ELSEVIER

Insect Biochemistry and Molecular Biology 34 (2004) 903–918

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

A diverse family of serine proteinase genes expressed in cotton boll weevil (*Anthonomus grandis*): implications for the design of pest-resistant transgenic cotton plants

Osmundo B. Oliveira-Neto [a,b], João A.N. Batista [a], Daniel J. Rigden [a], Rodrigo R. Fragoso [a,b], Rodrigo O. Silva [a,b], Eliane A. Gomes [c], Octávio L. Franco [a,d], Simoni C. Dias [a,b], Célia M.T. Cordeiro [a], Rose G. Monnerat [a], Maria F. Grossi-de-Sá [a,*]

[a] *Embrapa Recursos Genéticos e Biotecnologia, S.A.I.N. Parque Estação Biológica, Final W3, Asa Norte, Brasilia, DF 70770-900, Brazil*

[b] *Universidade de Brasília, Brasilia, DF 70910-900, Brazil*

[c] *Embrapa Milho e Sorgo, Sete Lagoas, MG 35701-970, Brazil*

[d] *Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Catolica de Brasilia, 916N Asa Norte, Brasilia, DF 70770-900, Brazil*

Received 22 May 2004; accepted 1 June 2004

Abstract

Fourteen different cDNA fragments encoding serine proteinases were isolated by reverse transcription-PCR from cotton boll weevil (*Anthonomus grandis*) larvae. A large diversity between the sequences was observed, with a mean pairwise identity of 22% in the amino acid sequence. The cDNAs encompassed 11 trypsin-like sequences classifiable into three families and three chymotrypsin-like sequences belonging to a single family. Using a combination of 5′ and 3′ RACE, the full-length sequence was obtained for five of the cDNAs, named *Agser2*, *Agser5*, *Agser6*, *Agser10* and *Agser21*. The encoded proteins included amino acid sequence motifs of serine proteinase active sites, conserved cysteine residues, and both zymogen activation and signal peptides. Southern blotting analysis suggested that one or two copies of these serine proteinase genes exist in the *A. grandis* genome. Northern blotting analysis of *Agser2* and *Agser5* showed that for both genes, expression is induced upon feeding and is concentrated in the gut of larvae and adult insects. Reverse northern analysis of the 14 cDNA fragments showed that only two trypsin-like and two chymotrypsin-like were expressed at detectable levels. Under the effect of the serine proteinase inhibitors soybean Kunitz trypsin inhibitor and black-eyed pea trypsin/chymotrypsin inhibitor, expression of one of the trypsin-like sequences was upregulated while expression of the two chymotrypsin-like sequences was downregulated.

Keywords: *Anthonomus grandis*; Serine proteinases; cDNA cloning; Multigene family

1. Introduction

Serine proteinases are one of a diverse group of enzymes capable of cleaving peptide bonds, and are involved in various essential processes such as intra- and extra-cellular protein metabolism, blood coagulation, immune response, fertilization and developmental regulation, digestion, among others (Elvin et al., 1994; Barrett and Rawlings, 1995; Rao et al., 1998; Nagano et al., 2003).

These enzymes have been detected in several insect orders including Coleoptera (Zhu and Baker, 1999), Diptera (Jiang et al., 1997) and Lepidoptera (Gate-

Abbreviations: Agser, *Anthonomus grandis* serine proteinase; ARE, AU-rich element; BTCI, black-eyed pea trypsin/chymotrypsin inhibitor; CUB domain, complement-Uegf-BMP-1 domain; $CI_{0.95}$, 95% confidence interval; RT-PCR, reverse transcription-polymerase chain reaction; SKTI, soybean Kunitz trypsin inhibitor; UTR, untranslated region

* Corresponding author. Cenargen/Embrapa, S.A.I.N. Parque Rural, Final W5, Asa Norte, 70770-990, Brasilia, DF, Brazil. Tel.: +55-61-448-4705; fax: +55-61-340-3624.

E-mail address: fatimasa@cenargen.embrapa.br (M.F. Grossi-de-Sá).

0965-1748/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2004.06.001

house et al., 1999) and are one of the major classes of endopeptidases, strongly implicated in the digestion of proteins in the gut (Applebaum, 1985; Terra and Ferreira, 1994). Larvae and adult insects of the coleopteran *Anthonomus grandis*, the boll weevil, a key pest of cotton, *Gossypium hirsutum*, are widespread in tropical and temperate regions of the American continent. The female weevil feeds, deposits its eggs and develops primarily in the flower buds and the larvae remain within the bud after hatching from the egg, using the buds as a food source and as a protective habitat until development is completed and the adult emerges. The control of this insect is difficult due to the inability to achieve contact between the control agent and the insect into the plant buds.

One form of natural defense of plants against insect pests is mediated by proteinase inhibitors (Ryan, 1990). Several reports show evidence that proteinase inhibitors have potential to reduce insect growth and development by blocking the digestive proteinases in the larval gut, thereby limiting the release of amino acids from food protein (Broadway et al., 1986; Johnston et al., 1993; Pompermayer et al., 2001). Genes encoding proteinase inhibitors have been used to produce transgenic plants resistant to insects (Hilder et al., 1987; Thomas et al., 1995; Duan et al., 1996; Christeller et al., 2002; Fabrick et al., 2002). However, in response to inhibitor ingestion, the range of gut proteinases can be shifted toward enzymes that are insensitive to the plant inhibitors, resulting in adaptation of gut proteolysis to proteinase inhibitors (Jongsma et al., 1995; Paulillo et al., 2000; Cloutier et al., 2000; Brito et al., 2001).

We have recently demonstrated that two inhibitors, one from soybean (SKTI) and other from cowpea seeds (BCTI), affect the larval development of the boll weevil in artificial diet (Franco et al., 2003, 2004). In this study, we have isolated several serine proteinase genes of *A. grandis* larvae and we evaluated the effects of SKTI and BCTI on the expression profiles of these genes on larvae feeding in the presence of these inhibitors.

2. Materials and methods

2.1. Insect rearing

A population of *A. grandis* (Coleoptera: Curculionidae) was maintained at 27 ± 1 °C, $70 \pm 10\%$ relative humidity at 14 h day length. Insects were routinely maintained on standard rearing diet (Monnerat et al., 2000). All components were purchased from Sigma (St. Louis, USA). Wild insects were collected in Unaí, Brazil. The BTCI and SKTI were incorporated in the diet at concentrations of 50 and 150 μM, respectively. Protein concentrations were calculated according to Bradford (1976).

2.2. Gut extracts and serine proteinase activity assays

Midguts from third-instar larvae and 10-day-old adults were excised from cold-anesthetized larvae and adults and ground in cold 0.15 M NaCl using a 1:10 (w/v) ratio. The homogenate was centrifuged twice at $10,000 \times g$ for 15 min at 4 °C and stored at −20 °C until use. The proteinase assay procedure was adapted from Sarath et al. (1990). To determine serine proteinase activity, aliquots of midgut homogenates were incubated with 1 mM N_{α}-benzoyl-D-arginine-*p*-nitroanilide (BApNA) as substrate for 30 min at 37 °C in Tris–HCl buffer pH 8.6 (Borovsky and Schlein, 1988). The reaction was stopped with 30% (v/v) acetic acid. Absorbance at 410 nm was measured after 15 min at room temperature. For each individual assay, 100 $ng.ml^{-1}$ of protein extract were used. Assays were carried out in triplicate.

2.3. Purification of soybean Kunitz trypsin (SKTI) and Bowman–Birk trypsin (BTCI) inhibitors

An SKTI enriched fraction was purchased from Sigma Co. (St. Louis, USA). Proteins were precipitated with ammonium sulfate in a range of 0–100% saturation. Precipitation was followed by analysis in an HPLC reversed-phase analytical column (Vydac 218 TP 1022 C-18) at a flow rate of 1.0 ml min^{-1}. SKTI fraction eluted with a linear gradient of acetonitrile (0–100%) was collected, lyophilized and stored at −20 °C. BTCI was purified in accordance with Freitas et al. (1999).

2.4. RT-PCR, 5′ and 3′ RACE amplifications and sequence analysis

Reverse transcription of *A. grandis* total RNA was done using an oligo d(T)-anchor primer and AMV-RT (Boehringer, Mannheim) according to the manufacturer's protocol. cDNA fragments encoding serine proteinase-like peptides were isolated by RT-PCR with total RNA isolated from second-instar larvae and degenerate primers based in the conserved His_{57} (5′-ACTGCTGCHCAYTG-3′) and Ser_{195} (5′-GGRCCACCAGAGTCRCC-3′) domains found in serine proteinases. Amplification was done in a PTC-100™ programmable thermal controller (MJ Research) using Taq DNA polymerase (Gibco) under the following conditions: 2 min at 94 °C, then 30 cycles of 30 s at 94 °C, 45 s at 45 °C and 1 min at 72 °C plus an extension step for 5 min at 72 °C. To obtain the complete cDNA sequences, the 5′ and 3′ ends were amplified using a 5′/3′ RACE kit (Boehringer, Mannheim)

according to the manufacturer's instructions using specific primers. The 5′ RACE were performed with specific antisense primers derived from the RT-PCR sequences, and the 3′ RACE with specific sense primers derived from the 5′ end of the 5′ RACE sequences. The amplified cDNAs were cloned into the plasmid vectors pCR2.1 (Invitrogen) or pGEM-T Easy (Promega) and recombinant clones were sequenced in both strands in an automated DNA sequencer. Computer analysis of the DNA and amino acid sequences was done using the GCG package (Genetics Computer Group, Inc.), bioinformatics resources of the NCBI homepage (http://www.ncbi.nlm.nih.gov), the EBI website (http://www.ebi.ac.uk/) and MODELLER (Sali and Blundell, 1993). The initial database searches, from whose results tentative functional annotations for the partial sequences were obtained, were carried out using FASTA (Pearson, 1990) in the Swissprot and Trembl databases (Boeckmann et al., 2003). A dendrogram was calculated using the programs PROTDIST and NEIGHBOR of the Phylip package (Felsenstein, 1989). Analysis of signal peptide sequences, including cleavage site localization, was made using SignalP software (Nielsen et al., 1997) at servers of the Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark (http://www.cbs.dtu.dk/services/SignalP/).

2.5. Southern blots

A. grandis genomic DNA was isolated according to Sambrook et al. (1989). DNA digests (15 μg per lane) were separated on a 0.8% agarose gel and transferred to Hybond-N^+ (Amersham) nylon membranes using standard procedures (Sambrook et al., 1989). The entire cDNAs were labeled with [α-^{32}P]dCTP to a high specific activity using the Rediprime DNA-labeling kit (Amersham) and used to probe the blots. The filters were washed with 0.1× SSC at 65 °C.

2.6. Northern blots

Total RNA from *A. grandis* in various developmental stages (neonate larvae, third-instar larvae, midguts from third-instar larvae, pupae, non-feeding adults, 10-day-old adults, and midguts from 10-day-old adults) was prepared using the RNeasy RNA extraction kit (Qiagen). Gel electrophoresis (Fourney et al., 1988) (8 μg per lane) and hybridizations (Sambrook et al., 1989) were done according to standard procedures. The filters were washed with 0.1× SSC at 65 °C.

2.7. Reverse northern

A 100 ng aliquot of a selected clone from each of the 14 cDNA sequences of the PCR fragments was dot blotted into nylon membranes (Hybond-N^+) and hybridized to labeled total cDNA probes prepared from *A. grandis* third-instar whole larvae midguts, and whole larvae feed with the serine proteinase inhibitors SKTI or BTCI. Total RNA was prepared using the RNeasy RNA extraction kit (Qiagen). The cDNA probes were synthesized by reverse transcription of 5 μg of total RNA using MMLV-RT (Invitrogen), 50 μCi [α-^{32}P]dCTP and 0.5 μg of oligo-d$(T)_{16}$ primer. Two replicates of the hybridizations were done; about $12–13 \times 10^6$ cpm of each labeled cDNA was used in the first replicate and $8–10 \times 10^6$ cpm in the second. The normalization of the amount of DNA loaded in the blots was done by hybridizing a replicate filter with the labeled backbone of the vectors carrying the cDNA inserts, and using these data to adjust the values from the hybridizations with the labeled cDNAs. The pGEM-T Easy (Promega) or pCR2.1 (Invitrogen) vectors were labeled with [α-^{32}P]dCTP using the Rediprime DNA-labeling kit (Amersham). Filters were washed with 0.1× SSC at 60 °C. The hybridized and washed membranes were exposed to a Biomax film (Kodak) and the image scanning was performed with a transparency scanner (Duo-scan AG-1200 AGFA). The sample pixel intensity was measured with ZeroDscan program (Scanalytics). The density data were analyzed using a model considering the following sources of variations: the cDNA probes from *A. grandis*, the serine proteinases genes, the interaction between these factors, repetitions (two) and experimental error. Repetition and experimental error were treated as random and the others as fixed effects. The model also included as covariate the density observed within the normalization control experiment. The model was analyzed using the PROC MIXED of SAS (2000).

3. Results

3.1. Proteolytic activity in the gut of A. grandis

Previous results (Purcell et al., 1992; Franco et al., 2004) have indicated that serine proteinase activity predominates in *A. grandis*. To confirm these results, we have measured the serine proteinase activity in the larval and adult midguts of both laboratory reared and field collected insects. Serine proteinase activity was slightly higher in the midgut of larvae when compared to that in the midgut of adult insects (Fig. 1). No significant difference was observed in the proteolytic activity between insects reared either artificially or in natural conditions and all further analysis was then performed with laboratory-reared insects.

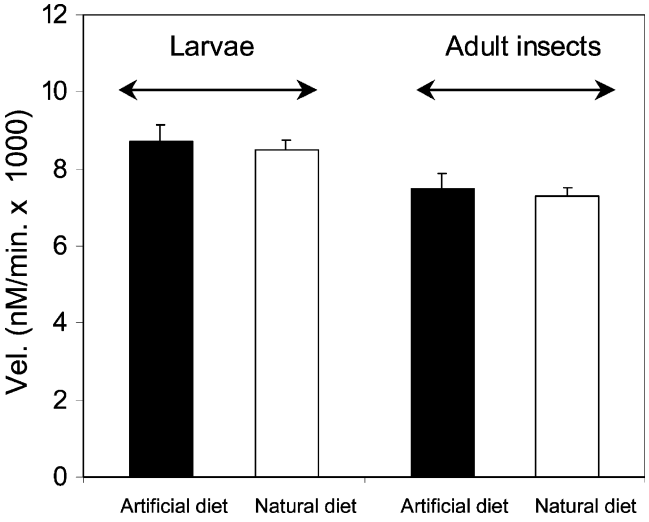

Fig. 1. Serine proteinase activity assays of midgut extracts of third-instar larvae and adults of *A. grandis* fed on artificial (black) and natural (white) diets. Natural diet corresponds to floral buds and bolls in which insects depend on nutrients from pollen grains and ovules.

3.2. Cloning and characterization of a serine proteinase multigene family from A. grandis

The identification of serine proteinase activity in boll weevil prompted us to identify and isolate genes encoding this activity. For that we used a generic PCR approach using degenerate oligonucleotide primers based on conserved segments found in serine proteinases. A similar approach for the isolation of serine proteinases has been successfully applied to other arthropods (Muller et al., 1993; Elvin et al., 1993, 1994; Casu et al., 1994; Gaines et al., 1999). RT-PCR amplification from *A. grandis* whole larvae total RNA, with degenerate primers from conserved regions of amino acids surrounding the active site of the His_{57} and Ser_{195} residues, yielded a diffuse band around 500 bp, which was cloned and sequenced. Sequence analysis of some randomly selected clones yielded products similar to serine proteinases and additional clones were chosen for further sequencing. In total, 93 clones were sequenced, of which 74 encoded protein fragments similar to serine proteinases. One clone encoded a protein similar to mouse palmitoyl-protein thioesterase and three clones encoding elongation factor 1-alpha were found. None of these was investigated further. The remaining 15 clones had no significant similarity to other sequences in the databases. The number of clones sequenced was based on the appearance of new serine proteinase sequences. No novel sequences were detected in the last 20 clones, indicating a saturation of the system, and the sequencing was stopped.

Analysis of the 74 clones encoding protein fragments similar to serine proteinases, varying in size from 142 to 162 amino acids (aa), revealed 14 different sequences (Table 1 and Fig. 2). Each sequence has been named after the number of the first sequenced clone bearing that sequence. Final sequences for each different cDNA were based on the consensus alignment of all clones bearing the same sequence. However, a single or few clones represented some cDNAs and in these cases, the possibility of isolated sequence errors, generated during PCR, cannot be excluded. Of the residues forming the serine proteinase family catalytic triad (His_{57}, Asp_{102} and Ser_{195}), Asp_{102} was present in all sequences. The presence and position of His_{57} and Ser_{195}, however, could not be confirmed since they are found in the regions to which the PCR primers were designed. These regions of identity corresponding to primer

Table 1
Serine proteinase cDNA fragment sequences isolated from *Anthonomus grandis*

Sequence name	Length (aa)	Serine proteinase class	Genebank accession number	% ID with top	Top hit	Top hit code
Agser1p	142	Trypsin-like	AF377979	34.4	Human kallikrein 5 precursor	Q9Y337
Agser2p	151	Trypsin-like	AF377980	40.1	Broad-fingered crayfish trypsin	P00765
Agser5p	151	Chymotrypsin-like	AF377981	41.4	Penaeid shrimp chymotrypsin bi precursor	Q00871
Agser6p	147	Trypsin-like	AF377982	41.6	*Scirpophaga incertulas* putative serine proteinase	O45046
Agser8p	160	Trypsin-like	AF377983	46.3	African malaria mosquito trypsin 2 precursor	P35036
Agser9p	147	Chymotrypsin-like	AF377984	40.1	Penaeid shrimp chymotrypsin bii precursor	O18488
Agser10p	149	Chymotrypsin-like	AF377985	38.6	Fruit fly serine collagenase 1 precursor[a]	Q9VRT2
Agser12p	162	Trypsin-like	AF377986	59.0	Sugarcane rootstalk borer weevil trypsin precursor	O76498
Agser17p	152	Trypsin-like	AF377987	41.1	Fruit fly proclotting enzyme precursor[a]	Q9VZH5
Agser21p	156	Trypsin-like	AF377988	37.2	Yellow fever mosquito trypsin 3a1 precursor	P29786
Agser29p	148	Trypsin-like	AF377989	34.4	Fruit fly trypsin precursor[a]	Q9VXC9
Agser39p	148	Trypsin-like	AF377990	37.0	Cat flea trypsin-like serine protease	Q9XY60
Agser41p	143	Trypsin-like	AF377991	36.3	Fruit fly serine proteinase[a]	Q9VHF7
Agser46p	157	Trypsin-like	AF377992	41.3	Japanese horseshoe crab proclotting enzyme precursor	P21902

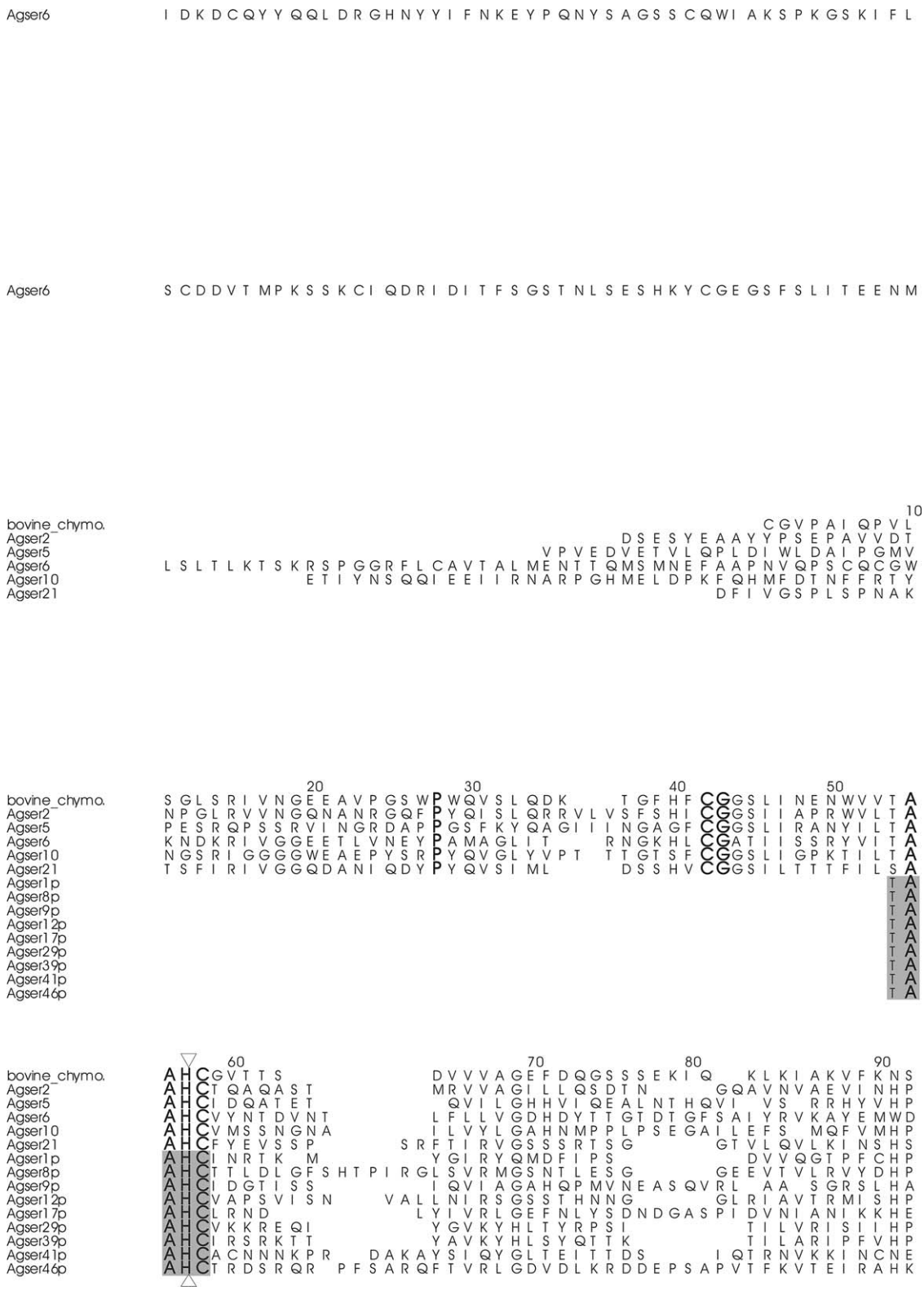

Fig. 2. Sequence alignment of the complete and PCR fragments (*p*) serine proteinase sequences obtained with the model enzyme bovine chymotrypsin. Numbering of bovine chymotrypsin is shown above the alignment. The catalytic triad is marked with arrowheads, the regions of the sequences of the PCR fragments corresponding to the primers used in PCR are shaded, and identical positions emboldened and italicized.

sequences were disregarded in all subsequent analyses. The overall identity between the amino acid sequences derived from the PCR fragments is remarkably low with a mean pairwise identity of 22.1% and just six

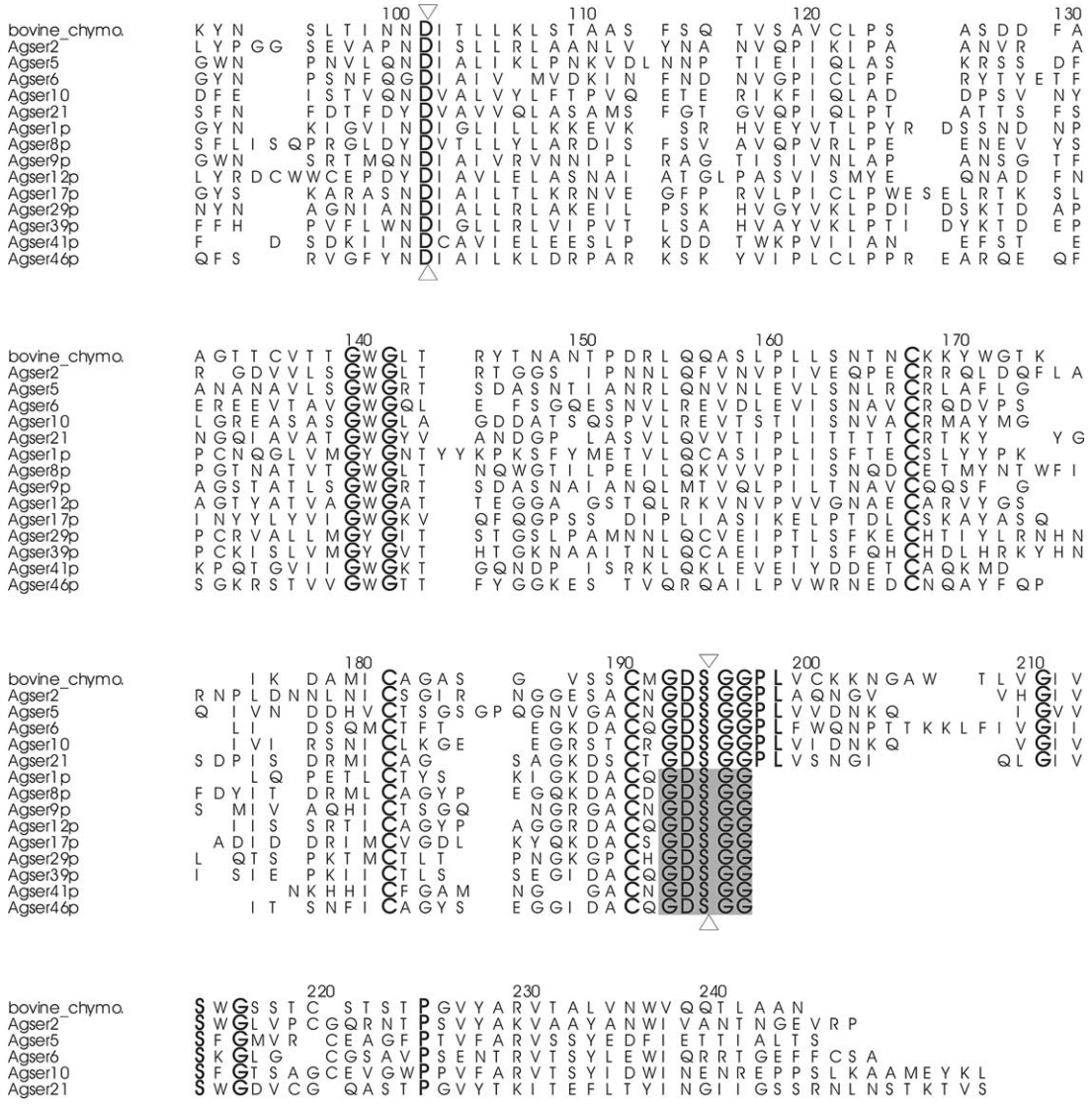

Fig. 2 (*continued*)

amino acid residues are conserved in all sequences. Invariant residues, shown in bold in Fig. 2, comprise the Asp_{102}, already discussed, and cysteine residues, along with two glycine residues presumably conserved for structural reasons.

Comparisons of the deduced amino acid sequences with the Swissprot + Trembl database were carried out. The database sequence most similar to each *A. grandis* sequence of the PCR fragments is shown in Table 1. These sequence comparisons reveal pairwise identities in the range 34–59%. The top hit sequences are mainly from arthropods, including insects, crustacea and chelicerata with the exception of *Agser1p*, which is most similar to human kallikrein.

Tentative functional annotations of the partial sequences as trypsin-like or chymotrypsin-like were obtained from consideration of the top database matches, giving higher weight to annotations in the manually curated Swissprot database. Although this approach may not be completely reliable, particularly since the partial sequences lack most of the C-terminal portion that is most predictive of specificity (Krem et al., 1999), strong independent support for the annotations came from the phylogenetic analysis (Fig. 3). Predicted trypsin-like sequences grouped together while the same was true of predicted chymotrypsin-like sequences. These two approaches enabled the identification of four groups within the set of sequences of the PCR fragments (Fig. 3). The first of these comprises three chymotrypsin-like sequences: *Agser5p*, *Agser9p* and *Agser10p*, which share 20–45% amino acid sequence identity, with 25 residues conserved between

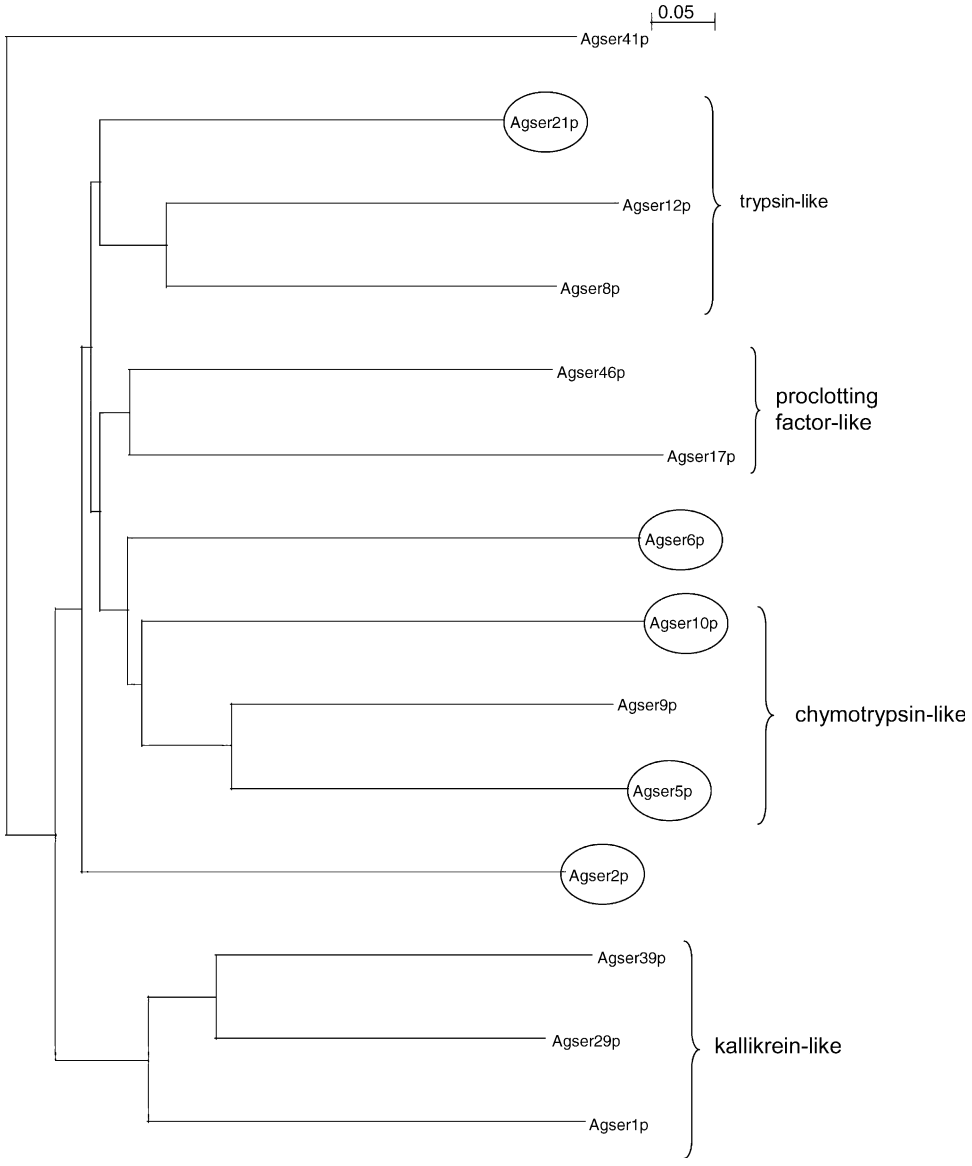

Fig. 3. Dendrogram derived from the predicted amino acid sequence alignment of the *A. grandis* 14 serine proteinase sequences of the PCR fragments. Horizontal distances are proportional to amino acid sequence identity. Groups of related sequences annotated according to the highest score database matches are identified. cDNAs for which the full-length sequence was obtained are circled.

them. A second group comprises *Agser8p*, *Agser12p* and *Agser21p*, which share 27–37% pairwise amino acid sequence identity and 30 identical residues among the three sequences. The three sequences are all similar to trypsins. The highest percentage identity with a database sequence of all amplified clones was obtained for *Agser12p*, 59% identical at the amino acid sequence to a trypsin precursor from *Diaprepes abbreviatus*, the sugarcane rootstalk borer weevil.

A third group contains *Agser17p* and *Agser46p*, two trypsin-like sequences, 29% identical. These sequences encode peptides with similarity to insect proclotting factors. Depending on the parameters used in the alignments, *Agser6p* also grouped here. A fourth group comprised *Agser1p*, *Agser29p* and *Agser39p*, which encode kallikrein-like sequences. This was the group with greatest similarity between the sequences, with 41 identical residues and pairwise sequence identities ranging from 41% to 49%. The two other clones *Agser2p* and *Agser41p* also encoded trypsin like sequences, but they were more divergent and did not group with any of the previous classes.

With the aim of extending the characterization of the isolated cDNAs, some sequences, including usually a representative clone of each class, were selected for the isolation of full-length clones. In total, five cDNAs were completely cloned and sequenced: *Agser2*, *Agser5*, *Agser6*, *Agser10* and *Agser21*. The complete sequences have been named after the corresponding RT-PCR sequences by exclusion of the "*p*".

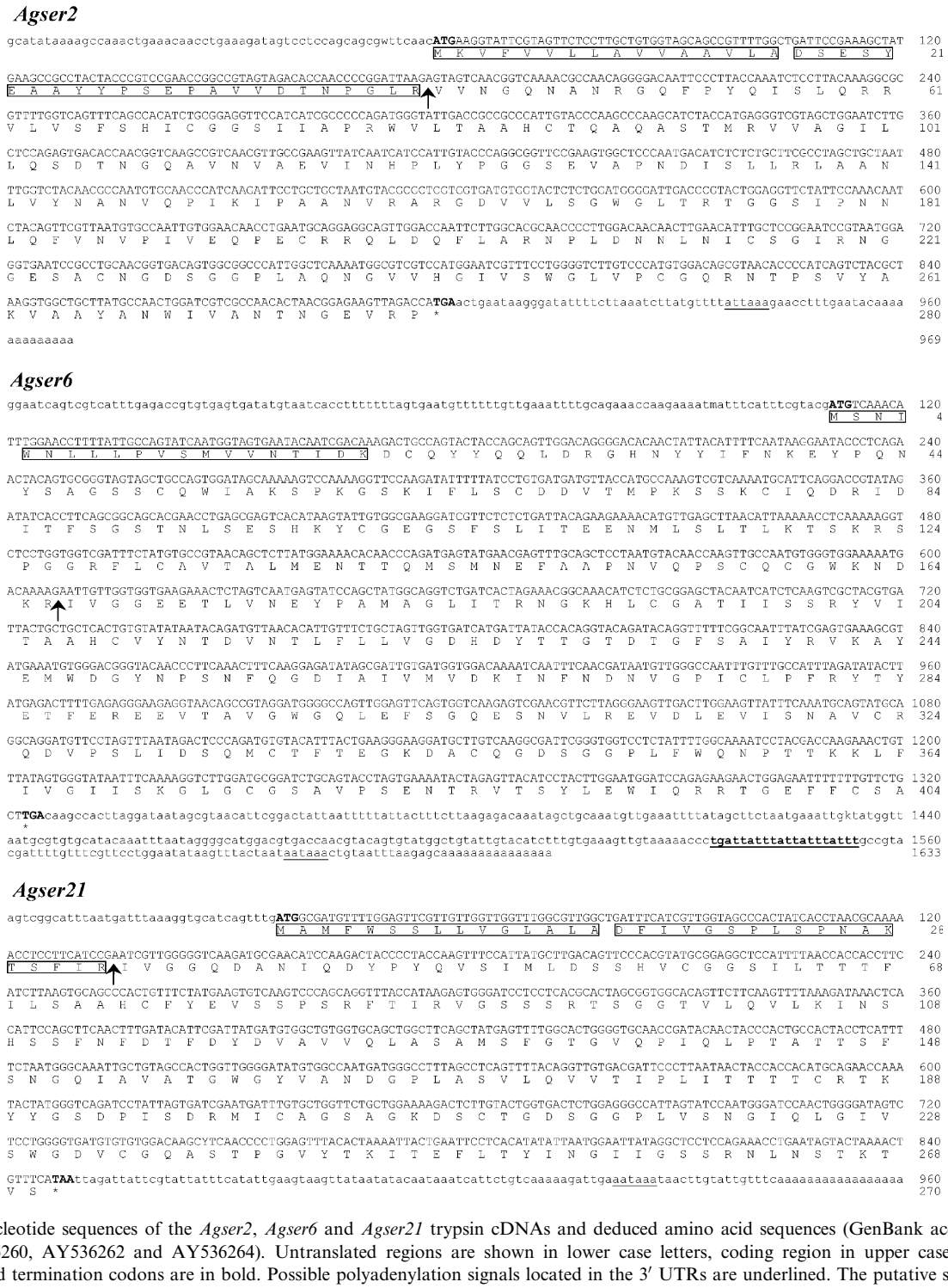

Fig. 4. Nucleotide sequences of the *Agser2*, *Agser6* and *Agser21* trypsin cDNAs and deduced amino acid sequences (GenBank accession numbers: AY536260, AY536262 and AY536264). Untranslated regions are shown in lower case letters, coding region in upper case letters. The initiation and termination codons are in bold. Possible polyadenylation signals located in the 3′ UTRs are underlined. The putative signal peptide and pro-region peptide in the predicted protein sequence are boxed. An arrow indicates the probable activation peptide cleavage site. A putative ARE in the 3′ UTR of *Agser6* is in bold and underlined.

3.3. Cloning of three trypsin genes from A. grandis and analysis of the deduced protein sequences

To obtain the complete sequence of the trypsin encoding cDNAs *Agser2p*, *Agser6p* and *Agser21p*, we first performed 5′ RACE using three antisense primers based on the RT-PCR sequences, followed by a 3′ RACE using two or three sense primers located at the 5′ end of the sequences derived from the 5′ RACE. The complete cDNA sequences and the deduced protein sequences are shown in Fig. 4.

Agser2 and *Agser21* are structurally very similar. The *Agser2* predicted protein sequence included a 16 amino acid signal peptide followed by a short prodomain 24 amino acids in length, producing a final mature protein with 240 amino acids with a calculated molecular weight of 25,450 Da. Prodomains were identified by sequence comparisons with other serine proteinases in which this domain had been previously characterized. In most of the sequences examined, including the ones from the boll weevil, the N-terminal end of the mature enzymes was characterized by the conserved sequence Ile/Val–Val/Ile/Gly–Gly/Asn–Gly, usually preceded by a negatively charged amino acid residue, mainly Arg, at the C-terminal end of the prodomain. Similarly, the *Agser21* sequence encoded a 15 amino acids signal peptide, a 18 amino acids prodomain and a mature protein of 237 amino acids and 24,716 Da. The final mature proteins of both *Agser2* and *Agser21* are composed of a single trypsin domain. The putative short secretion signal peptides, suggests that both enzymes should function as extracellular proteases.

In contrast, *Agser6* encoded a 404 amino acid protein containing an N-terminal region of 166 amino acids prior to the trypsin domain. Sequence comparisons of the N-terminal region revealed the presence of a CUB domain, a domain found mostly among developmentally regulated proteins (Bork and Beckmann, 1993). A similar domain was found in the most similar protein currently present in the sequence databases, a CUB-serine protease from the olfactory organ of the spiny lobster *Panulirus argus* (Levine et al., 2001).

The 3′ untranslated region of *Agser6*, with 296 bp, is significantly longer than the other two cDNAs, which ranged from 59 to 97 bp. Additionally the 3′ UTR of *Agser6* is remarkably rich in A + U residues (68%). Blast search results revealed that a 20 bp AU rich segment of *Agser6* 3′ UTR (bases 1534–1554) was 100% identical with an AU-rich element (ARE) found in the 3′ UTR of vertebrate tumor necrosis factor (TNF-alpha) superfamily genes and which are implicated in the control of mRNA stability and translation. Due to this similarity it is possible that putative regulatory elements may exist in the 3′ UTR of *Agser6*.

3.4. Cloning of two chymotrypsin genes from A. grandis and analysis of the deduced protein sequences

Among the chymotrypsin genes, *Agser5p* and *Agser10p* were selected and their complete sequence was obtained exactly as for the trypsin genes. The *Agser5* cDNA is 971 bp long, and presents an 846 bp open reading frame coding for a 282 amino acid protein (Fig. 5). The predicted protein sequence had a 16 amino acid signal peptide, followed by a 33 amino acid pro region, resulting in a final mature protein of 233 amino acids with a predicted molecular weight of 24,860 Da. The 1017 bp *Agser10* cDNA encoded a 307 amino acid protein, with signal peptides and pro region of 19 and 43 amino acids, respectively, resulting in a final mature protein of 245 residues with a predicted molecular weight of 26,240 Da. Identification of the prodomains was done as for the trypsin sequences. Similarly, the conserved sequence Ile/Val–Val/Ile/Gly–Gly/Asn–Gly was found at the predicted N-terminal of the mature enzymes. The 5′ and 3′ untranslated regions of both cDNAs were composed of short segments, with apparently no significant putative regulatory features beyond the polyadenylation signal.

3.5. Southern blot analysis

To analyze gene copy number, genomic Southern blots were done using two trypsin-like clones, *Agser2* and *Agser5*, and the chymotrypsin-like clones, *Agser6* and *Agser10*, α^{32}P-labelled cDNAs as probes. The results for the hybridization with these clones are shown in Fig. 6. Hybridization of the labeled *Agser2* cDNA probe with *Eco*RI and *Hin*dIII digestions, which do not cut the cDNA, yielded two bands with similar intensities, thus suggesting two copies for this gene. Hybridization of *Agser5* cDNA probe with DNA digested with *Eco*RI and *Xba*I, which do not cut the cDNA, yielded one stronger band and hybridization with *Hin*dIII, which cuts the amplified cDNA once, yielded two bands, suggesting one copy for this gene. Hybridizations with *Agser6* and *Agser10*, were less clear and accurate predictions of gene copy number were more difficult. Hybridizations with the *Agser6* cDNA, which is not cut by any of the restriction enzymes used, yielded one band for the *Eco*RI digestion and two stronger bands and some additional fainter bands for the *Hin*dIII and *Xba*I digestions. These results suggest there is one copy of *Agser6*. Hybridization of *Agser10* cDNA probe with DNA digested with *Hin*dIII and *Xba*I, each which cuts once the amplified cDNA, yielded more than two bands, and hybridization with *Eco*RI, which does not cut the amplified cDNA, yielded two stronger and three fainter bands, suggesting two copies for this gene. Additional fainter bands seen in both *Agser6* and *Agser10* hybridizations could correspond to a weaker hybridization of the probe with other similar genes or internal cuts at introns located close to the cDNA ends. In summary, these results indicate that the cDNAs analyzed are present in a low number of gene copies in the *A. grandis* genome.

Hybridizations of the labeled *A. grandis* complete cDNAs with digested DNA from *Acanthoscelides obtectus* (Coleoptera), *Spodoptera frugiperda* (Lepidoptera) and *Aedes aegypti* (Diptera), did not yield any hybridization band (not shown), further confirming the

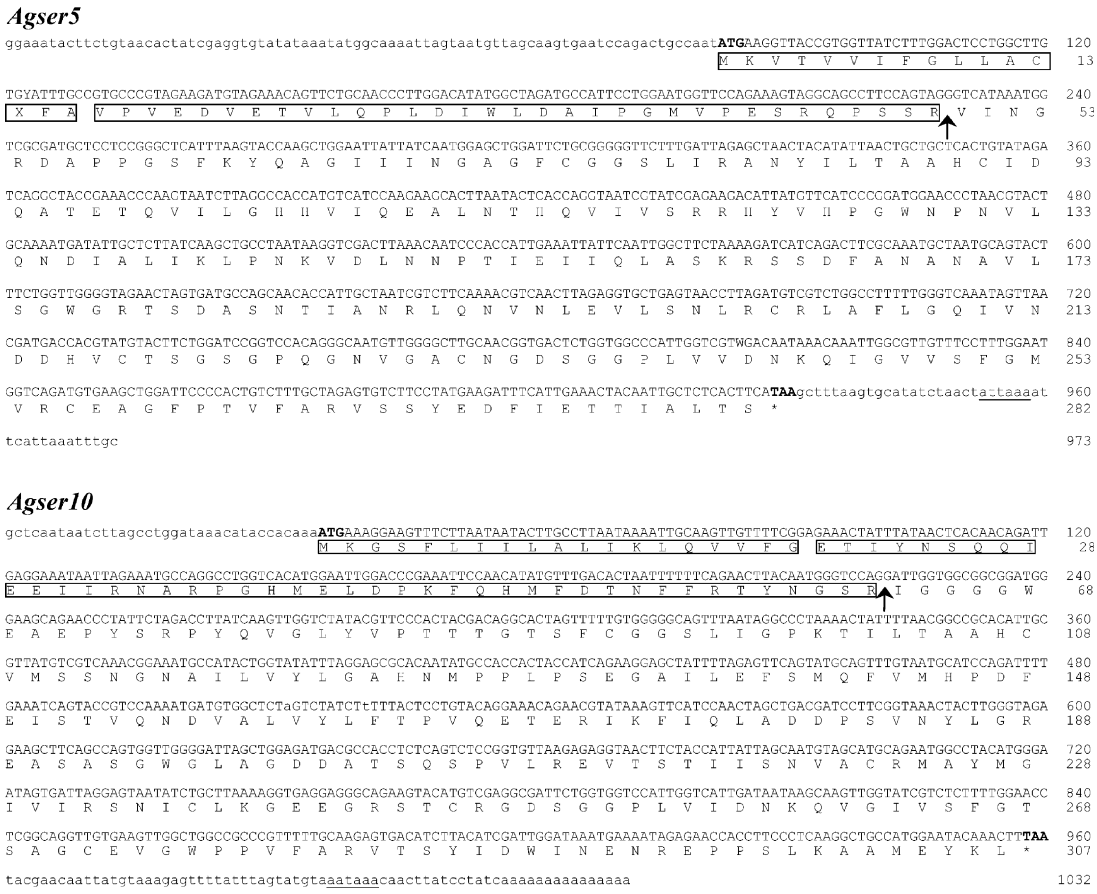

Fig. 5. Nucleotide sequence of the *Agser5* and *Agser10* chymotrypsin cDNAs and deduced amino acid sequences (GenBank accession numbers AY536261 and AY536263). Untranslated regions are shown in lower case letters, coding region in upper case letters. The initiation and termination codons are in bold. Possible polyadenylation signals located in the 3′ UTRs are underlined. The putative signal peptide and pro-region peptide in the predicted protein sequence are boxed. An arrow indicates the probable activation peptide cleavage site.

low similarity of the *A. grandis* serine proteinase sequences with those of other insects.

3.6. Expression analysis of Agser2 and Agser5 along A. grandis developmental stages

To analyze the expression of the cloned cDNAs during the *A. grandis* life cycle, northern blots were performed with total RNA extracted from neonate larvae, third-instar larvae, midguts from third-instar larvae, pupae, non-feeding adults, 10-day-old adults, and midguts from 10-day-old adults. The results are shown in Fig. 7. For both *Agser2* and *Agser5*, a single band of about 1.1 kb was observed, which agrees with the size of the cloned cDNAs. Hybridizations of the filters with the *Agser6* and *Agser10* labeled cDNAs did not yield any bands, even under low stringency washing conditions and long exposure times, indicating a low level of expression of these genes. The expression of *Agser2* was concentrated mainly in the gut from 10-day-old adults, feeding adult insects, gut from third-instar larvae, whole larvae and to a lesser extent in the neonate larvae, and in recently emerged adults. Expression was not detected in pupa. These results indicate that the expression of *Agser2* is induced by feeding and concentrated in the gut of adult insects. A similar trend was observed for *Agser5*, but expression was not detected in neonate larvae.

3.7. Expression of the cloned genes under effect of different serine proteinase inhibitors

The effect of proteinase inhibitors over the expression of insect proteinase genes has been subject of analysis by several studies (McManus and Burgess, 1995; Broadway and Villani, 1995; Bown et al., 1997; Gatehouse et al., 1997; Lara et al., 2000; De Leo et al., 2001; Brito et al., 2001). As one of the aims of this work is to exploit the use of serine proteinase inhibitors as an alternative for the control of the boll weevil, we have analyzed the effect of the serine proteinase inhibitors SKTI and BTCI on the expression of the cloned serine proteinase cDNAs. Selection of the inhibitors was based on previous studies (Franco et al., 2003,

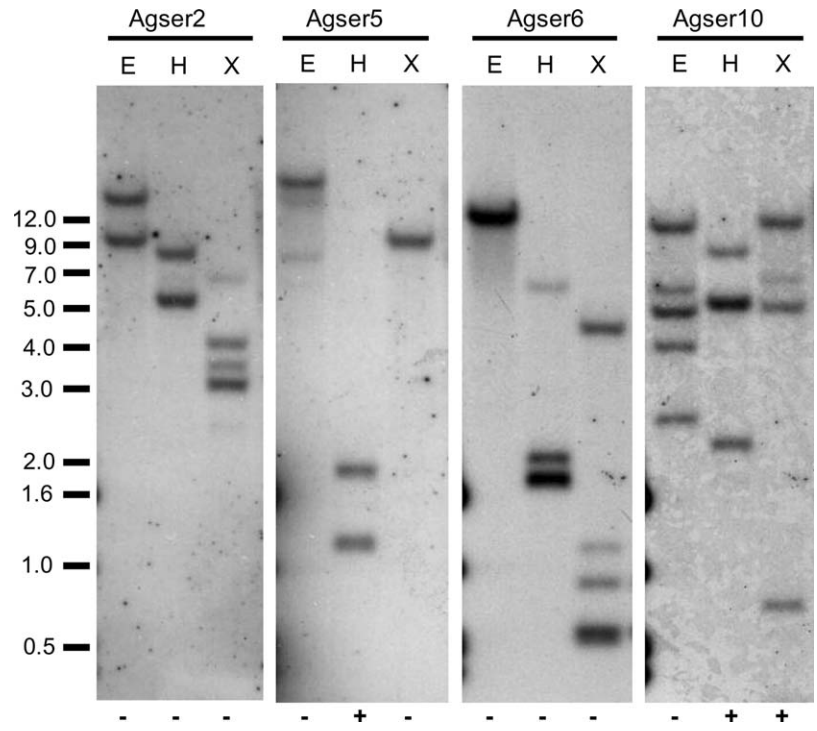

Fig. 6. Southern blot analysis of *A. grandis* genomic DNA digested with *Eco*RI (**E**), *Hin*dIII (**H**) and *Xba*I (**X**), and probed with the indicated ^{32}P-labelled complete cDNA sequences. DNA size markers are indicated in kb. Enzymes that cut each of the cDNAs are indicated by "+" and those that do not cut by "−".

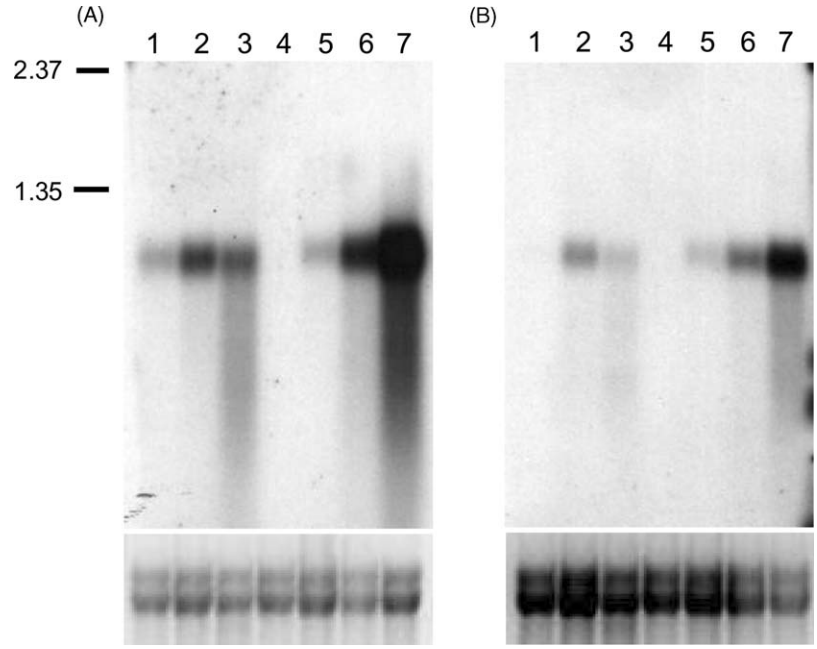

Fig. 7. Northern blot analysis of *Agser2* trypsin-like and *Agser5* chymotrypsin-like gene expression in different *A. grandis* developmental stages. RNA was extracted from: (1) neonate larvae; (2) whole third-instar larvae; (3) third-instar larvae midguts; (4) pupae; (5) recently emerged unfed adults; (6) 10-day-old feeding adults; (7) 10-day-old feeding adults midguts. The ethidium bromide staining of rRNA loaded per lane is shown at the bottom. The differences between the rRNA intensities in the two panels are due to differences in the image integration time for each figure and not to differences in the quantity of RNA loaded per lane in each gel. Size standards are indicated in kb.

Table 2
Reverse northern analysis of the expression of the trypsin-like *Agser2* and *Agser12*, and chymotrypsin-like *Agser5* and *Agser9* cDNAs in third-instar larvae. mRNA levels are represented by the relative intensity of the hybridisation signals of each clone probed with different ^{32}P-labelled total cDNAs. Hybridisation signals were quantified by densitometry and are presented as arbitrary units. Hybridisations were performed with labelled total cDNAs derived from larvae gut, whole larvae, and whole larvae fed on artificial diet added with 150 μM SKTI or 50 μM BTCI

cDNA source	Serine proteinase genes—density means[a] ($CI_{0.95}$)			
	Agser2	*Agser12*	*Agser5*	*Agser9*
Gut larvae	3160 (1309:7627)	5101 (2102:12380)	948 (393:2291)	1946 (799:4737)
Total larvae	1070 (443:2582)	3556 (1465:8630)	1150 (477:2779)	2043 (839:4974)
Larvae feed with SKTI	2398 (993:5788)	2943 (1213:7144)	590 (244:1425)	1005 (413:2447)
Larvae feed with BTCI	1841 (762:4443)	3018 (1244:7326)	487 (202:1177)	793 (326:1931)

[a] Back-transformed means of a log-transformed variable. Means adjusted for density of serine proteinase genes hybridized with vector (two repetitions). 95% confidence intervals are shown.

2004), which showed that these inhibitors are highly effective against the boll weevil.

The expression profile of the 14 serine proteinase cDNA sequences of the PCR fragments was analyzed by reverse northern blots. A representative clone of each cDNA was dot blotted to nylon membranes and hybridized with α^{32}P-labelled cDNA synthesized from total RNA from larvae, larvae midgut and larvae feed with either the serine proteinase inhibitors SKTI or BTCI. The results in the form of the mean densitometry values of the signals in the films are shown in Table 2. Of 14 serine proteinase cDNAs, only four, *Agser2p*, *Agser5p*, *Agser9p* and *Agser12p*, were detected by this analysis. Hybridization signals for the other cDNAs sequences were completely absent or very weak, indicating low levels of expression for these genes, and they were consequently not included in the analysis. These results are in accordance with the northern blots with the *Agser6* and *Agser10* cDNA probes in which the detection of the corresponding transcripts was not possible.

In whole larvae, expression of *Agser2* and *Agser5* was similar, while the expression of *Agser12* and *Agser9* was 3.20 ($CI_{0.95}$ 2.33:4.40) and 1.84 ($CI_{0.95}$ 1.32:2.57) times higher than the average expression of *Agser2* and *Agser5*, respectively. Expression of the trypsin cDNAs *Agser2* and *Agser12* was 2.95 ($CI_{0.95}$ 2.07:4.20) and 1.43 ($CI_{0.95}$ 1.01:2.04) times higher in gut than in the whole larvae, respectively. These results indicate that *Agser2* is predominantly intestinal; while for *Agser12* the results were less conclusive. Expression of the chymotrypsin cDNAs *Agser5* and *Agser9* in the gut did not differ significantly from that observed for the whole larvae, $P = 0.26$ and $P = 0.77$, respectively.

In larvae fed with 150 μM SKTI, the expression of *Agser2* was 2.24 ($CI_{0.95}$ 1.57:3.19) times higher than that observed in larvae fed without the inhibitor, while the expression levels of *Agser5* and *Agser9* were only 0.51 ($CI_{0.95}$ 0.36:0.73) and 0.49 (0.34:0.70), respectively, of that observed for larvae fed without the inhibitor. No significant differences in the expression of *Agser12* ($P = 0.27$) were observed in larvae fed with SKTI in relation to the larvae fed without inhibitors. Furthermore, in larvae fed with 50 μM BTCI, expression of *Agser2* was 1.72 ($CI_{0.95}$ 1.20:2.45) times higher than in larvae grown without the inhibitor, while expression of *Agser5* and *Agser9* was only 0.42 ($CI_{0.95}$ 0.30:0.60) and 0.38 ($CI_{0.95}$ 0.27:0.55), respectively, in relation to larvae fed without the inhibitors. Again, no significant difference in the expression of *Agser12* ($P = 0.33$) was observed in larvae under the effect of BTCI in relation to the larvae fed without inhibitors.

The same hybridizations performed with the complete cDNA sequences of *Agser2* and *Agser5* showed fundamentally the same results, although with higher values due to the longer extension in complementary sequence (results not shown). Hybridizations with the *Agser9* and *Agser12* complete sequences were not possible since these were not available. Hybridizations with the complete cDNA sequences of *Agser6*, *Agser10* and *Agser21*, as for the cDNA sequences of the PCR fragments, did not yield any hybridization signals.

To analyze if the inhibitors were inducing the expression of new serine proteinase genes, RT-PCR amplifications were performed using total RNA from larvae feed with SKTI. In total, 68 clones encoding serine proteinase cDNAs were sequenced. Sequence analysis did not reveal the presence of any novel clone different from the 14 cDNAs previously amplified (results not shown). In fact, upon inclusion of the inhibitor in the diet, diversity in the number of different sequences decreased, with just eight of the previously isolated 14 cDNAs being amplified.

4. Discussion

Early results suggested that cysteine proteinases were predominant in Coleoptera, while serine proteinases predominated in Lepidoptera (Murdock et al., 1987; Purcell et al., 1992). However, these results were based on a very small sampling and more recent results with

a broader sampling within Coleoptera, the largest order among insects, have shown that the use of different systems for midgut proteolytic digestion is extremely diverse among the order. These data indicate that cysteine proteinases are absent in the more primitive coleopterans and also in a few more recent lineages including some Cerambycidae and Curculionidae (Johnson and Rabosky, 2000). The results reported herein on both the proteolytic activity and cDNAs cloning further support previous results (Purcell et al., 1992; Franco et al., 2004), which indicate that serine proteinases are the predominant digestive enzymes in the boll weevil. In at least one other phytophagous coleopteran, *Rhyzopertha dominica*, serine proteinases are also predominant (Zhu and Baker, 1999).

Some authors have hypothesized that cysteine proteinases appeared in beetle lineages in response to dietary serine proteinase inhibitors commonly found in plant seeds and some other tissues (Murdock et al., 1987; Ryan, 1990). If this assumption is correct, it would explain why the boll weevil has predominantly serine proteinases, as it feeds in other parts of the plant, pollen grains and young ovules, that usually do not contain high levels of serine proteinase inhibitors.

Similar to many other arthropods, a multigene family encodes serine proteinase activity in *A. grandis*. However, to our knowledge, this is the first report of the characterization of a serine proteinase multigene family in a coleopteran species. Among insects, previous works have reported multi-gene families primarily for dipteran species, comprising the fruit fly *Drosophila melanogaster* (Davis et al., 1985), *Aedes aegypti* (Graf and Briegel, 1989), *Anopheles gambiae* (Muller et al., 1993), *Haematobia irritans exigua* (Elvin et al., 1993), *Lucilia curprina* (Casu et al., 1994), but also for the lepidopteran *Helicoverpa armigera* (Bown et al., 1997), cat flea *Ctenocephalides felis* (Gaines et al., 1999) and the orthopteran *Locusta migratoria* (Lam et al., 2000). In *L. cuprina* (Diptera) the number of serine proteinase genes has been estimated between 125 and 220 (Elvin et al., 1994). In the *D. melanogaster* genome, serine proteinases and their homologs constitute the second largest protein family, comprising 199 trypsin-like and 178 chymotrypsin-like proteins (Rubin et al., 2000). In our study, 14 different cDNAs were isolated by RT-PCR from total RNA from *A. grandis* second-instar larvae. Although sequencing proceeded until saturation of the system, we cannot exclude the possibility that the number of serine proteinase encoding genes in *A. grandis* may be higher. It could be that some other genes did not match well the primers used in the RT-PCR, were not expressed or were expressed at very low levels at the particular stage used for the RNA extraction.

The function of these gene families is not fully understood. Multigene families may have evolved to provide a more efficient mechanism for protein digestion as well as to provide an adaptive advantage for phytophagous species feeding on plants that contain proteinase inhibitors (Bown et al., 1997; Reeck et al., 1997). Multigene families may allow the induction of proteinases that are insensitive to dietary inhibitors, as noted by Jongsma et al. (1995), Broadway (1995) and Bolter and Jongsma (1995), or they may initiate proteolysis of proteinase inhibitors by non-target digestive proteinases (Michaud et al., 1995). Of course, proteinases have other functions beyond midgut proteolytic digestion and have been implicated in many other cellular processes such as development and apoptosis. Indeed, the similarity of some of the amplified cDNAs with proclotting factors and kallikrein, along with the presence of a CUB domain in one, suggests they are involved in processes other than proteolytic digestion in the gut.

The protein structure formed solely by a single serine proteinase domain for members of the first two groups (trypsin- and chymotrypsin-like in Fig. 3), along with the results that indicate that the expression of *Agser2* and *Agser5* is induced by feeding and concentrated in the gut of adult insects, supports the idea that they represent chymotrypsin- and trypsin-like digestive enzymes, since digestive serine proteinases typically have a short prodomain followed by catalytic domain 220–230 residues long. In contrast, the third group, showing high sequence similarity to blood clotting enzymes, may tentatively be assigned a role in haemolymph clotting.

The low level of identity of the *A. grandis* isolated sequences with the best serine proteinase matches probably reflects the small number of sequences from other coleopteran species available in the databases. The tentative annotation of the sequences derived from the PCR fragments as trypsin- or chymotrypsin-like in Table 1 is based primarily on whether trypsins or chymotrypsins were predominant in the high-scoring database matches. In order to improve confidence in the annotations of the database matches, we used exclusively the curated Swissprot database and its computer annotated companion database Trembl, giving higher weight to annotations from the former. With the exception of *Agser2p*, *Agser29p* and *Agser41p* the annotations are also consistent with the presence at position 189 of Asp in trypsins or Ser/Gly in chymotrypsins (Czapinska and Otlewski, 1999). Detailed functional assignment is made more difficult by the lack of the C-terminal region in the sequences derived from the PCR fragments. These residues 189–220 form most of the proteinase specificity sites and are necessary and sufficient to reliably assign function (Krem et al., 1999). Nevertheless, our tentative functional annotations as trypsin-like or chymotrypsin-like are supported by the grouping together of our sequences in the dendrogram;

predicted trypsin-like proteinases group together and predicted chymotrypsin-like proteinases group together.

As one of the main goals of this work is to explore the viability of using proteinase inhibitors for engineering resistance against *A. grandis* in transgenic cotton, we, therefore, determined the effect of the dietary inclusion of the trypsin serine proteinase inhibitors SKTI and BTCI on the expression of serine proteinases genes in *A. grandis*. In the overall, the inhibitors negatively affected expression of the chymotrypsin genes *Agser5* and *Agser9* that were not preferentially detected in the gut of larvae. In contrast, expression of the trypsin gene *Agser2*, that is preferentially intestinal, increased under the effect of the inhibitors. The effect of the two inhibitors was similar, but for BTCI the reduction in the expression of *Agser5* and *Agser9* was higher and the increase in the expression of *Agser2*, smaller.

In cotton boll weevil larvae fed with 150 μM SKTI, Franco et al. (2004) observed a reduction in weight of around 30%. It is possible that the small decrease observed in the expression of the serine proteinase genes may account for this result. At higher doses of the inhibitor of 500 μM, higher rates of weight reduction, mortality and deformity were observed, but after feeding at this higher concentration, surviving larvae were few and insufficient for the expression studies.

These results are in agreement with those of Bown et al. (1997) and others (McManus and Burgess, 1995; Gatehouse et al., 1997; De Leo et al., 2001), who found that the response to a protein inhibitor involved alteration of the relative expression of genes already being expressed rather than the expression of a new set of genes. However, the balance in the expression of trypsin and chymotrypsin genes upon the effect of the inhibitor seems to be variable. Similar to our results, McManus and Burgess (1995) have shown that SKTI stimulates the trypsin-like activity found in the gut of *Spodoptera litura*. On the other hand, Bown et al. (1997) and Gatehouse et al. (1997) have reported that trypsin mRNA levels decreased and chymotrypsin mRNA was increased in *H. armigera* larvae feed with SKTI. Other studies have described that insects counteract the effect of proteinase inhibitors with the synthesis of novel proteinases, which are inhibitor insensitive (Broadway and Villani, 1995; Lara et al., 2000; Brito et al., 2001). This is apparently not the case for the boll weevil when fed with the serine proteinase inhibitors SKTI and BTCI, although the new balance of expression observed in the presence of inhibitors may favor those proteinases less affected by them; as yet we have no information on the individual inhibition properties of our gene clones. Although we cannot completely exclude the possibility that these inhibitors may induce the expression of new proteinases different from the ones observed by us, amplifications of total RNA by RT-PCR from larvae fed with SKTI did not yield any different clone beyond the 14 cDNAs previously amplified. It is also possible that the boll weevil may respond to the serine proteinase inhibitors by increasing the expression level of other proteinase classes, but this possibility is still to be tested.

The fact that expression of eight of the cloned *A. grandis* cDNAs could not be detected in the reverse Northern analysis suggests that they are most probably expressed at very low levels. Some like *Agser6* and *Agser10* were not detected even on the northern blots. Due to the lower sensitivity of the reverse northern, particularly with the use of total RNA for labeling of the cDNAs, the detection of only some of the genes is not unexpected. The results of the northern blots are more surprising since this technique has greater sensitivity and covered all the developmental cycle of the insect. However, low expression of *Agser6* would obviously be consistent with the developmental role already strongly indicated by its CUB domain (Bork and Beckmann, 1993) and the putative regulatory elements found in its 3′ untranslated region. It is probable that other more sensitive and reliable techniques such as quantitative real-time PCR should be able to overcome the problem of detection of some of the boll weevil serine proteinase genes, apparently due to the low level of expression of these genes.

The isolation of genes involved in digestion is essential to understand how their expression is controlled at the molecular level. In this context, species that have a multigene family expressing serine proteinase are important models to understand the response of these genes' expression to the presence of proteinase inhibitors. Our results revealed that the majority of serine proteinases amplified from *A. grandis* are similar to trypsins and chymotrypsins and are most likely involved in extracellular digestion. Both types of serine proteinases may prove to be potential targets for engineering resistance against boll weevil.

Acknowledgements

The authors thank Dr. Sonia Maria Freitas for kindly providing a sample of BTCI inhibitor. This work was supported by grants from the Brazilian government (EMBRAPA and CNPq), FACUAL and FIALGO.

References

Applebaum, S.W., 1985. Biochemistry of digestion. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comparative Physiology Biochemistry and Pharmacology of Insects. Pergamon Press, Oxford, pp. 279–311.

Barrett, A.J., Rawlings, N.D., 1995. Families and clans of serine peptidases. Arch. Biochem. Biophys. 318, 247–250.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365–370.

Bolter, C.J., Jongsma, M.A., 1995. Colorado potato beetles adapt to proteinase inhibitors induced in potato leaves by methyl jasmonate. J. Insect Physiol. 41, 1071–1078.

Bork, P., Beckmann, G., 1993. The CUB domain. A widespread module in developmentally regulated proteins. J. Mol. Biol. 231, 539–545.

Borovsky, D., Schlein, Y., 1988. Quantitatve determination of trypsin-like and chymotrypsin like enzymes in insects. Arch. Insect Biochem. Physiol. 8, 249–260.

Bown, D.P., Wilkinson, H.S., Gatehouse, J.A., 1997. Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. Insect Biochem. Mol. Biol. 27, 625–638.

Bradford, M.M., 1976. A rapid and sensitive method from quantitation of microgram quantities of protein utilizing the principle of dye binding. Anal. Biochem. 72, 248–254.

Brito, L.O., Lopes, A.R., Parra, J.R.P., Terra, W.R., Silva-Filho, M.C., 2001. Adaptation of tobacco budworm *Heliothis virescens* to proteinase inhibitors may be mediated by the synthesis of new proteinases. Comp. Biochem. Physiol. B 128, 365–375.

Broadway, R.M., 1995. Are insects resistant to plant proteinase inhibitors? J. Insect Physiol. 41, 107–116.

Broadway, R.M., Villani, M.G., 1995. Does host range influence susceptibility of herbivorous insects to non-host plant proteinase inhibitors? Entomol. Exp. Appl. 76, 303–312.

Broadway, R.M., Duffey, S.S., Pearce, G., Ryan, C.A., 1986. Plant proteinase inhibitors: a defense against herbivorous insects? Entomol. Exp. Appl. 41, 33–38.

Casu, R.E., Pearson, R.D., Jarmey, J.M., 1994. Excretory/secretory chymotrypsin from *Lucilia cuprina*: purification, enzymatic specificity and amino acid sequence from mRNA. Insect Mol. Biol. 3, 201–211.

Christeller, J.T., Burgess, E.P., Mett, V., Gatehouse, H.S., Markwick, N.P., Murray, C., Malone, L.A., Wright, M.A., Philip, B.A., Watt, D., Gatehouse, L.N., Lovei, G.L., Shannon, A.L., Phung, M.M., Watson, L.M., Laing, W.A., 2002. The expression of a mammalian proteinase inhibitor, bovine spleen trypsin inhibitor in tobacco and its effects on *Helicoverpa armigera* larvae. Transgenic Res. 11, 161–173.

Cloutier, C., Jean, C., Fournier, M., Yelle, S., Michaud, D., 2000. Adult Colorado potato beetles, *Leptinotarsa decemlineata* compensate for nutritional stress on oryzacystatin I-transgenic potato plants by hypertrophic behavior and over-production of insensitive proteases. Arch. Insect Biochem. Physiol. 44, 69–81.

Czapinska, H., Otlewski, J., 1999. Structural and energetic determinants of the S1-site specificity in serine proteases. Eur. J. Biochem. 260, 571–595.

Davis, C.A., Riddell, D.C., Higgins, M.J., Holden, J.J., White, B.N., 1985. A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. Nucleic Acids Res. 13, 6605–6619.

De Leo, F., Bonade-Bottino, M., Ceci, L.R., Gallerani, R., Jouanin, L., 2001. Effects of a mustard trypsin inhibitor expressed in different plants on three lepidopteran pests. Insect Biochem. Mol. Biol. 31, 593–602.

Duan, X., Li, X., Xue, Q., Abo-el-Saad, M., Xu, D., Wu, R., 1996. Transgenic rice plants harboring an introduced potato proteinase inhibitor II gene are insect resistant. Nat. Biotechnol. 14, 494–498.

Elvin, C.M., Whan, V., Riddles, P.W., 1993. A family of serine proteases genes expressed in adult buffalo fly (*Haematobia irritans exigua*). Mol. Gen. Genet. 240, 132–139.

Elvin, C.M., Vuocolo, T., Smith, W.J.M., Eisemann, C.H., Riddles, P.W., 1994. An estimate of the number of serine proteases genes expressed in sheep blowfly larvae (*Lucilia cuprina*). Insect Mol. Biol. 3, 105–115.

Fabrick, J., Behnke, C., Czapla, T., Bala, K., Rao, A.G., Kramer, K.J., Reeck, G.R., 2002. Effects of a potato cysteine proteinase inhibitor on midgut proteolytic enzyme activity and growth of the southern corn rootworm, *Diabrotica undecimpunctata howardi* (Coleoptera: Chrysomelidae). Insect Biochem. Mol. Biol. 32, 405–415.

Felsenstein, J., 1989. PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Fourney, R.M., Miyakoshi, J., Day, III., R.S., Paterson, M.C., 1988. Northern blotting: efficient RNA staining and transfer. Focus 10, 5–7.

Franco, O.L., Santos, R.C., Batista, J.A.N., Mendes, A.C.M., Araújo, M.A.M., Monnerat, R.G., Grossi de Sá, M.F., Freitas, S.M., 2003. Effects of black-eyed pea trypsin/chymotrypsin inhibitor on proteolytic activity and on development of *Anthonomus grandis*. Phytochemistry 63, 343–349.

Franco, O.L., Dias, S.C., Magalhães, C.P., Monteiro, A.C.S., Bloch Jr., ., C., Melo, F.R., Oliveira-Neto, O.B., Monnerat, R.G., Grossi de Sá, M.F., 2004. Effects of soybean Kunitz trypsin inhibitor on cotton boll weevil (*Anthonomus grandis*). Phytochemistry 65, 81–89.

Freitas, S.M., Ikemoto, H., Ventura, M.M., 1999. Thermodynamics of the binding of chymotrypsin with the black-eyed pea trypsin and chymotrypsin inhibitor (BTCI). J. Protein Chem. 18, 307–313.

Gaines, P.J., Sampson, C.M., Rushlow, K.E., Stiegler, G.L., 1999. Cloning of a family of serine protease genes from the cat flea *Ctenocephalides felis*. Insect Mol. Biol. 8, 11–22.

Gatehouse, L.N., Shannon, A.L., Burgess, E.P., Christeller, J.T., 1997. Characterization of major midgut proteinase cDNAs from *Helicoverpa armigera* larvae and changes in gene expression in response to four proteinase inhibitors in the diet. Insect Biochem. Mol. Biol. 27, 929–944.

Gatehouse, A.M.R., Norton, E., Davison, G.M., Babbé, S.M., Newell, C.A., Gatehouse, J.A., 1999. Digestive proteolytic activity in larvae of tomato moth, *Lacanobia oleraceae*; effects of plant protease inhibitors in vitro and in vivo. J. Insect Physiol. 45, 545–558.

Graf, R., Briegel, H., 1989. The synthetic pathway of trypsin in the mosquito *Aedes aegypti* L. (Diptera: Culicidae) and in vitro stimulation in isolated midguts. Insect Biochem. 19, 129–137.

Hilder, V.A., Gatehouse, A.M.R., Sheerman, S.E., Barker, R.F., Boulter, D., 1987. A novel mechanism of insect resistance engineered into tobacco. Nature 330, 160–163.

Jiang, Q., Hall, M., Noriega, F.G., Wells, M., 1997. cDNA cloning and pattern of expression of an adult, female-specific chymotrypsin from *Aedes aegypti* midgut. Insect Biochem. Mol. Biol. 27, 283–289.

Johnson, K.S., Rabosky, D., 2000. Phylogenetic distribution of cysteine proteinases in beetles: evidence for an evolutionary shift to an alkaline digestive strategy in Cerambycidae. Comp. Biochem. Physiol. B 126, 609–619.

Johnston, K.A., Gatehouse, J.A., Anstee, J.H., 1993. Effects of soybean protease inhibitors on the growth and development of larval *Helicoverpa armigera*. J. Insect Physiol. 39, 657–664.

Jongsma, M.A., Bakker, P.L., Peters, J., Bosch, D., Stiekma, W.J., 1995. Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. Proc. Natl. Acad. Sci. 92, 8041–8045.

Krem, M.M., Rose, T., Di Cera, E., 1999. The C-terminal sequence encodes function in serine proteases. J. Biol. Chem. 274, 28063–28066.

Lam, W., Coast, G.M., Rayne, R., 2000. Characterization of multiple trypsins from the midgut of *Locusta migratoria*. Insect Biochem. Mol. Biol. 30, 85–94.

Lara, P., Ortego, F., Gonzalez-Hidalgo, E., Castañera, P., Carbonero, P., Diaz, I., 2000. Adaptation of *Spodoptera exigua* (Lepidoptera: Noctuidae) to barley trypsin inhibitor BTI-CMe expressed in transgenic tobacco. Transgenic Res. 9, 169–178.

Levine, M.Z., Walthall, W.W., Tai, P.C., Derby, C.D.J., 2001. A CUB-serine protease in the olfactory organ of the spiny lobster *Panulirus argus*. J. Neurobiol. 49, 277–302.

McManus, M.T., Burgess, E.P.J., 1995. Effects of the soybean (Kunitz) trypsin inhibitor on the growth and digestive proteases of larvae of *Spodoptera litura*. J. Insect Physiol. 41, 731–738.

Michaud, D., Bernier-Vadnais, N., Overney, S., Yelle, S., 1995. Constitutive expression of digestive cysteine proteinase forms during development of the Colorado potato beetle, *Leptinotarsa decemlineata* Say (Coleoptera: Chrysomelidae). Insect Biochem. Mol. Biol. 25, 1041–1048.

Monnerat, R.G., Dias, S.C., Oliveira-Neto, O.B., Nobre, S.D., Silva-Werneck, J.O., Grossi-de-Sá, M. F., 2000. Criação massal do bicudo-do-algodoeiro *Anthonomus grandis* em laboratório. Comunicado Técnico, Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brasil, 4 pp. (ISSN 0102-0099).

Muller, H.M., Crampton, J.M., Torre, A.D., Sinden, R., Crisanti, A., 1993. Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal. EMBO J. 12, 2891–2900.

Murdock, L.L., Brookhart, G., Dunn, P.E., Foard, D.E., Kelley, S., Kitch, L., Shade, E.R., Shukle, R.H., Wolfson, J.L., 1987. Cysteine digestive proteinases in Coleoptera. Comp. Biochem. Physiol. B 87, 783–787.

Nagano, I., Wu, Z., Nakada, T., Bonmans, T., Takahashi, Y., 2003. Molecular cloning and characterization of a sonine proteinase gene of *Trichinella spinallis*. J. Parasitol. 89, 92–98.

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.

Paulillo, L.C., Lopes, A.R., Cristofoletti, P.T., Parra, J.R., Terra, W.R., Silva-Filho, M.C., 2000. Changes in midgut endopeptidase activity of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) are responsible for adaptation to soybean proteinase inhibitors. J. Econ. Entomol. 93, 892–896.

Pearson, W.R., 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183, 63–98.

Pompermayer, P., Lopes, A.R., Terra, W.R., Parra, J.R.P., Falco, M.C., Silva, M.C., 2001. Effects of soybean proteinase inhibitor on development, survival and reproductive potential of the sugarcane borer, *Diatraea saccharalis*. Entomol. Exp. Appl. 99, 79–85.

Purcell, J.P., Greenplate, J.T., Sammons, R.D., 1992. Examination of midgut luminal proteinase activities in six economically important insects. Insect Biochem. Mol. Biol. 22, 41–47.

Rao, M.B., Tanksale, A.M., Ghatge, M.S., Deshpande, V.V., 1998. Molecular and biotechnological aspects of microbial proteases. Microbiol. Mol. Biol. Rev. 62, 597–635.

Reeck, G.R., Kramer, K.J., Baker, J.E., Kanost, M.R., Fabrick, J.A., Behnke, C.A., 1997. Proteinase inhibitors and resistance of transgenic plants to insects. In: Carozzi, N., Koziel, M.G. (Eds.), Advances in Insect Control: The Role of Transgenic Plants. Taylor and Francis, London, pp. 157–183.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., et al., 2000. Comparative genomics of the eukaryotes. Science 287, 2204–2215.

Ryan, C.A., 1990. Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Ann. Rev. Phytopathol. 28, 425–440.

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, New York.

Sarath, G., De la Motte, R.S., Wagner, F.W., 1990. Protease assay methods. In: Beynon, R.J., Bond, J.S. (Eds.), Proteolytic Enzymes: A Practical Approach. IRL Press, Oxford, pp. 25–55.

SAS Institute Inc., 2000. SAS/STAT®, User's Guide. Version 8, SAS Instiute Inc., Cary, NC.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalisation and function. Comp. Biochem. Physiol. B 109, 1–62.

Thomas, J.C., Adams, D.G., Keppenne, V.D., Wasmann, C.C., Brown, J.K., Kanost, M.R., Bohnert, H.J., 1995. Protease inhibitors of *Manduca sexta* expressed in transgenic cotton. Plant Cell Rep. 14, 758–762.

Zhu, Y.C., Baker, J.E., 1999. Characterization of midgut trypsin-like enzymes and three trypsinogen cDNA's from the lesser grain borer, *Rhyzopertha dominica* (Coleoptera: Bostrichidae). Insect Biochem. Mol. Biol. 29, 1053–1063.