Sun *et al. Microbiome* (2024) 12:241
https://doi.org/10.1186/s40168-024-01960-4

Microbiome

RESEARCH

Open Access

Integrated multi-approaches reveal unique metabolic mechanisms of Vestimentifera to adapt to deep sea

Qinglei Sun[1,2,3†], Zihao Yuan[1,2,3*†], Yuanyuan Sun[1,2,3] and Li Sun[1,2,3*]

Abstract

Background Vestimentiferan tubeworms are deep-sea colonizers, in which chemoautotrophic symbiosis was first observed. These animals are gutless and depend on endosymbiotic bacteria for organic compound synthesis and nutrition supply. Taxonomically, vestimentiferans belong to Siboglinidae and Annelida. Compared with other siboglinids, vestimentiferans are distinguished by high tolerance of the prevailing hydrogen sulfide in hydrothermal vents, rapid growth in local habitats, and a physical structure consisting of a thick chitinous tube. The metabolic mechanisms contributing to these features remain elusive.

Results Comparative genomics revealed that unlike other annelids, vestimentiferans possessed trehaloneogenesis and lacked gluconeogenesis. Transcriptome and metabolome analyses detected the expression of trehalose-6-phosphate synthase/phosphatase (TPSP), the key enzyme of trehaloneogenesis, and trehalose production in vestimentiferan tissues. In addition to trehaloneogenesis, glycogen biosynthesis evidenced by packed glycogen granules was also found in vestimentiferan symbionts, but not in other Siboglinidae symbionts. Data mining and analyses of invertebrate TPSP revealed that the TPSP in Vestimentifera, as well as Cnidaria, Rotifera, Urochordata, and Cephalochordata, likely originated from Arthropoda, possibly as a result of transposon-mediated inter-phyla gene transfer.

Conclusion This study indicates a critical role of bacterial glycogen biosynthesis in the highly efficient symbiont – vestimentiferan cooperation. This study provides a new perspective for understanding the environmental adaptation strategies of vestimentiferans and adds new insights into the mechanism of metabolic evolution in Metazoa.

Keywords Hydrothermal vents, Annelida, Vestimentiferan, Symbiosis, Comparative genomics, Trehaloneogenesis

Introduction

Since the discovery that the gutless giant tubeworm *Riftia pachyptila* of hydrothermal vents depended on chemosynthetic symbiosis for nutrition [1], deep-sea tubeworms, primarily Siboglinidae, have attracted much research attention around the world. To date, more than 200 species of siboglinids have been described, which belong to four main lineages, i.e., Vestimentifera, *Sclerolinum* (or Monilifera), *Osedax*, and Frenulata, all of which lack a gut and rely on bacterial symbionts for nutrition during adulthood [2, 3]. Of these four groups of Siboglinidae, Vestimentifera is the most studied, as they are ubiquitous in hydrothermal vents and cold seeps habitats,

†Qinglei Sun and Zihao Yuan contributed equally to this work.

*Correspondence:
Zihao Yuan
yuanzihao@qdio.ac.cn
Li Sun
lsun@qdio.ac.cn
[1] CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China
[2] Laboratory for Marine Biology and Biotechnology, Laoshan Laboratory, Qingdao 266200, China
[3] College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China

which have been extensively studied in the past 40 years. Currently, Vestimentifera consists of approximately 19 species, including *Arcovestia ivanovi* and *Lamellibrachia columna* [4, 5]. The adult body of a vestimentiferan consists of an obturacular region, a vestimentum, a trunk, and an opisthosoma. The bacteriocyte-containing trophosome is in the trunk. Recently, several vestimentiferan genomes have been released, and this has increased our understanding of the adaptation mechanisms of vestimentiferans to symbiosis [6–10]. For example, certain genes thought to be involved in symbiont – host interactions appear to be expanded [6]. There is also some evidence of horizontal gene transfer from bacteria to host in vestimentiferan genomes [6].

An accumulating number of studies show that the symbionts in frenulate, Monilifera, and Vestimentifera are generally chemoautotrophic, while the symbionts in *Osedax* are heterotrophic [11]. Currently, the vestimentiferan symbionts are the most studied, mainly in diversity and metabolism. Vestimentiferan symbionts are endosymbionts belonging to Gammaproteobacteria with sulfur-oxidizing potential [5, 12]. No symbiont has been cultured, but more than 10 metagenome-based assembled genomes have been obtained [13–17], which revealed that vestimentiferan symbionts have similar carbon and energy metabolic potentials. All symbionts harbor the Calvin-Benson-Bassham (CBB) cycle and the reductive tricarboxylic acid cycle [14–16]. Endosymbionts also have glycolysis and tricarboxylic acid cycle pathways [18]. Nitrate and oxygen possibly serve as electron acceptors for energy metabolism, while hydrogen, sulfide, and thiosulfate are possible electron donors [6]. The nutrient transport between the symbiont – host has been studied in *R. pachyptila*, with some evidence suggesting that the host acquires nutrients by symbiont digestion/lysis [19, 20], while other evidence has suggested that the symbiont provides the host with nutrients, such as succinate, by secretion [21, 22].

The central carbon metabolism includes glycolysis/glyconeogenesis, the tricarboxylic acid cycle, and the pentose phosphate pathway. In symbiosis, the central carbon metabolism is a key to understanding the carbon flows in microbe, host, and between microbe and host. In vestimentiferan holobionts, the central carbon metabolism of the host (vestimentiferans), especially in association with environmental adaptation, remains to be investigated. In this study, the genome of *A. ivanovi* and the genomes of the symbionts of *A. ivanovi* and *L. columna* were obtained. Comparative genomics was conducted to examine the central carbon metabolisms in both vestimentiferans and their symbionts. Metatranscriptome, transcriptome, metabolome, genetic analysis, and transmission electron microscopy analyses were performed to investigate the metabolic mechanisms and evolution of vestimentiferan holobionts. Our results revealed unique metabolic mechanisms of vestimentiferans and their symbionts and suggested that transposon-mediated gene transfer between different phyla may be the driving force for the evolution of the key metabolic gene involved in the environmental adaptation of vestimentiferans.

Materials and methods

Sample collection

A. ivanovi and *L. columna* individuals were collected using a ROV from the hydrothermal vents of Manus Basin and Iheya North, respectively, in 2015 and 2018, respectively. The collected tubeworms were kept in a biobox for thermal insulation. Once on board, the specimens were immediately washed thoroughly with sterile seawater, and samples from the plume, vestimentum, and trophosome were taken and stored at – 80 °C. The samples for transmission electron microscopy (TEM) analysis were fixed with 2.5% glutaric dialdehyde and stored at 4 °C.

16S rRNA gene amplicon sequencing

The trophosomes of nine *A. ivanovi* individuals and five *L. columna* individuals were used for the 16S rRNA gene amplicon analysis as reported previously [23] with the modification that the amplicon sequence variant (ASV) was calculated using QIIME2 [24]. Total DNA was extracted from the trophosomes of *A. ivanovi* and *L. columna* using the cetyltrimethylammonium bromide (CTAB) method as follows. The sample was lysed using CTAB buffer (NobleRyder, Beijing, China) and lysozyme at 65 °C, followed by centrifugation at 12,000 rpm for 10 min. The supernatant was mixed with an equal volume of phenol (pH 8.0):chloroform:isoamyl alcohol (25:24:1) mix. The mixture was centrifuged as above. The supernatant was combined with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged as above. The supernatant was transferred into a 1.5-mL centrifuge tube and mixed with a 3/4 volume of isopropanol, followed by precipitation at – 20 °C for 20 min. The sample was centrifuged as above, and the precipitate was collected. The precipitate, which contains DNA, was washed twice with 1 mL of 75% ethanol and then dried on a clean bench. The DNA was dissolved in 50-μL ddH_2O. RNase was added into the solution, followed by incubation at 37 °C for 15 min to eliminate any RNA present in the DNA sample. The V3–V4 regions of the 16S rRNA gene were amplified as reported previously [25].

Binning based on metagenome data

The trophosomes of one *A. ivanovi* individual and one *L. columna* individual were used to perform metagenomics.

Total DNA was extracted and purified with the MetaHIT protocol as described previously [26]. The DNA concentration was estimated by Qubit (Thermo Scientific, USA). The library was constructed as reported previously [27]. Briefly. the DNA was fragmented using a Covaris E220 (Covaris, Brighton, UK) to generate fragments of 300 to 700 bp, which were purified using the Axygen™ AxyPrep™ Mag PCR Clean-Up Kit (Axygen Corning Life Sciences, Glendale, USA). An equal volume of magnetic bead was added to the DNA sample, and the DNA was eluted with Tris–EDTA buffer. End repair and A-tailing were performed using T4 DNA polymerase (ENZYMATICS™ P708-1500, Qiagen, MD, USA), T4 polynucleotide kinase (ENZYMATICS™ Y904-1500), and rTaq DNA polymerase (TaKaRa™ R500Z, Dalian, China). Adapters were ligated to the DNA fragments using T4 DNA ligase (ENZYMATICS™ L603-HC-1500) at 23 °C. PCR was then performed, and the PCR products were ligated using T4 DNA ligase to form a single-stranded circular DNA library. The library was sequenced with the BGISEQ-500 protocol [28]. Raw reads were filtered to remove low-quality reads and host genomic DNA by using SOAPnuke [29] and SOAPdenovo [30], respectively, with default parameters. Then the clean data were assembled using metaSPAdes [31] with default parameters, and open reading frames (ORFs) were predicted using MetaGeneMark with default parameters [32]. All predicted genes were filtered by 100-bp length cutoff, and redundancies were removed by CD-HIT (95% identity) [33], resulting in a nonredundant gene catalogue. The gene profile was obtained as reported previously [26]. MetaWRAP was used for binning MAGs with default parameters [34]. CheckM was used for assessing bins contamination and completeness with default parameters [35].

Metatranscriptomic analysis

The trophosomes of two *A. ivanovi* individuals and one *L. columna* individual were used for metatranscriptome sequencing. RNA was extracted with E.Z.N.A.® Soil RNA Kit (Omega Bio-Tek, USA). The cDNA library was generated through a process of double-strand synthesis, followed by end repair and dA-tailing. The mRNA was enriched using VAHTS mRNA Capture Beads (Vazyme, Nanjing, China). The mRNA was then fragmented at 94 °C and used for double-stranded cDNA synthesis with NR604-VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina (Vazyme). The cDNA was subsequently purified and ligated with TruSeq adapters. The cDNA library of each sample was sequenced with the HiSeq 3000 platform (Illumina, CA, USA). Adapters and low-quality reads (base quality ≤ 20) were trimmed with Cutadapt. The clean reads were assembled, and the genes were predicted with MEGAHIT [36] and MetaGeneMark [32] using default parameters. All predicted genes were filtered by 100-bp length cutoff and were removed redundancies by cd-hit (95% identity). Transcript expression levels were quantified in transcript per million (TPM) using kallisto with default parameters [37]. The metagenome and metatranscriptome gene catalogs were searched against NR database, COG database, and KEGG database to obtain the taxonomy and function of the genes using BLAST (version 2.2.28+, http://blast.ncbi.nlm.nih.gov/Blast.cgi, *E*-value ≤ 1e − 5).

Transmission *electron* microscopy (TEM)

For TEM analysis, the sample was fixed as described in the above section of the " Sample collection" and washed three times with 0.1-M PBS at 4 °C. The sample was then fixed with 1% osmium tetroxide (Ted Pella, USA) in PBS for 1.5 h, followed by washing as above. The sample was dehydrated in alcohol (Hushi, China), infiltrated with acetone (Tieta, China) and epoxy resin (SPI-CHEM, USA) mixture, and embedded and polymerized in epoxy resin. Ultrathin sections were obtained with a Leica EM UC7 ultramicrotome (Leica Microsystems, Germany) and transferred onto copper grids covered with the formvar membrane (Electron Microscopy China, China). Uranyl acetate (2%) and lead citrate (Ted Pella Inc., USA) were used for contrast staining. The sections were photographed with a transmission electron microscope (HT7700, Hitachi, Japan).

Genome and transcriptome sequencing of A. ivanovi

Genomic DNA from one *A. ivanovi* individual was isolated using the phenol–chloroform method as described above for the CTAB method, except that the sodium dodecyl sulfate (SDS) buffer was used instead of the CTAB buffer. DNA degradation and contamination were monitored by electrophoresis in 1% agarose gels and pulsed-field gel electrophoresis. DNA purity was checked using a NanoPhotometer spectrophotometer (Implen, USA). DNA concentration was measured using the Qubit DNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, USA). The partial mitochondrial cytochrome c oxidase I (COI) gene was amplified by PCR using the primer pair COI L1490/COI H2198 [38]. The PCR reaction was carried out in a 30-μL reaction volume with 15-μL PrimeSTAR Max Premix (TaKaRa, Dalian, China), 0.2-μM forward and reverse primers, and 10-ng template DNA. Thermal cycling consisted of denaturation at 98 °C for 5 min, followed by 30 cycles of 98 °C for 10 s, 50 °C for 15 s, and 72 °C for 20 s, and finally 72 °C for 5 min. The PCR product was purified and cloned into pEASY-T1 Simple vector (TransGen, Beijing, China). DNA sequencing was carried out by Tsingke Biotech (Beijing, China). The sequences of other metazoans were collected from

published data and Coarbitrator_COI_nuc.fa [39]. The phylogenetic analysis was performed using IQ-TREE 2 with 1000 bootstrap and best-fit model GTR + F + I + G4. Genome sequencing, assembly, annotation, and transcriptome were performed in Novogene Bioinformatics Institute (Beijing, China). Both the Illumina HiSeq platform and the PacBio Sequel were used for the sequencing. For the Illumina HiSeq sequencing platform, 1.5-μg DNA per sample was used as the input material. Sequencing libraries were generated using NEBNext Ultra DNA Library Prep Kit (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The short but accurate reads from the Illumina platform were analyzed for genome survey and base-level correction after the long-reads assembly (Table S1). For the PacBio platform, genomic sequencing libraries were constructed according to the protocol recommended by PacBio. Long reads generated from the PacBio platform were used for genome assembly (Table S1). RNA was extracted from the plume of six A. ivanovi individuals using RNAiso Pure RNA Isolation Kit (Takara, Japan) and assessed for quality using NanoVue Plus spectrophotometer (GE Healthcare, USA). RNA-seq libraries were constructed as reported previously [40] and sequenced with Illumina HiSeq in paired-end 150-bp mode.

The genome size of *A. ivanovi* was estimated using the Illumina sequencing data with the Kmer-based method [41]. The genome was assembled using the FALCON genome assembler with parameter set as "overlap_filtering_setting = –max_diff 500 –max_cov 500 –min_cov 3 –n_core 24 –bestn 10" [42]. The assembled sequence was polished using Quiver (SMRT Analysis v2.3.0) [43] with default parameters. The completeness of the genome assembly was evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO) [44]. The accuracy of the genome was evaluated using the Illumina short reads mapping with BWA [45], with the parameters set as "mem -k 32 -w 10 -B 3 -O 11 -E 4 -t 16." Additionally, the transcriptome reads were also mapped to the genome assembly using BLAT [46] with default parameters.

Repetitive element annotation

The repetitive elements were identified via RepeatModeler 1.0.8 containing RECON [47] and RepeatScout with default parameters [48]. The derived repetitive elements were searched against Repbase by using RepeatMasker-4.1.2-p1 (https://www.repeatmasker.org) [49]. The average number of substitution per site (K) for repeat unit was subtotaled. The *K*-value was calculated based on the Jukes-Cantor formula: $K = -300/4 \times \mathrm{Ln}(1\text{-}D \times 4/300)$, with D representing the proportion of each repeat unit differing from the consensus sequences (Mouse Genome Sequencing Consortium, 2002).

Protein-coding gene prediction and function annotation

Structural annotation of the genome incorporates homology-based prediction, ab initio prediction, and RNA-Seq assisted prediction. For homogenous comparison prediction, the genome information of *Caenorhabditis elegans* (GCA_000002985.3), *Capitella teleta* (GCA_000328365.1), *Drosophila melanogaster* (GCA_000001215.4), *Helobdella robusta* (GCA_000326865.1), *Lingula anatine* (GCA_001039355.2), and *Octopus bimaculoides* (GCA_001194135.1) was obtained from the Ensembl database and NCBI database. The gene structures were predicted with TBLASTN [50] (E-value ≤ 1e−5) and GeneWise [51]. For gene prediction based on ab initio, AUGUSTUS [52], GeneID [53], Genscan [54], GlimmerHMM [55], and SNAP [56] were used with default parameters. To facilitate genome annotation, transcriptome reads assemblies were generated using Trinity with parameters set as "–normalize_reads –full_cleanup –min_glue 2 –min_kmer_cov 2" [57]. To optimize the genome annotation, the RNA-Seq reads were aligned to genome sequence using Hisat2 [58]/TopHat [59] with default parameters to identify exon region and splice positions. The alignment results were then used as input for StringTie [60]/Cufflinks [61] with default parameters for genome-based transcript assembly. The nonredundant reference gene set was generated by merging genes predicted by three methods with EvidenceModeler using PASA (Program to Assemble Spliced Alignment) [62]. Functional annotation of protein coding genes was evaluated with BLASTP (e-value 1e-05). The annotation information of the best BLAST hit, which was derived from database, was transferred to our gene set. Protein domains were annotated by searching SwissProt, InterPro, KEGG, and Pfam databases.

Gene expression and metabolomics analyses

The RNA-seq reads of tubeworms were processed and quality controlled with Trimmomatic to remove adapters and trim low-quality bases and reads shorter than 30 bases using the parameters ILLUMINACLIP: adapters.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:30 [63]. All trimmed reads were first mapped to the tubeworm genomes via HISAT2 [58] and HTSeq [64], and the final reads were processed with DESeq2 [65] with adjusted p-value cutoff of 0.05. The expression was plotted in Prism for visualization. Metabolomic analyses were performed with the trophosomes and vestimentums of three tubeworms as detailed in Supplemental data—method.

Invertebrate TPSP data-mining and Alien Index analysis

A total of 289 TPSP sequences were used as queries. The queries were searched against the NCBI RefSeq

database in 32 invertebrate phyla via BLASTP with an E-value of 1e-5 to ensure, in part, accuracy. The sequences were manually curated using conserved domain database (https://www.ncbi.nlm.nih.gov/cdd/). The putative TPSP homologs were further de-duplicated based on their genomic locations. Alien Index analysis was performed as reported previously [66], and Alien Index = bbhO/maxB–bbhG/maxB. BbhO, bbhG, and maxB are described in Table S2.

Results

Identification and characterization of hydrothermal vent vestimentiferans and their symbionts

Hydrothermal vent vestimentiferans *A. ivanovi* and *L. columna* were identified based on the COI gene analysis, which showed that they shared high identities with *A. ivanovi* (99.40%) and *L. columna* (99.06%), respectively, and were phylogenetically clustered together with the *A. ivanovi* and *L. columna*, respectively, of Vestimentifera in the Siboglinidae family (Fig. 1A). Since vestimentiferans rely on symbiotic bacteria for nutrition, we examined the symbiont community structures of *A. ivanovi* and *L. columna*. Fourteen 16S rRNA V3 – V4 region amplicon libraries were constructed from the trophosomes of nine *A. ivanovi* individuals and five *L. columna* individuals. A community structure analysis showed that 98.5 – 99.8% and 59.4 – 87.7% tags of the *A. ivanovi* and *L. columna* symbiont libraries, respectively, were annotated to be of the family Sedimenticolaceae (Fig. S1A). For all 14 of the libraries, most tags were unclassified at the genus level; however, SUP 05 (0.2 – 19.1%) and *Sulfurovum* (1.0 – 13.8%) were relatively abundant in the *A. ivanovi* symbiont libraries (Fig. S1B). All of the libraries were mainly composed of one ASV. ASV 42296 was

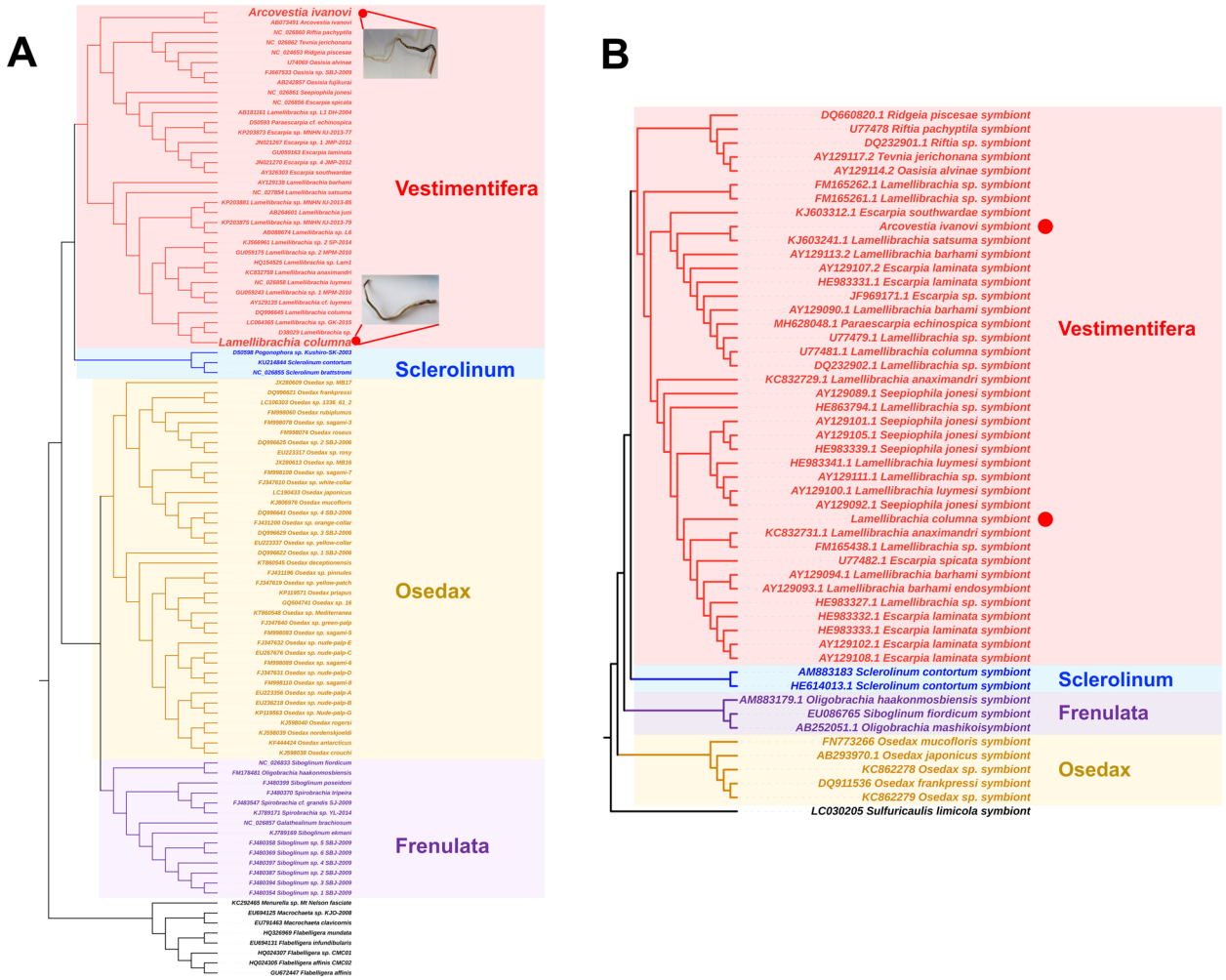


Fig. 1 The phylogenetic analysis of siboglinids and their symbionts. **A** The COI gene-based phylogenetic tree of siboglinids constructed with IQ-TREE 2. **B** The 16S rRNA gene-based phylogenetic tree of siboglinid symbionts constructed as above. The red dots in both panels indicate the *A. ivanovi* and *L. columna* from this study, and the photographs of these two species of tubeworms are shown in **A**

most abundant in the *A. ivanovi* symbiont libraries, while ASV 24308 was most abundant in the *L. columna* symbiont libraries (Fig. S1C). These results indicated that *Arcovestia* and *Lamellibrachia* each possessed primarily one ribotype of symbiont. The trophosomes of *A. ivanovi* and *L. columna* were each used to construct a metagenomic library. Only one nearly complete 16S rRNA gene (>1500 bp) of the symbiont was obtained from every library. The 16S rRNA gene sequences from the two libraries shared a 98.4% identity and were clustered within the clade of vestimentiferan symbionts but segregated into different groups (Fig. 1B). Only one high-quality metagenome-assembled genome (MAG) was obtained and used for further analysis for both *A. ivanovi* and *L. columna*.

Metabolic features of the vestimentiferan symbionts revealed by meta-omics and *electron* microscopy

Genome analysis based on the above obtained MAGs indicated that both *A. ivanovi* and *L. columna* symbionts harbored all of the genes of the CBB cycle, except for fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase, which were absent. However, this may be functionally substituted by ATP-dependent phosphofructokinase (K21071) in the CBB cycle and gluconeogenesis [67, 68] (Fig. S1D). Metatranscriptome analysis indicated that the key gene of the CBB cycle, ribulose-bisphosphate carboxylase, was expressed at very high levels in the *A. ivanovi* and *L. columna* symbionts, with the TPM value ranging from 81.11 to 1986.90 (top 1%) (Table S3), suggesting that the CBB cycle was highly active. The *A. ivanovi* and *L. columna* symbionts also possessed all of the genes of the rTCA cycle (Fig. S1D). In addition, ATP-dependent citrate lyase, the key gene of the rTCA cycle, was expressed at a high level (TPM of 14.06 – 25.28) (top 5%) (Table S3), suggesting that the rTCA cycle was active as well. In both the *A. ivanovi* and *L. columna* symbionts, the gluconeogenesis pathway was not complete, and glucose-6-phosphatase, which transforms glucose-6-phosphate to glucose, was absent (Fig. S1D). However, the glycogen biosynthesis pathway was complete, and glucose-6-phosphate could be transformed into glycogen (Fig. S1D). The metatranscriptome indicated that all of the genes of glycogen biosynthesis were expressed. In particular, glucose-1-phosphate adenylyltransferase was expressed at a high level (TPM of 15.51 – 26.28) (top 5%) in both the *A. ivanovi* and *L. columna* symbionts (Table S3). Consistently, glycogen was detected in the holobiont by TEM and metabolome analysis. TEM also revealed lobule-like structures rich in bacteria in the trophosome (Fig. 2A). Depending on their locations in the lobule, the symbiotic bacteria exhibited different morphologies. In the center of the lobule, the symbionts were small, with diameters of 1 – 2 μm, and some bacteria were apparently in the stage of replication (Fig. 2A). In the periphery, the symbionts were relatively large (2 – 4 μm in diameter) and filled with dark glycogen granules (Fig. 2A). Some glycogen-packed cells appeared to undergo a process of lysis accompanied by glycogen release, and the cells eventually broke apart (Fig. 2B, C). A comparative genomic analysis indicated that the glycogen biosynthesis pathway was present in most vestimentiferan symbionts, but it was absent in the frenulate and *Osedax* symbionts (Fig. S1E). In addition to the glycogen biosynthesis pathway, the complete riboflavin biosynthesis pathway was also present in the *A. ivanovi* and *L. columna* symbionts, and all of the genes of this pathway were expressed (Table S3). Further, the riboflavin level in the trophosome was significantly and strikingly higher (1247-fold) than that in the vestimentum (Table S4).

Assembly and annotation of the A. ivanovi genome

In order to explore the host-symbiont cooperation in metabolism, the genome of *A. ivanovi* was sequenced and assembled. Prior to assembly, the genome size of *A. ivanovi* was estimated to be approximately 853.7 Mb, with a heterozygous rate of 0.89%. The final assembled genome was 792.8 Mb and contained 9469 contigs, with an N50 of 571.01 kb (Table S5). A BUSCO assessment against the eukaryotes conserved core genes showed 100% completeness (249 complete, 6 fragmented, and no missing). The completeness of the genome was validated by a mapping rate of 98.1%, based on aligning the Illumina short reads against the assembly. All together, a total of 17,904 protein-coding genes were predicted in the genome, and 16,140 genes were annotated (Table S6).

The repeatome landscape of A. ivanovi

Because repetitive elements comprise a large proportion of eukaryote genomes, we comprehensively analyzed the repeatomes of the tubeworms covering three families (Table S7). The *A. ivanovi* genome contained 477,439,658 bp (60.2% of the genome) repetitive element sequences that were primarily class I transposable elements (124,047,905 bp), which accounted for 15.7% of the genome. The abundance of the class I transposable elements in *A. ivanovi* was comparable to that in other Siboglinidae species (14.4% in *Lamellibrachia luymesi* and 22.6% in *Paraescarpia echinospica*) but much higher than that in non-Siboglinidae species (5.6% in *Hydroides elegans* and 4.1% in *Owenia fusiformis*). Of the class I transposons, the long interspersed nuclear element (LINE) was markedly expanded in *A. ivanovi* (75.9% of the total transposable elements) as well as two other Siboglinidae species (71.3% and 61.7%, respectively, of the total transposable elements) and Serpulidae (59.5%

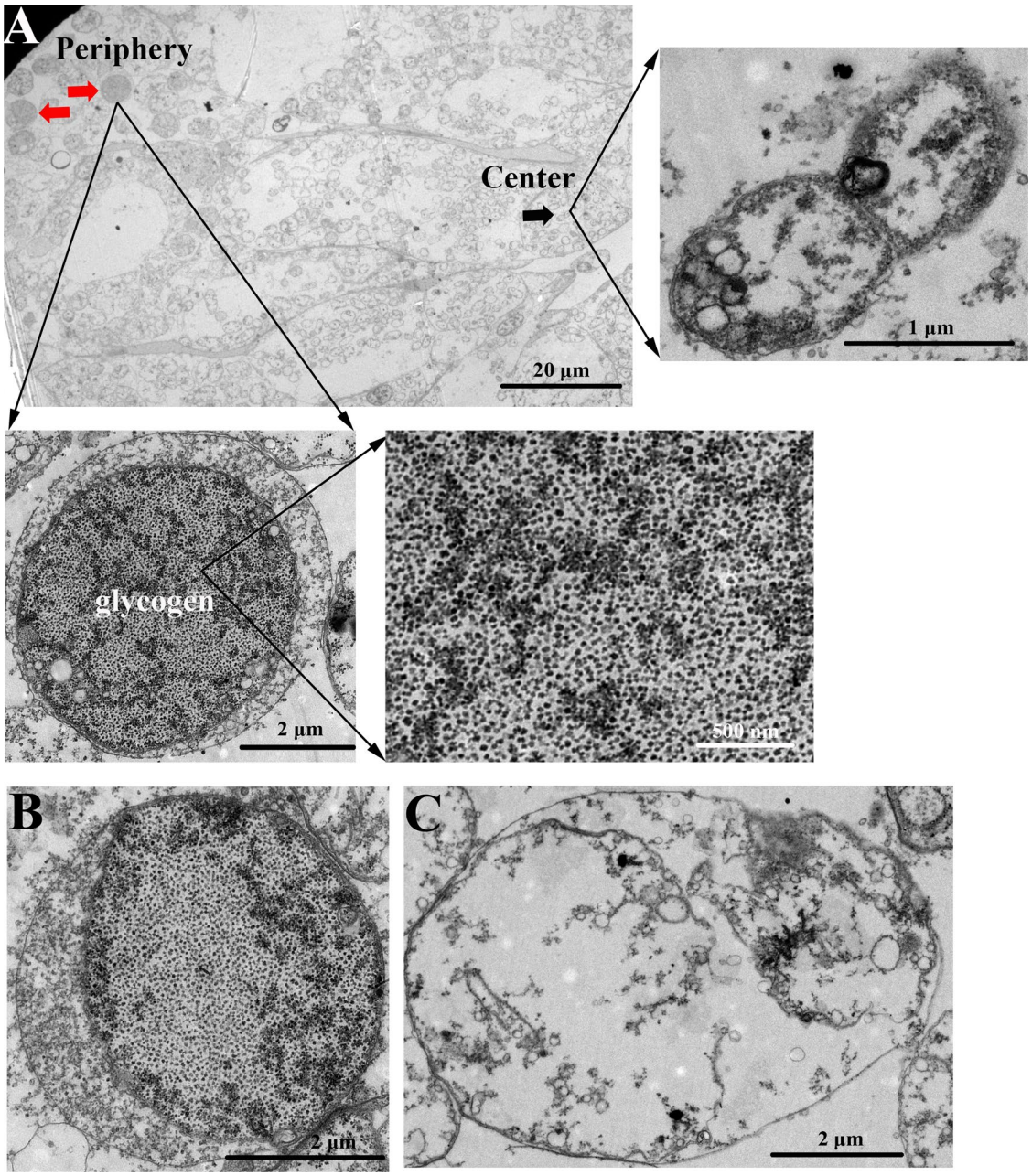


Fig. 2 The symbionts in the *L. columna* trophosome observed with a transmission electron microscope. **A** The overall lobule-like structure. The black and red arrows indicate the smaller bacteria in the center and the larger bacteria in the periphery, respectively. The images of a dividing bacterium, a glycogen-filled bacterium, and the glycogen granules inside the bacterium are enlarged on the right, left bottom, and right bottom, respectively. **B** The symbiont in the process of releasing glycogen-containing cellular contents. **C** The representative of a symbiont with disintegrated structure after emptying glycogen contents

of the total transposable elements), but it made up only 36.2% of the total transposable elements in Oweniidae (Fig. 3A). The Siboglinidae species were rich especially in LINE/CR1 and LINE/L2. Further, the LINE in *A. ivanovi* was not only abundant but also highly consensus, as reflected by the low substitution rates (Fig. 3B).

The unique metabolic pathways in vestimentiferans

Currently, little is known about the central carbon metabolisms in vestimentiferans. To explore the carbon metabolism pathways, comparative genomics was conducted, and this showed that the glycolysis pathway, tricarboxylic acid cycle, and pentose phosphate pathway

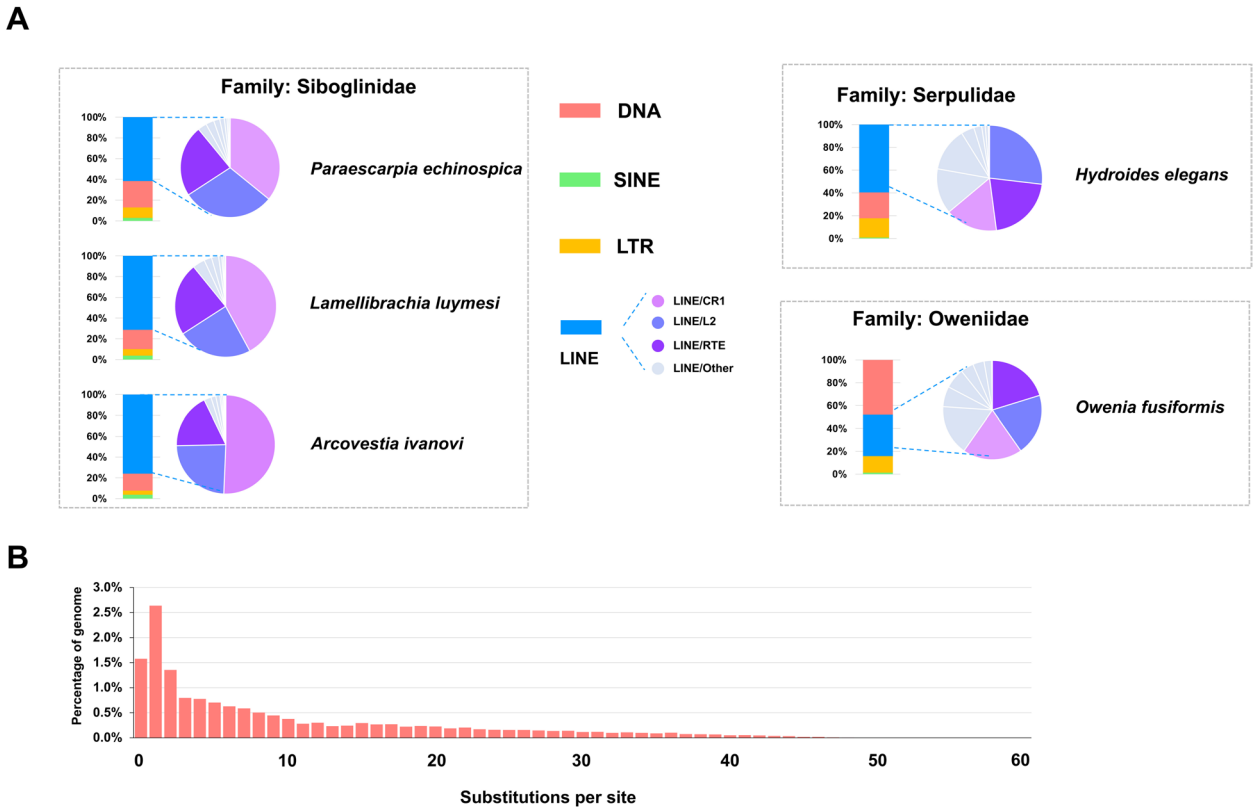


Fig. 3 Analysis of the repetitive elements in tubeworms. **A** The abundance of the repetitive elements in Siboglinidae compared with that in Serpulidae and Oweniidae. **B** The divergence of LINE/CR1 in *A. ivanovi*. The *y*-axis represents the percentage of the genome comprised of repeat classes (%), and the *x*-axis represents the substitution rate from consensus sequences (%)

were complete in all of the analyzed Polychaeta genomes (Table S8). However, the sugar-synthesizing pathways differed markedly between the symbiotic and nonsymbiotic Polychaeta, as well as between the different groups of Siboglinidae. All of the nonsymbiotic Polychaeta possessed complete gluconeogenesis pathway but lacked trehalose-6-phosphate synthase/phosphatase (TPSP), the key enzyme for trehalose synthesis, and thus were probably unable to synthesize trehalose (Fig. 4A). In contrast, all of the vestimentiferans lacked glucose-6-phosphatase, the key enzyme for glucose synthesis, but possessed TPSP, and therefore were able to synthesize trehalose (Fig. 4A, B). Surprisingly, *Osedax*, another member of Siboglinidae, possessed neither TPSP nor glucose-6-phosphatase (Fig. 4A). Transcriptome analysis indicated that TPSP was expressed in vestimentiferan tissues (Fig. 4C, Fig. S2). Comparing with vestimentum, trophosome and plume exhibited significantly higher levels of TPSP expression (Fig. S2). In agreement, the metabolome detected trehalose in the vestimentiferan trophosome and vestimentum, especially the former (Table S4, Fig. 4D). In contrast, although glucose is an important source of energy, it was not detected in the vestimentiferan metabolome.

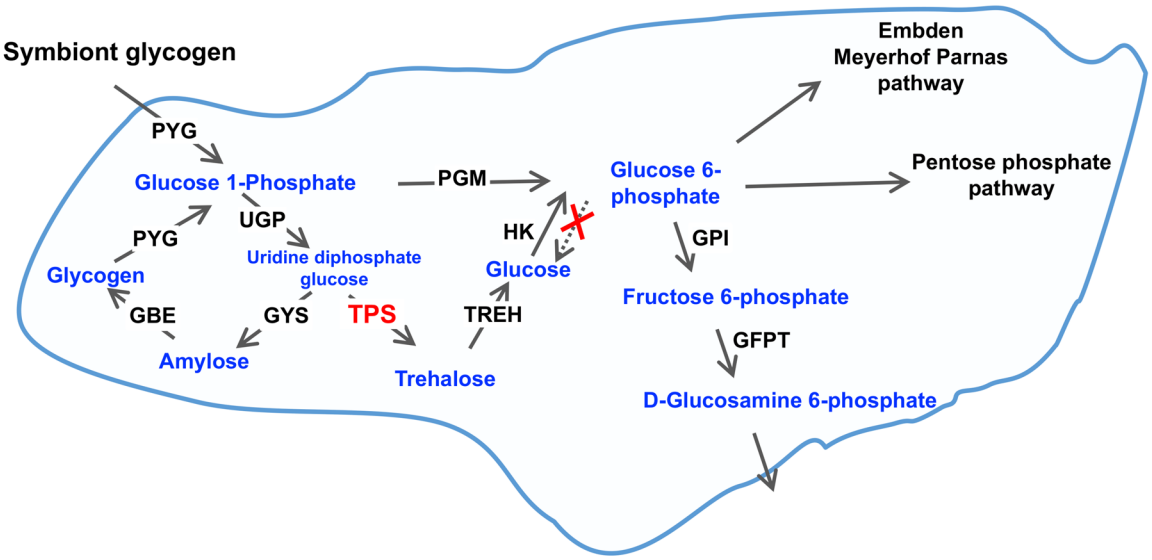
The origin and transfer of *Vestimentifera* TPSP

Sequence analysis showed that the TPSP of five Siboglinidae species shared high similarities with each other (87.8 – 96.1%) and with the TPSP of Arthropoda (58.5 – 73.1% in the top 10 alignments). Since in Annelida TPSP is present only in Vestimentifera, we examined whether the vestimentiferan TPSP was of a foreign source via an Alien Index analysis. The results showed that the vestimentiferan TPSP scored the highest index with Arthropoda (0.54 – 0.59) (Table S2), suggesting that the vestimentiferan TPSP may have an Arthropoda origin. Like arthropods, in which TPSP is a fused protein consisting of trehalose-6-phosphate synthase (TPS) and trehalose-6-phosphate phosphatase (TPP), the vestimentiferan TPSP is also a fused protein (Fig. 5A). In order to explore the evolutionary dynamics of TPSP, a comprehensive mining of TPSP in invertebrate phyla was conducted. A total of 963 TPSP were identified from 6 phyla (Table S9), including Rotifera and Urochordata, in

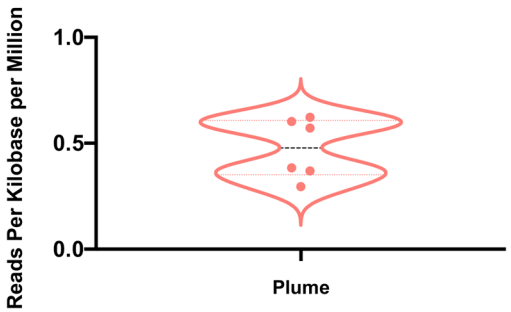
A

	Non-symbiotic Polychaeta				Symbiotic Polychaeta (Siboglinidae family)					
					Osedax	Vestimentifera				
	Capitella teleta	*Dimorphilus gyrociliatus*	*Owenia fusiformis*	*Hydroides elegans*	*Osedax frankpressi*	*Paraescarpia echinospica*	*Lamellibrachia luymesi*	*Arcovestia ivanovi*	*Oasisia alvinae*	*Riftia pachyptila*
Glucose-6-phosphatase (G6P)	+	+	+	+	−	−	−	−	−	−
Trehalose 6-phosphate synthase (TPS)	−	−	−	−	−	+	+	+	+	+

B



C



D

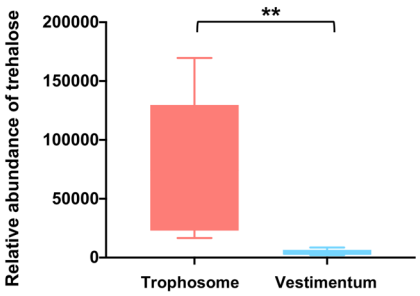


Fig. 4 The trehalose biosynthesis in Vestimentifera. **A** The distribution of trehalose-6-phosphate synthase/phosphatase (TPSP) and glucose-6-phosphatase (G6P) genes in symbiotic and nonsymbiotic Polychaeta. **B** The proposed carbon flow in Vestimentifera. The abbreviations of the enzymes are as follows: PYG, glycogen phosphorylase; UGP, UTP-glucose-1-phosphate uridylyltransferase; GYS, glycogen synthase; GBE, 1,4-alpha-glucan-branching enzyme; PGM, phosphoglucomutase; TREH, trehalase; HK, hexokinase; GPI, glucose-6-phosphate isomerase; GFPT1, glutamine-fructose-6-phosphate aminotransferase. **C** The expression of the TPSP gene in *A. ivanovi* plume. The *y*-axis represents the expression in reads per kilobase per million. **D** The relative abundance of trehalose in vestimentiferan tissues detected by metabolome
$^{**}p < 0.01$

which no TPSP had been reported previously. Of these six phyla, Cnidaria and Cephalochordata lacked the fused form of TPSP but possessed truncated TPSP with only the TPS domain. Phylogenetic analysis based on TPSP revealed that metazoan TPSP formed two major clades, one of which contained only Nematoda, and the other consisted of predominately Arthropoda (Fig. 5B). All of the Vestimentifera, Rotifera, and Urochordata TPSP fell into the Arthropoda clade (Fig. 5B). Similar results were obtained in phylogenetic analysis based on the TPS domain, which showed that the truncated TPSP (i.e., containing only the TPS domain) of Cnidaria and Cephalochordata also fell into the Arthropoda clade (Figure S3). These results indicated that the TPSP in Vestimentifera as well as Cnidaria, Rotifera, Cephalochordata, and Urochordata were probably of Arthropoda origin. Because transposons are common mediators of gene transfer, we examined the existence of the transposable elements surrounding TPSP. High densities of transposable elements, dominated by LINEs, were found adjacent (500-K base pairs) to the TPSP loci of *A. ivanovi* as well as other vestimentiferans tubeworms (Fig. 5C). These results suggested that transposable elements may have driven the horizontal transfer of TPSP.

Discussion

Vestimentiferans in hydrothermal vents can grow at an exceedingly high rate. For example, *R. pachyptila* and *Tevnia jerichonana* grow > 85 cm and > 30 cm, respectively, in tube length per year [69]. Vestimentiferans form very dense communities and accumulate a large amount of biomass, suggesting that a vestimentiferan – symbiont partnership is a highly successful strategy for adaptation to the vent surroundings. Because symbionts provide the host vestimentiferans with abundant carbon compounds, the carbon metabolism of the symbionts plays an important role in the holobiont system. Glycogen, an essential carbon reserve for many organisms, has been detected in vestimentiferan symbionts and hosts via TEM and chemical analyses [20, 70]. However, the glycogen biosynthesis pathway at the molecular level remains to be investigated. In this study, a genome analysis indicated that the glycogen biosynthesis pathway was conserved in most vestimentiferan symbionts but absent in the frenulate and *Osedax* symbionts. The presence of glycogen biosynthesis in vestimentiferans was verified by transcriptome, metabolome, and TEM analyses. The TEM showed that the symbionts located in the lobule periphery accumulated much more glycogen than the symbionts in the lobule center, which was similar to the observation in *R. pachyptila*, another Vestimentifera member [70, 71]. In our study, broken bacterial cells and glycogen release were observed in the periphery of the lobule. These results suggested that glycogen was an important carbon reserve in vestimentiferan symbionts, and that, after lysis/digestion of the symbionts, glycogen could be supplied to the host as a vital nutrient. Hence, glycogen biosynthesis is likely critical for the establishment and maintenance of the highly efficient symbiont – vestimentiferan cooperation. It is interesting that in addition to glycogen biosynthesis, this study also found that riboflavin biosynthesis was active in both *A. ivanovi* and *L. columna*. Consistently, riboflavin was detected for the first time in Vestimentifera. Riboflavin, also called vitamin B2, is essential for all organisms; however, it is only known to be produced by plants and microbes. Its detection in Vestimentifera suggested that in addition to glycogen, riboflavin may be another important nutrient supplied by the symbionts to the host vestimentiferans. In a previous report, riboflavin was shown to stimulate metamorphosis in the larvae of polychaete *C. teleta* [72]. It is possible that the riboflavin produced by the vestimentiferan symbionts observed in our study may be involved in the metamorphosis, growth, and development of the host tubeworms.

Because sugar synthesis is critical for the survival of all animals, especially under conditions of nutrient deficiency, sugar-synthesizing pathways are expected to exist in all metazoan species [73]. Gluconeogenesis is important for mammals and fish, because glucose, the product of this pathway, is utilized as blood sugar [73]. In contrast, trehaloneogenesis is important for many insects, because trehalose (a nonreducing disaccharide comprised of two glucose molecules), the product of trehaloneogenesis, is the principal hemolymph sugar of insects [74]. In some insects, such as the fruit fly, both trehaloneogenesis and gluconeogenesis are present, with the latter participating in neuronal signaling [75]. To date, the sugar-synthesizing pathways in Annelida are largely unknown. In this study, we unexpectedly found that unlike nonsymbiotic Polychaeta, vestimentiferans possessed only trehaloneogenesis and lacked gluconeogenesis, suggesting that trehalose, rather than glucose, was likely a vital energy source for vestimentiferans. Consistently, trehalose was detected in vestimentiferan organs, especially the trophosome. Because trophosome is the site of hematopoiesis and rich in blood vessels [4, 9], trehalose may be the principal blood sugar of vestimentiferans. Trehalose is also known as a cryoprotectant and a heat bioprotectant for cells and proteins [76]. This property may enable trehalose to confer a survival advantage upon vestimentiferans. Previous reports have shown that the vestimentiferan, *Riftia*, grew optimally at 25 °C but could not tolerate prolonged incubation in higher

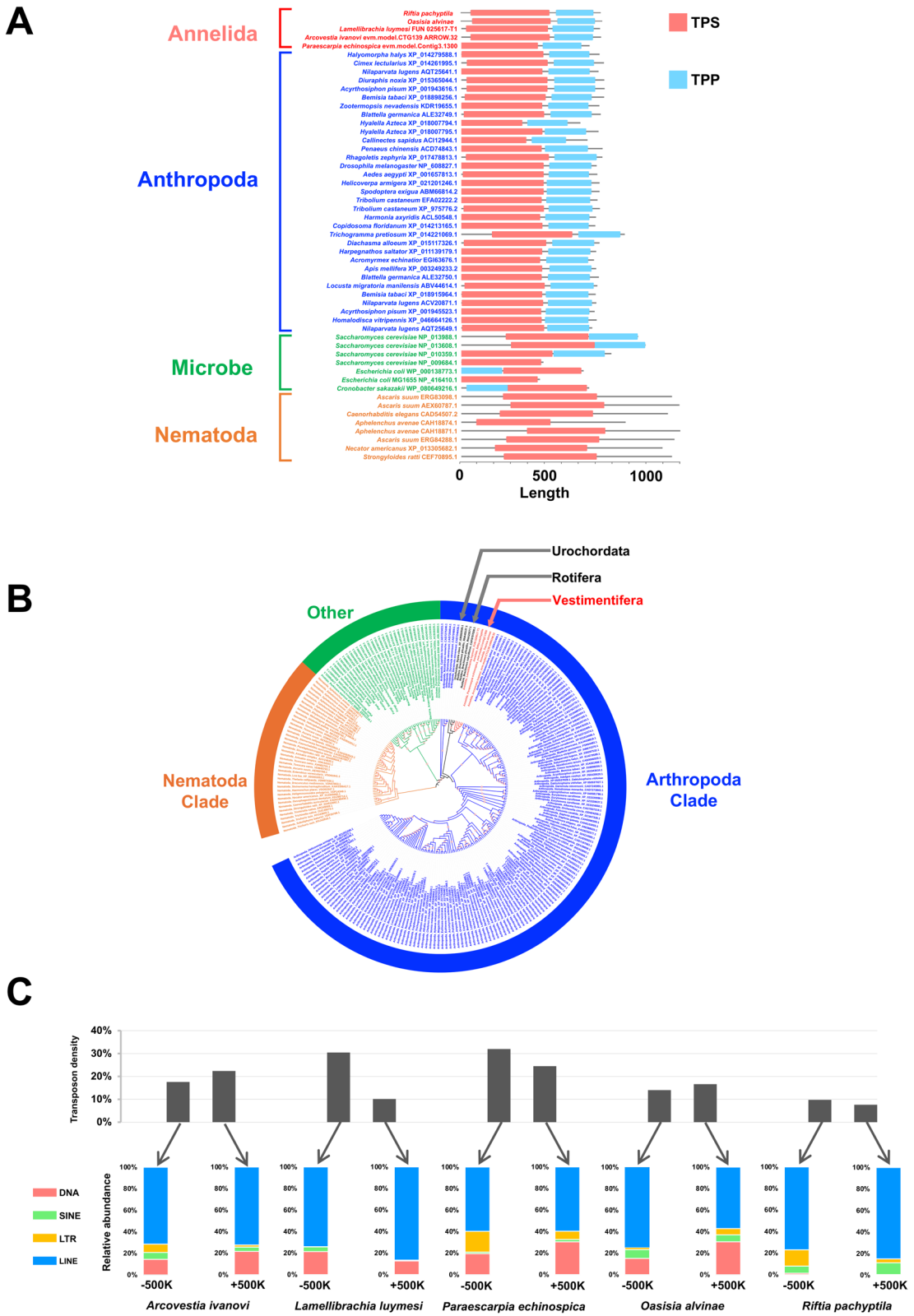


Fig. 5 The evolutionary analysis of TPSP. **A** The comparison of the conserved domains of TPSP from Annelida, Arthropoda, Nematoda, and microbes. **B** The phylogenetic analysis of the TPSP from Annelida, Arthropoda, Rotifera, Nematoda, Arthropoda, and others (archaea, bacteria, fungi, plants, and protists). **C** The comparison of the transposons, including DNA transposons (DNA), long terminal repeat (LTR), long interspersed nuclear element (LINE), and short interspersed nuclear element (SINE), adjacent to the TPSP loci (± 500-K base pairs) in vestimentiferan genomes

temperatures of 32 to 35 °C [77]. Because in hydrothermal vents the distribution of H_2S tends to be more abundant at regions nearer the vent (and hence higher temperatures), vestimentiferans are inevitably exposed to lethal high temperatures in order to obtain the sulfide required to sustain the chemoautotrophic metabolism of their symbionts [77]. Therefore, trehalose, as a heat protectant, may function to shield the vestimentiferans from the harm of high temperature in the vents. Trehalose has also been reported to participate in H_2S resistance in maize [78]. It is possible that in addition to contributing to heat resistance, trehalose may play a role in the sulfide tolerance of vestimentiferans. Further, in insects, trehalose is known to be an essential substrate for chitin synthesis, which is directly affected by TPSP [79]. Similar to insects, vestimentiferans, which are characterized by conspicuous chitinous tubes, require chitin synthesis. Hence, the acquisition of TPSP probably endows vestimentiferans with growth and survival advantages that are conducive to the flourishing of vestimentiferans in deep-sea hydrothermal vents. These results support trehaloneogenesis as a critical mechanism for vestimentiferans to adapt to the specific conditions of hydrothermal vents.

In this study, TPSP was detected only in vestimentiferans and not in other annelids, which raised the question of the origin of the vestimentiferan TPSP. Using multiple approaches, we obtained strong evidence indicating that the TPSP of Vestimentifera, as well as four other phyla, likely had an origin in Arthropoda, from which TPSP may have spread out via transposon-mediated horizontal gene transfer (HGT). HGT is known to widely occur from bacteria to bacteria and from bacteria to eukaryotes [80]. A few HGT events between animal species have also been proposed, such as the transfer of the AFP gene in fish [81, 82]. In our study, we found that the inter-phyla transmission of TPSP likely occurred from Arthropoda to not only Vestimentifera but also Cnidaria, Rotifera, Urochordata, and Cephalochordata. In addition to these phyla, potential HGT from Arthropoda to Nematoda (*Pristionchus pacificus*) has also been reported [83]. These observations favor the notion that Arthropoda might be an important source of horizontally transferred genes in Metazoa. Previous studies have indicated that of the different types of transposons, LINE was a significant mediator of HGT [80, 81]. In Vestimentifera, abundant LINEs were identified surrounding the TPSP loci, thus suggesting that these transposable elements may have played a vital role in HGT. Taken together, these results support the idea that transposon-mediated inter-phyla HGT represents an evolutionary shortcut for a fast and wide dissemination of certain trait-conferring genes, in particular those, such as TPSP, that are beneficial to the hosts for surviving specific conditions, across the Metazoa phylogeny.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01960-4.

Supplementary Material 1: Supplemental data-mMethod. Metabolomic analysis Supplemental data. -fFigures. Figure S1. The diversity and central carbon metabolism pathways of the *Arcovestia ivanovi* and *Lamellibrachia columna* symbionts. (A, B, C) The sequence tags are classified at the family (A), genus (B), and ASV (C) levels. Each color represents the percentage of the taxon in the total assemblage. A-1 to -9 represent *A. ivanovi* individuals 1 to 9; L-1 to -5 represent *L. columna* individuals 1 to 5. (D) Overview of the central carbon metabolisms in *A. ivanovi* and *L. columna*symbionts. G, glucose; G-6-P, D-glucose 6-phosphate; F-6-P, D-fructose 6-phosphate; F-1,6-P, Dfructose 1,6-bisphosphate; PGAL, D-glyceraldehyde 3-phosphate; DHAP, glycerone phosphate; 1,3-DPG, 3-phospho-D-glyceroyl phosphate; 3-PGA, 3-phospho-Dglycerate; 2-PGA, 2-phospho-D-glycerate; PEP, phospho-enolpyruvate; Pyru, pyruvate; Ace-CoA, acetyl-CoA; Rbu-1,5-P, D-ribulose 1,5-bisphosphate; Rbu-5-P, D-ribulose 5phosphate; Rbo-5-P, D-ribose 5-phosphate; Sedo-7P, sedoheptulose 7-phosphate; Sedo1,7-P, sedoheptulose 1,7-bisphosphate; Ery-4P, D-erythrose 4-phosphate; G-1-P, Dglucose 1-phosphate; ADP-G, ADP-glucose; Amy, amylose; Glyco, glycongen; Oxalo, oxaloacetate; Mal, malate; Fum, fumarate; Suc, succinate; Suc-CoA, succinyl-CoA; 2oxog, 2-oxoglutarate; Isocit, isocitrate; Aco, aconitate; Cit, citrate; 1, ribulosebisphosphate carboxylase; 2, phosphoglycerate kinase; 3, glyceraldehyde 3-phosphate dehydrogenase ; 4, fructose-bisphosphate aldolase; 5, ATP-dependent phosphofructokinase; 6, transketolase; 7, triosephosphate isomerase; 8, fructosebisphosphate aldolase; 9, ATP-dependent phosphofructokinase; 10, transketolase; 11, ribose 5-phosphate isomerase A; 12, ribulose-phosphate 3-epimerase; 13, phosphoribulokinase; 14, 2,3-bisphosphoglycerate-independent phosphoglycerate mutase; 15, enolase; 16, pyruvate, water dikinase; 17, pyruvate ferredoxin oxidoreductase; 18, citrate synthase; 19, malate dehydrogenase; 20, fumarate hydratase; 21, succinate dehydrogenase / fumarate reductase; 22, succinate-CoA ligase; 23, 2oxoglutarate ferredoxin oxidoreductase; 24, isocitrate dehydrogenase; 25/26, aconitate hydratase; 27, citrate lyase; 28, glucose-6-phosphate isomerase; 29, glucokinase; 30, phosphoglucomutase; 31, glucose-1-phosphate adenylyltransferase; 32, starch synthase; 33, 1,4-alpha-glucan branching enzyme; 34, pyruvate carboxylase. The solid and dashed lines represent the reactions that can and cannot, respectively, occur in vestimentiferan symbionts. (E) The glycogen pathway in siboglinid symbionts. GlgC, glucose-1-phosphate adenylyltransferase; GlgA, starch synthase; GlgB, 1,4-alphaglucan branching enzyme. +, present; -, absent. Figure S2. The expression of the trehalose-6-phosphate synthase/phosphatase gene in *Paraescarpia echinospica* tissues. The Y-axis represents the expression in Reads Per Kilobase per Million. Significant ($p<0.05$) differential expressions between tissues are indicated. Figure S3. The evolutionary analysis of TPS. The phylogenetic analysis of the TPS from Annelida, Cnidaria, Cephalochordata, Urochordata, Arthropoda, Rotifera, Nematoda, Arthropoda and others (archaea, bacteria, fungi, plants and protists). Supplemental data-tables. Table S1. The sequencing data used for the assembly of the Arcovestia ivanovi genome. The coverage was calculated using the estimated genome size with the Kmer-based method. -, not available. Table S2. The Alien Iindex score of vVestimentiferan TPSP. The Alien Index was calculated based on the gene similarities (Max Score) between specific outgroups (bbhO) and ingroups (bbhG). Table S3. Expression analysis of the central carbon metabolism genes in the symbionts of Arcovestia ivanovi (A1 and A2) and Lamellibrachia columna (L1). Transcripts per million (TPM) was used to assess the gene expression level. Table S4. The metabolites detected in the metabolomics of Lamellibrachia columna. Fold change indicates the ratio of the metabolites in the trophosome and vestimentitum of the tubeworm. Statistical significance was determined with unpaired Student's t test. Table S5. The statistics of the assembled Arcovestia ivanovi genome. Table S6. The number of genes in the Arcovestia ivanovi genome annotated with different databases. -, not available. Table S7. The categories of the repetitive elements in tubeworms. Table S8. The central carbon metabolisms in Polychaeta. Table S9. The accessions of identified TPSP

Acknowledgements
The authors thank Prof. Ivan Berg (University of Münster, Germany) for valuable suggestions on microbial metabolism. The authors acknowledge Oceanographic Data Center, IOCAS, and System Simulation of Laoshan Laboratory for providing the computing resources.

Authors' contributions
QS, ZY, and LS conceived the study and designed the experiments; QS collected the samples; YS and QS performed the experiments; QS and ZY obtained and analyzed the data; QS, LS, and ZY obtained the funding; QS and ZY wrote the first draft of the manuscript; LS edited the manuscript. All authors have read the manuscript and approved its submission.

Funding
This work was funded by the grants of Science and Technology Innovation Project of Laoshan Laboratory (LSKJ202203100), the National Natural Science Foundation of China (32201219, 42221005, and 41806202), and the Science Fund Program for Distinguished Young Scholars of Shandong Province (Overseas) (2022HWYQ-087).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 June 2024 Accepted: 26 October 2024
Published online: 16 November 2024

References

1. Felbeck H. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). Science. 1981;213:336–8.
2. Li Y, Kocot KM, Schander C, Santos SR, Thornhill DJ, Halanych KM. Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). Mol Phylogenet Evol. 2015;85:221–9.
3. Li Y, et al. Phylogenomics of tubeworms (Siboglinidae, Annelida) and comparative performance of different reconstruction methods. Zool Scr. 2017;46:200–13.
4. Rimskaya-Korsakova NN, Galkin SV, Malakhov VV. The anatomy of the blood vascular system of the giant vestimentiferan tubeworm Riftia pachyptila (Siboglinidae, Annelida). J Morphol. 2017;278:810–27.
5. Zimmermann J, et al. Dual symbiosis with co-occurring sulfur-oxidizing symbionts in vestimentiferan tubeworms from a Mediterranean hydrothermal vent. Environ Microbiol. 2014;16:3638–56.
6. Sun Y, 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.
7. Moggioli G, et al. Distinct genomic routes underlie transitions to specialised symbiotic lifestyles in deep-sea annelid worms. Nat Commun. 2023;14:2814.
8. Wang M, et al. The genome of a vestimentiferan tubeworm (*Ridgeia piscesae*) provides insights into its adaptation to a deep-sea environment. BMC Genomics. 2023;24:72.
9. de Oliveira AL, Mitchell J, Girguis P, Bright M. Novel insights on obligate symbiont lifestyle and adaptation to chemosynthetic environment as revealed by the giant tubeworm genome. Mol Biol Evol. 2022;9:msab347.
10. Li Y, Tassia MG, Waits DS, Bogantes VE, David KT, Halanych KM. Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm *Lamellibrachia luymesi*. BMC Biol. 2019;17:91.
11. Lösekann T, Robador A, Niemann H, Knittel K, Boetius A, Dubilier N. Endosymbioses between bacteria and deep-sea siboglinid tubeworms from an Arctic cold seep (Haakon Mosby mud volcano, Barents Sea). Environ Microbiol. 2008;10:3237–54.
12. Hand SC. Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. Biol Bull. 1987;173:260–76.
13. Zvi-Kedem T, Shemesh E, Tchernov D, Rubin-Blum M. The worm affair: fidelity and environmental adaptation in symbiont species that co-occur in vestimentiferan tubeworms. Environ Microbiol Rep. 2021;13:744–52.
14. Yang Y, et al. Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. ISME J. 2020;14:135–50.
15. Li Y, Liles MR, Halanych KM. Endosymbiont genomes yield clues of tubeworm success. ISME J. 2018;12:2785–95.
16. 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.
17. Perez M, Juniper K. Insights into symbiont population structure among three vestimentiferan tubeworm host species at eastern Pacific spreading centers. Appl Environ Microbiol. 2016;82:5197–205.
18. Rubin-Blum M, Dubilier N, Kleiner M. 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. 2019;4:e00394-18.
19. Hinzke T, et al. Host-microbe interactions in the chemosynthetic Riftia pachyptila symbiosis. mBio. 2019;10:e02243-19.
20. Bright M, Sorgo A. Ultrastructural reinvestigation of the trophosome in adults of Riftia pachyptila (Annelida, Siboglinidae). Invertebr Biol. 2005;122:345–66.
21. Bright M, Keckeism H, Fisher CR. An autoradiographic examination of carbon fixation, transfer and utilization in the Riftia pachyptila symbiosis. Mar Biol. 2000;136:621–32.
22. Felbeck H, Jarchow J. Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. Physiol Zool. 1998;71:294–302.
23. Sun QL, et al. High-throughput sequencing reveals a potentially novel *Sulfurovum* species dominating the microbial communities of the seawater-sediment interface of a deep-sea cold seep in South China Sea. Microorganisms. 2020;8:687.
24. Bolyen E, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:852–7.
25. Zhang J, Sun QL, Zeng ZG, Chen S, Sun L. Microbial diversity in the deep-sea sediments of Iheya North and Iheya Ridge. Okinawa Trough Microbiol Res. 2015;177:43–52.
26. Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. Appl Environ Microbiol. 1997;63:2802–13.
27. Fang C, et al. Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. Gigascience. 2018;7:1–8.
28. Huang J, et al. A reference human genome dataset of the BGISEQ-500 sequencer. Gigascience. 2017;6:gix024.
29. Chen Y, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7:1–6.
30. Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1:18.
31. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.
32. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38:e132.
33. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.
34. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.
35. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
36. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.

37. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.
38. Verrastro L, Maneyro R, Da Silva CM, Farias I. A new species of lizard of the *L wiegmannii* group (Iguania: Liolaemidae) from the Uruguayan Savanna. Zootaxa. 2017;4294:43.
39. Heller P, Casaletto J, Ruiz G, Geller J. A database of metazoan cytochrome c oxidase subunit I gene sequences derived from GenBank with CO-ARBitrator. Sci Data. 2018;5:1–7.
40. Liu H, et al. Draft genome of Glyptosternon maculatum, an endemic fish from Tibet Plateau. Gigascience. 2018;7:giy104.
41. Liu B, et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv. 2013. https://arxiv.org/abs/1308.2012.
42. Chin CS, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nature methods. 2016;13:1050–4.
43. Chin CS, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10:563.
44. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
45. Durbin LR. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
46. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res. 2002;12:656–64.
47. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 2002;12:1269–76.
48. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21(Suppl 1):i351-8.
49. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.
50. Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. 2006;4:41.
51. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14:988–95.
52. Mario S, Burkhard M. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:W465–7.
53. Guigó R, Knudsen S, Drake N, Smith T. Prediction of gene structure. J Mol Biol. 1992;226:141–57.
54. Burge C. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 1997;268:78–94.
55. Majoros W, Pertea M, Salzberg S. TigrScan & GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20:2878–9.
56. Korf I. Gene finding in novel genomes. BMC bioinformatics. 2004;5:59.
57. Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
58. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown Nat Protoc. 2016;11:1650–67.
59. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
60. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
61. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–78.
62. Haas BJ, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.
63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
64. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
65. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
66. Shen XX, et al. Tempo and mode of genome evolution in the budding yeast subphylum. Cell. 2018;175(1533–1545): e20.
67. Kleiner M, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Natl Acad Sci USA. 2012;109:E1173–82.
68. Khmelenina VN, Rozova ON, Trotsenko YA. Characterization of the recombinant pyrophosphate-dependent 6-phosphofructokinases from Methylomicrobium alcaliphilum 20z and Methylococcus capsulatus Bath. Method Enzymol. 2011;495:1–14.
69. Lutz R, et al. Rapid growth at deep-sea vents. Nature. 1994;371:663–4.
70. Sorgo A, Gaill F, Lechaire JP, Arndt C, Bright M. Glycogen storage in the *Riftia pachyptila* trophosome: contribution of host and symbionts. Mar Ecol Prog Ser. 2002;231:115–20.
71. Hinzke T, et al. Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. Elife. 2021;10:e58371.
72. Shikuma NJ. Bacteria-stimulated metamorphosis: an ocean of insights from investigating a transient host-microbe interaction. Systems. 2021;6:e00754-21.
73. Miyamoto T, Amrein H. Gluconeogenesis: an ancient biochemical pathway with a new twist. Fly (Austin). 2017;11:218–23.
74. Matsuda H, Yamada T, Yoshida M, Nishimura T. Flies without trehalose. J Biol Chem. 2015;290:1244–55.
75. Miyamoto T, Amrein H. Neuronal gluconeogenesis regulates systemic glucose homeostasis in Drosophila melanogaster. Curr Biol. 2019;29:1263-1272.e5.
76. Arguelles JC. Thermotolerance and trehalose accumulation induced by heat shock in yeast cells of Candida albicans. FEMS Microbiol Lett. 1997;146:65–71.
77. Girguis PR, Childress JJ. Metabolite uptake, stoichiometry and chemoautotrophic function of the hydrothermal vent tubeworm *Riftia pachyptila*: responses to environmental variations in substrate concentrations and temperature. J Exp Biol. 2006;209:3516–28.
78. Li ZG, Luo LJ, Zhu LP. Involvement of trehalose in hydrogen sulfide donor sodium hydrosulfide-induced the acquisition of heat tolerance in maize (Zea mays L) seedlings. Bot Stud. 2014;55:20.
79. Chen JX, Lyu ZH, Wang CY, Cheng J, Lin T. RNA interference of a trehalose-6-phosphate synthase gene reveals its roles in the biosynthesis of chitin and lipids in Heortia vitessoides (Lepidoptera: Crambidae). Insect Sci. 2020;27:212–23.
80. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008;9:605–18.
81. Galbraith JD, Ludington AJ, Suh A, Sanders KL, Adelson DL. New environment, new invaders—repeated horizontal transfer of LINEs to sea snakes. Genome Biol Evol. 2020;12:2370–83.
82. Graham LA, Davies PL. Horizontal gene transfer in vertebrates: a fishy tale. Trends Genet. 2021;37:501–3.
83. Rödelsperger C, Sommer RJ. Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol. 2011;11:239.

Publisher's Note