



# Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts

Yi Yang<sup>1</sup> · Jin Sun<sup>1</sup> · Yanan Sun<sup>1</sup> · Yick Hang Kwan<sup>1</sup> · Wai Chuen Wong<sup>1</sup> · Yanjie Zhang<sup>2</sup> · Ting Xu<sup>2</sup> · Dong Feng<sup>3,4</sup> · Yu Zhang<sup>5</sup> · Jian-Wen Qiu<sup>2</sup> · Pei-Yuan Qian<sup>1</sup>

Received: 5 June 2019 / Revised: 11 August 2019 / Accepted: 25 August 2019 / Published online: 8 October 2019  
© The Author(s) 2019. This article is published with open access

## Abstract

Deep-sea hydrothermal vents and methane seeps are often densely populated by animals that host chemosynthetic symbiotic bacteria, but the molecular mechanisms of such host-symbiont relationship remain largely unclear. We characterized the symbiont genome of the seep-living siboglinid *Paraescarpia echinospica* and compared seven siboglinid-symbiont genomes. Our comparative analyses indicate that seep-living siboglinid endosymbionts have more virulence traits for establishing infections and modulating host-bacterium interaction than the vent-dwelling species, and have a high potential to resist environmental hazards. Metatranscriptome and metaproteome analyses of the *Paraescarpia* holobiont reveal that the symbiont is highly versatile in its energy use and efficient in carbon fixation. There is close cooperation within the holobiont in production and supply of nutrients, and the symbiont may be able to obtain nutrients from host cells using virulence factors. Moreover, the symbiont is speculated to have evolved strategies to mediate host protective immunity, resulting in weak expression of host innate immunity genes in the trophosome. Overall, our results reveal the interdependence of the tubeworm holobiont through mutual nutrient supply, a pathogen-type regulatory mechanism, and host-symbiont cooperation in energy utilization and nutrient production, which is a key adaptation allowing the tubeworm to thrive in deep-sea chemosynthetic environments.

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-019-0520-y>) contains supplementary material, which is available to authorized users.

✉ Jian-Wen Qiu  
qiujw@hkbu.edu.hk

✉ Pei-Yuan Qian  
boqianpy@ust.hk

<sup>1</sup> Department of Ocean Science, Division of Life Science and Hong Kong Branch of The Southern Science and Engineering Guangdong Laboratory, The Hong Kong University of Science and Technology, Hong Kong, China

<sup>2</sup> Department of Biology, Hong Kong Baptist University, Hong Kong, China

<sup>3</sup> CAS Key Laboratory of Ocean and Marginal Sea Geology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, 510301 Guangzhou, China

<sup>4</sup> Laboratory for Marine Geology, Qingdao National Laboratory for Marine Science and Technology, 266061 Qingdao, China

<sup>5</sup> College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China

## Introduction

Siboglinid tubeworms are often conspicuous members of the benthic communities of deep-sea hydrothermal vents and cold seeps [1, 2]. They are mouthless and gutless yet can have high productivity [3]. Symbiosis with  $\gamma$ -proteobacteria, a group of chemosynthetic bacteria, is a key adaptation allowing tubeworms to thrive in vent and seep ecosystems [4]. Larvae of the tubeworms obtain free-living  $\gamma$ -proteobacteria from the ambient environment through a symbiont-specific infection process [5, 6]. In the adult stage, the symbiotic bacteria are housed in a specialized organ of their host called the trophosome and no longer in direct contact with the ambient environment [5]. Substrates for chemosynthesis including sulfide and oxygen are obtained from ambient seawater through the branchial plume of the host or from the sediment through the posterior end of the host and delivered to the symbiont through the host's circulation system which uses hemoglobin [7, 8].

The genomes of both the host and the symbiont contain critical genetic information about the symbiosis. Among

the 194 species in 34 genera of Siboglinidae, none has a published genome, although the genomes of several species are being sequenced by multiple groups. Only eight endosymbiont genomes (from *Escarpia spicata*, *Lamellibrachia luymesii*, *Galathealinum brachiosum*, *Ridgeia piscesae*, *Riftia pachyptila*, *Seepiophila jonesi*, *Tevnia jerichonana* and *Osedax frankpressi*) have been sequenced [9–13]. Previous studies of the symbiosis in siboglinids have primarily focused on the giant tubeworm *R. pachyptila*, which revealed that its symbiont uses both the Calvin–Benson cycle and the rTCA cycle for carbon fixation; meanwhile, the symbiont has a complete pathway of heterotrophic metabolism, and thus can live mixotrophically [7, 14]. Siboglinid symbionts vary in their capability of using the Calvin–Benson and rTCA cycles [9, 10], indicating a greater number of comparative genomic analyses is required to fully understand the mechanisms by which tubeworms adapt to different environmental conditions and thrive in vent and seep habitats. A comparative analysis of the metaproteomics of two vent-dwelling tubeworms (i.e., *R. pachyptila* and *T. jerichonana*) living in quite different geofluid environments at the host’s plume level shows highly consistent protein expression profiles in sulfur metabolism, carbon fixation and oxidative stress [11]. Although previous studies found that the symbionts in the trophosome have evolved a pathogen-type defense mechanism to protect themselves from the host, symbiont virulence intensity and its regulation of host immunity for symbiosis maintenance have not been studied; many studies have investigated siboglinid symbiosis only from the perspective of either the host or symbiont [9, 14, 15]. Given that different siboglinid tubeworms live in various chemosynthesis habitats from hydrothermal vents and cold seeps to sunken wood, it is necessary to analyze the host and symbiont as a holobiont in order to reveal the adaptive mechanisms of tubeworm symbiosis.

In the present study, we sequenced the endosymbiont genome, metatranscriptome, and metaproteome of the siboglinid tubeworm *Paraescarpia echinospica* (the holobiont) inhabiting cold seeps in the South China Sea of the West Pacific Ocean [16, 17]. As the first integrated genomic, transcriptomic, and proteomic analysis of cold-seep tubeworm, the present study aimed to decipher the interdependence between the host and symbiont with particular emphases on how the symbiont uses various metabolic pathways to generate energy, how the host and symbiont cooperate in nutrient provisions, and how the two partners regulate each other. Furthermore, a comparative analysis was conducted with other published vestimentiferan symbiont genomes to reveal genetic basis of adaption to the vent- and seep-environments [10].

## Materials and methods

### Sampling *Paraescarpia echinospica* and metagenomic sequencing

*P. echinospica* individuals were collected from a cold-seep area situated on the northern continental slope of the South China Sea at a water depth of 1147 m (22.11619° N, 119.2856° E). Sampling was conducted using the remotely operated vehicle (ROV) ROPOS onboard the R/V Tan Kah Kee on 19 April 2018 (see Supplementary Fig. S1 showing the tubeworms prior to sampling and a complete individual preserved in 95% ethanol). The tubeworms were placed into an insulated “Biobox” with a closed lid to minimize changes in temperature in the water inside the container. It took ~40 min for the ROV to ascend from the seabed to the main deck of the research vessel. Once the worms were brought onboard the research vessel, they were dissected, with their trophosome (an organ that harbors symbionts), plume (a gill-like organ) and vestimentum (primarily made up of muscle) fixed separately in RNAlater® (Invitrogen, USA), and then stored at –80 °C. Total DNA of the trophosome was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol. The symbiont genome was sequenced with both the Oxford Nanopore Technology and Illumina platforms and assembled. Briefly, an 8–10 kb Nanopore DNA library was constructed using the Ligation Sequencing Kit 1D (Oxford Nanopore, Oxford, UK) according to the manufacturer’s protocol and sequenced with the FLO-MIN106 R9.4 flow cell coupled to the MinION™ platform (Oxford Nanopore Technologies, Oxford, UK) at the Hong Kong University of Science and Technology. The raw reads were base-called according to the protocol in the MinKNOW and written into *fast5* files. Illumina DNA sequencing was performed using the Illumina HiSeq™ X-Ten to produce 150 bp paired-end reads at the Beijing Genomics Institute (BGI) in Shenzhen.

### Symbiont genome assembly and functional annotation

Trimmomatic v0.33 [18] was used to trim the Illumina adapters and low-quality bases (base quality ≤ 20). The clean reads were assembled using SPAdes v3.9.1 [19] with *k*-mer sizes of 21, 33, 55, 77, 99, and 127 bp, and the products were pooled. Genome binning was then conducted as in previous studies [20, 21]. Briefly, the clean reads were first mapped to the assembled contigs using Bowtie2 v2.2.9 [22], and the coverage of each contig was calculated using SAMTOOLS v1.3.1. Open reading frames (ORFs) were then predicted using Prodigal v2.6.3 [23] and protein functional domains were predicted using HMMER 3.1b2

[24] under the 100 + HMM model. Taxonomic affiliation of all HMM positive ORFs were determined using BLASTp [25] against NCBI nonredundant (NR) protein database, and the taxonomic assignment of each protein was imported to MEGAN v5.7.0 [26] using the lowest common ancestor (LCA) method with the parameters of Min Score 50, Max Expected 0.01, Top Percent 5 and LCA Percent 100. The results were analyzed in RStudio (<https://www.rstudio.com/>) with the libraries of vegan, plyr, RColorBrewer and alphahull. Sequences representing the draft symbiont genome were then extracted from the assembled contigs of both the host and the symbiont, based on the combination of sequencing coverage and GC content (Supplementary Fig. S2a) [21]. Contigs belonging to the potential bacterial genome were further determined using principal component analysis (PCA) of tetranucleotide frequencies (Supplementary Fig. S2b), assessed using CheckM v1.0.6 [27], and further scaffolded using SSPACE-LongRead v1.1 [28] by adding Nanopore long reads. The newly assembled scaffolds were binned again using the above-mentioned pipeline. GapFiller v1.10 [29] was then used to fill the gaps in the binned symbiont genome, and the completeness and potential contamination of the binned genome were estimated using CheckM v1.0.6 [27]. Coding sequences (CDS) in the *P. echinospica* symbiont genome were predicted and translated using Prodigal v2.6.3 [30]. The translated protein sequences were functionally annotated with RPS-BLAST v2.2.15 ( $e$ -value  $< 10^{-05}$ ) against the databases of Clusters of Orthologous Groups (COGs) for prokaryotes, Gene Ontology (GO) and Pfam using WebMGA online analysis [31]. Sequences were annotated with KEGG (Kyoto Encyclopedia of Genes and Genomes) numbers against the KEGG database using BLASTp, and KEGG Mapper was run on the KAAS to construct the metabolic pathways of the symbiont from these sequences [32].

Raw sequencing data of the *P. echinospica* metagenome have been deposited in NCBI's Sequence Read Archive database under BioProject PRJNA472657 and BioSample SAMN09239911. The 16S rRNA gene sequence of the *P. echinospica* symbiont has been deposited in GenBank under the accession number MH628048. The complete genome sequences of the symbiont have been deposited in DDBJ/ENA/GenBank under the accession number RZUD00000000.

### Genomic comparison and phylogenomic analysis of siboglinid symbionts

Four seep-living and three vent-dwelling vestimentiferans endosymbiont genome sequences (Table 1) were compared using BLASTn 2.2.26 [33], then visualized using BRIG [34] and Circletto [35] to provide an overview of genome sequence similarity (Supplementary Fig. S3). In particular, the orthologous groups (OGs) from the above

seven endosymbiont genomes and a mud-dwelling siboglinid endosymbiont genome were detected using Proteinortho v5.16b [36] (BLAST threshold  $E = 1 \times 10^{-10}$ ). Only single-copy genes in each OG that were found in all taxa were retained for phylogenomic analysis, resulting in 1305 OGs. Sequences of each OG were aligned using MUSCLE and trimmed using TrimAL v1.4 [37]. After concatenating these alignments, a phylogenetic tree was constructed using RaxML version 8.2.4 [38] under the GTR +  $\Gamma$  model with the partition information of each orthologous gene and 1000 bootstrap replicates. Similarly, a phylogenetic analysis of the siboglinids based on the 13 concatenated mitochondrial genes [39] was performed using RaxML version 8.2.4 [38] under the GTR + CAT model (Fig. 1a). PCA analysis on the orthologous proteins in 1305 OGs was performed using Jalview [40] under the BLOSUM62 model, the similarity scores between each pair of sequences were calculated to form the matrix, the components were generated and then visualized using BioVinci (Bioturing, San Diego, CA, USA) (Fig. 1b). A Venn diagram was constructed using Venn webtool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to illustrate the shared and unique orthologous genes among the seep- and vent-dwelling vestimentiferan endosymbionts (Fig. 1c). Based on the results from Proteinortho, orthologous genes that were only present in symbionts from a particular habitat were classified as unique genes to that habitat, e.g. vent-unique or seep-unique genes. In addition, a hidden Markov model (HMM)-based approach delta-bitscore (DBS) [41, 42] was used to identify the functional divergence of shared orthologous proteins in seep- and vent-dwelling vestimentiferan endosymbiont genomes, their adaptability to the host and habitat was mined through these loss-of-function mutations.

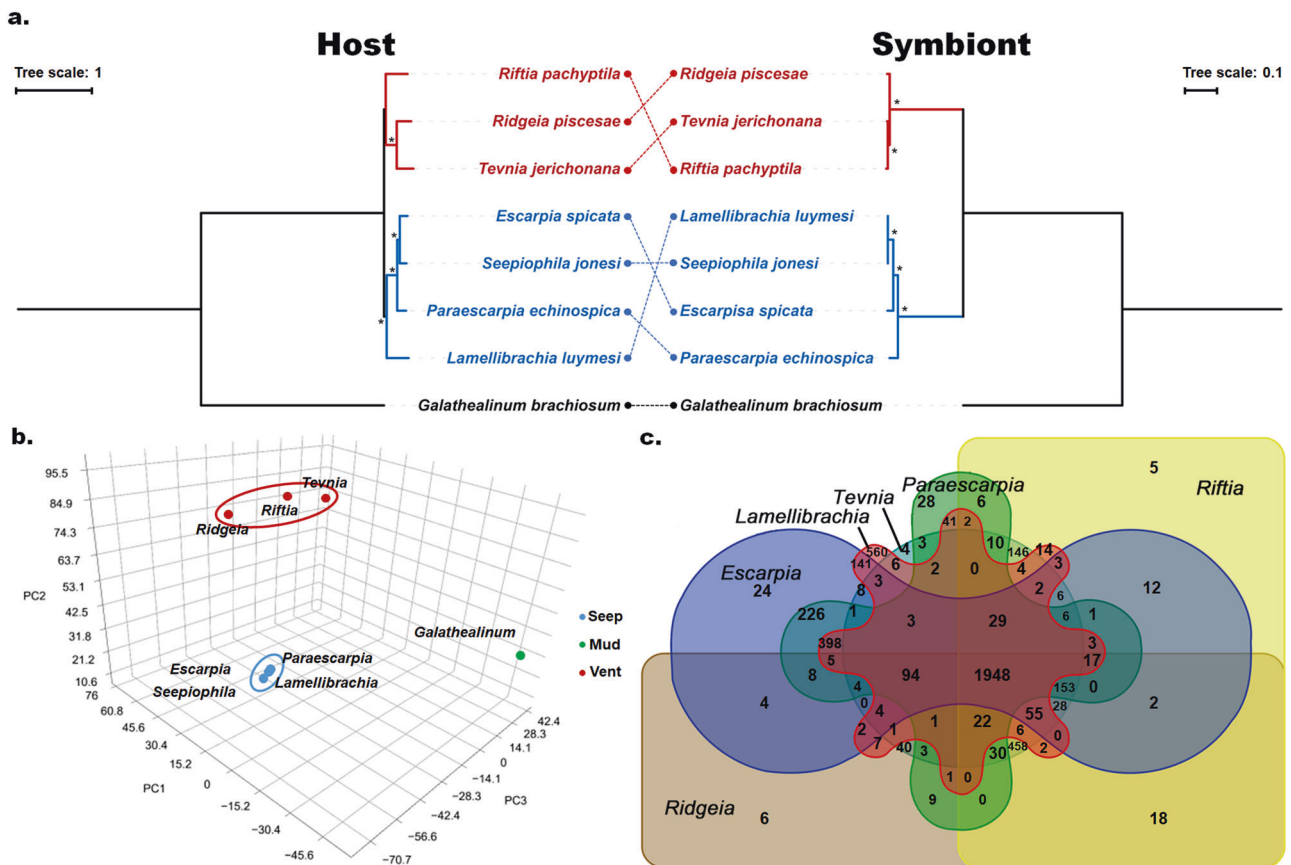
### Tubeworm holobiont transcriptome sequencing

The same *P. echinospica* individual used for metagenome sequencing was also subjected to transcriptome sequencing. Total RNA of the plume, that of the vestimentum and that of the trophosome were extracted using Trizol (Invitrogen, USA) following the manufacturer's protocol. A cDNA library of each body region was then constructed and sequenced on the HiSeq™ 4000 platform (Illumina, San Diego, CA) at BGI in Shenzhen to produce 100 bp paired-end reads. Since the trophosomal RNA includes the sequences from both the host and the symbiont, another library of the trophosome was constructed after removing the prokaryotic RNA so as to sequence the remaining eukaryotic RNA [43]. Therefore, two sets of sequencing data were produced for the trophosome: one including the transcripts of both the host and the symbiont, and the other including only host transcripts.

**Table 1** The general genomic features of vestimentiferan endosymbionts

Symbiont	Accession No.	Genome size (Mb)	No. of contigs	$N_{50}$ (kb)	GC%	No. of CDS	No. of functions assigned	Host
<i>Paraescarpia</i> symbiont	RZUD00000000	4.06	14	381.7	54.8	3 525	2 906	Seep-living vestimentiferans
<i>Escarpia</i> symbiont	QFXE00000000	4.06	23	313.6	54.2	—	—	
<i>Lamellibrachia</i> symbiont	QFXD00000000	3.53	337	20.6	54.3	—	—	
<i>Seepiophila</i> symbiont	QFXF00000000	3.53	323	20.6	54.3	—	—	
<i>Ridgeia</i> symbiont	LDXT00000000	3.44	97	84.0	58.9	3 158	2 699	Vent-dwelling vestimentiferans
<i>Riftia</i> symbiont	AFOC00000000	3.48	197	29.7	58.8	3 488	2 890	
<i>Tevnia</i> symbiont	AFZB00000000	3.64	184	92.7	58.2	3 367	2 827	

Genome data—symbiont of *Paraescarpia*: this study; symbionts of *Escarpia*, *Lamellibrachia* and *Seepiophila*: Li et al. [10]; symbiont of *Ridgeia*: Perez and Juniper [12]; symbionts of *Riftia* and *Tevnia*: Gardebrecht et al. [11]



**Fig. 1** **a** Cophylogeny analysis of bacterial symbionts (right side) and their associated siboglinid hosts (left side). Vent- and seep-living vestimentiferans are colored in red and blue respectively. All nodes have 100% bootstrap support. **b** PCA analysis on the orthologous

proteins in 1305 OGs of the endosymbionts of Siboglinidae under the BLOSUM62 model. **c** Venn diagram depicting unique and shared orthologous gene clusters in each of the six endosymbiont genomes

## De novo holobiont transcriptome assembly and sequence analysis

Adapters and low-quality reads (base quality  $\leq 20$ ) were trimmed with Trimmomatic v0.33 [18]. Clean reads from the

plume, vestimentum, and trophosome (including prokaryotic reads) were pooled and assembled using Trinity version 2.1.0 [44] under default settings. Only the highest expressed isoforms were retained. CD-HIT-EST [45] was used to further reduce redundant sequences with a threshold of 90%



similarity. TransRate [46] was used to detect errors such as chimeric artifacts, incomplete assembly and base errors in the assembled holobiont transcriptome. TransDecoder [47] was used to detect coding regions while BLAST-2.2.31+ [25] was applied to search holobiont proteins against NCBI NR protein database. Taxonomical assignment of the annotated transcripts was then performed using the LCA assignment algorithm in MEGAN v5.2.3 [26] with the top 10 hits of each transcript in the NR database, which allowed the sorting of both the host and symbiont transcripts. The transcriptomes of the host and symbionts were produced and separated. BUSCO v3 was used to evaluate the comprehensiveness of the *P. echinospica* transcriptome assembly [48]. Blast2GO v4.0.7 [49] was applied to assign GO terms to the transcripts. Transcript expression levels in each region (plume, vestimentum, and trophosome) were quantified and expressed in transcripts per million (TPM) using Salmon [50]. To understand the region-specific gene functions, the transcripts of each region with at least ten parts per million were retained [51] and blasted against the databases of Eukaryotic Orthologous Groups of proteins (KOG), COGs of proteins for prokaryotes and KEGG using RPS-BLAST v2.2.15 [31] ( $e$ -value  $< 10^{-05}$ ). The resultant KEGG Orthology assignments were mapped to KEGG pathways with KEGG Mapper on the KEGG Automatic Annotation Server v2.0 (KAAS) [32]. A gene was considered to be specifically expressed in a particular region if its TPM value in the region accounted for more than 75% of the total TPM of all three regions [51]. Besides, transcriptome data of *R. pachyptila*, *R. piscesae*, *E. spicata*, *L. luymesii*, *S. jonesii* and *G. brachiosum* were obtained from the NCBI SRA database for phylogenomic analyses (Supplementary Table S1, see Supplementary Information Methods for details).

Raw sequence data of the three regions have been deposited in NCBI Sequence Read Archive database under BioProject PRJNA494962 and SAMN09239911. Holobiont metatranscriptome sequences have been deposited in DDBJ/EMBL/GenBank under the accession number GHDL00000000. *P. echinospica* transcriptome sequences (without prokaryotes) have been deposited in DDBJ/EMBL/GenBank under the accession number GHDM00000000.

### Metaproteomic analysis

Three individuals of *P. echinospica* were used to determine the protein expression pattern in the trophosome. Specific details of protein extraction, SDS-PAGE, in-gel trypsin digestion, LC-MS/MS, protein identification and quantitation can be found in Supplementary Information Methods. In brief, tissues (~0.1 g of wet weight) were collected from the trophosome region of three *P. echinospica* individuals, which served as three replicates. Proteins were extracted, purified, quantified, separated using SDS-PAGE and in-gel digested

with trypsin. Resulting peptides were separated and analyzed on a liquid chromatography system coupled with mass spectrometry. The host and symbiont proteins were identified and quantified using Mascot version 2.3.0, and all converted mass spectrometry.mgf data were searched against the translated protein databases of *P. echinospica* and its endosymbionts. Protein abundance was represented as an emPAI value, and the 70 most abundant proteins in the trophosome and its endosymbiont were chosen to visualize protein expression. The mass spectrometry metaproteomic dataset has been deposited in the ProteomeXchange Consortium via PRIDE [52] with the accession number PXD013944.

### Real-time PCR validation

Real-time PCR was employed to validate the expression patterns of selected genes in the trophosome, as well as the plume and vestimentum for comparison. The primers of each gene were designed using the on-line NCBI Primer-BLAST tool (Supplementary Table S2, see Supplementary Information Methods for details). Total RNA was extracted from each region from three *Paraescarpia* individuals using the Trizol method. Residual contaminant DNA was removed using the TURBO DNA-free kit (Thermo Fisher Scientific). The first strand cDNA was then synthesized using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Real-time PCR was performed with the SYBR<sup>®</sup> Green RT-PCR Reagents Kit (Applied Biosystems) on LightCycler 480 II (Roche) (see Supplementary Information Methods for procedural details). All samples and negative controls were amplified in triplicate. Triplicates were applied for each gene, and the relative gene expression level was calculated based on the  $2^{\Delta\Delta C_t}$  method [53]. The standard deviation (SD) was calculated and Student's  $t$  tests were performed with Microsoft Excel.

## Results and discussion

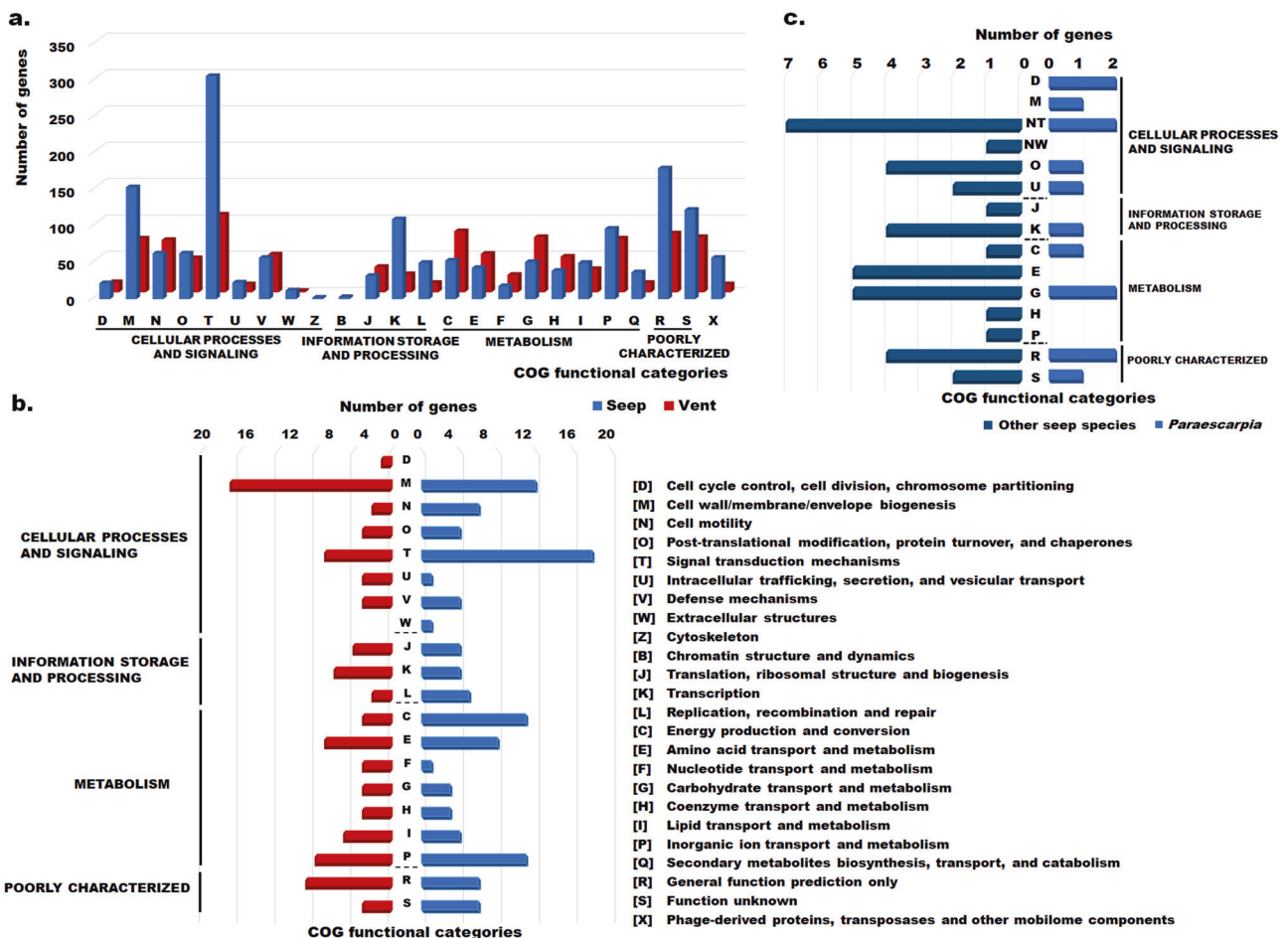
### The symbiont genome and comparative genomics

An examination of the 16S rRNA microbial community data revealed a single bacterial ribotype in the trophosome of *P. echinospica* with its phylogenetic position shown in Supplementary Fig. S4. Sequencing the trophosomal genomic DNA using the Illumina platform produced 235,260,418 paired-end reads. After assembling the reads, binning was conducted on contigs over 500 bp, and the results showed that the sequences of *P. echinospica* and its symbiont were well separated by sequencing coverage and GC content (Supplementary Fig. S2). There was only one 16S rRNA gene sequence among the potential symbiont contigs, which was identical to that obtained from the 16S rRNA gene clone library

sequence, further confirming that *P. echinospica* potentially harbored a single genotype of bacterial endosymbiont. Meanwhile, sequencing the trophosomal genomic DNA using the Nanopore MinION platform produced 1,158,101 reads, with an average length of 2.1 kb and an N50 statistic of 3.3 kb. A draft genome of the *P. echinospica* symbiont, assembled using both the Illumina and Nanopore reads, was 4.06 Mb in total length with 14 scaffolds. The maximum scaffold length was 942.6 kb, and the N50 length was 381.7 kb (Table 1). CHECKM analysis showed that this genome was 97.4% of completeness with 2.6% contamination and encoded 3525 predicted CDS. Among those CDS, 2906 (82.4%) had at least one significant hit in the COG, KEGG, Pfam and GO databases (Supplementary Table S3). Both the percentage and the number of genes in different GO and COG categories are shown in Supplementary Fig. S5. The general genomic features of the endosymbiont of *Paraescarpia* and its close relatives based on 16S rRNA gene analysis (Supplementary Fig. S4) are shown in Table 1, which indicates that the assembled *P. echinospica* symbiont is of high quality.

Our phylogenomic analysis of siboglinid holobionts showed that siboglinids and their endosymbionts did not cospicuate, while the endosymbionts were well clustered into two clades by vent and seep habitats (Fig. 1a, Supplementary Fig. S6). PCA analysis also showed that the endosymbionts of Siboglinidae were clustered strictly according to habitat type, with the seep- and vent-vestimentiferans being well separated (Fig. 1b). Gene orthology analysis showed that 1430 OGs and 677 OGs were unique to seep- and vent-dwelling vestimentiferan endosymbionts, respectively. Together, these results indicate independent evolutionary history of endosymbiosis between seep- and vent-dwelling vestimentiferans.

To understand the genetic basis of such habitat-specific endosymbiosis, we analyzed the functional composition of the seep- and vent-unique genes. The seep-unique genes contributed more than vent-unique genes to cell wall/membrane/envelope biogenesis [M], signal transduction [T] and mobile genetic elements [X] (Fig. 2a). The [M] category contains a lipid II flippase MurJ (*murJ*) which transports lipid in cell wall formation [54], a mechanosensitive channel of small



**Fig. 2** a Number of seep-unique and vent-unique orthologous genes in different COG categories. b Number of vent and seep loss-of-function genes in different COG categories. Red and blue colors represent genes belonging to vent- and seep-living symbionts respectively.

c Number of seep loss-of-function genes in different COG categories. Light and dark blue colors represent genes belonging to the *Paraescarpia* symbiont and other seep-living symbionts respectively

conductance (*mscS*) [55] and an outer membrane efflux protein TolC (*tolC*) [56, 57], which control the efflux of solutes and solvent from the outer membrane. Due to their potential for material transportation and osmosis regulation, these proteins are considered to be critical for bacterial survival, including antimicrobial resistance, symbiosis, and adaptation to adverse environments [54–57]. Their presence in seep-dwelling vestimentiferans may mediate adaptation of the symbionts to changes in their osmotic environment. The [T] category contains signal transduction histidine kinases (*baeS*, *ntrY*), which are known to sense and transmit environmental stimuli to response regulator containing CheY receiver, GGDEF domain, DNA-binding domains (*citB*, *atoSC*, *ompR*) which may activate symbiont responses to environmental signals and then the chemotaxis protein CheC (*cheC*) may enable symbionts to move toward more favorable environments [58]. Genes in [T] category are critical for bacterial chemotactic adaptation.

In natural habitats, vent-vestimentiferans grow at the basaltic base of vent chimneys and exclusively use their plumes for substances acquisition from the seawater, while seep-vestimentiferans grow in soft sediments and bury their posterior ends at greater depths in the sediments for absorbing substances [7, 59, 60] (Supplementary Fig. S1b). The endosymbionts of seep- and vent-vestimentiferans likely come from the free-living population in sediments and seawater, respectively [61, 62]. Vestimentiferans have the potential to adopt symbionts that optimally adapt to the local environment [60, 63]. In comparison to vent fluids, seep sediments often contain high concentrations of dissolved inorganic nutrient (e.g. nitrates, nitrites, ammonium, and phosphorus) [64–66] and harbor higher microbial community diversity and richness [67–70]. Therefore, the complex environment of the seep sediment may have driven the molecular adaptation in the endosymbiont of seep-dwelling vestimentiferans. Consequently, it led to a large number of seep-unique genes in the [X] category in the endosymbionts of seep-vestimentiferans, indicating these endosymbionts have the capacity to acquire more foreign genetic elements than the endosymbionts of vent-vestimentiferans. Gene acquisition is important in the adaptive evolution of prokaryotes as the acquisition of mobile genetic elements (e.g. genes encoding integrase, transposase, and phage-related proteins) may change the virulence potential of symbionts and confer their resistance to toxic compounds and other virulence factors in seep sediments [71, 72]. In addition, seep-unique genes encoding phage or phage-derived proteins have the potential to aid the symbionts in host immune evasion [73], and virulence genes that participate in bacterial capsules (e.g. capsule polysaccharide export proteins KpsS, KpsC, KpsE) may enable the symbionts to defend against phagocytosis as well as other aspects of the host immune system [72]. These results show that the seep-living siboglinid endosymbionts are more prone than the vent-dwelling siboglinid

endosymbionts to resist environmental stress and use pathogen-like mechanisms to evade host immune responses to survive intracellularly. A list of seep- and vent-unique genes is included in Supplementary Excel Table S4.

To show the effects of variation in shared orthologous clusters in symbionts, we used a profile HMM-based method DBS to capture functional genetic changes in conserved domains within shared orthologous protein sequences [41, 42]. Figure 2b shows the number of genes, classified by functions, with loss-of-function mutations in the endosymbionts of seep-living and vent-dwelling vestimentiferans. Notably, a CRISPR-associated protein Cse1 (*cseI*) in the endosymbionts of seep-living vestimentiferans, a bacterial defense protein against foreign genetic elements [74], has lost its function, which may have caused a larger number of foreign genetic elements in seep-unique genes than in vent-unique genes (Fig. 2a). Furthermore, 28 genes were unique to the endosymbiont of *Paraescarpia* (Fig. 1c) and another 15 genes lost their functions (Fig. 2c). In the *Paraescarpia* symbiont-unique genes, nitroreductase has the potential to degrade or transform toxic nitro-containing compounds from the environment [75], which provides an advantage for *Paraescarpia* symbionts adapting to seep sediments with highly total nitrogen concentrations [64, 65]. Many symbionts (e.g., siboglinid symbionts, rhizobia, enteropathogenic *Escherichia coli*) use a type II (T2SS) or a type III secretion system to evade phagocytosis and facilitate infection [10]. Instead, genes encoding type VI secretion system (T6SS) proteins (*vasD*, *impJKL*) and the hemolysin activation/secretion protein were found in the list of the *Paraescarpia* symbiont-unique genes. T6SS and the hemolysin activation/secretion protein play an important role as a transporter or pore in transporting proteins within a bacterial cell or between cells across the cell envelope. Furthermore, as an important virulence factor in Gram-negative bacteria allowing them to defend against competing organisms, T6SS mediates only bacterial intercellular interactions during symbiosis establishment rather than host cells [76, 77].

We concluded that the adaptation of vestimentiferan holobionts to vents and seeps with different physical-chemical and biotic factors is largely an attribute to differences in their symbiont genetic components and that the endosymbionts of seep-living vestimentiferans have more advantages than the endosymbionts of vent-dwelling vestimentiferans in terms of both environmental adaptation (free-living stage) and host-bacterium interaction (symbiotic stage). Among the seep-vestimentiferan endosymbionts, one with *Paraescarpia* has a higher potential than those with other species to reduce the toxicity of organic nitrogen compounds from the environment and transport proteins (including virulence factors) between cells for provision of intermediate product and nutrients. The loss-of-function orthologous genes from vent- and seep-living species are listed in Supplementary Excel Table S5.

## The holobiont metatranscriptome and metaproteome

RNA-Seq of the plume, the vestimentum and the two sets of trophosome (the holobiont and only the host) produced 39,111,919, 40,807,720, 36,212,117 and 23,460,651 paired-end reads, respectively. After trimming, 37,098,488, 38,670,009 and 34,158,322 clean and high-quality reads from the plume, vestimentum, and trophosome (holobiont reads) were assembled to produce the holobiont transcriptome. Similarly, after trimming, 21,380,499 reads from the trophosome (almost without prokaryotes) were assembled with the reads of the plume and vestimentum to produce the *P. echinospica* transcriptome. In the holobiont transcriptome, over 90.5% of the 1112 bacterial transcripts were specific to the trophosome [5]. Among the 142,750 transcripts retained, 23,810 coding regions were predicted by TransDecoder while functional annotation matched 1087 translated proteins of the symbiont, each of which had at least one significant hit in the NCBI NR, COG, KEGG or GO databases. On the other hand, for the *P. echinospica* transcriptome, among the 118,820 transcripts retained, 22,284 coding regions were predicted by TransDecoder. Functional annotation matched 20,733 translated proteins, each of which had at least one significant hit in the NCBI NR, KOG, KEGG, or GO databases (Supplementary Table S6). BUSCO analysis shows that the *P. echinospica* transcriptome is 97.7% of completeness assessed with 978 metazoan BUSCOs, which compares favorably with the completeness of transcriptome assemblies of several other species of vestimentiferans [78]. The 50 most highly expressed genes in the *P. echinospica* and their respective expression levels in the plume, vestimentum and trophosome, as well as the 50 most highly expressed genes in the symbiont are shown in Fig. 3a.

To find the protein evidence in the holobiont, total proteins extracted from the trophosome region of another three *Paraescarpia* individuals were identified and quantified by LC-MS/MS, resulted in 1767 host proteins and 474 endosymbiont proteins (Supplementary Information Methods). The 70 most abundant proteins of the trophosome and symbionts are shown in the heat map along with their relative abundances (Fig. 3b). Correlation of proteomic data and RNA profiling is shown in Supplementary Fig. S7.

## Host-microbe interdependence

### Energy sources

Similar to other vestimentiferans endosymbionts in previous study [10], the *P. echinospica* symbiont genome included the genes responsible for all of the essential metabolic

pathways for energy production and conversion in free-living chemoautotrophic sulfur-oxidizing bacteria (Fig. 4). Our results indicate that the *P. echinospica* symbiont was highly versatile in its energy use, with the ability to use thiosulfate, carbon monoxide (CO), and hydrogen as alternative energy sources. Interestingly, anaerobic oxidation of CO has not been reported in the endosymbiont of siboglinids before but has been found in the symbiont of the gutless marine worm *Olavius algarvensis* [79]. The identification of anaerobic carbon monoxide dehydrogenase (CODH) gene (*cdhA*) in *P. echinospica* symbiont genome indicates CO can be a potential energy source for tubeworm symbionts. This ability is likely a key adaptation allowing *P. echinospica* to thrive in more reducing habitats. However, unlike the abundant expression of anaerobic CODHs in the *O. algarvensis* symbiont (high CO concentrations (17–51 nM) in its habitat), anaerobic CODHs were not found in the transcriptome or proteome of the *P. echinospica* symbiont. We speculate that this might be due to the low content of CO in its habitat (currently no data on CO concentrations). The transportation of host-supplied substrates in *P. echinospica* holobiont is given in Supplementary Information (see the section on substrates supply and energy conversion, Supplementary Figs. S8–S11 and Supplementary Tables S9, S10).

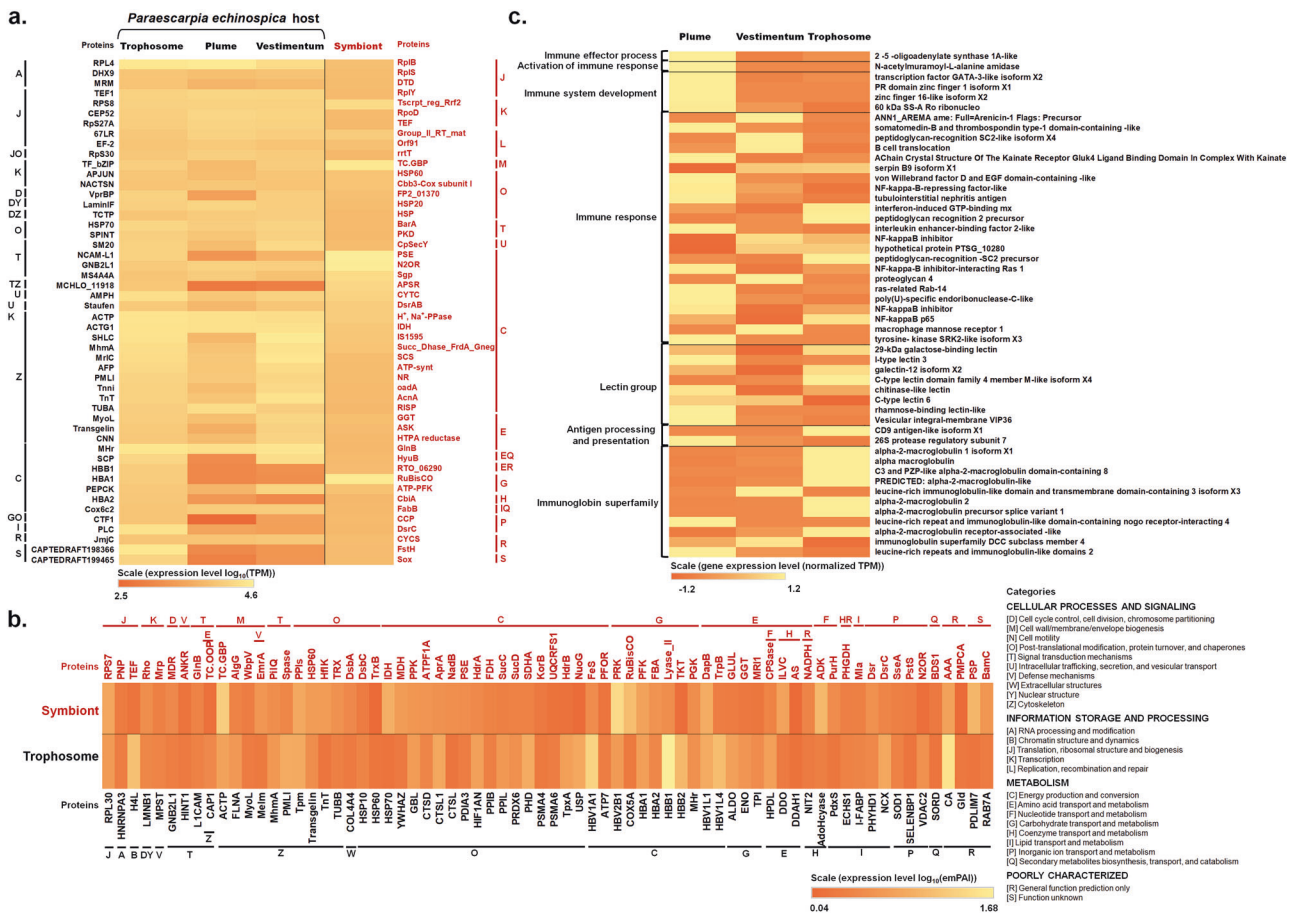
### Carbon fixation

Two carbon fixation pathways (i.e., the Calvin–Benson–Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle) have been reported in the endosymbionts of the siboglinids [9, 10]. The CBB and rTCA cycles also coexisted in the symbiont of *P. echinospica* (Fig. 4). In the CO<sub>2</sub> fixing process by CBB cycle, pyrophosphate-dependent phosphofructokinases (*PPi-PFK*) were co-encoded with proton-translocating pyrophosphatase (*hppA*) in the symbiont of *P. echinospica* and their co-transcription was confirmed in our transcriptome analysis which allows the symbiont to consume less energy at least 9.25% [80, 81]. Furthermore, the rTCA cycle is more energetically efficient than the CBB cycle [80, 82], and the rTCA cycle genes (*korB*, *por*, *sdhA*) were highly expressed in the *P. echinospica* symbiont, suggesting an active rTCA cycle in this symbiont (Fig. 3b and Fig. 4). These observations indicated that such metabolic strategy with a low energy demand could give the *P. echinospica* an advantage when living under energy- and nutrient-poor environmental conditions.

### Holobiont nutrition

The symbiont of *P. echinospica* possessed the typical metabolic pathways of nutrient generation, including the





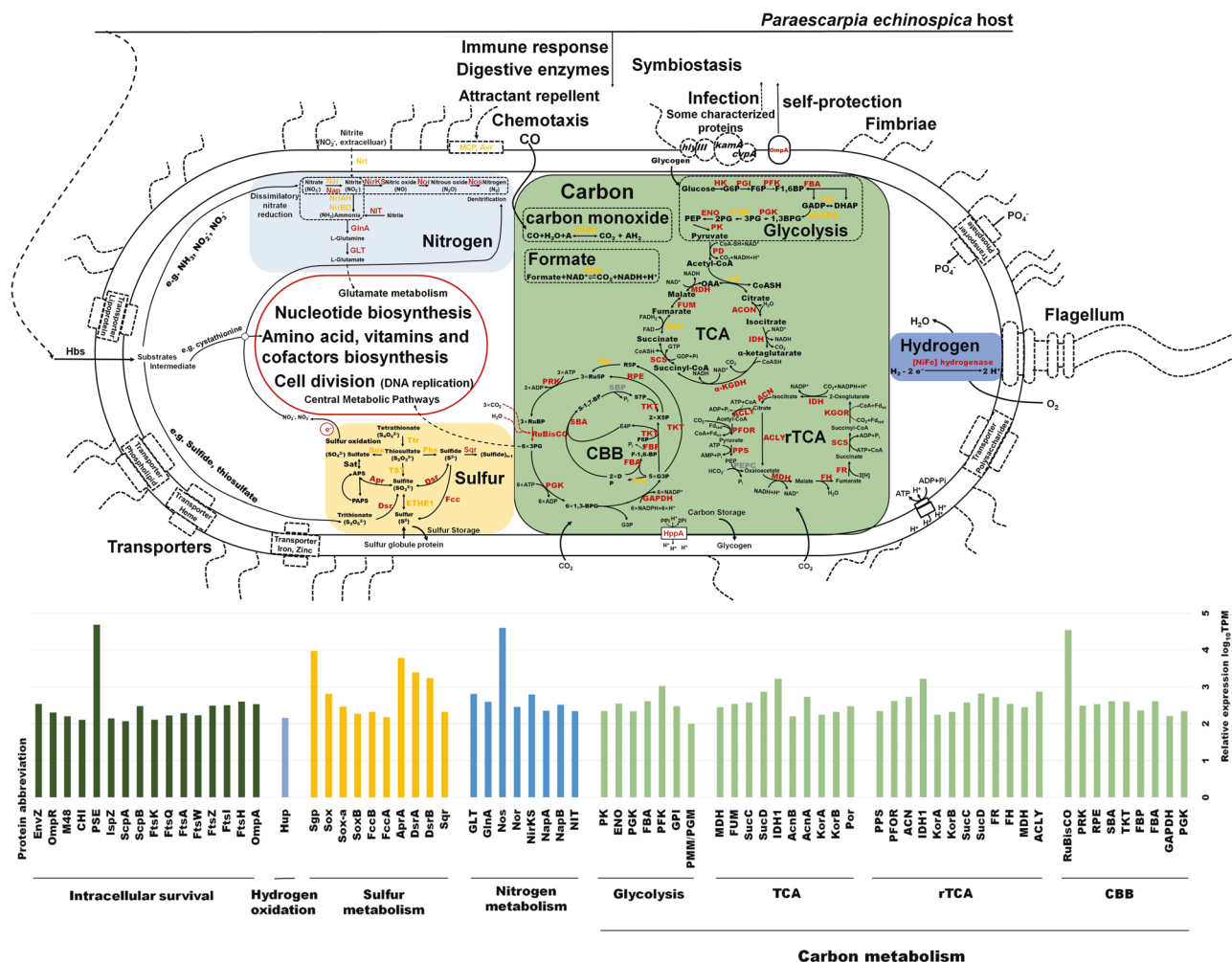
**Fig. 3** Heat map of **a** the 50 most highly expressed genes of *P. echinospica* and those of its symbiont as identified in the meta-transcriptome analysis, **b** the 70 most abundant proteins of the trophosome and those of its symbiont as identified in the metaproteome analysis, and **c** the top 50 most highly expressed immune-related genes in the plume, vestimentum and trophosome. Each grid represents an identified gene/protein in the respective sample. The color represents the gene expression level (based on Log-transformed and normalized TPM/emPAI values of the selected genes/proteins). Protein

abbreviations annotated from the host and the symbiont are listed on the two sides (see the list of abbreviations for the full names of proteins in Supplementary Information and Supplementary Tables S7 for details). Based on KOG and COG annotation, proteins are classified as shown in the lower right of the graph. Functionally redundant genes/proteins and genes/proteins of unknown function are excluded from this figure. The complete dataset is shown in Supplementary Excel Table S8

biosynthesis of carbohydrates, amino acids and vitamins/cofactors, which supply nutrients to the host. Specifically, *P. echinospica* could produce 4 vitamins/cofactors and 14 amino acids at least, whereas its symbionts can produce 13 vitamins/cofactors and 18 amino acids (Table 2). Nutrient interdependence between *P. echinospica* and its symbiont could be demonstrated by their cooperation in nutrient production, for example, the biosynthesis of methionine. Methionine that serves as an essential amino acid of most metazoans was detected in *Lamellibrachia* sp. and *Escarpia* sp. [83]. Currently, the endosymbionts of Siboglinidae cannot use cystathionine to synthesize methionine because of the lack of gene *metC* or *patB*, instead they use homoserine to synthesize methionine (Supplementary Fig. S12a). In *Paraescarpia* holobiont, host transcriptome contained the key genes (e.g., *CBS* and

*serB*, etc.) responsible for producing cystathionine, and the symbiont genome contains *patB* and *metH* genes for using cystathionine to synthesize methionine (Supplementary Fig. S12b). Thus, we hypothesize that the *Paraescarpia* symbiont has the potential to use the cystathionine produced by the host to synthesize methionine for the holobiont use which indicates the complementary ability in nutrient production in *Paraescarpia* tubeworm holobiont.

The genome of the *P. echinospica* symbiont encoded only ten transporters, including transporters for minerals, polysaccharides and lipids (Supplementary Table S9), it had no substrate-specific transporters for amino acids and vitamins, suggesting that the *P. echinospica* symbiont cannot transport the nutrients to the host efficiently. The symbiont genomes of *Riftia pachyptila* [14], *Calyptogenia* clam [84]



**Fig. 4** An overview of metabolic pathways of the *P. echinospica* endosymbiont. Different metabolic pathways are presented in squares of different colors. Nitrogen metabolism is in a light blue square, including dissimilatory nitrate reduction, denitrification, and ammonia assimilation. Carbon metabolism is in a green square, including CBB and rTCA cycles for carbon fixation, TCA and glycolysis cycles for organic carbon utilization and bidirectional reactions of carbon monoxide and formate. Sulfur metabolism is in a yellow square. The sulfur oxidation depends on the Dsr, Apr and Sox systems. The sulfur globule protein is highly expressed and acts as an energy storage compound. The hydrogen oxidation is in a blue violet square. The above energy-conversion pathways provide substrates and energy for the production of nutrients such as amino acids and vitamins (Table 2). Enzymes found in both the symbiont genome and transcriptome are

shown in red, whereas those found in the symbiont genome only are shown in yellow, and the missing enzymes are shown in gray. The histogram at the bottom shows the relative gene expression levels (log<sub>10</sub>TPM) of enzymes in different metabolic pathways and key proteins involved in intracellular survival mechanisms. The membrane transport proteins, bacterial chemotaxis proteins and some of the characterized proteins for bacterial infection, which were encoded in the symbiont genome but not expressed, are marked with dashed circles. The flagellum, fimbriae and pilus of the symbiont, which were encoded in the symbiont genome but not expressed, are indicated in dashed line. The full names of enzymes are given in the list of abbreviations in Supplementary Information, and the involved genes are listed in Supplementary Table S9

and the Bathymodiolin mussel [85] encode few substrate-specific transporters as well, which suggests the symbionts are either leaky or digested by their host for nutrients. Furthermore, transcriptome and real-time PCR analyses of *P. echinospica* showed that several digestive enzymes were specific to the trophosome (Supplementary Table S11, Fig. 5), and these enzymes can aid in the digestion of symbionts [86]. Thus, in the tubeworm *P. echinospica*, digestion of symbionts and carbon translocation from

symbionts to the host are key processes by which the host acquires nutrients from the symbionts and controls the symbiont population.

**Virulence and nutrient acquisition of symbiotic bacteria**

A large number of genes were found to encode proteases in the symbiont genome, and multiple bacterial proteinases

**Table 2** Capability of biosynthesis of amino acids, vitamins and cofactors in *Paraescarpia echinospica* and its symbiont

Nutrients	Description	Symbiont	<i>P. echinospica</i>
Biosynthesis of amino acids			
NEFAAs	A, C, D, E, G, P, Q, R, S, Orn	+	+
EFAAs	H, I, K, L, M, V, W, bA	+	–
	Ta, hypoTa, T, Y	–	+
	F, N	–	–
Biosynthesis of vitamins and cofactors			
Vitamin B1	Thiamine	+	–
Vitamin B2	Riboflavin	+	–
Vitamin B3	Nicotinate and nicotinamide	–	+
Vitamin B5	Pantothenate	+	+
Vitamin B6	Pyridoxine	+	+
Vitamin B7	Biotin	+	–
Vitamin B9	Folate	+	–
Vitamin B12	Cobalamin	–	–
Vitamin K2	Menaquinone	+	–
Coenzyme A	CoA	+	+
Coenzyme Q	Ubiquinone	+	–
	Protoheme (heme)	+	–
	Siroheme	+	–

Amino acids: A—Alanine; bA— $\beta$ -Alanine; C—Cysteine; D—Aspartate (aspartic acid); E— Glutamic acid; F—Phenylalanine; G—Glycine; H—Histidine; hypoTa—Hypotaurine; I—Isoleucine; K—Lysine; L—Leucine; M—Methionine; N—Asparagine; Orn—Ornithine; P—Proline; Q—Glutamine; R—Arginine; S—Serine; T—Threonine; Ta—Taurine; V—Valine; W—Tryptophan; Y—Tyrosine; Complete and missing pathways are indicated by ‘+’ and ‘–’, respectively

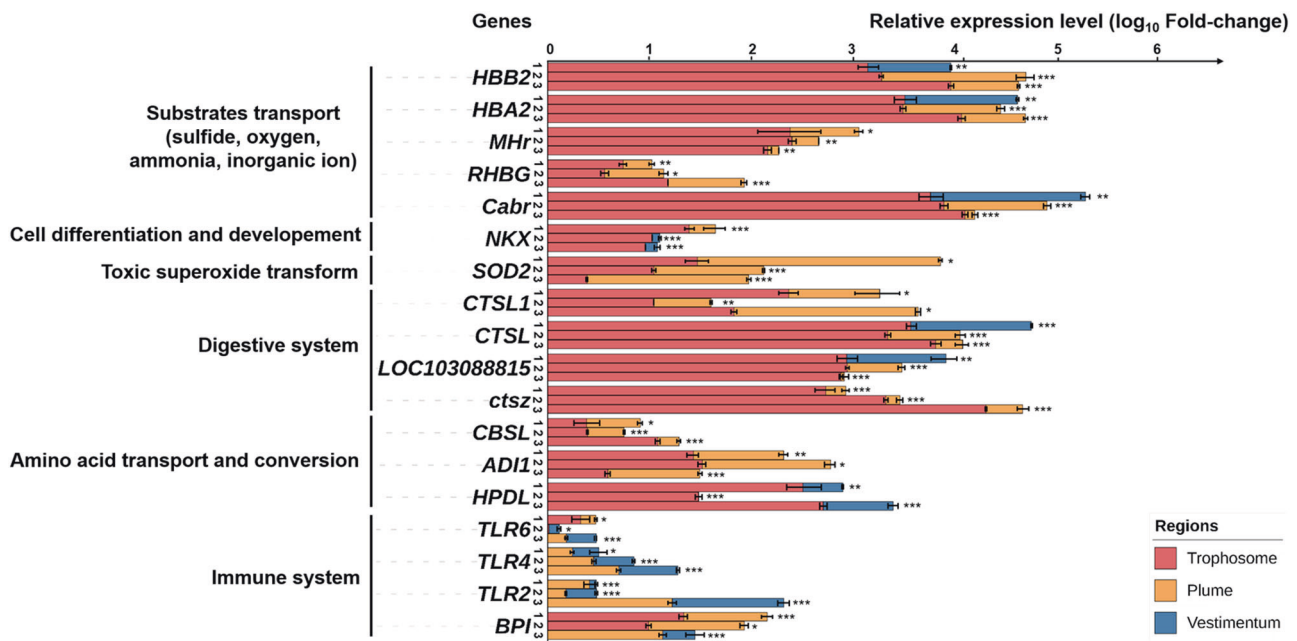
were found expressed highly in the symbiont transcriptome (Table 3). The proteinases have broad specificity, as they can degrade host proteins, including those associated with immune response proteins such as various immunoglobulins, cytokines and chemokins, etc [87]. Notably, the *Paraescarpia* symbiont transcriptome and proteome contained high expression (2nd in transcriptome, 31th in proteome) of putative secreted esterase (PSE) (Fig. 3a, b), which is important in the bacterial virulence and pathogenesis [88] and may also function as a digestive enzyme for degradation of host animal cells [87]. Similarly, the symbiont transcriptome of *P. echinospica* contained the endochitinase ChiA (EC 3.2.1.14) and peptidase M48 (EC 3.4.24) responsible for the degradation of the structural barriers of the host by pathogens [89, 90] as well as various proteolytic enzymes in the symbiont proteome, such as the HtrA (DegP) serine protease (RLJ17779.1), peptidase M16 (RRS32426.1) and peptidase S41 (RDH90197.1). Serine protease is a cell envelope proteinase that can diminish

function of the signal proteins manufactured by the host, it modulates the host immune response as a virulence factor by being anchored to the cell by sortase A that inactivates the complement factor of the host cell which is a key component of innate immune response [87, 91]. Cysteine proteases enhance bacterial ability to evade host innate immune response by degradation of host extracellular matrix material [87]. Thus, we hypothesize that the endosymbionts have the ability to modulate host immune response, degrade host cells and obtain nutrients by using their highly expressed proteases as virulence factors.

The symbiont genome of *P. echinospica* contained the OmpA-OmpF porin (OOP family) and OmpR families (OmpR-EnvZ and PhoP/Q systems) (Supplementary Table S12), which play important pathogenic roles such as bacterial adhesion and invasion in symbiotic-pathogenic bacteria [92–94]. These proteins can also promote bacterial intracellular survival and evasion of host defenses. Moreover, genes *ompA*, *ompR*, and *envZ* were highly expressed in the symbiont transcriptome and proteome (Fig. 3a, b). OmpR-EnvZ controls the bacterial virulence as a result of its ability to survive intracellularly [92] and PhoP/Q promotes bacterial resistance to the host’s innate immune response through bacterial surface modification [95]. The modified bacterial surface can enhance bacterial ability of immune evasion by the effects on activating factors and not activate host immune response [92, 93, 95]. These two are indispensable to the self-protection of bacteria after entering the host epidermal cells. Thus, the high expression of the OmpR-EnvZ system and OmpA proteins represents an adaptation in the symbionts that mediates host tolerance of symbiotic bacteria for their intracellular surviving, which may be critical for maintaining a stable symbiotic relationship in the *P. echinospica* holobiont. The bacterial infection process and proliferation post infection in *P. echinospica* are explained in the section of symbiont infection in Supplementary Information.

### Host innate immune responses

Unexpectedly, the expression level of immune-related genes in the bacteria-concentrated region of the trophosome was not higher than other two regions (Fig. 3c, Fig. 5). Among them, besides the bactericidal/permeability-increasing (BPI) protein highly expressed in the trophosome and plume as part of the innate immune system (Fig. 5) [96], Toll-like receptors (TLR2, TLR4, and TLR6) are key to the innate immune system and can recognize intruders and activate immune responses. TLR6 functionally interacts with TLR2 to mediate the cellular response to bacterial lipoproteins [97]. In the present study, expression level of the genes encoding TLRs was higher in the plume (the region in contact with free-living bacteria) and vestimentum than in the trophosome (Fig. 5). However, previous study showed the expression of TLR and



**Table 3** Highly expressed bacterial proteinases in the endosymbiont of *Paraescarpia echinospica*

Transcript ID	Bacterial proteinases	TPM value
	Annotation	Description
Pec_DN78589C0G1I1	Putative secreted esterase	Proteolysis; serine-type endopeptidase activity
Pec_DN33952C0G1I1	Membrane protease subunit HflK	Peptidase activity; membrane; proteolysis; integral component of membrane
Pec_DN70443C1G1I1	ATP-dependent Lon protease	ATP-dependent Serine peptidase MEROPS family S16
Pec_DN73964C0G1I1	ATP-dependent metalloprotease	Cell division protease FtsH ( <i>ftsH</i> , <i>hflB</i> )
Pec_DN5083C0G2I1	ATP-dependent Clp protease ATP-binding subunit	Peptidase activity; proteolysis; ATP-dependent peptidase activity
Pec_DN103576C0G1I1	ATP-dependent HslUV protease ATP-binding subunit HslU	HslUV protease complex; proteolysis; peptidase activity, acting on L-amino acid peptides
Pec_DN101171C0G1I1	ATP-dependent HslUV protease, peptidase subunit HslV	HslUV protease complex; proteolysis;
Pec_DN4612C0G1I1	Peptidase S49	<i>sppA</i> ; protease IV; peptidase activity; proteolysis
Pec_DN49575C1G1I1	ATP-dependent chaperone	<i>clpB</i> ; ATP-dependent Clp protease ATP-binding subunit ClpB
Pec_DN109493C0G1I1	Peptidase S41	Serine-type peptidase activity
Pec_DN26577C0G1I1	Peptidase C-terminal protease	Proteolysis; serine-type peptidase activity
Pec_DN49575C0G1I1	Disaggregation chaperone	<i>clpB</i> ; ATP-dependent Clp protease ATP-binding subunit ClpB
Pec_DN11727C0G1I1	Cysteine protease	Proteolysis; cysteine-type peptidase activity
Pec_DN40710C0G1I1	Membrane protease subunit HflC	Peptidase activity; membrane; proteolysis; integral component of membrane
Pec_DN109409C0G1I1	Zn-dependent protease	Peptidase activity; proteolysis; chaperone-mediated protein folding



PGRP genes in the trophosome was between five and 100-fold higher than that in the plume of vent-vestimentiferan *R. piscesae*, and expression of innate immunity genes in the trophosome of *R. piscesae* were higher than in the plume and as a whole may regulate the immune response to shape the symbiosis and to maintain symbiostasis [15]. Combining these observations with the above findings in the *Paraescarpia* endosymbionts, we speculate that the regulatory mechanism of *Paraescarpia* endosymbionts on host immunity make the trophosome possess a relatively weak immune system. More detailed information of the immune-related genes in the three regions is shown in the section on host innate immune responses in Supplementary Information (Supplementary Table S13). Consequently, we hypothesize that the endosymbiont of *Paraescarpia* has evolved elaborate strategies to distract the host's protective immunity and evade its defenses.

## Conclusions

Our integrated genomic, transcriptomic, and proteomic analysis of the *P. echinospica* holobiont revealed metabolic, nutritional, and regulatory interdependencies in symbiosis, a key adaptation allowing the tubeworm to thrive in cold-seep chemosynthetic ecosystems. Genomic comparisons of vestimentiferan endosymbionts showed that the *Paraescarpia* symbionts had a high potential to evade the host immune response, reduce the toxicity of organic nitrogen compounds and transport proteins between cells. Analyses of the energy and nutrient pathways indicated a strong interdependence between *P. echinospica* and its symbionts in energy consumption and nutrient production. The bacterial symbionts may be able to degrade host proteins and use host cells as a nutrient source by using various virulence factors as digestive enzymes. The symbiont of *Paraescarpia* is believed to have evolved strategies to mediate host innate immunity as immune response genes performed not prominently as expected in the trophosome. Our findings suggest that the maintenance of host-microbiota dynamics is determined by the holobiont's evolved interdependence, which provides a new insight into the adaptation of deep-sea chemosynthetic holobionts.

**Acknowledgements** We thank the captain and crew of R/V Tan Ka Kee and the operation team of ROPOS for collecting the sample, and Dr Weipeng Zhang from the Ocean University of China and Dr Yuanning Li from Yale University for their helpful comments on an earlier draft of the paper. This study was supported by China Ocean Mineral Resource Research and Development Association (DY135-E2-1-03), and Hong Kong Branch of South Marine Science and Engineering Guangdong Laboratory (SMSEGL20Sc01) to PYQ, and National Key R&D Program of China (2018YFC0310005) to JWQ. Alice Cheung edited the final version of the paper.

**Author contributions** PYQ and JWQ conceived the project. JS and YY designed the experiments. JWQ and DF collected the samples. YJZ, TX and YY extracted the RNA. YY extracted the DNA. YHK, WCW and YY performed the meta-proteomic and real-time PCR experiments. YY performed the remaining bioinformatics analysis and drafted the paper. All of the authors have read and edited the paper and approved its submission.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Sibuet M, Olu K. Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins. *Deep-Sea Res Pt II*. 1998;45:517–67.
2. Takishita K, Kakizoe N, Yoshida T, Maruyama T. Molecular evidence that phylogenetically diverged ciliates are active in microbial mats of deep-sea cold-seep sediment. *J Eukaryot Microbiol*. 2010;57:76–86.
3. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol*. 2008;6:725–40.
4. Halanych KM. Molecular phylogeny of siboglinid annelids (a.k.a. pogonophorans): a review. *Hydrobiologia*. 2005;535:297–307.
5. Nussbaumer AD, Fisher CR, Bright M. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature*. 2006;441:345–8.
6. Thornhill DJ, Wiley AA, Campbell AL, Bartol FF, Teske A, Halanych KM. Endosymbionts of *Siboglinum fiordicum* and the phylogeny of bacterial endosymbionts in Siboglinidae (Annelida). *Biol Bull-US*. 2008b;214:135–44.
7. Minic Z, Hervé G. Biochemical and enzymological aspects of the symbiosis between the deep-sea tubeworm *Riftia pachyptila* and its bacterial endosymbiont. *Eur J Biochem*. 2004;271:3093–102.
8. Girguis PR, Lee RW, Desaulniers N, Childress JJ, Pospesel M, Felbeck H, et al. Fate of nitrate acquired by the tubeworm *Riftia pachyptila*. *Appl Environ Microbiol*. 2000;66:2783–90.
9. Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh C, Huber JA. Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome*. 2018;6:19.
10. Li Y, Liles MR, Halanych KM. Endosymbiont genomes yield clues of tubeworm success. *ISME J*. 2018;12:2785–95.

11. Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thurmer A, Albrecht D, et al. Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J.* 2012;6:766–76.
12. Perez M, Juniper K. Insights into symbiont population structure among three vestimentiferan tubeworm host species at eastern Pacific spreading centers. *Appl Environ Microbiol.* 2016;82:5197–205.
13. Goffredi SK, Yi H, Zhang Q, Klann JE, Struve IA, Vrijenhoek RC, et al. Genomic versatility and functional variation between two dominant heterotrophic symbionts of deep-sea *Osedax* worms. *ISME J.* 2014;8:908–24.
14. Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, et al. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ Microbiol.* 2008;10:727–37.
15. Nyholm SV, Song PF, Dang JN, Bunce C, Girguis PR. Expression and putative function of innate immunity genes under in situ conditions in the symbiotic hydrothermal vent tubeworm *Ridgeia piscesae*. *Plos One.* 2012;7:e38267.
16. Sun Y, Liang Q, Sun J, Yang Y, Tao J, Liang J, et al. The mitochondrial genome of the deep-sea tubeworm *Paraescarpia echinospica* (Siboglinidae, Annelida) and its phylogenetic implications. *Mitochondr DNA Pt B.* 2018;3:131–2.
17. Feng D, Qiu JW, Hu Y, Peckmann J, Guan H, Tong H, et al. Cold seep systems in the South China Sea: an overview. *J Asian Earth Sci.* 2018;168:3–16.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
19. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77.
20. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol.* 2013;31:533–8.
21. Tian R-M, Sun J, Cai L, Zhang W-P, Zhou G-W, Qiu J-W, et al. The deep-sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. *Environ Microbiol.* 2016;18:2481–94.
22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
23. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics.* 2012;28:2223–30.
24. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inf.* 2009;23:205–11.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
26. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 2011;21:1552–60.
27. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
28. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform.* 2014;15:211.
29. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol.* 2012;13:R56.
30. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 2010;11:119.
31. Wu ST, Zhu ZW, Fu LM, Niu BF, Li WZ. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics.* 2011;12:444.
32. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35:W182–W185.
33. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
34. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* 2011;12:402.
35. Darzentas N. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics.* 2010;26:2620–1.
36. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinform.* 2011;12:124.
37. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–3.
38. Stamatakis A, Ludwig T, Meier H. RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics.* 2005;21:456–63.
39. Li Y, Kocot KM, Schander C, Santos SR, Thornhill DJ, Halanych KM. Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). *Mol Phylogenetics Evol.* 2015;85:221–9.
40. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25:1189–91.
41. Wheeler NE, Barquist L, Kingsley RA, Gardner PP. A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *Bioinformatics.* 2016;32:3566–74.
42. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. *Nat Rev Genet.* 2018;19:549–65.
43. Daniels C, Baumgarten S, Yum LK, Michell CT, Bayer T, Arif C, et al. Metatranscriptome analysis of the reef-building coral *Orbicella faveolata* indicates holobiont response to coral disease. *Front Mar Sci.* 2015;2:62.
44. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52.
45. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22:1658–9.
46. Smith-Unna R, Bournsnel C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 2016;26:1134–44.
47. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013;8:1494–512.
48. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–2.
49. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21:3674–6.
50. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14:417–9.
51. Sun J, Zhang Y, Xu T, Zhang Y, Mu HW, Zhang YJ, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat Ecol Evol.* 2017;1:0121.

52. Vizcaíno JA, Csordas A, del-Toro N, Dianas JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 2016;44:D447–D456.
53. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods.* 2001;25:402–8.
54. Zheng S, Sham LT, Rubino FA, Brock KP, Robins WP, Mekalanos JJ, et al. Structure and mutagenic analysis of the lipid II flippase MurJ from *Escherichia coli*. *Proc Natl Acad Sci USA.* 2018;115:6709–14.
55. Naismith JH, Booth IR. Bacterial mechanosensitive channels-MscS: evolution's solution to creating sensitivity in function. *Annu Rev Biophys.* 2012;41:157–77.
56. Cosme AM, Becker A, Santos MR, Sharypova LA, Santos PM, Moreira LM. The outer membrane protein TolC from *Sinorhizobium meliloti* affects protein secretion, polysaccharide biosynthesis, antimicrobial resistance, and symbiosis. *Mol Plant Microbe Interact.* 2008;21:947–57.
57. Zgurskaya HI, Krishnamoorthy G, Ntrel A, Lu S. Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria. *Front Microbiol.* 2011;2:189.
58. Galperin MY. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol.* 2006;188:4169–82.
59. Freytag JK, Girguis PR, Bergquist DC, Andras JP, Childress JJ, Fisher CR. A paradox resolved: sulfide acquisition by roots of seep tubeworms sustains net chemoautotrophy. *Proc Natl Acad Sci USA.* 2001;98:13408–13.
60. Vrijenhoek RC, Duhaime M, Jones WJ. Subtype variation among bacterial endosymbionts of tubeworms (Annelida: Siboglinidae) from the Gulf of California. *Biol Bull.* 2007;212:180–4.
61. Patra AK, Cho HH, Kwon YM, Kwon KK, Sato T, Kato C, et al. Phylogenetic relationship between symbionts of tubeworm *Lamellibrachia satsuma* and the sediment microbial community in Kagoshima Bay. *Ocean Sci J.* 2016;51:317–32.
62. Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, DeChaine EG, et al. Free-living tube worm endosymbionts found at deep-sea vents. *Appl Environ Microbiol.* 2008;74:3895–8.
63. Meo CAD, Wilbur AE, Holben WE, Feldman RA, Vrijenhoek RC, Cary SC. Genetic variation among endosymbionts of widely distributed vestimentiferan tubeworms. *Appl Environ Microbiol.* 2000;66:651–8.
64. Bowles M, Joye S. High rates of denitrification and nitrate removal in cold seep sediments. *ISME J.* 2011;5:565–7.
65. Joye SB, Boetius A, Orcutt BN, Montoya JP, Schulz HN, Erickson MJ, et al. The anaerobic oxidation of methane and sulfate reduction in sediments from Gulf of Mexico cold seeps. *Chem Geol.* 2004;205:219–38.
66. Damm KLV, Edmond JM, Measures CI, Grant B. Chemistry of submarine hydrothermal solutions at Guaymas Basin. *Gulf Calif Geochim Cosmochim Acta.* 1985;49:2221–37.
67. Watanabe H, Fujikura K, Kojima S, Miyazaki J, Fujiwara Y. Japan: vents and seeps in close proximity. In: Kiel S, editor. *The vent and seep biota: aspects from microbes to ecosystems.* Dordrecht: Springer; 2010. p. 379–401.
68. Levin LA. Ecology of cold seep sediments: Interactions of fauna with flow, chemistry and microbes. *Oceanogr Mar Biol: Annu Rev.* 2005;43:1–46.
69. Pachiadaki MG, Kormas KA. Interconnectivity vs. isolation of prokaryotic communities in European deep-sea mud volcanoes. *Biogeosciences.* 2013;10:2821–31.
70. Boetius A, Ravensschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature.* 2000;407:623–6.
71. Frank SA. Host-symbiont conflict over the mixing of symbiotic lineages. *Proc Biol Sci.* 1996;263:339–44.
72. Dobrindt U, Hochhut B, Hentschel U, Hacker J. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol.* 2004;2:414–24.
73. Jahn MT, Arkhipova K, Markert SM, Stigloher C, Lachnit T, Pita L, et al. A symbiont phage protein aids in eukaryote immune evasion. *bioRxiv* 2019; 608950.
74. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007;315:1709–12.
75. Roldán MD, Pérez-Reinado E, Castillo F, Moreno-Vivián C. Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *FEMS Microbiol Rev.* 2008;32:474–500.
76. Jani AJ, Cotter PA. Type VI secretion: not just for pathogenesis anymore. *Cell Host Microbe.* 2010;8:2–6.
77. Guckes KR, Cecere AG, Wasilko NP, Williams AL, Bultman KM, Mandel MJ, et al. Incompatibility of *Vibrio fischeri* strains during symbiosis establishment depends on two functionally redundant *hcp* genes. *J Bacteriol.* 2019. <https://doi.org/10.1128/JB.00221-19>.
78. Li Y, Kocot KM, Whelan NV, Santos SR, Waits DS, Thornhill DJ, et al. Phylogenomics of tubeworms (Siboglinidae, Annelida) and comparative performance of different reconstruction methods. *Zool Scr.* 2016;46:200–13.
79. Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, et al. Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from seagrass sediments. *Environ Microbiol.* 2015;17:5023–35.
80. Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, et al. Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics.* 2011;11:3106–17.
81. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci USA.* 2012;109:E1173–E1182.
82. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, et al. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science.* 2007;315:247–50.
83. Pruski AM, Fiala-Medioni A, Fisher CR, Colomines JC. Composition of free amino acids and related compounds in invertebrates with symbiotic bacteria at hydrocarbon seeps in the Gulf of Mexico. *Mar Biol.* 2000;136:411–20.
84. Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, et al. The *Calyptogenia magnifica* chemoautotrophic symbiont genome. *Science.* 2007;315:998–1000.
85. Ponnudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, et al. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. *Stand Genom Sci.* 2017;12:50.
86. Wippler J, Kleiner M, Lott C, Gruhl A, Abraham PE, Giannone RJ, et al. Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. *BMC Genomics.* 2016;17:942.
87. Hynes W, Sloan M. Secreted extracellular virulence factors. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. *Streptococcus pyogenes: basic biology to clinical manifestations*, internet. OK, USA: University of Oklahoma Health Sciences Center; 2016. p. 405–44.
88. Zhu H, Liu M, Sumbly P, Lei B. The secreted esterase of group A *Streptococcus* is important for invasive skin infection and dissemination in mice. *Infect Immun.* 2009;77:5225–32.
89. Sirakova TD, Markaryan A, Kolattukudy PE. Molecular cloning and sequencing of the cDNA and gene for a novel elastolytic metalloproteinase from *Aspergillus fumigatus* and its expression in *Escherichia coli*. *Infect Immun.* 1994;62:4208–18.

90. Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J, et al. Chitinases: an update. *J Pharm Bioallied Sci.* 2013;5:21–29.
91. Ruiz-Perez F, Nataro JP. Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence. *Cell Mol Life Sci.* 2014;71:745–70.
92. Forst SA, Nealson K. Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdus* spp. and *Photorhabdus* spp. *Microbiol Rev.* 1996;60:21–43.
93. Weiss BL, Wu Y, Schwank JJ, Tolwinski NS, Aksoy S. An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. *Proc Natl Acad Sci USA.* 2008;105:15088–93.
94. Groisman EA. The pleiotropic two-component regulatory system PhoP-PhoQ. *J Bacteriol.* 2001;6:1835–42.
95. Ernst RK, Guina T, Miller SI. How intracellular bacteria survive: surface modifications that promote resistance to host innate immune responses. *J Infect Dis.* 1999;179:S326–S330.
96. Gonzalez M, Gueguen Y, Destoumieux-Garzón D, Romestand B, Fievet J, Pugnère M, et al. Evidence of a bactericidal permeability increasing protein in an invertebrate, the *Crassostrea gigas* Cg-BPI. *Proc Natl Acad Sci USA.* 2007;104:17759–64.
97. Akira S, Yamamoto M, Takeda K. Toll-like receptor family: receptors essential for microbial recognition and immune responses. *Arthritis Res Ther.* 2003;5:S5.