nature portfolio

Corresponding author(s): Kim M. Handley

Last updated by author(s): Apr 7, 2026

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	Metagenomics and metatranscriptomics sequencing data were previously generated within our group. No software was used to retrieve this data. The IMG/VR database was downloaded via globus from the following website: https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=IMG_VR Taxonomy information from the viralRefSeq database (v223) was downloaded directly from the following website (accessed 16 Sept 2023): https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&SourceDB_s=RefSeq Taxonomy information from the ICTV database (ICTV_Master_Species_List_2022_MSL38.v3) was downloaded directly from the following website (accessed 16 Sept 2023): https://ictv.global/taxonomy
Data analysis	Sequence data processing: Trimmomatic v0.38 (metagenomics data) and v0.39 (metatranscriptomic data); SortMeRNA v4.3.6 (metatranscriptomic data); metaSPAdes v3.11.1 (metagenomics data). Prokaryote genome recovery and analysis: CONCOCT v0.4.1; MetaBAT v2.12.1; MaxBin v.2.2.4; BBMap v37.93; DAS_Tool v1.1.1; dRep v1.4.3; CheckM v1.2.1; GTDB-Tk v2.4.0 (database v214); DRAM v1.3.5; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341. Virus identification: VIBRANT v1.2.1; VirSorter2 v2.2.3; deepVirFinder v1.0; virome_per_sample_derep.py (custom script, available at https://

zenodo.org/doi/10.5281/zenodo.19197341); Cluster_genomes_5.1.pl; CheckV v0.7.0; DRAM-v v1.4.6; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341.

vOTU taxonomy and Caudoviricetes core genes phylogeny: vConTACT2 v2.0.11.3; Cytoscape v3.8.2; HMMR v3.3.2; Clustal Omega v1.2.4; IQ-TREE v2.2.2.2; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341.

vOTU ecosystem clustering analyses: vConTACT2 v2.0.11.3; IQ-TREE v2.2.2.2; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341

vOTU host identification: crass v1.0.1; BLAST v2.13.0; CRISPRDetect v2.4; Aragorn v1.2.38; VirHostMatcher v1.0.0; RaFAH v0.3; HostG (accessed 06 Dec 2021); additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341.

Spatial distribution and diversity: BBMap v39.01; featureCounts (subread v2.0.6); R (v4.4.0) packages gplots, vegan, ggplot2; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341.

Antiphage defence mechanisms: CRISPRDetect v2.4; PADLOC v2.0.0 (database v2.0.0); R (v4.4.0) packages vegan, ggplot2, gplots; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341.

Gene GC and protein amino acid proportions and isoelectric point: DRAM v1.4.6; pepstats (EMBOSS v6.6.0); R (v4.4.0) packages stats, ggplot2, ggpmisc; additional custom scripts available at https://zenodo.org/doi/10.5281/zenodo.19197341.

Host tRNA analyses: DRAM v1.3.5 (prokaryote data); DRAM v1.4.6 (vOTU data); BBMap v39.01; featureCounts (subread v2.0.6); R (v4.4.0) packages gplots, vegan, ggplot2, ggpmisc.

Protein structural prediction and surface electrostatic potential: DRAM v1.4.6; Alphafold 3 server; UCSF ChimeraX v1.9.

Closely related vOTUs: minimap2 v2.28; BLAST v2.13.0; DRAM-v v1.4.6; R (v4.4.0) package gggenomes.

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

Sequence data used in this study are available in the NCBI database under BioProject accession code PRJNA668816 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA668816]. Further analyses were conducted on data retrieved from the IMG/VR and viralRefSeq databases, as outlined in the methods. Additional data generated in this study are provided in the Supplementary Tables and Source Data files.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	Field data: To sample microbial communities across an estuarine salinity gradient, nine sites were sampled along a 5-km long salinity gradient in the Waiwera river and estuary (2 sites freshwater, 5 sites brackish, and 2 sites marine) with samples at each site comprising water and underlying sediment, as detailed previously by our group (DOI: 10.1186/s40168-021-01145-3). Database sampling: viral genomic and taxonomic data were obtained from IMG/VR and viralRefSeq versions as detailed below.
Research sample	Samples from each site comprised: (1) biomass filtered on-site from at least 10 L of water collected from each of the nine sites (from 0.2 and 1 m below the surface) on to mixed cellulose ester filters (0.22 micrometers); and (2) benthic sediment (top 2 cm) homogenized and sieved to a particle size of <= 1mm to remove detritus, and collected per site in triplicate (~5 m apart). Samples were filtered/sieved upon collection at each site, and immediately placed in Lifeguard. In total there were nine filtered biomass samples from water and 27 sediment samples.
Sampling strategy	Sufficient samples numbers to detect reproducible trends were estimated in advance based on prior estuarine sampling and analysis of heterogeneity (e.g., DOI:10.1111/1462-2920.15550).
Data collection	Site sampling data were recorded in field notebooks as sampling occurred by KMH, HST, and DW. Viral data from databases were obtained by MH from viralRefSeq version 211, viralRefSeq taxonomy version 223, and IMG/VR version 7.1.
Timing and spatial scale	Sampling was conducted on the 12th of November 2018 along a 5-km long salinity gradient in the Waiwera river and estuary.
Data exclusions	No data were excluded.
Reproducibility	Samples were collected along a salinity gradient with multiple samples per salinity zone and sample type (water/sediment), including spatial replicates for sediment at each sampling site. Findings on genetic traits in viruses were verified by analysis of viruses in the global IMG/VR dataset, and comparison.
Randomization	Taxa were assigned to groups based on habitat found in (salinity zone and water column or benthic sediment). Co-variance of nutrient availability and salinity is acknowledged in the manuscript.
Blinding	No blinding was used. Predictions necessitating blinding (e.g. of sampling habitat from genomic data) were not a study goal.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Approximately 18-19 degrees Celsius and clear weather (sunny; no wind)
Location	Waiwera River, Auckland, New Zealand (36°32'29.3"S, 174°42'37.0"E)
Access & import/export	Sampling occurred in public waterways with no permitting required from the local authorities
Disturbance	Undisturbed water and sediment samples were collected

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	The study did not involve laboratory animals
Wild animals	The study did not involve wild animals
Reporting on sex	Not applicable. The study is based on prokaryotes and viruses.
Field-collected samples	Sediment and filtered biomass were stored in the laboratory at -80 degrees Celsius the day of collection.
Ethics oversight	No ethical approval or guidance was required. The study subject matter is unmodified prokaryotes and viruses.

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*