

933 **Supplementary Information**

934

935 **Ecogenomics reveals viral communities across the Challenger Deep oceanic trench**

936

937

938 Ying-Li Zhou^{1,2, #}, Paraskevi Mara^{3, #}, Dean Vik⁴, Virginia P. Edgcomb³, Matthew B.
939 Sullivan^{4,5}, Yong Wang^{1,6, *}

940 ¹ Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya,
941 Hainan, China

942 ² University of the Chinese Academy of Sciences, Beijing, China

943 ³ Department of Geology and Geophysics, Woods Hole Oceanographic Institution,
944 Woods Hole, MA, USA

945 ⁴ Department of Microbiology and Center of Microbiome Science, The Ohio State
946 University, Columbus, OH, USA

947 ⁵ Department of Civil, Environmental and Geodetic Engineering, The Ohio State
948 University, Columbus, OH, USA

949 ⁶ Institute for Ocean Engineering, Shenzhen International Graduate School, Tsinghua
950 University, Shenzhen, China

951 # These authors contributed equally

952 * Yong Wang

953 **Email:** wangyong@sz.tsinghua.edu.cn

954

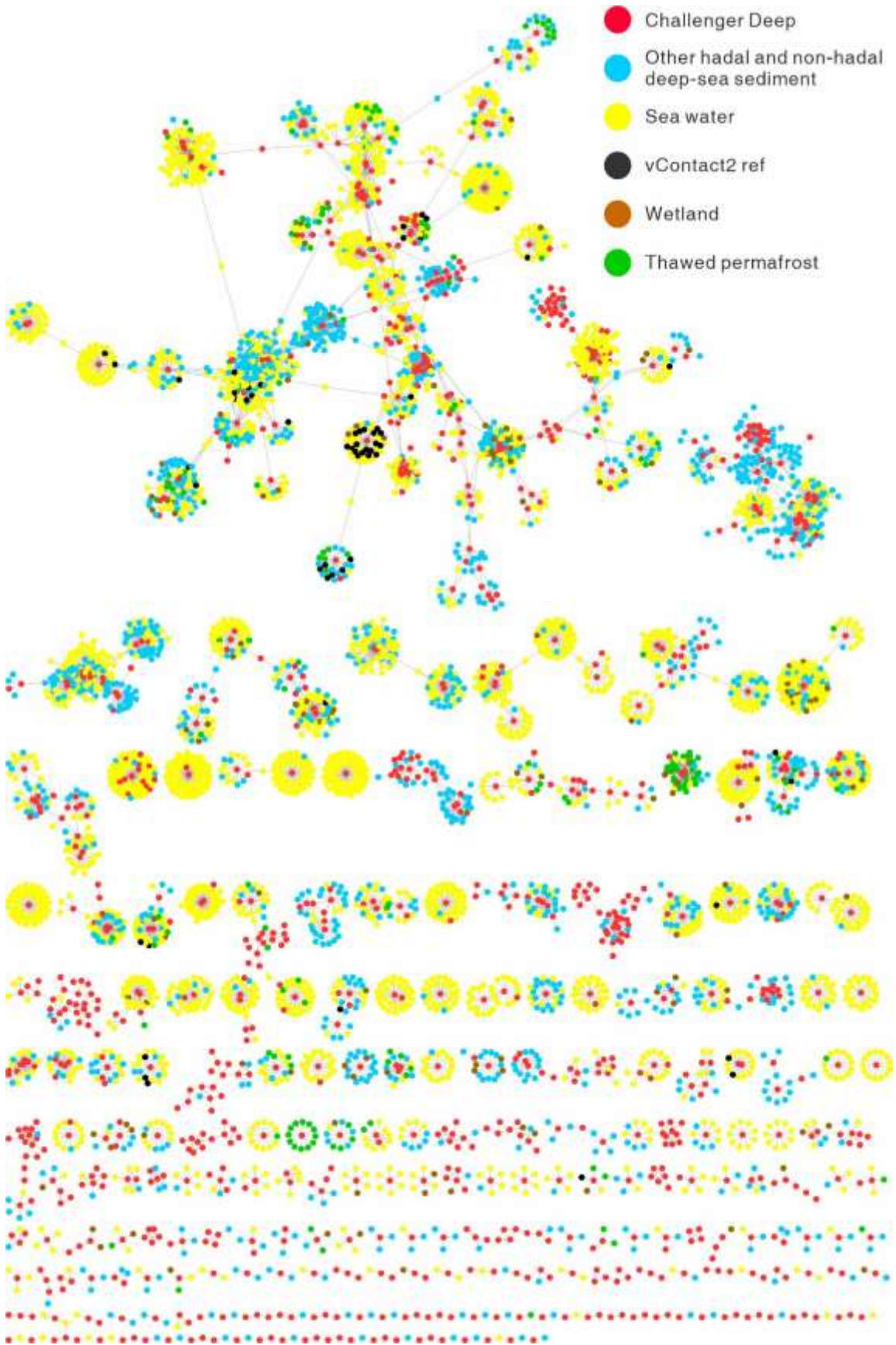
955 **Supplementary Information includes:**

956 Supplementary Figures 1-10.

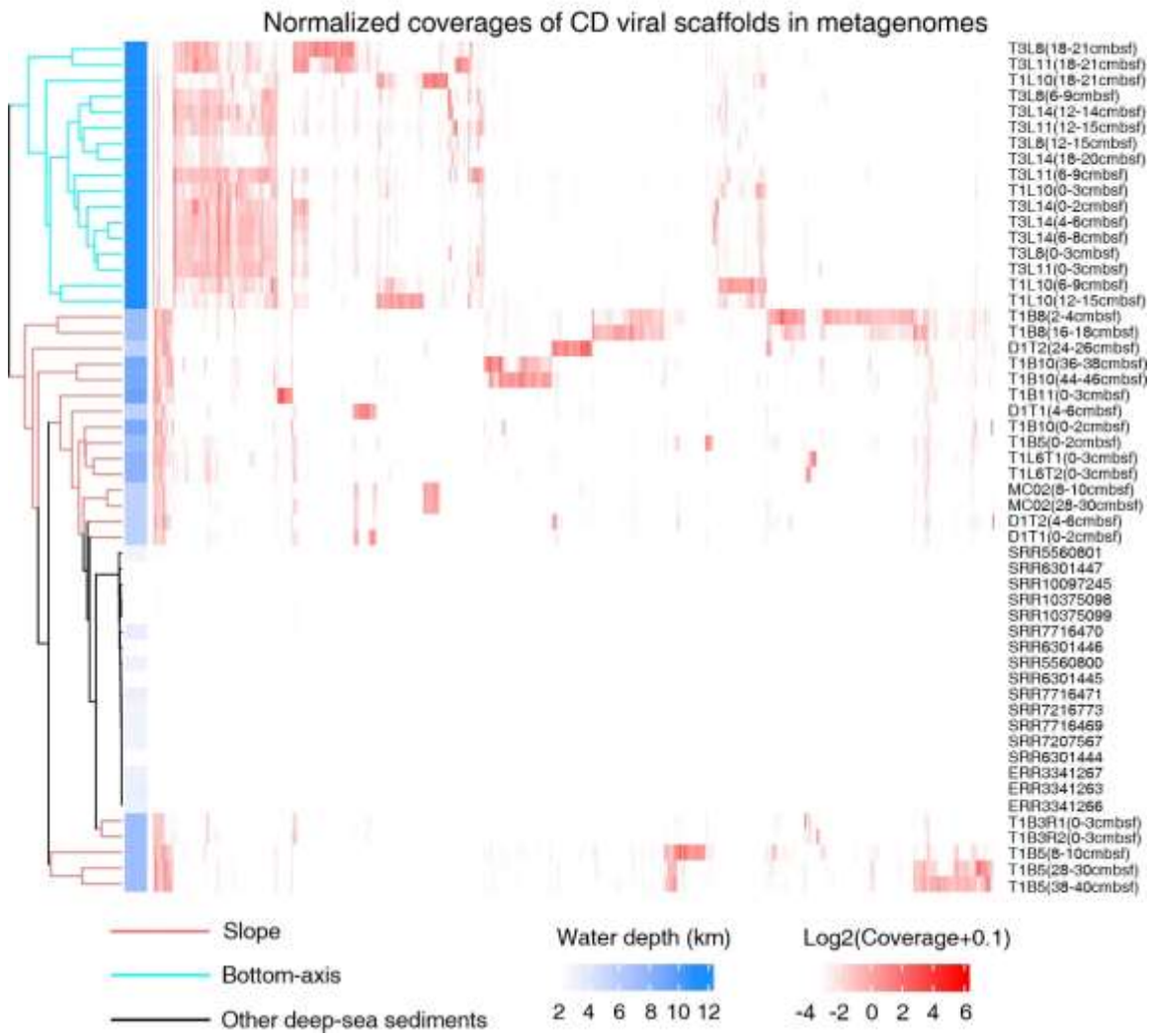
957 Supplementary Tables 1 and 2.

958 Legends for Supplementary Data 1-11. Supplementary Data 1-11 are provided in separate
959 excel files.

960



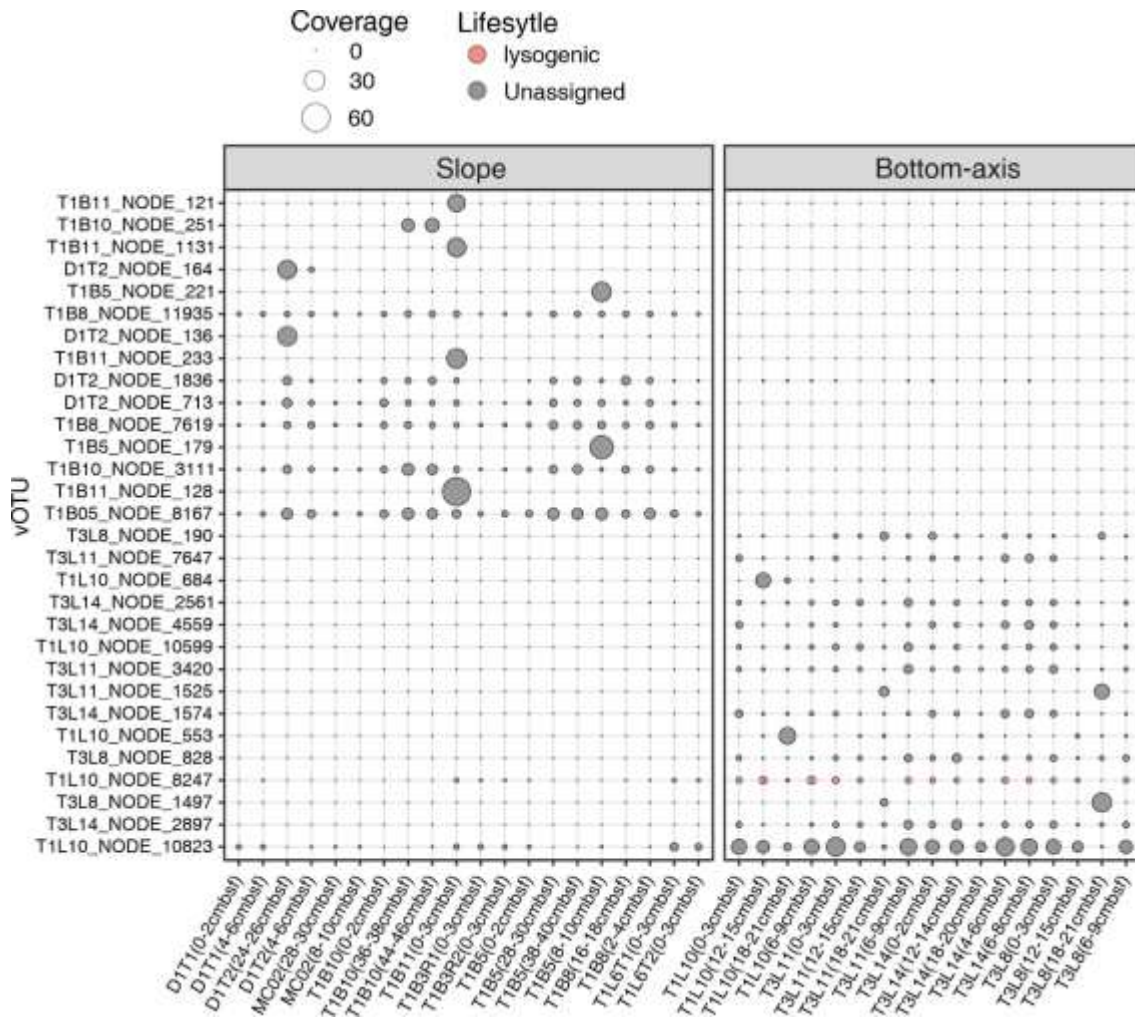
962 **Supplementary Fig. 1.** Protein-sharing network between viral contigs from Challenger
963 Deep (n = 1 628), other hadal and non-hadal deep-sea (n = 2 885) and wetland sediments
964 (n = 1 212), Global Ocean Virome 2.0 (n = 195 728), thawed permafrost soils (n=1 907),
965 and vContact2 RefSeq prokaryotic viral genomes (n = 2 010). Nodes (circles) represent
966 viral contigs. Shared edges (lines) indicate shared protein content. Well-supported viral
967 clusters containing CD vOTUs were shown in this figure.



968

969 **Supplementary Fig. 2.** Normalized coverages of viral populations (species level) in CD
 970 and reference deep-sea metagenomes. Each row represents an individual CD metagenome,
 971 while each column represents an individual CD viral population. The normalized relative
 972 abundance values (log₂-transformed) are shown in color scale. Red, blue and black lines
 973 represent the hierarchical clustering of slope, bottom-axis and other deep-sea sediments,
 974 respectively.

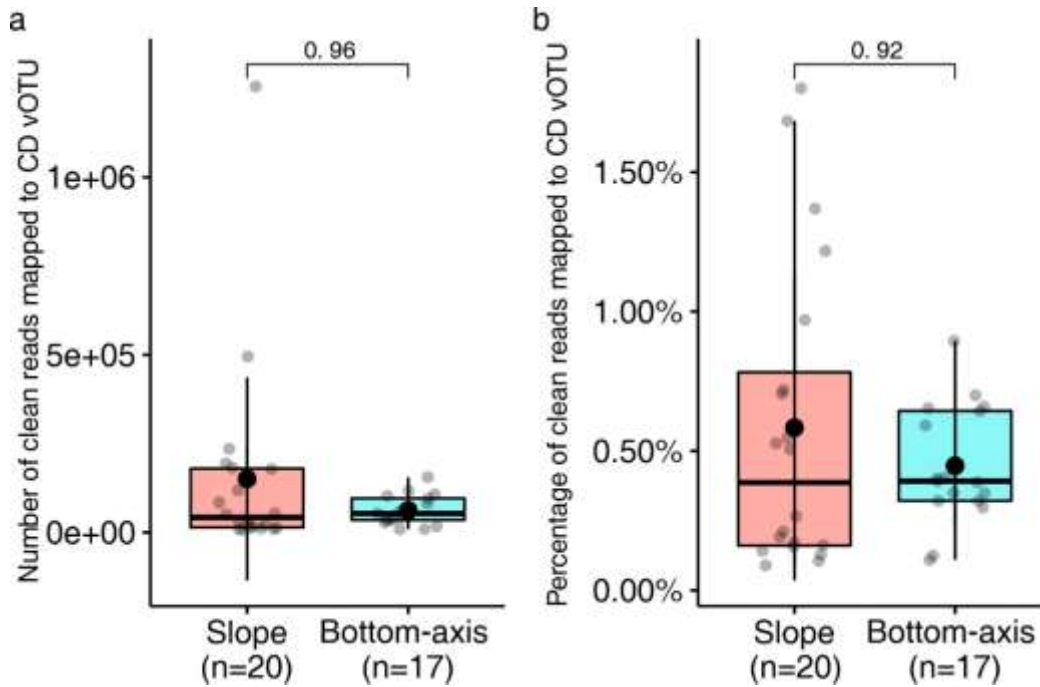
975



976

977 **Supplementary Fig. 3.** Relative abundance of the 15 most abundant viral populations for
 978 slope and bottom-axis samples. The bubbles indicate the normalized coverage (coverage
 979 per giga base), and the predicted lifestyle of viruses for each sample.

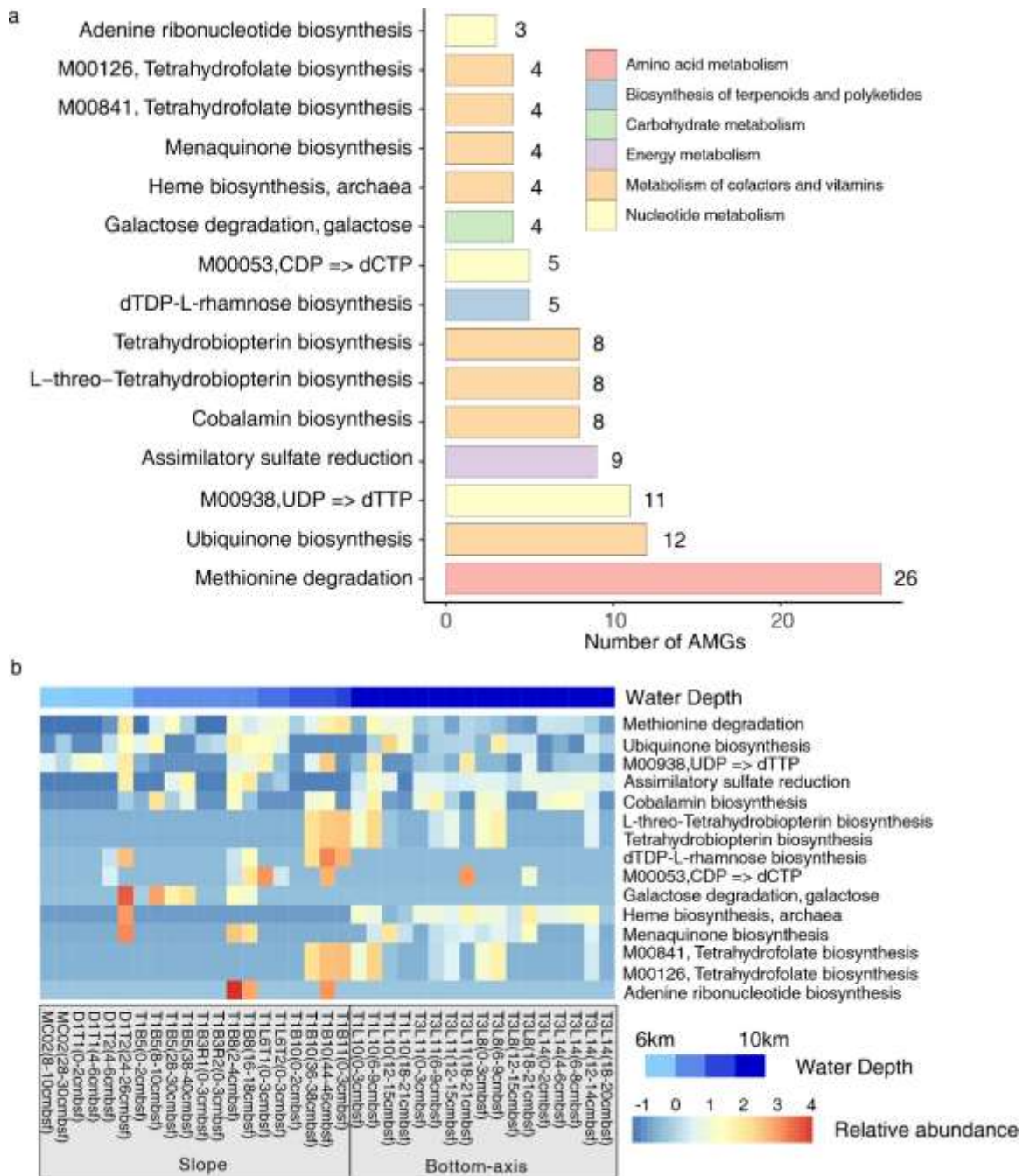
980



981

982 **Supplementary Fig. 4.** Sequencing depth of CD viromes in number of total paired reads
 983 per metagenome. (a) The number of paired-end reads mapped to CD vOTUs. (b)
 984 Percentage of paired-end reads mapped to CD vOTUs in metagenomes. *p*-value was
 985 estimated using the two-sided Wilcoxon test. For boxplots, centre line indicates median,
 986 bounds of box indicate 25th and 75th percentiles, and whiskers indicate minimum and
 987 maximum.

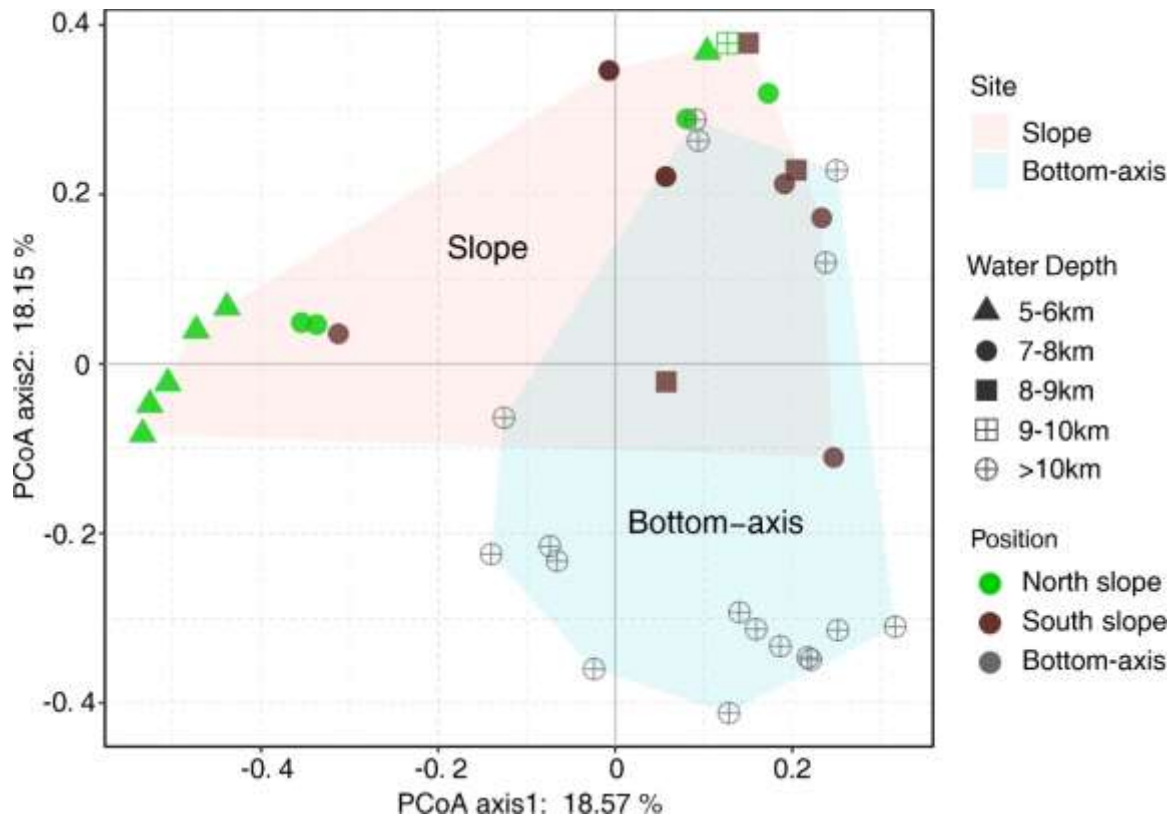
988



989

990 **Supplementary Fig. 5** Overview of putative AMGs in CD viral communities. (a) Number
 991 of putative AMGs involved ranked by AMGs prevalence in CD virome (top fifteen shown).
 992 (b) Abundance of viruses that have genes belonging to the fifteen most prevalent AMGs
 993 identified in CD sediment viromes. The genes identified in various KEGG-described
 994 metabolisms are listed in Supplementary Data 9. Relative abundances were log₂
 995 transformed and normalized with a z-score method based on normalized vOTUs coverage.

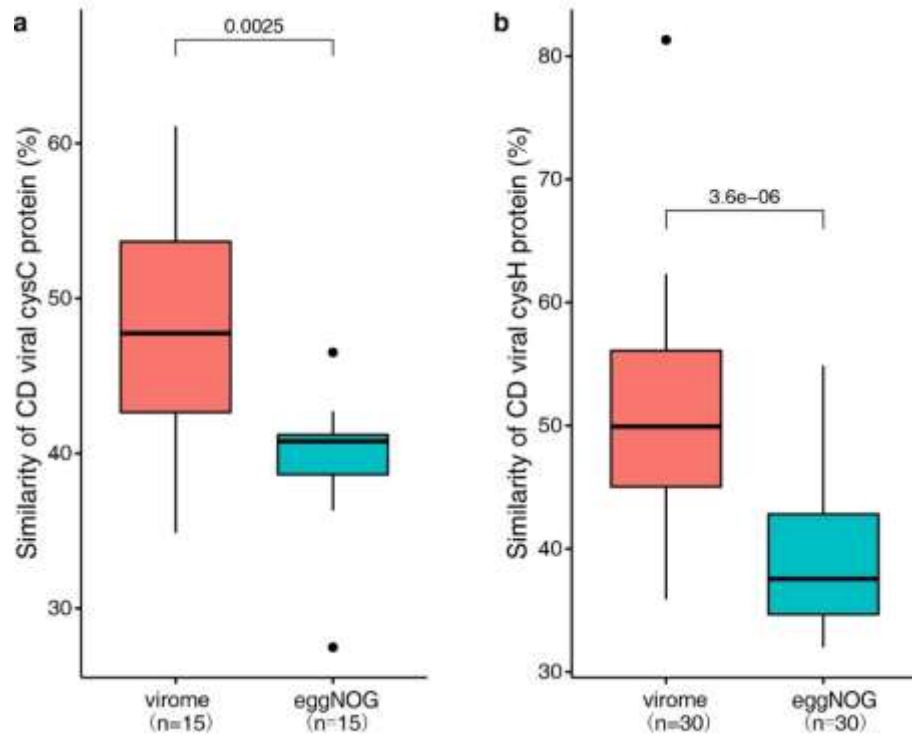
996



997

998 **Supplementary Fig. 6.** Bray-Curtis dissimilarity principal coordinate analyses (PCoA) of
 999 the viral AMGs. The pink and blue areas covered slope and bottom-axis samples,
 1000 respectively.

1001

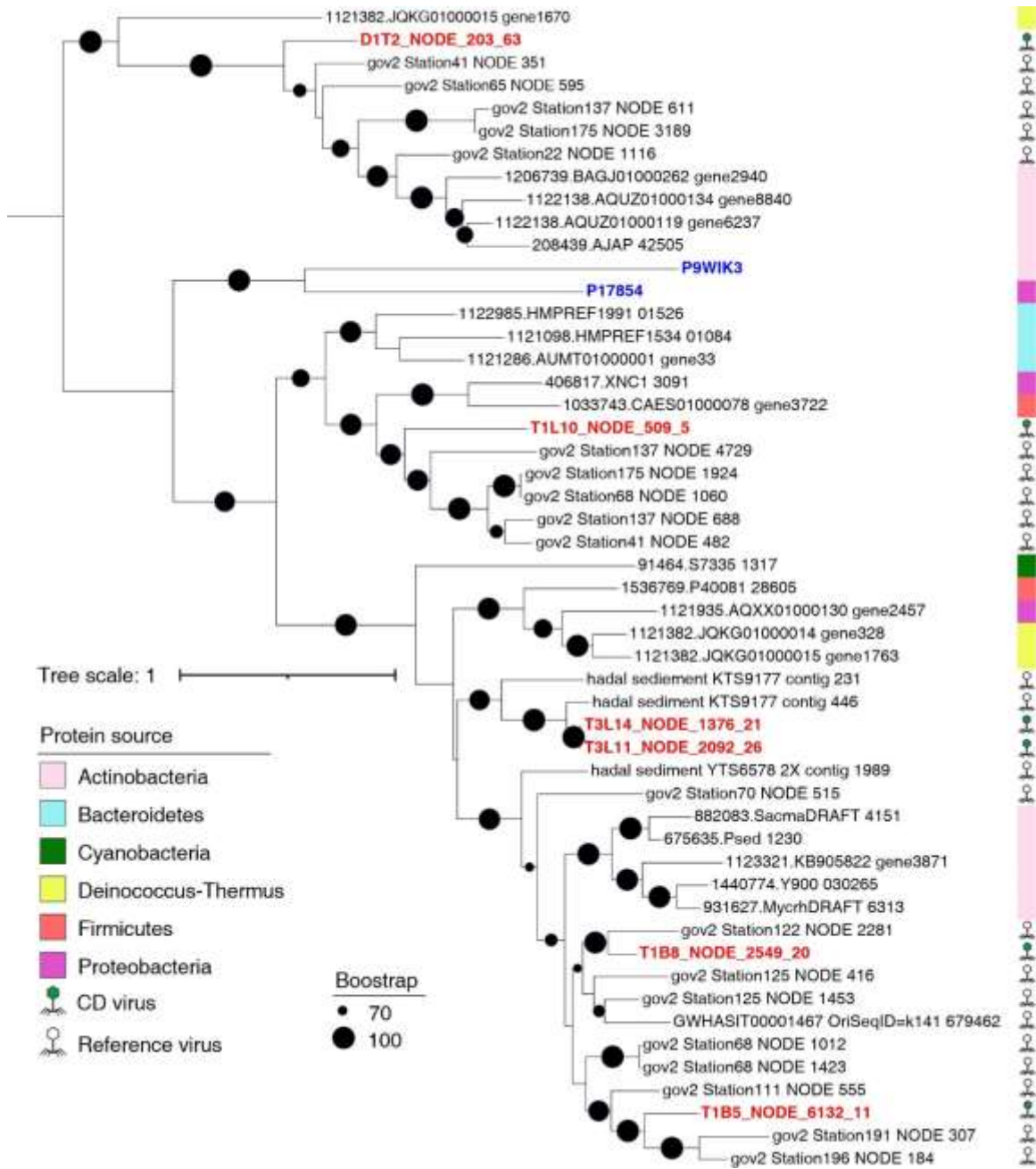


1002

1003 **Supplementary Fig. 7.** Similarity of each Cys protein (CysC/CysH) with the five closest
 1004 homologues recruited from public viromes and the eggNOG database, respectively. Three
 1005 CysC (**a**) and six CysH proteins (**b**) from CD viral contigs. Boxplots represent the median
 1006 with 25th and 75th percentiles, and whiskers show the 1.5 interquartile range.

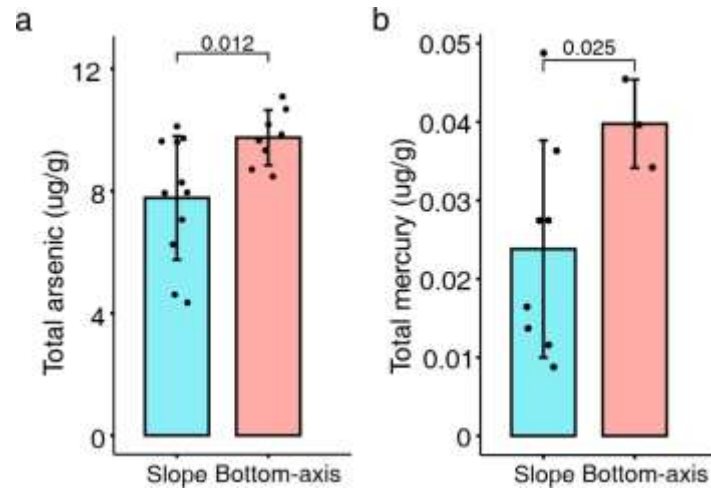
1007

1008



1010
 1011 **Supplementary Fig. 8.** Maximum-likelihood phylogenetic tree between the identified CD
 1012 CysH proteins (this study) and the CysH proteins found in eggNOG database (V5.0). The
 1013 tree is rooted in the mid-point. Two CysH homologs from Uniport database are in blue,
 1014 and indicate protein function based on experimental evidence. Bootstrap values (1 000
 1015 replicates) $\geq 70\%$ are indicated at nodes.
 1016

1017

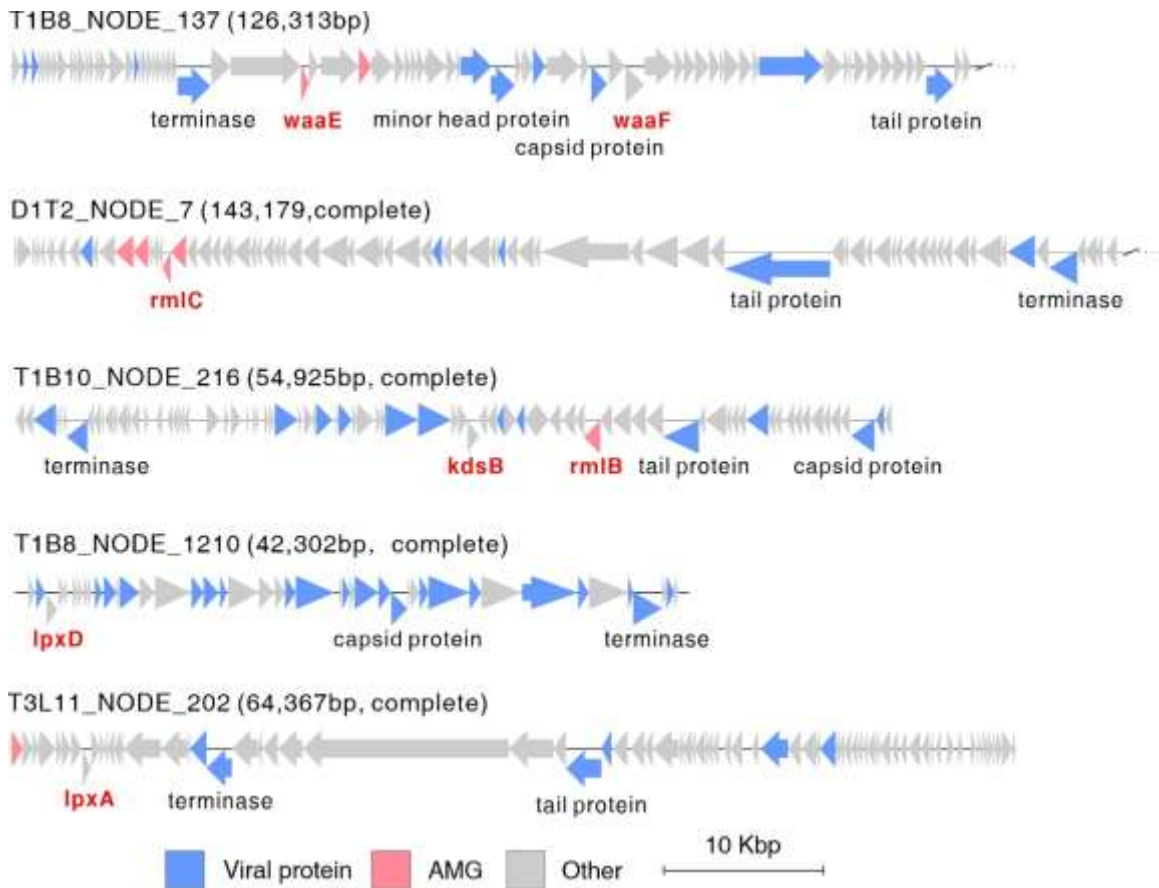


1018

1019 **Supplementary Fig. 9.** Total arsenic and mercury concentrations ($\mu\text{g g}^{-1}$ of dry sediment)
1020 in 13 CD sites (5 400-10 911 m depth). p values were estimated using two-sided Wilcoxon
1021 test for pairwise comparisons. CD sediments were analyzed for total arsenic and mercury
1022 using Atomic Fluorescence Spectrometry. Bar graphs identify mean and standard error bars.

1023

1024



1025

1026 **Supplementary Fig. 10.** Genome maps of the viral contigs that contain LPS and dTDP-
1027 L-rhamnose synthetic genes. The maps indicate the positions of these genes in the
1028 viromes, and the upstream and/or downstream phage genes. Viral genes were annotated
1029 by CheckV and VIBRANT. AMGs, non-phage-like and uncharacterized genes are shown
1030 in blue, pink and gray, respectively. Detailed functional annotation of the viral scaffolds
1031 can be found in Supplementary Data 8.

1032 **Supplementary Table 1.** Sampling sites and sampling methods used for collecting the sediment samples in CD

Cruise	Sample station	Sampling method	Latitude [North]	Longitude [East]	Water Depth [m]	Position in trench	Layers for metagenome [cmbsf]	Layers for metatranscriptome [cmbsf]
DY37II	MC02	push core (submersible)	11.764	141.976	5 400	slope	8-10; 28-30	\
DY37II	D1T1 [#]	push core (submersible)	11.801	142.117	5 533	slope	0-2; 4-6	\
DY37II	D1T2 [#]	push core (submersible)	11.801	142.117	5 533	slope	4-6; 24-26	\
TS01	T1B5	box core (hydrographic cable)	10.924	141.799	7 061	slope	0-2; 8-10; 28-30; 38-40	\
TS01	T1B3	box core (hydrographic cable)	11.553	141.873	7 082	slope	0-3	\
TS01	T1B8	box core (hydrographic cable)	11.602	142.228	7 143	slope	2-4; 16-18	\
TS01	T1L6	push core (lander)	11.091	142.073	7 850	slope	0-2 (two push cores)	\
TS01	T1B10	box core (hydrographic cable)	11.195	141.812	8 638	slope	0-2; 36-38; 44-46	\
TS01	T1B11	box core (hydrographic cable)	11.228	141.689	9 150	slope	0-3	\
TS03	T3L11	push core (lander)	11.325	142.191	10 908	bottom-axis	0-3; 6-9; 12-15; 18-21	6-9; 12-15; 18-21
TS03	T3L8	push core (lander)	11.327	142.194	10 909	bottom-axis	0-3; 6-9; 12-15; 18-21	\
TS03	T3L14	push core (lander)	11.325	142.189	10 911	bottom-axis	0-2; 4-6; 6-8; 12-14; 18-20	\
TS01	T1L10	push core (lander)	11.328	142.202	10 900	bottom-axis	0-3; 6-9; 12-15; 18-21	\

1033 [#]: sediment samples from two slope sites (~ 100m apart) collected in a single dive of the submersible “Jiaolong”.

1034

1035 **Supplementary Table 2. List of marine viruses related datasets**

Environment source	Ocean virome dataset	Ocean and sea regions (number of samples)	Total number of viromes	Sequencing technology	Data size (Gb)	Reference
Ocean water	The Marine Viromes of Four Oceanic Regions	Arctic Ocean (56; Canadian Arctic and Chukchi Sea), North Atlantic Ocean (42; Sargasso Sea (1), Gulf of Mexico (41)), North Pacific Ocean (85; British Columbia coastal waters)	4 viromes from 183 integrative samples	Life Sciences 454 pyrosequencing	0.181	¹
	Temperate and subtropical seawater	Saanich Inlet (11), Strait of Georgia (85), Gulf of Mexico (46)	142	Roche 454 Titanium	1.9 *	²
	Tara Oceans Viromes (TOV)	Mediterranean Sea (7), Red Sea (5), Indian Ocean (16), North Atlantic Ocean (0), South Atlantic Ocean (12), Southern Ocean (1), North Pacific Ocean (2), South Pacific Ocean (0)	43	Illumina HiSeq 2000	218 *	³
	Global Ocean Viromes (GOV)	Mediterranean Sea (8), Red Sea (6), Indian Ocean (27), North Atlantic Ocean (4), South Atlantic Ocean (19), Southern Ocean (4), North Pacific Ocean (15), South Pacific Ocean (21)	104	Illumina HiSeq 2000	925	⁴
	Baltic Sea Viromes	Baltic Sea (11)	21	Roche 454 pyrosequencing	NA	⁵
	Global Ocean Viromes 2.0 (GOV 2.0) [#]	Arctic Ocean (38), North Atlantic Drift Province (3; station 155), all GOV samples (104)	145	Illumina HiSeq 2000	3950	⁶
Hadal and non-hadal deep-sea sediment	Hadal sediment viromes	Izu-Ogasawara Trench (1), Mariana Trench (1)	2	454 GS FLX Titanium	0.029	⁷
	Arctic Ocean metagenome	Arctic Ocean: Southern Knipovich Ridge (3) Gakkel Ridge, Loki's castle (1)	4	Illumina HiSeq 2500	244	⁸
	Southwest Indian Ocean bulk metagenome	Southwest Indian Ocean (6)	6	Illumina HiSeq-2500	373	⁹
	Cold seep bulk metagenome [#]	Haakon Mosby mud volcano (6); Eastern North Pacific ODP site 1244 (7); Mediterranean Sea-Amon mud volcano (1); Santa Monica Mounds (1); Eastern Gulf of Mexico (3); Scotian Basin (8); and Western Gulf of Mexico (7)	40	Illumina (MiSeq, NextSeq 500, HiSeq 1500, HiSeq 2500, HiSeq 1000)	411	¹⁰

	Hadal sediment bulk metagenome [#]	Marina Trench (2), Kermadec Trench (2), Yap Trench (3)	7	Illumina HiSeq 2500, HiSeq X Ten	250	¹¹
	Challenger Deep bulk metagenome	Slope (20) and bottom-axis (17) sites of Challenger Deep in Mariana Trench	37 metagenomes and 3 metatranscriptomes	Illumina MiSeq, Illumina HiSeq 2500,	283	This study

1036 * Estimated by multiplying the number of reads by the average read length

1037 # Shows datasets used for comparisons in this study

1038

1039 **Supplementary references**

- 1040 1. Angly FE, *et al.* The marine viromes of four oceanic regions. *PLOS Biol* **4**, 2121-2131
1041 (2006).
- 1042 2. Labonté JM, Suttle CA. Previously unknown and highly divergent ssDNA viruses
1043 populate the oceans. *ISME J* **7**, 2169-2177 (2013).
- 1044 3. Brum JR, *et al.* Patterns and ecological drivers of ocean viral communities. *Science*
1045 **348**, 1261498 (2015).
- 1046 4. Roux S, *et al.* Ecogenomics and potential biogeochemical impacts of globally
1047 abundant ocean viruses. *Nature* **537**, 689-693 (2016).
- 1048 5. Allen LZ, *et al.* The Baltic sea virome: diversity and transcriptional activity of DNA
1049 and RNA viruses. *mSystems* **2**, e00125-00116 (2017).
- 1050 6. Gregory AC, *et al.* Marine DNA viral macro- and microdiversity from pole to pole.
1051 *Cell* **177**, 1109–1123 (2019).
- 1052 7. Yoshida M, Takaki Y, Eitoku M, Nunoura T, Takai K. Metagenomic analysis of viral
1053 communities in (hado)pelagic sediments. *PLoS One* **8**, e57271 (2013).
- 1054 8. Bäckström D, *et al.* Virus genomes from deep sea sediments expand the ocean
1055 megavirome and support independent origins of viral gigantism. *mBio* **10**, e02497-
1056 02418 (2019).
- 1057 9. Zheng X, *et al.* Extraordinary diversity of viruses in deep-sea sediments as revealed by
1058 metagenomics without prior virion separation. *Environ Microbiol* **23**, 728-743 (2021).
- 1059 10. Li Z, *et al.* Deep sea sediments associated with cold seeps are a subsurface
1060 reservoir of viral diversity. *ISME J* **15**, 2366–2378 (2021).
- 1061 11. Jian H, *et al.* Diversity and distribution of viruses inhabiting the deepest ocean on
1062 Earth. *ISME J* **15**, 3094–3110 (2021).
- 1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075