Extracellular chemosymbiont populations in a shallow-water thyasirid clam potentially shaped by priority effect

Menggong Li[1,2#], Yunlong Li[1,2#], Shi-Hai Mao[3], Xu Liu[1,2], Hui Wang[1,2], Chong Chen[4], Xiaoshou Liu[1], Guang-Chao Zhuang[3], Weipeng Zhang[1], Jin Sun[1,2]*

[1] Key Laboratory of Evolution & Marine Biodiversity (Ministry of Education) and Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China

[2] Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Laoshan Laboratory, Qingdao 266237, China

[3] Frontiers Science Center for Deep Ocean Multispheres and Earth System, and Key Laboratory of Marine Chemistry Theory and Technology, Ministry of Education, Ocean University of China, Qingdao, 266100, China

[4] X-STAR, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa Prefecture 237-0061, Japan

[#] equal contribution

* Corresponding authors: Prof. Jin Sun, jin_sun@ouc.edu.cn

Competing interests, The authors declare no competing interests.

Abstract

Chemosymbiosis is a highly successful strategy that evolved in several animal groups, best known for dominating chemosynthetic ecosystems such as deep-sea hot vents and hydrocarbon seeps but also found in other systems such as reducing sediments in shallow water. The symbiont population structure can be determined by the host genetic inheritance, geographic partitions, and environmental factors, among others if the symbionts are acquired horizontally from the environment. Here, we suggest that the earliest colonies can also influence the episymbiont population, known as the "priority effect", using the thyasirid cleftclam *Thyasira tokunagai*. This species is abundant in reducing sediments in Yellow Sea, hosting an extracellular sulfide-oxidizing symbiont (genus *Sedimenticola*) in the pouch-like structure in the gill. From samples taken across the whole Yellow Sea, complete symbiont genomes revealed only two dominant strains which were also verified by phylogenomic analysis. The two strains share key functional genes but exhibit a single-base difference in the 16S rDNA. We found mutually exclusive communities of these two strains in the 30 individual hosts studied, typically dominated or even monopolized by one strain. Inconsistent phylogenetic relationships between hosts and symbiont indicate the genetic heterogeneity of the holobiont, and the mean fixation index F_{ST} of each symbiont population within sampling sites showed a lack of correlation with geographic distance and environmental factors, suggesting another driving force. We deduce that the likely scenario is 1) the dominant symbiont strain is determined during initial acquisition; 2) the propagation of this initial strain as the host develops; 3) the continuous but restricted recruitment of the other strain during the adult stage. Overall, our results highlight the importance of "priority effect" in shaping the chemosymbiont population structures even in the horizontally transmitted episymbiont in a whole marginal sea area.

Keywords chemosymbiosis; priority effect; thyasirid; SNPs; Yellow Sea

Introduction

The close association between eukaryotes and microorganisms, or symbiosis, has been ubiquitously reported across various host groups and both terrestrial and aquatic ecosystems [1, 2]. Symbionts play a crucial role in helping their host in nutritional acquisition and pathogen defense, in turn gain advantages in the development, local adaptation, speciation, and evolution [3, 4]. A key example of symbiotic mutualism is between invertebrate animals and chemosynthetic bacteria found in chemosynthesis-based marine ecosystems, best known for deep-sea hydrothermal vents and hydrocarbon seeps in the absence of sunlight. The bacteria oxidize reducing substances (e.g., hydrogen gas H_2, hydrogen sulfide H_2S, and methane CH_4) and produce organic matter for their hosts [5], such as siboglinid tubeworms such as *Riftia* [6-8], abyssochrysoidean and peltospirid snails [9-11], bathymodioline mussels [12, 13], vesicomyid and thyasirid clams [14, 15], forming the base of diverse and flourishing communities.

Most organisms host just one specific symbiont lineage, indicating strong selection and cooperation within the holobiont. Different hosts in the same environment often host phylogetically distinct symbionts, even when the symbionts are functionally similar [11, 15]. The chemosymbionts are mostly methane- or sulfide-oxidizing bacteria (MOB and SOB); Gammaproteobacteria is the most common while Alphaproteobacteria or Epsilonproteobacteria have also been reported [16, 17]. Some hosts are capable of hosting multiple clades of symbionts, for instance, pliocardiine clams in Vesicomyidae exclusively host two clades of SOB (genera *Ruthia* and *Vesicomyococius*) [18, 19], and a similar pattern has also been reported in vestimentiferan tubeworms and most lucinid clams [20, 21]. The peltopsirid snail *Gigantopelta aegis* has dual symbiosis [11], while highly diverse symbiotic communities are known from the mussel *Bathymodiolus azoricus* [22, 23].

Chemoysmbiosis is often obligatory for the host's development due to their heavy energy dependence on symbionts, typically with reduced digestive systems compared to their asymbiotic relatives [24-26]. The mode of symbiont acquisition is variable. For instance, stringent vertical acquisition has been proposed in pliocardiine clams, evidenced by the presence of symbionts in the gonad and larvae, and the overall congruent coupled phylogeny between host and symbiont [27-29], though some phylogenetic incongruencies in species such as *Turneroconcha magnifica* and *Calyptogena fausta* suggest occasional hybrid or horizontal acquisition across different host

taxa [19]. Most other hosts, however, use horizontal transfer where they acquire symbionts from the environmental pool post-settlement. Horizontal acquisition is typified by a more diverse symbiont community, with the absence of symbionts in gonads and at the larval stages, and also the inconsistent or even disordered phylogeny between host and symbiont seen in bathymodioline mussels [12, 30, 31], lucinid clams [32], and *Alviniconcha* snails [33]. A mixed mode of the above two acquisition modes has been proposed in solemyid clams and the peltospirid snail *Chrysomallon squamiferum* [34, 35].

It has been hypothesized that locally adapted symbiont strains colonize the host symbiotic organ, which may aid hosts to better survive in its immediate environment and allow for success in a wider range of habitat conditions. This assumes that there are relatively homogeneous communities of symbionts at a local scale but heterogeneous among sites. Previous studies demonstrated that the symbiont community in horizontally acquired symbiosis is linked to biogeographic distances, environment, and host genetics [35]. The geographic difference explains most of the symbiont composition in the mussel inhabiting hydrothermal vents across the Mid-Atlantic Ridge [36]. In this case, 16 strains can co-existing in the gill tissue of the host, while the host genetic influence and geographic division play important roles in structuring the symbiont population [37]. Similarly, the mixed acquisition of symbiont strains promotes the local adaptation and evolutionary success of *Chrysomallon squamiferum* snails in different vents [9]. Moreover, the genetic diversity of horizontally acquired symbionts in a single host is reported as varying with the acquisition period in the life span, reflecting intra-host homogeneity and inter-host heterogeneity. The high pairwise F_{ST} values in *Bathymodiolus brooksi* mussels indicated the uptake of environmental bacteria in a restricted process and the communities in them might be determined by the earlier colonization via self-infection [38]. However, for these reported studies, these symbionts are exclusively endosymbiotic within host cell's cytoplasm, it is unknown whether the episymbionts population structures are shaped similarly or by more complicated factors.

The episymbionts live on the surface of the epithelial area, meaning they are more able to communicate with other bacteria and the surrounding environment. For instance, the shrimp *Rimicaris exoculate* could acquire epibiotic bacteria along the life cycle [39]. Here, our study focuses on the process of episymbiont population structure in chemosynthetic holobionts, using

the thyasirid cleftclam *Thyasira tokunagai* in the Yellow Sea as a case study. *Thyasira tokunagai* is a member of the widely distributed *T. gouldii* species complex commonly found in reducing sediments. Through various analyses, including examining the chemosymbiotic basis, the 16S rRNA gene of symbionts, transmission electron microscopy (TEM) and fluorescence *in situ* hybridization (FISH) analyses on the symbiotic mode, and the carbon fixation rate via radioactive carbon tracing, we gained insights into the mechanisms that facilitate these interactions. By sequencing 30 individuals from nine populations across the whole Yellow Sea, we discovered that despite the homogeneity of the thyasirid clam, it hosts two dominant strains with a single base difference in the 16S rRNA gene. This allowed us to decipher the symbiont's pan- and core-genomic features, the symbiont, the strain-level population structures of the symbiont via single nucleotide polymorphisms (SNPs) analyses, population structure via full mitogenome sequencing of the host, and the gene expression. Our findings shed light on the major internal factors that shape intra-host structure for episymbionts and elucidate the potential host-symbiont interactions in the chemosymbiotic marine invertebrates.

Materials and methods

Sampling description

Thyasira tokunagai (Fig. 1a) were collected from a total of twelve sites in the Yellow Sea between 42-72 m depth from three cruises on-board the R/V *Lanhai 101* in August 2020, April 2021, and October 2021, with the site details and environmental parameters shown in Table S1 andTable S2. A 0.1 m^2 box corer was employed to collect the surface sediments (~2 cm). The *T. tokunagai* were manually picked out from the sediments via a 0.5 mm aperture mesh once they were on board and then were properly fixed immediately according to the morphology. Sample storage details can be found in supplementary information.

Nucleic acid extraction, library construction, and sequencing

The soft tissues of *T. tokunagai* were used for DNA extraction using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the company’s protocol. Meanwhile, DNA was also

extracted from approximately 0.5g of ambient surface sediments (wet weight) by using the PowerSoil DNA Isolation Kit (Qiagen, Hilden, Germany). NanoDrop Lite (Thermo Scientific, USA) and 1% agarose gel electrophoresis were used to check the DNA quantity and quality, respectively. The mitochondrial cytochrome *c* oxidase subunit I (COI) gene fragment was amplified by the universal degenerated primers [40]. Similarly, the full-length 16S rRNA gene of bacteria from sediments and *T. tokunagai* gill were also amplified by the primers 27F and 1492R [41]. The PCR product of the COI gene was sequenced by BGI (Qingdao, China) from both ends using Sanger sequencing. The PCR product of the 16S rRNA gene was sequenced on the platform of the PacBio RS II System by Novogene (Beijing, China) with CCS mode. For the Metagenome sequencing, DNA libraries were constructed using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. For oxford nanopore technologies (ONT) sequencing, the DNA of whole tissue was extracted using QIAGEN Genomic-tip 100/G (QIAGEN, Germany) according to the manufacturer's instructions, and the library was generated using SQK-LSK109 (Oxford Nanopore Technologies, UK) following the manufacturer's recommendations.

Total RNA was extracted using TRIzol reagent (Invitrogen, CA, USA) with the guidance of the manufacturer's protocol. RNA integrity and quantity were measured using the Bioanalyzer 5400 system (Agilent Technologies, CA, USA). cDNA was obtained by removing the prokaryotic and eukaryotic ribosomal RNA from the total RNA for the construction of a meta-transcriptomic library. The nucleic acid for metagenome and meta-transcriptome sequencing was subjected to Illumina NovaSeq 6000 sequencing at Novogene (Beijing, China) with paired-end mode and a read length of 150bp, and the ONT library was subjected to PromethION platform at Novogene (Beijing, China). All sequencing details are found in Table S3

Mitogenome assembly, annotation, and population structure analysis

Raw reads were trimmed to remove low-quality sequences and adapters using Trimmomatic v.0.39 [42] with the following parameters: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10, LEADING:20, TRAILING:20, SLIDINGWINDOW:4:15, MINLEN:100. NOVOPlasty v.4.3.1 [43] were employed to construct the mitochondrial genomes with default parameter (160M reads randomly selected per sample). Assembled mitogenomes were annotated on the MITOS web

server [44] with the default setting except "the genetic code: 5 invertebrates", and further checked manually. Afterwards, 13 protein-coding genes (PCGs) of 30 individuals were aligned separately using MAFFT v.7.515 [45] with the default parameter. Population structure analysis was performed by STRUCTURE v.2.3.4 [46] with the settings of "*K*: from 2 to 7, 2,000,000 iterations and 10% of burnin". The most optimal *K* was determined using Structure Harvester [47] web server with the delta *K* method.

Stable isotope analysis

The soft tissues of three *T. tokunagai* were freeze-dried for two hours at -60□ and ground into powder. The 4 × 6 tin cup packaging samples were used to weigh approximately 0.1 mg of powder sample and then subjected to the isotope ratio mass spectrometer (IRMS, Thermo Fisher Scientific Inc., Waltham, USA) for stable isotope determination in Third Institute of Oceanography (Xiamen, China). The carbon isotope abundance ratio of international standard VPDB (Vienna Peedee Bellemnite) was used to calculate the $\delta^{13}C$ value, with an analysis accuracy of the $\delta^{13}C$ value of ± 0.2 ‰. Similarly, the nitrogen isotope abundance ratio of nitrogen in the air was used to calculate the $\delta^{15}N$ value with an analysis accuracy of ± 0.25 ‰.

DIC assimilation rate measurement based on radioactive carbon analysis

Radiotracer assays were used to determine the rate of dissolved inorganic carbon (DIC) assimilation by introducing a ^{14}C-labeled DIC tracer to the homogenized gill in a solution of MMJHS medium [48] and quantifying the amount of ^{14}C incorporated into total organic carbon (TOC) [49]. Four replicates of gill tissue samples (three experimental samples and one negative control) were filled into 10 ml serum vials without headspace and sealed with sterile PTFE septa and aluminum caps. After that, 100 μL of ^{14}C-DIC solution (~4×10^4 Becquerel, Bq) was injected into each serum vial through the stopper by displacing the same volume of water. Before injecting the ^{14}C-DIC tracer, the microorganisms of negative controls were removed by adding 0.5 mL of 100% trichloroacetic acid (TCA). All samples were incubated in the dark at 28 □ for 72 h, and the microorganisms of the experimental group were removed with the addition of 0.5 mL TCA and filtered onto 0.2 μm GSWP membranes (polyethersulfone, Millipore) after

195 incubation. The filters were rinsed with 35 ‰ sodium chloride (NaCl) solution [50] and
196 transferred into 7 mL scintillation vials containing a 6 mL scintillation cocktail (Ultima Gold ™
197 Cocktail, PerkinElmer). The radioactivity of the filters was determined using a Tri-Carb 3110TR
198 liquid scintillation counter [51]. The turnover rate constant (k, 1/day) of DIC was calculated
199 using equation (1), and the assimilation rate (Ass-rate, μmol/L/day) of DIC was calculated using
200 equation (2):

201 $$k = -\ln (1-\text{DPM-}^{14}\text{C-POC} / (\text{DPM-}^{14}\text{C-DIC}))/t \quad (1)$$

202 $$\text{Ass-rate} = k \times [\text{DIC}] \quad (2)$$

203 Here, DPM-^{14}C-POC is the radioactivity on the filter, DPM-^{14}C-DIC is the total activity of the
204 added DIC tracer, t is the incubation time (day), and [DIC] is the DIC concentration (μmol/L) in
205 the sample.

206

207 **Fluorescence *in situ* hybridization**

208 Symbiont specific-probe with Cy5-labeled (5'- TCCTCTATCACACTCTAGCTCAGCAGTATC-
209 3'), sense probe with CY3-labeled (5'- GATACTGCTGAGCTAGAGTGTGATAGAGGA-3'),
210 and bacterial universal probe EUB338 with CY5-labeled were designed based on the
211 corresponding representative 16S rRNA genes [52]. Gill tissues of *T. tokunagai* were dissected
212 and further dehydrated in 100% methanol for 30 min each and embedded in paraffin. Then, a
213 section with 7 μm thick was cut using a semiautomatic microtome (Lecia, Germany). After the
214 removal of paraffin in xylene and ethanol, the sections were rehydrated in a decreasing ethanol
215 series (100, 95, 80%, and 70%) for 15 min each, and then hybridization was processed at 46 □
216 with a hybridization buffer (work concentration: 5 μg/mL probe in 0.9 M NaCl, 0.02 M Tris-HCl,
217 0.01% sodium dodecyl sulfate and 30% Formamide) for 1h. These slides were washed in a
218 washing buffer (0.1 M NaCl, 0.02 M Tris-HCl, 0.01 % sodium dodecyl sulfate, and 5 mM EDTA)
219 for 10 min each at 48 □. Then, 4′,6-diamidino-2-phenylindole (DAPI, Solabio) and Alexa Fluor
220 488 Conjugate Concanavalin-A (Invitrogen, CA, USA) were added to each slide and incubated
221 for 5 min at room temperature. After washing using PBST (Tween-20: PBS=1: 1000), the slides
222 were mounted by ProLong™ Diamond Antifade Mountant (Invitrogen). Images were captured

by a confocal laser scanning microscope (ZEISS LSM900, Germany) and pre-treatment with ZEN v.3.1 software.

Transmission electron microscope

Pre-fixed dissected gill tissue was firstly washed with 0.1 M PBS three times for 15 min each time and fixed with 1 % osmic acid after being washed again with 0.1 M PBS three times for 15 min. Then, the tissue was dehydrated increasingly in a methanol series (50 %, 70%, 90 % once, and 100 % three times) for 15 min each and embedded in Epon 812 resin. Afterward, the embedded blocks were cured at different temperatures (37 □, 45 □, and 60 □), lasting for 1 day at each temperature. A Reichert ULTRACUT Ultrathin slicer (Austria) was used to slice the ultrathin slides at 70 nm thick. Slides were stained using the uranium acetate lead citrate double staining method that uranium acetate and lead citrate staining for 15 min each. Images were captured by a JEM 1200-EX (Japan) Transmission Electron Microscope at an accelerating voltage of 80 kV.

Symbiont genome assembly, binning and annotation

Clean reads obtained were assembled using Megahit v.1.2.9 [53] with the setting of "k-list = 21, 29, 39, 59, 79, 99, 119, 141". The metagenome-assembled genomes (MAGs) were obtained using MaxBin v.2.2.7 [54], with the cutoff of contig length from 1000 to 2000 for more optimal bins. To reduce host contamination, we conducted a decontamination process using Blobtools v.1.1.1 [55] with default settings, and then only sequences belonging to the phylum Proteobacteria were selected for downstream analyses. To obtain the circular genome of the symbiont, firstly, the ONT long reads were mapped to the highest quality MAGs; secondly, these long reads mapped were assembled using NextDenovo v.2.5.2 [56], thirdly, clean reads (i.e. Illumina) were input into NextPolish v.1.4.1 [57] to polish the complete genome. Then, these complete genomes were evaluated the completeness and contamination using CheckM v.2.0.1.3 [58]. GTDB-tk v.2.1.1 [59] was used to determine the taxonomy of symbionts at the genome level. The matrix of pairwise average nucleotide identity (ANI) of 30 MAGs was generated using FastANI v.1.34 [60]. 16S rRNA genes and open reading frames (ORFs) of MAGs were predicted

by Prokka v.1.14.6 [61] in the single genome mode. The functional category of predicted genes was retrieved using BlastKOALA [62] to get the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations.

Pangenome generation, strain decomposition, and phylogenomic analysis

Pangenome was generated using PanPhlAn v.3.1 [63] and in-built scripts of StrainPanDA [64]. These predicted genes of pangenome were searched against the NCBI NR database by using BLASTp in DIAMOND v.2.0.15.153 [65] with an E-value cut-off of $1e^{-5}$. The results further were used for Gene Ontology annotation by Blast2GO v.6.0 [66]. Meanwhile, Clusters of Orthologous Group 2020 (COG2020) [67] was adopted to classify the functional groups of genes in the pangenome. The genes of the pangenome were annotated using BlastKOALA [62] by searching against the KEGG database.

To decompose the diversity of symbiont at the strain resolution, StrainPanDA [64] was adopted based on the newly constructed pangenome of 30 MAGs and the full set of clean reads. In contrast, the former software STRONG [68] infers strain composition based on core COGs but it could not handle large-number datasets (over 10M reads in our case) due to its memory-consuming issue. In this work, STRONG was also employed for comparison, with less than 10M clean reads in each sample were used as input and the following parameter: bayespaths: nb_strains: 16, nmf_runs: 10, min_orf_number_to_merge_bins: 18. To verify the correlation of relative abundance of two symbiont strains and inferred from the software above, the 16S rRNA gene of *Ca.* Sedimenticola endothyasira oligotype G (G type) was mapped to 30 metagenomic datasets using Bowtie v.2.3.5 [69], and the coverage of 16S rRNA gene of *Ca.* Sedimenticola endothyasira of each dataset was checked by Integrative Genomics Viewer v.2.4.1 (IGV) [70]. To check whether the diversity of strains was caused by recombination, Rhometa v.1.0.3 [71] was employed to estimate the population recombination rate with the input of clean reads.

In addition, to find the taxonomy of two strains of symbiont in chemosynthetic sulfur-oxidizing bacteria (SOB), a total of 57 symbiont genomes assigned by GTDB-tk v.2.1.1 [59] were downloaded for phylogenomic analysis with *Bathymodiolus azoricus* symbiont (MOB) as an outgroup. A pipeline of phylogenomic analysis was modified from the published works [72, 73].

SNP calling

To understand the symbiotic population structure in *T. tokunagai* in the Yellow Sea, the single nucleotide polymorphisms (SNPs) analysis was adopted to find the nucleotide diversity within (π_{within}) and between host ($\pi_{between}$), with the assumption of each host as a symbiotic population. The whole genome sequencing reads were aligned to the pangenome (as a reference genome) using Bowtie v.2.3.5 [69]. A pipeline of Genome Analysis Toolkit (GATK) v.4.2.6.1 [74] was used to call SNPs in the symbionts with a standard pipeline. In detail, the *.bam* files were formatted using GATK AddOrReplaceReadGroups and GATK SortSam. Then, the duplicated reads were filtered by GATK MarkDuplicates. GATK HaplotypeCaller was used to call SNPs with the setting of "ploidy: 2". Afterward, only qualified SNPs were kept using GATK VariantFiltration with the setting of "QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 4.0 || MQRankSum < –12.5 || ReadPosRankSum < −8.0". The SNPs number of each individual (i.e., 30 *.vcf* files) was counted using GATK CountVariants. All *.vcf* were merged using GATK CombineGVCFs for plotting the PCA of SNPs. To understand the population structure at the SNP level tentatively, plink v.1.9 [75] was used for principal component analysis (PCA). The SNPs of core genes were extracted for downstream analyses. The per-gene nucleotide diversity (π) within single host individuals (π_{within}), between host individuals ($\pi_{between}$), and the fixation index (F_{ST}) were calculated following the codes (https://github.com/deropi/BathyBrooksiSymbionts/tree/master/Population_structure_analyses) from Ansorge et al [37]. PCA of the pairwise per-gene $\pi_{between}$ and π_{within} of core genes supported by PERMANOVA on pairwise Bray-Curtis dissimilarities with 9999 permutations using PAST version 4.03 [76]. In addition, to understand the influence of geographic distance and environmental factors (i.e. depth, temperature, and dissolved oxygen) on the symbiont population formation, the correlation between the mean F_{ST} of the symbiont population within each sampling site, and geographic distance and environmental factors, meanwhile, we performed the redundancy analysis (RDA) analysis in vegan R package based on the symbiont allele frequencies [AD divided by DP], with environmental factors, geographic distance and oligotype followed the pipeline [36].

Differences between two symbiont oligotype in genome-level and gene expression

To find the difference between two oligotypes (i.e. G and A type), based on the pangenome presence-absence result, the specific genes only can be found in the oligotype A or G were determined at the gene level. Meanwhile, the difference also can be checked in the gene expression of host and symbiont with the following steps: clean reads were mapped to pangenome using Bowtie v.2.3.5 [69] and then were quantified using Samtools v.1.9 [77], resulting in the symbiont-derived reads and symbiont-free reads. All the eukaryotic reads were subjected to Trinity v.2.13.2 [78] for *de novo* assembling a transcriptional profile in the host. The section in transcripts with coding potential was predicted using TransDecoder v.5.5.0 (https://github.com/TransDecoder/TransDecoder). CD-HIT v.4.8.1 [79] was used to remove the redundant sequence with the setting of "c = 0.8". The quality of transcripts was checked using BUSCO v.5.4.4 [80]. Salmon v.1.9.0 [81] was performed to quantify the gene expression levels, and the normalized gene expression matrix (TMM) obtained was further utilized in the PCA analysis normalized by log-transformed as well as further added one to uncover the gene expression pattern of each individual. In addition, the longest CDS sequences were annotated by searching against the NR, KEGG, and GO databases above. For symbiont gene expression, the pangenome was used as a reference to quantify the gene expression level with the rest analyses the same as the part in the host.

Results and Discussion

The undifferentiated population of *Thyasira tokunagai* and its chemosymbiotic capacity

Compared to other benthic fauna and environmental samples from the Yellow Sea (Table S4), *Thyasira tokunagai* exhibited the lowest $\delta^{15}N$ value (-0.23 ± 0.22 ‰, n = 3), suggesting a potential autotrophic supplement on its nitrogen source (Fig. 1b and Table S4). The $\delta^{13}C$ value of *T. tokunagai* (-20.52 ± 0.43 ‰, n = 3) was higher than that of particulate organic matter and phytoplankton, but lower than other benthic fauna. These results indicate that *T. tokunagai* relies heavily on the chemosymbiont for nutrition. The assimilation rates of inorganic carbon in *T. tokunagai* were quantified using radioactive ^{14}C-labeled bicarbonate (Table S5), which were 0.32, 0.29, and 0.55 μmol DIC per day per individual (average: 0.39 μmol DIC/day). According to the

species' average density (~75 samples/m^2) based sampling from our three research cruises, the carbon fixation flux in the *T. tokunagai* population was estimated to be 0.35 milligram (mg) Carbon every square meter per day in the Yellow Sea. Considering that *T. tokunagai* is a dominant species in the Yellow Sea and Japan Sea [82-85], and its close relative *Thyasira* cf. *gouldii* is a pan-arctic species widely discovered in the Atlantic Ocean [86], the role of DIC fixation by thyasirids in the *T. gouldii* complex cannot be neglected (Supplementary Note 1), and chemosynthesis potentially contributes to ocean carbon budgets [87].

STRUCTURE assigned the individuals into two lineages using the alignment of 13 protein-coding genes (PCGs) of 30 mitogenomes based on the delta K (K = 2; Fig. 1c), suggesting the individuals in the Yellow Sea originated from the same ancestor. Meanwhile, the pairwise comparisons of the mitochondrial *cox1* gene revealed an average of 99.84 % similarity (Table S6 and Fig. S1b) and the haplotype network (Fig. S1c and Table S7) was of a 'star-burst' type, although these specimens collected from nine distinct sampling sites. The haplotype network showed that nearly all haplotypes lack a clear geographical affinity, indicating that there is no significant differentiation among the nine sampling populations, showing frequent gene exchange and a panmixia for *T. tokunagai* in the Yellow Sea.

A species (*Candidatus* Sedimenticola endothyasira) in the genus *Sedimenticola* is the dominant clade among the bacterial community in the gills of *T. tokunagai* based on the full-length 16S rRNA gene amplicon sequencing, accounting for over 80% except for two individuals (i.e., 70.6% in NYS5_3 and 67.7% in NYS7_4). This *Sedimenticola* symbiont could not be found in the ambient surface sediment (Fig. 2a and Fig. S2), Notably, there was only a single base pair difference (G vs A) at the 563rd position of the top two ASVs in this clade with 46.28 % and 44.55 % (average percent)**,** respectively (Fig. 2b), which was further verified by 16S rDNA clone library and Sanger sequencing (Supplementary Note 2). The phylogenetic tree of the 16S rRNA gene of symbiont confirmed that its closest relative was a symbiont of *Thyasira* cf. *goudlii* (Fig. S3). Combining ONT and Illumina sequencing, a total of 30 high-quality circular genomes (i.e. MAGs) were assembled, with completeness > 99.23%, contamination rate < 0.34%, and size of 4.5Mb (Table S9). The average nucleotide identity (ANI) of these 30 MAGs ranged from 98.90

to 99.94% (Fig. 4c and Table S10). Based on the 563rd base pair of the 16S rRNA gene, they could be categorized into two groups, with 12 in *Ca.* Sedimenticola endothyasira oligotype G (with G in the 563rd position) and the rest 18 for *Ca.* Sedimenticola endothyasira oligotype A (with A in the same position). Results from phylogenomic analysis based on 1745 single-copy core genes of these 30 MAGs showed the two distinct phylogenetic clusters corresponding well to the two oligotype defined by 16S amplicon sequencing (Fig. S5). We also found that those within same oligotype group had a higher ANI identity (average of between two oligotypes group: 99.16%; average within oligotype G group: 99.77%; average within oligotype A group: 99.58%). Collectively, we show the dominant *Sedimenticola* symbiont encapsulating two distinct strains but with very high genomic similarity. The second clade *Spirochaeta*_2 account for 3.69 ± 4.36%. The co-existence of *Sedimenticola* and *Spirochaeta_2* was also reported in a lucinid clam [88].

The presence of the symbiont was examined by FISH analysis, showing the *Sedimenticola* symboint was concentrated in the bacteriocytes located at the middle part of the gill filament apart from the ciliated filament tip (Fig. 2c), and symbiont was enveloped by cell membranes. No signal was detected based on negative control (Fig. S6). In addition, TEM observation (Fig. 2d) of gill tissue showed that symbionts were localized in extracellular pouch-like structures among the microvilli, together with FISH results, implying an episymbiotic mode with the gill in *T. tokunagai* where bacteria are maintained outside of the host cytoplasm [89].

Two strains without major metabolic differences co-existing in a single host individual

Regarding symbiotic diversity at the strain level, *T. tokunagai*, there were two representative strains classified including *Ca.* Sedimenticola endothyasira oligotype G and *Ca.* Sedimenticola endothyasira oligotype A and their abundance was quantified in StrainPanDA (Fig. 4a). Notably, there was a positive correlation between the percentage of G type from StrainPanDA and the percentage of metagenomic reads of G base pair at the 563rd position in 16S rRNA gene (R^2 = 0.969, $p < 0.001$; Fig. S7c). Congruent result was also observed in STRONG, another software for strain decomposition but with limitations in data input and the identification of representative

three strains (Fig. S7a), these results showed that a positive correlation between STRONG and StrainPanDA ($R^2 = 0.991$, $p < 0.001$; Fig. S7b). The placement of these two representative strains in genus *Sedimenticola* were also verified by phylogenetic reconstruction at the genomic level (Fig. 4b and Fig. S4).

The metabolic potential in the symbiont of *T. tokunagai* was highly conserved (Fig. 3b). It encodes the full enzymes in carbon fixation and utilization, including the reductive pentose phosphate cycle (Calvin cycle), glycolysis/gluconeogenesis, tricarboxylic acid cycle (TCA), and oxidative phosphorylation, enabling the symbiont assimilating DIC in the former part. In the Calvin cycle, the ribulose-bisphosphate carboxylase large chain (K01601) was found in both oligotypes. The complete dissimilatory nitrate reduction pathway enables the symbiont to proceed with respiration under an anaerobic or hypoxic environment but hinders it from producing ammonia due to the lack of the nitrite reductase (NADH) large subunit (K00362). A complete dissimilatory sulfate reduction pathway and a partial SOX system were also found, mainly containing *soxA*, *B*, *X*, *Y*, and, *Z*, but *soxCD* was lacking. The incomplete assimilatory sulfate reduction pathway was detected, which contained *sat* and *cysC*. We also found the genes and enzymes related to hydrogen oxidation containing *hoxF*, *U*, *Y*, *H*, *hybC*, and *hyaB*. The symbiont encodes ABC transporters and PTS pathways, indicating the capacity of heterotrophy. The bacterial chemotaxis and flagellar assembly pathways were found. Additionally, the symbiont has a relatively complete capacity for the biosynthesis capacity of amino acids (20), and vitamins and cofactors (6) (Table S11), suggesting a capacity for chemosymbiosis.

Based on the previous results from phylogenetic and SNP analyses, we looked into the potential differences in the two oligotypes of the *T. tokunagai* symbiont (i.e., oligotype A and oligotype G). At the coding-gene level, we compared the two oligotypes groups based on the pangenome information, showing that there were 11 specific genes identified for oligotype A group and a different 11 for the oligotype G group (Table S12). Among those, the oligotype A specific genes were related to chemotaxis protein, ammonium transporter, and alpha/beta fold hydrolase, while those specifically belonging to oligotype G were related to cytochrome *c*, phosphodiesterase and ArsJ-associated glyceraldehyde-3-phosphate dehydrogenase. As for gene expression, we

quantified and annotated the host and symbiont transcripts (Table S13 and Table S14), indicating there was no congruent pattern with the different oligotypes shown in PCA, no matter at the host (Fig. 6a) or symbiont levels (Fig. 6b). Moreover, no key functional difference in the top 50 of the highest expressed chemosymbiosis-related genes could be found, in both the host (Fig. 6c) and symbiont (Fig. 6d). Taken together, the differences between these two major oligotypes might be limited to gene sequence variants instead of gene functions.

Horizontal acquisition of symbionts with strong selection

Our results revealed distinct bacterial communities in the gills of *T. tokunagai* compared to their surrounding sediments, and the bacterial communities in sediment from the Yellow Sea were roughly consistent with a previous study [90]. Though the lack of symbionts detected in the sediments may superficially appear to suggest a vertical transmission mode where the symbiont is passed down from parents to the offspring, it is not uncommon for horizontally transferred symbionts to be undetected from environmental samples [91-93]. Previous studies reported that thyasirid clams use their super extensile foot to ‘mine’ sulfide; and the magnetosome in its symbionts [94, 95] imply that they might be derived from the specific niche of sediments, which may not be fully covered in the present study. The metabolic potential of the symbiont genomes is suggestive of horizontal transmission. Vertical transmission typically leads to the loss of essential genes (e.g., flagellar and chemotaxis genes) in the symbiont genome required for a free-living lifestyle, known as genomic reduction [25]. The genomic sizes of the *T. tokunagai* symbiont MAGs were 4.5Mb (Fig. 3a), much larger than that typical of vertically transmitted symbionts which mostly exhibit genome sizes less than 2Mb [2, 14, 96]. Similar to previous metagenomic research of *T.* cf. *gouldii* [97], a set of genes encoding flagella and chemotaxis were identified in both oligotypes of the *T. tokunagai* symbiont, indicating that they have the potential to survive outside their host. The presence of the TCA cycle in their genomes also suggests that they can use extraneous carbon sources from the environment directly during a non-symbiotic or free-living stage. It was also evidenced by the inconsistency of phylogeny between host mitochondria and the corresponding symbionts (Fig. S5). In addition, the sampling site of *T. tokunagai* covered most of the extent of the Yellow Sea, with the distance from north to south extending over 500 km. Besides, there were only two strains observed in 30 individuals.

Therefore, we interpret that *T. tokunagai* most likely acquired the *Sedimenticola* symbionts from the environment via horizontal transmission, controlled by a highly selective mechanism between the host and symbionts.

Heterogeneous symbiotic populations among host individuals

The quality of pangenome is strongly correlated with the assembly level of MAGs. The use of fragmented MAGs might result in some incorrect comparisons among strains. Here, a comprehensive pangenome of *T. tokunagai* symbionts was constructed from 30 circular MAGs of each host, which consisted of 3256 core genes (79% in 4106 on average), 2008 accessory genes, and 371 specific genes (Fig. S8a; Table S15 and Table S16). Comparatively, previous studies retrieved lower portions of genes as core genes, such as 40-53% in bathymodioline mussels and 62.6%–68.8% in the peltospirid snail *Chrysomallon squamiferum* [9, 37]. Most of the elements participating in key processes of chemosymbiosis and fundamental metabolism were detected in the core genes, such as the TCA cycle, the Calvin cycle for carbon fixation, glycolysis, sulfide-oxidization, biosynthesis of amino acids and cofactors, sugar transfer, chemotaxis, and flagellar assembly (Fig.3b). The accessory genes and specific genes were annotated by the search against the COG database (Fig. S8b and Table S14), of which several COG categories were mainly involved: signal transduction mechanisms (T); energy production and conversion (C); cell wall/membrane/envelope biogenesis (M); translation, ribosomal structure and biogenesis (J); replication, recombination, and repair (L); amino acid transport and metabolism (E).

Furthermore, with the 3256 core genes as the reference sequences, we deduced the structure of symbiont populations at the SNP level. In total, 101,688~199,595 SNPs were detected from 30 populations with the setting of “ploidy = 2” (Fig. 4a and Table S17). The shared SNPs of oligotype A (SNP density: 26.62~39.75 SNPs/Kb) dominant groups account for 0.032% of total SNPs in the third position (Fig. S9a), while the shared SNPs of oligotype G dominant groups (SNP density: 22.28~43.72 SNPs/Kb) account for 0.025% in the first position (Fig. S9b). PCA with a PERMANOVA test based on pairwise Bray–Curtis dissimilarities based on the total SNPs,

showed the symbiotic community in *T. tokunagai* was shaped following their dominant oligotype rather than sampling sites (PERMANOVA test, F = 782.3, p = 0.22, Fig. 5a). The pairwise host individuals belonging to the same oligotype groups have a relatively low F_{ST} value and a relatively high ANI value with the oligotype group clustered closely (Fig. 4c), further confirming that the symbiont communities were divided into two groups (i.e. oligotype G and oligotype A) at the SNP level, which was consistent with the aforementioned 16S rRNA gene sequencing and phylogenomic tree results. The same oligotype group has similar levels of intra-host π (average oligotype A: 1.23; average oligotype G: 1.51) and F_{ST} within the same oligotype individuals (average oligotype A: 0.0023, average oligotype G: 0.0021; Table S18). Meanwhile, within the same oligotype (i.e. either within oligotype A symbionts or within oligotype G symbionts), the inter-host π is larger than the intra-host π (PERMANOVA test, F = 83.47, $p < 0.0001$, Fig. 5e), suggesting there is more difference between individuals compared to within a single individual. In addition, the population recombination rate of thirty metagenome datasets was estimated, with the undetectable event of population recombination for distinct symbiont strains within single individuals, which further suggested that the highly diverse community at the strain level is likely to be ascribed to the accumulation of mutation.

Priority effect on the communities of episymbiont in thyasirid clam

The population structure of horizontally transmitted symbionts could be shaped by a series of factors, including host genetics, environmental factors, and geographic factors [98]. Based on the symbiont population variant data, we revealed almost no correlation between the mean fixation index F_{ST} of pairwise symbiont population from each sampling site and environmental factors (i.e. water depth: r_s = -0.200; p = 0.61; temperature: r_s = 0.467; p = 0.21; dissolved oxygen: r_s = -0.550; p = 0.13; Fig. 5c). Similarly, the correlation between the mean F_{ST} of pairwise symbiont population from each sampling site and geographic subdivision (r_s = -0.085; p = 0.83; Fig. 5c) was performed. The redundancy analysis (RDA) showed that the variable oligotype contributed the dominant variation (42.5%) in the differentiated population, and more than half of it (23.1%) was regarded as the sole contributor (Fig. 5d). By contrast, only 3% and 0.1% variation were ascribed to the geographic distance and environmental factors as the sole factor. Overall, the analyses between the symbiont population versus the host genetics (as represented by

mitogenome genotyping), environmental factors (depth, temperature, and DO), and geographic factors (distance) have shown a lack of correlation between them and symbiont population structure, indicating some other factors could shape the symbiont population structure.

Here, we suggest that the population structure of two highly similar symbionts in *T. tokunagai* is likely controlled by the earliest type of symbionts colonized in the gill (e.g., oligotype A or oligotype G), reported in the literature as the "priority effect" [38, 99]. Our result showed that *T. tokunagai* tended to harbor either one of two oligotype (G or A) instead of a mixed mode between the two (Fig. 2a and Fig. 4a). Nonetheless, the average nucleotide identity between two oligotype' genomes was 98.71%, above the proposed threshold of inter-species variation of prokaryotes (95%) [60]. They shared most metabolic pathways, which indicated they were almost indistinguishable from each other. In contrast to the symbiont population of the peltospirid snail *Chrysomallon squamiferum* [9] and *Bathymodiolus* mussels [37], the F_{ST} and intra-host both revealed the same oligotype have a lower difference at the SNPs level (average of intra-host π:1.44; average of intra-host π:2.50; average of F_{ST} within oligotype: 0.0022; average of F_{ST} between two oligotypes: 0.0035), indicating the presence of two strains at the population level. Given that the environmental pool of symbionts (two oligotypes: A and G) is a homogeneous [100] and all the microbes share a similar chance to be acquired by newly settling *T. tokunagai* host, there should be a maximum frequency around the equivalent value (e.g., 50% in terms of oligotype G or A). Surprisingly, we detected the potential anti-normal distribution in the ratio of A:G, with only 2 in 30 samples (Fig. 5b), suggesting the restricted process in the first colonization, as reported in the mussel *Bathymodiolus brooksi* [38]. It is possible that the thyasirid only acquires symbionts in a very short span after settling to the seafloor, though this remains unclear. As mentioned before, there was a high density of *T. tokunagai* in the Yellow Sea within the aggregated reducing habitats. Symbionts would be released from dead individuals, which could provide a source of symbionts for acquisition. We also cannot reject the possibility of acquiring symbionts contained within water currents. Therefore, assuming horizontal transmission (discussed above), we propose the plausible scenario where: 1) the dominant strain of *Sedimenticola* in *T. tokunagai* is determined during the initial symbiont acquisition stage just after larval settlement and early gill development; 2) the propagation of the initial type of symbiont as the *T. tokunagai* develops into adulthood; 3) the continuous but restricted recruitment of the other type of symbionts during the adult stage.

Conclusion

In this study, we characterized chemosymbiosis in the thyasirid clam *Thyasira tokunagai* sampled from the Yellow Sea. We show that the symbiont was likely mixotrophic, localised in a pouch-like structure delimited by microvilli on the gill, capable of carbon fixation verified by radioactive isotope tracing, and horizontally transmitted from the environment. With just two functionally equivalent strains and just one dominate in each individual, the thyasirid becomes a perfect model to study the factors determining the episymbiotic communities in chemosynthetic holobionts. We employed population genetics to compare the symbiotic diversity among 30 individuals, where host inherited genes, geographic distance, and environmental factors could not reasonably explain the differences in the composition of the two strains among individuals. We deduced that the formation of the symbiont population in the thyasirid clam is possibly governed by the "priority effect." These findings are meaningful in understanding the determination of microbial community assembly in other holobionts and how the association between host and microorganism forms in chemosymbiosis.

Acknowledgments

This work was financially supported by Science and Technology Innovation Project of Laoshan Laboratory (LSKJ202203104), the Fundamental Research Funds for the Central Universities (202172002 and 202241002), and the Young Taishan Scholars Program of Shandong Province (tsqn202103036). We thank Prof. Suzanne C. Dufour from Memorial University of Newfoundland and Dr. Yi Lan and Dr. Hao Wang from Hong Kong University of Science and Technology for their constructive comments on the early version of the manuscript.

Author contribution

JS conceived the project. ML collected the samples, extracted the nucleic acid, and performed most of the experiments and bioinformatic analysis. YL performed the strain decomposition and SNP analysis. ML and YL drafted the original manuscript. XL contributed to the phylogenomic

analysis. HW contributed to the 16S rDNA sequence analysis. SM estimated the DIC assimilation rate of the symbiont. CC identified the samples. YL, XL, MS, CC, YL, XL, GZ, WZ, and JS contributed to the writing and editing of the manuscript.

Data availability

The data of this study were deposited into the NCBI database under the BioProject ID PRJNA995037 and PRJNA1016492. And the commands of this study used were deposited on GitHub (https://github.com/Menggong-li/Chemosymbiosis-in-the-thyasirid-from-Yellow-Sea).

References

1. Bascompte J. Mutualism and biodiversity. Curr Biol. 2019;29:R467-R70.
2. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. The evolution of host-symbiont dependence. Nat Commun. 2017;8:15973.
3. Bennett GM, Moran NA. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci USA. 2015;112:10169-76.
4. Cornwallis CK, van 't Padje A, Ellers J, Klein M, Jackson R, Kiers ET, et al. Symbioses shape feeding niches and diversification across insects. Nat Ecol Evol. 2023;7:1022-44.
5. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Micro. 2008;6:725-40.
6. Hinzke T, Kleiner M, Breusing C, Felbeck H, Häsler R, Sievert Stefan M, et al. Host-microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis. mBio. 2019;10:e02243-19.
7. Yang Y, Sun J, Sun Y, Kwan YH, Wong WC, Zhang Y, et al. Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. ISME J. 2020;14:135-50.
8. Sun Y, Sun J, Yang Y, Lan Y, Ip JC-H, Wong WC, et al. Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*. Mol Biol Evol. 2021;38:4116-34.
9. Lan Y, Sun J, Chen C, Wang H, Xiao Y, Perez M, et al. Endosymbiont population genomics sheds light on transmission mode, partner specificity, and stability of the scaly-foot snail holobiont. ISME J. 2022;16:1-12.
10. Breusing C, Mitchell J, Delaney J, Sylva SP, Seewald JS, Girguis PR, et al. Physiological dynamics of chemosynthetic symbionts in hydrothermal vent snails. ISME J. 2020;14:2568-79.
11. Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. Nat Commun. 2021;12:1165.
12. Franke M, Geier B, Hammel JU, Dubilier N, Leisch N. Coming together—symbiont acquisition and early development in deep-sea bathymodioline mussels. Proc R Soc B. 2021;288:20211044.
13. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nat Ecol Evol. 2017;1:0121.
14. Ip JC-H, Xu T, Sun J, Li R, Chen C, Lan Y, et al. Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. Mol Biol Evol. 2020;38:502-18.

15. Li Y, He X, Lin Y, Li Y-X, Kamenev GM, Li J, et al. Reduced chemosymbiont genome in the methane seep thyasirid and the cooperated metabolisms in the holobiont under anaerobic sediment. Mol Ecol Resour. 2023;23:1853-67.
16. Sogin EM, Kleiner M, Borowski C, Gruber-Vodicka HR, Dubilier N. Life in the dark: phylogenetic and physiological diversity of chemosynthetic symbioses. Annu Rev Microbiol. 2021;75:695-718.
17. Suzuki Y, Kojima S, Sasaki T, Suzuki M, Utsumi T, Watanabe H, et al. Host-symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the southwest pacific. Appl Environ Microbiol. 2006;72:1388-93.
18. Johnson SB, Krylova EM, Audzijonyte A, Sahling H, Vrijenhoek RC. Phylogeny and origins of chemosynthetic vesicomyid clams. Syst Biodivers. 2017;15:346-60.
19. Ozawa G, Shimamura S, Takaki Y, Takishita K, Ikuta T, Barry JP, et al. Ancient occasional host switching of maternally transmitted bacterial symbionts of chemosynthetic vesicomyid clams. Genome Biol Evol. 2017;9:2226-36.
20. Osvatic JT, Wilkins LGE, Leibrecht L, Leray M, Zauner S, Polzin J, et al. Global biogeography of chemosynthetic symbionts reveals both localized and globally distributed symbiont groups. Proc Natl Acad Sci USA. 2021;118:e2104378118.
21. 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.
22. Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, et al. A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. Environ Microbiol. 2006;8:1441-47.
23. Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, et al. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J. 2017;11:463-77.
24. Vrijenhoek RC. Genetics and evolution of deep-sea chemosynthetic bacteria and their invertebrate hosts. The Vent and Seep Biota: Aspects from Microbes to Ecosystems, 2010:15-49.
25. Newton I, Woyke T, Auchtung T, Dilly G, Dutton R, Fisher M, et al. The *Calyptogena magnifica* chemoautotrophic symbiont genome. Science. 2007;315:998-1000.
26. Hilario A, Capa M, Dahlgren TG, Halanych KM, Little CT, Thornhill DJ, et al. New perspectives on the ecology and evolution of siboglinid tubeworms. PLoS ONE. 2011;6:e16309.
27. Perez M, Breusing C, Angers B, Beinart RA, Won Y-J, Young CR. Divergent paths in the evolutionary history of maternally transmitted clam symbionts. Proc R Soc B. 2022;289:20212137.
28. Igawa-Ueda K, Ikuta T, Tame A, Yamaguchi K, Shigenobu S, Hongo Y, et al. Symbiont transmission onto the cell surface of early oocytes in the deep-sea clam *Phreagena okutanii*. Zool Sci. 2021;38:140-47.
29. Ikuta T, Igawa K, Tame A, Kuroiwa T, Kuroiwa H, Aoki Y, et al. Surfing the vegetal pole in a small population: extracellular vertical transmission of an 'intracellular' deep-sea clam symbiont. R Soc Open Sci. 2016;3:160130.
30. Laming SR, Gaudron SM, Duperron S. Lifecycle ecology of deep-sea chemosymbiotic mussels: a review. Front Mar Sci. 2018;5:282.
31. Pennec ML, Beninger PG. Ultrastructural characteristics of spermatogenesis in three species of deep-sea hydrothermal vent mytilids. Can J Zool. 1997;75:308-16.
32. Zauner S, Vogel M, Polzin J, Yuen B, Mußmann M, El-Hacen E-HM, et al. Microbial communities in developmental stages of lucinid bivalves. ISME Commun. 2022;2:56.
33. Breusing C, Genetti M, Russell S, Corbett-Detig R, Beinart R. Horizontal transmission enables flexible associations with locally adapted symbiont strains in deep-sea hydrothermal vent symbioses. Proc Natl Acad Sci USA. 2022;119.
34. Russell SL, Corbett-Detig RB, Cavanaugh CM. Mixed transmission modes and dynamic genome evolution in an obligate animal-bacterial symbiosis. ISME J. 2017;11:1359-71.

35. Russell SL, Pepper-Tunick E, Svedberg J, Byrne A, Ruelas Castillo J, Vollmers C, et al. Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. PLOS Genet. 2020;16:e1008935.
36. Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. ISME J. 2021;15:3076-83.
37. Ansorge R, Romano S, Sayavedra L, Porras MÁG, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019;4:2487-97.
38. Romero Picazo D, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. ISME J. 2019;13:2954-68.
39. Guri M, Durand L, Cueff-Gauchard V, Zbinden M, Crassous P, Shillito B, et al. Acquisition of epibiotic bacteria along the life cycle of the hydrothermal *shrimp Rimicaris*. ISME J. 2012;6:597-609.
40. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3:294-9.
41. Bahobail A, GadEl-Rab, S.M.F. and Amin, G.A. Locally isolated bacterial strains with multiple degradation potential capabilities on petroleum hydrocarbon pollutants. Adv Microbiol. 2016;6:852-66.
42. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20.
43. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2016;45:e18-e18.
44. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69:313-19.
45. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res. 2002;30:3059-66.
46. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003;164:1567-87.
47. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359-61.
48. Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, et al. Isolation and phylogenetic diversity of members of previously uncultivated ε-Proteobacteria in deep-sea hydrothermal fields. FEMS Microbiol Lett. 2003;218:167-74.
49. Zhuang G-C, Montgomery A, Joye SB. Heterotrophic metabolism of C1 and C2 low molecular weight compounds in northern Gulf of Mexico sediments: controlling factors and implications for organic carbon degradation. Geochim Cosmochim Acta. 2019;247:243-60.
50. Zhuang GC, Peña-Montenegro TD, Montgomery A, Montoya JP, Joye SB. Significance of acetate as a microbial carbon and energy source in the water column of Gulf of Mexico: implications for marine carbon cycling. Global Biogeochem Cycles. 2019;33:223-35.
51. Mao S-H, Zhang H-H, Zhuang G-C, Li X-J, Liu Q, Zhou Z, et al. Aerobic oxidation of methane significantly reduces global diffusive methane emissions from shallow marine waters. Nat Commun. 2022;13:7309.
52. Daims H, Brühl A, Amann R, Schleifer K-H, Wagner M. The domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. Syst Appl Microbiol. 1999;22:434-44.
53. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674-76.

54. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2015;32:605-07.
55. Laetsch D, Blaxter M. BlobTools: Interrogation of genome assemblies. F1000Research. 2017;6.
56. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. An efficient error correction and accurate assembly tool for noisy long reads. bioRxiv. 2023:531669.
57. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2019;36:2253-55.
58. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods. 2023;20:1203-12.
59. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38:5315-16.
60. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.
61. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-69.
62. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726-31.
63. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife. 2021;10:e65088.
64. Hu H, Tan Y, Li C, Chen J, Kou Y, Xu ZZ, et al. StrainPanDA: Linked reconstruction of strain composition and gene content profiles via pangenome-based decomposition of metagenomic data. iMeta. 2022;1:e41.
65. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59-60.
66. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674-76.
67. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2020;49:D274-D81.
68. Quince C, Nurk S, Raguideau S, James R, Soyer OS, Summers JK, et al. STRONG: metagenomics strain resolution on assembly graphs. Genome Biol. 2021;22:1-34.
69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357-59.
70. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings Bioinf. 2013;14:178-92.
71. Krishnan S, DeMaere MZ, Beck D, Ostrowski M, Seymour JR, Darling AE. Rhometa: population recombination rate estimation from metagenomic read datasets. PLOS Genet. 2023;19:e1010683.
72. Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, et al. Phylogenomics of Lophotrochozoa with consideration of systematic error. Syst Biol. 2017;66:256-82.
73. Liu X, Sigwart JD, Sun J. Phylogenomic analyses shed light on the relationships of chiton superfamilies and shell-eye evolution. Marine Life Science & Technology. 2023;5:525-37.
74. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20:1297-303.
75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.
76. Hammer Ø, Harper DA, Ryan PD. PAST: paleontological statistics software package for education and data analysis. Palaeontol Electronica. 2001;4:9.
77. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078-79.

78. 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.
79. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658-59.
80. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210-12.
81. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417-19.
82. Kim Y-R, Lee S, Kim J, Kim C-J, Choi K-Y, Chung C-S. *Thyasira tokunagai* as an ecological indicator for the quality of sediment and benthic communities in the East Sea-Byeong, Korea. Mar Pollut Bull. 2018;135:873-79.
83. Xu J, Lu X, Liu X. Patterns of species and functional diversity of macrofaunal assemblages and the bioassessment of benthic ecological quality status in the southern Yellow Sea. Mar Pollut Bull. 2021;171:112784.
84. Xu Y, Ma L, Sui J, Li X, Wang H, Zhang B. Potential effects of climate change on the habitat suitability of macrobenthos in the Yellow Sea and East China Sea. Mar Pollut Bull. 2022;174:113238.
85. Zhang J, Xiao N, Zhang S, Xu F, Zhang S. A comparative study on the macrobenthic community over a half century in the Yellow Sea, China. J Oceanogr. 2016;72:189-205.
86. Batstone RT, Laurich JR, Salvo F, Dufour SC. Divergent chemosymbiosis-related characters in *Thyasira* cf. *gouldi* (Bivalvia: Thyasiridae). PLoS ONE. 2014;9:e92856.
87. Ricci F, Greening C. Chemosynthesis: a neglected foundation of marine ecology and biogeochemistry. Trends Microbiol. 2024.
88. Lim SJ, Davis BG, Gill DE, Walton J, Nachman E, Engel AS, et al. Taxonomic and functional heterogeneity of the gill microbiome in a symbiotic coastal mangrove lucinid species. ISME J. 2019;13:902-20.
89. Dufour SC. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. Biol Bull. 2005;208:200-12.
90. Liu J, Liu X, Wang M, Qiao Y, Zheng Y, Zhang XH. Bacterial and archaeal communities in sediments of the north Chinese marginal seas. Microb Ecol. 2015;70:105-17.
91. Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. Nat Rev Microbiol. 2010;8:218-30.
92. Brissac T, Gros O, Merçot H. Lack of endosymbiont release by two Lucinidae (Bivalvia) of the genus *Codakia*: consequences for symbiotic relationships. FEMS Microbiol Ecol. 2009;67:261-67.
93. Martin BC, Middleton JA, Fraser MW, Marshall IP, Scholz VV, Hausl B, et al. Cutting out the middle clam: lucinid endosymbiotic bacteria are also associated with seagrass roots worldwide. ISME J. 2020;14:2901-05.
94. Dufour SC, Felbeck H. Sulphide mining by the superextensile foot of symbiotic thyasirid bivalves. Nature. 2003;426:65-67.
95. Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliott A, Poduska KM. Magnetosome-containing bacteria living as symbionts of bivalves. ISME J. 2014;8:2453-62.
96. Jäckle O, Seah BKB, Tietjen M, Leisch N, Liebeke M, Kleiner M, et al. Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*. Proc Natl Acad Sci USA. 2019;116:8505-14.
97. McCuaig B, Peña-Castillo L, Dufour SC. Metagenomic analysis suggests broad metabolic potential in extracellular symbionts of the bivalve *Thyasira* cf. *gouldi*. Anim Microbiome. 2020;2:7.
98. Breusing C, Xiao Y, Russell S, Corbett-Detig R, Li S, Sun J, et al. Ecological differences among hydrothermal vent symbioses may drive contrasting patterns of symbiont population differentiation. mSystems. 2023;8.
99. Debray R, Herbert RA, Jaffe AL, Crits-Christoph A, Power ME, Koskella B. Priority effects in microbiome assembly. Nat Rev Microbiol. 2022;20:109-21.

816 100. Breusing C, Genetti M, Russell SL, Corbett-Detig RB, Beinart RA. Horizontal transmission enables
817 flexible associations with locally adapted symbiont strains in deep-sea hydrothermal vent symbioses.
818 Proc Natl Acad Sci USA. 2022;119:e2115608119.

819

Figure legends

Fig. 1 Chemosynthetic capacity and population structure analysis of *Thyasira tokunagai* in the Yellow Sea. (a) Shell and external anatomy of *Thyasira tokunagai* from the Yellow Sea. (g, gill; f, foot; d, digestive diverticula; m, mantle; aa, anterior adductor; pa: posterior adductor), and a total of twelve sampling sites in the Yellow Sea during three research cruises in August 2020, April 2021, and October 2021. (NYS: North Yellow Sea; SYS: South Yellow Sea) (b) The stable isotopic niche of *Thyasira tokunagai* in the macrobenthic community of the Yellow Sea. The red solid circle represents *Thyasira tokunagai*. (c) Host population structure analysis based on the 13 protein coding genes of the mitogenome of 30 individuals from nine sampling sites in the Yellow Sea.

Fig. 2 Bacterial community composition and symbiont distribution in the gill of *Thyasira tokunagai* revealed by 16S rRNA gene, fluorescence *in situ* hybridization, and transmission electron microscope. (a) Genus-level relative abundance based on the full-length 16S rRNA gene of the bacterial community in the gill of *Thyasira tokunagai*. All ASVs other than showed ASV in gill and sediment were merged under 'Others'. (b) Two ASVs belonging to *Sedimenticola* genus and their (i.e. *Ca.* Sedimenticola endothyasira oligotype G and *Ca.* Sedimenticola endothyasira oligotype A) minor differences are one base at the 563rd position. (c) Fluorescence *in situ* hybridization of *Thyasira tokunagai* showing the dominant bacteria in the gill tissue. All cell nuclei were stained with DAPI, the cell membrane was stained by concanavalin-A, and symbionts were hybridized by the specific probe of *Ca.* Sedimenticola endothyasira. (d) The morphology and position of the symbiont were observed by transmission electron microscope. (MV: microvilli; S: symbiont).

Fig. 3 Circle genomes and metabolic capacity of symbiont. (a) A circular genome was reconstructed by the combination of ONT and Illumina reads. The genome size was 4.5M with a GC content of 52.16%. (b) Metabolic pathway related to chemosymbiosis of symbiont was constructed based on the pangenome information. Here, *Ca.* Sedimenticola endothyasira is highly conserved with a free-living status, which encodes the full enzymes in carbon fixation and utilization and has the potential to utilize sulfides and hydrogen as energy sources.

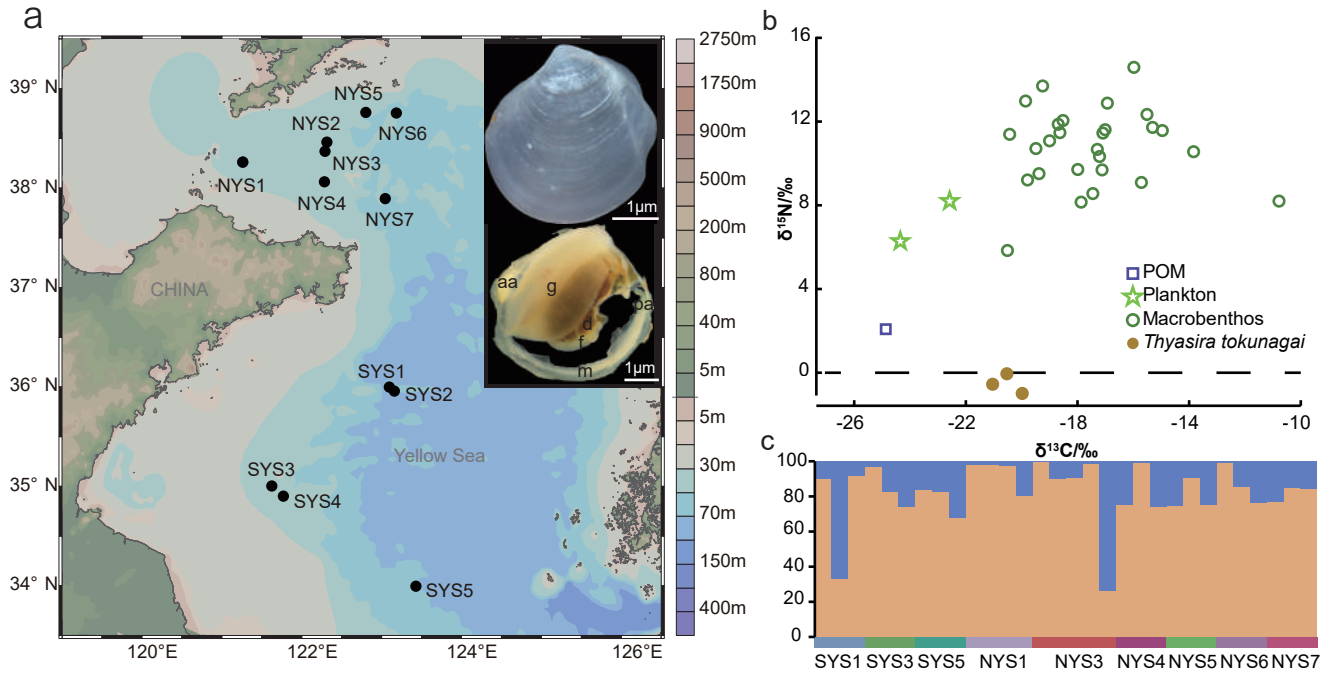
Fig. 4 Two representative strains of symbiont were confirmed based on the strain decomposition, phylogenomic analysis, and pairwise ANI and F_{ST}. (a) Strain decomposition

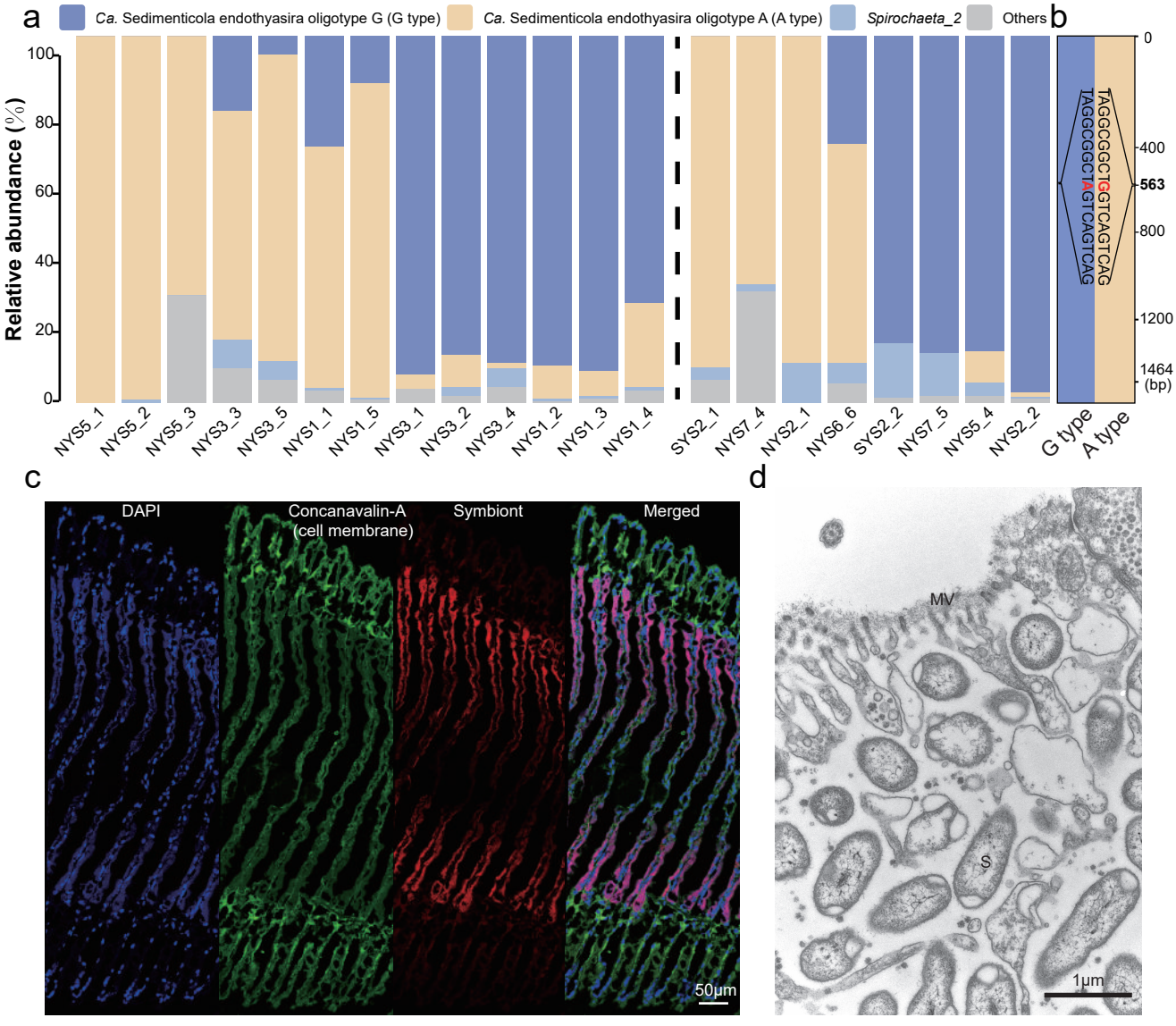
and SNP density of core genes in two representative strains of each symbiont population. A total of two strains were obtained, including G type and A type. Here, ploidy = 2 was set to uncover the SNP density of core genes of symbiont population. (b) Phylogenomic analysis of two symbiont oligotypes confirmed the taxonomy in Sedimenticolaceae (SOB), closely related to the cultured *Sedimenticola* bacterium, and the *Bathymodiolus azoricus* MOB symbiont was an outgroup. The solid black dot represents the bootstrap value of 100. The scale bar indicates 0.1 substitutions per site. (c) Average nucleotide identity (ANI) of pairwise MAGs and fixation index (F_{ST}) of each symbiont population matrix showed obvious partitioning based on the pairwise oligotypes, indicating there existed population differentiations when treating each host as a symbiont population. The G type represents the symbiont population that was dominated by the *Ca.* Sedimenticola endothyasira oligotypes G bacterium. The A type represents the symbiont population was dominated by the *Ca.* Sedimenticola endothyasira oligotypes A bacterium. The GA represents the pairwise F_{ST} between two oligotypes. Others are the same as above.

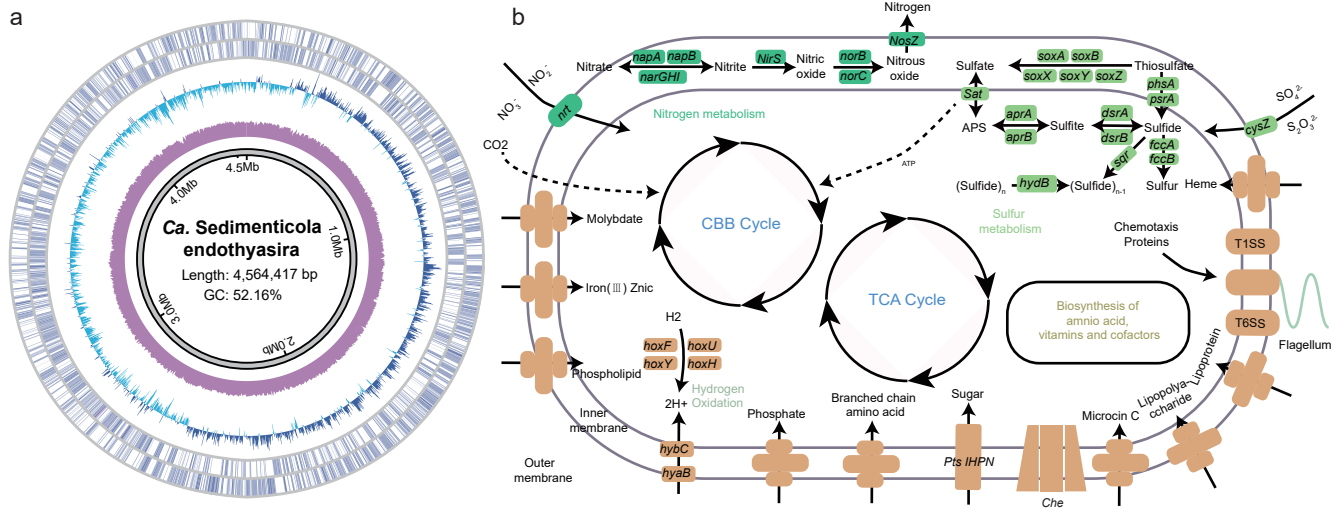
Fig. 5 Population structure analysis at the SNP level uncovered the difference of symbiont population. (a) PCA based on the total SNPs revealed symbiont population differentiation related to oligotype (G or A type) instead of sampling sites (two oligotypes: *pseudo-F* = 782.3 and *p*-value = 0.22). (b) Frequency of *Ca.* Sedimenticola endothyasira oligotype A (A type) percentage from StrainPanDA in 30 samples. Here, we divided into five groups, and the fitting curve was the non-normal distributions curve with *p*-value < 0.05. (c) The correlation between the mean F_{ST} of the symbiont population within each sampling site, and geographic distance and environmental factors (containing water depth, temperature and dissolved oxygen), showing that the formation of symbiont correlation were weakly influenced by geographic distance and environmental factors. (d) Variation partitioning of explanatory variables in RDA. Here, a total of 3 responding variables included: geographic distance, environmental factor and symbiont oligotype. Values less than 0 are hidden. *** = $p < 0.001$; ** = $p < 0.01$. (e) Principal component analysis (PCA) of intra-host (filled dots) nucleotide diversity (π values) and pairwise inter-host (empty dots) nucleotide. PCA was supported by PERMANOVA on pairwise Bray–Curtis dissimilarities with 9999 permutations. The GA type represents the pairwise π F_{ST} between two oligotypes. Others are the same as above. The bubble size represents the different π values, showing that each symbiont population have different nucleotide diversity (within-host π and between-host π: *pseudo-F* = 83.47 and *p*-value < 0.0001).

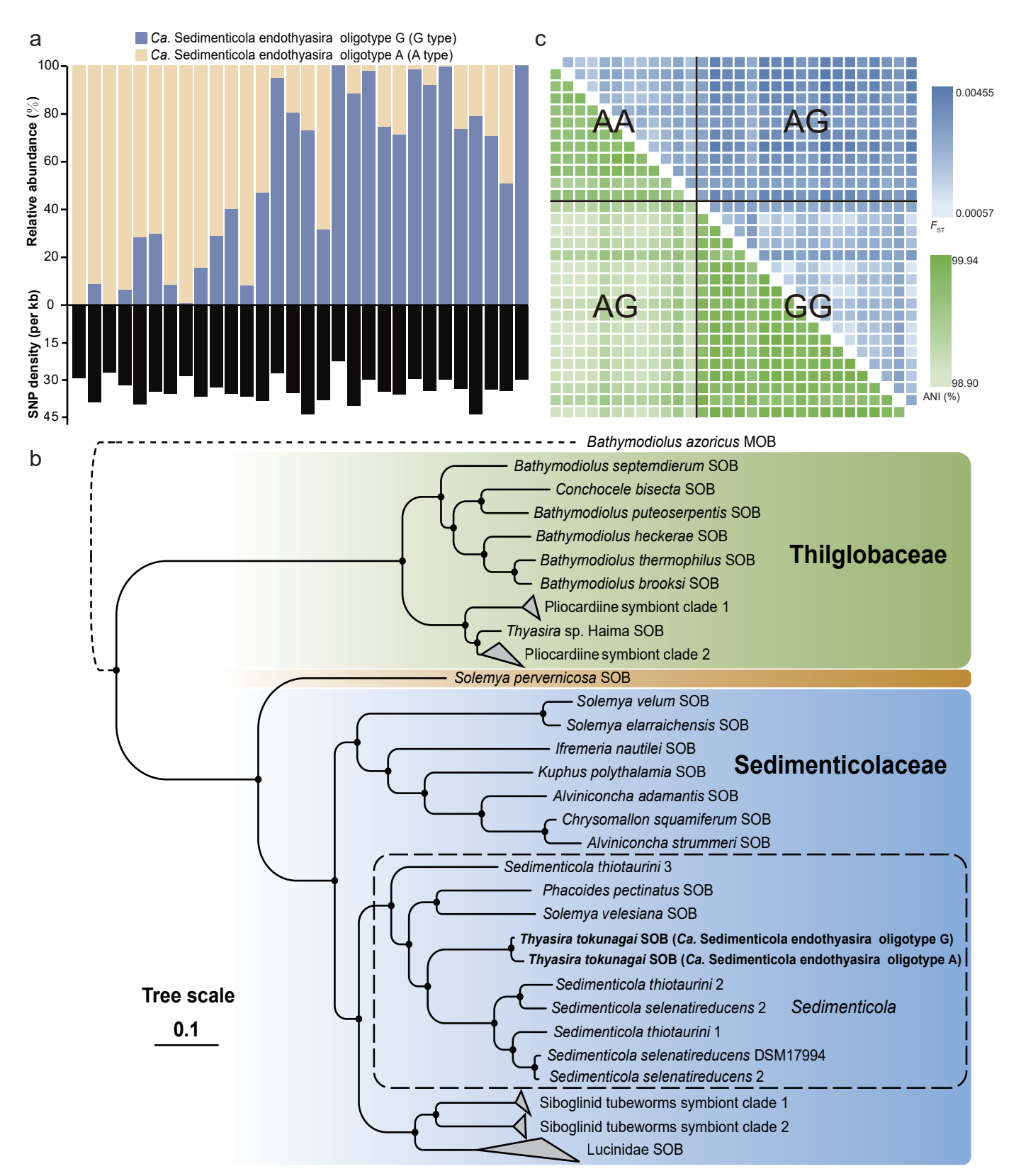
881 **Fig. 6 Gene expression difference of the two oligotypes individuals with the top 50 TPM**
882 **related to chemosymbiosis.** (a) PCA supported by the normalized gene expression data of the
883 host. (b) PCA supported by the normalized gene expression data of symbiont. (c, d) Comparative
884 gene expression differences related to chemosymbiosis were shown with the top 50 transcripts
885 per kilobase million (TPM) of host and symbiont, respectively. TPM was normalized by the z-
886 score method. G type individuals were represented by the blue color, while A type individuals
887 were represented by the yellow color.

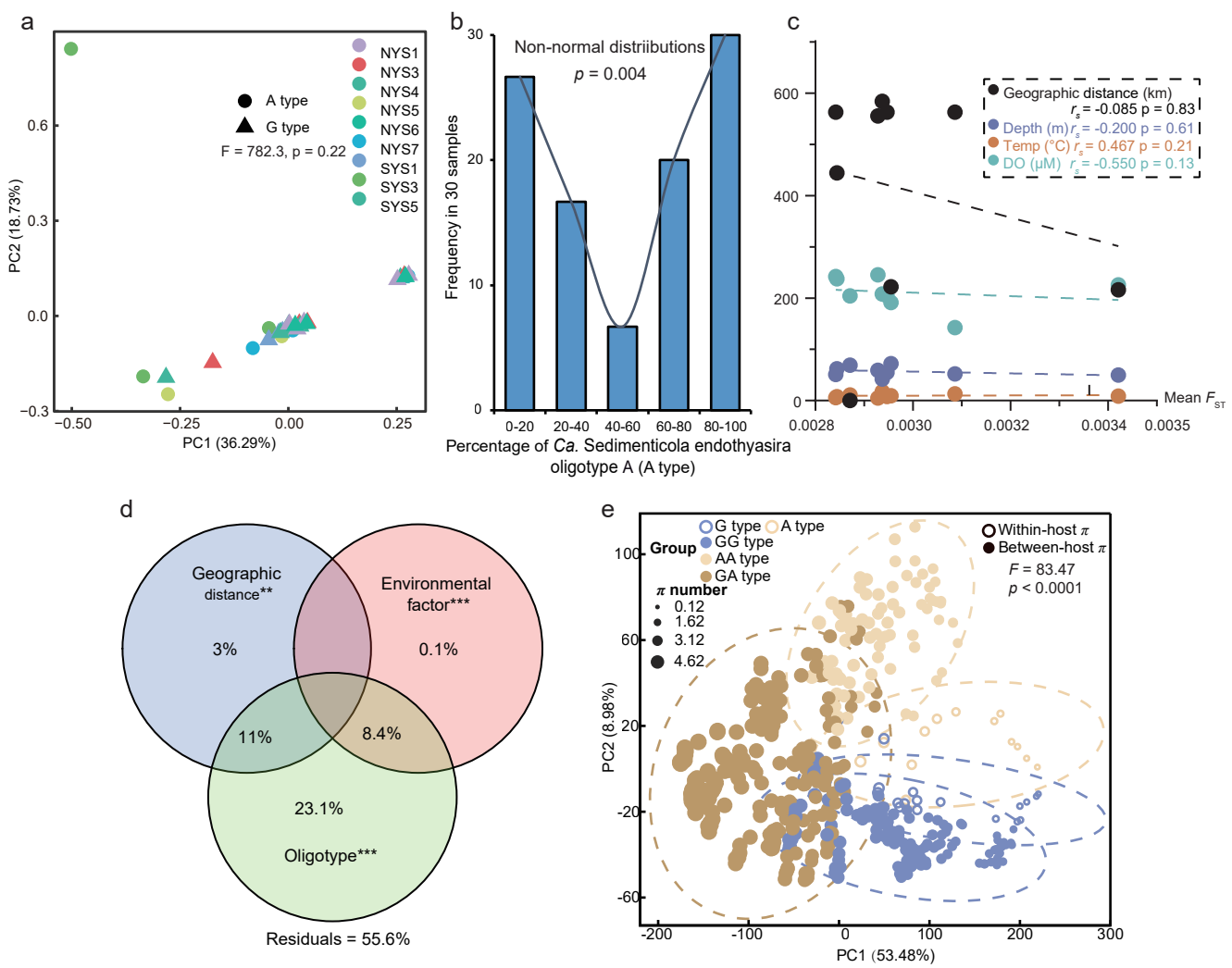
888

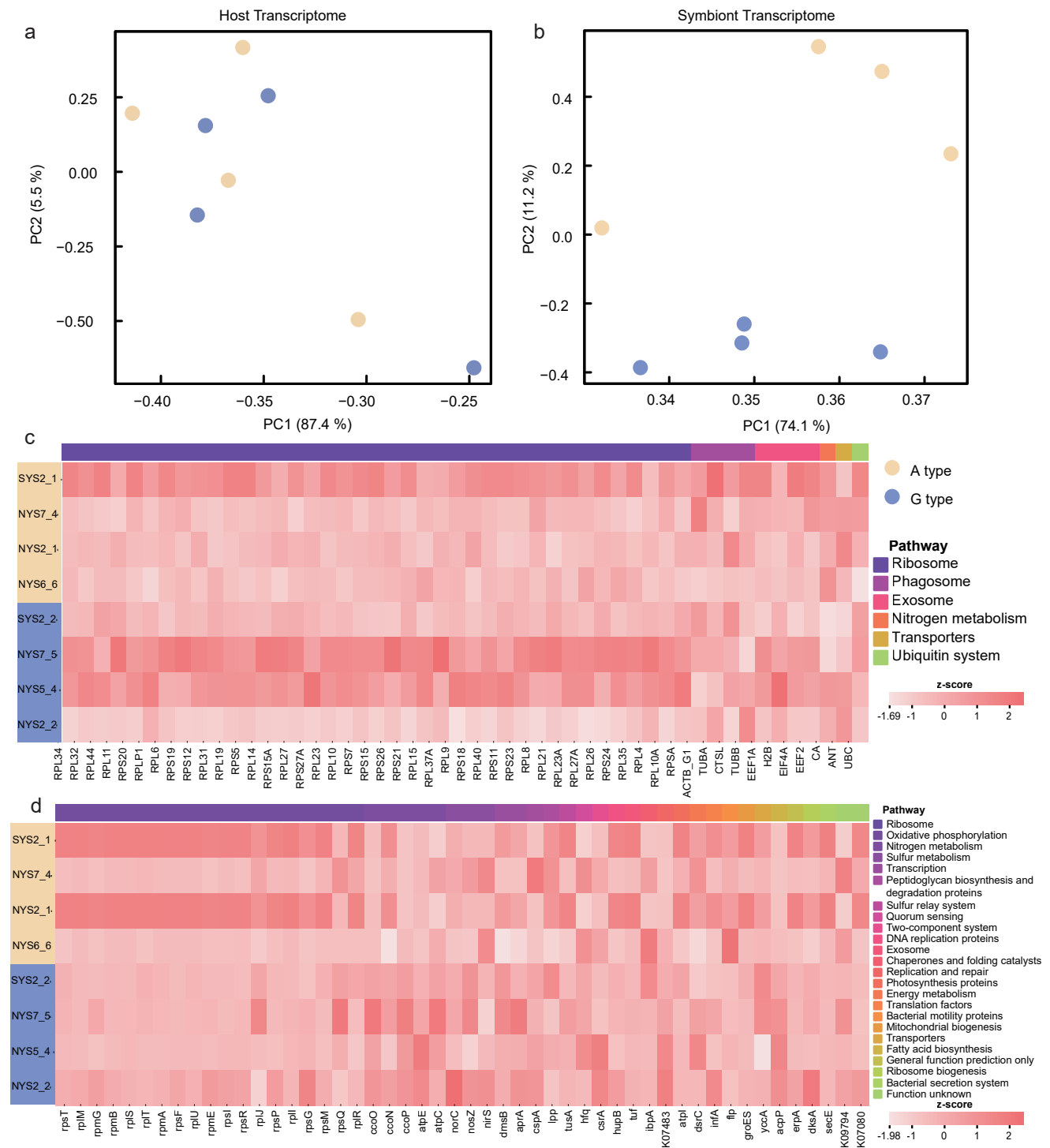

a
39° N
38° N
37° N
36° N
35° N
34° N
120°E
122°E
124°E
126°E
NYS1
NYS2
NYS3
NYS4
NYS5
NYS6
NYS7
SYS1
SYS2
SYS3
SYS4
SYS5
CHINA
Yellow Sea
2750m
1750m
900m
500m
200m
80m
40m
5m
5m
30m
70m
150m
400m
1μm
1μm
aa
g
pa
d
f
m
b
δ15N/‰
16
12
8
4
0
-26
-22
-18
-14
-10
POM
Plankton
Macrobenthos
Thyasira tokunagai
c
δ13C/‰
100
80
60
40
20
0
SYS1 SYS3 SYS5 NYS1 NYS3 NYS4 NYS5 NYS6 NYS7


a
Ca. Sedimenticola endothyasira oligotype G (G type)
Ca. Sedimenticola endothyasira oligotype A (A type)
Spirochaeta_2
Others
Relative abundance (%)
100
80
60
40
20
0
NYS5_1
NYS5_2
NYS5_3
NYS3_3
NYS3_5
NYS1_1
NYS1_5
NYS3_1
NYS3_2
NYS3_4
NYS1_2
NYS1_3
NYS1_4
SYS2_1
NYS7_4
NYS2_1
NYS6_6
SYS2_2
NYS7_5
NYS5_4
NYS2_2
b
TAGGCGGCTGGTCAGTCAG
TAGGCGGCTAGTCAGTCAG
0
400
563
800
1200
1464
(bp)
G type
A type
c
DAPI
Concanavalin-A
(cell membrane)
Symbiont
Merged
50μm
d
MV
S
1μm


a
4.5Mb
1.0Mb
2.0Mb
3.0Mb
4.0Mb
Ca. Sedimenticola endothyasira
Length: 4,564,417 bp
GC: 52.16%
b
Nitrogen
NosZ
Nitrate
napA napB narGHI
Nitrite
NirS
Nitric oxide
norB norC
Nitrous oxide
NO2-
NO3-
nrt
Nitrogen metabolism
CO2
Sulfate
Sat
APS
aprA aprB
Sulfite
soxA soxB soxX soxY soxZ
Thiosulfate
phsA psrA
dsrA dsrB
Sulfide
fccA fccB
sqr
cysZ
SO4 2-
S2O3 2-
(Sulfide)n
hydB
(Sulfide)n-1
Sulfur
Heme
ATP
Molybdate
Iron(Ⅲ) Znic
Phospholipid
CBB Cycle
TCA Cycle
Sulfur metabolism
Chemotaxis Proteins
T1SS
T6SS
Flagellum
Biosynthesis of amnio acid, vitamins and cofactors
H2
hoxF hoxY
hoxU hoxH
Hydrogen Oxidation
2H+
Inner membrane
Outer membrane
hybC
hyaB
Phosphate
Branched chain amino acid
Sugar
Pts IHPN
Che
Microcin C
Lipopolysaccharide
Lipoprotein


a
Ca. Sedimenticola endothyasira oligotype G (G type)
Ca. Sedimenticola endothyasira oligotype A (A type)
Relative abundance (%)
100
80
60
40
20
0
SNP density (per kb)
15
30
45
c
AA
AG
AG
GG
0.00455
0.00057
F_{ST}
99.94
98.90
ANI (%)
b
Bathymodiolus azoricus MOB
Bathymodiolus septemdierum SOB
Conchocele bisecta SOB
Bathymodiolus puteoserpentis SOB
Bathymodiolus heckerae SOB
Bathymodiolus thermophilus SOB
Bathymodiolus brooksi SOB
Pliocardiine symbiont clade 1
Thyasira sp. Haima SOB
Pliocardiine symbiont clade 2
Thilglobaceae
Solemya pervernicosa SOB
Solemya velum SOB
Solemya elarraichensis SOB
Ifremeria nautilei SOB
Kuphus polythalamia SOB
Alviniconcha adamantis SOB
Chrysomallon squamiferum SOB
Alviniconcha strummeri SOB
Sedimenticolaceae
Sedimenticola thiotaurini 3
Phacoides pectinatus SOB
Solemya velesiana SOB
Thyasira tokunagai SOB (Ca. Sedimenticola endothyasira oligotype G)
Thyasira tokunagai SOB (Ca. Sedimenticola endothyasira oligotype A)
Sedimenticola thiotaurini 2
Sedimenticola selenatireducens 2
Sedimenticola
Sedimenticola thiotaurini 1
Sedimenticola selenatireducens DSM17994
Sedimenticola selenatireducens 2
Siboglinid tubeworms symbiont clade 1
Siboglinid tubeworms symbiont clade 2
Lucinidae SOB
Tree scale
0.1


a
A type
G type
F = 782.3, p = 0.22
NYS1
NYS3
NYS4
NYS5
NYS6
NYS7
SYS1
SYS3
SYS5
PC2 (18.73%)
PC1 (36.29%)
b
Non-normal distriibutions
p = 0.004
Frequency in 30 samples
0-20
20-40
40-60
60-80
80-100
Percentage of Ca. Sedimenticola endothyasira oligotype A (A type)
c
Geographic distance (km)
rs = -0.085 p = 0.83
Depth (m) rs = -0.200 p = 0.61
Temp (°C) rs = 0.467 p = 0.21
DO (μM) rs = -0.550 p = 0.13
Mean FST
d
Geographic distance**
3%
Environmental factor***
0.1%
11%
8.4%
23.1%
Oligotype***
Residuals = 55.6%
e
G type
A type
Group
GG type
AA type
GA type
π number
0.12
1.62
3.12
4.62
Within-host π
Between-host π
F = 83.47
p < 0.0001
PC2 (8.98%)
PC1 (53.48%)


a
Host Transcriptome
PC2 (5.5 %)
PC1 (87.4 %)
b
Symbiont Transcriptome
PC2 (11.2 %)
PC1 (74.1 %)
c
A type
G type
Pathway
Ribosome
Phagosome
Exosome
Nitrogen metabolism
Transporters
Ubiquitin system
z-score
SYS2_1
NYS7_4
NYS2_1
NYS6_6
SYS2_2
NYS7_5
NYS5_4
NYS2_2
RPL34 RPL32 RPL44 RPL11 RPS20 RPLP1 RPL6 RPS19 RPS12 RPL31 RPL19 RPS5 RPL14 RPS15A RPL27 RPS27A RPL23 RPL10 RPS7 RPS15 RPS26 RPS21 RPL15 RPL37A RPL9 RPS18 RPL40 RPS11 RPS23 RPL8 RPL21 RPL23A RPL27A RPL26 RPS24 RPL35 RPL4 RPL10A RPSA ACTB_G1 TUBA CTSL TUBB EEF1A H2B EIF4A EEF2 CA ANT UBC
d
Pathway
Ribosome
Oxidative phosphorylation
Nitrogen metabolism
Sulfur metabolism
Transcription
Peptidoglycan biosynthesis and degradation proteins
Sulfur relay system
Quorum sensing
Two-component system
DNA replication proteins
Exosome
Chaperones and folding catalysts
Replication and repair
Photosynthesis proteins
Energy metabolism
Translation factors
Bacterial motility proteins
Mitochondrial biogenesis
Transporters
Fatty acid biosynthesis
General function prediction only
Ribosome biogenesis
Bacterial secretion system
Function unknown
z-score
rpsT rplM rpmG rpmB rplS rplT rpmA rpsF rplU rpmE rpsI rpsR rplJ rpsP rplI rpsG rpsM rpsQ rplR ccoO ccoN ccoP atpE atpC norC nosZ nirS dmsB aprA cspA lpp tusA hfq csrA hupB tuf ibpA K07483 atpI dsrC infA flp groES yccA acpP erpA dksA secE K09794 K07080