



Published in final edited form as:

*J Proteome Res.* 2010 August 6; 9(8): 3820–3831. doi:10.1021/pr1000169.

## An insight into the sialome of the bed bug, *Cimex lectularius*

Ivo M.B. Francischetti<sup>1</sup>, Eric Calvo<sup>1</sup>, John F. Andersen<sup>1</sup>, Van M. Pham<sup>1</sup>, Amanda J. Favreau<sup>2</sup>, Kent D. Barbian<sup>2</sup>, Alvaro Romero<sup>3</sup>, Jesus G. Valenzuela<sup>1</sup>, and José M.C. Ribeiro.<sup>1,\*</sup>

Ivo M.B. Francischetti: ifrancischetti@niaid.nih.gov; Eric Calvo: eric.calvo@fda.hhs.gov; John F. Andersen: jandersen@niaid.nih.gov; Van M. Pham: vpham@niaid.nih.gov; Amanda J. Favreau: favreua@niaid.nih.gov; Kent D. Barbian: KBarbian@niaid.nih.gov; Alvaro Romero: alvromero@uky.edu; Jesus G. Valenzuela: jvalenzuela@niaid.nih.gov; José M.C. Ribeiro.: jribeiro@niaid.nih.gov

<sup>1</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway, Rockville MD 20852

<sup>2</sup>Genomics Unit, Research Technologies Section, Rocky Mountain Laboratories, Hamilton, MT 59840

<sup>3</sup>Department of Entomology, University of Kentucky, S-225 Agriculture Science Center, Lexington, KY, 40546

### Abstract

The evolution of insects to a blood diet leads to the development of a saliva that antagonizes their hosts' hemostasis and inflammation. Hemostasis and inflammation are redundant processes, and thus a complex salivary potion comprised of dozens or near one hundred different polypeptides is commonly found by transcriptome or proteome analysis of these organisms. Several insect orders or families evolved independently to hematophagy creating unique salivary potions in the form of novel pharmacological use of endogenous substances, and in the form of unique proteins not matching other known proteins, these probably arriving by fast evolution of salivary proteins as they evade their hosts' immune response. In this work we present a preliminary description of the sialome (from the Greek Sialo = saliva) of the common bed bug *Cimex lectularius*, the first such work from a member of the Cimicidae family. This manuscript is a guide for the supplemental database files [http://exon.niaid.nih.gov/transcriptome/C\\_lectularius/S1/Cimex-S1.zip](http://exon.niaid.nih.gov/transcriptome/C_lectularius/S1/Cimex-S1.zip) and [http://exon.niaid.nih.gov/transcriptome/C\\_lectularius/S2/Cimex-S2.xls](http://exon.niaid.nih.gov/transcriptome/C_lectularius/S2/Cimex-S2.xls)

### Keywords

Bedbug; saliva; salivary transcriptome; salivary proteome

### Introduction

The habit of vertebrate blood feeding evolved independently in many insect families within 5 different orders.<sup>1</sup> Within the true bugs (Heteroptera) the habit developed independently within the Cimicomorpha families Cimicidae (the bed bugs and its smaller sister group Polyctenidae, bat bugs) and Reduviidae (kissing bugs), possibly from predacious ancestors,<sup>2</sup> producing, as examples, the common bed bug *Cimex lectularius* and the several genera of reduvid vectors of Chagas' disease.<sup>1</sup> These insects feed exclusively on blood throughout all

\*Corresponding author. José M.C. Ribeiro, Laboratory of Malaria Research, NIAID, 12735 Twinbrook Parkway, Rockville MD 20852. jribeiro@niaid.nih.gov.

**Supporting Information Available:** Supplemental Tables S1 and S2, and supplemental materials S3 and S4 are herein hyperlinked.

immature instars and as adults. The Cimicomorpha is an ancient Heteroptera branch that dates back to Triassic/Jurassic border, over 250 million years ago (MYA).<sup>1</sup>

Although *C. lectularius* can harbor viruses, bacteria and protozoal parasites, it is not generally considered a human disease vector.<sup>3</sup> Although *C. lectularius* prevalence worldwide decreased in the last half of the past century, more recently it has made reappearances in megalopolis such as New York City and London,<sup>4-7</sup> producing an increase in the literature associated with allergic responses to bed bug bites.<sup>8-12</sup>

Among the adaptations to blood feeding, hematophagous insects developed specialized saliva that counteracts their hosts' hemostasis (comprised of platelet aggregation, vasoconstriction and blood clotting) and inflammation.<sup>13, 14</sup> Previous studies with *C. lectularius* salivary gland homogenates has identified a novel type of secreted apyrase enzyme that hydrolyzed ATP and ADP.<sup>15, 16</sup> This enzyme destroys these nucleotide agonists of platelet and neutrophil aggregation that are released by injured cells.<sup>17</sup> A still molecularly uncharacterized factor X activation inhibitor<sup>18</sup> was also identified, as well as a nitric oxide (NO) carrier, named *Cimex* nitrophorin,<sup>19-21</sup> that carries the unstable NO gas molecule to the host tissues, promoting vasodilatation and inhibiting platelet aggregation.<sup>22</sup> Recombinant *Cimex* nitrophorin was recently identified as an allergen in patients with severe allergy to bed bug bites.<sup>9</sup>

In the past 8 years, salivary transcriptomes analysis of blood feeding arthropods started revealing the complex composition of this secretion, the sialome. Mosquitoes have near 100 different proteins, many of which are products of gene duplications of unique families. Kissing bug sialomes have over 100 different proteins including a large expansion of the lipocalin family of proteins that play different functions, such as carriers of nitric oxide,<sup>23</sup> chelators of inflammation and hemostasis agonists (named kratagonists)<sup>13</sup> such as histamine,<sup>24</sup> serotonin<sup>25</sup> and adenosine nucleotides,<sup>26, 27</sup> and as anticlotting mediators.<sup>28-30</sup> No sialome has been described so far for any member of the Cimicidae family. This paper attempts a preliminary description of the sialome of the common bed bug, *Cimex lectularius*.

## Materials and Methods

### Bed bugs

Adult insects were obtained from a colony (Fort Dix) maintained at 26° C, 65 ± 5% RH, and a photoperiod of 14:10 (L:D) h. This colony has not been exposed to insecticides for more than 30 years.<sup>31</sup> Insects were fed weekly in the laboratory through a parafilm–membrane feeder with citrated rabbit blood heated to 39° C with a circulating water bath.<sup>32</sup> Blood was supplied by Hemaresources (Oregon).

### Salivary Gland Isolation and Library Construction

Salivary glands were dissected from both males and female adult bugs that have been offered a blood meal 3 days earlier. The bugs were immersed in phosphate buffered saline (potassium phosphate 10 mM, NaCl 150 mM, pH 7.0) and their heads were pulled from the thorax with fine forceps, producing the orange colored glands attached to the head. These were further isolated from the head and transferred to 100 µl of RNA later (Qiagen, Valencia CA). The mRNA from 60 pairs of salivary glands was isolated using the Micro-FastTrack mRNA isolation kit (Invitrogen, San Diego, CA). The PCR-based cDNA library was made following the instructions for the SMART cDNA library construction kit (Clontech, Palo Alto, CA). This system utilizes oligoribonucleotide (SMART IV) to attach an identical sequence at the 5' end of each reverse-transcribed cDNA strand. This sequence is then utilized in subsequent PCR reactions and restriction digests.

First-strand synthesis was carried out using PowerScript reverse transcriptase at 42°C for 1 hour in the presence of the SMART IV and CDS III (3') primers. Second-strand synthesis was performed using a long distance (LD) PCR-based protocol, using Advantage™ Taq polymerase (Clontech) mix in the presence of the 5' PCR primer and the CDS III (3') primer. The cDNA synthesis procedure resulted in creation of *Sfi*I A and B restriction enzyme sites at the ends of the PCR products that are used for cloning into the phage vector. PCR conditions were as follows: 95°C for 20 sec; 24 cycles of 95°C for 5 sec., 68°C for 6 min. A small portion of the cDNA obtained by PCR was analyzed on a 1.1% agarose gel to check quality and range of cDNA synthesized. Double-stranded cDNA was immediately treated with proteinase K (0.8 µg/ml) at 45°C for 20 min, and the enzyme was removed by ultrafiltration through a Microcon (Amicon Inc., Beverly, CA) YM-100 centrifugal filter device. The cleaned, double-stranded cDNA was then digested with *Sfi*I at 50°C for 2 hours, followed by size fractionation on a ChromaSpin– 400 column (Clontech). The profile of the fractions was checked on a 1.1% agarose gel, and fractions containing cDNAs of more than 400 bp were pooled and concentrated using a Microcon YM-100.

The cDNA mixture was ligated into the λ TriplEx2 vector (Clontech), and the resulting ligation mixture was packaged using the GigaPack® III Plus packaging extract (Stratagene, La Jolla, CA) according to the manufacturer's instructions. The packaged library was plated by infecting log-phase XL1- Blue *Escherichia coli* cells (Clontech). The percentage of recombinant clones was determined by blue-white selection screening on LB/MgSO<sub>4</sub> plates containing X-gal/IPTG. Recombinants were also determined by PCR, using vector primers (5' λ TriplEx2 sequencing primer and 3' λ TriplEx2 sequencing) flanking the inserted cDNA, with subsequent visualization of the products on a 1.1% agarose/EtBr gel.

### Sequencing of the *C. lectularius* cDNA Library

The *C. lectularius* salivary gland cDNA library was plated on LB/MgSO<sub>4</sub> plates containing X-gal/IPTG to an average of 250 plaques per 150-mm Petri plate. Recombinant (white) plaques were randomly selected and transferred to 96-well MICROTTEST™ U-bottom plates (BD BioSciences, Franklin Lakes, NJ) containing 100 µl of SM buffer [0.1 M NaCl; 0.01 M MgSO<sub>4</sub>; 7 H<sub>2</sub>O; 0.035 M Tris-HCl (pH 7.5); 0.01% gelatin] per well. The plates were covered and placed on a gyrating shaker for 30 min at room temperature. The phage suspension was either immediately used for PCR or stored at 4°C for future use.

To amplify the cDNA using a PCR reaction, 4 µl of the phage sample was used as a template. The primers were sequences from the λ TriplEx2 vector and named PT2F1 (AAG TAC TCT AGC AAT TGT GAG C) and PT2R1 (CTC TTC GCT ATT ACG CCA GCT G), positioned at the 5' end and the 3' end of the cDNA insert, respectively. The reaction was carried out in 96 well PCR microtiter plate (Applied Biosystems) using FastStart Taq polymerase (Roche diagnostics) on a GeneAmp® PCR system 9700 (Perkin Elmer Corp., Foster City, CA). The PCR conditions were: one hold of 75°C for 3 min; 1 hold 94°C for 4 min, 33 cycles of 94°C for 1 min, 49°C for 1 min; 72°C for 2 min. The amplified products were analysed on a 1.5% agarose/EtBr gel. Clones were PCR amplified, and the ones showing single band were selected for sequencing. Approximately 200–250 ng of each PCR product was transferred to a 96 well PCR microtiter plate (Applied Biosystems) and frozen at –20°C. Samples were shipped on dry ice to the Rocky Mountain Laboratories Genomics Unit with primer (PT2F3, TCT CGG GAA GCG CGC CAT TGT) and template combined together in an ABI 96-well Optical Reaction Plate (P/N 4306737) following the manufacturers recommended concentrations. Sequencing reactions were setup as recommended by Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit by adding 1 µl ABI BigDye® Terminator Ready Reaction Mix v3.1 (P/N 4336921), 1.5 µl 5× ABI Sequencing Buffer (P/N 4336699), and 3.5 µl of water for a final volume of 10 µl. Cycle sequencing was performed at 96 °C for 10 seconds, 50 °C for 5 seconds, 60 °C for 4

minutes for 27 cycles on either a Bio-Rad Tetrad 2 (Bio-Rad Laboratories, Hercules, CA) or ABI 9700 (Applied Biosystems, Inc., Foster City, CA) thermal cycler. Fluorescently-labeled extension products were purified following Applied Biosystems BigDye® XTerminator™ Purification protocol and subsequently processed on an ABI 3730xL DNA Analyzer (Applied Biosystems, Inc., Foster City, CA). In addition to the sequencing of the cDNA clones, primer extension experiments were performed in selected clones to further extend sequence coverage.

### Bioinformatic Tools and Procedures Used

ESTs were trimmed of primer and vector sequences, clustered, and compared with other databases as previously described.<sup>33</sup> The CAP3 assembler was used to assemble EST sequences,<sup>34</sup> the BLAST tool was used to identify similar sequences in various databases,<sup>35</sup> the ClustalW<sup>36</sup> tool was used to align multiple sequences and TreeView version 1.6.6 software<sup>37</sup> was used to visualize phylogenetic trees. Dendrograms were drawn by the neighbor-joining (NJ) method implemented in MEGA package (version 4.0), and bootstrap pseudoreplicate was performed to evaluate statistical significance of tree topology.<sup>38</sup> For functional annotation of transcripts, we used the tool BlastX<sup>39</sup> to identify similar protein sequences to the NR protein database of the National Center for Biotechnology Information (NCBI) and to the Gene Ontology (GO) database.<sup>40</sup> The tool, Reverse Position Specific Blast (RPSBlast),<sup>39</sup> was used to search for conserved protein domains in the Pfam,<sup>41</sup> SMART,<sup>42</sup> Kog,<sup>43</sup> and conserved domains databases (CDD)<sup>44</sup>. The tool Seedtop, included in the stand-alone blast package, was used to search for PROSITE motifs.<sup>45</sup> We have also compared the transcripts with other subsets of mitochondrial and rRNA nucleotide sequences downloaded from NCBI. Segments of the three-frame translations of the EST (because the libraries were unidirectional, we did not use six-frame translations), starting with a methionine found in the first 40 predicted amino acids (AA) of any open reading frame (ORF), or to the predicted protein translation in the case of complete coding sequences, were submitted to the SignalP server<sup>46</sup> to help identify translation products that could be secreted. O-glycosylation sites on the proteins were predicted with the program NetOGlyc<sup>47</sup>. Functional annotation of the transcripts was based on all the comparisons above. Following inspection of all these results, transcripts were classified as either Secretory (S), Housekeeping (H) or of Unknown (U) function, with further subdivisions based on function and/or protein families.

### Proteomic characterization using one-dimensional gel electrophoresis and tandem mass spectrometry (MS)

The soluble protein fraction from salivary gland homogenates from *C. lectularius* corresponding to approximately 50 µg of protein was brought up in reducing Laemmli gel-loading buffer. The sample was boiled for 10 min and resolved on a NuPAGE 4-12% Bis-Tris precast gel. The separated proteins were visualized by staining with SimplyBlue (Invitrogen). The gel was sliced into 25 individual sections (Fig 1) that were destained and digested overnight with trypsin at 37°C. Identification of gel separated proteins was performed on reduced and alkylated, trypsin digested samples prepared by standard mass spectrometry protocols. Tryptic digests were concentrated and desalted on a C18 Atlantis Nanoease trapping column at 15 µl/min (0.180 mm ID × 23.5 mm, Waters Corp., Milford, MA). The recovered peptides were chromatographed and separated at 300 nl/min using a Waters Corp. XBridge BEH130 C18 nanoease LC reverse phase column (0.075 mm ID × 150 mm). The mobile phase consisted of a linear gradient prepared from solvent A (0.1% formic acid, 2% acetonitrile) and solvent B (0.2% formic acid, 98% acetonitrile) at room temperature. Nano LC-tandem MS (LC-MS/MS) was performed with a Tempo MDLC, a temperature controlled NanoSpray II ion source and a quadrupole-time of flight (TOF) mass spectrometer- QStarXL (Applied Biosystems/MDS Sciex, Framingham, MA). TOF mass

calibration was performed when needed to maintain mass accuracy. From a solution of Glu-Fibrinopeptide B at 1 pm/μl (F3261, Sigma-Aldrich, St. Louis, MI) the  $m+2H/2 = 785.8$   $m/z$  was acquired in product ion mode to obtain the fragment 175.1190 and 1285.5444  $m/z$  ions used for the two point calibration. Computer controlled data dependent automated switching to MS/MS provided peptide sequence information. AnalystQS 1.1 software (Applied Biosystems/MDS Sciex) was used for data acquisition. Data processing and databank searching were performed with Mascot software (Matrix Science, Beachwood, OH). The protein database provided by the 3 frame translations of the clusters shown in supplemental table S1, plus proteins shown in supplemental table S2 were used for the search analysis.

The peptides identified by MS were converted to Prosite block format<sup>45</sup> by a program written in Visual Basic. Each MS/MS derived sequence can be identified by 2 fields: (1) the gel band it has originated, and (2) the order of acquisition of the sequence information for that gel band. For example, peptide 2\_72 represents fragment number 72 derived from gel band 2. This database was used to search matches in the Fasta-formatted database of salivary proteins, using the poorly documented program Seedtop, which is part of the Blast package. The result of the Seedtop search is piped into the hyperlinked spreadsheet to produce a text file, such as the one shown for the nitrophorin CL-4-23-17, shown as a hyperlink here. Because the same tryptic fragment can be found in many gel bands, another program was written to count the number of fragments for each gel band, displaying a summarized result in an Excel table, such as in column AB of Supplemental Table S2. The summary, in the form of 13 -> 17| 14 -> 13| 12 -> 12|, indicates that 17 fragments were found in Band 13, while 13 and 12 peptides were found in bands 14 and 12, respectively. Furthermore, this summary included protein identification only when two or more peptide matches to the protein were obtained from the same gel slice. The summary program also produces additional spreadsheet cells with the larger number of peptides found in a single gel band, and the percent AA sequence coverage of the sum of the peptide matches, thus facilitating data analysis.

## Results and Discussion

### cDNA Library Characteristics

A total of 1,960 clones were sequenced and assembled to produce 960 clusters of related sequences, 837 of which contained only one EST (Supplemental Table S1). The consensus sequence of each cluster is named either a contig (deriving from two or more sequences) or a singleton (deriving from a single sequence). For sake of simplicity, this paper uses 'cluster' to denote sequences derived from both consensus sequences and singletons. The 960 clusters were compared using the program BlastX, BlastN, or RPSBlast<sup>39</sup> to the nonredundant protein database of the NCBI (NR database), the gene ontology database,<sup>40</sup> the conserved domains database of the NCBI,<sup>44</sup> and a custom-prepared subset of the NCBI nucleotide database containing either mitochondrial or rRNA sequences.

To help identify putative secreted proteins from the assembled cDNA database, we first obtained the protein sequences derived from the nucleotide sequences. Because the libraries used are unidirectional, three-frame translations of the dataset were derived; peptide sequences starting with a methionine and longer than 40 AA residues were submitted to SignalP server<sup>46</sup>. The EST assembly, BLAST, and signal peptide results were loaded into an Excel spreadsheet for manual annotation and are provided in Supplemental Table S1.

Six categories of expressed genes derived from the manual annotation of the contigs were created (Table 1). The putatively secreted (S) category contained 7% of the clusters and 36% of the sequences, with an average number of 10.3 sequences per cluster. The housekeeping (H) category had 47% and 40% of the clusters and sequences, respectively, and an average

of 1.8 sequences per cluster. Forty four percent of the clusters, containing 23% of all sequences, were classified as unknown (U), because no functional assignment could be made. This category had an average of 1.1 sequences per cluster. A good proportion of these transcripts could derive from 3' or 5' untranslated regions of genes of the above two categories, as was recently indicated for a sialotranscriptome of *An. gambiae*.<sup>48</sup> We additionally identified sequences associated to transposable elements, virus (V) and bacteria, as will be described further below (Table 1).

### Housekeeping (H) Genes

The 452 clusters (comprising 796 ESTs) attributed to H genes expressed in the salivary glands of *C. lectularius* were further divided into 19 subgroups according to function (Table 2). Not surprisingly for an organ specialized for the secretion of polypeptides, among the top four sets were transcripts associated with protein synthesis machinery (106 clusters containing 247 ESTs), protein modification machinery (42 clusters and 79 ESTs), energy metabolism (38 clusters containing 73 ESTs), a pattern also observed in other sialotranscriptomes.<sup>49-51</sup> We have arbitrarily included a group of 111 ESTs (88 clusters) in the H category that represent highly conserved proteins of unknown function, presumably associated with cellular function. They are named conserved proteins of unknown function in Supplemental Table S1. These sets may help functional identification of the 'conserved hypothetical' proteins as previously reviewed by Galperin and Koonin.<sup>52</sup> The complete list of all 452 gene clusters, along with further information about each, is given in Supplemental Table S1.

### Transposable elements (TE) and viral (V) and microbial (M) sequences

A total of 14 clusters, all comprised by singletons, were associated with parasitic or microbial sequences, as follows: Cimex-contig\_591 matches strongly the gene transfer agent protein of bacteria, possibly representing a contaminant or symbiont. Cluster 184 produces strong similarity to *Oita rhabdovirus* nucleoprotein, and to the PFAM domain indicative of Rhabdovirus nucleocapsid protein. Finally, 14 singletons match diverse transposable elements, both class I (Bel family) and class II (Penelope and Mariner) indicative of active transposition, or more likely, of messages that regulate or suppress TE mobilization.

### Possibly Secreted (S) Class of Expressed Genes

Inspection of Supplemental Table S1 indicates the expression of ubiquitous gene families, including those coding for Serpins, antigen-5 family members, odorant-binding proteins (the ESTs of which constitute 42% of the S class), lysozyme, and enzymes such as the *Cimex* type apyrase, serine proteases, esterases and diadenosine phosphatases (Table 3). This last enzyme, albeit ubiquitous as a housekeeping agent, is a first find in hematophagous arthropod sialomes. Additionally, ESTs coding for the unique *Cimex* nitrophorin contributes to 37 % of the S group, where several isoforms are identifiable, as well as for proteins with unique sequences.

### Analysis of the *C. lectularius* Sialome

Several clusters of sequences coding for housekeeping and putative secreted polypeptides indicated in Supplemental Table S1 are abundant and complete enough to extract novel consensus sequences. A total of 100 novel sequences, 46 of which code for putative secreted proteins, are grouped together in Supplemental Table S2. With this database in hand, we characterized the salivary proteome via analysis of SDS-PAGE separated proteins and MS (Figure 1). The results of this run (referred ad PAGE/MS/MS run in the text below) are integrated within the description of the deduced proteins from the transcriptome analysis, as outlined below.

## Enzymes

### **Cimex nitrophorin (NP) and the salivary inositol polyphosphate phosphatase (IPPase)**

**family:** IPPases are ubiquitous normally intracellular enzymes associated with the catabolism of the inositol phosphates, important in the regulation of these signaling molecules.<sup>53, 54</sup> Remarkably, salivary IPPases with clear signal peptide of secretion have been discovered in the sialotranscriptomes of *Rhodnius prolixus*,<sup>50</sup> *Triatoma brasiliensis*<sup>55</sup> and *T. infestans*<sup>56</sup>, but not in other insect (mosquitoes, sand flies, black flies) or tick sialotranscriptomes. The *Rhodnius* recombinant enzyme was expressed, and its kinetics studied.<sup>57</sup> Also structurally associated with the IPPase family is the salivary vasodilator of *C. lectularius*; nearly 15 years ago it was identified as the unstable gas NO,<sup>21</sup> carried from the salivary glands to the host tissues by a heme protein, the mRNA of which was cloned and sequenced.<sup>20</sup> The crystal structure of the recombinant protein was later determined,<sup>19</sup> showing it to be similar to inositol phosphate phosphatase, which are proteins not containing heme. The *C. lectularius* sialotranscriptome revealed 14 clusters totaling 258 ESTs matching the previously described *Cimex* NP (gi3388165), plus two additional clusters, totaling only 3 EST's, matching primarily salivary IPPases previously described in triatomine sialotranscriptomes.

Phylogenetic analysis based on the alignment of 8 deduced (except for one full length, the remaining are truncated products) protein sequences most similar to *Cimex* NP, with 2 other *Cimex* sequences most similar to IPPases, and published triatomine and bee IPPases produces a clade with strong bootstrap support containing the triatomine, bee (an intracellular enzyme) and the two *Cimex* proteins most similar to IPPase (marked as IPPase in Figure 2A), and one additional clade containing the original *Cimex* NP (marked Nitrophorin on Figure 2). The nitrophorin clade divides itself in two subclades (marked I and II in Figure 2A), suggesting the existence of at least 2 genes coding for the salivary NP in *Cimex*. The sequence CL-4-43-1 does not group significantly with any of the clades. From the determined crystal structure of *Cimex* NO, ten amino acids were identified making contact with the heme moiety, responsible for NO ligation.<sup>19</sup> The NP I clade have all these 10 amino acids (CL-11, CL-4-19-11 are truncated and show 9 and 8 amino acids only), while the NP II clade had substitutions on the last 3 residues (Figure 2B). Importantly, Cys 60 is conserved in both clades, this residue being identified to form a S-nitroso derivative during NO release from the nitrophorin.<sup>19</sup> The position of this Cys in the NP clade is substituted by a conserved Trp in all other non-NP clades, which also have a conserved Pro in the location of the first conserved Ala (Figure 2B). The *Cimex* IPPases and the orphan CL-4-43-1 do not have conservation of the heme-contacting 10 amino acids, except for the second Val in CL-4-43-1, indicating they may not have a heme group (Figure 2B).

Members of the nitrophorin family were detected in the PAGE/MS/MS run (Figure 1) on fractions 12-14. The strong band on fraction 13, with a gel location compatible with the expected mass of nitrophorins (32 kDa), most probably represent members of the nitrophorin family. The fragment coverage by all NP members reported on supplemental table S2 ranged from 36 to 95%.

As indicated above, the *Cimex* sialotranscriptome reveals 3 transcripts coding for 5' truncated proteins similar to the *Triatoma* and *Rhodnius* enzymes. Peptidic sequences deduced from the PAGE/MS/MS run reported in Figure 1 matched *Cimex* IPPase (on fraction 11), at a gel region indicative of 40-45 kDa masses. The function of this enzyme in blood feeding remains unknown, because the substrates, which are important in signal transduction, are intracellular.

**Apyrase:** A total of 15 ESTs were found coding for the previously described apyrase of *C. lectularius*. The translated products were all equal or higher than 95% identical to the

described enzyme (gi|74764847),<sup>16, 58</sup> suggesting that a single gene may code for this activity. The PAGE/MS/MS run (Figure 1 and supplemental table S1) reported many tryptic fragments for the enzyme at fraction 11, coincident with a well stained band above the 38 kDa marker, in accordance with the predicted 39 kDa mature molecular weight of the enzyme. As previously described,<sup>16, 58</sup> this enzyme is an inhibitor of platelet aggregation.

**Diadenosine phosphatase:** Eight transcripts code for members of the KOG family for Diadenosine and diphosphoinositol polyphosphate phosphohydrolase, a relatively ubiquitous protein family associated with signal transduction and never described before in hematophagous arthropod sialomes. Two protein sequences can be deduced from these transcripts, one of which, deriving from a single EST, is truncated. These two proteins are less than 60% identical at the amino acid sequence level, indicating that at least two genes code for these proteins. The PAGE/MS/MS run (Figure 1) identified 4 ions on fraction 19 matching the most abundantly transcribed product. The location of fraction 19 lies just below the 14 kDa marker, consistent with the predicted molecular weight of the mature polypeptide of 14.5 kDa. This enzyme may be related to the hydrolysis of the proinflammatory diadenosine nucleotides Ap4A and Ap5A, released by platelets and neutrophils.<sup>59-61</sup>

**Esterase:** At total of 40 ESTs encode proteins having high similarity to the carboxyterminus domain of insect proteins annotated as acetylcholinesterases, a PFAM domain for carboxyl esterases, and KOG domain for Acetyl/butyryl cholinesterases. One full length sequence and one truncated, but near full length, sequence can be deduced from these ESTs. These two sequences are 96% identical at the amino acid level, and may result as alleles from the same gene, or from gene duplication. Alignment of the *Cimex* sequences with vertebrate and invertebrate sequences for which the crystal structure is known (*Torpedo californica*, mouse and *D. melanogaster*)<sup>62-64</sup> reveals conservation of the 3 amino acids involved in catalysis, as well as conserved amino acids in the deep cliff where the enzyme's active center is located (Figure 3). A model of the *C. lectularius* salivary acetylcholinesterase protein was constructed using the automated Phyre fold recognition server (Figure 4).<sup>65</sup> Direct comparisons with the structure of the well-characterized enzyme from the electric eel *Torpedo californica*<sup>66</sup> show that while the catalytic triad consisting of Ser 201, His 443 and Glu 329 would be expected to be present in any enzyme with an esterase function, the *Cimex* protein also has a clearly defined "anionic" site needed to accommodate the formal positive charge on the quaternary nitrogen of the choline moiety. Trp 83, which is conserved in both the *Torpedo* and *Cimex* sequences has been shown to form a cation- $\pi$  interaction involving the quaternary nitrogen of choline. A subsidiary acetylcholine binding site has also been detected in the *Torpedo* enzyme which is thought to contribute to the extremely high catalytic rates seen for this enzyme. In this case, the quaternary nitrogen is also stabilized by a cation- $\pi$  interaction with a tryptophan residue. This residue, Trp 280, is also conserved in the *Cimex* protein, suggesting that this enzyme hydrolyzes either acetylcholine, or a closely related compound. While carboxylesterases and lipases have been commonly found in other insect sialotranscriptomes, such as the salivary carboxylesterase of *Aedes aegypti*, none found so far are so clearly of the acetyl/butyryl cholinesterase subfamily as the *Cimex* enzyme. The function of this enzyme in blood feeding remains to be identified.

**Serine proteases:** Trypsin or chymotrypsin-like enzymes are a common finding in hematophagous insect sialotranscriptomes.<sup>13</sup> These enzymes are postulated to act either as activators of immune cascades or to directly affect hemostasis by destroying fibrin, as is the case of tabylsins from a tabanid.<sup>67</sup> The sialotranscriptome of *C. lectularius* provides 18 ESTs coding for trypsin-like enzymes, including 16 that fall into a single cluster, from which a truncated product with 301 amino acid residues can be deduced. The PAGE/MS/MS



run (Figure 1) provided 40% coverage for the deduced protein, in a gel location (fraction 12) consistent with a mass near 38 kDa.

**Serpins**—Serpins constitute a protein family containing serine proteinase inhibitors that are widespread in all three domains of life, and participate in the regulation of immune activation and clotting cascades in animals.<sup>68, 69</sup> The main salivary anticlotting agent of the mosquito *Ae. aegypti*, an inhibitor of the clotting factor Xa, was shown to be a serpin.<sup>70</sup> The *Cimex* sialotranscriptome provides for 6 transcripts coding for a serpin, from which an amino truncated peptide of 311 residues can be deduced. Sixteen peptidic fragments covering 72% of these 311 residues were found in fraction 10 of the PAGE/MS/MS run, which also detected 13 fragments in fraction 9, indicating a molecular mass of 45-49 kDa. The actual target of this serpin is not clear, because the main anticlotting of *C. lectularius* was reported to have a 16 kDa mass, and to inhibit activation of X to Xa, but not Xa or other clotting proteases directly.<sup>18</sup>

**Antigen 5 related proteins**—This ubiquitous protein family belong to the CAP family (Cys-rich secretory proteins; AG5 proteins of insects; pathogenesis-related protein 1 of plants).<sup>71</sup> Virtually all hematophagous arthropod sialotranscriptomes reported thus far contain members of this family. The majority of these animal proteins have no known function. The notable exceptions include proteolytic activity in cone snail,<sup>72</sup> smooth muscle-relaxing activity in snake venoms,<sup>73, 74</sup> salivary neurotoxin activity in the venomous lizard *Heloderma horridum*<sup>75</sup> and immunoglobulin binding by a salivary protein of the stable fly, *Stomoxys calcitrans*.<sup>76</sup> Recently, an antigen-5 protein from the saliva of a tabanid fly<sup>77</sup> was shown to inhibit platelet aggregation by the unusual acquisition of a typical RGD domain that is known to prevent fibrinogen binding to platelets and its ensuing aggregation.<sup>78</sup> No function has been determined for any hemipteran antigen-5 protein. The sialotranscriptome of *C. lectularius* contain 14 ESTs producing matches to members of this protein family, including matches to this family's PFAM domain. All these transcripts are truncated at their 5' end, but allow indication of at least 3 different members of this family to be present in the transcriptome.

**Odorant binding protein family**—The odorant binding protein (OBP) family is ubiquitous in insects where it acts as carrier of lipids in their hemolymph, or of odorants in chemoreceptors.<sup>79</sup> Hematophagous Nematocera have abundant expression of a modified member of this protein family in their salivary glands, known as the D7 family, where long and short forms exist, the long form containing two repeated domains.<sup>80</sup> There are multiple genes coding for D7 proteins in mosquitoes. In *An. gambiae*, 3 genes code for the long forms and 5 genes code for short D7 proteins.<sup>48</sup> *Aedes aegypti* similarly has multiple genes coding for this protein family.<sup>81</sup> The crystal structure of classic OBP and mosquito D7 proteins has been solved, showing that the D7 fold has acquired 2 additional helices when compared to canonical OBP.<sup>82, 83</sup> In mosquitoes, these proteins function as kratoagonists of biogenic amines and lipidic mediators of inflammation. The *Cimex* sialotranscriptome presented 296 EST's coding for proteins having the OBP domain, representing 42 % of the transcripts associated to the S class (Table 3). Fourteen coding sequences were deduced from the assembled clusters (supplemental table S2). Phylogenetic analysis, using a *Drosophila mojavensis* sequence as an out group, shows that these 14 sequences group into 4 major clades labeled I-IV in Figure 5. The PAGE/MS/MS run shows matches for several OBP proteins in gel fractions 16-19 (Figure 1, supplemental tables S1 and S2). The strong band in fraction 18 (Figure 1) most probably represents members of this family, located above the 14 kDa marker, as expected by the predicted MW of the mature proteins (15-15.6 kDa). The role of this abundant salivary protein family is not known, but they may function

as kratagonists.<sup>13</sup> On the other hand, the main ant clotting of *C. lectularius* was reported to have a 16 kDa mass, which falls in the range of this protein family.

### Other peptides

**Lysozyme:** Lysozyme and other antibacterial peptides have been commonly found in hematophagous insect sialotranscriptomes. The *Cimex* sialotranscriptome revealed 27 ESTs assembling in 14 clusters that match lysozymes when the tool blastx is used to query the sequences against the NR protein database. The deduced coding sequences reveal at least 4 truncated polypeptides that are more than 40% diverse among themselves (supplemental table S2), indicating the expression of a multi-gene family. Lysozyme expression was detected by the PAGE/MS/MS run in bands 19 and 20, consistent with the predicted mass of this class of proteins. Lysozyme-like immune activity and antimicrobial activity were recently identified in the ejaculate of male *C. lectularius*.<sup>84</sup>

**Mucin:** Mucins are serine and/or threonine rich proteins that have O-linked N-acetyl-galactosamine residues. These proteins are commonly expressed in moist epithelia where they confer surface protection.<sup>85, 86</sup> They are commonly found in arthropod sialotranscriptomes, where they may have a yet unknown pharmacological function, and also may help to protect the inner portions of the salivary ducts. The *Cimex* sialotranscriptome provided for 13 ESTs that assembled into a single cluster from where the CDS for a truncated serine rich protein is deduced. The amino terminal sequence is missing thus it is not possible to determine whether this protein is secreted.

**Hypothetical Cimicomorpha secreted protein:** The deduced truncated protein sequence named CL-4-24-29 matches the carboxyterminus of a previously reported secreted polypeptide from the salivary glands of *T. infestans*, which provides no significant matches to other animal proteins in the NR database. This *Cimex* sequence thus deorphanizes the previously described kissing bug protein, indicating it may be a protein typical to Cimicomorpha.

**Hypothetical conserved secreted insect protein:** The protein sequence CL-4-43-48 derives from a cluster of 7 ESTs coding for a secreted peptide of predicted mature molecular mass of 12.7 kDa that is similar to previously described proteins found in *Triatoma infestans* sialotranscriptome, sand fly midgut transcriptomes,<sup>87, 88</sup> and many other hypothetical proteins of similar size deduced mostly from insect genomes, of unknown function. Psiblast analysis of this protein sequence against the NR database, using the switches  $-e\ 1e-4 -h\ 1e-10$  converges after 6 iterations, retrieving 121 peptides of similar size (120-150 amino acids), all from insects, except two that are derived from the fungus *Streptomyces clavuligerus* (Psiblast results available on supplemental file S3, and clustal alignment on file S4). Expression of this peptide was detected in fraction 17 of the PAGE/MS/MS run. This peptide ubiquity and relative abundant expression and a signal sequence indicative of secretion suggests an antimicrobial role.

**Cimex orphan putative secreted peptide:** CL-4-42-4 codes for a putative secreted peptide that has no significant matches when compared to the NR database. Its function is unknown.

### Housekeeping Proteins

Supplemental Table S2 presents sequence information on 54 proteins classified as housekeeping, several of which were identified in the proteome run (Supplemental table S2 - not shown on Figure 1). These housekeeping transcripts can be used to contrast their codon usage to those coding for secreted products, which are exposed to the pressure of their host's immune system.

Previous hematophagous insect sialotranscriptomes suggested that salivary proteins are at a fast pace of evolution, explaining the large degree of polymorphism among salivary proteins, and the appearance of completely novel sequences.<sup>13</sup> It has been suggested that codon volatility (the proportion of the point-mutation neighbors of a codon that encode different amino acids) could be a measure of selection for fast evolving proteins, as could occur in pathogens in a constant avoidance of antibody recognition.<sup>89</sup> Although this intuitive idea has created strong opposition,<sup>90-92</sup> it is supported by published models.<sup>93, 94</sup> We accordingly measured the average codon volatility for the 54 sequences coding for housekeeping and 46 sequences coding for putative secreted proteins shown in table S2. The average codon volatility for the H class genes was  $0.7717 \pm 0.0017$  while the S class had an average volatility of  $0.7820 \pm 0.0020$  (Average  $\pm$  SE), a highly significant result ( $P < 0.001$ , Mann-Whitney rank sum test). Regardless of the discussion of the value of this index, the result above indicates that a single point mutation on an S class gene has a significantly higher chance of producing a non-synonymous amino acid substitution than in an H class gene. Similar results were also found recently on the sialotranscriptomes of the mosquitoes *Culex tarsalis* and *Ochlerotatus triseriatus* (submitted).

## Conclusion

The evolution to a blood diet occurred several times within arthropods, and for this reason there is a great diversity of salivary composition among arthropods not sharing a common blood feeding ancestor. Even with a common hematophagous ancestor, different genera within the same family may have unique salivary components, because the salivary proteins are at a fast pace of evolution due to the immune pressure of their hosts, and because hosts also evolve their hemostatic physiology. For example, the mosquito genera *Aedes* and *Anopheles* share a common ancestor ~ 150 million years ago,<sup>95</sup> near 100 million years before the radiation of mammals and the feeding problem caused by the efficient mammalian blood platelet. Accordingly, following the menu change occurring after the dinosaur extinction, *Aedes* and *Anopheles* had to “invent” new ways to fight the more efficient hemostasis of mammals, possibly explaining the appearance of unique salivary proteins in each of these two genera.<sup>13</sup>

Because *Cimex lectularius* does not share a common ancestor with any other blood feeding insect for which the sialome is known, a number of unique proteins characterize this insect sialome, such as the previously reported nitrophorin,<sup>20</sup> which we here show to be a multi gene family. Two ubiquitous enzymes never described before in sialotranscriptomes were also identified: (a) the abundantly expressed esterase of the acetyl/butryl cholinesterase, which may hydrolyze PAF, and (b) a secreted Nudix type hydrolase that may hydrolyze diadenosine nucleotides. A novel peptide (CL-4-42-4) of unknown function was also identified.

Uniquely Cimicomorpha are the finding of (a) inositol phosphate polyphosphatase enzymes, found in the present work and also in the sialotranscriptomes of *Rhodnius prolixus*, *Triatoma infestans* and *T. brasiliensis* and in no other known sialotranscriptome, and (b) of a *Cimex* protein (Cl-4-24-29) uniquely matching a *T. infestans* salivary protein of unknown function.

It is interesting that the salivary apyrase of *Cimex*<sup>58</sup>, belonging to an ubiquitous and most intracellular protein family, was also recruited to function as a salivary apyrase in sand flies (but not on any other known blood feeder, so far),<sup>96, 97</sup> and that the multi gene expansion of the odorant binding family found in *Cimex* is also found expanded in the black fly *Simulium vittatum* sialome.<sup>98</sup> The OBP gene expansion is paralleled by the D7 gene family in mosquitoes, a protein family related to the OBP, and by the lipocalin family in triatomines

and ticks.<sup>13</sup> These coincidences can be ascribed as products of convergent evolution since *Cimex* does not share a common blood feeding ancestor with any of the named insects.

Ubiquitously found in hematophagous sialomes are the finding of antigen-5 related proteins, of salivary function mostly unknown (except, as mentioned above, for a protein that incorporated an RGD domain in tabanids and is an inhibitor of platelet aggregation, and a stable fly protein that binds immunoglobulins),<sup>76, 77</sup> serine proteases, lysozyme and mucin. We also identified in *Cimex* a small protein of unknown function, similar to proteins previously described in salivary and midgut transcriptomes of blood sucking insects and to hypothetical proteins discovered in insect genomes, abundantly so in *Drosophila*, which we postulate to be a novel antimicrobial peptide family.

The finding of multigenic families reaffirms the importance of gene duplication events in evolution in general and in the evolution of blood feeding in particular.<sup>99</sup> Gene duplication initially affects transcript abundance, which may lead to a beneficial increase in protein expression. The newly acquired paralogous genes are then free to evolve divergently, while their regulatory regions may independently evolve to increase or regulate their transcription. It is interesting that one multi family that we identified as expanded in *Cimex* sialome, the OBP family, is a family normally associated with odorant binding receptors, which is also a fast evolving multi family, as recently reviewed.<sup>100</sup> The fast evolution of salivary proteins is supported by the increased codon volatility of salivary messages when compared to housekeeping-associated sequences.

Since this is a discovery based work, the discussion should propose useful hypothesis or questions for conventional, Baconian hypothesis-driven research: (1) does *Cimex* saliva inhibits PAF, and if so, is it due to the acetylcholinesterase? (2) Does *Cimex* saliva hydrolyzes adenosine dinucleotides, and if so is it due to the Nudix hydrolase unique to this sialome? (3) Does the peptide CL-4-43-8, similar to other peptides found in sialomes and expressed in insect midguts, is part of a novel antimicrobial family? (4) What is the target of the *Cimex* salivary serpin? (5) Are the OBPs found in *Cimex* also working as kratagonists of hemostasis and inflammation, and if so, which ones? How prevalent is the virus coding for the singleton EST most certainly from a rhabdovirus? Additionally, there are mysterious proteins that we cannot at present even articulate a hypothesis, such as the additional properties that might have been acquired during evolution by the OBPs, or AG-5 or the uniquely culicomorpha protein (CL-4-24-29). Independent of their function, these proteins may also be used for immune detection of humans and animals to bed bug exposure, or as part of desensitization vaccines.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are grateful for the expert assistance of Dr. Carl Hammer from the NIAID Research Technology Branch. This work was supported by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, and funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the government of the United States of America.

Because J.F.A., I.M.B.F, J.G.V., V.M.P., Z.M., and J.M.C.R. are government employees and this is a government work, the work is in the public domain in the United States. Notwithstanding any other agreements, the NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may

tag or modify the work consistent with its customary practices. You can establish rights outside of the U.S. subject to a government use license.

## References

1. Grimaldi, D.; Engel, M. Evolution of the insects. Cambridge University Press; New York: 2005. p. 772
2. Schofield CJ, Galvao C. Classification, evolution, and species groups within the Triatominae. *Acta Trop.* 2009; 110(2-3):88–100. [PubMed: 19385053]
3. Reinhardt K, Siva-Jothy MT. Biology of the bed bugs (Cimicidae). *Annu Rev Entomol.* 2007; 52:351–74. [PubMed: 16968204]
4. Doggett SL, Geary MJ, Russell RC. The resurgence of bed bugs in Australia: with notes on their ecology and control. *Environ Health.* 2004; 4:30–38.
5. Ter Poorten MC, Prose NS. The return of the common bedbug. *Pediatr Dermatol.* 2005; 22(3):183–7. [PubMed: 15916561]
6. Lee IY, Ree HI, An SJ, Linton JA, Yong TS. Reemergence of the bedbug *Cimex lectularius* in Seoul, Korea. *Korean J Parasitol.* 2008; 46(4):269–71. [PubMed: 19127335]
7. Anderson AL, Leffler K. Bedbug infestations in the news: a picture of an emerging public health problem in the United States. *J Environ Health.* 2008; 70(9):24–7. 52–3. [PubMed: 18517150]
8. Abdel-Naser MB, Lotfy RA, Al-Sherbiny MM, Sayed Ali NM. Patients with papular urticaria have IgG antibodies to bedbug (*Cimex lectularius*) antigens. *Parasitol Res.* 2006; 98(6):550–6. [PubMed: 16425065]
9. Leverkus M, Jochim RC, Schad S, Brocker EB, Andersen JF, Valenzuela JG, Trautmann A. Bullous allergic hypersensitivity to bed bug bites mediated by IgE against salivary nitrophorin. *J Invest Dermatol.* 2006; 126(1):91–6. [PubMed: 16417223]
10. Scarupa MD, Economides A. Bedbug bites masquerading as urticaria. *J Allergy Clin Immunol.* 2006; 117(6):1508–9. [PubMed: 16751024]
11. Stucki A, Ludwig R. Images in clinical medicine. Bedbug bites. *N Engl J Med.* 2008; 359(10): 1047. [PubMed: 18768948]
12. Reinhardt K, Kempke D, Naylor RA, Siva-Jothy MT. Sensitivity to bites by the bedbug, *Cimex lectularius*. *Med Vet Entomol.* 2009; 23(2):163–6. [PubMed: 19292820]
13. Ribeiro JMC, Arca B. From sialomes to the sialoverse: An insight into the salivary potion of blood feeding insects. *Adv Insect Physiol.* 2009; 37:59–118.
14. Ribeiro JM, Francischetti IM. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu Rev Entomol.* 2003; 48:73–88. [PubMed: 12194906]
15. Valenzuela JG, Charlab R, Galperin MY, Ribeiro JM. Purification, cloning, and expression of an apyrase from the bed bug *Cimex lectularius*. A new type of nucleotide-binding enzyme. *J Biol Chem.* 1998; 273(46):30583–90. [PubMed: 9804829]
16. Valenzuela JG, Chuffe OM, Ribeiro JMC. Apyrase and anti-platelet activities from the salivary glands of the bed bug *Cimex lectularius*. *Insect Biochem Mol Biol.* 1996; 21:557–562.
17. Ribeiro JMC. Blood-feeding arthropods: Live syringes or invertebrate pharmacologists? *Infect Agents Dis.* 1995; 4:143–152. [PubMed: 8548192]
18. Valenzuela JG, Guimaraes JA, Ribeiro JM. A novel inhibitor of factor X activation from the salivary glands of the bed bug *Cimex lectularius*. *Exp Parasitol.* 1996; 83(2):184–90. [PubMed: 8682187]
19. Weichsel A, Maes EM, Andersen JF, Valenzuela JG, Shokhireva T, Walker FA, Montfort WR. Heme-assisted S-nitrosation of a proximal thiolate in a nitric oxide transport protein. *Proc Natl Acad Sci U S A.* 2005; 102(3):594–9. [PubMed: 15637157]
20. Valenzuela JG, Ribeiro JM. Purification and cloning of the salivary nitrophorin from the hemipteran *Cimex lectularius*. *J Exp Biol.* 1998; 201(Pt 18):2659–64. [PubMed: 9716517]
21. Valenzuela JG, Walker FA, Ribeiro JM. A salivary nitrophorin (nitric-oxide-carrying hemoprotein) in the bedbug *Cimex lectularius*. *J Exp Biol.* 1995; 198(Pt 7):1519–26. [PubMed: 7658188]
22. Moncada S, Palmer RMJ, Higgs EA. Nitric oxide: Physiology, pathophysiology, and pharmacology. *Pharmacol Rev.* 1991; 43:109–142. [PubMed: 1852778]

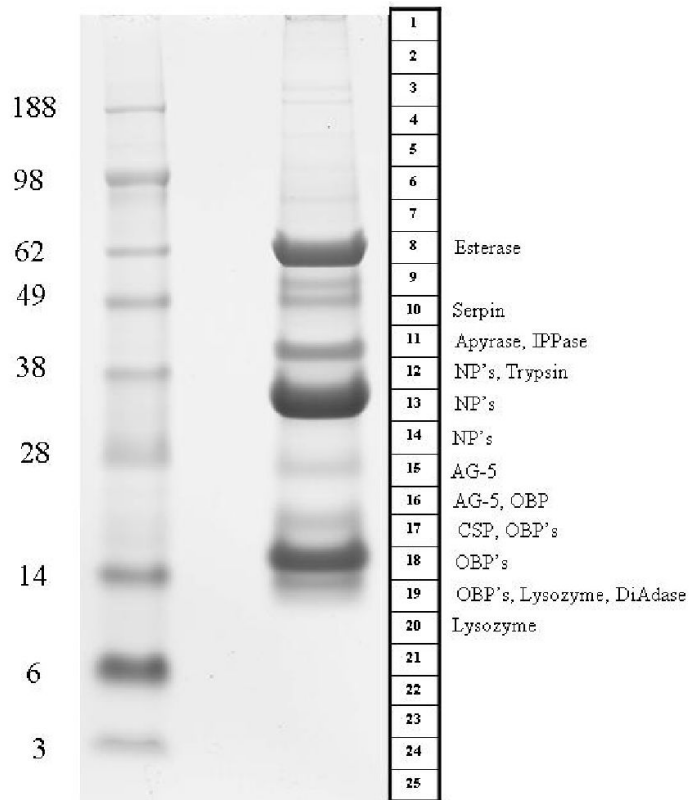
23. Champagne D, Nussenzweig RH, Ribeiro JMC. Purification, characterization, and cloning of nitric oxide-carrying heme proteins (nitrophorins) from salivary glands of the blood sucking insect *Rhodnius prolixus*. *J Biol Chem*. 1995; 270:8691–8695. [PubMed: 7721773]
24. Ribeiro JMC, Walker FA. High affinity histamine-binding and anti-histaminic activity of the salivary NO-carrying heme protein (Nitrophorin) of *Rhodnius prolixus*. *J Exp Med*. 1994; 180:2251–2257. [PubMed: 7964498]
25. Andersen JF, Francischetti IM, Valenzuela JG, Schuck P, Ribeiro JM. Inhibition of hemostasis by a high affinity biogenic amine-binding protein from the saliva of a blood-feeding insect. *J Biol Chem*. 2003; 278(7):4611–7. [PubMed: 12464610]
26. Francischetti IM, Andersen JF, Ribeiro JM. Biochemical and functional characterization of recombinant *Rhodnius prolixus* platelet aggregation inhibitor 1 as a novel lipocalin with high affinity for adenosine diphosphate and other adenine nucleotides. *Biochemistry*. 2002; 41(11): 3810–8. [PubMed: 11888300]
27. Francischetti IM, Ribeiro JM, Champagne D, Andersen J. Purification, cloning, expression, and mechanism of action of a novel platelet aggregation inhibitor from the salivary gland of the blood-sucking bug, *Rhodnius prolixus*. *J Biol Chem*. 2000; 275(17):12639–50. [PubMed: 10777556]
28. Gudderra NP, Ribeiro JM, Andersen JF. Structural determinants of factor IX(a) binding in nitrophorin 2, a lipocalin inhibitor of the intrinsic coagulation pathway. *J Biol Chem*. 2005; 280(26):25022–8. [PubMed: 15866866]
29. Zhang Y, Ribeiro JM, Guimaraes JA, Walsh PN. Nitrophorin-2: a novel mixed-type reversible specific inhibitor of the intrinsic factor-X activating complex. *Biochemistry*. 1998; 37(30):10681–90. [PubMed: 9692958]
30. Ribeiro JMC, Schneider M, Guimaraes JA. Purification and characterization of Prolixin S (Nitrophorin 2), the salivary anticoagulant of the blood sucking bug, *Rhodnius prolixus*. *Biochem J*. 1995; 308:243–249. [PubMed: 7755571]
31. Bartley JD, Harlan HJ. Bed bug infestation: its control and management. *Military Medic*. 1974 November.;884–6.
32. Montes C, Cuadrillero C, Vilella D. Maintenance of a laboratory colony of *Cimex lectularius* (Hemiptera: Cimicidae) using an artificial feeding technique. *J Med Entomol*. 2002; 39(4):675–9. [PubMed: 12144303]
33. Guo Y, Ribeiro JM, Anderson JM, Bour S. dCAS: a desktop application for cDNA sequence annotation. *Bioinformatics*. 2009; 25(9):1195–6. [PubMed: 19318425]
34. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res*. 1999; 9(9):868–77. [PubMed: 10508846]
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3):403–10. [PubMed: 2231712]
36. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997; 25(24):4876–82. [PubMed: 9396791]
37. Page RD. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*. 1996; 12(4):357–8. [PubMed: 8902363]
38. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*. 2004; 5(2):150–63. [PubMed: 15260895]
39. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25(17):3389–402. [PubMed: 9254694]
40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25(1):25–9. [PubMed: 10802651]
41. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. *Nucleic Acids Res*. 2000; 28(1):263–6. [PubMed: 10592242]
42. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res*. 2000; 28(1):231–4. [PubMed: 10592234]

43. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 2003; 4(1):41. [PubMed: 12969510]
44. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res*. 2002; 30(1):281–3. [PubMed: 11752315]
45. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ. The PROSITE database. *Nucleic Acids Res*. 2006; 34(Database issue):D227–30. [PubMed: 16381852]
46. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*. 1997; 10(1):1–6. [PubMed: 9051728]
47. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J*. 1998; 15(2):115–30. [PubMed: 9557871]
48. Arca B, Lombardo F, Valenzuela JG, Francischetti IM, Marinotti O, Coluzzi M, Ribeiro JM. An updated catalogue of salivary gland transcripts in the adult female mosquito, *Anopheles gambiae*. *J Exp Biol*. 2005; 208(Pt 20):3971–86. [PubMed: 16215223]
49. Francischetti IM, Valenzuela JG, Pham VM, Garfield MK, Ribeiro JM. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *J Exp Biol*. 2002; 205(Pt 16):2429–51. [PubMed: 12124367]
50. Ribeiro JM, Andersen J, Silva-Neto MA, Pham VM, Garfield MK, Valenzuela JG. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem Mol Biol*. 2004; 34(1):61–79. [PubMed: 14976983]
51. Ribeiro JM, Charlab R, Pham VM, Garfield M, Valenzuela JG. An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*. *Insect Biochem Mol Biol*. 2004; 34(6):543–63. [PubMed: 15147756]
52. Galperin MY, Koonin EV. ‘Conserved hypothetical’ proteins: prioritization of targets for experimental study. *Nucleic Acids Res*. 2004; 32(18):5452–63. [PubMed: 15479782]
53. Kashiwada M, Lu P, Rothman PB. PIP3 pathway in regulatory T cells and autoimmunity. *Immunol Res*. 2007; 39(1-3):194–224. [PubMed: 17917066]
54. Harris SJ, Parry RV, Westwick J, Ward SG. Phosphoinositide lipid phosphatases: natural regulators of phosphoinositide 3-kinase signaling in T lymphocytes. *J Biol Chem*. 2008; 283(5):2465–9. [PubMed: 18073217]
55. Santos A, Ribeiro JM, Lehane MJ, Gontijo NF, Veloso AB, Sant’Anna MR, Nascimento Araujo R, Grisard EC, Pereira MH. The sialotranscriptome of the blood-sucking bug *Triatoma brasiliensis* (Hemiptera, Triatominae). *Insect Biochem Mol Biol*. 2007; 37(7):702–12. [PubMed: 17550826]
56. Assumpcao TC, Francischetti IM, Andersen JF, Schwarz A, Santana JM, Ribeiro JM. An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas’ disease. *Insect Biochem Mol Biol*. 2008; 38(2):213–32. [PubMed: 18207082]
57. Andersen JF, Ribeiro JM. A secreted salivary inositol polyphosphate 5-phosphatase from a blood-feeding insect: allosteric activation by soluble phosphoinositides and phosphatidylserine. *Biochemistry*. 2006; 45(17):5450–7. [PubMed: 16634626]
58. Valenzuela JG, Charlab R, Galperin MY, Ribeiro JM. Purification, cloning, and expression of an apyrase from the bed bug *Cimex lectularius*. A new type of nucleotide-binding enzyme. *J Biol Chem*. 1998; 273(46):30583–90. [PubMed: 9804829]
59. Kisselev LL, Justesen J, Wolfson AD, Frolova LY. Diadenosine oligophosphates (Ap(n)A), a novel class of signalling molecules? *FEBS Lett*. 1998; 427(2):157–63. [PubMed: 9607303]
60. Schluter H, Tepel M, Zidek W. Vascular actions of diadenosine phosphates. *J Auton Pharmacol*. 1996; 16(6):357–62. [PubMed: 9131416]
61. Stavrou BM. Diadenosine polyphosphates: postulated mechanisms mediating the cardiac effects. *Curr Med Chem Cardiovasc Hematol Agents*. 2003; 1(2):151–69. [PubMed: 15320695]

62. Harel M, Kryger G, Rosenberry TL, Mallender WD, Lewis T, Fletcher RJ, Guss JM, Silman I, Sussman JL. Three-dimensional structures of *Drosophila melanogaster* acetylcholinesterase and of its complexes with two potent inhibitors. *Protein Sci.* 2000; 9(6):1063–72. [PubMed: 10892800]
63. Bourne Y, Taylor P, Bougis PE, Marchot P. Crystal structure of mouse acetylcholinesterase. A peripheral site-occluding loop in a tetrameric assembly. *J Biol Chem.* 1999; 274(5):2963–70. [PubMed: 9915834]
64. Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I. Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science.* 1991; 253(5022):872–9. [PubMed: 1678899]
65. Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins.* 2008; 70(3):611–25. [PubMed: 17876813]
66. Colletier JP, Fournier D, Greenblatt HM, Stojan J, Sussman JL, Zaccai G, Silman I, Weik M. Structural insights into substrate traffic and inhibition in acetylcholinesterase. *Embo J.* 2006; 25(12):2746–56. [PubMed: 16763558]
67. Ma D, Wang Y, Yang H, Wu J, An S, Gao L, Xu X, Lai R. Anti-thrombosis repertoire of blood-feeding horsefly salivary glands. *Mol Cell Proteomics.* 2009
68. Rau JC, Beaulieu LM, Huntington JA, Church FC. Serpins in thrombosis, hemostasis and fibrinolysis. *J Thromb Haemost.* 2007; 5 1:102–15. [PubMed: 17635716]
69. Kanost MR. Serine proteinase inhibitors in arthropod immunity. *Dev Comp Immunol.* 1999; 23(4-5):291–301. [PubMed: 10426423]
70. Stark KR, James AA. Isolation and characterization of the gene encoding a novel factor Xa-directed anticoagulant from the yellow fever mosquito, *Aedes aegypti*. *J Biol Chem.* 1998; 273(33):20802–9. [PubMed: 9694825]
71. Megraw T, Kaufman TC, Kovalick GE. Sequence and expression of *Drosophila* Antigen 5-related 2, a new member of the CAP gene family. *Gene.* 1998; 222(2):297–304. [PubMed: 9831665]
72. Milne TJ, Abbenante G, Tyndall JD, Halliday J, Lewis RJ. Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J Biol Chem.* 2003; 278(33):31105–10. [PubMed: 12759345]
73. Yamazaki Y, Hyodo F, Morita T. Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. *Arch Biochem Biophys.* 2003; 412(1):133–41. [PubMed: 12646276]
74. Yamazaki Y, Morita T. Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon.* 2004; 44(3):227–31. [PubMed: 15302528]
75. Mochca-Morales J, Martin BM, Possani LD. Isolation and characterization of helothermine, a novel toxin from *Heloderma horridum horridum* (Mexican beaded lizard) venom. *Toxicon.* 1990; 28(3):299–309. [PubMed: 1693019]
76. Ameri M, Wang X, Wilkerson MJ, Kanost MR, Broce AB. An immunoglobulin binding protein (antigen 5) of the stable fly (Diptera: Muscidae) salivary gland stimulates bovine immune responses. *J Med Entomol.* 2008; 45(1):94–101. [PubMed: 18283948]
77. Xu X, Yang H, Ma D, Wu J, Wang Y, Song Y, Wang X, Lu Y, Yang J, Lai R. Toward an understanding of the molecular mechanism for successful blood feeding by coupling proteomics analysis with pharmacological testing of horsefly salivary glands. *Mol Cell Proteomics.* 2008; 7(3):582–90. [PubMed: 18087067]
78. Scarborough RM, Naughton MA, Teng W, Rose JW, Phillips DR, Nannizzi L, Arfsten A, Campbell AM, Charo IF. Design of potent and specific integrin antagonists. Peptide antagonists with high specificity for glycoprotein IIb-IIIa. *J Biol Chem.* 1993; 268(2):1066–73. [PubMed: 8419315]
79. Hekmat-Safe DS, Dorit RL, Carlson JR. Molecular evolution of odorant-binding protein genes OS-E and OS-F in *Drosophila*. *Genetics.* 2000; 155(1):117–27. [PubMed: 10790388]
80. Valenzuela JG, Charlab R, Gonzalez EC, de Miranda-Santos IK, Marinotti O, Francischetti IM, Ribeiro JM. The D7 family of salivary proteins in blood sucking diptera. *Insect Mol Biol.* 2002; 11(2):149–55. [PubMed: 11966880]

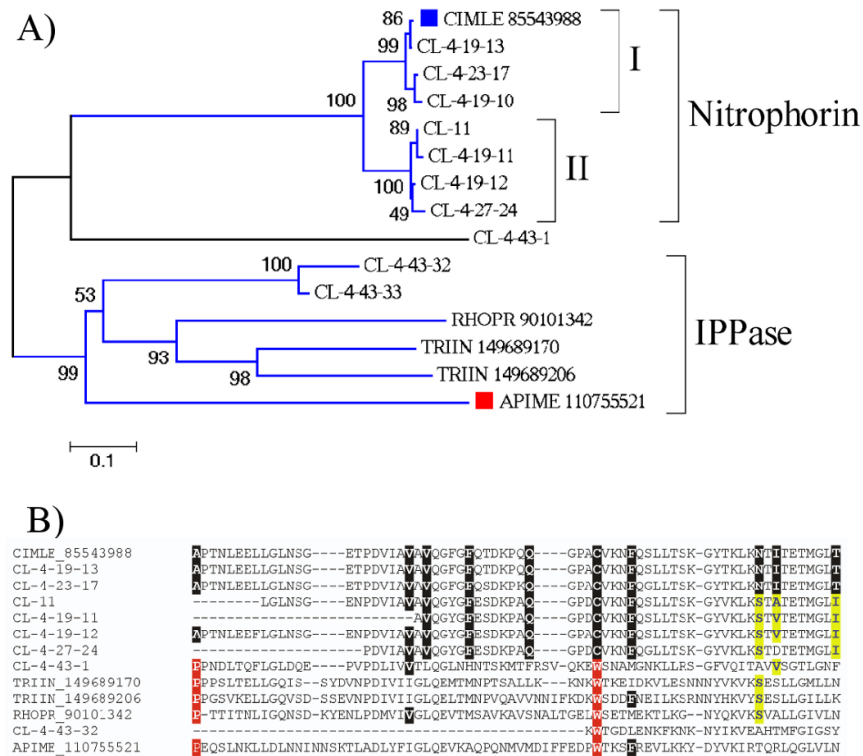


81. Ribeiro JM, Arca B, Lombardo F, Calvo E, Phan VM, Chandra PK, Wikel SK. An annotated catalogue of salivary gland transcripts in the adult female mosquito, *Aedes aegypti*. *BMC Genomics*. 2007; 8(1):6. [PubMed: 17204158]
82. Mans BJ, Calvo E, Ribeiro JM, Andersen JF. The crystal structure of D7r4, a salivary biogenic amine-binding protein from the malaria mosquito *Anopheles gambiae*. *J Biol Chem*. 2007; 282(50):36626–33. [PubMed: 17928288]
83. Calvo E, Mans BJ, Ribeiro JM, Andersen JF. Multifunctionality and mechanism of ligand binding in a mosquito antiinflammatory protein. *Proc Natl Acad Sci U S A*. 2009; 106(10):3728–33. [PubMed: 19234127]
84. Otti O, Naylor RA, Siva-Jothy MT, Reinhardt K. Bacteriolytic activity in the ejaculate of an insect. *Am Nat*. 2009; 174(2):292–5. [PubMed: 19548839]
85. Hang HC, Bertozzi CR. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem*. 2005; 13(17):5021–34. [PubMed: 16005634]
86. Fukuda M. Roles of mucin-type O-glycans in cell adhesion. *Biochim Biophys Acta*. 2002; 1573(3):394–405. [PubMed: 12417424]
87. Jochim RC, Teixeira CR, Laughinghouse A, Mu J, Oliveira F, Gomes RB, Elnaiem DE, Valenzuela JG. The midgut transcriptome of *Lutzomyia longipalpis*: comparative analysis of cDNA libraries from sugar-fed, blood-fed, post-digested and *Leishmania infantum* chagasi-infected sand flies. *BMC Genomics*. 2008; 9:15. [PubMed: 18194529]
88. Ramalho-Ortigao M, Jochim RC, Anderson JM, Lawyer PG, Pham VM, Kamhawi S, Valenzuela JG. Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania*-major-infected sandflies. *BMC Genomics*. 2007; 8:300. [PubMed: 17760985]
89. Friedman R, Hughes AL. Codon volatility as an indicator of positive selection: data from eukaryotic genome comparisons. *Mol Biol Evol*. 2005; 22(3):542–6. [PubMed: 15525698]
90. Pillai SK, Kosakovsky Pond SL, Woelk CH, Richman DD, Smith DM. Codon volatility does not reflect selective pressure on the HIV-1 genome. *Virology*. 2005; 336(2):137–43. [PubMed: 15892955]
91. Chen Y, Emerson JJ, Martin TM. Evolutionary genomics: codon volatility does not detect selection. *Nature*. 2005; 433(7023):E6–7. discussion E7-8. [PubMed: 15662371]
92. Sharp PM. Gene “volatility” is most unlikely to reveal adaptation. *Mol Biol Evol*. 2005; 22(4):807–9. [PubMed: 15616138]
93. Archetti M. Genetic robustness at the codon level as a measure of selection. *Gene*. 2009; 443(1-2):64–9. [PubMed: 19477246]
94. Plotkin JB, Dushoff J, Desai MM, Fraser HB. Codon usage and selection on proteins. *J Mol Evol*. 2006; 63(5):635–53. [PubMed: 17043750]
95. Krzywinski J, Grushko OG, Besansky NJ. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol*. 2006; 39(2):417–23. [PubMed: 16473530]
96. Valenzuela JG, Belkaid Y, Rowton E, Ribeiro JM. The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel *Cimex* family of apyrases. *J Exp Biol*. 2001; 204(Pt 2):229–37. [PubMed: 11136609]
97. Valenzuela JG, Garfield M, Rowton ED, Pham VM. Identification of the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*, vector of *Leishmania chagasi*. *J Exp Biol*. 2004; 207(Pt 21):3717–29. [PubMed: 15371479]
98. Andersen JF, Pham VM, Meng Z, Champagne DE, Ribeiro JM. Insight into the sialome of the Black Fly, *Simulium vittatum*. *J Proteome Res*. 2009; 8(3):1474–88. [PubMed: 19166301]
99. Kondrashov FA, Koonin EV. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet*. 2004; 20(7):287–90. [PubMed: 15219392]
100. Smadja C, Shi P, Butlin RK, Robertson HM. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol*. 2009; 26(9):2073–86. [PubMed: 19542205]

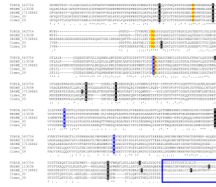


**Figure 1.**

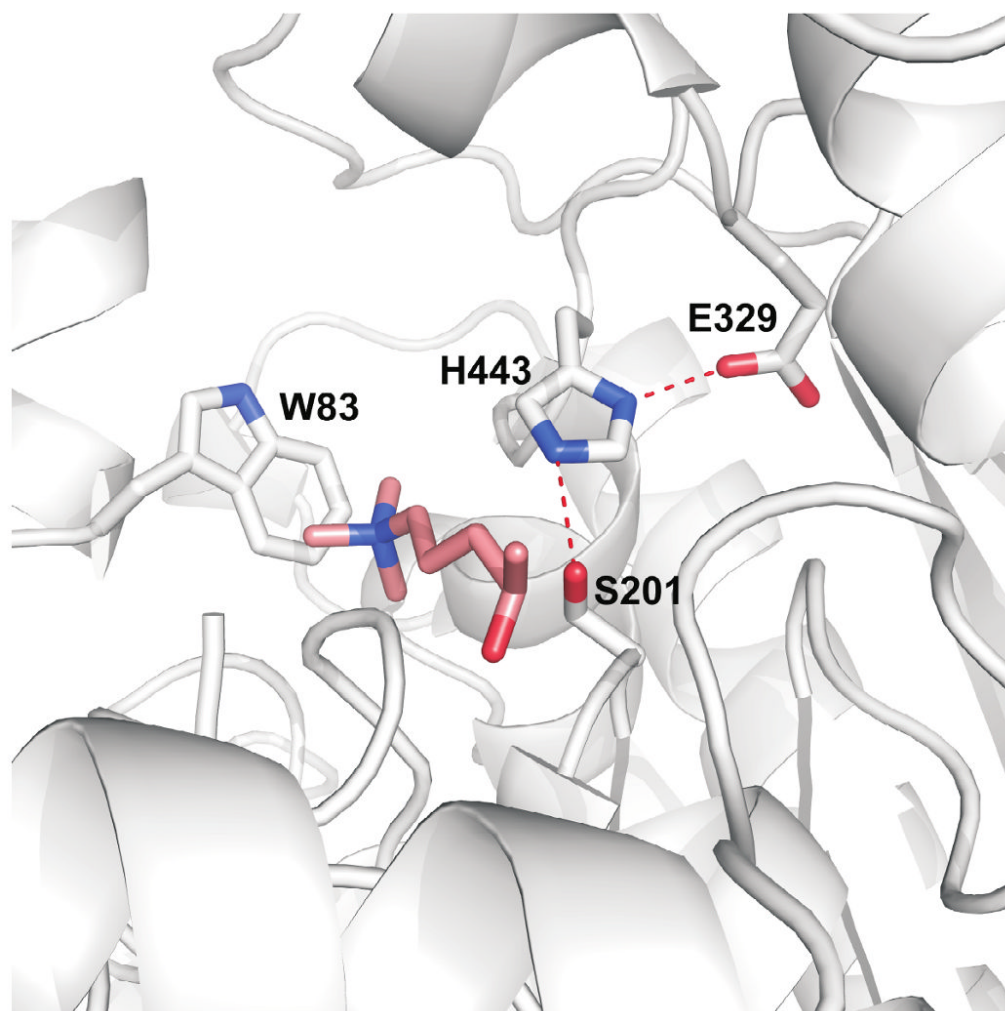
1D gel electrophoresis of *Cimex lectularius* salivary gland homogenates. The left lane shows the protein MW markers (kDa). The numbered template at right indicates how the gel containing the separated salivary gland proteins (right lane) were sliced. The names to the right of the template indicate the protein matches found by tryptic digestion/mass spectrometry. More detailed information about the proteins can be found in Supplemental Table S2. Abbreviations for the protein class are IPPase (Inositol phosphatase), NP (Nitrophorin), AG-5 (Antigen-5), CSP (conserved salivary peptide), OBP (odorant binding protein), DiAdase (dinucleotide phosphodiesterase).

**Figure 2.**

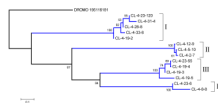
The nitrophorin/inositol polyphosphate phosphatase family of *Cimex lectularius*. A) Phylogram of the inositol polyphosphate phosphatase/nitrophorin family of *Cimex lectularius*, with triatomine (*Rhodnius prolixus* and *Triatoma infestans*) and one bee sequence. The numbers on the tree nodes represent the percent bootstrap support in 10,000 trials. The bar at the bottom indicates 10 % amino acid divergence. The *Cimex* sequences presented in this paper are recognized by CL-. Remaining sequences are named by a five letter code, the first 3 being from the genus name followed by the first two of the species name, followed by the NCBI accession number. The Previously described *Cimex* nitrophorin, as well as the bee protein are indicated with a square. For more details, see text. B) Clustal alignment of selected sequences shown in A) containing the region of amino acids providing contact to the heme moiety in the *Cimex* nitrophorin gi|85543988 (marked with black background). Residues marked in yellow background characterize the second subclade of the *Cimex* nitrophorins. Residues marked in red background characterize the IPPase clade, as well as CL-4-43-1.

**Figure 3.**

Alignment of the acetylcholinesterases from *Torpedo californica*, *Mus musculus*, *Drosophila melanogaster* and *Cimex lectularius*. Cysteine residues are marked in black background. The serine, glutamic acid and histidine residues important for catalysis are marked, as well as the amino acids tryptophan, the doublet glycine and alanine that are part of the active center cliff. The box in the carboxyterminus indicates the region of attachment of the glycosylphosphatidylinositol membrane anchor, lacking in the *Cimex* sequences. The number after the non-*Cimex* sequences are their NCBI accession number. The symbols below the alignment indicate: \* identical residues, : conserved residues, and. less conserved residues.



**Figure 4.** Molecular model of *Cimex lectularius* salivary acetylcholinesterase based on homology with the enzyme from *Torpedo californica*. A ligand model of 4-oxo-N,N,N-trimethylpentaminium, a non-hydrolyzable acetylcholine analog is built into the active site. The side chain of Trp 83 stabilizes the positive charge of the ligand via a cation- $\pi$  interaction. The catalytic triad, Ser 201, His 443, Glu 329 is also highlighted. Hydrogen bonds in the triad are shown as red dashed lines.



**Figure 5.** Phylogram of the odorant binding protein family of *Cimex lectularius*, with a *Drosophila mojavensis* outgroup sequence. The numbers on the tree nodes represent the percent bootstrap support in 10,000 trials (only values above 50% are shown). The bar at the bottom indicates 10 % amino acid divergence. The number following DROMO represents the NCBI gi| access for the drosophilid sequence. For more details, see text.

**Table 1**Functional classification of transcripts originating from *Cimex lectularius* salivary glands

Class	Number of Contigs	Number of ESTs	EST's / Contig
Housekeeping	452 (47)	796 (40)	1.8
Secreted	68 (7)	698 (36)	10.3
Unknown	426 (44)	452 (23)	1.1
Microbial, Virus and Transposable elements	14 (2)	14 (1)	1.0
<b>Total</b>	960	1960	

**Table 2**

Functional classification of putative housekeeping transcripts originating from *Cimex lectularius* salivary glands

Class	Number of Contigs	Number of ESTs	EST's / Contig
Protein synthesis machinery	106	247	2.3
Unknown, conserved	88	111	1.3
Protein modification machinery	42	79	1.9
Metabolism, energy	38	73	1.9
Transcription machinery	37	58	1.6
Protein export machinery	27	52	1.9
Signal transduction	38	51	1.3
Cytoskeletal	13	36	2.8
Transporters/storage	17	27	1.6
Transcription factor	3	15	5.0
Metabolism, lipid	9	13	1.4
Nuclear regulation	8	8	1.0
Metabolism, carbohydrate	8	8	1.0
Proteasome machinery	6	6	1.0
Metabolism, amino acid	6	6	1.0
Extracellular matrix/cell adhesion	3	3	1.0
Oxidant metabolism/detoxification	1	1	1.0
Metabolism, intermediate	1	1	1.0
Immunity	1	1	1.0
<b>Total</b>	452	796	



**Table 3**

Classification of transcripts associated with secreted products

Class	Number of ESTs	Percent of secreted class
Serpins	6	0.9
Odorant binding protein family	296	42.4
Immunity - Lysozyme	27	3.9
<b>Enzymes</b>		
Serine protease	18	2.6
Diadenosine phosphatase	8	1.1
Apyrase	16	2.3
Esterase	27	3.9
Antigen 5 family	14	2.0
Cimex nitrophorin family	261	37.4
Mucin-like	1	0.1
Other putative secreted peptides	24	3.4
<b>Total</b>	698	100