresse biótico (fungo) e abiótico (seca), para analisar o efeito desses estresses na expressão dos siRNAs (Figura 8 à Figura 13).

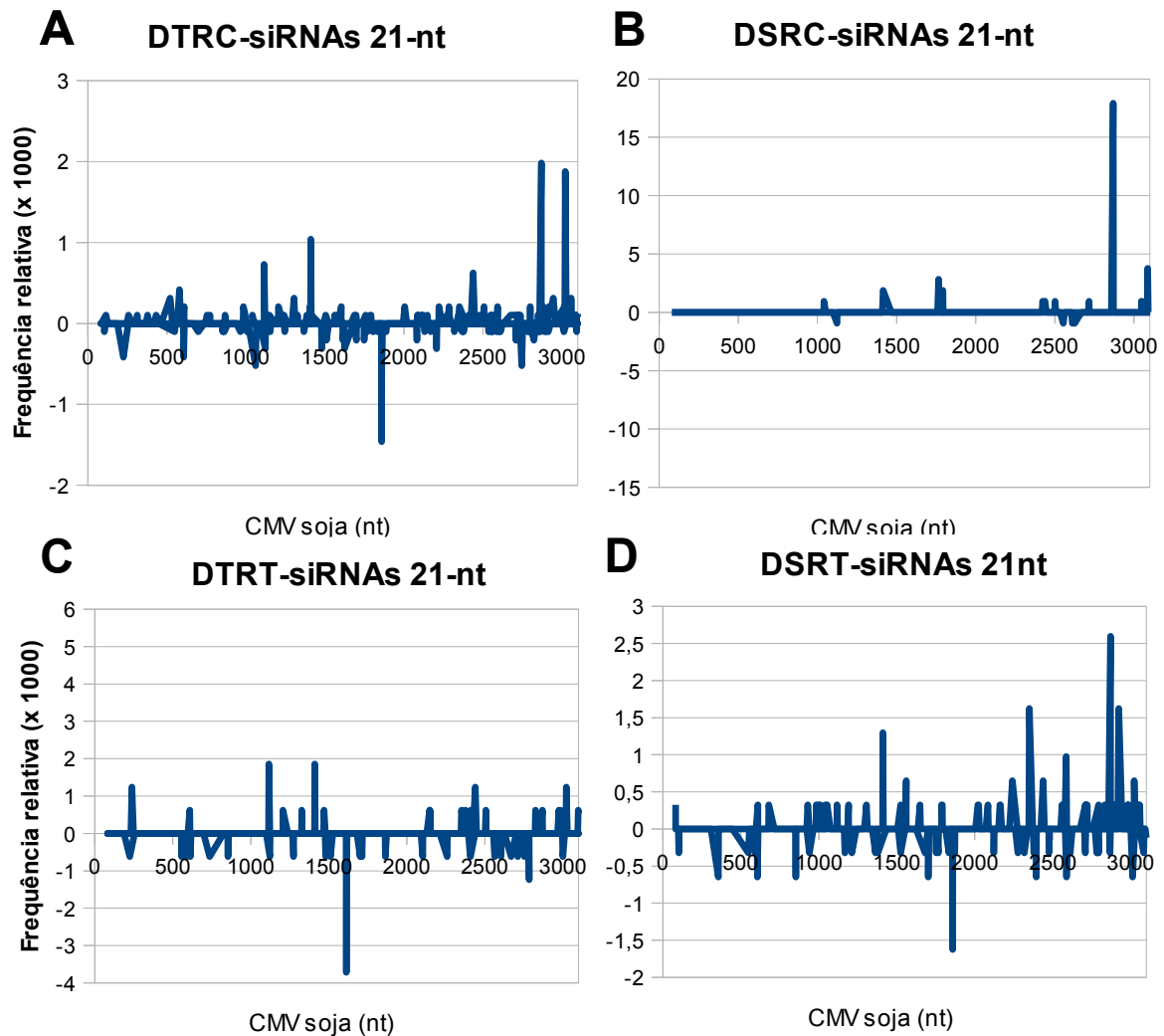


Figura 8. Distribuição dos siRNAs de 21 nt ao longo da sequência do RNA 1 do CMV para as bibliotecas de estresse abiótico (seca). A frequência relativa foi calculada a partir do total de siRNAs de cada biblioteca. **A)** Mapeamento dos siRNAs de 21 nt da biblioteca DTRC (cultivar tolerante à seca não submetida ao estresse). **B)** Mapeamento dos siRNAs de 21 nt da biblioteca DSRC (cultivar sensível à seca não submetida ao estresse). **C)** Mapeamento dos siRNAs de 21 nt da biblioteca DTRT (cultivar tolerante à seca submetida ao estresse). **D)** Mapeamento dos siRNAs de 21 nt da biblioteca DSRT (cultivar sensível à seca submetida ao estresse).

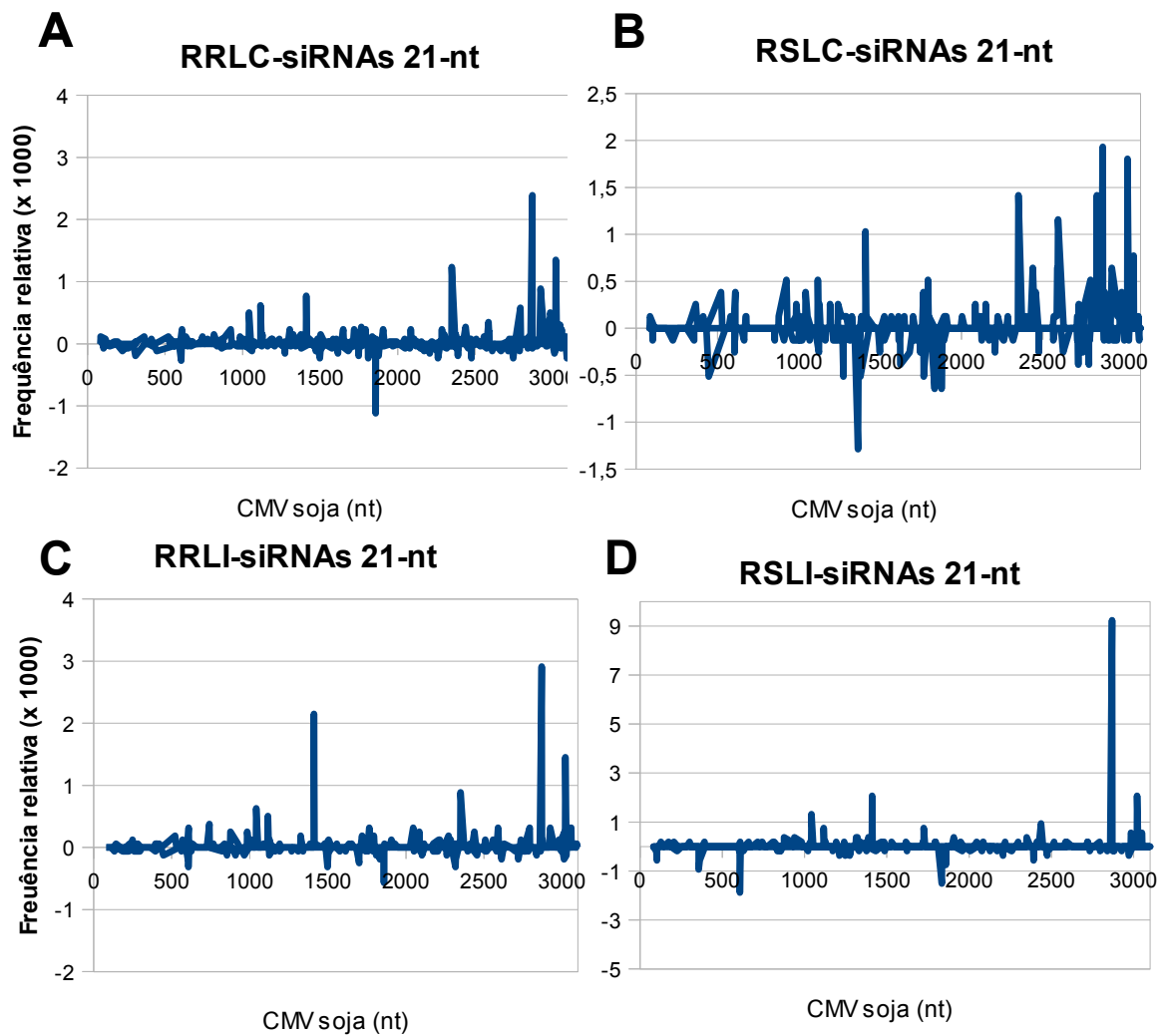


Figura 9. Distribuição dos siRNAs de 21 nt ao longo da sequência do RNA 1 do CMV para as bibliotecas de estresse biótico (fungo). A frequência relativa foi calculada a partir do total de siRNAs de cada biblioteca. **A)** Mapeamento dos siRNAs de 21 nt da biblioteca RRLC (cultivar resistente à infecção não submetida ao estresse). **B)** Mapeamento dos siRNAs de 21 nt da biblioteca RSLC (cultivar suscetível à infecção não submetida ao estresse). **C)** Mapeamento dos siRNAs de 21 nt da biblioteca RRLI (cultivar resistente à infecção submetida ao estresse). **D)** Mapeamento dos siRNAs de 21 nt da biblioteca RSLI (cultivar suscetível à infecção submetida ao estresse).

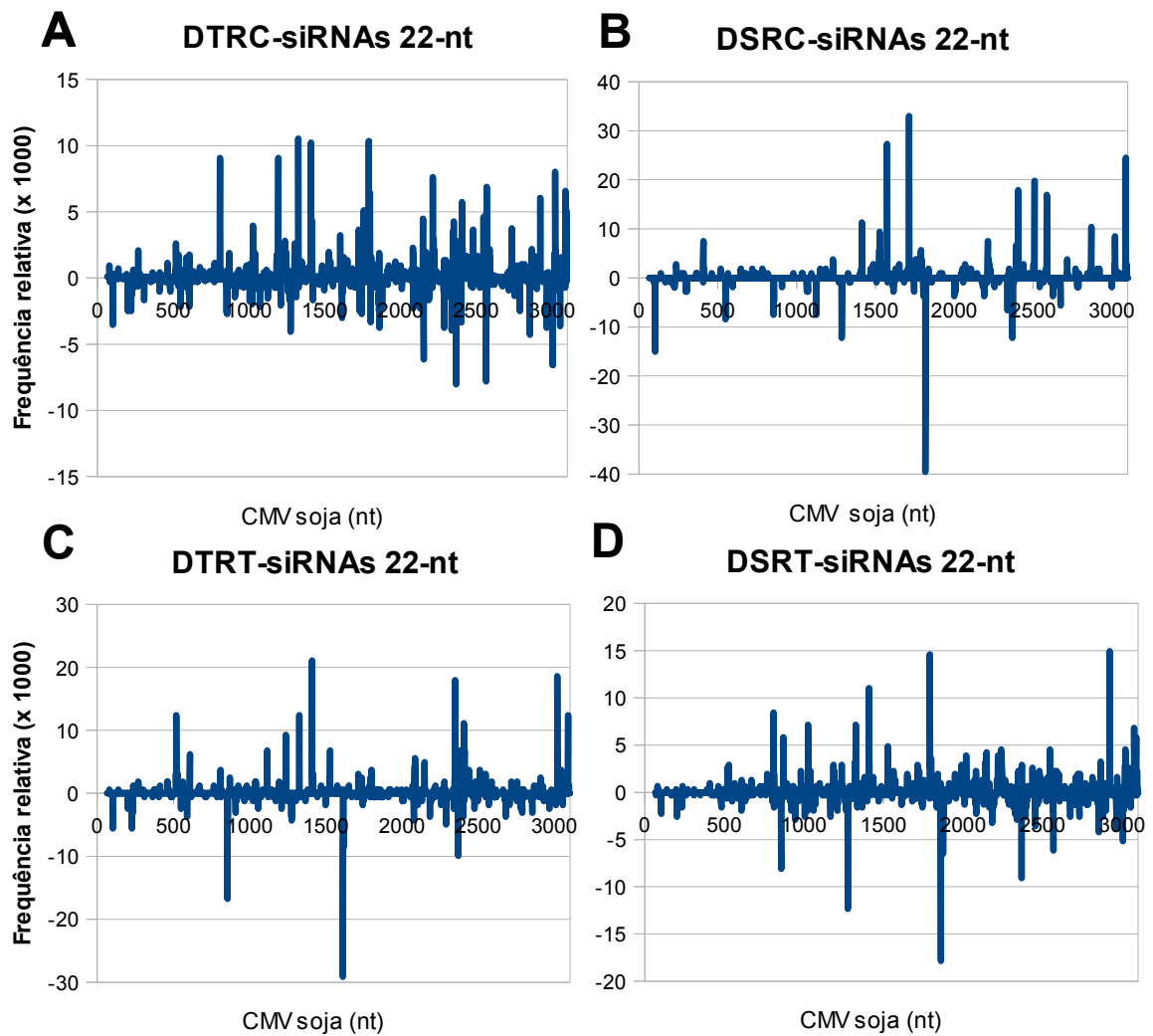


Figura 10. Distribuição dos siRNAs de 22 nt ao longo da sequência do RNA 1 do CMV para as bibliotecas de estresse abiótico (seca). A frequência relativa foi calculada a partir do total de siRNAs de cada biblioteca. **A)** Mapeamento dos siRNAs de 22 nt da biblioteca DTRC (cultivar tolerante à seca não submetida ao estresse). **B)** Mapeamento dos siRNAs de 22 nt da biblioteca DSRC (cultivar sensível à seca não submetida ao estresse). **C)** Mapeamento dos siRNAs de 22 nt da biblioteca DTRT (cultivar tolerante à seca submetida ao estresse). **D)** Mapeamento dos siRNAs de 22 nt da biblioteca DSRT (cultivar sensível à seca submetida ao estresse).

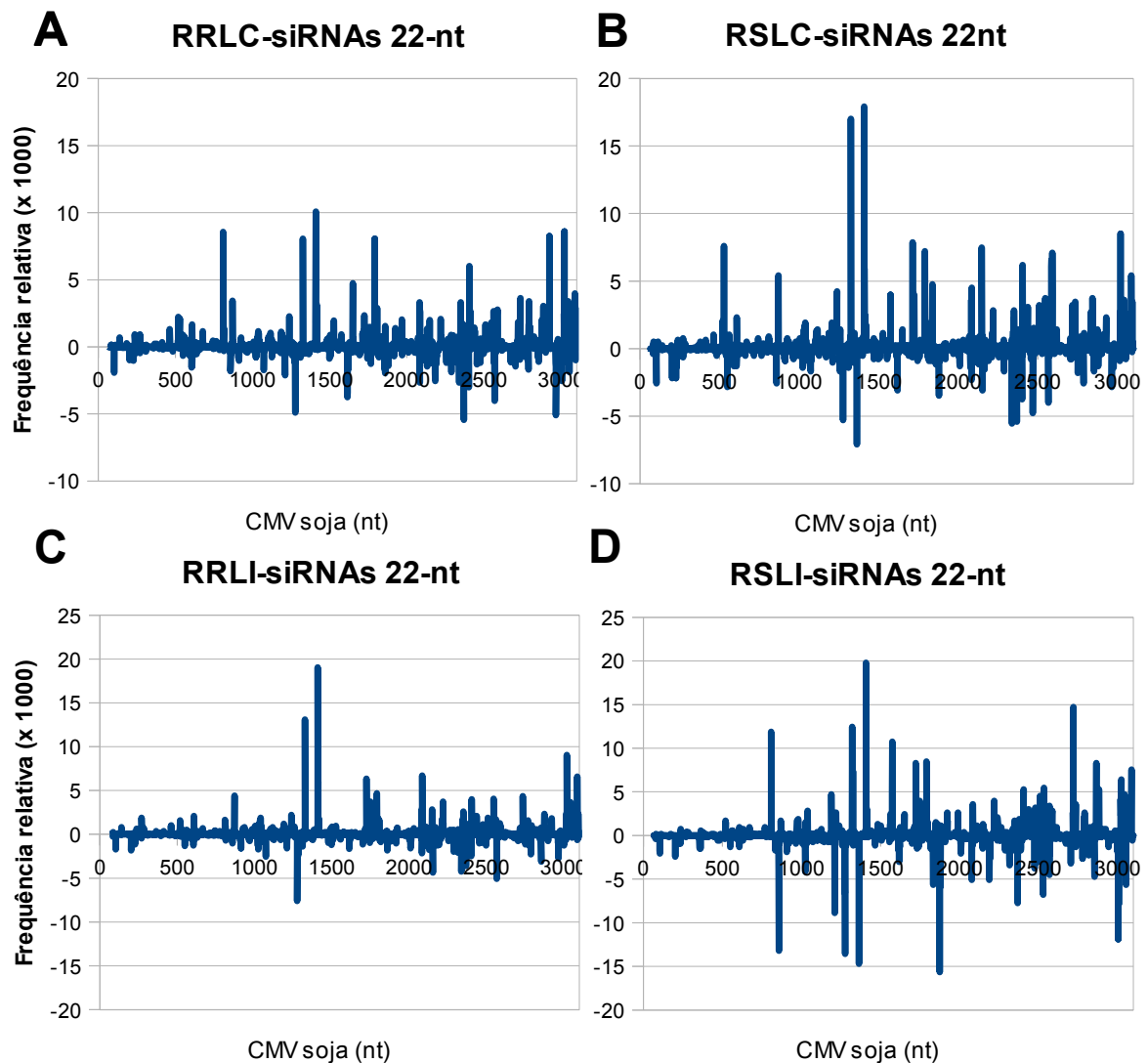


Figura 11. Distribuição dos siRNAs de 22 nt ao longo da sequência do RNA 1 do CMV para as bibliotecas de estresse biótico (fungo). A frequência relativa foi calculada a partir do total de siRNAs de cada biblioteca. **A)** Mapeamento dos siRNAs de 22 nt da biblioteca RRLC (cultivar resistente à infecção não submetida ao estresse). **B)** Mapeamento dos siRNAs de 22 nt da biblioteca RSLC (cultivar suscetível à infecção não submetida ao estresse). **C)** Mapeamento dos siRNAs de 22 nt da biblioteca RRLI (cultivar resistente à infecção submetida ao estresse). **D)** Mapeamento dos siRNAs de 22 nt da biblioteca RSLI (cultivar suscetível à infecção submetida ao estresse).

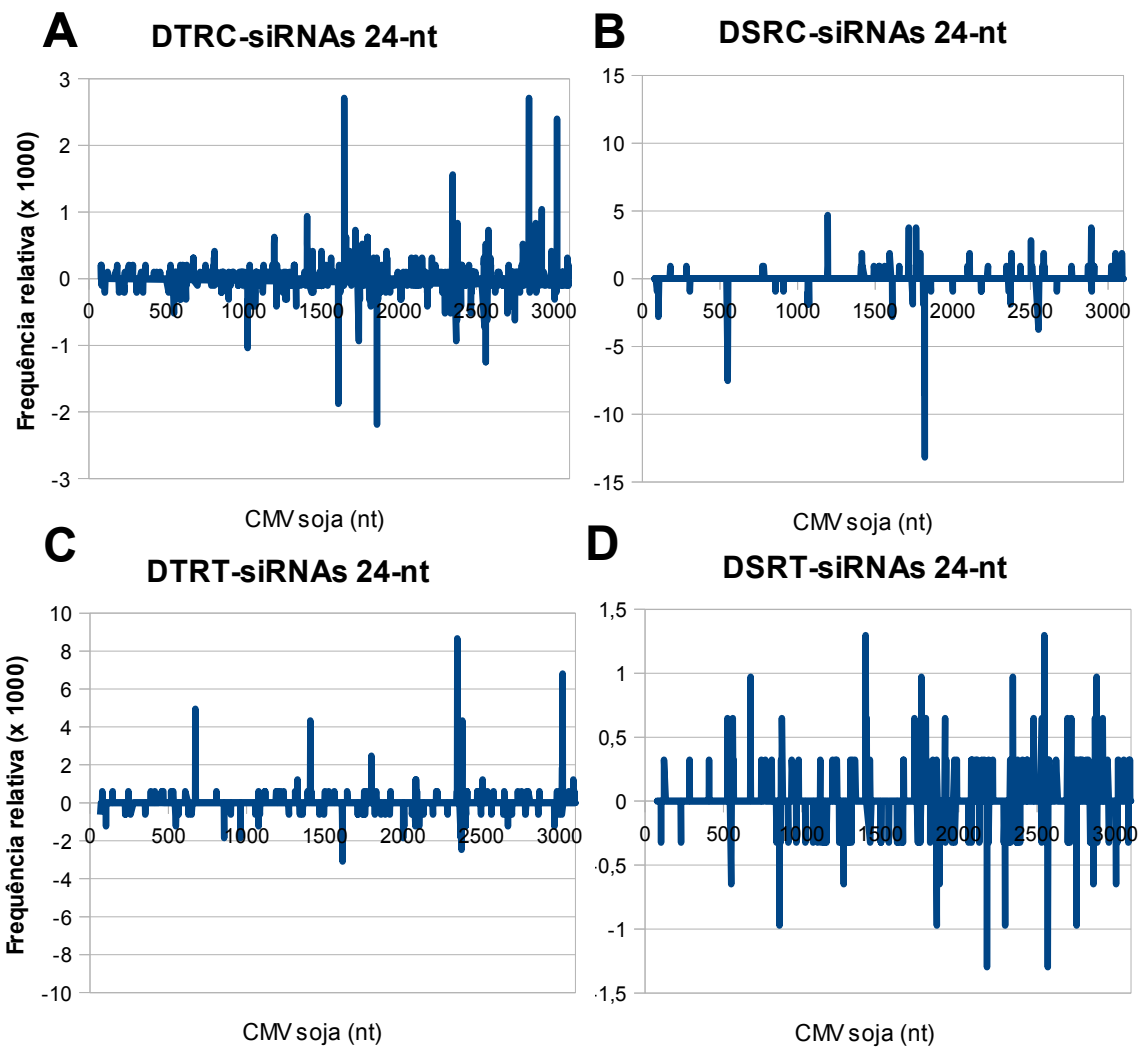


Figura 12. Distribuição dos siRNAs de 24 nt ao longo da sequência do RNA 1 do CMV para as bibliotecas de estresse abiótico (seca). A frequência relativa foi calculada a partir do total de siRNAs de cada biblioteca. **A)** Mapeamento dos siRNAs de 24 nt da biblioteca DTRC (cultivar tolerante à seca não submetida ao estresse). **B)** Mapeamento dos siRNAs de 24 nt da biblioteca DSRC (cultivar sensível à seca não submetida ao estresse). **C)** Mapeamento dos siRNAs de 24 nt da biblioteca DTRT (cultivar tolerante à seca submetida ao estresse). **D)** Mapeamento dos siRNAs de 24 nt da biblioteca DSRT (cultivar sensível à seca submetida ao estresse).

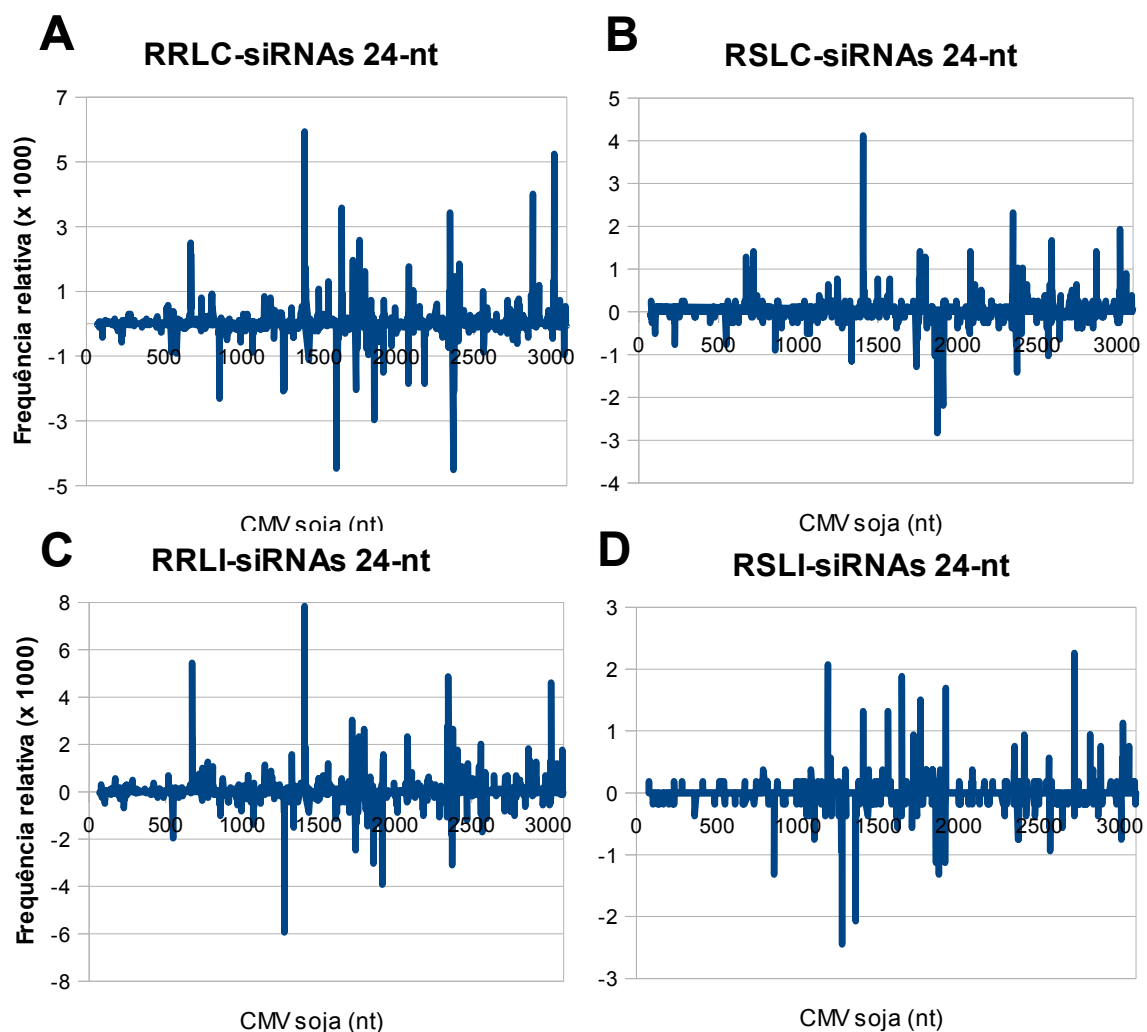


Figura 13. Distribuição dos siRNAs de 24 nt ao longo da sequência do RNA 1 do CMV para as bibliotecas de estresse biótico (fungo). A frequência relativa foi calculada a partir do total de siRNAs de cada biblioteca. **A)** Mapeamento dos siRNAs de 24 nt da biblioteca RRLC (cultivar resistente à infecção não submetida ao estresse). **B)** Mapeamento dos siRNAs de 24 nt da biblioteca RSLC (cultivar suscetível à infecção não submetida ao estresse). **C)** Mapeamento dos siRNAs de 24 nt da biblioteca RRLI (cultivar resistente à infecção submetida ao estresse). **D)** Mapeamento dos siRNAs de 24 nt da biblioteca RSLI (cultivar suscetível à infecção submetida ao estresse).

O mapeamento dos siRNAs de 19 a 24 nt revela que a sequência do RNA 1 do CMV é coberta em sua quase totalidade na maioria das bibliotecas (dados não apresentados) à exceção dos extremos 5' e 3' das sequências. Geralmente o RNA 1 do CMV apresenta 3.350 nt, comparando a sequência do RNA 1 do CMV obtida nesse trabalho com a do RNA 1 da linhagem “Legume” não foi possível mapear siRNAs para os primeiros 60 nt da extremidade 5' e nem para os últimos

200 nt do final 3' dessa sequência. A distribuição dos siRNAs ao longo do RNA 1 do CMV não é uniforme e apresenta vários pontos aonde os siRNAs são mais frequentes, sendo que algumas dessas regiões parecem ser similares entre as diferentes bibliotecas enquanto que outras variam entre as cultivares de soja ou em resposta ao tratamento aplicado (Figura 8 à Figura 13).

4.6. Análise dos siRNAs do CMV diferencialmente expressos

Foram feitas comparações das expressões dos siRNAs do RNA 1 do CMV entre as diferentes bibliotecas de estresse biótico e abiótico. Para as bibliotecas de estresse hídrico a cultivar “BR 16” (sensível à seca) apresentou dois siRNAs significativamente mais expressos na condição tratada (sob estresse) e seis na condição controle (Figura 14). Para a cultivar “Embrapa 48” (tolerante à seca) o resultado foi oito siRNAs mais expressos na condição tratada e dois na condição controle (Figura 14).

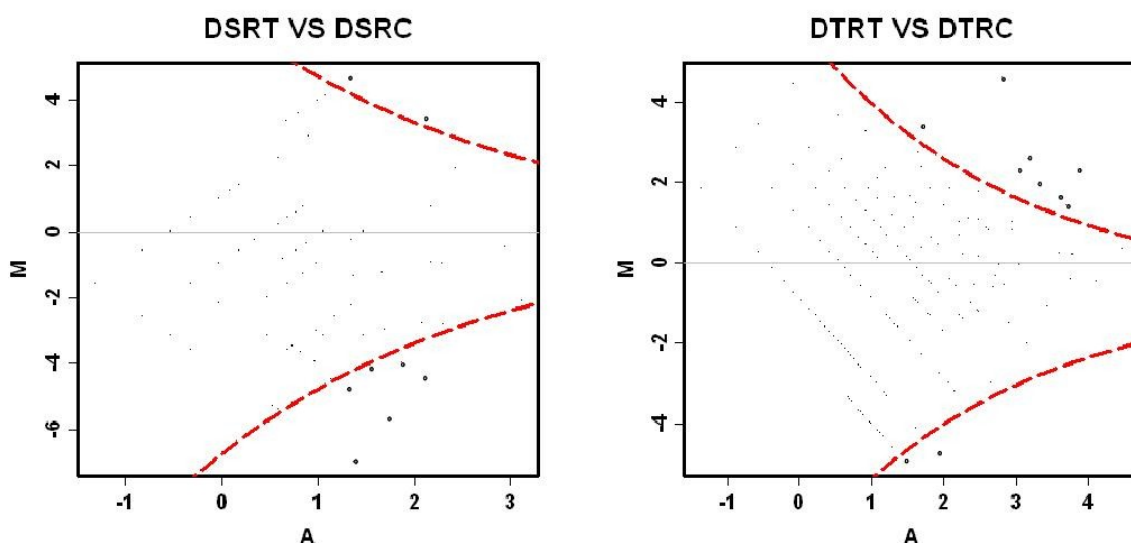


Figura 14. Gráficos apresentando os siRNAs do RNA 1 do CMV diferencialmente expressos entre as bibliotecas controle e de estresse hídrico para as cultivares “BR 16” (sensível à seca, gráfico à esquerda) e “Embrapa 48” (tolerante à seca, gráfico à direita). O eixo Y corresponde aos dados da biblioteca sob estresse e o eixo X à biblioteca controle. Os pontos acima (para biblioteca sob estresse) ou abaixo (para biblioteca controle) da linha tracejada vermelha correspondem aos siRNAs significativamente mais expressos para cada biblioteca.

Comparando as duas cultivares em condições controle ou sob estresse hídrico a “BR 16” apresentou 13 siRNAs mais expressos na condição controle e 3 no tratamento, enquanto a “Embrapa 48” apresentou 4 siRNAs mais expressos no controle e 6 sob estresse (Figura 15).

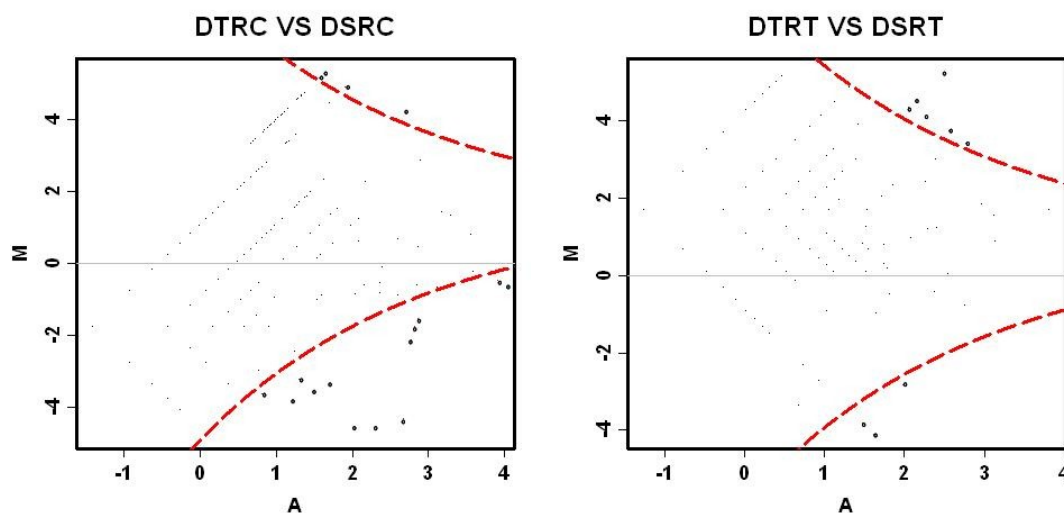


Figura 15. Gráficos apresentando os siRNAs do RNA 1 do CMV diferencialmente expressos entre as bibliotecas das cultivares “Embrapa 48” (tolerante à seca) e “BR 16” (sensível à seca) na condição controle (gráfico à esquerda) ou sob estresse hídrico (gráfico a direita). O eixo Y corresponde aos dados das bibliotecas da cultivar “Embrapa 48” e o eixo X aos das bibliotecas da cultivar “BR 16”. Os pontos acima (para as bibliotecas da cultivar “Embrapa 48”) ou abaixo (para as bibliotecas da cultivar “BR 16”) da linha tracejada vermelha correspondem aos siRNAs significativamente mais expressos para cada biblioteca.

Para às comparações entre as bibliotecas de estresse biótico a cultivar “PI561356” (resistente à ferrugem asiática) apresentou 6 siRNAs mais expressos na condição tratada e 11 siRNAs na condição controle (Figura 16). Enquanto que a cultivar “Embrapa 48” (suscetível a ferrugem asiática) apresentou 11 siRNAs mais expressos na condição sob estresse e 3 na condição controle (Figura 16).

Na comparação entre essas duas cultivares para as condições de estresse biótico e controle a “PI561356” apresentou 6 siRNAs mais expressos na biblioteca correspondente às plantas inoculadas pelo fungo e 11 na condição controle, enquanto a cultivar “Embrapa 48” obteve 11 siRNAs mais expressos na biblioteca sob tratamento e 3 na condição controle (Figura 17).

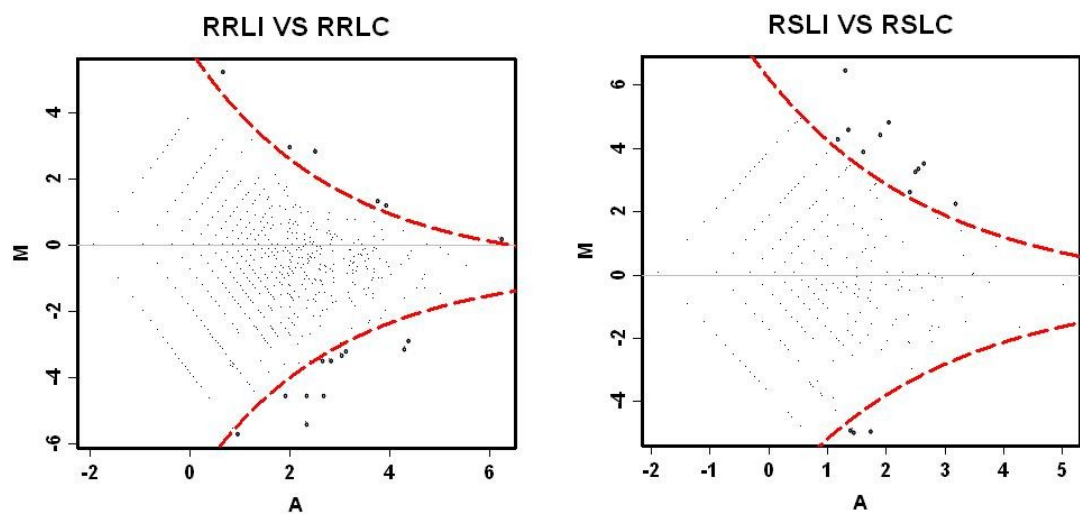


Figura 16. Gráficos apresentando os siRNAs do RNA 1 do CMV diferencialmente expressos entre as bibliotecas controle e de estresse biótico para as cultivares “PI561356” (resistente à infecção pelo fungo, gráfico à esquerda) e “Embrapa 48” (susceptível à infecção pelo fungo, gráfico à direita). O eixo Y corresponde aos dados da biblioteca sob estresse e o eixo X à biblioteca controle. Os pontos acima (para biblioteca sob estresse) ou abaixo (para biblioteca controle) da linha tracejada vermelha correspondem aos siRNAs significativamente mais expressos para cada biblioteca.

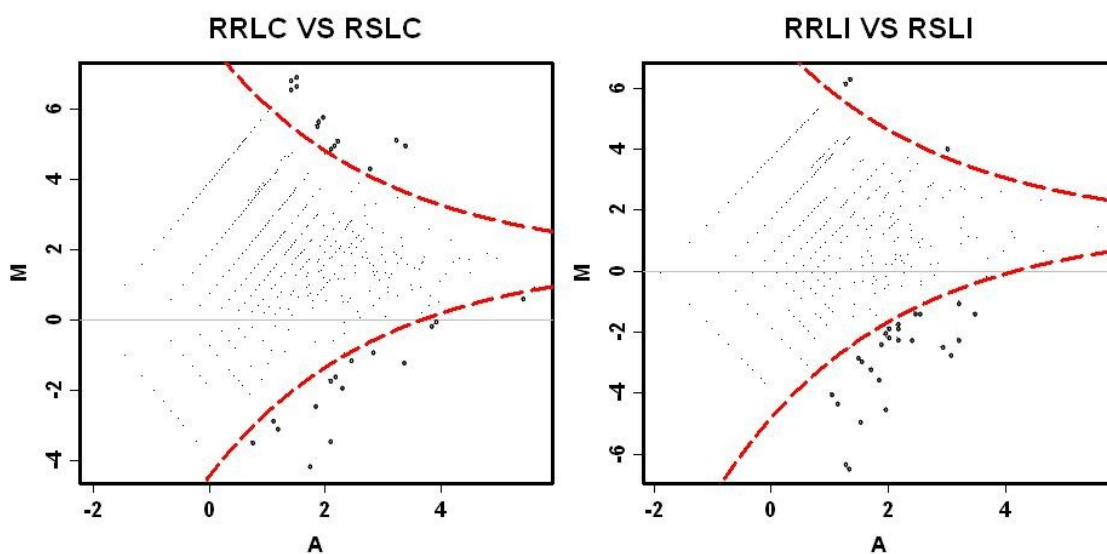


Figura 17. Gráficos apresentando os siRNAs do RNA 1 do CMV diferencialmente expressos entre as bibliotecas das cultivares “PI561356” (resistente à infecção pelo fungo) e “Embrapa 48” (susceptível à infecção pelo fungo) na condição controle (gráfico à esquerda) ou sob estresse biótico (gráfico a direita). O eixo Y corresponde aos dados das bibliotecas da cultivar “PI561356” e o eixo X aos das bibliotecas da cultivar “Embrapa 48”. Os pontos acima (para as bibliotecas da cultivar “PI561356”) ou abaixo (para as bibliotecas da cultivar “Embrapa 48”) da linha tracejada vermelha correspondem aos siRNAs significativamente mais expressos para cada biblioteca.

4.7. Prováveis alvos dos siRNAs diferencialmente expressos.

Todos os siRNAs diferencialmente expressos nas comparações entre as bibliotecas de estresse biótico e abiótico foram analisados quanto à possibilidade de terem como alvos RNAs endógenos de soja (*Glycine max*). Na comparação entre os siRNAs diferencialmente expressos das bibliotecas controle e sob estresse hídrico da cultivar “Embrapa 48” (tolerante à seca) foram encontrados 3 siRNAs mais expressos na condição tratada que possuem prováveis alvos no transcriptoma da soja. Um deles de 24 nt (lib24-6226861) possui como provável alvo um mRNA que codifica para um proteína desconhecida. Os outros dois, de 22 e 23 nt (lib22-472840 e lib23-985085, respectivamente), possuem uma sequência de nucleotídeos muito similar, com a exceção de uma adenina extra na extremidade 5' do último. Esses dois possuem como possível alvo um mRNA que codifica uma proteína desidrogenase/reductase de cadeia curta. Na comparação entre os siRNAs diferencialmente expressos das bibliotecas controle e inoculada com o fungo da ferrugem asiática da cultivar “PI 561356” (resistente à infecção pelo fungo) 2 siRNAs mais expressos na condição sob o estresse biótico e 3 na condição controle possuem prováveis alvos. Na condição sob estresse biótico um siRNA de 22 nt (lib22-413149) tem como alvo um mRNA que codifica uma proteína desconhecida da família das transferases. Um outro siRNA de 24 nt (lib24-5226416) possui como alvo um mRNA que codifica a subunidade 9 do proteossoma 26S. Curiosamente esse mRNA foi encontrado em bibliotecas de cDNA de soja sob estresse hídrico. Na condição controle um siRNA de 22 nt (lib22-931147) tem 3 possíveis alvos, dois que codificam uma serina-treonina quinase e um que codifica uma proteína de função desconhecida. O mRNA de uma das serina-treonina quinases foi encontrado em bibliotecas de cDNA de soja sob a infecção do fungo *P. pachyrhizi*. Um outro siRNA de 22 nt (lib22-472099) possui como alvo dois mRNAs também encontrados em bibliotecas sob estresse hídrico. Um que codifica uma ATPase da família AAA e outro que codifica uma N-acetilglicosamina-transferase hipotética. O último desses siRNAs possui 24 nt (lib24-5594589) e tem uma sequência muito similar ao lib22-472099 com exceção de uma timina e uma guanina na extremidade 5' e tem como alvo a ATPase da

família AAA. Para os siRNAs diferencialmente expressos entre as bibliotecas sob tratamento controle e sob estresse abiótico da cultivar “BR 16” (sensível à seca) e controle e estresse biótico da cultivar “Embrapa 48” (suscetível ao fungo) não foram encontrados prováveis alvos.

Também foram encontrados alvos para os siRNAs diferencialmente expressos nas comparações entre as cultivares sensíveis e resistentes à estresse sob o mesmo tratamento. Na comparação entre a cultivar “BR 16” e “Embrapa 48” sob estresse abiótico foram encontrados os siRNAs já mencionados lib24-6226861, lib22-472840 e lib23-985085 mais expressos na cultivar resistente à seca. Na comparação entre as cultivares “Embrapa 48” e “PI 561356” sob estresse biótico 5 siRNAs de 22 nt mais expressos na cultivar suscetível à inoculação pelo fungo (“Embrapa 48”) possuem prováveis alvos no transcriptoma da soja. O lib22-304330 tem como alvo um RNA desconhecido de 220 nt. O lib22-176284 tem como alvo um RNA que codifica uma proteína similar a ZEITLUPE de *Arabidopsis thaliana*, que é uma proteína que possui o domínio PAS que tem como função ser um receptor de sinais do ciclo circadiano da planta (SOMERS *et al.*, 2000). O siRNA Lib22-856527 tem como alvo um mRNA que codifica uma proteína citocromo P450 monoxigenase, esse mRNA foi encontrado em bibliotecas de cDNA de soja sob estresse hídrico. O siRNA lib22-455649 tem como alvo um mRNA que codifica uma proteína serina-treonina fosfatase. E por fim o lib22-337567 possui um sequência muito similar à um outro siRNA já descrito o lib24-5226416 e também possui o mesmo alvo um mRNA que codifica a subunidade 9 do proteassoma 26S. Na comparação entre as mesmas cultivares mas sob o tratamento controle, foram encontrados 3 siRNAs mais expressos na cultivar “PI 561356”. Dois deles já descritos anteriormente, o lib22-931147 e o lib24-5594589. O terceiro siRNA lib22-1252098 tem uma sequência muito similar ao lib22-931147 e também tem como alvo o mesmo mRNA, encontrado em bibliotecas de cDNA de plantas de soja inoculadas pelo fungo *P. pachyrhizi*, que codifica uma serina-treonina quinase.

4.8. Análise da expressão do RNA e dos siRNAs do RNA 1 do CMV.

Por RT-qPCR foram medidas as expressões do RNA 1 do CMV para as amostras dos experimentos de estresse biótico e abiótico. Para as amostras de estresse abiótico a expressão do RNA 1 do CMV foi significativamente maior ($P < 0,05$) nas plantas sob estresse em comparação ao controle (Figura 18B). Sob estresse hídrico a cultivar “Embrapa 48” (tolerante à seca) apresentou uma expressão significativamente maior em relação a cultivar “BR 16” (Figura 18B).

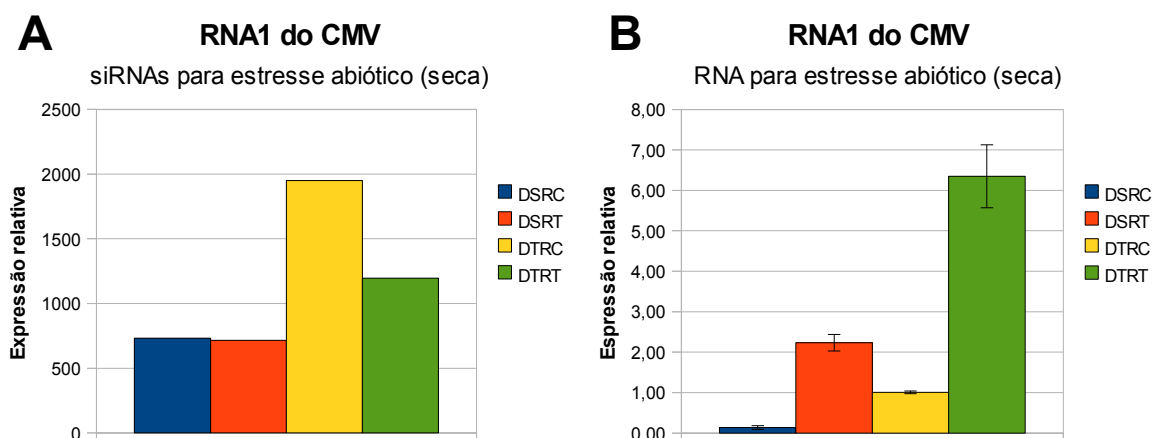


Figura 18. A) Expressão dos siRNAs de 19-24 nt para as bibliotecas de estresse abiótico. Os dados de expressão relativa correspondem ao total de siRNAs de 19-24 nt do RNA 1 do CMV normalizados pelo total de sRNAs sequenciados da biblioteca. **B)** Resultado da análise da expressão do RNA 1 do CMV por RT-qPCR. Os dados de cada amostra foram normalizados pela expressão dos miRNAs 156a, 156b e 1520d.

Para as amostras de estresse biótico a expressão do RNA 1 do CMV foi significativamente maior na cultivar “Embrapa 48” sob estresse biótico em comparação ao controle (Figura 19B).

Com relação a expressão dos siRNAs originados do RNA 1 do CMV integrado ao genoma da soja, utilizando o teste do qui-quadrado, foi verificada uma diminuição da expressão sempre nas bibliotecas das amostras que foram submetidas ao estresse biótico e abiótico com a exceção das amostras da cultivar “BR 16” (sensível à seca), na qual a diferença não foi significativa (Figura 18A e Figura 19A). As cultivares resistentes ao estresse biótico e abiótico também apresentaram a expressão dos siRNAs do CMV significativamente maior em relação as cultivares sensíveis ao estresse sob o mesmo tratamento (Figura 18A e Figura 19A).

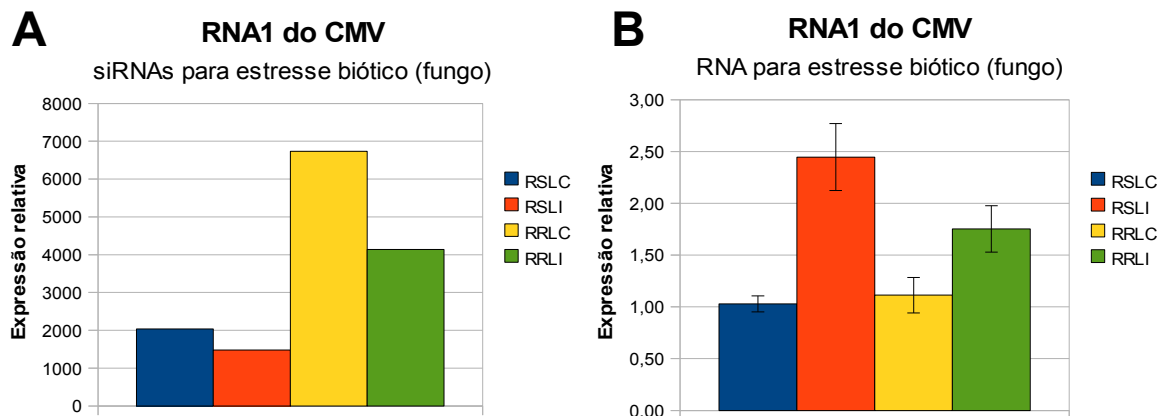


Figura 19. A) Expressão dos siRNAs de 19-24 nt para as bibliotecas de estresse biótico. Os dados de expressão relativa correspondem ao total de siRNAs de 19-24 nt do RNA 1 do CMV normalizados pelo total de sRNAs sequenciados da biblioteca. **B)** Resultado da análise da expressão do RNA 1 do CMV por RT-qPCR. Os dados de cada amostra foram normalizados pela expressão dos miRNAs 156a e 156b.

A expressão dos siRNAs de 19-24 nt do RNA 1 e RNA 2 do BPMV em comparação ao RNA 1 do CMV encontrados na biblioteca Semente também foi analisada. A expressão dos siRNAs do RNA 1 do BPMV foi cerca de 20 vezes maior e a do RNA 2 cerca de 10 vezes maior do que a do RNA 1 do CMV (Figura 20).

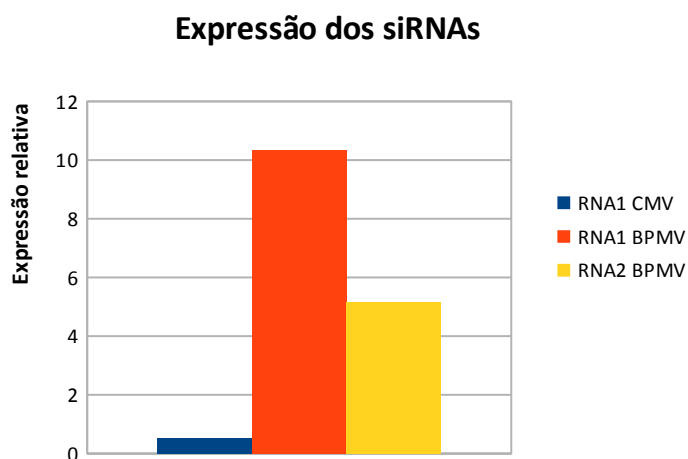


Figura 20. A expressão relativa corresponde ao total dos siRNAs de 19-24 nt do RNA 1 do CMV, RNA 1 do BPMV ou RNA 2 do BPMV normalizado pelo tamanho da sequência de cada RNA. Todos esses siRNAs são provenientes da biblioteca Semente.

5. DISCUSSÃO

5.1. Integração do CMV no genoma da soja (*Glycine max*)

Neste trabalho identificamos a presença de sRNAs de 19 a 24 nt com homologia à sequência do RNA 1 do CMV em 17 bibliotecas de sRNAs diferentes de plantas de soja de 5 cultivares. Foi feita a montagem *de novo* com os sRNAs do CMV e o resultado foram sequências idênticas e homólogas ao RNA 1 desse vírus. As cinco cultivares - “Embrapa 48”, “Conquista”, “BR 16”, “Urano” e PI 561356” - foram desenvolvidas no Brasil. Como esses sRNAs foram encontrados em plantas diferentes, cultivadas em locais distintos e ainda tendo sido encontrada homologia apenas ao RNA 1 e não aos outros RNAs desse vírus foi cogitada a hipótese desses sRNAs serem provenientes de uma integração do vírus no genoma da soja e não de uma eventual infecção viral. Corroborando com essa hipótese trabalhos recentes demonstraram que infecções do CMV e de outros vírus em plantas resultam na geração de vsiRNAs em todo o genoma viral (DONAIRE *et al.*, 2009, WANG *et al.*, 2010). O CMV é um vírus de RNA não retroviral e portanto incapaz de integrar-se ao genoma do hospedeiro, por não codificar uma proteína transcriptase reversa e nem a integrase que em retrovírus catalizam a síntese de um cDNA a partir do RNA viral e a integração desse ao DNA nuclear, respectivamente (DEFARGES & CIUFFI, 2010). Entretanto, foi demonstrado que em mamíferos o vírus da coriomeningite linfocítica (LCMV, do inglês, “Lymphocytic Choriomeningitis Virus”), que é um vírus de RNA não retroviral, é capaz de gerar cDNA que persiste nas células mesmo após a eliminação do RNA viral (KLENERMAN *et al.*, 1997). Esse cDNA é capaz de codificar proteínas virais, mas é presente em poucas células do hospedeiro (uma molécula de DNA a cada 10^4 a 10^5 células) e parece servir como uma vacina de DNA auxiliando na manutenção da memória imunológica, explicando em parte a presença de antígenos derivados do LCMV mesmo após longos períodos da infecção viral (KLENERMAN *et al.*, 1997). A ação de uma transcriptase reversa proveniente de elementos retrovirais endógenos (que correspondem de 10 a 20% do genoma de mamíferos) ou de uma infecção paralela de um retrovírus foi

sugerida como a responsável pela síntese desse cDNA (KLENERMAN *et al.*, 1997). Entretanto evidências para tal atividade só foram demonstradas anos mais tarde. O mesmo grupo demonstrou que o RNA do LCMV se recombina com um elemento de partícula-A intracisternal (IAP), que é um retrotransposon, que por sua vez não só forma o cDNA quimérico como promove a sua integração ao genoma do hospedeiro (GEUKING *et al.*, 2009). Essa recombinação se dá em todos os casos relatados entre o final 3' do RNA viral, após o codón de terminação do gene, e a região 3' do elemento IAP, além disso não é encontrada nenhuma duplicação no sítio alvo da integração, indicando que não há uma ação canônica da integrase retroviral. Foram verificadas múltiplas deleções flanqueando o ponto de integração do cDNA quimérico o que pode ser um indício de que a inserção ocorra em sítios aonde o DNA foi danificado (GEUKING *et al.*, 2009). Por fim a transcrição e a transposição dos elementos IAP ocorre mais eficientemente no timo e durante a espermatogênese, o que para o último caso pode levar a integração do LCMV para as gerações futuras do hospedeiro (GEUKING *et al.*, 2009). Evidências para a persistência dessa integração na prole só foram reveladas em trabalhos posteriores. O vírus da doença de Borna (BDV) é um vírus de RNA não retroviral que possui genoma de orientação negativa e um ciclo de vida associado majoritariamente ao núcleo da célula, sendo o único vírus de RNA humano com essa característica (TOMONAGA *et al.*, 2002). Assim como para o LCMV em infecções do BDV também é encontrada a presença de cDNA proveniente do mRNA viral e elementos endógenos do gene da nucleoproteína (N) do BDV (EBLN) foram encontrados no genoma de mamíferos (HORIE *et al.*, 2010). A proteína N é a responsável pela encapsidação e formação do nucleocapsídeo do BDV. Os EBLNs são capazes de gerar mRNAs em humanos e macacos e há indícios de que alguns desses sejam traduzidos à proteínas (EWING *et al.*, 2007, HORIE *et al.*, 2010). Foram encontradas sequências duplicadas no sítio de integração dos EBLNs, assim como deleções e rearranjos nas sequências adjacentes a integração, o que indica a ação de uma transcriptase reversa e da integrase de retrotransposons. De fato foram encontradas evidências para a recombinação de mRNAs do BDV com RNAs dos elementos nucleares interespaçados longos (LINE), uma classe de retrotransposons (HORIE *et al.*,

2010). Reforçando essa possibilidade já foi demonstrado que os LINE-1s são capazes de gerar pseudogenes através de sua recombinação com mRNAs em mamíferos (ESNAULT *et al.*, 2000). Outros casos de vírus de RNA não retroviral integrado (NIRV) ao genoma do hospedeiro foram relatados em mamíferos. RNAs virais das proteínas NP e L de vírus da família *Filoviridae* também se integraram ao genoma de mamíferos, aparentemente através de retrotransposons da classe LINE-1 de maneira similar ao já relatado anteriormente (TAYLOR *et al.*, 2010). NIRVs provenientes de vírus de dsRNA também foram encontrados em fungos, em alguns casos o genoma viral inteiro e a sua integração muito provavelmente foi realizada pelos retroelementos Ty (TAYLOR & BRUENN, 2009). Em insetos do gênero *Aedes* através de um sistema de integração que também envolve retrotransposons foram encontrados NIRVs para algumas das proteínas estruturais virais do gênero *Flavivirus* (CROCHU *et al.*, 2004). Entretanto não há relatos desse tipo de integração em plantas. Esse trabalho é o primeiro a mostrar evidências de um NIRV no genoma de plantas. Além dos siRNAs encontrados para o RNA 1 do CMV, por PCR foi identificada a presença dessa sequência no DNA de pelo menos mais 7 cultivares de soja além das já mencionadas (Figura 4). Essas evidências comprovam a existência do RNA 1 do CMV na forma de DNA, como já foi relatado para outros vírus. Também foi demonstrado a presença do RNA 1 do CMV na soja através de RT-qPCR (Figura 18B e Figura 19B). Plantas possuem retrotransposons similares aos encontrados em outros eucariotos que possuem NIRVs em seu genoma (MARCO & MARÍN, 2005), possibilitando a síntese do cDNA e integração de RNAs virais em plantas. Apesar de ser detectada a presença do cDNA do RNA 1 do CMV em 12 cultivares de soja não foi encontrado nenhum amplicon na cultivar “Willians” (Figura 4). A “Willians” é uma cultivar desenvolvida nos Estados Unidos e teve o seu genoma completamente sequenciado (SCHMUTZ *et al.*, 2010). Usando a ferramenta BLASTn (<http://www.phytozome.net/search.php?show=blast>) não foi possível achar nenhuma sequência com homologia ao RNA 1 do CMV em seu genoma (dados não apresentados). Interessantemente em outras duas cultivares também provenientes dos Estados Unidos a “Davis” e a “Hill” foi detectada a presença de DNA amplificável e com sequência homóloga a do RNA 1 do CMV, o que indica

que a integração pode ter ocorrido recentemente. Apesar de existirem ancestrais asiáticos em comum às 3 cultivares norte americanas, existem algumas cultivares asiáticas em comum apenas à “Davis” e “Hill” o que pode levar a data de integração à um período mais remoto. Para uma datação mais precisa desse evento tais ancestrais devem ser verificados quanto a presença dessa sequência por PCR e posterior sequenciamento.

5.2. Perfil dos siRNAs do RNA1 do CMV e dos RNAs 1 e 2 do BPMV.

Os siRNAs do RNA 1 do CMV foram distribuídos continuamente ao longo de boa parte da sequência quando comparada a outros RNAs 1 de diferentes linhagens desse vírus encontrados no banco de dados do NCBI. Entretanto para uma região de aproximadamente 60 nt que corresponde ao final 5' das sequências do RNA 1 do CMV não foi encontrado nenhum siRNA. No outro extremo quando comparado às outras sequências presentes no banco de dados há a falta de uma região conservada de mais de 200 nt (o comprimento dessa região varia entre as linhagens podendo chegar à quase 600 nt). Curiosamente regiões de aproximadamente 200 nt no final 3' dos RNAs genômicos do CMV e de outros vírus de RNA de plantas apresentam uma estrutura secundária similar em forma aos RNAs transportadores (tRNAs), que possui como provável função auxiliar a replicase viral no processo de início da transcrição da fita de orientação negativa (DREHER, 2009). Outros relatos apontam, no entanto, que a replicação viral não é dependente de tais estruturas e pode ocorrer simplesmente a partir de regiões iniciadoras conservadas encontradas nas extremidades dos RNAs virais (DREHER, 2009). Estruturas similares a tRNAs podem ser reconhecidas pela DCL 1, já que essa proteína em *A. thaliana* tem como alvo principal estruturas secundárias em forma de grampo, sendo a principal responsável pela síntese de miRNAs em plantas (MOISSIARD & VOINNET, 2006, XIE & QI, 2008). Assim uma das hipóteses para a ausência de siRNAs para uma região de aproximadamente 200 nt do terminal 3' é a de que esse RNA primeiramente foi processado pela DCL 1 e posteriormente se recombinau à algum retrotransposon que proporcionou a sua integração ao genoma da planta. Isso porém não explica a falta dos 60 nt

no terminal 5'. Uma possibilidade seria a ação de exonucleases. Outra hipótese para a falta de siRNAs dessas regiões vem da constatação que em infecções virais vsRNAs são encontrados em menor abundância nas extremidades das sequências dos RNAs e em alguns casos não cobrem completamente essas regiões (DONAIRE *et al.*, 2009, QI *et al.*, 2009). Então a cobertura total do RNA 1 do CMV pode não ter sido atingida pelo número baixo de sRNAs encontrados para essa sequência nas nossas bibliotecas (variando de 1061 a 25915). Entretanto para o RNA 2 do BPMV que possui um tamanho similar (3662 nt) ao RNA 1 do CMV a cobertura foi completa com a exceção de apenas 4 nucleotídeos (quatro timinas) e o número de sRNAs homólogos à essa sequência foi de 18.874 o que é até mais baixo o que o encontrado para o CMV na biblioteca RRLC. Para o RNA 1 do BPMV a cobertura também foi próxima de 100% com a exceção de apenas algumas adeninas da cauda poli A. Maiores explicações que justifiquem a falta de siRNAs para as extremidades da sequência do RNA 1 do CMV só serão obtidas com o sequenciamento da região adjacente ao ponto de integração no genoma da soja.

No caso do BPMV, as classes mais abundantes de siRNAs foram de 21 nt e 22 nt, correspondendo a 46% e 40% do total de siRNAs provenientes do RNA 1 e RNA 2, respectivamente (Figura 7A). Os resultados encontrados para o BPMV são similares ao encontrado para outros vírus de plantas aonde as classes de 21 e 22 nt são as mais abundantes em infecções virais (DONAIRE *et al.*, 2009). Em plantas a DCL4 é a principal geradora de vsRNAs, seguida da DCL2 ambas produzindo siRNAs de 21 e 22 nt respectivamente (LLAVE, 2010). Portanto ortólogos dessas enzimas em soja devem ser os responsáveis pela geração dos vsRNAs do BPMV encontrado na biblioteca Semente. O perfil para os siRNAs do CMV foi bem diferente. Em todas as bibliotecas a classe de siRNAs de 22 nt (em média 73%) foi a mais abundante entre os siRNAs homólogos à sequência do CMV seguido pela de 24 nt (em média 16%) (Figura 5). Considerando a hipótese de que o CMV se integrou ao genoma através de uma recombinação com um retrotransposon, esse perfil é inesperado. Em *A. thaliana* a classe mais abundante de siRNAs para esses elementos é a de 24 nt, que desencadeia a rota de RdDM, que tem por objetivo diminuir a transcrição e conseqüentemente a atividade dos

transposons (JAMALKANDI & MASOUDI-NEJAD, 2009). Isso pode explicar o fato de a classe de 24 nt ser a segunda mais abundante, podendo ser uma resposta da célula vegetal à transcrição de um RNA aberrante similar ao que ocorre com os transposons e retrotransposons. A RDR2 é uma enzima necessária para a geração de siRNAs de 24 nt junto da DCL3 (LU *et al.*, 2005, JAMALKANDI & MASOUDI-NEJAD, 2009). Em milho, ao contrário do que ocorre em *A. thaliana*, mutantes para o gene da RDR2 geram siRNAs de 22 nt ao invés de 24 nt em regiões de transposons, retrotransposons e de sequências altamente repetitivas no genoma (NOBUTA *et al.*, 2008). Indicando uma rota alternativa para a geração de siRNAs para essas regiões, o que pode ser o caso para a geração dos siRNAs de 22 nt do CMV encontrado nas nossas bibliotecas. Dados similares são encontrados em *Populus trichocarpa*. Em uma análise de bibliotecas de sRNAs dessa planta retrotransposons do tipo LTR (Repetições Terminais Longas ou “Long Terminal Repeat”) apresentam siRNAs de 22 nt como a classe dominante e em alguns casos uma representação quase idêntica das classes de 22 e 24 nt (KLEVEBRING *et al.*, 2009). Evidências de distintas rotas para a geração de siRNAs derivados de transposons em outras plantas que contrastam com as encontradas em *A. thaliana* podem ser indícios que em soja processos similares aos encontrados em álamo e milho podem estar atuando para a geração dos siRNAs do CMV.

A maioria dos siRNAs tanto derivados do CMV quanto do BPMV possuem orientação senso (Figura 6 e Figura 7). Esses dados coincidem com os encontrados para vsiRNAs em infecções virais de plantas (DONAIRE *et al.*, 2009 QI *et al.*, 2009). Já que moléculas de dsRNAs sintetizadas por RDRs ou replicases de vírus de RNA ao serem clivadas por DCLs deveriam gerar uma quantidade igual de siRNAs para ambas as orientações uma das explicações para a maior produção de siRNAs de orientação senso em RNAs virais é a de que estruturas secundárias em forma de grampo são a fonte para a geração dessas moléculas, provavelmente com o envolvimento da DCL 1, conhecida por ter maior afinidade para a clivagem de dsRNAs provenientes desses tipos de estruturas (MOISSIARD & VOINET, 2006). No entanto estudos recentes não tem encontrado relação direta entre estruturas em forma de grampo e regiões

altamente produtoras de siRNAs (LLAVE, 2010). Confrontando os resultados relatados nesses estudos, dos 5 siRNAs derivados do CMV mais abundantes das nossas bibliotecas de sRNAs somente um deles é encontrado em uma estrutura em forma de grampo similar as encontradas em pre-miRNAs (dados não apresentados). Portanto a principal fonte para a geração dos siRNAs encontrados nas bibliotecas estudadas parece ser dsRNAs sintetizados pelas diferentes RDRs de plantas ou por uma replicase viral (no caso do BPMV) e posteriormente clivados por DCLs.

5.3. Expressão do RNA 1 e dos siRNAs derivados do CMV.

A expressão do RNA 1 do CMV foi significativamente maior nas plantas sob estresse (biótico ou abiótico) em comparação ao tratamento controle, com a exceção da cultivar “PI 561356” (resistente à inoculação pelo *P. pachyrhizi*) que também apresentou um aumento, porém não significativo (Figura 18B e 19B). Com relação aos siRNAs totais de 19 a 24 nt do CMV a sua expressão parece diminuir nas plantas sob estresse, à exceção da cultivar “BR 16” (tolerante à seca) que não apresenta uma diminuição significativa (Figura 18A e 18B). Uma possível explicação para esses resultados se baseia na hipótese de o RNA 1 do CMV ter se integrado no genoma da soja via recombinação com um retrotransposon. Vários transposons e principalmente retrotransposons parecem ter a sua transcrição ativada em plantas sob estresse tanto biótico quanto abiótico (GRANDBASTIEN, 1998). Como já dito anteriormente, em plantas, transposons e retrotransposons tem a sua transcrição reprimida pela rota de RdRM, que produz hc-siRNAs que promovem a metilação de regiões homólogas a sua sequência no genoma. Em arroz dois siRNAs (siR441 e siR446) derivados de um transposon da classe MITE (“miniature inverted-repeat transposable element”) apresentam a sua expressão diminuída ao contrário de seus precursores que são regulados positivamente em diversas situações de estresse (YAN *et al.*, 2011). Entretanto apesar desses siRNAs serem sintetizados pela DCL 3, a sua produção é independente de RDR 2. Tais siRNAs parecem ser formados a partir de estruturas em forma de grampo, similar aos pre-miRNAs, porém como seus precursores

também dão origem à outras espécies de sRNAs eles não são classificados como miRNAs (MEYERS *et al.*, 2008). Apesar de o mecanismo pelo qual uma maior expressão dos precursores inibe a síntese desses siRNAs não estar bem elucidado, um processo similar pode estar ocorrendo para os siRNAs do CMV. Por motivos já descritos anteriormente a maioria dos siRNAs do CMV parecem ser produzidos a partir de um precursor de dsRNA polimerizado por uma RDR (provavelmente uma ortóloga da RDR 2 em soja), ao contrário dos siRNAs do arroz supracitados, contudo a maior abundância de siRNAs de orientação senso derivados do RNA 1 do CMV pode indicar que alguns desses siRNAs sejam formados a partir de estruturas secundárias de maneira similar aos siRNAs siR441 e siR446 de arroz. Outra possibilidade para explicar a diminuição dos siRNAs sob estresse apesar do aumento da expressão do RNA vem do próprio mecanismo de regulação da rota de RdRM. Para a síntese dos hc-siRNAs de transposons há a necessidade da atividade da RNA Pol IV transcrevendo a região aonde eles se encontram. A RNA Pol IV polimeriza uma fita simples de RNA a partir de regiões de DNA metilado e já foi relatado que a ativação de transposons pode estar associada a demetilação do DNA induzida por estresse (CHINNUSAMY & ZHU, 2009). Se um processo de demetilação do DNA similar estiver ocorrendo para a região ao qual está integrado o RNA 1 do CMV a transcrição de RNA via RNA Pol IV pode ser reprimida e portanto a subsequente geração de hc-siRNAs.

5.4. Expressão diferencial dos siRNAs e possíveis alvos.

A distribuição dos siRNAs ao longo da sequência do RNA 1 do CMV não é homogênea apresentando regiões aparentemente preferenciais para o processamento pelas DCLs (Figura 8 à Figura 13). Alguns autores em trabalhos de sequenciamento em larga escala de vsRNAs sugerem que essas regiões teriam uma proporção maior de nt guanina (G) e/ou citosina (C), indicando que as DCLs podem ter um maior processamento nesses *loci* para a geração de vsiRNAs (HO *et al.*, 2007, DONAIRE *et al.*, 2009). Dados preliminares comparando a quantidade de G/C em regiões da sequência do CMV à quantidade de vsiRNAs

formados também indicam uma relação similar, porém não para todas as regiões genômicas e inclusive muitos dos vsiRNAs mais expressos apresentam uma quantidade de G/C menor ou igual à da sequência do RNA 1 do CMV (45%) (dados não apresentados). Alguns desses siRNAs podem ser originados de estruturas secundárias de dupla fita do RNA transcrito (ex: em forma de grampo), isso explicaria a proporção maior de siRNAs de orientação senso encontrados nas nossas bibliotecas (Figura 6), entretanto como já discutido anteriormente, essa correlação nem sempre é verificada para vsiRNAs (LLAVE, 2010). Outros fatores podem contribuir para essa heterogeneidade da distribuição dos siRNAs ao longo da sequência do RNA 1 do CMV. Já foi demonstrado que características termodinâmicas dos siRNAs definem qual das fitas será carregada pela AGO (TOMARI *et al.*, 2004), provavelmente a fita presente no complexo RISC terá maior meia-vida na célula e maior chance de ser encontrada em bibliotecas de sequenciamento de sRNAs. Modificações químicas e estruturais nos siRNAs também podem levar a uma maior estabilidade dessas moléculas na célula, como a metilação por HEN 1 que parece ocorrer mais frequentemente no siRNA que é carregado pela AGO, tornando-o mais resistente a atividade de exonucleases, enquanto que o siRNA preterido recebe adição de uma a 5 uridinas (U) no seu terminal 5' o que facilita a sua degradação (LI *et al.*, 2005).

Analisando o perfil da distribuição dos siRNAs ao longo da sequência do RNA 1 do CMV, foi possível identificar determinados “picos” de geração de siRNAs que variavam de acordo com o tratamento aplicado (controle, estresse biótico ou abiótico) ou conforme a cultivar (resistente ou sensível ao estresse) em um mesmo tratamento (Figura 8 a 13). Com base nessas observações, testes estatísticos utilizando os dados de abundância de cada siRNA normalizados pelo total de siRNAs derivados do CMV sequenciados para cada biblioteca, foram realizados para identificar possíveis alterações na frequência dessas moléculas entre as bibliotecas. Alguns siRNAs para todas as comparações apresentaram variações em sua frequência que foram estatisticamente significativas, indicando que tanto o estresse hídrico quanto a infecção pelo fungo *P. pachyrhizi* podem estar de certa forma modulando a geração de siRNAs (Figura 14 e 16). Na comparação entre cultivares diferentes sob o mesmo tratamento também foi

possível observar siRNAs diferencialmente expressos (Figura 15 e 17). Contudo esses dados precisam ser analisados com cautela, já que o número amostral utilizado foi só de uma biblioteca por tratamento ou cultivar diferentes.

Uma outra característica importante dos vsRNAs, evidenciada recentemente, é a sua potencial capacidade de alterar a expressão de RNAs do hospedeiro em infecções virais, via PTGS. Foi possível obter o silenciamento gênico do gene fito-eno dessaturase (PDS, do inglês, “phytoene desaturase”) ao recombinar o seu cDNA em vetores virais híbridos com sequências do vírus do mosaico do tabaco e vírus do mosaico do tomate em *Nicotiana benthamiana* (KUMAGAI et al., 1995). Essa foi a primeira evidência de que vsRNAs podem realizar com sucesso o PTGS em plantas, essa mesma técnica foi aplicada com sucesso em mais de 30 outras espécies vegetais com o intuito de se estudar o funcionamento de genes (BECKER & LANGE, 2009). A partir dessas observações muitos trabalhos começaram a identificar possíveis alvos de vsRNAs utilizando ferramentas de bioinformática que levam em conta a termodinâmica e o pareamento ótimo necessário entre essas moléculas e os RNA endógenos das plantas para uma efetiva atividade de silenciamento gênico mediada por AGOs (LLAVE, 2010). Foram encontrados centenas de potenciais alvos para vsRNAs em diferentes espécies de plantas, porém evidências experimentais da clivagem do RNA ou da diminuição da expressão do gene alvo foram observadas em apenas alguns casos e mesmo assim sem uma comprovação da sua função durante a infecção viral (MOISSARD & VOINET, 2006, QI et al., 2009).

Com o intuito de verificar se alguns dos siRNAs derivados do CMV tem como alvos RNAs endógenos de soja análises de bioinformática foram realizadas e alguns resultados interessantes foram obtidos. Uma determinada região da sequência do CMV (do nt 739 ao 763) produziu 3 siRNAs de 22 nt em *tandem* (diferença de um nt para cada) com uma alta expressão entre as bibliotecas nomeados lib22-1040878, lib22-1252098 e lib22-921147, todos de orientação senso. Apesar da semelhança na sequência de nucleotídeos desses três siRNAs somente o lib22-125098 e o lib22-921147 apresentaram um provável alvo no genoma da soja, uma serina treonina quinase encontrada em bibliotecas de cDNA de folha de soja inoculada por esporos do fungo *P. pachyrhizi*, o número de

acesso no banco de dados do NCBI é FJ014790.1. Curiosamente na descrição desse RNA no banco de dados do NCBI os autores classificam a proteína codificada por essa mRNA de proteína de resistência à ferrugem asiática (doença causada pelo *P. pachyrhizi* em soja). Essa proteína possui uma maior homologia na sequência de aminoácidos à proteína quinase 1 de *Arabidopsis thaliana* (APK1, do inglês, “*Arabidopsis thaliana* protein kinase 1”) (dados não apresentados) que fosforila resíduos de serina e treonina de outras proteínas e de si mesma, porém os alvos da atividade dessa enzima *in vivo* ainda não foram identificados (HIRAYAMA & OKA, 1992). Apesar de diferenças significativas na expressão desses siRNAs não terem sido encontradas para todas as comparações testadas uma análise dos dados normalizados apresenta uma interessante tendência. Na cultivar “PI 561356” (resistente à infecção) todos os três siRNAs citados apresentam diminuição de sua expressão quando inoculados pelo fungo, somente os dados para o lib22-125 2098 não foram significativos, apesar de apresentarem uma diminuição maior de 3 vezes quando sob estresse (dados não apresentados). O contrário ocorreu na cultivar suscetível à infecção (“Embrapa 48”), sob estresse ocorreu um aumento da expressão de dois dos 3 siRNAs (somente significativo para lib22-1040878). Exatamente o mesmo padrão para os 3 siRNAs ocorreu também nas bibliotecas de estresse hídrico, porém nenhuma alteração estatisticamente significativa. O que parece ocorrer é que as cultivares resistentes já apresentam uma expressão maior desses siRNAs quando no tratamento controle e eles diminuem ao serem submetidas ao estresse, tanto biótico quanto abiótico, e o contrário ocorre para as cultivares sensíveis. Entretanto o fato de nem todas essas diferenças serem estatisticamente significativas sugere que novos testes deverão ser feitos com um número amostral maior de plantas sequenciadas para cada cultivar e tratamento para a confirmação dessa tendência. E por fim somente foi encontrada a fita senso desses 3 siRNAs em todas as bibliotecas. A fita anti-senso que deveria ser produzida em quantidade igual considerando o processamento das DCLs em dsRNAs não foi encontrada em nenhuma das bibliotecas, isso indica que os siRNAs senso estão sendo carregados preferencialmente ao RISC e portanto adquirindo uma maior estabilidade na célula. Isso aumenta as chances de esses

siRNAs estarem realmente agindo sobre o seu mRNA alvo.

5.5. Possíveis benefícios da integração do CMV ao genoma da soja.

De acordo com outros estudos sobre integração de vírus ao genoma de eucariotos essas sequências podem estar agindo de maneira importante no metabolismo celular, assumindo o papel de genes funcionais (KOONIN, 2010). Em fungos transcritos de NIRVs que codificam a proteína CP de *Totivirus* são detectados por RT-qPCR na célula. Já foi demonstrado que a superexpressão de genes CP em plantas promove resistência perante uma infecção viral, seja em nível de transcrição ou em nível de proteína, com as CPs endógenas interferindo na ação das proteínas virais (REIMANN-PHILIPP, 1998). Reforçando essa tese os CPs do NIRV de *Saccharomyces cerevisiae* possuem alterações de aminoácidos que impedem a atividade de remoção do cap 5' de mRNA endógenos, essa atividade é importante para a infecção do *Totivirus*, indicando a adaptação desse novo gene a uma possível nova função na célula da levedura. Em *Debaryomyces hansenii* um NIRV que codifica para uma RdRp mantém conservados todos os resíduos catalíticos dessa proteína, sugerindo que ela possa ter atividade enzimática (KOONIN, 2010). Essa RdRp pode servir de maneira análoga às RDRs amplificando transcritos virais de modo a gerar dsRNAs que possam promover uma resistência antiviral pelo mecanismo de RNAi. Não está descartada, porém, a possibilidade dos produtos dos NIRVs terem outras funções dentro da célula que não as de atividade antiviral, contudo devido ao grande número de NIRVs diferentes e inclusive outros elementos de diferentes vírus que parecem ter se integrado ao DNA eucarioto por processos similares em diferentes organismos a hipótese de essas sequências promoverem resistência frente a uma infecção viral parece ser a mais provável (KOONIN, 2010).

Colaborando com a ideia de que NIRVs podem servir como elementos importantes na resistência à infecções virais um estudo demonstrou que a expressão de um transgene com a sequência do RNA 1 do CMV em tabaco reduz a acumulação do vírus nas folhas das plantas inoculadas e impede a infecção

sistêmica (CANTO & PAKULATIS, 2001). Entretanto baseado nos dados obtidos os autores descartam a ativação da via de PTGS como a responsável por essa resposta antiviral. Basicamente duas observações suportam essa hipótese. A primeira é de que o nível do mRNA transgênico se mantém o mesmo enquanto o RNA 1 viral é suprimido nas plantas que apresentam a resistência a infecção. E a segunda é que em uma infecção nessas plantas transgênicas pelo vírus do chocolate do fumo (TRV, do inglês, "Tobacco Rattle Virus"), que teve o seu RNA 1 recombinado de maneira a ter parte da sequência do RNA 1 do CMV não é verificada a supressão dos RNAs virais do TRV. O mecanismo de resistência parece ocorrer devido a um bloqueio da replicação do RNA 1 ou de uma interferência na formação do complexo contendo a replicase viral e o RNA 1 (CANTO & PAKULATIS, 2001).

Entretanto, como no presente trabalho foram encontrados siRNAs derivados do RNA 1 do CMV, possivelmente as linhagens de soja que possuem a integração podem ser resistentes à uma infecção pelo CMV por meio da rota de PTGS. Alternativamente a expressão do RNA 1 do CMV encontradas nas plantas utilizadas nesse trabalho (Figura 18B e 19B) também pode interferir na replicação do RNA 1 viral de maneira similar ao que ocorre no tabaco transgênico. Curiosamente a expressão do RNA 1 do CMV aumenta quando submetida aos estresses biótico e abiótico testados (Figura 18B e 19B) o que facilitaria a ação desse último mecanismo (bloqueio da replicação viral) considerando que a infecção pelo CMV produza o mesmo efeito.

6. CONCLUSÕES E PERSPECTIVAS

Neste trabalho foram descritas evidências para o que pode ser o primeiro vírus de RNA não retroviral integrado ao genoma (NIRV) de plantas. Em 17 bibliotecas de sRNAs de diferentes cultivares de soja foram encontrados siRNAs derivados de uma sequência com alta homologia ao RNA 1 do CMV. Reações de PCR confirmaram a presença do DNA referente a essa sequência do CMV em 12 cultivares e foi possível detectar a presença do RNA em diferentes tecidos de três dessas cultivares por RT-qPCR. Porém não foi encontrado essa sequência no DNA da cultivar “Willians”, desenvolvida nos Estados Unidos. Esse dado pode ser um indício de que essa integração foi um evento muito recente, possivelmente o mais recente que já tenha sido relatado para um NIRV. Análises em cultivares ancestrais as encontradas nas Américas, provenientes da Ásia, precisam ser feitas para uma datação precisa dessa integração. A hipótese mais provável para a integração desse RNA ao genoma da soja é a de uma recombinação com um retrotransposon, devido ao perfil dos siRNAs derivados do CMV, principalmente quando comparados aos vsRNAs do BPMV também encontrado em uma das bibliotecas analisadas. A literatura também aponta essa como a principal causa para a geração dos NIRVs. Entretanto somente com o sequenciamento das regiões que flanqueiam o RNA 1 será possível identificar se ocorreu a recombinação com um retrotransposon ou se outro evento foi o responsável pela integração. A expressão do RNA 1 do CMV parece aumentar em resposta ao estresse biótico (infecção pelo fungo *P. pachyrhizi*) e abiótico (estresse hídrico) e em contrapartida os siRNAs derivados do CMV tem uma tendência a diminuir a sua expressão quando expostos aos mesmos tratamentos. Foram encontrados siRNAs diferencialmente expressos entre as diferentes bibliotecas de sRNAs analisadas. Dois desses siRNAs tem como alvo potencial um gene de resistência a doença da ferrugem asiática induzida pelo fungo *P. pachyrhizi* em soja, indicando que siRNAs derivados do CMV podem realizar um importante papel na resposta à estresses bióticos. Entretanto esses dados precisam ser confirmados em experimentos com um número amostral maior de plantas sequenciadas. Por fim esse provável NIRV de soja pode realizar uma importante função de defesa

antiviral seja pela via de PTGS através dos siRNAs derivados do CMV ou por uma interferência direta na replicação do vírus pela molécula de RNA 1 do CMV expressa no genoma dessas plantas. Contudo a confirmação dessa atividade só pode ser realizada após experimentos de infecção pelo CMV em cultivares de soja que não possuem esse NIRV (como a “Willians”) e cultivares em que ele é encontrado.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABIOVE Complexo Soja – Esportações. Disponível em: <http://www.abiove.com.br/exporta_br.html> Acesso em: 7 de Dezembro de 2010.

ALLEN, E., XIE, Z., GUSTAFSON, A. M. & CARRINGTON, J. C. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell*, 121 (2): 207-221, 2005.

ANDERSSON, J. O., DOOLITTLE, W. F. & NESBO, C. L. *Science*, 292: 1848–1850, 2001.

AXTELL, M. J., JAN, C., RAJAGOPALAN, R. & BARTEL, D. P. A two-hit trigger for siRNA biogenesis in plants. *Cell*, 127: 565–577, 2006.

BECKER, A. & LANGE, M. VIGS – genomics goes functional. *Trends in Plant Science*, 15 (1): 1-4, 2009.

BEJARANO, E. R., KHASHOGGI, A., WITTY, M. & LICHTENSTEI, 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: 759–764, 1996.

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*, 10: 651–655, 2000

BORSANI, O. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*, 123: 1279–1291, 2005.

CANTO, T. & PAKULATIS, 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.

COSTA, J.A. Cultura da Soja. Porto Alegre, 233 p, 1996.

CHEN, C., RIDZON, D. A., BROOMER, A. J., ZHOU, Z., LEE, D. H., NGUYEN, J. T., BARBASIN, M., XU, N. L., MAHUVAKAR, V. R. & ANDERSEN, M. R. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33 (20): e179, 2005.

CHINNUSAMY, V. & ZHU, J-K. RNA-directed DNA methylation and demethylation in plants. *Science in China Series C: Life Sciences*, 52 (4): 331-343, 2009.

CROCHU, S., COOK, S., ATTOUI, H., CHARREL, R. N., DE CHESSE, R.,

BELHOUCHE, M., LEMASSON, J. J., DE MICCO, P. & DE LAMBALLERIE, X. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes spp.* mosquitoes. *Journal of General Virology*, 85: 1971-1980, 2004.

DANIELS, S. B., PETERSON, K. R., STRAUSBAUGH, L. D., KIDWELL, M. G. & CHOVNICK, A. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics*, 124: 339-355, 1990.

DE KONING, A. P., BRINKMAN, F. S. L., JONES, S. J. M. & KEELING, P. J. *Molecular Biology and Evolution*, 17: 1769-1773, 2000.

DELERIS, A., GALLEGRO-BARTLOME, J., BAO, J., KASSCHAU, K. D., CARRINGTON J. C. & VOINNET, O. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science*, 313: 68-71, 2006.

DEFARGES, S. & CIUFFI, A. Retroviral integration site selection. *Viruses*, 2: 111-130, 2010.

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.

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: 203-214, 2009.

DREHER, T. W. Role of tRNA-like structures in controlling plant virus replication. *Virus Research*, 139 (2): 217-229, 2009.

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 cucumovirus *Cucumber mosaic virus* present in Brazil. *Journal of Plant Pathology*, 86 (1): 61-69, 2004.

EMBRAPA Usos. Disponível em: <http://www.cnpso.embrapa.br/index.php?op_page=25&cod_pai=29> Acesso em : 7 Dezembro de 2010d.

EMBRAPA Composição da soja. Disponível em: <http://www.cnpso.embrapa.br/soja_alimentacao/index.php?pagina=7> Acesso em: 7 Dezembro de 2010b.

EMBRAPA História. Disponível em: <http://www.cnpso.embrapa.br/index.php?op_page=112&cod_pai=33> Acesso em: 7 Dezembro de 2010a.

EMBRAPA Soja uso industrial. Disponível em: <http://www.cnpso.embrapa.br/index.php?op_page=27&cod_pai=31> Acesso em:

7 Dezembro de 2010c.

EMBRAPA Soja em números (safra 2008/2009). Disponível em : <http://www.cnpso.embrapa.br/index.php?op_page=294&cod_pai=17> Acesso em : 7 Dezembro de 2010e.

ESNAULT, C., MAESTRE, J. & HEIDMANN, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics*, 24: 363–367, 2000.

EWING, R. M., CHU, P., ELISAM, F., LI, H., TAYLOR, P., CLIMIE, S., MCBROOM-CERAJEWSKI, L., ROBINSON, M. D., O'CONNOR, L., LI, M., TAYLOR, R., DHARSEE, M., HO, Y., HEILBUT, A., MOORE, L., ZHANG, S., ORNATSKY, O., BUKHMAN, Y. V., ETHIER, M., SHENG, Y., VASILESCU, J., ABU-FARHA, M., LAMBERT, J. P., DUEWEL, H. S., STEWART, I. I., KUEHL, B., HOGUE, K., COLWILL, K., GLADWISH, K., MUSKAT, B., KINACH, R., ADAMS, S. L., MORAN, M. F., MORIN, G. B., TOPALOGLOU, T. & FIGEYS, D. Large-scale mapping of human protein–protein interactions by mass spectrometry. *Molecular Systems Biology*, 3 (89): 1-17, 2007.

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: 806–811, 1998.

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: 481–496, 2010.

GASCIOLLI, V. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing transacting siRNAs. *Current Biology*, 15: 1494–1500, 2005.

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: 393-396, 2009.

GRANDBASTIEN, M-A. Activation of plant retrotransposons under stress conditions. *Trends in Plant Science*, 3 (5): 181-187, 1998.

GUNSTONE, F.D. Vegetable oils. In: SHAHIDI, F. *Bailey's industrial oil and fat products: edible oils*. 6. ed. New York: Wiley-Interscience, 2005. v.1.

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

HEUER, H. & SMALLA, K. Horizontal gene transfer between bacteria.

Environmental Biosafety Research, 6: 3–13, 2007.

HIRAYAMA, T. & OKA, A. Novel protein kinase of *Arabidopsis thaliana* (APK1) that phosphorylates tyrosine, serine and threonine. *Plant Molecular Biology*, 20(4): 653-662. 1992.

HO, T., WANG, H., PALLETT, D. & DALMAY, T. Evidence for targeting common siRNA hotspots and GC preference by plant Dicer-like proteins. *FEBS Letters*, 581: 3267–3272, 2007.

HOTOPP, J. C. D., CLARK, M. E., OLIVEIRA, D. C. S. G., FOSTER, J. M., FISCHER, P., TORRES, M. C. M., GIEBEL, J. D., KUMAR, N., ISHMAEL, N., WANG, N., 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*, 317 (21): 1753-1756, 2007.

HORIE, M., HONDA, T., SUZUKI, Y., KOBAYASHI, 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: 84-88, 2010.

HYMOWITZ, T. On the domestication of soybean. *Economic Botany*, 24: 408–421, 1970.

HYMOWITZ, T. & HARLAN, J.R. Introduction of soybean to North America by Samuel Bowen in 1765. *Economic Botany*, 37: 371–379, 1983.

KATIYAR-AGARWAL, S., MORGAN, R., DAHLBECK, D., BORSANI, O., VILLEGAS, A., ZHU, J. K., STASKAWICZ, B. J. & JIN, H. L. A pathogen-inducible endogenous siRNA in plant immunity. *Proceedings of the National Academy of Sciences of the United States of America* 103(47): 18002–18007, 2006.

KLENERMAN, P., HENGARTNER, H. & ZINKERNAGEL, R. M. A non-retroviral RNA virus persists in DNA form. *Nature*, 390: 298-301, 1997.

KLEVEBRING, D., STREET, N. R., FAHLGREN, N., KASSCHAU, K. D., CARRINGTON, J. C., LUNDEBERG, J. & JANSSON, S. Genome-wide profiling of *Populus* small RNAs. *BMC Genomics*. 10 (620): 1-18, 2009.

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. Taming of the shrewd: novel eukaryotic genes from RNA viruses. *BMC Biology*, 8: 1-4, 2010.

KOSHIMIZU, Y. & IIZUKA, N. Soybean stunt disease. *Annals of the Phytopathological Society of Japan*, 23: 27, 1958.

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: 1–7, 2009.

KULCHESKI, F. R., MARCELINO-GUIMARRÃES, F. C., NEPOMUCENO, A. L., ABDELNOOR, R. V. & MARGIS, R. The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Analytical Biochemistry*, 406 (2): 185-192, 2010.

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 of the United States of America*, 92: 1679–1683, 1995.

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., SUN, S. S. & ZHANG, G. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*, 42: 1053-1059, 2010.

LEAL, J.C. Plantas da Lavoura Sul Rio-grandense. Porto Alegre, 274 p, 1967.

LI, R., YU, C., LI, Y., LAM, T. W., YIU, S. M., KRISTIENSEN, K. & WANG, J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25 (15): 1966-1967, 2009.

LI, J. YANG, Z., YU, B., LIU, J. & CHEN, X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Current Bioogy*. 15: 1501–1507, 2005.

LIPPMAN, Z. & MARTIENSSEN, R. The role of RNA interference in heterochromatic silencing. *Nature*, 431: 364–370, 2004.

LIVAK, K. J. & SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25 (4): 402-408, 2001.

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

LU, C., TEJ, S. S., LUO, S., HAUDENSCHILD, C. D., MEYERS, B. C. & GREEN, P. J. Elucidation of the small RNA component of the transcriptome. *Science* 309: 1567–1569, 2005.

MARCO, A. & MARÍN, I. Retrovirus-like elements in plants. *Recent Research*

Developments in Plant Science, 3: 15-24, 2005.

MAUCK, K. E., DE MORAES, C. M. & MESCHER, M. C. Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8): 3600-3605, 2010.

MEYERS, B. C., AXTELL, M. J., BARTEL, B., BARTEL, D. P., BAULCOMBE, D., BOWMAN, J. L., CAO, X., CARRINGTON, J. C., CHEN, X., GREEN, P. J., GRIFFITHS-JONES, S., JACOBSEN, S. E., MALLORY, A. C., MARTIENSSEN, R. A., POETHIG, R. S., QI, Y., VAUCHERET, H., VOINNET, O., WATANABE, Y., WEIGEL, D. & ZHU, J. K. Criteria for annotation of plant microRNAs. *Plant Cell*, 20: 3186–3190, 2008.

MILLAR, A. A. & WATERHOUSE, P. M. Plant and Animal microRNAs: similarities and differences. *Functional and 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: 19593–19598, 2006.

MOLNAR, A., CSORBA, T., LAKATOS, L., VARALLYAY, E., LACOMME, C. & BURGYÁN, J. Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. *Journal of Virology*, 79: 7812–7818, 2005.

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

MOXON, S., SCHWACH, F., DALMAY, T., MACLEAN, D., STUDHOLME, D. J. & MOULTON, V. A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics*, 24 (19): 2252-2253, 2008.

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: 279-289, 1990.

NOBUTA, K., LU, C., SHRIVASTAVA, R., PILLAY, M., DE PAOLI, E., ACCERBI, M., ARTEAGA-VAZQUEZ, M., SIDORENKO, L., JEONG, D. H., YEN, Y., GREEN, P. J., CHANDLER, V. L. & MEYERS, B. C. Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (39): 14958–14963, 2008.

JAMALKANDI, S. A. & MASOUDI-NEJAD, A. Reconstruction of *Arabidopsis thaliana* fully integrated small RNA pathway. *Functional and Integrative Genomics*, 9: 419–432, 2009.

PALUKAITIS, P., ROOSSINCK, M. J., DIETZGEN, R. G., FRANCKI, R. I. B. *Cucumber mosaic virus. Advances in virus Research* 41: 281-341, 1992.

PERAGINE, A. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes & Development*, 18: 2368–2379, 2004.

PONTES, O., LI, C. F., NUNES, P. C., HAAG, J., REAM, T., VITINS, A., JACOBSEN, S. E. & PIKAARD, C. S. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, 126: 79–92, 2006.

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-e4981, 2009.

REIMANN-PHILIPP, U. Mechanisms of resistance. Expression of coat protein. *Methods in Molecular Biology*, 81: 521-532, 1998.

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.

ROZEN, S. & SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, 132: 365-386, 2000.

SCHMUTZ, J., CANNON, S. B., SCHLUETER, J., MA, J., MITROS, T., NELSON, W., HYTEN, D. L., SONG, Q, THELEN, J. J., CHENG, J., XU, D., HELLSTEN, U., MAY, G. D., YU, Y., SAKURAI, T., UMEZAWA, T., BHATTACHARYYA, M. K., SANDHU, D., VALLIYODAN, B., LINDQUIST, E., PETO, M., GRANT, D., SHU, S., GOODSTEIN, D., BARRY, K., FUTRELL-GRIGGS, M., ABERNATHY, B., DU, J., TIAN, Z., ZHU, L., GILL, N., JOSHI, T., LIBAULT, M., SETHURAMAN, A., ZHANG, X. C., SHINOZAKI, K., NGUYEN, H. T., WING, R. A., CREGAN, P., SPECHT, J., GRIMWOOD, J., ROKHSAR, D., STACEY, G., SHOEMAKER, R. C. & JACKSON, S. A. Genome sequence of the palaeopolyploid soybean. *Nature*, 463: 178-183, 2010.

SCHWAB, R., PALATNIK, J. F., RIESTER, M., SCHOMMER, C., SCHMID, M. & WEIGEL, D. Specific effects of microRNAs on the plant transcriptome. *Developmental Cell*, 8: 517-527, 2005.

SCHWARZ, D. S., HUTVAGNER, G., DU, T., XU, Z., ARONIN, N. & ZAMORE, P. D. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115: 199–208, 2003.

SIMON, A. E., ROOSINCK, M. J. & HAVELDA, Z. Plant virus satellite and

defective interfering RNAs: New Paradigms for a New Century. *Annual Review of Phytopathology*, 42: 415–437, 2004.

SOMERS, D. E., SCHULTZ, T. F., MILNAMOW, M. & KAY, S. A. ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell*, 101: 319-329, 2000.

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. *Bulletin of Tohoku National Agricultural Experiment Station*, 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. *Miscellaneous Publications of the Tohoku National Agricultural Experiment Station*, 7: 1–35, 1987.

TAYLOR, D. J. & BRUENN, J. The evolution of novel fungal genes from nonretroviral RNA viruses. *BMC Biology*, 7 (88): 1-11, 2009.

TAYLOR, J. D., LEACH, R. W. & BRUENN, J. Filoviruses are ancient and integrated into mammalian genomes. *BMC Evolutionary Biology*, 10: 193-202, 2010.

TOMARI, Y. MATRANGA, C., HALEY, B., MARTINEZ, N., ZAMORE, P. D. A protein sensor for siRNA asymmetry. *Science*, 306: 1377–1380, 2004.

TOMONAGA, K., KOBAYASHI, T. & IKUTA, K. Molecular and cellular biology of Borna disease virus infection. *Microbes and Infection*. 4: 491–500, 2002.

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

VAZQUEZ, F., VAUCHERET, H., RAJAGOPALAN, R., LEPERS, C., GASCIOLLI, V., MALLORY, A. C., HILBERT, J-L, BARTEL, D. P. & CRÉTÉ, P. (2004) Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Molecular Cell*, 16: 69–79, 2004.

WANG, L., FENG, Z., WANG, X. & ZHANG, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26 (1): 136-138, 2010.

WANG, X. B., WU, Q., ITO, T., GILLO, F., LI, W. X., CHEN, X., YU, J.L. & DING, S. W. RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 1(197): 484-489, 2010.

XIE, Z., ALLEN, E., WILKEN, A. & CARRINGTON, J. C. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 12984–12989, 2005.

XIE, Z. & QI, X. Diverse small RNA-directed silencing pathways in plants. *Biochimica et Biophysica Acta*, 1779: 720-724, 2008.

YAN, Y., ZHANG, Y., YANG, K., SUN, Z., FU, Y., CHEN, X. & FANG, R. Small RNAs from MITE-derived stem-loop precursors regulate abscisic acid signaling and abiotic stress responses in rice. *The Plant Journal*, In Press.

YU, B., YANG, Z., LI, J., MINAKHINA, S., YANG, M., PADGETT, R. W., STEWARD, R. & CHEN, X. Methylation as a crucial step in plant microRNA biogenesis. *Science* 307: 932–935, 2005.

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

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, 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: 3255–3268, 2006

ZHENG, X., ZHU, J., KAPOOR, A. & ZHU, J. K. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO Journal*, 26: 1691–1701, 2007.

ZITTER, T. A. & MURPHY, J. F. Cucumber mosaic. Disponível em: <<http://www.apsnet.org/edcenter/intropp/lessons/viruses/Pages/Cucumbermosaic.aspx>> Acesso em: 8 Dezembro de 2010.

ZUKER, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31 (13): 3406-3415, 2003.

CURRICULUM VITÆ

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: cordenonsi85@hotmail.com

2. FORMAÇÃO: Bacharelado em Ciências Biológicas (Universidade Federal do Rio Grande do Sul - UFRGS) – 2003 a 2008

3. ESTÁGIOS

Bolsa de Iniciação Científica – Orientador: Rogerio Margis – Resumo: Uso da metodologia RNAi para a produção de plantas resistentes à insetos. - BIC/PROPEQ – 01/04/2006 à 31/12/2006.

Bolsa de Iniciação Científica – 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 à 31/07/2008

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

4. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

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.

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.

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.

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.