SUPPLEMENTARY FIGURES:

Supplementary Figure 1, Related to Figure 1:

Squid light organ transcriptome construction

(A-B) Quality controls for RNAseq collection. (A) Number of bacteria aggregating on the light organ surface (9 replicates of 4-13 squid). (B) Number of hemocytes trafficking to the appendage of the light organ (13 replicates of 4-7 squid). Box-plots represent median, interquartiles, and 5-95 % (GLM, Poisson distribution; **, $p \leq 0.01$).

(C) Details of libraries sequenced by 454 technology.

(D-L) Statistics from the Blast2go annotation. (D) Sequence similarity distribution during BLAST annotation. (E) E-value distribution during BLAST annotation. (F) Species distribution during BLAST annotation. (G) Top-Hit species distribution during BLAST annotation. (H) Mapping database sources used during the BLAST annotation process. UniProtKB: UniProt Knowledgebase, MGI: Mouse-Genome Informatics, RGD: Rat Genome Database, ZFIN: Zebrafish Model Organism Database, FB: FlyBase, RefSeq: Reference Sequences NCBI, WB: Wormbase, TAIR: The *Arabidopsis* Information Resource, GR_protein: Gramene Proteins, GeneDB_Tbrucei: GeneDataBase *Trypanosoma brucei,* TIGR_CMR: TIGR Comprehensive Microbial Resource; PseudoCAP: *Pseudomonas* Community Annotation Project. (I) Evidence code distribution for annotated sequences. IEA: Inferred from Electronic Annotation, IDA: Inferred from Direct Assay, TAS: Traceable Author Statement, IMP: Inferred from Mutant Phenotype, ISO: Inferred from Sequence Orthology, ND: No biological Data available, ISS: Inferred from Sequence or Structural Similarity, IPI: Inferred from Physical Interaction, NAS: Non-traceable Author Statement, EXP: Inferred from Experiment, IEP: Inferred from Expression Pattern, IGI: Inferred from Genetic Interaction, IC: Inferred by Curator, ISA: Inferred from Sequence Alignment, RCA: inferred from

Reviewed Computational Analysis, ISM: Inferred from Sequence Model. (J) Annotation distribution of GO terms per isotigs. (K) Level distribution of the GO terms associated to isotig annotations (green: biological processes, blue: molecular functions, yellow: cellular localizations). Low-level numbers are more general and high-level numbers are more specific.

(L) Distribution of annotated isotigs in Clusters of Orthologous Groups (COGs).

Supplementary Figure S2, Related to Figure 3:

Eschitotriosidase: sequence characterization, and negative controls for *in situ* hybridization and immunocytochemistry.

(A) Alignment of Eschitotriosidase (EsChit) sequence. *Euprymna scolopes* chitotriosidase is composed of a signal peptide (purple), a chitotriosidase catalytic domain (blue), and two chitin-binding/peritrophin A domains (red) separated by regions enriched in Proline, Glycine, Serine and Threonine. Catalytic residues are circled in blue, and cysteines of the chitin-binding/peritrophin A domains are circled in red. Alignment was performed using Muscle (Edgar, 2004) and visualized by CLC Viewer (amino acids are color coded by polarity, dots correspond to similar amino acids, dashes represent a gap character). Es: *Euprymna scolopes* (this study), Cg: *Crassostrea gigas* (gi|405974134), Sk: *Saccoglossus kowalevskii* (gi|291240487), Bf: *Branchiostoma floridae* (gi|260805432), Hs: *Homo sapiens* (gi|47480958).

(B) Negative control of *in situ* hybridization of Eschit in the light organ (sense probe). No purple staining is visible in the proximity of the pores (triangles, inset), in the epithelium of the appendages, and in the hindgut.

(C) Coomassie staining of an SDS-PAGE gel of squid-tissue soluble-protein extract (left, same as in Figure 3E) and its associated western blot against non-specific rabbit IgG (Genscript), which was used because the polyclonal antibody was generated in rabbit. (D-F) Immunocytochemistry of rabbit IgG in the light organ (D). No stain is visible in the ciliated field and the cytosol of the appendage epithelium (E), nor in the pores/duct region (F). Colored boxes in (D) show representative locations of pictures enlarged in panels (E) and (F). Green, α-rabbit IgG; red, rhodamine phalloidin (f-actin); blue, TOTO-3 (nuclei). (G) Immunocytochemistry of rabbit IgG in the mucus coating the appendage epithelium. The inset magnifies the region highlighted by the box. Green, α-rabbit IgG; blue, WGA (mucus).

Supplementary Figure S3, Related to Figure 6:

***V. fischeri* does not robustly undergo chemotaxis towards chitobiose under laboratory conditions.**

Chemotaxis of *V. fischeri* towards chitobiose, when directly exposed to 100 μM, 1 mM or 10 mM chitobiose (mean ± SE). A non-chemotactic mutant (strain MB08701; ES114 *cheA*::Tn*erm* (Mandel et al., 2012)) was used as a control for non-directed motility. To confirm the chemotactic ability of the wild-type strain, chemotaxis towards 1 mM GlcNAc was determined. Measurements were made in duplicate and the experiment was performed three times.

Supplementary Table S1, Related to Figure 2:

Differentially represented genes from the 454 database.

List of differentially represented annotated isotigs between libraries from aposymbiotic and symbiotic squid light organs (sheet 1), libraries from hatchling and aposymbiotic squid light organs (sheet 2), and libraries from hatchling and symbiotic squid light organs (sheet 3),

based on the Stekel's method (detail of statistic calculation process in sheet 4). Isogroups in green have a Rj > 3, whereas those in light green have a Rj > 2. Isogroups in bold are those that have been studied by qRT-PCR. Isogroups in italic have several isotigs with different annotations, which have been pooled in the table, as well as the GO terms associated with. See excel file TableS1_Kremer_et_al.xlsb

SUPPLEMENTAL EXPERIMENTAL PROCEDURES:

General procedures

- Adult *E. scolopes* were bred in the laboratory in a circulating system of Filter-Sterilized Instant Ocean (FSIO, Aquarium systems), maintained at 24 °C on a 12 h/12 h light/dark cycle. In experiments with hatchling squid, released from clutches laid by the collected captive females, the animals were anesthetized in 2 % ethanol in FSIO prior to sacrifice.
- *V. fischeri* cells were grown in LBS medium (Luria Bertani with 2 % wt/vol NaCl) with shaking at 28 °C to an $OD_{600nm} \sim 0.2$ - 0.4, unless otherwise noted. The wild-type strain used in this study is the GFP-labeled ES114 strain (Boettcher and Ruby, 1990). Squid were infected with an inoculum of $9.43 \pm 0.9 \times 10^3$ CFU/mL.
- Confocal experiments were performed on a Zeiss 510 laser scanning confocal microscope. Other microscopic observations were performed on an epifluorescence microscope (Zeiss, Axio Imager M2).
- Fluorochromes were obtained from Life Technologies. Unless otherwise noted, all other chemicals were purchased from Sigma-Aldrich.

Transcriptomic database using 454 pyrosequencing

- ***Sample collection***

An additional condition (MotB; infection with the *motB* mutant of VF_0715 gene, coding for a flagellar motor) was obtained simultaneously at 3 h for an independent experiment, used for the construction of the transcriptome of reference, and included for completeness.

Animals were anesthetized and placed into RNAlater (Life Technologies) either immediately for the hatchling condition or at 3 h for the apo- and symbiotic conditions.

- ***Quality controls for collection***

To determine whether animals were associating with *V. fischeri* normally, a subset of the animals was sampled 3 h after hatching, and viewed for aggregate formation and hemocyte trafficking (Supplementary Figure S1A and S1B). To visualize bacterial aggregates, squid were incubated for 5 min in in HOSW containing 10 μg/mL Alexa633 wheat germ agglutinin (WGA), a fluorochrome that stains the mucus. The animals were then anesthetized and the tissues were dissected to reveal the light organ surface. The number of GFP-expressing bacteria present in the ciliated field of one appendage per squid was counted under a confocal microscope.

To visualize hemocyte trafficking, a hallmark of initial colonization (Koropatnick et al., 2007), squid were fixed overnight in 4 % paraformaldehyde at 4°C. Samples were washed four times with mPBS (50 mM sodium phosphate pH 7.4, 0.4 M NaCl) for 30 min each. Squid light organs were dissected out and permeabilized for 24 h at 4°C in mPBST (mPBS with 1 % Triton-X100). Staining for globular actin, a highly abundant protein in hemocytes, was used to localize hemocytes. Samples were thus incubated with 1 mg/mL FITC-DNAseI in mPBST for 24 h at 4°C. Rhodamine phalloidin (25 μg/mL) was added to counterstain the samples and the incubation was continued for 24 h at 4°C. Samples were washed four times with mPBS for 15 min each, and mounted on glass slides with Vectashield (Vector laboratories) before imaging on a confocal microscope.

- ***Sample preparation***

RNA was extracted using the 'RNeasy protocol for animal tissues' (Qiagen) with the following modifications: after thawing the light organs and removing the RNA later solution, samples were homogenized over liquid nitrogen in lysis buffer. The homogenate was applied to a Qiashredder column (Qiagen) to macerate the tissues more fully. The ink contained in the flow through was pelleted twice at 10,000 *g* for 3 min and the ink-free homogenate was used for the RNeasy procedure. To remove any DNA contamination, an on-

column DNAse treatment was performed. To ensure limited degradation and high purity of the extracted RNA, the samples were analyzed by gel electrophoresis, and their quantity estimated spectrophotometrically by a NanoDrop ND1000 or fluorometrically by a Qubit 2.0 (Life Technologies), respectively.

In addition to the libraries generated with the SMART cDNA synthesis kit, and to assemble longer transcripts for the transcriptomic map, an additional library was created using the cDNA Rapid Library Preparation method (RACE, Roche Life Sciences), according to manufacturer's instructions.

- ***454 sequencing, EST assembly, and statistics***

The sequences were first extracted from the 454 output file (sffinfo script, Roche), then processed through Newbler assembler (v2.6, -cDNA option, minimum overlap length = 40 bp, minimum overlap identity = 90 bp) for *de novo* reconstruction of the squid light-organ transcriptome. Annotation was performed by blast2go software (Conesa et al., 2005), using a blastx against the non-redundant database nr (minimum query coverage = 30 %). A high e-value cut-off was chosen (0.1), because of the high divergence between *E. scolopes* (cephalopod) and current genomes in public databases (Supplementary Figure S1). Loci quantification was performed using the RSEM package (RNA-Seq by Expectation Maximization), v1.1.18 with bowtie 0.12.7 (Li and Dewey, 2011). Differential expression of isotigs was estimated using the method developed by Stekel et al. (2000). This method is based on the entropy of a partitioning of isotigs, among the cDNA libraries that have been prepared by the SMART protocol only. For Rj > 3, the number of isotigs was much above the theoretical exponential curve generated from a random distribution, suggesting that isotigs with a Rj > 3 were differentially expressed. These isotigs were then analyzed for a functional enrichment analysis using FatiGO software (Al-Shahrour et al., 2004) on the babelomics 4.3 platform (http://babelomics.bioinfo.cipf.es). The analysis based on the Cluster of

Orthologous Groups (COG) was performed as follow: (1) isotigs were blasted against the eukaryotic COG database comprising Clusters of Orthologous Groups from 7 eukaryotic complete genomes (*Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens, Saccharomyces cerevisiae, Schizosaccharomyces pombe* and *Encephalitozoon cuniculi*)(e-value = 10) and only the best BLAST hit was selected, (2) a quality filter was applied (minimum identity = 40 %, high-scoring segment pairs (HSP) > 30, query coverage > 40 %), (3) individual COG annotations were extracted for the graph reconstruction. 39.3% of the isotigs with a BLAST hit (based on the Blast2GO analysis) also possess a COG annotation.

Expression of candidate genes

- ***Quantitative Reverse Transcription-PCR (qRT-PCR)***

Juveniles were separated according to their clutch of origin. Each clutch was laid by a different captive female, originally collected in the same location (Maunalua bay) in Hawaii. Animals were collected within 15 min of hatching and randomly segregated into three experimental conditions: 'sterile' (squid in FSIO), 'aposymbiotic' (squid in HOWS), and 'symbiotic' (squid in HOWS containing ~5000 cells/mL of WT ES114 strain). At 3 h after hatching, animals were anesthetized and placed into RNAlater. To test the robustness of the experiment, a subset of the animals was sampled and viewed for aggregate formation (n = 5 / condition / collecting day) and hemocyte trafficking (n = 10 / condition / collecting day), as described above.

Light organs were dissected in RNAlater (n = 20 / replicate / condition; 7 biological replicates / condition) and frozen at -80 °C until RNA extraction. RNA extraction was performed as for the 454 sequencing, except that (i) samples were homogenized using the tissueLyserLT (Qiagen) for 4 min at 30 Hz (5 mm stainless bead) before being applied to a

Qiashredder column (Qiagen), (ii) DNAse treatment was performed using TurboDNAse (Ambion) following manufacturer's intructions. Reverse Transcription (RT) reactions were performed starting from 500 ng of total RNA, using SMART MMLV Reverse Transcriptase (Clontech) and oligodT primers, according to the Manufacturer's instructions, and reactions were diluted 1/8 fold. Negative controls were performed in the same manner but without MMLV enzyme. Primers used for the PCR amplification were designed by Primer3 software (see table below). The reaction mixture for quantitative RT-PCR (qRT-PCR) consisted of 0.5 µL of each primer (10 mM), 5 µL of LC480 SYBR-Green master mix (Roche), and 4 µL of a 1/8 dilution of the cDNA reaction. qRT-PCR was performed on a Roche Lightcycler LC480 system (2 technical replicates / biological replicate) as follows: 5 min at 95 °C, 40 times [15 s at 95 °C, 10 s at 59 °C, 15 s at 72 °C], 20 s at 70 °C. A melting curve was recorded at the end of the PCR amplification (from 70 °C to 95 °C) to confirm that a unique transcript product had been amplified. In order to calculate PCR efficiencies, standard curves were plotted using seven dilutions (10-10^7 copies) of a previously amplified PCR product purified using QiaQuick kit (Qiagen). Primer sets exhibit PCR efficiencies between 1.83 and 2.06 (mean ± SE = 1.94 ±0.02). Expression values were calculated by E^{-Cp}, where E corresponds to the efficiency of the PCR reaction and Cp to the crossing point (Pfaffl, 2001). Candidate gene expression was first normalized by the geometric mean of the expression of three housekeeping genes (40S ribosomal protein S19, ß-tubulin and Serine hydroxymethyltransferase (serine HMT)), and then standardized within each clutch to 'Sterile' condition equal to 1. As normalized data followed a normal distribution (Shapiro's test), comparison between aposymbiotic and symbiotic expression were performed using a paired t-test with unequal variances (pairing by clutch). The p-values were then adjusted using false discovery rate's correction (FDR, R software, version 2.14.1).

				Contig information		Best blast hit					qRT-PCR		
Gene name	Database isotig	Database isogroup	Blast sequence description	Sequence length (bp)	# of reads in contig	Organism	Hit accession	E-value	Simi-larity (%)	Score	Forward primer	Reverse primer	Amplicon size (bp)
Lipase	09765	04931	pancreatic triacylglycerol lipase-like	1593	1871	*Branchiostoma floridae*	XP_002605989	9.6e-92	54	301	lipase51302-qF: CTTGAAACACGCCGACTACA	lipase51302-qR: TGGGCTCCGACATTACTACC	189
Ferritin	03687	00662	Ferritin	813	824	*Hyriopsis schlegelii*	AEK27025	6.4e-77	79	242	Ferrit-52730qF: TATCCTGAAACCCGTCGAAA	Ferrit-52730qR: GCTAGGAACGTGGTGCAAAT	117
Chymotrypsin protease	10467	08855	s1 type peptidase	926	7862	*Loligo bleekeri*	BAI66447	6.5E-144	83	418	ChyP50959-qF: TGGTCTCGCTAAAGTCAGCA	ChyP50959-qR: GGTGTGCAAAGGATTCTGGT	153
Cathepsin L	02080	00196	cathepsin L	1186	2190	*Pinctada fucata*	ADC52431	1.2e-112	66	344	CysPr61126-qF: GTCCATTCCCCATTGTTGTC	CysPr61126-qR: AACTGTTGGTCCCGTCTCTG	196
Chitotriosidase	09111	04277	acidic mammalian chitinase-like	1832	1108	*Branchiostoma floridae*	XP_002597591	5.7e-159	63	476	chitin-50669-qF: CATTTTGGCAGAGCACGTAA	chitin-50669-qR: AGATTCCGTCGTTCCAAGTG	152
Legumain	10092	05258	legumain precursor	1487	309	*Ixodes ricinus*	AAS94231	5.7e-145	69	1174	legum51953-qF: GACATCTGCCACGCTTATCA	legum51953-qR: CCCACTTTTTGTCCCTGAAG	227
Lysozyme	17002	12168	Lysozyme	692	50	*Crassostrea virginica*	BAE47520	1.9e-19	56	142	lysoz33499-qF: CCGCTCTTTTATTGGTCAGC	lysoz33499-qR: ATTTCTCCCAACTCCCCATC	201
Housekeeping genes													
Serine HMT			Serine Hydroxymethyl transferase								Serine HMT-qF: GTCCTGGTGACAAGAGTGCAA TGA	Serine HMT-qR: TTCCAGCAGAAAGGCACGATA GGT	148
40S			40S ribosomal protein S19								40S-F2: AATCTCGGCGTCCTTGAGAA	40S-R2: GCATCAATTGCACGACGAGT	103
β-tubulin			β-tubulin								Btub-qF: TGGGAACCCTGTTGATTAGC	Btub-qR: GACAATGTGGCGTTGTATGG	124

Table: Characteristics of candidate genes and primers used for qRT-PCR.

- ***In situ hybridization***

Rapid Amplification of cDNA Ends (RACE) was performed using Superscript III RT kit, Platinium Taq polymerase and TOPO TA cloning kit for sequencing, as recommended by the manufacturer (Life Technologies). Specific PCR products for ISH were amplified using the following primers: chitin-F = ACC TTT GGT GGT GTC TCC TG and chitin-T7R = TAA TAC GAC TCA CTA TAG GGG GCT TTG ATG GTT TGG ATG T (anti-sense chitotriosidase probe, 471 bp); chitin-T7F = TAA TAC GAC TCA CTA TAG GGA CCT TTG GTG GTG TCT CCT G and chitin–R = GGC TTT GAT GGT TTG GAT GT (sense chitotriosidase probe, 471 bp). These PCR products were amplified from 10 ng of cDNA using the Platinium Taq polymerase following manufacturer's instructions, and purified using the Qiaquick kit (Qiagen). *In situ* probes were *in vitro* synthesized by combining 10 ng of the purified PCR product, 1 µL of digoxigenin-UTP RNA labeling mix 10X (Roche), 2 µL of 5X transcription buffer, 0.5 µL of RNAsin (Promega), 1 µL of 100 mM dithiothreitol (DTT), 0.75 µL of T7 polymerase (Promega) for a final volume of 10 µL. Probes were diluted 5 fold in water, purified using Illustra ProbeQuant G-50 micro column (GE Healthcare) and adjusted to 50 % formamide before storage at -20 °C.

Determination of mucus pH

SNARF was covalently linked to WGA by a diimine bond between the amino groups on WGA and carboxyl groups on SNARF, catalyzed by addition of the crosslinker EDAC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide).

Protein detection and localization

- ***Antibodies***

Synthetic peptides were made to regions in the proteins of interest without significant match to any sequence in the squid databases or the non-redundant database of NCBI. Affinity-purified polyclonal antibodies against chitotriosidase (EsChit) and chitin synthase (EsCS) were produced in rabbit and chicken, respectively, against 15-amino acid peptides chosen for their predicted hydrophilicity and antigenicity (GenScript). The chitotriosidase peptide was part of the glycoside hydrolase domain (CSNKAKTQPGDTTKG) and not found in other known *E. scolopes* chitinases. The chitin synthase peptide was located at the C-terminal region of the protein (CVIHKTSSKNHNEQM), which blasted against chitin synthase 3 in *Crassostrea gigas*.

- ***Western blot***

20 µg of squid soluble proteins extracted in PBS containing protease inhibitors (Sigma) were loaded for SDS-polyacrylamide gel electrophoresis. One gel of 10 wells was run for the entire experiment. The gel was cut into three sections, each section containing a set of standards. One section was stained with Coomassie as a companion gel to indicate the range of protein to which the antibody was reacting, one section was used for the specific antibody (α-EsChit or α-EsCS), and the last section was used for the α-IgG/IgY control. Proteins were transferred to PVDF (for α-EsChit) or nitrocellulose (for α-EsCS) membranes

using a Trans-Blot Electrophoretic Transfer Cell system (Biorad), or stained with ProtoBlue Safe (National Diagnostics) for the companion gel. Membranes were blocked overnight in TBS-Tween (TBS-T) containing 4 % milk, incubated with the antibodies (1/3000 in TBS-T containing 1 % milk) for 1 h (for α-EsChit and IgG) or 3 h (for α-EsCS and IgY) at RT, washed 3 times in TBS-T, incubated with Goat-anti-rabbit (for α-EsChit) or Goat-anti-chicken (for α-EsCS) coupled to horseradish peroxidase (1/3000 in TBS-T containing 1 % milk) for 45 min at RT, and washed 3 times before visualization by chemiluminescence (Thermo Scientific).

- ***Immunocytochemistry***

Immunocytochemistry experiments were performed as previously described (Troll et al., 2010), with the following modifications. Squid were incubated for 7 days with the primary antibody 1:1000 in blocking solution, or IgG/IgY at the same concentration for negative controls. Samples were incubated overnight in secondary FITC-conjugated goat anti-rabbit/chicken antibody (Jackson ImmunoResearch Laboratories) at a 1:40 dilution in blocking solution. Samples were counterstained with 25 µg/mL rhodamine phalloidin in mPBST (50 mM sodium phosphate pH 7.4, 0.4 M NaCl, 1 % Triton-X100) overnight for staining of the actin cytoskeleton, and with TOTO-3 1:500 in mPBS for 20 min at room temperature for nuclear staining. The squid mantle cavity was opened to reveal the light organ and samples were mounted in Vectashield (Vector Laboratories) before observation by confocal microscopy.

Protein biochemistry of EsChitotriosidase

- ***Protein purification by affinity chromatography***

The soluble proteins of the squid were extracted in PBS containing protease inhibitors and incubated, with mixing, for 1 h at 4 °C with 50 µL of chitin bound to magnetic beads (New Englands Biolabs), which had been previously washed with binding buffer (500 mM NaCl,

20 mM Tris-HCl pH 8. 1 mM EDTA, 0.05 % Triton-X100). Unbound proteins were detached by 7 washes in binding buffer, followed by 2 washes in PBS. Protein-bead complexes were boiled for 10 min in 1X gel loading buffer containing 25 mM Tris(2-carboxyethyl)phosphine (TCEP, ThermoScientific), and the supernatant was loaded into a NuPAGE 4-12 % Bis-Tris gel (Life Technologies) for electrophoresis. Proteins were then stained with SYPRO Ruby (Life Technologies).

- ***Protein activity***

Because chitin binding domains from chitotriosidase bind chitin with such a high affinity, it was not possible to detach the protein from the beads without affecting its activity. Hence, protein-bead complexes were tested for their chitotriosidase activity using a chitinase assay kit (Sigma-Aldrich) against 4-methylumbelliferyl-ß-D-N,N',N"-triacetylchitotriose either in the assay buffer, or in various buffers (100mM) containing 2.5 % NaCl: sodium acetate for pHs 3, 4, 5, and 6; Tris for pHs 7 and 8; sodium carbonate for pHs 9, 10, and 11. Protein-bead complexes were washed 3 times in the appropriate buffer, split into triplicates, re-suspended in 100 µL of buffer, and incubated with 50 µg of the fluorescent substrate for 30 min at 37 °C. Reactions were stopped by the addition of 200 µL of sodium carbonate, and fluorescence was measured at 535 nm (Tecan GENios Pro, excitation: 340 nm). A standard curve was prepared using 4-methylumbelliferone at 5 dilutions of substrate: 0, 10, 100, 400, 700, 1000 ng.

Capillary assay for chemoattraction:

The capillary assay is based on Adler's methods (Adler and Epstein, 1974), and adapted from Brennan et al. (2013). Strains were grown in SWT liquid medium (per liter, 5 g Bacto-tryptone, 3 g yeast extract, 3 mL glycerol, 700 mL Instant Ocean at a salinity of 33-35 ppt, and 300 ml distilled water), supplemented with either 1 mM GlcNAc or $(GlcNAc)_2$ when

indicated, to an optical density of OD_{600} ~0.3. Cells were pelleted gently for 5 min at 800 *g*, washed to further remove remaining media components, and resuspended in buffered artificial seawater (HEPES-ASW: 100 mM $MgSO_4$, 20 mM $CaCl_2$, 20 mM KCl, 400 mM NaCl, and 50 mM HEPES, pH 7.5). One microliter capillary tubes (Drummond Scientific) were sealed at one end, filled with either HEPES-ASW alone or HEPES-ASW containing the indicated attractant at a concentration of 10 mM, and inserted into microcentrifuge tubes containing the cell suspension. The tubes were incubated on their side for 5 min at room temperature (23-24°C), after which the capillary tubes were removed from the cell suspension and washed. The contents were expelled into 150 µL buffer (either HEPES-ASW or 70 % Instant Ocean), and dilutions were plated for colony counts on LBS plates. Measurements were done in duplicate and the experiment was performed three times. Because data did not follow a normal distribution, they were analyzed by a Wilcoxon's test (R software, version 2.14.1).

SUPPLEMENTARY REFERENCES:

Adler, J. and Epstein, W. (1974) Phosphotransferase-system enzymes as chemoreceptors for certain sugars in *Escherichia coli* chemotaxis. Proc Natl Acad Sci U S A *71,* 2895-9.

Al-Shahrour, F., Díaz-Uriarte, R., and Dopazo, J. (2004) FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics *20,* 578-80.

Boettcher, K. J. and Ruby, E. G. (1990) Depressed light emission by symbiotic *Vibrio fischeri* of the sepiolid squid *Euprymna scolopes*. J Bacteriol *172,* 3701-6.

Brennan, C.A., Deloney-Marino, C.R., and Mandel, M.J. (2013) The chemoreceptor VfcA mediates amino acid chemotaxis in *Vibrio fischeri*. Appl Environ Microbiol *79,* 1889-96.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21,* 3674-6.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5,* 113.

Koropatnick, T.A., Kimbell, and McFall-Ngai, M.J. (2007) Responses of host hemocytes during the initiation of the squid-vibrio symbiosis. Bio Bull *212,* 29-39.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res *29,* e45.

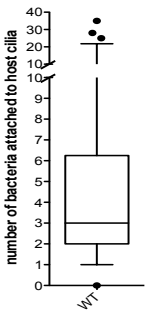
Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12,* 323.

Stekel, D.J., Git, Y., and Falciani, F. (2000) The comparison of gene expression from multiple cDNA libraries. Genome Res *10,* 2055-61.

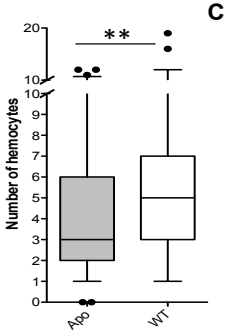
Troll, J.V., Bent, E.H., Pacquette, N., Wier, A.M., Goldman, W.E., Silverman, N., and McFall-Ngai, M.J. (2010) Taming the symbiont for coexistence: a host PGRP neutralizes a bacterial symbiont toxin. Environ Microbiol *12,* 2190-2203.

Supplementary Figure S1

A



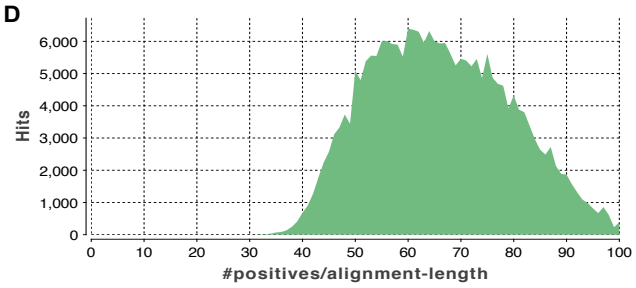
B



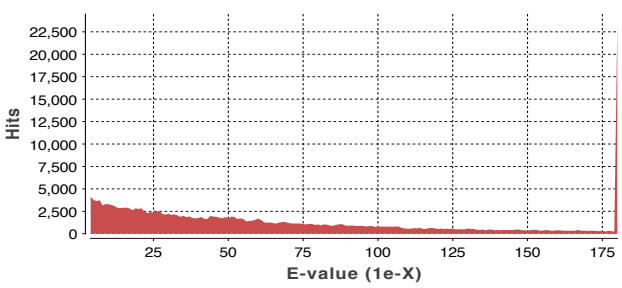
C

Sequencing lane 454	Biological condition	Protocol for cDNA preparation	Number of assembled bases	% of bases mapped	Number of assembled reads	% of reads mapped	Inferred read error
334Squid20100903.sff	Symbiotic WT	SMART	34664055	80.82	116306	72.75	1.92
334Squid20100916.sff	Symbiotic WT	SMART	52053382	78.98	160025	74.83	1.91
336Squid20100916.sff	Hatchling	SMART	45670541	77.43	143280	73.47	1.98
336Squid20101006.sff	Hatchling	SMART	31279560	75.20	140049	71	2.39
338Squid20100916.sff	Aposymbiotic	SMART	81000484	77.70	243419	73.43	1.81
339Squid20101006.sff	Symbiotic WT	Rapid	177643629	91.07	588403	90.05	1.53
337Squid20100916.sff	*MotB*	*SMART*	*42369988*	*79.35*	*121496*	*76.07*	*1.79*
337Squid20101006.sff	*MotB*	*SMART*	*34515689*	*78.05*	*132976*	*75.4*	*2.23*

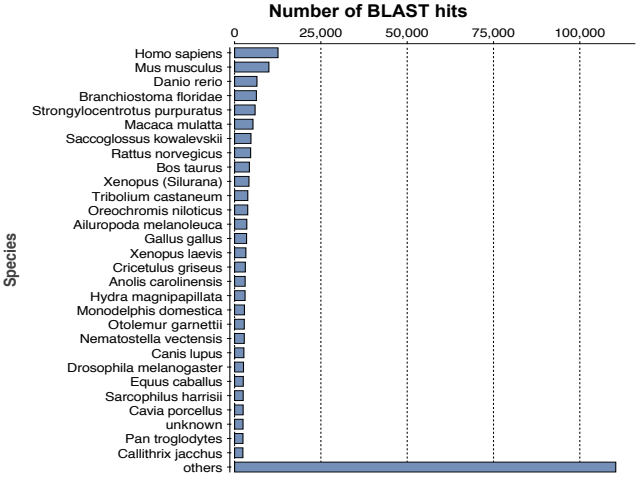
D



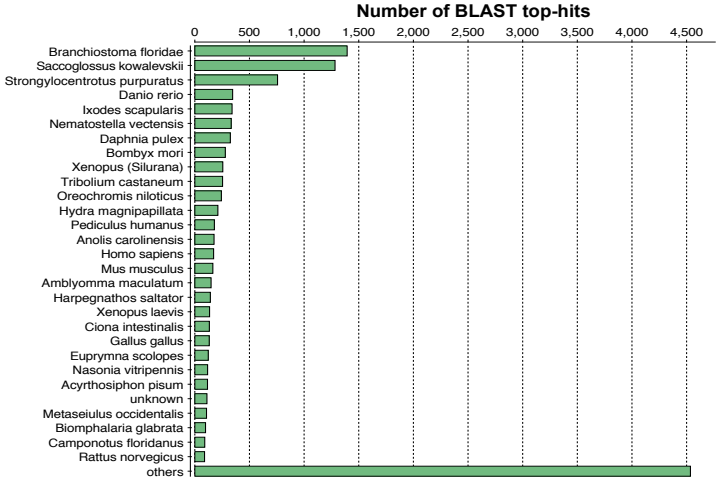
E



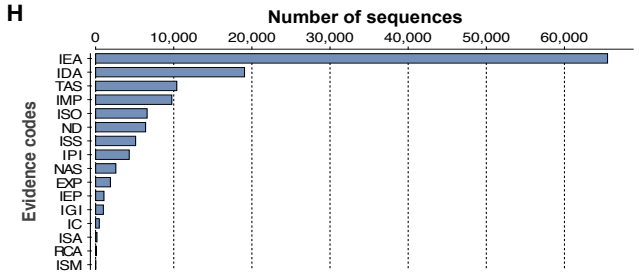
F



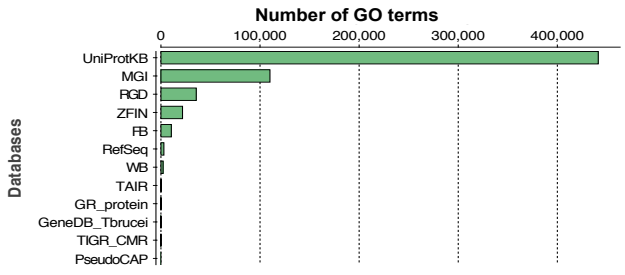
G



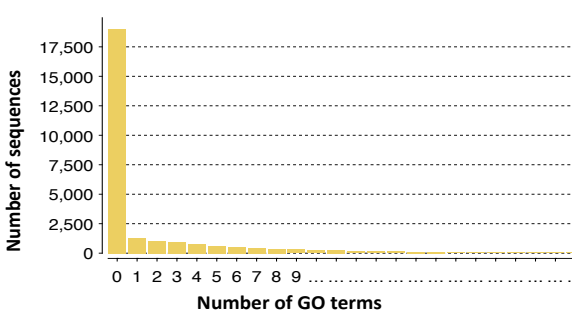
H



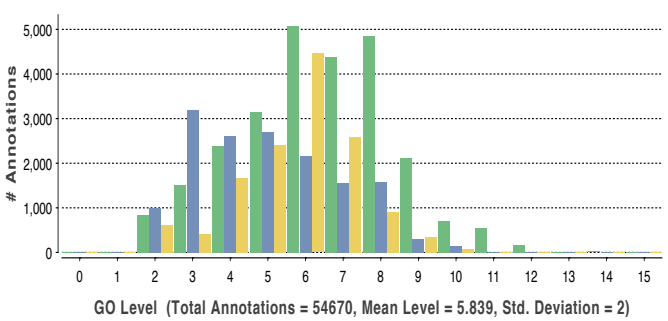
I



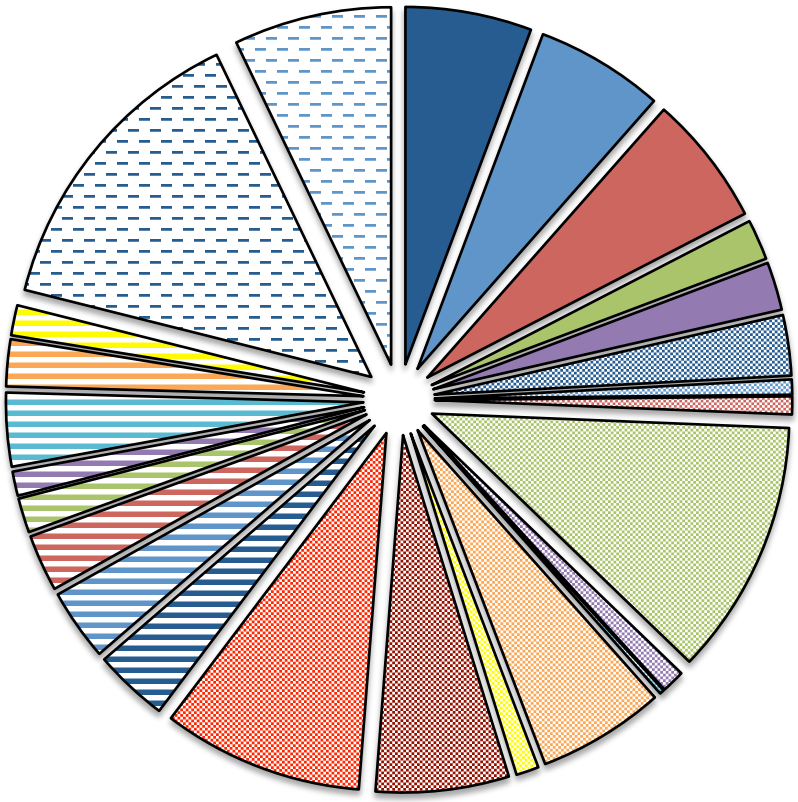
J



K



L



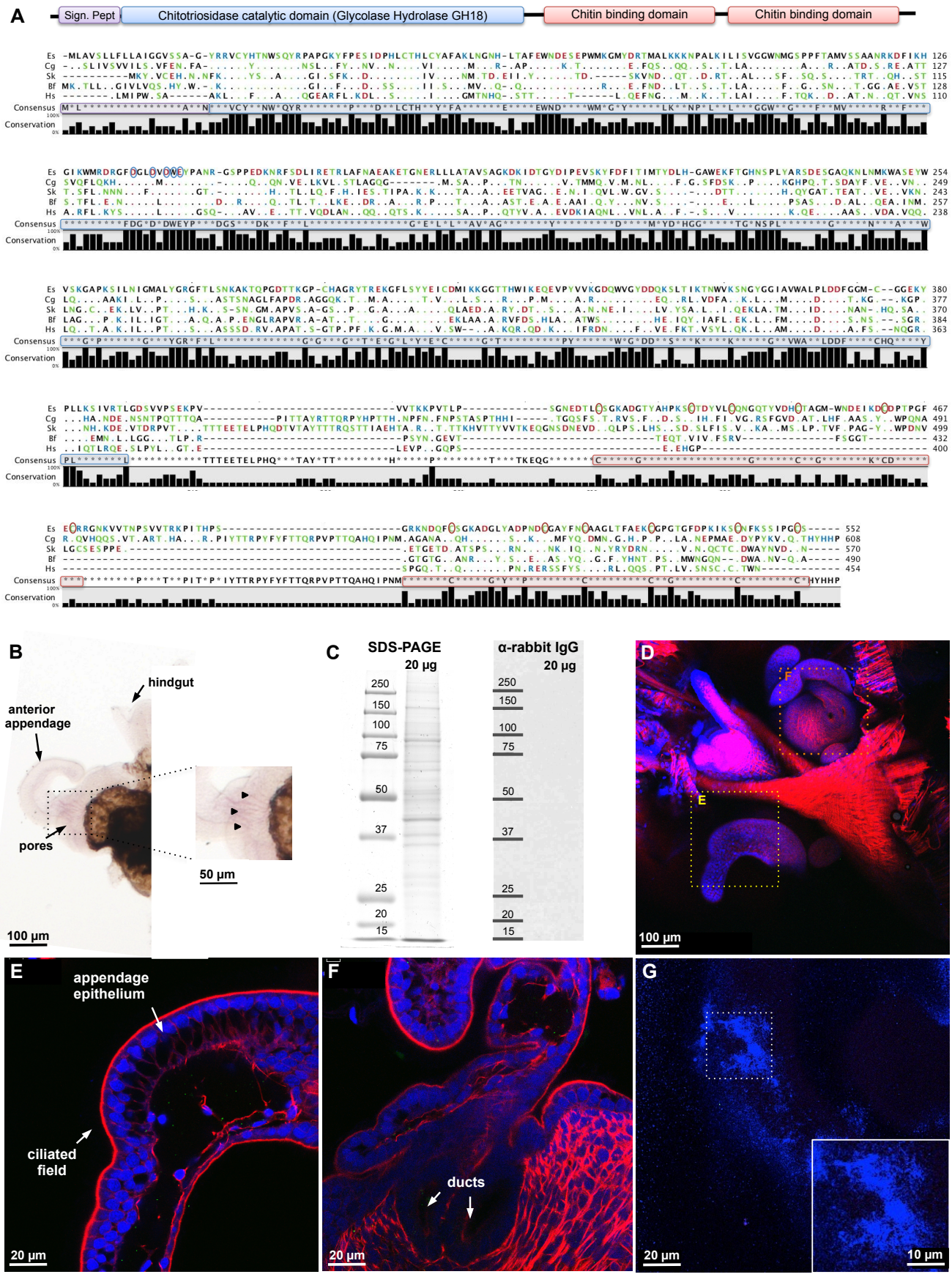
Information storage & processing
- Translation, ribosomal structure and biogenesis
- RNA processing and modification
- Transcription
- Replication, recombination and repair
- Chromatin structure and dynamics

Cellular processes & signaling
- Cell cycle control, cell division, chromosome partitioning
- Nuclear structure
- Defense mechanisms
- Signal transduction mechanisms
- Cell wall/membrane/envelope biogenesis
- Cell motility
- Cytoskeleton
- Extracellular structures
- Intracellular trafficking, secretion, and vesicular transport
- Posttranslational modification, protein turnover, chaperones

Metabolism
- Energy production and conversion
- Carbohydrate transport and metabolism
- Amino acid transport and metabolism
- Nucleotide transport and metabolism
- Coenzyme transport and metabolism
- Lipid transport and metabolism
- Inorganic ion transport and metabolism
- Secondary metabolites biosynthesis, transport and catabolism

Poorly characterized
- General function prediction only
- Function unknown

Supplementary Figure S2



Supplementary Figure S3