

LÍVIA SOARES ZARAMELA

**CARACTERIZAÇÃO DO SECRETOMA DE *Phakopsora pachyrhizi*
EXPRESSO DURANTE A INTERAÇÃO COM A SOJA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2011

LÍVIA SOARES ZARAMELA

**CARACTERIZAÇÃO DO SECRETOMA DE *Phakopsora pachyrhizi*
EXPRESSO DURANTE A INTERAÇÃO COM A SOJA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA: 19 de julho de 2011.

Prof. Eduardo Seiti Gomide Mizubuti

Prof^a. Marisa Vieira de Queiroz
(Coorientadora)

Prof. Sérgio Hermínio Brommonschenkel
(Orientador)

AGRADECIMENTOS

Aos meus pais, Sérgio e Waldete, pelo apoio incondicional as minhas escolhas, por sempre estarem ao meu lado, por tudo.

A minha irmã, Luiza, pelo carinho ao longo destes anos, mesmo estando tão distante. A minha irmã de coração, Ancély, pelo companheirismo, pela amizade e pelas longas, longas conversas. A minha amiga Renata pela companhia, pelas “brincas”, por chegar em casa e saber que teria alguém para contar como foi o dia.

Ao meu namorado e melhor amigo, Ricardo, por estar sempre comigo, pela preocupação, pelo amor, pelo carinho, pelos conselhos e pelo constante incentivo.

A todos os amigos do Laboratório de Genômica, Sandra, Jamile, Gustavo, Michele, Yukari, Dani “Tombi”, Poliene, Daniel, Gabi, Elías, Tadeu, Ricardo, Marcos, Cynthia, Thiago, Tácio, Rafinha, Josney, Luiz pelo convívio e pela amizade. Um agradecimento especial a Sandra, o trabalho é tão meu quando dela, por transmitir calma nos momentos mais difíceis.

Ao meu orientador, Professor Sérgio Hermínio Bromonschenkel, pela oportunidade e orientação durante a execução deste trabalho. Por todas as lições de bancada e por mostrar o caminho do mundo da pesquisa.

A Universidade Federal de Viçosa, por toda a minha formação, pela infraestrutura e pelo apoio para realização deste trabalho.

A CAPES pela bolsa concedida.

ÍNDICE

RESUMO	v
ABSTRACT	vii
1.0 – Introdução	1
2.0 – Revisão de literatura.....	3
2.1 – Aspectos gerais do patossistema soja – <i>Phakopsora pachyrhizi</i>	3
2.1.1 – Soja (<i>Glycine max</i>).....	3
2.1.2 – Ferrugem da soja.....	4
2.1.2.1 – Ciclo de vida das ferrugens.....	5
2.1.2.2 – Ferrugem asiática da soja (<i>P. pachyrhizi</i>)	7
2.1.2.3 – Sintomatologia e controle	11
2.2 – Interação planta-patógeno.....	12
2.2.1 – Sistema imune vegetal	12
2.2.2 – Efeitos: identificação e caracterização.....	14
3.0 – Objetivos	22
3.1 – Objetivo geral.....	22
3.2 – Objetivos específicos	22
4.0 – Materiais e Métodos.....	23
4.1 – Material vegetal, cultivo das plantas e inoculação.....	23
4.2 – Construção da biblioteca YST e obtenção de DNA plasmidial	23
4.3 – Transformação de levedura pelo método AcLi/PEG e triagem da biblioteca YST	24
4.4 – Extração plasmidial de levedura	25
4.5 – Reação de PCR, purificação do produto de PCR e sequenciamento dos insertos dos clones positivos.....	26
4.6 – Análises de bioinformática.....	27
5.0 – Resultados	29
6.0 – Discussão	50
7.0 – Conclusões	59
8.0 – Referências bibliográficas.....	60
ANEXO.....	71

RESUMO

ZARAMELA, Livia Soares, M.Sc., Universidade Federal de Viçosa, julho de 2011. **Caracterização do secretoma de *Phakopsora pachyrhizi* expresso durante a interação com a soja.** Orientador: Sérgio Hermínio Brommonschenkel. Coorientadores: Gleiber Quintão Furtado e Marisa Vieira de Queiroz.

Durante a interação com o hospedeiro, fungos biotróficos, como as espécies causadoras das ferrugens, secretam um arsenal de proteínas (efetores) envolvidas na supressão da resposta de defesa e manipulação do metabolismo do hospedeiro. Essas proteínas efetoras são direcionadas para o apoplasto ou para o citoplasma das plantas por uma rota ainda desconhecida, onde atuam promovendo a infecção. Algumas dessas proteínas efetoras, denominadas proteínas Avr, são reconhecidas por proteínas codificadas por genes de resistência, o que desencadeia a resposta de hipersensibilidade e fenótipo de resistência. Já foram descritos sete genes *R* que conferem resistência à ferrugem asiática da soja causada pelo fungo *P. pachyrhizi*. Todavia, ainda não foram identificadas as proteínas efetoras (Avr) reconhecidas pelas proteínas codificadas pelos genes *Rpp*, ou efetuada uma análise detalhada dos genes de *P. pachyrhizi* que codificam proteínas secretadas durante a sua interação com a soja. Desta forma, o objetivo deste trabalho foi identificar genes de *P. pachyrhizi* que codificam proteínas secretadas utilizando o sistema armadilha de sinal de secreção em leveduras (YST). A partir da triagem da biblioteca YST, foram selecionados 2054 clones positivos. A partir desses clones foram obtidas 1777 sequências da extremidade 5' com qualidade phred > 20, que foram agrupadas em 95 contíguos e 66 singletos. Análises comparativas com sequências depositadas nos bancos de dados de ESTs de soja e de proteínas não redundantes do Genbank/NCBI resultaram na identificação de 48 contíguos e 16 singletos sem similaridade a sequências de soja (*no hit*). Estas sequências *no hit* são potenciais genes de origem fúngica. Para corroborar com esta hipótese, as sequências *no hit* foram comparadas com o

banco de dados do NCBI de sequências genômicas de *P. pachyrhizi* NCBI, sendo que 20 contíguos e 4 singletos apresentaram similaridade significativa com sequências depositadas neste banco de dados. Quando comparadas com sequências de proteínas do *Puccinia Database Protein* pelo programa *BlastX*, 12 ORFs preditas codificam proteínas com similaridade significativa com proteínas de *Puccinia*. A predição de peptídio sinal através do SignalP foi positiva para 88,9% das ORFs deduzidas. Além disso, 87,7% das ORFs deduzidas também apresentaram predição para secreção pelo programa TargetP, e 92,6% não apresentam predição para domínio transmembrana. Dentre as proteínas codificadas por estes genes, duas apresentaram motivo semelhante ao RxLx-dEER de oomicetos, e uma proteína com motivo semelhante ao Y/F/WxC encontrado em candidatos a efetores de oídios. Foram também identificadas três regiões protéicas conservadas, ainda não relatadas na literatura. A predição para pontes de dissulfeto foi positiva para 14 proteínas deduzidas. A sequência completa foi obtida para 29 genes, que foram selecionados para caracterização funcional, sendo que destes, 12 já haviam sido previamente identificados em estudos anteriores realizados no Laboratório de Genômica. A confirmação da secreção em levedura já foi realizada para nove clones. Ensaio de secreção para os demais genes estão sendo realizados. O monitoramento da expressão dos genes identificados e a sua análise de expressão transiente em genótipo resistentes serão utilizados para determinação a sua função na patogênese.

ABSTRACT

ZARAMELA, Livia Soares, M.Sc. Universidade Federal de Viçosa, July, 2011. **Characterization of the *Phakopsora pachyrhizi* secretome expressed during the interaction with soybean.** Advisor: Sergio Hermínio Brommonschenkel. Co-Advisors: Gleiber Quintão Furtado and Marisa Vieira de Queiroz.

During the interaction with host plants rust fungi secrete an arsenal of effector proteins involved with suppressing host defense responses and redirecting normal host metabolism to facilitate infection. These effector proteins are directed to the apoplast and/or into plant cell cytoplasm by a unknown mechanism. Some effector proteins, called Avr proteins, are recognized by proteins encoded by resistance R genes. This recognition triggers defense responses that lead to a resistance phenotype. Seven R genes (*Rpp* genes) that confer resistance to Asian soybean rust caused by *P. pachyrhizi* have been described. However, the Avr proteins recognized by the proteins encoded by these *Rpp* genes have not yet been identified. Also, *P. pachyrhizi* genes that encode proteins secreted during the interaction with soybean have not yet been described. Therefore, the objective of this work was to identify genes of *P. pachyrhizi* that encode proteins secreted during the interaction with soybean using the Yeast Secretion Trap system (YST system), aiming the future identification of Avr genes. From the screening of one YST library constructed using cDNA synthesized from mRNA isolated from infected soybean leaves, 2054 positive clones were selected. From these clones, 1777 single pass 5'-end cDNA sequences with a Phred quality score > 20 were obtained, which were grouped into 95 contigs and 66 singlets. Comparative analyses with sequences deposited in soybean ESTs (BlastN) and Genbank/NCBI non-redundant protein databases (BlastX) resulted in the identification of 48 contigs and 16 singlets no hits that are potential *P. pachyrhizi* genes. Corroborating this hypothesis, 20 contigs and 4 singlets were similar to trace genome sequences from *P. pachyrhizi* (1X coverage) deposited at NCBI. Deduced

ORFs from twelve genes encode proteins with similarity to *Puccinia* proteins. Signal peptide prediction by SignalP algorithm was positive for 88.9% of the deduced proteins. In addition to, 87.7% of these deduced proteins showed prediction for secretion by TargetP algorithm, and 92.6% did not showed transmembrane domain prediction. Two deduced proteins have a motif similar to oomycete RxLx-dEER translocation domain, and one protein has a motif similar to the Y/F/WxC motif found in powdery mildew's effector candidates. Three new conserved protein regions, not yet reported in the literature, were also identified. Disulfide bond prediction was positive for 14 deduced proteins. The complete sequence of 29 genes that were selected for functional characterization was obtained, and yeast secretion of proteins encoded by nine selected genes has been confirmed. Secretion assays for proteins encoded by the other genes are in progress. Time course expression analyses and transient expression studies in resistant genotypes will be used to determine the function of the identified genes during the pathogenesis.

1.0 – Introdução

A soja (*Glycine max* L. Merrill) originária da Ásia é cultivada há milhares de anos e, é plantada em todo o mundo. Seus grãos dão origem a grande variedade de produtos e subprodutos utilizados pela agroindústria, indústria química e de alimentos. A produção nacional de soja, safra 2009/2010, é estimada em 68,47 milhões de toneladas (Conab, 2010). Entretanto, as perdas em produção decorrentes de problemas de ordem fitossanitária restringem o aumento da produtividade.

Uma das doenças mais severas que incidem sobre a cultura da soja é a ferrugem asiática. Desde sua detecção no Brasil, estima-se que os custos com a ferrugem já ultrapassem US\$15 bilhões (Consórcio Antiferrugem – Embrapa Soja, 2010). Esta doença é causada pelo fungo *Phakopsora pachyrhizi*, pertencente ao filo Basidiomycota, classe Urediniomycetes, ordem Uredinales, família Phakopsoraceae (Loehrer *et al.*, 2010). A ferrugem não reduz somente o rendimento, mas também afeta negativamente a qualidade dos grãos, diminuindo seu peso (Webb e Fellers, 2006).

Durante a interação com o hospedeiro, fungos biotróficos, como os causadores de ferrugem, secretam proteínas envolvidas na supressão da resposta de defesa e manipulação do metabolismo do hospedeiro (Panstruga e Dodds, 2009). Essas proteínas, denominadas proteínas efetoras, são direcionadas para o apoplasto ou para o citoplasma das plantas por uma rota ainda desconhecida, onde atuam promovendo a infecção. Algumas dessas proteínas efetoras, denominadas proteínas Avr, são reconhecidas direta ou indiretamente por proteínas codificadas por genes de resistência, o que desencadeia a resposta de hipersensibilidade e o fenótipo de resistência (Mendgen e Hahn, 2002). Já foram descritos sete genes em cinco locos designados Rpp1 a Rpp5 (*Rpp - Resistance to Phakopsora pachyrhizi*), que conferem resistência à ferrugem asiática da soja causada por *P. pachyrhizi* (Bromfield e Hartwig, 1980; Mclean e Byth, 1980; Hartwig e Bromfield, 1983; Hartwig, 1986; Hyten *et al.*, 2007; Calvo *et al.*, 2008; Silva *et al.*, 2008; Chakraborty *et al.*, 2009). Todavia, ainda não foram identificadas as

proteínas efetoras (Avr) reconhecidas pelas proteínas codificadas por esses genes *Rpp*.

Os secretomas de fungos fitopatogênicos são complexos e embora alguns componentes individuais tenham sido revelados, os processos de interação e translocação de efetores para o hospedeiro permanecem desconhecidos (Dodds *et al.*, 2009). O secretoma da interação soja - *P. pachyrhizi* ainda não foi caracterizado. Apesar de existirem etiquetas de sequências expressas (ESTs) de *P. pachyrhizi* depositadas no GenBank, pouco se sabe sobre o seu processo de infecção a nível molecular (Luster *et al.*, 2010).

Uma abordagem que pode ser utilizada na caracterização de secretomas é o sistema armadilha de sinal de secreção em leveduras (YST, *Yeast Secretion Trap*), que se baseia na ligação de um *pool* da extremidade 5' de cDNA heterólogos nos vetores pYST0-1. Estes vetores apresentam a sequência truncada do gene da invertase (*suc2*), cuja região codificadora do peptídeo sinal é removida. Apenas a clonagem *in frame* de cDNAs com região codificadora para sinal de secreção com o gene da invertase, permitirá a síntese de uma proteína recombinante capaz de ser secretada e capaz de metabolizar sacarose. As construções, são posteriormente transformadas em uma linhagem de *Saccharomyces cerevisiae* SUC⁻ e a seleção dos transformantes é realizada em meio de cultura contendo sacarose como única fonte de carbono (Lee *et al.*, 2006). Este método tem sido utilizado com sucesso na identificação de genes de oomicetos e de fungos filamentosos que codificam proteínas secretadas (Lee *et al.*, 2006, Catanzariti *et al.*, 2006).

Diante do exposto, os objetivos deste trabalho foram identificar genes de *P. pachyrhizi* que codificam proteínas secretadas, através da construção e triagem de uma biblioteca YST; selecionar, por meio de ferramentas de bioinformática, dentre os clones positivos no sistema YST, aqueles que apresentam cDNA de origem fúngica; obter a sequência completa (cDNA) dos genes identificados; confirmar a secreção da proteína codificada pela ORF completa dos genes selecionados pelo sistema YST.

2.0 – Revisão de literatura

2.1 – Aspectos gerais do patossistema soja – *Phakopsora pachyrhizi*

2.1.1 – Soja (*Glycine max*)

A soja (*Glycine max*) é uma planta da família *Fabaceae* originária de plantas rasteiras ancestrais que se desenvolviam na costa Leste da Ásia, principalmente na China. No Brasil, a soja foi introduzida no ano de 1882, porém, seu cultivo só teve início no ano de 1914, no município de Santa Rosa, RS, onde a cultura encontrou condições para se desenvolver e expandir. O interesse comercial no cultivo de soja iniciou-se na década de 60. Na década de 70, com o aumento do preço da soja no mercado mundial, o país passou a investir em tecnologias para adaptação da cultura às condições brasileiras e a soja se consolidou como a principal cultura do agronegócio brasileiro. Atualmente, o país é o segundo maior produtor de soja do mundo com produção de 68,7 milhões de toneladas e área plantada de 23,6 milhões de hectares na safra de 2009/2010, ficando atrás apenas dos Estados Unidos, maior produtor mundial do grão, com produção de 91,4 milhões de toneladas e área plantada de 30,9 milhões de hectares, na mesma safra (Embrapa Soja, 2011).

Problemas de ordem fitossanitária estão entre os principais fatores que limitam a obtenção de alta produtividade em soja. A grande expansão da área cultivada no mundo proporcionou aumento do número de doenças que afetam essa cultura, sendo que há mais de 35 espécies de patógenos relatadas para essa oleaginosa (Embrapa Soja, 2011). Dentre as doenças que afetam o plantio de soja, a ferrugem asiática é a que causa maior preocupação. Por se tratar de uma doença policíclica com alto poder destrutivo, pode causar perdas significativas no rendimento da safra. Esta doença causa grandes impactos econômicos à cultura de soja no Brasil, onde as condições ambientais são favoráveis ao desenvolvimento de epidemias (Scherm *et al.*, 2009).

2.1.2 – Ferrugem da soja

A ferrugem foi inicialmente relatada no Japão em 1902 e na Austrália em 1934. Desde então está estabelecida em todas as áreas de cultivo de soja no mundo (Edwards e Bonde, 2011). Na América, o primeiro relato da ferrugem foi em Porto Rico, em 1976, já na América do Sul foi relatada pela primeira vez em 2001, na divisa entre Paraguai e Brasil, causando severas perdas na produção de soja (Zambolim, 2006). Após a primeira constatação no Brasil em 2001, a ferrugem asiática espalhou-se gradativamente por todo país. Na safra de 2001/2002 já atingia 60% de toda a área brasileira de soja, e 90% da área na safra seguinte (Yorinori, 2005). Atualmente, a ferrugem asiática está disseminada em todas as regiões produtoras do país, exceto no estado de Roraima (Consórcio Antiferrugem – Embrapa Soja, 2011).

Na década de 80, acreditava-se que havia uma única espécie e diferentes populações de *Phakopsora* causando a ferrugem da soja (Bonde e Brown, 1980; Bromfield e Hartwig, 1980). A taxonomia do fungo causador da ferrugem da soja somente foi esclarecida em 1992, com o trabalho realizado pelo pesquisador Ono e colaboradores que demonstraram a ocorrência de duas espécies de *Phakopsora* causando ferrugem em plantas de soja, e que as diferenças de agressividade e virulência não se deviam apenas as características fisiológicas de isolados. Além de *P. pachyrhizi*, originária da região Australasiana, a ferrugem da soja pode ser causada por *P. meibomiae* (Arthur) Arthur, endêmica nas Américas (Ono *et al.*, 1992, Buritica *et al.*, 1994; Hennen *et al.*, 1995). Apenas as características morfológicas das duas espécies de *Phakopsora* muitas vezes não permitem sua diferenciação, a qual somente é feita por meio de análises moleculares. Estudos envolvendo a determinação da sequência de nucleotídeos da região ITS (*Internal Transcribed Spacer*) do rDNA revelaram que os isolados dentro de *P. pachyrhizi* e *P. meibomiae* possuem identidade maior que 99%, mas apenas 80% de identidade entre as duas espécies. Estas comparações possibilitaram a construção de oligonucleotídeos iniciadores para PCR (*Polymerase Chain Reaction*) que permitem a diferenciação das duas espécies (Frederick *et al.*, 2002).

2.1.2.1 – Ciclo de vida das ferrugens

Os patógenos biotróficos caracterizam-se por parasitismo específico em que são absolutamente dependentes do tecido vivo do hospedeiro para completar seu ciclo de vida. Fungos biotróficos distinguem-se dos fungos hemibiotróficos ou necrotróficos por: (a) não serem cultivados *in vitro*; (b) formarem estruturas de infecção altamente especializadas; (c) apresentarem atividade secretória de enzimas líticas limitada; (d) conseguirem suprimir a longo prazo as respostas de defesa da planta; (e) formarem haustório (Voegelé *et al.*, 2009).

Fungos causadores de ferrugens possuem ciclo de vida complexo, podendo formar até cinco tipos de estruturas: pícnio, écio, urédia, télia e pró-micélio (Ono *et al.*, 1992). Na urédia são produzidos os uredósporos, esporos assexuais responsáveis pela dispersão e multiplicação do patógeno em condições favoráveis à ocorrência da doença. O processo infeccioso dos fungos causadores de ferrugens pode ser dividido em três fases: fase de penetração, na qual o patógeno tem acesso ao interior do hospedeiro; fase biotrófica, na qual o patógeno estabelece as relações parasitárias com o hospedeiro e fase reprodutiva ou de esporulação, na qual o fungo se reproduz no tecido infectado (Bromfield *et al.*, 1980; Koch *et al.*, 1983). A fase de penetração tem início com a germinação do uredósporo depositado na superfície do hospedeiro. Em condições favoráveis de temperatura e umidade, o uredósporo germina, produzindo um tubo germinativo que forma em sua extremidade o apressório. Essa estrutura permite que o fungo penetre diretamente na superfície do hospedeiro. No interior do tecido do hospedeiro, a hifa fúngica forma uma estrutura especializada para a absorção de nutrientes denominada haustório, dando início à fase biotrófica. O haustório é uma hifa altamente especializada que penetra a parede celular vegetal e cresce dentro da célula do hospedeiro. Esta estrutura é circundada por uma membrana plasmática e parede celular. Todavia, ela não é formada diretamente no citoplasma da célula vegetal. Ao invés disso, a membrana plasmática da célula do hospedeiro invagina e forma uma membrana extra-haustorial ao redor do haustório. Na interface entre a parede celular do haustório e a membrana extra-haustorial é formada uma

camada de substância gelatinosa, rica em carboidratos, denominada matriz extra-haustorial (Catanzariti *et al.*, 2007). Assim, uma substância produzida pelo haustório deve atravessar sucessivamente a membrana plasmática do haustório, a matriz extra-haustorial e a membrana plasmática do hospedeiro, para ter acesso ao citoplasma vegetal (**Figura1**).

Os haustórios constituem o principal sítio de absorção de nutrientes a partir do hospedeiro e são determinantes para a fase biotrófica da infecção. Essas estruturas contêm transportadores de aminoácidos, transportadores de açúcares e bombas H⁺-ATPases envolvidas no transporte ativo de nutrientes a partir das células do hospedeiro. A formação do haustório também induz mudanças estruturais na célula hospedeira como, o rearranjo do citoesqueleto, migração do núcleo e condensação da cromatina (Mendgen e Hahn, 2002). Acredita-se que essas modificações são induzidas pela atuação de efetores produzidos nos haustórios, secretados na matriz extrahaustorial e translocados para o interior das células vegetais. Esta hipótese de translocação foi confirmada para fungos causadores de ferrugem, *Uromyces fabae* (Kemen *et al.*, 2005) e *Melampsora lini* (Wang *et al.*, 2007).

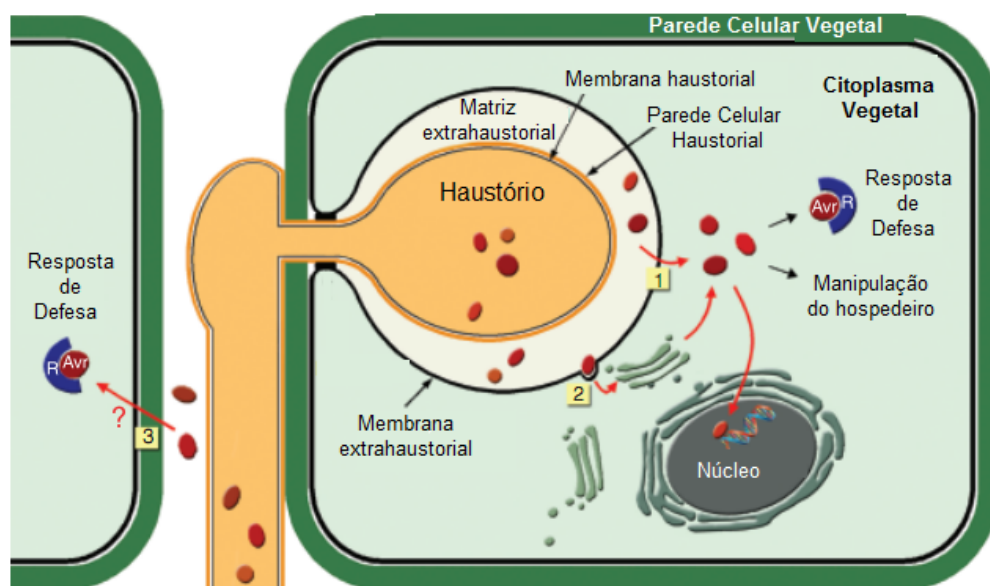


Figura 1 – A interface haustório-hospedeiro. Esquema de um haustório dentro de uma célula hospedeira mostrando a membrana extra-haustorial e a matriz extra-haustorial. Proteínas efectoras são secretadas a partir do haustório para a matriz extra-haustorial. Um subconjunto de proteínas é transportado para dentro da célula hospedeira, atravessando diretamente a membrana extra-haustorial (1) ou através de vesículas do sistema de endomembranas do hospedeiro (2). Uma vez dentro do citoplasma do hospedeiro, os efetores podem alterar seu metabolismo e vias de defesa. As proteínas efectoras que são reconhecidas pelas proteínas de resistência (R) são proteínas de avirulência (Avr). Este reconhecimento desencadeia as respostas de defesa. Outros efetores podem ser direcionados para organelas do hospedeiro, como o núcleo, onde podem alterar a transcrição. Os efetores secretados a partir das hifas podem também entrar nas células do hospedeiro através de um mecanismo ainda desconhecido (3) e, quando reconhecidos por uma proteína de resistência, podem também desencadear respostas de defesa (Adaptado de Catanzariti *et al.*, 2007).

2.1.2.2 – Ferrugem asiática da soja (*P. pachyrhizi*)

A ferrugem asiática da soja tem como agente etiológico o fungo *Phakopsora pachyrhizi*, pertencente ao filo Basidiomycota, classe Urediniomycetes, ordem Uredinales e família Phakopsoraceae. Este fungo pode infectar mais de 95 espécies de mais de 42 gêneros, incluindo a soja e outras

plantas da família *Fabaceae* (Rytter *et al.*, 1984; Ono *et al.*, 1992). Esta ampla gama de hospedeiros, incomum para a maioria das espécies de ferrugem, provavelmente é resultado de genes que contribuem para um complexo padrão de virulência (Hartman *et al.*, 2005).

A espécie *P. pachyrhizi* forma uredósporos em urédias sobre folhas colonizadas, cerca de 5 a 8 dias após a inoculação (**Figura 2d**). Os uredósporos são liberados da urédia e dispersos pelo vento. Sob condições apropriadas, temperatura entre 18 e 26°C e período mínimo de molhamento foliar de 6 horas, os uredósporos germinam produzindo um tubo germinativo. Este tubo forma em sua extremidade uma estrutura globosa, o apressório, aproximadamente do mesmo tamanho que o uredósporo (**Figura 2b**). O patógeno penetra diretamente no tecido foliar, iniciando com a formação de uma estrutura afunilada, denominada de cone apressorial, dentro do apressório (**Figura 3**). Este cone é contíguo com a parede celular da hifa de penetração formada dentro da célula epidérmica, também denominada de vesícula transepidermal (Bromfield, 1984). As células da epiderme penetradas pelo fungo tornam-se desorganizadas e necrosam. A hifa de penetração cresce através da célula epidérmica e atinge o espaço intercelular, sendo separada da hifa primária que emerge por um septo. A hifa primária se ramifica formando hifas secundárias que formam as células mãe do haustório nos locais de contato com células do mesófilo (**Figura 2c, Figura 3**). A formação do haustório primário é observada cerca de 24 a 48h após a inoculação (**Figura 2c, Figura 3**) (Koch *et al.*, 1983).

Estudos da interação entre *P. pachyrhizi* - *A. thaliana*, planta não hospedeira, comprovaram que espécies selvagens desta planta infectados por este fungo, acumulam H_2O_2 , que resulta na morte celular epidérmica (Loehrer *et al.*, 2008). Esta morte celular geralmente não impede a penetração do fungo no mesófilo foliar, e pode ser induzida pelo fungo para evitar a ativação das respostas de defesa da planta e para facilitar o seu ingresso nos espaços intercelulares da folha (Hoefle *et al.*, 2009). Assim, na etapa inicial de infecção, *P. pachyrhizi* apresenta característica de fungo necrotrófico e, quando atinge a camada foliar do mesófilo, passa a comportar-se como patógeno biotrófico por meio da formação

de haustórios (Loehrer *et al.*, 2008).

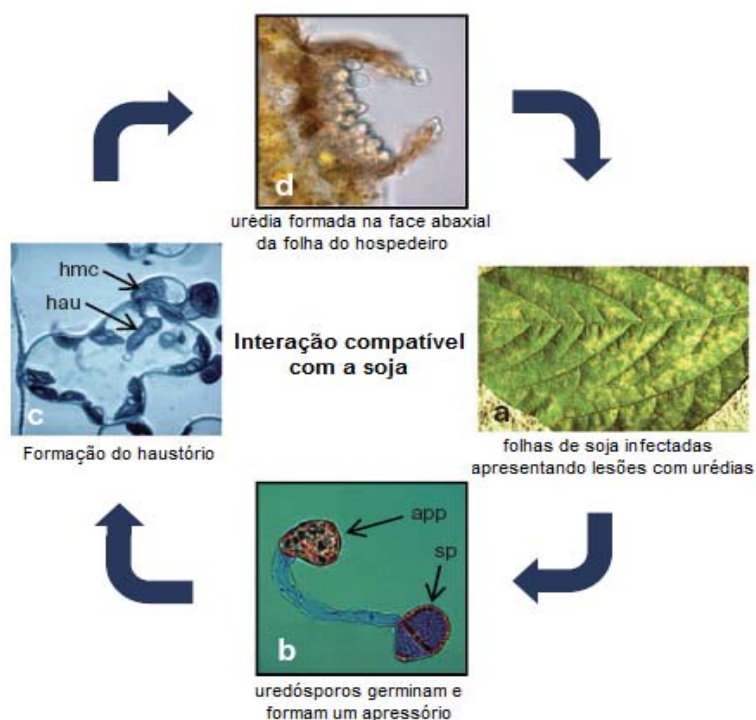


Figura 2 – Interação compatível entre *P. pachyrhizi* e a soja. a) Durante o processo de infecção, o patógeno forma urédias localizadas na face abaxial da folha. Os uredósporos formados são dispersos pelo vento. b) Esporos depositados sobre a superfície foliar germinam e formam apressório. c) Hifa intercelular forma células-mãe haustoriais de onde o haustório desenvolve-se dentro das células do mesófilo. d) O ciclo de infecção de *P. pachyrhizi* completa-se com a formação de novos uredósporos na urédia. (Adaptado de Goellner *et al.*, 2010)

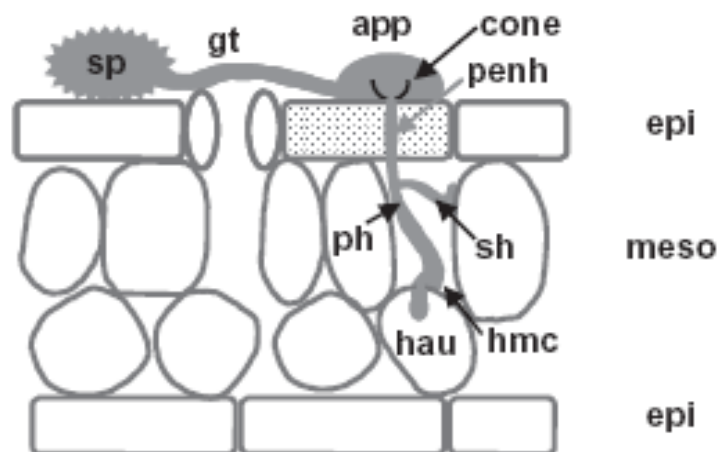


Figura 3 – Estágios iniciais do desenvolvimento de *P. pachyrhizi* durante a interação com planta hospedeira. Os uredósporos (sp) germinam produzindo um tubo germinativo (gt), que forma na sua extremidade, o apressório (app). Uma estrutura afunilada, o cone apressorial, se forma dentro do apressório. Este origina a hifa de penetração (penh) que perfura as células epidérmicas (epi) provocando sua morte (pontilhado). No espaço intercelular do mesófilo, a hifa primária (ph) é separada da hifa de penetração por um septo, e ramifica-se em várias hifas secundárias (sh). Finalmente, a célula-mãe haustorial (hmc), formada nos locais de contato da hifa secundária com as células do mesófilo (meso), diferencia o haustório (hau) no interior da célula do mesófilo (Adaptado de Koch *et al.*, 1983).

Além da formação de urédia e uredósporos, télia e teliósporos já foram reportados em hospedeiros de *P. pachyrhizi*, incluindo a soja (Yeh *et al.*, 1981; Sinclair, 1982). A formação dos teliósporos é influenciada pelo genótipo do cultivar, temperatura (médias diurnas acima de 20°C, com mínima diária de 16°C), umidade relativa e intensidade de luz (Tan-Yujun *et al.*, 2001a; Tan-Yujun *et al.*, 2001b). A germinação dos teliosporos, com formação de basidiósporos, tem sido observada em condições de laboratório, todavia, o seu papel no ciclo de infecção de *P. pachyrhizi* é desconhecido. Os estágios de pícnio e écio ainda não foram relatados (Saksirirat e Hoppe, 1991).

2.1.2.3 – Sintomatologia e controle

Os sintomas da ferrugem podem aparecer em qualquer estágio de desenvolvimento da planta, como em cotilédones, folhas e hastes, sendo mais característicos nas folhas. Os primeiros sintomas da ferrugem são minúsculos pontos de coloração esverdeada a cinza-esverdeada, mais escuros do que o tecido sadio da folha. No início da infecção, a folha permanece verde, dificultando a identificação da doença quando a lavoura é observada de forma superficial. Embora doenças como a mancha parda (*Septoria glycines*), o crestamento bacteriano (*Pseudomonas savastanoi* pv. *glycines*) e a pústula bacteriana (*Xanthomonas axonopodis* pv. *glycines*) causem sintomas semelhantes, a confirmação da ferrugem é feita pela constatação na face abaxial da folha, de saliências semelhantes a pequenas bolhas, que correspondem às estruturas de reprodução do fungo, as urédias. Progressivamente, a protuberância adquire coloração castanho-clara à castanho-escura, na qual se abre um minúsculo poro, expelindo os uredósporos. À medida que prossegue a esporulação, o tecido foliar ao redor das primeiras urédias, adquire coloração castanho-clara à castanho-avermelhada, formando lesões que são facilmente visíveis em ambas as faces da folha. Plantas infectadas apresentam desfolha precoce, comprometendo a formação e o enchimento de vagens, reduzindo o peso final dos grãos (Consórcio Antiferrugem, 2011).

São relatados três tipos de reações decorrentes da infecção de *P. pachyrhizi* em plantas hospedeiras: TAN (lesões castanho-clara), RB (lesões castanho-escuro) e imunidade (sem sintomas visíveis da doença). A lesão TAN contém urédias esporulantes e reflete a suscetibilidade do hospedeiro, enquanto que em cultivares resistentes se observa as lesões RB com pouca ou nenhuma esporulação. A reação de imunidade se caracteriza pela ausência de sintomas macroscópicos nas folhas do hospedeiro, após a inoculação. Com o decorrer do desenvolvimento da doença a distinção entre reações TAN e RB podem se tornar menos óbvia (Bromfield, 1984).

O uso de fungicidas é atualmente o principal método de controle da

ferrugem asiática da soja, o que acarreta aumento nos custos de produção e possíveis impactos negativos no meio ambiente e na saúde humana. Além do controle químico, a adoção do vazio sanitário, o plantio no início da época recomendada, o uso de cultivares de ciclo precoce e de variedades resistentes nas regiões para as quais estão disponíveis são outras medidas recomendadas para o manejo da doença. A utilização de variedades resistentes a ferrugem é a melhor alternativa para o controle desta doença devido à redução dos custos de produção, a facilidade de manejo, e o menor impacto ambiental (Garcia *et al.*, 2008).

Já foram identificados pelo menos sete genes de resistência a *P. pachyrhizi* (*Rpp1*, *Rpp1b*, *Rpp1?*, *Rpp2*, *Rpp3*, *Rpp4* e *Rpp5*) em cinco locos distintos (Bromfield e Hartwig, 1980; Mclean e Byth, 1980; Hartwig e Bromfield, 1983; Hartwig, 1986; Hyten *et al.*, 2007; Calvo *et al.*, 2008; Silva *et al.*, 2008; Chakraborty *et al.*, 2009), porém, todas esses genes de resistência conferem resistência raça-específica (Meyer *et al.*, 2009). Em função dessa limitação dos genes *Rpp* disponíveis, existe grande interesse em novas estratégias de desenvolvimento de variedades de soja com resistência durável.

2.2 – Interação planta-patógeno

2.2.1 – Sistema imune vegetal

No contexto das interações planta-patógeno, já é bem estabelecido que patógenos empregam um complexo arsenal molecular para penetrar e colonizar os tecidos da planta hospedeira (Stergiopoulos & de Wit, 2009). Estas moléculas são usualmente denominadas efetores, e a elas são conferidas as funções de alterar a estrutura e função da célula hospedeira, para facilitar a infecção e suplantar as respostas de defesa da planta (Hogenhout *et al.*, 2009). Neste mesmo cenário, as plantas desenvolveram ao longo do processo coevolutivo um complexo sistema de defesa, que apresenta duas linhas principais. A primeira linha, que confere defesa basal contra todos os patógenos em potencial, é baseada no reconhecimento de características microbianas conservadas, denominadas padrões moleculares

associados a patógenos (PAMPs – *Pathogen Associated Molecular Patterns*). Receptores localizados na membrana plasmática reconhecem PAMPs e ativam as respostas de defesa da planta. Esta primeira linha de defesa que é denominada imunidade desencadeada por PAMPs ou PTI (*PAMP triggered immunity*) previne a colonização posterior do hospedeiro (Nurnberger *et al.*, 2004; Jones e Dangl, 2006; de Wit, 2007). Um dos PAMPs de fungos mais bem conhecidos é a quitina, um importante componente estrutural da parede celular, para o qual dois receptores, semelhantes a cinase do tipo LysM, foram caracterizados em arroz e em *Arabidopsis* (Kaku *et al.*, 2006; Miya *et al.*, 2007).

Durante o processo infeccioso os fitopatógenos secretam um arsenal de proteínas efetoras que são translocadas para o citoplasma vegetal e suprimem PTI, permitindo assim ao patógeno infectar seu hospedeiro e causar doença (Stergiopoulos e de Wit, 2009). A segunda linha de defesa baseia-se na percepção das proteínas efetoras por proteínas R e subsequente ativação de imunidade desencadeada por efetores ETI (*effector triggered immunity*). Este reconhecimento resulta em uma resposta de defesa rápida, denominada resposta de hipersensibilidade (HR). Isto provoca uma segunda onda de competição coevolutiva entre patógenos e plantas, durante a qual os patógenos respondem por mutação ou perda de efetores, ou por meio do desenvolvimento de novos efetores que podem evitar ou reprimir ETI, enquanto que as plantas desenvolvem novas proteínas R capazes de reconhecer estes novos efetores (Jones e Dangl, 2006; de Wit, 2007).

De acordo com a teoria gene-a-gene de Flor (Flor, 1942), para cada gene dominante de avirulência (*Avr*) no patógeno, existe um gene cognato de resistência (R) no hospedeiro, e a interação dos produtos desses dois genes leva à ativação de respostas de defesa do hospedeiro. A natureza física do reconhecimento Avr-R tem potencial impacto na evolução dessas proteínas. Até o momento, existem dois modelos hipotéticos que descrevem como proteínas efetoras do patógeno interagem com proteínas de resistência da planta. O primeiro modelo sugere uma interação direta receptor-ligante, e tem sido descrita para o patossistema arroz – brusone, onde Pita interage com Avr-Pita, no patossistema

Arabidopsis thaliana – *Ralstonia solanacearum*, onde RRS1 interage com a proteína de avirulência PopP2 e no patossistema tabaco – TMV (*Tobacco mosaic virus*), onde N interage com a replicase viral TMV (Bernoux *et al.*, 2008; Deslandes *et al.*, 2003; Jia *et al.*, 2000; Ueda *et al.*, 2006). No segundo modelo, proteínas R monitoram a presença da proteína Avr indiretamente, pela detecção de mudanças em outras proteínas da planta. A hipótese guarda tem sido demonstrada no patossistema *A. thaliana* – *Pseudomonas syringae*, onde RPM1, RPS2 e RPS5 reconhecem alterações nas proteínas RIN4 e PBS1 induzidas por efetores bacterianos. A interação direta das proteínas R com proteínas Avr impõem uma forte pressão de seleção sobre estas proteínas efetoras, contribuindo para seleção de proteínas que não sejam reconhecidas pelas proteínas R. Quando a região de reconhecimento de Avr não está relacionada com função do efector, é possível a seleção de mutações que perturbem o reconhecimento sem que haja prejuízo considerável na adaptabilidade do indivíduo mutante. Desta forma, proteínas R do tipo guarda, podem fornecer uma resistência mais estável a longo prazo, enquanto que proteínas R que interagem diretamente com proteínas Avr podem acelerar a evolução de novas fenótipos de virulência (Ellis e Dodds, 2003; van der Hoorn *et al.*, 2002).

Estudos moleculares conduzidos nos últimos anos demonstraram que os genes R codificam proteínas com domínios LRR (*leucine-rich repeats*) e NBS (*nucleotide-binding site*) que reconhecem direta ou indiretamente proteínas AVR produzidas no apoplasto ou transferidas para o interior da célula vegetal enquanto que as proteínas Avr, de modo geral, não apresentam características moleculares comuns (Stergiopoulos e de Wit, 2009).

2.2.2 – Efetores: identificação e caracterização

Os efetores de fitopatógenos são proteínas secretadas que interferem em diferentes componentes do sistema imune vegetal, incluindo enzimas apoplásticas (Song *et al.*, 2009), receptores de superfície celular (Meng *et al.*, 2006; Shan *et al.*, 2008), cascatas de transdução de sinal (de Torres *et al.*, 2006), maquinaria de

transcrição (Greenshields e Jones, 2008; Kay e Bonas, 2009), tráfego de vesículas (Schlumberger e Hardt, 2006), degradação de proteínas via proteassoma (Groll *et al.*, 2008). Logo, os efetores podem ser definidos como moléculas que alteram a estrutura e função dos componentes celulares do hospedeiro, facilitando a infecção e a manutenção do parasitismo e/ou que causam a ativação de respostas de defesa da planta quando são reconhecidos pelas proteínas R (Kamoun, 2009). Desde a clonagem e caracterização do primeiro gene *Avr* em 1984 (Staskawicz *et al.*, 1984), vários efetores de bactérias, nematóides, oomicetos e fungos biotróficos foram identificados e caracterizados (Arnold *et al.*, 2009; Elling *et al.*, 2009; Kamoun, 2006; Bailey *et al.*, 2011).

O transporte de proteínas efetoras do citoplasma do patógeno para interior da célula do hospedeiro é uma etapa fundamental do processo de patogênese. Muitas bactérias gram-negativas desenvolveram um sistema especializado de secreção de proteínas, o sistema de secreção do Tipo III (TTSS), que está envolvido na translocação de proteínas efetoras através da parede celular e membrana plasmática do hospedeiro, lançando-as diretamente no citoplasma vegetal. A identificação de efetores bacterianos foi acelerada dramaticamente nos últimos anos pelo uso combinado de estratégias computacionais e experimentais para identificar proteínas secretadas via TTSS (Greenberg e Vinatzer, 2003; Chang *et al.*, 2005; Vencato *et al.*, 2006). A fitobactéria *Pseudomonas syringae* pv. *tomato* DC3000 transloca cerca de 30 efetores diferentes via TTSS (Boch *et al.*, 2010). Porém, o sequenciamento do genoma de outras três estirpes de *P. syringae* (*P. syringae* pv. *phaseolicola* 1448A, *P. syringae* pv. *syringae* B728a, *P. syringae* pv. *tomato* T1), mostrou que cada uma delas apresenta um conjunto particular de efetores. Muitos destes efetores atuam de maneira redundante, de forma que a deleção de um destes genes tem um impacto insignificante na agressividade do patógeno (Cunnac *et al.*, 2009). Diversos efetores secretados pelo TTSS também têm sido identificados no gênero *Xanthomonas*. Em especial pode-se citar os efetores TAL, que são secretados durante interação de *Xanthomonas* com várias espécies de plantas. Estas proteínas são translocadas para o interior do núcleo do hospedeiro, aonde se ligam a sequências específicas no DNA e ativam a expressão

de genes da planta que contribuem para a infecção bacteriana (Boch *et al.*, 2010). Estas proteínas são de grande interesse para estudo, tanto devido ao seu papel na patogenicidade, quanto em aplicações biotecnológicas, dada a relativa facilidade de manipulação visando o reconhecimento de sequências específicas de DNA (Boch *et al.*, 2010). Proteínas semelhantes foram encontradas em *R. solanacearum* (Heuer *et al.*, 2007).

Evidências da transferência direta de proteínas para o citoplasma do hospedeiro durante infecção e o papel importante do haustório neste processo de translocação foram obtidas para ferrugens e oomicetos. Através de ensaios de imunofluorescência e microscopia eletrônica, a proteína Uf-RTP1p de *Uromyces fabae* foi detectada não só na matriz extrahaustorial, mas também no interior das células das plantas infectadas. Uma proteína homóloga a Uf-RTP1p foi identificada em *Uromyces striatus* (Us-RTP1p). A localização da Uf-RTP1p e Us-RTP1p dentro de células infectadas da planta foi confirmada usando quatro anticorpos policlonais independentes. Dependendo do estágio de desenvolvimento do haustório, Uf-RTP1p foi encontrado em quantidades crescentes nas células hospedeiras, incluindo o núcleo. A localização nuclear de RTP1p sugere papel importante desta proteína na manutenção da interação biotrófica (Kemen *et al.*, 2005). A translocação de proteínas para o citoplasma é corroborada pelos estudos de clonagem e caracterização das proteínas Avr567, AvrP123, AvrP4 e AvrM da ferrugem do linho (*Melampsora lini*) (Wang *et al.*, 2007). A expressão dessas proteínas no citoplasma vegetal induziu uma reação de hipersensibilidade dependente das proteínas citoplasmáticas codificadas pelos genes de resistência correspondentes do linho (*Linum usitatissimum*). A AvrP123 é uma proteína de 127 aminoácidos com similaridade a inibidores de serinoproteases do tipo *kazal*, sugerindo um possível papel dessa proteína na patogênese de *M. lini*, por meio da inibição de proteases do hospedeiro. No entanto, o mecanismo pelo qual estas proteínas atravessam a membrana extrahaustorial e se acumulam no citoplasma do hospedeiro permanece desconhecido.

Tradicionalmente, os efetores têm sido identificados por purificação bioquímica e análises genéticas, porém, com o advento da genômica, novas

estratégias têm surgido. A implementação da genômica funcional e a consolidação do conceito de que as proteínas efetoras são proteínas secretadas tem contribuído para a descoberta de efetores com o uso de ferramentas computacionais. A estratégia típica para triagem de genes que codificam efetores consiste em duas etapas principais: uso de ferramentas de bioinformática para identificar genes candidatos que cumpram uma lista de critérios específicos em banco de dados de sequências, seguida de análise e validação destes genes candidatos por ensaios funcionais, tais como a expressão em plantas para avaliação de suas atividades efetoras (Kamoun, 2006). A aplicação dessas estratégias bioquímicas, genéticas e computacionais, muitas vezes em combinações, têm sido utilizadas para a identificação de genes que codificam para efetores em oomicetos e fungos filamentosos (Rehmany *et al.*, 2005; Win *et al.*, 2007; Wang *et al.*, 2007; Link e Voegelé, 2008).

Oomicetos e fungos filamentosos, como outros organismos eucariotos, secretam proteínas através do sistema de secreção do tipo II, uma via de secreção tradicional que envolve o retículo endoplasmático e a exocitose de vesículas secretoras derivadas do complexo de Golgi (Griffiths e Simons, 1986). A maioria dos efetores de fungos e oomicetos, até então conhecidos, apresenta um peptídeo sinal na extremidade N-terminal (O'Connell e Panstruga, 2006; Allen *et al.*, 2004; Catanzarite *et al.*, 2007; Rehmany *et al.*, 2005; Dodds *et al.*, 2004). Ferramentas computacionais, particularmente o programa SignalP (Bendtsen *et al.*, 2004), podem ser utilizadas para prever a presença de peptídeo sinal em proteínas deduzidas de sequências abertas de leituras, ou ORFs (*Open Reading Frame*), com função desconhecida, com alto grau de precisão (Bendtsen *et al.*, 2004; Kamoun, 2006). Portanto, com o acúmulo de sequências genômicas e de cDNA em bancos de dados, listas de proteínas potencialmente secretadas candidatas podem ser facilmente geradas utilizando ferramentas de bioinformática.

A presença do sinal para secreção é suficiente apenas para a translocação até o ambiente extracelular. Embora alguns efetores apresentem funções no ambiente apoplástico, como os efetores Avr2 e Avr4 de *Cladosporium fulvum* (Joosten *et al.*, 1997; Shen e Jacobs-Lorena, 1999; Luderer *et al.*, 2002; van den

Burg *et al.*, 2006; Shabab *et al.*, 2008) e Six3 de *Fusarium oxysporum* (Rep *et al.*, 2004; Houterman *et al.*, 2009), muito outros provavelmente possuem alvos no citoplasma da célula vegetal, uma vez que são reconhecidos por proteínas de resistência (R) intracelulares da planta (Rehmany *et al.*, 2005; Armstrong *et al.*, 2005). Porém, o mecanismo envolvido na translocação desses efetores para o citoplasma vegetal ainda permanece desconhecido.

Recentemente, foram descobertos motivos protéicos conservados que estão diretamente envolvidos na translocação de proteínas de oomicetos para o interior da célula vegetal. Whisson *et al.* (2007) utilizando transformantes de *Phytophthora infestans*, demonstraram o envolvimento dos motivos RxLR-dEER na translocação de efetores de oomicetos para o citoplasma do hospedeiro. A ligação fosfatidilinositol-3-fosfato (PI3P) ao motivo RxLR-dEER e posterior endocitose do efector é um dos mecanismos bioquímicos propostos para a translocação mediada por este motivo (Kale *et al.*, 2010).

A identificação do motivo RxLR-dEER têm permitido a utilização de ferramentas de bioinformática na mineração de sequências genômicas de oomicetos visando a identificação *in silico* de efetores. Em *Phytophthora* e *Hyaloperonospora* foram identificados aproximadamente 700 e 150 candidatos a efetores, respectivamente (Jiang *et al.*, 2008; Win *et al.*, 2007). As proteínas CRN (*Crinklers*) de *P. infestans*, também apresentam um motivo conservado, LXLFLAK, envolvido na translocação desta classe de efetores, e o papel destas proteínas na indução de morte celular já foi verificado (Schornack *et al.*, 2010).

Godfrey *et al.* (2010), por meio da análise de uma biblioteca de cDNA de haustório de *Blumeria graminis*, fungo causador do oídio da cevada, verificou que a maioria dos genes altamente expressos codificavam para proteínas com peptídeo sinal na região N-terminal. Através de uma análise mais detalhada da região N-terminal, descobriu-se um motivo Y/F/WxC conservado em 107 proteínas deduzidas para este fungo. A mesma análise foi conduzida para *Puccinia graminis*, agente causal da ferrugem do colmo do trigo. Por meio da análise do genoma de *P. graminis* foram identificadas 178 proteínas deduzidas com sinal de secreção e com presença do motivo Y/F/WxC. Por meio de análises de RT-PCR

semi-quantitativo para 23 genes altamente representados na biblioteca de ESTs de *B. graminis*, Godfrey *et al.*, (2010) verificaram que os genes que codificam para proteínas Y/F/WxC são expressos predominantemente durante a sua formação do haustório. Por isso, acredita-se que proteínas Y/F/WxC desempenham funções importantes na formação haustorial e no estabelecimento da relação biotrófica, uma vez que 19% do transcriptoma de *B. graminis* é constituído por genes que codificam proteínas secretadas com o motivo Y/F/WxC.

Estas novas descobertas ilustram a complexidade dos secretomas de fitopatógenos e, embora componentes individuais desses secretomas estejam sendo revelados, o progresso ainda é limitado e há muito a ser descoberto. Uma caracterização mais extensa dos secretomas de fitopatógenos eucarióticos fornecerá novos cenários dos processos de infecção e da natureza dos efetores secretados. Diante disto, é crescente o interesse do estudo do secretoma de patógenos de plantas. Por meio de análises *in silico* de bibliotecas de cDNA, tem-se estudado o secretoma de diversos fitopatógenos, como os oomicetos *P. infestans* (Raffaele *et al.*, 2010), *H. arabidopsidis* (Cabral *et al.*, 2011) e fungos como *Magnaporthe grisea* (Dean *et al.*, 2005), *Melampsora lini* (Catanzariti *et al.*, 2006) e *M. larici-populina* (Joly *et al.*, 2010). O sequenciamento em larga escala e análises *in silico* permitiram a identificação e caracterização de efetores de *H. arabidopsidis*. A partir do sequenciamento do cDNA da interação *A. thaliana* - *H. arabidopsidis*, foram identificados 2164 genes derivados do patógeno. Destes, 198 codificam para proteínas secretadas, sendo que 75 delas são específicas desse patógeno e 6 foram específicas a um determinado isolado (Waco-9). O motivo RxLR foi verificado para vários efetores de *H. arabidopsidis*. A translocação da proteína RXLR29 utilizando o TTSS de *P. syringae* (Sohn *et al.*, 2007) revelou que este efector pode suprimir a imunidade desencadeada por PAMP, aumentando a suscetibilidade de *arabidopsis* a este patógeno (Cabral *et al.*, 2011).

O sistema de armadilha de sinal de secreção em leveduras (sistema YST, *Yeast Signal Secretion Trap*) é um método para identificação de genes que codificam proteínas que apresentam sinal de secreção, e tem sido utilizado com

grande sucesso na caracterização do secretoma de fungos e oomicetos (Link e Voegelé, 2008; Birch *et al.*, 2006; Lee *et al.*, 2006). O sistema consiste em um vetor de expressão que carrega um gene que codifica para a proteína invertase (SUC2) em que o códon de iniciação e a região que codifica o peptídeo sinal foram removidos. A invertase codificada pelo gene *SUC2* é uma sacarase que catalisa a quebra de sacarose em frutose e glicose. Neste vetor a transcrição da invertase é mediada pelo promotor ADH1, do gene que codifica a enzima álcool desidrogenase. Para que o gene da invertase seja transcrito e a invertase secretada, uma sequência (cDNA ou DNA genômico) contendo um códon iniciador e a região codificadora para o peptídeo sinal deve ser clonada entre o promotor ADH1 e o gene *SUC2*, *in frame* com a sequência desse gene. Os plasmídeos recombinantes contendo a sequência promotor-cDNA-SUC2, são utilizados para transformação de linhagem de levedura *S. cerevisiae* contendo uma deleção genômica no loco *SUC2*(Δ *suc2*) sendo, portanto, incapaz de secretar a enzima invertase. A seleção é realizada em meio contendo sacarose como única fonte de carbono. Assim, apenas os clones que apresentarem sequências de cDNA ou DNA genômico que codificam proteínas secretadas, são selecionados (Jacobs *et al.*, 1997).

Lee *et al.* (2006) utilizando o sistema YST identificaram 23 genes que codificam proteínas secretadas pelo oomiceto *P. infestans*. Apenas seis dos 23 genes identificados puderam ser anotados com base na similaridade com sequências de genes previamente caracterizados. Os transcritos de todos os genes identificados só foram detectados em folhas infectadas, não sendo observada a sua expressão em micélio proveniente de cultivo *in vitro*. Em continuação ao trabalho desenvolvido por Lee *et al.* (2006), Kelley *et al.* (2010) demonstraram que a proteína SNE1(PiE12) secretada por *P. infestans* suprime a HR induzida por PsojNIP e PiNPP1.1, proteínas que conhecidamente ativam as respostas de defesa da planta e causam a subsequente morte celular no tecido vegetal (Qutob *et al.*, 2002; Pemberton e Salmond, 2004; Kanneganti *et al.*, 2006). Também ficou demonstrado que a proteína SNE1 suprime a HR mediada pela interação das proteínas cognatas Avr/R de um amplo espectro de patossistemas, incluindo

oomiceto (Avr3a/R3a), bactéria (AvrPto/Pto), fungo (Avr9/Cf9) e vírus (CP/Rx2). Com base nestes resultados, foi formulada a hipótese de que as rotas de sinalização de defesa mediada pelos diferentes genes *R* convergem em um ponto comum que é suprimido pela proteína SNE1. Link & Voegelé (2008) por meio sistema YST, identificaram 62 genes do fungo *U. fabae*, que codificam proteínas secretadas pelo haustório e 42 genes que codificam proteínas secretadas em esporos germinados. Apenas quatro genes foram identificados em haustório e esporos germinados, indicando uma forte regulação fase-específica na secreção de proteínas. Para 39 sequências analisadas, foi observada similaridade a proteínas previamente identificadas, 28 das quais apresentaram similaridade a proteínas identificadas apenas entre os membros da ordem Uredinales, fato que foi interpretado como uma indicação de possíveis papéis dessas proteínas na virulência e especificidade de hospedeiro exclusivo das espécies causadoras de ferrugens.

O secretoma do fungo *P. pachyrhizi* ainda não foi caracterizado. Posada-Buitrago e Frederick (2005) visando o conhecimento da expressão de genes na fase de germinação dos uredósporos, construíram uma biblioteca de cDNA e compararam a sequência destes clones ESTs com aquelas existentes em banco de dados. Apesar de mais de 52% dos clones não mostrarem similaridades significantes às sequências contidas no banco de dados, 189 ESTs únicas foram identificadas com funções ligadas ao metabolismo primário, expressão gênica, crescimento e organização celular, divisão e sinalização celular e respostas de defesa do organismo. Ainda não foram identificadas as proteínas efetoras (Avr) reconhecidas pelas proteínas codificadas pelos genes *Rpp*, ou efetuada uma análise detalhada dos genes de *P. pachyrhizi* que codificam proteínas secretadas durante a sua interação com a soja. Devido à escassez de informação em torno dos efetores moleculares de *P. pachyrhizi* envolvidos na interação soja-ferrugem, a identificação de proteínas secretadas será de grande auxílio para a elucidação do processo de infecção causado por este patógeno.

3.0 – Objetivos

3.1 – Objetivo geral

Este estudo teve por objetivo identificar e caracterizar genes que codificam proteínas secretadas por *P. pachyrhizi* durante a interação com a soja, visto que a caracterização destes genes pode auxiliar no entendimento do processo de infecção do patógeno.

3.2 – Objetivos específicos

- i – Identificar genes que codificam para proteínas secretadas, por meio da triagem de uma biblioteca de secreção em levedura;
- ii – Selecionar através de ferramentas de bioinformática dentre os clones positivos do sistema YST, aqueles que apresentam cDNA de origem fúngica;
- iii – Obter clones com as sequências completas (cDNA) dos genes em estudo;
- iv – Confirmar a secreção das proteínas codificadas pelas ORFs completas.

4.0 – Materiais e Métodos

4.1 – Material vegetal, cultivo das plantas e inoculação

Após 30 dias de cultivo, plantas de soja da variedade Conquista foram inoculadas com uma suspensão de $5,0 \times 10^5$ uredósporos/mL do isolado PPUFV02 de *P. pachyrhizi* em água destilada + Tween 20 (0,01%), e mantidas em câmara de nevoeiro, a 25°C no escuro, durante 24 horas. Em seguida, as plantas foram transferidas para a casa de vegetação, onde permaneceram durante todo o período de coleta.

4.2 – Construção da biblioteca YST e obtenção de DNA plasmidial

O RNA total foi extraído de folhas de soja coletadas às 24, 48, 72, 96, 120 horas após a inoculação do isolado PPUFV-02, utilizando-se o protocolo do RNA *isolation reagent* (Invitrogen). O RNA mensageiro foi purificado e a primeira fita de cDNA foi sintetizada utilizando um oligonucleotídeo iniciador contendo sítio para *NotI*. Após a síntese da segunda fita, os fragmentos de cDNA foram ligados a adaptadores *EcoRI*, e submetidos a digestão com *NotI*. Os fragmentos de cDNA com tamanhos entre 0,3 - 1,0 Kb, foram separados por eletroforese em gel de agarose, purificados e ligados aleatoriamente em uma mistura dos três vetores pYST0-2 (**Figura 4**) previamente digeridos com *EcoRI* e *NotI*. O produto da reação de ligação foi transformado por eletroporação em *Escherichia coli* EC100.

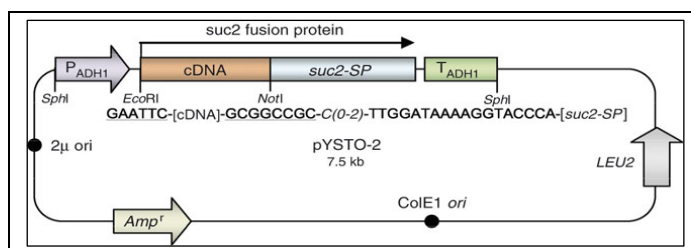


Figura 4: Representação esquemática do pYST(0-1). O vetor foi utilizado para a clonagem das ORFs selecionadas, amplificadas a partir do cDNA, para comprovação da secreção pelo sistema YST. *Amp^r* – marca de resistência à ampicilina. LEU2 - gene que codifica uma enzima envolvida na biossíntese de leucina. ColE1 *ori* – origem de replicação em bactéria. 2 μ *ori* – origem de replicação em levedura. P_{ADH1} – promotor do gene que codifica a enzima álcool desidrogenase. *suc2-SP* – gene da invertase sem a região que codifica para o sinal de secreção. T_{ADH} – região de terminação do gene que codifica a enzima álcool desidrogenase (Lee, *et al.*, 2006)

Os transformantes foram selecionados em meio LB contendo ampicilina 100 μ g/mL, e foi realizada uma purificação em massa do DNA plasmidial, utilizando o QIAGEN *Plasmid Midi Kit*, de acordo com as instruções do fabricante (QIAGEN).

4.3 – Transformação de levedura pelo método AcLi/PEG e triagem da biblioteca YST

Os plasmídeos da biblioteca de cDNA foram transformados em levedura (*Saccharomyces cerevisiae*) estirpe BY4742 (MAT α , Δ *suc2*, *his3 Δ* , *leu2 Δ* , *lys Δ* , *ura3 Δ*), de acordo com o protocolo descrito por (Gietz e Woods 2002). Para o preparo de células competentes foram inoculados 5mL meio YPD com uma colônia isolada da estirpe de levedura a ser transformada, que foi incubada *overnight* a 30°C sob agitação de 200rpm. Em seguida, 45mL do meio YST foi inoculado e incubado a 30°C até atingir 10⁷ células/mL (OD₆₀₀ \approx 0,3 – 0,5). A suspensão celular foi transferida para tubos de 50mL, que foram centrifugados a 4000g por 5 minutos. O sobrenadante foi descartado e as células foram ressuspensas em 25mL de água estéril. Uma nova centrifugação foi realizada, a água foi removida e foi adicionado 1mL de AcLi (acetato de lítio) 100mM. O volume total foi transferido para tubos de microcentrífuga de 1,5mL. Os tubos foram centrifugados por 15 segundos e o AcLi removido com o auxílio de uma pipeta. As células competentes, por fim, foram ressuspensas em 400 μ L de AcLi 100mM.

Para cada 50 μ L de célula competente foram adicionados os reagentes na seguinte ordem: 240 μ L de PEG 50% (polietileno glicol), 36 μ L de AcLi 1M, 50

μL de esperma de salmão ($2 \mu\text{g}/\mu\text{L}$), $34 \mu\text{L}$ de água deionizada estéril contendo $1 \mu\text{g}$ de DNA. Após a homogeneização da mistura por pipetagem, a mesma foi incubada por 60 minutos a 30°C , com homogeneização a cada 15 min com auxílio de vortex. Para cada placa de petri de 150 mm de diâmetro contendo o meio seletivo, foram plaqueados $160 \mu\text{L}$ da mistura de transformação. A seleção foi realizada em meio YPS sólido (1% extrato de levedura, 2% peptona e 2% sacarose), as placas foram incubadas por sete dias em estufa, a 30°C . As colônias positivas foram repicadas em poços individuais de placas de estoque de fundo “U” em meio YST contendo glicerol 25% e as placas foram deixadas a 30°C , 280 rpm, por 24 horas e depois estocadas a -80°C .

4.4 – Extração plasmidial de levedura

Para a purificação do DNA plasmidial das colônias positivas, $1,25 \text{ mL}$ de meio YPS contendo $100 \mu\text{L}/\text{mL}$ de gentamicina foram inoculados com $10 \mu\text{L}$ do estoque de levedura em placas de fundo “U”. As placas foram mantidas por 24h a 28°C sob agitação de 280 rpm e em seguida foram centrifugadas a $4200 \times g$ durante 15 minutos e o sobrenadante foi descartado. As placas foram mantidas no gelo e as células sedimentadas foram ressuspensas em $140 \mu\text{L}$ da solução 1 (1,0 M sorbitol, 10 mM Tris-HCl pH7.5, 10 mM CaCl_2 , 10 mM DTT). Em seguida foi realizada uma centrifugação, 4200rpm, 25°C , durante 10 minutos e as células foram ressuspensas em $140 \mu\text{L}$ da solução 1 + enzima: 1,0 M sorbitol, 10 mM Tris-HCl pH7.5, 10 mM CaCl_2 , 10 mM DTT, RNase ($100 \mu\text{g}/\text{mL}$) + Glucanex ($40 \text{ mg}/\text{mL}$). As placas foram agitadas por 5 minutos, 280 rpm, 37°C , e em seguida foram agitadas por uma hora, 200 rpm, 37°C . Após este intervalo foi adicionado $140 \mu\text{L}$ de tampão de lise (0,2N NaOH e 1% SDS). As placas foram seladas, invertidas 20 vezes e pulsadas na centrífuga a 3700rpm durante um minuto para remover o lisado do plástico. A lise foi realizada à temperatura ambiente por no máximo 5 minutos. Em seguida foram adicionados $140 \mu\text{L}$ de tampão de neutralização gelado (3M acetato de potássio, pH 5,5). As placas foram seladas, invertidas 20 vezes e incubadas no gelo durante 10 minutos. Em seguida foi

realizada uma centrifugação durante 25 minutos a 4200xg. Alíquotas de 280µL do sobrenadante foram transferidos para uma placa filtro de 0,45µm de acetato de celulose, empilhada em cima da placa coletora contendo 330µL de isopropanol 100%. As placas empilhadas foram centrifugadas a 3000xg, durante 5 minutos. Então foram adicionados 330µL de isopropanol 100%, gelado, e as placas foram invertidas 10 vezes, incubadas a -80°C por 20 minutos e centrifugadas a 4200g por 30 minutos. O precipitado foi lavado adicionando 250µL de etanol 80% em cada poço, e as placas foram centrifugadas a 4000rpm durante 5 minutos, o sobrenadante foi descartado e a lavagem foi repetida. Em seguida foi realizada uma centrifugação a 4200xg por 30 minutos. As placas ficaram em repouso cobertas com papel de filtro por aproximadamente 20 minutos na capela de fluxo laminar, para secagem dos precipitados. Foram adicionados 30µL de água Mili-Q autoclavada por poço, e as placas foram centrifugadas por um minuto a 900 rpm e os precipitados foram ressuspendidos *overnight* a 4°C. As placas foram estocadas a -20°C.

4.5 – Reação de PCR, purificação do produto de PCR e sequenciamento dos insertos dos clones positivos

Para a obtenção de produto de PCR (*Polymerase Chain Reaction*) em quantidade suficiente para purificação e posterior sequenciamento, foram realizadas reações de PCR com um volume final de 40µL por amostra. A solução para cada reação era formada por 2,0µL da amostra de DNA, 4,0µL de tampão 10x, 3,2µL de dNTP 2,5mM, 1,0µL *primer* F, 1,0µL *primer* R, 0,2µL *Taq*, 28,6µL de água Mili-Q autoclavada. As reações foram realizadas no termociclador *GeneAmp® PCR System 9700 – Applied Biosystems*, em placa de 96 poços, adequadas ao equipamento, empregando-se as seguintes condições de amplificação: 95°C – 30", (95°C – 30", 52°C – 30", 72°C – 1'30")x42, 72°C – 7', 4°C - ∞. Após realizada a reação de PCR, o volume total de cada poço foi transferido para placas para purificação *MultiScreen® PCR (Millipore)* e submetidas a vácuo através do equipamento *MultiScreen® Resist Vacuum*

Manifold (Millipore), sendo a seguir realizada uma lavagem com 200µL de água Mili-Q autoclavada, o DNA foi ressuspensionado em 45µL de água Mili-Q autoclavada. A concentração final do DNA foi aferida em gel de agarose 1% e as amostras foram sequenciadas.

As reações de sequenciamento foram realizadas com o kit *DYEnamic™ ET Dye Terminators* (GE Healthcare, Freiburg, Germany). Foram utilizados nestas reações 100 a 150ng de produto da PCR, 0,5µM do oligonucleotídeo Y5 e/ou pYST-R1, 2,0 µL de *DYEnamic™ ET Dye Terminator Sequencing Pre-mix* em volume final de 5 µL. As condições da PCR foram: (95°C -10", 50°C - 5" , 60°C – 2')x35. Em seguida, o DNA foi precipitado adicionando-se 27,5µL de etanol absoluto e acetato de amônio para uma concentração final de 0,75M. Após 10 minutos a temperatura ambiente, as reações foram centrifugadas por 45 minutos a 3700 rpm. O DNA foi então lavado com 100µL de etanol 70%, centrifugado a 3700 por 10 minutos e deixado secar por 10 minutos. Após a precipitação, o DNA foi ressuspensionado em 5µL de *Loading Buffer* (GE Healthcare). A leitura das reações foram realizadas no sequenciador automático MegaBACE™1000 de 96 capilares (GE Healthcare) no Laboratório de Genômica – BIOAGRO-UFV.

4.6 – Análises de bioinformática

Os eletroferogramas gerados foram analisados pelo pacote phred/phrap/consed (www.phrap.org). O programa phred foi inicialmente utilizado para selecionar as sequências que apresentavam qualidade phred > 20. Em seguida, o programa cross-match foi utilizado para marcar as sequências correspondentes ao vetor e um *script* perl foi utilizado para remover estas sequências. Por meio de um *script* Perl (*Practical Extraction and Report Language*) foram selecionadas dentre as sequências restantes apenas aquelas com mais de 70 nucleotídeos, que foram alinhadas utilizando o programa CAP3 (<http://pbil.univ-lyon1.fr/cap3.php>). Os contíguos e singletos formados foram inicialmente comparados com sequências do genoma de soja (<http://soybase.org/SequenceIntro.php>) depositadas no Soybase (cobertura 8X) e

com 1468417 ESTs de soja depositados no GenBank do *National Center for Biotechnology Information* (NCBI - <http://www.ncbi.nlm.nih.gov>) por meio do algoritmo *BlastN*. Apenas *hits* com o *e-value* menor que 10^{-3} foram considerados similares. O resultado do *BlastN* foi utilizado para triar os ESTs quanto à origem fúngica. As sequências que não apresentavam similaridade com sequências de soja depositadas no Soybase/GenBank foram comparadas por meio do algoritmo *BlastX* com o banco de proteínas não redundantes do NCBI, para avaliar a similaridade com proteínas já anotadas. Para reforçar a origem fúngica das sequências *no hits* a soja, estas foram contrastadas ao rascunho (*draft*) do genoma de *P. pachyrhizi* (cobertura 1X), depositados no banco de sequências TRACE/NCBI (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi>).

As sequências *no hit* foram comparadas a 34394 ESTs de esporos germinados de *P. pachyrhizi* depositados no banco de nucleotídeos do NCBI, 14133 ESTs da interação soja-*P. pachyrhizi* depositados no NCBI e com 6846 ESTs da interação soja-*P. pachyrhizi* depositadas do banco de dados interno do Laboratório de Genômica. Estas comparações foram efetuadas para confirmar a origem fúngica desses cDNAs e também para obter as sequências completas dos genes identificados

A predição das sequências abertas de leitura (ORFs) para os contíguos e singletos de suposta origem fúngica foi realizada através do programa ORFfinder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) e de *scripts* Perl que geram arquivos contendo a tradução de cada transcrito nos três possíveis quadros de leitura positivos e que selecionam as ORFs que codificam em sequências polipeptídicas com mais de 30 aminoácidos.

As proteínas deduzidas selecionadas foram a seguir submetidas à predição de sequências codificadoras de sinal de exportação celular e de hélices transmembrana pelos programas SignalP, Targetp (<http://www.cbs.dtu.dk/services/index.php>) e Phobius (<http://phobius.sbc.su.se/>). A composição de aminoácidos das proteínas deduzidas também foi analisada, utilizando o programa ProtParam (<http://ca.expasy.org/tools/protparam.html>) e os possíveis sítios de formação de pontes dissulfeto foram obtidos utilizando o

programa DISULFIND (<http://disulfind.dsi.unifi.it>). A busca por motivos protéicos conservados entre os genes identificados foi verificada pelo programa MEME (http://meme.sdsc.edu/meme4_6_1/cgi-bin/meme.cgi). As sequências de aminoácidos para as quais foram encontradas similaridades entre si foram alinhadas utilizando-se o programa ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). As sequências protéicas também foram comparadas com proteínas deduzidas a partir do genoma de *Puccinia graminis tritici*, constituído de 20.566 sequências, disponível no banco de dados de *Puccinia* (http://www.broadinstitute.org/cgi-bin/annotation/fgi/blast_page.cgi).

Para os genes de origem fúngica, cujas sequências estavam completas e as ORFs preditas não apresentavam similaridade com nenhuma proteína depositada no banco de dados, foram desenhados *primers* para a clonagem no vetor pYST1 para confirmação da secreção do peptídeo codificado. O procedimento adotado foi semelhante ao descrito no item 4.2, e foi utilizado o vetor pYST-1, por permitir a clonagem *in frame* dos genes completos com o gene *SUC2*, sem o peptídeo sinal.

5.0 – Resultados

A partir da triagem de aproximadamente 10^6 transformantes foram identificados 2054 clones positivos por meio do sistema armadilha de secreção em levedura. A partir desses clones foram obtidas 1777 sequências da extremidade 5', com qualidade phred > 20 e com mais de 70 nucleotídeos. Por meio do programa

CAP3 foi realizado o alinhamento e a montagem das sequências redundantes. Foram obtidas 161 sequências únicas, que correspondem a 95 contíguos e 66 singletos. Estas sequências foram comparadas com o banco de ESTs (*expressed sequence tags*) de soja depositadas no GenBank/NCBI usando o algoritmo *BlastN* e ao banco de proteínas não redundantes do NCBI para obter informações a cerca das funções desencadeadas por essas proteínas. Das 161 sequências analisadas, 74 (45,9%) apresentaram similaridade com genes de soja (**Anexo 1**), enquanto 87 (54,1%) sequências não apresentaram similaridade com ESTs de soja, classificadas como *no hits*. Como estas 87 sequências *no hits* poderiam representar sequências de genes de soja ainda não caracterizados, as sequências foram comparadas com as sequências do genoma de soja disponíveis no *Soybase* (cobertura 8X) visando identificar as sequências que seriam, de fato, de origem fúngica. Após essa análise restaram 64 sequências *no hits*, que representam potenciais genes de origem fúngica. Para avaliar se esses genes *no hits* codificam para alguma proteína com função já descrita, essas sequências foram comparadas com proteínas depositadas no banco de proteínas não redundantes do NCBI por meio do algoritmo *BlastX*. Apenas cinco sequências codificam proteínas que apresentam similaridade com alguma proteína anotada (**Tabela 1**). Porém, dentre estas cinco proteínas, apenas aquela codificada pelo singleto YST_003_04_A05 apresentou similaridade com uma proteína de função descrita, uma quitinase, cujo número de acesso é AAP42832.

Tabela 1: Similaridade de proteínas codificadas por singletos e contíguos *no hit* com sequências de proteínas depositadas no NCBI, detectadas com a utilização do algoritmo *BlastX*.

Contíguos / Clone	n° ESTs	Acesso / NCBI	Organismo	Descrição	Função	e-value
Contig10 / Contig17 / Contig72	52	XP_002956216	<i>Volvox carteri</i>	unknown	Desconhecida	9,00E-05
Contig49	6	XP_001615225	<i>Plasmodium vivax</i>	unknown	Desconhecida	4,00E-05
Contig51 / 60	17	XP_002648161	<i>Caenorhabditis briggsae</i>	unknown	Desconhecida	6,00E-05
YST_003_01_A03	1	XP_003307688	<i>Puccinia graminis</i>	unknown	Desconhecida	1,00E-38
YST_003_04_A05	1	AAP42832	<i>Puccinia triticina</i>	chitinase	Expansão parede celular	1,00E-34

Um sumário representativo da rotina (*pipeline*) utilizado na anotação das sequências obtidas a partir da triagem da biblioteca YST é apresentado na **Figura 5**.

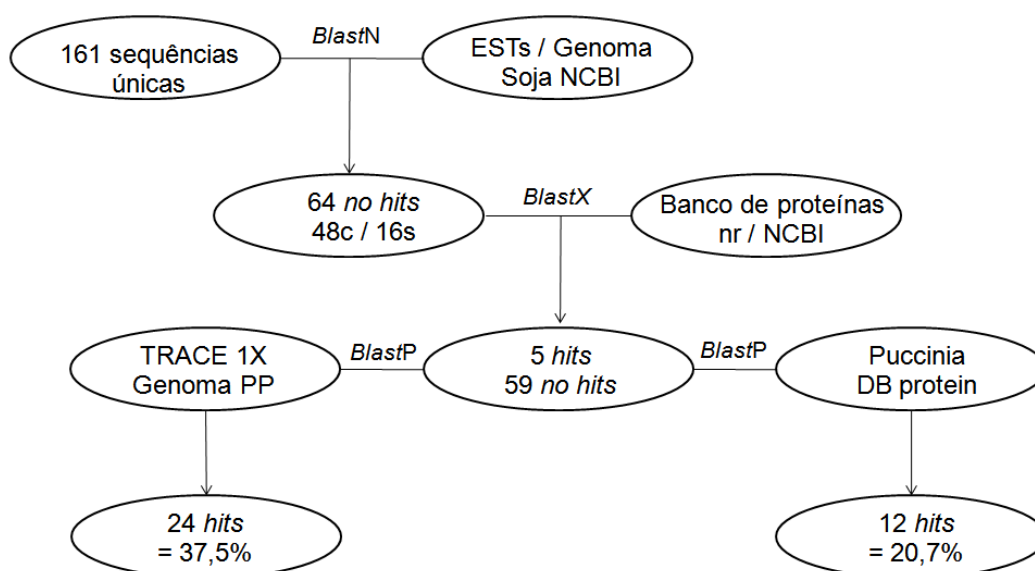


Figura 5: Fluxograma da estratégia utilizada para identificação de potenciais genes de *P. pachyrhizi* que codificam proteínas secretadas a partir da triagem de uma biblioteca YST da interação *P. pachyrhizi* – *G. max*. EST - *expressed sequence tag*. C – contíguos. S – singletos. nr – não redundante. PP – *P. pachyrhizi*. DB – *Database*.

A partir da sequência dos 64 genes *no hits* selecionados foram deduzidas ORFs (*open reading frame*) nas três fases de leitura possíveis através de um *script* Perl adaptado de Tisdall (2001). A tradução das sequências resultou na identificação de 89 ORFs com mais de 30 aminoácidos. Estas foram analisadas quanto a presença de sinal de secreção por meio do algoritmo SignalP e 55 ORFs codificam proteínas com predição positiva para secreção, com valor-D $\geq 0,55$, Sprob $\geq 0,65$. Observou-se que as 84 ORFs correspondiam a 58 genes *no hit*, e que para três deles (Contig13, Contig71, Contig91) a predição para secreção foi negativa (**Tabela 2**). As proteínas codificadas pelas 58 ORFs deduzidas foram

subsequentemente comparadas ao *Puccinia Database Protein* por meio do algoritmo *BlastP*. Observou-se a similaridade das proteínas codificadas por 12 ORFs deduzidas com proteínas de *Puccinia* já depositadas (**Tabela 3**).

Para tentar demonstrar por análise *in silico* que os 64 genes *no hits* eram de origem fúngica, as sequências foram comparadas com o rascunho do genoma (*draft*) de *P. pachyrhizi* (cobertura 1X) http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=TraceArchive&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch, por meio do algoritmo *BlastN* resultando em 24 sequências com similaridade com sequências depositadas (**Tabela 3**).

Tabela 2: Resultados da análise comparativa dos 64 genes *no hit* identificados com sequências depositadas nos bancos de dados do NCBI/Soybase por meio dos algoritmos *BlastX* e *BlastN* e da predição de sinal de secreção para as proteínas deduzidas a partir das ORFs dos genes *no hit* utilizando os algoritmos SignalP, Phobius e TargetP .

Contíguo / Clone	n° ESTs	NCBI / <i>BlastN</i> Soybase / <i>BlastX</i> nr			evalue	SignalP					Phobius	TargetP	
		Acesso / NCBI	Descrição	Função predita		Tam.	NN		HMM			Loc	RC
							Valor D	Pred.	Sprob	Pred.			
Contig3	4	-	no hit	-	-	24	820	Y	994	Y	0	S	1
Contig5	3	-	no hit	-	-	24	879	Y	998	Y	0	S	1
Contig8 / Contig24	235	-	no hit	-	-	20	855	Y	976	Y	0	S	1
Contig9	2	-	no hit	-	-	23	881	Y	995	Y	0	S	1
Contig10 / Contig17 / Contig72	52	XP_002956216	Unknown	Desconhecida	9e-05	20	898	Y	1000	Y	0	S	1
Contig12 / Contig16	264	-	no hit	-	-	19	882	Y	999	Y	0	S	1
Contig13	3	-	no hit	-	-	27	430	N	0	N	0	-	2
Contig15	2	-	no hit	-	-	23	849	Y	940	Y	0	S	1
Contig19	23	-	no hit	-	-	20	713	Y	968	Y	0	S	2
Contig20	7	-	no hit	-	-	23	782	Y	1000	Y	0	S	1
Contig21	2	-	no hit	-	-	18	800	Y	999	Y	0	S	2
Contig22	37	-	no hit	-	-	19	864	Y	999	Y	0	S	1
Contig25	2	-	no hit	-	-	24	728	Y	911	Y	0	S	1
Contig30	2	-	no hit	-	-	18	813	Y	999	Y	1	S	1
Contig32 Contig68	4	-	no hit	-	-	21	853	Y	994	Y	0	S	1
Contig36	2	-	no hit	-	-	26	795	Y	998	Y	0	S	1
Contig37	37	-	no hit	-	-	23	857	Y	999	Y	0	S	1
Contig38	2	-	no hit	-	-	20	896	Y	999	Y	0	S	1
Contig39	3	-	no hit	-	-	20	892	Y	1000	Y	0	S	1
Contig42	11	-	no hit	-	-	29	783	Y	915	Y	0	S	1
Contig44	4	-	no hit	-	-	21	841	Y	999	Y	0	S	1
Contig45 / 84	5	-	no hit	-	-	19	816	Y	999	Y	0	S	1
Contig47	39	-	no hit	-	-	21	755	Y	945	Y	0	S	1
Contig49	6	XP_001615225	Unknown	Desconhecida	4e-05	20	896	Y	1000	Y	0	S	1
Contig51 / 60	17	XP_002648161	unknown	Desconhecida	6e-05	20	928	Y	1000	Y	0	S	1

nr – banco de proteínas não redundante; Tam. – tamanho do sinal de secreção; D-score – Média da estimativa do ponto de clivagem do peptídeo sinal e da média dos valores de probabilidade da região atribuída ao peptídeo sinal, discrimina entre a região do peptídeo sinal do restante da proteína; Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q); Loc – Predição da localização celular, (M) mitocondria, (S) peptídeo sinal e – outro; RC – “Coeficiente de confiabilidade”, dado pela razão entre o maior e o segundo maior valor de predição de localização. 1 – 1-0,8; 2 – 0,8-0,6; sucessivamente até o valor 5; Y – Predição positiva de peptídeo sinal pelo SignalP; N – Predição negativa de peptídeo sinal pelo SignalP; S – Predição de secreção pelo TargetP; M – Predição de localização mitocondrial pelo TargetP. (-) não determinado ou sem informação disponível.

Tabela 2 cont.: Resultados da análise comparativa dos 64 genes *no hit* identificados com sequências depositadas nos bancos de dados do NCBI/Soybase por meio dos algoritmos *BlastX* e *BlastN* e da predição de sinal de secreção para as proteínas deduzidas a partir das ORFs dos genes *no hit* utilizando os algoritmos SignalP, Phobius e TargetP

Contig53	4	-	no hit	-	-	-	-	-	-	-	-	-	-	1
Contig56 / 57	34	-	no hit	-	-	20	890	Y	1000	Y	0	S	1	
Contig58	55	-	no hit	-	-	21	913	Y	993	Y	0	S	1	
Contig59	2	-	no hit	-	-	35	696	Y	695	Y	1	S	1	
Contig61	3	-	no hit	-	-	19	814	Y	998	Y	0	S	1	
Contig66	3	-	no hit	-	-	20	847	Y	986	Y	0	S	1	
Contig69	3	-	no hit	-	-	23	870	Y	994	Y	0	S	1	
Contig70	3	-	no hit	-	-	22	852	Y	998	Y	0	S	1	
Contig71	3	-	no hit	-	-	23	139	N	0	N	5	-	3	
Contig73	26	-	no hit	-	-	21	936	Y	1000	Y	0	S	1	
Contig75	10	-	no hit	-	-	21	888	Y	986	Y	0	S	1	
Contig76	10	-	no hit	-	-	20	851	Y	994	Y	0	S	1	
Contig79	3	-	no hit	-	-	-	-	-	-	-	-	-	1	
Contig83	3	-	no hit	-	-	19	618	Y	899	Y	1	S	4	
Contig86	2	-	no hit	-	-	25	941	Y	1000	Y	0	S	1	
Contig90	23	-	no hit	-	-	20	908	Y	1000	Y	0	S	1	
Contig91	3	-	no hit	-	-	21	621	Y	2	N	0	M	4	
Contig92	19	-	no hit	-	-	21	859	Y	996	Y	0	S	1	
Contig93	4	-	no hit	-	-	25	724	Y	992	Y	0	S	1	
Contig94	4	-	no hit	-	-	19	891	Y	999	Y	0	S	1	
Contig101	6	-	no hit	-	-	20	905	Y	999	Y	0	S	1	
Contig102	4	-	no hit	-	-	22	770	Y	999	Y	0	S	1	
Contig103	2	-	-	-	-	28	834	Y	995	Y	0	S	1	
YST_003_02_B04	1	-	no hit	-	-	20	871	Y	999	Y	0	S	1	
YST_003_01_A03	1	XP_003307688	Unknown	Desconhecida	1e-38	23	910	Y	1000	Y	0	S	2	
YST_004_08_A05	1	-	no hit	-	-	20	919	Y	1000	Y	0	S	1	
YST_004_05_D02	1	-	no hit	-	-	20	947	Y	1000	Y	0	S	1	
YST_003_04_A05	1	AAP42832	Chitinase	Expansão parede celular	1e-34	24	865	Y	999	Y	0	S	1	

nr – banco de proteínas não redundante; Tam. – tamanho do sinal de secreção; D-score – Média da estimativa do ponto de clivagem do peptídeo sinal e da média dos valores de probabilidade da região atribuída ao peptídeo sinal, discrimina entre a região do peptídeo sinal do restante da proteína; Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q); Loc – Predição da localização celular, (M) mitocondria, (S) peptídeo sinal e – outro; RC – “Coeficiente de confiabilidade”, dado pela razão entre o maior e o segundo maior valor de predição de localização. 1 – 1-0,8; 2 – 0,8-0,6; sucessivamente até o valor 5; Y – Predição positiva de peptídeo sinal pelo SignalP; N – Predição negativa de peptídeo sinal pelo SignalP; S – Predição de secreção pelo TargetP; M – Predição de localização mitocondrial pelo TargetP. (-) não determinado ou sem informação disponível.

Tabela 2 cont.: Resultados da análise comparativa dos 64 genes *no hit* identificados com sequências depositadas nos bancos de dados do NCBI/Soybase por meio dos algoritmos *BlastX* e *BlastN* e da predição de sinal de secreção para as proteínas deduzidas a partir das ORFs dos genes *no hit* utilizando os algoritmos SignalP, Phobius e TargetP

YST_004_01_B11	1	-	no hit	-	-	23	548	Y	869	Y	1	S	5
YST_003_03_G04	1	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_004_14_E09	1	-	no hit	-	-	24	896	Y	996	Y	0	S	1
YST_003_03_D11	1	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_004_01_D01	1	-	no hit	-	-	25	550	Y	886	Y	0	S	2
YST_004_07_C04	1	-	no hit	-	-	20	887	Y	1000	Y	1	S	1
YST_004_12_G08	1	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_003_03_H11	1	-	no hit	-	-	21	912	Y	993	Y	0	S	1
YST_004_10_E09	1	-	no hit	-	-	27	813	Y	999	Y	0	S	1
YST_003_04_G09	1	-	no hit	-	-	24	816	Y	997	Y	0	S	1
YST_004_02_C05	1	-	no hit	-	-	-	-	-	-	-	--	-	-

nr – banco de proteínas não redundante; Tam. – tamanho do sinal de secreção; D-score – Média da estimativa do ponto de clivagem do peptídeo sinal e da média dos valores de probabilidade da região atribuída ao peptídeo sinal, discrimina entre a região do peptídeo sinal do restante da proteína; Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q); Loc – Predição da localização celular, (M) mitocondria, (S) peptídeo sinal e – outro; RC – “Coeficiente de confiabilidade”, dado pela razão entre o maior e o segundo maior valor de predição de localização. 1 – 1-0,8; 2 – 0,8-0,6; sucessivamente até o valor 5; Y – Predição positiva de peptídeo sinal pelo SignalP; N – Predição negativa de peptídeo sinal pelo SignalP; S – Predição de secreção pelo TargetP; M – Predição de localização mitocondrial pelo TargetP. (-) não determinado ou sem informação disponível.

Tabela 3: Comparação dos genes selecionados com anotações do genoma de *P. pachyrhizi* por meio do algoritmo *BlastN* e análises de similaridade entre as proteínas deduzidas por estes genes, e proteínas deduzidas a partir do genoma de *Puccinia graminis tritici*, por meio do algoritmo *BlastP*.

Contíguo / Clone	n° ESTs	NCBI / Trace		Puccinia Database Protein		
		Acesso	E-value	Acesso	E-value	Função
Contig3	4	gnl_ti_452553078	9e-99	PTTG_02372	6e-07	Desconhecida
Contig5	3	gnl_ti_453233193	3e-55			
Contig8 / Contig24	235	gnl_ti_452180878	2e-22			
Contig9	2	gnl_ti_452501395	3e-13	PTTG_04588	3e-18	Desconhecida
Contig10 / Contig17 / Contig72	52	gnl_ti_452196760	1e-126			
Contig12 / Contig16	264	gnl_ti_453328477	4e-38			
Contig13	3			PTTG_00747	9e-08	Desconhecida
Contig15	2			PGTG_14060	1e-14	Desconhecida
Contig19	23			PGTG_07812	4e-08	Desconhecida
Contig20	7			PGTG_13910	8e-11	Desconhecida
Contig21	2			PTTG_06013	2e-05	Desconhecida
Contig22	37					
Contig25	2			PGTG_13504	7e-36	Desconhecida
Contig30	2			PGTG_10900	2e-24	Desconhecida
Contig32 Contig68	4	gnl_ti_453325403	3e-91			
Contig36	2					
Contig37	37	gnl_ti_452715553	1e-154			
Contig38	2					
Contig39	3					
Contig42	11					
Contig44	4			PTTG_07924	3e-42	Desconhecida
Contig45 / 84	5					
Contig47	39					
Contig49	6	gnl_ti_453268421	1e-104			

Tabela 3 cont.: Comparação dos genes selecionados com anotações do genoma de *P. pachyrhizi* por meio do algoritmo *BlastN* e análises de similaridade entre as proteínas deduzidas por estes genes, e proteínas deduzidas a partir do genoma de *Puccinia graminis tritici*, por meio do algoritmo *BlastP*.

Contig51 / 60	17	gnl_ti_453028962	9e-37			
Contig53	4					
Contig56 / 57	34	gnl_ti_452403809	1e-136			
Contig58	55					
Contig59	2	gnl_ti_452403809	1e-149			
Contig61	3	gnl_ti_452107387	2e-22			
Contig66	3					
Contig69	3					
Contig70	3			PTTG_08794	2e-08	Desconhecida
Contig71	3	gnl_ti_453244249	1e-155			
Contig73	26					
Contig75	10					
Contig76	10					
Contig79	3	gnl_ti_452237685	1e-47			
Contig83	3					
Contig86	2	gnl_ti_452114268	1e-154			
Contig90	23					
Contig91	3	gnl_ti_452403809	1e-140			
Contig92	19	gnl_ti_452245402	1e-07			
Contig93	4	gnl_ti_453076733	0.0			
Contig94	4					
Contig101	6	gnl_ti_453234213	4e-14			
Contig102	4					
Contig103	2					
YST_003_02_B04	1	gnl_ti_452866827	5e-08			
YST_003_01_A03	1			PGTG_00638	3-e41	Desconhecida
YST_004_08_A05	1					
YST_004_05_D02	1					
YST_003_04_A05	1					

YST_004_01_B11

1

Tabela 3 cont.: Comparação dos genes selecionados com anotações do genoma de *P. pachyrhizi* através do algoritmo *BlastN* e análises de similaridade entre as proteínas deduzidas por estes genes, e proteínas deduzidas a partir do genoma de *Puccinia graminis tritici*, através do algoritmo *BlastP*.

YST_003_03_G04	1	gnl_ti_454699170	3e-10
YST_004_14_E09	1		
YST_003_03_D11	1		
YST_004_01_D01	1		
YST_004_07_C04	1		
YST_004_12_G08	1		
YST_003_03_H11	1	gnl_ti_452237685	2e-30
YST_004_10_E09	1		
YST_003_04_G09	1	gnl_ti_454648963	8e-33
YST_004_02_C05	1		

As sequências protéicas deduzidas de 58 ORFs foram reanalisadas quanto à localização celular e presença de domínios transmembrana utilizando os algoritmos TargetP e Phobius, respectivamente. Pelo programa TargetP, 53 proteínas deduzidas apresentaram predição para secreção (S) com *score* (RC) \geq 2, obteve-se uma proteína (Contig91) com predição para localização mitocondrial e para duas proteínas deduzidas (Contig13 e Contig71) a predição não foi conclusiva (O) (**Tabela 2**).

Para obtenção de sequências completas dos 64 genes *no hits* identificados, foram utilizadas sequências oriundas de bancos de dados públicos e do banco de dados interno do Laboratório de Genômica. O banco de dados de esporos germinados de *P. pachyrrhizi* é constituído de 34394 ESTs. Análises anteriores demonstraram uma alta redundância de sequências. O alinhamento das sequências pelo programa CAP3 resultou em 3499 contíguos e 2419 singletos. Estas 5918 sequências foram utilizadas como banco de dados de referência denominado (eg_pp – esporo germinado *P. pachyrrhizi*). O banco de dados de ESTs oriundos da interação *P. pachyrrhizi* – *G. max* também foi utilizado na análise. O banco do NCBI da interação *P. pachyrrhizi* – *G. max* é constituído por 14133 ESTs, enquanto que o banco interno do Laboratório de Genômica possui 6846 sequências com qualidade phred >20 . Para facilitar a análise e apresentação dos dados, estes 20.979 ESTs foram agrupados pelo programa CAP3, resultando em um único banco de dados com 6410 contíguos e 3267 singletos. Estas 9677 sequências únicas foram comparadas com o banco de sequências de soja (ESTs / genoma), por meio do algoritmo *BlastN*. Esta análise identificou 2004 sequências *no hits*, potencialmente de origem fúngica, que foram então utilizadas como banco de dados de referência denominado nh_inter – *no hit* interação. As 64 sequências dos genes *no hits* foram então comparadas com as sequências dos bancos eg_pp e nh_inter, e 47 sequências apresentaram similaridade com sequências depositadas nesses bancos, sendo que 7 sequências foram similares exclusivamente a sequências do banco de esporos germinados, 15 sequências mostraram similaridade exclusivamente com sequências do banco de sequências da interação e 25 sequências foram similares a sequências de ambos bancos eg_pp e nh_inter (**Anexo 2**). Feita esta análise comparativa, os genes da biblioteca YST foram

agrupados individualmente com as sequências dos bancos de dados eg_pp e nh_inter utilizando-se o programa CAP3, e foi realizada a predição da ORF pelo ORFinder, culminando na identificação de 29 sequências gênicas completas. Para as 35 sequências restantes foram obtidas apenas sequências parciais dos genes, sem a presença da região 3'. As 29 sequências gênicas completas foram renomeadas com a sigla PHPA_RSP_ (*Phakopsora pachyrhizi* - Rust Secreted Proteins) de acordo com a nomenclatura de candidatos a efetores de *P. pachyrhizi* utilizada no Laboratório de Genômica (**Figura 6, Tabela 4**). Destes 29 genes com ORF completa, 10 são oriundos exclusivamente de cDNA de tecido infectado, 4 são oriundos de cDNA de esporo germinados e 15 são constituídos de sequências de ambos os bancos. Dos 29 genes completos 13 genes já haviam sido identificados em análises anteriores (**Tabela 4**, hachurados), e 16 genes codificam para proteínas ainda não identificadas.

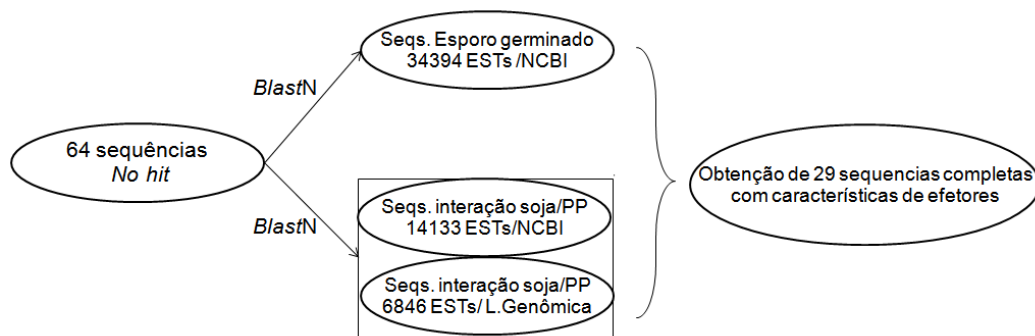


Figura 6: Fluxograma da estratégia utilizada para obtenção da sequência completa das ORFs dos genes de *P. pachyrhizi* identificados a partir da triagem de uma biblioteca YST da interação *P. pachyrhizi* – *G. max*. EST - *expressed sequence tag*. PP – *P. pachyrhizi*.

Tabela 4: Predição de sinal de secreção para as proteínas deduzidas a partir das ORFs dos 29 genes *no hit* com sequência completa.

Identificação Biblioteca		Origem	Contíguo / Clone	n° ESTs	SignalP				Phobius	TargetP		
Candidatos a efetores	PHPA's				Tam.	NN		HMM		Loc	RC	
						Valor D	Pred.	Sprob				Pred.
	PHPA_RSP_02	A	Contig12 / Contig16	264	19	882	Y	999	Y	0	S	1
	PHPA_RSP_03	A	Contig58	55	21	913	Y	993	Y	0	S	1
	PHPA_RSP_04	A	Contig45 / 84	5	19	816	Y	999	Y	0	S	1
	PHPA_RSP_05	A	Contig69	3	23	870	Y	994	Y	0	S	1
	PHPA_RSP_07	A	Contig9	2	23	881	Y	995	Y	0	S	1
	PHPA_RSP_12	A	Contig8 / Contig24	235	20	855	Y	976	Y	0	S	1
	PHPA_RSP_13	I	Contig44	4	21	841	Y	999	Y	0	S	1
	PHPA_RSP_43	A	Contig73	26	21	936	Y	1000	Y	0	S	1
	PHPA_RSP_46	E	Contig49	6	20	896	Y	1000	Y	0	S	1
	PHPA_RSP_52	A	Contig10 / Contig17 / Contig72	52	20	898	Y	1000	Y	0	S	1
	PHPA_RSP_54	A	Contig51 / 60	17	20	928	Y	1000	Y	0	S	1
	PHPA_55_Contig2083	A	Contig39	3	20	892	Y	1000	Y	0	S	1
	PHPA_55_contig2923	A	YST_004_05_D02	1	20	947	Y	1000	Y	0	S	1
	PHPA_RSP_64	E	Contig3	4	24	820	Y	994	Y	0	S	1
	PHPA_RSP_65	E	Contig15	2	23	849	Y	940	Y	0	S	1
	PHPA_RSP_66	A	Contig20	7	23	782	Y	1000	Y	0	S	1
	PHPA_RSP_67	A	Contig30	2	18	813	Y	999	Y	1	S	1
	PHPA_RSP_68	E	Contig32 / Contig68	4	21	853	Y	994	Y	0	S	1
	PHPA_RSP_69	I	Contig101	6	20	905	Y	999	Y	0	S	1
	PHPA_RSP_70	I	Contig102	4	22	770	Y	999	Y	0	S	1
	PHPA_RSP_71	I	Contig19	23	20	713	Y	968	Y	0	S	2
	PHPA_RSP_72	I	Contig37	37	23	857	Y	999	Y	0	S	1
	PHPA_RSP_73	A	Contig90	23	20	908	Y	1000	Y	0	S	1
	PHPA_RSP_74	I	Contig92	19	21	859	Y	996	Y	0	S	1
	PHPA_RSP_75	A	Contig21	2	18	800	Y	999	Y	0	S	2
	PHPA_RSP_76	I	Contig42	11	29	783	Y	915	Y	0	S	1
	PHPA_RSP_77	I	Contig47	39	21	755	Y	945	Y	0	S	1
	PHPA_RSP_78	I	Contig70	3	22	852	Y	998	Y	0	S	1
	PHPA_RSP_79	I	YST 004 01 B11	1	23	548	Y	869	Y	1	S	5

nr – banco de proteínas não redundante; Tam. – tamanho do sinal de secreção; D-score – Média da estimativa do ponto de clivagem do peptídeo sinal e da média dos valores de probabilidade da região atribuída ao peptídeo sinal, discrimina entre a região do peptídeo sinal do restante da proteína; Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q); Loc – Predição da localização celular, (M) mitocondria, (S) peptídeo sinal e – outro; RC – “Coeficiente de confiabilidade”, dado pela razão entre o maior e o segundo maior valor de predição de localização. 1 – 1-0,8; 2 – 0,8-0,6; sucessivamente até o valor 5; Y – Predição positiva de peptídeo sinal pelo SignalP; N – Predição negativa de peptídeo sinal pelo SignalP; S – Predição de secreção pelo TargetP; M – Predição de localização mitocondrial pelo TargetP. (-) não determinado ou sem informação disponível. A – contíguo formado por cDNAs de esporos germinados e tecido infectado, I - contíguo formado por cDNAs de tecido infectado, E - contíguo formado por cDNAs de esporos germinados.

As sequências protéicas deduzidas de 29 ORFs completas foram posteriormente avaliadas quanto ao conteúdo de resíduos de aminoácidos, em especial ao conteúdo de cisteínas, pelo programa ProtParam (**Tabela 5, Anexo 3**). Dos 29 candidatos a efetores, 6 (PHPA_RSP_05, PHPA_RSP_13, PHPA_RSP_69, PHPA_RSP_72, PHPA_RSP_76, PHPA_RSP_78) apresentaram 3,0% ou mais resíduos de cisteínas em sua composição (**Tabela 6, destacados**). Sítios passíveis de formação de pontes de dissulfeto foram verificados para 14 clones (PHPA_RSP_03, PHPA_RSP_05, PHPA_RSP_07, PHPA_RSP_13, PHPA_RSP_66, PHPA_RSP_67, PHPA_RSP_69, PHPA_RSP_70, PHPA_RSP_71, PHPA_RSP_72, PHPA_RSP_75, PHPA_RSP_76, PHPA_RSP_78, PHPA_RSP_79) com o emprego do programa Disulfind (**Tabela 5, Figura 7**).

Tabela 6: Conteúdo de cisteína das proteínas codificadas pelos 29 genes selecionados

Clones PHPAs	Tamanho	% cisteína	Pontes de dissulfeto
PHPA_RSP_02	208	0.5%	-
PHPA_RSP_03	125	0.8%	Predição positiva
PHPA_RSP_04	391	0.3%	-
PHPA_RSP_05	232	3.0%	Predição positiva
PHPA_RSP_07	231	2.6%	Predição positiva
PHPA_RSP_12	97	0.0%	-
PHPA_RSP_13	208	3.8%	Predição positiva
PHPA_RSP_43	91	0.0%	-
PHPA_RSP_46	186	0.5%	-
PHPA_RSP_52	172	0.6%	-
PHPA_RSP_54	178	0.0%	-
PHPA_55_Contig2083	171	0.6%	-
PHPA_55_contig2923	169	0.6%	-
PHPA_RSP_64	247	2.4%	-
PHPA_RSP_65	218	0.0%	-
PHPA_RSP_66	105	2.9%	Predição positiva
PHPA_RSP_67	211	0.9%	Predição positiva
PHPA_RSP_68	76	0.0%	-
PHPA_RSP_69	146	4.8%	Predição positiva
PHPA_RSP_70	195	2.1%	Predição positiva
PHPA_RSP_71	163	1.8%	Predição positiva
PHPA_RSP_72	121	5.0%	Predição positiva
PHPA_RSP_73	169	0.6%	-
PHPA_RSP_74	110	0.0%	-
PHPA_RSP_75	216	1.9%	Predição positiva
PHPA_RSP_76	65	4.6%	Predição positiva
PHPA_RSP_77	105	0.0%	-
PHPA_RSP_78	137	4.4%	Predição positiva
PHPA_RSP_79	150	2.0%	Predição positiva


```

PHPA_46 MQSFILSTLVLIVAQLVSAVPLAERSEVQSSVGHSEQKFAGLGGLPCPPVGGILPPVGGI 60
PHPA_52 MQSFILSTLVLIVAQLVSAVPLAERSEVQSSGGHSEQKFAGLGGLPCPPVGGILPPVGGI 60
PHPA_73 MQSLILSLLLVVAQLTGAIPLSERSEVHSSSESHSEQKFAGIGGLPCPPVGGLLPPFGGI 60
PHPA_54 MQSLFLSTLILVVAQLVSAIPLTERSEIHSSSESHSNQKFAGLGGLLTP-VTGLLGGLGGM 59
***:.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:

PHPA_46 LPPVGGVFPVGGI-----LPPVGGVFPVGGILPPVGGVFPVGGIPPPVLPF---GG 111
PHPA_52 LPPVGGVLPVGGI-----LPPVGGVFPVGG-----GIPPPVLPF---GG 97
PHPA_73 PPPI---LPPVGGLG-----LPPP--ILPPVGG-----FLPPPVLPI---GG 94
PHPA_54 LPIGG--LPVGGGLIPTAGGVIPGGGLLPPVGG-----IPPPGLPLPPVGG 103
*      :* ***:      : *      ::* ***:      :*** **:* **

PHPA_46 IGGIGGIGGAGGFGGGFGGASGFNRNSFGSSVQSSSSSFQTGGGRFGGAGGVGGLGGM 171
PHPA_52 IGGIGGIGGAGGFGGGFGGASGFNRNSFGSSVQSSSSSFQTGGGRFGGAGGVGGLGGM 157
PHPA_73 FGGIGGMGGMGFGGGGFGGASGFNRNSFGSSVQSSSSSFNNGAQFGGAGGAGGI 154
PHPA_54 LGGIGGMGGAGGFGGAFGGASGYNRNTFGSAVQSSSSSFQTGGGRFGGAGGAGGI 163
:*****.* ** *****.*****.*:*:*:*:*:*****.*.***:*****.***.***

PHPA_46 GGFLKDSEHKNKNS 186
PHPA_52 GGFLKDSEHKNKNS 172
PHPA_73 GGFLKDNEQTKNNS 169
PHPA_54 GGFLKDNHKNKNEA 178
*****:***:***:

```

Figura 8 – Alinhamento múltiplo de sequências protéicas codificadas pela ORF dos genes PHPA_RSP_46, PHPA_RSP_52, PHPA_RSP_54 e PHPA_RSP_73, utilizando o programa CLUSTAL 2.0.12, evidenciando a grande similaridade entre as sequências. Cores: Vermelho – resíduo pequeno (pequeno positivo, hidrofóbico); Azul – ácido; Magenta – básico; verde – hidroxil, amina, básico; Cinza – Outros. Pontos: “*” - identidade; “.” - substituição conservada, de acordo com as cores; “-” – substituições semi-conservadas.

```

PHPA_02 MQFLTLVTLIIASQLASSVPLVERAETRISA EHVNQK----- 37
PHPA_04 MHSFILATLLVTSQLASSVPIVERAESHS SDHVKQKGLIGADILTSGLGGIGGSDCLD 60
      * : * .***:*****:*****: * :*:**

PHPA_02 -----GFFGGLP 44
PHPA_04 GTGGNGVIGGGGGLGGLGGLGGIGKTLGGLGGLGGLGGGGGLGGLGGLGGGGALGGLP 120
      * :****

PHPA_02 CAPIG--GMLPPPIG--GFLPPPIGGF-----LPPPIGGFG----- 78
PHPA_04 PPIIGGLGGLPPPIIGGLGGLPPPIGGFGLRPPPIGGFGLPPPIGGFGLPPPLP 180
      .*** * : ***** * : ***** *****

PHPA_02 --GFGGL-----GGFGGL-----GGFGLPPP-PIGGFGL-----GFGGLPPPI 116
PHPA_04 IGGFGLPPPLPIGGFGLPPPLPIGGFGLPPPLPIGGFGLPPPLPIGGFGLPPPI 240
      ***** ***** ***** ***** *****

PHPA_02 GGLG-----FGGLPPMIGGFGLG----- 136
PHPA_04 GGFGLPPPIGGFGLPPPIGGFGGIGGGIRGIGGFGLPPPIGGFGLPP 300
      **.* ***** *****:*

PHPA_02 -----GGIGMGFGMGGGFGGASGFNRNSFSSVQSGSSSSGFK 177
PHPA_04 PPIGGFGLPPPIGGFGGIGGLGGMGGFGMGGGFGGASGFNRNSFNSVQSGATSSGFN 360
      **:*****:***:*****:****:

PHPA_02 AGGGQYGGAGGLGGIGMGFLKKNKSEA 208
PHPA_04 AGGGQYGGAGGMGGIGGLGGFLKKNKSEA 391
      *****:*****:*****:***:****

```

Figura 9 – Alinhamento múltiplo de seqüências protéicas codificadas pelas ORFs dos genes PHPA_RSP_02 e PHPA_RSP_04, utilizando o programa CLUSTAL 2.0.12 evidenciando grande similaridade entre as seqüências. Cores: Vermelho – resíduo pequeno (pequeno positivo, hidrofóbico); Azul – ácido; Magenta – básico; verde – hidroxil, amina, básico; Cinza – Outros. Pontos: “*” - identidade; “:” - substituição conservada, de acordo com as cores; “.” – substituições semi-conservadas.

```

PHPA_55_EE_Contig2083F1 MRCFILSAFLLVILLAASASPIVERAEIQSSNDQTKKFFGPGIPAAGLP 50
PHPA_55_EE_Contig2923F1 MRCFILSAILLVIIQAASASPIVERAEIQSSNEQTKKFFGPGIPAAGLP 50
      *****:***** *****:*****:*****:*****

PHPA_55_EE_Contig2083F1 IGGVPAFGPGMGGIGLGGGFGLGGFGGLGGFGGLGGFGGMGGFG 100
PHPA_55_EE_Contig2923F1 IGGVPAFGPGTFGPG-IGG-IGLGGFGGLGGFGGLGGFGGMGGFG 98
      ***** * * :** :*****:*****:*****:*****

PHPA_55_EE_Contig2083F1 GMGGARGFNNNNFNAAQSSASTSGFSAGGSRAMGGGLGGMGGIGGI 150
PHPA_55_EE_Contig2923F1 GMGGARGFNNNNFNAAQSSASTSGFSAGGARAGGGGLGGMGGIGGI 148
      *****:*****:*****:*****:*****

PHPA_55_EE_Contig2083F1 GGLGGLGGFLKDNEAKNQAA 171
PHPA_55_EE_Contig2923F1 GGIGLGGFLKDNEAKNKQAE 169
      **:*****:*****:***

```

Figura 10 – Alinhamento múltiplo das seqüências das proteínas codificadas pelas ORFs dos genes PHPA_RSP_55_Contig2083 e PHPA_RSP_55_Contig2923, utilizando o programa CLUSTAL 2.0.12, evidenciando grande similaridade entre as seqüências. Cores: Vermelho – resíduo pequeno (pequeno positivo, hidrofóbico); Azul – ácido; Magenta – básico; verde – hidroxil, amina, básico; Cinza – Outros. Pontos: “*” - identidade; “:” - substituição conservada, de acordo com as cores; “.” – substituições semi-conservadas.

possível identificar padrões nas sequências de aminoácidos, que podem representar motivos protéicos que se repetem em diferentes clones (**Figura 13**). Para 15 sequências protéicas foi verificada uma região conservada de 29 resíduos de aminoácidos na região N-terminal, identificada pela expressão regular $L[IV][VT][AS]QL[VA]S[AS]VP[LI][VA]ER[SA]EV[QH]SSx[GE]H[SV][EN]QE]K [FG]$, $\text{evalve } 5,0e-201$ (**Figura 13a**). Para 11 proteínas deduzidas foi verificada uma região conservada de 11 resíduos de aminoácidos internos, identificada pela sequência $GLPPPPIGGFG$, $\text{evalve } 6,9e-127$ (**Figura 13b**). Para 17 proteínas deduzidas foi verificada uma região conservada de 11 resíduos de aminoácidos internos, identificada pela expressão regular $GG[IL]GG[AIM]GG[FL]GG$, $\text{evalve } 2,7e-70$, **Figura 13c**.



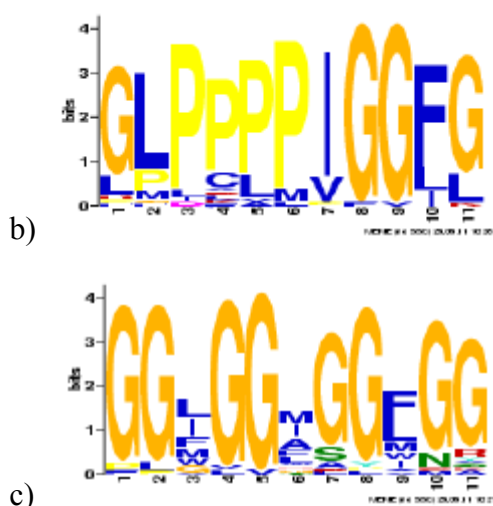


Figura 13 – Esquemas das regiões conservadas das sequências protéicas referentes aos genes *no hit* identificados pelo sistema YST. a) Região conservada de 29 aminoácidos na região N-terminal de 15 proteínas deduzidas. b) Região conservada de 11 aminoácidos na região interna de 17 proteínas deduzidas. c) Região conservada de 11 aminoácidos na região interna de 11 proteínas deduzidas. Cores: Azul – aminoácidos hidrofóbicos; Verde – aminoácidos polares não carregados; Magenta – aminoácidos polares com carga negativa; Vermelho – aminoácidos polares com carga positiva; Rosa – resíduos de histidina; Laranja – resíduos de glicina; Amarelo – resíduos de prolina; Turquesa – resíduos de tirosina.

Para os 29 genes com sequências completas foram desenhado oligonucleotídeos iniciadores com sítios para *EcoRI* e *NotI*, para clonagem no vetor pYST1 e confirmação da secreção em levedura. Para nove deles a secreção em levedura foi confirmada, (**Tabela 7**). Os demais clones ainda estão em processo de confirmação.

Tabela 7 – Confirmação da secreção em levedura das proteínas codificadas pela ORF deduzidas de 29 de genes selecionados identificados na triagem utilizando o sistema armadilha de sinal de secreção em leveduras.

Clones PHPAs	Secreção
PHPA_RSP_02	positiva
PHPA_RSP_03	positiva
PHPA_RSP_04	positiva

PHPA_RSP_05	positiva
PHPA_RSP_07	positiva
PHPA_RSP_12	positiva
PHPA_RSP_13	nd
PHPA_RSP_43	nd
PHPA_RSP_46	nd
PHPA_RSP_52	nd
PHPA_RSP_54	nd
PHPA_55_Contig2083	positiva
PHPA_55_contig2923	positiva
PHPA_RSP_64	nd
PHPA_RSP_65	nd
PHPA_RSP_66	nd
PHPA_RSP_67	nd
PHPA_RSP_68	nd
PHPA_RSP_69	nd
PHPA_RSP_70	nd
PHPA_RSP_71	positiva
PHPA_RSP_72	nd
PHPA_RSP_73	nd
PHPA_RSP_74	nd
PHPA_RSP_75	nd
PHPA_RSP_76	nd
PHPA_RSP_77	nd
PHPA_RSP_78	nd
PHPA_RSP_79	nd

nd – não determinado

6.0 – Discussão

A utilização do sistema de armadilha de sinal de secreção em leveduras, em conjunto com o sequenciamento dos insertos dos clones identificados e

análises de bioinformática usando recursos genômicos disponíveis em bancos de dados, permitiu a identificação e caracterização de vários genes de soja e de *P. pachyrhizi* que codificam proteínas secretadas. Por meio desta abordagem foram identificados 161 genes que codificam para proteínas secretadas, sendo que 64 (39,75%) destes não foram similares a quaisquer sequências de genes de soja depositados nos bancos de dados (genoma e ESTs). A comparação destas sequências ao *draft* do genoma de *P. pachyrhizi*, revelou que 24 genes apresentam similaridade a fragmentos de sequências depositadas neste banco, demonstrando a sua origem fúngica. Os demais genes possivelmente não estão representados nas sequências disponíveis em função da baixa cobertura do genoma disponível (1X). O grande número de genes de origem vegetal identificados, em relação a genes de origem fúngica é devido à maior quantidade de tecido vegetal presente nas folhas infectadas utilizadas para extração do RNA.

A inferência de função com base na comparação das proteínas deduzidas com proteínas depositadas em bancos de dados só foi possível para a proteína codificada pelo gene correspondente ao singlete YST_003_04_A05, que foi similar a uma quitinase. A parede celular dos fungos é uma estrutura altamente dinâmica, em constante mudança, seja na expansão celular durante a divisão, durante a germinação de esporos, crescimento de hifas ramificadas e formação dos septos em fungos filamentosos. Para isso os fungos dependem das atividades de uma série de enzimas hidrolíticas encontradas associadas a parede da celular fúngica. Dentre as hidrolases fúngicas, a quitinase é uma das proteínas mais bem caracterizadas até hoje (Adams *et al.*, 2004).

A maioria dos efetores identificados até o momento compartilha pouca similaridade entre si ou com sequências de domínio público, o que dificulta a caracterização funcional de novos efetores identificados. Além disso, isolados da mesma espécie ou isolados de diferentes espécies podem apresentar conjuntos específicos de efetores (Ellis *et al.*, 2009). Por exemplo, o gene efector Avr3 (Six1) é encontrado em todos os isolados de *Fusarium oxysporum* f.sp. *lycopersici* e de outras *formae speciales* de *F. oxysporum* patogênicos e não patogênicos avaliadas (Ellis *et al.*, 2009). De modo geral, as proteínas efetoras de fungos compartilham características gerais como o fato de serem proteínas pequenas, com menos de

400 aminoácidos de comprimento quando maduras, que podem ser ricas em resíduos do aminoácido cisteína e geralmente possuem um peptídeo sinal na região N-terminal que sinaliza para a secreção, embora existam exceções, como AVRa10 e AVRK1 de *Blumeria graminis f. sp. hordei* (Ridout *et al.*, 2006). Assim, é possível que muitos dos genes identificados sejam genes exclusivos de *P. pachyrhizi*, representando efetores em potencial.

Para a identificação das ORFs dos 64 genes *no hit* identificados foi utilizado um *script* Perl que permite a tradução simultânea de milhares de sequências, que pode gerar previsões incorretas (Kleemann *et al.*, 2008). Depois da predição *ab initio* das ORFs pela tradução nos três quadros de leitura positiva de cada sequência única, obtiveram-se mais de 84 ORFs potenciais. A predição das ORFs na fita antisenso não foi efetuada, uma vez que a clonagem e o sequenciamento foram direcionais (sentido 5'-3'). Após a análise das proteínas deduzidas das 84 ORFs pelo programa SignalP, 55 genes apresentaram predição positiva de peptídeo sinal para a proteína deduzida da ORF predita. Os outros nove genes não apresentaram ORF com proteína deduzida com predição de peptídeo de sinal de secreção ou não apresentaram predição de ORF. Para três deles por meio de análise mais detalhada foi possível determinar o início da ORF (presença ATG) e a sequência protéica codificada pelos transcritos. Para os outros seis genes, não foi possível identificar o início da ORF, o que pode ser explicado pelo tratamento inicial das sequências no script usado, onde se efetuou a remoção das bases de qualidade inferior ao valor phred 20 a partir das extremidades. Desta forma, a predição da ORF foi comprometida inviabilizando as demais análises *in silico*.

Os três genes que codificam proteínas que não apresentam sinal de secreção predito pelo SignalP, mas que foram identificados no sistema YST, provavelmente são genes que codificam proteínas que não são secretadas pela via de secreção convencional do RE/Complexo de Golgi. A secreção não convencional de proteínas já é bem conhecida em eucariotos. A diversidade de proteínas de *S. cerevisiae* que são secretadas e que não apresentam sinal para secreção N-terminal, sugere a existência de vários mecanismos de translocação dessas proteínas para o ambiente extracelular. Por exemplo, pesquisas em *S.*

cerevisiae demonstram que proteínas sem peptídeo sinal podem ser exportadas pelos transportadores ABC (Nombela *et al.*, 2006). O reconhecimento de sinais de secreção alternativos nas proteínas, assim como a utilização de rotas alternativas de secreção são mecanismos propostos para explicar a secreção de proteínas de mamíferos e parasitas que não utilizam a via secretória clássica (Nickel, 2003).

A análise comparativa das proteínas codificadas pelos genes *no hit* com as proteínas anotadas de *Puccinia* revelou baixa similaridade entre as proteínas putativas de *P. pachyrhizi* e de *Puccinia*. Além disso, apenas 12 proteínas das duas espécies foram similares, e nenhuma delas apresentou função descrita. Isto sugere que as espécies causadoras de ferrugens podem contar com proteínas secretadas específicas, fato que torna a busca e caracterização dos efetores mais intrigante e complexa. A recente caracterização do secretoma do fitopatógeno *H. arabidopsidis* revelou que 38% de proteínas preditas não apresentam similaridade a proteínas de outros oomicetos depositadas no NCBI, sugerindo que estas proteínas são específicas deste patógeno. Além disso, quando o secretoma de dois isolados diferentes de *H. arabidopsidis* (Emoy9 e Waco9) foram comparados, 6 proteínas foram encontradas unicamente no isolado Waco9 (Cabral *et al.*, 2011). Por outro lado, parte do “fracasso” na constatação de similaridade de proteínas de diferentes fitopatógenos pode ser atribuído ao número reduzido de sequências desses organismos depositadas em bancos de dados. À medida que esses bancos crescerem, a probabilidade de similaridade entre proteínas poderá ser maior, tornando possível sugerir funções e inferir estratégias comuns de patogênese.

Durante as comparações das sequências gênicas parciais obtidas pelo sequenciamento da biblioteca YST com os ESTs de *P. pachyrhizi* provenientes de bancos de dados públicos e local, procurou-se observar se as sequências utilizadas para completar os genes eram oriundas de esporos germinados ou de folhas de soja infectadas. A diferenciação das estruturas de infecção fúngica na superfície da planta, o desenvolvimento da infecção, bem como a resposta da planta à infecção por patógenos são o resultado da expressão de conjuntos específicos de genes que contribuem para a patogenicidade do fungo, ou para resistência da planta. Para compreender geneticamente um patossistema é importante monitorar os perfis de expressão gênica em diferentes estágios da interação (Sugui e

Deising, 2002). Dos 29 genes com sequência completa, 10 são constituídos exclusivamente de cDNAs de tecido infectado, quatro são constituídos de cDNA de esporo germinado e 15 são constituídos por sequências dos dois tipos de cDNA. Este resultado pode fornecer indício da expressão diferencial dos genes identificados. Porém, como as bibliotecas representam etapas específicas do processo infeccioso, apenas um monitoramento temporal da expressão pode fornecer dados concretos dessa expressão diferencial. Genes de patógenos que são expressos em momentos ou estruturas determinantes do processo de infecção, como os haustórios, poderiam ser priorizados para estudos funcionais (Hu *et al.*, 2007). Hu *et al.*, (2007) construíram bibliotecas de cDNA para diferentes estágios de infecção do fungo *P. triticina*, agente causal da ferrugem da folha do trigo. Análises *in silico* revelaram alta similaridade entre os ESTs oriundos das bibliotecas de cDNA de esporos germinados e não germinados, sugerindo sobreposição entre os transcriptomas destes estágios. Porém, pouca sobreposição foi encontrada com os outros estágios analisados (formação do apressório, desenvolvimento do haustório, interações compatíveis e incompatíveis), o que sugere uma indução de genes fase-específica. Durante a germinação dos esporos observou-se aumento no número de genes envolvidos na conversão de energia, incluindo seis *hits* para frutose bifosfato aldolase. No estágio haustorial, um grande número de genes de proteína ribossomal foi revelado, sugerindo elevada atividade de síntese protéica, semelhante ao que tem sido relatado em estudos anteriores envolvendo os fungos causadores das ferrugens do feijão e do trigo (Broeker *et al.*, 2006; Jakupovic *et al.*, 2006; Thara *et al.*, 2003; Zhang *et al.*, 2003). Transcritos de genes envolvidos na biossíntese de vitaminas e alguns outros compostos só foram identificados no estágio haustorial, alguns dos quais já foram identificados como PIG (*plant induced genes*) (Voegelé, 2006). Poucos genes de origem fúngica foram identificados como específicos para interações incompatíveis, um *hit* de interesse foi para uma monooxigenase P450, possivelmente envolvida na desintoxicação de compostos de defesa, semelhante ao encontrado para *U. maydis* (Miller *et al.*, 2003). Porém, grande parte dos transcritos estágio-específicos codificam para proteínas ainda não caracterizadas, de função desconhecida.

Várias proteínas que são secretadas por fungos fitopatogênicos e reconhecidas por proteínas de resistência são ricas em resíduos de aminoácido cisteína. Exemplos clássicos são os efetores Avr2, Avr4 e Avr9 de *C. fulvum*, que são reconhecidos pelas proteínas de resistência Cf-2, Cf-4 e Cf-9 de tomate, respectivamente (Dixon *et al.*, 1996; Joosten *et al.*, 1997; van den Hooven *et al.*, 1993). Outros exemplos são os efetores de *F. oxysporum*, Six1 e Six4, os efetores de *M. grisea* Avr-Pita e Avr-Pita2 e os efetores do fungo causador da ferrugem do linho *M. lini* AvrP123 e AvrP4 (Stergiopoulos e de Wit, 2009). Os resíduos cisteínas nestas proteínas formam pontes dissulfeto que contribuem para sua estabilidade estrutural no apoplasto da planta, um ambiente rico em proteases (Cabral *et al.*, 2011). Desta forma, a presença de resíduos cisteínas em proteínas secretadas tem sido utilizado como uma característica adicional para identificar proteínas efetoras dentro do conjunto de proteínas que constituem o secretoma de um fungo fitopatogênico. Seis genes de *P. pachyrhizi* caracterizados nesse trabalho codificam proteínas deduzidas secretadas que apresentaram pelo menos 3% de resíduos de cisteínas e 14 genes codificam proteínas que apresentam predição positiva para formação de pontes de dissulfeto. Apesar de já terem sido descritos pelo menos sete genes *R* que conferem resistência à ferrugem asiática da soja, ainda não foram identificadas as proteínas efetoras (Avr) reconhecidas pelas proteínas codificadas pelos genes *Rpp*. Assim, estas proteínas ricas em cisteína são candidatas potenciais para estudos funcionais visando a identificação de genes Avr de *P. pachyrhizi*.

A grande similaridade das proteínas codificadas por alguns genes identificados sugerem que esses genes representem alelos do mesmo loco, uma vez que *P. pachyrhizi* é um fungo dicariótico, ou mesmo que sejam genes parálogos. Sequências parálogas de efetores distribuídas ao longo do genoma têm sido descritas em outras espécies, como os genes que codificam as proteínas PiE12, PiE14 e PiE23 de *P. infestans* descritas por Lee *et al.* (2006), que também foram observadas na sequência do genoma de *P. sojae*. As sequências das proteínas PiE12, PiE14 e PiE23 apresentaram 65, 54, e 61% de similaridade com Ps_135684, Ps_135689 e Ps_135685, respectivamente, ao nível de aminoácidos. Todas as sequências homólogas codificam peptídeos sinal de secreção, motivos

RXLR-like, e sequências de localização nuclear preditos e foram encontradas agrupadas em uma sequência contígua de 20 kb no genoma de *P. sojae*, sugerindo que eles podem representar membros de uma família de genes. Análises das sequências de efetores (Avr) do patógeno causador de ferrugem do linho revelou famílias de proteínas altamente polimórficas, e que parecem estar sob seleção diversificadora. Por exemplo, as proteínas variantes do efector AvrM contêm 14 sítios polimórficos, bem como uma série de deleções e truncamentos (Catanzariti *et al.*, 2006). Um estudo inicial de AvrP4 revelou sete sítios polimórficos concentrados na região C-terminal da proteína (Catanzariti *et al.*, 2006), e a comparação com proteínas homólogas de AvrP4 em 22 espécies *Melampsora* revelou seleção diversificadora significativa em 15 códons localizados na região 3' (Van der Merwe *et al.*, 2009). A proteína AvrL567 apresenta alta variabilidade na sequência de aminoácidos, com 27,5% de resíduos de aminoácidos polimórficos entre suas proteínas variantes (Dodds *et al.*, 2004). Análises mutacionais em AvrL567 confirmaram o papel de vários destes resíduos polimórficos no controle da especificidade de reconhecimento por proteínas R de linho, e parece que essas especificidades são mediadas por múltiplos contactos de aminoácidos de uma maneira quantitativa (Wang *et al.*, 2007). A evolução de formas virulentas de AvrL567 poderia ocorrer de maneira gradual, onde mudanças em um único aminoácido de proteínas avirulentas reduzisse o reconhecimento pela proteína R do hospedeiro. Isto seria seletivamente vantajoso ao fungo causador da ferrugem, e mudanças subsequentes em outros aminoácidos poderia eventualmente, resultar na ausência de reconhecimento, e conseqüentemente, na perda total da efetividade do gene R correspondente (Ravensdale *et al.*, 2011).

O entendimento do mecanismo de translocação de proteínas efetoras de fungos patogênicos para o citoplasma do hospedeiro permanece limitado. Catanzariti *et al.*, (2006) mostraram que o secretoma do haustório de *M. lini* é rico em proteínas efetoras (Avr). A identificação de proteínas secretadas por essas estruturas especializadas e de motivos estruturais conservados entre as proteínas efetoras pode representar um caminho na caracterização do efectoroma de fungos fitopatogênicos, em especial das espécies causadores das ferrugens. Análises *in silico* mais refinadas tornaram a descoberta de candidatos a efetores mais precisa

em oomicetos. Estudos recentes mostraram uma gama de efetores distintos contendo motivos RXLR que são membros de uma superfamília de proteínas que partilham três motivos C-terminais conservados adicionais chamados W, Y e L. Muitas vezes, estes motivos, se repetem em módulos (Dou *et al.*, 2008a, b), que ocorrem em uma região sujeita a seleção positiva para evadir a detecção pelas proteínas R do hospedeiro, assim como para se adaptar a diferentes alvos do metabolismo do hospedeiro. Análise do secretoma de outro oomiceto *H. arabidopsidis* também revelou uma gama de proteínas contendo o motivo RxLR-like. Embora este motivo seja comum entre efetores de oomicetos, entre fungos filamentosos, aparentemente, está ausente (Li *et al.*, 2008). Estudos recentes com o transcriptoma de *B. graminis* identificaram um motivo Y/F/WxC comum entre candidatos a efetores. Este mesmo motivo foi identificado entre candidatos a efetores de *Puccinia graminis*. A descoberta de um motivo similar a RxLR e Y/F/WxC em proteínas putativas de *P. pachyrhizi* identificadas pelo sistema YST pode indicar a contribuição desses motivos na translocação de efetores do patógeno causador da ferrugem asiática da soja. Todavia, essa hipótese tem que ser validada por meio de estudos funcionais.

Como o conhecimento a cerca de motivos conservados em proteínas codificadas pelos genes de fungos causadores de ferrugens é restrita, procurou-se avaliar a presença regiões conservadas entre os efetores putativos de *P. pachyrhizi*. Sendo identificadas três regiões conservadas entre as proteínas deduzidas que podem, ou não, representar motivos protéicos. Estudos futuros são necessários para verificar se essas regiões apresentam algum papel relevante durante a infecção.

A secreção de nove das 29 proteínas deduzidas das ORFs dos genes selecionados foi confirmada em levedura. As outras 20 ORFs estão sendo clonadas no vetor pYST para confirmação da secreção. Como a abordagem inicial para identificação de proteínas secretadas foi o próprio sistema de secreção em levedura espera-se que as ORFs de todos os genes selecionados codifiquem proteínas que sejam secretadas. Todavia, é possível que isso não aconteça, uma vez que o sistema inicialmente trabalha com sequências parciais de genes e não com sua ORF completa. A presença de domínios transmembrana ou a própria

fusão com a invertase pode interferir na secreção da ORF completa clonada, ou ainda, a proteína completa pode ter efeito tóxico na levedura.

Com a identificação de proteínas secretadas por *P. pachyrhizi*, o próximo desafio é comprovar o papel efetor dessas proteínas. Como não existe ainda um sistema de transformação para *P. pachyrhizi*, uma estratégia para identificar as proteínas efetoras desse fungo é a expressão das proteínas secretadas no citoplasma de variedades resistentes (Shan *et al.*, 2004; Vleeshouwers *et al.*, 2008). As proteínas efetoras denominadas proteínas de avirulência são reconhecidas por proteínas de resistência do hospedeiro e assim podem ser identificadas em bioensaios como moléculas que desencadeiam respostas dependentes da proteína R do hospedeiro como a morte celular (reação de hipersensibilidade ou HR).

Diversos métodos podem ser utilizados para a análise de expressão transiente em plantas, como a agroinfiltração (Janssen & Gardner, 1990), a utilização do sistema EDV (*Effector Detector Vector*) baseado no sistema de secreção tipo III de bactérias (Sohn *et al.*, 2007), a introdução de cassetes de expressão em células de plantas utilizando biobalística (Sanford *et al.*, 1993), a introdução de cassetes de expressão em protoplastos de plantas via eletroporação (Guerche *et al.*, 1987) e a utilização de vetores virais (Jeong *et al.*, 2005). Para isso, é necessário o estabelecimento de protocolos de expressão transiente de genes candidatos em folhas de genótipos de soja resistentes ao isolado PPUFV02 de *P. pachyrhizi* utilizado neste trabalho a fim de se identificar genes que codificam proteínas secretadas que são reconhecidos pelas proteínas R codificadas pelos diferentes genes de resistência correspondentes em soja.

7.0 – Conclusões

Por meio do sistema de armadilha de sinal de secreção e análises de bioinformática identificaram-se 64 genes putativos de *P. pachyrhizi* que codificam proteínas secretadas. Foram obtidas ORFs completa para 29 genes. Três genes identificados codificam proteínas com motivos estruturais semelhantes a motivos envolvidos na translocação de efetores de oomicetos e fungos filamentosos para o citoplasma vegetal. Duas proteínas apresentam motivo semelhante ao RxLx-dEER de oomicetos, e uma proteína a um motivo semelhante ao Y/F/WxC encontrado em candidatos a efetores de *Blumeria graminis*. Pela análise comparativa das proteínas deduzidas dos demais genes foram identificadas três regiões

conservadas, que podem ou não representar motivos protéicos.

Das 29 novas proteínas hipotéticas de *P. pachyrhizi*, 6 são ricas em resíduos de cisteína e 14 tem predição positiva para formação de pontes de dissulfeto. A secreção de 9 proteínas foi confirmada pelo sistema YST e a clonagem dos demais genes e confirmação da secreção das proteínas codificadas pelos demais genes aqui identificados encontram-se em andamento. O monitoramento da expressão dos genes identificados durante as diferentes etapas da patogênese e estudos de expressão transiente das proteínas identificadas em fenótipos de soja resistente serão futuramente realizados para identificação das proteínas secretadas com atividade efetora no citoplasma da soja.

Este estudo representa uma etapa essencial nos esforços de caracterização do secretoma de *P. pachyrhizi* visando a identificação de genes de avirulência do patógeno e o entendimento dos mecanismos geradores de novas raças fisiológicas, o que pode contribuir para o desenvolvimento de novas estratégias visando a obtenção de resistência durável à ferrugem asiática.

8.0 – Referências bibliográficas

Adams, D. J. Fungal cell wall chitinases glucanases. *Microbiology*, v.150, p.2029-2035, 2004.

Allen, R. L.; Bittner-Eddy, P. D.; Grenville-Briggs, L. J.; Meitz, J. C.; Rehmany, A. P.; Rose, L. E.; Beynon, J. L. Host-Parasite Coevolutionary Conflict Between *Arabidopsis* and Downy Mildew. *Science*, v.10, p.1957-1960, 2004.

Armstrong, M. R.; Whisson, S. C.; Pritchard, L.; Bos, J. I B.; Venter, E.; Avrova, A. O.; Rehmany, A. P.; Böhme, U.; Brooks, K.; Cherevach, I.; Hamlin, N.; White, B.; *et al.* An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. *Proceedings of the National Academy of Sciences U.S.A.*, v.102, 7766-7771, 2005.

Arnold, R., Brandmaier, S., Kleine, F., Tischler, P., Heinz, E., Behrens, S., Niinikoski, A., Mewes, H. W., Horn, M., and Rattei, T. Sequence based prediction of type III secreted proteins. *PLoS Pathogens*, 5:e1000376. Published online, 2009.

Bailey, K.; Çevik, V.; Holton, N.; Byrne-Richardson, J.; Sohn, K. H.; Coates, M.; Woods-Tör, A.; Aksoy, H. M.; Hughes, L.; Baxter, L.; Jones, J. D. G.; Beynon, J.; Holub, E.; Tör, M. Molecular cloning of ATR5Emoy2 from *Hyaloperonospora arabidopsidis*, an avirulence determinant that triggers RPP5-mediated defense in Arabidopsis. *Molecular Plant-Microbe Interactions*, v.24, p.827–838, 2011.

Bendtsen, J. Improved prediction of signal peptides: SignalP 3.0. *Journal Molecular Biology*, v. 340, p. 783-95, 2004.

Bernoux, M., Timmers, T., Jauneau, A., Briere, C., deWit, P., Marco, Y. and Deslandes, L. RD19, an Arabidopsis cysteine protease required for RRS1-R-mediated resistance, is relocalized to the nucleus by the *Ralstonia solanacearum* Popp2 effector. *Plant Cell*, v.20, p.2252–2264, 2008.

Birch, P.R.; Rehmany, A.P.; Pritchard, L.; Kamoun, S.; Beynon J.L. Trafficking arms: Oomycete effectors enter host plant cells. *Trends in Microbiology*, v.14, p.8-11, 2006.

Boch, J. and Bonas, U. *Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function. *Annual Review Phytopathology*, v.48, p.419–436, 2010.

Bonde, M. R.; Brown, M. F. Morphological comparison of isolates of *Phakopsora pachyrhizi* from different areas of the world. *Canadian Journal of Microbiology*, v. 26, n. 12, p. 1443-1449, 1980.

Broeker, K.; Bernard, F. and Moerschbacher, B.M. An EST library from *Puccinia graminis* f. sp. *tritici* reveals genes potentially involved in fungal differentiation. *FEMS Microbiology Letters*, v.256, p.273–281, 2006.

Bromfield, K.R. Soybean Rust. St. Paul, MN: American Phytopathological Society, 1984.65p (American Phytopathological Society. Monograph, 11).

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

Buritica, P.; Hennen, J. F. Familia Phakopsoraceae (Uredinales). Generos anamorficos y telemorficos. *La Revista de la Academia Colombiana Ciencias Exactatas, Fisicas, y Naturales*. v. 19, p. 47-62, 1994.

Cabral, A.; Stassen, J.H.M.; Seidl, M.F.; Bautor, J.; Parker, J.E.; Van der Ackerveken, G. Identification of *Hyaloperonospora arabidopsidis* transcript sequences expressed during infection reveals isolate-specific effectors. *PLoS ONE*, v.6, e19328, 2011.

Calvo, E.S.; Kiihl, R.A.S.; Garcia, A.; Harada, A.; Hiromoto, D.M. Two major recessive soybean genes conferring soybean rust resistance. *Crop Science*, v.48, p.1350-1354, 2008.

Catanzariti, A-M.; Dodds, P. N.; Lawrence, G. J.; Ayliffe, M. A.; Ellis, J. G. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *The Plant Cell*, v.18, p.243-256, 2006.

Catanzariti, A. M.; Dodds, P. N.; Ellis, J. G. Avirulence proteins from haustoria-forming pathogens. *FEMS Microbiology Letters*. v.269, p.181–88, 2007.

Chakraborty, N.; Curley, J.; Frederick, R.D.; Hyten, D.L.; Nelson, R.L.; Hartman, G.L.; Diers, B.W. Mapping and confirmation of a new allele at *Rpp1* from soybean PI 594538A conferring RB lesion-type resistance to soybean rust. *Crop science*, v.49, p.783-790, 2009.

Chang, J. H., Urbach, J. M., Law, T. F., Arnold, L. W., Hu, A., Gombar, S., Grant, S. R., Ausubel, F. M., and Dangl, J. A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proceedings of the National Academy of Sciences, U.S.A.*, v.102, p.2549-2554, 2005.

CONAB – Companhia Nacional de Abastecimento. Disponível em: <http://www.conab.gov.br/>. Acesso em junho de 2011.

Cunnac, S.; Lindeberg, M.; Collmer, A. *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. *Current Opinion in Microbiology*, v.12, p.53-60, 2009.

Dean, R.A.; Talbot, N.J.; Ebbole, D.J.; Farman, M.L.; Mitchell, T.K.; Orbach, M.J.; Thon, M., Kulkarni, R.; Xu, J.R.; Pan, H.; Read, N.D.; Lee, Y.H., Carbone, I.; Brown, D.; Oh, Y.Y.; Donofrio, N.; Jeong, J.S.; Soanes, D.M.; Djonovic, S.; Kolomiets, E.; Rehmeier, C.; Li, W., Harding, M.; Kim, S., Lebrun, M.H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E. and Birren, B.W. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, v.434, p.980–986, 2005.

de Torres, M., Mansfield, J. W., Grabov, N., Brown, I. R., Ammounh, H., Tsiamis, G., Forsyth, A., Robotzek, S., Grant, M., and Boch, J. *Pseudomonas syringae* effector AvrPtoB suppresses basal defence in *Arabidopsis*. *Plant Journal*. v.47, p.368-382, 2006.

de Wit, P.J.G. How plants recognize pathogens and defend themselves. *Cellular and Molecular Life Sciences*, v.64, p.2726-2732, 2007.

Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, L., Genin, S. and Marco, Y. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proceedings of the National Academy of Sciences, U.S.A.*, v.100, p.8024–8029, 2003.

Dixon, M. S.; Jones, D.A.; Keddie, J.S.; Thomas, C.M.; Harrison, K.; Jones, J.D.G. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell*, v.84, p.451–459, 1996.

Dodds, P. N.; Lawrence, G. J.; Catanzariti, A.-M.; Ayliffe, M. A.; Ellis, J.G. The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell*, v.16, p.755–768, 2004.

Dodds, P. N.; Rafiqi, M.; Gan, P. H. P.; Hardham, A. R.; Jones, D. A.; Ellis, J. G. Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytologist*, v.183, p.993-1000, 2009.

Dou, D.; Kale, S.D.; Wang, X.; Chen, Y.; Wang, Q.; Jiang, R.H.; Arredondo, F.D.;

Anderson, R.G.; Thakur, P.B.; McDowell, J.M.; Wang, Y.; Tyler, B.M. Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell*, v.20, p.1118–1133, 2008a.

Dou, D.; Kale, S.D.; Wang, X.; Jiang, R.H.Y.; Bruce, N.A.; Arredondo, F.D.; Zhang, X.; Tyler, B.M. RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen encoded machinery. *Plant Cell*, v.20, p.1930-1947, 2008b.

Edwards, H. H. and Bonde, M. R. Penetration and Establishment of *Phakopsora pachyrhizi* in Soybean Leaves as Observed by Transmission Electron Microscopy. *Phytopathology*. v. 101, p. 894-900, 2011.

Elling, A. A., Mitreva, M., Gai, X., Martin, J., Recknor, J., Davis, E. L., Hussey, R. S., Nettleton, D., McCarter, J. P., and Baum, T. J. Sequence mining and transcript profiling to explore cyst nematode parasitism. *BMC Genomics*, v.10, p.58, 2009.

Ellis, J. G.; Rafiqi, M.; Gan, P.; Chakrabarti, A. and Dodds, P. N. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Current Opinion in Plant Biology*, v.12, p.399–405, 2009.

Ellis, J. and Dodds, P. Plant pathology: monitoring a pathogen targeted host protein. *Current Opinion in Plant Biology*, v.13, p. R400–R402, 2003.

Embrapa Soja. Disponível em:

http://www.cnpso.embrapa.br/index.php?op_page=294&cod_pai=17. Acesso junho 2011.

Embrapa Soja – Consórcio Antiferrugem. Disponível em: <http://www.consorcioantiferrugem.net>. Acesso em junho 2011.

Flor, H.H. Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology*, v.32, p.653-669, 1942.

Frederick, R. D.; Snyder, C.L.; Peterson, G.L.; Bonde, M.R. Polymerase chain reaction assays for the detection and discrimination of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomia*. *Phytopathology*, v. 92, p. 217-227, 2002.

Garcia, A.; Calvo, E.S.; Kiihl, R.A.S.; Harada, A.; Hiromoto, D.; Vieira, L. 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.

Gietz, R.D.; Woods, R.A. Transformation of yeast by the LiAc/SS carrier DNA/PEG method. *Methods in Enzymology*, v.350, p.87-96, 2002.

Godfrey, D.; Böhlenius, H.; Pedersen, C.; Zhang, Z.; Emmersen, J.; Tordal-Christensen, H. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics*, v.11, p.317, 2010.

Goellner, K.; Loehrer, M.; Langenbach, C.; Conrath, U.; Koch, E.; Schaffrath, U. *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Molecular Plant Pathology*, v.11, n.2, p.169-177, 2010.

Greenberg, J. T. and Vinatzer, B. A. Identifying type III effectors of plant pathogens and

analyzing their interactions with plant cells. *Current Opinion Microbiology*, v.6, p.20-28, 2003.

Greenshields, D. L., and Jones, J. D. Plant pathogen effectors: Getting mixed messages. *Current Opinion in Plant Biology*. v.18, p.R128-130, 2008.

Griffiths, G. and Simons, K. The trans Golgi network: sorting at the exit site of the Golgi complex. *Science*, v.234, p.438-444, 1986.

Groll, M., Schellenberg, B., Bachmann, A. S., Archer, C. R., Huber, R., Powell, T. K., Lindow, S., Kaiser, M., and Dudler, R. A plant pathogen virulence factor inhibits the eukaryotic proteasome by a novel mechanism. *Nature*, v.452, p.755-758, 2008.

Guerche, P.; Bellini, C.; Le Moullec, J.M.; Caboche, M. Use of a transient expression assay for the optimization of direct gene transfer into tobacco mesophyll protoplasts by electroporation. *Biochemistry*, v.69, p.621-628, 1987.

Hartman, G. L.; Miles, M.R.; Frederick, R.D. Breeding for resistance to soybean rust. *Plant Disease*, v.89, n.6, p.664-666, 2005.

Hartwig, E.E. Identification of a fourth major gene conferring resistance to rust in soybeans. *Crop Science*, v.26, p.1135-1136, 1986.

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.

Hennen, J. F. The taxonomy of the soybean rusts. *Soybean rust workshop*. v. 1, p. 29-32, 1995.

Heuer, H.; Yin, Y-N.; Xue, Q-Y.; Smalla, K.; Guo, J-H. Repeat domain diversity of avrBs3-like genes in *Ralstonia solanacearum* strains and association with host preferences in the field. *Applied and Environmental Microbiology*, v.73, p.4379-84, 2007.

Hoefle, C.; Loehrer, M.; Schaffrath, U.; Frank, M.; Schultheiss, H.; Hüchelhoven, 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.

Hogenhout, S. A.; Van der Hoorn, R. A.; Terauchi, R.; and Kamoun, S. Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interaction*, v.22, p.115-122, 2009.

Houterman, P.; Ma, L-S.; Van Ooijen, G.; de Vroomen, M.; Cornelissen, B.; Takken, F.; Rep, M. The effector protein Avr2 of the xylem colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *The Plant Journal*, v.58, p.970-978, 2009.

Hu, G.; Linning, B.; McCallum, B.; Banks, T.; Cloutier, S.; Butterfield, Y.; Liu, J.; Kirkpatrick, R.; Stott, J.; Yang, G.; Smailus, D.; Jones, S.; Marra, M.; Schein, J.; Bakkeren, G. Generation of a wheat leaf rust, *Puccinia triticina*, EST database from stage-specific cDNA libraries. *Molecular Plant Pathology*, v.8, p.451-467, 2007.

Hyten, D.L.; Hartman, G.L.; Nelson, R.L.; Frederick, R.D.; Concibido, V.C.; Narvel, J.M.; Cregan, P.B. Map location of the *Rpp1* locus that confers resistance to soybean rust in soybean. *Crop Science*, v.47, p.837–838, 2007.

Jacobs, J. A. A genetic selection for isolating cDNAs encoding secreted proteins. *Gene*, v.198, p.289-296, 1997.

Jakupovic, M., Heintz, M., Reichmann, P., Mendgen, K. and Hahn, M. Microarray analysis of expressed sequence tags from haustoria of the rust fungus *Uromyces fabae*. *Fungal Genetics Biology*, v.43, p.8–19, 2006.

Janssen, B.-J. & Gardner, R. Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. *Plant Molecular Biology*, v.14, p.61-72, 1990.

Jeong, R.; Hwang, S.; Kang, S., Choi, H.; Park, J., Kim, K. Virus-induced gene silencing as tool for functional genomics in a *Glycine max*. *The Plant Pathology Journal*, v.21, p.158-163, 2005.

Jia, Y.,McAdams, S.A., Bryan, G.T., Hershey, H.P. and Valent, B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO Journal*. v.19, p.4004–4014, 2000.

Jiang, R. H. Y.; Tripathy, S.; Govers, F.; Tyler, B. M. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proceedings of the National Academy of Sciences U.S.A.*, v.105, p.4874-4879, 2008.

Joly, D. L.; Feau, D.; Tanguay, P.; Hamelin, R. C. Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). *BMC Genomics*, v.11, p.422, 2010.

Joosten, M. H. A. J.; Vogelsang, R.; Cozijnsen, T. J.; Verberne, M. C.; DeWit, P. J. G. M. The biotrophic fungus *Cladosporium fulvum* circumvents Cf-4-mediated resistance by producing unstable AVR4 elicitors. *Plant Cell*, v.9, p.367–79, 1997.

Jones, J. D. & Dangl J. L. The plant immune system. *Nature*, v.444, p.323-329, 2006.

Kaku, H. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences of the United States of America*, v. 103, p. 11086-11091, 2006.

Kale, S. D.; Gu, B.; Capelluto, D. G. S.; Dou, D.; Feldman, E.; Rumore, A.; Arredondo, F. D.; Hanlon, R.; Fudal, I.; Rouxel, T., *et al.* External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell*, v.142, p.284-295, 2010.

Kamoun, S. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review of Phytopathology*, v.44, p.41-60, 2006.

Kamoun, S. The secretome of plant-associated fungi and oomycetes. In: Deising, H. & Essey, K. *The Mycota*, p.173-180, 2009.

Kanneganti, T.D.; Huitema, E.; Cakir, C.; Kamoun, S. Synergistic interactions of the plant cell death pathways induced by *Phytophthora infestans* Nep1-like protein PiNPP1.1 and INF1 elicitor. *Molecular Plant-Microbe Interactions*, v.19, p.854–863, 2006.

Kay, S., and Bonas, U. How *Xanthomonas* type III effectors manipulate the host plant. *Current Opinion in Microbiology*, v.12, p.37-43, 2009.

Kelley, B.S.; Lee, S-J.; Damasceno, C.M.B.; Chakravarthy, S.; Kim, B-D.; Martin, G.B.; Rose, J.K.C. A secreted effector protein (SNE1) from *Phytophthora infestans* is a broadly acting suppressor of programmed cell death. *The Plant Journal*, v.62, p.357-366, 2010.

Kemen, E.; Kemen, A. C.; Rafiqi, M.; Hempel, U.; Mendgen, K.; Hahn, M.; Voegelé, R. T. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Molecular Plant Microbe Interaction*, v.18, p.1130–1139, 2005.

Kleemann, J.; Takahara, H.; Stuber, Kurt.; O’Connell, R. Identification of soluble secreted proteins from appressoria of *Colletotrichum higginsianum* by analysis of expressed sequence tags. *Microbiology*, v.154, p.1204–1217, 2008.

Koch, E.; Ebrahimnesbat, F.; Hoppe, H. H. Light and electron microscopic studies on the development of soybean rust (*Phakopsora pachyrhizi* Syd) in susceptible soybean leaves. *Phytopathology*, v. 106, p. 302–320, 1983.

Lee, S-J.; Kelley, B. S.; Damasceno, C. M. B; ST. John, B. ; Kim, B-S.; Kim, B-D.; Rose J. K. C. A functional screen to characterize the secretomes of eukaryotic pathogens and their hosts *in planta*. *Molecular Plant Microbe Interaction*, v.19, p.1368–1377, 2006.

Li, M.; Yuan, J.; Lin, L.; Yue, Z.; Chui-Si, K.; Qing-Liang, Y.; You-Yong, Z.; Cheng-Yun, L. Recognition and Analysis of a Host-Targeting Motif in the Genome of Fungus *Magnaporthe Grisea*. *Bioinformatics and Biomedical Engineering. The 2nd International Conference*, v.16, p.286 –292, 2008.

Link, T. I. and Voegelé, R. T. Secreted proteins of *Uromyces fabae*: similarities and stage specificity. *Molecular Plant Pathology*, v. 9, p. 59–66, 2008.

Loehrer, M.; Langenbach, C.; Goellner, K.; Conrath, U.; Schaffrath, U. Characterization of nonhost resistance of *Arabidopsis* to the Asian soybean rust. *Molecular Plant-Microbe Interactions*, v.21, p.1421-1430, 2008.

Luderer, R.; Takken, F. L. W.; de Wit, P. J. G. M.; Joosten, M. H. A. J. *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing truncated AVR2 elicitor proteins. *Molecular Microbiology*, v.45, p.875–884, 2002.

Luster, D. Proteomic analysis of germinating urediniospores of *Phakopsora pachyrhizi*, causal agent of Asian soybean rust. *Proteomics*, v. 10, p. 3549-57, 2010.

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.

Mendgen, K. & Hahn, M. Plant infection and the establishment of fungal biotrophy. *Trends in Plant Science*, v.7, p.352-356, 2002.

- Meng, X., Bonasera, J. M., Kim, J. F., Nissinen, R. M., and Beer, S. V. Apple proteins that interact with DspA/E, a pathogenicity effector of *Erwinia amylovora*, the fire blight pathogen. *Molecular Plant-Microbe Interaction*, v.19, p.53-61, 2006.
- Meyer, J.D.F.; Silva, D.C.G.; Yang, C.; Pedley, .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.
- Miller, E.A.; Beilharz, T.H.; Malkus, P.N.; Lee, M.C.; Hamamoto, S.; Orci, L.; Schekman, R. Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles. *Cell*, v.114, p.497–509, 2003.
- Miya, A.; Albert, P.; Shinya, T.; Desaki, Y.; Ichimura, K.; Shirasu, K.; Narusaka, Y.; Kawakami, N.; Kaku, H.; Shibuya, N. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proceedings of the National Academy of Sciences USA*, v.104, p.19613-19618, 2007.
- Nickel, W. The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. *European Journal of Biochemistry*, v.270, p.2109–2119, 2003.
- Nombela, C.; Gil, C.; Chaffin, W. L. Non-conventional protein secretion in yeast. *Trends in Microbiology*, v.14, p.15-21, 2006.
- Nurnberger, T. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunological Reviews*, v. 198, p. 249-266, Apr 2004.
- O’Connell, R. J.; Panstruga, R. Tête à tête inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes. *New Phytologisty*, v.171, p.699-718, 2006.
- 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.
- Panstruga, R.; Dodds, P. N. Terrific proteins traffic: the mystery of effector protein delivery by filamentous plant pathogens. *Science*, v. 324, p. 748-750, 2009.
- Pemberton, C.L. & Salmond, G.P.C. The Nep1-like proteins: a growing family of microbial elicitors of plant necrosis. *Molecular Plant Pathology*, v.5, p.353-359, 2004.
- 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.
- Qutob, D.; Kamoun, S.; Gijzen, M. Expression of a *Phytophthora sojae* necrosis inducing protein occurs during transition from biotrophy to necrotrophy. *Plant Journal*, v.32, p.361–373, 2002.
- Raffaele, S.; Win, J.; Cano, J. M.; Kamoun, S. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora*

infestans. *BMC Genomics*, v.11, p.637, 2010.

Ravensdale, M.; Nemri, A.; Thrall, P. H.; Ellis, J. G.; Dodds, P. N. Co-evolutionary interactions between host resistance and pathogen effector genes in flax rust disease. *Molecular Plant Pathology*, v.12, p.93–102, 2011.

Rehmany, A. P., Gordon, A., Rose, L. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun, S., Tyler, B. M., Birch, P. R. J., and Beynon, J. L. Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two Arabidopsis lines. *Plant Cell*, v.17, p.1839-1850, 2005.

Rep, M. Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiology Letters*, v.253, p.19-27, 2005.

Ridout, C. J.; Skamnioti, P.; Porritt, O.; Sacristan, S.; Jones, J. D.; Brown, J. K. Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell*, v.18, p.2402-2414, 2006.

Rytter, J. L., Dowler, W. M., and Bromfield, K. R. Additional alternative hosts of *Phakopsora pachyrhizi*, causal agent of soybean rust. *Plant Disease*, v. 68, p. 818-819, 1984.

Saksirirat W, Hoppe HH Teliospore Germination of Soybean Rust Fungus (*Phakopsora pachyrhizi* Syd.). *J. Phytopathology*, v. 132, p. 339 – 342, 1991.

Sanford, J.; Smith, F.D.; Russel, J. Optimizing the biolistic process for different biological applications. *Methods in Enzymology*, v.217, p.483-509, 1993.

Scherm, H. Quantitative review of fungicide efficacy trials for managing soybean rust in Brazil. *Crop Protection*, v. 28, n. 9, p. 774-782, 2009.

Schlumberger, M. C., and Hardt, W. D. *Salmonella* type III secretion effectors: Pulling the host cell's strings. *Current Opinion of Microbiology*, v.9, p.46-54, 2006.

Schornack, S.; van Dammea, M.; Bozkurta, T. O.; Canoa, L. M.; Smokera, M.; Thinesb, M.; Gaulinc, E.; Kamouna, S.; Huitemaa, S. Ancient class of translocated oomycete effectors targets the host nucleus. *Proceedings of the National Academy of Sciences USA*, v.107, p.17421-17426, 2010.

Shabab, M.; Shindo, T.; Gu, C.; Kaschani, F.; Pansuriya, T.; *et al.* Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato. *Plant Cell*, v.20, p.1169–83, 2008.

Shan, L., He, P., Li, J., Heese, A., Peck, S. C., Nurnberger, T., Martin, G. B., and Sheen, J. 2008. Bacterial effectors target the common signaling partner BAK1 to disrupt multiple MAMP receptor-signaling complexes and impede plant immunity. *Cell Host Microbe*, v.4, p.17-27, 2008.

Shan, W.X., Cao, M., Dan, L.U.; Tyler, B.M. The Avr1b locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene Rps1b. *Molecular Plant–Microbe Interactions*, v.17, p.394–403, 2004.

Shen, Z.; Jacobs-Lorena, M. Evolution of chitin-binding proteins in invertebrates. *Journal of Molecular Evolution*, v.48, p.341–47, 1999.

Silva, D.C.G.; Yamanaka, N.; Brogin, R.L.; Arias, C.A.A.; Nepomuceno, A.L.; Mauro, A.D.; Pereira, S.S.; Nogueira, L.M.; Passianotto, A.L.L.; Abdelnoor, R.V. Molecular mapping of two loci that confer resistance to Asian rust in soybean. *Theoretical and Applied Genetics*, v.117, p.57–63, 2008.

Sinclair JB Rust. 24 - 26. In: Sinclair JB *Compendium of soybean diseases*. The American Phytopathological Society, St. Paul, Minnesota, 1982.

Sohn, K. H.; Lei, R.; Nemri, A.; Jones, J. D. The downy mildew effector proteins ATR1 and ATR13 promote disease susceptibility in *Arabidopsis thaliana*. *Plant Cell*, v.19, p.4077–4090, 2007.

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., van der Hoorn, R. A., and Kamoun, S. Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. *Proceedings of the National Academy of Sciences, U.S.A.*, v.106, p.1654-1659, 2009.

Staskawicz, B. J.; Dahlbeck, D.; Keen, N. T. Cloned avirulence gene of *Pseudomonas syringae* pv. *Glycinea* determines race-specific incompatibility on *Glycine max* (L.) Merr. *Proceedings of the National Academy of Sciences, U.S.A.*, v.81, p.6024–28, 1984.

Stergiopoulos, I. and de Wit, P. J. G. M. Fungal Effector Proteins. *Annual Review of Phytopathology*, v.47, p.233–63, 2009.

Sugui, J. A. and Deising, H.B. Isolation of infection-specific sequence tags expressed during early stages of maize anthracnose disease development. *Molecular Plant Pathology*, v.3, p.197–203, 2002.

Tan-Yujun; Fei-Fuhua; Shan-Zhihui; Zhou-Lecong; Shen-Migzhen; Li-Song; Tan-YJ; Fei-FH; Shan-H; Zhou, L. C.; Shen, M. Z.; Li, S. The role of teliospore of soybean rust pathogen (*Phakopsora pachyrhizi* Sydow) in the infection cycle. *Chinese Journal of Oil Crop Sciences*, v.23, n.1, p.49-51, 2001a.

Tan-Yujun; Fei-Fuhua; Shan-Zhihui; Zhou-Lecong; Shen-Migzhen; Li-Song; Tan-YJ; Fei-FH; Shan-H; Zhou, L. C.; Shen, M. Z.; Li, S. . Study on teliospore formation of *Phakopsora pachyrhizi* Sydow. *Chinese Journal of Oil Crop Sciences*, v.23, n.1, p.57- 59, 2001b.

Thara, V. K.; Fellers, J. P. and Zhou, J.M. In planta induced genes of *Puccinia triticina*. *Molecular Plant Pathology*, v.4, p.51–56, 2003.

Tisdall, J. Beginning Perl for Bioinformatics . *O'Reilly*. First Edition, 2001; ISBN: 0-596-00080-4, 384p.

Ueda, H., Yamaguchi, Y. and Sano, H. Direct interaction between the tobacco mosaic virus helicase domain and the ATP-bound resistance protein, N Factor, during the hypersensitive response in tobacco plants. *Plant Molecular Biology*, v.61, p.31–45, 2006.

van der Merwe, M. M.; Kinnear, M.W.; Barrett, L. G.; Dodds, P. N.; Ericson, L.; Thrall, P.

- H.; Burdon, J.J. Positive selection in AvrP4 avirulence gene homologues across the genus *Melampsora*. *Proceedings of the Royal Society B: Biological Sciences*, v.276, p.2913–2922, 2009.
- van den Burg, H. A.; Harrison, S. J.; Joosten, M. H. A. J.; Vervoort, J.; de Wit, P. J. G. M. *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Molecular Plant Microbe Interaction*, v. 19, p.1420–30, 2006.
- van der Hoorn, R.A.L., de Wit, P. J. G. M. and Joosten, M. H. A. J. Balancing selection favors guarding resistance proteins. *Trends Plant Science*, v.7, p.67–71, 2002.
- van den Hooven, H. W.; Appelman, A. W. J.; Zey, T.; de Wit, P. J. G. M.; Vervoort, J. Folding and conformational analysis of AVR9 peptide elicitors of the fungal tomato pathogen *Cladosporium fulvum*. *European Journal of Biochemistry*, v.264, p.9–18, 1999.
- Vencato, M.; Tian, F.; Alfano, J. R.; Buell, C. R.; Cartinhour, S.; DeClerck, G. A.; *et al.* Bioinformatics-enabled identification of the HrpL regulon and Type III secretion system effector proteins of *Pseudomonas syringae* pv. phaseolicola 1448A. *Molecular Plant-Microbe Interaction*, v.19, p.1193–1206, 2006.
- Vleeshouwers, V.G.A.A.; Driesprong, J.D.; Kamphuis, L.G.; Torto-Alalibo, T.; Van't Slot, K.A.E.; Govers, F.; Visser R.G.F.; Jacobsen, E.; Kamoun, S. Agroinfection-based high throughput screening reveals specific recognition of INF elicitors in *Solanum*. *Molecular Plant Pathology*, v.7, p.499–510, 2008.
- Voegelé, R.T. *Uromyces fabae*: development, metabolism, and interactions with its host *Vicia faba*. *FEMS Microbiology Letters*, v.259, p.165–173, 2006.
- Voegelé, R.T.; Hahn, M.; Mendgen, K. The Uredinales: cytology, biochemistry and molecular biology. In: Deising, H. & Essey, K. *The Mycota*, p. 69-98, 2009.
- Wang, C. I.; Guncar, G.; Forwood, J. K.; Teh, T.; Catanzariti, A. M.; Lawrence, G. J.; Loughlin, F. E.; Mackay, J. P.; Schirra, H. J.; Anderson, P. A.; Ellis, J. G.; Dodds, P. N.; Kobe, B. Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *The Plant Cell*, v. 19, p. 2898–2912, 2007.
- Webb C.A. & Fellers J.P. Cereal rust fungi genomics and the pursuit of virulence and avirulence factors. *Minireview. FEMS Microbiology Letters*, v.1, p.1-7, 2006.
- Whisson, S. C.; Boevink, P. C.; Moleleki, L.; Avrova, A. O.; Morales, J. G.; Gilroy, E. M.; Armstrong, M. R.; Grouffaud, S.; van West, P.; Chapman, S. *et al.*: A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*, v.450, p.115-118, 2007.
- Win, J.; Morgan, W.; Bos, J.; Krasileva, K.V.; Cano, L.M.; Chaparro-Garcia, A.; Ammar, R.; Staskawicz, B.J.; Kamoun, S. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell*, v.19, p.2349–2369, 2007.
- Yeh CC, Tschanz AT, Sinclair JB Induced teliospores formation by *Phakopsora pachyrhizi*

on soybeans and other hosts . *Phytopathology*, v. 71, p. 1111-1112, 1981.

Yorinori, J.T.; Pavia, W.M.; Frederick, R.D.; Costamilan, L.M.; Bertagnolli, P.F.; Hartman, G.L.; Godoy, C.; J. Nunes, Jr. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. *Plant Disease*, v.89, p.675-677, 2005.

Zambolim, L. Ferrugem asiática da soja. 1. ed. Viçosa: Departamento de Fitopatologia-UFV, 2006.

Zhang, L.; Meakin, H. and Dickinson, M. Isolation of genes expressed during compatible interactions between leaf rust (*Puccinia triticina*) and wheat using cDNA-AFLP. *Molecular Plant Pathology*, v.4, p.469-477, 2003.

ANEXO

Anexo 1: Análise dos 161 genes identificados pelo sistema YST

Contíguo / Clone	n° ESTs	Soybase/NCBI nr	Descrição	Função predita	avaliada	SignalP					Phobius	TargetP	
						Size	NN		HMM			Loc	RC
							Valor D	Pred.	Sprob	Pred.			
Contig1	30	AAD33696	PR1a precursor	Defesa contra patogenos	1e-42	27	903	Y	999	Y	0	S	1
Contig2	5	AAD33696	PR1a precursor	Defesa contra patogenos	2e-08	28	779	Y	993	Y	0	S	1
Contig3	4	-	no hit	-	-	24	820	Y	994	Y	0	S	1
Contig4	2	Gm18 ^{Soybase}	unknown	Desconhecida	1e-105	-	-	-	-	-	-	-	1
Contig5	3	-	no hit	-	-	24	879	Y	998	Y	0	S	1
Contig6	5	Gm18 ^{Soybase}	no hit	-	5e-94	27	872	Y	976	Y	1	S	1
Contig7	4	CAA40474	chitinase	Defesa contra patogenos	2e-16	31	803	Y	811	Y	0	S	1
Contig8 / Contig24	235	-	no hit	-	-	20	855	Y	976	Y	0	S	1
Contig9	2	-	no hit	-	-	23	881	Y	995	Y	0	S	1
Contig10 / Contig17 / Contig72	52	XP_002956216	unknown	Desconhecida	9e-05	20	898	Y	1000	Y	0	S	1
Contig 11	12	ACU17113	unknown	Desconhecida	1e-91	22	886	Y	1000	Y	0	S	1
Contig12	217	-	no hit	-	-	19	882	Y	999	Y	0	S	1
Contig13	3	-	no hit	-	-	27	430	N	0	N	0	-	2
Contig14	4	ACU17875	unknown	Desconhecida	8e-17	24	901	Y	1000	Y	0	S	1
Contig15	2	-	no hit	-	-	23	849	Y	940	Y	0	S	1
Contig16	46	-	no hit	-	-	19	891	Y	998	Y	0	S	1
Contig18	2	ACN97419	thaumatin-like protein	Defesa contra patogenos	1e-21	22	906	Y	1000	Y	0	S	1
Contig19	23	-	no hit	-	-	20	713	Y	968	Y	0	S	2
Contig20	7	-	no hit	-	-	23	782	Y	1000	Y	0	S	1
Contig21	2	-	no hit	-	-	18	800	Y	999	Y	0	S	2
Contig22	37	-	no hit	-	-	19	864	Y	999	Y	0	S	1
Contig23	2	ACU20499	unknown	Desconhecida	1e-35	28	789	Y	840	Y	0	S	1
Contig24	31	-	no hit	-	-	ok	856	Y	975	Y	0	S	1

Anexo 1cont.: Análise dos 161 genes identificados pelo sistema YST

Contig25	2	-	no hit	-	-	24	728	Y	911	Y	0	S	1
Contig26	2	ACU23540	unknown	Desconhecida	5e-39	17	670	Y	997	Y	0	S	4
Contig27	218	AAD33696	PR1a precursor disease resistance	Defesa contra patogenos	2e-55	28	807	Y	997	Y	0	S	1
Contig28	2	ABV89606	response protein stress-induced	Defesa contra patogenos	8e-29	22	834	Y	998	Y	0	S	1
Contig29	2	ACD45980	receptor-like kinase 2	Resposta ao estresse	1e-27	24	885	Y	989	Y	0	S	1
Contig30	2	-	no hit	-	-	18	813	Y	999	Y	1	S	1
Contig31	9	ACU15001	unknown	Desconhecida	3e-35	24	901	Y	1000	Y	0	S	1
Contig32	4	-	no hit	-	-	21	853	Y	994	Y	0	S	1
Contig33	11	BAA77676	acidic chitinase	Defesa contra patogenos	9e-21	26	890	Y	999	Y	0	S	1
Contig34	2	ADN05146	PsbA - Photosynthetic reaction centre protein cysteine proteinase	Metabolismo	5e-27	25	573	Y	944	Y	1	S	1
Contig35	3	CAB17076	precursor	Resposta ao estresse	8e-65	31	771	Y	994	Y	0	S	1
Contig36	2	-	no hit	-	-	26	795	Y	998	Y	0	S	1
Contig37	37	-	no hit	-	-	23	857	Y	999	Y	0	S	1
Contig38	2	-	no hit	-	-	20	896	Y	999	Y	0	S	1
Contig39	3	-	no hit	-	-	20	892	Y	1000	Y	0	S	1
Contig40	4	P29136	Metalloendoproteinas e 1	Degradação / Modificação de proteínas	4e-52	29	546	Y	595	Y	0	-	2
Contig41	2	AAK13589	rRNA intron-encoded homing endonuclease	Proteina ribossomal	6e-22	-	-	-	-	-	-	-	1
Contig42	11	no hit	no hit	-	-	29	783	Y	915	Y	0	S	1
Contig43	5	ACU23646	unknown	Desconhecida	2e-22	25	919	Y	1000	Y	0	-	3
Contig44	4	-	no hit	-	-	21	841	Y	999	Y	0	S	1
Contig45 / 84	5	-	no hit	-	-	19	816	Y	999	Y	0	S	1
Contig46	5	ACU16265	unknown	Desconhecida	2e-05	33	859	Y	1000	Y	0	S	1
Contig47	39	-	no hit	-	-	21	755	Y	945	Y	0	S	1
Contig48	4	AAR26001	endo-1,3-beta- glucanase	Defesa contra patogenos	1e-15	31	729	Y	992	Y	0	S	1

Anexo 1cont.: Análise dos 161 genes identificados pelo sistema YST

Contig49	6	XP_001615225	asparagine-rich antigen	Sintese de proteinas	4e-05	20	896	Y	1000	Y	0	S	1
Contig50	6	P29136	Metalloendoproteinas e 1	Degradação / Modificação de proteinas	9e-35	29	840	Y	991	Y	0	S	1
Contig51 / 60	17	XP_002648161	unknown	Desconhecida	6e-05	20	928	Y	1000	Y	0	S	1
Contig52	52	BAF95881	putative pathogenesis-related protein 1	Defesa contra patogenos	1e-18	27	852	Y	995	Y	0	S	1
Contig53	4	-	no hit	-	-	-	-	-	-	-	-	-	1
Contig54	2	ACU15747	unknown	Desconhecida	1e-77	23	62	N	1	N	0	-	2
Contig55	11	XP_002531443	serine-threonine protein kinase	Sinalização celular	1e-09	24	709	Y	994	Y	0	S	1
Contig56 / 57	34	-	no hit	-	-	20	890	Y	1000	Y	0	S	1
Contig58	55	-	no hit	-	-	21	913	Y	993	Y	0	S	1
Contig59	2	-	no hit	-	-	35	696	Y	695	Y	1	S	1
Contig61	3	-	no hit	-	-	19	814	Y	998	Y	0	S	1
Contig62	8	BAE71282	putative receptor-like GPI-anchored monodehydroascorbat e reductase	Divisao celular	2e-33	32	878	Y	1000	Y	0	S	1
Contig63	7	ACH87167	e reductase	Metabolismo	2e-48	20	454	Y	647	Y	0	S	5
Contig64	2	ACU20492	unknown	Desconhecida	4e-69	22	816	Y	994	Y	0	S	1
Contig65	2	Gm15 ^{Soybase}	unknown	Desconhecida	5e-66	-	-	-	-	-	-	-	1
Contig66	3	-	no hit	-	-	20	847	Y	986	Y	0	S	1
Contig67	4	ACU15834	unknown	Desconhecida	1e-70	25	906	Y	995	Y	0	S	1
Contig68	18	-	no hit	-	-	21	808	Y	984	Y	0	S	1
Contig69	3	-	no hit	-	-	23	870	Y	994	Y	0	S	1
Contig70	3	-	no hit	-	-	22	852	Y	998	Y	0	S	1
Contig71	3	-	no hit	-	-	23	139	N	0	N	5	-	3
Contig73	26	-	no hit	-	-	21	936	Y	1000	Y	0	S	1
Contig74	2	ABY48136	glycine-rich protein	Desconhecida	2e-19	24	170	N	2	N	0	-	4
Contig75	10	-	no hit	-	-	21	888	Y	986	Y	0	S	1
Anexo 1cont.: Análise dos 161 genes identificados pelo sistema YST													
Contig76	10	-	no hit	-	-	20	851	Y	994	Y	0	S	1
Contig77	3	ABK63195	Hevein-like protein	Defesa contra patogenos	3e-25	21	916	Y	999	Y	0	S	1

Contig78	6	ACU15142	unknown	Desconhecida	2e-38	24	908	Y	1000	Y	0	S	1
Contig79	3	-	no hit	-	-	-	-	-	-	-	-	-	1
Contig80	37	AAF61436	lipid transfer protein precursor	Defesa contra patogenos	2e-26	24	822	Y	997	Y	0	S	1
Contig81	28	ACU14216	unknown	Desconhecida	3e-47	24	923	Y	999	Y	0	S	1
Contig82	5	ACU21372	unknown	Desconhecida	4e-10	33	795	Y	943	Y	0	S	1
Contig83	3	-	no hit	-	-	19	618	Y	899	Y	1	S	4
Contig85	35	Gm11 ^{Soybase}	unknown	Desconhecida	1e-73	23	858	Y	996	Y	0	S	1
Contig86	2	-	no hit	-	-	25	941	Y	1000	Y	0	S	1
Contig87	3	ACJ06541	cyclophilin [Camellia oleifera]	Degradação / Modificação de proteínas	2e-37	24	910	Y	998	Y	0	S	1
Contig88	7	AAQ63884	putative serine carboxypeptidase	Degradação / Modificação de proteínas	4e-22	28	869	Y	997	Y	0	S	1
Contig89	2	ACU24190	unknown	Desconhecida	9e-08	-	-	-	-	-	-	-	1
Contig90	23	-	no hit	-	-	20	908	Y	1000	Y	0	S	1
Contig91	3	-	no hit	-	-	21	621	Y	2	N	0	M	4
Contig92	19	-	no hit	-	-	21	859	Y	996	Y	0	S	1
Contig93	4	-	no hit	-	-	25	724	Y	992	Y	0	S	1
Contig94	4	-	no hit	-	-	19	891	Y	999	Y	0	S	1
Contig95	3	ACV70142	putative translocon-associated protein	Transporte celular	4e-32	28	808	Y	1000	Y	0	S	1
Contig96	21	ACU14241	unknown	Desconhecida	2e-92	20	930	Y	1000	Y	0	S	1
Contig97	2	XP_002513051	serine/threonine-protein kinase bri1	Sinalização celular	1e-36	32	623	Y	479	N	0	M	5
Contig98	3	YP_358636	unknown	Desconhecida	1e-45	-	-	-	-	-	-	-	1
Contig99	3	ACU14752	unknown	Desconhecida	2e-74	24	877	Y	998	Y	0	S	2
Contig100	12	Gm06 ^{Soybase}	unknown	Defesa contra patogenos	2e-68	21	877	Y	995	Y	0	S	1
Contig101	6	-	no hit	-	-	20	905	Y	999	Y	0	S	1

Anexo 1cont.: Análise dos 161 genes identificados pelo sistema YST

Contig102	4	-	no hit	-	-	22	770	Y	999	Y	0	S	1
Contig103	2	-	no hit	-	-	28	834	Y	995	Y	0	S	1

YST_003_02R1_1_B04.esd	-	no hit	-	-	20	871	Y	999	Y	0	S	1
YST_003_01R1_A03.esd	XP_001834281	unknown	Desconhecida	3e-06	23	910	Y	1000	Y	0	S	2
YST_004_08_Y5_1_A05.esd	-	no hit	-	-	20	919	Y	1000	Y	0	S	1
YST_004_05_R1_2_D02.esd	-	no hit	-	-	20	947	Y	1000	Y	0	S	1
YST_003_04_R1_A05.esd	AAP42832	chitinase	Parede celular	1e-34	24	865	Y	999	Y	0	S	1
YST_004_01_R1_B11.esd	-	no hit	-	-	23	548	Y	869	Y	1	M	5
YST_003_03_R1_1_G04.esd	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_004_14_R1_1_E09.esd	-	no hit	-	-	24	896	Y	996	Y	0	S	1
YST_003_03_R1_2_D11.esd	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_004_01_R1_D01.esd	-	no hit	-	-	25	550	Y	886	Y	0	S	2
YST_003_04_R1_H03.esd	-	no hit	-	-	23	874	Y	986	Y	0	S	1
YST_004_07_R1_1_C04.esd	-	no hit	-	-	20	887	Y	1000	Y	1	S	1
YST_004_12_R1_2_G08.esd	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_003_03_R1_1_H11.esd	-	no hit	-	-	21	912	Y	993	Y	0	S	1
YST_004_10_R1_2_E09.esd	-	no hit	-	-	27	813	Y	999	Y	0	S	1
YST_003_04_R1_G09.esd	-	no hit	-	-	24	816	Y	997	Y	0	S	1
YST_004_02_R1_C05.esd	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_003_01R1_H01.esd	-	no hit	-	-	-	-	-	-	-	-	-	1
YST_003_01R1_A03.esd	XP_001834281	unknown	Desconhecida	3e-06	23	910	Y	1000	Y	0	S	2
YST_004_10_R1_2_G05.esd	BAB86897	syringolide-induced protein	Defesa contra patogenos	1e-10	34	776	Y	887	Y	1	S	1

Anexo 1cont.: Análise dos 161 genes identificados pelo sistema YST

YST_003_01R1_D02.esd	BAA77676	acidic chitinase	Defesa contra patogenos	2e-95	27	346	N	628	Y	0	-	2
YST_003_02R1_1_D10.esd	ACM89542	S-locus lectin protein kinase family protein	Resposta ao estresse	1e-108	49	337	N	0	N	0	-	5

YST_003_02R1_1_E11.esd	ACU20034	unknown ectonucleotide pyrophosphatase/phos phodiesterase	Desconhecida	3e-29	23	72	N	0	N	0	_	2
YST_003_02R1_1_B03.esd	XP_002514102		Transcrição / Tradução	8e-92	26	559	Y	140	N	0	S	3
YST_003_04_R1_A05.esd	AAP42832	chitinase	Parede celular	1e-34	20	68	N	0	N	0	_	2
YST_004_10_R1_2_C04.esd	ABR26094	retrotransposon protein	Transposons	3e-16	33	17	N	0	N	0	_	3
YST_004_11_R1_2_G09.esd	XP_002515387	Aspartic proteinase nepenthesin-2 precursor	Degradação / Modificação de proteínas	6e-06	28	702	Y	991	Y	0	S	1
YST_003_03_R1_1_E09.esd	BAE71287	putative early nodulin-like 2 GPI- anchored protein	Metabolismo	2e-22	-	-	-	-	-	-	_	1
YST_003_01R1_F07.esd	ABV23277	leucine-rich repeat- like protein	Sinalização celular	5e-25	30	79	N	0	N	0	_	5
YST_004_07_R1_1_A10.esd	ADC80503	non-specific lipid transfer protein 1b precursor	Defesa contra patogenos	2e-22	26	861	Y	1000	Y	0	S	1
YST_004_10_R1_2_E08.esd	ACH48223	tumor differentially expressed protein		9e-08	24	328	N	141	N	0	_	3
YST_004_07_R1_1_D01.esd	XP_002510165	skip-2, putative conserved	Sinalização celular	2e-45	19	566	Y	997	Y	0	S	3
YST_003_03_R1_1_G04.esd	ABM53545	hypothetical protein	Desconhecida	4e-06	-	-	-	-	-	-	_	1
YST_003_01R1_E03.esd	ACJ37414	protein kinase	Defesa contra patogenos	4e-73	19	66	N	0	N	0	_	1
YST_003_04_R1_B05.esd	ACU14858	unknown	Desconhecida	7e-89	22	933	Y	999	Y	0	S	1
YST_003_02R1_1_B08.esd	XP_002529978	gulonolactone oxidase	Metabolismo	3e-66	29	850	Y	999	Y	0	S	1
YST_003_02R1_1_F04.esd	ACU13668	unknown cysteine proteinase precursor	Desconhecida	4e-69	27	719	Y	968	Y	0	S	1
YST_003_04_R1_C10.esd	CAB17076		Resposta ao estresse	7e-05	-	-	-	-	-	-	_	1

Anexo 1cont.: Análise dos 161 genes identificados pelo sistema YST

YST_004_10_R1_2_C02.esd	XP_002886322	hypothetical protein	Desconhecida	7e-08	29	351	N	4	N	0	_	3
YST_003_03_R1_1_D08.esd	XP_002530644	aberrant large forked product	Sinalização celular	4e-07	-	-	-	-	-	-	_	1

YST_004_07_R1_1_F12.esd	ACU19114	unknown	Desconhecida	4e-14	19	410	N	689	Y	0	M	3
YST_004_01_R1_B04.esd	AAC04711	beta-1,3-glucanase 3	Defesa contra patogenos	8e-49	18	875	Y	999	Y	0	S	1
YST_004_06_R1_2_C12.esd	AAV44205	unknow hypothetical protein	Desconhecida	4e-18	18	82	N	0	N	0	_	5
YST_004_01_R1_G11.esd	CAK18872	precursor protein disulfide	Desconhecida	4e-35	28	175	N	2	N	0	M	5
YST_003_04_R1_B06.esd	BAD24714	isomerase-like protein	Metabolismo	5e-54	28	877	Y	995	Y	0	S	1
YST_004_09_R1_1_A06.esd	ACU15426	unknown	Desconhecida	2e-04	-	-	-	-	-	-	_	1
YST_003_02R1_1_G10.esd	AAO23072	R 14 protein	Defesa contra patogenos	2e-37	24	879	Y	1000	Y	0	S	1
YST_004_06_R1_2_A02.esd	AAD33696	PR1a precursor hydroproline-rich	Defesa contra patogenos	3e-06	21	632	Y	929	Y	0	S	1
YST_004_10_R1_2_B11.esd	AAA33969	protein unnamed protein	Parede celular	2e-08	-	-	-	-	-	-	_	1
YST_004_02_R1_F01.esd	CBI22404	product pore calcium channel	Desconhecida	6e-10	19	881	Y	999	Y	0	S	1
YST_004_01_R1_E08.esd	Q75VR0	protein 1B R 13 protein [Glycine	Transporte celular	8e-20	35	240	N	1	N	1	M	3
YST_003_04_R1_B02.esd	AAO23073	max]	Defesa contra patogenos	2e-12	23	854	Y	1000	Y	0	S	1

nr – banco de proteínas não redundante; Tam. – tamanho do sinal de secreção; D-score – Média da estimativa do ponto de clivagem do peptídeo sinal e da média dos valores de probabilidade da região atribuída ao peptídeo sinal, discrimina entre a região do peptídeo sinal do restante da proteína; Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q); Loc – Predição da localização celular, (M) mitocondria, (S) peptídeo sinal e – outro; RC – “Coeficiente de confiabilidade”, dado pela razão entre o maior e o segundo maior valor de predição de localização. 1 – 1-0,8; 2 – 0,8-0,6; sucessivamente até o valor 5; Y – Predição positiva de peptídeo sinal pelo SignalP; N – Predição negativa de peptídeo sinal pelo SignalP; S – Predição de secreção pelo TargetP; M – Predição de localização mitocondrial pelo TargetP. (-) não determinado ou sem informação disponível.

Anexo 2 – Composição dos contíguos quanto a origem das seqüências

Identificação seqüência	Acesso Banco de Dados	E- value	Orige m	Identificação seqüência	Acesso Banco de Dados	E- value	Orige m
Contig3	eg_pp_Contig_2756	1e-171	E	Contig66	nh_inter_GMPP_CD_04_099_B08.esd	1e-102	I
Contig8 / Contig24	nh_inter_Contig352	1e-151	A	Contig68	eg_pp_Contig_848	1e-109	E
Contig8 / Contig24	eg_pp_Contig_3311	1e-20		Contig69	nh_inter_Contig615	1e-147	A
Contig9	nh_inter_Contig615	0.0	A	Contig69	eg_pp_Contig_871	1e-45	

Contig9	eg_pp_Contig_871	0.0		Contig70	nh_inter_GMPP_CD_04_093e_C11.esd	1e-159	I
Contig10 / Contig17							
/Contig72	nh_inter_Contig526	2e-99	A	Contig71	eg_pp_Contig_2192	0.0	E
Contig10 / Contig17							
/Contig72	eg_pp_Contig_1727	2e-86		Contig73	eg_pp_Contig_3090	1e-155	A
Contig15	eg_pp_Contig_2392	1e-139	E	Contig73	nh_inter_Contig898	1e-155	
					nh_inter_gi_120533493_gb_EH261626.1_EH		
Contig12 / Contig 16	nh_inter_Contig420	0.0	A	Contig75	261626	1e-144	A
Contig12 / Contig 16	eg_pp_Contig_2979	0.0		Contig75	eg_pp_Contig_2150	1e-138	
Contig19	nh_inter_Contig864	0.0	I	Contig76	nh_inter_GMPP_CD_04_088_B06.esd	0.0	I
					gi 120527564 gb EH255698.1 EH255698		
Contig20	eg_pp_Contig_3019	1e-122	A	Contig79	JGI_ACAC2550	1e-148	E
Contig20	nh_inter_Contig2451	1e-122		Contig83	eg_pp_Contig_1354	3e-96	A
	gi 120526387 gb EH254521.1 EH254						
Contig21	521 JGI_ACAC616	1e-118	A	Contig83	nh_inter_gmpp-cd-02-013-F09.esd	2e-33	
Contig21	nh_inter_Contig528	1e-118		Contig90	nh_inter_Contig745	0.0	A
Contig22	nh_inter_Contig949	0.0	A	Contig90	eg_pp_Contig_2722	9e-08	
Contig22	eg_pp_Contig_313	0.0		Contig91	eg_pp_Contig_1044	1e-113	A
Contig30	eg_pp_Contig_1209	0.0	A	Contig91	nh_inter_GMPP_CD_04_099_H11.esd	7e-87	
	nh_inter_gi_120494813_gb_EH2229						
Contig30	80.1_EH222980	2e-17		Contig92	nh_inter_Contig374	0.0	I
Contig32	eg_pp_Contig_848	1e-114	E	Contig93	eg_pp_Contig_1511	1e-171	E
Contig37	nh_inter_Contig330	0.0	I	Contig94	nh_inter_Contig420	1e-134	A
Contig38	eg_pp_Contig_2054	1e-100	A	Contig94	eg_pp_Contig_2979	1e-128	
	nh_inter_GMPP_CD_04_035_G11.e						
Contig38	sd	1e-10		Contig101	nh_inter_Contig680	1e-179	I
Contig39	eg_pp_Contig_2083	0.0	A	Contig102	nh_inter_Contig476	1e-106	I
Contig39	nh_inter_Contig363	0.0		YST_003_02_B04	nh_inter_Contig565	0.0	I
	nh_inter_GMPP_CD_04_107_H12.e						
Contig42	sd	1e-127	I	YST_004_08_A05	eg_pp_Contig_1275	1e-157	A

Anexo 2 cont. – Composição dos contíguitos quanto a origem das sequências

Contig42	nh_inter_Contig754	1e-127		YST_004_08_A05	nh_inter_GMPP_CD_04_092_D11.esd	1e-155	
Contig44	nh_inter_Contig524	1e-118	I	YST_004_05_D02	eg_pp_Contig_2923	1e-128	A
Contig45 / Contig84	eg_pp_Contig_1075	1e-160	A	YST_004_05_D02	nh_inter_Contig363	6e-88	
Contig45 / Contig84	nh_inter_Contig420	1e-118		YST_004_01_B11	nh_inter_Contig2911	1e-148	I
Contig47	nh_inter_Contig629	1e-180	I	YST_004_01_D01	nh_inter_Contig328	1e-108	I

Contig51 / Contig60	eg_pp_Contig_2722	1e-115	A	YST_004_07_C04	nh_inter_GMPP_CD_04_083_H08.esd	0.0	A
Contig51 / Contig60	nh_inter_Contig844	1e-115		YST_004_07_C04	nh_inter_Contig619	1e-116	
Contig53	nh_inter_Contig605	1e-113	I	YST_004_07_C04	eg_pp_Contig_3252	1e-109	
Contig56 / Contig57	nh_inter_GMPP_CD_04_035_G11.e sd	1e-102	A	YST_003_03_H11	nh_inter_Contig302	1e-133	A
Contig56 / Contig57	eg_pp_Contig_2054	2e-28		YST_003_03_H11	gi 120499928 gb EH228062.1 EH228062		
Contig58	nh_inter_Contig302	1e-143	A		JGI ACAC4422	2e-66	
Contig58	gi 120499928 gb EH228062.1 EH228 062 JGI ACAC4422	2e-75					

nh-inter – biblioteca de cDNA interação soja-ferrugem, eg-pp – biblioteca de cDNA esporos germinados, A – contíguo formado por cDNAs de esporos germinados e tecido infectado, I - contíguo formado por cDNAs de tecido infectado, E - contíguo formado por cDNAs de esporos germinados.

Anexo 3: Composição de aminoácido pelo programa ProtParam

Contíguo	Aminoácidos																							
	Ala (A)		Arg (R)		Asn (N)		Asp (D)		Cys (C)		Gln (Q)		Glu (E)		Gly (G)		His (H)		Ile (I)		Leu (L)		Lys (K)	
PHPA_RSP_02	9	4.3%	3	1.4%	4	1.9%	0	0.0%	1	0.5%	5	2.4%	5	2.4%	70	33.7%	1	0.5%	10	4.8%	20	9.6%	6	2.9%
PHPA_RSP_03	6	4.8%	9	7.3%	4	3.2%	1	0.8%	1	0.8%	2	1.6%	4	3.2%	32	25.8%	0	0.0%	2	1.6%	2	1.6%	3	2.4%
PHPA_RSP_04	10	2.6%	3	0.8%	6	1.5%	4	1.0%	1	0.3%	5	1.3%	3	0.8%	151	38.6%	6	1.5%	29	7.4%	45	11.5%	6	1.5%

PHPA_RSP_05	26	11.3%	7	3.0%	12	5.2%	6	2.6%	7	3.0%	5	2.2%	13	5.6%	26	11.3%	5	2.2%	7	3.0%	25	10.8%	15	6.5%
PHPA_RSP_07	26	11.3%	7	3.0%	12	5.2%	6	2.6%	6	2.6%	5	2.2%	13	5.6%	26	11.3%	5	2.2%	7	3.0%	25	10.8%	15	6.5%
PHPA_RSP_12	8	8.2%	0	0.0%	2	2.1%	1	1.0%	0	0.0%	3	3.1%	0	0.0%	27	27.8%	9	9.3%	8	8.2%	3	3.1%	5	5.2%
PHPA_RSP_13	14	6.7%	10	4.8%	10	4.8%	12	5.8%	8	3.8%	10	4.8%	8	3.8%	17	8.2%	6	2.9%	10	4.8%	15	7.2%	11	5.3%
PHPA_RSP_43	9	9.9%	8	8.8%	6	6.6%	3	3.3%	0	0.0%	3	3.3%	3	3.3%	8	8.8%	0	0.0%	4	4.4%	5	5.5%	2	2.2%
PHPA_RSP_46	7	3.8%	3	1.6%	4	2.2%	2	1.1%	1	0.5%	6	3.2%	4	2.2%	53	28.5%	2	1.1%	10	5.4%	14	7.5%	4	2.2%
PHPA_RSP_52	7	4.1%	3	1.7%	4	2.3%	2	1.2%	1	0.6%	6	3.5%	4	2.3%	50	29.1%	2	1.2%	9	5.2%	14	8.1%	4	2.3%
PHPA_RSP_54	11	6.2%	2	1.1%	6	3.4%	2	1.1%	0	0.0%	6	3.4%	4	2.2%	51	28.7%	3	1.7%	9	5.1%	21	11.8%	4	2.2%
PHPA_55_Contig2083	16	9.4%	4	2.3%	8	4.7%	2	1.2%	1	0.6%	6	3.5%	4	2.3%	60	35.1%	0	0.0%	10	5.8%	13	7.6%	4	2.3%
PHPA_55_contig2923	16	9.5%	4	2.4%	8	4.7%	1	0.6%	1	0.6%	6	3.6%	6	3.6%	57	33.7%	0	0.0%	13	7.7%	10	5.9%	5	3.0%
PHPA_RSP_64	20	8.1%	7	2.8%	20	8.1%	10	4.0%	6	2.4%	15	6.1%	10	4.0%	29	11.7%	4	1.6%	10	4.0%	13	5.3%	10	4.0%
PHPA_RSP_65	13	6.0%	6	2.8%	11	5.0%	14	6.4%	0	0.0%	18	8.3%	12	5.5%	8	3.7%	8	3.7%	11	5.0%	18	8.3%	15	6.9%
PHPA_RSP_66	10	9.5%	6	5.7%	6	5.7%	5	4.8%	3	2.9%	3	2.9%	3	2.9%	11	10.5%	1	1.0%	6	5.7%	6	5.7%	2	1.9%
PHPA_RSP_67	24	11.4%	2	0.9%	11	5.2%	5	2.4%	2	0.9%	8	3.8%	1	0.5%	12	5.7%	1	0.5%	9	4.3%	17	8.1%	6	2.8%
PHPA_RSP_68	6	7.9%	3	3.9%	4	5.3%	5	6.6%	0	0.0%	1	1.3%	1	1.3%	8	10.5%	4	5.3%	8	10.5%	3	3.9%	4	5.3%
PHPA_RSP_69	6	4.1%	4	2.7%	1	0.7%	2	1.4%	7	4.8%	3	2.1%	0	0.0%	21	14.4%	3	2.1%	9	6.2%	6	4.1%	8	5.5%
PHPA_RSP_70	7	3.6%	13	6.7%	5	2.6%	3	1.5%	4	2.1%	9	4.6%	9	4.6%	53	27.2%	7	3.6%	2	1.0%	11	5.6%	9	4.6%
PHPA_RSP_71	15	9.2%	0	0.0%	12	7.4%	6	3.7%	3	1.8%	4	2.5%	1	0.6%	16	9.8%	2	1.2%	7	4.3%	6	3.7%	4	2.5%
PHPA_RSP_72	11	9.1%	4	3.3%	6	5.0%	7	5.8%	6	5.0%	3	2.5%	3	2.5%	12	9.9%	0	0.0%	9	7.4%	6	5.0%	10	8.3%
PHPA_RSP_73	7	4.1%	1	0.6%	6	3.6%	1	0.6%	1	0.6%	6	3.6%	5	3.0%	45	26.6%	2	1.2%	9	5.3%	17	10.1%	5	3.0%
PHPA_RSP_74	6	5.5%	4	3.6%	3	2.7%	0	0.0%	0	0.0%	5	4.5%	0	0.0%	32	29.1%	2	1.8%	9	8.2%	4	3.6%	0	0.0%
PHPA_RSP_75	22	10.2%	5	2.3%	7	3.2%	3	1.4%	4	1.9%	10	4.6%	2	0.9%	36	16.7%	4	1.9%	12	5.6%	24	11.1%	3	1.4%
PHPA_RSP_76	4	6.2%	5	7.7%	2	3.1%	2	3.1%	3	4.6%	3	4.6%	3	4.6%	6	9.2%	1	1.5%	3	4.6%	2	3.1%	2	3.1%
PHPA_RSP_77	8	7.6%	3	2.9%	3	2.9%	3	2.9%	0	0.0%	2	1.9%	6	5.7%	4	3.8%	4	3.8%	9	8.6%	21	20.0%	5	4.8%
PHPA_RSP_78	8	5.8%	8	5.8%	6	4.4%	9	6.6%	6	4.4%	4	2.9%	3	2.2%	14	10.2%	3	2.2%	8	5.8%	10	7.3%	7	5.1%
PHPA_RSP_79	5	3.3%	5	3.3%	8	5.3%	5	3.3%	3	2.0%	7	4.7%	12	8.0%	12	8.0%	0	0.0%	7	4.7%	8	5.3%	6	4.0%

Ala (A) – Alanina; Arg (R) – Arginina; Asn (N) – Asparagina; Asp (D) – Aspartato; Cys (C) – Cisteína; Gln (Q) – Glutamina; Gly (G) – Glicina; His (H) – Histidina; Ile (I) – Isoleucina; Leu (L) – Leucina.

Anexo 3 cont.: Composição de aminoácido pelo programa ProtParam

	Aminoácidos													Aminoácidos carregados										
	Met (M)	Phe (F)	Pro (P)	Ser (S)	Thr (T)	Trp (W)	Tyr (Y)	Val (V)	Pyl (O)	Sec (U)	Asp + Glu	Arg + Lys												
PHPA_RSP_02	6	2.9%	19	9.1%	25	12.0%	14	6.7%	4	1.9%	0	0.0%	1	0.5%	5	2.4%	0	0.0%	0	0.0%	5	9		
PHPA_RSP_03	3	2.4%	5	4.0%	17	13.7%	5	4.0%	2	1.6%	16	12.9%	2	1.6%	8	6.5%	0	0.0%	0	0.0%	5	12		
PHPA_RSP_04	4	1.0%	25	6.4%	65	16.6%	16	4.1%	6	1.5%	0	0.0%	0	0.0%	6	1.5%	0	0.0%	0	0.0%	7	9		

PHPA_RSP_05	5	2.2%	7	3.0%	7	3.0%	31	13.4%	7	3.0%	5	2.2%	6	2.6%	9	3.9%	0	0.0%	0	0.0%	19	22
PHPA_RSP_07	5	2.2%	7	3.0%	7	3.0%	31	13.4%	7	3.0%	5	2.2%	7	3.0%	9	3.9%	0	0.0%	0	0.0%	19	22
PHPA_RSP_12	2	2.1%	9	9.3%	2	2.1%	11	11.3%	1	1.0%	0	0.0%	1	1.0%	5	5.2%	0	0.0%	0	0.0%	1	5
PHPA_RSP_13	1	0.5%	8	3.8%	14	6.7%	12	5.8%	18	8.7%	3	1.4%	6	2.9%	14	6.7%	0	0.0%	0	0.0%	20	21
PHPA_RSP_43	3	3.3%	7	7.7%	7	7.7%	3	3.3%	4	4.4%	3	3.3%	7	7.7%	6	6.6%	0	0.0%	0	0.0%	6	10
PHPA_RSP_46	2	1.1%	13	7.0%	22	11.8%	17	9.1%	2	1.1%	0	0.0%	0	0.0%	20		0	0.0%	0	0.0%	6	7
PHPA_RSP_52	2	1.2%	11	6.4%	18	10.5%	17	9.9%	2	1.2%	0	0.0%	0	0.0%	16	9.3%	0	0.0%	0	0.0%	6	7
PHPA_RSP_54	4	2.2%	8	4.5%	15	8.4%	14	7.9%	7	3.9%	0	0.0%	1	0.6%	10	5.6%	0	0.0%	0	0.0%	6	6
PHPA_55_Contig2083	7	4.1%	14	8.2%	6	3.5%	11	6.4%	2	1.2%	0	0.0%	0	0.0%	3	1.8%	0	0.0%	0	0.0%	6	8
PHPA_55_contig2923	6	3.6%	13	7.7%	7	4.1%	10	5.9%	3	1.8%	0	0.0%	0	0.0%	3	1.8%	0	0.0%	0	0.0%	7	9
PHPA_RSP_64	6	2.4%	9	3.6%	11	4.5%	22	8.9%	11	4.5%	1	0.4%	4	1.6%	29		0	0.0%	0	0.0%	20	17
PHPA_RSP_65	3	1.4%	6	2.8%	11	5.0%	30	13.8%	18	8.3%	0	0.0%	4	1.8%	12	5.5%	0	0.0%	0	0.0%	26	21
PHPA_RSP_66	3	2.9%	6	5.7%	3	2.9%	15	14.3%	8	7.6%	0	0.0%	1	1.0%	7	6.7%	0	0.0%	0	0.0%	8	8
PHPA_RSP_67	3	1.4%	11	5.2%	13	6.2%	39	18.5%	29	13.7%	1	0.5%	4	1.9%	13	6.2%	0	0.0%	0	0.0%	6	8
PHPA_RSP_68	1	1.3%	4	5.3%	4	5.3%	8	10.5%	3	3.9%	1	1.3%	4	5.3%	4	5.3%	0	0.0%	0	0.0%	6	7
PHPA_RSP_69	2	1.4%	10	6.8%	20	13.7%	22	15.1%	7	4.8%	1	0.7%	0	0.0%	14	9.6%	0	0.0%	0	0.0%	2	12
PHPA_RSP_70	1	0.5%	9	4.6%	8	4.1%	14	7.2%	6	3.1%	12	6.2%	8	4.1%	5	2.6%	0	0.0%	0	0.0%	12	22
PHPA_RSP_71	7	4.3%	3	1.8%	21	12.9%	17	10.4%	23	14.1%	1	0.6%	2	1.2%	13	8.0%	0	0.0%	0	0.0%	7	4
PHPA_RSP_72	4	3.3%	9	7.4%	1	0.8%	12	9.9%	7	5.8%	1	0.8%	4	3.3%	6	5.0%	0	0.0%	0	0.0%	10	14
PHPA_RSP_73	4	2.4%	11	6.5%	20	11.8%	19	11.2%	2	1.2%	0	0.0%	0	0.0%	8	4.7%	0	0.0%	0	0.0%	6	6
PHPA_RSP_74	3	2.7%	7	6.4%	4	3.6%	15	13.6%	5	4.5%	6	5.5%	1	0.9%	4	3.6%	0	0.0%	0	0.0%	0	4
PHPA_RSP_75	4	1.9%	19	8.8%	8	3.7%	32	14.8%	5	2.3%	1	0.5%	0	0.0%	15	6.9%	0	0.0%	0	0.0%	5	8
PHPA_RSP_76	3	4.6%	6	9.2%	3	4.6%	4	6.2%	3	4.6%	2	3.1%	5	7.7%	3	4.6%	0	0.0%	0	0.0%	5	7
PHPA_RSP_77	2	1.9%	5	4.8%	4	3.8%	11	10.5%	4	3.8%	0	0.0%	2	1.9%	9	8.6%	0	0.0%	0	0.0%	9	8
PHPA_RSP_78	3	2.2%	6	4.4%	10	7.3%	13	9.5%	4	2.9%	2	1.5%	5	3.6%	8	5.8%	0	0.0%	0	0.0%	12	15
PHPA_RSP_79	4	2.7%	7	4.7%	11	7.3%	20	13.3%	10	6.7%	2	1.3%	11	7.3%	7	4.7%	0	0.0%	0	0.0%	15	11

Lys (K) – Lisina; Met (M) – Metionina; Phe (F) – Fenilalanina; Pro (P) – Prolina; Ser (S) – Serina; Thr (T) – Treonina; Trp (W) – Triptofano; Tyr (Y) – Tirosina; Val (V) – Valina; Pyl (O) – Prirrolisina; Sec (U) – Selenocistei