Supplemental information

The genome of a hadal sea cucumber reveals novel

adaptive strategies to deep-sea environments

Guangming Shao, Tianliang He, Yinnan Mu, Pengfei Mu, Jingqun Ao, Xihuang Lin, Lingwei Ruan, YuGuang Wang, Yuan Gao, Dinggao Liu, Liangsheng Zhang, and Xinhua Chen

1   **Supplementary Materials**

2   **Supplementary figures**

3   **Figure S1.** GenomeScope plot for YHSC genome. Related to STAR Methods.

4   **Figure S2.** Scatterplot of GC content on read depth in the YHSC genome. Related to STAR Methods.

5   **Figure S3**. The quality assessment of the D152 DNA sample detected by Agilent 4200 Bioanalyzer.
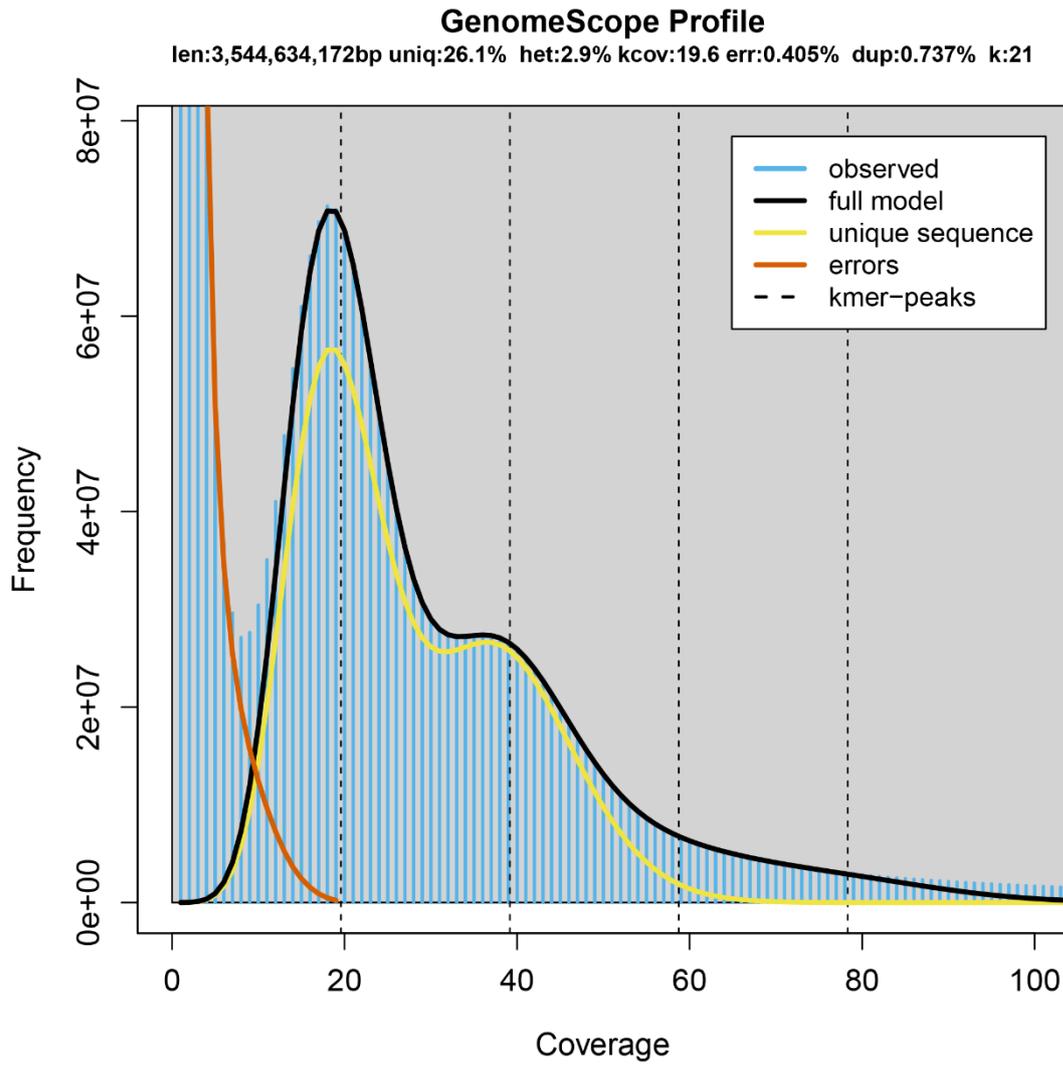6   Related to STAR Methods.

7   **Figure S4.** Bayesian phylogenies of extant Holothuroidea based on (a) three concatenated nuclear genes
8   (18S rRNA, 28S rRNA, and histone H3) and (b) two mitochondrial genes (16S rRNA and COI). Related
9   to Figure 1.

10   **Figure S5.** The whole paranome Ks distributions using histograms and kernel density estimates for
11   YHSC and *A. japonicus* together with the Ks distribution of their one-to-one orthologs. Related to STAR
12   Methods.

13   **Figure S6.** The gain or loss of gene families and the statistics for the unique domains across seven
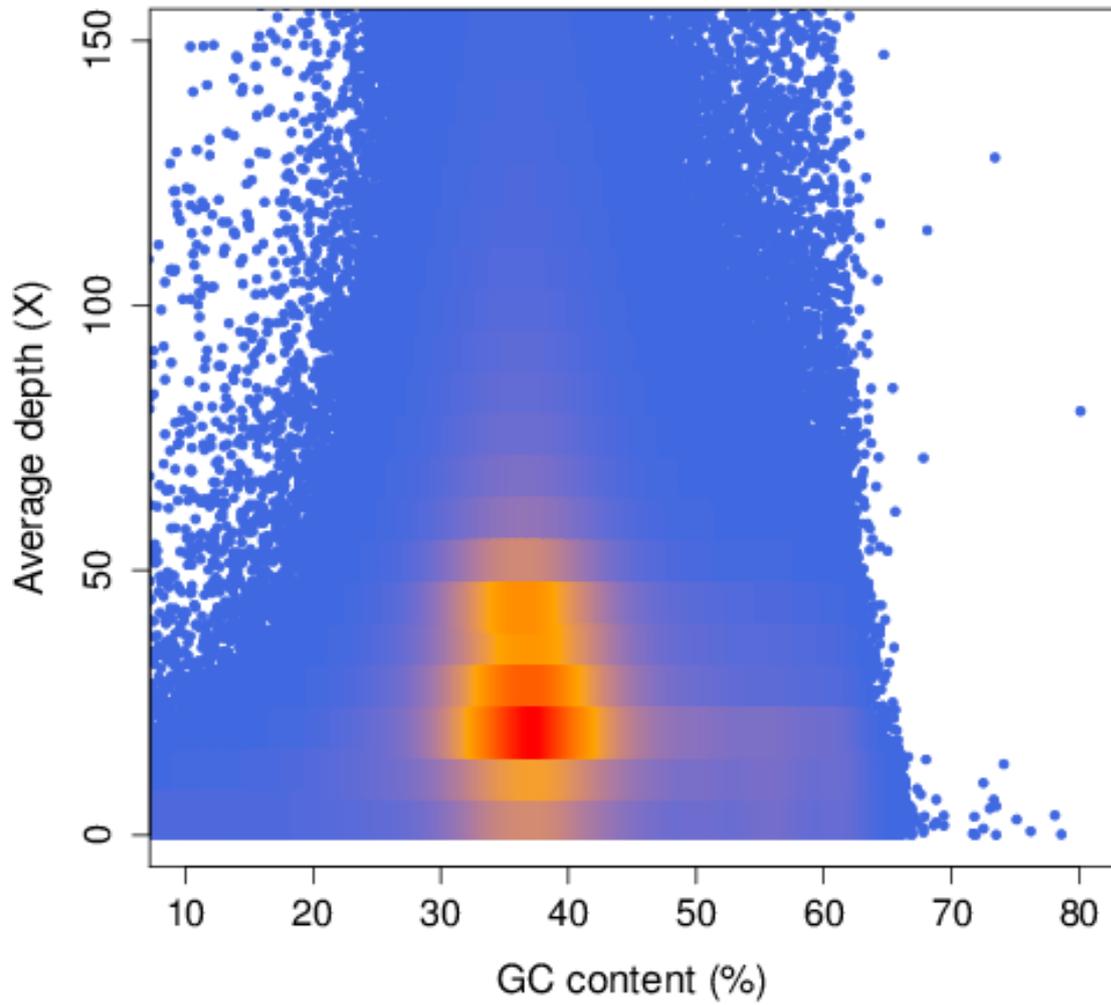14   invertebrate genomes. Related to Figure 2 to Figure 4.

15   **Figure S7.** The most common functions of the expanded domains in the YHSC genome. Related to
16   Figure 2 to Figure 4.

17   **Figure S8.** Heatmap depicting the quantitative variation of the pfam domains involved in translation
18   factors and ribosomal proteins among the species analyzed. Related to Figure 3.

GenomeScope Profile

len:3,544,634,172bp uniq:26.1% het:2.9% kcov:19.6 err:0.405% dup:0.737% k:21

19

20 **Figure S1.** GenomeScope plot for YHSC genome. Related to STAR Methods.

21

Figure S2. Scatterplot of GC content on read depth in the YHSC genome. Related to STAR Methods..
The x-axis shows the GC content of nonoverlapping 10 KB sliding windows along the genome; the y-axis shows the average per-base sequencing coverage.
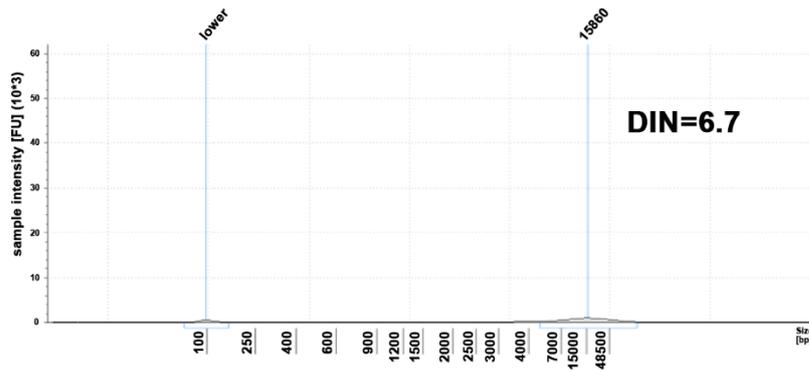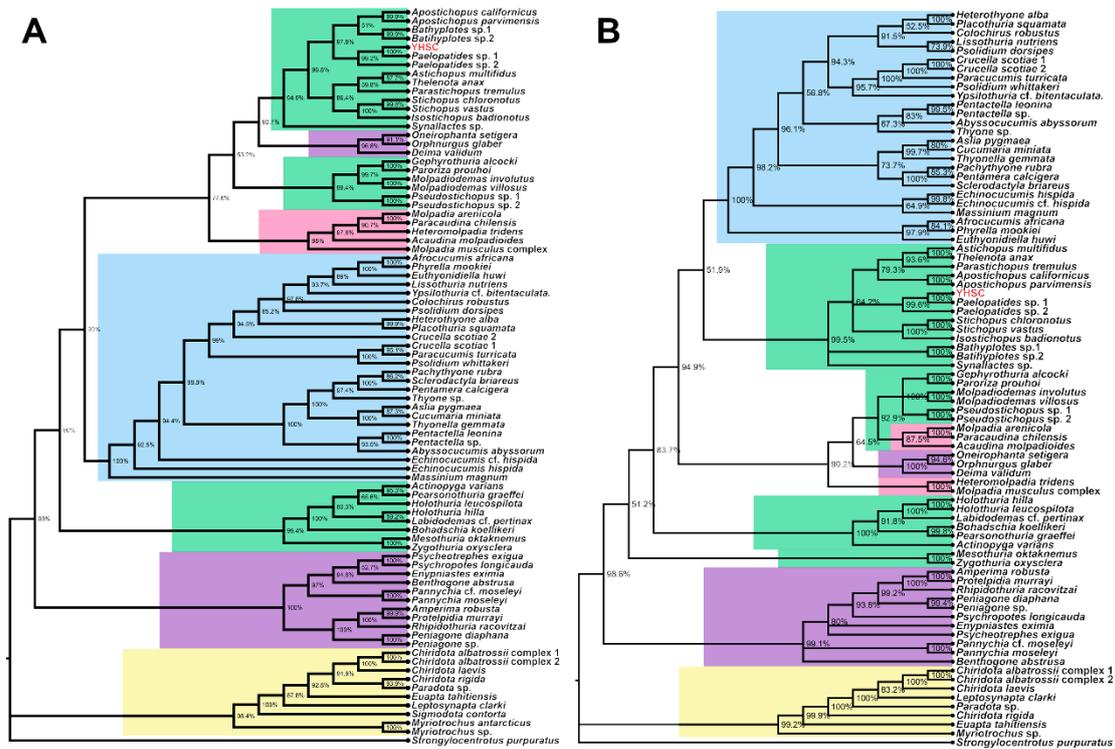
30

Figure S3. The quality assessment of the D152 DNA sample detected by Agilent 4200 Bioanalyzer.
Related to STAR Methods. The character "lower" in the Figure indicated DNA marker. The peak with a
size of 15860bp represents the D152 DNA sample.

Figure S4. Bayesian phylogenies of extant Holothuroidea based on (a) three concatenated nuclear genes
(18S rRNA, 28S rRNA, and histone H3) and (b) two mitochondrial genes (16S rRNA and COI). Related
to Figure 1. The colored boxes surround clades indicated different taxa according to previous report:
blue= Dendrochirotida; orange = Dactylochirotida; yellow = Apodida; purple = Elasipodida; red =
Molpadida; and green = Aspidochirotida.

43
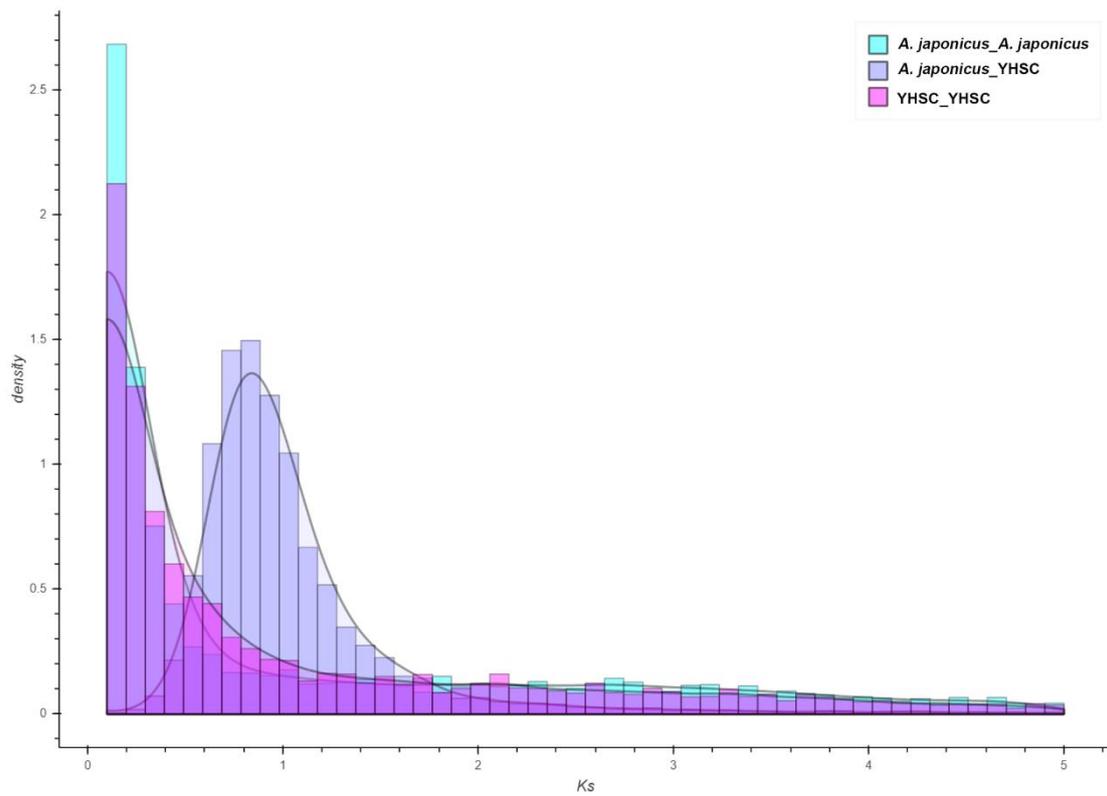
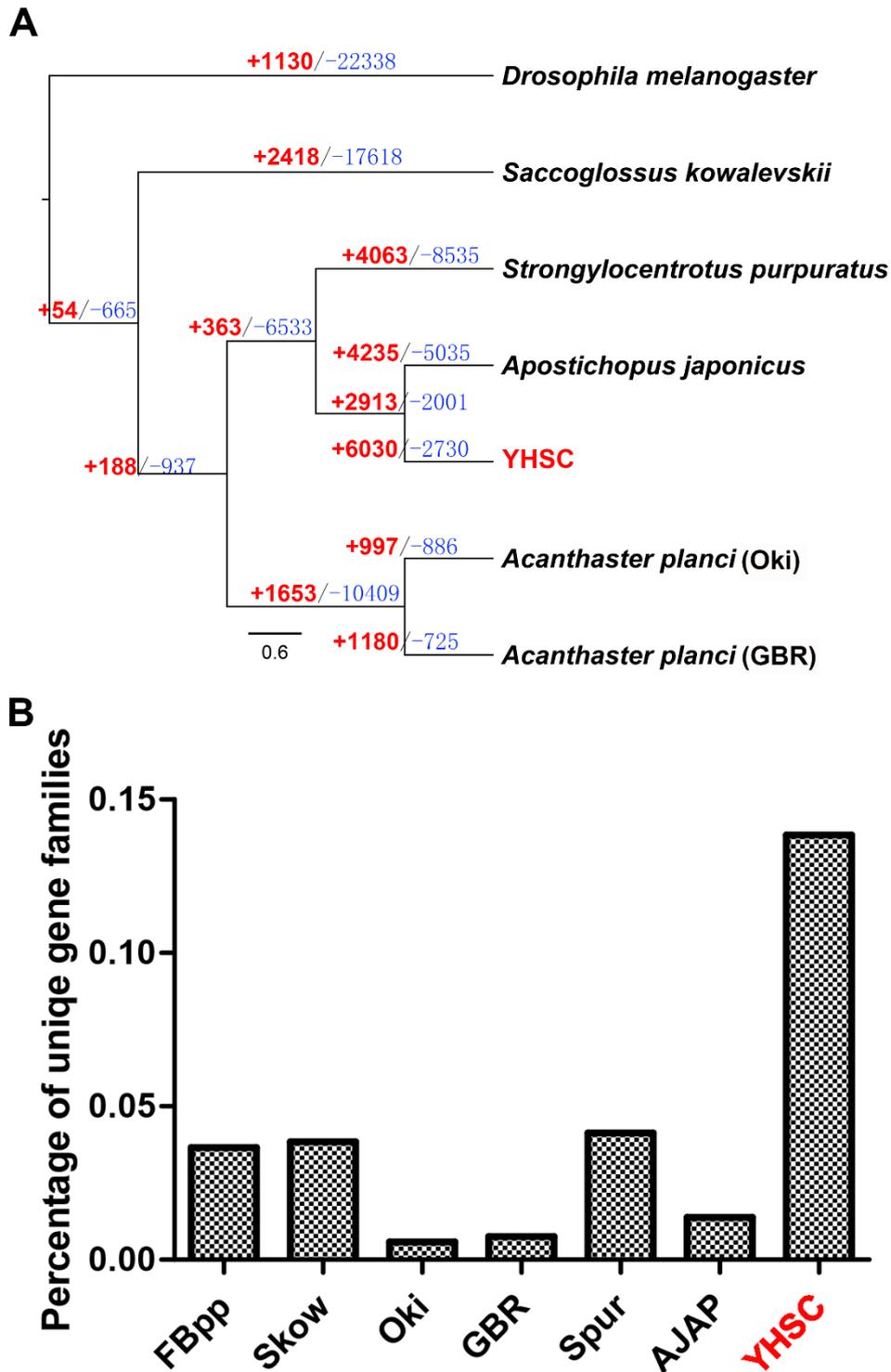
44

Figure S5. The whole paranome *Ks* distributions using histograms and kernel density estimates for YHSC and *A. japonicus* together with the *Ks* distribution of their one-to-one orthologs. Related to STAR Methods.

A

B

48

Figure S6. The gain or loss of gene families and the statistics for the unique domains across seven invertebrate genomes. Related to Figure 2 to Figure 4: "Comparative genomics". a, Phylogenetic tree constructed using the species tree root inference from duplication event (STRIDE) algorithm in OrthoFinder (version 2.3.8) based on the protein sequences from seven invertebrate genomes. The red and blue numbers on each branch show the numbers of gene families gained and lost. The scale bar indicates sequence divergence. b, the percentage of unique Pfam domains in each species analyzed.

55

56
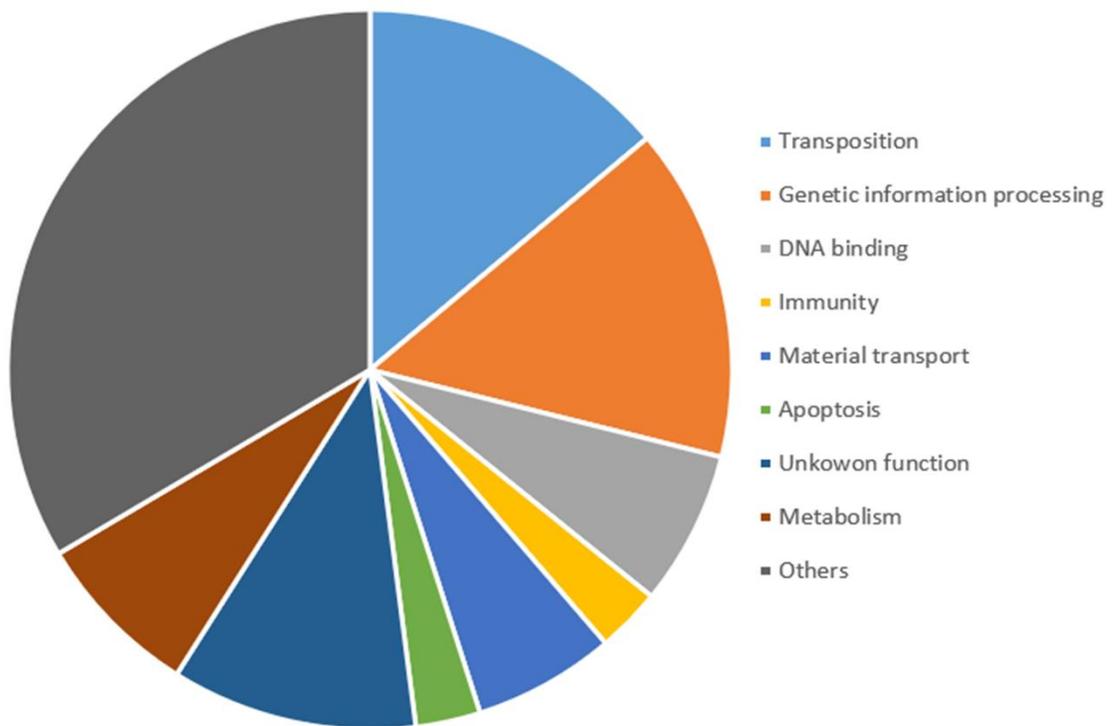
- Transposition
- Genetic information processing
- DNA binding
- Immunity
- Material transport
- Apoptosis
- Unkowon function
- Metabolism
- Others

57

Figure S7. The most common functions of the expanded domains in the YHSC genome. Related to Figure 2 to Figure 4.
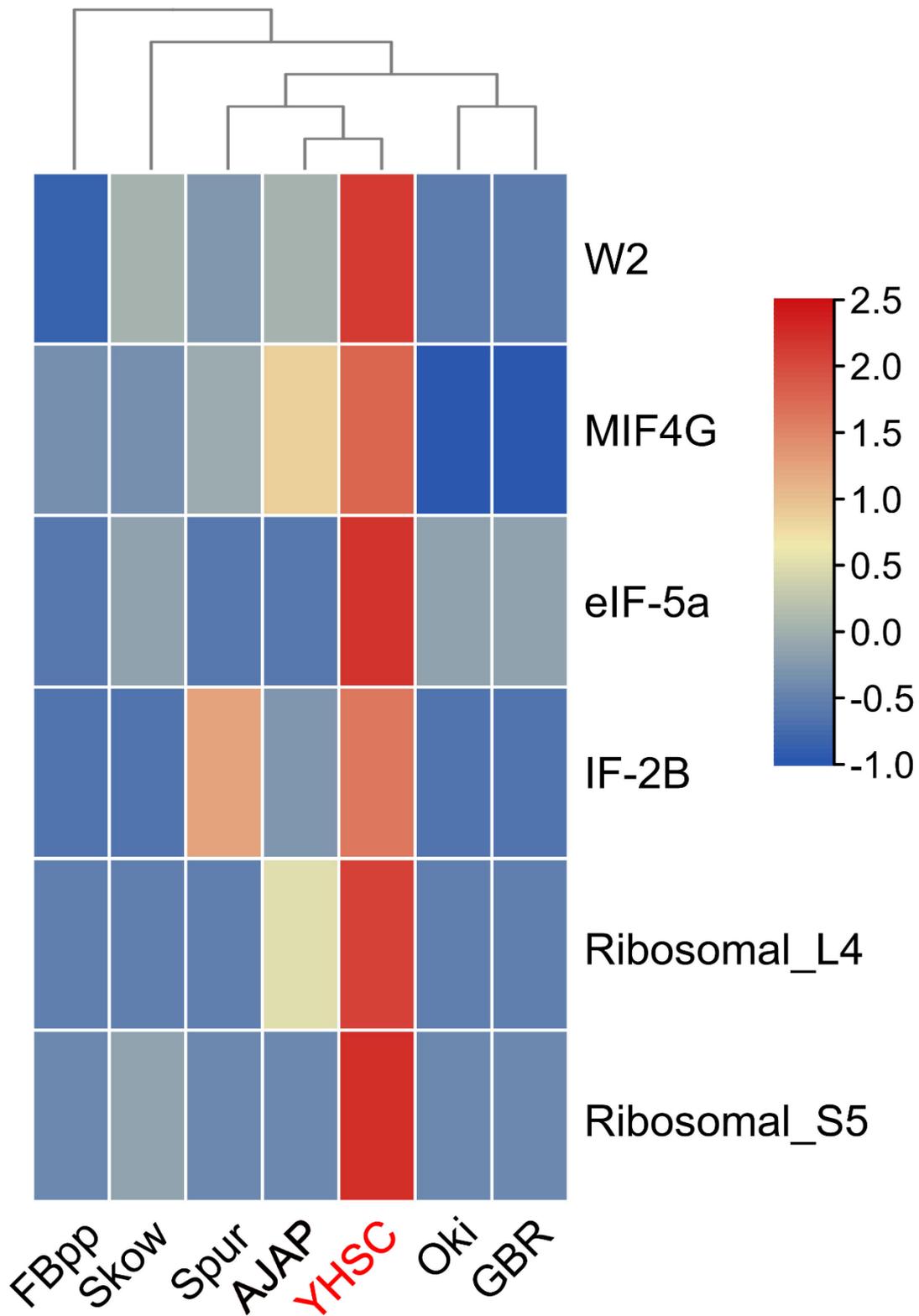
Figure S8. Heatmap depicting the quantitative variation of the pfam domains involved in translation factors and ribosomal proteins among the species analyzed. Related to Figure 3. The variation of the scale bar in color corresponds to the z-score of pfam domain number. The bigger the z-score, the redder the color. The smaller the z-score, the bluer the color. Only the domains in which YHSC genome has

65   more members than the other genomes are shown. Noting that the domains of W2, elf-5a, Ribosomal_L4

66   and Ribosomal_S5 reach the level of expansion, while MIF4G and IF-2B do not.

67

68   **Supplementary tables**

69   **Table S2.** Genome characteristics estimated by GenomeScope. Related to STAR Methods.

70   **Table S3.** Statistics of the PacBio subreads. Related to STAR Methods.

71   **Table S4.** Contamination detection and exclusion. Related to STAR Methods.

72   **Table S5.** Properties of the YHSC genome assembly. Related to STAR Methods.

73   **Table S6.** Assessment of YHSC genome assembly quality based on the successful mapping of the

74   Illumina clean reads to the genome. Related to STAR Methods.

75   **Table S7.** BUSCO evaluation of the YHSC genome assembly. Related to STAR Methods.

76   **Table S8.** Numbers of genes obtained using the three different prediction methods. Related to STAR

77   Methods.

78   **Table S10.** Percentage divergence between the YHSC and species of *Paelopatides*, *Apostichopus*, and

79   *Bathyplotes.* Related to Figure 1.

80   **Table S14.** Relative abundances of various fatty acids in membrane phospholipids of *A. japonicus* and

81   the YHSC. Related to Figure 2.

82

83   **Table S2.** Genome characteristics estimated by GenomeScope. Related to STAR Methods.

Property	Min	Max
Heterozygosity	2.89%	2.90%
Genome Haploid Length	3,542,080,715 bp	3,544,634,172 bp
Genome Repeat Length	2,618,594,627 bp	2,620,482,350 bp
Genome Unique Length	923,486,088 bp	924,151,822 bp
Model Fit	87.95%	98.52%
Read Error Rate	0.40%	0.40%

84

85   **Table S3.** Statistics of the PacBio subreads. Related to STAR Methods.

Sample ID	MB1	MB2	MB3	Total
Subreads number	25,755,660	26,309,430	10,991,427	63,056,517
Subreads base (bp)	197,864,419,961	201,749,020,982	91,261,630,410	490,875,071,353
Average length	7,682	7,668	95,093	
Subreads n50	11,077	10,315	12,683	

86

87 **Table S4.** Contamination detection and exclusion[1]. Related to STAR Methods.

	Gene number	Contig number	Percentage (%)
Total	60469	59054	
Genes from bacteria[2]	2801		4.63
Genes from deuterostomia[3]	52257		86.42
Genes with no sequence similarity	5411		8.95
Contigs containing bacterial genes		1872	
Removed contigs[4]		876	
Removed bacterial genes	2151		

88 [1]Contamination detection was performed by searching *Paelopatides* sp. proteins against the non-
89 redundant bacterial and deuterostome proteins retrieved from the NCBI nr database with a e-value of 1e-
90 5. [2,3]Taxonomy assignment for each query sequence was based on the best hits. A gene is assigned to
91 bacteria if it encoded a protein that is most similar to bacterial protein, while assigned to YHSC if most
92 similarity to deuterostome proteins.[4]Contigs with ≥50% bacterial proteins or no deuterostome proteins
93 were removed.

94

95 **Table S5.** Properties of the YHSC genome assembly. Related to STAR Methods.

Stat Type	Contig Length (bp)	Contig Number
Before redundant and contaminated sequences were removed		
Total	3,821,742,837	59,054
Max	2,300,492	
Number>=2000bp	-	59,018
N50	133,867	
N60	100,351	
N70	71,896	

N80	48,202	
N90	27,861	
Redundant contigs		2740
Contaminated contigs		876
After redundant and contaminated sequences were removed		
Total	3,704,652,769	55,447
Max	2,300,492	
Number>=2000bp		55,425
N50	137,082	
N60	102,343	
N70	73,213	
N80	49,190	
N90	28,545	

96

97 **Table S6.** Assessment of YHSC genome assembly quality based on the successful mapping of the
98 Illumina clean reads to the genome. Related to STAR Methods.

Sample	HS
Clean Reads	1,394,029,234
Clean Bases(bp)	209,104,385,100
Mapped Reads	1,381,633,889
Mapped Reads Rate (%)	99.11%
Mapped Bases	207,245,083,350
Mean Depth	56.44
Coverage Rate (%)	95.98
Coverage at least 4X (%)	94.05
Coverage at least 10X (%)	90.48
Coverage at least 20X (%)	76.35

99

100 **Table S7.** BUSCO evaluation of the YHSC genome assembly. Related to STAR Methods.

Items	Number	Percent (%)

Complete BUSCOs (C)	424	89.4
Complete and single-copy BUSCOs (S)	228	76.9
Complete and duplicated BUSCOs (D)	196	12.5
Fragmented BUSCOs (F)	32	8.2
Missing BUSCOs (M)	21	2.4
Total BUSCO groups searched	255	100

101

102 **Table S8.** Numbers of genes obtained using the three different prediction methods. Related to STAR
103 Methods.

	De novo	Homolog	Transcriptome-based	Total
numbers	43841	47881	11482	57,522
weights	72.50%	79.18%	18.99%	

104

105 **Table S10.** Percentage divergence between the YHSC and species of *Paelopatides*, *Apostichopus*, and
106 *Bathyplotes.* Related to Figure 1.

Paelopatides sp. 1	*Paelopatides* sp. 2	*Apostichopus japonicus*	*Apostichopus californicus*	*Apostichopus parvimensis*	*Bathyplotes* sp. 1	*Bathyplotes* sp. 2
1.05	11.87	24.14	27.47	25.57	18.62	18.32

107

108

109

110

111 **Table S14.** Relative abundances of various fatty acids in membrane phospholipids of *A. japonicus* and
112 the YHSC. Related to Figure 2. Asterisks indicate a significant difference in relative content between the
113 two taxa: *, $P < 0.05$; **, $P < 0.01$ (Student's t test). Red type indicates the fatty acid that differed most
114 substantially in relative content between *A. japonicus* and YHSC.

Fatty acid type	*Apostichopus japonicus* (N=6)	YHSC (N=3)	p-value

C6:0	0.00401±0.00188	0±0	0.18766
C8:0	0.00219±0.00142	0±0	0.32786
C10:0	0.00035±0.00035	0±0	0.51649
C11:0	0±0	0±0	-
C12:0	0.00319±0.00276	0±0	0.4548
C13:0	0±0	0±0	-
C14:1	0.0091±0.00659	0±0	0.37703
C14:0*	0.02871±0.01568	0.19478±0.09891	0.04627
C15:1	0.00592±0.00533	0.02613±0.02436	0.29245
C15:0	0.00051±0.00051	0.03486±0.03486	0.17622
C16:1	0.0907±0.04513	0.13178±0.12191	0.70355
C16:0*	0.28381±0.11719	7.48512±4.04031	0.0299
C17:1	0.00205±0.00105	0.01222±0.012	0.24466
C17:0*	0.01628±0.00363	0±0	0.01821
C18:3-cis-6,9,12	0.00073±0.0004	0±0	0.24883
C18:2-cis-9,12	0.14296±0.04189	0±0	0.0525
C18:1-trans-9	0.05668±0.01643	0.02401±0.01105	0.23467
C18:3-cis-9,12,15	0.00681±0.00368	0±0	0.24727
C18:1-cis-9	0.2648±0.08715	0.27697±0.26249	0.95625
C18:0	0.15231±0.07031	3.06367±1.89066	0.05153
C20:4-cis-5,8,11,14	1.11441±0.52437	0.60275±0.30362	0.53931
C20:5-cis-5,8,11,14,17	0.287±0.11066	0.71701±0.37478	0.18697
C20:3-cis-8，11,14	0.00094±0.00094	0±0	0.51649
C18:2-trans-9,12	0.00001±0.00001	0.00002±0	0.49223
C20:1-cis-11	0.13576±0.11258	0.06358±0.06197	0.68341
C20:3-cis-11,14,17	0.00001±0.00001	0.00002±0	0.50021
C20:0	0.06501±0.02083	0±0	0.07048
C20:2-cis-11,14	0.04081±0.01823	0±0	0.17015
C22:5n6- 4,7,10,13,16	0.01267±0.00693	0±0	0.25178
C22:6**	0.04292±0.00735	0.3119±0.04211	0.00004
C22:4	0.01988±0.01077	0±0	0.24762
C22:5n3-7,10,13,16,19	0.03586±0.0178	0±0	0.21124
C22:2-cis-13,16	0±0	0±0	-

C22:1-cis-13	0.02083±0.00642	0.04209±0.01991	0.22762
C22:0	0.0389±0.0155	0±0	0.13022
C23:0	0±0	0±0	-
C24:1	0.06239±0.02597	0.07151±0.00798	0.81903
C24:0	0±0	0±0	-