**Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"**

**Caracterização evolutiva das serina peptidases digestivas em
insetos holometábolos**

Renata de Oliveira Dias

Tese apresentada para a obtenção do título de
Doutora em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2014**

Renata de Oliveira Dias
Licenciada em Ciências Biológicas

**Caracterização evolutiva das serina peptidases digestivas em insetos
holometábolos**

Orientador:
Prof. Dr. **MARCIO DE CASTRO SILVA FILHO**

Tese apresentada para a obtenção do título de
Doutora em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2014**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Dias, Renata de Oliveira

Caracterização evolutiva das serina peptidases digestivas em insetos
holometábolos / Renata de Oliveira Dias.- - Piracicaba, 2014.
106 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Tripsina 2. Quimotripsina 3. Díptera 4. Coleóptera 5. Lepidóptera 6. Protease
I. Título

CDD 595.701
D541c

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte -O autor"

*Com amor
aos meus maiores incentivadores
Marino, Lizete e Adriano
Dedico*

AGRADECIMENTOS

Ao professor Dr. Marcio de Castro Silva Filho pela oportunidade, confiança e orientação;

Ao Dr. Marcelo Mendes Brandão, pela inestimável colaboração e ensinamentos;

À professora Dra. Anna Tramontano pela oportunidade e supervisão do meu estágio de pesquisa no exterior;

À Dra. Allegra Via pela contribuição na execução e discussão desse trabalho;

Ao professor Dr. Walter Ribeiro Terra pela discussão dos resultados;

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela concessão de bolsa de Doutorado e de Estágio de Pesquisa no Exterior;

Aos meu pais (Marino e Lizete) e meu irmão (Adriano), por toda a dedicação e amor;

Aos colegas do Laboratório de Biologia Molecular de Plantas, Marcelo Brandão, Thais Souza, Wellington Campos, Elaine Castelhana, Flávia Franco, Larissa Spoladore, Lucas Lopes, Marcos Araújo, Horácio Montenegro, Aline Borges, Karina Lopes, Fabiana Mingossi e Poliene Costa;

Aos demais amigos de Piracicaba, em especial a Mauro, Anna, Tania e Edemir Bellotti e Enéas Konzen;

Ao técnico do Laboratório de Biologia Molecular de Plantas, Rafael Colombi, pela dedicação e organização do laboratório;

Aos amigos distantes, mas sempre presentes, Cíntia Medeiros, Tiago Trento, Aliciane Fogaça, Daniela David, Juliana Girelli, Gabriel Girardi, Ana Paula Colonetti, Luzia Colonetti, Adérico Pimentel, Ludmilla Oliveira, Gustavo Pereira, Maria Carolina Camargo, dentre tantos outros;

Aos colegas do Laboratório de Biologia Computacional da Sapienza Università de Roma;

Aos meus familiares, em especial à minha tia Dóris, com quem sempre pude contar;

A todos que contribuíram, direta ou indiretamente, muito obrigada!

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS.....	13
LISTA DE TABELAS	15
LISTA DE SIGLAS.....	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Relação entre os grupos de insetos estudados	23
2.2 Composição enzimática no sistema digestivo de insetos	25
2.3 Tripsinas e tripsinogênio.....	27
2.4 Quimotripsinas e quimotripsinogênio	28
2.5 Peptidases de insetos herbívoros versus inibidores de peptidases de plantas ..	29
3 MATERIAL E MÉTODOS	33
3.1 Relação entre os grupos de insetos estudados	33
3.2 Análises filogenéticas	38
3.3 Definição das regiões proteicas estudadas.....	38
3.4 Análises de pressão de seleção.....	39
3.5 Modelagem das proteases de insetos.....	41
4 RESULTADOS	45
4.1 Inferência filogenética para tripsinas e quimotripsinas.....	45
4.2 Caracterização do sítio de ligação ao substrato das tripsinas	50
4.3 Caracterização do sítio de ligação ao substrato das quimotripsinas.....	54
4.4 Caracterização do motivo de ativação de tripsinas e quimotripsinas.....	56
4.6 Composição e conservação da região de peptídeo sinal.....	58
4.5 Tripsinas e quimotripsinas e o seu putativo modo de interação com sBBI	59
5 DISCUSSÃO	65
5.1 Caracterização evolutiva geral das tripsinas e quimotripsinas.....	65
5.2 História evolutiva das tripsinas	66
5.3 Auto-ativação do tripsinogênio	69
5.4 O papel de outras regiões proteicas na história evolutiva das tripsinas de Lepidóptera.....	70
5.5 História evolutiva divergente entre tripsinas e quimotripsinas.....	71
6 CONSIDERAÇÕES FINAIS	75

REFERÊNCIAS.....77

ANEXOS87

RESUMO

Caracterização evolutiva das serina peptidases digestivas em insetos holometábolos

Tripsinas e quimotripsinas são classes de serina peptidases amplamente estudadas e fortemente responsáveis pela digestão proteica, pela clivagem de ligações peptídicas no lado carboxila de L-aminoácidos de cadeia lateral básica e hidrofóbica, respectivamente. Três processos regulam finamente a ação dessas peptidases: secreção, ativação do precursor (zimogênio) e o sítio de reconhecimento do substrato. No presente trabalho é apresentada uma análise filogenética detalhada das tripsinas e quimotripsinas de três ordens de insetos holometábolos, revelando características divergentes nas enzimas de Lepidóptera em relação a Coleóptera e Díptera. Em particular, o sub-sítio S1 das tripsinas foi observado como mais hidrofílico em Lepidóptera do que em Coleóptera e Díptera, enquanto os sub-sítios S2-S4 parecem mais hidrofóbicos, sugerindo diferentes preferências pelo substrato. Além disso, Lepidóptera mostrou um grupo de tripsinas bastante específico a um grupo taxonômico, compreendendo somente proteínas de espécies da família Noctuidae. Evidências de eventos de auto-ativação facilitada foram também observadas em todas as ordens de insetos estudadas, com as características do motivo de ativação do zimogênio complementárias ao sítio ativo das tripsinas. Em contraste, as quimotripsinas de insetos não parecem ter uma história evolutiva peculiar com respeito a, por exemplo, seus homólogos em mamíferos. Em geral, os presentes resultados sugerem que a necessidade de uma rápida taxa de auto-ativação fez os insetos holometábolos selecionarem grupos especializados de tripsinas com altas taxas de auto-ativação e também destacam que a evolução das tripsinas culminou em um grupo especializado de enzimas em Lepidóptera.

Palavras-chave: Tripsina; Quimotripsina; Díptera; Coleóptera; Lepidóptera; Protease

ABSTRACT

Evolutionary characterization of digestive serine peptidases in holometabolous insects

Trypsins and chymotrypsins are well-studied classes of serine peptidases largely responsible for the digestion of proteins by cleavage of the peptide bond at the carboxyl side of basic and hydrophobic L-amino acids, respectively. Three processes mainly regulate the action of these peptidases: secretion, precursor (zymogen) activation and substrate-binding site recognition. In the present work is presented a detailed phylogenetic analysis of trypsins and chymotrypsins in three orders of holometabolous insects revealing divergent characteristics in the Lepidoptera enzymes in relation to Coleoptera and Diptera. In particular, trypsin subsite S1 was observed to be more hydrophilic in Lepidoptera than in Coleoptera and Diptera, whereas subsites S2-S4 appeared more hydrophobic, suggesting different substrate preferences. Furthermore, Lepidoptera displayed a very specific taxonomic trypsin group, only encompassing proteins from the Noctuidae family. Evidences for facilitated trypsin auto-activation events were also observed in all the insect orders at hand, with the characteristic zymogen activation motif complementary to the trypsin active site. In contrast, insect chymotrypsins did not seem to have a peculiar evolutionary history with respect to e.g. their mammal counterparts. Overall, the present findings suggest that the need for fast digestion made holometabolous insects evolve specialized groups of trypsins with high auto-activation rates and highlight that the evolution of trypsins culminated in a specialized group of enzymes in Lepidoptera.

Keywords: Trypsin; Chymotrypsin; Diptera; Coleoptera; Lepidoptera; Protease

LISTA DE FIGURAS

Figura 1 - Esquema mostrando as diferentes etapas envolvidas na ação das peptidases. O motivo de ativação do zimogênio descrito na figura corresponde ao consenso encontrado em tripsinas de mamíferos.	20
Figura 2 – Árvore representando a melhor estimativa das relações entre insetos, baseada na literatura atual. As linhas pontilhadas indicam relações suportadas tenuamente ou possíveis monofilias (em caso de braços terminais)	24
Figura 3 – Estrutura covalente e pontes dissulfeto do inibidor do tipo Bowman-Birk de soja.	31
Figura 4 – Estrutura cristalizada (PDB: 3RU4) mostrando o inibidor de serina peptidase do tipo Bowman-Birk (vermelho) de <i>Vigna unguiculata</i> em complexo com uma quimotripsina (verde) e uma tripsina (azul)	31
Figura 5 – Alinhamento das tripsinas de <i>Spodoptera frugiperda</i> e <i>Alabama argillacea</i> a tripsina bovina (PDB: 4I8H).....	34
Figura 6 - Alinhamento das quimotripsinas de <i>Spodoptera frugiperda</i> e <i>Alabama argillacea</i> a quimotripsina de <i>Solenopsis invicta</i> (Insecta – PDB: 1EQ9)	35
Figura 7 – Árvores provenientes de análises de Neighbor-Joining usadas na classificação das tripsinas das ordens: Coleóptera, Díptera e Lepidóptera	36
Figura 8 – Árvores provenientes de análises de Neighbor-Joining usadas na classificação das quimotripsinas das ordens: Coleóptera, Díptera e Lepidóptera	37
Figura 9 – As regiões das peptidases utilizadas nesse trabalho.....	39
Figura 10 – Inferência filogenética Bayesiana das sequências proteicas e os grupos de tripsinas para: (a) Coleóptera – C, (b) Díptera – D e (c) Lepidóptera – L.....	46
Figura 11 – Inferência filogenética Bayesiana das sequências proteicas e os grupos de quimotripsinas de: (a) Coleóptera – C, (b) Díptera – D e (c) Lepidóptera – L.....	47
Figura 12 – Proposição filogenética Bayesiana para as sequências de tripsinas de Coleóptera, Díptera e Lepidóptera	48
Figura 13 – Proposição filogenética Bayesiana para as sequências de quimotripsinas de Coleóptera, Díptera e Lepidóptera	49
Figura 14 – Consenso do sítio de ligação ao substrato dos grupos de tripsinas de Coleóptera (C), Díptera (D) e Lepidóptera (L)	50
Figura 15 – Padrão de hidrofobicidade no sítio de ligação ao substrato das tripsinas dos grupos L-I (a) e L-X (b) de Lepidóptera	52
Figura 16 – Sítios do sítio de ligação ao substrato encontrados sob pressão de seleção nas tripsinas de Lepidóptera.....	53

Figura 17 – Consenso do sítio de ligação ao substrato dos grupos de quimotripsinas de Coleóptera (C), Díptera (D) e Lepidóptera (L).....	55
Figura 18 – Sequência consenso na região de ativação dos grupos de similaridade de (a) tripsinogênio e (b) quimotripsinogênio de Coleóptera, Díptera e Lepidóptera..	57
Figura 19 – Alinhamento das tripsinas de <i>S. frugiperda</i> com as tripsinas de <i>Bos taurus</i> (BovTry – PDB: 4l8H) e <i>Fusarium oxysporum</i> (FoTry – PDB: 1pq7)	60
Figura 20 – Alinhamento das quimotripsinas de <i>S. frugiperda</i> com as quimotripsinas de <i>Bos taurus</i> (BovChy – PDB: 1YPH) e <i>Solenopsis invicta</i> (SiChy – PDB: 1EQ9)	61
Figura 21 – Superposição dos resultados da docagem da tripsina LSfru6 com o <i>loop</i> 1 do sBBI e da quimotripsina LcSfru2 ao <i>loop</i> 2 do sBBI à estrutura cristalizada 3RU4.....	62
Figura 22 – Frequência observada de resíduos em interação hidrofóbica com o sBBI, nas análises de docagem estrutural das tripsinas de Coleóptera e Lepidóptera.....	63
Figura 23 – Filogenia das tripsinas da ordem Lepidóptera, destacando as sequências proteicas provenientes de espécies da família Noctuidae (verde)	68
Figura 24 – Representação gráfica das posições estruturais conservadas na superfície de tripsinas e quimotripsinas das diferentes ordens de insetos estudadas	73

LISTA DE TABELAS

Tabela 1 – Parâmetros utilizados nos modelos de substituição por sítio.....	41
Tabela 2 – Sequência consenso para o peptídeo sinal dos grupos de tripsinas de Coleóptera, Díptera e Lepidóptera	58

LISTA DE SIGLAS

AIC – *Akaike Information Criterion*

BEB – *Bayes Empirical Bayes*

BTCl – Inibidor de serina peptidase do tipo Bowman-Birk de *Vigna unguiculata*

EC – *Enzyme Commission number*

gi – *GenInfo Identifier*

IP – Inibidor de peptidase

K_{cat} – Constante catalítica

K_m – Constante de Michaelis

M0 – Modelo de seleção de taxa única

M1a – Modelo de seleção neutro

M2a – Modelo de seleção positiva

M3 – Modelo de seleção discreto

M7 – Modelo de seleção com distribuição beta

M8 – Modelo de seleção com distribuição beta e omega

MEROPS – *Peptidase Database*

ML – Máxima verossimilhança

PDB – *Protein Data Bank*

sBBI – Inibidor de peptidase de soja do tipo Bowman-Birk

SE – *Single-end reads*

SP – Serina peptidase

ω – Medida de pressão de seleção, pela razão dN/dS, ou seja, substituição não-sinônima/sinônima

1 INTRODUÇÃO

Os insetos compreendem um dos grupos mais abrangentes de organismos no mundo e muito desse sucesso é devido a sua habilidade de explorar diversos tipos de ecossistemas e fontes de energia. No entanto, apesar da ampla diversidade de hábitos alimentares, mesmo em famílias ou espécies de insetos proximamente correlacionadas, a composição do conjunto de enzimas digestivas foi observada como sendo frequentemente específica a ordens, i.e. grupos taxonômicos maiores do que espécies e famílias (TERRA; FERREIRA, 1994). Entretanto, ainda não está claro se processos de divergência evolutiva ou pressão de seleção exercida por diferentes hábitos alimentares levaram essas enzimas a mudanças na especificidade ao substrato e/ou mecanismo de ação ao longo da evolução dos insetos.

Enzimas digestivas de insetos são compostas principalmente por hidrolases, compreendendo entre outras classes, as peptidases (também conhecidas como proteases – EC 3.4), as quais quebram cadeias polipeptídicas em aminoácidos livres, pela catálise de ligações peptídicas. Aproximadamente um terço das peptidases conhecidas são serina peptidases (SPs), as quais possuem um resíduo de serina nucleofílico no seu sítio ativo (HEDSTROM, 2002). Uma das mais importantes famílias de SPs compreende enzimas tais como tripsinas, quimotripsinas e elastases, cuja a tríade catalítica é composta por um resíduo de ácido aspártico, um de histidina e um de serina.

Tripsinas (EC 3.4.21.4) são conhecidas como enzimas de especificidade restrita, uma vez que elas clivam ligações peptídicas preferencialmente no lado carboxila de somente dois tipos de aminoácidos: lisina e arginina (CRAIK et al., 1985). Enquanto isso, quimotripsinas (EC 3.4.21.1) apresentam uma maior variabilidade em seus substratos preferenciais, os quais incluem ligações peptídicas no lado carboxila de resíduos de: fenilalanina, tirosina, triptofano e leucina (HEDSTROM; SZILAGYI; RUTTER, 1992).

Outra propriedade relevante dessas SPs é que elas necessitam passar por um processo de ativação (Figura 1). Tripsinogênios e quimotripsinogênios, os quais compreendem os zimogênios (formas inativas) de tripsinas e quimotripsinas, respectivamente, são ativados pela clivagem de uma sequência N-terminal de

tamanho variável. A presença de um aminoácido básico (arginina ou lisina) imediatamente anterior ao resíduo de isoleucina inicial das peptidases faz dessa região um substrato potencial para as próprias tripsinas. De fato, uma grande parte dos zimogênios das serina peptidases (incluindo os quimotripsinogênios) são ativados por tripsinas (RINDERKNECHT, 1986), o que, portanto, dá a essas enzimas um papel chave na cascata de ativação de diversas peptidases.

Figura 1 - Esquema mostrando as diferentes etapas envolvidas na ação das peptidases. O motivo de ativação do zimogênio descrito na figura corresponde ao consenso encontrado em tripsinas de mamíferos. As posições dos resíduos do substrato (P2', ..., P3) e dos subsítios da enzima (S2', ..., S3) seguem a nomenclatura proposta por Schechter e Berger (1968)

Entretanto, em alguns organismos, este processo de ativação do tripsinogênio é finamente regulado por diferentes enzimas. A regulação da ativação das tripsinas por enzimas especializadas, mesmo sendo mais lento em termos de suplemento de enzimas funcionais, traz como benefício o decréscimo no risco de efeitos indesejáveis sobre alvos inespecíficos, devido ao excesso de tripsinas ativas (KITAMOTO et al., 1994). Em mamíferos, a deficiência na função da enteroquinase e mutações na região de ativação do tripsinogênio, as quais causam aumento nas taxas de auto-ativação das tripsinas, são conhecidas como causas de pancreatites (HADORN et al., 1969). No entanto, sabe-se que a sequência do sítio de ativação do tripsinogênio de insetos, crustáceos, outros invertebrados e fungos diverge do consenso presente na região correspondente ao de vertebrados, sugerindo a

possibilidade de que possa ocorrer a auto-ativação nas enzimas desses organismos (CHEN et al., 2003).

No entanto, apesar da importância das tripsinas e quimotripsinas na digestão de insetos, o processo de ativação do seus respectivos zimogênios ainda não está claramente entendido. Mais genericamente, a região de pré-ativação como um todo, incluindo o peptídeo sinal e a via de secreção dessas enzimas, está ainda pobremente elucidada.

Esse trabalho objetiva, portanto, entender a história evolutiva das tripsinas e quimotripsinas e seus respectivos precursores inativos em insetos holometábolos, i.e., insetos que passam por metamorfose completa, focando-se no sítio de ligação ao substrato, o motivo de ativação do zimogênio e o peptídeo sinal, nas ordens Coleóptera, Díptera e Lepidóptera. Esse estudo está baseado na análise e comparação das histórias evolutivas dessas duas enzimas.

2 REVISÃO BIBLIOGRÁFICA

2.1 Relação entre os grupos de insetos estudados

Os insetos formam um dos táxons mais abrangentes da Terra, possuindo um grande número de indivíduos e espécies, os quais são observados explorando diversos ambientes e recursos.

Estima-se que aproximadamente 80% da riqueza de espécies de insetos esteja concentrada em quatro super radiações: besouros (Coleóptera: Coleoptera), moscas (Díptera: Diptera), abelhas e vespas (Himenóptera: Hymenoptera), e mariposas e borboletas (Lepidóptera: Lepidoptera); todos eles holometábolos (TRAUTWEIN et al., 2012).

A mais recente e amplamente aceita filogenia de Holometabola divide o grupo em duas principais linhagens: Neuropteroidea, incluindo a ordem Coleóptera entre outras, e Mecopteridea, a qual compreende as ordens Lepidóptera e Díptera, entre outras (TRAUTWEIN et al., 2012) (Figura 2).

Dentro dos insetos, algumas espécies herbívoras vêm recebendo substancial atenção devido a sua associação a grandes perdas nas lavouras e, conseqüentemente, afetando a produção de suprimentos agrícolas em todo o mundo (WATERFIELD; ZILBERMAN, 2012; MARTIN et al., 2013; OLIVEIRA et al., 2013; OLIVEIRA et al., 2014).

A ordem Lepidóptera, em particular, representa um dos mais importantes grupos de insetos herbívoros. Composta de borboletas e mariposas e com mais de 160.000 espécies descritas, a ordem compreende famílias bastante conhecidas, como a Noctuidae, maior família da ordem, com mais de 35 mil espécies descritas (NEW, 2004). Uma das espécies de maior relevância agrônômica do grupo é uma importante praga nas Américas, conhecida como lagarta-do-cartucho (*Spodoptera frugiperda*), relatada em uma ampla gama de culturas de diferentes famílias de plantas, incluindo mono e dicotiledôneas (BARROS et al., 2010). Outros importantes membros da família são *Heliothis virescens*, capaz de usar hospedeiros de 14 famílias de dicotiledôneas (WALDVOGEL; GOULD, 1990) e *Alabama argillacea*, uma importante praga do algodoeiro, especialista na família Malvaceae. A diversidade no comportamento alimentar de espécies tão próximas evolutivamente é alvo de estudos, mas pouco ainda se sabe sobre os mecanismos moleculares que permitiram essa adaptação, uma vez que grande parte das características do

sistema digestório desses indivíduos parece estar associada ao grupo taxonômico amplo (Ordem) ao qual eles pertencem (TERRA; FERREIRA, 1994).

Figura 2 – Árvore representando a melhor estimativa das relações entre insetos, baseada na literatura atual. As linhas pontilhadas indicam relações suportadas tenuamente ou possíveis monofilias (em caso de braços terminais). Os tipos de dados que suportam cada nó são mostrados, identificando se um nó foi observado por uma linha de evidência particular sozinha ou por análises combinadas. Dados de filogenômica se referem a conjuntos de dados moleculares de pelo menos 20kb, para dados coletados por sequenciamento de EST (*expressed sequence tag*), ou por comparação de genomas em larga escala. Na figura, as abreviações correspondem a: mtDNA, DNA mitocôndrial, rDNA, DNA ribossomal, Amph., Amphiesmenoptera; Coleop., Coleopterida; Neurop.; Neuropterida; Psoco., Psocodea; Xeno., Xenonomia.

Fonte: Extraído e adaptado de Trautwein et al. (2012).

2.2 Composição enzimática no sistema digestivo de insetos

As enzimas digestivas de insetos são compostas por hidrolases, tais como as peptidases (EC 3.4), as quais são responsáveis pela clivagem de ligações peptídicas. A digestão de proteínas é uma etapa crucial para a nutrição dos organismos. As cadeias polipeptídicas ingeridas precisam ser hidrolisadas em aminoácidos livres, os quais serão absorvidos pelo intestino e fornecerão substrato para a formação de novas proteínas.

A digestão de proteínas é realizada em duas etapas, sendo a inicial realizada pela ação das endopeptidases e a final pelas exopeptidases. As endopeptidases, responsáveis pela despolimerização das proteínas, são classificadas em pelo menos quatro principais classes: serina, cisteína, aspártico e metalo peptidases, de acordo com o mecanismo catalítico utilizado por elas (RAWLINGS; BARRET, 1993). Enquanto isso, as exopeptidases, responsáveis pela quebra de oligômeros e dímeros e a disponibilização de aminoácidos livres, englobam as aminopeptidases, carboxipeptidases e dipeptidases.

As proporções de diferentes peptidases no interior do intestino dos insetos podem alterar de acordo com o pH local, uma característica variável entre diferentes grupos de insetos (TERRA; FERREIRA, 1994). Um exemplo disso é a observação de que serina peptidases são encontradas em maior número nos pHs de neutro à alcalino (como o presente na ordem Lepidóptera) e as cisteína e aspártico peptidases em maior volume nos pHs mais ácidos (tais como os encontrados em Coleóptera) (TERRA; FERREIRA, 1994).

Aproximadamente um terço de todas as peptidases são classificadas como serina peptidases (SPs) (EC 3.4.21), as quais apresentam um resíduo de serina nucleofílico no seu sítio ativo (HEDSTROM, 2002). SPs são associadas em diversos grupos de organismos a uma ampla gama de importantes funções fisiológicas, tais como digestão, ativação de zimogênios, desenvolvimento e defesa (NEURATH, 1984).

Um dos grupos mais amplamente estudados de SPs compreende a família S1 (também chamada de família das quimotripsinas), a qual engloba enzimas tais como quimotripsinas, tripsinas e elastases (RAWLINGS; BARRET, 1993). Essas enzimas possuem em comum a presença de uma tríade catalítica composta por resíduos de ácido aspártico, histidina e serina, nas posições 102, 57 e 195, respectivamente

(numeração de acordo com a quimotripsina bovina (Uniprot: P00766), aqui e ao longo do trabalho). Além disso, essas enzimas possuem diferenças estruturais e na composição de aminoácidos em seus sítios, as quais permitem a sua especificidade a diferentes substratos e, por isso, são necessárias para que as peptidases exerçam diferentes funções no organismo (NERATH, 1984; PERONA; CRAIK, 1995).

As proteínas da família das quimotripsinas geralmente são compostas por três domínios: a) catalítico, b) de reconhecimento do substrato e c) de ativação do zimogênio. Esses domínios são interligados pela necessidade de muitas das suas características estruturais em todos os processos descritos (HEDSTROM, 2002).

A estrutura proteica dessas enzimas é bastante similar entre os membros da família, os quais possuem dois β -barris muito similares, compostos por seis folhas betas ligadas entre si e uma α -hélice C-terminal (BAPTISTA et al., 1998).

De acordo com Schechter e Berger (1968), as posições do substrato contatadas pelo sítio de ligação enzimático são numeradas como P1, P2, ..., Pn da região C- à N-terminal, e P1', P2', ..., Pn' da região N- à C-terminal. O sítio de ligação ao substrato está localizado entre as posições P1 e P1'. Cada uma dessas regiões do substrato pode contatar vários resíduos no sítio de ligação enzimático. O conjunto de resíduos da enzima contatados por uma única posição do substrato é chamado "sub-sítio" e indicado como S1, S2, ..., Sn na direção C- à N-terminal, e S1', S2', ..., Sn' da região N- à C-terminal. Por exemplo, o sub-sítio S1 na elastase pancreática bovina inclui Val190, Phe192, Ala213, Val216, Phe228, Cys191, Cys220, Gly193, e Ser195 e está contatado pelo resíduo do substrato P1 (BODE; MEYER; POWERS, 1989; HEDSTROM, 2002). O sub-sítio S2 está contatado pela posição P2, e assim por diante.

Outra importante questão em relação a atividade das SPs, é o requerimento dessas enzimas serem secretadas para o meio extracelular para agir na digestão. As vias de secreção de enzimas digestivas de insetos são pouco entendidas ainda e a composição do peptídeo sinal das tripsinas não é um tema inexplorado. No entanto, sabe-se que peptídeos sinais tendem a ser mais similares em genes próximos e drasticamente diferentes em ortólogos distantes (VEITIA; CABURET, 2009).

Como observado em outras famílias gênicas, eventos de duplicação são especulados como sendo os responsáveis pela expansão da família das serina

peptidases e seus homólogos (ROSS et al., 2003; ZHAO et al., 2010). Estudos em insetos vêm demonstrando uma alta representatividade de genes da família das SPs, como demonstrado em *Anopheles gambiae* (ZDOBNOV et al., 2002), *Bombyx mori* (ZHAO et al., 2010) e *Drosophila melanogaster* (ROSS et al., 2003), com 306, 143 e 204 genes relacionados a SPs, respectivamente. Além disso, clusters gênicos compostos por genes similares a tripsinas foram preditos em *A. gambiae* como envolvidos na adaptação desses insetos à hematofagia, por serem induzidos em conjunto, mediante a ingestão de sangue (MULLER et al., 1993).

2.3 Tripsinas e tripsinogênio

As tripsinas (EC 3.4.21.4) compreendem um grupo de SPs, especializado na clivagem de ligações peptídicas no lado carboxila de L-aminoácidos básicos (CRAIK et al., 1985). Além disso, em vários grupos de organismos, as tripsinas são preditas por preferencialmente clivarem ligações peptídicas no lado carboxila de resíduos de arginina mais do que de lisina (CRAIK et al., 1985). Em tripsinas pancreáticas de ratos, a atividade catalítica (mensurada em termos de k_{cat}/K_m) foi encontrada como sendo 11 vezes maior sobre ligações adjacentes à arginina do que à lisina (CRAIK et al., 1985). Entretanto, essa tendência não foi observada para todos os grupos de insetos. Espécies da ordem Lepidóptera não apresentam uma clara preferência por resíduos de Arg ou Lys nessa região (LOPES et al., 2004). Além disso, esses insetos apresentam uma maior preferência por resíduos hidrofóbicos em alguns subsítios em relação a outros grupos de insetos, tais como Díptera e Coleóptera, mostrando que existem algumas mudanças na evolução do sítio de ligação das tripsinas de insetos (LOPES et al., 2006).

Tripsinas são sintetizadas como tripsinogênio, um precursor inativo contendo uma extensão N-terminal, a qual deve ser removida. A remoção desse seguimento permite a formação de uma ponte salina entre o resíduo terminal de isoleucina (posição 16) e o resíduo de ácido aspártico (posição 194) na região do sítio catalítico. Após a formação da ponte salina, ocorre a indução de uma mudança de conformação no sítio de ligação enzimático, disparando a atividade catalítica e estabelecendo as taxas normais de atividade triptica (HEDSTROM, 2002). A presença de Lys/Arg adjacente a Ile16 faz dessa região um alvo potencial para a clivagem mediada por outras tripsinas. Entretanto, como uma maior concentração de peptidases no organismo pode ser perigoso, devido a possibilidade dessas enzimas

agirem sobre proteínas não-alvos, em alguns organismos várias etapas de regulação surgiram ao longo da evolução, dentre as quais, a dependência de complexas cascatas de ativação (KITAMOTO et al., 1994). Em mamíferos, a auto-ativação de tripsinas ocorre a taxas muito baixas e a maior parte da ativação é produzida pela ação de enteroquinases específicas.

A enteroquinase suína ativa o tripsinogênio bovino mais eficientemente do que a tripsina bovina o faz (K_m 6 vezes menor e k_{cat} 2000 vezes maior) (MAROUX; BARATTI; DESNUELLE, 1971). Isto porque, a enteroquinase possui especificidade não apenas pela ligação peptídica que cliva, mas também reconhece um motivo de quatro resíduos de ácido aspártico consecutivos, os quais estão adjacentes ao sítio de clivagem e são altamente conservados nos tripsinogênios de mamíferos (MAROUX; BARATTI; DESNUELLE, 1971). Vale a pena mencionar que, em humanos, a deficiência congênita de enteroquinase causa severas desordens pancreáticas (HADORN et al., 1969), e que mutações no motivo de ativação do zimogênio, tais como D22G (TEICH et al., 2000) ou K23R (FEREC et al., 1999), estão associadas com altas taxas de ativação do zimogênio, as quais são conhecidas como causas de pancreatites.

As sequências do motivo de ativação de tripsinogênios em insetos e em alguns outros invertebrados foram observadas como divergentes das encontradas em mamíferos (CHEN et al., 2003). Essas sequências englobam inclusive algumas das mutações conhecidas por aumentarem as taxas de auto-ativação em mamíferos, induzindo a ideia de que a auto-ativação é o processo mais provável para essas enzimas (CHEN et al., 2003). Em especial, também como observado em *Drosophila melanogaster*, a ausência dos quatro resíduos de ácido aspártico consecutivos na região de ativação do zimogênio sugere que pode não haver uma enzima específica (tal como a enteroquinase) envolvida na ativação das tripsinas de insetos e que, portanto, a auto-ativação seja o mecanismo normal de ativação dessas proteínas (CHEN et al., 2003).

2.4 Quimotripsinas e quimotripsinogênio

Quimotripsinas (EC 3.4.21.1) apresentam especificidade sobre ligações peptídicas no lado carboxila de aminoácidos hidrofóbicos de cadeia lateral larga. Nesse sentido, as quimotripsinas apresentam maior variabilidade em seus substratos preferenciais, incluindo ligações peptídicas no lado carboxila de resíduos

de: fenilalanina, tirosina, triptofano e leucina (HEDSTROM; SZILAGYI; RUTTER, 1992).

Assim como as quimotripsinas, as elastases também apresentam preferência por resíduos hidrofóbicos. No entanto, no caso das elastases, a preferência se dá por resíduos de cadeia lateral curta. A principal diferença entre estas duas enzimas é a composição dos resíduos na entrada do sítio catalítico de ambas. Enquanto as quimotripsinas (e tripsinas) apresentam tipicamente duas glicinas nas posições 216 e 226, permitindo a entrada de aminoácidos de cadeia lateral maior no sub-sítio S1, as elastases apresentam duas valinas (ou uma valina e uma treonina) nessas posições, limitando a passagem desses resíduos (PERONA; CRAIK, 1995).

Estudos de atividade enzimática, mostraram que quimotripsinas pertencentes a diferentes ordens de insetos não apresentam variações peculiares na especificidade pelo substrato: com todas as enzimas analisadas possuindo preferência por clivar ligações peptídicas no lado carboxila de resíduos de tirosina e não mostraram nenhuma tendência clara em aumento ou decréscimo de hidrofobicidade nos diferentes sub-sítios analisados (SATO et al., 2008).

2.5 Peptidases de insetos herbívoros versus inibidores de peptidases de plantas

Insetos herbívoros estão associados a grandes perdas nas lavouras de diversas culturas. Durante anos de evolução, as plantas desenvolveram várias estratégias para resistirem aos ataques sofridos pelos insetos, dentre as quais estão a liberação de voláteis, a produção de inibidores de peptidase, inibidores de alfa-amilase, polifenol oxidases, peptidases e proteínas associadas à via de defesa (FALCO et al., 2001). Enquanto alguns desses compostos, como os voláteis, atuam na mudança do comportamento dos insetos na procura por alimentos, outros atuam dificultando a sua digestão, provocando danos ao desenvolvimento e, conseqüentemente, à sobrevivência do animal, como os inibidores de peptidases (MELLO; SILVA-FILHO, 2002).

Analisados como fontes de resistência a insetos herbívoros, os inibidores de peptidase (IPs) estão presentes constitutivamente nas plantas, mas têm sua expressão aumentada em função da presença de herbivoria, danos mecânicos ou exposição a fitohormônios exógenos (RAKWAL; KUMAR AGRAWAL; JWA, 2001). Diferentes IPs de plantas vêm sendo associados a danos no desenvolvimento larval

de insetos herbívoros, agindo como poderosos efetores anti-nutricionais (RYAN et al., 1990; BOULTER, 1993; BROADWAY, 1995; WOLFSON; MURDOCK, 1995; SRINIVASAN et al., 2005).

Esse mecanismo anti-herbivoria é bem caracterizado (GREEN; RYAN, 1972; JONGSMA et al., 1995), sendo conhecido que a efetividade desses inibidores está intimamente correlacionada à sua afinidade, ou especificidade, pelas peptidases intestinais dos insetos alvos (BURGESS et al., 1991; GATEHOUSE et al., 1993; MCMANUS; WHITE; MCGREGOR, 1994).

Apesar da eficiência dos IPs na proteção das plantas contra ataques de herbívoros, esses animais respondem aumentando a expressão de peptidases (BROADWAY, 1996; PAULILLO et al., 2000; BRITO et al., 2001; BOWN; WILKINSON; GATEHOUSE, 2004; BRIOSCHI et al., 2007). O lepidóptera *Spodoptera frugiperda*, um inseto altamente polífago, mostra capacidade em contornar os efeitos deletérios dos IPs de soja, por uma capacidade intrínseca de sintetizar *de novo* enzimas do tipo tripsina e de aumentar a secreção de enzimas sintetizadas antes do ataque (PAULILLO et al., 2000; BRIOSCHI et al., 2007).

Essas novas enzimas são observadas como sensíveis (DE LEO et al., 1998) ou insensíveis (BOLTER; JONGSMA, 1997; BROADWAY, 1997) aos IPs, mas independente dessa natureza, elas parecem estar na base da capacidade desses insetos em sobrepor os efeitos dessa barreira protetiva das plantas.

Um dos inibidores de peptidases mais conhecidos é o inibidor do tipo Bowman-Birk de soja (sBBI), o qual possui especificidade por tripsinas e quimotripsinas e massa molecular aproximada de 8kDa (BOWMAN, 1946; BIRK; GERTLER; KHALEF, 1963). O sBBI possui dois loops, um com características mais hidrofílicas (loop 1 – Lys16), o qual possui alta especificidade por tripsinas, e outro mais hidrofóbico, o qual se liga preferencialmente a enzimas similares a quimotripsinas e elastases (loop 2 – Leu43) (Figura 3) (KOEPEKE et al., 2000).

Inibidores do tipo sBBI podem se ligar a duas proteínas simultaneamente, como pode ser observado na estrutura cristalizada do inibidor de serina peptidase do tipo Bowman-Birk de *Vigna unguiculata* (BTCI) (Figura 4) (ESTEVEES et al., 2007). Entretanto, alguns estudos tem mostrado que pode haver uma interferência entre os dois eventos de inibição (GLADYSHEVA et al., 1999). Em outras palavras, tem sido sugerido que a ligação de uma peptidase a um loop do sBBI pode interferir na ligação de uma outra peptidase ao outro loop e *vice versa*. Gladysheva et al. (1999)

encontraram que o potencial de inibição do sBBI a duodenases (um outro tipo de peptidase, a qual é inibida pelo loop 2) sofre um decréscimo quando o inibidor já está em complexo com uma peptidase no loop 1.

Figura 3 – Estrutura covalente e pontes dissulfeto do inibidor do tipo Bowman-Birk de soja
Fonte: Extraído e adaptado de Koepke et al. (2000)

Figura 4 – Estrutura cristalizada (PDB: 3RU4) mostrando o inibidor de serina peptidase do tipo Bowman-Birk (vermelho) de *Vigna unguiculata* em complexo com uma quimotripsina (verde) e uma tripsina (azul)

3 MATERIAL E MÉTODOS

3.1 Relação entre os grupos de insetos estudados

As sequências proteicas das tripsinas de insetos das ordens Coleóptera, Lepidóptera e Díptera foram identificadas usando a busca por PSI-BLAST (ALTSCHUL et al., 1997) no banco de dados não-redundante de proteínas do GenBank do NCBI (National Center for Biotechnology Information). Para isso, foram usadas sequências de tripsinas anotadas, representando cada um dos grupos (gi: 270006343, 17986087, 397910050 para as ordens Coleóptera, Díptera e Lepidóptera, respectivamente) como objeto de pesquisa. Enquanto isso, para as quimotripsinas foi utilizada uma busca por termo no banco de dados do NCBI.

A busca por homólogos de tripsinas e quimotripsinas em espécies de Díptera produziu um grande número de dados provenientes de espécies do gênero *Drosophila*. Essa excessiva representação de homólogos em *Drosophila* – a qual depende em grande parte da quantidade de informação disponível para esse gênero – mascarou os homólogos provenientes de outros gêneros de Dípteros. Portanto, todo o gênero *Drosophila* foi substituído por somente três espécies representando grupos distantes de *Drosophila*, de acordo com a proposição filogenética mais aceita para o grupo (SEETHARAM; STUART, 2012). Essas espécies englobam: uma do grupo melanogaster (*D. melanogaster*), uma do grupo obscura (*D. pseudoobscura pseudoobscura*) e uma do grupo virilis (*D. virilis*). As sequências de tripsinas e quimotripsinas para essas espécies foram obtidas no banco de dados Refseq do NCBI. Além disso, a família Culicidae, pertencente a ordem Díptera não foi considerada na busca por PSI-BLAST, uma vez que essa família contém apenas insetos que se alimentam de sangue (*Anopheles*, *Aedes*, etc.), os quais estão provavelmente sujeitos a diferentes pressões de seleção em termos de hábitos alimentares.

Com o intuito de obter as sequências das peptidases estudadas para as espécies *Spodoptera frugiperda* e *Alabama argillacea*, as quais não estão disponíveis em banco de dados públicos, foi realizado o sequenciamento do transcriptoma dessas espécies. Para isso, o RNA total, proveniente do intestino médio de um conjunto de 50 lagartas de quinto instar foi extraído, utilizando-se protocolo de extração por Trizol®. A biblioteca de cDNA construída a partir dessas amostras foram então sequenciadas, produzindo leituras de 200pb de sequências

únicas (“single reads” – SE). A construção e sequenciamento das bibliotecas de cDNA foi realizada por empresa privada, utilizando a plataforma Illumina Genome Analyzer II. Diferentes programas para a montagem *de novo* de leituras SE foram utilizados para gerar um conjunto de dados não redundante. Além disso, dois diferentes softwares para a montagem de transcriptomas foram utilizados: 1) Velvet Assembler (ZERBINO; BIRNEY, 2008), utilizando os parâmetros padronizados sugeridos pelo programa, com o corte de cobertura ótimo (“optimum coverage cutoffs”) e os valores de *k-mer* (sequência de k-nucleotídeos) propostos pelo pacote VelvetOptimizer; 2) CLC Bio Genomics Workbench, utilizando os parâmetros padronizados, exceto para o tamanho mínimo de *contig*. Ambos os programas foram executados duas vezes. A primeira vez com o tamanho mínimo de *contig* de 150pb e a segunda com 200pb. Depois da montagem, o software Dupwipe (parte do pacote BIOSPHA) (BRANDÃO; SILVA-FILHO, 2011) foi utilizado para remover as sequências duplicadas. Finalmente, os conjuntos de sequências gerados foram imergidos utilizando-se o software Minimus2 (SOMMER et al., 2007). As sequências foram então anotadas usando Blast2GO (CONESA; GÖTZ, 2008; GOTZ et al., 2008) e armazenadas em um banco de dados local MySQL. As análises identificaram 11 sequências de tripsinas e 20 quimotripsinas (Figura 5 e 6).

Figura 5 – Alinhamento das tripsinas de *Spodoptera frugiperda* e *Alabama argillacea* a tripsina bovina (PDB: 4I8H)

Figura 6 - Alinhamento das quimotripsinas de *Spodoptera frugiperda* e *Alabama argillacea* a quimotripsina de *Solenopsis invicta* (Insecta – PDB: 1EQ9)

Com o intuito de filtrar apenas as sequências confiáveis de tripsinas, três filtros foram aplicados: 1) primeiramente, as sequências sem a tríade catalítica completa (His57, Asp102 e Ser195), Asp189 (o qual é pertencente ao sítio de ligação ao substrato e responsável pela especificidade dessas proteínas) e a isoleucina inicial foram removidas; 2) posteriormente, essas sequências foram filtradas para a remoção de peptidases não envolvidas na digestão (como prophenoloxidasas, fatores de coagulação, neuropsinas e peptidases do fluido seminal), usando inferência de Neighbor-Joining (NJ). Para isso, essas sequências foram alinhadas usando o software MUSCLE (EDGAR, 2004) e as árvores foram preditas utilizando o software ClustalW2-Phylogeny (THOMPSON; HIGGINS;

GIBSON, 1994). O cluster contendo somente peptidases classificadas como “trypsin”, “trypsin-like” ou “serine peptidase” foi selecionado para o trabalho (Figura 7); 3) finalmente, as sequências redundantes, com mais de 97% de identidade na sequência de nucleotídeos, foram descartadas.

Figura 7 – Árvores provenientes de análises de Neighbor-Joining usadas na classificação das tripsinas das ordens: Coleóptera, Díptera e Lepidóptera. O grupo de proteínas agrupados nos ramos sombreados em cinza foram selecionados para o trabalho. A escala de cor representa as sequências anotadas como proteases não-digestivas, denominadas de acordo com a legenda

As sequências de quimotripsinas foram filtradas usando os mesmos critérios utilizados para tripsinas (Figura 8), exceto pela restrição a sequências sem Asp189, posição na qual as quimotripsinas são mais variáveis.

Figura 8 – Árvores provenientes de análises de Neighbor-Joining usadas na classificação das quimotripsinas das ordens: Coleóptera, Díptera e Lepidóptera. O grupo de proteínas agrupados nos ramos sombreados em cinza foram selecionados para o trabalho. A escala de cor representa as sequências anotadas como peptidases não-digestivas, denominadas de acordo com a legenda

Além disso, é importante destacar que a presente classificação, identificando as sequências de tripsinas e quimotripsinas, foi validada pela pesquisa por BLAST por proteínas anotadas como “trypsin” ou “chymotrypsin” no banco de dados de peptidases MEROPS (RAWLINGS; BARRETT, 1993). Nessa análise, as sequências proteicas com maior correspondência as tripsinas ou quimotripsinas depositadas no MEROPS foram observadas como presentes apenas nos conjuntos selecionados para o trabalho.

As sequências de tripsinas foram nomeadas com base na ordem taxonômica, espécie e um número crescente. Por exemplo, em CTcas6, “C” corresponde a Coleóptera e “Tcas” a *Tribolium castaneum*. Enquanto isso, as quimotripsinas foram nomeadas de maneira similar, mas um “c” foi incluído entre a letra correspondente à ordem e a sigla da espécie (ex.: CcTcas6).

Todas as sequências e espécies usadas nesse trabalho estão descritas nos Anexos A e B.

3.2 Análises filogenéticas

As sequências de tripsinas e quimotripsinas ativas foram alinhadas usando o software T-coffee (NOTREDAME; HIGGINS; HERINGA, 2000) e submetidas a proposição filogenética bayesiana, usando o software MrBayes (versão 3.2.1 *mpi-enabled*) (RONQUIST; HUELSENBECK, 2003). Os modelos de substituição usados nas análises foram escolhidos usando o software Prottest (versão 3) (ABASCAL; ZARDOYA; POSADA, 2005), seguindo o parâmetro *Akaike Information Criterion* (AIC). 2×10^6 gerações de Markov Chains foram conduzidas e amostradas a cada 100 gerações. 25% das árvores iniciais foram descartadas como *burn in* para assegurar a amostragem das árvores após a convergência da topologia. A árvore resultante foi organizada e analisada no software FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

3.3 Definição das regiões proteicas estudadas

O sítio de ligação ao substrato compreende as seguintes regiões (de acordo com a numeração da quimotripsina bovina, Uniprot: P00766): 40-42, 57-58, 99, 143, 151, 189-195, 213-217, 220, 223-224 e 228 (Figura 9) (BODE; MEYER; POWERS, 1989; HEDSTROM, 2002).

O motivo de ativação do zimogênio inclui aproximadamente seis resíduos imediatamente anteriores ao primeiro resíduo N-terminal das tripsinas/quimotripsinas (Ile16). O motivo é numerado P1-P6 na direção da região C- para a N-terminal. A posição P1 é sempre ocupada por um resíduo positivamente carregado (lisina ou arginina) e a clivagem ocorre entre esse resíduo e a primeira isoleucina (Ile16) das tripsinas/quimotripsinas (Figura 9). É importante ressaltar que o motivo de ativação é o alvo das proteases no processo de ativação. Essa é a razão pela qual o número

usado (P1, P2,...) é o mesmo adotado para os substratos das proteases (SCHECHTER; BERGER, 1968).

Figura 9 – As regiões das peptidases utilizadas nesse trabalho. a) Posições dos substratos (P5, ..., P2') e os sub-sítios correspondentes das tripsinas/quimotripsinas (S5, ..., S2') no sítio de ligação ao substrato. Somente o número do resíduo (ex.: 217 na posição S5), e não o tipo do resíduo, esta especificado em posições variáveis. As posições conservadas são reportadas usando o tipo e o número do resíduo (ex.: Gly216 na posição S4). Os resíduos marcados por um asterisco (Ser214 na posição S2 e Gly216 na posição S4) não são variáveis em tripsinas, mas eles são em quimotripsinas; b) as sequências do peptídeo sinal e do motivo de ativação no tripsinogênio humano (usado como referência). As flechas vermelhas indicam o sítio de clivagem

O peptídeo sinal é a subsequência do tripsinogênio/quimotripsinogênio responsável pela secreção da enzima e está claramente definido nos tripsinogênios humanos (CHEN et al., 2003), onde ele é parte da proteína chamada de pré-tripsinogênio e estende-se da primeira metionina na extremidade N-terminal até a região do motivo de ativação (Figura 9). Em insetos essa região não está bem definida e foi considerada como os demais resíduos imediatamente anteriores ao motivo de ativação (da posição -7 até a metionina).

3.4 Análises de pressão de seleção

Com o intuito de identificar os resíduos do sítio de ligação ao substrato de tripsinas e quimotripsinas evoluindo sob forças adaptativas em Lepidópteros, uma

detecção sítio por sítio foi conduzida por análises de Máxima Verossimilhança (*Maximum Likelihood* – ML). Em particular, o software CodeML do pacote PALM (YANG, 2007) foi usado para produzir os testes evolutivos.

As análises foram conduzidas apenas para as enzimas de Lepidóptera porque o número de sequências disponíveis para Coleóptera e Díptera não foram considerados suficientes para obter resultados robustos para ambas as proteínas analisadas.

As sequências de cDNA das tripsinas e quimotripsinas de Lepidóptera foram obtidas do banco de dados de nucleotídeos do NCBI, exceto por *S. frugiperda* e *A. argillacea*, as quais foram obtidas pela análise prévia do transcriptoma desses insetos, de acordo com a metodologia descrita no item 3.1. O alinhamento dos códons foi realizado utilizando o software TranslatorX (ABASCAL; ZARDOYA; TELFORD, 2010), usando como guia o alinhamento múltiplo das proteínas correspondente, o qual foi previamente obtido como descrito no item 3.2.

A árvore por inferência Bayesiana produzida para as sequências de cDNA foi obtida pela análise no software MrBayes (versão 3.2.1 mpi-enable) (RONQUIST; HUELSENBECK, 2003), usando o modelo evolutivo de substituição GTR+I+G de acordo com o proposto nas análises do software JModelTest (versão 2.1.3) (DARRIBA et al., 2012), seguindo o parâmetro AIC. 2×10^6 gerações de *Markov Chains* foram conduzidas e amostradas a cada 100 gerações. 25% das árvores iniciais foram descartadas como *burn in* para assegurar a amostragem das árvores após a convergência da topologia.

Para os testes de pressão de seleção por códon, a razão ω foi mensurada como medida da seleção natural agindo sobre a proteína, onde $\omega = d_N/d_S$, i.e., taxa de substituição não-sinônima/sinônima (GOLDMAN; YANG, 1994; YANG; NIELSEN, 1998). De acordo com esse modelo, as categorias de seleção foram definidas como: a) seleção positiva ($\omega > 1$): a variabilidade de aminoácidos na posição é mais frequente do que o esperado por deriva genética; b) seleção negativa ($\omega < 1$): forças evolutivas tem atuado contra mudanças no aminoácido que ocupa essa posição; c) seleção neutra ($\omega = 1$): a variabilidade de aminoácidos que ocupam essa posição é tão frequente quanto o esperado por deriva genética (YANG; NIELSEN, 2002).

Com o intuito de produzir uma detecção de pressão de seleção sítio por sítio, três pares de modelos constituindo três testes de Máxima Verossimilhança foram produzidos, usando os modelos de substituição descritos na Tabela 1. Primeiro, o

modelo de uma única razão de ω (M0) foi testado e comparado com o modelo de variáveis discretas (M3 – com K=3 classes discretas). Esse primeiro par de modelos (M0-M3) testa a presença de pressão de seleção variável entre os sítios. Então, dois pares de modelos foram usados para testar a probabilidade de seleção positiva nos sítios de códons (M1a vs M2a e M7 vs M8). O modelo aproximadamente neutro M1a, o qual inclui duas categorias de sítios (seleção neutra, i.e. $\omega = 1$, e seleção negativa, i.e. $0 < \omega < 1$), foi comparado com o modelo de seleção M2a, o qual contém uma classe de sítio adicional, i.e. seleção positiva ($\omega > 1$). Posteriormente, o modelo beta M7, o qual usa uma distribuição beta de ω (em um intervalo de 0-1), foi comparado ao modelo M8, o qual utiliza uma distribuição beta de ω e inclui uma classe adicional de ω , i.e. a de sítios selecionados positivamente ($\omega > 1$). Os modelos foram comparados usando testes de razão de verossimilhança (LRT) para determinar o ajuste dos modelos mais próximos, com 4 graus de liberdade para a comparação M0/M3, e 2 para a comparação M1a/M2a e M7/M8. Quando os testes foram significativos, o modelo mais simples (M0, M1a ou M7) foi rejeitado em favor do mais complexo (M3, M2a ou M8). Além disso, quando o teste foi significativo, o teste de probabilidade posterior *Bayes Empirical Bayes* (BEB) foi aplicado para calcular a probabilidade posterior de cada classe de sítio ($\omega < 1$, $\omega = 0$ e $\omega > 1$).

Tabela 1 – Parâmetros utilizados nos modelos de substituição por sítio

Modelo	Nº de sítios variáveis	Nº de parâmetros livres	Parâmetros
M0 (razão única)	0	1	ω
M1a (neutro)	1	2	p_0 ($p_1 = 1 - p_0$), $\omega_0 < 1$, $\omega_1 = 1$
M2a (seleção)	2	4	p_0 , p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$
M3 (discreto)	3	5	p_0 , p_1 ($p_2 = 1 - p_0 - p_1$) ω_0 , ω_1 , ω_2
M7 (beta)	7	2	p , q
M8 (beta e ω)	8	4	p_0 ($p_1 = 1 - p_0$), p , q , $\omega_s > 1$

*Os parâmetros nos parênteses não são livres

3.5 Modelagem das proteases de insetos

A predição estrutural das proteínas foi produzida usando modelagem por homologia. O software HHpred (SODING; BIEGERT; LUPAS, 2005) foi usado para a

busca de sequências moldes e geração do alinhamento alvo-molde. Os modelos foram produzidos pelo software Modeller (ESWAR et al., 2007).

Uma vez que algumas inserções estão presentes em alguns loops do sítio de ligação das tripsinas analisadas e nenhum molde cobrindo todas as inserções foi encontrado, uma metodologia de múltiplos moldes foi escolhida. Dois moldes foram usados: as estruturas cristalizadas das tripsinas de *Fusarium oxysporum* (PDB: 1PQ7) e *Bos taurus* (PDB: 3MFJ), ambas com 0,8 Å de resolução e com média de 37 e 35% de similaridade com as tripsinas de *S. frugiperda*. Nas sequências com inserções no loop entre os resíduos 174 e 175 (correspondentes a tripsina bovina), a região foi refinada usando a ferramenta de refinamento de loops do software Modeller, usando o inibidor Bowman-Birk de soja (sBBI) como restrição estrutural.

Para as quimotripsinas, dois moldes foram usados: as estruturas cristalizadas da quimotripsina de *Solenopsis invicta* (PDB: 1EQ9) e a colagenase de *Hypoderma lineatum* (PDB: 2HLC), ambas com 1.7 Å de resolução e com média de 36 e 39% de identidade, respectivamente, com as quimotripsinas de *S. frugiperda*.

A identificação dos resíduos do sítio ativo das tripsinas e da região de interação do sBBI, usados no direcionamento da docagem, foram preditas pela superposição das estruturas preditas à estrutura cristalizada da tripsina bovina em complexo com o sBBI (PDB 1D6R com 2,30Å de resolução) (KOEPEKE et al., 2000) usando o software PyMOL.

Para as quimotripsinas, devido a falta de estrutura cristalizada de uma quimotripsina em complexo ao sBBI, a superposição seguiu duas etapas. Primeiro, com o intuito de encontrar a orientação do sBBI com relação a quimotripsina, a estrutura do sBBI (extraído de 1D6R) foi superposta a estrutura do inibidor do tipo Bowman-Birk de *Vigna unguiculata* (BTCL), o qual possui estrutura cristalizada em complexo com uma quimotripsina bovina (PDB: 3RU4 – ESTEVES et al., 2007). Em seguida, os modelos de quimotripsinas de *S. frugiperda* foram superpostos a quimotripsina bovina ligada ao sBBI (complexo produzido pela superposição anterior, entre a tripsina bovina da estrutura PDB: 3RU4 e o sBBI da estrutura PDB:1D6R).

Com o intuito de preparar as estruturas para os estudos de docagem, 100 etapas de minimização de energia foram conduzidas usando o software Chimera/Amber (PETTERSEN et al., 2004).

Os resíduos potencialmente envolvidos na interação entre as tripsinas e quimotripsinas e o sBBI foram preditos usando o webserver PIC (TINA; BHADRA; SRINIVASAN, 2007). Os experimentos de docagem entre o sBBI e as peptidases preditas foram conduzidos com o Web Server HADDOCK (DE VRIES; VAN DIJK; BONVIN, 2010). As análises foram conduzidas usando os resíduos interagentes (“resíduos ativos”) como preditos pelo Web Server PIC. Os resíduos circundantes aos resíduos ativos (“resíduos passivos”) foram automaticamente identificados usando o HADDOCK.

A identificação de resíduos *hotspots* para a interação peptidase-inibidor foram preditos pelo *scanning* de alanina, usando o Software Robetta (KORTEMME; KIM; BAKER, 2004; KIM; CHIVIAN; BAKER, 2004), considerando como *hotspots* somente os resíduos para os quais a troca por alanina causou alteração na energia livre do complexo ($\Delta\Delta G_{\text{bind}}$) de mais de 1kcal/mol.

4 RESULTADOS

Com o intuito de estudar a história evolutiva das tripsinas e quimotripsinas de insetos, árvores filogenéticas foram preditas para as enzimas encontradas nas ordens Lepidóptera, Díptera e Coleóptera. Além disso, as sequências enzimáticas foram caracterizadas em termos de composição em três regiões funcionais: sítio de ligação ao substrato, motivo de ativação do zimogênio e peptídeo sinal (ver Figura 9). Para cada grupo de similaridade detectado nas análises e cada uma das regiões estudadas, uma sequência consenso foi identificada. Finalmente, uma análise de pressão de seleção foi realizada para as peptidases de Lepidóptera com o intuito de identificar resíduos sob forças evolutivas adaptativas.

4.1 Inferência filogenética para tripsinas e quimotripsinas

As sequências de tripsinas de cada ordem de inseto foram agrupadas de acordo com a inferência bayesiana. Esse procedimento identificou cinco grupos de similaridade, C-I, ..., C-V, para Coleóptera, doze, D-I, ..., D-XII, para Díptera e dez, L-I, ..., L-X, para Lepidóptera (Figura 10). Para cada grupo e para cada região funcional estudada, uma sequência consenso foi obtida.

As tripsinas pertencentes a uma mesma espécie foram, em alguns casos, encontradas em diferentes grupos de similaridade, tais como em *Tribolium castaneum* em Coleóptera, e em outros casos no mesmo grupo de similaridade junto com tripsinas de espécies diferentes, tais como no grupo C-III (Figura 10). A presença de tripsinas de diferentes espécies no mesmo grupo de similaridade, tais como CTcas12 de *T. castaneum*, CRdom3 de *R. dominica* e CDpon5 de *D. ponderosae* no grupo de similaridade C-III, indica que eventos de duplicação de gênica mais ancestrais podem ter ocorrido, talvez antes da separação dessas três espécies. Por outro lado, eventos mais recentes de duplicação também podem ser inferidos observando-se as filogenias dos grupos: por exemplo, a maior identidade de sequência encontrada entre CTcas4 e CTcas6 (C-III) do que entre essas enzimas e CTcas12 da mesma espécie e CRdom3 e CDpon5 de diferentes espécies, suporta a hipótese de que CTcas4 e CTcas6 podem ser resultado de um evento de duplicação gênica recente.

Figura 10 – Inferência filogenética Bayesiana das sequências proteicas e os grupos de tripsinas para: (a) Coleóptera – C, (b) Díptera – D e (c) Lepidóptera – L. Os números nos nós representam as probabilidades posteriores

Interessantemente, os grupos L-VI a L-X de Lepidóptera foram observados como monofilias, contendo somente tripsinas pertencentes a insetos da família Noctuidae, indicando que esses grupos de similaridade podem não possuir representantes em outras famílias de Lepidóptera (Figura 10). Esses resultados sustentam a hipótese de que as tripsinas desses grupos (L-VI a L-X) provavelmente possuem um ancestral comum em algum ponto do desenvolvimento evolutivo da família Noctuidae. Além disso, conforme discutido a seguir, os resíduos envolvidos na atividade enzimática dessas tripsinas, apresentam características peculiares, sugerindo uma especificidade diferenciada nos diferentes sub-sítios enzimáticas, das que foram observadas para os demais grupos.

A análise filogenética das quimotripsinas produziu resultados similares (Figura 11), nos quais também foi possível identificar grupos de similaridade contendo

apenas genes de espécies da família Noctuidae (Lc-X a Lc-XV). Entretanto, conforme descrito abaixo, a análise do sítio de ligação dessas enzimas não identificou nenhuma característica peculiar para esses grupos de enzimas.

Figura 11 – Inferência filogenética Bayesiana das sequências proteicas e os grupos de quimotripsinas de: (a) Coleóptera – C, (b) Díptera – D e (c) Lepidóptera – L. Os números nos nós representam as probabilidades posteriores

Adicionalmente, as análises filogenéticas das tripsinas e quimotripsinas de insetos, incluindo todas as três ordens de insetos estudadas, apontaram uma maior similaridade entre as proteínas de Díptera e Coleóptera em relação a Lepidóptera (Figura 12 e 13, respectivamente). As filogenias desses dois grupos proteicos foram, portanto, contrárias ao esperado, uma vez que a filogenia mais aceita para holometábolos propõe Díptera e Lepidóptera como ordens mais próximas, com

ambas igualmente distantes de Coleóptera (TRAUTWEIN et al., 2012). Esses resultados sugerem que, por alguma razão adaptativa, essa classe de proteínas tem seguido um caminho evolutivo peculiar em Lepidóptera.

Figura 12 – Proposição filogenética Bayesiana para as sequências de tripsinas de Coleóptera, Díptera e Lepidóptera. Táxons em cinza claro, cinza e preto representam sequências de espécies de Coleóptera, Díptera e Lepidóptera, respectivamente. Os números juntos aos nós representam a probabilidade posterior dos ramos

Figura 13 – Proposição filogenética Bayesiana para as sequências de quimotripsinas de Coleóptera, Díptera e Lepidóptera. Táxons em cinza claro, cinza e preto representam sequências de espécies de Coleóptera, Díptera e Lepidóptera, respectivamente. Os números juntos aos nós representam a probabilidade posterior dos ramos

4.2 Caracterização do sítio de ligação ao substrato das tripsinas

Com o intuito de elucidar as características do sítio de ligação das tripsinas, a sequência consenso nessa região – entre os sub-sítios S5 e S2' – de cada grupo de similaridade foi comparada ao consenso dos outros grupos (Figura 14).

Figura 14 – Consenso do sítio de ligação ao substrato dos grupos de tripsinas de Coleóptera (C), Díptera (D) e Lepidóptera (L). Os números romanos na primeira coluna de cada tabela representam os números dos grupos de tripsinas. O tipo de resíduo e a posição na estrutura foram obtidos da tripsina bovina (PDB: 4I8H). As posições conservadas em todos os grupos foram omitidas, exceto pela tríade catalítica (His57, Asp102 e Ser195) destacada em azul na estrutura. Xs indicam posições onde um consenso não foi observado em mais do que 50% das sequências do grupo. Posições marcadas com um “-” e “+” foram preditas como sob pressão de seleção negativa ($\omega < 1$) e positiva ($\omega > 1$), respectivamente, em Lepidóptera

Uma vez que o objetivo era estudar as diferenças entre os grupos e ordens, as posições das sequências onde o tipo de resíduo foi observado como conservado

em todos os grupos e ordens não serão apresentadas nos resultados. Essas posições estão reportadas na Figura 9 como o tipo de resíduo e o número da posição (ex.: His57), enquanto as posições consideradas variáveis estão simplesmente indicadas como o número da posição (ex.: 217).

Como apresentado na Figura 14, a posição 41 (pertencente ao sub-sítio S1'), as posições 143 e 151 (S2'), a posição 99 (S2) e as posições 217 e 223 (S5) foram observadas como pouco conservadas em ambas as comparações, intra- e inter-grupos de tripsinas. Em contraste, a posição 40 (S2'), as quatro posições 190, 192, 213 e 228 pertencentes ao sub-sítio S1, e a posição 215 (S2) são altamente conservadas em todos os grupos, mas mostram diferenças dentro dos grupos e ordens. Essas posições são discutidas em detalhes a seguir.

Posição 40 (S2') – Nessa posição, um resíduo de histidina é observado como mais frequente nas tripsinas conhecidas de mamíferos. Esse também é o caso das tripsinas de Coleóptera e Díptera. Entretanto, nessa posição, todos os grupos de similaridade de tripsinas de Lepidópteros analisados, exceto um (L-III), mostram uma glutamina, enquanto as tripsinas do grupo L-III apresentam uma leucina. A glutamina também foi encontrada sob pressão de seleção negativa ($\omega < 1$) na análise de pressão de seleção sítio-por-sítio (os parâmetros resultantes da análise de pressão de seleção das tripsinas estão disponíveis no Anexo C). Isso sugere que uma glutamina nessa posição é um resíduo chave para as tripsinas de Lepidóptera e que qualquer mutação ocorrida durante a evolução tem sido contra-selecionada.

Posições 190, 192, 213, 228 (S1) – S1 é o sub-sítio das tripsinas conhecido por determinar a especificidade da enzima por Arg ou Lys na posição P1 do substrato, sendo a posição 190 a posição mais relevante para essa especificidade (SICHLER et al., 2002). Em todas as sequências de tripsinas analisadas, a posição 192 foi a mais conservada, dentre essas posições, em todas as ordens, a qual é principalmente ocupada por uma glutamina. Além disso, nas posições 190, 213 e 228 Lepidóptera mostrou uma forte preferência por resíduos mais hidrofílicos (A/Q em 190, C/T em 213, N/S/Y em 228) do que Coleóptera e Díptera, as quais possuem uma alanina em 190, usualmente uma valina em 213, e uma tirosina em 228. Essa menor hidrofobicidade é mais pronunciada nos grupos L-VI a L-X (Figura 10). As análises de pressão de seleção mostraram que todos os resíduos envolvidos no sub-sítio S1 das tripsinas de Lepidóptera estão sob pressão de seleção negativa ($\omega < 1$), i.e. elas são menos variáveis do que o esperado por deriva genética.

Posições 215 (S2 e S4) – Seis dos dez grupos de similaridade de Lepidóptera mostram um triptofano na posição 215, consistentemente com todos os grupos de Coleóptera, Díptera (Figura 14) e mamíferos. Entretanto, nesta posição, as tripsinas dos grupos de similaridade de Lepidóptera L-VII, L-VIII, L-IX e L-X, os quais contem apenas enzimas de insetos da família Noctuidae, mostram ao invés disso, uma fenilalanina, a qual é mais hidrofóbica que o triptofano, de acordo com a escala de hidrofobicidade de Eisenberg et al. (1984). Não obstante, a análise de pressão de seleção indicou que a posição 215 está sujeita a seleção negativa ($\omega < 1$).

As diferenças entre o sítio de ligação das tripsinas de Lepidóptera nos grupos L-I e L-X, os quais representam grupos mais distantes na árvore filogenética para a ordem, podem ser também apreciadas pela comparação da superfície proteica de duas enzimas representativas, uma de cada grupo (Figura 15).

Figura 15 – Padrão de hidrofobicidade no sítio de ligação ao substrato das tripsinas dos grupos L-I (a) e L-X (b) de Lepidóptera. A escala de hidrofobicidade varia de verde escuro (baixa hidrofobicidade) a branco (alta hidrofobicidade). O SBSI é representado por linhas pretas e mimetiza a posição do substrato no sítio de ligação. A cadeia lateral da lisina na posição P1 do SBSI também é apresentada na figura, destacando a profundidade do sítio S1

As estruturas das tripsinas foram previstas usando técnicas de modelagem por homologia e o inibidor de soja do tipo Bowman-Birk (sBBI) – o qual mimetiza a posição do substrato dessas enzimas (KOEPEKE et al., 2000) – foi acoplado ao sítio de ligação do substrato (ver Material e Métodos). Uma diferença na hidrofobicidade entre as proteínas analisadas pode ser observada na posição 190, na posição 215 e na posição 217. As duas últimas posições são claramente mais hidrofóbicas no grupo L-X do que no grupo L-I.

Finalmente, a análise de pressão de seleção mostrou duas posições sujeitas a seleção positiva ($\omega > 1$) (Figura 16). Uma é a posição 151, a qual pertence ao subsítio S2', e a outra é a 146, a qual não está localizada no sítio de ligação ao substrato. Entretanto, ambas pertencem a um mesmo loop, o qual é conhecido por estar envolvido na especificidade das peptidases (SCHELLENBERGER; TURCK; RUTTER, 1994).

Figura 16 – Sítios do sítio de ligação ao substrato encontrados sob pressão de seleção nas tripsinas de Lepidóptera. a) probabilidades posteriores de *Bayes Empirical Bayes* (BEB) para as 3 classes de pressão de seleção testadas pelo modelo M2a nas tripsinas de Lepidópteros e b) a localização correspondente desses resíduos na estrutura cristalizada da tripsina bovina (PDB: 4I8H). Resíduos em vermelho e azul indicam as posições encontradas sob pressão de seleção purificadora ($\omega < 1$) e diversificadora ($\omega > 1$), respectivamente. As posições conservadas para todas as tripsinas foram omitidas da figura

4.3 Caracterização do sítio de ligação ao substrato das quimotripsinas

As quimotripsinas de insetos analisadas mostraram maior variabilidade na composição do sítio catalítico do que as tripsinas. Em particular, posições tais como 189, 214, 216 e 226, as quais foram encontradas como altamente conservadas nas tripsinas, não demonstraram o mesmo resultado nas quimotripsinas (Figura 17). As posições variáveis no sítio de ligação ao substrato das quimotripsinas são discutidas abaixo.

Posição 40 (S2') – Similarmente as tripsinas, a posição 40 também é ocupada por uma histidina nas quimotripsinas de mamíferos. Uma histidina também foi observada nessa posição em alguns grupos de Coleóptera e Díptera, mas não nas quimotripsinas de Lepidóptera, as quais demonstram quase que exclusivamente uma serina ou uma glicina nessa posição. Esta posição foi também predita como sujeita a pressão de seleção negativa ($\omega < 1$) (os escores de Máxima Verossimilhança das análises de pressão de seleção para as análises com as sequências de quimotripsinas estão apresentados no Anexo D).

Posições 189, 190, 192, 213 e 228 (S1) – Diferentemente das tripsinas, as posições do sub-sítio S1 envolvidas na especificidade ao substrato das quimotripsinas foram observadas como extremamente conservadas. Em Lepidóptera, nem todas essas posições foram observadas sob pressão de seleção negativa. Por exemplo, a posição 192, a qual é principalmente ocupada por uma glutamina conservada nas tripsinas, é altamente variável nas quimotripsinas e sujeita a pressão de seleção neutra ($\omega = 1$) em Lepidóptera e conservada em Díptera, onde uma histidina é predominante. Além disso, a posição 189, a qual é invariável nas tripsinas, ocupando o fundo do sub-sítio S1, nas quimotripsinas de insetos é ocupada por uma glicina, uma serina ou também a não tão comum asparagina. As maiores diferenças entre as ordens de insetos são observadas nas posições 190 e 213, onde as quimotripsinas de Lepidóptera mostram uma tendência em manter resíduos de treonina, diferentemente das proteínas de Coleóptera e Díptera, as quais mostram resíduos mais comuns e menos hidrofílicos A/S em 190 e V em 213.

Posições 214 e 215 (S2) – Na posição 214, a qual é ocupada por uma serina em todas as tripsinas e nas quimotripsinas de Coleóptera e Lepidóptera, é ocupada por um resíduo de asparagina nas quimotripsinas de Díptera. Na posição 215, as

quimotripsinas de Díptera apresentam exclusivamente resíduos de triptofano, enquanto Coleóptera e Lepidóptera possuem predominantemente uma fenilalanina.

Figura 17 – Consenso do sítio de ligação ao substrato dos grupos de quimotripsinas de Coleóptera (C), Díptera (D) e Lepidóptera (L). Os números romanos na primeira coluna de cada tabela representam os números dos grupos de tripsinas. O tipo de resíduo e a posição na estrutura foram obtidos da tripsina bovina (PDB: 4I8H). As posições conservadas em todos os grupos foram omitidas, exceto pela tríade catalítica (His57, Asp102 e Ser195) destacada em azul na estrutura. Xs indicam posições onde um consenso não foi observado em mais do que 50% das sequências do grupo. Posições marcadas com um “-” e “+” foram preditas como sob pressão de seleção negativa ($\omega < 1$) e positiva ($\omega > 1$), respectivamente, em Lepidóptera

Posições 216 e 226 – As posições 216 e 226 são conhecidas por determinar a especificidade diversa entre elastases e quimotripsinas. Nas tripsinas e quimotripsinas, ambas as posições são ocupadas principalmente por uma glicina, diferentemente das elastases, as quais apresentam duas valinas ou uma valina e uma treonina nessas posições.

Nas sequências de insetos analisadas neste trabalho, outros tipos de resíduos também podem ser observados ocupando as posições 216 e/ou 226. Por exemplo, na única quimotripsina de insetos a qual possui uma estrutura cristalizada determinada (PDB: 1EQ9), uma glicina é observada na posição 216, mas um aspartato foi observado na posição 226, substituindo a glicina usual (BOTOS et al., 2000). Além disso, ambas as posições 216 e 226 foram preditas como estando sob pressão de seleção negativa ($\omega < 1$) em Lepidóptera.

4.4 Caracterização do motivo de ativação de tripsinas e quimotripsinas

Com o objetivo de entender como as tripsinas e quimotripsinas são ativadas em cada grupo de similaridade dentro das ordens estudadas, a região do sítio de ativação do zimogênio, composta pelas seis posições imediatamente anteriores ao sítio de clivagem, foi analisada (Figura 18).

Chen et al. (2003) mostraram que duas mutações nos tripsinogênios humanos, D→G em P2 e K→R em P1, estão associadas com doenças relacionadas a auto-ativação de tripsinas. Isso sugere que a presença de uma arginina em P1 pode ser relacionada ao aumento da auto-ativação. Todos os tripsinogênios de insetos analisados nesse trabalho tem uma arginina conservada em P1, onde usualmente os tripsinogênios bovinos apresentam uma lisina. Isso sugere que a taxa de auto-ativação das tripsinas pode ser aumentada em insetos.

A região do sítio de ativação mostrou um consenso similar no tripsinogênio de Coleóptera e Díptera: com uma glicina em P2 e um resíduo carregado negativamente em P3, enquanto os tripsinogênios de Lepidóptera tem um resíduo polar não-carregado em P2 e uma prolina ou outro resíduo mais hidrofóbico em P3. Esse padrão de conservação pode estar correlacionado com um diferente processo de ativação em Lepidóptera comparado com as outras duas ordens. Em particular, como discutido anteriormente, o sub-sítio S3, contatando a posição P3, é mais hidrofóbico nas tripsinas de Lepidóptera do que nas outras duas ordens. Isso indica que a interação entre o sítio de ligação das tripsinas e o motivo de ativação do

tripsinogênio pode ser favorecida em Lepidóptera. Este fato pode claramente implicar em uma auto-ativação facilitada nessa ordem de inseto com respeito as outras ordens. Além disso, uma alta conservação de serina ou asparagina foi observada na posição P4 em Lepidóptera. De um modo geral, exceto pela posição P1, a região de ativação consensual em Lepidóptera fortemente difere daquela observada em Coleóptera e Díptera (Figura 18).

a		Tripsina	
Bovina		↓ I →	
C-I	P Q I G G R		
C-II	P L P D G R		
C-III	P X I D V R		
C-IV	P R L D G R		
C-V	L X P D G R		
D-I	N X X X G R		
D-II	P Q L D G R		
D-III	D F P F G R		
D-IV	X X X X G R		
D-V	P Q L D G R		
D-VI	A S I S T H		
D-VII	P X P D G R		
D-VIII	X X X X G R		
D-IX	X X X D G R		
D-X	R X P D X R		
D-XI	P R L D G R		
D-XII	P R L D G R		
L-I	C X X A X R		
L-II	X X X P X R		
L-III	X X X X X R		
L-IV	P R S P N R		
L-V	P K S I N R		
L-VI	X P N I Q R		
L-VII	P S N P Q R		
L-VIII	P S N P Q R		
L-IX	P T N P Q R		
L-X	P X N P Q R		
b		Quimotripsina	
Bovina		↓ I →	
Cc-I	T P I N V R		
Cc-II	X X P G L R		
Cc-III	X N X X W R		
Cc-IV	P X X D W R		
Cc-V	R E I X X R		
Cc-VI	X X X G P R		
Cc-VII	K H F X X R		
Cc-VIII	X Q S G L R		
Cc-IX	X X P S P R		
Cc-X	X P S K V R		
Cc-XI	F X X N S R		
D-I	X X W E G R		
D-II	Q Y T X X R		
D-III	K X L D N R		
D-IV	V K P E T R		
D-V	X X X E X R		
D-VI	X X X X G R		
D-VII	A X K D X R		
L-I	D F D G X R		
L-II	D F D G S R		
L-III	D F D G S R		
L-IV	A T R N G R		
L-V	X X X X X R		
L-VI	X A A Q N R		
L-VII	X P X X N R		
L-VIII	X E X X X R		
L-IX	S F N G Q R		
L-X	A R N P S R		
L-XI	D R N P S R		
L-XII	E R N P S R		
L-XIII	E X X X S R		
L-XIV	A S S A S R		
L-XV	X X X P S R		
L-XVI	Q M T X E R		
L-XVII	Y L Q Q Q R		
L-XVIII	K X X X S R		

Figura 18 – Sequência consenso na região de ativação dos grupos de similaridade de (a) tripsinogênio e (b) quimotripsinogênio de Coleóptera, Díptera e Lepidóptera. Xs indicam as posições onde não houve consenso em mais de 50% das sequências do grupo. Os resíduos foram coloridos de acordo com as suas características físico-químicas, onde: vermelho = carregados; alaranjado = histidina; roxo = polar não-carregados; verde = hidrofóbico não-aromático; azul = hidrofóbico aromático

Nas quimotripsinas uma arginina também substitui a lisina observada nos homólogos de mamíferos na posição P1 e os quatro resíduos de aspartato consecutivos não foram observados na região imediatamente anterior a ela (Figura 18). Entretanto, essa região não está bem conservada nas quimotripsinas analisadas. E além disso, as quimotripsinas de Lepidóptera possuem uma alta frequência de resíduos hidrofílicos em P2 (S/N/Q).

4.6 Composição e conservação da região de peptídeo sinal

As sequências dos zimogênios, incluindo o peptídeo sinal e o motivo de ativação de todos os grupos de tripsinas e quimotripsinas estudados, estão descritas no Anexos E e F.

Em muitos dos grupos de similaridade analisados uma sequência consenso para o peptídeo sinal não pode ser observada. Em particular, um consenso geral compartilhado por todos os grupos e ordens não existe, mesmo que algumas regiões conservadas tenham sido encontradas em alguns grupos próximos de tripsinas (C-IV e C-V; D-III e D-IV; e L-V, L-VII, L-VIII, L-IX, e L-X) (Tabela 2). Além disso, como observado no consenso das tripsinas, essa região das peptidases também corrobora com uma maior distância entre as características das sequências proteicas de Lepidóptera em relação a Coleóptera e Díptera.

Tabela 2 – Sequência consenso para o peptídeo sinal dos grupos de tripsinas de Coleóptera, Díptera e Lepidóptera

	Peptídeo sinal
C-IV	MLKFVVISLLVAXAVGAPQKRAXX
C-V	MKSVLLXVFLVATASAVPXFLRKNS
D-III	MLKFVILLSAVACALGGTIPEGLL
D-IV	MLKFVILVCSVAYVFGAVVPGGML
L-V	MRAVILLALLGSALAV
L-VII	MRFLALLALCFAAFAAV
L-VIII	MRILALVALCFAAFAAV
L-IX	MRIFALLALCFAAFAAV
L-X	MRXLAXLALCIAAFAAV

*Os grupos sem conservação de tamanho foram omitidos da tabela para deixar mais claros os resultados. Todas as sequências podem ser encontradas no Anexo E

**X: Indica posição onde não houve consenso em mais de 50% das sequências dos grupos

4.5 Tripsinas e quimotripsinas e o seu putativo modo de interação com sBBI

Com o intuito de compreender se as mudanças observadas no sítio de ligação ao substrato das proteínas estudadas podem ter alterado a sua interação com o inibidor de peptidase do tipo Bowman-Birk de soja (sBBI), uma estratégia de docagem foi aplicada. Nessa análise, o sBBI foi inicialmente docado as tripsinas e quimotripsinas de uma única espécie de Lepidóptera, *Spodoptera frugiperda*. As posições identificadas como pontos de interação com o sBBI estão descritas nos alinhamentos apresentados nas Figuras 19 e 20.

As análises de docagem estrutural não foram conclusivas para mostrar como mudanças na região do sítio de ligação ao substrato de ambas as peptidases podem ter alterado a ligação do sBBI ao sítio. Entretanto, uma posição em particular foi observada como bastante distinta entre tripsinas e quimotripsinas: a posição 192, a qual pertence ao sub-sítio S1. Como previamente discutido, essa posição é extremamente conservada entre as sequências de tripsinas dos insetos estudados e altamente variável nas quimotripsinas. Nos resultados de docagem das quimotripsinas de *S. frugiperda*, quando presentes nessa posição, os resíduos de His, Arg e Gln foram preditos como *hotspots* para a interação com o sBBI, enquanto os resíduos de Ser, Asn, Ala e Gly não o foram. Esses resultados sugerem que a alta variabilidade nessa posição do sítio de ligação ao substrato das quimotripsinas, pode possuir papel chave na maior ou menor afinidade de ligação dessas enzimas com o sBBI.

Outra observação relevante foi a de uma inserção no *loop* localizado na região entre as posições 169 e 174 das tripsinas. As sequências observadas nesse *loop* mostraram uma alta concentração de resíduos hidrofóbicos em todas as tripsinas de *S. frugiperda* analisadas. Os resultados sugerem que essa inserção pode aumentar o número de ligações hidrofóbicas entre essas enzimas e o sBBI. Nas análises de interação com o inibidor, esse loop foi observado como interagindo principalmente com os resíduos Phe50, Val52 e Ile54 no subdomínio 2 do sBBI, o qual tipicamente interage com as quimotripsinas (Figura 21). Notadamente, a posição Phe50 foi predita como *hotspot* na ligação peptidase-inibidor para todas as quimotripsinas estudadas e quase todas as tripsinas (exceto uma). O compartilhamento de uma mesma região do inibidor por ambas as peptidases,

sugere que a ligação de uma enzima a um *loop*, pode influenciar na ligação da outra enzima no outro *loop*. Análises bioquímicas futuras podem comprovar se a manutenção desse *loop* nesses insetos pode estar correlacionada a uma variação na eficiência do sBBI em atuar sobre as duas enzimas simultaneamente.

BovTry-4I8H	IVGGYTTCGANTVP-----YQVSLNSGYHFCGGSLINSQWVVSAAHCY-----KSGIQVR	15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66
FoTry-1PQ7	IVGGTSASAGDFPFIVSI--SRNGGPW---CGGSLLNANTVLTAACHVSGYA-QSGFQIR	
SfTry6	IVGGTVTTIDRYPTIAALLYNSWGSTFTQTCCGGTILNNRSILTAAHCTLGDL-ANWWRIR	
SfTry7	IVGGNPTVIENYPSIVQV-ESRTGIIWQSCAASILNTRFVVSAAHCFAGILNTRNRIR	
SfTry8	IMGGQLTTIDQYPSIATLLYSENLSFFQRCGGIIINNRAILSAAHCFYGRDPINRWIR	
SfTry9	IVGGSVTTIDRYPTIASLLYSWNLSYWQACGGTILNNRAILTAHCTVGDA-ANRWIR	
SfTry10	IVGGSVTTIDQYPTIAALLYTSNWSAYWQACGGTILNNRSILTAAHCTHGDA-VGRWRIR	
SfTry12	IVGGTVTTIDQYPSITALLYTVNFNTYFQACGGTILNNRAVLTAHCTVGDDQ-VNRWRIR	
BovTry-4I8H	LGEDNINVVEGNEQFISASKSIVHPSYNSNT-LNNDIMLIKLSAASLNSRVASISLPTS	67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127
FoTry-1PQ7	AGSLSRT---SGGITSSLSSVRVHPSYSGNN---NDLAILKLSTSIIPSGGNIGYARLAAS	
SfTry6	LGSTWAN---SGGVVHNVVQNIHHPQYNQGVRFNNDIAILRSASTFSFNNNVRAASIAGA	
SfTry7	AGTSNRA---SGGVVVSINRIVNHPTYSNQ-MDGDISVIQLASALVYNDRIQQVPIVPO	
SfTry8	VGSSFAH---SGGVMHAPNNVIMHPDFNYST-LDSDICIVHSPSAFTFNNNVQPAISIAGS	
SfTry9	LGSTWAN---SGGVVHNVTNIVHPSYNSAT-LNNDIAILRSATTFNNDNVQAASIAGA	
SfTry10	VGSTWAN---SGGVVHNVNQNIHVSYSNT-MDNDIAILRSASTFSFNNNVRAASIAGA	
SfTry12	VGSTFAN---SGGIVNDVNRIIHPHSDYST-LDSDLSILHSASAFNNDNVQAASIAGS	
BovTry-4I8H	CASAG--TQCLISGWGNTKSSGTSYDPVLKCLKAPILSDSSCKSAYPG-----QITSNMF	128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181
FoTry-1PQ7	GSDPVAGSSATVAGWGATSEGGSSTPVNLLKVTVPVIVSRATCRAQYGTGTS---AITNQMF	
SfTry6	NYFPGDNQAVWAAGWGDYSGSNAGSEQLRHVQVQVINQNICRNQYANTVY--TINDNML	
SfTry7	GFELPDNLPPVHSGWGATSQGL-ASAILRDVTIYTINHDVCASRYASLLFPRRVTQNM	
SfTry8	NYNVGDYQPVWAAGWGSQFDGGP-GSEQLRHVMVLTVNQDICSRYGLVFI--QITQNM	
SfTry9	NYLPGDNTAAWAAGWGTTASAGGS-SSEQLRHVELRIINQATCRNNYATRGI--TITDNML	
SfTry10	NYNVADNQAVWAAGWGTTASGASSGSEQLRHVQLVIMNQNSCANNYRIVGI--TITANML	
SfTry12	NYVLGDFQPVWAAGWGTTYYGAPTGSEQLRHVMVTVNQDICRANYGLVFL--TITQNM	
BovTry-4I8H	CAGYL-EGGKDSCQGDSSGGPVV-CSGKLQIGIVSWGSGCAQKNKPGVYTKVCNVVSWIKQT	182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241
FoTry-1PQ7	CAGVS-SGGKDSQGDSSGGPIVDSSNTLIGAVSWGNGCARPNYSGVYASVGALRSFIDTY	
SfTry6	CSGWP-SGGRDQCQGDSSGGPLY-HNGIVVGVC SFGIGCADVRYPGVNARVSRYSWIFSN	
SfTry7	CAGILDVGGKDACQGDSSGGPLY-YDGILVGIVSWGRGCADPVYPGVSTKVASYTDWIVQT	
SfTry8	CSGWP-IGGRDQCNGDSSGGPLY-HNGVVVGVC SFGGLGCLDSLHGVNVRVSRFSSWIAAN	
SfTry9	CSGWP-TGGRDQCQGDSSGGPLY-HNGIVVGVC SFGIGCAQAAFPGVNARVSRYTAWISSN	
SfTry10	CSGWP-SGGRDQCQGDSSGGPLY-HNGIVIGVC SFGGLGCLVNFPGVNARVSRYTTWISSN	
SfTry12	CSGWP-TGGRDQCQGDSSGGPLY-HNGVLVGVC SFGGLGCLDSLPGVNVRVSRFTSWISSN	

Figura 19 – Alinhamento das tripsinas de *S. frugiperda* com as tripsinas de *Bos taurus* (BovTry – PDB: 4I8H) e *Fusarium oxysporum* (FoTry – PDB: 1pq7). Letras em azul e vermelho indicam resíduos preditos como em interações do tipo pontes de hidrogênio e hidrofóbicas com o sBBI, respectivamente. Posições sublinhadas em preto, verde, laranja e rosa indicam posições preditas em interações do tipo iônica, aromática-aromática, aromática-sulfur e cátion-PI com o sBBI, respectivamente. Os números estão de acordo com a quimotripsina bovina e os pontos pretos nas regiões de interação mostram regiões com alinhamento estrutural a ambas as estruturas do PDB. As interações descritas para BovTry são correspondentes a análise do complexo cristalizado disponível no PDB: 1D6R

Figura 20 – Alinhamento das quimotripsinas de *S. frugiperda* com as quimotripsinas de *Bos taurus* (BovChy – PDB: 1YPH) e *Solenopsis invicta* (SiChy – PDB: 1EQ9). Letras em azul e vermelho indicam resíduos preditos como em interações do tipo pontes de hidrogênio e hidrofóbicas com o sBBI, respectivamente. Posições sublinhadas em preto, verde, laranja e rosa indicam posições preditas em interações do tipo iônica, aromática-aromática, aromática-sulfur e cátion-PI com o sBBI, respectivamente. Os números estão de acordo com a quimotripsina bovina e os pontos pretos nas regiões de interação mostram regiões com alinhamento estrutural a ambas as estruturas do PDB

Figura 21 – Superposição dos resultados da docagem da tripsina LSfru6 com o *loop* 1 do sBBI e da quimotripsina LcSfru2 ao *loop* 2 do sBBI à estrutura cristalizada 3RU4. A região de interação comum entre a quimotripsina e a tripsina com a Phe50 do sBBI é mostrada em destaque no canto superior direito

Finalmente, visando comparar o modo de interação do sBBI as tripsinas de diferentes ordens de insetos, os resultados da docagem estrutural das tripsinas de Lepidóptera com o sBBI foram contrastados com os das enzimas da ordem Coleóptera. No entanto, os resultados mostraram que as posições de resíduos dessas peptidases, as quais estão envolvidas na interação com o sBBI, são bastante similares, exceto pela região da inserção descrita anteriormente, como demonstrado na Figura 22, pelas regiões preditas em interação hidrofóbica com o inibidor.

Figura 22 – Frequência observada de resíduos em interação hidrofóbica com o SBBI, nas análises de docagem estrutural das tripsinas de Coleoptera e Lepidoptera. a) gráfico mostrando a frequência observada de interações hidrofóbicas em cada posição do sítio de interação enzima-inibidor; b) alinhamento das sequências utilizadas na docagem estrutural, mostrando apenas as posições observadas em interação com o SBBI. O alinhamento está colorido de acordo com a hidrofobicidade dos resíduos, variando de vermelho (mais hidrofóbico) a azul (mais hidrofílico)

5 DISCUSSÃO

5.1 Caracterização evolutiva geral das tripsinas e quimotripsinas

No presente trabalho, análises filogenéticas de tripsinas e quimotripsinas de insetos das ordens Coleóptera, Díptera e Lepidóptera foram conduzidas com o objetivo de investigar a presença de *imprints* evolutivos nessas enzimas digestivas.

Díptera e Lepidóptera são sabidamente mais próximas evolutivamente e igualmente distantes de Coleóptera (TRAUTWEIN et al., 2012). Idealmente era esperado que as filogenias das tripsinas e quimotripsinas refletissem essa relação, a não ser devido à existência de alguma restrição adaptativa excepcional que tivesse ocorrido.

A árvore filogenética obtida com um grande conjunto de sequências heterogêneas de tripsinas e quimotripsinas, contudo, mostrou que, para essa classe específica de enzimas digestivas, uma história diferente está sendo contada. As enzimas de Lepidóptera formaram um clado separado das enzimas de Coleóptera e Díptera, as quais são mais similares entre si. Tais achados indicam que Lepidóptera, por alguma razão desconhecida, evoluiu para um conjunto mais divergente de sequências de peptidases. Este resultado está suportado pela análise de três regiões das sequências dessas peptidases: o sítio de ligação ao substrato, o motivo de ativação do zimogênio e o peptídeo sinal, o qual é necessário para a secreção celular.

Com o propósito de estudar essas regiões funcionais, três filogenias específicas para cada ordem foram obtidas, para elucidar se as diferentes sequências de tripsinas e quimotripsinas de uma mesma espécie se agrupariam juntas ou separadas. Interessantemente, uma situação variada foi observada nessas análises: cada ordem se divide em um número variável de grupos de similaridade, todos contendo proteínas de diferentes espécies, ainda que em alguns casos mais de uma proteína de uma dada espécie tenha sido observada dentro de um mesmo grupo. A presença de tais proteínas proximamente relacionadas é consistente com a hipótese de eventos de duplicação gênica para essa classe de enzimas (MULLER et al., 1993; WANG; MAGOULAS; HICKEY, 1999). Eventos de duplicação também podem ser os responsáveis pelo grande número de diferentes tripsinas e quimotripsinas observadas em cada espécie, família e ordem.

Assim, os presentes resultados indicam uma divergência entre a filogenia taxonômica e a observada para a família gênica em estudo. Com isso, propõe-se a existência de características evolutivas diferenciadas para esses grupos de proteínas, mediadas por diferentes forças evolutivas, tais como eventos de expansão gênica e pressão de seleção.

5.2 História evolutiva das tripsinas

Algumas diferenças relevantes foram encontradas no sítio de ligação ao substrato das tripsinas nas três ordens de insetos analisadas. Enquanto que nessa região proteica Coleóptera e Díptera compartilham vários resíduos similares, em Lepidóptera tal região mostrou características singulares. Em particular, as enzimas de Lepidóptera apresentaram aumento do caráter hidrofílico do sub-sítio S1 e aumento da hidrofobicidade dos sub-sítios S2 e S4 em relação às demais ordens. Esses resultados se assemelham aos encontrados por Lopes et al. (2006), os quais ressaltaram uma maior hidrofobicidade no sub-sítio S3 de Lepidóptera em relação às outras duas ordens de insetos estudadas. Nos presentes resultados, uma maior hidrofobicidade não foi observada no sub-sítio S3 dessas proteínas, mas sim nos sub-sítios vizinhos a ele (S2 e S4), os quais podem ter alterado as características da região. Ademais, esses mesmos autores observaram que as tripsinas de Lepidóptera não apresentam preferência por arginina ou lisina na posição P1, enquanto Coleóptera e Díptera mostram a preferência usual por arginina (LOPES et al., 2006). Essa variação na especificidade do sub-sítio S1 em Lepidóptera pode estar correlacionada ao caráter mais hidrofílico deste, encontrado nas análises descritas no presente trabalho. Isso porque, como observado em tripsinas de mamíferos, uma troca de um aminoácido hidrofílico por um outro hidrofóbico, na região do sub-sítio S1, pode aumentar a preferência por arginina em relação a lisina (SICHLER et al., 2002). Mais especificamente, em mamíferos, observou-se que uma alanina na posição 190 aumenta em 5 vezes a preferência por uma arginina em P1 em relação a uma serina (mais hidrofílica) na mesma posição (SICHLER et al., 2002).

As tripsinas de Lepidóptera formaram dez agrupamentos de similaridade. Dentro dessa ordem, as sequências provenientes dos insetos da família Noctuidae foram encontradas em nove desses dez grupos, mostrando uma alta diversificação

nas tripsinas dessa família de Lepidóptera (Figura 23). Todavia, notadamente, seis dos dez grupos de Lepidóptera foram exclusivamente ocupados por sequências provenientes de insetos dessa família. Essa impressionante diversidade observada em Noctuidae pode ser o resultado de diversos mecanismos de adaptação, compreendendo desde a especialização a diferentes tipos de alimentos ao enriquecimento de um reservatório enzimático para estratégias de escape a inibidores de peptidases. Além disso, desses seis grupos, específicos da família Noctuidae, quatro deles (L-VI a L-X) formam um grupo de peptidases com características distintas de todas as demais, quando se observa as variações ocorridas no sítio de ligação ao substrato dessas proteínas. Interessantemente, a existência de dois grupos de tripsinas com características bastante diversas já foi anteriormente observada em insetos da família Noctuidae (PAULILLO et al., 2000; LOPES et al., 2004).

Lopes et al. (2004), observaram grupos distintos de sequências proteicas entre as tripsinas de insetos da família Noctuidae. Além disso, os autores identificaram uma correlação desses grupos com as sequências proteicas propostas como oriundas de genes sensíveis ou insensíveis a ingestão de inibidores de peptidases. Segundo essa classificação, grupos de genes que apresentam um aumento nos níveis de expressão são considerados insensíveis aos inibidores. Assim, genes que não alteram seu padrão de expressão ou são regulados negativamente, mediante à ingestão de inibidores, são considerados sensíveis. A comparação dos resultados obtidos no presente trabalho com os de Lopes et al. (2004) mostra que as sequências proteicas identificadas por eles como pertencentes a genes insensíveis à IPs (com regulação positiva na expressão) estão localizadas nos quatro últimos grupos de tripsinas, os quais foram identificados apenas na família Noctuidae (L-VI a L-X).

Souza (2013) também observou a presença de dois grupos de tripsinas com perfis de regulação da expressão gênica distintos em *S. frugiperda* (Lepidóptero da família Noctuidae) em resposta à inibidores de peptidase de soja, por tratamento crônico. Os resultados obtidos por Souza (2013) também corroboram os grupos apresentados por Lopes et al. (2004), mostrando que o grupo de tripsinas encontrado somente na família Noctuidae é composto por proteínas insensíveis ao inibidor de peptidase de soja. Adicionalmente, Castelhana (2014) testando a resposta de *Diatraea saccharalis* (Lepidóptero da família Crambidae) a inibidores de

Eventos de expansão gênica linhagem específica (*lineage-specific expansion* – LSE) são definidos como a proliferação de famílias proteicas em linhagens específicas, quando comparadas a linhagens irmãs (JORDAN et al., 2001; LESPINET et al., 2002). Esses eventos já foram identificados em diferentes proteínas (LESPINET et al., 2002; SHUKLA et al., 2014), incluindo enzimas da família das tripsinas em *Drosophila melanogaster* – quando comparada a outras quatro espécies de eucariotos (*Saccharomyces cerevisiae*, *Schizasaccharomyces pombe*, *Caenorhabditis elegans* e *Arabidopsis thaliana*; LESPINET et al., 2002) – e em *Anopheles gambiae* quando comparado à *D. melanogaster* (WU et al., 2009). Além disso, em *A. gambiae*, a expansão de enzimas do tipo tripsina é proposta como correlacionada à evolução da hematofagia, uma vez que os genes identificados como resultantes de expansão específica na linhagem, compreendem os genes com maior aumento nos níveis de expressão após a ingestão de sangue (WU et al., 2009).

Assim, nossos resultados sugerem que eventos de expansão gênica específicos à linhagem da família Noctuidae podem ter ocorrido durante a evolução do grupo. Não obstante, a presença de tal expansão parece estar correlacionada à adaptação desses insetos à ingestão de inibidores de peptidase, uma vez que tais genes sofrem aumento nos níveis de expressão quando esses insetos são submetidos a dietas contendo IPs.

5.3 Auto-ativação do tripsinogênio

A necessidade de uma digestão eficiente é particularmente forte em insetos com o hábito de alimentação incessante e com o uso de diversas fontes de energia. A adaptação a esses hábitos pode ter induzido esses insetos a evoluir um grande número de enzimas diversificadas, necessárias para o processo digestivo, as quais podem oferecer, por um lado, auto-ativação, como uma estratégia simples para aumentar a velocidade de ativação e, por outro lado, oportunidades para escapar dos inibidores.

Chen et al. (2003) mostraram que a sequência de ativação do tripsinogênio de muitos invertebrados diverge do consenso observado em mamíferos, sugerindo que um processo de auto-ativação pode ocorrer nas tripsinas desses organismos.

Em humanos, o sítio de ativação do tripsinogênio é composto por um motivo conservado de quatro aspartatos consecutivos seguidos por uma lisina. Mutações nessa região, as quais incluem D22G em P2 (TEICH et al., 2000) e K23R em P1 (FEREC et al., 1999), estão correlacionadas a pancreatites causadas pelo excesso de tripsinas ativas, provenientes do aumento nas taxas auto-ativação. Nos insetos estudados, esse motivo conservado está ausente e a composição de resíduos observada em casos de auto-ativação elevada em humanos está presente em grande parte das sequências (Gly em P2 e Arg em P1). Além disso, o sítio de ativação do tripsinogênio de Lepidópteros, apresenta características convergentes as observadas no sítio de ligação ao substrato das tripsinas desses organismos, ressaltando a ideia de que a auto-ativação deve ser facilitada nessa ordem.

Assim, as análises do motivo de ativação do tripsinogênio dos insetos estudados mostraram que as tripsinas de Lepidóptera tem uma maior tendência à auto-ativação do que as enzimas de Coleóptera e Díptera, as quais, por sua vez, também parecem ser mais propensas a apresentar auto-ativação do que as tripsinas de mamíferos.

Não obstante, essa hipótese ainda corrobora com o fato de uma enzima semelhante à enteroquinase de mamíferos, especializada na ativação do tripsinogênio, ainda não ter sido identificada em insetos.

O presente estudo mostra que as tripsinas de insetos tem claramente seguido um caminho evolutivo peculiar quando comparadas com os seus homólogos em mamíferos. Em particular, os resultados fortemente suportam a hipótese de que as tripsinas são enzimas chaves na adaptação de insetos aos seus distintos hábitos alimentares, por estarem envolvidas no processo de ativação de outras peptidases e delas próprias, proporcionando assim o rápido suprimento de enzimas funcionais.

5.4 O papel de outras regiões proteicas na história evolutiva das tripsinas de Lepidóptera

Duas regiões de *loops* foram observadas como possuindo características singulares nas tripsinas de Lepidóptera. A primeira, compreendendo o *loop* composto pelos resíduos 146 e 151, os quais foram observados sob pressão de seleção positiva; e a segunda, compreendendo o *loop* entre as posições 169 e 174,

o qual apresentou uma grande e conservada inserção nas enzimas desses organismos.

A pressão de seleção positiva, também conhecida como pressão de seleção diversificadora, tende a manter mutações que ocorram em uma determinada posição da proteína. As razões pelas quais as proteínas apresentem sítios sob pressão de seleção positiva, incluem: corrida armamentista evolutiva, tal como na interação hospedeiro-patógeno; substituições compensatórias, para compensar uma mutação deletéria com outra; e adaptações da molécula a novas funções ou ambientes (PÁL; PAPP; LERCHER, 2006).

O *loop* que engloba as posições 141 e 151 é conhecido como correlacionado com variações na especificidade de diferentes peptidases (SCHELLENBERGER; TURCK; RUTTER, 1994). Isso leva a hipótese de que esse *loop* possa estar envolvido em um aumento na diversidade de substratos das tripsinas de Lepidóptera.

Finalmente, no que se refere à inserção no *loop* entre as posições 169 e 174, especula-se que essa região possa estar envolvida na interação dessas peptidases com o sBBI. Além disso, de acordo com os resultados de docagem estrutural essa região pode desencadear uma “competição” entre tripsinas e quimotripsinas por uma mesma região do sBBI. Uma das hipóteses levantadas é que a interação desse *loop* com a região do sBBI, ligada à inibição das quimotripsinas, possa causar um desbalanceamento nas proporções de inibição de ambas as proteínas. No entanto, essa hipótese é fracamente suportada pelos presentes resultados e estudos bioquímicos ainda precisam ser realizados.

Ambas as teorias, envolvendo essas duas regiões de *loops* das tripsinas, ressaltam que regiões de alta variabilidade e mais distantes do sítio catalítico, podem possuir papel relevante na evolução dessas enzimas.

5.5 História evolutiva divergente entre tripsinas e quimotripsinas

Em contraste à alta conservação observada nas tripsinas, as quimotripsinas mostraram uma maior variabilidade em seus sub-sítios de ligação ao substrato (Figura 24), como também já foi demonstrado por análises bioquímicas (SATO et al., 2008). Essa observação sugere uma história evolutiva divergente da encontrada para as tripsinas.

De acordo com Perona e Craik (1997), organismos que divergiram primeiro (tais como procariotos e invertebrados) mostram menor número de enzimas da família das quimotripsinas em relação a eucariotos e vertebrados. No entanto, segundo os autores, as enzimas desses primeiros organismos possuem maior abrangência de substratos em contraste com as dos segundos, os quais possuem enzimas com funções mais especializadas. Essa teoria pode explicar a grande variabilidade observada na especificidade das quimotripsinas dos insetos estudados, assim como, a falta de uma peptidase específica para a ativação das tripsinas.

A maior plasticidade nos sub-sítios das quimotripsinas parece oferecer maior diversidade de especificidade pelo substrato, assim como, pode representar uma estratégia de escape a inibidores. Isso porque, assim como discutido nos resultados, posições de grande variação de resíduos, tais como a 192 (localizada no sub-sítio S1), podem ser responsáveis por alterações na interação dessas peptidases com inibidores, tais como o sBBI.

Nas quimotripsinas, a presença de uma monofilia exclusiva da família Noctuidae, formada pelos grupos Lc-X a Lc-XII, também sugere que uma expansão gênica específica para essa linhagem pode ter ocorrido nesse grupo de proteínas. Contudo, essas proteínas não apresentaram variações relevantes nas suas características, que suportem uma adaptação específica mediada por essas enzimas.

As histórias evolutivas das tripsinas e quimotripsinas dos insetos estudados parecem ser resultantes de diferentes forças evolutivas. E embora ambos os conjuntos proteicos tenham características que indiquem a expansão da família por duplicação gênica, a pressão de seleção parece atuar diferentemente em ambas as proteínas. Como observado nos sítios de ligação ao substrato, as tripsinas estão sujeitas a pressão de seleção negativa em quase todos as posições de resíduos analisadas, razão pela qual, mesmo com diferentes cópias, essas proteínas não apresentem grande diversificação funcional ou de especificidade.

Figura 24 – Representação gráfica das posições estruturais conservadas na superfície de tripsinas e quimotripsinas das diferentes ordens de insetos estudadas. A figura foi produzida pelo Webserver ConSurf (ASHKENAZY *et al.*, 2010). As posições destacadas em amarelo mostram posições onde nenhum resíduo foi alinhado a sequência molde: PDB: 4I8H (tripsina bovina) e PDB: 1EQ9 (quimotripsina de *S. invicta*)

6 CONSIDERAÇÕES FINAIS

- Sequências de peptidases da ordem Lepidóptera possuem características diversas nas suas regiões funcionais, quando comparadas a Coleóptera e Díptera, as quais são mais similares entre si;
- Tripsinas mostram maior conservação no sítio de ligação ao substrato em relação as quimotripsinas;
- Tripsinas de lepidópteros são mais hidrofílicas na região do sub-sítio S1 e mais hidrofóbicas nas regiões S2-S4 do que os seus homólogos em Coleóptera e Díptera;
- A família Noctuidae apresentou a maior diversidade de sequências dentro da ordem Lepidóptera, possuindo inclusive um grupo de proteínas com características similares, o qual engloba apenas sequências dessa família;
- Todas as tripsinas de insetos analisadas parecem ser ativadas por um mecanismo de auto-ativação, uma vez que acumulam características no sítio de ativação do zimogênio que favorecem a ligação de outras tripsinas a região;
- Em geral, nossos resultados levantam uma hipótese de como insetos das ordens Lepidóptera, Coleóptera e Díptera puderam se adaptar aos seus diferentes hábitos de alimentação e explica porque enzimas especializadas na ativação de tripsinas, tais como a enteroquinase de porcos, não tem sido observadas nesses organismos.

REFERÊNCIAS

- ABASCAL, F.; ZARDOYA, R.; POSADA, D. ProtTest: selection of best-fit models of protein evolution. **Bioinformatics**, Oxford, v. 21, n. 9, p. 2104-2105, May 2005.
- ABASCAL, F.; ZARDOYA, R.; TELFORD, M.J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. **Nucleic Acids Research**, London, v. 38, p. W7-W13, July 2010.
- ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z. MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, London, v. 25, n. 17, p. 3389-402, Sept. 1997.
- ASHKENAZY, H.; EREZ, E.; MARTZ, E.; PUPKO, T.; BEN-TAL, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. **Nucleic Acids Research**, London, v. 38, p. W529-W533, July 2010.
- BAPTISTA, A.M.; JONSON, P.H.; HOUGH, E.; PETERSEN, S.B. The origin of trypsin: evidence for multiple gene duplications in trypsins. **Journal of Molecular Evolution**, Berlin, v. 47, n. 3, p. 353-362, Sept. 1998.
- BARROS, E.M.; TORRES, J.B.; RUBERSON, J.R.; OLIVEIRA, M.D. Development of *Spodoptera frugiperda* on different hosts and damage to reproductive structures in cotton. **Entomologia Experimentalis Et Applicata**, Amsterdam, v. 137, n. 3, p. 237-245, Dec. 2010.
- BIRK, Y.; GERTLER, A.; KHALEF, S. A pure trypsin inhibitor from soybean. **Journal of Biochemistry**, Tokyo, n. 87, p. 281-284, 1963.
- BODE, W.; MEYER, E. JR.; POWERS, J.C. Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors: review. **Biochemistry**, Washington, v. 28, n. 5, p. 1951-1963, Mar. 1989.
- BOLTER, C.; JONGSMA, M.A. The adaptation of insects to plant protease inhibitors. **Journal of Insect Physiology**, London, v. 43, n. 10, p. 885-895, Oct. 1997.
- BOTOS, I.; MEYER, E.; NGUYEN, M.; SWANSON, S.M.; KOOMEN, J.M.; RUSSELL, D.H.; MEYER, E.F. The structure of an insect chymotrypsin. **Journal of Molecular Biology**, London, v. 298, n. 5, p. 895-901, May 2000.
- BOULTER, D. Insect pest control by copying nature using genetically engineered crops. **Photochemistry**, New York, v. 34, p. 1453-1466, 1993.
- BOWMAN, D.E. Differentiation of soybean antitryptic factor. **Proceedings of the Society for Experimental Biology and Medicine**, Malden, v. 63, p. 547-550, 1946.
- BOWN, D.P.; WILKINSON, H.S.; GATEHOUSE, J.A. Regulation of expression of genes encoding digestive proteases in the gut of a polyphagous lepidoptera larva in response to dietary protease inhibitors. **Physiological Entomology**, London, v. 29, p. 278-290, 2004.

BRANDÃO, M.M.; SILVA, M.C. Evolutionary history of *Arabidopsis thaliana* aminoacyl-tRNA synthetase dual-targeted proteins. **Molecular Biology and Evolution**, Chicago, v. 28, n. 1, p. 79-85, Jan 2011.

BRIOSCHI, D.; NADALINI, L.D.; BENGTSON, M.H.; SOGAYAR, M.C.; MOURA, D. S.; SILVA-FILHO, M.C. General up regulation of *Spodoptera frugiperda* trypsins and chymotrypsins allows its adaptation to soybean proteinase inhibitor. **Insect Biochemistry and Molecular Biology**, Oxford, v. 37, n. 12, p. 1283-1290, Dec. 2007.

BRITO, L.O.; LOPES, A.R.; PARRA, J.R.P.; TERRA, W.R.; SILVA-FILHO, M.C. Adaptation of tobacco budworm *Heliothis virescens* to proteinase inhibitors may be mediated by the synthesis of new proteinase. **Comparative Biochemistry and Physiology**, Oxford, v. 128B, p. 365-375, 2001.

BROADWAY, R.M. Are insects resistant to plant proteinase inhibitor? **Journal of Insect Physiology**, Oxford, v. 41, p. 107-116, 1995.

_____. Dietary proteinase inhibitors alter complement of midgut proteases. **Archives of Insect Biochemistry and Physiology**, New York, v. 32, p. 39-53, 1996.

_____. Dietary regulation of serine proteinases that are resistant to serine proteinase inhibitors. **Journal of Insect Physiology**, London, v. 43, n. 9, p. 855-874, Sept. 1997.

BURGESS, E.P.J.; STEVENS, P.S.; KEEN, G.K.; LAING, W.A.; CHRISTELLER, J.T. Effects of protease inhibitors and dietary-protein level on the black field cricket *Teleogryllus-Commodus*. **Entomologia Experimentalis Et Applicata**, London, v. 61, n. 2, p. 123-130, Nov. 1991.

CASTELHANO, E.C. **Efeito dos inibidores de peptidase de soja no padrão de expressão e atividade enzimática intestinal de lagartas de *Diatraea saccharalis* (Fabricius, 1794) (Lepidoptera: Crambidae)**. 2014. 91 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2013.

CHEN, J.M.; KUKOR, Z.; LE MARECHAL, U.; TOTH, M.; TSAKIRIS, L.; RAGUENES, O.; FEREC, C.; SAHIN-TOTH. Evolution of trypsinogen activation peptides. **Molecular Biology and Evolution**, Chicago, v. 20, n. 11, p. 1767-1777, Nov. 2003.

CONESA, A.; GÖTZ, S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. **International journal of plant genomics**, New York, v. 2008, p. 1-12, 2008.

CRAIK, C.S.; LARGMAN, C.; FLETCHER, T.; ROCZNIAK, S.; BARR, P.J.; FLETTERICK, R.; RUTTER, W.J. Redesigning trypsin - alteration of substrate-specificity. **Science**, Washington, v. 228, n. 4697, p. 291-297, Apr. 1985.

DARRIBA, D.; TABOADA, G.L.; DOALLO, R.; POSADA, D. jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, New York, v. 9, n. 8, p. 772-772, Aug. 2012.

DE LEO, F.; BONADE-BOTTINO, M.A.; CECI, L.R.; GALLERANI, R.; JOUANIN, L. Opposite effects on *Spodoptera littoralis* larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants. **Plant Physiology**, Lancaster, v. 118, n. 3, p. 997-1004, Nov. 1998.

DE VRIES, S.J.; DIJK, M. van; BONVIN, A. The HADDOCK web server for data-driven biomolecular docking. **Nature Protocols**, London, v. 5, n. 5, p. 883-897, May 2010.

EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, London, v. 32, n. 5, p. 1792-1797, Mar. 2004.

EISENBERG, D.; SCHWARZ, E.; KOMAROMY, M.; WALL, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. **Journal of Molecular Biology**, London, v. 179, n. 1, p. 125-142, 1984.

ESTEVEES, G.F.; TELES, R.C.; CAVALCANTE, N.S.; NEVES, D.; VENTURA, M.M.; BARBOSA, J.A.; FREITAS, S.M. de. Crystallization, data collection and processing of the chymotrypsin-BTCl-trypsin ternary complex. **Acta Crystallografica. Section F, Structural Biology and Crystallization Communications**, Chester, v. 63, n. 12, p. 1087-1090, Dec. 2007.

ESWAR, N.; WEBB, B.; MARTI-RENOM, M.A.; MADHUSUDHAN, M.S.; ERAMIAN, D.; SHEN, M.Y.; PIEPER, U.; SALI, A. Comparative protein structure modeling using Modeller. **Current protocols in bioinformatics**, New York, p. 5.6.1-5.6.30, Nov. 2007.

FALCO, M.C.; MARBACH, P.A.S.; POMPERMAYER, P.; LOPES, F.C.C.; SILVA-FILHO, M.C. Mechanisms of sugarcane response to herbivory. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1-4, p. 113-122, Jan. 2001.

FEREC, C.; RAGUENES, O.; SALOMON, R.; ROCHE, C.; BERNARD, J.P.; GUILLOT, M.; QUERE, I.; FAURE, C.; MERCIER, B.; AUDREZET, M.P.; GUILLAUSSEAU, P.J.; DUPONT, C.; MUNNICH, A.; BIGNON, J.D.; LE BODIC, L. Mutations in the cationic trypsinogen gene and evidence for genetic heterogeneity in hereditary pancreatitis. **Journal of Medical Genetics**, London, v. 36, n. 3, p. 228-232, Mar. 1999.

GATEHOUSE, A.M.R.; YING, S.; POWELL, K.S.; BROUGH, C.; HILDER, V.A.; HAMILTON, W.D.O.; NEWELL, C.A.; MERRYWEATHER, A.; BOULTER, D.; GATEHOUSE, J.A. Approaches to insect resistance using transgenic plants. **Philosophical Transactions of the Royal Society of London. Series B-Biological Sciences**, London, v. 342, n. 1301, p. 279-286, Nov. 1993.

GLADYSHEVA, I.P.; ZAMOLODCHIKOVA, T.S.; SOKOLOVA, E.A.; LARIONOVA, N.I. Interaction between duodenase, a proteinase with dual specificity, and soybean inhibitors of Bowman-Birk and Kunitz type. **Biochemistry**, Washington, v. 64, n. 11, p. 1244-1249, Nov. 1999.

GOLDMAN, N.; YANG, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Molecular Biology and Evolution**, Chicago, v. 11, n. 5, p. 725-736, Sept. 1994.

GOTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; WILLIAMS, T.D.; NAGARAJ, S.H.; NUEDA, M.J.; ROBLES, M.; TALON, M.; DOPAZO, J.; CONESA, A. High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic Acids Research**, London, v. 36, n. 10, p. 3420-3435, June 2008.

GREEN, T.R.; RYAN, C.A. Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. **Science**, Washinton, v. 175, n. 4023, p. 776-7, Feb. 1972.

HADORN, B.; TARLOW, M.J.; LLOYD, J.K.; WOLFF, O.H. Intestinal enterokinase deficiency. **The Lancet**, London, v. 293, n. 7599, p. 812-813, Apr. 1969.

HEDSTROM, L. Serine protease mechanism and specificity. **Chemical Reviews**, Easton, v. 102, n. 12, p. 4501-4523, Dec. 2002.

HEDSTROM, L.; SZILAGYI, L.; RUTTER, W.J. Converting trypsin to chymotrypsin - the role of surface loops. **Science**, New York, v. 255, n. 5049, p. 1249-1253, Mar. 1992.

JONGSMA, M.A.; BAKKER, P.L.; PETERS, J.; BOSCH, D.; STIEKEMA, W.J. Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 92, n. 17, p. 8041-5, Aug. 1995.

JORDAN, I.K.; MAKAROVA, K.S.; SPOUGE, J.L.; WOLF, Y.I.; KOONIN, E.V. Lineage-specific gene expansions in bacterial and archaeal genomes. **Genome Research**, Cold Spring Harbor, v. 11, n. 4, p. 555-565, Apr. 2001.

KIM, D.E.; CHIVIAN, D.; BAKER, D. Protein structure prediction and analysis using the Robetta server. **Nucleic Acids Research**, London, v. 32, n. 2, p. W526-W531, 2004.

KITAMOTO, Y.; YUAN, X.; WU, Q.; McCOURT, D.W.; SADLER, J.E. Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 91, n. 16, p. 7588-7592, Aug. 1994.

KOEPKE, J.; ERMLER, U.; WARKENTIN, E.; WENZL, G.; FLECKER, P. Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 angstrom resolution. Structural basis of Janus-faced serine protease inhibitor specificity. **Journal of Molecular Biology**, London, v. 298, n. 3, p. 477-491, May 2000.

KORTEMME, T.; KIM, D.E.; BAKER, D. Computational Alanine Scanning of Protein-Protein Interfaces. **Science's STKE: Signal Transduction Knowledge Environment**, Washington, v. 219, p. 1-8, Feb. 2004.

LESPINET, O.; WOLF, Y.I.; KOONIN, E.V.; ARAVIND, L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. **Genome Research**, Cold Spring Harbor, v. 12, n. 7, p. 1048-1059, July 2002.

LOPES, A.R.; JULIANO, M.A.; JULIANO, L.; TERRA, W.R. Coevolution of insect trypsins and inhibitors. **Archives of Insect Biochemistry and Physiology**, New York, v. 55, n. 3, p. 140-152, Mar. 2004.

LOPES, A.R.; JULIANO, M.A.; MARANA, S.R.; JULIANO, L.; TERRA, W.R. Substrate specificity of insect trypsins and the role of their subsites in catalysis. **Insect Biochemistry and Molecular Biology**, Oxford, v. 36, n. 2, p. 130-140, Feb. 2006.

MAROUX, S.; BARATTI, J.; DESNUELLE, P. Purification and specificity of porcine enterokinase. **Journal of Biological Chemistry**, Baltimore, v. 246, n. 16, p. 5031-5039, Aug. 1971.

MARTIN, E.A.; REINEKING, B.; SEO, B.; STEFFAN-DEWENTER, I. Natural enemy interactions constrain pest control in complex agricultural landscapes. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 110, n. 14, p. 5534-5539, Apr. 2013.

MCMANUS, M. T.; WHITE, D. W. R.; MCGREGOR, P. G. Accumulation of a chymotrypsin inhibitor in transgenic tobacco can affect the growth of insect pests. **Transgenic Research**, London, v. 3, n. 1, p. 50-58, Jan. 1994.

MELLO, M.O.; SILVA-FILHO, M.C. Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms. **Brazilian Journal of Plant Physiology**, Londrina, v. 14, n. 2, p. 71-81, May 2002.

MULLER, H.M.; CRAMPTON, J.M.; TORRE, A. della; SINDEN, R.; CRISANTI, A. Members of a trypsin gene family in anopheles-gambiae are induced in the gut by blood meal. **Embo Journal**, Eynsham, v. 12, n. 7, p. 2891-2900, July 1993.

NEURATH, H. Evolution of proteolytic enzymes. **Science**, Washington, v. 224, p. 350-357, Apr. 1984.

NEW, T.R. Moths (Insecta: Lepidoptera) and conservation: background and perspective. **Journal of Insect Conservation**, London, v. 8, n. 2-3, p. 79-94, June 2004.

NOTREDAME, C.; HIGGINS, D. G.; HERINGA, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. **Journal of Molecular Biology**, London, v. 302, n. 1, p. 205-17, Sept. 2000.

OLIVEIRA, C.M.; AUAD, A.M.; MENDES, S.M.; FRIZZAS, M.R. Economic impact of exotic insect pests in Brazilian agriculture. **Journal of Applied Entomology**, Berlin, v. 137, p. 1-15, 2013.

_____; AUAD, A.M.; MENDES, S.M.; FRIZZAS, M.R. Crop losses and the economic impact of insect pests on Brazilian agriculture. **Crop Protection**, Guildford, v. 56, p. 50-54, Feb. 2014.

PÁL, C.; PAPP, B.; LERCHER, M.J. An integrated view of protein evolution. **Nature Reviews. Genetics**, London, v. 7, n. 5, p. 337-348, May 2006.

PAULILLO, L.C.; LOPES, A.R.; CRISTOFOLETTI, P.T.; PARRA, J.R.; TERRA, W.R.; SILVA-FILHO, M.C. Changes in midgut endopeptidase activity of *Spodoptera frugiperda* (Lepidoptera : Noctuidae) are responsible for adaptation to soybean proteinase inhibitors. **Journal of Economic Entomology**, College Park, v. 93, n. 3, p. 892-896, June 2000.

PERONA, J.J.; CRAIK, C.S. Structural basis of substrate specificity in the serine proteases: review. **Protein Science**, New York, v. 4, n. 3, p. 337-360, Mar. 1995.

_____. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. **The Journal of Biological Chemistry**, Baltimore, v. 272, n. 48, p. 29987-29990, Nov 1997.

PETTERSEN, E.F.; GODDARD, T.D.; HUANG, C.C.; COUCH, G.S.; GREENBLATT, D.M.; MENG, E.C.; FERRIN, T.E. UCSF chimera: a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, New York, v. 25, n. 13, p. 1605-1612, Oct. 2004.

POMPERMAYER, P.; LOPES, A.R.; TERRA, W.R.; PARRA, J.R.P.; FALCO, M.C.; SILVA-FILHO, M.C. Effects of soybean proteinase inhibitor on development, survival and reproductive potential of the sugarcane borer, *Diatraea saccharalis*. **Entomologia Experimentalis et Applicata**, Amsterdam, v. 99, n. 1, p. 79-85, Apr. 2001.

RAKWAL, R.; KUMAR AGRAWAL, G.; JWA, N.-S. Characterization of a rice (*Oryza sativa* L.) Bowman-Birk proteinase inhibitor: tightly light regulated induction in response to cut, jasmonic acid, ethylene and protein phosphatase 2A inhibitors. **Gene**, Amsterdam, v. 263, n. 1, p. 189-198, Jan. 2001.

RAWLINGS, N. D.; BARRETT, A. J. Evolutionary families of peptidases. **Biochemical Journal**, London, v. 290, p. 205-218, Feb. 1993.

RINDERKNECHT, H. Activation of pancreatic zymogens. **Digestive diseases and sciences**, New York, v. 31, n. 3, p. 314-321, Mar. 1986.

RONQUIST, F.; HUELSENBECK, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics**, Oxford, v. 19, n. 12, p. 1572-4, Aug. 2003.

ROSS, J.; JIANG, H.; KANOST, M.R.; WANG, Y. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. **Gene**, Amsterdam, v. 304, p. 117-131, Jan. 2003.

RYAN, C. Protease inhibitors in plants: genes for improving defenses against insects and pathogens. **Annual Review of Phytopathology**, Palo Alto, v. 28, p. 425-449, 1990.

SATO, P.M.; LOPES, A.R.; JULIANO, L.; JULIANO, M.A.; TERRA, W.R. Subsite substrate specificity of midgut insect chymotrypsins. **Insect Biochemistry and Molecular Biology**, Oxford, v. 38, n. 6, p. 628-633, June 2008.

SCHECHTER, I.; BERGER, A. On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. **Biochemical and biophysical research communications**, New York, v. 32, n. 5, p. 898-902, Sept. 1968.

SCHELLENBERGER, V.; TURCK, C. W.; RUTTER, W. J. Role of the s' subsites in serine-protease catalysis - active-site mapping of rat chymotrypsin, rat trypsin, alpha-lytic protease, and cercarial protease from schistosoma-mansoni. **Biochemistry**, Washington, v. 33, n. 14, p. 4251-4257, Apr. 1994.

SEETHARAM, A.; STUART, G. W. Whole genome phylogenies for multiple *Drosophila* species. **BMC Research Notes**, London, v. 5, n. 1, p. 670, Dec. 2012.

SHUKLA, V.; HABIB, F.; KULKARNI, A.; RATNAPARKHI, G.S. Gene duplication, lineage-specific expansion, and subfunctionalization in the MADF-BESS family patterns the *Drosophila* wing line. **Genetics**, Austin, v. 196, n. 2, p. 481-496, Feb. 2014.

SICHLER, K.; HOPFNER, K.P.; KOPETZKI, E.; HUBER, R.; BODE, W.; BRANDSTETTER, H. The influence of residue 190 in the S1 site of trypsin-like serine proteases on substrate selectivity is universally conserved. **Febs Letters**, Amsterdam, v. 530, n. 1/3, p. 220-224, Oct. 2002.

SODING, J.; BIEGERT, A.; LUPAS, A.N. The HHpred interactive server for protein homology detection and structure prediction. **Nucleic Acids Research**, London, v. 33, p. W244-W248, July 2005.

SOMMER, D.D.; DELCHER, A.L.; SALZBERG, S.L.; POP, M. Minimus: a fast, lightweight genome assembler. **Bmc Bioinformatics**, London, v. 8, n. 64, p. 1-11, Feb. 2007.

SOUZA, T.P. de. **Efeito dos inibidores de proteinase de soja no padrão de expressão de proteinases de *Spodoptera frugiperda***. 2013. 85 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2013.

SRINIVASAN, A.; GIRI, A.P.; HARSULKAR, A.M.; GATEHOUSE, J.A.; GUPTA, V.S. A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae. **Plant Molecular Biology**, Boston, v. 57, p. 359-374, 2005.

TEICH, N.; OCKENGA, J.; HOFFMEISTER, A.; MANNS, M.; MÖSSNER, J.; KEIM, V. Chronic pancreatitis associated with an activation peptide mutation that facilitates trypsin activation. **Gastroenterology**, Baltimore, v. 119, n. 2, p. 461-465, Aug. 2000.

TERRA, W. R.; FERREIRA, C. Insect digestive enzymes - properties, compartmentalization and function. **Comparative Biochemistry and Physiology. B. Biochemistry & Molecular Biology**, Oxford, v. 109, n. 1, p. 1-62, Sept. 1994.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, London, v. 22, n. 22, p. 4673-4680, Nov. 1994.

TINA, K.G.; BHADRA, R.; SRINIVASAN, N. PIC: Protein Interactions Calculator. **Nucleic Acids Research**, London, v. 35, p. W473-W476, July 2007.

TRAUTWEIN, M.D.; WIEGMANN, B.M.; BEUTEL, R.; KJER, K.M.; YEATES, D.K. Advances in insect phylogeny at the dawn of the postgenomic era. **Annual Review of Entomology**, Palo Alto, v. 57, p. 449-468, Jan. 2012.

VEITIA, R.A.; CABURET, S. Extensive sequence turnover of the signal peptides of members of the GDF/BMP family: exploring their evolutionary landscape. **Biology Direct**, London, v. 4, n. 22, p. 1-9, July 2009.

WALDVOGEL, M.; GOULD, F. Variation in oviposition preference of *Heliothis virescens* in relation to macroevolutionary patterns of heliothine host range. **Evolution**, New York, v. 44, n. 5, p. 1326-1337, Aug. 1990.

WANG, S.J.; MAGOULAS, C.; HICKEY, D. Concerted evolution within a trypsin gene cluster in *Drosophila*. **Molecular Biology and Evolution**, Chicago, v. 16, n. 9, p. 1117-1124, Sept. 1999.

WATERFIELD, G.; ZILBERMAN, D. Pest management in food systems: an economic perspective. **Annual Review of Environment and Resources**, Palo Alto, v. 37, p. 223-245, May 2012.

WOLFSON, J.L.; MURDOCK, A.L. Potential use of protease inhibitors for host plant resistance: a test case. **Environmental entomology**, College Park, v. 24, n. 1, p. 52-57, Feb. 1995.

WU, D.D.; WANG, G.D.; IRWIN, D.M.; ZHANG, Y.P. A profound role for the expansion of trypsin-like serine protease family in the evolution of hematophagy in mosquito. **Molecular Biology and Evolution**, Chicago, v. 26, n. 10, p. 2333-2341, Oct. 2009.

YANG, Z.; NIELSEN, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. **Journal of Molecular Evolution**, Berlin, v. 46, p. 409-418, Apr. 1998.

_____. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. **Molecular Biology and Evolution**, Chicago, v. 19, p. 908-917, Jun. 2002.

_____. PAML 4: phylogenetic analysis by maximum likelihood. **Molecular biology and evolution**, Chicago, v. 24, n. 8, p. 1586-1591, May 2007.

ZDOBNOV, E. M.; MERING, C. von; LETUNIC, I.; TORRENTS, D.; SUYAMA, M.; COPLEY, R.R.; CHRISTOPHIDES, G.K.; THOMASOVA, D.; HOLT, R.A.; SUBRAMANIAN, G.M.; MUELLER, H.M.; DIMOPOULOS, G.; LAW, J.H.; WELLS, M.A.; BIRNEY, E.; CHARLAB, R.; HALPERN, A.L.; KOKOZA, E.; KRAFT, C.L.; LAI, Z.; LEWIS, S.; LOUIS, C.; BARILLAS-MURY, C.; NUSSKERN, D.; RUBIN, G.M.; SALZBERG, S.L.; SUTTON, G.G.; TOPALIS, P.; WIDES, R.; WINCKER, P.; YANDELL, M.; COLLINS, F.H.; RIBEIRO, J.; GELBART, W.M.; KAFATOS, F.C.; BORK, P. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. **Science**, Washington, v. 298, n. 5591, p. 149-159, Oct. 2002.

ZERBINO, D.R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, Cold Spring Harbor, v. 18, n. 5, p. 821-829, May 2008.

ZHAO, P.; WANG, G.H.; DONG, Z.M.; DUAN, J.; XU, P.Z.; CHENG, T.C.; XIANG, Z.H.; XIA, Q.Y. Genome-wide identification and expression analysis of serine proteases and homologs in the silkworm *Bombyx mori*. **Bmc Genomics**, London, v. 11, n. 405, p. 1-11, June 2010.

ANEXOS

Anexo A – Descrição das sequências de tripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continua)		
Família	Espécie	Tripsina ID	GI da Proteína		
Coleóptera					
Bostrichidae	<i>Rhyzopertha dominica</i>	CRdom1	4868339		
		CRdom2	4868337		
		CRdom3	4868335		
Chrysomelidae	<i>Zabrotes subfasciatus</i>	CZsub1	91984091		
	<i>Phaedon cochleariae</i>	CPcoc1	74834620		
Curculionidae	<i>Dendroctonus ponderosae</i>	CDpon1	478249814		
		CDpon2	478249812		
		CDpon3	478249813		
		CDpon4	478249778		
		CDpon5	478252183		
		CDpon7	332373666		
		<i>Diaprepes abbreviatus</i>	CDabb1	3288964	
Lampyridae	<i>Pyrocoelia rufa</i>	CPruf1	60499217		
Scarabaeidae	<i>Costelytra zealandica</i>	CCzea2	166919225		
		CCzea3	166919223		
		<i>Trypoxylus dichotomus</i>	CTdic1	313056339	
Tenebrionidae	<i>Tribolium castaneum</i>	CTcas3	91087943		
		CTcas4	270002930		
		CTcas6	189234734		
		CTcas8	270006344		
		CTcas9	91081845		
		CTcas10	282721014		
		CTcas11	270002837		
		CTcas12	189241758		
		<i>Tenebrio molitor</i>	CTmol2	61393532	
		Díptera			
		Calliphoridae	<i>Lucilia cuprina</i>	DLcup1	464952
		Cecidomyiidae	<i>Mayetiola destructor</i>	DMdes1	49472792
DMdes2	49472796				
DMdes3	50812469				
Ceratopogonidae	<i>Culicoides sonorensis</i>	DCson1	56199540		
		DCson2	56199546		
		DCson3	56199550		
		DCson4	56199562		
Drosophilidae	<i>Drosophila melanogaster</i>	DDmel1	116008020		
		DDmel2	17136562		

Anexo A – Descrição das sequências de tripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

Família	Espécie	Tripsina ID	(continuação)
			GI da Proteína
		DDmel3	17648133
		DDmel4	17933738
		DDmel5	17986085
		DDmel6	17986087
		DDmel7	17986089
		DDmel8	19922026
		DDmel9	20129689
		DDmel10	21358525
		DDmel11	221330452
		DDmel12	24581073
		DDmel13	24581075
		DDmel14	24581087
		DDmel15	24581459
		DDmel16	24582982
		DDmel17	24652706
		DDmel18	24652708
		DDmel19	24653378
		DDmel20	24653960
		DDmel21	78707268
	<i>Drosophila pseudoobscura pseudoobscura</i>	DDpse1	125810456
		DDpse2	125810648
		DDpse3	125810650
		DDpse4	125810654
		DDpse5	125810656
		DDpse6	125810660
		DDpse7	125811872
		DDpse8	125985879
		DDpse9	198449735
		DDpse10	198453877
		DDpse11	198456899
		DDpse12	198458797
		DDpse13	198458836
		DDpse14	198459961
		DDpse15	198459963
		DDpse16	198459968
		DDpse17	198459972
		DDpse18	198472090
	<i>Drosophila virilis</i>	DDvir1	195380341

Anexo A – Descrição das sequências de tripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continuação)
Família	Espécie	Tripsina ID	GI da Proteína
		DDvir2	195381035
		DDvir3	195381037
		DDvir4	195381041
		DDvir5	195381045
		DDvir6	195381047
		DDvir7	195381053
		DDvir8	195381055
		DDvir9	195381057
		DDvir10	195381063
		DDvir11	195381363
		DDvir12	195381365
		DDvir13	195381371
		DDvir14	195381723
		DDvir15	195381751
		DDvir16	195384784
		DDvir17	195387724
		DDvir18	195395510
		DDvir19	195398586
Glossinidae	<i>Glossina morsitans morsitans</i>	DGmor1	182676658
		DGmor2	9651986
Muscidae	<i>Stomoxys calcitrans</i>	DScal1	3319320
Oestridae	<i>Hypoderma diana</i>	DHdia1	194719377
	<i>Hypoderma lineatum</i>	DHlin1	3123228
		DHlin2	547699
		DHlin3	587137
Psychodidae	<i>Lutzomyia longipalpis</i>	DLlon1	120556922
		DLlon2	120556924
		DLlon3	157674409
		DLlon4	157674425
	<i>Phlebotomus papatasi</i>	DPpap1	32394738
		DPpap2	32394740
		DPpap3	32394742
		DPpap4	32394744
Sarcophagidae	<i>Neobellieria bullata</i>	DNbul1	1717788
Simuliidae	<i>Simulium vittatum</i>	DSvit1	464967
Stratiomyidae	<i>Hermetia illucens</i>	DHill1	311302297
Tephritidae	<i>Ceratitis capitata</i>	DCcap1	498925884
		DCcap2	498925900

Anexo A – Descrição das sequências de tripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continuação)
Família	Espécie	Tripsina ID	GI da Proteína
		DCcap3	498925902
		DCcap4	498925904
		DCcap5	498925906
		DCcap6	498925908
		DCcap7	498926291
		DCcap8	498926307
		DCcap9	498926315
		DCcap10	498926327
		DCcap11	498926357
		DCcap12	498926385
		DCcap13	498926387
		DCcap14	498929186
		DCcap15	498950046
		DCcap16	498950061
		DCcap17	498955830
		DCcap18	499014248
Lepidóptera			
Bombycidae	<i>Bombyx mori</i>	LBmor4	402483747
Crambidae	<i>Chilo suppressalis</i>	LCsup1	388540232
		LCsup3	388540240
		LCsup4	388540236
	<i>Ostrinia nubialis</i>	LOnub1	62082423
		LOnub4	41057924
		LOnub5	209395376
		LOnub6	392514708
		LOnub7	392514700
		LOnub8	392514714
		LOnub9	392514698
		LOnub10	392514696
		LOnub11	392514694
		LOnub14	62082413
		LOnub17	392514706
Noctuidae	<i>Alabama argillacea</i>	LAarg1	*
		LAarg2	*
		LAarg3	*
	<i>Agrotis ipsilon</i>	LAips1	9892203
		LAips2	8347636

Anexo A – Descrição das sequências de tripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continuação)
Família	Espécie	Tripsina ID	GI da Proteína
	<i>Helicoverpa armigera</i>	LHarm1	2463058
		LHarm2	145843824
		LHarm4	2463068
		LHarm5	2463072
		LHarm6	297340776
		LHarm10	2463056
		LHarm11	297340778
		LHarm12	156968305
		LHarm13	171740889
		LHarm14	156968295
	<i>Helicoverpa punctigera</i>	LHpun1	402235790
		LHpun3	402235798
	<i>Helicoverpa zea</i>	LHzea1	8347656
	<i>Heliothis virescens</i>	LHvir1	151200972
		LHvir2	151201030
		LHvir4	7248888
		LHvir5	397910056
		LHvir6	397910058
		LHvir7	397910054
		LHvir8	397910050
		LHvir9	397910038
		LHvir10	397910040
		LHvir11	151201040
		LHvir12	151201026
		LHvir14	397910042
		LHvir15	151201028
		LHvir16	397910044
		<i>Lacanobia oleracea</i>	LLole1
<i>Mamestra configurata</i>		LMcon1	304443643
	LMcon2	237700844	
	LMcon3	304443631	
	LMcon4	315139344	
	LMcon5	315139348	
	LMcon6	237700823	
	LMcon7	315139328	
	LMcon8	237700772	
<i>Spodoptera frugiperda</i>	Sfru1	*	
	Sfru2	*	
	Sfru6	*	
	Sfru7	*	

Anexo A – Descrição das sequências de tripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(conclusão)
Família	Espécie	Tripsina ID	GI da Proteína
		Sfru8	*
		Sfru9	237865270
		Sfru10	*
		Sfru11	*
		Sfru12	*
	<i>Spodoptera litura</i>	Slit1	392494298
		Slit2	156891153
	<i>Sesamia nonagrioides</i>	LSnon1	51094376
		LSnon2	51094388
		LSnon3	51094384
		LSnon5	51094370
		LSnon6	51094386
		LSnon7	51094358
Nymphalidae	<i>Danaus plexippus</i>	LDple5	357630235
		LDple6	357614056
		LDple7	357604385
Plutellidae	<i>Plutella xylostella</i>	LPxyl1	301153729
Pyralidae	<i>Plodia interpunctella</i>	LPint1	6690632
		LPint3	3153853
		LPint4	6690636
	<i>Galleria mellonella</i>	LGmel1	15072548
Saturniidae	<i>Antheraea pernyi</i>	LAper1	295682680
	<i>Antheraea yamamai</i>	LAyam1	398359533
Sphingidae	<i>Manduca sexta</i>	LMsex1	2738867
		LMsex2	146327866
		LMsex3	293224
		LMsex4	464957
		LMsex5	293226
Tortricidae	<i>Choristoneura fumiferana</i>	LCfum1	532085

*Sequências ainda não disponíveis no NCBI

Anexo B – Descrição das sequências de quimotripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

(continua)			
Família	Espécie	Quimotrip. ID	GI da Proteína
Coleóptera			
Bostrichidae	<i>Rhyzopertha dominica</i>	CcRdom1	4868165
Chrysomelidae	<i>Leptinotarsa decemlineata</i>	CcLdec1	372467915
Curculionidae	<i>Phaedon cochleariae</i>	CcPcoc1	74961114
	<i>Anthonomys grandis</i>	CcAgra1	47078376
		CcAgra2	47078372
		CcAgra3	47078370
	<i>Dendroctonus ponderosae</i>	CcDpon1	546683574
		CcDpon2	546679936
		CcDpon3	546677164
		CcDpon4	546674359
		CcDpon5	478256819
		CcDpon6	478253580
		CcDpon7	332376831
		CcDpon8	332376138
		CcDpon9	332375510
		CcDpon10	332375064
		CcDpon11	332374466
		CcDpon12	332374188
		CcDpon13	332374094
Meloidae	<i>Mylabris cichorii</i>	CcMcic1	529274764
Scarabaeidae	<i>Costelytra zealandica</i>	CcCzea1	166919247
		CcCzea2	166919245
		CcCzea3	166919243
		CcCzea4	166919237
		CcCzea5	166919235
		CcCzea6	166919231
		CcCzea7	166919227
Tenebrionidae	<i>Holotrichia oblita</i>	CcHobl1	255046044
	<i>Tribolium castaneum</i>	CcTcas1	91093949
		CcTcas2	91086153
		CcTcas3	91085343
		CcTcas4	91084917
		CcTcas5	91084807

Anexo B – Descrição das sequências de quimotripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continuação)	
Família	Espécie	Quimotrip. ID	GI da Proteína	
		CcTcas6	91084805	
		CcTcas7	91084803	
		CcTcas8	91083173	
		CcTcas9	91079312	
		CcTcas10	282721154	
		CcTcas11	282721150	
		CcTcas12	270015821	
		CcTcas13	270014250	
		CcTcas14	270012790	
		CcTcas15	270009217	
		CcTcas16	270008124	
		CcTcas17	270006456	
		CcTcas18	264681482	
		CcTcas19	264681480	
		CcTcas20	263191198	
		CcTcas21	263190964	
		CcTcas22	262316929	
		CcTcas23	262316925	
		CcTcas24	189241234	
		CcTcas25	189236982	
		CcTcas26	189235467	
		CcTcas27	189235465	
		<i>Tenebrio molitor</i>	CcTmol1	86279303
			CcTmol2	86279301
			CcTmol3	86279287
			CcTmol4	86279283
			CcTmol5	86279277
Diptera				
Cecidomyiidae	<i>Mayetiola destructor</i>	DcMdes1	49472788	
	<i>Sitodiplosis mosellana</i>	DcSmos1	313762885	
Ceratopogonidae	<i>Culicoides sonorensis</i>	DcCson1	56199440	
Drosophilidae	<i>Drosophila melanogaster</i>	DcDmel1	161078358	
		DcDmel2	21406635	
		DcDmel3	24647630	
		DcDmel4	24647632	

Anexo B – Descrição das sequências de quimotripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continuação)
Família	Espécie	Quimotrip. ID	GI da Proteína
Muscidae	<i>Drosophila pseudoobscura</i>	DcDmel5	320542958
		DcDmel6	525313415
		DcDmel7	85857732
		DcDpse1	125773443
		DcDpse2	198131026
		DcDpse3	198131027
		DcDpse4	198131029
		DcDpse5	198131030
		DcDpse6	198131033
	<i>Drosophila virilis</i>	DcDvir1	194142739
		DcDvir2	194142741
		DcDvir3	194142744
		DcDvir4	194142745
		DcDvir5	195394802
	<i>Musca domestica</i>	DcMdom1	557759522
		DcMdom2	557759524
		DcMdom3	557759540
		DcMdom4	557759542
		DcMdom5	557759544
		DcMdom6	557776542
Psychodidae	<i>Lutzomyia longipalpis</i>	DcLlon1	157674411
	<i>Phebotomus papatasi</i>	DcPpap1	157361541
Tephritidae	<i>Ceratitis capitata</i>	DcCcap1	499009610
		DcCcap2	499009849
		DcCcap3	499009853
		DcCcap4	499009859
Chironomidae	<i>Chironomus riparius</i>	DcCrip1	193527454
Lepidóptera			
Bombycidae	<i>Bombyx mandarina</i>	LcBman1	61191881
	<i>Bombyx mori</i>	LcBmor1	112982747
		LcBmor2	114051742
		LcBmor3	114051802
		LcBmor4	512892098
		LcBmor5	512892106
		LcBmor6	512896531
		LcBmor7	512912576

Anexo B – Descrição das sequências de quimotripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(continuação)
Família	Espécie	Quimotrip. ID	GI da Proteína
Crambidae	<i>Diatraea saccharalis</i> <i>Ostrinia nubialis</i>	LcBmor8	512915059
		LcBmor9	512915634
		LcBmor10	512936899
		LcBmor11	512936903
		LcDsac1	411101108
		LcOnub1	209395382
		LcOnub2	392514716
		LcOnub3	392514718
		LcOnub4	392514722
		LcOnub5	392514724
		LcOnub6	392514726
		LcOnub7	392514728
		LcOnub8	392514730
		LcOnub9	392514732
		LcOnub10	392514734
		LcOnub11	392514742
		LcOnub12	62082397
		LcOnub13	62082399
		LcOnub14	62082403
Noctuidae	<i>Alabama argillacea</i>	LcAarg1	*
		LcAarg4	*
		LcAarg6	*
		LcAarg8	*
		LcAarg9	*
		LcAarg11	*
	<i>Agrotis ipsilon</i>	LcAips1	8037814
		LcAips2	8037817
		LcAips3	8037819
	<i>Helicoverpa armigera</i>	LcHarm1	171740891
		LcHarm2	2463064
		LcHarm3	2463076
		LcHarm4	2463092
	<i>Helicoverpa punctigera</i>	LcHpun1	54310832
		LcHpun2	54310834
		LcHpun3	54310838
		LcHpun4	54310840
		LcHpun5	54310844
	<i>Helicoverpa zea</i>	LcHzea1	8037821
		LcHzea2	8037824
		LcHzea3	8037826

Anexo B – Descrição das sequências de quimotripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

Família	Espécie	Quimotrip. ID	(continuação)
			GI da Proteína
	<i>Heliothis virescens</i>	LcHvir1	151199960
		LcHvir2	151199962
		LcHvir3	390627044
		LcHvir4	390627046
		LcHvir5	390627050
		LcHvir6	390627064
		LcHvir7	390627066
		LcHvir8	390627070
		LcHvir9	390627072
	<i>Mamestra configurata</i>	LcMcon1	237700784
		LcMcon2	237700792
		LcMcon3	237700794
		LcMcon4	237700815
		LcMcon5	237700851
		LcMcon6	304443597
		LcMcon7	304443599
		LcMcon8	304443601
		LcMcon9	304443603
		LcMcon10	304443609
		LcMcon11	304443613
		LcMcon12	304443619
		LcMcon13	304443623
		LcMcon14	304443625
		LcMcon15	304443637
		LcMcon16	304443639
		LcMcon17	304443641
		LcMcon18	315139330
		LcMcon19	315139334
		LcMcon20	315139336
		LcMcon21	315139346
		LcMcon22	315139350
		LcMcon23	315139352
		LcMcon24	315139354
	<i>Spodoptera frugiperda</i>	LcSfru2	*
		LcSfru4	*
		LcSfru5	*
		LcSfru7	*
		LcSfru8	*
		LcSfru9	*
		LcSfru11	*
		LcSfru12	*

Anexo B – Descrição das sequências de quimotripsinas utilizadas no trabalho, com o número do GI da sequência proteica no NCBI

			(conclusão)
Família	Espécie	Quimotrip. ID	GI da Proteína
Nymphalidae	<i>Spodoptera litura</i>	LcSfru15	*
		LcSfru17	*
		LcSfru19	*
		LcSfru21	*
		LcSfru22	*
		LcSfru23	*
		LcSlit1	255046231
	<i>Danaus plexippus</i>	LcSlit2	300680014
		LcDple1	357605460
		LcDple2	357610295
		LcDple3	357610887
		LcDple4	357610888
		LcDple5	357610912
		LcDple6	357612227
		LcDple7	357615631
		LcDple8	357616417
		LcDple9	357617001
		LcDple10	357617985
		LcDple11	357619486
		LcDple12	357624259
		LcDple13	357627733
Plutellidae	<i>Plutella xylostella</i>	LcPxyl1	526841187
	<i>Plodia</i>		
Pyrallidae	<i>interpunctella</i>	LcPint1	2353159
Saturniidae	<i>Antheraea pernyi</i>	LcAper1	534468140
Sphingidae	<i>Manduca sexta</i>	LcMsex1	609526

*Sequências ainda não disponíveis no NCBI

Anexo C – Descrição dos parâmetros estimados e os escores de verossimilhança para as tripsinas de Lepidóptera sob diferentes modelos de seleção com razões de ω variáveis entre os sítios

Modelo de sítio	ω	$p(\omega)$	lnL	LRT	p-value
M0 (uma razão)	$\omega=0.16915$	-	-41651.93	3685.38	0.000
M3 (discreto)	$\omega_0=0.02,$ $\omega_1=0.14,$ $\omega_2=0.47$	$p_0=0.26, p_1=0.43,$ $p_2=0.31$	-39809.24		
M1a (proximamente neutro)	$\omega_0=0.13,$ $\omega_1=1.00$	$p_0=0.69, p_1=0.31$	-40599.63	10.04	0.007
M2a (seleção positiva)	$\omega_0=0.13,$ $\omega_1=1.00,$ $\omega_2=1.90$	$p_0=0.69, p_1=0.30,$ $p_2=0.01$	-40594.61		
M7 (beta)		$p=0.53, q=1.65$	-39671.88	-	-
M8 (beta e ω)	$\omega=2.73$	$p_0=1.00,$ $(p_1=0.00),$ $p=0.53, q=1.65$	-39671.88		
Sítios com $d_N/d_S > 1^*$			146 e 151		

*Sítios com probabilidade posterior BEB > 0.95, de acordo com o modelo M2a

Anexo D – Descrição dos parâmetros estimados e os escores de verossimilhança para as quimotripsinas de Lepidóptera sob diferentes modelos de seleção com razões de ω variáveis entre os sítios

Modelo de sítio	Ω	p (ω)	lnL	LRT	p-value
M0 (uma razão)	$\omega=0.174$ 99	-	-56357.68	4025.61	0.000
M3 (discreto)	$\omega_0=0.15$, $\omega_1=0.12$, $\omega_2=0.37$	$p_0=0.20$, $p_1=0.37$, $p_2=0.42$	-54344.87		
M1a (proximamente neutro)	$\omega_0=0.15$, $\omega_1=1.00$	$p_0=0.64$, $p_1=0.36$	-55604.44	8.88	0.012
M2a (seleção positiva)	$\omega_0=0.15$, $\omega_1=1.00$, $\omega_2=1.89$	$p_0=0.64$, $p_1=0.36$, $p_2=0.004$	-55599.99		
M7 (beta)		$p=0.66$, $q=2.41$	-54179.09	-	-
M8 (beta e ω)	$\omega=2.876$ 86	$p_0=1.00$, ($p_1=0.00$), $p=0.66$, $q=2.40$	-54179.09		
Sítios com $d_N/d_S > 1^*$	Nenhum				

Anexo E – Alinhamento do peptídeo sinal e do motivo de ativação do zimogênio de cada grupo de tripsina dos insetos estudados.

(continua)

			Trypsin →
C-I	<i>CTdic1</i>	MHLLRIECIGFI-----LT-VYR ICRVQSQGDGR I IGG	
	<i>CTcas3</i>	MVVPAI-----AKPRTRI IGG	
	<i>CDabb1</i>	MPSR-----DFENTN--GTVFYHPTIGGR I VGG	
	<i>CDpon4</i>	MKVL-IAFLGLVSLGLAAPEGHNPKLLHPQIGGR I VGG	
C-II	<i>CPcoc1</i>	MIRFT--LALAVIGVTFA-----ASTPQI-E-----TNPNI IGG	
	<i>CDpon1</i>	MTRMLLVVFFAIVAVCSQQYVLIDKREDLLEPEPLVPLPDGR I VGG	
	<i>CDpon2</i>	MLWLV--I-ISVL--SSALS---EYANVEESGSLVRHVESR I VGG	
	<i>CDpon3</i>	MLKLV--FWLGLVARSLCLT---VAADEIFEKPLVPLLDGR I VGG	
C-III	<i>CTcas12</i>	MNPLWL VVFLTV-----SVSAEYENSSLPEDR I VGG	
	<i>CDpon5</i>	MTVIYLSGN-----PISNSSLRRI SQNSFVR I VGG	
	<i>CTcas4</i>	MVVFLFGTFWEQRLGITSWGDCGAANNIHGDLKMP TIDVR I IGG	
	<i>CTcas6</i>	MHVFFLFLTSL-----LVTNAHVQVFKPQIDVR I VGG	
C-IV	<i>CPruf1</i>	MFKIVL I AALAALAVDARF-LSRAPQLDGR I VGG	
	<i>CCzea2</i>	MLKFIVLSLFVAYAVGAPQKRAALPRLDGR I VGG	
	<i>CCzea3</i>	MLKFVVISLLVACAFGAPQKRAPVPRLDGR I VGG	
C-V	<i>CTcas9</i>	MKFVVLFAVVLATASAVPAFLRK NKFFPDGR I VGG	
	<i>CTcas10</i>	MKSFLLI VFLVATASAVPSYLR TNSLLPDGR I IGG	
	<i>CTmol2</i>	MKSVL FVFLVASASAVPPFLRKNSLMPDGR I VGG	
D-I	<i>DDmel1</i>	MLLHWLVLV A-SVT-----LISAGSSSPER I VGG	
	<i>DDmel2</i>	MFVQWIFLAF-SVT-----VVSSNWI PER I VGG	
	<i>DDvir4</i>	MANTSSIPNL-ALLVLCCAGGLAKPANQDYGR I VGG	
	<i>DDmel3</i>	MAYVGIVATLVLLLLL---G-DASDEATGR I IGG	
	<i>DDpse1</i>	MTATGIVLL-LVLLIL---G-VTTANDTKGR I VGG	
D-II	<i>DDmel8</i>	MSL-LIATFLALLALTNGAVIPIGLEPQTSSLGGR I VGG	
	<i>DDpse6</i>	MFL-PIATLLALLASSYGAALPSGLPEQAP-LSGR I VGG	
	<i>DDvir8</i>	MYCQTLMVLLGLLAAASGAIMPLGLSPQP---DGR I VGG	
D-III	<i>DLcup1</i>	MLKFVVLCAI SCALGAAPVEGMVPQLDGR I VGG	
	<i>DDpse3</i>	MLKFVILL SAVACALGGTIPEGLLPQLDGR I VGG	
	<i>DDpse4</i>	MLKFVILL SAVACALGGTIPEGLLPQLDGR I VGG	
	<i>DDvir6</i>	MFKFGIILL SVVACAFGAALPEGLLPQLDGR I VGG	
	<i>DDvir7</i>	MFKFVILL SVVACAFGAALPEGLLPQLDGR I VGG	
	<i>DDpse5</i>	MLKFVILL SAVACALGGTIPEGLLPQLDGR I VGG	
	<i>DDvir5</i>	MFKFVILL SVVACAFGAAPVEGLLPQLDGR I VGG	
	<i>DDmel7</i>	MLKIVILL SAVVICALGGTVPEGLLPQLDGR I VGG	
	<i>DDmel5</i>	MLKFLILL SAVACALGGTIPEGLLPQLDGR I VGG	
D-IV	<i>DDmel6</i>	MLKFVILL SAVACALGGTVPEGLLPQLDGR I VGG	
	<i>DHdia1</i>	MLKFVILVCTVAYVFSAVVPGGMLPQLDGR I VGG	
	<i>DHdia2</i>	MLKFVILLCTVAYVFSAVVPGGMLPQLDGR I VGG	
	<i>DHlin3</i>	MLKFVILLCSIAYVFGAVVPLGMLSQSDGR I VGG	
	<i>DHlin1</i>	MLKFVILVCSVACVFGAVVPGGMLPQLDGR I VGG	
	<i>DHlin2</i>	MLKFVILVCSVACVFGAVVPGGMLPQLDGR I VGG	
D-V	<i>DGmor2</i>	MFRYFLVFVA--LSTVVAGLL-----HNDHTAGR I VNG	
	<i>DCcap3</i>	MFRFWVLSAL--LNALAA-AT-----VPDIFAGGR I VGG	
	<i>DCcap4</i>	MS-----TDL DGR I VGG	
	<i>DDvir9</i>	MQGLEALLLC LVGTTVAGTISP A-PNPFER EGR I VGG	
	<i>DDmel10</i>	MHRLVLLVCLAVGSACAGTVGVSN GDFPEREGR I VGG	
	<i>DDpse7</i>	MHGLVALLACLAVGS AWAGTVSVA-RSPFAREGR I VGG	
D-VI	<i>DScal1</i>	MLRFVILFALVSTSLA----GVSRNDF-YGR I VNG	
	<i>DNbul1</i>	MLRFIAVFALVNCALA----GTL PNDL-DGR I VNG	
	<i>DCcap5</i>	MLGLFTLFALASAITA----GTVPTSLFVGR I VKG	
	<i>DCcap6</i>	MFRFVALFAVSLAYG----ATLPDDL-DGR I VNG	
	<i>DDvir10</i>	MNRLIVSFVVVLACIAA-CQGLPDTEFPFGR I VNG	
	<i>DDmel9</i>	MNRLLLSVVALVALAAS-CHGNPGLDFPFGR I VNG	
	<i>DDpse8</i>	MNRLLLCVVAVIALGAGCCQARPDIDFPFGR I VNG	

[illegible]

Anexo F – Alinhamento do peptídeo sinal e do motivo de ativação do zimogênio de cada grupo de quimotripsina dos insetos estudados.

(continua)

[illegible]

Anexo F – Alinhamento do peptídeo sinal e do motivo de ativação do zimogênio de cada grupo de quimotripsina dos insetos estudados.

(conclusão)

LcBman1		MKVFA-AVLMALA	AVVVAAEEA	ELDYHTKI	GIPRAESL	KRAEEAAD	FDGTR	IVGGS
LcBmor1		MKVFA-AVLMALA	AVVVAAEEA	ELDYHTKI	GIPRAESL	KRAEEAAD	FDGTR	IVGGS
LcBmor7		MRGFNV	LALLALATAAHAFV	IEINHYHVI	GIPMAAMEL	KSAAEEAD	FDGAR	IVGGS
LcOnub14		MKSA-VLFL	LVVAAAAAERLO	PNTRYHETEG	IPKMEIQ	RLLEEGT	DFDGR	IVGGQ
LcOnub8		MKSS-LLFL	LVVAAAAAELLO	PNTRYHETEG	IPKFMQL	QLLEEGT	DFDGS	IVGGQ
LcOnub9		MOSS-LLFL	LVVAAAAAELLO	PNTRYHETEG	IPRLELOR	RLLEEGS	DFDGR	IVGGQ
LcDple1		MN-FLLV	TVALSLAVSAEES	IRNDYHETV	GIPMAARI	REAAAL	DFDGT	IVSGGQ
LcDple10		MK-LLLV	VVCLAVAVSAYEP	INIDYHNTIGV	YEARIKQ	AEAAAD	FDGSR	IVGGN
LcDple12		MK-LLMV	IVCLSVAVTAYEP	INIDYHNTIGV	YEARIKQ	AEES	TDGSR	IVGGN
LcDple13		MHYFL	LVVVGLAVAVSAYEP	INIDYHNTIGV	YEARIKQ	AEES	TDGSR	IVGGN
LcDple2		MK-LLLV	VVGLAVAVSAYEP	INIDYHNTIGV	YEARIKQ	AEAAAD	FDGSR	IVGGN
LcDple5		MK-LLLV	VVGLAVAVSAYEP	INIDYHNTIGV	YEARIKQ	AEAAAD	FDGSR	IVGGN
LcDple6		MK-LLLV	VVCLAVAVSAYEP	INIDYHNTIGV	YEARIKQ	AEES	TDGSR	IVGGN
LcDple8		MK-LLLV	VVGLAVAVSAYEP	INIDYHNTIGV	YEARIKQ	AEES	TDGSR	IVGGN
LcBmor4		MKVLLLI-FL	GAVSLASAGLLP	AMLPYHDTV	GIPLAESL	KKSEEAR	DFDGS	IVNGN
LcBmor5		MKVLLLI-LL	GAVSLASAGLLP	AMLPYHDTV	GIPLAESL	KKSEEAR	DFDGS	IVNGN
LcMcon13		MKYLLML	CALAALALAE--EP	IMLHYHENV	GIPASRIK	QALAT	DFDGS	IVGGS
LcOnub7		MKWNVLV-	FLAAVAVEAGP	FKCSQWSYHES	SVGIPRSEQL	KKLELAA	--NST	IVGGT
LcAarg6		---	---	---	---	---	---	---
LcHvir6		MRTLLI	ICLFFVVS	SWAHETT	GDE---	---	---	---
LcMcon6		---	---	---	---	---	---	---
LcSfru19		MKVLAVAL	LALVAVSSAR-HI	GLDIVDLES	LAVDYFNKVG	GIPLAEEI	RAAIEA	AAANPS
LcBmor3		MKTAVL	FGIFFVGYTAYV	VPV--ETFYH	NGVIGIPLAKS	IRAAETAK	LDSSVQ--	PONAA
LcDple9		MHGVL	FLCLSLVAAT	SAYLRV--PK	DYHLVEV	GIPATRIK	EAEDK	FLSS--Q--
LcOnub5		MKLVF	-VVVALVSLG	SVFVFGPGY	HEEVGIP	PLAKELK	AAERNV	ASIA
LcOnub6		MKLILV	IVVALVASA	AFVEVL	GPRGYHEE	IGIPLATAL	KEAEQM	IASMAR--
LcMcon14		---	---	---	---	---	---	---
LcMcon16		MRALLI	GVVLVASASA--	LGAGYHDEV	GIPLAERL	KKSEEAR	DFDGS	IVGGS
LcMcon17		MRALLLV	GLALVASAAFAEV	GAGYHQEV	GIPLYEKL	RAAEK	LLAE	TAQ--
LcSfru23		---	---	---	---	---	---	---
LcAarg9		MRALLI	IGLALVASA	AFVKVTS	DYHADV	GIPLAARI	KADEDA	VF
LcHvir5		MRVLLF	LGLGLLASA	AFVNVNFD	YHNGVIGI	PAAAKIK	EIEDK	LLAEAP--
LcMcon10		MRALVL	FGLSLVVS	SASALV	VQTNHYDS	VGIPAAER	IRVLEEE	ELAKNPQV
LcMcon11		---	---	---	---	---	---	---
LcMcon23		---	---	---	---	---	---	---
LcMcon5		---	---	---	---	---	---	---
LcAarg1		MCRTEI	ICLFI	LLQYCYG	GLQDRPMY	GYHMKIG	VPKHY	VKLQTTY--
LcBmor10		MLRSFI	LINILS	FTFA--	---	---	---	---
LcBmor11		MRYSVL	FLTC	AVIFANG	YAFNGPAY	GYHLRF	GPAYEL	LKRL--
LcDple11		MRVLLF	-VLO-LC-	GC--	---	---	---	---
LcOnub4		MAVR	FLLYFF	LVMTTHG	---	---	---	---
LcPint1		MASKI	LLCILF	-GVQSE-	---	---	---	---
LcAips1		MKFLG	VCCLAVAA	SAVE--	---	---	---	---
LcHarm2		MKFLG	VCCLAVAA	SAVE--	---	---	---	---
LcHpun1		---	---	---	---	---	---	---
LcHpun2		MKFLG	VCCLAVAA	SAVE--	---	---	---	---
LcHvir4		MKFLG	VCCLAVAA	SAVE--	---	---	---	---
LcMcon12		MKFLG	VCCLAVAA	SAVE--	---	---	---	---
LcSfru2		MKFLAV	-VCLAVAA	SAVE--	---	---	---	---
LcHarm3		---	---	---	---	---	---	---
LcHpun4		---	---	---	---	---	---	---
LcHvir9		MKLLAV	TLLAFAAV	VSARNID	HEAID	LEDITAY	GYHTK	IGFSLAEK
LcSfru17		MKVLAV	TLLAFAAV	VSARNID	HEAID	LEDITAY	GYHTK	IGFSLAEK
LcSfru4		MK-LAV	ALLALVAVS	AAKNID	EDVIDLES	NVFGYHEK	IGAPLAEK	KIAAEAAAL
LcHzea3		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcMcon19		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcMcon20		---	---	---	---	---	---	---
LcMcon21		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcSfru2		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHarm4		MKLLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHpun3		MKLLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHvir1		MKLLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHzea1		MKLLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHzea2		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcMcon3		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcAips2		MKFLAL	ALLALT	VFAAHV	SFEDA	IDLEDITAY	GYLSK	IGVAAER
LcMcon2		MKFLAL	ALLALT	VFAAHV	SFEDA	IDLEDITAY	GYLSK	IGVAAER
LcMcon22		---	---	---	---	---	---	---
LcMcon4		MKFLAL	ALLALT	VFAAHV	SFEDA	IDLEDITAY	GYLSK	IGVAAER
LcSfru12		MVVE	PLTLLA	AAVATANN	IVEDA	IDLEDITAY	GYLAK	IGKPLADE
LcSfru5		MKFLAL	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcSfru1		MKFLAL	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHarm1		MKFEAL	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHpun5		---	---	---	---	---	---	---
LcHvir3		MKFEAL	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcMcon15		---	---	---	---	---	---	---
LcMcon9		MKFLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcSfru21		MKLLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcSfru8		---	---	---	---	---	---	---
LcSfru9		MKLLAV	TLLAFAAV	VSARNID	LEDVIDLEDIT	AYDYHRA	IGGPLADK	IRLAEEAD
LcHvir7		MKVIV	VLALAS	LAYANV	LPSEGN	TAYGY	IQNYAI	PLAEKIR
LcMcon18		---	---	---	---	---	---	---
LcMcon8		MKVIV	VLALAS	LAYANV	LPSEGN	TAYGY	IQNYAI	PLAEKIR
LcSfru22		MKVIV	VLALAS	LAYANV	LPSEGN	TAYGY	IQNYAI	PLAEKIR
LcSfru7		MKVIV	VLALAS	LAYANV	LPSEGN	TAYGY	IQNYAI	PLAEKIR
LcOnub12		MKFLV	VLAVAS	LAHGK	VV-PD	NHTAF	GYLKN	SIVEAEK
LcOnub13		MKFLV	VLAVAS	LAHGK	VV-PD	NHTAF	GYLKN	SIVEAEK
LcOnub2		MKLLV	VVLAVAA	-AYGE	SPAST	SAYGYL	ANSVAH	AEKLR
LcBmor8		MKFI	I--FL	GLNLV	CGINL	SETFRN	SFGYH	ARVGI
LcHvir2		MKTFI	VACLALV	SFACAE	ESL-F	PGSTY	GYLVE	YGI
LcMcon7		MKTFI	VACLALV	SFACAE	ESL-F	PGSTY	GYLVE	YGI
LcSfru11		MKAFI	VVCLALV	SFACAE	ELR-Y	PSTAY	GYLVK	YGI