



REDE NORDESTE DE BIOTECNOLOGIA  
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO DE CIÊNCIAS DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**LEIDY JOHANA MADROÑERO**

**TRANSCRIPTOMA DIFERENCIAL E MECANISMOS  
MOLECULARES ASSOCIADOS AO  
DESENVOLVIMENTO DOS SINTOMAS DA MELEIRA  
DO MAMOEIRO**

VITÓRIA

2018

**LEIDY JOHANA MADROÑERO**

**TRANSCRIPTOMA DIFERENCIAL E MECANISMOS  
MOLECULARES ASSOCIADOS AO DESENVOLVIMENTO  
DOS SINTOMAS DA MELEIRA DO MAMOEIRO**

Tese de doutorado apresentada ao Programa de Pós-graduação em Biotecnologia da Rede nordeste de Biotecnologia (RENORBIO) do ponto focal Universidade Federal do Espírito Santo (UFES), como parte dos requisitos necessários à obtenção do título de Doutor em Biotecnologia.

Orientadores: Profa Dr.a Patricia Machado Bueno Fernandes e Prof. Dr. Silas Pessini Rodrigues

**VITÓRIA-ES**

**2018**

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Setorial do Centro de Ciências da Saúde da Universidade Federal do  
Espírito Santo, ES, Brasil)

---

M178t Madroñero, Leidy Johana, 1988 -  
Transcriptoma diferencial e mecanismos moleculares associados ao  
desenvolvimento dos sintomas da meleira do mamoeiro / Leidy Johana  
Madroñero - 2018.  
103 f. : il.

Orientador: Patricia Machado Bueno Fernandes.  
Coorientador: Silas Pessini Rodrigues.

Tese (Doutorado em Biotecnologia) – Universidade Federal do Espírito  
Santo, Centro de Ciências da Saúde.

1. Carica. 2. Transcriptoma. 3. Ácido Salicílico. 4. Biotecnologia.  
I. Fernandes, Patricia Machado Bueno. II. Rodrigues, Silas Pessini.  
III. Universidade Federal do Espírito Santo. Centro de Ciências da Saúde.  
IV. Título.

CDU: 61

---

LEIDY JOHANA MADROÑERO

**TRANSCRIPTOMA DIFERENCIAL E MECANISMOS  
MOLECULARES ASSOCIADOS AO DESENVOLVIMENTO DOS  
SINTOMAS DA MELEIRA DO MAMOEIRO**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia da Rede Nordeste de Biotecnologia (RENORBIO) e Universidade Federal do Espírito Santo (UFES), como requisito para obtenção do título de Doutor em Biotecnologia.

**Banca examinadora:**

Prof<sup>a</sup>. Dr<sup>a</sup>. Patricia Machado Bueno Fernandes  
Universidade Federal do Espírito Santo  
Orientadora

Prof. Dr. Silas Pessini Rodrigues  
Universidade Federal do Rio de Janeiro  
Coorientador

Prof. Dr. Alexandre Martins Costa Santos  
Universidade Federal do Espírito Santo  
Examinador

Prof. Dr. Francisco Murilo Zerbini Júnior,  
Universidade Federal de Viçosa  
Examinador

Prof. Dr José Aires Ventura  
Instituto Capixaba de Pesquisa, Assistência Técnica e  
Extensão Rural / Universidade Federal do Espírito Santo  
Examinador

Prof. Dr. Teodiano Freire Bastos Filho  
Universidade Federal do Espírito Santo  
Examinador

VITÓRIA-ES

2018

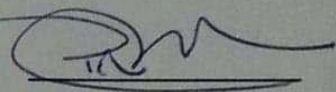
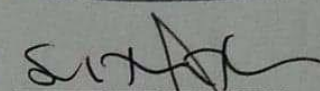
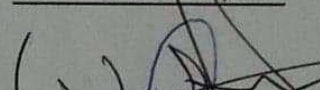
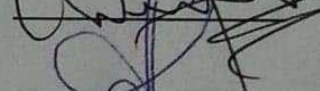
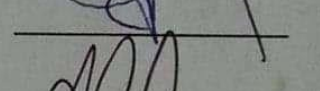
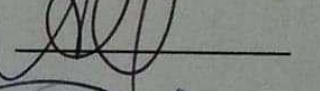
**Programa de Pós-Graduação em Biotecnologia**  
Ponto Espírito Santo - Universidade Federal do Espírito Santo

**DEFESA DE TESE**

**ALUNO(A):** LEIDY JOHANA MADROÑERO

**TÍTULO DO PROJETO:** "TRANSCRIPTOMA DIFERENCIAL E MECANISMOS MOLECULARES ASSOCIADOS AO DESENVOLVIMENTO DOS SINTOMAS DA MELEIRA DO MAMOEIRO"

**PROFESSOR(A) ORIENTADOR(A):** Patricia Machado Bueno Fernandes

<b>BANCA EXAMINADORA:</b>	<b>CONCEITO</b>	<b>ASSINATURA</b>
Prof. <sup>a</sup> . Dr. <sup>a</sup> . Patricia Machado Bueno Fernandes RENORBIO/UFES (Orientadora)	<u>Satisfatório</u>	
Prof. Dr. Silas Pessini Rodrigues UFRJ (Coorientador)	<u>satisfatório</u>	
Prof. Dr. José Aires Ventura RENORBIO/UFES (Titular)	<u>satisfatório</u>	
Prof. Dr. Teodiano Freire Bastos Filho RENORBIO/UFES (Titular)	<u>satisfatório</u>	
Prof. Dr. Alexandre Martins Costa Santos UFES (Titular)	<u>Satisfatório</u>	
Prof. Dr. Francisco Murilo Zerbini Júnior UFV (Titular - Participação Remota) 19	<u>Satisfatório</u>	

**TIPO DE SESSÃO:** Aberta

**DIA:** 06 de julho de 2018

**HORÁRIO:** 09:30 horas

**LOCAL:** Universidade Federal do Espírito Santo – Auditório do Núcleo de Biotecnologia – 2º Andar.

## DEDICATÓRIA

*A minha Família e os meus  
amigos*

## **AGRADECIMENTOS**

À Universidade Federal do Espírito Santo, ao Programa de Pós-Graduação em Biotecnologia da UFES e à Renorbio por me propiciar os espaços e as ferramentas necessárias para a execução do meu projeto de Doutorado.

À agência de apoio CAPES pelo financiamento de minha bolsa de estudos.

Às agências de apoio e financiamento: CNPq, FAPES, CAPES e FINEP pelo financiamento de reagentes, materiais e serviços requeridos para o desenvolvimento do projeto.

Aos professores José Aires, Antônio Alberto, por toda orientação e oportunidades de aprendizado.

Aos professores Patrícia Fernandes e Silas Rodrigues pelo constante acompanhamento e apoio.

A Miriam Cristine de Araújo Assis, Secretária do Programa de Pós-graduação em Biotecnologia da RENORBIO e da UFES, pela sua disposição e suporte nas questões administrativas.

Aos colegas de trabalho do Laboratório de Biotecnologia Aplicada ao Agronegócio. Especialmente a Tathiana, Eduardo e Paolla, que foram fundamentais para o desenvolvimento desta Tese.

A minha família, a Gustavo e os meus amigos: Juancho, Sebastián, Jessica e João, Lais, Karla, Joss, María, Vale, Oscar, Victor, Mile, Mafe, Jenny, Arelis, Alex, Eddie, Eduardo. Quem tem bons amigos tem tudo. Eu sou grata por ter vocês na minha vida.

“We no longer need to fear arguments, confrontations or any kind of problems with ourselves or others. Even stars collide, and out of their crashing new worlds are born. Today I know: THAT IS LIFE”

— Charlie Chaplin



## RESUMO

A meleira é uma doença que afeta gravemente a cultura do mamoeiro no Brasil e no México, e que está associada à infecção combinada do Papaya meleira vírus (PMeV) e Papaya meleira virus 2 (PMeV2) (complexo PMeV). Os sintomas da doença manifestam-se apenas após a floração. Para compreender os mecanismos envolvidos neste fenômeno, o transcriptoma diferencial de *C. papaya* inoculada com o complexo PMeV foi analisado e resultou na modulação de 633 e 88 genes na pré- e pós-floração, respectivamente. Na prefloração, a análise funcional mostrou, principalmente, que enquanto genes relacionados à defesa e transporte são induzidos, os genes relacionados à estruturação da parede celular, receptores quinase ricos em leucina (RLK-LRR) e ciclo celular são reprimidos. Em relação aos genes relacionados às vias de defesa, foi observada a indução de vários genes envolvidos no metabolismo de calose, detoxificação das espécies reativas de oxigênio (ROS) e, genes responsivos ao ácido salicílico (AS), tais como *PR1*, *PR2*, *PR5* e *WRKY*. Estes resultados indicaram o envolvimento da sinalização mediada pelo AS na tolerância de *C. papaya* ao desencadeamento dos sintomas durante a prefloração. Portanto, o papel do AS na resistência ao complexo PMeV foi avaliada mediante a aplicação exógena de AS em plantas inoculadas com o complexo PMeV, resultando numa tendência à diminuição no acúmulo do PMeV e o PMeV2 nas plantas tratadas com AS. Entretanto, durante a prefloração, as plantas também acumularam diferencialmente transcritos que codificam proteínas envolvidas no metabolismo de etileno, UDP-glicosiltransferases (UGTs) e, a proteína inibidora da proteína *Non-Expresser of Pathogenesis Related Gene 1* (NPR1) NPR1-I/NIM1, componentes que possuem um papel antagônico na indução de genes de defesa mediada pelo AS. Adicionalmente, a diminuição na sinalização de AS parece ser acentuada durante a pós-floração, já que foi observada a repressão de *PR1* e a indução do gene *BSMT1* e de genes envolvidos no metabolismo de ácido jasmônico (AJ), que também são reguladores negativos na acumulação de AS. Estes resultados em conjunto sugerem que vias de defesa mediadas por AS funcionam nas plantas infectadas pelo complexo PMeV durante a prefloração, e poderiam retardar o desenvolvimento dos sintomas, porém, a indução dos seus reguladores negativos prejudica a ativação total e duradoura da resposta de defesa.

## Palavras chave

Papaya meleira virus; *C. papaya*; RNA-Seq; WRKY70; NPR1-interacting protein; interação planta-virus; respostas de defesa; Ácido salicílico.

## ABSTRACT

The papaya sticky disease (PSD) severely affects the papaya crops in Brazil and México. PSD is associated with the combined infection of Papaya meleira virus (PMeV) and Papaya meleira virus 2 (PMeV2) (PMeV complex). Interestingly, PSD symptoms appear only after flowering. To understand the mechanisms involved in this phenomenon, the differential transcriptome of *C. papaya* inoculated with the PMeV complex was analyzed and resulted in the modulation of 633 and 88 genes in the pre- and post-flowering stages, respectively. At pre-flowering, functional analysis showed an up-regulation of genes related to defense and transport in parallel to the down-regulation of several genes coding to cell wall, leucine rich kinase receptors (RLK-LRR) and cell cycle proteins. Regarding the genes related to the defense pathways, it was observed the up-regulation of several genes involved in callose metabolism, detoxification of reactive oxygen species (ROS) and salicylic acid (SA) responsive genes, such as *PR1*, *PR2*, *PR5* and *WRKY*. These results suggest the involvement of SA-mediated signaling in the tolerance of *C. papaya* to the symptom development at pre-flowering. Hence, the role of SA in the resistance to the PMeV complex was evaluated by the exogenous application of SA in PMeV complex inoculated plants and resulted in a tendency to decrease the viral loads of PMeV and PMeV2 in the SA-treated plants. However, at pre-flowering, the plants also accumulated transcripts encoding proteins involved in the metabolism of ethylene, UDP-glycosyltransferases (UGTs), and the protein-inhibitor of the Non-Expresser of Pathogenesis Related Gene 1 (NPR1) NPR1-I / NIM1 protein, whose components play an antagonistic role in the induction of SA-mediated defense genes. In addition, the decrease in SA-signaling appears to be accentuated during post-flowering, since it was observed the down-regulation of *PR1* repression and the up-regulation *BSMT1* and jasmonic acid (AJ) metabolism genes, which are also negative regulators in accumulation of SA. These results together suggest that SA-mediated defense pathways are likely involved in the delayed symptoms at pre-flowering, but the induction of their negative regulators impairs the full and long-lasting activation.

### Keywords

Papaya meleira virus; *C. papaya*; RNA-Seq; WRKY70; NPR1-interacting protein; plant-virus interaction; defence responses; salicylic acid.

## LISTA DE FIGURAS

<b>Figura 1.</b> Flores e frutos relacionados aos três sexos morfológicamente distintos encontrados no mamoeiro.....	20
<b>Figura 2.</b> Evolução da produção de mamão no Brasil.....	21
<b>Figura 3.</b> Sintomas da infecção pelo complexo PMeV em <i>Carica papaya</i> .....	22
<b>Figura 4.</b> Alterações em diferentes processos biológicos causados após a infecção em interações virais compatíveis.. ..	26
<b>Figura 5.</b> Inter-relação hormonal em interações planta-virus.....	29
<b>Figura 6.</b> Resposta de defesa desencadeada durante infecções virais compatíveis em <i>Arabidopsis</i> .....	31
<b>Figura 7.</b> Metodologia do sequenciamento de RNA usando a plataforma Illumina...	41
<b>Figura 8.</b> Desenho experimental e linha temporal.. ..	50
<b>Figura 9.</b> Estratégia experimental. ....	50
<b>Figura 10.</b> Análise de componentes principais (PCA).....	52
<b>Figura 11.</b> Plantas de <i>C. papaya</i> no estágio de preflorescência (4 MAG).....	54
<b>Figura 12.</b> Plantas de <i>C. papaya</i> no estágio de pós-floração (9 MAG).....	55
<b>Figura 13.</b> Diagrama de Venn representando a distribuição e ocorrência dos transcritos nos estágios de pré- e pós-floração.....	63
<b>Figura 14.</b> Diagramas de relação entre os conjuntos de GDE e anotação funcional em categorias designadas pelo programa MapMan.....	65
<b>Figura 15.</b> Diagrama de cores mostrando os níveis de expressão ( $Log_2$ Fold changes) dos genes diferencialmente expressos.....	69
<b>Figura 16.</b> Correlação entre os níveis de expressão relativa (Inoculadas/controle) obtidos usando qRT-PCR e RNAseq.. ..	75
<b>Figura 17.</b> Alinhamento das sequências das proteínas cpNPR1-I, jcNIM1-I e rNRR.....	77
<b>Figura 18.</b> Níveis de expressão de nove genes relacionados à via do AS nas plantas infectadas de <i>C. papaya</i> .....	78
<b>Figura 19.</b> Efeito da aplicação do AS na resposta de <i>C. papaya</i> à infecção pelo PMeV e PMeV2.....	79
<b>Figura 20.</b> Modelo proposto para representar as vias envolvidas na resposta de <i>C. papaya</i> à infecção pelo PMeV.....	83

## LISTA DE TABELAS

<b>Tabela 1.</b> Resumo comparativo entre as principais plataformas de sequenciamento de nova geração. Adaptada de (BARBA; CZOSNEK; HADIDI, 2014).....	37
<b>Tabela 2.</b> Lista de genes de <i>C. papaya</i> selecionados para a validação da expressão gênica relativa usando qRT-PCR e primers usados para a quantificação de RNA viral do PMeV e PMeV2.....	60
<b>Tabela 3.</b> Estatísticas do set de dados produzidos no sequenciamento RNA-Seq...	62
<b>Tabela 4.</b> Análises de enriquecimento das ontologias gênicas (GO) nos GDEs nas plantas de <i>C. papaya</i> infectadas pelo complexo PMeV durante a prefloração.....	67
<b>Tabela 5.</b> Análises de enriquecimento de ontologia gênicas nos GDEs nas plantas de <i>C. papaya</i> infectadas pelo complexo do PMeV durante a pós-floração. ....	68
<b>Tabela 6.</b> Genes envolvidos no mecanismo de ROS que foram diferencialmente expressos nas plantas de <i>C. papaya</i> infectadas pelo complexo PMeV durante a pré- e o pós-floração.....	70
<b>Tabela 7.</b> Genes envolvidos na sinalização e metabolismo do ET e JA que foram diferencialmente expressos nas plantas de <i>C. papaya</i> infectadas pelo complexo PMeV durante a pré- e a pós-floração. ....	71
<b>Tabela 8.</b> Genes codificando receptores de membrana com repetições ricas em Leucina (LRR-RK) que foram diferencialmente expressos nas plantas de <i>C. papaya</i> infectadas pelo complexo PMeV durante a prefloração. ....	72
<b>Tabela 9.</b> Genes codificando proteínas de parede celular e metabolismo de calose que foram diferencialmente expressos nas plantas de <i>C. papaya</i> infectadas pelo complexo PMeV durante a pré- e o pós-floração. ....	73
<b>Tabela 10.</b> Genes envolvidos na sinalização e resposta ao AS que foram diferencialmente expressos nas plantas de <i>C. papaya</i> infectadas pelo complexo PMeV durante a pré- e a pós-floração. ....	76

## LISTA DE ABREVIATURAS E SIGLAS

AJ	Ácido jasmônico (do inglês, <i>Jasmonic acid</i> )
AS	Ácido salicílico
BSMT1	Carboxil-metil transferase do ácido salicílico do inglês, <i>Benzoic acid/salicylic acid methyltransferase</i> ).
CC	motivo super-hélice (do inglês <i>coiled-coil</i> )
cDNA	DNA complementar (do inglês, <i>Complementary DNA</i> )
Ct	Limiar do ciclo (do inglês, <i>cycle threshold</i> )
DIR1	Proteína falha na indução de Resistencia 1 (do inglês, <i>defect in induced-resistance protein 1</i> )
dsRNA	RNA dupla-fita (do inglês, <i>Double-strand RNA</i> )
ESTs	Etiquetas de Sequência Expressa (do inglês, <i>expressed sequence tag</i> )
FDR	Taxa de falsa descoberta (do inglês, <i>False discovery rate</i> )
FMO1	Monooxygenase 1 flavinho-dependente (do inglês, <i>flavin-dependent monooxygenase 1 monooxygenase 1 do flavin</i> )
FPKM	Fragments por kilobase de transcrito por milhões de fragmentos mapeados (do inglês, <i>Fragments Per Kilobase of transcript per Million fragments mapped</i> )

GO	Ontologia Gênica (do inglês, <i>Gene ontology</i> )
GRP	Proteína rica em glicina (do inglês, <i>Glycine rich protein</i> ).
HR	Reação de hipersensibilidade (do inglês, <i>Hypersensitive response</i> )
LRR	Repetições ricas em leucine (do inglês, <i>leucine-rich repeat</i> )
LTPs	Proteínas de transferência de lipídeos (do inglês, <i>lipid transfer protein</i> ).
MeSA	Salicilato de metilo (do inglês, <i>methyl-salicylate</i> ).
NBS	sítio de ligação de nucleotídeo (do inglês, <i>Nucleotide binding site</i> )
NGS	Sequenciamento de nova geração (do inglês, <i>Next Generation Sequencing</i> )
NO	Óxido nítrico (do inglês, <i>Nitric oxide</i> )
PMeV	<i>Papaya meleira virus</i>
RNA seq	Sequenciamento do RNA (do inglês, <i>RNA sequencing</i> )
ROS	Espécies reativas de oxigênio (do inglês, <i>Reactive oxygen species</i> )
RT	Transcrição reversa (do inglês, <i>reverse transcription</i> )
RT-PCR	Transcrição reversa-Reação em cadeia da polimerase (do inglês, <i>Reverse transcription-Polymerase chain reaction</i> )

SAGE	Análise serial da expressão gênica (do inglês, <i>serial analysis of gene expression</i> )
SAR	Resposta sistêmica adquirida (do inglês, <i>Systemic acquired resistance</i> )
TFs	Fatores de transcrição (do inglês, <i>Transcription fator</i> )
TIR	<i>Drosophila</i> Toll, animal interleukin-1 receptors
TSWV	Tomato spotted wilt vírus
UGT	udp-glucosyltransferase (do inglês, <i>UDP-glucosyltransferases</i> )



## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	19
1.2 BIOLOGIA DE <i>C. PAPAYA</i> E A SUA IMPORTÂNCIA NA ECONOMIA DA REGIÃO .....	19
1.3 MELEIRA DO MAMOEIRO: UMA VIROSE RELACIONADA AO DESENVOLVIMENTO .....	21
1.4 ASPECTOS DA GENÔMICA E ABORDAGENS COMPUTACIONAIS NA PREDIÇÃO DE GENES ASSOCIADOS AO DESENVOLVIMENTO EM <i>C. PAPAYA</i> .. .....	24
1.5 RESPOSTAS DA PLANTA À INFECÇÃO VIRAL .....	25
1.6 MODULAÇÃO DA EXPRESSÃO DE GENES DE DEFESA EM INTERAÇÕES VIRAIS .....	27
1.7 AVANÇOS NAS ANÁLISES DO TRANSCRIPTOMA EM <i>C. PAPAYA</i> .....	32
1.8 SEQUENCIAMENTO TOTAL DO TRANSCRIPTOMA: RNA-SEQ E AS NOVAS TECNOLOGIAS DE SEQUENCIAMENTO.....	34
1.9 PROCESSAMENTO DE DADOS DE RNA-SEQ POR BIOINFORMÁTICA .....	42
<b>2 OBJETIVOS</b> .....	<b>48</b>
2.1 OBJETIVO GERAL .....	48
2.2 OBJETIVOS ESPECÍFICOS .....	48
<b>3 MATERIAIS E MÉTODOS</b> .....	<b>49</b>
3.1 DESENHO EXPERIMENTAL E OBTENÇÃO DO MATERIAL VEGETAL .....	49
3.2 EXTRAÇÃO DO RNA E SEQUENCIAMENTO .....	51
3.3 MAPEAMENTO DE LEITURAS OBTIDAS POR RNA-SEQ NO GENOMA DE REFERÊNCIA .....	51
3.4 ANÁLISE DA EXPRESSÃO DIFERENCIAL DE GENES (EDG).....	52
3.5 CLASSIFICAÇÃO FUNCIONAL DOS GDE E ANÁLISE DE VIAS DE RESPOSTA A DEFESA.....	56

3.6 VALIDAÇÃO DA EXPRESSÃO GÊNICA ATRAVÉS DE PCR EM TEMPO REAL (qRTPCR).....	56
3.7 APLICAÇÃO EXÓGENA DO ACIDO SALICÍLICO.....	58
3.8 QUANTIFICAÇÃO ABSOLUTA DAS CÓPIAS DE RNA DOS VÍRUS PMEV E PMEV2 .....	58
<b>4 RESULTADOS.....</b>	<b>61</b>
4.1 ANÁLISE DO TRANSCRIPTOMA DE <i>C. PAPAYA</i> USANDO RNA-SEQ.....	61
4.2 ANÁLISE DA EXPRESSÃO DIFERENCIAL DE GENES (EDG) DURANTE A INFECÇÃO EM <i>C. PAPAYA</i> PELO COMPLEXO PMEV NA PRÉ- E PÓS-FLORAÇÃO .....	63
4.3 CLASSIFICAÇÃO FUNCIONAL DOS GDES DURANTE A INFECÇÃO EM <i>C. PAPAYA</i> PELO COMPLEXO PMEV NA PRÉ- E PÓS-FLORAÇÃO.....	64
4.4 ANÁLISE DE VIAS DE RESPOSTA DE <i>C. PAPAYA</i> À INFECÇÃO PELO COMPLEXO PMEV NA PRÉ- E PÓS-FLORAÇÃO.....	70
4.5 VALIDAÇÃO DA EXPRESSÃO GÊNICA MEDIANTE PCR EM TEMPO REAL (QRTPCR).....	74
4.6 SINALIZAÇÃO MEDIADA PELO AS NA RESPOSTA DE <i>C. PAPAYA</i> À INFECÇÃO PELO COMPLEXO PMEV .....	75
<b>5 DISCUSSÃO .....</b>	<b>80</b>
<b>6 CONCLUSÕES .....</b>	<b>88</b>
<b>REFERÊNCIAS.....</b>	<b>90</b>
<b>APÊNDICES .....</b>	<b>101</b>
APÊNDICE A.....	101
APÊNDICE B.....	101
APÊNDICE C. ....	101
APÊNDICE D. ....	101
<b>ANEXOS .....</b>	<b>101</b>

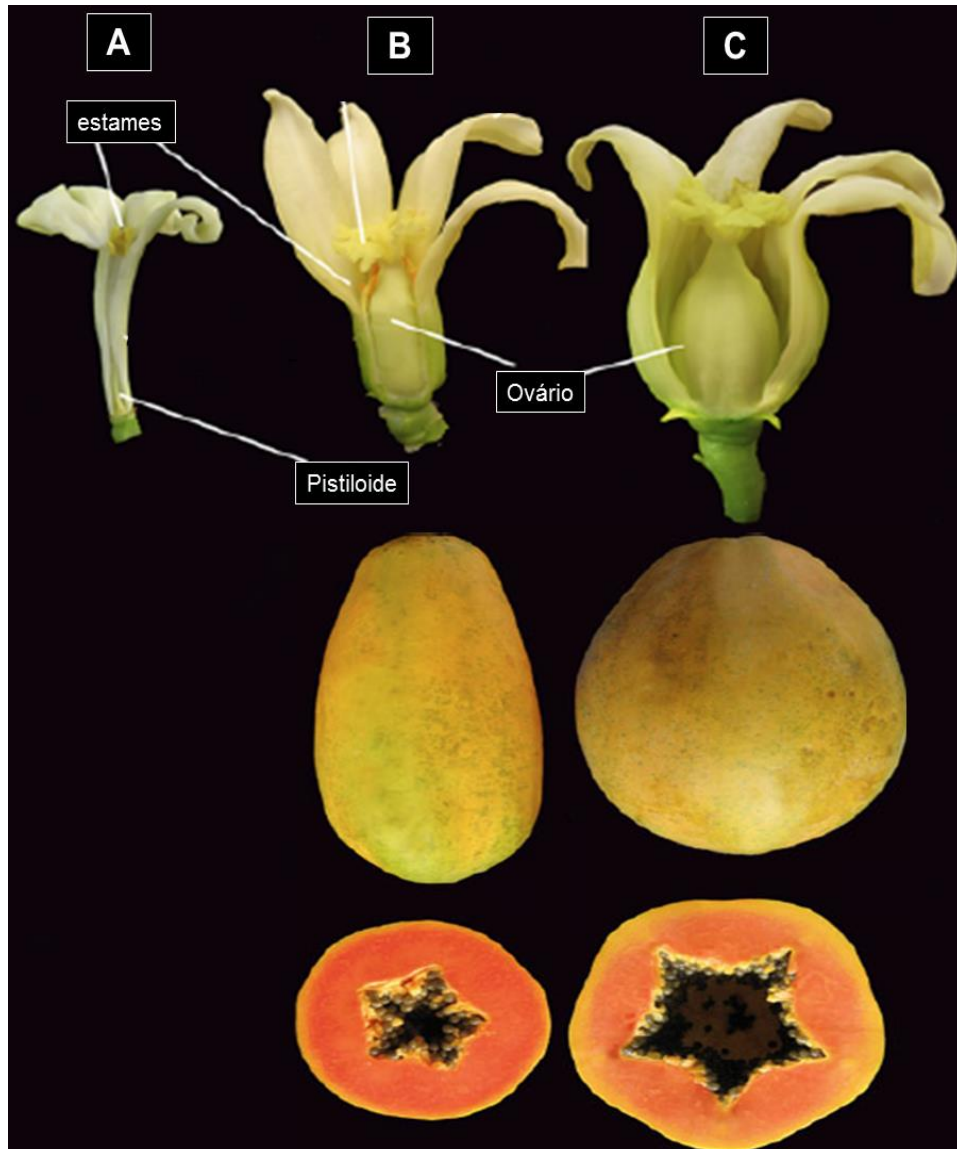
## 1 INTRODUÇÃO

### 1.2 BIOLOGIA DE *C. Papaya* E A SUA IMPORTÂNCIA NA ECONOMIA DA REGIÃO

*Carica papaya* é uma planta perene, semi-lenhosa e produtora de látex, que possui um tempo de desenvolvimento relativamente rápido, sendo que o tempo necessário para completar a fase juvenil, entre a germinação e a floração, leva em média cinco meses enquanto que o tempo requerido para a colheita é de nove a quinze meses (PATERSON et al., 2008). No mamoeiro são observados estágios sucessivos de desenvolvimento, porém, as plantas adultas podem manter simultaneamente crescimento vegetativo, flores e dúzias de frutos em diferentes fases de desenvolvimento (JIMÉNEZ; MORA-NEWCOMER; GUTIÉRREZ-SOTO, 2014).

As plantas do mamoeiro possuem três tipos de flores morfológicamente distintas, podendo produzir plantas com flores femininas, masculinas ou hermafroditas (Figura 1) (YU et al., 2008). O fruto derivado de flores femininas e hermafroditas é composto por cinco carpelos fusionados que formam uma cavidade central contendo numerosas sementes. As plantas do sexo masculino apresentam inflorescências com flores pedunculadas que possuem um pistilo rudimentar, sem estigma e conseqüentemente infértil. A forma do fruto é determinada pelo sexo da planta, desta forma, os frutos que surgem de flores femininas são de forma esférico-ovóide, enquanto que os frutos provenientes de flores hermafroditas são cilíndrico-piriformes (Figura 1). Pela sua forma piriforme e polpa mais espessa (Figura 1), frutos hermafroditos possuem uma maior preferência no mercado mundial (TRIPATHI et al., 2014).

O mamoeiro possui um fruto climatérico típico, caracterizado pelo aumento simultâneo entre a atividade respiratória e a produção de etileno durante a maturação (MING; MOORE, 2014). Bioquímica, fisiológica e, estruturalmente, conforme progride a maturação do fruto, ocorrem várias alterações que incluem a modificação da composição de carboidratos, uma maior produção de metabólitos secundários, tais como o licopeno e a  $\beta$ -criptoxantina, e pronunciadas alterações na cor, aroma e o amolecimento da polpa (GIOVANNONI, 2004).

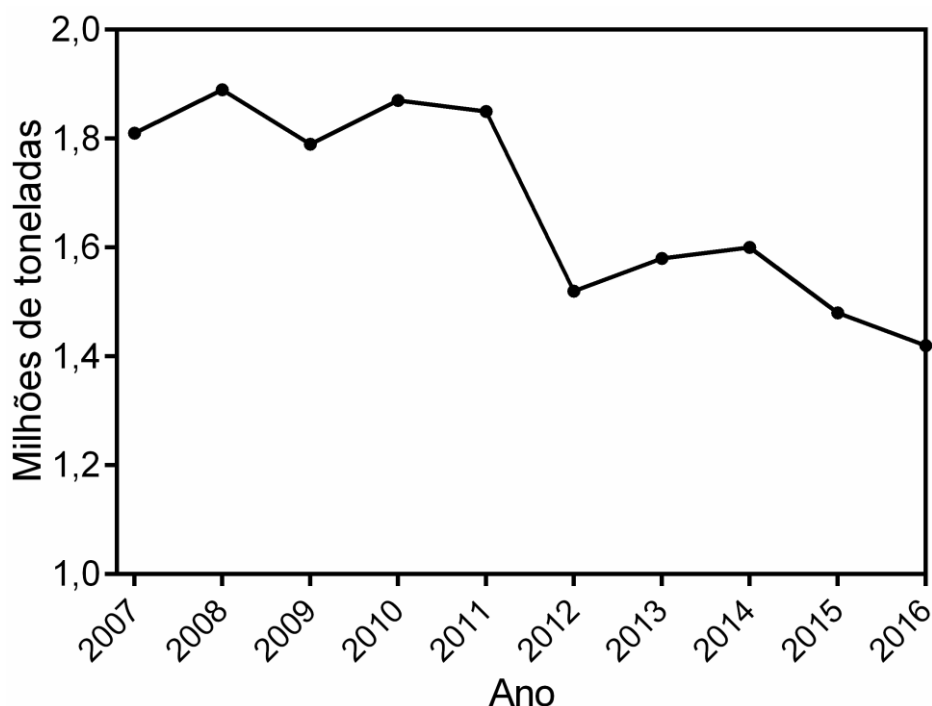


**Figura 1.** Flores e frutos relacionados aos três sexos morfologicamente distintos encontrados no mamoeiro. (A) Flores estaminadas (masculinas) inférteis; (B) Flor perfeita ou hermafrodita e seu fruto piriforme derivado; (C) Flores pistiladas (femininas) e seu fruto ovóide derivado. Adaptado de (JIMÉNEZ; MORA-NEWCOMER; GUTIÉRREZ-SOTO, 2014).

O mamão, uma fruta com um alto conteúdo de antioxidantes, vitamina B, minerais e fibras possui um alto valor nutricional e grandes benefícios para a saúde. Portanto, o comércio mundial de mamão é de grande importância econômica e financeira. O Brasil, segundo maior produtor de mamão, atingiu uma produção de 1,42 milhões de toneladas no ano 2016, o que representa 10% da produção mundial. Destas, 28 mil toneladas (2% da produção total) foram exportadas, gerando uma renda de 41 milhões de dólares, colocando o Brasil como o terceiro maior exportador mundial de mamão (FAO, 2018). No âmbito Nacional, o Estado do Espírito Santo destaca-se por

ser o segundo maior produtor de mamão no Brasil, com uma área colhida de 6.035 há, e uma produção de 251.365 toneladas em 2016 (IBGE, 2018).

Embora o Brasil tenha uma posição destacada no mercado mundial de mamão, na última década o país tem experimentado uma diminuição acelerada na sua produção (Figura 2). De fato, segundo a FAO-TRADE-MARKETS-DIVISION, 2016, o Brasil já foi responsável por 20% da produção mundial no ano 2000, no entanto, esta produção tem diminuído, sendo esperado que diminua ainda mais, abaixo dos 10%, no ano 2024. Dentre os vários fatores que afetam a produção de mamão, destacam-se as doenças causadas por patógenos, especialmente as viroses, que podem levar à destruição completa das plantações infectadas (ABREU et al., 2015).



**Figura 2.** Evolução da produção de mamão no Brasil. Produção de mamão no Brasil em milhões de toneladas durante os últimos dez anos (FAO, 2018).

### 1.3 MELEIRA DO MAMOEIRO: UMA VIROSE RELACIONADA AO DESENVOLVIMENTO

A meleira que foi oficialmente relatada no Brasil e México, dois dos maiores produtores e exportadores de mamão do mundo (ABREU et al., 2015), caracteriza-se pela exsudação espontânea de um látex fluído e aquoso que se acumula

principalmente em folhas e frutos e, quando em contato com o ar, é oxidado e escurece-se rapidamente, dando aos frutos um aspecto de “melado” (Figura 3). Este fator aumenta a suscetibilidade à infestação de insetos oportunistas como a mosca das frutas (*Ceratitis capitata*), uma praga de importância quarentenária que usualmente infesta os frutos em estágios avançados de maturidade (KITAJIMA et al., 1993; VENTURA; COSTA; TATAGIBA, 2004).

A queima ou necrose nas extremidades de folhas jovens, mancha zonada ou clorose nos frutos são também sintomas característicos desta doença (Figura 3), que em conjunto com o aspecto melado no mamoeiro comprometem a qualidade dos frutos, inviabilizando sua exportação. Interessantemente, estes sintomas só aparecem após o estágio de pós-floração, o que permite a permanência de uma planta contaminada no campo por vários meses, servindo de fonte de inoculo às demais plantas do pomar (ABREU et al., 2015).



**Figura 3.** Sintomas da infecção pelo complexo PMeV em *Carica papaya*. (A, B) Aspecto melado; (C) Mancha zonada ou clorose em frutos; (D) Lesões necróticas na ponta de folhas jovens. Adaptado de (VENTURA; COSTA; TATAGIBA, 2004).

Inicialmente, o agente causal da meleira foi identificado como um vírus de partícula isométrica que possui um genoma constituído por RNA de dupla fita (dsRNA), chamado Papaya meleira virus (PMeV). Recentemente, foi demonstrado, usando sequenciamento e purificação de partículas virais, a presença de dois diferentes RNAs, o que levou à indentificação de um segundo vírus de RNA de fita simples (ssRNA) associado à meleira, o Papaya meleira virus 2 (PMeV2). PMeV foi proposto ser classificado como um *toti-like* vírus e, PMeV2 como um *umbra-like* vírus (SÁ ANTUNES et al., 2016).

Até o momento, PMeV e o PMeV2 são os únicos vírus descritos em laticíferos (ARAÚJO et al., 2007; MACIEL-ZAMBOLIM et al., 2003; RODRIGUES et al., 2009a). Os laticíferos do mamoeiro são séries de células anastomosadas que são especializadas na produção e armazenamento de cisteíno-proteases, e de um fluido rico em metabolitos secundários denominado látex (HAGEL; YEUNG; FACCHINI, 2008). Os laticíferos do mamoeiro com meleira acumulam peróxido de hidrogênio (H<sub>2</sub>O<sub>2</sub>), e um látex com níveis aumentados de cristais de oxalato de cálcio e alcalóides (RODRIGUES et al., 2012). A análise global de miRNAs e proteínas em folhas de mamão com meleira mostrou uma alteração no controle do estresse oxidativo e do sistema do proteossoma 26S (ABREU et al., 2014; RODRIGUES et al., 2011). Todavia, componentes chaves na resposta de mamão à infecção pelos dois vírus ainda não foram identificados, principalmente pela carência de genótipos resistentes à doença (MEISSNER FILHO et al., 2017) e ao limitado conhecimento das condições fisiológicas de tolerância na planta (ABREU et al., 2015).

‘Sunrise Solo’ e ‘Golden’ são os cultivares de mamão, economicamente mais relevantes para o Brasil. O desenvolvimento dos sintomas da meleira tem como particularidade a dependência da transição juvenil-adulto, marcada pela floração (MACIEL-ZAMBOLIM et al., 2003). Mesmo as plantas exibindo uma alta carga viral, permanecem assintomáticas até a floração, que ocorre entre 6-8 meses após a transferência de mudas para o campo (ABREU et al., 2015). Este fenômeno sugere que mecanismos de tolerância ao desenvolvimento dos sintomas de meleira operam antes da floração.

#### 1.4 ASPECTOS DA GENÔMICA E ABORDAGENS COMPUTACIONAIS NA PREDIÇÃO DE GENES ASSOCIADOS AO DESENVOLVIMENTO EM *C. Papaya*

*C. papaya* possui um genoma relativamente pequeno, inicialmente descrito, com aproximadamente 372 Mpb (ARUMUGANATHAN; EARLE, 1991), mas recentemente ajustado a 442.5 Mpb (GSCHWEND et al., 2013). O genoma de *C. papaya*, que em 2008 incluía 24.746 genes (MING et al., 2008), hoje alcança quase 28.000 genes (PHYTOZOME, 2017). Estes genes foram agrupados em categorias que envolvem principalmente, crescimento e lignificação celular, biossíntese de carboidratos, resposta ao foto-período e, metabolismo secundário. Estas características na organização do genoma colocam a *C. papaya* numa posição intermediária entre erva e árvore (MING et al., 2008).

Usando abordagens computacionais, genes homólogos associados ao desenvolvimento floral em *Arabidopsis* têm sido identificados em *C. papaya*. Dentre estes, o gene *PAG* (*PAPAYA AGAMOUS*), o qual é o homólogo do gene *AGAMOUS* em *Arabidopsis*, está associado ao desenvolvimento dos carpelos e ao controle do tempo de floração. O gene *PFL*, homólogo de *LEAFY* em *Arabidopsis*, o qual é positivamente regulado pelo gene *AGAMOUS* e integra os sinais ambientais e endógenos para controlar o tempo de floração. O gene *Phua*, homólogo do *hual*, que é um regulador do desenvolvimento do estame em *Arabidopsis* (TRIPATHI et al., 2014; YU et al., 2005).

Uma análise de dados, derivados de ESTs depositados em *The Floral Genome Project* mostrou que aproximadamente 50% dos 200 genes envolvidos em desenvolvimento e regulação floral, controle do tempo de floração, e genes florais órgão-específico que estão presentes em *Arabidopsis*, encontram-se também representados em bibliotecas de cDNA florais em mamoeiro (TRIPATHI et al., 2014).

Em relação aos genes associados à maturação, principalmente foram preditos genes envolvidos na síntese de etileno, respiração, degradação de clorofila e, síntese de carotenoides, processos que podem também impactar na acumulação de açúcar. Os genes *S-adenosyl-L-methionine (S-SAM) sintase*, *1-aminocyclopropane-1-*



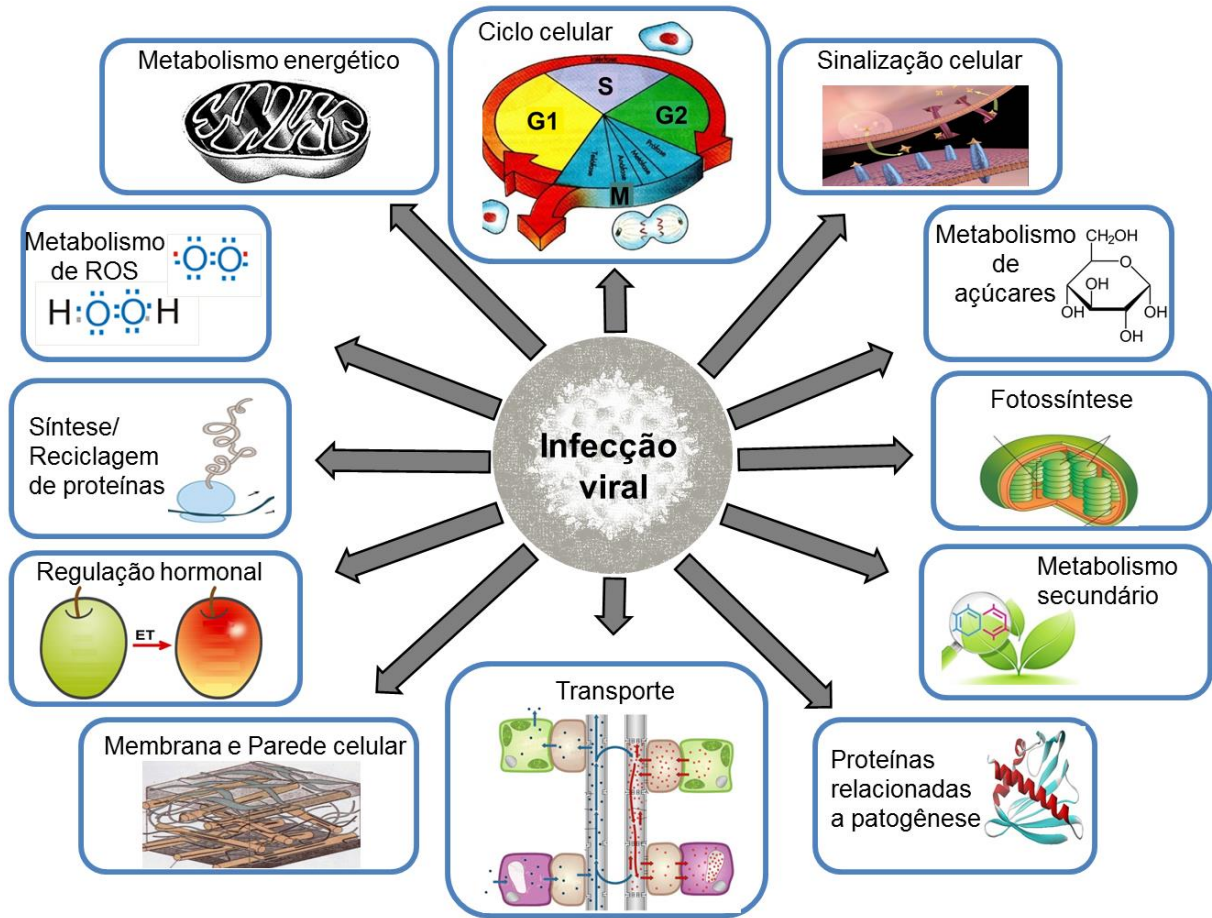
*carboxylate (ACC) sintase* e *ACC oxidase* que estão relacionados à síntese de etileno em *Arabidopsis*, assim como os fatores responsivos do etileno *AP2/ERF*, também se encontram representados no genoma do mamão. Adicionalmente, foram preditos dois genes que codificam uma oxidasa alternativa AOX1 (MING et al., 2008; TRIPATHI et al., 2014), e um grande número de  $\beta$ -1-3-glucanases e  $\beta$ -galactosidases foi predito no genoma do mamoeiro, os quais poderiam estar associados com uma grande necessidade de degradação e ajustamento da parede celular durante a maturação (MING et al., 2008).

## 1.5 RESPOSTAS DA PLANTA À INFECÇÃO VIRAL

As respostas da defesa da planta que conduzem à resistência em interações virais incompatíveis são bem caracterizadas (GOUVEIA et al., 2017; MANDADI; SCHOLTHOF, 2013; NICAISE, 2014). Iniciam-se com o reconhecimento de efetores virais mediante proteínas de resistência (R), e inclui a acumulação de ácido salicílico (AS), espécies reativas de oxigênio (ROS) e proteínas relacionadas à patogênese (PR), indução da resposta hipersensitiva (HR), deposição de calose e morte celular programada (BAEBLER et al., 2014).

Ao contrário das interações virais incompatíveis, nas quais a infecção viral aciona mecanismos de resposta específicos que bloqueiam a replicação e movimentação viral próximo ao local da infecção, as interações compatíveis resultam em infecções virais sistêmicas que são tipicamente acompanhadas pelo desencadeamento de sintomas que caracterizam uma determinada doença (KANG; YEAM; JAHN, 2005).

Embora nas interações compatíveis a planta hospedeira acabe se tornando susceptível à infecção viral, alterações moleculares, bioquímicas e fisiológicas acontecem em células, tecidos e inclusive na planta inteira em resposta à infecção (Figura 4). Contudo, tem sido difícil distinguir os componentes alterados que são requeridos para a replicação e expressão do genoma viral daqueles que estão associados com a resposta da defesa ou prejuízo do hospedeiro (MAULE; LEH; LEDERER, 2002).



**Figura 4.** Alterações em diferentes processos biológicos causados após a infecção em interações virais compatíveis. Processos comumente alterados nas interações virais compatíveis identificados usando estudos fisiológicos, proteômicos e de expressão gênica que incluem metabolismo de espécies reativas de oxigênio (ROS), regulação hormonal, fotossíntese, sinalização celular, metabolismo energético, metabolismo secundário, síntese e reciclagem de proteínas, metabolismo de açúcares e transporte endógeno de macromoléculas. Adaptado de (DI CARLI; BENVENUTO; DONINI, 2012).

Nas interações compatíveis, proteínas de membrana e parede celular são componentes do hospedeiro usados por alguns vírus na sua replicação e movimentação. O ciclo celular é também controlado para promover uma síntese ativa de DNA favorável na replicação viral (MAULE; LEH; LEDERER, 2002). Paralelamente, também são alterados componentes que evocam respostas de defesa clássicas das interações incompatíveis como são a HR, e a resposta sistêmica adquirida (SAR). Sendo, portanto, nas interações compatíveis, sugerido a ativação de uma resposta de estresse celular genérica (*defense-like*) (MAULE; LEH; LEDERER, 2002; WHITHAM; YANG; GOODIN, 2006).

Mediante estudos fisiológicos, proteômicos e de expressão gênica têm sido identificados vários processos que são comumente alterados nas interações virais

compatíveis, e que incluem o metabolismo ROS, regulação hormonal, fotossíntese, sinalização celular, metabolismo energético, metabolismo secundário, síntese e reciclagem de proteínas, metabolismo de açúcares e transporte endógeno de macromoléculas (Figura 4) (DI CARLI; BENVENUTO; DONINI, 2012; MAULE; LEH; LEDERER, 2002; WHITHAM; YANG; GOODIN, 2006).

O efeito da infecção viral e a severidade dos sintomas é o resultado de uma rede complexa de interações entre o vírus, e os componentes da célula hospedeira. Estas interações desencadeiam defesas específicas da planta, e reações de estresse não específicas, onde centenas de genes são modulados, causando assim, a reprogramação transcricional de diferentes vias de sinalização (DI CARLI; BENVENUTO; DONINI, 2012; MAULE; LEH; LEDERER, 2002; WHITHAM; YANG; GOODIN, 2006).

## 1.6 MODULAÇÃO DA EXPRESSÃO DE GENES DE DEFESA EM INTERAÇÕES VIRAIS

A expressão de genes envolvidos em respostas de defesa em ambas as interações compatíveis e incompatíveis é controlada por vias de sinalização que são mediadas principalmente por moléculas sinalizadoras como o ácido salicílico (AS), o ácido jasmônico (AJ), o etileno (ET), ácido abscísico (ABA), auxinas, citocininas (CKs) e os brassinosteróides (BRs), moléculas que agem sinérgica e antagonicamente e influenciam-se mutuamente através de complexas redes regulatórias para produzir um fenótipo de resistência ou susceptibilidade (Figura 5) (ALAZEM; LIN, 2015; CULVER; PADMANABHAN, 2007).

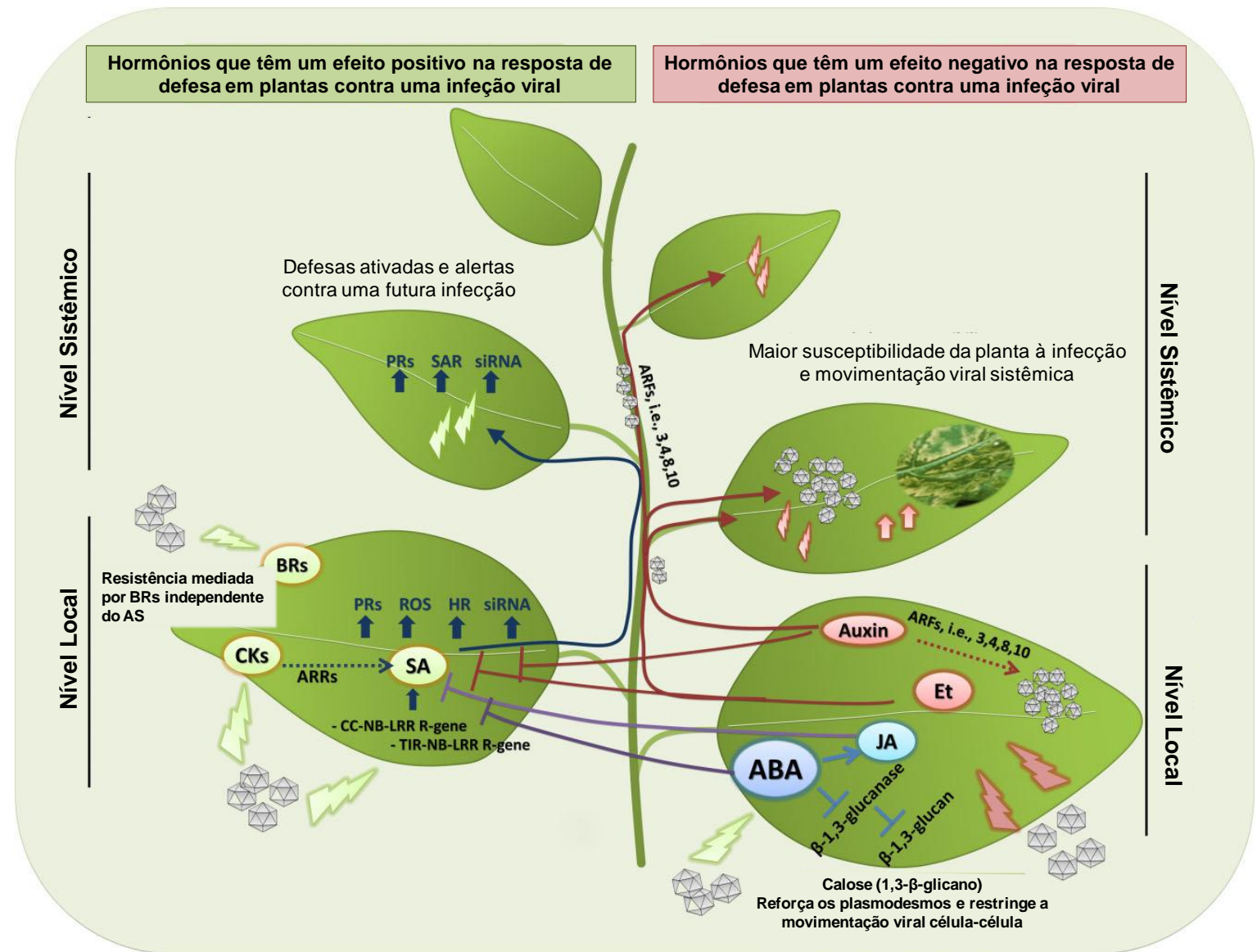
A sinalização mediada por AS constitui a maior via de defesa contra vírus em plantas, e está ligada estreitamente à maioria de genes de defesa da família de receptores NB-LRR (*Nucleotide-Binding Leucine-Rich-Repeat*) envolvidos na ativação de respostas de defesa locais que confinam à infecção viral em lesões necróticas e acionam o mecanismo antiviral de silenciamento (siRNA) e SAR nos tecidos distais (ALAMILLO; SAÉNZ; GARCÍA, 2006; LOAKE; GRANT, 2007; VLOT;

DEMPSEY; KLESSIG, 2009). As CKs contribuem para a defesa da planta em uma maneira dependente de AS. Fatores responsivos às CKs, tais como *ARR3* (ARABIDOPSIS RESPONSE REGULATOR 3), *ARR4*, *ARR5*, *ARR6*, *ARR8* e *ARR9*, estão envolvidos no *cross-talking* entre AS e as CKs (ARGUESO et al., 2012). Por outro lado, os BRs agem independentemente de AS e melhoram a resistência.

As auxinas são conhecidas por sua função antagonista à via do AS. Um grupo de fatores responsivos às auxinas (*ARFs*) é importante na replicação e movimentação de alguns vírus, tais como o *Tobacco mosaic virus* (TMV) (PADMANABHAN et al., 2005, 2008). O etileno também antagoniza a via desencadeada após a sinalização de AS, e está envolvido no desenvolvimento de sintomas durante a infecção pelo *Cauliflower mosaic virus* (CaMV), a movimentação sistêmica da estirpe do TMV que infecta crucíferas, e à formação de lesões necróticas após a infecção com outros vírus (CHEN et al., 2013; GERI et al., 2004).

JA e ABA desempenham um papel ambíguo na ativação de respostas de defesa da planta contra os vírus. JA parece apoiar a defesa da planta em estágios iniciais da infecção, porém, se for induzido ou aplicado em estágios avançados, este hormônio diminui a resistência da planta (GARCÍA-MARCOS et al., 2013; PACHECO et al., 2012). ABA possui muitos papéis durante a defesa das plantas contra infecções virais. Por um lado, aumenta a deposição de calose nos plasmodesmos (PD) e limita a movimentação viral célula-célula em alguns vírus, tais como TMV e *Tobacco necrosis virus* (TNV), enquanto que, por outro lado, antagoniza a via do AS e reduz a resistência ao nível local da infecção (ALAZEM; LIN; LIN, 2014; IRITI; FAORO, 2007; WHENHAM et al., 1986).

**Figura 5.** Inter-relação hormonal em interações planta-vírus. De modo geral, existem hormônios que possuem um papel positivo (em cor verde claro) ou negativo (em cor vermelho claro) na ativação de respostas de defesa contra vírus em plantas. O AS induz a expressão de genes de defesa local e sistêmica mediante a ativação de ROS, HR, silenciamento de RNA (siRNA) e SAR. O ET e as auxinas usualmente antagonizam a via desencadeada após a sinalização de AS, e estão envolvidos no desenvolvimento de sintomas em várias interações planta-vírus. As CKs e os BRs contribuem à defesa da planta, dependente e independente de AS, respectivamente. O AJ (em cor azul) possui um papel ambíguo, sendo que em estágios iniciais da infecção possui um papel positivo na ativação de vias de defesa. Porém, em estágios avançados da infecção, desempenha um papel antagonístico ao AS. ABA (em cor azul claro) possui vários papéis durante a defesa das plantas contra infecções virais. Por um lado, aumenta a deposição de calose nos plasmodesmos (PD) e limita a movimentação viral célula-célula em alguns vírus, enquanto que, por outro lado, antagoniza a via do AS e reduz a resistência ao nível local da infecção. Adaptado de (ALAZEM; LIN, 2015).

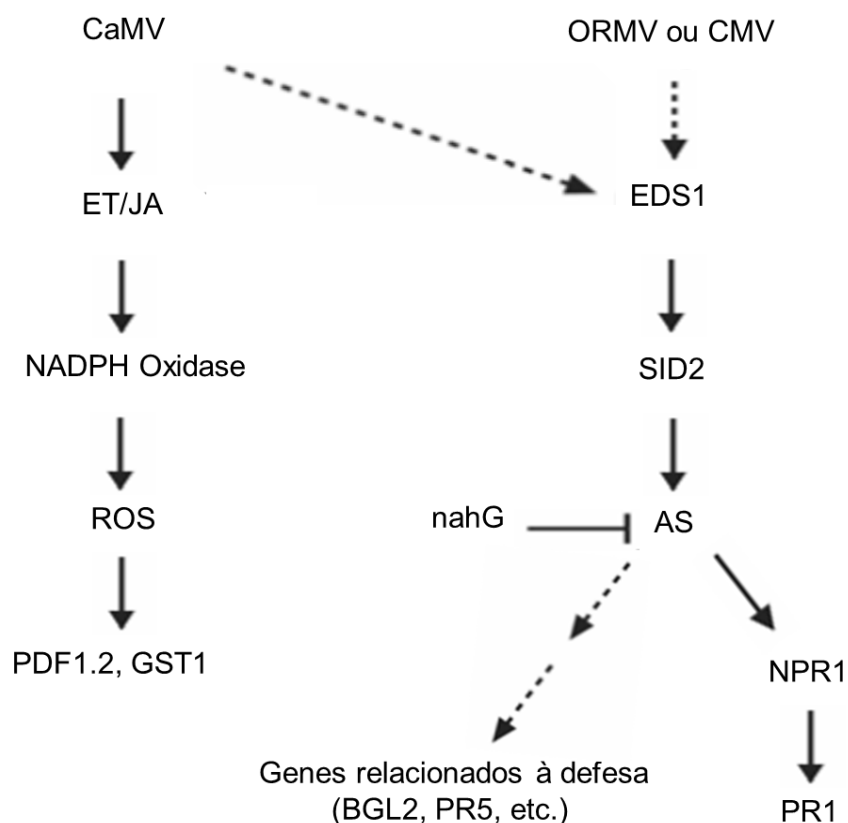


Através do uso de diferentes técnicas como a análise serial da expressão gênica (*serial analysis of gene expression -SAGE*), os microarranjos e RNA-seq aplicados principalmente em plantas modelo como *Arabidopsis* e *Nicotiana benthamiana* demonstraram que, de maneira similar às interações virais incompatíveis, a expressão de diversos genes envolvidos em vias de defesa, também é alterada nas interações virais compatíveis (MAULE; LEH; LEDERER, 2002; WHITHAM; YANG; GOODIN, 2006).

Genes que codificam enzimas envolvidas na remoção e desintoxicação de ROS, como as catalases, a glutationa-S-transferase (GST), a peroxidase, a superóxido dismutase, genes que codificam para várias proteínas relacionadas a patogêneses (PR) como PR-1, PR-2 ( $\beta$ -1,3 glucanase), PR-3 (quitinase), PR-4, PR-5 (taumatina e osmotina), e fatores de transcrição que são tipicamente marcadores de vias de defesa em interações incompatíveis, são induzidos sistemicamente em interações compatíveis (ITAYA et al., 2002; LOVE et al., 2005; MANDADI; SCHOLTHOF, 2012; SENTHIL et al., 2005; SUN et al., 2016). A indução destes genes representa a ativação de vias *stress-like* e *defense-like* que não são consideradas propriamente vias de defesa, devido à carência de um elicitor específico no hospedeiro, ou seja, aparentemente são respostas genéricas que ocorrem na ausência de um reconhecimento específico de um efetor viral (Avr) mediante uma proteína R (WHITHAM; YANG; GOODIN, 2006).

A caracterização das vias de defesa desencadeadas em plantas modelo como *Arabidopsis* em resposta aos vírus *Cauliflower mosaic virus* (CaMV), *Oilseed rape mosaic virus* (ORMV) e *Cucumber mosaic virus* (CMV) (Figura 6) tem ampliado o conhecimento relacionado à indução de genes envolvidos em respostas antivirais nas interações compatíveis. De modo geral, nas interações compatíveis, é requerido um incremento nos níveis de AS para a acumulação de proteínas e genes PRs, porém, a indução de genes *PR* parece não ser mediada por um incremento nos níveis de AS. De fato, em interações compatíveis em *Nicotiana* e *Arabidopsis*, não tem sido observada a síntese de AS, portanto, níveis basais de AS parecem ser suficientes para mediar a expressão de genes relacionados à defesa nas interações compatíveis (DEMPSEY et al., 1997; HUANG et al., 2005; MALAMY; HENNIG; KLESSIG, 1992).

A indução de genes relacionados à defesa mediada por AS em interações compatíveis sugere que as plantas hospedeiras poderiam dividir uma via de sinalização comum com as interações incompatíveis, no entanto, os baixos ou não existentes níveis de AS impedem a amplificação da sinalização necessária para uma robusta indução de genes PR, debilitando ou suprimindo a resposta de defesa. Portanto, as diferenças entre as duas interações está baseada na sincronização e magnitude da resposta induzida, mais do que na expressão de diferentes conjuntos de genes (ELVIRA et al., 2008; HUANG et al., 2005; WHITHAM; YANG; GOODIN, 2006; YANG et al., 2007).



**Figura 6.** Resposta de defesa desencadeada durante infecções virais compatíveis em *Arabidopsis*. O vírus *Cauliflower mosaic virus* (CaMV) induz a expressão de genes responsivos ao JA e o ET (*PDF1.2* e *GST1*), assim como a expressão do gene *PR1* dependente de AS. As linhas pontilhadas indicam que não é conhecido se *SID2* ou *EDS1* são requeridos para começar a expressão de genes dependentes de AS durante a infecção pelo vírus CaMV. Os vírus *Oilseed rape mosaic virus* (ORMV) e *Cucumber mosaic virus* (CMV) induzem a expressão de genes relacionados à defesa através da via dependente de *SID2*, *EDS1* e AS. A expressão de *PR1* e outros genes relacionados à defesa, como (*BGL2* e *PR5*), parece ser ativada após a indução de AS, já que o aumento na expressão de *PR1* é altamente dependente de NPR1. Adaptado de (WHITHAM; YANG; GOODIN, 2006).

Proteínas e enzimas reguladoras do complexo proteossomal 26S, 20S, espécies reativas de oxigênio (ROS) e calreutilina, são acumuladas nas folhas de mamoeiro

com meleira (ABREU et al., 2014; RODRIGUES et al., 2011). Mecanismos de resistência como SAR, que envolve moléculas sinalizadoras, hormônios como o AJ, AS e etileno ET e a proteína *non-expressor of pathogenesis-related protein 1* (NPR1/NIM1), são conservados em *C. papaya* (MING et al., 2008; PORTER et al., 2009). Entretanto, ainda não foi determinado o envolvimento destas proteínas e mecanismos de resistência na interação entre *C. papaya* e o complexo PMeV.

Para obter uma visão global das complexas redes regulatórias desencadeadas nas interações planta-vírus, é essencial o uso de tecnologias de alto rendimento como a proteômica e a transcriptômica.

### 1.7 AVANÇOS NAS ANÁLISES DO TRANSCRIPTOMA EM *C. papaya*

O sequenciamento e a análise do transcriptoma é essencial na validação de dados genômicos, descoberta de variantes genômicas, determinação de *splicing* alternativo e desenvolvimento de marcadores de DNA. Apesar de já existir o genoma de *C. papaya* sequenciado (MING et al., 2008), atualmente as análises de expressão gênica em mamoeiro ainda se encontram em um estágio inicial, e baseiam-se primariamente no sequenciamento de ESTs em algum estágio específico de desenvolvimento (TRIPATHI et al., 2014). A bibliografia mais abundante corresponde ao processo de maturação, e há pouca informação das alterações na expressão gênica das interações planta-patógeno em mamoeiro.

Atualmente estão disponíveis no *GenBank* 77.393 ESTs de *C. papaya*. Os transcritos cobrem aproximadamente 3,6% do genoma inteiro, e 48% da região codificante (MING et al., 2008). O primeiro estudo usando ESTs em *C. papaya* analisou a expressão de genes durante o desenvolvimento do fruto, gerando 1171 ESTs derivados de duas bibliotecas independentes de cDNA procedentes do fruto de um cultivar híbrido de Taiwan 'Tainung' (polpa vermelha) e de um híbrido australiano 1B (polpa amarela). As sequências foram principalmente associadas a enzimas envolvidas na degradação de quitina (quitinases), controle do metabolismo de etileno (ACC oxidase), decomposição do peróxido de hidrogênio (catalase) e,



biossíntese de metionina. ESTs com similaridade significativa, os quais estão relacionados com o amaciamento, aroma, e biossíntese da cor também foram identificados (DEVITT et al., 2006).

Outras abordagens no estudo de transcriptoma de mamão como os microarranjos também têm sido usadas. FABI e colaboradores, 2012 avaliaram a expressão de genes em frutos maduros e imaturos da *C. papaya* variedade Golden, usando uma plataforma de arranjos do genoma de *Arabidopsis*. Adicionalmente, o perfil de expressão foi comparado com dados de microarranjos derivados do tomate e uva, os quais mostraram que, em geral, existem muitas similaridades nos perfis de expressão entre o mamoeiro e tomateiro, principalmente na regulação do metabolismo primário, processo de transcrição, resposta ao estresse biótico e abiótico, e metabolismo da parede celular. Fatores de transcrição, membros das famílias MADs box, NAC e AP2/ERF, que são reguladores chaves na maturação, são também induzidos no mamoeiro durante a maturação.

Recentemente, usando a tecnologia de sequenciamento massivo de RNA (*RNA-seq*), foi comparado o transcriptoma da variedade transgênica do *C. papaya* 'SunUp', resistente ao *Papaya ringspot virus* (PRSV), com a sua variedade progenitora 'Sunset', suscetível ao PRSV. No total foram identificados 20.700 transcritos e 842 genes diferencialmente expressos. Os genes que foram diferencialmente expressos foram associados principalmente a categorias relacionadas aos microtúbulos, fatores de transcrição, transportadores, e biossíntese de hormônios (FANG et al., 2016).

Em relação à interação planta-vírus, em interações compatíveis, a maioria dos trabalhos é realizada utilizando-se plantas modelo, principalmente *Arabidopsis*. Poucos trabalhos abordando o estudo da expressão gênica global em *C. papaya* foram desenvolvidos. Conforme revisão de literatura (TRIPATHI et al., 2014), dentre os trabalhos que envolvem estudos de transcriptômica em mamoeiro, apenas um (ARYAL et al., 2012) aborda a interação planta-hospedeiro. Portanto, é importante o uso de abordagens de sequenciamento em larga escala em *C. papaya* para identificação de alvos que possam ser usados na obtenção de genótipos resistentes a doenças.

## 1.8 SEQUENCIAMENTO TOTAL DO TRANSCRIPTOMA: RNA-Seq E AS NOVAS TECNOLOGIAS DE SEQUENCIAMENTO

Várias tecnologias para conhecer e quantificar o transcriptoma têm sido desenvolvidas, as quais estão fundamentadas principalmente na hibridização e no sequenciamento de cDNA. Tecnologias baseadas em hibridação, tipicamente envolvem a incubação de cDNA derivados de amostras biológicas marcados fluorescentemente com uma série de sondas de DNA imobilizadas em uma matriz sólida que representa os microarranjos de genes de interesse que podem ser customizados ou comerciais. Estas técnicas são consideradas de alto rendimento e são relativamente baratas, com exceção de arranjos desenhados para grandes genomas, contudo, possuem várias limitações como a dependência de um prévio conhecimento da sequência genômica, a possibilidade de hibridação cruzada e, a incerteza relacionada à intensidade do sinal, devido a que a técnica apresenta uma escala limitada de detecção (WANG; GERSTEIN; SNYDER, 2009).

Ao contrário dos microarranjos, os métodos baseados no sequenciamento determinam diretamente a sequência de cDNA. Inicialmente foi, comumente usado o sequenciamento Sanger de cDNA ou ESTs (*expressed sequence tag*), porém, esta metodologia é relativamente de limitado desempenho, custosa, e geralmente não quantitativa (WANG; GERSTEIN; SNYDER, 2009). Para superar estas limitações, métodos que se baseiam em pequenas sequências *Tags*, incluindo a SAGE (VELCULESCU et al., 1995), a análise de expressão de genes *cap* (do inglês *cap analysis of gene expression* –CAGE), e o sequenciamento massivo em paralelo (do inglês *massively parallel signature sequencing* –MPSS) foram desenvolvidos.

Estes métodos são de alto desempenho e podem proporcionar uma quantificação digital acurada dos níveis de expressão gênica, porém, estão baseados na onerosa tecnologia de sequenciamento Sanger. Adicionalmente, apenas uma porção de transcritos é analisada, sendo que, uma fração significativa de pequenos *Tags* não pode ser mapeada de forma exclusiva no genoma de referência, e as isoformas são geralmente indistinguíveis uma da outra (WANG; GERSTEIN; SNYDER, 2009).

O desenvolvimento das tecnologias de sequenciamento de nova geração (do inglês *Next generation sequence* –NGS) tem permitido sequenciar, mapear e quantificar transcriptomas através de um método de alto desempenho conhecido como *RNAseq* (*RNA sequencing* ou *whole transcriptome sequencing* –WTSS). Este método supera amplamente as limitações das outras metodologias existentes, sendo que, os reduzidos custos, tempo e volume de dados produzidos fazem desta uma metodologia altamente acessível (BARBA; CZOSNEK; HADIDI, 2014; WANG; GERSTEIN; SNYDER, 2009).

Atualmente, a tecnologia RNA-seq encontra-se disponível comercialmente em seis plataformas, que estão classificadas em dois grupos. (I) Tecnologias nas quais, prévio ao sequenciamento, deve ser feita uma amplificação do cDNA usando algum tipo de PCR, agrupando quatro plataformas: Roche GS FLX 454 *sequencer* (Roche Diagnostics Corp., Branford, CT, USA), *Illumina genome analyzer* (Illumina Inc., San Diego, CA, USA), ABI SOLiD *System* (Life Technologies Corp., Carlsbad, CA, USA), e *Ion Personal Genome Machine* (Life Technologies, South San Francisco, CA, USA). (II) Tecnologias que estão baseadas no sequenciamento de uma molécula única e, portanto, não requer uma amplificação prévia ao sequenciamento. Pertencem a este grupo a HeliScope (Helicos BioScience Corp., Cambridge, MA, USA) e, *PacBio RS single-molecule real-time (SMRT) system* (Pacific Biosciences, Menlo Park, CA, USA).

Entre estas seis plataformas disponíveis, o *Illumina/Solexa Genome Analyzer*, a Roche 454 GS FLX *sequencer*, o *Applied Biosystems SOLiD Analyzer* e a HeliScope (que pertence às tecnologias de sequenciamento de segunda geração) dominam o mercado, enquanto que a *Pacific Biosciences PacBio RS SMRT system* e a *Ion Personal Genome Machine* da *Life Technologies* (terceira geração) têm sido recentemente incorporadas no mercado, e portanto, ainda não são de amplo uso (Tabela 1) (BARBA; CZOSNEK; HADIDI, 2014; ESCALONA; ROCHA; POSADA, 2016; JAIN et al., 2014).

Cada plataforma de sequenciamento possui vantagens e desvantagens. Exemplo, a plataforma *Illumina* oferece custos menores, porém, as leituras geradas são de tamanho menor do que as geradas pela Roche/454. Quando for requerida uma montagem “*de novo*” (do inglês, *de novo assembly*) leituras com tamanhos maiores

são recomendáveis. Entretanto, quando for preciso ter uma alta cobertura (do inglês, *high depths of coverage*), o rendimento e a acurácia, são priorizados, em vez do tamanho das leituras (BARBA; CZOSNEK; HADIDI, 2014; OZSOLAK; MILOS, 2011).

Enquanto que Roche 454 gera as leituras com maior tamanho, a plataforma SOLiD 5500 xls oferece a maior acurácia no sequenciamento; já a plataforma Illumina apresenta a maior capacidade de sequenciamento, menores custos e tempos, e uma boa acurácia (LIU et al., 2012) (Tabela 1). Em conjunto, estas variáveis têm posicionado a illumina como, provavelmente, a plataforma que domina o mercado dos sequenciadores.

A Illumina desenvolveu a série de plataformas HiSeq® 2500, HiSeq 2000, HiSeq 1500 e HiSeq 1000. A *HiSeq 2500* tem a capacidade de sequenciar um genoma em 24 horas, 20 exomas num dia ou 30 amostras para *RNA sequencing* em aproximadamente 5 horas, enquanto que a *HiSeq 2000* é capaz de sequenciar 600 bilhões de pares de bases (pb) per *lane* (BARBA; CZOSNEK; HADIDI, 2014).

**Tabela 1.** Resumo comparativo entre as principais plataformas de sequenciamento de nova geração. Adaptada de (BARBA; CZOSNEK; HADIDI, 2014).

Método de amplificação	Plataforma	Química do sequenciamento	Tamanho das leituras (pb)	Máxima produção per corrida	Rendimento /hora	Acurácia (%)
PCR de emulsão	Ion Torrent (Life Technologies)	Detecção da liberação do H <sup>+</sup>	100-400	100 Mb a 64 Gpb	25 Mpb–16 Gpb	99
	454 (Roche)	Piro-sequenciamento	400-700	700 Mpb	13 Mpb	99.9
	SOLiD (Life Technologies)	Ligação	75-85	80 a 360 Gpb	21–28 Mpb	99.99
Amplificação em ponte (Bridge PCR)	Illumina (ILLUMINA)	Terminadores reversíveis	100-300	600 Gpb	25 Mpb	99.9
Sem amplificação. Molécula única em tempo real (No amplification Single molecule real-time: SMRT)	PacBio (Pacific Biosciences)	Nucleotídeos marcados fluorescentemente	4000-5000	200 Mpb a 1 Gpb	50–115 Mpb	95
	Helicos (Helicos Biosciences)	Terminadores reversíveis	25-55	35 Gpb	83 Mpb	97
	Nanopore (Oxford Technologies)	-	Leituras muito longas até de 50 kpb	Dezenas de Gpb	150 Mpb	96

pb=par de bases

kpb= 1.000 pb

Mpb=1.000.000 pb

Gpb=1.000.000.000 pb

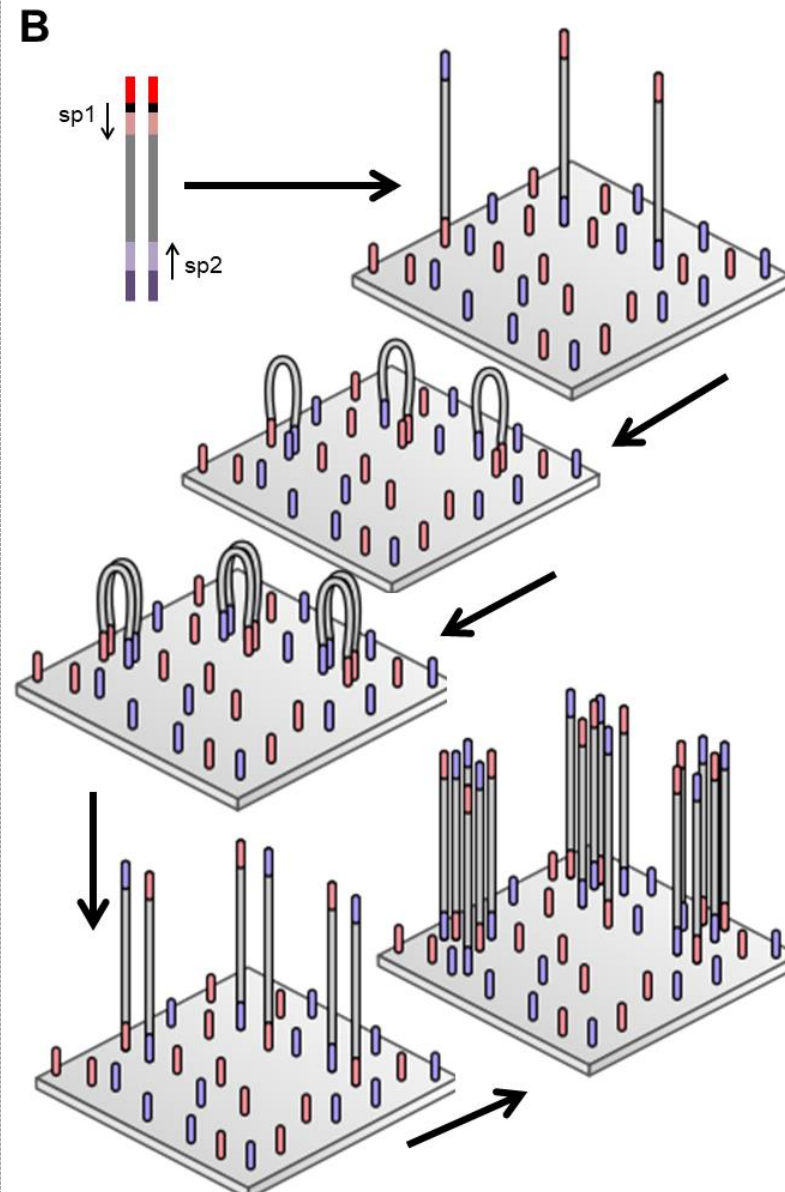
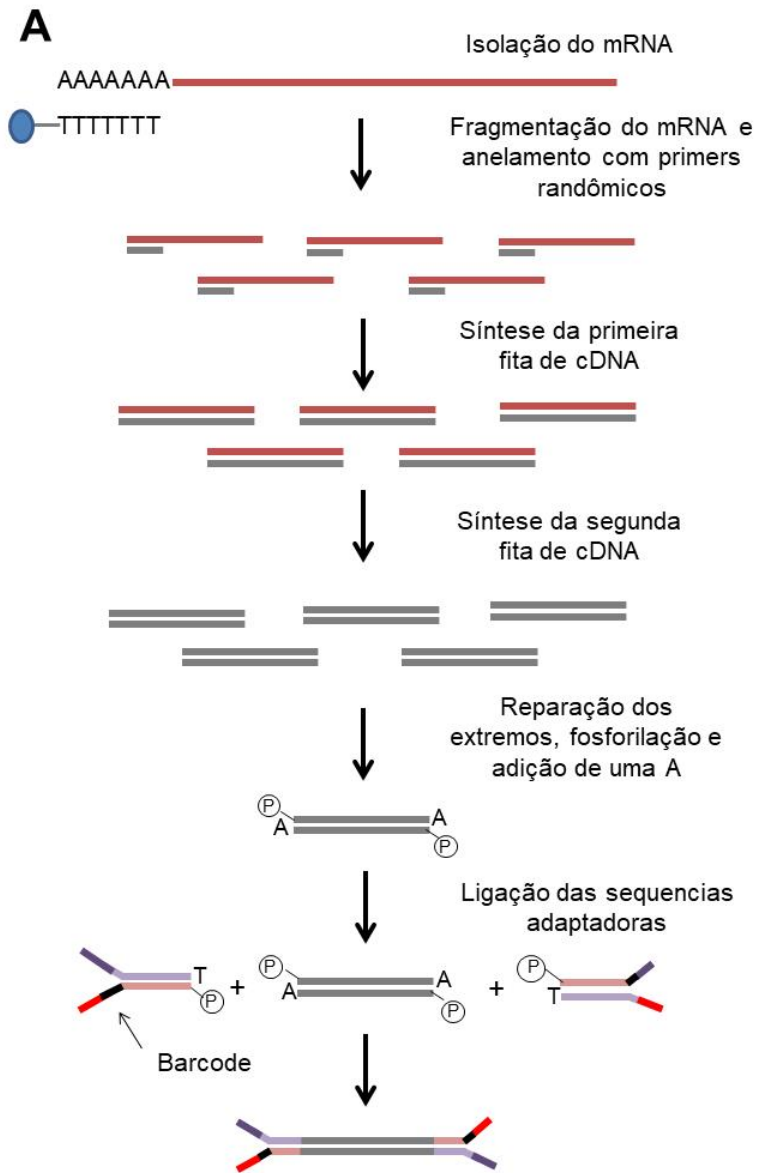
A metodologia de sequenciamento do RNA apresenta algumas variações, dependendo da plataforma usada, contudo, três passos são fundamentais: (1) preparação de uma biblioteca de cDNA: durante este passo uma população de RNA (total ou fracionado) é convertido numa biblioteca de fragmentos de cDNA com adaptadores ligados em um ou nos dois extremos da dupla fita. (2) Amplificação da biblioteca de cDNA: quando são usadas plataformas que não requerem amplificação, este passo é omitido. (3) Sequenciamento: neste passo, cada plataforma apresenta variações tanto no tamanho das leituras ou *reads*, que podem variar de 25 pb a milhares de pb (kpb), quanto no método utilizado, que pode ser mediante sequenciamento por ligação (no caso da plataforma SOLiD), sequenciamento por síntese com terminação reversível a cada ciclo (no caso das plataformas da Illumina), sequenciamento por síntese com detecção de sinal na incorporação de nucleotídeo único (nas plataformas 454 GS e Ion Torrent) e sequenciamento de moléculas únicas em tempo real (empregado nas mais recentemente introduzidas, PacBio, Helicos e Nanopore) (Tabela 1).

Os passos seguidos no sequenciamento do RNA são mostrados em detalhe na Figura 7, usando como exemplo o procedimento da plataforma Illumina. Brevemente, a preparação da biblioteca de cDNA começa a partir da seleção do mRNA do RNA total, usando esferas magnéticas ricas em Poli-T que se ligam às caudas poli A. Para a obtenção de fitas menores, as moléculas de mRNA são fragmentadas usando sonicação ou métodos enzimáticos. Estas moléculas são submetidas à transcrição reversa usando a enzima transcriptase reversa e primers randômicos. Após a síntese da primeira e segunda fita de cDNA, os extremos são reparados, fosforilados nos extremos 5' e adenilados com uma única A nos extremos 3'. Este procedimento prepara as moléculas de cDNA para a ligação com umas sequências curtas de DNA dupla fita adaptadoras (do inglês, *adapters*) que contêm vários elementos, entre estas sequências de identificação *Barcode* e sítios de ligação para os primers (sp1 e sp2) (Figura 7A).

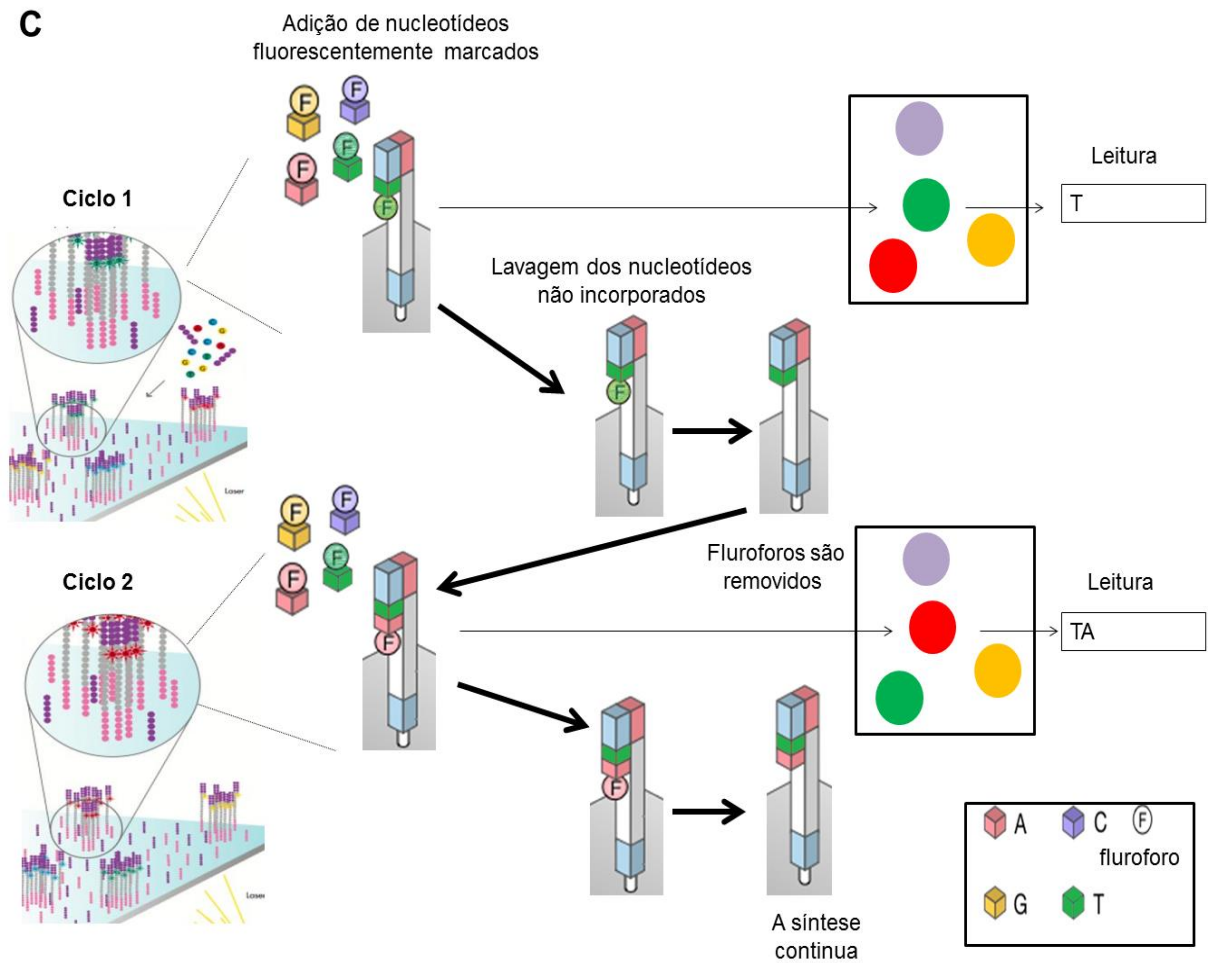
Os fragmentos que possuem os adaptadores em ambas as extremidades permitem o pareamento com *templates* complementares fixos numa superfície sólida chamada célula de fluxo (do inglês, *flow cell*). Os *templates* ligados aos adaptadores dos fragmentos de cDNA agem como iniciadores senso-antisenso, originando pontes

que favorecem a amplificação na presença de nucleotídeos e da enzima polimerase. Os amplicons ficam imobilizados e, após uma desnaturação, formam uma nova ponte. Estes passos são repetidos sucessivamente, gerando milhões de grupos ou *clusters* de cada fragmento (Figura 7B). A formação de *clusters* é importante para a obtenção de um sinal fluorescente forte o suficiente para fazer a leitura do nucleotídeo incorporado.

Uma vez formados os milhões de *clusters*, uma nova desnaturação é feita para dar início a ciclos automatizados de extensão e captura de imagem, sendo que nucleotídeos marcados com fluroforos reversíveis são introduzidos na reação. Estes nucleotídeos têm propriedades de terminação, o que permite parar a síntese quando a DNA polimerase incorpora o correspondente nucleotídeo na fita nascente. Uma vez integrado o nucleotídeo na fita nascente, os fluróforos são ativados por um laser, de tal forma que a luz emitida será diferencial, dependendo do nucleotídeo incorporado. Esta informação é capturada por uma câmera, sendo suas imagens interpretadas por um programa e armazenadas. Uma vez terminado o processo anterior, os nucleotídeos não incorporados são lavados, e os fluróforos "terminadores" são cortados enzimaticamente para que um novo ciclo permita a incorporação do seguinte nucleotídeo (Figura 7C) (ATDBIO, 2017; CORNEY, 2013; ILLUMINA, 2010).







**Figura 7.** Metodologia do sequenciamento de RNA usando a plataforma Illumina. (A) A preparação da biblioteca de cDNA e ligação das sequências adaptadoras. (B) Pareamento dos fragmentos que possuem as sequências adaptadoras com os templates fixos à célula de fluxo, amplificação em ponte e formação de *clusters*. (C) Amplificação na presença de nucleotídeos marcados com fluóforos reversíveis, captura e processamento de imagem. Adaptado de (ATDBIO, 2017; CORNEY, 2013; ILLUMINA, 2010).

## 1.9 PROCESSAMENTO DE DADOS DE RNA-Seq POR BIOINFORMÁTICA

O processamento de dados de RNA-Seq engloba um conjunto de técnicas computacionais que permitem estimar e comparar a abundância dos transcritos de RNA em diferentes amostras biológicas, em um determinado estágio de desenvolvimento, e/ou sobre determinada condição fisiológica (KORPELAINEN et al., 2018; WANG; GERSTEIN; SNYDER, 2009). A análise de expressão diferencial permite inferir os mecanismos de regulação gênica, inclusive para aqueles organismos em que o genoma não se encontra disponível em bancos públicos de dados.

Existem ferramentas de bioinformática comerciais, por exemplo, *Geneious* ou *CLC Bio*, e programas gratuitos, tais como, *Galaxy* (GOECKS et al., 2010) e *RobinRNA* (LOHSE et al., 2012), que possuem interfaces gráficas que permitem a condução de análises mesmo se o usuário não for especialista em bioinformática. Já, quando grandes quantidades de dados devem ser processadas (dezenas a centenas de Gigabytes (GBs), até mesmo Terabytes (TB)), e resultados com maior acurácia são desejados, é recomendável usar ferramentas que estejam baseadas em algoritmos robustos e eficientes.

De um modo geral, o processamento computacional de dados de RNA-Seq usa *pipelines* que envolvem etapas essenciais, tais como o mapeamento ou montagem das leituras, identificação, anotação, e quantificação de genes e/ou transcritos. Dependendo dos objetivos do trabalho, podem ser usadas principalmente 3 estratégias de montagem: (1) montagem *de novo*, caso não exista um genoma de referência para a espécie de interesse, e/ou quando existem polimorfismos, nucleotídeos ou haplótipos que podem ser perdidos pela comparação com o genoma de referência; (2) montagem usando um transcriptoma de referência, caso as sequências sejam muito curtas (< 50 pares de bases); (3) montagem usando um genoma de referência, quando não se aplica nenhum dos casos anteriores.

Três fatores são fundamentais durante o processo de montagem: (1) acelerar o processo, (2) minimizar a quantidade de memória consumida e (3) maximizar a

qualidade. Atualmente existem várias opções de *softwares* tanto livres como pagos que são adequados para realizar o processo de montagem e que variam em relação à estratégia de montagem usada (FONSECA et al., 2017). Na montagem usando um genoma de referência, merecem destaque três ferramentas gratuitas, *TopHat* (TRAPNELL; PACHTER; SALZBERG, 2009) *Star* (DOBIN et al., 2013) e *HiSat* (KIM; LANGMEAD; SALZBERG, 2015). Esta última, apesar de ser a mais recente, vem ganhando cada vez mais usuários.

Se compararmos a eficiência dos anteriores montadores, tomando como exemplo o processamento de aproximadamente 100 milhões de leituras observamos que, o *TopHat* consome muitas horas na montagem dos dados. Entretanto, ele não requer muita memória, sendo possível fazer a montagem com apenas 4 GB de memória. O *Star* gasta menos tempo para a etapa de montagem, 24 minutos, porém requer de 4-28 GB de memória. Isto dificulta o uso de máquinas locais, tornando necessário o uso de servidores. Já o *HiSat* leva menos tempo de processamento (~20 min), requer pouca memória, apenas 4 GB. Sendo assim, o *HiSat*, atualmente a melhor opção para a etapa de montagem de dados de RNA-Seq.

*TopHat* foi implementado no 2009 (TRAPNELL et al., 2009), e teve sua segunda versão, *TopHat2*, implementada no 2013 (KIM et al., 2013). Trata-se de uma ferramenta gratuita, de grande confiabilidade, eficiência, e aceitação na montagem de transcriptomas que possuem um genoma de referência. Em conjunto, *TopHat* e *TopHat2*, possuem mais de 12.000 citações, sendo, 805 (*TopHat*) e, 962 (*TopHat2*) no ano 2018 (mecanismo de pesquisa Google acadêmico, data 01-08-2018).

Após o mapeamento das leituras no genoma de referência, um dos objetivos mais frequentes é a comparação estatística dos níveis de abundância de genes e transcritos em diferentes condições biológicas. Esta comparação recebe o nome de análise da expressão diferencial. Ela envolve a quantificação de cada gene ou transcrito. Atualmente, existem varios métodos de quantificação de genes e transcritos, os quais utilizam a contagem do número de leituras sequenciadas que pertencem a um determinado gene ou transcrito, como uma medida de abundância, ou nível de expressão.

Diversos programas e pacotes de *software* têm sido desenvolvidos para realizar as

contagens. Em geral, estes métodos podem ser divididos em duas categorias: *union exon- and transcript-based*. Pela sua simplicidade, programas tais como, *featureCounts* (LIAO; SMYTH; SHI, 2014) e *HTSeq-count* (ANDERS; PYL; HUBER, 2015), os quais estão baseados no método de quantificação *union exon*, são amplamente usados. A aproximação *transcript-based* é intrinsecamente mais difícil, já que diferentes isoformas de um mesmo gene possuem uma alta proporção de sobreposição genômica (do inglês, *genomic overlap*). Porém, um gene é expresso em uma ou mais isoformas, tornando o método *transcript-based*, biologicamente, mais relevante (ZHANG et al., 2017; ZHAO; XI; ZHANG, 2015).

Métodos de quantificação *transcript-based* podem ser conduzidos utilizando-se as ferramentas *Cufflinks* (TRAPNELL et al., 2012), *RSEM* (LI; DEWEY, 2011) e, *BitSeq* (GLAUS; HONKELA; RATTRAY, 2012). Estes programas oferecem métodos robustos de quantificação, porém, requerem muito tempo de processamento.

Mais recentemente, foram implementados os métodos de quantificação *transcript-based* *Sailfish* (PATRO; MOUNT; KINGSFORD, 2014), *RapMap* (SRIVASTAVA et al., 2016), *kallisto* (BRAY et al., 2016) e *Salmon* (PATRO et al., 2017). Estas ferramentas trabalham mediante a quebra das leituras em sequências menores, chamadas de *K-mers*, as quais são pareadas a transcritos pré-indexados. Estes métodos são chamados de métodos *alignment-free* já que não usam o alinhamento clássico das leituras com a sequência de referência. Em vez disto, são usados algoritmos mais sofisticados como o *pseudoalignment*, *lightweight mapping* ou o *quasi-mapping*. Estas características tornaram estes métodos muito mais rápidos, eficientes e acurados na estimativa da abundância.

No entanto, a quantificação de isoformas é muito mais complexa devido ao alto grau de sobreposição entre transcritos, e atualmente não existe um consenso em relação ao método de quantificação, e às análises complementares mais apropriadas (SONESON; LOVE; ROBINSON, 2015; WU et al., 2018; ZHAO; XI; ZHANG, 2015). Uma vez estimada a abundância dos genes, estes dados são integrados em programas desenvolvidos para conduzir a análise de expressão diferencial de genes. Para este objetivo, também vários programas têm sido desenvolvidos, os quais utilizam diferentes métodos estatísticos. Estes métodos podem ser classificados em dois grupos: paramétricos e não paramétricos.

Métodos paramétricos permitem prever um valor desconhecido, a partir da observação de um modelo e os seus parâmetros. Quando métodos paramétricos são aplicados na expressão diferencial de genes, se assume que cada valor de expressão para um determinado gene é mapeado a uma distribuição particular, tais como Poisson ou binomial negativa (COSTA-SILVA; DOMINGUES; LOPES, 2017). Por outro lado, métodos não paramétricos podem capturar mais detalhes relacionados à distribuição dos dados, já que estes modelos não são limitados por um modelo rígido, e a distribuição dos dados não são definidos por um conjunto finito de parâmetros, portanto, a quantidade de informação relacionada os dados pode incrementar com seu volume (COSTA-SILVA; DOMINGUES; LOPES, 2017).

Ferramentas como *edgeR* (ROBINSON; MCCARTHY; SMYTH, 2010), *baySeq* (HARDCASTLE; KELLY, 2010) e *DESeq* (ANDERS; HUBER, 2010; LOVE; HUBER; ANDERS, 2014) utilizam a distribuição binomial negativa como seu principal método de análise. Entretanto, ferramentas como *NOIseq* (TARAZONA et al., 2015) e *SAMseq* (LI; TIBSHIRANI, 2013) adotaram métodos não paramétricos. Outros programas tais como, *EBSeq* (LENG et al., 2013) e *Cuffdiff* (TRAPNELL et al., 2013) baseiam-se em métodos que foram desenvolvidos para a quantificação e análise de expressão mais no nível de transcritos ou isoformas do que no nível de genes.

Dentre esta variedade de programas, *DESeq*, *edgeR* e *Cuffdiff* são as três ferramentas mais usadas. Enquanto, *DESeq* e *edgeR* são conhecidas por ter uma melhor performance no controle de falsos positivos. *Cuffdiff* é uma das ferramentas mais apropriadas para a análise diferencial no nível de isoformas. *Cuffdiff* é menos conservativo, permitindo encontrar um alto número de verdadeiros positivos, porém, um número maior de falsos positivos podem estar presentes no resultado final (JIA et al., 2015; TRAPNELL et al., 2013; ZHANG et al., 2014). No entanto, não há um consenso no que refere à metodologia mais apropriada para validar os resultados de análise diferencial das abundâncias, em termos de acurácia, robustez e reprodutibilidade (COSTA-SILVA; DOMINGUES; LOPES, 2017; JIA et al., 2015).

TRAPNELL et al. (2012) desenvolveram a *pipeline* formada por duas ferramentas, *TopHat* e *Cufflinks*, que em conjunto resolvem as 3 etapas essenciais na análise de dados de RNA-seq: (i) mapeamento das leituras no genoma de referência (ii) anotação de transcritos, e (iii) quantificação e análise de expressão diferencial de

genes e transcritos. *TopHat* alinha as leituras no genoma e descobre *splice sites*. *Cufflinks* cruza este mapa contra o genoma para montar as leituras nos transcritos. *Cuffdiff*, como parte de *Cufflinks*, toma as leituras alinhadas a partir de dois ou mais condições e reporta os genes e transcritos que são diferencialmente expressos.

*Cufflinks* usa o método de contagem fragmentos por kilobase de transcrito por milhão de fragmentos mapeados (do inglês, *Fragments Per Kilobase of Exon Per Million Mapped Reads* FPKM). *Cuffdiff2* assume que a expressão de um transcrito em cada condição pode ser medida pela contagem do número de FPKMs. Portanto, a alteração nos níveis de expressão de um transcrito é medida pela comparação estatística da contagem dos fragmentos em cada condição.

*Cuffdiff2* usa um modelo de variabilidade na contagem de fragmentos dos transcritos, em função de sua expressão, e de sua estrutura de *splicing*. Primeiro, determina-se o grau de sobredispersão através do ajuste da variância da contagem dos fragmentos em função de uma média obtida das réplicas experimentais. Segundo, estima-se a medida de incerteza através do cálculo do grau de confiança que um determinado fragmento seja atribuído corretamente ao transcrito ao qual foi associado. Transcritos com mais *exons* compartilhados e com alguns fragmentos exclusivamente atribuídos terão um maior grau de incerteza. Finalmente, combina-se a incerteza obtida em cada contagem dos fragmentos dos transcritos com a sobredispersão prevista para cada contagem.

A Incerteza da medição é calculada mediante um algoritmo que interpreta a contagem de fragmentos para um determinado transcrito, como uma distribuição beta. Já a sobredispersão na contagem é interpretada como uma distribuição binomial negativa. O mesmo algoritmo mistura as duas distribuições, para interpretar tudo como uma distribuição beta binomial negativa. Finalmente a alteração na expressão de cada gene e transcrito entre duas ou mais condições é reportada com sua correspondente significância estatística (valor  $p$ ) ajustados ao método estadístico *false discovery rate* (FDR) (BENJAMINI e HOCHBERG, 1995).

Em resumo, a metodologia do RNA-Seq integra a extração do RNA total, checagem da qualidade e integridade das amostras, preparação das bibliotecas de cDNA, sequenciamento, e análises de bioinformática. Os dados que são gerados,

analisados, e interpretados viabilizam o estudo e a inferência de mecanismos de regulação gênica em diferentes condições biológicas e/ou fisiológica.

O desenvolvimento dos sintomas da meleira somente após a floração sugere a hipótese de que a infecção pelo complexo PMeV desencadeia mecanismos moleculares durante a prefloração, os quais são alterados, ou diferentes daqueles que são desencadeados na pós-floração, em resposta à infecção. Um conhecimento detalhado sobre os conjuntos de genes envolvidos na resposta à infecção viral pode conduzir a novas conclusões relacionadas funcionamento de células hospedeiras durante a infecção, quais mecanismos de estresse e defesa são desencadeados e por que os sintomas da doença são desenvolvidos. Portanto, visando entender estes mecanismos, nesta tese de doutorado foi analisada a alteração no transcriptoma do mamoeiro em resposta à inoculação com o complexo PMeV em dois estágios de desenvolvimento: pré- e pós-floração.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Avaliar os genes diferencialmente expressos ao longo do desenvolvimento de *C.papaya* infectada pelo complexo PMeV que possam estar envolvidos nos possíveis mecanismos de defesa durante a prefloração e ao desencadeamento dos sintomas de meleira na pós-floração.

### 2.2 OBJETIVOS ESPECÍFICOS

- Identificar os genes diferencialmente expressos baseados em comparações pares (do inglês, *pairwise comparisons*) nas duas condições controle versus tratamento em duas idades diferentes (4 e 9 meses pós-germinação);
- Classificar os genes diferencialmente em categorias funcionais e conduzir análises estatísticas como o enriquecimento de categorias funcionais;
- Analisar o envolvimento dos genes diferencialmente expressos em vias e processos relacionados à defesa;
- Validação experimental dos resultados obtidos nas análises de expressão diferencial de genes;
- Relacionar e discutir os resultados no contexto do fenômeno observado quanto ao desenvolvimento dos sintomas de meleira somente após a floração;
- Propor possíveis alvos “hubs” e/ou mecanismos que possam ser usados para obstruir o desencadeamento dos sintomas de meleira e a obtenção de um mamoeiro melhorado geneticamente.



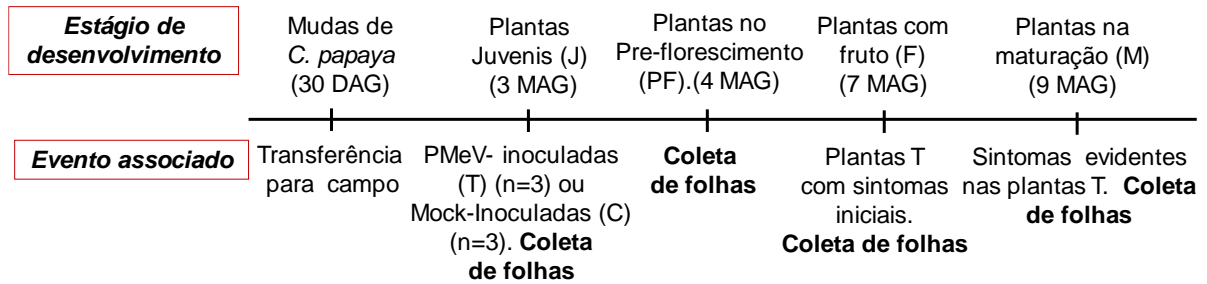
### 3 MATERIAIS E MÉTODOS

#### 3.1 DESENHO EXPERIMENTAL E OBTENÇÃO DO MATERIAL VEGETAL

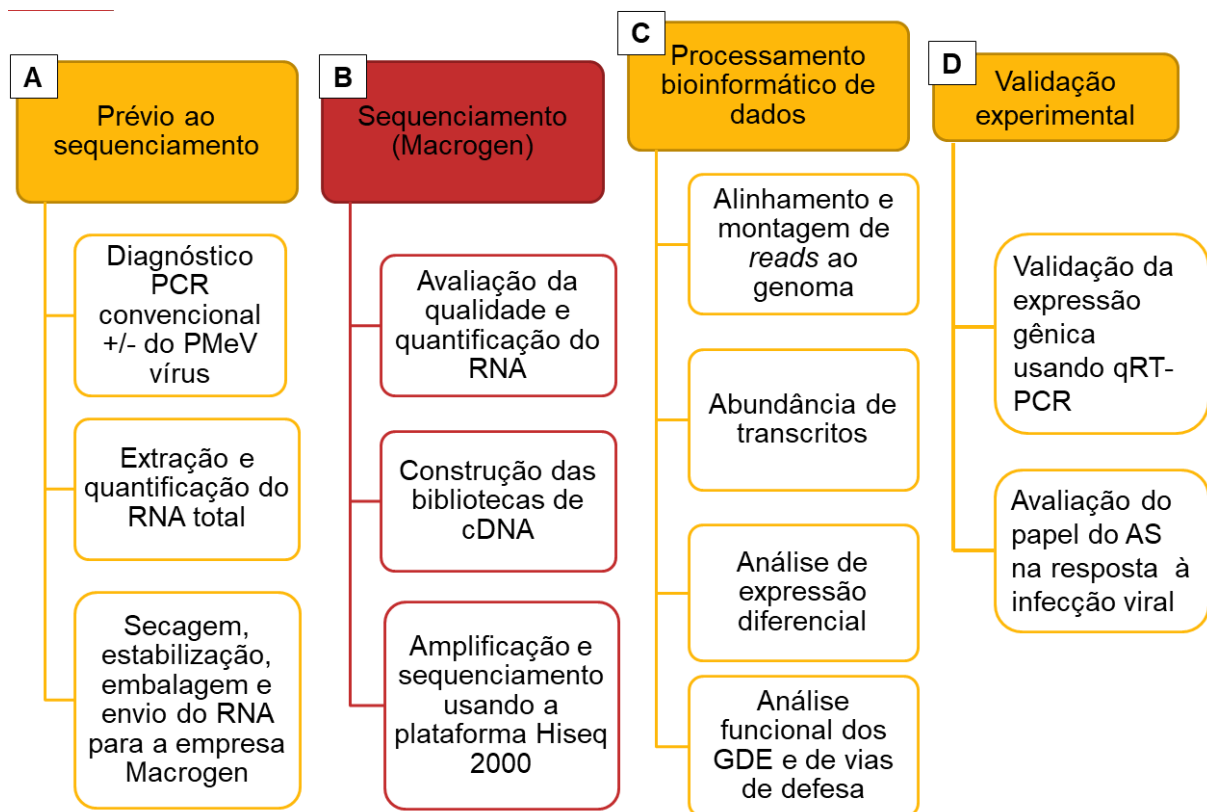
O desenho e o resumo da estratégia experimental que foram seguidos no desenvolvimento desta tese de doutorado são mostrados na Figura 8 e Figura 9, respectivamente. As etapas prévias ao sequenciamento e o sequenciamento conduzido pela empresa MacroGen (Seoul, South Korea) (Figura 9) encontram-se detalhados na dissertação de mestrado intitulada "Análise Transcriptômica da interação mamoeiro-Papaya Meleira Virus" (MADROÑERO, 2014).

Brevemente, mudas de *C. papaya* cultivar 'Golden', com aproximadamente 30 dias após germinação (DAG) foram plantadas no campo na Fazenda Experimental do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER), localizada no município de Sooretama, ao Norte do estado do Espírito Santo. Após dois meses, um grupo de plantas (n=3) foi inoculado no ápice do caule com 20 µL de uma suspensão (1:1, v/v) de látex coletado a partir de frutos com meleira tamponado em fosfato de sódio 50 mM, pH 7.0 (tratamento (T)), e em outro grupo de plantas (n=3) foi injetado unicamente 20 µL do tampão fosfato (controle (C)).

As plantas permaneceram no campo durante mais de nove meses, e diferentes coletas de material vegetal, a partir de folhas, foram feitas ao longo do desenvolvimento, como descrito na Figura 8. As amostras foram congeladas em nitrogênio líquido e armazenadas a -80 °C até o uso.



**Figura 8.** Desenho experimental e linha temporal. Mudanças de *C. papaya* com aproximadamente 30 dias após germinação (DAG) foram transferidas para uma estação experimental em condições de campo. Após dois meses, um grupo de plantas em estágio juvenil (J) com aproximadamente 3 meses após germinação (MAG) (n=3) foi inoculado no ápice do caule com 20 µl de uma suspensão (1:1, v/v) de látex coletado a partir de frutos com meleira tamponado em fosfato de sódio 50 mM, pH 7.0 (tratamento (T)), e em outro grupo de plantas (n=3) foi injetado unicamente 20 µl do tampão fosfato (controle (C)). As plantas permaneceram no campo durante mais de nove meses, e coletas a partir de material vegetal derivado da folha foram feitas ao longo do desenvolvimento nos estágios juvenil (J), prefloração (PF) aos 4 MAG, frutificação (F) 7 MAP e na maturação do fruto (M) aos 9 MAG, nas plantas T e C.



**Figura 9.** Estratégia experimental. A estratégia experimental utilizada é resumida em quatro etapas: (A) Experimentos prévios ao sequenciamento e após a obtenção do material biológico. (B) Metodologia utilizada para o sequenciamento das amostras na empresa MacroGen. (C) Processamento bioinformático dos dados obtidos no sequenciamento. (D) Validação dos dados obtidos usando técnicas complementares e abordagens fisiológicas. Em amarelo estão descritas as etapas que foram desenvolvidas no laboratório de Biotecnologia aplicada ao agronegócio, e em vermelho a etapa executada por um serviço contratado.

### 3.2 EXTRAÇÃO DO RNA E SEQUENCIAMENTO

O RNA foi isolado a partir de ~100 mg de tecido folhar usando o *RNAeasy plant mini kit* (Qiagen GmbH, Hilden, Germany). A quantidade e controle da qualidade foram determinados usando um *NanoDrop 2000 spectrophotometer* (Thermo Scientific, NanoDrop products, Wilmington, DE, USA) e um *Agilent 2100 Bionalyzer* (Agilent Technologies, Santa Clara, CA, USA). Amostras de RNA com uma quantidade superior a 4 µg, e uma relação entre RNAr 28S:18S e *RIN* (RNA integrity number) próximo ou superior a 1,5 e 7,0, respectivamente, foram usadas para a construção das bibliotecas individuais de ADNc. As bibliotecas de cDNA foram subsequentemente sequenciadas em ambas as extremidades usando a plataforma *Illumina HiSeq 2000* (Macrogen, Seoul, South Korea), e foram produzidas leituras *paired-end* de ~101 nt.

### 3.3 MAPEAMENTO DE LEITURAS OBTIDAS POR RNA-Seq NO GENOMA DE REFERÊNCIA

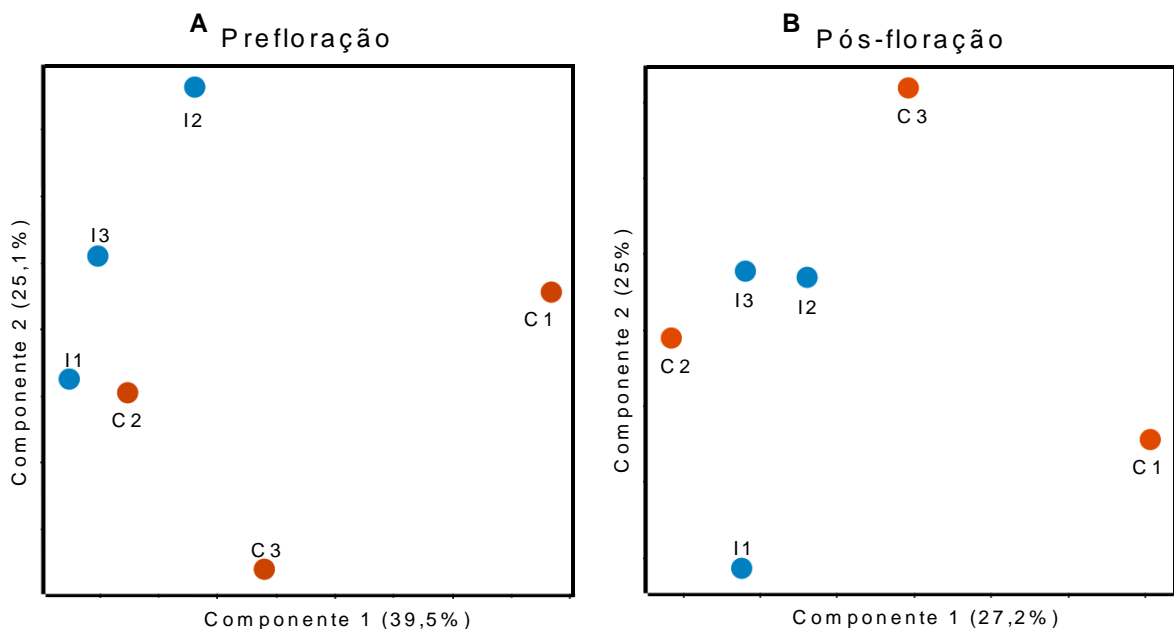
As leituras obtidas no sequenciamento (dado bruto em formato fastq) foram processadas usando o *software* FastQC v0.10.0 (BABRAHAM INSTITUTE, [s.d.]). Somente as leituras que apresentaram qualidade alta (*i.e.* *Phred quality score*  $\geq 20$  -  $\geq Q20$ ) foram utilizadas na montagem. A montagem do transcriptoma foi feita usando como referência o genoma de *C. papaya* v. 9.0 (MING *et al.*, 2008), o qual foi obtido da base de dados *Phytozome* (GOODSTEIN *et al.*, 2012). O genoma foi obtido a partir do sequenciamento de uma planta feminina do cultivar transgênico 'SunUp', e possui 370 Mega pares de bases (Mb) organizados em *contigs* e *scaffolds*.

A montagem das leituras (217 milhões de leituras) foi feita usando o *software* *TopHat* v.1.3.3 (TRAPNELL; PACHTER; SALZBERG, 2009) e o alinhador *Bowtie* v.1.1.0 (LANGMEAD, 2010). O baixo consumo de memória requerido pelo programa *TopHat* permitiu fazer a montagem usando um computador equipado com processador intel core I5, memória RAM 8 GB, disco rígido (HD) de 750 GB, levando apenas um tempo aproximado de 38 horas.

Os arquivos produzidos após a montagem (arquivos .BAM) foram utilizados como *inputs* para estimar a abundância relativa de transcritos usando o pacote *cuffnorm* disponível no *software Cufflinks v.2.0.2* (TRAPNELL et al., 2012). Esta tarefa foi concluída em aproximadamente duas horas. O *output* do *cuffnorm* são arquivos .TSV contendo a lista de genes ou transcritos e a suas respectivas abundâncias normalizadas em FPKMs.

### 3.4 ANÁLISE DA EXPRESSÃO DIFERENCIAL DE GENES (EDG)

A análise de componentes principais (PCA), que é uma análise exploratória de dados que fornece informação relacionada à similaridade das réplicas biológicas usadas em cada condição. Esta análise que houve uma alta variabilidade biológica entre as plantas, durante a pré- (Figura 10A), e a pós-floração (Figura 10B), principalmente entre as plantas que não foram inoculadas com o complexo PMeV (C).



**Figura 10.** Análise de componentes principais (PCA). (A) Réplicas biológicas no estágio de prefloração dentro das duas condições avaliadas: plantas inoculadas (I) em cor azul, e plantas não inoculadas ou controle (C), em cor laranja. (B) Réplicas biológicas no estágio de pós-floração dentro das duas condições avaliadas: plantas inoculadas (I) em cor azul, e plantas não inoculadas ou controle (C), em cor laranja.

Esta alta variabilidade entre as réplicas influencia a análise estatística de expressão diferencial de genes. Em consequência, quando são usados programas muito restritivos no controle de falsos positivos, um número muito baixo de genes são considerados diferencialmente expressos. Portanto, o pacote *cuffdiff*, o qual é menos restritivo foi usado na análise de expressão diferencial de genes.

Comparações por pares foram feitas entre os dois grupos, plantas controle (n=3) e plantas inoculadas (n=3) nos estágios de desenvolvimento: (1) prefloração, quando o botão floral foi formado aos 4 meses após germinação (MAG) (Figura 8 e Figura 11); (2) pós-floração (9 MAG), quando as plantas tiveram o fruto em fase de maturação (Figura 8 e Figura 12). Foram considerados como diferencialmente expressos os genes que obtiveram um valor p ajustado a uma FDR (*False Discovery Rate*)  $\leq 0,05$  e um  $\log_2$  |Fold change|  $\geq 0,5$  (FC).





**Figura 11.** Plantas de *C. papaya* no estágio de prefloração (4 MAG) antes da coleta de folhas para a extração de RNA e próximas análises. Plantas com botão floral (ampliado na Figura 10-1) que previamente (3 MAG) tinham sido inoculadas no ápice do caule com 20  $\mu$ l de uma suspensão (1:1, v/v) de látex coletado a partir de frutos com meira tamponado em fosfato de sódio 50 Mm, pH 7.0 (tratamento (T)) plantas 1-3, em que foram injetadas unicamente 20  $\mu$ l do tampão fosfato (controle (C)) plantas 4-6. Fotos tiradas pelo grupo de Biotecnologia aplicada ao agronegócio na Fazenda Experimental do INCAPER no município de Sooretama, estado do Espírito Santo.





**Figura 12.** Plantas de *C. papaya* no estágio de pós-floração (9 MAG) antes da coleta de folhas para a extração de RNA e próximas análises. Plantas com o fruto no estágio de maturação que foram inoculadas no ápice do caule com 20  $\mu$ l de uma suspensão (1:1, v/v) de látex coletado a partir de frutos com meleira tamponado em fosfato de sódio 50 Mm, pH 7.0 (tratamento (1-3 (T)), e plantas que foram injetadas unicamente com 20  $\mu$ l do tampão fosfato (4-6) (controle (C)). Fotos tomadas pelo grupo de Biotecnologia aplicada ao agronegócio na Fazenda Experimental do INCAPER no município de Soretama, estado do Espírito Santo.

### 3.5 CLASSIFICAÇÃO FUNCIONAL DOS GDE E ANÁLISE DE VIAS DE RESPOSTA A DEFESA

As análises de classificação funcional de genes diferencialmente expressos em plantas infectadas pelo complexo do PMeV e de envolvimento de vias de sinalização relacionadas à defesa foram realizadas usando principalmente o programa MapMan v. 3.6.0 (<http://mapman.gabipd.org>). Porém, uma alta porcentagem (aproximadamente o 15%) dos genes não foram designados a nenhuma categoria funcional. Por este motivo, as análises das descrições e anotações dos genes diferencialmente expressos foram complementadas utilizando-se o programa Blast2GO v. 2.7.1 (<https://www.blast2go.com>). Em alguns casos, foram manualmente corrigidas usando como referência as informações obtidas diretamente da base de dados Phytozome (GOODSTEIN et al., 2012) e do NCBI (*National center for Biotechnology information*).

As ontologias gênicas (do inglês, *Gene Ontology* -GO) foram obtidas usando a opção “GOslim” no Blast2GO, filtrando a classe funcional “plant”. A análise de enriquecimento das ontologias gênicas (do inglês, *GO enrichment analysis*) foi realizada usando o teste exato de Fisher, aplicando o método de correção de Bonferroni (FDR) (BENJAMINI; HOCHBERG, 1995). Os diagramas de Venn foram obtidos usando a ferramenta online “venny tool” (OLIVEROS, 2007).

### 3.6 VALIDAÇÃO DA EXPRESSÃO GÊNICA ATRAVÉS DE PCR EM TEMPO REAL (qRT-PCR)

Valores de expressão gênica foram obtidos para 22 genes do conjunto de dados “prefloração” e 9 genes do conjunto “pós-floração” (Tabela 2). Os ácidos nucleicos totais isolados a partir de aproximadamente 100 mg de tecido foliar de *C. papaya* usando solventes orgânicos foram precipitados com acetato de sódio 3 M, pH 5,2 e



etanol absoluto frio por ~12hrs, a -20 °C. A quantidade e qualidade do RNA foi determinada usando o aparelho NanoDrop 1000. Amostras de RNA (1 µg) foram tratadas com DNase I (Invitrogen, Carlsbad, CA, USA). A qualidade e quantidade foram novamente testadas.

As reações de transcrição reversa (RT) foram realizadas usando o kit “High Capacity cDNA” (Life Technologies, Grand Island, NY, USA). A mistura para a reação (~600 ng de RNA tratado com DNase, 2 µL de tampão 10X RT, 0.8 µL de dNTPs 100 mM, 2 µL de primers randômicos 10X RT, 1 µL de transcriptasa reversa MultiScribe™, 1 µL de inibidor de RNase e água ultrapura até completar um volume final de 20 µL) foi incubada sequencialmente a 25 °C por 10 min, 37 °C por 120 min e 85 °C por 5 min.

As misturas para a reação de PCR (1.33 µL do produto de RT, 5 µL de SYBR Green PCR Master Mix (Life Technologies, Grand Island, NY, USA), 1 µL de primer *forward* e *reverse* 10 µM e água ultrapura até completar 10 µL de volume final) foram incubadas sequencialmente (95 °C por 20 s, seguidos por 40 ciclos consecutivos a 95 °C por 3 s e 60 °C por 30 s).

A amplificação dos cDNA alvo foi monitorada usando um equipamento “7500 Fast Real-time PCR” (Life Technologies, Grand Island, NY, USA), e as curvas *melting* dos produtos de amplificação foram obtidos mediante a incubação sequencial das amostras a (95 °C por 15 s, 60 °C por 60 s e 95 °C por 15 s).

Os dados de RNA-seq direcionaram a seleção de três genes: (1) “translation initiation factor 4E” (*eIF4*, ETS\_17.108), (2) “band 7 family protein” (*B4FP*, ETS\_5.273), e (3) cyclophilin (*CYC*, ETS\_37.186), cujas abundâncias foram similares nas folhas de plantas de *C. papaya* tratadas e controle. A análise dos três genes escolhidos usando o algoritmo geNorm (VANDESOMPELE et al., 2002) resultou na escolha dos genes *eIF4* e *B4FP* como os mais adequados para a normalização dos dados de qRT-PCR.

A expressão gênica relativa comparando as plantas tratamento (n=3) e as plantas controle (n=3) foi calculada usando o método 2- $\Delta\Delta$ Ct (LIVAK; SCHMITTGEN, 2001). O teste estatístico t de Student foi aplicado para examinar a significância estatística

na expressão dos genes na pré- e pósfloração.

### 3.7 APLICAÇÃO EXÓGENA DO ACIDO SALICÍLICO

Para avaliar o efeito do tratamento com AS, plantas de *C. Papaya* com aproximadamente 2.5 meses de idade foram inteiramente pulverizadas durante sete dias consecutivos usando uma solução controle [etanol 0,05% (v/v)] (n=3) ou com uma solução de AS 1 mM diluído em etanol 0,05% (v/v) (n=3). Após este período pré-tratamento, as plantas foram inoculadas com látex coletado a partir de plantas de *C. papaya* exibindo os sintomas característicos da meleira, como foi descrito anteriormente. As plantas permaneceram em casa de vegetação até completarem aproximadamente os quatro meses de idade.

### 3.8 QUANTIFICAÇÃO ABSOLUTA DAS CÓPIAS DE RNA DOS VÍRUS PMEV E PMEV2

A acumulação de moléculas de RNA dos vírus PMeV e PMeV2 foi determinada nas plantas com aproximadamente quatro meses de idade usando PCR quantitativa em tempo real (qPCR). Primers específicos (Tabela 2) foram desenhados usando o programa Primer Express 3.0, e a amplificação do vírus foi monitorada em um equipamento de PCR 7500 Fast Real-time (Life Technologies, Grand Island, NY, USA).

Na quantificação do DNA viral, curvas padrão foram preparadas usando diluições seriais de plasmídeos de DNA contendo o fragmento do genoma do PMeV ou do PMeV2 (3 a  $3 \times 10^6$  cópias do genoma viral por reação). Curvas padrão foram obtidas usando a análise de regressão dos valores limiares do ciclo (cycle threshold -Ct) de cada uma das três replicatas biológicas em diluição em relação ao logaritmo da quantidade de DNA em cada diluição.

Na quantificação absoluta do número de moléculas de RNA viral, 50 ng do CDNA

total, obtido conforme descrito anteriormente, foram usados nas reações de PCR contendo 5 µL de SYBR Green PCR Master Mix (Life Technologies, Grand Island, NY, USA), 1 µL de cada um dos primers específicos para os vírus, com uma concentração de 10 µM e água ultrapura até um volume final de 10 µL. A reação de PCR foi incubada sequencialmente a 95 °C por 20 s, seguida de 40 ciclos consecutivos a 95 °C por 3 s e 60 °C por 30 s. As curvas de *melting* da amplificação do produto foram obtidas por incubação sequencial das amostras a 95 °C por 15 s, 60 °C por 60 s e 95 °C por 15 s), sendo que cada amostra foi analisada em duplicata técnica.

Para garantir que o tratamento com AS foi efetivo, a expressão relativa dos genes *PR1*, *CHIA* e *PR5* foi usada como um marcador de resposta ao AS, conforme anteriormente descrito na seção 3.6. O teste estatístico t de Student foi aplicado para examinar a significância estatística na diferença das cargas virais e expressão dos genes *PR1*, *CHIA* e *PR5*.

**Tabela 2.** Lista de genes de *C. papaya* selecionados para a validação da expressão gênica relativa usando qRT-PCR e primers para a quantificação de RNA viral do PMeV e PMeV2.

ID do gene <sup>1</sup>	Descrição <sup>2</sup>	Descrição curta do gene <sup>3</sup>	Primers qRT-PCR <sup>4</sup>
ETS_12.208 <sup>5</sup>	osmotin 34	<i>PR5</i>	AATGGGCCTTGTCCAGTTTTT CCACAGTTGCCAGAGTTGCA
ETS_1476.2 <sup>5</sup>	hevein-like prepro	<i>PR4</i>	GCGCTTTTTGCTCGACATG TGTCTCCTCCGCCAAGCTA
ETS_20.79 <sup>5</sup>	pathogenesis-related protein 1-like	<i>PR1</i>	CGCCAACCAACGCAAAG CGTAAGGACCACCGGAGTGT
ETS_120.1 <sup>5</sup>	glucan endo- -beta- basic isoform-like	<i>PR2</i>	ATCCCATTTCTTACCTCCCATG AGCAAACACCTATCTGAGCAG
ETS_19.44 <sup>5</sup>	wrky transcription factor 70	<i>WRKY70</i>	CCAATTCCTCCGATTCATACACA TGCCTCAAGCCTGACAAGCT
ETS_6164 <sup>5</sup>	salicylate o-methyltransferase	<i>BSMT1</i>	TTGCTCATCTGGACCGAACA CCGTGCAAGCTTGTCAACTG
ETS_1096.2 <sup>5</sup>	npr1/nim1-interacting protein 1	<i>NPR1-I</i>	TCCACTTGGATTCCGTCGTT TGAGAGGCGGTCTTCTGAACTC
ETS_61.5 <sup>5</sup>	NPR1/NIM1 like defence protein C terminal	<i>NPR1</i>	CGAGCAGAACTATGGATTTGA TTCAGGTGCTCCTCTTGGATTC
ETS_44.71 <sup>5</sup>	ethylene-responsive transcription factor 1b	<i>ERF1</i>	ATGAGAGGATCATCGGCGATT TCCCTCAGCGACTCCTCAA
ETS_1127.3	peroxidase	<i>POD</i>	GCCCTCAGGATGGAAGCAA TCGGAGTTGTCACGTCGAGAT
ETS_79.59	not available	<i>NA1</i>	AGCGGCTTGGGTTTCAAGT CCTGCTGCCCTGACTCCTT
ETS_8.191	glycine-rich cell wall structural protein 1	<i>GRP</i>	CACGCCACTTTAGCTGCTAGAA TCTGGTCATCGAGACCTTTGG
ETS_65.1	not available	<i>NA2</i>	TTGATCGAGATGATGAGGATGTCT GCCTGAGTGACGATATGACAACCTT
ETS_1999.2	calmodulin-binding receptor kinase 3	<i>CRCK3</i>	GTTACTACAGCAACCCACAATTTCTC TCCACCTGAGCCTTGTAACAGT
ETS_70.79	weak similarity to lip1 gene product gi	<i>LIP1</i>	GCCGGTGAGCCTGATATCTC ACATATTGGGCGCGTTTACAT
ETS_3.364	glutamine synthetase	<i>GLNA</i>	GCCTGGCCAGTGGGAATT TCATCACCGGCAGAAATGC
ETS_26.53	cyclophilin	<i>CYC 3</i>	TGGAGCTATTGCGGGATACG CGTGCACAGAGCCCTGAAG
ETS_37.186	Cyclophilin	<i>CYC</i>	TGGCTCGATGGGAAGCA TCACGTCCATGCCCTCAAC
ETS_49.84	peroxidase 12	<i>POD 12</i>	TGGACACCGTCGACATGGTA AGTGGCCAATTCCGATGGT
ETS_58.133	papaya latex serine protease inhibitor	<i>PR6</i>	CCTTGCAGCGATGTGGGTAT TCCTCTTAAAGTACCAGACGCCTAA
ETS_65.85	probable fructose-bisphosphate aldolase	<i>ALDO</i>	AACGTCCCGGATCGTCAA GGATCGGGACGACAGAAACTC
ETS_184.10	acidic endochitinase-like	<i>CHIA</i>	TCTCGCCGGCCACTGT AGCAAAGCGTATGCAGCTGTT
PMeV			CCCGATCCTGGTTTGAGAAA AGTTAATAGGAGCAGTCGCTAAC
PMeV2			GAAGGCTGTAGAGAAACCTGTC GCTGTCTCCTGGGCATTAAA

<sup>1</sup> ID do gene: Identificador do gene em *C. papaya* de acordo à base de dados Phytosome. ETS representa o nome curto do evm.TU.supercontig.

<sup>2</sup> Descrição do gene obtida após as análises usando o programa Blast2GO.

<sup>3</sup> Descrição curta do gene baseada nas anotações obtidas com o programa Blast2GO.

<sup>4</sup> Sequências dos primers usadas para os experimentos de qRT-PCR.

<sup>5</sup> Genes cujos valores de expressão gênica foram avaliados também na pós-floração.

## 4 RESULTADOS

### 4.1 ANÁLISE DO TRANSCRIPTOMA DE *C. papaya* USANDO RNA-Seq

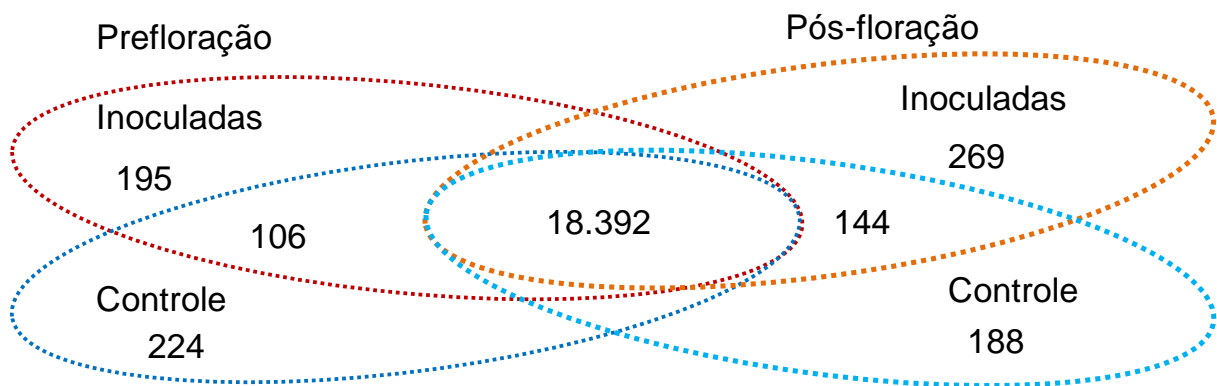
Plantas em estágio juvenil de *C. papaya* em condições de campo foram inoculadas com látex diluído em tampão fosfato obtido a partir de plantas mostrando os sintomas característicos de meleira. Como era esperado, as plantas durante a prefloração não mostraram nenhum dos sintomas de Meleira (Figura 11), enquanto que as plantas durante a pós-floração desenvolveram os sintomas (Figura 12). As plantas que foram inoculadas tanto na pré- como no pós-floração resultaram positivas para infecção com PMeV e PMeV2.

O transcriptoma de *C. papaya* foi analisado usando a tecnologia de sequenciamento RNA-seq em amostras de RNA (n=12) a partir de folhas de plantas que foram submetidas ao tratamento, e de plantas controle durante a pré (n=6) e a pós-floração (n=6). O sequenciamento gerou um total de 217.114.876 leituras pareadas (*paired-end reads*) com um tamanho de 101nt e com uma pontuação de qualidade Q20 (*Phred quality score 20*) (Tabela 3). 61,6% das leituras (133.732.106) foram mapeadas ao genoma de referência de *C. papaya*. As leituras foram alinhadas e mapeadas a 19.518 transcritos dos 27.796 genes preditos no genoma de *C. papaya* (PHYTOZOME, 2017), sendo 14.600 e 4918 com e sem anotação funcional, respectivamente (APÊNDICE A). Ao todo, 525 transcritos foram expressos unicamente durante a prefloração enquanto que 601 foram unicamente expressos durante a pós-floração (APÊNDICE B e Figura 13).

**Tabela 3.** Estatísticas do conjunto de dados produzidos no sequenciamento RNA-Seq.

Estágio	Condições das Plantas	Repetições Biológicas	Número de leituras com alta qualidade	Q20 (% leituras) <sup>1</sup>	Número de leituras mapeadas ao genoma de <i>C. papaya</i>	% de leituras mapeadas ao genoma de <i>C. papaya</i>
Prefloração	Controle	1	18414981	96,8	12147441	65,96
		2	18181510	96,8	10983014	60,41
		3	18519255	96,7	11532450	62,27
	Inoculadas	1	17899918	96,6	10879205	60,78
		2	17260243	96,6	10672029	61,83
		3	18900453	96,8	11433959	60,50
Pós-floração	Controle	1	18627050	96,8	11614220	62,35
		2	18055213	96,8	10719576	59,37
		3	17334426	96,8	10619436	61,26
	Inoculadas	1	18219651	97,0	11360444	62,35
		2	18339373	96,8	11185225	60,99
		3	17362803	96,9	10585107	60,96
		Média	18092906,33	<b>96,79</b>	11144342,17	<b>61,59</b>
		Total	<b>217114876</b>		<b>133732106</b>	

Q20(% leituras)<sup>1</sup>: Porcentagem de leituras que alcançaram uma pontuação de qualidade Q20 (Phred quality score 20) ou uma acúrcia do 99%.



**Figura 13.** Diagrama de Venn representando a distribuição e ocorrência dos transcritos nos estágios de pré- e pós-floração. As leituras mapearam para 19.518 dos 27.796 preditos no genoma de *C. papaya*. 18.392 foram encontrados em comum. 525 e 601 transcritos foram encontrados expressos unicamente na pré- e na pós-floração, respectivamente.

#### 4.2 ANÁLISE DA EXPRESSÃO DIFERENCIAL DE GENES (EDG) DURANTE A INFECÇÃO EM *C. Papaya* PELO COMPLEXO PMEV NA PRÉ- E PÓS-FLORAÇÃO

As análises de EDG em plantas infectadas pelo complexo do PMeV durante a prefloração resultaram em 633 genes (351 induzidos e 282 reprimidos) com uma alteração significativa (valores  $p$  ajustados a uma  $FDR \leq 0,05$ ) nos níveis de expressão  $Log_2 fold change$  (FC) quando comparados com as plantas controle (APÊNDICE C e Figura 14A). Três genes com funções desconhecidas mostraram os maiores níveis de acumulação (FC entre 5,8 e 5,9) nos tecidos infectados, seguidos por genes codificando a proteína relacionada à patogênese 1 (PR1) (FC= 4,7), uma heveína (PR4) (FC= 2,8) e uma esterase/lipase GDSL (FC=2,8). Em paralelo, genes codificando para uma proteína estrutural de membrana rica em glicina (GRP1) (FC=-6,3), uma proteína de função desconhecida (FC=-5,7) e um receptor quinase citoplasmático com domínio de ligação à calmodulina (FC=-4,9) apresentaram os níveis de expressão mais reprimidos nas plantas infectadas (APÊNDICE C).

Paralelamente, durante a pós-floração, nas plantas inoculadas com o PMeV e PMeV foram obtidos 88 genes diferencialmente expressos (44 induzidos e 44 reprimidos) (APÊNDICE D e Figura 14A). Enquanto que genes com os maiores níveis de acumulação codificaram para uma proteína com função desconhecida (FC=5,4),

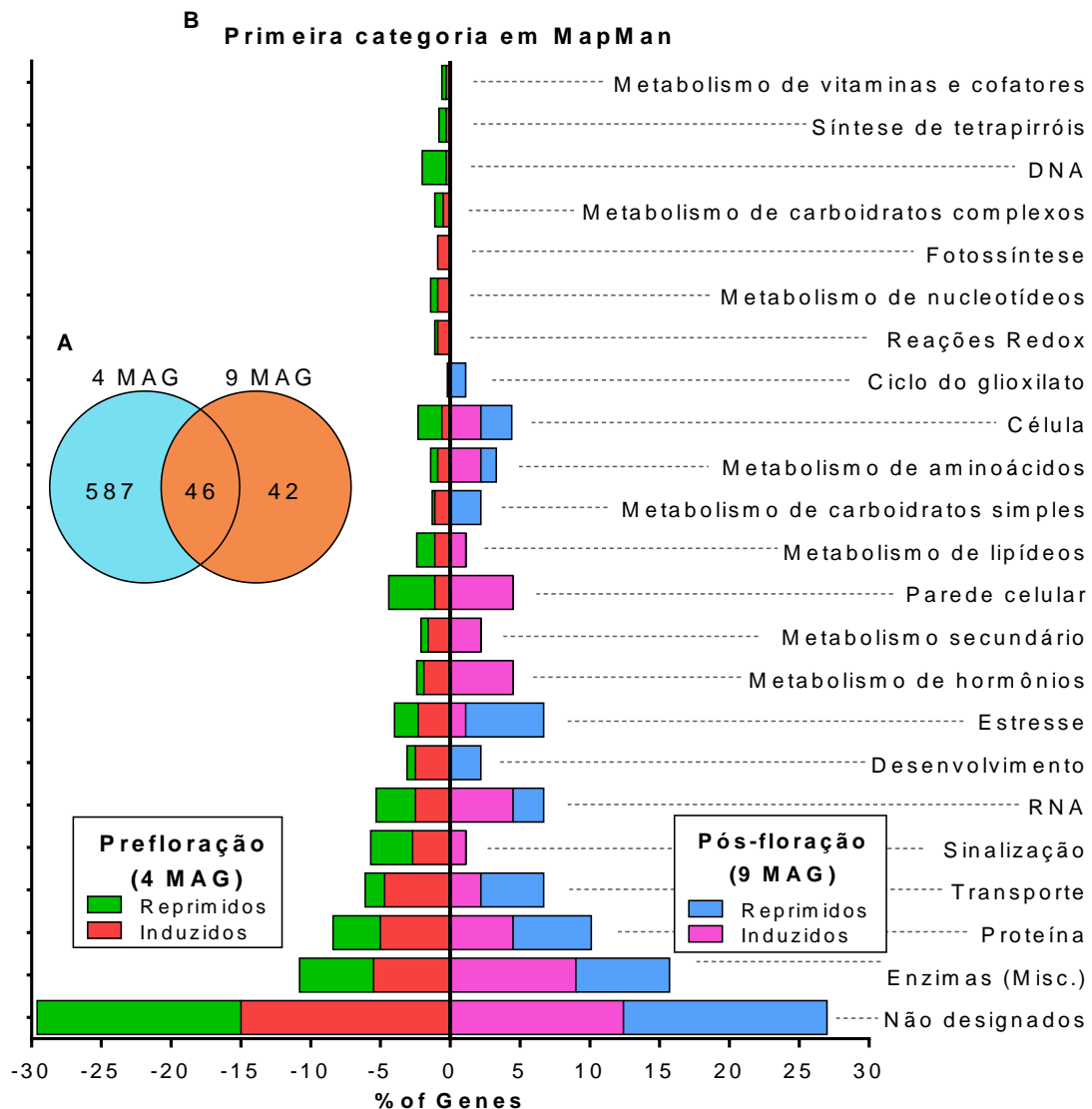
uma proteína estrutural de membrana rica em glicina (GRP1) (FC=4,0), uma fosfatase SGT1 (FC=3,8), uma proteína contendo o domínio *u-box* (FC=3,2), uma enzima metiltransferase do ácido salicílico (FC=2,2) e uma ciclina d3 (FC=2,0) (APÊNDICE D). Genes com os níveis de expressão mais reprimidos codificaram para um transportador de zinco (FC=-2,6), uma isomerase peptidil-prolyl (FC=-2,5), duas proteínas com função desconhecida (FC=-1,9 e 1,7), e uma proteína relacionada à patogênese 1 (PR1) (FC=-1,4) (APÊNDICE D).

#### 4.3 CLASSIFICAÇÃO FUNCIONAL DOS GDEs DURANTE A INFECÇÃO EM *C. papaya* PELO COMPLEXO PMEV NA PRÉ- E PÓS-FLORAÇÃO

Os genes diferencialmente expressos nas plantas infectadas durante a pré- e pós-floração foram atribuídos a 32 categorias funcionais disponíveis no programa MapMan usando as anotações do genoma de *C. papaya*. Como nove categorias foram representadas por baixos percentuais de genes, apenas as 23 categorias majoritariamente representadas são mostradas na Figura 14. Aproximadamente 15% de genes induzidos e reprimidos em ambos os estágios de desenvolvimento não foram atribuídos a uma categoria (Figura 14). Categorias tais como resposta ao estresse, desenvolvimento, metabolismo de carboidratos simples, reações redox, metabolismo de nucleotídeos e fotossíntese foram principalmente representadas por genes que foram induzidos durante a prefloração, e reprimidos ou não alterados durante a pós-floração (Figura 14).

Paralelamente, categorias como regulação de DNA e modificação/degradação de parede celular foram representadas principalmente por genes que foram reprimidos durante a prefloração e induzidos na pós-floração (Figura 14). Metabolismo de hormônios e metabolitos secundários foram principalmente representados por genes que foram induzidos em ambos os estágios de desenvolvimento (Figura 14).





**Figura 14.** Diagramas de relação entre os conjuntos de GDE e anotação funcional em categorias designadas pelo programa MapMan. (A) Diagrama de Venn mostrando a organização dos genes nas plantas infectadas pelo complexo do PMeV durante os estágios de prefloração (4 MAG) e pós-floração (9 MAG). Na prefloração foram exclusivamente expressos 587 genes, 42 unicamente expressos na pós-floração e, 46 foram expressos em comum nos dois estágios. (B) Porcentagem de genes induzidos (barras vermelhas) ou reprimidos (barras verdes) na prefloração, e induzidos (barras lilás) ou reprimidos (barras azuis) na pós-floração, designados pelo programa MapMan a 23 categorias funcionais.

Alternativamente, usando o programa Blast2GO, foi designado pelo menos um termo de ontologia gênica (do inglês, *Gene Ontology term* -GO term) a 418 (66,03%) dos 633 (APÊNDICE C) genes diferencialmente expressos nas plantas infectadas durante a prefloração. 20 termos GO resultaram significativamente ( $P \leq 0,005$ ) enriquecidos (Tabela 4). Três termos relacionados com transporte, e transporte em

membrana foram associados a 246 genes induzidos nas plantas infectadas, incluídos genes codificando transportadores de amônia, nitrato, fosfato, potássio, sódio, e açúcares (APÊNDICE C). Os outros 17 termos significativamente enriquecidos associados aos 172 genes reprimidos durante a infecção incluem genes codificando enzimas relacionadas ao metabolismo, especialmente hidrolases envolvidas na biologia de parede celular (Tabela 4 e APÊNDICE C).

As análises realizadas com Blast2GO nos genes diferencialmente expressos nas plantas infectadas na pós-floração resultaram na designação de, pelo menos, um termo GO em 78 (86,64%) dos 88 genes diferencialmente expressos (APÊNDICE D). Os treze termos GO significativamente enriquecidos ( $P \leq 0,05$ ) (Tabela 5) foram associados a genes induzidos envolvidos principalmente com metabolismo e desenvolvimento (Tabela 5).

**Tabela 4.** Análises de enriquecimento das ontologias gênicas (GO) nos GDEs nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a preflorescência.

ID do GO <sup>1</sup>	Descrição <sup>2</sup>	Classe do GO <sup>3</sup>	Nº genes reprimidos <sup>4</sup>	reprimidos (%) <sup>5</sup>	Nº genes induzidos <sup>6</sup>	Induzidos (%) <sup>7</sup>	FDR <sup>8</sup>	Valor p <sup>9</sup>
<b>Prevalentes na lista de genes reprimidos</b>								
GO:0071704	organic substance metabolic process	P	89	51,74	81	32,93	0,0247	0,00017
GO:0044238	primary metabolic process	P	86	50,00	73	29,67	0,0207	0,00004
GO:0016787	hydrolase activity	F	53	30,81	42	17,07	0,0824	0,00130
GO:0005515	protein binding	F	38	22,09	27	10,98	0,1431	0,00248
GO:0005975	carbohydrate metabolic process	P	23	13,37	13	5,28	0,2186	0,00449
GO:0016798	hydrolase activity, acting on glycosyl bonds	F	21	12,21	7	2,85	0,0247	0,00023
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compc	F	20	11,63	6	2,44	0,0247	0,00027
GO:0044424	intracellular part	C	16	9,30	6	2,44	0,1589	0,00301
GO:0043229	intracellular organelle	C	13	7,56	1	0,41	0,0207	0,00007
GO:0043226	organelle	C	13	7,56	1	0,41	0,0207	0,00007
GO:0032991	macromolecular complex	C	13	7,56	2	0,81	0,0247	0,00031
GO:0043234	protein complex	C	13	7,56	2	0,81	0,0247	0,00031
GO:0043231	intracellular membrane-bounded organelle	C	10	5,81	1	0,41	0,0566	0,00084
GO:0043227	membrane-bounded organelle	C	10	5,81	1	0,41	0,0566	0,00084
GO:0006259	DNA metabolic process	P	9	5,23	0	0,00	0,0247	0,00030
GO:0005634	nucleus	C	9	5,23	0	0,00	0,0247	0,00030
GO:0030246	carbohydrate binding	F	6	3,49	0	0,00	0,2186	0,00461
<b>Prevalentes na lista de genes Induzidos</b>								
GO:0005215	transporter activity	F	3	1,74	26	10,57	0,0247	0,0003
GO:0022857	transmembrane transporter activity	F	3	1,74	21	8,54	0,1431	0,0026
GO:0055085	transmembrane transport	P	2	1,16	23	9,35	0,0247	0,0003

<sup>1</sup>Identificador do termo de ontologia gênica.

<sup>2</sup>Descrição do termo de ontologia gênica.

<sup>3</sup>Categoria na qual são classificados os termos de ontologia gênica. P: processo Biológico, F: função molecular, C: componente celular.

<sup>4</sup>Número genes reprimidos que foram designados ao termo ontológico.

<sup>5</sup>Porcentagem de genes reprimidos que foram asignados ao termo ontológico.

<sup>6</sup>Número genes induzidos que foram designados ao termo ontológico.

<sup>7</sup>Porcentagem de genes induzidos que foram designados ao termo ontológico.

<sup>8</sup>Valores p ajustados a uma taxa de descobrimento falso (false discovery rate -FDR)  $\leq 0,5$ .

<sup>9</sup>Valores p obtidos após as análises usando o teste exato de Fisher com um valor de corte de 0,005.

**Tabela 5.** Análises de enriquecimento de ontologia gênica nos GDEs nas plantas de *C. papaya* infectadas pelo complexo do PMeV durante a pós-floração.

ID do GO <sup>1</sup>	Descrição <sup>2</sup>	Classe do GO <sup>3</sup>	No genes reprimidos <sup>4</sup>	reprimidos (%) <sup>5</sup>	No genes induzidos <sup>6</sup>	Induzidos (%) <sup>7</sup>	FDR <sup>8</sup>	Valor p <sup>9</sup>
<b>Prevalentes na lista de genes Induzidos</b>								
GO:0009791	post-embryonic development	P	0	0,0	6	17,1	0,717	0,0095
GO:0006629	lipid metabolic process	P	1	2,6	8	22,9	0,717	0,0116
GO:0044710	single-organism metabolic process	P	1	2,6	8	22,9	0,717	0,0116
GO:0071704	organic substance metabolic process	P	9	23,7	18	51,4	0,717	0,0170
GO:0044238	primary metabolic process	P	9	23,7	18	51,4	0,717	0,0170
GO:0009719	response to endogenous stimulus	P	0	0,0	4	11,4	0,790	0,0481
GO:0016043	cellular component organization	P	0	0,0	4	11,4	0,790	0,0481
GO:0071840	cellular component organization or biogenesis	P	0	0,0	4	11,4	0,790	0,0481
GO:0044767	single-organism developmental process	P	1	2,6	6	17,1	0,790	0,0497
GO:0007275	multicellular organismal development	P	1	2,6	6	17,1	0,790	0,0497
GO:0032502	developmental process	P	1	2,6	6	17,1	0,790	0,0497
GO:0032501	multicellular organismal process	P	1	2,6	6	17,1	0,790	0,0497
GO:0044707	single-multicellular organism process	P	1	2,63	6	17,14	0,790	0,0497

<sup>1</sup>Identificador do termo de ontologia gênica.

<sup>2</sup>Descrição do termo de ontologia gênica.

<sup>3</sup>Categoria na qual são classificados o termos de ontologia gênica. P: processo Biológico, F: função molecular, C: componente celular.

<sup>4</sup>Número genes reprimidos que foram designados ao termo ontológico.

<sup>5</sup>Porcentagem de genes reprimidos que foram designados ao termo ontológico.

<sup>6</sup>Número genes induzidos que foram designados ao termo ontológico.

<sup>7</sup>Porcentagem de genes induzidos que foram designados ao termo ontológico.

<sup>8</sup>Valores p ajustados a uma taxa de descobrimento falso (false discovery rate -FDR)  $\leq 0,8$ .

<sup>9</sup>Valores p obtidos após as análises usando o teste exato de Fisher com um valor de corte de 0,05.

Quarenta e seis genes em comum foram modulados nas plantas infectadas pelo complexo PMeV durante os dois estágios de desenvolvimento analisados. Dentre estes, 39 apresentaram um perfil oposto nos níveis de expressão (Figura 15). Por exemplo, o transcrito codificando para uma proteína similar a PR1 foi induzida durante a prefloração e reprimida durante a pós-floração (Figura 15), e 587 e 42 genes foram observados modulados unicamente na pré- e pós-floração, respectivamente (Figura 15).

Gene ID	Log <sub>2</sub> FC 4 MAG	Log <sub>2</sub> FC 9 MAG	Descrição
ETS_20.79	4.71	-1.41	pathogenesis-related protein 1-like (PR1-like)
ETS_79.59	3.24	-1.70	not available
ETS_307.1	2.31	-2.65	ZIP Zinc transporter
ETS_10.202	2.02	-1.01	boron transporter 4-like
ETS_525.1	1.88	-1.08	peroxidase 21-like
ETS_9.83	1.52	-1.44	heterogeneous nuclear ribonucleoprotein a3 homolog 2-like
ETS_14.246	1.52	-1.34	auxin-induced protein 5ng4-like
ETS_179.2	1.43	-0.81	regulator of vps4 activity in the mvb pathway
ETS_56.74	1.38	-0.83	cytosolic class i small heat shock protein type 2
ETS_2.75	1.34	-0.91	glutamate decarboxylase
ETS_112.29	1.23	-0.82	dehydrodolichyl diphosphate synthase 2-like
ETS_49.140	1.20	-0.79	mlp-like protein 31/ Pathogenesis-related protein Bet v I family
ETS_16.74	1.18	-0.75	phosphate chloroplastic
ETS_5.13	0.86	-0.87	bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin-3-like
ETS_49.9	0.86	-0.87	calcium-binding ef-hand family
ETS_17.158	0.85	-1.23	protease inhibitor seed storage lipid transfer family protein (LTP)
ETS_55.94	0.80	-0.80	adp-ribosylation factor a1 f isoform 2
ETS_14.143	0.80	-0.92	ring u-box superfamily protein isoform 2/protein isoform 1
ETS_14.144	0.80	-0.92	ring u-box superfamily protein isoform 2/protein isoform 1
ETS_103.73	-0.69	0.84	bifunctional inhibitor lipid-transfer protein seed storage 2s albumin superfamily (LTP)
ETS_58.60	-0.76	0.80	golgin subfamily b member 1-like isoform x1
ETS_318.2	-0.77	0.85	rop interactive partner 5 isoform 2
ETS_84.25	-0.79	0.85	zinc finger ccch domain-containing protein 53-like isoform x2
ETS_2.89	-0.96	0.82	indeterminate -domain isoform 2
ETS_21.150	-1.11	1.23	probable isoaspartyl peptidase l-asparaginase 2
ETS_107.51	-1.14	0.91	myosin heavy chain-related isoform partial
ETS_151.20	-1.18	1.36	endoglucanase 17-like (Cell wall)
ETS_3.5	-1.20	1.51	abc transporter g family member 8-like
ETS_21.52	-1.27	1.69	gdsl esterase lipase at1g33811-like
ETS_761.1	-1.42	1.91	germacrene d synthase
ETS_541.5	-1.45	0.95	pollen ole e 1 allergen and extensin family protein
ETS_46.44	-1.56	1.41	kda proline-rich (LTP)
ETS_132.40	-1.60	1.50	gdsl esterase lipase at2g42990-like
ETS_131.7	-1.73	1.10	transcription factor bhlh87-like
ETS_34.142	-1.80	1.30	expansin-b3-like (Cell wall)
ETS_83.49	-2.08	1.28	o-acyltransferase wsd1-like
ETS_8.192	-2.76	3.40	not available
ETS_200.13	-3.16	2.01	cyclin d3
ETS_8.191	-6.34	4.00	Glycine rich protein family (Cell wall)

**Figura 15.** Diagrama de cores mostrando os níveis de expressão (*Log<sub>2</sub> Fold changes*) dos genes diferencialmente expressos em comum entre os estágios de pré- e pósfloração que apresentaram um perfil de expressão oposto em ambos os estágios de desenvolvimento. Em vermelho são representados os genes induzidos e em verde os genes reprimidos.

#### 4.4 ANÁLISE DE VIAS DE RESPOSTA DE *C. papaya* À INFECÇÃO PELO COMPLEXO PMEV NA PRÉ- E PÓS-FLORAÇÃO

A análise dos genes diferencialmente expressos atribuídos à via de estresse biótico, usando as anotações do genoma de *C. papaya* em MapMan e a análise das ontologias gênicas relacionadas ao estresse usando Blast2GO, mostraram principalmente a modulação de vários genes associados a vias que envolvem a sinalização pelo AS, tais como o sistema de detoxificação de espécies reativas de Oxigênio (ROS) (Tabela 6), sinalização e metabolismo do ET e AJ (Tabela 7), receptores quinase ricos em repetições de leucina (LRR-RK) (Tabela 8), estruturação e remodelamento de parede celular, e metabolismo e degradação de calose (Tabela 9).

**Tabela 6.** Genes envolvidos no mecanismo de ROS que foram diferencialmente expressos nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a pré- e o pós-floração.

ID do Gene <sup>1</sup>	Log <sub>2</sub> FC <sup>2</sup>	Descrição <sup>3</sup>	Descrição curta/Similaridade médio <sup>4</sup>
<b>Prefloração</b>			
ETS_1127.3	2,76	peroxidase	POD (71,35%)
ETS_42.87	2,58	cationic peroxidase 2	POD2 (79,15%)
ETS_525.1	1,88	peroxidase 21-like	POD21 (88,15%)
ETS_73.70	1,7	glutathione s-transferase-like	GST (82,15%)
ETS_56.64	1,67	tau class glutathione transferase gstu45	GST45 (78,45%)
ETS_56.63	1,67	tau class glutathione transferase gstu45	GST45 (79,6%)
ETS_44.56	1,59	glutathione s-transferase	GST (83,8%)
ETS_56.92	1,42	heme-binding-like protein chloroplastic-like	HBPC (86,75%)
ETS_9.178	1,14	glutaredoxin family protein	GRX (92,15%)
ETS_17.27	1,13	glutathione s-transferase zeta class family protein	GSTZ (82,5%)
ETS_18.114	1,03	glutathione s-transferase zeta class-like	GSTZ (85,45%)
ETS_5.286	0,99	thioredoxin h2-like	TRX (72,75%)
ETS_7.187	0,97	glutaredoxin-like protein	GRX (75,4%)
ETS_12.261	0,82	gdp-l-galactose phosphorylase 1-like	GDPG (87,4%)
ETS_148.23	0,79	phospholipid hydroperoxide glutathione peroxidase chloroplastic-like	GPOD (89,1%)
ETS_17.237	0,65	glutathione s-transferase u9-like	GSTU9 (76,15%)
ETS_50.116	-0,82	glutathione s-transferase f9-like	GSTF9 (87,85%)
ETS_49.84	-0,87	peroxidase 12	POD12 (77,4%)
ETS_2.138	-1,06	transmembrane ascorbate ferrereductase 3	ASC (73,25%)
ETS_152.14	-1,39	peroxidase 42-like	POD42 (96%)
<b>Pós-floração</b>			
ETS_525.1	-1,1	peroxidase 21-like	POD21 (88,15%)

<sup>1</sup>Identificador do gene em *C. papaya*, de acordo com a base de dados Phytozome. ETS representa o nome curto do evm.TU.supercontig.

<sup>2</sup>Log<sub>2</sub> da razão nos níveis de expressão (fold change -FC) das plantas tratamento em relação com as plantas controle.

<sup>3</sup>Descrição do gene baseada nas anotações obtidas com o programa Blast2GO.

<sup>4</sup>Descrição curta do gene/similaridade médio entre as sequências alvo e o gene de *C. papaya* obtido após as análises usando o programa Blast2GO.

**Tabela 7.** Genes envolvidos na sinalização e metabolismo do ET e JA que foram diferencialmente expressos nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a pré- e a pós-floração.

ID do Gene <sup>1</sup>	Log <sub>2</sub> FC <sup>2</sup>	Descrição <sup>3</sup>	Descrição curta/Similaridade médio <sup>4</sup>
<b>Prefloração</b>			
ETS_44.71	2,38	ethylene-responsive transcription factor 1b	ERF1 (83.65%)
ETS_3.160	1,41	flavanone 3-dioxygenase-like	2OG-OXY (88.4%)
ETS_198.6	1,17	gibberellin 2-beta-dioxygenase 8-like	GA2OX8 (78.5%)
ETS_18.32	1,13	universal stress protein a-like protein	USP (82.45%)
ETS_2.209	1,1	1-aminocyclopropane-1-carboxylate oxidase	ACO4 (77.25%)
ETS_132.27	1,09	1-aminocyclopropane-1-carboxylate oxidase	ACO_1(95%)
ETS_29.56	0,74	probable aminotransferase acs10-like	ACS10 (81.6%)
ETS_152.58	0,65	1-aminocyclopropane-1-carboxylate oxidase	ACO_2 (94.95%)
<b>Pós-floração</b>			
ETS_458.4	1,9	linoleate 13s-lipoxygenase 2- chloroplastic-like	LOX2 (83.8%)
ETS_179.27	0,9	allene oxide cyclase chloroplastic-like	AOC3 (87.7%)
ETS_4.188	0,8	gibberellin 20-oxidase	GA20OX1 (83.2%)

<sup>1</sup>Identificador do gene em *C. papaya* de acordo com a base de dados Phytozome. ETS representa o nome curto do evm.TU.supercontig.

<sup>2</sup> Log<sub>2</sub> da razão nos níveis de expressão (fold change -FC) das plantas tratamento em relação com as plantas controle.

<sup>3</sup>Descrição do gene baseada nas anotações obtidas com o programa Blast2GO.

<sup>4</sup>Descrição curta do gene/similaridade médio entre as sequências alvo e o gene de *C. papaya* obtida após as análises usando o programa Blast2GO.

**Tabela 8.** Genes codificando receptores de membrana com repetições ricas em Leucina (LRR-RK) que foram diferencialmente expressos nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a preflorescência.

ID do Gene <sup>1</sup>	Log <sub>2</sub> FC <sup>2</sup>	Descrição <sup>3</sup>	Descrição curta/Similaridade médio <sup>4</sup>
<b>Preflorescência</b>			
ETS_54.61	-1,6	leucine-rich repeat receptor-like protein kinase at3g03770-like	LRR-RK (72,25%)
ETS_13.290	-1,31	leucine-rich repeat receptor-like protein kinase tdr-like	LRR-RK (86,35%)
ETS_106.88	-1,23	receptor protein kinase-like protein at4g34220-like	LRR-RK (70,2%)
ETS_106.88	-1,23	receptor protein kinase-like protein at4g34220-like	LRR-RK (70,2%)
ETS_6.204	-1,2	leucine-rich repeat family protein	LRR-RK (81,85%)
ETS_65.54	-1,12	leucine-rich repeat receptor-like protein kinase at1g68400-like	LRR-RK (84,8%)
ETS_5.100	-1,08	leucine-rich repeat receptor-like protein kinase at1g68400	LRR-RK (85,85%)
ETS_92.1	-1,05	leucine-rich repeat receptor-like protein kinase at5g06940-like	LRR-RK (85,8%)
ETS_21.161	-1,03	receptor kinase at5g67200-like	LRR-RK (78,95%)
ETS_21.162	-1,03	receptor kinase at5g67200-like	LRR-RK (86,1%)
ETS_85.116	-0,93	leucine-rich repeat receptor-like protein kinase at1g66830-like	LRR-RK (80,45%)
ETS_19.54	-0,77	leucine-rich repeat receptor-like protein kinase at3g28040-like	LRR-RK (85,5%)
ETS_126.17	-0,75	Irr receptor-like serine threonine-protein kinase erecta	LRR-RK (90,8%)
ETS_26.309	1,03	receptor protein kinase clavata1	LRR-RK (82,35%)
ETS_42.48	1,56	leucine-rich repeat receptor-like protein kinase family	LRR-RK (67,8%)

<sup>1</sup>Identificador do gene em *C. papaya* de acordo com a base de dados Phytozome. ETS representa o nome curto do evm.TU.supercontig.

<sup>2</sup>Log<sub>2</sub> da razão nos níveis de expressão (fold change -FC) das plantas tratamento em relação com as plantas controle.

<sup>3</sup>Descrição do gene baseada nas anotações obtidas com o programa Blast2GO.

<sup>4</sup>Descrição curta do gene/similaridade médio entre as sequências alvo e o gene de *C. papaya* obtida após as análises usando o programa Blast2GO.



**Tabela 9.** Genes codificando proteínas de parede celular e metabolismo de calose que foram diferencialmente expressos nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a pré- e o pós-floração.

ID do Gene <sup>1</sup>	Log <sub>2</sub> FC <sup>2</sup>	Descrição <sup>3</sup>	Descrição curta/Similaridade médio <sup>4</sup>
<b>Prefloração</b>			
ETS_1660.1	2,31	callose synthase 10	GSL10 (95,4%)
ETS_189.7	0,74	callose synthase 3	GSL3 (92,2%)
ETS_8.191	-6,34	glycine-rich cell wall structural protein 1	GRP (74%)
ETS_1461.3	-2,24	pectin methylesterase	PMES (82,65%)
ETS_102.19	-1,84	xyloglucan endotransglucosylase hydrolase protein 32	XTH32 (95,25%)
ETS_150.21	-1,81	endo- -beta-glucanase	GH (93,45%)
ETS_34.142	-1,8	expansin-b3-like	EXPB3 (89,1%)
ETS_25.194	-1,63	pectin methylesterase	PMES (84,8%)
ETS_3.250	-1,59	alpha-l-fucosidase 1-like	FUC1 (84,9%)
ETS_81.134	-1,38	expansin	EXP (92,65%)
ETS_0.14	-1,33	beta-xylosidase alpha-l-arabinofuranosidase 2-like	XYL2 (90,35%)
ETS_151.19	-1,31	pectinesterase pectinesterase inhibitor 40-like	PMESI40 (81,75%)
ETS_151.20	-1,18	endoglucanase 17-like	GH17 (92,3%)
ETS_69.65	-1,18	polygalacturonase at1g48100-like	PGLR (86,85%)
ETS_260.6	-1,12	expansin alpha	EXPA (85,5%)
ETS_180.28	-1,07	glycoside hydrolase family 28 protein	GH28 (88%)
ETS_146.54	-1,02	polygalacturonase at1g48100-like	PGLR (83,8%)
ETS_18.139	-0,92	fasciclin-like arabinogalactan protein 8	FLA8 (88,25%)
ETS_1.46	-0,92	alpha-expansin 1	EXPA1 (93,55%)
ETS_987.3	-0,85	fasciclin-like arabinogalactan protein 7 isoform 1	FLA7 (78,6%)
ETS_8.183	-0,83	pectin lyase-like superfamily protein	PECL (89,7%)
ETS_38.50	-0,82	exgt1 family protein	EXGT1 (91,1%)
ETS_12.146	-0,69	fasciclin-like arabinogalactan protein 1-like	FLA1 (84,55%)
ETS_42.121	-0,67	pectinesterase pectinesterase inhibitor 51-like	PMESI51 (81,65%)
ETS_117.87	0,75	nad-dependent epimerase dehydratase family protein	EPM (97,6%)
ETS_1540.1	0,8	pectinesterase pectinesterase inhibitor 12-like	PMESI12 (85%)
ETS_14.161	0,81	uridine diphosphate glucose dehydrogenase	UGD (97,2%)
ETS_198.15	0,81	pectinesterase pectinesterase inhibitor 18-like	PMESI18 (72,7%)
ETS_344.3	0,91	pectinesterase pectinesterase inhibitor 54-like	PMESI54 (79,9%)
ETS_129.52	0,94	xyloglucan endotransglucosylase hydrolase	XTH (92,05%)
ETS_2.137	1,18	xyloglucan endotransglycosylase family protein	XTH32 (88,75%)
<b>Pós-floração</b>			
ETS_8.191	4,0	glycine-rich cell wall structural protein 1	GRP (74%)
ETS_151.20	1,4	endoglucanase 17-like	GH17 (92,3%)
ETS_2.231	1,3	endoglucanase 4-related	GH4 (86,4%)
ETS_34.142	1,3	expansin-b3-like	EXPB3 (89,1%)
ETS_69.35	0,8	polygalacturonase 2	PGLR2 (78,45%)

<sup>1</sup>Identificador do gene em *C. papaya* de acordo com a base de dados Phytozome. ETS representa o nome curto do evm.TU.supercontig.

<sup>2</sup> Log<sub>2</sub> da razão nos níveis de expressão (fold change -FC) das plantas tratamento em relação com as plantas controle.

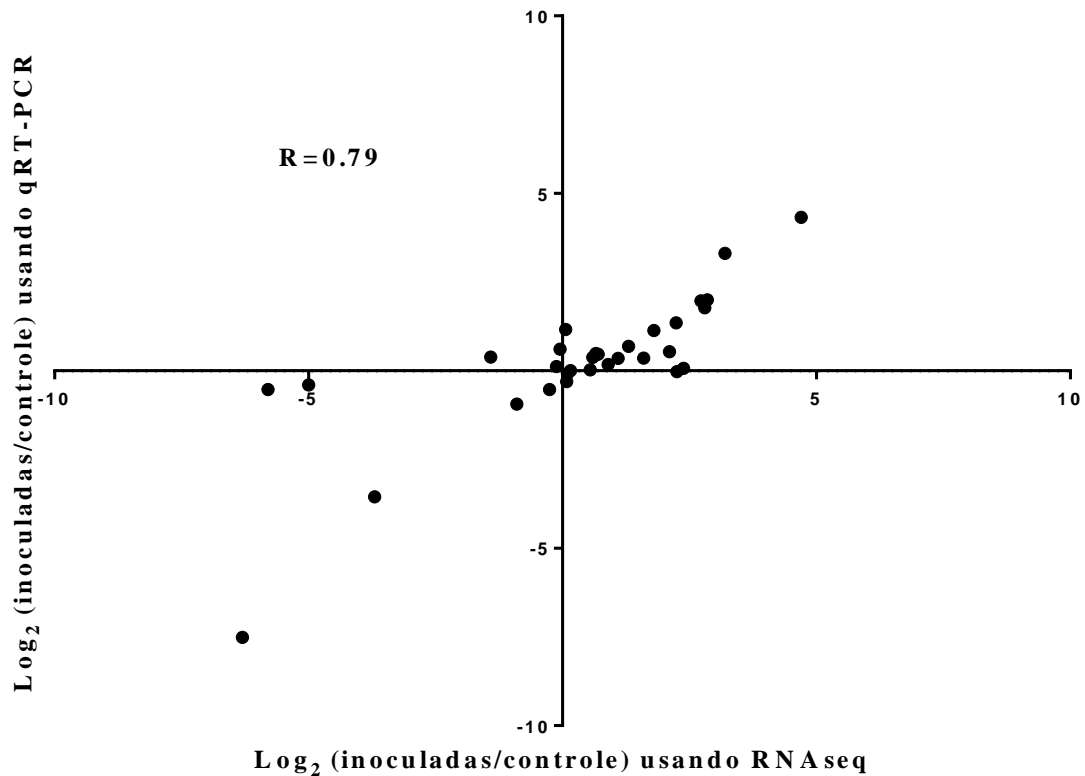
<sup>3</sup>Descrição do gene baseada nas anotações obtidas com o programa Blast2GO.

<sup>4</sup>Descrição curta do gene/similaridade médio entre as sequências alvo e o gene de *C. papaya* obtida após as análises usando o programa Blast2GO.

#### 4.5 VALIDAÇÃO DA EXPRESSÃO GÊNICA MEDIANTE PCR EM TEMPO REAL (qRT-PCR)

Os níveis de expressão gênica relativa de 22 genes na preflorescência (*POD\_1127.3*, *PR4\_1476.2*, *PR1\_20.79*, *NA1\_79.59*, *GRP\_8.191*, *NA2\_65.1*, *CRCK3\_1999.2*, *LIP1\_70.79*, *glnA\_3.364*, *CYC\_3\_26.53*, *CYC\_37.186*, *POD12\_49.84*, *PR6\_58.133*, *ALDO\_65.85*, *CHIA\_184.10*, *BMST1\_6.164*, *WRKY70\_19.44*, *ERF1\_44.71*, *NIMIN-1\_1096.2*, *NPR1\_61.5*, *PR5\_12.208*, *PR2\_120.1*) e, 9 genes na pós-floração (*BMST1\_6.164*, *WRKY70\_19.44*, *ERF1\_44.71*, *NIMIN-1\_1096.2*, *NPR1\_61.5*, *PR5\_12.208*, *PR4\_1476.2*, *PR1\_20.79*, *PR2\_120.1*) que foram induzidos, reprimidos ou inalterados na infecção pelo complexo PMeV (APÊNDICE C e D e, Tabela 2) foram testados usando qRT-PCR.

Os níveis de expressão relativos obtidos a partir das duas técnicas resultaram em uma correlação satisfatória, com um coeficiente de correlação de *Pearson*,  $R=0,79$  (Figura 16) (JIANG; WONG, 2009; WANG et al., 2010). Sendo que, os níveis de expressão de dois genes *NA2\_65.1* (FC=-5,8) e *CRCK3\_1999.2* (FC=-5,0) que foram observados fortemente reprimidos nas plantas que foram inoculadas durante a preflorescência, usando a técnica RNA-seq, não foram concordantes com os valores obtidos usando qRT-PCR.



**Figura 16.** Correlação entre os níveis de expressão relativa (Inoculadas/control) obtidos usando qRT-PCR e RNAseq. Comparação entre os níveis de expressão (Log<sub>2</sub> fold change) das plantas inoculadas em relação às plantas controle obtidos mediante duas técnicas diferentes, qRT-PCR e RNAseq.

#### 4.6 SINALIZAÇÃO MEDIADA PELO AS NA RESPOSTA DE *C. papaya* À INFECÇÃO PELO COMPLEXO PMEV

O grupo de genes envolvidos em sinalização e resposta ao AS inclui os marcadores clássicos de respostas mediadas por AS, *PR1*, *PR2* e *PR5*, um gene que codifica para a proteína monooxygenase 1 do flavin (FMO1), que é requerida para a acumulação sistêmica de AS em *Arabidopsis* (GRUNER et al., 2013), dois genes codificando a enzima glicosiltransferase dependente de UDP (UGT), e um gene codificando a proteína NPR1-I, envolvida com a diminuição da sinalização mediada por AS (Tabela 10) (GRUNER et al., 2013; VON SAINT PAUL et al., 2011; WEIGEL; PFITZNER; GATZ, 2005).

**Tabela 10.** Genes envolvidos na sinalização e resposta ao AS que foram diferencialmente expressos nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a pré- e a pós-floração.

ID do Gene <sup>1</sup>	Log <sub>2</sub> FC <sup>2</sup>	Descrição <sup>3</sup>	Descrição curta/Similaridade médio <sup>4</sup>
<b>Prefloração</b>			
ETS_20.79	4,71	Pathogenesis-related protein 1-like	PR1_1 (79.5%)
ETS_1476.2	2,86	hevein-like prepro	PR4 (85.4%)
ETS_12.208	2,72	osmotin 34	PR5_1 (84.35%)
ETS_184.10	1,78	acidic endochitinase-like	CHIB1 (81.6%)
ETS_120.1	2,11	glucan endo- -beta- basic isoform-like	PR2_1 (83.6%)
ETS_132.21	-1,81	glycosyl hydrolase family 17 family protein	PR2_2 (88.6%)
ETS_136.19	-1,44	glycosyl hydrolase family protein	PR2_3 (85.5%)
ETS_3.403	-2,86	thaumatin-like protein	PR5_2 (92.7%)
ETS_37.166	-1,76	pathogenesis-related thaumatin superfamily protein	PR5_3 (76.85%)
ETS_95.71	-1,34	Pathogenesis-related protein pr-1-like	PR1_2 (86.55%)
ETS_190.4	-1,05	osmotin-like protein	PR5_4 (85.15%)
ETS_106.99	1,25	limonoid udp-glucosyltransferase	UGT (80.8%)
ETS_8.229	0,69	udp-glucose flavonoid 3-o-glucosyltransferase	UGT (71.5%)
ETS_42.123	2,61	udp-rhamnose:rhamnosyltransferase 1-like	UGT (72.25%)
ETS_42.124	2,67	udp-rhamnose:rhamnosyltransferase 1-like	UGT (72.25%)
ETS_86.65	1,3	wrky transcription factor 70	WRKY70_1 (72.2%)
ETS_919.2	1,79	wrky transcription factor 51-like	WRKY51 (84.05%)
ETS_19.44	1,1	wrky transcription factor 70	WRKY70_2 (62.5%)
ETS_1096.2	2,24	npr1/nim1-interacting protein 1	NPR1-I (74.75%)
ETS_8.74	-1,75	lipid-transfer protein dir1-like	DIR1 (76.6%)
ETS_20.191	0,69	probable flavin-containing monooxygenase 1	FMO1 (84.05%)
<b>Pós-floração</b>			
ETS_20.79	-1.40	Pathogenesis-related protein 1-like	PR1_1 (79.5%)
ETS_1205.1	-1,1	udp-glycosyltransferase 74e2-like	UGT (65.4%)
ETS_6.164	2,2	salicylate o-methyltransferase	BSMT1 (82.25%)

<sup>1</sup>Identificador do gene em *C. papaya* de acordo com a base de dados Phytozome. ETS representa o nome curto do evm.TU.supercontig.

<sup>2</sup>Log<sub>2</sub> da razão nos níveis de expressão (fold change -FC) das plantas tratamento em relação com as plantas controle.

<sup>3</sup>Descrição do gene baseada nas anotações obtidas com o programa Blast2GO.

<sup>4</sup>Descrição curta do gene/similaridade médio entre as sequências alvo e o gene de *C. papaya* obtida após as análises usando o programa Blast2GO.

A sequência do gene codificando para o NPR1-I foi analisada com a ferramenta *blast* no banco de dados *GeneBank* e mostrou homologia com outras proteínas que interagem com NPR1, incluindo a proteína jcNIM1-I de *Jatropha curcas*. Similar à proteína NRR (NPR1-I) em arroz (CHERN et al., 2005), a proteína NPR1-I em *C. papaya* apresentou as sequências LDL-K/N-L e K-K/R-X-K/R, as quais evocam os motivos EAR (*ERF-associated amphiphilic repression*) (OHTA et al., 2001) e NLS (nuclear localization sequence) (CHERN et al., 2005) (Figura 17).

```

1
cpNPR1-I MKTEGMEHVKKVDGGSFNNGEGNRGEEGERGEEGDDEEKMEKFYALIRKRFREAREKRREELR
jcNIM1-I -----MENEKVDNGE--DEQEEQKVEEFFALIRSFQEARNRKRKQAL
rNRR      -MDATTTDATTAKRKRPAASDIADDAPT TVDEVSDAEVEEFYAILRRMRDAT--RRLGAR

61
cpNPR1-I ESEKKKMRRLNG-----EQSTWIPSFWRWEDFAQEIEFRRPPLIFPSPCNHKIDDDGKQQ
jcNIM1-I LEEKDRKKKTKLRKLNAAEQASWVPSFEWEDFNQEIQFRKSPLIYPRPCNNQNEKQKQ-
rNRR      PPPR-----APAWRPSFSWEDFADAPPKQAPP P P P P Q P A D H E R V A E N A T P P

115
cpNPR1-I HQSHKQQGADLDDAGLDLKLTL-----
jcNIM1-I -----EEDGGLDLNLTL-----
rNRR      R-----RPAPGLDLNVEPPSDAPATPR SARAPA

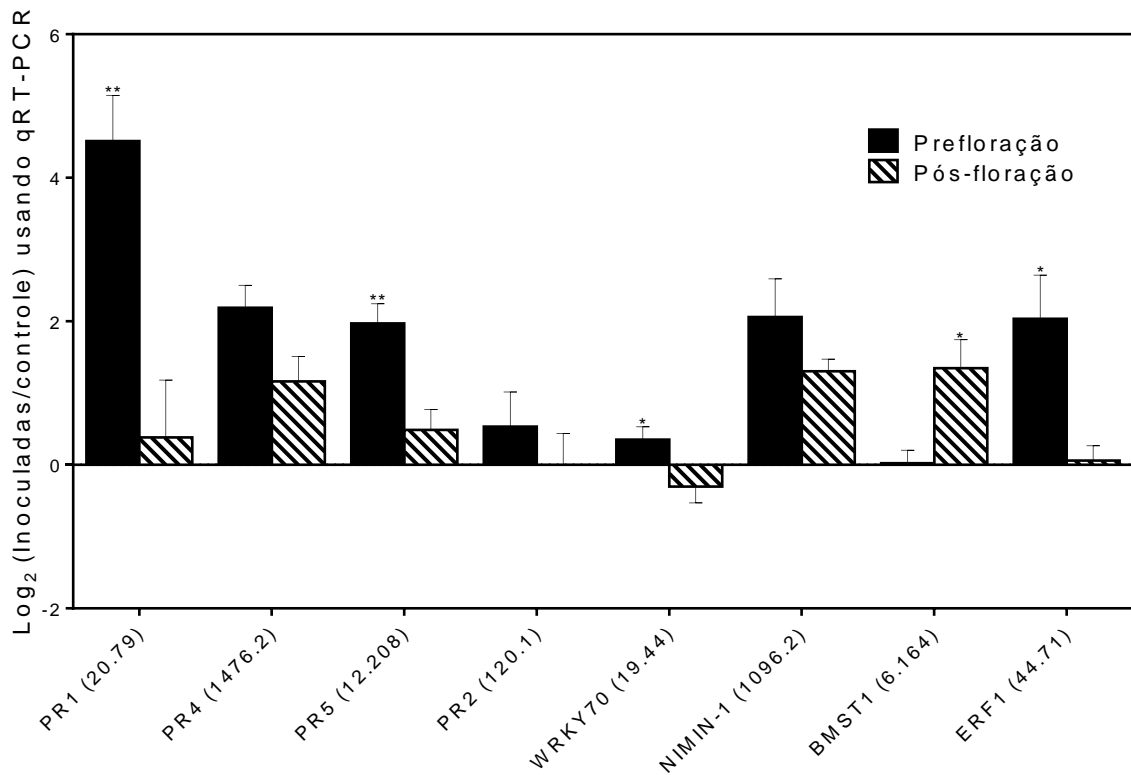
```

**Figura 17.** Alinhamento das seqüências das proteínas cpNPR1-I, jcNIM1-I e rNRR. A proteína que interage com NPR1 em *C. papaya* (cpNPR1-I, evm.TU.supercontig \_1096.2), *Jatropha curcas* (jcNIM1-I), e o regulador negativo de resistência em arroz (rNRR) foram alinhados usando CLUSTAW. Gaps presentes no alinhamento foram preenchidos com hifens. Aminoácidos idênticos foram sombreados. Os motivos putativos de localização nuclear (NLR) estão em negrito e sublinhados, enquanto que o motivo putativo de repressão anfifílica associado à ERF (EAR) está em negrito, sublinhado e sombreado.

Três transcritos codificando fatores de transcrição (TFs) WRKY (WRKY70\_1, WRKY70\_2 e WRKY51) (Tabela 10), que são responsivos ao AS, também foram acumulados nas plantas infectadas durante a prefloração, enquanto que nas plantas infectadas durante a pós-floração um transcrito codificando para a enzima metil transferase do ácido salicílico (BSMT1) foi induzido (Tabela 10). BSMT1 converte o AS em Salicilato de metilo (MeSA), prejudicando a sinalização de respostas de defesa mediada pela acumulação de AS (KOO et al., 2007; SESKAR; SHULAEV; RASKIN, 1998).

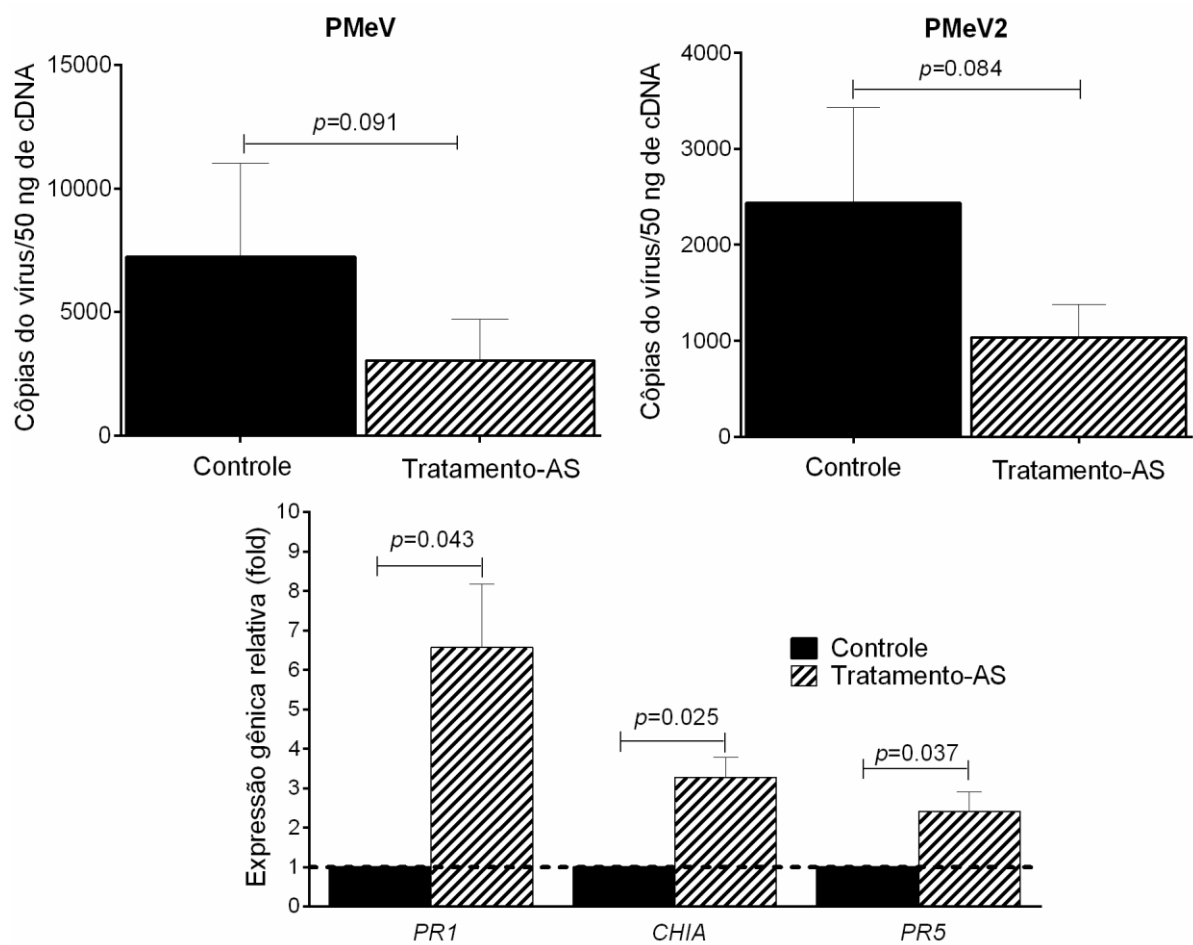
A avaliação da expressão relativa de nove genes em plantas infectadas com o complexo PMeV em relação às plantas controle mediante qRT-PCR mostrou a diminuição nos níveis de expressão de genes envolvidos em sinalização (WRKY70\_19.44) e resposta ao AS (PR1\_20.79, PR2\_120.1, PR4\_1476.2, PR5\_12.208) na pós-floração, em paralelo com o incremento nos níveis de expressão de genes que são reguladores negativos da sinalização por AS (BMST1\_6.164) quando comparados com a prefloração. Reguladores negativos da sinalização de AS, como ERF1\_44.71, NIMIN-1\_1096.2, foram também altamente induzidos na prefloração (Figura 18). Os valores de abundância relativa foram diferencialmente significativos ( $p \leq 0.05$ ) entre os estágios de pré- e pós-floração para os genes PR1\_20.79 ( $p = 0,0023$ ), PR5\_12.208 ( $p = 0,0038$ ), WRKY70\_19.44

( $p = 0,0464$ ), ERF ( $p = 0,0117$ ) e *BMST1\_6.164* ( $p = 0.0126$ ) (Figura 18).



**Figura 18.** Níveis de expressão de nove genes relacionados à via do AS nas plantas infectadas de *C. papaya*. Níveis de expressão relativa dos genes apresentados como o Log<sub>2</sub> da relação nos níveis de expressão (*fold change* -FC) das plantas tratamento e plantas controle na prefloração (barras em preto) e pós-floração (barras listradas paralelamente). Os valores de expressão gênica estão em log<sub>2</sub> (FC) ± EP (erro padrão). Um asterisco indica uma significância  $p \leq 0.05$ , enquanto que dois asteriscos indicam uma significância  $p < 0.005$ .

Para examinar o papel desempenhado pelo AS na resposta de defesa de *C. Papaya* contra a infecção do PMeV e o PMeV2, as plantas foram pulverizadas com uma solução de AS em etanol ou simplesmente com etanol. Sete dias após os tratamentos, estas foram inoculadas com látex coletado a partir de plantas com sintomas de meleira. A presença do vírus foi monitorada, e apesar de não ter obtido uma diferença significativa, os resultados mostraram uma tendência à diminuição das cargas virais do PMeV e do PMeV2 nas plantas que foram tratadas com AS quando comparadas com as plantas tratadas unicamente com etanol (Figura 19). Como era esperado, os genes *PR1*, *CHIA* e *PR5*, marcadores da resposta ao AS, aumentaram seus níveis de expressão nas plantas tratadas que foram tratadas com AS quando comparadas com as plantas controle (Figura 19).



**Figura 19.** Efeito da aplicação do AS na resposta de *C. papaya* à infecção pelo PMeV e PMeV2. Quantificação absoluta de cópias de RNA do PMeV e PMeV2 em plantas inoculadas com látex coletado a partir de plantas exibindo os sintomas característicos de Meleira. As plantas foram tratadas com uma solução controle [etanol 0,05% (p/v)] (n=3) (barras pretas) ou com uma solução de AS 1 mM diluído em etanol 0,05% (p/v) (n=3) (barras listradas). A expressão relativa do gene *PR1* foi usada com um marcador de resposta ao AS. Os dados são apresentados como as cópias do cDNA viral  $\pm$  SE (PMeV e PMeV2) e a expressão gênica relativa (valores fold)  $\pm$  EP (erro padrão). Os valores *p* mostrados na figura foram obtidos usando o teste t.

## 5 DISCUSSÃO

Neste trabalho foram analisadas as alterações no transcriptoma do mamoeiro durante a resposta à inoculação com o complexo PMeV. Interessantemente, os sintomas da meleira aparecem somente após a floração (RODRIGUES et al., 2009b). A expressão diferencial de genes e as vias regulatórias associadas a este fenômeno foram analisadas em mamoeiros inoculados com o complexo PMeV nos estágios de pré- e pós-floração.

No total, 61,6% das leituras foram mapeadas ao genoma de *C. papaya* e usadas para a construção do perfil de expressão gênica na interação entre *C. papaya* e o complexo PMeV. 70,2% (19.518) dos genes de *C. papaya* foram quantificados; dentre estes, 633, e 88 genes foram diferencialmente expressos nas plantas inoculadas durante a pré- e a pós-floração, respectivamente. Um coeficiente de correlação satisfatório ( $R=0,79$ ) entre os níveis de expressão gênica obtido usando RNA-seq e qRT-PCR validou a metodologia usada neste trabalho para a análise quantitativa do transcriptoma de *C. papaya*.

As respostas às infecções virais em interações compatíveis têm sido amplamente discutidas, porém, a maior parte destes estudos envolve a interação entre um vírus específico e uma planta modelo, comumente *Arabidopsis thaliana*. Por outro lado, pesquisas relacionadas com as respostas virais em plantas com um relativo ciclo de vida longo, quando comparado com *Arabidopsis*, tomate e tabaco, são limitadas, principalmente pelas dificuldades de manter as plantas em uma casa de vegetação.

*C. papaya* representa um bom modelo para o estudo de interações planta-patógeno em espécies frutais, devido às suas características, como organização e sequenciamento do genoma (MING et al., 2008), ciclo curto de reprodução e um sistema de transformação estabelecido (GSCHWEND et al., 2013; MING et al., 2008).

O complexo PMeV causa uma infecção sistêmica nas plantas de *C. papaya*, porém, os sintomas são induzidos tão somente após a floração (RODRIGUES et al., 2009b). Neste estudo, as vias regulatórias (Figura 20) associadas a este fenômeno foram



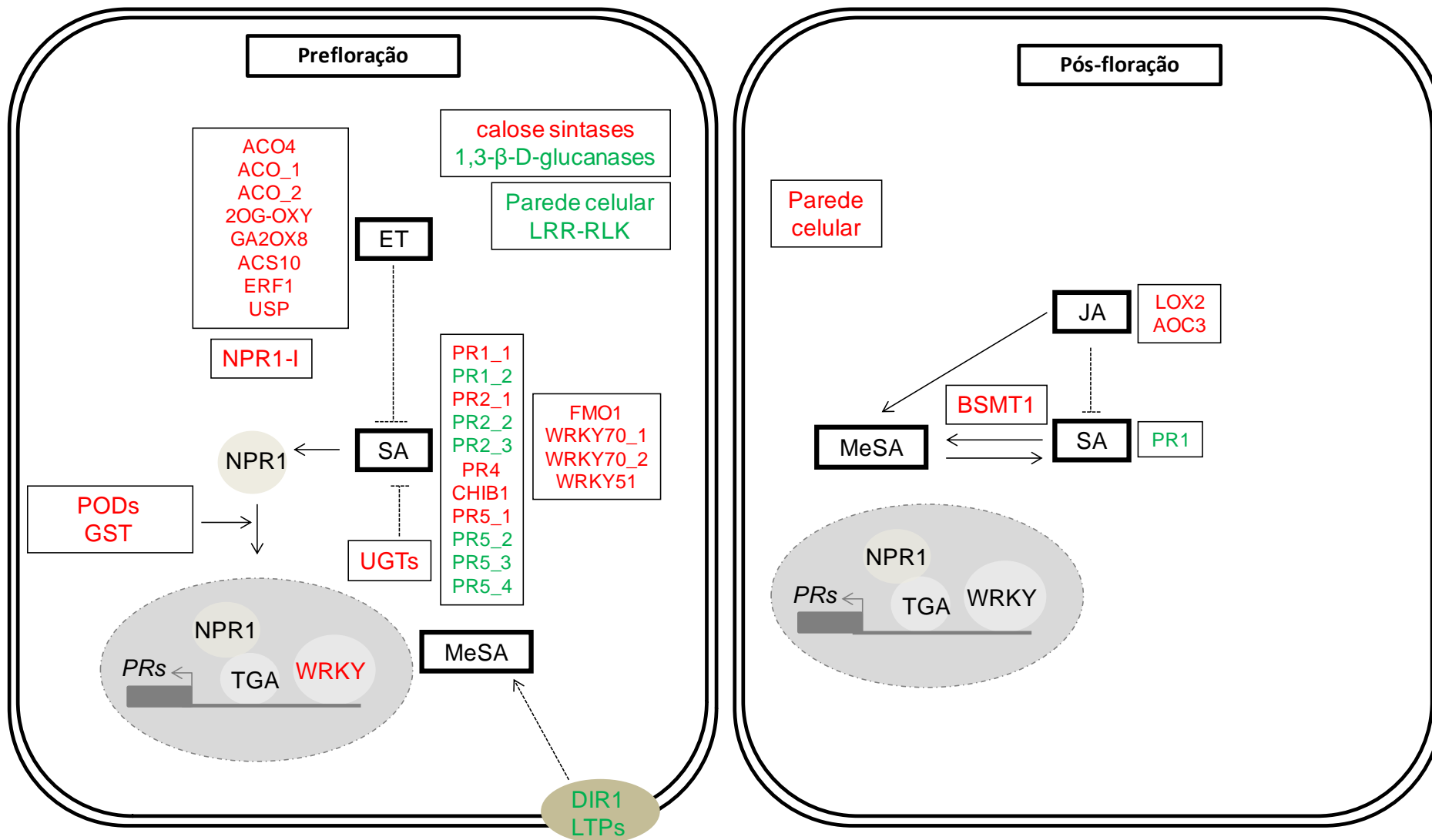
acessadas, principalmente ao nível transcriptômico, usando plantas de *C. papaya* infectadas pelo complexo PMeV na pré- (quatro meses de idade) e pós-floração (nove meses de idade) em condições de campo.

No estágio de prefloração foram observados indícios de uma resposta de resistência diante da infecção pelo complexo PMeV, incluindo a indução de genes que codificam para enzimas envolvidas na síntese de calose (Tabela 9) e desintoxicação de ROS (Tabela 6). Durante a interação planta-patógeno, no começo da infecção é desencadeada uma rápida acumulação das ROS que agem como moléculas sinalizadoras em respostas de defesa que podem induzir tolerância nas plantas infectadas (HANCOCK et al., 2002).

Em *C. papaya*, a infecção pelo complexo PMeV induz o aumento na produção de peróxido de hidrogênio ( $H_2O_2$ ) nos laticíferos (RODRIGUES et al., 2009a) e altera os níveis de expressão de miRNAs associados ao metabolismo de ROS em folhas de mamoeiro (ABREU et al., 2014). Adicionalmente, mudas de *C. papaya* inoculadas com o complexo PMeV imediatamente incrementam e mantêm uma alta atividade das enzimas peroxidase e superóxido dismutase (BUSS et al., 2011).

Consistente com o envolvimento de ROS associado com o fenômeno de tolerância ao desenvolvimento dos sintomas observados durante a prefloração, a expressão de genes codificando várias peroxidases e outras enzimas envolvidas no sistema de detoxificação de ROS foi induzida neste estágio, incluindo um dos genes (ETS\_525.1, Peroxidase 21-like) que foi reprimido durante a pós-floração (Tabela 6 e Figura 14B).

Paralelamente à modulação nos níveis de expressão de genes codificando enzimas envolvidas no sistema ROS, e similar ao observado em outras interações planta-vírus compatíveis e.g. *Solanum lycopersicum* e Tomato spotted wilt vírus (TSWV)/Citrus Exocortis Viroid (CEVd) (LÓPEZ-GRESA et al., 2016), *A. thaliana*/Nicotiana tabacum e Alfalfa mosaic virus (AMV) (APARICIO; PALLÁS, 2016), *Brachypodium distachyon* e Panicum mosaic virus (PMV)/Panicum mosaic satellite virus (SPMV) (MANDADI; SCHOLTHOF, 2012), as plantas infectadas pelo complexo PMeV na prefloração mostraram indícios de defesas mediadas por AS (Figura 20).



**Figura 20.** Modelo proposto para representar as vias envolvidas na resposta de *C papaya* à infecção pelo PMeV. Durante a preflorescência, quando as plantas permanecem assintomáticas, a percepção de partículas virais ou proteínas efetoras do PMeV e/ou PMeV2, provavelmente envolve receptores de membrana da família LRR-RLK ou transportadores de nutrientes de tipo *transceptors*, os quais, poderiam modular respostas de defesa, principalmente aquelas mediadas por ROS, deposição de calose e sinalização por AS. Textos em vermelho e verde indicam os genes induzidos e reprimidos, respectivamente. A indução de genes *PRs*, especialmente de *PR1*, é um marcador da ativação de vias de defesa mediadas por AS. A ativação completa de genes *PRs* envolve a interação entre fatores de transcrição da família TAG e a proteína NPR1. Os níveis incrementados de genes GSTs e peroxidases poderiam regular o balance redox, o qual é requerido para a translocação de NPR1, desde o citoplasma ao núcleo, onde interage com os fatores de transcrição TGAs e ativa a expressão de genes *PRs*. Os níveis reduzidos de transcritos codificando proteínas envolvidas em parede celular e proteínas associadas aos plasmodesmas, tais como 1,3-β-D-glucanases, assim como níveis incrementados de transcritos codificando para calose sintases, poderiam limitar a propagação dos vírus na planta. Os níveis de expressão aumentados em transcritos codificando proteínas envolvidas na síntese e metabolismo de ET, UGTs, e uma proteína que interage com NPR1 (NPR1-I), as quais poderiam prejudicar a sinalização de AS. Na pós-floração, quando os sintomas são evidentes, a repressão do gene *PR1* acompanhado da indução do gene *BSMT1* e genes relacionados ao metabolismo de JÁ, os quais favorecem a biossíntese de MeSA, poderiam estar envolvidos na diminuição da sinalização de vias de defesa mediadas por AS em *C. papaya*, levando finalmente ao desencadeamento dos sintomas.

Um transcrito similar ao gene relacionado a patogêneses 1 (*PR1*) esteve entre os transcritos com os maiores níveis de acumulação, assim como vários genes codificando outras proteínas *PRs* também foram notoriamente modulados nos tecidos infectados pelo complexo PMeV na preflorescência. A modulação de genes *PRs*, principalmente *PR1*, é considerada como um marcador para respostas mediadas por AS (RYALS et al., 1996). Além disso, as plantas de *C papaya* tratadas com AS mostraram uma tendência à diminuição da carga viral do PMeV e PMeV2, quando comparados com as plantas controle (Figura 19) no estágio de preflorescência. Estes resultados, em conjunto, sugerem que vias de defesa mediadas por AS poderiam operar no estágio de preflorescência.

Os níveis de expressão dos transcritos codificando para proteínas de transporte e proteínas relacionadas à defesa foram induzidos, enquanto que a expressão de transcritos codificando proteínas relacionadas a metabolismo e hidrolases, especialmente aquelas que estão envolvidas com a remodelação de parede celular, foi reprimida no estágio de preflorescência (Figura 15 e APÊNDICE C). Em *Arabidopsis*, uma reprogramação transcripcional similar foi mostrada ser dependente da proteína monooxygenase 1 do flavin (FMO1) durante a resposta de defesa SAR (GRUNER et al., 2013). Um gene codificando para uma proteína similar a FMO1 foi também induzido em *C. papaya* infectada pelo complexo PMeV na preflorescência (Tabela 10).

Os dados sugerem que a atividade da membrana também está envolvida com a resposta de *C. papaya* à infecção pelo complexo PMeV. Os termos GO relacionados com transporte/transmembranal foram significativamente ( $P \leq 0,005$ ) representados nos genes ativados durante a infecção na prefloração. De fato, as respostas celulares contra patógenos baseiam-se em receptores de membrana (AFZAL; WOOD; LIGHTFOOT, 2008; GÓMEZ-GÓMEZ; BOLLER, 2002; UEKI; CITOVSKY, 2014).

Receptores de membrana incluem uma grande subfamília de receptores quinase transmembranares ricos em repetições de leucina (LRR-RLK) (KEMMERLING, 2011). Os LRR-RKs são modulados em respostas a infecções virais tanto a nível transcricional como pós-traducional (COSTA et al., 2016; FONTES et al., 2004; MANDADI; SCHOLTHOF, 2012). Em geral, na prefloração, a maior parte de genes *LRR-RLK* foi reprimida nos tecidos infectados (Tabela 8). Este grupo de receptores (LRR-RLK) representam candidatos para efetores de imunidade em *C. papaya*.

Em paralelo, membros de famílias envolvidos em transporte de nutrientes, que agem como sensores-receptores em imunidade celular em plantas (GOJON et al., 2011), foram ativados durante a prefloração nas plantas infectadas (APÊNDICE C e Figura 15). Este grupo inclui genes que codificam para transportadores de amônia, potássio, sódio, fosfato, e sulfato (APÊNDICE C). Em *Arabidopsis*, a disrupção nos transportadores de nitrato NRT2.1 (CAMANES et al., 2012) e amônia AMT1.1 (PASTOR et al., 2014) melhora a resistência contra infecções bacterianas. Conseqüentemente, os níveis de expressão aumentados nos transportadores de nutrientes durante a infecção pelo complexo PMeV são consistentes com o perfil observado durante interações compatíveis.

O co-ativador transcricional NPR1/NIM1 é um regulador chave na ativação de defesas mediadas por AS (ZHANG et al., 1999). O fato de que *NPR1* é expresso constitutivamente, e os níveis de expressão não são dramaticamente aumentados após tratamentos com AS, sugere que *NPR1* é regulada ao nível de proteína (DONG, 2004). Fatores de transcrição TGA mediam a ativação de genes *PR* através da interação com *NPR1* (DONG, 2004; JOHNSON; BODEN; ARIAS, 2003; ZHANG et al., 1999).

Alternativamente, os fatores de transcrição WRKY reconhecem a caixa w (TTGACC/T), um elemento cis (do inglês, *Cis-regulatory element* -CRE), comumente presente nos reguladores do gene *PR1* (MALECK et al., 2000), sugerindo o envolvimento destes TFs na regulação de genes *PR* (DONG, 2004). Três genes *WRKY*, incluindo o *WRKY70*, foram acumulados nos tecidos infectados pelo complexo PMeV na prefloração em *C. papaya* (Tabela 10).

Linhagens S55 que sobreexpressam o gene *WRKY70* mostraram uma acumulação constitutiva de PRs, indicando que *WRKY70* é um ativador da expressão de genes *PRs* (LI; BRADER; PALVA, 2004). A indução de respostas de defesas mediadas por SA através de *WRKY70* acontece de uma maneira independente da via NPR1 (LUNA et al., 2014). Os altos níveis de expressão de *WRKY70* nas folhas de *C. papaya* infectadas com o complexo PMeV poderiam resultar na ativação de *PR1*, *PR4*, e de alguns genes *PR2* e *PR5* que foram encontrados induzidos nos tecidos infectados.

Estes resultados sugerem que respostas mediadas por AS, contrário à clássica resposta antiviral do silenciamento de RNA, cujos componentes, similar ao publicado por MANDADI; SCHOLTHOF, 2012, estiveram presentes, mas foram inalterados em nossos conjuntos de dados (APÊNDICES A, C, D), estão operando nas plantas de *C. papaya* infectadas pelo complexo PMeV e assintomáticas de meleira durante a prefloração.

Apesar dos níveis de expressão incrementados nos transcritos *PR1* e *WRKY70*, níveis de expressão reprimidos em alguns *PR2* e *PR5* foram observados (Tabela 10), sugerindo a ativação parcial de respostas mediadas pela sinalização de AS em folhas de *C. papaya* infectadas pelo complexo PMeV na prefloração.

As plantas também mostraram níveis incrementados na expressão de genes codificando para as proteínas NPR1-I/NIM1-I, UGTs, e proteínas envolvidas no metabolismo do ET (Tabela 7 e Tabela 10). As proteínas UGTs, as quais estão envolvidas na redução da acumulação e sinalização de AS (GRUNER et al., 2013; VON SAINT PAUL et al., 2011; WEIGEL; PFITZNER; GATZ, 2005), e a sinalização mediada por ET, que é associada na maioria das interações planta-vírus com um papel antagônico à sinalização de AS (ALAZEM; LIN, 2015), poderiam prejudicar a

ativação completa de defesas mediadas por AS durante a prefloração nas plantas infectadas.

A proteína NPR1-NIM1-I modula negativamente a sinalização de AS mediante sua interação com NPR1, danificando a ativação de respostas de defesa (WEIGEL; PFITZNER; GATZ, 2005). No arroz, a proteína NRR, que está estruturalmente relacionada à NIM1-I, contém sequências que evocam tanto o motivo EAR, que está envolvido com a repressão transcricional (OHTA et al., 2001), como o motivo NLS (CHERN et al., 2005), que interage com NPR1.

Adicionalmente, a superexpressão de NRR afeta a resistência relacionada à idade e aumenta a susceptibilidade a *Xanthomonas* (CHERN et al., 2005). A ocorrência de ambos os motivos EAR e NLS no gene *NPR1-I/NIM1-I* de *C. papaya* (Figura 17) sugere uma regulação negativa de NPR1-I sob a atividade de NPR1. Portanto, é razoável inferir que a ativação de *NPR1-I* seja continuada pela redução na atividade de NPR1, impedindo a ativação de defesas mediadas por NPR1, e prejudicando a indução completa de genes *PRs* através de WRKY70 (LI; BRADER; PALVA, 2004).

Genes codificando para membros da família de proteínas transportadoras de lipídios (LTP) que inclui elicitores de defesa em plantas e a proteína DIR1 (MALDONADO et al., 2002), que tem sido sugerida favorecer a acumulação da forma esterificada do AS, o Salicilato de metilo (MeSA) (LIU et al., 2011), foram encontrados reprimidos nas plantas infectadas durante a prefloração (APÊNDICE C). Estes resultados, confirmam nosso postulado de que durante a prefloração a infecção do complexo PMeV em *C.papaya* altera os níveis de AS, e portanto reduz a amplificação do sinal de AS na indução de respostas de defesa.

Na pós-floração, as plantas de *C. papaya* inoculadas com o complexo PMeV mostraram níveis reprimidos na expressão dos genes codificando a PR1, transportadores de zinco e boro e, uma peroxidase 21, contrário ao perfil observado durante a prefloração (Figura 14B). Paralelamente, transcritos codificando uma proteína rica em glicina (GRP), enzimas envolvidas na remodelagem da parede celular e uma ciclina d3 que foram reprimidas nas plantas infectadas com o complexo PMeV na prefloração, foram induzidos durante a pós-floração (Figura 14B).

Ciclo celular e processos associados ao DNA (Figura 15) são comumente alterados em interações compatíveis, sendo sugeridos serem manipulados pelos vírus para favorecer o processo infeccioso através da indução de genes associados à fase S e G do ciclo celular (ASCENCIO-IBÁÑEZ et al., 2008; JADA et al., 2014; PIERCE et al., 2013; SUN et al., 2016).

Endo-1,4- $\beta$ -glucanases e GRP (Tabela 9) são importantes para a estruturação da parede celular (MANGEON; JUNQUEIRA; SACHETTO-MARTINS, 2010). Alterações na parede durante a interação entre *C-papaya* e o complexo PMeV, que também têm sido observadas ao nível proteômico (SOARES et al., 2017) e ao nível morfológico (MAGAÑA-ÁLVAREZ et al., 2016), poderiam estar relacionados com o desencadeamento dos sintomas na pós-floração, especialmente associados à ruptura celular que leva à exsudação espontânea de látex atribuída ao incremento na absorção de água e à alteração na pressão interna dos laticíferos (RODRIGUES et al., 2009a).

Em contrapartida, já que as osmotinas são proteínas induzidas frente ao estresse osmótico (ANIL KUMAR et al., 2015) a modulação de genes PR5, especialmente aqueles que codificam para osmotinas (Tabela 10), poderiam ser uma resposta da planta para se proteger e combater uma possível estratégia de movimentação do vírus através dos laticíferos (RODRIGUES et al., 2009a).

As respostas de defesa mediadas pela sinalização de AS poderiam ser limitadas durante a pós-floração. As plantas mostraram níveis aumentados na expressão dos genes *LOX2* e *AOC* (Tabela 7), os quais estão envolvidos na síntese e sinalização de JA e afetam positivamente a biossíntese de MeSA (CHEN et al., 2003).

De fato, um gene codificando para uma proteína homóloga à carboxil-metil transferase do ácido salicílico (*BSMT1*) que converte AS em MeSA, foi induzido nas plantas inoculadas com o complexo PMeV durante a pós-floração (Tabela 10). Em plantas transgênicas, a superexpressão de um gene codificando uma metil transferase impede a ativação de genes de defesa e o desencadeamento de resistência (KOO et al., 2007; SESKAR; SHULAEV; RASKIN, 1998). Portanto, a indução de *BSMT1* nas plantas infectadas na pós-floração poderia estar relacionada com o desenvolvimento dos sintomas da meleira.

Exceto para a maturação, há uma escassez de informação relacionada aos grupos de genes e vias de sinalização que são ativadas em resposta ao desenvolvimento fisiológico em *C. papaya*. Apesar de terem sido preditos vários genes que estão associados ao desenvolvimento, poucos ou inexistentes estudos relacionados à pesquisa do perfil de expressão gênica nos diferentes estágios de desenvolvimento foram publicados até agora.

O desconhecimento dos processos que são alterados na planta como uma resposta às suas alterações próprias durante seu desenvolvimento dificulta uma melhor compreensão dos mecanismos moleculares que estão envolvidos no desencadeamento dos sintomas da meleira durante a pós-floração. Uma vez que temos mostrado que a infecção em *C. papaya* pelo complexo PMeV causa a alteração de genes relacionados a vários processos biológicos e ativa e modula a expressão de genes envolvidos em respostas de defesa, um dos desafios para futuras pesquisas é mostrar quais processos biológicos são ligados ou alterados em relação às distintas etapas do desenvolvimento de *C. papaya*, e estabelecer quais alterações que causadas pelo complexo PMeV são normalmente observadas em alguma etapa de desenvolvimento.

## 6 CONCLUSÕES

Estes resultados em conjunto confirmam a tese de que apesar de que vários processos são alterados durante a infecção, vias de defesa que são mediadas por AS estão ativas durante a prefloração e agem efetivamente evitando o desencadeamento dos sintomas da doença durante a prefloração. Contudo, a ativação de genes que antagonizam a sinalização de vias de defesa mediadas por AS, tais como *NPR1-I*, *ET* e *UGTs*, poderiam limitar a sua ativação completa e de longa duração. Este antagonismo é acentuado na pós-floração com a ativação de genes como *BSMT1* e genes envolvidos na síntese e metabolismo de AJ, como *LOX2* e *AOC*, o que finalmente leva ao desenvolvimento dos sintomas durante este último estágio.

Neste trabalho indicamos alvos que podem ser usados na modificação genética de



um mamoeiro que seja resistente à meleira, e para a aplicação de novas estratégias de manejo que podem vir a ser implantadas nas lavouras de mamão para reduzir as perdas com a doença.

O estudo do perfil de expressão gênica relacionado aos processos próprios do desenvolvimento fisiológico de *C. papaya* facilitará o entendimento dos processos e vias de sinalização que podem ser ou não alvos de modificação genética em *C. papaya*.

## REFERÊNCIAS

- ABREU, P. M. V et al. Carica papaya microRNAs are responsive to Papaya meileira virus infection. **PLoS ONE**, v. 9, n. 7, 2014.
- ABREU, P. M. V et al. A current overview of the Papaya meileira virus, an unusual plant virus. **Viruses**, v. 7, n. 4, p. 1853–70, 8 abr. 2015.
- AFZAL, A. J.; WOOD, A. J.; LIGHTFOOT, D. A. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. **Molecular plant-microbe interactions : MPMI**, v. 21, n. 5, p. 507–517, 2008.
- ALAMILLO, J. M.; SAÉNZ, P.; GARCÍA, J. A. Salicylic acid-mediated and RNA-silencing defense mechanisms cooperate in the restriction of systemic spread of plum pox virus in tobacco. **The Plant Journal**, v. 48, n. 2, p. 217–227, out. 2006.
- ALAZEM, M.; LIN, K.-Y.; LIN, N.-S. The Abscisic Acid Pathway Has Multifaceted Effects on the Accumulation of *Bamboo mosaic virus*. **Molecular Plant-Microbe Interactions**, v. 27, n. 2, p. 177–189, fev. 2014.
- ALAZEM, M.; LIN, N.-S. Roles of plant hormones in the regulation of host-virus interactions. **Molecular Plant Pathology**, v. 16, n. 5, p. 529–540, 2015.
- ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biology**, v. 11, n. 10, p. R106, 27 out. 2010.
- ANDERS, S.; PYL, P. T.; HUBER, W. HTSeq--a Python framework to work with high-throughput sequencing data. **Bioinformatics**, v. 31, n. 2, p. 166–169, 15 jan. 2015.
- ANIL KUMAR, S. et al. Osmotin: a plant sentinel and a possible agonist of mammalian adiponectin. **Frontiers in Plant Science**, v. 6, p. 163, 2015.
- APARICIO, F.; PALLÁS, V. The coat protein of Alfalfa mosaic virus interacts and interferes with the transcriptional activity of the bHLH transcription factor ILR3 promoting salicylic-dependent defense signaling response. **Molecular plant pathology**, p. n/a-n/a, fev. 2016.
- ARAÚJO, M. M. M. DE et al. Molecular detection of Papaya meileira virus in the latex of Carica papaya by RT-PCR. **Journal of Virological Methods**, v. 146, n. 1–2, p. 305–310, 2007.
- ARGUESO, C. T. et al. Two-Component Elements Mediate Interactions between Cytokinin and Salicylic Acid in Plant Immunity. **PLoS Genetics**, v. 8, n. 1, p. e1002448, 26 jan. 2012.
- ARUMUGANATHAN, K.; EARLE, E. Nuclear DNA content of some important plant species. **Plant Molecular Biology Reporter**, v. 9, n. 3, p. 208–218, 1991.
- ASCENCIO-IBÁÑEZ, J. T. et al. Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. **Plant physiology**, v. 148, n. 1, p. 436–54, set. 2008.

ATDBIO. **Next generation sequencing.** Disponível em: <<http://www.atdbio.com/content/58/Next-generation-sequencing#figure-bridging-pcr>>. Acesso em: 23 fev. 2017.

BABRAHAM INSTITUTE. **FastQC.** Disponível em: <<http://www.bioinformatics.babraham.ac.uk/index.html>>. Acesso em: 5 jul. 2016.

BAEBLER, Š. et al. Salicylic acid is an indispensable component of the Ny-1 resistance-gene-mediated response against Potato virus Y infection in potato. **Journal of experimental botany**, v. 65, n. 4, p. 1095–109, mar. 2014.

BARBA, M.; CZOSNEK, H.; HADIDI, A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. **Viruses**, v. 6, n. 1, p. 106–136, 6 jan. 2014.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289–300, 1995.

BRAY, N. L. et al. Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology**, v. 34, n. 5, p. 525–527, 4 maio 2016.

BUSS, D. S. et al. Oxidative Stress Defence Response of Carica papaya Challenged by Nitric Oxide, Papaya meleira virus and Saccharomyces cerevisiae. **The Open Nitric Oxide Journal**, v. 3, p. 55–64, 2011.

CAMANES, G. et al. A Deletion in NRT2.1 Attenuates Pseudomonas syringae-Induced Hormonal Perturbation, Resulting in Primed Plant Defenses. **Plant Physiology**, v. 158, n. 2, p. 1054–1066, 2012.

CHEN, F. et al. An Arabidopsis thaliana gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. **Plant Journal**, v. 36, n. 5, p. 577–588, 2003.

CHEN, L. G. et al. WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in Arabidopsis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 21, p. E1963–E1971, 2013.

CHERN, M. et al. Rice NRR, a negative regulator of disease resistance, interacts with Arabidopsis NPR1 and rice NH1. **The Plant journal : for cell and molecular biology**, v. 43, n. 5, p. 623–635, 2005.

CORNEY, D. C. RNA-seq Using Next Generation Sequencing. **Materials and Methods**, v. 3, n. 203, 2013.

COSTA-SILVA, J.; DOMINGUES, D.; LOPES, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. **PLOS ONE**, v. 12, n. 12, p. e0190152, 21 dez. 2017.

COSTA, A. T. et al. The receptor-like kinase SISOBR1 is differentially modulated by

virus infection but its overexpression in tobacco has no significant impact on virus accumulation. **Plant Cell Reports**, v. 35, n. 1, p. 65–75, 2016.

CULVER, J. N.; PADMANABHAN, M. S. Virus-Induced Disease: Altering Host Physiology One Interaction at a Time. **Annual Review of Phytopathology**, v. 45, n. 1, p. 221–243, 2007.

DEMPSEY, D. A. et al. Identification of an Arabidopsis locus required for resistance to turnip crinkle virus. **The Plant Journal**, v. 11, n. 2, p. 301–311, fev. 1997.

DEVITT, L. C. et al. Discovery of genes associated with fruit ripening in *Carica papaya* using expressed sequence tags. **Plant Science**, v. 170, n. 2, p. 356–363, 2006.

DI CARLI, M.; BENVENUTO, E.; DONINI, M. Recent Insights into Plant–Virus Interactions through Proteomic Analysis. **Journal of Proteome Research**, v. 11, n. 10, p. 4765–4780, 5 out. 2012.

DOBIN, A. et al. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics**, v. 29, n. 1, p. 15–21, 1 jan. 2013.

DONG, X. NPR1, all things considered. **Current Opinion in Plant Biology**, v. 7, n. 5, p. 547–552, 2004.

ELVIRA, M. I. et al. Proteomic analysis of pathogenesis-related proteins (PRs) induced by compatible and incompatible interactions of pepper mild mottle virus (PMMoV) in *Capsicum chinense* L3 plants. **Journal of Experimental Botany**, v. 59, n. 6, p. 1253–1265, 19 mar. 2008.

ESCALONA, M.; ROCHA, S.; POSADA, D. A comparison of tools for the simulation of genomic next-generation sequencing data. **Nat Rev Genet**, v. advance on, 20 jun. 2016.

FABI, J. P. et al. Analysis of ripening-related gene expression in papaya using an Arabidopsis-based microarray. **BMC plant biology**, v. 12, n. 1, p. 242, 2012.

FANG, J. et al. Transcriptome profiling revealed stress-induced and disease resistance genes up-regulated in PRSV resistant transgenic papaya. **Frontiers in Plant Science**, v. 7, n. 855, 2016.

FAO-TRADE-MARKETS-DIVISION. Tropical fruits outlook. In: (FAO), F. A. A. O. O. T. U. N. (Ed.). . **Medium-Term Prospects for Raw Materials, Horticulture and Tropical Products**. Rome: FAO, 2016. p. 31–37.

FAO. **Food and Agriculture Organization of the United Nations-FAOSTAT**. Disponível em: <<http://www.fao.org/faostat/en/#data>>. Acesso em: 16 jun. 2018.

FONSECA, N. A. et al. **Tools for mapping high-throughput sequencing data**. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts605>>. Acesso em: 30 jul. 2018.

FONTES, E. P. B. et al. The geminivirus nuclear shuttle protein is a virulence factor

that suppresses transmembrane receptor kinase activity. **Genes and Development**, v. 18, n. 20, p. 2545–2556, 2004.

GARCÍA-MARCOS, A. et al. Oxylin biosynthesis genes positively regulate programmed cell death during compatible infections with the synergistic pair potato virus X-potato virus Y and Tomato spotted wilt virus. **Journal of virology**, v. 87, n. 10, p. 5769–83, maio 2013.

GERI, C. et al. Arabidopsis mutants that suppress the phenotype induced by transgene-mediated expression of cauliflower mosaic virus (CaMV) gene VI are less susceptible to CaMV-infection and show reduced ethylene sensitivity. **Plant Molecular Biology**, v. 56, n. 1, p. 111–124, set. 2004.

GIOVANNONI, J. J. Genetic regulation of fruit development and ripening. **The Plant Cell**, v. 16, n. Suppl, p. S170–S180, 9 mar. 2004.

GLAUS, P.; HONKELA, A.; RATTRAY, M. Identifying differentially expressed transcripts from RNA-seq data with biological variation. **Bioinformatics**, v. 28, n. 13, p. 1721–1728, 1 jul. 2012.

GOECKS, J. et al. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. **Genome Biology**, v. 11, n. 8, p. R86, 25 ago. 2010.

GOJON, A. et al. Nitrate transceptor(s) in plants. **Journal of Experimental Botany**, v. 62, p. 2299–2308, 2011.

GÓMEZ-GÓMEZ, L.; BOLLER, T. Flagellin perception: A paradigm for innate immunity. **Trends in Plant Science**, v. 7, n. 6, p. 251–256, 2002.

GOODSTEIN, D. M. et al. Phytozome: A comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, 2012.

GOUVEIA, B. C. et al. **Immune receptors and co-receptors in antiviral innate immunity in plants** *Frontiers in Microbiology*, 2017.

GRUNER, K. et al. Reprogramming of plants during systemic acquired resistance. **Frontiers in plant science**, v. 4, n. July, p. 252, 2013.

GSCHWEND, A. R. et al. Genome size variation among sex types in dioecious and trioecious Caricaceae species. **Euphytica**, v. 189, n. 3, p. 461–469, 2013.

HAGEL, J. M.; YEUNG, E. C.; FACCHINI, P. J. **Got milk? The secret life of laticifers** *Trends in Plant Science*, 2008.

HANCOCK, J. T. et al. Cell signalling following plant/pathogen interactions involves the generation of reactive oxygen and reactive nitrogen species. **Plant Physiology and Biochemistry**, v. 40, n. 6–8, p. 611–617, 2002.

HARDCASTLE, T. J.; KELLY, K. A. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. **BMC Bioinformatics**, v. 11, n. 1, p. 422, 10 ago. 2010.

HUANG, Z. et al. Salicylic acid-dependent expression of host genes in compatible Arabidopsis-virus interactions. **Plant Physiology**, v. 137, n. 3, p. 1147–1159, mar. 2005.

IBGE. **Instituto Brasileiro de Geografia e Estatística-Metadados**. Disponível em: <<https://metadados.ibge.gov.br/consulta/estatisticos/operacoes-estatisticas/PA>>. Acesso em: 16 jun. 2018.

ILLUMINA. **Illumina Sequencing Technology. Highest data accuracy, simple workflow, and a broad range of applications**. USA: Illumina, Inc., 2010.

IRITI, M.; FAORO, F. Review of innate and specific immunity in plants and animals. **Mycopathologia**, v. 164, n. 2, p. 57–64, 2007.

ITAYA, A. et al. Potato spindle tuber viroid Strains of Different Pathogenicity Induces and Suppresses Expression of Common and Unique Genes in Infected Tomato. **Molecular Plant-Microbe Interactions MPMI**, v. 990, n. 10, 2002.

JADA, B. et al. Multiple Different Defense Mechanisms Are Activated in the Young Transgenic Tobacco Plants Which Express the Full Length Genome of the Tobacco Mosaic Virus, and Are Resistant against this Virus. **PLoS ONE**, v. 9, n. 9, p. e107778, 22 set. 2014.

JAIN, P. K. et al. Generation Sequencing Principle and Applications to Crops . In: BENKEBLIA, N. (Ed.). . **Omics Technologies and Crop Improvement**. New York: CRC Press Taylor & Francis Group, 2014.

JIA, C. et al. MetaDiff: differential isoform expression analysis using random-effects meta-regression. **BMC Bioinformatics**, v. 16, n. 1, p. 208, 2 dez. 2015.

JIANG, H.; WONG, W. H. Statistical inferences for isoform expression in RNA-Seq. **Bioinformatics (Oxford, England)**, v. 25, n. 8, p. 1026–1032, 2009.

JIMÉNEZ, V. M.; MORA-NEWCOMER, E.; GUTIÉRREZ-SOTO, M. V. Biology of the Papaya Plant. In: **Genetics and Genomics of Papaya**. [s.l: s.n.]. v. 10p. 17–33.

JOHNSON, C.; BODEN, E.; ARIAS, J. Salicylic acid and NPR1 induce the recruitment of trans-activating TGA factors to a defense gene promoter in Arabidopsis. **The Plant cell**, v. 15, n. August 2003, p. 1846–1858, 2003.

KANG, B.-C.; YEAM, I.; JAHN, M. M. Genetics of Plant Virus Resistance. **Annual Review of Phytopathology**, v. 43, n. 1, p. 581–621, set. 2005.

KEMMERLING, B. A genome-wide survey for Arabidopsis leucine-rich repeat receptor kinases implicated in plant immunity. **Frontiers in Plant Science**, v. 2, n. November, p. 1–6, 2011.

KIM, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome biology**, v. 14, n. 4, p. R36, 25 abr. 2013.

KIM, D.; LANGMEAD, B.; SALZBERG, S. L. HISAT: a fast spliced aligner with low

memory requirements. **Nature Methods**, v. 12, n. 4, p. 357–360, 9 abr. 2015.

KITAJIMA, E. W. et al. Association of isometric viruslike particles, restricted to lactifiers, with "meleira" ("sticky disease") of papaya (*Carica papaya*). **Fitopatologia Brasileira**, v. 18, p. 118, 1993.

KOO, Y. J. et al. Overexpression of salicylic acid carboxyl methyltransferase reduces salicylic acid-mediated pathogen resistance in *Arabidopsis thaliana*. **Plant Molecular Biology**, v. 64, n. 1–2, p. 1–15, 2007.

KORPELAINEN, E. et al. **RNA-seq data analysis: a practical approach**. Disponível em: <[https://www.ebi.ac.uk/~nf/hts\\_mappers/#timeline](https://www.ebi.ac.uk/~nf/hts_mappers/#timeline)>. Acesso em: 27 jul. 2018.

LANGMEAD, B. Aligning short sequencing reads with Bowtie. **Current Protocols in Bioinformatics**, n. SUPP.32, 2010.

LENG, N. et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. **Bioinformatics**, v. 29, n. 8, p. 1035–1043, 15 abr. 2013.

LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC Bioinformatics**, v. 12, n. 1, p. 323, 2011.

LI, J.; BRADER, G.; PALVA, E. T. The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. **The Plant cell**, v. 16, n. 2, p. 319–331, 2004.

LI, J.; TIBSHIRANI, R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. **Statistical Methods in Medical Research**, v. 22, n. 5, p. 519–536, 28 out. 2013.

LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, 1 abr. 2014.

LIU, L. et al. Comparison of next-generation sequencing systems. **J Biomed Biotechnol**, v. 2012, p. 251364, 2012.

LIU, P. P. et al. Interconnection between Methyl Salicylate and Lipid-Based Long-Distance Signaling during the Development of Systemic Acquired Resistance in *Arabidopsis* and Tobacco. **Plant Physiology**, v. 155, n. 4, p. 1762–1768, 2011.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC(T)</sup> Method. **Methods (San Diego, Calif.)**, v. 25, n. 4, p. 402–8, 2001.

LOAKE, G.; GRANT, M. Salicylic acid in plant defence—the players and protagonists. **Current Opinion in Plant Biology**, v. 10, n. 5, p. 466–472, 2007.

LOHSE, M. et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. **Nucleic Acids Research**, v. 40, n. W1, p. W622–W627, 1 jul. 2012.

LÓPEZ-GRESA, M. P. et al. Salicylic Acid Is Involved in the Basal Resistance of Tomato Plants to Citrus Exocortis Viroid and Tomato Spotted Wilt Virus. **PLOS ONE**, v. 11, n. 11, p. e0166938, 28 nov. 2016.

LOVE, A. J. et al. Cauliflower mosaic virus, a compatible pathogen of Arabidopsis, engages three distinct defense-signaling pathways and activates rapid systemic generation of reactive oxygen species. **Plant Physiol**, v. 139, n. 2, p. 935–948, 2005.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, 5 dez. 2014.

LUNA, E. et al. Role of NPR1 and KYP in long-lasting induced resistance by  $\beta$ -aminobutyric acid. **Frontiers in plant science**, v. 5, n. May, p. 184, 2014.

MACIEL-ZAMBOLIM, E. et al. Purification and some properties of Papaya meleira virus, a novel virus infecting papayas in Brazil. **Plant Pathology**, v. 52, n. 3, p. 389–394, 2003.

MACIEL-ZAMBOLIM, E. et al. Purification and some properties of Papaya meleira virus, a novel virus infecting papayas in Brazil. **Plant Pathology**, v. 52, n. 3, p. 389–394, 2003.

MADROÑERO, L. J. **Análise Transcriptômica da interação mamoeiro-Papaya Meleira Virus**. [s.l.] Universidade Federal Do Espírito Santo, 2014.

MAGAÑA-ÁLVAREZ, A. et al. Physical Characteristics of the Leaves and Latex of Papaya Plants Infected with the Papaya meleira Virus. **International Journal of Molecular Sciences**, v. 17, n. 4, p. 574, 15 abr. 2016.

MALAMY, J.; HENNIG, J.; KLESSIG, D. F. Temperature-Dependent Induction of Salicylic Acid and Its Conjugates during the Resistance Response to Tobacco Mosaic Virus Infection. **The Plant Cell**, v. 4, p. 359–366, 1992.

MALDONADO, A. M. et al. A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. **Nature**, v. 419, n. 6905, p. 399–403, 2002.

MALECK, K. et al. The transcriptome of Arabidopsis thaliana during systemic acquired resistance. **Nature genetics**, v. 26, n. december 2000, p. 403–410, 2000.

MANDADI, K. K.; SCHOLTHOF, K. B. G. Characterization of a Viral Synergism in the Monocot Brachypodium distachyon Reveals Distinctly Altered Host Molecular Processes Associated with Disease. **Plant Physiology**, v. 160, n. 3, p. 1432–1452, 2012.

MANDADI, K. K.; SCHOLTHOF, K. B. G. Plant immune responses against viruses: how does a virus cause disease? **Plant Cell**, v. 25, p. 1489–1505, 2013.

MANGEON, A.; JUNQUEIRA, R. M.; SACHETTO-MARTINS, G. Functional diversity of the plant glycine-rich proteins superfamily. **Plant signaling & behavior**, v. 5, n. 2, p. 99–104, 2010.



MAULE, A.; LEH, V.; LEDERER, C. The dialogue between viruses and hosts in compatible interactions. **Current Opinion in Plant Biology**, v. 5, n. 4, p. 279–284, 2002.

MEISSNER FILHO, P. E. et al. **Avaliação da resistência de genótipos de mamoeiro ao vírus da meleira no Semiárido Embrapa Mandioca e Fruticultura-Boletim de Pesquisa e Desenvolvimento (INFOTECA-E)**. Cruz das Almas, BA: [s.n.].

MING, R. et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). **Nature**, v. 452, n. 7190, p. 991–996, 2008.

MING, R.; MOORE, P. H. **Genetics and genomics of papaya**. [s.l.] Springer, 2014. v. 10

NICAISE, V. Crop immunity against viruses: outcomes and future challenges. **Frontiers in Plant Science**, 2014.

OHTA, M. et al. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. **The Plant cell**, v. 13, n. 8, p. 1959–1968, 2001.

OLIVEROS, J. C. **VENNY. An interactive tool for comparing lists with Venn Diagrams**. Disponível em: <<http://bioinfogp.cnb.csic.es/tools/venny/index.html>>.

OZSOLAK, F.; MILOS, P. M. RNA sequencing: advances, challenges and opportunities. **Nat Rev Genet**, v. 12, n. 2, p. 87–98, 2011.

PACHECO, R. et al. Comparative Analysis of Transcriptomic and Hormonal Responses to Compatible and Incompatible Plant-Virus Interactions that Lead to Cell Death. **Molecular Plant-Microbe Interactions**, v. 25, n. 5, p. 709–723, 2012.

PADMANABHAN, M. S. et al. Interaction of the Tobacco Mosaic Virus Replicase Protein with the Aux/IAA Protein PAP1/IAA26 Is Associated with Disease Development. **Journal of Virology**, v. 79, n. 4, p. 2549–2558, 2005.

PADMANABHAN, M. S. et al. Tobacco Mosaic Virus Replicase-Auxin/Indole Acetic Acid Protein Interactions: Reprogramming the Auxin Response Pathway To Enhance Virus Infection. **Journal of Virology**, v. 82, n. 5, p. 2477–2485, 2008.

PASTOR, V. et al. Disruption of the ammonium transporter AMT1.1 alters basal defenses generating resistance against *Pseudomonas syringae* and *Plectosphaerella cucumerina*. **Frontiers in plant science**, v. 5, n. May, p. 231, 2014.

PATERSON, A. H. et al. The Fruits of Tropical Plant Genomics. **Tropical Plant Biology**, v. 1, p. 3–19, 2008.

PATRO, R. et al. Salmon provides fast and bias-aware quantification of transcript expression. **Nature methods**, v. 14, n. 4, p. 417–419, abr. 2017.

PATRO, R.; MOUNT, S. M.; KINGSFORD, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. **Nature**

**Biotechnology**, v. 32, n. 5, p. 462–464, 20 maio 2014.

PHYTOZOME. **Carica papaya ASGPBv0.4 (Papaya)**. Disponível em: <[https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Cpapaya](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Cpapaya)>. Acesso em: 10 fev. 2017.

PIERCE, E. J. et al. Assessing Global Transcriptome Changes in Response to South African Cassava Mosaic Virus [ZA-99] Infection in Susceptible *Arabidopsis thaliana*. **PLoS ONE**, v. 8, n. 6, p. e67534, 27 jun. 2013.

PORTER, B. W. et al. Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. **Molecular Genetics and Genomics**, v. 281, n. 6, p. 609–626, 5 jun. 2009.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 1 jan. 2010.

RODRIGUES, S. P. et al. Effects of the Papaya meleira virus on papaya latex structure and composition. **Plant Cell Reports**, v. 28, n. 5, p. 861–871, 2009a.

RODRIGUES, S. P. et al. Papaya meleira virus is neither transmitted by infection at wound sites nor by the whitefly *Trialeurodes variabilis*. **Journal of Plant Pathology**, v. 91, n. 1, p. 87–91, 2009b.

RODRIGUES, S. P. et al. Proteomic analysis of papaya (*Carica papaya* L.) displaying typical sticky disease symptoms. **Proteomics**, v. 11, n. 13, p. 2592–2602, 2011.

RODRIGUES, S. P. et al. Label-free quantitative proteomics reveals differentially regulated proteins in the latex of sticky diseased *Carica papaya* L. plants. **Journal of Proteomics**, v. 75, n. 11, p. 3191–3198, 2012.

RYALS, J. A. et al. Systemic Acquired Resistance. **The Plant cell**, v. 8, n. 10, p. 1809–1819, 1996.

SÁ ANTUNES, T. F. et al. The dsRNA Virus Papaya Meleira Virus and an ssRNA Virus Are Associated with Papaya Sticky Disease. **PloS one**, v. 11, n. 5, p. e0155240, 11 jan. 2016.

SENTHIL, G. et al. Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses. **Journal of General Virology**, v. 86, n. 9, p. 2615–2625, 2005.

SESKAR, M.; SHULAEV, V.; RASKIN, I. Endogenous Methyl Salicylate in Pathogen-Inoculated Tobacco Plants1. **Plant physiology**, v. 116, n. 1, p. 387–392, 1998.

SOARES, E. DE A. et al. Label-free quantitative proteomic analysis of pre-flowering PMeV-infected *Carica papaya* L. **Journal of Proteomics**, v. 151, p. 275–283, 2017.

SONESON, C.; LOVE, M. I.; ROBINSON, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. **F1000Research**, v. 4, p. 1521, 2015.

SRIVASTAVA, A. et al. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. **Bioinformatics (Oxford, England)**, v. 32, n. 12, p. i192–i200, 2016.

SUN, F. et al. RNA-seq-based digital gene expression analysis reveals modification of host defense responses by rice stripe virus during disease symptom development in Arabidopsis. **Virology Journal**, v. 13, n. 1, p. 202, 2 dez. 2016.

TARAZONA, S. et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. **Nucleic Acids Research**, v. 43, n. 21, p. gkv711, 16 jul. 2015.

TRAPNELL, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 7, n. 3, p. 562–578, 2012.

TRAPNELL, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. **Nature biotechnology**, v. 31, n. 1, p. 46–53, 2013.

TRAPNELL, C.; PACHTER, L.; SALZBERG, S. L. TopHat: Discovering splice junctions with RNA-Seq. **Bioinformatics**, v. 25, n. 9, p. 1105–1111, 2009.

TRIPATHI, S. et al. Advances in Papaya Genomics. In: **Omics Technologies and Crop Improvement**. [s.l.] CRC Press, 2014. p. 219–252.

UEKI, S.; CITOVSKY, V. Plasmodesmata-associated proteins: can we see the whole elephant? **Plant signaling & behavior**, v. 9, n. 2, p. e27899, 2014.

VANDESOMPELE, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome Biology**, v. 3, n. 7, 2002.

VELCULESCU, V. E. et al. Serial Analysis of Gene Expression. **Science**, v. 270, n. 5235, p. 484–487, 1995.

VENTURA, J. A.; COSTA, H.; TATAGIBA, J. S. Papaya diseases and integrated control. In: NAQVI, S. A. H. M. (Ed.). **Diseases of Fruits and Vegetables**. [s.l.] Springer, 2004. v. IIp. 201–268.

VLOT, A. C.; DEMPSEY, D. A.; KLESSIG, D. F. Salicylic Acid, a Multifaceted Hormone to Combat Disease. **Annual Review of Phytopathology**, v. 47, n. 1, p. 177–206, 2009.

VON SAINT PAUL, V. et al. The Arabidopsis Glucosyltransferase UGT76B1 Conjugates Isoleucic Acid and Modulates Plant Defense and Senescence. **The Plant Cell**, v. 23, n. 11, p. 4124–4145, 2011.

WANG, K. et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. **Nucleic acids research**, v. 38, n. 18, p. e178, 2010.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews. Genetics**, v. 10, n. 1, p. 57–63, 2009.

- WEIGEL, R. R.; PFITZNER, U. M.; GATZ, C. Interaction of NIMIN1 with NPR1 modulates PR gene expression in Arabidopsis. **The Plant cell**, v. 17, n. 4, p. 1279–1291, 2005.
- WHENHAM, R. J. et al. Tobacco-mosaic-virus-induced increase in abscisic-acid concentration in tobacco leaves: **Planta**, v. 168, n. 4, p. 592–598, set. 1986.
- WHITHAM, S. A.; YANG, C.; GOODIN, M. M. Global Impact: Elucidating Plant Responses to Viral Infection. v. 19, n. 11, p. 1207–1215, 2006.
- WU, D. C. et al. Limitation of alignment-free tools in total RNA-seq quantification. **bioRxiv**, p. 246967, 2018.
- YANG, C. et al. Spatial Analysis of Arabidopsis thaliana Gene Expression in Response to Turnip mosaic virus Infection. **Molecular Plant-Microbe Interactions MPMI**, v. 358, n. 4, p. 358–370, 2007.
- YU, Q. et al. Cloning and characterization of a FLORICAULA/LEAFY ortholog, PFL, in polygamous papaya. **Cell Res**, v. 15, n. 8, p. 576–584, ago. 2005.
- YU, Q. et al. Recent Origin of Dioecious and Gynodioecious Y Chromosomes in Papaya. **Tropical Plant Biology**, v. 1, n. 1, p. 49–57, 2008.
- ZHANG, C. et al. Evaluation and comparison of computational tools for RNA-seq isoform quantification. **BMC genomics**, v. 18, n. 1, p. 583, 2017.
- ZHANG, Y. et al. Interaction of NPR1 with basic leucine zipper protein transcription factors that bind sequences required for salicylic acid induction of the PR-1 gene. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 11, p. 6523–6528, 1999.
- ZHANG, Z. H. et al. A comparative study of techniques for differential expression analysis on RNA-Seq data. **PloS one**, v. 9, n. 8, p. e103207, 2014.
- ZHAO, S.; XI, L.; ZHANG, B. Union Exon Based Approach for RNA-Seq Gene Quantification: To Be or Not to Be? **PLOS ONE**, v. 10, n. 11, p. e0141910, 11 nov. 2015.

## APÊNDICES

### APÊNDICE A.

Lista de transcritos e sua abundância estimada (FPKMs) nas plantas de *C. papaya* tratamento e controle nos estágios de pré- e pósfloração. ([Presente no arquivo .xls “Tabelas anexas Tese Doutorado Johana Madroño”](#)).

### APÊNDICE B.

Transcritos encontrados unicamente na pré- ou na pósfloração nas plantas tratamento ou nas plantas controle ao longo das triplicatas. ([Presente no arquivo .xls “Tabelas anexas Tese Doutorado Johana Madroño”](#)).

### APÊNDICE C.

Lista de genes diferencialmente expressos (GDEs) nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a prefloração e suas correspondentes ontologias gênicas (GO). ([Presente no arquivo .xls “Tabelas anexas Tese Doutorado Johana Madroño”](#)).

### APÊNDICE D.

Lista de genes diferencialmente expressos (GDEs) nas plantas de *C. papaya* infectadas pelo complexo PMeV durante a pós-floração e suas correspondentes ontologias gênicas (GO). ([Presente no arquivo .xls “Tabelas anexas Tese Doutorado Johana Madroño”](#)).

## ANEXOS

### ANEXO A

Link: [Artigo científico publicado \(Plant Cell Reports\)](#). (Primeira folha a continuação)

### ANEXO B.

Link: [Artigo científico publicado \(Journal of proteomics\)](#). (Primeira folha a continuação).



# Transcriptome analysis provides insights into the delayed sticky disease symptoms in *Carica papaya*

Johana Madroñero<sup>1</sup> · Silas P. Rodrigues<sup>1,2</sup> · Tathiana F. S. Antunes<sup>1</sup> · Paolla M. V. Abreu<sup>1</sup> · José A. Ventura<sup>1,3</sup> · A. Alberto R. Fernandes<sup>1</sup> · Patricia Machado Bueno Fernandes<sup>1</sup>

Received: 21 December 2017 / Accepted: 16 March 2018 / Published online: 21 March 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

**Key message** Global gene expression analysis indicates host stress responses, mainly those mediated by SA, associated to the tolerance to sticky disease symptoms at pre-flowering stage in *Carica papaya*.

**Abstract** *Carica papaya* plants develop the papaya sticky disease (PSD) as a result of the combined infection of papaya meleira virus (PMeV) and papaya meleira virus 2 (PMeV2), or PMeV complex. PSD symptoms appear only after *C. papaya* flowers. To understand the mechanisms involved in this phenomenon, the global gene expression patterns of PMeV complex-infected *C. papaya* at pre- and post-flowering stages were assessed by RNA-Seq. The result was 633 and 88 differentially expressed genes at pre- and post-flowering stages, respectively. At pre-flowering stage, genes related to stress and transport were up-regulated while metabolism-related genes were down-regulated. It was observed that induction of several salicylic acid (SA)-activated genes, including *PRI*, *PR2*, *PR5*, *WRKY* transcription factors, ROS and callose genes, suggesting SA signaling involvement in the delayed symptoms. In fact, pre-flowering *C. papaya* treated with exogenous SA showed a tendency to decrease the PMeV and PMeV2 loads when compared to control plants. However, pre-flowering *C. papaya* also accumulated transcripts encoding a NPR1-inhibitor (NPR1-I/NIM1-I) candidate, genes coding for UDP-glucosyltransferases (UGTs) and several genes involved with ethylene pathway, known to be negative regulators of SA signaling. At post-flowering, when PSD symptoms appeared, the down-regulation of *PR-1* encoding gene and the induction of *BSMT1* and JA metabolism-related genes were observed. Hence, SA signaling likely operates at the pre-flowering stage of PMeV complex-infected *C. papaya* inhibiting the development of PSD symptoms, but the induction of its negative regulators prevents the full-scale and long-lasting tolerance.

**Keywords** *Carica papaya* · Papaya meleira virus · Transcriptome · Plant–virus interaction · Defense responses · SA signaling

---

Communicated by Baochun Li.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00299-018-2281-x>) contains supplementary material, which is available to authorized users.

✉ Patricia Machado Bueno Fernandes  
patricia.fernandes@ufes.br

<sup>1</sup> Núcleo de Biotecnologia, Universidade Federal do Espírito Santo, Av. Marechal Campos, 1468, Vitória, ES 29040-090, Brazil

<sup>2</sup> Núcleo Multidisciplinar de Pesquisa-Polo de Xerém, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

<sup>3</sup> Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural, Vitória, ES, Brazil

## Introduction

Papaya sticky disease (PSD) is characterized by a spontaneous exudation of a fluid and translucent latex, which oxidizes when in contact with air and accumulates mainly on the plant's fruits and leaves. PSD was officially reported in Brazil and Mexico, two major papaya producers. Sticky diseased papaya fruits are unfit for consumption and affect the control of fruit fly populations, impairing fruit marketing and exportation (Abreu et al. 2015).

PSD is associated to a combined infection of PMeV, an isometric toti-like virus with a double-stranded RNA, and PMeV2, an umbra-like virus with single stranded RNA genome (Sá Antunes et al. 2016). PMeV complexes (PMeV and PMeV2) are the only laticifers-occurring viruses that



## Label-free quantitative proteomic analysis of pre-flowering PMeV-infected *Carica papaya* L.



Eduardo de A. Soares<sup>a</sup>, Emily G. Werth<sup>b</sup>, Leidy J. Madroño<sup>a</sup>, José A. Ventura<sup>a,c</sup>, Silas P. Rodrigues<sup>a,d,\*</sup>, Leslie M. Hicks<sup>b</sup>, Patricia M.B. Fernandes<sup>a</sup>

<sup>a</sup> Núcleo de Biotecnologia, Universidade Federal do Espírito Santo, Av. Marechal Campos, 1498, Vitória, ES 29040-090, Brazil

<sup>b</sup> Department of Chemistry, University of North Carolina at Chapel Hill, 125 South Road, Chapel Hill, NC 27599, USA

<sup>c</sup> Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural, Rua Afonso Sarlo 160, Vitória, ES 29052-010, Brazil

<sup>d</sup> Unidade de Espectrometria de Massas e Proteômica e Núcleo Multidisciplinar de Pesquisa e Extensão de Xerém, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, CCS Bloco H2 Sala 04, Rio de Janeiro, RJ 21941-902, Brazil

### ARTICLE INFO

#### Article history:

Received 30 January 2016

Received in revised form 27 May 2016

Accepted 18 June 2016

Available online 23 June 2016

#### Keywords:

Label-free quantitative proteomics

Mass spectrometry

Papaya meleira virus

### ABSTRACT

*Papaya meleira virus* (PMeV) infects papaya (*Carica papaya* L.) and leads to Papaya Sticky Disease (PSD) or “Meleira”, characterized by a spontaneous exudation of latex from fruits and leaves only in the post-flowering developmental stage. The latex oxidizes in contact with air and accumulates as a sticky substance on the plant organs, impairing papaya fruit’s marketing and exportation. To understand pre-flowering *C. papaya* resistance to PMeV, an LC-MS/MS-based label-free proteomics approach was used to assess the differential proteome of PMeV-infected pre-flowering *C. papaya* vs. uninfected (control) plants. In this study, 1333 proteins were identified, of which 111 proteins showed a significant abundance change (57 increased and 54 decreased) and supports the hypothesis of increased photosynthesis and reduction of 26S-proteasome activity and cell-wall remodeling. All of these results suggest that increased photosynthetic activity has a positive effect on the induction of plant immunity, whereas the reduction of caspase-like activity and the observed changes in the cell-wall associated proteins impairs the full activation of defense response based on hypersensitive response and viral movement obstruction in pre-flowering *C. papaya* plants.

**Biological significance:** The papaya (*Carica papaya* L.) fruit’s production is severely limited by the occurrence of *Papaya meleira virus* (PMeV) infection, which causes Papaya Sticky Disease (PSD). Despite the efforts to understand key features involved with the plant × virus interaction, PSD management is still largely based on the observation of the first disease symptoms in the field, followed by the elimination of the diseased plants. However, *C. papaya* develops PSD only after flowering, *i.e.* about six-months after planting, and the virus inoculum sources are kept in field. The development of PMeV resistant genotypes is impaired by the limited knowledge about *C. papaya* resistance against viruses. The occurrence of a resistance/tolerance mechanism to PSD symptoms development prior to *C. papaya* flowering is considered in this study. Thus, field-grown and PMeV-infected *C. papaya* leaf samples were analyzed using proteomics, which revealed the modulation of photosynthesis-, 26S proteasome- and cell-wall remodeling-associated proteins. The data implicate a role for those systems in *C. papaya* resistance to viruses and support the idea of a partial resistance induction in the plants at pre-flowering stage. The specific proteins presented in the manuscript represent a starting point to the selection of key genes to be used in *C. papaya* improvement to PMeV infection resistance. The presented data also contribute to the understanding of virus-induced disease symptoms development in plants, of interest to the plant-virus interaction field.

© 2016 Elsevier B.V. All rights reserved.

### 1. Introduction

*Papaya meleira virus* (PMeV) infects papaya (*Carica papaya* L.) and leads to Papaya Sticky Disease (PSD) or “Meleira” [1]. PSD is characterized by a spontaneous exudation of latex from fruits and leaves [2] which oxidizes in contact with air and accumulates as a sticky substance

on the plant organs [3]. PSD is officially reported to occur in Brazil and Mexico, two major papaya fruit producing countries [4].

PMeV has been observed in the laticifers of *C. papaya* [3] where it induces an extensive production of H<sub>2</sub>O<sub>2</sub> [5]. The analysis of PMeV-infected latex samples has previously allowed inferences about other local PMeV-associated effects such as higher levels of potassium, phosphorus and water, likely associated with an osmotic imbalance in PMeV-infected laticifers [5], along with reduced cysteine-protease abundance and activity [6]. In parallel, the accumulation of H<sub>2</sub>O<sub>2</sub> in the phloem [5]

\* Corresponding author.

E-mail address: [srodrigues@xerem.ufrrj.br](mailto:srodrigues@xerem.ufrrj.br) (S.P. Rodrigues).