nature research

Corresponding author(s): Martin Polz, Libusha Kelly

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Experiments comparing Autolykiviridae and tailed viruses included either all available members meeting quality control requirements (Infection assay presented in Fig. 2, Fig. 3D, and Extended Data Fig. 6) as limited by recovery during initial sampling protocol (described in the the Methods in the section "Isolation, culturing, and sequencing of bacteria and viruses"), or one selected representative (Chloroform Assay, Fig. 3A; Density, Fig. 3B; Protease Assay, Fig. 3C) of each of the major diversity groups of the Autolykiviridae and the three tail morphotypes for the Caudovirales. No statistical methods were used to predetermine sample sizes.

2. Data exclusions

Describe any data exclusions.

The criteria for inclusion & exclusion of viruses and hosts in the presented infection analyses are described in the methods section on pages 25-26, in the section headed "Characterization of Autolykiviridae host range". The infection dataset presented in Fig. 2 and Extended Data Fig. 6 includes 247 viruses, excluded were viruses from the original dataset that did not infect their host of isolation again in the large scale host range assay or that did not derive from independent plaques in the original isolation. For statistical comparisons of infections of Autolykiviridae and tailed viruses, the 241 sequenced viruses were included, with four sequenced Autolykiviridae excluded from infection analyses because they represent either genomically-identical sublineages of a member included in the analyses (1.107.A, 1.107.B, and 1.249.B), or because they did not infect their original host of isolation in the large-scale host range assay (1.095.O).

3. Replication

Describe whether the experimental findings were reliably reproduced.

Large scale infection assay (Fig. 2, Extended Data Fig. 6) - The large scale infection assay included 3 replicates of each interaction; the entire experiment was performed once as described. Observations in subsequent smaller-scale host range assays with members of these collections have been consistent with those described here.

Chloroform assay (Fig. 3a) - The chloroform assay included 3 replicates of each interaction and the 5 Autolykiviridae included in these experiments are representative of the diversity of this group and represent biological replicates; the entire experiment was performed once as described in the manuscript. Observations in subsequent similar experiments are consistent with those described here.

Density gradient (Fig. 3b) - The density gradient determination was performed once as described in the manuscript, the 5 Autolykiviridae included in these experiments are representative of the diversity of this group and represent biological replicates. Observations in subsequent similar experiments are consistent with those described here.

Protease treatment (Fig. 3c) - The protease treatment comparisons were performed in three separate experiments, each with a single replicate of each virus and treatment, a representative gel from one of these experiments is shown; the 5

Autolykiviridae included in these experiments are representative of the diversity of this group and represent biological replicates.

Infection timing (Fig. 3d) - The large scale infection assay included 3 replicates of each interaction; the entire experiment was performed once as described. Observations in subsequent similar experiments are consistent with those described here.

Decay assay (inline in Main & Methods) - The decay assay included 4 replicates of each virus; the entire experiment was performed once as described.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

All experiments comparing Autolykiviridae and tailed viruses included either all available members meeting quality control requirements (Infection assay presented in Fig. 2, Fig. 3D, and Extended Data Fig. 6), or one selected representative (Chloroform Assay, Fig. 3A; Density, Fig. 3B; Protease Assay, Fig. 3C) of each of the major diversity groups of the Autolykiviridae and the three tail morphotypes of the Caudovirales. Sample order was haphazardly assigned for each of the three independent replicates of the protease assay (Fig. 3C), position in each of three plate sectors was haphazardly assigned for each of three virus lysate replicates in the large scale infection assay (Fig. 2, Extended Data Fig. 6), otherwise samples were not randomized for the experiments.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Large scale infection assay (Fig. 2, Extended Data Fig. 6) - Results of the host range assay were performed blinded insofar as 1) they were recorded without reference to position of the three haphazardly assigned replicates on a given assay plate, and 2) a large number of assays with different sets of viruses were recorded at the same time reducing likelihood of pattern detection. This is briefly indicated in the methods.
In no other experiments were investigators blinded to group allocation during data collection or analyses.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Commercial and publicly available open-source software were used to analyze data in this study, these are indicated in their associated sections in the Methods, and include: BLASTp v.2.2.29+, CLC Assembly Cell v.4.4.2.133896, CLC Genomics Workbench v.8.5.1, Clustal Omega (EMBL-EBI web portal), EggNOG-Mapper v.4.5.1, ETE v.3.0.0b36 (implementing Clustal Omega, trimAl, MUSCLE, PhyML

v.3.0, MAFFT v5, M-Coffee, Gephi v.0.9.1, HHpred (MPI Bioinformatics Toolkit webportal), hmmer v.3.1b2, ImageJ, InterProScan v.5.17-56.0, iTOL v.4, MAFFT, MCL v.14.137, NCBI Batch Web Conserved Domain search tool, PhyML v.3.0 with SMS v.1.8.1, Phye2 webportal, Prodigal v.2.6.1 and v.2.6.3, Python with package NetworkX v.1.1.10, RAxML, R v.3.3.0 with packages GenoPlotR, data.table, ggplot2, cowplot, igraph, rgexf, lme4.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Bacteria and virus strains described as isolated in this work are available from the authors upon request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The study did not involve human research participants.