Published in final edited form as:

Insect Biochem Mol Biol. 2014 January ; 44: 23–32. doi:10.1016/j.ibmb.2013.10.006.

An insight into the sialome of the frog biting fly, *Corethrella appendiculata*

José M.C. Ribeiro^{a,*}, Andrezza C. Chagas^a, Van M. Pham^a, L.P. Lounibos^b, and Eric Calvo^a

^aLaboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway, Rockville, MD 20852, USA

^bFlorida Medical Entomology Laboratory, University of Florida, 200 9th Street SE, Vero Beach, FL 32962-4657, USA

Abstract

The Nematocera infraorder Culicomorpha is believed to have descended from bloodfeeding ancestors over 200 million years ago, generating bloodfeeding and non-bloodfeeding flies in two superfamilies, the Culicoidea—containing the mosquitoes, the frog-feeding midges, the Chaoboridae, and the Dixidae—and the Chironomoidea—containing the black flies, the ceratopogonids, the Chironomidae, and the Thaumaleidae. Bloodfeeding requires many adaptations, including development of a sophisticated salivary potion that disarms host hemostasis, the physiologic mechanism comprising platelet aggregation, vasoconstriction, and blood clotting. The composition of the sialome (from the Greek *sialo* = saliva) from bloodfeeding animals can be inferred from analysis of their salivary gland transcriptome. While members of the mosquitoes, black flies, and biting midges have provided sialotranscriptome descriptions, no species of the frog-biting midges has been thus analyzed. We describe in this work the sialotranscriptome of *Corethrella appendiculata*, revealing a complex potion of enzymes, classical nematoceran protein families involved in bloodfeeding, and novel protein families unique to this species of frog-feeding fly. Bacterial (*Wolbachia*) and novel viral sequences were also discovered.

Keywords

Evolution; Culicomorpha; Corethrellidae; Salivary gland; Hematophagy; Medical entomology

1. Introduction

The bloodfeeding mode appeared independently at least ten times within insects (Grimaldi and Engel, 2005; Ribeiro and Arca, 2009). Among the adaptations to this peculiar diet, these organisms evolved a sophisticated salivary potion that disarms their hosts' hemostasis, a complex and redundant system based on the triad of platelet aggregation, vasoconstriction,

*Corresponding author. Vector Biology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway, Room 2E-32D, Rockville, MD 20852, USA. Tel.: +1 301 496 9389; fax: +1 301 480 2571. jribeiro@niaid.nih.gov (J.M.C. Ribeiro).

Appendix A. Supplementary material

Supplementary data related to this article can be found online at <http://dx.doi.org/10.1016/j.ibmb.2013.10.006>.

and blood clotting (Ribeiro et al., 2010; Ribeiro and Arca, 2009). In the lower flies (suborder Nematocera), bloodfeeding evolved at least within the sand flies (family Psychodidae within the Psychodomorpha) and independently within the Culicomorpha, where several families are adept at hematophagy—including mosquitoes (Culicidae), black flies (Simuliidae), biting midges (Ceratopogonidae), and the frog-biting midges (Corethrellidae) (Grimaldi and Engel, 2005). Only adult females are blood feeders in this context. Adults also take sugar meals and have salivary glycosidases and antimicrobials that may assist digestion and microbial containment of the sugar meal that is stored in their crops, a cuticle lined, inert organ (Ribeiro et al., 2010).

The common ancestor of all Culicomorpha must have existed in the Triassic period, over 200 million years ago (MYA), producing 8 extant families of which 4 are blood feeders; the remaining having lost this ancestral feeding mode (Grimaldi and Engel, 2005). Accordingly, these flies were bloodfeeding on vertebrates before the radiation of mammals 60 MYA, giving ample time for diversification of their salivary potion. Additionally, the immune pressure imposed by their hosts on salivary antigens may have further accelerated the evolution of these proteins, leading to further salivary protein diversification (Ribeiro et al., 2010; Ribeiro and Arca, 2009). For these reasons, the protein composition of the salivary glands (sialome, from the Greek sialo = saliva)—even from closely related organisms—is quite divergent, and even between related genera from the same family there are two to five new protein families found, new in the sense that they have no sequence similarities to any known protein.

Blood feeders' sialomes can be revealed by transcriptome analysis of their salivary glands (SGs), revealing nearly 40 proteins in sand flies and ~100 in mosquitoes, black flies, and biting midges (Ribeiro et al., 2010); however, no sialome from frog-biting midges exists so far. This work describes the sialome of *Corethrella appendiculata*, an autogenous fly “that has well-developed biting mouthparts and are attracted to frog calls, indicating that in nature it undergoes additional ovarian cycles” (McKeever and French, 1991). As the Corethrellidae are regarded as evolutionarily older than their sister families the Culicidae and Chaoboridae (Grimaldi and Engel, 2005), their blood meal specificity for anuran hosts may have evolved in the late Jurassic period and conceivably required biochemical adaptations different from culicomorphs that blood feed from endothermic vertebrates.

2. Material and methods

2.1. Fly rearing, salivary gland (SG) collection, and mRNA extraction

Larvae from *C. appendiculata*, reared as described previously (Lounibos et al., 2008) at the Florida Medical Entomology Laboratory (University of Florida in Vero Beach), were obtained from a Florida colony originating from discarded tires and tree-hole collections of 200–300 immatures. Although bloodfeeding from tree frogs by this species has been observed in both the field and laboratory, this colony is maintained autogenously, i.e., without access to a blood source. Fourth instar larvae were sent to the NIH laboratory where they were fed *Aedes aegypti* first and second instar larvae as food. Pupae were collected 2–3 times per day and allowed to emerge in small plastic cups with deionized water covered by a fine mesh in an insectary maintained at 26 ± 0.5 °C. Adults had access to a cotton swab

containing 20% corn syrup. Adult females had their SGs removed at day 0, 1, or 2 by dissection under phosphate buffered saline. Dissected glands were transferred to 0.05 ml RNAlater (Invitrogen, San Diego, CA) with a small needle. Glands in RNAlater were stored at 4 °C for 48 h before being transferred to -70 °C until RNA extraction. SG RNA was extracted and isolated using the Micro-FastTrack mRNA isolation kit (Invitrogen) per manufacturer's instructions. The integrity of the total RNA was checked on a Bioanalyser (Agilent Technologies, Santa Clara, CA).

2.2. Next-generation sequencing and bioinformatic analysis

mRNA library construction and sequencing were done by the NIH Intramural Sequencing Center. The SG library was constructed using the TruSeq RNA sample prep kit, v. 2 (Illumina Inc., San Diego, CA). The resulting cDNA was fragmented using a Covaris E210 (Covaris, Woburn, MA). Library amplification was performed using eight cycles to minimize the risk of over-amplification. Sequencing was performed on a HiSeq 2000 (Illumina) with v. 3 flow cells and sequencing reagents. One lane of the HiSeq machine was used for this and two other libraries, distinguished by bar coding. These raw data are available at the Sequence Read Archives of the National Center for Biotechnology Information under bioproject number PRJNA213247 and raw data file SRR951913. A total of 151,646,242 sequences of 101 nucleotides in length were obtained. A paired-end protocol was used. Raw data were processed using RTA 1.12.4.2 and CASAVA 1.8.2. Reads were trimmed of low quality regions (<10), and only those with an average quality of 20 or more were used, comprising a total of 121,205,872 high-quality reads. These were assembled with the ABySS software (Genome Sciences Centre, Vancouver, BC, Canada) (Birol et al., 2009; Simpson et al., 2009) using various kmer (k) values (every even number from 24 to 96). Because the ABySS assembler tends to miss highly expressed transcripts (Zhao et al., 2011b), the SOAPdenovo-Trans assembler (Luo et al., 2012) was also used, again with odd kmers from 23 to 95. The resulting assemblies were joined by an iterative BLAST and cap3 assembler (Karim et al., 2011). Sequence contamination between bar-coded libraries were identified and removed when their sequence identities were over 98% but their abundance of reads were >10 fold between libraries. Coding sequences (CDS) were extracted using an automated pipeline based on similarities to known proteins or by obtaining CDS containing a signal peptide (Nielsen et al., 1999). CDS and their protein sequences were mapped into a hyperlinked Excel spreadsheet (presented as Supplemental File 1). Signal peptide, transmembrane domains, furin cleavage sites, and mucin-type glycosylation were determined with software from the Center for Biological Sequence Analysis (Technical University of Denmark, Lyngby, Denmark) (Duckert et al., 2004; Julenius et al., 2005; Nielsen et al., 1999; Sonnhammer et al., 1998). Reads were mapped into the contigs using blastn (Altschul et al., 1997) with a word size of 25, masking homonucleotide decamers and allowing mapping to up to three different CDS if the BLAST results had the same score values. Mapping of the reads was also included in the Excel spreadsheet. Automated annotation of proteins was based on a vocabulary of nearly 250 words found in matches to various databases, including Swissprot, Gene Ontology, KOG, Pfam, and SMART, and a subset of the non-redundant protein database of the National Center for Biotechnology Information containing proteins from vertebrates. Further manual annotation was done as required. Detailed bioinformatics analysis of our pipeline can be found in our previous

publication (Karim et al., 2011). Sequence alignments were done with the ClustalX software package (Thompson et al., 1997). Phylogenetic analysis and statistical neighbor-joining bootstrap tests of the phylogenies were done with the Mega package (Kumar et al., 2004). BLAST score ratios were done as indicated previously (Rasko et al., 2005).

2.3. Data availability

The raw reads used in this manuscript were deposited at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under bioproject PRJA213247 and run SRR951913. A total of 5052 coding sequences have been deposited to DDBJ/EMBL/GenBank through the Transcriptome Shotgun Annotation portal under the accession GANO000000000. The version described in this paper is the first version, GANO01000000.

3. Results

3.1. General assembly characteristics

The depth of coverage of this transcriptome allowed for the extraction of 8943 CDS (see Methods). Of these, 4276 covered 90% or more of the full length of deduced proteins from *Culex quinquefasciatus* with a blastp e value of $1e-15$ or better. Reads were mapped to each CDS, thus allowing an estimate of transcript abundance. Accordingly, 51,167,679 reads were mapped to the extracted CDS, representing 42% of the ~121 million high-quality reads. Automatic and manual annotation of the contigs allowed their general classification as belonging to the salivary secreted (S) class, housekeeping (H), unknown (U), transposable element (TE), viral (V) or bacterial (B) classes (Table 1). Although the S class contained only 9.5% of the 8971 CDS, it mapped over 18 million reads amounting to 36% of the total. H-class CDS accounted for 78% of the total and 58% of the mapped reads. Transposable elements were relatively highly expressed, with 532 CDS and 2% of all reads.

As a relative index of CDS expression, we normalized to 100 the CDS that had the higher number of reads (a member of the D7 family of proteins), and we will refer to other CDS as having an expression index (EI) relative to this maximally expressed contig. Thus a transcript having EI = 10 has 10% of the expression level as the most-expressed CDS, as measured by its number of mapped

3.2. Bacterial sequences

Eleven extracted CDS produced best match to *Wolbachia pipientis* (identities of 56–98% at the amino acid level), the endosymbiont bacteria of *Culex pipiens* (Supplemental File S1). These included several ribosomal proteins, elongation factors, ATP synthases, and other proteins, indicating *C. appendiculata* to have a *Wolbachia* symbiont, as commonly found with many insect species (Stouthamer et al., 1999). It is also possible and more likely (due to the polyadenylation selection of the RNA used for sequencing) that these sequences derive from *Wolbachia* genes that became integrated into the fly genome, an increasingly common finding as insect and other invertebrate genomic sequences accumulate (Klasson et al., 2009; McNulty et al., 2012; Woolfit et al., 2009).

3.3. Viral sequences

Two extracted CDS, coding for proteins over 1000 amino acids (aa), best match Midway and Nyamanini viruses RNA-dependent RNA polymerases—two related viruses cultured from birds, cattle, and their ticks—that define the genus Nyavirus (Kuhn et al., 2013; Mihindukulasuriya et al., 2009). The two extracted *C. appendiculata* CDS are only 41% identical among themselves. They are relatively highly expressed, having 28 and 37 thousand mapped reads each and EIs of 1.3 and 1.7. Phylogenetic analysis (Supplemental Fig. S1) shows the *C. appendiculata* viral sequences to share a clade of strong bootstrap support with Bornaviridae, the above-mentioned pair of Nyavirus, and also an unclassified Mononegavirales from the soybean cyst nematode (Bekal et al., 2011). Whether these viruses infect *Corethrella* or have become integrated as highly expressed endogenous viruses remains to be determined.

3.4. Transposable elements

Full-length and near full-length polyproteins of Class I elements were identified, including LTR retrotransposons of the BEL, Gypsy, and Copia subfamilies, some with over 100 representatives (Gypsy). On the NLTR family, the CR1, Crack, Daphne, I, L1, Kiri, and Jockey subfamilies are also well represented, with over 150 CDS. Class II elements of the hAT, Mariner, and P families are also found, including what appear to be full-length Mariner transposases. These sequences can be inspected in Supplemental File S1 under the Transposable element header.

3.5. Housekeeping sequences

Near 30 million reads mapped to CDS with a putative cellular function, of which the subclass of “Unknown conserved” had 1119 CDS and 19% of the reads (Table 2). Not surprisingly, protein synthesis-related CDS had 13% of the total reads with 399 CDS, followed by signal transduction-related CDS with 9.5% of the reads. Somewhat surprising was the storage subclass, with only 13 CDS to account for 6% of all reads. This is caused, however, by the presence of several contigs coding for vitellogenin, which reflect the autogenous character of this fly. Two of the contigs have EI > 23 and another 2 with EI > 13; these transcripts may have derived from fat body contamination that adhered to the SGs upon dissection. Indicative of the power of deep sequencing is the recovery of many neurohormones such as allatostatin, allatotropin, adipokinetic hormone, dromyosuppressin, short neuropeptide F, diuretic hormone, and insulin-like peptide precursor, which were assembled from 90 to 4902 reads.

3.6. Putative secreted proteins

The complexity of the *C. appendiculata* sialome is somewhat surprising, including protein families found in mosquitoes and black flies in addition to having its unique protein families, totaling 819 CDS classified into 57 subclasses arranged under 7 major categories, as follows (Table 3): Enzymes, Protease inhibitor domains, Immunity related, Other ubiquitous families, Insect conserved families, Nematocera-specific salivary domains, and Putative novel secreted proteins. While most of the 57 subclasses represent well defined gene families (such as serpins, or cystatins), others are open ended and could contain several

unrelated gene families, such as the subclass “Other proteins unique of nematocera”. This increased complexity may derive from the depth of coverage achieved in this transcriptome compared with previous sialotranscriptomes that were performed with a few thousands Sanger-derived ESTs. Indeed, when we restrict counting of the CDS to those having an EI = 1 or higher (e.g. having at least 1% of the number of mapped reads found in the most expressed transcript), the number of CDS falls from 819 to 118, although these 118 account for 94% of the total reads of the S class. Notice that an EI = 1 in our study represents over 20,000 reads mapped to that CDS. Some highlights follow where insights regarding the evolution of bloodfeeding can be derived from phylogenetic analysis or from other comparisons to other blood-feeding Nematocera. We will pay particular attention to CDS of the S class that have an EI > 1 as being of additional relevance.

3.6.1. Enzymes—A complete list of the putative salivary secreted enzymes found in the *C. appendiculata* sialotranscriptome is provided in Table 3 and Supplemental File S1.

3.6.1.1. Apyrase/5' nucleotidase: *C. appendiculata* has abundantly expressed salivary 5' nucleotidase/apyrases, two of which have over 1.2 million mapped reads. This enzyme family can destroy ADP, a mediator of platelet aggregation (Champagne et al., 1995; Sarkis et al., 1986), ATP, an inducer of neutrophil aggregation (Kuroki and Minakami, 1989; O'Flaherty and Cordes, 1994) and inflammasome activation (Bours et al., 2011), or AMP, which may induce mast cell degranulation (Ribeiro et al., 2010); however, the hemostatic/inflammatory system of frogs is not that well known. Supplemental Fig. S2 shows the phylogenetic tree of five *C. appendiculata* sequences resulting from their alignment with their GenBank matches. CorSigP-95137, coding for a highly expressed transcript (EI = 5.7), clusters with strong bootstrap support with mosquito and black fly enzymes that were found in previous sialotranscriptomes (clade I in Supplemental Fig. S2). Notice that this clade has a single *C. appendiculata* sequence, but individual mosquito species have at least two members in the clade, indicative of relatively recent gene duplication. The finding of two apyrase coding genes expressed in mosquito SGs has been noted previously (reviewed in (Ribeiro et al., 2010)). The second abundant *C. appendiculata* enzyme sequence (EI = 6.1) clusters with single-copy sequences found in mosquitoes and fruit flies (Clade III in Supplemental Fig. S2). A less-abundant *C. appendiculata* sequence is also present in this clade. The remaining two least-expressed *C. appendiculata* sequences cluster with additional single-copy clades (II and IV, Supplemental Fig. S2). This analysis suggests that clade I produced the primordial salivary enzyme in the ancestral bloodfeeding Culicomorpha; after Culicidae separated from Corethrellidae, a gene duplication in clade I produced two mosquito salivary-expressed apyrase transcripts, while in *C. appendiculata*, the gene duplication arose in clade III with recruitment of this transcript for a salivary secretory fate.

3.6.1.2. Adenosine deaminase/purine nucleosidase: These two enzymes promote further catabolism of adenosine to inosine and then to hypoxanthine plus ribose. They are found highly expressed in the *C. appendiculata* sialotranscriptome, with near 100,000 reads each (EI = 4.4 and 4.7). Transcripts coding for these enzymes were reported previously in mosquitoes and sand flies but not black flies or *Culicoides* (Ribeiro et al., 2010).

3.6.1.3. Inositol polyphosphate phosphatase: Over 100,000 reads mapped to a CDS encoding a member of this protein family (EI = 5.9), including a typical signal sequence indicative of secretion. This class of enzymes was never annotated as secreted in Diptera sialotranscriptomes but they abound in Triatomine and bed bug sialomes (Ribeiro et al., 2012). Their function is unknown, unless they are able to be delivered intracellularly to their hosts, and in this case they should shut down the inositol phosphate signaling pathway that is important in smooth muscle and immune cell activation (Kashiwada et al., 2007; Majerus et al., 1991). Perhaps the intracellular delivery could be promoted by exosome-based secretion (O'Loughlin et al., 2012).

3.6.1.4. DNase/hyaluronidase: This enzyme pair is commonly found in sand flies, black flies, and in *Culex*, but not *Aedes* or *Anopheles* mosquitoes. Their EI is similar, 1.2 and 1.0, respectively. They may help decrease skin viscosity and promote diffusion of pharmacologically active peptides deeper into the pre-capillary arterioles (Ribeiro et al., 2010). DNases also counteract neutrophil extracellular traps (Wartha et al., 2007), if their equivalent exists in frogs.

3.6.1.5. Serine proteases and trypsin-like proteins: Sixty-four serine protease-coding transcripts were found expressed in the *C. appendiculata* sialotranscriptome. Serine proteases are commonly found in Culicomorpha sialotranscriptomes (but not in sand flies, where only one salivary trypsin was found in *Phlebotomus ariasi*) (Ribeiro et al., 2010). Their specific salivary function is still unknown, but in tabanids it is associated with fibrinolytic activity (Xu et al., 2008). *C. appendiculata* has a highly expressed trypsin (CorSigP-94556; EI = 5.36) that is 46% identical to a digestive enzyme of *Aedes taeniorhynchus* (Zhao et al., 2011a), its nearest match from the NR database. Several serine proteases were found containing the clip domain indicative of specialized proteases, and one contains two complement control modules (CorSigP-1748) and possibly is implicated in complement activation. In addition to these 64 transcripts—which have rpsblast matches to the trypsin module (Tryp_SPc) of the CDD database with e values <1e-15—a group of CDS have much higher e values to this same domain (0.5–0.0001) and poor matches by blastp to insect enzymes, with 23–28% identities and 39–46 similarities, indicating great sequence divergence. Seven of these CDS have high expression (EI ranging from 2.1 to 6.9).

3.6.1.6. Dipeptidyl-peptidases: These enzymes are common in tick sialomes, where it is associated with a kininase activity (Ribeiro and Mather, 1998). Transcripts are also common in sand fly sialotranscriptomes, but not in Culicomorpha. The *C. appendiculata* sialotranscriptome reveals a highly expressed CDS coding for a member of the Angiotensin I-converting enzymes of the M2 family, with an EI = 4.6. This enzyme may be involved with the breakdown of inflammatory/ hemostatic peptides.

3.6.2. Protease inhibitors—Salivary proteins with protease inhibitor domains are associated with inhibition of vertebrate clotting or complement cascades as well as with antimicrobial activity. Members of the serpin, TIL, Kazal, metalloprotease inhibitor, pacifastin, and cystatin were found in *C. appendiculata*, with highly expressed members of the serpin, pacifastin, and cystatin families. No Kunitz domain peptides were found

belonging to the S class, these being abundant in *Culicoides* and black fly sialotranscriptomes (Three CDS were found to contain the Kunitz domain in *C. appendiculata*, but they are very similar to *Drosophila* conserved proteins, being poorly expressed and unlikely to be secreted). Recently a black fly Kunitz peptidewas shown to be a factor Xa inhibitor (Tsujimoto et al., 2012).

3.6.2.1. Serpins: The salivary anti-clotting proteins of culicine mosquitoes are serpins, but not those of *Anopheles* or black flies (Ribeiro et al., 2010). *Ae. aegypti* and *Aedes albopictus* salivary serpins have been characterized as factor Xa inhibitors (Calvo et al., 2011; Stark and James, 1998). Fifteen CDS coding for this protein family were identified in the *C. appendiculata* sialotranscriptome, two of which are highly transcribed with EI = 3.4 and 12.16, suggesting these proteins may work as anti-clotting agents or perhaps inhibit frog complement activation (Nonaka and Kimura, 2006; Romano et al., 1973).

3.6.2.2. Cystatins: Salivary cystatins were characterized from *Ixodes scapularis* ticks and shown to be immunosuppressive and anti-inflammatory (Schwarz et al., 2012). They were found expressed in *Ae. aegypti* sialotranscriptome, but its role in feeding, if any, is unknown. *C. appendiculata* has a highly expressed peptide (EI = 6.5) having a CDD match to the CY domain and matches to proteins annotated as cystatin in the Swissprot database.

3.6.2.3. Pacifastins: Pacifastins are serine protease inhibitors found in invertebrates and associated with control of the immune response, development, and reproduction (Breugelmans et al., 2009). They were not previously associated with bloodfeeding arthropod sialotranscriptomes. CorSigP-101443 is relatively well expressed, with over 23,000 reads and an EI = 1.04, and may play a role in bloodfeeding.

3.6.3. Immunity-related CDS—This class includes pathogen recognition proteins such as lectins, Gram-negative binding protein, peptidoglycan recognition proteins, and ML domain proteins as well as the complement-like thioester proteins, lysozymes, and antimicrobial peptides of the cecropin, defensin, and attacin classes. Within this group of transcripts, the cecropin encoded by Co-95050 is highly expressed, mapping 184,451 reads with an EI = 8.3; the peptidoglycan recognition protein (CorSigP-95,922) is also highly expressed, with and EI = 5.9. A phenoloxidase inhibitor proteinwas also discovered, but it may have a housekeeping role. The power of the RNAseq assembly can be appreciated by the achievement in this transcriptome analysis of having assembled two near full-length thioester proteins that are over 1400 aa in length from 5000 to 8000 reads each (Co-788 and CorSigP-93172), both covering >99% of the length of their mosquito homologs.

3.6.4. Other ubiquitous protein families

3.6.4.1. CAP/antigen-5 family: Members of this family are ubiquitously found in animal and plants (Gibbs et al., 2008) and have been consistently identified in sialotranscriptomes of hematophagous arthropods (Ribeiro et al., 2010; Ribeiro and Arca, 2009). Recently, a member of the family Triatomidae was shown to be antiinflammatory by virtue of its superoxide dismutase activity (Assumpcao et al., 2013). The *C. appendiculata* sequences

CorSigP-94647 and Co-94486 are highly expressed (EI = 23.5 and 7.72, respectively) members of the antigen 5 family, among seven other related CDS.

3.6.4.2. Lipocalins: Lipocalins are widespread in nature and abundantly found in tick and triatomine sialotranscriptomes but are occasionally found in other organisms, usually at low expression levels. Three CDS coding for apolipoprotein D-like lipocalins were found in the *C. appendiculata* sialome and may have a housekeeping function.

3.6.4.3. Other hypothetical secreted conserved proteins: Supplemental File S1 presents an additional 63 CDS coding for conserved hypothetical proteins with a signal peptide indicative of secretion. Of note, Co-99492, Co-94589, Co-100254, and CorSigP-99625 have EI values varying from 1.2 to 10 and code for His-rich polypeptides. These may function as antimicrobials by their metal (Zn, Cu, or Ni) chelating abilities (Lai et al., 2004; Loomans et al., 1998; Rothstein et al., 2002; Wiesner and Vilcinskas, 2010).

3.6.5. Insect-specific protein families

3.6.5.1. Yellow/major royal jelly family: This is typically a multifamily protein among Insecta. One of its members in *Drosophila* causes a phenotype of yellow cuticle when the gene is inactivated, and it has been shown that, in mosquitoes, one of these genes has dopachrome isomerase activity (Han et al., 2002; Johnson et al., 2001). Two members of this family are well expressed in sand fly SGs, where they function as scavengers, or kratagonists, of serotonin (Xu et al., 2011). This protein family was not found expressed in previously reported Culicomorpha sialomes but may have been found in this work by virtue of its deep-sequencing technology.

3.6.5.2. Insect allergen repeat family: Three CDS from the *C. appendiculata* sialotranscriptome code for proteins containing the Pfam Ins_allergen_rp motif associated with nitrile-specifier detoxification. This protein family has not been found previously in sialotranscriptomes, but transcripts have been previously found expressed in sand fly midgut (Ramalho-Ortigao et al., 2007). Co-96715 is relatively well expressed, with an EI = 1.24, and may be secreted in saliva.

3.6.5.3. Simulium/Culicoides conserved insect family: A conserved but diverse insect protein family of average molecular weight (75 kDa) of unknown function was previously characterized from black fly and *Culicoides* sialotranscriptomes (Ribeiro et al., 2010). This family is also found highly expressed in *C. appendiculata*, five members of which have EI varying from 1.2 to 9.1. Phylogenetic analysis (Supplemental Fig. S3) shows a clade with strong bootstrap support containing *C. appendiculata* and *Simulium* sequences quite divergent from other insect sequences where clear clades of Brachycera, mosquito, and Hymenoptera converge to a single clade with 99% bootstrap support. The phylogenetic tree also shows that *C. appendiculata* has at least six very divergent members, indicating fast evolution/divergence of this gene cluster.

3.6.5.4. OBP/D7 protein family: The odorant-binding protein family is widespread in insects, where it is well described as part of the olfaction machinery (Sanchez-Gracia et al.,

2009). The D7 family apparently derived from the OBP family by exon and gene duplications; the canonical OBP family has six α -helices, whereas D7 proteins have two additional helices (Mans et al., 2007). Additionally, two types of OBP/D7 exist—those with only one and those with two domains—defining the short and long D7 forms (Valenzuela et al., 2002). Both canonical OBPs and true D7 proteins have been described in *Nematocera* sialotranscriptomes (Ribeiro et al., 2010). These proteins function as chelators of inflammation and hemostasis agonists such as serotonin, histamine, thromboxane A₂, and leukotrienes. One member is known to inhibit kallikrein activation and prevent bradykinin synthesis (Isawa et al., 2007). They are normally highly expressed, mainly those binding serotonin and/or histamine. Mosquitoes have three genes coding for long forms and five coding for short forms, although only two genes coding for long forms and four coding for short forms are well expressed (Arca et al., 2005). Phylogenetic analysis of the *C. appendiculata* OBP/D7 proteins derived from their alignments with matching sequences found on the NR database (Supplemental Fig. S4) indicates that Co-94700 and Co-94890 are *bona fide* long D7 proteins, sharing a clade with strong bootstrap support (98%) containing *Anopheles* and *Aedes* sequences. This is unusual because to date, black fly and *Culicoides* sequences have not been found in this clade. Co-94700 has the motif C-x(14,17)-W-x(2)-W-x(9,11)-C-x(3)-C-x(52,56)-C found in lipid-binding D7 proteins. Frog thrombocytes produce cysteinyl leukotrienes (Gronert et al., 1995; Wang and Herman, 1997), and these could be targeted by this D7 protein. The motif C-E-x(14,17)-R-x-Y-x(10)-H-x(2)-C-x(55,58)-Y-x(15)-F-D-x(2)-E recognizes mosquito serotonin binding proteins but does not identify any *C. appendiculata* protein from our transcriptome analysis. This does not mean that there are no biogenic amine binding proteins in *C. appendiculata* which may have evolved beyond this constrained pattern. Co-94850, CorSigP-94843, and CorSigP-95009 are classical single-domain OBPs, clustering in a clade containing mosquito and *Drosophila* sequences. Remaining *C. appendiculata* sequences are too divergent to assign phylogeny, but CorSigP-95575 and 96459 appear most related to *Culicoides* sequences. CorSigP-94885—which forms a clade with Co-100513 and CorSigP-94677 (*Corethrella* I clade in Supplemental Fig. S4)—is the most expressed transcript in this sialotranscriptome, mapping over 2 million reads. Co-100513 from the same clade is also highly expressed, mapping over 1 million reads. The classical OBP Cor-SigP-94843 is highly expressed, with an EI = 17. Several additional CDS of the OBP/D7 family are highly expressed with EIs ranging from 1 to 11.

3.6.5.5. Mucins: This heterogeneous group of proteins are serine/threonine rich and have multiple sites where N-acetyl galactosylation may occur (Hang and Bertozzi, 2005; Hegedus et al., 2009; Tian and Ten Hagen, 2009). They are commonly found in insect and tick sialotranscriptomes. Of note, CorSigP-94726, Cor-SigP-95092, and CorSigP-95413 are highly expressed (EI = 83, 11, and 4.5, respectively).

3.6.5.6. Other conserved insect protein families: Eleven CDS are similar to conserved insect families of unknown function, some of which have been found in previous insect sialotranscriptomes. Co-94661 is highly expressed (EI = 16.9) and is similar to deduced salivary proteins from *Anopheles darlingi* (Calvo et al., 2009). PSI-BLAST of the protein

against the NR database reveals an extensive but divergent family of insect proteins that might be evolving at a rapid pace.

3.6.6. Nematocera-specific families

3.6.6.1. 41-kDa family: This protein family has been identified in nearly all nematoceran sialotranscriptomes completed so far, including canonical sequences found in culicines, black flies and *Culicoides*, and divergent members constituting a separate mucin mosquito family (Ribeiro et al., 2010). Their function is unknown, but they are also found expressed in male mosquitoes and may have a function unspecific to bloodfeeding (Ribeiro et al., 2010). Phylogenetic analysis (Supplemental Fig. S5) indicates that CorSigP-95454 forms a clade having strong bootstrap support (98%) with *Simulium* and *Culicoides* sequences that themselves group with a Culicomorpha clade of culicines, albeit with relatively low (57%) bootstrap support. CorSigP-95454 is relatively well expressed, having over 216,000 mapped reads with an EI = 9.7.

3.6.6.2. Culex WRP 16-kDa family: This protein family has been described solely from *Culex* mosquitoes (Ribeiro et al., 2010), and also from *Psorophora albipes* (Campos-Chagas et al, submitted). The protein products have 150–200 aa. It is largely expanded in *C. quinquefasciatus*, with more than 20 genes, mostly uniexonic (Arensburger et al., 2010; Calvo et al., 2010). They present weak Pfam matches to the RicinB_lectin_2 motif. PSI-BLAST of these sequences retrieve lectins and bacterial proteins, suggesting it may have been acquired by horizontal transfer. *C. appendiculata* has several CDS matching these *Culex* proteins (Supplemental File S1) including a large transcript coding for a 700-aa-long protein (Cor- SigP-94987). PSI-BLAST of this larger protein against the NR database using an inclusion value of 1e–6 retrieves, after four iterations, *Wolbachia* proteins at the top of the match list with an e value of 5e–124 and additional mosquito proteins that include several *Culex* members of the 16-kDa family, as expected. Surprisingly, it also retrieves *Aedes* sequences of the salivary 62-kDa family (Supplemental File S2), which was also marked previously as a horizontal transfer candidate gene family. Several other arthropod proteins were also retrieved by PSI-BLAST. Phylogenetic analysis resulting from alignment of CorSigP-94987 with matching *Wolbachia* and mosquito sequences allows the identification of a clade with strong bootstrap support (92%) containing the bacterial and *Culex* members of the 16-kDa family, including two *Aedes* sequences identified by the *Ps. albipes* transcriptome (Campos-Chagas et al, submitted), plus another clade of strong bootstrap support (94%) containing the mosquito 62-kDa family members (Supplemental Fig. S6). The *C. appendiculata* sequence—as a proper “missing link”—sits at the boundary of the two clades. Some *C. appendiculata* members of this family are well expressed, CorSigP-94624, 95021, and 94887 having EIs of 12.2, 6.7, and 1.2, respectively.

3.6.6.3. Other proteins unique to Nematocera: Supplemental File S1 lists several other *C. appendiculata* CDS that belong to known Nematocera families, including the GQ-rich, salivary protein 16, hyp8.2, and 37.7-kDa mosquito families. One member of this last family, Co-95438, is well expressed with an EI = 4.9 and produces best match for a *Wolbachia* protein when compared with the NR database, followed by matches to other

mosquito members of the family, suggesting this family to have originated from horizontal transfer.

3.6.7. Putative *C. appendiculata*-specific salivary proteins—Supplemental File S1 presents 387 CDS encoding putative secreted proteins that bear no significant similarities to known proteins deposited in the NR database. Many of these are relatively small transcripts having low EI that may have been artifactually recovered as CDS. Even so, we identify two novel families, each with at least two genes named Capp 130 kDa and Capp 10-kDa families, both relatively well expressed with members having EIs from 1 to 7. There are additionally 9 CDS without *C. appendiculata* relatives that are well expressed with EIs ranging from 5 to 44, the higher being CorSigP-94762, coding for a 475-aa protein, mapping 974,236 reads.

4. Conclusion

The bloodsucking mode of adult feeding is believed to be ancestral to the infraorder Culicomorpha, the non-bloodfeeding families having lost the primitive trait (Grogan and Szadziowski, 1988). The Culicomorpha is classically divided into two superfamilies, the Culicoidea (containing the Culicidae, Chaoboridae, Corethrellidae, and Dixidae) and the Chironomidea (containing the Thaumaleidae, Simuliidae, Ceratopogonidae, and Chironomidae) (Wood and Borkent, 1989). The proximity of Corethrellidae with Culicidae was also confirmed by 28S RNA phylogenetic analysis (Pawloski et al., 1996). The genus *Corethrella*—the single genus within the Corethrellidae—is found worldwide where there are frogs, the male frog call attracting hungry female flies. Sound-based traps that use recorded “calls” of male frogs are the most effective collecting method of host-questing corethrellids (McKeever and French, 1991; McKeever and Hartberg, 1980). Borkent (2008) proposed that four morphologic features of female corethrellids represent adaptations for host-seeking and bloodfeeding from anurans. Although there are identifications of endothermic vertebrate blood in a few female corethrellids caught in traps, the sharing of aquatic habitats and zoogeographic distributions also supports an ancient relationship between frogs and Corethrellidae (Borkent, 2008). The frog feeding mode is thought to be ancient (certainly before the Gondwanaland breakup >60 MYA), at least at the Early Cretaceous (Borkent, 2008) and before the great mammal radiation. Mosquitoes, on the other hand, are mostly bird or mammal feeders (Clements, 1992). Accordingly, within the Culicoidea superfamily, mosquitoes and frog-biting midges shared a common ancestor before the mammal and bird radiation to which the mosquitoes had to adapt as their hosts evolved their hemostasis systems—in particular, the mammals with the invention of the efficient platelet (Brass, 2005)—while *Corethrella* probably remained with their evolutionarily older amphibian hosts. Although the evolution of salivary genes is at a fast pace (Mans and Francischetti, 2010; Ribeiro et al., 2010) and gene recruitment and loss could have occurred independently by Culicidae and Corethrellidae, several protein families putatively associated to bloodfeeding were found in the sialotranscriptome of *C. appendiculata* that are also found in mosquitoes, including apyrases (although one gene may have been recruited from a different constitutive gene when compared to mosquitoes, see above), trypsin, adenine deaminase, purine hydrolase, DNAase, Hyaluronidase (these last

four only found in Culicines), serpins (Culicines only), cystatin (only found in Aedes so far), D7, antigen 5, 41 kDa mosquito family, 37 kDa mosquito family, among others. Notably missing are members of the Aegyptin/30 kDa family that were identified as one of only 2 shared families between sand flies and Culicomorpha (the other being the 41 kDa family) (Ribeiro et al., 2010). It may be possible that this gene function (collagen inhibitor) in *Corethrella* may have been substituted by a different gene family.

While the biology of vertebrate hemostasis and inflammation is well studied, the equivalent literature for amphibians is quite scanty. Indeed there are 155,994 PubMed articles retrieved with the words “platelet” and “human,” but only 104 with the words “thrombocyte” and “amphibian”. Nonetheless, *Rana* thrombocytes are known to produce thromboxane and cysteinyl-leukotrienes (Gronert et al., 1995; Stiller et al., 1974; Wang and Herman, 1997). Turtle thrombocytes were shown to have serotonin (Maurer-Spurej, 2005), and snake and fish thrombocytes are aggregated by ADP (Gregory and Jagadeeswaran, 2002; Sanomartins et al., 1994), indicating the conservation of the main agonists of platelet aggregation and vasoconstriction between fish, amphibians, reptiles, and mammals. A thrombin/fibrinogen clotting system is also found in amphibians (Ahmad et al., 1979; Jordan, 1983), as is a complement system (Nonaka and Kimura, 2006; Romano et al., 1973). Within this context, the *C. appendiculata* sialome revealed putative ADP antagonism in the form of apyrases of the 5' nucleotidase family and canonical D7 proteins, some of which have a lipid-binding motif. Similar to *Culex* and *Aedes* (Ribeiro and Valenzuela, 2003), adenosine deaminase and purine nucleosidase were found with relatively high expression rates, indicating further purine catabolism. Uniquely, an inositol polyphosphate phosphatase was here identified for the first time in Culicomorpha sialomes, despite being a common finding in triatomines and *Cimex*. Two highly transcribed serpins are candidates for blood-clotting or complement inhibitors. Additionally, a cystatin is highly transcribed, a novel finding in insect or tick sialotranscriptomes. Many other protein families of unknown function were found, some of which are unique to mosquitoes. Among these, *C. appendiculata* has members of the WRP-16-kDa protein family, which was previously thought to be unique to *Culex* (Ribeiro et al., 2010). Protein families unique to *C. appendiculata* were also found, some relatively highly expressed. Despite the apparently primitive frog hemostatic system, the *C. appendiculata* sialome is quite complex and appears well prepared to deal with host hemostasis and to have all the basic components to move on to bird or mammal feeding.

Over 100 species of *Corethrella* are known, females of all but one having mouthparts adapted to bloodsucking (Borkent, 2008). Sialomic analyses of additional species will reveal more about salivary adaptations within this family and possible evolutionary relationships to bloodfeeding among other Culicomorpha.

It should also be noted that this has been, to our knowledge, the first annotated sialome of bloodfeeding Diptera performed with RNAseq technology leading to the identification of over 800 CDS possibly associated with salivary function, about 10× the number found in previous conventional Culicomorpha sialotranscriptomes. Upgrade of these past studies may reveal nearly 10 times more coding sequences, perhaps uncovering potent pharmacologically active agents that are expressed at low levels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA, and by NIH grants R01 AI044793 and R21 AI095780 to L.P.L. We thank B. R. Marshall, DPSS, NIAID, for editing.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Because J.M.C.R., A.C.C., V.M.P., and E.C. 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

- Ahmad N, Dube B, Agarwal GP, Dube RK. Comparative studies of blood coagulation in hibernating and non-hibernating frogs (*Rana tigrina*). *Thromb. Haemostasis*. 1979; 42:959–964. [PubMed: 505409]
- 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; 25:3389–3402. [PubMed: 9254694]
- 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; 208:3971–3986. [PubMed: 16215223]
- Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F, Campbell CL, Campbell KS, Casola C, Castro MT, Chandramouliswaran I, Chapman SB, Christley S, Costas J, Eisenstadt E, Feschotte C, Fraser-Liggett C, Guigo R, Haas B, Hammond M, Hansson BS, Hemingway J, Hill SR, Howarth C, Ignell R, Kennedy RC, Kodira CD, Lobo NF, Mao C, Mayhew G, Michel K, Mori A, Liu N, Naveira H, Nene V, Nguyen N, Pearson MD, Pritham EJ, Puiu D, Qi Y, Ranson H, Ribeiro JM, Roberston HM, Severson DW, Shumway M, Stanke M, Strausberg RL, Sun C, Sutton G, Tu ZJ, Tubio JM, Unger MF, Vanlandingham DL, Vilella AJ, White O, White JR, Wondji CS, Wortman J, Zdobnov EM, Birren B, Christensen BM, Collins FH, Cornel A, Dimopoulos G, Hannick LI, Higgs S, Lanzaro GC, Lawson D, Lee NH, Muskavitch MA, Raikhel AS, Atkinson PW. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science (New York, N.Y.)*. 2010; 330:86–88.
- Assumpcao TC, Ma D, Schwarz A, Reiter K, Santana JM, Andersen JF, Ribeiro JM, Nardone G, Yu LL, Francischetti IM. Salivary antigen-5/CAP family members are Cu²⁺-dependent antioxidant enzymes that scavenge O₂⁻ and inhibit collagen-induced platelet aggregation and neutrophil oxidative burst. *J. Biol. Chem*. 2013; 288:14341–14361. [PubMed: 23564450]
- Bekal S, Domier LL, Niblack TL, Lambert KN. Discovery and initial analysis of novel viral genomes in the soybean cyst nematode. *J. Gen. Virol*. 2011; 92:1870–1879. [PubMed: 21490246]
- Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, Horsman DE, Connors JM, Gascoyne RD, Marra MA, Jones SJ. De novo transcriptome assembly with ABySS. *Bioinformatics (Oxf., Engl.)*. 2009; 25:2872–2877.
- Borkent A. The frog-biting midges of the World (Corethrellidae: Diptera). *Zootaxa*. 2008; 1804:1–456.
- Bours MJ, Dagnelie PC, Giuliani AL, Wesselius A, Di Virgilio F. P2 receptors and extracellular ATP: a novel homeostatic pathway in inflammation. *Front. Biosci. (Schol. Ed.)*. 2011; 3:1443–1456. [PubMed: 21622280]
- Brass LF. Did dinosaurs have megakaryocytes? New ideas about platelets and their progenitors. *J. Clin. Invest*. 2005; 115:3329–3331. [PubMed: 16322776]

- Breugelmans B, Simonet G, van Hoef V, Van Soest S, Vanden Broeck J. Pacifastin-related peptides: structural and functional characteristics of a family of serine peptidase inhibitors. *Peptides*. 2009; 30:622–632. [PubMed: 18775459]
- Calvo E, Mizurini DM, Sa-Nunes A, Ribeiro JM, Andersen JF, Mans BJ, Monteiro RQ, Kotsyfakis M, Francischetti IM. Alboserpin, a factor Xa inhibitor from the mosquito vector of yellow fever, binds heparin and membrane phospholipids and exhibits antithrombotic activity. *J. Biol. Chem.* 2011; 286:27998–28010. [PubMed: 21673107]
- Calvo E, Pham VM, Marinotti O, Andersen JF, Ribeiro JM. The salivary gland transcriptome of the neotropical malaria vector *Anopheles darlingi* reveals accelerated evolution of genes relevant to hematophagy. *BMC Genom.* 2009; 10:57.
- Calvo E, Sanchez-Vargas I, Favreau AJ, Barbian KD, Pham VM, Olson KE, Ribeiro JM. An insight into the sialotranscriptome of the West Nile mosquito vector, *Culex tarsalis*. *BMC Genom.* 2010; 11:51.
- Champagne DE, Smartt CT, Ribeiro JM, James AA. The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 50-nucleotidase family. *Proc. Natl. Acad. Sci. U. S. A.* 1995; 92:694–698. [PubMed: 7846038]
- Clements AN. Development, Nutrition and Reproduction. Vol. 1. London: Chapman & Hall; 1992. Biology of mosquitoes.
- Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Select.* 2004; 17:107–112.
- Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins - roles in reproduction, cancer, and immune defense. *Endocr. Rev.* 2008; 29:865–897. [PubMed: 18824526]
- Gregory M, Jagadeeswaran P. Selective labeling of zebrafish thrombocytes: quantitation of thrombocyte function and detection during development. *Blood Cell. Mol. Dis.* 2002; 28:418–427.
- Grimaldi D, Engel M. Evolution of the Insects. New York: Cambridge University Press; 2005.
- Grogan WL, Szadziowski R. A new biting midge from Upper Cretaceous (Cenomanian) amber of New Jersey (Diptera: Ceratopogonidae). *J. Paleontol.* 1988; 62:808–812.
- Gronert K, Virk SM, Herman CA. Thrombocytes are the predominant source of endogenous sulfidopeptide leukotrienes in the American bullfrog (*Rana catesbeiana*). *Biochim. Biophys. Acta.* 1995; 1259:203–210. [PubMed: 8541326]
- Han Q, Fang J, Ding H, Johnson JK, Christensen BM, Li J. Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome conversion enzymes. *Bio. chem. J.* 2002; 368:333–340.
- Hang HC, Bertozzi CR. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg. Med. Chem.* 2005; 13:5021–5034. [PubMed: 16005634]
- Hegedus D, Erlandson M, Gillott C, Toprak U. New insights into peritrophic matrix synthesis, architecture, and function. *Annu. Rev. Entomol.* 2009; 54:285–302. [PubMed: 19067633]
- Isawa H, Orito Y, Iwanaga S, Jingushi N, Morita A, Chinzei Y, Yuda M. Identification and characterization of a new kallikrein-kinin system inhibitor from the salivary glands of the malaria vector mosquito *Anopheles stephensi*. *Insect Biochem. Mol. Biol.* 2007; 37:466–477. [PubMed: 17456441]
- Johnson JK, Li J, Christensen BM. Cloning and characterization of a dopachrome conversion enzyme from the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* 2001; 31:1125–1135. [PubMed: 11520691]
- Jordan RE. Antithrombin in vertebrate species: conservation of the heparindependent anticoagulant mechanism. *Arch. Biochem. Biophys.* 1983; 227:587–595. [PubMed: 6607710]
- Julenius K, Molgaard A, Gupta R, Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology.* 2005; 15:153–164. [PubMed: 15385431]
- Karim S, Singh P, Ribeiro JM. A deep insight into the sialotranscriptome of the Gulf Coast tick, *Amblyomma maculatum*. *PLoS ONE.* 2011; 6:e28525. [PubMed: 22216098]
- Kashiwada M, Lu P, Rothman PB. PIP3 pathway in regulatory T cells and autoimmunity. *Immunol. Res.* 2007; 39:194–224. [PubMed: 17917066]

- Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. BMC Genom. 2009; 10:33.
- Kuhn JH, Bekal S, Cai Y, Clawson AN, Domier LL, Herrel M, Jahrling PB, Kondo H, Lambert KN, Mihindukulasuriya KA, Nowotny N, Radoshitzky SR, Schneider U, Staeheli P, Suzuki N, Tesh RB, Wang D, Wang LF, Dietzgen RG. Nyamiviridae: proposal for a new family in the order Mononegavirales. Arch. Virol. 2013
- Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 2004; 5:150–163. [PubMed: 15260895]
- Kuroki M, Minakami S. Extracellular ATP triggers superoxide production in human neutrophils. Biochem. Biophys. Res. Commun. 1989; 162:377–380. [PubMed: 2546551]
- Lai R, Takeuchi H, Lomas LO, Jonczy J, Rigden DJ, Rees HH, Turner PC. A new type of antimicrobial protein with multiple histidines from the hard tick, *Amblyomma hebraeum*. FASEB J. 2004; 18:1447–1449. [PubMed: 15247144]
- Loomans HJ, Hahn BL, Li QQ, Phadnis SH, Sohnle PG. Histidine-based zinc-binding sequences and the antimicrobial activity of calprotectin. J. Infect. Dis. 1998; 177:812–814. [PubMed: 9498472]
- Lounibos LP, Makhni S, Alto BW, Kesavaraju B. Surplus killing by predatory larvae of *Corethrella appendiculata*: prepupal timing and site-specific attack on mosquito prey. J. Insect Behav. 2008; 21:47–54. [PubMed: 19081802]
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 2012; 1:18. [PubMed: 23587118]
- Majerus PW, Bansal VS, Lips DL, Ross TA, Mitchell CA, Caldwell KK, Cunningham TW. The phosphatidylinositol pathway of platelets and vascular cells. Ann. N.Y. Acad. Sci. 1991; 614:44–50. [PubMed: 2024892]
- 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; 282:36626–36633. [PubMed: 17928288]
- Mans BJ, Francischetti IMB. Sialomic perspectives on the evolution of blood-feeding behavior in arthropods: future therapeutics by natural design. In: Kini RM, Clemetson KJ, Markland FS, McLane MA, Morita T, editors Toxins and Hemostasis: From Bench to Bed-side. Springer; 2010.
- Maurer-Spurej E. Circulating serotonin in vertebrates. Cell. Mol. Life Sci. 2005
- McKeever S, French FE. *Corethrella* (Diptera, Corethrellidae) of Eastern North-America - laboratory life-history and field responses to Anuran calls. Ann. Entomol. Soc. Am. 1991; 84:493–497.
- McKeever S, Hartberg WK. An effective method for trapping adult female *Corethrella* (Diptera, Chaoboridae). Mosq. News. 1980; 40:111–112.
- McNulty SN, Abubucker S, Simon GM, Mitreva M, McNulty NP, Fischer K, Curtis KC, Brattig NW, Weil GJ, Fischer PU. Transcriptomic and proteomic analyses of a *Wolbachia*-free filarial parasite provide evidence of trans-kingdom horizontal gene transfer. PLoS ONE. 2012; 7:e45777. [PubMed: 23049857]
- Mihindukulasuriya KA, Nguyen NL, Wu G, Huang HV, da Rosa AP, Popov VL, Tesh RB, Wang D. Nyamanini and midway viruses define a novel taxon of RNA viruses in the order Mononegavirales. J. Virol. 2009; 83:5109–5116. [PubMed: 19279111]
- Nielsen H, Brunak S, von Heijne G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng. 1999; 12:3–9. [PubMed: 10065704]
- Nonaka M, Kimura A. Genomic view of the evolution of the complement system. Immunogenetics. 2006; 58:701–713. [PubMed: 16896831]
- O'Flaherty J, Cordes JF. Human neutrophil degranulation responses to nucleotides. Lab. Invest. 1994; 70:816–821. [PubMed: 8015286]
- O'Loughlin AJ, Woffindale CA, Wood MJ. Exosomes and the emerging field of exosome-based gene therapy. Curr. Gene Ther. 2012; 12:262–274. [PubMed: 22856601]
- Pawloski J, Szadziowski R, Kmiecik D, Fahrni J, Bittar G. Phylogeny of the infraorder Culicomorpha (Diptera: Nematocera) based on 28S RNA gene sequences. Syst. Entomol. 1996; 21:167–178.

- 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 Genom.* 2007; 8:300.
- Rasko DA, Myers GS, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinform.* 2005; 6:2.
- Ribeiro JM, Mans BJ, Arca B. An insight into the sialome of blood-feeding Nematocera. *Insect Biochem. Mol. Biol.* 2010; 40:767–784. [PubMed: 20728537]
- Ribeiro JM, Mather TN. *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. *Exp. Parasitol.* 1998; 89:213–221. [PubMed: 9635445]
- Ribeiro JM, Valenzuela JG. The salivary purine nucleosidase of the mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* 2003; 33:13–22. [PubMed: 12459196]
- 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.
- Ribeiro JMC, Assumpcao TCF, Francischetti IMB. An insight into the sialomes of bloodsucking Heteroptera. *Psyche.* 2012; 2012:16.
- Romano EL, Geczy CL, Steiner LA. Reaction of frog antiserum with guinea pig complement. *Immunochemistry.* 1973; 10:655–657. [PubMed: 4586187]
- Rothstein DM, Helmerhorst EJ, Spacciapoli P, Oppenheim FG, Friden P. Histatin-derived peptides: potential agents to treat localised infections. *Expert Opin. Emerg. Drugs.* 2002; 7:47–59. [PubMed: 15989535]
- Sanchez-Gracia A, Vieira FG, Rozas J. Molecular evolution of the major chemosensory gene families in insects. *Heredity (Edinb.).* 2009; 103:208–216. [PubMed: 19436326]
- Sanomartins IS, Jared C, Brunner A. Ultrastructural alterations in thrombocytes of the snake *Waglerophis nerremii* after activation by ADP. *Comp. Haematol. Int.* 1994; 4:226–231.
- Sarkis JJ, Guimaraes JA, Ribeiro JM. Salivary apyrase of *Rhodnius prolixus* Kinetics and purification. *Biochem. J.* 1986; 233:885–891. [PubMed: 3010945]
- Schwarz A, Valdes JJ, Kotsyfakis M. The role of cystatins in tick physiology and blood feeding. *Ticks Tick-borne Dis.* 2012; 3:117–127. [PubMed: 22647711]
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 2009; 19:1117–1123. [PubMed: 19251739]
- Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1998; 6:175–182. [PubMed: 9783223]
- 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; 273:20802–20809. [PubMed: 9694825]
- Stiller RA, Belamarich FA, Shepro D. Frog thrombocytes aggregation and the release reaction. *Thromb. Diath. Haemorrh.* 1974; 32:685–694. [PubMed: 4548774]
- Stouthamer R, Breeuwer JA, Hurst GD. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 1999; 53:71–102. [PubMed: 10547686]
- 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; 25:4876–4882. [PubMed: 9396791]
- Tian E, Ten Hagen KG. Recent insights into the biological roles of mucintype O-glycosylation. *Glycoconj. J.* 2009; 26:325–334. [PubMed: 18695988]
- Tsujimoto H, Kotsyfakis M, Francischetti IM, Eum JH, Strand MR, Champagne DE. Simukunin from the salivary glands of the black fly *Simulium vittatum* inhibits enzymes that regulate clotting and inflammatory responses. *PLoS ONE.* 2012; 7:e29964. [PubMed: 22383955]
- 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; 11:149–155. [PubMed: 11966880]

- Wang X, Herman CA. Eicosanoid synthesis by purified thrombocytes and erythrocytes from warm- and cold-acclimated American bullfrogs (*Rana catesbeiana*). *Gen. Comp. Endocrinol.* 1997; 106:138–144. [PubMed: 9126474]
- Wartha F, Beiter K, Normark S, Henriques-Normark B. Neutrophil extracellular traps: casting the NET over pathogenesis. *Curr. Opin. Microbiol.* 2007; 10:52–56. [PubMed: 17208512]
- Wiesner J, Vilcinskas A. Antimicrobial peptides: the ancient arm of the human immune system. *Virulence.* 2010; 1:440–464. [PubMed: 21178486]
- Wood DM, Borkent A, McAlpine JF, Wood DM. Phylogeny and classification of the Nematocera. *Manual of Nearctic Diptera 3. Research Branch, Agriculture Canada, Monograph.* 1989; 32:1333–1370.
- Woolfit M, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL. An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipientis*. *Mol. Biol. Evol.* 2009; 26:367–374. [PubMed: 18988686]
- Xu X, Oliveira F, Chang BW, Collin N, Gomes R, Teixeira C, Reynoso D, My Pham V, Elnaiem DE, Kamhawi S, Ribeiro JM, Valenzuela JG, Andersen JF. Structure and function of a “yellow” protein from saliva of the sand fly *Lutzomyia longipalpis* that confers protective immunity against *Leishmania major* infection. *J. Biol. Chem.* 2011; 286:32383–32393. [PubMed: 21795673]
- 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. Proteom.* 2008; 7:582–590.
- Zhao L, Chen J, Becnel JJ, Kline DL, Clark GG, Linthicum KJ. Identification and transcription profiling of trypsin in *Aedes taeniorhynchus* (Diptera: Culicidae): developmental regulation, blood feeding, and permethrin exposure. *J. Med. Entomol.* 2011a; 48:546–553. [PubMed: 21661315]
- Zhao QY, Wang Y, Kong YM, Luo D, Li X, Hao P. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinform.* 2011b; 12(Suppl. 14):S2.

Table 1

General classification and abundance of extracted coding sequences (CDS) from the sialotranscriptome of *Corethrella appendiculata*.

Class	No. of CDS	% Total	No. of reads	% Total
Secreted	819	9.158	18,575,254	36.303
Housekeeping	6988	78.139	29,641,996	57.931
Unknown product	580	6.486	1,534,093	2.998
Transposable element	531	5.938	1,064,007	2.079
Viral product	4	0.045	65,909	0.129
Bacterial product	21	0.235	286,420	0.560
Total	8943	100	51,167,679	100

Table 2

Classification and abundance of coding sequences of putative housekeeping function extracted from the sialotranscriptome of *Corethrella appendiculata*.

Class	No. of CDS	No. of reads	% Total
Unknown conserved	1119	5,764,917	19.448
Protein synthesis machinery	399	3,963,087	13.370
Signal transduction	1100	2,825,398	9.532
Storage	13	1,846,207	6.228
Energy metabolism	190	1,788,293	6.033
Transcription machinery	731	1,596,419	5.386
Cytoskeletal protein	331	1,584,853	5.347
Extracellular matrix	236	1,456,986	4.915
Lipid metabolism	250	1,216,368	4.104
Protein export	376	1,177,701	3.973
Protein modification	150	1,075,646	3.629
Transporter and channel	456	962,760	3.248
Nuclear regulation	327	902,957	3.046
Proteasome machinery	279	691,606	2.333
Carbohydrate metabolism	165	534,212	1.802
Protein modification, protease	111	435,401	1.469
Transcription factor	190	422,998	1.427
Amino acid metabolism	105	333,125	1.124
Immunity	72	225,746	0.762
Nucleotide metabolism	102	176,660	0.596
Oxidant metabolism/detoxification	67	156,456	0.528
Intermediary metabolism	55	134,845	0.455
Nuclear export	38	129,302	0.436
Signal transduction, apoptosis	61	126,152	0.426
Detoxification	65	113,901	0.384
Total	6988	29,641,996	100

Table 3

Classification and abundance of coding sequences of putative secretory function extracted from the sialotranscriptome of *Corethrella appendiculata*. Bold text indicates subtotals for each group of coding sequences.

Class	No. of CDS ^a	No. of reads	% Total	No. of CDS with EI > 1% ^b	No. of reads	% Total
<i>Ubiquitous domains</i>						
Enzymes			11.997			11.091
Trypsin-like unique to Corethrella	8	711,082	3.828	7	699,346	4.016
5' nucleotidase/apyrase	6	281,194	1.514	2	264,085	1.517
Amylase and maltase	13	268,877	1.448	2	243,425	1.398
Adenine deaminase	2	216,995	1.168	2	216,995	1.246
Inositol polyphosphate phosphatase	3	135,539	0.730	1	130,558	0.750
Serine protease	64	228,687	1.231	1	119,643	0.687
Purine nucleosidase	4	106,124	0.571	1	104,692	0.601
Terminal peptidases	14	132,744	0.715	1	103,673	0.595
DNAse	5	44,429	0.239	1	25,894	0.149
Hyaluronidase	1	22,997	0.124	1	22,997	0.132
Lipase	9	35,331	0.190			
Esterases	6	18,528	0.100			
Alkaline phosphatase	5	12,496	0.067			
Metalloprotease	4	6582	0.035			
RNAse	3	3915	0.021			
Ceramidases	2	2595	0.014			
Chitinase	1	433	0.002			
Protease inhibitor domains			3.229			2.962
Serpins	15	396,099	2.132	2	347,757	1.997
Cystatin	1	144,865	0.780	1	144,865	0.832
Pacifastin	2	26,903	0.145	1	23,217	0.133
TIL domain	7	24,687	0.133			
Kazal domain	4	6886	0.037			
Metalloproteinase inhibitor	1	334	0.002			
Immunity related			2.965			2.457
Cecropins	3	237,092	1.276	2	237,025	1.361
Peptidoglycan recognition protein	7	139,137	0.749	1	130,831	0.751
Lysozyme	4	37,468	0.202	1	31,008	0.178
Defensins and GGY peptide	15	44,145	0.238	1	29,000	0.167
ML domain containing protein	13	25,495	0.137			
Chitinase-like lectin	3	24,204	0.130			
Gram negative binding protein	2	16,079	0.087			
Thioester proteins	2	14,781	0.080			
C-type lectin	3	4333	0.023			

Class	No. of CDS ^a	No. of reads	% Total	No. of CDS with EI > 1% ^b	No. of reads	% Total
Phenoloxidase inhibitor protein	2	3978	0.021			
Galectin	3	2595	0.014			
Attacin	4	1484	0.008			
Other ubiquitous families			6.193			5.980
Antigen 5 protein	9	719,488	3.873	2	697,129	4.003
Lipocalins	3	4568	0.025			
Hypothetical secreted conserved protein families			0.000			
Histidine rich conserved Diptera families	4	344,128	1.853	4	344,128	1.976
CysHis rich protein family	2	1387	0.007			
Cys-rich secreted peptide	2	1855	0.010			
Other hypothetical conserved protein	55	78,996	0.425			
Insect conserved families			44.342			46.680
OBP/D7 family	35	4,873,688	26.238	16	4,844,619	27.821
Mucin related proteins	22	2,261,503	12.175	4	2,212,379	12.705
Simulium/Culicoides insect conserved secreted protein family	10	667,781	3.595	6	653,216	3.751
Insect allergen repeat	3	35,299	0.190	1	27,730	0.159
Yellow protein family	2	7785	0.042			
Other conserved Insect families	11	390,511	2.102	11	390,511	2.243
Nematocera specific salivary domains			4.803			4.862
Culex WRP/16 kDa family	14	456,712	2.459	3	447,620	2.571
41 kDa family	1	216,812	1.167	1	216,812	1.245
37.7 kDa mosquito family	3	117,876	0.635	1	108,574	0.624
Salivary protein 16 family	2	77,846	0.419	1	73,686	0.423
GQ rich salivary secreted protein	3	9414	0.051			
hyp8.2 culicine family	2	519	0.003			
Other proteins unique of Nematocera	8	12,989	0.070			
Putative novel secreted proteins			26.471			25.968
Capp13 kDa family	2	313,121	1.686	2	313,121	1.798
Capp10 kDa	2	63,391	0.341	1	41,415	0.238
Other proteins unique of <i>Corethrella</i>	383	4,540,472	24.444	37	4,167,371	23.932
Total	819	18,575,254		118	17,413,322	94

^aCDS, coding sequence.

^bEI, the expression index, is defined as 100 times the ratio of the number of reads mapped to a coding sequence divided by the number of reads observed for the most expressed transcript.