Supplementary Information Appendix for

Symbiotic Organs Shaped by Distinct Modes of Genome Evolution in Cephalopods

Mahdi Belcaid[1], Giorgio Casaburi[2], Sarah J. McAnulty[3], Hannah Schmidbaur[4], Andrea M. Suria[3], Silvia Moriano-Gutierrez[1], M. Sabrina Pankey[5], Todd H. Oakley[5], Natacha Kremer [6], Eric J. Koch[1], Andrew J. Collins[3], Hoan Nguyen[7], Sai Lek[7], Irina Goncharenko-Foster[2], Patrick Minx[8], Erica Sodergren[7], George Weinstock[7], Daniel Rokhsar[9,10,11], Margaret McFall-Ngai[1], Oleg Simakov[2,9*], Jamie S. Foster[2*], Spencer V. Nyholm[3*]

[1]Pacific Biosciences Research Center, University of Hawaii, Honolulu, Hawaii, 96822, USA. [2]Department of Microbiology and Cell Science, University of Florida, Space Life Science Lab, Merritt Island, Florida, 32953, USA. [3]Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA 06269. [4]Department of Molecular Evolution and Development, University of Vienna, Austria. [5]Ecology, Evolution, and Marine Biology Department, University of California, Santa Barbara, CA 93106. [6] Laboratory of Biometry and Evolutionary Biology, University of Lyon, Villeurbanne, France. [7]Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06032, USA. [8]McDonnell Genome Institute, Washington University, 63108, [9]Molecular Genetics Unit, Okinawa Institute of Science and Technology, Japan. [10]Department of Molecular and Cell Biology, University of California Berkeley, California 94720, USA. [11]Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA.

Correspondence to: oleg.simakov@univie.ac.at, jfoster@ufl.edu, spencer.nyholm@uconn.edu

This PDF file includes:

Supplementary methods
Figs. S1 to S9
Tables S1 to S2
Dataset S1

www.pnas.org/cgi/doi/10.1073/pnas.1817322116

Materials and Methods

Genome sequencing and assembly

Genomic DNA generated in this study was derived from a single adult male *E. scolopes*. The optic lobe and gill tissues were dissected, frozen in liquid nitrogen, and then ground with a mortar and pestle. Genomic DNA was isolated from a total of 20 replicate DNA extractions using a MasterPure™ DNA Purification Kit according to the manufacturer's instructions (Epicentre® Madison, Wisconsin). The DNA was pooled and treated with RNAse A at 37°C for and then concentrated according to the manufacturer's instructions, with a final resuspension in 930 µl of TE buffer. The DNA was quantified using a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA) and quality was assessed using agarose gel. Genomic DNA (21.3 µg) was shipped to the Genome Institute of Washington University for Illumina library construction. Illumina libraries were prepared using Illumina V3 chemistry generating three fragment libraries (180, 500 and 700 bp) as well as two mate pair libraries (3 and 8 kb inserts). The libraries were sequenced on the HiSeq2000 platform at the Genome Institute of Washington University. A 1x PacBio library was also generated using DNA from the same samples using P6C4 chemistry and run on the PacBio RS II Single Molecule, Real-Time (SMRT) DNA Sequencing System at the Jackson Laboratory for Genomic Medicine (Farmington, CT) on four SMRT Cells, generating an estimated 606,065 reads. Genomic DNA from the same animal used for Illumina and PacBio sequencing was used to prepare a Chicago library (Dovetail Genomics, Santa Cruz, CA) as described previously (1). Briefly, ultra-high molecular weight (500 kb) was generated by shearing to an average 300 – 500 bp in size before ligation to adapters for sequencing on the Illumina HiSeq 2500 as paired 150-bp reads. The final assembly was generated using sequences from the Illumina and Chicago libraries.

For the Illumina HiSeq2000 and HiSeq2500 reads initial FASTQ quality assessment, demultiplexing, and adapter trimming was performed using BaseSpace software (Illumina, San Diego, CA) and TrimGalore (2) . The resulting filtered reads were pooled and assembled using Meraculous in diploid_mode 1 (3) with a k-mer size of 41. Additionally, the haplotype split peak was defined at 30X, and any k-mers below 7x coverage were discarded from the assembly. Gaps were then filled in with the fragment libraries for the final assembly. The resulting assembly had a N50 of 98 kb and was furthermore complemented with Dovetail Chicago library and assembly (Table S2). The initial genome size was estimated from the 41-mers to be 5.6 Gb based on the equation described in Albertin et al., (4).

RepeatModeller 1.0.10 (http://www.repeatmasker.org/RepeatModeler/) was used to construct a repeat library for *E. scolopes*, using the final genome assembly. This resulted in the reconstruction of 1,603 elements that were then annotated with previously published methods (5). The results of using this library to mask the genome are presented in *SI Appendix*, Fig. S1b.

We assessed the age distribution of repeat copies with a new consensus-free method of dating. Briefly, we searched the genome with BLASTN to find the matching loci of all families in the RepeatModeller library. For each family, we then extracted the genomic sequence of the match and ran the BLASTN alignments to estimate the pairwise distance as measured by the number of substitutions (excluding gaps) in the non-CpG regions. The distance was then corrected with the Jukes-Cantor (JC) formula. All pairwise distances were then screened to find the most recently diverging pair of repeat loci. Proceeding iteratively, we were able to construct a neighbor-joining tree for all repeat loci. The distances between individual coalescences in the tree reflect the insertion history. The dynamics for each family from DNA, LINE, LTR, and

SINE classes are plotted in *SI Appendix*, Fig. S1c identifying, unlike in octopus, a single large expansion peak at JC distance 0.1.

Transcriptome sequencing

Transcriptomes derived from 31 different tissues and developmental stages of *E. scolopes* were incorporated into a reference transcriptome to facilitate genome assembly. The metadata associated with tissue type, developmental stage, RNA extraction protocol and sequencing platform are listed in Supplementary Table S1. For the tissue-specific transcriptomes, RNA was extracted from adult ANG, brain, eyes, gills, hemocytes, light organ, and skin tissues, as well as juvenile head (white body, optic nerve, brain), eyes, gills, and light organ separately using the RNA extraction kits listed in *SI Appendix*, Table S2. For the PacBio IsoSeq library, RNA was extracted, normalized and pooled from adult ANG, brain, eyes, white body, optical lobe, gills, light organ, and skin as well as from whole juvenile hatchling, 24 h aposymbiotic and 24 h symbiotic animals (*SI Appendix*, Table S1) (6-9).

The RNA was quantified using a Qubit RNA High Sensitivity kit with a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA) and quality was assessed using either a BioAnalyzer 2100 or 2200 TapeStation (Agilent, Santa Clara, CA). RNA was treated with Turbo DNase (Ambion, Calsbad, CA) using the rigorous protocol and then the rRNA was removed using the Ribo-Zero Gold rRNA removal kit (Epidemiology, Illumina, San Diego, CA) according to the manufacturer's instructions. RNA underwent polyA selection (with the exception of the ANG-specific tissue) and cDNA library synthesis using either the TruSeq stranded mRNA sample prep kit (Illumina, San Diego, CA) or the NEBNext® Ultra™ RNA Library Prep Kit for Illumina (New England Biolab, Ipswich, MA). The PacBio IsoSeq libraries were generated by the National Center for Genome Resources (Santa Fe, NM). A table of recovered reads from each of the libraries is provided in *SI Appendix*, Table S1. All reads generated as part of this study have been deposited at the NCBI under Bioproject PRJNA470951 (genomic data), PRJNA320238 (RNA-seq paired-end adult light organs), PRJNA205147 (RNA-seq paired-end 4-week light organ), and PRJNA473394 (RNA-seq paired-end 24 hours LO, eyes, gills, and head). Some of the transcriptomes had been previously published under the accession numbers SAMN06159576, PRJNA257113, SRR329677.8, SRR329678.5.

Illumina, PacBio and 454 libraries were pooled and reads were quality filtered and digitally normalized before assembling using the Trinity de novo assembly package v2.4.0 (10, 11). The resulting assembly was merged with PacBio high-quality polished assembled isoforms and redundancy was removed using CD-HIT (12) at 100% identity. Assembly integrity and validation was estimated using BUSCO (13) against the metazoan core protein dataset. Hybrid assembly was screened for peptide-coding regions with TransDecoder in the Trinity package, retaining ORFs that were at least 100 amino acids long. Coding transcripts were annotated using BLASTX against the SwissProt (v.2016) database following the Trinotate v. 2.0.6 package pipeline (http://trinotate.github.io/). Our final combination of assemblies has passed through CD-HIT and contained 134,352 transcripts.

Gene model predictions

Available transcriptomes were mapped onto the genome to produce a training set for AUGUSTUS using PASA. Briefly, the training set given by PASA was filtered for completeness (complete tag by PASA), at least three exons, and no detectable self-redundancy as assessed by BLASTP of all peptides against each other. This information was used to train AUGUSTUS and

resulted in 45,359 models (after filtering out models that had more than 50% overlap with repetitive elements).

A total of 94% of the transcripts could be aligned against the reference genome using GMAP (14) and Splign (15). Only alignments with high quality (85%-similarity over 0.75-length) were subsequently retained for downstream analysis. The gene models inferred from the GMAP and Splign alignment were combined with the de novo-predicted and dereplicated such that short alignments where exons are completely included in longer alignments were discarded. This resulted in a total of 43,025 unique transcripts. Pairs of transcripts that shared at least one exon were subsequently grouped into clusters. This resulted in the final set of total 29,259 high-quality genes. Overall, the set of de-novo predicted and assembled transcripts identified 99% of the BUSCO (13) metazoan dataset protein, with 96.9% being complete and 2.1 identified as fragmented.

Synteny analysis

Using previously published methods of phylogeny-informed clustering (16) we constructed the sets of orthologous gene families between the following species: *Capitella teleta*, *Helobdella robusta*, *Lottia gigantea*, *Octopus bimaculoides*, *Euprymna scolopes*, *Crassostrea gigas*, *Nematostella vectensis*, and *Branchiostoma floridae*. To account for differential gene loss that may have impeded quantification of synteny loss/gain, we focused only on 3,547 clusters. We implemented a micro-synteny detection algorithm as described in Simakov et al. (16) and found, in accordance to previous results (4, 5) approximately 600 microsyntenic blocks that could be traced back to the bilaterian ancestor because they were shared between either or both ingroups (protostome and deuterostome) or an ingroup and an outgroup species (*Nematostella*). We encoded presence or absence of micro-synteny block as 1 or 0, respectively and ran a MrBayes analysis on the constrained species topology. Our results were consistent with the previous finding that cephalopod genomes show an accelerated pace of micro-synteny turnover (long branch), yet also revealed that the vast majority of those rearrangements happened before the split of *O. bimaculoides* and *E. scolopes*. This result suggests that many blocks should be lost (and gained) in the ancestor of both species (coleoid ancestor).

Data Access

Genome and transcriptome sequencing reads have been deposited in the SRA as Bioproject PRJNA 470951.

General trends in *E. scolopes* gene family evolution

Annotation methods

Each RNA-seq sample (transcript-per-million, TPM) was compared against all other samples and genes present in one sample and absent from the others were flagged as tissue-specific. We initially used normalized expression cutoffs of 20 and 2 to flag expressed and repressed transcripts respectively. These values, while stringent, minimized any false positives (i.e., identifying lowly expressed genes as tissue-specific) at the cost of potentially a higher false negatives rate. The transcripts identified were initially grouped using their Gene Ontology-Slim (generic GO-Slim) (17) functional classes. The resulting categories were then manually refined to correct for duplicated or closely related functions. The resulting 19 functional classes are presented in Fig. 3a.

PFAM enrichment analysis
PFAM annotations for the species shown in *SI Appendix*, Fig. S2 were obtained using Interpro-Scan (14). Counts for each PFAM category were obtained by counting genes with a given PFAM domain. The same PFAM domain present in more than one copy per gene was counted once. As described in Albertin et al., (4), the Fisher's exact test (and Bonferroni correction) was conducted to test for the overrepresentation of a given PFAM domain in a set of species compared to the background average. For a PFAM to be listed in the heatmap (*SI Appendix*, Fig. S2) we required that all species in a given group have significant (corrected p-value < 1E-5) enrichment against the background average. To improve the readability of the plot, we excluded transposon-related and uncharacterized PFAMs, such as "unknown", "reverse transcriptase","transposase", "endonuclease", "gag-poly", "retrotransposon", "parvovirus", "integrase core", "helicase", "GAG-pre-integrase". We also merged counts from related PFAMs, such as Cadherins, Zinc-fingers, GSTs, and Ankyrins.

Analysis of Symbiosis-specific Tissues

Reflectins, peroxidases, and crystallin identification and phylogeny
Reflectins were identified by BLASTP similarity searches across the available deposited sequences at the NCBI. Species included in the analyses were *Doryteuthis pealeii*, *Octopus bimaculoides* and *Sepia officinalis*. Publicly available sequences of *Euprymna scolopes* were also included in addition to sequences from our genome (Fig. 4, *SI Appendix*, Fig. S4-S5). Heme peroxidases were identified by the PFAM domain (PF03098) (18) and extracted from annotated genomes from NCBI. Species included in the analyses were *Aplysia californica, Branchiostoma floridae*, *Caenorhabditis elegans*, *Capitella teleta*, *Crassostrea gigas*, *Drosophila melanogaster*, *Homo sapiens*, *Hydra vulgaris*, *Lottia gigantea*, *Nematostella vectensis* and *Octopus bimaculoides*. A subset of the tree, only including *Octopus bimaculoides*, *Crassostrea gigas* and sequences of *Euprymna scolopes* from our genome, was used in Fig. 4 and *SI Appendix*, Fig. S6. Crystallins were identified by three PFAM domains (PF00171, PF02798, PF14497) and extracted from annotated genomes from NCBI. PF02798 and PF14497 are PFAM entries for glutathione S-transferases (GST) and lens crystallins in cephalopods are a subfamily of glutathione S-transferases (19). Species included in the analyses were *Branchiostoma floridae*, *Capitella teleta, Crassostrea gigas*, *Drosophila melanogaster*, *Homo sapiens, Hydra vulgaris*, *Lottia gigantea*, *Nematostella vectensis, Octopus bimaculoides.* Alignments were conducted using MAFFT (20) followed by phylogenetic reconstruction using FastTree (21), also using the default options. IQTREE was additionally run for reflectins to confirm tree topology (*SI Appendix*, Fig. S5). Sequences clustered on the tree were identified as S-crystallins or glutathione S-transferases using TBLASTN similarity searches against the entire NCBI database. Gene expression heatmaps were done in R with the heatmap.2 function of the gplots package (22).

Immunity related genes
The sampled tissues were analyzed to identify genes with immune-related PFAM categories that are highly expressed in a single tissue and lowly or not expressed in the remaining tissues. A total of 233 genes assigned to one or more of the protein families (PFAM) described below were identified. These genes were subsequently manually grouped based on their annotations (PFAM domain, KEGG pathways, eggnog orthologous groups and functional annotations) into 21 high-

level classes, as described in *SI Appendix*, Fig. S9. The set of immune-related PFAM ids identified in genes with tissue-specific expression were:

PF00047	Immunoglobulin domain
PF00058	Low-density lipoprotein receptor repeat class B
PF00084	Sushi repeat (SCR repeat)
PF00092	von Willebrand factor type A domain
PF00093	von Willebrand factor type C domain
PF00400	WD domain, G-beta repeat
PF00435	Spectrin repeat
PF00530	Scavenger receptor cysteine-rich domain
PF00560	Leucine Rich Repeat
PF05790	Immunoglobulin C2-set domain
PF07654	Immunoglobulin C1-set domain
PF07679	Immunoglobulin I-set domain
PF07686	Immunoglobulin V-set domain
PF08205	CD80-like C2-set immunoglobulin domain
PF11465	Natural killer cell receptor 2B4
PF12662	Complement Clr-like EGF-like
PF12799	Leucine Rich repeats (2 copies)
PF13516	Leucine Rich repeat
PF13519	von Willebrand factor type A domain
PF13768	von Willebrand factor type A domain
PF13855	Leucine rich repeat
PF13895	Immunoglobulin domain
PF13927	Immunoglobulin domain
PF14580	Leucine-rich repeat
PF15494	Scavenger receptor cysteine-rich domain
PF16680	T-cell surface glycoprotein CD3 delta chain
PF16681	Ig-like domain on T-cell surface glycoprotein CD3 epsilon chain
PF16706	Izumo-like Immunoglobulin domain

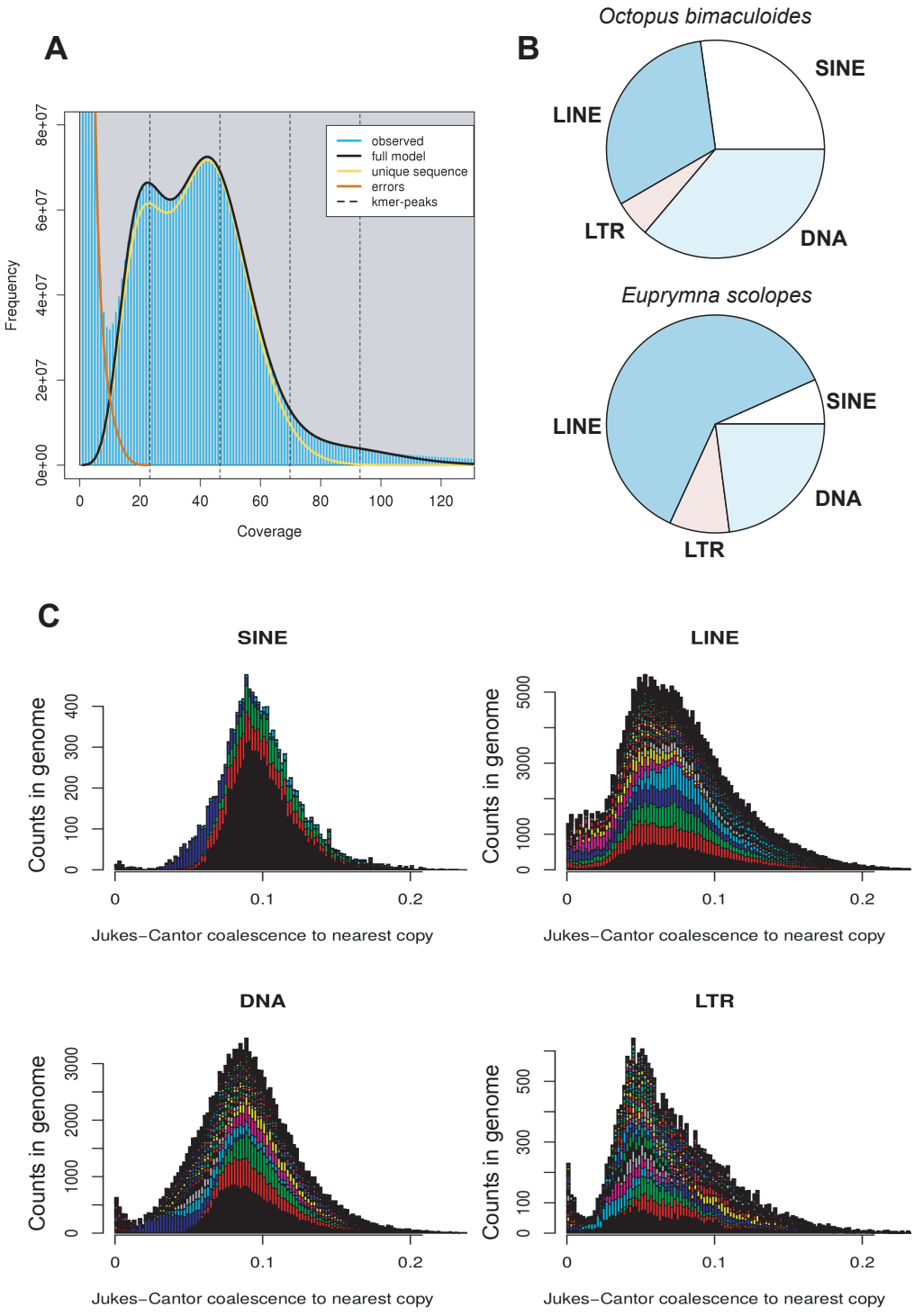

Fig. S1. Genome assembly statistics. (*A*) Kmer profile using GenomeScope. (*B*) Total repeat content in the genome of *O. biomaculoides* and *E. scolopes*. (*C*) Repeat age distribution using corrected (Jukes-Cantor) distances. SINE, LINE – short and long interspersed nuclear elements, respectively. LTR – long-terminal repeat retrotransposons; DNA – DNA transposons.

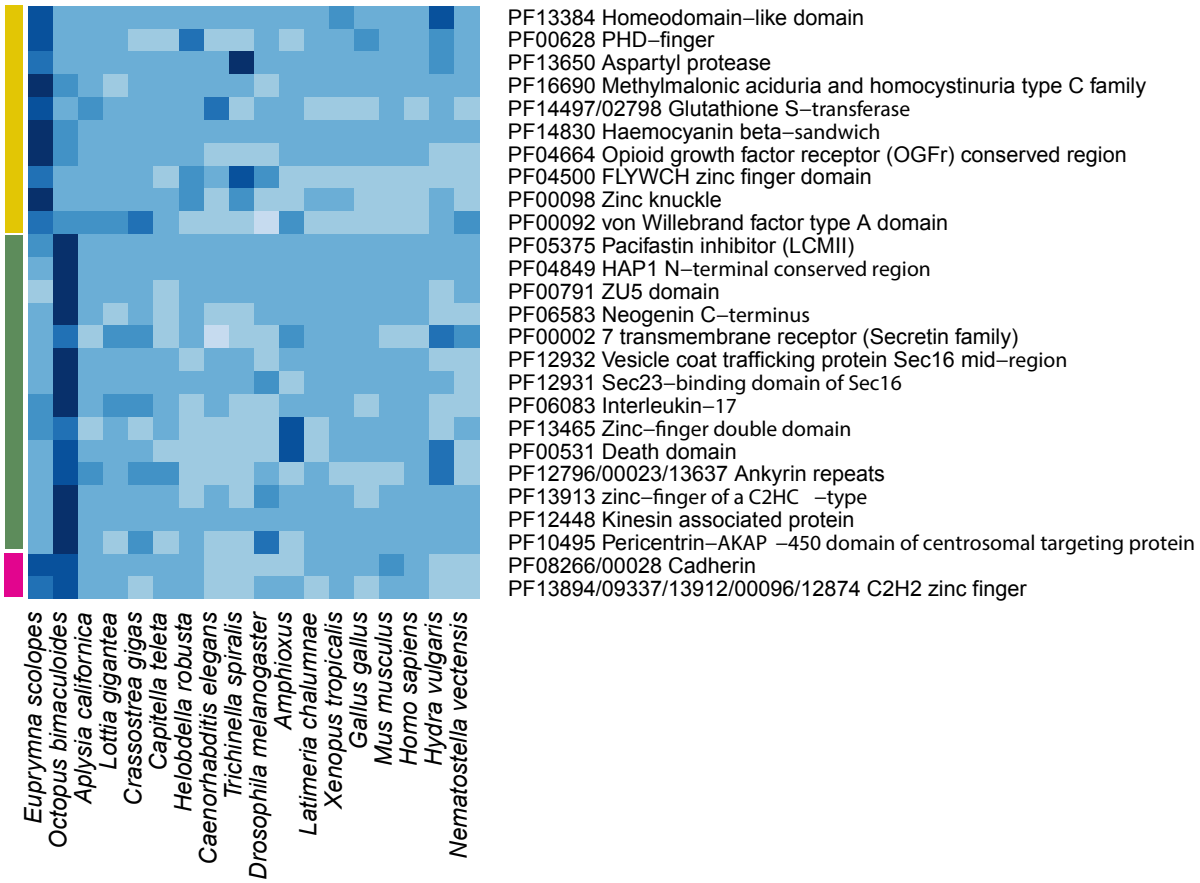

Fig. S2. PFAM domain enrichments. PFAM domain enrichment in *E. scolopes* (upper), *O. bimaculoides* (mid), and shared octopus-squid enrichment (lower) highlighting common zinc finger and protocadherin expansions. PFAM enrichment analysis as described in Albertin et al. (4), using Fischer's exact test and Bonferroni multiple test correction. Darker blue – higher enrichment, pale blue – no enrichment.

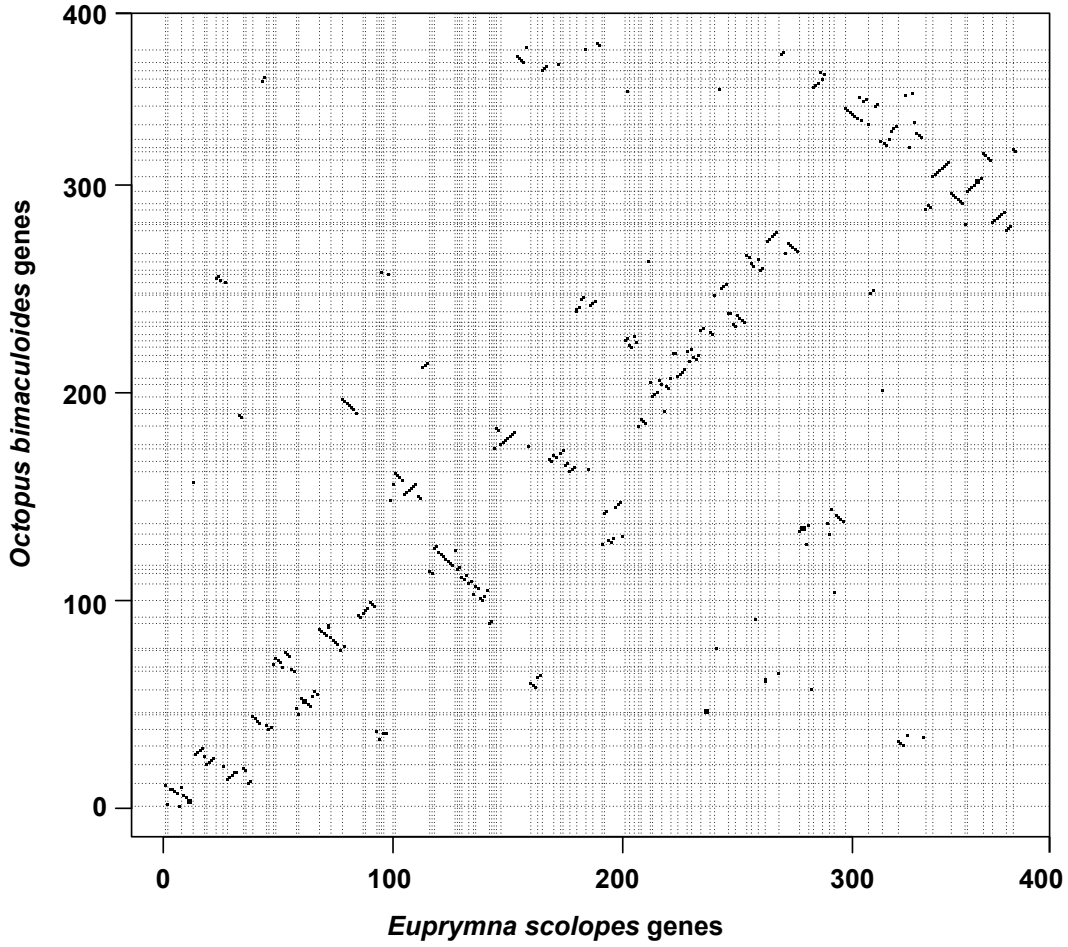

Fig. S3. Overlay of co-linearity of genes on octopus and bobtail squid scaffolds in a single region. Each dot indicates an orthologous gene family. Vertical and horizontal lines demarcate scaffold boundaries.

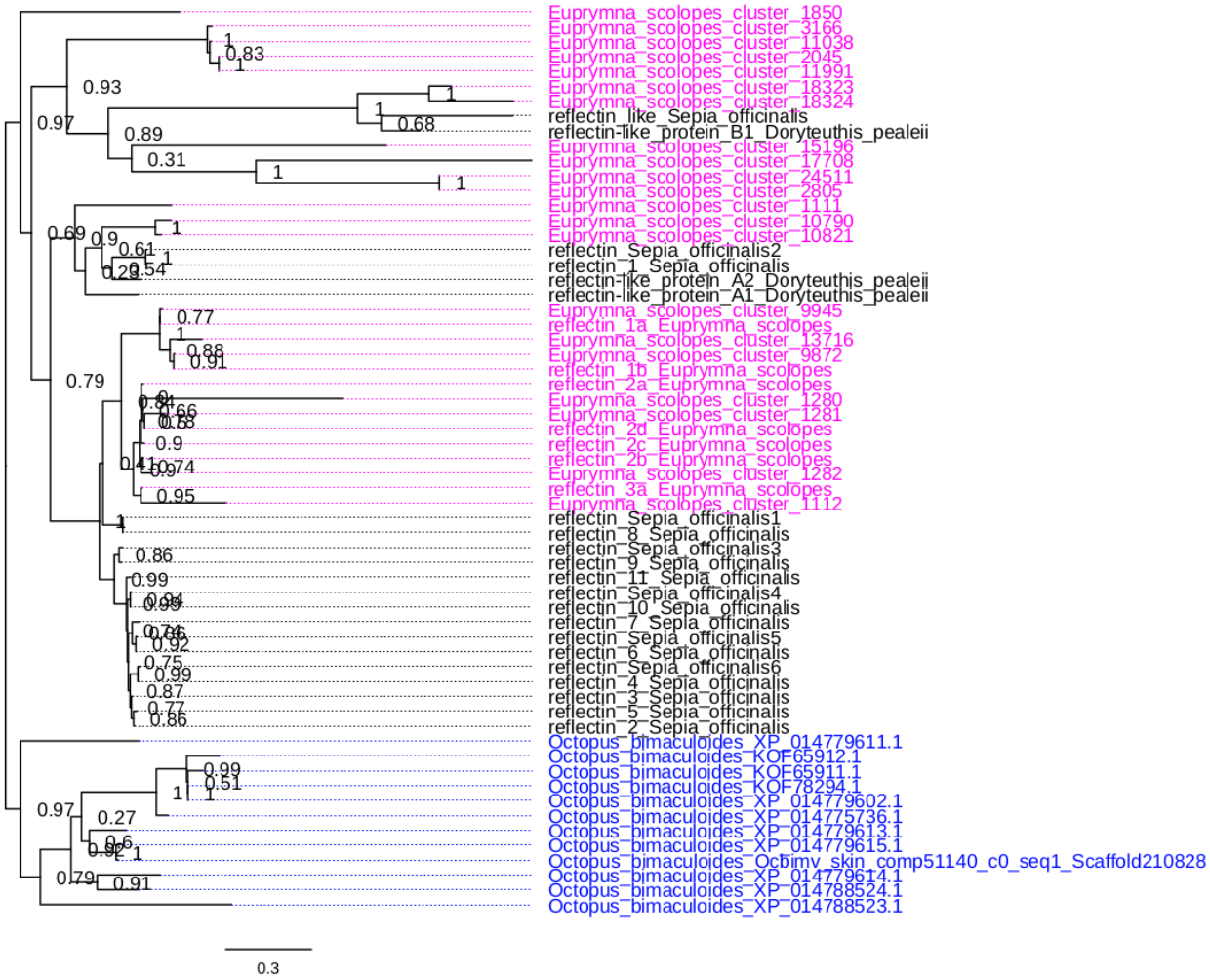

Fig. S4. FastTree analysis for reflectins. Branches in the main figure are labeled. Bootstrap support is shown for each node. Gene identifiers are at the end of each branch label, from the respective databases (*i.e.*, NCBI, except for *E. scolopes*).

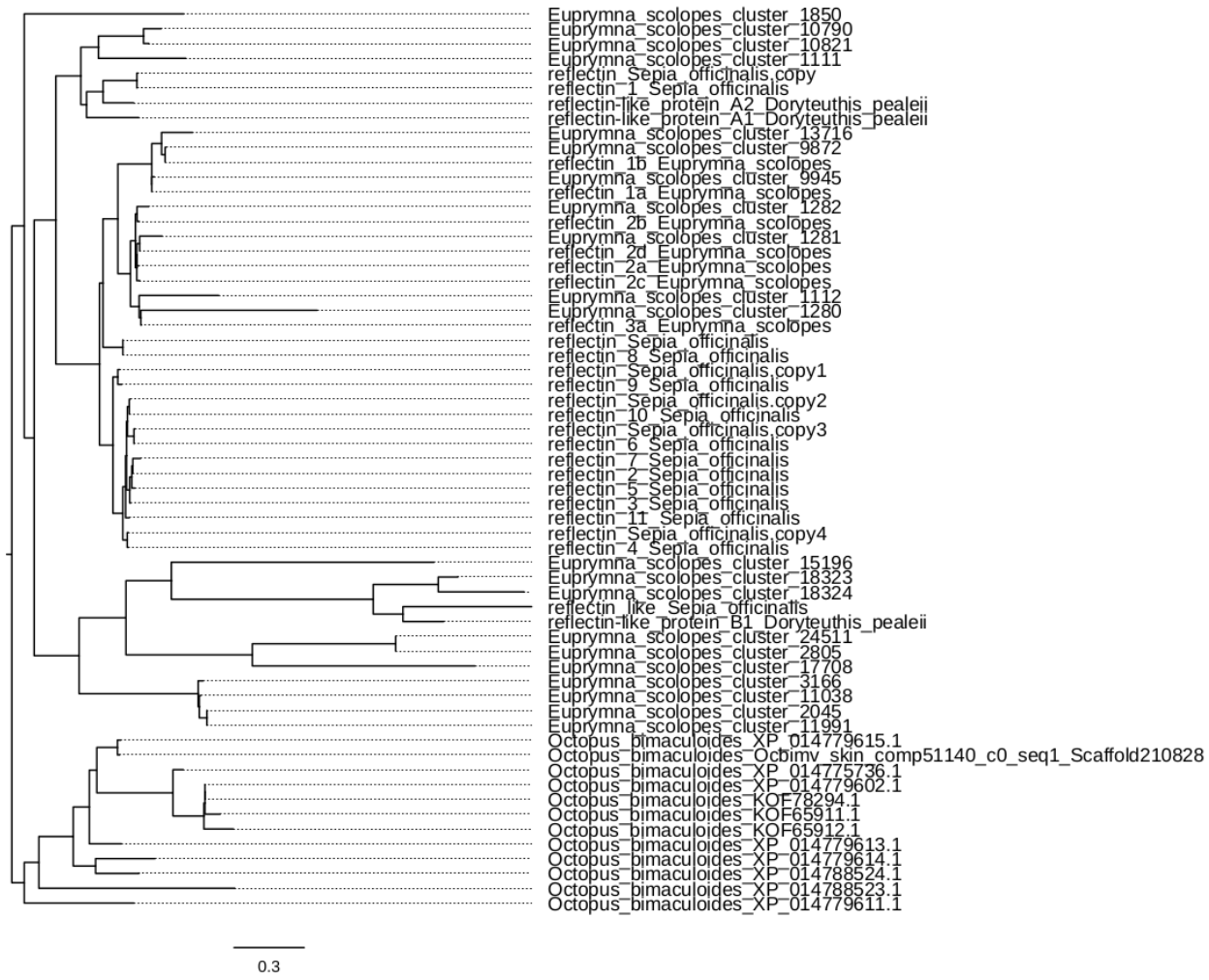

Fig. S5. IQTREE analysis for reflectins. *Euprymna*-specific expansions were detected within the assembled genome.

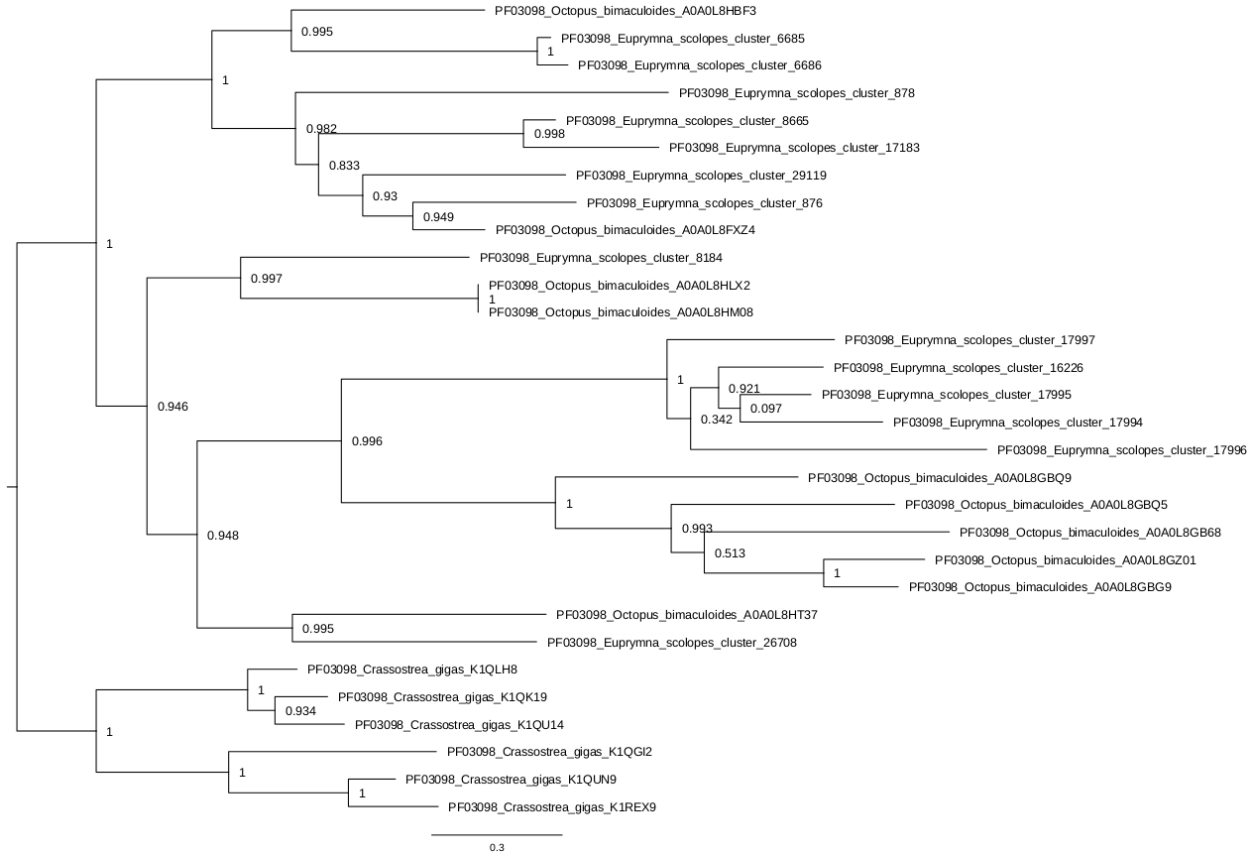

Fig. S6. FastTree analysis for peroxidases. Branches in the main figure are labeled. Bootstrap support is shown for each node. PF indicates the peroxidase PFAM domain. Gene identifiers are listed at the end of each branch label, from the respective databases (*i.e.*, NCBI, except for *E. scolopes*).

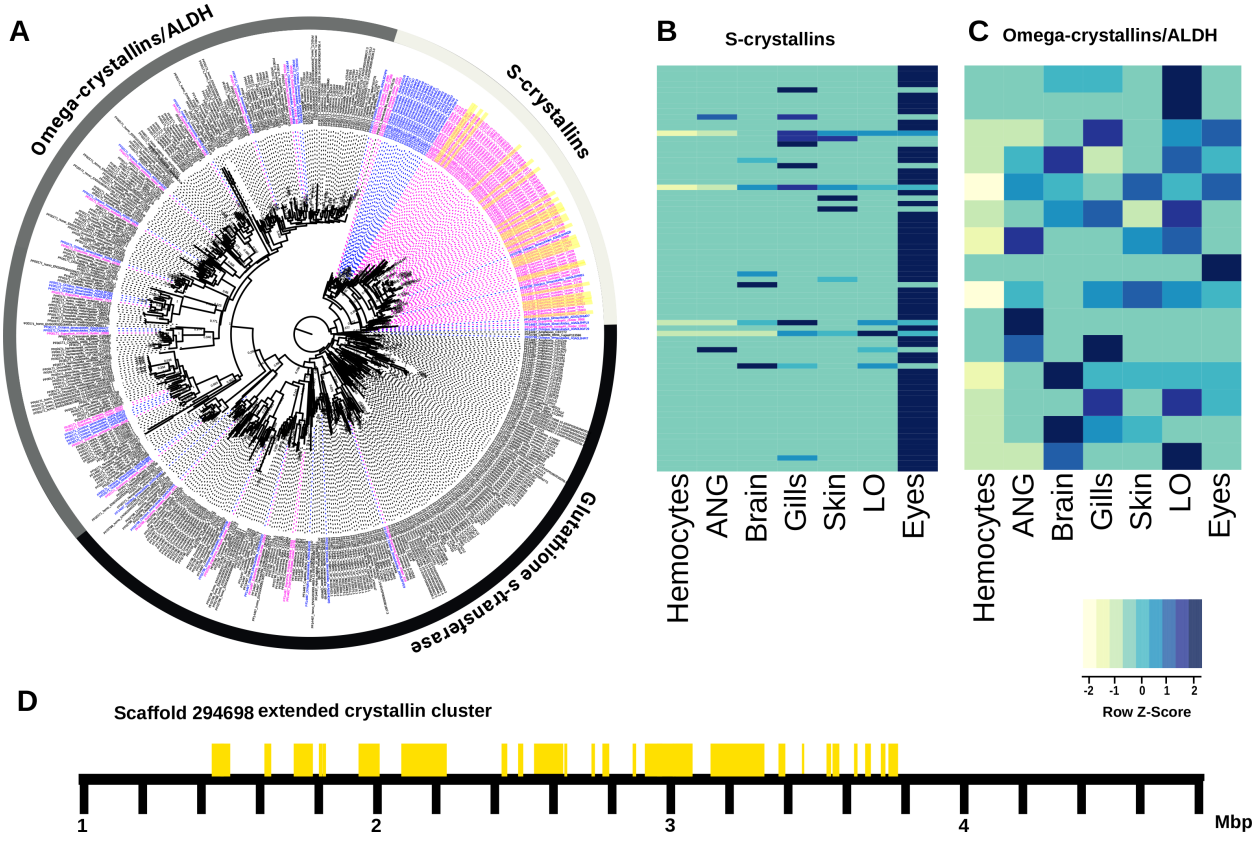

Fig. S7. Crystallins and ANG specific gene linkages. (*A*) FastTree maximum likelihood tree of aldehyde dehydrogenases (ALDH), glutathione-S-transferases (GST), and crystallins indicating expansions of S-crystallins in *E. scolopes* (pink) and *O. bimaculoides* (blue), shared (yellow). (*B*) Expression heatmap is shown for S- and (*C*) ALDH/omega-crystallins in E. scolopes showing eye-specific expression for S-crystallins and enriched LO-expression for ALDH/omega-crystallins. (*D*) Location of the ~2 Mb S-crystallin gene cluster in *E. scolopes*.

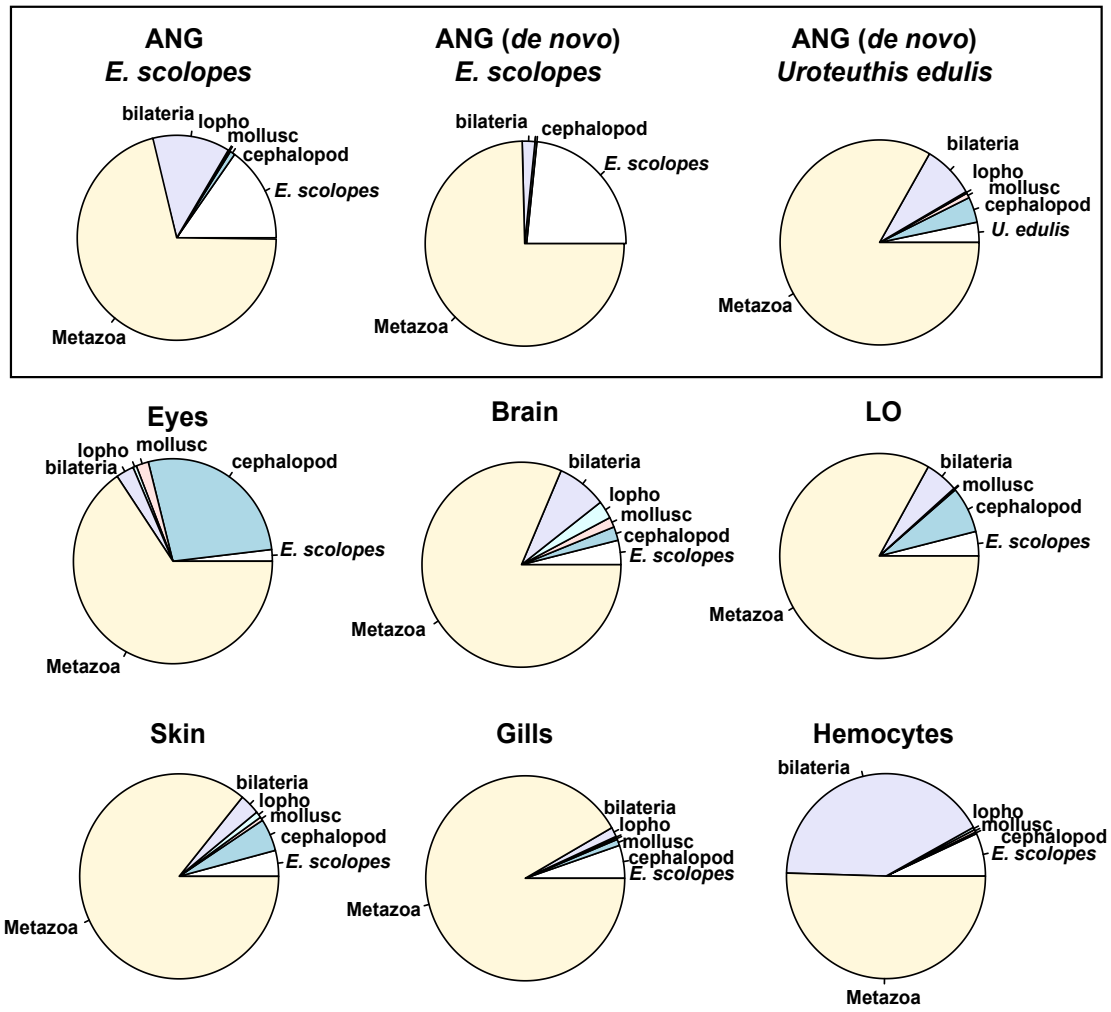

Fig. S8. Contribution of genes from different age categories (i.e., phylostrata) towards tissue transcriptome total expression. The proportion of the total expression is plotted for each phylostratum. For the ANG, we computed the proportions based on the genome (left), de-novo transcriptome assembly (middle), and de-novo transcriptome assembly for ANGs from *Uroteuthis* ANG from Pankey et al. (6).

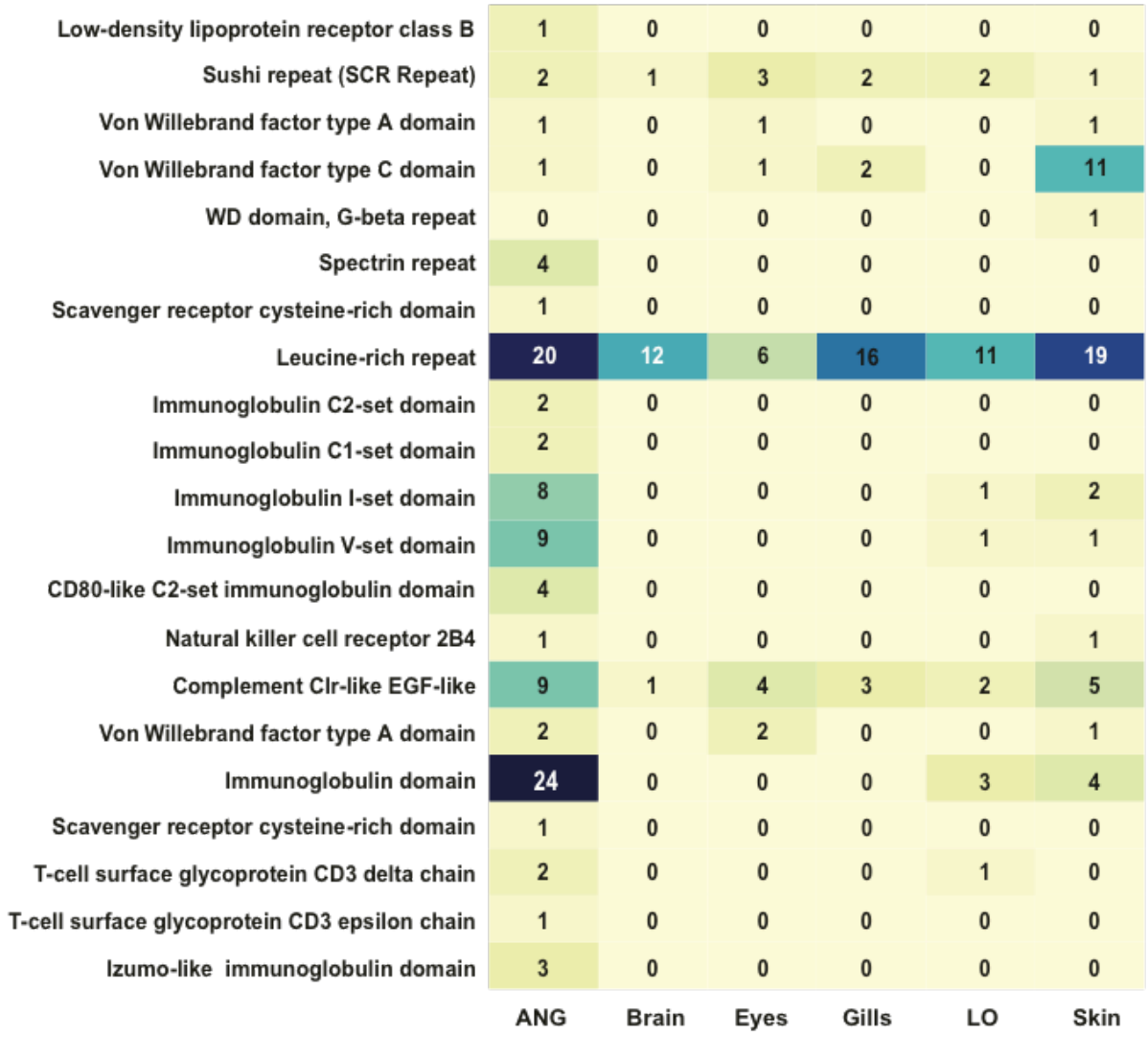

Fig. S9. Immune-related PFAM domains showing distribution among the six different adult tissues analyzed. Distribution of immune-related PFAM domains among the six different adult tissues analyzed. Isoform counts for a given PFAM domain/category are shown for the genes specifically expressed in each of the six adult tissues.

Table S1. Metadata associated with the transcriptomic libraries used for the *Euprymna scolopes* reference transcriptome.

Table S1. Metadata associated with the transcriptomic libraries used for the *Euprymna scolopes* reference transcriptome.

SQL Database name	Tissue Type	Biological Replicates	*V. fischeri* strain exposed to host	Age at sacrifice	RNA Extraction Kit	Sequencing platform	Team generating library	Publication	Additional Information
ANG_Adult	Accessory Nidamental Gland	1	natural population	Adult	MasterPure RNA Purification Kit	Illumina HiSeq	University of Connecticut	This study	Field caught animals
Brain_Adult	Brain	1	natural population	Adult	RiboPure (Ambion)	Illumina HiSeq	University of California Santa Barbara	Pankey et al., 2014	field caught animals
Eyes_Adult	Eyes	1	natural population	Adult	RiboPure (Ambion)	Illumina HiSeq	University of California Santa Barbara	Pankey et al., 2014	field caught animals
Gills_Adult	Gills	1	natural population	Adult	RiboPure (Ambion)	Illumina HiSeq	University of California Santa Barbara	Pankey et al., 2014	field caught animals
Hemocytes_Adult	Hemocytes	1	natural population	Adult	RNAqueous-Microkit (Ambion)	454 platform	University of Connecticut	Collins et al., 2012	field caught animals
Skin_Adult	Skin	1	natural population	Adult	RiboPure (Ambion)	Illumina HiSeq	University of California Santa Barbara	Pankey et al., 2014	field caught animals
LO_Adult	Light Organ	1	natural population	Adult	RiboPure (Ambion)	Illumina HiSeq	University of California Santa Barbara	Pankey et al., 2014	field caught animals
LO_APO_4week	Light Organ	1	none	4 weeks	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	Kremer et al., 2018	4 week old laboratory raised animal
LO_LUX_4week	Light Organ	1	*V. fischeri* Δlux	4 weeks	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	Kremer et al., 2018	4 week old laboratory raised animal
LO_WT_4week	Light Organ	1	*V. fischeri* ES114	4 weeks	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	Kremer et al., 2018	4 week old laboratory raised animal
Eyes_APO_24h	Eyes	1	none	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Eyes_LUX_24 h	Eyes	1	*V. fischeri* Δlux	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Eyes_WT_24h	Eyes	1	*V. fischeri* ES114	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Gills_APO_24h	Gills	1	none	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Gills_LUX_24h	Gills	1	*V. fischeri* Δlux	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Gills_WT_24h	Gills	1	*V. fischeri* ES114	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Head_APO_24h	Head - White Body, brain, optical nerve	1	none	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Head_LUX_24h	Head - White Body, brain, optical nerve	1	*V. fischeri* Δlux	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
Head_WT_24h	Head - White Body, brain, optical nerve	1	*V. fischeri* ES114	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
LO_APO_24h	Light Organ	1	none	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
LO_LUX_24h	Light Organ	1	*V. fischeri* Δlux	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
LO_WT_24h	Light Organ	1	*V. fischeri* ES114	24 hours	RNAeasyMini (Qiagen)	Illumina HiSeq	University of Hawaii	McFall-Ngai unpublished	24 h animals
LO_APO_24h_Gravity	Light Organ	2	none	24 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	24 h animals maintained in high aspect ratio vessels - gravity control position
LO_SYM_24h_Gravity	Light Organ	2	*V. fischeri* ES114	24 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	24 h animals maintained in high aspect ratio vessels - gravity control position
LO_APO_24h_LSMMG	Light Organ	3	none	24 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	24 h animals maintained in high aspect ratio vessels - low shear modeled microgravity
LO_SYM_24h_LSMMG	Light Organ	3	*V. fischeri* ES114	24 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	24 h animals maintained in high aspect ratio vessels - low shear modeled microgravity
LO_APO_12h_Gravity	Light Organ	3	none	12 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	12 h animals maintained in high aspect ratio vessels - gravity control position
LO_SYM_12h_Gravity	Light Organ	3	*V. fischeri* ES114	12 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	12 h animals maintained in high aspect ratio vessels - gravity control position
LO_APO_12h_LSMMG	Light Organ	2	none	12 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	12 h animals maintained in high aspect ratio vessels - low shear modeled microgravity
LO_SYM_12h_LSMMG	Light Organ	3	*V. fischeri* ES114	12 hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	12 h animals maintained in high aspect ratio vessels - low shear modeled microgravity
LO_Hatchling1	Light Organ	2	none	01hours	RNAeasyMini (Qiagen)	Illumina NextSeq500	University of Florida	Casaburi et al., 2017	collected within 30 min hatching
PacBio_Merged	Adults - brain, eyes, white body, optical lobe, gills, skin, light organ; Juveniles 24 h aposymbiotic; 24 h symbiotic; hatchling	1	mixed	mixed ages	RNAeasyMini (Qiagen)	PacBio IsoSeq	University of Florida	This Study	mixed tissues were normalized before library preparation

Table S2. Assembly statistics comparison between *Octopus bimaculoides* and *Euprymna scolopes*

	Euprymna scolopes, min 50 kb* (this study)	*Euprymna scolopes* (this study)	*Octopus bimaculoides* (4)	*Callistoctopus minor* (23)
Size, Gbp	5.11	5.71	2.37	5.09
N50, kbp	3,724	3,171	466	196
Longest scaffold, Mbp	29.65	29.65	4.06	3.03
Total scaffolds above 2kb	3,876	50,192	15,798	41,584
Gaps	35%	33%	15%	0%

*Assembly with scaffolds of 50 kb and longer in length.

Dataset S1. Microsynteny analysis of *Euprymna scolopes* genome. (see attached excel sheet). Columns from left to right: synteny block id, species name, number of orthologous blocks, list of orthologous blocks, number of other species in orthologous blocks, classification (if available), location in the genome, length in nucleotides, syntenic genes ids. Species abbreviations include the following: *Branchiostoma floridae* (Bfl), *Capitella teleta* (Cte)*, Crassostrea gigas* (Cgi), *Euprymna scolopes* (Esc), *Homo sapiens* (Hsa)*, Helobdella robusta* (Hro), *Lottia gigantea* (Lgi), *Nematostella vectensis* (Nve)*, Octopus bimaculoides* (Obi).

Supplemental References:

1. Putnam NH*, et al.* (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* 26(3):342-350.
2. Krueger F (2015) Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files).
3. Chapman JA*, et al.* (2011) Meraculous: de novo genome assembly with short paired-end reads. *PLoS One* 6(8):e23501.
4. Albertin CB*, et al.* (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524(7564):220-224.
5. Simakov O*, et al.* (2013) Insights into bilaterian evolution from three spiralian genomes. *Nature* 493(7433):526-531.
6. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH (2014) Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proc Natl Acad Sci USA* 111(44):E4736-4742.
7. Collins AJ*, et al.* (2012) Diversity and partitioning of bacterial populations within the accessory nidamental gland of the squid *Euprymna scolopes*. *Appl Environ Microbiol* 78(12):4200-4208.
8. Kremer N*, et al.* (2018) Persistent interactions with bacterial symbionts direct mature-host cell morphology and gene expression in the squid-vibrio symbiosis. *mSystems* 3(5):1-17.
9. Casaburi G, Goncharenko-Foster I, Duscher AA, Foster JS (2017) Transcriptomic changes in an animal-bacterial symbiosis under modeled microgravity conditions. *Sci Rep* 7:46318.
10. Grabherr MG*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol* 29(7):644-652.
11. Haas BJ*, et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Prot* 8(8):1494-1512.
12. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659.
13. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210-3212.
14. Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9):1859-1875.

15. Kapustin Y, Souvorov A, Tatusova T, Lipman D (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct* 3:20.
16. Simakov O*, et al.* (2015) Hemichordate genomes and deuterostome origins. *Nature* 527(7579):459-465.
17. Ashburner M*, et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25-29.
18. Finn RD*, et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279-285.
19. Tomarev SI*, et al.* (1993) Abundant mRNAs in the squid light organ encode proteins with a high similarity to mammalian peroxidases. *Gene* 132(2):219-226.
20. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
21. Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLOS ONE* 5(3):e9490.
22. Warnes MGR, Bolker B, Bonebakker L, Gentleman R (2016) Package gplots. *Various R Programming Tools for Plotting Data*.
23. Kim BM*, et al.* (2018) The genome of common long-arm octopus *Octopus minor*. *Gigascience*.