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*

Follow this and additional works at: [http://digitalcommons.unl.edu/entomologyfacpub](http://digitalcommons.unl.edu/entomologyfacpub)

Part of the [Entomology Commons](http://digitalcommons.unl.edu/entomologyfacpub)

Siegfried, Blair D.; Waterfield, N.; and ffrench-Constant, R. H., "Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins" (2005). *Faculty Publications: Department of Entomology*. 74.

[http://digitalcommons.unl.edu/entomologyfacpub/74](http://digitalcommons.unl.edu/entomologyfacpub/74)

This Article is brought to you for free and open access by the Entomology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Department of Entomology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
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, *Diabrotica virgifera virgifera* 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 important 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; Loguerchio 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 \(2.53 \times 10^{10}\) 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](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/](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 processes 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](http://staff.bath.ac.uk/bssrfc/EST.html)

**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](http://staff.bath.ac.uk/bssrfc/EST.html)

**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).
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
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, Callosobruchus maculatus (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
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), *Man- 

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

Figure 5. Phylogenetic 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* (AA917127). 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 proteinases. 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; Koiba et al., 2000). Examination of the EST sequences reveals a large diversity of cysteine proteinases within the rootworm midgut that apparently are highly expressed given that > 10% of the clones were identified as cathepsin hits. Previously, Koiba 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 micro-quantity 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 × 10¹⁰) 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/](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


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.


