Functional annotation and analysis of the *Ornithodoros moubata* midgut genes differentially expressed after blood feeding.

Ana Oleaga*¹, Prosper Obolo-Mvoulouga¹, Raúl Manzano-Román¹, Ricardo Pérez-Sánchez¹

¹ Parasitology Laboratory, Institute of Natural Resources and Agrobiology (IRNASA, CSIC). Cordel de Merinas, 40-52, 37008 Salamanca, Spain.

Corresponding author:

Dr. Ana Oleaga

Parasitology Laboratory

Instituto de Recursos Naturales y Agrobiología (IRNASA, CSIC)

Cordel de Merinas, 40-52

37008 Salamanca, Spain

Tel.: +34 923219606; fax: +34 923219609

E-mail address: ana.oleaga@irnasa.csic.es

Abstract

The argasid tick *Ornithodoros moubata* is the main vector of the African swine fever and the human relapsing fever in Africa. As part of the host-parasite-pathogen interface, the tick midgut expresses key proteins for tick survival and tick-borne pathogen transmission. Accordingly, midgut proteins are potential targets for the development of new drugs and vaccines aimed at tick control, and obtaining proteomic and transcriptomic data from the *O. moubata* midgut would facilitate the identification of such target candidates. With this aim, we have assembled and characterized the midgut transcriptome of *O. moubata* females before and 48 hours after a blood meal, and identified the genes that are differentially expressed in the midgut after feeding. Overall, 23,863 transcripts were obtained, and of them, 9,164 were identified and annotated. The most represented molecular functions were catalytic and binding activities, and the most represented biological processes were metabolic, cellular and single-organism processes. KEGG analysis of the annotated sequences assigned up to 3,053 of them to 130 active pathways, among which, the top 30 pathways were mostly metabolic routes. Differential expression analysis between unfed and fed ticks detected 8,026 Differentially Expressed Genes (DEGs), 4,093 up-regulated and 3,933 down-regulated, respectively. The biological significance of these DEGs was further investigated using the KEGG, Pfam and GO databases. The functional groups of the genes/proteins predicted to be involved in the processes of blood digestion, nutrient transport and metabolism, and in responses related to defence and oxidative stress are discussed in more detail. This work reports the first midgut transcriptome analysis of an argasid tick species, and provides a wealth of novel molecular information about the argasid machinery involved in blood feeding and digestion. This information represents a starting point for the development of alternative strategies for tick control.

119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177

45

46

47

48

49

Keywords

Ornithodoros moubata; Soft tick; Midgut; Transcriptome; Blood digestion; Differential expression.

1. Introduction

Soft (Argasidae) and hard ticks (Ixodidae) are two families of hematophagous arthropods that transmit a wide range of pathogenic microorganisms which cause severe diseases to livestock, pets and people (de la Fuente et al., 2008; Manzano-Román et al., 2012a).

Even though haematophagy is a common feature to soft and hard ticks, both tick families exhibit important morphological and biological differences. Most ixodids are non-nidicolous organisms (exophiles), which fed during several days and ingest amounts of blood higher than 100 times their own body weight. Ixodid females fed only once and die after laying several thousands of eggs (Apanaskevich and Olivier, 2014). Unlike ixodids, argasids are nidicolous ticks, which typically complete their blood meal within minutes or hours, expanding to about 5 to 10 times their pre-fed body weight. Adult argasids are intermittent feeders since they can feed and reproduce up to 10 times, laying several hundreds of eggs *per* trophogonic cycle. Moreover, argasids are notably resistant to starvation as they can survive for several years without feeding (Oleaga-Pérez et al., 1990; Vial, 2009).

In Africa, the argasid tick *Ornithodoros moubata* is the main vector of the African swine fever virus (Sánchez-Vizcaíno et al., 2015) and the East African human relapsing fever spirochete, *Borrelia duttonii* (Cutler, 2010). In the countries where these diseases are endemic, the *O. moubata* populations inhabiting anthropic habitats contribute to the persistence of both diseases and, potentially, to their spreading to other countries (Cutler, 2010; Sánchez-Vizcaíno et al., 2015; Quembo et al., 2016). Thus, to be effective, any program aimed at the prevention and control of these diseases shall

require the elimination of, at least, the synanthropic tick populations (Manzano-Román et al., 2012a).

Chemical acaricides have traditionally been used as first election method for tick control, but their use has raised a range of concerns including selection of tick-resistant strains, environmental pollution and contamination of animal products. Instead, anti-tick vaccines have emerged as sustainable and cost-effective methods for tick control, and their inclusion in integrated control programs to ticks and tick-borne diseases would allow reducing acaricide use (de la Fuente, 2012; Guerrero et al., 2012).

Rational searching of argasid and ixodid tick organs, physiological processes and involved molecules to be targeted by anti-tick vaccines has driven the tick midgut and the midgut physiological processes to be considered one of the most suitable sources of vaccine candidate antigens in both the soft (Manzano-Roman et al., 2006, 2007; Díaz-Martín et al., 2015) and the hard ticks (Guerrero et al., 2012; Maritz-Olivier et al. 2012; de la Fuente et al., 2016). The tick midgut is where the blood meal is digested, and it constitutes the first defensive barrier faced by the tick-borne pathogens (Kocan et al., 2004). This barrier must be overcome by the pathogens to effectively infect the tick and be subsequently transmitted to the next tick stage, the tick progeny or a new vertebrate host. Thus the tick midgut is an important part of the pathogen-tick-host interface, and there proteins that play vital functions for ticks and tick invasion by pathogens are expressed. Hence, these intestinal proteins would be potential antigen targets for anti-tick vaccines and pathogen transmission-blocking vaccines. In fact, the only two anti-tick vaccines marketed up to date (TickGARD[®] and GAVAC[®]), both developed against the ixodid tick *Rhipicephalus microplus*, are based on the antigen Bm86, which is a glycoprotein expressed on the luminal membrane of midgut epithelial cells (Willadsen, 2008).

Regarding the ornithodoros ticks, our previous results indicated that the midgut membranes of *O. moubata* and *Ornithodoros erraticus* also contain protective antigens, although they are hitherto unidentified (García-Varas, 2004; Manzano-Román et al., 2006, 2007; Díaz-Martín et al., 2015). These results proved that the argasid midgut is also a suitable source of vaccine candidate antigens.

The tick midgut consists of a monolayer of midgut cells or enterocytes, which digest the blood intracellularly, inside lysosomal vesicles, at acidic pH (Sojka et al., 2013). In ixodid ticks, and particularly in *Ixodes* spp., the molecular physiology of blood digestion and the histological changes of the midgut epithelium throughout the blood feeding and digestion processes have been deeply studied and, as a result, they are quite well established (Akov, 1982; Coons et al., 1986; Mans and Neitz, 2004; Caperucci et al., 2010; Sojka et al., 2013). The recent obtaining and characterization of the midgut transcriptomes and proteomes from several ixodid species have greatly helped in the understanding of these processes at a molecular level (Anderson et al., 2008; Rachinsky et al., 2008; Kongsuwan et al., 2010; Heekin et al., 2013; Schwarz et al., 2014; Kotsyfakis et al., 2015).

By contrast, in argasid ticks, the physiology and biochemistry of blood digestion have been little studied and the scant information available has been drawn from a handful studies on *O. moubata* (Granjean, 1983, 1984; Grunclová et al., 2006). More recently, the description of the *O. erraticus* midgut proteome in fed and unfed females has contributed to expand our knowledge on this topic. However, none argasid midgut transcriptome has been published to date (Oleaga et al., 2015).

With this in mind, and as pointed out by Kongsuwan et al. (2010), the investigation of the *O. moubata* midgut transcriptome might provide a comprehensive understanding of the essential cellular processes of the digestive physiology of argasids,

affording valuable information about potential targets for drug and vaccine design. Thus, the aims of this work were (i) the characterization of the midgut transcriptome of *O. moubata* in females before and after a blood meal, and (ii) the identification and analysis of the genes that are differentially expressed in midgut after feeding.

To achieve this goal, midgut tissues from unfed and engorged *O. moubata* females taken at 48 hours post-feeding were dissected and the polyA⁺ RNAs were purified and sequenced by RNA-seq. In total, in both fed and unfed ticks, 23,863 transcripts were assembled and 9,164 of them were functionally annotated and classified. Additionally, 4,093 up-regulated and 3,933 down-regulated genes were detected in the midgut of fed ticks as compared to unfed ticks. These genes were further classified into functional groups and their biological significance was analyzed and commented, especially those predicted to participate in the process of blood digestion and other processes related to blood feeding, such as the defense response and the response to oxidative stress produced during blood digestion.

2. Material and Methods

2.1. Ticks and sample collection

The *O. moubata* specimens used came from our laboratory colony, where they were regularly fed on rabbits and kept in an environmental chamber at 28°C, 85% relative humidity, and a 12 hours light-dark cycle. Tick feeding and animal manipulation were done according to the regulations established by the Ethical and Animal Welfare Committee of the institution where the experiments were conducted (IRNASA, CSIC), following the corresponding EU legislation.

Tick midguts were obtained from unfed females (U samples) and from 48 hours post-feeding engorged females (F samples). The mean weight of unfed *O. moubata* females and that of fed females at 48 hours post-feeding were 63.5 ± 3.1 mg and 253.5 ± 53.2 mg, respectively. So that, the mean increase in weight after feeding was 189.9 ± 23.7 mg. The ticks were dissected in sterile phosphate buffered saline (PBS), pH 7.4, at 4 °C, and the midguts were removed, rinsed several times in PBS to eliminate the host's blood and examined to remove all contaminant tissues (mostly Malpighian tubules and tracheae) (Manzano-Román et al., 2006). The midgut tissues were collected in batches of 20 specimens each, and immediately preserved in RNAlater until RNA extraction.

Two biological replicates from each physiological condition (U1/U2 and F1/F2) were obtained and processed in order to obtain the corresponding transcriptomes.

2.2. RNA extraction and library construction for Illumina sequencing

All of the midgut samples (U1, U2, F1 and F2) were equally processed. They were lysed and homogenized using the TissueLyser II (Qiagen), and total RNA extraction was performed with the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. RNeasy Protect® (Qiagen) was used to ensure the integrity of RNA during isolation and to stop transcription at given time points. RNA was treated with Turbo DNase-free (Ambion) to remove possible traces of contaminant DNA. RNA quality was assessed by measuring RNA integrity number (RIN) in the Agilent 2100 bioanalyzer, resulting in an average RIN of 9.4 (8.8–9.8). Then, polyA+ RNA was purified using NEBNext® Poly(A) mRNA Magnetic Isolation Module (Biolabs, Inc.), and RNA libraries were prepared according to standard procedures using the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (Biolabs, Inc.). Libraries were pooled in equimolar amounts and sequenced using a Genome Analyzer

173 Iix sequencer (Illumina). Sequences were double read 125 bases long. Quality filtering
 174 was automatically performed according to Illumina specifications, and, on average, 30
 175 million reads per sample were obtained.

177 **2.3. Data analysis: bioinformatics processing and assembly**

178 Before assembly, artefacts and paired-end reads were filtered for adapter
 179 sequences using the FastQC tool
 180 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed using the
 181 Trimalore software (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
 182 to improve read quality, accuracy and the computational efficiency of the assembly;
 183 suboptimal read quality nucleotides were assessed using a PHRED quality score above
 184 30. To reduce apparent redundancy, sequences >300 nucleotides were clustered using
 185 the Oases contig assembly program (Schulz et al., 2012) and the Trinity package,
 186 (Grabherr et al., 2011) employing a minimum overlap length of 25 nucleotides.

187 The raw sequence reads, derived from each library, were mapped to the non-
 188 redundant transcriptomic data using the mapping program Bowtie2 (Langmead and
 189 Salzberg, 2012). To quantify the transcriptional expression, the relative abundance of
 190 each transcript was analysed and normalized per length using the RSEM software (Li
 191 and Dewey, 2011), which is based on the number of reads of each gene in the four
 192 libraries with the RPKM method (reads per kilobase per million reads, RPKM)
 193 (Mortazavi et al., 2008).

195 **2.4. Transcriptome characterization: gene identification and function assignment.**

196 The non-redundant datasets for *O. moubata* were annotated based on BLASTx
 197 and BLASTn for sequence homology using a permissive search strategy (E -value < 10^{-5})

⁰⁵) against sequences in the available databases: the NCBI non-redundant sequence database (<http://www.ncbi.nlm.nih.gov/>), Swiss-Prot database (<http://www.uniprot.org/>) and the non-redundant genome-wide sequence database for *Ixodes scapularis* (http://metazoa.ensembl.org/Ixodes_scapularis/Info/Index).

Proteins were translated *in silico* from the predicted coding domains of individual nucleotide sequences using the Transdecoder software and all the ORFs being longer than 150 nucleotides were selected (<https://transdecoder.github.io/>). Protein-coding sequences were functionally classified using the HMMER hmmscan tool against the Pfam protein domain database (Finn et al., 2011, 2016), employing the default search parameters. Gene ontology (GO) categories and protein classes were assigned using the Uniprot tools (<http://www.uniprot.org/>) and the Gene Ontology database (<http://www.geneontology.org/>). A Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis was performed using the Blast2GO software (<https://www.blast2go.com/>).

2.5. Differential expression analysis

The differentially expressed genes (DEGs) were screened by comparing the transcript expression levels of the two Illumina libraries made from fed ticks (F1, F2) *versus* the two Illumina libraries from unfed ticks (U1, U2). In this study, differential expression analysis between these two conditions, unfed and fed, was performed using the DEseq algorithm (Love et al., 2014), considering those genes with a fold change of > 2 or < -2 as being differentially expressed. A false discovery rate (FDR) correction, using the Bonferroni method, was employed to determine statistical significance taking an adjusted *p*-value (*p*Adj) < 0.05 as limit.

2.6. Experimental validation

In order to evaluate the accuracy of the assembled transcripts, 6 genes (named Om99, Om03, Om17, Om28, Om29 and Om86) were selected to be amplified, cloned and sequenced on both strands. Om86 was chosen because it is the orthologue of the vaccine antigen Bm86 (Willadsen, 2008) in *O. moubata*. The other 5 genes were randomly chosen among the overexpressed genes that showed a complete open reading frame that code for a peptide of at least 75 amino acids long.

Hence, the first-strand cDNAs were synthesized from the same RNA samples which were used for RNA-Seq by using the first Strand cDNA Synthesis kit (Roche) according to the manufacturer's instruction. The primer sequences designed *ad hoc* from the Illumina sequences and the reaction conditions used for the amplification of cDNAs are shown in the Supplementary Table 1.

The PCR products were purified from agarose gels using the StrataPrep DNA Gel Extraction kit (Stratagene) and cloned into the pSC-A sequencing vector (Stratagene). The sequencing was performed on both strands at the DNA Sequencing Service of the University of Salamanca (Spain) using the T3 and T7 primers. The sequences obtained from the pSC-A clones were compared to the sequences of the corresponding open reading frames that were predicted for each transcript/contig by the ORF finder tool (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Sequence comparisons were performed with Multalin software (Corpet, 1988).

Additionally, to validate the expression levels given by the RNA-seq analysis we performed a quantitative real-time PCR (qRT-PCR) analysis on the above six selected genes. Primers were designed using the Primer3Plus software (Untergasser et al., 2007) selecting a product size range of 100-150 nucleotides and a T_m close to 60 °C. The qRT-PCR reactions were performed in triplicate in 96-well plates with a sequence

detector system (ABI PRISM 7900HT, Applied Biosystems), using Power SYBR Green PCR Master Mix (Applied Biosystems), cDNA and gene-specific primers in a total volume of 20 µl. The primer sequences and the qRT-PCR conditions are indicated in the Supplementary Table 1. The *O. moubata* actin gene (GenBank: AB208021.1) was used as the reference for normalization purposes in order to make a comparison of mRNA levels between the fed and unfed samples. For this, the mRNA levels of the selected genes were normalized against actin using the comparative Ct method (Schmittgen and Livak, 2008). The fold change in the expression level was calculated by dividing the normalized mRNA level in fed samples by the normalized mRNA level in the unfed samples. The relative fold change in transcript levels obtained by qRT-PCR was compared to that provided by the RNA-seq analysis.

3. Results and discussion

3.1. Illumina sequencing and *de novo* assembly of the midgut transcriptome

To analyse and compare the midgut transcriptomes of unfed and fed *O. moubata* female ticks, RNA was extracted from two replicated biological samples of midgut tissue from each physiological condition and the purified PolyA⁺ RNAs were sequenced on the Illumina HiSeq system.

A total of 71,214,831 (71.2 Mb) and 61,219,673 (61.2 Mb) of 125-bp paired-ends reads were obtained from the U1 and U2 unfed tick midguts, respectively, and 65,232,011 (65.2 Mb) and 52,668,669 (52.7 Mb) 125-bp paired-ends reads were obtained from the F1 and F2 fed tick midguts, respectively.

After trimming the adapters and removing all low-quality and ambiguous reads, 65,896,668 (U1), 57,304,776 (U2), 61,849,703 (F1) and 49,650,121 (F2) high-quality, clean reads were obtained, which accounted for 92.5% - 94.8 % of total raw reads.

After *de novo* assembly, 23,863 transcripts were obtained showing a N50 length of 1,475 bp, an average length of 1,989 nucleotides, and the longest sequence of 28,505 bp.

The coding sequences were deposited in DDBJ/ENA/GenBank under accession number GFJQ000000000 as a Transcriptome Shotgun Assembly project (BioProject: PRJNA377416). The version described in this paper is the first version, GFJQ01000000.

3.2. Functional annotation and classification

For functional annotation of the assembled transcripts, the sequences were launched against the NCBItr, Swiss-Prot and *Ixodes scapularis* genome databases.

In total, 9,164 transcripts were annotated, accounting for 38.4% of all the assembled sequences. The remaining 14,699 transcripts (61.6%) were classified as unknown because they did not produce matches within any of the searched databases, or produced non-significant or irrelevant matches, indicating that many new genes and non-coding RNA sequences had been identified. This was expected considering that *O. moubata* is a non-model species and its genome is essentially unknown.

Afterwards, all of the annotated sequences from the tick midgut were classified according to their Molecular Function and Biological Process using the UniProt tools and the Gene Ontology (GO) database.

With respect to the classification of the genes based on Molecular Function, we only considered those categories that were represented by more than 30 genes, which were the following: catalytic, binding, structural, transporter, antioxidant, electron carrier, molecular transducer, regulator, and transcription factor activity. As can be observed in Fig. 1A, the proteins with catalytic activity (n=2,936, 32%), binding activity

(n=2,882, 31%), and unknown function (n=2,215, 24%) were the most represented. Similarly, the midgut transcriptome was classified into 13 different biological processes, which were again represented by at least 30 genes for each category (Fig. 1B). Of these 13 categories, the proteins involved in metabolic processes (n=3,584, 23%), cellular processes (n=3,383, 21%) and single-organism processes (n=2,431, 34%) were the most abundant. The remaining categories were less numerous, except those classified as having an unknown Biological Process, which represented up to 14% of the annotated transcriptome. The available GO data regarding Molecular Function, Biological Process and Cellular Component for all of the annotated proteins are included in the Supplementary Table 2.

To detect active biological pathways in the midgut of *O. moubata*, the 9,164 annotated genes were analysed in the KEEG pathway database. Overall, we assigned 3,053 sequences to 130 biological pathways that clustered into 16 different “Classes” (Supplementary Table 3).

Table 1 shows the top 30 pathways in the transcriptome of *O. moubata*, with purine metabolism (428 sequences), thiamine metabolism (302 sequences), biosynthesis of antibiotics (167 sequences) and aminobenzoate degradation (123 sequences) being the most represented. Also, these top 30 pathways mostly included metabolic routes (26, 86.6%) involved in the metabolism of amino acids, carbohydrates, lipids, xenobiotics, nucleotides, cofactors and vitamins, and energy metabolism. The purine metabolism pathway was represented by the highest number of enzymes (428). As suggested by Stutzer et al. (2009), a number of these nucleotide-metabolising enzymes may be modulating the extracellular levels of nucleotides, such as ATP and ADP, which would prevent platelet activation and aggregation, thus helping to maintain the fluidity of the ingested blood.

3.3. Differential expression analysis of the midgut transcriptomes from fed and unfed ticks.

In order to compare the differentially expressed genes (DEGs) between both conditions (fed *versus* unfed), gene expression levels were first assessed using the RPKM method. According to the RPKM values, 8,026 significantly DEGs were detected in the midgut of fed ticks as compared to the basal condition (unfed ticks), specifically, 4,093 up-regulated genes and 3,933 down-regulated genes. Fig. 2 shows the fold change in expression in fed *versus* unfed (y-axis) against expression level (x-axis). Each data point represents a transcript, and those showing significant differential expression (fold change > 2 or < -2, adjusted *p* value < 0.05) are coloured in red. The fold change of up- and down- regulated DEGs are summarised in Fig. 3. Most of the DEGs (n=4,817, 60.0%) showed a 2 to 5 fold change in expression level between fed and unfed ticks, while 3,010 DEGs (37.5%) showed a fold change between 5 and 50, and 199 DEGs (2.5%) showed a fold change higher than 50. This fold change distribution is in concordance with that reported from a similar differential expression analysis of the midgut transcriptome of the hard tick *Haemaphysalis flava* (Xu et al., 2016).

To investigate in more detail the biological significance of the DEGs identified in the transcriptome of fed ticks, we analysed them using the KEGG pathways, Pfam protein domain and GO (Molecular Function and Biological Process) databases. The analysis of these genes and the metabolic routes in which they are involved indicated that several metabolic processes and cellular functions were differentially regulated (Table 1).

When considering the top 30 biological pathways represented in the midgut transcriptome, we detected that the numbers of up-regulated genes involved in amino acid metabolism (excluding tryptophan metabolism), energy metabolism, thiamine metabolism, nucleotide metabolism and amonoacyl-tRNA biosynthesis (translation) were higher than the number of down regulated genes in these same pathways (Table 1). In contrast, the number of genes down-regulated related to carbohydrate metabolism (excluding the amino sugar and nucleotide sugar metabolism) and signal transduction pathways was higher than the number of up-regulated genes. Finally, for the lipid metabolism, T cell receptor signalling and xenobiotics metabolism pathways, the number of up- and down-regulated DEGs was similar (Table 1).

The search using the Pfam domain database identified 5,528 and 4,867 protein families for the up- and down-regulated DEGs, respectively (Supplementary Table 4 and Supplementary Figure 1). Table 2 summarizes the top 30 Pfam domains detected in the significantly DEGs.

Among the up-regulated DEGs, the protein domains most represented were the “RNA recognition motifs RRM_1, RRM_6 and RRM_5”, the transmembrane “ABC transporter”, the “WD40 domain”, the membrane transporters “Major Facilitator Superfamily (MFS)” and the “50S ribosome-binding GTPase motif” (Table 2). The first three domains belong to large protein families found in all eukaryotes cells, which are involved in several processes of cell biology such as signal transduction and transcription, regulation to cell cycle, apoptosis or protein synthesis (Maris et al., 2005; Smith et al., 2008; Shields et al., 2009).

The ABC and MFS transporters are two superfamilies involved in the uptake or export of a wide variety of substrates across membranes. In this context, these

transporters might facilitate the fed female to acquire the necessary nutrients from the ingested blood (Saier et al., 1999; Wang et al., 2009; Stutzer et al., 2013).

The Pfam protein domains more abundant in down-regulated DEGs were the “Protein kinase domains”, “Reverse transcriptase domain” and up to 9 different “Zinc finger motifs (Znf)” (Table 2, Supplementary Table 4 and Supplementary Figure 1). This last protein superfamily has been linked to several biological processes (Laity et al., 2001). Recently, it has been demonstrated that a tumour necrosis factor receptor-associated factor-type zinc finger domain containing protein 1 (TRAFD1) from the hard tick *Haemaphysalis longicornis* strongly affects both the tick humoral and cellular immunity against bacterial infection suggesting its participation in defence responses against pathogens ingested with the host blood (Takechi et al., 2016).

The annotated up- and down-regulated DEGs were analysed separately and categorized according to their “Molecular Function” and “Biological Process” using the tools available at the UniProt website (Fig. 4). This analysis revealed that DEGs coding for proteins involved in catalytic and binding activities were the most numerous, which is in agreement with their role in the intracellular processing of the blood meal (Anderson et al., 2008). Among the remaining categories, all of them with significant lower proportions, the most numerous corresponded to genes encoding proteins with transporter activity (Fig. 4A). A remarkably high number of structural proteins were detected among the up-regulated DEGs (98 up-regulated versus 26 down-regulated), of which 83% were structural constituents of ribosomes (e.g. ribosomal RNA 40S, 60S, and other ribosomal genes), suggesting that the metabolic and cellular processes involved in protein synthesis were the most active in fed ticks. Additionally, it is known that these ribosomal genes could also be associated with stress responses triggered by the blood digestion process (Anderson et al., 2008; Villar et al., 2014).

Regarding the classification based on Biological Process, the more numerous up- and down-regulated DEGs after blood feeding were those involved in cellular, metabolic and single-organism processes (Fig. 4B). The DEGs included among the remaining categories, all with lower ratios, were those proteins involved in biological regulation, localization, response to stimulus and cellular component organization and biogenesis. The higher number of up-regulated *versus* down-regulated genes was detected in the category of localization process (GO: 0051179) (Figure 4B).

3.4. Genes involved in feeding, blood meal digestion and defence.

After having characterized and analysed the midgut transcriptomes from fed and unfed ticks, we carried out a more detailed analysis of particular groups of proteins/genes that are involved in the process of blood digestion and other processes related to blood feeding.

Thus, considering the GO classification and the Pfam database annotation we selected representative genes involved in the following functional groups: (i) genes involved in the metabolism of nutrients, such as protein digestion and carbohydrate and lipid metabolism and transport (Table 3), and (ii) genes involved in defence response and response to oxidative stress (Table 4).

3.4.1. Nutrient transport and metabolism.

The uptake and metabolism of nutrients ingested with the blood meal implicate a range of genes involved in the processing and transport of proteins, carbohydrates and lipids (Stutzer et al., 2013) (Table 3, Supplementary Table 5).

In ticks, the intracellular digestion of nutrients from the blood meal is mainly directed at proteins, mainly haemoglobin. In ixodid ticks, the dynamics of the

haemoglobinolytic system during blood digestion is known in considerable detail. It is initiated by endopeptidases of the aspartic and cysteine classes (cathepsin D, cathepsin L and legumain) and is continued by cathepsins C and B, by a carboxypeptidase and by a leucine aminopeptidase (Horn et al., 2009; Sojka et al., 2013). In argasid ticks, the information about the machinery of blood digestion has been limited to the description of protease activity in the midgut of *Ornithodoros tholozani* (Akov et al., 1976), and the identification of two cystatins in *O. moubata* (Grunclová et al., 2006). More recently, the description of the midgut proteome of *O. erraticus* has allowed the identification of several proteins likely involved in this process (Oleaga et al., 2015).

In the current work, numerous proteases have been identified in the midgut transcriptome of *O. moubata*. Many of them are homologues of the digestive enzymes described in ixodids and belong to the following families: Aspartic peptidases (8 transcripts), cysteine peptidases (22 transcripts), carboxypeptidases (15 transcripts), metallopeptidases (37 transcripts) and serine peptidases (33 transcripts). Other non-specific peptidases have been not included in this selection (Table 3, Supplementary Table 5).

It has been suggested that most of the proteins involved in blood digestion are up-regulated after feeding (Anderson et al., 2008). Our current results support this notion, and also parallel the dynamics of the digestive proteolytic system of ixodids (Franta et al., 2010; Sojka et al., 2013). Actually, we found up-regulated genes in the transcriptome of *O. moubata* fed ticks that code for several aspartic proteases (a aspartic protease, cathepsin D, cathepsin E and several uncharacterized proteins), cysteine proteases (cathepsin B, cathepsin L and a cysteine protease) and several carboxypeptidases. Regarding the metalloproteases, we identified up to 5 up-regulated DEGs and 2 additional transcripts, not differentially expressed, that code for a leucine

aminopeptidase. Several proteins with serine protease activity, mostly uncharacterised, were up-regulated after feeding, suggesting that these proteins could participate in the haemoglobin digestion (Miyoshi et al., 2008). (Table 3).

The studies regarding carbohydrases in ticks are very limited and the enzymes involved in the soft tick carbohydrate digestion are unknown. The only data on this topic is the description of a digestive α -L-fucosidase activity in *Amblyomma cajennense* (Moreti et al., 2013), and the identification of several putative carbohydrate-metabolizing enzymes in the midgut transcriptomes of *Ixodes ricinus* (Kotsyfakis et al., 2015), *Rhipicephalus microplus* (Stutzer et al., 2013), *Haemaphysalis flava* (Xu et al., 2016) and *Dermacentor variabilis* (Anderson et al., 2008). In the *O. moubata* midgut transcriptome analysed in this study, several carbohydrases, carbohydrate transporters and enzymes with chitinase activity have been identified (Table 3, Supplementary Table 5). Regarding putative carbohydrate-metabolizing enzymes, up to 111 transcripts were identified (37 up-regulated, 36 down-regulated and 38 not differentially expressed), playing diverse enzymatic functions such as kinase, hydrolase, transferase, isomerase, lyase and oxidoreductase activity. Several digestive enzymes (such as α -amylase, α -galactosidase, α -mannosidase, β -galactosidase, β -glucuronidase, β -hexosaminidase, glucosidase and glucosylceramidase) are involved in the hydrolysis of O-glycosyl compounds and removal of glycan groups from carbohydrates, glycoproteins and glycolipids. As suggested by Moreti et al. (2013), these enzymes might play a role in the digestion of these compounds, and also in the defence against pathogens by removing glycan groups that could participate in tick-pathogen interactions. This is the case of infections caused by *Anaplasma phagocytophilum*, in which the bacterium colonizes its tick host through interactions with α -1,3-fucose (Pedra et al., 2010).

The carbohydrate transmembrane transporters identified, two glucose transporters and other sugar transporters, are up-regulated in the midgut of fed ticks (Table 3). The up-regulated expression of genes in the tick midgut associated with carbohydrate metabolism and their transport support the idea that carbohydrate digestion may be important at the midgut in the blood digestion process.

Eight enzymes putatively involved in the chitin metabolism were identified. They are predicted to have chitinase activity, chitin binding activity, and some of them contained the chitin binding peritrophic-A domain (Pfam code: Pf1607.19:CBM-14), which may be involved in peritrophic matrix formation and maintenance (Kotsyfakis et al., 2015) (Table 3, Supplementary Table 2). The presence of a peritrophic membrane in ticks has been described in *Ixodes scapularis* (Grigor'eva and Amosova, 2004), *Ixodes ricinus* (Zhu et al., 1991), *Haemaphysalis longicornis* (Matsuo et al., 2003) and *Ornithodoros moubata* (Granjean, 1983). It is believed that this structure provides protection to the microvilli of epithelial cells of the digestive tract against injurious particulates and ingested microbes (Lehane, 1997). In ixodids, the peritrophic membrane forms within 9 – 12 hours after the initiation of feeding (Grigor'eva and Amosova, 2004). The finding that 12 out of the 19 identified transcripts, potentially involved in the peritrophic membrane formation, are highly up-regulated in the *O. moubata* midgut (Table 3) suggests that in argasids, or at least in *O. moubata*, the peritrophic membrane formation is still active at 48 hours post-feeding.

The mechanism of lipid digestion in ticks is unclear. In hard ticks, a number of overexpressed midgut genes involved in lipid transport and metabolism have been identified, suggesting an important role of lipids as a nutrient source (Stutzer et al., 2013; Kotsyfakis et al., 2015; van Zyl et al., 2015). In *O. moubata*, approximately 112 midgut transcripts were identified with a putative function in lipid metabolism and

transport, most of them not differentially expressed (56 transcripts) or down-regulated (29 transcripts) after blood feeding (Table 3, Supplementary Table 5). The putative lipid metabolic enzymes that have been identified have several molecular functions including hydrolase, oxidoreductase, ligase, isomerase, transferase, lipid binding and phospholipase activity. The proteins with phospholipase activity were phospholipase A2 (PLA2), phosphoinositide phospholipase C, phospholipase D1 and a putative salivary secreted protein (Q6B8F9).

Secreted PLA2 is a component of tick saliva, where it functions not only as a lipolytic enzyme, but also as a regulator in many other processes including the host haemostatic, defensive and inflammatory responses (Kini, 2005; Manzano-Román et al., 2012b) and novel ways of cellular signalling and communication (Ibeas et al., 2009; Charalambous et al., 2012). The role of phospholipase A2 in tick midgut has been less investigated, but it has been suggested that PLA2 may play a role in the production of lipid mediators, the regulation of membrane remodelling, the degradation of foreign phospholipids, and the lysis of host blood cells in the gut lumen (Zhu et al., 1998; Murakami et al., 2015; Perner et al., 2016; Xu et al., 2016).

Regarding erythrocyte lysis, two transcripts annotated as putative saposin (B7Q6Z1) have been identified among the genes involved in lipid metabolic process in the *O. moubata* midgut (Table 3). This protein, which had not been previously identified in ticks, is expressed in the gastrodermis of adult schistosomes and other parasitic trematodes, where it would be involved in the lysis of ingested host cells and in the acquisition and transport across the gastrodermal cells of host lipids as a source of nutrition (Lee et al., 2002; Grams et al., 2006; Don et al., 2008). In ticks, the mechanisms for haemolysis of the red blood cell in the gut lumen and haemoglobin release are still poorly understood (Sojka et al., 2013). However, since saposin functions

as a haemolytic agent and transporter in schistosomes (a blood feeding trematode) it might be that the saposin homologue from ticks could play a similar role in the blood digestion process in the tick midgut.

Several transcripts coding for putative lipid carrier proteins were also identified (Table 3). Apolipoprotein and vitellogenin belong to the large lipid protein transfer superfamily. Vitellogenin, a precursor of vitellin that is produced in midgut cells and fat bodies in ticks, is essential for egg development and oviposition, and has been shown to play a role in heme sequestration (Smolenaars et al., 2007; Boldbaatar et al., 2010). In *O. moubata*, it has been demonstrated that vitellogenin expression increases after engorgement (Horigane et al., 2010). In the present work, 22 transcripts coding for vitellogenins have been identified, and 5 of them were upregulated. Additionally, the heme lipoprotein and hemelipoglycoprotein identified contain a β -sheet shell domain that represents a conserved region also found in vitellogenin and apolipoprotein B-100 (Anderson et al., 1998). Furthermore, it is known that hemelipoglycoproteins are involved in the detoxification process and energy production of larval hard tick stages (Donohue et al., 2008).

Among the lipid transporter proteins, an up-regulated fatty acid binding protein was also identified. According to that described by Stutzer et al. (2013) for *R. microplus*, this protein may play a role in intracellular transport of long-chain fatty acids. Other putative lipid transporter proteins not previously identified in the tick midgut were a oxysterol-binding protein, a phosphoinositol 4-phosphate adaptor protein and a phospholipid-transporting ATPase. The oxysterol-binding protein would be involved in cholesterol cycling and, possibly, in additional functions related to signal transduction pathways and cell cycle regulation (Fu et al., 2011).

1358
1359
1360 546 3.4.2. Defence response and response to oxidative stress.
1361

1362 547 The midgut represents the first major internal tick-tissue barrier encountered by
1363
1364 548 the pathogens ingested with blood. Additionally, the hydrolysis of haemoglobin in the
1365
1366 549 digestive vesicles of digestive cells results in the release of large amounts of heme, a
1367
1368
1369 550 pro-oxidant compound, which leads to the production of hydroxyl radicals (Citelli et al.,
1370
1371 551 2007). Thus, ticks need efficient protective mechanisms against potential invading
1372
1373 552 pathogens and oxidative stresses produced during blood digestion.
1374

1375 553 Up to 21 transcripts, most of them up-regulated after blood feeding, encoding
1376
1377 554 proteins putatively associated with defence responses were identified (Table 4,
1378
1379 555 Supplementary Table 5). Among them, 8 up-regulated transcripts corresponded to
1380
1381 556 antimicrobial peptides such as lysozyme and defensins. Lysozyme is the most potent
1382
1383 557 molecule active against Gram-positive bacteria and defensins are widespread
1384
1385 558 antibacterial peptides in invertebrates, mammals and plants (Nakajima et al., 2002;
1386
1387 559 Kopáček et al., 1999). Lysozyme and defensins have been previously reported from the
1388
1389 560 midgut of *O. moubata*, where they are constitutively expressed and up-regulated in
1390
1391 561 response to blood feeding (Grunclová et al., 2003; Nakajima et al., 2002), which is in
1392
1393 562 accordance with our current results. Another two transcripts have been annotated as
1394
1395 563 peptidoglycan-recognition protein, a protein which was also identified in the mialome
1396
1397 564 of *Ixodes ricinus* and that may function as pathogen recognition receptor (Kotsyfakis et
1398
1399 565 al., 2015). The remaining 12 transcripts also code for proteins involved in defence
1400
1401 566 responses according to the GO database (Table 4).
1402

1403
1404 567 Ticks have developed several antioxidant defences and heme detoxification
1405
1406 568 mechanisms such as the accumulation of heme into hemosomes (Lara et al., 2003).
1407
1408 569 Regarding the genes involved in these processes, in the current work 78 transcripts
1409
1410 570 coding for several chaperones with protein folding function, and numerous antioxidant
1411
1412
1413
1414
1415
1416

proteins with oxidoreductase activity have been identified. As was the case for the mialome of *Dermacentor marginatus*, most of these genes are up-regulated (Anderson et al., 2008) (Table 4, Supplementary Table 5).

Several of the antioxidant enzymes identified here, such as catalase, glutathione peroxidase, GSTs, thioredoxins, and superoxide dismutase (SOD), have already been identified in the midgut of *O. erraticus* and in other ixodid species, in which they play an important role in cellular stress responses, such as those occurring as a result of blood feeding, as well as in innate immunity (Dreher-Lesnick et al., 2006; Citelli et al., 2007; Anderson et al., 2008; Heekin et al., 2013; Stutzer et al., 2013; Oleaga et al., 2015).

Besides their involvement in the defense against the oxidative stress, these antioxidant enzymes could be also involved in the pathogens colonization and development in ticks. Actually, it has been demonstrated that the enzyme SOD aids in the colonization of *Amblyomma maculatum* tissues by *Rickettsia parkeri*, and that Salp25D, a homologue of peroxiredoxin, plays a critical role in the acquisition of *Borrelia burgdorferi* by the *Ixodes scapularis* tick (Narasimhan et al., 2007; Crispell et al., 2016) . Similarly, the GST may be important during the *Anaplasma marginale* life cycle in *Dermacentor variabilis* and the thioredoxin reductase is known to play an essential role in the maintenance of the native microbial community associated with *Amblyomma maculatum* (Kocan et al., 2009; Budachetri and Karim, 2015).

Numerous heat shock proteins (HSPs) and other chaperones with a predicted protein folding function (peptidyl-prolyl cis-trans isomerase, protein disulphide isomerase and T-complex protein) were highly up-regulated in the *O. moubata* midgut after feeding. HSPs stabilize proteins and protect tick cells from damage when temperature changes from approximately 20 °C to 38 °C occur as the tick is ingesting

1476
1477
1478 596 warm host blood (Guilfoile and Packila, 2004). In hard ticks, it is known that heat shock
1479
1480 597 proteins are overexpressed during feeding and that they play crucial roles in the
1481
1482 598 mechanisms evolved by ticks in order to survive during blood feeding and to tolerate
1483
1484
1485 599 pathogen infection (Busby et al., 2012).

1486
1487 600 All of these results indicate that the genes involved in defence responses,
1488
1489 601 detoxification and stress responses seem to be tightly regulated and their expression
1490
1491 602 levels significantly increase upon blood feeding.

1493 603 1494 1495 604 **3.5. Experimental validation of RNA-seq results by PCR and qRT-PCR**

1496
1497 605 The correctness of assembled transcripts was evaluated for six randomly
1498
1499 606 selected genes by PCR amplification, cloning and sequencing. All the amplicons
1500
1501 607 showed the expected size (Supplementary Fig. 2), and their nucleotide sequences were
1502
1503 608 100% identical to those of the corresponding transcripts obtained by RNA-seq,
1504
1505 609 confirming the accuracy of their assembly.

1506
1507 610 To confirm experimentally the consistency of the gene expression levels given
1508
1509 611 by the RNA-seq analysis, the expression of these six selected genes were evaluated in
1510
1511 612 fed and unfed female ticks by qRT-PCR using actin as a reference gene. This analysis
1512
1513 613 showed that the genes named Om 99 (Protein NEF1, putative), Om03 (Uncharacterized
1514
1515 614 protein), Om17 (Receptor expression-enhancing protein), Om28 (Organic
1516
1517 615 cation/carnitine transporter, putative) and Om29 (KOG3882-like protein) were up-
1518
1519 616 regulated after feeding, while the Om86 (OS86 protein) was not differentially expressed
1520
1521 617 after feeding (Fig.5). The comparison between the qRT-PCR and RNA-seq results
1522
1523 618 confirmed that all genes showed a similar trend in transcriptional expression changes to
1524
1525 619 that obtained by RNA-seq, although the magnitude of fold change in expression may
1526
1527 620 have been different between both methods, particularly for gene Om17 (Fig. 5). For this

gene, both methods indicated that it was significantly up-regulated (fold change >2), although the fold change given by qRT-PCR is 26.3 and by RNA-seq is 2.04 (Fig. 5). This discrepancy between the results of both analyses is not uncommon and may be due to several reasons including variability in the expression of the reference gene in different samples, differences in amplification efficiency in different genes, or differences in data standardization method between qRT-PCR and RNA-Seq analyses (Xu et al., 2016).

4. Conclusions

In this work we reported for the first time the next generation transcriptome sequencing, assembly, characterization and functional annotation of the midgut transcriptome of an argasid tick species. The study also includes the identification and functional analysis of differentially expressed genes (DEGs) after a blood meal, which in turn has allowed the identification of metabolic processes and cellular functions that were differentially expressed in the *O. moubata* midgut after feeding. The biological significance of such a differential expression is analysed and commented on, and special attention is given to the functional groups of genes/proteins predicted to be involved in the processes of blood digestion, nutrient transport and metabolism, and in defence response and response to the oxidative stress that is produced during blood digestion. Overall, these analyses have provided a wealth of novel information allowing a more in-depth understanding of the biochemistry and physiology of blood digestion in argasids. This information could be a starting point for the identification and selection of new targets for the development of alternative strategies for tick and tick-borne disease control.

Finally, this work also includes the experimental validation of the RNA-seq results by conventional PCR and qRT-PCR analysis for six randomly selected genes as regards the accuracy of the transcript sequence assembly and the assessment of transcript expression levels.

Acknowledgements

The authors are grateful to Rocío Vizcaíno Marín, María González Sánchez and Sergio González Pérez, from the Instituto de Recursos Naturales y Agrobiología de Salamanca (IRNASA, CSIC) (Spain), for their skilful technical assistance. The Illumina sequencing and bioinformatics assembly were carried out, respectively, in the Genomics Unit of the Madrid Science Park (Spain) and the Cenit Support Systems Company (Salamanca, Spain). This research was funded by project AGL2013-42745-P granted by the Spanish Ministry of Economy and Competitiveness.

References

- Akov, S., 1982. Blood digestion in ticks. In: Oberchain, F.D. and Gallun, R., (Ed.), Physiology of ticks. Pergamon Press Ltd, Oxford, England, pp.197–212.
- Akov, S., Samish, M., Galun, R., 1976. Protease activity in female *Ornithodoros tholozani* ticks. Acta Trop. 33, 37-52.
- Anderson, J.M., Sonenshine, D.E., Valenzuela, J.G., 2008. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick, *Dermacentor variabilis* (Acari: Ixodidae). BMC Genomics 9, 552.
- Anderson, T.A., Levitt, D.G., Banaszak, L.J., 1998. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. Structure 6, 895-909.
- Apanaskevich, D.A., Oliver, J.H. Jr., 2014. Life Cycles and Natural History of Ticks. In: Sonnenshine, D.E., Roe, R.M. (eds.) Biology of ticks. Oxford University Press, pp: 59-73.
- Boldbaatar, D., Umemiya-Shirafuji, R., Liao, M., Tanaka, T., Xuan, X., Fujisaki, K., 2010. Multiple vitellogenins from the *Haemaphysalis longicornis* tick are crucial for ovarian development. J. Insect Physiol. 56, 1587-98.
- Budachetri, K., Karim, S., 2015. An insight into the functional role of thioredoxin reductase, a selenoprotein, in maintaining normal native microbiota in the Gulf Coast tick (*Amblyomma maculatum*). Insect Mol. Biol. 24, 570–581.
- Busby, A.T., Ayllón, N., Kocan, K.M., Blouin, E.F., de la Fuente, G., Galindo, R.C., Villar, M., de la Fuente, J., 2012. Expression of heat shock proteins and subolesin affects stress responses, *Anaplasma phagocytophilum* infection and questing behaviour in the tick, Ixodes scapularis. Med Vet Entomol. 26, 92-102.
- Caperucci, D., Bechara, G.H., Camargo Mathias, M.I., 2010. Ultrastructure features of the midgut of the female adult *Amblyomma cajennense* ticks Fabricius, 1787 (Acari: Ixodidae) in several feeding stages and subjected to three infestations. Micron. 41, 710–21.
- Charalambous, K., Booth, P.J., Woscholski, R., Seddon, J.M., Templer, R.H., Law, R.V., Barter, L.M., Ces, O., 2012. Engineering de novo membrane-mediated protein-protein communication networks. Journal of the American Chemical Society 134, 5746-5749.
- Citelli, M., Lara, F.A., da Silva Vaz I. Jr, Oliveira, P.L., 2007. Oxidative stress impairs heme detoxification in the midgut of the cattle tick, *Rhipicephalus (Boophilus) microplus*. Mol. Biochem. Parasitol. 151, 81-8.

- 693 Coons, L.B., Rosell-Davis, R., Tarnowski, B., 1986. Blood meal digestion in ticks. In:
694 Sauer, J.R., Hair, J.A., (Ed.), Morphology, physiology, and behavioural biology of
695 ticks, Ellis Horwood limited, Chichester, England, pp. 248–73.
- 696 Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucl. Acids
697 Res. 16, 10881-10890.
- 698 Crispell, G., Budachetri, K., Karim, S., 2016. *Rickettsia parkeri* colonization in
699 *Amblyomma maculatum*: the role of superoxide dismutases. Parasit. Vectors 9,
700 291.
- 701 Cutler, S.J., 2010. Relapsing fever—a forgotten disease revealed. J. Appl.
702 Microbiol. 108, 1115–1122.
- 703 de la Fuente, J., 2012. Vaccines for vector control: exciting possibilities for the future.
704 Vet. J. 194, 139–140.
- 705 de la Fuente, J., Estrada-Peña, A., Venzal, J.M., Kocan, K.M., Sonenshine, D.E., 2008.
706 Overview: ticks as vectors of pathogens that cause disease in humans and animals.
707 Front. Biosci. 13, 6938–6946.
- 708 de la Fuente, J., Kopáček, P., Lew-Tabor, A., Maritz-Olivier, C., 2016. Strategies for
709 new and improved vaccines against ticks and tick-borne diseases. Parasite
710 Immunol. 0, 1-16
- 711 Díaz-Martín, V., Manzano-Román, R., Obolo-Mvoulouga, P., Oleaga, A., Pérez-
712 Sánchez, R., 2015. Development of vaccines against *Ornithodoros* soft ticks: an
713 update. Tick Tick Borne Dis. 6, 211–220.
- 714 Don, T.A., Bethony, J.M., Loukas, A., 2008. Saposin-like proteins are expressed in the
715 gastrodermis of *Schistosoma mansoni* and are immunogenic in natural infections.
716 Int J. Infect. Dis. 12, 39-47.
- 717 Donohue, K.V., Khalil, S.M., Mitchell, R.D., Sonenshine, D.E., Roe, R.M., 2008.
718 Molecular characterization of the major hemelipoglycoprotein in ixodid ticks.
719 Insect Mol. Biol. 17, 197-208.
- 720 Dreher-Lesnick, S.M., Mulenga, A., Simser, J.A., Azad, A.F., 2006. Differential
721 expression of two glutathione S-transferases identified from the American dog
722 tick, *Dermacentor variabilis*. Insect Mol. Biol. 15, 445-453.
- 723 Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence
724 similarity searching. Nucleic Acids Res. 39 (Web Server issue), W29-37.
- 725 Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter,
726 S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J.,

- Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, 279-85.
- Franta, Z., Frantová, H., Konvičková, J., Horn, M., Sojka, D., Mareš, M., Kopáček, P., 2010. Dynamics of digestive proteolytic system during blood feeding of the hard tick *Ixodes ricinus*. *Parasit Vectors.* 3, 119.
- Fu, Q., Lynn-Miller, A., Lan, Q., 2011. Characterization of the oxysterol-binding protein gene family in the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol.* 20, 541-52.
- García-Varas, S., (Doctoral thesis) 2004. Vacuna Anti-*Ornithodoros moubata*. Universidad de Salamanca, Spain <http://hdl.handle.net/10261/11031>
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644-52.
- Grams, R., Adisakwattana, P., Ritthisunthorn, N., Eursitthichai, V., Vichasri-Grams, S., Viyanant, V. The saposin-like proteins 1, 2, and 3 of *Fasciola gigantica*. *Mol Biochem Parasitol*, 148 (2006), pp. 133–143
- Grandjean, O., 1983. Blood digestion in *Ornithodoros moubata* Murray sensu strictu Walton females (Ixodoidea: Argasidae) II. Modification of midgut cells related to digestive cycle and triggering action of mating. *Ann. Parasitol. Hum. Comp.* 58, 493–14.
- Grandjean, O., 1984. Blood digestion in *Ornithodoros moubata* Murray sensu strictu Walton females (Ixodoidea: Argasidae). I. Biochemical changes in midgut cells related to intracellular digestion. *Acarologia* 25, 147–65.
- Grigor'eva, L.A., Amosova, L.I., 2004. Peritrophic matrix in the midgut of tick females of the genus *Ixodes* (Acari: Ixodidae)]. *Parazitologiya.* 38, 3-11.
- Grunclová, L., Fouquier, H., Hypsa, V., Kopáček, P., 2003. Lysozyme from the gut of the soft tick *Ornithodoros moubata*: the sequence, phylogeny and post-feeding regulation. *Dev. Comp. Immunol.* 27, 651-60.
- Grunclová, L., Horn, M., Vancová, M., Sojka, D., Franta, Z., Mares, M., Kopáček, P., 2006. Two secreted cystatins of the soft tick *Ornithodoros moubata*: differential expression pattern and inhibitory specificity. *Biol. Chem.* 387, 1635-1644.
- Guerrero, F., Miller, R.J., Pérez de León, A., 2012. Cattle tick vaccines: many candidates antigens, but will a commercial product emerge? *Int. J. Parasitol.* 42, 421–424.

- 761 Guilfoile, P.G., Packila, M., 2004. Identification of four genes expressed by feeding
762 female *Ixodes scapularis*, including three with sequence similarity to previously
763 recognized genes. *Exp Appl Acarol.* 32, 103-10.
- 764 Heekin, A.M., Guerrero, F.D., Bendele, K.G., Saldivar, L., Scoles, G.A., Dowd, S.E., et
765 al., 2013. Gut transcriptome of replete adult female cattle ticks, *Rhipicephalus*
766 (*Boophilus*) *microplus*, feeding upon a *Babesia bovis*-infected bovine host.
767 *Parasitol. Res.* 112, 3075–90.
- 768 Horigane, M., Shinoda, T., Honda, H., Taylor, D., 2010. Characterization of a
769 vitellogenin gene reveals two phase regulation of vitellogenesis by engorgement
770 and mating in the soft tick *Ornithodoros moubata* (Acari: Argasidae). *Insect.*
771 *Mol. Biol.* 19, 501-15.
- 772 Horn, M., Nussbaumerová, M., Sanda, M., Kovárová, Z., Srba, J., Franta, Z., Sojka, D.,
773 Bogyo, M., Caffrey, C.R., Kopáček, P., Mares, M., 2009. Hemoglobin digestion
774 in blood-feeding ticks: mapping a multi-peptidase pathway by functional
775 proteomics. *Chem Biol.* 16, 1053-1063.
- 776 Ibeas, E., Fuentes, L., Martin, R., Hernandez, M., Nieto, M.L., 2009. Secreted
777 phospholipase A2 type IIA as a mediator connecting innate and adaptive
778 immunity: new role in atherosclerosis. *J. Cardiovasc. Transl. Res.* 81, 54-63.
- 779 Kini, R.M., 2005. Structure-function relationships and mechanism of anticoagulant
780 phospholipase A2 enzymes from snake venoms. *Toxicon.* 45, 1147-61.
- 781 Kocan, K.M., de la Fuente, J., Blouin, E.F., Garcia-Garcia, J.C., 2004. *Anaplasma*
782 *marginale* (Rickettsiales: Anaplasmataceae): recent advances in defining host-
783 pathogen adaptations of a tick-borne rickettsia. *Parasitology* 129, 285-300.
- 784 Kocan, K.M., Zivkovic, Z., Blouin, E.F., Naranjo, V., Almazan, C., Mitra, R., de la
785 Fuente, J., 2009. Silencing of genes involved in *Anaplasma marginale*-tick
786 interactions affects the pathogen developmental cycle in *Dermacentor variabilis*.
787 *BMC Dev Biol* 9, 42.
- 788 Kongsuwan, K., Josh, P., Zhu, Y., Pearson, R., Gough, J., Colgrave, M.L., 2010.
789 Exploring the midgut proteome of partially fed female cattle tick (*Rhipicephalus*
790 (*Boophilus*) *microplus*). *J. Insect Physiol.* 56, 212–226.
- 791 Kopáček, P., Vogt, R., Jindrák, L., Weise, C., Safarik, I., 1999. Purification and
792 characterization of the lysozyme from the gut of the soft tick *Ornithodoros*
793 *moubata*. *Insect Biochem. Mol. Biol.* 29, 989-97.

- 794 Kotsyfakis, M., Schwarz, A., Erhart, J., Ribeiro, J.M., 2015. Tissue- and time-dependent
795 transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on
796 the vertebrate host. *Sci. Rep.* 5, 9103.
- 797 Laity, J.H., Lee, B.M., Wright, P.E., 2001. Zinc finger proteins: new insights into
798 structural and functional diversity. *Curr. Opin. Struct. Biol.* 11, 39–46.
- 799 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat.*
800 *Methods.* 9, 357-9.
- 801 Lara, F.A., Lins, U., Paiva-Silva, G., Almeida, I.C., Braga, C.M., Miguens, F.C. et al.,
802 2003. A new intracellular pathway of haem detoxification in the midgut of the
803 cattle tick *Boophilus microplus*: aggregation inside a specialized organelle, the
804 hemosome. *J. Exp. Biol.* 206:1707–15.
- 805 Lee, J.Y., Cho, P.Y., Kim, T.Y., Kang, S.Y., Song, K.Y., Hong, S.J., 2002. Hemolytic
806 activity and developmental expression of pore-forming peptide, clonorin.
807 *Biochem. Biophys. Res. Commun.* 296, 1238-44.
- 808 Lehane, M.J., 1997. Peritrophic matrix structure and function. *Annu. Rev. Entomol.* 42,
809 525-50.
- 810 Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq
811 data with or without a reference genome. *BMC Bioinformatics.* 12, 323.
- 812 Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and
813 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- 814 Mans, B.J., Neitz, A.W., 2004. Adaptation of ticks to a blood-feeding environment:
815 evolution from a functional perspective. *Insect Biochem. Mol. Biol.* 34, 1–17.
- 816 Manzano-Román, R., Díaz-Martín, V., de la Fuente, J., Ricardo Pérez-Sánchez, R.,
817 2012a. Soft ticks as pathogen vectors: distribution, surveillance and control. In:
818 Manjur S., (Ed.), *Parasitology*. Intech, Rijeka (Croatia), pp. 125-162.
- 819 Manzano-Román, R., Díaz-Martín, V., González-González, M., Matarraz, S., Álvarez-
820 Prado, A.F., LaBaer, J., Orfao, A., Pérez-Sánchez, R., Fuentes, M., 2012b. Self-
821 assembled protein arrays from an *Ornithodoros moubata* salivary gland
822 expression library. *J. Proteome Res.* 11, 5972–5982.
- 823 Manzano-Román, R., Encinas-Grandes, A., Pérez-Sánchez, R., 2006. Antigens from the
824 midgut membranes of *Ornithodoros erraticus* induce lethal anti-tick immune
825 responses in pigs and mice. *Vet. Parasitol.* 135, 65–79.
- 826 Manzano-Román, R., García-Varas, S., Encinas-Grandes, A., Pérez-Sánchez, R., 2007.
827 Purification and characterization of a 45-kDa concealed antigen from the midgut

1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006

828 membranes of *Ornithodoros erraticus* that induces lethal anti-tick immune
829 responses in pigs. Vet. Parasitol. 145, 314–325.

830 Maris, C., Dominguez, C., Allain, F.H., 2005. The RNA recognition motif, a plastic
831 RNA-binding platform to regulate post-transcriptional gene expression. FEBS J.
832 272, 2118-31.

833 Maritz-Olivier, C., van Zyl, W., Stutzer, C., 2012. A systematic, functional genomics,
834 and reverse vaccinology approach to the identification of vaccine candidates in the
835 cattle tick, *Rhipicephalus microplus*. Ticks Tick Borne Dis. 3, 179–187.

836 Matsuo, T., Sato, M., Inoue, N., Yokoyama, N., Taylor, D., Fujisaki, K., 2003.
837 Morphological studies on the extracellular structure of the midgut of a tick,
838 *Haemaphysalis longicornis* (Acari: Ixodidae). Parasitol. Res. 90, 243-8.

839 Miyoshi, T., Tsuji, N., Islam, M.K., Alim, M.A., Hatta, T., Huang, X., Fujisaki, K.,
840 2008. A set of serine proteinase paralogs are required for blood-digestion in the
841 ixodid tick *Haemaphysalis longicornis*. Parasitol. Int. 57, 499-505.

842 Moreti R, Perrella NN, Lopes AR. Carbohydrate digestion in ticks and a digestive α -L-
843 fucosidase. J Insect Physiol. 2013 Oct;59(10):1069-75.

844 Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and
845 quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 5, 621-8.

846 Murakami, M., Sato, H., Miki, Y., Yamamoto, K., Taketomi, Y., 2015. A new era of
847 secreted phospholipase A2 (sPLA2). J. Lipid Res., 56, 1248-1261.

848 Nakajima, Y., van der Goes van Naters-Yasui, A., Taylor, D., Yamakawa, M., 2002.
849 Antibacterial peptide defensin is involved in midgut immunity of the soft tick,
850 *Ornithodoros moubata*. Insect Mol. Biol. 11, 611-8.

851 Narasimhan, S., Sukumaran, B., Bozdogan, U., Thomas, V., Liang, X., DePonte, K.,
852 Marcantonio, N., Koski, R.A., Anderson, J.F., Kantor, F., Fikrig, E., 2007. A tick
853 antioxidant facilitates the Lyme disease agent's successful migration from the
854 mammalian host to the arthropod vector. Cell Host Microbe 2, 7–18.

855 Oleaga, A., Obolo-Mvoulouga, P., Manzano-Román, R., Pérez-Sánchez, R., 2015.
856 Midgut proteome of an argasid tick, *Ornithodoros erraticus*: a comparison
857 between unfed and engorged females. Parasit. Vectors 8, 525.

858 Oleaga-Perez, A., Perez-Sanchez, R., Encinas-Grandes, A., 1990. Distribution and
859 biology of *Ornithodoros erraticus* in parts of Spain affected by African swine
860 fever. Vet. Record 126, 32-37.

2007
2008
2009 861 Pedra, J.H., Narasimhan, S., Rendić, D., DePonte, K., Bell-Sakyi, L., Wilson, I.B.,
2010 862 Fikrig, E., 2010. Fucosylation enhances colonization of ticks by *Anaplasma*
2011 863 *phagocytophilum*. Cell Microbiol. 12, 1222-34.
2012 864 Perner J, Provazník J, Schrenková J, Urbanová V, Ribeiro JM, Kopáček P, 2016. RNA-
2015 865 seq analyses of the midgut from blood- and serum-fed Ixodes ricinus ticks. Sci
2016 866 Rep. 6:36695.
2017 867 Quembo, C.J., Jori, F., Pérez-Sánchez, R., Heath, L., Vosloo, W., 2015. Investigation
2020 868 into the epidemiology of African swine fever virus at the wildlife-domestic
2021 869 interface of the Gorongosa National Park, Central Mozambique. Transbound.
2022 870 Emerg. Dis. 63, 443-451.
2023 871 Rachinsky, A., Guerrero, F.D., Scoles, G.A., 2008. Proteomic profiling of
2024 872 *Rhipicephalus (Boophilus) microplus* midgut responses to infection with *Babesia*
2025 873 *bovis*. Vet. Parasitol. 152, 294-313.
2026 874 Saier MH Jr, Beatty JT, Goffeau A, Harley KT, Heijne WH, Huang SC, Jack DL, Jähn
2027 875 PS, Lew K, Liu J, Pao SS, Paulsen IT, Tseng TT, Virk PS. The major facilitator
2028 876 superfamily. J Mol Microbiol Biotechnol. 1999 Nov;1(2):257-79. Review.
2029 877 Sánchez-Vizcaíno, J.M., Mur, L., Gomez-Villamandos, J.C., Carrasco, L., 2015.
2030 878 Anupdate on the epidemiology and pathology of African swine fever. J.
2031 879 Comp.Pathol. 152, 9e21.
2032 880 Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative
2033 881 C(T) method. Nat Protoc 3, 1101-8.
2034 882 Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: robust de novo
2035 883 RNA-seq assembly across the dynamic range of expression levels. Bioinformatics
2036 884 28, 1086-92.
2037 885 Schwarz, A., Tenzer, S., Hackenberg, M., Erhart, J., Gerhold-Ay, A., Mazur, J., et al.,
2038 886 2014. A systems level analysis reveals transcriptomic and proteomic complexity
2039 887 in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding.
2040 888 Mol. Cell Proteomics 13, 2725-35.
2041 889 Shields, M.J., Fischer, J.J., Wieden, H.J., 2009. Toward understanding the function of
2042 890 the universally conserved GTPase HflX from Escherichia coli: a kinetic approach.
2043 891 Biochemistry 48, 10793-802.
2044 892 Smith, T.F., 2008. Diversity of WD-repeat proteins. Subcell Biochem. 48, 20-30.

893 Smolenaars, M.M., Madsen, O., Rodenburg, K.W., Van der Horst, D.J., 2007.
894 Molecular diversity and evolution of the large lipid transfer protein superfamily. J.
895 Lipid Res. 48, 489-502.

896 Sojka, D., Franta, Z., Horn, M., Caffrey, C.R., Mareš, M., Kopáček, P., 2013. New
897 insights into the machinery of blood digestion by ticks. Trends Parasitol. 29, 276-
898 285.

899 Stutzer, C., Mans, B.J., Gaspar, A.R., Neitz, A.W., Maritz-Olivier, C., 2009.
900 *Ornithodoros savignyi*: soft tick apyrase belongs to the 5'-nucleotidase family.
901 Exp. Parasitol. 122, 318-27.

902 Stutzer, C., van Zyl, W.A., Olivier, N.A., Richards, S., Maritz-Olivier, C., 2013. Gene
903 expression profiling of adult female tissues in feeding *Rhipicephalus microplus*
904 cattle ticks. Int. J. Parasitol. 43, 541-54.

905 Takechi, R., Galay, R.L., Matsuo, T., Maeda, H., Kusakisako, K., Talactac, M.R.,
906 Mochizuki, M., Fujisaki, K., Tanaka, T., 2016. Role of the tumor necrosis factor
907 receptor-associated factor-type zinc finger domain containing protein 1
908 (TRAFD1) from the hard tick *Haemaphysalis longicornis* in immunity against
909 bacterial infection. Ticks Tick Borne Dis. 7, 36-45.

910 Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., 2007.
911 Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35(Web
912 Server issue), W71-4.

913 van Zyl, W.A., Stutzer, C., Olivier, N.A., Maritz-Olivier, C., 2014. Comparative
914 microarray analyses of adult female midgut tissues from feeding *Rhipicephalus*
915 species. Ticks Tick Borne Dis. 6, 84-90.

916 Vial, L., 2009. Biological and ecological characteristics of soft ticks (Ixodida:
917 Argasidae) and their impact for predicting tick and associated disease distribution.
918 Parasite 16, 191-202.

919 Villar, M., Popara, M., Ayllón, N., Fernández de Mera, I.G., Mateos-Hernández, L.,
920 Galindo, R.C., Manrique, M., Tobes, R., de la Fuente, J., 2014. A systems
921 biology approach to the characterization of stress response in *Dermacentor*
922 *reticulatus* tick unfed larvae. PLoS One 9, e89564.

923 Wang, B., Dukarevich, M., Sun, E.I., Yen, M.R., Saier, M.H. Jr., 2009. Membrane
924 porters of ATP-binding cassette transport systems are polyphyletic. J Membr Biol.
925 231, 1-10.

926 Willadsen, P., 2008. Anti-tick vaccines. In: Bowman, A.S., Nuttall, P. (Ed.), Ticks:
 927 Biology, Disease and Control. Cambridge University Press, Cambridge, pp. 424–
 928 446.

929 Xu, X.L., Cheng, T.Y., Yang, H., Liao, Z.H., 2016. De novo assembly and analysis of
 930 midgut transcriptome of *Haemaphysalis flava* and identification of genes involved
 931 in blood digestion, feeding and defending from pathogens. Infect. Genet. Evol. 38,
 932 62-72.

933 Zhu, K., Bowman, A.S., Dillwith, J.W., Sauer, J.R., 1998. Phospholipase A2 activity in
 934 salivary glands and saliva of the lone star tick (Acari: Ixodidae) during tick
 935 feeding. J. Med. Entomol. 35, 500-4.

936 Zhu, Z., Gern, L., Aeschlimann, A., 1991. The peritrophic membrane of *Ixodes ricinus*.
 937 Parasitol. Res. 77, 635-41.

938

939 **Figure legends**

940

941 Figure 1. *Ornithodoros moubata* midgut transcriptome. Classification of the annotatable
942 transcripts according to their molecular function (A) and biological process (B). Only
943 the categories represented by more than 30 genes have been included in the pie charts.
944 The number of transcripts is indicated inside the parentheses.

945

946 Figure 2. Scatter plot showing the fold change in gene expression in fed *versus* unfed.
947 The log₂ fold change of each gene is plotted against the total number of counts recorded
948 for that gene. Differentially expressed genes (fold change > 2 or < -2 and $p < 0.05$) are
949 highlighted in red.

950

951 Figure 3. Fold change distribution of the up- and down-regulated DEGs. Blue and red
952 bars refer to the up- and down-regulated genes, respectively, in the midgut from fed *O.*
953 *moubata* females as compared to unfed females.

954

955 Figure 4. Classification of the differential expressed genes (DEGs) in the midgut
956 transcriptome of fed *versus* unfed *O. moubata* females by molecular function (A) and
957 biological process (B).

958

959 Figure 5. qRT-PCR validation of six differentially expressed genes analysed by RNA-
960 seq. qRT-PCR was performed on 6 genes that were identified as being differentially
961 expressed between fed (blue) and unfed (red) ticks. The expression level of each gene
962 was normalised to its level in unfed ticks, and their relative mRNA expression levels
963 (fold change in fed *versus* unfed) are shown in the Y axis.

964

965 Supplementary Figure 1. Pfam domains in the up- and down-regulated DEGs in the
966 *Ornithodoros moubata* midgut after feeding.

967

968 Supplementary Figure 2. PCR amplification products from the cDNAs of genes: Om99,
969 Om03, Om17, Om28, Om29 and Om86. The sizes of the predicted cDNA coding
970 sequences are indicated inside parentheses, and the amplicons showing the expected
971 size are boxed. These amplicons were purified from agarose gels, cloned into the pSC-A
972 sequencing vector and then sequenced in both strands.

973 **Supplementary Tables:**

974 Supplementary Table 1. Primer sequences (5' to 3') and reaction conditions used for
975 conventional PCR and quantitative real time PCR (qRT-PCR) analysis for validation of
976 RNA-Seq results.

977

978 Supplementary Table 2. Annotated genes expressed in the midgut of *Ornithodoros*
979 *moubata*. U1, U2: midgut transcriptome from Unfed ticks. F1, F2: midgut transcriptome
980 from Fed ticks. Differently expressed genes (DEGs) in FED ticks as compared to the
981 UNFED ticks are those having a fold change > 2 (up-regulated) or < -2 (down-
982 regulated), and a FDR < 0.05 .

983

984 Supplementary Table 3. Biological pathways represented in the midgut of *Ornithodoros*
985 *moubata*, number of sequences and enzymes involved in each pathway. The analysis of
986 the midgut annotated genes was done in the KEEG pathway database.

987

988 Supplementary Table 4. Pfam domain families in the up- and down-regulated genes in
989 the midgut of *Ornithodoros moubata* after feeding.

990

991 Supplementary Table 5. Overview of RKPM values of putative genes involved in
992 protein digestion carbohydrate metabolism and transport, lipid metabolism and
993 transport, defense response and response to oxidative stress. U1, U2: midgut
994 transcriptome from Unfed ticks. F1, F2: midgut transcriptome from Fed ticks.

995

996

B

BIOLOGICAL PROCESS

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Table 1. Top 30 Biological Pathways represented in the transcriptome of *O. moubata* midgut.

Class	Pathway	Pathway ID	Sequences in Pathway
-	Biosynthesis of antibiotics	map01130	167
Amino acid metabolism	Cysteine and methionine metabolism	map00270	38
	Glutathione metabolism	map00480	43
	Tryptophan metabolism	map00380	40
	Lysine degradation	map00310	36
	Glycine, serine and threonine metabolism	map00260	33
Carbohydrate metabolism	Amino sugar and nucleotide sugar metabolism	map00520	39
	Starch and sucrose metabolism	map00500	36
	Citrate cycle (TCA cycle)	map00020	31
	Pyruvate metabolism	map00620	30
	Glycolysis / Gluconeogenesis	map00010	39
Energy metabolism	Oxidative phosphorylation	map00190	37
	Carbon fixation pathways in prokaryotes	map00720	32
	Other glycan degradation	map00511	28
Immune system	T cell receptor signaling pathway	map04660	92
Lipid metabolism	Arachidonic acid metabolism	map00590	42
	Glycerophospholipid metabolism	map00564	40
	Fatty acid degradation	map00071	38
	Steroid hormone biosynthesis	map00140	38
	Linoleic acid metabolism	map00591	32
	Sphingolipid metabolism	map00600	32
Metabolism of cofactors and vitamins	Thiamine metabolism	map00730	302
Nucleotide metabolism	Purine metabolism	map00230	428

	Pyrimidine metabolism	map00240	89
Signal transduction	Phosphatidylinositol signaling system	map04070	40
Translation	Aminoacyl-tRNA biosynthesis	map00970	60
	Aminobenzoate degradation	map00627	123
Xenobiotics metabolism	Drug metabolism - other enzymes	map00983	71
	Drug metabolism - cytochrome P450	map00982	35
	Metabolism of xenobiotics by cytochrome P450	map00980	35

Table 2. Top 30 Pfam domains families represented in the up- and down-regulated DGEs in the midgut of *Ornithodoros moubata* after feeding.

Pfam No	Name	Counts	Summary
Up-regulated genes			
PF00076.17	RRM_1	76	RNA recognition motif
PF14259.1	RRM_6	73	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
PF13893.1	RRM_5	71	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
PF00005.22	ABC_tran	68	ABC transporter
PF00400.27	WD40	67	WD domain, G-beta repeat
PF07690.11	MFS_1	65	Major Facilitator Superfamily
PF01926.18	MMR_HSR1	62	50S ribosome-binding GTPase
PF00004.24	AAA	51	ATPase family associated with various cellular activities
PF00069.20	Pkinase	48	Protein kinase domain
PF08477.8	Miro	47	Ras of Complex, Roc, domain of DAPkinase
PF04156.9	IncA	46	The Inclusion Membrane Protein (IncA) Family
PF00071.17	Ras	45	Ras family
PF07714.12	Pkinase_Tyr	44	Protein tyrosine kinase
PF00009.22	GTP_EFTU	41	Elongation factor Tu GTP binding domain
PF02985.17	HEAT	41	HEAT repeat
PF00025.16	Arf	40	ADP-ribosylation factor family
PF00096.21	zf-C2H2	40	Zinc finger, C2H2 type
PF07719.12	TPR_2	38	Tetratricopeptide repeat
PF13646.1	HEAT_2	36	HEAT repeats
PF14531.1	Kinase-like	36	Kinase-like domain
PF01078.16	Mg_chelatase	35	Magnesium chelatase, subunit ChII
PF03193.11	DUF258	35	Protein of unknown function, DUF258
PF13191.1	AAA_16	34	ATPase domain
PF00270.24	DEAD	32	DEAD/DEAH box helicase
PF00910.17	RNA_helicase	32	RNA helicase family
PF03205.9	MobB	32	Molybdopterin guanine dinucleotide synthesis protein B
PF00085.15	Thioredoxin	31	Thioredoxin
PF00515.23	TPR_1	30	Tetratricopeptide repeat
PF00106.20	adh_short	29	Short chain dehydrogenase family
PF01636.18	APH	29	Phosphotransferase enzyme family
Down-regulated genes			
PF00069.20	Pkinase	65	Protein kinase domain
PF07714.12	Pkinase_Tyr	57	Protein tyrosine kinase
PF00078.22	RVT_1	46	Reverse transcriptase
PF00096.21	zf-C2H2	43	Zinc finger, C2H2 type
PF13894.1	zf-C2H2_4	39	C2H2-type zinc finger
PF00076.17	RRM_1	35	RNA recognition motif.
PF01636.18	APH	30	Phosphotransferase enzyme family
PF00515.23	TPR_1	29	Tetratricopeptide repeat
PF00665.21	rve	29	Integrase core domain
PF12874.2	zf-met	29	Zinc-finger of C2H2 type

PF13465.1	zf-H2C2_2	29	Zinc-finger double domain
PF14259.1	RRM_6	29	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
PF14531.1	Kinase-like	29	Kinase-like
PF00400.27	WD40	28	WD domain, G-beta repeat
PF07719.12	TPR_2	27	Tetratricopeptide repeat
PF12171.3	zf-C2H2_jaz	27	Zinc-finger double-stranded RNA-binding
PF00067.17	p450	26	Cytochrome P450
PF00097.20	zf-C3HC4	26	Zinc finger, C3HC4 type (RING finger)
PF00098.18	zf-CCHC	26	Zinc knuckle
PF00169.24	PH	26	Pleckstrin homology domain (PH domain)
PF06293.9	Kdo	26	Lipopolysaccharide kinase (Kdo/WaaP) family
PF00005.22	ABC_tran	25	ATP-binding domain of ABC transporters
PF01926.18	MMR_HSR1	25	50S ribosome-binding GTPase
PF13639.1	zf-RING_2	25	Ring finger domain
PF13893.1	RRM_5	25	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
PF13923.1	zf-C3HC4_2	25	Zinc finger, C3HC4 type (RING finger)
PF04156.9	IncA	23	IncA protein
PF04582.7	Reo_sigmaC	22	Reovirus sigma C capsid protein
PF00041.16	fn3	21	Fibronectin type III domain
PF00130.17	C1_1	21	Phorbol esters/diacylglycerol binding domain (C1 domain)

Table 3. Selected genes predicted to be involved in protein digestion, carbohydrate metabolism and transport, and lipid metabolism and transport. It is shown the number of up- and down-regulated transcripts as well as the number of transcripts whose expression did not significantly change (NS) after feeding. The RKPM values are shown in the Supplementary Table 5.

UniProt ID	Protein name	No. of transcripts			Gene ontology (molecular function)
		up-regulated	down-regulated	NS	
PROTEIN DIGESTION					
Aspartic peptidases					
B5G4Y2	Aspartic protease	1	0	0	aspartic-type endopeptidase activity [GO:0004190]
C1FXZ1	Cathepsin D preproprotein	1	0	0	
G1SGA0	Cathepsin E	1	0	0	
E7E821	Putative cathepsin D3	0	0	1	
D7EJ79, B7QM32, B7QBI4, E9IFF2	Putative uncharacterized proteins	2	2	0	
Cysteine peptidases					
B7PEU5, B7QCA9	Calcium-dependent cysteine protease	0	2	1	cysteine-type endopeptidase activity [GO:0004197]
A4GTA7	Cathepsin B-like cysteine protease	5	1	0	
A9Z0S7	Cathepsin L	1	0	0	
B0EIS2	Cysteine proteinase 3, putative	0	1	0	
B7PG68_IXOSC	Cystein protease, putative (Fragment)	0	0	2	
B7P7I7_IXOSC	Cysteine protease and A protease inhibitor, putative (EC 3.4.22.42)	1	0	0	
Q6NYJ7, Q6PRC7	Legumain	0	1	1	
B7PF28_IXOSC	Longipain, putative (EC 3.4.22.1)	0	0	1	
B7PF70, H3BIL1, F7DRZ9, H2ZK39, I1GAL0	Uncharacterized proteins	0	5	0	
Carboxypeptidases					
G3MKP9, W4YJH6, B7PC00, A0A087UIL3, B7Q1Z4, B7PTG1, B7PHS3	Carboxypeptidase	5	1	1	carboxypeptidase activity [GO:0004180]
B7QAP0	D-alanyl-D-alanine carboxypeptidase	0	2	0	
B7P458, B7PI32	Glutamate carboxypeptidase	0	1	1	
F0J8C4, B7Q8V9, B7QLB7, B7PI01	Serine carboxypeptidase	3	0	1	
Metallopeptidases					
E1ZWX5	ADAM 11	0	1	0	metallopeptidase activity [GO:0004222]
B7PBB5	Intermediate peptidase	0	0	3	
B7QLQ7	Leucine aminopeptidase	0	0	2	
E9HP49, G3MHG8	Metalloendopeptidase	0	2	2	
B7QBX2	Metalloendopeptidase OMA1	0	0	1	
Q09JL6, Q09JE8, B7P464	Metalloprotease	1	0	3	

B1GTW4	Metis2 protein	1	0	0	
B7P9R8	Oligopeptidase	1	1	0	
A7RJZ8	Predicted protein	1	0	0	
B7P573	Processing peptidase beta subunit	0	0	1	
B7PFC2	Protease	0	0	1	
B7PC82	Thimet oligopeptidase	0	0	1	
A6NA12	Truncated metalloprotease	0	2	0	
B7PFQ3, B7QM92, E9GFV5, D6WJP0, E9FQX7, B7PBF0, B7P8S4, E9H0Q2, G6CN60, G6DD40	Uncharacterized proteins	1	5	7	

Serine peptidase

F7BBQ0	ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92)	1	0	0	serine-type endopeptidase activity [GO:0004252]; serine-type exopeptidase activity [GO:0070008]
B7QMK0	Dipeptidyl peptidase IV	0	0	1	
Q8T4N4	Midgut serine proteinase-1	1	0	0	
A7RL44	Predicted protein (Fragment)	1	0	0	
B7PDF5	Prolyl endopeptidase, putative (EC 3.4.21.26)	0	0	1	
O96089	Serin proteinase 2	1	0	0	
B7PC98	Serin proteinase	0	0	1	
Q6L7Z5	Serine protease	1	0	0	
E2AUP0	Serine protease HTRA2	1	0	0	
Q6U8A8	Serine protease-like protein	6	0	0	
Q5DHM3	SJCHGC01895 protein	1	0	0	
B7X9Z8	Tick serine proteinase	0	1	1	
B7PQH8	Tripeptidyl-peptidase II	0	0	1	
E9FS53, O93424, E9FXX1, C3Z5C3, G3MIC1, G3MIU4, G3MG39, H0VU83, H2M4P3, H2YTU9, G3QLN3, H2UQQ4	Uncharacterized proteins	10	4	0	

CARBOHYDRATE METABOLISM AND TRANSPORT

Carbohydrate metabolism

A0A087UZ87	ADP-dependent glucokinase	0	0	1	kinase activity [GO:0016301]
Q7Q4V0	AGAP000862-PA	1	0	0	hydrolase activity [GO:0016787]
H2LYL1, H2Y4R9, H2Y4R7	Alpha-1,4 glucan phosphorylase	0	3	0	transferase activity [GO:0016740]
B7PY71	Alpha2-glucosyltransferase	1	0	0	transferase activity [GO:0016740]
E9GG04, H2U5F1, B7QMB4	Alpha-amylase	0	2	1	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]
B7PDZ5, B7PDZ6	Alpha-D-galactosidase	1	0	2	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]
E5S0Y1	Alpha-galactosidase A	0	1	0	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]
A0JPA7, A0A067R124, R7T6U8, G5CAB5, B7PF17, A0A087UQ06, A0A087UC61,	Alpha-mannosidase	1	4	7	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]

A0A087UC63, B7PL97						
A8P9E0	ATP citrate lyase isoform 2	1	0	0	lyase activity [GO:0016829]	
A0A087UFY7	Acetylgalactosaminyltransferase bre-4	0	0	1	transferase activity, transferring glycosyl groups [GO:0016757]	
E9FTN1, B7PW84	Beta-galactosidase	2	0	1	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]	
Q8VNV4, B7PJE8	Beta-glucuronidase	0	2	0	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]	
B7PE53, A0A087TZD0, A0A087TZD0	Beta-hexosaminidase	0	0	3	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]]	
B7PL94	Beta-mannosidase	0	1	0	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]]	
B7QBA2	Chondroitin 4-sulfotransferase	0	0	1	transferase activity [GO:0016740]	
G5APN7	FGGY carbohydrate kinase domain-containing protein	1	0	0	transferase activity [GO:0016740]	
E6RF59, D0UPT3	Glucosamine-6-phosphate isomerase	0	1	1	isomerase activity [GO:0016853]	
G5BNZ1	Glucosamine-fructose-6-phosphate aminotransferase [isomerizing] 1	1	0	0	transferase activity [GO:0016740]	
B7QCA7	Glucosidase II	1	0	0	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]	
B7PK21, B7PNW7, E7EZM1	Glucosylceramidase	0	2	3	hydrolase activity, hydrolyzing O-glycosyl compounds [GO:0004553]	
A2V728	Glutamine: fructose-6-phosphate aminotransferase	0	0	1	trasferase activity [GO:0016740]	
B7QLX0	Glycerol kinase	0	0	1	trasferase activity [GO:0016740]	
F1KZX3	Glycerol-3-phosphate dehydrogenase [NAD(+)]	0	2	0	oxidoreductase activity [GO:0016491]	
B7PH44	Malate dehydrogenase	0	0	2	oxidoreductase activity [GO:0016491]	
B7QDE7	Mannosyl-oligosaccharide glucosidase	1	0	0	hydrolase activity [GO:0016787]	
H3AUP9	N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) (GlcNAc 6-P deacetylase)	0	0	1	hydrolase activity [GO:0016787]	
A0A0C9Q5I1	Pgls_0 protein	0	0	1	hydrolase activity, acting on ester bonds [0016788]	
A9P774, F8U8U4, B7P9M5	Phosphoenolpyruvate carboxykinase	0	2	1	lyase activity [GO:0016829]	
B7PKT6	Poly(ADP-ribose) glycohydrolase	0	1	0	hydrolase activity, acting on glycosyl bonds [GO:0016798]	
A9V3R2	Predicted protein	1	0	0	isomerase activity [GO:0016853]	
B7P559	Ribulose-phosphate 3-epimerase	1	0	0	isomerase activity [GO:0016853]	
B4F740, B7PHV2, B7Q8M5, D6WWI8, D8PNS7, E9FVC6, F6WCE6, G1N4C4, G3MK74, G3MM76, G8BW61, H2YDG1, H2YRG6, I1CM06, Q3URN2	Uncharacterized proteins	7	11	1	—	
Carbohydrate transport						
E2C2M2	Glucose transporter type 1	1	0	0	substrate-specific transmembrane transporter activity [GO:0022891]	
B7PS38	Sodium-dependent glucose transporter	1	0	1	transmembrane transporter activity [GO:0022857]	
B7P811, B7Q2F5, B7PXB1, B7PIM0, B7PST3, B7P5I5, B7PPW3	Sugar transporter	4	2	3	transmembrane transporter activity [GO:0022857]	

Chitin metabolism					
Q179H4	AAEL005622-PA	1	0	0	chitin binding [GO:0008061]
Q7PX44	AGAP001203-PA	1	0	0	chitin binding [GO:0008061]
A0A084VDZ0	AGAP006898-PA-like protein	2	0	2	chitinase activity [GO:0004568]; chitin binding [GO:0008061]
B7P5R3, B7P5R4, D0E0G4, B7P5R3, I4DQ04, D0E0G4	Chitinases	1	2	3	chitinase activity [GO:0004568]
F4WFW0	JNK1-associated membrane protein	1	0	0	-
B7QKX7	Peritrophic membrane chitin binding protein, putative (Fragment)	1	0	0	chitinase activity [GO:0004568]
A8CW99	Putative peritrophin	1	0	0	chitin binding [GO:0008061]
G3MR91, B7Q1A6, H0ZHH9	Uncharacterized proteins	3	0	0	chitinase activity [GO:0004568]

LIPID METABOLISM AND TRANSPORT
Lipid metabolism

D3BK46	Acyl-CoA dehydrogenase	1	0	0	oxidoreductase activity [GO:0016491]
A0A087U9L2	Choline/ethanolaminephosphotransferase 1	0	0	1	transferase activity [GO:0016780]
Q4RNP3	Chromosome 2 SCAF15010	1	0	0	transferase activity [GO:0016780]
B7QG88	Dfg10 protein	0	1	0	oxidoreductase activity [GO:0016491]
A0A087UGN3, B7PBT7, B7PKN6	Glycerol-3-phosphate acyltransferase	0	1	4	transferase activity [GO:0016780]
A0A067QFB7	Glycerophosphodiester phosphodiesterase domain-containing protein 1	0	0	1	hydrolase activity [GO:0008889]
B7PNT8	Glycerophosphoryl diester phosphodiesterase	1	0	0	hydrolase activity [GO:0008889]
B7Z150	Inactivation no afterpotential E	0	1	0	hydrolase activity [GO:0008889]
B7PG19	Inositol-1-phosphate synthetase	0	1	1	isomerase activity [GO:0016853]
B7Q0I4, B7PJA8	Long chain fatty acid CoA ligase	0	0	6	ligase activity [GO:0016874]
B7Q8J7	N-acylsphingosine amidohydrolase (Acid ceramidase)	0	0	1	hydrolase activity [GO:0008889]
Q1LXK8	Novel protein containing an Acyl-CoA dehydrogenase	1	0	0	oxidoreductase activity [GO:0016491]
B7PJJ9	Phosphatidylcholine transfer protein	0	0	1	lipid binding [GO:0008289]
B7QN53	Phosphatidylinositol-specific phospholipase C X domain-containing protein	0	0	1	hydrolase activity [GO:0008889]
F5HMB9, B4Q064, B7P6Q5, B7Q2N6, D2HGF6	Phosphoinositide phospholipase C	0	7	0	phospholipase activity [GO:0004620]
B7QGM6, A0A087U096, E2AUV7, B7P0P3	Phospholipases A2	1	1	3	phospholipase activity [GO:0004620]
A0A067R529	Phospholipase D1	0	0	1	phospholipase activity [GO:0004620]
A0A087US76	Putative cardiolipin synthase	0	0	1	transferase activity [GO:0016780]
Q6B8F9	Putative salivary secreted peptide	2	0	0	phospholipase activity [GO:0004620]
B7PYU9	Rat sphingolipid delta 4 desaturase, putative	0	0	2	oxidoreductase activity [GO:0016491]
B7Q6Z1	Saposin, putative (EC 3.4.23.40)	0	1	1	hydrolase activity [GO:0016787]
A0A067R0A6	Sn1-specific diacylglycerol lipase alpha	1	0	1	hydrolase activity [GO:0016787]

B7QE92	Steroidogenic acute regulatory protein	1	0	0	lipid binding [GO:0008289]
B7PP53	Triacylglycerol lipase, putative	0	0	3	hydrolase activity [GO:0016787]
G6CQP9, E9IAX1, E9GSQ1, D2A2L1, E9FT86, B7P4U8, R7T915, A0A087TT40, B5M6F6, B7Q3J2, G1NWU5, E9IWS1, F7E622, E9GAX7, E9IF21	Uncharacterized proteins	5	8	4	—
Lipid transport					
A0A087UIX1	Apolipoprotein B-100	0	1	0	lipid transporter activity [GO:0005319]
B7QMW0	Fatty acid-binding protein FABP	1	0	0	transporter activity [GO:0005215]
A0MVX0_	Heme lipoprotein	0	0	1	lipid transporter activity [GO:0005319]
B7QIL0, B7Q402	Hemelipoglycoprotein	1	1	0	lipid transporter activity [GO:0005319]
I3J0T7, E0VXN7, B7P7L6, B7P2I1	Oxysterol-binding protein	2	0	2	lipid transporter activity [GO:0005319]
B7Q2X5, B7PLB7	Phosphoinositol 4-phosphate adaptor protein	1	0	3	lipid transporter activity [GO:0005319]
V4B0M8	Phospholipid-transporting ATPase	0	0	1	lipid transporter activity [GO:0005319]
B6ZIV7, B7Q7E5	Vitellogenin	2	1	9	lipid transporter activity [GO:0005319]
S5G8C9, E1CAX9	Vitellogenin 1	3	0	2	lipid transporter activity [GO:0005319]
B1B544	Vitellogenin-2	0	1	0	lipid transporter activity [GO:0005319]
G9M4L6	Vitellogenin-B	0	2	2	lipid transporter activity [GO:0005319]
B7Q381, E9HQQ7, B7PDP4, G3ML22, B7PH02	Uncharacterized proteins	3	2	4	lipid transporter activity [GO:0005319]

B7XFT1	Persulcatusin (defensin)	1	0	0	–
B7PUU4, B7Q5T0	Toll, putative	1	1	4	binding [GO:0005488]
G3PHS3, G3MHR5, H2TVT4	Uncharacterized protein	2	1	0	–
RESPONSE TO OXIDATIVE STRESS					
B7G713, C4JIX9	Catalase	2	0	0	oxidoreductase activity [GO:0016491]
G3HF60, B7PTW0, Q5RJH8, B7PZQ2	Glutathione peroxidase	2	1	1	oxidoreductase activity [GO:0016491]
B7PB71, B7PG84, B7PPH2, B7PR98, B7PRA0, B7PXH0, B7Q278, B7Q917, B7Q9K1, B7Q9K1, B7QMI2	Glutathione S-transferase	7	3	4	transferase activity [GO:0016740]
E2C5D4	Heat shock factor-binding protein	1	0	0	protein folding [GO:0006457]
A1SXX9	Heat shock protein DnaJ domain protein	0	1	0	protein folding [GO:0006457]
B7PYP5, B7P1Z8	Heat shock protein, putative	2	0	0	protein folding [GO:0006457]
G8FUE4, B7PPP4, B7Q0B3, Q0Z7Z6, Q98899	Heat shock protein 70 (HSP70)	5	2	0	protein folding [GO:0006457]
E5DVR6, G9MD87, A2TE72	Heat shock protein 90 (HSP90)	5	0	0	protein folding [GO:0006457]
Q86G69	Putative heat shock-related protein	1	0	0	-
B7Q709	Oxidative stress-induced growth inhibitor	2	0	0	oxidoreductase activity [GO:0016491]
B5XDJ1	Peptidyl-prolyl cis-trans isomerase	1	0	0	protein folding [GO:0006457]
B7QGB0, B7QCX2	Peroxidase	0	1	1	oxidoreductase activity [GO:0016491]
B7QAL3, B7PUM7, B7P9B9	Peroxinectin, putative	0	1	4	oxidoreductase activity [GO:0016491]
B7Q7K9	Peroxiredoxin (PRX)	1	0	0	oxidoreductase activity [GO:0016491]
E9FUV9, Q3TGS0	Protein disulfide-isomerase	2	0	0	protein folding [GO:0006457]
C3KK50, B7P328, H3CX53, G5BEA6, B7QEW8	Superoxide dismutase	1	2	4	oxidoreductase activity [GO:0016491]
B7FXH5, B7PRH5, G7YA28, E1G0P0, H3CG97, Q4UCM0	T-complex protein 1 subunit alpha	5	0	1	protein folding [GO:0006457]
E2ASR9, Q685V4	Thioredoxin-1	2	0	0	oxidoreductase activity [GO:0016491]

Q802D6	Thioredoxin 2	0	1	0	oxidoreductase activity [GO:0016491]
H2KVH0, H2KP99, B7PQP2, B7P4N2	Thioredoxin domain-containing proteins	4	0	0	cell redox homeostasis [0045454]
B7PTG8	Thioredoxin reductase, putative	0	0	3	oxidoreductase activity [GO:0016491]
B7QN17, G9KIF6	Thioredoxin-dependent peroxide reductase	2	0	0	oxidoreductase activity [GO:0016491]
Q29C01, A7RPZ4, E9H172	Uncharacterized proteins	1	1	1	oxidoreductase activity [GO:0016491]

MIDGUT TRANSCRIPTOME

Ornithodoros moubata

Unfed

48 hours post-feeding

ILLUMINA sequencing

Bioinformatics assembly

Transcriptome annotation and characterization

Identification of differentially
expressed genes

Supplementary Figure 1.

Supplementary Figure 2.

Highlights

- The differentially expressed genes after blood feeding are identified.
- Genes involved in blood digestion, metabolism and stress responses are discussed.
- This mialome helps understanding the molecular physiology of digestion in argasids.
- The midgut transcriptome of an argasid tick is assembled for the first time.