PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Klimovich A, Bosch TCG. 2024 Novel technologies uncover novel 'anti'-microbial peptides in *Hydra* shaping the species-specific microbiome. *Phil. Trans. R. Soc. B* **379**: 20230058.
https://doi.org/10.1098/rstb.2023.0058

Received: 17 July 2023
Accepted: 16 November 2023

One contribution of 18 to a theme issue 'Sculpting the microbiome: how host factors determine and respond to microbial colonization'.

Subject Areas:
developmental biology, evolution, bioinformatics

Keywords:
taxonomically restricted genes, machine learning, scRNA-seq, holobiont

Authors for correspondence:
Alexander Klimovich
e-mail: aklimovich@zoologie.uni-kiel.de
Thomas C. G. Bosch
e-mail: tbosch@zoologie.uni-kiel.de

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7105351.

THE ROYAL SOCIETY PUBLISHING

Novel technologies uncover novel 'anti'-microbial peptides in *Hydra* shaping the species-specific microbiome

Alexander Klimovich and Thomas C. G. Bosch

Zoological Institute, Christian-Albrechts University of Kiel, Am Botanischen Garten 1-9, Kiel 24118, Germany

AK, 0000-0003-1764-0613; TCGB, 0000-0002-9488-5545

The freshwater polyp *Hydra* uses an elaborate innate immune machinery to maintain its specific microbiome. Major components of this toolkit are conserved Toll-like receptor (TLR)-mediated immune pathways and species-specific antimicrobial peptides (AMPs). Our study harnesses advanced technologies, such as high-throughput sequencing and machine learning, to uncover a high complexity of the *Hydra*'s AMPs repertoire. Functional analysis reveals that these AMPs are specific against diverse members of the *Hydra* microbiome and expressed in a spatially controlled pattern. Notably, in the outer epithelial layer, AMPs are produced mainly in the neurons. The neuron-derived AMPs are secreted directly into the glycocalyx, the habitat for symbiotic bacteria, and display high selectivity and spatial restriction of expression. In the endodermal layer, in contrast, endodermal epithelial cells produce an abundance of different AMPs including members of the arminin and hydramacin families, while gland cells secrete kazal-type protease inhibitors. Since the endodermal layer lines the gastric cavity devoid of symbiotic bacteria, we assume that endodermally secreted AMPs protect the gastric cavity from intruding pathogens. In conclusion, *Hydra* employs a complex set of AMPs expressed in distinct tissue layers and cell types to combat pathogens and to maintain a stable spatially organized microbiome.

This article is part of the theme issue 'Sculpting the microbiome: how host factors determine and respond to microbial colonization'.

1. Introduction: diversity and role of antimicrobial peptides in *Hydra*

Antimicrobial peptides (AMPs) are small cationic peptides that play a crucial role in the innate immune defence of a wide range of organisms from bacteria to humans [1,2]. These peptides exhibit broad-spectrum activity against various microorganisms, including bacteria, fungi, viruses and parasites.

The freshwater polyp *Hydra*, a member of the phylogenetically ancient phylum Cnidaria (figure 1*a–c*), has long been used as a model organism for the study of the immune response evolution [4–6]. Major components of the *Hydra* immune toolkit are highly conserved immune pathways mediated by Toll-like receptors (TLR) [5,7] and nucleotide-binding and oligomerisation domain-like receptors (NLR) [8]. They are complemented by a rich repertoire of immune effector molecules—secreted AMPs. While the first AMP in *Hydra* was discovered using traditional biochemical approaches [9], the advance of molecular biology techniques fueled the identification of multiple novel AMPs, such as the arminins, periculins, kazal-like inhibitors and the neuron-derived antimicrobial peptide NDA-1 [10–13]. AMPs in *Hydra* share several common features: active AMPs are derived from larger precursors through a post-translational proteolytic cleavage of a signal peptide (figure 1*d*). Most

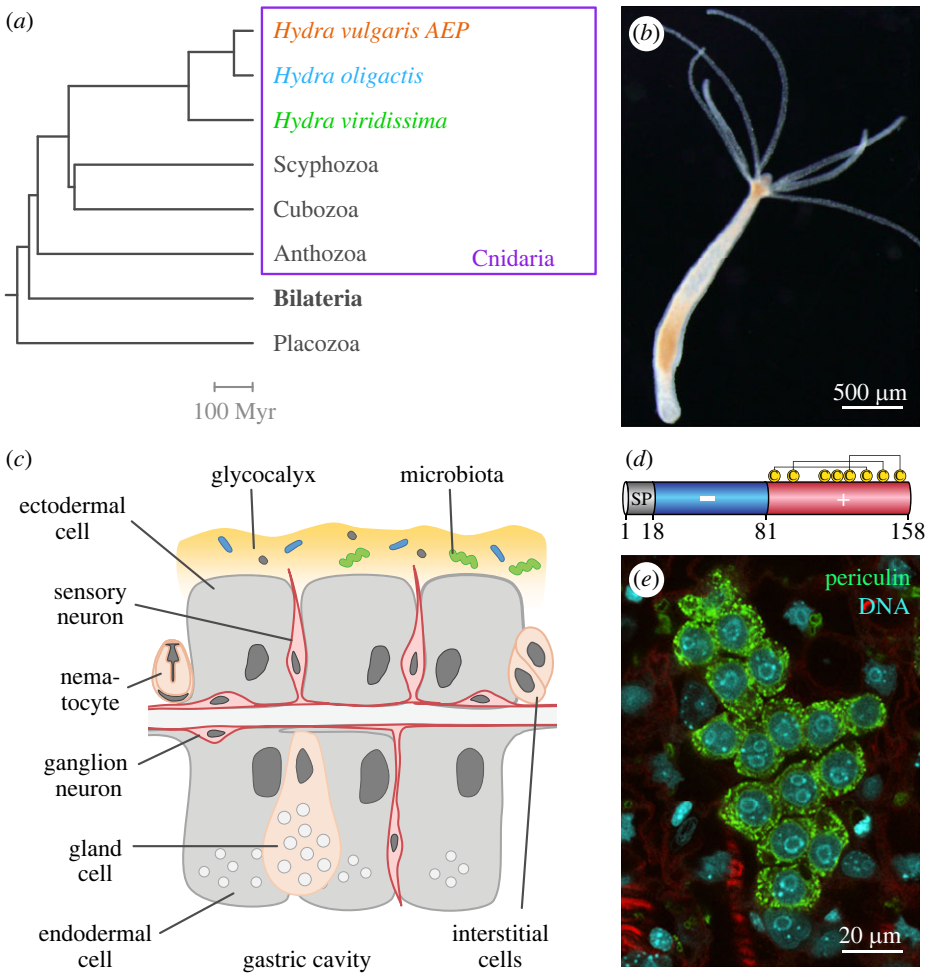


Figure 1. (*a*) Phylogenetic tree demonstrating the position of *Hydra*. High-quality genome datasets have become recently available for three *Hydra* species—*H. vulgaris* AEP, *H. oligactis* and *H. viridissima*. The divergence of the crown group *Hydra* took place about 193 Ma, and two species of brown hydras, *H. vulgaris* and *H. oligactis*, diverged over 100 Ma [3]. (*b*) A polyp of *H. vulgaris* AEP strain. It is composed of a tube-shaped body column, a basal disc attaching to a substratum, and an oral end with a hypostome and ring of tentacles. (*c*) The *Hydra* body is composed of the ectodermal and endodermal epithelial layers separated by the extracellular matrix. The outer surface of the ectoderm is covered by a glycocalyx that serves as a habitat for symbiotic bacteria. The endoderm lining the gastric cavity is free of glycocalyx and stable microbiota. Cells of the interstitial lineage, including the stem cells, nematocytes, gland cells and the neurons, are embedded within both epithelia. (*d*) Hydra-restricted periculin protein demonstrates key features of *Hydra* AMPs—small size, presence of a signal peptide (SP), bi-partite charge distribution and complex pattern of Cys-bridges. (*e*) Periculin is specifically expressed in the female gamete precursor cells of *Hydra*. Immunochemical detection of periculin 1a, DNA stained with TO-PRO3.

AMPs are characterized by a clear bipartite structure with a strongly biased distribution of positively and negatively charged amino acids, and a complex cysteine pattern. Another notable property of *Hydra* AMPs is that they are typically encoded by a number of paralogous genes, hence they represent distinct gene families. Importantly, the phylogenetic analysis of AMP genes in *Hydra* uncovered that no homologues of these genes can be found in other animals, outside of the *Hydra* genus. Therefore, most AMPs of *Hydra* appear to be species-specific and, hence, represent so called taxonomically restricted genes (TRGs) or orphans [14]. This suggests that the taxonomically restricted AMPs have evolved relatively recently in the evolution of *Hydra* and specifically in response to the unique challenges faced by this animal.

Studies on the *Hydra* AMPs function provided evidence that mature secreted peptides possess a specific and often remarkably strong antibacterial activity, and are able to effectively inhibit growth of gram-positive and -negative bacteria *in vitro* [9,11,13,15,16]. These observations led to a hypothesis that AMPs protect the *Hydra* from foreign microbes. Later, it was recognized that, *in vivo*, they are equally important for maintaining the diversity of the species-specific bacterial community stably associated with *Hydra*, the *Hydra* microbiome [1,17]. This has been convincingly demonstrated in experiments where genetic knock-down of individual AMP genes or their families resulted in profound changes in the *Hydra* microbiome composition [11,12,18].

Thus far, AMP genes and their products have been identified and functionally characterized individually, and no systematic study attempted to integrate the findings on the entire suite of AMPs present in each *Hydra* species. To understand the evolutionary dynamics of *Hydra*-specific AMPs and their functional role in maintaining microbiome homeostasis, a comprehensive, whole-genome-scale survey of the AMPs repertoire and their expression in *Hydra* is needed.

Here, we demonstrate how novel technologies, including high-throughput transcriptome and genome sequencing and machine learning, provide insights into a high complexity of the *Hydra*'s AMP repertoire. Further, we uncover a shared feature of AMP genes genomic organization and common principles that govern the tissue and cell type-specific expression of these genes. Furthermore, we explore the evolutionary significance of

these genes and their role in sculpturing the *Hydra*-specific microbiome. Finally, we outline a few open question and perspectives for further research on this enigmatic group of genes.

2. Insights from genomes: AMPs are encoded by fast evolving genes

The first AMPs discovered in *Hydra*, hydramacin and hydralysin, were initially identified through biochemical purification from *Hydra* tissue extracts [9,19]. Recent advancements in molecular biology techniques, such as expressed sequence tag analysis (EST) [9,10] and high-throughput transcriptome sequencing (RNAseq) [13,15], have greatly facilitated the systematic discovery of novel AMPs in *Hydra*. The utilization of these technologies has greatly expanded our understanding of the diversity and complexity of AMP families in *Hydra*. However, it remained unclear how complete was the repertoire of AMPs in each *Hydra* species, and whether all members of AMP families have been discovered. Recently, high-quality genome sequences became publicly available (see figure 1*a*) for two species of the 'brown hydra' phylogenetic group (*Hydra vulgaris* AEP and *Hydra oligactis*) [20], and one green hydra species (*Hydra viridissima*) [21], hence providing a glimpse into 200 Myr of evolutionary radiation within the *Hydra* crown group [3]. Additionally, a number of high-quality genomes of other hydrozoan cnidarians, scyphozoans and anthozoans became available [22–26]. Together, these resources allow accurate analysis of AMP genes and may provide novel insights into the role of AMPs in the biology of *Hydra*.

To uncover the complete repertoire of AMPs in *Hydra*, we first identified all paralogues of known AMP gene families in the genome of *Hydra vulgaris* strain AEP [20] (see electronic supplementary material text for details; electronic supplementary material, data). This strain is of particular interest, since it is the only one where functional gene manipulation by transgenesis is available [27,28]. In the *H. vulgaris* AEP genome, we discovered, to our surprise, a very high number of paralogues within each AMP family, often substantially higher than previously reported. For instance, we were able to identify at least 28 paralogues of *periculin* family genes (figure 2*a*,*b*; electronic supplementary material, table S1), in contrast to previously reported five *periculin* isoforms [13]. Although the nucleotide sequences of these 28 paralogues were clearly different (electronic supplementary material, figure S1), all these genes demonstrated similar exon-intron structure (figure 2*a*), and the amino acid sequences of peptides encoded by these genes were very remarkably similar (electronic supplementary material, figure S2). Most intriguingly, numerous *periculin* paralogues were found clustered in a few genomic loci (figure 2*a*). For instance, in *H. vulgaris* AEP, two clusters on chromosome 10 contained 14 and 9 *periculin* paralogues, and the other five paralogues were scattered among three other chromosomes. A very similar pattern was observed for other AMP families. We identified a total of nine *arminin* paralogues, seven genes of the Kazal-like family, and five genes encoding Hym357-like neuropeptides with antimicrobial activity (figure 2*d–f*; electronic supplementary material, table S1).

Taken together, these observations point to a substantial expansion of AMP gene families in *H. vulgaris* AEP. The genomic organization of the AMP gene clusters suggests that, during evolution, the peptide families were formed through several rounds of tandem gene duplications. This vast gene expansion appears particularly surprising given the relatively recent origin of the founder genes: for instance, *periculin* and *arminin* genes are strictly confined to the genus *Hydra* and, hence, their origin cannot date back longer that 200 Ma, and the duplication might have occurred much more recently. The mechanisms that may have contributed to the rapid evolution of the AMP gene complement in the recent history of the *Hydra* genus remain poorly understood.

To explore further the phylogenetic history of the duplicated AMP genes, we used available high-quality genomes of other *Hydra* species, as well as other cnidarians (see electronic supplementary material, text). This analysis of orthologues yielded three essential observations. First, the general trend of the presence of multiple paralogues has been confirmed. For instance, similar to *H. vulgaris* AEP, the genome of *H. oligactis* contained 21 paralogues of *periculin* family genes and 12 *arminin* orthologues (figure 2*b*,*e*; electronic supplementary material, figures S3 and S4; electronic supplementary material, table S1). These numerous paralogues of AMP genes were also clustered on the chromosomes of *H. oligactis* and *H. viridissima*, like in the *H. vulgaris* AEP genome (this is reflected in close numbers of the gene models from all three species; figure 2*b*,*e*; electronic supplementary material, table S1).

Second, the phylogenetic reconstruction uncovered that, in every *Hydra* species, AMP paralogues from each *Hydra* species tend to cluster together, forming species-specific clades (figure 2*b*,*e*; electronic supplementary material, figures S3 and S4). Typically, AMP genes from one species code for very similar or virtually identical proteins, distinct from AMPs from other species. For instance, 23 *periculin* paralogues in *H. vulgaris* AEP represent a solid cluster on the phylogenetic tree (figure 2*b*; electronic supplementary material, figure S3), and most likely have emerged from one ancestral sequence within *H. vulgaris* AEP. A set of 18 *periculin* paralogues in *H. oligactis* was formed independently (figure 2*b*; electronic supplementary material, figure S3).

This clear affinity of paralogues strongly supports their origination through a repeated and recent gene duplication within each *Hydra* species. In addition, in every *Hydra* species we uncovered individual representatives of AMP gene families that were clustering separately from closely related paralogues (figure 2*b*,*e*). These sequences represent, most likely, the ancestral, founder members of AMP families. Taken together, our cross-species analysis suggests that the ancestral state of the AMP gene complement was in fact very small, composed of three *periculin* and two *arminin* paralogues (figure 2*b*,*e*). These gene families underwent a major expansion later, upon radiation of *Hydra* species.

To our surprise, we were not able to uncover any *hydramacin* orthologues in *H. viridissima* using BLAST and hidden Markov model (HMM)-based searches (see electronic supplementary material, text), although two orthologues were confidently detected in the *H. oligactis* genome. Moreover, the synteny analysis (see electronic supplementary material, text) identified only a non-coding sequence in the syntenic *H. viridissima* chromosomal region where the *hydramacin* orthologue would be anticipated (electronic supplementary material, figure S5). This suggest that the ancestral arsenal of AMPs in the last common ancestor of green and brown hydras was very limited and did not include any hydramacin

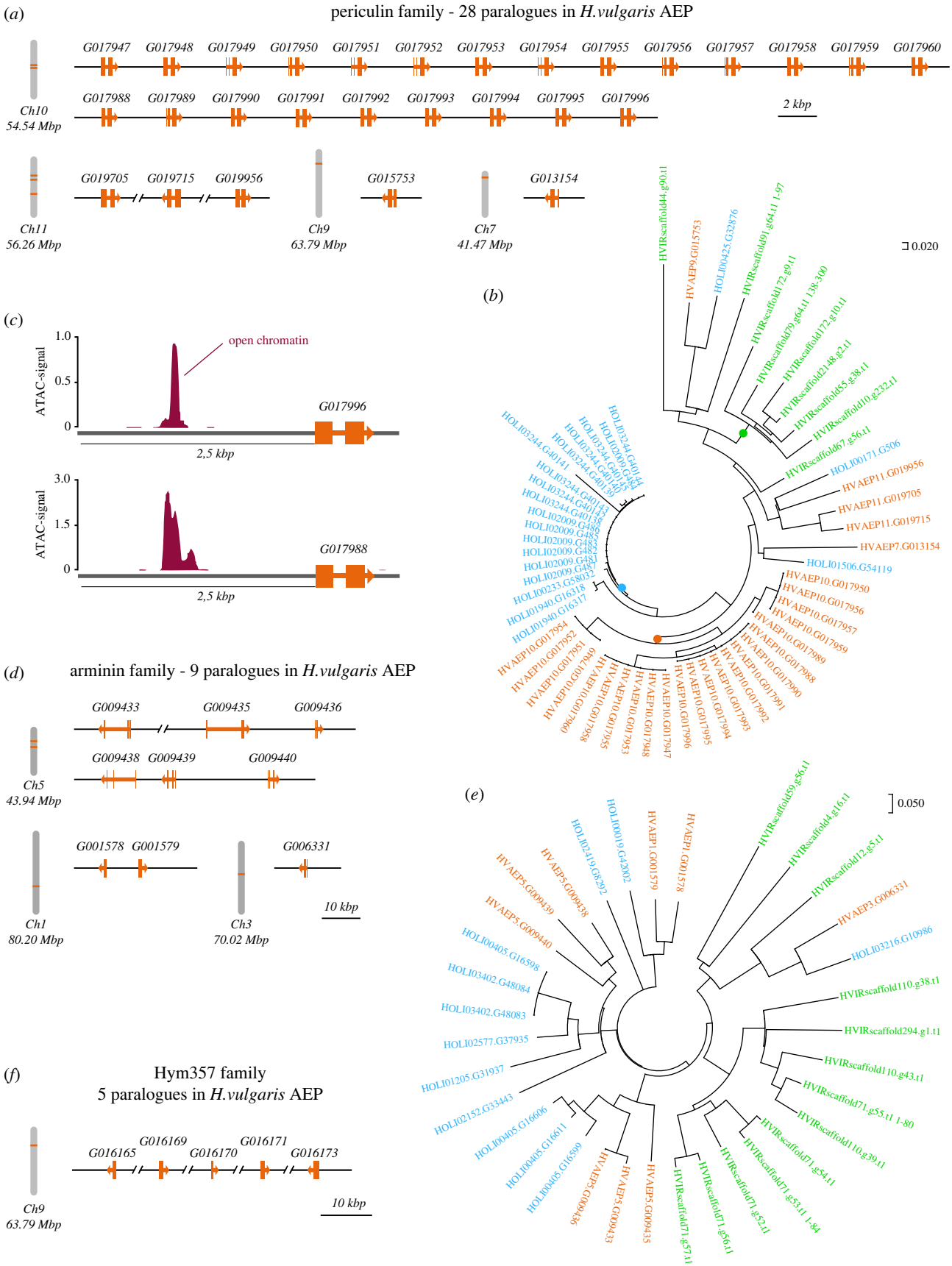


Figure 2. Complexity of AMP gene families in *Hydra* species. (*a*) Twenty-eight paralogues of the *periculin* AMP family are located on four chromosomes of *H. vulgaris* AEP, whereby 23 genes form one dense cluster on chromosome 10. (*b*) Phylogenetic analysis of *periculin* orthologues from three *Hydra* species. Genes are coloured according to species and numbers correspond to gene models: *H. vulgaris* AEP (HVAEP; orange), *H. oligactis* (HOLI; blue) and *H. viridissima* (HVIR; green). Not compressed bootstrapped tree is shown in electronic supplementary material, figure S3. (*c*) Chromatin accessibility analysis using ATAC-seq approach uncovers open chromatin regions within 2.5 kbp upstream from poorly expressed *periculin* genes, suggesting that these genes are not pseudogenes. Visualization based on data from Cazet *et al.* [20]. (*d*) Most *arminin* family paralogues are also present in one genomic locus on chromosome 5 in *H. vulgaris* AEP. (*e*) Phylogenetic analysis of *arminin* orthologues from three *Hydra* species. Not compressed bootstrapped tree is shown in electronic supplementary material, figure S4. (*f*) All five paralogues of Hym357 genes are found in one genomic cluster on chromosome 9 of *H. vulgaris* AEP. A complete list of accession numbers is presented in electronic supplementary material, table S1.

peptides, which evolved later, after the radiation of the crown *Hydra* group.

Finally, our screening for putative orthologues of AMP genes in the genomes of other cnidarians revealed no homologues even in closely related hydrozoans—*Hydractinia* and *Clytia*. These observations support the notion that AMPs are truly lineage-restricted genes confined to the *Hydra* genus. They evidently emerged about 200 Ma and diverged further following the radiation of *Hydra* species. The absence of any orthologues with at least partial similarity in animals outside of *Hydra* genus strongly suggests that the ancestral AMP genes have emerged *de novo* [29] from a non-coding sequence through one of multiple gene birth mechanism [30]. Although the origin of the founder AMP genes and the mechanisms of their further expansion in the *Hydra* lineage represent a substantial interest, they are beyond the scope of this study.

Similar to *Hydra*, the repertoire of AMPs in other animals and plants is dominated by lineage-specific genes. For instance, the cathelicidin peptide family is restricted to vertebrates [31–33], and diptericins are peptides confined to Diptera [34]. However, one AMP family, the defensins [35], appears to be omnipresent in the animal kingdom, in plants and fungi. Numerous *defensin* genes were *in silico* predicted from the genomes of Cnidaria as well [36–38] and few of them were empirically validated [39]. However, no members of the defensin family have been described in *Hydra* so far. We attempted to mine the genomes of three *Hydra* species for genes encoding putative defensins using BLAST and HMM-based approaches (see electronic supplementary material, text). To our surprise, we were not able to identify any genes in *Hydra* genomes coding for peptides with attributes of canonical defensin family members—mammalian defensins, arthropod defensins or protostome big defensins (electronic supplementary material, figure S6). Given that *defensin* orthologues are present in other cnidarians, placozoans and sponges, the most parsimonious explanations would be that the ancestral *defensin* genes were either lost in the *Hydra* lineage or evolved beyond recognition. We note that the *Hydra*-specific AMP hydramacin, in fact, shares some similarity with defensins (including the presence of six cystein residues), as previously suggested [9]. It is thus possible that hydramacin represents a far derived version of an ancestral defensin AMP.

Although we were not able to detect any *bona fide* defensins encoded in the *Hydra* genomes, our analysis uncovered a family of genes encoding secreted cysteine-rich peptides with partial similarity to defensins. Similar to defensins, these peptides possess six Cys residues, likely linked into three disulphide bonds, yet the spacing between these residues is clearly different from that characteristic for defensins (electronic supplementary material, figure S6, S7*a*). Additionally, these peptides are rich in tryptophan, and hence, we refer to them as *Hydra* cysteine/tryptophan-rich peptides, the HyCWR peptides. Intriguingly, the predicted HyCWR peptides demonstrated a strongly biased charge distribution, with the C-terminal portion being strongly positively charged, however no conventional cleavage site was found to separate these two portions (electronic supplementary material, figure S7*a*). We also note that the HyCWR genes represent a family of related genes, which comprises at least five orthologues in *H. vulgaris* AEP, seven in *H. oligactis* and one in *H. viridissima* (electronic supplementary material, figure S7*a*; electronic supplementary material, table S1), whereby several paralogues are typically located in the same genomic locus. Therefore, the HyCWR peptides in their structure and the genomic architecture of their genes follow similar trends described for AMPs in *Hydra*. However, we emphasize that it remains unclear whether the HyCWR peptides indeed display antimicrobial activity *in vitro* and *in vivo*. It is plausible that, in the absence of *bona fide* defensins, the non-related yet structurally similar HyCWR peptides take over their function. Taken together, a genome-wide mining for AMP sequences and cross-species comparison of AMP genes reveal a high complexity of AMP families in *Hydra* and suggest a complex gene family evolution within the *Hydra* genus.

3. Insights from scRNAseq—AMP genes are selectively expressed in certain cell types

Previous findings uncovered that most AMP genes are constitutively transcribed at a very high level. For instance, *arminin* mRNAs were reported to be more abundant that *β-actin* transcripts [12]. Similarly, *periculin* transcripts were among the most abundant transcripts in female polyps [13,40]. Additionally, AMP genes were reported to be expressed in certain tissue layers of *Hydra*. Most *arminin* paralogues, for instance, were expressed exclusively in the endodermal epithelial layer [12], while *periculin* transcripts were rather restricted to the female germline precursor cells within the interstitial cell lineage [13,40] (figure 1*e*). More recently, several AMP genes with neuron-restricted expression were discovered [11,16], but a comprehensive overview of the AMP cell-type specific expression pattern is still missing. Whole-genome expression atlases with single-cell resolution, which recently became available [16,41], uncovered a high diversity of cell types in *Hydra*. For instance, five types of ectodermal epithelial cells with specific transcriptional profiles and localization in the body were identified using scRNA sequencing. Even more surprisingly, up to 11 distinct spatially restricted neuronal cell types have been characterized [16,41,42]. Given this diversity of cell types, whole-genome expression atlases may provide a more comprehensive understanding of AMPs expression pattern and valuable insights into their function.

Our mapping of AMP genes expression using the scRNA-seq atlas of *H. vulgaris* AEP [20,41] fully corroborated and expanded earlier observations (figure 3). Indeed, the *hydramacin*, all *arminin* and most *kazal-like* genes are expressed exclusively in the endodermal epithelial cells (figure 3; electronic supplementary material, figure S8). Moreover, several other *kazal-like* transcripts are expressed in the gland cells, also located strictly in the endodermal layer. The ectodermal epithelial cells, on the contrary, were generally devoid of any AMP gene transcripts (figure 3). Our preliminary observations suggest that the genes encoding HyCWR peptides might be the only group of AMPs expressed in the ectodermal cells (electronic supplementary material, figure S7*b*,*c*). The cells of the interstitial lineage localized in the ectodermal layer (figure 1*c*), however, do express a variety of AMP genes. First, female germline precursor cells produce transcripts encoding the hydralysin, several periculin and Kazal-like peptides (figure 3). Neurons localized in the ectodermal and endodermal layers (figure 1*c*) express distinct sets of

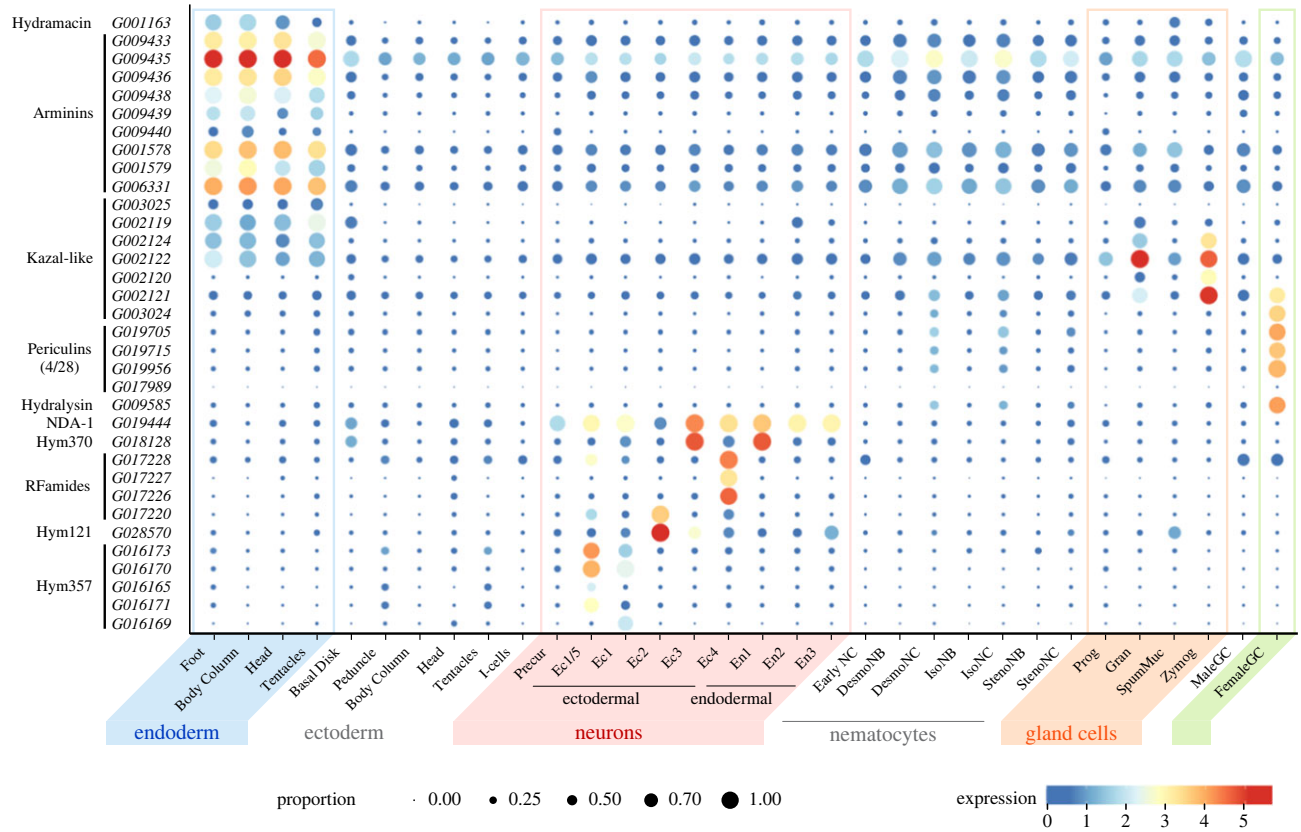


Figure 3. Expression of genes coding for AMPs across all cell types of *Hydra*. Visible are only the genes constitutively expressed in a *H. vulgaris* AEP polyp in homeostatic conditions, while inducible AMP genes expression is not illustrated. Note that only four out of 28 *periculin* paralogues are expressed at detectable levels. Visualization is based on data from Cazet *et al.* [20], gene expression is normalized and log-transformed.

dual function peptides, such as Hym370, Hym176, RFamides [11] and Hym121 [16]. Only one of these neuron-specific AMPs, NDA-1, is produced by both ectodermal and endodermal neurons.

The scRNA-seq data also provided insights into the spatial expression of AMP genes along the body axis of *Hydra*. AMP genes expressed in the endodermal epithelial cells do not show any expression bias and their transcripts are equally abundant in the polyp's foot, body column, head and tentacles (figure 3). Ectodermally expressed genes coding for putative HyCWR peptides show more distinct expression patterns, whereby one of them (G021955; electronic supplementary material, figure S7*b*,*c*), for instance, is strongly expressed in the basal disc, while another paralogue is not expressed in the foot at all (G0114589, electronic supplementary material, figure S7*b*,*c*). Expression of Kazal-like AMPs is confined to the upper body column, since zymogen and granular gland cells are abundant in the upper gastric region, but virtually absent from the polyp's foot and tentacles [41,43]. Intriguingly, since most of the neuronal populations are spatially restricted [16,41], the expression of neuron-derived AMP genes is also confined to a particular body compartment of *Hydra*. For instance, two RFamide precursor genes are expressed only in the hypostome and tentacles (population Ec4, figure 3), while Hym121 precursor is strictly present in the tentacles (neuronal population Ec2). Therefore, in each part of a polyp, a complex cocktail of AMPs is produced collectively by a variety of cell types.

The scRNA-seq datasets along with *in situ* hybridizations provide valuable insights into the expression of AMP genes on mRNA level. However, the localization of mature peptides translated from these mRNAs remains poorly investigated. Owing to the availability of specific antibodies, the localization of periculin peptides has been studied in most detail [13,40]. Mature periculins are produced in the female germline cells (figure 1*e*) located in the polyps ectoderm, are secreted and found on the outer surface of the epithelium. Even more intriguingly, periculins are also accumulated in the vesicles within the nurse cells, incorporated into an oocyte and released onto the embryo surface beneath the cuticle layer at early gastrulation stages [13]. Additionally, a fusion protein periculin-GFP expressed in the ectodermal epithelial cells recapitulates the vesicular accumulation and release of the peptide on the surface [13]. Similarly, with the help of specific antibodies, deposition of the neuronally expressed peptide NDA-1 into the glycocalyx of *Hydra* has been also demonstrated [11]. Therefore, the glycocalyx appears impregnated with diverse AMPs. Further proteome studies using mass spectrometry approaches, and particularly, spatial proteomics [44], should provide a more comprehensive view of the AMP localization in diverse cells, tissues and body compartments of *Hydra*.

Another evident observation emerging from the scRNA-seq data is that a substantial fraction of AMP genes is actually not expressed in homeostatic conditions. For instance, 24 out of 28 *periculin* paralogues have no evidence for transcription in the scRNAseq atlas, while several other AMP genes demonstrate a barely detectable expression in a small proportion of cells (figure 3). This is consistent with earlier observations of Franzenburg and co-authors [12], who reported expression of some *arminin* paralogues to be below detection level of microarray hybridization.

A plausible explanation for this observation might be that numerous paralogues of AMP genes actually represent

pseudogenes. However, several lines of evidence speak against this assumption. First, all paralogues, including the non-expressed ones, display features of protein-coding genes, such as an open-reading frame with a defined transcription start site, a start and a stop codon. Second, the paralogues show very similar exon-intron structure (particularly evident in case of *periculins*, figure 2*a*). Third, the sequences of these paralogs do not overlap with coding sequences of other genes. Finally, ATAC-seq profiling of the accessible chromatin states [20] identifies distinct peaks about 2.5 kbp upstream from the coding sequences of AMP genes, even the ones that have no expression evidence (figure 2*f*). Such a pattern of ATAC-signal is characteristic for most *Hydra* promoters [20] and suggests that *cis*-regulatory elements upstream from the AMP genes are located in the open chromatin and are accessible for binding of transcription factors. In homeostatic conditions, although the genes appear silent, their promoters are primed, and the transcription of AMP genes can be effectively activated upon specific stimulus. Taken together, cumulative evidence clearly indicates that numerous poorly expressed *periculin* and *arminin* paralogues are true protein-coding genes, whose expression is silenced in homeostatic conditions (see also §6c).

4. Advances in artificial intelligence: AMPs can be predicted *ab initio*

Until recently, identification of AMPs in diverse organisms relied mainly on homology-based screenings using known peptides as a 'bait'. Numerous AMP databases, such as the APD3, DBAASP, GRAMPA and InverPep [45–48], contain thousands of identified AMPs from animals and fungi, plants and microorganisms, provide a rich source of reference peptides for similarity searches and offer diverse build-in tools to perform such screenings. However, the homology-based approach has clear limitations, particularly given that, across the animal tree, AMPs are typically encoded by species-restricted genes [49]. Recent advances in artificial intelligence (AI), including the deep- and machine learning algorithms, provide a new opportunity of systematic *ab initio* discovery of novel AMPs [50–56]. Similar to BLAST-based homology searches, AI tools are dependent on rich datasets of known AMPs. However, in contrast to other approaches, AI predictive tools do not rely specifically on the amino acid sequence of AMPs. Instead, they identify essential physicochemical determinants of AMP functionality in the known AMPs present in the training dataset (so-called structure–function relations, which often are much more complex than simply a presence of a given amino acid in a certain position) and screen the target dataset to uncover proteins with similar structure–function correlations and rank them by likelihood of being *bona fide* AMPs. We have previously used one of these machine-learning algorithms (MLA) [57] to identify putative transcripts encoding α-helical AMPs among genes specifically expressed in *Hydra* neurons [16]. This approach turned out to be very effective and resulted in identification of dozens of putative neuronally expressed secreted AMPs encoded in Cnidaria-specific TRGs (figure 4*a*). These hitherto uncharacterized peptides, such as the product of a TRG *cluster131995* (figure 4*b*) demonstrate a very distinct pattern of charge and secondary structure distribution as well as strong predicted membrane activity. One of these genes, a *Hydra*-specific TRG *cluster62692*, was predicted to encode a precursor of a secreted short peptide with strong antimicrobial activity. Our minimal growth inhibitory concentration (MIC) assays confirmed that the active peptide Hym121 encoded within *cluster62692* was indeed a neuron-derived AMP highly potent against gram-positive and negative bacteria [16]. Hence, our functional analysis confirmed the accuracy of the MLA prediction. Intriguingly, a similar approach and the same MLA were used to identify a novel antimicrobial factor PACAP in the mammalian brain [59]. This dual-function neuropeptide known to regulate neurodevelopment, emotion and stress responses has been recently demonstrated to function as an AMP. Together, these observations demonstrate the power of AI tools in discovering novel functionally relevant AMPs. They also provide additional evidence, from the evolutionary perspective, for the structural similarity and functional reciprocity of AMPs and neuropeptides [60–62].

In our previous study, we focused on discovering putative AMPs exclusively expressed in neurons of *H. vulgaris* AEP. A high computational demand of the MLA precluded us from a deeper and more extensive analysis of AMP coding genes in *Hydra*. Nowadays, with the complete genomes for several *Hydra* species available and dramatically increased computational power, a whole-genome survey of AMPs encoded in *Hydra* genomes is feasible. It will be instrumental in uncovering novel, previously uncharacterized and very likely species-restricted AMP.

5. Lessons from the *Hydra* holobiont

(a) Expansion of AMP families in the phylogenetically younger *Hydra* species

As any other animal, each *Hydra* species forms a stable association with a specific multispecies bacterial community and hence functions as a metaorganism [12,63]. Understanding the mechanisms and molecular interactions involved in long-term maintenance of the metaorganism homeostasis remains a major challenge. Since AMPs are key factors regulating bacterial colonization, it is imperative to consider our findings on the AMP complexity in *Hydra* in the holobiont framework.

First, our observations clearly indicate that the majority of *Hydra* AMPs are encoded in *Hydra*-restricted genes. The forces that propelled the emergence of these TRGs at the root *Hydra* (figure 5*a*) radiation about 190 Ma [3] remain unclear. It is, however, plausible that the transition of a *Hydra* ancestor from the marine into the freshwater habitat has exposed the host to a totally new microbial environment. Additionally, in the new freshwater, low ion-strength environment, some ancient AMPs might become inefficient (e.g. defensins are generally known to be highly effective in a saltwater environment and tend to expand in the context of marine habitats [64]). Together, these factors might have fuelled an elaboration of a new molecular language for communication between the host and the microbes.

Our genome-wide survey of AMP-encoding genes in *Hydra* uncovered a high complexity of lineage-restricted AMP families (figure 5*a*). While comparing different *Hydra* species, one interesting tendency became obvious: the size of AMP families was generally larger in the representatives of

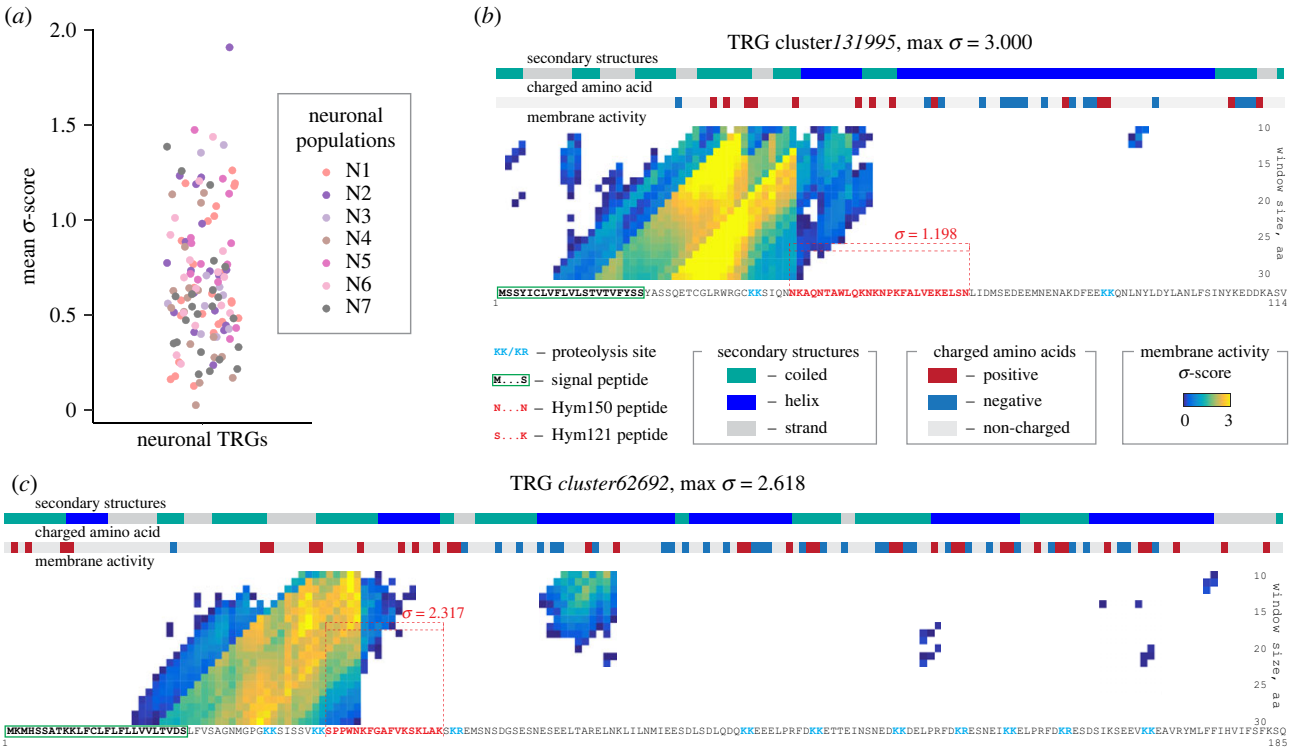


Figure 4. Machine learning algorithms allow for unbiased genome-wide prediction of putative AMPs. (*a*) Distribution of mean σ-score values for individual secreted peptides encoded by neuron-specific TRGs in seven neuronal subpopulations illustrates a high likelihood of containing active AMPs for the peptides. Data are from Klimovich *et al.* [16]. (*b*) *Hydra*-specific TRG *cluster131995* expressed exclusively in endodermal neurons N5 encodes a putative hitherto uncharacterized AMP. Moving-window protein scan prediction map with residue charge and secondary structure annotations. The heat map reflects the peptide's probability (σ-score) of being membrane active as predicted by the MLA [57]. High σ-scores (yellow) suggest that *cluster131995* peptide codes for a potent AMP. N-terminal signal peptide, putative proteolysis sites and a sequence identical to a previously discovered peptide Hym150 [58] are found within the cluster131995 peptide, providing evidence that a precursor *cluster131995* is processed and gives rise to a secreted active AMP. (*c*) The predicted profile of the peptide encoded in *cluster131995* resembles that of the TRG *cluster62692*, which has been previously demonstrated to contain a highly potent neuron-derived AMP Hym121 in *Hydra* [16].

the 'brown hydra' group compared to the green hydra *H. viridissima* (figures 2 and 5). For instance, only eight *periculin* genes and a single *Hym357* orthologue were found in the *H. viridissima* genome, and *hydramacin* appears to be missing in this species. This trend suggests that a major expansion of AMP gene families has occurred after the segregation of the 'green' and 'brown' hydra groups, which took place about 193 Ma [3]. This phylogenetic radiation coincides with a major change in the *Hydra* biology—the loss of its algal photosymbiont *Chlorella*. Given the tight metabolic co-dependence between *H. viridissima* and its endosymbiont *Chlorella* [65], such a transition must have been reflected in the entire holobiont biology and most likely had an impact on the relation with the extracellularly located microbiota. It is plausible that, upon the partner switch, certain function(s) previously allocated to the photosymbiont might have been re-allocated (outsourced) to the bacterial symbionts. This, in turn, necessitated a more elaborate system of control exerted by *Hydra* on its microbiome in the form of AMPs. This scenario is supported by the observation that symbiotic *H. viridissima* harbour a distinct microbiome clearly different from that of aposymbiotic (algae-free) polyps [66]. We also note that colonization of *H. viridissima* with *Chlorella* algae is associated with an up-regulation of multiple *Hydra*-restricted TRGs [65], which remain uncharacterized, but some of them might code for putative AMPs.

These observations prompt a hypothesis that the loss of a photosynthetic endosymbiont might be associated with the increasing role of the extracellular bacterial microbiome, which demands a more sophisticated control via complex AMP cocktails. To test the hypothesis whether the bi- or tripartite holobionts architecture is reflected in the complexity and evolutionary history of their AMP genes repertoire, a comprehensive analysis across members of the Cnidaria phylum is imperative. While virtually all Anthozoa species form stable association with intracellular photobionts and species-specific bacterial communities colonizing the surface mucus layer, the gastrovascular system, and the skeleton [67], members of other Cnidaria classes, such as Scyphozoa and Hydrozoa, rarely harbour photobionts. *H. viridissima* and *Cassiopea xamachana* are, in fact, rather exceptions among the hydrozoans [68]. Although some recent studies attempt to create a comprehensive survey of AMPs in Cnidaria [36], their focus remains bound to exclusively conserved gene families. Implementation of novel highly automatized algorithms for AMP detection and annotation, such as the MLAs, promises a major progress in understanding the link between AMP repertoire and holobiont architecture in Cnidaria.

In this context, it is particularly interesting to compare the diversification of AMPs to the evolutionary history of other immune genes in *Hydra* and other symbiotic and non-symbiotic Cnidaria. The diversity of TLR genes in *Hydra* is very low. In fact, only a single functional TLR is assembled from the products of two genes—*hyLRR* and *hyTRR* [5,7]. Genes coding for putative NOD-like receptors, on the contrary, have undergone expansion in *Hydra* indeed [8]. Intriguingly, the broadest repertoire of genes encoding NACHT- and

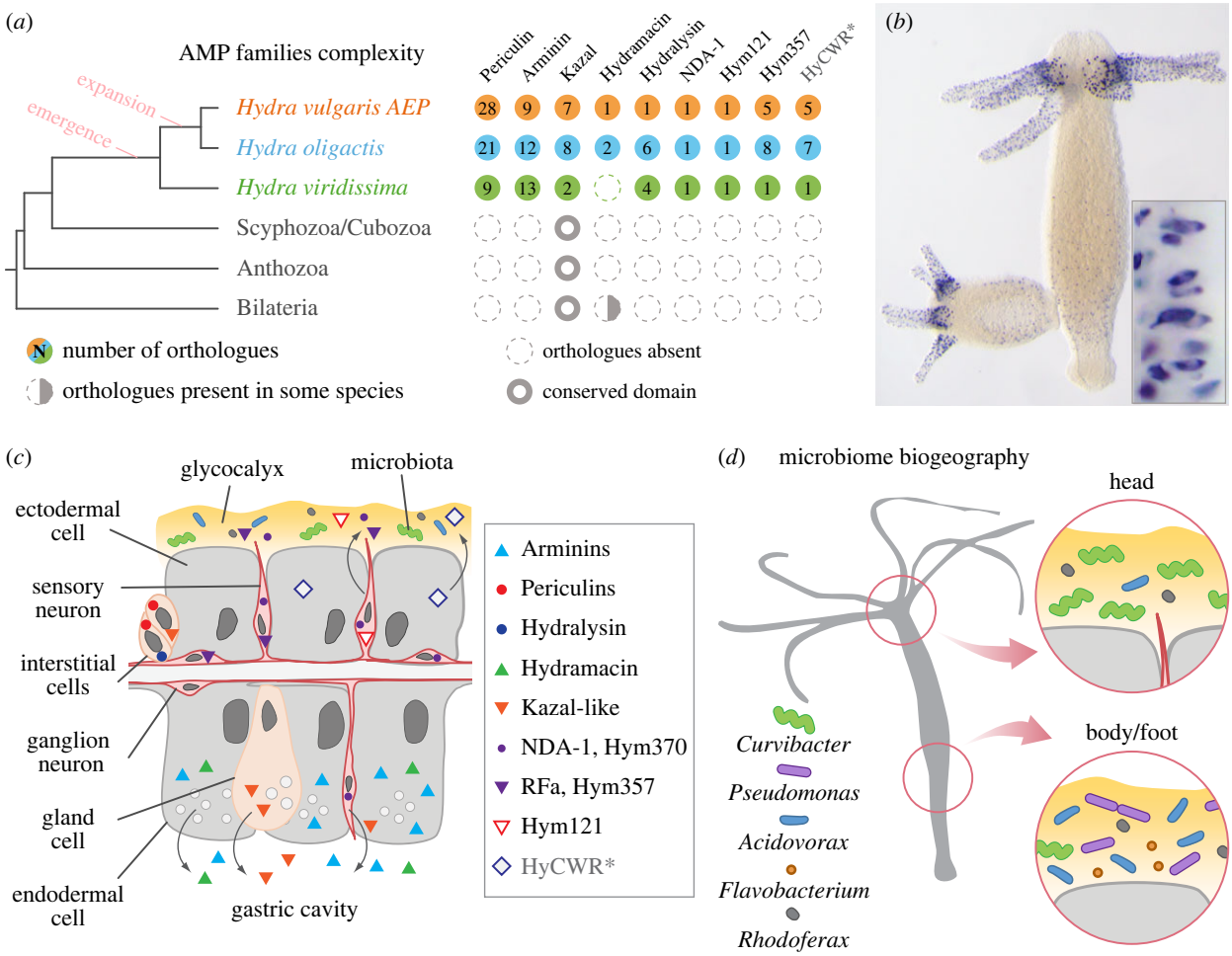


Figure 5. Complex species-specific and spatially restricted cocktails of AMPs sculpt the microbiome of *Hydra*. (*a*) Overview of the AMP gene family complexity in *Hydra* species. Note that *H. viridissima* possesses generally less AMP genes compared to *H. oligactis* and *H. vulgaris* AEP. Most AMP families are restricted to *Hydra* and only in a few cases can proteins with similar domains be detected in other cnidarians and/or bilaterian animals. (*b*) *Hydra*-specific AMP Hym121 is expressed in a distinct population of sensory neurons (N7) confined to the tentacles of *Hydra*, where it creates a selective microenvironment for specific members of the microbiome. *In situ* hybridization reveals the presence of Hym121 mRNAs. (*c*) Tissue and cell type-specific expression pattern of *Hydra*'s AMPs. While numerous AMPs are secreted into the gastric cavity by the endodermal layer, ectodermal epithelial cells may produce only few AMPs. This is particularly surprising given that the ectoderm is facing the environment. The AMPs produced by the neurons in the ectoderm are secreted directly into the glycocalyx, the habitat for symbiotic bacteria. (*d*) The sharp spatially controlled expression patterns of AMPs control the spatial organization of the *Hydra* microbiome—its biogeography. Asterisk, the antimicrobial activity of HyCWR peptides and their role in the *Hydra* holobiont remain to be validated.

NB-ARC-domain containing NOD-like receptors (over 260 in total) is observed in the green *H. viridissima* [21]. The arsenal of NLRs in the brown *H. vulgaris* is substantially smaller (about 89–101 genes). Hence, we observe here an inverse trend, compared to the AMP families—expansion of a gene family in the context of algal symbiosis and contraction in algae-free hydras. Remarkably, this trend is also evident on the scale of the phylum Cnidaria: symbiotic cnidarians, like the anemone *Acropora*, possess over 400 NLR genes, while a symbiont-free jellyfish *Morbakka* has only 24 genes, and *Nematostella* has only six genes for NLRs. Similarly, TIR-domain-containing proteins are substantially more abundant in *H. viridissima* (49) and *Acropora* (49) compared to symbiont-free *H. vulgaris* (11) and *Nematostella* (17) [21]. Therefore, the evolutionary development of symbiosis with algae by certain cnidarians likely required expansion and greater sophistication of genes encoding innate immunity pathway genes, critical for recognition and maintenance of symbiotic organisms in cnidarian tissues. The loss of photosynthetic symbionts resulted in contraction of the receptor-encoding gene families and expansion of the families encoding the effector molecules for communication with the prokaryotic partners—the AMPs.

In sum, the emergence and 'recent' elaboration of the AMP repertoire in the brown *Hydra* might be a signal of a change in the holobiont complexity and biology. The complexity of AMP families in diverse *Hydra* species, hence, represents a genomic footprint of a co-evolution between the host, similar to other species (e.g. the fly [69]) and its microbiome and reflects the species' adaptations to their unique microbial environments.

(b) AMPs shape the spatiotemporal structure of the *Hydra* microbiome

For several decades, it has been accepted that AMPs, as 'killers' and hence often referred to as host-defence peptides (HDP; [70]), protect an animal from noxious microorganisms. More recently, as stated above, we start appreciating a broader role of AMPs in shaping the commensal microbiome [71]. The *Hydra* host imposes strong selective forces on its microbiome via section of diverse AMPs [72] and thereby maintains

species-specific microbiota communities over extended periods [12,63,73]. Our observations expand this view and add a spatial dimension to these host–microbiome interactions. We provide evidence that AMPs in *Hydra* are expressed in a tightly regulated spatially controlled manner (figure 5*b*). A plethora of AMPs are expressed throughout the endoderm of *Hydra* along the entire body (figure 3). These peptides, most likely, keep the gastric cavity of a polyp essentially sterile and protect *Hydra* from pathogens. In the ectodermal layer, AMPs expressed mostly in distinct spatially restricted neuronal populations are, hence, confined to certain body domains (figures 3 and 5). This suggests that spatially confined AMP cocktails are secreted into the glycocalyx of *Hydra* and generate a complex chemical landscape on the polyp's surface. These distinct microhabitats shape locally the microbiome of *Hydra*. They not only regulate the density of the bacterial communities a healthy polyp harbours (so called carrying capacity; [74]) but also control the composition of these communities. As a result, certain species of bacteria, such as *Pseudomonas*, *Flavobacterium* and *Acidovorax*, are confined to the lower part of the *Hydra* polyp and virtually absent from the hypostome [75]. On the contrary, other members of the microbiome such as *Curvibacter* [11,76], are found more abundantly on the polyp's head and tentacles. Our analysis of AMP genes expression uncovered their strict spatially restricted production and suggested their contribution to the specific regionalization of the microbiome (figure 5*b–d*).

We have mechanistically proven the role of *Hydra*'s AMPS in shaping the microbial biogeography by genetically modifying the expression pattern of a nerve cell-specific AMP, NDA-1 [11]. Using a knockdown approach, we observed that the absence of NDA-1 peptides results in both a shift in the composition of the microbiome and a perturbation of the microbial biogeography [11].

6. Perspectives and open questions

Our bioinformatic analysis uncovered a remarkable expansion of AMPs families encoded in *Hydra*-specific TRGs. However, to fully understand the evolution of the AMP gene complement and implications of this complexity for the *Hydra* holobiont, further systematic studies are needed.

(a) Puzzling redundancy of AMP genes

Our analysis uncovered high evolutionary dynamics of AMP families in *Hydra*. Generally, duplication of species-specific AMP genes or their loss through pseudogenization are not uncommon in the animal kingdom [49,77,78], and most animals indeed possess a broad array of AMPs. The clustered genomic organization of AMP genes has also been recognized as characteristic for numerous AMP families across animal species [79–83]. However, we find it truly puzzling that numerous paralogues of AMP genes in *Hydra* though having slightly different nucleotide sequences, code for identical precursor polypeptides and, hence, give rise to identical active peptides. This is particularly evident in the case of the periculin family (figure 2*b*; electronic supplementary material, figure S2). The biological relevance of this apparent redundancy as well as the evolutionary mechanisms that lead to it remain unclear. A deep analysis of the paralogues' nucleotide sequences, such as the dN/dS estimation, may reveal sign of negative or positive selection. Additionally, comparison of gene complement and genomic organization between polyps from different geographically isolated populations of the same species might be informative. One can anticipate that such survey may even uncover single amino acid polymorphisms (similar, for instance, to the functionally crucial polymorphism S69R in diptericin A sequences [69]) or copy number variation in AMP genes within different *Hydra* clones. The current state of accuracy in genome sequencing and assembly allows detecting such genomic events.

In sum, the genome-wide survey of AMP repertoire in *Hydra* provides evidence for an expansion of AMP gene families. Together with observations on other invertebrate animals, plants and fungi [64,78,84,85], these findings support the view that elaboration of the AMP arsenal through novel family emergence, gene duplication and diversitifcation is a common, universal principle in AMP genes evolution.

(b) Uncovering further AMP families in *Hydra*

Our analysis was focused on a detailed analysis of previously identified AMP families in *Hydra*. Beyond that, we demonstrate how additional, novel tools allow discovering novel members of known families or even new families. For instance, using a hidden Markov model-based approach, we uncovered a novel family of putative secreted AMPs—the HyCWR family. Novel AI-based tools also allow unbiased genome-wide screening and *ab initio* detection of AMPs. Our preliminary analysis suggests that dozens of novel, previously not characterized AMPs and their families are still hidden in the genome unrecognized (figure 4*a*). This hypothesis is supported by our finding that clusters of tandemly repeated *Hydra*-specific TRGs, architecturally similar to, for instance, the *periculin* cluster (figure 2*a*), are scattered through the *Hydra* genome. For example, a dense cluster of over 30 relatively short collinear uncharacterized genes with no homologues outside *Hydra* (*G009076 – G009116*) can be found on chromosome 5 of *H. vulgaris* AEP.

Testing the hypothesis whether this plethora of genes encode novel AMPs and characterizing them represent a major analytical challenge. However, this analysis may be streamlined by applying improved AI tools. In our previous efforts, we trained the MLA using a dataset of predominantly human AMPs [57]. Therefore, our analysis had a certain bias and likely, favoured identifying AMPs with features common to those of Bilateria. However, the dynamic nature of MLA allows re-training them on additional or expanded datasets. Addition of already discovered and functionally validated AMPs from *Hydra* into the training dataset may substantially increase the accuracy of the MLAs. Moreover, the rapidly evolving tools for three-dimensional protein modelling, such as the AlphaFold and similar template-independent tools [86–88], offer an opportunity to predict with high confidence the folding of peptides and, hence, may greatly streamline the *in silico* analysis of putative AMPs and facilitate selection of candidates for testing *in vivo* and *in vitro*. We particularly emphasize that testing the function of candidate AMPs remains a major bottleneck. Not all peptides can be synthetized effectively in their active form and tested *in vitro*, and recombinant expression may be also challenging due to toxicity for cells. Finally, the *in vivo* studies of AMPs by manipulating the genes in the host though transgenesis are very laborious and require smart selection of candidates. AI algorithms represent an excellent tool for making an 'educated guess' and selecting candidates for in-depth validation.

The majority of AMPs in *Hydra* are encoded in *Hydra*-restricted TRGs (figure 5*a*), yet the *hydramacin* family is an exception. It appears to be confined to the brown *Hydra* group, since no orthologues were found in *H. viridissima* (electronic supplementary material, figure S5). This suggests that *hydramacin* either has emerged after the split of brown and green hydras, or has been lost in *H. viridissima*. The latter appears more plausible, since genes coding for proteins similar to hydramacin were found in several bilaterian species, such as leeches and molluscs [5,9,89,90]. Our synteny analysis (electronic supplementary material, figure S5), which indicates that the entire locus containing the *hydramacin* gene might have been lost in the green hydra lineage, provides an additional support for this hypothesis. Hence, the most parsimonious explanation of the mosaic *hydramacin* distribution on the phylogenetic tree is that the hydramacin family is ancient and likely common for all Eumetazoa, but its members have been either lost in some lineages or evolved beyond the level of detection. This gene loss might be not the only example of reduction in AMP repertoire in *Hydra*. For instance, our analysis provided no evidence for the presence of canonical defensins in *Hydra*. This appears surprising given a broad phylogenetic distribution of these peptides. However, partial or complete absence of some AMP families has been described in vertebrate and invertebrate species [34,77,91–93], supporting the high evolutionary dynamics of AMP families. To resolve the paradoxical absence of some AMP families in *Hydra* and identify the factors that might have caused this gene loss, a deep cross-species and, possibly, cross-isolate comparison of the genomic organization (exon-intron structure, synteny) are needed along with a survey of the microbiomes and the ecology of these species and isolates. Extensive implementation of AI tools may facilitate genome-wide discovery and comparison of AMP repertoires.

(c) Uncovering expression of 'silent' AMP genes

Although many AMP genes discovered in *Hydra* are characterized by a constitutive expression, a substantial fraction of AMP genes appears not to be expressed in homeostatic conditions (figure 3; electronic supplementary material, figure S7). This suggests that AMP expression is under a tight developmental as well as environmental control. In fact, some *periculin* genes are developmentally regulated and expressed at particularly high level in mature oocytes. As maternal antimicrobial peptides, they control bacterial colonization of the *Hydra* embryos [13]. Absence of expression in adult polyps also indicates that some AMPs might be inducible, and their expression is triggered upon a specific signal, such as encounter of a bacterial species or metabolite. Indeed, earlier observations provide strong evidence that expression of some *hydramacin*, *arminin* and *periculin* paralogues can be up-regulated in the presence of diverse bacterial products (LPS, flagellin) or danger signals (dsRNA) [5]. Moreover, interference with the upstream signalling pathways [18], and tissue manipulations such as amputation-induced regeneration [94] and elimination of neurons [95,96], also result in modulation of AMP gene expression. This may cause a concomitant enhanced antimicrobial activity of the tissue [97]. However, it remains unclear, whether the transcription of already expressed paralogues is elevated, or additional previously silent AMP genes (figures 2*c* and 3) are turned on. A particularly exciting possibility is that ectodermal cells that do not produce any AMPs in homeostatic condition (figures 3 and 5*c*), start expressing certain AMP genes following developmental or environmental signals.

(d) For AMPs the name no longer fits the function

As outlined in detail in another paper in this issue [98], from the beginning of animal (and plant) evolution, AMPs serve a crucial role in regulating the composition of the microbiome [1]. These findings make it quite clear that AMPs do much more than just kill pathogens. They play a 'silent' role in plant, animal and human health by permitting coexistence with environmental and symbiotic microbes, shaping the microbiome according to the susceptibility to particular AMPs, contributing to the spatial organization of the microbiota. Instead of being 'anti'-microbial, one could just as well speak of 'pro'-microbial peptides. The function of AMPs goes far beyond just killing bacteria. It is generally accepted that AMPs inhibit growth of microbes, through interfering with a diversity of cellular function in bacterial cells [2]. However, they can also interfere with the microbes physiology in plethora of other ways. Accumulating evidence indicates that AMPs may modulate formation of biofilms and swarming behaviour of microbes [99], or act as immunomodulators [100]. AMPs produced by *Hydra* may appear to display similar multifunctionality. Most of them do demonstrate strong growth-inhibiting activity in minimal inhibitory concentration (MIC) assays [9–11]. However, we noticed that some peptides have milder effects on target bacteria and rather change their physiology. For instance, that the Hym121 peptide effectively inhibits growth of *Curvibacter* and *Acidovorax*, but does not kill *Bacillus megaterium* and only alters its colony morphology, likely by reducing cells motility [16]. Therefore, this AMP actually acts as a signalling molecule (somehow reminiscent of the signalling role of microbe-derived antibiotics [101]). Similar observations remain very scarce, and no systematic survey of non-conventional roles of *Hydra* AMPs has been performed. Since MIC assays have been the main tools to test AMPs' activity and infer function, behaviour-modulating aspects of AMP activity have escaped detection so far. We emphasize the urgent need to develop and implement novel methods, such as motility assays and microcosm setups [74], to gain a comprehesive view of diverse AMP roles in animals. This thinking may also shape the development of *in silico* tools, such as activity predictors and AI-based algorithms (in line with the current efforts [56,102], whose logic has been mainly built around the membrane disruptive and bacteria killing properties. These developments may also fuel discovery and a guided design of novel antibiotics [56,103–105].

In sum, our survey of AMPs in *Hydra* uncovered a fascinating diversity and complex role of these TRGs in *Hydra* biology. It is generally accepted that emergence of novel, taxon-restricted genes may promote emergence of novel traits allowing access to a new environment. As demonstrated here, families of AMPs appear to represent an attractive system for experimentally dissecting the link between gene emergence and expansion, and a (meta)organism's phenotype and its adaptation to the environment. *Hydra* offers a unique experimental platform for testing how the host sculpts its microbiome, and the microbiome shapes the genome of its host. Hence, the studies on *Hydra* provide an evolutionary informed perspective onto the principles governing the intricate host–microbiome interactions and the molecular mechanisms behind them [17,106–108]. They enrich our understanding of the critical

factors maintaining the metaorganism homeostasis and health across the animal kingdom.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All data and references, such as Genbank accessions, are included in the manuscript and the electronic supplementary material.

The sequence datasets supporting this article have been uploaded as part of the electronic supplementary material, table S1 and Data S1 and S2.

Supplementary material is available online [109].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. A.K.: conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, visualization, writing—original draft, writing—review and editing; T.C.G.B.: conceptualization, funding acquisition, writing—original draft.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We have no competing interests to declare.

Funding. Research in the laboratory of T.C.G.B. is supported in part by grants from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), the CRC 1182 'Origin and Function of Metaorganisms' (to T.C.G.B.), and the CRC 1461 'Neurotronics: Bio-Inspired Information Pathways' (Project-ID 434434223 – SFB 1461) to T.C.G.B. and A.K. A.K. is supported by the DFG (grant no. KL3475/2-1). T.C.G.B. appreciates support from the Canadian Institute for Advanced Research.

Acknowledgements. We appreciate the help of Andrea Murillo and Jinru He in preparing illustrations. We apologize to our colleagues whose work we have not cited, owing to constraints on the length of this article.

References

1. Bosch TCG, Zasloff M. 2021 Antimicrobial peptides—or how our ancestors learned to control the microbiome. *MBio* **12**, e01847-21. (doi:10.1128/mBio.01847-21)
2. Zasloff M. 2002 Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395. (doi:10.1038/415389a)
3. Schwentner M, Bosch TCG. 2015 Revisiting the age, evolutionary history and species level diversity of the genus *Hydra* (Cnidaria: Hydrozoa). *Mol. Phylogenet. Evol.* **91**, 41–55. (doi:10.1016/j.ympev.2015.05.013)
4. Augustin R, Fraune S, Franzenburg S, Bosch TCG. 2012 Where simplicity meets complexity: *Hydra*, a model for host–microbe interactions. In *Recent advances on model hosts* (eds E Mylonakis, FM Ausubel, M Gilmore, A Casadevall), pp. 71–81. Berlin, Germany: Springer.
5. Bosch TCG *et al.* 2009 Uncovering the evolutionary history of innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defence. *Dev. Comp. Immunol.* **33**, 559–569. (doi:10.1016/j.dci.2008.10.004)
6. Klimovich AV, Bosch TCG. 2018 Rethinking the role of the nervous system: lessons from the *Hydra* holobiont. *Bioessays* **40**, 1800060. (doi:10.1002/bies.201800060)
7. Franzenburg S, Fraune S, Kunzel S, Baines JF, Domazet-Lošo T, Bosch TCG. 2012 MyD88-deficient *Hydra* reveal an ancient function of TLR signaling in sensing bacterial colonizers. *Proc. Natl Acad. Sci. USA* **109**, 19 374–19 379. (doi:10.1073/pnas.1213110109)
8. Lange C, Hemmrich G, Klostermeier UC, Lopez-Quintero JA, Miller DJ, Rahn T, Weiss Y, Bosch TCG, Rosenstiel P. 2010 Defining the origins of the NOD-like receptor system at the base of animal evolution. *Mol. Biol. Evol.* **28**, 1687–1702. (doi:10.1093/molbev/msq349)
9. Jung S *et al.* 2009 Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan *Hydra*. *J. Biol. Chem.* **284**, 1896–1905. (doi:10.1074/jbc.M804713200)
10. Augustin R, Siebert S, Bosch TCG. 2009 Identification of a kazal-type serine protease inhibitor with potent anti-staphylococcal activity as part of *Hydra*'s innate immune system. *Dev. Comp. Immunol.* **33**, 830–837. (doi:10.1016/j.dci.2009.01.009)
11. Augustin R *et al.* 2017 A secreted antibacterial neuropeptide shapes the microbiome of *Hydra*. *Nat. Commun.* **8**, 698. (doi:10.1038/s41467-017-00625-1)
12. Franzenburg S, Walter J, Künzel S, Wang J, Baines JF, Bosch TCG, Fraune S. 2013 Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc. Natl Acad. Sci. USA* **110**, E3730–E3738. (doi:10.1073/pnas.1304960110)
13. Fraune S, Augustin R, Anton-Erxleben F, Wittlieb J, Gelhaus C, Klimovich VB, Samoilovich MP, Bosch TCG. 2010 In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proc. Natl Acad. Sci. USA* **107**, 18 067–18 072. (doi:10.1073/pnas.1008573107)
14. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009 More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* **25**, 404–413. (doi:10.1016/j.tig.2009.07.006)
15. Augustin R, Anton-Erxleben F, Jungnickel S, Hemmrich G, Spudy B, Podschun R, Bosch TCG. 2009 Activity of the novel peptide arminin against multiresistant human pathogens shows the considerable potential of phylogenetically ancient organisms as drug sources. *Antimicrob. Agents Chemother.* **53**, 5245–5250. (doi:10.1128/AAC.00826-09)
16. Klimovich A *et al.* 2020 Prototypical pacemaker neurons interact with the resident microbiota. *Proc. Natl Acad. Sci. USA* **117**, 17 854–17 863. (doi:10.1073/pnas.1920469117)
17. Bosch TCG. 2012 Understanding complex host–microbe interactions in *Hydra*. *Gut Microbes* **3**, 345–351. (doi:10.4161/gmic.20660)
18. Mortzfeld BM, Taubenheim J, Fraune S, Klimovich AV, Bosch TCG. 2018 Stem cell transcription factor FoxO controls microbiome resilience in *Hydra*. *Front. Microbiol.* **9**, 629. (doi:10.3389/fmicb.2018.00629)
19. Sher D, Fishman Y, Zhang M, Lebendiker M, Gaathon A, Mancheno JM, Zlotkin E. 2005 Hydralysins, a new category of β-pore-forming toxins in Cnidaria. *J. Biol. Chem.* **280**, 22 847–22 855. (doi:10.1074/jbc.M503242200)
20. Cazet JF *et al.* 2023 A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state. *Genome Res.* **33**, 283–298. (doi:10.1101/gr.277040.122)
21. Hamada M, Satoh N, Khalturin K. 2020 A reference genome from the symbiotic hydrozoan *Hydra viridissima*. *G3* **10**, 3883–3895. (doi:10.1534/g3.120.401411)
22. Khalturin K *et al.* 2019 Medusozoan genomes inform the evolution of the jellyfish body plan. *Nat. Ecol. Evol.* **3**, 811–822. (doi:10.1038/s41559-019-0853-y)
23. Kim H-M *et al.* 2019 The genome of the giant Nomura's jellyfish sheds light on the early evolution of active predation. *BMC Biol.* **17**, 28. (doi:10.1186/s12915-019-0643-7)
24. Kon-Nanjo K, Kon T, Horkan HR, Febrimarsa SG, Steele RE, Cartwright P, Frank U, Parker J. 2023 Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*. *G3* **13**, jkad107. (doi:10.1093/g3journal/jkad107)
25. Santander MD, Maronna MM, Ryan JF, Andrade SCS. 2022 The state of Medusozoa genomics: current evidence and future challenges. *Gigascience* **11**, giac036. (doi:10.1093/gigascience/giac036)
26. Ying H *et al.* 2019 The whole-genome sequence of the coral *Acropora millepora*. *Genome Biol. Evol.* **11**, 1374–1379. (doi:10.1093/gbe/evz077)
27. Klimovich A, Wittlieb J, Bosch TCG. 2019 Transgenesis in *Hydra* to characterize gene function and visualize cell behavior. *Nat. Protoc.* **14**, 2069–2090. (doi:10.1038/s41596-019-0173-3)
28. Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TCG. 2006 Transgenic Hydra allow *in vivo* tracking of individual stem cells during

morphogenesis. *Proc. Natl Acad. Sci. USA* **103**, 6208–6211. (doi:10.1073/pnas.0510163103)

29. Neme R, Tautz D. 2014 Evolution: dynamics of *de novo* gene emergence. *Curr. Biol.* **24**, R238–R240. (doi:10.1016/j.cub.2014.02.016)
30. Rödelsperger C, Prabh N, Sommer RJ. 2019 New gene origin and deep taxon phylogenomics: opportunities and challenges. *Trends Genet.* **35**, 914–922. (doi:10.1016/j.tig.2019.08.007)
31. Tomasinsig L, Zanetti M. 2005 The cathelicidins—structure, function and evolution. *Curr. Protein Peptide Sci.* **6**, 23–34. (doi:10.2174/1389203053027520)
32. Hao X, Yang H, Wei L, Yang S, Zhu W, Ma D, Yu H, Lai R. 2012 Amphibian cathelicidin fills the evolutionary gap of cathelicidin in vertebrate. *Amino Acids* **43**, 677–685. (doi:10.1007/s00726-011-1116-7)
33. Uzzell T, Stolzenberg ED, Shinnar AE, Zasloff M. 2003 Hagfish intestinal antimicrobial peptides are ancient cathelicidins. *Peptides* **24**, 1655–1667. (doi:10.1016/j.peptides.2003.08.024)
34. Hanson MA, Lemaitre B. 2020 New insights on *Drosophila* antimicrobial peptide function in host defense and beyond. *Curr. Opin. Immunol.* **62**, 22–30. (doi:10.1016/j.coi.2019.11.008)
35. Ganz T. 2003 Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* **3**, 710–720. (doi:10.1038/nri1180)
36. Leal E, Múnera M, Suescún-Bolívar LP. 2022 *In silico* characterization of cnidarian's antimicrobial peptides. *Front. Mar. Sci.* **9**, 1065717. (doi:10.3389/fmars.2022.1065717)
37. Mitchell ML, Shafee T, Papenfuss AT, Norton RS. 2019 Evolution of cnidarian trans-defensins: sequence, structure and exploration of chemical space. *Proteins Struct. Funct. Bioinformat.* **87**, 551–560. (doi:10.1002/prot.25679)
38. Fu J, He Y, Peng C, Tang T, Jin A, Liao Y, Shi Q, Gao B. 2022 Transcriptome sequencing of the pale anemones (*Exaiptasia diaphana*) revealed functional peptide gene resources of sea anemone. *Front. Mar. Sci.* **9**, 856501. (doi:10.3389/fmars.2022.856501)
39. Ovchinnikova TV, Balandin SV, Aleshina GM, Tagaev AA, Leonova YF, Krasnodembsky ED, MenâShenin AV, Kokryakov VN. 2006 Aurelin, a novel antimicrobial peptide from jellyfish *Aurelia aurita* with structural features of defensins and channel-blocking toxins. *Biochem. Biophys. Res. Commun.* **348**, 514–523. (doi:10.1016/j.bbrc.2006.07.078)
40. Domazet-Lošo T, Klimovich A, Anokhin B, Anton-Erxleben F, Hamm MJ, Lange C, Bosch TCG. 2014 Naturally occurring tumours in the basal metazoan *Hydra*. *Nat. Commun.* **5**, 4222. (doi:10.1038/ncomms5222)
41. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. 2019 Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution. *Science* **365**, eaav9314. (doi:10.1126/science.aav9314)
42. Primack AS, Cazet JF, Little HM, Mühlbauer S, Cox BD, David CN, Farrell JA, Juliano CE. 2023 Differentiation trajectories of the *Hydra* nervous system reveal transcriptional regulators of neuronal fate. *bioRxiv* 531610. (doi:10.1101/2023.03.15.531610)
43. Siebert S, Anton-Erxleben F, Bosch TCG. 2008 Cell type complexity in the basal metazoan *Hydra* is maintained by both stem cell based mechanisms and transdifferentiation. *Dev. Biol.* **313**, 13–24. (doi:10.1016/j.ydbio.2007.09.007)
44. Mund A, Brunner A-D, Mann M. 2022 Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* **82**, 2335–2349. (doi:10.1016/j.molcel.2022.05.022)
45. Pirtskhalava M *et al.* 2021 DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Res.* **49**, D288–D297. (doi:10.1093/nar/gkaa991)
46. Wang G, Li X, Wang Z. 2016 APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44**, D1087–D1093. (doi:10.1093/nar/gkv1278)
47. Gómez EA, Giraldo P, Orduz S. 2017 InverPep: a database of invertebrate antimicrobial peptides. *J. Glob. Antimicrob. Resist.* **8**, 13–17. (doi:10.1016/j.jgar.2016.10.003)
48. Witten J, Witten Z. 2019 Deep learning regression model for antimicrobial peptide design. *bioRxiv* 692681. (doi:10.1101/692681)
49. Lazzaro BP, Zasloff M, Rolff J. 2020 Antimicrobial peptides: application informed by evolution. *Science* **368**, eaau5480. (doi:10.1126/science.aau5480)
50. Aguilera-Puga M DC, Cancelarich NL, Marani MM, De La Fuente-Nunez C, Plisson F. 2024 Accelerating the discovery and design of antimicrobial peptides with artificial intelligence. In *Computational drug discovery and design* (eds M Gore, UB Jagtap), pp. 329–352. New York, NY: Springer.
51. Huang J *et al.* 2023 Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nat. Biomed. Eng.* **7**, 797–810. (doi:10.1038/s41551-022-00991-2)
52. Lata S, Sharma BK, Raghava GPS. 2007 Analysis and prediction of antibacterial peptides. *BMC Bioinf.* **8**, 1–10. (doi:10.1186/1471-2105-8-263)
53. Porto WF, Pires ÁS, Franco OL. 2012 CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides. *PLoS ONE* **7**, e51444. (doi:10.1371/journal.pone.0051444)
54. Torrent M, Andreu D, Nogués VM, Boix E. 2011 Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS ONE* **6**, e16968. (doi:10.1371/journal.pone.0016968)
55. Fjell CD, Jenssen H, Hilpert K, Cheung WA, Panté N, Hancock REW, Cherkasov A. 2009 Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.* **52**, 2006–2015. (doi:10.1021/jm8015365)
56. Wong F, De La Fuente-Nunez C, Collins JJ. 2023 Leveraging artificial intelligence in the fight against infectious diseases. *Science* **381**, 164–170. (doi:10.1126/science.adh1114)
57. Lee EY, Fulan BM, Wong GCL, Ferguson AL. 2016 Mapping membrane activity in undiscovered peptide sequence space using machine learning. *Proc. Natl Acad. Sci. USA* **113**, 13 588–13 593. (doi:10.1073/pnas.1609893113)
58. Fujisawa T. 2008 Hydra peptide project 1993–2007. *Dev. Growth Differ.* **50**, S257–S268. (doi:10.1111/j.1440-169X.2008.00997.x)
59. Lee EY *et al.* 2021 PACAP is a pathogen-inducible resident antimicrobial neuropeptide affording rapid and contextual molecular host defense of the brain. *Proc. Natl Acad. Sci. USA* **118**, e1917623117. (doi:10.1073/pnas.1917623117)
60. Lee EY, Srinivasan Y, Anda J de, Nicastro LK, Tükel Ç, Wong GCL. 2020 Functional reciprocity of amyloids and antimicrobial peptides: rethinking the role of supramolecular assembly in host defense, immune activation, and inflammation. *Front. Immunol.* **11**, 1629. (doi:10.3389/fimmu.2020.01629)
61. Schluesener HJ, Su Y, Ebrahimi A, Pouladsaz D. 2012 Antimicrobial peptides in the brain: neuropeptides and amyloid. *Front. Biosci.* **4**, 1375–1380. (doi:10.2741/s339)
62. Abbott A. 2020 Are infections seeding some cases of Alzheimer's disease? *Nature.* **587**, 22–25. (doi:10.1038/d41586-020-03084-9)
63. Fraune S, Bosch TCG. 2007 Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. *Proc. Natl Acad. Sci. USA* **104**, 13 146–13 151. (doi:10.1073/pnas.0703375104)
64. Gerdol M, Schmitt P, Venier P, Rocha G, Rosa RD, Destoumieux-Garzón D. 2020 Functional insights from the evolutionary diversification of big defensins. *Front. Immunol.* **11**, 758. (doi:10.3389/fimmu.2020.00758)
65. Hamada M *et al.* 2018 Metabolic co-dependence drives the evolutionarily ancient *Hydra–Chlorella* symbiosis. *Elife* **7**, e35122. (doi:10.7554/eLife.35122)
66. Bathia J, Schröder K, Fraune S, Lachnit T, Rosenstiel P, Bosch TCG. 2022 Symbiotic algae of *Hydra viridissima* play a key role in maintaining homeostatic bacterial colonization. *Front. Microbiol.* **13**, 869666. (doi:10.3389/fmicb.2022.869666)
67. Van Oppen MJH, Blackall LL. 2019 Coral microbiome dynamics, functions and design in a changing world. *Nat. Rev. Microbiol.* **17**, 557–567. (doi:10.1038/s41579-019-0223-4)
68. Kayal E, Bentlage B, Sabrina Pankey M, Ohdera AH, Medina M, Plachetzki DC, Collins AG, Ryan JF. 2018 Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evol. Biol.* **18**, 68. (doi:10.1186/s12862-018-1142-0)
69. Hanson MA, Grollmus L, Lemaitre B. 2023 Ecology-relevant bacteria drive the evolution of host antimicrobial peptides in *Drosophila*. *Science* **381**, eadg5725. (doi:10.1126/science.adg5725)
70. Brown KL, Hancock REW. 2006 Cationic host defense (antimicrobial) peptides. *Curr. Opin. Immunol.* **18**, 24–30. (doi:10.1016/j.coi.2005.11.004)
71. Bosch T, Blaser M, Ruby E, McFall-Ngai M. 2024 A new lexicon in the age of microbiome research. *Phil.*

Trans. R. Soc. B **379**, 20230060. (doi:10.1098/rstb.2023.0060)

72. Deines P, Lachnit T, Bosch TCG. 2017 Competing forces maintain the *Hydra* metaorganism. *Immunol. Rev.* **279**, 123–136. (doi:10.1111/imr.12564)
73. Taubenheim J, Miklós M, Tökölyi J, Fraune S. 2022 Population differences and host species predict variation in the diversity of host-associated microbes in hydra. *Front. Microbiol.* **13**, 799333. (doi:10.3389/fmicb.2022.799333)
74. Deines P, Hammerschmidt K, Bosch TCG. 2020 Microbial species coexistence depends on the host environment. *MBio* **11**, 10–128. (doi:10.1128/mbio.00807-20)
75. Nawroth JC, Giez C, Klimovich A, Kanso E, Bosch TCG. 2023 Spontaneous body wall contractions stabilize the fluid microenvironment that shapes host-microbe associations. *Elife* **12**, e83637. (doi:10.7554/eLife.83637)
76. Giez C *et al.* 2023 Multiple neuronal populations control the eating behavior in *Hydra* and are responsive to microbial signals. *Curr. Biol.* **33**, 5288–5303. (doi:10.1016/j.cub.2023.10.038)
77. Fjell CD, Jenssen H, Fries P, Aich P, Griebel P, Hilpert K, Hancock REW, Cherkasov A. 2008 Identification of novel host defense peptides and the absence of α-defensins in the bovine genome. *Proteins Struct. Funct. Bioinformat.* **73**, 420–430. (doi:10.1002/prot.22059)
78. Vilcinskas A, Mukherjee K, Vogel H. 2013 Expansion of the antimicrobial peptide repertoire in the invasive ladybird *Harmonia axyridis*. *Proc. R. Soc. B* **280**, 20122113. (doi:10.1098/rspb.2012.2113)
79. Huttner KM, Lambeth MR, Burkin HR, Burkin DJ, Broad TE. 1998 Localization and genomic organization of sheep antimicrobial peptide genes. *Gene* **206**, 85–91. (doi:10.1016/S0378-1119(97)00569-6)
80. Clemmons AW, Lindsay SA, Wasserman SA. 2015 An effector peptide family required for *Drosophila* Toll-mediated immunity. *PLoS Pathog.* **11**, e1004876. (doi:10.1371/journal.ppat.1004876)
81. Sinner MP, Masurat F, Ewbank JJ, Pujol N, Bringmann H. 2021 Innate immunity promotes sleep through epidermal antimicrobial peptides. *Curr. Biol.* **31**, 564–577. (doi:10.1016/j.cub.2020.10.076)
82. Belov K *et al.* 2007 Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res.* **17**, 982–991. (doi:10.1101/gr.6121807)
83. Amid C, Rehaume LM, Brown KL, Gilbert JGR, Dougan G, Hancock REW, Harrow JL. 2009 Manual annotation and analysis of the defensin gene cluster in the C57BL/6 J mouse reference genome. *BMC Genomics* **10**, 606. (doi:10.1186/1471-2164-10-606)
84. Campos ML, De Souza CM, De Oliveira KBS, Dias SC, Franco OL. 2018 The role of antimicrobial peptides in plant immunity. *J. Exp. Bot.* **69**, 4997–5011. (doi:10.1093/jxb/ery294)
85. Wu J, Gao B, Zhu S. 2014 The fungal defensin family enlarged. *Pharmaceuticals* **7**, 866–880. (doi:10.3390/ph7080866)
86. Varadi M *et al.* 2022 AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444. (doi:10.1093/nar/gkab1061)
87. Jumper J *et al.* 2021 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. (doi:10.1038/s41586-021-03819-2)
88. Meng Q, Guo F, Tang J. 2023 Improved structure-related prediction for insufficient homologous proteins using MSA enhancement and pre-trained language model. *Brief. Bioinform.* **24**, bbad217. (doi:10.1093/bib/bbad217)
89. Tasiemski A, Vandenbulcke F, Mitta G, Lemoine J, Lefebvre C, Sautiere P-E, Salzet M. 2004 Molecular characterization of two novel antibacterial peptides inducible upon bacterial challenge in an annelid, the leech *Theromyzon tessulatum*. *J. Biol. Chem.* **279**, 30 973–30 982. (doi:10.1074/jbc.M312156200)
90. Xie Z, Yao T, Ye L, Wang J. 2021 Molecular characterization and expression analysis of an antimicrobial peptide, mytimacin-6, in the small abalone, *Haliotis diversicolor*. *Isr. J. Aquacult.* **73**, 1–11. (doi:10.46989/001c.25815)
91. Shelomi M, Jacobs C, Vilcinskas A, Vogel H. 2020 The unique antimicrobial peptide repertoire of stick insects. *Dev. Comp. Immunol.* **103**, 103471. (doi:10.1016/j.dci.2019.103471)
92. Vilcinskas A. 2013 Evolutionary plasticity of insect immunity. *J. Insect. Physiol.* **59**, 123–129. (doi:10.1016/j.jinsphys.2012.08.018)
93. Hanson MA, Lemaitre B, Unckless RL. 2019 Dynamic evolution of antimicrobial peptides underscores trade-offs between immunity and ecological fitness. *Front. Immunol.* **10**, 2620. (doi:10.3389/fimmu.2019.02620)
94. Wenger Y, Buzgariu W, Reiter S, Galliot B. 2014 Injury-induced immune responses in *Hydra*. *Semin. Immunol.* **26**, 277–294. (doi:10.1016/j.smim.2014.06.004)
95. Wenger Y, Buzgariu W, Galliot B. 2016 Loss of neurogenesis in *Hydra* leads to compensatory regulation of neurogenic and neurotransmission genes in epithelial cells. *Phil. Trans. R. Soc. B* **371**, 20150040. (doi:10.1098/rstb.2015.0040)
96. Fraune S, Abe Y, Bosch TCG. 2009 Disturbing epithelial homeostasis in the metazoan *Hydra* leads to drastic changes in associated microbiota. *Environ. Microbiol.* **11**, 2361–2369. (doi:10.1111/j.1462-2920.2009.01963.x)
97. Kasahara S, Bosch TCG. 2003 Enhanced antibacterial activity in *Hydra* polyps lacking nerve cells. *Dev. Comp. Immunol.* **27**, 79–85. (doi:10.1016/S0145-305X(02)00073-3)
98. Bosch T, Blaser M, Ruby E, McFall-Ngai M. 2024 A new lexicon in the age of microbiome research. *Phil. Trans. R. Soc. B* **379**, 20230060. (doi:10.1098/rstb.2023.0060)
99. Hancock REW, Alford MA, Haney EF. 2021 Antibiofilm activity of host defence peptides: complexity provides opportunities. *Nat. Rev. Microbiol.* **19**, 786–797. (doi:10.1038/s41579-021-00585-w)
100. Hilchie AL, Wuerth K, Hancock REW. 2013 Immune modulation by multifaceted cationic host defense (antimicrobial) peptides. *Nat. Chem. Biol.* **9**, 761–768. (doi:10.1038/nchembio.1393)
101. Yim G, Huimi Wang H, Davies J. 2007 Antibiotics as signalling molecules. *Phil. Trans. R. Soc. B* **362**, 1195–1200. (doi:10.1098/rstb.2007.2044)
102. Wan F, De La Fuente-Nunez C. 2023 Mining for antimicrobial peptides in sequence space. *Nat. Biomed. Eng.* **7**, 707–708. (doi:10.1038/s41551-023-01027-z)
103. Silveira GGOS, Torres MDT, Ribeiro CFA, Meneguetti BT, Carvalho CME, De La Fuente-Nunez C, Franco OL, Cardoso MH. 2021 Antibiofilm peptides: relevant preclinical animal infection models and translational potential. *ACS Pharmacol. Transl. Sci.* **4**, 55–73. (doi:10.1021/acsptsci.0c00191)
104. Torres MDT, De La Fuente-Nunez C. 2019 Toward computer-made artificial antibiotics. *Curr. Opin. Microbiol.* **51**, 30–38. (doi:10.1016/j.mib.2019.03.004)
105. Li C *et al.* 2022 AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens. *BMC Genomics* **23**, 1–15. (doi:10.1186/s12864-022-08310-4)
106. Obata Y, Pachnis V. 2020 Linking neurons to immunity: lessons from *Hydra*. *Proc. Natl Acad. Sci.* **117**, 19 624–19 626. (doi:10.1073/pnas.2011637117)
107. Hadfield MG, Bosch TCG. 2020 Cellular dialogues between hosts and microbial symbionts. In *Cellular dialogues in the holobiont* (eds TCG Bosch, MG Hadfield), pp. 287–290. Boca Raton, FL: CRC Press.
108. He J, Bosch TCG. 2022 *Hydra*'s lasting partnership with microbes: the key for escaping senescence? *Microorganisms* **10**, 774. (doi:10.3390/microorganisms10040774)
109. Klimovich A, Bosch TCG. 2024 Novel technologies uncover novel 'anti'-microbial peptides in *Hydra* shaping the species-specific microbiome. Figshare. (doi:10.6084/m9.figshare.c.7105351)

Supplementary materials for the manuscript

Novel technologies uncover novel “anti”-microbial peptides in *Hydra* shaping the species-specific microbiome

Alexander Klimovich*, Thomas C.G. Bosch*

Zoological Institute, Christian-Albrechts University of Kiel, Am Botanischen Garten 1-9, 24118 Kiel, Germany

ORCID:

Alexander Klimovich 0000-0003-1764-0613

Thomas C.G. Bosch 0000-0002-9488-5545

*To whom correspondence may be addressed: aklimovich@zoologie.uni-kiel.de (A.K.) and tbosch@zoologie.uni-kiel.de (T.C.G.B.)

Supplementary Text

BLAST-based identification of paralogues and orthologues

To uncover the repertoire of AMP coding genes in the genome of *H. vulgaris AEP*, we performed extensive BLAST search. As initial seed queries, we used predicted amino acid sequences of previously reported AMPs from diverse *Hydra* species deposited in GenBank: arminins accession numbers KC701494–KC701520; periculins FJ517724–FJ517733; hydramacin-1 XM_047274969; hydralysins AY655142, *AY967765, AY967764*; kazal FJ496649–FJ496653; NDA-1 XM_002162825; Hym357 AB018544, Hym-121 (reported in (1)). We performed a BLAST search (tblastn) against the genome of *H. vulgaris AEP* (2) available at https://research.nhgri.nih.gov/HydraAEP/. Matches with expectation e-value <10e-5 were considered as signs of homologue presence and were verified by manual domain composition analysis using SMART (3), and reciprocal BLAST against the NCBI database. To ensure that no paralogues are missing, me performed iterative a reciprocal BLAST search of *H. vulgaris* AEP uncovered AMPs against *H. vulgaris* AEP genome. In a few cases, some hits with high BLAST confidence were detected in *H. vulgaris* AEP genome but the peptide sequence was found to be a part of a bigger non-related protein (for instance, BLAST analysis of arminins consistently identified the gene G009888 as a putative arminin homologue, yet in fact that gene encodes a much longer enzyme gamma-glutamyltransferase). Such false hits were discarded from the analysis.

To detect AMP orthologues in other *Hydra* species and other cnidarians, we used AMP genes from *H. vulgaris AEP* as seed queries. As target datasets, we used the genome of *H. oligactis* (2) available at https://research.nhgri.nih.gov/HydraAEP/, *H. viridissima* genome (4) available at https://marinegenomics.oist.jp/hydra_viridissima_a99/blast/, the genome of *Clytia hemisphaerica* genome GCA902728285v1 (5) available at ENSEMBL. Finally, to test for the presence of AMP genes orthologues in other cnidarian species outside Hydrozoa, we used a set of orthologues from numerous cnidarian species previously published by Khalturin and co-authors (6). Similarly, matches with expectation e-value <10e-5 were considered as signs of homologue presence and were verified by manual domain composition analysis using SMART (3), and a reciprocal BLAST against the *H. vulgaris AEP* genome.

HMM-based identification of paralogues and homologues

In order to verify the completeness of the AMP genes repertoire in *Hydra* species, uncover putative distant homologues that might have evolved beyond the recognition by BLAT approach, we made use of a search based on hidden Markov model (HMM) profiles. First, we collected amino acid sequences of all reliably detected AMPs (see **Suppl. Table 1**) from 3 *Hydra* species – *H. vulgaris*, *H. oligactis*, and *H. viridissima*. After removal of the signal peptide

regions, the sequences were aligned using Clustal Omega algorithm (7). The resulting alignments in STOCKHOLM format were used to generate HMM profiles using the HMMER v.3.4 package (8). The HMM profiles are presented in **Suppl. Data 1**. The HMM profiles were further screened against a dataset made by combining predicted proteomes of three *Hydra* species using the HMMER package and default parameters. The lists of hits were manually curated and the sequences of were inspected for the presence of a signal peptide, overall length, absence of other domains and similarity to other, conserved proteins.

Similar HMM-based approach was used to test whether any AMPs of defensin family are encoded in genomes of three *Hydra* species. Defensins is broad family of AMPs present in animals, plants and fungi, and their sequence is poorly conserved apart from the specific cysteine-bridge motifs. Hence, the HMM approach which accounts for such as conserved spacing of motifs or Cys residues, appears more productive. To generate defensin HMM profiles (**Suppl. Data 2**), we used extensive sets of defensin peptides represented in the PFAM dataset by the following entries: PF00323 – mammalian defensins (257 peptides), PF13841 - beta-defensins (1944 peptides), PF00711 – beta-defensins (3509 peptides), PF14862 – big defensins (79 peptides), and PF01097 – arthropod defensins (356 peptides). The alignments of these peptides (with signal peptides removed) were exported in STOCKHOLM format and used for HMM profile generation (**Suppl. Data 2**). Profile logos (Suppl. Fig. 6) were generated using Skylign (9). We used these HMM profiles to screen the following datasets: proteomes of three *Hydra* species, proteomes of *Clytia hemisphaerica* (5) and *Nematostella vectensis* (10), and proteomes of diverse cnidarians and Bilateria (6). The stringency for HMM search was reduced to an E-value threshold <10e-5. The results of HMM screen were manually inspected for the presence of the defensin characteristic features: N-terminal signal peptide, total length of mature peptide <100 amino acids, characteristic pattern of at least 6 cysteins, absence of other domains.

Synteny analysis of the *hydramacin* locus

To verify the absence of the *hydramacin* gene in *H. viridissima* genome, we performed synteny analysis. First, we identified the genomic locus on the chromosome 1 of *H. vulgaris* AEP (360 kbp), which contains the *hydramacin* gene itself (*G001163*), five protein-coding genes upstream (*G001157-G001162*) and six genes downstream (*G001164-G001169*; see **Suppl. Fig. 5**). Next, using the BLAST tool on the https://research.nhgri.nih.gov/HydraAEP/, we identified a syntenic region on the scaffold 0031 in the genome of *H. oligactis*. Unfortunately, since the genome of this species is more fragmented, the locus was not complete, and the genes downstream from the *hydramacin* itself (*G22718*) were missing. However, all five genes upstream of *hydramacin* (*G22708-G22712*) were present in the same order and orientation as

in *H. vulgaris AEP*, indicating a solid synteny. Finally, we used BLAST on the https://marinegenomics.oist.jp/hydra_viridissima_a99/ resource and identified a syntenic region on the scaffold 17 of *H. viridissima* genome (**Suppl. Fig. 5**). The first five genes (*s17.g88-s17.g94*) demonstrated clear synteny to *H. vulgaris AEP*, whereby only one gene was in the opposite orientation but still in the conserved order. Two downstream genes (*s17.g95* and *s17.g96*) were demonstrated also similarity on position and orientation. However, no protein-coding genes were detected between the gens *s17.94* and *s17.95*, where the *hydramacin* sequence would be expected. In fact, we noticed that the distance between these two genes was substantially shorter (6,5 kbp, **Suppl. Fig. 5**) compared to the corresponding genomic region between the syntenic genes *G001162* and *G001166* on *H. vulgaris* chromosome (126 kbp). This suggests that the *hydramacin* gene either has emerged upon an insertion of a 100 kbp region in the brown hydra region or was deleted in the *H. viridissima*. Since no additional genomes of *Hydra* and closely related species are available, further reconstruction of the genomic events that lead to the absence of the *hydramacin* gene in *H. viridissima* remains not feasible.

Generation of phylogenetic trees

To infer the evolutionary history of the *periculin* and *arminin* families in *Hydra* genus, we generated phylogenetic trees. We performed an alignment of all orthologues using Clustal W algorithm (11,12) and identified the best model for the phylogenetic relation inference based on the alignment using the Mega7 software (13). Finally, we generated phylogenetic trees using Maximum Likelihood method based on the Jones-Taylor-Thornton with Gamma Distributed substitution rate (JTT+G) model and 1,000 bootstrap iterations. The trees with the highest log likelihood are shown in **Suppl. Fig. 3 and 4**, and simplified compressed trees are shown in **Fig. 2 B,E**. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Only fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position.

Analysis of AMP genes genomic organization

To gain insight into the genomic organization of AMP genes, we made use of the *H. vulgaris AEP* genome (2) available at https://research.nhgri.nih.gov/HydraAEP/. From the genome browser and individual gene's pages, we extracted the chromosomal position of each AMP gene, orientation of its ORF and position of exons and introns. These data were plotted using Adobe Illustrator and presented in **Fig. 2A,D,F**. Chromatin accessibility data were also retrieved and plotted using the *H. vulgaris AEP* genome browser, and later converted into Adobe Illustrator vector graphics.

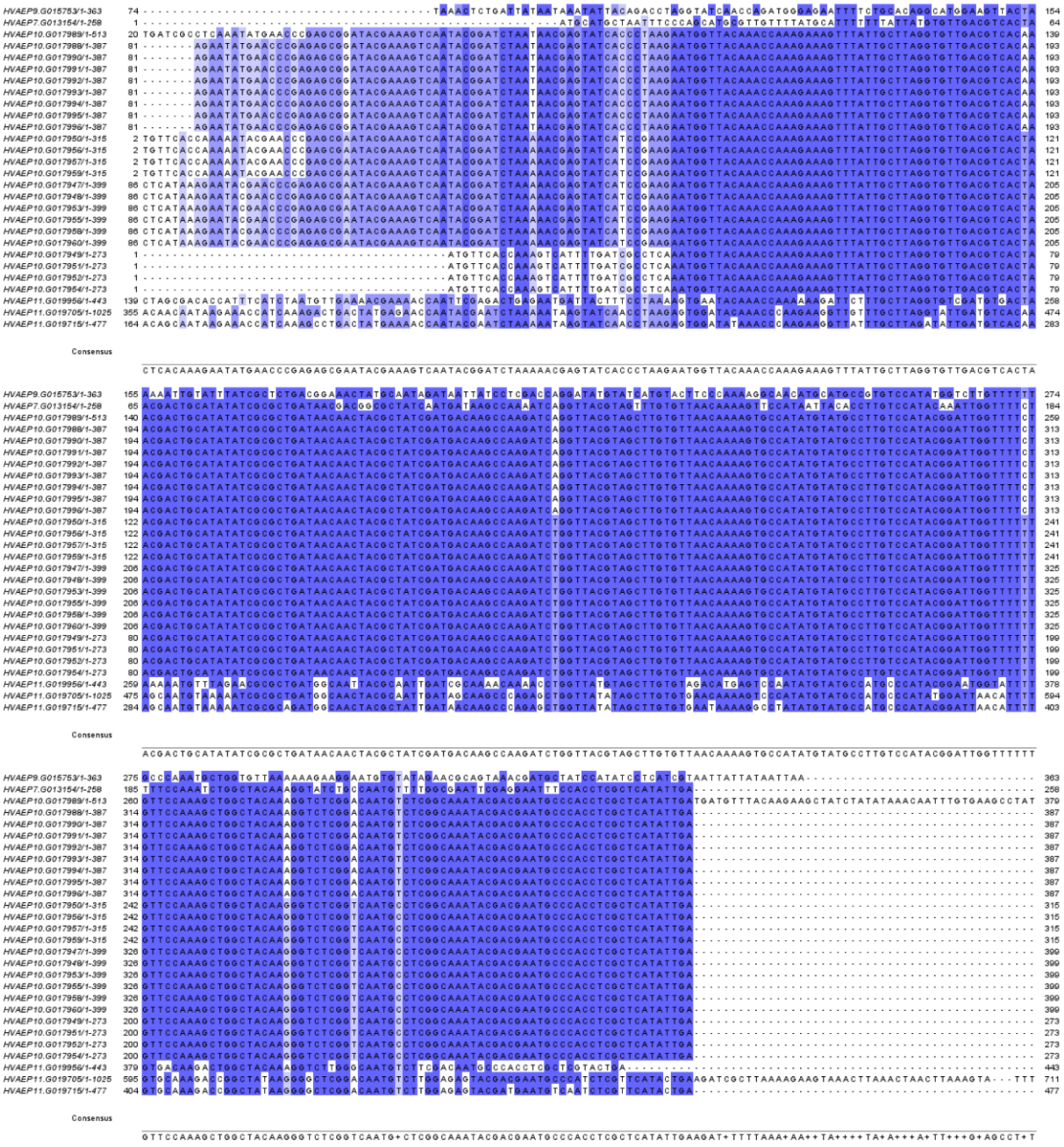
Analysis of gene expression with single cell resolution

To visualise the expression of AMP genes in individual cell types of *Hydra*, we used the single-cell expression atlas (2) available at https://research.nhgri.nih.gov/HydraAEP/SingleCellBrowser/. All genes for which expression data were available in the dataset, were plotted and grouped by categorical cell information (cell cluster) and family. The normalised expression values were averaged, log-transformed and then plotted (**Fig. 3**). Additionally, raw expression values for two genes, *arminin G009435* and *periculin G019705* (see **Suppl. Fig. 8**) were extracted and plotted as violin diagrams using R package (14).

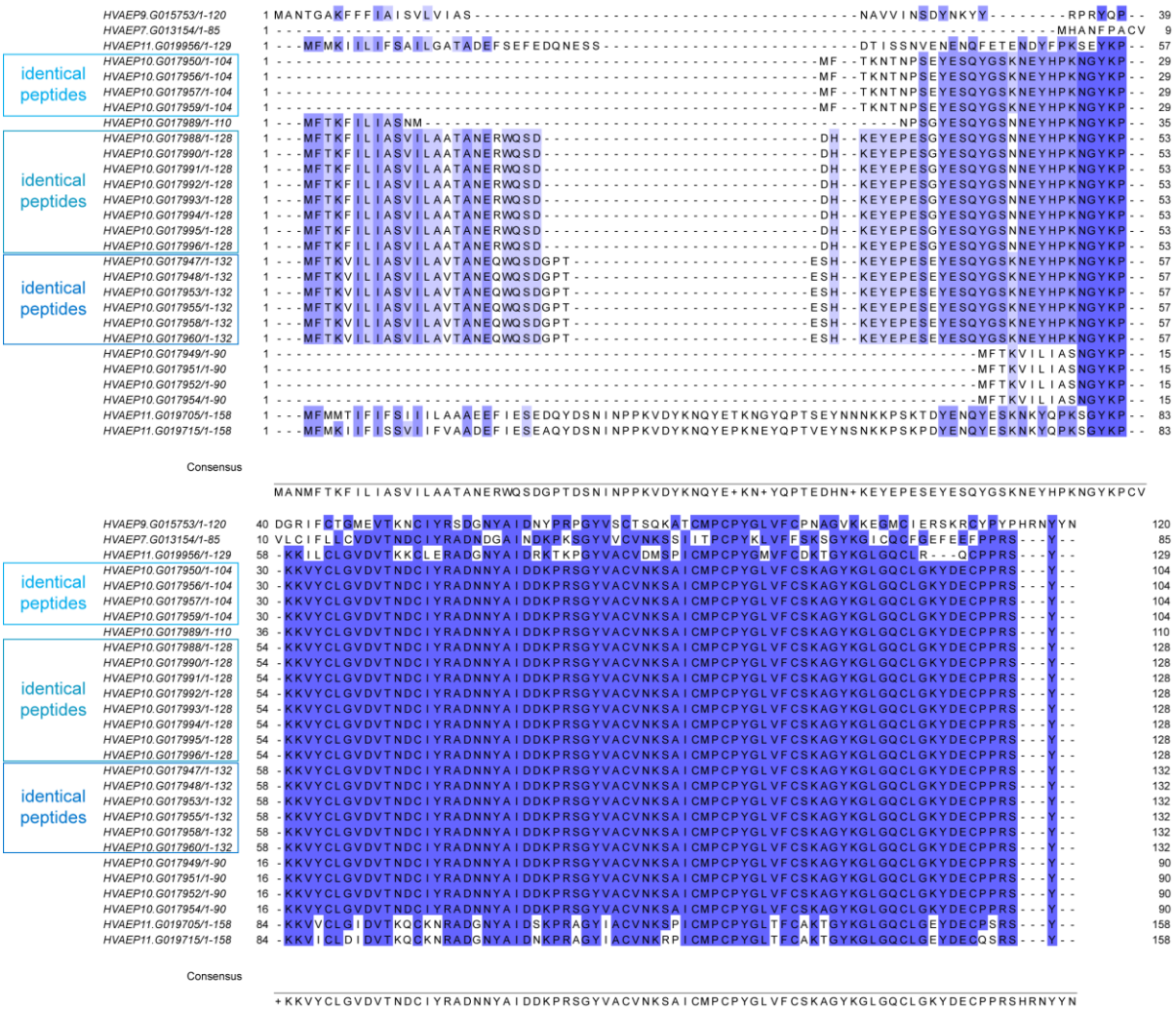
***In situ* hybridization**

Expression pattern of two genes encoding putative HyCWR peptides (*G014589* and *G021955*) was detected in whole mount *Hydra* preparations by *in situ* hybridization with an anti-sense digoxigenin (DIG) -labelled RNA probe as previously described (15). DIG-labelled sense-probe was used as a control. Signal was developed using anti-DIG antibodies conjugated to alkaline phosphatase (1:2000, Roche) and NBT/BCIP staining solution (Roche). Images of the *in situ* preparations were captured on a Zeiss Axioscope microscope with an Olympus DP74 camera.

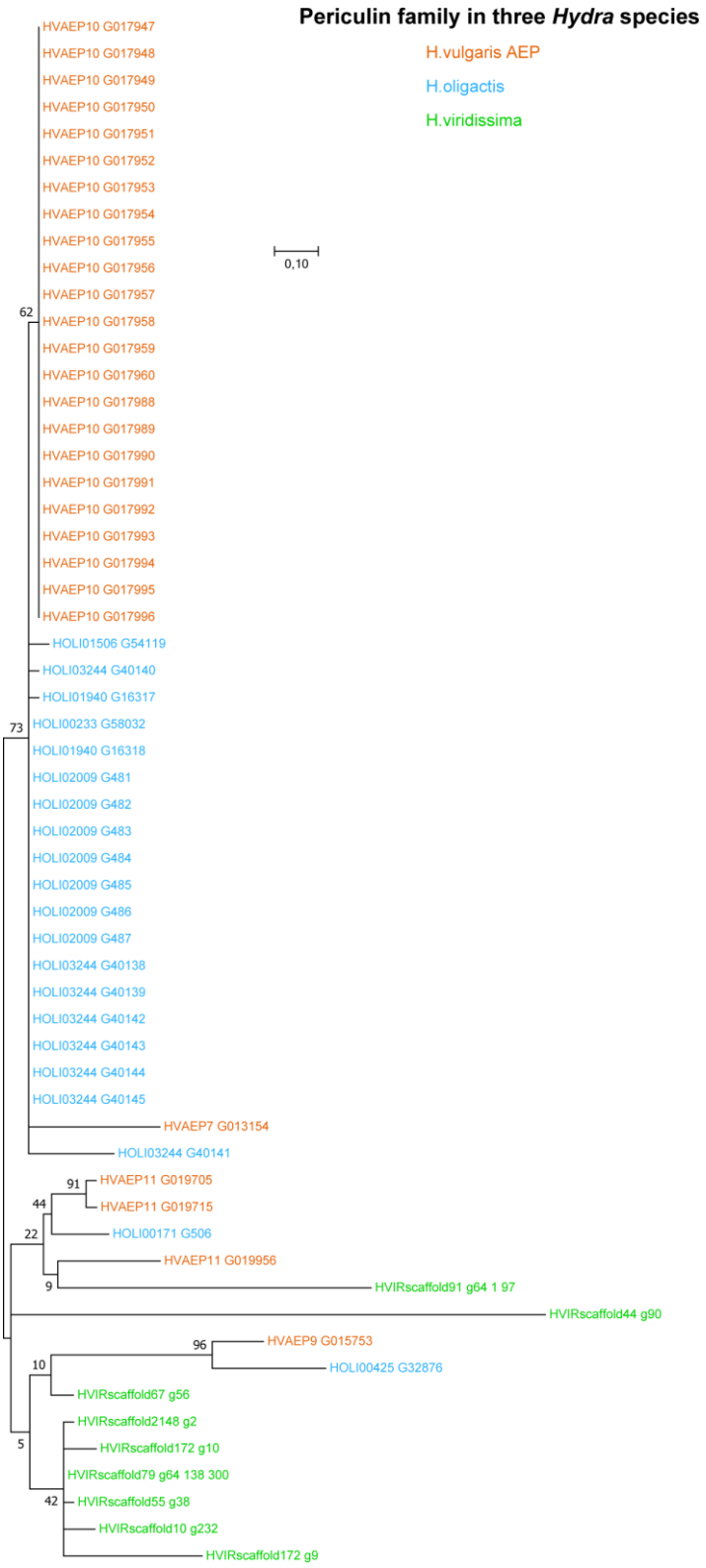
Supplementary Figures



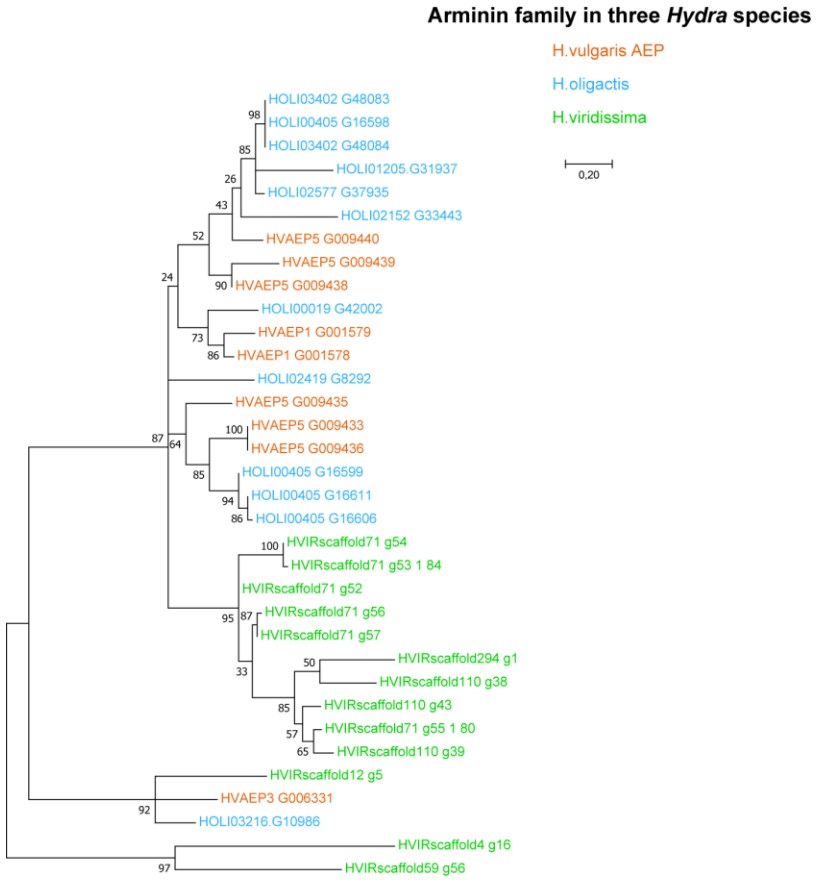
Supplementary Fig. 1. Nucleotide sequence alignment of 28 paralogues of the *periculin* family from the *H. vulgaris AEP* genome. Positions are colored by percentage of identity. Parts of alignment with extended 5´- and 3´- sequences of individual genes that could not be aligned are truncated.



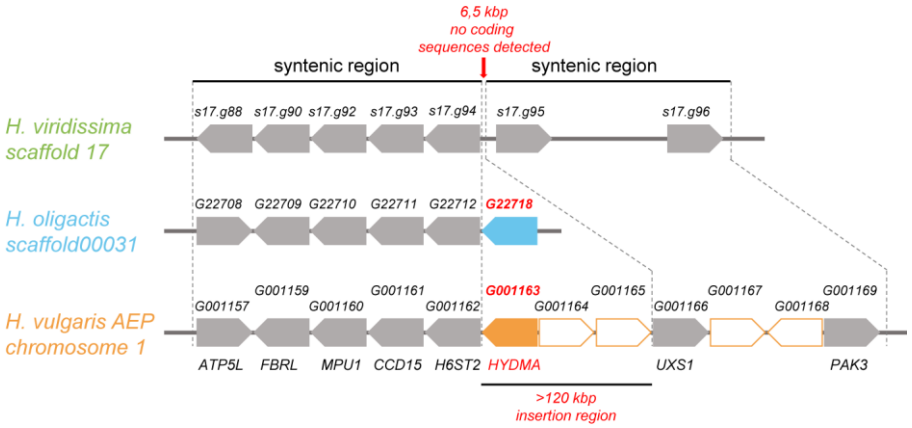
Supplementary Fig. 2. Amino acid sequence alignment of 28 paralogues of periculin family from *H. vulgaris AEP* genome. Positions are colored by percentage of identity. Note that some groups of paralogues (boxes) encode peptides with identical amino acid sequences.



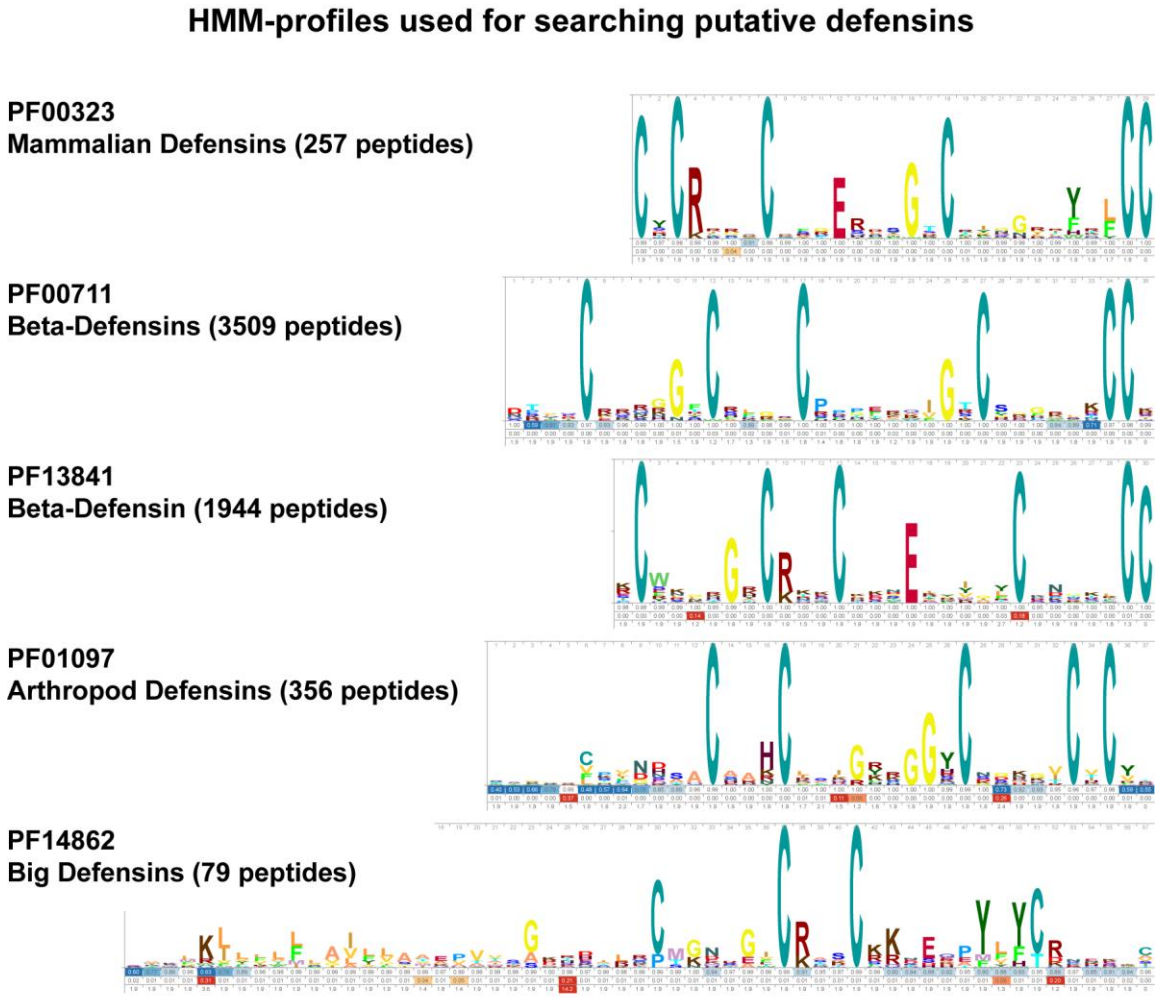
Supplementary Fig. 3. Phylogenetic analysis of periculin family orthologues from three *Hydra* species. The evolutionary history was inferred by comparing amino acid sequences using the Maximum Likelihood method based on the Jones-Taylor-Thornton with Gamma Distributed substitution rate (JTT+G) model (16) and 1,000 bootstrap iterations. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA7 (13).



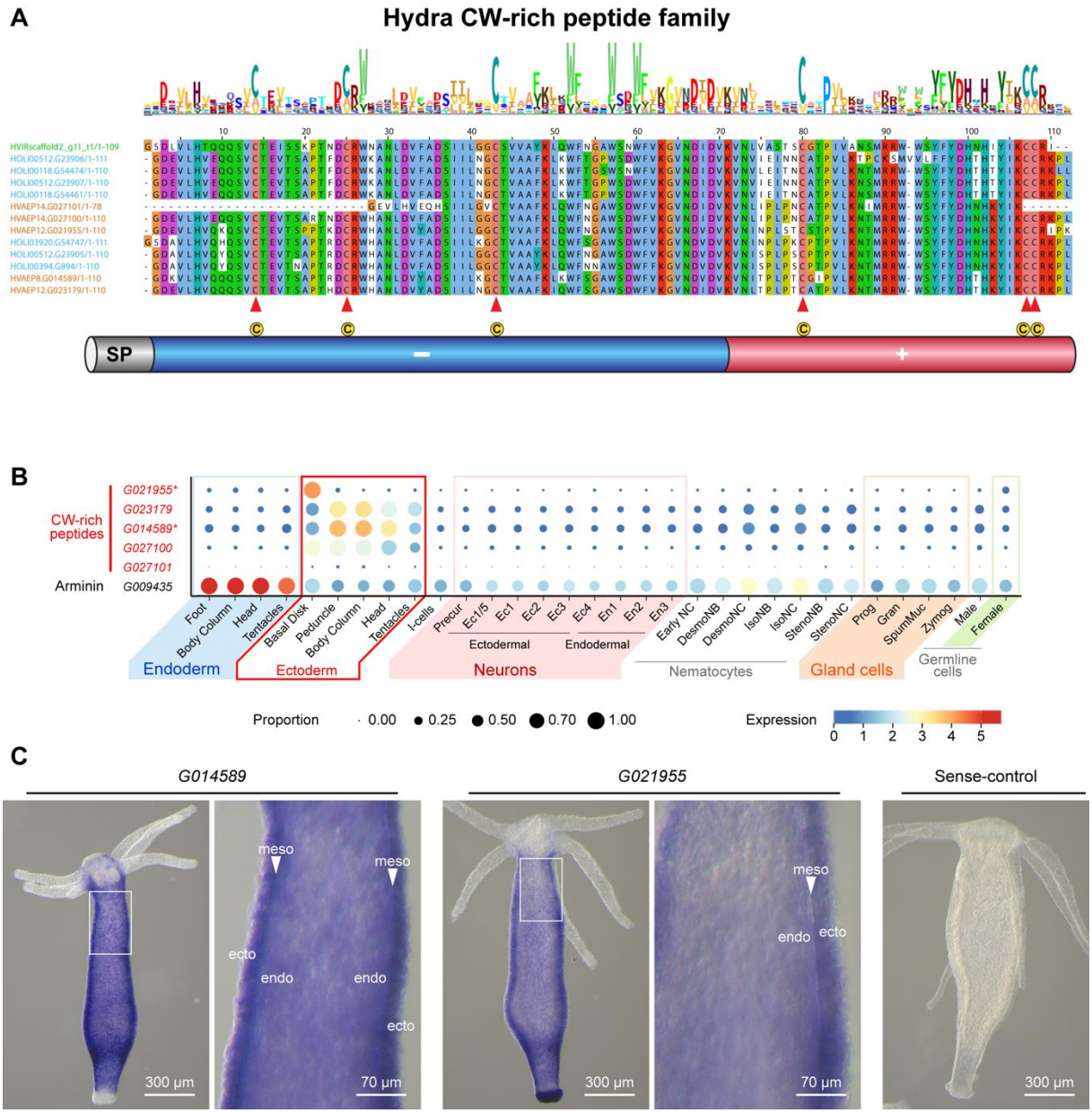
Supplementary Fig. 4. Phylogenetic analysis of arminin family orthologues from three *Hydra* species. The evolutionary history was inferred by comparing amino acid sequences using the Maximum Likelihood method based on the Jones-Taylor-Thornton with Gamma Distributed substitution rate (JTT+G) model (16) and 1,000 bootstrap iterations. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA7 (13).



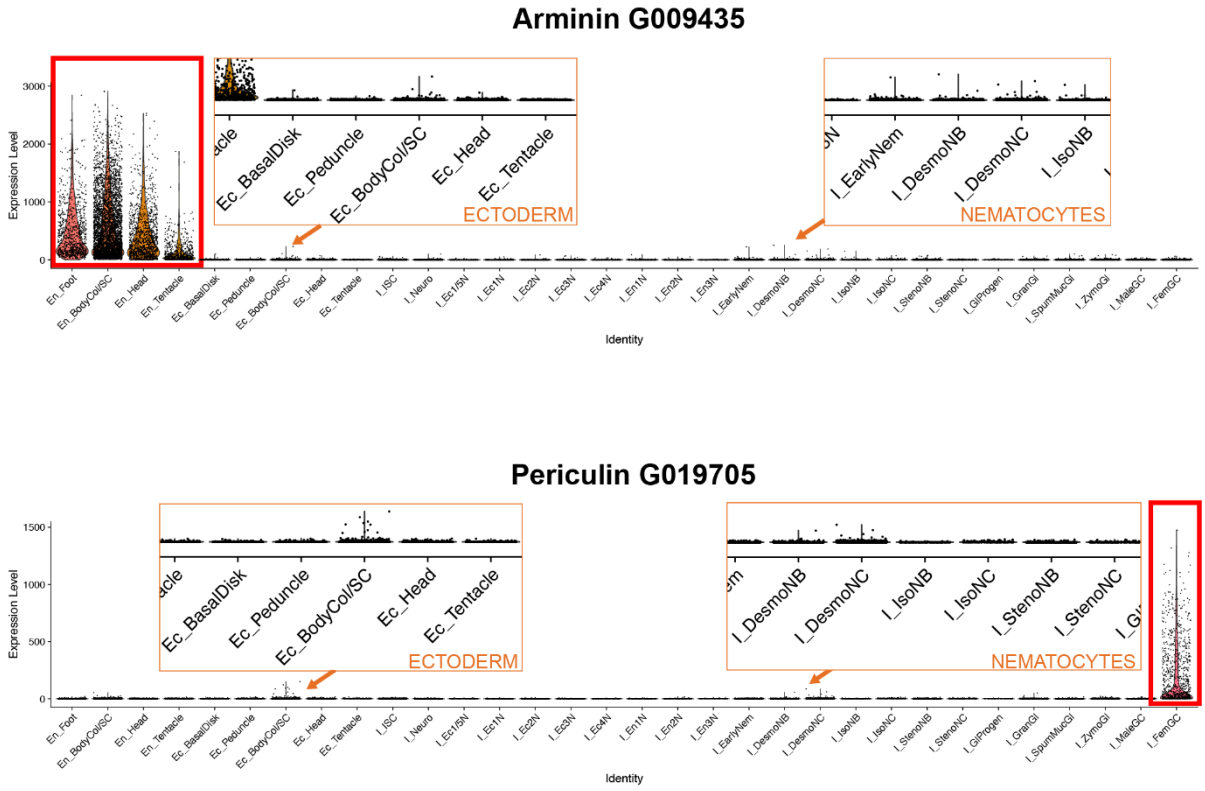
Supplementary Fig. 5: Synteny analysis supports absence of a *hydramacin* gene in *H. viridissima*. Schematic representation of predicted protein-coding genes in the syntenic loci of *H viridissima*, *H. oligactis* and *H. vulgaris AEP*. Almost identical order and orientation of 7 conserved genes (grey) indicates a syntenic genomic context. While in the chromosomes of *H. vulgaris AEP* and *H. oligactis* the *hydramacin* gene (orange and blue, respectively) is adjacent and collinear with the gene encoding heparan-sulfate 6-O-sulfotransferase 2 (H6ST2), no protein coding-sequence can be found in the similar location on the chromosome of *H. viridissima*. The substantial difference in the distance between two syntenic genes, H6ST2 and UXS1 (G001162 and G001166, 126 kbp in *H. vulgaris AEP*; s17.g94 and s17.g95, 6,5 kbp in *H. viridissima*) suggests that a major chromosomal rearrangement occurred along the evolution of the *hydramacin* gene's locus: a fragment of over 100 kbp was either lost (deletion) in *H. viridissima* or aquired in the brown *Hydra* lineage (insertion).



Supplementary Fig. 6: HMM profile logos of defensin families used to screen *Hydra* genomes for putative defensin-like peptides.



Supplementary Fig. 7: A family of cysteine and tryptophane rich secreted peptides in *Hydra* (HyCWR) demonstrates the closest similarity to canonical defensins. **A:** Alignment of HyCWR orthologues from *H. vulgaris AEP*, *H. oligactis* and *H. viridissima*. N-terminal signal peptide is removed from the alignment. ClustalX color scheme is used to emphasize the biased charge distribution: negatively charged amino acids (purple) are mostly present in the N-terminal half of the peptide, while the C-terminal part is rich in positively charged amino acids (red). Conserved Cys residues are highlighted by arrowheads. They demonstrate spacing somehow reminiscent of conventional defensins, yet not identical to any of the major families we used for HMMer analysis (Suppl. Fig. 6). **B:** Expression of genes encoding HyCWR peptides in cell types of *H. vulgaris AEP*. Note that HyCWR genes are expressed in the ectodermal cells, in contrast to all other AMP genes described to date. Endodermally-expressed *arminin* gene G009435 shown for comparison. Visualization is based on data from (2), gene expression is normalized and log-transformed. **C:** *In situ* hybridization with anti-sense probes against genes encoding two HyCWR members in *H. vulgaris AEP* (G014589 and G021955) provides evidence that both genes are strongly expressed in the polyps ectoderm (eco – ectoderm; meso – mesoglea; endo – endoderm). Hybridization with a sense probe (right) served as a control.



Supplementary Fig. 8 Raw expression data for two AMP genes – *arminin* G009435 and *periculin* G019705. It might appear from Fig. 3 that some nematocytes and ectodermal cells express *arminin* and *periculin* genes, in addition to the endodermal cells and female germline cells, respectively. Note that data on Fig. 3 are based on normalized and transformed values. Plotting the raw expression data clearly indicates that transcripts of both, *periculin* and *arminin* genes, are absent from the vast majority of ectodermal cells and nematocytes. Strongest and most consistent signal is observed in the endodermal epithelial cells - for *arminin* G009435, and in the female germline precursors - for the *periculin* G019705.

Supplementary Tables

Supplementary Table1: Repertoire of genes encoding AMPs in the genomes of three Hydra species - H.vulgaris AEP, H.oligactis, and H.viridissima

Note: Genes in the lists are sorted numerically, and occurrence of genes from different species in the same raw does not imply their similarity.

Family	H.vulgaris AEP (Cazet et al., 2023)	H.oligactis (Cazet et al., 2023)	H.viridissima (Hamada et al., 2020)
Periculin	**28**	**21**	**9**
	HVAEP10.G017947	HOLI00171.G506	HVIRscaffold10.g232.t1
	HVAEP10.G017948	HOLI00233.G58032	HVIRscaffold172.g10.t1
	HVAEP10.G017949	HOLI00425.G32876	HVIRscaffold172.g9.t1
	HVAEP10.G017950	HOLI01506.G54119	HVIRscaffold2148.g2.t1
	HVAEP10.G017951	HOLI01940.G16317	HVIRscaffold44.g90.t1
	HVAEP10.G017952	HOLI01940.G16318	HVIRscaffold55.g38.t1
	HVAEP10.G017953	HOLI02009.G481	HVIRscaffold67.g56.t1
	HVAEP10.G017954	HOLI02009.G482	HVIRscaffold79.g64.t1
	HVAEP10.G017955	HOLI02009.G483	HVIRscaffold91.g64.t1
	HVAEP10.G017956	HOLI02009.G484	
	HVAEP10.G017957	HOLI02009.G485	
	HVAEP10.G017958	HOLI02009.G486	
	HVAEP10.G017959	HOLI02009.G487	
	HVAEP10.G017960	HOLI03244.G40138	
	HVAEP10.G017988	HOLI03244.G40139	
	HVAEP10.G017989	HOLI03244.G40140	
	HVAEP10.G017990	HOLI03244.G40141	
	HVAEP10.G017991	HOLI03244.G40142	
	HVAEP10.G017992	HOLI03244.G40143	
	HVAEP10.G017993	HOLI03244.G40144	
	HVAEP10.G017994	HOLI03244.G40145	
	HVAEP10.G017995		
	HVAEP10.G017996		
	HVAEP11.G019705		
	HVAEP11.G019715		
	HVAEP11.G019956		
	HVAEP7.G013154		
	HVAEP9.G015753		
Arminin	**9**	**12**	**13**
	HVAEP1.G001578	HOLI00019.G42002	HVIRscaffold110.g38.t1
	HVAEP1.G001579	HOLI00405.G16598	HVIRscaffold110.g39.t1
	HVAEP3.G006331	HOLI00405.G16599	HVIRscaffold110.g43.t1
	HVAEP5.G009433	HOLI00405.G16606	HVIRscaffold12.g5.t1
	HVAEP5.G009435	HOLI00405.G16611	HVIRscaffold294.g1.t1
	HVAEP5.G009436	HOLI01205.G31937	HVIRscaffold4.g16.t1
	HVAEP5.G009438	HOLI02152.G33443	HVIRscaffold59.g56.t1
	HVAEP5.G009439	HOLI02419.G8292	HVIRscaffold71.g52.t1
	HVAEP5.G009440	HOLI02577.G37935	HVIRscaffold71.g53.t1
		HOLI03216.G10986	HVIRscaffold71.g54.t1
		HOLI03402.G48083	HVIRscaffold71.g55.t1
		HOLI03402.G48084	HVIRscaffold71.g56.t1
			HVIRscaffold71.g57.t1
Kazal	**7**	**8**	**2**
	HVAEP1.G002119	HOLI02750.G58520	HVIRscaffold13.g191.t1
	HVAEP1.G002120	HOLI02750.G58522	HVIRscaffold13.g192.t1
	HVAEP1.G002121	HOLI02750.G58524	
	HVAEP1.G002122	HOLI02750.G58525	
	HVAEP1.G002124	HOLI02750.G58526	
	HVAEP2.G003024	HOLI02896.G35283	
	HVAEP2.G003025	HOLI02896.G35284	
		HOLI02896.G35285	
		HOLI02896.G35287	
		HOLI02896.G35289	
Hydramacin	**1**	**2**	**0**
	HVAEP1.G001163	HOLI00031.G22718	
		HOLI00638.G147	
Hydralysin	**1**	**6**	**4**
	HVAEP5.G009585	HOLI00574.G60360	HVIRscaffold124.g25.t1
		HOLI01625.G60380	HVIRscaffold124.g27.t1
		HOLI00029.G34696	HVIRscaffold124.g24.t1
		HOLI00574.G60372	HVIRscaffold85.g6.t1
		HOLI00574.G60366	
		HOLI00574.G60361	
NDA-1	**1**	**1**	**1**
	HVAEP11.G019444	HOLI02108.G30666	HVIRscaffold14.g21.t1
Hym121	**1**	**1**	**1**
	HVAEP15.G028570	HOLI02720.G34319	HVIRscaffold59.g24.t1
Hym357	**5**	**8**	**1**
	HVAEP9.G016165	HOLI02613.G54832	HVIRscaffold26.g50.t1
	HVAEP9.G016169	HOLI02613.G54833	
	HVAEP9.G016170	HOLI02613.G54834	
	HVAEP9.G016171	HOLI02613.G54835	
	HVAEP9.G016173	HOLI04210.G34832	
		HOLI04210.G34833	
		HOLI04210.G34834	
		HOLI04210.G34835	
HyCWR family			
	HVAEP12.G021955	HOLI00118.G54461	HVIRscaffold2.g.11.t1
	HVAEP12.G023179	HOLI00118.G54474	
	HVAEP14.G027100	HOLI00394.G894	
	HVAEP14.G027101	HOLI00512.G23905	
	HVAEP8.G014589	HOLI00512.G23906	
		HOLI00512.G23907	
		HOLI03920.G54747	

Supplementary Table 1: Accession numbers of AMP genes from three *Hydra* species: *H. vulgaris AEP*, *H. oligactis* and *H. viridissima*. The Accession numbers correspond to the genes IDs from the original datasets (2) available at https://research.nhgri.nih.gov/HydraAEP/

- for *H. vulgaris AEP* and *H. oligactis*, and gene IDs from *H. viridissima* genome (4) https://marinegenomics.oist.jp/hydra_viridissima_a99/.

Supplementary Data

Suppl_Data_1.txt

Supplementary Data 1: Hidden Markov Model profiles of *Hydra* AMPs.

Suppl_Data_2.txt

Supplementary Data 2: Hidden Markov Model profiles of diverse defensin family AMPs.

References

1. Klimovich A, Giacomello S, Björklund Å, Faure L, Kaucka M, Giez C, et al. Prototypical pacemaker neurons interact with the resident microbiota. Proc Natl Acad Sci U S A. 2020 Jan 1;117(30):17854–63.

2. Cazet JF, Siebert S, Little HM, Bertemes P, Primack AS, Ladurner P, et al. A chromosome-scale epigenetic map of the Hydra genome reveals conserved regulators of cell state. Genome Res. 2023 Feb;33(2):283–98.

3. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res [Internet]. 2017 Oct 10;46(D1):D493–6. Available from: https://doi.org/10.1093/nar/gkx922

4. Hamada M, Satoh N, Khalturin K. A Reference Genome from the symbiotic hydrozoan, Hydra viridissima. G3 Genes, Genomes, Genet [Internet]. 2020 Jan 1;10(11):3883–3895. Available from: http://biorxiv.org/content/early/2020/05/29/2020.05.26.117382.abstract

5. Leclère L, Horin C, Chevalier S, Lapébie P, Dru P, Peron S, et al. The genome of the jellyfish Clytia hemisphaerica and the evolution of the cnidarian life-cycle. Nat Ecol Evol [Internet]. 2019;3(5):801–10. Available from: https://doi.org/10.1038/s41559-019-0833-2

6. Khalturin K, Shinzato C, Khalturina M, Hamada M, Fujie M, Koyanagi R, et al. Medusozoan genomes inform the evolution of the jellyfish body plan. Nat Ecol Evol [Internet]. 2019;3(5):811–22. Available from: https://doi.org/10.1038/s41559-019-0853-y

7. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Oct;7:539.

8. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46(W1):W200–4.

9. Wheeler TJ, Clements J, Finn RD. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics. 2014 Jan;15:7.

10. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. Science (80-) [Internet]. 2007 Jul 6;317(5834):86–94. Available from:

http://www.sciencemag.org/content/317/5834/86.abstract

11. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007 Nov;23(21):2947–8.

12. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res [Internet]. 1994 Nov 11;22(22):4673–80. Available from: https://doi.org/10.1093/nar/22.22.4673

13. Kumar S, Stecher G, Tamura K, Le SQ, Gascuel O. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.

14. Team RC. R: A language and environment for statistical computing. 2013;

15. Domazet-Lošo T, Klimovich A, Anokhin B, Anton-Erxleben F, Hamm MJMJ, Lange C, et al. Naturally occurring tumours in the basal metazoan Hydra. Nat Commun [Internet]. 2014;5:4222. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24957317

16. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992 Jun;8(3):275–82.