


RESEARCH ARTICLE

Open Access



Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*

Juliane Wippler^{1,6*†} , Manuel Kleiner^{1,2*†}, Christian Lott^{1,3}, Alexander Gruhl¹, Paul E. Abraham⁴, Richard J. Giannone⁴, Jacque C. Young^{4,5}, Robert L. Hettich⁴ and Nicole Dubilier¹

Abstract

Background: The gutless marine worm *Olavius algarvensis* has a completely reduced digestive and excretory system, and lives in an obligate nutritional symbiosis with bacterial symbionts. While considerable knowledge has been gained of the symbionts, the host has remained largely unstudied. Here, we generated transcriptomes and proteomes of *O. algarvensis* to better understand how this annelid worm gains nutrition from its symbionts, how it adapted physiologically to a symbiotic lifestyle, and how its innate immune system recognizes and responds to its symbiotic microbiota.

Results: Key adaptations to the symbiosis include (i) the expression of gut-specific digestive enzymes despite the absence of a gut, most likely for the digestion of symbionts in the host's epidermal cells; (ii) a modified hemoglobin that may bind hydrogen sulfide produced by two of the worm's symbionts; and (iii) the expression of a very abundant protein for oxygen storage, hemerythrin, that could provide oxygen to the symbionts and the host under anoxic conditions. Additionally, we identified a large repertoire of proteins involved in interactions between the worm's innate immune system and its symbiotic microbiota, such as peptidoglycan recognition proteins, lectins, fibrinogen-related proteins, Toll and scavenger receptors, and antimicrobial proteins.

Conclusions: We show how this worm, over the course of evolutionary time, has modified widely-used proteins and changed their expression patterns in adaptation to its symbiotic lifestyle and describe expressed components of the innate immune system in a marine oligochaete. Our results provide further support for the recent realization that animals have evolved within the context of their associations with microbes and that their adaptive responses to symbiotic microbiota have led to biological innovations.

Keywords: RNA-Seq, Annelida, Oligochaeta, Phalloporinae, PGRP, FREP, SRCR, Respiratory pigment, Carbon monoxide, Immunology, Chemosynthetic symbiosis

* Correspondence: jwippler@mpi-bremen.de; manuel.kleiner@ucalgary.ca

†Equal contributors

¹Symbiosis Department, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

Full list of author information is available at the end of the article



Background

Most, if not all, animals are associated with a species-specific microbial assemblage that profoundly affects their evolution, ecology, development and health [1–3]. Animals and their microbiota have evolved molecular mechanisms to recognize and maintain these stable associations, and on the host side, these mechanisms are largely mediated by their immune system [4]. The mechanisms that govern host-symbiont interactions have been studied in a number of model organisms [4, 5], but remain unexplored in many animal phyla.

Olavius algarvensis is a gutless oligochaete worm (Annelida; Oligochaeta; Phalloporilinae) that lives in an obligate nutritional symbiosis with at least four bacterial species [6]. These extracellular endosymbionts thrive in a dense bacterial layer between the cuticle and the epidermis of the worm (Fig. 1). Over the course of their

symbiotic evolution, the gutless oligochaetes have lost their digestive and excretory organs, and rely solely on their bacterial symbionts for nourishment and removal of their waste products [7–9]. *O. algarvensis* harbors two gammaproteobacterial symbiont species that are chemoautotrophic sulfur oxidizers, and two deltaproteobacterial symbionts that are sulfate-reducing bacteria [6]. Together, these symbionts engage in a syntrophic sulfur cycle that enables autotrophic carbon fixation by the sulfur-oxidizing symbionts and provision of organic carbon to the host [8, 9]. Many worm individuals also harbor a spirochaetal symbiont, whose function has not yet been resolved [10].

Metagenomic and metaproteomic studies of the symbionts have revealed much about their metabolic capabilities, highlighted their immense capacity to use and recycle the host’s waste products and led to the

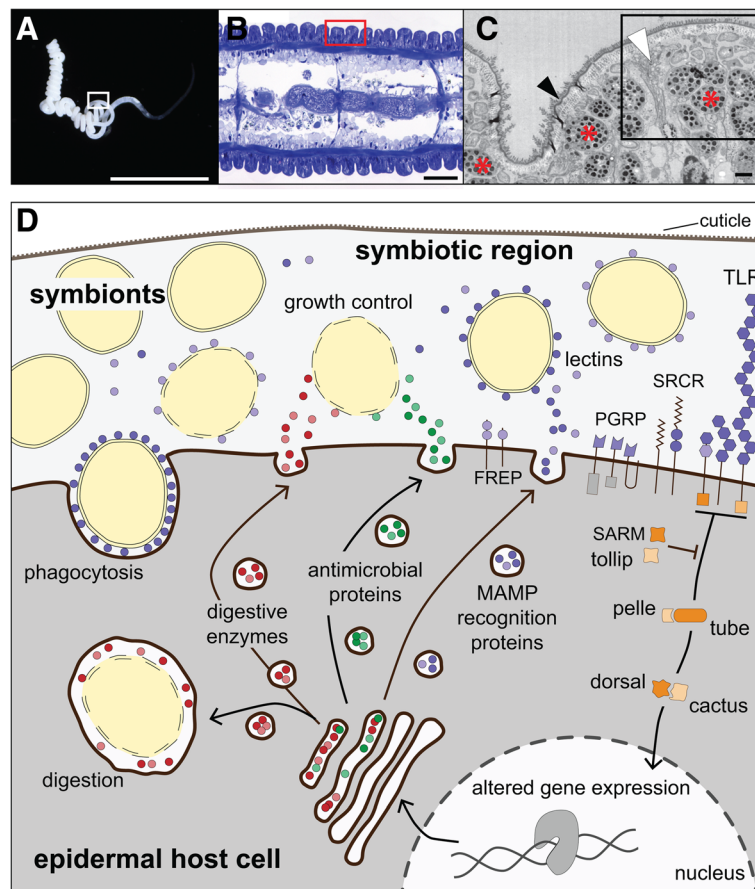


Fig. 1 Schematic overview of proposed molecular host-symbiont interactions. **a**) Light micrograph of an *Olavius algarvensis* worm. Scale bar 5 mm. **b**) Light micrograph of a longitudinal section through *O. algarvensis*, tissue stained with toluidine blue. Scale bar 50 μ m. **c**) Transmission electron micrograph of the symbiotic region, longitudinal section. Red asterisks, symbiont cells; black arrow, cuticle; white arrow, epidermal cell extensions. Scale bar 5 μ m. The boxes in **a**), **b**) and **c**) frame regions corresponding to the image to their right (in **a**) and **b**) or below (in **c**). Images **a**), **b**) and **c**) do not show the same worm specimen. **d**) Schematic overview of the main groups of expressed pattern recognition molecules, components of the Toll immune signaling pathway and proposed interactions between the host and its symbionts. Ig, immunoglobulin domain proteins; PGRP, peptidoglycan recognition proteins; SRCR scavenger receptor-like cysteine rich proteins; TLR, Toll-like receptors; FREP, fibrinogen-related proteins; AMPs, antimicrobial proteins

discovery of novel, energy-efficient pathways to fix both inorganic and organic carbon into biomass [8, 9]. Research aimed at a better understanding of the host, on the other hand, has been hampered by the fact that the worms are very small (0.1–0.2 mm in diameter and 10–20 mm in length), cannot be cultivated, and by a lack of sequence data. Recent advances in sequencing technology have made it possible to sequence and assemble comprehensive *de novo* transcriptomes of uncultured, non-model organisms collected in the environment. These transcriptomes provide a reference database for identifying the proteins organisms express using mass-spectrometry-based proteomic approaches. This methodological advance has opened the door for in depth studies of the molecular repertoire used by *O. algarvensis* and other non-cultivable organisms to establish and maintain a successful symbiosis.

All animals employ mechanisms for selecting and maintaining a specific microbial consortium over the course of their lives, while avoiding overgrowth by their own microbiota or infection by detrimental bacteria from the environment. The innate immune system is crucial in the establishment and maintenance of healthy symbiotic interactions, but has so far not been studied in gutless oligochaetes. These hosts face additional challenges because they obligately rely on their symbionts and therefore must provide conditions under which they can thrive, while also dealing with the physiological challenges caused by their symbiotic lifestyle. For example, *O. algarvensis* must be able to live in anoxic sediment layers for extended periods of time to enable sulfate reduction by its anaerobic sulfate-reducing symbionts [8]. Additionally, it must be able to deal with the hydrogen sulfide that is produced during sulfate reduction. It must also be able to endure the relatively high carbon monoxide concentrations in its environment, which both the sulfate-reducing and sulfur-oxidizing symbionts use as an energy source [9, 11]. Another challenge occurs when *O. algarvensis* inhabits the upper oxygenated sediment layers where it competes for oxygen with its aerobic sulfur-oxidizing symbionts.

Here, we used transcriptomics and proteomics to elucidate how *O. algarvensis* fulfills the physiological requirements outlined above and how it obtains nutrition from its symbionts. We exposed worms collected from the environment to two types of conditions that they naturally encounter to increase transcriptome and proteome coverage. Our identification and analysis of proteins expressed by *O. algarvensis* provide insights into their molecular mechanisms for microbe recognition, interaction and regulation, as well as their physiological adaptations to living in symbiosis with sulfur-oxidizing and sulfate-reducing bacteria.

Methods

Sample collection and incubations

For proteomic analyses, sediment that contained gutless oligochaete worms was collected at 7 m water depth in the Bay of Sant'Andrea, Elba, Italy (42° 48' 29.38' N, 10° 08' 31.57' E) in October 2007 and 2008. Worms were carefully washed out of the sediment at the HYDRA field station (Fetovaia, Elba, Italy) by hand (for details see [9]). To increase proteome coverage we treated the worms in the following manner. Worms were either immediately frozen in liquid nitrogen in batches of 150–200 worms (called "fresh" worms in the following) or were kept for 8 days in glass petri dishes filled with a thin layer (2–3 mm) of washed sediment and 0.2 µm-filtered sea water and then frozen in liquid nitrogen (called "starved" worms in the following, because no external electron donor for energy conservation and autotrophic carbon fixation was provided). The sulfur-oxidizing symbionts of *O. algarvensis* store large amounts of sulfur and polyhydroxyalkanoate granules, which give the worms a bright white appearance. Under prolonged exposure to oxygen without access to an electron donor like the sulfide produced anaerobically by the sulfate-reducing symbionts, these storage granules become depleted, the worms turn transparent, and are effectively starved of nutrition. Transparent worms are regularly found in the environment, especially during the reproductive season of the worms. All samples were stored in liquid N₂ and later at –80 °C until further use.

For transcriptomics, 100–120 worms were collected in April 2012 from the same site as for proteomics. The live worms were kept in washed sediment and transported to the lab in Bremen where they were washed out of the sediment again, washed in petri dishes with filtrated seawater, then flash-frozen in liquid nitrogen and stored at –80 °C until they were used to prepare the cDNA library "A". A second collection of worms was used for library "B" to identify genes expressed under prolonged anoxia, a condition that the worms often experience. For library "B", 100–120 live worms were collected in March 2013 from the same site as above, transported to Bremen in the same way as for library "A", and incubated in anoxic serum bottles for 43 h. Serum bottles were filled with sediment and sea water from Elba, and were flushed with nitrogen to remove oxygen from the headspace. The sediment and sea water were not sterilized, so that fully anoxic conditions could develop quickly through microbial metabolism. Oxygen concentrations were measured at the end of the incubation with an oxygen microelectrode and were below 0.1 µM. Worms were fixed overnight in RNAlater (Thermo Fisher Scientific, Braunschweig, Germany) at 4 °C and stored at –80 °C until they were used to prepare the cDNA library "B".

Illumina library preparation and sequencing

Total RNA was isolated using peqGOLD TriFast reagent (PEQLAB, Erlangen, Germany) and treated with DNase. Poly(A)⁺ RNA was isolated from the total RNA, fragmented with ultrasound (2 pulses of 30 s at 4 °C) and used for cDNA synthesis with random hexamer primers. Illumina TruSeq adaptors were ligated to the ends of the cDNA fragments and amplified according to the manufacturer's instructions (Illumina Inc., USA). Library DNA fragments of 300–500 bp were eluted from a preparative agarose gel and paired-end sequenced on an Illumina HiSeq2000 sequencer (2x 100 bp). A standard PhiX174 DNA spike-in of 1% was used for sequencing quality control. We sequenced ~170 million read pairs from library A, and ~6 million read pairs from library B (Additional file 1: Table S1). For library B, a much smaller number of reads was sequenced because the purpose of this library was to detect abundant transcripts expressed under anoxic conditions.

De novo transcriptome assembly and sequence analysis

The raw reads were trimmed to remove Illumina adapters, filtered for PhiX174 spike-in DNA and quality trimmed with nsoni clip version 0.109 (parameters used: --match 7 -quality 27 -trim-start 10) [12]. The cleaned reads from library A and B were co-assembled *de novo* using Trinity release 2013-02-25 with default parameters [13]. Trinity reports individual assembled transcript sequences ("isoforms") as members of transcript families ("components"), which can represent fragments of the same transcript, chimeric artifacts, or actual biological splice variants of a gene. In our reports of numbers of different transcripts from a certain family of proteins (e.g. number of identified peptidoglycan recognition proteins), we stick to the more conservative number of "components" rather than reported isoforms, as these are difficult to reliably verify without a confirmed reference. Transcripts were quantified with RSEM as implemented in Trinity using default parameters [14].

De novo assembled transcripts were annotated with blast2GO [15]. Transcripts of particular interest were searched against the invertebrate division of EST (Expressed Sequence Tags) and TSA (Transcriptome Shotgun Assembly) sequences of NCBI with tblastx [16] to determine their similarity to genes expressed in other annelids.

Hemoglobin sequences were assigned to families, if possible, based on sequence homology and specific conserved amino acid patterns as described in [17]. The secondary structures of the putative sulfide binding domains in *O. algarvensis* hemoglobin chains were predicted with hydrophobic cluster analysis using the program drawhca [18].

Host species identification

Two species of gutless oligochaete co-occur at the sampling site, which can only be distinguished under the dissecting scope when sexually mature. The majority of worms used in this study were not sexually mature, and as a result, could not be morphologically identified. Therefore, we used EMIRGE [19] to estimate the relative abundance of the different species in our samples based on the read coverage of the mitochondrial cytochrome c oxidase I (COI) gene. We determined that the contamination with species other than *Olavius algarvensis* was less than 3.5% in library A and less than 0.1% in library B.

2D-LC-MS/MS

Protein was extracted from frozen worms, and peptides prepared as previously described using a single-tube small processing method [9, 20]. We analyzed three biological replicates for each condition (fresh and starved). All samples were analyzed in technical duplicates via 24 h nano-2D-LC MS/MS with a split-phase column (RP-SCX-RP) [21, 22] on a hybrid linear ion trap-Orbitrap (Thermo Fischer Scientific), as previously described [9].

Peptide and protein identifications

Coding sequences (CDS) were predicted from the transcriptomes using FrameDP [23] and getorf of the EMBOSS package using the standard genetic code [24]. Transcriptome CDS were combined into a single protein sequence database with the symbiont protein sequence database used by Kleiner et al. 2012 [9]. To remove potential chimeric sequences and redundant CDS from the database we used sequence clustering with CD-HIT (version 4.5.4, [25]). Experimental peptide fragmentation spectra (MS/MS) generated from Xcalibur v2.0.7 were compared with theoretical peptide fragmentation spectra obtained from the protein sequence database to which protein sequences of common contaminant proteins (e.g., human keratin and trypsin) were added to a total of 1,318,114 entries. To determine the false-discovery rate (FDR), a decoy database, generated by reversing the sequences of the target database, was appended.

MyriMatch v2.1.111 [26] was configured to derive fully-tryptic peptides with the following parameters: 2 missed cleavages, parent mass tolerance of 10 ppm, and a fragment mass tolerance of 0.5 m/z units. For protein inference, peptide identifications were merged together in IDPicker v.3 [27]. Only protein identifications with at least two identified spectra and a maximum q-value of 0.02 were considered for further analysis. The number of distinct peptides (i.e., a peptide with a unique series of amino acids, but does not relate to its uniqueness to the protein reference database) required for identifications

was set to 1 to allow for the identification of small anti-microbial proteins and/or small, fragmented protein sequences in the transcriptome assembly. Based on these settings, protein-level FDR was < 3% for all samples.

To deal with sequence redundancy, post-search protein grouping was performed by clustering all protein sequences in the protein sequence database by sequence similarity ($\geq 90\%$) using the UCLUST component of the USEARCH v5.0 software platform [28]. As described previously [29], identified proteins were then consolidated into their defined protein groups. Protein groups were represented by the longest protein sequence (i.e., the seed sequence), which shares $\geq 90\%$ sequence similarity to each member of the protein group. Peptide uniqueness was re-assessed and classified as either unique (i.e., only belonging to one protein group) or non-unique (i.e., shared among multiple protein groups). We required each protein group to have at least two distinct peptides, with at least one of these being a unique peptide. For shared peptides belonging to multiple protein groups, their spectral counts were recalculated based on the proportion of uniquely identified peptides between the protein groups sharing the peptide. Following spectra balancing, total spectral counts of a protein group were converted to normalized spectra counts (nSpC) [30], which are derived from normalized spectral abundance factors [31]. Relative protein abundances of host proteins are listed in tables as nSpC values multiplied by 10,000 i.e. the sum of all host protein nSpC values in one sample is 10,000 and the nSpC values are thus given as a fraction of 10,000 ($^0/_{000}$).

Results and Discussion

Transcriptome/proteome measurement metrics

To generate our protein sequence database for host protein identification, we sequenced the transcriptomes of untreated whole worms (library A), and of worms kept under anoxic conditions for 43 h (library B). We chose these two conditions, which the worms encounter regularly, to obtain a larger range of host transcripts for the generation of a comprehensive reference sequence database for improved protein identification. After trimming and error correction, 159,551,509 (library A) and 5,745,537 (library B) read pairs remained, which were co-assembled into 173,602 contigs, with an N50 of 1236 bp, and 23,719 contigs of at least 1 kbp length (for more details on sequencing and assembly metrics, see Additional files 1: Table S1 and Table S2). Of these contigs, 31,913 could be functionally annotated (see Additional file 1: Figure S1 for annotation summary). The sequencing depth and assembly metrics are comparable to other, well-covered, transcriptomes of recently sequenced annelid taxa [32]. We analyzed proteomes of freshly collected worms, and worms that

had been starved for 8 days - that is kept under oxic conditions without an external electron donor for energy conservation and autotrophic carbon fixation (see Methods). The purpose of creating these two conditions was to identify as many proteins as possible, including those expressed in worms that are starved. We identified a total of 4355 protein groups (see Methods for details on protein grouping) with a per sample protein-level FDR < 3%. Of these, 2562 were host proteins and 1793 were symbiont proteins. The annotated host transcriptomes and proteomes were manually screened for sequences relevant for host-symbiont interactions. We identified 316 transcriptome sequences and 60 proteins potentially involved in microbe recognition, microbial growth regulation, symbiont digestion, immune modulation and physiological interactions (see Table 1 and Fig. 1).

Physiological adaptations of the host to the symbiosis *Nutrients are transferred from the symbionts to the host via digestion*

Previous to this study, it was not clear how gutless oligochaetes gain nutrition from their bacterial symbionts. Two transfer modes, which are not mutually exclusive, have been suggested for symbioses with endosymbionts [33]: (1) “milking” of the symbionts (uptake of small compounds leaked or actively released by the symbionts), and (2) symbiont digestion through endocytosis. Endocytosis can include phagocytosis of symbiont particles or whole cells, as well as uptake of extracellularly digested and dissolved compounds by pinocytosis.

Several results from this study indicate that the main mode of nutrient transfer from the symbionts to *O. algarvensis* is through their digestion. First, we measured significantly less symbiont protein relative to host protein in the proteomes of starved worms compared to freshly collected worms (*t*-test, $p < 0.01$). Symbiont protein accounted for only 18.7% of the total holobiont protein in starved worms, while freshly collected worms had 29.5% symbiont protein (Table 2 and Additional file 1: Table S3). In starved worms, the symbionts had no access to external sources of energy and carbon. Since these worms gain all their nutrition from their symbionts, the absence of external energy and carbon sources meant that no net growth of the symbiosis was possible. Thus, both the worms and their symbionts were starved of nutrition. We therefore hypothesize that the observed decrease in total symbiont protein relative to total host protein in starved worms occurred because the symbionts were digested by the host. An alternative explanation for the decrease in relative symbiont protein is that the symbionts, but not the host, degraded their own proteins in response to starvation.

Table 1 Overview of proposed host-symbiont interaction proteins

Functional category	Protein family	Transcripts ^{a)}	Proteins ^{b)}
<i>Pattern recognition proteins</i>			
	peptidoglycan recognition proteins (PGRPs)	6 (16)	1
	C-type lectins	33 (119)	5
	R-type lectins	6 (18)	4
	SUEL/rhamnose-binding lectins	7 (22)	1
	galectins	3 (3)	1
	fucolectin	1 (23)	0
	fibrinogen-related proteins (FREPs)	27 (161)	1
	toll-like/variable lymphocyte receptor-like (TLR/VLR)	13 (52)	1
	scavenger receptor cysteine-rich (SRCR) domain proteins	25 (164)	1
	beta-1,3-glucan binding protein	1 (1)	1
	novel immunoglobulin I-set proteins	16 (17)	1
	novel immunoglobulin V-set proteins	9 (22)	1
<i>Antimicrobial proteins</i>			
	lumbricin	1 (1)	1
	invertebrate-type lysozyme	1 (1)	0
	bactericidal permeability increasing protein BPI	1 (3)	1
	insect defensin/reeler-like proteins	4 (8)	1
	cysteine-rich secretory proteins (CRSPs)	6 (28)	2
	membrane attack complex/perforin	2 (23)	0
<i>Other immune effectors</i>			
	ROS modulator 1	2 (3)	1
	alpha-2-macroglobulin	10 (24)	1
	kazal-type serpin	2 (8)	0
	kunitz-type serpin	1 (2)	0
	leukocyte elastase inhibitors	5 (25)	1
	phosphatidylethanolamine-binding protein PEBP	3 (5)	1
<i>Immune response regulators</i>			
	Toll/interleukin-1 receptor (TIR) domain proteins	5 (9)	0
	NF-kappa-B inhibitor Cactus	2 (6)	0
	dorsal protein	2 (3)	1
	evolutionarily conserved signaling intermediate in Toll (ECSIT)	1 (1)	0
	Pelle protein	1 (1)	0
	Relish protein	1 (6)	0
	mitogen-activated protein kinase kinase kinase 7 (TAK1)	1 (1)	0
	I-kappa-B-kinase alpha (IKK α)	1 (1)	0
	I-kappa-B-kinase beta (IKK β)	1 (5)	0
	interleukin-1 receptor-associated kinase 1 (IRAK1)	1 (2)	0
	mitogen-activated protein kinase kinase kinase 4 (MEKK4)	1 (1)	0
	sterile alpha and TIR motif-containing protein (SARM)	1 (3)	0
	Toll-interacting protein Tollip	1 (2)	1
	(LPS-induced) tumor necrosis factor (TNF)	3 (7)	0
	Tumor necrosis factor (TNF)	4 (6)	0
	tumor necrosis factor alpha-induced protein 3 (TNFAIP3)	1 (3)	0

Table 1 Overview of proposed host-symbiont interaction proteins (Continued)

tumor necrosis factor receptor associated proteins (TRAF)	9 (16)	0
IFN regulatory factor	8 (10)	0
IFN regulatory factor-binding protein	1 (1)	0
IFN-induced GTPase	7 (23)	1
macrophage migration inhibitory factor (MIF)	3 (24)	1
ILN enhancer binding factor 2	2 (4)	1
ILN-16	1 (1)	0
<i>Digestive enzymes</i>		
carboxypeptidases	11 (22)	0
cathepsins total	15 (28)	4
cathepsin B	3 (12)	1
cathepsin C	2 (4)	0
cathepsin F	3 (3)	1
cathepsin L	5 (6)	2
cathepsin O	1 (2)	0
cathepsin Z	1 (1)	0
chymotrypsins	3 (17)	1
pancreatic elastase	1 (1)	1
alpha amylase	2 (3)	1
lysosomal alpha glucosidase	1 (3)	1
acid trehalase	2 (4)	1
sucrase-isomaltase	1 (1)	0
lysosomal acid lipase	1 (1)	0
<i>Respiration</i>		
hemerythrin	2 (2)	2
giant extracellular hemoglobin, globin chains	12 (16)	8
giant extracellular hemoglobin, linker chains	6 (18)	5
sum	316 (1032)	60

For more extensive details see Additional file 2: Table S8

^aNumber of transcripts defined as Trinity components, which approximately correspond to genes; see [13]; in parentheses: number of contigs (isoforms or fragments)

^bNumber of unique proteins

Second, we identified 15 digestive enzymes predicted to occur in lysosomes, indicating their role in endocytosis, and 28 digestive enzymes involved in general secretory pathways, which could be targeted to phagolysosomes or to the extracellular region (Table 3). If these enzymes are not directed to phagolysosomes, but rather secreted

extracellularly, they could also aid in the digestion of symbionts in the extracellular space just below the worm's cuticle, and precede endocytotic digestion by the epidermal cells. The digestive proteins included various proteases for the degradation of polypeptides and oligopeptides, glucosidases with specificity for $\alpha 1 \rightarrow 4$, $\alpha 1 \rightarrow 6$ and $\beta 1 \rightarrow 4$ glycosidic bonds, and enzymes involved in lipid and peptidoglycan degradation (Table 3).

Table 2 Difference in symbiont protein content in fresh worms compared to starved worms

	Symbionts fresh worms	Symbionts starved worms
average nSpC	3050.30	1869.77
standard deviation	421.88	122.47
# replicates	3	3
p-value t-test	0.00963	

Significant differences between fresh and starved samples were determined with the Student's *t*-Test; nSpC, normalized spectral counts. For more extensive details see Additional file 1: Table S3

The third line of evidence that indicates that *O. algarvensis* digests its symbionts is that it expressed three different types of intestinal digestive enzymes, despite the fact that it does not have a mouth or gut. (i) The first type were digestive proteases (Table 3), namely pancreatic carboxypeptidase A, chymotrypsins A and B, cathepsins B, F and L, and pancreatic elastase. These enzymes are most often found in the intestinal tract of animals with a digestive system (Additional file 1: Table S4). Most of

Table 3 Digestive enzymes expressed in *Olavius algarvensis*

Transcript ID ^{a)}	Protein ID	Annotation	Consensus localization evidence ^{b)}	Substrate or function
<i>Protein digestion</i>				
comp310626_c3	n.d.	pancreatic carboxypeptidase A1	secreted	release of C-terminal amino acids
comp330196_c1	n.d.	pancreatic carboxypeptidase A2	secreted	release of C-terminal amino acids
comp209868_c0	n.d.	pancreatic carboxypeptidase A2	secreted	release of C-terminal amino acids
comp329532_c3	n.d.	pancreatic carboxypeptidase A2	secreted or membrane	release of C-terminal amino acids
comp328734_c12	n.d.	carboxypeptidase	secreted	peptides
comp328734_c1	n.d.	carboxypeptidase	secreted	peptides
comp328734_c4	n.d.	carboxypeptidase	secreted	peptides
comp326419_c0	n.d.	uncharacterized carboxypeptidase	secreted	peptides
comp330196_c2	n.d.	uncharacterized carboxypeptidase	secreted	peptides
comp319717_c3	n.d.	lysosomal Pro-X carboxypeptidase	lysosomal	proline - amino acid bonds
comp320275_c1	n.d.	chymotrypsin A	secreted	Tyr-/Trp-/Phe-/Leu- -Xaa bonds
comp306409_c1	n.d.	chymotrypsin B	secreted	Tyr-/Trp-/Phe-/Leu- -Xaa bonds
comp334148_c2	BF11_334148_c2_seq1_11 BF11_334148_c2_seq2_10	chymotrypsin-like protease ctrl-1	secreted	proteins
comp331491_c1	BF11_331491_c1_seq1_7	pancreatic elastase	secreted	proteins
comp334775_c2	n.d.	cathepsin B	secreted	Arg - Arg/-Xaa bonds
comp335560_c0	BF11_335560_c0_seq5_5 BF11_335560_c0_seq9_5	cathepsin B	lysosomal	Arg - Arg/-Xaa bonds
comp306522_c2	n.d.	cathepsin F	lysosomal	peptides, cleaves Phe/Leu
comp306522_c3	n.d.	cathepsin F	lysosomal	peptides, cleaves Phe/Leu
comp308536_c0	BF11_308536_c0_seq1_12	cathepsin F	lysosomal	peptides, cleaves Phe/Leu
comp283346_c1	n.d.	cathepsin L	lysosomal	proteins
comp306922_c1	BF11_306922_c1_seq1_34	cathepsin L	lysosomal	proteins
comp315575_c0	n.d.	cathepsin L	lysosomal	proteins
comp315575_c2	FD_315575_c2_seq1:64:270:1:+	cathepsin L	lysosomal	proteins
comp328653_c0	n.d.	cathepsin L	lysosomal	proteins
comp329800_c6	n.d.	cathepsin O	lysosomal	peptides (endopeptidase)
comp314408_c1	n.d.	cathepsin Z	lysosomal	C-terminal amino acids (not Pro)
comp308700_c0	n.d.	cathepsin C (dipeptidyl peptidase I)	lysosomal	release of an N-terminal dipeptide
comp329456_c1	n.d.	cathepsin C (dipeptidyl peptidase I)	lysosomal	release of an N-terminal dipeptide
comp328746_c4	n.d.	alpha amylase	secreted	α1- > 4 glycosidic bonds
comp324906_c1	BF11_324906_c1_seq1_9 BF11_324906_c1_seq2_12	alpha amylase	secreted or membrane	α1- > 4 glycosidic bonds
comp335205_c1	BF11_335205_c1_seq1_20	lysosomal alpha glucosidase	secreted or membrane	α1- > 4 glycosidic bonds
comp320084_c0	n.d.	lysosomal beta-mannosidase	lysosomal	cleaves terminal β-D-mannose
comp334411_c3	n.d.	sucrase-isomaltase	secreted or membrane	α1- > 6 glycosidic bonds
comp329957_c8	BF11_329957_c8_seq2_17	acid trehalase	secreted or membrane	trehalose - > glucose
comp335402_c7	n.d.	acid trehalase	secreted or membrane	trehalose - > glucose
<i>Lipid degradation</i>				
comp22535_c0	n.d.	lysosomal acid lipase	secreted	hydrolyzes steryl esters
comp249291_c0	n.d.	lysozyme	secreted	peptidoglycan (glycosidic bonds)
comp250229_c0	BF11_250229_c0_seq1_15 BF11_250229_c0_seq2_17	peptidoglycan recognition protein	secreted	peptidoglycan (peptide bonds)

Table 3 Digestive enzymes expressed in *Olavius algarvensis* (Continued)

comp335695_c10	n.d.	peptidoglycan recognition protein	secreted or membrane	peptidoglycan (peptide bonds)
comp330541_c4	n.d.	peptidoglycan recognition protein	secreted	peptidoglycan (peptide bonds)
comp314994_c0	n.d.	peptidoglycan recognition protein	secreted or membrane	peptidoglycan (peptide bonds)
comp332570_c2	n.d.	peptidoglycan recognition protein	secreted or membrane	peptidoglycan (peptide bonds)

^aDefined as Trinity components; see [13]

^bProbable subcellular localization of proteins based on the results of TMHMM [117, 118], SignalP-4 [119], Phobius [120], TargetP [121], DISTILL [122], LocTree3 [123], BaCellLo [124], and iLoc-Animal [125]. *Secreted* – enters secretory pathway and is either excreted to the extracellular space or confined to non-cytoplasmic insides of intracellular compartments; *membrane* – predicted to be membrane integral; *lysosomal* – predicted to be targeted towards the lysosome. See Additional file 3: Table S14 for details on localization evidence

the *O. algarvensis* digestive proteases were highly similar to enzymes expressed in the midgut of the oligochaete *Eisenia andrei* (Additional file 1: Table S5). (ii) *O. algarvensis* also expressed a number of digestive glucosidases: two alpha amylases, with best BLAST hits to salivary gland and pancreatic amylases, an intestinal sucrase-isomaltase and two enzymes similar to pancreatic acid trehalase (Additional file 1: Table S6). (iii) *O. algarvensis* expressed five peptidoglycan recognition proteins (PGRPs) with predicted amidase activity (Fig. 2) and a lysozyme, all proteins that degrade peptidoglycans. Although PGRPs and lysozyme are known for their role in immune defense [34], they can also aid in the digestion of food bacteria [35, 36]. The five *O. algarvensis* PGRP sequences were highly similar to PGRPs expressed by the annelid *Eisenia andrei* in its midgut (Additional file 1: Table S5). The expression levels of these enzymes in starved worms were comparable to those in fresh worms, suggesting that there was no overall increase in host digestion rates of symbionts during starvation.

Taken together, these results strongly indicate that *O. algarvensis* obtains nutrition from its symbionts by digestion, rather than milking, using a wide range of digestive enzymes, many of which are known to be expressed in the digestive tissues of animals. Given that the symbiotic bacteria are only found in the body wall of their host, it is highly likely that, in adaptation to the symbiosis, the expression of these “intestinal” enzymes has been redirected from the gut to the epidermis. This assumption is supported by ultrastructural analyses that

show the lysis of symbionts in the epidermal cells of the worm [37]. Additional support for the digestion of symbionts instead of “milking” stems from the observation that some of the *O. algarvensis* symbionts abundantly expressed high-affinity uptake transporters for organic substrates [9]. If ‘milking’ were the main manner in which the hosts gained their nutrition, they would compete with their symbionts for the uptake of small organic compounds.

Giant hemoglobins are likely involved in sulfide tolerance and transport

O. algarvensis abundantly expressed giant extracellular hemoglobins, which are respiratory pigments produced exclusively by annelids [38]. They are large multiprotein complexes (3.8 MDa in earthworms [39]), each consisting of more than a hundred copies of heme-containing globin chains and non-heme linker chains [38]. We found 12 globin chains and 6 linker chains from giant extracellular hemoglobins in our proteomes and transcriptomes (Additional file 1: Table S7). A signal peptide was predicted for all complete coding sequences, lending further support that these hemoglobins are indeed extracellular. Of the twelve *O. algarvensis* hemoglobin chain sequences, five could be unequivocally assigned to their respective families (3x family A, 2x family B).

We found that one of the three chains assigned to family A contained a free cysteine residue (Fig. 3). Free cysteine residues do not participate in the formation of disulfide bonds in proteins, and therefore may unintentionally react with other blood components and disturb

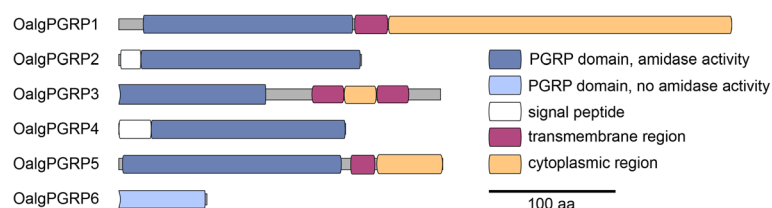
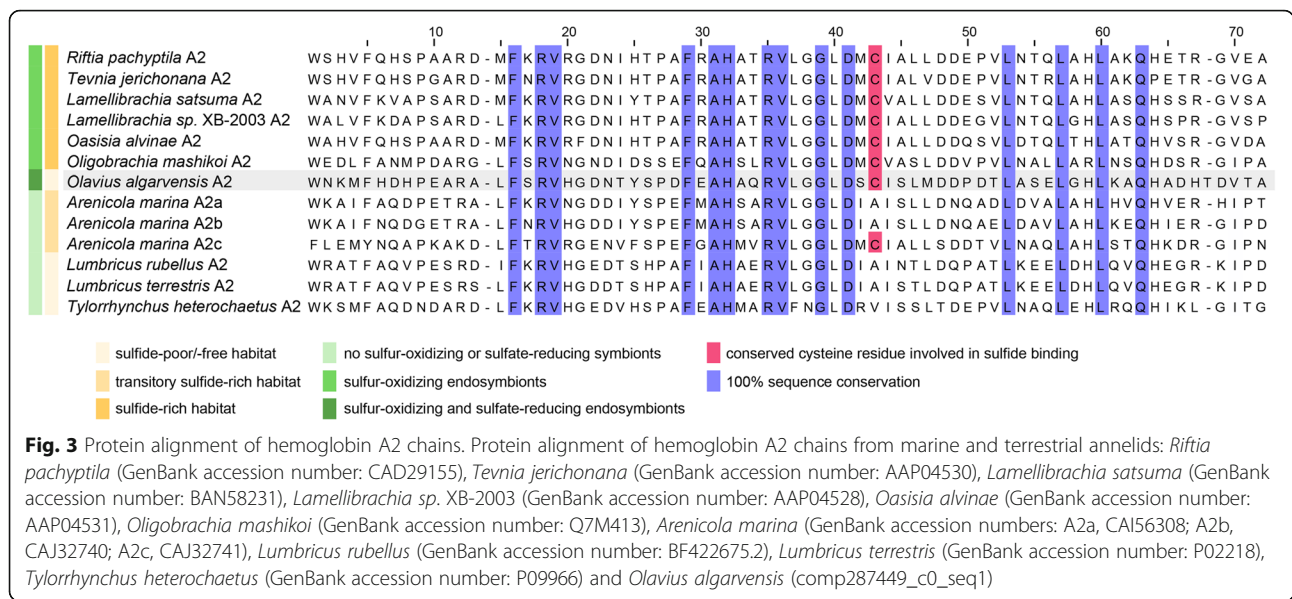


Fig. 2 Domain structures of peptidoglycan recognition proteins. Structure of conserved functional domains in *Olavius algarvensis* peptidoglycan recognition proteins; OalgPGRP1: comp330541_c4; OalgPGRP2: comp250229_c0; OalgPGRP3: comp335695_c10; OalgPGRP4: comp314994_c0; OalgPGRP5: comp332570_c2; OalgPGRP6: comp1100768_c0



blood homeostasis [40, 41]. Extracellular hemoglobins are therefore under strong selective pressure to avoid the incorporation of free cysteines. The exception are annelids that experience high concentrations of sulfide in their habitats (Fig. 3, [17]). In these worms, free cysteine residues in the A2 and B2 hemoglobin chains may allow them to reversibly bind environmental hydrogen sulfide and oxygen simultaneously [42]. It has been argued that this could mitigate the toxic effects of hydrogen sulfide for these worms. In hydrothermal vent tube worms, which also have free cysteine residues in their hemoglobin, it is assumed that these also allow them to bind and transport sulfide to their sulfur-oxidizing endosymbionts [43]. In these worms, sulfide-binding to hemoglobin could also be mediated by zinc ions rather than free cysteine [44, 45]; however zinc does not appear to play a role in sulfide-binding in other annelids [46].

In *O. algarvensis*, the free cysteine residue is located in the conserved position that allows sulfide binding, and hydrophobic cluster analysis showed that the molecular environment of this free cysteine is highly similar to the sulfide-binding domain of A2 chains in other annelids (Additional file 1: Figure S2). It is therefore plausible that the *O. algarvensis* hemoglobin can also bind sulfide.

O. algarvensis lives in oligotrophic sediments with very low environmental sulfide concentrations [6, 9]. However, its sulfate-reducing symbionts are a considerable internal source of sulfide under anoxic conditions [6]. With its sulfide-binding hemoglobin, the host could store this internally produced sulfide for use by the SOX symbionts once they return to oxic conditions. Furthermore, the sulfide-binding hemoglobin might keep sulfide levels low in sensitive tissues of *O. algarvensis* such as the central nervous system.

Hemerythrin may enable respiration in the absence of O₂ and in the presence of CO

In addition to hemoglobin, the host expressed two hemerythrins, which are also respiratory proteins, but without heme groups. One of these hemerythrins was by far the most abundant protein in both fresh and starved worms, and accounted for 11–15% of total host protein (Additional file 2: Table S8). In comparison, the second most abundant protein, a histone, accounted only for less than 3%. Both hemerythrins were more highly expressed than any of the hemoglobin chains; expression levels of the most abundant hemerythrin were almost 32 times higher than the most abundant globin chain in the proteome (Additional file 2: Table S8). Such abundant expression of hemerythrin is unknown from gut-bearing oligochaetes and other annelids (Additional file 1: Table S9).

Hemerythrin is an oxygen-carrying protein in sipunculids, priapulids and brachiopods, and also in a few polychaete annelids [47, 48]. In addition to oxygen transport, annelids might use hemerythrins for heavy metal resistance and antibacterial defense, or as an egg yolk protein [49–51]. In the only study that found hemerythrin expression in an oligochaete, it was assumed to be involved in heavy metal detoxification [50]. Since the environment of the *O. algarvensis* sampled for this study is not contaminated with high levels of heavy metals or pathogenic bacteria, and the worms in our experiments were not exposed to such conditions, it is unlikely that the high expression levels of hemerythrin are related to heavy metal resistance or antibacterial defense. We can also exclude its role in egg yolk protein, because the worms for proteomics were sampled in the fall, a time of the year when *O. algarvensis* does not reproduce (Kleiner, Lott, Wippler, unpublished observation). Therefore, it

seems most likely that the hemerythrin in *O. algarvensis* is used to bind oxygen. This raises the question why *O. algarvensis* has two abundant oxygen binding proteins - hemoglobin and hemerythrin.

The fact that hemerythrin expression is unusual in oligochaetes suggests that there is a considerable selective advantage for its expression in *O. algarvensis*. One intriguing property of hemerythrin is that it is insensitive to carbon monoxide (CO) [52]. In contrast, heme proteins such as hemoglobin and myoglobin have much higher affinities for CO than for oxygen [53, 54]. This makes CO highly toxic to organisms that rely on heme proteins for oxygen transport. Considerable *in situ* CO concentrations of up to 51 nM were regularly measured in the *O. algarvensis* environment [11], and CO serves as an energy source for its sulfur-oxidizing and sulfate-reducing symbionts [9]. Thus, the selective advantage of using hemerythrin for oxygen binding could be that it mitigates the adverse effects of carbon monoxide for the host.

The question remains why hemoglobin is also expressed in *O. algarvensis*, in parallel to hemerythrin. We speculate that hemerythrin and hemoglobin fulfill different functions in these worms. We propose that hemerythrin is used for oxygen storage to bridge the frequent and extended periods of anoxia that *O. algarvensis* is exposed to in the reduced sediment layers it mainly inhabits. Hemerythrin is well suited for oxygen storage because its oxygen binding capacity is stable under varying concentrations of O₂, CO₂ and protons [55, 56], and has been shown to play a key role in oxygen storage for bridging hypoxic episodes in sipunculids [57]. In contrast, hemoglobin, due to cooperative binding of oxygen and the Bohr effect, is well suited for gas exchange with the environment, which occurs in the upper oxic layer of the sediment where CO concentrations are much lower [11].

Interestingly, hemerythrin was also co-expressed with hemoglobin in the sulfur-oxidizing symbiont-bearing trophosome tissue of the deep-sea hydrothermal vent tube worm *Ridgeia piscesae*, a polychaete annelid that is not closely related to *O. algarvensis* [58]. The function of hemerythrin in *Ridgeia* is at present unknown. It is intriguing that the two animals currently known to abundantly express both hemoglobin and hemerythrin, *O. algarvensis* and *R. piscesae*, live in symbiosis with sulfur-oxidizing bacteria.

Interactions between the host innate immune system and its microbiome

We analyzed the proteins of the host innate immune system in our transcriptomes and proteomes, because these receptors, regulators and effectors are essential for sensing and responding to microbes [59], and are thus crucial for establishing and maintaining bacterial symbiosis [4]. The immune system must be able to distinguish

beneficial symbionts from harmful intruders, and must respond appropriately, avoiding chronic inflammation in the presence of symbionts, while allowing rapid elimination of non-symbiotic bacteria.

Multitude of pattern recognition molecules for differential responses to microbes

Pattern recognition receptors (PRRs) are proteins that recognize microbe-associated molecular patterns (MAMPs) by binding to surface molecules specific to microbes, like peptidoglycan or lipopolysaccharide [60]. PRRs are essential for sensing the presence of different microbial species and initiating an appropriate response, either via activation of immune signaling pathways and the synthesis of antimicrobial compounds, or by dampening or silencing the immune response in the case of bacterial symbionts [4]. We identified many different types of classical pattern recognition receptors, as well as proteins potentially involved in pattern recognition via conserved domains (Table 1).

PGRPs

Six peptidoglycan recognition proteins (OalgPGRP1-OalgPGRP6) were expressed in the *O. algarvensis* transcriptomes, and one of these was detected in the proteomes (OalgPGRP2, Table 1). PGRPs were first described as an important component of the innate immune defense [61], but are now known to play a major role in many animal-bacteria symbioses, mediating symbiont tolerance [62, 63], controlling symbiont populations [64], and regulating symbiosis establishment and maintenance [63, 65]. Elevated expression of PGRPs was also observed in the symbiont-bearing tissues of hydrothermal vent tube worms and mussels; however their precise function within these symbioses remains unknown [4, 66].

Specific PGRP function can not be determined from sequence information alone and depends on the molecular context in which they are expressed. However, some assumptions can be made and are discussed in the following. OalgPGRP1, OalgPGRP3 and OalgPGRP5 contained N-terminal transmembrane domains (indicating that they are membrane integral), as well as novel cytoplasmic domains (Fig. 2). As is typical for PGRPs, the poorly conserved cytoplasmic domains had no similarity to known sequences [34]. PGRPs that integrate into the cell membrane and carry intracellular domains often induce an antimicrobial response by activating immune signaling pathways like Toll and IMD (immune deficiency) [67, 68]. However, some PGRP receptors bind peptidoglycans, but do not pass on an intracellular signal, thus effectively down-regulating the immune response and mediating tolerance towards resident bacteria [69].

insertion of two amino acids in the binding-groove region, and the OalgPGRP6 fragment did not contain the binding-relevant region.

Lectins

We detected six different classes of lectins in the transcriptome and proteome (Table 1, Table 4). They included C-type lectins, R-type lectins, fucoslectin, SUEL/rhamnose-binding lectins, galectins, a beta-1,3-glucan binding protein and fibrinogen-like proteins. Lectins are proteins with widely differing molecular structures and physiological functions. They are unified by their ability to strongly, yet reversibly, bind specific carbohydrate residues on the surfaces of cells and proteins, without exhibiting enzymatic activity [78].

Lectins are often associated with immune functions because of their molecular pattern recognition properties. For instance, they aid in microbe recognition and elimination through agglutination or direct antibacterial activity [79, 80], but, similar to PGRPs, are often also involved in modulating interactions between hosts and their beneficial symbionts. Lectins were, for example, shown to play major roles in symbiont acquisition and maintenance in sponges [81], corals [82, 83], clams [84], mice [85], and stilbonematid nematodes [86]. The sulfur-

oxidizing symbionts of stilbonematine nematodes are very closely related to the primary symbionts of gutless oligochaetes [7, 87]. However, the stilbonematine lectins have no notable sequence similarity to the *O. algarvensis* lectins, as expected given the long independent evolutionary histories of these two animal groups [87].

The domain architectures of *Olavius* lectins and their potential functions in host-symbiont interaction are summarized in Table 4. C-type lectins were particularly diverse, and 33 different forms were found in the transcriptome. Some of these C-type lectins have significant sequence similarity to lectins implicated in host-microbe interactions (Additional file 1: Table S10), for example to CD209 antigen-like proteins, macrophage mannose receptors, and C-type lectin receptor B – all MAMP receptors and phagocytosis enhancers of bacteria in vertebrates [88–90], and to immunolectin A, a microbe-inducible C-type lectin in *Manduca sexta* (tobacco hornworm) that is also involved in phagocytosis [91].

Another highly diverse group of lectins found in *O. algarvensis* were fibrinogen-related proteins (FREPs), which are almost exclusively involved in host-microbe interactions in invertebrates [92]. They were represented by 27 different unigenes (“components” in Trinity assembler terminology) in the transcriptome (Table 1,

Table 4 Lectins expressed in *Olavius algarvensis*

Lectin group	Transcripts a)	Proteins b)	Domain architectures	Potential functions
<i>Fucoslectins</i>	1	0		glycan recognition and host defense [126]
<i>C-type lectins</i>	33	5		glycan recognition [85], symbiont recognition and acquisition [127]
<i>R-type lectins</i>	6	4		glycan recognition [128], antimicrobial [129, 130], host defense [131]
<i>Galectins</i>	3	1		[132, 133], host defense [134, 135]
<i>SUEL rhamnose-binding lectins</i>	7	1		MAMP recognition [136, 137], egg fertilization [138], development [139]
<i>Fibrinogen-like proteins</i>	27	1		[92], MAMP recognition and antimicrobial activity [79], development [140], egg fertilization [141]



a) Number of transcripts, defined as Trinity components; see [13]
 b) Number of identified proteins in the proteomes

Table 4). For most of these, several isoforms with varying amino acid sequences were predicted, indicating that they may form an even more diverse array of proteins, possibly allowing very high specificity in the recognition of microbes.

Scavenger receptor cysteine rich proteins

In the transcriptomes we found a large group of sequences containing single or tandem scavenger receptor cysteine rich (SRCR) domains, often in association with other conserved domains, such as C-type lectin, trypsin, epidermal growth factor, low density lipoprotein (LDL) receptor, and immunoglobulin domains (Additional file 1: Figure S3). One of these proteins, which contained an additional universal stress protein A and four LDL receptor class B domains, was also identified in the proteome (Table 1).

The SRCR domain is an ancient and highly conserved module often found in proteins of the innate immune system that are involved in the recognition of microbial patterns and phagocytosis of bacteria in vertebrates [93]. In invertebrates, SRCR proteins have been implicated in host-symbiont interaction [94] and MAMP recognition [95].

Many SRCR sequences we identified had significant similarity to the MARCO scavenger receptor, DMBT1, CD163/M130, sea urchin scavenger receptors, and lamprey Pema-SRCR protein (Additional file 1: Table S11); all of these proteins are known or have been implicated to be involved in immune functions [93, 96]. Similar to the *Olavius* FREPs, the SRCR sequences identified in the transcriptome were represented by a considerable number of unigenes (FREP: 27, SRCR: 25), but many more different isoforms were predicted by the assembly. We therefore expect a high variability in the final proteins, possibly supporting highly specific recognition of microbes in *Olavius*, as has been observed in other invertebrates [97].

Toll-like receptors

We identified two Toll-like receptors (TLRs) consisting of the typical intracellular Toll/interleukin-1 receptor (TIR) homology domain and extracellular leucine- and cysteine-rich domains [98]. One of them was also detected in the proteome. Furthermore, we identified two sequences with only a TIR domain, one sequence with a TIR and transmembrane domain, and eight sequences containing leucine-rich repeats with high sequence similarity to TLRs from other animals and the variable lymphocyte receptors (VLRs) of agnate fish (Additional file 1: Table S12). VLRs are immune receptors that experience somatic recombination and convey a form of adaptive immunity in jawless vertebrates [99].

Toll-like receptors (TLRs) are microbial pattern recognition receptors and intracellular signaling transducers

that play a vital role in sensing and responding to microbiota in many animals [100]. They also play a role in many beneficial host-microbe symbioses [101, 102]. TLRs have long been thought to be absent from annelids [103, 104]. However, their presence and importance in host-microbe interactions has recently been recognized in polychaetes, leeches and earthworms [105, 106], where some were shown to be involved in the innate immune response against pathogens [107, 108] or were constitutively expressed in the gut [109].

We identified all the major components of the Toll signaling pathway in *O. algarvensis*, indicating that Toll signaling is active (Additional file 1: Table S13). We identified SARM (sterile alpha and TIR motif containing protein), an inhibitor of Toll signaling [110], that could aid in down-regulating the immune response against symbionts. Tollip, another inhibitor of Toll signaling [111], was also detected in the proteome, suggesting that these two inhibitors of Toll signaling may protect *O. algarvensis* against constant inflammation in response to its symbionts.

Interactions between symbionts and host may be regulated by different immune effectors and modulators

We detected several different types of antimicrobial proteins in the host transcriptome and proteome (Table 1), some of which were very abundant (Additional file 2: Table S8). The antimicrobials expressed in both transcriptome and proteome were lumbricin, an antimicrobial protein first discovered in earthworms [112], BPI (bactericidal permeability increasing protein), perforin/membrane attack complex-like proteins, insect defensin-like reeler proteins and cysteine-rich secretory proteins (Table 1). Antimicrobials combat infection by pathogenic microbes [113], but are also important in beneficial host-microbe interactions [85, 114], where they are used to modulate and control symbiont populations [115, 116]. In *O. algarvensis* they might be used to prevent symbionts and pathogens from invading non-symbiotic tissues, or to regulate symbiont growth.

Conclusions

This study provides insights into the physiological and molecular mechanisms that allow *Olavius algarvensis* to live in a stable beneficial association with its microbial consortium. Our results indicate that these animals have undergone a number of evolutionary changes in adaptation to their symbiotic lifestyle, apart from a complete reduction of the excretory and digestive organs. Examples of such adaptations are host proteins involved in symbiont digestion and nutrient uptake, with likely relocalization of the expression sites of some of these enzymes, and unconventional proteins for gas exchange and storage.

Since a mouth and anus are absent in gutless oligochaetes, and their epidermis is covered by a thick layer of symbionts, foreign microbes can only invade these hosts if they have the ability to penetrate the egg integument, or the cuticle in a juvenile or adult worm, and pass through the symbiont layer just under the worm's cuticle. As a result, the complexity of the *O. algarvensis* microbiome is quite low and consists primarily of its five symbiotic phylotypes. Despite this low microbial diversity, we found that *O. algarvensis* expresses a highly diverse array of pattern recognition receptors, comparable to other invertebrates that are associated with a much more complex community of microbes on their skin and in their digestive system. The high number of MAMP recognition proteins expressed in the transcriptome and proteome that clearly originated from different genes demonstrate the need of *Olavius algarvensis* to differentially sense and respond to both its symbiotic microbiota as well as environmental bacteria, although direct contact with the latter may be limited. The transcriptomes generated in this study contained small amounts of contamination with other *Olavius* species (0.1 – 3.5%) and minor contaminations are also expected to be present in the proteomes. Therefore, transcripts and proteins with very low expression levels should be treated with caution, as they alternatively may have originated from closely related *Olavius* species. Particularly, if several variants of a transcript or protein were expressed, it is possible that some of the variants that were considerably less abundant than the most abundant variant could be derived from the contaminating species.

This is also the first comprehensive transcriptomic and proteomic analysis of the innate immune system of a marine oligochaete. It shows how genes common to a wide array of invertebrates have evolved to enable the intricate communication and interactions that occur between animals and their symbiotic microbiota. The analyses described here lay the foundation for future experimental studies of immune processes and physiological responses that are essential in the functioning of this symbiosis.

Additional files

Additional file 1: Table S1. Summary of transcriptome sequencing. **Table S2.** Summary of transcriptome assembly and protein database. **Table S3.** Total protein of symbionts from fresh compared to starved whole worms. **Table S4.** Expression patterns of typical intestinal proteases in other animals. **Table S5.** High similarity of host digestive proteins to earthworm midgut enzymes. **Table S6.** Similarity of host digestive glucosidases to intestinal glucosidases of other animals. **Table S7.** *O. algarvensis* giant extracellular hemoglobin sequences. **Table S9.** Hemerythrin expression in annelids. **Table S10.** Host c-type lectins with similarity to known immune lectins. **Table S11.** Host SRCR proteins with similarity to SRCR proteins involved in immune processes. **Table S12.** Host Toll-like receptor sequences. **Table S13.** Toll immune signaling

pathway in *Olavius*. **Figure S1.** Transcriptome annotation statistics.

Figure S2: Hydrophobicity cluster analysis plots of annelid hemoglobin chains. **Figure S3.** Domain structures of proteins with scavenger domains. (DOCX 920 kb)

Additional file 2: Table S8. Proteins potentially involved in symbiont interaction in *O. algarvensis*, all transcripts and proteins. (XLSX 92 kb)

Additional file 3: Table S14. Subcellular localization evidence of host digestive proteins. (XLSX 17 kb)

Abbreviations

AMP: Antimicrobial protein/peptide; cDNA: Complementary DNA; CDS: Coding DNA sequence; CO: Carbon monoxide; FREP: Fibrinogen related protein; PGRP: Peptidoglycan recognition protein; SRCR: Scavenger receptor cysteine-rich; TLR: Toll-like receptor

Acknowledgements

We thank the team of the HYDRA Institute on Elba for their extensive support with sample collection and on site experiments, and Silke Wetzel for excellent technical assistance. Sequencing was performed by the Max Planck Society Genome Center at the Max Planck Institute for Plant Breeding in Cologne, Germany.

Funding

The study was funded by the Max Planck Society and by the Gordon and Betty Moore Foundation through Grant GBMF3811 to ND. MK was supported by a PhD scholarship of the Studienstiftung des Deutschen Volkes and a NSERC Banting Postdoctoral Fellowship.

Availability of data and materials

The datasets supporting the conclusions of this article are available from the following repositories and its Additional files 1, 2 and 3: Assembled transcript sequences are available from the European Nucleotide Archive (ENA) under the accession numbers HACZ01000001-HACZ01173602 (TSA project HACZ01000000 data, <http://www.ebi.ac.uk/ena/data/view/HACZ010000000>). Raw reads were deposited under the study ID PRJEB10952 (<http://www.ebi.ac.uk/ena/data/view/PRJEB10952>). The complete protein sequence database is available from the MassIVE data repository under accession MSV000079512 and available for download via FTP: <ftp://MSV000079512@massive.ucsd.edu>. All proteomics data sets used in this study were deposited at the MassIVE data repository under accession numbers: MSV000079512 [MassIVE] & PXD003626 [ProteomeXchange] and available for download via FTP: <ftp://MSV000079512@massive.ucsd.edu>.

Authors' contributions

JW and MK contributed equally to this manuscript. JW conceived and wrote the manuscript, analyzed and interpreted the proteomic data and provided ideas, did in-depth bioinformatic analyses of proteins of interest, prepared all figures and tables, collected all worms for transcriptomics experiments, conceived and performed all transcriptomics experiments, assembled and annotated the transcriptomes, predicted CDS from transcriptome data and did all transcriptome statistics and analyses. MK conceived the study and the manuscript, edited the manuscript and provided ideas, conceived proteomics experiments, collected worms for proteomics experiments and performed two of the starvation experiments, compiled the protein reference database, did statistical analyses and processed and analyzed the proteomics data. CL provided ideas and collected worms for proteomics and transcriptomics experiments and performed one of the starvation experiments. AG provided all microscopic images used in Fig. 1, and commented on the manuscript. PEA and RJG processed proteomics data and commented on the manuscript. JCY provided ideas and ran 2D-LC-MS/MS experiments. RLH provided access to the proteomics equipment, provided conceptual input and coordinated the data processing at the ORNL. ND was involved in the organization and coordination of this study, provided ideas and commented on the manuscript. All authors reviewed and revised the final manuscript before submission.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The animals used in this study underlie no national or international guidelines or regulations. Consent to participate: not applicable.

Author details

¹Symbiosis Department, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany. ²Energy Bioengineering and Geomicrobiology Research Group, University of Calgary, Calgary T2N 1N4, AB, Canada. ³HYDRA Institute for Marine Sciences, Elba Field Station, Via del Forno 80, 57034 Campo nell' Elba, (LI), Italy. ⁴Oak Ridge National Laboratory, Chemical Sciences Division, Oak Ridge, Tennessee, 1 Bethel Valley Rd, Oak Ridge, TN 37831, USA. ⁵Present Address: Saul Ewing LLP, 1500 Market Street, 37th Floor, Philadelphia, PA 19102-2186, USA. ⁶Symbiosis Department, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany.

Received: 4 June 2016 Accepted: 15 November 2016

Published online: 21 November 2016

References

- Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science*. 2012;336:1268–73.
- McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A*. 2013;110:3229–36.
- Gilbert SF, Bosch TC, Ledón-Rettig C. Eco-Evo-Devo: Developmental symbiosis and developmental plasticity as evolutionary agents. *Nat Rev Genet*. 2015;16:611–22.
- Nyholm SV, Song P, Dang J, Bunce C, Girguis PR. Expression and putative function of innate immunity genes under *in situ* conditions in the symbiotic hydrothermal vent tubeworm *Ridgeia piscesae*. *PLoS One*. 2012;7:e38267.
- Chu H, Mazmanian SK. Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat Immunol*. 2013;14:668–75.
- Dubilier N, Mülders C, Ferdelman T, de Beer D, Pernthaler A, Klein M, Wagner M, Erséus C, Thiermann F, Krieger J, Giere O, Amann R. Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature*. 2001;411:298–302.
- Dubilier N, Blazejak A, Rühland C. Symbioses between bacteria and gutless marine oligochaetes. In: Overmann J, editor. *Molecular Basis of Symbiosis*. Berlin Heidelberg: Springer; 2005. p. 251–75.
- Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson JJ, Barry KW, Shapiro HJ, Szeto E, Kyripides NC, Musmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature*. 2006;443:950–5.
- Kleiner M, Wentrup C, Lott C, Teeling H, Wetzels S, Young J, Chang YJ, Shah M, Verberkmoes NC, Zarzycki J, Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebecke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier N. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci U S A*. 2012;109:E1173–82.
- Ruehland C, Blazejak A, Lott C, Loy A, Erséus C, Dubilier N. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environ Microbiol*. 2008;10:3404–16.
- Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, Littmann S, Kuypers MM, Dubilier N. Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ Microbiol*. 2015;17:5023–35.
- Harrison P, Seemann T. From high-throughput sequencing read alignments to confident, biologically relevant conclusions with Nasoni. 2009. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.509.7046&rep=rep1&type=pdf>. Accessed 25 Apr 2016
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 2011;29:644.
- Li W, Li S, Zhong J, Zhu Z, Liu J, Wang W. A novel antimicrobial peptide from skin secretions of the earthworm, *Pheretima guillelmi* (Michaelsen). *Peptides*. 2011;32:1146–50.
- Conesa A, Götz S, Garcá-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674–6.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
- Bailly X, Jollivet D, Vanin S, Deutsch J, Zal F, Lallier F, Toulmond A. Evolution of the sulfide-binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. *Mol Biol Evol*. 2002;19:1421–33.
- Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Moron J. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci*. 1997;53:621–45.
- Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol*. 2011;12:R44.
- Thompson MR, Chourey K, Froelich JM, Erickson BK, Verberkmoes NC, Hettich RL. Experimental approach for deep proteome measurements from small-scale microbial biomass samples. *Anal Chem*. 2008;80:9517–25.
- Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*. 2001;19:242–7.
- McDonald WH, Oh I, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int J Mass Spectrom*. 2002;219:245–51.
- Gouzy J, Carrere S, Schiex T. FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics*. 2009;25:670–1.
- Rice P, Longden I, Bleasby A. EMBL-EBI: the European molecular biology open software suite. *Trends Genet*. 2000;16:276–7.
- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
- Tabb DL, Fernando CG, Chambers MC. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res*. 2007;6:654–61.
- Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL. IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J Proteome Res*. 2009;8:3872–81.
- Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
- Abraham P, Adams R, Giannone RJ, Kalluri U, Ranjan P, Erickson B, Shah M, Tuskan GA, Hettich RL. Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of *Populus* using shotgun proteomics. *J Proteome Res*. 2011;11:449–60.
- Giannone RJ, Huber H, Karpinetz T, Heimerl T, Küper U, Rachel R, Keller M, Hettich RL, Podar M. Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans*-*Ignicoccus hospitalis* relationship. *PLoS One*. 2011;6:e22942.
- Zybailov B, Mosley AL, Sardu ME, Coleman MK, Florens L, Washburn MP. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *J Proteome Res*. 2006;5:2339–47.
- Andrade SC, Novo M, Kawachi GY, Worsaae K, Pleijel F, Giribet G, Rouse GW. Articulating “archiannelids”: Phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. *Mol Biol Evol*. 2015; 32(11):2860–75.
- Cavanaugh CM, McKiness ZP, Newton IL, Stewart FJ. Marine chemosynthetic symbioses. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The Prokaryotes*. Berlin Heidelberg: Springer; 2013. p. 579–607.
- Dziarski R, Gupta D. The peptidoglycan recognition proteins (GPRs). *Genome Biol*. 2006;7:232.
- Van Herreweghe JM, Michiels CW. Invertebrate lysozymes: diversity and distribution, molecular mechanism and *in vivo* function. *J Biosci*. 2012;37:327–48.
- Mellroth P, Karlsson J, Steiner H. A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J Biol Chem*. 2003;278:7059–64.

37. Giere O, Erséus C, Stuhlmacher F. A new species of *Olavius* (Tubificidae) from the Algarve coast in Portugal, the first East Atlantic gutless oligochaete with symbiotic bacteria. *Zool Anz.* 1998;237:209–14.
38. Royer WE, Sharma H, Strand K, Knapp JE, Bhyravhatla B. *Lumbricus erythroruorin* at 3.5 Å resolution: architecture of a megadalton respiratory complex. *Structure.* 2006;14:1167–77.
39. Fushitani K, Matsuuru M, Riggs A. The amino acid sequences of chains A, B, and C that form the trimer subunit of the extracellular hemoglobin from *Lumbricus terrestris*. *J Biol Chem.* 1988;263:6502–17.
40. Concha NO, Herzberg O, Rasmussen BA, Bush K. Crystal structures of the cadmium- and mercury-substituted metallo- β -lactamase from *Bacteroides fragilis*. *Protein Sci.* 1997;6:2671–6.
41. Leder A, Wiener E, Lee MJ, Wickramasinghe SN, Leder P. A normal beta-globin allele as a modifier gene ameliorating the severity of alpha-thalassemia in mice. *Proc Natl Acad Sci U S A.* 1999;96:6291–5.
42. Zal F, Leize E, Lallier FH, Toulmond A, Van Dorsselaer A, Childress JJ. S-Sulfohemoglobin and disulfide exchange: the mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. *Proc Natl Acad Sci U S A.* 1998;95:8997–9002.
43. Childress J, Fisher C, Favuzzi J, Kochevar R, Sanders N, Alayse AM. Sulfide-driven autotrophic balance in the bacterial symbiont-containing hydrothermal vent tubeworm, *Riftia pachyptila* Jones. *Biol Bull.* 1991;180:135–53.
44. Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, Schaeffer SW, Royer WE. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc Natl Acad Sci U S A.* 2005;102:2713–8.
45. Flores JF, Hourdez SM. The zinc-mediated sulfide-binding mechanism of hydrothermal vent tubeworm 400-kDa hemoglobin. *Cah Biologie.* 2006;47:371.
46. Chabasse C, Bailly X, Rousselot M, Zal F. The multigenic family of the extracellular hemoglobin from the annelid polychaete *Arenicola marina*. *Comp Biochem Physiol B.* 2006;144:319–25.
47. Kurtz Jr D. Molecular structure/function relationships of hemerythrins. In: Mangum CP, editor. *Blood and Tissue Oxygen Carriers*. Berlin Heidelberg: Springer; 1992. p. 151–71.
48. Manwell C, Baker CA. *Magelona* haemerythrin: tissue specificity, molecular weights and oxygen equilibria. *Comp Biochem Physiol B.* 1988;89:453–63.
49. Baert JL, Britel M, Sautière P, Malecha J. Ovohemerythrin, a major 14-kDa yolk protein distinct from vitellogenin in leech. *Eur J Biochem.* 1992;209:563–9.
50. Nejmeddine A, Wouters-Tyrou D, Baert JL, Sautière P. Primary structure of a myohemerythrin-like cadmium-binding protein, isolated from a terrestrial annelid oligochaete. *C R Acad Sci.* 1997;320:459–68.
51. Deloffre L, Salzet B, Vieau D, Andries JC, Salzet M. Antibacterial properties of hemerythrin of the sand worm *Nereis diversicolor*. *Neuroendocrinol Lett.* 2003;24:39–45.
52. Florin K. Recherches sur les hémérythrine. In: *Archives Internationales de Physiologie*. Liège: Vaillant-Carmanne; 1933. p. 247–329.
53. Haldane J, Smith JL. The absorption of oxygen by the lungs. *J Physiol.* 1897;22:231.
54. Rossi-Fanelli A, Antonini E. Studies on the oxygen and carbon monoxide equilibria of human myoglobin. *Arch Biochem Biophys.* 1958;77:478–92.
55. Kubo M. The oxygen equilibrium of hemerythrin and Bohr effect. *Bull Chem Soc Jpn.* 1953;26:189–92.
56. Ordway GA, Garry DJ. Myoglobin: an essential hemoprotein in striated muscle. *J Exp Biol.* 2004;207:3441–6.
57. Mangum CP. Physiological function of the hemerythrins. In: Mangum CP, editor. *Blood and Tissue Oxygen Carriers*. Berlin Heidelberg: Springer; 1992. p. 173–92.
58. Sanchez S, Hourdez S, Lallier FH. Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. *BMC Genomics.* 2007;8:1.
59. Medzhitov R. Recognition of microorganisms and activation of the immune response. *Nature.* 2007;449:819–26.
60. Mackey D, McFall AJ. MAMPs and MIMPs: proposed classifications for inducers of innate immunity. *Mol Microbiol.* 2006;61:1365–71.
61. Royet J, Gupta D, Dziarski R. Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat Rev Immunol.* 2011;11:837–51.
62. Wang J, Wu Y, Yang G, Aksoy S. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proc Natl Acad Sci U S A.* 2009;106:12133–8.
63. Troll JV, Bent EH, Pacquette N, Wier AM, Goldman WE, Silverman N, McFall-Ngai MJ. Taming the symbiont for coexistence: a host PGRP neutralizes a bacterial symbiont toxin. *Environ Microbiol.* 2010;12:2190–203.
64. Anselme C, Vallier A, Balmand S, Fauvarque MO, Heddi A. Host PGRP gene expression and bacterial release in endosymbiosis of the weevil *Sitophilus zeamais*. *Appl Environ Microbiol.* 2006;72:6766–72.
65. Saha S, Jing X, Park SY, Wang S, Li X, Gupta D, Dziarski R. Peptidoglycan recognition proteins protect mice from experimental colitis by promoting normal gut flora and preventing induction of interferon- γ . *Cell Host Microbe.* 2010;8:147–62.
66. Bettencourt R, Rodrigues M, Barros I, Cerqueira T, Freitas C, Costa V, Pinheiro M, Egas C, Santos RS. Site-related differences in gene expression and bacterial densities in the mussel *Bathymodiolus azoricus* from the Menez Gwen and Lucky Strike deep-sea hydrothermal vent sites. *Fish Shellfish Immunol.* 2014;39:343–53.
67. Leulier F, Parquet C, Pili-Floury S, Ryu JH, Caroff M, Lee WJ, Mengin-Lecreux D, Lemaître B. The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. *Nature Immunol.* 2003;4:478–84.
68. Kaneko T, Goldman WE, Mellroth P, Steiner H, Fukase K, Kusumoto S, Harley W, Fox A, Golenbock D, Silverman N. Monomeric and polymeric gram-negative peptidoglycan but not purified LPS stimulate the *Drosophila* IMD pathway. *Immunity.* 2004;20:637–49.
69. Maillet F, Bischoff V, Vignal C, Hoffmann J, Royet J. The *Drosophila* peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK pathway activation. *Cell Host Microbe.* 2008;3:293–303.
70. Michel T, Reichhart JM, Hoffmann JA, Royet J. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature.* 2001;414:756–9.
71. Lu X, Wang M, Qi J, Wang H, Li X, Gupta D, Dziarski R. Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *J Biol Chem.* 2006;281:5895–907.
72. Tydel CC, Yuan J, Tran P, Selsted ME. Bovine peptidoglycan recognition protein-S: antimicrobial activity, localization, secretion, and binding properties. *J Immunol.* 2006;176:1154–62.
73. Zaidman-Rémy A, Hervé M, Poidevin M, Pili-Floury S, Kim MS, Blanot D, Oh BH, Ueda R, Mengin-Lecreux D, Lemaître B. The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. *Immunity.* 2006;24:463–73.
74. Wang ZM, Li X, Cocklin RR, Wang M, Wang M, Fukase K, Inamura S, Kusumoto S, Gupta D, Dziarski R. Human peptidoglycan recognition protein-L is an N-acetylmuramoyl-L-alanine amidase. *J Biol Chem.* 2003;278:49044–52.
75. Garver LS, Wu J, Wu LP. The peptidoglycan recognition protein PGRP-SC1a is essential for Toll signaling and phagocytosis of *Staphylococcus aureus* in *Drosophila*. *Proc Natl Acad Sci U S A.* 2006;103:660–5.
76. Swaminathan CP, Brown PH, Roychowdhury A, Wang Q, Guan R, Silverman N, Goldman WE, Boons GJ, Mariuzzo RA. Dual strategies for peptidoglycan discrimination by peptidoglycan recognition proteins (PGRPs). *Proc Natl Acad Sci U S A.* 2006;103:684–9.
77. Schleifer KH, Kandler O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev.* 1972;36:407.
78. Varki A, Cummings R, Esko J, Freeze H, Stanley P, Bertozzi CR, Hart GW, Etzler ME. *Essentials of Glycobiology*. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 2009.
79. Fan C, Zhang S, Li L, Chao Y. Fibrinogen-related protein from amphioxus *Branchiostoma belcheri* is a multivalent pattern recognition receptor with a bacteriolytic activity. *Mol Immunol.* 2008;45:3338–46.
80. Mukherjee S, Zheng H, Derebe MG, Callenberg KM, Partch CL, Rollins D, Propher DC, Rizo J, Grabe M, Jiang QX, Hooper LV. Antibacterial membrane attack by a pore-forming intestinal C-type lectin. *Nature.* 2014;505:103–7.
81. Müller W, Zahn R, Kurelec B, Lucu C, Müller I, Uhlenbruck G. Lectin, a possible basis for symbiosis between bacteria and sponges. *J Bacteriol.* 1981;145:548–58.
82. Wood-Charlson EM, Hollingsworth LL, Krupp DA, Weis VM. Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell Microbiol.* 2006;8:1985–93.
83. Kvennefors ECE, Leggat W, Hoegh-Guldberg O, Degnan BM, Barnes AC. An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. *Dev Comp Immunol.* 2008;32:1582–92.

84. Gourdine JP, Smith-Ravin EJ. Analysis of a cDNA-derived sequence of a novel mannose-binding lectin, codakine, from the tropical clam *Codakia orbicularis*. *Fish Shellfish Immunol.* 2007;22:498–509.
85. Cash HL, Whitham CV, Behrendt CL, Hooper LV. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science.* 2006;313:1126–30.
86. Bulgheresi S, Gruber-Vodicka HR, Heindl NR, Dirks U, Kostadinova M, Breiteneder H, Ott JA. Sequence variability of the pattern recognition receptor Mermaid mediates specificity of marine nematode symbioses. *ISME J.* 2011;5:986–98.
87. Zimmermann J, Wenstrup C, Sadowski M, Blazejak A, Gruber-Vodicka H, Kleiner M, Ott J, Cronholm B, De Wit P, Erséus C, Dubilier N. Closely coupled evolutionary history of ecto- and endosymbionts from two distantly-related animal phyla. *Mol Ecol.* 2016; doi: 10.1111/mec.13554
88. Park CG, Takahara K, Umemoto E, Yashima Y, Matsubara K, Matsuda Y, Clausen BE, Inaba K, Steinman RM. Five mouse homologues of the human dendritic cell C-type lectin, DC-SIGN. *Int Immunol.* 2001;13:1283–90.
89. Ezekowitz R, Sastry K, Bailly P, Warner A. Molecular characterization of the human macrophage mannose receptor: demonstration of multiple carbohydrate recognition-like domains and phagocytosis of yeasts in Cos-1 cells. *J Exp Med.* 1990;172:1785–94.
90. Soanes KH, Figueroide K, Richards RC, Mattatall NR, Ewart KV. Sequence and expression of C-type lectin receptors in Atlantic salmon (*Salmo salar*). *Immunogenetics.* 2004;56:572–84.
91. Yu XQ, Gan H, Kanost MR. Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. *Insect Biochem Mol Biol.* 1999;29:585–97.
92. Hanington PC, Zhang SM. The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. *J Innate Immun.* 2010;3:17–27.
93. Sarrías MR, Gronlund J, Padilla O, Madsen J, Holmskov U, Lozano F. The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. *Crit Rev Immunol.* 2004;24:1–37.
94. Steindler L, Schuster S, Ilan M, Avni A, Cerrano C, Beer S. Differential gene expression in a marine sponge in relation to its symbiotic state. *Mar Biotechnol.* 2007;9:543–9.
95. Liu F, Li J, Fu J, Shen Y, Xu X. Two novel homologs of simple C-type lectin in grass carp (*Ctenopharyngodon idellus*): potential role in immune response to bacteria. *Fish Shellfish Immunol.* 2011;31:765–73.
96. Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proc Natl Acad Sci.* 2000;97:13156–61.
97. Buckley KM, Rast JP. Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev Comp Immunol.* 2015;49:179–89.
98. Zheng L, Zhang L, Lin H, McIntosh M, Malacrida A. Toll-like receptors in invertebrate innate immunity. *Invert Surviv J.* 2005;2:105–13.
99. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature.* 2004;430:174–80.
100. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol.* 2010;11:373–84.
101. Kubinak JL, Round JL. Toll-like receptors promote mutually beneficial commensal-host interactions. *PLoS Pathogens.* 2012;8:e1002785.
102. Venkatesh M, Mukherjee S, Wang H, Li H, Sun K, Benechet AP, Qiu Z, Maher L, Redinbo MR, Phillips RS, Fleet JC. Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity.* 2014;41:296–310.
103. Cooper E, Kvell K, Engelmann P, Nemeth P. Still waiting for the Toll? *Immunol Lett.* 2006;104:18–28.
104. Francis J, Wreesman S, Yong S, Reigstad K, Krutzik S, Cooper EL. Analysis of the earthworm coelomocyte cell surface for the presence of Toll-like immune receptors. *Eur J Soil Biol.* 2007;43:592–6.
105. Davidson CR, Best NM, Francis JW, Cooper EL, Wood TC. Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). *Dev Comp Immunol.* 2008;32:608–12.
106. Halaných KM, Kocot KM. Repurposed transcriptomic data facilitate discovery of innate immunity toll-like receptor (TLR) genes across lophotrochozoa. *Biol Bull.* 2014;227:201–9.
107. Cuvillier-Hot V, Boidin-Wichlacz C, Slomianny C, Salzet M, Tasiemski A. Characterization and immune function of two intracellular sensors, HmTLR1 and HmNLR, in the injured CNS of an invertebrate. *Dev Comp Immunol.* 2011;35:214–26.
108. Fjøsnes TF, Stenseth EB, Myrsmølién F, Rudi K. Gene expression of TLR homologues identified by genome-wide screening of the earthworm *Dendrobaena veneta*. *Innate Immun.* 2014;21:161–6.
109. Škanta F, Roubalová R, Dvořák J, Procházková P, Bilej M. Molecular cloning and expression of TLR in the *Eisenia andrei* earthworm. *Dev Comp Immunol.* 2013;41:694–702.
110. Zhang Q, Zmasek CM, Cai X, Godzik A. TIR domain-containing adaptor SARM is a late addition to the ongoing microbe-host dialog. *Dev Comp Immunol.* 2011;35:461–8.
111. Zhang G, Ghosh S. Negative regulation of Toll-like receptor-mediated signaling by Tollip. *J Biol Chem.* 2002;277:7059–65.
112. Cho JH, Park CB, Yoon YG, Kim SC. Lumbricin I, a novel proline-rich antimicrobial peptide from the earthworm: purification, cDNA cloning and molecular characterization. *Biochim Biophys Acta.* 1998;1408:67–76.
113. Vizioli J, Salzet M. Antimicrobial peptides from animals: focus on invertebrates. *Trends Pharmacol Sci.* 2002;23:494–6.
114. Hooper LV. Do symbiotic bacteria subvert host immunity? *Nat Rev Microbiol.* 2009;7:367–74.
115. Vaishnava S, Behrendt CL, Ismail AS, Eckmann L, Hooper LV. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc Natl Acad Sci U S A.* 2008;105:20858–63.
116. Login FH, Balmant S, Vallier A, Vincent-Monégat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A. Antimicrobial peptides keep insect endosymbionts under control. *Science.* 2011;334:362–5.
117. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
118. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Intell Conf Syst Mol Biol.* 1998;6:175–82.
119. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8:785–6.
120. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol.* 2004;338:1027–36.
121. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 2000;300:1005–16.
122. Baú D, Martin AJM, Mooney C, Vullo A, Walsh I, Pollastrì G. Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. *BMC Bioinform.* 2006;7:402.
123. Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, Altermann U, Angerer P, Ansorge S, Balasz K, Bernhofer M, Betz A, Cizmadjia L, Do KT, Gerke J, Greil R, Joerdens V, Hastreiter M, Hembach K, Herzog M, Kalemánov M, Kluge M, Meier A, Nasir H, Neumaier U, Prade V, Reeb J, Sorokoumov A, Troshani I, Vorberg S, Waldraff S, Zierer J, Nielsen H, Rost B. LocTree3 prediction of localization. *Nucleic Acids Res.* 2014;42:W350–5.
124. Pierleoni A, Martelli PL, Fariselli P, Casadio R. BaCello: a balanced subcellular localization predictor. *Bioinformatics.* 2006;22:e408–16.
125. Lin WZ, Fang JA, Xiao X, Chou KC. iLoc-Animal: a multi-label learning classifier for predicting subcellular localization of animal proteins. *Mol Biosyst.* 2013;9:634–44.
126. Honda S, Kashiwagi M, Miyamoto K, Takei Y, Hirose S. Multiplicity, structures, and endocrine and exocrine natures of eel fucose-binding lectins. *J Biol Chem.* 2000;275:33151–7.
127. Bulgheresi S, Schabussova I, Chen T, Mullin NP, Maizels RM, Ott JA. A new C-type lectin similar to the human immunoreceptor DC-SIGN mediates symbiont acquisition by a marine nematode. *Appl Environ Microbiol.* 2006;72:2950–6.
128. Katzin BJ, Collins EJ, Robertus JD. Structure of ricin A-chain at 2.5 Å. *Proteins.* 1991;10:251–9.
129. Nakano M, Tabata S, Sugihara K, Kouzuma Y, Kimura M, Yamasaki N. Primary structure of hemolytic lectin CEL-III from marine invertebrate *Cucumaria echinata* and its cDNA: structural similarity to the B-chain from plant lectin, ricin. *Biochim Biophys Acta.* 1999;1435:167–76.
130. Hisamatsu K, Nagao T, Unno H, Goda S, Hatakeyama T. Identification of the amino acid residues involved in the hemolytic activity of the *Cucumaria echinata* lectin CEL-III. *Biochim Biophys Acta.* 1830;2013:4211–7.
131. Kim DH, Patnaik BB, Seo GW, Kang SM, Lee YS, Lee BL, Han YS. Identification and expression analysis of a novel R-type lectin from the coleopteran beetle, *Tenebrio molitor*. *J Invertebr Pathol.* 2013;114:226–9.
132. Yang RY, Rabinovich GA, Liu FT. Galectins: structure, function and therapeutic potential. *Expert Rev Mol Med.* 2008;10:e17.

133. Vasta GR. Galectins as pattern recognition receptors: structure, function, and evolution. *Adv Exp Med Biol.* 2012;946:21–36.
134. Bao Y, Shen H, Zhou H, Dong Y, Lin Z. A tandem-repeat galectin from blood clam *Tegillarca granosa* and its induced mRNA expression response against bacterial challenge. *Genes Genom.* 2013;35:733–40.
135. Shi XZ, Wang L, Xu S, Zhang XW, Zhao XF, Vasta GR, Wang JX. A galectin from the kuruma shrimp (*Marsupenaeus japonicus*) functions as an opsonin and promotes bacterial clearance from hemolymph. *PLoS One.* 2014;9:e91794.
136. Tateno H, Ogawa T, Muramoto K, Kamiya H, Saneyoshi M. Rhamnose-binding lectins from steelhead trout (*Oncorhynchus mykiss*) eggs recognize bacterial lipopolysaccharides and lipoteichoic acid. *Biosci Biotechnol Biochem.* 2002;66:604–12.
137. Watanabe Y, Tateno H, Nakamura-Tsuruta S, Kominami J, Hirabayashi J, Nakamura O, Watanabe T, Kamiya H, Naganuma T, Ogawa T, Naudé RJ, Muramoto K. The function of rhamnose-binding lectin in innate immunity by restricted binding to Gb3. *Dev Comp Immunol.* 2009;33:187–97.
138. Mourão PAS. A carbohydrate-based mechanism of species recognition in sea urchin fertilization. *Braz J Med Biol Res.* 2007;40:5–17.
139. Ozeki Y, Yokota Y, Kato KH, Titani K, Matsui T. Developmental expression of D-galactoside-binding lectin in sea urchin (*Anthocidaris crassispina*) eggs. *Exp Cell Res.* 1995;216:318–24.
140. Yamada S, Hotta K, Yamamoto TS, Ueno N, Satoh N, Takahashi H. Interaction of notochord-derived fibrinogen-like protein with Notch regulates the patterning of the central nervous system of *Ciona intestinalis* embryos. *Dev Biol.* 2009;328:1–12.
141. Harada Y, Takagaki Y, Sunagawa M, Saito T, Yamada L, Taniguchi H, Shoguchi E, Sawada H. Mechanism of self-sterility in a hermaphroditic chordate. *Science.* 2008;320:548–50.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

