Cooperation between bacteriocytes and endosymbionts drives function and development of symbiotic cells in mussel holobionts

Hao Chen[1†*], Mengna Li[1,4†], Zhaoshan Zhong[1†], Minxiao Wang[1,3*], Chao Lian[1], Guanghui Han[1], Hao Wang[1], Li Zhou[1], Huan Zhang[1], Lei Cao[1], Chaolun Li[1,2,3*]

[1] Center of Deep Sea Research, and CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

[2] South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China

[3] Laoshan Laboratory, Qingdao 266071, China

[4] National Deep Sea Center, Qingdao 266071, China

[†]These authors contributed equally to this work

[*]Correspondence and requests for materials should be addressed to H.C. (email: chenhao@qdio.ac.cn), M.W. (email: wangminxiao@qdio.ac.cn) or C.L. (email: lcl@qdio.ac.cn)

Abstract

Symbiosis drives the adaptation and evolution in multicellular organisms. Modeling the function and development of symbiotic cells/organs in holobionts is yet challenging. Here, we surveyed the molecular function and developmental trajectory of bacteriocyte lineage in non-model deep-sea mussels by constructing a high-resolution single-cell expression atlas of gill tissue. We show that mussel bacteriocytes optimized immune genes to facilitate recognition, engulfment, and elimination of endosymbionts, and interacted with them intimately in sterol, carbohydrate, and ammonia metabolism. Additionally, the bacteriocytes could arise from three different stem cells as well as bacteriocytes themselves. In particular, we showed that the molecular functions and developmental process of bacteriocytes were guided by the same set of regulatory networks and dynamically altered regarding to symbiont abundance via sterol-related signaling. The coordination in the functions and development of bacteriocytes and between the host and symbionts underlies the

interdependency of symbiosis, and drives the deep-sea adaptation of mussels.

Keywords

Chemosynthetic endosymbiosis, function and development of symbiotic cell, host-symbiont interaction, metabolic remodeling, development plasticity

Introduction

Symbiosis is a major force in the acquisition of novel adaptive traits, expansion of ecologic range, and shaping the biodiversity and evolution of eukaryotic organisms [1]. To establish symbiotic associations, the host has formed highly complex symbiotic cells and organs, which harbor the symbionts and provide mutualist relationships. The formation and development of symbiotic cells and organs are therefore regarded as sources of phenotypic complexity and direct outcome of developmental plasticity of the holobionts [2]. Along with a growing number of studies, the function of, and symbiotic interactions within, symbiotic cells and organs are becoming clear in some model holobionts [3-5]. However, characterizing the formation and development of symbiotic cells and organs remains challenging in both model and non-model holobionts [6]. Furthermore, questions on whether symbionts influence the formation and development processes in the host, and the possible mechanisms involved in this activity, remain debated and may vary greatly across species [7-9].

The phylum Mollusca is the second most species rich phylum. Mollusks thrive in a wide range of habitats including fresh water, intertidal region, and even the deep-sea environments. While mollusks possess vastly diverse body plans, some mollusks, especially bivalves from the *Mytilidae, Vesicomyidae, Solemyidae, Thyasiridae,* and *Lucinidae* families, have also formed nutritional symbiosis with chemosynthetic microbes in adaptation to harsh environments such as deep-sea cold seeps and hydrothermal vents [10]. While the chemosymbiosis in mollusks can vary greatly depending on the symbiont location (extracellular or intracellular), symbiont type (methanotroph, thiotroph, or both), and transmission mode (vertical or horizontal), such associations may also be lost during evolution when organic matter became more accessible [11,12]. The mollusks are therefore regarded as a promising model to investigate the formation and development of symbiotic cells and organs, and to examine the participation of symbionts in these processes. As endemic species in deep-sea chemosynthetic ecosystems, the deep-sea *Bathymodiolinae* mussels are known to harbor endosymbiotic methanotrophs and/or thiotrophs in their specialized gill epithelial cells (bacteriocytes) [13-15]. It was also estimated that adult deep-sea mussels contained 10^8 to 10^9 bacteriocytes with hundreds of methanotrophic endosymbionts or thousands of thiotrophic endosymbionts harbored inside a single bacteriocyte [16]. The chemosynthetic endosymbionts further provide up to ninety percent of the total energy budget of mussel hosts, depending on the body size of the mussel and varying with environmental particulate organic matter and reduced chemicals [17].

The chemosymbiosis between deep-sea mussels and endosymbionts further sustains their flourishing communities in worldwide deep-sea chemosynthetic ecosystems, including cold-seeps, hydrothermal vents, and whale falls. For these reasons, how deep-sea mussels establish and maintain the symbiotic associations with their endosymbionts is therefore vital for understanding both the phenotypic complexity and developmental plasticity of holobionts as well as their ecological success in the deep sea. With great efforts by omics tools, such as whole-genome sequencing, bulk RNA-seq, proteomics analysis and spatial metabolomics, several pathways and genes potentially involved in the nutritional and immune interaction between mussel and endosymbionts are proposed [18-22]. However, due to the extensive heterogeneity of gill tissue and limits in the isolation and *in vitro* culture of bacteriocytes and endosymbionts, the exact molecular mechanisms and cellular processes underlying the establishment and maintenance of chemosymbiosis still remain obscure and debatable. With recent advances in single cell transcriptome and spatial transcriptome technologies, it is now possible to reveal the function and development of symbiotic cells and organs in either model or non-model organisms, and with and without host cell lineages and culturable symbionts [23-25]. In this study, we used high-resolution single-cell transcriptome and meta-transcriptome data to construct and survey a population atlas and the molecular function and development trajectory of bacteriocytes in the methanotrophic deep-sea mussel *Gigantidas platifrons*. We then compared our data with those obtained on the decolonized mussel to evaluate the potential influence of symbionts on the function and development of symbiotic cells. Finally, we compared our data with those obtained on other mollusks in order to determine whether the toolkit guiding the function and development of symbiotic cells is universal and conserved in the mollusk.

Results

Functional landscape of bacteriocyte lineage

Single cell transcriptome and spatial transcriptome technologies provide the best ever opportunity to characterize the function and development of bacteriocytes in non-model species. Recently, we have successfully characterized the cell atlas of *G. platifrons* gill tissue using the microwell-based BD Rhapsody single-cell sequencing platform and identified one bacteriocyte lineage along with another 12 cell types [26]. We have noticed that the varying number of cell nuclei subjected to sequencing (ranging from less than 10,000 to nearly 30,000 nuclei per sample) has affected the characterization of host genes (with 631 identified genes per cell per sample, median value) and identification of candidate marker genes for either bacteriocytes (only 25 genes expressed by more than 25% of target cell cluster, and with expression levels at least 2-fold higher than that of other cell cluster; *p*-value less than 0.01) or stem cells (only 15 genes for 3 stem cell lineages), which further hampered the characterization of function and development of bacteriocytes. To better characterize the expression pattern of bacteriocyte lineages, we employed two groups of deep-sea mussels in the present study and conducted both single-cell transcriptome (scRNA-seq, Chromium platform of 10x Genomics) and spatial transcriptomics (ST-seq, Visium platform of 10x Genomics) analysis (Fig. 1A). In particular, mussels collected from the seepage region (methane concentration up to 31,227 ppm, designated as the InS group, n=14) were used to represent normal mussels with a fully symbiotic state; mussels transplanted *in situ* to a low methane region (methane concentration about 800 ppm) for 604 days with an average 76.5% decrease of endosymbionts were used to represent partially decolonized mussels (DeC group, n=5). Principal co-ordinates analysis (PCoA) based on the meta-transcriptome data of the two groups showed robust expressional stability within individuals of the same group and strong variations between individuals of different groups (Supplementary Fig. S1A–B). Two individual mussels with similar body size were selected from each group and subjected for scRNA-seq and ST-seq. To include more possible cell types identified previously and examine their molecular functions precisely, the dorsal–middle region of the gill (containing both descending and ascending gill filaments and adjacent to posterior end of mussel) was subjected to scRNA-seq after cell nuclei extraction. In addition, to better annotate gill cell types, particularly those of bacteriocytes, the cross-sectioned middle region of the gill (dorsal view) was used for ST-seq (Fig. 1A). As a result, over 3,600 barcoded spots with 172,705 mean reads per spot, and approximately 26,707 cells with 37,585 mean reads per cell, were obtained for the two samples evaluated using ST-seq and scRNA-seq (Supplementary Table 1). Transcript coverage profiling indicated that sequencing depth of all samples was saturated for ST-seq and scRNA-seq (Supplementary Fig. S1C–F), ensuring the reliability of subsequent analyses. In addition, the median number of identified genes per cell and total number of candidate marker genes (Supplementary Table 1, 2) have been improved substantially in comparison with our previous study [26], which is necessary for the characterization of function and development of bacteriocytes.

With knowledge on the cell atlas and marker genes of mussel gill, a combined strategy based on the maker gene expression pattern and cell morphology was employed to annotate bacteriocytes and other cell lineages. In harmony with our previous findings [26], a total of 14 cell clusters, including the bacteriocytes (cluster 1), three types of stem cells [dorsal end proliferation cells (DEPCs, cluster 0), posterior end budding zone cells (PEBZCs, cluster 2 and 6)], five types of ciliary and smooth muscle cells [intercalary cells (ICCs, cluster 4), lateral ciliary cells (LCCs, cluster 9), food grove ciliary cells (FGCCs, cluster 13), and smooth muscle cells (SMCs, cluster 12)], five supportive cells [basal membrane cell 1 (BMC1, cluster 5 and 11), BMC2 (cluster 7), mucus cell (cluster 8), and inter lamina cells (ILCs, cluster 10)] and one unknown cells (cluster 3), were identified here based on scRNA-seq data. The annotation results were further verified by combining scRNA-seq- and ST-seq-derived data using Seurat-v3.2 anchor-based integration and projecting all identified cell clusters from the scRNA-seq data onto gill tissue data of ST-seq (Fig. 1B–C). Particularly, fluorescent *in situ* hybridization of endosymbionts and key marker genes confirmed the successful identification of bacteriocytes (Fig. 1D), which in consistency with previous findings are distributed on the middle part of the gill filaments as well as on the inner (abfrontal) edge of gills.

In addition to their distinct anatomical distribution pattern, the bacteriocytes were also distinguishable at the molecular level, especially with respect to their marker genes (Supplementary Table 2) and molecular function (Supplementary Fig. S2A). The bacteriocytes transcribed a total of 199 candidate marker genes based on scRNA-seq data of both the InS and DeC groups, which increased greatly than our previous study and consisted 10 out of 25 candidate marker genes that identified previously [26]. In support of the scRNA-seq data, majority of the annotated marker genes of bacteriocytes (83/199) were also found abundantly expressed (with expression levels at least 1.5-fold higher than that of other cell cluster; *p*-value less than 0.01) in ST-seq data of InS group. Function enrichment analysis of marker genes further showed that the bacteriocytes largely encoded genes involved in metabolic processes (such as biosynthesis and transport of carbohydrates, lipids, amino acids, and vitamins) and symbiosis-related immune responses (including GO: 0002376, GO: 0009617, GO: 0044111, and GO: 0044403) (Supplementary Fig. S2). Particularly, we noticed that a total of 67 out of 199 marker genes, such as excitatory amino acid transporter 1 (*SLC1A3*), monocarboxylate transporter 13 (*SLC16A14*), cathepsin-L (*CTL*), acid phosphatase type 7 (*ACP7*), and myoneurin (*MYNN*), were consistently expressed by bacteriocytes in the InS and DeC groups regardless of fluctuations in symbiont abundance. Comparatively, the rest of the 132 out of 199 marker genes, including solute carrier family 40 member 1 (*SLC40A1*), ammonium transporter Rh type A (*RHBG-A*), solute carrier family 26 member 10 (*SLC26A2*), 24-hydroxycholesterol 7 alpha-hydroxylase (*CYP39A1*, EC1.14.14.26), sugar phosphate exchanger 2 (*SLC37A2*), zinc finger protein 271 (*ZNF271*), and ETS-related transcription factor (*EHF*) were downregulated in the DeC group in response to symbiont decolonization (Supplementary Table 3).

Symbiosis-related immune processes in bacteriocytes

With merely the innate immunity, how invertebrate host forged the symbiotic association with the chemosynthetic bacteria remains intriguing. We, therefore, screened for the possible genes and pathways that participate in the establishment and maintenance of symbiosis in bacteriocyte lineages. Notably, we found that only several pattern recognition receptors (PRRs) were highly expressed in bacteriocytes of fully-symbiotic mussels (InS group) or decolonized mussels (DeC group), while numerous PRRs showed a relatively low level of expression or even failed to be characterized (e.g., approximately 55 out of 146 TLRs were not characterized by scRNA-seq). Notably, one *TLR2* gene were annotated as markers for bacteriocytes, respectively (Fig. 2A, Supplementary Table 2). In addition, two PRRs (*rhamnose-binding lectin* and *TLR2*) showed increased expression levels in bacteriocytes after decolonization (Supplementary Table 3). PRRs, such as TLRs and PGRPs, play vital roles in the recognition of symbionts in other holobionts [27]; the aforementioned PRRs may, therefore, act as main effectors in the establishment or maintenance of symbiotic association in deep-sea mussels.

In addition to these PRRs, we also detected abundant expression of endocytosis- and lysosome-related marker genes in bacteriocytes of fully-symbiotic mussels (Fig. 2A, Supplementary Table 2). The coordination among these genes may facilitate phagocytosis and lysosome-mediated digestion of bacteria (Fig. 2B, Supplementary Fig. S2B), which promotes the establishment of symbiosis and controls symbiont population. We additionally observed that endocytosis and lysosome-mediated digestion were vigorously regulated by the host depending on symbiont abundance (Fig. 2C, Supplementary Table 3), suggesting that a flexible toolkit was used to maintain this association.

Metabolic interaction between the host and symbionts

While the deep-sea mussels are known to be highly dependent on endosymbionts as nutrition sources for sterol intermediates, essential amino acids, vitamins, and carbohydrates, the exact metabolites and mechanism beneath the nutritional interaction between bacteriocytes and endosymbionts remains largely obscure due to the limits of the bulk-seq method. Bearing this in mind, we screened scRNA-seq and ST-seq data of mature bacteriocytes along with the meta-transcriptome data of endosymbionts to uncover possible metabolic interactions within the deep-sea mussel holobionts. As expected, both scRNA-seq and ST-seq data showed that mature bacteriocytes of fully symbiotic mussels (InS group) were massively transcribing the *CYP39A1* (EC1.14.14.26) gene (Fig. 3A), implying a high turnover of sterol metabolism and demand for intermediates such as presqualene-PP, squalene, (S)-squalene-2,3-epoxide, and 4,4-dimethyl-cholesta-8,14,24-trienol. While the unique expression pattern of *CYP39A1* (EC1.14.14.26) gene inside bacteriocytes was further confirmed by FISH and IHC assay, we also noticed that the mussel host is in need of related enzymes such as EC2.5.1.21, EC1.14.1417, EC1.14.14154, and EC1.14.1536. Consistently, meta-transcriptome data showed that methanotrophic endosymbionts were robustly transcribing enzymes such as EC2.5.1.21 and EC1.14.14154 under normal condition (the top 10% highly expressed genes in the InS group) in a complementary manner with the host (Fig. 3A, Supplementary Fig. S3A–D, Supplementary Table 4). This finding confirmed that symbionts could provide sterol intermediates to the host [28] and showed the reliability of scRNA-seq and ST-seq in investigating metabolic interactions in the deep-sea mussel.

We then focused on carbohydrate metabolism to identify possible interactions between both partners. Although massive lysosome-related genes were abundantly expressed in bacteriocytes while some endosymbionts were being digested by lysosomes, both our EdU labeling assay and 3D electron microscopy analysis showed that few endosymbionts (3 out of 408 symbionts for an individual bacteriocyte) were proliferating inside bacteriocytes of adult mussels and even fewer were being captured from environment (Supplementary Fig. S2C, Supplementary Video.1), making the “farming of endosymbionts” unsustainable for adult mussels to acquire nutrition. It is therefore speculated that the endosymbionts can also provide the host with carbohydrates directly without being digested (a “milking” way). Yet, what carbohydrates are exchanged, and how this exchange occurs, remains unknown. Meta-transcriptome data of endosymbionts showed that multiple key genes (including EC2.4.1.18, 2.4.1.21, 2.7.7.27, and 5.4.2.2) involved in gluconeogenesis and glycogen biosynthesis were abundantly expressed (the top 10% of highly expressed genes in the InS group, Fig. 3B, Supplementary Fig. S4, Supplementary Fig. S5, Supplementary Table 4), suggesting a robust production of glycogen in endosymbionts. However, multiple enzymes in the gluconeogenesis pathway were either insufficiently expressed (EC5.3.1.9) or completely absent from the endosymbiont genome (EC3.1.3.11, EC2.7.1.11, and EC4.1.2.13), implying overproduction of fructose-6P and shortages of glucose-6P, fructose-1,6P_2, glyceraldehyde-3P, and glucose. Conversely, we found that all of these enzymes were encoded by mature bacteriocytes, while a sugar phosphate exchanger gene *SLC37A2* (responsible for transmembrane transport of fructose-6P, glucose-6P, fructose-1,6P_2, and glyceraldehyde-3P) were abundantly expressed by mature bacteriocytes even under normal conditions (scRNA-seq and ST-seq of InS group) with proteins co-localizing with endosymbionts (Fig. 3B, Supplementary Fig. S5). Complementation in carbohydrate metabolism strongly suggested that symbionts could supply fructose-6P directly to the host in exchange for gluconeogenesis intermediates. In support of this speculation, we also noted the abundant expression of phosphate and carboxylate transporters (including the phosphate-binding protein PstS, maltose 6-phosphate phosphatase, phosphate permease, sodium-dependent dicarboxylate transporter SdcS, and bicarbonate transporter BicA) in endosymbionts (Supplementary Table 4), which might be responsible for the reciprocal interaction with mussel bacteriocytes. While the lack of gluconeogenesis-related enzymes is not detrimental to the symbiont, coordination between both partners in the exchange of sugar phosphates can provide the host with a more efficient and diverse mechanism of acquiring carbohydrates.

Analysis of TEM bacteriocyte images showed that endosymbionts were contained in separate vacuoles designated as symbiosomes [29,30]. While the presence of symbiosomes creates a suitable micro-niche for endosymbionts, it also limits their access to environmental nutrients [31]. Nutrients such as oxygen, carbon dioxide, and methane can pass freely through the membrane, but other nutrients, such as ammonia, must be supplied by the host [32]. We also observed high expression of ammonia consumption-related genes (top 10%) rather than the ammonia production-related genes in endosymbionts (Fig. 3C, Supplementary Fig. S6, Supplementary Table 4), suggesting a robust ammonia demand of endosymbionts. Ammonia, while important for the symbiont, is detrimental to the host as a by-product of protein digestion and nucleotide metabolism. Here, with scRNA-seq and ST-seq data, we noted that the mussel host was massively transcribing ammonium transporter gene *RHBG-A* in mature bacteriocytes under normal conditions (Fig. 3C), which might facilitate the transport of ammonia into symbiosomes. In favor of our speculation, we noted that RHBG-A proteins co-located with endosymbionts (Fig. 3C) while the endosymbionts were abundantly transcribing ammonium transporter genes (the top 10% of highly expressed genes in the InS group). While the endosymbionts were also transcribing genes involved in the glutamine synthetase/glutamate synthase cycle [33], the intimate interaction via ammonium transporter provided a more direct and efficient way to obtain ammonia and was also reported in cnidarian-algal symbiosis [34,35].

While the above findings collectively demonstrate an intimate interaction in the metabolism of both partners, we also observed that the interaction could be compromised when methane supply was limited and when symbiont abundance dropped (Supplementary Fig. S7). Notably, the expression level of these genes involved in metabolic interactions, including *CYP39A1*, *RHBG-A*, and *SLC37A2*, were markedly downregulated in mature bacteriocytes, demonstrating a reciprocal modulation in metabolic coordination.

Coordinated regulatory networks guide molecular functions in bacteriocytes

While the above results collectively show the host optimizing its metabolism and immune processes to facilitate symbiosis, we questioned whether there were coordinated regulatory networks guiding these processes. It was speculated that such regulatory networks would contain the majority of aforementioned metabolic and immune genes, which could share a coordinated expression pattern across all cell types and samples. We then constructed a regulatory network using weighted gene co-expression network analysis (WGCNA) of scRNA-seq data obtained on all the cell clusters in fully-symbiotic (InS group) and decolonized mussels (DeC group). Among all co-expression modules, the Mod12 modules were of interest because they contained 112/199 marker genes of bacteriocytes (Supplementary Fig. S8A, Supplementary Table 5, Supplementary Table 6). Moreover, biological processes undertaken by these 112 element marker genes in module Mod12 were largely identical with those of the 199 marker genes of bacteriocytes (Supplementary Fig. S8B, Supplementary Table 7), suggesting that these markers could be core elements in the regulatory networks of bacteriocytes.

We then screened the highly expressed transcription factors and signal transducers in the regulatory network, which may function as hub regulators of bacteriocyte function. Within the regulatory networks of bacteriocytes, we identified seven transcription factors (zinc finger protein *ZNF271*, ETS-related transcription factor *ELF-3*, autism susceptibility gene *AUST2*, histone-lysine N-methyltransferase *EZH2*, GATA zinc finger domain-containing protein *GATAD14*, hepatocyte growth factor receptor *MET*, and myoneurin) (Fig. 4A). These seven transcription factors were abundantly expressed in bacteriocytes as marker genes and correlated with the expression of 107/199 marker genes (including the aforementioned metabolic genes such as 24-hydroxycholesterol 7 alpha-hydroxylase *CYP39A1*, sugar phosphate exchanger *SLC37A2*, and ammonium transporter *RHBG-A*) and immune genes (such as *TLR2*, rabenosyn-5, lysosomal-trafficking regulator, and cathepsins) (Fig. 4A). Among these seven transcription factors, four genes (*ZNF271*, *ELF-3*, *AUST2*, and *EZH2*) were also vigorously downregulated during decolonization in conjunction with the differential expression of another 63 marker genes (Fig. 4A). These findings highlight the regulatory role of marker gene expression in bacteriocytes. Moreover, strong interconnectivity (weight higher than 0.3) was observed between these transcription factors, indicating that hub genes in bacteriocytes could work cooperatively to promote the efficiency of the regulatory network. The identification of the regulatory network of bacteriocytes, especially the hub regulators, has revealed the molecular basis beneath the biological functions of bacteriocytes, which also provided a unique opportunity to uncover the development of bacteriocyte lineages.

Successive trajectory of the development of bacteriocyte lineages

Without culturable samples of post-larval and juvenile deep-sea mussels, the development process of bacteriocyte lineages remains largely elusive. ScRNA-seq and ST-seq, however, provide a unique opportunity to solve this mechanism in non-model symbiotic associations. As reported in our previous study, there are three types of stem cells in gills of *G. platifrons*, including PEBZCs, DEPCs and VEPCs. By conducting EdU-labeling assay, we confirmed the proliferating activity of DEPCs, which were located at gill base (specifically at the dorsal part of the gill connected to the mantle, with positive EdU signals). Meanwhile, we also noticed that some gill cells dispersed along the gill filaments retained proliferating activity (Fig. 4B) in a manner similar to that observed in fish gills [36]. In addition to the previously identified markers gene, our scRNA-seq and ST-seq data further identified massive cell cycle-, proliferation-, and differentiation-related genes in PEBZCs (cluster 2/6) and DEPCs (cluster 0) [including several markers of stem cells such as DEK [37,38], Serine/arginine-rich splicing factor 1 [39], protein hedgehog [40,41], nesprin-1 [42], protein crumbs-like 2 [43,44], repulsive guidance molecule A [45], G1/S-specific cyclin-E gene [46,47], calcium/calmodulin-dependent protein kinase type II [48,49], prospero homeobox protein 1 [50], Neuroglian [51], SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1 [52,53], bone morphogenic protein type 2 receptor [54], and protein strawberry notch homolog 1 [55]] (Supplementary Table 2). Using an expression atlas of stem cells and bacteriocyte lineages, we then characterized the bacteriocyte development trajectory using three different models: pseudo-time analysis by Monocle (Fig. 4C) and partition-based graph abstraction (PAGA) analysis by Scanpy and RNA velocity analysis (Supplementary Fig. S8C, D). Results by Monocle show that the gill cells were distributed at three distinct pseudo-time states in which most stem cells, as the starting point of differentiation, were in state 3 and 1 of the trajectory (cluster 6 PEBZCs: 92.02% state 3; cluster 2 PEBZCs: 71.05% state 3, 25.22% state 1; cluster 0 DEPCs: 28.44% state 3, 64.52% state 1). Conversely, as differentiation endpoints, approximately 88.19% of bacteriocytes were in state 1 while only 2.52% were in state 3, and 9.29% were in state 1 (Fig. 4C). In addition, as comparison, the ciliated cells (e.g, cluster 4 ICCs) were exclusively in state 3 (59.78%) and state 1 (40.13%). A similar result was also observed in PAGA analysis, in which PEBZCs were found in a more primitive state than DEPCs, ICCs and bacteriocytes (Supplementary Fig. S8C). These distinct pseudo-time trajectory states collectively confirmed developmental heterogeneity within gill cells and suggested the bacteriocytes as one of the most differentiated cells in gill. Besides, it is noticeable that a succession of processes occurred in differentiation (cell type transitions from stem cells to bacteriocytes) and maturation (subtype transitions from state 3 to state 1, or state 1 to state 2) of bacteriocytes (Fig. 4C, Supplementary Fig. S8C).

To further identify the differentiation processes in bacteriocyte lineages, we then examined the abundantly expressed genes in different pseudo-time states (Supplementary Table 8). We speculated that progenitor states would encode several hub marker genes (especially transcription factors) of descendent cells, while the expression levels of marker genes would increase during further maturation, showing maximal expression levels in the function-matured state. Our results indicate that bacteriocytes in state 2 were in the function-matured state, transcribing more marker genes (including hub marker genes) compared with the levels transcribed in the rest of the states (Supplementary Fig. S8E, Supplementary Table 8). Bacteriocytes in state 1 were in function-maturing state and showed moderate expression of marker genes. Interestingly, we also noted that most of the genes abundantly expressed in state 2 PEBZCs (cluster 2: 100/314, cluster 6: 29/32) and DEPCs (129/755) were the marker genes expressed by bacteriocytes (a total of 199 markers); also, 29/43 of DEPCs marker genes were abundantly expressed in state 1 PEBZCs (cluster 2) (Supplementary Fig. S8E, Supplementary Table 8). These findings suggest that the state 2 cells of both PEBZCs and DPECs may have been progenitors of mature bacteriocytes, while DEPCs may have descended from state 1 PEBZCs. The conclusion was also supported by RNA velocity analysis, in which a positive velocity from PEBZCs and DPECs to bacteriocytes was observed in corresponding cells of adjacent region (Supplementary Fig. S8D). We also observed that state 3 mature bacteriocytes encoded multiple DNA replication- and cell cycle-related genes (Supplementary Table 8), indicating that mature bacteriocytes could proliferate directly. In support of this hypothesis, we observed increased proliferation-related signaling in several mature bacteriocytes, as assessed using our EdU-labeling assay and 3D electron microscopy (Fig. 4D, Supplementary Video.2) [56].

Co-option of conserved transcription factors in bacteriocyte development

While the development of symbiotic cells and organs may be a highly organized process, the regulatory networks guiding these developmental processes are difficult to identify in either model and non-model symbiotic associations. In this study, using our bacteriocyte developmental trajectory, we explored the mechanisms underlying bacteriocyte differentiation and maturation while focusing on the genes abundantly expressed in the progenitor state. By conducting t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis, we noted a co-overlap in the t-SNE distribution of state 1 bacteriocytes and state 2 cells of DEPCs/PEBZCs and a co-overlap between state 1 cells of DEPCs and state 1 cells of PEBZCs (Fig. 5A), reconfirming that the state 2 cells of DEPCs/PEBZCs progenitor were the progenitor cells of bacteriocytes. Specifically, our results indicate that mRNA transcripts of histone-lysine N-methyltransferase *EZH2*, one of the hub marker genes in the regulatory network of mature bacteriocytes, were abundantly expressed in state 2 of all types stem cells (Fig. 5B, Supplementary Table 8). Additionally, the expression of dedicator of cytokinesis protein 11, acid phosphatase type 7, cathepsin D, cathepsin L and prosaposin, which serve as element marker genes in the mature bacteriocyte regulatory network, and as crucial regulators of cell differentiation and lysosomal function [57-59], was also highly upregulated in state 2 of all types stem cells (Supplementary Table 8). These genes may collectively promote the differentiation of state 2 stem cells into bacteriocytes. In addition, the expression of another three hub transcription factors in the regulatory network of bacteriocytes (*ZNF271*, *ELF-3*, and *MET*), was also robustly promoted in state 2 DEPCs (Supplementary Table 8), thereby contributing to the differentiation of DEPCs into bacteriocytes. Similarly, we observed a robust expression of *myoneurin*, *ZNF271*, *ELF-3 and MET* in state 2 of cluster 2 PEBZCs (Supplementary Table 8). These genes may facilitate the transition of stem cells into bacteriocytes by transducing intracellular signals and activating the expression of bacteriocyte-specific genes.

During further maturation of bacteriocytes, we observed a gradual increase in the expression of four hub marker genes (*ZNF271*, *EZH2*, *MET,* and *AUST2*) from state 1 to state 2 (Fig. 5C, Supplementary Table 8). Additionally, 21 transcription factors (or activators/repressors) were also highly expressed in state 1 bacteriocytes (Supplementary Table 8). These genes may collaboratively promote function maturation in bacteriocytes. Interestingly, we also observed that delta(14)-sterol reductase (*TM7SF2*, EC:1.3.1.70*)* was abundantly expressed in state 1 (Fig. 5C, Supplementary Table 8). The *TM7SF2* gene plays crucial role in the sterol metabolism of mussel host by transforming symbiont-deprived 4,4-dimethyl-cholesta-8,14,24-trienol into 14-Demethyl-lanosterol, demonstrating the participation of symbionts in bacteriocyte maturation.

While the aforementioned hub transcription factors and signaling transducers may be the pioneer molecules promoting bacteriocyte development, we questioned whether they were evolutionarily novel genes diverged from non-symbiotic ancestors or conserved across evolution. Our phylogenetic analysis of protein sequences obtained from homologues of these genes in Lophotrochozoa indicated that all of these genes were highly conserved across the mollusk and in harmony with the general host phylogenies (Supplementary Fig. S9). We therefore conclude that bacteriocyte development occurred via co-option of conserved genes rather than via evolutionarily novel genes (Fig. 5E).

Discussion

In addition to being an evolutionary novelty, symbiotic cells and organs provide a suitable niche for symbiosis and endow holobionts with a unique adaptation to surrounding environments. In this study, we integrated the data obtained on single-cell /spatial transcriptomes, meta-transcriptomes, ISH/IHC of key genes, EdU labeling assay and 3D electron microscopy to uncover the function and development of symbiotic cells in deep-sea mussels. Our findings indicate that these processes were guided by an ancestral intrinsic toolkit via co-option of conserved transcription factors and modulation by endosymbionts. Our results show how both partners interact with each other closely and shape the function and development of symbiotic cells cooperatively. The present study also propose a roadmap to address the formation and development of symbiotic cells and organs in all non-model symbiotic deep-sea mollusks, which is crucial in understanding and protecting the biodiversity of marine fauna from the growing anthropogenic activities in habitats, such as the deep sea [60,61].

The establishment and maintenance of endosymbiosis with exogenous bacteria is challenging for multicellular organisms because of the host immune system, which requires an adjustment that entails an inoculation with, and proliferation of, symbionts inside the cell [62]. Using a single-cell expression atlas, we characterized the immune strategy adopted by bacteriocyte lineages, and observed that only a few PRRs were abundantly expressed in bacteriocytes despite their expansion in the deep-sea mussel genome [18]. While these PRRs might participate in other process, such as early symbiotic acquisition in juveniles and immune defense again pathogenic bacteria, our findings suggested that only a few PRRs are needed to maintain symbiosis in adult mussels. We also confirmed that bacteriocytes can phagocytize exogenous bacteria, which directly facilitate the entrance of endosymbionts [30]. It is also noticeable that phagocytotic ability of gill cells may be conserved in most mollusks, including non-symbiotic ones [30,63,64], which may contribute importantly to the establishment of endosymbiosis.

It is recognized that the formation of symbiotic associations in deep-sea mollusks are mostly nutrition driven [65]. However, whether and how molluscan symbiotic cells adjusted their metabolism to acquire nutrition from symbionts remains unclear. Previous studies have described the nutritional contributions of symbionts, but the mechanisms used by molluscan hosts to support symbiont metabolism also remain uncharacterized [18,19,22,28,66]. With the help of state-of-the-art single-cell sequencing and IHC/FISH assay, we here showed that deep-sea mussels have remarkably reshaped bacteriocyte metabolism to maximize symbiotic profits for both partners. For example, bacteriocytes encode dozens of genes involved in the biosynthesis and transport of carbohydrates, lipids, amino acids, and vitamins to improve the acquisition of nutrition from symbionts. Indeed, bacteriocytes can retrieve fructose-6P in direct cooperation with methanotrophic symbionts via expression of sugar phosphate exchanger genes. Fructose-6P is then converted into glucose-6P, fructose-1,6P_2, and glyceraldehyde-3P before being redistributed back to both the host and symbionts as supplements of gluconeogenesis and TCA cycles. We have also shown that the host directly supplies ammonia to symbionts via ammonium transporters. The bioconversion and exchange of sugar phosphate and ammonia between the host and symbionts demonstrates an efficient and direct way of coordinating the metabolism of both partners, which greatly increases the efficiency and profits of symbiosis and could be widely adapted by other symbiosis systems[35].

Notably, we rarely observed proliferating endosymbionts or free symbionts that being engulfed in bacteriocytes of adult mussels, as evidenced by EdU labeling assay and 3D electron microscopy. Since both our scRNA-seq/ST-seq data and previous studies revealed the massive transcription of lysosome-related genes and high lysosomal activity in mature bacteriocytes[67], we speculated that lysosome-mediated symbiont digestion may be a secondary option to obtain nutrition under normal circumstance, but an efficient way to control the symbiont population and obtain nutrition under abrupt stresses or other emergencies. In support of this speculation, we observed a significant repression of lysosome activity in mussels containing a limited number of symbionts after long-term methane starvation (DeC group). Moreover, lysosome-mediated degradation of symbionts was found to be augmented shortly after depressurization stress or in combined treatment of methane deprivation and atmospheric cultivation [68,69].

The development of symbiotic cells provides an elegant and stable niche for endosymbiosis, and is crucial for understand the formation of symbiosis [2,6]. Our single-cell transcriptome data demonstrate the successive development trajectory of bacteriocytes in deep-sea mussels. This trajectory was controlled by several conserved transcription factors (including *EZH2*, *ELF-3*, *ZNF271*, and *AUST2*) that were also responsible for bacteriocyte molecular functions. The synergistic modulation shown by these transcription factor genes confirmed that the formation of symbiotic cells in deep-sea mussels occurred via co-option of conserved, rather than evolutionarily novel genes [70]. Interestingly, similar phenomena are also observed in the deep-sea scaly-foot snail, in which the formation of biomineralized skeleton is driven by an ancestral intrinsic toolkit that is conserved across the mollusk [71]. These findings collectively demonstrate the developmental plasticity of the mollusk. The function and development of mussel bacteriocytes may have been influenced by symbionts via sterol metabolism. Symbiont participation in the development of symbiotic cells and organs has attracted increasing attention for several years, but has remained largely unclear at the mechanistic level [72,73]. Sterol metabolism in symbionts may be a common mechanism in the development of symbiotic cells and organs since several animal hosts rely on their symbionts for sterol intermediates [74-76]. In support of this theory, we noted a suppression of sterol/steroid biosynthesis in the symbiont-depleted deep-sea mussels after long-term atmospheric cultivation without methane supply; furthermore, a crucial sterol metabolic gene was differentially expressed during the maturation of bacteriocytes, highlighting possible participation of sterol metabolism in the phenotypic plasticity of gill tissue [69]. Other studies have also shown the modulation of host development and reproduction by symbiotic *Wolbachia* via steroid-nuclear receptor signaling pathway [77]. While the regulatory role of symbionts in host development remains to be fully elucidated, our findings highlight the robust plasticity of molluscan symbiotic cells, which assist in the distribution of mollusk across a wide range of habitats including the deep sea.

There are also some limitations for the present study. For example, it should be noted that we employed an *in situ* transplantation assay with experimental period of almost two years to obtain the decolonized mussels, which complicated the characterization of biological response of bacteriocytes during symbiont depletion since previous studies showed that both methane deprivation and tidal cycles could influence the expression pattern of mussel host [20,78]. While the experimental design reflects a common challenge for studies on deep-sea organisms, the symbiosis-related processes concluded from our single-cell transcriptome and meta-transcriptome data are distinctly different with the rhythm-related processes, suggesting that the expression shifts after symbiont decolonization were more directly associated with methane alternations rather than temporal variations. In addition, some hypotheses raised by the present study still needs to be verified in other symbiotic deep-sea holobionts in future study and with more mussel samples in different symbiotic or aposymbiotic states including recolonized mussels as well as post-larval and juvenile mussels. Nevertheless, by combing high-resolution single-cell transcriptome data and IHC/FISH/EM data, we have successfully decoded the molecular function, and development trajectory of bacteriocyte lineages in deep-sea mussels. Our results show that the function and development of bacteriocytes were highly coordinated via co-option of conserved genes and could be dynamically modulated by symbionts. The intimate symbiotic associations and robust plasticity of symbiotic cells have rendered the deep-sea mussel one of the most successful organisms in the deep sea.

Materials and Methods

Experimental design

The goal of the present study was to address the molecular function and developmental trajectory of bacteriocytes in the deep-sea mussel *G. platifrons*. To achieve this goal, we employed methanotrophic *G. platifrons* collected from cold seeps as our model, and used the spatial transcriptome, single-cell transcriptome, and meta-transcriptome to characterize both the bacteriocytes and endosymbionts. We also used an *in situ* transplantation assay to construct the decolonized mussel to elucidate the dynamic response of symbiotic cells in response to symbiont depletion. We employed EdU-labeling, phagocytosis, *in situ* hybridization (ISH) assays, immunofluorescence (IF) assays, transmission electron microscopy (TEM), 3D electron microscopy and phylogenetic analysis to verify hypotheses obtained from the sequencing data.

In situ transplantation and animal collection

G. platifrons mussels were collected from the cold seeps (22°06′N, 119°17′E) of South China Sea during the cruises of 2018 and 2020. The mussel fauna live at 1,120 m beneath the surface at a temperature of approximately 3.35°C, salinity of approximately 35.54 psu, dissolved oxygen of 2.98–3.17 mg/L, and methane concentrations of up to 31,227 ppm (seepage region) [79]. To shelter the specimens from temperature and pressure fluctuations during sampling, all mussel samples were collected using a self-designed isothermal isobaric sampler and a self-designed manually controlled macrofauna *in situ* sampling device as described previously [68]. In brief, a total of 20 mussels living in close proximity to active fluid seepages were collected as representatives of normal mussels (designated as the InS group). Among them, seven mussels were collected using an isothermal isobaric sampler (in 2018), seven mussels were collected using a multipurpose *in situ* sampling device (in 2020) after treating with RNAsafer stabilizer reagent (Omega Bio-Tek, Norcross GA, USA) *in situ*, and the remaining six mussels were collected using a multipurpose *in situ* sampling device after being treated with 4% paraformaldehyde *in situ*. Only one mussel from the InS group was used for single cell transcriptome sequencing (scRNA-seq, Chromium platform of 10x Genomics, Pleasanton CA, USA) and spatial transcriptome sequencing (ST-seq, Visium platform of 10x Genomics), and all 14 mussels collected by isothermal isobaric sampler or after RNA stabilizing treatment were subjected to meta-transcriptome sequencing. Mussels collected after paraformaldehyde treatment were subjected to ISH, IF, and TEM imaging. During the 2018 cruise, we also performed an *in situ* transplantation assay, in which dozens of mussels in the seepage region were translocated to an authigenic carbonate region having low concentration of CH_4 (at approximately 100 m away from seepage, methane concentration had decreased to 800 ppm). Five of the transplanted mussels (designated as the DeC group) were retrieved 604 days later during the 2020 cruise using an isothermal isobaric sampler (one mussel, used for scRNA-seq, ST-seq and meta-transcriptome) or multipurpose *in situ* sampling device (four mussels, used for meta-transcriptome) after treatment with RNAsafer stabilizer reagent. After the retrieval and depressurization of the isothermal isobaric sampler, the mussels were instantly dissected into two halves to remove excess seawater and quickly frozen in an isopentane bath (Macklin, Shanghai, China) with liquid nitrogen for scRNA-seq and ST-seq.

ScRNA-seq

Two individual mussels having similar size (approximately 80 mm in length) from the InS and DeC groups were used for scRNA-seq. Because of the difficulties involved in isolation and *in vitro* culture of gill cells, and because mussel sampling and single cell preparation can potentially influence gene expression, scRNA-seq analysis of *G. platifrons* gill tissue was performed using nuclei isolated from fresh frozen samples collected in an isothermal isobaric manner. The isolated nuclei were assayed per 10×Genomics single cell protocol by generating a Single-cell Gel Bead-In-EMulsion (GEM) on a GemCode single-cell instrument (https://www.10xgenomics.com/support). Full-length cDNA was synthesized using Chromium Next GEM Single Cell 3’ Reagent Kit v3.1 and sequenced on an Illumina HiSeq X Ten platform (Gene Denovo Biotechnology Co., Guangzhou, China).

Cryosectioning and ST-seq

To help determine the taxonomy of all gill cells, especially bacteriocytes, we also conducted ST-seq with the cross-sectioned middle region of the gill (ventral view) from the InS group. The gill tissue used for 10×Visium ST-seq was first incubated with precooled methanol (4°C) for 15 min to reduce potential damage by the cryostat blade during cryosectioning. The tissue was then embedded in OCT (Sakura, Torrance, CA, USA) and cross-sectioned at the thickness of 10 μm through the middle part of the gill using a cryostat (Leica CM1950, Heidelberger, Germany). Gill sections were then stained with 4,6-diamidino-2-phenylindole (DAPI, Thermo Fisher, Waltham, MA, USA) and HE (Sangon Biotech, Shanghai, China), and morphology of the gill-tissue sections was assessed under light microscopy (Nikon ECLIPSE Ni, Tokyo, Japan). Successive sections, obtained from the same blocks of embedded tissue, were placed on Visium Spatial Tissue Optimization Slides within the capture area for tissue optimization assay. For the sequencing assay, gill sections were first mounted on Visium Spatial Gene Expression Slides and permeabilized according to previously described parameters (permeabilization time of 9 min) after HE staining and imaging using light microscopy. After reverse transcription and library preparation, all samples were sequenced on an Illumina HiSeq X Ten platform (Gene Denovo Biotechnology Co., Guangzhou, China). The rest sections of the same tissue were further subjected to ISH assays to assist the cell type annotation.

Cell clustering, cell type annotation, and analysis of differentially expressed genes

Single-cell clustering of *G. platifrons* gill tissue was first performed using scRNA-seq—derived data. After quality control, all raw reads were mapped onto a *G. platifrons* reference genome. The genome was first reported by Sun et al. [18] and updated recently with Hi-C and high-depth PacBio long-read sequencing by us (GenBank accession NO. JAOEFJ000000000, unpublished data). Unique molecular identifiers (UMIs) in each sequenced read were counted and corrected for sequencing errors. Using valid barcodes that were identified based on EmptyDrops method, the gene matrices of all the cells were then produced and imported into Seurat (version 3.1.1) for cell clustering using a graph-based clustering approach.

Single-cell clustering of *G. platifrons* gill tissue was also performed using ST-seq—derived data, which have the advantage of cell type annotation. A slide image obtained before permeabilization was first imported into Space Ranger software for fiducial and Visium barcoded spot alignment. After decoding correlations between tissue, capture spots, and barcodes, a splice-aware alignment of sequencing reads to the *G. platifrons* genome was performed using STAR in the Space Ranger software package.

An integrated analysis of data derived using scRNA-seq, ST-seq, and *in situ* hybridization assay was then used for cell type annotation in the InS group. In brief, anchor-based integration was first used to integrate ST-seq data with scRNA-seq data using the FindIntegrationAnchors command in Seurat-v3.2. All cell-type labels in scRNA-seq were then transferred to spatial data using the TransferData command. Cell type prediction scores indicating the similarity between ST-seq spots and scRNA-seq cell clusters were calculated simultaneously, and only spot-cluster pairs with highest scores were considered for further cell-type annotation. For further annotation and cell-type verification, cell types in gill tissue were inspected using HE stained images of gill cryosections and tissue plot with spots colored by clustering of ST-seq data. Several cell markers obtained using the FindAllMarkers function in Seurat were also cloned and subjected to *in situ* hybridization using gill tissues from the InS group.

To further investigate the biological function of all identified cell clusters, differentially expressed genes (DEGs) were surveyed in both scRNA-seq and ST-seq. For scRNA-seq, expression values of all identified genes in a given cluster were compared against the rest of the cell clusters using Wilcoxon rank sum test. Genes expressed mainly in the target cluster (more than 25% of cells were designated as a target cluster), and showing at least 1.28-fold upregulated expression levels and *p*-values less than 0.01, were considered as DEGs per cell cluster. For ST-seq, the Model-based Analysis of Single cell Transcriptomics (MAST) in the R package was used to determine DEGs in a single cell cluster using the same criteria as those used for scRNA-seq. All the genes in ST-seq were further analyzed for spatially-specific DEGs using markvariogram in the Seurat R package and mark-segregation hypothesis testing in the trendsceek R package. Genes with r.metric.5 parameter values of less than 0.8 and *p*-values of less than 0.01 were designated as spatially DEGs.

Meta-transcriptome sequencing

To explore the expression atlas of symbionts in the InS and DeC groups, we performed meta-transcriptome sequencing using the 14 samples from the InS group and five samples from the DeC group as described previously. Total RNA extraction, removal of eukaryotic and prokaryotic rRNA, and synthesis of cDNA library using bulk-seq were conducted as described previously [68]. cDNA libraries belonging to the InS and DeC groups was finally sequenced using an Illumina HiSeq 2500 platform with paired-end reads performed by Novogene (Tianjing, China). After quality control, the filtered reads were aligned against both the endosymbiont and *G. platifrons* genomes using HISAT (v2.0.4). The genome of methanotrophic symbionts was first reported by Takishita et al., (2017) and we updated it with high-depth PacBio long-read sequencing (raw data deposited in NCBI with accession NO. PRJNA891367) [80]. The expression levels of host and symbiont genes were calculated using HTSeq (v0.6.1), while significant differences between groups were determined using DESeq2 (v1.10.1). Additionally, the top 10% of genes with most abundant mRNA transcripts in the InS group were designated as abundantly expressed genes in order to survey the biological processes that actively occurred under normal conditions *in situ*. For PCoA analysis, Bray-Curtis distance between samples was first calculated and subjected for subsequent analysis and visualization.

GO/KEGG analysis, cell trajectory construction, and WGCNA of bacteriocytes

Gene Ontology (GO) annotation of *G. platifrons* genes was obtained using Blast2GO software (version 5.2) and employed for GO enrichment analysis using homemade scripts. DEG numbers for every GO term were first calculated and the significantly enriched GO terms were determined using a hypergeometric test (FDR less than 0.05). KEGG pathway enrichment analysis was conducted using homemade scripts with KEGG annotations obtained from the KEGG database (http://www.genome.jp/kegg/pathway.html).

Single cell trajectory analysis was conducted using Monocle (Version 2.6.4) with reversed graph embedding algorithm and confirmed by PAGA analysis using Scanpy (v1.6.0). Gene expression matrix of stem cells and bacteriocytes was used in the analysis and visualized using the orderCells function of Monocle (sigma = 0.001, lambda = NULL, param. gamma = 10, tol = 0.001). For the PAGA analysis, the connectivity of each cell cluster was calculated based on the partition-based graph abstraction algorithm. After cell embedding with ForceAtlas2, the pseudo-time value of each cell was then calculated with the DPT algorithm. Weighted gene co-expression network analysis (WGCNA) was conducted using the WGCNA package (v1.47) in R with gene expression values of power = 7 and minModuleSize = 50, and visualized using Cytoscape (v3.8.2).

EdU-labeling, phagocytosis, ISH assays, IF assays, 3D electron microscopy, and phylogenetic analysis

The EdU-labeling assay was carried out *in situ* using mussels collected from the seepage region. In brief, mussels were incubated with 5-ethynyl 2’-deoxyuridine (EdU, final concentration of 40 μM) in a self-designed, manually controlled macrofauna *in situ* experiment device for approximately 18 h, and then retrieved using an isothermal isobaric sampler. The Click-iT Plus EdU Imaging Kit (Thermo Fisher) was then used to visualize the EdU signal.

Phagocytosis assay was conducted using primary gill cells obtained from the fresh collected mussels using previously described methods with modifications [63,81]. Because endosymbiotic methanotrophs are nonculturable, *Vibrio alginolyticus*, an environmental bacterium isolated from the cold seep, was used in this assay and labeled with fluorescein isothiocyanate (FITC, Sigma, St. Louis, MO, USA) before use. To collect primary gill cells, gill tissue was first treated with 1% trypsin (diluted in sterilized seawater) at 4°C for 30 min, and then successively centrifuged at 300 g (4°C, 5 min) and 800 g (4°C, 5 min). Cell pellets were resuspended in modified L15 medium (Gibco, Carlsbad, CA, USA; supplemented with 0.54 g/L KCl, 0.6 g/L $CaCl_2$, 1 g/L $MgSO_4$, 3.9 g/L $MgCl_2$, and 20.2 g/L NaCl) to a final concentration of 1×10^6 cells mL^{-1} and incubated with the same volume of FITC-labeled *V. splendidus* (1×10^8 cells mL^{-1}) for 30 min at 4°C in the dark. Primary cells were then washed three times with modified L15 medium to remove extracellular bacteria. Cells were stained with DAPI and DiI perchlorate, and were then imaged using a laser scanning confocal microscope (Zeiss LSM710, Jena, Germany).

ISH assays were conducted as described previously [82]. For cell-type verification, cell markers were cloned using gene-specific primers to synthesize digoxigenin-labeled ISH probes. Fluorescent ISH analysis of symbionts was performed using a Cyanine 3 (Cy3)-labeled Eub338 eubacteria probe (5’-GCTGCCTCCCGTAGGAGT-3’) or FITC-labeled pmoB (methanotroph-specific gene) probe (5’-CGAGATATTATCCTCGCCTG-3’).

For the IF assay, unique peptide fragments of the 24-hydroxycholesterol 7 alpha-hydroxylase *CYP39A1*, sugar phosphate exchanger *SLC37A2*, and ammonium transporter *RHBG-A* were first synthesized by Sangon Biotech (Shanghai, China) and employed as antigens to produce rabbit polyclonal antibodies. After antigen affinity purification and specificity verification by ELISA and western blot, the antibodies were then used in an IF assay using paraffin embedded gill tissue (5 μm) according to methods described previously [82] and with the assistance of ServiceBio (Wuhan, China). Specifically, anti-lipid A antibody (ab8467, Abcam, Cambridge, MA, USA) was used to indicate endosymbionts. After incubation with FITC- and Cy3-labeled secondary antibodies (Servicebio, Wuhan, China), gill sections were visualized and imaged using fluorescent microscope.

For the 3D electron microscopy assay, gill tissues collected from adult mussels (obtained in isothermal way) were fixed by paraformaldehyde-glutaraldehyde, treated with 1.0% osmium tetroxide (OsO4) and infiltrated with acrylic resin successively. The embedded tissues were then serially sectioned with a thickness of 100 nm using an ultramicrotome (EM UC7, Leica, Vienna, Austria) and mounted onto a silicon wafer before imaging. The scanning electron microscope micrographs of mounted sections were captured using Helios NanoLab 600i FIB-SEM (FEI, Hillsboro, USA). The obtained serial section images were then aligned in Amira (version 2019.3) for the 3D reconstruction of bacteriocytes and cell proliferation analysis.

For phylogenetic analysis of hub transcription factors, homologue proteins were first obtained using NCBI blastp and aligned using Seaview. Maximum likelihood phylogenetic trees of these proteins were then constructed using Mega software (v11) in Jones-Taylor-Thornton model with bootstrap of 100.

Acknowledgments

We thank the crewmembers of the R/V Kexue for their assistance in sample collection, and all the laboratory staff for continuous technical advice and helpful discussions. We also thank Prof. Qiang Lin from the South China Sea Institute of Oceanology, Chinese Academy of Sciences for insightful comments and suggestions that improved the manuscript.

Funding: The present study was supported by the National Natural Science Foundation of China (42106134, 42030407, and 42076091), the Key Research Program of Frontier Sciences (ZDBS-LY-DQC032) and Strategic Priority Research Program of the Chinese Academy of Sciences (XDA22050303, XDB42020401), Key

Deployment Project of Centre for Ocean Mega-Research of Science, CAS (COMS2020Q02) and Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2022QNLM030004).

Author contributions: Conceptualization: H.C., Chaolun Li; Sample collection: M.W., M.L., Z.Z., Chao Lian., L.Z.; Environment investigation: M.W., L.C.; Transcriptome analysis: H.C., Z.Z., M.L.; EdU labeling, phagocytosis, ISH and IF assays: H.C., M.L., G.H., Z.Z., H.W.; Phylogenetic analysis: H.C., H.Z.; Supervision: C.L.; Writing—original draft: H.C.; Writing—review & editing: H.C., M.W., Chaolun Li.

Conflict of interest: The authors declare that they have no competing interests.

Data Accessibility: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials**.** All sequencing data were deposited in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession NO. PRJNA838712. Raw images for 3D electron microscopy were deposited in Figshare (https://figshare.com/articles/figure/Raw_image_for_3D_electron_microscopy_of_mussel_gill/23575734).

References

1. Douglas, A.E. (2014). Symbiosis as a general principle in eukaryotic evolution. Cold Spring Harbor perspectives in biology ***6***, 10.1101/cshperspect.a016113.
2. Gilbert, S.F., Bosch, T.C.G., and Ledón-Rettig, C. (2015). Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. Nature Reviews Genetics ***16***, 611-622, 10.1038/nrg3982.
3. Nyholm, S.V., and McFall-Ngai, M.J. (2021). A lasting symbiosis: how the *Hawaiian bobtail* squid finds and keeps its bioluminescent bacterial partner. Nature Reviews Microbiology ***19***, 666-679, 10.1038/s41579-021-00567-y.
4. Smith, T.E., and Moran, N.A. (2020). Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and Buchnera. Proceedings of the National Academy of Sciences ***117***, 2113-2121, 10.1073/pnas.1916748117.
5. Rosset, S.L., Oakley, C.A., Ferrier-Pagès, C., et al. (2021). The molecular language of the cnidarian–dinoflagellate symbiosis. Trends in Microbiology ***29***, 320-333, 10.1016/j.tim.2020.08.005.
6. Douglas, A.E. (2020). Housing microbial symbionts: evolutionary origins and diversification of symbiotic organs in animals. Philosophical Transactions of the Royal Society B: Biological Sciences ***375***, 20190603, 10.1098/rstb.2019.0603.
7. McFall-Ngai Margaret, J., and Ruby Edward, G. (1991). Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. Science ***254***, 1491-1494, 10.1126/science.1962208.
8. Braendle, C., Miura, T., Bickel, R., et al. (2003). Developmental origin and evolution of bacteriocytes in the aphid–Buchnera symbiosis. PLoS Biology ***1***, e21, 10.1371/journal.pbio.0000021.
9. Matsuura, Y., Kikuchi, Y., Miura, T., and Fukatsu, T. (2015). Ultrabithorax is essential for bacteriocyte development. Proceedings of the National Academy of Sciences ***112***, 9376-9381, 10.1073/pnas.1503371112.
10. Dubilier, N., Bergin, C., and Lott, C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nature Reviews Microbiology ***6***, 725, 10.1038/nrmicro1992.

11. Duperron, S., Gaudron, S.M., Rodrigues, C.F., et al. (2013). An overview of chemosynthetic symbioses in bivalves from the North Atlantic and Mediterranean Sea. Biogeosciences ***10***, 3241-3267, 10.5194/bg-10-3241-2013.

12. Rodrigues, C.F., Laming, S.R., Gaudron, S.M., et al. (2015). A sad tale: has the small mussel Idas argenteus lost its symbionts. Biological Journal of the Linnean Society ***114***, 389-405.

13. Sun, J., Zhang, Y., Xu, T., et al. (2017). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature ecology & evolution ***1***, 121, 10.1038/s41559-017-0121.

14. Fontanez, K.M., and Cavanaugh, C.M. (2014). Evidence for horizontal transmission from multilocus phylogeny of deep-sea mussel (Mytilidae) symbionts. Environmental microbiology ***16***, 3608-3621, 10.1111/1462-2920.12379.

15. Miyazaki, J.-I., Martins, L.d.O., Fujita, Y., et al. (2010). Evolutionary process of deep-sea bathymodiolus mussels. PloS one ***5***, e10363, 10.1371/journal.pone.0010363.

16. Duperron, S., Quiles, A., Szafranski, K.M., et al. (2016). Estimating symbiont abundances and gill surface areas in specimens of the hydrothermal vent mussel *Bathymodiolus puteoserpentis* maintained in pressure vessels. Frontiers in Marine Science ***3***, 10.3389/fmars.2016.00016.

17. Martins, I., Colaço, A., Dando, P.R., et al. (2008). Size-dependent variations on the nutritional pathway of *Bathymodiolus azoricus* demonstrated by a C-flux model. Ecological Modelling ***217***, 59-71, 10.1016/j.ecolmodel.2008.05.008.

18. Sun, J., Zhang, Y., Xu, T., et al. (2017). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature Ecology & Evolution ***1***, 121.

19. Ponnudurai, R., Kleiner, M., Sayavedra, L., et al. (2016). Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. The ISME journal 10.1038/ismej.2016.124.

20. Wang, H., Zhang, H., Wang, M., et al. (2019). Comparative transcriptomic analysis illuminates the host-symbiont interactions in the deep-sea mussel *Bathymodiolus platifrons*. Deep Sea Research Part I: Oceanographic Research Papers ***151***, 103082, 10.1016/j.dsr.2019.103082.

21. Sun, Y., Wang, M., Chen, H., et al. (2022). Insights into symbiotic interactions from metatranscriptome analysis of deep-sea mussel *Gigantidas platifrons* under long-term laboratory maintenance. Molecular ecology ***n/a***, 10.1111/mec.16765.

22. Geier, B., Sogin, E.M., Michellod, D., et al. (2020). Spatial metabolomics of in situ host–microbe interactions at the micrometre scale. Nature Microbiology 1-13.

23. Levy, S., Elek, A., Grau-Bové, X., et al. (2021). A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. Cell ***184***, 2973-2987.e2918, 10.1016/j.cell.2021.04.005.

24. Alacid, E., and Richards, T.A. (2021). A cell–cell atlas approach for understanding symbiotic interactions between microbes. Current Opinion in Microbiology ***64***, 47-59, 10.1016/j.mib.2021.09.001.

25. Hu, M., Zheng, X., Fan, C.-M., and Zheng, Y. (2020). Lineage dynamics of the endosymbiotic cell type in the soft coral Xenia. Nature ***582***, 534-538, 10.1038/s41586-020-2385-7.

26. Wang, H., He, K., Zhang, H., et al. (2023). Deciphering deep-sea chemosynthetic symbiosis by single-nucleus RNA- sequencing. eLife 12:RP88294, 10.7554/eLife.88294.1.

27. Chu, H., and Mazmanian, S.K. (2013). Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nature Immunology ***14***, 668-675, 10.1038/ni.2635.

28. Takishita, K., Takaki, Y., Chikaraishi, Y., et al. (2017). Genomic evidence that methanotrophic endosymbionts likely provide deep-sea bathymodiolus mussels with a sterol intermediate in cholesterol biosynthesis. Genome Biology and Evolution ***9***, 1148-1160, 10.1093/gbe/evx082.

29. Ikuta, T., Amari, Y., Tame, A., et al. (2021). Inside or out? Clonal thiotrophic symbiont populations occupy deep-sea mussel bacteriocytes with pathways connecting to the external environment. ISME Communications ***1***, 38, 10.1038/s43705-021-00043-x.

30. Tame, A., Maruyama, T., and Yoshida, T. (2022). Phagocytosis of exogenous bacteria by gill epithelial cells in the deep-sea symbiotic mussel *Bathymodiolus*

japonicus. Royal Society Open Science ***9***, 211384, 10.1098/rsos.211384.

31. Mohd Noor, S.N., Day, D.A., and Smith, P.M. (2015). The symbiosome membrane. Biological Nitrogen Fixation 683-694, 10.1002/9781119053095.ch68.

32. Sproles, A.E., Kirk, N.L., Kitchen, S.A., et al. (2017). Phylogenetic characterization of transporter proteins in the cnidarian-dinoflagellate symbiosis. Molecular Phylogenetics and Evolution ***120***, 307-320, 10.1016/j.ympev.2017.12.007.

33. Cui, G., Liew, Y.J., Li, Y., et al. (2019). Host-dependent nitrogen recycling as a mechanism of symbiont control in Aiptasia. PLoS genetics ***15***, e1008189, 10.1371/journal.pgen.1008189.

34. Fochi, V., Chitarra, W., Kohler, A., et al. (2017). Fungal and plant gene expression in the *Tulasnella calospora–Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. New Phytologist ***213***, 365-379, 10.1111/nph.14279.

35. Cui, G., Konciute, M.K., Ling, L., et al. Molecular insights into the Darwin paradox of coral reefs from the sea anemone Aiptasia. Science Advances ***9***, eadf7108, 10.1126/sciadv.adf7108.

36. Stolper, J., Ambrosio, E.M., Danciu, D.P., et al. (2019). Stem cell topography splits growth and homeostatic functions in the fish gill. eLife ***8***, e43747, 10.7554/eLife.43747.

37. Koleva, R.I., Ficarro, S.B., Radomska, H.S., et al. (2012). C/EBPα and DEK coordinately regulate myeloid differentiation. Blood ***119***, 4878-4888, 10.1182/blood-2011-10-383083.

38. Yue, L., Wan, R., Luan, S., et al. (2020). Dek modulates global intron retention during muscle stem cells quiescence exit. Developmental cell ***53***, 661-676.e666, 10.1016/j.devcel.2020.05.006.

39. Chung, Y.R., Cho, H., Kim, Y.J., et al. (2016). Serine/Arginine-Rich Splicing Factor 1 (SRSF1) is required for adult and embryonic hematopoiesis and has non-overlapping roles with SRSF2 in hematopoiesis. Blood ***128***, 1478, 10.1182/blood.V128.22.1478.1478.

40. Katoh, Y., and Katoh, M. (2006). Hedgehog signaling pathway and

gastrointestinal stem cell signaling network (Review). International Journal of Molecular Medicine ***18***, 1019-1023, 10.3892/ijmm.18.6.1019.

41. Parisi, M.J., and Lin, H. (1998). The role of the hedgehog/patched signaling pathway in epithelial stem cell proliferation: from fly to human. Cell research ***8***, 15-21, 10.1038/cr.1998.2.

42. Yang, W., Zheng, H., Wang, Y., et al. (2013). Nesprin-1 plays an important role in the proliferation and apoptosis of mesenchymal stem cells. International Journal of Molecular Medicine ***32***, 805-812, 10.3892/ijmm.2013.1445.

43. Pellissier, L.P., Alves, C.H., Quinn, P.M., et al. (2013). Targeted ablation of crb1 and crb2 in retinal progenitor cells mimics leber congenital amaurosis. PLoS genetics ***9***, e1003976, 10.1371/journal.pgen.1003976.

44. Boroviak, T., and Rashbass, P. (2011). The apical polarity determinant crumbs 2 is a novel regulator of ESC-derived neural progenitors. Stem Cells ***29***, 193-205, 10.1002/stem.567.

45. Isaksen, T.J., Fujita, Y., and Yamashita, T. (2020). Repulsive guidance molecule a suppresses adult neurogenesis. Stem Cell Reports ***14***, 677-691, https://doi.org/10.1016/j.stemcr.2020.03.003.

46. Gonnot, F., Langer, D., Bourillot, P.-Y., et al. (2019). Regulation of Cyclin E by transcription factors of the naïve pluripotency network in mouse embryonic stem cells. Cell cycle ***18***, 2697-2712, 10.1080/15384101.2019.1656475.

47. Ables, E.T., and Drummond-Barbosa, D. (2013). Cyclin E controls Drosophila female germline stem cell maintenance independently of its role in proliferation by modulating responsiveness to niche signals. Development ***140***, 530-540, 10.1242/dev.088583.

48. Aji, K., Maimaijiang, M., Aimaiti, A., et al. (2016). Differentiation of human adipose derived stem cells into smooth muscle cells is modulated by CaMKII γ. Stem Cells International ***2016***, 1267480, 10.1155/2016/1267480.

49. Gu, Y., Chen, T., Meng, Z., et al. (2012). CaMKII γ, a critical regulator of CML stem/progenitor cells, is a target of the natural product berbamine. Blood ***120***, 4829-4839, 10.1182/blood-2012-06-434894.

50. Choksi, S.P., Southall, T.D., Bossing, T., et al. (2006). Prospero acts as a binary switch between self-renewal and differentiation in drosophila neural stem cells. Developmental cell ***11***, 775-789, 10.1016/j.devcel.2006.09.015.

51. Resnik-Docampo, M., Cunningham, K.M., Ruvalcaba, S.M., et al. (2021). Neuroglian regulates Drosophila intestinal stem cell proliferation through enhanced signaling via the epidermal growth factor receptor. Stem Cell Reports ***16***, 1584-1597, 10.1016/j.stemcr.2021.04.006.

52. Alajem, A., Biran, A., Harikumar, A., et al. (2015). Differential association of chromatin proteins identifies BAF60a/SMARCD1 as a regulator of embryonic stem cell differentiation. Cell Reports ***10***, 2019-2031, 10.1016/j.celrep.2015.02.064.

53. Saha, S., Samal, P., Madhulika, S., et al. (2022). SMARCD1 negatively regulates myeloid differentiation of leukemic cells via epigenetic mechanisms. Blood Advances ***6***, 3106-3113, 10.1182/bloodadvances.2021006235.

54. Bhatia, M., Bonnet, D., Wu, D., et al. (1999). Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells. Journal of Experimental Medicine ***189***, 1139-1148, 10.1084/jem.189.7.1139.

55. Watanabe, Y., Miyasaka, K.Y., Kubo, A., et al. (2017). Notch and Hippo signaling converge on Strawberry Notch 1 (Sbno1) to synergistically activate Cdx2 during specification of the trophectoderm. Scientific reports ***7***, 46135, 10.1038/srep46135.

56. Piquet, B., Lallier, F.H., André, C., et al. (2020). Regionalized cell proliferation in the symbiont-bearing gill of the hydrothermal vent mussel Bathymodiolus azoricus. Symbiosis ***82***, 225-233, 10.1007/s13199-020-00720-w.

57. Samardzic, T., Gerlach, J., Muller, K., et al. (2002). CD22 regulates early B cell development in BOB.1/OBF.1-deficient mice. European journal of immunology ***32***, 2481-2489, 10.1002/1521-4141(200209)32:9<2481::AID-IMMU2481>3.0.CO;2-C.

58. O'Brien, J.S., Carson, G.S., Seo, H.C., et al. (1994). Identification of prosaposin as a neurotrophic factor. Proceedings of the National Academy of Sciences ***91***, 9593-9596, 10.1073/pnas.91.20.9593.

59. O'Brien, J.S., and Kishimoto, Y. (1991). Saposin proteins: structure, function, and role in human lysosomal storage disorders. The FASEB Journal ***5***, 301-308,

10.1096/fasebj.5.3.2001789.

60. Laming, S.R., Gaudron, S.M., and Duperron, S. (2018). Lifecycle ecology of deep-sea chemosymbiotic mussels: a review. Frontiers in Marine Science ***5***, 282.

61. Weaver, P.P., Billett, D.S., and Van Dover, C.L. (2018). Environmental risks of deep-sea mining. In Handbook on marine environment protection, (Springer), pp. 215-245.

62. Gross, R., Vavre, F., Heddi, A., et al. (2009). Immunity and symbiosis. Molecular microbiology ***73***, 751-759, 10.1111/j.1365-2958.2009.06820.x.

63. Li, Y., Song, X., Wang, W., et al. (2017). The hematopoiesis in gill and its role in the immune response of Pacific oyster *Crassostrea gigas* against secondary challenge with *Vibrio splendidus*. Developmental & Comparative Immunology ***71***, 59-69, 10.1016/j.dci.2017.01.024.

64. Pichon, D., Cudennec, B., Huchette, S., et al. (2013). Characterization of abalone *Haliotis tuberculata–Vibrio harveyi* interactions in gill primary cultures. Cytotechnology ***65***, 759-772, 10.1007/s10616-013-9583-1.

65. Sogin, E.M., Leisch, N., and Dubilier, N. (2020). Chemosynthetic symbioses. Current Biology ***30***, R1137-R1142, 10.1016/j.cub.2020.07.050.

66. Ponnudurai, R., Sayavedra, L., Kleiner, M., et al. (2017). Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. Standards in genomic sciences ***12***, 50, 10.1186/s40793-017-0266-y.

67. Kádár, E., Davis, S.A., and Lobo-da-Cunha, A. (2008). Cytoenzymatic investigation of intracellular digestion in the symbiont-bearing hydrothermal bivalve *Bathymodiolus azoricus*. Marine Biology ***153***, 995-1004, 10.1007/s00227-007-0872-0.

68. Chen, H., Wang, M., Li, M., et al. (2021). A glimpse of deep-sea adaptation in chemosynthetic holobionts: depressurization causes DNA fragmentation and cell death of methanotrophic endosymbionts rather than their deep-sea Bathymodiolinae host. Molecular ecology 10.1111/mec.15904.

69. Li, M., Chen, H., Wang, M., et al. (2022). Robust phenotypic plasticity of symbiotic organs facilitates the adaptation and evolution of deep-sea mussels. bioRxiv

2022.2008.2011.503589, 10.1101/2022.08.11.503589.

70. Jacob, F. (1977). Evolution and tinkering. Science ***196***, 1161-1166, 10.1126/science.86013.

71. Sun, J., Chen, C., Miyamoto, N., et al. (2020). The Scaly-foot Snail genome and implications for the origins of biomineralised armour. Nature Communications ***11***, 1-12.

72. Altura, M.A., Stabb, E., Goldman, W., et al. (2011). Attenuation of host NO production by MAMPs potentiates development of the host in the squid-vibrio symbiosis. Cellular microbiology ***13***, 527-537, 10.1111/j.1462-5822.2010.01552.x.

73. Foster, J.S., Apicella, M.A., and McFall-Ngai, M.J. (2000). Vibrio fischeri lipopolysaccharide induces developmental apoptosis, but not complete morphogenesis, of the Euprymna scolopes symbiotic light organ. Developmental Biology ***226***, 242-254, 10.1006/dbio.2000.9868.

74. Bien, T., Hambleton, E.A., Dreisewerd, K., and Soltwisch, J. (2021). Molecular insights into symbiosis—mapping sterols in a marine flatworm-algae-system using high spatial resolution MALDI-2-MS imaging with ion mobility separation. Analytical and Bioanalytical Chemistry ***413***, 2767-2777, 10.1007/s00216-020-03070-0.

75. Kok, L.T., Norris, D.M., and Chu, H.M. (1970). Sterol metabolism as a basis for a mutualistic symbiosis. Nature ***225***, 661-662, 10.1038/225661b0.

76. Hambleton, E.A., Jones, V.A.S., Maegele, I., et al. (2019). Sterol transfer by atypical cholesterol-binding NPC2 proteins in coral-algal symbiosis. eLife ***8***, 10.7554/eLife.43923.

77. Negri, I. (2012). Wolbachia as an “infectious” extrinsic factor manipulating host signaling pathways. Frontiers in Endocrinology ***2***, 10.3389/fendo.2011.00115.

78. Mat, A.M., Sarrazin, J., Markov, G.V., et al. (2020). Biological rhythms in the deep-sea hydrothermal mussel *Bathymodiolus azoricus*. Nature Communications ***11***, 3454, 10.1038/s41467-020-17284-4.

79. Cao, L., Lian, C., Zhang, X., et al. (2021). In situ detection of the fine scale heterogeneity of active cold seep environment of the Formosa Ridge, the South China

Sea. Journal of Marine Systems ***218***, 103530, 10.1016/j.jmarsys.2021.103530.

80. Sun, Y., Wang, M., Zhong, Z., et al. (2022). Adaption to hydrogen sulfide-rich environments: Strategies for active detoxification in deep-sea symbiotic mussels, *Gigantidas platifrons*. Science of The Total Environment ***804***, 150054, 10.1016/j.scitotenv.2021.150054.

81. Chen, H., Zhou, Z., Wang, H., et al. (2016). An invertebrate-specific and immune-responsive microRNA augments oyster haemocyte phagocytosis by targeting CgIkappaB2. Scientific reports ***6***, 29591, 10.1038/srep29591.

82. Wang, H., Zhang, H., Zhong, Z., et al. (2021). Molecular analyses of the gill symbiosis of the bathymodiolin mussel *Gigantidas platifrons*. iScience ***24***, 101894, 10.1016/j.isci.2020.101894.

Fig. 1. Spatially resolved single-cell transcriptomic atlas of deep-sea mussel gill

(A) Experimental workflow for, and analysis of, single-cell transcriptome (scRNA-seq) and spatial transcriptome (ST-seq) of gill tissue in fully symbiotic (InS group) and partially decolonized (DeC) deep-sea mussels. A marked decrease of the endosymbiont is observed after *in situ* translocation assay (n=14 for the InS group, n=5 for the DeC group).

(B) t-SNE projection of spatial transcriptome clustered by gene expression in the InS group, with color assigned by cell type (left). Projection of cell clusters onto spatial transcriptome barcoded spots (right) in the InS group. Two successive sections were used in the same capture region of spatial transcriptome as technique replicates.

(C) t-SNE projection of single-cell transcriptome clustered by gene expression in the InS group, with color assigned by cell type (left). Spatial transcriptome barcoded spots labeled using scRNA-seq cell type with maximum prediction score (right).

(D) Fluorescent *in situ hybridization* (FISH) of endosymbionts with successive gill sections for ST-seq (InS group) and ISH of bacteriocytes marker genes.

Fig. 2. Immune genes and processes involved in symbiosis

(A) Schematic drawing of immune-related marker genes of bacteriocytes.

(B) Engulfment of symbionts (labeled with star) by bacteriocytes as revealed by TEM and phagocytosis of FITC-labeled *Vibrio alginolyticus* by bacteriocytes *in vitro*.

(C) Radar chart showing expression pattern of endocytosis- and lysosome-related genes of bacteriocytes in the InS and DeC groups. The size of blue/yellow colored apical circles represents the absolute value of $\log_2$(fold changes), where blue indicates downregulation, and yellow indicates upregulation, in partially decolonized mussels. Green and dark green colored inner diagram represents the abundance of a given gene in decolonized and fully symbiotic mussels, respectively.

Fig. 3. Metabolic interactions between mussel host and endosymbionts

(A) ScRNA/ST-seq and meta-transcriptome data show an intimate interaction of sterol metabolism between the host and symbionts. In support of this, immunofluorescence (IF) assay (using gills of the InS group) of the 24-hydroxycholesterol 7 alpha-hydroxylase (CYP39A1) protein shows that CYP39A1 proteins are widely distributed across bacteriocytes. In addition, a more intensive signal of CYP39A1 protein could also be observed at the apical region that enriched with endosymbionts (indicated by lipid A signals).

(B) ScRNA/ST-seq and meta-transcriptome data show intimate interaction of glucose/glycogen metabolism between the host and symbionts. In support of this, IF assay of the sugar phosphate exchanger (SLC37A2) protein shows the co-location of SLC37A2 proteins with endosymbionts inside bacteriocytes.

(C) ScRNA/ST-seq and meta-transcriptome data show intimate interactions of ammonia metabolism between the host and symbionts. In support of this, IF assay of the ammonium transporter Rh (RHBG-A) protein shows co-location of RHBG-A proteins with endosymbionts inside bacteriocytes.

Fig. 4. Development trajectory of deep-sea mussel bacteriocyte lineages

(A) Genetic regulatory network of mature bacteriocytes constructed by weighted gene co-expression network analysis (WGCNA) using scRNA-seq data. Marker genes with similar function or in same biological process are labeled in the same color (A1). Several hub transcription factors of the genetic regulatory network were found downregulated in the DeC group, accompanied by the differential expression of other 63 element markers (A2).

(B) Distribution pattern of stem cells in the InS group as revealed by a 5-ethynyl 2’-deoxyuridine (EdU) staining assay. Newly synthesized DNA is labeled with EdU signals (green) while cell nuclei are stained by DAPI (blue) and symbionts are labeled using a Cy3-labeled Eub338 probe (orange).

(C) Development trajectory of bacteriocyte lineages. Pseudo-time analysis in Monocle shows the successive development process from stem cells to bacteriocytes (C1), which could be further divided into three distinct states (C2). Stem cells and ciliated cells are distributed mostly in states 3 and 1 (C3) while bacteriocytes are mostly distributed in state 2.

(D) Cell proliferation of bacteriocytes observed by EdU assay and 3D electron microscopy. EdU signals indicating DNA replication are observed in the nucleus region of some mature bacteriocytes of the InS group (white triangle, D1). Nuclear division in bacteriocytes could also be observed with 3D reconstruction of gill tissue (a total of 91 serial images with thickness at 100 nm per section).

Fig. 5. Co-option of conserved transcription factors in function and development of bacteriocytes

(A) A t-SNE plot of stem cells and bacteriocytes that in different pseudo-time states. Co-overlaps in the t-SNE distribution of state 1 bacteriocytes and state 2 cells of DEPCs/PEBZCs and between state 1 cells of DEPCs and state 1 cells of PEBZCs were observed (shown in magnified fields).

(B) Expression pattern of the hub transcription factors EZH2 in color-coded t-SNE plots (left) and different states of stem cells.

(C) Expression pattern of hub transcription factors in different states of bacteriocytes.

(D) Schematic diagram shows function and development of bacteriocyte lineages guided by co-option of conserved transcription factors identified from the genetic regulatory networks. Arrows represent the deduced development trajectory (including cell type transition and maturation) in bacteriocytes.

Supplementary Fig. S1. PCoA and saturation analysis of transcriptome data

(A) Principal co-ordinates analysis (PCoA) of the meta-transcriptome of the InS (isobaric samples, n=7) and DeC groups (isobaric samples, n=1, and RNAsafer-preserved samples, n=4) using all mussel mRNA transcripts. (B) PCoA of the meta-transcriptome of the InS (RNAsafer-preserved samples, n=7) and DeC groups using all mussel mRNA transcripts. (C) Saturation analysis of ST-seq data in the InS group. (D) Saturation analysis of ST-seq data in the DeC group. (E) Saturation analysis of scRNA-seq data in the InS group. (F) Saturation analysis of scRNA-seq data in the DeC group.

Supplementary Fig. S2. Molecular function of bacteriocyte lineages

(A) Gene Ontology (GO) enrichment analysis of marker genes in bacteriocytes, DEPCs, PEBZCs (cluster 2) and ICCs. Size of the dot represents number of enriched genes in pathway, and color represents q-value of enrichment result. (B) Phagocytosis of FITC-labeled beads by bacteriocytes *in vitro*. (C) Cell division signals (black triangles) of endosymbionts revealed by TEM analysis.

Supplementary Fig. S3. Sterol biosynthesis pathways of endosymbionts and deep-sea mussel host

(A) Metabolic pathway of steroid biosynthesis in the symbiont as revealed using the genome (left) and meta-transcriptome (right, data from isobaric samples of the InS group, n=7). For genomic data, all the identified genes are colored in grey. For meta-transcriptome data, all identified genes are colored in green, while abundantly expressed genes (top 10%) are colored in pink. (B) Metabolic pathway of steroid biosynthesis in the host as revealed using the genome (left) and meta-transcriptome (right). For transcriptome data, only abundantly expressed genes (top 10%) are colored in green.

Supplementary Fig. S4. Glucose/glycogen metabolic pathways of endosymbionts

Glucose/glycogen metabolic pathways of the symbiont are colored based on the genome and meta-transcriptome data (isobaric samples of the InS group, n=7). For genomic data, all the identified genes are colored in grey. For meta-transcriptome data, all identified genes are colored in green, while abundantly expressed genes (top 10%) are colored in pink.

Supplementary Fig. S5. Glucose/glycogen metabolic pathways of mussel host

Glucose/glycogen metabolic pathways of the host are colored based on the genome and meta-transcriptome data (isobaric samples of the InS group, n=7). For genomic data, all the identified genes are colored in grey. For transcriptome data, only abundantly expressed genes (top 10%) are colored in green.

Supplementary Fig. S6. Ammonia-related metabolic pathways of endosymbionts

Ammonia-related metabolic pathways of the symbiont are colored based on the genome and meta-transcriptome data (isobaric samples of the InS group, n=7). For genomic data, all the identified genes are colored in grey. For meta-transcriptome data, all identified genes are colored in green, while abundantly expressed genes (top 10%) are colored in pink.

Supplementary Fig. S7. Differentially expressed genes of mature bacteriocytes in the InS group in comparison

(A) Differentially expressed genes of mature bacteriocytes in the InS group in comparison with the DeC group. (B) GO enrichment analysis shows multiple metabolic pathways were modulated dot size represents the overall number of differentially expressed genes, up-down normalization = [number of upregulated genes] / [number of downregulated genes] - 1). (C) Genes involved in carbohydrate, vitamin, lipid, and amino acid biosynthesis are mostly downregulated in decolonized mussels.

Supplementary Fig. S8. WGCNA modules of gill tissue and the developmental process of bacteriocytes

(A) Clustering of WGCNA module eigengenes based on scRNA-seq data of the InS and DeC group. Eigengenes represent the overall expression level of genes in each module of a given cell type and sample, with green indicating a negative, and red indicating a positive correlation. (B) GO enrichment analysis using 112 element marker genes of mature bacteriocytes annotated from module Mod12. (C) Development trajectory of bacteriocyte lineages revealed by PAGA (C1-2) and RNA velocity analysis (C3). Ongoing development from stem cells to mature bacteriocytes is observed by PAGA analysis in which PEBZCs were the most primitive cells in development (C1). Interconnectivity between each cell pair is calculated and visualized, showing the successive differentiation process (C2). A positive velocity from PEBZCs and DEPCs to bacteriocytes was also observed in the adjacent regions of corresponding cells (C3). (D) Venn diagrams of marker genes in bacteriocytes (D1), DEPCs (D2), PEBZCs (D3 for cluster 2, D4 for cluster 6) with high expression genes (HEGs) of corresponding cell types in states 1, 2, and 3. (E) Venn diagrams of DEPC marker genes with state-related HEGs of PEBZCs (cluster 2). (F) Expression pattern of key transcription factors in different states of bacteriocytes, DEPCs and PEBZCs.

Supplementary Fig. S9. Phylogenetic analysis of hub transcription factors

(A) Phylogenetic analysis of EZH2, ELF-3, ZNF271 and AUST2 proteins. (B) Phylogenetic analysis of all hosts using cytochrome C oxidase subunit I (mitochondrial) proteins.

Supplementary Table 1. Statistical analysis of ST-seq and scRNA-seq data

Supplementary Table 2. scRNA-seq cell markers

Supplementary Table 3. Differentially expressed markers in the DeC and InS groups as assessed using scRNA-seq

Supplementary Table 4. Feature count information of meta-transcriptome of host and symbionts

Supplementary Table 5. WGCNA sample expression pattern

Supplementary Table 6. WGCNA networks in bacteriocyte lineages

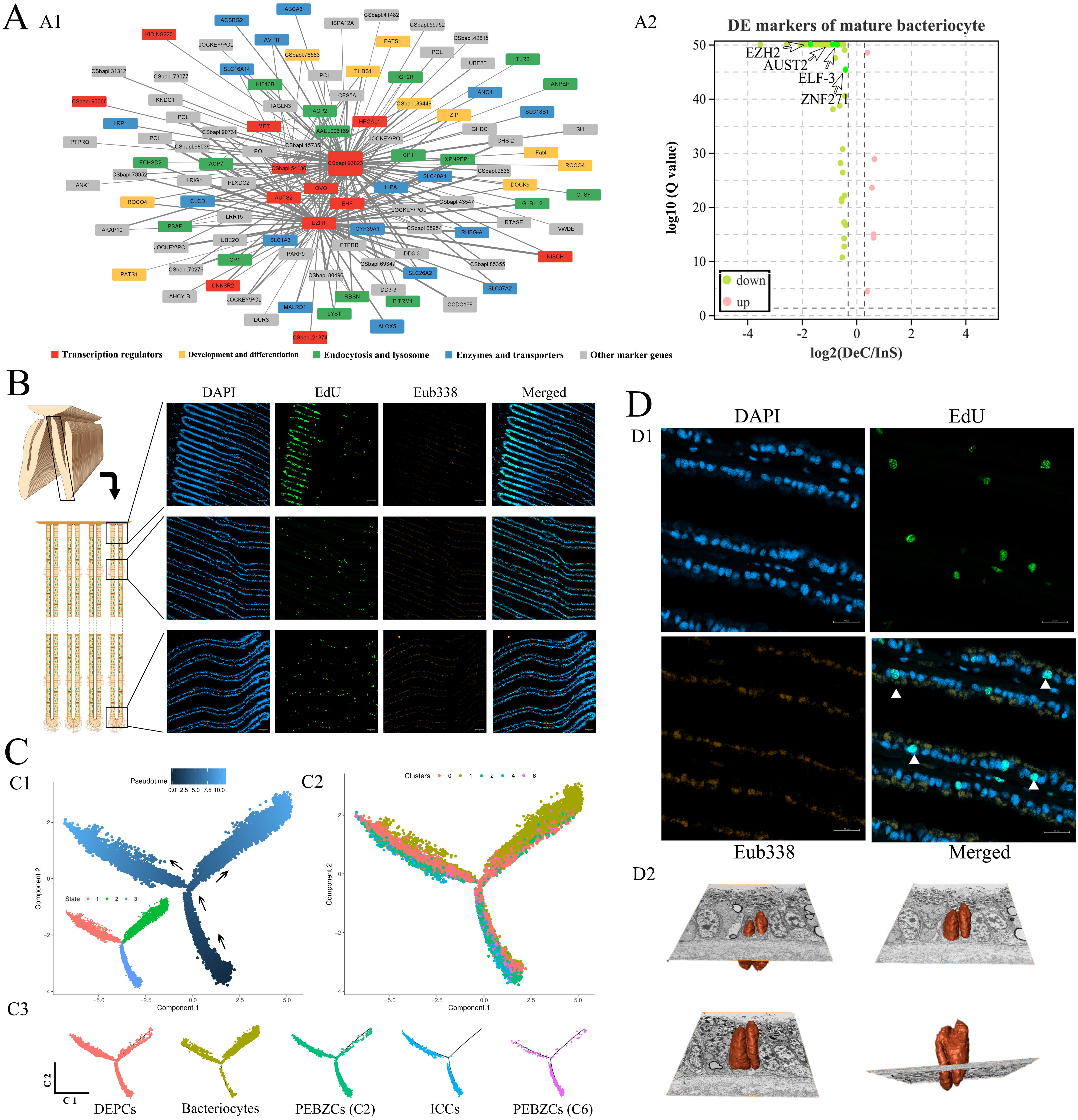
Supplementary Table 7. GO enrichment of WGCNA and all other markers in mature bacteriocytes

Supplementary Table 8 State-related differentially expressed genes in stem cells and bacteriocyte lineages

A
ST-seq
Cryosection
Dorsal
Ventral
Cell nuclei
ScRNA-seq
InS
DeC
30
25
20
15
8
6
4
2
0
InS
DeC
Endosymbiont Relative Abundance
Decolonization by Translocation
Methane concentration up to 31,227 ppm
Methane concentration about 800 ppm
B
5
CCs
6
2
3
BMC+MC
BCs
tSNE2
tSNE1
Cell cluster
0 1 2 3 4 5 6 7 8
C
ILCs
LCCs
ICCs
SMCs
PEBZCs
MC
DEPCs
PEBZCs
BMC 2
PEBZCs
BMC 1
BC
VEPCs
BMC 1
FGCCs
tSNE2
tSNE1
Cell cluster
0 1 2 3 4 5 6 7 8 9 10 11 12 13
D
FISH of gill tissue employed in ST-seq
500 μm
Blue: Dapi Green: FITC-labeled pmoB probe
Organe: Cy3-labeled Eub338 probe
ammonium transporter Rh (CSbapl.37883)
SLC40A1 (CSbapl.42154)
SLC26A2 (CSbapl.20694)
excitatory amino acid transporter (CSbapl.50748)
24-hydroxycholesterol 7 alpha-hydroxylase (CSbapl.66085)
sugar phosphate exchanger (CSbapl.90257)

A
TLR2*
TLR2
rabenosyn-5
EEA1
SNX2
CH$_4$
endophilin-A3
ELMO
Rab-11A
cathepsin D/F/L, ACP2
rhamnose-binding lectin*
B
Blue: Dapi Green: FITC-labeled *Vibrio* bacteria Red: Dil
10 μm
C
Endocytosis-Related Genes
ACAP2
AP2A2
arpc4
Bicd2
brat
Canx
EHD1
Hip1
IGF2R
LDLR
MFGE8
Pard6b
PRKC1
pten
Ptpn9
rabenosyn-5
RABGEF1
SNX12
Sra-1
TM9SF4
WASHC2A
abundance of InS group
abundance of DeC group
log2(fc) > 0
log2(fc) < 0
Lysosome-Related Genes
ACP2(CSbapl.3066)
ACP2(CSbapl.37088)
AP3D1
ATP6V0A1
CDIP1
CTSA
CTSC
CTSD
CTSL(CSbapl.44450)
CTSL(CSbapl.68766)
FUCA1(CSbapl.46497)
FUCA1(CSbapl.46506)
FUCA1(CSbapl.79438)
GLB1
HEXA
HEXB
Hgsnat
IGF2R(CSbapl.36552)
IGF2R(CSbapl.61538)
IGF2R(CSbapl.68191)
LYST
NPC1
PRCP
VhaSFD

A
● High Expression in scRNA-seq & ST-seq
● High Expression in bulk-seq
● Low expression in bulk-seq
● Not detected in bulk-seq or genome
CH_4
SYMBIOSOME
CH_4
CH_3OH
CH_2O
Terpenoid Biosynthesis
Glycolysis
EDD-RuMP
Farnesyl-PP
Presqualene-PP
Squalene
(S)-squalene-2,3-epoxide
14-Demethyl-lanosterol
4,4-dimethyl-cholesta-8,14,24-trienol
Lanosterol
Farnesyl-PP
4,4-dimethyl-cholesta-8,14,24-trienol
14-Demethyl-lanosterol
Zymosterol
Bacteriocyte
24-hytroxycholesterol 7 alpha-hydroxylase
Primary Bile Acid
Cholesterol
CYP39A1 protein
lipid A
DAPI
Merged
B
● High Expression in scRNA-seq & ST-seq
● High Expression in bulk-seq
● Low expression in bulk-seq
● Not detected in bulk-seq or genome
CH4
SYMBIOSOME
CH4
CH2OH
CH2O
EDD-RuMP
Glucose-6P
Fructose-6P
MOBs
Glucose-1P
Glucose
Fructose-1,6P
Glyceraldehydc-3P
Glycogen
TCA-cycle
Bacteriocyte
Glycogen
Sugar phosphate exchanger
TCA-cycle
Glucose
Glucose-1P
Glyceraldehydc-3P
Glucose-6P
Fructosc-6P
Fructosc-1,6P
SLC37A2 protein
lipid A
DAPI
Merged
C
● High Expression in scRNA-seq & ST-seq
● High Expression in bulk-seq
● Low expression in bulk-seq
● Not detected in bulk-seq or genome
Nitrate/nitrite
SYMBIOSOME
Nitrate
Nitrite
Glucosamine-6P
Glutamine
Ammonia
MOBs
Hydroxylamine
5-phospho-ribosylamine
Carbamoyl-phosaphate
Amino/nucleotide sugar metabolism
Purine metabolism
Arginine/Pyrimidine metabolism
Bacteriocyte
Ammonia transporter
Protein digestion
Nucleotide metabolism
Ammonia
Glutamate
Glutamina
RHBG-A protein
lipid A
DAPI
Merged


A
A1
A2
DE markers of mature bacteriocyte
EZH2
AUST2
ELF-3
ZNF271
log10 (Q value)
log2(DeC/InS)
down
up
Transcription regulators
Development and differentiation
Endocytosis and lysosome
Enzymes and transporters
Other marker genes
CSbapl.93823
EZH1
AUTS2
OVO
EHF
MET
HPCAL1
KIDINS220
CSbapl.96066
CSbapl.54136
CNKSR2
CSbapl.21874
NISCH
B
DAPI
EdU
Eub338
Merged
C
C1
Pseudotime
0.0 2.5 5.0 7.5 10.0
State 1 2 3
Component 1
Component 2
C2
Clusters 0 1 2 4 6
C3
C2
C1
DEPCs
Bacteriocytes
PEBZCs (C2)
ICCs
PEBZCs (C6)
D
D1
DAPI
EdU
Eub338
Merged
D2

A
Cluster-State
0–1
0–2
0–3
1–1
1–2
1–3
2–1
2–2
2–3
4–1
4–2
4–3
6–1
6–2
6–3
tSNE_1
tSNE_2
40
20
0
−20
−25
0
25
B
EZH2
tSNE-1
tSNE-2
DEPCs (C0)
PEBZCs (C2)
PEBZCS (C6)
Expression
State
1
2
3
C
EZH2
ZNF271
MET
AUST2
TM7SF2
Expression
State
D
state 1 PEBZCs
encode 29/43 of DEPCs marker genes
state 2 DEPCs
encode 129/199 of bacteriocyte maker genes
state 2 PEBZCs
encode 100/199 of bacteriocyte maker genes
Differentiation
EZH2, ZNF271, ELF-3, MET,
EZH2, ELF-3, ZNF271, MET ,myoneurin
107/199 marker genes
ZNF271, ELF-3, AUST2, EZH2
GATAD14, MET, myoneurin
state 2 bacteriocytes
Maturation
ZNF271, EZH2, ELF-3, AUST2 etc.
state 1 bacteriocytes
PATS1, PRKC1, MSI2
ROBO1, daf-36
Proliferation
state 3 bacteriocytes