nature portfolio

Corresponding author(s): Jin-tian Li

Last updated by author(s): Aug 13, 2014

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The NCLDV and phage genome sequences were sourced from the Giant Virus Database (https://faylward.github.io/GVDB/, accessed on 2022/10/19) and the CheckV (v1.5) database (https://portal.nersc.gov/CheckV/, accessed on 2024/03/18) respectively. To construct gene trees, orthologs of the NCLDV ARGs in prokaryotes and eukaryotes were collected from the eggNOG (v5.0) database and NCBI database.
Data analysis	To annotate ARGs, the following databases and softwares are used: DeepARG (v1.0.2), CARD (https://card.mcmaster.ca/, accessed on 2024.03.10), SARG (v3.2.1-S, https://smile.hku.hk/ARGs/Indexing, accessed on 2024.03.14), NCBI NDARO (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/, accessed on 2024.03.19), SARGfam (https://smile.hku.hk/SARGs, accessed on 2024.03.21), and Reference HMM Catalog of the NCBI NDARO platform (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/, accessed on 2024.03.19). To annotate MGEs, the following databases and softwares are used: ISEScan (v1.7.2.3), CONJscan (v2.0.1), Phage_Finder (v2.1), ICEberg (v2.0), DRAM (v1.3.5) implemented with KEGG and PFAM databases. VFDB (http://www.mgc.ac.cn/VFs, accessed on 2022.06.07) is used for VF annotation. Other softwares for general bioinformatics analysis include MEGAHIT (v1.2.9), Bowtie (v2.4.5), Samtools (v1.15.1), MetaBAT2 (v 2.12.1), dRep (v3.3.0), IQ-TREE (v2.1.2), VirSorter2 (v2.2.3), geNomad with genomad_db_v1.7, Perl (v5.16.3), diamond (v2.1.8.162), HMMER (v3.1b2), Prodigal (v2.6.3), MAFFT (v7.490) ,trimAl (v1.4),ParaAT (v2.0), KaKs (v3.0) ,ESPript 3, Phyre2. Statistical analysis and plotting are performed using R software (v4.1.0), and the codes used in this study are available on GitHub (https://doi.org/10.5281/zenodo.13234118).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Metagenomic sequencing data for the tailings samples used in this study have been deposited in the NCBI BioProject database under the accession number PRJNA1085405. The giant virus sequences and phage contigs generated from tailings metagenomes in this study have been deposited in the ENA Sequence Read Archive database under the accession number PRJEB74361 and PRJEB78842, respectively. Source data of this paper are provided on Github (https://doi.org/10.5281/zenodo.13234118).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This study does not involve human participants or human data.
Reporting on race, ethnicity, or other socially relevant groupings	This study does not involve human participants or human data.
Population characteristics	This study does not involve human participants or human data.
Recruitment	This study does not involve human participants or human data.
Ethics oversight	This study does not involve human participants or human data.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study investigates the prevalence and diversity of antibiotic resistance gene (ARG) sequences within NCLDV genomes and explores their evolutionary relationships with ARGs in phages, bacteria, archaea, and eukaryotes. We gathered viral genomes from public databases, representing a wide range of viral taxonomy and habitats, and supplemented this with additional viral genomes assembled from mine tailings metagenomes collected by our team. Our analysis focuses on the distribution of ARGs across different viral taxonomies and habitats, in the context of ARG quantity, composition, and resistance mechanisms. To understand why giant viruses harbor ARGs, we constructed gene trees of specific ARGs across viruses, prokaryotes, and eukaryotes, and analyzed correlations between ARGs, mobile genetic elements, and virulence factors in viral genomes. Additionally, we conducted microbial experiments to evaluate the function of selected ARG sequences from giant viruses. We used a comparative design with three groups: E. coli transformed with plasmids carrying each of the two distinct giant virus genes and a control group with empty plasmids. Each group had 8 independent replicates to ensure robust and reliable results. This experimental design facilitates direct comparison of the antibiotic resistance expression between the two giant virus genes and the control, highlighting differences in functional outcomes.
Research sample	The majority of genome sequences used in this study were sourced from public databases, including 1,383 NCLDV genomes from Aylward's previous publication and 39,689 phage genomes from the CheckV database. These datasets were selected to cover genomes from nearly all currently available cultured isolates and representative metagenome-assembled genomes (MAGs) across diverse global habitats. To further broaden the range of habitats, we included and assembled 115 metagenomic datasets from mine tailings, sampled from 39 locations with three replicates each. This additional data generated 33 NCLDV MAGs and 588 phage genomes from the mine tailings habitat. No specific sample size was predetermined, as these datasets represent the most comprehensive collections of virus genomes currently available.
Sampling strategy	The metagenomes of tailings have been published in a previous study. The sample sites were determined to cover diverse mine sites across China, representing different provinces and climatic conditions. At each mine site, three mine tailings samples were collected from a drained tailings pond (i.e. an abandoned tailings disposal site) at a depth of 0–20cm using a stainless steel trowel. Specifically,

	three separate plots (1 × 1 m, at least 10 m apart) were designated in each pond, with one tailings sample collected from each plot.
Data collection	The NCLDV and phage genome sequences were sourced from the Giant Virus Database (https://faylward.github.io/GVDB/) and the CheckV database (https://portal.nersc.gov/CheckV/) by Xinzhu Yi. Tailings metagenomic sequencing was conducted by Guangdong Magigene Biotechnology Co., Ltd, with assembly performed by Shi-wei Feng using the bioinformatics pipeline detailed in the manuscript. Cloning, transformation, and microbial antimicrobial susceptibility tests were carried out by Shen-yan Liu, Yuan-yue Zhuang and Yu-qian Guo as described in the manuscript.
Timing and spatial scale	Field sampling of the mine tailings was conducted in July and August 2018, with each site being sampled once. The sampling sites cover a wide range of latitude and longitude (22° 8′ 19″ N–48°15′ 54″ N, 86° 19′ 47″ E–29° 17′ 29″ E).
Data exclusions	A total of 117 tailings samples were collected and subjected to DNA extraction, resulting in 115 samples meeting the quality criteria for metagenomic sequencing. The entirety of the sequenced data was utilized for the assembly of virus genomes.
Reproducibility	This study primarily hinges on rigorous bioinformatics analysis. To ensure the reproducibility of our findings, meticulous organization and preservation of all codes were maintained. Additionally, microbial experiments on antimicrobial susceptibility were conducted at least three times by different students. Each experimental batch consisted of eight replicates per group, yielding consistent results across all batches.
Randomization	In the microbial antimicrobial susceptibility experiments, each group of E. coli was exposed to a different plasmid—either containing a giant virus gene or an empty control plasmid. All other factors, including strain, vector, growing conditions, and antibiotic concentrations, were kept consistent across the groups. The allocation process ensured a balanced distribution of these factors, emphasizing comparability between the experimental groups.
Blinding	For the antimicrobial susceptibility test, individuals involved in the experiments were blinded to the experiment group to ensure unbiased measurements.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The field data have been published previously. Briefly, the 39 mine sites covered wide geographical locations across China. The climatic conditions of them also varied considerably, with mean annual precipitation (MAP) of 25–1917mm and mean annual temperature (MAT) of −0.09 to 22.8°C.
Location	The sampled mine sites covered a wide range of latitude and longitude (22° 8′ 19″ N–48°15′ 54″ N, 86° 19′ 47″ E–29° 17′ 29″ E).
Access & import/export	Import/export is not relevant to our study.
Disturbance	No disturbance due to the small sample amount.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not relevant to this study.
Novel plant genotypes	Not relevant to this study.
Authentication	Not relevant to this study.