PHILOSOPHICAL TRANSACTIONS
OF THE ROYAL SOCIETY B
BIOLOGICAL SCIENCES

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

Journal:	*Philosophical Transactions B*
Manuscript ID	RSTB-2023-0058.R2
Article Type:	Research
Date Submitted by the Author:	16-Nov-2023
Complete List of Authors:	KLIMOVICH, Alexander; Christian-Albrechts-Universitat zu Kiel, Zoological Institute Bosch, Thomas; Christian-Albrechts University Kiel, Zoological Institute
Issue Code (this should have already been entered and appear below the blue box, but please contact the Editorial Office if it is not present):	SCULPT
Subject:	Developmental biology < BIOLOGY, Evolution < BIOLOGY, Bioinformatics < BIOLOGY
Keywords:	taxonomically-restricted genes, machine learning, scRNA-seq, holobiont

Author-supplied statements

Relevant information will appear here if provided.

Ethics

Does your article include research that required ethical approval or permits?:
This article does not present research with ethical considerations

Statement (if applicable):
CUST_IF_YES_ETHICS :No data available.

Data

It is a condition of publication that data, code and materials supporting your paper are made publicly available. Does your paper present new data?:
Yes

Statement (if applicable):
All data and references, such as Genbank accessions, are included in the manuscript and the supplementary material.

Conflict of interest

I/We declare we have no competing interests

Statement (if applicable):
CUST_STATE_CONFLICT :No data available.

Use of AI

Please provide a statement of any use of AI technology in the preparation of the paper.

No, we have not used AI-assisted technologies in creating this article
CUST_IF_YES_DECLARATION_OF_AI_USE :No data available.

Contribution to a special issue of Philosophical Transactions Royal Society B

"Sculpting the microbiome: How host factors determine and respond to microbial colonization."

RSTB-2023-0058.R1

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.)

Abstract

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.

Key words

Taxonomically-restricted genes, machine learning, scRNA-seq, holobiont

Key findings

- Novel technologies, including high-throughput sequencing and machine learning, allow for the discovery of a surprisingly high level of complexity in *Hydra*'s AMP repertoire.
- *Hydra* possesses a rich repertoire of AMPs encoded in taxonomically-restricted species-specific genes.
- *Hydra* uses a complex set of AMPs expressed in distinct tissue layers, cell lineages, and cell types to combat pathogens and maintain a stable spatially-organized microbiome.
- The tissue and cell type-specific expression patterns of *Hydra*'s AMPs highlight the importance of understanding the spatial organization of host innate immune responses and the microbiome.

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 (**Fig. 1A-C**), has long been used as a model organism for the study of the immune response evolution (3–5). Major components of the *Hydra* immune toolkit are highly conserved immune pathways mediated by Toll-like receptors (TLR) (4,6) and nucleotide-binding and oligomerisation domain-like receptors (NLR) (7). They are complemented by a rich repertoire of immune effector molecules – secreted AMPs. While the first AMP in *Hydra* has been discovered using traditional biochemical approaches (8), 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 (9–12). AMPs in *Hydra* share several common features: active AMPs are derived from larger precursors through a post-translational proteolytic cleavage of a signal peptide (**Fig. 1D**). Most 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 (13). This suggests that the taxonomically-restricted AMPs have evolved relatively recently in 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* (8,10,12,14,15). These observations lead 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 (1,16). 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 (10,11,17).

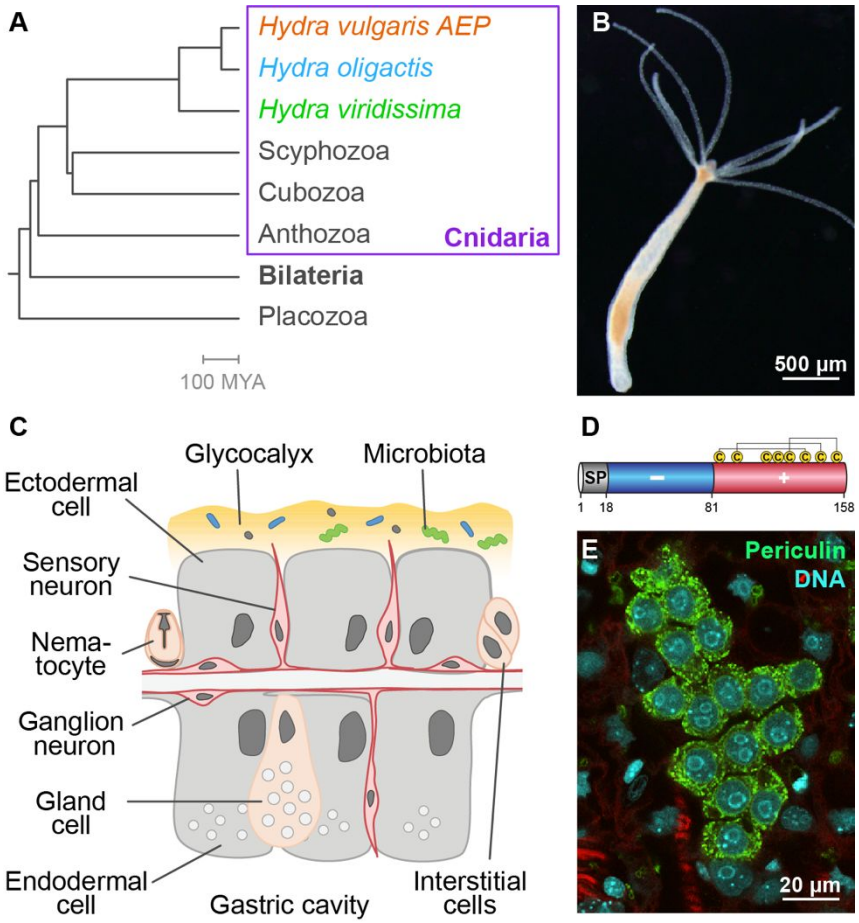

Figure 1: A: Phylogenetic tree demonstrating the position of *Hydra*. High-quality genome datasets have become recently available for three *Hydra* species – *H. vulgaris* AEP, *H. oligactis*, and *H. viridissima*. The divergence of the crown group *Hydra* took place about 193 MYA, and two species of brown hydras, *H. vulgaris* and *H. oligactis*, diverged over 100 MYA (18). **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.

Thus far, AMPs 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 AMPs repertoire. Further, we uncover 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 (8,19). Recent advancements in molecular biology techniques, such as expressed sequence tag analysis (EST) (8,9) and high-throughput transcriptome sequencing (RNAseq) (12,14), 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 **Fig. 1A**) 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 MYA of evolutionary radiation within the *Hydra* crown group (18). 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 AMPs gene families in the genome of *Hydra vulgaris strain AEP* (20) (see Suppl. Text for details, Suppl. 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 (**Fig. 2A,B; Suppl. Table 1**), in contrast to previously reported 5 *periculin* isoforms (12). Although the nucleotide sequences of these 28 paralogues were clearly different (**Suppl. Fig.1**), all these genes demonstrated similar exon-intron structure (**Fig. 2A**), and the amino acid sequence of peptides encoded by these genes were very remarkably similar (**Suppl. Fig.2**). Most intriguingly, numerous *periculin* paralogues were found clustered in a few genomic loci (**Fig. 2A**). For instance, in *H. vulgaris* AEP, two clusters on chromosome 10 contained 14 and 9 *periculin* paralogues, and the rest 5

paralogues were scattered among three other chromosomes. Very similar pattern was observed for other AMPs families. We identified a total of 9 *arminin* paralogues, 7 genes of the Kazal-like family, and 5 genes encoding Hym357-like neuropeptides with antimicrobial activity (**Fig. 2D-F; Suppl. Table 1**).

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 can not date back longer that 200 MYA, and the duplication might have occurred much more recently. The mechanisms that my have contributed to the rapid evolution of the AMP gene complement in the recent history of 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 Suppl. 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 (**Fig. 2B,E; Suppl. Fig. 3 and 4; Suppl. Table 1**). 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; **Fig. 2B,E; Suppl. Table 1**).

Second, the phylogenetic reconstruction uncovered that, in every *Hydra* species, AMP paralogues from each *Hydra* species tend to cluster together, forming species-specific clades (**Fig. 2B,E; Suppl. Fig. 3 and 4**). 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 (**Fig. 2B; Suppl. Fig. 3**), 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 (**Fig. 2B; Suppl. Fig. 3**).

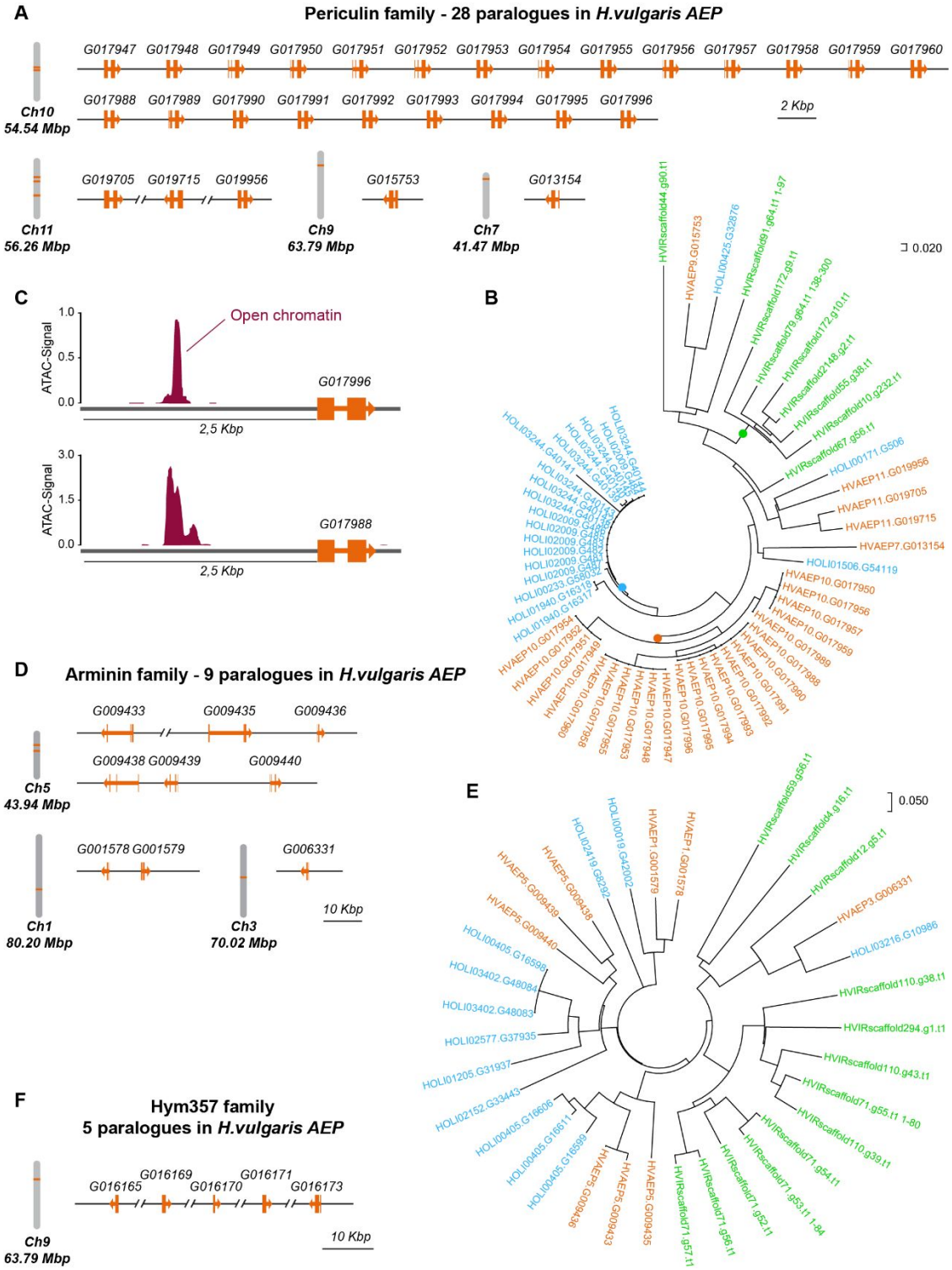

Figure 2: Complexity of AMP gene families in *Hydra* species. A: Twenty-eight paralogues of the *periculin* AMP family are located on 4 chromosomes of *H. vulgaris* AEP, whereby 23 genes form one dense cluster on chromosome 10. **B:** Phylogenetic analysis of *periculin* orthologues from three *Hydra* species. Genes are colored 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 Suppl. Fig. 3. **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 (20). **D:** Most of *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 Suppl. Fig. 4. **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 Suppl. Table 1.

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 (**Fig. 2B,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 3 *periculin* and 2 *arminin* paralogues (**Fig. 2B,E**). These gene families undergone 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 Suppl. Text), although two orthologues were confidently detected in the *H. oligactis* genome. Moreover, the synteny analysis (see Suppl. Text) identified only a non-coding sequence in the syntenic *H. viridissima* chromosomal region where the *hydramacin* orthologue would be anticipated (**Suppl. Fig. 5**). 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 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 cnadiarians 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 have emerged about 200 MYA 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 origination 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 are 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 Suppl. 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 (**Suppl. Fig. 6**). 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 6 cystein residues), as previously suggested (8). 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 6 Cys residues, likely linked into three disulphide bonds, yet the spacing between these residues is clearly different from that characteristic for defensins (**Suppl. Fig. 6,7A**). 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 (**Suppl. Fig. 7A**). We also note that the HyCWR genes represent a family of related genes, which comprises at least 5 orthologues in *H. vulgaris AEP*, 7 in *H. oligactis* and 1 in *H. viridissima* (**Suppl. Fig. 7A**; **Suppl. Table 1**), 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 display indeed 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 an 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 AMPs genes are constitutively transcribed at a very high level. For instance, *arminin* mRNAs were reported to be more abundant that *β-actin* transcripts (11). Similarly, *periculin* transcripts were among the most abundant transcripts in

female polyps (12,40). 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 (11), while *periculin* transcripts were rather restricted to the female germline precursor cells within the interstitial cell lineage (12,40) (**Fig. 1E**). More recently, several AMP genes with neuron-restricted expression were discovered (10,15), 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 (15,41), uncovered a high diversity of cell types in *Hydra*. For instance, 5 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 (15,41,42). 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 (**Fig. 3**). Indeed, the *hydramacin*, all *arminin* and most *kazal-like* genes are expressed exclusively in the endodermal epithelial cells (**Fig. 3; Suppl. Fig. 8**). 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 AMPs gene transcripts (**Fig. 3**). Our preliminary observations suggest that the genes encoding HyCWR peptides might be the only group of AMPs expressed in the ectodermal cells (**Suppl. Fig. 7B,C**). The cells of the interstitial lineage localized in the ectodermal layer (**Fig. 1C**), however, do express a variety of AMP genes. First, female germline precursor cells produce transcripts encoding the hydralysin, several periculin and Kazal-like peptides (**Fig. 3**). Neurons localized in the ectodermal and endodermal layers (**Fig. 1C**) express distinct sets of dual function peptides, such as Hym370, Hym176, RFamides (10) and Hym121 (15). Only one of these neuron-specific AMPs, the NDA-1, is produced by both, ectodermal and endodermal neurons.

The scRNA-seq data also provided insights into the spatial expression of AMPs genes along the body axis of *Hydra*. AMPs 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 (**Fig 3**). Ectodermally-expressed genes coding for putative HyCWR peptides show more distinct expression patterns, whereby one of them (G021955; **Suppl. Fig. 7B,C**), for instance, is strongly expressed in the basal disc, while another paralogue is not expressed in the foot at all (G0114589, **Suppl. Fig. 7B,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 polyps foot and tentacles (41,43). Intriguingly, since most of the neuronal populations are spatially restricted (15,41), 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, **Fig. 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 was studied in most details (12,40). Mature periculins are produced in the female germline cells (**Fig. 1E**) 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 embryos surface beneath the cuticle layer at early gastrulation stages (12). Additionally, a fusion protein periculin-GFP expressed in the ectodermal epithelial cells recapitulates the vesicular accumulation and release of the peptide on the surface (12). Similarly, with the help of specific antibodies, deposition of the neuronally-expressed peptide NDA-1 into the glycocalyx of *Hydra* has been also demonstrated (10). Therefore, the glycocalyx appears impregnated with diverse AMPs. Further proteome studies using mass spectrometry approaches, and particularly – the spatial proteomics (44), should provide a more comprehensive view of the AMP localization in diverse cells, tissues, and body compartments of *Hydra*.

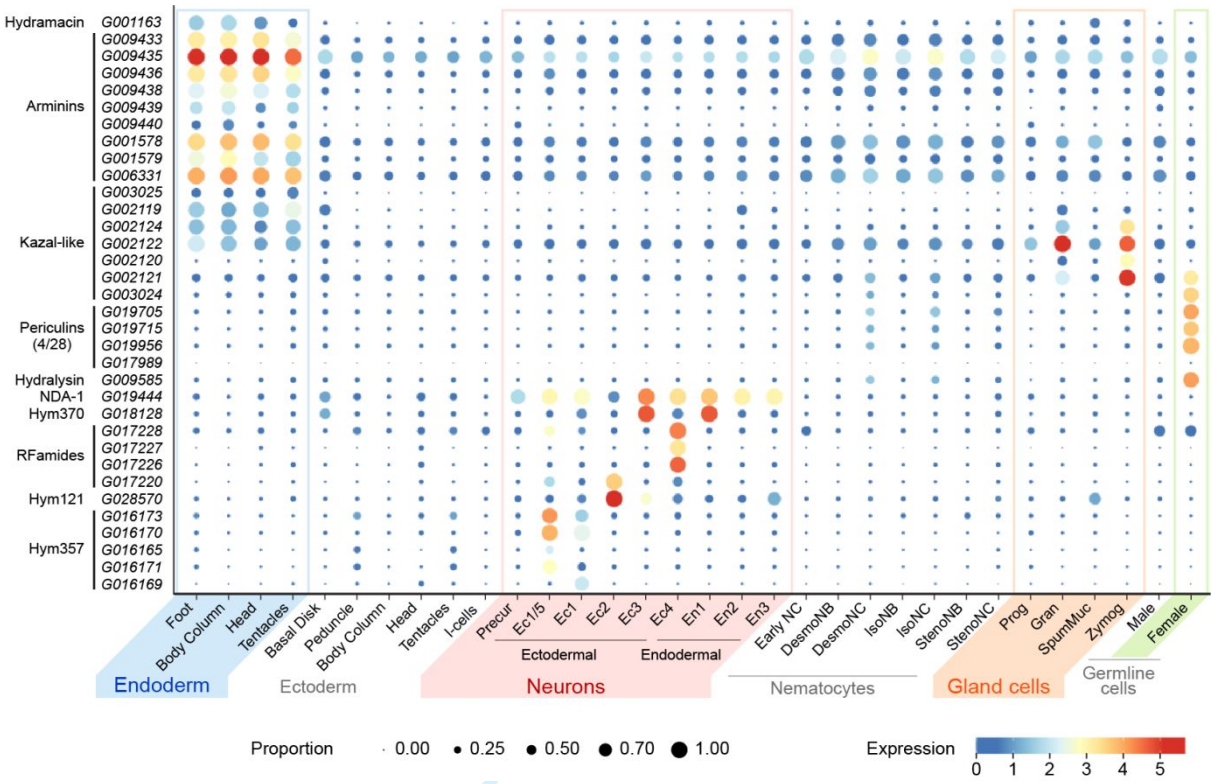

Figure 3: Expression of genes coding for AMPs across all cell types of *Hydra*. Visible are only the genes constitutively expressed in a *H. vulgaris* AEP polyp in homeostatic conditions, while inducible AMP genes expression is not illustrated. Note that only 4 out of 28 *periculin* paralogues are expressed at detectable levels. Visualization is based on data from (20), gene expression is normalized and log-transformed.

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 small proportion of cells (**Fig. 3**). This is consistent with earlier observations of Franzenburg and co-authors (11), 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 codons. Second, the paralogues show very similar exon-intron structure (particularly evident in case of *periculins*, **Fig. 2A**). 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 (**Fig. 2F**). Such a

pattern of ATAC-signal is characteristic for most *Hydra* promoters (20) and suggests that *cis*-regulatory elements upstream from the AMP genes 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 *Chapter 6.3*).

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 (15). 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 (**Fig. 4A**). These *hitherto* uncharacterized peptides, such as the product of a TRG *cluster131995* (**Fig. 4B**) 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 (15). 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 (58). 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 (59–61).

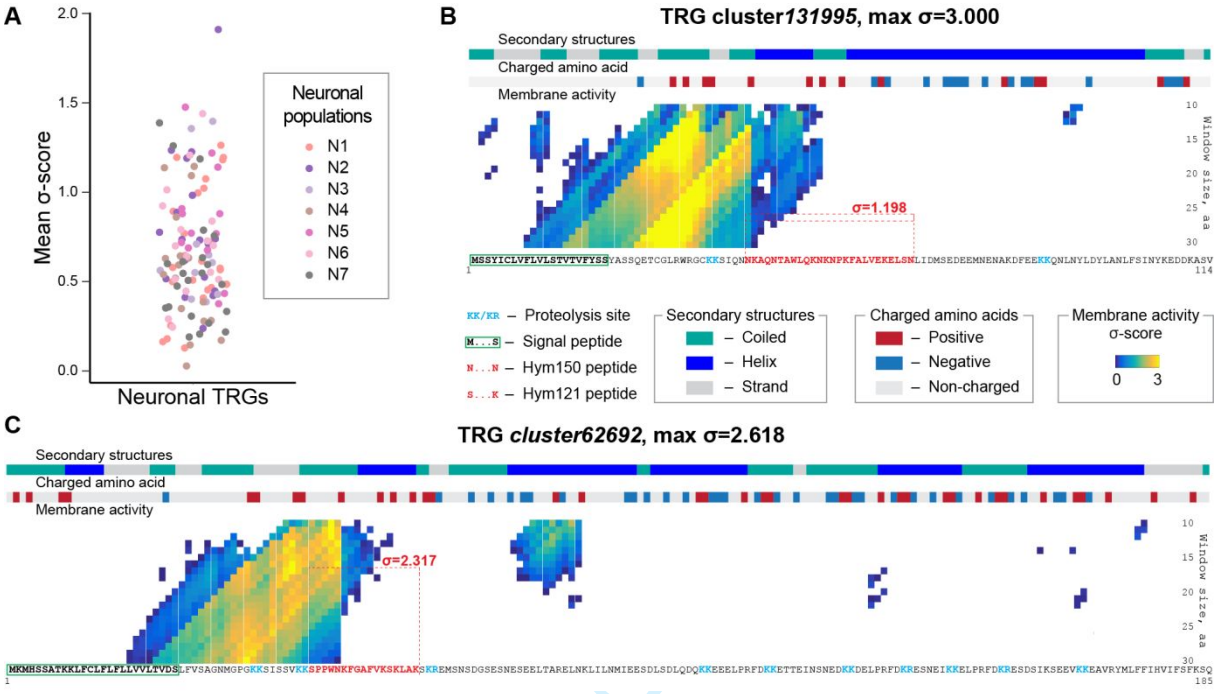

Figure 4: Machine learning algorithms allow for unbiased genome-wide prediction of putative AMPs. A: Distribution of mean σ-score values for individual secreted peptides encoded by neuron-specific TRGs in seven neuronal subpopulations illustrates a high likelihood of containing active AMPs for the peptides. Data are from (15). **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 (62) 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* (15).

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

5.1 Expansion of AMPs 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 (11,63). 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* (**Fig. 5A**) radiation about 190 MYA (18) remain unclear. It is, however, plausible that the transition of a hydra ancestor from the marine into 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 became inefficient (*e.g.* defensins are generally known to be highly effective in salt water 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 AMPs families (**Fig. 5A**). While comparing different *Hydra* species, one interesting tendency became obvious: the size of AMPs families was generally larger in the representatives of the “brown hydra” group compared to the green hydra *H. viridissima* (**Fig. 2** and **5**). For instance, only 8 *periculin* genes and a single *Hym357* orthologue were found in *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 MYA (18). 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 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 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* (4,6). Genes coding for putative NOD-like receptors, on the contrary, have undergone expansion in *Hydra* indeed (7). Intriguingly, the broadest repertoire of genes encoding NACHT- and 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 6 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.

5.2 AMPs shape the spatiotemporal structure of the Hydra microbiome

Since 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 (see contribution by Bosch et al in this issue). The *Hydra* host imposes strong selective forces on its microbiome via section of diverse AMPs (71) and thereby maintains species-specific microbiota communities over extended periods (11,63,72). 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 (**Fig. 5B**). A plethora of AMPs are expressed throughout the endoderm of *Hydra* along the entire body (**Fig. 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 and are, hence, confined to certain body domains (**Fig. 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;(73)) 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 (74). On the contrary, other members of the microbiome such as *Curvibacter* (10,75), 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 (**Fig. 5B-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 (10). 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 (10).

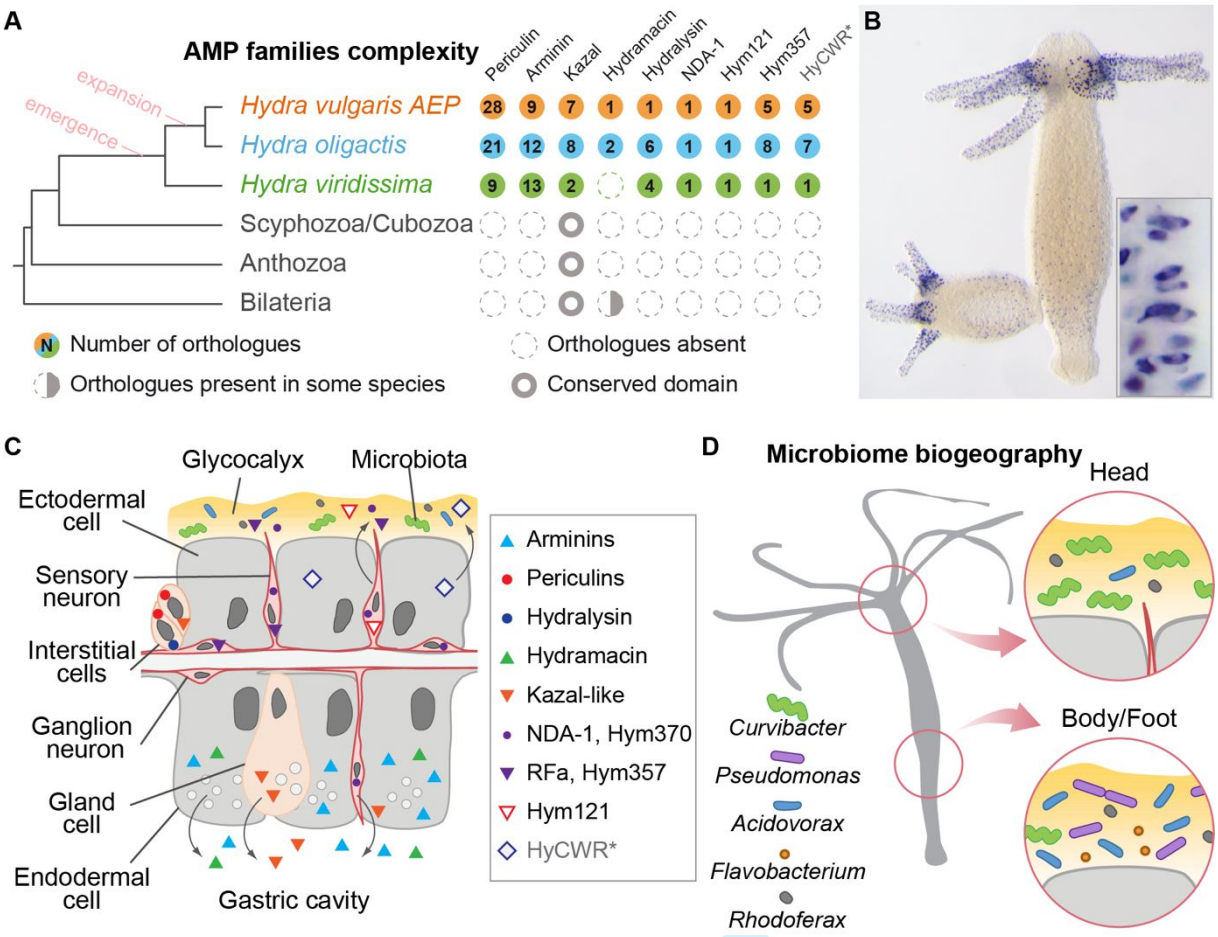

Figure 5: Complex species-specific and spatially restricted cocktails of AMPs sculpt the microbiome of *Hydra*. A: Overview of the AMP gene family complexity in *Hydra* species. Note that *H. viridissima* possesses generally less AMP genes compared to *H. oligactis* and *H. vulgaris AEP*. Most of AMP families are restricted to *Hydra* and only in few cases proteins with similar domains can 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 *Hydra* microbiome – its biogeography. * - the antimicrobial activity of HyCWR peptides and their role in the *Hydra* holobiont remain to be validated.

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.

6.1 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 animal kingdom (49,76,77), and most animals indeed possess a broad array of AMPs. The clustered genomic organization of AMP genes has been also recognized characteristic for numerous AMP families across animal species (78–82). However, we find it truly puzzling that numerous paralogues of AMP genes in *Hydra* have though slightly different nucleotide sequences, code for identical precursor polypeptides and, hence, give rise to identical active peptides. This is particularly evident in case of the periculin family (**Fig. 2B**; **Suppl. Fig. 2**). 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 ant fungi (64,77,83,84), these findings support the view that elaboration of AMP arsenal through novel family emergence, gene duplication and diversitifcation is a common, universal principle in AMP genes evolution.

6.2 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 novel tools also allow unbiased genome-wide screening and *ab initio* detection on AMPs. Our preliminary analysis suggests that dozens of novel, previously not characterized

AMPs and their families are still hidden in the genome unrecognized (**Fig. 4A**). This hypothesis is supported by our finding that clusters of tandemly-repeated *Hydra*-specific TRGs, architecturally similar to, for instance, the *periculin* cluster (**Fig. 2A**), 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, favored 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 3D protein modelling, such as the AlphaFold and similar template-independent tools (85–87), 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 bottle neck. Not all peptides can be synthetized effectively in their active form and tested *in vitro*, and recombinant expression may be also challenging to due 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 “educated guess” and selecting candidates for in-depth validation.

The majority of AMPs in *Hydra* are encoded in *Hydra*-restricted TRGs (**Fig. 5A**), 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* (**Suppl. Fig. 5**). 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 mollusks (4,8,88,89). Our synteny analysis (**Suppl. Fig. 5**) 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 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 (34,76,90–92), supporting the high evolutionary dynamics of AMPs 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.

6.3 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 to be not expressed in homeostatic conditions (**Fig. 3; Suppl. Fig. 7**). 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 (12). 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) (4). Moreover, interference with the upstream signalling pathways (17), and tissue manipulations, such as amputation-induced regeneration (93) and elimination of neurons (94,95), also result in modulation of AMP gene expression). This may cause a concomitant enhanced antimicrobial activity of the tissue (96). However, it remains unclear, whether the transcription of already expressed paralogues is elevated, or additional previously silent AMP genes (**Fig. 2C** and **3**) are turned on. A particularly exciting possibility is that ectodermal cells that do not produce any AMPs in homeostatic condition (**Fig. 3** and **5C**), start expressing certain AMP genes following developmental or environmental signals.

6.4 For AMPs the name no longer fits the function

As outlined in detail in another paper in this issue (Bosch, Blaser, Ruby and McFall-Ngai), 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 (97), or act as immunomodulators (98). 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 (8–10). 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 (15). Therefore, this AMP actually acts as a signaling molecule (somehow reminiscent of the signaling role of microbe-derived antibiotics, (99)). Similar observations remain very scarce, and no systematic survey of non-conventional roles of *Hydra* AMPs has been performed. Since MIC assays have been main tools to test AMPs activity and infer function, behavior-modulating aspects of AMPs activity have escaped detection so far. We emphasize the urgent need to develop and implement novel methods, such as motility assays and microcosm setups (73), 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 (56,100), 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 (56,101–103)

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 (16,104–106). They enrich our understanding of the critical factors maintaining the metaorganism homeostasis and health across the animal kingdom.

Acknowledgements

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 TCGB), 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 KL3475/2-1. T.C.G.B. appreciates support from the Canadian Institute for Advanced Research. 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.

Data accessibility

The sequence datasets supporting this article have been uploaded as part of the electronic supplementary material – Supplementary Table 1 and Supplementary Data 1 and 2.

Author’s contribution

A.K.: conceptualization, data curation, formal analysis, funding acquisition, investigation, visualization, writing original draft and revised manuscript; T.C.G.B.: conceptualization, funding acquisition, writing original draft.

Conflict of interest declaration

We have no competing interests to declare.

References

1. Bosch TCG, Zasloff M. Antimicrobial Peptides—or How Our Ancestors Learned to Control the Microbiome. MBio [Internet]. 2021 May 15;12(5):e01847-21. Available from: https://doi.org/10.1128/mBio.01847-21

2. Zasloff M. Antimicrobial peptides of multicellular organisms. Nature [Internet]. 2002;415(6870):389–95. Available from: https://doi.org/10.1038/415389a

3. Augustin R, Fraune S, Franzenburg S, Bosch TCG. Where simplicity meets complexity: hydra, a model for host–microbe interactions. In: Recent Advances on Model Hosts. Springer; 2012. p. 71–81.

4. Bosch TCG, Augustin R, Anton-Erxleben F, Fraune S, Hemmrich G, Zill H, et al. Uncovering the evolutionary history of innate immunity: the simple metazoan Hydra uses epithelial cells for host defence. Dev Comp Immunol. 2009;33(4):559–69.

5. Klimovich A V., Bosch TCG. Rethinking the Role of the Nervous System: Lessons From the Hydra Holobiont. BioEssays. 2018;40(9):1800060.

6. Franzenburg S, Fraune S, Kunzel S, Baines JF, Domazet-Loso T, Bosch TCG. MyD88-deficient Hydra reveal an ancient function of TLR signaling in sensing bacterial colonizers. Proc Natl Acad Sci [Internet]. 2012;109(47):19374–9. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1213110109

7. Lange C, Hemmrich G, Klostermeier UC, López-Quintero JA, Miller DJ, Rahn T, et al. Defining the origins of the NOD-like receptor system at the base of animal evolution. Mol Biol Evol. 2010;28(5):1687–702.

8. Jung S, Dingley AJ, Augustin R, Anton-Erxleben F, Stanisak M, Gelhaus C, et al. Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan Hydra. J Biol Chem. 2009;284(3):1896–905.

9. Augustin R, Siebert S, Bosch TCG. Identification of a kazal-type serine protease inhibitor with potent anti-staphylococcal activity as part of Hydra's innate immune system. Dev Comp Immunol [Internet]. 2009;33(7):830–7. Available from: http://www.sciencedirect.com/science/article/pii/S0145305X0900038X

10. Augustin R, Schröder K, Murillo Rincón AP, Fraune S, Anton-Erxleben F, Herbst E-MM, et al. A secreted antibacterial neuropeptide shapes the microbiome of Hydra. Nat Commun [Internet]. 2017;8(1):698. Available from: https://doi.org/10.1038/s41467-017-00625-1

11. Franzenburg S, Walter J, Künzel S, Wang J, Baines JF, Bosch TCG, et al. Distinct antimicrobial peptide expression determines host species-specific bacterial associations. Proc Natl Acad Sci. 2013;110(39):E3730–8.

12. Fraune S, Augustin R, Anton-Erxleben F, Wittlieb J, Gelhaus C, Klimovich VB, et al. In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. Proc Natl Acad Sci U S A [Internet]. 2010 Oct 19 [cited 2013 Jan 29];107(42):18067–72. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2964230&tool=pmcentrez&rendertype=abstract

13. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. More than just orphans: are taxonomically-restricted genes important in evolution? Trends Genet. 2009;25(9):404–13.

14. Augustin R, Anton-Erxleben F, Jungnickel S, Hemmrich G, Spudy B, Podschun R, et al. Activity of the novel peptide arminin against multiresistant human pathogens shows the considerable potential of phylogenetically ancient organisms as drug sources.

Antimicrob Agents Chemother. 2009;53(12):5245–50.

15. 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.

16. Bosch TCG. Understanding complex host-microbe interactions in Hydra. Gut Microbes [Internet]. 2012;3(4):345–51. Available from: http://europepmc.org/abstract/MED/22688725

17. Mortzfeld BM, Taubenheim J, Fraune S, Klimovich A V, Bosch TCG. Stem Cell Transcription Factor FoxO Controls Microbiome Resilience in Hydra. Front Microbiol [Internet]. 2018;9:629. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2018.00629

18. Schwentner M, Bosch TCG. Revisiting the age, evolutionary history and species level diversity of the genus Hydra (Cnidaria: Hydrozoa). Mol Phylogenet Evol. 2015;91:41–55.

19. Sher D, Fishman Y, Zhang M, Lebendiker M, Gaathon A, Mancheño J-M, et al. Hydralysins, a New Category of β-Pore-forming Toxins in Cnidaria*. J Biol Chem [Internet]. 2005;280(24):22847–55. Available from: https://www.sciencedirect.com/science/article/pii/S0021925820614545

20. 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.

21. 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

22. 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

23. Kim H-M, Weber JA, Lee N, Park SG, Cho YS, Bhak Y, et al. The genome of the giant Nomura's jellyfish sheds light on the early evolution of active predation. BMC Biol [Internet]. 2019 Dec 29 [cited 2019 Sep 12];17(1):28. Available from: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-019-0643-7

24. Kon-Nanjo K, Kon T, Horkan HR, Steele RE, Cartwright P, Frank U, et al. Chromosome-level genome assembly of Hydractinia symbiolongicarpus. G3 Genes|Genomes|Genetics [Internet]. 2023 May 18;jkad107. Available from: https://doi.org/10.1093/g3journal/jkad107

25. Santander MD, Maronna MM, Ryan JF, Andrade SCS. The state of Medusozoa genomics: current evidence and future challenges. Gigascience [Internet]. 2022 Jan 1;11:giac036. Available from: https://doi.org/10.1093/gigascience/giac036

26. Ying H, Hayward DC, Cooke I, Wang W, Moya A, Siemering KR, et al. The Whole-Genome Sequence of the Coral Acropora millepora. Genome Biol Evol [Internet]. 2019 Apr 26;11(5):1374–9. Available from: https://doi.org/10.1093/gbe/evz077

27. Klimovich A, Wittlieb J, Bosch TCGTCG. Transgenesis in Hydra to characterize gene function and visualize cell behavior. Nat Protoc [Internet]. 2019;14(7):2069–90. Available from: https://doi.org/10.1038/s41596-019-0173-3

28. Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TCG. Transgenic Hydra

allow in vivo tracking of individual stem cells during morphogenesis. Proc Natl Acad Sci U S A [Internet]. 2006 Apr 18;103(16):6208–11. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1458856&tool=pmcentrez&rendertype=abstract

29. Neme R, Tautz D. Evolution: dynamics of de novo gene emergence. Curr Biol. 2014;24(6):R238–40.

30. Rödelsperger C, Prabh N, Sommer RJ. New Gene Origin and Deep Taxon Phylogenomics: Opportunities and Challenges. Trends in Genetics. 2019.

31. Tomasinsig L, Zanetti M. The cathelicidins-structure, function and evolution. Curr Protein Pept Sci. 2005;6(1):23–34.

32. Hao X, Yang H, Wei L, Yang S, Zhu W, Ma D, et al. Amphibian cathelicidin fills the evolutionary gap of cathelicidin in vertebrate. Amino Acids. 2012;43:677–85.

33. Uzzell T, Stolzenberg ED, Shinnar AE, Zasloff M. Hagfish intestinal antimicrobial peptides are ancient cathelicidins. Peptides [Internet]. 2003;24(11):1655–67. Available from: https://www.sciencedirect.com/science/article/pii/S0196978103003401

34. Hanson MA, Lemaitre B. New insights on Drosophila antimicrobial peptide function in host defense and beyond. Curr Opin Immunol [Internet]. 2020;62:22–30. Available from: https://www.sciencedirect.com/science/article/pii/S0952791519300937

35. Ganz T. Defensins: antimicrobial peptides of innate immunity. Nat Rev Immunol. 2003;3(9):710–20.

36. Leal E, Múnera M, Suescún-Bolívar LP. In silico characterization of Cnidarian's antimicrobial peptides [Internet]. Vol. 9, Frontiers in Marine Science . 2022. Available from: https://www.frontiersin.org/articles/10.3389/fmars.2022.1065717

37. Mitchell ML, Shafee T, Papenfuss AT, Norton RS. Evolution of cnidarian trans-defensins: Sequence, structure and exploration of chemical space. Proteins Struct Funct Bioinforma. 2019 Jul;87(7):551–60.

38. Fu J, He Y, Peng C, Tang T, Jin A, Liao Y, et al. Transcriptome Sequencing of the Pale Anemones (Exaiptasia diaphana) Revealed Functional Peptide Gene Resources of Sea Anemone. Front Mar Sci. 2022;9:856501.

39. Ovchinnikova T V, Balandin S V, Aleshina GM, Tagaev AA, Leonova YF, Krasnodembsky ED, et al. Aurelin, a novel antimicrobial peptide from jellyfish Aurelia aurita with structural features of defensins and channel-blocking toxins. Biochem Biophys Res Commun. 2006;348(2):514–23.

40. 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

41. Siebert S, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, et al. Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. Science (80-). 2019 Jul 26;365(6451):eaav9314.

42. Primack AS, Cazet JF, Little HM, Mühlbauer S, Cox BD, David CN, et al. Differentiation trajectories of the Hydra nervous system reveal transcriptional regulators of neuronal fate. bioRxiv [Internet]. 2023 Jan 1;2023.03.15.531610. Available from: http://biorxiv.org/content/early/2023/04/04/2023.03.15.531610.abstract

43. Siebert S, Anton-Erxleben F, Bosch TCG. Cell type complexity in the basal metazoan Hydra is maintained by both stem cell based mechanisms and transdifferentiation. Dev Biol [Internet]. 2008 Jan 1 [cited 2013 Jan 29];313(1):13–24. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/18029279

44. Mund A, Brunner A-D, Mann M. Unbiased spatial proteomics with single-cell resolution in tissues. Mol Cell. 2022;82(12):2335–49.

45. Pirtskhalava M, Amstrong AA, Grigolava M, Chubinidze M, Alimbarashvili E, Vishnepolsky B, et al. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. Nucleic Acids Res. 2021;49(D1):D288–97.

46. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res. 2016;44(D1):D1087–93.

47. Gómez EA, Giraldo P, Orduz S. InverPep: a database of invertebrate antimicrobial peptides. J Glob Antimicrob Resist. 2017;8:13–7.

48. Witten J, Witten Z. Deep learning regression model for antimicrobial peptide design. bioRxiv [Internet]. 2019 Jan 1;692681. Available from: http://biorxiv.org/content/early/2019/07/12/692681.abstract

49. Lazzaro BP, Zasloff M, Rolff J. Antimicrobial peptides: Application informed by evolution. Science (80-) [Internet]. 2020 May 1;368(6490):eaau5480. Available from: https://doi.org/10.1126/science.aau5480

50. Aguilera-Puga M d. C, Cancelarich NL, Marani MM, de la Fuente-Nunez C, Plisson F. Accelerating the Discovery and Design of Antimicrobial Peptides with Artificial Intelligence BT - Computational Drug Discovery and Design. In: Gore M, Jagtap UB, editors. New York, NY: Springer US; 2024. p. 329–52. Available from: https://doi.org/10.1007/978-1-0716-3441-7_18

51. Huang J, Xu Y, Xue Y, Huang Y, Li X, Chen X, et al. Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. Nat Biomed Eng [Internet]. 2023;7(6):797–810. Available from: https://doi.org/10.1038/s41551-022-00991-2

52. Lata S, Sharma BK, Raghava GPS. Analysis and prediction of antibacterial peptides. BMC Bioinformatics. 2007;8(1):1–10.

53. Porto WF, Pires ÁS, Franco OL. CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides. PLoS One. 2012;7(12):e51444.

54. Torrent M, Andreu D, Nogués VM, Boix E. Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. PLoS One. 2011;6(2):e16968.

55. Fjell CD, Jenssen H, Hilpert K, Cheung WA, Pante N, Hancock REW, et al. Identification of novel antibacterial peptides by chemoinformatics and machine learning. J Med Chem. 2009;52(7):2006–15.

56. Wong F, de la Fuente-Nunez C, Collins JJ. Leveraging artificial intelligence in the fight against infectious diseases. Science (80-) [Internet]. 2023 Jul 14;381(6654):164–70. Available from: https://doi.org/10.1126/science.adh1114

57. Lee EY, Fulan BM, Wong GCL, Ferguson AL. Mapping membrane activity in undiscovered peptide sequence space using machine learning. Proc Natl Acad Sci [Internet]. 2016 Nov 10;113(48):13588–93. Available from: http://www.pnas.org/content/early/2016/11/09/1609893113.abstract

58. Lee EY, Chan LC, Wang H, Lieng J, Hung M, Srinivasan Y, et al. PACAP is a pathogen-inducible resident antimicrobial neuropeptide affording rapid and contextual molecular host defense of the brain. Proc Natl Acad Sci U S A. 2021;

59. Lee EY, Srinivasan Y, de Anda J, Nicastro LK, Tükel Ç, Wong GCL. Functional

Reciprocity of Amyloids and Antimicrobial Peptides: Rethinking the Role of Supramolecular Assembly in Host Defense, Immune Activation, and Inflammation. Frontiers in Immunology. 2020.

60. Schluesener HJ, Su Y, Ebrahimi A, Pouladsaz D. Antimicrobial peptides in the brain: Neuropeptides and amyloid. Front Biosci. 2012;4(4):1375–80.

61. Abbott A. Are infections seeding some cases of Alzheimer's disease? Nature. 2020;

62. Fujisawa T. Hydra peptide project 1993-2007. Dev Growth Differ [Internet]. 2008 Jun [cited 2011 May 9];50 Suppl 1:S257-68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18459981

63. Fraune S, Bosch TCG. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra Proc Natl Acad Sci [Internet]. 2007 Aug 7;104(32):13146 LP – 13151. Available from: http://www.pnas.org/content/104/32/13146.abstract

64. Gerdol M, Schmitt P, Venier P, Rocha G, Rosa RD, Destoumieux-Garzón D. Functional insights from the evolutionary diversification of big defensins. Front Immunol. 2020;11:758.

65. Hamada M, Schröder K, Bathia J, Kürn U, Fraune S, Khalturina M, et al. Metabolic co-dependence drives the evolutionarily ancient Hydra–Chlorella symbiosis. Elife. 2018;7:e35122.

66. Bathia J, Schröder K, Fraune S, Lachnit T, Rosenstiel P, Bosch TCG. Symbiotic Algae of Hydra viridissima Play a Key Role in Maintaining Homeostatic Bacterial Colonization [Internet]. Vol. 13, Frontiers in Microbiology . 2022. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2022.869666

67. van Oppen MJH, Blackall LL. Coral microbiome dynamics, functions and design in a changing world. Nat Rev Microbiol [Internet]. 2019;17(9):557–67. Available from: https://doi.org/10.1038/s41579-019-0223-4

68. Kayal E, Bentlage B, Sabrina Pankey M, Ohdera AH, Medina M, Plachetzki DC, et al. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. BMC Evol Biol [Internet]. 2018;18(1):68. Available from: https://doi.org/10.1186/s12862-018-1142-0

69. Hanson MA, Grollmus L, Lemaitre B. Ecology-relevant bacteria drive the evolution of host antimicrobial peptides in Drosophila. Science (80-). 2023;381(6655):eadg5725.

70. Brown KL, Hancock REW. Cationic host defense (antimicrobial) peptides. Curr Opin Immunol [Internet]. 2006;18(1):24–30. Available from: https://www.sciencedirect.com/science/article/pii/S0952791505001998

71. Deines P, Lachnit T, Bosch TCG. Competing forces maintain the Hydra metaorganism. Immunol Rev. 2017;279(1):123–36.

72. Taubenheim J, Miklós M, Tökölyi J, Fraune S. Population differences and host species predict variation in the diversity of host-associated microbes in hydra. Front Microbiol. 2022;13.

73. Deines P, Hammerschmidt K, Bosch TCG. Microbial Species Coexistence Depends on the Host Environment. MBio [Internet]. 2020 Jul 21;11(4):10.1128/mbio.00807-20. Available from: https://doi.org/10.1128/mbio.00807-20

74. Nawroth JC, Giez C, Klimovich A, Kanso E, Bosch TCG. Spontaneous body wall contractions stabilize the fluid microenvironment that shapes host-microbe associations. Goldstein RE, editor. Elife [Internet]. 2023;12:e83637. Available from:

https://doi.org/10.7554/eLife.83637

75. Giez C, Pinkle D, Giencke Y, Wittlieb J, Herbst E, Spratte T, et al. Microbes as part of ancestral neuronal circuits: Bacterial produced signals affect neurons controlling eating behavior in Hydra. bioRxiv. 2023;2004–23.

76. Fjell CD, Jenssen H, Fries P, Aich P, Griebel P, Hilpert K, et al. Identification of novel host defense peptides and the absence of α-defensins in the bovine genome. Proteins Struct Funct Bioinforma. 2008;73(2):420–30.

77. Vilcinskas A, Mukherjee K, Vogel H. Expansion of the antimicrobial peptide repertoire in the invasive ladybird Harmonia axyridis. Proc R Soc B Biol Sci. 2013 Jan;280(1750):20122113.

78. Huttner KM, Lambeth MR, Burkin HR, Burkin DJ, Broad TE. Localization and genomic organization of sheep antimicrobial peptide genes. Gene [Internet]. 1998;206(1):85–91. Available from: https://www.sciencedirect.com/science/article/pii/S0378111997005696

79. Clemmons AW, Lindsay SA, Wasserman SA. An effector peptide family required for Drosophila Toll-mediated immunity. PLoS Pathog. 2015;11(4):e1004876.

80. Sinner MP, Masurat F, Ewbank JJ, Pujol N, Bringmann H. Innate immunity promotes sleep through epidermal antimicrobial peptides. Curr Biol. 2021;31(3):564–77.

81. Belov K, Sanderson CE, Deakin JE, Wong ESW, Assange D, McColl KA, et al. Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. Genome Res. 2007;17(7):982–91.

82. Amid C, Rehaume LM, Brown KL, Gilbert JGR, Dougan G, Hancock REW, et al. Manual annotation and analysis of the defensin gene cluster in the C57BL/6J mouse reference genome. BMC Genomics [Internet]. 2009;10(1):606. Available from: https://doi.org/10.1186/1471-2164-10-606

83. Campos ML, de Souza CM, de Oliveira KBS, Dias SC, Franco OL. The role of antimicrobial peptides in plant immunity. J Exp Bot. 2018 Oct;69(21):4997–5011.

84. Wu J, Gao B, Zhu S. The Fungal Defensin Family Enlarged. Vol. 7, Pharmaceuticals. 2014. p. 866–80.

85. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res [Internet]. 2022 Jan 7;50(D1):D439–44. Available from: https://doi.org/10.1093/nar/gkab1061

86. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature [Internet]. 2021;596(7873):583–9. Available from: https://doi.org/10.1038/s41586-021-03819-2

87. Meng Q, Guo F, Tang J. Improved structure-related prediction for insufficient homologous proteins using MSA enhancement and pre-trained language model. Brief Bioinform. 2023 Jul;24(4).

88. Tasiemski A, Vandenbulcke F, Mitta G, Lemoine J, Lefebvre C, Sautiere P-E, et al. Molecular characterization of two novel antibacterial peptides inducible upon bacterial challenge in an annelid, the leech Theromyzon tessulatum. J Biol Chem. 2004;279(30):30973–82.

89. Xie Z, Yao T, Ye L, Wang J. Molecular characterization and expression analysis of an antimicrobial peptide, mytimacin-6, in the small abalone, Haliotis diversicolor. Isr J Aquac. 2021;73:1–11.

90. Shelomi M, Jacobs C, Vilcinskas A, Vogel H. The unique antimicrobial peptide repertoire

of stick insects. Dev Comp Immunol [Internet]. 2020;103:103471. Available from: https://www.sciencedirect.com/science/article/pii/S0145305X19303477

91. Vilcinskas A. Evolutionary plasticity of insect immunity. J Insect Physiol [Internet]. 2013;59(2):123–9. Available from: https://www.sciencedirect.com/science/article/pii/S0022191012002272

92. Hanson MA, Lemaitre B, Unckless RL. Dynamic evolution of antimicrobial peptides underscores trade-offs between immunity and ecological fitness. Front Immunol. 2019;10:2620.

93. Wenger Y, Buzgariu W, Reiter S, Galliot B. Injury-induced immune responses in Hydra. Semin Immunol [Internet]. 2014;26(4):277–94. Available from: https://www.sciencedirect.com/science/article/pii/S104453231400061X

94. Wenger Y, Buzgariu W, Galliot B. Loss of neurogenesis in Hydra leads to compensatory regulation of neurogenic and neurotransmission genes in epithelial cells. Philos Trans R Soc B Biol Sci [Internet]. 2016 Jan 5;371(1685). Available from: http://rstb.royalsocietypublishing.org/content/371/1685/20150040.abstract

95. Fraune S, Abe Y, Bosch TCG. Disturbing epithelial homeostasis in the metazoan Hydra leads to drastic changes in associated microbiota. Environ Microbiol. 2009;11:2361–9.

96. Kasahara S, Bosch TCG. Enhanced antibacterial activity in Hydra polyps lacking nerve cells. Dev Comp Immunol [Internet]. 2003;27(2):79–85. Available from: http://www.sciencedirect.com/science/article/pii/S0145305X02000733

97. Hancock REW, Alford MA, Haney EF. Antibiofilm activity of host defence peptides: complexity provides opportunities. Nat Rev Microbiol [Internet]. 2021;19(12):786–97. Available from: https://doi.org/10.1038/s41579-021-00585-w

98. Hilchie AL, Wuerth K, Hancock REW. Immune modulation by multifaceted cationic host defense (antimicrobial) peptides. Nat Chem Biol [Internet]. 2013;9(12):761–8. Available from: https://doi.org/10.1038/nchembio.1393

99. Yim G, Huimi Wang H, Davies Frs J. Antibiotics as signalling molecules. Philos Trans R Soc B Biol Sci. 2007;362(1483):1195–200.

100. Wan F, de la Fuente-Nunez C. Mining for antimicrobial peptides in sequence space. Nat Biomed Eng [Internet]. 2023;7(6):707–8. Available from: https://doi.org/10.1038/s41551-023-01027-z

101. Silveira GGOS, Torres MDT, Ribeiro CFA, Meneguetti BT, Carvalho CME, de la Fuente-Nunez C, et al. Antibiofilm Peptides: Relevant Preclinical Animal Infection Models and Translational Potential. ACS Pharmacol Transl Sci [Internet]. 2021 Feb 12;4(1):55–73. Available from: https://doi.org/10.1021/acsptsci.0c00191

102. Torres MDT, de la Fuente-Nunez C. Toward computer-made artificial antibiotics. Curr Opin Microbiol [Internet]. 2019;51:30–8. Available from: https://www.sciencedirect.com/science/article/pii/S1369527419300050

103. Li C, Sutherland D, Hammond SA, Yang C, Taho F, Bergman L, et al. AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens. BMC Genomics. 2022;23(1):1–15.

104. Obata Y, Pachnis V. Linking neurons to immunity: Lessons from Hydra. Proc Natl Acad Sci [Internet]. 2020 Aug 18;117(33):19624–6. Available from: https://doi.org/10.1073/pnas.2011637117

105. Hadfield MG, Bosch TCG. Cellular dialogues between hosts and microbial symbionts. In: Cellular Dialogues in the Holobiont. 2020.

106. He J, Bosch TCG. Hydra's Lasting Partnership with Microbes: The Key for Escaping Senescence? Microorganisms. 2022 Apr;10(4).

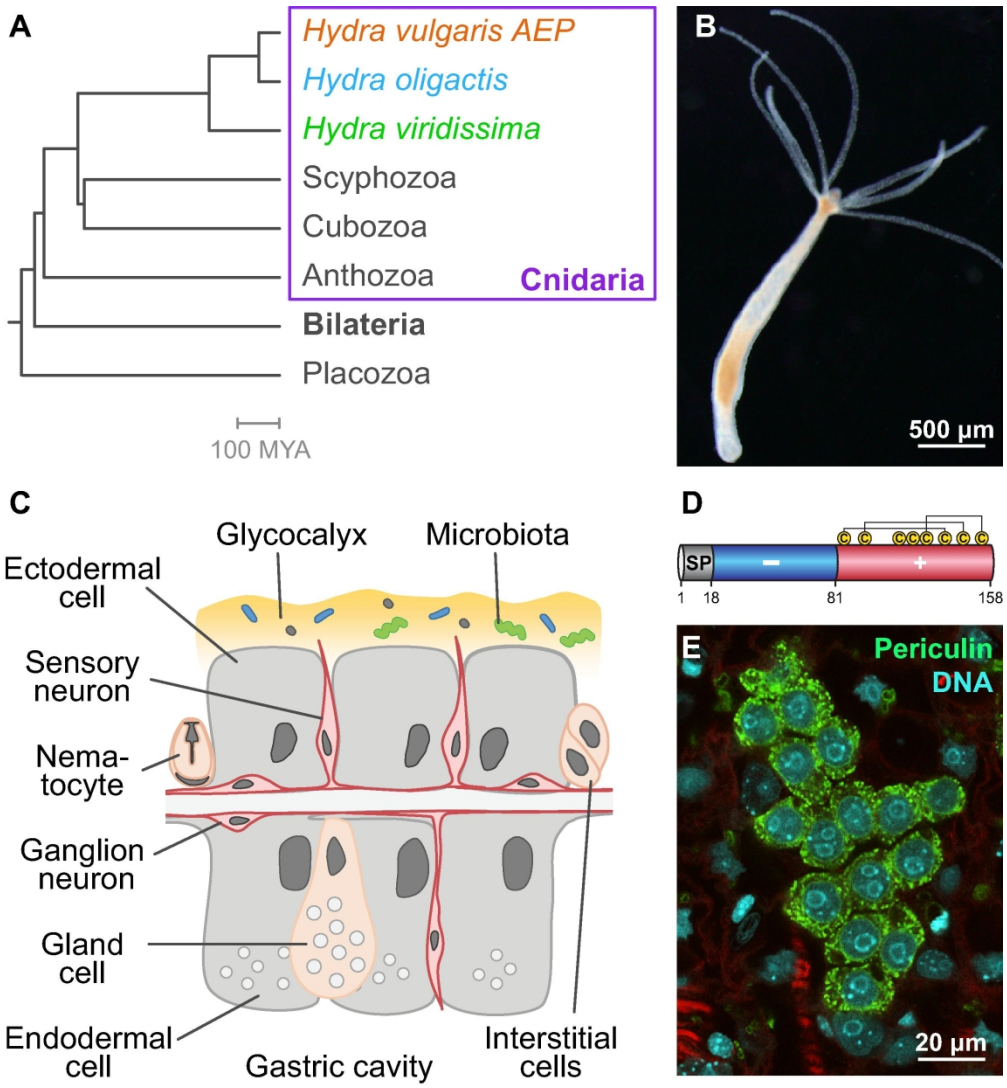

Figure 1: A: Phylogenetic tree demonstrating the position of *Hydra*. High-quality genome datasets have become recently available for three *Hydra* species – *H. vulgaris AEP*, *H. oligactis*, and *H. viridissima*. The divergence of the crown group *Hydra* took place about 193 MYA, and two species of brown hydras, *H. vulgaris* and *H. oligactis*, diverged over 100 MYA (18). **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 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.

112x121mm (600 x 600 DPI)

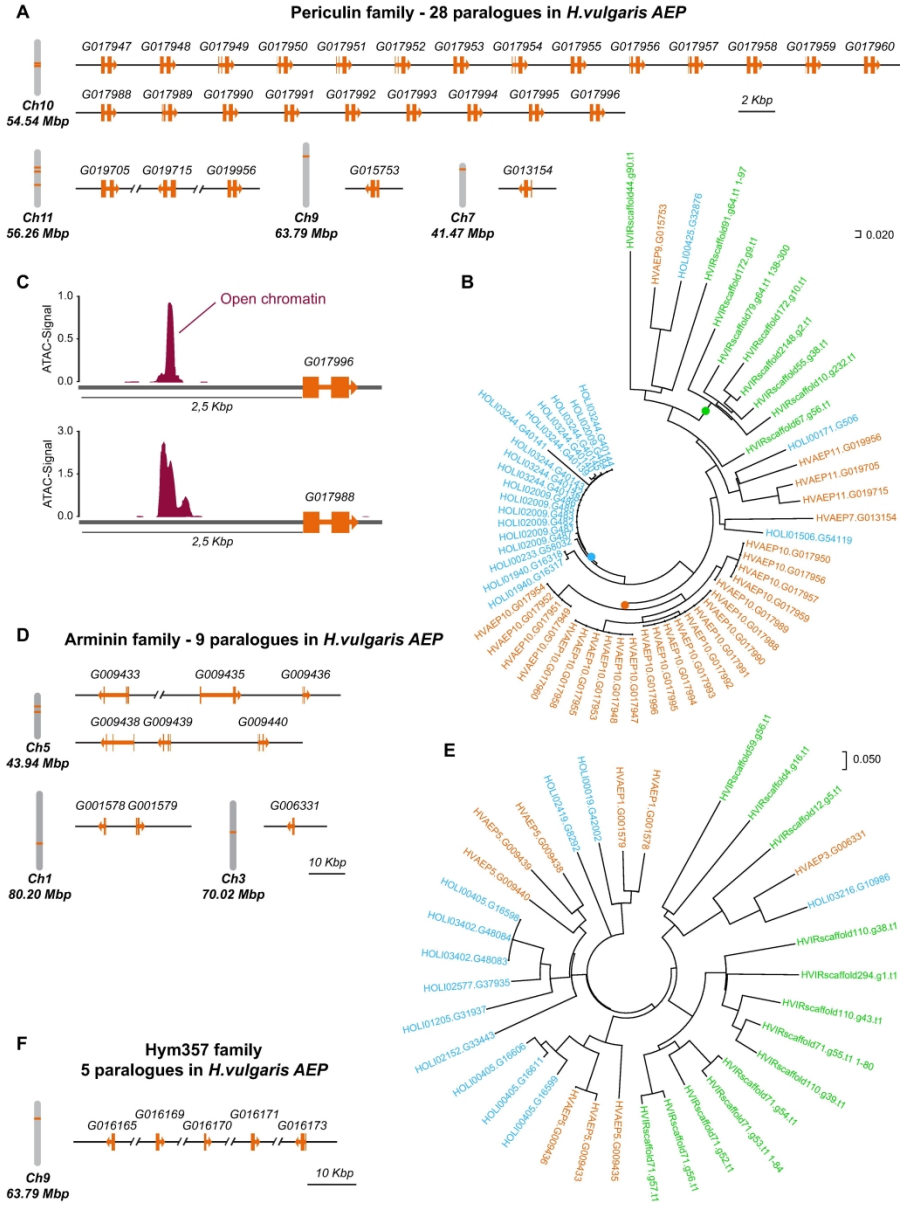

Figure 2: Complexity of AMP gene families in *Hydra* species. A: Twenty-eight paralogues of the periculin AMP family are located on 4 chromosomes of *H. vulgaris AEP*, whereby 23 genes form one dense cluster on chromosome 10. **B:** Phylogenetic analysis of periculin orthologues from three Hydra species. Genes are colored 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 Suppl. Fig. 3. **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 (20). **D:** Most of 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 Suppl. Fig. 4. **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 Suppl. Table 1.

169x227mm (600 x 600 DPI)

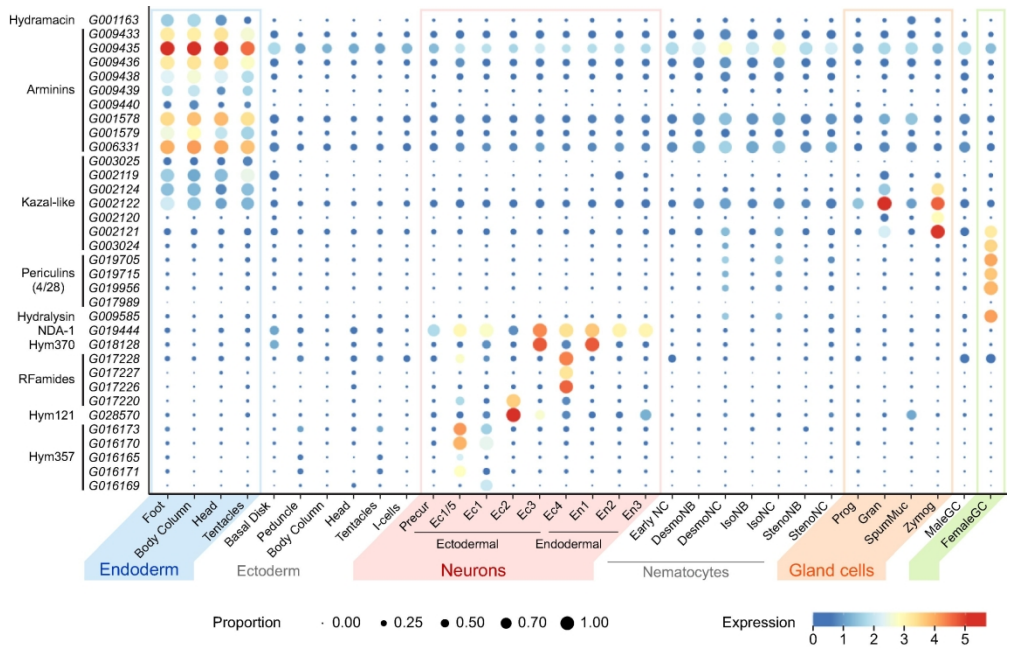

Figure 3: Expression of genes coding for AMPs across all cell types of *Hydra*. Visible are only the genes constitutively expressed in a *H. vulgaris AEP* polyp in homeostatic conditions, while inducible AMP genes expression is not illustrated. Note that only 4 out of 28 periculin paralogues are expressed at detectable levels. Visualization is based on data from (20), gene expression is normalized and log-transformed.

142x91mm (600 x 600 DPI)

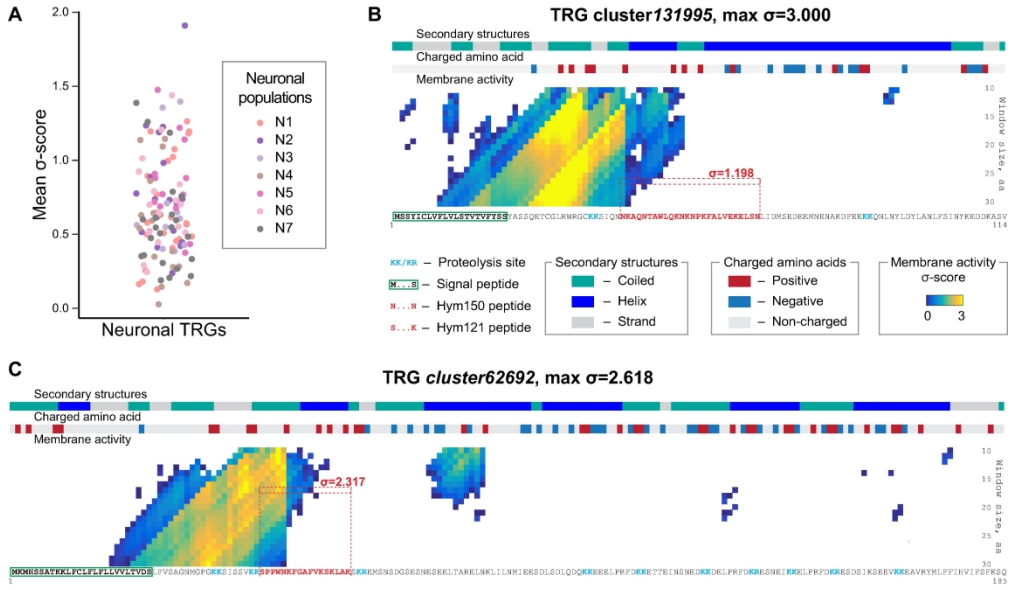

Figure 4: Machine learning algorithms allow for unbiased genome-wide prediction of putative AMPs. A: Distribution of mean σ-score values for individual secreted peptides encoded by neuron-specific TRGs in seven neuronal subpopulations illustrates a high likelihood of containing active AMPs for the peptides. Data are from (15). **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 (62) 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* (15).

176x102mm (600 x 600 DPI)

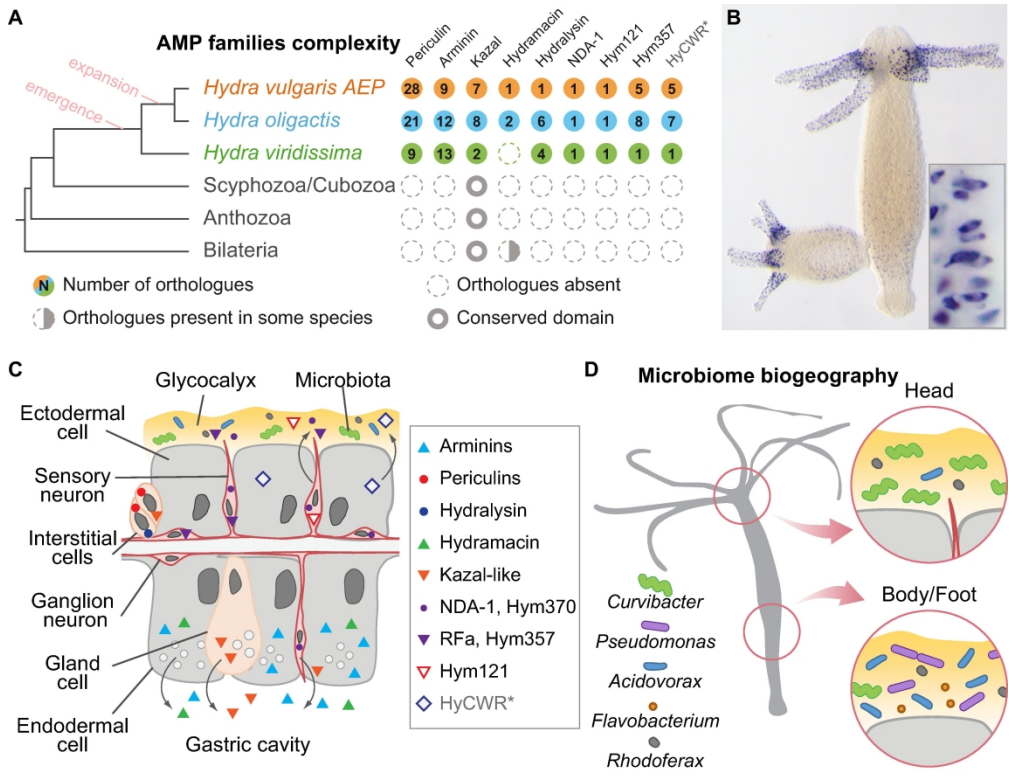

Figure 5: Complex species-specific and spatially restricted cocktails of AMPs sculpt the microbiome of *Hydra*. A: Overview of the AMP gene family complexity in *Hydra* species. Note that *H. viridissima* possesses generally fewer AMP genes compared to *H. oligactis* and *H. vulgaris AEP*. Most of AMP families are restricted to *Hydra* and only in few cases proteins with similar domains can 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 *Hydra* microbiome – its biogeography. * - the antimicrobial activity of HyCWR peptides and their role in the *Hydra* holobiont remain to be validated.

168x129mm (600 x 600 DPI)