

EVOLUÇÃO MOLECULAR DARWINIANA NAS
PROTEÍNAS RELACIONADAS À PATOGÊNESE
(PRs) EM PLANTAS

Nicole de Miranda Scherer

Orientador: Francisco M. Salzano

Co-orientador: Sandro L. Bonatto

Dissertação apresentada ao Programa de Pós
Graduação em Genética e Biologia
Molecular da Universidade Federal do Rio
Grande do Sul para a obtenção do Grau de
Mestre em Genética e Biologia Molecular.

Porto Alegre
2002

*Dedicado aos meus mestres,
Francisco Salzano e Loreta Freitas.*

Agradecimentos

A realização deste projeto não teria sido possível sem a ajuda da minha inigualável equipe de orientadores: O Dr. Sandro Bonatto, para dar asas a nossa imaginação, com suporte técnico, obviamente; a Dra. Loreta Freitas, para trazer de volta ao mundo real, com bons “puxões-de-orelha”, como só as mães sabem fazer; e o Sábio Dr. Francisco Salzano, que com seus toques de anjo, transforma essa fantasia em realidade.

OBRIGADA!!!!

Agradeço a todos que viveram ao meu lado todas as emoções deste projeto, que vibraram e se angustiaram junto comigo. Agradeço a minha família pelo apoio incondicional, principalmente ao meu pai, pelo apoio “logístico” e pelo entusiasmo com a biologia evolutiva e a bioinformática. Agradeço a minha mãe por sua determinação, e a minha maninha que se sentiu injustiçada.....

Aline, Valéria, Tati, Daisy e Nelson, obrigada pelo ombro amigo, quando mais precisei! Agradeço muito aos amigos do Lab 123, que quase não têm me visto, mas nunca esquecem de me chamar para as festas. Danke schön!

Obrigada professora Ana Lúcia Cetertich Bazzan, da Informática, pela amizade e suporte técnico. Agradecimentos mais do que técnicos também ao Ronaldo, o domador da onipotência.

Agradeço também as minhas várias outras famílias, de longe e de perto, pela compreensão da minha ausência. Muito obrigada a todos da “casa da sogra”, por todo o carinho dessa família encantadora.

Acima de tudo, obrigada Diego, por ser mais que um companheiro, e pela inestimável ajuda na realização dos programas e organização dos dados.

E quem um dia irá dizer, que existe razão nas coisas feitas pelo coração. E quem irá dizer, que não existe razão...

Sumário

I.	Introdução	5
I.1.	Bases Moleculares da Evolução	5
I.2.	Modelos de Substituição de Códon no Estudo da Evolução de Proteínas	5
I.2.1.	<i>Avaliação Global</i>	5
I.2.2.	<i>Modelos de distribuições</i>	9
I.3.	Mecanismos de Defesa em Plantas	18
I.3.1.	<i>Introdução</i>	18
I.3.2.	<i>Genes de Resistência e de Defesa Contra Patógenos</i>	19
I.4.	Proteínas Relacionadas à Patogênese	20
I.4.1.	<i>Informações Gerais</i>	20
I.4.2.	<i>Classificação</i>	22
I.5.	Evolução dos Genes de Resistência e de Defesa em Plantas	28
I.6.	Objetivos da Dissertação	29
II.	Material e Métodos	30
II.1.	Obtenção das Sequências, Bancos de Dados Utilizados	30
II.2.	Métodos de Análise	30
III.	Resultados	32
III.1.	Informações gerais	32
III.2.	Informações por Classe de PRs	35
III.3.	SGI	85
IV.	Discussão	89
V.	Resumo e Conclusões	93
VI.	Summary and Conclusions	95
VII.	Referências Bibliográficas	97

I. Introdução

I.1. Bases Moleculares da Evolução

Uma das causas primárias da evolução é a mutação dos genes. Um gene ou seqüência de DNA originado por substituição nucleotídica, inserção/deleção, recombinação, conversão gênica, ou outras formas, pode propagar-se na população por deriva genética e/ou seleção natural, e, eventualmente, ser fixado em uma espécie. Sendo possível a identificação dos genes envolvidos em um determinado fenótipo e o estudo de suas mudanças evolutivas, poderemos identificar o tipo de modificação molecular que gerou um caráter morfológico ou fisiológico em particular (Nei & Kumar, 2000).

O uso de seqüências de DNA em estudos evolutivos é vantajoso em decorrência do reduzido número de caracteres, apenas quatro tipos de nucleotídeos, com grande quantidade de informação, centenas ou milhares de nucleotídeos, que estão envolvidos na codificação de um gene. Uma vez que as mudanças evolutivas no DNA seguem padrões razoavelmente regulares, é possível o uso de modelos matemáticos para comparar essas seqüências nucleotídicas. Esse tipo de material pode ser muito bem aproveitado com o auxílio da informática, mas é na maioria dos casos inviável de ser analisado manualmente. Por este motivo, muitos pesquisadores passaram a desenvolver programas computacionais baseados nos princípios da evolução molecular e seus algoritmos, para trabalhar com esta enorme quantidade de informação.

I.2. Modelos de Substituição de Códon no Estudo da Evolução de Proteínas

I.2.1. Avaliação Global

Conforme a teoria neutralista da evolução molecular, a maior parte das substituições nucleotídicas e de aminoácidos não deveria gerar conseqüências adaptativas. Neste caso, a probabilidade de fixação de uma mutação no DNA seria a mesma para qualquer sítio nucleotídico. Mas em muitos casos, a substituição de aminoácidos pode provocar fortes alterações na estrutura da proteína e, conseqüentemente, em suas características fisiológicas, com repercussão no valor adaptativo do organismo. Quando uma mutação prejudica o organismo em questão, ela tende a ser eliminada, porém se ela potencializa o seu valor adaptativo, há grandes chances de permanecer e ser propagada na população, caracterizando um tipo de seleção positiva.

Uma mutação na proteína ocorre quando a alteração nos nucleotídeos ocasiona a substituição de um códon por outro não-sinônimo, isto é, quando este códon não codifica

para o mesmo aminoácido do códon preexistente. Uma mutação não-sinônima pode decorrer por uma alteração no número de nucleotídeos (inserção/deleção), quando o módulo de leitura é alterado, ou por substituição nucleotídica, quando ocorre em sítio não sinônimo do códon. Em geral as substituições nucleotídicas são mais facilmente encontradas, já que uma alteração no módulo de leitura pode causar a inativação total da proteína.

Devido às propriedades do código genético, a maioria das substituições sinônimas ocorre na terceira posição dos códons, mas algumas ocorrem na primeira. Todas as substituições na segunda posição resultam em mutações sinônimas ou sem sentido (quando codificam um dos códons de parada). Se todos os códons forem igualmente frequentes no genoma, e a probabilidade de substituição for a mesma para todos os pares de nucleotídeos, as proporções de mutações sinônimas, não sinônimas e sem sentido seriam, respectivamente, cerca de 25%, 71% e 4% (Nei & Kumar, 2000). Na realidade, a frequência dos códons e de substituição de pares de nucleotídeos é variável de gene para gene e de organismo para organismo, mas todas estas características podem ser consideradas quando incluídas nos algoritmos utilizados para as análises.

Os modelos de substituição de códons podem demonstrar como a seleção natural, que é exercida sobre a proteína (fenótipo), atua na estrutura gênica. Daí o porquê de buscar sinais de seleção positiva diretamente nas seqüências de DNA codificadoras.

As mutações sinônimas são consideradas livres de pressão seletiva em nível protéico, mas sabe-se que, em muitos casos, a neutralidade é influenciada pela utilização preferencial de códons.

Como foi indicado acima, as mutações não-sinônimas seriam alvos certos da seleção, por influenciarem a estrutura da proteína, podendo inclusive alterar sua função. Estas mutações são, em sua maioria, deletérias, e são rapidamente eliminadas por seleção negativa. Mas em alguns casos elas podem conferir um aumento na eficiência ou uma nova função à proteína. Nestes casos, elas podem representar uma novidade evolutiva que aumentaria o valor adaptativo de seus portadores, podendo, desta forma, virem a ser fixadas na população. Estes eventos são mais comuns de ocorrer em famílias multigênicas, quando há múltiplas cópias dos genes e a perda de uma função por mutação pode ser suprida por outro membro da família (Freeman & Herron, 1998).

Forças seletivas operando em nível de seqüências de aminoácidos têm sido detectadas por comparação do número de substituições não-sinônimas com o número de

substituições sinônimas por sítio (Hughes & Nei, 1988, 1989; Endo *et al.*, 1996; Tsunoyama & Gojobori, 1998). A pressão de seleção em proteínas pode ser medida pela proporção dessas taxas, indicada pela razão ômega ($\omega = d_N/d_S$) (Goldman & Yang, 1994), onde d_N : taxa de fixação das mutações não-sinônimas, e d_S : taxa de fixação das mutações sinônimas.

Se as trocas de aminoácidos forem vantajosas, mutações não-sinônimas serão fixadas a uma taxa maior do que as mutações sinônimas. Dessa forma obteremos $d_N > d_S \Rightarrow \omega > 1$. Assim, $\omega > 1$ é uma indicação de seleção positiva, ou evolução molecular adaptativa. Se, ao contrário, as trocas de aminoácidos forem totalmente deletérias, o resultado é que d_N seria zero e conseqüentemente $\omega = 0$. Por outro lado, caso as mutações não-sinônimas resultarem no mesmo valor adaptativo que as mutações sinônimas (forem neutras), então elas serão fixadas na mesma proporção, $d_N = d_S$ e $\omega = 1$. Já no caso das mutações não-sinônimas reduzirem o valor adaptativo destas proteínas, $d_N < d_S$, provocando um valor intermediário de ômega ($0 < \omega < 1$).

Inicialmente, este método era aplicado considerando a seqüência como um todo, com ômega sendo calculado sobre a média de todos os códons, o que implicaria na condição de pressão seletiva homogênea sobre todos os códons analisados. Isto leva a uma diminuição na eficiência do método, pois na maioria das proteínas, uma parte significativa dos aminoácidos está sob forte pressão seletiva purificadora para a conservação de sua estrutura e função (Yang *et al.*, 2000a). Assim, se a evolução adaptativa estiver ocorrendo em um número reduzido de códons, o teste será ineficiente na detecção deste tipo de seleção.

Em virtude dessa ineficiência, Nielsen & Yang (1998) desenvolveram um método de máxima verossimilhança baseado em modelos estatísticos, que permite a distinção de razões ω diferentes entre os códons. Tais modelos podem ser utilizados para testar e identificar os aminoácidos críticos em uma proteína sob seleção positiva. Nesta análise, os autores implementaram modelos simples que permitiam ω heterogêneos entre os códons. O primeiro deles, o modelo "neutro", supõe duas classes de códons, uma de códons conservados, na qual as mutações não-sinônimas seriam removidas por serem deletérias, e uma classe completamente neutra, na qual $\omega = 1$. O outro modelo acrescenta uma terceira classe de códons, na qual o ω é estimado a partir dos dados dos códons que não correspondem a $\omega = 0$, nem a $\omega = 1$. Mas em geral, o ω calculado acaba ficando entre

$0 < \omega < 1$ quando, na maior parte desses códons, as mutações não-sinônimas são levemente deletérias e o número de códons adaptativos é muito diluído.

Uma nova análise foi desenvolvida por Yang *et al.* (2000a), na qual foram implementados novos modelos para testar diversas distribuições de probabilidade para ω heterogêneo entre os códons. Essas diferentes distribuições possibilitam a identificação dos códons selecionados positivamente através da discretização da distribuição das probabilidades em diferentes classes de códons.

Paralelamente, Suzuki e Gojobori (1999) desenvolveram outro método para detectar a seleção positiva em sítios de aminoácidos, que também é baseado na proporção de substituições não-sinônimas e sinônimas. Porém, o parâmetro analisado não é ômega. Neste método os códons ancestrais são inferidos para cada nó da árvore filogenética, pelo método de máxima parcimônia. Os números de sítios sinônimos e não-sinônimos são estimados para cada códon ao longo da árvore filogenética e o número total de substituições sinônimas e não-sinônimas é obtido pela soma destes. O teste de neutralidade considera que, se não há força seletiva atuando, a razão do número de substituições sinônimas observadas sobre o total de substituições será equivalente à razão do número de códons sinônimos estimados pelo total de códons. O mesmo ocorrerá para as substituições não-sinônimas.

Recentemente, Suzuki & Nei (2001) realizaram uma simulação computacional para testar a confiabilidade dos dois métodos na detecção de seleção positiva em sítios de aminoácidos. Eles concluíram que a inferência pelo método de parcimônia é mais robusta que o uso de diferentes modelos evolutivos, calculados pelo método baseado na máxima verossimilhança. Porém, isso é significativo apenas quando o número de seqüências comparadas é muito grande, e o comprimento dos ramos da árvore filogenética relativamente pequeno. Assim sendo, em análises que envolvam um número pequeno de seqüências, o método da máxima verossimilhança é confiável e mais indicado, principalmente quando as seqüências são mais distantemente relacionadas, já que ele considera também as substituições múltiplas, não consideradas pelo método da parcimônia.

1.2.2. Modelos de Distribuições

O programa CODEML (Yang, 2000) usa 14 modelos distintos para as distribuições de probabilidade da razão ω entre a taxa de ocorrência de substituições não-sinônimas e sinônimas, isto é,

$$\omega = d_N / d_S$$

onde d_N é a frequência de substituições não-sinônimas e d_S a de sinônimas. A seguir, é apresentada a descrição sucinta dos 14 modelos:

Modelo M0

Este modelo assume um único valor de ω para todos os sítios.

Modelo M1

É o “modelo neutro”, no qual ω assume apenas os valores 0 ou 1. O único parâmetro é a fração (ou frequência) p_0 de sítios onde $\omega=0$.

Modelo M2

Este modelo adiciona uma classe de sítios com frequência $p_2=1-p_0-p_1$, para o qual a razão ω não é predeterminada, mas encontrada *a posteriori*. Tem-se, portanto, neste modelo, três parâmetros: p_0 , p_1 e ω_2 .

Modelo M3

Usa K distintos valores de ω , não pré-fixados, $\omega_0, \omega_1, \dots, \omega_K$, que ocorrem com frequências também não pré-fixadas, $p_0, p_1, \dots, p_K=1-(p_1+p_2+\dots)$, Tem-se, portanto, $2K-1$ parâmetros. Nos cálculos foi usado $K=3$, ou seja, cinco parâmetros.

Modelo M4

Este modelo fixa, *a priori*, K distintos valores para ω e deixa as $K-1$ frequências como parâmetros para determinação *a posteriori*. Para os cálculos do presente estudo foi usado $K=5$ e os seguintes valores para ω : 0, 1/3, 2/3, 1, 3.

Modelo M5

A partir do M5, todos os modelos admitem uma distribuição contínua para ω . O modelo M5, com distribuição chamada gama, usa a função gama da matemática (continuação analítica de fatorial) como fator de normalização. Sua densidade de probabilidade contém dois parâmetros, α e β , e é definida por:

$$f(\omega) = \beta^\alpha \exp(-\beta \omega) \omega^{\alpha-1} / \Gamma(\alpha).$$

Para melhor visualizar a distribuição gama e seu comportamento frente à variação dos parâmetros, mostra-se seu gráfico na Figura I.1. A curva em azul representa $\text{gama}(\omega; \alpha=1, \beta=1)$; a curva em verde representa $\text{gama}(\omega; \alpha=0,1; \beta=1)$; a curva em vermelho representa $\text{gama}(\omega; \alpha=2, \beta=2)$.

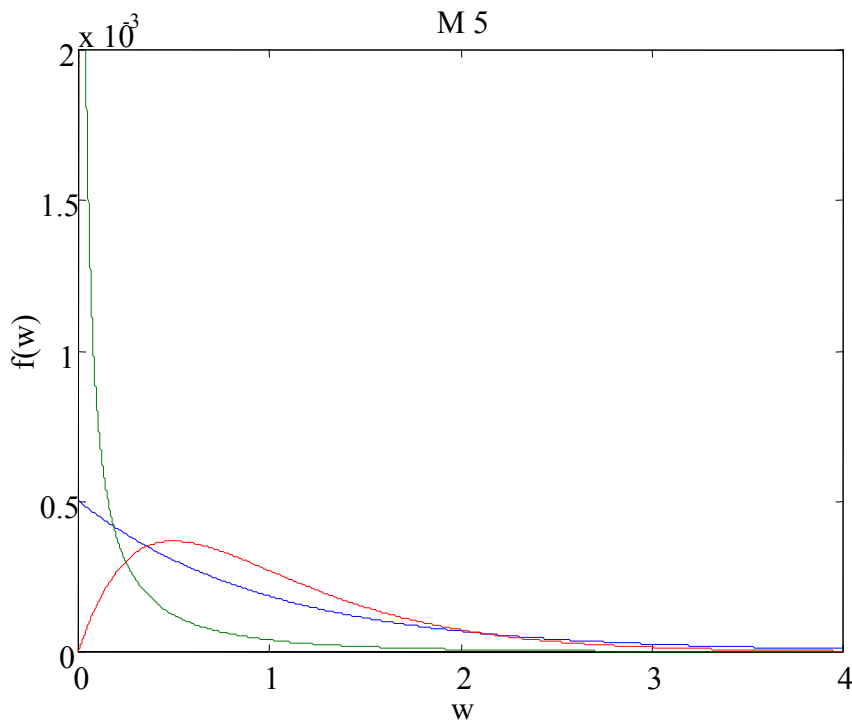


Figura I.1. Diversas distribuições apresentadas por ω em função de $f(\omega)$ no modelo 5.

Vemos que para pequenos valores de α a distribuição se concentra em pequenos valores de ω , deslocando-se para a direita à medida que aumenta α . Ela não é tão sensível ao valor de β , mas desloca-se para a esquerda ao aumentar β .

Modelo M6

É a combinação de duas distribuições gama. Serve para dar conta de uma concentração nas proximidades de $\omega=0$ e uma distribuição centrada em algum valor mais alto. A Figura I.2 mostra a função

$$f(\omega) = 0,2 \text{ gama}(\omega; 0,1; 1,0) + 0,8 \text{ gama}(\omega; 4,0; 8,0).$$

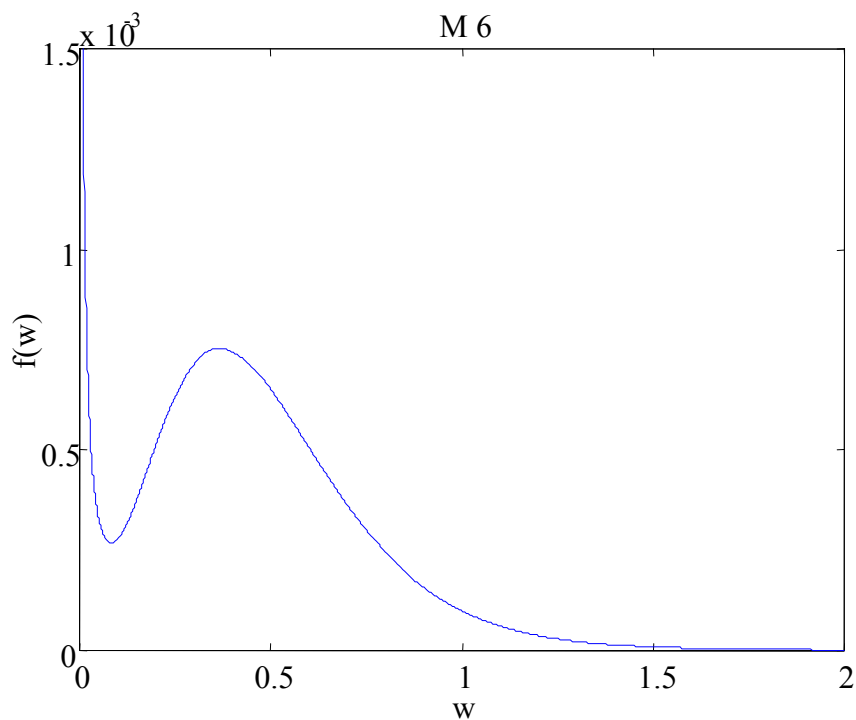


Figura I.2. Distribuição de ω em função de $f(\omega)$ no modelo 6.

Claramente, o pico próximo a $\omega=0$ vem de $\text{gama}(\omega; 0,1; 1,0)$ e a distribuição centrada em $\omega=0,5$ vem de $\text{gama}(\omega; 4,0; 8,0)$.

Modelo M7

Utiliza a distribuição beta, possui dois parâmetros, p e q , e é definido por

$$f(\omega) = \omega^{p-1} (1-\omega)^{q-1} / B(p,q),$$

onde $B(p,q)$ é a função beta da matemática. Esta função só pode ser usada como distribuição de probabilidade para ω no intervalo $0 < \omega < 1$, pois fora deste intervalo ela adquire valores negativos.

Na Figura I.3 são mostrados três gráficos da distribuição beta: A curva azul, com $p=q=0,2$; a curva verde, com $p=q=5$; e a curva vermelha, com $p=0,2$ e $q=5$. Quando $p=q$ a distribuição é simétrica em torno de $\omega=0,5$, tendo forma de U quando $p=q < 1$ e forma de chapéu quando $p=q > 1$. Quando $p < q$ ela se concentra à esquerda do intervalo (próximo a $\omega=0$) e quando $p > q$ ela se desloca para a direita.

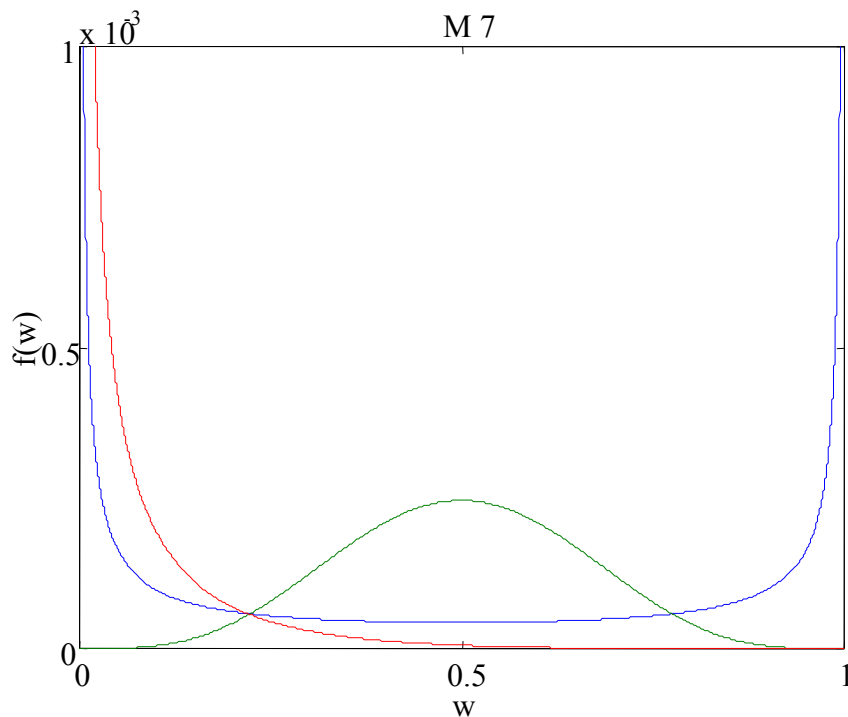


Figura I.3. Diversas distribuições apresentadas por ω em função de $f(\omega)$ no modelo 7.

Modelo M8

A maneira mais simples de corrigir o problema da distribuição beta, de não poder assumir valores fora do intervalo $0 < \omega < 1$, é a combinação de beta com um valor discreto em ω_1 como parâmetro. Este modelo tem, portanto, quatro parâmetros, p , q , ω_1 e a frequência $p_1=1-p_0$ de sítios com $\omega=\omega_1$. A densidade correspondente da probabilidade $f(\omega)$ para a seguinte escolha dos parâmetros: $p=0,005$; $q=2,0$; $\omega_1=5,0$ e $p_1=0,5$ é mostrada na Figura I.4

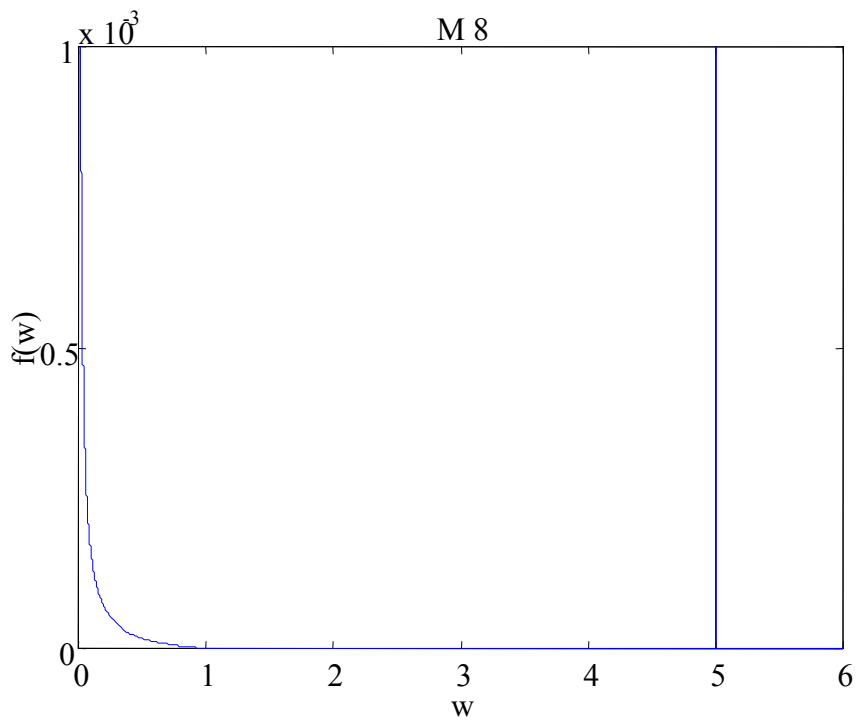


Figura I.4. Distribuição de ω em função de $f(\omega)$ no modelo 8.

Modelo M9

Este modelo combina as duas distribuições, beta e gama, isto é

$$f(\omega) = p_0 \text{beta}(\omega; p, q) + (1-p_0) \text{gama}(\omega; \alpha, \beta).$$

Possui, portanto, cinco parâmetros, p_0 , p , q , α , β . Com valores convenientes desses parâmetros, a distribuição beta dá conta de uma concentração próximo a $\omega=0$ e gama permite uma distribuição mais larga, em valores mais elevados de ω . Uma simulação para o caso em que $p=0,005$; $q=1,0$; $\alpha=8,0$; $\beta=5,0$; e $p_0=0,5$ está representada na Figura I.5.

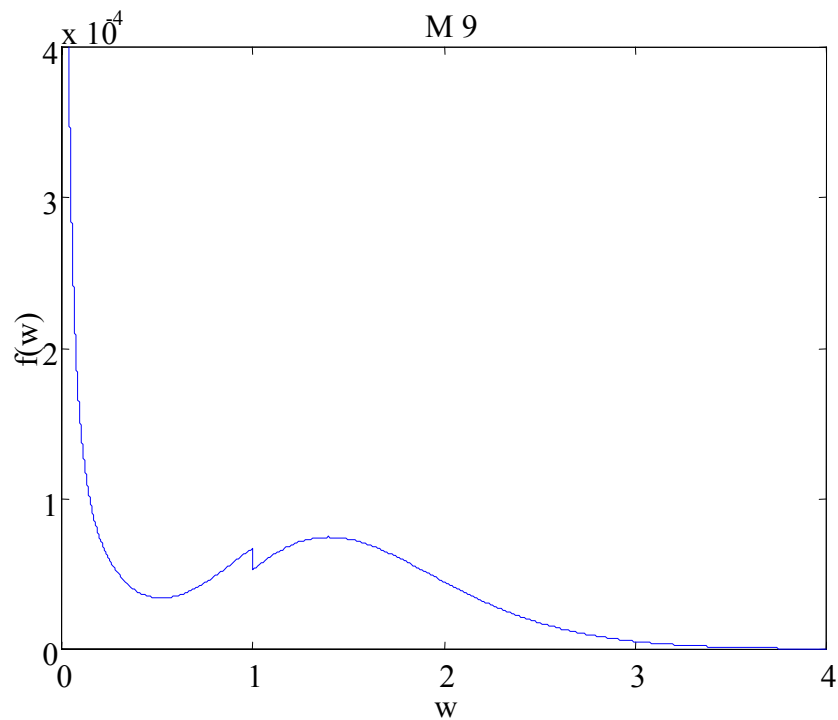


Figura I.5. Distribuição de ω em função de $f(\omega)$ no modelo 9.

Modelo M10

É semelhante ao M9, mas a distribuição gama é deslocada para a direita em uma unidade de ω . A intenção é separar melhor a distribuição em valores mais altos de ω do pico em $\omega=0$. Para os mesmos parâmetros usados em M9, mostra-se a distribuição de M10 na Figura I.6.

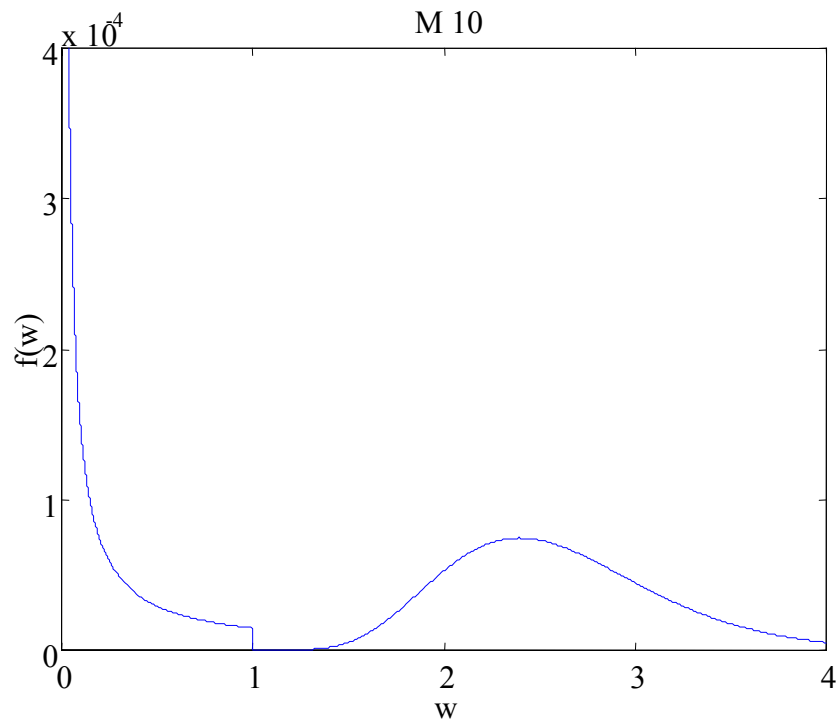


Figura I.6. Distribuição de ω em função de $f(\omega)$ no modelo 10.

Modelo M11

Este modelo também é semelhante ao M9, mas combina beta com uma distribuição normal em vez de gama. Permite caudas mais estendidas, como se vê na Figura I.7, onde foram usados os mesmos parâmetros da Figura I.5 para a distribuição beta, enquanto para a normal usou-se média $\mu=2,0$ e variância $\sigma = 1$.

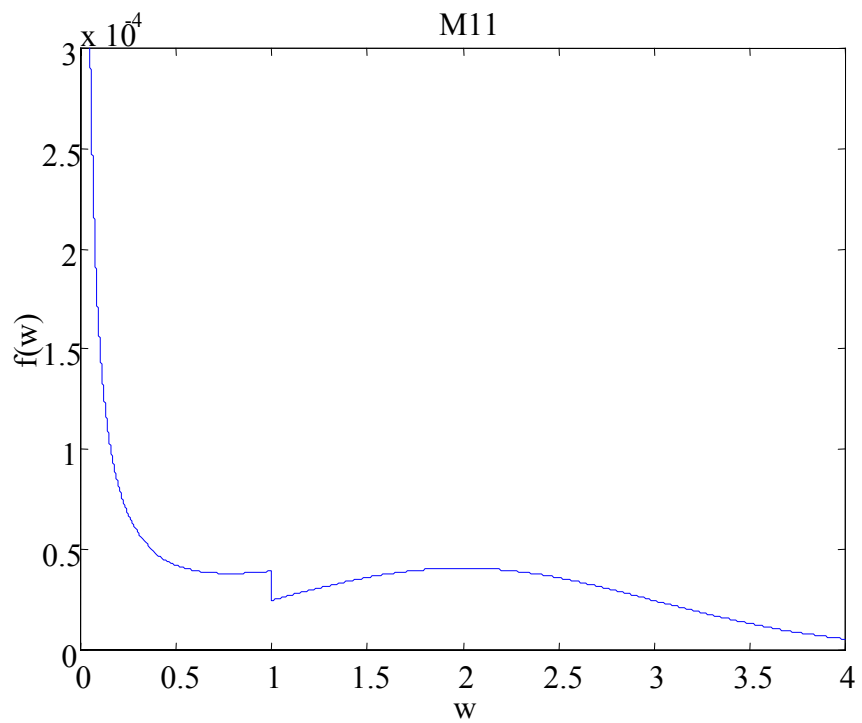


Figura I.7. Distribuição de ω em função de $f(\omega)$ no modelo 11.

Modelo M12

Combina duas distribuições normais com um pico discreto em $\omega_0=0$.

Uma das distribuições é centrada em $\mu_1=1$ e a outra em μ_2 a determinar. Possui, portanto, cinco parâmetros: μ_2 , σ_1 , σ_2 , p_1 e p_2 . Na Figura II.8 mostra-se o caso em que $\mu_2=0,5$; $\sigma_1=0,4$; $\sigma_2=9$; $p_1=0,2$ e $p_2=0,6$.

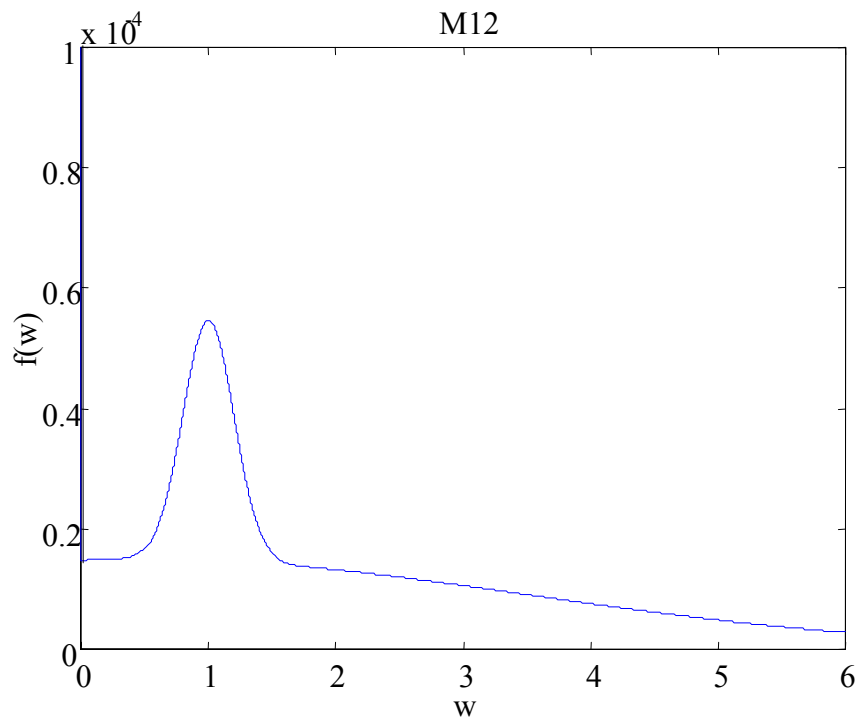


Figura I.8. Distribuição de ω em função de $f(\omega)$ no modelo 12.

Modelo M13

É o modelo que possui o maior número de parâmetros, seis ao todo. É semelhante ao M12, mas em lugar do pico discreto em $\omega=0$, possui outra distribuição normal com centro em $\omega=0$, definida somente para valores não negativos de ω . Uma simulação do modelo é apresentada na Figura I.9.

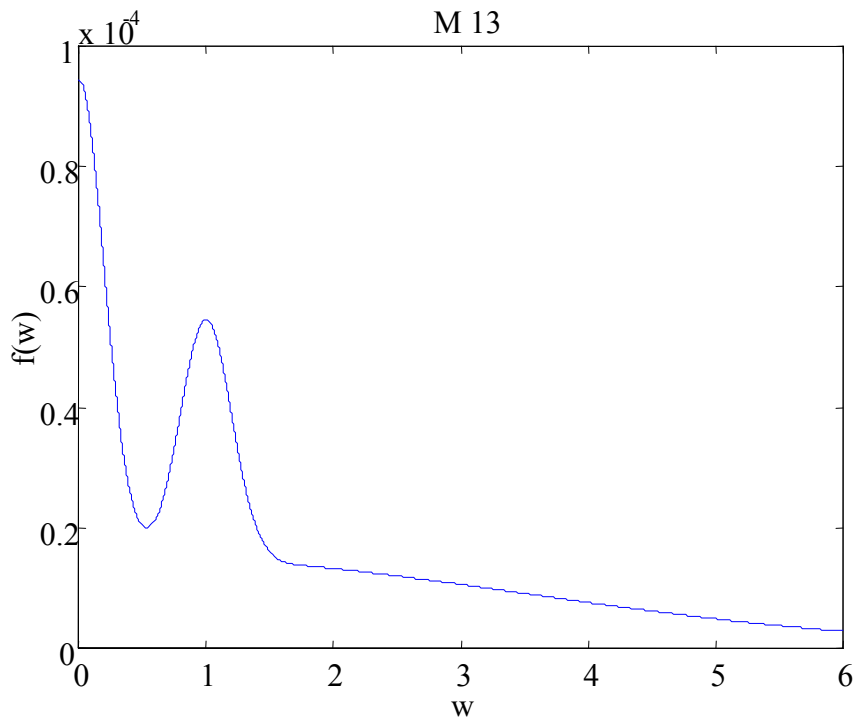


Figura I.9. Distribuição de ω em função de $f(\omega)$ no modelo 13.

As distribuições contínuas neste programa (CODEML) são convertidas em discretas, para a otimização do tempo computacional. Assim são construídas K categorias de classes, compreendendo, cada uma, uma área definida da distribuição proposta. Neste caso, M5 com $K=8$ apresenta oito classes, de frequência 0,125 cada, com o valor médio de ω entre todos os sítios locados em cada classe.

I.3. Mecanismos de Defesa em Plantas

I.3.1. Introdução

Plantas são, em regra geral, organismos fixos, e sua dispersão é usualmente mediada apenas pelo movimento das sementes. Na impossibilidade de fuga de um

ambiente que apresente riscos à sobrevivência, estes organismos necessitam desenvolver mecanismos bioquímicos ou morfológicos de defesa contra ameaças como herbivoria, patogênese, poluentes, fatores edáficos e climáticos, bem como de competição com outras plantas e microrganismos na conquista do ambiente. As defesas desenvolvidas dependem, no caso, principalmente de adaptações moleculares. A expressão de gomas e espinhos para a defesa física, substâncias detoxificadoras, toxinas para afastar herbívoros ou substâncias alelopáticas dependem da existência, no genoma, de genes relacionados a estes mecanismos, e da expressão adequada destes. O desenvolvimento da resistência depende também da variabilidade genética existente na população (Harborne, 1993).

1.3.2. Genes de Resistência e de Defesa Contra Patógenos

Fungos, bactérias e vírus são considerados os principais agentes patogênicos em plantas. Os genes do hospedeiro que estão relacionados à resistência e combate a estes agentes podem ser divididos em dois grandes grupos: os genes de resistência (*R*) e os genes de resposta defensiva (*PR*). Os patógenos, por sua vez, apresentam a contrapartida na forma de genes de avirulência (*avr*) e genes de patogenicidade (*P*) (Dixon & Harrison, 1990).

Genes *R* de plantas são todos aqueles que estão diretamente relacionados à detecção do patógeno por reconhecimento de seus sinais ou elicitores. Tais genes determinam a resistência específica a cada patógeno. Este, por sua vez, pode desenvolver mudanças para escapar dessa resistência através dos genes de avirulência. Um patógeno é considerado não-virulento a uma determinada planta quando é passível de reconhecimento pelo hospedeiro, resultando em uma interação incompatível. Este reconhecimento é devido à expressão de moléculas (elicitores) que são detectadas pelos produtos dos genes *R*. A virulência existe quando nenhuma de suas moléculas é reconhecida pelo hospedeiro impedindo, assim, a indução de outros mecanismos de defesa. Essa interação é considerada compatível (Dixon & Harrison, 1990).

Estes dois grupos de genes, *R* e *avr*, são conhecidos por seu processo de evolução gene-a-gene (Thompson & Burdon, 1992; Bishop *et al.*, 2000). A cada nova mutação que determina detecção de um elicitor pelo hospedeiro, evolui uma nova forma nos genes de avirulência para impedir esta detecção.

Os genes de resposta defensiva do hospedeiro são aqueles que determinam a expressão constitutiva de moléculas protetoras. Exemplos destas respostas são: produção

de compostos fenólicos, que são depositados na parede celular; expressão induzida de enzimas que interferem nas rotas metabólicas de produtos de expressão constitutiva; produção de metabólitos secundários; indução de enzimas hidrolíticas, que atacam diretamente as moléculas dos patógenos e que seriam consideradas fitotoxinas; fitoalexinas, que são sintetizadas e acumuladas nas células do hospedeiro após a exposição a microrganismos, ou constitutivamente expressas, relacionadas a produtos secundários; e a produção induzida ou constitutiva de inibidores de proteinases, que intoxicam os organismos invasores interferindo na função de suas enzimas (Dixon & Harrison, 1990).

A resposta determinada por estes genes não é específica para cada patógeno como os produtos dos genes *R*, e nesta condição a sua forma de evolução não pode ser explicada por um modelo tão simples como a evolução gene-a-gene, pois a interação depende de moléculas de diferentes agentes patogênicos e de outras moléculas do próprio organismo, e devem responder mais, portanto, às mudanças geográficas e temporais na prevalência de diferentes espécies de patógenos (Bishop *et al.*, 2000).

Da mesma forma que os genes de resposta defensiva do hospedeiro, os genes de patogenicidade expressos pelos organismos estão envolvidos com uma gama de interações patógeno-hospedeiro. São estes genes que desenvolvem a toxicidade do patógeno e a eficiência da desativação das defesas acionadas pelas plantas (Ham *et al.*, 1997; Hodge *et al.*, 1996).

I.4. Proteínas Relacionadas à Patogênese

I.4.1. Informações Gerais

As proteínas de defesa de plantas que são induzidas após infecções por agentes patogênicos e/ou estresse abiótico, são conhecidas como proteínas relacionadas à patogênese (Pathogenesis Related Proteins, PRs) (para revisão, ver Van Loon *et al.*, 1994; Van Loon, 1997; Van Loon & Van Strien, 1999). Elas conferem resistência local e sistêmica em resposta a patógenos, após o reconhecimento dos elicitores por receptores da planta (Dixon & Harrison, 1990).

Entre estas proteínas são encontradas hidrolases, fatores de transcrição, inibidores de proteases, enzimas relacionadas a várias rotas metabólicas e produtos alergênicos. Seus motivos funcionais estão relacionados a um número de proteínas intra- e extra-celulares de eucariotos, que estão envolvidas em funções muito distintas, como glicoproteínas relacionadas à maturação do esperma em roedores, proteínas de reserva em sementes,

osmotina em tabaco, e genes do desenvolvimento e diferenciação floral, o que leva a crer que a função defensiva evoluiu posteriormente ao surgimento destas famílias gênicas.

Plantas sistemicamente resistentes são caracterizadas pela presença de PRs específicas. O aparecimento, após a infecção por vírus, de PRs específicas expressas "de novo" foi originalmente demonstrado em tabaco e *Nicotiana glutinosa* por Van Loon e Van Kammen em 1970 (ver Van Loon, 1982). Após a infecção pelo vírus do mosaico do tabaco (TMV), PRs aparecem nas folhas inoculadas no princípio da necrose e aumentam em quantidade com o tempo. A partir de cerca de sete dias, elas também se tornam detectáveis em órgãos não inoculados da planta, onde o vírus está ausente e onde a resistência sistêmica adquirida é provocada após nova inoculação.

Aparentemente, a capacidade da planta hospedeira em reagir da forma particular que faz está presente, sob uma forma silenciosa, durante todo o tempo, e é invocada somente pela ação ativadora do agente infectante. A informação genética subjacente à expressão dos sintomas pode não ser expressa durante o desenvolvimento da planta, ou, se o for após a infecção, será de uma forma descoordenada e imprópria, não sujeita aos controles regulatórios comuns dando espaço, assim, aos sintomas característicos da doença (Van Loon, 1982).

Exemplificando através do que ocorre em infecções por vírus, durante a defesa ativa, são expressas informações para contra-atacar a multiplicação e/ou dispersão do vírus. Frequentemente, a defesa ativa toma a forma de uma reação hipersensível, na qual o vírus é efetivamente contido em uma mancha circular no tecido ao redor da lesão necrótica no sítio em que ele penetrou. A reação hipersensível é não específica, de forma que lesões necróticas basicamente similares são formadas após a infecção por vírus, fungos ou bactérias. A reação em si é sempre semelhante. Uma vez que a lesão local começa a se desenvolver, muitas mudanças na atividade enzimática se tornam aparentes. O aumento em enzimas oxidativas e hidrolíticas é notável, a respiração é aumentada, e a biossíntese aromática é fortemente ativada.

Inicialmente, pelo menos, o vírus não é impedido fisicamente de dispersar-se pelas barreiras de fenóis oxidados formados no tecido necrosado, porque: a) vírus são capazes de escapar do local da lesão por mudanças apropriadas nas condições ambientais como, por exemplo, por aumento da temperatura; b) vírus podem ser encontrados a mais de 1,5 mm da área necrosada, em células que de outra forma parecem saudáveis; e c) necrose localizada, induzida por tratamento com água quente, de células envolvidas na

multiplicação viral em plantas infectadas sistemicamente, não interrompem a dispersão sistêmica do vírus.

Essas observações apoiam a concepção de que o mecanismo de localização é operativo nas células vizinhas, ainda não infectadas, ao invés de nas células em que os vírus se multiplicam ativamente. Esta é uma observação muito importante e é evolutivamente significativa, pois a proteção das células sadias ao redor do tecido infectado é muito mais eficiente do que tentar combater o vírus em uma célula sob seu domínio.

1.4.2. Classificação

As PRs são classificadas em famílias com base nas suas seqüências de aminoácidos, relações sorológicas e/ou atividade enzimática ou biológica (Van Loon *et al.*, 1994). Elas devem estar bem caracterizadas e devem ter propriedades claramente distintas para garantir sua inclusão em uma nova família. Hoje estas proteínas já foram classificadas em 14 grupos distintos (Van Loon & Van Strien, 1999), e em breve, possivelmente mais três grupos serão anexados a estes (Van Loon, comunicação pessoal, 2001).

PR1

De todas as PRs já descritas, um grupo que continua até hoje como um mistério é o da PR1. O mecanismo de ação que resulta em suas propriedades antifúngicas permanece desconhecido. Entretanto, suas características estruturais já foram estudadas por ressonância magnética nuclear, o que revelou uma arquitetura molecular singular, onde quatro alfa-hélices e quatro bainhas-beta são arranjadas em paralelo. O núcleo da molécula, compacto e bipartido, é estabilizado por interações hidrofóbicas e múltiplas pontes de hidrogênio. Proteínas homólogas a estas PRs já foram descritas para os mais diversos tipos de organismos, incluindo fungos, insetos, répteis e humanos. O domínio funcional da PR1, conhecido por SCP (sperm-coating protein, glicoproteínas de revestimento do esperma, encontradas em roedores), também é encontrado em venenos de vespas e lagartos (Lu *et al.*, 1993), além de ter alta homologia com uma proteína especificamente expressa em um tipo de neoplasia que ocorre no tecido intersticial do cérebro humano, o glioblastoma multiforme. Szyperski *et al.* (1998) relatam que a GliPR apresenta 35% de similaridade na seqüência de aminoácidos com uma PR1 do tomate. Os

autores compararam a estrutura molecular de ambas as proteínas e identificaram um “cluster” de quatro resíduos de aminoácidos aparentemente conservados em todas as PR1 de plantas. Eles sugerem a existência de um sítio ativo comum para as proteínas PR do ser humano e de plantas, e uma ligação funcional entre o sistema imunológico humano e o de defesa das plantas. É possível que todas as proteínas desta superfamília operem de acordo com o mesmo mecanismo molecular.

Este é o grupo menos definido de PRs e acredita-se que esta classificação sofrerá uma reestruturação em breve. Embora a função destas proteínas em plantas ainda seja desconhecida, já foi comprovado que elas são fortemente induzidas no início da resistência sistêmica adquirida (SAR) em tabaco (Uknes *et al.*, 1993).

PR2

As proteínas PR2 correspondem à família 17 das glicosil hidrolases (Henrissat, 1991). A região mais conservada nos produtos desta família contém um resíduo estável de triptofano, que pode estar envolvido na interação com os substratos glucanos (Ori *et al.*, 1990); esta região também contém glutamato, que age como nucleófilo no mecanismo catalítico (Varghese *et al.*, 1994).

PR3

As proteínas da família 19 das quitinases, ou PR3, são enzimas que catalisam a hidrólise das ligações beta-1,4-N-acetil-D-glucosamina em polímeros de quitina (Flach *et al.*, 1992), principais componentes da parede celular de fungos e insetos. Pela similaridade de seqüência, as quitinases pertencem às famílias 18 e 19 na classificação das glicosil hidrolases. As quitinases da família 19 são enzimas que funcionam na defesa contra fungos e insetos destruindo suas paredes celulares. O domínio catalítico destas enzimas consiste de cerca de 220 a 230 resíduos de aminoácidos (Henrissat, 1991).

PR4

Barwin é uma proteína básica isolada de extratos aquosos de sementes de cevada. Nesta proteína foi caracterizado um domínio com seis resíduos de cisteína, que se combinam para formar três pontes dissulfídicas (Svensson *et al.*, 1992; Ludvigsen & Poulsen, 1992). Este domínio liga-se a sacarídeos e caracteriza as proteínas PR4.

Existem outras proteínas extraídas de tabaco, seringueira e batata, codificadas por genes induzidos por ferimentos, que apresentam domínios com altos índices de similaridade ao de barwin. Estas características sugerem que tal domínio possa estar envolvido em mecanismos de defesa mais gerais.

PR5

As PR5 são proteínas que apresentam o mesmo domínio da taumatina, uma proteína intensamente doce (cerca de cem mil vezes mais doce do que a sacarose), encontrada em *Thaumatococcus danielli*, um arbusto do oeste da África (Edens *et al.*, 1982). As proteínas deste grupo são estruturalmente diversas e aparentemente onipresentes em plantas (Ruiz-Mendrano *et al.*, 1992). A família inclui a taumatina, a osmotina (Singh *et al.*, 1989), duas formas de PRs de tabaco induzidas após a infecção por vírus, o inibidor alfa-amilase de tripsina, e proteínas P21 e PWIR2 de soja e trigo, respectivamente (Rebmann *et al.*, 1991). Embora a função de muitas PR5 seja desconhecida, sabe-se que estão envolvidas na resistência sistêmica adquirida e resposta ao estresse. Como outras proteínas PR, acredita-se que tenham uma estrutura beta com alto conteúdo de circunvoluções e pequenas hélices (Ruiz-Mendrano *et al.*, 1992).

PR6

Já em 1972, Green e Ryan relataram o extraordinário aumento na concentração de um inibidor de proteinases em tecidos de batata (*Solanum tuberosum*) após o ataque de suas folhas por besouros da batata do Colorado. Os autores observaram também que as folhas não atacadas pelos besouros, na mesma planta, apresentavam também um aumento na concentração dessas proteínas, em relação à planta controle. Em função destas observações, os autores propuseram que, uma vez sofrida a infecção, a planta se prepararia contra ataques subsequentes através da indução sistêmica de proteínas de defesa.

As PR-6 (inibidores de proteinase) são capazes de inibir enzimas proteolíticas de origem microbiana e de insetos, sem que exerçam essa função sobre as proteínas da própria planta. Esta característica lembra em muito o sistema imunológico humano, no funcionamento de seus anticorpos.

A forte interação entre as proteínas do trato digestivo de insetos com as PR-6 pode ser um fator de coevolução inseto-planta, servindo como um excelente objeto para o estudo da evolução por seleção positiva em sítios de interação entre as moléculas.

As proteínas da família Kunitz de inibidores da tripsina têm atividade inibitória contra proteinases de serina das famílias da tripsina e subtilisina, tiolproteinase e aspartico-proteinase. A existência, nestas famílias, de um membro sem atividade inibitória, sugere que esta função deva ter surgido em proteínas de reserva (Laskowski & Kato, 1980). Estas proteínas apresentam de 170 a 200 aminoácidos e uma ou duas pontes dissulfídicas intracadeia. A região mais conservada é encontrada na porção N-terminal e apresenta uma estrutura de 12 fitas beta arranjadas em três lobos próximos a um eixo central e seis fitas formando um cilindro beta anti-paralelo (Murzin *et al.*, 1992).

PR7

A maior proteína entre as PRs é a PR7. Designada inicialmente por P69 devido ao seu peso molecular, que varia entre 69.000 e 70.000 daltons (Granell *et al.*, 1987), foi identificada como uma endoproteinase por Vera & Conejero (1988).

Recentemente, Jordá *et al.* (1999; 2000) descreveram seis diferentes cópias deste gene formando dois “clusters” no genoma de tomate. As características mais marcantes destes genes são a ausência completa de íntrons e a sua regulação diferencial e tecido específica. Segundo os autores, quatro destes genes são induzidos como os das PRs clássicas, enquanto dois são expressos constitutivamente em plantas saudáveis e não apresentam acréscimo na expressão quando submetidos aos agentes indutores. Mesmo com essas características, esses dois últimos genes não podem ser desvinculados da atividade na defesa da planta, visto que nos tecidos com alta atividade mitótica, onde são expressos, podem ser necessários durante todo o tempo. Os quatro genes induzidos por agentes patogênicos também apresentam especificidade de tecido, como se otimizassem sua atividade com essa especificidade.

A análise da seqüência nucleotídica revela que essas PRs são sintetizadas como pré-pró-enzimas, na forma de um polipeptídeo, com 666 a 746 resíduos de aminoácidos (aa), composto de um peptídeo sinal de 22 a 28 aa, de um pro-polipeptídeo de 87 a 92 aa e do peptídeo maduro, que varia entre 552 e 631 aa.

PR8

A PR8 é uma quitinase que foi isolada primeiramente em pepino (*Cucumis sativum*) e não apresenta homologia significativa com as outras quitinases relacionadas à patogênese

previamente descritas (PR3, quitinases tipos I e II). Esta quitinase é classificada como tipo III e difere das outras duas na especificidade pelo substrato.

Os estudos de Métraux *et al.* (1988) apontaram uma similaridade de 68% na seqüência de aminoácidos postulada para esta proteína e a seqüência parcial de uma enzima bifuncional (lisoizima/quitinase) isolada de *Parthenocissus quinquefolia*.

PR9

O grupo PR9 compreende as peroxidases formadoras de lignina (Lagrimini *et al.*, 1987), que são fortemente ativadas após infecções. Estas isoenzimas catalisam a formação dos polímeros de lignina, que depositados na parede celular das plantas, constituem uma barreira física impermeável que serve como proteção nos casos de infecção.

PR10

Esta família, constituída por proteínas intracelulares de estrutura semelhante à ribonuclease, é amplamente difundida entre as plantas. Seus membros são descritos como os principais alérgenos de plantas, a família multigênica Bet v 1. A proteína Betv1 é o alérgeno mais potente do pólen de *Betula verrucosa* (ordem Fagales) e provoca alergias do tipo I, mediadas pela IgE. Ele afeta uma em cada cinco pessoas na Europa e América do Norte, causando febre, dermatite, asma e até choque anafilático (Breiteneder *et al.*, 1989). Sua estrutura é composta por seis fitas beta anti-paralelas e três alfa-hélices (Gajhede *et al.*, 1996). Quatro das fitas determinam o dobramento global e duas alfa-hélices formam um motivo helicoidal anfipático C-terminal, que determinaria o epítipo reconhecido pela célula T (Faber *et al.*, 1996; Kungl, 1996). Esta família inclui outras proteínas alérgênicas de plantas, como *Mald I* da maçã, ABR 17 e DRRG49-D do pêssego, e *Cora I* da avelã (Breiteneder *et al.*, 1993).

Inicialmente observadas em cultura celular de salsinha (*Petroselinum crispum*), após estímulo por elicitores, estas PRs foram descritas como PR1 ou PR2 (Somssich *et al.*, 1986), mas não apresentam qualquer semelhança com as PRs de tais tipos em tabaco.

PR11

A classe V de quitinases de plantas foi descrita por Melchers *et al.* (1994) e não apresenta qualquer similaridade com as outras classes de quitinases. Porém, apresenta similaridade com as exo-quitinases de certas bactérias (*Bacillus circulans*, *Serratia*

marcescens e *Streptomyces plicatus*). Os autores demonstraram que essas quitinases são induzidas após o tratamento com agentes patogênicos e diferentes formas de estresse abiótico. Foi também demonstrado que essas proteínas apresentam atividade de endo-quitinase e que atuam sinergisticamente com a beta-glucanase classe I (PR2) contra fungos. As duas seqüências encontradas em tabaco têm peso molecular de 41.000 e 43.000 daltons, respectivamente.

PR12

As PR12 são defensinas, ou gama-purotioninas, caracterizadas primeiramente em radite (Terras *et al.*, 1992). São proteínas de reconhecida atividade antifúngica, ricas em resíduos de cisteína e amplamente distribuídas entre as plantas.

PR13

As tioninas de plantas que caracterizam as PR13 são diferentes das gama-purotioninas do grupo PR12. A Thi2.1, tionina de *Arabidopsis thaliana*, isolada por Epple *et al.* (1995), é marcadamente induzida por patógenos e certos elicitores químicos, mas não por salicilatos (indutor comum de resistência sistêmica adquirida, SAR), indicando que este gene possui uma rota de ativação, ao menos parcialmente, diferente das outras PRs.

PR14

As LTPs (Nonspecific lipid-transfer proteins) de plantas são proteínas ubíquas e formam uma única família de polipeptídeos com 90 a 95 resíduos de aminoácidos, que foi identificada (em nível protéico ou de DNA), em uma variedade de tecidos, em espécies de mono- e dicotiledôneas (Yamada, 1992; García-Olmedo *et al.*, 1995).

As LTPs possuem a capacidade de transferir fosfolipídios entre duas membranas *in vitro*. São proteínas pequenas (~8,7kDa) estabilizadas por quatro pontes dissulfídicas e com uma cavidade central hidrofóbica formando uma espécie de túnel. As LTPs são ativas *in vitro* contra uma série de bactérias e fungos; entretanto, o mecanismo de sua ação como PR ainda não é conhecido. Cogita-se que elas se introduzam na membrana celular do patógeno e a sua cavidade hidrofóbica central forme um poro, causando a saída de íons intracelulares, o que acarreta a morte do microrganismo (Selitrennikoff, 2001).

PR15

Recentemente, em comunicação pessoal, Van Loon relatou que três novas proteínas poderiam vir a ser classificadas como PRs. Uma delas, uma oxalato-oxidase, seria classificada como PR15.

Essa proteína, conhecida como germin, é uma glicoproteína homopentamérica, de reconhecido envolvimento com a germinação de sementes (Lane *et al.*, 1993). As oxalato-oxidases são enzimas que catalisam a degradação oxidativa do oxalato em carbonato e peróxido de hidrogênio. Como a germin está relacionada com a alteração das propriedades da parede celular durante a germinação, acredita-se que essa função também esteja presente na defesa contra patógenos.

I.5. Evolução dos Genes de Resistência e de Defesa em Plantas

Os genes *R* pertencem a grandes famílias multigênicas que sofrem rápida diversificação adaptativa, aparentemente em resposta à seleção intensa para novas especificidades de resistência, o que sugere que seriam o alvo principal da seleção na coevolução hospedeiro-patógeno. A evolução gene-a-gene neste grupo, ajustar-se-ia mais a um modelo do tipo "batalha de trincheira" (Stahl *et al.*, 1999) do que ao de "corrida armamentista" proposto por P. R. Ehrlich e P. H. Raven (Thompson & Burdon, 1992), o qual pressupõe que as populações seriam inicialmente monomórficas para o loco considerado. O modelo de Stahl *et al.* (1999) propõe que alelos para resistência e suscetibilidade teriam coexistido por milhões de anos e, avanços e retrações na frequência dos alelos de resistência, manteriam a variabilidade para a resistência a doenças na forma de um polimorfismo dinâmico.

Alguns dos trabalhos publicados até o momento apresentam considerações quanto à origem evolutiva destas proteínas, porém poucos dispõem-se a analisá-las no contexto de suas seqüências codificadoras. Segundo Bishop *et al.* (2000), as proteínas relacionadas à patogênese em plantas têm sido muito estudadas do ponto de vista bioquímico e fisiológico, porém conhece-se pouco sobre os processos evolutivos que atuam nos genes que codificam para essas proteínas. Estes autores aplicaram modelos de substituição de códons em um estudo evolutivo sobre as quitinases. Seus resultados sugeriram uma evolução rápida por seleção positiva em sítios de atividade da quitinase classe III (classificada no artigo como classe I) em *Arabidopsis*, e que estes sítios seriam alvo da seleção na co-evolução planta-patógeno. No estudo, testes de neutralidade e de estrutura cristalina

combinados para analisar as conseqüências funcionais e estruturais da seleção sugeriram seleção positiva para alguns sítios de aminoácidos, e o posicionamento destes sítios na estrutura da molécula confirmou a localização prevista. Este trabalho, no entanto, concentra-se em apenas um grupo de plantas, o gênero *Arabis*, o grupo mais próximo do organismo modelo *Arabidopsis thaliana*. Os autores sugeriram que a evolução adaptativa rápida do sítio ativo seria um fenômeno geral nas quitinases de classe III das plantas dicotiledôneas, mas não puderam fazer interpretações sobre um padrão mais geral, pois as linhagens analisadas são filogeneticamente muito correlacionadas e as seqüências que eles dispunham eram apenas fragmentos dessas proteínas.

A efetividade variável das quitinases específicas contra diferentes patógenos e a existência de inibidores microbianos da quitinase sugerem a hipótese de que as quitinases podem co-evoluir com fungos em resposta à variação nas defesas dos patógenos contra a atividade quitinolítica (Bishop *et al.*, 2000).

Saliente-se que a magnitude da seleção imposta por patógenos não é conhecida para qualquer gene *PR*, nem foram identificadas as adaptações moleculares que compõem a resposta à seleção (co-evolução) para qualquer planta (Bishop *et al.*, 2000).

I.6. Objetivos da Dissertação

As PRs constituem um grupo de proteínas cuja a análise dos mecanismos de evolução pode fornecer informações significativas, não só sobre estas proteínas, mas também sobre outros genes de defesa em plantas. Assim, decidiu-se realizar um estudo comparativo e abrangente sobre a variação, em nível molecular, dessas proteínas. O estudo envolveu:

1. Busca das seqüências apropriadas nos bancos de dados de domínio público da internet, SWISS-PROT, TrEMBL e GenBank.
2. Alinhamento das seqüências, nas diferentes classes de PRs, de acordo com a sua similaridade.
3. Teste, através do método da máxima verossimilhança, de modelos de substituições de códons que resultam em mutações sinônimas e não-sinônimas em nível protéico.
4. Para um subgrupo de seqüências, testes através do método da parcimônia.

O *objetivo final* foi verificar o grau com que a seleção positiva está ocorrendo neste conjunto heterogêneo de seqüências.

II. Material e Métodos

II.1. Obtenção das Seqüências, Bancos de Dados Utilizados

Os bancos de dados utilizados como fonte de seqüências protéicas e de DNA foram o SWISS-PROT e o TrEMBL (Bairoch & Apweiler, 2000), desenvolvidos em parceria pelo Instituto Suíço de Bioinformática (SIB) e o Instituto Europeu de Bioinformática (EBI), ligado ao Laboratório Europeu de Biologia Molecular (EMBL). Em raros casos, as seqüências foram obtidas do GenBank (NCBI- Centro Nacional de Informação Biotecnológica, USA). Esses bancos de dados e diversas outras ferramentas são encontradas no site ExpASy (Expert Protein Analysis System): <http://www.expasy.org/>.

A busca das seqüências foi realizada com base na classificação das PRs de Van Loon e Van Strien (1999). Como palavras-chave para as buscas foram usadas as descrições e os nomes dos autores referidos para cada PR citados no artigo.

Foram utilizadas diversas ferramentas de busca relacionadas com os bancos de dados, com o objetivo de cobrir todas as seqüências registradas até o momento. A ferramenta que demonstrou ser mais eficiente para o reconhecimento e identificação das seqüências disponíveis foi o SRS – Sequence Retrieval System (Etzold & Argos, 1993), disponível na página do SWISS-PROT. Esse procedimento foi utilizado para grande parte das seqüências e, principalmente para as PRs imperfeitamente descritas, foi de inestimável importância, por permitir várias frentes de busca, por referência bibliográfica, organismo, palavra-chave, nome do gene ou código de acesso. O SIB BLAST Network Service (Altschul *et al.*, 1997) também foi empregado e demonstrou ser bastante abrangente, evitando ao máximo o recolhimento de seqüências que não seriam úteis nas análises. Essa ferramenta tem um link direto com a entrada do SWISS-PROT (<http://www.expasy.org/sprot>).

Uma vez encontradas as proteínas, suas seqüências foram divididas em arquivos, alinhadas e agrupadas em árvores filogenéticas. Foram então escolhidas, entre os grupos mais relacionados, aquelas que seriam utilizadas nas análises.

II.2. Métodos de Análise

Todas as seqüências protéicas, obtidas nas buscas por grupo de PRs, foram alinhadas com auxílio do programa CLUSTAL X versão 1.8 (Multiple Sequence Alignment Program - Thompson *et al.*, 1997; Jeanmougin *et al.*, 1998).

As seqüências de DNA foram retroalinhadas a partir das seqüências de aminoácidos obtidas nos bancos de dados, com o auxílio do programa GeneDoc (Multiple Sequence Alignment Editor & Shading Utility), versão 2.5.000 (Nicholas & Nicholas, 1997).

A conversão dos arquivos de seqüências de DNA alinhadas de formato GCG/MSF (Multiple Sequence File, Genetics Computer Group, Madison, USA), utilizado pelo programa GeneDoc, para PHYLIP (Felsenstein, 1993), utilizado no programa PAML, foi realizada através do programa ForCon (Raes & Van de Peer, 1999).

Para cada uma das famílias foram obtidas filogenias pelo método de neighbor-joining, usando o programa MEGA2 (Kumar *et al.*, 2000; 2001). Grupos de seqüências obtidos nestas filogenias foram selecionados e suas relações filogenéticas estabelecidas novamente usando o método de máxima verossimilhança, com o auxílio do programa BASEML, do pacote PAML (Yang, 1997). Com este programa, também foram determinados os tamanhos dos ramos, as freqüências de cada base nucleotídica, a proporção de sítios constantes e a taxa de transição/transversão (κ).

O programa CODEML, do pacote PAML, foi usado para determinar com que probabilidade os sítios individuais estão sofrendo seleção positiva em cada conjunto de dados. Os 14 modelos (M0-M13) listados na Tabela 2 de Yang *et al.* (2000a) e implementados pelo programa CODEML foram utilizados na análise. O máximo de parâmetros foi mantido para todas as análises realizadas.

Para testar os modelos do CODEML que melhor se adaptam a cada grupo de dados analisados, o Critério de Informação de Akaike (AIC - Sakamoto *et al.*, 1986) foi escolhido como método estatístico. Nesse teste, o valor absoluto do logaritmo da máxima verossimilhança, obtido para cada modelo, é somado ao número de parâmetros variáveis no modelo (penalidade maior para o maior número de parâmetros), e o valor obtido é, então, comparado ao dos outros modelos. O modelo com o menor valor de AIC é considerado o que melhor representa os dados.

Apenas como base para comparação, alguns conjuntos de seqüências utilizados no estudo foram analisados pelo método de máxima parcimônia proposto por Suzuki e Gojobori (1999). O programa utilizado foi o SGI, desenvolvido por Su (2000), e os resultados foram processados com a planilha sugerida pelo autor.

III. Resultados

III.1. Informações gerais

A disponibilidade de seqüências das 15 famílias de PRs analisadas foi muito variável. Proteínas como as dos grupos PR1 e PR5 já foram amplamente estudadas e seqüenciadas em diversos organismos. Por outro lado, para a família das PR11, um caso extremo, só são encontradas duas seqüências claramente identificadas nos bancos de dados. Em outras famílias, há disponibilidade de seqüências, mas estas são, em grande parte, representadas por fragmentos da região codificadora, que não podem ser utilizados, pois tornam a análise menos efetiva. Há também uma alta variabilidade no tamanho de algumas PRs, o que dificultou o alinhamento e a identificação dos sítios.

Devido a limitações das ferramentas de análise disponíveis (tempo de processamento aumentando exponencialmente com o aumento do número de seqüências) e em função do alto grau de divergência observado entre grupos de seqüências de algumas das 15 famílias de PRs, o estudo foi desenvolvido por subgrupos de proteínas dentro de cada uma delas. A escolha dos subgrupos a serem analisados foi baseada na análise filogenética das 15 famílias, obtida com o auxílio do programa MEGA2, e na similaridade de tamanho e seqüência dentro de cada família. Quarenta subgrupos de seqüências foram analisados (Tabela III.1). Em algumas famílias, foi possível o estabelecimento de diversos grupos homogêneos para a análise, reunindo-se seqüências de espécies de mesma família taxonômica, ou até variantes de uma mesma espécie. A família com o maior número de subgrupos foi a PR14, para a qual foi verificada a formação de oito grupos de seqüências, seguida pela PR5, com cinco agrupamentos. Para algumas PRs, apenas um subgrupo pode ser selecionado.

Foram obtidas, assim, 201 seqüências distribuídas em uma ampla gama de organismos. Nada menos do que 32 gêneros, com um variado número de espécies, estão listados na Tabela III.1.

Tabela III 1

Conjuntos de seqüências utilizados, espécies às quais pertencem, e respectivos números de acesso nos bancos de dados.

PR	Conjuntos de seqüências	OTUs	Espécies	Números de acesso
PR1	A	4	<i>Lycopersicon esculentum</i> <i>Solanum tuberosum</i>	Q04108, P04284 Q9SC15, Q941G6
	B	5	<i>Nicotiana tabacum</i>	P07053, Q40397, P08299, Q40557, P09042
	C	6	<i>Triticum aestivum</i> <i>Hordeum vulgare</i>	Q94F73, O82714 Q43489, Q05968, P35792, P35793
PR2	A	4	<i>Nicotiana tabacum</i>	P23432, P23433, P52399, P23547
	B	6	<i>Solanum tuberosum</i>	P52401, O65360
			<i>Lycopersicon esculentum</i>	Q01413
			<i>Nicotiana tabacum</i> <i>Nicotiana plumbaginifolia</i>	P27666, P23546 P23431
PR3	A	13	<i>Arabidopsis thaliana</i>	O22073, O22072, O22065, P19172, O22066, O22070, O22068, O22069, O22067, O22071, O22074, O22076, O22075
	B	5	<i>Chenopodium amaranticolor</i> <i>Beta vulgaris</i>	O23803, O23804, O23806, O23805 Q9XFW7
PR4	A	4	<i>Lycopersicon esculentum</i> <i>Nicotiana tabacum</i>	P32045, Q40558 P29062, P29063
	A	4	<i>Avena sativa</i>	P50695, P50696, P50697, P50698
PR5	B	7	<i>Hordeum vulgare</i>	P32937, CAB99485, P32938
			<i>Triticum aestivum</i> <i>Secale cereale</i>	P27357, Q94F70, Q9S776, Q9ZSN0
	C	5	<i>Pyrus pyrifolia</i>	O80327
			<i>Prunus avium</i>	P50694
			<i>Castanea sativa</i> <i>Malus domestica</i>	Q9SMH2 Q9FSG7, O82546
	D	8	<i>Solanum dulcamara</i>	Q9FT35
			<i>Solanum commersonii</i>	P50702, P50701
<i>Lycopersicon esculentum</i> <i>Nicotiana tabacum</i> <i>Petunia hybrida</i> <i>Capsicum annuum</i>			P12670 P14170, Q40529 Q94JN9 Q9ARG0	
E	4	<i>Nicotiana tabacum</i>	P25871	
		<i>Nicotiana sylvestris</i> <i>Solanum commersonii</i> <i>Lycopersicon esculentum</i>	Q40428 P50703 Q9M3X2	
PR6	A	4	<i>Solanum tuberosum</i>	Q07459, Q43648, P08454, Q00783
PR7	A	6	<i>Lycopersicon esculentum</i>	Q96478, O04678, Q9ZR46, Q9SAN2, Q9LWA4, Q9LWA3
	B	5	<i>Lycopersicon esculentum</i>	Q96478, O04678, Q9SAN2, Q9LWA4, Q9LWA3
PR8	A	7	<i>Arabis gemmifera</i>	Q9SXJ1
			<i>Arabidopsis thaliana</i>	Q9S7J5, Q9SXJ3, Q9SXJ4, Q9S838, P19171, Q9SXJ2
	B	18	<i>Arabis</i> (11)	Q9M7F5, Q9M7H2, Q9M7G6, Q9M7H5, Q9M7F7, Q9M7G0, Q9M7F9, Q9M7G7, Q9M7H3, Q9M7G2, Q9M7H1, Q9M7G5, Q9M7G1, Q9M7F4, Q9M7G4, Q9M7H0, Q9M7G9, Q9M7H4
C	17	<i>Arabis</i> (10)	Q9M7F9, Q9M7G9, Q9M7H4, Q9M7G7, Q9M7H3, Q9M7G1, Q9M7G0, Q9M7F7, Q9M7F4, Q9M7F5, Q9M7H0, Q9M7H2, Q9M7H5, Q9M7G6, Q9M7G4, Q9M7H1, Q9M7G2	

Tabela III 1(Cont.)

Conjuntos de seqüências utilizados, espécies às quais pertencem, e respectivos números de acesso nos bancos de dados.

PR	Conjuntos de seqüências	OTUs	Espécies	Números de acesso
PR9	A	6	Populus kitakamiensis Populus nigra Populus trichocarpa	Q43050 Q40949, Q40950 Q43099, Q43100, Q43101
	B	5	Populus nigra Populus trichocarpa	Q40949, Q40950 Q43099, Q43100, Q43101
PR10	A	12	Betula verucosa	P43178, P43183, P43179, P15494, P43177, P43185, CAA54486, P43180, P43176, P43184, P43186, P45431
	B	4	Petroselinum crispum Apium graveolens	P19418, P19417, P27538 P49372
	C	6	Pisum sativum Phaseolus vulgaris Glycine max	P14710, P13239, P27047 P25986, P25985 P26987
PR11	A	2	Nicotiana tabacum	Q43576, Q43591
	B		Nicotiana tabacum Arabidopsis thaliana	Q43576, Q43591 O81862, O81863
PR12	A	7	Wasabia japonica Raphanus sativus Brassica oleracea Brassica napus	Q9FS38 O24332, P30230, O24331 Q94IN7 Q39313, P30225
	B	5	Wasabia japonica Raphanus sativus Brassica oleracea Brassica napus	Q9FS38 P30230, O24331 Q94IN7 P30225
PR13	A	4	Hordeum vulgare	P08772, P09618, Q42838, P09617
	B	6	Secale cereale Hordeum vulgare Triticum aestivum	Q9ZNY5 P21742 P01543, Q9T0P2, P32032, Q43205
PR14	A	8	Hordeum vulgare Triticum aestivum	Q43767, Q43875, Q42842, Q42849, Q43762, Q43871 Q9ATG4, Q9FUK0
	B	4	Zea mays Sorghum bicolor	P19656, O24583 Q43193, Q43194
	C	6	Oryza sativa	P23096, Q42978, Q42999, O22483, O22482, O22485
	D	4	Lycopersicon esculentum Nicotiana tabacum	P93224, P27056 Q42952, Q03461
	E	4	Brassica oleracea Brassica napus	Q42642 Q42614, Q42615, Q42616
	F	5	Gossypium hirsutum	Q9M6B8, Q9M6B6, Q9FVA5, O49200, Q43129
	G	4	Prunus dulcis Prunus avium Malus domestica Pyrus communis	Q43017 Q9M5X8 Q9M5X7 Q9M5X6
	H	5	Brassica napus Brassica oleracea	Q42614, Q42615, Q42616 Q42642, Q43304
PR15	A	4	Lolium perenne Triticum aestivum Hordeum vulgare	Q9FEW6 P15290 P45851, O24004
	B	6	Lolium perenne Triticum aestivum Hordeum vulgare	Q9FEW6 P15290, P26759, P93600 P45851, O24004

OTUs - Operational Taxonomic Units

III.2 Informações por Classe de PRs

III.2.1. PR 1

Esta classe de PRs está muito bem representada nos bancos de dados utilizados, principalmente no que diz respeito ao número de seqüências. Essa grande disponibilidade de seqüências possibilitou uma seleção exigente e a formação de grupos similares em tamanho e conservação de seqüência. Três grupos foram obtidos, dois contendo espécies de Solanaceae (dicotiledôneas) e um grupo com espécies de Poaceae (monocotiledôneas). A presença de representantes destes dois grupos tão distintos na análise permite a melhor compreensão dos padrões evolutivos dessas proteínas.

O alinhamento de todas as seqüências de PR1 utilizadas é apresentado na Figura III.1. Note-se que as diferenças nos tamanhos das seqüências ocorrem apenas entre os grupos e que, apesar de distantes, elas permanecem muito conservadas. A árvore filogenética apresentada na Figura III.2 demonstra como os grupos estão bem definidos, sendo de notar que a seqüência Q43489 de *Hordeum vulgare* se associa com a O82714 de *Triticum aestivum*, e não com as outras três de sua espécie. Informações adicionais são fornecidas na Tabela III.2. O alinhamento feito envolve cerca de 500 pb, a proporção de CG situa-se entre 45% e 64%, e a percentagem de sítios constantes entre 87% e 89% nos três grupos. No grupo A, quinze códons foram indicados como possíveis alvos de seleção positiva, isto é, apresentam valores de ω acima de 1 (Figura III.1). Os modelos com menores valores de AIC são M1 e M5, porém seis outros (Tabela III.2), fornecem uniformemente indicações deste tipo de seleção para as 15 posições. O ω médio, entre todos os códons, foi de 0,208. Já com referência ao grupo B, 10 códons apresentaram valores de ω acima de 1; neste caso, a distribuição que melhor se adapta aos dados é a M3, porém, sete outras (Tabela III.2) mostraram valores concordantes para esses códons. O ω médio encontrado foi de 0,420. Por último, o conjunto C, formado por seqüências de plantas monocotiledôneas, teve 10 códons apontados como alvo de seleção positiva para dois modelos (M3 e M8), e aquele que melhor se adaptou aos dados, segundo o critério de informação de Akaike, foi o M1, que permite somente a ocorrência de classes de códons neutros ou em que as mutações são completamente deletérias.

PR 1

	*	20	*	40	*	60	*	80	*	100																																																							
Q04108	:	MGLFNIS----	LLL--	TCLMVLAI	FHSC	EAQNSPQD	YLAVHND	ARAQVGV	GPMSWD	AMLASRAQNYANSR	AGDCNLIHSG--	AGENLAKGG--	DFTGRA	:	90																																																		
P04284	:	MGLFNIS----	LLL--	TCLMVLAI	FHSC	EAQNSPQD	YLAVHND	ARAQVGV	GPMSWD	AMLASRAQNYANSR	AGDCNLIHSG--	AGENLAKGG--	DFTGRA	:	90																																																		
Q93C15	:	MGLFNIS----	LLL--	TCLMVLAI	FHSC	EAQNSPQD	YLAVHND	ARAQVGV	GPMSWD	AGLASRAQNYANSR	TGDCNLIHSG--	AGENLAKGS	G--DFTGRA	:	90																																																		
Q941G6	:	MGLFNIS----	LLL--	TCLMVLAI	FHSC	EAQNSPQD	YLAVHND	ARAQVGV	GPMSWD	AGLASRAQNYANSR	TGDCNLIHSG--	AGENLAKGS	G--DFTGRA	:	90																																																		
P07053	:	MGFFLFSQ	MPFFLV	STLLFL	II	ISHSHAQNS	QDYLDAHNT	ARADV	GVVEPL	TWDNGVAAY	AQNYVLSQA	AADCNLV	VHSHGQYGENLAQ	QSGG--	DFMTAA	:	98																																																
Q40397	:	MAFILFSQ	MPFLLV	STLLFL	II	ISHSHAQNS	QDYLDAHNT	ARADV	GVVEPL	TWDDEVAAY	AQNYVLSQA	AADCNLV	VHSHGQYGENLAM	SGG--	DFMTAA	:	98																																																
P08299	:	MGFVLF	SQPSFL	LVSTLL	FL	VISHS	CRAQNS	QDYLDAHNT	ARADV	GVVEPL	TWDDQVAAY	AQNYASQA	AADCNLV	VHSHGQYGENLA	EGSG--	DFMTAA	:	98																																															
Q40557	:	MGFVLF	SQPSFL	LVSTLL	FL	VISHS	CRAQNS	QDYLDAHNT	ARADV	GVVEPL	TWDDQVAAY	AQNYASQA	AADCNLV	VHSHGQYGENLA	EGSG--	DFMTAA	:	98																																															
P09042	:	MEFVLF	SQMSFF	LVSTLL	FL	ISHS	HAQNS	QDYLDAHNT	ARADV	GVVEPL	TWDDQVAAY	AQNYASQA	AADCNLV	VHSHGQYGENLA	SGG--	DFLTAA	:	98																																															
Q94F73	:	MQTPKLA	----	ILL--	--	ALAMSA	AMVLSQA	QNSPQD	YLSPHNA	ARA	AVGVAV	TWSTKLQGF	AQSYANQR	INDCKLQHS	GGPYGENIF	WGS	AGADWKAAD	:	94																																														
O82714	:	METPKLA	----	VLL--	--	ALAMAA	AVMLSQA	QNSPQD	YLSPHNA	ARA	AVGVAV	SWSTKLQGF	AQSYANQR	INDCKLQHS	GGPYGENIF	WRS	AGADWKAAD	:	94																																														
Q43489	:	METPKLA	----	VLL--	--	ALAMAA	AVMLSQA	QNSPQD	YLSPHNA	ARA	AVGVAV	SWSTKLQGF	AQSYANQR	INDCKLQHS	GGPYGENIF	WRS	AGADWKAAD	:	94																																														
Q05968	:	MQTPKLA	----	ILL--	--	ALAMAA	AVMLSQA	QNSPQD	YVSPHNA	ARS	AVGVAV	SWSTKLQGF	AQNYANQR	INDCKLQHS	GGPYGENIF	WGS	AGADWKAAD	:	94																																														
P35792	:	MQTPKLA	----	ILL--	--	ALAMAA	AVMLSQA	QNSPQD	YVSPHNA	ARS	AVGVAV	SWSTKLQGF	AQNYANQR	INDCKLQHS	GGPYGENIF	WGS	AGADWKAAD	:	94																																														
P35793	:	MQTPKLV	----	ILL--	--	ALAMSA	AMVLSQA	QNSPQD	YVSPHNA	ARA	AVGVAV	SWSTKLQGF	AQNYANQR	INDCKLQHS	GGPYGENIF	WGS	AGADWKAAD	:	94																																														
Q04108	:	-AVQLWV	SE	RP	YNYA	TNQC	VGK	MCH	YTVQ	WV	WNSV	R	LG	C	R	A	R	C	N	N	G	-	W	W	F	I	S	C	N	Y	D	P	V	G	N	W	V	G	E	R	P	:	159																						
P04284	:	-AVQLWV	SE	RP	YNYA	TNQC	VGK	K	CH	YTVQ	WV	WNSV	R	LG	C	R	A	R	C	N	N	G	-	W	W	F	I	S	C	N	Y	D	P	V	G	N	W	V	G	E	R	P	:	159																					
Q93C15	:	-AVQLWV	GE	FP	NY	G	T	QC	A	S	G	Q	V	C	G	H	T	Q	V	W	W	N	S	V	R	L	G	C	G	R	A	R	C	N	N	G	-	W	W	F	I	S	C	N	Y	D	P	V	G	N	W	V	G	E	R	P	:	159							
Q941G6	:	-AVQLWV	GE	FP	NY	G	T	QC	A	S	G	Q	V	C	G	H	T	Q	V	W	W	N	S	V	R	L	G	C	G	R	A	R	C	N	N	G	-	W	W	F	I	S	C	N	Y	D	P	V	G	N	W	V	G	E	R	P	:	159							
P07053	:	KAVEMW	V	DE	K	Q	Y	D	H	S	N	T	C	A	Q	G	V	C	G	H	T	Q	V	W	W	N	S	V	R	V	G	A	R	V	C	M	N	G	-	G	Y	V	V	S	C	N	Y	D	P	P	G	N	V	I	G	Q	S	P	:	168					
Q40397	:	KAVEMW	V	DE	K	Q	Y	D	H	S	N	T	C	A	Q	G	V	C	G	H	T	Q	V	W	W	N	S	V	R	V	G	A	R	V	C	M	N	G	-	G	Y	V	V	S	C	N	Y	D	P	P	G	N	V	I	G	Q	S	P	:	168					
P08299	:	KAVEMW	V	DE	K	Q	Y	D	H	S	N	T	C	A	Q	G	V	C	G	H	T	Q	V	W	W	N	S	V	R	V	G	A	R	V	C	M	N	G	-	G	Y	V	V	S	C	N	Y	D	P	P	G	N	V	I	G	Q	S	P	:	168					
Q40557	:	KAVEMW	V	DE	K	Q	Y	D	H	S	N	T	C	A	Q	G	V	C	G	H	T	Q	V	W	W	N	S	V	R	V	G	A	R	V	C	M	N	G	-	G	Y	V	V	S	C	N	Y	D	P	P	G	N	V	I	G	Q	S	P	:	168					
P09042	:	KAVEMW	V	NE	K	Q	Y	A	H	S	N	T	C	A	Q	G	V	C	G	H	T	Q	V	W	W	N	S	V	R	V	G	A	R	V	C	M	N	G	-	G	Y	I	V	S	C	N	Y	D	P	P	G	N	V	I	G	K	S	P	:	168					
Q94F73	:	-AVNAW	V	GE	K	K	D	Y	D	Y	G	S	N	T	C	A	G	K	V	C	G	H	T	Q	V	W	W	R	A	S	T	I	G	C	A	R	V	V	C	M	N	N	L	G	V	F	I	T	C	N	Y	E	P	R	G	N	I	I	G	Q	K	P	:	164	
O82714	:	-AVKLW	V	DE	K	K	D	Y	D	Y	G	S	N	T	C	A	G	K	V	C	G	H	T	Q	V	W	W	R	A	S	T	I	G	C	A	R	V	V	C	M	N	N	R	G	V	F	I	T	C	N	Y	E	P	A	G	N	V	V	G	Q	K	P	:	164	
Q43489	:	-AVKLW	V	DE	K	K	D	Y	D	Y	G	S	N	T	C	A	G	K	V	C	G	H	T	Q	V	W	W	R	A	S	T	I	G	C	A	R	V	V	C	M	N	N	R	G	V	F	I	T	C	N	Y	E	P	A	G	N	V	V	G	Q	K	P	:	164	
Q05968	:	-AVNSW	V	W	SE	K	K	D	Y	D	Y	G	S	N	T	C	A	G	K	V	C	G	H	T	Q	V	W	W	R	A	S	T	I	G	C	A	R	V	V	C	M	N	N	R	G	V	F	I	T	C	N	Y	E	P	R	G	N	I	I	G	Q	K	P	:	164
P35792	:	-AVNSW	V	W	NE	K	K	D	Y	D	Y	G	S	N	T	C	A	G	K	V	C	G	H	T	Q	V	W	W	R	A	S	T	I	G	C	A	R	V	V	C	M	N	N	R	G	V	F	I	T	C	N	Y	E	P	R	G	N	I	V	G	Q	K	P	:	164
P35793	:	-AVNSW	V	W	SE	K	K	D	Y	D	Y	G	S	N	T	C	A	G	K	V	C	G	H	T	Q	V	W	W	R	A	S	T	I	G	C	A	R	V	V	C	M	N	N	R	G	V	F	I	T	C	N	Y	E	P	R	G	N	I	V	G	Q	K	P	:	164

Figura III.1. Alinhamento das seqüências de aminoácidos dos três conjuntos da PR1. As seqüências com nomes em amarelo correspondem ao conjunto A, em rosa, B, e em azul, C. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva para cada grupo. Alinhamento realizado com ClustalX e editado com GeneDoc.

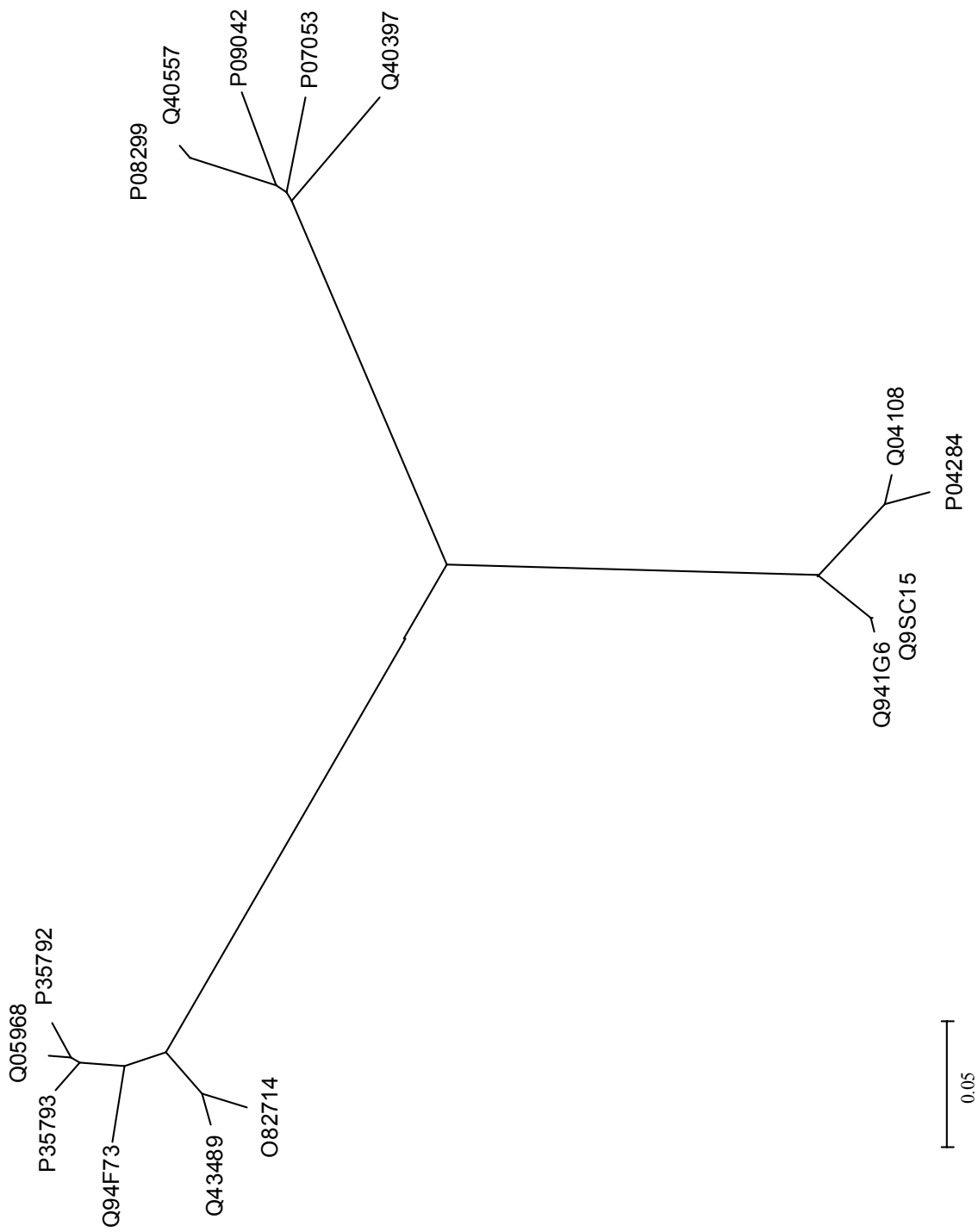


Figura III.2. Árvore filogenética das seqüências de PR 1 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 2
Resultados obtidos para os conjuntos de PR1 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	4	477 pb	49,74%	423 (88,68%)	0,1095	159	15 (9,43%)	0,208	M1 e M5	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13
B	5	504 pb	44,76%	441 (87,50%)	0,18333	168	10 (5,95%)	0,420	M3	M3, M4, M5, M6, M9, M10, M11, M12, M13
C	6	492 pb	64,30%	426 (86,59%)	0,17608	164	0 (0,00%)	0,125	M1	M3, M8

OTUs - *Operational Taxonomic Units*; AIC - *Akaike Information Criterion*

Modelos que são apresentados em itálico, indicam números bastante reduzidos de códons em relação aos outros modelos; modelos sublinhados mostravam $P > 0,95$ para a totalidade ou quase todos os códons inferidos

III.2.2. PR 2

Embora o grupo das endo-1,3-beta-glucanases esteja muito bem representado nos bancos de dados, só foi possível a escolha de dois subgrupos para a análise, devido à falta de homogeneidade entre as seqüências. Há uma grande quantidade de seqüências fragmentadas, principalmente pela ausência da extremidade amino-terminal, impedindo uma análise apropriada. Os resultados obtidos estão apresentados nas Figuras III.3 e III.4 e na Tabela III.3. Os dois grupamentos mostram boa congruência entre si, o número de pares de bases alinhados situa-se entre 1000 e 1100, a proporção de GC entre 37% e 41%, e os sítios nucleotídicos conservados ocorrem em uma freqüência de 76% a 82%. Quanto ao grupo A, cinco códons (135, 316, 329, 331 e 347) têm indicada a presença de seleção positiva em pelo menos cinco dos modelos testados, com especial destaque para o 135, que a apresentou em oito deles. O ω médio foi intermediário (0,362) aos obtidos nos grupos A e B de PR 1. Com relação ao conjunto B, 60 códons foram sugeridos como significativos para o modelo M10, porém não houve indicações de seleção positiva em qualquer outro modelo e esses resultados foram desconsiderados.

III.2.3. PR 3

As quitinases têm uma classificação própria bem estabelecida, que diverge da sua classificação como PRs. Isso representou um problema adicional, devido à dificuldade em enquadrar as seqüências obtidas nos bancos de dados nas diversas classes (3, 4, 8 e 11) de PRs que envolvem quitinases. Há falta de homogeneidade na descrição dessas proteínas nos bancos de dados, e as abreviaturas nem sempre são utilizadas com precisão. Segundo Van Loon e Van Strien (1999), a abreviatura para os genes de PR 3 seria *Chia*, para classe PR 4, *Chid*, para PR 8, *Chib*, e para PR 11 *Chic*; porém, poucas foram as seqüências que puderam ser recuperadas dos bancos de dados através da utilização destas abreviaturas como palavra-chave, pois raramente elas aparecem associadas às descrições, e algumas vezes os dados obtidos não conferem com a classificação sugerida. Outro ponto problemático é que quitinases do mesmo tipo pertencem a grupos diferentes de PRs, como, por exemplo, os tipos I e II que possuem representantes tanto na família PR 3 quanto na PR 4. Isso se deve ao fato de que a classificação das PRs está principalmente relacionada com a função da proteína na defesa da planta, enquanto que a das quitinases refere-se à estrutura molecular. Por esse motivo, o número de conjuntos de dados para essas PRs ficou reduzido e, muitas vezes, restrito às seqüências de referência de dado grupo.

PR 2

	*	20	*	40	*	60	*	80	*	100	*	120	*	
P23432	:	MALWYLFNKRSLCAAVLIIWG--LLMCMNIOMTCAOSNIQVYCYGKIANNLPSQDQVNLVKANGCIKRNRIYMSD TMI FKLSGNSNIRIILDDVPNODLEALAN--SSIAMCWQDNIRSHPPYKFKYTSIGN	:	127										
P23433	:	MALWYLFNKRSLCAAVLIIWG--LLMCMNIOMTCAOSNIQVYCYGKIANNLPSQDQVNLVKANGCIKRNRIYYPD TMI FKALNGSNIERILIEVFNODLEALAN--SSIAMCWQDNIRSHPPYKFKYTSIGN	:	127										
P52399	:	MALC--I-KMGFLAAALVWG--LLMCMNIOMTCAOS--IGVYCYGKIANNLPSQDQVNLVKANGCIKRNRIYYPD TMI FKALNGSNIERILIEVFNODLEALAN--SSIAMCWQDNIRSHPPYKFKYTSIGN	:	123										
P23547	:	MTLC--I-KMGFLAAALVWG--LLICSIQMTGAOS--IGVYCYGKIANNLPSDQDQVNLVKANGCIKRNRIYMPD TMI FVFNALRGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	123										
P27666	:	MSTS--DRHNTPQMAAIIILGCLLAVASTIEIAGAES--IGVYCYGKIANNLPSDQDQVNLVKANGCIKRNRIYMPD TMI FVFNALRGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	127										
P23431	:	MSTL--HKHNTPQMAAIIILGCLLAVASTIEIAGAES--IGVYCYGKIANNLPSDQDQVNLVKANGCIKRNRIYMPD TMI FVFNALRGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	127										
P23546	:	-----MAAIIILGCLLAVASTIEIAGAES--IGVYCYGKIANNLPSDQDQVNLVKANGCIKRNRIYMPD TMI FVFNALRGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	116										
P52401	:	-----MATSQIAVIVLLG--LLAVATNIHITEAQ--LGVYCYGKIANNLPSHSEVITQYKSRNIGRRLRYD PNHGALQALKGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	119										
065360	:	-----MATSQIAVIVLLG--LLAVATNIHITEAQ--LGVYCYGKIANNLPSHSEVITQYKSRNIGRRLRYD PNHGALQALKGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	119										
Q01413	:	-----MATSQIAVIVLLG--LLAVATNIHITEAQ--LGVYCYGKIANNLPSHSEVITQYKSRNIGRRLRYD PNHGALQALKGSNIERILIEVFNODLQSLTD--PSRAMCWQDNIRSHPPYKFKYLAVCN	:	119										
	*	140	*	160	*	180	*	200	*	220	*	240	*	260
P23432	:	EVSPSN--NGQYSQFLDHAMENVYMALAAAGLQDKIKVITATYSGLLANTYPPKDSIFRFRKSFINPIIEFLARNMLPLLANTYPPYFCHIYNTVDVPLSYALFNQO----GTNSTCYQMLFDALDLSIYF	:	252										
P23433	:	EVSPSN--NGQYSQFLDHAMENVYMALAAAGLQDKIKVITATYSGLLANTYPPKDSIFRFRKSFINPIIEFLARNMLPLLANTYPPYFCHIYNTVDVPLSYALFNQO----GTNSTCYQMLFDALDLSIYF	:	252										
P52399	:	EVSPIN--NGQYSQFLDHAMENVYMALAAAGLQDKIKVITATYSGLLANTYPPKASIFRGEFNSFINPIIQFLAQNNLPLLANTYPPYFVHISNTADVPLSYALFTQR----CKNSACYQMLFDALDLSIYF	:	248										
P23547	:	EVSPIN--NGQYAPFVAPAMQNVYMALAAAGLQDKIKVITATYSGLLANTYPPKDSIFRGEFNSFINPIIQFLVQHNLPLLANTYPPYFCHIYNTADVPLSYALFTQO----EANTFACYQMLFDALDLSIYF	:	248										
P27666	:	EISPVTCISYLTSLTPAMVNIYKAIKRNRIKVSISVDMTLIGNSYPPSQSFRNDARWFTDPIVCFRLDTRAPLLVNIYPPYFYSYGNPQGISLPPYSLFTAPNWWWQDCSRQYRMLFDALDLSIYF	:	257										
P23431	:	EISPVTCISYLTSLTPAMVNIYKAIKRNRIKVSISVDMTLIGNSYPPSQSFRNDARWFTDPIVCFRLDTRAPLLVNIYPPYFYSYGNPQGISLPPYSLFTAPNWWWQDCSRQYRMLFDALDLSIYF	:	257										
P23546	:	EISPVTCISYLTSLTPAMVNIYKAIKRNRIKVSISVDMTLIGNSYPPSQSFRNDARWFTDPIVCFRLDTRAPLLVNIYPPYFYSYGNPQGISLPPYSLFTAPNWWWQDCSRQYRMLFDALDLSIYF	:	246										
P52401	:	EISPVTCISYLTSLTPAMVNIYKAIKRNRIKVSISVDMTLIGNSYPPSQSFRNDARWFTDPIVCFRLDTRAPLLVNIYPPYFYSYGNPQGISLPPYSLFTAPNWWWQDCSRQYRMLFDALDLSIYF	:	249										
065360	:	EISPVTCISYLTSLTPAMVNIYKAIKRNRIKVSISVDMTLIGNSYPPSQSFRNDARWFTDPIVCFRLDTRAPLLVNIYPPYFYSYGNPQGISLPPYSLFTAPNWWWQDCSRQYRMLFDALDLSIYF	:	249										
Q01413	:	EISPVTCISYLTSLTPAMVNIYKAIKRNRIKVSISVDMTLIGNSYPPSQSFRNDARWFTDPIVCFRLDTRAPLLVNIYPPYFYSYGNPQGISLPPYSLFTAPNWWWQDCSRQYRMLFDALDLSIYF	:	249										
P23432	:	AVEKAGCFNVEIIVSSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	346										
P23433	:	AVEKAGCFNVEIIVSSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	346										
P52399	:	AVEKAGCFNVEIIVSSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	342										
P23547	:	AVEKAGCFNVEIIVSSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	342										
P27666	:	ALERSGGASVGIWVSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	370										
P23431	:	ALERSGGASVGIWVSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	365										
P23546	:	ALERSGGASVGIWVSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	359										
P52401	:	AMERTGGGSVGIWVSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	363										
065360	:	AMERTGGGSVGIWVSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	363										
Q01413	:	AMERTGGGSVGIWVSEGWPSSEGNSSAATIEMAQTYRNLVHVKGGACTPKRPGKVEIYTLFAMFDEENKMGCEVTEKHFGLFYFPMIAKYQLNF-----	:	360										

Figura III.3. Alinhamento das seqüências de aminoácidos dos conjuntos da PR2. As seqüências com nomes em amarelo correspondem ao conjunto A, e em azul, B. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

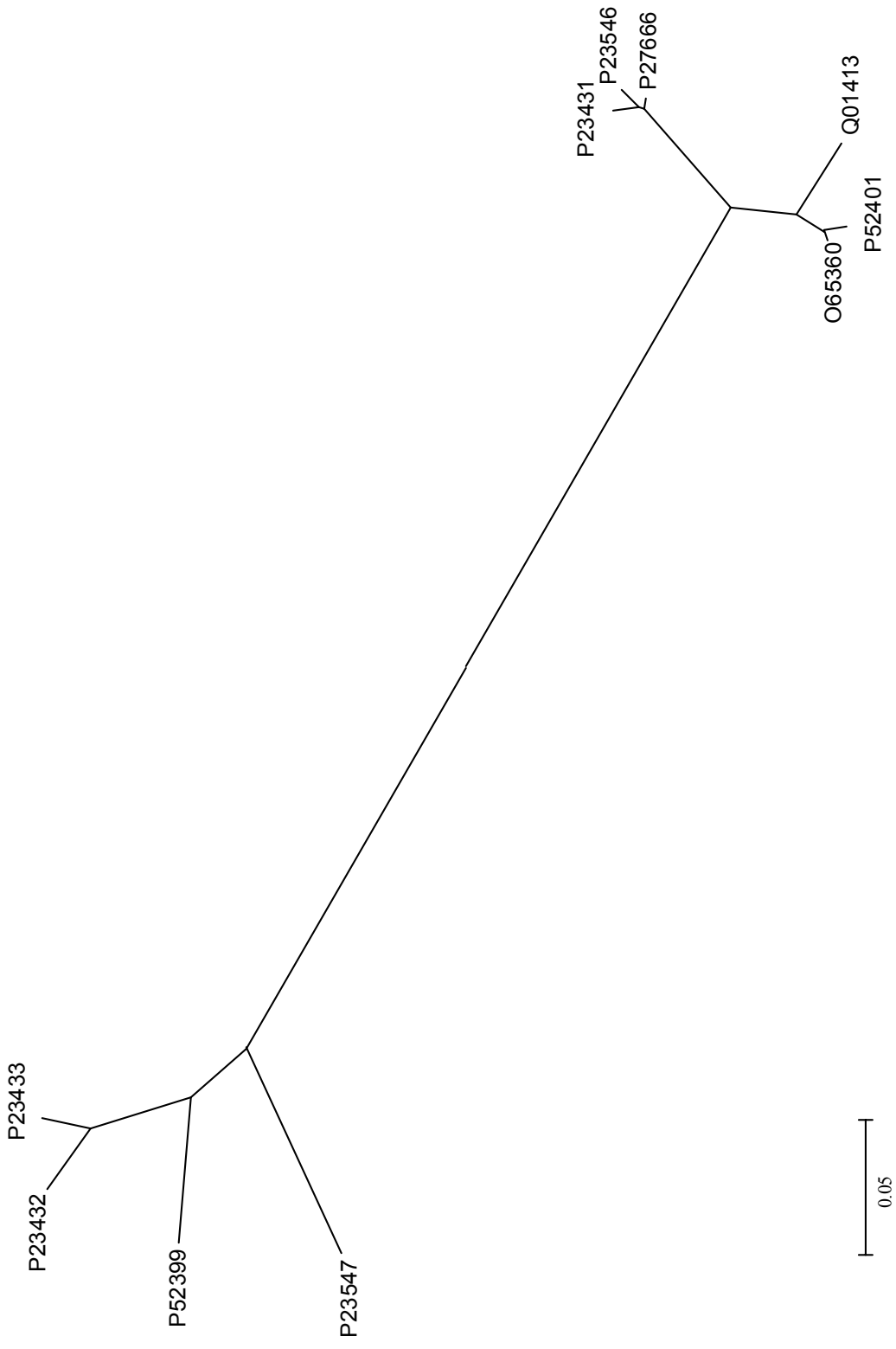


Figura III.4. Árvore filogenética das seqüências de PR 2 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III.3
Resultados obtidos para os conjuntos de PR2 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	4	1038 pb	37,01%	850 (81,89%)	0,21183	346	5 (1,45%)	0,362	M1	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13
B	6	1113	41,47%	851 (76,46%)	0,26579	371	0 (0,00%)	0,134	M1	M10

Ídem Tabela III.2

Para a PR 3 foram formados dois conjuntos: um com as 13 seqüências descritas por Kawabe *et al* (1997) num estudo sobre o polimorfismo nucleotídico na quitinase ácida ChiA de *Arabidopsis thaliana*; e outro, com cinco seqüências reunidas na busca por similaridade de seqüência, usando o programa SIB BLAST. Os dados obtidos estão apresentados nas Figuras III.5 e III.6 e na Tabela III.4. As seqüências de *Arabidopsis thaliana* agruparam-se compactamente entre si, tendo o alinhamento envolvido de 800 a 900 pb, sendo que neste grupo, 42 códons mostraram, uniformemente, indicações de seleção positiva, porém para apenas três classes de distribuição (M3, M8 e M12). De maneira mais ou menos arbitrária, adotando uma posição conservadora, decidiu-se considerar como apresentando indicações fortes de seleção positiva apenas aqueles códons em que havia valores significativos em pelo menos cinco dos modelos testados. Partindo deste critério, apenas o códon 244 fornece indicações deste tipo de seleção. O ω médio foi de 0,261. As seqüências do grupo B apresentaram-se bastante variáveis (apenas 69% de sítios constantes, contra 88% para as do grupo A, as proporções de GC sendo de, respectivamente, 47% e 46%) e 31 códons apresentaram indicação de seleção positiva em pelo menos cinco modelos de distribuição diferentes (M3, M6, M8, M11 e M13), sendo o ω médio obtido de 0,340.

III.2.4. PR 4

Apenas um subgrupo foi formado em PR 4. Nele estão reunidas duas seqüências de tomate (*Lycopersicon esculentum*) e duas de tabaco (*Nicotiana tabacum*). O alinhamento dessas seqüências é apresentado na Figura III.7 e a árvore filogenética correspondente, na Figura III.8. Nesse caso, curiosamente, uma das seqüências de *L. esculentum* agrupou-se com as de *N. tabacum* e não com a outra de sua espécie. O alinhamento incluiu 441 pb, dos quais 81% era constante, a proporção de GC sendo de 47%. Esse conjunto de seqüências não apresentou qualquer sinal de seleção positiva, e o modelo ao qual melhor corresponde foi o M1. O ω médio obtido foi de 0,203.

PR 3

	*	20	*	40	*	60	*	80	*	100	*	120
022074	:	MTNINLKHVIYLLFFISCF	LTSPDTSRGGIAIYWQSGNEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKYCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022076	:	MTNINLKHVIYLLFFISCF	LTSPDASRGGLAIYWGDLGIEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKYCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022073	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022072	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022066	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022068	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
P19172	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022065	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022070	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022069	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022071	:	MTNMTLRKHVIYLLFFISCS	LSKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
022075	:	MTNINLKHVIYLLFFISCF	VTKPSDASRGGLAIYWGQNGEGNLSATCATGRYAYVMVAVLVKFCMNGQTPELNLAGHCNPAANTCTHFGSQVKDCCSRGRIKAMLSLGGCI	NYISGSK	:	120						
023803	:	MSSIKQIPAILWALACHS	CTMTAAQNGC-CASGLCCSKYCYCCTTRAYCGDC-----	CKQPCYSSSTPTTPSGGGC	ASVQSLVTNAFFNCGILNQAGSCCAG	:	98					
023804	:	MSSLGQIPAILWALACHS	CTMTAAQNGC-CASGLCCSKYCYCCTTRAYCGDC-----	CKQPCYSSSTPTTPSGGGC	ASVQSLVTNAFFNCGILNQAGSCCAG	:	96					
023806	:	MSSLTOPFPAILWALACHS	CTMTAAQNGC-CASGLCCSKYCYCCTTRAYCGDC-----	CKQPCYSSSTPTTPSGGGC	ASVQSLVTNAFFNCGILNQAGSCCAG	:	95					
023805	:	MSSLGQIPAILWALACHS	CTMTAAQNGC-CASGLVCCSYCYCNGDAYSCKC-----	CQQPCYSSST-----	CSC--AVSVQSLVTDAFFNCGILNQAGSCCAG	:	90					
09XFW7	:	MSSFGPIPAILWALACHS	STLVVAQNGC-CASNLCCSFFCFCCGSDAYSCKC-----	CRGPCRSPTSS-----	GCG-----SVSSLVTDAPFNWRIINQASASACAG	:	89					
022074	:	DAKVVADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKLTSHRGRKVIYLTGAPQCFPPDLMGMAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022076	:	DAKVVADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKLTSHRGRKVIYLTGAPQCFPPDLMGMAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022073	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022072	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022066	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022068	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
P19172	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022065	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022070	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022069	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022071	:	DAKVIADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
022075	:	DAQVVADYLWNNFLGCKSS	RRPLGDAVLDCIDFNIELGSPQHWDLARSLSKFSHRGRKVIYLTGAPQCFPPDLMGSAALNTRRFDYVWIQFYNNPPCSYTS	SGNTQNL	FDSWNRKWTTSITA	:	240					
023803	:	-----KSYFTRSAFLNALG	WYYPQFKGCSDDTKREVAFAFHVHTGCHFCYIEELAKSTYQSSAQWPCNPMKQYTCRGPQITWNYNYGAAKSKICFDGINAPETVAN-NPVI	AFKTA	:	213						
023804	:	-----KSYFTRSAFLNALG	WYYPQFKGCSDDTKREVAFAFHVHTGCHFCYIEELAKSTYQSSAQWPCNPMKQYTCRGPQITWNYNYGAAKSKICFDGINAPETVAN-NPVI	AFKTA	:	211						
023806	:	-----KSYFTRSAFLNALG	WYYPQFKGCSDDTKREVAFAFHVHTGCHFCYIEELAKSTYQSSAQWPCNPMKQYTCRGPQITWNYNYGAAKSKICFDGINAPETVAN-NPVI	AFKTA	:	210						
023805	:	-----KRYFTRSAFLNALG	WYYPQFKGCSDDTKREVAFAFHVHTGCHFCYIEELAKSTYQSSAQWPCNPMKQYTCRGPQITWNYNYGAAKSKICFDGINAPETVAN-NPVI	AFKTA	:	205						
09XFW7	:	-----KRYFTRAAFLSALR	RYYPQFG-SGSSDVVRRREVAFAFHVHTGCHFCYIEELAKSTYQSSAAAFPCNPSKQYTCRGPQITWNYNYIPACRSICFCFDGINAPETVAN-NAV	TAFKTA	:	203						

Figura III.5. Alinhamento das seqüências de aminoácidos dos três conjuntos da PR3. As seqüências com nomes em amarelo correspondem ao conjunto A, e em azul, B. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

PR 3

```

022074 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILATV : 300
022076 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILATV : 302
022073 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022072 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022066 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022068 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
P19172 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022065 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSIVASV : 302
022070 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSISASV : 302
022069 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022067 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022071 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
022075 : QKIFLGLPAAPEAAGSCSYIPDPVLTSQLPTLKSRKRYGCVMLWSKFWDKMGYSSSSILASV : 302
023803 : FWFMMNVHSRIISCGFCSTIRAVNGECGGENTPAVMARVEYTYQCKQLGWSPCNNLSC : 275
023804 : FWFMMNVHSRIIVSCQFCSTIRAVNGECGGENTPAVMARVEYTYQCKQLGWSPCNNLSC : 273
023806 : FWFMMNVHSRIIVSCQFCSTIRAVNGECGGENTPAVMARVEYTYQCKQLGWSPCNNLSC : 272
023805 : FWFMMNVHSRIINSRCFCATIRAINSTECNCGNTCAVMSRVQLYROYCNQFCVSPCNNLSC : 267
09XF07 : FWFMMNVHSVIIVNCGFCASIRAINCTECNCGNSAAVTARVEYTYQCKQLGWSPCNNLSC : 265

```

Figura III.5. (continuação). Alinhamento das seqüências de aminoácidos dos três conjuntos da PR3. As seqüências com nomes em amarelo correspondem ao conjunto A, e em azul, B. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

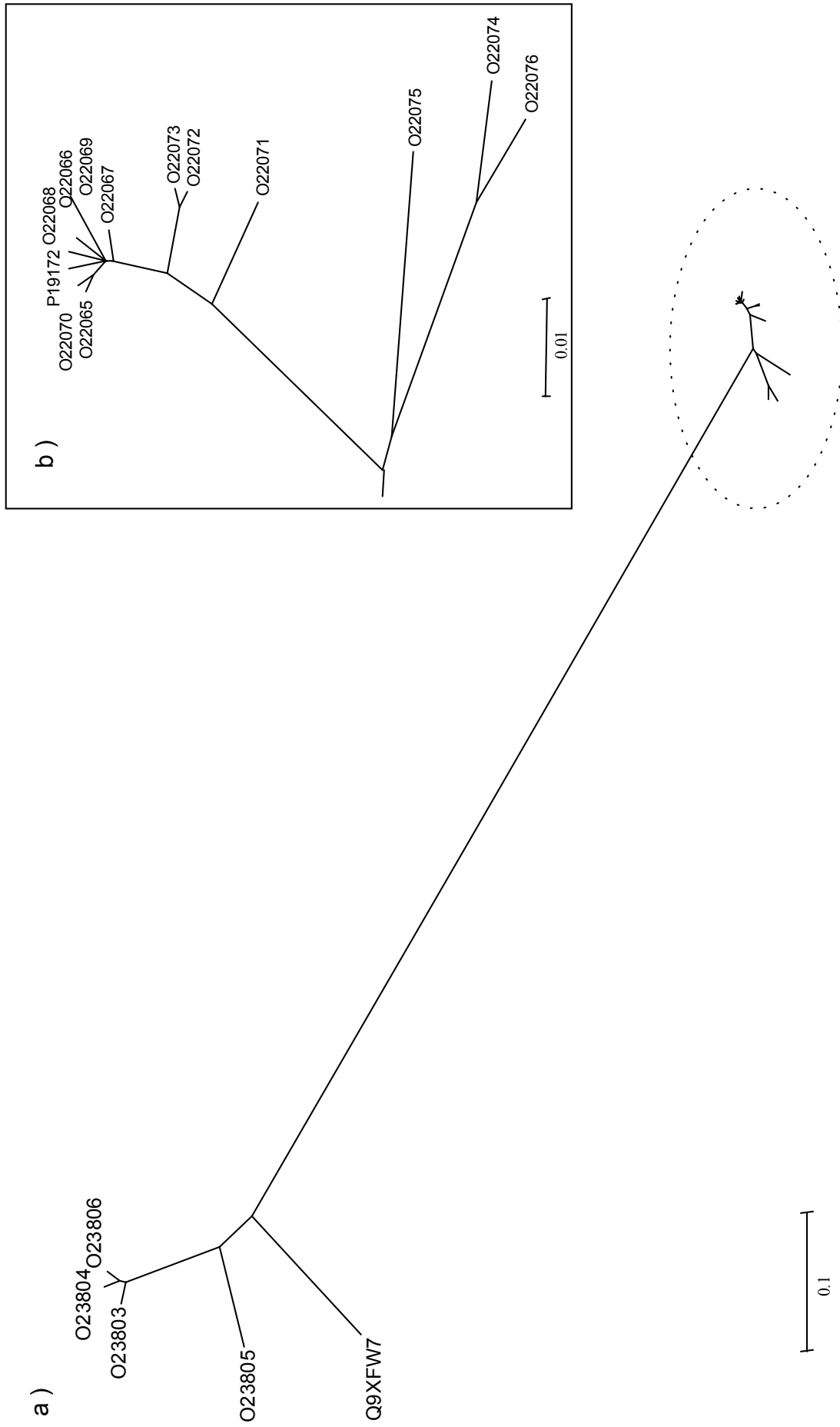


Figura III.6. a) Árvore filogenética das seqüências de PR 3 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*. b) detalhe do ramo circulado.

Tabela III 4
Resultados obtidos para os conjuntos de PR3 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados ω	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	13	906 pb	45,98%	796 (87,86%)	0,13669	302	1 (0,33%)	0,261	M1	M3, M5, M6, M8, M9, M10, M11, M12, M13
B	5	825 pb	46,82%	566 (68,61%)	0,40033	275	31 (11,27%)	0,340	M5	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13

Ídem Tabela III.2

PR 4

```

*          20          40          60          80
P29062 : MERVNNYKLCVALLII SMVVMAMAAAQSATNVRSTYHLYNPQINWDLRAASAFCATWDADKPLAMRQKYGWTAFCGPAGE : 80
Q40558 : MERVNNYKLCVALLII SMVVMAMAAAQSATNVRSTYHLYNPQINWDLRAASAFCATWDADKPLAMRQKYGWTAFCGPAGE : 80
P29063 : MERVNNYKLCVALLIIMSVMMAMAAAQSATNVRSTYHLYNPQINWDLRAASAFCATWDADKPLAMRQKYGWTAFCGPAGE : 80
P32045 : MERVN--KLCVAFFVINMMMAVAAAQSATNVRATYHLYNPQINWDLRTASVYCATWDADKPLEWRRRYGWTAFCGPAGE : 78

*          100         120         140
P29062 : RGQDSCGRCLRVNTGTGTQTTVRIVDQCSNGGLDLDVNVFNQLDTNGVGYQQGHLTVNYEFVNCND : 147
Q40558 : RGQVSCGRCLRVNTGTGTQTTVRIVDQCSNGGLDLDVNVFNQLDTNGVGYQQGHLTVNYEFVNCND : 147
P29063 : RGQDSCGRCLRVNTGTGTQATVRIVDQCSNGGLDLDVNVFNQLDTNGLGYQQGHLTVNYEFVNCND : 147
P32045 : TGQASCGRCLRVNTGTGTQETVRIVDQCRNGGLDLDVNVFNRLDTNGLGYQRGNLNVNYEFVNC-- : 143

```

Figura III.7. Alinhamento das seqüências de aminoácidos da PR4. As seqüências com nomes em rosa correspondem ao conjunto A. Alinhamento realizado com ClustalX e editado com GeneDoc.

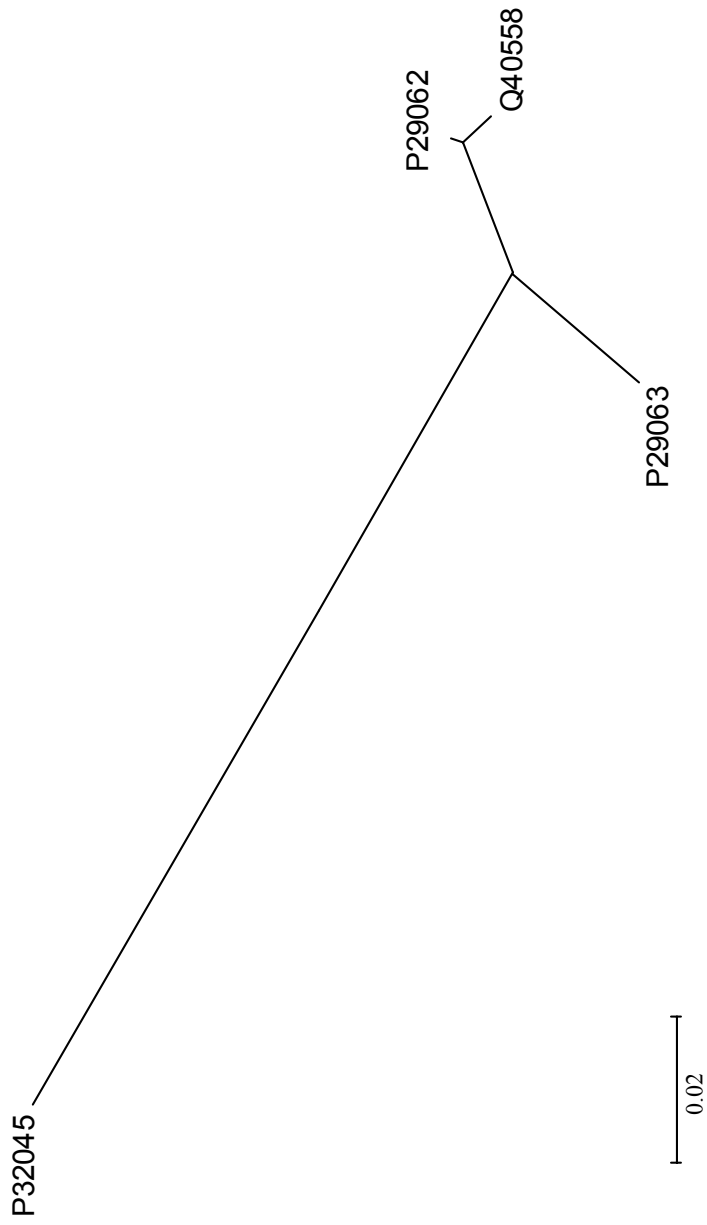


Figura III.8. Árvore filogenética das seqüências de PR 4 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

III.2.5. PR 5

As proteínas semelhantes à taumatina foram amplamente descritas para diversos organismos e estão muito bem representadas e classificadas nos bancos de dados, tanto que cinco grupos distintos puderam ser formados. Os resultados alcançados estão apresentados nas Figuras III.9 e III.10 e na Tabela III.5. As seqüências alinhadas envolveram de 500 a 750 pb, e a proporção de GC de 45% a 66%. As seqüências do grupo C apresentaram 48% de sítios constantes, enquanto que as do grupo B variaram muito menos (90% de sítios invariáveis). Três conjuntos de seqüências (A, B e C) demonstraram ser alvo de seleção positiva. Não menos do que 50 códons do conjunto C apresentaram indicações de seleção positiva, sendo que destes, 28 apresentaram valores significativos para oito modelos, oito para sete deles, e 14 em pelo menos cinco. O ω médio foi relativamente alto (0,342). Os conjuntos A e B apresentaram 12 e 19 códons, respectivamente, com evidências de seleção positiva. O conjunto C apresentou o menor valor de AIC para o modelo M5. Todos os outros conjuntos se adequaram melhor ao modelo M1.

III.2.6. PR 6

Um inibidor de proteinase de tabaco descrito por Van Loon e Van Strien (1999) foi utilizado como palavra-chave na busca de seqüências relacionadas à família PR 6. Das seqüências obtidas, apenas um conjunto de seqüências foi formado com clareza, enquanto que outros quatro se revelaram muito diferentes da seqüência proposta, sendo retirados da análise. O alinhamento e as relações filogenéticas observadas estão apresentados nas Figuras III.11 e III.12. O alinhamento incluiu 321 pb, a proporção de GC foi de 37%, e 90% dos sítios se mostraram constantes. Entre os sítios variáveis, para vinte deles foram observadas indicações significativas de seleção positiva nos modelos M4 e M12. Porém, apenas para seis códons (posições n^{os} 46, 70, 85, 99, 102 e 105) observou-se valores significativos para pelo menos cinco modelos. Especialmente sugestivos foram os resultados para os códons 46 e 105, nos quais valores significativos ocorreram para todos os 10 modelos que permitem $\omega > 1$ (M2 – M6; M8 – M13). O valor médio de ω foi bastante alto: 0,763.

PR 5

	*	20	*	40	*	60	*	80	*	100	*	120	*	140
P50695	:	-MAT--SSAV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CFQLNFKQSSNINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 122
P50696	:	-MAT--SSAV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CFQLNFKQSSNINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 122
P50697	:	-MAT--SSAV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CFQLNFKQSSNINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 122
P50698	:	-MAT--SSTV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CQQLDCCQQTWTINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 122
Q9S776	:	-MAT--S--AV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				AFALCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 124
Q923N0	:	-MAT--S--AV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				AFALCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 124
P32938	:	-MST--S--AV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				AFALCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 124
P27357	:	-MAT--S--AV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 124
Q94F70	:	-MAT--S--AV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 124
Q9F5C7	:	-MAT--S--AV--LFLLLAVFAACASAAATFTITNNCGFTVWPACIPVGC-----				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 124
Q82546	:	MKSQVAPRF--LALVFFSCAHLA--KLTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 137
Q9SMH2	:	MKTLALYGL--LALVFFSCAHLA--KLTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 136
P50694	:	MKTLVWVLS--LSTLISF--SCAHLA--KLTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 136
Q803Z7	:	-MKFEALIGL--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 135
P12670	:	-MGT--LTSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
P50702	:	-MGT--LRSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
Q9FT35	:	-MGT--LRSFL--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 127
P50701	:	-MAY--LRSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
P14170	:	-MGM--LRSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
Q40529	:	-MEY--LRSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
Q94IN9	:	-MAN--LTSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
Q9ARQ0	:	-MGT--LRSF--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 126
P25871	:	-MSH--LITL--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 127
Q40428	:	-MSH--LITCV--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 127
P50703	:	-MSH--LITCL--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 127
Q9M3X2	:	-MSH--LITCL--VLFVLSSEHGVYSA--KFTFTNNCPNIVWPCTLTDRKPOLSLTDFELASKASVDAAPSPW--CRFWGRTKCSDDAAAKGFTCEADCCSGQVACMGACAVPPATLVEITLAAAGCQDYDVSIVLWDCFN				CFELCSQTSISSINVPACTSACRIMWRTCCSRFN--CRSCATCDCA--GALSCTLSGQ-P--ATLARYTTCG--SQDFYDISVIDGFN								: 127

Figura III.9. Alinhamento das seqüências de aminoácidos dos cinco conjuntos da PR5. As seqüências com nomes em amarelo correspondem ao conjunto A; azul, B; rosa C; bege, D, e verde, E. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

PR 5

	*	160	*	180	*	200	*	220	*	240	*	260	*																																																																																										
P50695	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	MC	-	P	---	DA	Y	H	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	169																																																												
P50696	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	MC	-	P	---	DA	Y	H	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	169																																																												
P50697	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	MC	-	P	---	DA	Y	H	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	169																																																												
P50698	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	MC	-	P	---	DA	Y	H	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	169																																																												
Q9S776	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	MA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
Q9Z8N0	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	MA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
P32937	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
CAB99485	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
P32938	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
P27357	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
Q94F70	:	LAMD	F	S	---	CST	-	GV	AL	K	C	R	D	A	---	SC	-	PP	---	PQ	AY	Q	H	P	N	D	---	VA	TH	AC	M	G	-	NS	NY	Q	I	T	F	C	P	---	173																																																												
Q9F8C7	:	LPHS	V	A	P	-	Q	G	T	C	E	K	P	S	S	C	P	A	M	V	M	K	V	C	P	A	P	I	Q	V	K	A	D	G	S	V	I	S	C	K	S	A	C	L	A	F	D	S	K	Y	C	C	T	P	P	N	T	P	E	T	C	P	T	E	S	E	I	F	E	K	O	C	P	Q	A	Y	S	A	Y	D	D	K	N	S	T	I	T	C	S	G	-	G	D	Y	I	T	F	C	P	---	246		
082546	:	LPHS	V	A	P	-	Q	G	T	C	E	K	P	S	S	C	P	A	M	V	M	K	V	C	P	A	P	I	Q	V	K	A	D	G	S	V	I	S	C	K	S	A	C	L	A	F	D	S	K	Y	C	C	T	P	P	N	T	P	E	T	C	P	T	E	S	E	I	F	E	K	O	C	P	Q	A	Y	S	A	Y	D	D	K	N	S	T	I	T	C	S	G	-	G	D	Y	I	T	F	C	P	---	245		
Q9SMH2	:	LPHS	V	A	T	-	R	G	T	G	C	K	A	T	S	C	P	A	M	V	M	K	V	C	P	A	P	I	Q	V	K	A	D	G	S	V	I	S	C	K	S	A	C	L	A	F	D	S	K	Y	C	C	T	P	P	N	T	P	E	T	C	P	T	E	S	E	I	F	E	K	O	C	P	Q	A	Y	S	A	Y	D	D	K	N	S	T	I	T	C	S	G	-	A	D	Y	I	T	F	C	P	---	243		
P50694	:	LPHS	V	T	P	-	Q	G	T	G	C	K	A	T	S	C	P	A	M	V	M	K	V	C	P	A	P	I	Q	V	K	A	D	G	S	V	I	S	C	K	S	A	C	L	A	F	D	S	K	Y	C	C	T	P	P	N	T	P	E	T	C	P	T	E	S	E	I	F	E	K	O	C	P	Q	A	Y	S	A	Y	D	D	K	N	S	T	I	T	C	S	G	-	G	D	Y	I	T	F	C	P	---	245		
080327	:	LPHS	V	T	P	-	R	G	T	G	C	K	A	T	S	C	P	A	M	V	M	K	V	C	P	A	P	I	Q	V	K	A	D	G	S	V	I	S	C	K	S	A	C	L	A	F	D	S	K	Y	C	C	T	P	P	N	T	P	E	T	C	P	T	E	S	E	I	F	E	K	O	C	P	Q	A	Y	S	A	Y	D	D	K	N	S	T	I	T	C	S	G	-	G	D	Y	I	T	F	C	P	---	244		
P50702	:	IPHT	F	A	P	T	R	S	A	G	K	H	A	I	H	C	T	A	M	I	N	G	E	C	P	R	A	L	K	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	247
Q9FT35	:	IPHT	F	A	P	T	R	S	A	G	K	H	A	I	H	C	T	A	M	I	N	G	E	C	P	R	A	L	K	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	248
P14170	:	IPHT	F	A	P	T	R	S	A	G	K	H	A	I	H	C	T	A	M	I	N	G	E	C	P	R	A	L	K	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	246
040529	:	IPHT	F	A	P	T	R	S	A	G	K	H	A	I	H	C	T	A	M	I	N	G	E	C	P	R	A	L	K	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	246
Q94JN9	:	IPHT	F	A	P	T	R	S	A	G	K	H	A	I	H	C	T	A	M	I	N	G	E	C	P	R	A	L	K	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	246
Q9ARQ0	:	IPHT	F	A	P	T	R	S	A	G	K	H	A	I	H	C	T	A	M	I	N	G	E	C	P	R	A	L	K	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	246
P25871	:	IPHS	F	G	T	R	P	G	K	C	H	I	H	C	T	A	M	I	N	G	E	C	P	G	S	L	R	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	246		
Q40428	:	IPHS	F	G	T	R	P	G	K	C	H	I	H	C	T	A	M	I	N	G	E	C	P	G	S	L	R	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T	DE	VA	K	---	251		
P50703	:	IPHS	F	G	T	R	P	G	K	C	H	I	H	C	T	A	M	I	N	G	E	C	P	G	S	L	R	V	P	G	-	C	M	N	P	C	T	I	F	G	Q	Y	C	C	T	Q	---	PC	PT	L	S	K	F	F	K	K	C	P	A	Y	S	T	P	O	D	D	P	T	S	T	T	C	P	G	S	T	N	Y	V	V	F	C	P	M	G	V	A	D	---	PN	F	L	EM	P	AS	T							

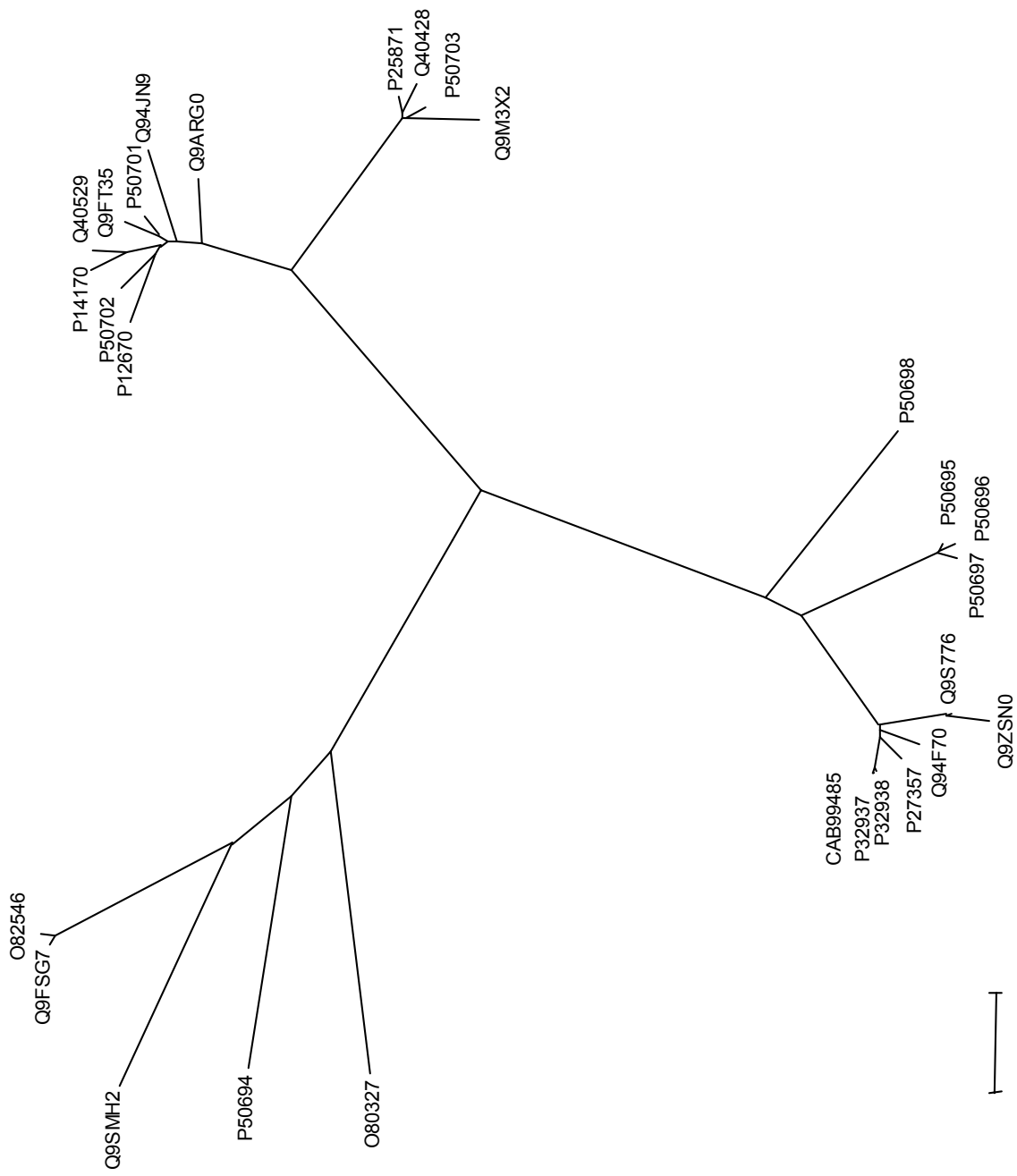


Figura III.10. Árvore filogenética das seqüências de PR 5 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 5
Resultados obtidos para os conjuntos de PR5 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	4	507 pb	66,32%	430 (84,81%)	0,21153	169	12 (7,10%)	0,173	M1	M5, M6, M9, M10, M12, M13
B	7	519 pb	65,07%	466 (89,79%)	0,13661	173	19 (10,98%)	0,135	M1	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13
C	5	741 pb	50,91%	355 (47,91%)	0,88183	247	50 (20,24%)	0,342	M5	M3, M4, M5, M6, M9, M10, M11, M12, M13
D	8	744 pb	46,90%	519 (69,76%)	0,58079	248	0 (0,00%)	0,136	M1	
E	4	753 pb	45,02%	625 (83,00%)	0,20309	251	0 (0,00%)	0,095	M1	

Ídem Tabela III.2

PR 6

```

*           20           40           60
Q07459 : MESKFAHIIVFFLLATSFETLLARKE SDGPEVIELLKEFEFCNGKQFMPPELIGVPTKLAKEIIEK : 64
Q43648 : MESKFAHIIVFFLLATSFETLLARKE SDGPEVIELLKEFEFCNGKQFMPPELIGVPTKLAKEIIEK : 64
P08454 : MESKFAHIIVFFLLATSFETLLARKE SDGPEVIELQKEFEFCNGKQRMPELIGVPTKLAKEIIEK : 64
Q00783 : MELKFAHIIVFFLLATSFETL MARKE SDGPEVIQLLKEFEFCNGKGLRMPPELIGVPTKLAKEIIEK : 64

*           80           *           100
Q07459 : ENSLI NN VQILLNGSPVAMDY R CNRVRRLF DNILG SVVQIPRVA : 107
Q43648 : ENSLI NN VQILLNGSPVTMDY R CNRVRRLF DNILG SVVQIPRVA : 107
P08454 : ENSLI TNVQILLNGSPVTMDY R SNRVRRLF DNILG DVVQIPRVA : 107
Q00783 : ENSLI SNVHILLNGSPVTLDI R CDRVRRLF DNILG YVVDIPVVG : 107

```

Figura III.11. Alinhamento das seqüências de aminoácidos da PR6. As seqüências com nomes em lilás correspondem ao conjunto A. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

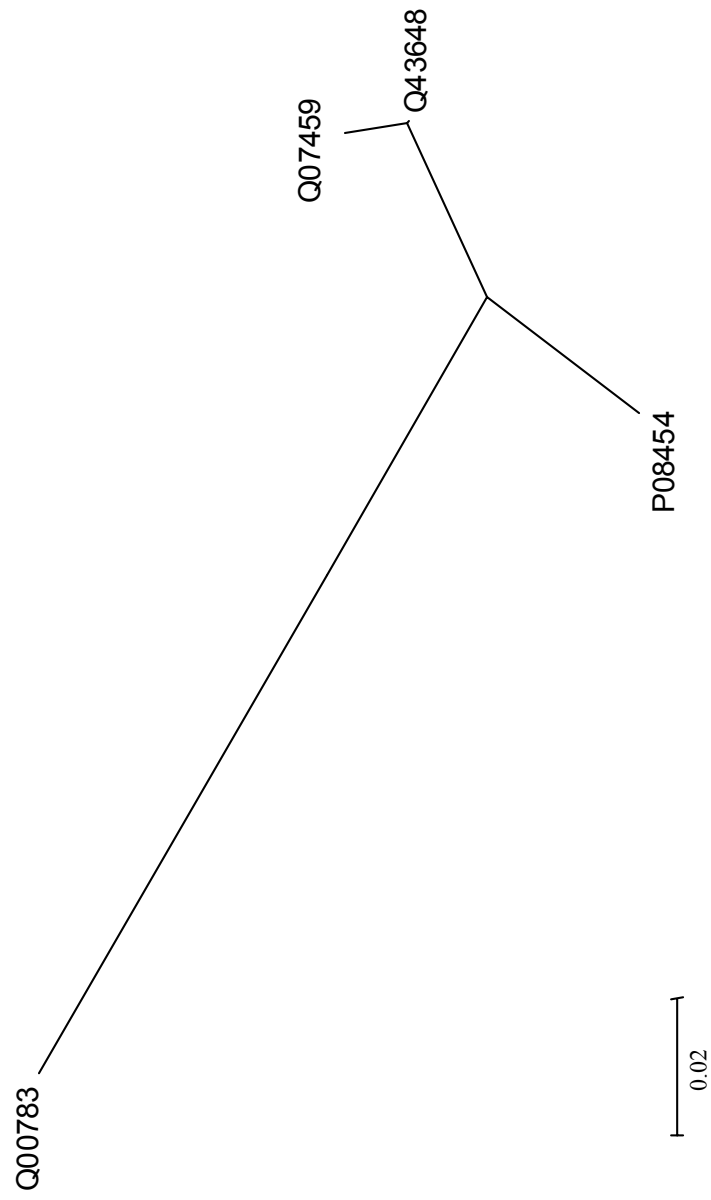


Figura III.12. Árvore filogenética das seqüências de PR 6 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

III.2.7. PR 7

As seqüências de PR 7 (ou P69) foram obtidas de Vera e Conejero (1988) e outros artigos publicados por seu grupo. Essas proteínas só foram claramente descritas em tomate, onde são representadas por seis cópias no genoma. Como uma das seqüências é substancialmente menor, esta foi excluída na formação de um segundo grupo de comparação. Os resultados obtidos são apresentados nas Figuras III.13 e III.14 e na Tabela III.6.

Há uma certa heterogeneidade na comparação das seqüências, além da diferença de tamanho da Q9ZR46. O alinhamento envolveu 2.268 pb, havendo alta variabilidade nessas seqüências (57% quando todas as seis são consideradas, e 65% quando excluída a seqüência de menor tamanho). O conteúdo de GC situou-se em 43%. Os códons com indicações significativas de seleção positiva foram praticamente os mesmos nos dois conjuntos, indicando que as cinco seqüências mantidas no segundo agrupamento foram suficientemente representativas do tipo de evolução que ocorre nas P69 de tomate. Setenta e oito códons foram observados como alvo de seleção positiva significativamente para ambos os conjuntos, sendo que sete são acrescentados quando todas as seqüências são analisadas (conjunto A). O valor médio de ω foi de 0,350.

III.2.8. PR 8

Três conjuntos de PR 8 foram formados. Um com seqüências de *Arabidopsis thaliana* e *Arabis gemmifera* e os dois outros com as seqüências das diversas espécies de *Arabis* utilizadas no artigo de Bishop *et al* (2000). O conjunto B é formado pelas 18 seqüências, considerando toda a informação disponível, enquanto no C as regiões mais fragmentadas foram retiradas, o que implicou na eliminação total de uma seqüência (Q9M7G5). Os resultados estão apresentados nas Figuras III.15 e III.16 e na Tabela III.7. A separação das seqüências não é muito clara (com exceção da Q9M7H4 e Q9M7G9). O alinhamento envolveu de 900 a 1000 pb, o conteúdo de GC entre 51% e 54%, com variação na freqüência de sítios nucleotídicos constantes de 64% a 94%. O conjunto A, com pouca variabilidade, não apontou qualquer códon com seleção positiva. Já com relação aos grupos B e C o número de códons com indicações de tal tipo de seleção é de respectivamente 19 e 21. O códon 191 do conjunto B destacou-se de maneira especial, com nove valores significativos. O ω médio, como esperado, foi praticamente igual nos dois conjuntos (0,393 e 0,394 respectivamente).

PR 7

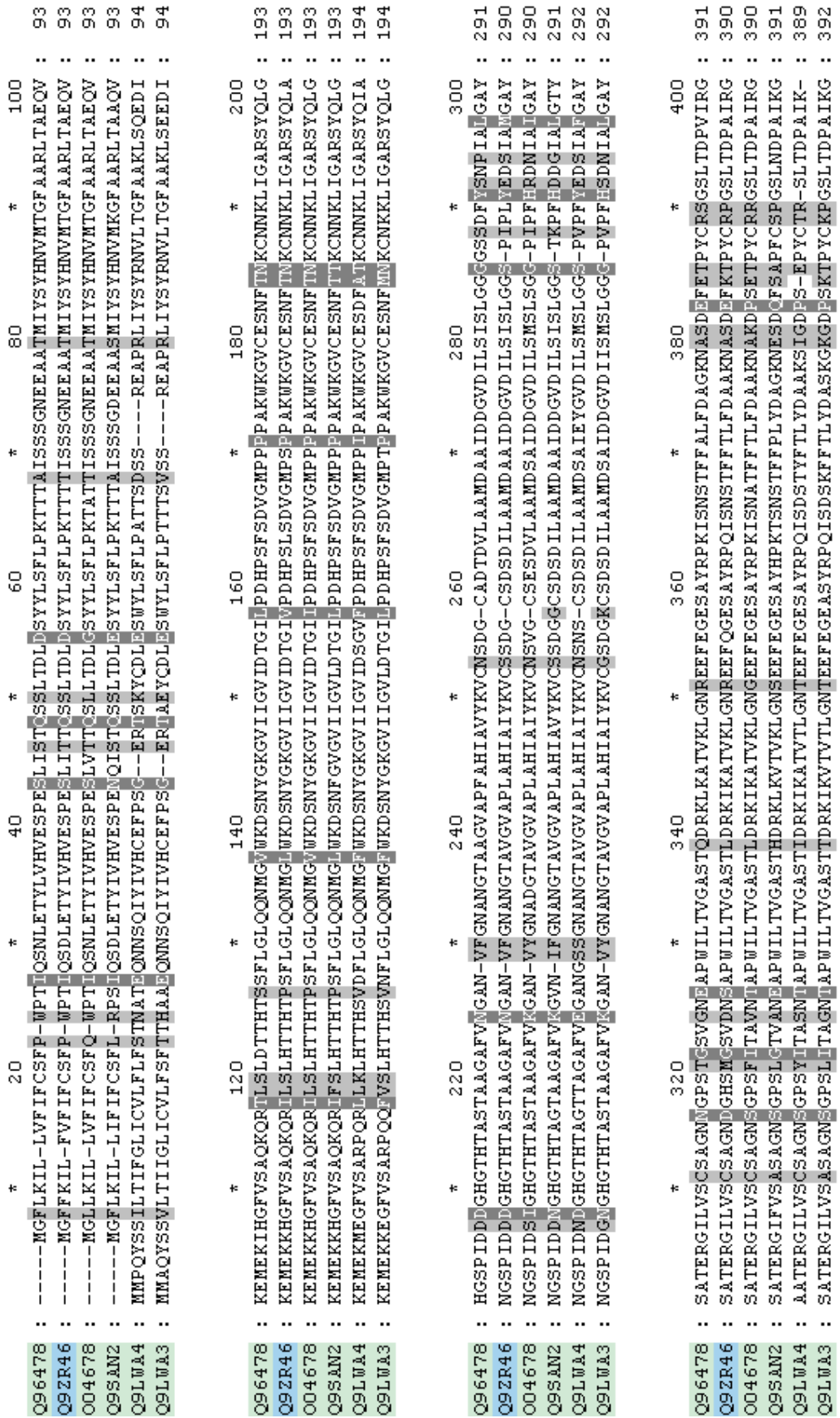


Figura III.13. Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR7. As seqüências com nomes em verde correspondem às presentes nos dois conjuntos (A e B), e em azul àquela presente apenas no conjunto A. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva para ambos, e em azul, os que foram indicados somente para o conjunto A. Alinhamento realizado com ClustalX e editado com GeneDoc.

PR 7

```

Q96478 : KIVICLAGGGVPRVVDKQAVKADAGGVGMIIINQORSVGTKSADAHVIPALDISDADGKILAYMNSTSNPVATITFQGTIIIGDKNAPIVAASFSSRGPSSGA : 491
Q9ZR46 : KIVLCLAFGGVTIVDKQAVKADAGGVGMIIINSPDDGVTKSADAHVLPALDVSDADGKILAYMNSTSNPVATIAFQGTIIIGDKNAPMVAASFSSRGPSSRA : 490
O04678 : KIVLCSALCHVANVDKQAVKADAGGVGMIIINFSQYGVTKSADAHVLPALVWSAADGKILAYMNSTSSPVATIAFQGTIIIGDKNAPMVAASFSSRGPSSRA : 490
Q9SAN2 : KIVLCLRSLSLLRVAQGQSVKADAGGVGMIIINEKQRGVTKSAEAHVLPALDVSNADGKILAYMNSTSNPVASITFHGTVIGDKNAPIVASFSRGPSSVA : 491
Q9LWA4 : KIAICQAG--DVSNIKQRQAVKADAGGVGHIVINHHIYGVTKSADAHVLPGLVWSAADGSKILDYTNISINPIATITIQGTIIIGDKNAPIVAASFSSRGPSSKP : 488
Q9LWA3 : KIVICYPG--VWSKVVKQAVKADAGGVGMIIINLPEDGVTKSADAHVLPALVWSAADGIRILTYTNSISNPTAKITFQGTIIIGDENAPIVASFSRGPSSKPK : 491

Q96478 : SIGILKPDIIIGPGWNILAAAMPTSVDDNKKTKSTFNIISGTSMSCPHLSGVRALLKSTHDPWSPAAIKSAMTTTADTLNLANSPILDERLLPADIYAIGAG : 591
Q9ZR46 : SPGILKPDIIIGPGWNILAAAMPTSVDDNKKTKSTFNIISGTSMSCPHLSGVAALLKSTHDPWSPAAIKSAIMTTTADTLNLANSPILDERLLPADIFATGAG : 590
O04678 : SPGILKPDIIIGPGANILAAAMPTSVDDNKKTKSTFNIISGTSMSCPHLSGVAALLKCTHDPWSPAVIKSAMTTTADTLNLANSPILDERLLPADIYAIGAG : 590
Q9SAN2 : SPGILKPDIIIGPGWNVLAAMPTSVDDNKKTKSTFNIISGTSMSCPHLSGVAALLKSAHPDWSPAAIKSAMTTTADTVNLANSPILDERLISADLFAMGAG : 591
Q9LWA4 : NPGILKPDIIIGPGWNILAAAMPTSVDDNKKTKSTFNIISGTSMSCPHLSGIAALLKSTHDPWSPAAIKSAIMTTTAYTLNLDSSPILDERLLPADIFAIAGAG : 588
Q9LWA3 : SPGILKPDIIIGPGWNILAAAMPTSVDDNKKTKSTFNIISGTSMSCPHLSGVAALLKSTHDPWSPAAIKSAIMTTTAYTLNLANSPILDERLLPADIFAIAGAG : 591

Q96478 : HVNPSRANDPGLVYDTPFEDYYPYLCGLNYTNRQVGNLLQRVWNCSEVKSILEAQLNYPFSFSDYDLGSTPQTYTRVTWVGDAKSSYKVEVASPEALPSK : 691
Q9ZR46 : HVNPSRANDPGLVYDIPFEDYLPYLCGLNYTNRQVGNLLQRVWNCSEVKIILEAQLNYPFCITELGS--RLFERLTI----- : 666
O04678 : HVNPSRANDPGLVYDTPFEDYYPYLCGLKYTDQVGNLIQRVWNCSEVKSILEAQLNYPFSFSDYDLGSTPQTYTRVTWVGDATSSYKVEVASPEGVAIE : 690
Q9SAN2 : HVNPSRASPGLVYDTPFEDYLPYLCGLNYTNREVGVKLVQRVWNCSEVKRIPEQQLNYPFSFSDYDLGSTPQTYTRVTWVGDATSSYKVEIVSPKGVVVK : 690
Q9LWA4 : HVNPSANDPGLVYDTPSEDYFPYLCGLGYTNAQVSSLLRRVWNCLEVNSIPEAQLNYPFSFSDYDLGSTPQTYTRVTWVGDATSSYKVKIASLIGVAWE : 688
Q9LWA3 : HVNPSANDPGLVYDTPSEDYFPYLCGLRYTNAQVSKLLQRVWNCLEVKSIPEAELNYPFSFSDYDLGSTPQTYTRVTWVGDVASSYKVEIASPIGVAIE : 691

Q96478 : LTLR--DNFSSD--QKLYQVTFSKTANSSNTEVIEGFLKWTNRSRSPIALLLIQ : 745
Q9ZR46 : ----- : -
O04678 : VEPSENFSEINQKLYQVTFSKTITSSNTEVIEGFLKWTNRSRSPIAWWSA- : 745
Q9SAN2 : VEPSANFSTINQKLYQVIFTKTITSSNTEVIEGFLKWTNRSRSPIAVRVSA : 746
Q9LWA4 : VVPTENFSEINQKLYQVTFSKTITSSSEVVVVEGFLKWTSTRHSRSPIAWLV- : 743
Q9LWA3 : VVPTENFSEINQKLYQVTFSKTITSSSEVVVVEGFLKWTSTRHSRSPIAWLV- : 746

```

Figura III.13 (continuação). Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR7. As seqüências com nomes em verde correspondem às presentes nos dois conjuntos (A e B), e em azul àquela presente apenas no conjunto A. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva para ambos, e em azul, os que foram indicados somente para o conjunto A. Alinhamento realizado com ClustalX e editado com GeneDoc.

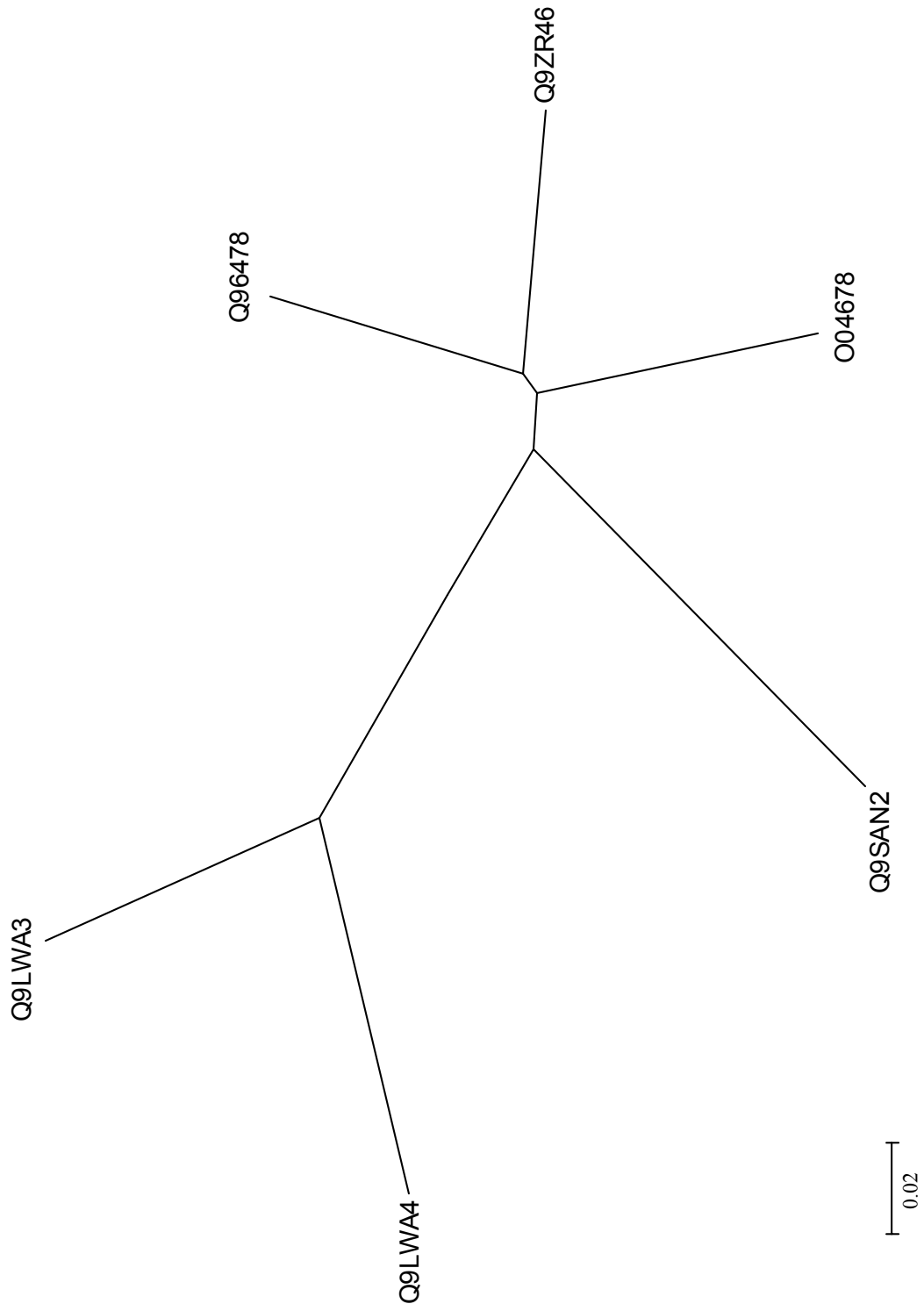


Figura III.14. Árvore filogenética das seqüências de PR 7 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

PR 8

```

260          280          300          320          340
Q9S7J5 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
Q9SXJ3 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
Q9SXJ4 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
Q9SXJ2 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
Q9S838 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
P19171 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
Q9M7F5 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 312
Q9M7H2 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 305
Q9M7G6 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 295
Q9M7H5 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 299
Q9M7F7 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 295
Q9M7G0 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 295
Q9M7H0 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 299
Q9M7H9 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 326
Q9M7G7 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 320
Q9M7H3 : ACIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 289
Q9M7G5 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 311
Q9M7G1 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 300
Q9M7G2 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 306
Q9M7H1 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 300
Q9M7G4 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 302
Q9M7F4 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 312
Q9SXJ1 : AAIWFMHTAQP PKPSCHAVIACQWQPSAADAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 335
Q9M7G9 : AAIWFMHTAQP PKPSCHAVIACQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 310
Q9M7H4 : AAIWFMHT PQSPKPSCHDVITGQWQPSDADAAAGRLP GYGVITWIIINGL ECGRCQD ERVAD RIGFYQRYCNI FGNP GGNLDCYNQSFV NCLLEAAI : 308

```

Figura III.15 (continuação). Alinhamento das seqüências de aminoácidos dos três conjuntos da PR8. As seqüências com nomes em vermelho correspondem ao conjunto A, e em azul, B e C, sendo que a seqüência com nome em azul escuro não está presente no conjunto C. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva para os conjuntos B e C, e em lilás, apenas para o conjunto C. Alinhamento realizado com ClustalX e editado com GeneDoc.

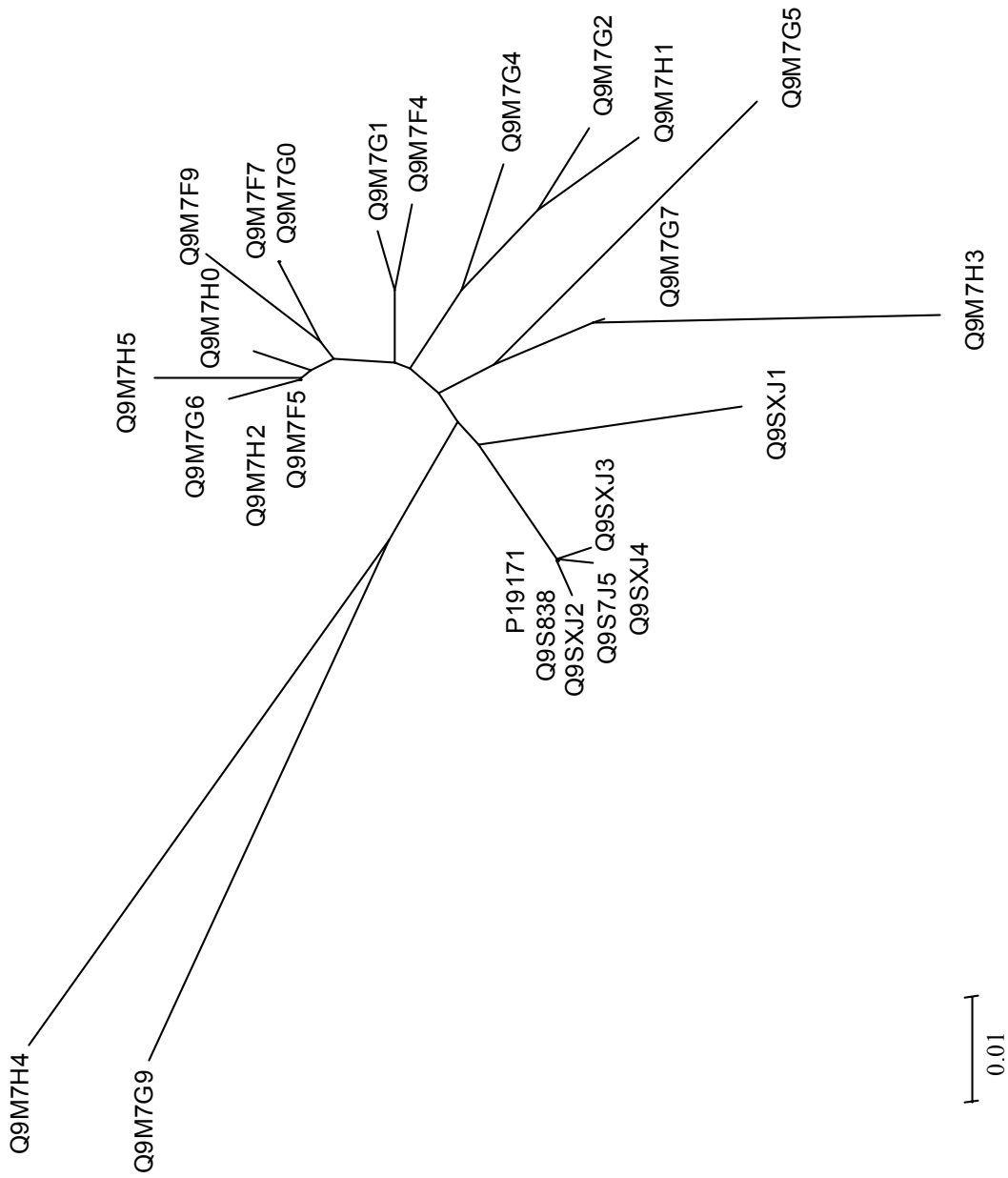


Figura III.16. Árvore filogenética das seqüências de PR 8 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 6
Resultados obtidos para os conjuntos de PR7 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados ω	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	6	2268 pb	42,78%	1293 (57,01%)	0,60497	756	88 (11,64%)	0,350	M3	M3, M4, M5, M6, M8, M9, M11, M12, M13
B	5	2268 pb	42,74%	1479 (65,21%)	0,53927	756	58 (7,67%)	0,349	M3 e M4	M3, M4, M5, M6, M9, M11, M12, M13

Ídem Tabela III.2

Tabela III 7
Resultados obtidos para os conjuntos de PR8 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados ω	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	7	1005 pb	51,48%	948 (94,33%)	0,07143	335	0 (0,00%)	0,226	M0 e M1	
B	18	1014 pb	53,71%	653 (64,40%)	0,40579	338	19 (5,62%)	0,393	M3 e M8	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13
C	17	900 pb	53,98%	660 (73,33%)	0,32836	300	21 (7,00%)	0,394	M3	M3, M4, M5, M6, M9, M11, M12, M13

Ídem Tabela III.2

III.2.9. PR 9

Os dois conjuntos de PR 9 foram formados pelas mesmas seqüências, com a diferença da exclusão, no conjunto B, de uma das seqüências do conjunto. Os resultados estão expressos nas Figuras III.17 e III.18 e na Tabela III.8. A diferença entre a seqüência Q43050 de *Populus kitakamensis* com relação às das outras espécies é ilustrada na Figura III.18. O alinhamento incluiu 1029 pb, o conteúdo de GC foi praticamente idêntico (44%), e a exclusão da seqüência acima indicada condiciona forte aumento na percentagem de sítios constantes (de 75% para 92%). Quando o conjunto A é considerado, observa-se 28 códons em que ocorreram indicações de seleção positiva para pelo menos 5 modelos, mas quando é retirada a seqüência mais divergente (o conjunto B de análise), encontra-se valores significativos para 50 códons em nove dos modelos utilizados. Pode-se concluir que a homogeneidade das seqüências leva a uma maior consistência na detecção da seleção, justificando o nível de exigência na formação dos conjuntos utilizados nesta pesquisa. O ω médio para o conjunto B (0,731) é quase duas vezes maior do que o encontrado para o A (0,399).

III.2.10. PR 10

Três conjuntos bastante distintos foram formados para essas proteínas. Os resultados estão apresentados nas Figuras III.19 e III.20 e na Tabela III.9. Há boa separação entre os conjuntos de seqüências, com maior dissimilaridade para o conjunto B. Entre 477 e 480 pb foram alinhados, o conteúdo de GC variou entre 44% e 48%, e a percentagem de sítios constantes foi maior no conjunto A (85%) do que nos outros dois (48%, 59%). No conjunto A há indicações significativas de seleção positiva para 27 códons com relação a pelo menos seis modelos de distribuição. Já para o conjunto B apenas dois códons (posições 46 e 142) mostraram tal tipo de indicação para, respectivamente, seis e sete modelos. Para o conjunto C foram 11 os códons com indicações de seleção positiva, com especial ênfase para os de número 74, 106 e 129. O ω médio situou-se entre 0,230 e 0,319.

PR 9

```

Q40949 : MHSKAIVAAFFFVWLLGGTLACQLTPFFYDQTCPNVSSIIRDVITETLVSDPRIGASLIRLHFHDCLVMGCDGSLLLDNTDTIVSEKAGGNNNSARG : 100
Q43100 : MHSKAIVEAFFFVWLLRGTLACQLTPFFYDQTCPNVSSIIRDVITETLVSDPRIGASLIRLHFHDCFVMGCDGSLLLDNTDTIVSEKACGNNNSARG : 100
Q40950 : MHSKAIVAAFFFVWLLGGTLAHQLTPFFYDQTCPNVSSIIRNVITETLVSDPRIAASLIRLHFHDCFVMGCDGSLLLDNTDTIESEKAAGNNNSARG : 100
Q43099 : MHSKAIVAAFFFVWLLGGTLAHQLTPFFYDQTCPNVSSIIRNVITETLVCDRRIGASLIRLHFHDCFVMGCDGSLLLDNTDTIESEKAAGNNNSARG : 100
Q43101 : MHSKAIVAAFFFVWLLGGTLAHQLTPFFYDRTCPWNVSSIISNVITETLVSDTRIGASLIRLHFHDCFVMGCDGSLLLDNTDTIESEKAAGNNNSARG : 100
Q43050 : MKLSKLMVALFYAFLVGGPLAYQLTPFFYDETCPYVISIIRPGVIAETLIFDPPRIGASLIRLHFHDCFVMGCDGSLLDKTATIDTEKEAAMNSARG : 100

Q40949 : FEVDRMKALESACPATVSCADILTIAAESVVLAGGPNWTVPLGRRDSTTASRDAANAFLPAPFFLDQLRESFTNVSLNNNTDLVALSGAHTFGRAK : 200
Q43100 : FEVDTMKALESACPATVSCADILTIAAESVVLAGGPNWTVPLGRRDSTTASRDAANAFLPAPFFLDQLRESFTNVSLNNNSDLVALSGAHTFGRAK : 200
Q40950 : FEVDRMKALESACPATVSCADILTIAAESVVLAGGPNWTVPLGRRDSTTASRAAANASLPAPFLPLDQLRESFTNVGLNNNTDLVALSGAHTFGRAQ : 200
Q43099 : FEVDRMKALESACPATVSCADILTIAAESVVLAGGPNWTVPLGRRDSTTASRAAANASLPAPFLPLDQLRESFTNVGLNNNTDLVALSGAHTFGRAK : 200
Q43101 : FEVDRMKALESACPTTVSCADILTIAAESVVLAGGPNWTVPLGRRDSTTASRDAANAFLPAPFFLDQLRESFTNVSLNNNSDLVALSGAHTFGRAQ : 200
Q43050 : FDVDIMKELEGVCPATVSCADILAIAAESVVLAGGPNWPIPLGRRDSLIANRAANAIPGPODTLERLRESFTNVSLNNNTDLVALSGAHTFGRAQ : 200

Q40949 : CSTFDRLYDFNSTGAPDQSLNTLLADLQELCPQGNGSVITDLDLTPDAFSNYYSNLQGNQLQTDQELFSTPGADDVIALVNAFSANQTAFFFES : 300
Q43100 : CSTFDRLYDFNSTGAPDPSLDTTLAALQELCPEGNGSVITDLDLSTPDAFSDYYSNLQGNRLQTDQELFSTPGADDVIALVNAFSANQTAFFFES : 300
Q40950 : CSTFDRLYDFNSTGAPDPSLDPTLAALQELCPQGNRSVITDLDLTPDAFSNYYSNLQGNRLQTDQELFSTPGADDVIALVNAFSANQTAFFFES : 300
Q43099 : CSTFNRLYDFNGTGAPDPTLDPPLAALQELCPQGNDSVITDLDLTPDAFSNYYSNLQGNRLQTDQELFSTPGADDVIALVNAFSANQTAFFFES : 300
Q43101 : CSTFFRLYDFNGTGAPDSTIDPPLEALQKLCPENGNGSVITDLDLVTADAFDSKYYSNLQGNRLQTDQELFSTPGADDVIALVNAFSANQTAFFFES : 300
Q43050 : CRNFLDRLYNFNTGLPDPLDTTLATLQELCPGGNTVLADLDETTPDPGFDNNYFSNLQASKGLLQSDQELFSTPEADDIELVDIFSTDETAFFFES : 300

Q40949 : FAESMIRMGNLSPLTGTEGEIRLNCRVVNANLAGPDSMLVSSI : 343
Q43100 : FVESMIRMGNLSPLTGTEGEIRLNCSVVNANLAGPDSMLVSSI : 343
Q40950 : FAESMIRMGNLSPLTGTEGEIRLNCRVVNANLAGPDSMLVSSI : 343
Q43099 : FVESMIRMGNLSPLTGTEGEIRLNCSVVNANLAGPDSMLVSSI : 343
Q43101 : FVESMIRMGNLSPLTGTEGEIRLNCRVVNANLAGPDSMLVSSI : 343
Q43050 : FVESMIRMGNLSPLTGTEGEIRLNCAVNADLAGKDSVLVSSV : 343

```

Figura III.17. Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR9. As seqüências com nomes em azul claro correspondem aos conjuntos A e B, e seqüência em azul escuro está ausente no conjunto B. Os sítios marcados em azul correspondem aos códons indicados como alvos de seleção positiva no conjunto A e em cinza, no conjunto B. Alinhamento realizado com ClustalX e editado com GeneDoc.

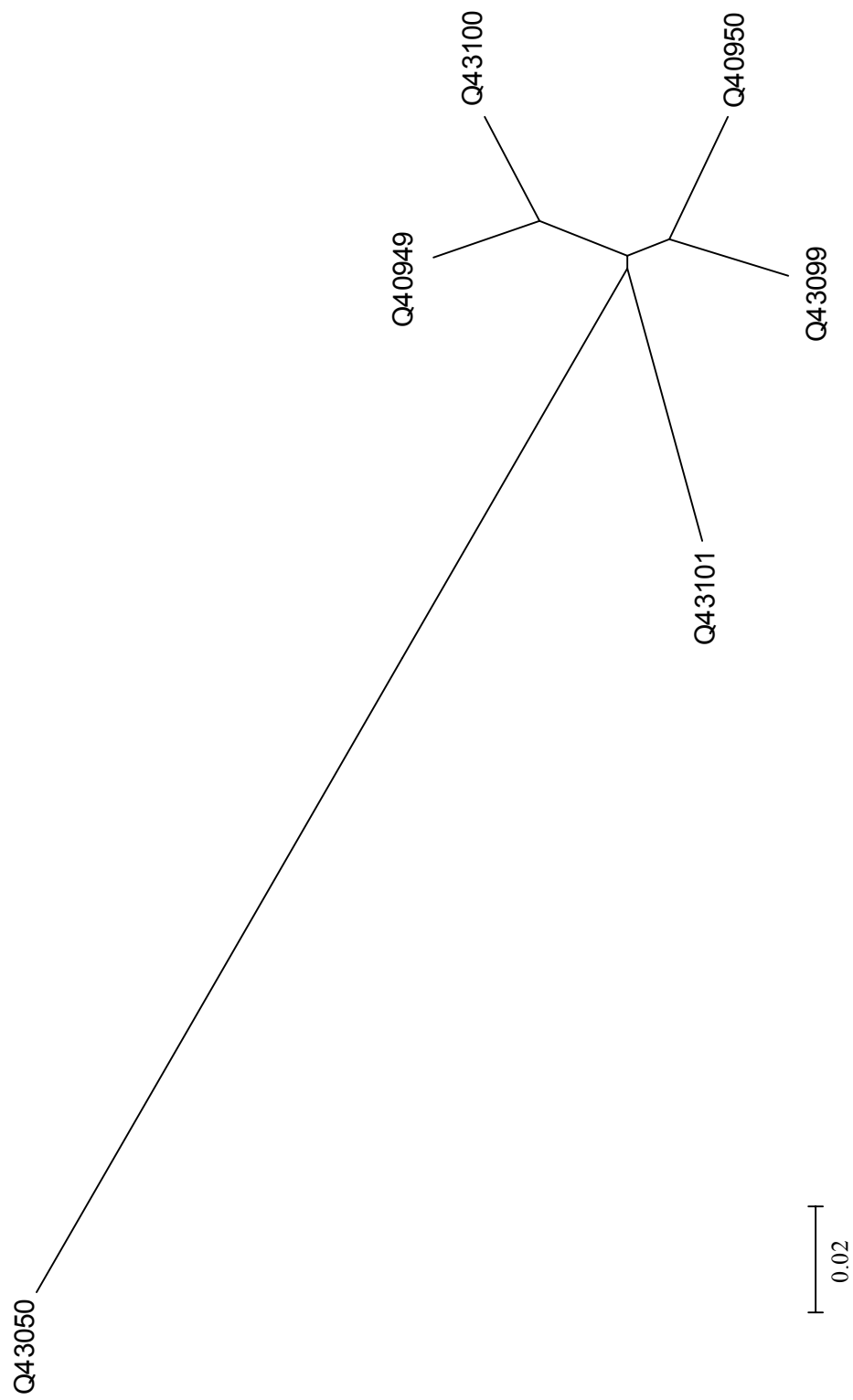


Figura III.18. Árvore filogenética das seqüências de PR 9 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

PR 10

	*	20	*	40	*	60	*	80	*	100	*	120																																																																																																
P43176	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	A	L	G	T	L	E	K	I	C	N	E	K	I	V	A	T	P	G	C	S	L	L	K	I	S	M	K	:	120				
P43184	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	A	L	G	T	L	E	K	I	C	N	E	K	I	V	A	T	P	G	C	S	L	L	K	I	S	M	K	:	120				
P45431	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	A	L	G	T	L	E	K	I	C	N	E	K	I	V	A	T	P	G	C	S	L	L	K	I	S	M	K	:	120				
P43186	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	A	L	G	T	L	E	K	I	C	N	E	K	I	V	A	T	P	G	C	S	L	L	K	I	S	M	K	:	120		
P43178	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	S	L	L	K	I	M	N	K	:	120		
P43183	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	C	V	L	K	I	S	M	K	:	120		
P43179	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	C	V	L	K	I	S	M	K	:	120		
CAA54486	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	C	V	L	K	I	S	M	K	:	120		
P43180	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	C	V	L	K	I	S	M	K	:	120		
P43185	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	C	V	L	K	I	S	M	K	:	120		
P15494	:	MCVFN	Y	E	T	S	V	I	P	A	A	R	J	F	K	A	F	I	L	E	G	T	L	D	N	L	P	K	V	A	P	A	I	S	S	V	E	N	I	E	C	M	C	P	C	T	I	K	I	T	F	P	E	C	S	F	P	K	Y	Y	K	E	R	V	E	D	H	A	M	F	K	Y	S	M	E	G	G	P	V	E	D	L	E	K	I	S	N	E	K	I	V	A	T	P	G	C	C	V	L	K	I	S	M	K	:	120		
P19418	:	MCVQ	S	E	V	E	A	T	S	S	A	E	K	L	F	K	G	L	C	L	D	I	D	L	L	P	V	L	P	G	A	K	S	S	E	T	L	E	D	C	G	V	T	V	K	L	H	L	G	A	S	F	P	K	T	M	K	Q	V	D	A	I	D	K	A	T	F	T	S	I	D	G	L	L	G	F	I	E	S	I	N	N	H	F	T	A	V	P	N	A	D	G	C	C	T	V	K	S	I	I	:	120						
P19417	:	MCVQ	S	E	V	E	T	S	S	A	E	K	L	F	K	G	L	C	L	D	I	D	L	L	P	V	L	P	G	A	K	S	S	E	T	L	E	D	C	G	V	T	V	K	L	H	L	G	A	S	F	P	K	T	M	K	Q	V	D	A	I	D	K	A	T	F	T	S	I	D	G	L	L	G	F	I	E	S	I	N	N	H	F	T	A	V	P	N	A	D	G	C	C	T	V	K	S	I	I	:	120							
P49372	:	MCVQ	T	H	L	E	L	T	S	S	A	E	K	L	F	Q	G	F	V	D	V	T	L	V	L	P	K	A	P	G	A	K	S	V	E	-	I	E	D	C	G	P	L	K	I	L	I	D	G	G	T	I	T	H	L	R	D	C	M	K	E	A	L	T	F	D	S	V	I	D	G	L	L	G	F	I	E	S	I	N	N	H	F	T	A	V	P	N	A	D	G	C	C	T	V	K	S	I	I	:	119							
P27538	:	MC	A	V	T	D	V	E	A	S	S	V	P	A	O	T	T	I	K	G	L	L	D	N	I	I	P	K	V	I	P	A	I	S	S	V	E	N	I	E	C	C	A	C	T	I	K	I	T	L	G	R	S	Q	T	V	V	K	R	I	D	E	I	D	A	E	A	L	K	Y	S	I	I	E	G	L	L	G	F	I	E	S	I	N	N	H	F	T	A	V	P	N	A	D	G	C	C	T	V	K	S	I	I	:	119			
P14710	:	MCVFN	V	E	I	T	S	V	A	P	A	I	L	K	A	L	V	D	A	D	L	L	P	K	V	I	-	D	A	I	K	S	E	I	V	E	C	M	C	G	A	C	T	I	K	I	T	F	V	E	D	C	E	T	H	V	L	H	K	V	E	L	V	D	V	A	M	L	A	N	Y	S	I	V	G	C	G	F	P	D	T	V	E	K	I	S	F	E	A	K	L	S	A	C	P	M	C	S	I	A	K	L	S	V	K	:	119	
P13239	:	MCVFN	F	R	E	A	T	S	V	A	P	A	I	L	K	A	L	V	D	A	D	L	L	P	K	V	I	-	D	A	I	K	S	E	I	V	E	C	M	C	G	A	C	T	I	K	I	T	F	V	E	D	C	E	T	H	V	L	H	K	V	E	L	V	D	V	A	M	L	A	N	Y	S	I	V	G	C	G	F	P	D	T	V	E	K	I	S	F	E	A	K	L	S	A	C	P	M	C	S	I	A	K	L	S	V	K	:	119
P27047	:	MCVFN	F	R	E	A	T	S	V	A	P	A	I	L	K	A	L	V	D	A	D	L	L	P	K	V	I	-	D	A	I	K	S	E	I	V	E	C	M	C	G	A	C	T	I	K	I	T	F	V	E	D	C	E	T	H	V	L	H	K	V	E	L	V	D	V	A	M	L	A	N	Y	S	I	V	G	C	G	L	P	D	T	V	E	K	I	S	F	E	A	K	L	S	A	C	P	M	C	S	I	A	K	L	S	V	K	:	119
P25986	:	M	A	V	T	F	E	Q	T	S	V	A	P	A	T	L	K	A	V	K	A	D	A	T	I	P	V	P	K	A	V	-	D	S	F	K	S	V	E	I	V	E	C	M	C	P	C	T	I	K	I	S	F	V	E	D	C	E	T	H	V	L	H	K	I																																											

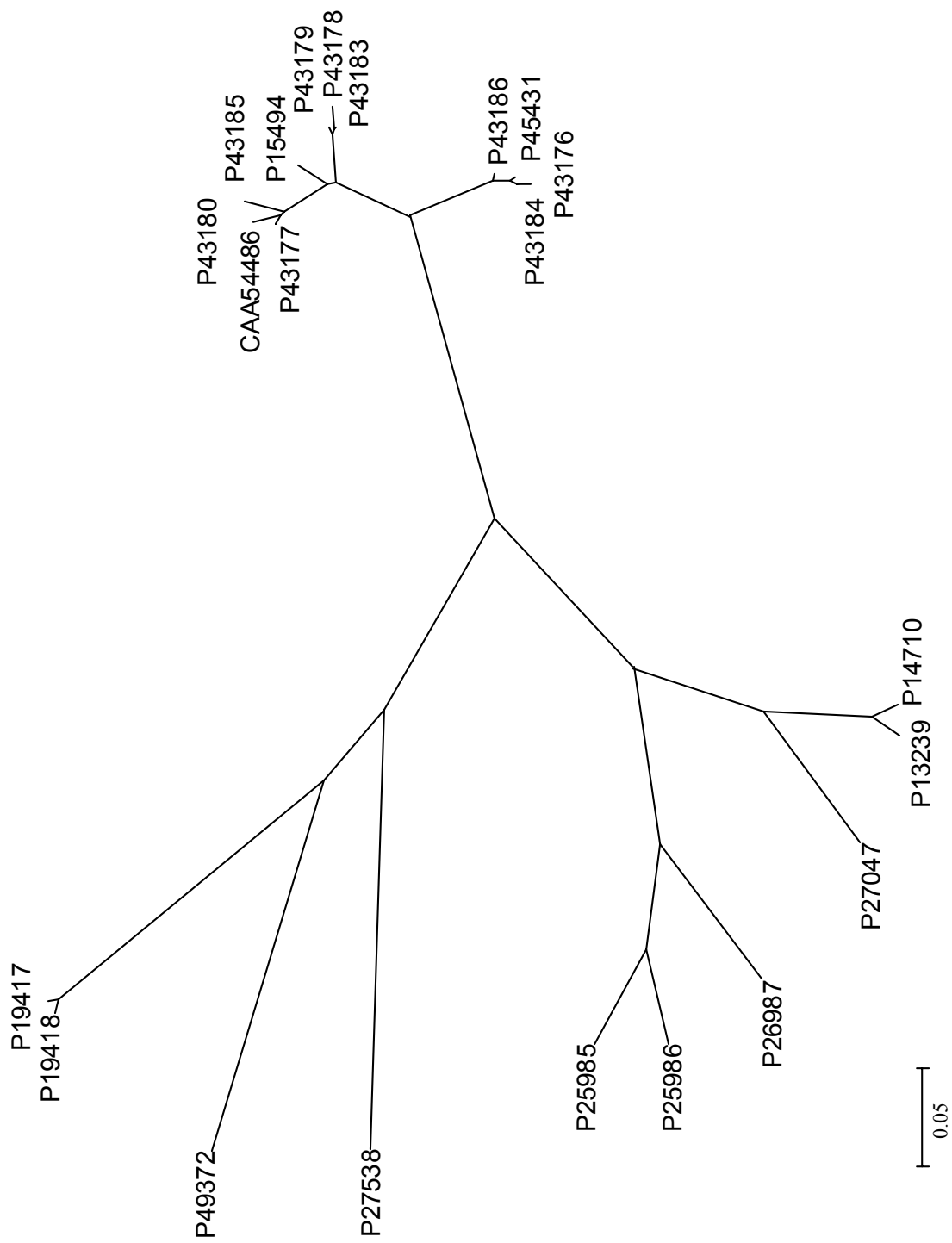


Figura III.20. Árvore filogenética das seqüências de PR 10 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 8
Resultados obtidos para os conjuntos de PR9 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	6	1029 pb	44,41%	771 (74,93%)	0,42541	343	28 (8,16%)	0,399	M3, M5 e M8	M3, M4 , M5, M6, M8 , M9, M10, M11, M12, M13
B	5	1029 pb	44,50%	947 (92,03%)	0,10873	343	50 (14,58%)	0,731	M5	M2, M3, M4, M5, M6, M8, M9, M10, M11, M12, M13

Ídem Tabela III.2

Tabela III 9
Resultados obtidos para os conjuntos de PR10 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	12	480 pb	48,33%	410 (85,42%)	0,26155	160	27 (16,88%)	0,230	M2	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13
B	4	477 pb	46,35%	230 (48,22%)	0,78117	159	2 (1,26%)	0,319	M5 e M7	M3, M5, M6, M8, M9, M10, M11, M12, M13
C	6	480 pb	44,53%	283 (58,96%)	0,63576	160	11 (6,88%)	0,253	M3	M3, M4 , M6, M9, M11, M12, M13

Ídem Tabela III.2

III.2.11. PR 11

A PR 11 representou um problema devido à reduzida quantidade de seqüências disponíveis. Na realidade, apenas as duas seqüências isoladas por Melchers *et al* (1994), no artigo que descreve pela primeira vez essas proteínas, estão claramente descritas nos bancos de dados. Duas hipóteses são possíveis: ou nenhum pesquisador teve interesse em isolar essas proteínas em outros organismos, ou elas são realmente exclusivas de tabaco. As duas seqüências mais semelhantes obtidas são homólogas a quitinases, obtidas por similaridade de seqüência a partir do seqüenciamento do genoma total de *Arabidopsis thaliana*. Os resultados obtidos estão indicados nas Figuras III.21 e III.22 e na Tabela III.10. O total de pares de bases considerados situou-se ao redor de 1100, o conteúdo de GC entre 40% e 43%, havendo, naturalmente, uma redução drástica na percentagem de sítios constantes (de 95% para 48%) quando as seqüências de *A. thaliana* são consideradas. Não há indicação de seleção positiva quando somente duas seqüências são utilizadas (*N. tabacum*) e apenas dois códons (n^{os} 26 e 121) fornecem tal indicação no conjunto ampliado. O ω médio situou-se entre 0,212 e 0,339.

III.2.12. PR 12

Foram formados dois conjuntos um pouco discrepantes: o primeiro com todas as seqüências disponíveis (A) e o outro com a eliminação de duas delas (B). Os resultados são mostrados nas Figuras III.23 e III.24 e na Tabela III.11. A seqüência P30230 de *Raphanus sativus* reuniu-se com a Q941N7 de *Brassica oleraceae* e não com as outras duas de sua espécie, que também não se agruparam entre si. Foram considerados 240 pb nas duas comparações, o conteúdo de GC foi praticamente o mesmo (46%) nas duas, e a proporção de sítios constantes variou entre 71% e 75%. Para os dois conjuntos, há sugestão de seleção positiva nos códons 9, 13, 17, 56, 59 e 71, sendo que, no conjunto A, o códon 8 também apresenta esta sugestão. A indicação mais forte ocorreu no códon 56, com valores significativos em oito e nove modelos, em A e B respectivamente. O ω médio situou-se entre 0,196 e 0,204.

PR 11

```

Q43576 : MANSVTLFA--IIFSCFLLQQLVCTNSQW-VKGGYWFKDSGLALNNIDSTLFTHLFCFAFADLNPQLNQLIIISPENQDSFRQFTSTVQRKNPVSVKTFLSIA : 97
Q43591 : MANSVTLFS--IIFSCFLLRQLVCTNSQWVIKGGYWFKNKSLALNNIDSTLFTHLFCFAFADLNPQSNQLIIISPENQDSFSQFTSTVQRKNPVSVKTFLSIA : 98
O81862 : -MSSTKLISLIVSITFFLT--LQCSMAQTVVKASYWFPASEFPVTDIDSSLFTHLFCFAFADLNSQTNQVTVSSANQPKFSTFTQTVQRNRPVSVKILLSIG : 97
O81863 : -MSSTKLIS----ITFFLSLLRFSSAQTVVKATYWF AESESPLAQIDSSLFTHLFCFAFADINLTLYQVIVSSRNKPKFSTFTQTVRRRNPTVKILLSIG : 95

Q43576 : GGRANSTAYGIMARQPNRSKSFIDSSIRLARQLGFHGLDLDWEYPLSAADMNLTGTLLEWERTAINTEARNSGRAALLLTAAVSNSPRVNGLNYPVESLA : 197
Q43591 : GGRADTTAYGIMARQPNRSKSFIDSSIRLARQFGFHGLDLDWEYPLSATDMNLTGILLNEWRTAINMEARNSGRAALLLTAAVSYSPRVNGLNYPVESVA : 198
O81862 : GGIADKTAYASMASNP TSKRKFIDSSIRVARSYGFHGLDLDWEYPSATEMNF GTLLREWRSAAVVAEASSSGKPRLLLAAAVFYSMNYYSVLYPVSAVA : 197
O81863 : GDFTYNF AFASMASNP TSKRKF ISSSIKLABRSCGFHGLDLDNWKYPSITTEMDFGKLLREWRLAVEAEARSSGKPRLLLTAAVFYSYYSVLYHPVNAVA : 195

Q43576 : RNLDMINLMAYDFYGNWSPSQTNSHAQLFDPVNH-VSGSDGINAMIQAQVPTKKLVLGIPFFGYAWRLVNIHGLRAPAAAGKSNVGVAVDDGSMTYMRI : 296
Q43591 : RNLNWINLMAYDFYGNWSPSQTNSHAQLFDPVNH- ISGSDGINAMIQAQVPTKKLVLGIPFFGYAWRLVNPNIHDLRAPAAAGKSNVGVAVDDGSMTYMRI : 297
O81862 : SSLDWNLMAYDFYGNWSPSQTNSHAQLFDPVNH-RVTGPPAALFDPNAGPSDAGTRSWIQAAGLPAKKAVALGFFPYGYAWRLTANSHSYAAPTTGAAIS---PDGSGIGYQI : 293
O81863 : DSLDWNLVAYDFYESGSS-RVTCSPAPLYDPITTGPSGDAGVRAWTQAAGLPAKKAVALGFFPYGYAWCLTDAKNHNYANSSGPAIS---PDGSGIGYDQI : 291

Q43576 : RDIYIVESRATTVMNATIVGDYCYSGSNWISYDDTQTVRNKVMYVKGKRGLLGYFAHVAGDQNWGLSRTASQTMGVSFQEMK : 377
Q43591 : RDIYIVQSRATTVMNATIVGDYCYSGSNWISYDDTQSVRNKVMYVKGKRGLLGYFAHVAGDQNWGLSRTASQTMGVSSQEMK : 378
O81862 : RKFIVDNGATTVMNISTVVGDYCYAGTNWIGYDDNQSIVTKVRYAKQKRGLLGYFSWHVVGADDNLSRAASQAAMDATTATR : 374
O81863 : RRFIVDNKATVMYNSMLVQNYCYAKKTIWIGYDDNQSIVMKVYAKQKRGLLGYFSWHIGADDNLSRAGSVLFFQF----- : 366

```

Figura III.21. Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR11. As seqüências com nomes em azul escuro correspondem ao conjunto A, e em azul claro, as seqüências acrescentadas no conjunto B. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

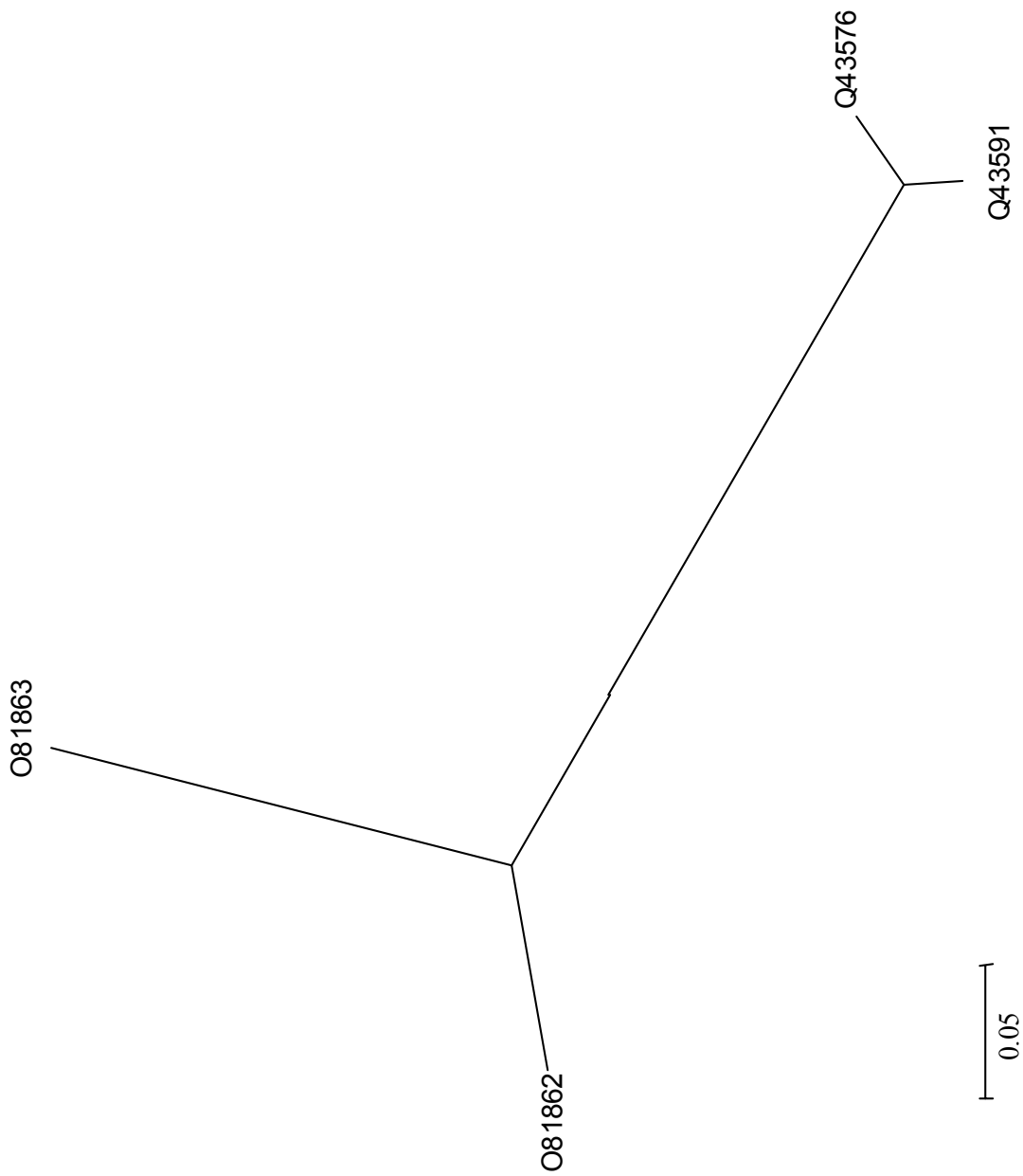


Figura III.22. Árvore filogenética das seqüências de PR 11 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

PR 12

```

*          20          40          60
O24332 : MAKEASIVALLFAALVVFVFAAFEAPTVEA-KLCERSSGTWMSGVCGNNAACKNQICIRLEGAQHGSC : 64
Q39313 : MAKEASIIITLFFAALVVFVFAAFEAPTVEA-KLCERSSGTWMSGVCGNNAACKNQICIRLEGAQHGSC : 64
Q94IN7 : MAKVASIVALLFPALVIFAALVVFVFAAFEAPTVEAQLCERPSGTWMSGVCGNNAACKNQICIRLEKARHGSC : 65
P30230 : MAKEASIIIVLLFVALVVFVFAAFEAPTVEAQLCQRPSTWMSGVCGNNAACKNQICIRLEKARHGSC : 65
P30225 : MAKEASIIIALLF AALVLFVFAAFEAPTVEAQLCERPSGTWMSGVCGNNAACKNQICINLEKARHGSC : 65
O24331 : MAKEVSIITLFFVALVLFVFAAFEAPTVEAQLCERSSGTWMSGVCGNNAACKNQICINLEGARHGSC : 65
Q9FS38 : MAKEASIIIALLF AALVLFSAFEAPSMVEAQLKCEKSSGTWMSGVCGNNAACKNQICINLEGARHGSC : 65

*          80
O24332 : NYVFP AHKCI CYFPC : 79
Q39313 : NYVFP AHKCI CYFPC : 79
Q94IN7 : NYVFP AHKCI CYFPC : 80
P30230 : NYVFP AHKCI CYFPC : 80
P30225 : NYVFP AHKCI CYFPC : 80
O24331 : NYIFPYHRCI CYFPC : 80
Q9FS38 : NYIFPYHRCI CYFPC : 80

```

Figura III.23. Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR12. As seqüências com nomes em azul escuro estão presentes somente no conjunto A, e em azul claro, no A e no B. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva em ambos os conjuntos, e em azul, apenas no A. Alinhamento realizado com ClustalX e editado com GeneDoc.

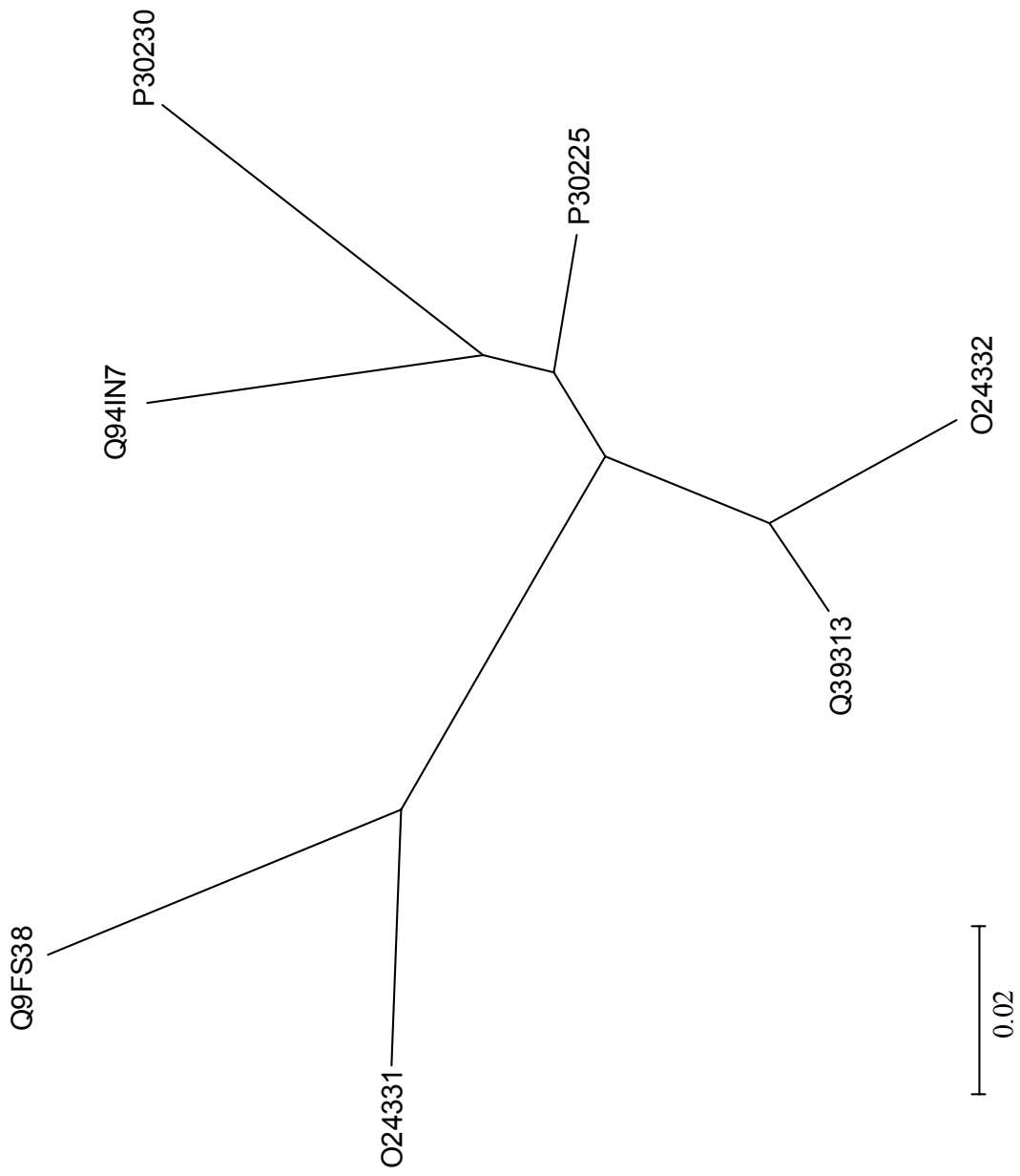


Figura III.24. Árvore filogenética das seqüências de PR 12 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 10
Resultados obtidos para os conjuntos de PR11 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	2	1134 pb	40,04%	1080 (95,24%)	0,06053	378	0 (0,00%)	0,212	M0	
B	4	1143 pb	43,27%	553 (48,38%)	0,80513	381	2 (0,52%)	0,339	M5	M3, M5, M6, M8, M9, M10, M11, M12, M13

Ídem Tabela III.2

Tabela III 11
Resultados obtidos para os conjuntos de PR12 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	7	240 pb	45,94%	171 (71,25%)	0,43253	80	7 (8,75%)	0,204	M4	M3, M4, M5, M6, M9, M10, M11, M12, M13
B	5	240 pb	45,92%	179 (74,58%)	0,35821	80	6 (7,50%)	0,196	M3 e M4	M3, M4, M5, M6, M9, M10, M11, M12, M13

Ídem Tabela III.2

III.2.13. PR 13

Os dois conjuntos formados para PR 13 compreendem espécies de monocotiledôneas. Os resultados são apresentados nas Figuras III.25 e III.26 e na Tabela III.12. Note-se, na Figura III.26, a boa separação entre os dois conjuntos e o fato de que a seqüência P21742, de *Hordeum vulgare*, se associa com as de *Triticum aestivum* e *Secale cereale*, não com as outras quatro de sua espécie. Um total de 408 a 411 pb foram considerados, o conteúdo de GC situou-se entre 48% e 50%, sendo a proporção de sítios constantes variando entre 85% e 89%. Seleção positiva só foi identificada para um códon (nº 36) no conjunto A, mas com referência ao B houve acentuada concordância dos modelos para tal indicação nos códons 47, 71, 118 e 137 (nove a 10 valores significativos entre todos os modelos). O ω médio distribuiu-se entre 0,256 e 0,650.

III.2.14. PR 14

O número de conjuntos formados para a PR14 foi realmente surpreendente. A quantidade de seqüências obtidas nos bancos de dados demonstra a grande distribuição dessas proteínas no Reino Vegetal.

Os resultados são apresentados nas Figuras III.27 e III.28 e na Tabela III.13. Há alguns agrupamentos interessantes na Figura III.28, entre as seqüências Q42952 e Q03461 de *Nicotiana tabacum* com duas outras de *Lycopersicon esculentum* (P27056, P93224); bem como de duas seqüências de *Triticum aestivum* (Q9FUK0, Q9ATG4) com seis outras espécies de *Hordeum vulgare*. O alinhamento envolveu entre 342 e 363 pb, o conteúdo de GC situou-se entre 50% e 72%, e a proporção de sítios constantes entre 63% e 91%. Há uma ampla heterogeneidade no que se refere à probabilidade de seleção positiva, de zero (conjuntos C, E e H) a 17% (conjunto F, formado por cinco seqüências de *Gossypium hirsutum*). O ω médio variou entre 0,153 e 1,259 (o único valor acima de 1 observado na presente investigação).

PR 13

	*	20	*	40	*	60	*	80
P08772	:	MAPSKIKSWVICVLILGLVLEQVQVEGKSCCKD	TL	ARNCYNTCHF	AGGSRPVCAGACRCKI	ISGPKCPSDY	PKLNLLPE	: 80
P09618	:	MAPSKIKSWVICVLILVLVLEQVQVEGKSCCKD	TL	ARNCYNTCRF	AGGSRPVCAGACRCKI	ISGPKCPSDY	PKLNLLPE	: 80
Q42838	:	MATNKSIKSWVICVLILGLVLEQVQVEGKSCCKNT	T	GRNCYNACRF	AGGSRPVCATACGCKI	ISGPTCPRD	YPKLNLLPE	: 80
P09617	:	MATNKSIKSWVICVLILGLVLEQVQVEAKSCCKNT	T	GRNCYNACRF	AGGSRPVCATACGCKI	ISGPTCPRD	YPKLNLLPE	: 80
P32032	:	MG-SKGLKGVWVCLLILGLVLEQVQVEGKSCCRT	TL	GRNCYNLCR-	SRGAQKLCSTVCRCKL	TSGLSCP	GFPKLALESN	: 78
Q43205	:	MG-SKGLKGVWVCLLILGLVLEQVQVEGKSCCRT	TL	GRNCYNLCR-	SRGAQKLCSTVCRCKL	TSGLSCP	GFPKLALESN	: 78
P01543	:	MG-SKGLKGVWVCLLILGLVLEQVQVEGKSCCK	STL	GRNCYNLCR-	ARGAQLCANVCRCKL	TSGLSCP	GFPKLVLESN	: 78
Q9TOP2	:	MG-SKGLKGVWVCLLILGLVLEQVQVEGKSCCK	STL	GRNCYNLCR-	ARGAQLCANVCRCKL	TSGLSCP	GFPKLVLESN	: 78
Q9ZNY5	:	MG-SKGLKGVWVCLLILGLVLEQVQVEGKSCCK	STL	GRNCYNLCR-	TRGAQKLCANFCRCKL	ISS	TSCPKEFPKLALESN	: 78
P21742	:	MG-SKGLKGVWVCLLILGLVLEHVQVEGKSCCR	STL	GRNCYNLCR-	VRGAQKLCANACRCKL	TSGLKC	PSSFPKLAALVSN	: 78
	*	100	*	120	*			
P08772	:	SSEPDTQYCTIGCRNSVCDNMDNVFRG-QEMK	FD	MGLCSNACARF	CND	GAVIQSVEA	:	137
P09618	:	SSEPDTQYCTIGCRNSVCDNMDNVFRG-QEMK	FD	MGLCSNACARF	CND	GAVIQSVEA	:	137
Q42838	:	SSEPNTVEYCTIGCRTSVCDNMDNVSRG-QEMK	FD	MGLCSNACARF	CND	GEVIQSVEA	:	137
P09617	:	SSEPNTVEYCTIGCRTSVCDNMDNVSRG-QEMK	FD	MGLCSNACARF	CND	GEVIQSVEA	:	137
P32032	:	SDEPDTIEYCNLGCRRSVCYMWNAADDEEM	KL	YVENCGD	ACVNF	CNGDAGL	TSLDA	: 136
Q43205	:	SDEPDTIEYCNLGCRRSVCYMWNAADDEEM	KL	YVENCGD	ACVNF	CNGDAGL	TSPDA	: 136
P01543	:	SDEPDTMEYCNLGCRRSLLCDYIVNAAADDE	EM	KLYVEQCGD	ACVNF	CNADAGL	TSLDA	: 136
Q9TOP2	:	SDEPDTMEYCNLGCRRSLLCDYMWNAADDE	EM	KLYVEQCGD	ACVNF	CNADAGL	TSLDA	: 136
Q9ZNY5	:	SDEPDTVEYCNLGCRRSVCYMWNSAAAADDE	EM	KLYVERCGD	ACVSF	CNGDAGL	PLSA	: 136
P21742	:	SDEPDTIDYCNLGCRAAMCDYMWNAADDEEM	KL	YVEH	CSD	ACVNF	CNGDVGLTSLTA	: 136

Figura III.25. Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR13. As seqüências com nomes em amarelo correspondem ao conjunto A, e em azul, B. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

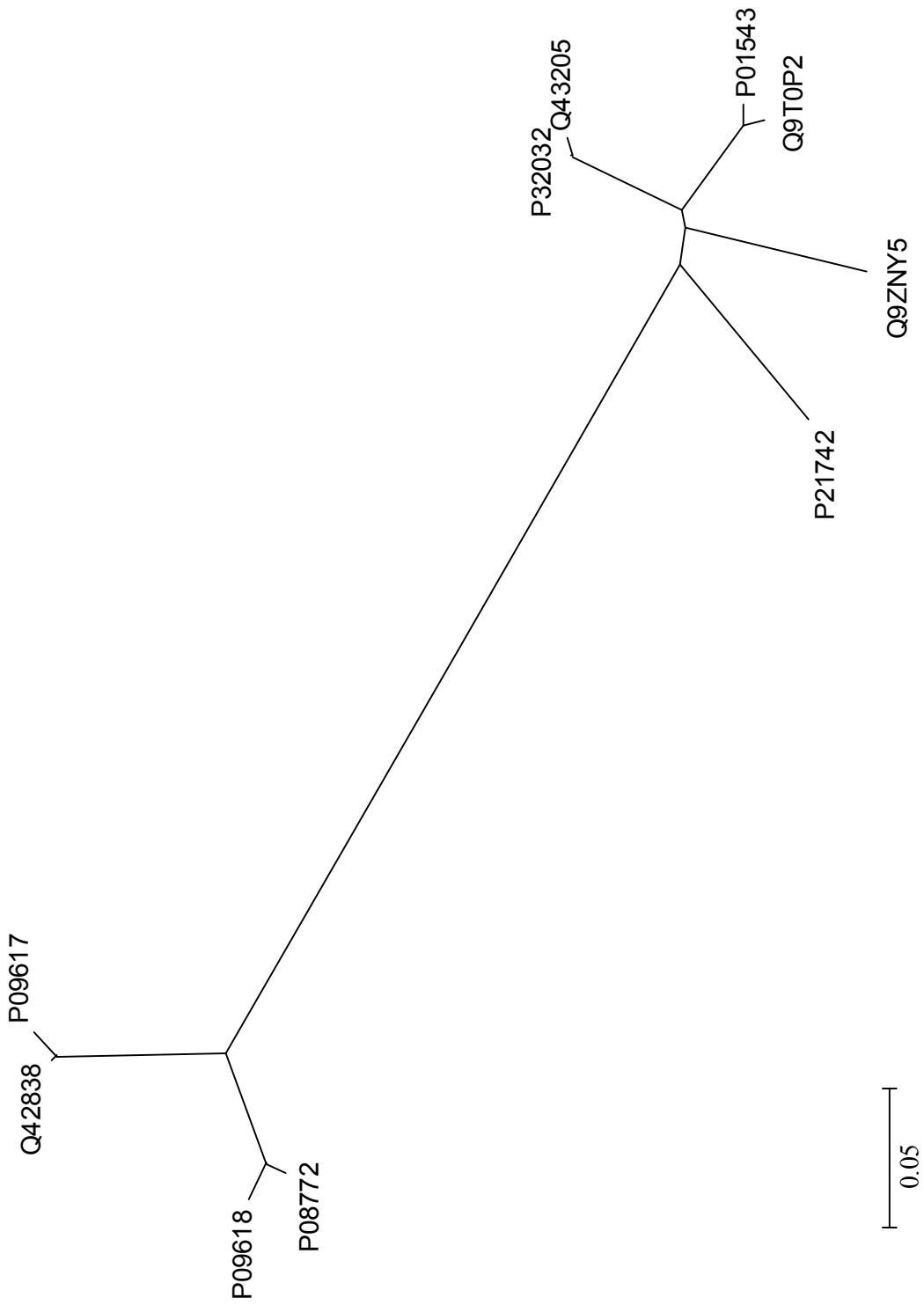


Figura III.26. Árvore filogenética das seqüências de PR 13 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 12
 Resultados obtidos para os conjuntos de PR13 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	4	411 pb	48,48%	364 (88,56%)	0,12943	137	1 (0,73%)	0,256	M1	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13
B	6	408 pb	49,96%	346 (84,80%)	0,20803	136	4 (2,94%)	0,650	M5	M3, M4, M5, M6, M8, M9, M10, M11, M12, M13

Ídem Tabela III.2

PR 14

	*	20	*	40	*	60	*	80	*	100	*	120
043875	:	MARAAATQLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
042842	:	MARAAATQLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
043767	:	MARAAASQLVLVAVMAA--MLLVA-----	ADAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
042849	:	MARSAATQLVLVAVMAA--MLLVA-----	ADAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
043762	:	MARSAATQLVLVAVMAA--MLLVA-----	ADAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
043871	:	MARAAATQLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
09ATC4	:	MARAAATQLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
09FUKO	:	MARAAATQLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
043193	:	MAR---LAVAIAVVAAA--VWVWLAATTS	EAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	118					
043194	:	MARSMK--LAVAIAVVAAA--VWVWLAATTS	EAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	122					
024583	:	MAR--TQ--SAAVAVVAAA--VWVWLAATTS	EAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	121					
P19656	:	MAR--T--QDLAVVAVVAAA--VWVWLAATTS	EAATSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	120					
P23096	:	MAR--A--QLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	116					
022485	:	MAR--A--QLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	116					
042978	:	MAR--A--QLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	118					
022483	:	MAR--A--QLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	118					
042999	:	MAR--A--QLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
022482	:	MAR--A--QLVLVAVMAA--MLLVA-----	TDAAITSCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
Q03461	:	MEM-----VQKIACFVVLCHVWVAP--	HAAVAITCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	114					
P93224	:	MEM-----VQKIACFVVLCHVWVAP--	HAAVAITCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	114					
P27056	:	MEM-----VQKIACFVVLCHVWVAP--	HAAVAITCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	114					
042952	:	MEI-----ACKIACFVVLCHVWVAP--	CAEAITCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	114					
042614	:	MAG-----LVKLSCLVLACHIVACPIATNAALSCG	VTSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
042642	:	MAG-----LVKLSCLVLACHIVACPIATNAALSCG	VTSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
042615	:	MAG-----LVKLSCLVLACHIVACPIATNAALSCG	VTSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
042616	:	MAG-----LVKLSCLVLACHIVACPIATNAALSCG	VTSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
043304	:	MAG-----LVKLSCLVLACHIVACPIATNAALSCG	VTSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	118					
09FVA5	:	MAS-----MSLKLACVWVLCMVWGAP--	LAQCAVTTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	120					
09M6B6	:	MAS-----MSLKLACVWVLCMVWGAP--	LAQCAVTTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	120					
09M6B8	:	MAS-----MSLKLACVWVLCMVWGAP--	LAQCAVTTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	120					
049200	:	MAS-----MSLKLACVWVLCMVWGAP--	LAQCAVTTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	120					
043017	:	MAYS-----AMTKLALVVALCHVWVSP--	IAQ--ALTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
09MSX8	:	MACS-----AMTKLALVVALCHVWVSP--	IAQ--ALTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	117					
09MSX7	:	MAS-----AVTKLALVVALCHVWVSP--	IAQ--ALTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					
09MSX6	:	MAS-----AVTKLALVVALCHVWVSP--	IAQ--ALTCCGVSSALSPCLSYARCN--GAKPFPACCCGWRRLAGAAQSTADKQAACKCKIKSAAG---	CLNAGKAAAGLPSMCCGVSPYAIASAVD	CSKIR	:	115					

Figura III.27. Alinhamento das seqüências de aminoácidos dos oito conjuntos da PR14. As seqüências com nomes em vermelho correspondem ao conjunto A; verde, B; amarelo, C; Lilás, D; azul claro, E, e os dois tons de azul, ao conjunto H; rosa, F, e laranja, G. Os sítios marcados em cinza correspondem aos códons indicados como alvos de seleção positiva. Alinhamento realizado com ClustalX e editado com GeneDoc.

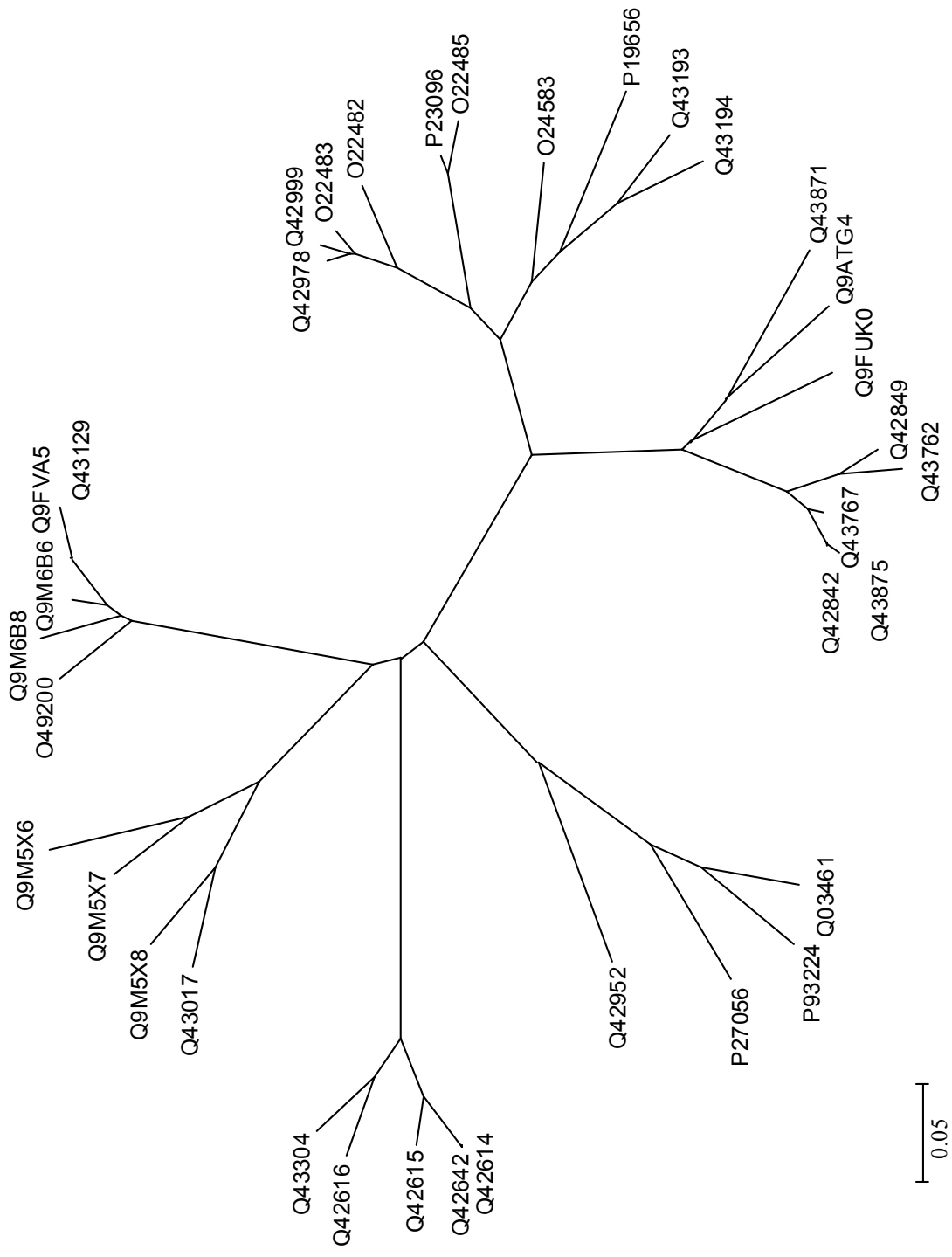


Figura III.28. Árvore filogenética das seqüências de PR 14 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 13
Resultados obtidos para os conjuntos de PR14 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	8	345 pb	65,87%	240 (69,57%)	0,47955	115	5 (4,35%)	0,399	M5	<u>M3</u> , M4, M5, M6, M8, M9, M10, M11, M12, M13
B	4	363 pb	72,26%	267 (73,55%)	0,26793	121	9 (7,44%)	0,177	M1 e M5	M3, M4, M5, <u>M6</u> , M9, M11, M12, M13
C	6	354 pb	70,52%	266 (75,14%)	0,28786	118	0 (0,00%)	0,153	M5 e M7	M3, M4, M8, M12
D	4	342 pb	50,51%	216 (63,16%)	0,49695	114	2 (1,75%)	0,300	M1	M3, M5, M6, M8, M9, M10, M11, M12, M13
E	4	351 pb	53,00%	315 (89,74%)	0,11522	117	0 (0,00%)	0,381	M0	
F	5	360 pb	55,95%	328 (91,11%)	0,10215	120	20 (16,67%)	1,259	M5	M0, <u>M2</u> , <u>M3</u> , M4, M5, <u>M6</u> , M8, M9, M10, <u>M11</u> , M12, M13
G	4	351 pb	56,25%	271 (77,21%)	0,25668	117	1 (0,85%)	0,480	M1 e M5	M2, M4, M5, M6, M8, M9, M10, M11, M12, M13
H	5	357 pb	53,47%	307 (85,99%)	0,22777	119	0 (0,00%)	0,382	M1 e M5	<u>M3</u> , M4, M5, M6, <u>M8</u> , M9, M10, M11, M12, M13

Ídem Tabela III.2

III.2.15. PR 15

Como esta PR ainda não foi oficialmente classificada em publicação científica, deve ser considerada à parte dos outros dados. Os resultados estão expressos nas Figuras III.29 e III.30 e na Tabela III.14. A árvore filogenética tem uma forma de estrela, mostrando acentuadas divergências entre as seqüências. Os dois conjuntos formados, são praticamente idênticos, com a diferença de duas seqüências acrescentadas ao conjunto B, o A reunindo seqüências um pouco mais homogêneas. Foram considerados 672 pb, o conteúdo de GC permanecendo praticamente idêntico (61%), e a proporção de sítios constantes variou entre 72% (B) e 78% (A). Nenhuma das análises apontou indícios de seleção positiva. Os valores de ω médio foram os mais baixos do presente estudo (0,058 e 0,084, respectivamente).

III.3. SGI

Quinze conjuntos de seqüências, de dez PRs estudadas, foram adaptados para a análise de máxima parcimônia realizada com o programa SGI (Su, 2000). Nenhum sítio em qualquer dos conjuntos analisados apresentou probabilidade significativa ($P > 0,95$) de seleção positiva (resultados não mostrados).

PR 15

```

P15290 : MGYSKTLVAGLFAMLLLAPAVLATAVLAADLQDFCVADLDGKAVSVNGHTCKPWEAGDDFLFSSKLA KAGNTSTPNGSAVTELDV : 84
P45851 : MGYSKTLAVSLFAVLLLAAPAVLADSDPDPLQDFCVADLDGKAVSVNGHCPKPWEAGDDFLFSSKLA KAGNTSTPNGSAVTELDV : 84
P26759 : MGYSKNIASGMFAMLLLASAVLSSNPHPLQDFCVADLDGKAVSVNGHCKPWEAGDDFLFSSKLA KAGNTSTPNGSAVTDLNV : 84
P93600 : MGYSKNLGAGLFGMLLAPGVLASDPDPLQDFCVADLDGKAVSVNGHCKPWEAGDDFLFSSKLA KAGNTSTPNGSAVTELDV : 84
O24004 : MGYSKNLGAGLFTMLLAPAIMATDPDPLQDFCVADLDGKAVSVNGHTCKPWEAGDDFLFSSKLA KAGNTSTPNGSAVTELDV : 84
Q9FEW6 : MAYFKTLAAGLFAALLFLAPFIMATDPDPLQDFCVADLDGKEVSVNGHCPKPWEAGDDFLFSSKLA KAGNTSTPNGSAVTELDV : 84

P15290 : AEWPGTNTLGVSMNRVDFAPGGTNPPIHPRATEIGIVMKGELLVGLGSLDSGNKLYSRVVRAGETFLIPRGLMHFQFNVGKT : 168
P45851 : AEWPGTNTLGVSMNRVDFAPGGTNPPIHPRATEIGIVMKGELLVGLGSLDSGNKLYSRVVRAGETFLIPRGLMHFQFNVGKT : 168
P26759 : AEWPGTNTLGVSMNRVDFAPGGTNPPIHPRATEIGIVMKGELLVGLGSLDSGNKLYSRVVRAGETFLIPRGLMHFQFNVGKT : 168
P93600 : AEWPGTNTLGVSMNRVDFAPGGTNPPIHPRATEIGIVLKGELLVGLGSLDSGNKLYSRVVRAGETFLIPRGLMHFQFNVGKT : 168
O24004 : AEWPGTNTLGVSMNRVDFAPGGTNPPIHPRATEIGIVMKGELLVGLGSLDSGNKLYSRVVRAGETFLIPRGLMHFQFNVGKT : 168
Q9FEW6 : AEWPGTNTLGVSMNRVDFAPGGTNPPIHPRATEIGIVMKGELLVGLGSLDSGNKLYSRVVRAGETFLIPRGLMHFQFNVGKT : 168

P15290 : EASHVVSFNQNPQIVFVPLTLFGSNPPIPTPVLTKALRVEARVVELLKSFAAGF : 224
P45851 : EASHVVSFNQNPQIVFVPLTLFGSNPPIPTPVLTKALRVEAGVVELLKSFAAGF : 224
P26759 : EASHVVSFNQNPQIVFVPLTLFGSNPPIPKPVLTKALRVEAGVVELLKSFAAGS : 224
P93600 : HASHVVSFNQNPQIVFVPLTLFFGSNPPPTPVLTKALRVEAGVVELLKSFAAGS : 224
O24004 : EAYHVVSFNQNPQIVFVPLTLFGSNPPIPTPVLTKALRVEAGVVELLKSFAAGS : 224
Q9FEW6 : EASHVVSFNQNPQIVFVPLTVFGSNPPIPTAVLTKALRVDAGVVELLKSFAAGF : 224

```

Figura III.29. Alinhamento das seqüências de aminoácidos dos dois conjuntos da PR15. As seqüências com nomes em azul claro correspondem aos conjuntos A e B, e em azul escuro, àquelas acrescentadas ao conjunto B. Alinhamento realizado com ClustalX e editado com GeneDoc.

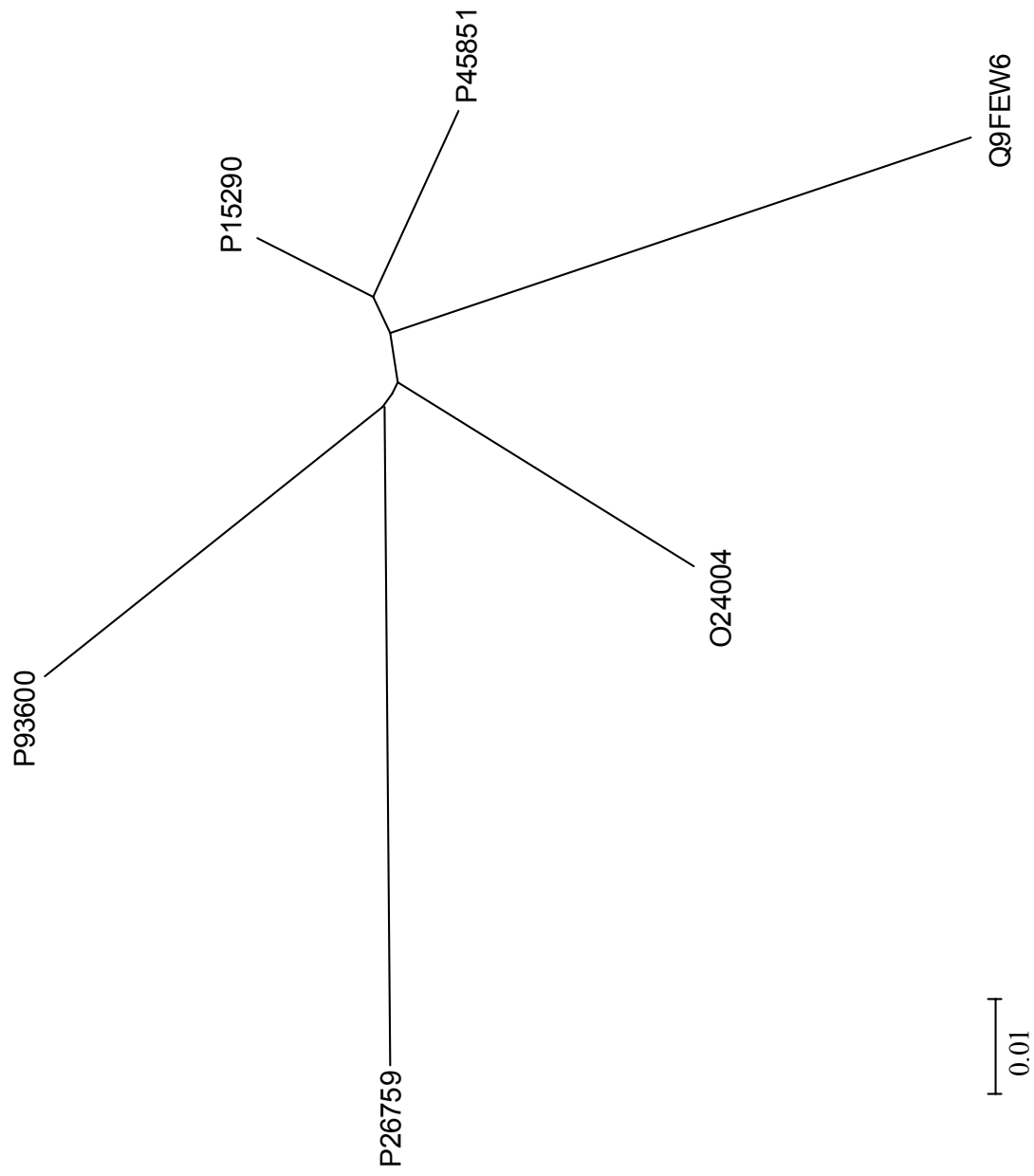


Figura III.30. Árvore filogenética das seqüências de PR 15 consideradas. Programa MEGA 2, Neighbor-Joining, distância p, *complete deletion*.

Tabela III 14
Resultados obtidos para os conjuntos de PR15 considerados

Conjuntos de seqüências	OTUs	Alinhamento	GC %	Sítios nucleotídicos constantes	Comprimento total dos ramos	Número de códons alinhados	Número de códons com ω acima de 1	ω médio entre todos os códons	Modelo de menor AIC	Modelos com indicações de seleção positiva
A	4	672 pb	61,27%	527 (78,42%)	0,35627	224	0 (0,00%)	0,058	M5 e M7	
B	6	672 pb	61,26%	487 (72,47%)	0,48462	224	0 (0,00%)	0,084	M5 e M7	

Ídem Tabela III.2

IV. Discussão

Inicialmente devem ser consideradas as limitações e eventuais méritos da presente Dissertação. Embora tenham sido consideradas 201 seqüências, de 50 espécies que representam 32 gêneros de 12 famílias, a quase totalidade das mesmas constitui-se de material cultivado, e portanto já bastante manipulado pela nossa espécie. Portanto, inferências sobre o mesmo, a respeito do que ocorre na natureza em geral devem ser feitas com cautela. Por outro lado, não há qualquer outro estudo disponível na literatura que tenha considerado, de maneira homogênea, tal número de seqüências através de 14 modelos diferentes de seleção.

O tamanho dos segmentos examinados variou de maneira acentuada, envolvendo desde somente 240 pares de bases (PR12, conjuntos A e B) até 2.268 (PR7, conjuntos A e B), estas últimas sendo 9,45 vezes maiores do que as primeiras. A constituição dessas seqüências é também bastante heterogênea. Assim, a proporção de dinucleotídeos GC variou entre 37% (PR6) e 72% (PR14, conjunto B), isto é, esta última apresenta quase o dobro da primeira no que se refere a este atributo. A percentagem de sítios constantes por sua vez, diferiu acentuadamente, os extremos ocorrendo em conjuntos diversos da mesma PR (PR11, conjunto B, 48%; conjunto A, 95%). Isto talvez possa ser devido ao fato das seqüências de *Arabidopsis thaliana* que foram incluídas para a comparação com *Nicotiana tabacum* no conjunto B serem demasiado diferentes dessas últimas. Mas também há diferenças acentuadas em outros conjuntos (por exemplo, PR8, conjunto B, seqüências de diferentes espécies de *Arabis*, 64%; PR3, conjunto A, isoformas de *Arabidopsis thaliana*, 88%).

Que evidências sobre o tipo de seleção que está ocorrendo nessas regiões protéicas foram obtidas? Considerando-se o segmento como um todo, o ω médio também variou bastante, de 0,06 (PR15, conjunto A) a 1,26 (PR14, conjunto F). Estritamente falando, apenas para este conjunto poder-se-ia inferir a ocorrência de seleção positiva para as seqüências como um todo.

A vantagem da metodologia empregada, no entanto, é que podem ser feitas análises envolvendo sítios individuais. E neste caso, apesar de que para muitos conjuntos (PR1, C; PR2, B; PR4, A; PR5, D, E; PR8,A; PR11; A; PR14 C, E, H; PR15, A, B) não foram detectados sítios que satisfizessem o critério rigoroso estabelecido neste estudo (indicações de seleção em pelo menos cinco dos 12 modelos que consideram tal tipo de seleção), para outros os resultados foram diferentes. Os casos mais extremos envolvem o conjunto A de

PR10, em que foram detectados 27 códons com valores acima de 1 em uma região de 160 códons e o conjunto F da PR14, em que foram encontrados 20 em 120 códons (frequência, em ambos os casos, de 17%). Nos outros agrupamentos foram encontrados valores intermediários entre esses extremos.

Que modelos melhor se adaptam aos tipos de distribuição observados? Se o Critério de Informação de Akaike (AIC; Sakamoto *et al.*, 1986) for utilizado, o modelo com o maior número de indicações (17) é o M5, que admite uma distribuição contínua de ω baseada na distribuição gama. Quase igualmente freqüente, no entanto (16 indicações), aparece M1, no qual o único parâmetro variável é a freqüência de códons onde $\omega = 0$.

Examinando-se os dados, agora, a partir dos resultados dos códons individuais, verifica-se que todos os modelos que admitem ω maior que 1 (portanto, com a exclusão de M1 e M7) estão representados. M13 (que possui o maior número de parâmetros) aparece 22 vezes, seguido de perto por M3 (valores e freqüências de ω não pré-fixados), M6 (combinação de duas distribuições γ), e M11 (combinação de distribuições β e normal), todos com 19 aparições.

Não há uma correlação clara entre a percentagem de códons com ω maior do que 1, o conteúdo de GC das seqüências consideradas, ou grau com que ela varia. Por exemplo, os conjuntos com maiores freqüências de códons com valores acima de 1 (PR10, A; PR14, F) apresentam conteúdos de GC e percentagem de sítios constantes de, respectivamente 48%, 85%; 56%, 91%; enquanto alguns com ausência de tais códons mostram valores similares (PR4, A: 47%, 81%; PR8, A.: 51%, 94%).

Situações interessantes ocorrem quando isoformas de uma espécie agrupam-se com as de outras. Tais situações ocorreram em cinco das árvores filogenéticas aqui apresentadas. Na Figura III.2, onde as relações filogenéticas das PR1 são apresentadas, observa-se uma associação entre duas seqüências de *Solanum tuberosum* com duas de *Lycopersicon esculentum*; na Figura III.8, associações entre isoformas de PR4 desta última espécie com as de *Nicotiana tabacum* podem ser verificadas; na Figura III.24 de seqüências de PR12 constata-se o agrupamento de seqüências de *Raphanus sativum*, *Brassica oleracea* e *Brassica napus*; na Figura III.26 verifica-se a associação das isoformas de *Hordeum vulgares* com as de *Triticum aestivum* / *Secale cereale*; e na Figura III.28, PR14, agrupamentos entre seqüências de *Lycopersicon esculentum* com *Nicotiana tabacum*, e de *Hordeum vulgare* com *Triticum aestivum* ficam evidentes. Como as estruturas moleculares dessas PRs são muito diferentes, bem como suas funções (PR1:

anti-fúngico; PR4: quitinase; PR12: defensinas anti-fúngicas; PR13: tioninas; PR14: transferência não-específica de lipídios), tais situações parecem ocorrer independentemente do tipo de molécula considerada ou sua função específica.

A não detecção de códons com probabilidade significativa para seleção positiva pelo método da máxima parcimônia com o programa SGI em qualquer dos conjuntos estudados, mesmo naqueles casos em que eles apresentaram alta probabilidade pelo método de máxima verossimilhança, não é inesperada. Suzuki & Gojobori (1999) e Suzuki & Nei (2001) alertaram para o fato de que a eficiência do seu método de parcimônia aumenta com o número de seqüências analisadas, e que portanto este método tem baixa sensibilidade quando poucas seqüências entram na análise, o que foi corroborado com os nossos resultados.

Como os dados aqui apresentados se relacionam com os já existentes na literatura? A idéia de testar-se a possível ocorrência de seleção positiva através das taxas de substituições sinônimas e não-sinônimas do DNA já é relativamente antiga. O método até pouco tempo mais usado para essas avaliações era o de Nei e Gojobori (1986), que subsequenteiramente foi modificado por Li (1993), Pamilo & Bianchi (1993), Ina *et al.* (1994) e Ina (1995). A aplicação do mesmo foi realizada em uma ampla gama de genes que codificam para proteínas ribossomais, do sistema contráctil, fatores de ativação e receptores, proteínas sangüíneas, apolipoproteínas, imunoglobulinas, interferons e enzimas, tendo os resultados respectivos sido sumariados em Graur & Li (2000). As taxas médias de todos eles foram de $0,74 \pm 0,67$ para as alterações não-sinônimas e de $3,51 \pm 1,01$ para as sinônimas. Isto é, as últimas seriam 4,7 vezes mais freqüentes do que as primeiras. Note-se, no entanto, o alto coeficiente de variação para as mudanças não-sinônimas (91%) comparado com o das sinônimas (29%).

Exemplos recentes da aplicação deste enfoque são o de Hellberg *et al.* (2000), que encontraram evidências de seleção positiva para quatro das dez comparações pareadas entre cinco espécies de gastrópodos dos gêneros *Tegula* e *Norrisia*, no que se refere às proteínas espermáticas; e o de Lukens & Doebley (2001), que estudando um fator de transcrição (*tb1*, *Teosinte branched 1*) de 23 espécies de *Andropogonae* (a tribo na qual o milho é classificado) e em cinco outras gramineas, não encontraram evidências de seleção positiva. Segundo esses últimos autores, os dados seriam melhor explicados por níveis desiguais de seleção purificadora (negativa) entre as linhagens.

O método da máxima verossimilhança, aqui utilizado, já foi aplicado em diversos outros estudos. Yang (1998) verificou, através dele, que as proporções d_N/d_S , no que se refere às lisozimas, diferiam significativamente entre linhagens de primatas. Stoltz *et al.* (2000) consideraram 22 proteínas inibidoras de poligalacturonase de plantas, que se ligavam a 19 poligalacturonases de fungos, e verificaram que substituições vantajosas dominavam a evolução molecular desses genes. Bishop *et al.* (2000), já citados anteriormente em outras seções desta dissertação, investigaram as quitinases classificadas como PR8, verificando uma sucessão de substituições de aminoácidos de caráter adaptativo. Identificaram eles resíduos individuais que estariam sujeitos à seleção positiva e os localizaram na estrutura tridimensional das quitinases de classe I. Comparadas com as lisozimas de primatas e as quitinases de classe III, as substituições nas quitinases de classe I parecem ocorrer de maneira desproporcionada na fenda do sítio ativo da molécula.

Outros estudos, desenvolvidos pelo grupo de Ziheng Yang, envolveram genes determinadores de lisina do esperma de *Haliothis* (um molusco), da globina β em vertebrados, codificadores de proteína da mitocôndria, da hemaglutinina do vírus A da influenza humana, os identificados como *env*, *vif* e *pol* do HIV-1, os das glicoproteínas ZP2 e ZP3, da zona pelúcida dos óvulos de mamíferos, e os de classe I do Complexo Maior de Histocompatibilidade (MHC) de humanos. (Yang *et al.*, 2000a, b; Yang, 2001; Yang & Swanson, 2002). Os resultados, em geral, mostram heterogeneidade de pressão seletiva entre linhagens e códons, fornecendo também evidências quanto a questões evolutivas específicas, como a competição entre espermatozóides e o conflito sexual.

De especial interesse é o estudo de Yang & Swanson (2002), por ter considerado uma classe de proteínas (alelos de classe I do MHC humano) que, como as PRs, estão envolvidas no processo de defesa contra organismos infectantes. Identificaram eles 23 códons com indicação consistente de seleção através dos modelos M2, M3 e M8.

Esses resultados mostram claramente que a evolução em nível molecular envolve uma intrincada rede de agentes de conservação estrutural e de divergência de seqüências (ver, por exemplo, Parisi & Echave, 2001). Há fatores relacionados com o uso de códons (Sueoka & Kawanishi, 2000; Morton, 2001), e a localização gênica em sítios próximos ou não dos sítios funcionais ativos tem importância óbvia. Transcorridos 143 anos da publicação da obra clássica de Charles Darwin, “A Origem das Espécies”, o seu paradigma principal, a evolução por seleção natural, permanece como o princípio norteador da evolução biológica, agora iluminado por detalhes íntimos da estrutura molecular.

V. Resumo e Conclusões

Um total de 201 seqüências de DNA, de 50 espécies pertencentes a 32 gêneros e 12 famílias, foi investigado através do método da máxima verossimilhança para identificar, nas proteínas respectivas, possíveis códons nos quais estivesse ocorrendo seleção positiva. Foram considerados 15 tipos de proteínas relacionadas à patogênese (PR1-PR15), quanto a 14 modelos diferentes de seleção. Tanto quanto se possa avaliar, não há qualquer estudo disponível na literatura que tenha examinado de maneira homogênea tal número de seqüências de forma tão abrangente.

Os resultados podem ser resumidos como segue:

1. Houve ampla variabilidade no tamanho das seqüências investigadas (de 240 a 2.268 pares de bases).
2. O conteúdo de GC das mesmas também variou bastante, de 37% a 72%.
3. A proporção de sítios nucleotídicos constantes distribuiu-se entre 48% e 95%.
4. Apenas um dos 40 ω s (proporção entre as taxas de substituições não sinônimas e sinônimas no DNA) médios calculados apresentou valor acima de 1 (1,26; conjunto F, PR14). O menor número desta medida (0,06) foi obtido para o conjunto A da PR15.
5. Quanto à ação da seleção em códons individuais, para muitos conjuntos (PR1, C; PR2, B; PR4, A; PR5, D, E; PR8, A; PR11, A; PR14, C, E, H; PR15, A, B) não foram detectados casos que satisfizessem o critério rigoroso estabelecido neste estudo [indicação significativa de seleção positiva ($P > 0,95$) em pelo menos cinco dos 12 modelos que consideram tal tipo de seleção.]. Para os outros, no entanto, os resultados foram diferentes. Os casos mais extremos ocorreram para o conjunto A da PR10, em que foram detectados 27 códons com valores de ω acima de 1 em uma região de 160 códons; e o conjunto F da PR14, em que foram encontrados 20 em 120 códons (frequência, em ambos os casos, de 17%). Nos outros conjuntos foram encontrados valores intermediários entre esses dois extremos.
6. Utilizando-se o Critério de Informação de Akaike, o modelo de seleção com o maior número de indicações (17) para os 40 conjuntos foi o M5, que pressupõe uma distribuição contínua de ω baseada na distribuição gama. Em segundo lugar (com 16 indicações) aparece M1, no qual o único parâmetro variável é a frequência de códons onde $\omega = 0$.

7. Quanto às análises relativas aos códons individuais, todos os modelos que admitem ω acima de 1 (excluindo-se, portanto, M1 e M7) estão representados. M13 (que possui o maior número de parâmetros variáveis) aparece 22 vezes. Com número um pouco menor de aparições (19) estão representados M3 (valores e frequências de ω não pré-fixados), M6 (combinação de duas distribuições gama) e M11 (combinação de distribuições beta e normal).
8. Não há uma correlação clara entre a percentagem de códons com ω maior do que 1, o conteúdo de GC das seqüências consideradas, ou o grau de variabilidade das seqüências.

A *conclusão geral* é a de que um número significativo de códons, nessas proteínas, está submetido à ação da seleção positiva, um resultado que não é inesperado devido à função genérica das mesmas, de defesa das plantas contra agentes patogênicos.

VI. Summary and Conclusions

A total of 210 DNA sequences from 50 species belonging to 32 genera and 12 families was investigated by the maximum likelihood method to identify, in the respective proteins, possible codons in which positive selection would be occurring. Fifteen types of pathogenesis related proteins (PR1 – PR15) were considered in connection with 14 different selection models. As far as can be ascertained, there is no study in the literature which had examined in a homogeneous and comprehensive way such a number of sequences.

The results can be summarized as follows:

1. There was wide variability in the sizes of the investigated sequences (from 240 to 2,268 base pairs).
2. Their GC content also varied considerably, from 37% to 72%.
3. The proportion of constant nucleotide sites was distributed between 48% and 95%.
4. Only one of the average ω s (proportion between the DNA non-synonymous and synonymous substitution rates) calculated showed a figure above 1 (1.26; set F, PR14). The lowest value of this measure (0.06) was obtained for set A of PR15.
5. In relation to the action of selection at individual codons, for many sets (PR1, C; PR2, B; PR4, A; PR5, D, E; PR8, A; PR11, A; PR14, C, E, H; PR15, A,B) no cases were detected that would satisfy the rigorous criterion established in this study [indication of positive selection ($P > 0.95$) in at least five of the 12 models which consider this type of selection]. For the others, however, the results were different. The most extreme cases occurred for set A of PR10, where 27 codons were detected with ω above 1 values, in a region of 160 codons; and set F of PR14, in which 20 of the 120 codons were found (frequency, in both cases, of 17%). In the other sets intermediate values between these extremes were found.
6. Using Akaike's Information Criterion, the selection model with the largest number (17) of indications for the 40 sets was M5, which presupposes a continuous ω distribution based on the gamma distribution. In second place (with 16 indications) appears M1, in which the only variable parameter is the frequency of codons with $\omega = 0$.
7. As for the analyses related to individual codons, all the models which allow ω above 1 (with the exclusion, therefore, of M1 and M7) are represented. M13 (which possesses the largest number of variable parameters) appears 22 times. With little less appearances (19) M3 (unconstrained ω values and frequencies) and M11 (combination of the beta and normal distributions) were represented.

8. There is no clear correlation between the percentage of codons with ω above 1, the GC content of the considered sequences; and their degree of variability.

The *general conclusion* is that there is a significant number of codons, in these proteins, which are being subjected to positive selection, a not unexpected result due to their generic defensive function against pathogenic agents in plants.

VII. Referências Bibliográficas

- ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.** **25**:3389-3402.
- BAIROCH, A. & APWEILER, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Res.** **28**:45-48.
- BISHOP, J. G., DEAN, A. M. & MITCHELL-OLDS, T. 2000. Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. **Proc. Natl. Acad. Sci. USA** **97**: 5322-5327.
- BREITENEDER, H., FERREIRA, F., HOFFMANN-SOMMERGRUBER, K., EBNER, C., BREITENBACH, M., RUMPOLD, H., KRAFT, D. & SCHEINER, O. 1993. Four recombinant isoforms of *CoraI*, the major allergen of hazel pollen, show different IgE-binding properties. **Eur. J. Biochem.** **212**:355-362.
- BREITENEDER, H., PETTENBURGER, K., BITO, A., VALENTA, R., KRAFT, D., RUMPOLD, H., SCHEINER, O & BREITENBACH, M. 1989. The gene coding for the major birch pollen allergen *BetvI*, is highly homologous to a pea disease resistance response gene. **EMBO J.** **8**:1935-1938.
- DARWIN, C. 1859. **On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.** John Murray, London.
- DIXON, R. A. & HARRISON, M. J. 1990. Genomic responses to environmental stress. **Adv. Genet.** **28**:165-233.
- EDENS, L., HESLINGA, L. KLOK, R., LEDEBOER, A. M., MAAT, J., TOONEN, M. Y., VISSER, C. & VERRIPS, C. T. 1982. Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in *Escherichia coli*. **Gene** **18**:1-12.
- ENDO, T., IKEO, T. & GOJOBORI, T. 1996. Large-scale search for genes on which positive selection may operate. **Mol.Biol. Evol.** **13**:685-690.
- EPPLE, P., APEL, K. & BOHLMANN, H. 1995. An *Arabidopsis thaliana* thionin gene is inducible via a signal transduction pathway different from that for pathogenesis-related proteins. **Plant Physiol.** **109**: 813-820.
- ETZOLD T. & ARGOS P. 1993. SRS - an indexing and retrieval tool for flat file data libraries. **Comput. Appl. Biosci.** **9**:49-57.

- FABER, C., LINDENMAN, A., STICHT, H., EJCHART, A., KUNGL, A., SUSANI, M., FRANK, R. W., KRAFT, D., BREITENBACH, M. & ROSCH, P. 1996. Secondary structure and tertiary fold of the birch pollen allergen Bet v 1 in solution. **J. Biol. Chem.** **271**:19243-19250.
- FELSENSTEIN, J. 1993. PHYLIP (Phylogeny Inference Package) User Manual. Department of Genetics, University of Washington, Seattle.
- FLACH, J., PILET, P.-E. & JOLLES, P. 1992. What's new in chitinase research?. **Experientia** **48**:701-716.
- FREEMAN, S. & HERRON, J. C. 1998. **Evolutionary Analysis**. Prentice-Hall. New Jersey.
- GAJHEDE, M., OSMARK, P., POULSEN, F. M., IPSEN, H., LARSEN, J. N., JOOST VAN NEERVEN, R. J., SCHOU, C., LOWENSTEIN, H. & SPANGFORT, M. D. 1996. X-ray and NMR structure of Betv1, the origin of birch pollen allergy. **Nat. Struct. Biol.** **3**:1040-1045.
- GARCÍA-OLMEDO, F., MOLINA, A., SEGURA, A. & MORENO, M. 1995. The defensive role of nonspecific lipid-transfer proteins in plants. **Trends Microbiol.** **3**: 72-74.
- GOLDMAN, N., & YANG, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Mol. Biol. Evol.****11**:725-736.
- GRANELL, A., BELLES, J. M. & CONEJERO, V. 1987. Induction of pathogenesis-related proteins in tomato by citrus exocortis viroid, silver ion and ethephon. **Physiol. Mol. Plant Pathol.** **31**: 83-90.
- GRAUR, D. & LI, W.-H. 2000. **Fundamentals of Molecular Evolution**, second edition. Sinauer Associates, Sunderland, MA.
- GREEN, T. R. & RYAN, C. A. 1972. Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. **Science** **175**: 776-777.
- HAM, K. S., WU, S. C., DARVILL, A. G., ALBERSHEIM, P. 1997. Fungal pathogens secrete an inhibitor protein that distinguishes isoforms of plant pathogenesis-related endo-beta-1,3-glucanases. **Plant J.** **11**:169-179.
- HARBORNE, J. B. 1993. **Introduction to Ecological Biochemistry**. Academic Press, London.

- HELLBERG, M. E., MOY, G. W. & VACQUIER, V. D. 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. **Mol. Biol. Evol.** **17**: 458-466.
- HENRISSAT, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochem. J.** **280**:309-316.
- HODGE, A., GOODAY, D. W. & ALEXANDER, I. J. 1996. Inhibition of chitinolytic activities from tree species and associated fungi **Phytochemistry** **41**:77-84.
- HUGHES, A. L. & NEI, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. **Nature** **335**:167-170.
- HUGHES, A. L. & NEI, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. **Proc. Natl. Acad. Sci. USA** **86**:958-962.
- INA, Y. 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. **J Mol Evol.** **40**:190-226.
- INA, Y., MIZOKAMI, M., OHBA, K., AND GOJOBORI, T. 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J. Mol. Evol.* **38**:50-56.
- JEANMOUGIN, F., THOMPSON, J. D., GOUY, M., HIGGINS, D. G. & GIBSON, T. J. 1998. Multiple sequence alignment with Clustal X. **Trends Biochem. Sci.** **23**:403-405.
- JORDÁ, L., COEGO, A., CONEJERO, V. & VERA, P. 1999. Genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. **J. Biol. Chem.** **274**: 2360-2365.
- JORDÁ, L., CONEJERO, V. & VERA, P. 2000. Characterization of *P69E* and *P69F*, two differentially regulated genes encoding new members of the subtilisin-like proteinase family from tomato plants. **Plant Physiol.** **122**: 67-73
- KAWABE, A., INNAN, H., TERAUCHI, R. & MIYASHITA, N. T. 1997. Nucleotide polymorphism in the acidic chitinase locus (ChiA) region of the wild plant *Arabidopsis thaliana*. **Mol. Biol. Evol.** **14**:1303-15.
- KUMAR, S. TAMURA, K., JAKOBSEN, I. & NEI, M. 2000. **MEGA: Molecular Evolutionary Genetics Analysis, Ver. 2.**, Pennsylvania State University, University Park.

- KUMAR, S., TAMURA, K., JAKOBSEN, I. B. & NEI, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. **Bioinform.** **17**:1244-1245.
- KUNGL, A. J. 1996. Evidence for alpha helical T cell epitope in the C-terminus of the main birch pollen allergen Bet v 1. **Biochem. Biophys. Res. Commun.** **223**:187-192.
- LAGRIMINI, L. M., BURKHART, W., MOYER, M. & ROTHSTEIN, S. 1987. Molecular cloning of complementary DNA encoding the lignin-forming peroxidase from tobacco: molecular analysis and tissue-specific expression. **Proc. Natl. Acad. Sci. USA** **84**: 7542-7546.
- LAMB, C. J., LAWTON, M. A., DRON, M. & DIXON, R. A. 1989. Signals and transduction mechanisms for activation of plant defense against microbial attack. **Cell** **56**:215-224.
- LANE, B. G., DUNWELL, J. M., RAY, J. A., SCHMITT, M. R. & CUMING, A. C. 1993. Germin, a protein marker of early plant development, is an oxalate oxidase. **J Biol Chem** **268**:12239-12242
- LASKOWSKI, M. & KATO, I. 1980. Protein inhibitors of proteinases. **Annu. Rev. Biochem.** **49**:593-626.
- LI, W. H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. **J. Mol. Evol.** **36**:96-9.
- LU, G., VILLALBA, M. COSCIA, M. R., HOFFMAN, D. R. & KING, T. P. 1993. Sequence-analysis and antigenic cross-reactivity of a venom allergen, antigen-5, from hornets, wasps and yellow jackets. **J. Immunol.** **150**:2823-2830.
- LUDVIGSEN, S. & POULSEN, F. M. 1992. Secondary structure in solution of Barwin from barley seed using ¹H nuclear magnetic resonance spectroscopy. **Biochemistry** **31**:8771-8782.
- LUKENS, L. & DOEBLEY, J. 2001. Molecular evolution of the teosinte branched gene among maize and related grasses. **Mol. Biol. Evol.** **18**:627-38.
- MELCHERS, L. S., APOTHEKER-DE GROOT, M., VAN DER KNAAP, J. A., PONSTEIN, A. S., SELA-BUURLAGE, M. B., BOL, J. F., CORNELISSEN, B. J. C., VAN DEN ELZEN, P. J. M. & LINTHORST, H. J. M. 1994. A new class of tobacco chitinases homologous to bacterial exo-chitinases displays antifungal activity. **Plant J.** **5**: 469-480.
- MÉTRAUX, J.-P., STREIT, L. & STAUB, TH. 1988. A pathogenesis-related protein in cucumber is a chitinase. **Physiol. Mol. Plant Pathol.** **33**: 1-9.

- MORTON, B. R. 2001. Selection at the amino acid level can influence synonymous codon usage: Implications for the study of codon adaptation in plastid genes. **Genetics** **159**:347-358.
- MURZIN, A. G., LESK, A. M. & CHOTHIA, C. 1992. Patterns of structure and sequence in the Kunitz inhibitors, interleukins-1 beta and alpha and fibroblast growth factors. **J. Mol. Biol.** **223**:531-543.
- NEI M, GOJOBORI T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. **Mol. Biol. Evol.** **3**:418-26.
- NEI, M. & KUMAR, S. 2000 **Molecular Evolution and Phylogenetics**, Oxford University Press, Oxford.
- NICHOLAS, K. B. & NICHOLAS, H. B. Jr. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments (distribuido pelo autor).
- NIELSEN, R., & YANG, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**, **148**: 929-936.
- ORI, N., SESSA, G., LOTAN, T. HIMMELHOCH, S. & FLUHR, R. 1990. A major stylar matrix polipeptide (sp41) is a member of the pathogenesis-related proteins superclass. **EMBO J.** **9**:3429-3436.
- PAMILO, P. & BIANCHI, N. O. 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. **Mol Biol Evol.** **10**:271-81.
- PARISI, G. & ECHAVE, J. 2001. Structural constraints and emergence of sequence patterns in protein evolution. **Mol Biol Evol** **18**:750-756.
- RAES, J. & VAN DE PEER, Y. 1999. ForCon : a software tool for the conversion of sequence alignments. **EMBnet.news** **6**:10-12. Distributed for free at: <http://vulcan.rug.ac.be/~jerae/ForCon/index.html>.
- REBMANN, G., MAUCH, F. & DUDLER, R. 1991. Sequence of a wheat cDNA encoding a pathogen induced thaumatin-like protein. **Plant Mol. Biol.** **16**
- RUIZ-MENDRANO, R., JIMENES-MORAILA, B., HERRERA-ESTRELLA, L. & RIVERA-BUSTAMANTE, R. F. 1992. Nucleotide sequence of an osmotin-like cDNA induced in tomato during viroid infection. **Plant Mol. Biol.** **20**:1199-1202.
- SAKAMOTO, Y., ISHIGURO, M. & KITAGAWA, G. 1986. **Akaike Information Criterion Statistics**, D. Reidel Pub. Co.

- SELITRENNIKOFF, C. P. 2001. Antifungal proteins. **Appl. Environ. Microbiol.** **67**:2883-2894
- SINGH, N. K., NELSON, D. E., KUHN, D., HASEGAWA, P. M. & BRESSAN, R. A. 1989. Molecular cloning of osmotin and regulation of its expression by ABA and adaptation to low water potential. **Plant Physiol.** **90**:1096-1101.
- SOMSSICH, I.E., SCHMELZER, E., BOLLMANN, J. & HAHLBROCK, K. 1986. Rapid activation by fungal elicitor of genes encoding "pathogenesis-related" proteins in cultured parsley cells. **Proc. Natl. Acad. Sci. USA** **83**: 2427-2430.
- STAHL, E. A., DWYER, G., MAURICIO, R., KREITMAN, M. & BERGELSON, J. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. **Nature** **400**:667-671.
- STOLTZ, H. U., BISHOP, J. G., BERGMANN, C. W., KOCH, M., ALBERSHEIM, P., DARVILL, A. G. & LABAVITCH, J. M. 2000. Identification of target amino acids that affect interactions of fungal polygalacturonases and their plant inhibitors. **Phys Mol Plant Path.** **56**: 117-130.
- SU, C. 2000. **SGI: Modified Suzuki and Gojobori's Method for Detecting Positive and Negative Selection at Individual Codon Sites**. The Pennsylvania State University, University Park, PA, USA.
- SUEOKA, N. & KAWANISHI, Y. 2000. DNA G+C content of the third codon position and codon usage biases of human genes. **Gene.** **261**:53-62.
- SUZUKI, Y., & GOJOBORI, T. 1999. A method for detecting positive selection at single amino acid sites. **Mol Biol Evol** **16**: 1315-1328.
- SUZUKI, Y. & NEI, M. 2001 Reliabilities of Parsimony-based and Likelihood-based Methods for Detecting Positive Selection at Single Amino Acid Sites. **Mol. Biol. Evol.** **18**:2179-2185.
- SVENSSON, B., SVENDSEN, I., HOEJRUP, P., LUDVIGSEN, S. & POULSEN, F. M. 1992. Primary structure of Barwin: a barley seed protein closely related to the C-terminal domain of proteins encoded by wound-induced plant genes. **Biochemistry** **31**:8767-8770.
- SWANSON, W. J., YANG, Z., WOLFNER, M. F. & AQUADRO, C. F. 2001. Positive Darwinian selection in the evolution of mammalian female reproductive proteins. **Proc. Natl. Acad. Sci. U.S.A.** **98**:2509-2514.

- SZYPERSKI, T., FERNANDEZ, C., MUMENTHALER, C. & WUTHRICH, K. 1998. Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system. **Proc Natl Acad Sci U S A** **95**:2262-2266.
- TERRAS, F. R. G., SCHOofs, H., DE BOLLE, M. F. C., VAN LEUVEN, F., REES, S. B., VANDERLEYDEN, J., CAMMUE, B. P. A. & BROEKAERT, W. F. 1992. Analysis of two novel classes of plant antifungal proteins from radish (*Raphanus sativus* L.) seeds. **J. Biol. Chem.** **267**: 15301-15309.
- THOMPSON, J. N. & BURDON, J. J. 1992. Gene-for-gene coevolution between plants and parasites. **Nature** **360**:121-125.
- THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D. G. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucl. Acids Res.** **24**:4876-4882.
- TSUNOYAMA, K. & GOJOBORI, T. 1998. Evolution of nicotinic acetylcholine receptor subunits. **Mol. Biol. Evol.** **15**:518-527.
- UKNES, S., DINCHER, S., FRIEDRICH, F., NEGROTTO, D., WILLIAMS, S., THOMPSON-TAYLOR, H., POTTER, S., WARD, E. & RYALS, J. 1993. Regulation of pathogenesis-related protein-1a gene expression in tobacco. **Plant Cell** **5**:159-169.
- VAN LOON, L. C. 1982. Regulation of changes in proteins and enzymes associated with active defense against virus infection. In: **Active Defense Mechanisms in Plants**. (R. K. S. Wood, ed.), pp. 247-273, Plenum Press, New York, USA.
- VAN LOON, L. C. 1997. Induced resistance and the role of pathogenesis-related proteins. **Eur. J. Plant Pathol.** **103**: 753-765.
- VAN LOON, L. C., PIERPOINT, W. S., BOLLER, T. & CONEJERO, V. 1994. Recommendations for naming plant pathogenesis-related proteins. **Plant Mol. Biol. Reporter** **12**: 245-264.
- VAN LOON, L. C. & VAN STRIEN, E. A. 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. **Physiol. Mol. Plant Pathol.** **55**: 85-97.
- VARGHESE, J. N., GARRETT, T. P. J., COLMAN, O. M., CHEN, L., HOJ, P. J. & FINCHER, G. B. 1994. Three-dimensional structures of two plant beta-glucan

- endohydrolases with distinct substrate specificities. **Proc. Natl. Acad. Sci. USA** **91**:2785-2789.
- VERA, P. & CONEJERO, V. 1988. Pathogenesis-related proteins of tomato. P-69 as an alkaline endoproteinase. **Plant Physiol.** **87**: 58-63.
- YAMADA, M. 1992. Lipid transfer proteins in plants and microorganisms. **Plant Cell Physiol.** **33**:1-6
- YANG, Z. 1997. **Phylogenetic Analysis by Maximum Likelihood (PAML), version 1.4.** Department of Integrative Biology, University of California, Berkeley.
- YANG, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. **Mol. Biol. Evol.****15**: 568-573.
- YANG, Z. 2000. **Phylogenetic Analysis by Maximum Likelihood (PAML), Version 3.0.** University College London, London, England.
- YANG, Z. 2001. Maximum likelihood analysis of adaptive evolution in HIV-1 gp120 env gene. **Pacific Symposium on BioComputing 2001** pages 226-237.
- YANG, Z., NIELSEN, R., GOLDMAN, N. & PEDERSEN, A.-M. K. 2000a. Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics** **155**: 431-449.
- YANG, Z., SWANSON, W. J & VACQUIER, V. D. 2000b. Maximum likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. **Mol. Biol. Evol.** **17**: 1446-1455.
- YANG, Z. & SWANSON, W. J. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. **Mol. Biol. Evol.** **19**:49-57.