

## Single cell atlas of *Xenoturbella bocki* highlights the limited cell-type complexity of a non-vertebrate deuterostome lineage

Helen E. Robertson<sup>1,2,3‡</sup>, Arnau Sebe-Pedros<sup>4,‡</sup>, Baptiste Saudemont<sup>5</sup>, Yann Loe Mie<sup>5</sup>, Anne Zakrzewski<sup>3</sup>, Xavier Grau-Bové<sup>4</sup>, Marie-Pierre Mailhe<sup>5</sup>, Philipp Schiffer<sup>3,5</sup>, Maximilian J. Telford<sup>3\*</sup> and Heather Marlow<sup>1,2\*</sup>

‡Equally contributing authors

\*Corresponding Authors: [hmarlow@uchicago.edu](mailto:hmarlow@uchicago.edu), [m.telford@ucl.ac.uk](mailto:m.telford@ucl.ac.uk)

### Author Affiliations:

<sup>1</sup> Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois, 60637, USA

<sup>2</sup> (Epi)genomics of Animal Development Unit, Department of Developmental and Stem Cell Biology, Institut Pasteur, 75015, Paris, France

<sup>3</sup> Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK

<sup>4</sup> Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

<sup>5</sup> Institute of Zoology, Section Developmental Biology, University of Cologne, Köln, Wormlab, Germany

### **Abstract**

Evidence for a sister relationship between Xenacoelomorpha and Ambulacraria (Xenambulacraria) has revived the longstanding debate surrounding the complexity of the Urdeuterostomian and Urbilaterian ancestors and has led to a reassessment of early bilaterian features. Employing whole organism single-cell RNA-seq in the marine xenacoelomorph worm *Xenoturbella bocki*, we show that Xenambulacrarian nerve nets share regulatory features and a peptidergic identity with those found in cnidarians and protostomes. We also suggest that Xenacoelomorpha muscles are likely to have evolved their “smooth” phenotype convergently. Furthermore, we identify pigmented cells as a potential synapomorphy of the Xenambulacraria. Taken together, these data are consistent with a simplification from an ancestral urbilaterian/urdeuterostomian body plan and a non-centralized nerve net, minimally differentiated contractile cells and evidence of deuterostome cell-type synapomorphies.

### **Introduction**

*Xenoturbella bocki* is a marine worm with long-contested phylogenetic affinities. Initially described as a turbellarian flatworm, *X. bocki* was first described by Westblad in 1949 from specimens collected from the muddy bottom of the Gullmarsfjord on the West coast of Sweden<sup>1</sup>. *X. bocki* is a small (~2cm long) yellowy-orange or brown slightly flattened worm (**Figure 1A**). It has a blind ended gut, no clear nervous condensations (nerve chords or a brain), coelomic cavities, discrete gonads or filtratory organs<sup>1-3</sup>. Five species have been described within the genus *Xenoturbella* in the past few years, all of which have a very similar morphology to *X. bocki*, differing most obviously in their colours and size with the largest, *X. monstrosa*, being as much as 20 cm long<sup>4</sup>.

Xenoturbellida (the higher-level group that contains *Xenoturbella*) is widely accepted to be the sister group of a second group of morphologically simple, free living marine worms - the Acoelomorpha, which together form the phylum Xenacoelomorpha. The Xenacoelomorpha are no longer considered to be close to the Platyhelminthes but their true affinities have been

highly contentious<sup>5-7</sup>. While a number of phylogenetic studies placed Xenacoelomorpha as the sistergroup of all other Bilateria, recent work suggests this tree is influenced by systematic error and the true affinities of Xenacoelomorpha is as the sister group of the Ambulacraria (Hemichordata and Echinodermata).<sup>8</sup>

Given their phylogenetic placement (whether as sistergroup of other Bilateria or of Ambulacraria), the morphological features of Xenoturbellids are critical to our understanding of animal body plan evolution<sup>6,9</sup>. Two major branches within animals constitute the “Bilateria”, a superphyletic assemblage of bilaterally symmetrical clades, the protostomes and the deuterostomes<sup>10</sup>. Within deuterostomes, the diversity of body plans and nervous system organizations have presented a challenge for the reconstruction of ancestral deuterostome characters. Recent evidence suggests that profound differences between Ambulacrarian, echinoderm and hemichordate, body plans and chordates could even be explained by paraphyly of the deuterostome clades<sup>8</sup>. It then follows that the position of Xenacoelomorphs as sister Ambulacraria, makes them increasingly relevant for our understanding of non-protostome body plan evolution. If indeed deuterostomes are a paraphyletic clade, the Xenacoelomorph body plan may serve as a proxy for a simple Xenambulacrarian ancestor. As such, investigations of *Xenoturbella*'s cellular diversity and transcriptional programs are of particular interest specifically in reference to excretory organs and nervous system cellular diversity.

Our current understanding of the morphology of *Xenoturbella* relies, for the most part, on classical morphological and histological descriptions. *Xenoturbella*'s most notable external features are a ventrally placed mouth acting as the single opening to a large central blind gut lumen, a circumferential ring furrow dividing front and rear of the body, and lateral furrows at the anterior end. Internally there is an anterior statocyst, a thick, ciliated epidermis with a basiepithelial nervous system and layers of muscles surrounding the gut. Our knowledge of the cell type diversity of *Xenoturbella bocki* at the level of either cellular morphology or gene expression is minimal.

Here, we use whole-organism single cell RNA sequencing (scRNA-seq) as a high-throughput approach to characterise the cell type expression programs in adult specimens of *Xenoturbella*. In order to validate our findings and to link the cell states we have identified to the locations and morphology of cells in the animal, we have developed an *in situ* hybridisation (ISH) protocol on sectioned animals.

## **Results**

### **Complement of cell states in *Xenoturbella***

We sampled single cell transcriptomes from four adult specimens of *Xenoturbella bocki* collected from the Gullmarsfjord, Sweden using the MARS-seq, microplate-based single-cell sequencing protocol<sup>11</sup> (for details see Methods). We sequenced single-cell libraries to an average depth of 33,870 reads per cell (Table S1). The average total mappability to the *Xenoturbella* genome was 78% of reads, with 26% uniquely mapped reads. 12,350 cells with more than 100 transcripts per cell (determined by Unique Molecular Identifiers, UMIs) were retained for downstream analyses. The median number of transcripts measured per retained cell was 418 and we detected expression for a total of 11,304 genes ( $\geq 10$  total UMIs) (Figure S1A-B)

We applied the Metacell algorithm<sup>12</sup> to group cells into transcriptionally coherent clusters (metacells) (**Figure 1C, S1C-E**). Each metacell contained between 41 and 291 single-cells. By examining a graph-based 2D projection (**Figure 1C**) and gene co-expression patterns (**Figure 1D**), we identified 26 distinct cell type/tissue-level groups of metacells characterised by specific transcriptional signatures. These metacell groups include putative muscles, neurons, epithelial and digestive gland cells, amongst others (**Figure 1C, D**). Gut cells (30%) and epidermal cells (22%) are the most abundant cell types in *Xenoturbella*, followed by muscle (8%), ciliated cells (7%), neurons (5%) and pigmented cells (5%) (**Figure 1C**).

To localise the populations of cells characterised by each of the cell states identified by their transcriptomic profile, we identified marker genes whose expression was highly specific for each metacell population. We characterised the expression of a subset of these markers in preserved tissue sections of adult *Xenoturbella* using *in situ hybridization* (ISH). This allowed us to associate the cell clusters we had identified with the corresponding cells in the adult animal providing information on the relative positions of different cell types and their morphology, size and abundance (**Figure 1E**).

### Gut, digestive, and excretory cells

The blind-ended *Xenoturbella* gut constitutes a large central tissue with a lumen that is clearly identifiable in horizontal-section (**Figure 1E**)<sup>2</sup>. Reflecting the large proportion of the animal that is made up of gut tissue, we found that the largest group of metacells, defined principally by the expression of a number of transporter proteins, is identified as the gut tissue by ISH (**Figure 1F**). We have identified four additional groups of cells that express digestive enzymes and proteases, and which are likely to belong to a digestive gland population. One of these secretory gland cell populations is characterised by expression of chymotrypsin, and ISH locates this cell type to the gastrodermis of *Xenoturbella* (**Figure 1G**).

We also found a population of cells that expresses a number of genes - sodium transporters, an Rh-type ammonium transporter, and an ortholog of aquaporin-8 (**Figure 1H**) indicative of an osmoregulatory and excretory function. These aquaporin and ammonium transporter genes are specifically and highly expressed within this cell population and are not found expressed in the population of cells expressing transporters, suggesting the presence of cells specialised for excretion/osmoregulation and different from other gut cells.

### Muscle cell types

Three discrete populations of muscles have been described in *Xenoturbella* with different orientations and relative positions - an inner layer of longitudinal muscle adjacent to the gut, a layer of circular muscle surrounding these, and radial muscles connecting the circular and longitudinal muscle layers to the epidermal layer (**Figure 2A**)<sup>8/18/22 2:07:00 AM</sup>. All of these have been described as smooth muscle<sup>13</sup>.

Metacells representing muscle cells in *Xenoturbella* were identified by the presence of conserved marker genes including troponin and myosin homologs. Most muscle cluster cells express calponin, calmodulin, muscle LIM and myosin light chain. A subset of metacells expresses the transcription factors musculin (*MyoR*), FoxF1 and FoxG and an additional group is specifically enriched for titin-like contractile elements (**Figure 2B**).

All eight muscle metacells identified express a “striated” myosin heavy chain (MHC) ortholog as well as three Troponin orthologs (Troponin I, C and T); only Troponin I and T are present in echinoderms and hemichordates. One of these eight clusters of cells (# 76) expressing muscle-effector specifically expresses a “smooth” muscle myosin type. These cells, as with the other muscle clusters, co-express other muscle-related genes including troponin members and titin-related genes (**Figure 2B**).

To confirm that the clusters identified by muscle-specific genes in our scRNA-seq analysis (**Figure 2B**) correspond to the muscles previously characterised by histological analyses, we performed ISH with *XbTroponin-C* and *XbTroponin-T* expressed across all of the putative *Xenoturbella* muscle clusters and found expression in the prominent muscle layer adjacent to the gut (**Figures 2C, D**).

### Neuronal cell types

The nerve plexus of *Xenoturbella* has been described as a simple basiepithelial nerve net, lacking discrete ganglia or nerve cords<sup>3,14</sup>. The location, density and possible identity of neurons has been characterised in a limited number of ultrastructural studies<sup>3,14-16</sup>, but there has been no molecular characterization of the diversity of cell types that might be present or their distribution. Both major lineages of the sister clade Ambulacraria - hemichordates and echinoderms - show some condensation of nerve cords<sup>17,18</sup>, and different degrees of condensation is seen in acoels<sup>9,19</sup>. It remains unclear whether the ‘nerve-net’ like organization observed in *Xenoturbella* is an ancestral or derived condition. We aimed to identify neuronal populations among *Xenoturbella* metacells, identify their putative functional identity and to examine their expression of transcriptional regulators.

Histological studies of *Xenoturbella* cross sections have shown that the neuronal cell bodies lie in the region between the bulk of the nerve plexus and the base of the epidermis (Anne Zakrzewski, unpublished) (**Figure 2E**). Presumptive neuronal metacells were identified by expression of the pan-neural gene *synaptotagmin* (**Figure 2F**). ISH using *synaptotagmin* labeled cells in the basal portion of the epidermis - in the presumptive nerve plexus - with higher levels of expression at the anterior end of the animal (**Figure 2G**). These neuronal metacells also expressed markers characteristic of vesicle secretion and of the pre-synaptic synapse, notably, *secretagogin* and *SNAP25* (**Figure 2F**). *Secretagogin* is localised to the nerve plexus at the very anterior of the animal (**Figure 2H**). In addition to these synaptic vesicle genes, enzymes required for neuropeptide processing, indicative of peptidergic neurons (prohormone convertase and peptidylglycine alpha-hydroxylating monooxygenase) were expressed across all neuron subtypes (**Figure 2F**). We also identified two broad groups of neurons expressing either the transcription factors *POU4* or *islet* (**Figure 2F**). Neurons marked by *POU4* are likely to be sensory neurons expressing members of the *erg* family of K<sup>+</sup> channel genes as well as a diversity of PKD channels. The neural bHLH member *achaete-scute* is also specifically upregulated in *POU4* positive neurons. A broader subset of neurons is identified by the expression of *islet*. We found *islet* to be co-expressed with the glutamate transporter *vgt*, and with *elav* and *irx*. A specific group of *islet* positive cells is also enriched for *lim1*, and adrenergic and acetylcholine receptors (**Figure 2F**).

We identified several genes expressed in specific neuronal cell clusters for which we could identify no orthologs outside of *Xenoturbella*. Dot plot mappings of these show expression in distinct neural metacells (**Figure 2F**). For one of these orphan genes, ISHs localised expression to a small and specific subset of cells within the anterior nerve plexus (**Figure 2I**). A second



orphan had a more diffuse expression across a defined anterior region of the animal's nervous system (**Figure 2J**).

### **Pigmented cells**

One group of several metacells was strongly enriched for *tyrosinase*-family genes (**Figure 3A**). Tyrosinases are found across the animal, plant and fungal kingdoms, where they act as the primary rate-limiting enzyme during melanin biosynthesis<sup>20</sup>. In vertebrates, melanin is found in pigment cells; in invertebrates, melanin has been reported to function in a number of different roles including wound healing, immune response, and photoreception<sup>21,22</sup>.

ISH for one of the *Xenoturbella* tyrosinase members localised expression across the epidermis, with higher expression at the anterior of the animal. This expression in the epidermis is consistent with the Tyrosinase expressing cells correspond to the subset of pigmented epidermal cells (**Figure 3B**).

Tyrosinase proteins and the related Tyrosinase-like proteins show independent lineage-specific expansions in a number of metazoan groups<sup>20,23,24</sup>. For example, Tyrosinase-related proteins (tyrp), which are identifiable by conserved residue changes in their metal binding domains, are uniquely found in chordates. A phylogenetic analysis of tyrosinase and tyrosinase-related sequences show that *Xenoturbella* tyrosinase sequences represent a *Xenoturbella* specific expansion (**Figure 3C**). Our tree of Metazoan tyrosinase related proteins groups *Saccoglossus*, vertebrate and *Branchiostoma* tyrosinases and tyrosinase-related proteins, and places the *Xenoturbella* tyrosinase-like sequences as a sister group to all other deuterostome tyrosinase family proteins.

Cells in the *Xenoturbella* tyrosinase-enriched metacells are also enriched for transcripts of a polyketide synthase-like (PKS) gene. PKSs and the related fatty acid synthases (FASs) are a diverse family, found across the eubacteria and eukaryotes<sup>25</sup>. In a phylogeny of PKS and FAS sequences from across eubacteria and eukaryotes the XbPKS1 sequence groups with high confidence with sequences (SpPKS) from an echinoderm (*Strongylocentrotus purpuratus*) (**Figure 3D**).

In *S. purpuratus*, *SP-PKS1* is necessary for the biosynthesis of echinochrome - one of the major pigment classes present in echinoderms<sup>26</sup>. Several other genes in addition to *PKS1* are required for echinochrome biosynthesis, including *glial cells missing (gcm)*, three *flavin monooxygenases (fmo1, 2 and 3)*, a *sulfotransferase (sult)* and a *dopachrome tautomerase (dopt)*<sup>26</sup>. No tyrosinase or tyrosinase-like members have been identified in the echinoderms<sup>24</sup>. All pigment cell clusters found in *Xenoturbella* express *dopt* and *sulfotransferase*, but we did not detect enrichment of any *fmo* members or *gcm*. In *Xenoturbella*, we also see the upregulation of *aristaless (alx1)* in the pigment cells (figure ref), in common with the mesodermally-derived larval skeleton in the sea urchin, which expresses *aristaless (alx1)*. The co-expression of *PKS* and *alx1* in the urchin and *Xenoturbella* pigment cell populations might suggest a common mesodermal differentiation pathway that has diverged in the specific mesodermal lineages leading to skeleton and pigment cells in the echinoderm.

### **Ciliary furrow**

We identified seven metacells marked by the expression of numerous ciliary-related genes - including *tubulin*, *calmodulin*, *rootletin*, *FoxJ1* and *tubby* (**Figure 3E**). *In situ* hybridisation for

*FoxJ1* localised expression to two opposing regions midway along the anteroposterior axis, which we identify as the ciliated circumferential furrow (**Figure 3G**) midway along the anteroposterior axis of *Xenoturbella* (**Figure 3F**). The function of the circumferential ciliary furrow in *Xenoturbella* is not known. ISH for an ortholog of the tektin gene *Tekt4*, necessary for microtubule stabilisation<sup>27</sup>, also localised expression to this circumferential furrow (**Figure 3H**). ISH for a *Xenoturbella* orphan gene specific to the ciliary metacell group, showed expression in both the ciliary furrow and the anterior nerve plexus (**Supp Fig S3**). We did not find expression by ISH of any other ciliary marker gene in the neuronal metacells, nor neuronal marker genes in the ciliary metacells.

We looked further into the genes expressed in the ciliary metacluster and found many markers associated with extracellular matrix (ECM) remodelling and maintenance, such as *hemicentin-1*, *ets* factors, *cadherin* members including *cadherin-23*, *collagen*, *integrin* and *fibropellin*, and a number of genes necessary for conveying information between the ciliary membrane and ECM, including *PKD1*, *AURKA*, *SRC*. Interestingly, we also find genes implicated in bacterial response, including *Saposin*, *LITAF*, and NFκB (**Figure 3E**).

### Cell type-specific endosymbiotic bacteria

*Xenoturbella* has been previously shown to harbour two types of bacteria in the gastrodermis; one related to *Chlamydia* and the other to Gammaproteobacteria<sup>28,29</sup>. We identified genes showing similarity to sequences from both *Chlamydia* and Gammaproteobacteria in our *Xenoturbella* scRNA-seq data supporting an intracellular endosymbiotic relationship. By measuring bacterial transcript abundances across metacells, we show that the *Chlamydia* is found both in the gastrodermis and the epidermis, while the proteobacterial signal is found primarily in gut metacells (**Figure 4A**). We found that the two different bacteria infect different subsets of cell populations within the gastrodermis, with a small population of cells that are infected by both types of bacteria (**Figure 4B, C**). We found that *Chlamydia* also infect a group of cells within the epidermal population (**Figure 4A, B**). Confirming this predicted pattern, *In situ* hybridisation against the *Chlamydia* 16S rRNA localised this bacterium to both the gut and epidermis of the animal (**Figure 4D**).

Interestingly, the percentage of cells we detect which are infected by either endosymbiont varies across the animals sampled (**Figure 4A**). However, the percentage of cells infected by either endosymbiont is inconsistent across the animals sampled. In three out of four of the animals sampled ~10% of cells contain bacteria; the fourth replicate had just ~5%. The relative contribution of each bacterial type to total infection is also variable between animals (**Figure 4E**). Discrepancies between frequency and type of bacterium did not obviously correlate with the time of year at which the different animals were collected although the samples were limited.

### Discussion

Here we report the use of whole-organism scRNA-seq in conjunction with *ISH* to characterise the cell type repertoire of *Xenoturbella bocki*. This single cell analysis provides the most comprehensive molecular profiling of cell states in this animal to date, and the parallel use of *ISH* for highly specific and enriched genes provides a spatial context for the cell types we identified. While the broad categorisation of cell types we find is broadly consistent with prior morphological descriptions of the animal, the molecular fingerprints of these cell types allow

us to identify potentially homologous cell types and give unique insights into the biology and life history of the animal.

The proportions of different cell types represented in our data reflect the abundance of these cell types in the animal: the gut and epidermis, which are the largest tissue layers seen in histological analysis, comprise over half of the metacells in our clustered dataset. Although we were able to assign an identity to the majority of metacells identified in our analysis, some metacells, representing between 0.8 and 1.8% of the sampled cells in our dataset, remain unidentified.

### **Blind ended gut with multiple distinct cell types**

We find populations of digestive gland cells in *X. bocki* that do not express transporters, indicating that *Xenoturbella* has distinct populations of digestive gland cells that are separate from the nutrient transporting and absorptive cell population of its blind gut. This differs from the acoel *Isodiametra pulchra*, which also possesses a blind-gut. In *I. pulchra*, an aquaporin member, the bilaterian ultrafiltratory gene nephrin/kirre and transporters were found to be co-expressed in a broad population of cell types<sup>30</sup>.

The aquaporin-8 ortholog in *Drosophila*, DRIP, is expressed exclusively in the fly's excretory Malpighian tubules, and the complement of ion transporters we find in the population of *Xenoturbella* gastrodermal cells is similar to that found in the distal collecting tubule of the flatworm protonephridium. Our data suggest that, unlike acoelomorphs, *Xenoturbella* has a discrete cell type specialised for osmoregulation. We have not been able to produce ISH data that allow us to identify these cells.

### **Conserved differentiators of neural and muscle cell types**

We find *Xenoturbella* neurons to be significantly less diverse than those characterized in acoelomorphs<sup>31</sup>, another of the Xenacoelomorph lineages. Interestingly, *Xenoturbella* neurons express hallmarks of metazoan neurons including *islet-1*, *brn3/pouIV* and the proneural bHLH transcription factor *achaete-scute*, indicating that they share some features of an ancestral neural regulatory program. PouIV has been identified as serve a neural identity factor in lineages across bilaterians including flies and mice and was recently identified to be functionally required for the development of neurons of the sea anemone *Nematostella vectensis*<sup>32–34</sup>, suggesting a role in the earliest nervous systems, those found in lineages which predate Bilateria. Similarly, the lim-homeodomain gene *islet-1* has been identified in cnidarian neurons<sup>35</sup> and in neurons and neuroendocrine cells in the Bilateria<sup>36,37</sup>.

Expression of the pre-synaptic complex components required for vesicle secretion as well as the neuropeptide processing pathway members in these cells strongly suggests that they function as peptidergic neurons. Recent comparative work in invertebrate nervous systems suggests that neuropeptidergic control of simple behavioural circuits as well as paracrine signaling of neuroendocrine-like cells was a feature of early nervous systems<sup>38–40</sup>. These features have been preserved in marine invertebrate nervous systems broadly supporting a “chemical brain hypothesis” for nervous system evolution<sup>41</sup>. The conservation of this organization in *Xenoturbella* strongly suggests that this simple peptidergic nerve plexus represents an ancestral ground plan for animals that has been preserved in the Xenacoelomorph lineage.

The *Xenoturbella* muscle metacells have the molecular hallmarks of striated muscle despite being morphologically smooth, but no representatives of any of the myogenic regulatory factors were found in our data. This finding provides some support for the independent evolution of striated muscle in separate lineages by the repurposing of core bilaterian-specific proteins.

### **Xenambulacraria synapomorphies**

Recent phylogenetic studies place Xenocoelomorpha (Xenoturbellids and Acoels) as the sister clade to the Ambulacraria<sup>7,9</sup> - a lineage which includes the hemichordates (enteropneusts and pterobranchs) and echinoderms - despite few putative morphological synapomorphies of the Xenambulacraria (Xenaxoelomorpha plus Ambulacraria clade). While Xenacoelomorpha lack characters such as gill slits, a coleom and a through gut that are likely to have existed in the Ambulacrarian ancestor, they share some characters with possible affinities to other ambulacrarian lineages. Among these are the pigment cell population, and the transverse ciliary band.

In the larval sea urchin, pigment cells develop from primary mesenchyme, a mesodermal cell type. These cells become pigmented by “echinochrome” generated by the PKS and sulfotransferase gene products. Similarly, we find that *Xenoturbella* pigmented cells express both PKS and sulfotransferase as well as a massively expanded repertoire of tyrosine hydroxylase genes. Further, the expression of the transcription factor *alx1* in the *Xenoturbella* pigment cell population is reminiscent of *alx* expression in derivatives of the primary mesenchyme in sea urchin - the skeletogenic mesenchyme.

Ciliary bands are present both on the larvae of echinoderms and on the tornaria larvae of hemichordates, where they are necessary for locomotion and feeding. In hemichordates, ciliary bands persist through metamorphosis into the adult. The molecular fingerprint of the *Xenoturbella* ciliary cell types is unlike that found in the hemichordates, echinoderms, or that reported in the acoel *Hofstenia*. The primary ciliary markers, *foxj1* and *tubby*, suggest that the circumferential ciliary furrow in *Xenoturbella* may be important for integrating environmental cues, such as fluid-flow or external compression, but we do not find the expression of genes that are otherwise expressed in the neural cluster, or any marker genes that have been found in the neural ciliary band of larval echinoderms. ISH for one *Xenoturbella*-specific gene, with no orthologs in other taxa, showed expression in the ciliary furrow and anterior nerve plexus, but ISH for other ciliary marker genes (including *Tubby* and *Tektin*) showed no expression outside of the ciliary furrow. This may indicate that the ciliary band in *Xenoturbella* is a lineage-specific innovation, unrelated to those found in other ambulacrarians.

### **Methods**

#### **Experimental model and subject details**

Adult *Xenoturbella bocki* were collected from the Gullmarsfjorden on the west coast of Sweden. A Warén dredge aboard the Oscar von Sydow research vessel (Sven Loven Centre, Kristineberg, Gothenburg University) was used to collect soft mud at a depth of ~60-100m. Mud was subsequently filtered through two sieves of decreasing pore sizes, so that larger organisms were removed first, and *Xenoturbella* (of approximately 50-100mm) were retained on the mesh of the second sieve with all mud washed away. The contents of the second sieve

were washed into large shallow trays, and *Xenoturbella* specimens identified by eye. Animals were kept unfed in sealed falcon tubes filled with deep-sea fjord water at 4°C.

Four adult specimens were sampled at different times for FACS sorting. Animals were dissociated by placing them in full strength calcium/magnesium-free and EDTA-free artificial sea-water (CMFSW). Animals were then transferred to 300 $\mu$ L CMFSW supplemented with Liberase<sup>TM</sup> (Roche, #05401119001) at a concentration of 50  $\mu$ g/ml. Dissociation was carried out at room temperature by pipetting using gelatine-coated pipette tips of decreasing diameter to disrupt the cell suspension over a period of 15 minutes. Dissociation was stopped by the addition of one-tenth volume of 500 mM EDTA solution to a final concentration of 50 mM. 3  $\mu$ g Calcein AM (Thermo #C3100MP) and 2.25  $\mu$ g Propidium Iodide (PI) (Thermo #P3566) were added to the cell suspension to identify live cells. Cells were then placed on ice and not washed out of the Calcein/PI solution before sorting.

Dissociated cells were sorted using a BD FACSAria III fluorescently-activated cell sorting machine into 384-well capture plates with each well containing 2 $\mu$ l of cell lysis solution. Live cells were selected by sorting Calcein positive/PI negative cells. Doublet/multiplet exclusion was performed using FSC-W versus FSC-H. Lysis solution contained 0.2% Triton X-100, RNaseIN and barcoded poly(T) reverse-transcription (RT) primers for single cell RNA-seq. Immediately after cell distribution, plates were briefly spun down to ensure complete immersion of cells in the lysis solution and stored at -80°C. Four empty wells were left in each plate as a control.

### Single-cell RNA-seq

Single cell libraries were prepared as previously described <sup>8</sup>/18/22 2:07:00 AM. All 42 plates (a total of 15,960 single cell libraries) were prepared using the same conditions and reagents. Using a Bravo automated liquid handling platform (Agilent), mRNA was reverse transcribed into cDNA using the lysis oligonucleotide containing unique molecular identifiers (UMIs) and cell barcodes. Residual, unused oligonucleotides were removed by treatment with Exonuclease I (NEB). cDNAs were pooled and linearly amplified using T7 *in vitro* transcription. The resulting RNA was fragmented and ligated to an oligo containing the pool barcode and Illumina sequences, using T4 ssDNA:RNA ligase. RNA was reverse transcribed into DNA and PCR-amplified with 17 cycles. The size distribution and concentration of DNA in the libraries were measured using TapeStation (Agilent) and Qubit (Invitrogen). scRNA-seq libraries were pooled at equimolar concentrations and sequenced as a paired end run on an Illumina NextSeq 500 with a 75 cycle v2 kit. Read one was 59bp (covering the pool-specific barcode and cDNA) and read two was 17bp (covering the well-specific barcode and UMI). We obtained a total of 1,086M reads, with an average depth of 30,330 reads per cell and 19.73 reads/UMI on average.

### MARS-seq reads processing and filtering

Reads were mapped onto the *Xenoturbella* genome (Schiffer et al, unpublished) using bowtie2 (with parameters: `-D 200 -R 3 -N 1 -L 20 -i S,1,0.50 --gbar`) and assigned to gene intervals associated with a custom gene annotation (Supp File X). We extended gene intervals up to 1kb downstream or until the next gene on the same DNA strand is found. This accounts for the poor 3'UTR annotation, which causes many of the MARS-seq (a 3' biased RNAseq method) reads to map outside genes. Additionally, in order to account for putative unannotated genes, we defined 2000bp bins (not covered by our gene intervals) genome-wide. We retained 592 such regions with  $\geq 100$  uniquely mapping reads and used them in the cell clustering process.



Mapped reads were further processed and filtered as previously described<sup>42</sup>. UMI filtering includes two components, one eliminating spurious UMIs resulting from synthesis and sequencing errors, and the other eliminating artifacts involving implausible IVT product distributions that are likely a consequence of second strand synthesis or IVT errors. The minimum FDR q-value required for filtering was 0.2.

### **Metacell and clustering analysis**

We used the MetaCell package<sup>12</sup> to select gene features, to construct gene modules and to create projected visualization of the data, using parameters as described below.

We applied preliminary cell filtering based on total UMI counts using a permissive threshold of 100 UMIs. For gene feature selection we used a normalized depth scaling correlation threshold of -0.05, and total UMI count of more than 100 molecules. For metacell balanced k-nn graph construction we used K=150. We computed 1,000 graph partitions resampling 75% of the cells to obtain a co-occurrence matrix. The final metacell partition was calculated on this co-occurrence matrix using K=30, alpha=2 and min\_mc\_size=30 parameters. For 2D metacell projections, a force-directed projection was calculated using a K-nn constant of 20 and restricted the module graph degree by at most 4.

We filtered 21 metacells (representing 1,571 single cells) that lacked specific expression signatures. We note that these filtered cells may represent weaker signals that are in fact part of other, more robust modules, but that for our goals in the analysis here, robustness of the reported transcriptional states and the subsequent genomic analyses is key. Overall, our final dataset consisted of 12,350 single-cells organized in 97 metacells.

### **Bacterial endosymbiont analysis**

MARS-seq reads were mapped against the genomes of the two *Xenoturbella bocki* bacterial endosymbionts: related to *Chlamydia* and *Gammaproteobacteria*. The entire genome of each bacteria (in both frames) was considered for molecule counting, under the assumption that our polyA-capture based scRNA-seq method might be detecting at least some bacterial transcripts (even if inefficiently) and, therefore, these aggregated counts constitute a proxy for the presence of these symbionts within single-cells. Despite the fact that bacterial transcripts are non-polyadenylated, we indeed detected hundreds of bacterial molecules (genome-wide) in some single-cells (0-4960 bacterial UMIs, median 10). Chlamydial and gammaproteobacterial UMI counts were pooled using our 97-metacell clustering solution and the signal for each bacteria was normalized by the total bacterial transcripts per cell and also the *Xenoturbella* transcripts (to account for differences in sampling efficiency).

### **Functional gene annotation**

We used blastp (with parameters `-evalue 1e-5` and `-max_target_seqs 1`) to find the most similar, if any, human, fruit fly and yeast homologs (retrieved from Uniprot) for each protein of the *Xenoturbella* predicted proteomes. Additionally, we predicted, for each protein, the Pfam domain composition using Pfamscan<sup>43</sup> with default curated gathering threshold. TFs were identified using univocal Pfam domains for each structural TF family<sup>44</sup>.

### **Animal fixing and sectioning**

Animals for *in situ* hybridisation were washed on ice progressively from sea-water into 3.6% MgCl<sub>2</sub> and anaesthetised in the dark at 4°C for approximately four hours. Prior to complete loss of movement, the orientation of the animals was noted (anterior/posterior; left/right; dorsal/ventral axis) and individuals were photographed. Fully anaesthetised animals were placed on blotting paper on top of a block of frozen fixative (freshly prepared 4% paraformaldehyde (PFA) in 1x phosphate buffered saline (PBS)) and left for ~12 hours in the dark at 4°C to fix. Fixed animals were washed several times in 1x PBS and subsequently washed into progressively increasing concentrations of methanol (MeOH). Specimens were stored in 100% MeOH at -20°C prior to use.

Fixed animals stored in 100% MeOH were removed from -20°C, and MeOH was exchanged with 100% EtOH over a number of quick washes. Animals were washed into HistoSol, three times for 20 minutes each at room temperature. After a total of one hour of HistoSol incubation, samples were washed into a 50:50 mix of HistoSol:paraffin wax. Two 30 minute washes into the HistoSol:paraffin mix were carried out at room temperature. Finally, specimens were transferred into 100% paraffin wax and incubated overnight at 60°C. The following day, samples were washed five times in 100% paraffin wax at 60°C. Paper moulds were filled with 100% molten wax and the specimens placed into the mould. Wax was left to solidify at room temperature.

Embedded animals were horizontally sectioned at either 6 or 7µm onto Superfrost Plus slides along the dorso-ventral axis using a rotary microtome. Slides were stored at room temperature.

### **Total RNA extraction and cDNA synthesis for ISH probe PCRs**

Animals from which total RNA was to be extracted were placed directly into RNA-later REF for a minimum of one week for later RNA extraction. RNA was extracted using a standard Trizol/Chloroform RNA extraction protocol, quantified on Qubit and stored at -80°C until required. Total RNA was used to make cDNA under RNAase-free conditions with the Ambion RETROscript RT kit (Invitrogen AM1710). A 12µl reaction using 2µl RNA and 1.2 µM oligo(dT) primers was incubated at 80°C for three minutes and then removed onto ice. A final reaction volume of 20µl was made with 1x M-MLV RT buffer, 10 mM dNTPs, 100U M-MLV reverse transcriptase and 10U RNase Inhibitor. Reverse transcription was carried out at 42°C for two hours and the reaction finally heated to 90°C for ten minutes to inactivate the reverse transcriptase.

### **Probe design and synthesis**

cDNA from total RNA was diluted 1:10 for use in a PCR for probe synthesis PCR. PCR primer sequences for making *in situ* hybridisation probes are shown in table X. For each probe synthesis PCR, 20µl first round reactions were prepared using Qiagen HotStarTaq DNA polymerase (Qiagen # 203205). The cycling conditions were as set as 95°C for 15 minutes; 30 cycles at 94°C for 30 s, 60°C for 30 s, 72°C for 1 min 30 s; a final elongation at 72°C for 10 min. Resulting PCR products were diluted 1:50, and 1ul of the diluted reaction used in a second round PCR amplification, using reverse primers which also included a T7 promoter sequence. The cycling protocol was the same as that used in first round PCR. Second round PCR products were cleaned using Agencourt AMPure XP beads (Beckman Coulter #A63881) and visualised on a 1% agarose gel to verify amplicon size.

Antisense RNA probes were transcribed from the purified PCR products using T7 RNA polymerase with digoxigenin-conjugated dUTPs, with the Megascript T7 Transcription kit (Ambion, AM 1334). Final RNA probes were purified using RNA XP (Beckman, A63987) quantified on Qubit and stored at -80°C in hybridization buffer.

### ***In situ* hybridisation (ISH)**

All preparatory and pre hybridisation stages were carried out in coplin jars treated with RNaseZap (Sigma Aldrich #R2020) and rinsed in DEPC-water. The protocol for *in situ* hybridisation on sections was adapted from Gillis et al. (2012).

Slides were dewaxed by two five-minute washes in HistoClear, followed by two five-minute washes in 100% EtOH and HOW MANY two-minute washes through an ethanol gradient into diethylpyrocarbonate (DEPC)-treated phosphate-buffered saline (PBS) (90%, 70% and 50% EtOH in DEPC-PBS). Slides were rinsed once in DEPC-treated water; once in PBTw (0.1% Tween-20 in DEPC-PBS), and once in 2x SSC (0.3 M NaCl, 0.03 M sodium citrate).

Probes were diluted in 250µl hybridisation solution (1x salt solution [0.2 M NaCl, 10 mM Tris pH 7.5, 5 mM NaH<sub>2</sub>PO<sub>4</sub>.H<sub>2</sub>O, 5 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM EDTA], 50% formamide, 10% dextran sulphate, 1 mg/ml yeast tRNA, 1x Denhardt's solution) to a final concentration of 1ng/µl. Hybridisation buffer containing probe was applied to each slide and incubated overnight at 60°C under a glass coverslip in a sealed chamber humidified with 50% formamide/2x SSC. Slides were rinsed twice for 30 minutes each in wash solution at the hybridisation temperature, followed by three ten minute washes at room temperature in MABT (0.1 M maleic acid, 150 mM NaCl, 0.1% Tween-20, pH 7.5).

Slides were blocked for two hours at room temperature in blocking buffer composed of 1% Roche blocking reagent (Roche #11096176001) and 20% sheep serum, diluted in MABT. Anti-digoxigenin-AP Fab fragments (Roche #11093274910) were diluted 1:1000 in the same blocking buffer. 130µl of the anti-DIG solution was applied to each slide and incubated overnight under a paraffin coverslip in a humidified chamber at room temperature.

Slides were rinsed twice in MABT, followed by five washes of progressive length in MABT on an orbital shaker, and equilibrated in NTMT (100 mM NaCl, 100 mM Tris, pH 9.5, 50 mM MgCl<sub>2</sub>, 0.1% Tween-20). The colour reaction was initiated using BM Purple ready-to-use AP substrate (Roche #1142074001) in the dark at room temperature and stopped by washing in PBS. Slides were post-fixed in 4% PFA for 10 minutes, transferred back to PBS and mounted using Fluoromount G.

1. Westblad, E. *Xenoturbella bocki* n.g, n.sp, a peculiar, primitive turbellarian type. *Arkiv för Zoologi* **1**, (1949).
2. Gavilán, B., Sprecher, S. G., Hartenstein, V. & Martinez, P. The digestive system of xenacoelomorphs. *Cell and tissue research* **377**, 369–382 (2019).

3. Raikova, O. I., Reuter, M., Jondelius, U. & Gustafsson, M. K. An immunocytochemical and ultrastructural study of the nervous and muscular systems of *Xenoturbella westbladi* (Bilateria inc. sed.). *Zoomorphology* **120**, 107–118 (2000).
4. Rouse, G. W., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. *Nature* **530**, 94–97 (2016).
5. Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89–93 (2016).
6. Philippe, H. *et al.* Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria. *Curr Biol* **29**, 1818–1826.e6 (6).
7. Philippe, H. *et al.* Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature* **470**, 255–258 (2011).
8. Kapli, P. *et al.* Lack of support for Deuterostomia prompts reinterpretation of the first Bilateria. *Sci. Adv.* **7**, eabe2741 (2021).
9. Kapli, P. & Telford, M. J. Topology-dependent asymmetry in systematic errors affects phylogenetic placement of Ctenophora and Xenacoelomorpha. *Sci. Adv.* **6**, eabc5162 (2020).
10. Dunn, C. W. *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–9 (2008).
11. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
12. Baran, Y. *et al.* MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* **20**, (2019).
13. Ehlers, U. & Sopott-Ehlers, B. Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. *Zoomorphology* **117**, 71–79 (1997).
14. Stach, T. 8 XENOTURBELLA. *Structure and evolution of invertebrate nervous systems* 62 (2015).

15. Stach, T. *et al.* Nerve cells of *Xenoturbella bocki* (phylum uncertain) and *Harrimania kupfferi* (Enteropneusta) are positively immunoreactive to antibodies raised against echinoderm neuropeptides. *Journal of the Marine Biological Association of the United Kingdom* **85**, 1519–1524 (2005).
16. Israelsson, O. Ultrastructural aspects of the ‘statocyst’ of *Xenoturbella* (Deuterostomia) cast doubt on its function as a georeceptor. *Tissue and Cell* **39**, 171–177 (2007).
17. Nomaksteinsky, M. *et al.* Centralization of the deuterostome nervous system predates chordates. *Current biology* **19**, 1264–1269 (2009).
18. Lowe, C. J. *et al.* Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
19. Hulett, R. E., Potter, D. & Srivastava, M. Neural architecture and regeneration in the acoel *Hofstenia miamia*. *Proceedings of the Royal Society B* **287**, 20201198 (2020).
20. Esposito, R. *et al.* New Insights into the Evolution of Metazoan Tyrosinase Gene Family. *PLOS ONE* **7**, e35731 (2012).
21. Porchet-Henneré, E. & Vernet, G. Cellular immunity in an annelid (*Nereis diversicolor*, Polychaeta): production of melanin by a subpopulation of granulocytes. *Cell and tissue research* **269**, 167–174 (1992).
22. Sugumaran, M. Comparative biochemistry of eumelanogenesis and the protective roles of phenoloxidase and melanin in insects. *Pigment cell research* **15**, 2–9 (2002).
23. Aguilera, F., McDougall, C. & Degnan, B. M. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. *Acta biomaterialia* **10**, 3855–3865 (2014).
24. Aguilera, F., McDougall, C. & Degnan, B. M. Origin, evolution and classification of type-3 copper proteins: lineage-specific gene expansions and losses across the Metazoa. *BMC evolutionary biology* **13**, 1–12 (2013).
25. Castoe, T. A., Stephens, T., Noonan, B. P. & Calestani, C. A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* **392**, 47–58 (2007).



26. Calestani, C. & Wessel, G. M. These Colors Don't Run: Regulation of Pigment-Biosynthesis in Echinoderms. *Results Probl Cell Differ* **65**, 515–525 (2018).
27. Amos, L. A. The tektin family of microtubule-stabilizing proteins. *Genome biology* **9**, 229 (2008).
28. Israelsson, O. Chlamydial symbionts in the enigmatic Xenoturbella (Deuterostomia). *J Invertebr Pathol* **96**, 213–20 (2007).
29. Kjeldsen, K. U., Obst, M., Nakano, H., Funch, P. & Schramm, A. Two types of endosymbiotic bacteria in the enigmatic marine worm *Xenoturbella bocki*. *Applied and environmental microbiology* **76**, 2657–2662 (2010).
30. Andrikou, C., Thiel, D., Ruiz-Santesteban, J. A. & Hejnol, A. Active mode of excretion across digestive tissues predates the origin of excretory organs. *PLoS biology* **17**, e3000408 (2019).
31. Duruz, J. *et al.* Acoel Single-Cell Transcriptomics: Cell type analysis of a deep branching bilaterian. *Molecular biology and evolution* **38**, 1888–1904 (2021).
32. Leyva-Díaz, E., Masoudi, N., Serrano-Saiz, E., Glenwinkel, L. & Hobert, O. Brn3/POU-IV-type POU homeobox genes—Paradigmatic regulators of neuronal identity across phylogeny. *WIREs Dev Biol* **9**, (2020).
33. Serrano-Saiz, E., Leyva-Díaz, E., De La Cruz, E. & Hobert, O. BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice. *Current Biology* **28**, 2813-2823.e2 (2018).
34. Tournière, O. *et al.* NvPOU4/Brain3 Functions as a Terminal Selector Gene in the Nervous System of the Cnidarian *Nematostella vectensis*. *Cell Reports* **30**, 4473-4489.e5 (2020).
35. Sebé-Pedrós, A. *et al.* Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* **173**, 1520-1534.e20 (2018).
36. Ericson, J., Thor, S., Edlund, T., Jessell, T. M. & Yamada, T. Early stages of motor neuron differentiation revealed by expression of homeobox gene *Islet-1*. *Science* **256**, 1555–1560 (1992).

37. Thor, S. & Thomas, J. B. The *Drosophila* islet Gene Governs Axon Pathfinding and Neurotransmitter Identity. *Neuron* **18**, 397–409 (1997).
38. Williams, E. A. *et al.* Synaptic and peptidergic connectome of a neurosecretory center in the annelid brain. *eLife* **6**, e26349 (2017).
39. Varoqueaux, F. *et al.* High Cell Diversity and Complex Peptidergic Signaling Underlie Placozoan Behavior. *Current Biology* **28**, 3495-3501.e2 (2018).
40. Sachkova, M. Y. *et al.* Neuropeptide repertoire and 3D anatomy of the ctenophore nervous system. *Current Biology* **31**, 5274-5285.e6 (2021).
41. Jékely, G. The chemical brain hypothesis for the origin of nervous systems. *Phil. Trans. R. Soc. B* **376**, 20190761 (2021).
42. Keren-Shaul, H. *et al.* MARS-seq2. 0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nature protocols* **14**, 1841–1862 (2019).
43. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2021).
44. de Mendoza, A. & Sebé-Pedrós, A. Origin and evolution of eukaryotic transcription factors. *Current Opinion in Genetics & Development* **58–59**, 25–32 (2019).

**Fig.1: *Xenoturbella* single-cell RNA-seq identifies spatially distinct cell types and a minimally complex neural cell type complement.** A. A light micrograph of a live *Xenoturbella bocki*. B. Phylogenetic tree showing the recently proposed position of *X. bocki* as sister to Ambulacraria. C. A T-SNE projection of 12,350 *X. bocki* cells. D. A cell by gene heatmap of cells from (C) showing expression similarity within the distinct tissue types. E. A schematized diagram of a cross-section of *Xenoturbella* tissues: ci (cilia), sf (supporting filament), esc (epidermal supporting cell), gc (gland cell), nu (nucleus), np (nerve plexus), m (muscle), ecm (extracellular matrix), gu (gut), smc (sub-epidermal membrane complex), ep (epidermis). Right, schema of a horizontal cross-section of a *Xenoturbella* individual in which endodermal (red), ectodermal (aqua) and mesodermal (yellow) derivatives are indicated. F.-G. In situ hybridization of cell-type markers identified from the scRNA-seq clustering. H. Markers associated with excretion co-localize in a metacell.

**Figure 2: Muscle and neural metacell markers localize to histologically distinct muscle cell types and cell bodies in the subepithelial nerve plexus, respectively.** A. Schematic indicating the position of muscle in *X. bocki*. B. Muscle markers are enriched in a group of

eight metacells (yellow). C. In situ hybridization of troponin transcripts localization to muscle. C'. High magnification image of striations in stained tissue from (C). D. Troponin staining in muscle. E. Schematic indicating the position of the subepidermal membrane complex (smc). F. A group of metacells (blue) expressing transcripts encoding channel proteins, neuropeptide processing proteins, secretory proteins and transcription factors. G-H. Synaptotagmin I (G) and secretogin (H) expression are localized to the anterior nerve plexus. I and J. *X. bocki* specific transcripts are localized to the nerve plexus.

**Figure 3: Hallmarks of *X. bocki*, the pigmented epidermal cells and the sensory, ciliated transverse band, are associated with metacell groups by in situ hybridization.** A. An expansive repertoire of tyrosinases as well as polyketide synthase (pks) are enriched in pigmented epithelial cells. B. Tyrosinase transcript expression in pigmented epithelial cells. Maximum likelihood trees indicating a large, species-specific, expansion of tyrosinase transcripts in *X. bocki* (C) and the position of *X. bocki* PKS as nested within animal PKS genes (D). E. Transverse sensory band enriched markers. F. Schematic of the position of the transverse sensory band on an intact animal. Expression of the ciliary sensory band markers foxJ1 (G) and tectin (H).

**Figure 4: Intracellular microbial populations of *Chlamydia* and Gammaproteobacteria are associated with gut and epidermal tissues.** A. Abundance of each microbial population within *X. bocki* cells. The proportion of *Chlamydia* vs. Gammaproteobacteria show in the four sampled individuals. B. *Chlamydia* are associated with gut and epidermal cells. C. Gammaproteobacteria are associated with gut populations only. D. An in situ hybridization of 16s rRNA for *Chlamydia* expressed in both gut and epidermal populations of cells.











