Universidade Estadual de Londrina

MICHELLE PIRES RINCÃO

TRANSCRIPTOMA *IN PLANTA* DE *PHAKOPSORA PACHYRHIZI* E CARACTERIZAÇÃO FUNCIONAL DE GENES ENVOLVIDOS EM MECANISMOS BASAIS DE SOBREVIVÊNCIA E PATOGENICIDADE

Londrina
2017

Universidade Estadual de Londrina

Instituto Agronômico do Paraná

Empresa Brasileira de Pesquisa Agropecuária

MICHELLE PIRES RINCÃO

TRANSCRIPTOMA *IN PLANTA* DE *PHAKOPSORA PACHYRHIZI* E CARACTERIZAÇÃO FUNCIONAL DE GENES ENVOLVIDOS EM MECANISMOS BASAIS DE SOBREVIVÊNCIA E PATOGENICIDADE

Londrina
2017

MICHELLE PIRES RINCÃO

TRANSCRIPTOMA *IN PLANTA* DE *PHAKOPSORA PACHYRHIZI* E CARACTERIZAÇÃO FUNCIONAL DE GENES ENVOLVIDOS EM MECANISMOS BASAIS DE SOBREVIVÊNCIA E PATOGENICIDADE

Tese de doutorado apresentada ao Programa de Pós Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Doutora.

Orientador: Prof. Dr. Ricardo Vilela Abdelnoor.
Co-orientadora: Dra. Francismar Correa Marcelino-Guimarães.

Londrina
2017

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Rincão, Michelle Pires.
Transcriptoma in planta de Phakopsora pachyrhizi e caracterização funcional de genes envolvidos em mecanismos basais de sobrevivência e patogenicidade / Michelle Pires Rincão. - Londrina, 2017.
212 f. : il.

Orientador: Ricardo Vilela Abdelnoor.
Coorientador: Francismar Correa Marcelino-Guimarães.
Tese (Doutorado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2017.
Inclui bibliografia.

1. Ferrugem-asiática da soja - Tese. 2. Phakopsora pachyrhizi - transcriptoma - Tese. 3. Validação funcional de genes de Phakopsora pachyrhizi - Tese. I. Abdelnoor, Ricardo Vilela. II. Marcelino-Guimarães, Francismar Correa. III. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. IV. Título.

MICHELLE PIRES RINCÃO

TRANSCRIPTOMA *IN PLANTA* DE *PHAKOPSORA PACHYRHIZI* E CARACTERIZAÇÃO FUNCIONAL DE GENES ENVOLVIDOS EM MECANISMOS BASAIS DE SOBREVIVÊNCIA E PATOGENICIDADE

Tese de doutorado apresentada ao Programa de Pós Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Doutora.

BANCA EXAMINADORA

_____

Orientador Dr. Ricardo Vilela Abdelnoor
Empresa Brasileira de Pesquisa Agropecuária – Embrapa Soja

_____

Dra. Mayra C. C. Gallo de Carvalho
Universidade Estadual do Norte do Paraná - UENP

_____

Dr. Eduardo Chumbinho de Andrade
Empresa Brasileira de Pesquisa Agropecuária - Embrapa Mandioca e Fruticultura

_____

Dra. Claudia Vieira Godoy
Empresa Brasileira de Pesquisa Agropecuária – Embrapa Soja

_____

Prof. Dr. Laurival Antonio Vilas-Boas
Universidade Estadual de Londrina - UEL

Londrina. 29 de Maio de 2017.

Ao meu esposo, Leandro,
pelo companheirismo, paciência e amor incondicional.

Aos meus pais, Clair e Vilson,
pela base familiar, amor, cuidado, educação e incentivo.

Aos meus irmãos, Vinícius e Matheus,
pela amizade, carinho e apoio.

DEDICO

“O cansaço físico, mesmo que suportado forçosamente, não prejudica o corpo, enquanto o conhecimento imposto à força não pode permanecer na alma por muito tempo”.

Platão

“A ignorância gera mais frequentemente confiança do que o conhecimento: são os que sabem pouco, e não aqueles que sabem muito, que afirmam de uma forma tão categórica que este ou aquele problema nunca será resolvido pela ciência”.

Charles Darwin

AGRADECIMENTOS

A Deus, por todas as graças a mim concedidas, e por sempre me capacitar para a realização dos meus objetivos.

A toda minha família, em especial aos meus pais Vilson e Clair, pelo amor incondicional e pelas sábias palavras que sempre me fortaleceram nos momentos difíceis; e ao meu esposo Leandro por todo o apoio, companheirismo, amor e compreensão.

Ao meu orientador Prof. Dr. Ricardo Vilela Abdelnoor, pela orientação e pela oportunidade de crescimento profissional desde a graduação até o doutorado, pela confiança em mim depositada e pela paciência em partilhar seus conhecimentos.

À Dra. Francismar Correa Marcelino-Guimarães, pela co-orientação e sugestões, em especial, pelos conhecimentos, conversas e incentivos recebidos para melhor direcionar este projeto.

Aos doutores Mayra Costa da Cruz Gallo de Carvalho, Eduardo Chumbinho de Andrade , Claudia Vieira Godoy e Laurival Antonio Vilas-Boas, por gentilmente aceitarem avaliar esta tese com sugestões construtivas, enriquecendo este trabalho.

Aos doutores Marcelo Falzarella Carazolle (UNICAMP) e Leandro Costa Nascimento (UNICAMP), por toda a ajuda e por compartilharem seus conhecimentos em bioinformática.

A Profa. Dra. Ana Lúcia Dias e a Profa. Dra. Silvia Helena Sofia, pela sábia coordenação do curso de Pós Graduação em Genética e Biologia Molecular, e por estarem sempre à disposição para nos ajudar e nos orientar frente a qualquer dificuldade.

Aos professores do Programa de Pós Graduação em Genética e Biologia Molecular pela oportunidade, estrutura e conhecimento adquirido ao longo do doutorado.

A todos os pesquisadores da Embrapa Soja, em especial a Dra. Ivani de Oliveira Negrão Lopes, pela ajuda e conhecimentos que auxiliaram na conclusão deste trabalho.

Agradeço aos técnicos e analistas do laboratório de Biotecnologia Vegetal da Embrapa Soja por toda ajuda e amizade, assim como a todos os funcionários da Embrapa Soja que de alguma forma contribuíram para a realização deste projeto.

Aos queridos amigos do laboratório de Biotecnologia Vegetal da Embrapa Soja, por toda a ajuda e conhecimento partilhado ao longo destes anos, pelas ótimas conversas, conselhos e por tornarem agradável e divertida a rotina de laboratório. Em especial a Dra. Kenia de Carvalho pela ajuda na execução deste trabalho e por toda a amizade.

A Universidade Estadual de Londrina (UEL), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – (CAPES) e Empresa Brasileira de Pesquisa Agro pecuária - Embrapa Soja pelo apoio financeiro, concessão da bolsa e disponibilização da estrutura física para a realização deste trabalho, respectivamente.

E todos aqueles que contribuíram, direta ou indiretamente, para a realização e desenvolvimento deste trabalho, o meu sincero agradecimento e reconhecimento. Muito obrigada!

BIOGRAFIA

Michelle Pires Rincão, filha de Clair de Fátima Pires Rincão e Vilson Crispo Rincão, nasceu no dia 21 de outubro de 1987, na cidade de Auriflama, São Paulo.

Em 2006, iniciou o curso de Ciências Biológicas na Universidade Estadual do Norte do Paraná (UENP – *Campus* Luiz Meneghel). Em julho de 2010 obteve o grau de licenciatura e em dezembro de 2012 o título de bacharel em Ciências Biológicas.

Realizou estágio de iniciação científica no laboratório de Biotecnologia Vegetal da Empresa Brasileira de Pesquisa Agropecuária (Embrapa Soja) nos anos de 2007 a 2010, sob co-orientação do Dr. Ricardo Vilela Abdelnoor (Embrapa Soja) e orientação, inicialmente da Dra. Danielle Cristina Gregório da Silva, e posteriormente da Dra. Mayra Costa Cruz Gallo de Carvalho (UENP-CLM). O projeto de monografia (TCC) foi intitulado: “Validação de marcadores moleculares microssatélites ligados ao gene *Rps1k* de resistência à Podridão Radicular de Fitóftora em soja”.

Em 2011 ingressou no Programa de Mestrado em Genética e Biologia Molecular (UEL), sob orientação do Dr. Ricardo Vilela Abdelnoor e co-orientação da Dra. Lizandra Lucy Catelli, e defendeu a dissertação intitulada: “Validação de marcadores moleculares em soja ligados a genes de resistência à Podridão Radicular de Fitóftora” em 2013. No mesmo ano, ingressou no Programa de Doutorado em Genética e Biologia Molecular (UEL), sob orientação do Dr. Ricardo Vilela Abdelnoor e co-orientação da Dra. Francismar Correa Marcelino-Guimarães, desenvolvendo a tese intitulada: “Transcriptoma *in planta* de *Phakopsora pachyrhizi* e caracterização funcional de genes envolvidos em mecanismos basais de sobrevivência e patogenicidade”, com defesa em 29 de Maio de 2017.

RINCÃO, Michelle Pires. **Transcriptoma *in planta* de *Phakopsora pachyrhizi* e caracterização funcional de genes envolvidos em mecanismos basais de sobrevivência e patogenicidade**. 212 p. Tese de Doutorado (Pós Graduação em Genética e Biologia Molecular), Universidade Estadual de Londrina, Londrina, 2017.

RESUMO

O Brasil é o segundo maior produtor mundial de soja (*Glycine max* (L) Merrill), produto responsável por metade da produção agrícola nacional. Entretanto, a produção dessa importante oleaginosa é afetada, dentre outros fatores, pela ação do fungo *Phakopsora pachyrhizi* (Sydow & P. Sydow), causador da doença conhecida como ferrugem-asiática da soja (FAS). Por ser um fungo biotrófico obrigatório, o estudo de *P. pachyrhizi* é limitado ao seu contato com o hospedeiro durante o processo de infecção. Neste sentido, este trabalho utilizou a técnica de microdissecção a laser da lesão foliar, associada ao sequenciamento de alto desempenho, para a obtenção de sequências de transcritos do patógeno durante a interação com genótipos resistente e suscetível de soja (PI561356 e BRS 231, respectivamente). Os resultados obtidos através do sequenciamento do transcriptoma do fungo compreendem a identificação de 36.350 unisequências de *P. pachyrhizi*, que forneceram uma visão geral dos mecanismos moleculares e rotas biológicas ativas no patógeno aos dez dias após a infecção em soja, como metabolismo energético e de ácidos nucleicos, síntese proteica, além de processos de transcrição, biossíntese e transporte. Transcritos relacionados a processos de exocitose, sinalização de vias metabólicas, dentre outros, se apresentaram enriquecidos entre os genótipos resistente e suscetível utilizados neste trabalho em relação ao transcriptoma obtido de *P. pachyrhizi*. Adicionalmente, transcritos relacionados a processos de metilação, transporte de sulfato, e resposta a espécies reativas de oxigênio, entre outros, também foram enriquecidos entre diferentes estruturas e fases de infecção do fungo como esporos germinados, apressório, haustório e lesão foliar. Análises OrthoMCL comparando as unisequências geradas a partir do genoma de outras 15 espécies de fungos e oomicetos, incluindo outras espécies de ferrugens, identificaram famílias multigênicas conservadas entre as espécies analisadas, bem como famílias comuns a fungos causadores de ferrugens, e famílias encontradas exclusivamente em *P. pachyrhizi*, com destaque para famílias relacionadas a sinalização de vias metabólicas e transportadores de membrana. A predição de elementos de transposição ativos entre os 36.350 transcritos, identificou diferentes famílias de retrotransposons e de transposons de DNA, e análises de RT-qPCR confirmaram os níveis de expressão de genes específicos observados no sequenciamento de alto desempenho. Adicionalmente, sete genes potencialmente envolvidos em processos de sobrevivência e patogenicidade de *P. pachyrhizi*, selecionados a partir do transcriptoma do fungo e de análises *in silico*, foram caracterizados funcionalmente via silenciamento gênico por RNA interferente. Mediante a utilização da tecnologia de silenciamento gênico induzido pelo

hospedeiro (*Host-Induced Gene Silencing - HIGS*), construções utilizando plasmídeos baseados no genoma do vírus do mosqueado do feijoeiro (*Bean pod mottle vírus - BPMV*) foram obtidas para os sete genes do fungo previamente selecionados. As construções HIGS-BPMV foram inseridas em plantas suscetíveis de soja, e posteriormente inoculadas com *P. pachyrhizi*. Os resultados obtidos revelaram redução significativa nos níveis de transcritos e na patogenicidade do fungo para três dos sete genes silenciados. Os resultados obtidos neste trabalho contribuem com o aumento do conhecimento sobre o transcriptoma do fungo *P. pachyrhizi*, demostram a existência de uma maquinaria de silenciamento ativa nesse patógeno, e, consequentemente, que o silenciamento gênico baseado na técnica de HIGS-BPMV constitui uma ferramenta viável na caracterização funcional de genes do fungo, sendo identificados pelo menos três genes que ocasionaram reduções significativas na patogenicidade quando silenciados.

Palavras-chave: Soja. Ferrugem-asiática da soja. Transcriptoma. Silenciamento gênico.

RINCÃO, Michelle Pires. ***In planta* transcriptome of *Phakopsora pachyrhizi* and functional characterization of genes involved in survival and pathogenicity basal mechanisms**. P.212. PhD Thesis (Graduate in Genetics and Molecular Biology), State University of Londrina, Londrina, 2017.

ABSTRACT

Brazil is the second largest producer of soybean (*Glycine max* (L) Merrill), product responsible for half of the national agricultural production. However, the production of this important oilseed is affected, among other factors, by the fungus *Phakopsora pachyrhizi* (Sydow & P. Sydow) that causes the disease known as Asian Soybean Rust (ASR). Because it is an obligatory biotrophic fungus, the study of *P. pachyrhizi* is limited to its contact with the host during the infection process. In this sense, this work utilized foliar lesion laser microdissection technique, associated with high performance sequencing, to obtain pathogen transcript sequences during interaction with resistant and susceptible soybean genotypes (PI561356 and BRS 231, respectively). The results obtained through the sequencing of the fungus transcriptome comprise the identification of 36,350 unisequences of *P. pachyrhizi*. The unisequences provided an overview of the molecular mechanisms and biological pathways active in the pathogen at ten days after soybean infection, such as energetic and nucleic acid metabolism, protein synthesis, as well as transcription, biosynthesis and transport processes. Transcripts related to exocytosis processes, signaling of metabolic pathways, among others, were enriched among the resistant and susceptible genotypes used in this work in relation to the transcriptome obtained from *P. pachyrhizi*. Additionally, transcripts related to methylation processes, sulfate transport, and response to reactive oxygen species, among others, were also enriched among the different structures and phases of fungus infection, such as germinated spores, appressorium, haustorium and foliar lesion. OrthoMCL analyzes comparing the unisequences generated from the genome of other 15 species of fungi and oomycetes, including other species of rust, identified multigenic families conserved among the analyzed species, as well as families common to rust fungi, and families found exclusively in *P. pachyrhizi*, with emphasis on families related to signaling metabolic pathways and membrane transporters. The prediction of active transposable elements among the 36,350 transcripts identified different families of retrotransposons and DNA transposons, and RT-qPCR analyzes confirmed the expression levels of specific genes observed in high performance sequencing. In addition, seven genes potentially involved in *P. pachyrhizi* survival and pathogenicity, selected from the fungus transcriptome and *in silico* analyzes, were functionally characterized by gene silencing by RNA interference. Using Host-Induced Gene Silencing (HIGS) technology, constructs using plasmids based on the Bean pod mottle virus genome (BPMV) were obtained for the seven genes of the fungus previously selected. The HIGS-BPMV constructs were inserted into soybean susceptible plants, which were later inoculated with *P. pachyrhizi*. The results

showed a significant reduction in transcript levels and in the fungus pathogenicity for three of the seven silenced genes. The results obtained in this work contribute to the increase of knowledge on the transcriptome of the *P. pachyrhizi* fungus, demonstrate the existence of an active silencing machinery in this pathogen, and, consequently, that the gene silencing based on the HIGS-BPMV technique is a viable tool in the functional characterization of fungus genes, being identified at least three genes that caused significant reductions in pathogenicity when silenced.

Keywords: Soybean. Asian soybean rust. Transcriptome. Gene silencing.

LISTA DE FIGURAS

LISTA DE TABELAS

ABREVIATURAS E SIGLAS

18S	*18S ribosomal subunit*
AGO	*Argonaut*
ASR	*Asian Soybean Rust*
BPMV	*Bean Pod Mottle Virus*
CDD	*Conserved Domains Database*
DCL	*Dicer*
Dpi	*days post inoculation*
dsRNA	*double stranded RNA*
EST	*Expressed Sequence Tag*
ETI	*Effector-Triggered Immunity*
ETS	*Effector-Triggered Susceptibility*
FAS	Ferrugem-Asiática da Soja
FPKM	*Fragments Per Kilobase Million*
GO	*Gene Ontology*
Gpα	*G protein alpha subunit*
HIGS	*Host Induced Gene Silencing*
Hpi	*hours post infection*
HR	*Hypersensitivity Response*
HSS	*Small Heat Shock*
IR	*Inverted Repeat*
JGI	*Joint Genome Institute*
KEGG	*Kyoto Encyclopedia of Genes and Genomes*
LCM	*Laser Capture Microdissection*
LGE	Laboratório de Genômica e Expressão
LINE	*Long Interspersed Nuclear Element*
LTR	*Long Terminal Repeat*
MAPK	*Mitogen-Activated Proteins Kinase*
milRNAs	*miRNA-like sRNAs*
miRNAs	*microRNAs*
MSUD	*Meiotic Silencing by Unpaired DNA*
natsiRNAs	*natural antisense siRNAs*
NCBI	*National Center for Biotechnology Information*

NR	*Non Redundant GenBank database*
NtR	*Nitrate reductase*
NT	*Nucleotide GenBank database*
ORFs	*Open Reading Frame*
PLEs	*Penelope Elements*
Pp	*Phakopsora pachyrhizi*
PPI	*Peptidyl-prolyl cis/trans isomerase*
PTGS	*Post Transcriptional Gene Silencing*
Pv-SNARE	*Soluble NSF attachment receptor*
qiRNAs	*Quelling-associated small RNAs*
RB	*Reddish-Brown* (lesões típicas de resistência a *P. pachyrhizi*)
RdRP	*RNAse Dependent of RNA Polymerase*
RIP	*Repeat-Induced Point Mutation*
RISC	*RNA-Induced Silence Complex*
RNAi	*RNA interfering*
RNA-Seq	*RNA Sequencing* (Sequenciamento de alto rendimento de mRNAs)
ROS	*Reactive Oxygen Species*
RT-qPCR	*Real Time Quantitative Polimerase Chain Reaction*
S	*Spores*
SAGE	*Serial Analysis of Gene Expression*
SG	*Germinated Spores*
SINE	*Short Interspersed Nuclear Element*
siRNA	*small interfering RNAs*
SMV	*Soybean Mosaic Viruses*
sRNA	*small RNAs*
ssRNAs	*single-stranded RNAs*
TAN	(lesões típicas de suscetibilidade a *P. pachyrhizi*)
tasiRNAs	*trans-acting siRNAs*
TE	*Transposable Elements*
TGS	*Transcriptional Gene Silencing*
Thi	*Thiamine biosynthesis*
TIR	*Terminal Inverted Repeat*
TSD	*Target Site Duplication*
Tub	*Tubulin*
VIGS	*Virus-Induced Gene Silencing*

SUMÁRIO

1. CAPÍTULO 1: CONSIDERAÇÕES GERAIS

1.1 INTRODUÇÃO

Phakopsora pachyrhizi (Sydow & P. Sydow), causador da ferrugem-asiática da soja (FAS), é um fungo biotrófico pertencente a Ordem Uredinales, também conhecida como Pucciniales (GBIF, 2017). Esse fungo, pode infectar cerca de 31 espécies de plantas leguminosas (ONO et al., 1992), no entanto a soja (*Glycine max* (L.) Merrill) é seu principal hospedeiro. Seu ciclo infeccioso inicia-se com o contato dos esporos (urediniósporos) com a superfície da folha de soja, que sob condições propícias germinam e dão início ao processo infecioso, culminando com a liberação de novos urediniósporos recém-formados aproximadamente 10 dias após o início do processo de infecção, sendo observada a produção de urediniósporos apenas via reprodução assexual (ZAMBOLIN, 2006).

Devido ao elevado tamanho e complexidade, o genoma do fungo ainda não está disponível, de modo que grande parte dos estudos sobre genes e rotas metabólicas essenciais para sobrevivência, biotrofia e patogenicidade são derivados de estudos do seu transcriptoma e/ou análises comparativas com outras espécies de ferrugem (POSADA-BUITRAGO & FREDERICK, 2005; TREMBLEY et al., 2009, 2012, 2013; STONE et al., 2012; LINK et al., 2014). Tais trabalhos se limitam geralmente a uma visão restrita de genes expressos em estágios e/ou estruturas específicas durante a infecção da soja, de modo que trabalhos adicionais, que ampliem a compreensão dos mecanismos essenciais ao patógeno para infecção, estabelecimento da biotrofia e patogenicidade a nível molecular, ainda são essenciais neste patossistema.

Por se tratar de um fungo biotrófico obrigatório, métodos convencionais de estudo da função gênica via metodologias tradicionais de transformação não são aplicáveis. O silenciamento gênico indireto, baseado na produção e transferência de pequenos RNAs de interferência do hospedeiro para o patógeno, metodologia descrita como silenciamento gênico induzido pelo hospedeiro (*Host-Induced Gene Silencing - HIGS*), tem sido empregado com sucesso no estudo funcional de genes em fungos, incluindo espécies de fungos causadores de ferrugem como, por exemplo, em *Puccinia triticinia*, causador da ferrugem da folha do trigo e em *Uromyces appendiculatus*, causador da ferrugem no feijão comum (PANWAR et al., 2013a, 2013b, COOPER & CAMPBELL, 2017). O emprego de tal metodologia em *P. pachyrhizi* ainda não foi descrito, e a ocorrência de genes envolvidos na maquinaria de silenciamento gênico a nível transcricional e pós-transcricional se limita a algumas sequências identificadas no transcriptoma de haustório, mas cujas funções moleculares não foram comprovadas (LINK et al., 2014).

A identificação e caracterização funcional de genes envolvidos na sobrevivência, desenvolvimento e patogenicidade de *P. pachyrhizi*, mediante o sequenciamento de alta resolução do seu transcriptoma e a utilização da abordagem de HIGS, é de suma importância para o entendimento da doença e consequentemente para o desenvolvimento de estratégias viáveis de controle da FAS. Assim, neste trabalho o sequenciamento de alto desempenho do transcriptoma *in planta* de *P. pachyrhizi* obtido a partir da microdissecção de lesões foliares, permitiu uma visão geral dos transcritos obtidos e dos processos moleculares atuantes no fungo aos 10 dias após a infecção. Adicionalmente, a caracterização funcional de genes relacionados a mecanismos de sobrevivência e patogenicidade, revelou o importante papel destes durante o processo de infecção na soja, reduzindo

os sintomas da doença em plantas infectadas com o patógeno, quando os mesmos foram silenciados.

Os resultados obtidos são apresentados em dois capítulos. O primeiro apresenta a análise global do transcriptoma de *P. pachyrhizi* com ênfase na identificação das sequências mais expressas e transcritos enriquecidos entre genótipos e entre diferentes estruturas de infecção, assim como em análises comparativas entre os transcritos de *P. pachyrhizi* e outros fungos e oomicetos, na predição de elementos de transposição ativos, e por fim na validação dos níveis de expressão de genes do fungo observados nos resultados do sequenciamento. O segundo capítulo descreve a caracterização *in silico* de genes envolvidos em diferentes processos biológicos relacionados a sobrevivência e patogenicidade do fungo, tais como maquinaria de silenciamento gênico, síntese de proteínas, biossíntese de tiamina, sinalização de vias metabólicas, fusão vesicular, seguido do silenciamento gênico de sete genes via HIGS.

1.2 OBJETIVO GERAL

Obtenção e caracterização do transcriptoma *in planta* de *P. pachyrhizi* aos 10 dias após a infecção da soja, além da identificação e caracterização funcional de genes envolvidos em mecanismos basais de sobrevivência e infecção do fungo.

1.2.1 Objetivos Específicos

- Identificar transcritos de *P. pachyrhizi* expressos aos 10 dias após a infecção da soja a partir do sequenciamento de alto desempenho de lesões foliares obtidas por LCM (*Laser Capture Microdissection*);
- Caracterizar as principais rotas metabólicas e categorias funcionais ativas no patógeno aos 10 dias após a infecção da soja;
- Caracterizar categorias funcionais enriquecidas a partir da analise comparativa entre os transcritos expressos do fungo na integração com os diferentes genótipos PI561356 e BRS 231;
- Caracterizar categorias funcionais enriquecidas a partir da analise comparativa entre os transcritos expressos em diferentes estruturas de infecção do fungo, como esporos germinados, apressório, haustório e o transcriptoma da lesão foliar (obtido neste trabalho);
- Identificar famílias gênicas OrthoMCL conservadas entre *P. pachyrhizi* e diferentes fungos e oomicetos, bem como famílias conservadas entre fungos causadores de ferrugem e especificas de *P. pachyrhizi*;
- Identificar superfamílias de elementos de transposição presentes no transcriptoma de *P. pachyrhizi* aos 10 dias após a infecção da soja;

- Validar os níveis de expressão obtidos no sequenciamento por RNA-Seq por meio de análises de RT-qPCR;
- Selecionar *in silico* genes potencialmente envolvidos em mecanismos basais de sobrevivência e patogenicidade de *P. pachyrhizi* para estudos funcionais*;*
- Caracterizar funcionalmente os genes de *P. pachyrhizi* previamente selecionados.

1.3 FUNDAMENTAÇÃO TEÓRICA

1.3.1 A Ferrugem-Asiática da Soja

A soja (*Glycine max*) é uma *commodity* importante e a principal oleaginosa do mundo, sendo cultivada comercialmente em mais de 35 países, principalmente para o consumo humano e animal, uma vez que sua qualidade de proteínas e teor de óleo atende às necessidades dietéticas de ambos (LUSAS, 2004). Nas últimas décadas a maior expansão da produção mundial de soja vem ocorrendo em uma área quase contínua da América do Sul, abrangendo a Argentina, o Brasil, o Paraguai, o Uruguai e a Bolívia. Em todos esses países cresce igualmente a presença de grandes empresas multinacionais nos segmentos de comercialização e industrialização, que se estende também às áreas de produção de sementes e financiamento da produção do grão. Dentre outros fatores, a grande disponibilidade de terras apropriadas ao plantio da soja faz do Cone Sul da América do Sul uma das áreas preferenciais para a expansão do cultivo da soja (SCHLESINGER, 2008).

Os Estados Unidos se destacam como o maior produtor mundial de soja, e de acordo com o Departamento de Agricultura (USDA) o país fechou o ano de 2016 com uma produção de 117,1 milhões de toneladas (USDA, 2017). No Brasil, a soja ocupa 59% da área plantada e contribui com 49,8% da produção nacional de grãos, sendo cultivada em grande parte do território nacional. As estimativas de produção são de 107,61 milhões de toneladas para a safra 2016/2017 (CONAB, 2017). Assim, a soja tornou-se o motor da economia de várias regiões brasileiras e têm sido um dos maiores responsáveis pela expansão da receita cambial do país, conferindo ao

Brasil a posição de segundo maior produtor e exportador mundial desse grão (MISSÃO, 2006).

A ocorrência de doenças constitui uma das principais ameaças que acometem a cultura da soja, apresentando muitos entraves quanto aos métodos de controle, e assim, limitando em grandes proporções a produtividade das lavouras. Alguns fatores como a expansão da cultura para novas áreas e a utilização da monocultura, aliadas às más práticas agrícolas como utilização inadequada de manejo e de compostos químicos, têm aumentado o número de doenças causadas por fungos, bactérias, vírus e nematóides (EMBRAPA, 2011). A ferrugem-asiática da soja (FAS), doença fúngica, é um dos principais fatores bióticos responsáveis por perdas significativas no rendimento da safra, e tem causado grande preocupação por se tratar de uma doença policíclica e de elevado poder destrutivo (SCHERM et al., 2009). A FAS é causada por um fungo biotrófico obrigatório denominado *Phakopsora pachyrhizi* (Sydow & P. Sydow), pertencente ao Reino Fungi, Filo Basidiomicota, Ordem Uredinales, Classe Urediniomicetos, Família Phakopsoraceae, Gênero *Phakopsora* (GBIF, 2017).

A FAS foi descrita pela primeira vez no Japão em 1902 (HENNINGS, 1903). Em 1914 surgiu em caráter epidêmico em vários países no sudoeste da Ásia, disseminando-se em direção à África (LI, 2009). Em janeiro de 1998 a doença foi constatada em Uganda, Kenia e Ruanda, e em março de 2001, na África do Sul, atingido caráter epidêmico em 2002. No continente americano, a doença foi detectada em 2001 (FREIRE et al., 2008), e imediatamente passou a se tornar motivo de grande preocupação quando, na safra 2001/2002, causou danos de até 50% no Paraguai (YORINORI et al., 2002). Em 2002 a doença foi verificada na

Argentina, em 2003 na Bolívia, e no Uruguai, na Colômbia e nos Estados Unidos somente em 2004 (SCHNEIDER et al., 2005).

No Brasil, desde sua detecção em 2001 a FAS atingiu dispersão em praticamente 100% da área de cultivo, provocando danos que variaram entre 10 a 80% (YORINORI et al., 2005). Na safra de 2001/2002, as perdas registradas totalizaram 569,2 mil tonelada de grãos, com prejuízo estimado em US$125,5 milhões, e desde então as perdas decorrentes da ferrugem-asiática da soja tem sido motivo de preocupação para os produtores (YORINORI et al., 2004). A importância da FAS pode ser avaliada pela sua rápida expansão nas regiões produtoras de soja, por seus danos e pelo elevado custo no controle da doença (aproximadamente 2 bilhões de dólares), e pode ser considerada a principal ameaça à produtividade e competitividade da soja nacional (YORINORI et al., 2004).

A FAS destrói o tecido foliar, resultando em menor atividade fotossintética, desfolha prematura e redução do ciclo de vida da planta. A queda prematura das folhas impede a completa formação de grãos (SINCLAIR, 1989, YORINORI et al., 2005). Por fim, o efeito cumulativo da FAS na produção de soja acaba resultando em uma diminuição do número de vagens e de sementes, além de causar redução no peso das sementes (SINCLAIR, 1989).

O principal método de controle dessa doença é a utilização de fungicidas, porém o excesso de aplicação seleciona populações menos sensíveis, fazendo com que sua eficiência seja reduzida, além de representar um risco real ao meio ambiente. Além da aplicação de fungicidas, o vazio sanitário também tem auxiliado no controle da doença. O período conhecido como vazio sanitário constitui um intervalo de 60 a 90 dias nos quais não se pode semear ou manter plantas vivas de soja no campo. Uma vez que *P. pachyrhizi* é um fungo biotrófico obrigatório, o vazio

sanitário reduz a sobrevivência do fungo durante a entressafra, atrasando a ocorrência da doença na safra seguinte. Assim, o desenvolvimento de cultivares resistentes tem se mostrado uma estratégia viável e necessária. Até o momento, seis locos foram descritos em soja para resistência ao fungo *P. pachyrhizi*, denominados *Rpp1, Rpp2, Rpp3, Rpp4 e Rpp4-b, Rpp5 e Rpp6* (BROMFIELD & HARTWIG, 1980; MCLEAN & BYTH, 1980; HARTWIG & BROMFIELD, 1983; HARTWIG, 1986; GARCIA et al., 2008; LI et al., 2012; KING et al., 2017). No entanto, a estabilidade dessa resistência é duvidosa, devido ao patógeno apresentar uma alta diversidade genética, dificultando o desenvolvimento de cultivares que sejam efetivas por um longo período (HARTMAN et al., 2005). Assim, os genes identificados, conferem resistência a um conjunto limitado de isolados de *P. pachyrhizi* (BONDE et al., 2006; PAUL & HARTMAN, 2009; PHAM et al., 2009) e ainda não há disponível variedades de soja com resistência a todos os isolados do patógeno.

1.3.2 Ciclo de vida de *P. pachyrhizi*

A reprodução de *P. pachyrhizi* é predominantemente, se não exclusivamente, assexual (ANDERSON et al., 2008). O ciclo infeccioso se inicia quando urediniósporos, formas assexuadas de esporos, atingem a superfície superior ou inferior das folhas de soja. Em condições favoráveis de temperatura (entre 12 e 27°C) e umidade, os urediniósporos germinam, produzindo um único tubo germinativo que cresce sobre a superfície da folha até a formação do apressório, estrutura necessária para a penetração da hifa no interior do tecido (ZAMBOLIN, 2006). A penetração ocorre diretamente através da epiderme, ao contrário de outras

espécies de ferrugens que geralmente penetram através dos estômatos. Durante a penetração as células da epiderme entram em colapso e começam a mostrar sinais de desorganização e morte celular (PANSTRUGA, 2003, MENDGEN et al. 2006). A morte celular ocasionada inicialmente pelo fungo, além de facilitar seu ingresso nos espaços intercelulares da folha, também pode ser induzida para evitar a ativação de mecanismos de defesa da planta (HOEFLE et al., 2009).

No interior da folha, uma estrutura chamada de hifa de infecção se desenvolve entre as células epidérmicas e atinge o espaço intercelular. Ao entrarem em contato com as células do mesófilo formam a célula mãe do haustório, que realiza a penetração nas células hospedeiras e a diferenciação do haustório. Este é uma estrutura altamente ramificada, delimitada pela membrana celular da célula hospedeira que proporciona uma ampla superfície de contato entre o fungo e a célula infectada, através da qual são adquiridas quantidades massivas de açúcares e aminoácidos necessários para o desenvolvimento do patógeno (PANSTRUGA, 2003; MORALES et al., 2012). Após cinco a oito dias do início da infecção, novos urediniósporos são formados no interior de estruturas de esporulação denominadas urédias. Os urediniósporos, inicialmente de coloração hialina (cristalina), tornam-se beges e acumulam-se ao redor dos poros ou são carregados pelo vento. À medida que prossegue a esporulação, o tecido da folha ao redor das primeiras urédias adquire coloração castanho-clara (lesão do tipo TAN, observada em genótipos suscetíveis) a castanho-avermelhada (lesão do tipo *Reddish-Brown* – RB, observada em genótipos resistentes), formando as lesões típicas da doença e que são facilmente visíveis em ambas as faces da folha (SINCLAIR, 1989). O processo de infecção causado por fungos de ferrugem, incluindo *P. pachyrhizi* pode ser observado no esquema ilustrado na Figura 1.

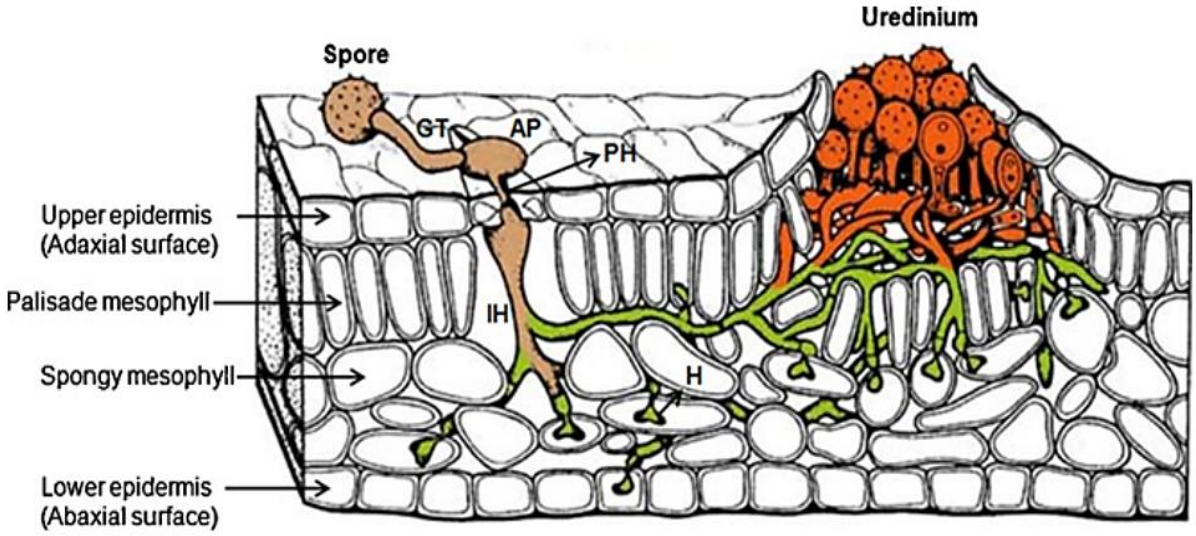


Figura 1. Esquema didático da estrutura interna de uma típica folha de dicotiledônea mostrando o processo de infecção causado por fungos de ferrugem. As estruturas de esporo, tubo germinativo (GT); apressório (AP), hifa de penetração (PH); hifa de infecção (IH); haustório (H), e a estrutura de esporulação (urédia) estão representadas na imagem. Esquema obtido a partir de Hahn (2000).

O ponto de contato mais íntimo entre o *P. pachyrhizi* e a soja é o haustório. Esta estrutura apresenta um papel fundamental na interação planta-patógeno, pois, além de permitir a captação de nutrientes da planta, é capaz de secretar proteínas para o interior da célula hospedeira, denominadas de proteínas efetoras. Essas proteínas efetoras podem ser definidas como moléculas que alteram a estrutura e função da célula hospedeira, podendo atuar como fatores de virulência, levando a alterações que facilitam a infecção, ou de avirulência, pela ativação das respostas de defesa. Em algumas situações uma mesma molécula efetora pode atuar tanto como fator de virulência como também de avirulência.

Quando uma proteína efetora é reconhecida por uma proteína de resistência, codificada por um gene de resistência (R) da planta, observa-se uma interação

incompatível, a qual desencadeia uma cascata de sinalização levando a resistência e morte celular programada da célula do hospedeiro (resposta de hipersensibilidade - HR), impedindo o crescimento do fungo. Neste caso a proteína efetora do fungo é identificada como um fator de avirulência e provoca uma cascata de reações denominadas de ETI (ETI - *effector-triggered immunity*) (GASSMANN & BHATTACHARJEE, 2012). Porém quando uma proteína efetora não é reconhecida por uma proteína de resistência da planta, observa-se uma interação compatível, na qual o fungo é capaz de colonizar a célula hospedeira, levando ao desenvolvimento doença. Neste tipo de interação a proteína efetora do fungo é identificada como um fator de virulência, resultando em reações denominadas de ETS (ETS - *effector-triggered susceptibility*) (JONES & DANGL 2006).

1.3.2 Estudos transcricionais em *P. pachyrhizi*

Fungos causadores de ferrugem constituem um grande grupo de fungos fitopatogênicos, com mais de 7.000 espécies já descritas (CUMMINS & HIRATSUKA, 2003). Porém destas, são poucas as espécies que já possuem seus genomas completamente sequenciados, incluindo *Blumeria graminis*, *Puccinia graminis* f.sp *tritici*, *Melampsora larici-populina, Puccinia striiformis* f.sp. *tritici*, e *Melampsora lini* (SPANU et al., 2010; DUPLESSIS et al., 2011; CANTU et al., 2011; NEMRI et al., 2014). Assim, grande parte das informações disponíveis acerca da biologia, assim como dos mecanismos que atuam durante o processo de infecção desses fungos, é proveniente de análises do transcriptoma desses patógenos. Análises do transcriptoma de uma espécie podem fornecer informações de perfis de expressão gênica e permitem inferir funções a genes específicos.

P. pachyrhizi é uma das espécies de fungos causadores de ferrugem que ainda não possui seu genoma sequenciado, concentrando muitas das informações disponíveis sobre sua biologia e estratégias de infecção, nos estudos de seu transcriptoma. Dentre os estudos que investigaram o transcriptoma de *P. pachyrhizi* expresso em vários estágios do ciclo de vida, podemos destacar o trabalho realizado por Posada-Buitrago e Frederick (2005), que por meio de uma biblioteca de cDNA seguida de sequenciamento Sanger, avaliou a expressão de genes durante a germinação de urediniósporos. Neste estudo foi identificado um total de 488 ESTs únicos (*Expressed Sequence Tag*), cuja maioria correspondeu a genes com função ainda desconhecida na época, e as sequências que apresentaram anotação conhecida, foram relacionadas predominantemente a proteínas putativas envolvidas no metabolismo primário, expressão de genes e proteínas e estrutura celular. Os resultados gerados neste trabalho contribuíram de maneira valiosa para o aumento do número de sequências disponíveis de *P. pachyrhizi*.

Posteriormente, Stone et al. (2012) identificaram 1.029 ESTs (238 ESTs não redundantes) e 119 diferentes proteínas obtidas a partir de amostras de urediniósporos germinados e apressórios. A maioria das sequências identificadas apresentou similaridade com sequências previamente disponíveis de *P. pachyrhizi* de urediniósporos germinados e de amostras de folhas de soja infectadas. Assim 29 ESTs não redundantes foram identificadas exclusivamente na biblioteca de cDNA enriquecida para transcritos especificamente envolvidos na formação de apressório. Dentre essas sequências, algumas apresentaram função desconhecida, e outras apresentaram papéis no metabolismo ou no ciclo celular, o que é esperado levando-se em conta a importância da autofagia e da mitose na formação do apressório.

Sequências expressas especificamente durante a formação do haustório em *P. pachyrhizi* também foram identificadas. Link et al. (2014) identificaram 4.483 ESTs em amostras obtidas de haustório provenientes de folhas de soja infectadas, e as compararam com sequências de urediniósporos em repouso e germinados, visando identificar os transcritos mais representados ou exclusivos no haustório. Como esperado, pela atuação do haustório na absorção de açúcares e aminoácidos, as sequências encontradas no haustório quando comparadas às sequências de urediniósporos em repouso e germinados, apresentaram maior destaque para processos de geração de precursores de metabólitos e energia, aminoácidos celulares e processo metabólico derivativo, metabolismo de carboidratos e transcrição. Neste estudo também foram identificadas sequências possivelmente secretadas pelo haustório na célula hospedeira, semelhantes a sequências encontradas no secretoma de outros fungos causadores de ferrugem, que podem atuar como efetores, envolvidos nos mecanismos de virulência.

Por ser um fungo biotrófico obrigatório, a obtenção apenas de células de *P. pachyrhizi* para o estudo de seu transcriptoma se torna tarefa difícil. Assim, a microdissecção por captura de laser (*Laser Capture Microdissection* - LCM) (EMMERT-BUCK et al., 1996), que permite a dissecção do tecido infectado com precisão suficiente para isolar o organismo e suas estruturas de infecção dentro do hospedeiro, foi utilizada por Trembley et al. (2009) para observar os transcritos de *P. pachyrhizi* expressos durante a formação de urédias. Um total de 130 ESTs foram obtidos, dos quais 117 apresentaram similaridade com sequencias de *P. pachyrhizi* e de outros fungos disponíveis no NCBI. Dentre as sequências anotadas contra o banco de dados de proteínas do NCBI, foram identificadas categorias funcionais relacionadas a metabolismo de modo geral, a processos de transcrição, crescimento

e divisão celular, síntese proteica, biogênese e organização celular, e ainda sequências relacionadas a comunicação e defesa da célula, dentre outros.

Métodos como análise em série da expressão gênica (*Serial Analysis of Gene Expression - SAGE*) (VELCULESCU et al., 1995), microarranjos (SCHENA et al., 1995), sequenciamento de ESTs, dentre outros, têm sido amplamente utilizados no estudo de transcriptomas. Entretanto, a técnica de RNA-Seq fornece uma detecção mais sensível dos transcritos encontrados na célula, e os níveis de expressão detectados apresentam melhor correlação com os níveis de proteínas (WANG et al., 2009; FU et al., 2009). Tal metodologia também foi utilizada por Trembley et al. (2012; 2013) no estudo do transcriptoma de *P. pachyrhizi*. Trembley et al. (2013), identificaram 27.715 contigs expressos no fungo durante quatro diferentes tempos de infecção (representando os estágios de pré-penetração, colonização e esporulação), fazendo deste, até o momento, o estudo de transcriptoma que mais abrange os processos de infecção desenvolvidos por esse patógeno na soja. Os resultados obtidos identificaram genes comuns expressos durante todos os tempos de infecção, como genes relacionados à produção de energia e ao metabolismo de carboidratos. Entretanto, alguns processos foram identificados com maior abundância em momentos específicos do processo de infecção, como o metabolismo de nucleotídeos, que apresentou elevados níveis de expressão durante a germinação de urediniósporos; o metabolismo de aminoácidos e a síntese de proteínas, que foram mais abundantes durante a formação do haustório; e ainda o metabolismo de carboidratos e ácidos graxos, que foram abundantes durante a esporulação.

Apesar de identificar mecanismos importantes da biologia do fungo durante o processo de infecção, os transcritos de *P. pachyrhizi* identificados por Tremblay et

al. (2013) foram obtidos apenas durante o contato do fungo com um genótipo de soja suscetível, desse modo alguns aspectos continuam ainda incompreendidos, como por exemplo, processos moleculares específicos desencadeados no fungo em resposta aos genes de resistência presentes nos genótipos de soja. As informações disponíveis a partir dos poucos estudos existentes sobre o transcriptoma de *P. pachyrhizi*, possibilitam uma visão ainda restrita sobre diversos processos biológicos que atuam na sobrevivência desse patógeno e durante a infecção de seu principal hospedeiro, a soja. Assim a técnica de RNA-Seq aliada a técnica de LCM, constituem uma poderosa ferramenta que pode auxiliar os estudos do transcriptoma de *P. pachyrhizi*.

A identificação de sequências ortólogas tem sido utilizada em abordagens comparativas, permitindo inferir funções específicas a proteínas baseando-se em informações disponíveis de outras espécies e possibilitando análises evolutivas entres as espécies correlacionadas (TATUSOV et al., 1997). Proteínas denominadas homólogas normalmente apresentam um ancestral comum, e podem ser classificadas como parálogas, quando derivam de um evento de duplicação gênica, ou ortólogas, quando divergiram de um ancestral comum devido à especiação (SONNHAMMER & KOONIN, 2002). De modo geral, sequências ortólogas desempenham funções semelhantes, ou até mesmo idênticas, em contrapartida, sequências parálogas podem assumir diferentes papéis funcionais. Nesse sentido, o desenvolvimento da ferramenta OrthoMCL (LI et al.; 2003), um algoritmo de agrupamento, tem possibilitado a realização deste tipo de análises com base na semelhança entre as sequências proteicas de diferentes espécies, auxiliando no entendimento de organismos ainda não sequenciados, cujas informações disponíveis são ainda limitadas. Essa ferramenta tem sido utilizada na identificação

de famílias gênicas em diferentes espécies de fungos, possibilitando a inferência de relações filogenéticas entre organismos distantes (CORNELL et al., 2007; RICHARDS et al., 2009; KELKAR & OCHMAN, 2012; MEERUPATI et al., 2013; OKAGAKI et al., 2016).

Após o sequenciamento dos genomas dos fungos causadores de ferrugem *M. larici-populina* e *Puccinia graminis,* Duplessis et al. (2011), agruparam as proteínas preditas desses fungos em famílias gênicas de acordo com a suas similaridades entre sequências ortólogas e parálogas de um conjunto diversificado de fungos ascomicetos e basidiomicetos. Tal análise possibilitou a identificação de um total de 3.984 famílias ortólogas compartilhadas entre ambas às espécies, compreendendo quase 8.000 genes em cada genótipo, enquanto 774 famílias de genes eram únicas para estes dois fungos de ferrugem. Além de famílias gênicas conservadas e específicas entre as espécies, famílias gênicas expandidas também foram identificadas em Pucciniales, como as que codificam proteases e transportadores de peptídeos, permitindo inferir sobre os mecanismos de assimilação de aminoácidos nestes patógenos, pela clivagem extracelular de proteínas, seguida de sua assimilação. A baixa capacidade de assimilação de nitrogênio e enxofre inorgânico também foi revelada pela analise comparativa, assim como a redução no complemento gênico codificando enzimas ativas no metabolismo de carboidratos e toxinas, a fim de evitar a percepção inicial do hospedeiro e desse modo retardar as respostas de defesa.

Em *P. pachyrhizi*, análises semelhantes, com base na homologia entre sequências do secretoma de outros fungos de ferrugem, também foram utilizadas para a identificação de famílias multigênicas entre sequências preditas do secretoma do fungo, resultando em 596 famílias multigênicas, das quais 13 apresentaram

características consistentes com a função de proteínas efetoras (CARVALHO et al., 2016). Neste trabalho, com o auxílio dessas análises comparativas, 22 transcritos foram selecionados a partir do transcriptoma obtido de *P. pachyrhizi* para análises do perfil de expressão gênica, dos quais 8 se caracterizavam por serem fortes candidatos a efetores. Seis destes candidatos foram submetidos a superexpressão transiente em folhas de tabaco, via vetores de clonagem que exploram o sistema de secreção bacteriano e transformação com *Pseudomonas syringae*, e os resultados obtidos revelaram a capacidade destes genes de suprimir respostas de ETI.

1.3.4 Silenciamento gênico mediado por pequenos RNAs

A regulação da expressão gênica pela ação de pequenos RNAs (*small RNAS - sRNA*) não codificadores foi descrita inicialmente em plantas como um mecanismo de co-supressão da expressão de transgenes (NAPOLI et al, 1990), sendo posteriormente caracterizado como um processo de silenciamento gênico pós-transcricional mediado por pequenos RNAs de interferência (*small interfering RNAs - siRNA*) (HAMILTON & BAULCOMBE, 1999). Tal mecanismo é conservado em eucariotos e a repressão da expressão gênica pode ocorrer por meio da supressão da transcrição (silenciamento de genes a nível transcricional [*Transcriptional Gene Silencing - TGS*]), via modificações epigenéticas na cromatina, ou pela ativação de processos de degradação de mRNA e/ou repressão da tradução (silenciamento gênico pós transcricional [*Post Transcriptional Gene Silencing - PTGS*]). Quando comparados ao controle transcricional, os mecanismos pós-transcricionais regulam a meia-vida do mRNA e as taxas de tradução, assim, ao ajustar as concentrações celulares de proteínas necessárias, promovem mudanças mais rápidas e

permanentes no estado fisiológico da célula (SONENBERG & HINNEBUSCH, 2009; PICHON et al., 2012).

O processo, também conhecido como RNA interferente (RNAi), parece ter surgido inicialmente como um mecanismo de defesa contra vírus de RNA e elementos de transposição, e posteriormente evoluiu para mecanismos de regulação da expressão de genes endógenos, participando de muitos aspectos da fisiologia dos organismos vivos, incluindo o desenvolvimento e a reprogramação transcricional (BARTEL, 2009; AMERES & ZAMORE, 2013; BORGES & MARTIENSSEN, 2015). A ativação da maquinaria de silenciamento se dá pela presença de moléculas de RNA endógenas ou exógenas de cadeia dupla (*double stranded RNA- dsRNA*), sendo que no caso de PTGS, tais moléculas desencadeiam uma resposta citoplasmática, que envolve a identificação e destruição de uma molécula de mRNA alvo de sequência específica. Neste processo estão envolvidas três principais enzimas: Dicer (DCL), Argonauta (AGO) e RNA polimerase dependente de RNA (RdRP).

Dicer é uma proteína multidomínio, que apresenta os domínios RNA helicase dependente de ATP, domínio PAZ, dois domínios RNase III em tandem, e um domínio de ligação a molécula de dsRNA (BERNSTEIN et al., 2001). A ação da enzima Dicer gera os sRNAs, que podem variar de 21 a 28 nucleotídeos (nt), e atuam como guias até o mRNA alvo do silenciamento. Os sRNA gerados associam-se com um complexo de proteínas chamado Complexo de Silenciamento Induzido por RNA (*RNA-Induced Silencing Complex - RISC*), que separa as fitas da molécula de sRNA, gerando duas moléculas de fita simples (*single-stranded RNAs - ssRNAs*) denominadas “guia” e “passageiro”. A molécula de fita simples denominada de “passageiro” é degradada, enquanto a “guia”, que constitui a fita antisense fica livre para o emparelhamento com o mRNA alvo (HUTVAGNER & ZAMORE, 2002). A

molécula “guia” é identificada por uma proteína chave do complexo RISC, a proteína Argonauta, cujo principal domínio é o domínio Piwi, conferindo a essa proteína sua atividade catalítica (HANNON, 2002). A proteína Argonauta, guiada pela molécula de RNA fita simples, reconhece então a molécula correspondente de mRNA pela complementariedade entre as moléculas de RNA, o que conduz a enzima a realizar a clivagem do mRNA alvo.

Em algumas espécies, como por exemplo, em plantas, insetos e animais, pode ocorrer um processo denominado “RNAi transitivo”, que constitui em uma amplificação adicional de sRNAs, gerada pela ação da enzima RdRP. Essas enzimas sintetizam RNAs complementares a sRNAs, a ssRNAs guias, ou mesmo complementares a fragmentos recém gerados das moléculas de mRNA que foram degradadas, formando assim novos dsRNAs que serão novamente clivados pela Dicer, e assim aumentam e disseminam sistematicamente a degradação do mRNA alvo no citoplasma (PARKER et al., 2006; LIU et al., 2003).

1.3.4.1 Silenciamento gênico em plantas

Em plantas, de modo geral, podemos encontrar duas classes de sRNA endógenos, os microRNAs (miRNAs) e siRNAs, ambas derivadas de moléculas de dsRNA clivadas em um ou vários sRNAs por uma das quatro enzimas dicer (DCL) que são encontradas em plantas (BAULCOMBE, 2004). Moléculas de miRNAS são produzidas pela ação das enzimas DCL1 e DCL4, enquanto que os siRNAs são gerados pela ação das enzimas DCL2, DCL3 e/ou DCL4. Os dsRNAs precursores de siRNAs são em sua maioria moléculas quase perfeitas, resultantes de transcritos de repetição invertida (*inverted repeat – IR*), sobreposição de transcrição

convergente, ou a partir da transformação de moléculas de ssRNAs em novas moléculas de dsRNAs, pela ação das enzimas RdRP (PARENT et al., 2012).

Diferentes proteínas argonautas também atuam nas plantas, e são determinantes para a atividade final do complexo de silenciamento. Os miRNAS e siRNAs produzidos com 21 nt, gerados pelas enzimas DCL1 e DCL4, associam-se a proteínas argonautas AGO1, AGO2, AGO7 ou AGO10, que associadas ao complexo RISC, participam do silenciamento por degradação ou por inibição de moléculas de mRNA por meio de processos de PTGS (BAUMBERGER & BAULCOMBE, 2005; BRODERSEN et al., 2008). Já as moléculas geradas pelas enzimas DCL3, que produzem siRNAs de 24 nt, são associadas a proteínas argonautas AGO4, AGO6 ou AGO9, e atuam em mecanismos de TGS (BROSNAN & VOINNET, 2011). As enzimas DCL2 produzem siRNAs de 22 nt que podem atuar como repositores de sRNAs para as moléculas de 21 e 24 nt (GASCIOLLI et al., 2005).

A classe mais abundante de sRNAs em plantas compreende os siRNAs de 24 nt (KASSCHAU & CARRINGTON, 2007), que atuam em mecanismos de silenciamento a nível transcricional, promovendo modificações epigenéticas no DNA, como por exemplo no controle de elementos de transposição (*transposable elements - TEs*), que são abundantes e podem constituir uma ameaça a integridade do genoma, uma vez que promovem inserções, deleções e rearranjos cromossômicos. De modo geral, os siRNAs que inativam TEs são derivados de transposons que foram transcritos pela RNA polimerase IV/V, e desencadeiam a metilação do DNA e assim a modificação da cromatina, o que por sua vez resulta na supressão da transcrição dos TEs (PARENT et al., 2012).

A regulação da expressão gênica por miRNAs, a segunda classe mais abundante de sRNAs em plantas, vai além do controle de genes relacionados ao

desenvolvimento (incluindo divisão meristemática, separação de órgãos, forma de folha, alongamento de raiz secundária, tempo de floração, entre outros) (VOINNET, 2009), uma vez que foram identificados miRNAs com alvos validados envolvidos no metabolismo primário e secundário de plantas (JONES-RHOADES & BARTEL, 2004; SUNKAR & ZHU, 2004). Estes sRNAs participam até mesmo do controle de suas próprias vias, regulando a expressão dos genes *DCL1* e *AGO1*, essenciais para sua biogênese e atividade, revelando a complexa relação existente estre essas moléculas e os mecanismos de regulação gênica existentes em plantas (PARENT et al., 2012). Uma classe de siRNAs endógenos de 21-nt, assim como os miRNAs, também atuam na regulação da expressão gênica por meio de ação *in trans*, e por isso são denominados de tasiRNAs (*trans-acting siRNAs*) (VAZQUEZ et al., 2004). Uma classe final de sRNAs endógenos também estão presentes nas plantas, são os natsiRNAs, siRNAs anti-senso naturais. Essas moléculas são resultantes da ação de enzimas DCL2, que produzem sRNAs de 21/22nt (natsiRNAs) e que podem mediar PTGS (PYOTT & MOLNAR, 2015).

Além dos sRNAs endógenos produzidos pela planta, a presença de moléculas exógenas de dsRNA ou siRNA, derivadas de transgenes e infecções virais, também ativam a maquinaria de PTGS. O silenciamento via sRNAs desempenha um papel natural atuando em mecanismos de defesa antiviral, em um processo chamado de silenciamento gênico induzido por vírus (*Virus-Induced Gene Silencing – VIGS*). A resposta das plantas a infecção viral é amplificada de modo sistêmico pela ação de enzimas RdRP6, garantindo que a maquinaria de silenciamento seja capaz de acompanhar as altas taxas de replicação do vírus, o que pode imunizar as células que estão prestes a ser infectadas e pode resultar na exclusão do vírus (BRODERSEN & VOINNET, 2006). Em plantas, o processo de VIGS é amplamente

conhecido e vem sendo utilizado como uma ferramenta biotecnológica, a partir do desenvolvimento de vetores baseados nos genoma virais, capazes de induzir a supressão da expressão de diferentes genes, quando um fragmento de um gene endógeno é inserido no genoma viral, que durante a infecção da planta produz siRNAs específicos para o gene alvo endógeno, resultando na degradação específica de seus transcritos (ROBERTSON, 2004; BURCH-SMITH et al., 2004).

Nos mecanismos de silenciamento ativados pela presença de sRNAs exógenos, além da amplificação do sinal de silenciamento (transitividade), as enzimas RdRP podem ser de extrema importância também em outros processos. Algumas dessas enzimas (RdRP6) podem atuar no reconhecimento de certos transgenes com características aberrantes, que incluem a falta de metilação na extremidade 5’ do transcrito, utilizando estas sequências como molde para a síntese de sRNAs complementares e consequentemente, resultando no silenciamento do transgene (DALMAY et al., 2000). A Figura 2 contém um esquema geral das principais vias de silenciamento que ocorrem em plantas.

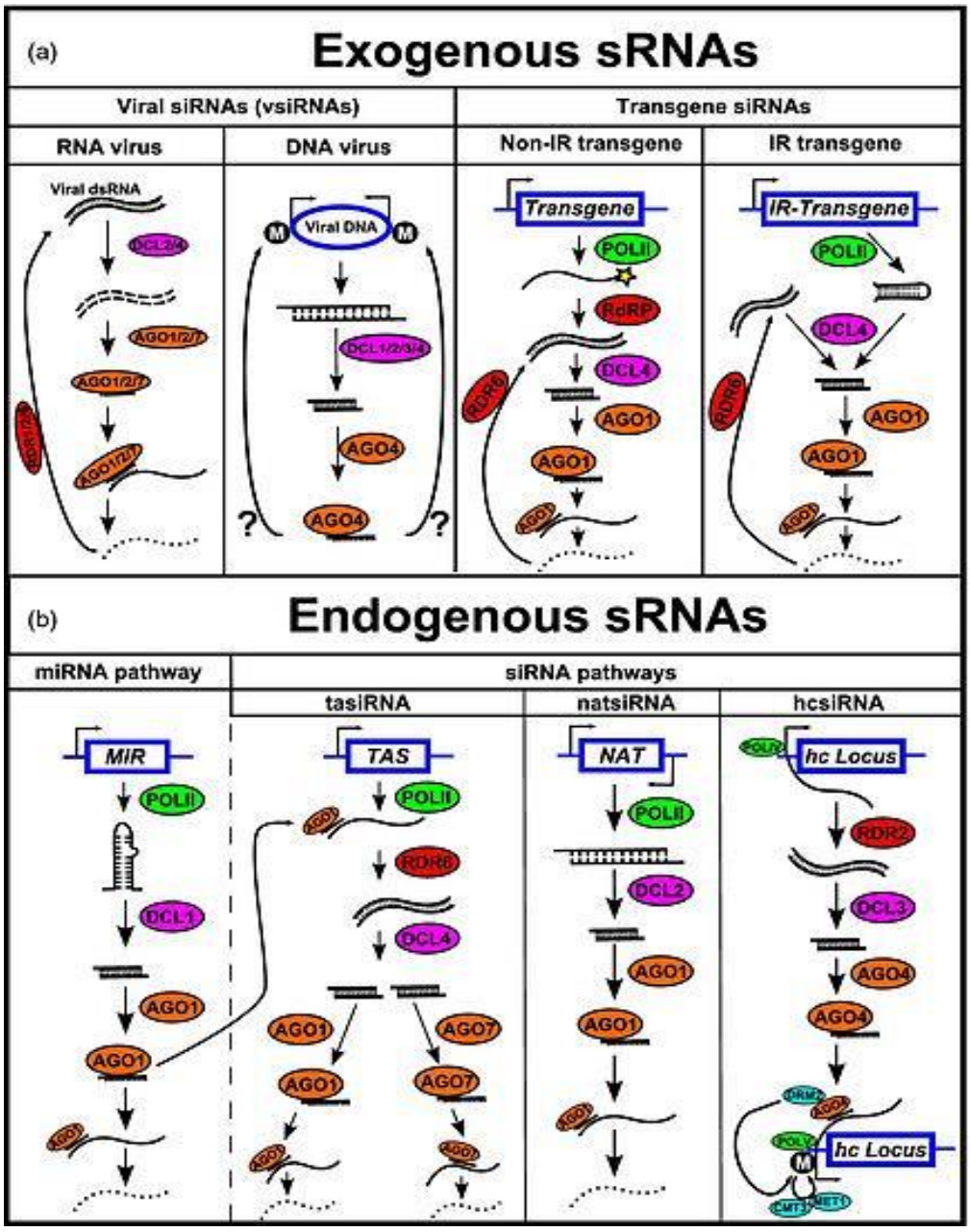


Figura 2. Resumo das principais vias de silenciamento em plantas com a ação das enzimas dicer (DCL), argonauta (AGO) e RNA dependente de RNA polimerase (RDR). (a) Vias de silenciamento iniciadas por DNA ou RNA exógeno. (b) Vias de silenciamento iniciadas a partir de loci endógenos. O DNA/RNA é representado por linhas azuis/pretas, respectivamente; A estrela amarela representa aberrância estrutural; O M em branco indica metilação de citosina. As linhas tracejadas indicam a geração de RNAs por clivagem de dsRNA; as linhas pontilhadas representam a degradação do mRNA. Imagem obtida de Pyott & Molnar (2015).

1.3.4.2 Silenciamento gênico em fungos

Enquanto as rotas de silenciamento gênico são razoavelmente bem conservadas em plantas e animais, tais rotas e maquinaria diferem entre as diferentes espécies de fungos. Especificamente nesses organismos, os aspectos moleculares da maquinaria de RNAi foram analisados minuciosamente em *N. crassa* (CATALANOTTO et al., 2000; COGONI & MACINO, 1997, 1999a, 1999b; ROMANO & MACINO, 1992). Esse fungo constitui um robusto organismo modelo de estudo, com muitas ferramentas genéticas disponíveis que permitiram a clonagem de vários genes envolvidos nessa via e ajudou a desvendar o núcleo principal da maquinaria RNAi em fungos. Assim como em plantas e animais, pelo menos três componentes fundamentais da maquinaria de silenciamento também estão presentes em fungos: as proteínas Dicer, Argonauta e RdRP. Com base nos estudos em *N. crassa*, três principais mecanismos de silenciamento gênico foram descritos: quelling, mutação pontual induzida por repetição (*repeat-induced point mutation - RIP*) e silenciamento meiótico de DNA não pareado (*meiotic silencing by unpaired DNA - MSUD*) (CHANG et al., 2012).

Queling foi o primeiro fenômeno de RNAi descrito em eucariotos e é o mais bem estudado até o momento, foi identificado quando múltiplas copias de genes envolvidos na produção de carotenóides foram inseridas no genoma do fungo *N. crassa*, resultando numa pigmentação albina, ao invés de laranjada, que se pretendia obter (ROMANO & MACINO, 1992). As enzimas envolvidas no processo foram denominadas de QDE-1, QDE-2 e codificam as enzimas RdRP e argonauta respectivamente. Além destas, duas proteínas Dicer com funções redundantes também foram identificadas (DCL-1 e DCL-2), e clivam moléculas de dsRNAs em sRNAs de ~25 nt.

O silenciamento por quelling ocorre durante o ciclo assexual de *N. crassa,* e é ativado pela introdução de sequências de DNA repetitivo, (TEs e sequências virais) que ativam os mecanismos de PTGS para o silenciamento de todos os genes similares a estas sequências. Moléculas aberrantes de RNA (aRNA), geradas a partir da transcrição dessas sequências repetitivas, são inicialmente reconhecidas por uma helicase (QDE-3), e posteriormente processadas pela RdRP (QDE-1), gerando moléculas de dsRNA. Estas são então clivados pela Dicer em siRNA de ~25 nt, e incorporados ao complexo RISC (DANG et al., 2011). O envolvimento das enzimas QDE-2 e Dicer na repressão da transposição foi demonstrado em *N. crassa* por Nolan et al., (2005), onde a transposição de *Tad,* um retrotransposon do tipo LINE, gera repetições invertidas, que podem originar dsRNA pela atuação de QDE1 e QDE-3, ativando a maquinaria de silenciamento.

Adicionalmente, danos no DNA também induzem a expressão da proteína argonauta QDE-2 e uma nova classe de sRNAs associados a QDE-2, chamados qiRNAs (LEE et al., 2009). Os qiRNAs derivam de sequências altamente repetitivas de rDNA, são mais curtos em comprimento (~21 nt) do que o siRNA padrão e requerem componentes semelhantes aos utilizados para o processo de quelling (QDE-1, QDE-3, DCL-1 e DCL-2), indicando que as duas vias são intimamente relacionadas (LEE et al., 2009, 2010a). Ambas as vias compartilham o mesmo sinal inicial, que são sequências altamente repetitivas, embora para quelling essas moléculas sejam sempre de origem exógena (TEs ou vírus), enquanto que para qiRNAs as moléculas são de origem endógena (rDNA repetitivo) (BILLMYRE et al., 2013).

O silenciamento gênico do tipo RIP, antes chamado de rearranjo induzido premeioticamente, é altamente regulado, pois ocorre em um período específico do

ciclo sexual, na fase premeiótica, quando os dois núcleos opostos compartilham um citoplasma comum em células dicarióticas formadas antes da cariogamia (SELKER, 1990). Esse mecanismo é capaz de detectar e mutar sequências duplicadas, com tamanho mínimo de 400 bp (WATTERS et al., 1999), tamanho que se assemelha ao comprimento mínimo requerido para a recombinação em leveduras, sugerindo que mecanismos semelhantes a recombinações podem ativar esse tipo de silenciamento (ROSSIGNOL et al., 1994; SELKER, 1990). As mutações geradas por RIP são geralmente suficientes para inativar completamente os genes alvo do silenciamento, apesar de geralmente ocasionarem alteração de bases de G:C para A:T, com posterior metilação das citosinas remanescentes, o que pode contribuir ainda mais para o silenciamento dos genes (COGONI, 2001). Adicionalmente, essas modificações também podem resultar na expansão de famílias gênicas (GALAGAN & SELKER, 2004)

Acredita-se que o mecanismo de RIP atue principalmente na defesa de fungos contra TEs, tornando-os inativos, e embora algumas evidências apontem para o envolvimento de metilação de citosinas seguida de desaminação (FREITAG et al., 2002), os aspectos moleculares que atuam nesse mecanismo ainda permanecem obscuros, uma vez que a metilação de citosinas ainda não foi relatada durante a fase sexual. Apesar de ainda não ser muito bem descrito, esse mecanismo já foi identificado em diferentes espécies de fungo além de *N. crassa*, como em *Podospora anserina* (GRAIA et al., 2001; BOUHOUCHE et al., 2004; ARNAISE et al., 2008), *Magnaporthe grisea* (IKEDA et al., 2002), *Leptosphaeria maculans* (IDNURM & HOWLETT, 2003), e *Nectria haematococca* (COLEMAN et al., 2009).

O mecanismo de silenciamento MSUD atua como um sistema de vigilância, que silencia de modo transiente os genes não pareados durante o estágio de emparelhamento da profase I na meiose, juntamente com qualquer outro DNA homólogo às sequências não pareadas. Atua também na proteção do genoma contra qualquer fator que possa resultar em segmentos de DNA não pareados, como vírus e transposons em movimento (NICOLÁS & RUIZ-VÁZQUEZ, 2013). O mecanismo de ação é semelhante a outras vias de RNAi, assim, um aRNA é transcrito de regiões de DNA não pareadas e usado como molde para sintetizar dsRNAs, que são processados em pequenos RNAs de interferência associados a MSUD (masiRNAs) (HAMMOND et al., 2013). Entretanto, algumas diferenças são encontradas quando comparado com a via canônica de RNAi, como por exemplo, a presença de dois novos componentes, o gene *sad-1* (*suppressor of ascus dominance*), que codifica uma RdRP, e *sad-2*, responsável pela localização adequada de SAD-1, recrutando essa proteína da região perinuclear (SHIU et al., 2001; SHIU et al., 2006). A dicer DCL-1 é comum entre as vias MSUD e RNAi tradicional, mas a proteína argonauta é codificada pelo gene *sms-2* (*suppressor of meiotic silencing 2*) em MSUD.

Além dos mecanismos descritos acima, a ocorrência de loci que codificam pequenos RNAs similares a microRNAs (*miRNA-like sRNAs – milRNAs*) já foi descrita no genoma de *N. crassa* (LEE et al., 2010b). Esses sRNAs são produzidos por pelo menos quatro diferentes vias de biogênese que utilizam diferentes combinações de componentes, incluindo Dicers, QDE-2, QIP (*QDE-2-interacting protein*) e uma proteína contendo o domínio RNase III (MRPL3). Pelo menos 25 milRNA foram identificados em *N. crassa* e várias evidências indicam que alguns destes sRNAs regulam a expressão gênica de modo similar aos miRNA de plantas e

animais (CHANG et al., 2012). Embora essa classe de sRNAs não tenha sido confirmada em outras espécies de fungos, em *Mycosphaerella graminicola*, fungo patogênico de trigo, outros 65 potenciais loci que codificam milRNA também foram identificados (GOODWIN, et al., 2011).

Assim, a maquinaria de RNAi já foi identificada em diferentes espécies de fungos, incluindo ascomicetos, basidiomicetos e zigomicetos, muitos dos quais possuem múltiplos componentes do silenciamento de RNA em seus genomas, entretanto alguns fungos ascomicetos e basidiomicetos não apresentam alguns componentes, ou mesmo apresentam ausência total da maquinaria de silenciamento (NAKAYASHIKI et al., 2006; LI et al., 2010; HU et al., 2013). Até o momento, em *P. pachyrhizi* foram descritos apenas três contigs semelhantes a sequências de proteínas argonautas (LINK et al., 2014), entretanto nada foi comprovado quanto a funcionalidade dessas sequências e nem de outros componentes que atuam nesse processo, e mesmo o tráfico de moléculas de siRNA entre hospedeiro e patógeno ainda são desconhecidos para esse fungo.

1.3.5 Elementos de transposição no genoma de fungos

O conhecimento relacionado à estrutura e composição dos genomas tem progredido em ritmo acelerado, principalmente com o crescente número de organismos completamente sequenciados. Os chamados elementos de transposição (TEs), fragmentos de DNA que podem se inserir em novas localizações cromossômicas, e muitas vezes fazem cópias duplicadas de si mesmos, podem afetar o genoma de um organismo devido a sua plasticidade, alterando desde estruturas cromossômicas até a expressão gênica. Os TEs têm sido encontrados em

praticamente todas as espécies de organismos eucarióticos investigadas até o momento, incluindo humanos, plantas e insetos (International Human Genome Sequencing Consortium, 2001; International Rice Genome Sequencing Project, 2005; ADAMS et al., 2000). Em alguns organismos, os TEs podem representar a maior parte do genoma, como por exemplo, em plantas, podendo representar 80% ou mais do DNA genômico total, entretanto em outros organismos como fungos, de modo geral representam menos de 50% do genoma, incluindo membros da maioria das superfamílias de TEs (DABOUSSI & CAPY, 2003; HUA-VAN et al., 2005).

Descobertos no milho por Barbara McClintock na década de 50, os elementos de transposição foram classificados inicialmente por Finnegan (1989) em basicamente duas classes, de acordo com seus intermediários na transposição. A primeira classe, denominada de retrotransposons (ou classe I), atua por intermédio de moléculas de RNA por meio do mecanismo conhecido como “copia e cola”, no qual o DNA é transcrito em RNA, e este novamente em DNA pela ação de uma enzima conhecida como transcriptase reversa codificada por um TE. Cada ciclo de replicação completo produz uma nova cópia, assim os retrotransposons são frequentemente os principais contribuintes para a fração repetitiva observada em grandes genomas (KUMAR & BENNETZEN, 1999; HAN & BOEKE, 2005; SABOT & SCHULMAN, 2006). Já a segunda classe, chamada de transposons de DNA (ou classe II), atua sobre o DNA por meio do mecanismo “corta e cola”, no qual a sequência de DNA é excisada da sua posição inicial e inserida numa nova localização, através da ação de uma enzima chamada transposase. Assim, de modo geral, para os elementos da classe I, o transcrito codificado (mRNA) forma o intermediário de transposição, enquanto que para os elementos da classe II, o

elemento em si se move de um local para outro no genoma (WESSLER, 2006; WICKER et al., 2007).

Cada classe de TEs possui elementos autônomos e não autônomos. Os elementos autônomos apresentam ORFs (*open reading frame*) que codificam os produtos necessários para a transposição, ao contrário dos elementos não autónomos, que não codificam proteínas de transposição, mas que são capazes de transpor por conterem sequências “cis” necessárias para transposição. A integração de quase todos os TEs resulta na duplicação de uma sequência genômica curta no local de inserção, chamada de duplicação de um local alvo, ou TSD (*target site duplication*) (WESSLER, 2006).

Os retrotransposons (classe I) podem ser divididos basicamente em duas subclasses com base nos mecanismos de transposição, estrutura e filogenia, de acordo com a classificação proposta por Wicker et al. (2007): uma subclasse de retrotransposons LTR (*long terminal repeat*), e uma subclasse de retrotransposons não-LTR, que inclui principalmente as ordens de longos elementos nucleares intercalados (*long interspersed nuclear elements* - LINEs) e curtos elementos nucleares intercalados (*short interspersed nuclear elements* - SINEs). Recentemente os elementos do tipo Penelope (PLEs) e os elementos do tipo DIRS também foram descobertos. Os retrotransposons LTR, flanqueados por longas sequências terminais, e os não-LTR basicamente diferem entre si pela capacidade de se copiarem para realizar uma nova inserção (LTR), da capacidade de deixar o seu sítio inicial e se inserir em outro local (não-LTR). Menos abundantes em animais, os retrotransposons LTR são a subclasse predominante em plantas, onde parecem ser o principal determinante da grande variação no tamanho do genoma.

Os transposons de DNA (classe II) também apresentam duas subclasses, que se distinguem pela clivagem ou não clivagem de cadeias de DNA durante a transposição (WICKER et al., 2007). A subclasse I, que cliva a dupla fita de DNA, é composta geralmente por uma estrutura simples com uma repetição terminal invertida curta (elementos TIR [*terminal inverted repeat*]) e um único gene que codifica a enzima transposase. Diferente da subclasse I, a subclasse II compreende elementos que sofrem um processo de transposição envolvendo a replicação sem a clivagem da dupla fita de DNA, onde ocorre o deslocamento de apenas uma das fitas da molécula de DNA (elementos Helitrons e Mavericks). Em ambas as classes de TE (retrotransposons e transposons de DNA), existem inúmeras superfamílias que compreendem diferentes tipos de elementos de transposição.

Os TEs em fungos foram identificados por uma variedade de estratégias, principalmente por meio da caracterização de sequências repetitivas dispersas ou pela sua presença em um gene alvo, e à medida que segmentos de genomas começaram a ser clonados e sequenciados, observou-se um aumento significativo na descoberta de novos TEs (CAMBARERI et al., 1998; HUA-VAN et al.; 2000). Elementos de transposição já foram encontrados em três ordens de fungos, Basidiomicetos, Zigomicetos e Ascomicetos, mas a maioria foi identificada nesta última ordem, provavelmente devido ao numero de estudos envolvendo espécies consideradas modelo pertencentes a essa ordem, como por exemplo, *N. crassa* (DABOUSSI & CAPI, 2003).

No caso especial dos fungos causadores de ferrugem, o sequenciamento de diferentes espécies tem revelado o elevado tamanho de seus genomas, normalmente relacionado ao estilo de vida biotrófico (CANTU et al., 2011; DUPLESSIS et al., 2011). Uma estimativa do tamanho do genoma de 32 espécies

de ferrugem pela análise de citometria de fluxo, revelou que as ferrugens apresentam os maiores tamanhos de genoma entre os fungos, variando de 6.9 e 77.4 Mbp para *P. triticina* e *P. graminis f. sp. tritici*, respectivamente, a 893.2 Mbp em *Gymnosporangium confusum*, com uma média global de 305.5 Mbp. Tal tamanho é significativamente superior ao encontrado em qualquer outra ordem do reino Fungi que tenha três ou mais genomas sequenciados. Em especial, no caso de *P. pachyrhizi*, o genoma foi estimado em 720 Mbp (TAVARES et al, 2014).

Com base no sequenciamento de espécies de fungos causadores de ferrugem disponíveis, ficou evidente que o elevado tamanho dos genomas não implica em elevado número de genes estruturais, mas, invariavelmente, como resultado da proliferação de TEs e DNA repetitivo (CANTU et al., 2011; DUPLESSIS et al., 2011). Tal ambiente genômico torna-se uma fonte de variabilidade relevante, especialmente em espécies sem ciclo sexual (SPANU, 2012). Nos fungos *M. larici-populina* e *P. graminis* f. sp. *tritici*, cujos genomas haploides são de 101,1 Mb e 88,6 Mb, respectivamente, a presença de TEs, foi de 45%, em ambas as espécies (DUPLESSIS et al., 2011).

Interessantemente, a organização de algumas famílias gênicas em regiões ricas em transposons tem sido descrita em fungos. Evidências indicam que genes efetores em fungos filamentosos fitopatogênicos, estão geralmente localizados próximos a elementos de transposição (TEs), ou regiões do genoma ricas em TEs, como cromossomos dispensáveis ou telômeros (SCHMIDT & PANSTRUGA, 2011), o que já foi observado em espécies como *Magnaporthe oryzae* (ORBACH et al., 2000), *Phytophthora infestans* (HAAS et al., 2009), *Fusarium oxysporum* (MA et al., 2010), e *Leptosphaeria maculans* (BALESDENT et al., 2013). Tal localização preferencial destes genes neste ambiente gênico pode trazer vantagens seletivas,

permitindo uma resposta rápida de seleção de genes de resistência gerada por mutações e/ou recombinação (DAVERDIN et al., 2012), como ocorre em *P. infestans* (RAFFAELE et al., 2010). Especificamente para *P. pachyrhizi*, transposons e retrotransposons foram identificados em análises de transcriptoma por Link et al. (2014), entretanto muito pouco foi descrito sobre a classificação destes elementos e sobre o papel que desempenham no genoma e na biologia desse fungo.

1.3.6 Silenciamento gênico induzido pelo hospedeiro como ferramenta para a validação de genes em fitopatógenos

Uma vez identificada a maquinaria de silenciamento em diferentes espécies de fitopatógenos, os mecanismos de RNAi vem sendo cada vez mais utilizados com sucesso como uma valiosa ferramenta para o estudo funcional de genes nestes organismos. O silenciamento gênico induzido pelo hospedeiro (*Host-Induced Gene Silencing - HIGS*) constitui uma ferramenta potente semelhante à abordagem empregada no processo de VIGS, entretanto na metodologia de HIGS os genes alvo do silenciamento são provenientes de patógenos e pragas que infectam a planta. Nessa metodologia, o RNA de silenciamento pode ser proveniente de diferentes fontes, como por exemplo, do resultado da expressão de um vetor viral na planta hospedeira, ou da expressão em plantas geneticamente modificadas de um dsRNA homólogo ao mRNA alvo no patógeno, e mesmo da aplicação direta de moléculas de dsRNA ou siRNA direcionadas ao silenciamento. Assim, fragmentos complementares aos transcritos de agentes patogênicos são produzidos no citoplasma das células da planta, e são adquiridos pelo patógeno durante a infecção,

ativando então a maquinaria de RNAi presente nas células destes organismos (BAULCOMBE, 2015).

A metodologia de HIGS tem mostrado que diferentes organismos que crescem no interior ou que desenvolvem contato íntimo com plantas hospedeiras são sensíveis a sRNAs gerados por seus hospedeiros e que são direcionados ao silenciamento de transcritos dos patógenos, como já foi descrito em bactérias (ESCOBAR et al., 2001, 2002), nematoides (HUANG et al., 2006), insetos (BAUM et al., 2007; MAO et al., 2007), e espécies parasitas de plantas (TOMILOV et al., 2008). Essa abordagem também tem se revelado uma estratégia valiosa na identificação de genes de fungos e oomicetos relacionados ao processo de infecção em plantas hospedeiras, incluindo espécies de fungos causadores de ferrugem (TINOCO et al., 2010; NOWARA et al., 2010; ZHANG et al., 2012; KOCH et al., 2013; PLIEGO et al., 2013; PANWAR et al., 2013a, 2013b; GHAG et al., 2014; HU et al., 2015; YIN et al., 2011, 2015; CHENG et al., 2015; ANDRADE et al., 2015; JAHAN et al., 2015; SANJU et al., 2015; GOVINDARAJULU et al., 2015; SONG & THOMMA, 2016; CHEN et al., 2016; ZHANG et al., 2016; ZHOU et al., 2016; COOPER & CAMPBELL, 2017).

O primeiro estudo envolvendo HIGS na análise de fitopatógenos filamentosos, foi realizado por Tinoco et al. (2010), por meio do silenciamento do gene *gus*, previamente inserido no genoma do fungo *Fusarium verticillioides*, causador da podridão de orelha observada em grãos de milho. Os resultados obtidos neste trabalho revelaram que a expressão do transgene *gus* no fungo foi abolida após o silenciamento pela indução de células miceliais em plantas de tabaco transgênicas, manipuladas para expressar siRNAs a partir de um dsRNA correspondente ao transgene. Este estudo inicial revelou que a utilização da via de RNAi poderia

constituir uma poderosa ferramenta para a regulação específica da expressão gênica e demonstrou ser de grande importância nos estudos moleculares básicos e aplicados da interação entre plantas e patógenos.

Ainda no mesmo ano, Nowara et al. (2010) realizaram os primeiros estudos utilizando essa metodologia para o silenciamento de genes em um fungo causador de ferrugem. Neste trabalho, 76 genes candidatos de *Blumeria graminis* expressos *in planta* durante a interação do fungo com a cevada, foram testados através do silenciamento por HIGS utilizando um vetor BSMV para a transformação de plantas de cevada. Dentre os genes testados, o silenciamento de 16 genes resultou em uma redução significativa da porcentagem de conídios do fungo. Dentre os 16 genes que apresentaram resultados significativos, encontram-se sequencias semelhantes a enzima 1,3-b-glucanosiltransferase, transportadores de ADP/ATP, proteínas de choque térmico, NADH oxidoredutase, protease vacuolar, receptor de carga endossomal, proteína ribossomal 40S, e proteínas sem função conhecida.

Posteriormente, Pliego et al. (2013) confirmaram a participação de oito genes candidatos a efetores em *B. graminis* no desenvolvimento do fungo durante o processo de infecção, dentre os quais foram identificadas sequências relacionadas a β-1,3-glucosiltransferases e metalo-proteases, e sequências que codificam ribonucleases, também utilizando a metodologia HIGS.

Subsequentemente, outros estudos envolvendo fungos de ferrugem demonstraram a funcionalidade de HIGS na supressão da infecção causadas por esses patógenos. Em *P. striiformis*, um dos fungos causadores de ferrugem em trigo, a metodologia foi validada por Yin et al. (2011), que apesar de não ter observado redução significativa nos sintomas da doença, observou redução dos níveis de expressão de seis genes haustoriais, que codificam proteínas similares a proteínas

chitinases, a transportadores de hexose, e a proteínas secretadas sem função conhecida. Neste trabalho, os autores sugerem que o silenciamento gênico possa ser limitado a genes expressos em células haustoriais, o que restringiria a seleção de genes para a caracterização funcional por meio dessa abordagem.

Diferentes tipos de vetores virais podem ser utilizados para a expressão transiente de fragmentos de interesse em plantas hospedeiras, e recentemente um vetor baseado no genoma vírus BPMV (*Bean pod mottle vírus*) foi desenvolvido para aplicação de VIGS em plantas de feijão e soja (ZHANG et. al., 2009; ZHANG et. al., 2010). O BPMV é um vírus de RNA fita simples de cadeia positiva, membro do gênero Comovirus e tem um genoma bipartido, encapsulado separadamente em partículas isométricas, constituído por RNA1 (aproximadamente 6,0 kb), que codifica cinco proteínas essenciais para a replicação e manutenção viral, e RNA2 (aproximadamente 3,6 kb), que codifica uma possível proteína de movimentação célula a célula (MP) e duas proteínas de revestimento (L-CP e SCP) (ZHANG et al., 2009). O gene de interesse é inserido na estrutura de leitura aberta da poliproteína codificada por RNA2, entre as regiões de codificação MP e L-CP, e construindo locais adicionais de clivagem de proteinase para flanquear a proteína alvo. O vetor BPMV-VIGS foi utilizado com sucesso para o estudo de genes de soja envolvidos na resistência à ferrugem asiática da soja (ASR) (MEYER et. al., 2009; PANDEY et al., 2011).

O vetor BPMV foi utilizado também através da metodologia de HIGS para o silenciamento de seis genes candidatos a efetores em *U. appendiculatus* (COOPER & CAMPBELL, 2017). Após o silenciamento os autores observaram redução significativa dos sintomas da doença em plantas de feijão infectadas com o fungo para quatro dos seis genes silenciados, que incluem sequências relacionadas a

proteínas trealose fosfatase, do tipo quitinase, hidrolase glicosídica, e uma proteína de função desconhecida. Com os resultados deste trabalho os autores concluíram que a inibição destas quatro proteínas pode conferir proteção às plantas de feijão contra a infecção pelo fungo de ferrugem *U. appendiculatus.* Até o momento a metodologia de HIGS ainda não foi aplicada em *P. pachyrhizi* para a caracterização funcional de genes do patógeno, mas mediante os resultados obtidos em outros fungos causadores de ferrugem, essa abordagem, aliada à utilização do vetor BPMV, se destaca como uma estratégia promissora para o estudo funcional de genes desse fungo através do silenciamento gênico pós transcricional.

Em relação aos mecanismos pelos quais fungos causadores de ferrugem podem adquirir moléculas de siRNAs produzidas no tecido de seus hospedeiros, Panwar et al. (2013b), estudando o silenciamento gênico via RNAi em *Puccinia triticinia*, causador da ferrugem da folha trigo, propuseram esquematicamente como esse transporte pode ocorrer (Figura 3). Moléculas fúngicas de dsRNA, produzidas dentro da célula hospedeira, são clivadas pela maquinaria da planta (enzimas Dicer) em moléculas de siRNAs. Proteínas da membrana plasmática e outros materiais específicos destinados à secreção extracelular são internalizados por endocitose, capturando pequenos RNAs ao longo do caminho para formar endossomos, que subsequentemente se compartimentalizam como vesículas intraluminais dentro de corpos multivasculares. Estes por sua vez, podem fundir-se com lisossomos para permitir a degradação das vesículas intraluminais ou seguir uma via retrógrada, fundindo-se com a membrana plasmática e liberando as vesículas intraluminais como exossomos para o espaço paramural. Esse espaço atua como um nexo de comunicação e transporte entre células hospedeiras e patogênicas e seus respectivos ambientes externos. A passagem através da parede celular haustorial,

seja ativa ou passiva, ocorre, e uma vez dentro do haustório, as moléculas de siRNAs desencadeiam o silenciamento de seus mRNA alvos por meio da maquinaria de RNAi do fungo (PANWAR et al., 2013b).

Apesar de Panwar et al. (2013b) sugerir que essa aquisição das moléculas de siRNA ocorra durante a fase de desenvolvimento do haustório no fungo, Hückelhoven et al. (2000) demonstrou que vesículas de secreção como exossomos são também observadas durantes os primeiros contatos do fungo com a planta hospedeira, durante o desenvolvimento do tubo germinativo, como foi observado durante a infecção da cevada por *B. graminis*.

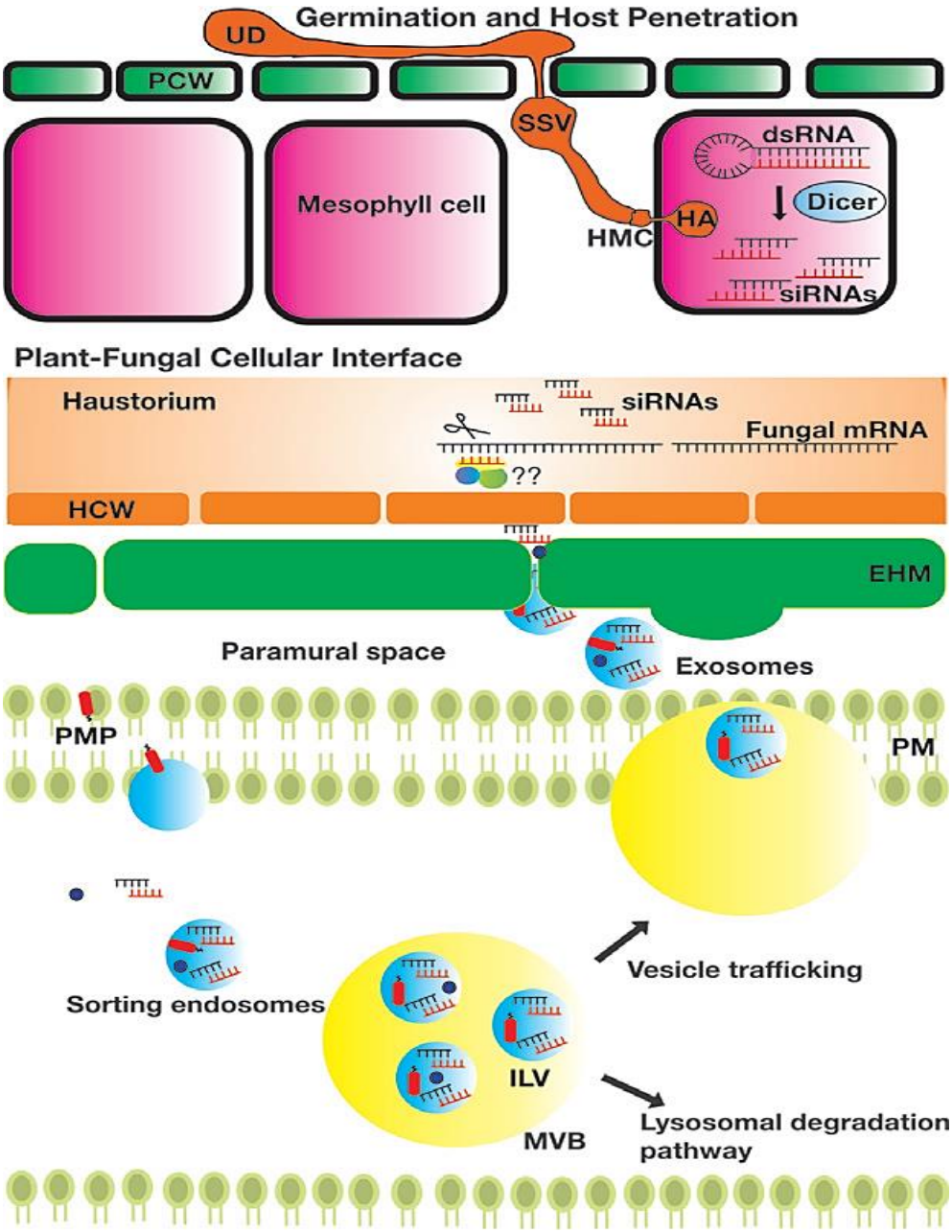


Figura 3. Esquema didático de possíveis mecanismos de silenciamento gênico (PTGS) em fungos induzidos *in planta*. Moléculas de siRNA geradas em células de plantas hospedeiras podem ser enviadas para células do fungo. Membrana plasmática (PM); proteínas da membrana plasmática (PMPs); vesículas intraluminais (ILVs); corpos multivasculares (MVBs); célula mãe de haustório (HMC); parede celular haustorial (HCW); membrana de hifas extra-invasiva (EHM); parede celular da planta (PCW); haustório (HÁ); vesícula subestomática (SSV); urediniósporo (UD). Imagem obtida a partir de Panwar et al. (2013b).

1.4 REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, M. CELNIKER, S.E.; HOLT, R.A.; EVANS, C.A.; et al. The genome sequence of *Drosophila melanogaster*. **Science**, v. 287, p. 2185–2195, 2000.

AMERES, S.L. & ZAMORE, P.D. Diversifying microRNA sequence and function. **Nature Review Molecular Cell Biology**, v. 14, p. 475–488, 2013.

ANDERSON, S.J.; STONE, C.L.; POSADA-BUITRAGO, M.L.; BOORE, J.L.; NEELAM, B.A; STEPHENS, R.M.; LUSTER, D.G.; FREDERICK. R.D.; PEDLEY, K.F. Development of simple sequence repeat markers for the soybean rust fungus, *Phakopsora pachyrhizi*. **Molecular Ecology Resources**, v. 8, p. 1310–1312, 2008.

ANDRADE, C.; TINOCO, M.; RIETH, A.; MAIA, F.; ARAGÃO, F. Host-induced gene silencing in the necrotrophic fungal pathogen *Sclerotinia sclerotiorum*. **Plant Pathology**, v. 65, p. 626–632, 2015.

ARNAISE, S.; ZICKLER, D.; BOURDAIS, A.; DEQUARD-CHABLAT, M.; DEBUCHY, R. Mutations in mating-type genes greatly decrease repeat-induced point mutation process in the fungus *Podospora anserina*. **Fungal Genetics Biology**, v. 45, p. 207-220, 2008.

BALESDENT, M.H.; FUDAL, I.; OLLIVIER, B.; BALLY, P.; GRANDAUBERT, J.; EBER, F.; CHÈVRE, A.M.; LEFLON, M.; ROUXEL, T. The dispensable chromosome

of Leptosphaeria maculans shelters an effector gene conferring avirulence towards Brassica rapa. **New Phytologist**, v. 198, p. 887-898, 2013.

BARTEL, D.P. MicroRNAs: target recognition and regulatory functions. **Cell**, v. 136, p. 215-233, 2009.

BAULCOMBE, D.C. VIGS, HIGS and FIGS: small RNA silencing in the interactions of viruses or filamentous organisms with their plant hosts. **Current Opinion in Plant Biology**, v. 26, p. 141-146, 2015.

BAULCOMBE, D. RNA silencing in plants. **Nature**, v. 431, p. 356-363, 2004.

BAUM, J.A.; BOGAERT, T.; CLINTON, W.; HECK, G.R.; FELDMANN, P.; ILAGAN, O.; JOHNSON, S.; PLAETINCK, G.; MUNYIKWA, T.; PLEAU, M. Control of coleopteran insect pests through RNA interference. **Nature Biotechnology**, v. 25, p. 1322-1326, 2007.

BAUMBERGER, N. & BAULCOMBE, D.C. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, p. 11928-11933, 2005.

BERNSTEIN, E.; CAUDY, A. A.; HAMMOND, S. M.; HANNON, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. **Nature**, v. 409, p. 363–366, 2001.

BILLMYRE, R.B.; CALO, S.; FERETZAKI, M.; & WANG, X.; HEITMAN, J. RNAi function, diversity, and loss in the fungal kingdom. **Chromosome Research**, v. 21, p.561–572, 2013.

BONDE, M.R.; NESTER, S.E.; AUSTIN, C.N.; STONE, C.L.; FREDERICK, R.D.; HARTMAN, G.L.; MILES, M.R. Evaluation of virulence of *Phakopsora pachyrhizi* and *P. meibomiae* isolates. **Plant Disease**, v. 90, p. 708–716, 2006.

BORGES, F. & MARTIENSSEN, R.A. The expanding world of small RNAs in plants. **Nature Review Molecular Cell Biology**, v. 16, p. 727–741, 2015.

BOUHOUCHE, K.; ZICKLER, D.; DEBUCHY, R.; ARNAISE, S. Altering a gene involved in nuclear distribution increases the repeat-induced point mutation process in the fungus *Podospora anserina*. **Genetics**, v. 167, p. 151-159, 2004.

BRODERSEN, P. & VOINNET, O. The diversity of RNA silencing pathways in plants. **Trends in Genetics**, v. 22, n. 5, p. 268–280, 2006.

BRODERSEN, P.; SAKVARELIDZE-ACHARD, L.; BRUUN-RASMUSSEN, M.; DUNOYER, P.; YAMAMOTO, Y.Y.; SIEBURTH, L.; VOINNET, O. Widespread translational inhibition by plant miRNAs and siRNAs. **Science**, v. 320, p. 1185-1190, 2008.

BROMFIELD, K.R. & HARTWIG, E.E. Resistance to soybean rust and mode of inheritance. **Crop Science**, v. 20, p. 254–255, 1980.

BROSNAN, C.A. & VOINNET, O. Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications. **Current Opinion Plant Biology**, v. 14, p. 580-587, 2011.

BURCH-SMITH, T.M.; ANDERSON, J.C.; MARTIN, G.B.; DINESH-KUMAR, S.P. Applications and advantages of virus-induced gene silencing for gene function studies in plants. **Plant Journal**, v. 39, p. 734–746, 2004.

CAMBARERI, E.B.; AISNER, R.; CARBON, J. Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. **Molecular Cell Biology**, v. 18, p. 5465–77, 1998.

CANTU, D.; GOVINDARAJULU, M.; KOZIK, A.; et al. Next Generation Sequencing Provides Rapid Access to the Genome of *Puccinia striiformis* f. sp. *tritici*, the Causal Agent of Wheat Stripe Rust. **Plos One**, v. 6, e24230, 2011.

CARVALHO, M.C.C.G.; NASCIMENTO, L.C.; DARBEN, L.M.; et al. Prediction of the *in planta P. pachyrhizi* secretome and potential effector families. **Molecular Plant Pathology**, 2016.

CATALANOTTO, C.; AZZALIN, G.; MACINO, G.; COGONI, C. Transcription: Gene silencing in worms and fungi. **Nature**, v. 404, p. 245, 2000.

CHANG, S.S.; ZHANG, Z.; LIU, Y. RNA Interference Pathways in Fungi: Mechanisms and Functions. **Annual Review of Microbiology**, v. 66, p. 305-323, 2012.

CHEN, W.; KASTNER, C.; NOWARA, D.; OLIVEIRA-GARCIA, E.; RUTTEN, T.; ZHAO, Y.; DEISING, H.B.; KUMLEHN, J.; SCHWEIZER, P. Host-induced silencing of Fusarium culmorum genes protects wheat from infection. **Journal of Experimental Botany**, v. 67, p. 4979–4991, 2016.

CHENG, W.; SONG, X.S.; LI, H.P.; CAO, L.H.; SUN, K.; QIU, X.L.; XU, Y.B.; YANG, P.; HUANG, T.; ZHANG, J.B. Host-induced gene silencing of an essential chitin synthase gene confers durable resistance to *Fusarium* head blight and seedling blight in wheat. **Plant Biotechnology Journal**, v. 13, p. 1335–1345, 2015.

COGONI, C. Homology-dependent Gene Silencing Mechanisms in Fungi. **Annual Reviews of Microbiology**, v. 55, p. 381-406, 2001.

COGONI, C. & MACINO, G. Posttranscriptional gene silencing in *Neurospora* by a RecQ DNA helicase. **Science**, v. 286, p. 2342–2344, 1999a.

COGONI, C. & MACINO, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. **Nature**, v. 399, p. 166–169, 1999b.

COGONI, C. & MACINO, G. Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, p. 10233–10238, 1997.

COLEMAN, J.J.; ROUNSLEY, S.D.; RODRIGUEZ-CARRES, M.; KUO, A.; WASMANN, C.C.; GRIMWOOD, J.; et al. The Genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. **PloS Genetics**, v. 5, e1000618, 2009.

CONAB. **Acompanhamento da safra brasileira grãos**, v.4 – Safra 2016/17, n.6 – Sexto Levantamento, Brasília, p. 1-176, Março de 2017.

COOPER, B & CAMPBELL, K.B. Protection against Common Bean Rust Conferred by a Gene Silencing Method. **Phytopathology**, p. 1-24, 2017.

CORNELL, M.J.; ALAM, I.; SOANES, D.M.; WONG, H.M.; HEDELER, C.; PATON, N.W.; RATTRAY, M.; HUBBARD, S.J.; TALBOT, N.J.; OLIVER, S.G. Comparative genome analysis across a kingdom of eukaryotic organisms: Specialization and diversification in the Fungi. **Genome Research**, v.17, n. 12, p. 1809-1822, 2007.

CUMMINS, G.B. & HIRATSUKA, Y. **Illustrated Genera of Rust Fungi**. St Paul, MN: American Phytopathological Society, 2003.

DABOUSSI, M. & CAPY, P. Transposable elements in filamentous fungi. **Annual Review of Microbiology**, v. 57, p. 275–299, 2003.

DALMAY, T.; HAMILTON, A.; RUDD, S.; ANGELL, S.; BAULCOMBE, D.C. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. **Cell**, v. 101, p. 543-553, 2000.

DANG, Y.; YANG, Q.; XUE, Z.; LIU, Y. RNA interference in fungi: pathways, functions and applications. **Eukaryotic Cell**, v. 10, p. 1148-1155, 2011.

DAVERDIN, G.; ROUXEL, T.; GOUT, L.; AUBERTOT, J.N.; FUDAL, I.; MEYER, M.; PARLANGE, F.; CARPEZAT, J.; BALESDENT, M.H. Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. **PLoS Pathogens**, v. 8, e1003020, 2012.

DUPLESSIS, S.; CUOMO, C.A.; LIN, Y.C.; AERTS, A.; TISSERANT, E.; et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 22, p. 9166–9171, 2011.

EMBRAPA-**Tecnologias de produção de soja - Região Central do Brasil 2012 e 2013.** Sistemas de Produção / Embrapa Soja, ISSN 2176-2902; n.15, 2011.

EMMERT-BUCK, M.R.; BONNER, R.F.; SMITH, P.D.; CHUAQUI, R.F.; et al. Laser Capture Microdissection. **Science**, v. 274, p. 998-1001, 1996.

ESCOBAR, M.A.; LESLIE, C.A.; MCGRANAHAN, G.H.; DANDEKAR, A.M. Silencing crown gall disease in walnut (*Juglans regia* L.). **Plant Science**, v. 163, p. 591-597, 2002.

ESCOBAR, M.A.; CIVEROLO, E.L.; SUMMERFELT, K.R.; DANDEKAR, A.M. RNAi-mediated oncogene silencing confers resistance to crown gall tumorigenesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, p. 13437-13442, 2001.

FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends Genetics**, v. 5, p. 103–107, 1989.

FREIRE, M.C.M.; OLIVEIRA, L.O.; ALMEIDA, A.M.R.; SCHUSTER, I.; MOREIRA, M.A.; LIEBENBERG, M.M.; MIENIE, C.M.S. Evolutionary history of *Phakopsora pachyrhizi* (the Asian soybean rust) in Brazil based on nucleotide sequences of the internal transcribed spacer region of the nuclear ribosomal DNA. **Genetics and Molecular Biology**, v. 31, p. 920-931, 2008.

FREITAG, M.; WILLIAMS, R.L.; KOTHE, G.O.; SELKER, E.U. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, p. 8802-8807, 2002.

FU, X.; FU, N.; GUO, S.; YAN, Z.; XU, Y.; HU, H.; MENZEL, C.; CHEN, W.; LI,Y.; ZENG, R.; KHAITOVICH, P. Estimating accuracy of RNA-Seq and microarrays with proteomics. **BMC Genomics**, v. 10, p.161, 2009.

GALAGAN, J.E. & SELKER, E.U. RIP: the evolutionary cost of genome defense. **Trends Genetics**, v. 20, p. 417-423, 2004.

GARCIA, A.; CALVO, E.S.; SOUZA KIIHL, R.A.; HARADA, A.; HIROMOTO, D.M.; VIEIRA, L.G. Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. **Theoretical and Applied Genetics**, v. 117, p. 545–553, 2008.

GASCIOLLI, V.; MALLORY, A.C.; BARTEL, D.P.; VAUCHERET, H. Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. **Current Biology**, v. 15, p. 1494-1500, 2005.

GASSMANN, W. & BHATTACHARJEE, S. Effector-Triggered Immunity Signaling: From Gene-for Gene Pathways to Protein-Protein Interaction Networks. **Molecular Plant-Microbe Interactions Journal**, v. 25, n. 7, p. 862-868, 2012.

GBIF. Global Biodiversity Information Facility. Disponível em: http://www.gbif.org/species/2514154. Acesso em Março de 2017.

GHAG, S.B.; SHEKHAWAT, U.K.; GANAPATHI, T.R. Host-induced posttranscriptional hairpin RNA-mediated gene silencing of vital fungal genes confers efficient resistance against *Fusarium* wilt in banana. **Plant Biotechnology Journal**, v. 12, p. 541–553, 2014.

GOODWIN, S.B.; BEN M'BAREK, S.; DHILLON, B.; WITTENBERG, A.H.J.; CRANE, C.F.; et al. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. **PLoS Genetics**, v. 7, e1002070, 2011.

GOVINDARAJULU, M.; EPSTEIN, L.; WROBLEWSKI, T.; MICHELMORE, R.W. Host-induced gene silencing inhibits the biotrophic pathogen causing downy mil- dew of lettuce. **Plant Biotechnology Journal**, v. 13, p. 875-883, 2015.

GRAIA, F.; LESPINET, O.; RIMBAULT, B.; DEQUARD-CHABLAT, M.; COPPIN, E.; PICARD, M. Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. **Molecular Microbiology**, v. 40, n. 3, p. 586-595, 2001.

HAAS, B.J.; KAMOUN, S.; ZODY, M.C.; JIANG, R.H.; HANDSAKER, R.E.; et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. **Nature**, v. 461, p. 393-398, 2009.

HAHN, M. The rust fungi. Cytology, physiology and molecular biology of infection. In: KRONSTADT, J. (Ed.). **Fungal Pathology**, Kluwer Academic Publisher, Dordrecht, p. 267–306. 2000.

HAMILTON, A.J. & BAULCOMBE, D.C. A species of small antisense RNA in posttranscriptional gene silencing in plants. **Science**, v. 286, p. 950-952, 1999.

HAMMOND, T.M.; XIAO, H.; BOONE, E.C.; DECKER, L.M.; LEE, S.A.; PERDUE, T.D.; PUKKILA, P.J.; SHIU, P.K.T. Novel proteins required for meiotic silencing by unpaired DNA and siRNA generation in *Neurospora crassa*. **Genetics**, v. 1, p. 91-100, 2013.

HAN, J. S. & BOEKE, J. D. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? **Bio Essays**, v. 27, p. 775–784, 2005.

HANNON, G.J. RNA interference. **Nature**, v. 418, p. 244–51, 2002.

HARTMAN, G.L.; MILES, M.R.; FREDERICK, R.D. Breeding for Resistance to Soybean Rust. **Plant Disease**, v. 89, p. 664-666, 2005.

HARTWIG, E.E. & BROMFIELD, K.R. Relationships among three genes conferring specific resistance to rust in soybeans. **Crop Science**, v. 23, p. 237–239, 1983.

HARTWIG, E.E. Identification of a fourth major gene conferring resistance to soybean rust. **Crop Science**, v. 26, p. 1135–1136, 1986.

HENNINGS, V.P. A few new Japanese Uredinaceae. **Hedwigia**, v. 42, p. 107-108, 1903.

HOEFLE, C.; LOEHRER, M.; SCHAFFRATH, U.; FRANK, M.; SCHULTHEISS, H.; HUCHELHOVEN, R. Transgenic suppression of cell death limits penetration success of the soybean rust fungus *Phakopsora pachyrhizi* into epidermal cells of barley. **Phytopathology**, v. 99, p. 220-226, 2009.

HU, Y.; STENLID, J.; ELFSTRAND, M.; OLSON, A. Evolution of RNA interference proteins dicer and argonaute in Basidiomycota. **Mycologia**, v. 105, n. 6, p. 1489–1498, 2013.

HU, Z.; PAREKH, U.; MARUTA, N.; TRUSOV, Y.; BOTELLA, J.R. Down-regulation of *Fusarium oxysporum* endogenous genes by host-delivered RNA interference enhances disease resistance. **Frontiers in Chemistry**, v. 3, p. 1, 2015.

HUANG, G.; ALLEN, R.; DAVIS, E.L.; BAUM, T.J.; HUSSEY, R.S. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, p. 14302-14306, 2006.

HUA-VAN, A.; LE ROUZIC, A.; MAISONHAUTE, C.; CAPY, P. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. **Cytogenetic Genome Research**, v. 110, p. 426–440, 2005.

HUA-VAN, A.; DAVIÈRE, J.M.; LANGIN, T.; DABOUSSI, M.J. Genome organization in *Fusarium oxysporum*: clusters of class II transposons. **Current Genetics**, v. 37, p. 339–347, 2000.

HÜCKELHOVEN, R.; TRUJILLO, M.; KOGEL, K.H. Mutations in *Ror1* and *Ror2* genes cause modification of hydrogen peroxide accumulation in *mlo*-barley under attack from the powdery mildew fungus. **Molecular Plant Pathology**, v. 1, p. 287-292, 2000.

HUTVAGNER, G. & ZAMORE, P.D. RNAi: nature abhors a double-strand. **Current Opinion in Genetics Development**, v. 12, p. 225–232, 2002.

IDNURM, A. & HOWLETT, B.J. Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. **Fungal Genetics and Biology**, v. 39, n. 1, p. 31–37, 2003.

IKEDA, K.; NAKAYASHIKI, H.; KATAOKA, T.; TAMBA, H.; HASHIMOTO, Y.; TOSA, Y.; MAYAMA, S. Repeat-induced point mutation (RIP) in *Magnaporthe grisea* : implications for its sexual cycle in the natural field context. **Molecular Microbiology**, v. 45, n. 5, p. 1355-1364, 2002.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. **Nature**, v. 409, p. 860–921, 2001.

International Rice Genome Sequencing Project. The map-based sequence of the rice genome. **Nature**, v. 436, p. 793–800, 2005.

JAHAN, S.N.; ÅSMAN, A.K.M.; CORCORAN, P.; FOGELQVIST, J.; VETUKURI, R.R.; DIXELIUS, C. Plant-mediated gene silencing restricts growth of the potato late blight pathogen *Phytophthora infestans*. **Journal of Experimental Botany**, v. 66, n. 9, p. 2785-2794, 2015.

JONES, J.D. & DANGL, J.L. The plant immune system. **Nature**, v. 9, p. 444-323, 2006.

JONES-RHOADES, M.W. & BARTEL, D.P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. **Molecular Cell**, v. 14, p. 787-799, 2004.

KASSCHAU, K.D. & CARRINGTON, J.C. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. **Plant Cell**, v. 19, p. 926-942, 2007.

KELKAR, Y.D. & OCHMAN, H. Causes and Consequences of Genome Expansion in Fungi. **Genome Biology and Evolution**, v. 4, n. 1, p. 13-23, 2012.

KING, Z.R.; CHILDS, S.P.; HARRIS, D.K.; PEDLEY, K.F.; BUCK, J.W.; BOERMA, H.R.; LI, Z. A new soybean rust resistance allele from PI 423972 at the *Rpp4* locus. **Mol Breeding**, v. 37, p. 62, 2017.

KOCH, A.; KUMAR, N.; WEBER, L.; KELLER, H.; IMANI, J.; KOGEL, K.H. Host induced gene silencing of cytochrome P450 lanosterol C14a-demethylaseencoding genes confers strong resistance to *Fusarium* species. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, p. 19324–19329, 2013.

KUMAR, A. & BENNETZEN, J. Plant retrotransposons. **Annual Review of Genetics**, v. 33, p. 479–532, 1999.

LEE, H.C.; AALTO, A.P.; YANG, Q.; CHANG, S.S.; HUANG, G.; FISHER, D.; CHA, J.; PORANEN, M.M.; BAMFORD, D.H.; LIU, Y. The DNA/RNA-dependent RNA polymerase QDE-1 generates aberrant RNA and dsRNA for RNAi in a process requiring replication protein A and a DNA helicase. **PLOS Biology**, v. 8, e1000496, 2010a.

LEE, H.C.; LI, L.; GU, W.; XUE, Z.; CROSTHWAITE, S.K.; PERTSEMLIDIS, A.; LEWIS, Z.A.; FREITAG, M.; SELKER, E.U.; MELLO, C.C.; LIU, Y. Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. **Molecular Cell**, v. 38, p. 803-814, 2010b.

LEE, H.C.; CHANG, S.S.; CHOUDHARY, S.; AALTO, A.P.; MAITI, M.; BAMFORD, D.H.; LIU, Y. qiRNA is a new type of small interfering RNA induced by DNA damage. **Nature**, v. 459, p. 274-277, 2009.

LI, S.; SMITH, J.R; RAY, J.D.; FREDERICK, R.D. Identification of a new soybean rust resistance gene in PI567102B. **Theoretical Applied Genetics**, v. 125, p. 133-142, 2012.

LI, L.; CHANG, S.; LIU, Y. RNA interference pathways in filamentous fungi. **Cellular Molecular Life Sciences**, v. 67, p. 3849–3863, 2010.

LI, S. Reaction of soybean rust-resistance lines identified in Paraguay to Mississippi isolates of *Phakopsora pachyrhizi*. **Crop Science**, v. 49, p. 887-894, 2009.

LI, L.; STOECKERT, C.J.JR; ROOS, D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. **Genome Research**, v. 13, p. 2178–2189, 2003.

LINK, T.; LANG, P.; SCHEFFLER, B.E.; DUKE, M.V.; et al. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. **Molecular Plant Pathology**, v. 15, n. 4, p. 379–393, 2014.

LIU, Q.; RAND, T.; KALIDAS, S.; DU, F.; KIM, H.; SMITH, D.; WANG, X. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. **Science**, v. 301, p. 1921–1925, 2003.

LUSAS, E.W. Soybean processing and utilization. p 949-1036. In: BOERMA, H.R.; SPECHT, J.E. (eds.), Soybeans: Improvement, Production, and Uses. **Agronomy Monograph.** 3rd ed. No. 16, ASA- CSSA-SSSA, Madison, WI. 2004.

MA, L.J.; VAN DER DOES, H.C.; BORKOVICH, K.A.; COLEMAN, J.J.; DABOUSSI, M.J.; et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. **Nature**, v. 464, p. 367-373, 2010.

MAO, Y.B.; CAI, W.J.; WANG, J.W.; HONG, G.J.; TAO, X.Y.; WANG, L.J.; HUANG, Y.P.; CHEN, X.Y. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. **Nature Biotechnology**, v. 25, p. 1307-1313, 2007.

MCLEAN, R.J. & BYTH, D.E. Inheritance of resistance to rust (*Phakopsora pachyrhizi*) in soybeans. **Australian Journal of Agricultural Research**, v. 31, p. 951–956, 1980.

MEERUPATI, T.; ANDERSSON, K.M.; FRIMAN, E.; KUMAR, D.; TUNLID, A.; AHRÉN, D. Genomic Mechanisms Accounting for the Adaptation to Parasitism in Nematode-Trapping Fungi. **PLoS Genetic**, v. 9, n. 11, e1003909, 2013.

MENDGEN, K.; KRIEG, P.A.; REBAGLIATI, M.R.; MANIATIS, T.; ZINN, K.; GREEN, M.R. Volatiles modulate the development of plant pathogenic rust fungi. **Planta**, v. 224, p. 1353-1361, 2006.

MEYER, J.D.; SILVA, D.C.; YANG, C.; PEDLEY, K.F.; ZHANG, C.; VAN DE MORTEL, M.; HILL, J.H.; SHOEMAKER, R.C., ABDELNOOR, R.V.; WHITHAM, S.A.; GRAHAM, M.A. Identification and analyses of candidate genes for rpp4-mediated resistance to Asian soybean rust in soybean. **Plant Physiology**, v. 150, p. 295–307, 2009.

MISSÃO, M.R. Soja: origem, classificação, utilização e uma visão abrangente do mercado. **Maringá Management: Revista de Ciências Empresariais,** v. 3, n. 1, p. 7-15, jan./jun. 2006.

MORALES, A. M. A. P.; BORÉM, A.; GRAHAM, M. A.; ABDELNOOR, R. V. Advances on molecular studies of the interaction soybean - Asian rust. **Crop Breeding and Applied Biotechnology,** v. 12, p. 1-7, 2012.

NAKAYASHIKI, H.; KADOTANI, N.; MAYAMA, S. Evolution and diversification of RNA silencing proteins in fungi. **Journal of Molecular Evolution**, v. 63, p. 127–135, 2006.

NAPOLI, C.; LEMIEUX, C.; JORGENSEN, R. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. **Plant Cell**, v. 2, p. 279–289, 1990.

NEMRI, A.; SAUNDERS, D.G.O.; ANDERSON, C.; UPADHYAYA, N.M.; WIN, J.; LAWRENCE G.J.; JONES, D.A.; KAMOUN, S.; ELLIS, J.G.; DODDS, P.N. The

genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. **Frontiers in Plant Science**, v. 5, n. 98, p. 1-14, 2014.

NICOLÁS, F.E. & RUIZ-VÁZQUEZ, R.M. Functional Diversity of RNAi-Associated sRNAs in Fungi. **International Journal of Molecular Science**, v. 14, p. 15348-15360, 2013.

NOLAN, T.; BRACCINI, L.; AZZALIN, G.; DE TONI, A.; MACINO, G.; COGONI, C. The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in *Neurospora crassa*. **Nucleic Acids Research**, v. 33, p. 1564-1573, 2005.

NOWARA, D.; GAY, A.; LACOMME, C.; SHAW, J.; RIDOUT, C.; DOUCHKOV, D.; HENSEL, G.; KUMLEHN, J.; SCHWEIZERA, P. HIGS: Host-Induced Gene Silencing in the Obligate Biotrophic Fungal Pathogen *Blumeria graminis*. The Plant Cell, v. 22, p. 3130-3141, 2010.

OKAGAKI, L.H.; SAILSBERY, J.K.; EYRE, A.W.; DEAN, R.A. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. **BMC Genomics**, v. 17, p. 135, 2016.

ONO, Y.; BURITICA, P.; HENNEN, J.F. Delimitation of *Phakopsora, Physopella* and *Cerotelium* and their species on Leguminosae. **Mycological Research**, v. 96, p. 825-850, 1992.

ORBACH, M.J.; FARRALL, L.; SWEIGARD, J.A.; CHUMLEY, F.G.; VALENT, B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. **Plant Cell**, v. 12, p. 2019–2032, 2000.

PANDEY, A.K.; YANG, C.; ZHANG, C.; GRAHAM, M.A.; HORSTMAN, H.D.; LEE, Y.; ZABOTINA, O.A.; HILL, J.H.; PEDLEY, K.F.; WHITHAM, S.A. Functional analysis of the Asian soybean rust resistance pathway mediated by *Rpp2*. **Molecular Plant Microbe Interactions**, v. 24, p. 194–206, 2011.

PANSTRUGA, R. Establishing compatibility between plants and obligate biotrophic pathogens. **Current Opinion in Plant Biology**, v. 6, p. 320-326, 2003.

PANWAR, V.; MCCALLUM, B.; BAKKEREN, G. Host-induced gene silencing of wheat leaf rust fungus *Puccinia triticina* pathogenicity genes mediated by the *Barley stripe mosaic virus*. **Plant Molecular Biology**, v. 81, p. 595–608, 2013a.

PANWAR, V.; MCCALLUM, B.; BAKKEREN, G. Endogenous silencing of Puccinia triticina pathogenicity genes through in planta-expressed sequences leads to the suppression of rust diseases on wheat. **The Plant Journal**, v. 73, p. 521–532, 2013b.

PARENT, J.S.; ALBA, A.E.M.; VAUCHERET, H. The origin and effect of small RNA signaling in plants. **Frontiers in Plant Pathology**, v. 3, p. 1-9, 2012.

PARKER, G.; ECKERT, D.; BASS, B. RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA. **RNA**, v. 12, p. 807–818, 2006.

PAUL, C. & HARTMAN, G.L. Sources of soybean rust resistance challenged with single-spored isolates of *Phakopsora pachyrhizi*. **Crop Science**, v. 49, p. 1781-1785, 2009.

PHAM, T.A.; MILES, M.E.; FREDERICK, R.D.; HILL, C.B.; HARTMAN, G.L. Differential responses of resistant soybean entries to isolates of *Phakopsora pachyrhizi*. **Plant Disease**, v. 93, p. 224-228, 2009.

PICHON, X.; WILSON, L.A.; STONELEY, M.; BASTIDE, A.; KING, H.A.; SOMERS, J.; et al. RNA binding protein/RNA element interactions and the control of translation. **Current Protein & Peptide Science**, v. 13, p. 294–304, 2012.

PLIEGO, C.; NOWARA, D.; BONCIANI, G.; GHEORGHE, D.M.; XU, R.; SURANA, P.; WHIGHAM, E.; NETTLETON, D.; BOGDANOVE, A.J.; WISE, R.P.; SCHWEIZER, P.; BINDSCHEDLER, L.V.; SPANU, P.D. Host-Induced Gene Silencing in Barley Powdery Mildew Reveals a Class of Ribonuclease-Like Effectors. **Molecular Plant-Microbe Interactions**, v. 26, n. 6, p. 633–642, 2013.

POSADA-BUITRAGO, M.L. & FREDERICK, R.D. Expressed sequence tag analysis of the soybean rust pathogen *Phakopsora pachyrhizi*. **Fungal Genetics and Biology**, v. 42, p. 949–962, 2005.

RAFFAELE, S.; FARRER, R.A.; CANO, L.M.; STUDHOLME, D.J.; MACLEAN, D.; et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. **Science**, v. 330, p. 1540-1543, 2010.

RICHARDS, T.A.; SOANES, D.M.; FOSTER, P.G.; LEONARD, G.; THORNTON, C.R.; TALBOT, N.J. Phylogenomic Analysis Demonstrates a Pattern of Rare and Ancient Horizontal Gene Transfer between Plants and Fungi. **The Plant Cell**, v. 21, p. 1897-1911, 2009.

ROBERTSON, D. VIGS vectors for gene silencing: many targets, many tools. Annual **Review of Plant Biology**, v. 55, p. 495–519, 2004.

ROMANO, N. & MACINO, G. Quelling, transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. **Moecular. Microbiology**, v. 6, p. 3343–3353, 1992.

ROSSIGNOL, J.L. & FAUGERON, G. Gene inactivation triggered by recognition between DNA repeats. **Experientia**, v. 50, p. 307-317, 1994.

SABOT, F. & SCHULMAN, A. H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. **Heredity**, v. 97, p. 381–388, 2006.

SANJU, S.; SIDDAPPA, S.; THAKUR, A.; SHUKLA, P.K.; SRIVASTAVA, N.; PATTANAYAK, D.; SHARMA, S.; SINGH, B. Host-mediated gene silencing of a

single effector gene from the potato pathogen Phytophthora infestans imparts partial resistance to late blight disease. **Functional & Integrative Genomics**, v. 15, p. 697-706, 2015.

PYOTT, D.E. & MOLNAR, A. Going mobile: Non-cell-autonomous small RNAs shape the genetic landscape of plants. **Plant Biotechnology Journal**, v.13, p. 306-318, 2015.

SCHENA, M.; SHALON, D.; DAVIS, R.W.; BROWN, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, v. 270, p. 467-470, 1995.

SCHERM, H.; CHRISTIANO, R.S.; ESKER, P.D.; PONTE, E.M.; GODOY, C.V. Quantitative review of fungicide efficacy trials for managing rust in Brazil. **Crop Protection**, v. 28, p. 774-782, 2009.

SCHLESINGER, S. Soja: o grão que segue crescendo. **Grupo de Trabalho Sobre Desenvolvimento e Meio Ambiente nas Américas.** Documento de Discussão. n. 21, 23p., 2008.

SCHMIDT, S.M. & PANSTRUGA, R. Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? **Current Opinion in Plant Biology**, v. 14, p. 392-399, 2011.

SCHNEIDER, R.W.; HOLLIER, C.A.; WHITAM, H.K.; PALM, M.E.; MCKEMY, J.M.; HERNANDEZ, J.R.; LEVY, L.; DEVRIES-PATERSON, R. First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. **Plant Disease**, v. 89, p. 774, 2005.

SELKER, E.U. Premeiotic instability of repeated sequences in Neurospora crassa. **Annual Review of Genetics**, v. 24, p. 579-613, 1990.

SHIU, P.K.; ZICKLER, D.; RAJU, N.B.; RUPRICH-ROBERT, G.; METZENBERG, R.L. SAD-2 is required for meiotic silencing by unpaired DNA and perinuclear localization of SAD-1 RNA-directed RNA polymerase. **Proceedings of the National Academy of Sciences of the United States of America**, v. 7, p. 2243-2248, 2006.

SHIU, P.K.; RAJU, N.B.; ZICKLER, D.; METZENBERG, R.L. Meiotic silencing by unpaired DNA. **Cell**, v. 7, p. 905–916, 2001.

SINCLAIR, J.B. Threats to soybean production in the tropics: red leaf blotch and leaf rust. **Plant Disease**, v. 73, p. 604-606, 1989.

SONENBERG, N. & HINNEBUSCH, A.G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. **Cell**, v. 136, p. 731–745, 2009.

SONG, Y. & THOMMA, B.P.H.J. Host-induced gene silencing compromises Verticillium wilt in tomato and *Arabidopsis*. **Molecular Plant Pathology**, v. 1, p. 1-13, 2016.

SONNHAMMER, E.L. & KOONIN, E.V. Orthology, paralogy and proposed classification for paralog subtypes. **Trends Genetics**, v. 18, p. 619–620, 2002.

SPANU, P.D. The Genomics of Obligate (and Nonobligate) Biotrophs. **Annual Reviews Phytopathology**, v. 50, p. 91-109, 2012.

SPANU, P.D.; ABBOTT, J.C.; AMSELEM, J.; BURGIS, T.A.; SOANES, D.M.; et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. **Science**, v. 330, p. 1543–1546, 2010.

STONE, C.L.; MCMAHON, M.B.; FORTIS, L.L.; NUÑEZ, A.; SMYTHERS, G.W.; LUSTER, D.G.; FREDERICK, R.D. Gene expression and proteomic analysis of the formation of *Phakopsora pachyrhizi* appressoria. **BMC Genomics**, v. 13, p. 269, 2012.

SUNKAR, R. & ZHU, J.K. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. **Plant Cell**, v. 16, p. 2001-2019, 2004.

TATUSOV, R.L.; KOONIN, E.V.; LIPMAN, D.J. A genomic perspective on protein families. **Science**, v. 278, p. 631–637, 1997.

TAVARES, S.; RAMOS, A.P.; PIRES, A.S.; AZINHEIRA, H.G.; CALDEIRINHA, P.; LINK, T.; ABRANCHES, R.; SILVA, M.C.; VOEGELE, R.T.; LOUREIRO, J.;

TALHINHAS, P. Genome size analyses of Pucciniales reveal the largest fungal genomes. **Frontiers in Plant Science**, v. 5, p. 422, 2014.

TINOCO, M.L.; DIAS, B.B.; DALL'ASTTA, R.C.; PAMPHILE, J.A.; ARAGÃO, F.J. *In vivo* trans-specific gene silencing in fungal cells by in planta expression of a double-stranded RNA. **BMC Biology**, v. 8, p. 27, 2010.

TOMILOV, A.A.; TOMILOVA, N.B.; WROBLEWSKI, T.; MICHELMORE, R.; YODER, J.I. Trans-specific gene silencing between host and parasitic plants. **Plant Journal**, v. 56, p. 389-397, 2008.

TREMBLAY, A.; HOSSEINI, P.; LI, S.; ALKHAROUF, N.W.; MATTHEWS, B.F. Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. **BMC Genomics**, v. 14, p. 614, 2013.

TREMBLAY, A.; HOSSEINI, P.; LI, S.; ALKHAROUF, N.W.; MATTHEWS, B.F. Identification of genes expressed by *Phakopsora pachyrhizi*, the pathogen causing soybean rust, at a late stage of infection of susceptible soybean leaves. **Plant Pathology**, v. 61, p. 773–786, 2012.

TREMBLAY, A.; LI, S.; SCHEFFLE, B.E.; MATTHEWS, B.F. Laser capture microdissection and expressed sequence tag analysis of uredinia formed by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. **Physiological and Molecular Plant Pathology**, p. 12, 2009.

USDA. National Agricultural Statistics Service. **Crop Production 2016 Summary**, p. 1-123, January 2017.

VAZQUEZ, F.; VAUCHERET, H.; RAJAGOPALAN, R.; LEPERS, C.; GASCIOLLI, V.; MALLORY, A.C.; HILBERT, J.L.; BARTEL, D.P.; CRETE, P. Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. **Molecular Cell**, v. 16, p.69-79, 2004.

VELCULESCU, V.E.; ZHANG, L.; VOGELSTEIN, B.; KINZLER, K.W. Serial analysis of gene expression. **Science**, v. 270, p. 484-487, 1995.

VOINNET, O. Origin, biogenesis, and activity of plant microRNAs. **Cell**, v. 136, p. 669-687, 2009.

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

WATTERS, M.K.; RANDALL, T.A.; MARGOLIN, B.S.; SELKER, E.U.; STADLER, D.R. Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. **Genetics**, v. 153, p. 705-714, 1999.

WESSLER, S.R. Transposable elements and the evolution of eukaryotic genomes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 47, p. 17600-17601, 2006.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J.L.; et al. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics, v. 8, p. 973-982, 2007.

YIN, C.T.; DOWNEY, S.I.; KLAGES-MUNDT, N.L.; RAMACHANDRAN, S.; CHEN, X.; SZABO, L.J.; PUMPHREY, M.; HULBERT, S.H. Identification of promising host-induced silencing targets among genes preferentially transcribed in haustoria of *Puccinia*. **BMC Genomics**, v. 16, p. 579, 2015.

YIN, C.T.; JURGENSON, J.E.; HULBERT, S.H. Development of a host-induced RNAi system in the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. **Molecular Plant–Microbe Interactions**, v. 24, p. 554–561, 2011.

YORINORI, J. T.; PAIVA, W. M.; FREDERICK, R. D.; COSTAMILAN, L. M.; BERTAGNOLLI, P. F.; HARTMAN, G. L.; GODOY, C. V.; NUNES-JUNIOR, J. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. **Plant Disease**, Saint Paul, v. 89, n. 4, p. 675-677, 2005.

YORINORI, J.T.; JUNIOR, J.N.; LAZZAROTTO, J.J. Ferrugem “asiática” da soja no Brasil: evolução, importância, economia e controle. **Documentos 247**, 36p, 2004.

YORINORI, J.T.; PAIVA, W.M.; FREDERICK, R.D.; FERNANDEZ, P.F.T. Ferrugem da soja (*Phakopsora pachyrhizi*) no Brasil e no Paraguai, nas safras 2000/01 e 2001/02. **Resumos, II Congresso Brasileiro de Soja**, Foz do Iguaçu, PR., p. 94. 2002.

ZAMBOLIN, L. Manejo integrado da ferrugem asiática da soja. In: ZAMBOLIN, L. (ed.). **Ferrugem asiática da soja**. Suprema, Viçosa, 139p, 2006.

ZHANG, T.; JIN, Y.; ZHAO, J.H.; GAO, F.; ZHOU, B.J.; FANG, Y.Y.; GUO, H.S. Host-induced gene silencing of target gene in fungal cells confers effective resistance to cotton wilt disease pathogen *Verticillium dahliae*. **Molecular Plant**, doi: 10.1016/j.molp.2016.02.008, 2016.

ZHANG, H.; GUO, J.; VOEGELE, R.T.; ZHANG, J.; DUAN, Y.; LUO, H.; KANG, Z. Functional characterization of calcineurin homologs PsCNA1/PsCNB1 in *Puccinia striiformis* f. sp. *tritici* using a host-induced RNAi system. **PLoS One**, v. 7, e49262, 2012.

ZHANG, C.; BRADSHAW, J.D.; WHITHAM, S.A.; HILL, J.H. The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. **Plant Physiology**, v. 153, p. 52–65, 2010.

ZHANG, C.; YANG, C.; WHITHAM, S.A.; HILL, J.H. Development and use of an efficient DNA-based viral gene silencing vector for soybean. **Molecular Plant Microbe Interactions**, v. 22, p. 123–131, 2009.

ZHOU, B.; BAILEY, A.; NIBLETT, C.; QU, R. Control of brown patch (*Rhizoctonia solani*) in tall fescue (*Festuca arundinacea* Schreb.) by host induced gene silencing. **Plant Cell Reports**, v. 35, p. 791–802, 2016.

2. CAPÍTULO 2: New insights of *Phakopsora pachyrhizi* infection based on transcriptome analysis *in planta*

Short running title: *Phakopsora pachyrhizi* transcriptome

ABSTRACT

Asian soybean rust (ASR) is one of the most destructible diseases affecting soybeans. The causative agent of ASR, the fungus *Phakopsora pachyrhizi*, presents characteristics that make it difficult to study *in vitro*, limiting the knowledge access of plant-pathogen dynamics. Therefore, this work used the leaf lesion laser microdissection technique, associated with the high performance sequencing, to obtain pathogen transcriptome during the interaction with soybean. The 36,350 unisequences generated provided an overview of the main genes and biological pathways active in the fungus, among them the proteins synthesis, nucleic acids and energy metabolism. We also identified the most expressed transcripts, including sequences similar to other fungi virulence proteins and signaling proteins. Enriched *P. pachyrhizi* transcripts were found to resistant soybean genotype (PI561356), related to extracellular matrix organization and signaling metabolic pathways, and among infection structures, involved with amino acid metabolism and the intracellular transport. Unisequences were further grouped into gene families along with predicted sequences from 15 other fungi and oomycete species, including rust fungi, allowing the identification of multigenic families conserved, as well as specific to *P. pachyrhizi*. The results revealed important biological processes observed in *P. pachyrhizi*, contributing with information related to fungal biology and, consequently, a better understanding of ASR.

Keywords: Asian soybean rust, multigenic families, transposable elements

2.1 INTRODUCTION

The plant pathogenic basidiomycete fungi *Phakopsora pachyrhizi* (Sydow & P. Sydow) causes the disease known as Asian soybean rust (ASR). ASR is one of the diseases causing the most significant losses in soybean (*Glycine max* (L.) Merrill) crops and causes great concern because it is a polycyclic disease with a high destructive power (Scherm *et al*., 2009). *P. pachyrhizi* reproduction is assexual (Anderson *et al*., 2008). Once in contact with leaf surface, the urediniospores, an assexual form of spores, germinates initiating the infection process that progress rapidly, and within five to eight days new urediniospores are formed and released through sporulation structures (uredinias), initiating a new cycle of infection (Zambolin, 2006; Morales *et al*., 2012). As an obligatory biotrophic organism, the development of *P. pachyrhizi* occurs only in living tissue, thus, *in vitro* growth of this pathogen is impossible, which hampers the study of fungal biology (Voegele *et al*., 2009).

The infection strategies used by biotrophic organisms have received increased attention in the last decade as a result an increasing availability of genomic data for these pathogens, such as the sequencing of the genomes of *Blumeria graminis*, *Puccinia graminis* f.sp *tritici*, *Melampsora larici-populina, Puccinia striiformis* f.sp. *tritici*, and *Melampsora lini* (Spanu *et al*., 2010; Cantu *et al*., 2011; Duplessis *et al*., 2011; Nemri *et al*., 2014). These data allowed the identification of some adaptive characteristics preserved throughout the evolutionary process, such as those related to the extreme parasitic lifestyle leading to the loss of nitrate and sulfate assimilation pathways. On the other hand, these genomes shows expansion of gene families

related to pathogen nutrient acquisition and effectors delivery (Cantu *et al*., 2011; Duplessis *et al*., 2011; Nemri *et al*., 2014).

Since the nuclear genome of *P. pachyrhizi* has not yet been sequenced, most of the molecular information about this pathogen comes from work involving transcriptome sequencing associated to bioinformatics analyzes. In this way, a moderate repertoire of candidate gene sequences expressed in germinated urediniospores, appressorium, haustorium and uredinia became available (Posada-Buitrago and Frederick, 2005; Tremblay *et al*., 2009, 2012, 2013; Stone *et al*., 2012; Link *et al*., 2014; Carvalho *et al*, 2016). Most of these studies focused on the characterization of specific molecular mechanisms regarding the infection process, concentrating in the analysis of specific structures of the pathogen, such as urediniospores and haustorium, or the analysis of the total infected leaf. To date, studies aimed to characterize the *P. pachyrhizi* transcriptome associating high performance sequencing and sample enrichment with pathogen tissues, via LCM (Laser Capture Microdissection) (Emmert-Buck *et al*., 1996) of the lesion, are still scarce.

This work enabled the characterization of the *in plant* transcriptome of *P. pachyrhizi*, allowing an overview of the transcripts obtained and the molecular processes acting at 10 days post infection in the resistant (PI561356) and susceptible (BRS 231) soybean genotypes. Besides, allowed the identification of most expressed sequences in both genotypes, and enriched pathogen transcripts presents in the PI561356 and among different fungal infection structures. Comparative analysis between the transcripts of *P. pachyrhizi* and other fungi allowed grouping transcripts in multigenic families common to different species of fungi, including other rust species, revealing families conserved between the different

species analyzed and specific to *P. pachyrhizi*. The prediction of active transposable elements allowed the identification of different subclasses of retrotransposons and DNA transposons elements in *P. pachyrhizi* transcriptome. Finally, RT-qPCR analysis validated the expression levels of *P. pachyrhizi* genes observed in the results of deep sequencing.

2.2 MATERIALS AND METHODS

2.2.1 *P. pachyrhizi* transcriptome during host interaction

2.2.1.1 Experimental design and inoculation

The experimental design was completely randomized, with three replicates per treatment and three plants per pot, using genotypes PI561356, which has an R gene mapped close to the *Rpp1* gene (Kim *et al*., 2012), and susceptible soybean BRS 231 (Ribeiro *et al*., 2007). The plants were grown in a greenhouse under controlled conditions of temperature and humidity until reaching the developmental stage V2 (Fehr *et al*., 1971), when they were inoculated with the fungus. The inoculum used in this work came from a Brazilian population of *P. pachyrhizi*. Urediniospores were propagated for more than 10 generations in susceptible genotype BRS MS-Bacuri, under controlled greenhouse conditions. For inoculation, urediniospores were collected and resuspended in solution containing 0.05% Tween20 (v/v), to a final concentration of 1.3×10^5 spores mL^{-1}. The same solution, however without spores, was used for false inoculated leaves (mock plants) as a control for the inoculation procedure. The plants were covered with plastic bags for two days to optimize the infection of the plant by the pathogen and to avoid the

cross-contamination of the control plants falsely inoculated. At 10 days post inoculation (dpi) we observed TAN (BRS 231) and RB (PI561356) lesions on the underside of the leaf of the inoculated plants, but not on the falsely inoculated controls. At this time, pieces of leaves containing rust lesions were collected for laser capture microdissection (LCM) procedures.

2.2.1.2 Laser capture microdissection (LCM)

Leaf segments of 1cm^2 with lesion were randomly sampled from the third trifoliate of infected plants and immediately fixed on ice in Farmer's solution containing 75% ethanol (v/v) and 25% acetic acid (v/v) (Kerk *et al*., 2003). Pieces of leaves were fixed overnight at 4 °C in Farmer's solution. On the second day, the Farmer solution was removed and the leaf samples were washed twice with 75% (v/v) cold ethanol, and dehydrated by a series of ethanol: xylene solutions graduated. The samples were embedded in paraffin at 58 °C, transferring the samples into xylol: paraffin solutions in which the paraffin concentration increased with each transfer (every four hours), culminating in the transfer of the samples to pure paraffin in the fourth and final transfer, as described by Cai and Lashbrook (2006). All paraffin-embedded blocks were stored for three weeks at -20 °C until LCM. Immediately before the LCM procedure, 12-µm sections were cut into a rotating microtome and transferred to membrane microscope slides. The cuts were dewaxed with xylol in a series of xylol: ethanol solutions, dehydrated in a graduated ethanol series, and stained with fast green acid and fuchsin. Twenty sections containing a variable number of rust lesions were prepared for each biological replicate. For this study, the PixCell II LCM system (Arcturus, Sunnyvale, CA, USA) and CapSure Macro LCM (Arcturus) were used to collect fungal and leaf cells. Each cap was used for pulses

between 1500-2000 corresponding to the collection of approximately the same number of cells. Samples were collected for both plants, resistant and susceptible (Carvalho *et al*., 2016).

2.2.1.3 RNA isolation and sequencing

RNA extraction was performed from cells collected separately from each biological replicate, so the RNA extracted from 4,500 to 6,000 cells represents each genotype. Total RNA (<10ng) was extracted from mesophilic cells collected from rust infection sites using the PicoPure RNA isolation kit (Arcturus). Two rounds of RNA amplification were performed with RiboAmp HS Plus (Arcturus) to obtain a final yield of approximately 18 µg of amplified RNA (aRNA). The quality and quantity of the aRNA were determined using a 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA). The cDNA synthesis was performed before sequencing using the First Strand Super Script III kit (Invitrogen), following the manufacturer's recommendations. The high performance paired-end (54 pb, 250 pb insert size) and single-end (100 bp) sequencing was obtained through the Illumina Genome Analyzer GAiix platform (San Diego, CA, USA).

2.2.2 Construction of the *P. pachyrhizi* transcriptome

Initially, the low-quality sequences were removed and the remaining RNA-Seq reads were aligned with the soybean reference genome Williams 82 (Schmutz *et al*., 2010) using the TopHat software (Trapnell *et al*., 2009). Reads without alignments were de novo assembled using the Trinity assembler (Grabherr *et al*., 2011). Two different assemblies were performed: (i) using only the single-end reads (minimum

contig size of 200 bp); (ii) using only the paired-end reads (minimum contig size of 100 bp). In order to increase the set of transcript reference data, the *P. pachyrhizi* ESTs from Sanger sequencing available at National Center for Biotechnology Information (NCBI) were assembled into contigs and singlets. The ESTs were adjusted using the bdtrimmer software (Baudet and Dias, 2005) to discard low quality, vector, and short sequences, and possible soybean contamination was estimated using the BLASTN (Altschul *et al*., 1997) against the soybean genome. The remaining sequences were assembled, producing 6,105 sequences. The three assemblies were merged using CAP3 (Huang and Madan, 1999).

Finally, Sspace (Boetzer *et al*., 2011) was used to scaffold the CAP3 result, producing 53,405 scaffolds (mean size, 433.1 bp). All 53,405 sequences from de novo assembly were aligned with the soybean genome (BLASTN, e-value cutoff of 1e-10), with the GenBank nr database (BLASTX, e-value cutoff of 1e-5) and with two local fungal databases: (i) 78,105 proteins from the following plant interacting fungus: *L. bicolor*, *P. graminis* f. sp. *tritici*, *U. maydis*, *M. larici-populina* and *M. oryzae*; and (ii) 840,789 *P. pachyrhizi* genomic sequences (accession numbers 451441523–455661682). Sequences that aligned with the soybean genome and not with our local fungal database were considered to be soybean contigs. Sequences that showed high similarity to plant sequences and did not align to the soybean genome or to fungal sequences were also considered to be from soybean. Sequences with a positive match in soybean, but with high similarity to the fungal databases were considered to be *P. pachyrhizi* contigs. All the other sequences were considered to be *P. pachyrhizi* contigs.

The final *P. pachyrhizi* dataset consisted of 36,350 contigs. All reads used in the de novo assembly step were mapped into the contigs using SOAP2 (Li *et al*.,

2009) and the alignment result was used to calculate the expression level of each contig by FPKM (fragments per kilobase million) (Mortazavi *et al*., 2008). All generated sequences, from soybean and from *P. pachyrhizi*, are available in the *P. pachyrhizi* transcriptome LGE database (http://bioinfo03.ibi.unicamp.br/phakopsora/).

P. pachyrhizi transcripts were aligned and annotated against the GenBank non-redundant (NR) protein database (BLASTX, e-value cut-off of 1e-5), GenBank nucleotide database (NT) (BLASTN, e-value cutoff of 1e-5), Conserved Domains Database (CDD) (Marchler-Bauer *et al*., 2011) (RPSBLAST, e-value cut-off of 1e-3), the reference soybean genome Williams 82 (BLASTN, e-value cutoff of 1e-10), and finally against the local fungal databases. The functional annotations for the *P. pachyrhizi* transcripts were performed using the Blast2GO software (Conesa *et al*., 2005), refining the annotations using functions Annex and GO-Slim mapping for yeast, as well as the InterProScan function used to identify domains, and the Enzime Code and KEEG (Kyoto encyclopedia of genes and genomes) function for the identification of metabolic pathways and enzymatic sequences.

2.2.3 Enriched categories in *P. pachyrhizi* transcriptome

Enrichment analysis (Fisher's exact test) (p-value of 0,05) available in Blast2GO were performed to identify contigs overrepresented in each of both genotypes, susceptible and resistant (BRS 231 and PI561356, respectively), and contigs conserved among other *P. pachyrhizi* infection structures. To identify enriched transcripts among genotypes, we used contigs with FPKM values equal to or greater than 1 to each of the genotypes. To identify sequences common to different infection structures, a total of 1,029 ESTs from germinated urediniospores

and appressorium (Stone *et al*., 2013) and 4,483 ESTs from haustorium (Link *et al*., 2014) were separately aligned against the 36,350 *P. pachyrhizi* unisequences (leaf lesion) using WU-BLAST (Washington University BLAST - http://blast.wustl.edu). Sequences results (only the sequences that presented a similarity greater than or equal to 90%) were then used in Venn analysis. The sequences common to infection structures were then subjected to enrichment analysis by Blast2GO software.

2.2.4 Comparative analysis

To identify multigenic families conserved between *P. pachyrhizi* transcriptome and other fungi genome and transcriptome, we performed a clustering analysis using the OrthoMCL software (Li *et al*., 2003). The construction of multigenic families was performed using sets of proteins predicted from publicly available genomes: 12 fungi genomes [10 Basidiomycetes – *M. larici-populina* (Joint Genome Institute - JGI), *P. graminis* f. sp. *tritici* (Broad Institute), *Coprinopsis cinerea* (Broad Institute), *Cryptococcus neoformans* (Broad Institute), *Postia placenta* (JGI), *Lacaria bicolor* (JGI), *Malassezia globosa* (JGI), *Phanerochaete chrysosporium* (JGI), *Sporobolomyces roseus* (JGI, v1) and *Ustilago maydis* (Broad Institute); and two Ascomycetes - *Neurospora crassa* (Broad Institute) and *Magnaporthe grisea* (Broad Institute)] and two oomycetes genomes [*Phytopthora sojae* (JGI) and *Phytopthora infestans* (Broad Institute)]. In addition to the already sequenced genomes listed above, we also used the data sets of ESTs available for rust *Uromyces appendiculatus* and *P. pachyrhizi* (Link *et al*., 2014). These sets of sequences were combined with the transcriptome results from *P. pachyrhizi* obtained in this work. All data were compiled into a single FASTA file that was used in the independent

version of the OrthoMCL software version 1.4 (http://orthomcl.org/common/downloads/software/unsupported/v1.4/), using MCL standard and inflation parameter of 1.5. Thus, based on the similarity of the protein sequences, they were grouped into OrthoMCL multigenic families.

Comparative analyzes were performed using three parameters: a) families common to all 16 species used; b) families common to basidiomycetes (covering the species *M. larici-populina*, *P. graminis*, *C. cinerea*, *C. neoformans*, *P. placenta*, *L. bicolor*, *M. globosa*, *P. chrysosporium*, *S. roseus, U. maydis, U. appendiculatus* and *P. pachyrhizi*); and c) families common to rust fungi (Pucciniales) (covering the species *M. larici-populina*, *P. graminis*, *U. appendiculatus* and *P. pachyrhizi*). Finally, we also identified *P. pachyrhizi* exclusive families.

2.2.5 Prediction of transcriptionally actives transposable elements

P. pachyrhizi unisequences were compared against transposable elements (TEs) sequences available at Repbase protein transposable elements database (Kapitonov and Jurka, 2008) using a computer RepeatMasker tool version open 4.0.6. The query species was assumed to be fungi dataset, available in the Repbase database. Unigenes were considered related to TEs following the RepeatMasker parameters “selective and matches to coding sequences”, available in http://www.repeatmasker.org/webrepeatmaskerhelp.html.

2.2.6 Validation of RNA-Seq expression levels by RT-qPCR

Quantitative PCR (RT-qPCR) was performed using samples of spores and germinated spores (*in vitro* growth) and infected leaf tissue (*in vivo* growth). Spores (S) and germinated spores (GS) were obtained from the fresh spores of *P. pachyrhizi* cultivated on detached soybean leaves and maintained in Petri dishes under controlled temperature and humidity conditions in a heated chamber. Germinated spores were obtained from fresh spores deposited in Petri dish containing water solution with addition of 0.04% Tween and remained overnight for 16 hours. Infected leaf tissue were collected from susceptible soybean genotype (Williams 82) maintained in a greenhouse for fungal inoculation and subsequent samples collection. Inoculation was performed at stage V2 of developing, on the second foliolate, using a brazilian *P. pachyrhizi* isolate LD5511 (shows virulence in *Rpp1*, *Rpp2* e *Rpp3* and *Rpp5* soybean genotypes) (Darben *et al*., 2013) in a concentration of 1 x 10^5 spores.mL^{-1}. Infected leaves of compatible interactions were collected at different times after inoculation (0, 6, 12, 24, 48, 72, 96 and 192 hours post inoculation “hpi”) to represent the progression of infection and host tissue colonization by the fungus. Once harvested, fungus and plant were immediately frozen in liquid nitrogen and stored at -80 ° C. Experiment consisted of three biological replicates each containing three plants.

The total RNA was extracted from 100mg of frozen leaf tissue samples and 30mg spores and germinated spores using the RNeasy Plant Mini Kit (Qiagen). RNA contamination with genomic DNA was eliminated by the treatment of 1ug RNA with RNase-free DNase (Invitrogen). The cDNA was synthesized using the First Strand Super Script III kit (Invitrogen), following the manufacturer's recommendations. RT-

qPCR reactions were performed on the Real Time StepOnePlusTM equipment (Applied Biosystems) using SYBR green for the detection of double-strand PCR products. Primers were designed using Primer3Plus software based on the sequences of the six *P. pachyrhizi* transcripts (Table 1). All primers were first tested by standard PCR using DNA and soybean cDNA to ensure specific amplification for *P. pachyrhizi*. Three negative controls were used to ensure that only the cDNA of *P. pachyrhizi* was amplified. The efficiency of the primers was calculated based on the equation $[10^{(-1/slope)}] - 1$ (Pfaffl, 2001). Each PCR reaction was performed in triplicate and the specificity of the amplification products was validated by the analysis of the dissociation curve. Expression levels were determined by equation $2^{-\Delta Ct}$. The fungus endogenous gene tubulin was used as the normalizing gene (Maciel *et al*., 2010). RT-qPCR expression levels obtained were used to support deep sequencing results, thus the FPKM values obtained through RNA-Seq data were also normalized by the FPKM values of the endogenous tubulin gene.

Table 1. Sequences of RT-qPCR primers, amplicon size and primer efficiency of selected *Phakopsora pachyrhizi* genes.

Contigs	Primers (F - forward, R - reverse)	Amplicon size (bp)	Primer efficiency (%)
de_novo_595 (Thi – thiamine biosynthesis)	F - TCACTGAGCTAATCGGTACAGG R - GCCTTTCCAGCCAAAGTATG	73	97
de_novo_4668 (PPI – peptidyl-prolil-cys/trans isomerase)	F - TCGTCAGCTTTGCCTTAGAC R - CTCCCTGGAGCATAAATTGG	86	98
de_novo_939 (AGO – argonaut)	F – ACTCGCTCGGTTTCTATTGC R - CCGTTGAAATCGACGTTACC	92	90
de_novo_2740 (Pv_SNARE – soluble NSF attachment receptor)	F – CCGTCAATGATCCATACGTG R - AGAAGTGATGCCCGTTGTTC	113	90
de_novo_5382 (NtR – nitrate reductase)	F - CCAAGTCAAAGGTCCCAAAG R - CAGCATTGGAGTTATCCCTGTC	97	98
de_novo_57 (HSS – small heat shock)	F – GACGCAAGTTTGATGACTCG R - AACAGAGATAACCGGCTTGC	92	90
de_novo_380 (Tub– tubulin)	F – CCAAGGCTTCTTCGTGTTTCA R - AGAGAAGAGCGCCAAACC	65	93

2.3 RESULTS

2.3.1 Transcriptome overview

The *P. pachyrhizi* transcriptome on infected soybean leaves was obtained by inoculating a resistant genotype (PI561356 (*Rpp1b*) and a susceptible genotype (BRS 231) with a *P. pachyrhizi* isolate maintained in a greenhouse. At 10 dpi, sections of leafs containing RB (PI561356) or TAN (BRS 231) lesions were collected and the tissues were fixed and prepared for LCM. The LCM methodology was used to enrich the transcripts of the fungus in relation to the plant transcripts, dissecting fungal cells and adjacent cells of the mesophyll. RNAs present in the LCM samples were identified through Illumina sequencing. A total of 92 million of reads of 100 bp single-end and 54 bp paired-end (insert size of 250 bp) were generated from susceptible (BRS 231) and resistance (PI561356) soybean genotypes.

A total of 144,659 contigs were generated after the removal of low quality reads and reads that aligned with the soybean genome (*Glycine max Wm82.a2.v1*) or soybean predicted transcripts (Schmutz *et al*., 2010). The remaining reads were used for the *ab initio* assembly of *in plant P. pachyrhizi* transcripts at 10 dpi. To improve assembling quality, *P. pachyrhizi* ESTs from Sanger sequencing reads available from NCBI have been edited and assembled into 6105 contigs (with a mean size of 625 bp). The three assemblies were merged generating a total of 53,405 contigs, comprising 17,055 from soybean sequences and 36,350 unique *P. pachyrhizi* sequences (unisequences), expressed at 10 dpi *in plant*.

The combined assembly resulted in an increase in the size of the contigs and a higher percentage of contigs that were mapped to the local database of fungal

sequences. A total of 11,614 (31.95%) transcripts from *P. pachyrhizi* showed similarity to proteins encoded by other phytopathogenic fungi, especially rust fungi *P. graminis* (9,979 sequences) and *M. larici-populina* (9,362 sequences), which present their complete genomes available (Duplessis *et al*., 2011). It is important to note that 64.07% of the sequences (23,290 sequences) exhibited similarity with genomic reads of *P. pachyrhizi* (Table 2).

Table 2. General statistics of the *de novo* assembly of RNA-Seq data.

Parameters	***in planta* Pp transcriptome**
Number of contigs	36,350
Average contigs size (bp)	471.68
Size of largest contig (bp)	7,874
Size of smallest contig (bp)	100
Average number of reads per contig	2,077.11
Number of annotated contigs[1]	
Against NR	19,573
Against NT	12,520
Against CDD	11,626
Against local fungi database[2]	11,614
Against *P. pachyrhizi* genomic reads	23,290
Against soybean genome	106
Blast2GO general results	
Total GO terms	14,043
Contigs with functional annotation	5,622
Contigs with identified domains	4,187
Contigs in metabolic pathways	1,519
Contigs with enzymatic codes	1,129

[1]Number of annotated contigs: number of contigs aligned in NR (BLASTX, e-value cutoff of 1e-5), NT (BLASTN, e-value cutoff of 1e-5), Conserved Domais Databases (CDD) (BLASTX, e-value cutoff of 1e-5), banco local de dados de fungos (BLASTX, e-value cutoff of 1e-5), reads genômicas de *P. pachyrhizi* e genoma da soja (BLASTN, e-value cutoff of 1e-5).

[2] Local fungi database: 78,105 proteins of five phytopathogenic fungi – 23,.132 of *L. bicolor* (JGI, v. 2.0), 20,566 of *P. graminis* (Broad Institute), 6,522 of *U. maydis* (Broad Institute), 16,831 of *M. larici-populina* (JGI, v. 1.0) and 11,054 of *M. grisea* (Broad Institute) - and 840,789 genomic reads of *P. pachyrhizi* (NCBI).

Blast2GO software, with the support of the refinement annotation strategies ANNEX and GO-Slim mapping for yeast, assigned 14,043 GO (gene ontology) terms to the 36,350 *P. pachyrhizi* unisequences. The transcripts were submitted to the InterProScan function, which detected the presence of domains in 4,187 sequences, and the Enzime Code and KEGG function, which identified 1,519 contigs present in metabolic pathways and 1,129 showed known enzymatic functions (Table 2).

2.3.1.1 Gene ontology analysis and main functional categories

Blast2GO attributed functional annotation to 15.47% of the *P. pachyrhizi* unisequences, resulting in 5,622 transcripts associated with GO terms. These sequences were classified and grouped according to their characteristics into three broad categories: cellular components, molecular functions and biological processes. A total of 1,934 cellular components, 4,010 molecular functions and 3,469 biological process terms were associated with our sequences (Table 3). The GO classifications were distributed in 15 levels among these three categories. The most informative GO levels for *P. pachyrhizi* transcripts were three to eight, which include a high number of annotated GO terms.

Table 3. Summary of acyclic graphs for the arrangement of GO terms for cellular components, molecular functions and biological processes.

Categories	Nº of contigs	Categories	Nº of contigs	Categories	Nº of contigs
Cellular Component	**1,934**	**Molecular Functions**	**4,010**	**Biological Processes**	**3,469**
Membrane	697	Binding	2,289	Localization	609
part of membrane	474	Ion binding	968	transport	580
Cell	1,471	protein binding	574	single-organism localization	378
cytoplasm	573	organic cyclic compound binding	1,242	Biological Regulation	609
intracellular organelle	874	heterocyclic compound binding	1,242	regulation of biological process	383
ribonucleoproteins intracellular complex	338	small molecule binding	745	Single-Organism Process	1,693
Macromolecular Complex	854	carbohydrate derivative binding	557	single-organism localization	378
ribonucleoproteins complex	338	Catalytic Activity	2,353	single-organism celular process	1,293
protein complex	530	transferase activity	596	single-organism metabolic process	1,150
Organelle	880	Hidrolase activity	745	Cellular Process	2,363
part of organelle	452	Oxidorredutase activity	474	regulation of the cellular process	365
Intracellular organelle	874			single-organism celular process	1,293
non-membrane-bounded organelle	416			cellular metabolic process	1,848
membrane-bounded organelle	566			Metabolic Processes	2,693
				oxidation-reduction process	354
				cellular metabolic process	1,848
				primary metabolic process	1,853
				organic substances metabolic processes	2,041
				biosynthetic process	913
				nitrogen compound metabolic process	1,198
				Cell Component Organization or Biogenesis	437

Bold, underline, and simple font mean a progression from high to low hierarchical levels, respectively.

Cell components category presented four parental terms: membrane (GO: 0.016.020), cell (GO: 0.005.623), macromolecular complex (GO: 0.032.991) and organelle (GO: 0.043.226) (Table 3). The parental term with the highest number of sequences grouped was cell term, presenting sequences related to parts of the cytoplasm and intracellular organelles, as well as components of the intracellular complex of ribonucleoproteins including ribosome sequences. The parental term organelles presented sequences related to intracellular organelles, basically divided between organelles not delimited by membrane and organelles delimited by membranes, and for the latter, there are sequences related to the components of the nucleus. The parental terms membrane and macromolecular complex showed limited detail of the sequences covered.

Molecular functions category presented only two parental terms: binding (GO: 0.005.488) and catalytic activity (GO: 0. 003 824) (Table 3). The parental term binding basically comprised sequences whose molecular functions are related to linkages between proteins, ionic bonds and linkage between nucleic acids, especially among purine bases. Already the parental term catalytic activity comprised sequences detailed only by having metabolic activity of transferases, hydrolases and oxidoreductases.

Finally, biological processes were the category that presented the highest number of classes of parental terms. A total of six parental terms were identified: location (GO: 0.051.179), biological regulation (GO: 0.065.007), single-organism process (GO: 0.044.699), cellular process (GO: 0.009.987), metabolic process (GO: 0.008.152) and organization of cellular component or biogenesis (GO: 0.071.840)

(Table 3). Cellular process and metabolic process were the most strongly represented parent terms.

Cellular process, presented the majority of sequences related to the metabolic cellular process that includes sequences that act on the heterocyclic metabolism, phosphorus metabolism, aromatic compounds, nitrogen compounds, processes of cellular biosynthesis and metabolism of macromolecules (including protein metabolism and RNA metabolic processes). The majority of sequences grouped into the parental term metabolic process are also related to the metabolic cellular process (including the sequences mentioned above) and also to the metabolism of organic substances, which among others includes the biosynthesis of organic compounds and macromolecules, as well as processes related to gene expression. The others parental terms within the biological processes were also related to the functions of some sequences already mentioned, but they present a lower level of detail of the sequences found.

2.3.1.2 Conserved domains and metabolic pathways

The InterProScan function in the Blast2GO software identified 4,187 non-redundant domains in *P. pachyrhizi* unisequences (Figure 1). Among twenty most abundant domains identified in transcripts, the main domain found among fungal transcripts was the P-loop domain containing nucleotide phosphate hydrolase (IPR027417), present in 248 sequences. Proteins that present this type of domain are usually responsible for catalyzing the hydrolysis of the phosphate binding present in the nucleoside triphosphate (NTP). The energy resulting from the hydrolysis of NTP is usually used to promote conformational changes in other molecules, which

forms the basis of the biological functions of most of these enzymes. The other domains found cover different molecular functions such as, for example, oxidoreductases domains, protein-protein binding domains, nucleic acid recognition and cleavage domains, protein kinase domains, transporter domains, among others.

Through the Enzime Code and KEGG function available in the Blast2GO software, it was possible to identify metabolic pathways and enzymatic functions related to annotated transcripts of *P. pachyrhizi*, giving an overview of the pathogen metabolism during interaction with the host. A total of 1,519 contigs were grouped into 99 metabolic pathways identified by KEGG. Some of these pathways stand out because of the high number of transcripts distributed among different enzymatic classes, as in the case of the purine metabolism (ko00230), with 308 contigs and 26 enzymatic classes, and antibiotic biosynthesis (ko01130), with 96 contigs and 56 enzymatic classes. Among the other metabolic pathways, it was possible to find different types of metabolisms and processes, such as carbohydrate and lipid metabolism, cell cycle, cellular respiration processes, amino acid biosynthesis and degradation processes, protein interactions and regulation, biosynthesis of hormones, responses to stresses and processes of signaling metabolic pathways.

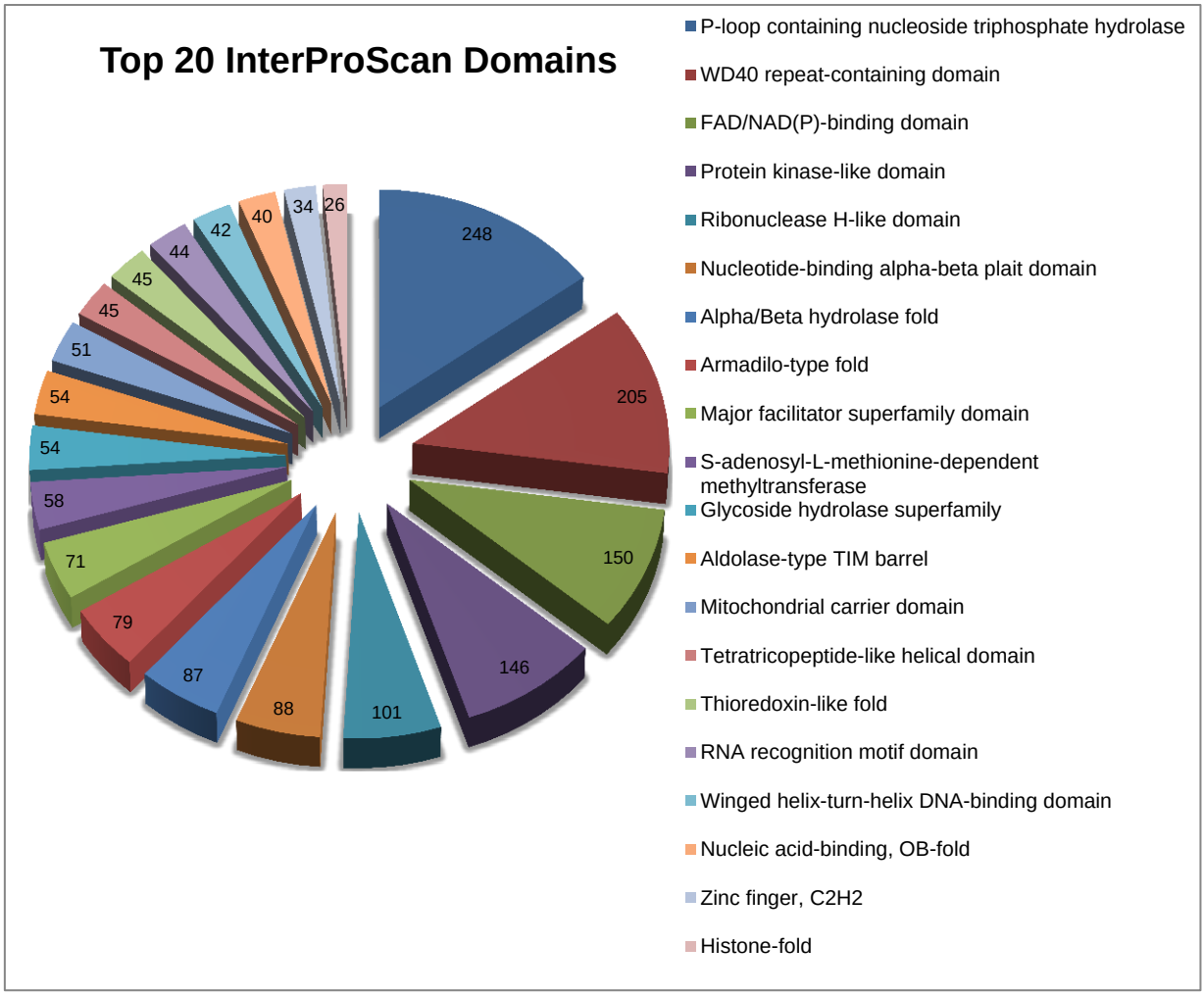


Figure 1. Most represented InterProScan domains associated with *Phakopsora pachyrhizi* transcripts. The number of sequences found for each domain is presented.

2.3.2 Top expressed genes in *P. pachyrhizi* transcriptome

From the set of *P. pachyrhizi* 36,350 unisequences, the 50 most expressed transcripts, based on the FPKM values, were identified for each of the soybean genotypes (PI561356 and BRS 231) (Table S1). Among the 50 sequences identified for each genotype, 42 were common among the both genotypes, resulting in 58

different transcripts identified in the total, which were not associated with any GO term. A total of 25 sequences presented annotations when aligned against the NR database and against the local database of phytopathogenic fungi, and these sequences showed similarity to GAS1 and GAS2 virulence proteins of *M. grisea*. When 58 transcripts were compared to the secreted sequences identified by Link *et al*. (2014) in the haustorium transcriptome of *P. pachyrhizi*, it was possible identify 44 sequences similar to 12 of the transcripts predicted as secreted. In addition, three transcripts found among the top expressed common to both genotypes (de_novo_2238, de_novo_5381, and de_novo_5849) were also functionally validated by Carvalho *et al*. (2016).

Some domains have been identified among the 58 most common transcripts present in almost all sequences, generally indicating their presence in membranes, as well as the formation of helix and coil structures (Coil – Coiled coil domain, Non-cytoplasmic domain, Transmembrane domain; Cytoplasmic domain, and TMhelix - Transmembrane helix domain). Eleven transcripts presented an uncharacterized domain, related to a family of eukaryotic proteins named DUF3129 family (IPR021476). Other five domains were identified among 42 contigs common to both genotypes: DNA polymerase III subunits gamma and tau (PRK07764); RNA polymerase I-associated factor PAF67 (PFAM10255); DUF70 domain (PFAM01901) present in archaebacterial and that may present transmembrane proteins function; MAEBL (PTZ00121) domain, that defines a family of erythrocyte binding proteins, those present in malaria parasites, with a role during a host invasion; and finally a domain present in prokaryotic organisms, related to membrane lipoproteins with specific lipid-binding topics (PS51257).

Specifically for the most expressed *P. pachyrhizi* transcripts in contact with BRS 231 (eight sequences), only three domains are found: a UL36 domain (PHA03247), typical of herpesvirus in tegument proteins, possibly acts shortly after the onset of infection; a fibronectin-attachment protein domain; and DNA polymerase III (PRK12323). To eight most expressed sequences found only in PI561356, four domains are identified: a DUF3129 domain, already mentioned above; a common fungal domain (IPR020100), which encodes a yet unknown function protein, but is related to the repression of carbohydrate metabolism, once increasing during a glucose breakdown; a non-catalytic Src homology (SH2) domain (PFAM14633), conserved among a series of cytoplasmic signaling proteins regulated by receptor protein-tyrosine kinases and involved in normal signaling and cellular transformation; and an internal membrane protein (Tol-a) related to cell envelope integrity (PRK09510). Some non-specific domains have also been found among the 58 most expressed sequences, for example, domains relating to galactose-3-O-sulfotransferase proteins (Pfam06990), synthesis of one form of natural vitamin k (cd13962), fibronectin binding proteins (Pfam07174), and keratin-associated proteins (Pfam11759), common mammalian domains.

2.3.3 Enriched categories in *P. pachyrhizi* transcriptome

We conducted enrichment analysis in Blast2GO to identify transcripts overrepresented in each genotyped, susceptible (BRS 231) and resistant (PI561356), and also contigs conserved and overrepresented among different infection

structures, such germinated urediniospores and appressorium (Stone *et al*., 2012), haustorium (Link *et al*., 2014) and leaf lesion.

Most *P. pachyrhizi* 36,350 unisequences were generated by assembling reads obtained from the two genotypes; however, some contigs were exclusive from reads obtained from susceptible (BRS 231) or resistant (PI561356) genotypes. Considering only transcripts with FPKM values equal to or greater than one (30,491 contigs), it was possible to identify 5,952 contigs composed exclusively of reads from *P. pachyrhizi* genes expressed in BRS 231, and 6,185 contigs composed exclusively of reads from *P. pachyrhizi* genes expressed in PI561356 (Figure 2). Among the 5,952 sequences found exclusively in the susceptible genotype, 1,349 presented hits in the GenBank non redundant (NR) database and 456 were associated with gene ontology (GO) terms. To the 6,185 sequences found exclusively in the resistant genotype, 1,904 presented hits in the GenBank non redundant (NR) database and 243 were associated with GO terms. To both genotypes the most of all contigs were similar to hypothetical or predicted proteins of *Puccinia* and/or *Melampsora*.

To verify the presence of transcripts common to different infection structures, 1,029 sequences expressed in germinated urediniospores and appressorium (Stone *et al*., 2012) and 4,483 sequences expressed in haustorium (Link *et al*., 2014) were aligned against the *P. pachyrhizi* unisequences generated in this work, which correspond to sequences expressed in leaf lesions. A total of 3,265 sequences from dataset of haustorium and 423 sequences from dataset of germinated urediniospores and appressorium showed high similarity with *P. pachyrhizi* unisequences; and of these two datasets 48 are common between all structures (Figure 2). To the common transcripts identified between the haustorium and leaf lesions sequences, only 154 of

the 3,265 sequences were associated to GO terms, to transcripts common between germinated urediniospores, appressorium and leaf lesions, 12 of the 423 sequences were associated with GO terms, and finally, to 48 transcripts common between all infections structures, eight sequences were associated to GO terms.

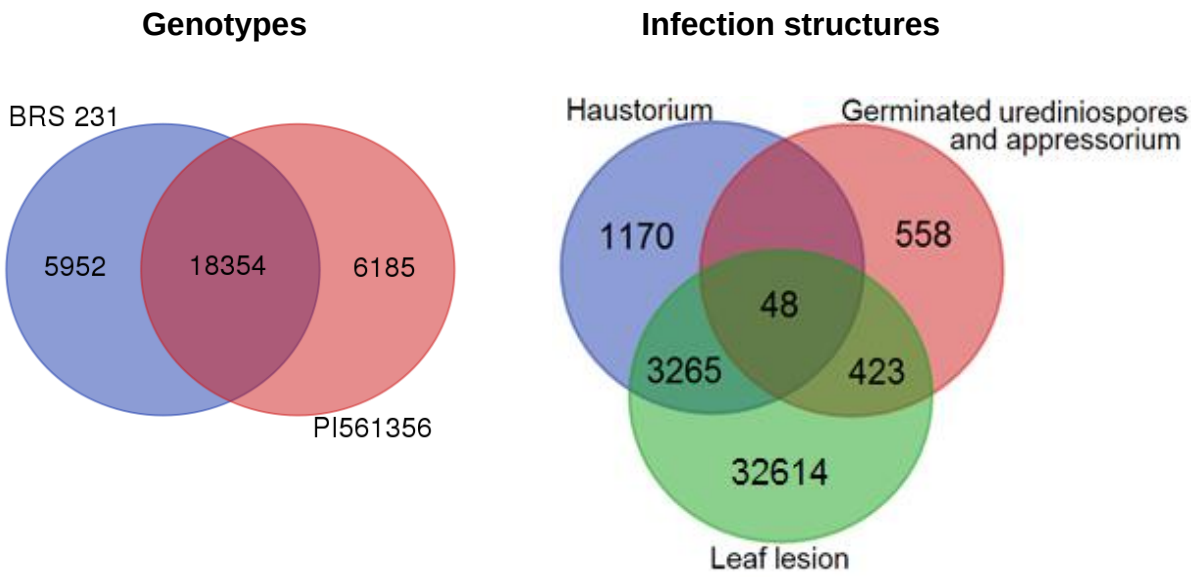


Figure 2. Number of *P. pachyrhizi* unisequences found exclusively among susceptible and resistant genotypes (BRS 231 and PI561356, respectively), and common among different fungal infection structures (germinated urediniospores and appressorium, haustorium and leaf lesion).

2.3.3.1 Enriched categories among genotypes

The enrichment analyzes performed with the transcripts from each genotype against the total set of unisequences of *P. pachyrhizi* (36,350 contigs) revealed no enriched molecular class among the sequences found in BRS 231, whereas for the sequences found in PI561356, it allowed the identification of 23 different enriched

molecular classes. Although it was not found any enriched class among BRS 231 and the total transcriptome, it was possible observe in the sequence exclusives to the susceptible genotype four different and general molecular classes for the 456 transcripts associate with GOs: catalytic activity, protein binding, metal cluster binding and metabolic process.

Among the 23 molecular classes enriched for *P. pachyrhizi* transcripts from resistant genotype, the majority draws attention because it is related to fungal pathogenic activity during host infection. Class related to exocytosis activity is the one with the highest number of contigs (18 sequences), followed by classes of glycosaminoglycan binding (14 sequences), positive regulation of signal transduction (nine sequences) and acute inflammatory response (nine sequences). Other molecular classes also stand out from the others, for example, the regulation of intracellular protein kinase cascade and extracellular matrix organization. The enriched molecular classes, as well as the amount of sequences present in resistant genotype (PI561356) from the GO terms identified are shown in Figure 3.

The other 16 molecular classes identified among the transcripts, but not enriched for the sequences from the resistant genotype, comprise different biological activities, for example, related to ribonucleoprotein complex, which contains the largest number of contigs (82 sequences), besides the integral membrane components, transcription processes, nucleic acid biosynthesis, RNA processing, phosphorylation and protein transport, regulation of cellular component organization, among others.

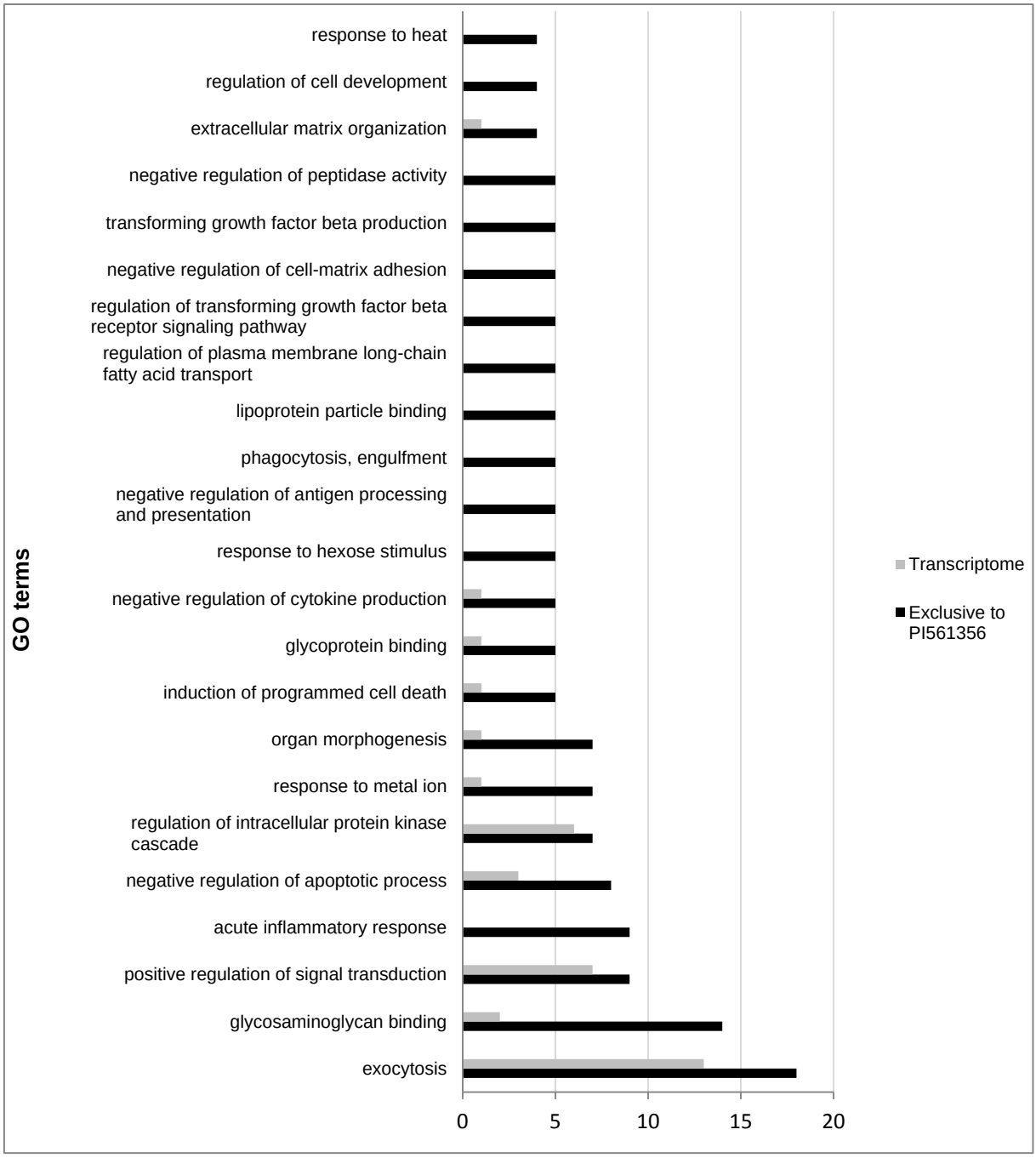


Figure 3. Differential gene ontology (GO) term distribution between the *Phakopsora pachyrhizi* transcriptome (36,350 contigs) and transcripts generated exclusively in resistant genotype PI561356 (6,185 contigs). In this graph are presented only the molecular classes enriched for the exclusive transcripts of PI561356. This graph was automatically generated after the enrichment analysis of the BLAST2GO tool (Fisher's exact test, P>0.05).

2.3.3.2 Enriched categories among fungal structures

The enrichment analyzes performed with the transcripts common to different infections structures against the *P. pachyrhizi* transcriptome (36,350 contigs) allowed the identification of three enriched molecular classes among sequences common to germinated urediniospores, appressorium and leaf lesion, 12 enriched molecular class between sequences common haustorium and leaf lesion, and finally, were identified one enriched molecular class between sequences common to all infection structures analyzed (Figure 4).

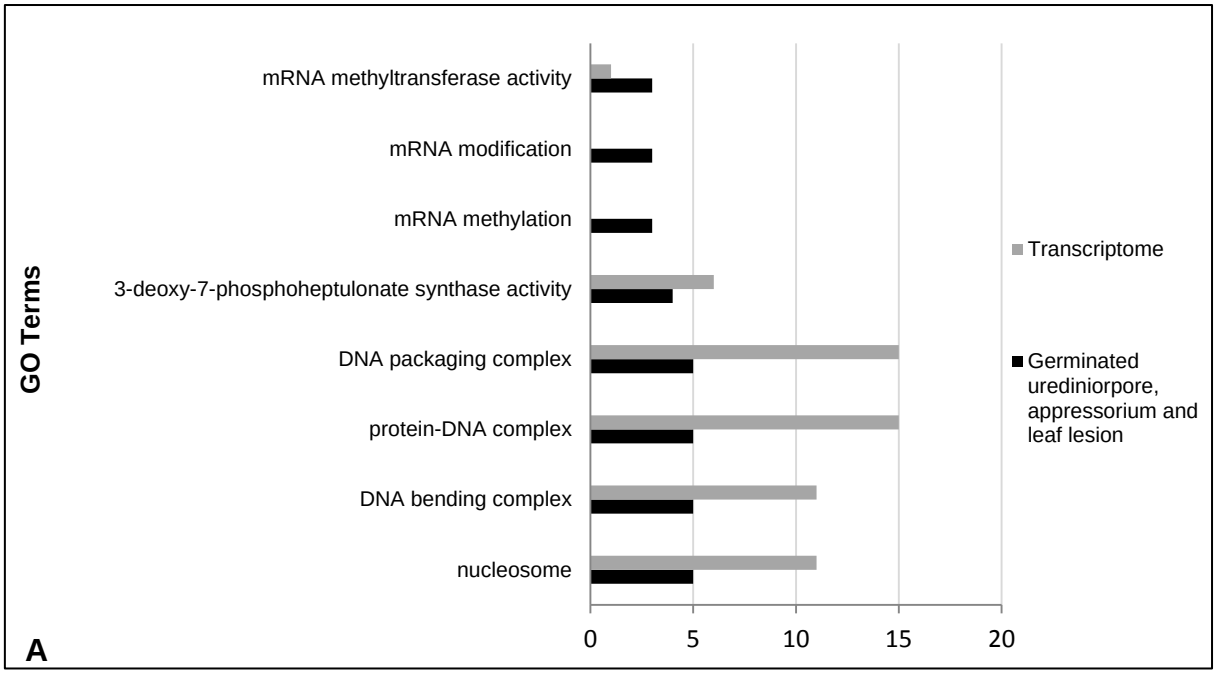
The 12 transcripts common to germinated urediniospores, appressorium and leaf lesion, which were associated with GO terms, were classified into eight different molecular classes, that comprises modification and methylations of mRNAs molecules, DNA binding and packaging complexes, protein-DNA complex, nucleosomes and enzymes of the shikimate pathway. To this analysis, only classes associated with mRNA modification (modifications and methylations of mRNAs and mRNA methyltransferase activity) were enriched in relation to the *P. pachyrhizi* transcriptome, with less than five sequences in each class (Figure 4A).

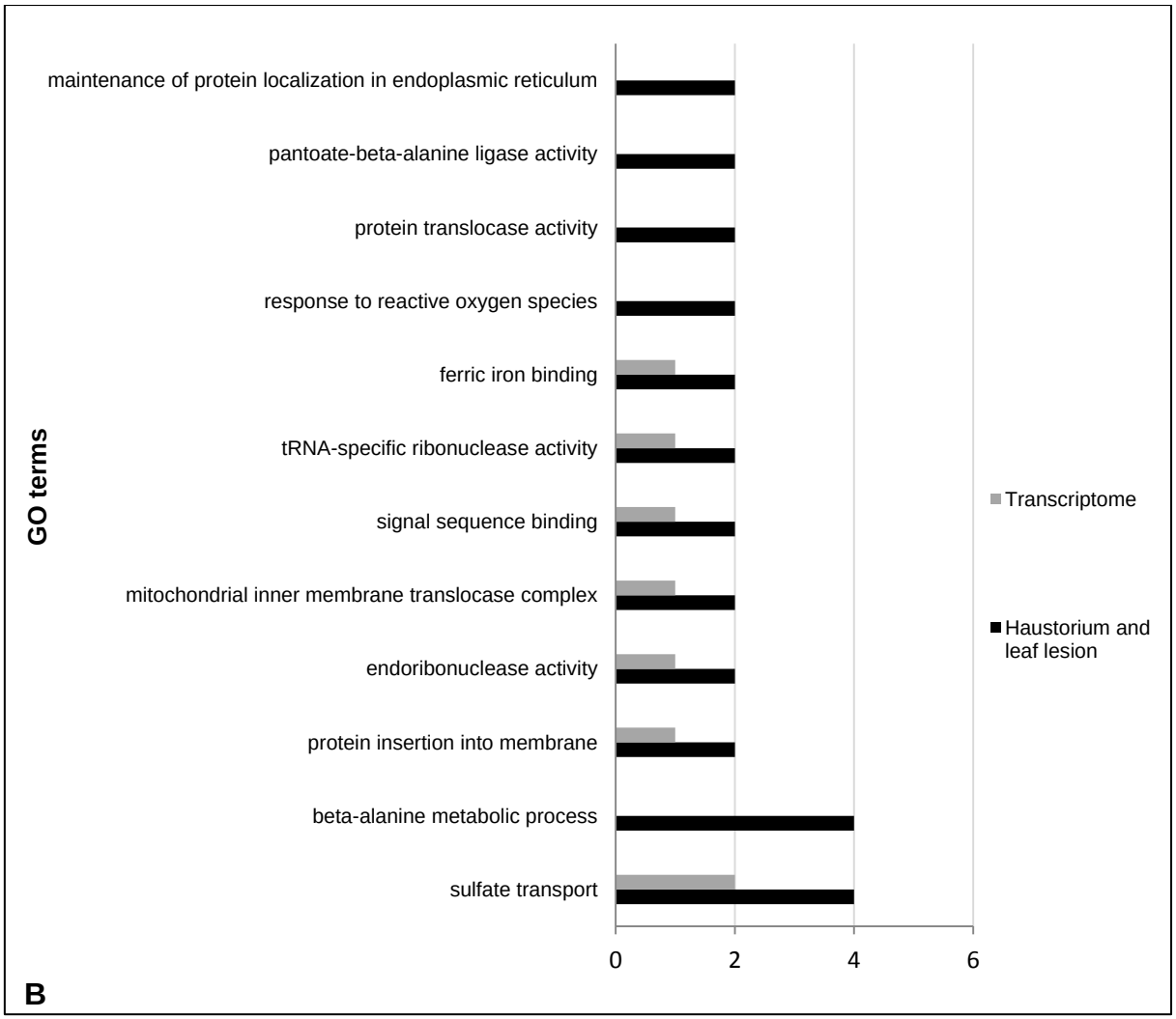
To 154 transcripts associated with GO terms which are common to haustorium and leaf lesion, were identified a total of 34 molecular classes. Among the 12 enriched classes, the largest number of sequences was observed into sulfate transport and beta-alanine metabolic process, each with four contigs. Besides these classes, it is also possible to observe molecular classes related to response to reactive oxygen species, endoribonuclease activity, and protein translocase activity, among others. The remaining 22 molecular classes comprise molecular functions related to different biosynthetic process, oxidation-reduction process,

ribonucleoprotein complex, intracellular transport, generation of precursor metabolites and energy, transcription processes and modification of RNA, for example (Figure 4B).

Finally, eight of 48 transcripts common between all infections structures were associated with GO terms and classified into 11 molecular classes, and only one of these classes is enriched in relation to transcriptome, the formaldehyde degradation (1 sequence). The other 10 molecular classes cover sequences basically related to catabolism and biosynthesis of amino acids, and with the transport of substances between the endoplasmic reticulum and the Golgi complex (Figure 4C).

Among the 40 remaining transcripts common among all infection structures and that were not associated with GO terms, 15 sequences did not show annotation against the NCBI database. To the transcripts with annotation in NCBI database, 14 sequences were similar to *M. larici-populina* and/or *P. graminis* hypothetical proteins, two sequences were similar to *P. pachyrhizi* clones, one sequence was related to secreted protein of *M. larici-populina*, and one sequence was similar to carbohydrate esterase family of *M. larici-populina*. Seven transcripts presented non-specific annotations.


GO Terms
mRNA methyltransferase activity
mRNA modification
mRNA methylation
3-deoxy-7-phosphoheptulonate synthase activity
DNA packaging complex
protein-DNA complex
DNA bending complex
nucleosome
0
5
10
15
20
Transcriptome
Germinated urediniorpore, appressorium and leaf lesion
A


GO terms
maintenance of protein localization in endoplasmic reticulum
pantoate-beta-alanine ligase activity
protein translocase activity
response to reactive oxygen species
ferric iron binding
tRNA-specific ribonuclease activity
signal sequence binding
mitochondrial inner membrane translocase complex
endoribonuclease activity
protein insertion into membrane
beta-alanine metabolic process
sulfate transport
0
2
4
6
Transcriptome
Haustorium and leaf lesion
B

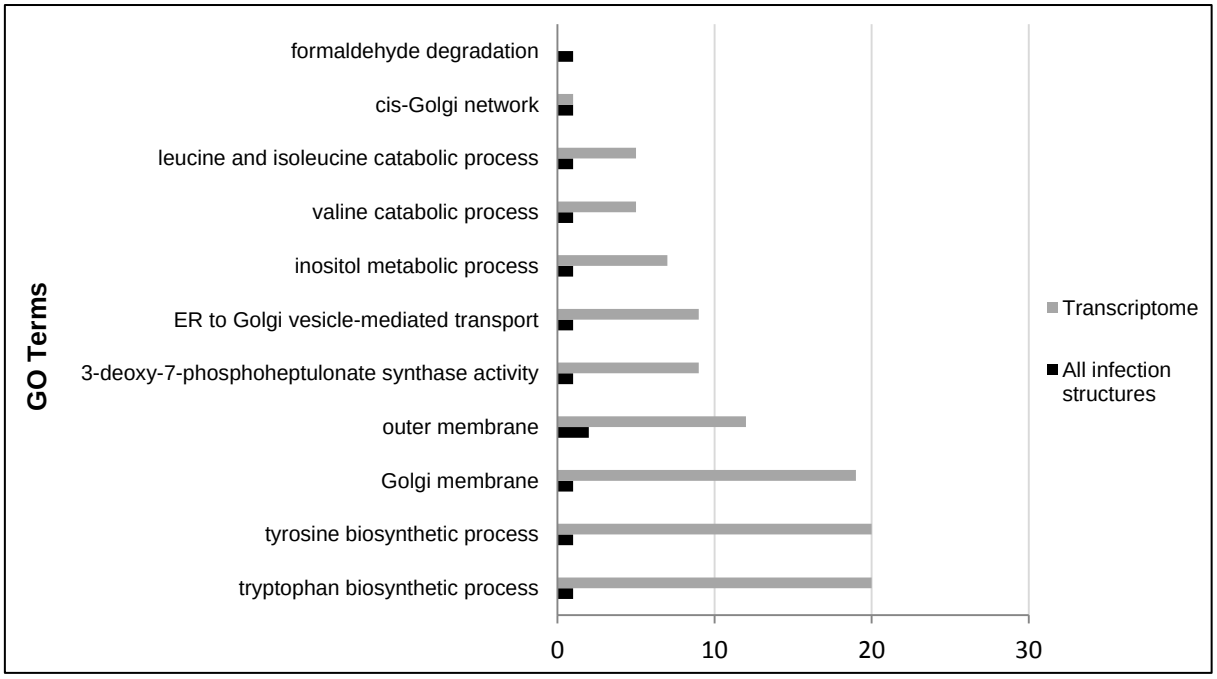


Figure 4. Differential gene ontology (GO) term distribution between the *Phakopsora pachyrhizi* transcriptome (36,350 contigs) and fungal infection structures. (A) transcripts common to germinated urediniospores, appressorium and leaf lesion (423 contigs); (B) transcripts common to haustorium and leaf lesion (3,265 contigs); and (C) transcripts common to all infection structures, germinated urediniospores, appressorium, haustorium and leaf lesion (48 contigs). In graph B are presented only the molecular classes enriched for transcripts common to haustorium and leaf lesion. This graph was automatically generated after the enrichment analysis of the BLAST2GO tool (Fisher's exact test, P>0.05).

2.3.4 Comparative analysis

Predicted protein sequences from 13 fungal genomes and two oomycetes genomes were grouped with the *P. pachyrhizi* unisequences (36,350 transcripts) based on the similarity of the sequences, thus constituting OrthoMCL gene families.

We identified 4,775 OrthoMCL multigenic families containing at least two sequences. OrthoMCL families comprise 52,380 protein sequences, of which 6,072 are *P. pachyrhizi* transcripts, accounting for 16.7% of total fungal transcripts.

The correlations between sequences of *P. pachyrhizi* and the other 15 species used for the construction of OrthoMCL families were analyzed and the highest values were observed to the families common to all species (with 260 OrthoMCL families and 5,475 genes distributed among them), and families common to rust-causing fungi (with 92 OrthoMCL families and 525 genes distributed among them), following by families common to basidiomycetes (with 7 OrthoMCL families and 130 genes distributed among them) (Table 4).

Hypothetical proteins and predicted proteins are the molecular categories that correspond to the largest portion of the sequences observed, except for families common to all species that present ribosomal proteins as most abundant sequences. Secreted proteins were observed to families common to basidiomycetes and families common to rust fungi. For both parameters, the amount of secreted sequences corresponded to approximately 25% of all sequences found in the families. Specifically for the families of secreted proteins among the 12 species of basidiomycetes, *P. pachyrhizi* was the species with the highest number of sequences, presenting 7 of the 14 predicted secreted sequences. The other families found among the parameters are mainly characterized by involvement in primary metabolic processes related to the survival and maintenance of the organism, such as protein and carbohydrate metabolism, and transport of substances, and a few families involved in processes of secondary metabolism, for example, signaling metabolic pathways.

Table 4. General characteristics of the comparative analysis between the OrthoMCL multigene families obtained from the predicted proteins of the *Phakopsora pachyrhizi* transcriptome and proteins predicted from other 15 species.

Molecular categories of OrthoMCL families	OrthoMCL families	Total of sequences	*P. pachyrhizi* sequences
Families common to all species[1]			
Ribosomal proteins	54	1,014	75
Predicted proteins	24	449	28
Protein synthesis	17	309	18
Desidrogenases	11	311	17
Cytoplasmic transporters	8	225	12
Hypothetical proteins	5	87	8
Membrane transporters	2	334	7
Others	139	2,746	169
Total	260	5,475	334
Families common to basidiomycetes			
Predicted proteins	2	32	3
Hypothetical proteins	1	22	1
Secreted proteins	1	35	7
Transcription factor binding domains	1	17	1
Methylation	1	12	1
Metabolic pathways signaling	1	12	1
Total	7	130	14
Families common to rust fungi			
Hypothetical proteins	66	374	77
Secreted proteins	11	93	17
Carbohydrate Metabolism	4	21	4
Predicted proteins	3	14	4
Transport of peptides	1	10	1
Spindle checkpoint signaling	1	6	1
Protein metabolism	1	5	1
Signaling and regulation of the circadian cycle	1	5	1
Vesicular fusion	1	4	1
Transmembrane transport	1	4	1
Nitrogen metabolism	1	4	1
No annotation	1	5	1
Total	92	525	110

Families exclusive of P. pachyrhizi[2]			
No annotation	510	867	867
Hypothetical proteins	101	179	179
Predicted proteins	45	80	80
Secreted proteins	8	18	18
Others	192	316	316
Total	856	1,460	1,460

[1]Families common to all species: for this parameter were listed only the eight molecular categories that had the largest number of families, or a greater number of sequences.
[2]Families exclusive of *P. pachyrhizi*: for this parameter were listed only the four molecular categories that had the largest number of families, or a greater number of sequences.

The molecular category with the largest number of families, representing 18.2% of the sequences in the families common to all species, was that of genes encoding ribosomal proteins (mainly for the 60S and 40S subunits), with 75 sequences identified in *P. pachyrhizi*. Other categories also presented high number of sequences, such as proteins involved in the translation process (mainly translation factors and tRNAs), in addition to general dehydrogenases and transporters. Two families specific for ABC membrane transporters stand out as being the two largest families common to all species in relation to the number of sequences, comprising alone 334 genes. The 139 families common to all species that were not presented in the table include genes involved in the most diverse biological processes, including metabolism of carbohydrates, proteins, nucleic acids and energy metabolism, transcription processes, signaling metabolic pathways, among others.

Few families were common among the species of basidiomycetes, however, the families identified are related to interesting functions such as transcription processes, nucleic acid methylation processes, and signaling pathways. Rust fungi share families with different molecular functions, most related to general processes of maintenance and cell survival, such as protein and carbohydrate metabolism, spindle checkpoint signaling and transport of peptides. Other families may have a more

direct relationship with pathogenicity processes common to these species, such as families related to transmembrane transporters and vesicle fusion.

It was possible to observe that among the sequences found in *P. pachyrhizi*, some were similar to those found in the haustorium transcriptome results obtained by Link *et al*. (2014). Most were observed in families common to all species, a total of 97 sequences that perform varied functions, since sequences related to histone proteins and cell cycle proteins, to proteins related to metabolic pathway signaling. In the families common to rust fungi, 20 sequences were found, of which 14 are in the hypothetical protein category, four were identified as secreted proteins, and two others are related to protein metabolism and circadian cycle signaling and regulation.

Exclusive families of *P. pachyrhizi* were also identified, totaling 856 families that added 1,460 genes distributed among them, and of these, 422 sequences were identified among the transcripts obtained by Link *et al*. (2014). Among the exclusive families of *P. pachyrhizi*, most do not have functional annotation (510 families), however, among those with a known function it is possible to note the presence of gene families, beyond those identified as hypothetical, predicted, or secreted sequences. To the 18 sequences found in families of secreted proteins, two (de_novo_3939 and de_novo_7164), belonging to the same family, were also present in the results of functional validation obtained by Carvalho *et al*. (2015). The other gene families not shown in the table, presented the most varied functions among which we highlight the sequences act in carbohydrate metabolism processes, protein and nucleic acid metabolism, cell cycle , splicing processes, as translation factors, in metabolic pathway signaling, and as membrane transporters (including MFS and ABC transporters).

2.3.5 Prediction of transcriptionally actives transposable elements

BLAST analyzes were performed against a reference database of transposable elements (Repbase) using the computer RepeatMasker tool and were identified 592 contigs with transposable elements (TE) fragments. From these contigs, 484 were annotated as retrotransposons (Class I) (81.76%) and 108 as DNA transposons (Class II) (18.24%) (Table 5). To the class I, 413 elements were classified as long terminal repeat elements (LTR elements), and 132 elements were classified as non-long terminal repeat elements (non-LTR elements), more specifically elements identified as long interspersed nuclear element (LINE). To the class II, we found 90 elements classified as terminal inverted repeats elements (TIR elements), two elements classified as Crypton, and 22 elements classified as Helitron.

Different superfamilies of TEs were found within the orders of each class. Into class I, the order of LINE elements presented three different superfamilies, Tad1, Deceiver and L1 superfamilies, and the order of LTR elements is composed of Copia and Gypsy superfamilies. The Copia and Gypsy superfamilies are most abundant mobile elements in fungal genomes. Already into classe II, the order of TIR elements showed eight diferent superfamilies, Tc1-Mariner, PIF-Harbinger, EnSpm, Zisupton, P-Fungi, Merlin, hAT-Ac and MuDR (MULE). Besides the transposable elements found in the unisequences of *P. pachyrhizi*, were also identified 10,012 simple repeat elements and others 1,376 elements detected were not identified because are classified as low complexity.

Table 5. Transcriptionally active transposable elements in *Phakopsora pachyrhizi* transcriptome.

TE Classification	Nº of TE elements	Nº of *P. pachyrhizi* contigs
Class I (retrotransposons)	545	484
LINE	*132*	*98*
Tad1	129	96
Deceiver	2	2
L1	1	1
LTR	*413*	*386*
Copia	214	195
Gypsy	188	183
Class II (DNA transposons)	124	108
TIR	*90*	*85*
Tc1-Mariner	37	35
PIF-Harbinger	26	24
EnSpm	18	18
Zisupton	3	3
P-Fungi	1	1
Merlin	2	2
hAT-Ac	2	2
MuDR (MULE)	1	1
Helitron	*22*	*20*
Crypton	*2*	*2*

Underline, italic and simple fonts represent the classes, orders and superfamilies, respectively.

2.3.6 Validation of RNA-Seq expression levels by RT-qPCR

RT-qPCR was conducted using six *P. pachyrhizi* contigs, selected preferentially involved in pathogenicity, showing expression levels in different *P. pachyrhizi* infection time points, corroborating the expression levels from FPKM values observed from the mRNA-Seq data (Table 6). Expression levels identified by

RT-qPCR for the genes Thi (thiamine biosynthesis), PPI (peptidyl-prolyl-cis/trans isomerase), AGO (argonaut), Pv-SNARE (soluble NSF attachment receptor) and HSS (small heat shock), show that at some point in the process of plant infection by the pathogen these genes were induced. Similar results can be observed for the transcripts of these genes obtained by RNA-Seq, whose already normalized values of FPKM are high.

The highest expression values from the Seq-mRNA results to induced genes were observed for the Thi and PPI genes, with already standardized FPKM values above 20, and the lowest values were observed for the Pv-SNARE gene, with FPKM values below 6. In contrast, the highest values obtained by the RT-qPCR analyzes to induced genes were observed for the Pv-SNARE gene, with fold change values up to 8.4, and the lowest values were observed for the Thi gene, with fold change values up to 1.6. Specifically for the NtR (nitrate reductase) gene, both levels of expression, by mRNA-Seq and by RT-qPCR, show that, unlike the other contigs, this presented low levels of expression. These results show that, despite RT-qPCR detected the expression level at different time-points which did not occur for the mRNA-Seq values, and despite differences between expression levels, RT-qPCR expression values of all genes tested were consistent with the values observed in the mRNA-Seq data.

Table 6. Validation of gene expression base on mRNA-Seq assay using RT-qPCR.

Genes	mRNA-Seq[3]		RT-qPCR[1,2]									
	BRS 231	PI561356	ES	EG	0hpi	6hpi	12hpi	24hpi	48hpi	72hpi	96hpi	192hpi
Thi	27.146	23.533	0.004	0.011	0.187	0.003	0.052	0.467	1.054	0.843	0.894	1.589
PPI	21.178	24.666	0.212	0.412	0.282	0.082	0.392	1.118	1.211	1.386	1.658	2.340
AGO	5.676	6.886	1.083	2.075	0.854	1.613	3.075	2.773	1.157	1.048	1.375	2.782
Pv-SNARE	3.209	5.561	2.636	8.464	3.144	7.926	5.174	0.829	0.397	0.293	0.268	0.191
HSS	6.780	3.387	0.144	0.457	1.245	1.138	2.985	1.103	1.677	1.617	1.204	0.836
NtR	0.288	0.356	0.097	0.200	0.117	0.140	0.389	0.110	0.272	0.191	0.195	0.231

[1]Main *P. pachyrhizi* infection time points: at the stages of spore (ES) and germinated spore (EG) before contact with soybean, and after soybean contact at 0, 6, 12, 24, 48, 72 , 96 and 192 hours post infection “hpi”.
[2]qPCR results are represented by fold change values obtain after normalization with the endogenous tubulin gene.
[3]mRNA-Seq results are represented by FPKM values obtained after normalization with the endogenous tubulin gene.

2.4 DISCUSSION

By using the combination of LCM technique, high-throughput sequencing and the merging of *P. pachyrhizi* NCBI ESTs to our contigs, we generated 36,360 *P. pachyrhizi* unisequences. The total number of transcripts obtained corresponds to 73.3% of the total *P. pachyrhizi* sequences available in the NCBI (49,596 ESTs), but only 23,290 contigs showed similarity between these sequences, suggesting that approximately 36% of the transcripts obtained in this work may still be unknown. Although a good level of similarity with previously identified transcripts has been observed, it is note that, mostly these do not yet present functional annotation. Additionally, the results suggest that about one-third of the generated sequences can be conserved, since they presented similarity to proteins encoded by other phytopathogenic fungi. This similarity is more specifically among other rust fungi, such as *P. graminis* and *M. larici-populina*, which presented more than half (56% and 57%, respectively) of their predicted sequences (Duplessis *et al*., 2011) to the transcripts of *P. pachyrhizi* obtained here.

Functional annotations were assigned to a total of 5,622 contigs, corresponding to only 15.5% of the identified transcripts, but this number of sequences was similar to that observed for other rust species (Garnica *et al*., 2013), and three times than that obtained by Link *et al*. (2014) for *P. pachyrhizi* haustorium transcriptome. In addition, although the number of total *P. pachyrhizi* transcripts identified has been high in our results (36,360 unisequences), the number of genes encoding proteins is expected to be much less, similar to that identified in other rust species, as in *M. larici-populina* and *P. graminis*, which presented 16,399 and 17,773 predicted protein-coding genes in their genomes (Duplessis *et al*., 2011). However,

the lack of access to the *P. pachyrhizi* whole-genome makes it difficult to understand its composition and functioning.

Most of the molecular processes and metabolic pathways observed in our results have already been described in other fungi studies, including rust fungi and *P. pachyrhizi* during the main stages of the infection process. Energy and carbohydrate metabolism were also identified by Tremblay *et al*. (2013), which observed the presence of the genes encoding enzymes involved in these metabolic processes from non-germinated urediniospores until the moment of sporulation. In the same study, after the contact of the non-germinated urediniospores with the host until the beginning of the germination process, it was observed the presence of transcripts involved in oxidative phosphorylation processes and transcription processes, suggesting great energy production for transcription of genes involved in the subsequent stages needs to pathogen development.

Transcriptome analysis of *P. striiformis* (Garnica *et al*., 2013) and *P. pachyrhizi* (Posada-Buitrago and Frederick, 2005; Tremblay *et al*., 2013; Link *et al*., 2014) demonstrated that nucleic acid metabolism (mainly DNA synthesis), transcription processes, cell cycle control and metabolic pathway signaling processes act during the germination of urediniospores. In addition, it is believed that *P. pachyrhizi* does not have access to host nutrients in the early stages of infection, therefore, the synthesis of glycerol (necessary for penetration) occurs from the nutrients obtained from lipids, glycogen and sugar catabolism present in urediniospores, indicating the activity of these metabolic pathways in the degradation of these compounds during germination until the penetration of host tissue (Both *et al*., 2005; Thomas *et al*., 2002).

Transcripts related to the uptake of sugars and amino acids (membrane transporters and carbohydrate metabolism), as well as to lipid metabolism, to more active biosynthetic and transcription processes, can mainly be found during haustorium formation (Both *et al*., 2005; Link *et al*., 2014; Tremblay *et al*., 2013). However, carbohydrate and lipid synthesis, as well as protein synthesis and amino acid metabolism, are also involved in the later stages of fungus development, during uredinia formation and later in sporulation (Both *et al*., 2005; Tremblay *et al*., 2013). Nucleic acid metabolism in *P. pachyrhizi* was mainly represented by transcripts associated with RNA metabolism and DNA replication and repair. Both processes reflect the proliferation of the fungus through the synthesis of proteins and cell division, mainly involved in the process of penetration and the production of urediniospores, as observed in the proteome of hyphae in sporulation of *B. graminis* (Bindschedler *et al*., 2009).

The metabolism of purines which presented the highest number of transcripts identified by the KEEG analysis was described in *M. orizae* as essential for the growth of the fungus in the host cell (Fernandez et al., 2013). In this study a mutant for the SAICAR synthetase-encoding gene (*MoADE1*) show no differences to the appressorium formation or rice cuticle penetration compared to the wild type, but presented greatly reduction in pathogenicity on rice leaves compared to the wild type, indicating that *de novo* adenine biosynthesis is essential for disease development by *M. orizae.* The authors suggest that the attenuated pathogen growth in rice cells observed for mutant strains may be a result of the fungus difficulty to obtaining more complex molecules such as purines, unlike sugars, from the host cells by the invaginated plant-derived plasma membrane, called extra-invasive hyphal membrane (EIHM).

Transcripts related to nitrogen metabolism possibly act after haustorium formation, during the assimilation of compounds from the host. The genome and transcriptome analysis of *M. lini* after the development of haustorium, at six days after inoculation, identified the presence of a putative gene for nitrate reductase, but no transcripts were identified with this function, indicating that this metabolic pathway may not be functional in this specie (Nemri *et al*., 2014), as predicted in most species of rust (Spanu *et al*., 2010, Cantu *et al*., 2011). However, in the same study, Nemri *et al*. (2014) identified homologues of the ammonia assimilation pathway, suggesting that most of the nitrogen acquired through the host is assimilated into the ammonia form. These results corroborate the results observed in this work, in which the sequence related to nitrate reductase gene showed low expression levels for both RNA-Seq and RT-qPCR analyzes, and transcripts identified in the nitrogen metabolism by KEGG, act at the end of the assimilation pathway of this compound, more specifically on ammonia compounds (data not shown). In addition, the analyzes of OrthoMCL families revealed that some genes related to nitrogen metabolism are conserved among the rust species used in this work.

Among the 58 transcripts top expressed in the transcriptome analysis, almost half showed similarity to virulence proteins GAS1 and GAS2 of *M. grisea*. These proteins are virulence factors that act mainly in the initial stages of the infection process during the penetration of host tissue and have already been identified in other fungi, including *P. pachyrhizi* secretoma (Xue *et al*., 2002; Carvalho *et al*., 2016). The DUF3129 domain was one of the most abundant domains found among the most expressed transcripts, and although its function is still unknown, it appears to be conserved among some species of rust and was also identified more specifically in sequences secreted by these fungi, as observed in *P. graminis* and *M.*

larici-populina (Saunders *et al*., 2012) and even in *P. pachyrhizi* (Stone *et al*., 2012, Carvalho *et al*., 2016). The SH2 domain found among transcripts most expressed in resistant genotype (PI561356), is a type of phosphotyrosine signaling, and although it is not much found in fungi, this domain is described as playing a central role in many cell-to-cell communication pathways, including those that regulate proliferation, differentiation, adhesion, hormone responses, and immune defense (Hunter, 2009; Lim and Pawson, 2010). Additionally, three of the top expressed transcripts common between both genotypes (de_novo_2238, de_novo_5381, and de_novo_5849), were also analyzed by Carvalho *et al*. (2016) as putative *P. pachyrhizi* effectors, and were functionally validated by transient overexpression in tobacco leaves, revealing the ability of these sequences to suppress ETI responses.

For the enrichment analyzes, the transcripts expressed in the resistant genotype PI561356, presented enriched molecular categories closely related to the process of infection of the host tissue. In addition to basic mechanisms for the development of the pathogen, such as regulation of cellular development, organization of the extracellular matrix and regulation of growth factor pathways, transcripts related more closely to the pathogenicity of the fungus as to processes of cellular secretion and signaling of metabolic pathways were also enriched. The process of exocytosis or cellular secretion is essential during infection of the host tissue, among other reasons, mainly for the secretion of proteins called effector during the haustorium formation (Catanzariti *et al*., 2006). These effector proteins are responsible for altering the structure and function of the host cell, leading to molecular and physiological changes that facilitate infection and nutrient uptake by the pathogen (Voegele and Mendgen, 2011). Several of putative effector proteins are identified in oomycetes, fungi and rust fungi, and some of these have already been

shown to be directly related to the infection process, as in *B. gramiis* (Pliego *et al*., 2013), in *P. infestans* (Sanju *et al*., 2015) and even in *P. pachyrhizi* (Carvalho *et al*., 2016).

Besides the processes of cellular secretion, the molecular categories of positive regulation of signal transduction and regulation of intracellular protein kinase cascade enriched in PI561356 may be connected in a metabolic signaling pathway in response to defense mechanisms acting on the host cell. The mitogen-activated proteins kinases (MAPKs) are one of the most well-known types of kinases. MAPKs act in general on the transmission of stress signals from receptors to specific effectors that regulate gene expression, cell growth and differentiation in various processes of development and adaptation of different organisms (Moustafa *et al*., 2014). The silencing of a gene encoding a MAPK in *P. triticinia* showed suppression of the disease in the host, revealing the intimate relationship of this gene with the pathogenicity of this rust fungus (Panwar *et al*., 2013).

Different molecular classes were also enriched among transcripts common to *P. pachyrhizi* infection structures. Modifications and methylations in mRNAs molecules were enriched processes between the sequences found in germinated urediniospores, appressorium, and transcripts obtained from leaf lesions. These molecular classes indicate that the common transcripts between these structures are basically involved in transcription and mRNA processing processes. As previously mentioned, transcription processes are very active during most of the infection process, mainly in the stages of germination until the penetration of the host tissue.

When we compared our transcriptome to the *P. pachyrhizi* germinated urediniospores and appressorium ESTs, and to the haustorium transcripts, it was possible note that we could access transcripts related to all these biological

structures, but mainly to the haustorium. This was expected due, besides the greater number of sequences coming from the dataset of haustorium transcriptome (Link et al., 2014), also as a result from the mesophyll cells isolation just below rust lesions, which were probably rich in haustorium structures at 10 days after inoculation. At this time, the pathogen had already completed one reproduction cycle and started the next one. Hacquard et al. (2010), also by using the LCM technique, isolate different portions of uredinia formed by *M. larici-populina* on the susceptible poplar leaves, as spores and sporogenous hyphae, fungal infection tissues in spongy mesophyll, and fungal infection tissues in palisade mesophyll. The exon oligoarrays were used to measure transcripts expression in these areas, and the results showed the identification of genes associated with biotrophy expressed in the last one. Among these sequences were identified a massive induction of sequences encoding putative effector proteins, supporting the maintenance of biotrophy at the late infection stages. As reported by these authors, the use of LCM to collect samples, provide a good preservation of plant and fungal cell structures, which determine integrity of RNA isolated from microdissected tissues.

Among the categories enriched for common transcripts between haustorium sequences and leaf lesions, the molecular categories of sulfate transport stand out. Usually, sulfate is taken up by the fungi and then is converted in cysteine precursor (Marzluf, 1997), but in some rust fungi species as *B. graminis* (Spanu *et al*., 2010) and *P. graminis* (Duplessis *et al*., 2011) the genes encoding enzymes related to sulfate uptake and reduction were not identified in their genomes. Despite this, Garnica *et al*. (2013) and Castillejo *et al*. (2010), found evidence of sulfur metabolism in *P. striiformis* and *Uromyces striatus*, respectively, corroborate our results.

Another molecular category enriched that draws attention is the response to reactive oxygen species (ROS). In plants are known that ROS act on defense mechanisms against pathogen infection, including programmed cell death (Dangl and Jones, 2001). ROS are generated by several different enzymes, but NADPH oxidases is one of the most known, and in fungi ROS are involved in regulation of variety of cellular physiological and differentiation processes, as defense and infection processes (Takemoto *et al*., 2007). Functional analyzes by deletions of the single NADPH oxidase gene from *Podospora anserina* (*Nox1*) (Malagnac *et al*., 2004) and *Neurospora crassa* (*nox-1*) (Aguirre *et al*., 2005), demonstrating that production of ROS is critical for sexual fruiting body development in filamentous fungi. In *M. grisea* the ROS production was observed mainly during appressorium development, and two genes that encode NADPH oxidases are proved as required for pathogenicity of this fungus (Egan *et al*., 2007). Sequences similar to NADPH oxidases are already identified in *P. pachyrhizi* transcripts found in germinated urediniospores and appressorium (Stone *et al*., 2012).

Formaldehyde degradation was the only enriched molecular category among transcripts common to all infection structures analyzed. The degradation of this compound has been reported in some species of fungus found in the industrial wastewater, as in *Aspergillus* sp. (Yu *et al*., 2015), but nothing has been described about the metabolism or degradation of formaldehyde in phytopathogenic fungi.

The comparative analysis between *P. pachyrhizi* and other 15 species of fungi and oomycetes identified that the largest number of conserved sequences among these species were grouped into ribosomal protein families. Ribosomal proteins play an important role in all organisms, making translation possible, and among the large number of families, many are conserved among species of Bacteria, Archaea and

Eucarya (including fungi species) (Lecompte *et al*., 2002). Among eukaryotic organisms, phylogenetic analysis involving sequences of ribosomal proteins showed that families of these proteins present in Plantae and Animalia species are more related and closer than other present in fungi species, which form a more distant clade (Veuthey and Bittar, 1998). In addition, specifically in fungi, Tanay *et al.*, (2005) suggest that some corregulated responses related to ribosomal proteins may be conserved even though the underlying regulatory mechanisms are changing, and this can be explained by the formation of a redundant intermediate program. These results make it possible to infer that in addition to conserving the sequence and structure of ribosomal proteins, other mechanisms may be involved in maintaining the function of these proteins between different fungi species.

The two families of membrane transporters identified among all species also deserve attention in this discussion because of the high sequence numbers they contained. Both are families of ABC (ATP binding cassette) membrane transporters, which plays an important role in the transport of various substances (Rees *et al*., 2009) and have already been described in different fungi species. In *P. striiformis* were identified twelve transcripts similar to ABC transporters, and although it has not been possible to establish their biological functions, three of these sequences were upregulated in germinated spores (Garnica *et al*., 2013). Already in *M. oryzae*, three ABC transporter genes (*ABC1*, *ABC3* and *ABC4*) seem to be directly related with the mechanisms of pathogenicity of the fungus, during appressorium formation and in the penetration of the host tissue, reflecting the possible role of these sequences in the protection of pathogen cells, excluding defense molecules secreted by plants, and also secreting secondary metabolites extremely necessary for the colonization of the

host tissue (Urban *et al*., 1999; Sun *et al*., 2006; Gupta and Chattoo, 2008; Soanes *et al*., 2012).

For the rust fungi used in the comparative analysis, conserved sequences among these species were identified in multigene families related to carbohydrate and protein metabolism, transmembrane transport, and vesicular fusion. Sequences related to the three first molecular processes were also previously mentioned as identified in other phytopathogenic fungi, including rust fungi, during the main infection structures development. The sequences involved in the vesicular fusion process are required to the *U. maydis* and *M. oryzae* development in host tissue and to the pathogenicity, acting on both uptake and secretion mechanisms of substances (Fuchs *et al*., 2006; Qi *et al*., 2016). Additionally, Carvalho et al. (2016) demonstrated that two sequences of putative effectors proteins, presents in one same secreted protein family conserved among fungi in our results, are able to suppress ETI responses in tobacco leaves by overexpression.

Transposable elements can drastically interfere in the composition and expression of a genome. The transposons or retrotransposons movement into or near genes, can contribute to partial or total gene inactivation, act in the regulation of gene expression, and may still contribute to a large genotypic and phenotypic variety. In the last decades the knowledge about TEs in fungi has increased greatly due to the growing number of studies involving fungi of medicinal, agronomic and biotechnological importance, including filamentous fungi (Daboussi *et al*., 2003). TEs were identified in sequenced genomes of other rust fungi, as example in *B. graminis*, which TEs correspond to 64% to genome size (Spanu *et al*., 2010), *M. larici-populina* and *P. graminis*, which TEs account for about 45% of both genomes (Duplessis et al., 2011), and in *P. striiformis*, which TEs represent 17,8% of contig sequences

generated (Cantu *et al*., 2011). TEs were also identified in *P. pachyrhizi* studies (Posada-Buitrago and Frederick, 2005; Tremblay *et al*., 2009; Stone *et al*., 2012; Link *et al*., 2014), but here we provide an overall classification of the active TEs in the transcriptome of this fungus for the first time, including the identification of superfamilies in each class. We found a total of 592 *P. pachyrhizi* sequences with TEs, representing 1.63% of all transcriptome, and of the TEs identified the majority (81.76%) was of retrotransposons. Many of the superfamilies identified among TE classes have already been identified in rust fungi *P. striiformis* genome (Cantu *et al*., 2011), as transposons superfamilies Tc1-Mariner, PIF-Harbinger, EnSpm, hAT, MuDR, P and Helitron, and retrotransposons superfamileis Tad1, Copia and Gypsy, the latter two which were the most representative superfamilies in our results (61.7% of all transposable elements), and are very common in other fungi (Daboussi *et al*., 2003).

TEs can interact with the genome by means of insertions, excisions, aberrant transpositions and even causing chromosomal rearrangements. This genomic environment becomes a source of relevant variability, especially in species with no sexual cycle (Spanu, 2012). In phytopathogenic filamentous fungal species, effector genes were identified close to regions rich in TEs, as dispensable chromosomes or telomeres (Orbach *et al*., 2000; Haas *et al*., 2009; Ma *et al*., 2010; Balesdent *et al*., 2013), which may result in selective advantages to these organisms, allowing a rapid response to selection of resistance genes, as observed by Raffaele *et al*. (2010) in *P. infestans*. In plants and animals, TEs are normally stabilized during growth and development processes and can be activated by stress (Grandbastien, 1998; Capy *et al*., 2000), but little is known about the mechanisms of control of the activity of these elements in the fungi genome. Some processes are proposed to explain this

regulation in fungi, as alternative splicing (Kempken and Kück, 1996) and homology-dependent process, such as quelling and repeat-induced point mutation (Selker, 1999; Faugeron, 2000; Cogoni, 2001; Daboussi *et al*., 2003). Although it is not as well-known as quelling, evidence of repeat-induced point mutation was already observed in other fungi species as *P. anserina* (Graia *et al*., 2001; Bouhouche *et al*., 2004; Arnaise *et al*., 2008), *M. grisea* (Ikeda *et al*., 2002), *Leptosphaeria maculans* (Idnurm and Howlett, 2003), and *Nectria haematococca* (Coleman *et al*., 2009).

RT-qPCR has been used frequently for the validation of expression levels observed for specific genes through the results of RNA-Seq. Hacquard *et al*. (2010), after obtaining the *M. larici-populina* transcriptome from uredinia stage dissected by LCM technique from poplar leaves infected, validated the exon oligoarray expression profile of the 29 transcripts encoding small secreted proteins and known rust protein homologs by RT-qPCR analyzes, validating the transcriptomic approach. In *P. pachyrhizi*, Tremblay *et al*. (2013) confirmed RNA-Seq differential gene expression for seven genes (alpha-tubulin, NADH dehydrogenase, ribulose-1,5-bisphosphate carboxylase oxygenase, pectin methylesterase, maturase-related, serine palmitoytransferase, and 60S ribosomal protein) also using RT-qPCR. Despite finding some differences between the genes expressions profiles, as the qPCR expressions detected at time-points where mRNA-Seq did not, both types of detection of expression levels were well related in general. Our validation of RNA-Seq expression profile by RT-qPCR show similar results. The observed expression pattern for the six genes selected in the RNA-Seq data was very similar in the data obtained after the RT-qPCR analyzes, in which five of the six genes analyzed were induced.

The results presented in this study enrich the available knowledge about the *P. pachyrhizi* transcriptome, corroborating molecular mechanisms already identified in previous studies, and also providing new perspectives on processes still unknown in this pathogen. The whole-genome sequencing of this fungus as well as the functional characterization of genes related to pathogenicity during the soybean infection process are extremely important, contributing even more to the search for molecular mechanisms that may aid in the control of the disease.

ACKNOWLEDGMENTS

This work was supported by grant for the Coordination of Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

2.5 REFERENCES

Aguirre J, Rios-Momberg M, Hewitt D, Hansberg W (2005) Reactive oxygen species and development in microbial eukaryotes. Trends Microbiol. 13: 111-118.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. Nucleic Acids Research 25(17):3389–3402.

Anderson SJ, Stone CL, Posada-Buitrago ML, Boore JL, Neelam BA, Stephens RM, Luster DG, Frederick RD and Pedley KF (2008) Development of simple sequence repeat markers for the soybean rust fungus, *Phakopsora pachyrhizi*. Molecular Ecology Resources 8:1310–1312.

Arnaise S, Zickler D, Bourdais A, Dequard-Chablat M, Debuchy R (2008) Mutations in mating-type genes greatly decrease repeat-induced point mutation process in the fungus *Podospora anserina*. Fungal Genetics Biology 45: 207-220.

Balesdent MH, Fudal I, Ollivier B, Bally P, Grandaubert J, Eber F, Chèvre AM, Leflon M and Rouxel T (2013) The dispensable chromosome of Leptosphaeria maculans shelters an effector gene conferring avirulence towards Brassica rapa. New Phytologist, 198: 887-898.

Baudet C and Dias Z (2005) New EST Trimming Strategy. Advances in Bioinformatics and Computational Biology 3594:206–209.

Bindschedler LV, Burgis TA, Mills DJS, Ho JTC, Cramer R and Spanu PD (2009) *In Planta* Proteomics and Proteogenomics of the Biotrophic Barley Fungal Pathogen *Blumeria graminis* f. sp. *hordei*. Molecular & Cellular Proteomics 8:2368-2381.

Boetzer M, Henkel CV, Jansen HJ, Butler D and Pirovano W (2011) Scaffolding Pre-Assembled Contigs Using SSPACE. Bioinformatics 27(4):578–579.

Both M, Csukai M, Stumpf MPH and Spanu PD (2005) Gene Expression Profiles of *Blumeria graminis* Indicate Dynamic Changes to Primary Metabolism during Development of an Obligate Biotrophic Pathogen. The Plant Cell 17:2107–2122.

Bouhouche K, Zickler D, Debuchy R, Arnaise S (2004) Altering a gene involved in nuclear distribution increases the repeat-induced point mutation process in the fungus *Podospora anserina*. Genetics 167: 151-159.

Cai S and Lashbrook CC (2006) Laser capture microdissection of plant cells from tape-transferred paraffin sections promotes recovery of structurally intact RNA for global gene profiling. The Plant Journal 48:628–637.

Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Jurka KKKJ, Michelmore RW and Dubcovsky J (2011) Next Generation Sequencing Provides Rapid Access to the Genome of *Puccinia striiformis* f. sp. *tritici*, the Causal Agent of Wheat Stripe Rust. Plos One 6:e24230.

Carvalho MCCG, Nascimento LC, Darben LM, Polizel-Podanosqui AM, Lopes-Caitar VS, Qi M, Rocha CS, Carazzolle MF, Kuwahara MK, Pereira GAG, Abdelnoor RV, Whitham SA and Marcelino-Guimarães FC (2016) Prediction of the *in planta P. pachyrhizi* secretome and potential effector families. Molecular Plant Pathology 1-15.

Capy P, Gasperi G, Biemont C and Bazin C (2000) Stress and transposable elements: co-evolution or useful parasites? Heredity 85:101–106.

Catanzariti AM, Dodds PNVET, Kobe B, Ellis JG and Staskawicz BJ (2010) The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. MPMI 23:49–57.

Cogoni C (2001) Homology-dependent gene silencing mechanisms in fungi. Annu. Rev. Microbiol. 55:381–406.

Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, Schmutz J, Taga M, White GJ, Zhou S, *et al*. (2009) The Genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PloS Genetics 5: e1000618.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics Applications Note 21(18):3674–3676.

Daboussi MJ and Capy P (2003) Transposable Elements in Filamentous Fungi. Annu. Rev. Microbiol. 57:275–99.

Dangl JL and Jones JDG (2001) Plant pathogens and integrated defense responses to infection. Nature 411:826–33.

Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, *et al*. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. U.S.A. 108(22):9166–9171.

Egan MJ, Wang ZY, Jones MA, Smirnoff N and Talbot NJ (2007) Generation of reactive oxygen species by fungal NADPH oxidases is required for rice blast disease. PNAS 104(28):11772–11777.

Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA (1996) Laser Capture Microdissection. Science 274:998-1001.

Faugeron G (2000) Diversity of homology-dependent gene silencing strategies in fungi. Curr. Opin. Microbiol. 3:144–48.

Fehr WR, Caviness CE, Burmood DT and Pennington SJ (1971) Stage of Development Descriptions for Soybeans, *Glycine Max* (L.) Merrill. Crop Science 11(6):929–931.

Fernandez J, Yang KT, Cornwell KM, Wright JD and Wilson RA (2013) Growth in rice cells requires *de novo* purine biosynthesis by the blast fungus *Magnaporthe oryzae*. Scientific Reports 3:2398.

Fuchs U, Hause G, Schuchardt I and Steinberga G (2006) Endocytosis Is Essential for Pathogenic Development in the Corn Smut Fungus *Ustilago maydis*. The Plant Cell 18:2066–2081.

Garnica DP, Upadhyaya NM, Dodds PN and Rathjen JP (2013) Strategies for Wheat Stripe Rust Pathogenicity Identified by Transcriptome Sequencing. Plos One 8:e67150.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al*. (2011) Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. Nature Biotechnology 29(7):644–652.

Graia F, Lespinet O, Rimbault B, Dequard-Chablat M, Coppin E, Picard M (2001) Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. Mol. Microbiol. 40:586–95.

Grandbastien MA (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci. 3:181–87.

Gupta A and Chattoo BB (2008) Functional analysis of a novel ABC transporter *ABC4* from *Magnaporthe grisea*. FEMS Microbiol Lett 278: 22-28.

Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, *et al*. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461: 393-398.

Hacquard S, Delaruelle C, Legué V, Tisserant E, Kohler A, Frey P, Martin F and Duplessis S (2010) Laser Capture Microdissection of Uredinia Formed by *Melampsora larici-populina* Revealed a Transcriptional Switch between Biotrophy and Sporulation. MPMI 23(10):1275–86.

Hahn M and Mendgen K (1997) Characterization of In Planta–Induced Rust Genes Isolated from a Haustorium-Specific cDNA Library. MPMI 10(4):427–437.

Huang X and Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Research 9(906):868–877.

Hunter T (2009) Tyrosine phosphorylation: thirty years and counting. Current Opinion in Cell Biology 21:140–146.

Idnurm A and Howlett BJ (2003) Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. Fungal Genetics and Biology 39(1): 31–37.

Ikeda K, Nakayashiki H, Kataoka T, Tamba H, Hashimoto Y, Tosa Y, Mayama S (2002) Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. Mol. Microbiol. 45:1355–64.

Kapitonov VV and Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nature Reviews Genetics 8:973-982.

Kempken F and Kück U (1996) Restless an active Ac-like transposon from the fungus *Tolypocladium inflatum*: structure, expression, and alternative RNA splicing. Mol. Cell Biol. 16:6563–72.

Kerk NM, Ceserani T, Tausta SL, Sussex IM and Nelson TM (2003) Laser Capture Microdissection of Cells from Plant Tissues. Plant Pathology 132(1):27–35.

Kim KS, Unfried JR, Hyten DL, Frederick RD, Hartman GL, Nelson RL, Song Q and Diers BW (2012) Molecular Mapping of Soybean Rust Resistance in Soybean Accession PI 561356 and SNP Haplotype Analysis of the *Rpp1* Region in Diverse Germplasm. Theoretical and Applied Genetics 125(6):1339–1352.

Lecompte O, Ripp R, Thierry JC, Moras D and Poch O (2002) Comparative analysis of ribosomal proteins incomplete genomes: an example of eductive evolution at the domain scale. Nucleic Acids Research 30(24):5382-5390.

Li L, Stoeckert JR CJ and Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13:2178–2189.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K and Wang, J. (2009) SOAP2: An Improved Ultrafast Tool for Short Read Alignment. Bioinformatics 25(15):1966–1967.

Lim WA and Pawson T (2010) Phosphotyrosine Signaling: Evolving a New Cellular Communication System. Cell 142:661-667.

Link T, Lang P, Scheffler BE, Duke MV, Graham MA, Cooper B, Tucker ML, Van de Mortel M, Voegele RT, Mendgen K, Baum TJ and Whitham SA (2014) The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. Molecular Plant Pathology 15(4):379–393.

Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, *et al*. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464: 367-373.

Maciel TEF, Freire MCM, Almeida AMR and Oliveira LO (2010) Molecular Characterization of Beta-Tubulin from *Phakopsora pachyrhizi*, the Causal Agent of Asian Soybean Rust. Molecular Plant Pathology 11(2):169–177.

Malagnac F, Lalucque H, Lepère G, Silar P (2004) Two NADPH oxidase isoforms are required for sexual reproduction and ascospore germination in the filamentous fungus *Podospora anserina*. Fungal Genet. Biol. 41: 982-997.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, *et al*. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39:225–229.

Marzluf GA (1997) Molecular genetics of sulfur assimilation in filamentous fungi and yeast. Annu Rev Microbiol 51:73–96.

Morales AMAP, Borém A, Graham MA and Abdelnoor RV (2012) Advances on molecular studies of the interaction soybean - Asian rust. Crop Breeding and Applied Biotechnology 12:1-7.

Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B (2008) Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. Nature Methods 5(7):621–628.

Moustafa K, AbuQamar S, Jarrar M, Al-Rajab AJ and Trémouillaux-Guiller J (2014) MAPK cascades and major abiotic stresses. Plant Cell Reports 9p.

Nemri A, Saunders DGO, Anderson C, Upadhyaya NM, Win J, Lawrence GJ, Jones DA, Kamoun S, Ellis JG and Dodds PN (2014) The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. Frontiers in Plant Science 5(98):1-14.

Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B (2000) A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. Plant Cell 12: 2019–2032.

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29(9)2002-2007.

Posada-Buitrago ML and Frederick RD (2005) Expressed sequence tag analysis of the soybean rust pathogen *Phakopsora pachyrhizi*. Fungal Genetics and Biology 42:949–962.

Qi Z, Liu M, Dong Y, Zhu Q, Li L, Li B, Yang J, Li Y, Ru Y, Zhang H, Zheng X, Wang P and Zhang Z (2016) The syntaxin protein (MoSyn8) mediates intracellular trafficking to regulate conidiogenesis and pathogenicity of rice blast fungus. New Phytologist 209:1655–1667.

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, *et al*. (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science 330: 1540-1543.

Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. Nat Rev Mol Cell Biol 10: 218-227.

Ribeiro AS, Moreira JUV, Pierozzi PHB, Rachid BF, Toledo JFF, Arias CAA, Soares RM and Godoy CV (2007) Genetic Control of Asian Rust in Soybean. Euphytica 157(1-2):15–25.

Sanju S, Siddappa S, Thakur A, Shukla PK, Srivastava N, Pattanayak D, Sharma S and Singh BP (2015) Host-mediated gene silencing of a single effector gene from the potato pathogen *Phytophthora infestans* imparts partial resistance to late blight disease. Funct Integr Genomics 15:697-706.

Saunders DGO, Win J, Cano LM, Szabo LJ, Kamoun S and Raffaele S (2012) Using Hierarchical clustering of Secreted Protein Families to Classify and Rank Candidate Effectors of Rust Fungi. PLoS One 7:e29847.

Scherm H, Christiano RS, Esker PD, Ponte EM and Godoy CV (2009) Quantitative review of fungicide efficacy trials for managing rust in Brazil. Crop Protection 28:774-782.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al*. (2010) Genome Sequence of the Palaeopolyploid Soybean. Nature 463:178–83.

Selker EU (1999) Epigenetic phenomena in filamentous fungi: useful paradigms or repeated-induced confusion? Trends Genet. 13:296–301.

Soanes DM, Chakrabarti A, Paszkiewicz KH, Dawe AL, Talbot NJ (2012) Genome-wide transcriptional profiling of appressorium development by the rice blast fungus Magnaporthe oryzae. PLoS Pathog 8: e1002514.

Spanu PD (2012) The Genomics of Obligate (and Nonobligate) Biotrophs. Annual Reviews Phytopathology 50:91-109.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ, *et al*. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330:1543–1546.

Stone CL, McMahon MB, Fortis LL, Nuñez A, Smythers GW, Luster DG and Frederick RD (2012) Gene expression and proteomic analysis of the formation of *Phakopsora pachyrhizi* appressoria. BMC Genomics 13:269.

Sun CB, Suresh A, Deng YZ, Naqvi NI (2006) A multidrug resistance transporter in *Magnaporthe* is required for host penetration and for survival during oxidative stress. Plant Cell 18: 3686-3705.

Tanay A, Regev A and Shamir R (2005) Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast. PNAS 102(20):7203-7208.

Thomas SW, Glaring MA, Rasmussen SW, Kinane JT and Oliver RP (2002) Transcript profiling in the barley mildew pathogen *Blumeria graminis* by serial analysis of gene expression (SAGE). MPMI 15:847–856.

Trapnell C, Pachter L and Salzberg SL (2009) TopHat: Discovering Splice Junctions with RNA-Seq. Bioinformatics 25(9):1105–1111.

Tremblay A, Hosseini P, Li S, Alkharouf NW and Matthews BF (2012) Identification of genes expressed by *Phakopsora pachyrhizi*, the pathogen causing soybean rust, at a late stage of infection of susceptible soybean leaves. Plant Pathology 61:773–786.

Tremblay A, Hosseini P, Li S, Alkharouf NW and Matthews BF (2013) Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. BMC Genomics 14:614.

Tremblay A, Li S, Scheffle BE and Matthews BF (2009) Laser capture microdissection and expressed sequence tag analysis of uredinia formed by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. Physiological and Molecular Plant Pathology 1-12.

Urban M, Bhargava T, Hamer JE (1999) An ATP-driven efflux pump is a novel pathogenicity factor in rice blast disease. EMBO J 18: 512-521.

Veuthey AL and Bittar G (1998) Phylogenetic Relationships of Fungi, Plantae, and Animalia Inferred from Homologous Comparison of Ribosomal Proteins. J Mol Evol 47:81-92.

Voegele RT and Mendgen KW (2011) Nutrient uptake in rust fungi: how sweet is parasitic life? Euphytica 179:41–55.

Voegele RT, Hahn M and Mendgen K (2009) The Uredinales: cytology, biochemistry, and molecular biology. In: Deising, H. (ed). The Mycota V. Plant Relationships 69–98.

Webb CA, Szabo LJ, Bakkeren G, Garry C, Staples RC, Eversmeyer M and Fellers JP (2006) Transient expression and insertional mutagenesis of *Puccinia triticina* using biolistics. Functional & Integrative Genomics 6:250–260.

Xue C, Park G, Choi W, Zheng L, Dean RA and Xu JR (2002) Two Novel Fungal Virulence Genes Specifically Expressed in Appressoria of the Rice Blast Fungus. The Plant Cell 14:2107–2119.

Yu DS, Song G, Song LL, Wang W and Guo CH (2015) Formaldehyde degradation by a newly isolated fungus *Aspergillus* sp. HUA. Int. J. Environ. Sci. Technol. 12:247–254.

Zambolin L (2006) Manejo integrado da ferrugem asiática da soja. In: Zambolin L (ed.) Ferrugem asiática da soja, Suprema, Viçosa, pp 139.

INTERNET RESOURCES

LGE database, http://bioinfo03.ibi.unicamp.br/phakopsora.

WU BLAST. Washington University BLAST - http://blast.wustl.edu (March 10, 2015)

OrthoMCL software version 1.4
(http://orthomcl.org/common/downloads/software/unsupported/v1.4/).

RepeatMasker parameters “selective and matches to coding sequences”, available in http://www.repeatmasker.org/webrepeatmaskerhelp.html.

2.6 SUPPLEMENTARY MATERIAL

Table S1. The 50 top expressed *Phakopsora pachyrhizi* transcripts at 10 days post soybean infection, based on the FPKM values, identified for each soybean genotypes PI561356 and BRS 231.

Contig	Annotation against NCBI database	FPKM values	
		BRS 231	PI561356
Common to both genotypes			
de_novo_11027	*P. pachyrhizi* clone JGIAFNA-33E19	98,342.72	48,807.68
de_novo_23913	*M. larici-populina* secreted protein with DUF3129 domain	19,815.81	29,193.11
de_novo_23959	*M. larici-populina* secreted protein with DUF3129 domain	20,080.94	27,905.95
de_novo_9895	*M. larici-populina* secreted protein with DUF3129 domain	15,173.93	21,918.25
de_novo_10151	*M. larici-populina* secreted protein with DUF3129 domain	13,719.74	18,582.40
de_novo_5849	*M. larici-populina* secreted protein with DUF3129 domain	11,467.69	16,627.02
de_novo_33003	Protein of unknown function with DUF3129 domain	11,039.74	15,825.92
de_novo_32708	No hits found	8,469.59	14,466.77
de_novo_4830	*P. pachyrhizi* clone JGIAFNA-13C12	8,744.68	13,596.40
de_novo_12325	*M. larici-populina* hypothetical protein with DUF3129 domain	9,392.71	13,079.84
de_novo_40849	No hits found	7,541.34	13,034.33
de_novo_9851	*M. larici-populina* secreted protein with DUF3129 domain	8,353.40	12,345.03
de_novo_7016	Conserved hypothetical protein with MAEBL domain	7,907.53	12,282.19
de_novo_3845	No hits found	7,849.31	12,225.83
de_novo_2238	No hits found	8,845.75	12,136.21
de_novo_5381	*M. larici-populina* secreted protein with DUF3129 domain	8,544.68	11,635.99

de_novo_6669	No hits found	5,936.03	11,509.40
de_novo_3100	No hits found	7,480.70	11,489.08
de_novo_3210	No hits found	7,388.56	11,382.46
de_novo_1402	No hits found	7,457.41	10,509.42
de_novo_3449	No hits found	5,986.86	10,209.79
de_novo_2152	No hits found	6,538.66	10,157.06
de_novo_2142	No hits found	6,573.82	10,099.46
de_novo_1770	No hits found	6,106.69	9,494.88
de_novo_2721	*P. pachyrhizi* clone JGIAFNA-5A11	5,945.71	9,249.17
de_novo_6942	No hits found	3,659.44	9,162.98
de_novo_6949	Conserved hypothetical protein with MAEBL domain	6,131.29	9,062.17
de_novo_2592	No hits found	5,847.87	9,060.92
de_novo_1478	No hits found	5,834.43	8,954.22
de_novo_8907	*P. pachyrhizi* clone JGIAFNA-33E19	1,9014.19	7,701.48
de_novo_2285	No hits found	3,954.02	7,675.55
de_novo_555	No hits found	4,633.16	7,167.71
de_novo_5657	Conserved hypothetical protein with MAEBL domain	4,650.95	6,833.45
de_novo_2037	No hits found	4,287.27	6,671.67
de_novo_38206	No hits found	4,250.22	6,600.62
de_novo_8211	DNA polymerase III subunits gamma and tau domain	6,113.70	6,181.58
de_novo_6732	*M. larici-populina* secreted protein with DUF3129 domain	3,837.32	5,672.38
de_novo_6277	Conserved hypothetical protein with MAEBL domain	3,730.86	5,508.55
de_novo_11704	Conserved hypothetical protein with MAEBL domain	3,733.27	5,302.30
de_novo_1254	No hits found	5,021.69	4,837.57
de_novo_1491	RNA polymerase I-associated factor PAF67	4,196.92	4,833.46
de_novo_9082	*P. pachyrhizi* clone JGIAFNA-33E19	7,804.53	4,547.20

Only for PI561356			
de_novo_13814	Galactose-3-O-sulfotransferase family	440.91	5,945.22
de_novo_4216	No hits found	3,111.34	4,925.88
de_novo_2312	No hits found	3,224.50	4,869.87
de_novo_9222	Membrane protein TolA domain (cell envelope biogenesis, outer membrane)	3,376.32	5,091.02
de_novo_6032	Protein of unknown function with DUF3129 domain	3,230.27	4,744.18
de_novo_1182	*M. larici-populina* hypothetical protein with SH2 domain	3,121.50	4,739.35
de_novo_3323	No hits found	3,020.47	4,637.93
de_novo_947	No hits found	3,454.19	4,616.23
Only for BRS 231			
de_novo_18573	Fibronectin-attachment protein domain	6,047.65	4,435.17
de_novo_3301	No hits found	5,827.95	2,238.76
de_novo_7858	No hits found	4,585.47	3,478.69
de_novo_2057	No hits found	4,390.65	3,960.86
de_novo_7893	Large tegument protein UL36 domain	3,995.37	3,909.36
de_novo_6455	DNA polymerase III subunits gamma and tau domain	3,965.85	2,812.26
de_novo_5057	No hits found	3,755.03	3,636.51
de_novo_7366	No hits found	3,751.33	3,755.50

3. CAPÍTULO 3: Functional characterization of *Phakopsora pachyrhizi* genes involved in the pathogenicity by Host-Induced Gene Silence

ABSTRACT

Rust fungi are destructive pathogens that affect production of many plant commodities worldwide. Asian soybean rust (ASR), caused by the fungus *Phakopsora pachyrhizi* (Pp), is one of the most important disease in soybean, reaching all soybean production regions. The development of resistant cultivars is the preferred strategy, but the ability of such pathogen to evolve new pathotypes makes obtaining these cultivars a difficult task. Host induced gene silencing (HIGS) constitute a promise alternative to identify pathogen genes that might be involved in infection process or be essential for the fungal survival, helping to find and develop more efficient strategies to control the disease. In this study, we mining *P. pachyrhizi* predicted transcripts involved in the survival and pathogenic processes, previously obtained from lesions of infected soybean leaves transcriptome, to be characterized by HIGS. Seven candidate genes were selected based on literature, expression profile and phylogenetic analysis. Initially, transcripts sequences were mined from Pp transcriptome and compared to other species of fungi and oomycetes, followed by expression profile analysis across pathogen infection cycle in soybean and in urediniospores of two Brazilian isolates. The selected genes potentially involved in post transcriptional gene silencing (PTGS) machinery, thiamine biosynthesis, signaling metabolic pathways, protein folding, vesicle fusion and ribosomal proteins were silenced using BPMV-HIGS strategy. The level of silencing was determinate by RT-qPCR, being more evident in specific time points, consistent with where the genes presented higher levels of induction in expression analysis. The phenotypic evaluation of four parameters revealed a significant reduction in disease symptoms for the same three genes that the presented higher levels compared with non-silenced plants (infected with empty vector). The reduction in disease symptoms indicated the probable relationship of these genes in the survival and pathogenicity processes of *P. pachyrhizi* during soybean infection. This study was the first to demonstrate that BPMV-HIGS approach can be an effective strategy for functional characterization of *P. pachyrhizi* genes.

Keywords: Asian soybean rust, gene silencing, BPVM

3.1 INTRODUCTION

Asian soybean rust (ASR) is caused by the rust fungi *Phakopsora pachyrhizi* (Sydow & P. Sydow), one of the main pathogens that reduce soybean productivity up to 80%, revealing a great destructive power (YORINORI et al., 2005; SCHERM et al., 2009). As well as other fungi that cause rust, the dynamics between *P. pachyrhizi* and its host is complex, mainly because it is an obligatory biotrophic fungus, making it difficult to understand and control the disease. The challenges presented by this rust fungus therefore demand a continuing search for innovative strategies to elucidate the mechanisms involved in the infection process, not only related to host responses, but also to pathogen behavior.

Host-induced gene silence (HIGS) tool is based on RNA interference (RNAi) mechanism, where small RNA targeting pathogen genes are produced in the host and sequentially transferred to the pathogen (BAULCOMBE, 2015), potentially by the haustorium, in the case of haustorium-forming by pathogens, and silencing target genes. It was initially demonstrated in *Fusarium verticilliodes* and tobacco interaction, when hairpin loops for a β-glucuronidase (GUS) intron expressed in tobacco were able to move and silence the target gene in haustoria-forming transgenic expressing GUS, resulting in a reduced GUS gene expression by the fungi (Tinoco et al. 2010). This means that a molecular silencing signal can move from a plant to a fungus. Nowara et al. (2010) proved this concept by using *Barley stripe mosaic virus* (BSMV) as a vector to transiently express RNA fragments of powdery mildew fungus *Blumeria graminis* 1,3-beta-glucanosyltransferase genes in wheat and by using transformation to stably express RNA hairpin loops of the same in barley. They observed reduced

severity of powdery mildew on the plants. The results implied that RNA produced by the plant can move to the fungus, possibly through the haustorium, to activate gene silencing pathways in the fungus. This can compromise fungal gene expression and protein production, resulting in reduced fungal accumulation and pathogenicity. Consequently, plant-mediated gene silencing appears to be a useful method to test the function of fungal genes.

Recently, similar strategy has been used in different pathosystem, being a valuable tool contributing to the identification genes related to infection process during plant-pathogen interactions, including rust species (ZHANG et al., 2012; KOCH et al., 2013; PLIEGO et al., 2013; PANWAR et al., 2013; GHAG et al., 2014; HU et al., 2015; YIN et al, 2015; CHENG et al., 2015; ANDRADE et al., 2015; SONG & THOMMA, 2016; CHEN et al., 2016; ZHANG et al., 2016; ZHOU et al., 2016; COOPER & CAMPBELL, 2017).

In *Blumeria graminis*, the barley powdery mildew fungus, 15 effector candidate genes were silenced by HIGS using a plasmid pTA30 developed based on *Cauliflower mosaic virus* (CaMV), and eight, among which stand out sequences similar to β-1,3 glucosyltransferases, metallo-proteases, and microbial secreted ribonucleases, presented important role in the fungus development (PLIEGO et al., 2013). Panwar et al. (2013) subsequently showed that a virus vector expressing RNA fragments of general pathogenicity genes of the wheat leaf rust fungus, *Puccinia triticina*, also led to a reduction of rust severity in wheat. They observed in the BSMV-infected wheat the presence of small interfering RNAs (siRNA) for the targeted genes, an indicator of the activation of RNA silencing pathways, and genes encoding other targets besides effector are potencially good candidates for HIGS.

In *Puccinia* species, another study showed that, after being silenced by HIGS using BSMV-derived vector, 10 genes of haustorium (related to glycolytic enzyme, sugar metabolism, thiazol biosynthesis, auxin biosynthesis, amino acid permease, and membrane transport), showed reduction in fungal development (YIN et al., 2015). In addition, Cooper & Campbell (2017) infected bean plants with virus expressing fragments of six *Uromyces appendiculatus* effectors candidate genes, and observed a reduction in disease symptoms for four of the silenced genes, including genes encoding trehalose phosphatase protein, chitinase-like protein, glycoside hydrolase and an protein with an unknown function.

Although transcript sequences of *P. pachyrhizi* has been available from urediniospores, germinated urediniospores, appressorium, haustorium and *in plant* (POSADA-BUITRAGO & FREDERICK, 2005; TREMBLEY et al., 2009, 2012, 2013; STONE et al., 2012; LINK et al., 2014, CARVALHO et al., 2016), no *P. pachyrhizi* gene was functionally tested so far. Additionally, the functional post transcriptional gene silencing (PTGS) machinery has not yet been described for *P. pachyrhizi*, and HIGS strategy was also not yet employed for the functional validation of *P. pachyrhizi* genes.

In this study, we characterized *in silico* genes involved in PTGS machinery *in P. pachyrhizi*, and other processes essentials for fungal survival and pathogenicity mechanisms. Based on *in silico* analysis, at least one codifying genes in each category was identified showing elevated similarities and grouping in phylogenetic trees with others fungi and rust sequences available. All the pathogen genes selected were induced at least once time point during the soybean infecting process in expression analysis using two different fungal isolates. Using the BPMV-based

HIGS approach, the expression of fragments of seven *P. pachyrhizi* target genes in soybean resulted in a significant reduction of expression levels of the three silenced genes (Pv-SNARE, Gpα and 18S), as well as in the disease symptoms. Here, we first demonstrate that BPMV-HIGS approach can be effective in functional characterization of *P. pachyrhizi* genes, and identified the involvement of selected *P. pachyrhizi* genes tested in the survival and pathogenicity processes during soybean infection.

3.2 MATERIALS AND METHODS

3.2.1 Selection of candidate genes

Phakopsora pachyrhizi genes were selected *in silico* for subsequent functional characterization by HIGS strategy. Initially, it was investigated the presence of the transcripts involved in the post transcriptional gene silencing (PTGS) machinery. Transcribed sequences of *Puccinia striiformis* f. sp. *tritici* dicer (JN033211) and argonaut (JN033210) (CANTU et al., 2011), and a *Puccinia graminis* f. sp. *tritici* hypothetical protein with RNAse dependent of RNA polymerase domain (RdRP domain) (XM_003336137.2) were obtained from NCBI (National Center for Biotechnology Information). The conserved domains of these sequences were confirmed with the online software Pfam. The sequences were then aligned against the *P. pachyrhizi* transcriptome (available in http://bioinfo03.ibi.unicamp.br/phakopsora/), previously obtained from Pp lesion in soybean leaves, combining laser microdissection capture and RNA-Seq

(CARVALHO et al, 2016), to search for similar sequences. An additional search for sequences of the PTGS machinery in the *P. pachyrhizi* transcriptome database was also performed using word searches based on the annotations obtained against the NCBI database.

Besides genes involved in the PTGS machinery, genes potentially involved in basal survival mechanisms and the pathogenicity of *P. pachyrhizi* were also selected based on orthologous genes previously described in the literature for other fungi, including rust fungi (HAHN & MENDGEN, 1997; FUCHS *et al*., 2006; WEBB et al., 2006; BAILEY et al., 2010; VIEIRA et al., 2012). The identified orthologous genes were then aligned (BLASTN, e-value cutoff of 1e-5) to *P. pachyrhizi* transcriptome database, allowing the selection of transcripts with greater similarity. Additional search using the annotations obtained against the NCBI database for sequences of the *P. pachyrhizi* transcriptome database was also performed. The FPKM (fragments per kilobase million) values for each transcript obtained from *P. pachyrhizi* transcriptome database were also considered, and all contigs selected was analyzed in the Pfam software (http://pfam.xfam.org/) to confirm the presence of the characteristic domains present in these transcripts.

The selected transcripts of *P. pachyrhizi* were also compared (BLASTP, e-value cut-off of 1e-5) against sets of proteins predicted from publicly available genomes of 15 others species, 12 fungi [10 Basidiomycetes – *M. larici-populina* (Joint Genome Institute - JGI), *P. graminis* f. sp. *tritici* (Broad Institute), *Coprinopsis cinerea* (Broad Institute), *Cryptococcus neoformans* (Broad Institute), *Postia placenta* (JGI), *Lacaria bicolor* (JGI), *Malassezia globosa* (JGI), *Phanerochaete chrysosporium* (JGI), *Sporobolomyces roseus* (JGI, v1) and *Ustilago maydis* (Broad Institute); and

two Ascomycetes - *Neurospora crassa* (Broad Institute) and *Magnaporthe grisea* (Broad Institute)] and two oomycetes genomes [*Phytopthora sojae* (JGI) and *Phytopthora infestans* (Broad Institute)]. In addition to the already sequenced genomes listed above, we also considered the data sets of ESTs (Expressed Sequence Tags) available for rust *Uromyces appendiculatus* and *P. pachyrhizi* (LINK et al., 2014). Phylogenetic tree was constructed using the MEGA 6.0 software (TAMURA et al., 2013). Evolutionary relations were inferred using the Neighbor-Joining method, with Bootstrap = 1,000, and evolutionary distances were calculated using the Poisson correction method.

3.2.3 Expression profiling of *P. pachyrhizi* candidate genes during an infection time course

To verify the expression profile of the selected *P. pachyrhizi* genes, spores (S) and germinated spores (GS) were obtained from the culture of *P. pachyrhizi* fresh spores on detached soybean leaves and maintained in Petri dishes under controlled temperature and humidity conditions in a heated chamber. Germinated spores were obtained from fresh spores deposited in Petri dish containing water solution with addition of 0.04% Tween and remained overnight for 16 hours. In addition, soybean plants with susceptible genotype (Williams 82) to *P. pachyrhizi* were maintained in a greenhouse for inoculation of the pathogen and subsequent collection of the samples. Plants were divided into two experiments, temporally separated and containing two different Brazilian monouredinial isolates of *P. pachyrhizi*, LUB112 and LD5511, obtained at Embrapa Soybean, Londrina, Brazil. These isolates differed

in relation to the virulence pattern in the soybean differentiating genotypes, where LUB112 presented virulence in the genotypes with resistance genes *Rpp1, Rpp5* and *Rpp6*, while the LD5511 presented virulence in the genotypes with resistance genes *Rpp1, Rpp2*, *Rpp3* and *Rpp5* (DARBEN et al., 2013). Both experiments followed a completely randomized design and consisted each of three biological replicates with three plants per pot. Inoculations were performed on the second trifoliate of the susceptible plants using water solution with addition of 0.04% Tween containing 10^5 spores mL^{-1}. To each experiment, infected leaves of these compatible interactions were collected at different times after inoculation (0, 6, 12, 24, 48, 72, 96 and 192 hours post inoculation “hpi") to represent the progression of infection and colonization of the host tissue by the fungus. Once collected, fungus structures and plant were immediately frozen in liquid nitrogen and stored at -80 °C until the RNA extraction and subsequently RT-qPCR analyzes.

3.2.4 Construction of HIGS plasmids

Seven primer pairs (Table S1) were used for PCR amplification using cDNA bulk from expression analysis experiment as template. The PCR products were directionally cloned into RNA2 of the BPMV (Bean Pod Mottle Virus) HIGS vector. The BPMV RNA2 vector used in this study, pBPMV-IA-V2 (1033) (ZHANG et. al., 2009; ZHANG et. al., 2010), have *BamH1* and *XhoI* cloning sites engineered between the cistrons encoding the viral movement protein and the large coat protein subunit. Inserts are cloned antisense in frame with the viral genes. We used fragments from seven contigs of a *P. pachyrhizi* transcriptome database

(http://bioinfo03.ibi.unicamp.br/phakopsora): de_novo_595, de_novo_939, de_novo_4668, de_novo_4692, de_novo_2740, de_novo_57 and de_novo_6 to generate seven HIGS constructs pBPMV-Thi, pBPMV-AGO, pBPMV-PPI, pBPMV-Gpα, pBPMV-Pv-SNARE, pBPMV-HSS and pBPMV-18S, respectively (Table S1). The BPMV vector without an insert (pBPMV empty vector) was used as a control.

3.2.5 BPMV and fungal inoculation

BPMV RNA1 and RNA2 DNA clones, as well as Soybean Mosaic Viruses (SMV) used to increase BPMV accumulation in host tissue (ZHANG et. al., 2009; ZHANG et. al., 2010; LIM et al., 2011), were used to bombard soybean susceptible leaves (Williams 82), healthy seedling with unifoliate leaves at 14 day after sowing, to generate inoculum for the HIGS experiments. BPMV-infected leaf tissue with virus symptoms was collected at 3 weeks after bombardment, lyophilized, and stored at -20 ºC. Plants of the same susceptible genotype were germinated in a greenhouse for two weeks, when were dusted with carborundum and the unifoliate leaves were rub inoculated with the lyophilized leaf tissue inoculum corresponding to the constructs pBPMV-Thi, pBPMV-AGO, pBPMV-PPI, pBPMV-Gpα, pBPMV-Pv-SNARE, pBPMV-HSS and pBPMV-18S, and pBPMV (empty vector). To each construct, three biological replicates, each composed of three plants from Williams 82, were infected. Twenty one days after BPMV infection, plants were inoculated with Brazilian *P. pachyrhizi* population (10^5 spores mL^{-1}). Infected leaves were collected at specific time points after fungal inoculation, according to the main induction times observed in the expression analyzes for the selected genes. Once harvested, fungus and plant

were immediately frozen in liquid nitrogen and stored at -80 ° C. Plants infected were maintained in greenhouse until 14 days after *P. pachyrhizi* inoculation to phenotypic evaluation.

3.2.6 RNA extraction and RT-qPCR analysis

Total RNA was extracted from 100mg of frozen leaf tissue samples and 30mg spores and germinated spores using the RNeasy Plant Mini Kit (Qiagen). RNA contamination with genomic DNA was eliminated by the treatment of 1ug RNA with RNase-free DNase (Invitrogen). The cDNA was synthesized using the First Strand Super Script III kit (Invitrogen), following the manufacturer's recommendations.

All RT-qPCR analyses were performed on the Real Time StepOnePlusTM equipment (Applied Biosystems) using SYBR green for the detection of double-strand PCR products. Primers were designed using Primer3Plus software based on the sequences of the selected *P. pachyrhizi* contigs (Table S1). All primers were first tested by standard PCR using DNA and soybean cDNA to ensure specific amplification for *P. pachyrhizi*. Three negative controls were used to ensure that only the cDNA of *P. pachyrhizi* was amplified. The efficiency of the primers was calculated based on the equation $[10^{(-1/slope)}] - 1$ (PFAFFL, 2001). Each PCR reaction was performed in triplicate and the specificity of the amplification products was validated by the analysis of the dissociation curve.

The expression levels were determined by equation $2^{-\Delta Ct}$, where 2 is the summation of the target and endogenous primers efficiency, and ΔCt = (target Ct – endogenous Ct). The fungus endogenous gene tubulin (Tub) was used as the

normalizing gene (MACIEL et al., 2010). The expression profile of 18S gene was not evaluated. To determine the silencing levels, relative quantification was conducted by $2^{-\Delta\Delta Ct}$ method (PFAFFL, 2001). The empty vector (pBPMV) was used for the calibration of the samples. For the expression profile analysis of the genes across the infection cycle, the delta Ct values were log2 transformed to construct a hierarchical grouping Heat map following a complete hierarchical clustering connection, using the Cluster3 and Tree View software (EISEN et al., 1999).

3.2.7 Phenotypic evaluation

Plants were evaluated phenotypically 14 days after fungal inoculation according to four distinct parameters: sporulation level, number of uredinia per lesion, number of open uredinia per lesion and disease severity. The phenotypic parameters sporulation level, number of uredinia per lesion and number of open uredinia per lesion were evaluated according to previous described by Yamanaka et al. (2010). The disease severity was determined by percentage of lesions using photographed leaves and the softwares Photoshop cs6 and ImageJ. Phenotypic data for all parameters were obtained from 18 soybeans leafs (nine different plants) infected with fungus for each construct, totaling at least 180 lesions evaluated per construct. To test the effect of each event in the development of the disease, contrasts between the effects of control pBPMV (empty vector) and each of the other constructs were tested considering the significance level (P>0,05%), identified by Tukey-Kramer test, described by Kramer (1956).

3.3 RESULTS

3.3.1 Selection of candidate genes

3.3.1.1 P. pachyrhizi genes involved in PTGS machinery

The *Puccinia* sequences of dicer, argonaut and a hypothetical RdRP protein were obtained from NCBI and analyzed for the presence of domains in Pfam software. Conserved domains were identified in these sequences, as Piwi domain (position 977-1,125) in argonaut sequence, DEAD/DEAH box helicase (DEAD) (position 22-161), Dicer dimerization (Dicer dimer) (position 791-883), and two Ribonuclease III domains (RNAse III) (position 1,208-1,315 and 1,396-1,513) in dicer sequence, and the RdRP domain (position 845-1,505) in hypothetical RdRP sequence. The function of Piwi domain is related to double stranded-RNA-guided hydrolysis to single-stranded RNA, which describes the main function of proteins related to the argonaut family. The domains present in the dicer sequence play a role in binding to dsRNA molecules (Dicer dimer), as well as helicase (DEAD), and the RNAse III domain, the most common domain found in dicer sequences, is a type of ribonuclease that cleaves the dsRNA at specific locations.

These *Puccinia* sequences were aligned against the *P. pachyrhizi* transcriptome to search for similar sequences. Only one transcript, similar to argonaut sequence (de_novo_939), was found using this strategy. However, we also carried out a search using the annotations obtained against the NCBI database. This second strategy made it possible to identify eight contigs with argonaut annotation, three contigs with dicer annotation, and four contigs with RdRP annotations. All the

P. pachyrhizi contigs identified similar to components of the silencing machinery are available in Table S2. Each transcript found was separately analyzed for the presence of specific domains, and from the total of 15 sequences identified, only eight actually presented the domains DEAD, ArgoN, Piwi and RdRP, and of these almost all domains are not completed.

The domains found in the *P. pachyrhizi* transcripts, most correspond to the domains also present in the *Puccinia* sequences, such as the DEAD, Piwi, and RdRP domains, but one domain was not identified in the *Puccinia* sequences, the N-terminal domain of argonaut (ArgoN). Although not present in the argonaut sequence of *Puccinia*, the ArgoN domain is present in most argonaut proteins, such as in the argonaut sequence of *N. crassa* (NCU04730). ArgoN domain is usually linked to the PAZ domain, another typical argonaut protein domain, and together appear to interact with the 3 'ends of single-stranded RNA, contributing to the specific incorporation of siRNAs and miRNAs into the RNAi pathway.

Among the contigs of *P. pachyrhizi* that did not present domains, most are composed of small sequences ranging from 232 to 610 base pairs. However, for the contigs that presented domains, the major sequences presented from 800 to 3,558 base pairs. The *P. pachyrhizi* contigs which we found with annotations for PTGS machinery genes were aligned with predicted protein sequences from 13 fungal and two oomycetes species, including other rust fungi and also a *P. pachyrhizi* haustorium transcriptome sequences previously described by Link et al (2014). For a total of 15 *P. pachyrhizi* sequences used in this alignment, only three sequences showed similarity with other sequences compared. The contig de_novo_30111, which present a partial RdRP domain, showed similarity with six sequences found in

the haustorium of *P. pachyrhizi*, while the contig de_novo_7, that also contains a partial RdRP domain, showed similarity with one sequence of *U. appendiculatus*. Finally, the contig de_novo_939 similar to *Puccinia* argonaut and presenting a Piwi domain, was similar to sequences of all species used, and was closer also to the sequence of *U. appendiculatus* (Figure 1A).

Thus the results obtained suggested that most *P. pachyrhizi* contigs found as putative components of the PTGS machinery can actually perform such functions based on the domains present in the sequences, although their expression during pathogen interaction with soybean presented low levels of FPKM, except for the observed for the contig de_novo_939 (Table S2). Based on the results, we selected the contig de_novo_939 (AGO) similar to Argonaut protein sequence to be tested by HIGS (Table 1).

3.3.1.2 P. pachyrhizi genes involved in survival and pathogenicity

Genes potentially involved in basal survival mechanisms and the pathogenicity of *P. pachyrhizi* were selected based on the literature and mining in our transcriptome dataset obtained from the lesion on infected soybean leaves. Sequences of five orthologous genes described for other fungi, and one gene already described in *P. pachyrhizi*, were used to find similar sequences in *P. pachyrhizi* transcriptome database. These genes comprise sequences related to different functions, such as thiamine biosynthesis, signaling metabolic pathways, protein folding, ribosomal protein and vesicle fusion (HAHN & MENDGEN, 1997; FUCHS *et al*., 2006; WEBB et al., 2006; BAILEY et al., 2010; VIEIRA et al., 2012). The orthologous sequences were also analyzed for the presence of domains in Pfam

software before to alignment against the *P. pachyrhizi* transcriptome database. Conserved domains were identified in five of six sequences: NMT1 domain (position 53-273) in thiamine biosynthesis (Thi) sequence; Pro isomerase domain (position 19-170) in peptidyl-prolyl isomerase (PPI) sequence; G-alpha domain (position 23-206) in G protein α subunit (Gpα) sequence; HSP70 domain (position 340-595) in small heat shock protein (HSS); and the PX domain (position 845-1,505) in soluble NSF attachment receptor protein (Pv-SNARE) sequence. No domains were found in the 18S ribosomal (18S) sequence. These domains found are typical of the selected genes, confirming their respective molecular functions.

The reference sequences were then aligned against the *P. pachyrhizi* transcriptome to search for similar sequences and were selected only transcripts with e-value cut off 1e-5. The searches identified one contig similar to Thi sequence, three contigs similar to PPI sequence, one contig similar to Gpα sequence, two contigs similar to 18S sequence, and seven contigs similar to HSS sequence. All the *P. pachyrhizi* contigs identified similar to orthologue sequences are available in Table S2. No transcripts were found similar to Pv-SNARE sequence, however only to this gene we also carried out a search using word searches based on the annotations obtained against the NCBI database. This second strategy made it possible to identify 18 contigs annotated as v-SNARE protein (data not shown), most identified as hypothetical proteins of *P. graminis* and *M. larici-populina*, and almost all presented very low FPKM values in transcriptome. To Pv-SNARE the selection was based on FPKM values, and the contig de_novo_2740 presenting higher value was selected. Based on phylogenetic analysis and expression levels based on FPKM values in the transcriptome, the contigs de_novo_595 (Thi), de_novo_4668 (PPI),

de_novo_4692 (Gpα), de_novo_6 (18S) and de_novo_57 (HSS) were selected for future analysis.

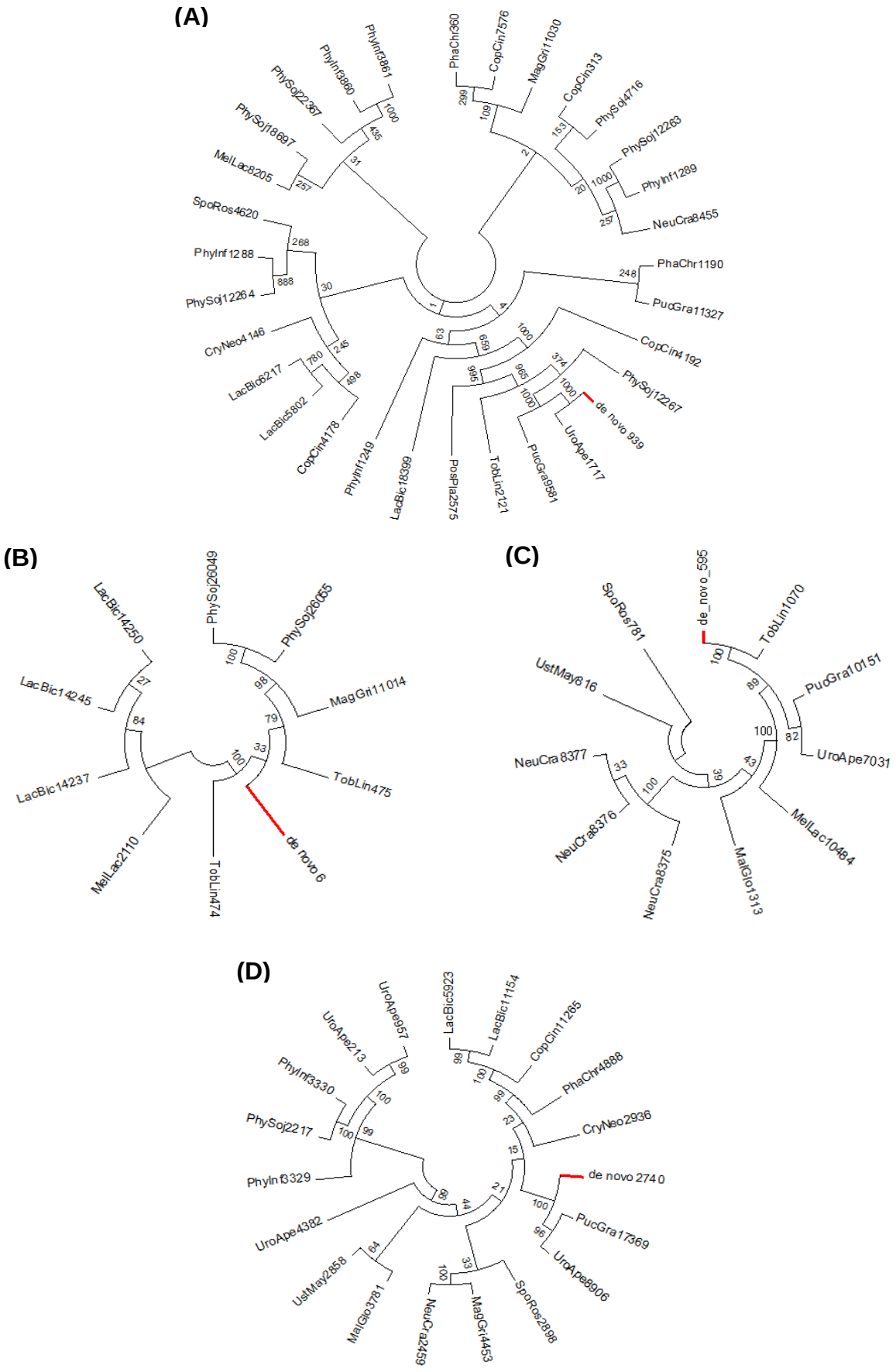
The six contigs selected as potentially involved in the basal survival and pathogenicity processes of *P. pachyrhizi* were carefully analyzed for the confirmation of their molecular functions by the presence of domains using the Pfam. The *P. pachyrhizi* transcripts selected based on alignment presented the same domains found in their respective orthologous sequences. The NMT1 domain (found in de_novo_595) is a family of domains that contains the NMT1 and THI5 proteins, which are homologous and are required for the biosynthesis of the hydroxymethylpyrimidine (HMP), a precursor of thiamine, essential for the functioning of many metabolic processes, for example, glucose metabolism. The Pro isomerase domain (found in de_novo_4668) is a typical domain found in peptidyl-prolyl cis/trans isomerase proteins, also known as cyclophilins, which act by accelerating protein folding by catalyzing the isomerization of imidic proline cis/trans peptide bonds in oligopeptides, and in some specific cases can perform functions similar to chaperone proteins.

The contig de_novo_4692 presented a G-alpha domain, a domain found in guanine nucleotide binding proteins (G proteins) that are membrane-associated. When receptors located on the cell surface bind to specific substances, they undergo a conformational change which in turn activates the bound G protein on the intracellular side of the membrane. Thus, in general, the G proteins end up regulating the intracellular concentrations of secondary messengers, and at the end of the process may result in a physiological response. The HSP70 domain (identified in de_novo_57) are characteristic of heat shock proteins, also known as chaperones,

which aid in the folding of many proteins, which involves repeated cycles of binding and release of the substrate. These proteins promote the folding of other proteins both under normal growth conditions and in response to stress conditions, preventing protein aggregation, remodeling folding pathways, and regulating activity of others proteins. The contig de_novo_2740 presented a domain different from the PX domain, found in the orthologous gene. The V-SNARE-C domain found in the Pp transcript is involved in anchoring in the vesicle membranes, being critical to the membrane fusion. Finally, the contig de_novo_6 identified as the most similar to the reference sequence predict to 18S ribossomal subunit showed no domain.

The comparison of the selected contigs sequences with predicted protein sequences from 15 other fungal and oomycetes species, revelead that four (de_novo_6, de_novo_595, de_novo_2740 and de_novo_57) of the six sequences of *P. pachyrhizi* showed similarity with the other sequences compared (Figure 1). Similar to observe to contig de_novo_939 (AGO), most of Pp sequences were also closer to the sequences found in *P. pachyrhizi* haustorium (LINK et al., 2014) and found in other rust species as *P. graminis, M. larici-populina* and *U. appendiculatus*.

Based on the analyzes performed for the search for genes of the PTGS machinery and genes involved in survival and pathogenicity mechanisms, seven *P. pachyrhizi* putative genes were selected for functional characterization using the HIGS strategy: a gene that encode argonaut protein (AGO); one that encodes 18S ribosomal subunit (18S); thiamine biosynthesis (Thi); peptidyl-prolyl isomerase (PPI); one that encodes a G protein alpha subunit (Gpα); a gene encoding a small heat shock protein (HSS); and another that encodes a vesicle fusion protein (Pv-SNARE) (Table 1).


(A)
PhaChr360
CopCin7576
MagGri11030
CopCin313
PhySoj4716
PhySoj12263
PhyInf1289
NeuCra8455
PhaChr1190
PucGra11327
CopCin4192
PhySoj12267
de novo 939
UroApe1717
PucGra9581
TobLin2121
PosPla2575
LacBic18399
PhyInf1249
CopCin4178
LacBic5802
LacBic6217
CryNeo4146
PhySoj12264
PhyInf1288
SpoRos4620
MelLac8205
PhySoj18697
PhySoj22367
PhyInf3860
PhyInf3861
(B)
LacBic14250
LacBic14245
LacBic14237
MelLac2110
TobLin474
de novo 6
TobLin475
MagGri11014
PhySoj26055
PhySoj26049
(C)
de_novo_595
TobLin1070
PucGra10151
UroApe7031
MelLac10484
MalGlo1313
NeuCra8375
NeuCra8376
NeuCra8377
UstMay6816
SpoRos781
(D)
UroApe957
UroApe213
PhyInf3330
PhySoj2217
PhyInf3329
UroApe4382
UstMay2858
MalGlo3781
NeuCra2459
MagGri4453
SpoRos2898
UroApe8906
PucGra17369
de novo 2740
CryNeo2936
PhaChr4888
CopCin11265
LacBic11154
LacBic5923

(E)

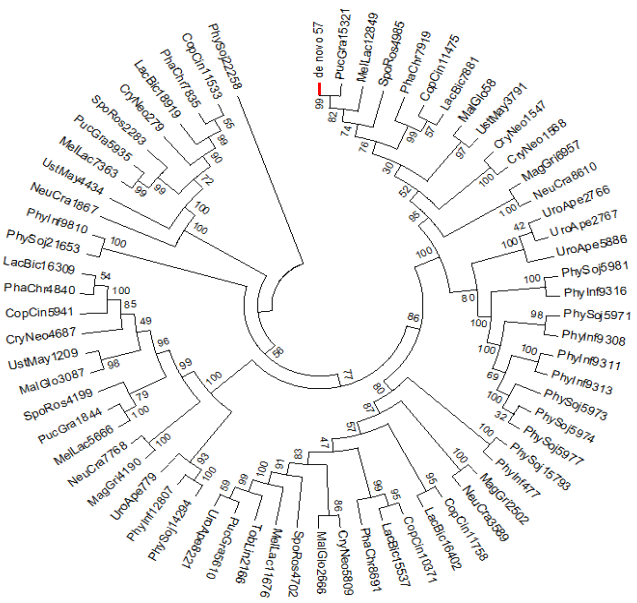


Figure 1. Phylogenetic analysis of transcripts predicted to encode genes involved in silencing, survival and pathogenesis of *Phakopsora pachyrhizi* compared with other fungi and oomycetes species. The analysis involved 31, 10, 11, 19, and 67 amino acid sequences obtained from the similarities analysis of the contig de_novo_939 (A), de_novo_6 (B) , de_novo_595 (C) , de_novo_2740 (D) , and de_novo_57 (E), respectively, with the predicted proteins from the genome and transcriptome sequence from 15 fungi and oomycetes species. The numbers near the branches indicate the percentages of replicates of the trees in which the sequences appear grouped in the bootstrap test (1,000 replicates). *P. pachyrhizi* sequences are described as “de_novo_” to identify the selected contig sequence and as “TobLin” to identify predicted Pp sequences of haustorium (Link et al., 2014). The sequences of the other species are identified by means of the following abbreviations: PucGra - *P. graminis*; UroApe - *U. appendiculatus*; MelLac - *M. larici-populina*; SpoRos - *S. roseus*; CryNeo - *C. neoformans*; CopCin - *C. cinerea*; LacBic – *L. bicolor*; MalGlo - *M. globosa*; UstMay - *U. maydis*; PosPla – *P. placenta*; MagGri - *M. grisea*; NeuCra – *N. crassa*; PhaChr: *P. chrysosporium*; PhyInf – *P. infestans*; PhySoj – *P. sojae*).

Table 1. Summary of the *Phakopsora pachyrhizi* selected genes to functional characterization by HIGS.

Gene	Contig ID	Putative gene functions	FPKM[1]		Orthologous genes[2]	Reference
			Susceptible	Resistant		
Thi	de_novo_595	Thiamine biosynthesis	1533.76	743.64	U81789.1	HAHN & MENDGEN, 1997
PPI	de_novo_4668	Peptidyl-prolyl cis/trans isomerase	1196.55	779.43	U81792.1	HAHN & MENDGEN, 1997
Gpα	de_novo_4692	Heterotrimeric G protein α subunit (signaling pathway)	7.95	3.97	FR851890.1	VIEIRA *et al*., 2012
AGO	de_novo_939	Argonaut protein	320.67	217.59	JN033210	CANTU *et al*., 2011
Pv-SNARE	de_novo_2740	Soluble NSF attachment receptor vesicle protein	181.33	175.74	AAF62178.1	FUCHS *et al*., 2006
HSS	de_novo_57	Small heat shock protein	383.09	107.03	U26597.1	WEBB *et al*., 2006
18S	de_novo_6	18S ribossomal protein	250.98	216.66	-	BAILEY et al., 2010

[1]FPKM values of *P. pachyrhizi* transcripts obtained from lesions from susceptible (BRS 231) and resistant (PI561356) genotypes leaves after 10 days of infection corresponding of the *P. pachyrhizi* transcriptome database, available in LGE database (http://bioinfo03.ibi.unicamp.br/phakopsora/).
[2]NCBI access number of orthologous genes from other fungi used to select *P. pachyrhizi* genes.

3.3.2 Expression profiling of *P. pachyrhizi* genes during infection time course

The transcriptional profiles of the candidate genes selected were evaluated in samples obtained from fresh spores and spores germinated, as well as during an infection time course in a susceptible soybean genotype (Williams 82) using two different isolates, LUB112 and LD5511, that present different virulence profile in the soybean differentials containing different *Rpp* background (DARBEN et al., 2013). The plants infected in a greenhouse showed TAN lesions typical of susceptible genotypes, as expected. The clusterization of expression profile levels in the heat map across the infection times detected a formation of two main clusters, one covering the initial time points of infection, from S to 24hpi, with a peak with in 12 and 24hpi, corresponding from the spores germination to the beginning of the host tissue colonization, and another group comprising the times from 48 to 192hpi, ranging from the formation of haustorium to the formation and liberation of new spores (Figure 2).

Additionally, the clusterization across the isolate types grouped the genes in four major groups (Figure 2). The first group comprises only the expression profile for both isolates to gene Pv-SNARE, which was induced from S up to 24hpi. The second group comprises Thi (both isolates) and PPI (only LD5511 isolated) genes, that were induced after 48hpi. The third group covers Gpα and HSS genes (both isolated), which did not show induction in the S and GS times and after were induced in specific time points, as 12hpi for example. The fourth and last group comprises AGO (both isolates) and PPI (only LUB112 isolated) genes, that were induced in almost all

time points evaluated. The fold change values and respective values of the standard deviations for each transcript are specified in the supplementary files (Table S3).

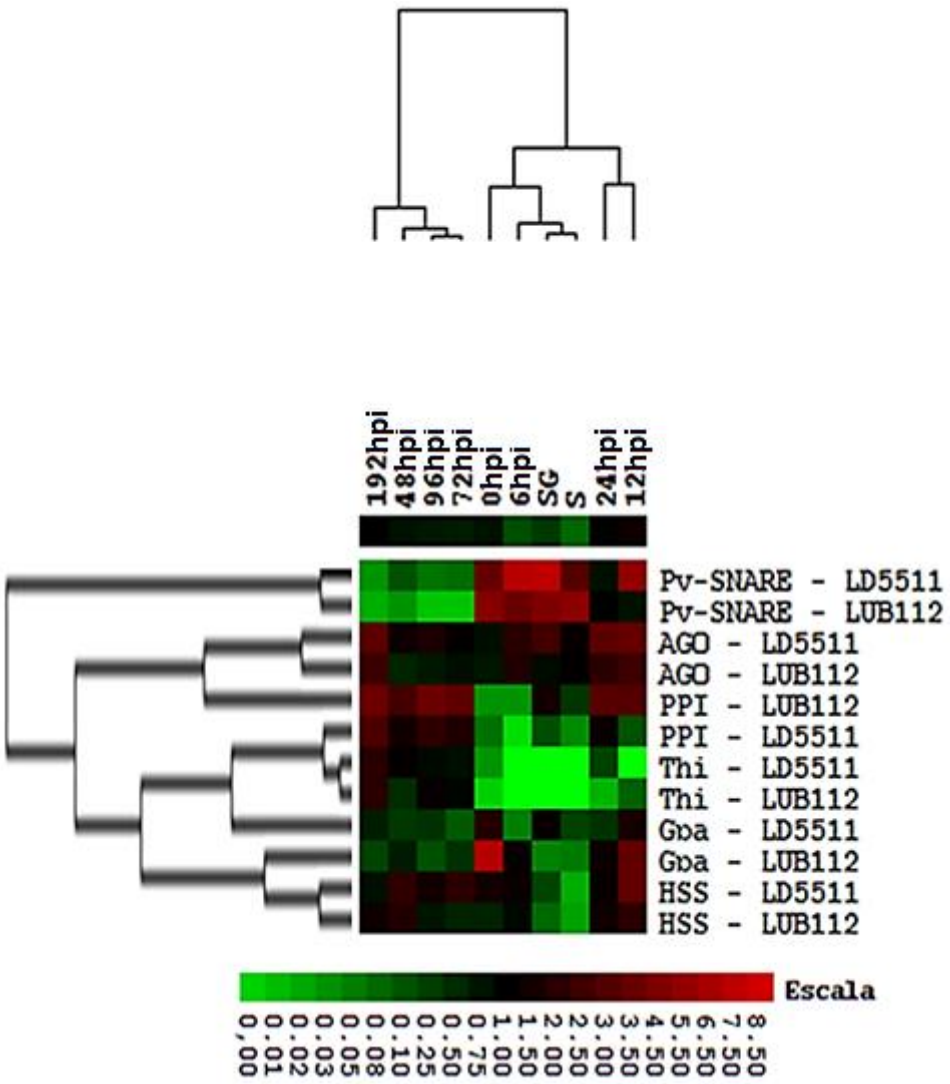


Figure 2. Heat map and hierarchical clustering of the expression profiles of six genes selected from *Phakopsora pachyrhizi* during the spore (S) and germinated spore (SG) stages, and during the main times of the soybean infection process (0, 6 , 12, 24, 48, 72, 96 and 192 hpi). Clustering was performed using Cluster3 software. Normalized fold change values were used for clustering. The color scale bar can be seen below the figure and covers the interval between fold change values.

Although gene clustering was observed based on the expression profile, there were generally no significant differences in the expression of the genes among the two isolates. The few differences observed among the isolates for the same gene are related to the infection times, where one isolate usually had induction in more time points compared with another isolate. This situation can be observed in the expression profile of the genes HSS and AGO, in which the genes were induced at more time points for the LD5511 isolate than for the LUB112 isolate, and also to the expression profile of the gene PPI, which are induced in more time points to LUB112 isolate.

3.3.3 BPMV-HIGS reduced transcript abundance of silenced *P. pachyrhizi* genes

The 3'-end coding sequences of seven *P. pachyrhizi* genes (Thi, PPI, AGO, Gpα, Pv-SNARE, HSS and 18S) were cloned in an antisense orientation in the BPMV RNA2 component to produce the silencing constructs. Soybean leaves with 14 days (susceptible genotype Willliams 82) were rub inoculated with the BPMV constructs carrying the respective fungal gene fragments in antisense conformation, or with the same BPMV empty vector as a control. After 21 days BPMV infection, when virus symptoms became apparent in third trifoliate leaves, plants were inoculated with Brazilian population of *P. pachyrhizi* urediniospores and observed for leaf rust disease symptoms on the following days.

The level of silencing was examined by RT-qPCR analysis. The leaf samples were collected in time points corresponding with the peak of the gene expression

according previously evaluated, in which it was possible to identify the induction in both isolates used. Thus, leaf infected samples with HIGS constructs to pBPMV-Thi, pBPMV-AGO, pBPMV-PPI, and pBPMV-18S were collected at 192hpi, pBPMV-Gpα was collected at 12 and 192hpi, pBPMV-Pv-SNARE was collected at 6 and 192hpi, pBPMV-HSS was collected at 48 and 192hpi, and finally the pBPMV, empty vector, was collected in all time points, at 6, 12, 48 and 192hpi. The RT-qPCR results showed significant reductions in transcript abundance of three out of seven corresponding *P. pachyrhizi* genes (Table 2). The greatest reduction in expression levels and consequently high level of silencing could be observed for the Gpα gene, reaching 67% of silencing at 12hpi. Pv-SNARE and 18S genes also showed a significant decrease of its transcripts, around 63% and 32%. Although also shown to reduced transcript levels, the silencing observed for the HSS and Thi genes were not significant in the t test at 0.05% probability in the time points evaluated.

Table 2. Relative quantification of *Phakopsora pachyrhizi* selected genes after BPMV-HIGS silencing.

Genes	***P. pachyrhizi* times infection[1]**			
	6hpi	**12hpi**	**48hpi**	**192hpi**
Pv-SNARE	0,370 (0,11)*	-	-	0,790 (0,26)*
Gpα	-	0,334 (0,39)*	-	0,890 (0,14)
HSS	-	-	0,600 (0,63)	0,701 (0,25)
Thi	-	-	-	0,752 (0,32)
PPI	-	-	-	1,025 (0,10)
AGO	-	-	-	0,938 (0,25)
PPA-18S	-	-	-	0,680 (0,21)*

Fold change values corresponding to the relative quantification are presented in the table followed by their respective standard deviation values, shown in parentheses.
*Significant values at P>0.05% (t test)
[1]*P. pachyrhizi* times infection after contact with soybean at 6, 12, 48, and 192 hours post infection.

3.3.4 Expression of target fungal gene fragments using BPMV-HIGS in soybean reduces *P. pachyrhizi* disease symptoms

To evaluate the effect of the silencing of *P. pachyrhizi* genes on the disease symptoms, and to investigate the role of these genes in the survival, development and pathogenicity of this fungus, silencing assays were conducted using BPMV-mediated HIGS and soybean plants infected with the *P. pachyrhizi* were phenotypically evaluated 14 days after rust infection. Four phenotypic parameters were evaluated comparing the silenced with the non-silenced plants, infected with empty virus vector (Figure 3). The sporulation level significantly reduced in plants silenced for the genes Pv-SNARE, 18S, Gpα, AGO and PPI (Figure 3A). The uredinia number per lesion presented significant reduction in silenced plants only for the genes 18S, Gpα e Pv-SNARE (Figure 3B), while the number of open uredinia per lesion shows significant reduction in almost all silenced plants, except those silenced for the gene AGO (Figure 3C). Disease severity was attenuated significantly in plants silenced for the genes Pv-SNARE, 18S, Gpα, Thi and PPI (Figure 3D). The values referring to the statistical analyzes performed for each phenotypic parameter are shown in Table S2.

Phenotypic parameters evaluated presented different results for the *P. pachyrhizi* silenced genes. The AGO and HSS silenced genes presented reduction of symptoms in only one of the four evaluated phenotypic parameters (sporulation level and number of open uredinia per lesion, respectively). The Thi silenced gene show reduction of symptoms in two of four evaluated phenotypic parameters (number of open uredinia per lesion and disease severity), and PPI silenced gene, in three of

four evaluated phenotypic parameters (sporulation level, number of open uredinia per lesion and disease severity). The remaining three silenced genes, Pv-SNARE, 18S and Gpα, showed significant reduction in all phenotypic parameters analyzed, as well as in the levels of transcripts detected by the RT-qPCR analysis. The soybean plants inoculated with the BPMV-HIGS constructs pBPMV-18S, pBPMV-Pv-SNARE and pBPMV-Gpα presented lower levels of fungus symptoms when compared to plants inoculated with the empty vector (pBPMV). The disease severity differences between plants inoculated with empty vector and these three constructions could be observed in the Figure 4. The silencing of the Pv-SNARE gene showed the greatest reduction of disease severity in relation to the control (pBPMV), followed by Gpα and 18S genes.

Additionally, an important aspect was observed in leaves infected with the constructs, including the empty vector. Most leaves that exhibited the symptoms of the virus did not exhibit the symptoms evenly distributed over the leaf, some leaves show the typical symptoms of the BPMV virus more concentrated in some portions of the leaf, such as from the middle to the tips or from the middle to the sides, for example.

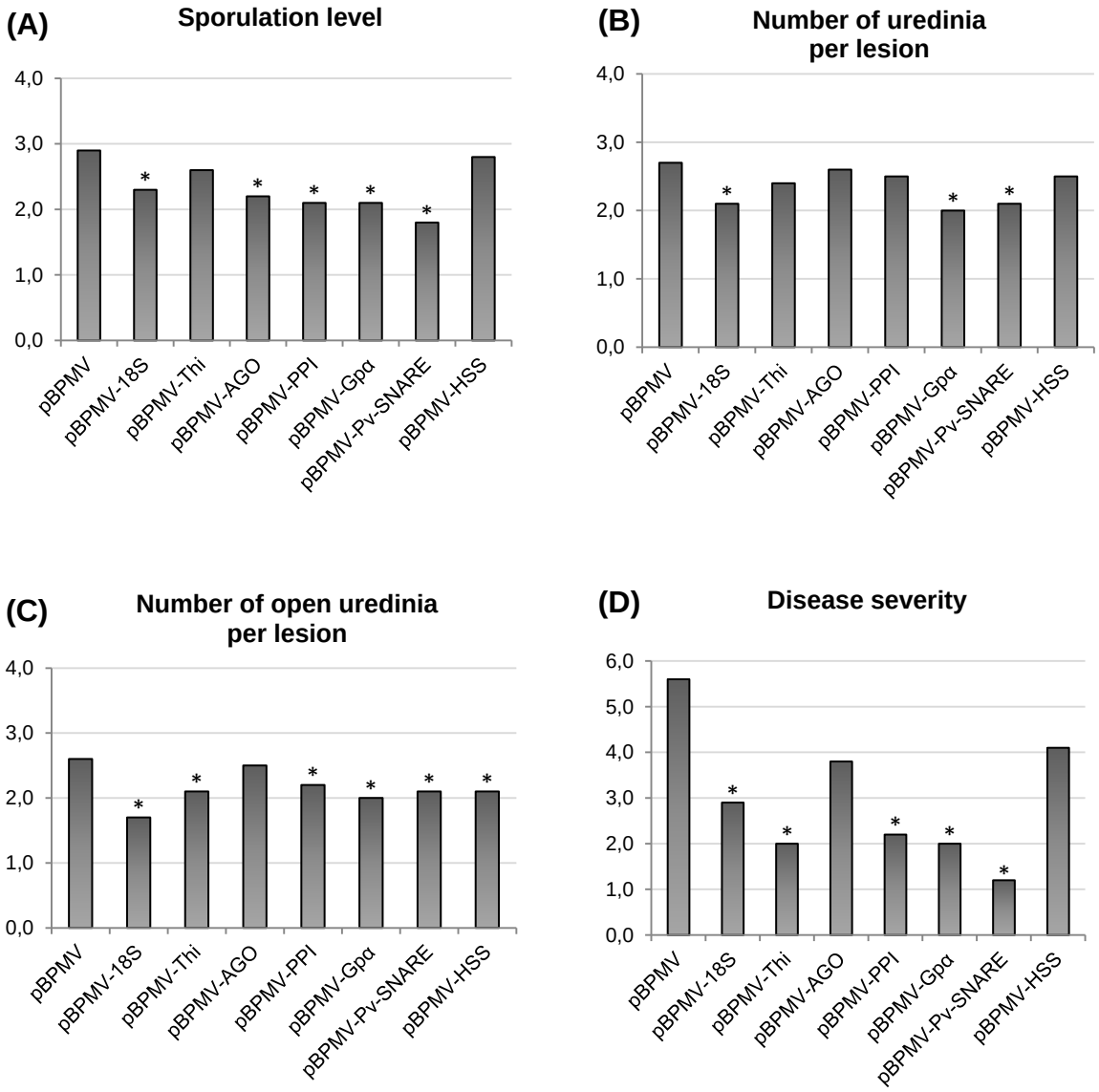


Figure 3. Phenotypic parameters evaluation 14 days after *Phakopsora pachyrhizi* inoculation: sporulation level (A); number of uredinia per lesion (B); number of open uredinia per lesion (C); and disease severity (D). Soybean plants inoculated with the HIGS-BPMV constructs pBPMV-Thi, pBPMV-AGO, pBPMV-PPI, pBPMV-18S, pBPMV-Gpα, pBPMV-Pv-SNARE, and pBPMV-HSS (containing the fragments of the selected *P. pachyrhizi* genes) were contrasted with plants inoculated with the empty vector (pBPMV). *Significant values at P>0.05% (Tukey-Kramer test).

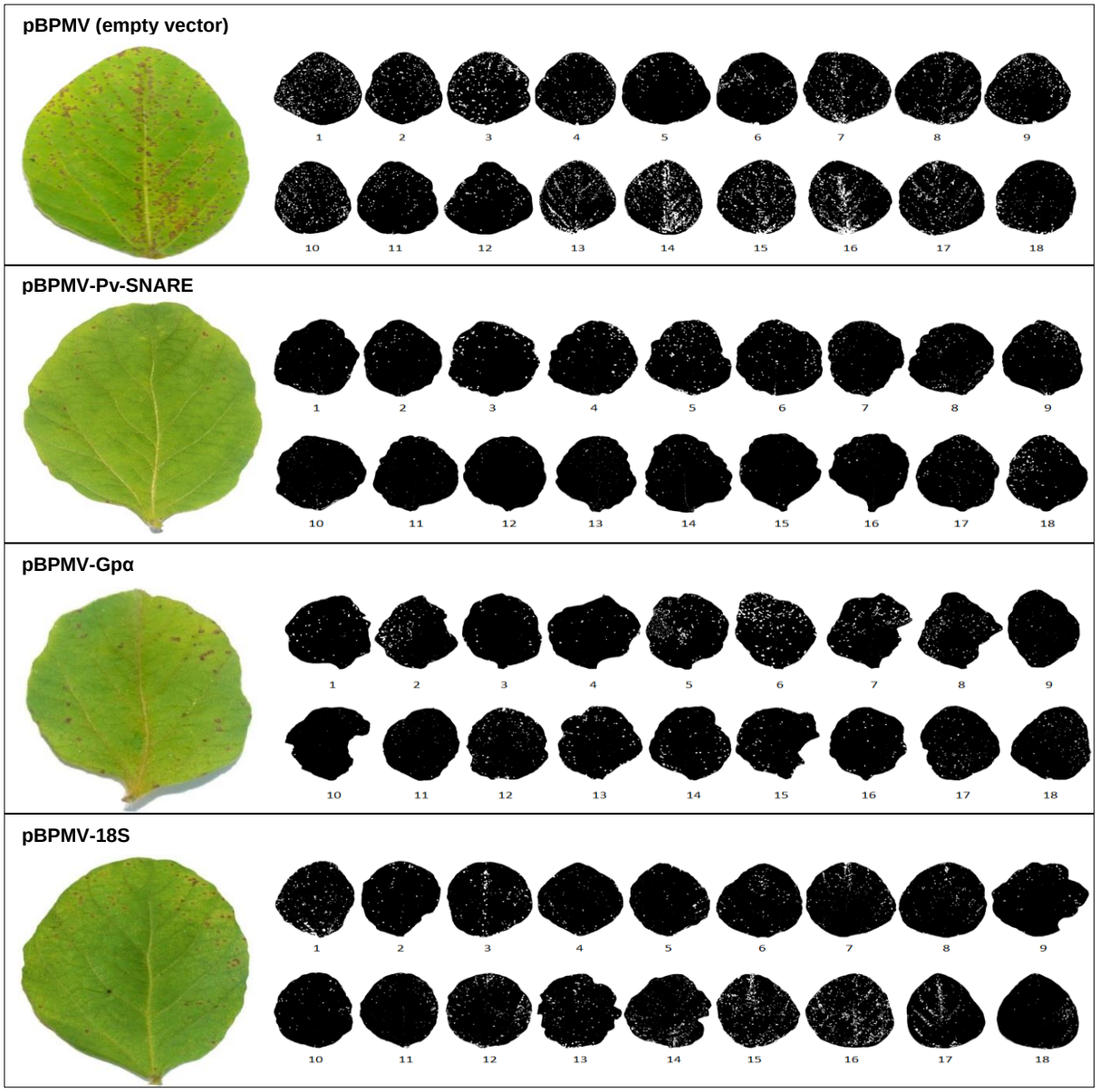


Figure 4. Effect of BPMV-HIGS in *Phakopsora pachyrhizi* genes on disease development in soybean. Rust disease severity phenotype in non-silenced (pBPMV empty vector) and silenced soybean leaves (pBPMV-Pv-SNARE, pBPMV-Gpα and pBPMV-18S). Plants inoculated with BPMV vectors containing segments of *P. pachyrhizi* transcripts as indicated, show disease suppression whereas control (pBPMV – empty vector) is heavily infected. Representative photos of each genotype, as well as the black and white contrast for the 18 phenotypically evaluated soybean leaves are presented.

3.4 DISCUSSION

Plant-mediated gene silencing has been a useful method to test the function of fungal genes. The HIGS tool has been recently used for functional characterization in fungi forming-haustoria, mainly focusing on the study of effector genes candidates (NOWARA et al., 2010; PLIEGO et al., 2013; YIN et al., 2015; COOPER & CAMPBELL, 2017). The study of non-effector genes that act in a determinant way in processes of development and pathogenicity are also important to a more comprehensive understanding of plant-pathogen interaction at molecular level. Furthermore, it also shows promise as a potential mechanism to protect plants from fungal disease. The silencing of non-effector genes in *P. triticina*, such as the endogenous MAP kinase (PtMAPK1), cyclophilin (PtCYC1) and calcineurin B (PtCNB) genes, mediated by *Agrobacterium tumefaciens*, resulted in a drastic reduction in the transcripts levels as well as in the pathogenicity of the fungus, affecting its normal development and even the sporulation (PANWAR et al., 2013b).

In this study, a transcriptome previously obtained by a combination of LCM of *P. pachyrhizi* lesions in soybean plants and deep sequencing (CARVALHO et al., 2016) was used to search for potentially target genes to be tested in HIGS experiments in the soybean and *P. pachyrhizi* pathosystem. Sequences from genes previously reported as essential in the biological process of silencing, thiamine biosynthesis, protein biosynthesis and folding, signaling metabolic pathways and vesicular fusion where selected. Rust orthologues sequences were used to mining Pp transcripts from the transcriptome dataset, and *in silico* analysis filter the best targets by confirming the presence of expected domains in six of the seven Pp

sequences selected (Table 1). The phylogenetic analysis of the Pp transcripts also demonstrated a specific cluster formation for five candidate genes (AGO, 18S, Thi, Pv-SNARE and HSS) with a more related Order, the Uredinales (Pucciniales), and also with other Pp transcripts obtained from haustorium (Figure 1) (Link et al, 2014).

The search for *P. pachyrhizi* genes, mainly based on the similarity between orthologous sequences, can be a complex work since the genome of the fungus is not yet available and the studies of its transcriptome, as well as the one carried out by Carvalho et al. (2016), generate transcripts that normally do not correspond to the complete coding sequence. In addition, many of the sequences obtained may be considerably shorter. This search may become more difficult when the genes of interest exhibit great sequence variation even among phylogenetically close species, as was observed for the dicer sequences, which presented divergent domain architecture among Basidiomycete fungi (HU et al., 2013).

The results obtained in our study by silencing the selected *P. pachyrhizi* genes revealed the important role of these genes in the development and pathogenicity of this fungus during soybean infection. Three of the seven silenced genes showed significant reductions both in transcript levels and in all phenotypic parameters evaluated. SNARE proteins, which here were the most successfully silenced targets (Figure 4), act as key proteins in the fusion of membranes, presenting four SNARE domains, anchored in opposite membranes, that form a tetrameric complex that approaches the membranes and thus facilitates it’s fusion (JAHN e SCHELLER, 2006). In addition to being essential in the initial stages of the infection process acting on endocytosis mechanisms, as in *U. maydis* (FUCHS et al., 2006), SNARE proteins are also involved in cellular secretion processes. In *M.*

oryzae, SNARE proteins are involved in regulating the transport of effector proteins secreted during the initial phase of the infection process, being essential for the normal physiology and pathogenicity of the fungus (QI et al., 2016).

In our analysis of gene expression profile, this gene was specifically induced from urediniospores germination until the penetration of the host tissue and beginning of the haustorium formation (Figure 2). Additionally, this sequence was also identified in haustorium transcriptome of *P. pachyrhizi* (LINK et al, 2014), as well as in the spore and germinating spores (STONE et al., 2012), corroborating with the profile observed here and its role in pathogen colonization and establishment in the host. The impact of the gene silencing on the Pp pathogenicity on soybean was evident for all parameters evaluated and could be related with its acting since the beginning of the infection process and its consequently, with silencing since 6 hours after rust inoculation in virus infected plants.

Signaling of metabolic pathways is extremely essential for host recognition and the integration of several essential stimuli to the establishment of biotrophy. In gene expression analysis the gene encoding the G protein alpha subunit (Gpα), presented induction during the first stages of the infection process, from the contact of the fungus with the plant, to the appressorium formation. Similar results were observed in other rust fungi, such as *M. larici-populina* (DUPLESSIS et al., 2011b) and *Hemileia vastatrix* (VIEIRA et al., 2012), in which this protein plays an important role in the regulation of crucial signaling pathways for the formation of appressorium. Specifically, in the coffee rust *H. vastatrix*, the Gpα gene has been described as involved in fungus penetration into the host tissue and haustorium formation (VIEIRA et al., 2012). In phytopathogenic fungi, mutations in the Gpα gene resulted in a

decrease in virulence and affected the development of appressorium (BOLKER, 1998; DEISING et al., 2000; LI et al., 2007), confirming the essential role of this gene in the infection process. In addition, this Pp Gpα sequence was also identified in haustorium transcriptome of *P. pachyrhizi* (LINK et al, 2014). The silencing of this gene in *P. pachyrhizi* also resulted in a decrease in pathogenicity, reducing the level of transcripts and the symptoms of the disease in infected soybean plants (Table 2 and Figures 3 and 4).

The 18S ribosomal protein, component of the smallest 40S eukaryotic ribosomal subunit, generally presents a low evolutionary rate being widely used in phylogenetic analyzes, including Uredinales fungi species (AIME, 2006). Due to its essential role in the composition of the ribosome and thus in the translation process, this gene was silenced in *P. pachyrhizi* and as well as the Pv-SNARE and Gpα genes, showed a significant reduction in transcript levels and disease symptoms evaluated (Table 2 and Figures 3 and 4) . In *B. graminis*, disrupting the 40S ribosomal protein resulted in a significant reduction of fungal development (NOWARA et al., 2010). The silencing of the 18S gene was previously reported successfully in *P. pachyrhizi* through the external and direct application of dsRNA molecules (BAILEY et al., 2010), but not by the HIGS approach.

The other silenced genes, although did not present significant reduction in transcript levels in the time points evaluated, they also showed reduction in some phenotypic parameters evaluated, revealing that could also play an important role in the development of *P. pachyrhizi*, such as the PPI and Thi genes. The enzymes peptidyl prolyl cis-trans isomerases (PPIs) constitute a superfamily, and can play the most different molecular roles, acting from folding and reactivation of proteins, to

large complexes of chaperones and transport of ions through the membrane (KROMINA et al., 2008). They are classified into subfamilies according to their inactivation by immunosuppressant, of which stands out mainly the cyclophilin binding proteins or cyclosporin A. The quantity and quality of PPI enzymes in the available genomes of pathogenic fungi is very diverse (PEMBERTON, 2006). In *M. grisea*, the *CYP1* gene encodes a cyclophilin, strains with mutation in this gene showed less virulence and the development of processes associated with pathogenicity was slower, including the development of appressorium and the penetration in vegetal tissues (VIAUD et al., 2002). Stone et al. (2012) also identified the presence of cyclophilins in the *P. pachyrhizi* appressorium transcriptome. The expression profile observed for the PPI gene revealed expression differences between the isolates, but both isolates showed induction after haustorium formation, with a gradual increase in expression levels until sporulation (Figure 2), similar to results observed by Hahn & Mendgen (1997).

The acquisition of nutrients from the host is an essential process for the lifestyle of obligate biotrophic organisms, however for some metabolic processes *P. pachyrhizi* is apparently able to synthesize its own nutrients, as in the case of thiamine, which probably may constitute a factor growth limit since it is not available in the host (HAHN e MENDGEN, 1997). Thiamine, also called vitamin B1, is essential in rust fungi, as a cofactor of several enzymes involved in carbohydrate metabolism, as reported in the haustorium of *U. fabae* (SOHN et al., 2000), and for spores germination, haustorium maturation and until the end of the infection process, presenting high levels of expression in these stages of development of the disease (HAHN e MENDGEN, 1997; DUPLESSIS et al., 2011; FERNANDEZ et al., 2012;

TREMBLAY et al., 2013; GARNICA et al., 2013). The results obtained in our gene expression analysis revealed a gradual induction of the Thi gene after the haustorium formation, with a higher peak of expression during the later stages of the infection process, during the formation of new urediniospores, corroborating the results observed previously in other rust studies. The reduction of the disease symptoms as number of open uredinia per lesion and disease severity, indicate that this gene can actually play an important role during the infection process in *P. pachyrhizi*.

The HSS and AGO genes presented lowers levels of silencing in the time points evaluated and a consequently low effect in the pathogenicity. In the plants infected with the virus followed by the rust, only one of the four evaluated phenotypic parameters was significantly reduced. HSS proteins, which act as molecular chaperones due to their positive regulation in response to thermal stress, can also act in the stabilization of biological substrates, in the assembly of macromolecules, in the degradation of polypeptides, as well as in the regulation of transcription, splicing and translation (BUKAU et al., 2006). A heat shock protein was silenced successfully in *B. graminis* (NOWARA et al., 2010) and in *P. pachyrhizi* these proteins were identified during urediniospores germination (LUSTER et al., 2010) and in secretome (CARVALHO et al., 2016). The gene expression analysis of the HSS gene revealed that in a more virulent isolate, the induction of this gene occurs during appressorium formation and penetration into the plant tissue, and remains active during almost the whole process of infection.

In the expression analysis, the AGO gene was induced almost throughout the infection process (mainly for the most virulent isolate), suggesting that the mechanisms of RNAi act continuously on the gene regulation of the pathogen, with

more expressive peaks during the penetration of host tissue and during sporulation. Argonaut is a main RISC (RNA induced silence complex) catalytic subunit, that cleavage short dsRNA in single-stranded RNA (ssRNA) and are ssRNA-guide to mRNA target for its degradation in RNAi pathways (HUTVAGNER & SIMARD, 2008). In filamentous fungi, RNAi pathways were more extensively studied in *N. crassa*, where there are several different pathways (LI et al., 2010). Gene silencing by RNAi was identified in many species of Ascomycetes, but in Basidiomycetes only some species have components related to this silencing machinery, including argonaut (NUNES et al., 2011). In *P. pachyrhizi* three putative argonaut-like sequences were also identified by Link et al. (2014), but so far nothing has been described about other components and even if this machinery is functional in this fungus. Our results show the presence of sequences similar to the PTGS machinery genes in the *P. pachyrhizi* transcriptome and the successful silencing for some of the tested genes, suggests that besides being present in the fungus, this machinery may be active.

According to Panwar et al. (2013b), the acquisition of siRNA molecules (produced in plant cells) by rust fungi may occur during the haustorium stage through the release of host cell exosomes into the paramural space that functions as a communication link between the plant and the pathogen. In this sense, gene silencing would occur preferentially for genes whose major expression occurs in the formation phase of the haustorium (YIN, et al., 2011). However, our results showed that the highest silencing rates were observed for the Pv-SNARE and Gpα genes in the initial time points, at 6 and 12 hours after rust inoculation in virus infected leaves, respectively. This result demonstrated that an alternative uptake of silencing signaling, possible by the siRNA uptake from the plant (COOPER & CAMPBELL,

2017), might be occurring before the haustorium formation, since there is a contact between the pathogen and the host. A possibility of a pre-haustorial traffic of molecules, for example during germ tube formation, where multivisicular bodies and exosomes are observed, can occur a few hours post infection (HUCKELHOVEN et al., 2000).

Differences observed in expression profile between two isolates used in the gene expression analyzes were also reported by Vieira et al. (2012) in the transcriptional profile of coffee rust genes (*H. vastatrix*) when compared among distinct isolates. Despite the differences found in the expression profile, all the genes analyzed were induced at some point in both isolates, indicating that these genes may play important roles during the development of the fungus in the infection process. It is important to note that the silencing was conducted from contigs obtained from the *P. pachyrhizi* transcriptome previous performed. Since the genome of this fungus has not yet been sequenced, limited access to complete gene sequences makes it difficult to seal the gene portions to be silenced, which may interfere with the final silencing result.

If the gene plays an important role in the survival of the fungus or to establish the pathogenicity in the host plant, it is expected that after the silencing of this gene a drastic reduction in the development of the disease or even death of the pathogen will occur. However, the silencing of target mRNAs depends primarily on the viral plasmids expression in the host cells. Variability in the virus expression is expected and may be generate an irregular pattern of symptoms by the plant leaf (NOWARA et al., 2010), which may also interfere in complete silencing. Viral expression variability was also been observed in other rust fungi studies of post transcriptional gene

silencing, as in *P. graminis* (YIN et al., 2015) and *U. appendiculatus* (COOPER & CAMPBELL 2017). In addition, Himber et al. (2003) reported that the silencing signal was able to spread for 10 to 15 cells, independent of the presence of the homologous transcripts, which can also contribute to the limitation of silencing in certain regions of the plant as in leaf sections. Thus, we cannot state that genes that did not present a significant reduction in transcript levels and/or in phenotypic parameters did not play an important role in the survival and infection processes of *P. pachyrhizi*, we just did not identify evidence of these roles in our results.

For those genes that it is possible that the silencing was not sufficient to reduce the encoded proteins to level that would interfere with fungal development. It is also possible that genes are sufficiently expressed in adjacent un-silenced cells and their expression is sufficient to supply the fungal tissues, and even if occurs functional redundancy, since the genes might be members of genes families, as proposed by Yin et al. (2015). Additionally, other parameters in the methodology, such as number of days between the virus infection and fungal inoculation might have important effects in the silencing and should be tested to optimize the results.

Here, we showed for the first time that the HIGS approach can be used successfully for the functional characterization of *P. pachyrhizi* genes. The reduction of the pathogenicity of this fungus after the silencing confirms the acting of the selected genes during fungus development and the soybean infection. Our results contribute with the knowledge already available regarding the molecular interactions between *P. pachyrhizi* and soybean, increasing the information for the generation of new strategies to combat ASR.

3.5 REFERENCES

AIME, M.C. Toward resolving family-level relationships in rust fungi (Uredinales). **Mycoscience**, v. 47, p. 112–122, 2006.

ANDRADE, C.; TINOCO, M.; RIETH, A.; MAIA, F.; ARAGÃO, F. Host-induced gene silencing in the necrotrophic fungal pathogen *Sclerotinia sclerotiorum*. **Plant Pathology**, v. 65, p. 626–632, 2015.

BAILEY, A.M & NIBLETT, C. BIOASSAY FOR GENE SILENCING CONSTRUCTS. Patent Application Publication. United States. Pub. No.: US 2010/0257634 A1, Pub. Date: Oct. 7, 2010.

BAULCOMBE, D.C. VIGS, HIGS and FIGS: small RNA silencing in the interactions of viruses or filamentous organisms with their plant hosts. **Current Opinion in Plant Biology**, v. 26, p. 141-146, 2015.

BOLKER, M. Sex and crime: heterotrimeric G proteins in fungal mating and pathogenesis. **Fungal Genetics and Biology**, v. 25, p. 143–156, 1998.

BUKAU, B.; WEISSMAN, J.; HORWICH, A. Molecular chaperones and protein quality control. **Cell**, v. 125, p. 443–451, 2006.

CANTU, D.; GOVINDARAJULU, M.; KOZIK, A.; et al. Next Generation Sequencing Provides Rapid Access to the Genome of *Puccinia striiformis* f. sp. *tritici*, the Causal Agent of Wheat Stripe Rust. **Plos One**, v. 6, e24230, 2011.

CARVALHO, M.C.C.G.; NASCIMENTO, L.C.; DARBEN, L.M.; et al. Prediction of the *in planta P. pachyrhizi* secretome and potential effector families. **Molecular Plant Pathology**, 2016.

CHEN, W.; KASTNER, C.; NOWARA, D.; OLIVEIRA-GARCIA, E.; RUTTEN, T.; ZHAO, Y.; DEISING, H.B.; KUMLEHN, J.; SCHWEIZER, P. Host-induced silencing of Fusarium culmorum genes protects wheat from infection. **Journal of Experimental Botany**, v. 67, p. 4979–4991, 2016.

CHENG, W.; SONG, X.S.; LI, H.P.; CAO, L.H.; SUN, K.; QIU, X.L.; XU, Y.B.; YANG, P.; HUANG, T.; ZHANG, J.B. Host-induced gene silencing of an essential chitin synthase gene confers durable resistance to *Fusarium* head blight and seedling blight in wheat. **Plant Biotechnology Journal**, v. 13, p. 1335–1345, 2015.

COOPER, B & CAMPBELL, K.B. Protection against Common Bean Rust Conferred by a Gene Silencing Method. **Phytopathology**, p. 1-24, 2017.

DARBEN, L. **Obtenção e caracterização molecular e fenotípica de isolados monouredíniais de Phakopsora pachyrhizi coletados em diferentes regiões do**

Brasil. Maringá. Universidade Estadual de Maringá-PR, 2013, p. 81. Tese (Doutorado em Genética e Melhoramento de Plantas).

DEISING, H.B.; WERNER, S.; WERNITZ, M. The role of fungal appressoria in plant infection. **Microbes and Infection**, v. 2, p. 1631–1641, 2000.

DUPLESSIS, S.; CUOMO, C.A.; LIN, Y.C.; AERTS, A.; TISSERANT, E.; et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 22, p. 9166–9171, 2011.

DUPLESSIS, S., HACQUARD, S., DELARUELLE, C., TISSERANT, E., FREY, P., MARTIN, F., KOHLER, A. *Melampsora larici-populina* transcript profiling during germination and time course infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. **Molecular Plant-Microbe Interactions**, v. 24, p. 808–818, 2011b.

EISEN, M.B.; SPELLMAN, P.T.; BROWN, P.O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, p. 14863-14868, 1999.

FERNANDEZ, D.; TISSERANT, E.; TALHINHAS, P., et al. 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-

expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction. **Molecular Plant Pathology**, v. 13, n. 1, p. 17–37, 2012.

FUCHS, U.; HAUSE, G.; SCHUCHARDT, I.; STEINBERGA, G. Endocytosis Is Essential for Pathogenic Development in the Corn Smut Fungus *Ustilago maydis*. **The Plant Cell**, v. 18, p. 2066–2081, 2006.

GARNICA, D.P.; UPADHYAYA, N.M.; DODDS, P.N.; RATHJEN, J.P. Strategies for Wheat Stripe Rust Pathogenicity Identified by Transcriptome Sequencing. **Plos One**, v. 8, e67150, 2013.

GHAG, S.B.; SHEKHAWAT, U.K.; GANAPATHI, T.R. Host-induced posttranscriptional hairpin RNA-mediated gene silencing of vital fungal genes confers efficient resistance against *Fusarium* wilt in banana. **Plant Biotechnology Journal**, v. 12, p. 541–553, 2014.

HAHN, M. E MENDGEN, K. Characterization of In Planta–Induced Rust Genes Isolated from a Haustorium-Specific cDNA Library. **Molecular Plant-Microbe Interactions : MPMI**, v. 10, n. 4, p. 427–437, 1997.

HIMBER, C.; DUNOYER, P.; MOISSIARD, G.; RITZENTHALER, C.; VOINNET, O. Transitivity-dependent and –independent cell-to-cell movement of RNA silencing. **EMBO J**, v. 22, p. 4523-4533, 2003.

HU, Y.; STENLID, J.; ELFSTRAND, M.; OLSON, Å. Evolution of RNA interference proteins dicer and argonaute in Basidiomycota. **Mycologia**, v. 105, n. 6, p. 1489-1498, 2013.

HU, Z.; PAREKH, U.; MARUTA, N.; TRUSOV, Y.; BOTELLA, J.R. Down-regulation of *Fusarium oxysporum* endogenous genes by host-delivered RNA interference enhances disease resistance. **Frontiers in Chemistry**, v. 3, p. 1, 2015.

HÜCKELHOVEN, R.; TRUJILLO, M.; KOGEL, K.H. Mutations in *Ror1* and *Ror2* genes cause modification of hydrogen peroxide accumulation in *mlo*-barley under attack from the powdery mildew fungus. **Molecular Plant Pathology**, v. 1, p. 287-292, 2000.

HUTVAGNER, G. & SIMARD, M.J. Argonaute proteins: key players in RNA silencing. **Nature**, v. 9, p. 22-32, 2008.

JAHN, R.; SCHELLER, R.H. SNAREs – engines for membrane fusion. **Nature Reviews Molecular Cell Biology**, v. 7, p. 631–643, 2006.

KOCH, A.; KUMAR, N.; WEBER, L.; KELLER, H.; IMANI, J.; KOGEL, K.H. Host induced gene silencing of cytochrome P450 lanosterol C14a-demethylaseencoding genes confers strong resistance to *Fusarium* species. **Proceedings of the National Academy of Science**, v. 110, p. 19324–19329, 2013.

KRAMER, C.Y. Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications, **Biometrics**, v. 12, p. 309-310, 1956.

KROMINA, K.A.; IGNATOV, A.N.; ABDEEVA, I.A. Role of Peptidyl-Prolyl-*cis*/*trans*-Isomerases in Pathologic Processes. **Biologicheskie Membrany**, v. 25, n. 4, p. 243–251, 2008.

LI, L.; CHANG, S.S.; LIU, Y. RNA interference pathways in filamentous fungi. **Cellular and Molecular Life Science**, v. 67, p. 3849-3863, 2010.

LI, L.; DING, S.-L.; SHARON, A.; ORBACH, M.; XU, J.-R. Mirl is highly upregulated and localized to nuclei during infectious hyphal growth in the rice blast fungus. **Molecular Plant-Microbe Interactions**, v. 20, p. 448–458, 2007.

LIM, H.S.; JANG, C.Y.; BAE, H.; KIM, J.K.; LEE, C.H.; HONG, J.S.; JU, H.J.; KIM, H.G.; DOMIER, L.L. *Soybean mosaic virus* Infection and Helper Component-protease Enhance Accumulation of *Bean pod mottle virus*-Specific siRNAs. **Plant Pathology Journal**, v. 27, n. 4, p. 315-323, 2011.

LINK, T.; LANG, P.; SCHEFFLER, B.E.; DUKE, M.V.; et al. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. **Molecular Plant Pathology**, v. 15, n. 4, p. 379–393, 2014.

LUSTER, D.G.; MCMAHON, M.B.; CARTER, M.L.; FORTIS, L.L.; NUNEZ, A. Proteomic analysis of germinating urediniospores of *Phakopsora pachyrhizi*, causal agent of Asian soybean rust. **Proteomics**, v. 10, p. 3549–3557, 2010.

MACIEL, T.E.F.; FREIRE, M.C.M.; ALMEIDA, A.M.R. DE; OLIVEIRA, L.O. DE. Molecular Characterization of Beta-Tubulin from *Phakopsora pachyrhizi*, the Causal Agent of Asian Soybean Rust. **Molecular Plant Pathology**, v. 11, n. 2, p. 169–177, 2010.

NOWARA, D.; GAY, A.; LACOMME, C.; SHAW, J.; RIDOUT, C.; DOUCHKOV, D.; HENSEL, G.; KUMLEHN, J.; SCHWEIZERA, P. HIGS: Host-Induced Gene Silencing in the Obligate Biotrophic Fungal Pathogen *Blumeria graminis*. The Plant Cell, v. 22, p. 3130-3141, 2010.

NUNES, C.C.; SAILSBERY, J.K.; DEAN, R.A. Characterization and application of small RNAs and RNA silencing mechanisms in fungi. **Fungal Biology Reviews**, v. 25, p. 172-180, 2011.

PANWAR, V.; MCCALLUM, B.; BAKKEREN, G. Host-induced gene silencing of wheat leaf rust fungus *Puccinia triticina* pathogenicity genes mediated by the *Barley stripe mosaic virus*. **Plant Molecular Biology**, v. 81, p. 595–608, 2013.

PANWAR, V.; MCCALLUM, B.; BAKKEREN, G. Endogenous silencing of Puccinia triticina pathogenicity genes through in planta-expressed sequences leads to the

suppression of rust diseases on wheat. **The Plant Journal**, v. 73, p. 521–532, 2013b.

PEMBERTON, T.J. Identification and Comparative Analy- sis of Sixteen Fungal Peptidyl-Prolyl cis/trans Isomerase Repertoires. **BMC Genomics**, n. 7, p. 244, 2006.

PFAFFL, M.W. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic Acids Research**, v. 29, n. 9, p. 2002-2007, 2001.

PLIEGO, C.; NOWARA, D.; BONCIANI, G.; GHEORGHE, D.M.; XU, R.; SURANA, P.; WHIGHAM, E.; NETTLETON, D.; BOGDANOVE, A.J.; WISE, R.P.; SCHWEIZER, P.; BINDSCHEDLER, L.V.; SPANU, P.D. Host-Induced Gene Silencing in Barley Powdery Mildew Reveals a Class of Ribonuclease-Like Effectors. **Molecular Plant-Microbe Interactions**, v. 26, n. 6, p. 633–642, 2013.

POSADA-BUITRAGO, M.L. & FREDERICK, R.D. Expressed sequence tag analysis of the soybean rust pathogen *Phakopsora pachyrhizi*. **Fungal Genetics and Biology**, v. 42, p. 949–962, 2005.

QI, Z.; LIU, M.; DONG, Y.; et al. The syntaxin protein (MoSyn8) mediates intracellular trafficking to regulate conidiogenesis and pathogenicity of rice blast fungus. **New Phytologist**, v. 209, p. 1655–1667, 2016.

SCHERM, H.; CHRISTIANO, R.S.; ESKER, P.D.; PONTE, E.M.; GODOY, C.V. Quantitative review of fungicide efficacy trials for managing rust in Brazil. **Crop Protection**, v. 28, p. 774-782, 2009.

SOHN, J.; VOEGELE, R.T.; MENDGEN, K.; HAHN. M. High Level Activation of Vitamin B1 Biosynthesis Genes in Haustoria of the Rust Fungus *Uromyces fabae*. **Molecular Plant-Microbe Interactions**, v. 13, n. 6, p. 629–636, 2000.

SONG, Y. & THOMMA, B.P.H.J. Host-induced gene silencing compromises Verticillium wilt in tomato and *Arabidopsis*. **Molecular Plant Pathology**, v. 1, p. 1-13, 2016.

STONE, C.L.; MCMAHON, M.B.; FORTIS, L.L.; NUÑEZ, A.; SMYTHERS, G.W.; LUSTER, D.G.; FREDERICK, R.D. Gene expression and proteomic analysis of the formation of *Phakopsora pachyrhizi* appressoria. **BMC Genomics**, v. 13, p. 269, 2012.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. Mega6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725-9, 2013.

TREMBLAY, A.; HOSSEINI, P.; LI, S.; ALKHAROUF, N.W.; MATTHEWS, B.F. Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four

time-points during infection of a susceptible soybean cultivar using deep sequencing. **BMC Genomics**, v. 14, p. 614, 2013.

TREMBLAY, A.; HOSSEINI, P.; LI, S.; ALKHAROUF, N.W.; MATTHEWS, B.F. Identification of genes expressed by *Phakopsora pachyrhizi*, the pathogen causing soybean rust, at a late stage of infection of susceptible soybean leaves. **Plant Pathology**, v. 61, p. 773–786, 2012.

TREMBLAY, A.; LI, S.; SCHEFFLE, B.E.; MATTHEWS, B.F. Laser capture microdissection and expressed sequence tag analysis of uredinia formed by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. **Physiological and Molecular Plant Pathology**, p. 12, 2009.

VIAUD, M.C.; BALHADERE, P.V.; TALBOT, N.J. A *Magnaporthe grisea* cyclophilin acts as a virulence determinant during plant infection. **Plant Cell**, v. 14, p. 917–930, 2002.

VIEIRA, A.; TALHINHAS, P.; LOUREIRO, A., et al. Expression profiling of genes involved in the biotrophic colonisation of *Coffea arabica* leaves by *Hemileia vastatrix*. **European Journal Plant Pathology**, v. 133, p. 261–277, 2012.

WEBB, C.A.; SZABO, L.J.; BAKKEREN, G.; GARRY, C.; STAPLES, R.C.; EVERSMEYER, M.; FELLERS, J.P. Transient expression and insertional

mutagenesis of *Puccinia triticina* using biolistics. **Functional & Integrative Genomics**, v. 6, p. 250–260, 2006.

YAMANAKA, N.; YAMAOKA, Y.; KATO, M.; LEMOS, N.G.; PASSIANOTTO, A.L.L.; DOS SANTOS, J.V.M.; BENITEZ, E.R.; ABDELNOOR, R.V.; SOARES, R.M.; SUENAGA, K. Development of classification criteria for resistance to soybean rust and differences in virulence among Japanese and Brazilian rust populations. **Tropical Plant Pathology**, v. 35, n. 3, p. 153-162, 2010.

YIN, C.T.; DOWNEY, S.I.; KLAGES-MUNDT, N.L.; RAMACHANDRAN, S.; CHEN, X.; SZABO, L.J.; PUMPHREY, M.; HULBERT, S.H. Identification of promising host-induced silencing targets among genes preferentially transcribed in haustoria of *Puccinia*. **BMC Genomics**, v. 16, p. 579, 2015.

YIN, C.T.; JURGENSON, J.E.; HULBERT, S.H. Development of a host-induced RNAi system in the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. **Molecular Plant–Microbe Interactions**, v. 24, p. 554–561, 2011.

YORINORI, J. T.; PAIVA, W. M.; FREDERICK, R. D.; COSTAMILAN, L. M.; BERTAGNOLLI, P. F.; HARTMAN, G. L.; GODOY, C. V.; NUNES-JUNIOR, J. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. **Plant Disease**, Saint Paul, v. 89, n. 4, p. 675-677, 2005.

ZHOU, B.; BAILEY, A.; NIBLETT, C.; QU, R. Control of brown patch (*Rhizoctonia solani*) in tall fescue (*Festuca arundinacea* Schreb.) by host induced gene silencing. Plant Cell Rep, v. 35, p. 791–802, 2016.

ZHANG, T.; JIN, Y.; ZHAO, J.H.; GAO, F.; ZHOU, B.J.; FANG, Y.Y.; GUO, H.S. Host-induced gene silencing of target gene in fungal cells confers effective resistance to cotton wilt disease pathogen *Verticillium dahliae*. **Molecular Plant**, doi: 10.1016/j.molp.2016.02.008, 2016.

ZHANG, H.; GUO, J.; VOEGELE, R.T.; ZHANG, J.; DUAN, Y.; LUO, H.; KANG, Z. Functional characterization of calcineurin homologs PsCNA1/PsCNB1 in *Puccinia striiformis* f. sp. *tritici* using a host-induced RNAi system. **PLoS One**, v. 7, e49262, 2012.

ZHANG, C.; BRADSHAW, J.D.; WHITHAM, S.A.; HILL, J.H. The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. **Plant Physiology**, v. 153, p. 52–65, 2010.

ZHANG, C.; YANG, C.; WHITHAM, S.A.; HILL, J.H. Development and use of an efficient DNA-based viral gene silencing vector for soybean. **Molecular Plant Microbe Interactions**, v. 22, p. 123–131, 2009.

3.6 SUPPLEMENTARY FILES

Table S1. RT-qPCR and HIGS primers used in this study, sequences, amplicon size and primer efficiency of selected *Phakopsora pachyrhizi* contigs.

Contig/Construct	Primers (F - forward, R - reverse)	Amplicon size (pb)	Primer efficiency (%)
RT-qPCR primers			
Thi (de_novo_595)	F - TCACTGAGCTAATCGGTACAGG R - GCCTTTCCAGCCAAAGTATG	73	96.8
Gpα (de_novo_4692)	F - GTACGAATCAGCGGCAAAG R - TTGAGAGTCAGTGGCACAGG	103	92.7
PPI (de_novo_4668)	F - TCGTCAGCTTTGCCTTAGAC R - CTCCCTGGAGCATAAATTGG	86	98.3
AGO (de_novo_939)	F – ACTCGCTCGGTTTCTATTGC R - CCGTTGAAATCGACGTTACC	92	90.2
Pv-SNARE (de_novo_2740)	F – CCGTCAATGATCCATACGTG R - AGAAGTGATGCCCGTTGTTC	113	89.6
HSS (de_novo_57)	F – GACGCAAGTTTGATGACTCG R - AACAGAGATAACCGGCTTGC	92	90.4
Tub (de_novo_380)	F – CCAAGGCTTCTTCGTGTTTCA R - AGAGAAGAGCGCCAAACC	65	93.5
Gpα (de_novo_4692)[a]	F – GGAACGCAAGAAGTGGATTC R - GCTTCAGCCATTCGATTCAC	120	88.7
18S (de_novo_6)[a]	F – TCTAAGAAGTCAGCAGCCAGCC R - GGTGCATGGCCGTTCTTAGT	115	91.6
HIGS primers			
pBPMV-Thi	F: AAATGGATCCTTGGAGTGTGTGTGTGGTC R: AAATGGATCCCTTTCATAGGTTGCCAAAGAC	306	-
pBPMV- Gpα	F: ATATCTCGAGAATGGCTGAAGCAGCAAC R: AAATGGATCCTGTACACCAGCAAGTACAACAG	291	-
pBPMV- PPI	F: ATATCTCGAGCACGCTCTAGATTTGTGCTC R: AAATGGATCCGGTGAATCAATGCTTCCTG	283	-
pBPMV- AGO	F- ATATCTCGAGGCAGCGATCAAGAAGTTG R - AAATGGATCCGAAGTTGCAAAGAGGAAGTG	258	-
pBPMV- Pv-SNARE	F – ATATCTCGAGTAACTCCAGGAGGAAACGTG R - AAATGGATCCGTGCCTGGTGACTATTTACATC	266	-
pBPMV- HSS	F – ATATCTCGAGAGCCGGTTATCTCTGTTGAG R - AAATGGATCCGATAAGGACATTGCGCTCTC	293	-
pBPMV-18S	F - ATATCTCGAGCTCCCACTTTAGTTGTGCTC R - AAATGGATCCTTGGAGTGTGTGTGTGGTC	282	-

[a]RT Primers used exclusively in the silencing RT-qPCR analyzes.

Table S2. *Phakopsora pachyrhizi* contigs related to PTGS machinery and to survival and pathogenicity mechanisms identified in *in silico* analysis.

Putative sequence function	Pp contigs ID	Contig length (bp)	Pfam domain	Domain position (start / stop) bp	FPKM values[1] BRS 231	FPKM values[1] PI561356
PTGS machinery						
Dicer	de_novo_33348	266	DEAD/DEAH	16 / 74	1.48	0.80
	de_novo_13436	507	No domain found	-	8.21	1.12
	de_novo_17407	421	No domain found	-	5.34	3.84
Argonaut	de_novo_939	1,522	Piwi	7 / 238	320.67	217.59
	de_novo_13083	516	Piwi	7 / 99	13.90	17.59
	de_novo_33823	264	Piwi	1 / 78	5.54	4.06
	de_novo_6063	800	ArgoN[2]	200 / 344	0.14	0.00
	de_novo_10102	610	No domain found	-	0.04	0.00
	de_novo_16828	432	No domain found	-	33.40	36.01
	de_novo_17902	413	No domain found	-	1.57	2.25
	de_novo_40389	232	No domain found	-	0.36	1.84
RNAse dependent of RNA polymerase	de_novo_7	3,558	RdRP[2]	721 / 1,157	0.00	0.00
	de_novo_43011	220	RdRP	2 / 71	0.00	1.84
	de_novo_30111	286	RdRP	13 / 82	0.68	1.16
	de_novo_1438	1364	no domain found	-	36.54	22.97

Survival and pathogenicity mechanisms						
Thiamine bioshyntesis	de_novo_595	1,702	NMT1[2]	131 / 351	1,533.76	743.64
Peptidyl-prolyl isomerase	de_novo_4668	890	Pro isomerase	54 / 193	1,196.55	779.43
	de_novo_15523	458	Pro isomerase	60 / 145	1.10	0.78
	de_novo_17691	416	Pro isomerase	8 / 91	3.51	0.11
G protein α subunit	de_novo_4692	888	G-alpha	1 / 170	7.95	3.97
18S ribosomal subunit	de_novo_6	3,586	No domain found	-	250.98	216.66
	de_novo_30180	286	No domain found	-	56.95	53.31
Small heat shock	de_novo_57	2,550	HSP70[2]	128 / 732	383.09	107.03
	de_novo_822	1,570	HSP70	96 / 351	743.64	204.24
	de_novo_2534	1,138	HSP70	58 / 337	0.02	0.00
	de_novo_12138	542	HSP70	123 / 180	6.33	2.20
	de_novo_3306	1,035	HSP70	124/308	20.9	14.9
	de_novo_46321	208	No domain found	-	0.00	2.29
	de_novo_25016	325	No domain found	-	1.99	0.95
Soluble NSF attachment receptor	de_novo_2740	1,107	V-SNARE-C[2]	157 / 220	181.33	175.74

[1]FPKM values of *P. pachyrhizi* transcripts obtained from lesions from susceptible (BRS 231) and resistant (PI561356) genotypes leaves after 10 days of infection corresponding of the *P. pachyrhizi* transcriptome database, available in LGE database (http://bioinfo03.ibi.unicamp.br/phakopsora/).
[2]Complete domain found in *P. pachyrhizi* transcripts sequences.

Table S3. Fold change and standard deviation values obtained during gene expression analyzes of the six genes selected from *Phakopsora pachyrhizi* by RT-qPCR.

Genes	*P. pachyrhizi* infection time points[1]									
	S	**GS**	**0hpi**	**6hpi**	**12hpi**	**24hpi**	**48hpi**	**72hpi**	**96hpi**	**192hpi**
Isolate LUB112										
Thi	0,043 (0,09)	0,033 (0,53)	0,097 (0,12)	0,036 (0,06)	0,330 (0,29)	0,134 (0,17)	0,607 (0,25)	0,943 (0,06)	1,106 (0,17)	1,549 (0,24)
PPI	0,545 (0,11)	1,236 (0,19)	0,189 (0,10)	0,188 (0,05)	2,579 (1,23)	2,710 (1,25)	2,129 (0,92)	2,813 (1,00)	3,376 (1,20)	3,305 (0,50)
Gpα	0,257 (0,09)	0,232 (0,36)	7,671 (2,94)	1,072 (0,96)	3,027 (2,27)	1,148 (0,45)	0,726 (0,57)	0,577 (0,34)	0,405 (0,26)	0,463 (0,26)
Pv-SNARE	4,839 (0,07)	4,151 (0,11)	4,513 (0,92)	3,262 (0,11)	0,820 (0,60)	0,971 (0,95)	0,192 (0,08)	0,111 (0,02)	0,109 (0,02)	0,142 (0,00)
HSS	0,161 (0,06)	0,307 (0,14)	0,683 (0,51)	0,895 (0,11)	1,628 (0,81)	1,205 (0,66)	1,477 (0,60)	0,676 (0,15)	0,783 (0,24)	1,186 (0,32)
AGO	1,016 (0,15)	0,878 (0,08)	0,773 (0,13)	1,437 (0,81)	2,010 (0,07)	1,615 (0,52)	0,672 (0,17)	0,851 (0,09)	0,759 (0,09)	1,918 (0,51)
Isolate LD5511										
Thi	0,004 (0,10)	0,011 (0,22)	0,187 (0,20)	0,003 (0,00)	0,052 (0,05)	0,467 (0,40)	1,054 (0,33)	0,843 (0,05)	0,894 (0,18)	1,589 (0,57)
PPI	0,212 (0,14)	0,412 (0,14)	0,282 (0,16)	0,082 (0,05)	0,392 (0,16)	1,118 (0,63)	1,211 (0,42)	1,386 (0,21)	1,658 (0,36)	2,340 (0,13)
Gpα	0,471 (0,24)	1,100 (0,49)	1,487 (1,25)	0,243 (0,09)	1,283 (0,65)	0,569 (0,21)	0,532 (0,11)	0,366 (0,11)	0,573 (0,18)	0,720 (0,10)
Pv-SNARE	2,636 (0,15)	8,464 (0,12)	3,144 (0,34)	7,926 (1,31)	5,174 (0,69)	0,829 (0,22)	0,397 (0,15)	0,293 (0,06)	0,268 (0,14)	0,191 (0,02)
HSS	0,144 (0,16)	0,457 (0,05)	1,245 (0,20)	1,138 (0,11)	2,985 (0,53)	1,103 (0,27)	1,677 (0,35)	1,617 (0,75)	1,204 (0,09)	0,836 (0,19)
AGO	1,083 (0,14)	2,075 (0,18)	0,854 (0,27)	1,613 (0,91)	3,075 (0,90)	2,773 (0,49)	1,157 (0,13)	1,048 (0,06)	1,375 (0,04)	2,782 (0,31)

Fold change values are shown in the table followed by their respective standard deviation values, shown in parentheses.
[1]Main *P. pachyrhizi* infection time points: spore (S) and germinated spore (GS) stages before contact with soybean, and after soybean contact at 0, 6, 12, 24, 48, 72 , 96 and 192 hours post infection.

Table S4. Accurate significance levels (p) for the contrasts between treatments and control pBPMV (empty vector).

Treatments	Sporulation Level	Number of uredinia	Number of open uredinia	Disease Severity
pBPMV-18S	0.0000*	0.0176*	0.0000*	0.0049*
pBPMV-Thi	0.0972	0.9445	0.0323*	0.0000*
pBPMV-AGO	0.0000*	0.9997	0.9989	0.1867
pBPMV-PPI	0.0000*	0.9591	0.0400*	0.0001*
pBPMV-Gpα	0.0000*	0.0034*	0.0014*	0.0000*
pBPMV-Pv-SNARE	0.0000*	0.0229*	0.0090*	0.0000*
pBPMV-HSS	0.9152	0.9755	0.0323*	0.4091

*Significant values at P>0.05% (Tukey-Kramer test)

4. CONSIDERAÇÕES FINAIS

Os resultados gerados neste estudo fornecem valiosas informações a respeito das interações moleculares existentes entre o fungo *P. pachyrhizi* e plantas de soja durante o processo infeccioso, contribuindo para o desenvolvimento de estratégias de combate e controle da ferrugem asiática da soja.

No capítulo 2, foram descritos os principais transcritos e rotas metabólicas que atuam no fungo após 10 dias do início da infecção, resultando em um banco de dados de transcriptoma composto por um elevado número de sequências, que constitui uma fonte confiável para futuras análises moleculares. A análise de transcritos do fungo encontrados em contato com o genótipo resistente de soja permitiu a identificação de importantes vias metabólicas envolvidas diretamente com as respostas do patógeno às vias de resistência encontradas nestas plantas. Adicionalmente, foi possível notar a grande quantidade de sequências de haustório encontradas entre as sequências geradas neste trabalho.

O capítulo 3 parte dos dados gerados pelo segundo capítulo, para a identificação e seleção de possíveis candidatos para a caracterização funcional de genes do fungo, por meio de HIGS. Os resultados apresentados mostraram a eficiência dessa metodologia no estudo funcional de genes neste patógeno, evidenciando que os genes selecionados estão, de fato, relacionados e são de extrema importância para o desenvolvimento de *P. pachyrhizi* no tecido hospedeiro durante o processo de infecção.

Novos estudos ainda são necessários para elucidar os mecanismos-chave que determinam o estabelecimento e a manutenção de *P. pachyrhizi* no tecido hospedeiro, principalmente em relação à caracterização funcional de genes

possivelmente relacionados a patogenicidade desse fungo. As análises de enriquecimento realizadas nesse estudo, por exemplo, podem fornecer fortes candidatos para o estudo funcional de genes, tanto a partir dos transcritos do fungo obtidos exclusivamente em contato com o genótipo de soja resistente (PI561356), quanto dos transcritos identificados entre as estruturas de infecção do fungo. Assim como para as análises de enriquecimento, os resultados obtidos a partir da análise comparativa entre *P. pachyrhizi* e outras 15 espécies de fungos e oomicetos, devem ser explorados mais cuidadosamente.

Além da utilização de vetores virais, como foi descrito neste trabalho, outras abordagens também podem constituir uma ferramenta viável para a caracterização funcional de genes do fungo via HIGS, como a aplicação direta de moléculas de dsRNA ou siRNA. Alguns estudos já estão sendo realizados no laboratório de Biotecnologia da Embrapa Soja nesse sentido, e os resultados tem se mostrado promissores, como a possível absorção de moléculas de dsRNA por urediniósporos de *P. pachyrhizi*, além da redução da patogenicidade observada em testes iniciais realizados para candidatos a efetores do fungo.