

RNAseq-based gene expression analysis of *Melolontha hippocastani* hindgut pockets and the surrounding hindgut wall tissue

Pol Alonso-Pernas¹ and Wilhelm Boland¹

¹Department of Bioorganic Chemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

Abstract

In this study, the metatranscriptome of newly-discovered structures attached at the distal end of the hindgut of the larvae of a coleopteran (*Melolontha hippocastani*), is compared with that of the surrounding hindgut wall. Larvae were collected in their natural habitat, RNA was extracted using a commercial kit and sequenced in a Illumina HiSeq2500 platform. 250 bp paired-end reads were used to de novo assemble the transcriptomes. Contig annotation was carried out with BLASTx and Blast2GO PRO and differential expression analysis was performed in edgeR. Contigs aligned mainly to *Achromobacter* sp. in the pockets and to the Firmicutes phylum in hindgut wall. Host RNAs were expressed in the pockets in higher amounts than in hindgut wall. Gene expression suggest that pocket bacteria undergo aerobic metabolism and are exposed to higher levels of oxidative stress than the population of the hindgut wall. Hypothetical functions for the pocket might be immune-stimulation and regulation of host development, while the hindgut wall appears to be devoted to degradation of dietary polysaccharides and host nitrogenous wastes. Further research is necessary to experimentally prove these suggested roles.

1. Introduction

Insects are one of the most successful animal groups on earth and they thrive on a plethora of different diets. Although their digestive tracts present a wide variety of morphologies as result of adaptation to disparate food sources (Engel & Moran 2013) they are generally divided in the same three sections: the foregut, where the food is preprocessed and digestion may start; the midgut, where host digestive enzymes are secreted into the gut lumen and, in combination with symbiotic activity, most of the digestion and nutrient absorption takes place; and the hindgut, where water and other small molecules are absorbed (Chapman 2013). However, in some insect groups, this scheme can be subjected to significant modifications. For example, in termites or scarabaeid beetles, the hindgut is enlarged in the so called paunch (in termites) or fermentation chamber (in scarabaeids) and it also participates in digestion with the aid of symbiotic microorganisms (Calderon-Cortes et al. 2012). Or in *Spodoptera* larvae, the foregut is enlarged and accumulates α - and β -carotene, apparently as an adaptation to feeding on toxic plants (Shao et al. 2011).

In some cases, the morphological diversification of the digestive tract creates additional structures exclusive of particular insect taxa. These structures may be devoted to the housing of a specific bacterial ectosymbiont: for example, the midgut crypts in bugs of the Alydidae family are colonized by the environmental bacterium *Burkholderia* sp. (Kikuchi 2005) or by Gammaproteobacteria in bugs of the Pentatomidae and Cydnidae families (Prado & Almeida 2009; Hosokawa et al. 2012); also, the pouch-like cavity located in the midgut-hindgut junction of *Tetraponera* ants is occupied by root-nodule nitrogen fixing related bacteria (van Borm et al. 2002). In other insects, bacterial or fungal endosymbionts inhabit specialized cells (bacteriocytes or mycetocytes) which usually are grouped into clusters called bacteriome or mycetome. This is the case of the lygaeid *Kleidocerys resedae* or the chrysomelid *Bromius obscurus* that house gammaproteobacterial symbionts within, respectively, midgut or foregut-midgut junction associated bacteriomes (Küchler et al. 2010; Fukumori et al. 2017). Similarly, the cerambycid beetles *Tetropium castaneum*, *Rhagium inquisitor* and *Leptura rubra* harbour, respectively, unidentified yeast, *Candida rhagii* and *Candida shehatae* endosymbionts within mycetomes associated with the proximal midgut (Grünwald et al. 2009). The purpose of the microbes inhabiting such symbiont-containing organs is unclear in most of the cases, although nitrogen

fixing has been suggested for the root nodule-related bacteria within the paunch of *Tetraponera* ants (Stoll et al. 2007) and *Burkholderia* sp. associated with the bean bug *Riptortus pedestris* has been proven capable of degrading the insecticide fenitrothion (Kikuchi et al. 2012) and promoting insect growth and egg production through modulation of host protein expression (Lee et al. 2017). Regarding the endosymbiotic microorganisms confined within bacteriomes or mycetomes, it is commonly accepted that they provide a variety of essential nutrients to the host (Bright & Bulgheresi 2010).

RNA sequencing (RNAseq) has been successfully applied to explore the gut function of insects under changing conditions (Roy et al. 2016) as well as the interaction of insect host with either symbiotic (Peterson & Scharf 2016; Emery et al. 2017) or pathogenic organisms (Chen et al. 2017; Yadav et al. 2017). Tissue-specific RNAseq can also unmask differentially expressed genes and functions that could otherwise remain unidentified in whole-gut analysis (Mamidala et al. 2012; Shelomi 2017; Nakayama et al. 2017). In the present experiment, we used Illumina HiSeq to survey the transcriptome of the hindgut pockets of the forest cockchafer (*Melolontha hippocastani*). These symbiont-containing organs, placed at both sides of the terminal segment of the larval (male and female) hindgut chamber, were discovered in the 50s (Wildbolz 1954) and their architecture and bacterial community have been recently characterized in detail utilizing modern techniques (Alonso-Pernas et al. 2017). However, little is known about the processes ongoing in these enigmatic structures, such as the mechanism of symbiotic colonization or their role in the context of insect physiology. We expect to get an insight into these matters by comparing the pocket transcriptome with that of the surrounding hindgut wall of larvae retrieved straight from their natural environment.

2. Materials and methods

2.1. Insect collection and RNA extraction

Second-instar (L2) *Melolontha hippocastani* larvae were collected from a deciduous forest next to Pfungstadt (Germany, 49°49'44" N 8°36'17" E) in June 2016. Larvae were carried to the laboratory in their natural soil and flash-frozen with liquid nitrogen immediately after arrival in order to

minimize changes in gene expression. Frozen larvae were stored at -80 °C until further processing. Diethyl pyrocarbonate (DEPC, Sigma Aldrich, Saint Louis, MO, USA) treated phosphate buffered saline solution (PBS) and ddH₂O (Nagy et al. 2007) were prepared by adding 0.1% v/v DEPC into PBS (composition per liter: 8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, 0.24 g KH₂PO₄ (pH 7.4)) or ddH₂O and incubating the mixtures at 37 °C for 24h before autoclaving. Dissection of thawed insects was carried out on ice in DEPC-treated PBS using ethanol sterilized tools rinsed with DEPC treated ddH₂O. Entire guts were carefully pulled out from the larvae. Pockets and 1x1 mm pieces of nearby hindgut wall were excised from the hindgut chamber and rinsed in DEPC-ddH₂O to remove debris and unattached bacteria. Tissues were pooled together in a 2 mL polypropylene screw cap micro tube (Sarstedt, Nümbrecht, Germany) forming 3 hindgut wall replicates, each consisting of 8 hindgut wall fragments from 8 different individuals, and 1 pocket replicate, consisting of approximately 100 pockets of 50 different individuals. Reduced size of pocket tissue prevented the obtainment of more replicates. Tubes containing dissected tissues were kept in liquid nitrogen until pooling was completed. RNA extraction was performed using the innuPREP RNA Mini Kit (Analytik Jena, Jena, Germany) with an extra homogenization step at the beginning of the protocol: 450 µl of Lysis Solution RL, 6 µL of lysozyme solution (50 mg/mL) and bashing beads were added to the sample 2 mL PP micro tube before homogenization with a 2010 Geno/Grinder® device (Spex®SamplePrep, Metuchen, NJ, USA) during 1 min at 1250 rpm. Samples were incubated for 3 min at RT and then homogenized again under abovementioned conditions. Rest of RNA extraction was conducted according to manufacturer's instructions. RNA quality and purity was tested with a Experion™ RNA chip (Bio-Rad, Hercules, CA, USA) on a 2100 Bioanalyzer (Agilent, Santa Clara, USA) following manufacturer's protocols.

2.2. RNA sequencing and de novo transcriptome assembly

RNA samples were sent on dry ice to the Max Planck Genome Centre in Cologne (Germany) for sequencing on an Illumina HiSeq2500 platform. Prior to library preparation ribosomal RNA was depleted from the samples. Random primed cDNA libraries were prepared with a TruSeq RNA Sample Prep kit (Illumina, San Diego, CA, USA) and sequenced, yielding 8 million (4 gigabases, each hindgut wall library), and 12 million (6 gigabases, pocket library) of 250 bp paired-end reads. Library quality control, adaptor trimming and *de novo* transcriptome assembly were carried out

using CLC Genomics Workbench (Qiagen, Redwood City, CA, USA). A reference transcriptome was generated by a combined assembly of both hindgut wall and pocket reads using CLC Genomics Workbench v10.1 software with standard settings and an additional CLC-based assembly with the following parameters: word size = 64, bubble size = 300; minimum contig length = 350 bp; nucleotide mismatch cost = 2; insertion = deletion costs = 2; length fraction = 0.7; similarity = 0.9. Conflicts among individual bases were resolved in all assemblies resolved by voting for the base with the highest frequency. Contigs shorter than 350 bp were removed from the final analysis. The two assemblies were compared according to quality criteria such as N50 contig size, total number of contigs and the number of sequence reads not included in the contig assembly. For each assembly, the 50 largest contigs were manually inspected for chimeric sequences. The presumed optimal consensus transcriptome was selected from the different assemblies based on the highest N50 contig size, lowest total number of contigs and lowest number of chimeric sequences in the 50 largest contigs. The parameters of contig assembly were the following: word size = automatic, bubble size = 300; minimum contig length = 350 bp; nucleotide mismatch cost = 2; insertion = deletion costs = 2; length fraction = 0.7; similarity = 0.9. The resulting final *de novo* reference assembly (backbone) contained 305,905 contigs (minimum contig size = 350 bp) with an N50 contig size of 1,037 bp and a maximum contig length of 32,815 bp. The raw sequencing data was deposited in the NCBI Sequence Read Archive under the accession numbers SRR6123467 (hindgut wall replicate 1 sequencing run 1), SRR6123468 (hindgut wall replicate 1 sequencing run 2), SRR6123465 (hindgut wall replicate 2), SRR6123466 (hindgut wall replicate 3) and SRR6123464 (pockets).

2.3. Annotation and differential expression analysis

BLAST searches were conducted on a local computer cluster against the NCBI nr database with BLASTx (e-value cut-off $1e-1$) and saved as .xml files. Further transcriptome annotation was carried out with Blast2GO PRO (Conesa & Stefan 2008) using Gene Ontology terms and EC numbers. Digital gene expression analysis was carried out using CLC Genomics workbench v10.1 to generate BAM mapping files, and finally by counting the sequences to estimate expression levels, using previously described parameters for read mapping and normalization. For read mapping, we used the following parameters: read assignment quality options required at least 60

% of the bases matching (the amount of mappable sequence as a criterion for inclusion) the reference with at least 94% identity (minimum similarity fraction, defining the degree of preciseness required) within each read to be assigned to a specific contig; mismatch cost = 2; insertion = deletion cost = 3; maximum number of hits for a read (reads matching a greater number of distinct places than this number are excluded) = 10; n-mer repeat settings were automatically determined and other settings were not changed. RPKMs (reads per kilo base per million mapped reads) (Mortazavi et al. 2008) were calculated from the raw read count to normalize the expression of each contig. Contigs with RPKM < 1 in all libraries were considered not transcriptionally active and were discarded from further analysis. Differential gene expression analysis between pocket and hindgut wall libraries was carried out using the R package edgeR (Robinson et al. 2010), using TMM (trimmed mean of M-values) normalized values of read CPM (counts per million) per contig (Oshlack & Wakefield 2009). Significance threshold of FDR (false discovery rate) corrected p value was set to 0.05. Since lack of replicates in one of the samples analyzed (pocket library) reduces statistical reliability in differential expression testing, the following components were included in the analysis: a) before TMM normalization, contigs with ≤ 1 CPM (counts per million) in all libraries were filtered out, regardless of their RPKM value. Therefore, genes with very low counts that may influence the accuracy of multiple testing were removed while maintaining tissue-specific expression patterns. b) Besides a FDR value ≤ 0.05 , the requirement of fold change (FC) value > 10 or < -10 was imposed to define differentially expressed genes, thus discarding genes with low FDR value due to consistency among hindgut wall libraries but with little difference between hindgut wall and pockets. Additionally, in order to discriminate the major processes ongoing in pocket and hindgut wall, highly and differentially expressed (HDE) genes were selected by defining high expression as any RPKM 10 times above the library mean. Hindgut wall contigs were considered highly expressed when meeting such requirement in all three hindgut wall libraries. The web-based KEGG pathway mapping tool Search Pathway (http://www.genome.jp/kegg/tool/map_pathway1.html) was used to functionally annotate differentially expressed (DE) enzymes with associated EC code and to reconstruct metabolic pathways according to the Kyoto Encyclopedia of Genes and Genomes (KEGG).

3. Results

3.1. de novo transcriptome assembly and annotation

A total of four sequencing libraries (three for hindgut wall, one for pocket) were produced from larval *M. hippocastani* RNA. Each hindgut wall library contained, approximately, 8 million reads (4 gigabases). Each pocket library contained about 12 million reads (6 gigabases). Reads were assembled into a total of 305.905 contigs (Tables 1 and 2). After read alignment and RPKM normalization, contigs with very low expression (RPKM < 1 in all libraries) were discarded. The taxonomic assignment of contigs performed via Blast2GO PRO revealed that the most common eukaryotic alignment was the insect *Tribolium castaneum* (fam. Tenebrionidae) in both hindgut wall and pocket libraries, while the most common prokaryotic alignments were *Achromobacter* sp. (fam. Alcaligenaceae) in the pocket library and *Clostridium* sp. (fam. Clostridiaceae) in hindgut wall libraries (Fig. 1).

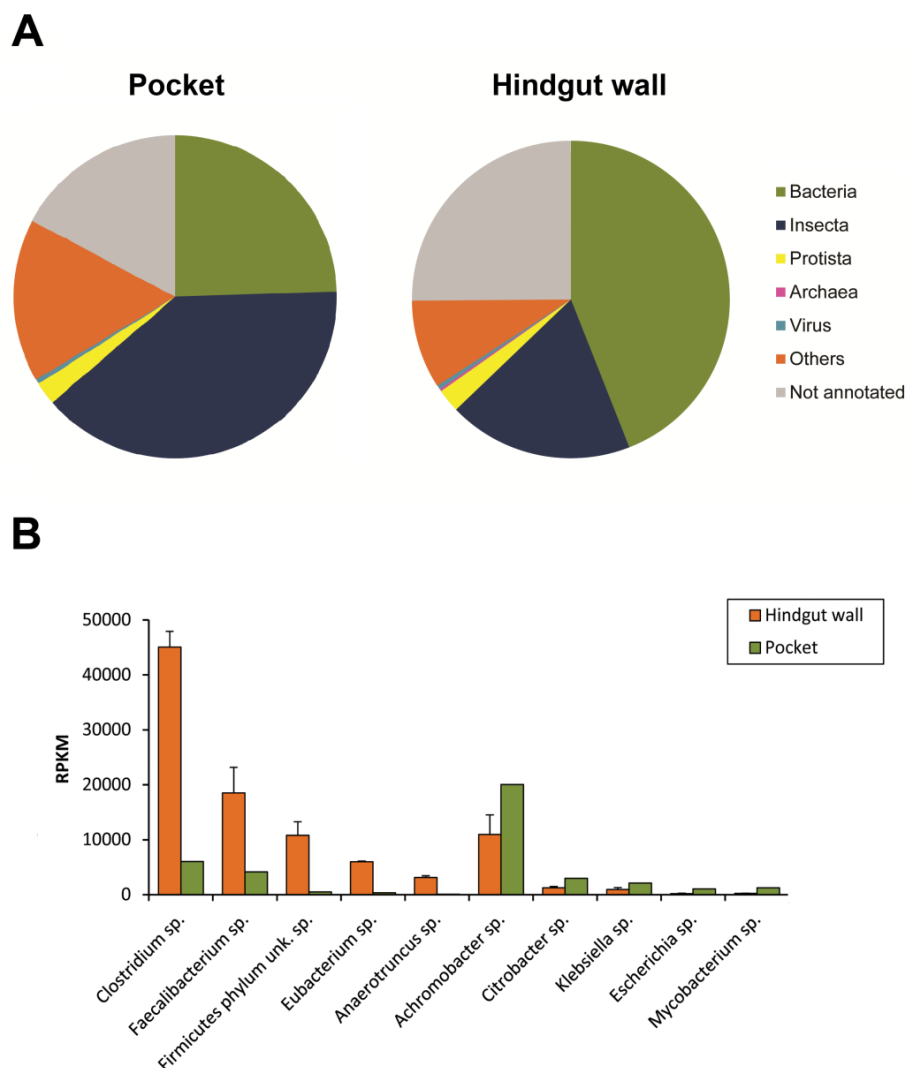


Figure 1 (previous page). Taxonomic distribution of annotated contigs. After RPKM normalization, contigs with RPKM < 1 in all libraries were discarded. A) Pie charts showing the distribution of contigs in taxonomic groups. Hindgut wall chart represents the mean of the 3 hindgut wall libraries. B) Bar chart showing the five bacterial genera with higher RPKM in hindgut wall libraries (left half) and pocket library (right half). Orange bars represent the mean of all three hindgut wall libraries. Error bars represent standard deviation.

Table 1. Contig length measurements

	With scaffolded regions	Without scaffolded regions
N75	636	631
N50	1,037	1,026
N25	1,984	1,948
Minimum	350	120
Maximum	32,815	32,815
Average	927	914
Count	305,905	309,982

Table 2. Transcriptome assembly statistics

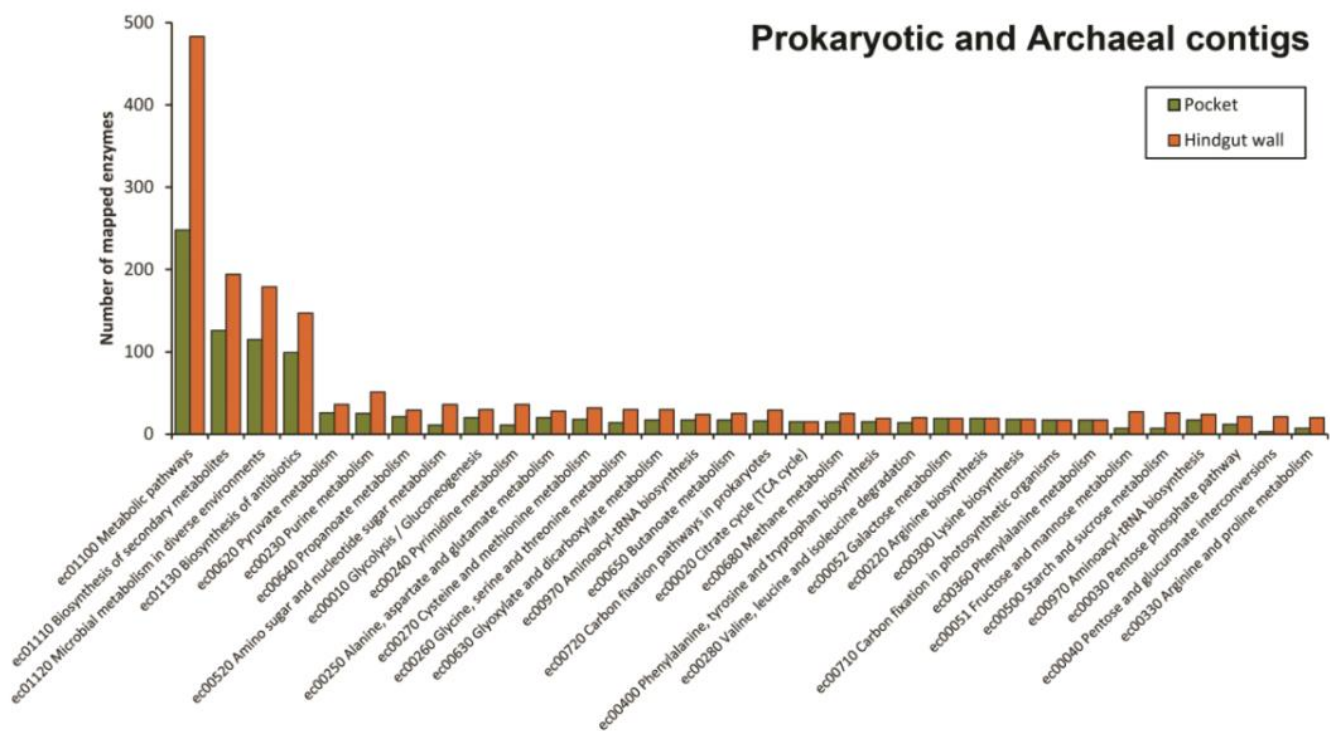
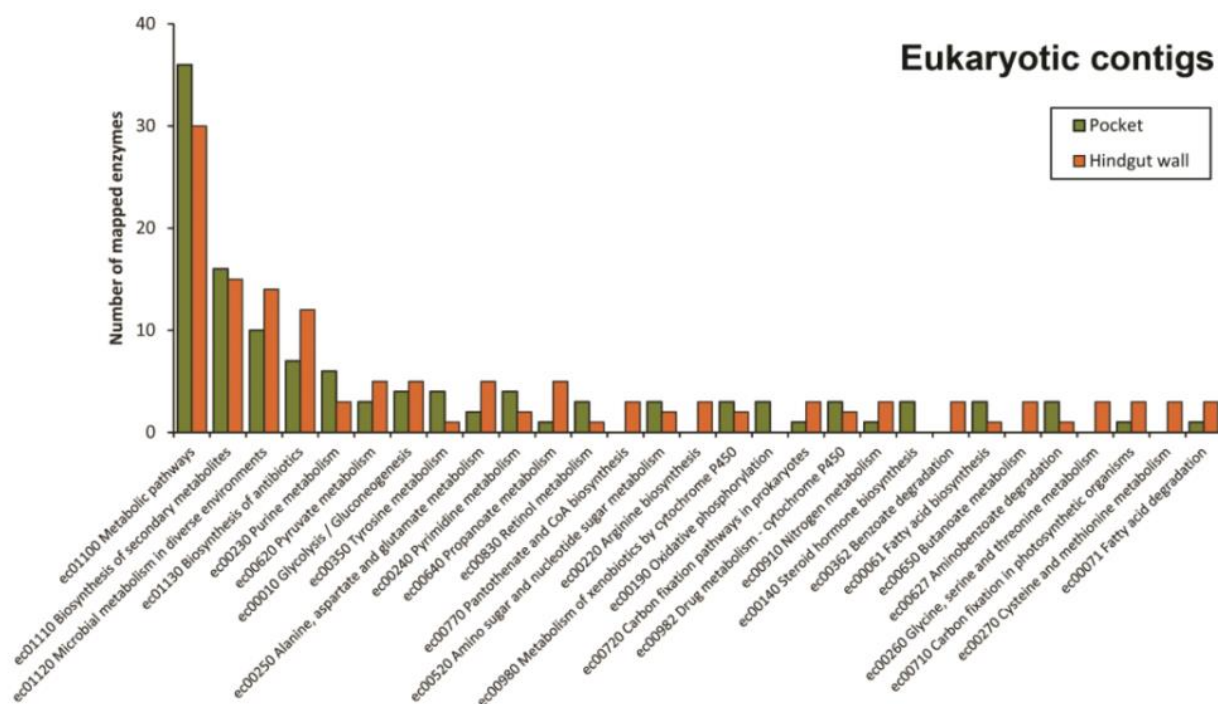
	Count	Average length	Total bases
Reads	64,067,071	227.73	14,590,016,540
Matched	51,171,136	227.1	11,620,962,785
Not matched	12,895,935	230.23	2,969,053,755
Reads in pairs	43,000,840	273.87	283,424,930
Broken paired reads	8,102,050	244.72	
Contigs	305,905	926	

3.2. Differential expression analysis

After filtering contigs with CPM ≤ 1 in all libraries, a total of 105,693 contigs were considered for differential expression analysis. Of these, edgeR identified 23,856 differentially expressed (DE) contigs (FDR <0.05 , FC > 10 or < -10) between hindgut wall and pocket libraries. Differentially expressed KEGG pathways in pocket and hindgut wall were reconstructed by the web-based tool Search Pathway using as input contig-associated EC codes. A summary of the top DE KEGG pathways is shown in Fig. 2. A comprehensive list of all DE contigs and the KEGG pathways to which they map is shown in Supplementary Table 1. Additionally, any contig with RPKM ten times above the mean of the library was considered as highly expressed. By applying this rationale, 1,106 contigs were found to be highly expressed in all three hindgut libraries, while 2,657 contigs were found to be highly expressed in the pocket library. By combining these statistics, 495 contigs were found to be highly and differentially expressed (HDE) in the hindgut wall compared to pocket, and 220 were found to be HDE in the pocket compared to hindgut wall (Fig. 3). The top identified HDE contigs in pocket and hindgut libraries are shown in Table 3. A comprehensive list of all HDE contigs and the KEGG pathways to which they map is shown in Table S2, and can be summarized as follows: unique HDE genes in pocket library were, among others, 20 contigs that mapped to insect cuticular proteins and precursors, 15 to eukaryotic transposases and transposable elements, 9 to insect chitinases and chitin-binding proteins, 8 to bacterial porins, , 6 to eukaryotic reverse transcriptases, 6 to insect CD109 antigen, 4 to putative eukaryotic growth factors (3 of which are described as multiple epidermal growth factor-like domains), 4 to putative insect sulfate transmembrane transporters (sodium-independent sulfate anion transporter-like), 4 to eukaryotic histone-lysine N-methyltransferase and 3 histone-lysine N-methyltransferase-like, 3 to insect glucose dehydrogenases, 3 to insect Osiris proteins and precursors, 2 to toxic bacterial proteins of Hok family, 2 to insect yellow protein, 2 to proteins involved in insect response to stimulus (cuticular analogous to peritrophins 1-J isoform X2 and odorant binding 4), 2 to insect peroxidases and one to a bacterial peroxiredoxin, an insect serine protease H164 (EC:3.4.21), a bacterial lipoyl synthase (EC:2.8.1.8), an insect phenoloxidase subunit A3, a bacterial genetic competence-related protein, an insect arylphorin subunit alpha, a bacterial Tu translation elongation factor, an insect C-type lectin and an insect gram-negative bacteria binding protein. Unique HDE genes in hindgut libraries were, among others, 10 contigs

that mapped to bacterial flagellar proteins, 10 to bacterial carbohydrate ABC transporters and transcriptional regulators, 8 to bacterial proteases (6 of which are serine-type (EC:3.4.21; EC:3.4.17; EC:3.4.16)) and regulators, 7 to bacterial sporulation-related proteins, 6 to bacterial chemotaxis-related proteins, 5 to bacterial amino acid transmembrane transporters, 5 to bacterial proteins of pF06949 family, 4 to bacterial collagen repeats, 3 to bacterial aldehyde oxidoreductases (EC:1.2.99.7), 3 to bacterial aldehyde ferredoxin oxidoreductases (EC:1.2.7.5), 3 to insect cystathionine beta-synthases, 2 to bacterial carbamate kinases (EC:2.7.2.2), 2 to bacterial C4-dicarboxylate ABC transporters, 2 to bacterial iron transmembrane transporters, 2 to bacterial alkaline-shock proteins, 2 to bacterial beta lactamases, 2 to bacterial pyrophosphatases, 2 to bacterial addiction module toxin system, one to an insect aquaporin AQPcic, a bacterial ferredoxin, a bacterial manganese-containing catalase, a negative regulator of bacterial genetic competence, a bacterial citrate transporter and an insect histone deacetylase complex subunit SAP18.

Figure 2 (next page). Top DE KEGG pathways in pocket and hindgut wall tissues sorted according to contig taxonomic assignment.



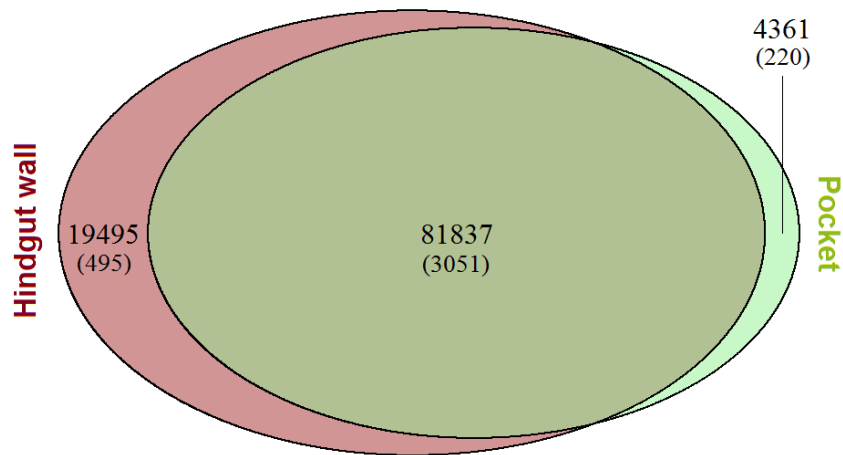


Figure 3. Venn diagram showing the distribution of contig expression between libraries. Brown area shows DE sequences in hindgut wall, green area shows DE sequences in pocket. Overlapping area shows sequences equally expressed in both tissues. Number of sequences (in parentheses, highly expressed sequences) contained in each area are given.

Table 3 (next page). Top identified HDE contigs in hindgut wall and pocket, sorted in decreasing order of fold change. Only contigs highly (RPKM > 10x library mean) and differentially expressed (FDR < 0.05, FC > 10 or < -10), annotated to proteins with known function are shown. In case of hindgut wall, log₂ RPKM column shows mean of log₂ transformed RPKMs of the three libraries. Fold change column shows the RPKM fold change compared to the other tissue.

HDE contigs in hindgut wall					
Contig name	Description (BLAST2GO)	Organism	Log ₂ RPKM	Fold change	Biological role
Mhippo_Comb_C34879	alkaline-shock	<i>Pelotomaculum thermopropionicum</i>	5.78	26,035	Alkaline pH tolerance
Mhippo_Comb_C28060	phage infection	<i>Rattus norvegicus</i>	5.46	20,856	Membrane spanning protein, required for phage infection
Mhippo_Comb_C224	3D domain-containing	<i>Desulfosporosinus</i> sp.	5.26	18,162	Hydrolyzing O-glycosyl compounds, peptidoglycan turnover
Mhippo_Comb_C14542	tRNA delta(2)-isopentenylpyrophosphate transferase	<i>Carboxydothemus ferrireducens</i>	5.07	15,957	tRNA dimethylallyltransferase activity
Mhippo_Comb_C11085	transposase	<i>Pseudoalteromonas citrea</i>	4.93	14,424	Transposition, DNA integration
Mhippo_Comb_C5942	phenolic acid decarboxylase subunit C	<i>Blautia producta</i>	6.13	1,308	Aromatic compound catabolic process, response to toxic substance
Mhippo_Comb_C5739	flagellin domain	<i>Paenibacillus sabiniae</i>	7.98	953	Bacterial-type flagellum filament
Mhippo_Comb_C6635	Cytochrome b561	<i>Tribolium castaneum</i>	5.40	480	Integral component of lysosomal membrane, transmembrane electron transport
Mhippo_Comb_C3434	Appr-1-p processing	<i>Dehalobacter</i> sp.	5.81	475	RNA-directed RNA polymerase activity, tRNA splicing
Mhippo_Comb_C7994	amino acid carrier	<i>Firmicutes bacterium</i>	5.76	448	Transmembrane transport, alanine:sodium symporter activity
Mhippo_Comb_C24847	alpha beta hydrolase	<i>Clostridium josui</i>	5.04	411	Hydrolytic activity (protease, lipase, peroxidase, esterase, epoxide hydrolase or dehalogenases)
Mhippo_Comb_C8677	uroporphyrinogen decarboxylase	<i>Treponema primitia</i>	5.44	408	Porphyrin-containing compound biosynthetic process
Mhippo_Comb_C4248	manganese containing catalase	<i>Anaerotruncus</i> sp.	5.67	406	Non-haem Mn-catalase, conversion of hydrogen peroxide

					to water and molecular oxygen
Mhippo_Comb_C23143	Na ⁺ /H ⁺ antiporter	<i>Eubacterium</i> sp.	5.00	381	Transmembrane transport, antiporter activity
HDE contigs in hindgut wall (continued)					
Contig name	Description (BLAST2GO)	Organism	Log ₂ RPKM	Fold change	Biological role
Mhippo_Comb_C2271	O-antigen polymerase	<i>Pelotomaculum thermopropionicum</i>	5.63	376	Lyase activity, lipopolysaccharide biosynthetic process
HDE contigs in pocket					
Mhippo_Comb_C263745	flexible cuticle 12	<i>Tribolium castaneum</i>	7.14	66,743	Structural constituent of cuticle
Mhippo_Comb_C276859	Hok Gef family	<i>Escherichia coli</i>	6.36	38,863	Bacterial toxin – antitoxin system
Mhippo_Comb_C205421	outer membrane porin	<i>Achromobacter piechaudii</i>	5.73	25,264	Porin activity, transmembrane transport
Mhippo_Comb_C28529	peptide of tmRNA	<i>Gordonia rhizosphaera</i>	5.63	23,461	Ribosomal rescue system
Mhippo_Comb_C276989	IS116 IS110 IS902 family	<i>Rhodococcus erythropolis</i>	5.60	23,020	Transposase activity
Mhippo_Comb_C205423	outer membrane porin	<i>Achromobacter piechaudii</i>	5.59	22,813	Porin activity, transmembrane transport
Mhippo_Comb_C277394	RNA-directed DNA polymerase	<i>Escherichia coli</i>	5.44	20,614	RNA-dependent DNA biosynthetic process
Mhippo_Comb_C227062	pseudouridine-5 - monophosphatase	<i>Tribolium castaneum</i>	5.26	18,251	Hydrolase activity, structural RNA degradation
Mhippo_Comb_C276863	S-(hydroxymethyl)glutathione dehydrogenase	<i>Escherichia fergusonii</i>	5.13	16,618	Alcohol dehydrogenase (NAD) activity, formaldehyde detoxification
Mhippo_Comb_C205422	outer membrane porin	<i>Achromobacter arsenitoxydans</i>	4.98	14,978	Porin activity, transmembrane transport
Mhippo_Comb_C276858	TPA: cuticle	<i>Coptotermes formosanus</i>	7.95	14,179	Structural constituent of cuticle
Mhippo_Comb_C28536	peptide of tmRNA	<i>Mycobacterium kansasii</i>	7.39	13,616	Ribosomal rescue system

Mhippo_Comb_C101570	50S ribosomal L21	<i>Escherichia coli</i>	4.79	13,088	Cytosolic large ribosomal subunit protein
Mhippo_Comb_C270914	outer membrane porin	<i>Achromobacter piechaudii</i>	4.75	12,781	Porin activity, transmembrane transport
Mhippo_Comb_C199511	outer membrane A	<i>Achromobacter piechaudii</i>	6.89	12,760	Porin activity, ion transmembrane transport

4. Discussion

4.1. Taxonomic distribution of reads

As shown in Fig. 1A, the majority of annotated reads in pocket and hindgut wall fell into insect or bacterial taxa. Protist-annotated reads were also detectable in both tissues, although they are likely to originate from transient or parasitic microorganisms or from wrongly annotated insect-derived RNAs, as their symbiotic association with terrestrial coleopterans is exceptional (Tanahashi et al. 2017). Archaeal reads were also detected in both tissues, although in negligible amounts (less than 0.5%) which is in concordance with previous observations in *Melolontha melolontha* (Egert et al. 2005). Finally, around 10 and 20% of reads in hindgut wall and pocket tissue, respectively, mapped to non-insect eukaryotic organisms. Considering the lack of reference genome for *Melolontha hippocastani* and the inherent alignment uncertainty, all eukaryotic-mapped reads were assumed to be of host origin. Bacteria-mapped contigs in pocket and hindgut wall libraries fell into different genera depending on the tissue (Fig. 1B). In hindgut wall, reads mapped mostly to anaerobic gram-positive bacteria, while in the pocket they mapped mostly to aerobic or facultative anaerobic gram-negative bacteria. Differences in community composition between pockets and hindgut wall were previously observed; however, the Clostridiales and the Clostridiaceae, the order and family with highest number of mapped reads in the hindgut wall, were only poorly detected in the hindgut wall of second-instar larvae (Alonso-Pernas et al. 2017). They were, nevertheless, very abundant in third-instar larvae. This suggests that, although the larvae used in the present study had the body size corresponding to second-instar larvae, the composition of their gut microbiome was already drifting to third-instar's. Following this hypothesis, negligible number of reads mapped to the

Micrococcaceae family in pocket libraries, although they were very abundant in second-instar larvae (Alonso-Pernas et al. 2017). This suggests that this bacterial family is either dormant or ultimately outgrown by the Alcaligenaceae family, concretely by the genus *Achromobacter*.

4.2. Highly and differentially expressed (HDE) genes in hindgut wall

The disparate annotation of highly and differentially expressed (HDE) contigs in pockets and hindgut wall is indicative of functional differences between the two tissues. In hindgut wall, a bacterial alkaline shock gene showed the highest fold change compared to pockets, which is in line with the alkaline pH (around 8) of this section of the gut (Egert et al. 2005) (Table 3). Also, many bacterial genes related to sporulation were HDE in hindgut wall (Table S2), which is not surprising due to the high abundance of Clostridial RNAs in the libraries. Sporulation-related genes are commonly expressed in gram-positive symbionts of insects and other organisms (Margulis et al. 1998; Paredes-Sabja et al. 2011) as a response to biofilm formation and high bacterial density (Dapa et al. 2013; McKenney et al. 2013). Other HDE genes suggesting ongoing bacterial colonization of the hindgut wall and microbe-microbe interactions were flagellar and chemotaxis-related contigs (Rawls et al. 2007; Wang & Wood 2011; Stephens et al. 2015), the toxin-antioxin system addiction module (Wang & Wood 2011), bacterial collagen repeats that might anchor bacteria in the gut wall and participate in host-microbe interactions (Yu et al. 2014) and beta-lactamases that may protect bacteria from antibiotics secreted by neighbouring microorganisms (Chen et al. 2017) or from toxic dietary compounds (Allen et al. 2009). All these HDE genes related to symbiotic colonization and interactions indicate that the hindgut wall community of *M. hippocastani* is not stable but subjected to continuous changes, which agrees with previous observations (Alonso-Pernas et al. 2017). HDE bacterial aldehyde oxidoreductases might neutralize antimicrobial agents contained in the diet (Correia et al. 2016). HDE bacterial transcripts for iron transporters were possibly a consequence of limitation of this element, whose availability in the gut is tightly controlled by the insect (Nichol et al. 2002). Congruent with iron scarcity is the presence of a HDE bacterial manganese-containing catalase transcript, a non-heme catalase expressed under conditions of microaerophilic oxidative stress and iron depletion (Whittaker 2012). Many ABC transporters and transcriptional regulators involved in transmembrane transport of carbohydrates were also HDE by hindgut wall bacteria, reflecting their role, along with insect

secreted enzymes, in the digestion of recalcitrant polysaccharides, which are degraded extracellularly and the resulting soluble saccharides are internalized for further processing (Martin 1983; Artzi et al. 2017). HDE bacterial protease transcripts (mostly serine proteases) and the alkaline pH occurring in the hindgut, which might contribute to the solubilization of dietary proteins, supports the digestive role of the resident community (Vinokurov et al. 2006). HDE bacterial amino acid transmembrane transporters (most of which were identified as sodium:solute symporters) suggest occurrence of free amino acids, probably due to ongoing proteolysis or *de novo* synthesis from host nitrogenous waste (Alonso-Pernas et al. 2017). Differentially expressed (DE) urease (EC:3.5.1.5) and carbamate kinase (EC:2.7.2.2.) in hindgut wall support presence of ammonia in hindgut (Table S1), as the former produces it by hydrolyzing urea and the latter uses it as substrate. HDE insect cystathionine beta-synthases (EC:4.2.1.22), which plays a role in the synthesis of the essential amino acid cysteine, might be related to the production of cysteine-rich proteins that regulate symbiotic population (Futahashi et al. 2013). HDE insect aquaporin might be involved in water absorption from feces, a well-reported hindgut function (Campbell et al. 2008; Chapman 2013). Finally, HDE transcripts for insect histone deacetylase complex may suggest repression of host genes, which is in line with the low number of detected host RNAs (Fig. 1) (Martin & Zhang 2005; Haberland et al. 2009). Alternatively, insufficient sequencing depth might have caused masking of low-abundant insect RNAs by highly-abundant bacterial RNAs, in which case some of the host's functions may have gone unnoticed.

4.3. Highly and differentially expressed (HDE) genes in pocket

An obvious increase in host-related HDE transcripts was observed in the pocket library compared to hindgut wall, possibly due to histone-lysine N-methyltransferase mediated upregulation (Table S2) (Martin & Zhang 2005). Most of the host HDE transcripts in pocket were cuticle-related. Insect cuticle constitutes a defensive barrier in the external body as well as in foregut and hindgut (Chapman 2013). Among pocket HDE transcripts both insect structural constituents of cuticle and chitinases are found, which suggest simultaneous cuticle synthesis and degradation. This may indicate either cuticular renovation, as the old cuticle must be degraded before being replaced by the new one (Willis et al. 2012; Chapman 2013), or cuticle thickening as a response to external challenge (Li & Denlinger 2009; Bascañán-García et al. 2010). The latter possibility is supported

by early observations of the tissue (Wildbolz 1954) and by the presence of HDE eukaryotic epidermal growth factors. It has been shown that other functions of these proteins are protecting the integrity of the intestinal barrier, reducing colonization by pathogens and attenuating the epithelial inflammatory response (Tang et al. 2016; Kim et al. 2016). Further hints pointing towards active host immune response in pocket tissue are the HDE insect transcripts for gram-negative bacteria binding protein, C-type lectin and HDE CD109 antigen-like proteins which might be involved in pathogen recognition and activation of host defenses (Kim et al. 2000; Zelensky & Gready 2005; Warr et al. 2008; Yazzie et al. 2015), for a phenoloxidase domain and a serine protease, both possibly involved in melanin synthesis (Ashida & Breyt 1995) and for a yellow protein, which positions the melanin pigment in the cuticle (Willis et al. 2012) among other defensive roles (Gretschler et al. 2016). Melanin synthesis in pockets is further supported by the fact that in some individuals the pockets appear stained black (Wildbolz 1954). Functions of this pigment are mostly defensive, i.e. wound healing or pathogen encapsulation (Nakhleh et al. 2017). This points towards a tight regulation of the pocket population by the innate immunity of the host which, in combination with the oxidative stress suggested by HDE insect peroxidases and bacterial peroxiredoxins (Lushchak 2010), may have the purpose of preventing hindgut anaerobic bacteria from colonizing the pockets. Additionally, host proteases may directly control symbiotic titer (Byeon et al. 2015) and yellow proteins appear to be up- and downregulated depending on seasonal conditions (Daniels et al. 2014; Vilcinskas & Vogel 2016) raising the question of whether seasonal changes have an influence in pocket gene expression. The high number of HDE contigs that mapped eukaryotic transposases, transposable elements and reverse transcriptases might be consequence of oxidative stress (Giorgi et al. 2011) or of host perception of pocket colonization as an infection event (Mhiri et al. 1997). In view of the high number of pocket reads assigned to *Achromobacter* sp. (Fig 1B), it is tempting to assume that host immune system is less detrimental to this genus than to the others. The mechanisms underlying this possible selectiveness are to be studied, but is it plausible that HDE insect C-type lectin is involved, since lectins can be used by certain bacteria to evade host anti-microbial peptides (Pang et al. 2016). Also, the presence of symbiotic *Achromobacter* in the pocket might stimulate host immune system as reported in other systems, thus conferring resistance against pathogens (Kim et al. 2016). HDE contigs mapping to porins were also very abundant in the pocket. These passive transporters are the most abundant proteins in the outer membrane of gram-negative bacteria and their high expression in pockets

suggests an intense exchange of nutrients and ions between symbionts and environment (Galdiero et al. 2012). Interestingly, porin-assigned reads aligned mainly to *Achromobacter* sp., which supports a significant role for this bacterium in the pocket. Porins are crucial for the symbiotic colonization of the squid light-emitting organ (Aeckersberg et al. 2001), for the infection of insect hemocoel by an entomopathogen and possibly for the colonization of the gut of its nematode vector as well (Van Der Hoeven & Forst 2009). Moreover, porins have been related to invasion of non-insect tissues and manipulation of host defenses by pathogenic bacteria (Provenzano & Klose 2000; Duperthuy et al. 2011). Collectively, these observations suggest that *Achromobacter* sp. is engaged in a porin-mediated colonization process that may contribute to the triggering of insect immune response. Unfortunately, considering the current data any speculation on the role of the pockets as a whole is still far-fetched. Nevertheless, previous reports (Alonso-Pernas et al. 2017) and the presence of HDE sulfate transmembrane transporters suggests that the pockets might participate in nutrient provisioning. It is also noteworthy the pocket HDE contigs mapping to the Osiris gene family. These proteins are exclusive of insects, and although their role is unclear, they show expression peaks at specific life stages (embryo, second instar larvae and pupae) and might be involved in insect development (Shah et al. 2012). Also the HDE insect arylphorin subunit alpha and glucose dehydrogenases point towards the pockets being involved in host development. Arylphorins are ubiquitous proteins in *Melolontha* sp. and other insects that act as amino acid storage proteins. Their concentration in hemolymph increases during each larval stage to quickly drop at each molt, as their amino acid components are used for the synthesis of body tissues (Delobel et al. 1993; Chapman 2013). Expression of glucose dehydrogenases is highly correlated with that of 20-hydroxyecdysone, a major insect molting hormone, and is increased during metamorphosis (Cox-Foster et al. 1990). HDE arylphorins, Osiris family proteins and glucose dehydrogenases make tempting to speculate that the role of the pockets and, perhaps, their bacterial symbionts, might go beyond insect immunity and nutrition and embrace other aspects of physiology such as development, as it happens in the *Riptortus-Burkholderia* system (Lee et al. 2017) or *Aedes* mosquitoes (Coon et al. 2017).

4.4. Differentially expressed (DE) metabolic pathways in hindgut wall

Taking into consideration all differentially expressed (DE) enzymes (FDR<0.05, FC>10 or <-10) with associated EC codes, DE KEGG pathways in pocket and hindgut libraries were reconstructed. The taxonomic assignment of the BLAST top hit allowed separation in host- and symbiont-derived pathways (Fig. 2, Table S1). Symbiotic DE enzymes belonging to the KEGG pathways purine and pyrimidine metabolism, and aminoacyl tRNA biosynthesis were DE in both pocket and hindgut wall libraries. This indicates ongoing DNA, RNA and protein synthesis, thus high symbiont growth and activity in both tissues. Within these pathways, allantoinases (EC 3.5.2.5), allophanate hydrolases (EC 3.5.1.54) and ureases (EC 3.5.1.5) were DE only in hindgut wall, suggesting degradation of insect nitrogenous waste, that is, uric acid and urea, to ammonia, as part of a nitrogen recycling mechanism (Alonso-Pernas et al. 2017). Proximity of the outlet of the Malpighian tubules, located in the midgut-hindgut junction, to the site of sampling of hindgut wall tissue supports this hypothesis, as Malpighian tubules collect hemolymph waste and pour it into the hindgut (Shelomi 2017). On the contrary, the glutamine synthetase pathway was DE by both pocket and hindgut wall symbionts, suggesting that pocket bacteria utilize hindgut-produced ammonia for the synthesis of amino acids. Symbiotic chitinases were DE only in hindgut wall, probably reflecting the usage of this cuticular polysaccharide by hindgut bacteria as nitrogen and carbon source, thus incidentally contributing in maintaining the optimal thickness of the intima layer for proper diffusion of nutrients (Indiragandhi et al. 2007). In the hindgut wall more DE symbiotic enzymes involved in KEGG pathways related to carbon metabolism compared to pockets were found (fructose and mannose metabolism and starch and sucrose metabolism). This is in line with the digestive role of hindgut bacteria discussed above. Cellulases were only DE by hindgut wall bacteria. Taken together, these observations indicate that pocket bacteria may not be directly involved in digestion of insect food. Symbiotic enzymes belonging to carbon fixation KEGG pathways such as the reductive TCA cycle and the Wood-Ljungdahl pathway were DE in hindgut wall. They may contribute to the supply of acetyl-CoA needed for bacterial metabolism, but also to the production of acetate which might be taken up by the insect. Bacterial acetogenesis is well documented within the digestive tract of wood-feeding termites and roaches (Warnecke et al. 2007; Zhang et al. 2010) as well as in other insects with disparate diets (Matson et al. 2011). High acetate concentration in the gut fluid of *Melolontha melolontha* and hydrogen accumulation in the midgut, but not in the hindgut, is in line with ongoing acetogenesis in the hindgut, as this process uses hydrogen as electron donor, thus preventing its accumulation (Egert et al. 2005).

Hindgut wall symbionts might also provide to the host, nutrients such as niacin, pantothenic acid, biotin and pyridoxine, as suggested, respectively, by the DE KEGG pathways nicotinate and nicotinamide metabolism, pantothenate and CoA biosynthesis, biotin metabolism and vitamin B6 metabolism (Gilmour 1961; Cohen 2015). Methanogenesis might happen in hindgut wall as well, as coenzyme-B sulfoethylthiotransferase (EC 2.8.4.1) is DE in libraries of this tissue. However, this is likely to be a minor process carried out by the small archaeal population (Fig. 1, Table S1) (Egert et al. 2005). Finally, the pathways yielding geranyl pyrophosphate, geranylgeranyl pyrophosphate and farnesyl pyrophosphate from pyruvate (within the KEGG pathway terpenoid backbone biosynthesis) were DE by hindgut wall symbionts, suggesting production of these precursors of monoterpenoids (geranyl-PP), diterpenoids and carotenoids (geranylgeranyl-PP) and sesquiterpenoids (farnesyl-PP). Terpenoids are commonly used by plants as defense against herbivores and many of them have antimicrobial properties (Mithöfer & Boland 2012). Furthermore, it has been shown that many bacteria are able to synthesize them (Yamada et al. 2014). However, to our knowledge there is no report on insect gut symbionts producing such compounds. A more plausible hypothesis is that geranylgeranyl-PP or farnesyl-PP might be taken up by the host and used as precursors for hormone synthesis. Geranylgeranyl-PP might be used to produce carotenoids as well. Carotenoids play a role in multiple physiological functions of the insect host (Heath et al. 2012; Sloan & Moran 2012). Host KEGG pathways having more DE enzymes in the hindgut wall, as compared to pocket were glycolysis/gluconeogenesis, pyruvate metabolism and pantothenate and CoA biosynthesis. Expression of these pathways, related to energy, CoA and acetyl-CoA production, is likely linked to the absorption of acetate and other short chain fatty acids through the hindgut epithelium (Bayon & Mathelin 1980; Terra & Ferreira 2009). Fatty acid biosynthesis KEGG pathways were DE as well, possibly related to high acetyl-CoA production rate in hindgut epithelial cells. Also KEGG pathways leading to synthesis of amino acids were DE in hindgut wall libraries (alanine, aspartate and glutamate metabolism, arginine biosynthesis and cysteine and methionine metabolism). This might be consequence of high availability of free ammonia and/or glutamate due to the degradation of insect waste nitrogen by hindgut wall bacteria (Alonso-Pernas et al. 2017). The KEGG pathway biosynthesis of antibiotics also showed higher number of DE mapped enzymes in hindgut wall, reflecting a possible expression of antimicrobial compounds in order to keep the symbiotic community under control (Garcia et al. 2010). However, the number of annotated contigs with assigned eukaryotic

taxonomy is surprisingly low in hindgut wall compared to pockets (Table S1). This might be consequence of either low host RNA synthesis or masking of host's transcripts by the highly abundant bacterial RNAs. In the latter case, it is possible that some host functions have been overlooked.

4.5. Differentially expressed (DE) metabolic pathways in pockets

In pocket libraries only a few KEGG pathways have more mapped enzymes of symbiotic origin compared to hindgut wall (Fig. 2). This might be a consequence of a lack of EC codes for many annotated enzymes, or might reflect that in second-instar larvae the pocket community is still ongoing a colonization phase and does not express yet all the particular functional pathways of the tissue. Two of the KEGG pathways DE by pocket symbionts were the TCA cycle and the synthetic pathway of protoheme groups from glutamate and Fe^{2+} (within porphyrin and chlorophyll metabolism). The expression of these pathways, together with HDE bacterial lipoyl synthase, supports occurrence aerobic metabolism and oxidative stress in pocket symbionts (Navari-Izzo et al. 2002; Girvan & Munro 2013). As mentioned before, no cellulases were DE in pocket, but enzymes processing oligo- and monosaccharides such as sucrose, fructose or maltose (within starch and sucrose metabolism) and glucose (within glycolysis/gluconeogenesis) were DE by pocket symbionts, suggesting that they might use the saccharides resulting from degradation of diet polysaccharides, such as cellulose and hemicellulose, in the hindgut. Pocket symbionts also DE more enzymes related to sulfur metabolism compared to hindgut wall. This, together with HDE insect sulfate transmembrane transporters in the pocket, suggest a role related to sulfur provisioning. Sulfur is needed for the synthesis of essential amino acids such as methionine and cysteine, the deficiency of which renders the insect vulnerable to plant defensive protease inhibitors (Broadway & Duffey 1986). In pockets, it appears that the insect differentially expresses all enzymes leading to eumelanin synthesis (within tyrosine metabolism pathway). This is in line with the host defensive deployment in this tissue already suggested by the analysis of HDE contigs (Nakhleh et al. 2017). An unexpected feature of the pockets is that the insect differentially expresses three enzymes of the KEGG pathway steroid hormone biosynthesis, one of which, the cholesterol monooxygenase (EC 1.14.15.6), is exclusive of the pathway. Steroid hormones may regulate the pocket cuticle synthesis as deduced from the HDE of insect cuticular proteins (Karlson

1989). Additionally, steroids might downregulate the synthesis of antimicrobial peptides, thus allowing symbiotic colonization (Gordya et al. 2016). Steroids are also involved in molting and metamorphosis (Niwa & Niwa 2016), supporting the abovementioned hypothesis of the pockets playing a role in host development. Insect chitin degrading and synthesizing enzymes were DE in pocket (within amino sugar and nucleotide sugar metabolism), in line with the pocket cuticular renovation or thickening previously discussed (Willis et al. 2012; Chapman 2013).

4.6. Conclusion

The present comparison of the gene expression of the pockets with the surrounding hindgut wall provided the first functional insight of these enigmatic structures. Combining the data of the present study with previous observations, we conclude that differences between the two tissues include, but may be not limited to: a) The active bacterial population in hindgut wall is mainly composed by anaerobes of the Firmicutes phylum, while those of the pockets is composed of aerobic and facultative anaerobes among which *Achromobacter* sp. is the most significant genus. b) The environmental origin of pocket symbionts is strongly supported by their taxonomy, their aerobic metabolism, their expression of transcripts related pathogen-like tissue colonization and the triggering of insect's innate immunity. c) The environmental conditions between the two tissues appear to be remarkably different, the pockets having a higher oxygen concentration and oxidative stress, and lower bacterial competition and concentration of dietary toxins than the hindgut wall. d) The hindgut wall is likely to be the site of bacterial degradation of dietary recalcitrant polysaccharides and host nitrogenous wastes, while the pockets might play a role in stimulating host immunity, regulating host development and/or micronutrient absorption. e) The pocket bacterial community probably varies across larval stages as does that of the hindgut wall; therefore, its gene expression may change depending on larval maturity. Time-course RNAseq experiments may be useful to clarify this question.

5. Acknowledgements

The authors thank Dr. Heiko Vogel for his support during sequence analysis.

6. References

- Aeckersberg F, Lupp C, Feliciano B et al. (2001) *Vibrio fischeri* outer membrane protein OmpU plays a role in normal symbiotic colonization. *Journal of Bacteriology*, 183(22), pp.6590–6597.
- Allen HK, Cloud-Hansen KA, Wolinski JM et al. (2009) Resident microbiota of the gypsy moth midgut harbors antibiotic resistance determinants. *DNA and cell biology*, 28(3), pp.109–117.
- Alonso-Pernas P, Arias-Cordero E, Novoselov A et al. (2017) Bacterial community and PHB-accumulating bacteria associated with the wall and specialized niches of the hindgut of the forest cockchafer (*Melolontha hippocastani*). *Frontiers in Microbiology*, 8(February), pp.1–13.
- Alonso-Pernas P, Bartram S, Arias-Cordero EM et al. (2017) *In vivo* isotopic labeling of symbiotic bacteria involved in cellulose degradation and nitrogen recycling within the gut of the forest cockchafer (*Melolontha hippocastani*). *Frontiers in Microbiology*, 8(October), p.1970.
- Artzi L, Bayer EA & Moraïs S (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology*, 15(2), pp.83–95.
- Ashida M & Breyt PT (1995) Role of the integument in insect defense: pro-phenol oxidase cascade in the cuticular matrix. *Proceedings of the National Academy of Sciences*, 92(November), pp.10698–10702.
- Bascuñán-García AP, Lara C & Córdoba-Aguilar A (2010) Immune investment impairs growth, female reproduction and survival in the house cricket, *Acheta domesticus*. *Journal of Insect Physiology*, 56(2), pp.204–211.
- Bayon C & Mathelin J (1980) Carbohydrate fermentation and by-product absorption studied with labelled cellulose in *Oryctes nasicornis* larvae (Coleoptera:Scarabaeidae). *Journal of Insect Physiology*, 26, pp.833–840.
- van Borm S, Buschinger A, Boomsma JJ et al. (2002) *Tetraponera* ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. *Proceedings of the Royal Society*, (269), pp.2023–2027.
- Bright M & Bulgheresi S (2010) A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol*, 8(3), pp.218–230.
- Broadway RM & Duffey SS (1986) Plant proteinase inhibitors: mechanism of action and effect on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exigua*. *Journal of Insect Physiology*, 32(10), pp.827–833.
- Byeon JH, Seo ES, Lee JB et al. (2015) A specific cathepsin-L-like protease purified from an insect midgut shows antibacterial activity against gut symbiotic bacteria. *Developmental and Comparative Immunology*, 53(1), pp.79–84.
- Calderón-Cortés N, Quesada M, Watanabe H et al. (2012) Endogenous plant cell wall digestion: a key mechanism in

- insect evolution. *Annual Review of Ecology, Evolution and Systematics*, (43), pp.45–71.
- Campbell EM, Ball A, Hoppler S et al. (2008) Invertebrate aquaporins: a review. *Journal of Comparative Physiology B*, 178(8), pp.935–955.
- Chapman R (2013) *The Insects, structure and function*. S. Simpson & A. Douglas, eds., Cambridge University Press.
- Chen D, Guo R, Xu X et al. (2017) Uncovering the immune responses of *Apis mellifera ligustica* larval gut to *Ascospaera apis* infection utilizing transcriptome sequencing. *Gene*, 621, pp.40–50. A
- Chen S, Blom J & Walker ED (2017) Genomic, physiologic and symbiotic characterization of *Serratia marcescens* strains isolated from the mosquito *Anopheles stephensi*. *Frontiers in Microbiology*, 8(August), p.1483.
- Cohen AC (2015) *Insect Diets: Science and Technology* 2nd ed., CRC Press.
- Conesa A & Stefan G (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, 2008:619832
- Coon KL, Valzania L, McKinney DA et al. (2017) Bacteria-mediated hypoxia functions as a signal for mosquito development. *Proceedings of the National Academy of Sciences*, 114(27), pp.E5362–E5369.
- Correia MA, Otrelo-Cardoso AR, Schwuchow V et al. (2016) The *Escherichia coli* periplasmic aldehyde oxidoreductase is an exceptional member of the xanthine oxidase family of molybdoenzymes. *ACS Chemical Biology*, 11(10), pp.2923–2935.
- Cox-Foster DL, Schonbaum CP, Murtha MT et al. (1990) Developmental expression of the glucose dehydrogenase gene in *Drosophila melanogaster*. *Genetics*, 124(4), pp.873–880.
- Daniels EV, Murad R, Mortazavi A et al. (2014) Extensive transcriptional response associated with seasonal plasticity of butterfly wing patterns. *Molecular Ecology* 23(24), pp.6123–6134.
- Đapa T, Leuzzi R, Ng YK et al. (2013) Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *Journal of Bacteriology*, 195(3), pp.545–555.
- Delobel B, Rahbé Y, Nardon C et al. (1993) Biochemical and cytological survey of tyrosine storage proteins in Coleoptera: diversity and strategies. *Insect biochemistry and molecular biology*, 23(3), pp.355–365.
- Duperthuy M, Schmitt P, Garzón E et al. (2011) Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus*. *Proceedings of the National Academy of Sciences*, 108(7), pp.2993–8.
- Egert M, Stingl U, Bruun LD et al. (2005) Structure and topology of microbial communities in the major gut compartments of *Melolontha melolontha* larvae (Coleoptera:Scarabaeidae) *Applied and environmental*

microbiology, 71(8).

Emery O, Schmidt K & Engel P (2017) Immune system stimulation by the gut symbiont *Frischella perrara* in the honey bee (*Apis mellifera*). *Molecular Ecology*, 26(9), pp.2576–2590.

Engel P & Moran NA (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, 37(5), pp.699–735.

Fukumori K, Koga R, Nikoh N et al. (2017) Symbiotic bacteria associated with gut symbiotic organs and female genital accessory organs of the leaf beetle *Bromius obscurus* (Coleoptera:Chrysomelidae). *Applied Entomology and Zoology*.

Futahashi R, Tanaka K, Tanahashi M et al. (2013) Gene expression in gut symbiotic organ of stinkbug affected by extracellular bacterial symbiont. *PLoS ONE*, 8(5).

Galdiero S, Falanga A, Cantisani M et al. (2012) Microbe-host interactions: structure and role of gram-negative bacterial porins. *Current Protein and Peptide Science*, 13(8), pp.843–854.

Garcia ES, Castro DP, Figueiredo MB et al. (2010) Immune homeostasis to microorganisms in the guts of triatomines (Reduviidae) - a review. *Memórias do Instituto Oswaldo Cruz* 2010 Aug;105(5):605-610.

Gilmour D (1961) *The biochemistry of insects*, New York, N.Y. & London, Academic Press.

Giorgi G, Marcantonio P & del Re B (2011) LINE-1 retrotransposition in human neuroblastoma cells is affected by oxidative stress. *Cell and Tissue Research*, 346(3), pp.383–391.

Girvan HM & Munro AW (2013) Heme sensor proteins. *Journal of Biological Chemistry*, 288(19), pp.13194–13203.

Gordya NA, Nesin AP, Simonenko NP et al. (2016) Regulation of antimicrobial peptide synthesis in larvae of *Calliphora vicina* (Diptera:Calliphoridae): a dose-dependent effect of ecdysteroids. *Journal of Evolutionary Biochemistry and Physiology*, 52(4), pp.292–298.

Gretschler RR, Streicher PE, Strauß AS et al. (2016) A common theme in extracellular fluids of beetles: extracellular superoxide dismutases crucial for balancing ROS in response to microbial challenge. *Scientific Reports*, 6(April), pp.1–13.

Grünwald S, Pilhofer M & Höll W (2009) Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles [Coleoptera:Cerambycidae]. *Systematic and Applied Microbiology*, 33(1), pp.25–34.

Haberland M, Montgomery RL & Olson EN (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews. Genetics*, 10(1), pp.32–42.

Heath JJ, Cipollini DF & Stireman JO (2012). The role of carotenoids and their derivatives in mediating interactions

- between insects and their environment. *Arthropod-Plant Interactions*, pp.1–20.
- van der Hoeven R & Forst S (2009) OpnS, an outer membrane porin of *Xenorhabdus nematophila*, confers a competitive advantage for growth in the insect host. *Journal of Bacteriology*, 191(17), pp.5471–5479.
- Hosokawa T, Kikuchi Y, Nikoh N et al (2012) Polyphyly of gut symbionts in stinkbugs of the family Cydnidae. *Applied and Environmental Microbiology*, 78(13), pp.4758–4761.
- Indiragandhi P, Anandham R, Madhaiyan M et al. (2007) Cultivable bacteria associated with larval gut of prothiofos-resistant, prothiofos-susceptible and field-caught populations of diamondback moth, *Plutella xylostella* and their potential for antagonism towards entomopathogenic fungi and host insect nutrition. *Journal of Applied Microbiology*, 103(6), pp.2664–2675.
- Karlson P (1989) *Ecdysone: from chemistry to mode of action*, Jan Koolman ed. Thieme Medical Pub.
- Kikuchi Y, Hayatsu M, Hosokawa T et al. (2012) Symbiont-mediated insecticide resistance. *Proceedings of the National Academy of Sciences*, 109(22), pp.8618–8622.
- Kikuchi Y, Meng X & Fukatsu T (2005) Gut symbiotic bacteria of the genus *Burkholderia* in the broad-headed bugs *Riptortus clavatus* and *Leptocorisa chinensis* (Heteroptera:Alydidae). *Applied and Environmental Microbiology*, 71(7), pp.4035–4043.
- Kim DH, Hwang JS, Lee IH et al. (2016) The insect peptide CopA3 increases colonic epithelial cell proliferation and mucosal barrier function to prevent inflammatory responses in the gut. *Journal of Biological Chemistry*, 291(7), pp.3209–3223.
- Kim JK, Lee JB, Jang HA et al. (2016) Understanding regulation of the host-mediated gut symbiont population and the symbiont-mediated host immunity in the *Riptortus-Burkholderia* symbiosis system. *Developmental and Comparative Immunology*, 64, pp.75–81.
- Kim YS, Ryu JH, Han SJ et al. (2000) Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and β -1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *Journal of Biological Chemistry*, 275(42), pp.32721–32727.
- Küchler SM, Dettner K & Kehl S (2010) Molecular characterization and localization of the obligate endosymbiotic bacterium in the birch catkin bug *Kleidocerys resedae* (Heteroptera:Lygaeidae, Ischnorhynchinae). *FEMS Microbiology Ecology*, 73(2), pp.408–418.
- Lee JB, Park KE, Lee SA et al. (2017) Gut symbiotic bacteria stimulate insect growth and egg production by modulating hexamerin and vitellogenin gene expression. *Developmental and Comparative Immunology*, 69, pp.12–22.

- Li A & Denlinger DL (2009) Pupal cuticle protein is abundant during early adult diapause in the mosquito *Culex pipiens*. *Journal of medical entomology*, 46(6), pp.1382–1386.
- Lushchak VI (2010) Adaptive response to oxidative stress: bacteria, fungi, plants and animals. *Comparative Biochemistry and Physiology*, 153(2), pp.175–190.
- Mamidala P, Wijeratne AJ, Wijeratne S et al. (2012) RNA-Seq and molecular docking reveal multi-level pesticide resistance in the bed bug. *BMC Genomics*, 13(6).
- Margulis L, Jorgensen JZ, Dolan S et al. (1998) The arthromitus stage of *Bacillus cereus*: intestinal symbionts of animals. *Proceedings of the National Academy of Sciences*, 95(3), pp.1236–1241.
- Martin C & Zhang Y (2005) The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology*, 6(11), pp.838–849.
- Martin M (1983) Cellulose digestion in insects. *Comparative Biochemistry and Physiology*, 75A(3), pp.313–324.
- Matson EG, Gora KG & Leadbetter JR (2011) Anaerobic carbon monoxide dehydrogenase diversity in the homoacetogenic hindgut microbial communities of lower termites and the wood roach. *PLoS ONE*, 6(4).
- McKenney PT, Driks A & Eichenberger P (2013) The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. *Nature Reviews Microbiology*, 11(1), pp.33–44.
- Mhiri C, Vernhettes S & Casacuberta JM (1997) The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Molecular Biology*, pp.257–266.
- Mithöfer A & Boland W (2012) Plant defense against herbivores: chemical aspects. *Annual Review of Plant Biology*, 63(1), pp.431–450.
- Mortazavi A, Williams BA, McCue K et al. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), pp.621–628.
- Nagy A, Gertsenstein M, Vintersten K et al. (2007) General procedures for avoiding contamination with RNase. *Cold Spring Harbor Protocols*, 2007 Sep 1;2007:pdb.ip43.
- Nakayama DG, Santos CD, Kishi LT et al. (2017) A transcriptomic survey of *Migdolus fryanus* (sugarcane rhizome borer) larvae. *PLoS ONE*, 12(3), pp.1–24.
- Nakhleh J, El Moussawi L & Osta MA (2017) *The Melanization Response in Insect Immunity* 1st ed., Elsevier Ltd.
- Navari-Izzo F, Quartacci MF & Sgherri C (2002). Lipoic acid: A unique antioxidant in the detoxification of activated oxygen species. *Plant Physiology and Biochemistry*, 40(6–8), pp.463–470.
- Nichol H, Law JH & Winzerling JJ (2002) Iron metabolism in insects. *Annu. Rev. Entomol.*, 47, pp.535–559.

- Niwa YS & Niwa R (2016) Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. *Development Growth and Differentiation*, 58(1), pp.94–105.
- Oshlack A & Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, 4(14).
- Pang X, Xiao X, Liu Y et al (2016) Mosquito C-type lectins maintain gut microbiome homeostasis. *Nature Microbiology*, pp.1–24.
- Paredes-Sabja D, Setlow P & Sarker MR (2011) Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. *Trends in Microbiology*, 19(2), pp.85–94.
- Peterson BF & Scharf ME (2016) Metatranscriptome analysis reveals bacterial symbiont contributions to lower termite physiology and potential immune functions. *BMC Genomics*, pp.1–12.
- Prado SS & Almeida RPP (2009) Phylogenetic placement of pentatomid stink bug gut symbionts. *Current Microbiology*, 58(1), pp.64–69.
- Provenzano D & Klose KE (2000) Altered expression of the ToxR-regulated porins OmpU and OmpT diminishes *Vibrio cholerae* bile resistance, virulence factor expression and intestinal colonization. *Proceedings of the National Academy of Sciences*, 97(18), pp.10220–4.
- Rawls JF, Mahowald MA, Goodman AL et al. (2007) *In vivo* imaging and genetic analysis link bacterial motility and symbiosis in the zebrafish gut. *Proceedings of the National Academy of Sciences*, 104(18), pp.7622–7627.
- Robinson MD, McCarthy DJ & Smyth GK, 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.
- Roy A, Walker WB, Vogel H et al (2016) Diet dependent metabolic responses in three generalist insect herbivores *Spodoptera* spp. *Insect Biochemistry and Molecular Biology*, 71, pp.91–105.
- Shah N, Dorer DR, Moriyama EN et al. (2012) Evolution of a large, conserved and syntenic gene family in insects. *Genes/Genomes/Genetics*, 2(2), pp.313–319.
- Shao Y, Spiteller D, Tang X et al. (2011) Crystallization of alpha- and beta-carotene in the foregut of *Spodoptera* larvae feeding on a toxic food plant. *Insect Biochemistry and Molecular Biology*, 41(4), pp.273–281.
- Shelomi M (2017) *De novo* transcriptome analysis of the excretory tubules of *Carausius morosus* (Phasmatodea) and possible functions of the midgut “appendices”. *PLOS one*, 12(4), p.e0174984.
- Sloan DB & Moran NA (2012) Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology letters*, pp.1–4.

- Stephens WZ, Wiles TJ, Martinez ES et al (2015) Identification of population bottlenecks and colonization factors during assembly of bacterial communities within the zebrafish intestine. *mBio*, 6(6), pp.1–11.
- Stoll S, Gadau J, Gross R et al. (2007) Bacterial microbiota associated with ants of the genus *Tetraponera*. *Biological Journal of the Linnean Society*, 90(3), pp.399–412.
- Tanahashi M, Meng XY & Fukatsu T (2017) A novel symbiotic ciliate (Ciliophora:Peritrichia) in the hindgut of a stag beetle (Coleoptera:Lucanidae). *Zoological Science*, 34(3), pp.217–222.
- Tang X, Liu H, Yang, S et al. (2016) Epidermal growth factor and intestinal barrier function. *Mediators of Inflammation*, 2016: 1927348.
- Terra W & Ferreira C (2009) Digestive system. In V. Resh & R. Carde, eds. *Encyclopedia of Insects*. Elsevier, pp. 273–281.
- Vilcinskas A & Vogel H (2016) Seasonal phenotype-specific transcriptional reprogramming during metamorphosis in the European map butterfly *Araschnia levana*. *Ecology and Evolution*, 6(11), pp.3476–3485.
- Vinokurov KS, Elpidina EN, Oppert B et al. (2006) Diversity of digestive proteinases in *Tenebrio molitor* (Coleoptera:Tenebrionidae) larvae. *Comparative Biochemistry and Physiology*, 145(2), pp.126–137.
- Wang X & Wood TK (2011) Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Applied and Environmental Microbiology*, 77(16), pp.5577–5583.
- Warnecke F, Luginbühl P, Ivanova N et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 450(November), pp.560–565.
- Warr E, Das S, Dong Y et al. (2008) The gram-negative bacteria-binding protein gene family: its role in the innate immune system of *Anopheles gambiae* and in anti-Plasmodium defence. *Insect Molecular Biology*, 17(1), pp.39–51.
- Whittaker JW (2012) Non-heme manganese catalase - the “other” catalase. *Archives of Biochemistry and Biophysics*, 525(2), pp.111–120.
- Wildbolz T (1954) Beitrag zur Anatomie, Histologie und Physiologie des Darmkanals der Larve von *Melolontha melolontha* L. *Mitteilungen der Schweizerischen Entomologischen Gesellschaft*, 27: 193-240
- Willis JH, Papandreou NC, Iconomidou VA et al. (2012) Cuticular proteins. *Insect Molecular Biology and Biochemistry* pp.134-166
- Yadav S, Daugherty S, Shetty AC et al. (2017) RNAseq analysis of the *Drosophila* response to the entomopathogenic nematode *Steinernema*. *Genes/Genomes/Genetics*, 7(June), pp.1955–1967.

Yamada Y, Kuzuyama T, Komatsu M et al. (2014) Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences*, 112(3), pp.857–862.

Yazzie N, Salazar KA & Castillo MG (2015) Identification, molecular characterization, and gene expression analysis of a CD109 molecule in the Hawaiian bobtail squid *Euprymna scolopes*. *Fish and Shellfish Immunology*, 44(1), pp.342–355.

Yu Z, An B, Ramshaw JA et al. (2014) Bacterial collagen-like proteins that form triple-helical structures. *Journal of Structural Biology*, 186(3), pp.451–461.

Zelensky AN & Gready JE (2005) The C-type lectin-like domain superfamily. *FEBS Journal*, 272(24), pp.6179–6217.

Zhang X, Matson EG & Leadbetter JR (2010) Genes for selenium dependent and independent formate dehydrogenase in the gut microbial communities of three lower, wood-feeding termites and a wood-feeding roach. *Environmental Microbiology*, 13(2), pp.307–323.

Supplementary Material

The supplementary material of this manuscript can be downloaded at:

<https://www.dropbox.com/s/mmqmcoedvidyaic/Supplementary%20materials%20RNAseq.zip?dl=0>