

Molecular Cloning and Characterization of a Midgut Chymotrypsin-Like Enzyme From the Lesser Grain Borer, *Rhyzopertha dominica*

Yu-Cheng Zhu and James E. Baker*

Grain Marketing and Production Research Center, ARS-USDA, Manhattan, Kansas

A cDNA encoding a chymotrypsinogen-like protein in midguts of the lesser grain borer, *Rhyzopertha dominica* (F.) (Coleoptera: Bostrichidae) was cloned and sequenced. The 901 bp cDNA contains an 816-nucleotide open reading frame encoding 272-amino acids. The predicted molecular mass and pI of the mature enzyme are 23.7 kDa and 4.64, respectively. The encoded protein includes amino acid sequence motifs that are conserved with 5 homologous chymotrypsinogen proteins from other insects. Features of the putative chymotrypsin-like protein from *R. dominica* include the serine proteinase active site (His⁹⁰, Asp¹³³, Ser²²⁶), conserved cysteine residues for disulfide bridges, the residues (Gly²²⁰, Gly²⁴³, Asp²⁵²) that determine chymotrypsin specificity, and both zymogen activation and signal peptides. A TPCCK-sensitive caseinolytic protein (P6) with an estimated molecular mass of 24 kDa is present in midgut extracts of *R. dominica* and can be resolved by electrophoresis on 4–16% polyacrylamide gels. The molecular mass of this caseinolytic enzyme is similar to that of the chymotrypsin deduced from cDNA. Midgut extracts of *R. dominica* readily hydrolyzed azocasein and N-succinyl-alanine-alanine-proline-phenylalanine-*p*-nitroanilide (SAAPF_pNA), a chymotrypsin-specific substrate. Properties of the enzymes responsible for these activities were partially characterized with respect to distribution in the gut, optimum pH, and sensitivity toward selected proteinase inhibitors. Optimal activity against both azocasein and SAAPF_pNA occurs in a broad pH range from about 7 to 10. Both azocasein and SAAPF_pNA activities, located primarily in the anterior midgut region, are inhibited by aprotinin, phenylmethylsulphonyl fluoride (PMSF), and soybean trypsin inhibitor (STI). TPCCK (N- α -tosyl-L-phenylalanine chloromethyl ketone) and chymostatin inhibited more than 60% of SAAPF_pNA but only about 10–20% of azocasein activity. These results provide additional evidence for the presence of serine proteinases, including chymotrypsin, in midguts of *R. dominica*. Arch. Insect Biochem. Physiol. 43:173–184, 2000. Published 2000 Wiley-Liss, Inc.†

*Correspondence to: James E. Baker, Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, KS 66502. E-mail: baker@usgmrl.ksu.edu

Received 14 October 1999; accepted 21 January 2000

Key words: *Rhyzopertha dominica*; lesser grain borer; digestion; midgut; serine proteinases; chymotrypsin; cDNA; TPCK; protein sequence

INTRODUCTION

The lesser grain borer, *Rhyzopertha dominica* (F.) (Coleoptera: Bostrichidae), is a major coleopteran pest of stored wheat in the United States. Both larvae and adults of this species can attack whole, intact grain kernels, a process that leaves grain susceptible to subsequent attack by a group of secondary pest species. As a primary insect pest, this species may be amenable to population suppression through development of genetically transformed cereals expressing an appropriate complex of inhibitors. Proteinase inhibitors may be effective in such a program (Baker and Kramer, 1996; Pittendrigh et al., 1997). However, because of the complexity of the interactions between proteinase inhibitors and proteinases in insect midguts (Reeck et al., 1997), a more complete understanding of protein digestion in this species is required before appropriate proteinase inhibitors can be formulated.

In contrast to grain weevils in which cysteine proteinases are major enzymes involved in protein digestion (Murdock et al., 1987; Liang et al., 1991; Houseman and Thie, 1993; Matsumoto et al., 1997), serine proteinases were shown to predominate in midguts of *R. dominica* (Konarev and Fomicheva, 1991). Because of the important pest status of *R. dominica*, and the need to document differences in digestive mechanisms between the grain borers and weevils, the mechanisms of protein digestion in *R. dominica* must be studied in more detail.

Zhu and Baker (1999) characterized trypsin-like enzyme in *R. dominica* against *N*- α -benzoyl-L-arginine *p*-nitroanilide (BAPNA), a substrate for trypsin-like enzymes. A multi-gene family of three cDNAs for trypsinogen-like proteins from midguts of this species were cloned and sequenced. These cDNAs *RdoT1*, *RdoT2*, and *RdoT3*, shared 77–81% sequence identity and the three encoded trypsinogens shared 51–62% identity in their amino acid sequences. Conserved features in these enzymes included the N-terminal residues IVGG, the active site triad (His¹⁰⁹, Asp¹⁵⁶, Ser²⁵⁷), three pairs of conserved cysteine residues, and the three residues (Asp²⁵¹, Gly²⁷⁴, Gly²⁸⁴) that determine trypsin specificity. In addition to the trypsin-like enzymes in midguts of *R. dominica*

with activity against BAPNA, there was also activity against *N*-succinyl-alanine-alanine-proline-phenylalanine-*p*-nitroanilide (SAAPFpNA), a chymotrypsin substrate. Zhu and Baker (1999) suggested that trypsin and chymotrypsin are the major serine proteinases in the midgut of *R. dominica*. Both trypsins and chymotrypsin are primarily luminal enzymes that initiate the hydrolytic breakdown of ingested dietary proteins to peptides of varying chain lengths. Trypsins cleave the protein chain on the carboxyl side of arginine or lysine, whereas chymotrypsin cleaves on the carboxyl side of aromatic amino acids. Terra and Ferreira (1994) provide a thorough review of the mechanisms and specificities of digestive proteinases in insects. In this report, activity against two substrates, SAAPFpNA and azocasein, is characterized in more detail and the cloning and sequencing of a cDNA from midgut mRNA that encodes a chymotrypsinogen-like protein in *R. dominica* is described.

MATERIALS AND METHODS

Insect Cultures

A stock culture of *R. dominica* was maintained on hard red winter wheat at 27°C and 55–65% RH. Lesser grain borers are cylindrically shaped, dark-brown beetles about 2.5 mm in length that weigh approximately 1.2 mg. Adults (1–3-week-old) were generally used in these studies.

Distribution of Proteinase Activity

Intestinal tracts were dissected in cold saline (128 mM NaCl, 4.7 mM KCl, 2.8 mM CaCl₂) by holding and pressing the abdomen and by removing the head simultaneously with forceps. Intact intestinal tracts were divided into four sections: foregut, anterior midgut, posterior midgut, and hindgut. Enzyme solution for each gut section was prepared as described by Walker et al. (1998). In brief, individual gut sections were immersed in 5 μ l/gut section of ice-cold 1 mM dithiothreitol (DTT), vortexed briefly, and centrifuged at 15,000g. Supernatant was transferred and frozen at –20°C until assayed.

General proteinase activity for each gut section was determined with azocasein with a slight modification of the method described by Walker et al. (1998). In this assay, 5 μ l enzyme solution (equivalent to enzyme from 1 gut section) was mixed with 10 μ l azocasein solution (1% w/v in 0.05% SDS), 20 μ l 200 mM tris-chloride pH 8, and 5 μ l H₂O. The reaction mixture was incubated for 2.5 h at 25°C during which time the hydrolysis of azocasein results in the release of the azo dye. Undigested protein was precipitated with 30 μ l 10% TCA, centrifuged, and the absorbance of a mixture of 60- μ l supernatant, plus 40 μ l 1 M NaOH was measured at 415 nm. Chymotrypsin activity was determined by using *N*-succinyl-alanine-alanine-proline-phenylalanine *p*-nitroanilide (SAAPFpNA) (Sigma, St. Louis, MO) as substrate (Oppert et al., 1996). In this assay, 5 μ l enzyme solution was mixed with 100 μ l SAAPFpNA in 100 mM tris-chloride pH 8 (final substrate concentration of 0.5 mg/ml). Absorbance at 405 nm was monitored continuously for 10 min at 37°C. Mean distribution of activities was determined from 5 replicates per gut section on each substrate.

Effect of pH on Midgut Proteinase Activity

Effect of buffer pH on midgut proteinase activities against azocasein and SAAPFpNA was measured by using the universal buffer system of Frugoni (1957). Midguts were dissected from 200 adults and pooled at a ratio of 1 gut/5 μ l 1 mM DTT. Enzyme solution was prepared by brief vortexing and centrifugation as described above. Five milliliters enzyme solution (1 midgut equivalent) were added to 40 μ l of azocasein in 0.05% SDS and 20 μ l buffer of each pH or to 100 μ l of SAAPFpNA in 10% *N,N*-dimethyl formamide and 50 μ l of each pH buffer. Results are based on 2 groups of midguts with triplicate analyses at each pH.

Inhibitors

Phenylmethylsulfonyl fluoride (PMSF), pepstatin, trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane (E-64), benzamidine, potato tuber carboxypeptidase inhibitor (PCPI), *N*- α -tosyl-L-phenylalanine chloromethyl ketone (TPCK), *N*- α -tosyl-L-lysine chloromethyl ketone (TLCK), and soybean trypsin inhibitor (STI, type I-S from soybean), were obtained from Sigma Chemical Company. Stock solutions were prepared in ethanol for pepstatin and TPCK, and in H₂O for the remaining inhibitors.

Enzyme solution (5 μ l) equivalent to 1 midgut was mixed with 45 μ l of universal buffer pH 8.0 and preincubated with 1 μ l of inhibitor stock solution for 15 min at 30°C before adding substrate. Residual activities against azocasein and SAAPFpNA were determined as above. Results are based on enzyme preparations from 2 groups of 150 midguts with triplicate analyses for each inhibitor.

Zymogram Analysis

Enzyme solution (5 μ l) equivalent to 1 midgut was mixed with 1 μ l of 100 mM TPCK and incubated at 30°C for 2 h. Subsequently, both inhibited and uninhibited enzyme samples were mixed with an equal volume of 2 \times sample buffer (0.125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.005% bromophenol blue), and loaded onto a 4–16% gradient tris-glycine gel with blue-stained β -casein (Novex, San Diego, CA). Enzymes were renatured in 2.5% Triton X-100, and the gel was developed by following the manufacturer's instructions.

Cloning Chymotrypsin-Like Protein cDNA

Cloning of the chymotrypsin-like cDNA was conducted in a near identical manner to that for the trypsinogen-like cDNAs from this species (Zhu and Baker, 1999) and similar to the procedure used by Yan et al. (1999) who cloned a trypsin cDNA from a curculionid weevil. Approximately 0.5 g of adult *R. dominica* were ground in liquid nitrogen. Total RNA was extracted with guanidine thiocyanate solution and precipitated with isopropanol (Titus, 1991). The poly(A) RNA was isolated from the total RNA using PolyATtract mRNA isolation system (Promega, Madison, WI). To obtain the 3'-end of chymotrypsin-like cDNA, reverse transcription was conducted, and amplification of the 3'-end of cDNA was performed using 3'-RACE (Gibco BRL Life-Technologies, Gaithersburg, MD). Polymerase chain reaction (PCR) was carried out by using Taq DNA polymerase and an oligo-dT reverse primer and a forward degenerate primer, 5'- TGYCARGGNGAYWSNNGGNGGCC-NYT-3', designed from a highly conserved region (CQGDSGGPL) in both *Manduca sexta* trypsin and chymotrypsin cDNAs located approximately 250 bp from the 3' end (Peterson et al., 1994, 1995). PCR-amplified DNA fragments (~250 bp)

were cloned into a pGEM-T vector (Promega). Sequences of these clones were determined by using an automated sequencer. cDNA sequence of the chymotrypsin-like protein was confirmed by homology search of GenBank provided by the National Center for Biotechnology Information using Blastx protocol (Altschul et al., 1990; Gish and States, 1993).

After the 3'-end sequence of chymotrypsin-like cDNA was obtained, three reverse primers, LGB13R0, LGB13R1, and LGB13R2, were designed (see Fig. 4). Reverse transcription was performed using LGB13R0 reverse primer and mRNA isolated from midgut tissue dissected from 300 adults. Two semi-nested amplifications were performed using the 5'-RACE system with a forward abridged anchor primer and two reverse primers, LGB13R1 and LGB13R2. DNA fragments (~700 bp) resolved from 5'-RACE amplification were cloned into a pGEM-T vector and sequenced from both directions using vector primers and 4 primers located on the insert (see Fig. 4).

To verify the cDNA sequence of chymotrypsinogen determined by using Taq DNA polymerase, a thermostable proof-reading Pfu DNA polymerase (Promega) was used to reamplify a full-length chymotrypsin cDNA fragment from 3'-RACE cDNA using a forward primer designed from the cDNA sequence determined with the Taq polymerase and an oligo-dT reverse primer. This fragment was A-tailed and cloned into a pGEM-T vector. The sequence of the insert was determined from both directions.

The Wisconsin Sequence Analysis Package GCG Unix version 9.0 (Genetics Computer Group, Madison, WI) including Pileup, Gap, Distances, and Growtree programs was used to analyze the similarity of chymotrypsin sequences (gap weight = 2, gap length weight = 1). Sequence analysis tools of the ExPASy Molecular Biology Server of Swiss Institute of Bioinformatics were used to process data of deduced protein sequences.

RESULTS

Distribution of Proteinase Activity

Both azocasein and SAAPFpNA were readily hydrolyzed by extracts prepared from guts of *R. dominica*. About 88% of general proteinase activity and 80% of SAAPFpNA activity in the intesti-

nal tract were located in the midgut (Fig. 1). The remaining activity was primarily in the hindgut. The small amount of activity in the hindgut may have resulted from the posterior movement of midgut fluid during dissection. The foregut, including a small crop region, had unmeasurable or very low activity against both substrates. Within the midgut, about 60% of the activity against both substrates was located in the anterior region and 40% in the posterior region.

Influence of pH on Proteinase Activity

Proteinase activity against azocasein and SAAPFpNA was optimal at neutral to alkaline pH (Fig. 2). When azocasein was used as substrate, there was no detectable activity below pH 4 but activity increased rapidly between pH 4 and 6, indicating the possible presence of cysteine or aspartic proteinases. Maximum activity against azocasein occurred across a broad pH range from about pH 6.5 to 9.5 and was reduced above pH 10. Optimum activity against SAAPFpNA was at about pH 9.5 but activity was near the maximum over a broad pH range from about pH 7 to 10.

Sensitivity of Proteinases to Inhibitors

The use of proteinase inhibitors to determine mechanistic class or specificity is complicated by the cross-reactions that many inhibitors have with different types of proteinases (Walker et al., 1998). Nevertheless, inhibitors can provide information about the more predominant enzymes in crude preparations. The effect of preincubation of enzymes prepared from *R. dominica* midguts

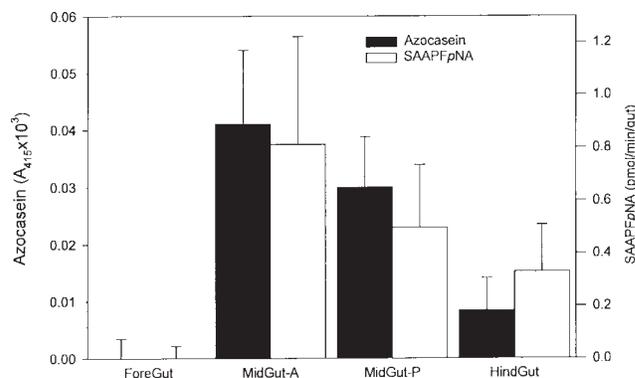


Fig. 1. Distribution of general proteinase and chymotrypsin activities (means \pm SE) in different gut regions of *R. dominica*. MidGut-A: anterior region of midgut; MidGut-P: posterior region of midgut.

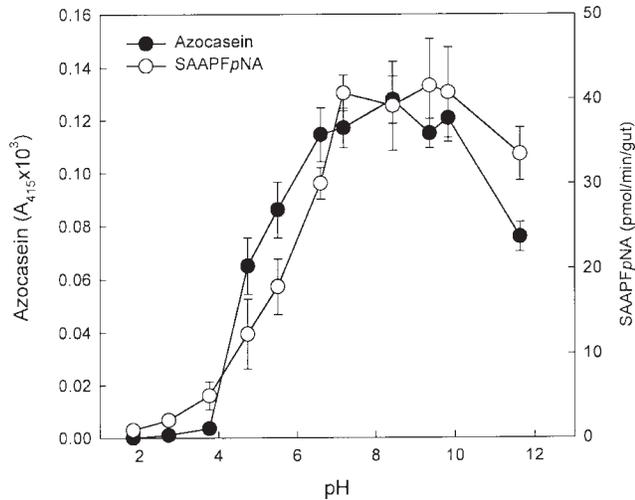


Fig. 2. Effect of pH on general proteinase and chymotrypsin activities (means \pm SE) in midguts of *R. dominica*.

with selected proteinase inhibitors on activity against azocasein and SAAPFpNA in our assays is presented in Table 1. The most effective inhibitor against azocasein was STI, which reduced activity by 92%. Aprotinin, PMSF, benzamidine, and leupeptin were also very effective inhibitors, providing evidence that serine proteinases were active in this species. E-64 and cystatin reduced azocasein activity by about 22 and 38%, respectively, but whether cysteine proteinases contribute to protein digestion remains to be determined. TLCK was much more effective than TPCK in the inhibition of activity against azocasein, providing evidence that trypsin-like activity is more prevalent than chymotrypsin-like activity in the ex-

tract. Based on inhibition by TPCK and chymostatin, chymotryptic activity constitutes about 10–17%, respectively, of the proteolytic activity in the crude extract. Pepstatin was ineffective, suggesting that aspartic proteinase inhibitors do not contribute to protein digestion in this species.

Chymotryptic activity against SAAPFpNA was inhibited by aprotinin, PMSF, STI, chymostatin, TPCK, and to a lesser extent by leupeptin. As expected, TLCK as well as the cysteine and aspartic proteinase inhibitors had no effect on chymotryptic activity against this substrate.

Proteinase Zymograms

To identify the enzymes inhibited by TPCK, midgut proteinases were separated on gradient polyacrylamide gels containing blue casein. Seven bands with proteinase activity (P1-P7) could be resolved in this gel system (Fig. 3, lane 1). When midgut extracts were preincubated with TPCK and then electrophoresed (Fig. 3, lane 2), the activity of proteinase P5 (~35 kDa) was partially suppressed, and activities of P6 (~24 kDa) and the minor proteinases P1-P4 (>50kDa) were completely suppressed. Activity of proteinase P7 (~22 kDa) was slightly reduced by preincubation with TPCK but P7 was not as sensitive to this inhibitor as was P6.

Chymotrypsinogen-Like Protein cDNA

Reverse transcription was conducted to generate the first strand cDNA from mRNA. A ~250 bp fragment was amplified from the first strand cDNA by using a degenerate forward primer and

TABLE 1. Effect of Selected Inhibitors on Activity of *R. dominica* Midgut Enzymes Against 2 Proteinase Substrates

Inhibitors	Conc.	Target enzymes	Azocasein relative activity*	SAAPFpNA pmol/min/gut	SAAPFpNA Relative activity*
Control	—	—	100 a	35.4	100 cd
Aprotinin	10 μ M	Serine protease	33 g	11.2	32 h
PMSF	5 mM	Serine protease	31 g	2.6	7 j
STI	100 μ M	Chymo-/trypsin	8 h	1.6	5 j
Chymostatin	10 μ M	Chymotrypsin	83 bc	8.2	23 i
TPCK	1 mM	Chymotrypsin	90 ab	14.5	41 g
TLCK	1 mM	Trypsin	48 ef	35.0	99 d
Leupeptin	200 μ M	Serine/cysteine	36 gf	23.8	67 f
E-64	100 μ M	Cysteine protease	78 bc	36.2	102 bcd
Cystatin	2 μ M	Cysteine protease	62 de	37.6	106 abc
Pepstatin	2 μ M	Aspartic protease	100 a	38.0	107 ab
Benzamidine	10 mM	Arg. bind pocket	36 gf	27.5	78 e
PCPI	5 μ M	Carboxypeptidase	72 cd	39.7	112 a

*Means followed by same letters are not significantly different ($P = 0.05$).

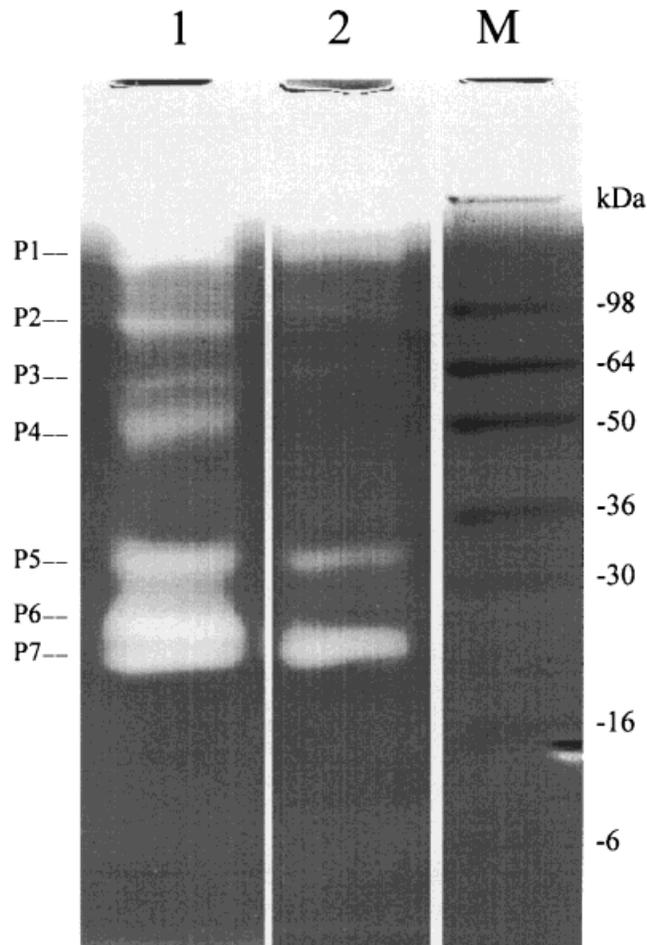


Fig. 3. Proteinase zymogram against blue casein. **Lane 1:** midgut enzymes of *R. dominica* (control); **lane 2:** midgut enzymes preincubated with TPCK. Proteinases designated as P1 to P7; **lane M:** protein molecular standard from Novex.

an oligo-dT reverse primer. This double-strand DNA fragment was cloned into pGEM-T vector, and its sequence was determined. A total of 25 recombinant clones from 3'-RACE was sequenced. Only one clone carried an insert of 229 bp nucleotides with a deduced protein sequence similar to chymotrypsin in the GenBank. Three reverse primers, LGB13R0, LGB13R1, and LGB13R2, were designed based on the 3'-end of the cDNA sequence. 5'-RACE RT was carried out using LGB13R0 reverse primer and gut tissue mRNA. Approximately 700 bp fragments were successfully amplified in both subsequent 1st PCR and 2nd semi-nested PCR amplifications. The fragment was cloned into pGEM-T vector and was sequenced from both directions. Sequences from

3'-RACE and 5'-RACE were combined. A GenBank homology search indicated that this cDNA contained a full sequence and that the putative amino acid sequence was similar to chymotrypsinogen-like protein.

To verify that these two fragments were amplified from one chymotrypsin mRNA species, reverse transcription was repeated. PCR amplification was performed using Pfu DNA polymerase with one forward primer flanking to the 1st 20 nucleotides, and a reverse oligo-dT primer. The clone was sequenced from both directions. The sequence of the new clone was identical to that of the previous clones except for 2 nucleotide mismatches. These two mismatches might have resulted from sequencer error. The cDNA sequence reported below was from the clone obtained by using Pfu DNA polymerase.

The 901 cDNA sequence included a start codon ATG at positions 10–12, a termination codon, TAA, at positions 826–828, and a polyadenylation signal, AATAAA, at positions 860–865 (Fig. 4). The open reading frame consisted of 816 nucleotides. The deduced amino acid sequence contained 272 residues. A 19-residue signal peptide was predicted using Signal1P program (Nielsen et al., 1997). The calculated molecular masses for the precursor of chymotrypsinogen (the prepro form of the enzyme including the signal and activation peptides) and the mature chymotrypsin were 29.3 and 23.7 kDa, respectively. There were 30 negative charged residues (Asp + Glu) and 15 positive charged residues (Arg + Lys) in the mature chymotrypsin. The calculated pI based on the deduced structure of the enzyme was 4.64.

The predicted amino acid sequence encoded by the *R. dominica* cDNA was aligned with five highly homologous insect chymotrypsin-like proteases obtained from GenBank (Table 2, Fig. 5). This sequence contained all of the conserved residues in the putative active site, His⁹⁰, Asp¹³³, Ser²²⁶. Six cysteine residues, predicted to occur in disulfide bridge configurations among trypsins and chymotrypsins, were located at positions 75, 91, 196, 212, 222, and 246.

DISCUSSION

Based on pH optima for general proteinase activity, activity against SAAPFpNA, specificity

```

CGATTTCGCC 9
LGB13F1→
ATGTCCTCCCAATCCCGCAATCTTGAAGTTGCTTGGTATGCTTTGGCTGCATTGCAGCTTTA 69
M S P I P A I L S C L V C F G C I A A L 20
                                     ↑
CCGCCTCCGATCCATAAAAAACAAACCCTTACTGATGAAGCACGATATGATCCACAACCTCT 129
P P P I H K N K P L L M K H D M I H N S 40
AGAAAGCATTTTCTGGTGATACGAATAAGATAGTAGGCGGAAGTGATGCTGAAGAAGCA 189
R K H F P G D T N K I V G G S D A E E A 60
                                     ▲
CAGTTCCTTTTATTGTGTCGTTGCAAACCTTTAGGACACAACCTGTGGTGGAAACAATTATA 249
Q F P F I V S L Q T L G H N C G G T I I 80
←LGB13R4
TCAGACAGATGGGTAGTATCAGCTGCCACTGTTTCGGCCACTCACCGGACTACAAGGTC 309
S D R W V V S A A H C F G H S P D Y K V 100
LGB13F2→
GTGGCAGGCGCAACCAAATTAAGTGAAGGAGGAGACAACCTATGGCGTGTCAAAGTGATA 369
V A G A T K L S E G G D N Y G V S K V I 120
GTACACGAAGAATATGACGATTTTGAAGATAGCAAACGATATTGCTCTTATCGAGACCAAT 429
V H E E Y D D F E I A N D I A L I E T N 140
TCTCCCATCTCATTCTCGTCTAAAGTTTCTCCATACCACTTGACGATTCCTACGTAGGC 489
S P I S F S S K V S S I P L D D S Y V G 160
AAGGACGTCAATGTGACGGCCATCGGTTGGGGTTTCACCGACTATCCATATGACCTTCCT 549
K D V N V T A I G W G F T D Y P Y D L P 180
←LGB13R3
GATCACCTGCAGTACATATCTCTGAAGACCATAGATAACAAGGATTGCGTAATTAGCCAC 609
D H L Q Y I S L K T I D N K D C V I S H 200
CCACTAGCACCACCCGTAACGGACGGCAACATTTGCACGCTCACCAAATTTGGAGAAGGA 669
P L A P P V T D G N I C T L T K F G E G 220
ACTTGCAAGGGAGATTTCAGGTGGACCACTGGTAGCCAATGGCAAATTTGGTGGTGTGTG 729
T C K G D S G G P L V A N G K L V G V V 240
←LGB13R2                                     ←LGB13R1
TCTTGGGGAAATCCATGTGCCAAGGGTGAGCCTGATGGGTATACAAGGGTATCGCACTAT 789
S W G N P C A K G E P D G Y T R V S H Y 260
LGB13R0
GTCGATTGGATAAGAGAGAAAACCTGGTTTGAAGTTTAAATTGTTTGATTAAGATTGT 849
V D W I R E K T G L E V 272
ATGAAGTTCTAATAAACGCAGTTTTCGGCATTAAAAAATAAAAAAAAAAAAAAAAAAAAA 901

```

Fig. 4. Nucleotide and deduced amino acid sequences of chymotrypsinogen-like cDNA isolated from *R. dominica* gut RNA. ATG = start codon; TAA = termination codon; AATAAAA = polyadenylation signal; ↑ = predicted signal peptide cleavage site; closed triangle = predicted activation peptide cleav-

age site. IVGG are conserved N-terminal residues. Primer sequences used for cloning and sequencing cDNA are underlined and labeled on the top. The sequence has been deposited in GenBank with accession number AF127088.

trophoresis on gradient gels, a minor band at 35 kDa and a major band at about 23 kDa. The major band of activity against SAAPF_pNA was similar in size to the caseinolytic band P6 at 24 kDa

that we found in adult midguts. Although the electrophoretically-resolved chymotryptic bands from *R. dominica* were nearly identical in size to the molecular mass of enzyme deduced from the

TABLE 2. Similarity Comparison and Evolutionary Distance Relationship (GCG) of *R. dominica* Chymotrypsin-Like Proteinase to Chymotrypsin From Other Insect Species*

Species	Proteinase type	Similarity (%)	Identity (%)	Evolutionary distance
<i>R. dominica</i>	chymotrypsin	—	—	
<i>A. gambiae</i>	chymotrypsin 1	60	50	
<i>A. gambiae</i>	chymotrypsin 2	59	48	
<i>V. orientalis</i>	chymotrypsin 2	56	47	
<i>V. crabro</i>	chymotrypsin P2	57	48	
<i>V. crabro</i>	chymotrypsin 2	57	48	

*This tree was generated using the GCG unrooted Growtree program using the Neighbor-joining method (Saitou and Nei, 1987).

of inhibitors, and deduced protein sequences, this study provides additional evidence that serine proteinases are the major mechanistic class of enzymes involved in protein digestion by *R. dominica*. Among the serine proteinases, both trypsin (Zhu and Baker, 1999) and chymotrypsin are present in adult midguts. These results generally confirm the prior studies of Konarev and Fomicheva (1991) and Oppert et al. (personal communication) showing the presence of serine proteinases in this species. Houseman and Thie (1993) documented the presence of serine proteinases in the larger grain borer, *Prostephanus truncatus*, another bostrichid beetle.

Results with the inhibitor pepstatin indicate that aspartic proteinases are apparently not active in midgut preparations. E-64 inhibited 22% of azocasein activity but did not affect SAAPFpNA activity in our assay. E-64, which is considered to be specific for cysteine proteinases, was shown to inhibit trypsin activity against BApNA (Sreedharan et al., 1996). E-64 inhibited 34% of BApNA activity in *R. dominica* (Zhu and Baker, 1999) providing evidence that the E-64 inhibition of azocasein activity may result from inhibition of serine proteinase activity and not cysteine proteinase activity. However, additional studies are needed to confirm the presence or absence of cysteine proteinases in this species.

Intestinal fluid pH in *R. dominica* ranges from 6.8 in the foregut, to 5.2–6.8 in the anterior midgut and 7.0 to 7.2 in the posterior midgut, to acidic (pH 3.6–4.6) in the hindgut (Sinha, 1959). Based on our in vitro assays, the neutral to alkaline proteinases, including chymotryptic activity against SAAPFpNA in *R. dominica*, have a rather broad pH optima and probably would be active throughout most of the midgut region

except perhaps at the very anterior connection to the foregut.

The distribution of enzymes within insect midguts can be arranged radially (i.e., from the endoperitrophic space outward to the epithelium) as well as longitudinally along the length of the midgut (Terra and Ferreira, 1994). The midgut of *R. dominica* is a tubular tract about 2 mm in length with the anterior portion slightly enlarged relative to the posterior section. The midgut tissue is fragile and we assume that the vortexing procedure effectively releases luminal enzymes. It was not possible to measure ectoperitrophic levels of enzymes. Despite the small size of the gut tract, we did find that the midgut is the primary source of the endoproteinases and higher activity against both azocasein and SAAPFpNA in the anterior region of the midgut compared with their activities in the posterior region.

Konarev and Fomicheva (1991) found 5 major zones of proteolytic activity against gelatin in zymograms prepared after isoelectric focusing of extracts prepared from adult *R. dominica* guts. The proteinases had pIs ranging from 3.5 to 10 and all were inhibited by STI. All bands were also active against BApNA, but the bands with more alkaline pI values were much less active against this substrate than were bands with acidic pI values. Our PAGE results with gradient gels containing casein also indicate a complex of proteinases present in midguts of *R. dominica*. The estimated molecular mass of proteinase band P6 that was one of those inhibited by TPCK was 24 kDa, a value similar to that calculated from the deduced amino acid sequence (23.7 kDa). Oppert et al. (unpublished data) demonstrated 2 bands of activity from *R. dominica* larval midguts against SAAPFpNA on zymograms following elec-

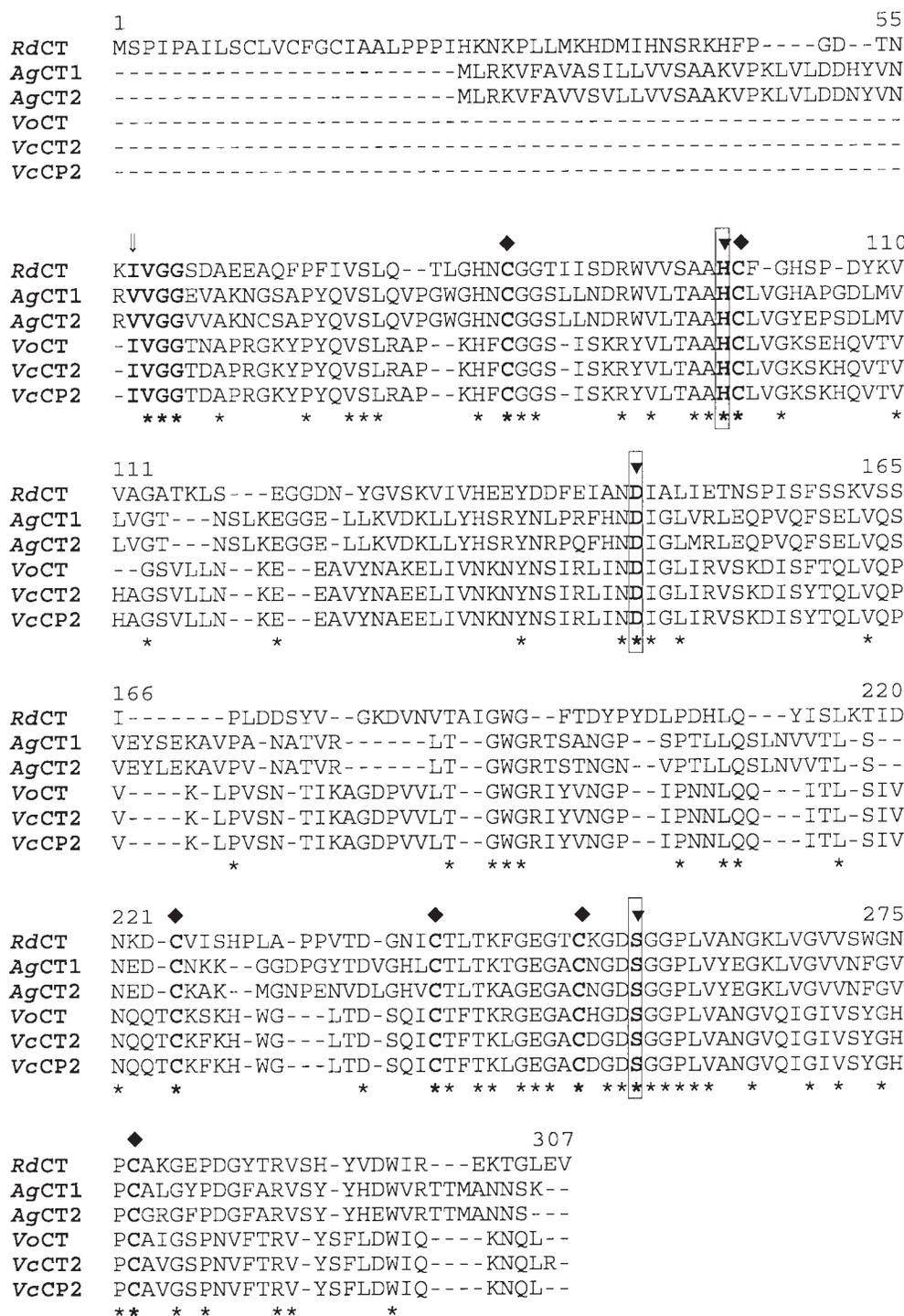


Fig. 5. Predicted amino acid sequence of chymotrypsinogen-like protein from *R. dominica* and alignment with 5 other insect chymotrypsin-like sequences. RdCT = chymotrypsinogen-like sequence of *R. dominica*; AgCT1 = chymotrypsin 1 precursor from the African malaria mosquito, *A. gambiae*. AgCT2 = chymotrypsin 2 precursor from the African malaria mosquito, *A. gambiae*. VoCT = chymotrypsin II from oriental hornet *V. orientalis*. VcCT2 = chymotrypsin II from European hornet *V. crabro*. VcCP2 = chymotrypsin-like

proteinase II from European hornet *V. crabro*. Functionally important residues are boxed and indicated by bold letters and solid triangles on the top of sequences. Cysteines corresponding to the sites of predicted disulfide bridges are marked with bold letters and solid diamonds on the top. Identical residues among all six sequences are indicated with stars at the bottom of sequences. The downward arrow indicates the N-terminal residues of the active enzymes. Hyphens represent sequence alignment gaps.

cDNA, we have no direct evidence that the chymotrypsin we cloned was responsible for the hydrolysis of either casein or SAAPFpNA.

The predicted amino acid sequence encoded by the *R. dominica* cDNA was aligned with five homologous insect chymotrypsin-like proteases (Fig. 5). This sequence contained all of the conserved residues in the putative active site, His⁹⁰, Asp¹³³, Ser²²⁶, which form the catalytic triad in serine proteases (Kraut, 1977; Wang et al., 1993; Peterson et al., 1994). Six cysteine residues, predicted to occur in disulfide bridge configurations among trypsins and chymotrypsins, exhibited the same conserved pattern of cysteine residues found in all six insect serine proteinases. The residues (Gly²²⁰, Gly²⁴³, Asp²⁵²) that define the substrate binding pocket were highly conserved in all of these chymotrypsin-like enzymes. Gly²²⁰ was predicted to be located at the bottom of the binding pocket, which determines specificity in bovine chymotrypsin B (Ser, Gly, Ala) (Hedstrom et al., 1992; Peterson et al., 1995).

Mature sequences of lepidopteran gut serine proteases identified previously contain no lysine residues but have high arginine levels, and it has been suggested that this is an adaptation necessary for stability in the high pH environment of these guts (Peterson et al., 1994; Wang et al., 1995). This adaptation apparently has not occurred in *R. dominica* whose midgut pH ranges from slightly acid in the anterior region to slightly alkaline in the posterior region (Sinha, 1959). The deduced sequence of the mature chymotrypsin from *R. dominica* included 5.9% lysine residues and 1.4% arginine residues.

The sequence IleValGlyGly at positions 51–54 is highly conserved in many trypsin- and chymotrypsin-like proteinases and marks the N-termini of the active enzymes (Jany and Haug 1983; Wang et al., 1993). The proenzyme for the chymotrypsinogen-like protein contains a typical long activation peptide (31 amino acid residues), which ends in a basic lysine residue, suggesting that tryptic cleavage (Arg-Ile or Lys-Ile bond) might be involved in activation. A scan of PROSITE database revealed that the deduced protein sequence belongs to the trypsin family of serine proteases that contains a conserved histidine active site sequence pattern [LIVM]-[ST]-A-[STAG]-H-C and a conserved serine active site sequence pattern

[DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-[LIVMFYSTANQH].

A search of the GenBank database using the Blastx non-redundant program revealed that the deduced sequence of the *R. dominica* protein was most similar to proteins in the serine proteinase trypsin/chymotrypsin family. The most similar sequences included chymotrypsin 2 precursor and chymotrypsin 1 precursor from the malaria mosquito, *Anopheles gambiae* (Swiss-Prot accession Q17025 and Q27289; Mueller, unpublished data). Other highly similar insect chymotrypsin-like sequences included chymotrypsin II (Swiss-Prot accession P00768) from the oriental hornet, *Vespa orientalis* (Jany et al., 1983), chymotrypsin-like proteinase II (PID g223690), and chymotrypsin II (Swiss-Prot accession P00769) from the European hornet, *Vespa crabro* (Jany and Haug, 1983). GCG Gap analysis indicated that the deduced chymotrypsinogen-like protein sequence from *R. dominica* was most similar to the chymotrypsin 1 precursor from *An. gambiae* with 60% similarity and 50% identity in the amino acid residues (Table 2). Sequence analyses using GCG Distances and Growtree (Neighbor-joining; Saitou and Nei, 1987) methods also indicated that the chymotrypsinogen-like protein sequence from *R. dominica* was most similar to the chymotrypsin 1 precursor from *An. gambiae* (Table 2).

R. dominica has a complex of at least 1 chymotrypsin and 3 trypsin genes involved in protein digestion. These and similar types of multiple gene complexes that encode digestive proteinases in insects may be an adaptive mechanism for reducing the deleterious effects of plant protease inhibitors (Bown et al., 1997). How the expression and activity levels of these and other proteinases in *R. dominica* fluctuate in response to dietary inhibitors, and provide a biochemical plasticity during food conversion processes, needs to be determined to maximize the potential for pest management using genes introduced into cereals.

ACKNOWLEDGMENTS

The authors thank Haobo Jiang of Kansas State University and B. Oppert and K. J. Kramer of USDA-ARS-GMPCR, Manhattan, KS, for their technical advice. The authors thank M. J. Chris-

peels, B. Oppert, and J. Reese for comments and suggestions on an early version of this manuscript.

LITERATURE CITED

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Baker JE, Kramer K J. 1996. Biotechnological approaches for stored-product insect pest management. *Postharvest News Info* 7:11N–18N.
- Bown DP, Wilkinson HS, Gatehouse JA. 1997. Differentially regulated inhibitor-sensitive and insensitive protease genes from phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. *Insect Biochem Mol Biol* 27:625–638.
- Frugoni JAC. 1957. Tampone universale di Britton e Robinson a forza ionica costante. *Gaz Chim Ital* 87:403–407.
- Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. *Nat Genet* 3:266–272.
- Hedstrom L, Szilagyi L, Rutter WJ. 1992. Converting trypsin to chymotrypsin: the role of surface loops. *Science* 255:1249–1253.
- Houseman JG, Thie NMR. 1993. Difference in digestive proteolysis in the stored maize beetles: *Sitophilus zeamais* (Coleoptera: Curculionidae) and *Prostephanus truncatus* (Coleoptera: Bostrichidae). *J Econ Entomol* 86:1049–1054.
- Jany KD, Haug H. 1983. Amino acid sequence of the chymotryptic protease II from the larvae of the hornet, *Vespa crabro*. *FEBS Lett* 158:98–102.
- Jany KD, Bekelar K, Pfliederer G, Ishay J. 1983. Amino acid sequence of an insect chymotrypsin from the larvae of the hornet, *Vespa orientalis*. *Biochem Biophys Res Commun* 110:1–7.
- Konarev AV, Fomicheva YV. 1991. Cross analysis of the interaction of α -amylase and proteinase components of insects with protein inhibitors from wheat endosperm. *Biokhimiya* 56:419–427.
- Kraut J. 1977. Serine proteases: structure and mechanism of catalysis. *Ann Rev Biochem* 46:331–358.
- Liang C, Brookhart G, Feng GH, Reeck GR, Kramer KJ. 1991. Inhibition of digestive proteinases of stored grain coleoptera by oryzacystatin, a cysteine proteinase inhibitor from rice seed. *FEBS Lett* 278:139–142.
- Matsumoto I, Emori Y, Abe K, Arai S. 1997. Characterization of a gene family encoding cysteine proteinases of *Sitophilus zeamais* (maize weevil), and analysis of the protein distribution in various tissues including alimentary tract and germ cells. *J Biochem* 121:464–476.
- Murdock LL, Brookhart G, Dunn PE, Foard DE, Kelley S, Kitch L, Shade RE, Shukle RH, Wolfson JL. 1987. Cysteine digestive proteinases in Coleoptera. *Comp Biochem Physiol* 87B:783–787.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 10:1–6.
- Oppert B, Kramer KJ, Johnson D, Upton SJ, McGaughey WH. 1996. Luminal proteinases from *Plodia interpunctella* and the hydrolysis of *Bacillus thuringiensis* Cry1A(c) protoxin. *Insect Biochem Mol Biol* 26:571–583.
- Peterson AM, Barillas-Mury CV, Wells MA. 1994. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. *Insect Biochem Mol Biol* 24:463–471.
- Peterson AM, Fernando GP, Wells MA. 1995. Purification, characterization and cDNA sequence of an alkaline chymotrypsin from the midgut of *Manduca sexta*. *Insect Biochem Molec Biol* 25:765–774.
- Pittendrigh BR, Hueusing JE, Shade RE, Murdock LL. 1997. Effects of lectins, CRY1A/CRY1B Bt delta-endotoxin, PAPA, proteinase and alpha-amylase inhibitors, on the development of the rice weevil, *Sitophilus oryzae*, using an artificial seed bioassay. *Entomol Exp Appl* 82:201–211.
- Reeck GR, Kramer KJ, Baker JE, Kanost MR, Fabrick JA, Behnke CA. 1997. Proteinase inhibitors and resistance of transgenic plants to insects. In: Carozzi N, Koziel M, editors. *Advances in insect control: the role of transgenic plants*. Bristol, PA: Taylor & Francis, p 157–183.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
- Sinha RN. 1959. The hydrogen-ion concentration in the alimentary canal of beetles infesting stored grain and grain products. *Ann Entomol Soc Am* 52:763–765.
- Sreedharan SK, Verma C, Caves LSD, Brocklehurst SM, Gharbia SE, Shah HN, Brocklehurst K. 1996. Demonstration that 1-*trans*-exopoxysuccinyl-L-leucylamido-(4-guanidino)butane (E-64) is one of the most effective low Mr inhibitors of trypsin-catalysed hydrolysis. Characterization by kinetic analysis and by energy minimization and molecular dynamics simulation of the E-64- trypsin complex. *Biochem J* 316:777–786.
- Terra WR and Ferreira C. 1994. Insect digestive enzymes: properties, compartmentalization and function. *Comp Biochem Physiol* 109B:1–62.

- Titus D. 1991. Promega protocols and application guide, 2nd ed. Madison, WI: Promega Corporation.
- Walker AJ, Ford L, Majerus MEN, Geoghegan IE, Birch N, Gatehouse JA, Gatehouse AMR. 1998. Characterisation of the mid-gut digestive proteinase activity of the two-spot ladybird (*Adalia bipunctata* L.) and its sensitivity to proteinase inhibitors. *Insect Biochem Molec Biol* 28:173–180.
- Wang S, Magoulas C, Hickey DA. 1993. Isolation and characterization of a full-length trypsin- encoding cDNA clone from the lepidopteran insect, *Choristoneura fumiferana*. *Gene* 136: 375–376.
- Wang S, Young F, Hickey DA 1995. Genomic organization and expression of a trypsin gene from the spruce budworm, *Choristoneura fumiferana*. *Insect Biochem Mol Biol* 25:899– 908.
- Yan XH, De Bondt HL, Powell CC, Bullock RC, Borovsky D. 1999. Sequencing and characterization of the citrus weevil, *Diaprepes abbreviatus*, trypsin cDNA. *Eur J Biochem* 262:627–636.
- Zhu YC, Baker JE. 1999. Characterization of midgut trypsin-like enzymes and three trypsinogen cDNAs from the lesser grain borer, *Rhyzopertha dominica* (Coleoptera: Bostrichidae). *Insect Biochem Mol Biol* 29:1053–1063.