

Research article

Open Access

An insight into the sialome of the oriental rat flea, *Xenopsylla cheopis* (Rots)

John F Andersen¹, B Joseph Hinnebusch², David A Lucas³,
Thomas P Conrads³, Timothy D Veenstra³, Van M Pham¹ and
José MC Ribeiro*¹

Address: ¹Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA, ²Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840, USA and ³Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., National Cancer Institute, P.O. Box B, Frederick, MD 21702, USA

Email: John F Andersen - jandersen@niaid.nih.gov; B Joseph Hinnebusch - JHINNEBUSCH@niaid.nih.gov; David A Lucas - dl258t@nih.gov; Thomas P Conrads - conrads@upmc.edu; Timothy D Veenstra - tv52i@nih.gov; Van M Pham - VPHAM@niaid.nih.gov; José MC Ribeiro* - jribeiro@nih.gov

* Corresponding author

Published: 16 April 2007

Received: 20 December 2006

BMC Genomics 2007, 8:102 doi:10.1186/1471-2164-8-102

Accepted: 16 April 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/102>

© 2007 Andersen et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The salivary glands of hematophagous animals contain a complex cocktail that interferes with the host hemostasis and inflammation pathways, thus increasing feeding success. Fleas represent a relatively recent group of insects that evolved hematophagy independently of other insect orders.

Results: Analysis of the salivary transcriptome of the flea *Xenopsylla cheopis*, the vector of human plague, indicates that gene duplication events have led to a large expansion of a family of acidic phosphatases that are probably inactive, and to the expansion of the FS family of peptides that are unique to fleas. Several other unique polypeptides were also uncovered. Additionally, an apyrase-coding transcript of the CD39 family appears as the candidate for the salivary nucleotide hydrolysing activity in *X.cheopis*, the first time this family of proteins is found in any arthropod salivary transcriptome.

Conclusion: Analysis of the salivary transcriptome of the flea *X. cheopis* revealed the unique pathways taken in the evolution of the salivary cocktail of fleas. Gene duplication events appear as an important driving force in the creation of salivary cocktails of blood feeding arthropods, as was observed with ticks and mosquitoes. Only five other flea salivary sequences exist at this time at NCBI, all from the cat flea *C. felis*. This work accordingly represents the only relatively extensive sialome description of any flea species. Sialotranscriptomes of additional flea genera will reveal the extent that these novel polypeptide families are common throughout the Siphonaptera.

Background

The blood-feeding habit evolved independently in many arthropod orders or even within insect families [1]. Although most orders containing hematophagous insects already existed in the Carboniferous period, fleas probably arose in the early Cretaceous, about 120 million years ago [2], but are thought to have expanded relatively recently, in the past 50 million years, where they spread simultaneously with mammals and birds [3]. For comparison, the Culicidae provided a common ancestor for anopheline and culicine mosquitoes over 150 million years ago [4], or 100 million years before the spread of fleas.

In their peculiar adaptation to blood feeding, arthropods had to prevent hemostasis and inflammatory mechanisms in the host, because these could disrupt the flow of blood or alert the host defense behavior. Therefore, arthropods evolved a salivary cocktail that prevents blood clotting and platelet aggregation, induces vasodilatation, and additionally modulates the immune and inflammatory responses. A large diversity of salivary pharmacologically active molecules is produced by both unrelated and related hematophagous insects. For example, in the Hemiptera, the salivary vasodilatation induced by the bed bug *Cimex lectularius* and kissing bug *Rhodnius prolixus* are both a result of nitric oxide, but they are carried out by completely different NO-carrying molecules; a lipocalin in the case of *Rhodnius* [5,6], and a modified inositol phosphatase in the case of *Cimex* [7,8]. In Diptera, the salivary anticlotting mechanisms also differ between anophelines and culicines, where a unique antithrombin molecule named anophelin is found in the *Anopheles* [9,10], and a serpin inhibitor of factor Xa occurs in *Aedes aegypti* [11]. The enzyme apyrase (ATP-diphosphohydrolase, catalysing the hydrolysis of both ATP and ADP to AMP) is ubiquitously found in the saliva of bloodsucking arthropods, where it destroys the inducer of platelet aggregation, ADP and the proinflammatory mediator ATP, released by damaged cells [12]; however, two very different protein families have been recruited to serve this function in insect saliva. In mosquitoes [13,14] and kissing bugs [15,16], 5' nucleotidase family members serve this function, while calcium-dependent *Cimex* apyrase family members are found in the bed bug [17] and sand flies [18]. It is clear that a convergent evolutionary scenario accounts for the diversity of pharmacologically active salivary components of hematophagous insects, generating a large diversity of novel compounds.

Four species of fleas are known to have salivary apyrase activity, but their enzymatic protein family is unknown [19,20], while *Xenopsylla cheopis* (Rots) and *Xenopsylla astia* were shown to have salivary anticlotting activity [21]. Hydrolysis of the lipidic agonist platelet activating factor

(PAF) was also reported for the cat flea, but its molecular nature is unknown [22]. Other studies on flea saliva emphasize its allergenic role in human and veterinary medicine [23,24], including the molecular characterization of several antigens [25,26]. The paucity of salivary molecular information in fleas is indicated by the presence of only five proteins deposited in GenBank belonging to Siphonaptera and having the keyword 'salivary' (as in December 6, 2006), all belonging to the cat flea *Ctenocephalides felis*.

Despite great progress in the last 20 years in the identification of salivary pharmacologically active compounds, recent salivary transcriptome analysis of different blood-sucking arthropods is revealing a large number of putative secreted polypeptides for which the modes of action are unknown, including many novel protein families, some of which appear to be genus specific [27,28]. In the mosquitoes *Anopheles gambiae* and *Ae. aegypti*, over 70 salivary proteins have been identified; however, the function of fewer than 20 of these has been identified [28,29]. Given the phylogenetic distance between fleas and other blood-sucking arthropods, this work attempts to fill a gap in the knowledge of the salivary repertoire of hematophagous animals by providing insight into the sialome of the human vector of plague, *X. cheopis*. In this work, several transcripts coding for novel protein and peptide families are described and their translation into polypeptides confirmed. This work expands our knowledge in the salivary diversity of bloodsucking animals and, at the same time, it increases and maps our ignorance on the potential pharmacologic role of these novel protein families. To the extent that immune responses to flea saliva can control plague or rickettsial diseases for which fleas are vectors—as is the case of some other arthropodborne diseases [30-32]—this study also provides a platform of potential candidate antigens for future studies.

Results and discussion

cDNA library characteristics

A total of 944 clones were sequenced and used to assemble a database [see additional file 1] that yielded 245 clusters of related sequences, 171 of which contained only one expressed sequence tag (EST). 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 simplicity sake, this paper uses 'cluster' to denote sequences deriving both from consensus sequences and from singletons. The 245 clusters were compared using the program BlastX, BlastN, or RPSBLAST [33] to the nonredundant protein database of the National Center of Biological Information (NCBI), to a gene ontology database [34], to the conserved domains database of the NCBI [35], and to a customprepared sub-

set of the NCBI nucleotide database containing either mitochondrial or rRNA sequences.

Because the libraries used are unidirectional, the three-frame translations of the dataset were also derived, and open reading frames (ORF) starting with a methionine and longer than 40 amino acid (AA) residues were submitted to SignalP server [36] to help identify putative secreted proteins. The EST assembly, BLAST, and signal peptide results were loaded into an Excel spreadsheet for manual annotation.

Four categories of expressed genes derived from the manual annotation of the contigs (Table 1). The putatively secreted (S) category contained 34% of the clusters and 75% of the sequences, with an average number of 8.4 sequences per cluster. The housekeeping (H) category had 43% and 17% of the clusters and sequences, respectively, and an average of 1.5 sequences per cluster. Twenty two percent of the clusters, containing 8.4% of all sequences, were classified as unknown (U), because no functional assignment could be made. Similar to the Hgroup, this category also had an average of 1.5 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* [29]. A possible transposable element originated 21 singletons representing either active transposition or, more likely, expression of transposable element regulatory transcripts in *X.cheopis*.

Housekeeping (H) genes

The 106 clusters (comprising 157 EST) attributed to Hgenes expressed in the salivary glands of *X. cheopis* were further characterized into 13 subgroups according to function (Table 2). Not surprisingly for an organ specialized in secreting polypeptides, the two larger sets were associated with protein synthesis machinery (82 EST in 44 clusters) and energy metabolism (15 clusters containing 19 EST), a pattern also observed in other sialotranscriptomes [37-39]. We have also included in the H category a group of 12 EST that grouped into eight clusters, representing conserved proteins of unknown function, presumably associated with cellular metabolism. The fourth most abundant

group, attributed to the protein modification function, is of interest due to the specialized function of the salivary glands. This group, comprising five clusters, includes a chaperone protein and enzymes that are associated with disulfide bridge formation. Several transporters were also identified, including those coding for two subunits of the VATPase complex, which has been shown to be necessary for mosquito salivation [40,41]. Transcripts coding for VATPases are a common finding in mosquito sialomes [37,39,42-44]. The complete list of all 106 gene clusters, along with further information about each, is given in additional file 2.

Possibly secreted (S) class of expressed genes

Inspection of additional file 1 indicates the expression of several expanded gene families, including a family encoding proteins producing a match to KOG3720, indicative of lysosomal and prostatic acid phosphatases. This family alone is responsible for 35 clusters comprised of 268 sequences, or over 25% of the salivary library. Mucins are also represented, as well as an expansion of a peptide family unique to fleas that is related to the cat flea antigen annotated as FSH precursor (gi|1575479). Several other novel peptide families, described in greater detail below, were also observed.

Preliminary characterisation of the salivary proteome of X. cheopis

To obtain information on protein expression in the salivary glands of *X. cheopis*, salivary gland homogenates (SGH) of approximately 100 pairs of glands were separated by SDS-PAGE. The stained protein bands were excised from the gel and submitted to tryptic digestion followed by MS/MS identification. Note that the electropherogram is dominated by two broad bands in the molecular mass range of 40–43 kDa (Figure 1). Only five slices of the gel shown in Figure 1 yielded useful information by being assigned to predicted sequences from our clustered database and to fulllength sequences disclosed in this work. These results are summarized in Figure 1. To explore the expression of lower molecular mass polypeptides that were not represented in the SDS-PAGE experiment, we filtered 51 µg of homogenized salivary glands through a 30-kDa cutoff filter and submitted this filtrate

Table 1: Transcript abundance according to functional class

Class	Clusters *	Sequences	Sequences/Cluster
Housekeeping	106 (43.3)	157 (16.6)	1.48
Secreted	84 (34.3)	707 (74.9)	8.42
TE	1 (0.4)	1 (0.1)	1.00
Unknown	54 (22.0)	79 (8.4)	1.46
Total	245	944	

* Number (Percent of total)

Table 2: Functional classification of housekeeping transcripts

Function	Clusters	Sequences	Sequences/Cluster
Protein synthesis	53	82	1.55
Energy metabolism	15	19	1.27
Unknown conserved	8	12	1.50
Protein modification	5	6	1.20
Proteasome machinery	5	7	1.40
Protein export	4	4	1.00
Transcription machinery	4	9	2.25
Transporters/storagers	4	4	1.00
Signal transduction	3	3	1.00
Cytoskeletal	2	8	4.00
Cuticle protein	1	1	1.00
Nuclear regulation	1	1	1.00
Oxidative metabolism	1	1	1.00
Total	106	157	

to direct LC-MS/MS analysis. In addition, an aliquot of this filtrate was digested with trypsin prior to LC-MS/MS analysis. The low molecular weight (MW) filtrate yielded 22 additional matches with secreted proteins, as indicated in the supplemental Tables S1 and S2. The description of the identified proteins are embedded in the following manuscript sections.

Analysis of the adult female *Xenopsylla cheopis* sialotranscriptome

Several clusters of sequences coding for housekeeping and putative secreted polypeptides indicated in additional file 1 are abundant and complete enough to extract consensus sequences of novel sequences. Additionally, we have performed primer extension studies in several clones to obtain full-or near full-length sequences of products of interest. A total of 76 novel sequences, 48 of which code for putative secreted proteins, are grouped together in additional file 2.

Following is a detailed description of the full-length transcripts found in the salivary glands of adult *X. cheopis*:

Putative salivary secreted proteins for which a protein family is known

Enzymes

Phosphatase family
Acid phosphatases catalyze the hydrolysis of phosphate monoesters and, in some cases, phosphoryl transfer between a phosphoester and alcohols [45]. These proteins are widely distributed in animal and plants, occurring in three different types. One type has a relatively small MW (18–20 kDa) and is found in mammalian liver, a second type has a higher MW (45–60 kDa), such as the enzymes of wheat germ, lysosomes and prostate. The third type are the purple acid phosphatases, which contain a binuclear iron center [46]. As listed in additional file 2, eight full-

length transcripts coding for mature proteins ranging from 36–45 kDa all having a basic (> 8.5) pI were identified in this study. We also report two truncated (at the 5' end) members of this family. All transcripts of this class produce similar matches to lysosomal/prostatic acid phosphatases in the NR and GO databases, as well as to KOG3720 indicative of lysosomal and prostatic phosphatases. Except for one of the truncated members, several peptides were identified for each protein originating from the major band at 40 kDa reported in Figure 1. Edman degradation of the same band also revealed a strong signal for the proteins annotated as phos, phos1 and phos 1A.

Alignment of the flea salivary phosphatase sequences with the vertebrates human, rat and chicken members, as well as with the venom phosphatase of the bee, a fly (*Drosophila melanogaster*) and a beetle (*Tribolium castaneum*) shows that, except for the first arginyl residue, all flea sequences do not have the conserved catalytic site residues as determined by crystal analysis of the human and rat sequences (shown in red color over yellow background in Figure 2) [46,47]. Remarkably absent is the histidyl residue located in the conserved motif RHGDR found in all non-flea sequences. This residue is known to be essential for catalysis, forming a phosphorylated intermediate during the reaction. All these active site residues experimentally determined for vertebrate enzymes are conserved in the fly, beetle and bee enzymes, indicating that fleas probably co-opted the acid phosphatase fold for a non-phosphatase function. Indeed, only a few other AA are conserved across the alignment shown in Figure 2, including the last three cysteines, a region of double aromatic AAs, and another region of double glycines, among others. Phylograms of the alignments show three distinct clades (Figure 3). Clade I has all non-flea enzymes, where three sub-families are found, the vertebrate subfamily

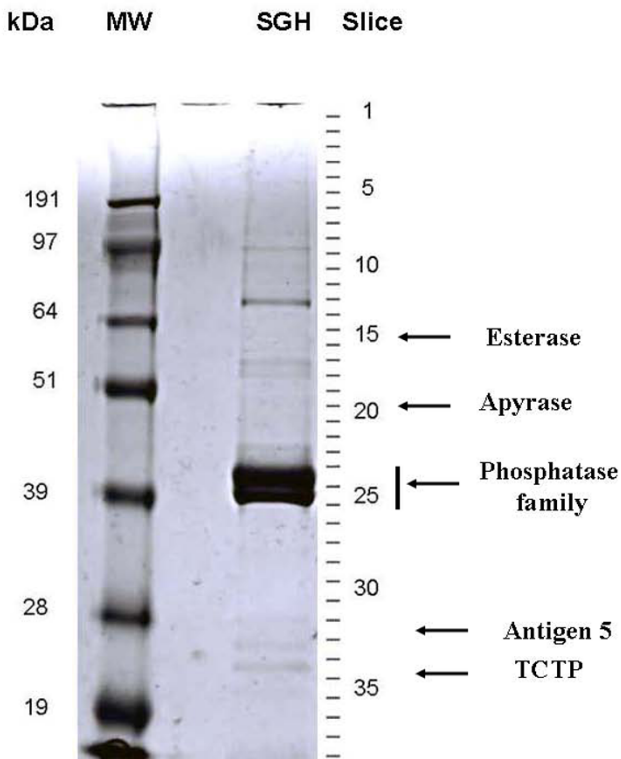


Figure 1
Gel electrophoresis of *Xenopsylla cheopis* salivary gland homogenates. The left portion of the gel shows molecular weight markers (MW) in kDaltons. The right portion shows the electrochromatogram of salivary gland homogenates, and the slices that were cut for tryptic digestion MS/MS experiments. The arrows point to the gel slices where the indicated enzymes and translationally controlled tumor protein (TCTP) were found. For more detail, see text.

(with six cysteines, four of which are involved in disulphide bonds) [46], the bee venom enzyme (with only three cysteines), and the fly and beetle enzymes (with five and seven cysteines, respectively). Clade II has eight of the ten flea sequences (with seven cysteines), and contains the flea sequences with the most conservation to the non-flea members, while clade III has the two most divergent flea sequences (with five and three cysteines). The odd number of cysteines in all invertebrate sequences indicate the possibility of reactive sulfhydryl groups in all enzymes, including those vertebrate enzymes where two of the six cysteines are not involved in disulphide bridges. It is apparent that fleas achieved a unique expansion of acid phosphatases, at least nine members of which were identified in the salivary glands using MS/MS. In contrast, only three genes are found for this class of enzymes in the *An. gambiae* proteome (by querying AnOxcel at the anobase site [48]) for the KOG motif lysosomal and prostatic

acid phosphatases), or four such enzymes in the human genome [49].

The real substrate of acid prostatic/lysosomal phosphatases (where the enzyme can achieve 1 mg/ml in semen) [50] is not known, but it has been suggested that it can act as a protein tyrosine phosphate phosphatase and affect cell growth when the enzyme occurs intracellularly [51,52]. Extracellular protein phosphorylation/dephosphorylation is known to affect many aspects of cellular signaling [53,54], and is involved in platelet aggregation [55-57]. The conserved basic nature of all flea salivary phosphatases points to interaction with a negatively charged target. Although the active center of acid phosphatases appear quite open, accepting a diversity of small substrates, it is possible that the large size of the enzyme confers some degree of selectivity to the larger phosphorylated protein substrates. This selectivity is analogous to the large serine proteases that are involved in vertebrate blood clotting or invertebrate prophenoloxidase activation, which accepts many different small substrates but are quite specific for their protein substrates. Accordingly, it is possible that binding some host phosphorylated protein substrate has been the target of flea salivary phosphatases. Perhaps loss of the enzymatic activity kept the substrate permanently blocked by an inactive interaction, resulting in a more advantageous complete receptor blockage. On the other hand, loss of enzymatic function would require larger amounts of the protein to interact stoichiometrically with the target host protein, which might have been the reason for large protein expression and gene duplication observed, where gene duplication immediately confers the benefit of increased transcription. With time, gene duplicates may diverge to different targets, and/or to avoid immune detection by hosts. It is also possible that this phosphatase family may have evolved to chelate polyphosphates released by platelets that recently have been shown to have important hemostatic functions [58]. These considerations should help to identify the function of the flea salivary phosphatases.

Esterase

Additional file 2 reports two mRNA sequences coding for esterases, both similar to many insect proteins annotated as carboxylesterases. The XC-184 translation product of 211 aa is similar to the aminoterminal region of carboxylesterases with sizes varying between 530 and 560 aa [59]. The full polyadenylated transcript has clear polyadenylation sites and cannot transcribe for the larger homolog. Unless the EST is an artifact, it appears that flea saliva codes for a truncated version of a carboxylesterase enzyme. On the other hand, evidence for expression of the full-length esterase, consistent with a protein product of ~ 60 kDa, was obtained by the MS/MS experiments from gel slices 15 and 16 shown in Figure 1. Another este-

rase is coded by Cluster-136, a truncated mRNA coding for the carboxy terminus of a different carboxylesterase. Because we did not obtain the 5' region of this mRNA, we cannot verify whether the protein product has a signal peptide indicative of secretion. These mRNA sequences are included in the putative secreted category due to the possibility these enzymes may function as PAF hydrolases, as was found with an esterase salivary activity of the cat flea [22]. Esterases have been also reported in the salivary glands and saliva of both male and female *Ae. aegypti* mosquitoes [60], but their natural substrate is unknown.

Apyrase

Five transcripts coding for a homolog of an apyrase member of the CD39 family is the first finding of this type of gene expressed in the salivary glands of any hematophagous arthropod to date [Additional file 1]. The consensus sequences shown in additional file 1 indicates the EST are truncated and match the 3' region of homologous

sequences, accounting for approximately 50% of the enzyme. This relatively abundant gene product may account for the salivary apyrase of this flea species. Because this activity was never described in *X. cheopis*, we investigated whether its SGH could hydrolyze adenosine nucleotides. Figure 4 indicates, as found for most salivary homogenates of bloodsucking arthropods, that *X. cheopis* has indeed a divalent cation dependent salivary apyrase activity, that can be activated by either Ca⁺⁺ or Mg⁺⁺. This result is consistent with a CD39 or a 5'-nucleotidase family member, who functions with either divalent cation, but is not consistent with the Cimex apyrase type found in bed bugs and sand flies, which are strictly Ca⁺⁺ dependent. Protocols involving rapid amplification of cDNA ends (RACE) were used to obtain the 5' region of the truncated putative apyrase transcripts, yielding the full-length sequence coding for XC-APY, shown in additional file 2. When XC-APY was compared with eight members of the human CD39 family, plus the soluble potato apyrase sequence, the four conserved regions of the enzyme fam-

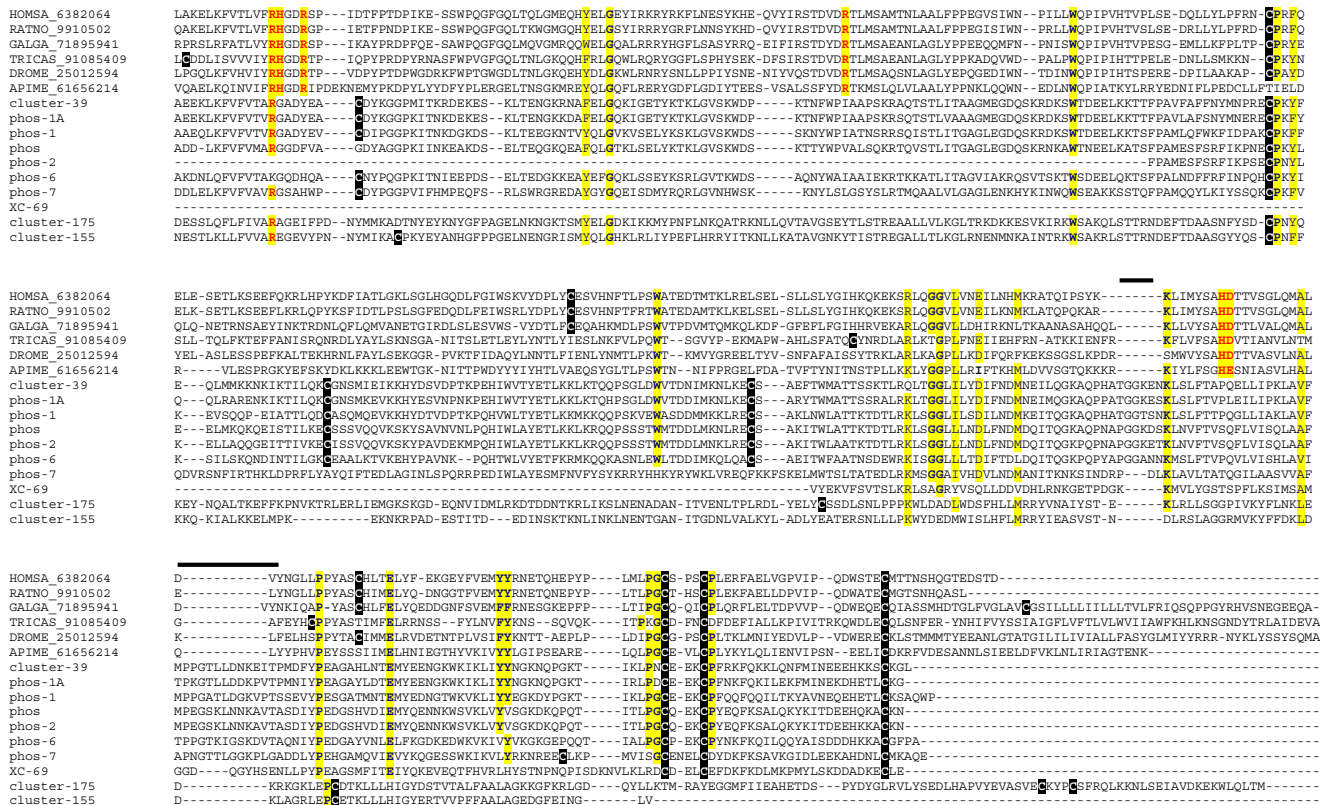


Figure 2
Alignment of human, rat, chicken, beetle (*Tribolium castaneum*), fly (*Drosophila melanogaster*), bee and flea acid phosphatase sequences. The red letters over black background indicate amino acids found in the active center of rat and human enzymes. Other yellow background indicate identical (in bold) or conserved amino acids. The bars indicate regions of insertion/deletion when the flea and remaining sequences are compared. Cysteines are shown in white font over black background.

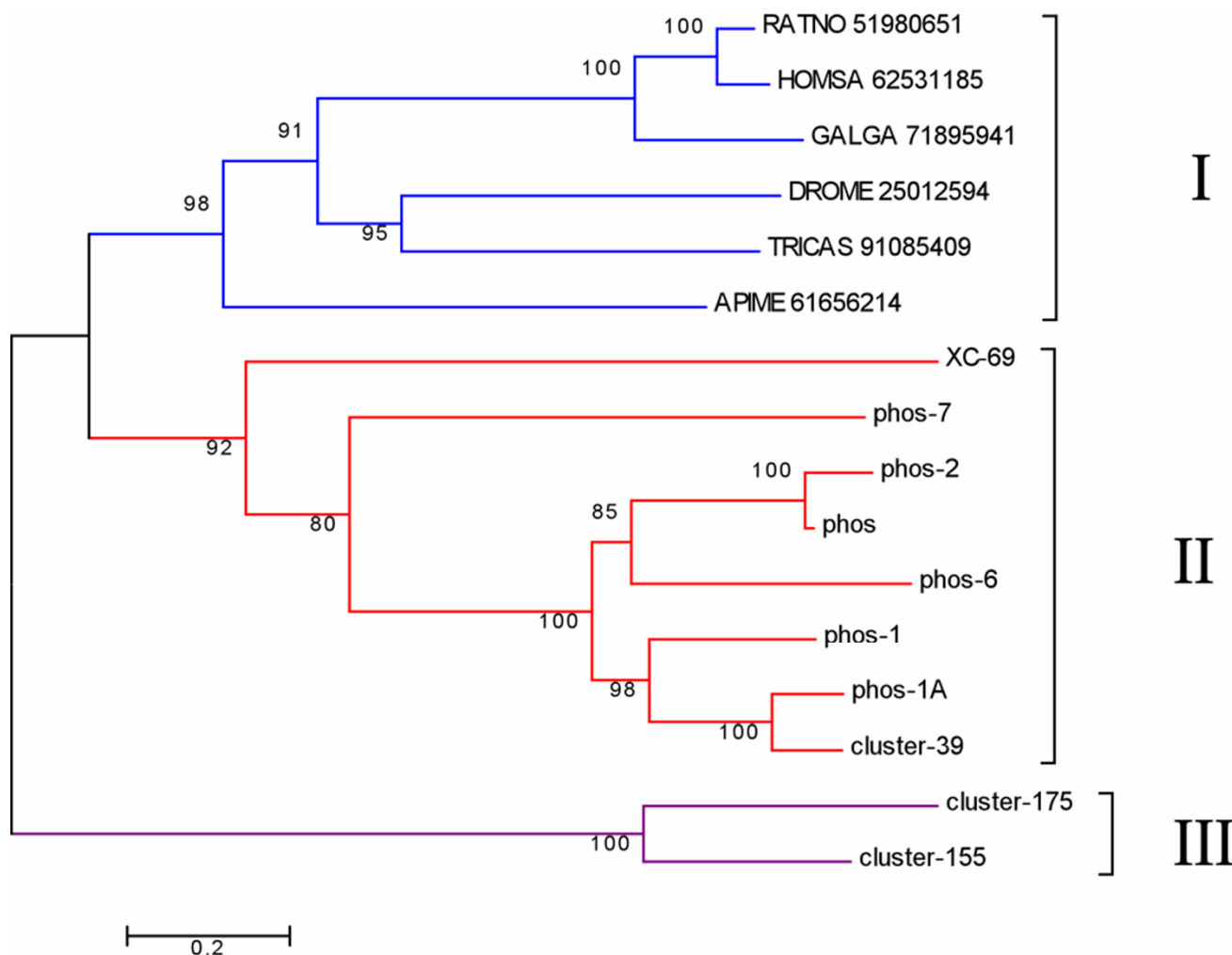


Figure 3
Phylogram of human, rat, chicken, beetle (*Triboleum castaneum*), fly (*Drosophila melanogaster*), bee and flea acid phosphatase sequences. The phylogram was deduced from the Clustal alignment from figure 2. The non flea sequences are indicated by 5 letters representing the first 3 letters of the genus and 2 letters of the species name, followed by the NCBI accession number. The numbers in the phylogram indicates the percentage of concordance in 10,000 bootstraps. The bar at the bottom indicates 20% amino acid distance.

ily are clearly found [61] (Figure 5 – boxed regions). Except for two human enzymes that are secreted, the remaining human sequences are membrane bound by two membrane anchors, one in each of the carboxy- or aminoterminal regions of the protein, indicated by the bars above the alignments in Figure 5, and by the predominance of aliphatic AA (shown in turquoise color in Figure 5). These hydrophobic regions coincide with membrane helices predicted by the TMHMM server, and can be seen in additional file 3[61]. These carboxyterminal helices are missing in the two secreted human enzymes [62,63], on the soluble potato apyrase, and in the flea apyrase (Figure 5). The flea apyrase also contains a clear signal peptide indicative of secretion, supporting the hypothesis that XC-

apy is responsible for the salivary apyrase of *X.cheopis*. Mass spectrometry experiments located this enzyme within bands 19 and 20 of the gel shown in Figure 1, consistent with a MW of 45-50 kDa, and matching the predicted mass of the mature secreted enzyme of 46.9 kDa. To the extent that XC-apy is responsible for the observed salivary apyrase activity of *X.cheopis*, this will be the first description of a hematophagous arthropod to have co-opted this protein family for this particular activity.

Adenosine deaminase

Three transcripts code for truncated versions of the enzyme adenosine deaminase [Additional file 1], for which both transcripts and enzymatic activity were dem-

onstrated in mosquito and sand fly salivary glands [64,65]. This salivary activity may help to convert mast cell degranulating adenosine into inosine [66].

Mucins

Three full-length transcripts with signal peptides indicative of secretion code for peptides of mature MW ranging from 5.5–13 kDa and having 6–12 putative galactosylation sites. They produce no similarity matches to any known proteins. Two of these products appear to result from a gene duplication event. Peptides with a high number of predicted galactosylation sites are regularly found in other hematophagous insect's sialotranscriptomes. These proteins may help to lubricate the insect food canal. Evidence for translation of XC-61 was found in bands 39 and 40 in the gel experiment of Figure 1, and also in the low MW filtrate. Expression of XC99 was indicated by a peptide match in the digested low MW filtrate.

Antimicrobial peptide

The full-length sequence for an antimicrobial peptide (AMP) of the defensin family is presented in additional file 2. Antimicrobial peptides are a regular finding in sialotranscriptomes of hematophagous insects and ticks. These peptides, when ingested with the blood meal, help control bacterial growth in the gut and may also protect their host-feeding lesions from infection.

Antigen 5 family

This protein family is widely found in the venom of vespids [67], in the salivary glands of many blood-sucking insects [37,44,68], and also as a multi-gene family in most animal genomes such as in *Drosophila* [69]. These proteins belong to the larger family of cysteine-rich extracellular proteins (CRISP) ubiquitously found in animals and plants [70]. Most proteins have no known function. We here report a salivary member of the antigen 5 family found in *X. cheopis*. Alignment and phylogenetic analysis of insect members of this family indicates, as expected, that *X. cheopis* salivary antigen 5 protein clusters with a related salivary protein from the cat flea (not shown).

Putative salivary secreted proteins belonging to novel polypeptide families

FS or antigen I family

The sialotranscriptome of *X. cheopis* revealed transcripts coding for several peptides with similarity to a previously described antigen of the cat flea named FS-H precursor, as well as two other cat flea larger proteins named FS-I and antigen 1 precursors. When the newly discovered peptide sequences were compared with the whole transcriptome database (using the tool BlastP with the low complexity filter off) other related sequences were found, thus creating a group of 15 related products that have in common solely the presence of eight cysteine residues (Figure 6A),

the presence of signal peptides indicative of secretion, and a mature molecular mass varying from 6.3 to 9.2 kDa.

The phylogram (Figure 6B) shows two unrelated groups of sequences, with strong bootstrap support for the relationship among several sequence pairs, including one pair consisting of one cat flea and one rat flea sequence. However, the roots of the several clades have weak bootstrap support indicating that these sequences either originated from unrelated genes, or more likely, they originated from gene duplication events with fast divergence from the ancestral gene. The programs hmmbuild and hmmcalibrate of the hmmer package (version 2.0) [71] were used to make a hidden Markov model from the alignment shown in Figure 6A (excluded of the first 20 AA containing most of the signal peptide region). The program hmmsearch from the same package was used to scan the NCBI nonredundant protein database with the flea sequences described here added. Only flea sequences were retrieved at a significant probability level ($< 1e-7$). Two additional sequences were retrieved at much lower significance levels, one being also a flea sequence, and one plant peptide (Figure 7). When the plant peptide was compared with the NR database (using BlastP) several defensinlike molecules from *Arabidopsis* and *Helianthus* were retrieved.

Plant and insect defensins as well as the scorpion toxins show a high degree of structural conservation despite considerable divergence of sequence [72]. These peptides adopt a cysteine-stabilized- $\alpha\beta$ fold containing a three stranded β -sheet and a single α -helix [72,73]. The cysteine sequence motif C...CXXXC...GXC...CXC is conserved in this group, where X is any amino acid and ... is a variable interval of AA [72]. The six cysteine residues contained in the motif form three disulfide bonds, and a fourth disulfide bond is often present, but its position is variable. Sequence alignments of FS family of flea salivary peptides with insect/plant defensins reveal only low overall identity. However, the three disulfide bond cysteine sequence motif characteristic of defensins and scorpion toxins is present (Figure 8), making it likely that these peptides belong to the cysteine stabilized $\alpha\beta$ fold group. Insect type defensins are antimicrobial molecules important for protection against bacteria and fungi [74], while scorpion toxins are neurotoxic molecules that impair the functions of sodium and potassium channels [75]. Based on this similarity, possible functions of this group of peptides could include, but not be limited to, controlling microbial growth or analgesic function.

The digested or undigested 30 kDa filtered SGH provided for MS/MS identification of 11 of the 15 peptides reported in additional file 2, while the front of the gel shown in Figure 1 allowed for identification of two peptides, also iden-

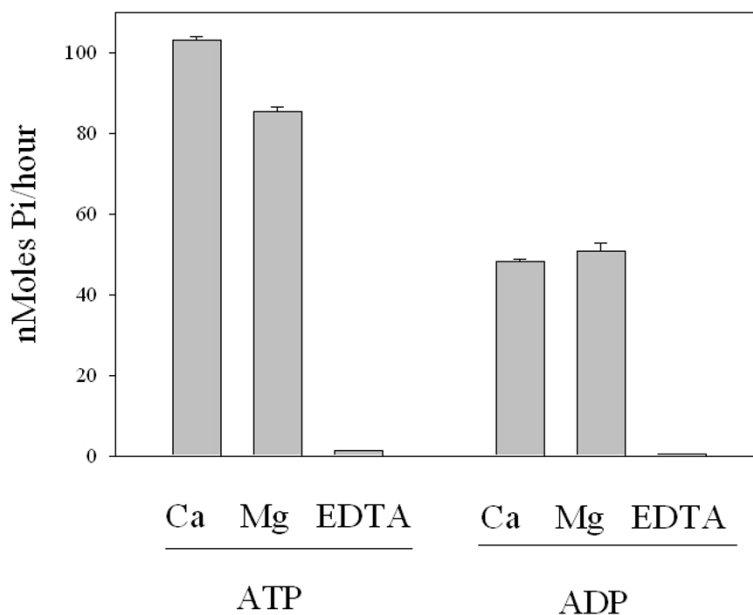
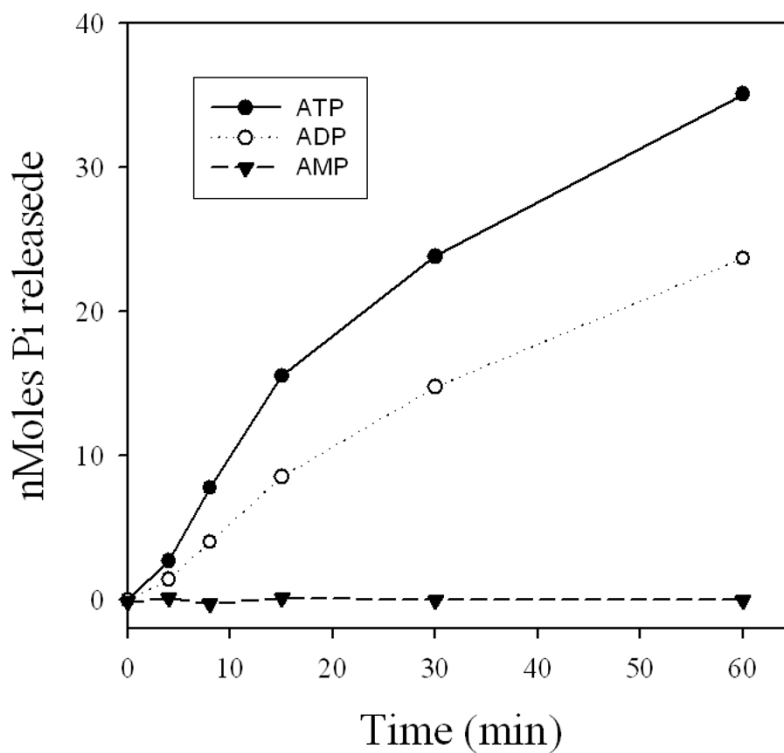


Figure 4

Apyrase activity of salivary homogenates of the flea, *Xenopsylla cheopis*. Reaction media contained 50 mM TrisCl pH 7.4, 150 mM NaCl, 2 mM indicated nucleotide in 100 μ l. The reaction started with addition of salivary gland homogenate to give 2.5 pairs per ml. (A) 1 mM CaCl₂ plus 1 mM MgCl₂ were added to the media before starting the reaction. (B) Either 2 mM CaCl₂, 2 mM MgCl₂ or 2 mM EDTA were added to the media before starting the reaction. The bars represent the average \pm SE of 3 determinations. All incubations were done at 37°C.

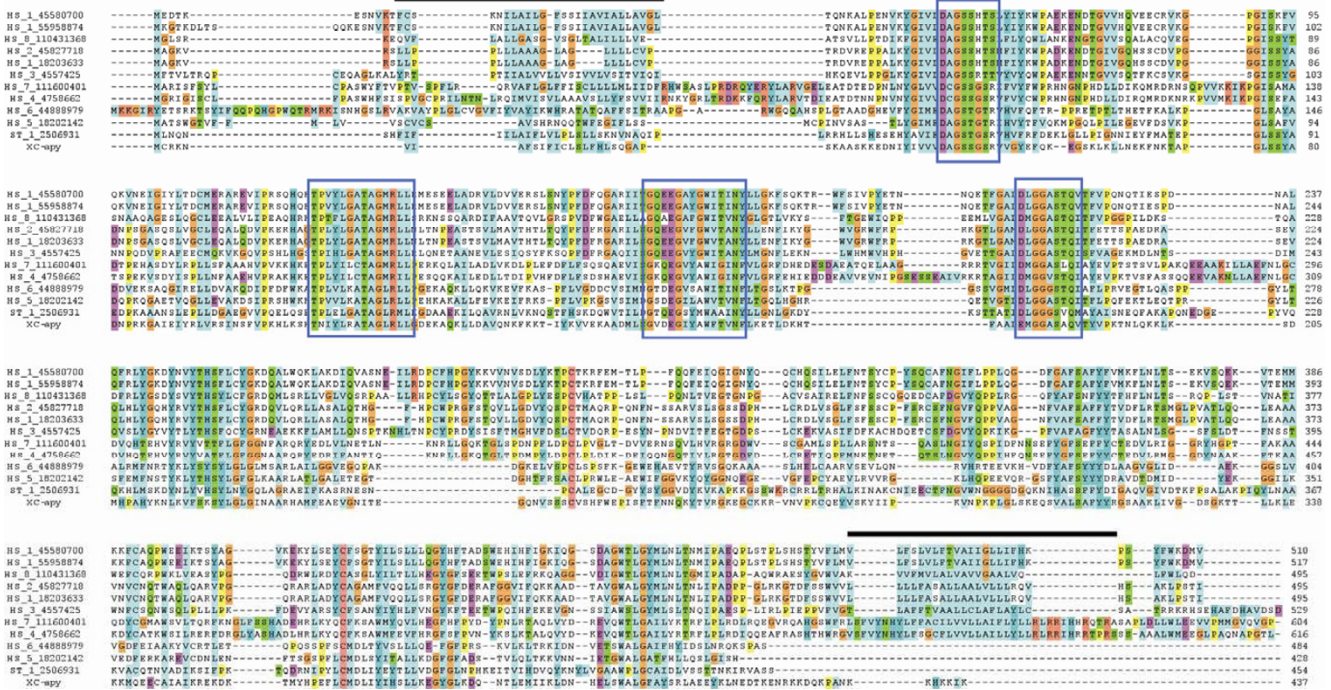


Figure 5
Alignment of human, potato and flea proteins of the apyrase/CD39 family. Human sequences are identified by HS_#_???? where # is the enzyme number and ??? refers to the NCBI gi accession number. The potato sequence is identified by ST_1 followed by the NCBI accession number. The flea sequence is indicated by the prefix XC. The boxes mark the 4 conserved domains identified by Wang et al [60]. The lines above the alignments indicate the regions of hydrophobic helices in the amino and terminal regions, the last of which is lacking in the four bottom sequences, representing the 2 human secreted enzymes, plus the potato and the flea enzymes.

identified by MS/MS of the filtrated sample (for details see Methods section).

Other novel peptides

Fifteen other polypeptides are described in additional file 2, all containing a signal peptide indicative of secretion, and without significant matches to other proteins in the NR database, except for the peptide named Cluster-149 that is weakly similar to low complexity bacterial proteins. Two of these peptides appear to be splice variants (XC-42 and XC-43) or products of recent gene duplication. Two other peptides (XC-3 and XC-56) share significant homology in their signal peptides, but not so much in their mature peptide sequences. Except for Cluster-149 that codes for a mature protein of 19.3 kDa, all other peptides have a predicted weight smaller than 7.8 kDa, six of which are less than five kDa. XC-105, coding for a mature peptide of 3.9 kDa has a weak SMART match to calcitonin or calcitonin gene related peptides, the last one being a potent vasodilator [76]. XC-56 is weakly similar to HIV

gp160 and to patented peptides with inhibitory activity towards the viral protein binding to the CCR5 receptor. Cluster-113 peptide has a cysteine scaffold found in the cysteine rich region of tick salivary metalloproteases, suggestive of an interaction with extracellular matrix elements. XC-2 has sequence homology to patented decapeptides similar to peptide hormone receptors, and to small sequence regions of these receptors. XC-68 has the cysteine scaffold of arthropod defensins, including the C-X-C carboxyterminus where X is a hydrophobic AA (Ile in this case). Mass spectrometry experiments indicated in additional file 2 produced evidence for expression of 9 of the 15 polypeptides.

Housekeeping proteins

Additional file 2 describes 28 full-length sequences coding for 20 proteins involved in protein synthesis (mostly ribosomal proteins), two that are part of the proteasome machinery, one involved in protein modification, one cytoskeletal protein, one involved in energy metabolism

A

	1	2	3	4	5	6	7	8				
CTEFE_1575479	MKSLLLFPTLCVLYQIVVAD	RRVSKT	OSGGKI-QSEEQVVIKSGQHI	LENY	SDGRN	NNNP	HLF	MRER	SRSGNGG	CGNGGR	TRPDSK	CEAPYSG
cluster-169	MKYIYAVLLTFLLCQLSLAD	RMVSLT	KTGGQQ-TGDNQVAIPSGTHI	LQGY	CK	NHND	COMF	MTES	CRSGNGG	CGNGGT	SRPNRD	CEAPYSG
XC-78	MKWFLLLALVCAIEFPCSC	WKASQK	FKTGT-DITPNIEVGNLAK	VPVG	NGAGDKDTKKNK	LEH	CK	CK	KATNGS	CGTSSI	AKPNPQ	CEYKQDPDK
XC-7-90-90-50-CLU	MKWILVLLALVCLAVEYSYS	WKVSEK	LKKGH-KFQADQEIIGNLAT	AKGQ	CKTDSQKQKAGK	CDKH	CGV	CL	GSGGS	CGDGSS	QKPNKE	CEYKSK
XC-58	MKCIPLFVLICLTVQFSYS	WKVSEK	LKKGH-KFQADQEIIGNLAT	AKGQ	CKTDSQKQKAGK	CDKH	CGV	CL	GSGGS	CGDGSS	QKPNKE	CEYKSK
XC-80	MSKSLVFLLECAFVLVNA	KRNSDK	KAGATG-GTPTTE-ENDTLR	LPGF	CKPDK	GSPQ	CKQ	CKE	GFDFGL	CGSYGTS	ADKNT	CEFTNNPTTIK
XC-7-90-90-51	MKPIFLLAIAFAFAVAVVA	AGSKDS	KA-RA-GVRLT-SQSLQ	VPNI	CKPNE	NDDK	NHF	CKNE	CAKKGTT	CGSGTT	AQNT	CEFTYN
XC-44	MKCLTLLILCLTFAAVFVSG	NRNSAE	MNNKP--KPNVLVIDDENA	VPDI	CKSND	TGDEA	CKN	CKNT	CSDFGF	CGNEASE	DKKVV	CEYQFN
cluster-145	MKLFVFLFLVFNQYICRRSRNTVEVTKK	KS	SGGKDTSMKEVEVSGSFYPTLELPDY	THRR	LDDE	CKSK	CKE	CKN	NPPFKR	CGVGVWD	SQAGN	CEYVHT
cluster-133	MKPLFGIFVIFAILCIKLIS	ARDNN	QYR-----PGIIV	VPGG	CKDK	NDVEG	CKQ	CKQ	GGYQKTVR	CGGTYS	PSTL	CEYVGP
XC-7-90-90-48	MKFAFAIFVVLAILHTELIS	ATEYK	CKVT-----GTGDMT	IPFG	CKDG	NDVQ	CKK	CKE	CKK	YTTTRQS	CKV	CEYV
XC-7-90-90-49	MKPLTIIFVAITIS-CFLG	AMGDT	CKTI-----SGTYQ	PKDE	CKTKL	YGDEG	CKSA	CKNT	CSGKNGY	CGNGDS	KTPDEB	CEYVGN
XC-54	MKVLVILFVIFLLQIVFSA	TVKKE	CKQD-----SAGDY	VPGG	CKDSK	DRNK	CKRL	CKE	CKSKVKDLMS	CGDKLL	BVDKD	CEYIDLTFPGEETDVK
XC-63	MKLVPIILVFLVLLQIVFAD	KQE	CKKD-----NIGDY	IPKG	CKDK	DNDK	CKKL	CKE	CKVLSKVKEPFGA	CGSNVHYENG	NINPDP	CEYFDLQVLGTSK
XC-92	MKFLVLSLIFVL-QIVFAD	EELQ	CKED-----GAGSY	IPKG	CKKM	DQNE	CKKL	CKL	CKSKVKNPKGY	CGSNVMR	NVDADS	CEYVGT
cluster-119	MRLLFAYTLILALLQITFAA	TANLK	CKQD-----NVGDYF	IPNG	CKDDP	TGSN	CKNL	CKSL	CKTKVKNPFG	CGVKEI	KVDPK	CEYVLEIGD

B

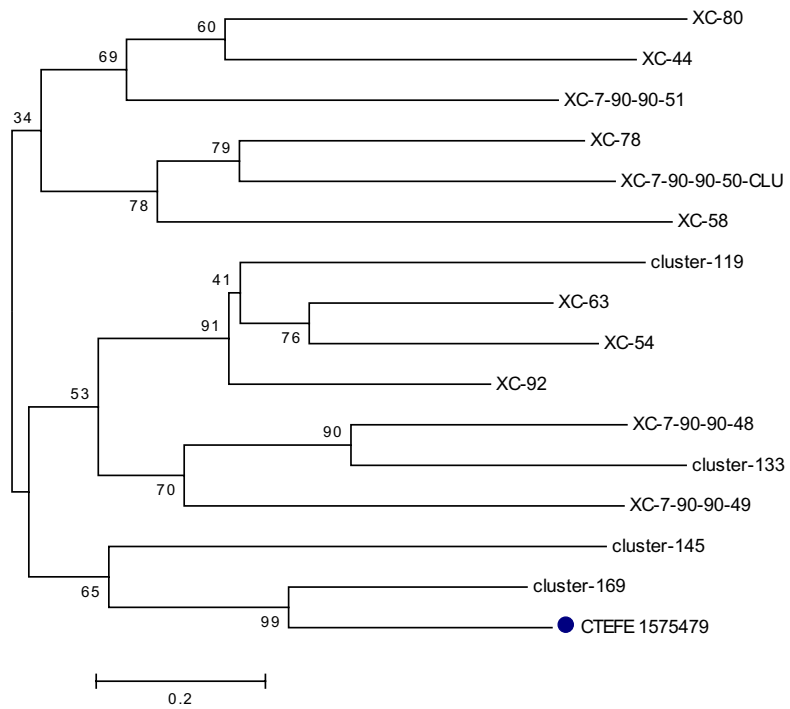


Figure 6
The FS peptide family expressed in the salivary glands of *Xenopsylla cheopis*. A) Clustal alignment of the Oriental rat flea sequences with the cat flea sequence with NCBI accession number 1575479. Conserved cysteines are shown in reverse black color, and the number above their locations B) Phylogram of the FS family displaying the divergence of the family, and the association of the cat flea sequence with the rat flea sequence annotated as cluster-169. The numbers represent the percent concordance of 10,000 bootstrap replicates. The bar at the bottom indicates 20% amino acid distance.

and two that are uncharacterised, conserved proteins. These sequences can help in phylogenetic studies and as controls in future studies with *X.cheopis*.

Conclusion

Analysis of the salivary transcriptome of the flea *X. cheopis* revealed the unique pathways taken in the evolution of the salivary cocktail of fleas. Gene duplication events led to large expansions of a family of probably inactive acidic

phosphatases, never found before in any other blood-sucking arthropod, and to the FS family of peptides unique to fleas. Additionally, over a dozen unique peptides were found. An apyrase-coding transcript of the CD-39 family appears as the candidate for the salivary nucleotide hydrolyzing activity in *X. cheopis*. If this is confirmed, this will be the first arthropod to have recruited this gene family for its salivary apyrase activity. Only five other flea salivary sequences exist at this time at NCBI, all from the

Scores for complete sequences (score includes all domains):

Sequence	Description	Score	E-value	N
XC-78	putative salivary p	120.6	1.9e-30	1
gi 1575479 gb AAB09540.1	FS-H precursor	115.9	4.7e-29	1
gi 1575479 gb AAB09540.1	FS-H precursor	115.9	4.7e-29	1
XC-58	similar to CF saliv	109.0	5.5e-27	1
XC-54	similar to fungal p	107.1	2.2e-26	1
cluster-169	putative secreted s	105.5	6.3e-26	1
XC-80	putative salivary s	98.6	7.7e-24	1
XC-7-90-90-50-CLU	putative salivary s	98.1	1.1e-23	1
XC-63salivary	secreted peptide of	97.0	2.2e-23	1
cluster-145	putative secreted s	90.7	1.9e-21	1
XC-44	putative salivary s	86.2	4.2e-20	1
XC-7-90-90-51	putative salivary s	81.3	1.2e-18	1
cluster-119	putative secreted s	80.7	1.9e-18	1
XC-92	putative salivary s	78.9	6.6e-18	1
XC-7-90-90-49	putative salivary s	70.4	2.4e-15	1
cluster-133	putative secreted s	64.5	1.4e-13	1
gi 4336703 gb AAD17905.1	salivary antigen 1	61.8	8.9e-13	2
gi 4336703 gb AAD17905.1	salivary antigen 1	61.8	8.9e-13	2
gi 14423664 sp Q94424 CTF1_CTEFE	Salivary antigen 1	61.8	8.9e-13	2
gi 3805687 gb AAC69105.1	FS-I [Ctenocephalid	52.8	4.6e-10	2
gi 3805687 gb AAC69105.1	FS-I [Ctenocephalid	52.8	4.6e-10	2
XC-7-90-90-48	putative salivary s	43.7	2.6e-07	1
XC-2	putative secreted s	8.5	2.2	1
gi 79324430 ref NP_001031491.1	unknown protein [Ar	6.1	3.7	1

Figure 7

Search of the non-redundant NCBI protein database for proteins similar to flea sequences found in this work.

A hidden Markov model was made from the alignment shown in Figure 6 (minus first 20 amino acids to exclude signal peptide) to search the non-redundant protein database.

cat flea *C. felis*. This work accordingly represents the only relatively extensive sialome description of any flea species. Sialotranscriptomes of additional flea genera will reveal the extent that these novel polypeptide families are common throughout the Siphonaptera.

Methods

Fleas

Intact salivary gland pairs were collected from adult female *X. cheopis* fleas. Individual fleas (anesthetized by chilling on ice) were dissected in 10 µl of PBS on a glass microscope slide on the stage of a dissecting stereomicroscope. By grasping the dorsal half of the flea above the forelegs with forceps, pressing down on the abdomen just posterior to the midgut with a bent dissecting needle, and pulling, the two pairs of salivary glands and the midgut would usually remain attached to the head and be pulled free of the rest of the body. The common lateral salivary ducts were cut to release each pair, which were then hooked with a dissecting pin and placed in PBS at 4°C.

Pools of 40 pairs of glands in 20 µl PBS were frozen at 70°C. Glands used for apyrase assays were dissected and stored in 10 mM TrisHCl and 150 mM NaCl, pH 7.4 rather than PBS.

Salivary gland isolation and library construction

X. cheopis salivary gland mRNA was isolated from 200 salivary gland pairs from adult fleas using the Micro-Fast-Track 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 hr in the presence of the SMART IV and CDS III (3') primers. Second strand synthe-

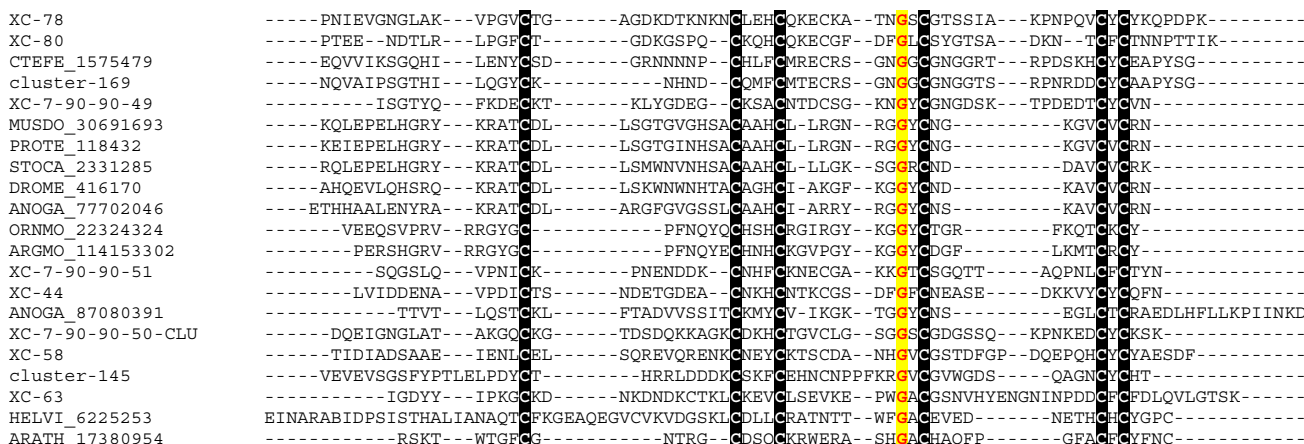


Figure 8
Alignment of selected *Xenopsylla cheopis* peptides of the FS family with defensin peptides. The conserved cysteine framework C...C-X-X-X-C...G-X-C ...C-X-C is indicated. Sequences from the present work start with XC or cluster; other sequences are from GenBank and consist of 5 letters and their gi| accession number. The five letters represent the genus name (first 3 letters) followed by 2 letters from the species name; Accordingly, CTEFE: *Ctenocephalides felix*, MUSDO: *Musca domestica*, PROTE: *Protophormia terraenovae*, STOCA: *Stomoxys calcitrans*, DROME: *Drosophila melanogaster*, ANOGA: *Anopheles gambiae*, ORNMO: *Ornithodoros moubata*, ARGMO: *Argas monolakensis*, HELVI: *Heliopsis virescens*, ARATH: *Arabidopsis thaliana*. The aminoterminal part of the peptides is not shown.

sis was performed by 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 the creation of *Sfi*I A & B restriction enzyme sites at the ends of the PCR products that are used for cloning into the phage vector. The PCR conditions were: 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 analysed on a 1.1% agarose gel to check for the quality and range of cDNA synthesised. 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) YM100 centrifugal filter device. The cleaned, doublestranded cDNA was then digested with *Sfi*I at 50 °C for 2 hrs, followed by size fractionation on a ChromaSpin-400 column (Clontech, Palo Alto, CA). 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 YM100.

The cDNA mixture was ligated into the λ TriplEx2 vector (Clontech, Palo Alto, CA) 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 *E. coli* cells (Clontech, Palo Alto, CA). The percentage of recombinant clones was determined by performing a blue-white selection screening on LB/MgSO₄ plates containing X-gal/IPTG. Recom-

binants were also determined by PCR, using vector primers (5' λ TriplEx2 Sequencing Primer and 3' λ TriplEx2 Sequencing) flanking the inserted cDNA and visualising the products on a 1.1% agarose/EtBr gel.

Sequencing of the *X. cheopis* cDNA library

The *X. cheopis* salivary gland cDNA library was plated on LB/MgSO₄ 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 MgSO₄; 7H₂ O; 0.035 M TrisHCl (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, four microliters of the phage sample was used as a template. The primers were sequences from the λ TriplEx2 vector and named pTEx2 5 seq (5'TCC GAG ATC TGG ACG AGC 3') and pTEx2 3 LD (5' ATA CGA CTC ACT ATA GGG CGA ATT GGC 3'), positioned at the 5' end and the 3' end of the cDNA insert, respectively. The reaction was carried out in 96 well flexible PCR plates (Fisher Scientific, Pittsburgh, PA) using the TaKaRa EX Taq polymerase (TAKARA Mirus Bio, Madison, WI), on a Perkin Elmer GeneAmp® PCR system 9700 (Perkin Elmer Corp., Foster City, CA). The PCR conditions were: one hold of 95 °C for 3 min; 25 cycles of

95°C for 1 min, 61°C for 30 sec; 72°C for 2 min. The amplified products were analysed on a 1.5% Agarose/EtBr gel. 1100 cDNA library 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 Thermo-Fast 96-well PCR plates (ABgene Corp., Epsom, Surrey, UK) and frozen at 20°C, before cycle sequencing using an ABI3730 XL machine.

Bioinformatic tools and procedures used

Expressed sequence tags (EST) were trimmed of primer and vector sequences, clusterized, and compared with other databases as described [44]. The BLAST tool [77], CAP3 assembler [78], ClustalW [79], and Treeview software [80] were used to compare, assemble, and align sequences and to visualise alignments. For functional annotation of the transcripts we used the tool BlastX [33] to compare the nucleotide sequences to the NR protein database of the National Center for Biotechnology Information (NCBI) and to the Gene Ontology (GO) database [34]. The tool RPSBlast [33] was used to search for conserved protein domains in the Pfam [81], SMART [82], Kog [83] and Conserved Domains Databases (CDD) [35]. We have also compared the transcripts with other subsets of mitochondrial and rRNA nucleotide sequences downloaded from NCBI, and to several organism proteomes downloaded from NCBI (yeast), Flybase (*Drosophila melanogaster*), or ENSEMBL (*An. gambiae*). 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 100 predicted AA, or to the predicted protein translation in the case of complete coding sequences, were submitted to the SignalP server [36] to help identify translation products that could be secreted. O-glycosylation sites on the proteins were predicted with the program NetOGlyc [84]. 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. Phylogenetic analysis and statistical Neighbor Joining (NJ) bootstrap tests of the phylogenies were done with the Mega package [85].

Gel electrophoresis studies

Flea salivary proteins representing approximately 100 gland pairs were resolved by one-dimensional (1D) sodium dodecylsulfate polyacrylamide gel electrophoresis (4–12% gradient gels) and visualised with Coomassie blue staining (Pierce, Rockford, IL). Excised gel bands were destained using 50% acetonitrile in 25 mM NH_4HCO_3 , pH 8.4 and vacuum dried. Trypsin (20 $\mu\text{g}/\text{mL}$ in 25 mM NH_4HCO_3 , pH 8.4) was added and the mixture was incubated on ice for one hr. The supernatant was

removed and the gel bands were covered with 25 mM NH_4HCO_3 , pH 8.4. After overnight incubation at 37°C, the tryptic peptides were extracted using 70% acetonitrile, 5% formic acid, and the peptide solution was lyophilised and desalted using ZipTips (Millipore, Bedford, MA).

Low MW fractionation of flea salivary proteins

A low molecular protein sample was prepared by resuspending 51 μg of total flea protein salivary homogenate into 2 mL of 100 mM NH_4HCO_3 , pH 8.4, containing 10% acetonitrile. Low MW proteins were obtained by centrifugal ultrafiltration using Centriplus 30 kDa ultrafilters (Millipore, Billerica, MA) spun at $750 \times g$ at 4°C. The low MW filtrate was lyophilised and resuspended in 50 μL of 25 mM NH_4HCO_3 , pH 8.4, and half of the solution was digested with trypsin (enzyme:protein ratio of 1:50) for 16 h at 37°C. The undigested and digested samples were desalted using C18 ZipTips (Millipore, Bedford, MA), lyophilised to dryness and resuspended in 14 μL 0.1% TFA for subsequent nanoRPLCMS/MS analysis.

Nanoflow reversedphase liquid chromatography tandem mass spectrometry (nanoRPLCMS/MS)

The tryptic peptides were analyzed using nanoRPLCMS/MS. A 75 μm i.d. \times 360 μm o.d. \times 10 cm long fused silica capillary column (Polymicro Technologies, Phoenix, AZ) was packed with 3 μm , 300 Å pore size C-18 silica bonded stationary RP particles (Vydac, Hesperia, CA). The column was connected to an Agilent 1100 nanoLC system (Agilent Technologies, Palo Alto, CA) that was coupled online with a linear iontrap (LIT) mass spectrometer (LTQ, ThermoElectron, San Jose, CA). The peptides were separated using a gradient consisting of mobile phase A (0.1% formic acid in water) and B was (0.1% formic acid in acetonitrile). The peptide samples were injected and gradient elution was performed under the following conditions: 2% B at 500 nL/min in 30 min; a linear increase of 2–42% B at 250 nL/min in 110 min; 42–98% in 30 min including the first 15 min at 250 nL/min and then 15 min at 500 nL/min; 98% at 500 nL/min for 10 min. The LIT-MS was operated in a data-dependent tandem MS (MS/MS) mode in which the five most abundant peptide molecular ions in every MS scan were selected for collision induced dissociation (CID) using a normalized collision energy of 35%. Dynamic exclusion was applied to minimize repeated selection of previously analyzed peptides. The capillary temperature and electrospray voltage were set to 160°C and 1.5 kV, respectively. Tandem MS spectra from the nanoRPLCMS/MS analyses were searched against a protein fasta database derived from the flea salivary gland, using SEQUEST operating on an 18 node Beowulf cluster. For a peptide to be considered legitimately identified, it had to achieve stringent charge state and proteolytic cleavage-dependent cross correlation (X_{corr}) and a minimum correlation (ΔC_n) score of 0.08.

Measurement of apyrase activity

Apyrase activity was measured as described previously [86]. Specific conditions are given in the legend accompanying Figure 4.

Authors' contributions

JFA made cDNA library, supervised DNA sequencing and RACE extensions, contributed to work supervision, data analysis and manuscript writing. BJH contributed salivary glands and manuscript writing.

DAL contributed to mass spectrometry part of the work and manuscript writing. TPC contributed to mass spectrometry part of the work and manuscript writing. TDV contributed to mass spectrometry part of the work and manuscript writing. VMP performed cloning and sequencing of the cDNA library. JMCR contributed to work supervision, apyrase experiments, data analysis and manuscript writing. All authors read and approved the final manuscript.

Additional material

Additional file 1

Supplemental table S1. Hyperlinked Microsoft Excel file with assembled EST's and various database comparisons. This file can be found at NCBI [87] and at BMC Genomics as supplemental data.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-102-S1.xls>]

Additional file 2

Supplemental table S2. Hyperlinked Microsoft Excel file with predicted salivary proteins and various database comparisons. This file can be found at NCBI [88] and at BMC Genomics as supplemental data.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-102-S2.xls>]

Additional file 3

Supplemental table S1. Hyperlinked Microsoft Excel file with CD39/apyrase proteins and various database comparisons. This file can be found at NCBI [89] and at BMC Genomics as supplemental data.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-102-S3.xls>]

Acknowledgements

This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases. This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract NOI-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. We are grateful to Nancy Shulman and Brenda Marshall for editorial assistance.

References

- Ribeiro JMC: **Blood-feeding arthropods: Live syringes or invertebrate pharmacologists?** *Infect Agents Dis* 1995, **4**:143-152.
- Grimaldi D, Engel M: **Evolution of the insects.** New York: Cambridge University Press; 2005.
- Rothschild M: **Recent advances in our knowledge of the order Siphonaptera.** *Annu Rev Entomol* 1975, **20**:241-259.
- Rai KS, Black WC IV: **Mosquito genomes: Structure, organization, and evolution.** *Adv Genet* 1999, **41**:1-33.
- Andersen JF, Weichsel A, Balfour CA, Champagne DE, Montfort WR: **The crystal structure of nitrophorin 4 at 1.5 Å resolution: transport of nitric oxide by a lipocalin-based heme protein.** *Structure* 1998, **6**(10):1315-1327.
- Champagne D, Nussenzveig 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.
- Valenzuela JG, Ribeiro JM: **Purification and cloning of the salivary nitrophorin from the hemipteran *Cimex lectularius*.** *J Exp Biol* 1998, **201**(Pt 18):2659-2664.
- 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 USA* 2005, **102**(3):594-599.
- Francischetti IM, Valenzuela JG, Ribeiro JM: **Anophelin: kinetics and mechanism of thrombin inhibition.** *Biochemistry* 1999, **38**(50):16678-16685.
- Valenzuela JG, Francischetti IM, Ribeiro JM: **Purification, cloning, and synthesis of a novel salivary anti-thrombin from the mosquito *Anopheles albimanus*.** *Biochemistry* 1999, **38**(34):11209-11215.
- 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):20802-20809.
- Ribeiro JM, Francischetti IM: **Role of arthropod saliva in blood feeding: Sialome and post-sialome perspectives.** *Annu Rev Entomol* 2002, **19**:19.
- Champagne DE, Smartt CT, Ribeiro JM, James AA: **The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5'-nucleotidase family.** *Proc Natl Acad Sci USA* 1995, **92**(3):694-698.
- Sun D, McNicol A, James AA, Peng Z: **Expression of functional recombinant mosquito salivary apyrase: A potential therapeutic platelet aggregation inhibitor.** *Platelets* 2006, **17**(3):178-184.
- Faudry E, Lozzi SP, Santana JM, D'Souza-Ault M, Kieffer S, Felix CR, Ricart CA, Sousa MV, Vernet T, Teixeira AR: ***Triatoma infestans* apyrases belong to the 5'-nucleotidase family.** *J Biol Chem* 2004, **279**(19):19607-19613.
- Faudry E, Santana JM, Ebel C, Vernet T, Teixeira AR: **Salivary apyrases of *Triatoma infestans* are assembled into homo-oligomers.** *Biochem J* 2006, **396**(3):509-515.
- 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):30583-30590.
- 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):229-237.
- Cheeseman MT: **Characterization of apyrase activity from the salivary glands of the cat flea *Ctenocephalides felis*.** *Insect Biochem Mol Biol* 1998, **28**(12):1025-1030.
- Ribeiro JMC, Vaughan JA, Azad AF: **Characterization of the salivary apyrase activity of three rodent flea species.** *Comp Biochem Physiol* 1990, **95B**:215-218.
- Deoras PJ, Prasad RS: **Feeding mechanisms of Indian fleas *X. cheopis* (Roths) and *X. astia* (Roths).** *Ind J Med Res* 1967, **55**:1041-1050.
- Cheeseman MT, Bates PA, Crampton JM: **Preliminary characterization of esterase and platelet-activating factor (PAF)-acetylhydrolase activities from cat flea (*Ctenocephalides felis*) salivary glands.** *Insect Biochem Mol Biol* 2001, **31**(2):157-164.
- Feingold BF, Benjamini E: **Allergy to flea bites. Clinical and experimental observations.** *Ann Allergy* 1961, **19**:1275-1289.
- Scheidt VJ: **Flea allergy dermatitis.** *Vet Clin North Am Small Anim Pract* 1988, **18**(5):1023-1042.
- McDermott MJ, Weber E, Hunter S, Stedman KE, Best E, Frank GR, Wang R, Escudero J, Kuner J, McCall C: **Identification, cloning, and characterization of a major cat flea salivary allergen (Cte f 1).** *Mol Immunol* 2000, **37**(7):361-375.

26. Lee SE, Jackson LA, Opdebeeck JP: **Salivary antigens of the cat flea, *Ctenocephalides felis felis*.** *Parasite Immunol* 1997, **19(1)**:13-19.
27. Ribeiro JM, Alarcon-Chaidez F, Francischetti IM, Mans BJ, Mather TN, Valenzuela JG, Wikel SK: **An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks.** *Insect Biochem Mol Biol* 2006, **36(2)**:111-129.
28. 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.
29. 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)**:3971-3986.
30. Bell JF, Stewart SJ, Wikel SK: **Resistance to tick-borne *Francisella tularensis* by tick-sensitized rabbits: Allergic kensidity.** *Am J Trop Med Hyg* 1979, **28**:876-880.
31. Edwards JF, Higgs S, Beatty BJ: **Mosquito feeding-induced enhancement of Cache Valley Virus (Bunyaviridae) infection in mice.** *J Med Entomol* 1998, **35(3)**:261-265.
32. Wikel SK, Ramachandra RN, Bergman DK, Burkot TR, Piesman J: **Infestation with pathogen-free nymphs of the tick *Ixodes scapularis* induces host resistance to transmission of *Borrelia burgdorferi* by ticks.** *Inf Immun* 1997, **65**:335-338.
33. 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)**:3389-3402.
34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25(1)**:25-29.
35. 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)**:281-283.
36. 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.
37. 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)**:2429-2451.
38. 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.
39. 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)**:543-563.
40. Novak MG, Ribeiro JMC, Hildebrand JG: **5-Hydroxytryptamine in the salivary glands of adult female *Aedes aegypti* and its role in regulation of salivation.** *J Exp Biol* 1995, **198**:167-174.
41. Novak MG, Rowley WA: **Serotonin depletion affects blood-feeding but not host-seeking in *Aedes triseriatus* (Diptera: Culicidae).** *J Med Entomol* 1994, **31**:600-606.
42. Calvo E, Andersen J, Francischetti IM, de LCM, deBianchi AG, James AA, Ribeiro JM, Marinotti O: **The transcriptome of adult female *Anopheles darlingi* salivary glands.** *Insect Mol Biol* 2004, **13(1)**:73-88.
43. Valenzuela JG, Pham VM, Garfield MK, Francischetti IM, Ribeiro JMC: **Toward a description of the sialome of the adult female mosquito *Aedes aegypti*.** *Insect Biochem Mol Biol* 2002, **32**:1101-1122.
44. Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Ribeiro JM: **Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito.** *Insect Biochem Mol Biol* 2003, **33(7)**:717-732.
45. Vincent JB, Crowder MW, Averill BA: **Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions.** *Trends Biochem Sci* 1992, **17(3)**:105-110.
46. Schneider G, Lindqvist Y, Vihko P: **Three-dimensional structure of rat acid phosphatase.** *Embo J* 1993, **12(7)**:2609-2615.
47. Jakob CG, Lewinski K, Kuciel R, Ostrowski W, Lebioda L: **Crystal structure of human prostatic acid phosphatase.** *Prostate* 2000, **42(3)**:211-218.
48. **Anobase** [<http://www.anobase.org>]
49. Moss DW, Raymond FD, Wile DB: **Clinical and biological aspects of acid phosphatase.** *Crit Rev Clin Lab Sci* 1995, **32(4)**:431-467.
50. Ronnberg L, Vihko P, Sajanti E, Vihko R: **Clomiphene citrate administration to normogonadotropic subfertile men: blood hormone changes and activation of acid phosphatase in seminal fluid.** *Int J Androl* 1981, **4(3)**:372-378.
51. Lin MF, Meng TC: **Tyrosine phosphorylation of a 185 kDa phosphoprotein (pp185) inversely correlates with the cellular activity of human prostatic acid phosphatase.** *Biochem Biophys Res Commun* 1996, **226(1)**:206-213.
52. Meng TC, Lin MF: **Tyrosine phosphorylation of c-ErbB-2 is regulated by the cellular form of prostatic acid phosphatase in human prostate cancer cells.** *J Biol Chem* 1998, **273(34)**:22096-22104.
53. Gerasimovskaya EV, Ahmad S, White CW, Jones PL, Carpenter TC, Stenmark KR: **Extracellular ATP is an autocrine/paracrine regulator of hypoxia-induced adventitial fibroblast growth. Signaling through extracellular signal-regulated kinase-1/2 and the Egr-1 transcription factor.** *J Biol Chem* 2002, **277(47)**:44638-44650.
54. Lautrette C, Loum-Ribot E, Petit D, Vermot-Desroches C, Wijdenes J, Jauberteau MC: **Increase of Fas-induced apoptosis by inhibition of extracellular phosphorylation of Fas receptor in Jurkat cell line.** *Apoptosis* 2006.
55. Babinska A, Hogan MV, Sobocki T, Sobocka MB, Ehrlich YH, Kornecki E: **Identification of ecto-PKC on surface of human platelets: role in maintenance of latent fibrinogen receptors.** *Am J Physiol Heart Circ Physiol* 2000, **278(6)**:H2008-2019.
56. Morgenstern E, Gnad U, Preissner KT, Dierichs R, Belleli A, Chesnutkin A, Schwartz I, Shaltiel S: **Localization of protein kinase A and vitronectin in resting platelets and their translocation onto fibrin fibers during clot formation.** *Eur J Cell Biol* 2001, **80(1)**:87-98.
57. Naik UP, Kornecki E, Ehrlich YH: **Phosphorylation and dephosphorylation of human platelet surface proteins by an ecto-protein kinase/phosphatase system.** *Biochim Biophys Acta* 1991, **1092(2)**:256-264.
58. Smith SA, Mutch NJ, Baskar D, Rohloff P, Docampo R, Morrissey JH: **Polyporphosphate modulates blood coagulation and fibrinolysis.** *Proc Natl Acad Sci USA* 2006, **103(4)**:903-908.
59. Krejci E, Duval N, Chatonnet A, Vincens P, Massoulie J: **Cholinesterase-like domains in enzymes and structural proteins: functional and evolutionary relationships and identification of a catalytically essential aspartic acid.** *Proc Natl Acad Sci USA* 1991, **88(15)**:6647-6651.
60. Argentine JA, James AA: **Characterization of a salivary gland-specific esterase in the vector mosquito, *Aedes aegypti*.** *Insect Biochem Mol Biol* 1995, **25(5)**:621-630.
61. Wang TF, Handa M, Guidotti G: **Structure and function of ectoapyrase (CD39).** *Drug Develop Res* 1998, **45**:245-252.
62. Ivanenkov VV, Murphy-Piedmonte DM, Kirley TL: **Bacterial expression, characterization, and disulfide bond determination of soluble human NTPDase6 (CD39L2) nucleotidase: implications for structure and function.** *Biochemistry* 2003, **42(40)**:11726-11735.
63. Mulero JJ, Yeung G, Nelken ST, Ford JE: **CD39-L4 is a secreted human apyrase, specific for the hydrolysis of nucleoside diphosphates.** *J Biol Chem* 1999, **274(29)**:20064-20067.
64. Charlab R, Rowton ED, Ribeiro JM: **The salivary adenosine deaminase from the sand fly *Lutzomyia longipalpis*.** *Exp Parasitol* 2000, **95(1)**:45-53.
65. Ribeiro JM, Charlab R, Valenzuela JG: **The salivary adenosine deaminase activity of the mosquitoes *Culex quinquefasciatus* and *Aedes aegypti*.** *J Exp Biol* 2001, **204(Pt 11)**:2001-2010.
66. Tilley SL, Wagoner VA, Salvatore CA, Jacobson MA, Koller BH: **Adenosine and inosine increase cutaneous vasopermeability by activating A(3) receptors on mast cells.** *J Clin Invest* 2000, **105(3)**:361-367.
67. King TP, Spangfort MD: **Structure and biology of stinging insect venom allergens.** *Int Arch Allergy Immunol* 2000, **123(2)**:99-106.
68. Li S, Kwon J, Aksoy S: **Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*.** *Insect Mol Biol* 2001, **10(1)**:69-76.
69. 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.
70. Schreiber MC, Karlo JC, Kovalick GE: **A novel cDNA from *Drosophila* encoding a protein with similarity to mammalian cysteine-rich secretory proteins, wasp venom antigen 5, and plant group I pathogenesis-related proteins.** *Gene* 1997, **191(2)**:135-141.
71. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14(9)**:755-763.
72. Bontems F, Roumestand C, Gilquin B, Menez A, Toma F: **Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins.** *Science* 1991, **254(5037)**:1521-1523.

73. Cociancich S, Goyffon M, Bontems F, Bulet P, Bouet F, Menez A, Hoffmann J: **Purification and characterization of a scorpion defensin, a 4 kDa antibacterial peptide presenting structural similarities with insect defensins and scorpion toxins.** *Biochem Biophys Res Commun* 1993, **194**(1):17-22.
74. Ganz T: **Defensins: antimicrobial peptides of innate immunity.** *Nat Rev Immunol* 2003, **3**(9):710-720.
75. Rodriguez de la Vega RC, Possani LD: **Overview of scorpion toxins specific for Na⁺ channels and related peptides: biodiversity, structure-function relationships and evolution.** *Toxicon* 2005, **46**(8):831-844.
76. Brain SD, Williams TJ, Tippins JR, Morris HR, MacIntyre I: **Calcitonin gene-related peptide is a potent vasodilator.** *Nature (Lond)* 1985, **313**:54-56.
77. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
78. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9**(9):868-877.
79. 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):4876-4882.
80. Page RD: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**(4):357-358.
81. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL: **The Pfam protein families database.** *Nucleic Acids Res* 2000, **28**(1):263-266.
82. 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):231-234.
83. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al.: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**(1):41.
84. Hansen JE, Lund O, Tolstrup NI, 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):115-130.
85. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**(2):150-163.
86. Marinotti O, James A, Ribeiro JMC: **Diet and salivation in female *Aedes aegypti* mosquitoes.** *J Insect Physiol* 1990, **36**:545-548.
87. **Supplemental table S1 for *X. cheopis sialome*** [http://www.ncbi.nlm.nih.gov/projects/omes/X_cheopis_sialome/Table_1/XC-SupTable1.xls]
88. **Supplemental table S2 for *X. cheopis sialome*** [http://www.ncbi.nlm.nih.gov/projects/omes/X_cheopis_sialome/Table_2/XC-SupTable2.xls]
89. **Supplemental table S3 for *X. cheopis sialome*** [http://www.ncbi.nlm.nih.gov/projects/omes/X_cheopis_sialome/Table_3/XC-SupTable3.xls]