

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure 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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	To retrieve marine-derived Autographiviridae genomes, approximately 7 million UViGs were downloaded from IMG/VR v4, Global Ocean Viromes (GOV and GOV 2.0), the MedDCM fosmid library, Station ALOHA assembly-free virus genomes 11 and the ALOHA 2.0 virome, San Pedro Ocean Time-series Viromes, and Red Sea Viromes, 34 viral fosmids obtained from the oxic surface (10 m) and oxygen-starved basin (200 m) waters of Saanich Inlet. A total of 220 marine viromic datasets, including Global Ocean Viromes, Pearl River estuary virome, Mariana Trench virome, Eastern Tropical North Pacific virome, Delaware Bay and Chesapeake Bay viromes, Black Sea virome, Red Sea virome, South China Sea DNA virome, were used for viromic read-mapping analysis.
Data analysis	Prodigal v2.6.3, MAFFT v7.505, HMMER v3.3.2, CheckV v0.9.0, CD-HIT v4.8.1, BLAST v2.11.0, OrthoFinder v2.5.4, Easyfig v2.2.2, RaFAH v0.3, ViPTree v1.1.3, trimAl v1.4.rev15, IQ-TREE v2.2.0.3, Interactive Tree Of Life (iTOL) v6, coverm v0.6.1

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](#)

This study did not generate any new data.

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

NA

Reporting on race, ethnicity, or other socially relevant groupings

NA

Population characteristics

NA

Recruitment

NA

Ethics oversight

NA

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](https://www.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 employed bioinformatics methods to recover 1,253 complete marine Autographiviridae UViGs from public databases, revealing 14 phylogenomic groups (including 6 novel ones) with conserved core genes and distinct genomic/ecological characteristics. Metagenomic analyses demonstrated their global distribution in tropical/temperate oceans, which correlated with host niches, underscoring their ecological impact on marine microbial communities. No biological samples were collected, and no new experimental data were generated in this study, as all analyses were performed using publicly available datasets.

Research sample

Approximately 7 million UViGs were downloaded from IMG/VR v4, Global Ocean Viromes (GOV and GOV 2.0), the MedDCM fosmid library, Station ALOHA assembly-free virus genomes 11 and the ALOHA 2.0 virome, San Pedro Ocean Time-series Viromes, and Red Sea Viromes, 34 viral fosmids obtained from the oxic surface (10 m) and oxygen-starved basin (200 m) waters of Saanich Inlet for Autographiviridae UViG recovery. A total of 220 marine viromic datasets, including Global Ocean Viromes, Pearl River estuary virome, Mariana Trench virome, Eastern Tropical North Pacific virome, Delaware Bay and Chesapeake Bay viromes, Black Sea virome, Red Sea virome, South China Sea DNA virome, were used for viromic read-mapping analysis.

Sampling strategy

NA

Data collection

The data were downloaded from the publicly available sources as specified in the published papers, with this task being performed by Sen Du and Zefeng Zhang.

Timing and spatial scale

The UViGs and marine viromic datasets were downloaded in March 2023. This study utilized the principal reference datasets in marine virome research, representing the most comprehensive and authoritative collections in the field.

Data exclusions

No data were excluded.

Reproducibility

Using the publicly available datasets specified in this study and applying the identical bioinformatics pipeline described in our

Reproducibility	Methods section, we successfully replicated the reported results.
Randomization	NA
Blinding	NA

Did the study involve field work? ☐ Yes ☒ No

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	NA
Novel plant genotypes	NA
Authentication	NA