

RICARDO ROBERTO DA SILVA

**IDENTIFICAÇÃO DE GENES QUE CODIFICAM PROTEÍNAS
SECRETADAS POR *Puccinia psidii*, AGENTE ETIOLÓGICO DA
FERRUGEM DAS MIRTÁCEAS**

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
2010

RICARDO ROBERTO DA SILVA

**Identificação de genes que codificam proteínas secretadas por
Puccinia psidii, agente etiológico da ferrugem das mirtáceas**

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: 11 de fevereiro de 2010.

Prof. Eduardo Seiti Gomide Mizubuti
(Co-orientador)

Prof^a. Marisa Vieira de Queiroz
(Co-orientadora)

Prof. Maurício Dutra Costa

Dr. Lúcio Mauro da Silva Guimarães

Prof. Sérgio Hermínio Brommonschenkel
(Orientador)

AGRADECIMENTOS

Aos meus pais, Carlos e Maria Aparecida, por todo sacrifício para que continuasse estudando, pelo imenso carinho, e pelo apoio incondicional as minhas escolhas.

Ao meu irmão, Diego, e aos meus amigos, Charles, Diego “de Nilda” e Rogério, pelo companheirismo e momentos de descontração que, sem dúvida foram decisivos para que eu tivesse força para dar seguimento a este trabalho.

A todos os colegas do laboratório de Genômica, Luiz, Tadeu, Cynthia, Gustavo, Yukari, Daniela, Polyene, Jamile, Sandra, Michele, Livia e Guilherme, pelo convívio e pela amizade.

À minha namorada, Livia, por estar sempre ao meu lado, pelo amor, pelo carinho, e pelo constante incentivo.

Ao meu orientador, professor Sérgio Hermínio Bromonschenkel, pela oportunidade e orientação durante a execução deste trabalho.

Ao professor Acelino Couto Alfenas, pelas oportunidades que me propiciou, possibilitando-me o ingresso à carreira de pesquisador, e pela constante disposição em auxiliar.

Aos Doutores, Douglas, Elene e Klaus, e novamente aos professores, Sérgio e Acelino, pelas lições de bancada e principalmente, por cultivarem em mim o amor à ciência.

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

Ao CNPq pela bolsa concedida.

ÍNDICE

RESUMO	iv
ABSTRACT	vi
1 - INTRODUÇÃO	1
2 - REVISÃO BIBLIOGRÁFICA	4
2.1 - Aspectos gerais do patossistema <i>Eucalyptus</i> spp.- <i>Puccinia psidii</i>	4
2.2 - Efetores: descoberta e caracterização.....	10
3 – METODOLOGIA	18
3.1 - Construção e sequenciamento da biblioteca de cDNA	18
3.2 - Análises de bioinformática	19
3.3 - Análise da secreção das proteínas codificadas pelos genes selecionadas no sistema YST	20
3.4 – Amplificação das ORFs a partir do DNA genômico de <i>E. grandis</i> e <i>P. psidii</i>	22
3.5 - Expressão transiente dos genes que codificam as proteínas comprovadamente secretadas.....	23
4 - RESULTADOS	25
5 – DISCUSSÃO E CONCLUSÕES	37
6 - REFERÊNCIAS BIBLIOGRÁFICAS.....	47
ANEXO.....	55

RESUMO

SILVA, Ricardo Roberto da. M.Sc. Universidade Federal de Viçosa, Fevereiro de 2010. **Identificação de genes que codificam proteínas secretadas por *Puccinia psidii*, agente etiológico da ferrugem das mirtáceas.** Orientador: Sérgio Hermínio Brommonschenkel. Co-orientadores: Eduardo Seiti Gomide Mizubuti e Marisa Vieira de Queiroz.

A ferrugem causada pelo basidiomiceto *Puccinia psidii* é uma das doenças mais destrutivas da cultura do eucalipto no Brasil, e motivo de restrições sanitárias onde a doença ainda não ocorre na Ásia e Oceania. Durante a interação com o hospedeiro os fungos biotróficos, como as ferrugens, secretam proteínas efetoras envolvidas na supressão da resposta de defesa e na manipulação do metabolismo do hospedeiro. Uma estratégia que pode ser utilizada para descoberta dessas proteínas é a predição *in silico* de sequências de exportação celular em sequências abertas de leitura deduzidas a partir do sequenciamento de cDNAs. O presente trabalho teve por objetivo identificar e caracterizar sequências de cDNAs de *P. psidii* que codificam proteínas secretadas por meio da análise de sequências de uma biblioteca de cDNA construída a partir de mRNA isolado de folhas de *Eucalyptus grandis* infectadas por *P. psidii*. Foram analisadas 9864 etiquetas de sequências expressas (ESTs). A partir de 1675 transcritos que codificam proteínas sem similaridade a sequências protéicas do banco não redundante do GenBank/NCBI e que não possuíam identidade a sequências de cDNA ou sequências genômicas depositadas em bancos de dados de eucalipto, foram obtidas 173 ORFs que apresentaram predição positiva em cinco ou mais parâmetros do algoritmo de predição de sequências sinal de exportação SignalP. Foi selecionado um subgrupo de 31 clones com melhores valores de predição de exportação celular em diferentes algoritmos para caracterização. Vinte e nove dos clones selecionados tiveram a região da ORF amplificada do genoma de *P. psidii*, gerando fragmentos iguais ou maiores aos amplificados a partir do cDNA. Oito clones apresentaram similaridade a sequências de *Puccinia graminis* e cinco a sequências de *Melampsora laricis*, dentre as quais, quatro têm similares em ambos os bancos. Os clones selecionados também foram comparados entre si utilizando o programa BlastP, e apresentaram similaridade, variando desde pequenos motivos compartilhados a extensa identidade. Foi confirmada a secreção *in vitro* das proteínas codificadas por 24 clones pelo sistema de armadilha de sinal de secreção em levedura

(YST). Vinte clones tiveram sua ORF clonada em um vetor para expressão transiente mediada por *Agrobacterium*. Análises de expressão transiente e do monitoramento da expressão das proteínas secretadas estão sendo otimizados para investigação do papel efetor das proteínas hipotéticas identificadas.

ABSTRACT

SILVA, Ricardo Roberto da. M.Sc. Universidade Federal de Viçosa, February 2010. **Identification of genes that encode proteins secreted by *Puccinia psidii*, etiologic agent of Myrtaceae rust.** Adviser: Sergio Herminio Brommonschenkel. Co-advisers: Eduardo Gomide Seiti Mizubuti and Marisa Vieira de Queiroz.

The rust caused by the basidiomycete *Puccinia psidii* is one of the most destructive diseases of eucalypt plantations in Brazil, and cause for sanitary barriers where this disease does not occur in Asia and Oceania. During the interaction with the host the biotrophic fungi, such as rusts, secrete effector proteins involved in the suppression of the defense response and in the manipulation of the host metabolism. A strategy that can be used to discovery of these proteins is the *in silico* prediction of cellular sequences export in open reading frames deduced from cDNA sequencing. This study aimed to identify and characterize cDNA sequences of *P. psidii* that encode secreted proteins through sequence analysis of a cDNA library constructed from mRNA isolated from *Eucalyptus grandis* leaves infected by *P. psidii*. We analyzed 9,864 expressed sequence tags (ESTs). From 1,675 transcripts that encode proteins with no similarity to protein sequences of non-redundant GenBank/NCBI database and lacked identity to cDNA or genome sequences deposited in the eucalypt databases, 173 ORFs that showed positive prediction in five or more parameters of the export signal sequence prediction algorithm SignalP were obtained. A subgroup of 31 clones with better predictive value of cell export in different algorithms was selected for characterization. Twenty-nine of the clones had the ORF amplified from the genome of *P. psidii* generating equal or larger fragments than those amplified from the cDNA. Eight clones showed sequences similarity to *Puccinia graminis* and five *Melampsora laricis* sequences, among which, four have similar in both banks. The selected clones were also compared with each other using the program BlastP, and presented similarity, ranging from small motifs shared to extensive identity. Secretion of proteins encoded for 24 clones was confirmed *in vitro* by the yeast secretion trap (YST). Twenty clones had their ORF cloned into a vector for Agrobacterium-mediated transient expression. Analysis of transient

expression and expression monitoring of secreted proteins are being optimized to investigate the effector role of hypothetical proteins identified.

1 - INTRODUÇÃO

A ferrugem das mirtáceas, causada por *Puccinia psidii*, esta amplamente dispersa pelo Brasil, com relatos de sua ocorrência em pelo menos 10 dos 23 gêneros de mirtáceas nativas. Dentre os hospedeiros economicamente importantes destacam-se as espécies de *Eucalyptus*, para as quais a ferrugem é considerada a doença mais severa, com potencial de causar perdas em áreas tropicais e subtropicais que possuam florestas de eucalipto (Glen *et al.*, 2007). A existência de variabilidade genética inter e intra-específica para resistência à ferrugem têm permitido, até o presente, o controle da doença por meio do plantio de clones, progênies ou espécies resistentes (Alfenas *et al.*, 2004).

Puccinia psidii possui um ciclo de vida complexo, pois pode formar até quatro tipos de estruturas reprodutivas: écio, urédia, télia e basídia. O processo infeccioso da ferrugem pode ser dividido em três fases: a fase de penetração, durante a qual o patógeno tem acesso ao interior do hospedeiro; a fase biotrófica, durante a qual o patógeno estabelece as relações parasitárias com o hospedeiro e a fase reprodutiva ou de esporulação, durante a qual o fungo se reproduz no tecido infectado (Alfenas *et al.*, 2004). A fase de penetração tem início com a germinação do urediniósporo depositado na superfície do hospedeiro. Em condições adequadas de temperatura e umidade, o urediniósporo 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 forma uma estrutura especializada para a absorção de nutrientes denominada haustório, dando início à fase biotrófica (Alfenas *et al.*, 2004). Além de atuar na absorção de nutrientes a partir da célula do hospedeiro, o haustório tem função na manutenção da fase biotrófica, provavelmente por meio da supressão de respostas de defesa da planta (Catanzariti *et al.*, 2007). Aparentemente, a resistência à ferrugem no eucalipto está associada ao desencadeamento da resposta de hipersensibilidade, uma vez que em indivíduos resistentes, visualiza-se a necrose das células localizadas no sítio de penetração do patógeno (Xavier *et al.*, 2001). O mapeamento de um gene conferindo a herança monogenética dominante para a resistência em *E. grandis* foi relatado (Junghans *et al.*, 2003). O reconhecimento de proteínas efetoras, secretadas por fungos causadores de ferrugens, e o

desencadeamento de respostas de morte celular, foram relatados em outros patossistemas (Wang *et al.*, 2007), onde genes de resistência (genes R) e os genes que codificam os efetores por eles reconhecidos (genes Avr) já foram clonados.

O plantio em larga escala de clones ou variedades resistentes normalmente leva à seleção de estirpes do patógeno que conseguem suplantar a resistência conferida pelos genes R. Essas raças virulentas possuem mutações nos genes AVR que resultam em uma proteína AVR não mais reconhecida pela proteína R (Jones & Dangl, 2006). Em função dessa limitação dos genes R, existe grande interesse na busca de fontes alternativas de resistência (Alves, 2008). Assim, há uma expectativa de que a compreensão da fisiologia do parasitismo e dos processos moleculares envolvidos na interação da ferrugem com o hospedeiro possa fornecer informações importantes para a compreensão dos mecanismos envolvidos na resposta de defesa.

Diversos estudos foram conduzidos para descoberta e caracterização de moléculas efetoras utilizadas pelos patógenos para manipular o metabolismo e evadir a resposta de defesa dos hospedeiros. Os resultados ressaltam o papel dos efetores na reprogramação dos mecanismos de defesa da célula vegetal e na virulência do patógeno e demonstram que a caracterização de efetores é base fundamental para a identificação dos determinantes envolvidos na virulência e especificidade de hospedeiros das ferrugens. Um exemplo recente é a proteína AVR2 (Six3) de *Fusarium oxysporum* f.sp. *lycopersici* que é secretada no xilema, mas reconhecida pela proteína R citoplasmática I2 de tomate (Houterman *et al.*, 2009). Evidências diretas para a translocação de proteínas efetoras de fungos, causadores de ferrugem, para células hospedeiras foram fornecidas por Kemen *et al.* (2005) utilizando imunolocalização. Whisson *et al.* (2007) utilizaram transformantes de *Phytophthora infestans*, com a sequência do motivo RXLR-dEER modificada e com a fusão protéica RXLR-dEER-GusA, para mostrar o envolvimento dos motivos RXLR-dEER na captação de AvrR3a pelo hospedeiro e da translocação da fusão protéica para células do hospedeiro infectadas com haustório (Ellis *et al.*, 2009). A análise em larga escala de genes expressos no haustório da ferrugem do linho, *Melampsora lini*, levou à identificação 21 proteínas secretadas (HESP, *haustorium-specific protein*, Catanzariti *et al.*, 2006).

O presente trabalho teve por objetivo identificar genes de *P. psidii* que possuem sequências de secreção celular, a partir da dedução de ORFs contidas em

sequências de cDNAs provenientes de uma biblioteca de folhas de eucalipto infectadas, que não possuem identidade a sequências dos bancos de dados do hospedeiro. Tais ORFs representam candidatos a efetores, e podem possibilitar a melhor compreensão das estratégias de infecção do fungo, subsidiando o desenvolvimento de variedades com resistência mais durável e, ou, de novas estratégias de controle da doença.

2 - REVISÃO BIBLIOGRÁFICA

2.1 - Aspectos gerais do patossistema *Eucalyptus* spp.-*Puccinia psidii*

Espécies arbóreas do gênero *Eucalyptus* figuram entre as mais plantadas do mundo, em função do seu crescimento acelerado, ampla adaptabilidade e propriedades da madeira úteis a diferentes indústrias. As florestas de alta produtividade de eucalipto suprem matéria prima de alta qualidade para fabricação de polpa, papel, energia e uma ampla gama de produtos associados à cadeia produtiva da madeira que de outra forma seriam extraídas de florestas nativas (Grattapaglia & Kirst, 2008). Com destaque histórico, as áreas de florestas plantadas com eucalipto no Brasil acumularam em 2008 o total estimado de 4.258.704 ha, constatando-se crescimento de 7,3 % na área plantada em relação a 2007 (ABRAF, 2009).

Fungos e bactérias são os principais agentes patogênicos que causam doenças em espécies florestais como o eucalipto (Alfenas *et al.*, 2004). Alguns dos fungos fitopatogênicos mais agressivos são parasitas biotróficos obrigatórios, como os causadores das ferrugens. O termo biotrófico caracteriza um estilo de parasitismo peculiar em que o patógeno causa o mínimo de dano, permitindo ao hospedeiro se manter vivo para que possa completar o ciclo de vida. O fungo biotrófico obrigatório diferencia-se dos fungos hemibiotróficos ou necrotróficos por apresentar: (a) estruturas de infecção altamente desenvolvidas; (b) atividade secretória limitada, especialmente de enzimas líticas; (c) uma zona de contato estreita entre a membrana do fungo e a da planta; (d) supressão em longo prazo da resposta de defesa da planta; (e) formação de haustório (Mendgen & Hahn, 2002).

Os fungos causadores de ferrugens (ordem: Pucciniales; filo: Basidiomycota) compreendem mais de 100 gêneros e por volta de 7000 espécies. O gênero *Puccinia* é constituído por mais de 4000 espécies, seguido pelo gênero *Uromyces* com mais de 600 espécies (Maier *et al.*, 2003). *Puccinia psidii*, descrito pela primeira vez por Winter em 1884, em *Psidium pomiferum*, no Brasil, possuía sua localização limitada às Américas do Sul e Central e ao Caribe, até o seu aparecimento na Flórida em 1977. Também houve relatos da ocorrência de *P. psidii* em Taiwan, em 1992, onde o patógeno nunca foi definitivamente identificado e não parece ter se estabelecido, e na

África do Sul, infectando *E. nitens*, em 1990, onde as características morfológicas descritas não foram compatíveis as de *P. psidii* e a mesma detecção não se repetiu posteriormente (Glen *et al.*, 2007). Além destes relatos incomuns, até abril de 2005 não houve evidência de dispersão adicional de *P. psidii*, quando o patógeno foi detectado em mirtáceas nativas no Havaí (Glen *et al.*, 2007). Em 2007 a ocorrência de *P. psidii* foi verificada no Japão incidindo em Ohia (*Metrosideros polymorpha*, Mirtaceae) (Kawanishi *et al.*, 2009).

O fungo *Puccinia psidii* está amplamente disperso pelo Brasil com relatos de sua ocorrência em pelo menos 10 dos 23 gêneros de mirtáceas nativas, e constitui o único patógeno confirmadamente capaz de causar os sintomas de ferrugem em espécies de eucalipto (Glen *et al.*, 2007). Em 1912, o fungo foi observado em *Eucalyptus citriodora*, mas somente em 1944 foi formalmente descrito como causador da doença neste hospedeiro (Coutinho *et al.*, 1998). A primeira ocorrência da ferrugem no Brasil causando danos consideráveis foi no Espírito Santo na década de 1970, em plantios de *Eucalyptus grandis* com idade inferior a dois anos e procedentes da África do Sul. Índices elevados de infecção de *Eucalyptus* spp. também foram verificados tanto em viveiros como em plantações na região do Vale do Rio Doce em Minas Gerais e Espírito Santo e no sul da Bahia (Coutinho *et al.*, 1998). O *Eucalyptus grandis*, espécie amplamente utilizada para confecção de híbridos clonais plantados em larga escala pela indústria de celulose e papel no Brasil, é suscetível a *P. psidii*, embora fontes de resistência venham sendo identificadas em populações de *Eucalyptus* sp. (Junghans *et al.*, 2003; Alves, 2008).

A ferrugem do eucalipto tem como principal sinal para sua diagnose a esporulação urediniospórica pulverulenta e de coloração amarela sobre os órgãos afetados. Em material altamente suscetível causa deformações, necrose, hipertrofia, minicancros e morte das brotações terminais. As espécies de *Eucalyptus* mais suscetíveis à ferrugem são *Eucalyptus grandis*, *E. cloeziana*, *E. phaeotricha*, *E. globulus* e *E. nitens*, nas quais o fungo ocorre em plantas jovens em viveiros e no campo. Além do eucalipto, a ferrugem ocorre em várias outras espécies nativas ou exóticas da família mirtáceas, como goiabeira, jambeiro, jabuticabeira, jamelãozeiro, melaleuca e calistemon, entre outras. Nesses hospedeiros, além de tecidos vegetativos meristemáticos, o fungo infecta botões florais e frutos em desenvolvimento, podendo ocasionar perdas significativas (Alfenas *et al.*, 2004).

Quando folhas de eucalipto que já atingiram 50% de seu crescimento total são infectadas por *P. psidii*, esporulações menos intensas são observadas e estas ficam restritas a superfície inferior da folha (Coutinho *et al.*, 1998). As temperaturas na faixa de 18-25° C (ótima a 23° C), períodos prolongados de molhamento foliar (orvalho noturno ou garoas por períodos superiores a 6 h por 5-7 dias consecutivos) e a existência de órgãos juvenis (folhas jovens e terminais em crescimento) são condições favoráveis à infecção (Alfenas *et al.*, 2004).

As ferrugens são pleomórficas podendo apresentar vários tipos de estruturas esporogênicas e esporos com diferentes morfologias e funções, que são produzidos nos diversos estágios ou fases de seus ciclos vitais, por vezes extremamente confusos. Essas estruturas manifestam-se em diferentes épocas do ano (estacionais). Mesmo as ferrugens que requerem dois hospedeiros diferentes para completar o seu ciclo (heteróicas), podem sobreviver em um único hospedeiro por meio da reprodução vegetativa ou clonal (urediniósporos). No caso das ferrugens denominadas autóicas, em que todos os estados esporíferos manifestam-se em um mesmo hospedeiro, a determinação dos ciclos de vida deveria ser mais simples, se não fosse a grande variabilidade existente em certas espécies. Nestas, a morfologia das estruturas esporíferas e esporos não coincidem com as funções normalmente desempenhadas no ciclo vital das ferrugens mais bem estudadas. Assim, o pleomorfismo, a complexidade dos ciclos de vida e as dificuldades inerentes aos estudos de parasitas biotróficos são os principais fatores que dificultam as pesquisas nessa área (Figueiredo & Passador, 2008).

Poucos estudos foram conduzidos para determinar a morfologia em diferentes estádios do ciclo de vida de *P. psidii*. Sabe-se que trata-se de uma ferrugem autóica com ciclo vital macrocíclico (**Figura 1**), com as seguintes fases: aécia (Estadio I), uredínia (Estadio II), télia (Estadio III) e basídia (Estadio IV). Ainda não foi encontrado o estadio de espermácia (Estadio 0), mas devido a ocorrência de aécia, acredita-se que esta fase ocorra na natureza. Os teliósporos possuem coloração marrom escuro e são pedicelados, bicelulares, possuem forma de cubo, com papilas achatadas na parede distal para a parede pedicelar. As pústulas consistem em uma mistura de teliósporos e urediniósporos. Ambas as células do teliósporo germinam para formar a basídia em condições de umidade favorável, apesar de relatos afirmarem que apenas 8 % dos teliósporos germinam. As basídias possuem 40 a 70

µm de tamanho e apenas 0,1 % produzem basidiósporos. Os aeciósporos, por definição, são os primeiros esporos formados após a germinação dos basidiósporos e possuem o mesmo formato dos urediniósporos, porém em ambos o formato pode apresentar variações. Os aeciósporos têm sido encontrados somente em *S. jambos*. Teliósporos e basidiósporos são raramente encontrados em eucaliptos infectados naturalmente, exceto durante os meses mais quentes do ano (dezembro a março). Muitos teliósporos são produzidos em eucalipto e jambo inoculados artificialmente, e a formação de aeciósporos ocorre entre o 10º e 18º dia após a infecção. Sabe-se também que os basidiósporos são infectivos apenas para jambo, e apenas a temperaturas superiores a 25°C. Os urediniósporos são as estruturas infectivas mais comuns (Coutinho *et al.* 1998).

As uredínias são formadas a partir do micélio dicariótico e formam esporos dicarióticos, denominados urediniósporos. Estes se originam por simples segmentação da parte final de uma hifa não ramificada e que é freqüentemente referida como pedicelo. Uma segunda geração de uredínias e de urediniósporos pode resultar de uma infecção por urediniósporos e são, por isso, chamadas estruturas repetitivas, funcionando como conídios. Esse tipo de reprodução é também chamada de anamórfica ou clonal, por repetir exatamente as características genéticas da geração anterior (Figueiredo & Passador, 2008).

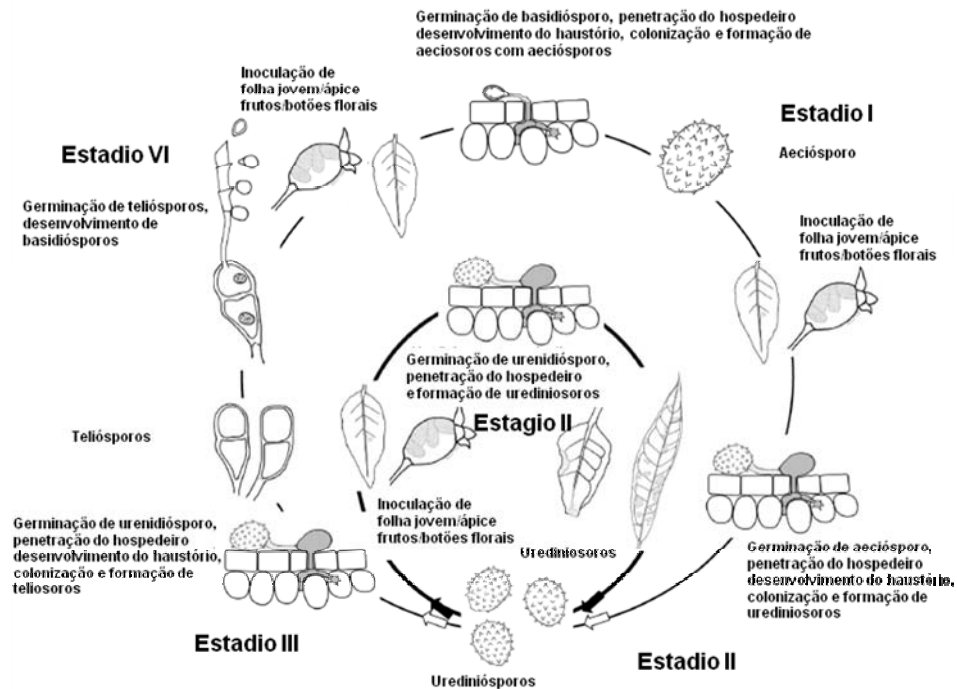


Figura 1 – Ciclo de vida esquemático de *Puccinia psidii*. Adaptado de Glen *et al.*, 2007.

Durante a infecção, o uredinióspero germina na superfície da planta e produz um tubo germinativo que se adere à cutícula (Xavier *et al.*, 2001). Características topográficas, e possivelmente também sinais químicos, induzem a formação do apressório que permite ao fungo penetrar diretamente no hospedeiro. A penetração ocorre geralmente entre duas células epidérmicas ou, menos comumente, nas extremidades entre três células epidérmicas (Coutinho *et al.*, 1998). Etapas posteriores da infecção vêm sendo descritas em espécies do gênero *Puccinia*, onde uma hifa de infecção emerge da vesícula iniciando o crescimento no espaço intracelular do tecido do hospedeiro, e um par de núcleos migra para esta hifa. Quando a extremidade da hifa alcança a parede celular da epiderme de células hospedeiras, a hifa pára de crescer e a célula mãe de haustório é formada. Esta célula é separada da hifa por um septo e contém de 2-4 núcleos. Uma extremidade da célula-mãe do haustório se desenvolve em direção à parede da célula hospedeira, e forma o haustório no interior da célula (Xavier *et al.*, 2001; Larous *et al.*, 2008). Durante a emissão dos haustórios no interior das células do hospedeiro ocorre a dissolução da parede celular, seguida da invaginação da membrana celular do

hospedeiro, que forma uma membrana extra-haustorial (**Figura 2, EHM**) ao redor do haustório, sendo esta desprovida de parede celular. O haustório é circundado por uma membrana plasmática (**Figura 2, EM**) e parede celular.

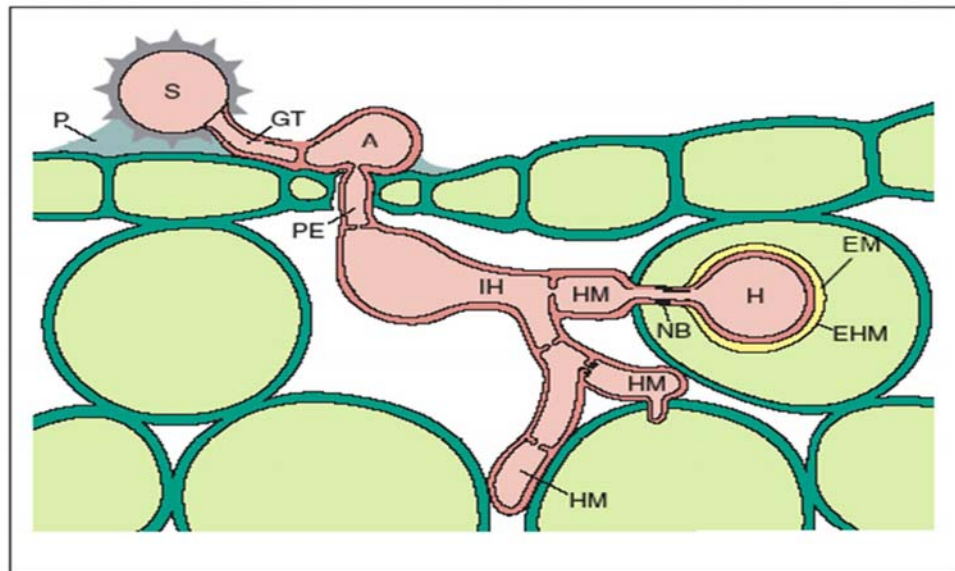


Figura 1 - Representação esquemática das estruturas infectivas produzidas pelas ferrugens na fase de penetração e biotrófica. Um tubo germinativo (GT) emerge a partir de um urediniósporo (S) aderido ao hospedeiro pelo *pad* de adesão (P). Depois do reconhecimento um apressório (A) se desenvolve. A hifa de penetração (PE) penetra e se alonga formando uma hifa de infecção. Quando a extremidade da hifa de infecção contata a parede celular do hospedeiro uma célula mãe do haustório (HM) é formada, a partir da qual o haustório (H) invade a célula do hospedeiro. O haustório dicariótico possui características únicas, em preto, uma estrutura semelhante a uma estria de caspary, conhecida como *neckband*, e a matriz extra-haustorial (amarelo) circundada pela membrana extra-haustorial (EHM). Fonte: Mendgen & Hahn (2002).

Os haustórios constituem o principal sítio de absorção de nutrientes a partir do hospedeiro. 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, por exemplo, rearranjo do citoesqueleto, migração do núcleo e condensação da cromatina (Mendgen & 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 é corroborada pelos estudos recentes 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*). 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. Resultados similares foram obtidos para as ferrugens *Uromyces fabae* e *U. striatus*, que infectam *Vicia faba* e *Medicago truncatula*, respectivamente, por Kemen *et al.* (2005). Esses autores relataram a identificação de novas proteínas haustoriais (Uf-RTP1p e sua homóloga Us-RPT1p) que foram imunolocalizadas na matriz extrahaustorial e no citoplasma vegetal. Estes estudos demonstram que a caracterização do secretoma de *P. psidii* é fundamental para a identificação de efetores envolvidos na virulência e especificidade de hospedeiros desse patógeno.

2.2 - Efetores: descoberta e caracterização

Em geral, patógenos empregam um complexo arsenal molecular para penetrar e colonizar os tecidos da planta hospedeira (Stergiopoulos & de Wit, 2009). Por sua vez, as plantas desenvolveram ao longo do processo coevolutivo, um sistema imune capaz de detectar as atividades de patógenos e ativar as respostas de defesa. Esse sistema é constituído por duas linhas de defesa principais: uma que atua antes da invasão da célula pelo patógeno, ou seja, numa etapa extracelular da interação planta-patógeno e a outra que atua após a secreção de moléculas efetoras (antes denominadas de virulência, avirulência, elicitoras ou toxinas) no citoplasma da célula vegetal (Stergiopoulos & de Wit, 2009). A linha de defesa que ocorre ainda na etapa extracelular após a detecção de padrões moleculares associados a patógenos (PAMPs, *Pathogen-Associated Molecular Patterns*) por proteínas receptoras localizadas na membrana plasmática, é denominada imunidade ativada por PAMPs (PTI, *PAMP Triggered Immunity*). Esta linha de defesa é suplantada pelos patógenos por meio da translocação de moléculas efetoras para o citoplasma vegetal. Em

contrapartida, as plantas evoluíram uma segunda linha de defesa baseada nas proteínas codificadas pelos genes R que são capazes de reconhecer diretamente ou indiretamente os efetores ou a sua interferência em proteínas específicas na célula e ativar as respostas de defesa que restringem o desenvolvimento do patógeno ao local inicial de infecção. Esta segunda linha de defesa é denominada de imunidade ativada por efetores (ETI, *Effector Triggered Immunity*) e os genes que codificam efetores que são reconhecidos pelas proteínas R são comumente denominados genes de avirulência, ou genes AVR (Stergiopoulos & de Wit, 2009). A existência de uma relação um-a-um entre genes de ataque e de defesa, respectivamente, no patógeno e no hospedeiro foi brilhantemente demonstrado por Flor (1956) em seus estudos genéticos no patossistema linho-*Melampsora lini*. Percebe-se ainda que pela atuação nas duas etapas da resposta imune das plantas, as moléculas efetoras desempenham um papel fundamental do fenótipo das interações patógeno-hospedeiro, determinando inclusive a gama de plantas hospedeiras ou não-hospedeiras de um determinado patógeno. Esta íntima interação molecular leva a uma coevolução entre patógeno e planta, durante a qual patógenos respondem por mutação ou perda de efetores, ou através do desenvolvimento de novos efetores que podem evitar ou reprimir ETI, enquanto as plantas desenvolvem novas proteínas R mediando reconhecimento de novos efetores (Dangl & McDowell, 2006).

As moléculas efetoras podem ser secretadas na interface intercelular entre o patógeno e a planta ou translocadas para o citoplasma da célula hospedeira (Lucas, 2004). As bactérias gram-negativas desenvolveram o sistema de secreção de proteínas do Tipo III (TTSS) para translocação de proteínas de virulência através da parede celular e membrana plasmática do hospedeiro lançando-as diretamente no citoplasma. 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). Em *Pseudomonas syringae* uma única cepa geralmente codifica 30 diferentes efetores, no entanto, diferentes cepas de *P. syringae* possuem um conjunto de efetores únicos. Muitos destes efetores atuam de maneira redundante, de forma que a perda de um destes genes tem um impacto insignificante na agressividade do patógeno (Cunnac *et al.*, 2009).

Em contraste, pouco se sabe sobre o secretoma ou a natureza de potenciais efetores de fitopatógenos microbianos eucarióticos. A clonagem molecular do primeiro gene *Avr* bacteriano foi relatada em 1984 (Staskawicz *et al.*, 1984) enquanto o primeiro gene *Avr* fúngico foi clonado somente em 1991 (van Kan *et al.*, 1991) e, em 2004, foi clonado o primeiro gene *Avr* de um oomiceto (Tian *et al.*, 2004). Embora a maioria das proteínas *Avr* de fitopatógenos eucarióticos que foram estudadas até o momento sejam apoplásticas (Rep, 2005), recentemente foi demonstrado que, de modo similar a fitopatógenos bacterianos, algumas proteínas *Avr* de fungos e oomicetos são translocados para o citoplasma vegetal (Armstrong *et al.*, 2005; Rehmany *et al.*, 2005; Catanzariti *et al.*, 2006). Um exemplo recente é a proteína AVR2 (Six3) de *Fusarium oxysporum* f.sp. *lycopersici* que é secretada no xilema, mas reconhecida pela proteína R citoplasmática I2 de tomate (Houterman *et al.*, 2009). Evidência direta para a translocação do efector RTP1 da ferrugem *Uromyces fabae* para células hospedeiras foram fornecidas por Kemen *et al.* (2005) utilizando imunolocalização. 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. Estas novas descobertas ilustram a complexidade dos secretomas de fitopatógenos e seus hospedeiros e, embora componentes individuais desses secretomas estejam sendo revelados por ensaios específicos ou triagens moleculares, o progresso é limitado e, claro, ainda há muito a ser descoberto. Por analogia ao notável progresso que foi feito na caracterização do secretoma de fitobactérias, um catálogo mais extenso dos secretomas de fitopatógenos eucarióticos/planta hospedeira forneceria novos esclarecimentos em processos apoplásticos e na extensão e natureza do conjunto de efetores secretados.

A análise do genoma completo de fungos fitopatógenos suporta a importância especial das proteínas secretadas, e programas visando a descoberta e identificação de genes que codificam proteínas extracelulares foram iniciados com sucesso (Torto *et al.*, 2003; Lucas, 2004; Dean *et al.*, 2005). Link & Voegelé, 2008, estimaram que o secretoma de *Uromyces fabae* é composto de cerca de 200 proteínas. No entanto, ressaltam os autores, o número total de proteínas secretadas de *U. fabae* só poderia ser inferido com precisão a partir de estudos genômicos, ou investigações adicionais de todos os estágios de desenvolvimento do fungo. Também foram estimados os

secretomas dos fungos *Magnaporthe grisea*, para o qual foi predita a secreção de 739 proteínas, e do saprófita *Neurospora crassa* com cerca de metade desse número (Dean *et al.*, 2005). O secretoma de *Candida albicans* compreende menos de 300 proteínas (Lee *et al.*, 2003), e Catanzariti *et al.* (2006) identificaram apenas 20 proteínas secretadas expressas no haustório (HESPs) da ferrugem *Melampsora lini*. A seqüência completa do genoma do fungo biotrófico *Ustilago maydis*, publicada recentemente, revelou a presença de 12 conjuntos de genes que codificam proteínas secretadas, compreendendo cerca de 20% das proteínas secretadas de *U. maydis*, que são em parte co-reguladas e estão envolvidas na patogenicidade, de uma forma semelhante a ilhas de patogenicidade bacterianas. A descoberta de que a deleção desses blocos gênicos afeta a virulência em cinco casos apoia a importância das proteínas extracelulares e indica que o foco sobre as proteínas secretadas promete ser um instrumento para aumentar a compreensão das estratégias de patogênese de fungos fitopatogênicos (Kamper *et al.* 2006).

Lee *et al.* (2006) utilizando o sistema de armadilha de sinal de secreção em levedura (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 seqüências de genes previamente caracterizados. Os transcritos de todos os genes identificados só foram detectados em folhas infectadas e não em micélio cultivado *in vitro*. Link & Voegelé (2008) utilizando o sistema YST identificaram 62 genes do fungo *Uromyces fabae*, agente etiológico da ferrugem de *Vicia faba*, que codificam proteínas secretadas pelo haustório e 42 genes que codificam proteínas secretadas em esporos germinados. Apenas quatro genes eram idênticos em ambas as bibliotecas, indicando uma forte regulação fase específica na secreção de proteínas. Foi observada similaridade a proteínas previamente identificadas para 39 das seqüências analisadas, 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 na virulência e especificidade de hospedeiro exclusivo para Uredinales.

Outra abordagem utilizada para caracterização de efetores em larga escala é a clonagem de fragmentos genômicos diretamente no sistema de armadilha de secreção em levedura. Müller *et al.*, (2008), fragmentaram aleatoriamente o DNA genômico da estirpe haplóide selvagem Um521 (a1b1) de *U. maydis*, em fragmentos do

tamanho médio desejado de ~300 bp, que foram clonados no vetor pRK18 (Klein *et al.* 1996). O seqüenciamento revelou 52 clones de DNA genômico únicos. Para testar o papel destes genes na infecção, mutantes nulos (*knock-outs*) foram induzidos para diferentes genes. Enquanto *knock-outs* únicos para os genes de hidrofobinas e proteínas tipo-repelente não afetaram a patogenicidade, foi observado um forte efeito de um *knock-out* duplo de *hum3* e *rsp1*. O crescimento tipo levedura, acasalamento, formação de hifas aéreas e hidrofobicidade de superfície não foram afetados neste duplo mutante. No entanto, o desenvolvimento patogênico em planta logo após a penetração foi afetado, levando a uma completa perda de patogenicidade.

Recentemente, Mosquera *et al.* (2009) empregaram a técnica de microarranjo para descobrir sequências diferencialmente expressas em células de arroz infectadas, identificando genes de *Magnaporthe oryzae* que codificam proteínas secretadas durante o crescimento biotrófico. Por meio da análise de transformantes de *M. oryzae* que expressavam as proteínas em estudo, fusionadas à uma proteína fluorescente, as proteínas identificadas assim como as proteínas de avirulência AvrPita1 e Pw12 foram localizadas em uma região distinta do espaço extra-invasivo de hifas, conhecida como o complexo interfacial biotrófico. Postula-se que este complexo funcione como uma base de montagem de efetores secretados destinados ao transporte para o interior da célula vegetal, embora a detecção direta destas proteínas nas células do hospedeiro ainda não tenha sido relatada (Ellis *et al.*, 2009).

A construção de bibliotecas de cDNA a partir de RNA proveniente do apressório maduro, formado *in vitro*, foi utilizada por Kleemann *et al.*, (2008) trabalhando no sistema *Colletotrichum higginsianum*-*Arabidopsis thaliana*. Os autores demonstraram, por análises de RT-PCR, que sete genes que codificam proteínas secretadas de função desconhecida, incluindo dois genes específicos de *Colletotrichum*, são induzidos em apressórios e expressos precocemente durante a infecção da planta, e, portanto, representam efetores candidatos. *Colletotrichum higginsianum* pode ser cultivado axenicamente e estavelmente transformado usando uma variedade de métodos, e, também, é um organismo haplóide com conídios uninucleados, o que facilita a análise mutacional. Este patossistema, portanto, fornece um modelo atraente para estudar a biotrofia, em que ambos os parceiros na interação são geneticamente tratáveis.

Bowen *et al.*, (2009) trabalhando com o fungo hemibiotrófico *Venturia inaequalis* Cooke (Wint.) causador de doenças em membros da família Maloideae, contruíram diferentes bibliotecas de cDNA para triagem de candidatos a efetores. Os autores descreveram dezesseis novos genes candidatos efetores de *V. inaequalis* com base em características comuns aos genes efetores caracterizados a partir fungos filamentosos, ou seja, codificam proteínas pequenas, novas, ricas em cisteína e com um peptídeo sinal putativo. Três dos 16 candidatos, em particular, mostraram níveis significativos de indução da transcrição durante o crescimento em planta.

Dentre os efetores melhor caracterizados até o momento figuram as proteínas Avr2 e Avr4 de *Cladosporium fulvum* e Avr3 e Avr1 de *Fusarium oxysporum*. *Cladosporium fulvum* (*syn. Passalora fulva*) é um fungo fitopatogênico extracelular que infecta tomate. O gene Avr2 de *C. fulvum* codifica uma pré-proteína de 78 aminoácidos (aa), que gera uma proteína madura de 58 aa, com oito resíduos de cisteína, e que induz HR em plantas de tomate que carregam o gene de resistência cognato Cf-2 (Luderer *et al.*, 2002). Durante a infecção, Avr2 inibe pelo menos quatro cisteína proteases de tomate, incluindo Rcr3, Pip1, aleurain e TDI65 e, como tal, contribui para virulência do patógeno, por inibir proteases que são importantes para defesa do hospedeiro (Shabab *et al.*, 2008). O gene Avr4 codifica uma pré-proteína de 135 aa que é processada nas extremidades Carboxi e N-terminal após a secreção no apoplasto em uma proteína madura de 86 aa com oito resíduos de cisteína (Joosten *al.*, 1997). Com base no seu padrão de ligação dissulfeto, Avr4 tem semelhança estrutural a proteínas com um domínio de ligação a quitina de invertebrados, como *tachycitin* (inv ChBD) (Shen & Jacobs-Lorena, 1999). A ligação da proteína Avr4 à quitina foi confirmada experimentalmente, e, além disso, foi demonstrado que Avr4 pode proteger fungos quitinosos, tal como *Trichoderma viride* e *Fusarium solani*, contra a ação de quitinases básicas vegetais (Van Den Burg *et al.*, 2006).

Fusarium oxysporum f. sp. *lycopersici* (Fol) é um patógeno extracelular que coloniza os vasos do xilema de tomate. Quatro proteínas pequenas produzidas durante a infecção, denominadas Six1 a Six4 (*secreted in xylem*), foram identificadas em Fol. Six1, atualmente renomeada como Avr3, uma proteína de 284 aa é processada em uma proteína madura de 189 aa e é necessária para a virulência completa em tomate, mas também se comporta como um fator de avirulência que

desencadeia HR, na presença do gene de resistência cognato *I-3* (Rep *et al.*, 2004). A presença de *Avr3* é necessária para a completa virulência em tomate, e mutantes em *Avr3* que evitam o reconhecimento por *I-3*, mas continuam a ser virulentos, não foram relatados. Levantou-se, então, a questão de como cepas de *Fol* poderiam superar a resistência mediada por *I-3*, uma vez que a perda de *Avr3* representaria uma séria penalidade no valor adaptativo. A caracterização funcional recente de *Avr1* (anteriormente conhecido como *Six4*) forneceu uma resposta a esta pergunta. *Avr1* é uma pequena proteína de 242 aa, processada em uma proteína madura de 184 aa, rica em cisteína, secretada, que confere a avirulência a raças de *Fol* em linhagens de tomate que carregam os genes de resistência *I* ou *I-1*. Entretanto, ao contrário de *Avr3*, *Avr1* não é exigida para a virulência completa de estirpes de *Fol* em plantas de tomate que não possuem o gene R cognato. Em contraste, se propõe que *Avr1* atue como um supressor da resistência mediada por *I-2* e *I-3* (Houterman *et al.*, 2008). O mecanismo molecular envolvido nesta supressão ainda não é conhecido (Houterman *et al.*, 2008).

Recentemente, um exemplo muito elegante do envolvimento de um efetor na virulência e estabelecimento da biotrofia foi demonstrado para a proteína *Pep1* no sistema *Zea mays-Ustilago maydis* (Doehlemann *et al.*, 2009). *Pep1* é uma proteína de 178 aa secretada, contendo 4 cisteínas, codificada por um gene de cópia única expresso somente durante a infecção, característica típica de muitos efetores sem similares em bancos de dados. A análise de microscopia confocal das fusões funcionais *Pep1-GFP* e *Pep1-mcherry* em transformantes de *U. maydis* indicaram que a proteína é expressa e secretada pelas hifas fúngicas no apoplasto do hospedeiro, mas não foi detectada a sua transferência para o citoplasma do hospedeiro. *U. maydis* penetra as paredes da célula hospedeira, permanecendo separado do citoplasma do hospedeiro por sua membrana e cruza de uma célula a outra por novos eventos de rompimento da parede celular. *Pep1* é expressa em níveis elevados no estágio de passagem de célula a célula e é absolutamente necessária para a virulência, uma vez que mutantes de deleção de *Pep1* penetram a parede celular no local inicial da infecção, mas a infecção não progride. Usando um gene sintético com um promotor que expressou *Pep1* nesta fase inicial, mas não mais tarde, os autores também mostram que *Pep1* é necessária durante o movimento célula a célula de hifas no interior das folhas de milho. Assim, *Pep1* é necessária para biotrofia e mutantes induziram (ou não suprimiram) a expressão de um conjunto de genes de defesa do

hospedeiro, bem como morte celular localizada nos sítios de infecção. Este relato está em consonância com o fato de a morte das células do hospedeiro ser uma reação padrão à tentativa de penetração de fungos na ausência de efetores que suprimam a ativação desta morte celular. A ausência de assinaturas de seqüências de aminoácidos que indicam uma função enzimática em Pep1 e muitos outros efetores, que muitas vezes têm menos que 200 resíduos de aminoácidos, sugere que esses efetores podem atuar tanto como inibidores/modificadores da função de proteínas do hospedeiro quanto como moléculas que interagem com DNA, afetando a expressão dos genes de defesa (Ellis *et al.*, 2009).

O avanço recente na descoberta e caracterização de efetores fornece expectativas animadoras da aplicação a curto prazo em programas de melhoramento. Vleeshouwers *et al.* (2008) trabalhando no patossistema *Phytophthora infestans-Solanum tuberosum*, realizaram um trabalho pioneiro de utilização da expressão transiente em larga escala, para a identificação de genes efetores candidatos para função de avirulência em um conjunto de diferentes genótipos de batata. Pelo fato de as interações planta-patógeno seguirem o modelo de interação gene-a-gene, a descoberta de efetores se traduz também na descoberta dos genes R correspondentes, alguns dos quais podem ser novos e úteis aos programas de melhoramento. Especificamente, este trabalho utiliza genes de *P. infestans*, diversas espécies hospedeiras de *Solanum* e um bioteste que detecta a morte celular em resposta a efetores, dependendo do genótipo do hospedeiro. A análise genética mostrou que a resposta ao efector ipi0 no hospedeiro co-segrega com a resistência de hospedeiros a estirpes de *P. infestans* expressando ipi0. Os autores demonstraram que o bioensaio de efetores pode mostrar que genes R de diversas origens codificam a mesma especificidade. Ensaio rápido da análise de especificidade de genes R tem um grande potencial de aplicação em biotecnologia, pois possibilitam a descoberta de genes R que poderiam ser suprimidos por outro efector em uma inoculação envolvendo o patógeno e evitam a clonagem redundante de genes diferentes que têm a mesma especificidade gene-a-gene. Isso possibilitaria o emprego simultâneo de diferentes genes R em estratégias de emprego de genes R como, por exemplo, multilinhas, dificultando a evolução de raças de patógenos que possuam uma combinação específica de efetores capaz de suplantar a resistência, aumentando assim a durabilidade da resistência no campo (Ellis *et al.*, 2009).

3 – METODOLOGIA

3.1 - Construção e sequenciamento da biblioteca de cDNA

O mRNA de folhas de *E. grandis* naturalmente infectadas por *P. psidii* e coletadas em campo foi isolado e serviu de molde para a síntese de cDNA. Os cDNAs foram clonados no vetor pSPORT I, da Invitrogen®, e transformados em *Escherichia coli* DH5 α por eletroporação. Os clones recombinantes foram selecionados em meio LB contendo ampicilina (50 $\mu\text{g.mL}^{-1}$) e X-gal (20 $\mu\text{g. mL}^{-1}$) após incubação a 37°C por 16 h. Os clones foram então transferidos para placas de 96 cavidades contendo meio “Freeze media” e ampicilina (50 $\mu\text{g.mL}^{-1}$), incubados a 37°C por 16 h e estocados a -80°C. Essa biblioteca foi denominada EUGR-PU, sendo a sigla EUGR derivada de *Eucalyptus grandis* e PU de *Puccinia psidii*, como parte do projeto da Rede Brasileira de Pesquisa do Genoma de Eucalipto (Projeto Genolyptus).

As reações de seqüenciamento foram realizadas com o kit DYEnamic™ ET Dye Terminators (GE Healthcare, Freiburg, Germany). Foram utilizados nestas reações 100 a 150 ng de DNA plasmidial, 0,5 μM do oligonucleotídeo M13F ou M13R, 2,0 μL de DYEnamic™ ET Dye Terminator Sequencing Pré-mix em um volume final de 5 μL . As condições da PCR foram: 95°C por 10 segundos, 50°C por 5 segundos e 60°C por 2 minutos repetidas 35 vezes. Em seguida, o DNA foi precipitado adicionando-se 27,5 μL de etanol absoluto e acetato de amônio para uma concentração final de 0,75 M. Após 10 min a temperatura ambiente, as reações foram centrifugadas por 45 min a 3700 rpm. O DNA foi então lavado com 100 μL de etanol 70%, centrifugado a 3700 por 10 min e deixado secar por 10 min. Após a precipitação, o DNA foi ressuspensão em 5 μL de *Loading Buffer* (GE Healthcare). A leitura das reações foi realizada em sequenciador automático MegaBACE™ 1000 de 96 capilares (GE Healthcare) no Laboratório de Genômica, do Instituto de Biotecnologia Aplicada a Agropecuária (BIOAGRO), da Universidade Federal de Viçosa -UFV.

3.2 - Análises de bioinformática

As seqüências da biblioteca EUGR-PU foram depositadas no banco de dados do projeto Genolyptus (www.lge.ibi.unicamp.br/eucalyptus) de acesso restrito. Após a remoção de seqüências do vetor, as seqüências foram comparadas às depositadas no GenBank do National Center for Biotechnology Information (NCBI - <http://www.ncbi.nlm.nih.gov>) por meio do algoritmo BlastX. O resultado do BlastX foi utilizado para triar os ESTs quanto à origem vegetal ou fúngica. Apenas *hits* com o valor E menor que 10^{-3} foram considerados similares.

Posteriormente as sequências que não apresentaram similaridade a proteínas depositadas no GenBank foram comparadas localmente com sequências da montagem parcial (*draft*) do genoma do eucalipto com 14.317 *scaffolds* compreendendo 530.775.863 pb (EUCAGEN - <http://www.eucagen.org/>), e com sequências derivadas de outras bibliotecas geradas no projeto Genolyptus com 159.220 ESTs de diferentes bibliotecas, além de um banco de dados público: a coleção completa de ESTs do NCBI, com 51.378.116 ESTs, excluídos ESTs de humanos e ratos (*E-value* < 10^{-10}).

A qualidade das seqüências foi verificada através do pacote phred/phrap/consed (www.phrap.org), instalado no Laboratório de Genômica – BIOAGRO-UFV. As seqüências identificadas como de origem fúngica e sem similaridade a seqüências depositadas em bancos de dados foram alinhadas pelo programa CAP3 (<http://bio.ifom-firc.it/ASSEMBLY/assemble.html>) com sobreposição mínima de 15 nucleotídeos e identidade mínima de 80%.

Os contíguos e os singletons obtidos foram novamente comparados com seqüências do GenBank/NCBI através do algoritmo BlastX. Os resultados dessa análise e a presença de domínios conservados foram utilizados para inferir a função dos genes identificados. As seqüências sem similaridade foram ainda comparadas com seqüências depositadas em bancos de seqüências de fungos (<http://www.broad.mit.edu/annotation/genome/> e <http://genome.jgi-psf.org/Mellp1/Mellp1.home.html>).

A predição das seqüências abertas de leitura (ORFs) para os contíguos e singletons de suposta origem fúngica foi realizada através do programa ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) e de *scripts* Perl (Practical extraction and report language) que geram um arquivo contendo a tradução de cada transcrito

nos três possíveis quadros de leitura positivos, no formato fasta, com a linha descritora da sequência original adicionada da letra F seguida pelo número do quadro de leitura correspondente, 1, 2 ou 3: >nome_da_sequenciaF1 e um segundo *script* Perl que encontra ORFs com mais de 40aa em sequências polipeptídicas, adaptados para análise local em larga escala. O arquivo de saída dos programas contem ORFs maiores que 40 aa com um traço e um número adicionados ao nome da sequência indicando que ORF se trata dentre as possíveis para um quadro de leitura:>nome_da_sequenciaF1-1 (Anexo). As ORFs obtidas foram a seguir submetidas à predição de sequências codificadoras de sinal de exportação celular e de hélices transmembrana pelos programas SignalP, TMHMM, targetp (<http://www.cbs.dtu.dk/services/index.php>) e Phobius (<http://phobius.sbc.su.se/>). A composição de aminoácidos dos clones selecionados também foi descrita, 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 METAL DETECTOR (<http://metaldetector.dsi.unifi.it/>).

Foram selecionadas sequências da biblioteca EUGR-PU que não apresentaram identidade a sequências dos bancos de DNA utilizados, e que possuíam a predição de sequências de exportação celular com alta probabilidade para os algoritmos utilizados. Para os clones selecionados foram desenhados oligonucleotídeos iniciadores, manualmente, com a inserção de um sítio de *EcoRI* no oligonucleotídeo iniciador direto (*Forward*) imediatamente a montante (*upstream*) do códon de iniciação predito (ATG), com três dinucleotídeos GC a montante do sítio de *EcoRI*, para estabilização do produto amplificado. Para o oligonucleotídeo iniciador reverso (*reverse*) foi inserido um sítio de *NotI* com a primeira base do sítio coincidindo com a primeira base do códon de parada predito (TAA, TAG ou TGA) e três dinucleotídeos GC a jusante (*downstream*) do sítio.

3.3 - Análise da secreção das proteínas codificadas pelos genes selecionadas no sistema YST

Os clones de cDNA que codificam as proteínas pressupostamente secretadas tiveram a ORF predita amplificada e o produto da amplificação submetido à digestão enzimática pelas endonucleases *EcoRI* e *NotI*, e posteriormente ligados ao vetor

pYST0-1 (**Figura 3**), previamente digerido com *EcoRI* e *NotI*. O produto da reação de ligação foi transformado por choque térmico em *Escherichia coli* EC100.

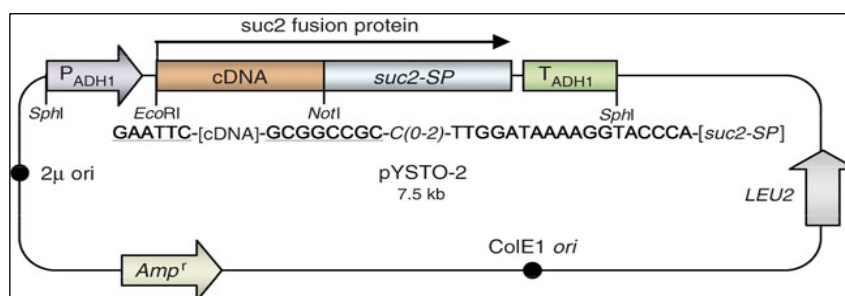


Figura 3 – Representação esquemática do vetor pYSTO-1 utilizado para a clonagem das ORFs selecionadas, amplificadas a partir de cDNA, para a comprovação da secreção pelo sistema YST (Lee. *et al.* 2006).

Os transformantes foram selecionados em meio LB contendo ampicilina (100µg/mL), sendo os plasmídeos purificados utilizando-se o kit Plasmid midi kit, de acordo com as instruções do fabricante (Qiagen).

A estirpe YIL162w (MAT α , *suc2*, *his3* Δ , *leu2* Δ , *lys* Δ , *ura3* Δ) de levedura (*Saccharomyces cerevisiae*) foi utilizada para expressão da proteína recombinante e confirmação da secreção. A transformação foi realizada de acordo com o protocolo descrito por Gietz & Woods (2002): para cada 50 µL de células foram adicionados os reagentes na seguinte ordem: 240 µL de PEG 50 % (polietilenoglicol), 36 µL de AcLi 1M, 50 µL de espermatozoides de salmão (2 µg/µL), 34 µL de água deionizada estéril contendo 1 µg de DNA a ser transformado. A solução foi incubada por 30 min a 30° C e em seguida incubada por 20 min a 42° C. A seguir, a suspensão foi centrifugada a 8000 g por 15 segundos. O precipitado foi lavado com 1 mL de água deionizada estéril e ressuspensão em 200 µL de água deionizada estéril. Para cada placa de Petri de 50 mL contendo o meio seletivo, foram plaqueados 200 µL de suspensão celular. As placas contendo meio YPS sólido (1 % extrato de levedura, 2 % peptona, 2 % sacarose, 15,0 g/L de ágar) foram mantidas por sete dias em a 30°C.

Após 6 a 10 dias de incubação foram selecionadas colônias no meio YPS. Para confirmação da identidade dos clones selecionados foi realizada a extração dos plasmídeos e amplificação por PCR da região fusionada ao gene da invertase. As colônias selecionadas na etapa anterior foram repicadas em meio YPS e incubadas a

30⁰ C por dois dias. Aproximadamente 50 µL de células foram raspadas com espátula estéril e ressuspensas em 1 mL de água deionizada estéril, em micro tubos de 1,5 mL. As células foram centrifugadas a 5.000 g, por 30 seg. A água deionizada foi descartada e o precipitado foi ressuspendido em 200 µL de solução de lise (SDS 1%, EDTA 50mM, Tris-HCl 50mM, pH 8,0). Em seguida foram adicionados 200 µL de fenol:clorofórmio:álcool isoamílico (25:24:1) e 200 µL de esferas de vidro. O micro tubo foi agitado vigorosamente por 2 min e em seguida deixado no gelo por mais 2 min. A suspensão foi centrifugada a 13.200 g por 30 min e o sobrenadante transferido cuidadosamente para outro micro tubo evitando misturas com a interface. Foi adicionado 3 vezes o volume pipetado de etanol 100 % gelado, misturado e centrifugado por mais 30 min a 13.200g. O precipitado foi lavado em etanol 70 % e ressuspendido em 30 µL de água deionizada estéril.

Os clones apresentando crescimento em meio YPS e que tiveram a ORF amplificada foram considerados positivos.

3.4 – Amplificação das ORFs a partir do DNA genômico de *E. grandis* e *P. psidii*

As extrações de DNA genômico de eucalipto foram realizadas conforme protocolo descrito em Grattapaglia & Sederoff (1994a).

O isolado monopustular de *P. psidii* UFV-2, gentilmente cedido pelo prof. Acelino Couto Alfenas, foi empregado para a extração de DNA a partir de urediniósporos. O isolado foi mantido em mudas de jameiro (*Syngisium jambos* (L.) Alston), mediante re-inoculações periódicas. Aproximadamente um grama de esporos foi macerado em almofariz com nitrogênio líquido. Após a maceração, o macerado foi suspenso em 5 mL do tampão de extração (Sorbitol 0,35 M, Tris 0,1 M, EDTA 5 mM, pH 7,5) ao qual foi adicionado bisulfito de sódio até atingir a concentração de 20 mM imediatamente antes do uso. A suspensão foi homogenizada suavemente, sem *vortex*, e foram adicionados 5 mL de tampão de lise nucléica (Tris 200 mM, EDTA 50 mM, NaCl 2 mM e CTAB 2%) e 2 mL de sarcosil 5%. A suspensão foi incubada em banho-maria a 65°C por 1 h, sendo suavemente invertida de 5 em 5 min. Após a incubação foram adicionados 15 mL de clorofórmio:álcool isoamílico (24:1) e a suspensão foi invertida suavemente 30 a 40 vezes até atingir uma emulsão e logo após centrifugada por 15 min a 6.500 g. O sobrenadante foi transferido para um tubo novo e misturado ao mesmo volume de isopropanol gelado.

O DNA foi precipitado por centrifugação a 11.000 g por 20 min, lavado duas vezes com etanol 70% e ressuspensão em 600 µL de água deionizada estéril contendo RNase (100 µg . mL⁻¹).

O conjunto de oligonucleotídeos para o sistema *gateway* (descrito abaixo) foi utilizado para testar a amplificação a partir do DNA genômico de *E. grandis* e *P. psidii*. As amplificações foram realizadas em condições de baixa estringência, com TM de 50° C, e 40 ciclos de anelamento, utilizando 20 ng de DNA por reação para *P. psidii* e 50 ng para *E. grandis*.

3.5 - Expressão transiente dos genes que codificam as proteínas comprovadamente secretadas

Para o ensaio de expressão transiente foi utilizada a agroinfiltração. Foram utilizadas folhas expandidas de plantas de tabaco com três semanas de idade (*Nicotiana tabacum* ‘Havana 425’ e *Nicotiana benthamiana* ‘selvagem’) e folhas tenras e expandidas dos clones eucalipto (*E. grandis*) G7D (resistente a ferrugem) e 1028 (sucetível a ferrugem). O cassete de expressão foi construído com os genes selecionados, como descrito acima, amplificados com oligonucleotídeos iniciadores F, desenhados com adição do sítio CACC, para clonagem pelo sistema *Gateway*® (Invitrogen), à montante do códon ATG e com o oligonucleotídeo iniciador R terminando com o respectivo códon de parada. Os amplicons obtidos a partir do cDNA dos clones selecionados foram clonados no vetor pENTR™/D-TOPO® segundo instruções do fabricante. As ORFs foram então clonadas no vetor de destino pK7FWG2 (Karimi *et al.*, 2002), segundo instruções do *kit Gateway*® LR Clonase™ II Enzyme Mix (Invitrogen).

A estirpe pRKYRI de *Agrobacterium tumefaciens* foi transformada por choque térmico com as construções citadas acima, para as ORFs de clones que tiveram sua secreção comprovada pelo sistema YST. O vetor pK7FWG2 vazio também foi utilizado para transformação, servindo como controle negativo para as análises de agroinfiltração. As agrobactérias contendo o cassete de expressão foram cultivadas em meio YEP (Bactotripton 1 %, NaCl 0,5% e extrato de levedura 0,5 %), contendo rifampicina 50 µg.mL⁻¹ e estreptomicina 50 µg.mL⁻¹, sendo as mesmas mantidas sob agitação (250 rpm) por 12 h a 28°C. Após o crescimento do pré-inóculo

2 mL foram utilizados para inocular 15 mL de meio YEP, incubado nas mesmas condições por 1 h e 30 min. A seguir, as agrobactérias foram coletadas por centrifugação a 11.000 g por 5 minutos e o sobrenadante descartado. As agrobactérias foram ressuspensas em 5 ml de tampão de infiltração (MES 50 mM, Na₃PO₄ 2 mM, glucose 0.5 % (w/v) e acetoseringona 200 µM) e ajustadas para uma absorbância a 600 nm entre 0,2 e 0,4 (Zottini *et al.*, 2008). A solução foi injetada com seringa de plástico de 1 ml na face abaxial das folhas, em um ponto, entre as nervuras secundárias. As plantas foram mantidas em casa de vegetação, e avaliadas diariamente para monitorar o desenvolvimento de necrose na área infiltrada.

4 - RESULTADOS

Foram analisadas 9864 ESTs da biblioteca EUGR-PU do projeto Genolyptus que apresentaram qualidade phred > 20. Cerca de 6737 (68,3%) dos ESTs apresentaram similaridade a seqüências de genes de plantas depositadas no GenBank/NCBI, 198 (2,0%) a genes de fungos e 2613 (28,9%) não apresentavam similaridade a genes previamente caracterizados. Os 115 contíguos de ESTs similares a genes de fungos foram previamente analisados por Marques (2007): 60,5% deles apresentaram similaridade a proteínas hipotéticas e 39,5% a proteínas de função conhecida, tais como: metabolismo de carboidrato (malato desidrogenase do ciclo de Krebs, por exemplo), de proteínas (serina endopeptidase), de lipídeos, de aminoácidos (glutamina sintetase), de DNA e de RNA. Foram identificadas também proteínas ribossomais, proteínas com domínios de interação proteína-proteína e proteínas envolvidas em transporte. A maior parte dos contigs codifica proteínas envolvidas no metabolismo primário do fungo, no entanto, também foram identificadas proteínas similares às proteínas RTP1 e HESP-329 sem função predita que são expressas no haustório (Kemen *et al.*, 2005; Catanzariti *et al.*, 2006).

Este trabalho foi focado na análise de seqüências de *P. psidii* sem similaridade com seqüências depositadas em bancos de dados, tendo em vista que a maioria das proteínas efetoras caracterizadas até o momento são proteínas novas, pequenas (< 400 aminoácidos) e com peptídeo sinal de secreção. Como as seqüências de ESTs sem similaridade com seqüências depositadas em bancos de dados poderiam incluir seqüências de genes de eucalipto ainda não caracterizados, efetuou-se a seleção de ESTs de origem fúngica e função desconhecida por meio da filtragem por similaridade (BlastN) com seqüências do rascunho inicial (*draft*) do genoma de *E. grandis*, do banco de dados de ESTs do projeto Genolyptus, além da coleção completa de ESTs do NCBI (**Figura 4**). A partir de 2613 transcritos que não codificavam proteínas com similaridade (BlastX) significativa a proteínas do banco não redundante do GenBank/NCBI, identificaram-se 1239 transcritos que também não apresentavam identidade significativa (e-value < 10⁻¹⁰) a seqüências do banco de ESTs/NCBI, 1228 não apresentavam identidade significativa a seqüências do *draft* do genoma do eucalipto e 1109 não apresentavam identidade significativa a seqüências do banco de ESTs do genolyptus (**Figura 4**).

Foram utilizados 1675 transcritos, denominados de *No hits* não redundantes, para a predição de ORFs nos três quadros de leitura positivos, uma vez que a clonagem foi direcional e apenas a extremidade 5' dos transcritos foi sequenciada.

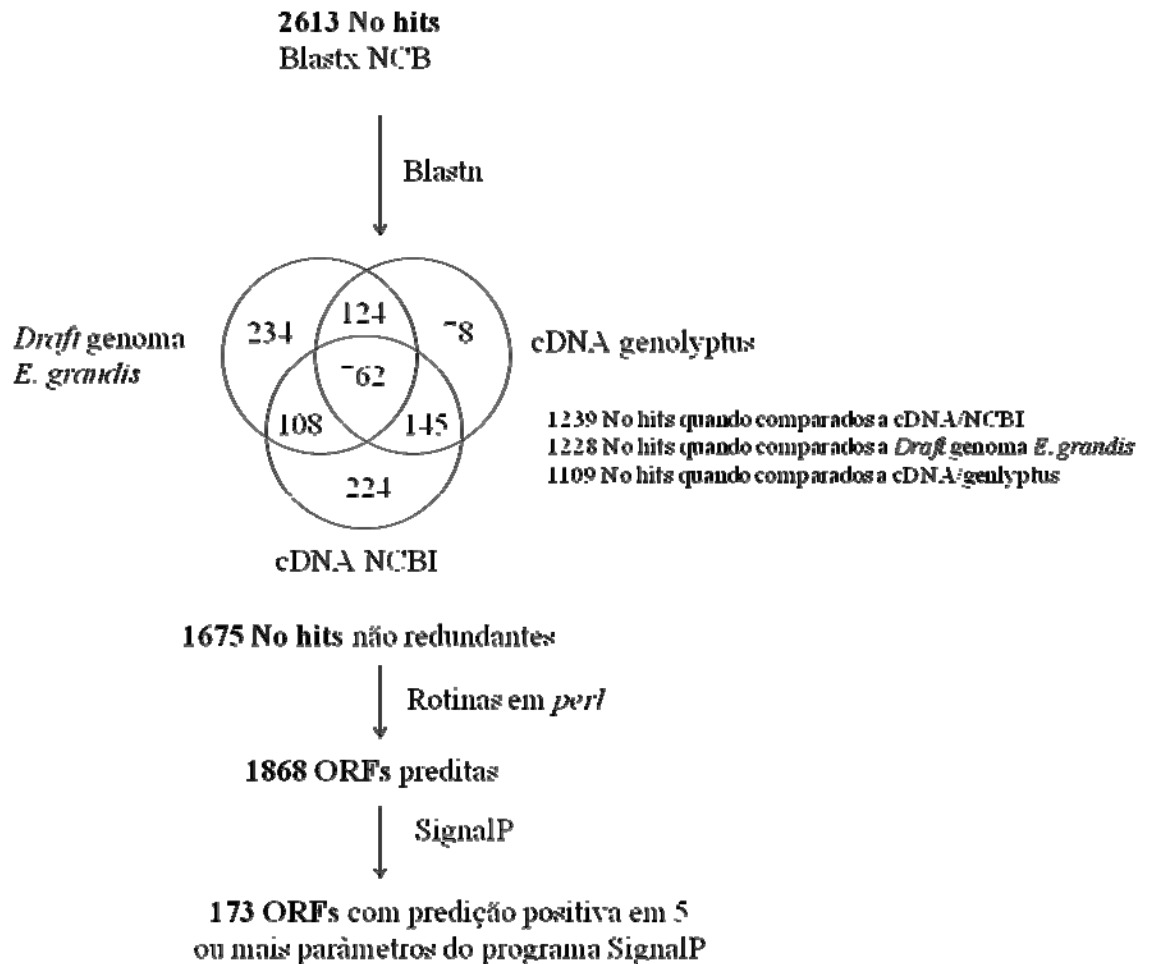


Figura 4 - Esquema da estratégia utilizada para descoberta de candidatos a efetores do fungo *Puccinia psidii* a partir de uma biblioteca de ESTs de folhas de *Eucalyptus grandis* infectadas.

A tradução das seqüências de nucleotídeos realizada com um *script* Perl (**Anexo**) nos três quadros positivos de leitura resultou na identificação de 1868 ORFs com mais de 40 aa. A análise das ORFs deduzidas por meio do algoritmo SignalP de predição de sinal de secreção resultou na identificação de 173 ORFs com predição positiva (Y) em cinco ou mais parâmetros do programa (**Anexo**). As seqüências positivas foram reanalisadas quanto à predição de sinal de exportação e hélices transmembrana utilizando outros algoritmos visando uma seleção mais estrigente

dos clones a serem completamente caracterizados (**Anexo**). Desta forma, foram selecionados clones que apresentaram valor-D ≥ 0.65 , Sprob ≥ 0.8 , com predição de secreção (S) e *score* (RC) 1 ou 2 para o programa TargetP e que não tivessem predição de hélice transmembrana confirmada pelos algoritmos TMHMM e Phobius (**Anexo**). Esses clones foram completamente sequenciados e as sequências foram montadas e analisadas com o pacote phred/phrap/consed para confirmação da predição, sendo selecionados 31 clones que apresentaram sequências com valor phred ≥ 20 , possuíam a ORF completa e com predição de seqüência codificadora de peptídeo sinal confirmada (**Tabela 1**). As ORFs codificadas por estes clones quando comparadas entre si por meio do programa BlastP, revelaram similaridade variando desde pequenos motivos compartilhados a extensa identidade (**Tabela 2**). Estas ORFs também foram comparados localmente a bancos de dados de proteínas deduzidas a partir do genoma de *Puccinia graminis tritici* constituído de 20.566 sequências, disponível em (http://www.broadinstitute.org/annotation/genome/puccinia_graminis/download/?sp=EATranscriptsGtf&sp=SPG_tritici_V2&sp=S.zip) e do genoma de *Melampsora laricis* com 87.121 sequências, disponível em (http://genome.jgi-psf.org/Mellp1/download/Mlaricis_populina.allModels.aa.fasta.gz). Quatro ORFs codificam proteínas com similaridade a proteínas preditas do banco de *P. graminis* e de *M. laricis*, quatro, exclusivamente, com proteínas preditas de *P. graminis* e uma somente com uma proteína predita de *M. laricis* (**Tabela 1**). As proteínas similares não tem função predita nos genomas de *P. graminis* e *M. laricis*, e apenas as proteínas preditas PGTG_10501, PGTG_16623 e PGTG_02680 (similares às ORFs preditas para os clones 06, 20 e 52, respectivamente) de *P. graminis* e jgi|Mellp1|95372 (similar ao clone 06) de *M. Laricis* possuem seqüência codificadora validada por transcritos, sendo as seqüências codificadoras das demais somente identificadas *in silico*. A ORF codificada pelo clone 06 foi a que apresentou maior similaridade com proteínas preditas, 66 % com a proteína PGTG_10501 e 73 % com a proteína jgi|Mellp1|95372 (**Figura 5**).

Clone	Nº AA	PS ¹	SignalP-NN	SignalP-HMM	targetp v1.1		Tmhmm	Phobius		blastp/ <i>P. Graminis</i>		blastp/ <i>M. Laricis</i>	
			Valor-D ²	Sprob ³	Loc ⁴	RC ⁵		TM ⁷	SP ⁸	Descrição da sequência	E-value	Descrição da sequência	E-value
02Pp_sp	145	22	0.878	1	S	1	PredHel=1 ⁶	0	Y		No hit		No hit
05Pp_sp	99	25	0.912	1	S	1	PredHel=0	0	Y		No hit		No hit
06Pp_sp	245	26	0.878	0.989	S	1	PredHel=1	0	Y	PGTG_10501 hypothetical protein (translation) (351 aa)	2e-39	jgi Mellp1 95372 EuGene.00070277 (438 aa)	3e-44
07Pp_sp	128	22	0.886	0.999	S	1	PredHel=0	0	Y		No hit		No hit
09Pp_sp	162	18	0.904	0.999	S	1	PredHel=0	0	Y	PGTG_00845 predicted protein (translation) (262 aa)	5e-09	jgi Mellp1 103832 fgenes2_pg.7_#_140 (336 aa)	3e-06
13Pp_sp	116	19	0.897	0.999	S	1	PredHel=0	0	Y		No hit		No hit
15Pp_sp	99	24	0.901	0.994	S	1	PredHel=1	0	Y		No hit		No hit
19Pp_sp	168	19	0.911	0.999	S	1	PredHel=0	0	Y		No hit		No hit
20Pp_sp	146	21	0.64	0.782	S	2	PredHel=0	0	Y	PGTG_16623 hypothetical protein (translation) (168 aa)	5e-08		No hit
22Pp_sp	159	26	0.653	0.967	S	2	PredHel=0	0	Y	PGTG_06297 predicted protein (translation) (184 aa)	2e-05		No hit
24Pp_sp	204	19	0.896	1	S	1	PredHel=0	0	Y		No hit		No hit
25Pp_sp	138	22	0.879	0.988	S	1	PredHel=0	0	Y		No hit		No hit
27Pp_sp	170	22	0.879	0.996	S	1	PredHel=0	0	Y		No hit		No hit
28Pp_sp	164	19	0.659	0.986	S	1	PredHel=0	0	Y		No hit		No hit
32Pp_sp	102	31	0.76	0.884	S	2	PredHel=0	0	Y		No hit		No hit
41Pp_sp	83	19	0.828	1	S	1	PredHel=2	0	Y		No hit		No hit
42Pp_sp	180	25	0.681	0.998	S	1	PredHel=0	0	Y		No hit		No hit
43Pp_sp	168	19	0.921	0.999	S	1	PredHel=0	0	Y		No hit		No hit
45Pp_sp	183	19	0.648	0.997	S	1	PredHel=0	0	Y	PGTG_06052 predicted protein (translation) (239 aa)	2e-15	jgi Mellp1 109863 fgenes2_pg.32_#_105 (195 aa)	1e-06
46Pp_sp	126	22	0.886	0.999	S	1	PredHel=0	0	Y		No hit		No hit
47Pp_sp	168	19	0.93	0.999	S	1	PredHel=0	0	Y		No hit		No hit
49Pp_sp	130	20	0.875	0.999	S	1	PredHel=0	0	Y		No hit	jgi Mellp1 109863 fgenes2_pg.39_#_96 (166 aa)	6e-08
50Pp_sp	183	19	0.682	0.998	S	1	PredHel=0	0	Y	PGTG_06052 predicted protein (translation) (239 aa)	5e-15	jgi Mellp1 108883 fgenes2_pg.32_#_105 (195 aa)	1e-06
51Pp_sp	180	19	0.715	0.991	S	1	PredHel=0	0	Y		No hit		No hit
52Pp_sp	139	21	0.892	0.999	S	2	PredHel=0	0	Y	PGTG_02680 hypothetical protein (translation) (134 aa)	5e-15		No hit
53Pp_sp	120	27	0.876	0.999	S	1	PredHel=1	0	Y	PGTG_01914 predicted protein (translation) (121 aa)	1e-11		No hit
56Pp_sp	104	32	0.695	0.955	S	1	PredHel=1	0	Y		No hit		No hit

57Pp_sp	148	24	0.706	0.989	S	1	PredHel=1	0	Y	No hit	No hit
58Pp_sp	80	22	0.74	0.967	S	1	PredHel=0	1	Y	No hit	No hit
59Pp_sp	136	22	0.851	1	S	1	PredHel=1	0	Y	No hit	No hit
60Pp_sp	131	22	0.914	0.999	S	1	PredHel=1	0	Y	No hit	No hit

Tabela 1 – Predição de peptídeo sinal de secreção e comparação com as proteínas deduzidas a partir dos genomas de *Puccinia graminis* e *Melampsora laricisa* para os clones selecionados por apresentarem cinco ou mais parâmetros positivos pelo programa SignalP.

1. Número de aminoácidos do peptídeo sinal predito.

2. 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;

3. Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q);

4. Loc – Predição da localização celular, (M) mitocondria, (S) peptídeo sinal e – outro;

5. 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;

6. PredHel – Número de α -hélices transmembrana preditas;

7. TM - Número de α -hélices transmembrana preditas;

8. SP – Predição de peptídeo sinal;

Tabela 2 – Similaridade das seqüências de aminoácidos das ORFs preditas evidenciadas pelo programa BlastP.

Clone	02Pp_sp	07Pp_sp	13Pp_sp	19Pp_sp	20Pp_sp	22Pp_sp	25Pp_sp	27Pp_sp	42Pp_sp	43Pp_sp	45Pp_sp	46Pp_sp	49Pp_sp	50Pp_sp	59Pp_sp	
02Pp_sp	100%															
07Pp_sp	82/141 ¹ (58%)	100%														
13Pp_sp	No hit	No hit	100%													
19Pp_sp	No hit	No hit	99/103 (96%)	100%												
20Pp_sp	No hit	No hit	No hit	No hit	100%											
22Pp_sp	No hit	No hit	No hit	No hit	58/131 (44%)	100%										
24Pp_sp	45/112 (40%)	No hit	No hit	No hit	No hit	No hit										
25Pp_sp	No hit	No hit	No hit	No hit	45/105 (42%)	61/155 (39%)	100%									
27Pp_sp	49/96 (51%)	38/79 (48%)	No hit	No hit	No hit	No hit	No hit	100%								
42Pp_sp	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	100%							
43Pp_sp	No hit	No hit	96/103 (93%)	157/169 (92%)	No hit	No hit	No hit	No hit	No hit	100%						
45Pp_sp	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	80/180 (44%)	No hit	100%					
46Pp_sp	84/143 (58%)	124/129 (96%)	No hit	No hit	No hit	No hit	No hit	62/155 (40%)	No hit	No hit	No hit	100%				
47Pp_sp	No hit	No hit	94/103 (91%)	154/169 (91%)	No hit	No hit	No hit	No hit	No hit	162/169 (95%)	No hit	No hit				
49Pp_sp	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	100%			
50Pp_sp	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	79/180 (43%)	No hit	181/184 (98%)	No hit	No hit	100%		
51Pp_sp	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	92/181 (50%)	No hit	91/187 (48%)	No hit	No hit	No hit	89/187 (47%)	
52Pp_sp	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	38/80 (47%)	No hit	
57Pp_sp	No hit	No hit	No hit	No hit	66/149 (44%)	65/162 (40%)	57/147 (38%)	No hit	No hit	No hit	No hit	No hit	No hit	No hit	No hit	
59Pp_sp	126/146 (86%)	86/132 (65%)	No hit	No hit	No hit	No hit	No hit	No hit	54/126 (42%)	No hit	No hit	No hit	83/135 (61%)	No hit	No hit	100%
60Pp_sp	74/122 (60%)	111/132 (84%)	No hit	No hit	No hit	No hit	No hit	No hit	63/156 (40%)	No hit	No hit	No hit	109/132 (82%)	No hit	No hit	77/120 (64%)

1 – Numero de aminoácidos similares/comprimento do alinhamento local.

```

06          -----MSSHFPFIIILLISLLSTTHLVIAG-RCCPGDSDICSGCWMCPGSDLRNCWDNA 53
Mellp      -----MPSLVRITAIIVSVV--LSLAHSISAGGRCCPGDNICSGCWMCKGSDLRDCWDQA 52
PGTG       MAYLFTRLTLSIAFLLAQLSVLSMVAPESPAQTCCPGDLTCSGCWMCPESDLGKCWATA 60
           : . : : : : . . ***** ***** * * * * *

06          IKDVDQCASASNKQMAQCYCLNYAKVAKCWSVGTCTEYAPSMAAATNLCNFANWPTEK 113
Mellp      TPMPKCEAISDLQEKAAACYCKGYAAIAKCSWRGTCTEYAPTAVATDLCNLAWSPSDQ 112
PGTG       VKEIDECVITDDLPKSADTFRAFAKLAKCWSVGTCCSEYAPTAAATNLCNLTWSSEL 120
           : * . : * : * : * : * : * * * * * * * * * * * * * * * * * * * * * *

06          INNQTLQNLVKEARNVKEAIWHNDSGYDVPKLSIHVDPLGS-----SNKKNP--SND 164
Mellp      ITSARLDHCVKEAARVADAIWHNDSGVDPKQSSIEPYAGGPVATGVAHVKKPKP--VHS 170
PGTG       ITREELEKYIEIANATRDAIWNNDSGIDVPKLTVSVEANTSNSSTESSPSNSTESSPSIST 180
           * . : * : * . : * * * * * * * * * * * * * * * * * * * * * *

06          ENLDNNYQD-----KNKKNFLPKKSYHHHSHNHQ----- 196
Mellp      ESYVGSPPADPPPTSSQNRKQKSSASHHPSHSHLAE-----KERKKK 216
PGTG       ESTQANSTEVKTNPESTNSTQHTYEKKSVVNPSQADDDQQTNPKEKIDRRLLSSDREGK 240
           * . : : * : * : * : * : * : * : * : * : * : * : * : * : * : *

06          -----KRNPFIHPKLS-----ITHS-----IQKRSSSQ 220
Mellp      AAMEKKKQHYEEKKKKEWEAKQRAKKLADKKWKEHHHQDKGAKHSEREGGCKEKRNHTK 276
PGTG       KDSEKANHRQSNEKANETSKNQNCVAVKANKHRTKVADQHHKSRSSKTKKSLSKRVENP 300
           * : : : : : : * * * * * * * * * * * * * * * * * * * * * *

06          -----CGIDFQGVYPKSVPTNLLKAYSQI 245
Mellp      RASYTTLNVSPAEMNDRMGKRNVLGIDFQGVKHPSTVPELVQKAYDEI 326
PGTG       FHKRMYHHQHHPDLFKRTAPGEGGCGLDQGVAYPTEVPQTVAQIFSEL 350
           * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

Figura 5 - Alinhamento múltiplo da sequência protéica deduzida a partir da ORF predita para o clone 06 de *Puccinia psidii* com as proteínas hipotéticas PGTG_10501 de *P. graminis* e jgi|Mellp1|95372 de *M. laricis*, utilizando o programa CLUSTAL 2.0.12. 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.

O primeiro grupo com extensa similaridade é composto pelos clones selecionados 02, 07, 46, 59 e 60. Analisando-se todo o alinhamento verifica-se que 02 e 59 são diferentes de 07, 46 e 60 (Figura 6).

```

07Pp_sp    MKAHFVALVQILCVFVGSAVSAPFSGDNSIS-GGFSSIPFGSGSG--FGNDDDDDEEL 57
46Pp_sp    MKAHFVALVQILCVFVGSAVSAPFSGDNSIS-GGFSSFPFGSGSG---FGNDDNDEEL 55
60Pp_sp    MKAHFVALVQILCIFVGSAVSAPFPDINSILPGLGSSIPFGSFAFGHHYGHNDDEEL 60
           ***** :***** .***** * : * * * * * * * * * * * * * * * *

07Pp_sp    SKKLGFSFDQGRQAGFLQAGSAAAGFD SNGNPQAGRAAGVKAAYGQERNKGYSEHR 117
46Pp_sp    SKKLGFSFDQGRQAGYLQAGSAAAGFD SNGNPQAGRAAGVNAAYGQERNKGYSEHR 115
60Pp_sp    HKKLAFAYDQGRQAGYQAVSGATAGFD SHGNPQATRADGVQAAAYGQQRNKGKGYSEPR 120
           * * * * * : * * * * * * * * * * * * * * * * * * * * * *

```

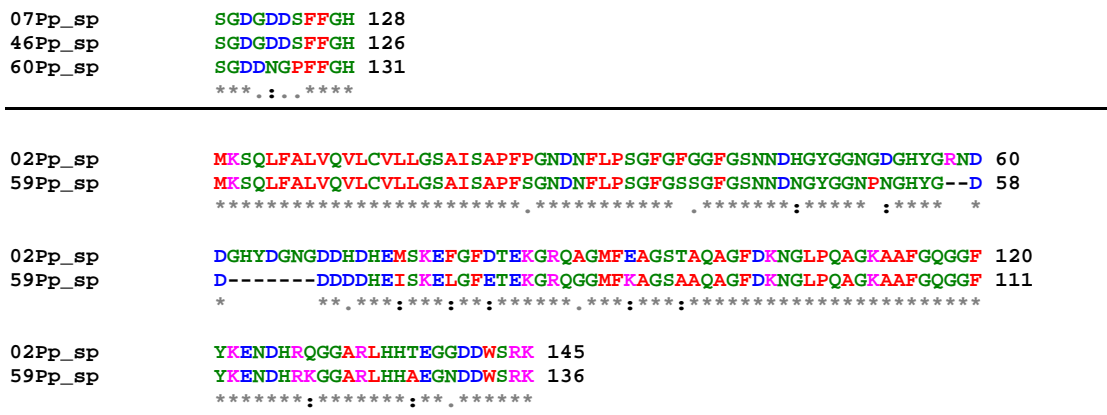


Figura 6 – Alinhamento múltiplo de seqüências protéicas, utilizando o programa CLUSTAL 2.0.12 para o sub-grupo de proteínas codificadas pela ORF dos clones 02, 07, 46, 59 e 60, evidenciando as maiores semelhanças entre os clones 07, 46 e 60, e entre 02 e 59. Cores: Vermelho – resíduo pequeno (pequeno positivo, hidrofóbico); Azul – ácido; Magenta – básico; verde – hidroxil, amina, básico; Cinza – Outros. Pontos: “*” - identidade; “:” - substituição conservada, de acordo com as cores; “.” – substituições semi-conservadas.

Outro sub-grupo com extensa similaridade observado foi o composto pelos clones 13, 19, 43 e 47, no qual, pode se observar que o clone 13 possui uma pequena região deletada na extremidade Carboxi-terminal (**Figura 7**).

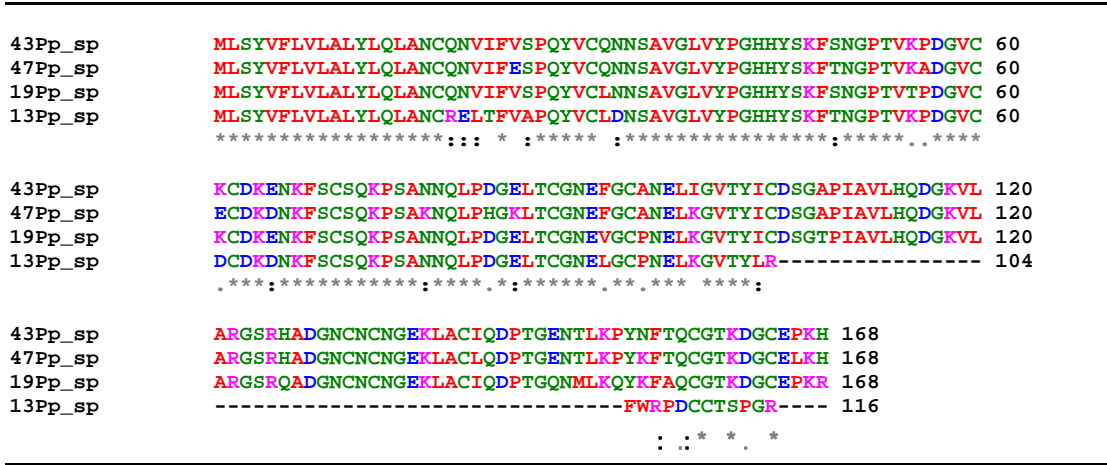


Figura 7 – Alinhamento múltiplo de seqüências protéicas, utilizando o programa CLUSTAL 2.0.12 para o sub-grupo de clones 13, 19, 43 e 47. Cores:

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.

Os clones 45Pp e 50Pp codificam proteínas que diferem em apenas em 5 aminoácidos. Os nucleotídeos correspondentes em ambas as sequências foram validadas por pelo menos duas leituras com valor phred ≥ 40 (Figura 8).

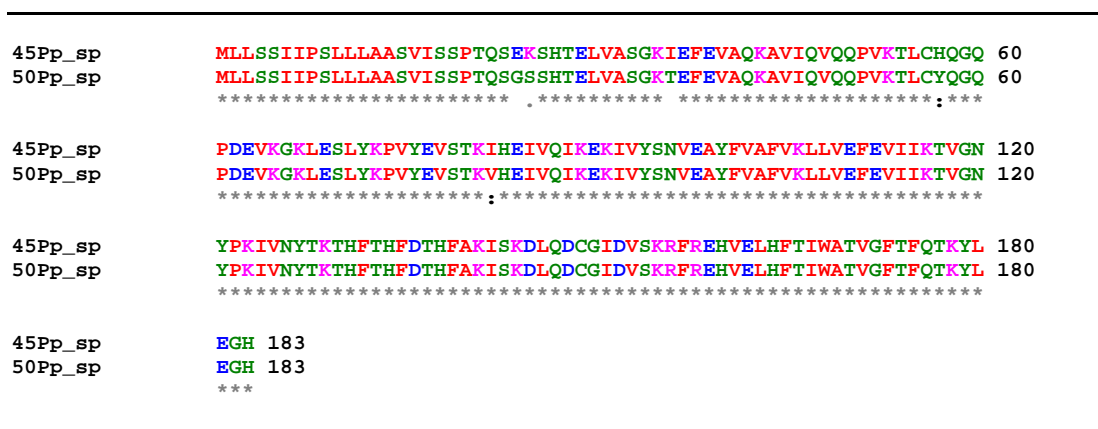


Figura 8 – Alinhamento múltiplo de sequências protéicas codificadas pelas ORFs dos clones 45 e 50, utilizando o programa CLUSTAL 2.0.12. Cores: Vermelho – resíduo pequeno (pequeno positivo, hidrofóbico); Azul – ácido; Magenta – básico; verde – hidroxil, amina, básico; Cinza – Outros. Pontos: “*” - identidade; “:” - substituição conservada, de acordo com as cores; “.” – substituições semi-conservadas.

A presença de cisteína foi verificada nos clones analisados, e foi possível observar que os clones 06 com 14 (5,7%), 13 com 9 (7,8%), 15 com 7 (7,1%), 19 com 13 (7,7%), 20 com 9 (6,2%), 22 com 9 (5,7%), 25 com 8 (5,8%), 28 com 10 (6,1%), 43 com 13 (7,7%), 47 com 13 (7,7%), 52 com 7 (5,0%) e 57 com 8 (5,4%), apresentam 5% ou mais de cisteínas na sua composição (Anexo). Sítios passíveis de formação de pontes dissulfeto foram observados para os clones 06, 13, 15, 19, 20, 22, 24, 25, 43, 45, 47, 49, 50, 52, 56 e 57, por meio da análise com o programa METAL DETECTOR (Figura 9).

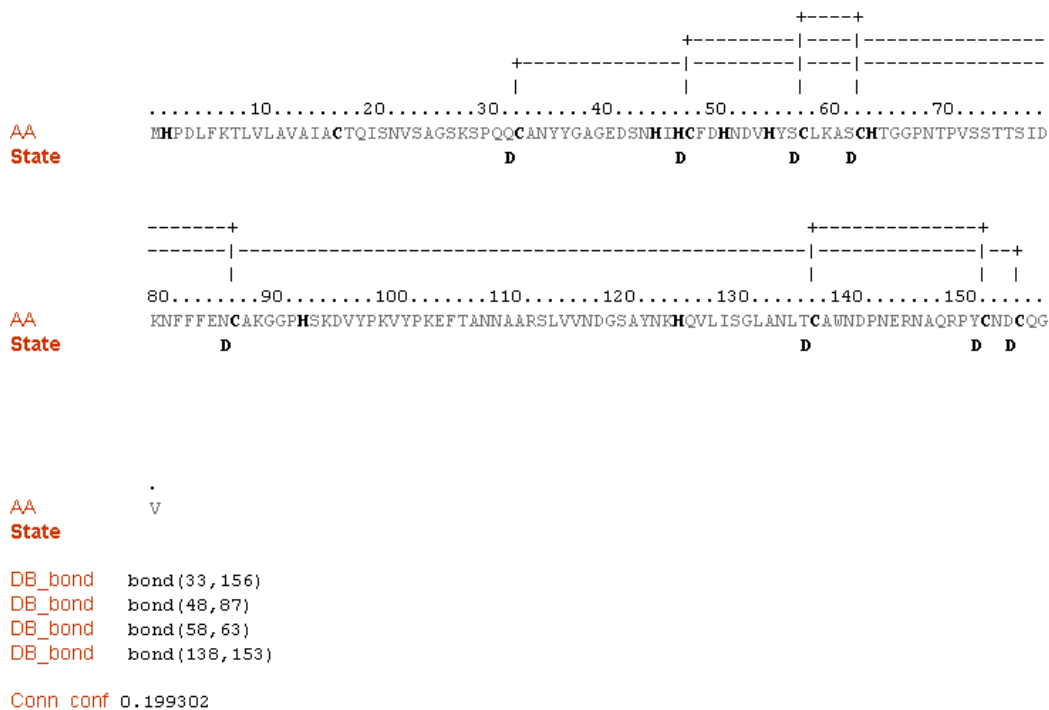


Figura 9 – Análise ilustrativa dos possíveis pontos de formação de pontes dissulfeto obtidas pelo programa METAL DETECTOR para a sequência protéica correspondente à ORF do clone 22.

Os clones selecionados tiveram suas ORFs amplificadas utilizando os oligonucleotídeos iniciadores desenhados para clonagem nos sistemas YST e no vetor de entrada pENTR™/D-TOPO®. Todos os oligonucleotídeos desenhados amplificaram fragmentos do tamanho médio predito. Os fragmentos amplificados pelos oligonucleotídeos desenhados para o sistema YST foram clonados no vetor pYST-1. A secreção foi confirmada *in vitro* para 24 clones (**Tabela 6**).

O conjunto de oligonucleotídeos para o sistema *gateway* também foi utilizado para testar a amplificação a partir do DNA genômico de *E. grandis* e *P. psidii*. As amplificações foram realizadas em condições de baixa estringência, com TM de 50° C, e 40 ciclos de anelamento, para minimizar as chances de falsos negativos. Em tais condições foi possível observar a amplificação de um fragmento único do mesmo

tamanho observado para o cDNA, ou ligeiramente maior, para 29 dos 31 clones, apenas a partir do DNA de *P. psidii* (**Figura 10**). O clone 58 não apresentou amplificação nas condições testadas, e o clone 56 apresentou amplificação apenas para o genoma do eucalipto.

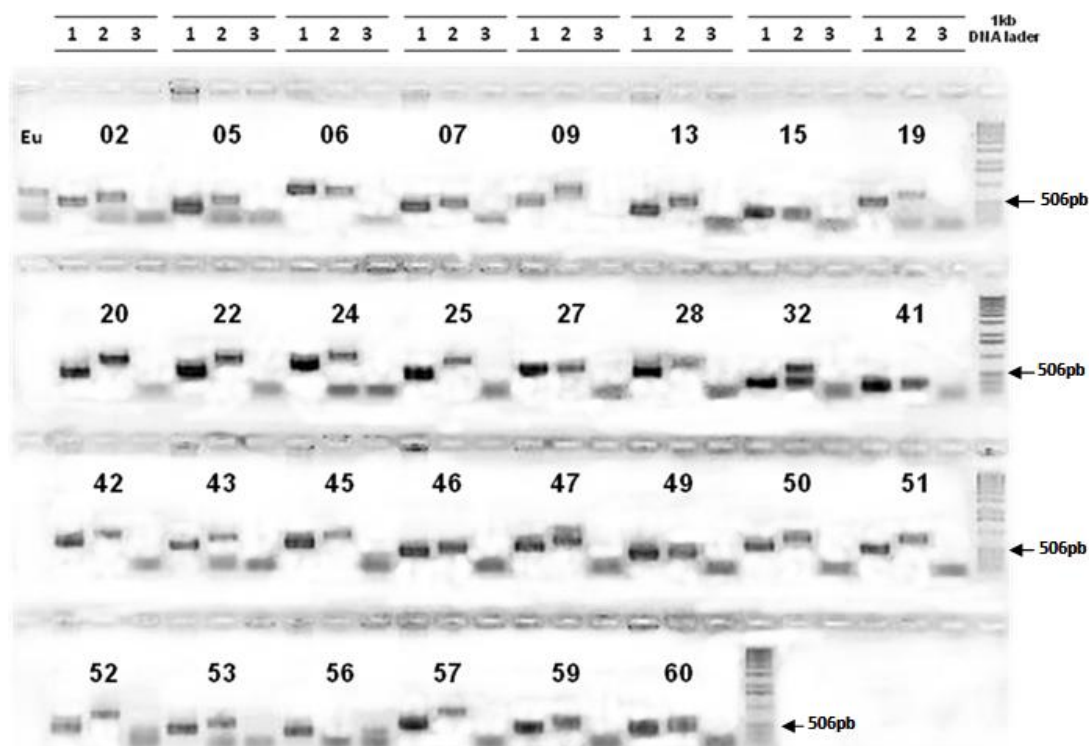


Figura 10 – Amplificação das ORF deduzidas a partir do DNA genômico de *E. grandis* e *P. psidii* utilizando o conjunto de oligonucleotídeos para o sistema *gateway*. Os números no topo da figura representam: 1- Clone de cDNA original selecionado; 2 – DNA genômico de *P. psidii*; 3 – DNA genômico de *E. grandis*. Os números no centro da figura representam a codificação dos clones de acordo com o descrito na Tabela 2. A sigla Eu representa o amplicon do gene putativo da enzima Lacase de *E. grandis*, amplificado como testemunha da qualidade do DNA.

Para testar a atividade efetora dos clones, vinte clones que tiveram sua secreção comprovada em levedura foram clonadas no vetor pK7FWG2 e transformados em *Agrobacterium* para os ensaios de expressão transiente em folhas de tabaco e eucalipto. Para eucalipto não foi possível infiltrar a suspensão de bactérias no

mesofilo foliar, impedindo a análise de expressão transiente por agroinfiltração. As duas espécies de tabaco testadas possibilitaram a infiltração do mesofilo foliar, no entanto, sete dias após a inoculação, foi observada apenas uma reação clorótica para todas as construções analisadas, o que não diferiu da reação observadas nas plantas infiltradas com a suspensão de *Agrobacterium* contendo o vetor pK7FWG2 vazio, nas condições de inoculação utilizadas.

Tabela 6 – Resultados das análises de secreção em levedura e expressão transiente em *Nicotiana tabacum*, *N. benthamiana* e *E. grandis* dos clones selecionados.

Clone	Secreção em Levedura	Expressão Transiente		
		<i>N. tabacum</i>	<i>N. benthamiana</i>	<i>E. grandis</i>
02Pp_sp	+ ¹	- ²	-	ND ³
05Pp_sp	+	-	-	ND
06Pp_sp	+	ND	ND	ND
07Pp_sp	+	-	-	ND
09Pp_sp	+	-	-	ND
13Pp_sp	+	ND	ND	ND
15Pp_sp	+	-	-	ND
19Pp_sp	+	-	-	ND
20Pp_sp	+	-	-	ND
22Pp_sp	+	-	-	ND
24Pp_sp	+	-	-	ND
25Pp_sp	+	-	-	ND
27Pp_sp	+	-	-	ND
28Pp_sp	+	-	-	ND
32Pp_sp	+	-	-	ND
41Pp_sp	-	ND	ND	ND
42Pp_sp	+	ND	ND	ND
43Pp_sp	+	-	-	ND
45Pp_sp	+	-	-	ND
46Pp_sp	+	-	-	ND
47Pp_sp	-	ND	ND	ND
49Pp_sp	+	-	-	ND
50Pp_sp	-	ND	ND	ND
51Pp_sp	-	ND	ND	ND
52Pp_sp	+	-	-	ND
53Pp_sp	+	-	-	ND
56Pp_sp	ND	ND	ND	ND
57Pp_sp	-	ND	ND	ND
58Pp_sp	-	ND	ND	ND
59Pp_sp	+	-	-	ND
60Pp_sp	+	ND	ND	ND

1 – (+) Análise positiva, segundo critérios descritos no texto; 2 – (-) Análise negativa, segundo critérios descritos no texto; 3 – (ND) Não determinado.

5 – DISCUSSÃO E CONCLUSÕES

Um grande obstáculo à descoberta de efetores fúngicos em bibliotecas de tecidos infectados é a sub-representação da biomassa fúngica, logo, o grande número de ESTs de origem vegetal na biblioteca da interação já era esperado (Lee. *et al.* 2006; Marques, 2007). Como relatado em trabalhos anteriores, um dos motivos para a grande proporção de genes sem qualquer similar nos bancos de dados se deve a ausência de sequências de fungos fitopatogênicos presentes em bancos de dados públicos, o que também acontece no patossistema *Eucalyptus grandis*-*P. psidii*, espécies de difícil manipulação genética. Embora as espécies de eucalipto tenham um grande apelo econômico, só recentemente foi firmado um consórcio público para o sequenciamento do seu genoma (Grattapaglia & Kirst, 2008).

O banco do NCBI apresentou menor poder de filtragem devido ao reduzido número de seqüências de eucalipto depositadas, apresentando ainda o problema de conter seqüências de transcritos de fungos ainda não caracterizadas, o que poderia filtrar indesejavelmente transcritos de *P. psidii*. Como esperado, o banco de ESTs do genolyptus, com a expectativa de amostragem de cerca de 40.000 mil genes (Grattapaglia, 2003), teve maior poder de filtragem, com apenas 78 transcritos não apresentando identidade as suas seqüências, mas apresentando nos outros dois bancos, sendo maior que o poder de filtragem do *draft* do genoma, que ainda apresenta um tamanho menor do que os 640 Mb estimados para o genoma (Grattapaglia & Bradshaw, 1994b).

Com o surgimento recente da caracterização de efetores fúngicos ficou clara a possibilidade de se encontrar proteínas de *P. psidii* ainda não caracterizadas em outros sistemas, ou mesmo, únicas para o fungo. Diante desta perspectiva, a estratégia traçada para encontrar os candidatos a efetores foi a filtragem por similaridade utilizando bancos de dados do hospedeiro, melhor caracterizado até então. A filtragem demonstrou potencial em revelar seqüências do fungo, uma vez que 762 transcritos não estavam presentes nos bancos de dados utilizados, o que corresponde a 8 % da biblioteca, número elevado, tendo em vista a estimativa anterior de apenas 2 % de seqüências fúngicas na biblioteca efetuada por Marques

(2007). No entanto, os bancos de dados não representam o genoma completo do hospedeiro e como demonstrado para o clone 56, algumas dessas sequências podem representar genes de eucalipto ainda não caracterizados por sequenciamento de cDNA ou genômico. A hipótese da subrepresentação de genes fungicos é corroborada pela baixa frequência de leituras observadas para as 29 ORFs supostamente fungicas.

Uma vez que o mecanismo de tradução esperado para *P. psidii* é o proposto para eucariotos, onde o ribossomo percorre o transcrito até encontrar o primeiro códon AUG para que inicie a tradução, e não há como para procariotos, uma sequência RBS (*Ribosomal Binding Site*) que marca o início da ORF, e além disso, não se conhece em que quadro de leitura ocorre a tradução, uma ORF, teoricamente, poderia ser predita em qualquer quadro de leitura, com um quadro de leitura podendo conter mais de uma ORF. Para tratar o problema de forma abrangente, foram utilizados *scripts* Perl para análise em larga escala. O primeiro foi adaptado do *script* descrito em Tisdall, 2001. A principal adaptação realizada foi a inserção de uma estrutura de repetição, conhecida em programação como “laço”, que permite a tradução simultânea de milhares de sequências. Além disso, foram inseridos códons com bases desconhecidas (N) para abranger os erros de sequenciamento, levando a um aminoácido desconhecido X, ou a um dos aminoácidos com mais de um códon quando a base N esta na última posição. Obtidas as traduções nos três quadros de leitura, estas foram submetidas a um programa que procura por ORFs que iniciem com o aminoácido metionina (M) e terminem ou não (para sequências 3' incompletas) com um dos três códons de parada do código genético (UGA, UAA, UAG) traduzidos como *, e que tenha pelo menos 40 aa.

A abordagem descrita acima permite a tradução simultânea de milhares de sequências, podendo, entretanto, levar a falsas predições. Kleemann *et al.*, 2008, depois da predição *ab initio* das ORFs pela tradução nos seis quadros de leitura de cada sequência única, obteve mais de 2200 potenciais ORFs, a maioria das quais supôs que seria artefato. No entanto, após a análise destas com o programa SignalP, apenas 167 ORFs foram preditas contendo um peptídeo sinal. Estas foram, então, cruzadas com as análises referências do BlastX das respectivas sequências originais

para eliminar traduções artefatuais de aminoácidos. Para merecer mais análises, as traduções candidatas tinham de ser seqüências órfãs ou terem resultados de BlastX consistentes, ou seja, uma seqüência homóloga significativa (E-value < 1e-5) utilizando o mesmo quadro de leitura. O exemplo citado acima ilustra o fato de que, a probabilidade de se obter uma ORF artefactual, e dentro desta ORF obter a predição de uma seqüência de exportação, é muito pequena, o que sugere que seja mais vantajoso fazer a tradução inicial de um número maior de seqüências, e só então realizar a predição, repetindo as comparações a bancos de dados se necessário.

Em contraste a oomicetos, proteínas efetoras de fungos identificadas até o momento, compartilham pouca similaridade entre si ou com seqüências de domínio público, e muitas vezes não têm seqüências estreitamente relacionadas em espécies diferentes do mesmo gênero, nem entre isolados da mesma espécie, mas com gamas de hospedeiros diferentes, ou agrupamentos de “*formae speciales*”. Por exemplo, o efetor Avr3 (Six1) está em todos os isolados de *Fusarium oxysporum* f.sp. *lycopersici* e ausentes em todos as outras *formae speciales* de *F. oxysporum* patogênicos e não patogênicos testados (Ellis *et al.*, 2009). Entretanto, de modo geral, as proteínas efetoras de fungos compartilham algumas características comuns: são normalmente pequenas, com menos de 400 aminoácidos de comprimento quando maduras, podem ser ricas em cisteína e geralmente possuem uma peptídeo sinal N-terminal para secreção, embora existam exceções, incluindo AVR_{a10} e AVR_{K1} de *Blumeria graminis* f. sp. *hordei* (Ridout *et al.*, 2006).

As ORFs dos candidatos a efetores de *P. psidii* identificados codificam peptídeos preditos variando de 80 a 245 aa, que apresentam peptídeo sinal, com pequena ou nenhuma similaridade a proteínas já caracterizadas, se enquadrando no perfil geral descrito para os efetores fúngicos que vem sendo caracterizados. Como discutido acima, estas proteínas também podem ser ricas em cisteína, o que poderia auxiliar em sua estabilidade pela formação de pontes dissulfídicas (Rep, 2005). As proteínas codificadas pelas ORFs dos clones 06, 13, 15, 19, 20, 22, 25, 43, 47, 52 e 57 apresentam 5% ou mais de cisteínas na sua composição, com predição positiva para formação de pontes dissulfeto, sendo portanto candidatos potenciais para a análise de sua função efetora na interação *P. psidii*-eucalipto.

Como a similaridade entre a maioria dos efetores de fungos com proteínas presentes em bancos públicos é limitada, não foi possível atribuir funções aos candidatos a efetores identificados. No entanto, esta abordagem mostrou ser bem sucedida com a identificação recente dos primeiros homólogos dos genes efetores de *C. fulvum* Avr4, ECP2 e ECP6 no genoma do fungo *Mycosphaerella fijiensis*, agente etiológico da Sigatoka Negra em bananeira (Stergiopoulos e de Wit, 2009) e a descoberta de um homólogo funcional do peptídeo tóxico ToxA de *Stagonospora nodorum* em *Pyrenophora tritici-repentis* (Friesen *et al.*, 2008), demonstrando que, na medida em que os bancos de dados de seqüências de fitopatógenos crescem, a similaridade entre proteínas de diferentes patógenos tende a ser revelada, tornando possível sugerir funções para essas, baseando-se nas estratégias comuns de infecção dos patógenos que as compartilham. Os genomas ainda não caracterizados dos fungos *P. graminis* e *M. laricis* apresentaram quatro proteínas hipotéticas, sem função descrita, similares às proteínas codificadas pelas ORFs de *P. psidii* identificadas, com destaque para o clone 06, que apresentou maior similaridade a proteínas dos dois genomas. Embora a coevolução das ferrugens com seus hospedeiros tenha levado a um alto nível de especificidade na suas estratégias de infecção é esperado que alguns dos genes sejam compartilhados a partir da descendência de ancestrais comuns, constituindo alvos interessantes para análises funcionais posteriores.

A extensa similaridade encontrada entre as proteínas codificadas por alguns dos clones analisados sugere que sejam seqüências de alelos diferentes, por se tratar de um fungo dicariótico, ou mesmo genes parálogos. No entanto, não se pode descartar a possibilidade de pertencerem a diferentes indivíduos, uma vez que, a biblioteca foi construída a partir de folhas infectadas naturalmente. Seqüências parálogas de efetores distribuídas ao longo do genoma tem sido descritas em outras espécies, como as proteínas PiE12, PiE14 e PiE23 de *P. infestans* descritas por Lee *et al.* (2006), que também foram observadas na seqüência do genoma de *Phytophthora sojae*. As seqüências 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 seqüências homólogas contêm seqüências que codificam peptídeos sinal de secreção, motivos RXLR-like, e seqüências de localização nuclear

preditos e foram encontrados agrupados em uma seqüência contígua de 20 kb no genoma de *P. sojae*, sugerindo que eles podem representar membros de uma família de genes. Embora seja possível discernir algumas seqüências com similaridade ao motivo RXLR em regiões N-terminais de efetores fúngicos (Li, *et al.*, 2008), nenhum sinal funcional para captação de efetores de fungos foi definido ainda, fato que dificulta uma predição mais apurada de efetores a partir da seqüência de cDNAs ou mesmo a partir de seqüências do genoma.

Uma bioinformática mais refinada tem reforçado a precisão de descoberta de efetores candidatos em oomicetos, e mostrou que muitos dos efetores contendo motivos RXLR são realmente distantemente relacionados e são membros de uma superfamília de proteínas que partilham três motivos C-terminais conservados adicionais chamados W, Y e L, muitas vezes repetidos em módulos (Dou *et al.*, 2008), que ocorrem em uma região sujeitas a seleção positiva para evadir a detecção de proteínas (R) do hospedeiro, assim como para se adaptar a diferentes alvos do metabolismo do hospedeiro. Várias centenas desses genes ocorrem no genoma de *P. sojae*, por exemplo, possivelmente como resultado de duplicação e diversificação de um único gene de um ancestral (Ellis *et al.*, 2009), o que também vem sendo observado em fungos patogênicos (Kamper *et al.* 2006), que possuem estratégias de infecção, reprodução e dispersão semelhantes, e portanto, devem estar expostos a pressões seletivas semelhantes.

A presença de alterações na seqüência de aminoácidos com a manutenção do quadro de leitura, observada para os candidatos efetores dos grupos de clones 02, 07, 46, 59 e 60, clones 13, 19, 43 e 47 e clones 45, 50, tem sido relatada como forma de evasão ao reconhecimento por proteínas R, mas com a manutenção da atividade efetora. Um papel direto na virulência foi demonstrado para os efetores Avr2 (Six3) e Avr3 (Six1) na interação tomate-*Fusarium oxysporum* f.sp. *lycopersici*. Mutantes com deleção induzidos de ambos os genes tornaram o patógeno menos virulento, especialmente em plantas mais velhas (Ellis *et al.*, 2009). Para *P. sojae*, um patógeno de soja, a perda de expressão de Avr1b não afetou a virulência, nem as mutações de mudança de quadro de leitura de Avr4 em *P. infestans*, provavelmente devido à redundância dos efetores (Ellis *et al.*, 2009). Para o gene Avr2 de *F. oxysporum*, isolados que superaram naturalmente o gene R correspondente em tomate, I2, contêm

apenas mudanças de aminoácidos que alteraram o reconhecimento pela proteína R, mas não afetam a virulência, separando estas 2 funções (Houterman *et al.*, 2009), o que é consistente com observações sobre a seqüência e estrutura em cristal das proteínas Avr567 de ferrugem do linho onde formas variantes são mantidas por raças virulentas e diferenças entre as proteínas reconhecidas ou não pelo proteína de resistência L6 de linho envolvem os resíduos expostos na superfície (Ellis *et al.*, 2009).

Em todos os casos descritos até agora, o ganho de virulência está relacionado com a diversificação de seqüência ou perda de genes efetores. No entanto, um cenário diferente foi observado para o efector Avr3 de *F. oxysporum* f. sp. lycopersici que é necessário para a virulência completa, onde a supressão do ETI é correlacionada com a aquisição de uma proteína adicional (Avr1) que suprime as vias de sinalização de defesa mediada por *I-3* e *I-2* (Houterman. *et al.*, 2008).

A elevada taxa de evolução molecular observada em efetores como um resultado de mutação, seleção e reprodução causou extensa diversificação de seqüência, expansão de genes, e outros rearranjos nos genes de efetores, que poderiam explicar a ausência de homologia entre proteínas efectoras e as suas especificidades a diferentes hospedeiros. Muitos genes efetores estão localizados em áreas altamente dinâmicas do genoma, tais como as extremidades do cromossomo ou em regiões ricas em elementos transponíveis que sofrem rearranjos genômicos freqüentes, permitindo assim maior variabilidade genética e facilidade de superar a resistência mediada por genes R, como já discutido anteriormente. Em muitos desses casos, o ganho de virulência devido a inserções de transposons, deleções de gene, e outros rearranjos genéticos tem sido frequentemente observados (Stergiopoulos & de Wit, 2009). Evidências iniciais sugerem que o genoma de *P. psidii* seja povoado de seqüências repetitivas que podem contribuir para a diversificação gênica deste patógeno (Milanez, 2009).

A secreção no sistema YST foi confirmada para 24 das 30 ORFs clonadas (o clone 56 não foi clonado neste sistema, pois trata-se de um gene de eucalipto), correspondendo a uma taxa de 20% de falsos positivos na predição *in silico*. Os clones 41, 57 e 58 foram selecionados com base na ausência de predição simultânea de hélices transmembrana pelos programas Phobius e TMHMM, mas haviam

apresentado predição de hélices em um dos dois programas, o que pode explicar a ausência da secreção, como observado em nosso laboratório para o homólogo da proteína HESP-329 clonado em *P. psidii*. No entanto, a estratégia de seleção foi vantajosa, uma vez que os clones 02, 06, 15, 53, 59 e 60 que também apresentavam predição para hélices transmembrana em apenas um algoritmo codificam proteínas que foram secretadas. A predição de seqüências de exportação celular tem permitido grandes avanços na clonagem e caracterização de efetores fúngicos, no entanto, como todas as ferramentas baseadas em previsões, há uma taxa de erro inerente e ocorrência de falsos positivos e negativos. Por exemplo, Torto *et al.* (2003) relataram uma taxa de erro de 17% de falsos positivos, o que não é incomum em tais programas. Como ilustrado com as análises do secretoma de bactérias, triagens experimentais fornecem um complemento importante para as análises de bioinformática por validar ou refutar funcionalmente as previsões. Isso enfatiza o valor da triagem por YST, para confirmação da secreção predita. No entanto, não podemos excluir a possibilidade de a predição ter eliminado genes que codificam proteínas secretadas. Lee *et al.* (2006) relataram que, dos 23 clones de *P. infestans* positivos pelo YST, apenas três (PiE1, PiE5 e PiE6) haviam sido previamente relatados nas análises bioinformática de Torto *et al.* (2003). Uma explicação para a identificação de seqüências sem peptídeo sinal é a existência de uma via de excreção "não-clássica" (ER Golgi ou independentes), como descrito em mamíferos e em eucariotos inferiores (Nickel, 2005). Uma revisão recente também catalogou mais de 40 proteínas secretadas, mas sem SP em levedura (Nombela *et al.* 2006), representando diversas rotas alternativas de secreção de proteínas. Outras exceções incluem os efetores AVR_{a10} e AVR_{K1} de *Blumeria graminis* f. sp. *hordei* que não apresentam peptídeo sinal de secreção (Ridout *et al.*, 2006).

As ORFs preditas foram amplificadas a partir do DNA genômico de *P. psidii* e *E. grandis*, e 29 das 31 ORFs apresentaram a amplificação de um fragmento igual ou ligeiramente maior que o predito para o transcrito, o que é coerente com a previsão de baixa frequência ou pequeno tamanho de íntrons relatados para eucariotos inferiores, com destaque para o fungo fitopatogênico biotrófico *U. maydis* que possui 70% dos seus genes sem íntrons (Kamper *et al.*, 2006). A clonagem posterior dos insertos genômicos poderá fornecer informações valiosas a respeito das

sequências das junções íntron-éxon, ajudando na predição de genes em sequências genômicas que serão geradas no futuro. O fato do clone 32 ter apresentado a amplificação de duas bandas pode se dever ao anelamento inespecífico ou mesmo a presença de genes parálogos no genoma, fato que também poderá ser elucidado pela clonagem e sequenciamento dos amplicons.

O protocolo de agroinfiltração utilizado no presente trabalho se encontra em fase de adaptação e foi utilizado pela primeira vez no laboratório para o ensaio de infecção descrito. A ausência de um controle que demonstre a taxa de infecção, como proteínas que sabidamente desencadeiam a morte celular, ou proteínas repórteres como GUS (Ridout *et al.*, 2006; Dou *et al.*, 2008), nos impede de extrair conclusões a respeito da ausência de reação do tabaco à infiltração com *Agrobacterium* contendo as ORFs dos vinte clones testados. A conservação das vias de reconhecimento e resposta a patógenos tem sido relatada para espécies pouco relacionadas filogeneticamente, o que torna possível a expressão heteróloga de proteínas de espécies de difícil manipulação genética em espécies modelo. A indução de reação hipersensível em *N. benthamiana* foi observada após a expressão transiente de um candidato efetor de *Magnaporthe oryzae* (Chen *et al.*, 2009), mas a expressão de efetores de *C. fulvum* em *N. tabacum* não induziram HR (Van der Hoorn *et al.*, 2000), demonstrando que a ausência de reação na expressão heteróloga pode refletir a especificidade ao hospedeiro e não a ausência de atividade efetora.

Muitas proteínas efetoras são reconhecidas por proteínas de resistência polimórficas do hospedeiro e assim podem ser identificadas em bioensaios como moléculas que desencadeiam respostas dependentes de proteína R do hospedeiro como a morte celular (reação de hipersensibilidade ou HR), quando expressa dentro ou direcionadas para o interior das células hospedeiras (Shan, *et al.*, 2004; Vleeshouwers *et al.*, 2008). Um gene dominante que confere a resistência a *Puccinia psidii* já foi mapeado no genoma de *Eucalyptus grandis* (Junghans *et al.*, 2003) e aparentemente, a resistência conferida por este gene está associada ao desencadeamento da resposta de hipersensibilidade, uma vez que em indivíduos resistentes, visualiza-se a necrose das células localizadas no sítio de penetração do patógeno (Xavier *et al.*, 2001). A região genômica que contém o gene já foi fisicamente clonada (Brommonschenkel *et al.*, 2009) e os genes candidatos

identificados codificam proteínas citoplasmáticas com domínios NBS-LRR, indicando possibilidade de identificar o efetor correspondente por meio de bioensaios de elicitação de morte celular. A expressão transiente por meio da agroinfiltração constitui uma alternativa atraente para obter indícios da atividade efetora de proteínas sem atividade conhecida, entretanto, algumas espécies como soja, arroz e milho apresentam dificuldades para infiltração do mesófilo foliar (Van der Hoorn *et al.*, 2000). O mesmo obstáculo foi encontrado para espécies de eucalipto tornando necessário o desenvolvimento de outras estratégias para investigação da atividade efetora.

Ensaio de co-bombardeamento de genes R e genes efetores, recentemente melhoradas com métodos que permitem um melhor controle interno (Dou *et al.*, 2008), figuram como uma alternativa atraente em espécies de eucalipto, pois independem da especificidade da infecção por *Agrobacterium*. Também seria promissor a co-inoculação de plantas de tabaco, nas quais é possível infiltrar o mesófilo foliar, com as ORFs dos candidatos a efetores e as ORFs dos genes R candidatos, permitindo o desencadeamento de HR em células que expressam as proteínas Avr e R, respectivamente (Van der Hoorn *et al.*, 2000; Ridout *et al.*, 2006). Recentemente um bioensaio para verificar a capacidade de indução das respostas de defesa a efetores candidatos foi desenvolvido, através da qual, proteínas heterólogas podem ser translocadas para o citoplasma das plantas através sistema de secreção do Tipo III via bactérias patogênicas geneticamente modificadas (Ellis *et al.*, 2009). Estes ensaios constituem uma alternativa para o eucalipto que é hospedeiro de espécies de *Xanthomonas* que causam lesões foliares (Gonçalves *et al.*, 2008). Contudo, na ausência de genótipos resistentes adequados, muitos efetores não podem ser identificados por função de avirulência. Essa restrição é particularmente pertinente para *U. maydis* onde nenhum gene de resistência correspondente de milho é conhecido (Ellis *et al.*, 2009).

Em conclusão, o presente trabalho empregou uma estratégia para identificação de ORFs sem função predita que possuem sequências para exportação celular descrevendo 29 novas proteínas hipotéticas de *P. psidii* que se enquadram nas características descritas para efetores fúngicos recentemente descritos. Nove destas ORFs apresentaram similares em genomas de fitopatógenos, corroborando o

predição *ab initio* realizada. A secreção de 24 proteínas foi confirmada pelo sistema YST, fornecendo indícios adicionais da sua secreção por *P. psidii* e fortalecendo o seu potencial de efetor. Novos ensaios se encontram em condução para otimização da expressão transiente das proteínas, bem como para o monitoramento da sua expressão ao longo do ciclo da doença, para comprovação da atividade efetora das proteínas identificadas.

6 - REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAF. **Anuário estatístico da ABRAF 2009, ano base 2008**. Associados da ABRAF 2009; 31p.
- ALFENAS, A.C.; ZAUZA, E. A. V.; MÁFIA, R. G.; ASSIS, T. F. **Clonagem e Doenças do Eucalipto**. Editora UFV, 2004; 442p.
- ALVES, A. A. **Herança e mapeamento genético da resistência à ferrugem (*Puccinia psidii*) em cruzamentos interespecíficos de *Eucalyptus***. Viçosa, MG: UFV, 2008. 39p. Dissertação de Mestrado (Genética e Melhoramento). Universidade Federal de Vicososa. 2008.
- ARMSTRONG M. R., WHISSON S. C., PRITCHARD L., BOS J. I., VENTER E., AVROVA A. O., REHMANY A. P., BOHME U., BROOKS K., CHEREVACH I., HAMLIN N., WHITE B., FRASER A., LORD A., QUAIL M. A., CHURCHER C., HALL N., BERRIMAN M., HUANG S., KAMOUN S., BEYNON J. L., AND BIRCH P. R. 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, USA**, v.102, p.7766-7771, 2005.
- BOWEN J. K.; MESARICH, C. H.; REES-GEORGE, J.; CUI, W.; FITZGERALD, A.; WIN, J.; PLUMMER K. M.; TEMPLETON, M. D. Candidate effector gene identification in the ascomycete fungal phytopathogen *Venturia inaequalis* by expressed sequence tag analysis. **Molecular Plant Pathology**, v.10, n.3, p.431–448, 2009.
- BROMMONSCHENKEL, S. H. ; MELO, J. DE O. ; CAMARGOS, J.O. Map-based cloning of the *Eucalyptus grandis* genomic region that encompass the rust resistance gene Ppr-1. **Internacional Congress on Molecular Plant-Microbe Interactions**, <http://www.ismpminet.org/meetings/abstracts/2009/c09ma10.asp>, 2009.
- 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. **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.
- 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, USA**, v.102, p.2549-2554,2005.
- CHEN, S.; GOWDA, M.; VENU, R. C.; SONGKUMARN, P.; PARK, C. H.; BELLIZZI, M.; EBBOLE, D. J.; WANG, G-L. Isolation and Functional Analysis of Putative Effectors from *Magnaporthe oryzae* Using Integrated Genomic Approaches. p93-103. In: Guo-Liang Wang and

Barbara Valent. **Advances in Genetics, Genomics and Control of Rice Blast Disease**, DOI 10.1007/978-1-4020-9500-9_10, 2009.

COUTINHO, T. A.; WINGFIELD, M. J.; ALFENAS, A. C.; CROUS, P. W. Eucalyptus rust: A disease with the potential for serious international implications. **Plant Disease**, v.82, p.819-825, 1998.

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.

DANGL, J. L. AND MCDOWELL, J. M. Two modes of pathogen recognition by plants. **Proceedings of the National Academy of Sciences, USA**, v.103, n.23, p.8575–8576, 2006.

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., REHMEYER, 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.

DOEHLEMANN, G.; VAN DER LINDE, K.; ABMANN, D.; SCHWAMMBACH, D.; HOF, A.; MOHANTY, A.; JACKSON, D.; KAHMANN, R. Pep1, a secreted effector protein of *Ustilago maydis* is required for successful invasion of plant cells. **PLoS Pathogens**, v.5, p.e100290, 2009.

DOU, D.; KALE, S. D.; WANG, X.; CHEN, Y.; WANG, Q.; WANG, X.; JIANG, R. H.; ARREDONDO, F. D.; ANDERSON, R. G.; THAKUR, P. B.; et al.: Conserved Cterminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. **Plant Cell**, v.20, p.1118-1133, 2008.

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.

FIGUEIREDO, M. B. E PASSADOR, M. M. Morfologia, funções dos soros e variações dos ciclos vitais das ferrugens. **Arquivos do Instituto Biológico**, v.75, n.1, p.117-134, 2008.

FLOR, H.H. The complementary genetic systems in flax and flax rust. **Advances Genetics**, v.8, n.1, p.29-54, 1956.

FRIESEN, T. L.; FARIS, J. D.; SOLOMON, P. S.; OLIVER, R. P. Host-specific toxins: effectors of necrotrophic pathogenicity. **Cell Microbiology**, v.10, p.1421–28, 2008.

GIETZ, R.D. AND R.A. WOODS. Transformation of yeast by the liac/ss carrier dna/peg method. **Methods in Enzymology**, v.350, p.87-96, 2002.

GLEN, M.; ALFENAS, A. C.; ZAUZA, E. A. V.; WINGFIELD, M. J.; MOHAMMED, C. *Puccinia psidii*: a threat to the Australian environment and economy - a review. **Australasian Plant Pathology**, v.36, p.1-16, 2007.

GONÇALVES, R. C., LAU, D.; OLIVEIRA, J. R.; MAFFIA, L. A., CASCARDO, J. C. M., ALFENAS, A. C. Etiology of bacterial leaf blight of eucalyptus in Brazil. **Tropical Plant Pathology**, v.33, n.30, p.180-188, 2008.

GRATTAPAGLIA, D. AND SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, v.137, p.1121–1137, 1994a.

GRATTAPAGLIA, D. AND BRADSHAW, H. D. Nuclear DNA Content Of Commercially Important *Eucalyptus* Species And Hybrids. **Canadian Journal of Forest Research**, v.24(5), p.1074-1078, 1994b.

GRATTAPAGLIA, D. Genolyptus. In: Aluizio Borém. (Org.). **Melhoramento Genômico**. Viçosa: Editora da UFV, 2003; v. 1, p. 51-71.

GRATTAPAGLIA, D. AND KIRST, M. Eucalyptus applied genomics: from gene sequences to breeding tools. **New Phytologist**, v.179, p.911–929, 2008.

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.

HOUTERMAN, P. M.; CORNELISSEN, B. J. C.; REP, M. Suppression of plant resistance gene-based immunity by a fungal effector. **PLoS Pathogens**, v.4, n.5, doi:10.1371/journal.ppat.1000061, 2008.

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.

JONES, J. D. AND DANGL, J. L. The plant immune system. **Nature**, v.444, p.323–329, 2006.

JOOSTEN, M. H. A. J.; VOGELANG, R.; COZIENSEN, T. J.; VERBERNE, M. C.; DE WIT, 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.

JUNGHANS, D. T.; ALFENAS, A. C.; BROMMONSCHENKEL, S. H.; ODA, S.; MELLO, E. J.; GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**, v.108, p.175–180, 2003.

KAMPER, J.; KAHMANN, R.; BOLKER, M.; MA, L. J.; BREFORT, T.; SAVILLE, B. J.; BANUETT, F.; KRONSTAD, J. W.; GOLD, S. E.; MULLER, O.; ET AL. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. **Nature**, v.444, p.97-101, 2006.

KARIMI, M.; INZE, D.; DEPICKER, A. GATEWAY™ vectors for Agrobacterium-mediated plant transformation. **Trends Plant Science**, v.7, p.193–195, 2002.

KAWANISHI, T.; UEMATSU, S.; KAKISHIMA, M.; KAGIWADA, S.; HAMAMOTO, H.; HORIE, H.; NAMBA, S. First report of rust disease on ohia and the causal fungus, *Puccinia psidii*, in Japan. **Journal of General Plant Pathology**, v.75, n.6, p.428-431, 2009.

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.

KEMEN, E.; KEMEN, A. C.; RAFIQI, M.; HEMPEL, U.; MENDGEN, K.; HAHN, M.; VOEGELE, 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.

KLEIN, R.D., GU, Q., GODDARD, A. AND ROSENTHAL, A. Selection for genes encoding secreted proteins and receptors. **Proceedings of the National Academy of Sciences, USA**, v.93, p.7108–7113, 1996.

LAROUS, L.; KAMELI, A.; LÖSEL, D. M. Ultrastructural observations on *Puccinia menthae* infections. **Journal of Plant Pathology**, v.90, n.2, p.185-190, 2008.

LEE, S.A., WORMSLEY, S., KAMOUN, S., LEE, A.F., JOINER, K. AND WONG, B. An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms. **Yeast**, v.20, p.595–610, 2003.

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, n.12, 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, 2008. ICBBE 2008. The 2nd International Conference**, v.16, n.18, p.286 –292, 2008.

LINK, T. I. AND VOEGELE, R. T. Secreted proteins of *Uromyces fabae*: similarities and stage specificity. **Molecular Plant Pathology**, v.9, n.1, p.59–66, 2008.

LUCAS, J.A. Survival, surfaces and susceptibility – the sensory biology of pathogens. **Plant Pathology**, v.53, p.679–691, 2004.

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.875–84, 2002.

MAIER, W.; BEGEROW, D.; WEIß, M.; OBERWINKLER, F. Phylogeny of the rust fungi: an approach using nuclear large subunit ribosomal DNA sequences. **Canadian Journal of Botany**, v.81, p.12–23, 2003.

MARQUES, V. M. **Identificação e caracterização “in silico” de genes de *Puccinia psidii* Winter expressos na interação com *Eucalyptus grandis***. Viçosa, MG:UFV, 2007. 26p. Monografia -. Universidade Federal de Viçosa, 2007.

MENDGEN, K. AND HAHN, M. Plant infection and the establishment of fungal biotrophy. **Trends in Plant Science**, v.7, p.352–356, 2002.

MILANEZ, G. P. **Sequenciamento amostral e análise genômica do fungo basidiomiceto *Puccinia psidii*, agente causal da ferrugem do eucalipto**. Viçosa, MG: UFV, 2009. 30p. Monografia - Universidade Federal de Viçosa, 2009.

MOSQUERA, G.; GIRALDO, M. C.; KHANG, C. H.; COUGHLAN, S.; VALENT, B. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as Biotrophy-Associated Secreted proteins in rice blast disease. **Plant Cell**, v.21, p.1273-1290, 2009.

MÜLLER, O.; SCHREIER, P. H.; UHRIG, J. F. Identification and characterization of secreted and pathogenesis-related proteins in *Ustilago maydis*. **Molecular Genetics and Genomics**, v.279, p.27–39, 2008.

NICKEL, W. Unconventional secretory routes: Direct protein export across the plasma membrane of mammalian cells. **Traffic**, v.6, p.607-614, 2005.

NOMBELA, C.; GIL, C.; CHAFFIN, W. L. Non-conventional protein secretion in yeast. **Trends in Microbiology**, v.14, p.15-21, 2006.

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.; VAN DER DOES, H. C.; MEIJER, M.; VAN WIJK, R.; HOUTERMAN, P. M.; ET AL. A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. **Molecular Microbiology**, v.53, p.1373-83, 2004.

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.

SHABAB, M.; SHINDO, T.; GU, C.; KASCHANI, F.; PANSURIYA, T.; ET AL. Fungal effector protein AVR2 targets diversifying defense-related cysteine proteases of tomato. **Plant Cell**, v.20, p.1169-83, 2008.

SHAN, W., CAO, M., LEUNG, D., AND 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 Interaction**, 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.

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, USA**, 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.

TIAN, M.; HUITEMA, E.; DA CUNHA, L.; TORTO-ALALIBO, T.; KAMOUN, S. A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. **The Journal of Biological Chemistry**, v. 279, p.26370–77, 2004.

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

TORTO, T. A., LI, S., STYER, A., HUITEMA, E., TESTA, A., GOW, N. A. R., VAN WEST, P., AND KAMOUN, S. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. **Genome Research**, v.13, p.1675-1685, 2003.

VAN DER HOORN, R. A. L.; LAURENT, F.; ROTH, R.; DE WIT, P. J. G. M. Agroinfiltration Is a Versatile Tool That Facilitates Comparative Analyses of Avr9/Cf-9-Induced and Avr4/Cf-4-Induced Necrosis. **Molecular Plant Microbe Interaction**, v.13(4), p.439–446, 2000.

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 KAN, J. A. L.; VAN DEN ACKERVEKEN, G. F. J. M.; DE WIT, P. J. G. M. Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. **Molecular Plant Microbe Interaction**, v.4, p.52–59, 1991.

VLEESHOUWERS, V. G.; RIETMAN, H.; KRENEK, P.; CHAMPOURET, N.; YOUNG, C.; OH, S. K.; WANG, M.; BOUWMEESTER, K.; VOSMAN, B.; VISSER, R. G.; et al. Effector genomics accelerates discovery and functional profiling of potato disease resistance and *phytophthora infestans* avirulence genes. **PLoS ONE**, v.3, p.e2875, 2008.

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.

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.

XAVIER, A. A.; ALFENAS, A. C.; MATSUOKA, K.; HODGES, C. S. Infection of resistant and susceptible *Eucalyptus grandis* genotypes by urediniospores of *Puccinia psidii*. **Australasian Plant Pathology**, v.30(3), p.277–281, 2001.

ZOTTINI, M.; BARIZZA, E.; COSTA, A., FORMENTIN, E.; RUBERTI, C.; CARIMI, F.; SCHIAVO, F. LO. Agroinfiltration of grapevine leaves for fast transient assays of gene expression and for long-term production of stable transformed cells. **Plant Cell Reports**, v.27, p.845–853, 2008.

ANEXO

```
#!/usr/bin/perl

#####
#                                                                 #
# Um programa que traduz uma seqüência de nucleotídeo nos três quadros de      #
# leitura positivos                                             #
# Adaptado de Tisdall, 2001                                       #
#                                                                 #
#####

    use strict;
    use warnings;

# Declaração das variáveis

    my @arquivo_de_sequencias = ();
    my @lista_de_nomes = ();
    my @lista_de_sequencias = ();
    my $proteína = "";
    my $contador_de_sequencias = 0;

# Lê o conteúdo de um arquivo de seqüências e armazena em uma "lista" na
# memória do computador

    @arquivo_de_sequencias = <>;

    foreach my $linha (@arquivo_de_sequencias) {

        if ( $linha =~ />/ ) {

            $linha =~ s/(>[^\s]{7,26})(\s)?.*$/1/;
            $lista_de_nomes[$contador_de_sequencias] = $linha;

            chomp($lista_de_nomes[$contador_de_sequencias] = $linha);
                ++$contador_de_sequencias;
            }
            else {
                chomp($lista_de_sequencias[$contador_de_sequencias] .= sprintf "%U$linha");
            }
        }
    }

# traduz o DNA para os três quadros de leitura positivos
# e imprime em linhas com no máximo 70 aa
# para executar essa tarefa são utilizadas funções definidas pelo usuário
# conhecidas em Perl como sub-rotinas
```

```

for (my $c = 0; $c < (@lista_de_sequencias - 1); ++$c) {

    print "$lista_de_nomes[$c]F1\n";
    $proteina = translate_frame($lista_de_sequencias[($c + 1)], 1);
    print_sequence($proteina, 70);

    print "$lista_de_nomes[$c]F2\n";
    $proteina = translate_frame($lista_de_sequencias[($c + 1)], 2);
    print_sequence($proteina, 70);

    print "$lista_de_nomes[$c]F3\n";
    $proteina = translate_frame($lista_de_sequencias[($c + 1)], 3);
    print_sequence($proteina, 70);
}

#####
#                                                                 #
#                               SUB-ROTINAS                       #
#                                                                 #
#####

# translate_frame
#
# Uma sub-rotina para traduzir um quadro de leitura do DNA

sub translate_frame {

    my($seq, $start, $end) = @_ ;

    my $protein;

    # Para tornar a utilização da sub-rotina mais fácil,
    # o final da seqüência não precisa ser especificado
    # ela irá para o fim da seqüência por padrão

    unless($end) {
        $end = length($seq);
    }

    # Calcule e retorne a tradução
    return dna2peptide ( substr ( $seq, $start - 1, $end - $start + 1 ) );
}

# dna2peptide
#
# Uma sub-rotina para traduzir uma seqüência de DNA em um peptídeo

sub dna2peptide {

    my($dna) = @_ ;

    use strict;

```



```

use warnings;

my $protein = "";

# Traduz cada codon de três bases para um aminoácido e adiciona a proteína
for(my $i=0; $i < (length($dna) - 2) ; $i += 3) {
    $protein .= codon2aa( substr($dna,$i,3) );
}

return $protein;
}

# codon2aa
#
# Uma sub-rotina para traduzir um códon de 3 caracteres de uma seqüência de # DNA para
# aminoácido
# Version 3, using hash lookup

sub codon2aa {
    my($codon) = @_ ;

    $codon = uc $codon;

    my(%genetic_code) = (
        'TCA' => 'S', # Serine
        'TCC' => 'S', # Serine
        'TCG' => 'S', # Serine
        'TCT' => 'S', # Serine
        'TCN' => 'S', # Serine
        'TCX' => 'S', # Serine
        'TTC' => 'F', # Phenylalanine
        'TTT' => 'F', # Phenylalanine
        'TTA' => 'L', # Leucine
        'TTG' => 'L', # Leucine
        'TAC' => 'Y', # Tyrosine
        'TAT' => 'Y', # Tyrosine
        'TAA' => '*', # Stop
        'TAG' => '*', # Stop
        'TGC' => 'C', # Cysteine
        'TGT' => 'C', # Cysteine
        'TGA' => '*', # Stop
        'TGG' => 'W', # Tryptophan
        'CTA' => 'L', # Leucine
        'CTC' => 'L', # Leucine
        'CTG' => 'L', # Leucine
        'CTT' => 'L', # Leucine
        'CTN' => 'L', # Leucine
        'CTX' => 'L', # Leucine
        'CCA' => 'P', # Proline
        'CCC' => 'P', # Proline
        'CCG' => 'P', # Proline
        'CCT' => 'P', # Proline
        'CCN' => 'P', # Proline
    );
}

```

```

'CCX' => 'P', # Proline
'CAC' => 'H', # Histidine
'CAT' => 'H', # Histidine
'CAA' => 'Q', # Glutamine
'CAG' => 'Q', # Glutamine
'CGA' => 'R', # Arginine
'CGC' => 'R', # Arginine
'CGG' => 'R', # Arginine
'CGT' => 'R', # Arginine
'CGN' => 'R', # Arginine
'CGX' => 'R', # Arginine
'ATA' => 'I', # Isoleucine
'ATC' => 'I', # Isoleucine
'ATT' => 'I', # Isoleucine
'ATG' => 'M', # Methionine
'ACA' => 'T', # Threonine
'ACC' => 'T', # Threonine
'ACG' => 'T', # Threonine
'ACT' => 'T', # Threonine
'ACN' => 'T', # Threonine
'ACX' => 'T', # Threonine
'AAC' => 'N', # Asparagine
'AAT' => 'N', # Asparagine
'AAA' => 'K', # Lysine
'AAG' => 'K', # Lysine
'AGC' => 'S', # Serine
'AGT' => 'S', # Serine
'AGA' => 'R', # Arginine
'AGG' => 'R', # Arginine
'GTA' => 'V', # Valine
'GTC' => 'V', # Valine
'GTG' => 'V', # Valine
'GTT' => 'V', # Valine
'GTN' => 'V', # Valine
'GTX' => 'V', # Valine
'GCA' => 'A', # Alanine
'GCC' => 'A', # Alanine
'GCG' => 'A', # Alanine
'GCT' => 'A', # Alanine
'GCN' => 'A', # Alanine
'GCX' => 'A', # Alanine
'GAC' => 'D', # Aspartic Acid
'GAT' => 'D', # Aspartic Acid
'GAA' => 'E', # Glutamic Acid
'GAG' => 'E', # Glutamic Acid
'GGA' => 'G', # Glycine
'GGC' => 'G', # Glycine
'GGG' => 'G', # Glycine
'GGT' => 'G', # Glycine
'GGN' => 'G', # Glycine
'GGX' => 'G', # Glycine
);

if(exists $genetic_code{$scodon}) {

```

```

        return $genetic_code{$scodon};
    }else{
        return 'X';
    }
}

# print_sequence
#
# Uma sub-rotina para formatar e imprimir uma seqüência

sub print_sequence {

    my($sequence, $length) = @_ ;

    use strict;
    use warnings;

    # Imprime a seqüência em linha de tamanho especificado por $length

    for ( my $pos = 0 ; $pos < length($sequence) ; $pos += $length ) {
        print substr($sequence, $pos, $length), "\n";
    }
}

```

```

#!/usr/bin/perl
#####
# Um programa que encontra ORFs com mais de 40aa em seqüências #
# polipeptídicas #
#####

```

```

    use strict;
    use warnings;

```

```

# Declaração das variáveis

```

```

    my @arquivo_de_sequencias = ();
    my @lista_de_nomes = ();
    my @lista_de_sequencias = ();
    my $contador_de_sequencias = 0;

```

```

# Lê o conteúdo de um arquivo de seqüências e armazene em uma "lista" na
# memória do computador

```

```

    @arquivo_de_sequencias = <>;

```

```

    foreach my $linha (@arquivo_de_sequencias) {

```

```

        if( $linha =~ />/ ) {

```

```

        chomp( $lista_de_nomes[$contador_de_sequencias] = $linha );
        ++$contador_de_sequencias;
    }
    else {
        chomp( $lista_de_sequencias[$contador_de_sequencias] .= $linha );
    }
}

# Percorre o conteúdo das listas em busca de "frases" que iniciem com m,
# com pelo menos 40aa e que terminem ou não com *, pode haver mais de uma
# ORF por sequência

    for (my $c = 0; $c < (@lista_de_sequencias -1); ++$c) {

my $numero_da_orf = 1;
my @numero_de_aa;

        while ( @lista_de_sequencias[($c + 1)] =~ /(\w*)(m+\w*(\w*)?)/i
                and defined ( @lista_de_sequencias[($c + 1)] ) ) {

if ( @lista_de_sequencias[($c + 1)] =~ /(\w*)(m+\w*(\w*)?)/i ) {

        @lista_de_sequencias[($c + 1)] = $';
        @numero_de_aa = split(" ", $2);

        my $numero_de_aa = @numero_de_aa;

        print "$lista_de_nomes[$c]-$numero_da_orf", " ", $numero_de_aa, "\n" if
@numero_de_aa > 40;

        print "$2\n" if @numero_de_aa > 40;

        ++$numero_da_orf;
        }
    }
}

```

Tabela 1 – Tabela de ORFs com 40 ou mais aminoácidos, que apresentaram predição positiva (Y) em cinco ou mais parâmetros do programa SignalP e foram analisadas com outros algoritmos de predição.

Nome do clone	Nº AA	SignalP-NN	SignalP-HMM	targetp v1.1		Tmhmm	Phobius	
		D-score ¹	Sprob ²	Loc ³	RC ⁴		TM ⁶	SP ⁷
Contig114F2-2	134	0,546	0,972	S	2	PredHel=0 ⁵	0	Y
Contig140F1-1	120	0,511	0,676	S	1	PredHel=1	1	0
Contig140F3-1	153	0,874	1	S	1	PredHel=1	0	Y
Contig217F3-1	151	0,891	0,999	S	1	PredHel=0	0	Y
Contig269F2-1	99	0,913	1	S	1	PredHel=0	0	Y
Contig437F1-1	215	0,709	0,992	S	1	PredHel=0	0	Y
Contig575F2-1	170	0,886	0,998	S	1	PredHel=0	0	Y
Contig585F2-1	107	0,749	0,999	S	1	PredHel=1	0	Y
Contig588F3-2	56	0,495	0,973	S	1	PredHel=0	1	0
Contig589F1-1	168	0,929	0,999	S	1	PredHel=0	0	Y
Contig621F2-1	162	0,903	0,999	S	1	PredHel=0	0	Y
Contig744F3-1	137	0,85	1	S	1	PredHel=1	0	Y
EUGR-PU-003-057-C08-GO.FF3-1	105	0,747	0,963	S	2	PredHel=0	0	Y
EUGR-PU-003-057-F09-GO.FF3-2	52	0,497	0,982	_	4	PredHel=0	0	Y
EUGR-PU-003-058-A04-GO.FF3-4	52	0,503	0,067	S	1	PredHel=1	1	0
EUGR-PU-003-058-A11-GO.FF1-5	55	0,442	0,99	S	3	PredHel=0	0	Y
EUGR-PU-003-058-G05-GO.FF1-2	43	0,622	0,946	S	4	PredHel=0	0	Y
EUGR-PU-003-060-E06-GO.FF1-2	49	0,466	0,867	_	4	PredHel=0	0	Y
EUGR-PU-003-064-C03-GO.FF1-1	87	0,612	0,939	S	1	PredHel=2	2	0
EUGR-PU-003-064-D07-GO.FF1-1	94	0,823	1	S	1	PredHel=0	0	Y
EUGR-PU-003-065-A11-GO.FF2-2	43	0,504	0,14	M	2	PredHel=1	1	0
EUGR-PU-003-065-H11-GO.FF3-1	90	0,686	0,998	S	1	PredHel=0	0	Y
EUGR-PU-003-066-A10-GO.FF3-1	106	0,851	0,999	S	1	PredHel=1	0	Y
EUGR-PU-003-066-C09-GO.FF1-1	144	0,875	0,999	S	1	PredHel=1	0	Y
EUGR-PU-003-066-D03-GO.FF3-1	95	0,885	0,999	S	1	PredHel=1	0	Y
EUGR-PU-003-067-A04-GO.FF1-1	47	0,504	0,395	S	5	PredHel=1	2	0
EUGR-PU-003-067-A04-GO.FF2-1	134	0,63	0,986	S	4	PredHel=1	1	Y
EUGR-PU-003-067-A04-GO.FF3-1	155	0,482	0,751	S	4	PredHel=4	3	Y
EUGR-PU-003-067-D07-GO.FF3-1	53	0,633	0,989	S	3	PredHel=0	0	0
EUGR-PU-003-067-D09-GO.FF2-1	176	0,92	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-068-C02-GO.FF1-1	104	0,88	0,999	S	1	PredHel=1	0	Y
EUGR-PU-003-068-H10-GO.FF1-1	172	0,597	0,785	S	1	PredHel=1	1	0
EUGR-PU-003-069-A04-GO.FF2-1	63	0,682	0,985	S	2	PredHel=0	0	Y
EUGR-PU-003-072-D12-GO.FF3-2	75	0,526	0,866	S	1	PredHel=0	0	Y
EUGR-PU-003-072-H12-GO.FF2-1	102	0,771	0,993	S	4	PredHel=1	0	Y
EUGR-PU-003-071-F01-GO.FF1-1	59	0,47	0,751	S	2	PredHel=0	0	0
EUGR-PU-003-071-G05-GO.FF2-2	60	0,719	0,996	S	1	PredHel=1	0	Y
EUGR-PU-003-055-C02-GO.FF2-2	43	0,526	0,675	S	2	PredHel=1	1	0
EUGR-PU-003-055-G03-GO.FF1-1	45	0,662	0,992	S	4	PredHel=0	0	Y
EUGR-PU-003-055-C02-GO.GF2-2	43	0,526	0,675	S	2	PredHel=1	1	0
EUGR-PU-003-055-G03-GO.GF1-1	45	0,662	0,992	S	4	PredHel=0	0	Y
EUGR-PU-003-056-B01-GO.FF2-1	62	0,432	0,659	M	4	PredHel=0	0	Y
EUGR-PU-003-056-D04-GO.FF1-3	60	0,76	1	S	4	PredHel=0	0	Y
EUGR-PU-003-056-H11-GO.FF1-1	49	0,733	0,977	S	2	PredHel=0	0	Y
EUGR-PU-003-062-B08-GO.FF1-1	104	0,699	0,968	S	1	PredHel=1	0	Y

EUGR-PU-003-062-E06-GO.FF1-1	65	0,616	0,785	S	1	PredHel=2	2	0
EUGR-PU-003-062-H11-GO.FF2-1	69	0,882	0,999	S	1	PredHel=1	0	Y
EUGR-PU-003-077-G12-GO.FF1-1	53	0,531	0,84	S	5	PredHel=0	0	Y
EUGR-PU-003-080-A07-GO.FF2-1	69	0,862	0,998	S	1	PredHel=0	0	Y
EUGR-PU-003-080-B01-GO.FF3-1	117	0,911	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-080-G09-GO.FF3-1	152	0,524	0,845	M	1	PredHel=0	0	0
EUGR-PU-003-075-B06-GO.FF2-1	141	0,527	0,877	S	4	PredHel=0	0	0
EUGR-PU-003-075-B12-GO.FF2-2	70	0,623	0,99	S	4	PredHel=0	0	Y
EUGR-PU-003-076-G05-GO.FF2-2	50	0,503	0,834	S	1	PredHel=1	0	Y
EUGR-PU-003-076-G07-GO.FF1-2	50	0,503	0,834	S	1	PredHel=1	0	Y
EUGR-PU-003-079-C09-GO.FF3-3	111	0,569	0,6	S	1	PredHel=1	1	0
EUGR-PU-003-079-H04-GO.FF1-1	113	0,628	0,976	S	1	PredHel=1	1	0
EUGR-PU-000-001-D10-MG.RF3-2	165	0,613	0,667	M	5	PredHel=0	0	Y
EUGR-PU-000-001-E03-MG.RF1-1	67	0,77	0,878	S	1	PredHel=2	1	Y
EUGR-PU-000-005-B03-MG.RF2-1	149	0,898	0,997	S	1	PredHel=0	0	Y
EUGR-PU-000-006-E03-MG.RF2-1	45	0,69	0,533	S	2	PredHel=1	1	0
EUGR-PU-000-008-C06-MG.RF3-1	54	0,595	0,912	S	1	PredHel=1	1	0
EUGR-PU-000-008-D10-MG.RF2-1	63	0,538	0,953	S	3	PredHel=1	0	Y
EUGR-PU-000-008-H12-MG.RF3-1	64	0,632	0,297	S	1	PredHel=0	1	0
EUGR-PU-000-009-E07-MG.RF3-1	78	0,667	0,946	S	4	PredHel=0	0	0
EUGR-PU-000-009-E12-MG.RF1-1	128	0,886	0,999	S	1	PredHel=0	0	Y
EUGR-PU-000-010-H03-MG.RF2-1	48	0,759	0,101	S	2	PredHel=1	1	0
EUGR-PU-000-011-B08-MG.RF2-1	102	0,762	0,914	S	2	PredHel=0	0	Y
EUGR-PU-000-011-E10-MG.RF1-2	44	0,659	0,159	S	3	PredHel=1	1	0
EUGR-PU-000-011-H09-MG.RF2-4	60	0,444	0,102	S	2	PredHel=0	0	0
EUGR-PU-000-012-G11-MG.RF3-2	71	0,605	0,8	S	3	PredHel=0	2	0
EUGR-PU-000-015-G03-MG.RF3-1	64	0,632	0,297	S	1	PredHel=0	1	0
EUGR-PU-000-016-D04-MG.RF3-1	62	0,457	0,818	S	5	PredHel=0	0	0
EUGR-PU-000-018-G08-MG.RF2-1	101	0,902	1	S	1	PredHel=3	0	Y
EUGR-PU-000-020-F11-MG.RF2-1	110	0,904	0,999	S	1	PredHel=0	0	Y
EUGR-PU-000-020-G11-MG.RF2-1	46	0,552	0,656	S	5	PredHel=0	0	Y
EUGR-PU-000-020-H12-MG.RF2-2	42	0,717	0,804	S	2	PredHel=0	0	Y
EUGR-PU-000-021-E06-MG.RF2-1	47	0,671	0,901	S	2	PredHel=0	0	Y
EUGR-PU-000-022-E03-MG.RF1-1	46	0,847	0,99	S	1	PredHel=1	0	Y
EUGR-PU-000-022-H06-MG.RF2-5	66	0,439	0,815	S	4	PredHel=1	0	Y
EUGR-PU-000-022-H10-MG.RF1-3	52	0,657	0,92	S	3	PredHel=0	0	Y
EUGR-PU-003-001-D12-MG.RF3-1	78	0,736	0,999	S	4	PredHel=0	0	Y
EUGR-PU-003-001-E11-MG.RF1-1	47	0,916	0,992	S	1	PredHel=1	0	Y
EUGR-PU-003-001-F06-MG.RF2-6	56	0,535	0,506	S	5	PredHel=0	0	Y
EUGR-PU-003-002-F09-MG.RF2-2	178	0,495	0,66	S	4	PredHel=0	0	Y
EUGR-PU-003-003-F08-MG.RF2-1	79	0,571	0,841	S	3	PredHel=0	0	Y
EUGR-PU-003-007-E09-MG.RF2-1	116	0,896	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-008-B06-MG.RF2-2	157	0,641	0,487	S	3	PredHel=0	0	Y
EUGR-PU-003-005-B05-MG.RF1-1	53	0,502	0,938	S	4	PredHel=0	0	Y
EUGR-PU-003-011-G04-MG.RF1-1	41	0,621	0,949	S	4	PredHel=0	0	Y
EUGR-PU-003-012-F05-MG.RF2-1	203	0,91	0,998	S	1	PredHel=0	0	Y
EUGR-PU-003-013-C11-MG.RF1-1	146	0,636	0,742	S	2	PredHel=0	0	Y
EUGR-PU-003-016-A01-MG.RF1-1	80	0,908	1	S	1	PredHel=1	0	Y
EUGR-PU-003-016-D08-MG.RF2-1	62	0,549	0,613	S	1	PredHel=2	2	0
EUGR-PU-003-016-E10-MG.RF3-1	207	0,838	0,999	S	2	PredHel=0	0	Y
EUGR-PU-003-017-D09-MG.RF2-3	53	0,752	0,844	S	5	PredHel=0	0	Y

EUGR-PU-003-017-G11-MG.RF3-5	86	0,54	0,439	M	3	PredHel=0	0	0
EUGR-PU-003-019-B02-MG.RF1-1	97	0,661	0,905	S	1	PredHel=1	0	Y
EUGR-PU-003-032-H09-MG.RF2-1	45	0,613	0,678	M	3	PredHel=0	0	Y
EUGR-PU-003-033-C07-MG.RF1-1	121	0,58	0,547	S	4	PredHel=0	0	0
EUGR-PU-003-034-A12-MG.RF3-1	53	0,508	0,959	S	5	PredHel=0	0	Y
EUGR-PU-003-034-E04-MG.RF2-3	41	0,57	0,943	S	2	PredHel=0	0	Y
EUGR-PU-003-035-D08-MG.RF1-1	47	0,46	0,17	S	5	PredHel=0	0	0
EUGR-PU-003-037-C02-MG.RF1-1	145	0,875	1	S	1	PredHel=1	0	Y
EUGR-PU-003-037-C02-MG.RF2-1	120	0,512	0,676	S	1	PredHel=1	1	0
EUGR-PU-003-037-G05-MG.RF1-4	47	0,552	0,966	S	4	PredHel=0	0	Y
EUGR-PU-003-038-E02-MG.RF2-3	51	0,529	0,98	S	2	PredHel=1	0	Y
EUGR-PU-003-039-H07-MG.RF1-1	111	0,447	0,796	S	2	PredHel=1	1	0
EUGR-PU-003-039-H07-MG.RF3-1	136	0,85	1	S	1	PredHel=1	0	Y
EUGR-PU-003-040-A03-MG.RF3-1	258	0,878	0,997	S	1	PredHel=1	0	Y
EUGR-PU-003-040-B05-MG.RF3-1	130	0,877	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-041-H08-MG.RF1-1	120	0,874	1	S	1	PredHel=1	0	Y
EUGR-PU-003-042-D10-MG.RF2-10	41	0,583	0,907	M	5	PredHel=0	0	Y
EUGR-PU-003-042-E08-MG.RF3-1	48	0,531	0,93	S	3	PredHel=0	0	Y
EUGR-PU-003-042-E09-MG.RF3-1	69	0,862	0,998	S	1	PredHel=0	0	Y
EUGR-PU-003-044-C05-MG.RF2-11	44	0,62	0,682	S	2	PredHel=1	0	0
EUGR-PU-003-045-C12-MG.RF3-1	131	0,914	0,999	S	1	PredHel=1	0	Y
EUGR-PU-003-046-G11-MG.RF1-1	163	0,86	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-047-F05-MG.RF3-1	45	0,479	0,564	S	5	PredHel=1	1	0
EUGR-PU-003-047-G09-MG.RF3-1	54	0,565	0,663	S	2	PredHel=0	0	Y
EUGR-PU-003-047-H04-MG.RF1-1	90	0,71	0,979	S	5	PredHel=1	0	Y
EUGR-PU-003-050-D12-MG.RF2-1	66	0,581	0,982	S	4	PredHel=0	1	Y
EUGR-PU-003-052-A07-MG.RF2-2	248	0,504	0,859	M	3	PredHel=0	0	Y
EUGR-PU-003-052-C11-MG.RF2-1	60	0,51	0,972	S	3	PredHel=0	0	Y
EUGR-PU-003-052-E06-MG.RF2-1	56	0,472	0,928	_	5	PredHel=0	0	Y
EUGR-PU-003-053-H06-MG.RF1-4	68	0,476	0,586	M	1	PredHel=0	0	0
EUGR-PU-003-053-H12-MG.RF1-5	82	0,606	0,977	S	1	PredHel=0	0	Y
EUGR-PU-003-054-A03-MG.RF2-1	180	0,73	0,978	S	1	PredHel=0	0	Y
EUGR-PU-003-049-A05-MG.RF1-1	80	0,8	0,992	S	1	PredHel=2	1	Y
EUGR-PU-001-001-D01-RS.RF1-1	137	0,455	0,889	_	4	PredHel=0	0	Y
EUGR-PU-001-001-D01-RS.RF2-2	41	0,734	0,922	S	1	PredHel=1	1	0
EGPP-CD-01-001-M13F_A10F2-2	52	0,706	0,996	S	1	PredHel=1	0	Y
EGPP-CD-01-001-M13F_A10F2-2	56	0,719	0,462	S	1	PredHel=0	0	0
EGPP-CD-01-004-M13F_D02F2-1	59	0,442	0,983	S	2	PredHel=0	0	Y
EGPP-CD-01-008-M13F_F10F1-1	40	0,64	0,902	S	1	PredHel=1	1	0
EUGR-PU-003-077-G10-GO.FF2-3	99	0,613	0,701	S	3	PredHel=1	1	0
EUGR-PU-000-011-A05-MG.RF3-1	138	0,878	0,978	S	1	PredHel=0	0	Y
EUGR-PU-000-017-E04-MG.RF1-1	67	0,577	0,988	_	5	PredHel=0	0	Y
EUGR-PU-000-020-E02-MG.RF1-1	67	0,496	0,855	S	1	PredHel=1	1	0
EUGR-PU-003-001-A07-MG.RF1-1	178	0,744	0,925	S	1	PredHel=3	1	Y
EUGR-PU-003-041-E10-MG.RF1-1	175	0,877	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-051-C08-MG.RF1-1	52	0,586	0,996	S	1	PredHel=1	0	Y
EUGR-PU-003-057-A07-GO.FF3-2	41	0,538	0,743	S	1	PredHel=1	0	Y
EUGR-PU-003-063-G09-GO.FF3-3	53	0,568	0,983	S	2	PredHel=1	0	Y
EUGR-PU-003-064-G01-GO.FF2-2	67	0,49	0,835	S	4	PredHel=2	2	0
EUGR-PU-003-068-C10-GO.FF3-1	42	0,581	0,974	S	1	PredHel=1	1	0
EUGR-PU-003-072-B09-GO.FF1-2	79	0,596	0,546	S	2	PredHel=1	1	0

EUGR-PU-003-073-B01-GO.FF2-1	61	0,44	0,978	M	5	PredHel=0	0	Y
EUGR-PU-003-074-A05-GO.FF1-1	86	0,815	0,992	S	1	PredHel=0	0	Y
EUGR-PU-003-073-B01-GO.GF2-1	61	0,44	0,978	M	5	PredHel=0	0	Y
EUGR-PU-003-056-E02-GO.FF2-2	134	0,619	0,896	S	1	PredHel=1	0	Y
EUGR-PU-003-075-A04-GO.FF1-1	53	0,808	0,996	S	2	PredHel=0	0	Y
EUGR-PU-000-001-C10-MG.RF3-2	67	0,818	1	S	2	PredHel=1	0	Y
EUGR-PU-000-003-D02-MG.RF1-2	49	0,667	0,975	S	1	PredHel=0	0	0
EUGR-PU-000-003-F09-MG.RF1-1	71	0,52	0,821	S	1	PredHel=2	1	Y
EUGR-PU-000-003-F09-MG.RF3-1	63	0,591	0,08	M	5	PredHel=1	1	Y
EUGR-PU-000-009-A10-MG.RF3-1	155	0,878	0,995	S	1	PredHel=1	0	Y
EUGR-PU-000-015-D12-MG.RF1-1	71	0,607	0,516	S	1	PredHel=1	2	0
EUGR-PU-000-021-A08-MG.RF1-1	226	0,805	1	S	1	PredHel=1	0	Y
EUGR-PU-000-021-B06-MG.RF3-2	49	0,667	0,975	S	1	PredHel=0	0	0
EUGR-PU-000-022-F05-MG.RF1-1	70	0,469	0,898	_	5	PredHel=0	0	0
EUGR-PU-003-001-E04-MG.RF1-1	252	0,774	0,904	S	2	PredHel=1	1	0
EUGR-PU-003-009-H10-MG.RF3-1	99	0,9	0,996	S	1	PredHel=1	0	Y
EUGR-PU-003-005-E11-MG.RF1-6	41	0,547	0,965	S	2	PredHel=0	0	Y
EUGR-PU-003-012-D02-MG.RF2-1	186	0,921	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-045-H07-MG.RF1-1	49	0,709	0,989	S	1	PredHel=0	0	Y
EUGR-PU-003-054-D02-MG.RF3-1	139	0,892	0,999	S	2	PredHel=0	0	Y
EUGR-PU-003-049-C09-MG.RF1-3	40	0,71	0,99	S	2	PredHel=1	0	Y
EUGR-PU-003-015-F02-MG.RF1-1	159	0,653	0,967	S	2	PredHel=0	0	Y
EUGR-PU-003-021-G12-MG.RF1-1	164	0,659	0,986	S	1	PredHel=0	0	Y
EUGR-PU-003-029-G04-MG.RF1-1	126	0,886	0,999	S	1	PredHel=0	0	Y
EUGR-PU-003-031-G09-MG.RF1-1	148	0,706	0,989	S	1	PredHel=1	0	Y
EUGR-PU-003-037-E10-MG.RF1-1	80	0,74	0,967	S	1	PredHel=0	1	Y

1. D-score – Media da estimativa do ponto de clivagem do peptídeo sinal e da media dos valores de probabilidade da região atribuída ao peptídeo sinal, discrimina entre a região do peptídeo sinal do restante da proteína;
2. Sprob – Valor de probabilidade do algoritmo de predição HMM, que distingue entre peptídeo sinal (S), sinal de ancoramento (A) e outro (Q);
3. Loc – Predição da localização celular, (M) mitocôndria, (S) peptídeo sinal e – outro;
4. RC – “Coeficiente de confiabilidade”, dado pela razão entre o maior e o segundo maior valores de predição de localização. 1 – 1-0,8; 2 – 0,8-0,6; sucessivamente até o valor 5;
5. PredHel – Numero de α -hélices transmembrana preditas;
6. TM - Numero de α -hélices transmembrana preditas;
7. SP – Predição de peptídeo sinal;

Tabela 2 – Clones de cDNA que codificam proteínas secretadas, selecionados de acordo com os critérios descritos no texto, e respectiva codificação simplificada.

Clone de Referência	Nº de clones com seqüência idêntica presentes na biblioteca EUGR-PU¹	Código
EUGR-PU-003-007-E04-MG.R	3	02Pp_sp
EUGR-PU-000-008-D01-MG.R	2	05Pp_sp
EUGR-PU-000-009-A10-MG.R	2	06Pp_sp
EUGR-PU-000-009-E12-MG.R	2	07Pp_sp
EUGR-PU-000-020-F11-MG.R	3	09Pp_sp
EUGR-PU-003-007-E09-MG.R	1	13Pp_sp
EUGR-PU-003-009-H10-MG.R	1	15Pp_sp
EUGR-PU-003-012-F05-MG.R	2	19Pp_sp
EUGR-PU-003-013-C11-MG.R	1	20Pp_sp
EUGR-PU-003-015-F02-MG.R	1	22Pp_sp
EUGR-PU-000-004-F08-MG.R	2	24Pp_sp
EUGR-PU-000-011-A05-MG.R	1	25Pp_sp
EUGR-PU-003-021-D07-MG.R	4	27Pp_sp
EUGR-PU-003-021-G12-MG.R	1	28Pp_sp
EUGR-PU-000-011-B08-MG.R	1	32Pp_sp
EUGR-PU-003-064-D07-GO.F	1	41Pp_sp
EUGR-PU-003-065-H09-GO.F	2	42Pp_sp
EUGR-PU-003-067-D09-GO.F	2	43Pp_sp
EUGR-PU-003-018-A02-MG.R	3	45Pp_sp
EUGR-PU-003-029-G04-MG.R	2	46Pp_sp
EUGR-PU-003-030-B01-MG.R	2	47Pp_sp
EUGR-PU-003-040-B05-MG.R	1	49Pp_sp
EUGR-PU-003-041-E05-MG.R	3	50Pp_sp
EUGR-PU-003-054-A03-MG.R	1	51Pp_sp
EUGR-PU-003-054-D02-MG.R	1	52Pp_sp
EUGR-PU-003-066-A10-GO.F	3	53Pp_sp
EUGR-PU-003-062-B08-GO.F	1	56Pp_sp
EUGR-PU-003-031-G09-MG.R	1	57Pp_sp
EUGR-PU-003-037-E10-MG.R	1	58Pp_sp
EUGR-PU-003-039-H07-MG.R	3	59Pp_sp
EUGR-PU-003-045-C12-MG.R	2	60Pp_sp

1 – Sequências que apresentaram E-value zero na análise com o algoritmo BlastN.

Tabela 3 – Composição de aminoácidos dos clones selecionados, obtida com o programa ProtParam.

Clones	AA																					
	Ala (A)		Arg (R)		Asn (N)		Asp (D)		Cys (C)		Gln (Q)		Glu (E)		Gly (G)		His (H)		Ile (I)		Leu (L)	
02Pp_sp	11	7.6%	5	3.4%	9	6.2%	14	9.7%	1	0.7%	7	4.8%	6	4.1%	30	20.7%	8	5.5%	1	0.7%	8	5.5%
05Pp_sp	9	9.1%	4	4.0%	6	6.1%	12	12.1%	0	0.0%	5	5.1%	3	3.0%	12	12.1%	9	9.1%	0	0.0%	7	7.1%
06Pp_sp	14	5.7%	5	2.0%	24	9.8%	12	4.9%	14	5.7%	11	4.5%	6	2.4%	9	3.7%	13	5.3%	15	6.1%	18	7.3%
07Pp_sp	14	10.9%	5	3.9%	5	3.9%	11	8.6%	1	0.8%	6	4.7%	5	3.9%	24	18.8%	3	2.3%	3	2.3%	5	3.9%
09Pp_sp	16	9.9%	4	2.5%	10	6.2%	9	5.6%	3	1.9%	5	3.1%	11	6.8%	11	6.8%	6	3.7%	3	1.9%	10	6.2%
13Pp_sp	5	4.3%	4	3.4%	8	6.9%	7	6.0%	9	7.8%	4	3.4%	4	3.4%	9	7.8%	2	1.7%	0	0.0%	14	12.1%
15Pp_sp	7	7.1%	3	3.0%	4	4.0%	4	4.0%	7	7.1%	3	3.0%	5	5.1%	8	8.1%	4	4.0%	2	2.0%	10	10.1%
19Pp_sp	9	5.4%	3	1.8%	14	8.3%	8	4.8%	13	7.7%	11	6.5%	6	3.6%	16	9.5%	3	1.8%	4	2.4%	15	8.9%
20Pp_sp	6	4.1%	6	4.1%	8	5.5%	7	4.8%	9	6.2%	5	3.4%	4	2.7%	12	8.2%	2	1.4%	5	3.4%	11	7.5%
22Pp_sp	15	9.4%	3	1.9%	16	10.1%	9	5.7%	9	5.7%	6	3.8%	4	2.5%	10	6.3%	8	5.0%	5	3.1%	8	5.0%
24Pp_sp	11	5.4%	2	1.0%	6	2.9%	16	7.8%	2	1.0%	15	7.4%	5	2.5%	16	7.8%	3	1.5%	5	2.5%	15	7.4%
25Pp_sp	7	5.1%	6	4.3%	10	7.2%	8	5.8%	8	5.8%	12	8.7%	2	1.4%	10	7.2%	3	2.2%	5	3.6%	10	7.2%
27Pp_sp	12	7.1%	3	1.8%	12	7.1%	24	14.1%	0	0.0%	6	3.5%	4	2.4%	19	11.2%	11	6.5%	1	0.6%	8	4.7%
28Pp_sp	9	5.5%	6	3.7%	14	8.5%	6	3.7%	10	6.1%	6	3.7%	5	3.0%	7	4.3%	3	1.8%	12	7.3%	17	10.4%
32Pp_sp	7	6.9%	10	9.8%	1	1.0%	3	2.9%	1	1.0%	5	4.9%	2	2.0%	8	7.8%	5	4.9%	4	3.9%	7	6.9%
41Pp_sp	21	25.3%	1	1.2%	1	1.2%	0	0.0%	0	0.0%	0	0.0%	1	1.2%	11	13.3%	1	1.2%	3	3.6%	3	3.6%
42Pp_sp	22	12.2%	1	0.6%	8	4.4%	6	3.3%	3	1.7%	15	8.3%	8	4.4%	5	2.8%	1	0.6%	12	6.7%	24	13.3%
43Pp_sp	10	6.0%	2	1.2%	15	8.9%	8	4.8%	13	7.7%	9	5.4%	7	4.2%	16	9.5%	5	3.0%	5	3.0%	14	8.3%
45Pp_sp	9	4.9%	2	1.1%	3	1.6%	5	2.7%	2	1.1%	10	5.5%	15	8.2%	7	3.8%	9	4.9%	15	8.2%	14	7.7%
46Pp_sp	14	11.1%	5	4.0%	7	5.6%	10	7.9%	1	0.8%	6	4.8%	5	4.0%	23	18.3%	3	2.4%	2	1.6%	5	4.0%
47Pp_sp	11	6.5%	2	1.2%	13	7.7%	8	4.8%	13	7.7%	9	5.4%	7	4.2%	16	9.5%	6	3.6%	3	1.8%	16	9.5%
49Pp_sp	9	6.9%	7	5.4%	4	3.1%	8	6.2%	6	4.6%	4	3.1%	5	3.8%	8	6.2%	4	3.1%	10	7.7%	8	6.2%
50Pp_sp	9	4.9%	2	1.1%	3	1.6%	5	2.7%	2	1.1%	10	5.5%	14	7.7%	8	4.4%	8	4.4%	13	7.1%	14	7.7%
51Pp_sp	13	7.2%	3	1.7%	9	5.0%	6	3.3%	1	0.6%	14	7.8%	10	5.6%	7	3.9%	3	1.7%	17	9.4%	26	14.4%
52Pp_sp	10	7.2%	4	2.9%	7	5.0%	5	3.6%	7	5.0%	7	5.0%	3	2.2%	11	7.9%	1	0.7%	6	4.3%	14	10.1%
53Pp_sp	11	9.2%	6	5.0%	5	4.2%	6	5.0%	0	0.0%	7	5.8%	4	3.3%	3	2.5%	3	2.5%	4	3.3%	13	10.8%
56Pp_sp	6	5.8%	3	2.9%	4	3.8%	1	1.0%	4	3.8%	1	1.0%	1	1.0%	20	19.2%	0	0.0%	2	1.9%	9	8.7%

57Pp_sp	6	4.1%	8	5.4%	10	6.8%	7	4.7%	8	5.4%	10	6.8%	3	2.0%	13	8.8%	3	2.0%	9	6.1%	11	7.4%
58Pp_sp	5	6.2%	2	2.5%	2	2.5%	1	1.2%	0	0.0%	2	2.5%	3	3.8%	0	0.0%	4	5.0%	4	5.0%	20	25.0%
59Pp_sp	12	8.8%	4	2.9%	10	7.4%	12	8.8%	1	0.7%	6	4.4%	6	4.4%	25	18.4%	5	3.7%	2	1.5%	9	6.6%
60Pp_sp	15	11.5%	5	3.8%	5	3.8%	11	8.4%	1	0.8%	8	6.1%	3	2.3%	22	16.8%	6	4.6%	4	3.1%	6	4.6%

Tabela 3 – Composição de aminoácidos dos clones selecionados, obtida com o programa ProtParam.

Clones	AA																						aa carregado	
	Lys (K)	Met (M)	Phe (F)	Pro (P)	Ser (S)	Thr (T)	Trp (W)	Tyr (Y)	Val (V)	Pyl (O)	Sec (U)	- (Asp + Glu)	+ (Arg + Lys)											
02Pp_sp	7	4.8%	3	2.1%	12	8.3%	4	2.8%	8	5.5%	3	2.1%	1	0.7%	4	2.8%	3	2.1%	0	0.0%	0	0.0%	20	12
05Pp_sp	6	6.1%	1	1.0%	4	4.0%	2	2.0%	7	7.1%	2	2.0%	2	2.0%	2	2.0%	6	6.1%	0	0.0%	0	0.0%	15	10
06Pp_sp	20	8.2%	4	1.6%	7	2.9%	14	5.7%	25	10.2%	9	3.7%	5	2.0%	10	4.1%	10	4.1%	0	0.0%	0	0.0%	18	25
07Pp_sp	5	3.9%	1	0.8%	12	9.4%	3	2.3%	17	13.3%	0	0.0%	0	0.0%	2	1.6%	6	4.7%	0	0.0%	0	0.0%	16	10
09Pp_sp	7	4.3%	5	3.1%	8	4.9%	15	9.3%	13	8.0%	17	10.5%	3	1.9%	1	0.6%	5	3.1%	0	0.0%	0	0.0%	20	11
13Pp_sp	6	5.2%	1	0.9%	5	4.3%	9	7.8%	7	6.0%	6	5.2%	1	0.9%	6	5.2%	9	7.8%	0	0.0%	0	0.0%	11	10
15Pp_sp	3	3.0%	3	3.0%	3	3.0%	11	11.1%	4	4.0%	8	8.1%	2	2.0%	2	2.0%	6	6.1%	0	0.0%	0	0.0%	9	6
19Pp_sp	12	7.1%	2	1.2%	5	3.0%	10	6.0%	10	6.0%	7	4.2%	0	0.0%	7	4.2%	13	7.7%	0	0.0%	0	0.0%	14	15
20Pp_sp	14	9.6%	2	1.4%	11	7.5%	3	2.1%	7	4.8%	12	8.2%	2	1.4%	10	6.8%	10	6.8%	0	0.0%	0	0.0%	11	20
22Pp_sp	9	5.7%	1	0.6%	6	3.8%	9	5.7%	14	8.8%	8	5.0%	1	0.6%	7	4.4%	11	6.9%	0	0.0%	0	0.0%	13	12
24Pp_sp	9	4.4%	1	0.5%	21	10.3%	5	2.5%	38	18.6%	10	4.9%	2	1.0%	5	2.5%	17	8.3%	0	0.0%	0	0.0%	21	11
25Pp_sp	9	6.5%	2	1.4%	4	2.9%	3	2.2%	13	9.4%	10	7.2%	2	1.4%	5	3.6%	9	6.5%	0	0.0%	0	0.0%	10	15
27Pp_sp	11	6.5%	2	1.2%	10	5.9%	5	2.9%	18	10.6%	5	2.9%	1	0.6%	10	5.9%	8	4.7%	0	0.0%	0	0.0%	28	14
28Pp_sp	10	6.1%	1	0.6%	7	4.3%	13	7.9%	16	9.8%	6	3.7%	0	0.0%	7	4.3%	9	5.5%	0	0.0%	0	0.0%	11	16
32Pp_sp	6	5.9%	3	2.9%	5	4.9%	10	9.8%	11	10.8%	4	3.9%	2	2.0%	2	2.0%	6	5.9%	0	0.0%	0	0.0%	5	16
41Pp_sp	2	2.4%	1	1.2%	1	1.2%	7	8.4%	0	0.0%	12	14.5%	2	2.4%	13	15.7%	3	3.6%	0	0.0%	0	0.0%	1	3
42Pp_sp	15	8.3%	3	1.7%	11	6.1%	3	1.7%	15	8.3%	13	7.2%	1	0.6%	3	1.7%	11	6.1%	0	0.0%	0	0.0%	14	16
43Pp_sp	11	6.5%	1	0.6%	6	3.6%	10	6.0%	10	6.0%	7	4.2%	0	0.0%	7	4.2%	12	7.1%	0	0.0%	0	0.0%	15	13
45Pp_sp	18	9.8%	1	0.5%	11	6.0%	6	3.3%	14	7.7%	13	7.1%	1	0.5%	7	3.8%	21	11.5%	0	0.0%	0	0.0%	20	20
46Pp_sp	5	4.0%	1	0.8%	12	9.5%	3	2.4%	15	11.9%	0	0.0%	0	0.0%	3	2.4%	6	4.8%	0	0.0%	0	0.0%	15	10
47Pp_sp	14	8.3%	1	0.6%	6	3.6%	8	4.8%	9	5.4%	8	4.8%	0	0.0%	7	4.2%	11	6.5%	0	0.0%	0	0.0%	15	16
49Pp_sp	7	5.4%	1	0.8%	7	5.4%	7	5.4%	11	8.5%	10	7.7%	0	0.0%	9	6.9%	5	3.8%	0	0.0%	0	0.0%	13	14
50Pp_sp	17	9.3%	1	0.5%	11	6.0%	6	3.3%	15	8.2%	14	7.7%	1	0.5%	8	4.4%	22	12.0%	0	0.0%	0	0.0%	19	19
51Pp_sp	12	6.7%	4	2.2%	13	7.2%	7	3.9%	11	6.1%	11	6.1%	2	1.1%	1	0.6%	10	5.6%	0	0.0%	0	0.0%	16	15
52Pp_sp	8	5.8%	2	1.4%	7	5.0%	9	6.5%	12	8.6%	14	10.1%	1	0.7%	1	0.7%	10	7.2%	0	0.0%	0	0.0%	8	12
53Pp_sp	12	10.0%	1	0.8%	4	3.3%	8	6.7%	14	11.7%	9	7.5%	3	2.5%	2	1.7%	5	4.2%	0	0.0%	0	0.0%	10	18

56Pp_sp	4	3.8%	2	1.9%	7	6.7%	11	10.6%	15	14.4%	4	3.8%	0	0.0%	3	2.9%	7	6.7%	0	0.0%	0	0.0%	2	7
57Pp_sp	5	3.4%	2	1.4%	10	6.8%	6	4.1%	4	2.7%	13	8.8%	1	0.7%	9	6.1%	10	6.8%	0	0.0%	0	0.0%	10	13
58Pp_sp	6	7.5%	3	3.8%	3	3.8%	4	5.0%	4	5.0%	7	8.8%	3	3.8%	2	2.5%	5	6.2%	0	0.0%	0	0.0%	4	8
59Pp_sp	9	6.6%	2	1.5%	10	7.4%	4	2.9%	11	8.1%	1	0.7%	1	0.7%	3	2.2%	3	2.2%	0	0.0%	0	0.0%	18	13
60Pp_sp	4	3.1%	1	0.8%	9	6.9%	7	5.3%	10	7.6%	2	1.5%	0	0.0%	6	4.6%	6	4.6%	0	0.0%	0	0.0%	14	9