January 2005

Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins

Blair D. Siegfried
University of Nebraska-Lincoln, bsiegfried1@ufl.edu

N. Waterfield
University of Bath, Bath, UK

R. H. ffrench-Constant
University of Bath, Bath, UK

Published in *Insect Molecular Biology* **14:**2 (2005), pp. 137–143; doi: 10.1111/j.1365-2583.2004.00538.x

Submitted July 6, 2003; accepted after revisions October 6, 2004.

Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins

B. D. Siegfried*, N. Waterfield§,

and R. H. ffrench-Constant§

* *Department of Entomology, University of Nebraska-Lincoln, Lincoln, NE, USA*

§ *Department of Biology and Biochemistry, University of Bath, Bath, UK*

Corresponding author: Dr. Blair Siegfried, Department of Entomology, 202 Plant Industry Building, University of Nebraska–Lincoln, Lincoln, NE, USA.Tel 402 472-8714; fax 402 472-4687; Email: bsiegfried1@unl.edu

Abstract

The Western corn rootworm is the major pest of corn in the USA and has recently become the target for insect-resistant transgenic crops. Transgenic crops have switched the focus for identifying insecticide targets from the insect nervous system to the midgut. Here we describe a collection of 691 sequences from the Western corn rootworm midgut, 27% of which predict proteins with no matches in current databases. Of the remaining sequences, most predict proteins with either catalytic (62%) or binding (19%) functions, as expected for proteins expressed in the insect midgut. The utility of this approach for the identification of targets for novel toxins is demonstrated by analysis of the first coleopteran cadherin gene, a putative Bt receptor, and a large class of cysteine-proteases, the cathepsins.

Keywords

ESTs, *Diabrotica*, insect midgut, cadherin, Bt receptor, cathepsin.

Introduction

The Western corn rootworm, *Diabrotica virgifera vigifera* LeConte (Coleoptera: Chrysomelidae), is the most economically important pest of field corn, *Zea mays* (L.) in the United States (Levine & Oloumi-Sadeghi, 1991). Recently, *D. v. virgifera* was also introduced into Eastern Europe and has spread rapidly, infesting more than 70,000 square miles in central Europe (Hummel, 2003). Rootworms feed on corn roots, and damaged plants are more susceptible to drought and disease, have decreased yield, and are prone to lodging. Crop rotation and chemical control have been the primary management strategies (Levine & Oloumi-Sadeghi, 1991), although *D. v. virgifera* has become increasingly difficult to control because of its sequential ability to evolve resistance to different classes of small molecule insecticides (Ball & Weekman, 1963; Meinke *et al*., 1998; Zhou *et al*., 2002). In areas where crop rotation has been the primary management strategy, rootworms have also evolved a “behavioural” resistance involving oviposition in non-host crops (O’Neal *et al*., 2001; Levine *et al*., 2002). Those eggs deposited outside cornfields can cause significant damage to corn planted in those fields in the following year. Soil insecticides such as organophosphates and pyrethroids are still effective but pose serious environmental and human health risks, and are therefore unlikely to provide viable and long-term management options.

Current alternative management technologies include the transgenic crops that express insecticidal proteins toxic to corn rootworm larvae (Moellenbeck *et al*., 2001; Ellis *et al*., 2002).The first transgenic corn hybrids expressing δ-endotoxins from *Bacillus thuringiensis* (Bt) were commercially available in 2003 (Environmental Protection Agency, 2003). This technology represents a significant departure from traditional crop protection chemistry, because Bt toxins act on receptors in the insect midgut rather than the insect nervous system. Although this approach offers a potentially important new direction for rootworm management, there are concerns that widespread adoption of Bt transgenic plants will rapidly lead to δ-endotoxin resistance. As a consequence, receptors associated with the insect gut have become an increasingly import target for development of novel toxins including protease inhibitors (Koiwa *et al*., 2000), the toxin complexes (Tc’s) from *Photorhabdus luminescens* (Liu *et al*., 2003) and other toxins from *B. thuringiensis* (Moelleneck *et al*., 2001; Ellis *et al*., 2002; Loguercio *et al*., 2002).

We are interested in identifying novel receptors and enzymes within the *Diabrotica* midgut as potential targets for Bt and alternative novel toxins (Gaines *et al*., 2002).

Here we describe an expressed sequence tag (EST) strategy (Hill & Guitierrez, 2000) for the identification of candidate receptors in the *Diabrotica* midgut. We discuss the relationship of these ESTs to known sequences from insects and other animals (Gaines *et al*., 2002) as well as their potential biological and molecular functions. The utility of this approach is illustrated by the isolation of the first coleopteran cadherin gene, encoding a putative Bt receptor (Ferré & Van Rie, 2002; Morin *et al*., 2003) and also a description of the diversity of cathepsins (cysteine-proteases) expressed in the rootworm midgut, which have been identified as potential targets of protease inhibitors (Koiwa *et al*., 2000).

Results

Diabrotica *midgut ESTs*

The cDNA library from mRNA extracted from dissected *D. v. virgifera* larval midguts contained 2.53 × 10^{10} primary clones. We sequenced 2880 clones, of which 1528 usable sequences were assembled into 190 contiguous sequences (contigs) and 501 singletons (unique sequences).The average length of readable sequences was 635 nucleotides. Each unique sequence was searched against the non-redundant GenBank database using the BLASTX algorithm. A listing of accession numbers and associated contig designations is available at http://staff.bath.ac.uk/bssrfc/EST.html .

Predicted biological and molecular EST functions

Of the 691 unique sequences, 27% (187) did not return any significant (E = 10^{-5}) BLASTX match (Figure 1A). Of the remaining 504 sequences, 71% had best matches to insect sequences, specifically 42% to *Drosophila melanogaster* and 29% to other insects (Figure 1B).Those sequences returning a significant BLASTX match were ascribed a putative biological process and molecular function (The Gene Ontology Consortium, 2001; http://www.geneontology.org/) based on the single "best hit" match. Tables 1 and 2 summarize the molecular functions and biological processes respectively. Molecular functions correspond to activities that can be performed by individual gene products, whereas biological process are accomplished by one or more ordered assemblies of molecular functions.The distinction between a biological process and a molecular function is based on the general rule that a process must have more than one distinct step (The Gene Ontology Consortium, 2001). Strikingly, 80% of the sequences predicted proteins with either catalytic activities (61.8%) or binding functions (18.8%) (Figure 2A). Correspondingly, 74% of sequences were predicted to encode proteins involved in either metabolism (64.5%) or transport (9.1%) (Figure 2B). Several proteins known to play a biological role in the insect midgut were identified by the ESTs. For example, at least five different peritrophin-like proteins, which comprise a major component of the peritrophic membrane, were identified.

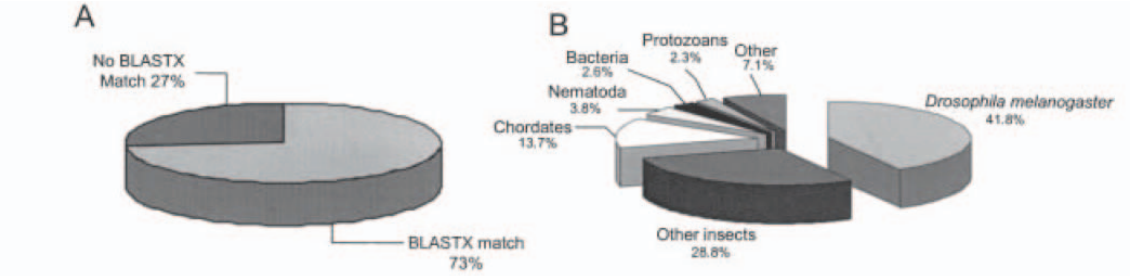

Figure 1. (A) The proportion of EST sequences with and without BLASTX matches (E < 10^{-5}) in GenBank. Note that 27% of the sequences did not predict the occurrence of similar proteins in current databases. (B) The percentages of sequences with matches from BLASTX classified by organism. Note that 70% of the identifiable sequences matched those found in other insects.

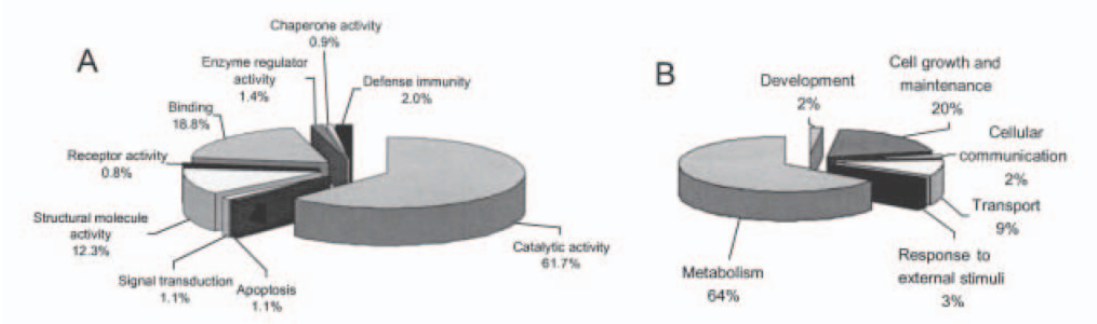

Figure 2. (A) Percentage analysis of the different biological processes ascribed to the 691 contigs by GeneOntology. Note the large percentage of predicted proteins involved in catalytic and binding activities. (B) Percentage analysis of the different molecular functions ascribed. Note the predominance of proteins involved in metabolism and transport (see text for discussion).

Table 1. Molecular function (The Gene Ontology Consortium, 2001; http://www.geneontology.org)

Function	No. of sequences	Function	No. of sequences
Antioxidant activity	4	Kinase activity	2
Apoptosis regulator activity	4	Lactase activity	2
Binding		Ligase activity	9
Actin binding	3	Lipase activity	5
Bacterial binding	7	Lyase activity	2
Calcium ion binding	3	Methyltransferase activity	3
Cytoskeletal binding	8	Nuclease activity	1
Imaginal disc growth factor activity	2	Oxidoreductase activity	52
Lipid binding	5	Peptidase activity	48
Metal binding	6	Peroxidase activity	1
Nucleic acid binding	21	Phosphoric ester hydrolase activity	1
Peptide binding	2	Protein phosphatase activity	2
Protein binding	6	Serine-type peptidase activity	1
Steroid binding	1	Transaminase activity	2
Phospholipid binding	1	Transferase activity	6
Drug binding	1	Glycosyl transferase activity	3
Catalytic activity		Transketolase activity	1
Acyl-transferase acivity	1	Trehalase activity	1
Aldo-keto reductase activity	2	Ubiquitin conjugating enzyme activity	2
Aldolase activity	3	Dopochrome isomerase activity	1
Amino acid kinase activity	3	Phosphorylase activity	1
Amylase activity	1	Aminopeptidase activity	1
Carbon-carbon lyase activity	1	Arginase activity	1
Carbon-oxygen lyase activity	2	Chaperone activity	
Carboxylesterase activity	12	Heat shock protein	3
Carboxypeptidase	4	Defence/immunity protein	7
Cellulase activity	12	Enzyme regulator activity	5
Chitinase activity	3	Receptor activity	2
Cytochrome P450	1	Signal transducer activity	4
Elastase activity	1	Structural molecule activity	
Endopeptidase	2	Structural constituent of cytoskeleton	18
Fatty-acid ligase activity	2	Structural constituent of peritrophic membrane	8
Fumarate hydratase activity	1	Structural constituent of ribosome	20
Galactosidase activity	2	Transcription regulator activity	6
Glucose-6-phosphate isomerase activity	1	Translation regulator activity	1
Glucosidase activity	3	Transporter activity	
Glutathione transferase activity	2	Amino acid transporter activity	2
GTPase	1	Carbohydrate transporter activity	1
Hydrolase activity acting on ether bonds	1	Electrochemical potential-driven transporter activity	18
Hydrolase activity acting on carbon–nitrogen bonds	1	Electron transporter activity	2
Hydrolase activity acting on glycosyl bonds	2	Hydrogen-exporting ATPase activity	10
Isomerse	5	Ion transporter activity	10

Classification is hierarchical: indented terms are children of parent terms listed above. Genes may be assigned to one or more than one term. Also note that child terms may have more than one parent term.

Candidate targets for gut-specific toxins

Two EST classes were examined in greater detail given their potential as targets for gut active toxins. These include a cadherin-like protein, which is the putative receptor for Bt toxins in Lepidoptera, and a family of cysteine proteases, the cathepsins, which have been proposed as targets for protease inhibitors. Sequence 2370 (accession number CN497270) contains a single clone predicting a protein with 29.9% identity and 48.7% similarity to a cadherin protein from the moth *Lymantria dispar*. We therefore sequenced this clone in its entirety. The complete clone predicts a partial cadherin protein showing regions of high identity to cadherin sequences from other insects (Figure 3). Phylogenetic analysis of the nucleotide sequences of all known insect cadherins (Figure 4) indicates that the dipteran, coleopteran and lepidoperan cadherins are all derived from a common ancestor although a greater sequence similarity between the Lepidoptera and Coleoptera is apparent.

The second class of ESTs that are candidates for gut-specific toxins predict a group of cathepsin-like proteins. The 171 individual sequences identified as cathepsin-like proteins comprised greater than 10% of the total sequences and assembled into eleven different contigs and four singletons.Visual comparison of the aligned nucleotide sequences and predictions of their consensus amino acid sequences suggest that these sequences belong to a large family of genes and are probably encoded by several independent loci. Phylogenetic analysis of the nucleotide sequences together with other *D. v. virgifera* cathepsin sequences available from the non-redundant GenBank database helped to clarify which proteins are likely to be encoded by allelic variants of the

Table 2. Biological function (The Gene Ontology Consortium, 2001; http://www.geneontology.org)

Process	No. of sequences
Cell communication	
Cell adhesion	1
Signal transduction	6
Cell growth and/or maintenance	
Carbohydrate transport	1
Cell ion homeostasis	2
Cell proliferation	1
Cell organization and biogenesis	30
Cytokinesis	1
ER organization and biogenesis	1
Nuclear organization and biogenesis	1
Proton transport	2
Receptor mediated endocytosis	1
Cell homeostasis	2
Development	
Morphogenesis	2
Organogenesis	2
Regulation of gene expression	4
Metabolism	12*
Alcohol metabolism	4
Aldehyde metabolism	2
Amine biosysntesis	1
Amino acid biosynthesis	8
Amino acid metabolism	2
Biogenic amine biosynthesis	3
Carbohydrate metabolism	35
Carboxylic acid metabolism	1
Catecholamine metabolism	3
Cuticle biosynthesis	2
DNA metabolism	1
Electron transport	14
Energy pathways	3
Glucose metabolism	7
Glycologysis	14
Glycoprotein biosynthesis	2
Hormone metabolism	1
Lipid metabolism	19
Nucleic acid metabolism	5
Nucleotide biosynthesis	1
Oxidative phosphorylation	1
Phosphate metabolism	6
Phospholipid metabolism	1
Protein biosynthesis	28
Protein folding	4
Protein metabolism	62
Regulation of transcription	3
RNA metabolism	6
Serotonin biosynthesis	2
Steroid biosynthesis	1
Formaldehyde metabolism	1
Pyruvate metabolism	1
Physiological process	17*
Chitin catabolism	2
Immune response	10
Response to external stimulus	8
Response to stress	2
Transcription	3
Transport	15*
Amino acid transport	1
ATP/ADP exchange	5
Lysosomal transport	1
Proton transport	9
Synaptic vesicle transport	1
Nucleocytoplasmic transport	1
Golgi vesicle transport	1

Classification is hierarchical: indented terms are children of parent terms listed above. Genes may be assigned to one or more than one term. Also note that child terms may have more than one parent term.
*Children of parent term could not be assigned.

same gene or by different independent loci (Figure 5).The tree was clearly divided among two classes of cathepsin genes (L and B) as was reported previously by Brown *et al*. (2004). A number of sequences were similar to cathepsin genes previously identified from *D. v. virgifera* (Koiwa *et al*., 2000; Brown *et al*., 2004). However, a few genes identified in these previous studies (e.g. *DvRs33* and *DvRs30*) apparently were not present in the cDNA library used in the present study, suggesting possible differences among populations in the composition of cathepsins.

Discussion

The present study highlights the utility of EST projects in identifying novel insecticide targets from candidate insect tissues. Previous insect EST projects have examined specific tissues such as the hindgut and Malpighian tubules of the cat flea, *Ctenocephalides felis* (Gaines *et al*., 2002), and gut tissue of cowpea weevil, *Callosobrucus maculates* (Pedra *et al*., 2003). All these these projects identified large numbers of ESTs (> 25%), without significant matches in current databases, suggesting that numerous proteins may be specific to the insect gut, and providing a useful starting point for identifying insecticide targets. Most of the *Diabrotica* ESTs returned matches to dipteran sequences, such as those from *D. melanogaster*. This probably reflects the current dominance of dipteran (fly and mosquito) sequences in GenBank, and the paucity of coleopteran (beetle) sequences, and is therefore not really indicative of any true phylogenetic relationships among the ESTs. The predicted biological functions of the *Diabrotica* ESTs support the dominant role of the midgut in binding (19% of ESTS) and catalytic activity (62%). These figures therefore stand in contrast to the dominance of ESTs involved in cell growth and communication found in collections of honey bee brain ESTs (Whitfield *et al*., 2002).

The *Diabrotica* ESTs predict a variety of proteins that are highly expressed in the midgut, such as pertitrophin, a major component of the peritrophic membrane. However, the collection also contains tags for specific proteins of lower abundance, such as the *Diabrotica* cadherin which is, to our knowledge, the first coleopteran cadherin to be identified. Although the predicted protein contains several regions of high identity to lepidopteran proteins (Figure 3), the overall level of amino acid identity suggests that this cadherin gene may have been hard to clone via low-stringency hybridization or the use of degenerate primers in the polymerase chain reaction. The isolation of this EST among a relatively small EST sample size therefore illustrates the utility of a tissue-specific approach. Although there is still considerable debate over the relative role of cadherins in the binding of Bt δ-endotoxins to the insect midgut, cloning and

```
 243  - E Q K L N L D S F E T N S K - - S G N S E A A L R A W L I G - V S V V L G I L V L L L L I T L I L  Rootworm cadherin EST
1545  A T R L L V L R D L F T D T S P A P D A G S A - A V L Y A L A V L S A L L A A L C L L L L V I F I I  Bombyx mori
1543  S E R N L V L Q D L S T G L G - - Q S A D P V Q T V I Y I L A G L C I L L T L L L L V L S V V F S F  Chilo supressalis
1555  E E Q N L Q L A D L F T G E T P I L G G D A Q A R A L Y A L A A V A A A L A L I V V V L L I V F F V  Helicoverpa armigera
1555  Q E E N L N L A D L F T G E T P I L G G E A Q A R A V Y A L A A V A A A L A L L C V V L L I L F F I  Heliothis virescens
1550  N M Q L V T L E D F V T G A S P D I S T D S A Q I I V Y V L I G L T V A L G F L C I V L L L T C I I  Lymantria dispar
1548  A A R S L V L Q D L L T N S S P D S A P D S S - L T V Y V L A S L S A V L G F M C L V L L L T F I I  Manduca sexta
1556  N Q R L L V L N D L V T G V S P D L G T A G V Q I T I Y V L A G L S A I L A F L C L I L L I T F I V  Pectinophora gossypiella
1727  F K E F N V L D T Q A S E L V Q T A E M D E L S V N I I W L F V T N I L L G A L L I V V I G L S I S  Anopholes gambiae
1822  F K E L N V L D T Q A A E A Q L L T A G P S R G P L F V W L I F T N L F L A T L L V V T I A L C A S  Drosophila melanogaster

 289  K T R Q L S N R I K K L T T P Q F G S Q E S G - - - - L N R M G I N A P T T N K H A I E G T N P V Y  Rootworm cadherin EST
1594  R T K K L N R R L E A L T V K K Y G S V D S G - - - - L N R V G I A A P G T N K H A V E G S N P I W  Bombyx mori
1591  K T S A L N R R M K A M S M T K F G S V D S G - - - - L N R M G I A P - G T N K H T V E G S N P I Y  Chilo suppressalis
1605  R T R T L N R R L Q A L S M T K Y S S Q D S G - - - - L N R V G L A A P G T N K H A V E G S N P I W  Helicoverpa armigera
1605  R T R A L N R R L E A L S M T K Y S S Q D S G - - - - L N R V G L A A P G T N K H A V E G S N P I W  Heliothis virescens
1600  R T R M L N R R L E A L S M T R Y G S V D S G - - - - L N R A G I A A P G T N K H A V E G S N P I W  Lymantria dispar
1597  R T R A L N R R L E A L S M T K Y G S L D S G - - - - L N R A G I A A P G T N K H T V E G S N P I F  Manduca sexta
1606  R T R A L N R R L E A L S M T K Y G S V D S G - - - - L N R V G I A A P G T N K H A I E G S N P I W  Pectinophora gossypiella
1777  Q R L S Y R R Q L R A A K I A A F G - - S T G P S R M Y Q E V L G A V P N T N K H S M K G S N P I W  Anopholes gambiae
1872  Q R N G Y R R Q L R A A K V N I F R G H S S M S L Q D A Q E P A T R V P N T N K H S V Q G S N P I W  Drosophila melanogaster

 335  N N N - E I K K P K N M N D F D T H S I R S G D S D L V G I E N N P E F D Y N F N T N E D K T T Y L  Rootworm cadherin EST
1640  N - E - T I K A P - - - - D F D S M S D A S N D S D L I G I E D L P H F G E N N Y F P R D V D E F K  Bombyx mori
1636  N - E - N I K A P - - - - D F D A I S D M S N D S D L I G I E D L P Q F R G D F L P P E D S N S T A  Chilo suppressalis
1651  N - E - T L K A P - - - - D F D A L S E Q S Y D S D L I G I E D L P Q F R N D Y F P P E E G S S M R  Helicoverpa armigera
1651  N - E - T L K A P - - - - D F D A L S E Q S Y D S G L I G I E D L P Q F R N D Y F P P D E E S S M R  Heliothis virescens
1646  N - E - T I K A P - - - - D F D A L S D V S N E S D L I G I E D M P Q F K L D Y N P S A D D G S L Q  Lymantria dispar
1643  N - E - A I K T P - - - - D L D A I S E G S N D S D L I G I E D L P H F G - N V F M D P E V N E K A  Manduca sexta
1652  N - E - Q I K A P - - - - D F D A I S D T S D E S D L I G I E D L P Q F K S D Y F P P E D S E S A H  Pectinophora gossypiella
1825  I G S G T P E G E W A K D E F D K C K D A I D A Q Y E R S L S S G - F F I D N C L Q Y E A R K G F A  Anopholes gambiae
1922  L K G - - Y D N E W F K S E E S G S I G G H D S L D D N F L A V A T Q D M H E T L K G T A K L F N N  Drosophila melanogaster
```

Figure 3. Amino acid alignment of known cadherin genes with *D. v. virgifera* cadherin sequence (AY769085). Amino acids shared with the *D. v. virgifera* sequence are highlighted in bold. Aligned sequences for cadherin genes with accession numbers are *Bombyx mori* (BAA99405), *Manduca sexta* (AAM21151), *Lymantria dispar* (AAL26896), *Helicoverpa armigera* (AAM65319), *Heliothis virescens* (AAK85198), *Chilo supressalis* (AAM75590), *Pectinophora gossypiella* (AAP31705), *Anopholes gambiae* (XM321513) and *Drosophila melanogaster* (AAF55082).

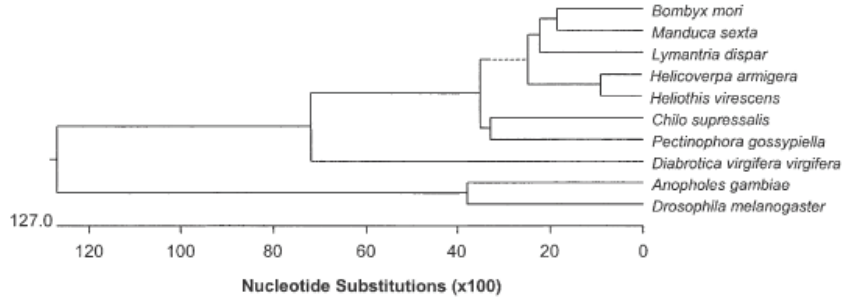

Figure 4. Phylogenetic tree of known cadherin sequences in relation to *D. v. virgifera* cadherin EST. See Figure 3 for accession numbers.

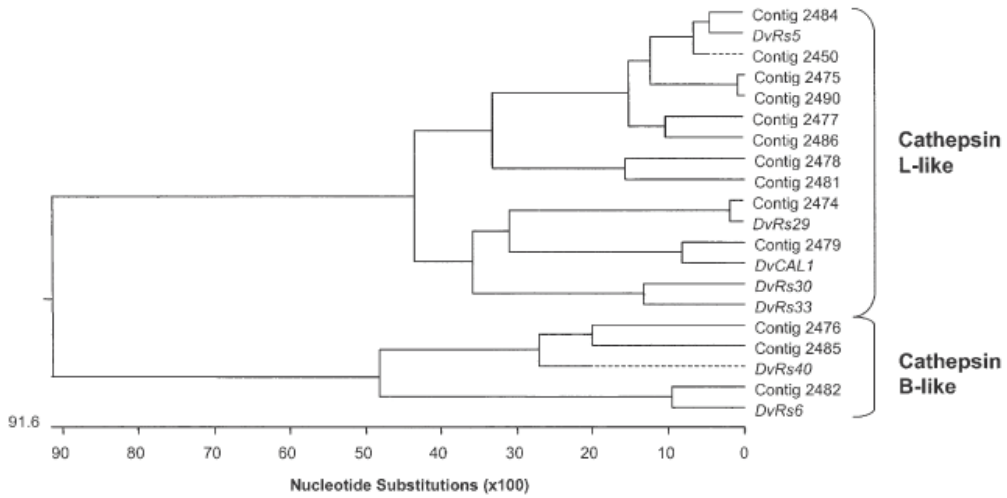

Figure 5. Phylogentic tree of cathepsin-like enzymes predicted by clones isolated from *D. v. virgifera* larval midgut cDNA library. Aligned sequences of *D. v. virgifera* cathepsins with accession numbers are *DvRs5* (AJ583508), *DvRs29* (AJ583510), *DvRs30* (AJ583511), *DvRs33* (AJ583512), *DvRs40* (AJ583513), *DvRs6* (AJ583509) and *DvCAL1* (AAG17127). Accession numbers for individual EST clones associated with each contig are available at http://staff.bath.ac.uk/bssrfc/EST.html.

functional expression of the *Diabrotica* cadherin gene will allow us to test its role in toxin binding directly and perhaps begin to dissect the relationship between taxonomic status and specificity of various cry toxins from *Bacillus thuringiensis*.

One of the largest groups of ESTs identified in the EST database was the cathepsin-like cysteine proteases. Unlike that seen in mammals and most insects, digestive proteolysis by coleopterans (including *D. v. virgifera*) is predominantly due to cysteine proteinase activity (Gatehouse *et al*., 1985; Murdock *et al*., 1988; Lilley *et al*., 1997; Koiwa *et al*., 2000). Examination of the EST sequences reveals a large diversity of cysteine proteases within the rootworm midgut that apparently are highly expressed given that > 10% of the clones were identified as cathepsin hits. Previously, Koiwa *et al*. (2000) resolved five major peptides by affinity chromatography which exhibited similar N-terminal sequences and suggested that the digestive proteolytic system of rootworms is derived from products of a multigene family. Recently, Brown *et al*. (2004) reported that a cDNA library representing *D. v. virgifera* larval gut tissue mRNA contained cysteine proteinase-encoding clones at high frequency (15%) although only seventy clones were sequenced. Sequence analysis revealed eleven cysteine proteinases, nine of which were cathepsin L-like enzymes and another two cathepsin B-like enzymes.The present study is in agreement with these results except that the diversity of cathepsins appears to be even greater than reported by Brown *et al*. (2004).

In the absence of complete genome sequences for beetle pests, EST sequencing is likely to remain a useful tool in the identification of candidate proteins for novel toxins. Tissue-specific EST projects in different beetle pests, such as *Tribolium*, should enable us to establish what percentage of the unidentified midgut ESTs correspond to beetle-specific gut proteins. In this fashion, we should be able to ascertain not only which proteins are specific to insects as a whole but also which are specific to certain insect tissues, presenting candidate insect and tissue-specific insecticide targets.

Experimental procedures

Library preparation and EST sequencing

Midguts were dissected from approximately 800 third-instar western corn rootworm larvae and flash frozen in liquid nitrogen. The dissected tissue was shipped on dry ice to Invitrogen (Carlsbad, CA, USA), where a microquantity cDNA library was prepared from polyA$^+$ selected mRNA using SuperScript II RNase H$^-$ reverse transcriptase (Invitrogen Life Technologies). The total number of primary clones (2.53×10^{10}) in the library and their average insert size (1.7 kb) were estimated. *Escherichia coli* (strain ElectroMax DH10b ton A) colonies containing individual cDNA clones were picked into ninety-six-well deep-well microtitre plates and grown overnight in 2× Yt supplemented with ampicillin (100 μg/ml) antibiotic. Glycerol (20% v/v) was added to aliquots of these overnight cultures for the preparation of freezer stocks. DNA was prepared from the remainder of the overnight culture using a plasmid-preparing robot (MWG RoboPrep 2500), and the clones were sequenced using the SP6 sequencing primer on an ABI3700 capillary sequencer. As the library was directionally cloned, this primer sequenced through the 5′ end of the cDNA.

EST annotation

Following automatic removal of vector sequences, nucleotide sequences were assembled using LaserGene software (DNAStar, Madison, WI, USA) to determine redundancy among the EST sequences. All ESTs have been submitted to the GenBank dbEST database (accession numbers CN497248–CN498776). Consensus sequences were submitted in batch to the Blastx server to search GenBank non-redundant databases current as of September 2003.The top hit from the BLASTX search was deposited in a spreadsheet for further annotation and ease of data handling. Biological processes and molecular functions were assigned using GeneOntology (The Gene Ontology Consortium, 2001; http://www.geneontology.org/) based on annotation of the single 'best hit' match in BLASTX searches.

For the cadherin EST, the complete cDNA was sequenced by primer walking using primers derived from internal sequence.This was repeated until the ESTs had been sequenced three times on each strand of DNA. For the cathepsin ESTs, the nucleotide sequences were assembled in LaserGene as contigs and then the assemblies viewed manually to look for nucleotide differences likely to correspond to allelic variation or differences likely to correspond to proteins encoded by different loci. Comparative protein sequence analyses was carried out using the CLUSTALW analysis program with default parameters (DNAStar) and phylogenetic trees derived from the resulting alignments.

Acknowledgements

We thank Paul Wilkinson for DNA sequencing. The Biotechnology and Biology Research Council supported the research through the Underwood Fund provided to R.f.-C. and B.D.S. This is journal series paper 14660, Nebraska Agricultural Research Division, and contribution no. 1191 of the Department of Entomology, University of Nebraska–Lincoln.

References

Ball, H.J. and Weekman, G.T. (1963) Insecticide resistance in the western corn rootworm in Nebraska. *J Econ Entomol* **55**: 439–441.

Brown, D.P., Wilkinson, H.S., Jongsma, M.A., and Gatehouse, J.A. (2004) Characterisation of cysteine proteinases responsible for digestive proteolysis in guts of larval western corn rootworm (*Diabrotica virgifera*) by expression in the yeast *Pichia pastoris*. *Insect Biochem Mol Biol* **34**: 305–320.

Ellis, R.T., Stockhoff, B.A., Stamp, L., Schnepf, H.E., Schwab, G.E., Knuth, M., Russell, J., Cardineau, G.A., and Narva, K.E. (2002) Novel *Bacillus thuringiensis* binary insecticidal crystal proteins active on western corn rootworm, *Diabrotica virgifera virgifera* LeConte. *Appl Environ Microbiol* **68**: 1137–1145.

Environmental Protection Agency (2003) *Bacillus thuringiensis Cry3Bb1 Protein and the Genetic Material Necessary for its Production (Vector ZMIR13L) in Event MON863 Corn Fact Sheet*. EPA Publication Number 730-F-03–01. Environmental Protection Agency, Washington, DC.

Ferré, J. and Van Rie, J. (2002) Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*. *Annu Rev Entomol* **47**: 501–533.

Gaines, P.J., Brandt, K.S., Eisele, A.M., Wagner, W.P., Bozic, C.M., and Wisnewski, N. (2002) Analysis of expressed sequence tags from subtracted and unsubtracted *Ctenocephalides felis* hindgut and Malpighian tubule cDNA libraries. *Insect Mol Biol* **11**: 299–306.

Gatehouse, A.M.R., Butler, K.J., Fenton, K.A., and Gatehouse, J.A. (1985) Presence and partial characterisation of a major proteolytic enzyme in the larval gut of *Callosobruchus maculatus*. *Entomol Exp Appl* **39**: 279–286.

Hill, C.A. and Guitierrez, J.A. (2000) Analysis of the expressed genome of the lone star tick, *Amblyomma americanum* (Acari: Ixodidae) using an expressed sequnece tag approach. *Microbial Comparative Genomics* **5**: 89–101.

Hummel, H.E. (2003) Introduction of *Diabrotica virgifera virgifera* into the Old World and its consequences: a recently acquired invasive alien pest species on *Zea mays* from North America. *Commun Agric Appl Biol Sci* **68**: 45–57.

Koiwa, H., Shade, R.E., Zhu-Salzman, K., D'Urzo, M.P., Murdock, L.L., Bressan, R.A., and Hasegawa, P.M. (2000) A plant defensive cystatin (soyasystatin) targets cathepsin L-like digestive cysteine proteinases (DvCALs) in the larval midgut of western corn rootworm (*Diabrotica virgifera virgifera*). *FEBS Lett* **471**: 67–70.

Levine, E. and Oloumi-Sadeghi, H. (1991) Management of diabroticite rootworms in corn. *Annu Rev Entomol* **36**: 229–255.

Levine, E., Spencer, S.A., Isard, Onstad, D.W., and Gray, M.E. (2002) Adaptation of the western corn rootworm, *Diabrotica virgifera virgifera* LeConte (Coleoptera: Chrysomelidae) to crop rotation: evolution of a new strain in response to a cultural management practice. *Am Entomol* **48**: 94–107.

Lilley, C.J., Urwin, P.E., McPherson, M.J., and Atkinson, H.J. (1997) Characterization of intestinally active proteinases of cyst-nematodes. *Parasitology* **113**: 415–424.

Liu, D., Burton, S., Glancy, T., Li, Z.S., Hampton, R., Meade, T., and Merlo, D.J. (2003) Insect resistance conferred by 283-kDa *Photorhabdus luminescens* protein TcdA in *Arabidopsis thaliana*. *Nat Biotechnol* **21**: 1222–1228.

Loguercio, L.L., Barreto, M.L., Rocha, T.L., Santos, C.G., Teixeira, F.F., and Paiva, E. (2002) Combined analysis of supernatant-based feeding bioassays and PCR as a first-tier screening strategy for VIP-derived activities in *Bacillus thuringiensis* strains effective against tropical fall armyworm. *J Appl Microbiol* **93**: 269–277.

Meinke, L.J., Siegfried, B.D., Wright, R.J., and Chandler, L.D. (1998) Adult susceptibility of Nebraska western corn rootworm (Coleoptera: Chrysomelidae) populations to selected insecticides. *J Econ Entomol* **91**: 594–600.

Moellenbeck, D.J., Peters, M.L., Bing, J.W., Rouse, J.R., Higgins, L.S., Sims, L., Nevshemal, T., Marshall, L., Ellis, R.T., Bystrak, P.G., Lang, B.A., Stewart, J.L., Kouba, K., Sondag, V., Gustafson, V., Nour, K., Xu, D., Swenson, J., Zhang, J., Czapla, T., Schwab, G., Jayne, S., Stockhoff, B.A., Narva, K., Schnepf, H.E., Stelman, S.J., Poutre, C., Koziel, M., and Duck, N. (2001) Insecticidal proteins from *Bacillus thuringiensis* protect corn from corn rootworms. *Nat Biotechnol* **19**: 668–672.

Morin, S., Biggs, R.W., Sisterson, M.S., Shriver, L., Ellers-Kirk, C., Higginson, D., Holley, D., Gahan, L.J., Heckel, D.G., Carriere, Y., Dennehy, T.J., Brown, J.K., and Tabashnik, B.E. (2003) Three cadherin alleles associated with resistance to *Bacillus thuringiensis* in pink bollworm. *Proc Natl Acad Sci USA* **100**: 5004–5009.

Murdock, L.L., Shade, R.E., and Pomeroy, M.A. (1988) Effects of E-64, a cysteine proteinase inhibitor, on cowpea weevil growth, development, and fecundity. *Eviron Entomol* **17**: 467–469.

O'Neal, M.E., Gray, M.E., Ratcliffe, S., and Steffey, K.L. (2001) Predicting western corn rootworm (Coleoptera: Chrysomelidae) larval injury to rotated corn with Pherocon AM traps in soybeans. *J Econ Entomol* **94**: 98–105.

Pedra, J.J.F., Brandt, A., Westerman, R., Lobo, N., Li, H.-M., Romero-Severson, J., Murdock, L.L., and Pittendrigh, B.R. (2003) Transcriptome analysis of the cowpea weevil bruchid: identification of putative proteinases and α-amylases associated with food breakdown. *Insect Molec Biol* **12**: 405–412.

The Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. *Genome Res* **11**: 1425–1433.

Whitfield, C.W., Band, M.R., Bondaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J.R., Robertson, H.M., Soares, M.B., and Robinson, G.E. (2002) Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res* **12**: 555–566.

Zhou, X., Scharf, M.E., Parimi, S., Meinke, L.J., Wright, R.J., and Chandlerr, L.D. (2002) Diagnostic assays based on esterase-mediated resistance mechanisms in western corn rootworms (Coleoptera: Chrysomelidae). *J Econ Entomol* **95**: 1261–1266.