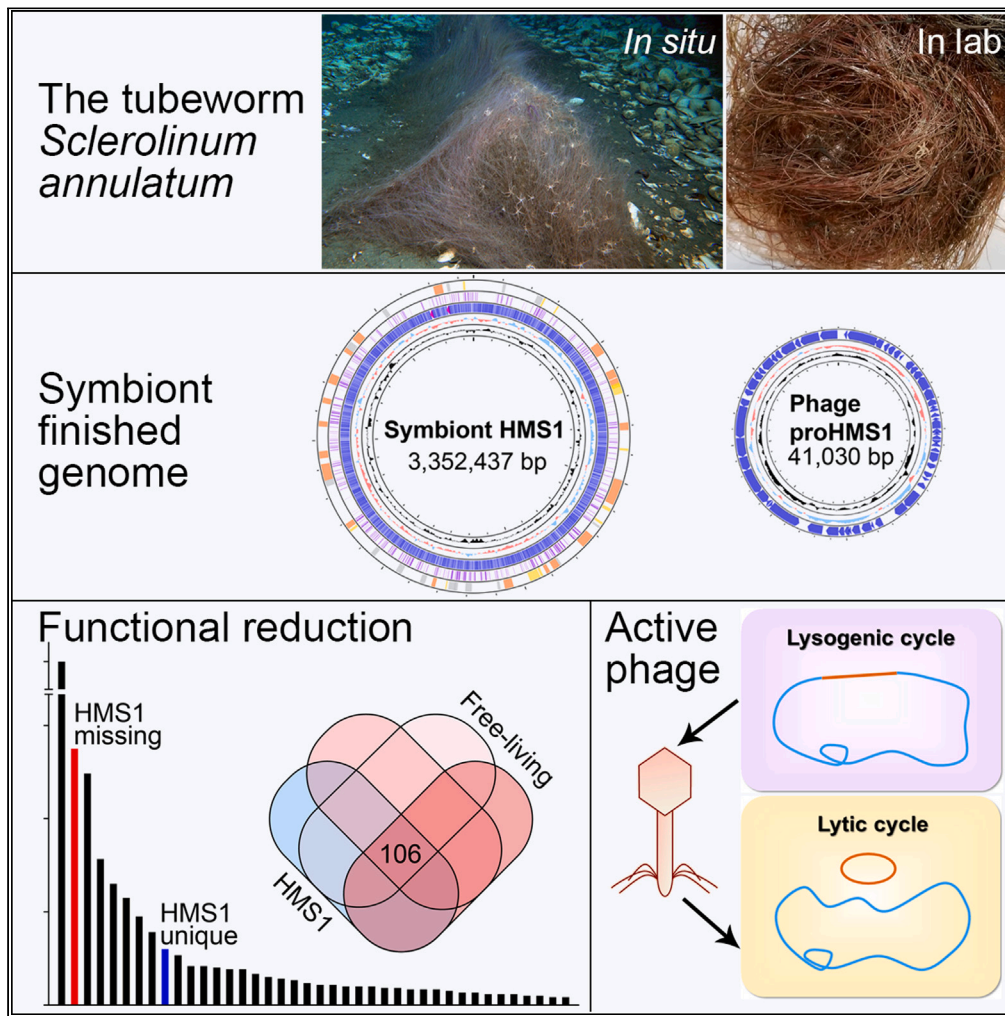


Article

Early genome erosion and internal phage-symbiont-host interaction in the endosymbionts of a cold-seep tubeworm



Zhao-Ming Gao,
Ting Xu, Hua-
Guan Chen, ...,
Hong-Bin Wang,
Jian-Wen Qiu,
Yong Wang

qiujuw@hkbu.edu.hk (J.-W.Q.)
wangyong@sz.tsinghua.edu.
cn (Y.W.)

Highlights

A finished endosymbiont genome of *Sclerolinum* tubeworm was assembled

It shows signs of early genome erosion and obligate endosymbiotic lifestyle

The symbiont evolution was mediated by mobile genetic elements

An active phage likely has interactions with the symbiont and the host



Article

Early genome erosion and internal phage-symbiont-host interaction in the endosymbionts of a cold-seep tubeworm

Zhao-Ming Gao,^{1,2,9} Ting Xu,^{3,4,9} Hua-Guan Chen,^{1,5} Rui Lu,^{1,5} Jun Tao,⁶ Hong-Bin Wang,⁶ Jian-Wen Qiu,^{4,7,*} and Yong Wang^{1,2,8,10,*}

SUMMARY

Endosymbiosis with chemosynthetic Gammaproteobacteria is widely recognized as an adaptive mechanism of siboglinid tubeworms, yet evolution of these endosymbionts and their driving forces remain elusive. Here, we report a finished endosymbiont genome (HMS1) of the cold-seep tubeworm *Sclerolium annulatum*. The HMS1 genome is small in size, with abundant prophages and transposable elements but lacking gene sets coding for denitrification, hydrogen oxidation, oxidative phosphorylation, vitamin biosynthesis, cell pH and/or sodium homeostasis, environmental sensing, and motility, indicative of early genome erosion and adaptive evolution toward obligate endosymbiosis. Unexpectedly, a prophage embedded in the HMS1 genome undergoes lytic cycle. Highly expressed ROS scavenger and LexA repressor genes indicate that the tubeworm host likely activates the lysogenic phage into lytic cycle through the SOS response to regulate endosymbiont population and harvest nutrients. Our findings indicate progressive evolution of *Sclerolium* endosymbionts toward obligate endosymbiosis and expand the knowledge about phage-symbiont-host interaction in deep-sea tubeworms.

INTRODUCTION

Due to the limited availability of organic matter from the ocean surface, the majority of the deep-sea floor is characterized by low biomass.¹ However, at hydrothermal vents, cold seeps and organic falls, several groups of invertebrates including siboglinid tubeworms, vesicomid clams, bathymodiolinid mussels, and peltospiroid snails have formed symbiotic relationships with chemosynthetic bacteria that use reducing substances such as hydrogen sulfide, methane and hydrogen.^{2,3} Such chemosymbiosis provides the hosts with most or even all of their energy and nutrition, allowing these specialized deep-sea invertebrates to develop into dense populations that rival the intertidal mussel beds in biomass.⁴ Studies over the last four decades have provided insights into the phylogenetic positions of these specialized invertebrates, and the molecular mechanisms of symbiosis between these invertebrates and their chemosynthetic bacteria.^{3,5}

It is widely known that deep-sea tubeworm species in the family Siboglinidae (Annelida) usually depend on sulfur-oxidizing bacteria for carbon fixation by using reductive fluids as an energy source.² Siboglinidae consists of four main lineages, including the giant tubeworms Vestimentifera, the bone-eating *Osedax*, and the slender tubeworms *Sclerolium* and *Frenulata*.⁶ Among them, vestimentiferans inhabiting hydrothermal vents and cold seeps are conspicuous and structure-forming, provide a habitat for many other invertebrates, and thus have been extensively studied.^{7–9} Because vestimentiferans obtain nutrition from their endosymbionts, they lack both the mouth and gut. Vestimentiferans usually host a single ribotype of sulfur-oxidizing Gammaproteobacteria,¹⁰ although polyclonal endosymbionts have also been reported.^{11,12} Vestimentiferan endosymbionts are horizontally acquired from free-living pools during the host's larval stage and maintained in a specialized organ of the adult termed trophosome.¹³ When the host dies, the endosymbionts are released back into the ambient environment.¹⁴ Nevertheless, no information is available on the transmission mode of the *Sclerolium*, *Osedax*, and *Frenulata* symbionts.

¹Institute of Deep Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

²HKUST-CAS Sanya Joint Laboratory of Marine Science Research, Chinese Academy of Sciences, Sanya 572000, China

³Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong, China

⁴Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

⁵University of Chinese Academy of Sciences, Beijing 101408, China

⁶MLR Key Laboratory of Marine Mineral Resources, Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou 511458, China

⁷Department of Biology, Hong Kong Baptist University, Hong Kong, China

⁸Institute for Ocean Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518000, China

⁹These authors contributed equally

¹⁰Lead contact

*Correspondence: qijuw@hkbu.edu.hk (J.-W.Q.), wangyong@sz.tsinghua.edu.cn (Y.W.)

<https://doi.org/10.1016/j.isci.2023.107033>



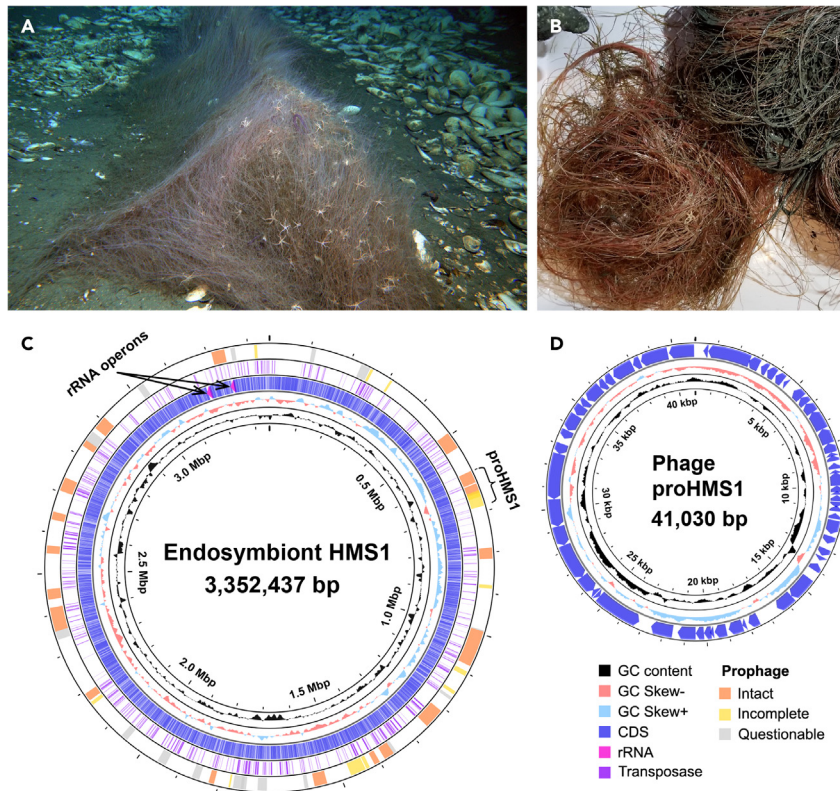


Figure 1. Photography of the tubeworm *Sclerolinum annulatum* and genomic view of its endosymbiont HMS1 and phage proHMS1

(A) *In situ* photograph of a population of *S. annulatum* on the Haima cold seep.

(B) Its specimens on aboard.

(C) Key features of the circular HMS1 genome.

(D) Key features of the circular proHMS1 genome. For both HMS1 and proHMS1, their genome size was indicated in the center. From the innermost to the outermost, the first circle represents GC content, the second circle represents GC skew, the third circle shows open reading frames (CDS, tRNA, and rRNA), the fourth circle shows genes encoding transposases, and the fifth circle shows genomic regions of prophages predicted using the PAST web server. Details of the intact, incomplete and questionable prophages are in Table S13. Locations of the rRNA operons and proHMS1 are labeled on the HMS1 genome. Genomic features were plotted with CGView Server (<http://cgview.ca/>). Abbreviation: CDS, coding sequence.

Siboglinid endosymbionts are currently unculturable. Their host/geographic specificity, genetic diversity, metabolic potentials and symbiotic mechanisms have only been investigated by culture-independent molecular methods. Metagenomic analyses have been used to assemble more than ten siboglinid endosymbiont genomes.^{10,15–22} Analyses of the known Vestimentiferan and Frenulata (the *Galathealinum* genus) gammaproteobacterial endosymbionts indicate that they likely employ mixotrophic lifestyles and have conserved mechanisms for host infection.^{16,17} In contrast, endosymbionts of the distantly related bone-eating *Osedax* tubeworms belong to the order Oceanospirillales and appear to be heterotrophic.²³ Nevertheless, because of a lack of sequencing effort on *Sclerolinum* species, the metabolic potential and genetic diversity of their symbionts remain unknown.

The genus *Sclerolinum* is the sister clade to vestimentiferans. They have been reported from deep-sea cold seep sediments and wood falls.^{24,25} Among the eight species of *Sclerolinum*, only *Sclerolinum contortum* has been examined for its endosymbionts through light and electron microscopy, bacterial 16S rRNA gene amplification and fluorescent *in situ* hybridization,^{26–28} yet no metagenomic studies have been carried out for them. The recent discovery of *Sclerolinum annulatum* (Figures 1A and 1B) from the Haima cold seep in the South China Sea has provided an opportunity to fill this knowledge gap.²⁹ Here, we employed culture-independent metagenomic and metatranscriptomic approaches to characterize the

Table 1. General genomic features of the endosymbiont HMS1 and its relatives

Genome ^a	Genome size (Mbp)	N ₅₀ (Kbp)	No. of contigs	% GC	No. of CDSs	No. of signal peptides	% coding density	Host
HMS1	3.35	NA	1	50.41	4,129	130	80.80	<i>S. annulatum</i>
MCR1	4.26	89.9	82	50.52	3,886	349	88.14	MCR vestimentiferans
MCR2	4.64	42.8	200	50.19	4,205	373	86.18	
<i>Paraescarpia echinospica</i>	4.06	381.7	14	53.5	3,522	366	87.28	Seep vestimentiferans
<i>Seepiophila jonesi</i>	3.53	20.7	323	54.39	3,561	312	86.05	
<i>Lamellibrachia luymesii</i>	3.53	20.6	337	54.36	3,559	311	85.84	
<i>Lamellibrachia barhami</i>	4.17	4.15	19	53.94	3,884	366	85.88	
<i>Escarpia laminata</i>	4.25	124.9	71	54.52	3,768	385	88.73	
<i>Escarpia piscesae</i>	4.06	313.6	23	54.18	3,700	395	89.62	
<i>Riftia pachyptila</i>	3.48	29.7	197	58.82	3,364	290	87.10	Vent vestimentiferans
<i>Ridgeia piscesae</i>	3.44	84.0	97	58.88	3,133	316	88.87	
<i>Tevnia jerichonana</i>	3.64	92.7	184	58.16	3,312	308	86.34	
<i>Galathealinum brachiosum</i>	3.77	726.6	14	38.89	3,450	404	88.66	Frenulata
<i>Sedimenticola thiotaurini</i>	3.96	NA	1	56.83	3,612	302	88.41	Free-living
<i>Sedimenticola selenitireducens</i>	4.56	223.5	41	56.65	4,240	384	87.88	
<i>Sedimenticola endophacoides</i>	2.98	27.7	172	63.86	3,421	197	87.36	Lucinidae

^aThe siboglinid symbionts were labeled using the names of their hosts. Data for HMS1 is from this study. Data for the relatives come from our annotations of their genomes. Detailed information is shown in [Tables S1](#) and [S2](#). Abbreviations: CDS, coding sequence; NA, not applicable.

endosymbionts of *S. annulatum*. We successfully obtained a complete genome of the endosymbiont. For the first time, we found remarkable functional reduction but abundant mobile genetic elements in a siboglinid endosymbiont indicative of the early stage of genome erosion and a host-restricted lifestyle. Unexpectedly, we found that a prophage embedded in the endosymbiont genome was activated to enter the lytic cycle and may help the tubeworm host to regulate the population size of the endosymbionts.

RESULTS

Endosymbiont genome recovery and metatranscriptomes

Specimens of *S. annulatum* were collected from the Haima cold seep at 1,425 m depth in May, 2019 ([Figures 1A](#) and [1B](#)). The tubeworms formed dense aggregations with empty shells of the vesicomyid clams scattered around ([Figure 1A](#)). A sediment core collected adjacent to the tubeworm colony was rich in hydrogen sulfide and gas hydrate. Using both Illumina and PacBio sequencing reads from one tubeworm, we successfully assembled a complete circular bacterial genome (hereafter referred to as HMS1) with a total length of 3.35 Mbp and a GC content of 50.41% ([Table 1](#), [Figure 1C](#), and [Table S1](#)). HMS1 harbors two rRNA operons with their 16S rRNA genes having two base mismatches ([Figures 1C](#) and [S1](#)). The two 16S rRNA genes entirely match the previously reported OTU-1 phylotypes that represent the endosymbiont of *S. annulatum*.²⁹ Phylogenetic analysis of the 16S rRNA gene sequences (1,495 bp) consistently showed high similarities between HMS1 and the endosymbionts hosted by *S. contortum* from the Gulf of Mexico (GenBank: HE614013) and the Haakon Mosby Mud Volcano (GenBank: AM883183) with > 99.6% sequence identity ([Figure S2](#)). These results indicate that HMS1 represents the endosymbiont of *S. annulatum*.²⁹

Phylogenomic analysis based on 43 conserved bacterial proteins revealed that HMS1 formed a well-supported clade with the endosymbionts of *Escarpia* sp. and *Lamellibrachia* sp. 2 (the MCR1 and MCR2 symbionts, [Figure 2A](#)) from a low-temperature hydrothermal diffuse vent along the Mid-Cayman Rise (MCR).¹⁶ HMS1 is also closely related to the free-living *Sedimenticola* species collected from estuarine or salt marsh sediment, the gill symbionts hosted by the mangrove lucinid *Phacoides pectinatus*³⁰ and the intertidal mudflat solemyid *Solemya velesiana*,³¹ whereas it is distant from other vestimentiferan endosymbionts

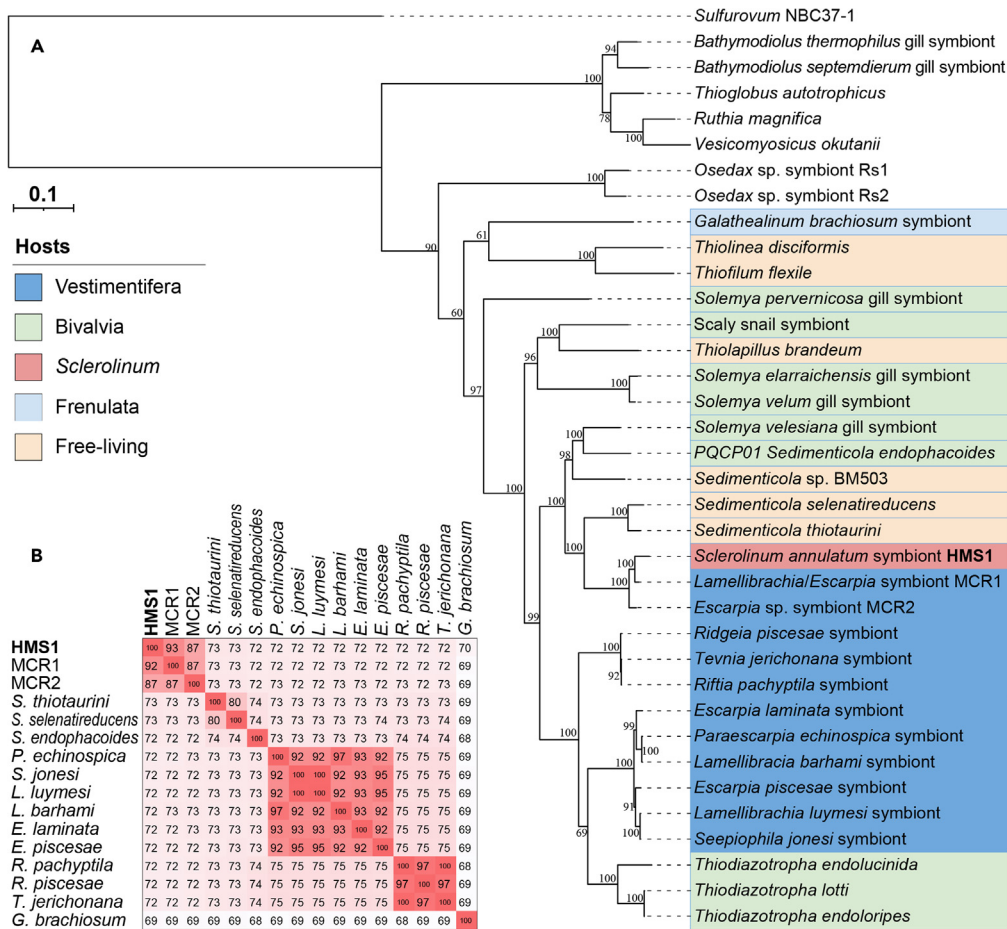


Figure 2. Phylogenomic analysis and ANI calculations for the endosymbiont HMS1 and its relatives

(A) Maximum-likelihood tree of tubeworm endosymbionts based on 43 concatenated conserved proteins derived from CheckM was constructed with the PROTGAMMALG model using raxmlGUI v1.5 for 100 replicates. Bootstrap values are shown as percentages on the branches. The scale bar indicates the mean number of amino acid substitutions per site. *Sulfurovum* sp. NBC37-1 was used as the outgroup. The GenBank accession number of genomes present in the tree and respective references are provided in Table S2.

(B) The ANI values between HMS1 and its relatives. A full version of ANI calculations can be found in Table S3.

from vents and seeps (Figure 2A and Table S2).²¹ HMS1 is even more distantly related to the *Galathealinum* endosymbiont.¹⁷ Although HMS1 is closely related to the MCR endosymbionts in the 16S rRNA gene (98.5–99.1% identity) and in the phylogenomic tree (Figure 2A), and is highly similar in GC content (50.41% vs. 50.19–50.52%, Table 1), genome rearrangement analysis indicates that the HMS1 genome is structurally highly distinct from the MCR genomes (Figure S3). Furthermore, the average nucleotide identity (ANI) values between the HMS1 and MCR genomes are < 92.5%, lower than the intraspecific ANI threshold of 95% (Figure 2B and Table S3).³² The ANI values between HMS1 and the other reference genomes are even lower (Figure 2B and Table S3). Therefore, HMS1 should be a novel species in the same genus as the MCR endosymbionts that is *Candidatus Vondammii*,¹⁶ and herein named as *Candidatus Vondammii deminuti* (*deminuti*: decreasing in Latin, referring to early genome reduction).

Three individuals of *S. annulatum* stored in RNAlater at -80°C were used for transcriptomic analysis. Metatranscriptomic sequencing produced 54 million, 109 million, and 105 million clean paired-end reads, respectively (Table S4). Among them, a total of 1,844,893, 1,250,911, and 2,780,863 reads were mapped to the coding sequences (CDSs) of HMS1, respectively. Of the total 4,129 CDSs, 3,782 CDSs (91.6%) mapped by metatranscriptomic reads of at least two of the three individuals were considered expressed (Figure S4 and Table S5). Average transcripts per million value (TPM_{av}) of the expressed genes varied from 0.21

to 26114.27, with the top 10% highly expressed genes having a TPM_{av} of > 540. Gene expression levels among the biological replicates were highly correlated (Pearson correlation coefficients ranging from 0.73 to 0.91, [Figure S4](#)), indicating their high transcriptional consistency.³³ Metatranscriptomic analysis revealed that two *iscR* genes involved in the transcription of iron-sulfur cluster assembly and a number of housekeeping genes encoding ribosomal proteins and chaperonins are the most actively expressed genes. Notably, the genes encoding antioxidant enzymes including a superoxide dismutase and a peroxiredoxin, as well as the *lexA* and *recA* genes that take part in the SOS response, also ranked among the actively expressed genes.³⁴

Reduced genome of the endosymbiont HMS1

In light of the phylogenetic inference ([Figure 2A](#)), we structurally and functionally compared HMS1 with those of the vestimentiferan endosymbionts and the representative *Sedimenticola* species (the free-living *Sedimenticola selenatireducens* and *Sedimenticola thiotaurini*, and the symbiotic *Sedimenticola endophacoides* of the lucinid *P. pectinatus*) ([Table 1](#)). As a far distinct relative, the *Galathealinum* endosymbiont was excluded from the comparison. The *S. velesiana* gill symbiont was also excluded because of the low completeness (86.18%) of its genome assembly ([Table S6](#)).

The HMS1 genome is smaller (3.35 Mbp) than those of the closely related MCR endosymbionts (4.26 Mbp and 4.64 Mbp) and the free-living *Sedimenticola* species (3.99 Mbp and 4.59 Mbp), but larger than the symbiotic *S. endophacoides* genome (2.98 Mbp) ([Table 1](#)). The HMS1 genome is also smaller than those of the other vent and seep siboglinid endosymbionts (3.44 Mbp to 4.06 Mbp). Meanwhile, HMS1 has a much lower number of signal peptide-containing genes ($n = 130$) compared to the vestimentiferan endosymbionts ($n = 290$ – 404), the free-living *Sedimenticola* species ($n = 302$ – 384) and the symbiotic *S. endophacoides* ($n = 197$) ([Table 1](#)). In addition, the HMS1 genome has a lower coding density (80.80%) than all the compared genomes (85.84–89.62%). However, HMS1 has more CDSs ($n = 4,129$) than most of the vestimentiferan endosymbionts ($n = 3133$ – 4180) and *Sedimenticola* species ($n = 3,421$ – $4,240$) ([Table 1](#)), mainly due to more transposase genes in HMS1 than in the compared genomes (1,057 vs. 8–223). When transposase genes are excluded, HMS1 has fewer CDSs than all the compared genomes ([Table S1](#)). HMS1 also has the smallest number of CDSs assigned to KEGG Orthologies (KOs) and COG categories after excluding transposase genes ([Table S1](#)). Functional clustering based on KEGG/COG/PFAM annotations showed that HMS1 is clearly separated from the vestimentiferan endosymbionts and the free-living *Sedimenticola* species ([Figures 3A](#) and [S5](#)). Instead, as phylogenetically close relatives of HMS1, the MCR endosymbionts exhibit a much closer relationship with the free-living *Sedimenticola* species and vestimentiferan endosymbionts than with HMS1 ([Figure 3A](#)).

UpSet diagram further illustrates the divergence between HMS1 and the vestimentiferan endosymbionts ([Figure 3B](#)). Among the 2,421 KOs identified in the analyzed genomes, 1,010 are present in all genomes. HMS1 has 36 unique KOs but lacks 165 KOs that are present in all vestimentiferan endosymbionts. Of interest, 134 of the 165 HMS1-missing KOs are present in the two free-living *Sedimenticola* species and 106 are even present in the symbiotic *S. endophacoides* ([Figure 3C](#)). Functional interpretation shows that the HMS1-missing KOs include genes with important functions, such as energy and nutrient metabolisms, cell pH and/or sodium homeostasis, environmental sensing, motility, and secretion ([Figures 4](#) and [S6](#)). HMS1 lacks extra 149 KOs when only compared to its two closely relative MCR endosymbiont genomes ([Figure 3B](#)).

Lost functions of the endosymbiont HMS1

Energy and nutrient metabolisms

The genes involved in key energy and nutrient metabolisms are summarized in [Figure 4A](#) and [Table S7](#). HMS1 has all the genes in the Calvin–Benson–Bassham (CBB) cycle and the reductive tricarboxylic acid (rTCA) cycle for carbon fixation, as well as genes encoding the sox system and the reverse sulfate reduction pathway for sulfur oxidation, confirming the conservation of these core metabolic pathways across siboglinid chemoautotrophic symbionts.^{17,20} Genes involved in the rTCA cycle were expressed at a higher level than those in the CBB cycle in HMS1 ([Figure 4B](#)), which is consistent with the findings of vestimentiferan endosymbionts.^{15,20} However, the group I NiFe type hydrogenases (HyaBA: KEGG KOs K06281 and K06282) that are present in all the vestimentiferan endosymbionts and *Sedimenticola* species are missing in HMS1. HMS1 retains the *napABC* and *nirBD* genes that constitute the intact dissimilatory nitrate reduction pathway, but lacks dissimilatory denitrifying genes (*nirS*, *norBC* and *nosZ*) that are present in all the

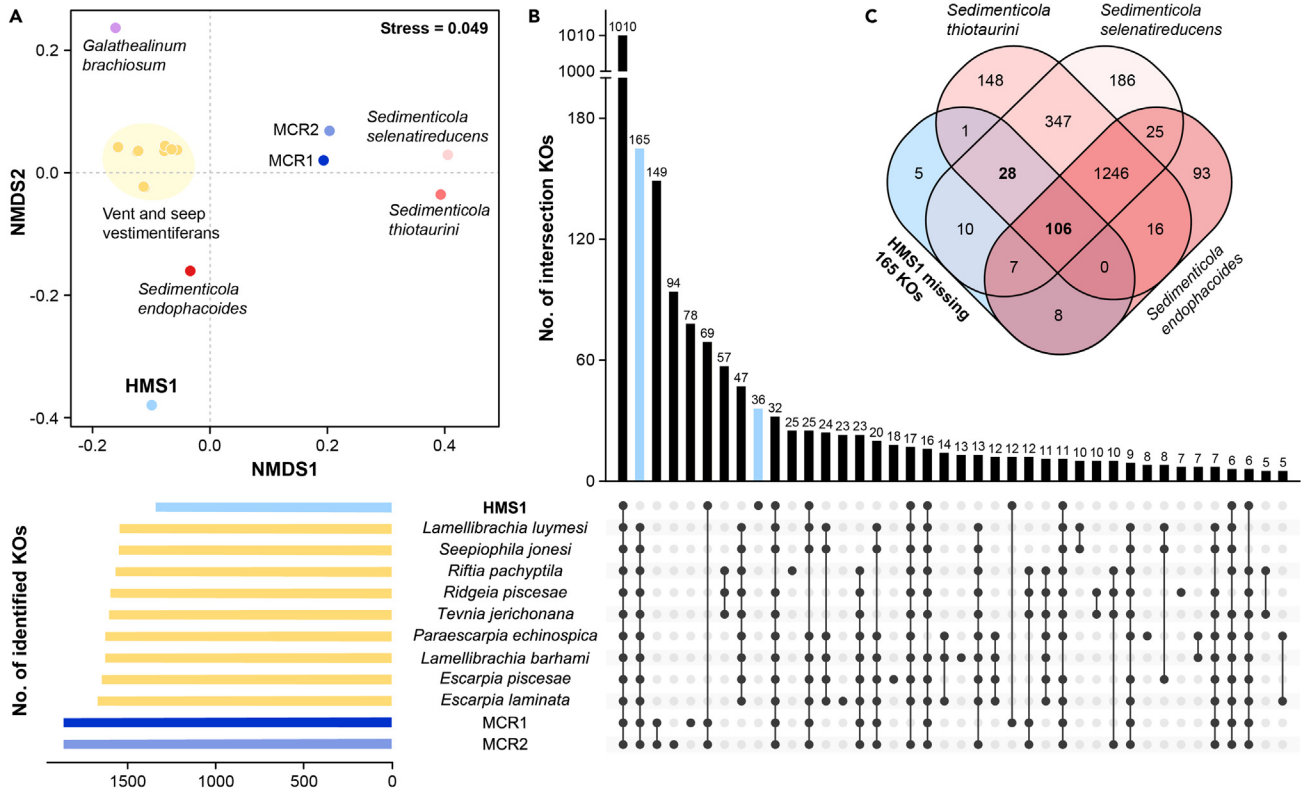


Figure 3. Functional comparison of the endosymbiont HMS1 and its relatives

(A) Non-metric multidimensional scaling (nMDS) ordination of HMS1 and reference genomes based on KEGG annotations. The two-dimensional stress value for the nMDS was 0.049 based on the Bray-Curtis distance. Analysis was performed using PRIMER-E based on the relative abundance of KOs.

(B) UpSet diagram of unique and shared KOs for HMS1 and the vestimentiferan endosymbionts.

(C) Venn diagram shows the number of intersected KOs between the 165 HMS1-missing KOs and the *Sedimenticola* species. Upset and Venn diagrams were drawn using the BioLadder cloud server (<https://www.bioladder.cn/>).

vestimentiferan endosymbionts and *Sedimenticola* species. All the available vestimentiferan endosymbiont genomes encode both the F- and V-type ATPases, but HMS1 lacks a whole set of *atpABCDEIK* genes encoding the V-type ATPase. In addition, HMS1 lacks the *men* and *cob* genes for menaquinone (vitamin K2) and cobalamin (B12) synthesis. The loss of these genes indicates a simpler metabolic capability of the *S. annulatum* endosymbiont.

Transporters

HMS1 has 119 genes assigned to transporters (KEGG Brite: ko02000), substantially fewer than the number of such genes in the vestimentiferan endosymbionts ($n = 154\text{--}200$), the free-living *Sedimenticola* species ($n = 202\text{--}218$) and the symbiotic *S. endophacoides* ($n = 146$) (Table S8). The genes encoding iron (III) transporter (*afuABC*) and peptides/nickel transporter (ABC.PE.E, ABC.PE.E, ABC.PE.A) that are present in all the vestimentiferan endosymbionts and free-living *Sedimenticola* species are missing in HMS1. Notably, the *nhaA* and *mnhBCDEFG* genes encoding Na^+/H^+ antiporters that function in cellular salt and pH homeostasis³⁵ are present in the vestimentiferan endosymbionts and *Sedimenticola* species but absent in HMS1 (Figure 4A).

Two-component systems (TCSs)

The TCSs genes are summarized in Table S9. HMS1 encodes 29 genes assigned to TCSs (KEGG Brite: ko02022), which is fewer than the number of such genes in the vestimentiferan endosymbionts ($n = 46\text{--}61$), the free-living *Sedimenticola* species ($n = 59\text{--}60$), and the symbiotic *S. endophacoides* ($n = 42$). The PhoR-PhoB and PhoQ-PhoP TCSs that have been proposed to mainly sense host-related signals and mediate adaptation of symbionts to the host's internal environment^{36,37} are present in all the analyzed tubeworm endosymbionts including HMS1, and the respective CDSs of HMS1 are expressed (Figure 4B),

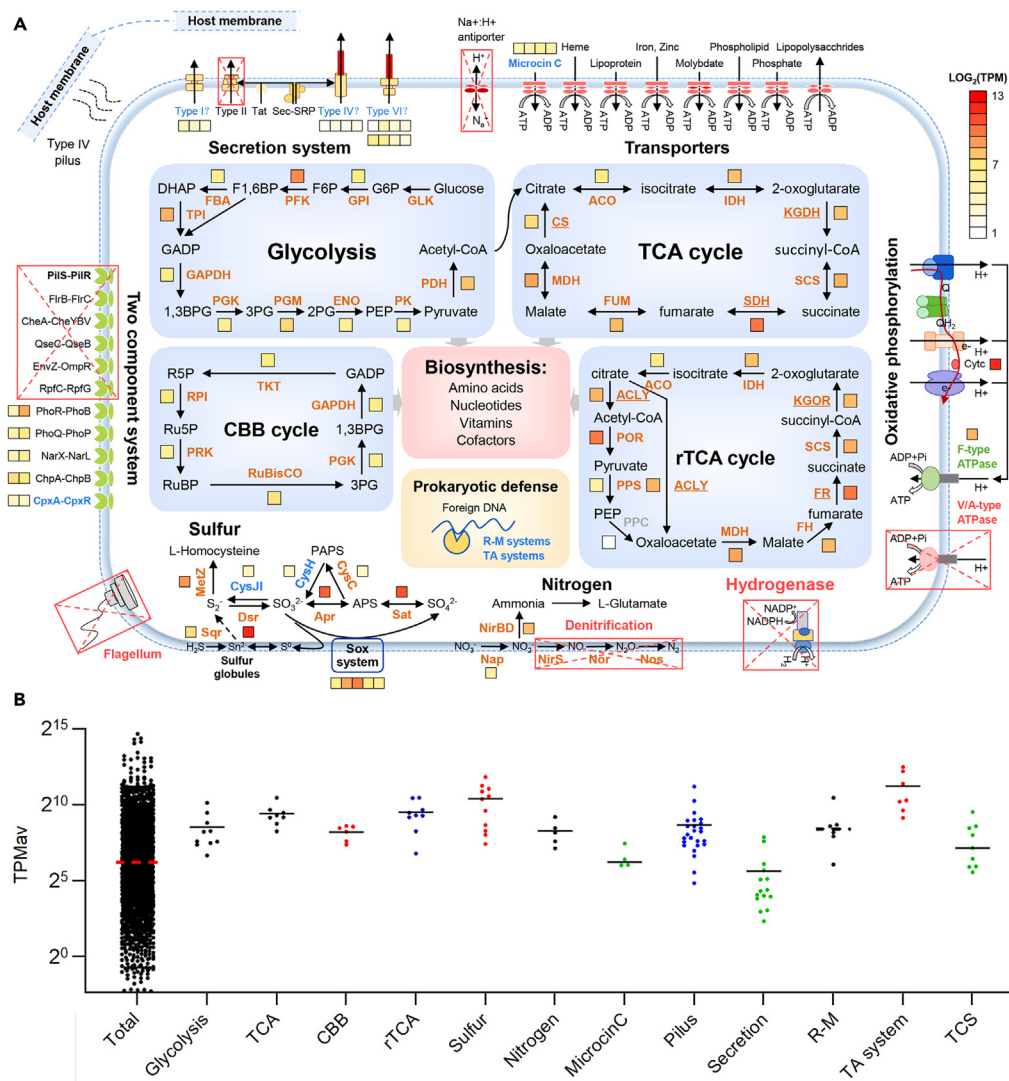


Figure 4. Schematic overview of the lifestyle of the *Sclerolium annulatum* endosymbiont HMS1

(A) Major cellular features and central metabolism pathways. Boxes adjacent to genes are colored according to the log-transformed TPM_{av} values (mean of three replicates) to represent expression levels. Functions missing in HMS1 are indicated by red rectangles with a cross. Functions present in HMS1 but absent in most or all other tubeworm endosymbionts are highlighted in blue.

(B) Expression of selected KEGG pathways and Brites. The median expression level in each group is represented by a solid line.

indicating their essential role in maintaining the endosymbiosis. Unlike the vestimentiferan endosymbionts and free-living *Sedimenticola* species, HMS1 has lost genes encoding the QseC-QseB, EnvZ-OmpR, PilS-PilR, CheA-CheYBV, RpfC-RpfG and FlrB-FlrC TCS systems (Figure 4A). These TCS modules, which function in bacteria's free-living stage, are responsible for regulating cell motility and biofilm formation through quorum sensing, and are required for the transcription of genes responsible for bacterial adherence and virulence.^{38–43} The loss of these TCSs indicates a reduced capability of the *S. annulatum* endosymbiont for environmental sensing.

Motility capability

All the available vestimentiferan endosymbiont genomes harbor the genes encoding flagella and T4P motility machinery as well as chemotaxis proteins (Figure 4A and Table S10). In contrast, HMS1 has lost all genes encoding the flagella and chemotaxis proteins (Table S10). As a two-component system,

PilS-PilR has been known to regulate the expression of the T4P major subunit PilA in *Pseudomonas aeruginosa*, and the T4P phenotype is missing in a *pilR* mutant.⁴⁴ HMS1 loses the *pilS* and *pilR* genes and thus likely has an incomplete T4P system. The lack of the flagella machinery and chemotaxis protein coupled with the incomplete T4P system all indicate a restricted motility capability of the *S. annulatum* endosymbiont.

Secretion systems

Type II secretion system (T2SS) has been suggested to export hemolysins and chitinases for the siboglinid endosymbionts to penetrate the host cells and migrate to the newly developed trophosome.¹⁷ A complete set of genes encoding T2SS system are present in the vestimentiferan endosymbionts and *Sedimenticola* species but missing in HMS1. Furthermore, gene sets encoding the T1SS, T4SS and T6SS systems for macromolecule export are either incomplete in HMS1 compared to vestimentiferan endosymbionts or lowly expressed (Figure 4A, Note S1, and Table S11).

Highly abundant and actively expressed mobile genetic elements

The HMS1 chromosome architecture is complex with highly abundant mobile genetic elements (MGEs), including prophages and transposable elements (TEs) (Figure 1C). A total of 71 genes encoding phage proteins were identified in HMS1 according to the COG annotation, which is much more abundant than 4–20 such genes in the vestimentiferan endosymbionts and 5–10 such genes in the *Sedimenticola* species (Table S12). Searching against the PHASTER web server⁴⁵ further revealed 47 prophage regions in HMS1, of which 19 regions are intact (Figure 1C and Table S13). These prophages, ranging from 4.0 to 88.8 kb in size, add up to 1.03 Mbp and account for 30.7% of the HMS1 genome. The density of prophage elements in the HMS1 genome is close to that of the insect parasite *Arsenophonus nasoniae*, whose genome encodes 55 prophage regions — the highest number of prophage complements reported to date.⁴⁶ Alignment view using the ViPTree online server⁴⁷ shows that the prophage elements 4, 6, 43 and 44 have syntenic regions that may come from the same ancestral phage through multiple invasions or transpositions (Figure S7). Several HMS1 genes belonging to the phage integrase family show high expression with TPM_{av} values up to 437.36 (Table S14).

HMS1 encodes 1,057 transposase genes as revealed by the COG annotation, compared to 8–223 and 20–49 such genes in the vestimentiferan endosymbionts and *Sedimenticola* species, respectively (Table S15), suggesting an extensive expansion of TEs in the *S. annulatum* endosymbiont. These transposase genes are located not just in the prophage regions, but relatively evenly distributed on the genome except for several regions having no insertions (Figure 1C). Annotation of the transposases in HMS1 using the ISfinder database and categorization by insertion sequence (IS) family⁴⁸ further revealed 986 transposase genes in 895 ISs (Table S16). Among them, 716 transposase genes are putatively truncated, while 201 transposase genes are complete and probably functional. Metatranscriptomic analysis showed that a large number of transposase genes were expressed, with some among the top 10% highly expressed genes (Table S17).

A prophage undergoing lytic cycle

Unlike the simple holobiont system between vestimentiferan tubeworms and their endosymbionts, the *S. annulatum* holobiont appears to be more complex with the participation of an active prophage here termed proHMS1 (Figure 1D). PacBio long reads mapping showed that proHMS1 could be integrated into the endosymbiont genome or be a circular genome, suggesting that it has both lysogenic and lytic lifestyles (Figure S8). The extra higher reads coverage on the phage region compared to the endosymbiont genome also indicates the lytic activity of proHMS1. The annular phage genome has a total length of 41,030 bp (Figure 1D). Taxonomic classification by ViPTree⁴⁷ showed that proHMS1 is affiliated with the family Siphoviridae (Figure S9). The proteomic tree also showed that proHMS1 is far distant from other previously known phages (Figure S9). Genes involved in phage core functions such as head, tail, capsid, terminase, recombinase and lysozyme were annotated, of which most were expressed, yet no auxiliary metabolic genes (AMGs) that modulate host cell metabolism were found (Table S18). Analyzing the meta-genomic data from ten *S. annulatum* individuals shows that proHMS1 is highly abundant in nine of them, but almost undetectable in the individual HM_W20 (Table S19). Correspondingly, the mean coverage of the endosymbiont HMS1 normalized by 100X mitochondrial sequence is 54.19X in HM_W20, higher than the values in eight other individuals (5.27–39.33X), which suggests a relatively higher abundance of the endosymbionts in proHMS1-missing tubeworm individuals.

Anti-viral defense systems

Most of the compared siboglinid endosymbionts and the two free-living *Sedimenticola* species have Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) spacers and encode CRISPR-Associated (Cas) proteins, but HMS1 lacks CRISPR arrays (Table S1, annotated through the Prokka v1.11 program) and Cas proteins (Table S20). CrisprMiner2⁴⁹ further confirms the missing of CRISPR-Cas system in HMS1. Instead, HMS1 has a number of genes encoding types I and II restriction-methylation (R-M) defense systems and types II (YoeB-YefM, HigB-HigA and VapC-VapB) and III (CptA-CptB) toxin-antitoxin (TA) systems (Table S20). Metatranscriptomic analysis further shows that most of the R-M and TA genes were expressed in HMS1 and some of them fell into the top 10% highly expressed genes with TPM_{av} up to 5781.71 (Tables S5 and S20). In addition, a gene encoding DNA methyltransferase Hhal in the proHMS1 genome was expressed with an extremely high TPM_{av} of 1692.34 (Table S18).

DISCUSSION

The rapid development of sequencing technologies has made it possible to produce high-quality genomes of tubeworm endosymbionts recently. Before this study, a highly contiguous siboglinid endosymbiont genome from *Lamellibrachia barhami*²¹ and a complete genome from *Riftia pachyptila* were available.²² In the present study, we successfully assembled a complete genome of the endosymbiont *Ca. Vondammii deminuti* (HMS1) from *S. annulatum*, which represents the first endosymbiont genome of *Sclerolium* tubeworms. Our identification of HMS1 as a novel species in the genus *Candidatus Vondammii*¹⁶ expands the genetic diversity of tubeworm endosymbionts and provides a useful reference for further systemic analyses of chemosynthetic bacteria.

Symbiosis in prokaryotes is a progressive process that could lead to genome reduction and functional deficiency.⁵⁰ Vertically transmitted symbionts of marine animals are apt to have a small genome and only retain essential functions that are beneficial to their hosts.^{51–53} However, the currently known siboglinid endosymbionts have a genome size of up to 4.64 Mbp, without obvious reduction when compared with their relatives living in natural environments.¹⁶ Such endosymbionts have functional traits that are necessary for their survival during the free-living stage.^{10,18,22} In comparison with the vestimentiferan endosymbionts and the free-living *Sedimenticola* species, HMS1 has undergone remarkable genome erosion. HMS1 has a smaller genome size compared to other tubeworm endosymbionts, lower coding density, a smaller number of CDSs and functional genes (when excluding transposase genes), and fewer signal-peptide-containing genes. All of these genome-level findings imply that the *S. annulatum* endosymbiont has likely undergone reductive evolution toward obligate endosymbiosis without a free-living stage,^{50,54} yet direct evidences for their vertical transmission lifestyle are yet unavailable.

Motility machines are typical traits to distinguish horizontally and vertically transmitted symbionts. For instance, the horizontally transmitted bathymodiolin symbionts and mixed-mode transmitted solemyid symbionts have genes encoding motility machines, yet the vertically transmitted vesicomid symbionts have lost them.^{31,53,55} The siboglinid tubeworms have been inferred to horizontally acquire their endosymbionts from the ambient environments.¹⁴ Flagella and type IV pilus (T4P) and associated chemotaxis proteins that are essential for tubeworm symbionts to approach their hosts¹⁷ and attach to the host surface⁵⁶ are encoded by their genomes. The loss of these genes in HMS1 implies that the *S. annulatum* endosymbiont is unable to reach tubeworm hosts like their free-living relatives. TCSs are used by bacteria to sense and respond to diverse environmental stresses and host signals.⁵⁷ The lack of those TCSs in HMS1 for quorum sensing, adherence and virulence of free-living bacteria indicates a reduced capability of the *S. annulatum* endosymbiont for environmental sensing. Na⁺/H⁺ antiporters in prokaryotic and eukaryotic membranes involve in the regulation of salt and pH homeostasis.³⁵ Tubeworm endosymbionts that are acquired from the external environments may confront various environmental challenges during their free-living stage, which necessitates the use of Na⁺/H⁺ antiporters to maintain cell homeostasis. The depletion of Na⁺/H⁺ antiporter genes in HMS1 indicates that the *S. annulatum* endosymbiont is likely always sheltered in the host's relatively stable internal environment. Furthermore, HMS1 lacks the CRISPR-cas system, consistent with the results of a previous study showing this system is missing in bacterial lineages adopting obligate endosymbiotic lifestyle.⁵⁸ Together, all these genomic features indicate that the *S. annulatum* endosymbiont likely has evolved toward obligate endosymbiosis, extending our knowledge on the lifestyle and evolutionary process of siboglinid endosymbionts.

The highly completed HMS1 genome allows accurate enumeration of MGEs. Transposases are often the main components of TEs that catalyze the movement of transposons within and between genomes, and are particularly abundant in pathogens and newly host-restricted symbionts.⁵⁹ Massive expansions of TEs have been reported to mediate the loss of unnecessary genes in the anglerfish symbiont genome.⁵¹ Phage elements could also promote prokaryotic genome evolution through gene disruption, duplication, transduction, or by acting as anchor points for major chromosomal rearrangements.⁶⁰ The highly abundant MGEs, including prophages and TEs in HMS1 together with their high expression activity, indicate that the *S. annulatum* endosymbiont is undergoing MGEs-mediated reductive evolution. However, the genome size of HMS1 remains substantially larger than those of strictly vertically transmitted symbionts such as the vesicomid endosymbiont (1.03 Mb)⁵³ and the flatworm symbiont (1.34 Mb).⁵² Because highly abundant TEs and MGEs are characteristic of obligate symbionts in the early stage,⁵⁰ we deduce that the *S. annulatum* endosymbiont probably remains in the early stage of genome erosion.

The *S. annulatum* endosymbiont is distinct from the vestimentiferan symbionts with regard to transmission mode, yet the ecological and evolutionary drivers of this difference remain unknown. The Haima cold seep has at least six disconnected seepage areas. The giant tubeworm *Paraescarpi echinospica* has been observed at three of these sites,⁶¹ whereas *S. annulatum* was only found at a site adjacent to the six active seeps.²⁹ This suggests a relatively low requirement of reductive fluids by *S. annulatum* when compared with vestimentiferans. Multiple *S. annulatum* individuals inhabit together and form a clump covering a large seafloor area, which might allow for full utilization of reductive fluids from the underlying sediment for chemoautotrophic biosynthesis. This capacity to utilize a low flux of cold seep fluids for survival may indicate a low growth rate and long-life span of *S. annulatum*. Tubeworms, especially those inhabiting cold seeps, could live for more than 200 years,⁶² and symbionts are released only when the host tubeworm dies. The restricted release of symbionts could lead to a shortage of symbiont bacteria in ambient environments and reduce the survival probability of tubeworm holobionts.⁶² Therefore, there is a possibility that the potentially long life span of *S. annulatum* allows the hosts to evolve into more innate relationships with their endosymbionts.

Although bacteriophages are widely distributed in natural habitats, their roles in symbiosis remain largely unknown.^{63,64} Even more rarely reported are active phages infecting bacterial endosymbionts in the deep ocean.^{65,66} Our identification of proHMS1 within the *S. annulatum* endosymbiont undergoing a lytic cycle offers an opportunity to understand the tripartite interactions among the phage, the endosymbiont and the animal host. To defend against the invasion of foreign genetic elements, prokaryotes possess defense systems such as CRISPR-Cas, R-M and TA immune systems.^{67,68} The missing of CRISPR-Cas system and the high expression levels of the R-M and TA genes indicate that the *S. annulatum* endosymbiont mainly relies on innate immunity to defend against the invasion of foreign viral particles, including the lytic phage proHMS1 and other potential phages in the host's trophosome environment.¹⁹ In addition, the highly expressed DNA methyltransferase HhaI gene of proHMS1 likely plays a counter-defense role in resisting defense systems of the symbiont like that happened in the deep-sea snail holobiont.⁶⁶ Our work thus provides a new case of internal phage-bacterial interactions in deep-sea chemosymbiosis ecosystems.

The mechanism that activates the lytic cycle of the lysogenic proHMS1 remains unclear. It has been reported that the SOS response induced by DNA damage can activate prophages into a lytic cycle,⁶⁹ and our work implies that the *lexA* and *recA* genes for the SOS response are highly expressed in the *S. annulatum* endosymbiont. As a main factor causing DNA damage and the SOS response,³⁴ a high level of oxidative stress in the symbiont is implied by the highly expressed ROS scavenger genes (among the top 50 expressed genes) in HMS1. A transcriptomic study of *R. pachyptila* symbionts with different sizes revealed higher expression of ROS scavengers in symbionts with a small size, and suggested that the symbionts confront host-induced oxidative stress in their division stage.⁷⁰ A recent study revealed that the complete genome of *Candidatus* Endoriftia persephone in the tubeworm *R. pachyptila* has two identical rRNA operons.²² HMS1 also has two rRNA operon copies that probably support its high growth rate and high frequency of cell division,⁷¹ which may cause a higher level of host-induced oxidative stress on the symbionts. Therefore, we propose a hypothesis that the tubeworm host-induced oxidative stress activates the lysogenic phage into a lytic cycle by triggering the SOS response. Because the relative abundance of endosymbionts is the highest in the proHMS1-missing *S. annulatum* individual, it is possible that the tubeworm host controls the density of the endosymbiont population and harvests nutrients from symbiotic chemoautotrophic bacteria through activating the lytic cycle of the prophages, resembling the role of WO phages infecting the symbiotic *Wolbachia* bacteria.⁶⁵ However, further studies are required to test this hypothesis.

Conclusions

In this study, we assembled *Ca. Vondammii deminuti* (HMS1) – the complete genome of the *S. annulatum* endosymbiont with characteristics of early genome erosion toward host-restricted lifestyle. Furthermore, we found a lytic phage of the *Sclerolium* symbiont and proposed a hypothesis of the complex symbiont-phage-host interactions. Given that HMS1 is the only assembled symbiont genome in the *Sclerolium* tubeworm genus, it remains uncertain whether genome reduction also occurs in the symbionts of other *Sclerolium* species. Although the vestimentiferan symbionts have been experimentally proved to be horizontally transmitted, it remains uncertain whether the Frenulata sulfur-oxidizing bacteria are obligate or facultative symbionts. The heterotrophic symbionts of *Osedax* have been proposed to be acquired secondarily after the loss of the original sulfur-oxidizing symbiont, yet more efforts are needed to test this hypothesis.²⁵ Thus, more representative symbionts of the *Sclerolium*, *Osedax* and Frenulata lineages should be sequenced to provide a better understanding of the diversity of symbioses in Siboglinidae, and to illustrate the evolution of such symbioses that enabled them to thrive in various deep-sea chemosynthetic habitats. The reported deep-sea interactions of active phage with symbionts are not fully illustrated, and thus require further analyses especially in detecting lytic prophage proHMS1 *in situ* and comparing the metatranscriptomic and proteomic data of tubeworm individuals with or without active prophages.

Limitations of the study

A limitation of our sampling design was a lack of surrounding seawater and sediment samples of the *Sclerolium* tubeworms, which prevented us from verifying the absence of the *Sclerolium* endosymbiont in the environment. Because the slender *S. annulatum* resides inside a very fragile tube, we could not obtain eggs and larvae for detecting the presence of endosymbiont using the fluorescence *in situ* hybridization method. Thus, research in the future should aim to provide such direct evidence to verify the obligate endosymbiosis of *Ca. Vondammii deminuti* with the *S. annulatum* host. As our samples were fixed on the deck of the research vessel (R/V), the metatranscriptome data might have been affected by changes in temperature and water pressure during the sample collection process, and thus could not comprehensively reflect the *in-situ* expression of the ROS scavenger and LexA repressor genes. Meanwhile, the incomplete tubeworm bodies and the limited number of proHMS1-missing individuals could weaken the comparison of the relative abundance of endosymbionts among individuals. Besides, the evidence chain to prove the hypothesis of tubeworm host controlling symbiont population through the active phage remains incomplete. Therefore, further efforts are also needed to clarify the relationships among the tubeworm host, the prophage proHMS1 and the endosymbiont HMS1.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Sample collection
 - DNA extraction
 - Metagenome sequencing
 - Metagenome assembly and genome binning
 - Genome annotation
 - Metatranscriptome sequencing and analyses
 - Phylogenetic analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.107033>.

ACKNOWLEDGMENTS

We are grateful to the captain and crew of the R/V *Haiyang 6* and the operation team of the ROV *Haima* for their professional support during the cruise. The study was supported by the National Natural Science Foundation of China (42076136), the Key Science and Technology Program of Hainan Province, China (ZDKJ2019011), the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (Guangzhou) (GML2019ZD0409 and SMSEGL20SC02) and the University Grants Committee of Hong Kong, Hong Kong SAR, China (GRF121010 and CRF C2013-22GF).

AUTHOR CONTRIBUTIONS

Y.W., J.W.Q., and H.B.W. conceived the project. Z.M.G. and T.X. designed the experiments. Z.M.G., T.X., and J.T. collected the samples and took the photos. T.X. and Z.M.G. extracted the DNA and RNA. T.X. assembled the metagenome of HM_W22 using a combination of the Illumina short reads and PacBio long reads, and binned the symbiont genome. Z.M.G. assembled and binned the metagenomes of the other individuals using the Illumina short reads. Z.M.G., R.L., H.G.C., and T.X. performed the downstream data analyses. Z.M.G. drafted the manuscript. All authors edited the manuscript and approved its submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 30, 2022

Revised: March 11, 2023

Accepted: May 31, 2023

Published: June 7, 2023

REFERENCES

- Weir, C.L., Rowe, G.T., Escobar-Briones, E., Boetius, A., Soltwedel, T., Caley, M.J., Soliman, Y., Huettmann, F., Qu, F., Yu, Z., et al. (2010). Global patterns and predictions of seafloor biomass using random forests. *PLoS One* 5, e15323. <https://doi.org/10.1371/journal.pone.0015323>.
- Dubilier, N., Bergin, C., and Lott, C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* 6, 725–740. <https://doi.org/10.1038/nrmicro1992>.
- Sogin, E.M., Kleiner, M., Borowski, C., Gruber-Vodicka, H.R., and Dubilier, N. (2021). Life in the dark: phylogenetic and physiological diversity of chemosynthetic symbioses. *Annu. Rev. Microbiol.* 75, 695–718. <https://doi.org/10.1146/annurev-micro-051021-123130>.
- Zhao, Y., Xu, T., Law, Y.S., Feng, D., Li, N., Xin, R., Wang, H., Ji, F., Zhou, H., and Qiu, J.W. (2020). Ecological characterization of cold-seep epifauna in the South China Sea. *Deep-Sea Res. PT. I* 163, 103361. <https://doi.org/10.1016/j.dsr.2020.103361>.
- Osman, E.O., and Weinnig, A.M. (2022). Microbiomes and obligate symbiosis of deep-sea animals. *Annu. Rev. Anim. Biosci.* 10, 151–176. <https://doi.org/10.1146/annurev-animal-081621-112021>.
- Hilário, A., Capa, M., Dahlgren, T.G., Halanych, K.M., Little, C.T.S., Thornhill, D.J., Verna, C., and Glover, A.G. (2011). New perspectives on the ecology and evolution of siboglinid tubeworms. *PLoS One* 6, e16309. <https://doi.org/10.1371/journal.pone.0016309>.
- Felbeck, H. (1981). Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* 213, 336–338. <https://doi.org/10.1126/science.213.4505.336>.
- Halanych, K.M. (2005). Molecular phylogeny of siboglinid annelids (a.k.a. pogonophorans): a review. *Hydrobiologia* 535–536, 297–307. <https://doi.org/10.1007/s10750-004-1437-6>.
- Bright, M., and Lallie, F. (2010). The biology of vestimentiferan tubeworms. *Oceanogr. Mar. Biol.* 48, 213–265. <https://doi.org/10.1201/Ebk1439821169-c4>.
- Robidart, J.C., Bench, S.R., Feldman, R.A., Novoradovsky, A., Podell, S.B., Gaasterland, T., Allen, E.E., and Felbeck, H. (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ. Microbiol.* 10, 727–737. <https://doi.org/10.1111/j.1462-2920.2007.01496.x>.
- Zimmermann, J., Lott, C., Weber, M., Ramette, A., Bright, M., Dubilier, N., and Petersen, J.M. (2014). Dual symbiosis with co-occurring sulfur-oxidizing symbionts in vestimentiferan tubeworms from a Mediterranean hydrothermal vent. *Environ. Microbiol.* 16, 3638–3656. <https://doi.org/10.1111/1462-2920.12427>.
- Polzin, J., Arevalo, P., Nussbaumer, T., Polz, M.F., and Bright, M. (2019). Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. *Proc. Biol. Sci.* 286, 20181281. <https://doi.org/10.1098/rspb.2018.1281>.
- Bright, M., and Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* 8, 218–230. <https://doi.org/10.1038/nrmicro2262>.
- Klose, J., Polz, M.F., Wagner, M., Schimak, M.P., Gollner, S., and Bright, M. (2015). Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. *Proc. Natl. Acad. Sci. USA* 112, 11300–11305. <https://doi.org/10.1073/pnas.1501160112>.
- Yang, Y., Sun, J., Sun, Y., Kwan, Y.H., Wong, W.C., Zhang, Y., Xu, T., Feng, D., Zhang, Y., Qiu, J.W., and Qian, P.Y. (2020). Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. *ISME J.* 14, 135–150. <https://doi.org/10.1038/s41396-019-0520-y>.
- Reveillaud, J., Anderson, R., Reves-Sohn, S., Cavanaugh, C., and Huber, J.A. (2018). Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome* 6, 19. <https://doi.org/10.1186/s40168-018-0411-x>.
- Li, Y., Liles, M.R., and Halanych, K.M. (2018). Endosymbiont genomes yield clues of

- tubeworm success. *ISME J.* 12, 2785–2795. <https://doi.org/10.1038/s41396-018-0220-z>.
18. Gardebrecht, A., Markert, S., Sievert, S.M., Felbeck, H., Thürmer, A., Albrecht, D., Wollherr, A., Kabisch, J., Le Bris, N., Lehmann, R., et al. (2012). Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J.* 6, 766–776. <https://doi.org/10.1038/ismej.2011.137>.
 19. Perez, M., and Juniper, S.K. (2016). Insights into symbiont population structure among three vestimentiferan tubeworm host species at eastern Pacific spreading centers. *Appl. Environ. Microbiol.* 82, 5197–5205. <https://doi.org/10.1128/AEM.00953-16>.
 20. Rubin-Blum, M., Dubilier, N., and Kleiner, M. (2019). Genetic evidence for two carbon fixation pathways (the Calvin-Benson-Bassham cycle and the Reverse Tricarboxylic Acid cycle) in symbiotic and free-living bacteria. *mSphere* 4, e00394-18–e00318. <https://doi.org/10.1128/mSphere.00394-18>.
 21. Breusing, C., Schultz, D.T., Sudek, S., Worden, A.Z., and Young, C.R. (2020). High-contiguity genome assembly of the chemosynthetic gammaproteobacterial endosymbiont of the cold seep tubeworm *Lamellibrachia barhami*. *Mol. Ecol. Resour.* 20, 1432–1444. <https://doi.org/10.1111/1755-0998.13220>.
 22. De Oliveira, A.L., Srivastava, A., Espada-Hinojosa, S., and Bright, M. (2022). The complete and closed genome of the facultative generalist *Candidatus* Endoriftia persephone from deep-sea hydrothermal vents. *Mol. Ecol. Resour.* 22, 3106–3123. <https://doi.org/10.1111/1755-0998.13668>.
 23. Goffredi, S.K., Yi, H., Zhang, Q., Klann, J.E., Struve, I.A., Vrijenhoek, R.C., and Brown, C.T. (2014). Genomic versatility and functional variation between two dominant heterotrophic symbionts of deep-sea *Osedax* worms. *ISME J.* 8, 908–924. <https://doi.org/10.1038/ismej.2013.201>.
 24. Mazumdar, A., Dewangan, P., Peketi, A., Gullapalli, S., Kalpana, M.S., Naik, G.P., Shetty, D., Pujari, S., Pillutla, S.P.K., Gaikwad, V.V., et al. (2019). The first record of active methane (cold) seep ecosystem associated with shallow methane hydrate from the Indian EEZ. *J. Earth Syst. Sci.* 128, 18. <https://doi.org/10.1007/s12040-018-1044-y>.
 25. Li, Y., Kocot, K.M., Schander, C., Santos, S.R., Thornhill, D.J., and Halanach, K.M. (2015). Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). *Mol. Phylogenet. Evol.* 85, 221–229. <https://doi.org/10.1016/j.ympev.2015.02.008>.
 26. Eichinger, I., Schmitz-Esser, S., Schmid, M., Fisher, C.R., and Bright, M. (2014). Symbiont-driven sulfur crystal formation in a thiotrophic symbiosis from deep-sea hydrocarbon seeps. *Environ. Microbiol. Rep.* 6, 364–372. <https://doi.org/10.1111/1758-2229.12149>.
 27. Eichinger, I., Klepal, W., Schmid, M., and Bright, M. (2011). Organization and microanatomy of the *Sclerolinum contortum* trophosome (polychaeta, Siboglinidae). *Biol. Bull.* 220, 140–153. <https://doi.org/10.1086/BBLv220n2p140>.
 28. Lösekann, T., Robador, A., Niemann, H., Knittel, K., Boetius, A., and Dubilier, N. (2008). Endosymbioses between bacteria and deep-sea siboglinid tubeworms from an arctic cold seep (Haakon Mosby Mud Volcano, barents sea). *Environ. Microbiol.* 10, 3237–3254. <https://doi.org/10.1111/j.1462-2920.2008.01712.x>.
 29. Xu, T., Sun, Y., Wang, Z., Sen, A., Qian, P.Y., and Qiu, J.W. (2022). The morphology, mitogenome, phylogenetic position, and symbiotic bacteria of a new species of *Sclerolinum* (Annelida: Siboglinidae) in the South China Sea. *Front. Mar. Sci.* 8, 793645. <https://doi.org/10.3389/fmars.2021.793645>.
 30. Lim, S.J., Davis, B.G., Gill, D.E., Walton, J., Nachman, E., Engel, A.S., Anderson, L.C., and Campbell, B.J. (2019). Taxonomic and functional heterogeneity of the gill microbiome in a symbiotic coastal mangrove lucinid species. *ISME J.* 13, 902–920. <https://doi.org/10.1038/s41396-018-0318-3>.
 31. Russell, S.L., Pepper-Tunick, E., Svedberg, J., Byrne, A., Ruelas Castillo, J., Vollmers, C., Beinart, R.A., and Corbett-Detig, R. (2020). Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. *PLoS Genet.* 16, e1008935. <https://doi.org/10.1371/journal.pgen.1008935>.
 32. Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* 106, 19126–19131. <https://doi.org/10.1073/pnas.0906412106>.
 33. Moitinho-Silva, L., Díez-Vives, C., Batani, G., Esteves, A.I., Jahn, M.T., and Thomas, T. (2017). Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J.* 11, 1651–1666. <https://doi.org/10.1038/ismej.2017.25>.
 34. Podlasek, Z., and Žgur Bertok, D. (2020). The DNA damage inducible SOS response is a key player in the generation of bacterial persister cells and population wide tolerance. *Front. Microbiol.* 11, 1785. <https://doi.org/10.3389/fmicb.2020.01785>.
 35. Ito, M., Morino, M., and Krulwich, T.A. (2017). Mrp antiporters have important roles in diverse bacteria and archaea. *Front. Microbiol.* 8, 2325. <https://doi.org/10.3389/fmicb.2017.02325>.
 36. Yuan, J., Jin, F., Glatter, T., and Sourjik, V. (2017). Osmosensing by the bacterial PhoQ/PhoP two-component system. *Proc. Natl. Acad. Sci. USA* 114, E10792–E10798. <https://doi.org/10.1073/pnas.1717272114>.
 37. Chakraborty, S., Sivaraman, J., Leung, K.Y., and Mok, Y.K. (2011). Two-component PhoB-PhoR regulatory system and ferric uptake regulator sense phosphate and iron to control virulence genes in type III and VI secretion systems of *Edwardsiella tarda*. *J. Biol. Chem.* 286, 39417–39430. <https://doi.org/10.1074/jbc.M111.295188>.
 38. Liu, J., Hu, L., Xu, Z., Tan, C., Yuan, F., Fu, S., Cheng, H., Chen, H., and Bei, W. (2015). *Actinobacillus pleuropneumoniae* two-component system QseB/QseC regulates the transcription of PilM, an important determinant of bacterial adherence and virulence. *Vet. Microbiol.* 177, 184–192. <https://doi.org/10.1016/j.vetmic.2015.02.033>.
 39. Kostakioti, M., Hadjifrangiskou, M., Pinkner, J.S., and Hultgren, S.J. (2009). QseC-mediated dephosphorylation of QseB is required for expression of genes associated with virulence in uropathogenic *Escherichia coli*. *Mol. Microbiol.* 73, 1020–1031. <https://doi.org/10.1111/j.1365-2958.2009.06826.x>.
 40. Li, S., Liang, H., Wei, Z., Bai, H., Li, M., Li, Q., Qu, M., Shen, X., Wang, Y., and Zhang, L. (2019). An osmoregulatory mechanism operating through OmpR and LrhA controls the motile-sessile switch in the plant growth-promoting bacterium *Pantoea alhagi*. *Appl. Environ. Microbiol.* 85, e00077-19–e00019. <https://doi.org/10.1128/AEM.00077-19>.
 41. Solano, C., Echeverez, M., and Lasa, I. (2014). Biofilm dispersion and quorum sensing. *Curr. Opin. Microbiol.* 18, 96–104. <https://doi.org/10.1016/j.mib.2014.02.008>.
 42. Luo, G., Huang, L., Su, Y., Qin, Y., Xu, X., Zhao, L., and Yan, Q. (2016). flrA, flrB and flrC regulate adhesion by controlling the expression of critical virulence genes in *Vibrio alginolyticus*. *Emerg. Microb. Infect.* 5, e85. <https://doi.org/10.1038/emi.2016.82>.
 43. Zhou, L., Lei, X.H., Bochner, B.R., and Wanner, B.L. (2003). Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *J. Bacteriol.* 185, 4956–4972. <https://doi.org/10.1128/jb.185.16.4956-4972.2003>.
 44. Kilmury, S.L.N., and Burrows, L.L. (2018). The *Pseudomonas aeruginosa* PilSR two-component system regulates both twitching and swimming motilities. *mBio* 9, e01310-18–e01318. <https://doi.org/10.1128/mBio.01310-18>.
 45. Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. <https://doi.org/10.1093/nar/gkw387>.
 46. Frost, C.L., Siozios, S., Nadal-Jimenez, P., Brockhurst, M.A., King, K.C., Darby, A.C., and Hurst, G.D.D. (2020). The hypercomplex genome of an insect reproductive parasite highlights the importance of lateral gene transfer in symbiont biology. *mBio* 11, e02590-19–e02519. <https://doi.org/10.1128/mBio.02590-19>.
 47. Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017). ViPTree: the viral proteomic tree server. *Bioinformatics* 33, 2379–2380. <https://doi.org/10.1093/bioinformatics/btx157>.
 48. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34,

- D32–D36. <https://doi.org/10.1093/nar/gkj014>.
49. Zhang, F., Zhao, S., Ren, C., Zhu, Y., Zhou, H., Lai, Y., Zhou, F., Jia, Y., Zheng, K., and Huang, Z. (2018). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun. Biol.* **1**, 180. <https://doi.org/10.1038/s42003-018-0184-6>.
 50. McCutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26. <https://doi.org/10.1038/nrmicro2670>.
 51. Hendry, T.A., Freed, L.L., Fader, D., Fenolio, D., Sutton, T.T., and Lopez, J.V. (2018). Ongoing transposon-mediated genome reduction in the luminous bacterial symbionts of deep-sea ceratioid anglerfishes. *mBio* **9**, e01033-18–e01018. <https://doi.org/10.1128/mBio.01033-18>.
 52. Jäckle, O., Seah, B.K.B., Tietjen, M., Leisch, N., Liebeke, M., Kleiner, M., Berg, J.S., and Gruber-Vodicka, H.R. (2019). Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*. *Proc. Natl. Acad. Sci. USA* **116**, 8505–8514. <https://doi.org/10.1073/pnas.1818995116>.
 53. Ip, J.C.H., Xu, T., Sun, J., Li, R., Chen, C., Lan, Y., Han, Z., Zhang, H., Wei, J., Wang, H., et al. (2021). Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. *Mol. Biol. Evol.* **38**, 502–518. <https://doi.org/10.1093/molbev/msaa241>.
 54. Hönigschmid, P., Bykova, N., Schneider, R., Ivankov, D., and Frishman, D. (2018). Evolutionary interplay between symbiotic relationships and patterns of signal peptide gain and loss. *Genome Biol. Evol.* **10**, 928–938. <https://doi.org/10.1093/gbe/evy049>.
 55. Ozawa, G., Shimamura, S., Takaki, Y., Takishita, K., Ikuta, T., Barry, J.P., Maruyama, T., Fujikura, K., and Yoshida, T. (2017). Ancient occasional host switching of maternally transmitted bacterial symbionts of chemosynthetic vesicomyid clams. *Genome Biol. Evol.* **9**, 2226–2236. <https://doi.org/10.1093/gbe/evx166>.
 56. Craig, L., Pique, M.E., and Tainer, J.A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**, 363–378. <https://doi.org/10.1038/nrmicro885>.
 57. He, Y.W., Wang, C., Zhou, L., Song, H., Dow, J.M., and Zhang, L.H. (2006). Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *J. Biol. Chem.* **281**, 33414–33421. <https://doi.org/10.1074/jbc.M606571200>.
 58. Burstein, D., Sun, C.L., Brown, C.T., Sharon, I., Anantharaman, K., Probst, A.J., Thomas, B.C., and Banfield, J.F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* **7**, 10613. <https://doi.org/10.1038/ncomms10613>.
 59. Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**, 722–732. <https://doi.org/10.1038/nrmicro1235>.
 60. Davies, E.V., Winstanley, C., Fothergill, J.L., and James, C.E. (2016). The role of temperate bacteriophages in bacterial infection. *FEMS Microbiol. Lett.* **363**, fnw015. <https://doi.org/10.1093/femsle/fnw015>.
 61. Xu, H., Du, M., Li, J., Zhang, H., Chen, W., Wei, J., Wu, Z., Zhang, H., Li, J., Chen, S., et al. (2020). Spatial distribution of seepages and associated biological communities within Haima cold seep field, South China Sea. *J. Sea Res.* **165**, 101957. <https://doi.org/10.1016/j.seares.2020.101957>.
 62. Sato, M., and Sasaki, A. (2021). Evolution and maintenance of mutualism between tubeworms and sulfur-oxidizing bacteria. *Am. Nat.* **197**, 351–365. <https://doi.org/10.1086/712780>.
 63. Oliver, K.M., Degnan, P.H., Hunter, M.S., and Moran, N.A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* **325**, 992–994. <https://doi.org/10.1126/science.1174463>.
 64. Pramono, A.K., Kuwahara, H., Itoh, T., Toyoda, A., Yamada, A., and Hongoh, Y. (2017). Discovery and complete genome sequence of a bacteriophage from an obligate intracellular symbiont of a cellulolytic protist in the termite gut. *Microbes Environ.* **32**, 112–117. <https://doi.org/10.1264/jsme2.ME16175>.
 65. Kent, B.N., and Bordenstein, S.R. (2010). Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol.* **18**, 173–181. <https://doi.org/10.1016/j.tim.2009.12.011>.
 66. Zhou, K., Xu, Y., Zhang, R., and Qian, P.Y. (2021). Arms race in a cell: genomic, transcriptomic, and proteomic insights into intracellular phage-bacteria interplay in deep-sea snail holobionts. *Microbiome* **9**, 182. <https://doi.org/10.1186/s40168-021-01099-6>.
 67. Dupuis, M.É., Villion, M., Magadán, A.H., and Moineau, S. (2013). CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat. Commun.* **4**, 2087. <https://doi.org/10.1038/ncomms3087>.
 68. Harms, A., Brodersen, D.E., Mitarai, N., and Gerdes, K. (2018). Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol. Cell* **70**, 768–784. <https://doi.org/10.1016/j.molcel.2018.01.003>.
 69. Filipiak, M., Łoś, J.M., and Łoś, M. (2020). Efficiency of induction of Shiga-toxin lambdaoid prophages in *Escherichia coli* due to oxidative and antibiotic stress depends on the combination of prophage and the bacterial strain. *J. Appl. Genet.* **61**, 131–140. <https://doi.org/10.1007/s13353-019-00525-8>.
 70. Hinzke, T., Kleiner, M., Meister, M., Schlüter, R., Hentschker, C., Pané-Farré, J., Hildebrandt, P., Felbeck, H., Sievert, S.M., Bonn, F., et al. (2021). Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. *Elife* **10**, e58371. <https://doi.org/10.7554/eLife.58371>.
 71. Roller, B.R.K., Stoddard, S.F., and Schmidt, T.M. (2016). Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* **1**, 16160. <https://doi.org/10.1038/nmicrobiol.2016.160>.
 72. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 73. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736. <https://doi.org/10.1101/gr.251087.116>.
 74. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Pribelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
 75. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158. <https://doi.org/10.1186/s40168-018-0541-1>.
 76. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
 77. Wick, R.R., and Holt, K.E. (2019). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res.* **8**, 2138. <https://doi.org/10.12688/f1000research.21782.3>.
 78. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963. <https://doi.org/10.1371/journal.pone.0112963>.
 79. Hunt, M., Silva, N.D., Otto, T.D., Parkhill, J., Keane, J.A., and Harris, S.R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **16**, 294. <https://doi.org/10.1186/s13059-015-0849-0>.
 80. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055. <https://doi.org/10.1101/gr.186072.114>.
 81. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.

82. Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
83. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. <https://doi.org/10.1093/nar/gky1085>.
84. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301. <https://doi.org/10.1093/nar/gkr1065>.
85. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. <https://doi.org/10.1093/nar/gkr485>.
86. Huang, Y., Gilna, P., and Li, W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25, 1338–1340. <https://doi.org/10.1093/bioinformatics/btp161>.
87. Katoh, K., and Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26, 1899–1900. <https://doi.org/10.1093/bioinformatics/btq224>.
88. Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
89. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
90. Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* 12, 335–337. <https://doi.org/10.1007/s13127-011-0056-0>.
91. Stewart, C.N., Jr., and Via, L.E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* 14, 748–750.
92. Bushnell, B. (2014). BBMap: a fast. Accurate, Splice-Aware Aligner (No. LBNL-7065E) (Lawrence Berkeley National Lab (LBNL)). <https://www.osti.gov/servlets/purl/1241166>.
93. Gao, Z.M., Huang, J.M., Cui, G.J., Li, W.L., Li, J., Wei, Z.F., Chen, J., Xin, Y.Z., Cai, D.S., Zhang, A.Q., and Wang, Y. (2019). *In situ* meta-omic insights into the community compositions and ecological roles of hadal microbes in the Mariana Trench. *Environ. Microbiol.* 21, 4092–4108. <https://doi.org/10.1111/1462-2920.14759>.
94. Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285. <https://doi.org/10.1007/s12064-012-0162-3>.
95. Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
RNAlater	Thermoisher Scientific	Cat#AM7021
Trizol	Invitrogen	Cat#15596026
Deposited data		
Raw metagenomic and metatranscriptomic data of the tubeworm <i>Sclerolimum annulatum</i>	This study	NCBI under BioProject PRJNA595466
The endosymbiont genome	This study	NCBI under accession number CP099567
Software and algorithms		
TRIMMOMATIC v0.38	Bolger et al., 2014 ⁷²	N/A
BBmap	https://www.osti.gov/servlets/purl/1241166	N/A
Canu v1.8	Koren et al., 2017 ⁷³	N/A
SPAdes v3.13	Bankevich et al., 2012 ⁷⁴	N/A
MetaWrap v1.2.2	Uritskiy et al., 2018 ⁷⁵	N/A
minimap2	Li, 2018 ⁷⁶	N/A
Miniasm/Minipolish	Wick & Holt, 2019 ⁷⁷	N/A
Pilon v.1.22	Walker et al., 2014 ⁷⁸	N/A
Circlator 1.5.5	Hunt et al., 2015 ⁷⁹	N/A
CheckM v1.1.0	Parks et al., 2015 ⁸⁰	N/A
Prokka v1.11	Seemann, 2014 ⁸¹	N/A
BlastKOALA	Kanehisa et al., 2016 ⁸²	N/A
eggNOG-mapper	Huerta-Cepas et al., 2019 ⁸³	N/A
pfam_scan.pl	Punta et al., 2012 ⁸⁴	N/A
PHAST web server	Zhou et al., 2011 ⁸⁵	N/A
CoverM 0.4.0	https://github.com/wwood/CoverM	N/A
Meta_RNA	Huang et al., 2009 ⁸⁶	N/A
MAFFT L-INS-i v7.294b	Katoh & Toh, 2010 ⁸⁷	N/A
Bandage 0.8.1	Wick et al. 2015 ⁸⁸	N/A
trimAl v1.4	Capellagutiérrez et al., 2009 ⁸⁹	N/A
raxmlGUI v1.5	Silvestro & Michalak, 2012 ⁹⁰	N/A
pyani	https://pypi.org/project/pyani/	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yong Wang (wangyong@sz.tsinghua.edu.cn).

Materials availability

All unique/stable reagents in this study are available from the [lead contact](#) without restriction.

Data and code availability

Raw metagenomic and metatranscriptomic reads are deposited in the NCBI Sequence Reads Archive (SRA) database under BioProject PRJNA595466. The endosymbiont genome is deposited in the NCBI GenBank database under the genome accession number CP099567.

This paper does not report original code.

Additional information required to reanalyze the data in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No experimental models were used in this study.

METHOD DETAILS

Sample collection

Specimens of *S. annulatum* were collected from the Haima cold seep (depth 1,425 m) in the South China Sea²⁹ using ROV *Haima* on board the R/V *Haiyang 6* in May 2019. Because the slender *S. annulatum* (i.e., diameter <0.6 mm) resides inside a very fragile tube, all of the sampled individuals were incomplete. Upon arrival at the main deck of the R/V, specimens were rinsed several times with 0.22 μ m membrane-filtered iced seawater to remove sediment, debris, and microbes attached to their tubes. The cleaned specimens were either frozen immediately at -80°C or preserved in 100% ethanol for metagenomic analyses, or fixed in RNAlater (ThermoFisher Scientific, Waltham, MA, USA) and later stored at -80°C for metatranscriptomic analyses. All the specimens were embedded in dry ice and transported to the laboratory at the end of the cruise.

DNA extraction

The slenderness of *S. annulatum* made it difficult to dissect enough intact soft tissues for DNA extraction. Therefore, DNA was extracted from tubeworm individuals along with their tubes. Individual HM_W22 (tube length 50.2 cm) was subjected to DNA extraction using the CTAB method⁹¹ for both Illumina and PacBio sequencing (Figure S10). A batch of mixed individuals was rinsed with 100% ethanol and then dissected into the pale white anterior part (HM_W02) and the brownish posterior part (HM_W03) for creating datasets with different sequencing coverages and facilitating genome binning (Figures S10 and S11). Both samples were further fragmented with sterile scissors and subjected to DNA extraction using the PowerSoil DNA isolation kit (MoBio, Carlsbad, USA) according to the manufacturer's protocol. Genomic DNA was also extracted from ten other individuals (HM_W11–HM_W20) using the PowerSoil DNA isolation kit for relative abundance analysis (Figure S10). DNA purity was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA). DNA quality and quantity were checked using gel electrophoresis and a Qubit 2.0 Fluorometer (Invitrogen, USA).

Metagenome sequencing

Metagenomic libraries of samples HM_W02 and HM_W03 with an insert size of 550 bp were separately constructed following the standard protocol of the Illumina TruSeq Nano DNA Sample Prep Kit (Illumina, USA), and sequenced on an Illumina HiSeq sequencer at Berry Genomics Co. Ltd. (Beijing, China) to produce 150 bp paired-end reads. Metagenomic libraries of the individuals HM_W11 to HM_W20 and HM_W22 with an insert size of 350 bp were constructed following the standard protocol of the NEBNext DNA Library Prep Kit (NEB, USA) and sequenced on an Illumina NovaSeq 6000 sequencer at Novogene Co. Ltd. (Beijing, China) to produce 150 bp paired-end reads. Illumina reads were trimmed using TRIMMOMATIC v0.38⁷² with the following parameters: LEADING = 10, TRAILING = 10, SLIDINGWINDOW = 4:15, and MINLEN = 40. Reads of HM_W22 were further normalized using the bbnorm.sh script of BBmap⁹² with a depth normalization setting of 30X (target = 30). A 20 kb SMRTbell library of individual HM_W22 (named HM_W22P) was also constructed and sequenced on a PacBio Sequel platform at Novogene Co. Ltd. (Beijing, China). PacBio reads were corrected and trimmed using Canu v1.8⁷³ to retain reads $\geq 15,000$ bp. Overall, a total of 13 Illumina, along with one PacBio dataset, were obtained for downstream metagenomics analyses (Table S4).

Metagenome assembly and genome binning

The metagenome assembly strategy is illustrated in Figure S10. Firstly, the filtered and cleaned Illumina and PacBio reads of HM_W22 were co-assembled using SPAdes v3.13⁷⁴ with a *k*-mer set of 21, 33, 55, 77, 99 and 127 under the “–only-assembler” mode. The assembled metagenome was subjected to genome binning using MetaWrap v1.2.2 under the default settings,⁷⁵ with the Illumina datasets HM_W02, HM_W03 and HM_W22 being used in the “Binning” module to provide distinct coverages for the metagenomic

contigs and to improve the binning quality. A high-quality draft genome (binC) was extracted from the co-assembled metagenome. Secondly, PacBio reads were remapped to binC using minimap2,⁷⁶ and the mapped reads were reassembled using the “Miniasm/Minipolish” pipeline⁷⁷ with results visualized using Bandage 0.8.1.⁸⁸ The genome assembly consisted of two long contigs and two variable regions. One of the variable regions is an unclosed gap caused by a fragment with unusually high coverage (region R1), and the other is an unsolved cross-interconnected region (region R2) (Figure S12A). The Illumina reads mapped to R1 were extracted from HM_W22 and reassembled by SPAdes v3.13, which produced a circular phage genome for proHMS1 (Figure 1D). The reads coverage on the phage region is much higher than the coverage on the endosymbiont genome (Figure S8). Alignment of long PacBio reads reveals that proHMS1 also could be integrated into the endosymbiont genome, suggesting that proHMS1 has both the lysogenic and lytic forms, representing an active prophage (Figure S8). When we tried to assemble raw PacBio reads of HM_W22 by the “Miniasm/Minipolish” pipeline, a linear contig could solve the cross interconnection of region R2 (Figure S12B). Therefore, we manually merged the two long contigs (Figure S12A) with the prophage region R1 and the linear region R2 to produce a finished genome for the endosymbiont (HMS1, Figure 1C). The genome was further polished using Pilon v.1.22⁷⁸ with three iterations using the paired-end Illumina short clean reads. The genome start position was determined using the “fixstart” module in Circulator 1.5.5.⁷⁹

Besides, the endosymbiont genome of *S. annulatum* was assembled using SPAdes v3.13 based on Illumina short clean reads of HM_W02 and HM_W03. The assembled metagenome was subjected to genome binning using MetaWrap v1.2.2 under the default settings, which produced the draft genome binI. Assembly graphs produced by SPAdes were explored with Bandage 0.8.1 and 150 nodes with length of longer than 2,000 bp were entirely consistent with the contigs of genome binI (Figure S13). To determine variations among individuals, metagenomic assembly and genome binning were separately carried out using the Illumina short clean reads for HM_W11 to HM_W20. Genome completeness and contamination were evaluated using CheckM v1.1.0.⁸⁰

Genome annotation

All the assembled genomes were automatically annotated using Prokka v1.11 under the default settings.⁸¹ Signal peptides were identified from the Prokka annotation. The predicted amino acid sequences were annotated by using the Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COGs) and PFAM databases. KEGG annotation was performed using the online BlastKOALA program against the species-level prokaryotic KEGG gene database.⁸² COG annotation was performed using the eggNOG-mapper tool against the eggNOG v5.0 database.⁸³ PFAM annotation was performed using the script pfam_scan.pl by searching against the PFAM 32.0 database.⁸⁴ Functional annotations of the three bins (binC, HMS1 and binI) based on KEGG/COG/PFAM annotations produced highly consistent results, indicating the high quality of the finished endosymbiont genome of *S. annulatum* and the reliability of our annotation results (Figure S14 and Tables S21–S23). The vestimentiferan endosymbionts and *Sedimenticola* species used for comparative analyses were re-annotated using the above methods. Principal components analysis (PCA) was performed to cluster the endosymbiont and the free-living genomes based on their KEGG/COG/PFAM annotations according to a method used in our recent study.⁹³ Potential prophages were predicated using the PHAST web server.⁸⁵

Metatranscriptome sequencing and analyses

Three individuals of *S. annulatum* (HM_W26, HM_W27 and HM_W28) were used for total RNA extraction using the Trizol reagent (Invitrogen, USA) following the manufacturer’s protocol. Quality and quantity of the extracted RNA were examined using 1% gel electrophoresis, and a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, USA), respectively. Qualified total RNA was then sent to Novogene Co. Ltd. (Beijing, China) for metatranscriptomic sequencing. Briefly, ribosomal RNA was removed from the total RNA of each individual using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, Ipswich, USA). The resultant mRNA was reverse-transcribed into cDNA and then fragmented into 250 to 300 bp. The constructed libraries were afterward sequenced on an Illumina NovaSeq 6000 sequencer (Illumina, San Diego, USA) to produce 150 bp paired-end reads. Adapters and low-quality reads were removed using Trimmomatic v.0.38⁷² under the following settings: LEADING = 10, TRAILING = 10, SLIDINGWINDOW = 4:15, MINLEN = 40. Clean reads for each individual were then mapped to the endosymbiont genome with BBmap⁹² with a minimum identity of 0.98. Reads counts for symbiont genes in each individual were calculated with CoverM 0.4.0 (<https://github.com/wwood/CoverM>) and normalized with the transcripts per million calculation (TPM) values.⁹⁴

Average TPM values (TPM_{av}) across the three tubeworm individuals were summarized for the estimated gene expression levels.

Phylogenetic analyses

The 16S rRNA gene sequence of the *S. annulatum* endosymbiont was identified using Meta_RNA under the default settings.⁸⁶ For phylogenetic analysis, the 16S rRNA gene sequence was queried against the NCBI GenBank database using BLASTN to recruit sequences of close relatives. These sequences were aligned using MAFFT L-INS-i v7.294b,⁸⁷ and the poorly aligned regions were trimmed using trimAl v1.4 under the -automated1 setting.⁸⁹ A maximum-likelihood (ML) tree was constructed using raxmlGUI v1.5⁹⁰ with the GTRGAMMA model for 1,000 replicates. For phylogenomic analyses, available tubeworm endosymbiont genomes were downloaded from the NCBI genome database. In addition, the 16S rRNA gene sequence of the *S. annulatum* endosymbiont was searched against the NCBI genome database to find the other close relative genomes. Concatenated alignment sequences of 43 conserved proteins were derived from these genomes using CheckM v1.1.0 program⁸⁰ and further treated using trimAl v1.4 under the default settings. An ML tree was built using raxmlGUI v1.5 with the PROTGAMMALG model for 100 replicates. The resulting trees were visualized using the Interactive Tree Of Life (iTOL) v5.⁹⁵ Average nucleotide identity (ANI) between genomes was calculated using pyani (<https://pypi.org/project/pyani/>).

QUANTIFICATION AND STATISTICAL ANALYSIS

Three metatranscriptomic datasets from separated individuals were used to evaluate the consistence of gene expression among biological replicated using the Pearson correlation coefficient (Results and [Figure S4](#)). nMDS ordination of HMS1 and reference genomes was performed based on the relative abundance of their KOs using the Bray-Curtis distance in PRIMER-E. The two-dimensional stress value of 0.049 indicate good fit ([Figure 3A](#)). Other statistical details are provided in the corresponding figure legends.