1	Supplementary figures for
2	
3	Giant viruses as reservoirs of antibiotic resistance genes
4	
5	Xinzhu Yi ^{1,†} , Jie-Liang Liang ^{1,†} , Ping Wen ¹ , Pu Jia ¹ , Shi-wei Feng ¹ , Shen-yan Liu ¹ , Yuan-
6	yue Zhuang ¹ , Yu-qian Guo ¹ , Jing-li Lu, Sheng-ji Zhong ¹ , Bin Liao ² , Zhang Wang ¹ , Wen-
7	sheng Shu ¹ , & Jin-tian Li ^{1,*}
8	
9	¹ Institute of Ecological Science, Guangzhou Key Laboratory of Subtropical Biodiversity
10	and Biomonitoring, Guangdong Provincial Key Laboratory of Biotechnology for Plant
11	Development, School of Life Sciences, South China Normal University, Guangzhou
12	510631, PR China
13	² School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, PR China
14	
15	[†] These authors contributed equally to this work.
16	
17	*Corresponding author:
18	School of Life Sciences, South China Normal University, Guangzhou 510631, PR China
19	E-mail: Jin-tian Li (<u>lijintian@m.scnu.edu.cn</u>)
20	Tel.: +86 20 85211850; Fax: +86 20 85211850
21	



Supplementary Fig. 1 | **Habitat distribution of viral genomes and their antibiotic resistance gene (ARG) contents. A, D:** Habitat distribution of the genomes of A) nucleocytoplasmic large DNA viruses (NCLDVs) and D) phages analyzed in this study. **B, E:** Possibility of ARG-carriage in different habitats for B) NCLDVs and E) phages. **C, F**: Genomic potential of ARG-carriage of C) NCLDVs and F) phages in different habitats. For clarity, only seven major habitats are shown individually, with other habitats being

referred to as "Other" and shown as a whole. Only habitats with more than three genomes are shown in B), C), E), and F). Lower-case letters above the bars in C) and F) represent significantly different groups assessed with two-sided Wilcoxon rank-sum test, and *P*-values indicate the overall difference among all habitats assessed with Kruskal–Wallis test. Data presented in C) and F) were mean values \pm standard error of the mean (SEM). Each data point represents an individual genome in the corresponding group. The number (*n*) of genomes in each group can be found in Source Data file on Github ¹. The y-axis was truncated to zoom in on values below 0.1% in panel F). ORF is the abbreviation for open reading frame, IG is for isolate genome, and MAG is for metagenome-assembled genome.



Supplementary Fig. 2 | Average numbers of ARG-like ORFs detected in different families of A) NCLDVs and B) phages. Data are presented as mean values \pm SEM. The unit of study is one genome. The number (*n*) of genomes in each group can found in Source Data file on Github ¹. Lower-case letters above the bars represent significantly different groups assessed with two-sided Wilcoxon rank-sum test, and *P*-values indicate the overall difference among all families assessed with Kruskal–Wallis test.



Supplementary Fig. 3 | Distribution of antimicrobial resistance gene families in different viral families of A) NCLDVs and B)

phages. All viral families with at least one ARG detected are displayed. Source data are provided as a Source Data file on Github¹.

	1	10 20	30	4 0	50	eo
P00379 <i>E_coli</i> dihydrofolate reductase	MISLIA	ALAVDRVIGMENAMP	W.NLPADLAWFKRNTLN.	KPVIMGRH	TWESIG RPL	PGRKNIILSSOPG
P31074 K. aerogenes dihydrofolate reductase	MISLIA	ALAVDRVIGMENAMP	W. DLPADLAWFKRNTLN	KPVVM <mark>G</mark> RL	TWESIG RPL	PGRKNIVISSKPG
P04174 N. gonorrhoeae dihydrofolate reductase	MLKITIIA	ACAENLCIGAGNAMP	W.HIPEDFAFFKVYTLG	KPVIMGRK	TWESLPVKPL	PGRRNIVISRQAD
P00381 L. casei dihydrofolate reductase	MTAFLW	AQDRDGLIGKDGHLP	W.HLPDDLHYFRAQTVG	KIMVV <mark>G</mark> RR	TYESFPK RPL	PERTNVVLTHQED
P11045 B. subtilis dihydrofolate reductase	MISFIF	AMDANRLIGKDNDLP	W.HLPNDLAYFKKITSG	HSIIM <mark>G</mark> RK	TFESIGRPL	PNRKNIVVTSAPD
P10167 S. aureus dihydrofolate reductase type I	MTLSILV	AHDLQRVIGFENQLP	W.HLPNDLKHVKKLSTG	HTLVM <mark>G</mark> RK	TFESIGKPL	PNRRNVVLTSDTS
P00380 E. faecium dihydrofolate reductase	MFIS M W	AQDKNGLIGKDGLLP	W.RLPNDMRFFREHTMD	KILVM <mark>G</mark> RK	TYEGMGKLSL	PYRHIIVLTTQKD
UTT44570 E. aurantiacum plasmid dfrE	MIISLIA	AISSNYVIGKDKDIP	W.KIPGEQVRFKDLTMG	KSVIM <mark>G</mark> RK	TFESIGQPL	PNRKTIIISKSKD
GVMAG-M-3300023174-193, dfrA16	MELVV	AFAKNGI IGNDNK IP	W.NIPEDLIRFKHMTYG	HVIVM <mark>G</mark> RT	TFESLPNGPL	KN <mark>R</mark> THIVLTRNPK
GVMAG-S-1026894-48, dfrA3	MRFSIIA	AINEQGGIGYNNTIP	W.HVPEDLKHFRKLTMG	KTIIM <mark>G</mark> RN	TWESIGCKPL	IGRKNIVLSSTLP
GVMAG-S-1101181-91, dfrA8	MIINIAV	AITKNGGIGLNGALP	WPHLKGDMALFSKRTTG	LGNNAVLM <mark>G</mark> KN	TWCSIPENRRPL	KNRTNIIISSSLP
GVMAG-M-3300024261-42, dfrA19	MNLIF	ACDKKYGIGIKNKLP	W.KIDNDLARFSKLTIG	NGSNVIIM <mark>G</mark> KN	TYLSLPNNYL	KNRRNIVISQTLF
NM.tailings.bin.2.fa, dfrA20	MMSIRQPTLTLIV	AATARNGIGHNGTLP	W P M L K K E M A Y F A R V T K R V I	JMGARRNA <mark>VIM</mark> GRK	TWESIPPKFRPL	KD <mark>R</mark> TNIVISSQSR
GVMAG-M-3300009149-34, dfrA36	MNIIV	AHCKNRGIGFKNKLP	W.ELSADLERFKQLTIG	NGNNAVIM <mark>G</mark> RN	TWRSLPSRYKPL	PKRENIVLTTEIE
GVMAG-S-3300013093-118, dfrA21	MNPRNF <mark>SIIA</mark>	AFRSDRGIGYENMIP	W. KKSNDLAFFRDETTKDC	CLVGEKNM <mark>VIM</mark> GRC	TFESLNCKPL	KHRINVVVTSQTT
GCA.003233935.1.ASM323393v1, dfrA3	MGALATVPFSIVV	AMTVSRAIGQDGKLP	WGRLPSEMADFRNLTRTTI	DPAKANALIM <mark>G</mark> RL	TFDSLPRR.RPL	PGRINVVLTRRPL
GVMAG-M-3300007236-6, dfrA24	MKDVGIIV	AATTNGGIGYKNALP	W.SIPEELKLFRKITTCVE	NDKKYNC <mark>IIMG</mark> KN	TWHSIPNKPL	KNRVNIIITSNEY
GVMAG-S-1091232-186, dfrA38	MPPLKVSAIL	AIDALGGVARDGKLP	W.SKTWDMRHFVETTRG	CTVVMGRA	TWETLEKPL	PGRSNLVLTRNPD
ERX552261.44.dc, dfrE	MKCIV	AHDNKFGIGKNNKLP	W.NIPEDLKRFRKLTEH	STMIMGSK	TFFSLPINKRPL	PGRKSIVITYEPN
GVMAG-S-1101165-81, dfrA38	MQIDNEIELIV	AFSKNNVLGNQNKLP	W.NIPEDLKRFKDLTTN	HIVVMGRK	TFESLPNGPL	KNRINLVLTNQIT
GX.T.bins.17.fa, dfrE	MALSLFSIVV	AVDMGNGIAKNGEIP	W.YSQEDMRFFRETTMG	KKKNAVIMGRV	TYESIDAKHRPL	EGRYNVVISRTWK
ERX556315.47.dc, dfrE	MIAANRIKFSLIA	AV TNNYGLGFKNGLP	WKKFPVDMKWFQETTTG	HTVIMGRK	TWESLPKTERPL	PNRRNFVLTSNFD
GVMAG-M-3300023174-188, dfrl	MCDFSIIA	AVSKDNGIGKNGELP	W. NIPEDLRFFOKITKITN	IDPNKCNALIMGRN	TFHSIG RPL	PGRLNVCISTSYT
ERX552270.62.dc, dfrA8	MKINIIV	AYCKNNGIGINNNLP	W. SIKSDMKKFKSLTIG	DGNNCVVMGSN	WISLNNKGL	VGRDNLILSKSQI
ERX556019.13.dc, <i>dfrL</i>	MRTFSIIV	AVCNNNGIGINGKLP	W. KNKEDMEFFKNITINTE	INMLKQNIVVMGRN	TFESMNNKTL	LLRKNIVISNNHL
GVMAG-S-1101173-83, dfrA36	MIV <u>NIIV</u>	AYCKNRGIGKNNTLL	M. DIKSDMAKFKKLIVG	NNNNGVIMGRK	TESLNN. IKGL	VNRDNLILSKSLT
	**	* *	** *		*** *	
	70	80	90 100) 11	0 12	o.
P00379 E. coli dihydrofolate reductase	TDDRVT	WVKSVDEAIAA.	.CGDVPEIMVIGGGRVY	EOFL. PK. AOK	LYLTHID AEV	EGD
P31074 K. aerogenes dihydrofolate reductase	SDDRVO	WVKSVDEAIAA.	.CGDAEEIMVIGGGRVY	EOFL. PK. AHK	LYLTHID AEV	EGD
P04174 N. gonorrhoeae dihydrofolate reductase	YCAAGAE	TVASLEVALAL.	.CAGAEEAVIMGGAOIY	GOAMPLATD	LRITEVDLSV	EGD
P00381 L. casei dihydrofolate reductase	YOAOGAV	VVHDVAAVFAY.	. AKOHPDOELVIAGGAOIF	TAFKDDVDT	LLVTRLAGSF	EGD
P11045 B. subtilis dihydrofolate reductase	SEFOGCT	VVSSLKDVLDI.	.CSGPEECFVIGGAOLY	TDLFPYADR	LYMTKIHHEF	EGD
P10167 S. aureus dihydrofolate reductase type I	FNVEGVD	VIHSIEDIYQL.	PGHVFIF <mark>GG</mark> QTLF	EEMIDKVDD	MYITVIEGKF	RGD
P00380 E. faecium dihydrofolate reductase	FKVEKNAE	VLHSIDELLAY.	.AKDIP.EDIYVSGGSRIF	QALLPETKI	IWRTLID AEF	EGD
UTT44570 E. aurantiacum plasmid dfrE	INYNNCL	TVESLERAFNL.	.LQQEDEIFIAGGEIY	KESLPFADR	IYLTIIEKEY	EGN
GVMAG-M-3300023174-193, dfrA16	PSNIPDVV	FVNT.DNLRKTLEK.	.YQKTKKIFLIGGREIY	DLLFDYCEI	FHITLVNAEP	EGN

4174 N. gonorfioeae dihydrofolate reductase YCAA....GAĒTVA...SLEVALAL..CAGA..ELAVIMGGAQIYGAM..PL..ATDLRITEVD..LSVEGD P00381 L. casei dihydrofolate reductase YCAQ....GAVVVH...DVAAVFAY.AKQHPDQELVIAGGAQIYTDLF..PY..ADLLYTRLA..GSFEGD P11045 B. subtils dihydrofolate reductase SEFQ....GCTVVS...SLKDVLDT..CSGP..EECFVIGGAQLYTDLF..PY..ADRLYTKLIA..GSFEGD P10380 E. faecium dihydrofolate reductase SEFQ....GCTVVS...SLKDVLDT...P..GHVFIFGQTFFEEMI..DK..VDDMYITVIE..GKFRGD P00380 E. faecium dihydrofolate reductase FKVE...KNAEVLH..SIDELLAY.AKDIP.EDIYVSGSRIFQALL..PE..TKIIWRTLLD.AEFEGD UTT44570 E. aurantiacum plasmid dffE INYN...NCLTVE..SLERAFNL.LQQE..DEIFIAGGCEIYKESL..PF..ADRIYLTIIE..KEYEGN GVMAG-M-3300023174-193, dfrA16 PSNI...PDVVFVNT.DNLRKTLEK.YQKT..KKIFLIGGREIYDLIF..DY..CEIFHITLVN.AEPEGN GVMAG-S-1026894-48, dfrA3 PSKT...DTLWFSH..SFETVIES.LKHE.DEVFIIGGSSLYKKAMMHPL.CSELYITVVV.SKCECD GVMAG-S-1026894-48, dfrA3 PSKT...DTLWFSH..SFETVIES.LKHE.DEVFIIGGSSLYKKAMMHPL..VNKIYLTVVK.SKCECD GVMAG-M-3300023174-193, dfrA16 PSNI...PDVVFVNT.DNLRKTIEK.YQKT.KKIFLIGGRSLYNDFLNTYCDKINRVYITCVC.SNHDCD GVMAG-S-1026894-48, dfrA3 PSKT...DTLWFSH..SFETVIES.LKHE.DEVFIIGGSSLYKAALDLPQ.TNRILLTRIS.KEYDCD GVMAG-S-1001894-34, dfrA3 PSKT...DTLWFSH.SLEVIES GCAA0323390014934, dfrA36 KPVFSDT.NDTPIWPSH.SLEVIESSLYKAAKDELWIIGGSSLYKAALDLPQ.TNRILLTRIS.KEYDCD GVMAG-M-3300013093-118, dfrA27 WSQSNDY.PDTYFVS.SLDKALGL.NVPNCKINRRFVIGGEKLYCEAIQHPR.CKELINIIIIIIICDVNEISCD GCAA03233935.1ASM3233934, dfrA36 GADT.Y.PEGVLVAG.SLDKALGL.NVPNCKINRRFVIGGEKLYKEAIQHPR.CKELINIIIIIIICDVNEISCD GVMAG-S-1001163-6, dfrA24 KKMKNEVDNDNIVVK.DLQEAINHN.NRTDSIENGFIIGGSQLYKEALKKLNKYVYISIIEV KKYKINVPL GVMAG-S-101136-6, dfrA38 SVEL...GASVVH...TAEALDA.VGAD.GELWVIGGSVIAAVIHPA.CCAIWVQITNLDYPEAD GVMAG-S-101163-81, dfrA38 SNE...SNVYTINM.ENFCTINKK.LNNSD.KKVFIIGGSQLYKEALDHPG.CKELIFKN.DIKYVYISIIEVED GXMAG-S-101163-81, dfrA38 SNE...SNVYTINM.ENFCTINKK.LNNSD.KKVFIIGGSQLYKAALRAKIYVTKKINE ERX552261.44.dc. dfrE DSKFDEYRKTENLFVY...NLDEFMTK.VTNNRGEQITAAERDYMYLCKKIYVTKFK.NDYDCD ERX552261.44.dc. dfrE MSDQ....GAFFP.SLDALNN.I

* Trimethoprim binding sites

Supplementary Fig. 4 | **Sequence alignment of the putative dihydrofolate reductase proteins in NCLDVs and reference sequences in bacteria.** The binding residues of trimethoprim (indicated by purple stars below the alignments) and the accessions of DFR sequences from bacteria were obtained from Coque *et al.*². NCLDV sequence labels are marked blue.



Supplementary Fig. 5 | Mobile genetic elements (MGEs) carried by NCLDVs and phages from different habitats. A, C: Possibility of MGE-carriage of A) NCLDVs and C) phages in different habitats. B, D: Genomic potential of ARG-carriage of B) NCLDVs and D) phages in different habitats. Lower-case letters above the bars represent significantly different groups assessed with two-sided Wilcoxon rank-sum test, and *P*-values indicate the overall difference among all families assessed with Kruskal–Wallis test. Data presented in C) and F) were mean values \pm SEM. Each data point represents an individual genome in the corresponding group. The number (*n*) of genomes in each group can be found in the Source Data file on Github ¹. For clarity, only seven major habitats are

shown individually, with other habitats being referred to as "Other". Only habitats with at least three genomes are shown.



Supplementary Fig. 6 | **Dependency between ARG-carriage and the carriage of different MGE types in the NCLDV genomes. A:** Overall incidence of individual MGE types. **B:** Possibility of ARG-carriage in genomes with or without a given MGE type. Twosided Chi-squared test was performed to test dependency between carriage of a given MGE type and ARG-carriage. **C:** Genomic potential of ARG-carriage in genomes with or without a given MGE type. Two-sided Wilcoxon rank-sum test was performed to evaluate the statistical significance of the differences in ARG content between genomes with or without a given MGE type. MGE+: genomes with at least one MGE belonging to a specific MGE type, MGE-: genomes without any MGE belonging to a specific MGE type. Data are presented as mean values ± SEM. Each data point is one genome. **D:** Histogram of distance

(kb) between ARGs and different types of MGEs. Source data are provided as a Source Data file on Github ¹.



Supplementary Fig. 7 | Dependency between ARG-carriage and the carriage of different MGE types in the phage genomes. A: Overall incidence of individual MGE types. B: Possibility of ARG-carriage in genomes with or without a MGE type. Two-sided Chi-squared test was performed to test dependency between carriage of a MGE type and ARG-carriage. C: Genomic potential of ARG-carriage in genomes with or without a MGE type. Two-sided Wilcoxon rank-sum test was performed to evaluate the statistical significance of the differences in ARG content between genomes with or without a MGE type. MGE+: genomes with at least one MGE belonging to a specific MGE type, MGE-: genomes without any MGE belonging to a specific MGE type. Data are presented as mean values \pm SEM. Each data point is one genome. D: Histogram of distance (kb) between

ARGs and different types of MGEs. Source data are provided as a Source Data file on Github ¹.



Supplementary Fig. 8 | Correlations between genomic potential of ARG carriage of isolated viruses and their genome size. Spearman coefficients are shown as heatmap in A) for isolated NCLDVs and B) for isolated phages. Colors represent Spearman's *rho*. Stars and cross symbols indicate significance of the Spearman correlation, which are FDR-adjusted for multiple comparisons. Only viral families with more than three ARG-carrying genomes and the overall patterns of NCLDVs and phages are displayed. Exact *n* and *P*-values for individual groups can be found in the Source Data file on Github ¹.

Figure S9



Supplementary Fig. 9 | Phylogenetic tree of dihydrofolate reductases from NCLDVs, eukaryotes, prokaryotes, and phages. This tree is the same as that in Figure 3A in the main text, with trimethoprim sequences from NCLDVs functionally validated in this study marked with stars.



Supplementary Fig. 10 | **Maximum-likelihood phylogenetic tree of NCLDVs.** Reference genome sequences were adopted from a previous publication ³ (https://faylward.github.io/GVDB). The tree was constructed using concatenated alignments of seven marker genes SFII, RNAPL, PolB, TFIIB, TopoII, A32 and VLTF3, employing the LG+I+F+G4 model and rooted with *Poxviridae*. Branches in different colors represent taxonomic information of the giant viruses at the order level. The giant virus genomes assembled in this study are highlighted in pink font. The outermost circle represents the taxonomic information of the giant viruses at the family level. The marker genes comprise: SFII (DEAD/SNF2-like helicase), RNAPL (DNA-directed RNA polymerase alpha subunit), PolB (DNA polymerase family B), TFIIB (transcription

initiation factor IIB), TopoII (DNA topoisomerase II), A32 (Packaging ATPase), and VLTF3 (Poxvirus late transcription factor VLTF3). For better presentation, branches not involving any genomes in this study are folded.

Supplementary references

- 1 Yi, X., *et al.* Giant viruses as reservoirs of antibiotic resistance genes. *Github.* <u>https://doi.org/10.5281/zenodo.13234118</u> (2024).
- 2 Coque, T. M., Singh, K. V., Weinstock, G. M. & Murray, B. E. Characterization of dihydrofolate reductase genes from trimethoprim-susceptible and trimethoprim-resistant strains of *Enterococcus faecalis*. *Antimicrob*. *Agents Chemother*. **43**, 141-147 (1999).
- 3 Aylward, F. O., Moniruzzaman, M., Ha, A. D. & Koonin, E. V. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol.* **19**, e3001430 (2021).