

Article

The Evolution, Gene Expression Profile, and Secretion of Digestive Peptidases in Lepidoptera Species

Lucas R. Lima ^{1,†}, Renata O. Dias ^{2,3,†}, Felipe Jun Fuzita ², Clélia Ferreira ², Walter R. Terra ² and Marcio C. Silva-Filho ^{1,4,*}

¹ Programa de Pós-Graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, Mato Grosso do Sul 79117-010, Brazil; lukas6923@gmail.com

² Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, São Paulo 05508-000, Brazil; diasrdo@gmail.com (R.O.D.); felipe.fuzita@gmail.com (F.J.F.); clfterra@iq.usp.br (C.F.); warterra@iq.usp.br (W.R.T.)

³ Departamento de Genética, Universidade Federal de Goiás, Goiânia, Goiás 74690-900, Brazil

⁴ Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo 13418-900, Brazil

* Correspondence: mdcsilva@usp.br; Tel.: +55-19-3429-4442

† These authors contributed equally to this work.

Received: 6 January 2020; Accepted: 6 February 2020; Published: 11 February 2020

Abstract: Serine peptidases (SPs) are responsible for most primary protein digestion in Lepidoptera species. An expansion of the number of genes encoding trypsin and chymotrypsin enzymes and the ability to upregulate the expression of some of these genes in response to peptidase inhibitor (PI) ingestion have been associated with the adaptation of Noctuidae moths to herbivory. To investigate whether these gene family expansion events are common to other Lepidoptera groups, we searched for all genes encoding putative trypsin and chymotrypsin enzymes in 23 publicly available genomes from this taxon. Phylogenetic analysis showed that several gene family expansion events may have occurred in the taxon's evolutionary history and that these events gave rise to a very diverse group of enzymes, including proteins lacking the canonical SP catalytic triad. The expression profile of these enzymes along the midgut and the secretion mechanisms by which these enzymes enter the luminal content were also analyzed in *Spodoptera frugiperda* larvae using RNA-seq and proteomics. These results support the proposal of a midgut countercurrent flux responsible for the direction of these proteins to the anterior portion of the midgut and show that these enzymes reach the midgut lumen via both exocytosis and microapocrine secretion mechanisms.

Keywords: Lepidoptera; trypsin; chymotrypsin; peptidase; peritrophic membrane

1. Introduction

In insects, expansion events in gene families encoding digestive enzymes including cathepsins L [1,2] and D [3], lysozymes [4], and trypsins [5] have been associated with the adaptation of diverse taxonomic groups to their food sources. The role of these duplication events in the adaptation of these insects is hypothesized to be related to dosage amplification selection and the diversification of the expression and biochemical properties of the expanded genes. In herbivorous Lepidoptera species, the expansion of the number of genes encoding serine peptidases (SPs) has been mainly associated with adaptation to overcome the effects of host serine peptidase inhibitors (PIs) [6].

Lepidopteran larvae rely on serine peptidases (mainly trypsin and chymotrypsin enzymes) for primary protein digestion [7]. These enzymes are well suited to the alkaline midgut of these insects but also make

them susceptible to dietary serine peptidase inhibitor (PI) ingestion. PIs improve plant defenses against insects by acting as pseudosubstrates and thereby limiting the effects of SPs on proteolysis [8]. Polyphagous *Spodoptera frugiperda* (Lepidoptera, Noctuidae) larvae exhibit a large set of genes encoding trypsin and chymotrypsin enzymes in their genomes (14 chymotrypsin and nine trypsin genes were identified in a previous assembly of the species transcriptome), which are differentially transcribed when the larvae are exposed to PIs [9,10]. In contrast, the less-generalist *Diatraea saccharalis* (Lepidoptera, Crambidae) larvae have a smaller set of genes encoding these enzymes (nine chymotrypsin and four trypsin genes were identified in the species transcriptome), and none of these genes are upregulated in response to soybean PI ingestion [6]. *Spodoptera frugiperda* larvae are able to physiologically adapt to soybean peptidase inhibitor ingestion [11], whereas *Diatraea saccharalis* larvae are significantly affected by these inhibitors [12]. These results show that Lepidoptera species rely on different strategies to overcome the deleterious effects of PIs, but in at least some of these species, SP gene expansion events must play an important role in adaptation to their host plants.

Approximately one-third of peptidases are serine peptidases. SP enzymes are typically composed of a catalytic Ser/His/Asp triad and function via a mechanism in which serine acts as a nucleophile, histidine acts as a general acid base and aspartate aids in histidine positioning and charge stabilization. However, enzymes exhibiting SP folds with noncanonical triads (such as Ser/His/Glu, Ser/His/His, Ser/Glu/Asp, or Gly/His/Asp) and dyads (Ser/Lys and Ser/His) have also been described [13]. These innovations have been shown to be useful in some cases; for example, the Ser/Glu/Asp triad is believed to allow the peptidase to function in low-pH environments [13]. In Arthropoda, serine peptidase homologs have been found to play important roles such as acting as antimicrobial molecules [14], collaborating in the activation of prophenoloxidasases [15], and mediating cell adhesion [16]. Nonenzyme proteins can regulate their functional enzyme homologs, as observed for PLA2 inhibitors stabilizing the inactive form of phospholipase A2 [17] and SOD1 copper chaperones necessary for the full activity of superoxide dismutase [18]. For these pairs of proteins, the nonenzyme protein is hypothesized to have evolved from its enzyme counterpart [19]. However, the evolutionary history and role of most SP protein homologs in insects remain poorly understood.

Serine peptidases are synthesized as inactive zymogen precursors that must undergo an activation process. In this process, an N-terminal pro-peptide region of variable length is removed by a second peptidase enzyme that cleaves the precursor between a basic amino acid (lysine or arginine) and an isoleucine residue. In Lepidoptera, the trypsin and chymotrypsin activation process is believed to be performed by trypsins, as the amino acid composition of the activation region of both enzymes includes a high-affinity to trypsin binding site [5].

In the Lepidoptera species *S. frugiperda*, digestive trypsins are secreted in the anterior midgut region and bound to the membrane via a microapocrine mechanism in which small vesicles bud from microvilli [20]. Another pathway for anchoring SPs to the membrane was identified on the basis of the presence of detergent-resistant domains in midgut microvilli, where these domains likely act as a mechanism to recruit proteins for microapocrine secretion [21]. In the posterior region, these enzymes are soluble and secreted via exocytosis [20].

Trypsin and chymotrypsin enzymes present decreasing activity along the lepidopteran larval midgut, an effect that can be eliminated by disrupting the peritrophic membrane in calcofluor-treated larvae [22–24]. Thus, it seems that the compartmentalization achieved by the peritrophic membrane, together with water counterfluxes in the ectoperitrophic space (outside PM), is responsible for recycling digestive enzymes in the midgut of this animal.

To better understand how the evolutionary history of trypsins and chymotrypsins is correlated with Lepidoptera species diversification, we searched for genes encoding these proteins in all available Lepidoptera genomes. Phylogenetic analysis revealed that several expansion events may have occurred in these genes during the evolutionary history of the order and also highlighted the emergence of several noncanonical peptidases over time. Moreover, we analyzed the expression profile of these genes along the midgut and the secretion mechanisms by which they reach the luminal space in *Spodoptera frugiperda* larvae via RNA-seq and proteomic analyses of different midgut regions.

2. Results

2.1. Lepidopteran Trypsin and Chymotrypsin Gene Evolution

To understand how the genes encoding trypsins and chymotrypsins evolved during Lepidoptera species diversification and in which taxonomic groups gene family expansion events may have occurred, we searched for all protein sequences related to these enzymes in 23 Lepidoptera genomes with public gene annotation datasets. These sequences were then subjected to phylogenetic and substrate-binding site analyses (see the analyzed sequence positions in Figure 1A).

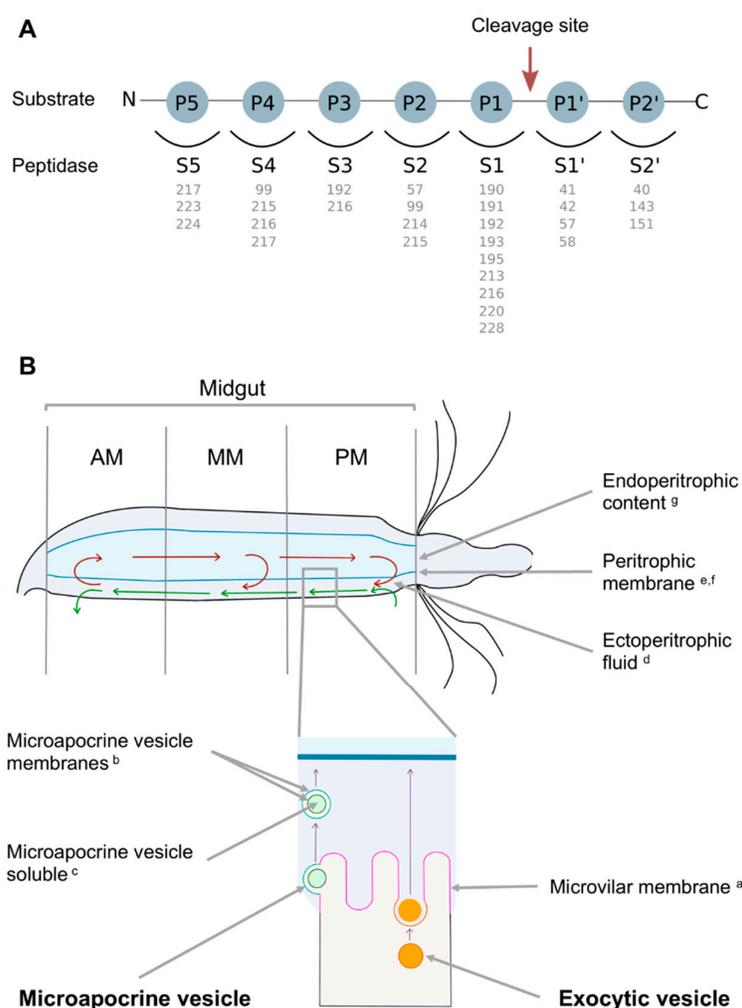


Figure 1. The peptidase subsites and *Spodoptera frugiperda* midgut regions explored in this work. **(A)** Substrate positions (P5, P2') and corresponding peptidase subsites (S5, S2') in the substrate-binding site. The residue positions belonging to each subsite are listed in gray below the subsite. **(B)** Schematic view of the *Spodoptera frugiperda* gut and the midgut regions analyzed by RNA-seq (anterior (AM), middle (MM), and posterior (PM) midgut) and proteomic analysis: ^a microvillar membrane, ^b microapocrine vesicles, ^c microapocrine vesicle soluble, ^d ectoperitrophic fluid, ^{e,f} peritrophic membrane wash and core, and ^g endoperitrophic content. Red arrows, enzyme flux; green arrows, water flux.

Branches with good support predicted for each phylogenetic tree (designated T1 to T18 for trypsins and C1 to C23 for chymotrypsins) were individually analyzed to facilitate the identification of gene family

expansion events and substitutions in the amino acid composition of the enzyme substrate-binding sites (Figures S1 and S2).

For trypsins, the largest numbers of gene copies are observed in the genomes of species from the Noctuidae (28 to 42), Plutellidae (37), and Sphingidae (35) moth families (Tables S1,S2). The number of genes encoding chymotrypsins is highest in the genomes of species from the moth families Plutellidae (71) and Noctuidae (32 to 58) and the butterfly family Pieridae (61) (Tables S3,S4). However, the number of genes encoding both trypsins and chymotrypsins is variable among Lepidoptera genomes, even between species within the same family, as observed for *Ostrinia furnacalis* and *Chilo suppressalis* from the Crambidae family.

The first six branches of both trees (T1–T6 and C1–C6) are included in a major branch together with homologous sequences from Diptera and/or *Homo sapiens* that were retrieved from the UniProt database (Figures S1 and S2). Several gene family expansion events may be speculated on other branches, such as the trypsin T7, T10, T15, T17, and T18 branches. These branches include multiple gene copies likely originated from several duplication events that probably occurred at different times during the evolutionary history of Lepidoptera species.

The trypsin and chymotrypsin similarity groups are highly variable in their substrate-binding sites (Figures S3 and S4). The only highly conserved sites are those related to the catalytic triad, disulfide bond-forming residues, and amino acids located at the S1 subsite. This high variability is also observed among the enzyme activation sites and pro-peptide regions (data not shown). Notably, the mammalian DDDD motif in the trypsinogen activation site, which shows high affinity for enterokinase enzyme activation, is not observed in any of the trypsin groups.

2.2. Trypsin and Chymotrypsin Genes: Expression and Secretion Pathways in the *Spodoptera frugiperda* Midgut

Protein sequences classified as trypsin and chymotrypsin sequences were used to search for homologous proteins in the *Spodoptera frugiperda* transcriptome. *S. frugiperda* transcripts were analyzed according to their expression profile in three midgut regions (anterior (AM), middle (MM), and posterior (PM)) and the carcass (CAR, larvae without the midgut) (Figure 1B).

In agreement with our classification, all *S. frugiperda* genes exhibited higher expression (transcripts per million (TPM)) values in at least one midgut tissue than in the carcass (Tables 1,2). Notably, most trypsin and chymotrypsin genes were most highly expressed in the middle midgut (MM). Moreover, with only one exception, the proteomics data showed that all sequences chosen for phylogenetic analyses were present in the midgut and not in the other analyzed tissues. Only group T4 seemed to be composed of nondigestive trypsins. Sf_186.1 was highly abundant in hemolymph cells in contrast to the midgut fractions, and the low levels observed only in microapocrine vesicles could be due to hemolymph contamination.

Protein sequencing was effective, and approximately 50% of the sampled sequences included in Tables 1,2 were identified by this technique (14 out of 26 from Table 1 and 19 out of 36 from Table 2). In addition, most of the sequences identified by proteomics were sequences with higher expression values. Using proteomic data, we determined the secretion mechanism that is likely used by each protein. The enzymes that were observed in microapocrine vesicles and reached the PM and/or endoperitrophic contents were considered to be secreted via a microapocrine mechanism. By exclusion, the protein was considered to present an exocytic mechanism if it was not present in these vesicles but reached the PM and/or its contents. Finally, we did not consider a protein to be present in an evaluated sample when its quantitative value was less than 10% across all fractions because minor cross-contamination is intrinsic to sample handling for obtaining the different midgut fractions.

Table 1. *Spodoptera frugiperda* trypsin gene expression and proteomic analysis results.

Branch	Protein	Expression (TPM)					Proteomics (NSAF) ‡									Secretory Route	GPI Anchor
		AM	MM	PM	CAR	EMM	mvMP	mvSP	EF	PMW	PMS	ECS	HC	HS	FB		
T4	Sf_186.1	14.8	11.8	11.1	10.0	-	194.7*	342.1*	-	-	-	-	6708.2	-	55.5*	HC	
T6	Sf_11739.1	16.4	7.9	24.8	0.1	-	-	-	-	-	-	-	-	-	-		
T7	Sf_10598.1 ^a	2555.6	200.0	4192.4	0.9	-	-	174.1*	-	9972.9	2990.9	7503.5	-	-	-	EC	
	Sf_20521.1	1953.7	183.8	2575.1	0.8	-	-	185.4*	302.3*	1264.3	936.8	2446.0	-	-	-	EC	
	Sf_281.1	159.5	687.6	565.2	0.5	-	1276.2*	186.8*	-	7644.0	7740.5	5587.0	-	-	-	EC	
T8	Sf_2549.1	436.3	353.5	481.1	0.4	66.5	366.3	-	-	-	-	282.9	-	-	-	MA	E, P
T11	Sf_2412.1	930.9	5906.7	991.3	2.0	-	-	-	-	-	-	-	-	-	-		
	Sf_1747.1	564.6	2790.7	523.5	1.1	-	-	-	-	-	-	-	-	-	-		
	Sf_12619.1	0.3	3.6	0.5	0.8	-	-	-	-	-	-	-	-	-	-		
	Sf_1575.1	46.2	16.8	56.2	-	-	-	-	-	-	-	-	-	-	-		
T12	Sf_918.1	4.3	0.1	0.1	-	-	-	-	-	-	-	-	-	-	-		
T18	Sf_21802.1	2539.4	9706.0	3475.6	4.5	S	S	S	-	2951.7	3159.0	7191.2	-	-	-	MA	
	Sf_225.1	3707.9	14586.9	5547.5	6.6	316.1*	11,975.0	2868.8	1714.7	27,125.0	24,543.0	24,894.0	-	-	-	MA	P
	Sf_1694.1	1841.1	4761.2	2011.9	2.1	-	S	S	-	S	S	S	-	-	-	MA	
	Sf_22890.1	1318.4	5789.0	2837.4	1.9	-	-	-	-	-	-	-	-	-	-		
	Sf_3096.1	270.7	1583.7	658.7	1.5	-	-	-	-	7763.5	5752.3	2169.6	-	-	-	EC	
	Sf_1720.1	422.9	1535.2	619.1	1.0	-	-	-	-	11,126.0	11,428.0	2935.2	-	-	-	EC	
	Sf_18450.1	133.1	238.9	69.3	0.2	-	-	-	-	-	966.3	504.6	-	-	-	EC	
	Sf_5016.1 ^b	3.1	6.9	3.9	-	-	-	-	-	-	-	-	-	-	-		
	Sf_17832.1	214.7	235.7	144.3	0.1	-	-	-	-	-	899.0	313.0	-	-	-	EC	
	Sf_6939.1	59.6	146.1	152.2	0.3	-	-	-	-	-	-	-	-	-	-		
	Sf_18981.1	34.1	275.1	450.4	0.4	-	-	-	-	-	-	-	-	-	-		
	Sf_11764.1	156.4	206.2	617.2	0.2	-	-	-	-	4676.3	5197.3	3183.4	-	-	-	EC	
	Sf_3144.1	29.7	15.6	14.5	-	-	213.5*	-	-	1278.9	1516.2	824.8	-	-	-	EC	
Sf_24638.1	3.0	1.0	2.2	-	-	-	-	-	-	-	-	-	-	-			
Sf_26131.1	2.6	1.0	1.4	-	-	-	-	-	-	-	-	-	-	-			
All groups		17,419.6	49,250.8	26,026.8	35.1												

AM, MM, and PM: anterior, middle, and posterior portions of the midgut, respectively; CAR: carcass (body without the midgut). ‡ Normalized spectral abundance factor varying from 0 to 1,000,000. * The value represents quantitatively less than 10% of the observed protein in relation to its identification across all analyzed samples. ECS, endoperitrophic content supernatant; EMM, enterocyte microvillar membrane; EF, ectoperitrophic fluid; FB, fat body; HS, hemolymph soluble proteins; mvMP, microapocrine vesicle membrane proteins; mvSP, microapocrine vesicle soluble proteins; PMW, peritrophic membrane washing; PMS, peritrophic membrane supernatant. EC, exocytosis; HC, hemolymph cells; MA, microapocrine secretion; DRM, protein found in detergent-resistant domains present in the microvillar membrane [21]; E, empirical observation of a GPI anchor [21]; P, positive for GPI anchor prediction using PredGPI [25]. Protein sequences with substitutions in the catalytic triad: ^a S195F and ^b S195A (with best BLAST hits in the tree: Sfrr_SFRICE004673-PA and Slit_XP_022821660.1, respectively, both with the same amino acid replacements in the catalytic triad as their *S. frugiperda* homologs).

Table 2. *Spodoptera frugiperda* chymotrypsin gene expression and proteomic analysis results.

Branch	Protein ID	Expression (TPM)			Proteomics (NSAF) †										Secretory Route	
		AM	MM	PM	CAR	EMM	mvMP	mvSP	EF	PMW	PMS	ECS	HC	HS		FB
C1	Sf_21444.1	63.7	47.4	431.2	0.7	-	-	-	-	-	-	-	-	-	-	-
	Sf_18384.1	2298.8	3666.5	1210.2	2.2	143.9 *	2180.4 *	870.6 *	-	20,183.0	16,010.0	26,339.0	-	-	-	EC
	Sf_290.1	706.9	595.1	523.6	0.3	9544.8	2513.8	679.4 *	-	1158.2 *	7894.9	-	-	-	-	MA/DRM
C2	Sf_26117.1	12.0	15.0	12.9	-	-	-	-	-	-	-	-	-	-	-	-
	Sf_7089.1	2.3	1.6	0.6	-	-	-	-	-	-	-	-	-	-	-	-
	Sf_18329.1 ^a	1.2	0.7	0.2	-	-	-	-	-	-	-	-	-	-	-	-
C4	Sf_8178.1	1.1	0.9	1.5	0.1	-	-	-	-	-	-	-	-	-	-	-
C5	Sf_6061.1	143.3	4.2	16.1	0.1	-	-	-	-	-	6693.5	-	-	-	-	EC
	Sf_22760.1	1.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C8	Sf_7342.1 ^b	392.5	430.7	344.1	0.3	6141.5	4539.9	8519.3	-	4696.7	9890.4	3028.9 *	-	-	-	MA/DRM
	Sf_3834.1	6.8	8.7	6.7	-	-	-	-	-	-	-	-	-	-	-	-
C12	Sf_3001.1	241.5	271.2	55.9	0.3	-	-	-	-	-	-	-	-	-	-	-
C14	Sf_15188.1	25.9	15.8	14.0	1.6	-	-	-	-	-	-	-	-	-	-	-
	Sf_2499.1	3116.3	14,163.1	4658.9	8.6	-	1506.9 *	1058.9 *	-	17,510.0	10,031.0	13,155.0	-	-	-	EC
C17	Sf_4253.1 ^c	98.2	392.7	148.2	0.2	-	-	-	-	-	-	-	-	-	-	-
	Sf_8620.1	32.5	31.1	63.3	-	-	-	-	-	-	-	-	-	-	-	-
C18	Sf_5271.1	3.9	191.5	138.8	0.2	-	-	-	-	-	-	-	-	-	-	-
	Sf_5915.1	227.7	788.4	124.7	0.5	-	-	-	-	5.5 *	11.6 *	255.0	-	-	-	EC
C19	Sf_1602.1	6.0	92.6	5.0	0.1	-	-	-	-	-	-	-	-	-	-	-
	Sf_24363.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C20	Sf_12403.1	1053.4	5584.5	1209.6	1.5	-	-	902.4 *	-	820.4 *	2127.6	4497.3	-	-	-	EC
	Sf_1643.1	2172.9	3428.5	1026.9	1.3	-	-	-	-	3598.2	6429.8	5596.4	-	-	-	EC
	Sf_1786.1	586.3	1145.4	286.0	0.3	-	-	-	-	1107.8	4104.2	3000.7	-	-	-	EC
	Sf_10797.1	3.4	8.8	1.7	-	-	-	-	-	-	-	-	-	-	-	-
C21	Sf_1782.1	8835.3	39,557.9	10,240.5	19.6	453.2 *	17840.0	6894.9 *	9068.5 *	50,007.0	31,352.0	58,987.0	-	-	178.0 *	MA
	Sf_3419.1	3536.5	15,615.7	3760.8	7.8	-	S	S	349.4 *	1558.8	2454.3	1759.2	-	-	-	MA
	Sf_2384.1	701.4	2381.5	591.1	0.9	-	3176.2	1148.8 *	-	3581.0	4311.6	10,104.0	-	-	-	MA
	Sf_14897.1	738.7	2140.7	676.2	1.0	-	10,901.0	3736.8 *	380.7 *	9555.0	10,384.0	23,211.0	-	-	-	MA
	Sf_3021.1	123.6	604.0	116.5	0.3	-	S	S	S	S	S	S	-	-	-	MA
	Sf_11251.1	22.5	15.8	11.5	-	-	-	-	-	-	-	313.7	409.5	-	-	-

C22	Sf_17877.1	35.8	282.4	131.8	0.2	-	-	-	-	1089.6	-	1264.9	-	-	-	EC
	Sf_18487.1	5.8	15.1	6.4	-	-	-	-	-	-	-	428.7	-	-	-	EC
	Sf_21341.1	0.4	3.6	1.0	-	-	-	-	-	-	-	-	-	-	-	
	Sf_2548.1	0.2	0.3	0.2	-	-	-	-	-	-	-	-	-	-	-	
C23	Sf_373.1	129.2	1460.2	175.5	0.4	-	-	-	-	7436.0	3339.2	2470.4	-	-	-	EC
	Sf_1788.1	31.3	1389.7	114.0	0.5	-	180.1 *	-	-	2805.3	959.3	2922.5	-	-	-	EC
	Sf_1755.1	7.3	723.5	59.6	0.5	-	-	-	-	-	-	-	-	-	-	
	Sf_5768.1	1.0	2.3	0.2	1.3	-	-	-	-	-	-	-	-	-	-	
TOTAL		7481.8	145,287.1	38,605.7	76.5											

AM, MM, and PM: anterior, middle, and posterior portions of the midgut, respectively; CAR: carcass (body without the midgut). ‡ Normalized spectral abundance factor varying from 0 to 1,000,000. * The value represents quantitatively less than 10% of the observed protein in relation to its identification across all analyzed samples. ECS, endoperitrophic content supernatant; EMM, enterocyte microvillar membrane; EF, ectoperitrophic fluid; FB, fat body; HS, hemolymph soluble proteins; mvMP, microapocrine vesicle membrane proteins; mvSP, microapocrine vesicle soluble proteins; PMW, peritrophic membrane washing; PMS, peritrophic membrane supernatant. EC, exocytosis; HC, hemolymph cells; MA, microapocrine secretion; DRM, protein found in detergent-resistant domains present in the microvillar membrane [21]; E, empirical observation of a GPI anchor [21]; P, positive for GPI anchor prediction using PredGPI [25]. Protein sequences with substitutions in the catalytic triad: ^a H57Y, ^b H57S, and S195V; ^c S195T (with best blast hits in the tree: Sfrs_SFRICE019800-PA, Sfrs_SFRICE005241-PA, Sfrs_SFRICE000209-PA, respectively, both with the same amino acid replacements in the catalytic triad as their *S. frugiperda* homologs). GPI anchor was not predicted for any chymotrypsin.

Some of the proteins that were secreted in microapocrine vesicles could also be found in detergent-resistant domains, which is apparently a way to recruit these molecules for release into the lumen. Trypsins were not observed via this route, but three putative chymotrypsins and two SP homologs were enriched in these domains [21]. Two of these putative chymotrypsins are shown in Table 2. The other sequences did not meet the established criteria to chymotrypsins identification used in the present work (see Material and Methods section).

There is a tendency of proteins on the same branch to follow the same secretory route (Tables 1,2). Exceptions to this pattern are observed in clades T18 (Figure S1) and C1 (Figure S2). Among the six sequences in group C21, only the low-abundance protein Sf_11251.1 was assigned to an exocytic route, whereas all other proteins seemed to be secreted through a microapocrine mechanism. The absence of Sf_11251.1 in microapocrine fractions could be due to its low abundance. Exocytosis seems to be the route used by the majority of both trypsins (70%) and chymotrypsins (63%). Chymotrypsins are also secreted in detergent-resistant membranes present in microapocrine vesicles through an unclear anchoring mechanism.

Other protein sequences from the *S. frugiperda* transcriptome predicted to include a trypsin domain (according InterProScan analysis, IPR001254), but not classified according our criteria as trypsins or chymotrypsins, are presented in Table S5. Some of these proteins were detected by proteomics analysis in the midgut lumen and are encoded by genes that are more highly expressed in the midgut than in the carcass. These proteins may represent unsampled digestive trypsin and chymotrypsin groups.

2.3. Noncanonical Peptidases and Their Putative Role in Lepidoptera Species

Several sequences with high similarity to trypsin and chymotrypsin sequences but with noncanonical amino acids occupying at least one of the positions constituting the catalytic triad (His57, Asp102, and Ser195) were observed. These sequences are highlighted in red in both phylogenetic trees (Figures S1 and S2).

Branches composed only of sequences exhibiting substitutions in the catalytic triad from different species were observed in both trees, showing that some of these sequences may have originated from duplications in ancient ancestral species and been preserved in these genomes to the present day.

Two trypsin and three chymotrypsin-homologous genes encoding protein sequences with amino acid replacements at the putative catalytic triad positions were expressed in the *S. frugiperda* midgut. Two genes encoding trypsin homologs with substitutions of the nucleophilic serine (S195F and S195A) and three genes encoding chymotrypsin homologs with substitutions of serine and histidine residues (H57Y, H57S, S195V, and S195T) were expressed in the *S. frugiperda* midgut. These substitutions were also present in the most similar sequences of these proteins in the enzyme phylogenetic trees. For example, the trypsin Sf_10598.1 transcript encodes a protein with an S195F substitution (i.e., it exhibits a phenylalanine residue instead of the catalytic serine residue). This amino acid replacement is also present in the most similar sequence to this protein in the trypsin phylogenetic tree (Sfrr_SFRICE004673-PA) (Figure S1). According to the phylogenetic analysis, this substitution may have occurred in a sequence of the ancestral Noctuidae species and been preserved in *S. frugiperda*, *S. litura*, *H. virescens*, and *T. ni* orthologous sequences. This gene is highly expressed in *S. frugiperda* midgut tissues (PM > AM > MM) and was identified by proteomics as being present in the peritrophic membrane and endoperitrophic space contents, showing that it may be an important protein in *S. frugiperda* digestion.

3. Discussion

3.1. Trypsin and Chymotrypsin Gene Families are Highly Expanded in Almost All Lepidoptera Groups

In this work, we searched for genes encoding trypsins and chymotrypsins in all available Lepidoptera genomes with protein annotation and analyzed their evolutionary history and sequence divergence. Phylogenetic analysis showed that gene family expansion events may have occurred in all studied lepidopteran taxonomic groups and at different times during the evolutionary history of these insects. Gene

duplication events involving these gene families seem to be very common in Lepidoptera species, and many of these copies have survived in these genomes for several millions of years (Figures S1 and S2).

Our previous study highlighted the presence of four lepidopteran trypsin groups, designated L-VII, L-VIII, L-IX, and L-X, which are exclusively composed of sequences from the Noctuidae family and exhibited unusual characteristics in their substrate-binding sites. These groups were hypothesized to have resulted from a specific lineage expansion event originating in Noctuidae family-specific ancestors in the recent past. In this work, these four groups of sequences were merged into group T18, which also included sequences from the Noctuidae family. However, the present analyses based on complete genome datasets from five Noctuidae species (and two *Spodoptera frugiperda* lineages) showed that these sequences may be a result of duplication events that occurred only in more recent ancestral species. This conclusion is based on the fact that the genome of *T. ni* (from the Plusiinae subfamily) exhibits five sequences in the T18 group, but all of them are more similar to each other than to the sequences from subfamilies Heliothinae (*H. virescens* and *H. armigera*) and Amphipyriinae (*S. litura* and *S. frugiperda*) and are probably derived from gene duplication events that occurred after the speciation event that gave rise to these lineages. The groups previously reported to have resulted from Noctuidae lineage-specific expansion events may in fact have resulted from several duplication events that occurred in the common ancestral species of Heliothinae and Amphipyriinae and in its descendant species. Finally, substrate-binding site characteristics that were considered to be exclusive to this Noctuidae group are also reported in other sequences from different trypsin groups and families (Figures S3 and S4).

Finally, based on the estimated average time for the silencing of one copy of a duplicated gene pair of approximately 4 million years [26], many of the duplication events observed in our trees may have already been silenced or will be in years to come.

3.2. Trypsin and Chymotrypsin Genes are Mainly Expressed in the Middle Midgut, and the Proteins are Mainly Secreted by Exocytic and Microapocrine Vesicles.

Trypsin and chymotrypsin enzymes exhibit decreasing activities along the midgut of *S. frugiperda* larvae [22]. However, according to our RNA-seq analysis, the genes encoding these proteins are mainly expressed in the middle midgut of this insect. One explanation for these findings is that the water counterflux in the ectoperitrophic space (outside PM) [22] must carry the newly synthesized proteins in the direction of the anterior midgut (see the schematic model of this mechanism in Figure 1B).

Biochemical and immunohistochemical data indicated that trypsin is secreted bound to the membrane via a microapocrine mechanism in the anterior midgut, whereas in the posterior region, trypsin is soluble and is observed in exocytic vesicles [20]. A further study suggested that detergent-resistant domains present in midgut microvilli could act as recruiters for protein secretion in microapocrine vesicles [21]. Although we did not obtain data about different regions along the midgut in our proteomic analyses, both microapocrine and exocytosis mechanisms can be proposed for different trypsins, as shown in Table 1. In addition, we observed trypsins in membrane fractions, confirming previous biochemical data.

It has been proposed that membrane-bound trypsins are anchored by a hydrophobic peptide [20]. However, the transmembrane prediction was only correct for a few low-abundance sequences, and none of these sequences were identified by mass spectrometry. The presence of a GPI anchor was empirically determined for Sf_2549.1, which exhibits low expression [21], and predicted for the highly abundant Sf_225.1. In both cited studies, phospholipase C treatment was used to solubilize GPI-anchored enzymes. Although Jordão and collaborators [20] did not observe differences between the trypsin activities of control and enzymatically treated samples, in another study, in which a highly sensitive mass spectrometer was used, five serine endopeptidase homologs were shown to be differentially released from midgut microvilli after phospholipase C incubation [21]. Thus, it seems likely that a third mechanism, such as lipidation, could explain how SPs with neither a transmembrane prediction nor a GPI anchor are bound to the membrane. This newly proposed mechanism could explain the trypsin activity observed in membrane fractions as well as how some proteins are associated with detergent-resistant domains (Table 1, Sf_290.1 and Sf_7342.1). Analytical efforts aimed at addressing this subject and the type of enzymes involved should be considered

for future research in the field. In particular, the prediction tools for lipidation do not seem to be effective for insect sequences, which require the development of mass spectrometry protocols for these specific targets.

Proteomic analyses confirmed that at a minimum, approximately 50% of the *S. frugiperda* proteins selected in this work as putative trypsin and chymotrypsin sequences may be considered digestive enzymes, as they were found in the midgut contents (ECS) with a single exception (Sf_186.1). In groups T7, T18, C1, C20, C21, and C23, most or all of the proteins have been sequenced, suggesting an important role of these enzymes and their closely related forms in the digestive process in *S. frugiperda*. Note that even products of putatively recent gene duplication events are translated into proteins and may be actively involved in digestion. In contrast, T11, a group with an average transcription expression level, was not identified via proteomic analysis. In addition, the presence of Sf_1782.1 chymotrypsin associated with the fat body seems to be an artifact, as it was also detected in all midgut fractions.

In general, the secretory route (exocytic or microapocrine secretion) is common to both trypsins and chymotrypsins from the branches of the corresponding trees, with the exception of trypsin T18. We could not identify common signatures within enzyme sequences associated with the type of secretory route.

3.3. Trypsin- and Chymotrypsin-Homologous Sequences must Play an Important Role in Lepidopteran Evolution

We found several SP-homologous protein sequences with high similarity to trypsins and chymotrypsins, but with changes in at least one of the residues occupying the positions of the enzyme catalytic triad (Figures S1 and S2). SP homologs were found in all Lepidoptera genomes, and some of them were grouped on branches composed of closely related species, showing that these sequences may have originated from a duplication event of an SP gene in a common ancestor. In addition, the residue changes were maintained in these genes upon lineage speciation. This conservation over time and the presence of these enzymes in *S. frugiperda* tissues, as observed by RNA-seq and proteomics analyses (Tables 1,2), show that these enzymes might play important roles in these insects.

Inactive enzyme homologs have been widely found in different Metazoan species where they play different roles, particularly in regulatory processes [27]. Enzyme catalytic site residues play an essential role in maintaining the activity of enzymes toward their canonical substrates, and must therefore be subjected to high purifying selective pressure. The high frequency of SP homologs observed in our work may be related to the large number of gene expansion events observed during trypsin and chymotrypsin gene family evolution, which may lead to newly generated variants.

The putative role of these SP homologs may be linked to the regulation of other SP enzymes, antimicrobial activities and/or the inactivation of plant PIs. These roles must be further elucidated.

4. Materials and Methods

4.1. Chymotrypsin and Trypsin Sequence Identification

Protein sequence datasets obtained from Lepidoptera genome sequencing projects were retrieved from the NCBI Genome [28], BIPAA (https://bipaa.genouest.org/sp/spodoptera_frugiperda_pub/), and OrthoDB databases [29]. Quantitative measures for the assessment of the gene set completeness of the 23 analyzed genomes according to BUSCO results [30] are shown in Figure S5. For genes with transcript isoforms, only the lengthiest protein sequence was retained for further analysis.

For all genomes, protein sequences were annotated using InterProScan software [31]. All sequences annotated as containing a trypsin domain (IPR001254) were selected for further analysis (see workflow in Figure S6). These sequences were grouped according their sequence similarity using the MAFFT online service with the UPGMA hierarchical clustering method [32]. Insect and *Homo sapiens* sequences previously annotated as encoding serine peptidase proteins were also included in the tree prediction to guide the selection of the branches, including putative trypsin, and chymotrypsin sequences.

Putative trypsin and chymotrypsin sequences were further classified according to the amino acid residue occupying position 189 (numbered according to bovine chymotrypsinogen, UniProt ID: P00766,

here and throughout the text). Based on this criterion, sequences with an aspartate at position 189 were selected as putative trypsins, and the remaining sequences were selected as putative chymotrypsins. These two groups of sequences were further filtered via a neighbor-joining clustering method [32] and according to their best hit annotation in BLASTp searches against the Refseq protein database. The predicted tree branches were classified according to their domain (predicted by InterproScan) and functional (predicted by RefSeq) annotation, and sequences within the majority of branches composed of proteins annotated as encoding nondigestive peptidases were removed from the next analysis. The classification and the number of sequences discarded for each insect genome in this step are provided in Table S6.

4.2. Gene Phylogenetic Tree Inference

Trypsin and chymotrypsin protein sequences were individually aligned using MAFFT software [33] and submitted for phylogenetic tree prediction using RAxML [34]. A total of 100 mL and 100 bootstrap searches were carried out using the best protein model automatically determined by RAxML by using the PROTGAMMAAUTO option. The resultant trees were visualized and annotated using Figtree software.

Branches with good support predicted for both enzyme trees were individually analyzed. These analyses included the number of genes per species genome and their consensus amino acids at positions related to the enzyme substrate-binding site (described in Figure 1A and based on studies by Bode et al. [35] and Hedstrom et al. [36]).

4.3. *Spodoptera frugiperda* Expression Analyses

Transcript sequencing analysis for *Spodoptera frugiperda* was conducted using the HiSeq2500 platform (Illumina/Solexa) and the paired-end strategy (2×100 bp). Three biological replicates, including ten insects each, divided into three midgut regions (anterior, middle, and posterior) and the carcass (larvae less midgut), were used in the analysis. Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Transcriptome assembly and RNA-seq analysis were conducted according to the method described by Dias et al. [37].

4.4. Midgut Fractionation and Proteomic Analyses

Figure 1B shows a schematic representation of the fractions analyzed by proteomics. Midgut microvillar membranes and microapocrine vesicles were obtained as previously described using pools of 25 last-instar larvae [21,38]. The ectoperitrophic fluid is equivalent to the last supernatant obtained from midgut washing for microapocrine vesicle isolation. The peritrophic membrane and the endoperitrophic content were separated after dissection. Then, the PM was washed twice in 100 μ L of two different solutions of 25 mM ammonium bicarbonate buffer, pH 7.5 (proteomics buffer), for 1 min at 3000 RPM. The solutions obtained from 25 PMs were pooled and centrifuged at $600 \times g$ for 20 min at 4 °C. The supernatant was collected and used as a peritrophic membrane washing (PMW) sample source. The remaining PM samples were homogenized in proteomics buffer using a Potter–Elvehjem device, followed by centrifugation for 20 min at $600 \times g$ and 4 °C, and the supernatant was collected for proteomic analyses. The isolated endoperitrophic contents of 25 animals were subjected to the same subsequent procedure as the PM remnants prior to analyses.

Hemolymph (20 μ L) was collected from a single animal after a small puncture was introduced near its third leg. Quickly thereafter, the sample was subjected to centrifugation at $16,100 \times g$ for 5 min at 4 °C. The supernatant was collected, and the pellet was resuspended in 20 μ L of deionized water. Both fractions were immediately frozen at -80 °C. Fat bodies from the same larvae were then dissected and homogenized in deionized water using a Potter–Elvehjem device.

Due to the different nature of the protein sources, two different protocols were used for protein digestion. In-gel and in-solution digestion were performed as previously described using 30 or 15 μ g of protein, respectively [21]. Each gel lane was cut into 2 mm bands and pooled into five samples for analysis, each of which included six subjacent bands starting from the top of the gel. Two biological replicates of each midgut fraction, hemolymph and fat body, were analyzed after digestion by the two protocols. Analyses

were conducted in a Q-Exactive Plus instrument (Thermo Scientific) equipped with a nanoAcquityUPLC System (Waters). Tryptic digests were resuspended in solution A (0.1% formic acid, 1% acetonitrile). The samples were first trapped in a nanoACQUITY UPLC Symmetry™ C18 Trap Column (100 Å, 5 µm, 180 µm × 200 mm, Waters) and then separated in a Peptide BEH C18 nanoACQUITY Column (130 Å, 1.7 µm, 150 µm × 100 mm, Waters). Trapping was performed for 5 min with a constant flow rate of 3 µl/min in solution A. Peptides were separated through a gradient from 100% solution A to 40% solution B (0.1% formic acid, 99% acetonitrile) over 90 min at a constant flow rate of 250 nL/min. The column was subsequently cleaned for 10 min with 95% solution B. The Q-Exactive Plus MS was operated in positive-ion mode with data-dependent acquisition. MS spectra were acquired in the Orbitrap with an automatic gain control (AGC) target value of 10×10^6 ions and a maximum fill time of 80 ms, and the full-scan event was performed over the mass range of m/z 380 to 1800 at a resolution of 70,000. Subsequently, ions selected for MS/MS were dynamically excluded for 30 s. For MS2 scans, higher-energy collisional dissociation (HCD) fragmentation was used, and the spectra were acquired at a resolution of 20,000; MS2 scans were acquired for the 10 most abundant ions, with an activation time of 0.1 ms over a mass range of m/z 100 to 2000. The minimal signal intensity required to trigger MS/MS fragmentation was 2.5×10^5 . Raw files were accessed using Xcalibur software (Thermo Scientific).

Generic Mascot files were generated using msConvert [39] with default parameters. The abovementioned transcriptome database was used for protein identification. The searches in Mascot (Matrix Sciences) allowed errors of 10 ppm and 0.1 Da for MS1 and MS2, respectively. Cysteine carbamidomethylation was set as a fixed modification, whereas methionine oxidation and asparagine deamidation were set as variable modifications. Two trypsin miss cleavages were allowed. The search results for the replicates of each midgut fraction, hemolymph, or fat body tissues obtained via both protocols were loaded as a single experiment onto Scaffold 4 [40] for comparisons between them using multidimensional protein identification technology. In the overall experiment, the proteins that exhibited positive identification presented at least two sequenced peptides. Under these conditions, a false discovery rate of 0.2% was observed. Quantification was performed using the normalized spectral abundance factor [41] in Scaffold.

5. Conclusions

We showed that the trypsin and chymotrypsin gene families are expanded in almost all Lepidoptera groups and that the large numbers of genes in these families are probably a result of duplication events that occurred at different times during the evolutionary history of Lepidoptera. We also identified a large number of genes encoding proteins with high similarity to trypsins and chymotrypsins but with an amino acid replacement at a minimum of one of the positions related to the enzyme's catalytic triad. Some of these proteins were confirmed to occur in different regions of *S. frugiperda* larvae by RNA-seq and proteomics analysis, suggesting a functional role.

Finally, we demonstrated the relevance of a recycling mechanism allowing newly synthesized enzymes to be carried toward the anterior midgut region in *S. frugiperda*. In addition, we propose that these enzymes are secreted via both exocytosis and microapocrine secretion mechanisms.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Maximum likelihood phylogenetic tree based on putative trypsin protein sequences, Figure S2: Maximum likelihood phylogenetic tree based on putative chymotrypsin protein sequences, Figure S3: Trypsin-binding site consensus for each similarity group, Figure S4: Chymotrypsin-binding site consensus for each similarity group, Figure S5: BUSCO assessment results for the 23 Lepidoptera species protein datasets, Figure S6: Overview of the bioinformatic workflow used in the putative trypsin and chymotrypsin protein sequences annotation, Table S1: Distribution of the number of genes related to each trypsin similarity group by genome, Table S2: Description of the trypsin protein sequences grouped in each similarity group, Table S3: Distribution of the number of genes related to each chymotrypsin similarity group by genome, Table S4: Description of the chymotrypsin protein sequences grouped in each similarity group, Table S5: Other sequences with trypsin domains in the *S. frugiperda* transcriptome, Table S6: Number of sequences with an identified trypsin domain that were not classified as trypsin or chymotrypsin sequences for each analyzed genome

Author Contributions: Conceptualization, R.O.D., W.R.T, and M.C.S.F.; Methodology, L.R.L., R.O.D., and F.J.F.; Validation, L.R.L., R.O.D., F.J.F., C.F., W.R.T, and M.C.S.F.; Formal Analysis, L.R.L., R.O.D., F.J.F., C.F., W.R.T, and M.C.S.F.; Investigation, L.R.L., R.O.D., and F.J.F.; Resources, C.F., W.R.T, and M.C.S.F.; Data Curation, L.R.L. and R.O.D., F.J.F., C.F., W.R.T, and M.C.S.F.; Writing—Original Draft Preparation, R.O.D., C.F., W.R.T, and M.C.S.F.; Writing—Review & Editing, L.R.L., R.O.D., F.J.F., C.F., W.R.T, and M.C.S.F.; Visualization, L.R.L., R.O.D., F.J.F., C.F., W.R.T, and M.C.S.F.; Supervision, R.O.D., C.F., W.R.T, and M.C.S.F.; Project Acquisition, L.R.L., R.O.D., F.J.F., C.F., W.R.T, and M.C.S.F.; Funding Acquisition, C.F., W.R.T, and M.C.S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grants 2014/50275-9 and 2017/08103-4, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant 482737/2012-3 to M.C.S.F., and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Finance Code 001. L.R.L. is the recipient of a CAPES fellowship. R.O.D. and F.J.F. are postdoctoral fellows of FAPESP. C.F. and W.R.T. are staff members of their respective departments and members of the Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular (INCT-EM). M.C.S.F, C.F. and W.R.T. are also research fellows of the CNPq.

Acknowledgments: We are grateful to Christiane Cardoso for technical assistance with the *Spodoptera frugiperda* biological samples, transcriptome assembly and RNA-seq analyses.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cristofolletti, P.T.; Ribeiro, A.F.; Terra, W.R. The cathepsin L-like proteinases from the midgut of *Tenebrio molitor* larvae: Sequence, properties, immunocytochemical localization and function. *Insect Biochem. Mol. Biol.* **2005**, *35*, 883–901.
2. Damasceno, T.F.; Dias, R.O.; de Oliveira, J.R.; Salinas, R.K.; Juliano, M.A.; Ferreira, C.; Terra, W.R. Active subsite properties, subsite residues and targeting to lysosomes or midgut lumen of cathepsins L from the beetle *Tenebrio molitor*. *Insect Biochem. Mol. Biol.* **2017**, *89*, 17–30.
3. Pimentel, A.C.; Fuzita, F.J.; Palmisano, G.; Ferreira, C.; Terra, W.R. Role of cathepsins D in the midgut of *Dysdercus peruvianus*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2017**, *204*, 45–52.
4. Lemos, F.J.A.; Terra, W.R. Digestion of bacteria and the role of midgut lysozyme in some insect larvae. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **1991**, *100*, 265–268.
5. Dias, R.O.; Via, A.; Brandão, M.M.; Tramontano, A.; Silva-Filho, M.C. Digestive peptidase evolution in holometabolous insects led to a divergent group of enzymes in Lepidoptera. *Insect Biochem. Mol. Biol.* **2015**, *58*, 1–11.
6. Souza, T.P.; Dias, R.O.; Castelhana, E.C.; Brandão, M.M.; Moura, D.S.; Silva-Filho, M.C. Comparative analysis of expression profiling of the trypsin and chymotrypsin genes from Lepidoptera species with different levels of sensitivity to soybean peptidase inhibitors. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2016**, *196–197*, 67–73.
7. Terra, W.R.; Ferreira, C. Insect digestive enzymes—Properties, compartmentalization and function. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **1994**, *109*, 1–62.
8. Ryan, C.A. Protease inhibitors in plants: Genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* **1990**, *28*, 425–449.
9. 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 Biochem. Mol. Biol.* **2007**, *37*, 1283–1290.
10. de Oliveira, C.F.R.; de Paula Souza, T.; Parra, J.R.P.; Marangoni, S.; de Castro Silva-Filho, M.; Macedo, M.L.R. Insensitive trypsins are differentially transcribed during *Spodoptera frugiperda* adaptation against plant protease inhibitors. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2013**, *165*, 19–25.
11. Paulillo, L.C.M.; Lopes, A.R.; Cristofolletti, P.T.; Parra, J.R.P.; 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. *J. Econ. Entomol.* **2000**, *93*, 892–896.
12. Pompermayer, P.; Falco, M.C.; Parra, J.R.P.; Silva-Filho, M.C. Coupling diet quality and Bowman-Birk and Kunitz-type soybean proteinase inhibitor effectiveness to *Diatraea saccharalis* development and mortality. *Entomol. Exp. Appl.* **2003**, *109*, 217–224.

13. Ekici, O.D.; Paetzel, M.; Dalbey, R.E. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci.* **2008**, *17*, 2023–2037.
14. Kawabata, S.I.; Tokunaga, F.; Kugi, Y.; Motoyama, S.; Miura, Y.; Hirata, M.; Iwanaga, S. Limulus factor D, a 43-kDa protein isolated from horseshoe crab hemocytes, is a serine protease homologue with antimicrobial activity. *Febs Lett.* **1996**, *398*, 146–150.
15. Yu, X.Q.; Jiang, H.; Wang, Y.; Kanost, M.R. Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol.* **2003**, *33*, 197–208.
16. Huang, T.S.; Wang, H.; Lee, S.Y.; Johansson, M.W.; Soderhall, K.; Cerenius, L. A cell adhesion protein from the crayfish *Pacifastacus leniusculus*, a serine proteinase homologue similar to *Drosophila* masquerade. *J. Biol. Chem.* **2000**, *275*, 9996–10001.
17. Devedjiev, Y.; Popov, A.; Atanasov, B.; Bartunik, H.D. X-ray structure at 1.76 Å resolution of a polypeptide phospholipase A2 inhibitor. *J. Mol. Biol.* **1997**, *266*, 160–172.
18. Wong, P.C.; Waggoner, D.; Subramaniam, J.R.; Tessarollo, L.; Bartnikas, T.B.; Culotta, V.C.; Price, D.L.; Rothstein, J.; Gitlin, J.D. Copper chaperone for superoxide dismutase is essential to activate mammalian Cu/Zn superoxide dismutase. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2886–2891.
19. Bartlett, G.J.; Borkakoti, N.; Thornton, J.M. Catalysing new reactions during evolution: Economy of residues and mechanism. *J. Mol. Biol.* **2003**, *331*, 829–860.
20. Jordao, B.P.; Capella, A.N.; Terra, W.R.; Ribeiro, A.F.; Ferreira, C. Nature of the anchors of membrane-bound aminopeptidase, amylase, and trypsin and secretory mechanisms in *Spodoptera frugiperda* (Lepidoptera) midgut cells. *J. Insect Physiol.* **1999**, *45*, 29–37.
21. Fuzita, F.J.; Pimenta, D.C.; Palmisano, G.; Terra, W.R.; Ferreira, C. Detergent-resistant domains in *Spodoptera frugiperda* midgut microvillar membranes and their relation to microapocrine secretion. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2019**, *235*, 8–18.
22. Bolognesi, R.; Terra, W.R.; Ferreira, C. Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models. *J. Insect Physiol.* **2008**, *54*, 1413–1422.
23. Bolognesi, R.; Ribeiro, A.F.; Terra, W.R.; Ferreira, C. The peritrophic membrane of *Spodoptera frugiperda*: Secretion of peritrophins and role in immobilization and recycling digestive enzymes. *Arch. Insect Biochem. Physiol.* **2001**, *47*, 62–75.
24. Santos, C.D.; Ferreira, C.; Terra, W.R. Consumption of food and spatial-organization of digestion in the cassava hornworm, *Erimnys-ello*. *J. Insect Physiol.* **1983**, *29*, 707–714.
25. Pierleoni, A.; Martelli, P.L.; Casadio, R. PredGPI: A GPI-anchor predictor. *BMC Bioinform.* **2008**, *9*, 392.
26. Lynch, M.; Conery, J.S. The evolutionary demography of duplicate genes. *J. Struct. Funct. Genom.* **2003**, *3*, 35–44.
27. Pils, B.; Schultz, J. Inactive enzyme-homologues find new function in regulatory processes. *J. Mol. Biol.* **2004**, *340*, 399–404.
28. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13.
29. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; Simão, F.A.; Zdobnov, E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **2018**, *47*, D807–D811.
30. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212.
31. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240.
32. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166.
33. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780.
34. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313.
35. 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. *Biochemistry* **1989**, *28*, 1951–1963.
36. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **2002**, *102*, 4501–4524.

37. Dias, R.O.; Cardoso, C.; Pimentel, A.C.; Damasceno, T.F.; Ferreira, C.; Terra, W.R. The roles of mucus-forming mucins, peritrophins and peritrophins with mucin domains in the insect midgut. *Insect Mol. Biol.* **2018**, *27*, 46–60.
38. Capella, A.N.; Terra, W.R.; Ribeiro, A.F.; Ferreira, C. Cytoskeleton removal and characterization of the microvillar membranes isolated from two midgut regions of *Spodoptera frugiperda* (Lepidoptera). *Insect Biochem. Mol. Biol.* **1997**, *27*, 793–801.
39. Chambers, M.C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D.L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30*, 918–920.
40. Searle, B.C. Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10*, 1265–1269.
41. Florens, L.; Carozza, M.J.; Swanson, S.K.; Fournier, M.; Coleman, M.K.; Workman, J.L.; Washburn, M.P. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **2006**, *40*, 303–311.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).