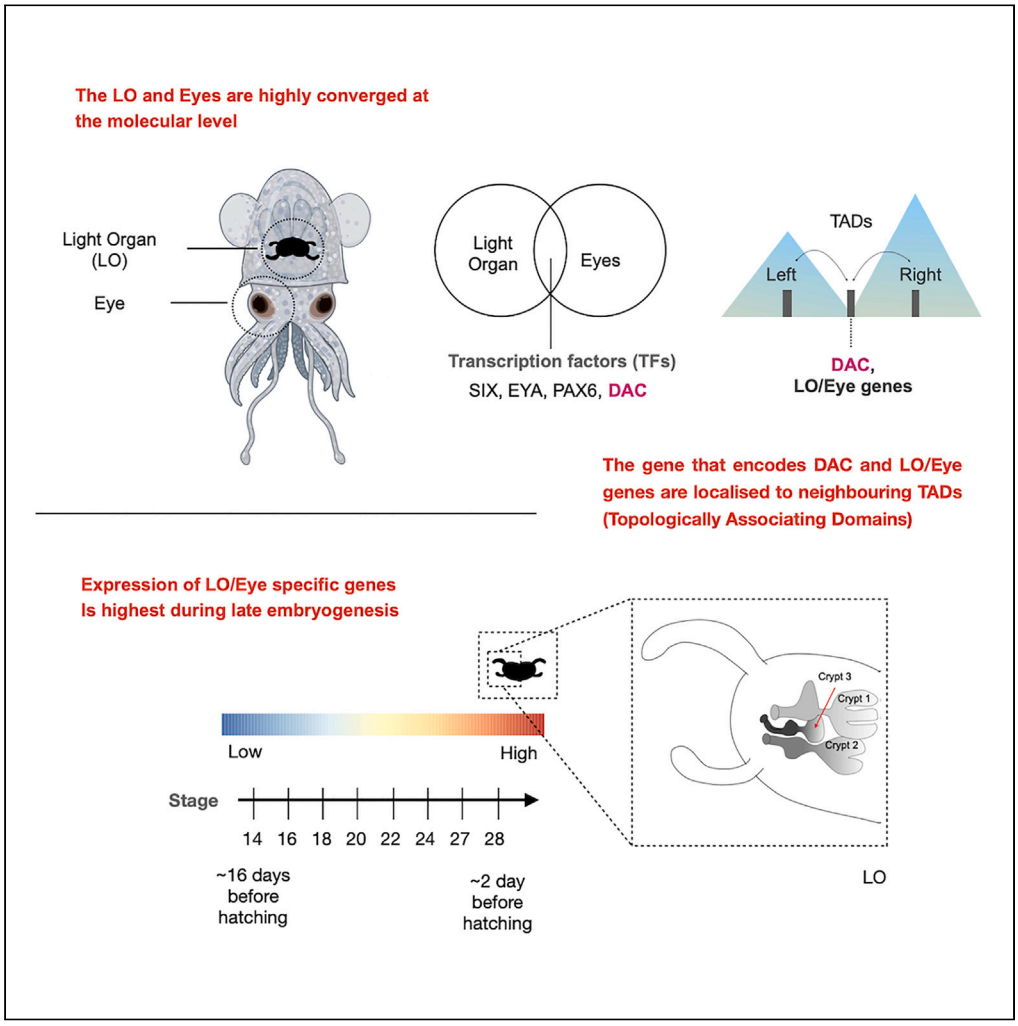
iScience

CellPress
OPEN ACCESS

Article

# Emergence of novel genomic regulatory regions associated with light-organ development in the bobtail squid




Lisa Rouressol, John Briseno, Nidhi Vijayan, ..., Spencer V. Nyholm, Margaret J. McFall-Ngai, Oleg Simakov

lisa.rouressol@univie.ac.at (L.R.)
mmcfallngai@carnegiescience.edu (M.J.M.-N.)
oleg.simakov@univie.ac.at (O.S.)

## Highlights

Chromosomal-conformation capture reveals novel networks of co-regulated LO/eye genes

Most LO/eye genes and the new DAC network are located at boundaries between two TADs

LO/eyes genes are highly expressed in late/stage embryogenesis as LO crypt 3 forms

A new DAC network is present in *Euprymna scolopes* suggesting a link with LO emergence

Rouressol et al., iScience *26*, 107091
July 21, 2023 
https://doi.org/10.1016/j.isci.2023.107091

iScience

CellPress
OPEN ACCESS

Article

# Emergence of novel genomic regulatory regions associated with light-organ development in the bobtail squid

Lisa Rouressol,[1,2,6,*] John Briseno,[3] Nidhi Vijayan,[3] Grischa Y. Chen,[2] Elena A. Ritschard,[1,4] Gustavo Sanchez,[5] Spencer V. Nyholm,[3] Margaret J. McFall-Ngai,[2,*] and Oleg Simakov[1,*]

## SUMMARY

**Light organs (LO) with symbiotic bioluminescent bacteria are hallmarks of many bobtail squid species. These organs possess structural and functional features to modulate light, analogous to those found in coleoid eyes. Previous studies identified four transcription factors and modulators (SIX, EYA, PAX6, DAC) associated with both eyes and light organ development, suggesting co-option of a highly conserved gene regulatory network. Using available topological, open chromatin, and transcriptomic data, we explore the regulatory landscape around the four transcription factors as well as genes associated with LO and shared LO/eye expression. This analysis revealed several closely associated and putatively co-regulated genes. Comparative genomic analyses identified distinct evolutionary origins of these putative regulatory associations, with the DAC locus showing a unique topological and evolutionarily recent organization. We discuss different scenarios of modifications to genome topology and how these changes may have contributed to the evolutionary emergence of the light organ.**

## INTRODUCTION

The coleoid cephalopod clade, emerging around 400 million years ago (mya),[1,2] (octopus, squid, cuttlefish) possesses the largest invertebrate nervous system, a feature associated with extensive behavioral and camouflage abilities.[3,4] Although certain genomic features have been associated with morphological innovations in cephalopods,[5] because of the large genetic distances between cephalopods and other mollusks, as well as the lack of regulatory genomic data in this group; it has been challenging to identify and reconstruct the genomic innovations within this clade. Only recently, potential links between large-scale genomic rearrangements in the coleoid ancestor and organismal novelties[6–8] were uncovered.

One of the most prominent morphological novelties in the cephalopod clade is the convergent evolution of the symbiotic light organ (LO) associated with two families, the Sepiolidae (bobtail squid) and Loliginidae (pencil squid). Although the LO is present in several genera, including *Uroteuthis* spp. from the Loliginidae lineage and *Euprymna* spp. of the Sepiolidae lineage[9] (Figures 1A and 1B), its origin remains elusive.[10] These cephalopods use their LOs in a nocturnal camouflage behavior called 'counterillumination', in which the animals mask themselves against a background of downwelling moonlight and starlight to avoid predators.[11] One of the most established models in LO symbiosis and evolution research is the Hawaiian bobtail squid, *Euprymna scolopes*.[12] *E. scolopes* is easily raised in the laboratory and amenable to experimentation.[12] The adult LO that is located in the center of the mantle cavity of *E. scolopes* comprises a set of tissues that is used to modulate light. *E. scolopes* has been extensively studied due its binary relationship with the bioluminescent bacterium *Vibrio fischeri*.[12,13] Shortly after hatching, juvenile *E. scolopes* recruit *V. fischeri* from the surrounding environment and house them in the LO in two sets of epithelium-lined crypts, three on each side of the LO.[14,15]

Intriguingly, the LO resembles the cephalopod camera eye with lens, choroid, iris, and tapetum analogues.[16] The evolutionary origins of the light organ are still enigmatic. Several papers have provided a "gene list" characterization (e.g., Belcaid et al., 2019[17]), and transcriptomic analyses in *E. scolopes* revealed

[1]Department for Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria

[2]Department of Biosphere Sciences and Engineering, Carnegie Institution for Science, Pasadena, CA 91125, USA

[3]Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA

[4]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, 80121 Napoli, NA, Italy

[5]Molecular Genetics Unit, Okinawa Institute of Science and Technology, Okinawa 904-0495, Japan

[6]Lead contact

*Correspondence: lisa.rouressol@univie.ac.at (L.R.), mmcfallngai@carnegiescience.edu (M.J.M.-N.), oleg.simakov@univie.ac.at (O.S.)

https://doi.org/10.1016/j.isci.2023.107091

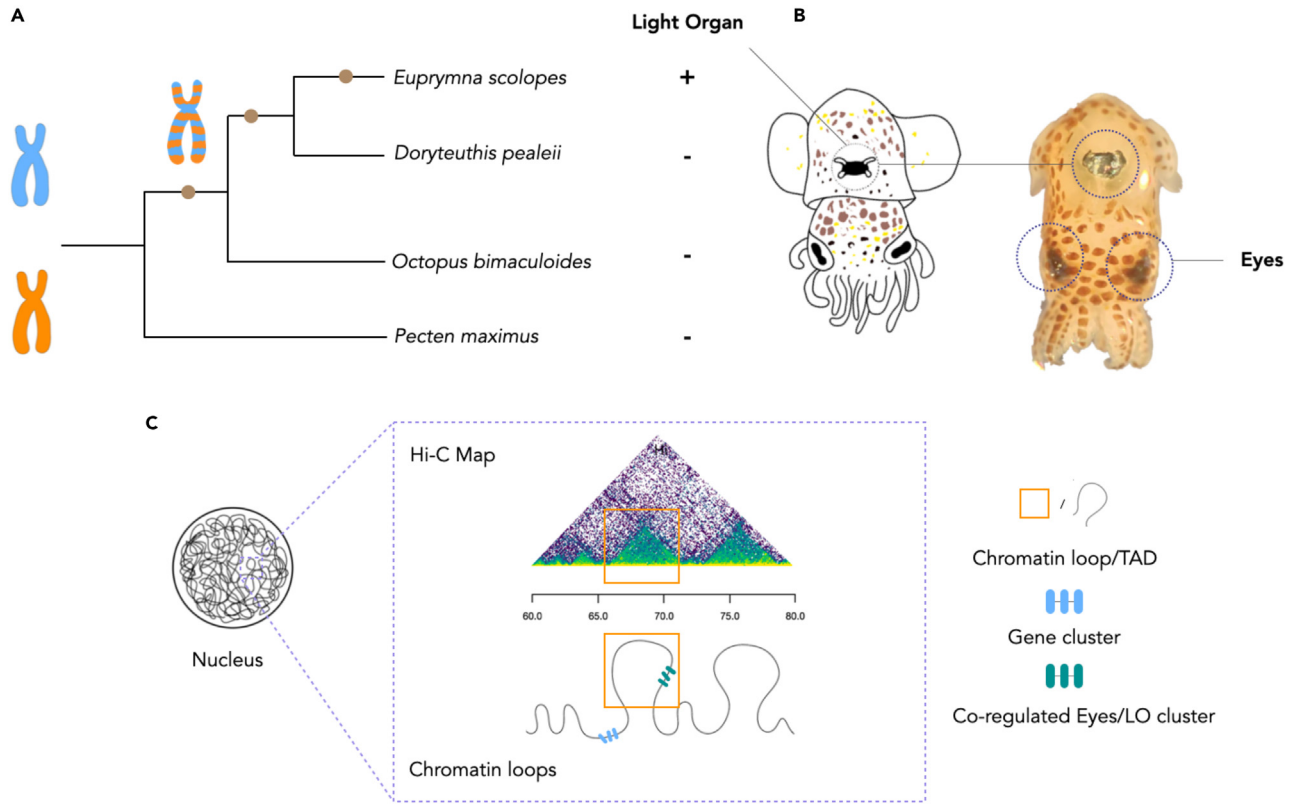



**Figure 1. Ancient coleoid cephalopod genome rearrangements and their role in the emergence of novel regulatory neighborhoods**

(A) Phylogenetic tree depicting how a series of genomic rearrangements (from left to right) in the coleiod cephalopod clade resulted in development of an LO in *E. scolopes*. LO presence (+)/absence (–) in species: *E. scolopes* (bobtail squid), *D. pealeii* (longfin squid),*O. bimaculoides* (two-spot octopus) and *P. maximus* (king scallop); the brown dots on the phylogenetic tree represent the novel regulatory gene linkages.

(B) Diagram (left) and photograph (right) of the position of the LO in the mantle cavity of a juvenile *E. scolopes*.

(C) Left, an ultrastructural depiction of condensed chromatin in a nucleus; middle upper, a Hi-C Map zoomed in; middle lower, a depiction of the chromatin loops which contribute to TAD formation. Right, the square represents a TAD unit formed by chromatin loop. Inside the TAD, genes clusters may interact with each other more frequently than with genes outside the TAD. The arrangements suggest that co-regulated LO/EYA gene clusters occur inside the TADs.

expression of genes from the rhabdomeric and ciliary visual transduction pathways.[18] One of the most crucial insights came from a study that showed that in *E. scolopes* the genes encoding four transcription factors (TFs) and modulators: EYA (eyes absent), SIX (sine oculis), PAX6, and DAC (dachshund) are upregulated during both eye and LO development.[19,20] The network and interaction of these genes have long been implicated in various organogenesis processes, most prominently during eye development of Drosophila[21] and early sensory placode development in vertebrates.[22]

EYA is a transcriptional modulator, its phosphatase activity has been shown to act on SIX and DACH[23] during various organogenetic processes, including kidney, branchial arches and general sensory organ (eyes, ears) development. In mice, EYA is expressed in the eyes, ears, nasal placode, and metanephric mesenchyme.[24–26] In *Drosophila*, EYA is expressed in different tissues including eyes and muscle.[27–29] In *E. scolopes*, EYA is expressed in the eyes, optic lobes, statocysts, olfactory organs, tentacles, mantle, gills and light organ[20]; in worms (*Schmidtea polychroa, Dugesia japonica*) EYA is expressed in the eyes.[30,31]

Pax6 is a transcription factor that belongs to the highly conserved paired box family (Noll, 1993). In *Drosophila*, mice, zebrafish, PAX6 is involved in eye morphogenesis and mostly expressed in the optic region, and brain.[32–34] In coleoids, it is known that PAX6 is expressed in the eyes (*Sepia officinalis*, *Loligo opalescens*, *E. scolopes*, *Nautilus pompilius*), in the gills (*S. officinalis*, *E. scolopes*), in sensory organs, as well as in olfactory organs (*L. opalescens*, *E. scolopes*), arms (*S. officinalis*, *L. opalescens*, *E. scolopes*) and mantle (*L. opalescens*, *E. scolopes*).[20,35–38]

SIX is a member of the homeobox family of transcription factors. In *Drosophila*, it is required for compound eye formation[39,40] and muscle development.[29] In vertebrates, SIX genes play critical roles in tissue formation and organogenesis, such as for the head, ear, retina, nose, brain, skeletal muscle, and kidney.[26,41,42] In these tissues, the SIX family of transcription factors function in regulation of progenitor cell maintenance and differentiation. In jellyfish (*Cladonema radiatum*), SIX is expressed in the eyes, manubrium, tentacles, gonads, and umbrella[43]; in worms (*S. polychroa*, *D. japonica*) SIX is expressed in the eyes.[30,31] In cephalopods (*E. scolopes*) SIX is expressed in eyes, optic lobe, statocysts, olfactory organ, tentacles, mantle, gills and light organ.[20]

DAC is a transcription factor involved in the development of the compound eye. In mice, DAC is expressed in the eyes, optic cup, neural crest, brain, limb, otic vesicle, and genitalia.[44–46] In *Drosophila*, DAC is expressed in eyes, limbs and muscles.[29,47] In fish (*Oryzias latipes*) DAC is expressed in the eyes, central nervous system, pancreas and finbuds.[48] In *E. scolopes* DAC is expressed in eyes, optic lobes, statocysts, olfactory organs, tentacles, mantle, gills and light organ[20] and in *S. officinalis*, DAC is expressed in the growing limb.[49] In polychaetes, DAC is expressed in eyes.[31]

Together with the more recent global transcriptomic findings ([Belcaid et al. ref]), the expression of these genes in the developing bobtail squid eye and LO suggests a deep evolutionary link between the developmental programs in these organs and leads to the hypothesis that co-option of the eye regulatory network may be responsible for the development of the LO.

Of interest, evolutionary tinkering through co-option of these genes in the LO development is an evolutionarily recent event. Recent studies suggest that the LO first evolved ~50 million years ago (mya) in the ancestral sepiolinid.[11] To begin investigating the genomic changes leading to the LO emergence, it is crucial to consider the modality of cephalopod genome evolution. Coleoid cephalopod genomes share several derived changes relative to other, more ancestral, molluscan genomes. One of the most striking of these changes is the evolution of a new chromosomal complement.[7,8] Despite the large genome size of coleoid cephalopods, this chromosomal reorganization did not involve whole genome duplication, unlike in vertebrates, and is rather a result of a seemingly random combination of ancestral chromosomal fragments. This pattern indicates a period of large genomic restructuring in the coleoid lineage.[7,8]

Such rearrangements may constitute one of the main drivers behind novel traits emergence within coleoids.[8] A recent study suggested that whole genome rearrangements produced novel local co-regulated gene neighborhoods in both octopus and squid lineages.[8] Co-regulation occurs either through common regulatory elements that drive co-expression of neighboring genes, or through regulatory regions found in an intron of one gene that controls expression of a neighboring target gene (genomic regulatory block, GRB).[50] Often, such co-regulated gene neighborhoods are co-localized in the same topological vicinity, known as topologically associated domains (TADs). Over the past several years, TADs have emerged as units of chromosomal topology, in which the genes located inside the TADs are more likely to form regulatory interactions between each other, whereas the genes located in neighboring TADs, separated by insulator elements, are prohibited from coming into the same regulatory vicinity[51,52] (Figure 1C). Coleoid cephalopod genomes were found to be enriched in such novel co-regulated regions, especially of the GRB-type co-regulation,[8] however their role in specific morphological innovations has not been explored.

The principal aim of the paper is to obtain a better understanding of the emergence of the light organ by characterizing the regulatory genomic neighborhood and its evolutionary history for the genes involved in both eye and LO development. Specifically, by reassessing the published information on both developmental and functional identity genes involved in LO development and function, we investigate their evolutionary history in light of what is known about global trends in cephalopod genome evolution. Using available and newly generated transcriptomic data, we reassess previous findings on transcriptional activity of genes involved in eye and LO development and function (Peyer et al., 2014).[17,20] We then take this targeted approach by examining genes in direct vicinity of these transcription factors to shed light on potential novel co-regulation. We investigate the expression, regulation, and genome topology of the four proposed transcription factors (EYA, SIX, PAX6, DAC) and other genes that define eye and LO. We also investigate and reevaluate LO-specific gene signature, revealing that, similar to previous studies, most of them are terminal functional genes (e.g., crystallins). Neither gene duplication of these four transcription factors nor

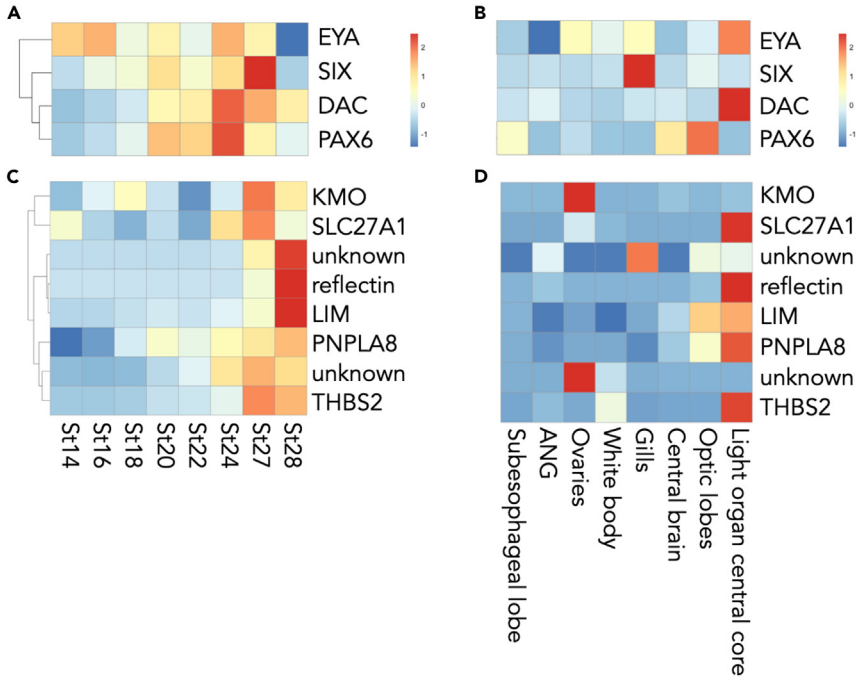



**Figure 2. Expression dynamics of shared eye-light organ genes during development and in adult tissues**

(A–D) The relative gene expression of four proposed co-opted transcription factors (PAX6, SIX, DAC, EYA) (A) over developmental stages 14–28 (14 corresponding to day 5 and 28 to day 19 post fertilization) before hatching (around day 21), and (B) expression in adult tissues: light organ central core, optic lobes, central brain, gills, white body, ovaries, ANG (accessory nidamental gland), and subesophageal lobe. Relative gene expression of eight select eye/LO co-expressed genes according to new transcriptomic dataset over (C) developmental time course and (D) adult tissues. A *Z* score normalization was performed on the normalized read counts across samples for each gene. Z-scores were computed on a gene-by-gene (row-by-row) basis by subtracting the mean and then dividing by the standard deviation.

enrichment of orphan genes can adequately explain LO emergence.[17] Therefore, we test the hypothesis that, instead, gene co-option through regulatory region modification and formation of novel putatively functional gene neighborhoods are responsible for LO emergence.

Our study reveals potential novel regulatory interactors and provides for several testable hypotheses around the four transcription factors implicated in LO and eye development, highlighting potential key players such as DAC. This study thus lays the foundation for future mechanistic studies of these genes.

## RESULTS

### Expression of LO and eye genes during embryonic development

To identify gene sets corresponding to eye and/or LO development, we analyzed the embryonic transcriptome datasets from Schmidbaur et al., 2022, covering developmental stages 14, 16, 18, 20, 22, 24, 27, 28 – hatching day corresponds to stage 30 – described in Lee et al., 2009,[15] and two adult transcriptomic datasets, the one from Belcaid et al., 2019[17] and a new adult tissue sampling for this study. We treated the two datasets separately because of technical biases in their generation (such as extraction and sequencing procedures), which might have introduced batch effects in our comparisons. We therefore ran analyses separately on both transcriptomes and use them to cross-validate each other. One notable distinctive feature is that Belcaid et al., 2019[17] provides transcriptome data for the whole LO tissue, whereas our new data contains transcriptomic data only from the central core of the LO's.

First we examined the expression of the LO- and eye-specific transcription factors (SIX, EYA, PAX6 and DAC)[20] over *E. scolopes* (1) embryological development (stages 14–28, which represent day 5–19 post fertilization)[15] and (2) in several adult tissues. All LO- and eye-specific transcription factors were most highly expressed late in development, around stage 24 and stage 27, which is characteristic of eye and optic lobe development[15] (Figure 2A). We found that EYA is expressed not only in eye tissue and the LO but also in the

ovaries and gills of adults. DAC is expressed predominantly in the LO; PAX6 is expressed in the brain tissue and eyes, whereas SIX is expressed only in gills (Figure 2B).

Next we identified a total of eleven genes with high expression in both the LO and eye (nine genes according to the Belcaid et al., 2019[17] dataset) (Figure S1B, Table S1) or LO and optic lobe (two genes according to the tissue dataset presented here: PNPLA8 and LIM, Figures 2C and 2D). The remaining eight LO- and eye-specific genes were highly expressed during late embryogenesis, around stage 24, 27 and 28 (Figure 2C) and most were highly expressed in the LO (Figure 2D). Unlike the strong eye expression in the Belcaid et al., 2019[17] dataset, these genes were not highly expressed in the optic lobes (Figure 2D). Several gene categories were represented, including: a reflectin gene, which contributes to cephalopod dynamic pigmentation and iridescence[53]; the THBS2 gene (thrombospondin-2), which mediates cell-to-cell and cell-to-matrix interactions in humans[54]; the KMO gene (kynurenine 3-monooxygenase), involved in the activation of cytokine mediated changes in behavior because of inflammatory stimuli such as infections in mammals[55]; the SLC27A1 gene, which encodes the long-chain fatty acids transport protein[56]; and, five genes of unknown function in *E. scolopes* (no annotation: uncharacterized protein and/or no significant similarity found when blasting the sequences). Other genes identified only from newly generated adult transcriptomic data (Figure S2A) include the PNPLA8 gene encoding the protein calcium-independent phospholipase A2-gamma,[57] and the gene encoding LIM domain-containing protein A isoform X2. The LIM-containing proteins are involved in diverse biological processes including cytoskeleton organization, cell lineage specification and organ development.[58]

Using our new dataset, we also identified seventeen genes uniquely expressed in the LO tissue (Figure S2A and Table S2). The LO-specific genes were mainly expressed in late embryogenesis, in stage 28 (approx. day 19; Figure S2B), which corresponds to the beginning of the formation of the third and final pair of LO crypts.[14,15] The majority of LO-specific genes were halide peroxidase[59] (also referred to as melanogenesis genes) melanin-producing genes and the galaxin gene EsGal1 – which is implicated in the selection and modulating growth of *V. fischeri* during symbiosis onset.[60] We also found SULT1B1 gene, used to catalyze the sulfation of many hormones and neurotransmitters[61]; a selenium-dependent glutathione peroxidase, belonging to a class of antioxidant enzymes with the capacity to scavenge free radicals[62]; the omega crystallin gene, which encodes structural components of squid and octopus eye lenses,[63,64] as well as light-organ lenses[65] the Corin gene, which regulates blood volume and pressure in vertebrates[66]; and the AGL gene, known for breakdown of glycogen.[67] There were also 3 genes of unknown function in this dataset (no annotation: uncharacterized protein and/or no significant similarity found when blasting the sequences).

In summary, LO and shared LO/eye-specific genes were expressed during the very late stages of embryogenesis, after the earlier peak of the four developmental TFs. This expression suggests that transcriptional regulation of these LO/eye-specific genes by the DAC, EYA, PAX6, or SIX TFs may be required for LO organogenesis, though more evidence is needed to validate this hypothesis.

## Topological organization of genomic eye/LO loci reveals potentially co-regulated genes

Localization of co-regulated genes in the same genomic vicinity is a key aspect of gene regulation.[51,68,69] Schmidbaur et al., *2022*[8] provided first insights that topological domains, i.e., TADs, in the *E. scolopes* genome contain several co-regulated genes associated with putative regulatory sequences, as profiled by an assay for transposase accessible chromatin (ATAC). Closely positioned genes often contain regulatory regions that control the expression of their neighbors. Topological organization around LO/eye-specific genes may thus reveal the regulatory landscape and identify a possible convergent evolutionary mechanism for their emergence.

Using the available data from chromosomal conformation capture (Hi-C) from Schmidbaur et al., *2022*,[8] we profiled interaction and insulation scores around the four transcription-factor genes. HiC interaction plots and insulation score analysis showed that DAC was at the boundary between two compartments (Figures 3A and 3B) whereas SIX, PAX6, EYA were located centrally in a compartment (Figures S3A, S3B, S4A, S4B, S5A and S5B).

We annotated interacting genes in the same and neighboring topological vicinity and profiled them based on their putative co-regulation during development (Table S3). Our conditions for proposing

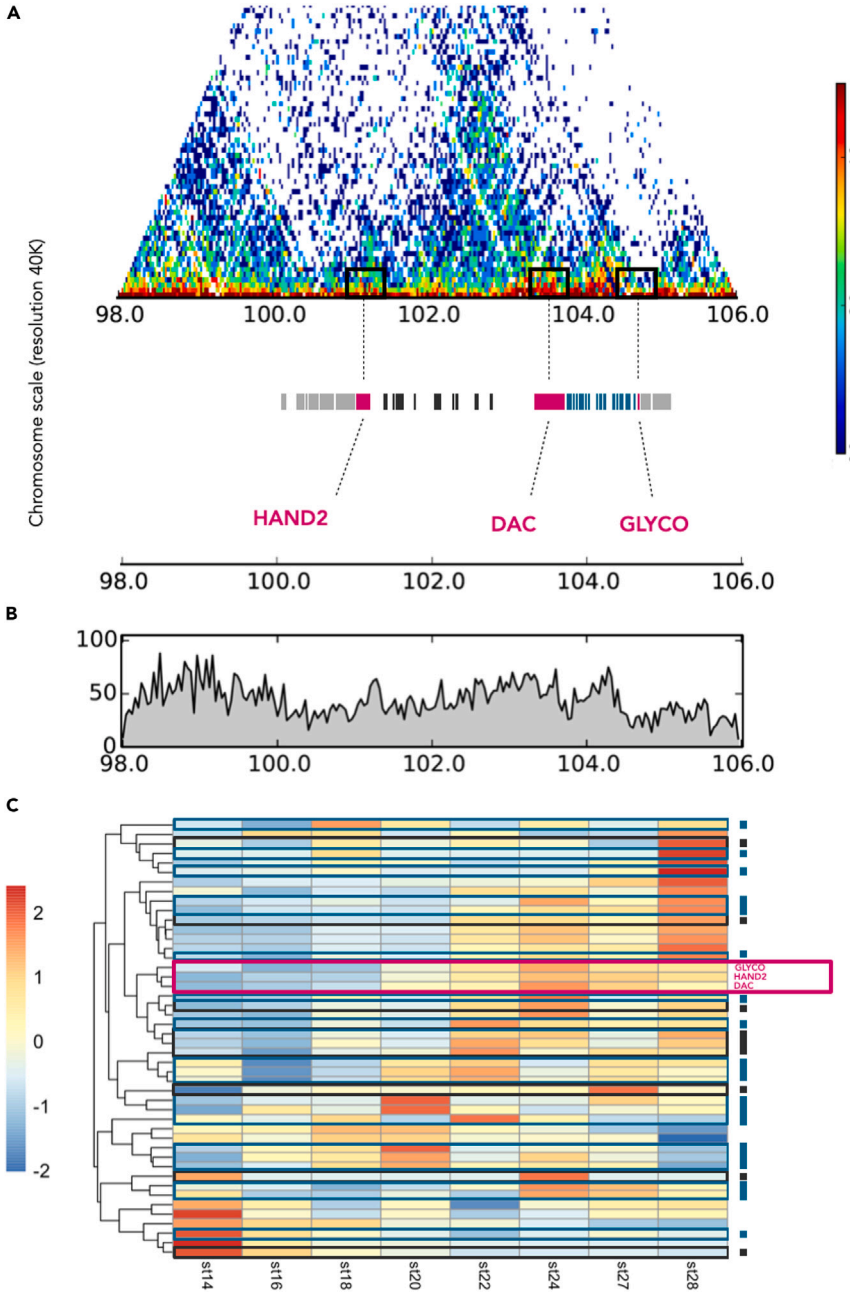



**Figure 3. Genome topology and putative co-regulated genes around DAC**

(A) Hi-C map of the DAC network of a portion of chromosome 3. Genes within the DAC containing TAD are plotted as well as of the DAC coregulated genes, HAND2 and GLYC.

(B) The calculated insulation score across chromosome 3.

(C) Heatmap of genes within the DAC containing TAD and adjacent gene clusters over developmental stages 14–28. Coexpressed genes, HAND2 and GLYCO are highlighted in purple. Genes located between HAND2 and DAC are highlighted in black; genes located between DAC and GLYCO are highlighted in green. We selected HAND2 and GLYCO that most closely share developmental gene expression profile, as determined by the heatmap clustering. For the heatmap, a *Z* score normalization was performed on the normalized read counts across samples for each gene. Z-scores were computed on a gene-by-gene (row-by-row) basis by subtracting the mean and then dividing by the standard deviation.

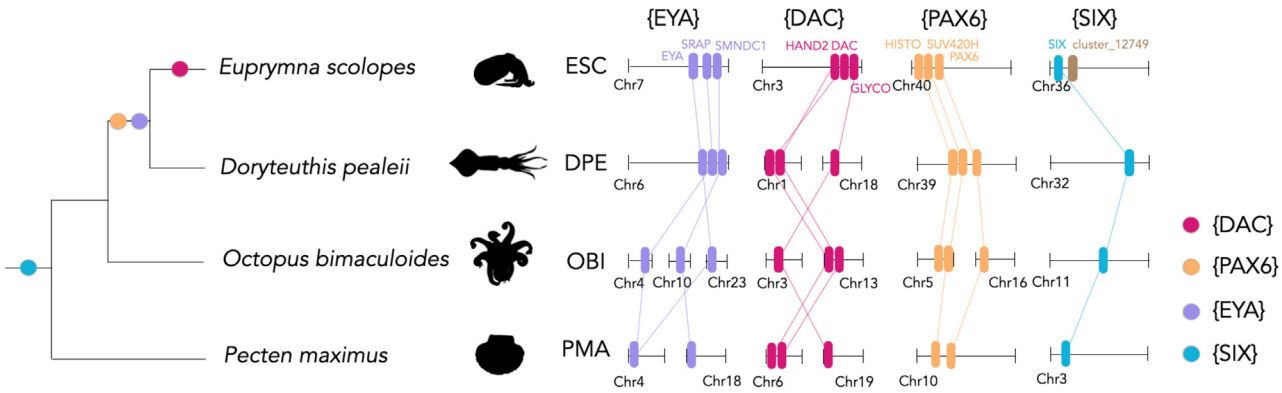



**Figure 4. Evolutionary emergence of co-regulated eye/LO regions**

Genomic comparison of the newly predicted co-regulated genes and the 4 eye/LO specific TFs in *E. scolopes* (bobtail squid), *D. pealeii* (longfin squid), *O. bimaculoides* (two-spot octopus) and *P. maximus* (king scallop). Genes are positioned according to their locations within the chromosomes. {DAC} is represented in magenta, {PAX6} in yellow, {EYA} in purple and {SIX} in blue.

co-regulation are: (1) genomic vicinity within the same or neighboring TAD; and (2) developmental co-expression. We chose the genes that most closely share developmental gene expression profile, as determined by the heatmap clustering. For DAC, we found two co-regulated genes: GLYCO, glucosyltransferase enzymes that are encoded by members of the GTF gene family,[70] and the HAND2 gene, which is part of the helix-loop-helix family of transcription factors that play an essential role in cardiac morphogenesis[71] (Figures 3C and S6A–S6C). We discovered two genes co-regulated with EYA (Figures S3C and S7A–S7C): SraP, which is a serine-rich adhesin that is hypothesized to mediate ligand binding and/or cell adhesion of platelets,[72] and SMNDC1, which encodes the survival motor-neuron protein.[73] For PAX6 (Figures S4C and S8A–S8C), we found two methylation regulators: SUV420H, which is responsible for chromatin compaction and gene repression,[72] HISTO, which are histone methyltransferases that have the same function,[74] and another unknown/novel gene. Finally, we found that SIX is co-regulated only with an orphan gene that does not have any similarity to other sequences and is not associated with repetitive elements (Figures S5C and S9A–S9C). We denote these novel putative co-regulated interactors with parenthetical brackets respective to their locus, i.e., {EYA}, {DAC}, {SIX}, {PAX6}.

We next investigated topological organization around genes expressed in eyes and LO. We found that 5 out of 8 such genes were located at the TAD boundaries (Figure S10), whereas the remaining 3 were inside a TAD. For the LO specific genes, 12 out of 17 were found inside a TAD and 5 were at the TAD boundaries (Figure S11). We also found varying gene density, especially in the regions of low gene density, presence of ATAC peaks indicating putative regulatory elements. Although gene regulation within TADs has been reported in *E. scolopes,*[8] the finding that some of the genes, most prominently DAC, can be located at the TAD boundaries has not been reported. Such arrangement may have been instrumental in LO evolution, as co-option between eye and LO regulatory cascades were driven not only through addition of novel regulatory elements, but through bridging available and neighboring regulatory units (TADs).

## Evolutionary emergence of the LO-eye co-regulated loci

To assess the evolutionary emergence of {PAX6}, {DAC}, {SIX}, {EYA} loci, we conducted genomic comparisons between the available genomes of squids (*E. scolopes*, *Doryteuthis pealeii*), the octopus *Octopus bimaculoides* and the scallop *Pecten maximus* (Figure 4 and Table S4).[7,75] We found that DAC clustering with GLYCO and HAND2 genes was present only in *E. scolopes,* suggesting that the emergence of its unique topological organization also correlates with the emergence of the LO in this clade. For {EYA} and {PAX6} loci, we found the associated genes in the same genomic vicinity only in *E. scolopes* and *D. pealeii,* but not in *O. bimaculoides* or *P. maximus* where they were located on different chromosomes. These data indicate ancient chromosomal translocations at the stem of the squid lineage that brought the co-expressed and putatively co-regulated {EYA} and {PAX6} genes together. These translocations occurred on top of the conserved local gene neighborhood of those genes, such as the otoferlin gene linkages with EYA and

scavenger-receptor like gene linkage with PAX6 that exist broadly across more deeply branching metazoan species,[8] indicating ongoing evolutionary process of bringing new genes into local co-regulatory vicinity. Finally, {SIX} neighborhood is unique as it occupies the center of an 4Mb TAD of almost no coding gene content. Without more genomes of closely related species, it is currently not possible to infer whether this large gene desert is shared across LO-containing bobtail squids. An indication that this expansion may be a novelty can be seen in the presence of one putative *Euprymna*-specific orphan and SIX co-expressed gene (Figure S3). The different evolutionary origin of these associations highlights the iterative process of regulatory modifications and, together with topological data, may point to the different mechanisms at play during LO evolution.

## DISCUSSION

### The mode of symbiotic organ evolution inferred from squid genome

The genomic changes associated with the origin and evolution of novel organs are often difficult to discern, as there are several paths to generate complex gene regulation, including, but not limited to: novel genes, gene duplication, and evolution of noncoding regulatory elements.[5] A recent genomic study[17] pointed to divergent modes of symbiotic organ evolution in the bobtail squid. Of interest, the light organ, which is present only in some species of the squid lineages, did not show any transcriptomic evidence of duplicated gene families or novel genes. The major genomic driving force behind its evolution has thus been elusive.

In this study we explore the hypothesis that LO evolution has been associated with the expansion and novel regulatory associations of the core eye gene regulation network.[20] Our paper discusses the transcriptomic complement across adult tissues and developmental time of LO-specific genes, LO/eye-shared genes, and the four developmental transcription factors (DAC, EYE, PAX6, SIX), previously proposed to be co-opted in both LO and eye development.[20] We show that the four transcription factors are co-regulated at later developmental stages, followed by the activation of the LO and LO/eye-shared genes. The LO and LO/eye shared genes have specific functional categories represented, involved in basic LO and eye functionality. Of interest, LO-only genes had similar predicted functional categories, indicating that similar sets of genes are involved in both LO and eye differentiation and functionality.[76]

The gene expression related to functionality is established at very late developmental stages, coinciding with the development of the final (third) crypt pair of the LO. In total, we predict that 42 genes are transcriptionally involved in LO development (Tables S1, S2, and S3). Although our methods rely on several published and newly available transcriptomes of adult tissues, further validation using *in situ* expression or broader transcriptomic sampling is needed to validate these claims and tissue specificity.

### Modification to the topological organization as a driving force behind gene co-option

To test whether local gene regulation may have been impacted during the evolution of the LO, we furthermore investigated topological organization around the developmental transcription factors and the LO/eye genes using published HiC data.[8] We found that, unlike most cephalopod genes investigated,[8] some of the LO/eye genes and at least one of the four proposed TFs were often located toward the boundaries of topological domains. Such localization may have been crucial in extending regulatory interactions in the local vicinity beyond a single topologically associating domain. To this end, we found several closely positioned putatively co-regulated interactors, such as GLYCO and HAND2 in the vicinity of DAC. Such novel interactors may form the core developmental cassette required for LO emergence (Figure 4). In this context, the recent origin of the {DAC} association (Figure 4) is striking and may indicate a key event that enabled LO evolution. On the contrary, the {SIX} genomic vicinity (shared with non-cephalopod mollusks) occupies a central position in a TAD (Figure S5A). We thus find specific modification to the regulatory regions of some of the involved LO/eye TFs.

Similar organization, i.e., positioning at a TAD boundary, has been famously observed for the hoxD cluster of vertebrates involved in limb and digit development through sequential activation of either telomeric or centromeric TAD.[68,69,77,78] During vertebrate development, first one and then another TAD is active and results in different hox gene expression. According to this regulatory model, we propose that a similar scenario may be the case for DAC and other genes involved in LO and eye development (Figure 5A); each of the TADs around the shared LO-eye genes would harness regulatory sequences and co-regulated genes

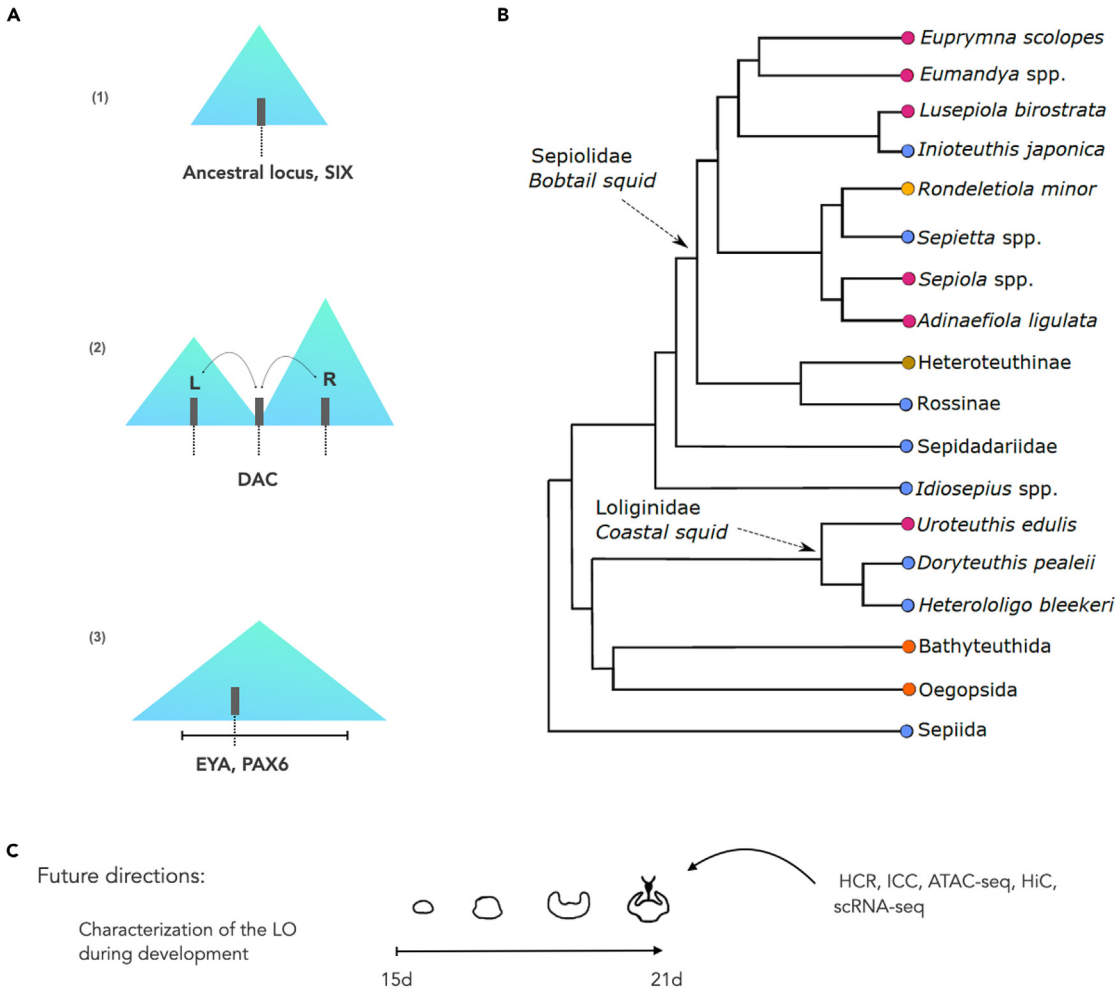



**Figure 5. Scenarios of genome topology changes associated with gene co-option for both LO and eye development, their testability across species and molecular approaches**

(A) The three main scenarios of genome topology around LO/eye genes, including (1) conservation of the original (molluscan) regulation (e.g., in {SIX}), (2) positioning at the TAD boundary (e.g, {DAC}), (3) or expansion of the non-coding regions within the same TAD (e.g., {EYA}, {PAX6}).

(B) A general cladogram showing the relationship between some of the main squid lineages or species and their ventral light organ (LO). Pink dots indicate species with symbiotic and bilobed LO; the yellow dot shows species possessing symbiotic and rounded LO; brown dot indicate the species present autogenic LO; blue dots indicate the species with no LO. Orange dots show lineages without LO or autogenic LO in locations other than the ventral side.

(C) Future directions and emerging methodology to study LO development.

responsible for either LO or eye development. Hox clusters are also genomically unique in vertebrates, compared to other locally conserved gene clusters, in that they also show significantly higher gene density than the average genomic gene density in those genomes.[79] Similar changes in gene densities are observed around the {DAC} cluster. However, as we do not yet have tissue-resolved ATAC-seq or RNA-seq during embryonic development, we cannot fully test this hypothesis. The initial investigation of the genes in the upstream or downstream TAD of the regions remains inconclusive (Figure S10).

In our study, as other genes show locations well within their respective TADs, it is also possible that a combination of different evolutionary modalities was involved, such as an expansion of the non-coding regions around the genes and evolution of novel enhancer elements. This latter scenario is likely for the case of EYA with at least 2 Mb space lacking any coding genes until the closest putatively co-regulated gene (Figure S3). To this end, other modes of regulatory innovations may have contributed to the LO emergence, e.g., through expansion of the regulatory regions and presence of extensive gene deserts/regulatory regions, as is the case for {EYA} and {PAX6} (Figure 5A).

## Toward an LO evolutionary genomic "profile"

These observations, both from an evolutionary and regulatory genomics perspective, allow us to formulate a possible genomic profile for the development and evolution of the LO. As such, the rearranged nature of coleoid cephalopod genomes sets the stage for the exploration of novel local gene linkages and their eventual co-regulation. As genes that were previously on separate chromosomes were brought together through fusion of chromosomal parts, the newly formed chromosomes began exploring novel combinations of co-regulation of neighboring genes. We thus propose that genome rearrangements have been a lasting signature of coleoid cephalopod genomes and they may have contributed to clade-specific innovations that appeared only later in coleoid evolution, including emergence of the LO.

The testable signature of LO emergence could comprise the following (Figure 5A): (1) localization of genes toward TAD boundaries, (2) co-regulation of genes both in a left and right TAD, and/or (3) expansion of regulatory vicinity with extra DNA that comprises "gene deserts" and harbors regulatory elements (as identified by open chromatin profiling). An alternative scenario for co-option may include genes with faster evolutionary rates or local (tandem) duplication of regulatory regions. In that case, however, the genes would still be expected to be found within the same original topological domain vicinity.

To test this hypothesis, we also would expect to find LO or eye specific regulatory regions in left or right TADs. Loss of function studies[80] could then target such regions and should result in ablation of the LO or eye phenotype or development.

Finally, several modes of evolution and resulting genomic "profiles" could complement each other in LO evolution: although we do not find signatures of gene duplication behind the proposed key transcription factors and, rather an expansion of their regulatory regions, genes expressed at later developmental stages may be more affected by gene duplication processes and sub-functionalization. Such gene families in both LO-specific and LO/eye gene lists, include crystallins, reflectins, and so on. The ongoing sequencing of multiple closely related bobtail squid species as a part of the Aquatic Symbiosis Genomics consortium[81] (https://www.aquaticsymbiosisgenomics.org/) (Figure 5B), will be crucial in testing the proposed evolutionary model as well as identifying currently undetectable conserved regulatory regions.

## Dissecting LO functionality

The most immediate research direction would be to investigate the expression, via HCR (*In situ* hybridization chain reaction,[82] Figure 5C), the new EYA, DAC, PAX6 and SIX co-regulated genes to validate their expression domain during development. Our study also has identified novel candidate genes specific to the LO, including the largely expanded omega crystallin gene family, reflectin, and halide peroxidase, that were expressed in the LO, with their paralogs expressed in the eyes. Furthermore, tissue-resolved ATAC-seq, HiC (e.g., micro-C), and similar approaches will be required to test for proposed topological and regulatory changes during LO development (Figure 5C).

We furthermore found several previously unreported genes in the LO transcriptomes. For example, we found that a gene with homology to SraP, so far only reported in *Staphylococcus aureus* association with platelet cells,[71] which is co-regulated with EYA during development and expressed in the eye, optic lobe, and central core in adults. The function of this gene in those tissues is largely unknown. Similarly, we found corin, reported to be involved in the cardiac morphogenesis in humans,[83,84] as a gene expressed only in the LO.

The discovery of the adult LO genes expressed at very late embryonic developmental stages, coinciding with the appearance of the final (third) crypt pair of the LO may suggest their key role in crypt 3 functionality. Indeed, crypt 3 is distinguished from crypt 1 and 2 in that it is smaller,[85] and has a different symbiont venting frequency.[86] Since the other two crypts appear earlier in embryonic development, they may be associated with genes with broader expression that have undergone similar evolutionary co-option as the four transcription factors involved in LO and eye development.

Taken together, our analyses provide a first glimpse into the regulatory and evolutionary context of key developmental transcription factors and functional genes that are involved in LO and eye development. The observed different modalities of regulatory evolution around these loci provide for several testable

predictions that will help in the understanding how novel organs evolve in bobtail squids and potentially other animal lineages.

### Limitations of the study

The insights into LO development and evolution rely on the known reported and co-opted transcription factors. Our study aimed to extend this finding by (1) assessing transcriptomic profiles of LO and eye genes using newly available data, (2) characterizing the regulatory interactions in the vicinity of these genes, and (3) their evolutionary emergence. Our data provides glimpses into possible co-regulated genes and their activity during development. A major obstacle is still the understanding of whether and how co-localized genes are co-regulated in cephalopods. Higher resolution HiC (micro-C) data will be required to reveal local regulatory interactions. Our study also provides a set of candidates that further can be tested via gene expression and gene knock-out studies.[80] Furthermore, availability of chromosomal-scale genomes[81] of other cephalopods and bobtail squids in particular, will become crucial in validating our hypothesis about the recent DAC locus emergence.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animal material
- METHOD DETAILS
  - RNA-extraction methods
  - RNA-extraction in embryos
  - Genomic and transcriptomic analysis
  - Genes annotations
  - Visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107091.

## ACKNOWLEDGMENTS

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 945026 to Oleg Simakov); competitive fellowship funds from the (Kuffler Research Awards, Spiegel Research Awards Fund, L. & A. Colwin Summer Research Fellowship) Fund of the Marine Biological Laboratory in Woods Hole, MA to MBL Whitman Fellowships; National Institutes of Health R37-AI050661 to Margaret McFall-Ngai and Edward Ruby and National Institutes of Health R01-GM135254 to Edward Ruby and Margaret McFall-Ngai; and NSF IOS-1557914 to SVN. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health. Computation was done using the Life Sciences Cluster at the University of Vienna. Sequencing was conducted at the Vienna Biocenter Core Facility.

## AUTHOR CONTRIBUTIONS

L.R., M.M.-N., and O.S. designed the study. L.R. conducted analyses. J.B., N.V., S.V.N., and E.A.R. conducted RNA isolation. G.S. and G.Y.C. conducted data evaluation. All authors contributed to manuscript writing.

## DECLARATION OF INTERESTS

All the authors declare no competing interests.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research. One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community.

Received: November 30, 2022
Revised: March 25, 2023
Accepted: June 7, 2023
Published: June 9, 2023

## REFERENCES

1. Tanner, A.R., Fuchs, D., Winkelmann, I.E., Gilbert, M.T.P., Pankey, M.S., Ribeiro, Â.M., Kocot, K.M., Halanych, K.M., Oakley, T.H., da Fonseca, R.R., et al. (2017). Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. Proc. Biol. Sci. *284*, 20162818.

2. Sanchez, G., Simakov, O., and S Rokhsar, D. (2022). Beyond "living fossils": can comparative genomics finally reveal novelty? Mol. Ecol. Resour. *22*, 9–11.

3. Shigeno, S., Andrews, P.L.R., Ponte, G., and Fiorito, G. (2018). Cephalopod brains: an overview of current knowledge to facilitate comparison with vertebrates. Front. Physiol. *9*, 952.

4. Ponte, G., Chiandetti, C., Edelman, D.B., Imperadore, P., Pieroni, E.M., and Fiorito, G. (2021). Cephalopod behavior: from neural plasticity to consciousness. Front. Syst. Neurosci. *15*, 787139.

5. Ritschard, E.A., Whitelaw, B., Albertin, C.B., Cooke, I.R., Strugnell, J.M., and Simakov, O. (2019). Coupled genomic evolutionary histories as signatures of organismal innovations in cephalopods: Co-evolutionary signatures across levels of genome organization may shed light on functional linkage and origin of cephalopod novelties. Bioessays *41*, e1900073.

6. Albertin, C.B., and Simakov, O. (2020). Cephalopod biology: at the intersection between genomic and organismal novelties. Annu. Rev. Anim. Biosci. *8*, 71–90.

7. Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., and Rokhsar, D.S. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. Nat. Commun. *13*, 2427.

8. Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O.P., Zimmermann, B., Ritschard, E.A., Weissenbacher, A., Foster, J.S., Nyholm, S.V., et al. (2022). Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat. Commun. *13*, 2172.

9. Haddock, S.H.D., Moline, M.A., and Case, J.F. (2010). Bioluminescence in the sea. Ann. Rev. Mar. Sci *2*, 443–493.

10. Sanchez, G., Fernández-Álvarez, F.Á., Taite, M., Sugimoto, C., Jolly, J., Simakov, O., Marlétaz, F., Allcock, L., and Rokhsar, D.S. (2021). Phylogenomics illuminates the evolution of bobtail and bottletail squid (order Sepiolida). Commun. Biol. *4*, 819.

11. Jones, B.W., and Nishiguchi, M.K. (2004). Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes berry* (Mollusca: cephalopoda). Mar. Biol. *144*, 1151–1155.

12. Nyholm, S.V., and McFall-Ngai, M.J. (2021). A lasting symbiosis: how the Hawaiian bobtail squid finds and keeps its bioluminescent bacterial partner. Nat. Rev. Microbiol. *19*, 666–679.

13. Visick, K.L., Stabb, E.V., and Ruby, E.G. (2021). A lasting symbiosis: how Vibrio fischeri finds a squid partner and persists within its natural host. Nat. Rev. Microbiol. *19*, 654–665.

14. Montgomery, M.K., and McFall-Ngai, M. (1993). Embryonic development of the light organ of the sepiolid squid *Euprymna scolopes berry*. Biol. Bull. *184*, 296–308.

15. Lee, P.N., Callaerts, P., and de Couet, H.G. (2009). The embryonic development of the Hawaiian bobtail squid (*Euprymna scolopes*). Cold Spring Harb. Protoc. *2009*, db.ip77.

16. McFall-Ngai, M., and Montgomery, M.K. (1990). The anatomy and morphology of the adult bacterial light organ of *Euprymna scolopes berry* (Cephalopoda:Sepiolidae). Biol. Bull. *179*, 332–339.

17. Belcaid, M., Casaburi, G., McAnulty, S.J., Schmidbaur, H., Suria, A.M., Moriano-Gutierrez, S., Pankey, M.S., Oakley, T.H., Kremer, N., Koch, E.J., et al. (2019). Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proc. Natl. Acad. Sci. USA *116*, 3030–3035.

18. Tong, D., Rozas, N.S., Oakley, T.H., Mitchell, J., Colley, N.J., and McFall-Ngai, M.J. (2009). Evidence for light perception in a bioluminescent organ. Proc. Natl. Acad. Sci. USA *106*, 9836–9841.

19. Donner, A.L., and Maas, R.L. (2004). Conservation and non-conservation of genetic pathways in eye specification. Int. J. Dev. Biol. *48*, 743–753.

20. Peyer, S.M., Pankey, M.S., Oakley, T.H., and McFall-Ngai, M.J. (2014). Eye-specification genes in the bacterial light organ of the bobtail squid *Euprymna scolopes*, and their expression in response to symbiont cues. Mech. Dev. *131*, 111–126.

21. Epstein, J.A., and Neel, B.G. (2003). An Eye on Organ Development (Nature Publishing Group UK). https://doi.org/10.1038/426238a.

22. Christophorou, N.A.D., Bailey, A.P., Hanson, S., and Streit, A. (2009). Activation of Six1 target genes is required for sensory placode formation. Dev. Biol. *336*, 327–336.

23. Li, X., Oghi, K.A., Zhang, J., Krones, A., Bush, K.T., Glass, C.K., Nigam, S.K., Aggarwal, A.K., Maas, R., Rose, D.W., and Rosenfeld, M.G. (2003). Eya protein phosphatase activity regulates Six1–Dach–Eya transcriptional effects in mammalian organogenesis. Nature *426*, 247–254. https://doi.org/10.1038/nature02083.

24. Abdelhak, S., Kalatzis, V., Heilig, R., Compain, S., Samson, D., Vincent, C., Weil, D., Cruaud, C., Sahly, I., Leibovici, M., et al. (1997). A human homologue of the Drosophila eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. Nat. Genet. *15*, 157–164.

25. Xu, P.X., Woo, I., Her, H., Beier, D.R., and Maas, R.L. (1997). Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. Development *124*, 219–231.

26. Xu, P.-X., Zheng, W., Huang, L., Maire, P., Laclef, C., and Silvius, D. (2003). Six1 is required for the early organogenesis of mammalian kidney. Development *130*, 3085–3094.

27. Chen, R., Amoui, M., Zhang, Z., and Mardon, G. (1997). Dachshund and eyes absent proteins form a complex and function synergistically to induce ectopic eye development in Drosophila. Cell *91*, 893–903.

28. Pignoni, F., Hu, B., Zavitz, K.H., Xiao, J., Garrity, P.A., and Zipursky, S.L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development. Cell *91*, 881–891.

29. Heanue, T.A., Reshef, R., Davis, R.J., Mardon, G., Oliver, G., Tomarev, S., Lassar, A.B., and Tabin, C.J. (1999). Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation. Genes Dev. *13*, 3231–3243.

30. Dong, Z., Yuwen, Y., Wang, Q., Chen, G., and Liu, D. (2012). Eight genes expression patterns during visual system regeneration in *Dugesia japonica*. Gene Expr. Patterns *12*, 1–6.

31. Martín-Durán, J.M., Monjo, F., and Romero, R. (2012). Morphological and molecular development of the eyes during embryogenesis of the freshwater planarian *Schmidtea polychroa*. Dev. Gene. Evol. *222*, 45–54.

32. Walther, C., and Gruss, P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. Development *113*, 1435–1449.

33. Püschel, A.W., Gruss, P., and Westerfield, M. (1992). Sequence and expression pattern of pax-6 are highly conserved between zebrafish and mice. Development *114*, 643–651.

34. Quiring, R., Walldorf, U., Kloter, U., and Gehring, W.J. (1994). Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. Science *265*, 785–789.

35. Tomarev, S.I., Callaerts, P., Kos, L., Zinovieva, R., Halder, G., Gehring, W., and Piatigorsky, J. (1997). Squid Pax-6 and eye development. Proc. Natl. Acad. Sci. USA *94*, 2421–2426.

36. Hartmann, B., Lee, P.N., Kang, Y.Y., Tomarev, S., de Couet, H.G., and Callaerts, P. (2003). Pax6 in the sepiolid squid *Euprymna scolopes*: evidence for a role in eye, sensory organ and brain development. Mech. Dev. *120*, 177–183.

37. Navet, S., Andouche, A., Baratte, S., and Bonnaud, L. (2009). Shh and Pax6 have unconventional expression patterns in embryonic morphogenesis in *Sepia officinalis* (Cephalopoda). Gene Expr. Patterns *9*, 461–467.

38. Imarazene, B., Andouche, A., Bassaglia, Y., Lopez, P.-J., and Bonnaud-Ponticelli, L. (2017). Eye development in *Sepia officinalis* embryo: what the uncommon gene expression profiles tell us about eye evolution. Front. Physiol. *8*, 613.

39. Cheyette, B.N., Green, P.J., Martin, K., Garren, H., Hartenstein, V., and Zipursky, S.L. (1994). The Drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system. Neuron *12*, 977–996.

40. Serikaku, M.A., and O'Tousa, J.E. (1994). Sine oculis is a homeobox gene required for Drosophila visual system development. Genetics *138*, 1137–1150.

41. Ohto, H., Takizawa, T., Saito, T., Kobayashi, M., Ikeda, K., and Kawakami, K. (1998). Tissue and developmental distribution of Six family gene products. Int. J. Dev. Biol. *42*, 141–148.

42. Lagutin, O.V., Zhu, C.C., Kobayashi, D., Topczewski, J., Shimamura, K., Puelles, L., Russell, H.R.C., McKinnon, P.J., Solnica-Krezel, L., and Oliver, G. (2003). Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. Genes Dev. *17*, 368–379.

43. Stierwald, M., Yanze, N., Bamert, R.P., Kammermeier, L., and Schmid, V. (2004). The Sine oculis/Six class family of homeobox genes in jellyfish with and without eyes: development and eye regeneration. Dev. Biol. *274*, 70–81.

44. Hammond, K.L., Hanson, I.M., Brown, A.G., Lettice, L.A., and Hill, R.E. (1998). Mammalian and Drosophila dachshund genes are related to the Ski proto-oncogene and are expressed in eye and limb. Mech. Dev. *74*, 121–131.

45. Caubit, X., Thangarajah, R., Theil, T., Wirth, J., Nothwang, H.G., Rüther, U., and Krauss, S. (1999). Mouse Dac, a novel nuclear factor with homology to Drosophila dachshund shows a dynamic expression in the neural crest, the eye, the neocortex, and the limb bud. Dev. Dynam. *214*, 66–80.

46. Davis, R.J., Shen, W., Heanue, T.A., and Mardon, G. (1999). Mouse Dach, a homologue of Drosophila dachshund, is expressed in the developing retina, brain and limbs. Dev. Gene. Evol. *209*, 526–536.

47. Mardon, G., Solomon, N.M., and Rubin, G.M. (1994). Dachshund encodes a nuclear protein required for normal eye and leg development in Drosophila. Development *120*, 3473–3486.

48. Loosli, F., Mardon, G., and Wittbrodt, J. (2002). Cloning and expression of medaka Dachshund. Mech. Dev. *112*, 203–206.

49. Tarazona, O.A., Lopez, D.H., Slota, L.A., and Cohn, M.J. (2019). Evolution of limb development in cephalopod mollusks. Elife *8*, e43828.

50. Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J.C., Suzuki, H., Daub, C.O., Hayashizaki, Y., and Lenhard, B. (2009). Transcriptional features of genomic regulatory blocks. Genome Biol. *10*, R38.

51. Gonzalez-Sandoval, A., and Gasser, S.M. (2016). On TADs and LADs: spatial control over gene expression. Trends Genet. *32*, 485–495.

52. Beagan, J.A., and Phillips-Cremins, J.E. (2020). On the existence and functionality of topologically associating domains. Nat. Genet. *52*, 8–16.

53. Crookes, W.J., Ding, L.-L., Huang, Q.L., Kimbell, J.R., Horwitz, J., and McFall-Ngai, M.J. (2004). Reflectins: the unusual proteins of squid reflective tissues. Science *303*, 235–238.

54. Hirose, Y., Chiba, K., Karasugi, T., Nakajima, M., Kawaguchi, Y., Mikami, Y., Furuichi, T., Mio, F., Miyake, A., Miyamoto, T., et al. (2008). A functional polymorphism in THBS2 that affects alternative splicing and MMP binding is associated with lumbar-disc herniation. Am. J. Hum. Genet. *82*, 1122–1129.

55. Alberati-Giani, D., Cesura, A.M., Broger, C., Warren, W.D., Röver, S., and Malherbe, P. (1997). Cloning and functional expression of human kynurenine 3-monooxygenase. FEBS Lett. *410*, 407–412.

56. Martin, G., Nemoto, M., Gelman, L., Geffroy, S., Najib, J., Fruchart, J.C., Roevens, P., de Martinville, B., Deeb, S., and Auwerx, J. (2000). The human fatty acid transport protein-1 (SLC27A1; FATP-1) cDNA and gene: organization, chromosomal localization, and expression. Genomics *66*, 296–304.

57. Wilson, P.A., Gardner, S.D., Lambie, N.M., Commans, S.A., and Crowther, D.J. (2006). Characterization of the human patatin-like phospholipase family. J. Lipid Res. *47*, 1940–1949.

58. Kiss, H., Kedra, D., Yang, Y., Kost-Alimova, M., Kiss, C., O'Brien, K.P., Fransson, I., Klein, G., Imreh, S., and Dumanski, J.P. (1999). A novel gene containing LIM domains (LIMD1) is located within the common eliminated region 1 (C3CER1) in 3p21.3. Hum. Genet. *105*, 552–559.

59. Weis, V.M., Small, A.L., and McFall-Ngai, M.J. (1996). A peroxidase related to the mammalian antimicrobial protein myeloperoxidase in the Euprymna–Vibrio mutualism. Proc. Natl. Acad. Sci. USA *93*, 13683–13688.

60. Heath-Heckman, E.A.C., Gillette, A.A., Augustin, R., Gillette, M.X., Goldman, W.E., and McFall-Ngai, M.J. (2014). Shaping the microenvironment: evidence for the influence of a host galaxin on symbiont acquisition and maintenance in the squid-Vibrio symbiosis. Environ. Microbiol. *16*, 3669–3682.

61. Meinl, W., and Glatt, H. (2001). Structure and localization of the human SULT1B1 gene: neighborhood to SULT1E1 and a SULT1D pseudogene. Biochem. Biophys. Res. Commun. *288*, 855–862.

62. Kiss, C., Li, J., Szeles, A., Gizatullin, R.Z., Kashuba, V.I., Lushnikova, T., Protopopov, A.I., Kelve, M., Kiss, H., Kholodnyuk, I.D., et al. (1997). Assignment of the ARHA and GPX1 genes to human chromosome bands 3p21.3 by in situ hybridization and with somatic cell hybrids. Cytogenet. Cell Genet. *79*, 228–230.

63. Weis, V.M., Montgomery, M.K., and McFall-Ngai, M.J. (1993). Enhanced production of ALDH-like protein in the bacterial light organ of the sepiolid squid *Euprymna scolopes*. Biol. Bull. *184*, 309–321.

64. Zinovieva, R.D., Tomarev, S.I., and Piatigorsky, J. (1993). Aldehyde dehydrogenase-derived omega-crystallins of squid and octopus. Specialization for lens expression. J. Biol. Chem. *268*, 11449–11455.

65. Montgomery, M.K., and McFall-Ngai, M.J. (1992). The muscle-derived lens of a squid bioluminescent organ is biochemically convergent with the ocular lens. Evidence for recruitment of aldehyde dehydrogenase as a predominant structural protein. J. Biol. Chem. *267*, 20999–21003.

66. Yan, W., Sheng, N., Seto, M., Morser, J., and Wu, Q. (1999). Corin, a mosaic transmembrane serine protease encoded by a novel cDNA from human heart. J. Biol. Chem. *274*, 14926–14935.

67. Yang-Feng, T.L., Zheng, K., Yu, J., Yang, B.Z., Chen, Y.T., and Kao, F.T. (1992). Assignment of the human glycogen debrancher gene to chromosome 1p21. Genomics *13*, 931–934.

68. Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marlétaz, F., Gómez-Marín, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.-M., et al. (2016). A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nat. Genet. *48*, 336–341.

69. Rodríguez-Carballo, E., Lopez-Delisle, L., Willemin, A., Beccari, L., Gitto, S., Mascrez, B., and Duboule, D. (2020). Chromatin topology and the timing of enhancer function at the HoxD locus. Proc. Natl. Acad. Sci. USA *117*, 31231–31241.

70. Williams, G.J., and Thorson, J.S. (2009). Natural product glycosyltransferases: properties and applications. Adv. Enzymol. Relat. Area Mol. Biol. *76*, 55–119.

71. Siboo, I.R., Chambers, H.F., and Sullam, P.M. (2005). Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. Infect. Immun. *73*, 2273–2280.

72. Lai, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S., and Lin, W. (2000). Identification of novel human genes evolutionarily conserved in *Caenorhabditis elegans* by comparative proteomics. Genome Res. *10*, 703–713.

73. Talbot, K., Miguel-Aliaga, I., Mohaghegh, P., Ponting, C.P., and Davies, K.E. (1998). Characterization of a gene encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome complex. Hum. Mol. Genet. *7*, 2149–2156.

74. Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., et al. (2007). New nomenclature for chromatin-modifying enzymes. Cell *131*, 633–636.

75. Kenny, N.J., McCarthy, S.A., Dudchenko, O., James, K., Betteridge, E., Corton, C., Dolucan, J., Mead, D., Oliver, K., Omer, A.D., et al. (2020). The gene-rich genome of the scallop *Pecten maximus*. GigaScience *9*, giaa037.

76. Nishiguchi, M.K., Lopez, J.E., and Boletzky, S.v. (2004). Enlightenment of old ideas from new investigations: more questions regarding the evolution of bacteriogenic light organs in squids. Evol. Dev. *6*, 41–49.

77. Duboule, D. (1998). Vertebrate hox gene regulation: clustering and/or colinearity? Curr. Opin. Genet. Dev. *8*, 514–518.

78. Acemel, R.D., Maeso, I., and Gómez-Skarmeta, J.L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. WIREs Dev. Biol. *6*. https://doi.org/10.1002/wdev.265.

79. Robert, N.S.M., Sarigol, F., Zimmermann, B., Meyer, A., Voolstra, C.R., and Simakov, O. (2022). Emergence of distinct syntenic density regimes is associated with early metazoan genomic transitions. BMC Genom. *23*, 143.

80. Crawford, K., Diaz Quiroz, J.F., Koenig, K.M., Ahuja, N., Albertin, C.B., and Rosenthal, J.J.C. (2020). Highly efficient knockout of a squid pigmentation gene. Curr. Biol. *30*, 3484–3490.e4.

81. McKenna, V., Archibald, J.M., Beinart, R., Dawson, M.N., Hentschel, U., Keeling, P.J., Lopez, J.V., Martín-Durán, J.M., Petersen, J.M., Sigwart, J.D., et al. (2021). The Aquatic Symbiosis Genomics Project: Probing the Evolution of Symbiosis across the Tree of Life (Wellcome Open Research).

82. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753.

83. Yan, W., Wu, F., Morser, J., and Wu, Q. (2000). Corin, a transmembrane cardiac serine protease, acts as a pro-atrial natriuretic peptide-converting enzyme. Proc. Natl. Acad. Sci. USA *97*, 8525–8529.

84. Moriano-Gutierrez, S., Koch, E.J., Bussan, H., Romano, K., Belcaid, M., Rey, F.E., Ruby, E.G., and McFall-Ngai, M.J. (2019). Critical symbiont signals drive both local and systemic changes in diel and developmental host gene expression. Proc. Natl. Acad. Sci. USA *116*, 7990–7999.

85. Sycuro, L.K., Ruby, E.G., and McFall-Ngai, M. (2006). Confocal microscopy of the light organ crypts in juvenile *Euprymna scolopes* reveals their morphological complexity and dynamic function in symbiosis. J. Morphol. *267*, 555–568.

86. Essock-Burns, T., Bongrand, C., Goldman, W.E., Ruby, E.G., and McFall-Ngai, M.J. (2020). Interactions of symbiotic partners drive the development of a complex biogeography in the squid-Vibrio symbiosis. mBio *11*, 008533-20–e920.

87. Team, R.C. (2013). R: A Language and Environment for Statistical Computing.

88. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

89. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

90. Kolde, R. (2012). Pheatmap: pretty heatmaps. R package version *1*, 726.

91. Akdemir, K.C., and Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. Genome Biol. *16*, 198.

92. UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49*, D480–D489.

93. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

94. Behera, A.K., Kumar, M., Lockey, R.F., and Mohapatra, S.S. (2002). 2′-5′ oligoadenylate synthetase plays a critical role in interferon-γ inhibition of respiratory syncytial virus infection of human epithelial cells. J. Biol. Chem. *277*, 25601–25608.

95. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Biological samples | | |
| *Euprymna scolopes* | Oahu, Hawaii | N/A |
| Deposited data | | |
| *E. scolopes* whole genome and transcriptome sequence data | (Belcaid et al., 2019)[17] | PRJNA470951 |
| *E. scolopes* adult transcriptomic data | This paper | PRJNA907048 |
| *E. scolopes* genome assembly, annotation report and Hi-C data | NCBI, (Schmidbaur et al., 2022)[8] | PRJNA661684 |
| *O. bimaculoides* reference genome | NCBI | PRJNA305125 |
| *D. pealeii* reference genome | NCBI | PRJNA641326 |
| *P. maximus* reference genome | NCBI | PRJEB35330 |
| Accession code | This paper | https://bitbucket.org/viemet/public/src/master/LO_topology |
| Software and algorithms | | |
| RStudio | RStudio 2021.09.2 Build 382 software © 2009–2022 RStudio, PBC, R version 4.2.1 (2022-06-23) (Team, 2013)[87] | https://posit.co/download/rstudio-desktop/ |
| STAR aligner | (Dobin et al., 2013)[88] | https://github.com/alexdo bin/STAR |
| FeatureCounts | (Liao et al., 2014)[89] | https://subread.sourcefor.ge.net/ |
| Pheatmap package | (Kolde et al., 2019)[90] | https://cran.r-project.org/web/packages/pheatmap/index.html |
| HiCPlotter biotool | (Akdemir & Chin, 2015)[91] | https://github.com/akdemi rlab/HiCPlotter |
| UniProt | (The UniProt Consortium, 2021)[92] | https://www.uniprot.org/blast |
| BLAST NCBI | (Altschul et al., 1990)[93] | https://blast.ncbi.nlm.nih.g.ov/Blast.cgi |
| Euprymna Genome Browser | N/A | http://metazoa.csb.univie.ac.at:8000/euprymna/query |
| Reagents | | |
| Microcentrifuge bead-beating tubes | Fisher Scientific | 02-682-558 |
| 0.1mm Zirconia/Silica beads | BioSpec | 11079101Z |
| RNeasy Mini kit | Qiagen | 74104 |
| TURBO DNA-*free*™ Kit | ThermoFisher Scientific | AM1907 |
| TRIzol™ Reagent | Invitrogen | 15596026 |

### RESOURCE AVAILABILITY

All the resources are available with this manuscript.

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead contact: Lisa Rouressol (lisa.rouressol@univie.ac.at) and by the corresponding authors: Lisa Rouressol (lisa.rouressol@univie.ac.at), Margaret McFall-Ngai (mcfallng@hawaii.edu), and Oleg Simakov (oleg.simakov@univie.ac.at).

### Materials availability

There are no restrictions to the availability of newly generated materials in this study.

### Data and code availability

- RNA seq used in the study has been published or deposited at NCBI (see STAR Methods) and will be publicly available as of the date of publication except for transcriptomic data under temporary embargo. Accession numbers are listed in the key resources table.
- All original code has been deposited in bitbucket: https://bitbucket.org/viemet/public/src/master/LO_topology and is publicly available as of the date of publication.
- All bioinformatic analyses were conducted with published protocols (see key resources table) and any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Animal material

All adult animal experiments were conducted in compliance with protocol number A18-029 approved by the Institutional Animal Care and Use Committee, University of Connecticut. Adult Hawaiian bobtail squid were collected from Maunalua Bay, Oahu, Hawaii (21°16′51.42″N, 157°43′33.07″W), and were transported to the University of Connecticut where they were maintained in recirculating artificial seawater. Animals were euthanized and tissues were sampled for RNA as described below. Generation of RNA data from embryonic stages is described in Schmidbaur et al., 2022 and data is deposited under PRJNA903148.

## METHOD DETAILS

### RNA-extraction methods

Animals were anesthetized in 2% ethanol and organs (central brain, subesophageal, brain lobes, optic lobes, white bodies, arms, tentacles, skin, mantle, ovaries, testes, hectocotylus) ANGs, central cores, and gills were dissected, and either placed in a cryotube and flash frozen in liquid nitrogen or submerged in 1 mL of TRIzol within a freestanding 2 mL bead-beating tube with 0.1mm Zirconia/Silica beads. The RNA from ANGs, central cores and gills were extracted with Qiagen Rneasy mini kit as per manufacturer's instructions. The remaining samples were processed for RNA extraction following the TRIzol manufacturer's protocol.[94] The samples were first homogenized with a Qiagen PowerLyzer at settings: 4500 rpm for 35 s. The final RNA pellet was washed three times with 75% ethanol at 4°C and resuspended in 30 μL of nuclease-free water. Samples were then treated with Ambion's Turbo DNA-free kit and quality controlled with an Agilent 4200 TapeStation Automated Electrophoresis system. Libraries were prepared with polyA selection and sequenced using NovaSeq S4 PE150 mode at Vienna Core Facility.

### RNA-extraction in embryos

Animals for RNA-seq were all collected from the same clutch. Five embryos were used per replicate and three replicates per stage were sampled. As much of the seawater as possible was removed, 50μL of TRIzol Reagent (Life Technologies, Carlsbad, California, United States) was added and embryos were flash frozen in liquid nitrogen and kept at −80°C until library preparation. Method described in Schmidbaur et al., 2022.[8]

### Genomic and transcriptomic analysis

The genome and transcriptome sequencing of *E. scolopes* has been deposited in the Sequence Read Archive under the Bioproject PRJNA470951, published in Belcaid et al., 2019.[17] New transcriptomic data from adult tissues has been deposited in Sequence Read Archive under the Bioproject (PRJNA907048); Briseno, Vijayan et al., pers. comm. The reads were mapped with STAR aligner[88] and counts were obtained with featureCounts tools from subread package[89] against the annotated chromosomal reference genome.[7,8] For chromatin conformation analysis, and HiC maps for *E. scolopes* we used the HiCPlotter bio-tool[91] and the Hi-C data are deposited under the Bioproject PRJNA661684 and published in Schmidbaur et al., 2022.[8] For genomic comparison to other species, for *P. maximus* we used the NCBI Annotation Release ID:100 of xPecMax1.1, under the Bioproject PRJEB35330. For *D. pealeii* we used the NCBI Annotation, under the Bioproject PRJNA641326. For *O. bimaculoides* we used the NCBI Annotation, under the

Bioproject PRJNA305125. For *E. scolopes*, *D. pealeii* and *O. sinensis*, we also used the *Euprymna* Genome Browser Blast Tool; *D. pealeii*: =http://metazoa.csb.univie.ac.at:4570/?seldb = dorPea; *O. sinensis*: http://metazoa.csb.univie.ac.at:4570/?seldb=Oct_sinensis. We used the *Euprymna* Genome Browser (http://metazoa.csb.univie.ac.at:8000/euprymna/query) to identify, find, the genes and localize their position in the genome.

### Genes annotations

For genes annotations, we used the NCBI's blast, Tblastn tool,[93] the UniProt website[92] and the EggNOG Database.[95]

### Visualization

The statistical analyzes and the production of the heatmap graphs were carried out using the RStudio 2021.09.2 Build 382 software © 2009–2022 RStudio, PBC, R version 4.2.1 (2022-06-23)[87] and the pheatmap package.[90]

## QUANTIFICATION AND STATISTICAL ANALYSIS

For the RNA data on embryological development, the samples were taken from eight time points representing development stages 14–28, with 3 replicates for each; for the final figures we used the average of the replicates. For RNA expression analysis in adult tissues, 4 biological replicates per organ were used. Tissue-specific expression in a given set of tissues were identified using cutoffs (TPM>20). Such cutoffs were previously employed in Belcaid et al., 2019[17] and used for comparison across datasets to obtain the initial list of candidate genes. For the heatmaps, a *Z* score normalization was performed on the normalized read counts across samples for each gene. Z-scores were computed on a gene-by-gene (row-by-row) basis by subtracting the mean and then dividing by the standard deviation.