

From hepatitis misdiagnosis to zoonotic false alarms: a metagenomic blacklist framework for the parvo-like hybrid viral group

Peng Zhao,¹ Hong Liu,² Jie Dong,¹ Haoxiang Su,¹ Qi Jin,¹ Fan Yang^{1,3,4}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

KEYWORDS metagenomic sequencing, parvo-like hybrid viruses (PHVs), reagent contamination, false-positive signals, silica membrane spin columns, pathogen surveillance bias, PVDDC

Metagenomic sequencing has become a cornerstone of pathogen discovery and emerging infectious disease surveillance, enabling unbiased detection of known and unknown viruses directly from clinical, animal, and environmental samples (1–5). However, the same sensitivity that underpins these approaches also increases susceptibility to laboratory-derived nucleic acid contaminants, which may be misinterpreted as evidence of novel viral circulation or cross-species transmission (6, 7). Such artifacts pose a particular challenge for early-warning systems, where spurious signals can complicate biological interpretation and downstream risk assessment.

Parvo-like hybrid viral (PHV) sequences exemplify this problem. One representative strain, NIH-CQV, was initially reported as a candidate etiological agent of seronegative hepatitis (8) but was subsequently shown to originate from nucleic acid contaminants embedded in silica-based extraction reagents (7, 9, 10). Despite this clarification, related PHV sequences continue to be detected in metagenomic data sets from diverse wildlife and environmental sources and are frequently annotated as putative infectious or zoonotic viruses, sustaining uncertainty regarding their biological relevance (6, 11).

Here, we provide systematic evidence that PHV sequences detected by metagenomics represent reagent-derived artifacts rather than genuine infectious viruses. Using the Panoramic Virus Discovery Data Chain (PVDDC) developed by us (6), we curated 93 global PHV genomes with detailed metadata, including 10 newly assembled sequences recovered directly from silica membranes of 29 commercial nucleic acid extraction kits representing diverse manufacturers and intended applications (Table S1). Phylogenetic analysis resolved two major clades spanning substantial sequence divergence (Fig. 1a); however, neither clade exhibited ecological structure consistent with host taxonomy, sampling context, or geographic origin. Instead, sequences recovered from extraction reagents were interspersed throughout both clades alongside PHVs previously attributed to birds, fish, plants, and other hosts (Fig. 1b).

Multiple PHVs reported as wildlife-associated viruses showed implausibly high nucleotide identity (>90%–95%) to sequences obtained directly from extraction reagents. Such extreme sequence conservation across highly divergent host taxa is incompatible with established principles of viral evolution and host adaptation and cannot be readily explained by repeated natural spillover events (Fig. 1c). Similar contamination-driven host misassignments have been documented for other parvovirus-related sequences (6), suggesting that PHVs represent a prominent manifestation of a broader technical artifact affecting metagenomic virome studies.

Editor Uladzimir Karnychuk, The Ohio State University College of Veterinary Medicine, Columbus, Ohio, USA

Address correspondence to Haoxiang Su, suruemail@ipbcams.ac.cn, Qi Jin, zdsys@vip.sina.com, or Fan Yang, yangfan@ipbcams.ac.cn.

Peng Zhao, Hong Liu, and Jie Dong contributed equally to this article. Author order was determined based on the duration of time spent working on this specific project.

The authors declare no conflict of interest.

See the funding table on p. 3.

Published 9 April 2026

Copyright © 2026 Zhao et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

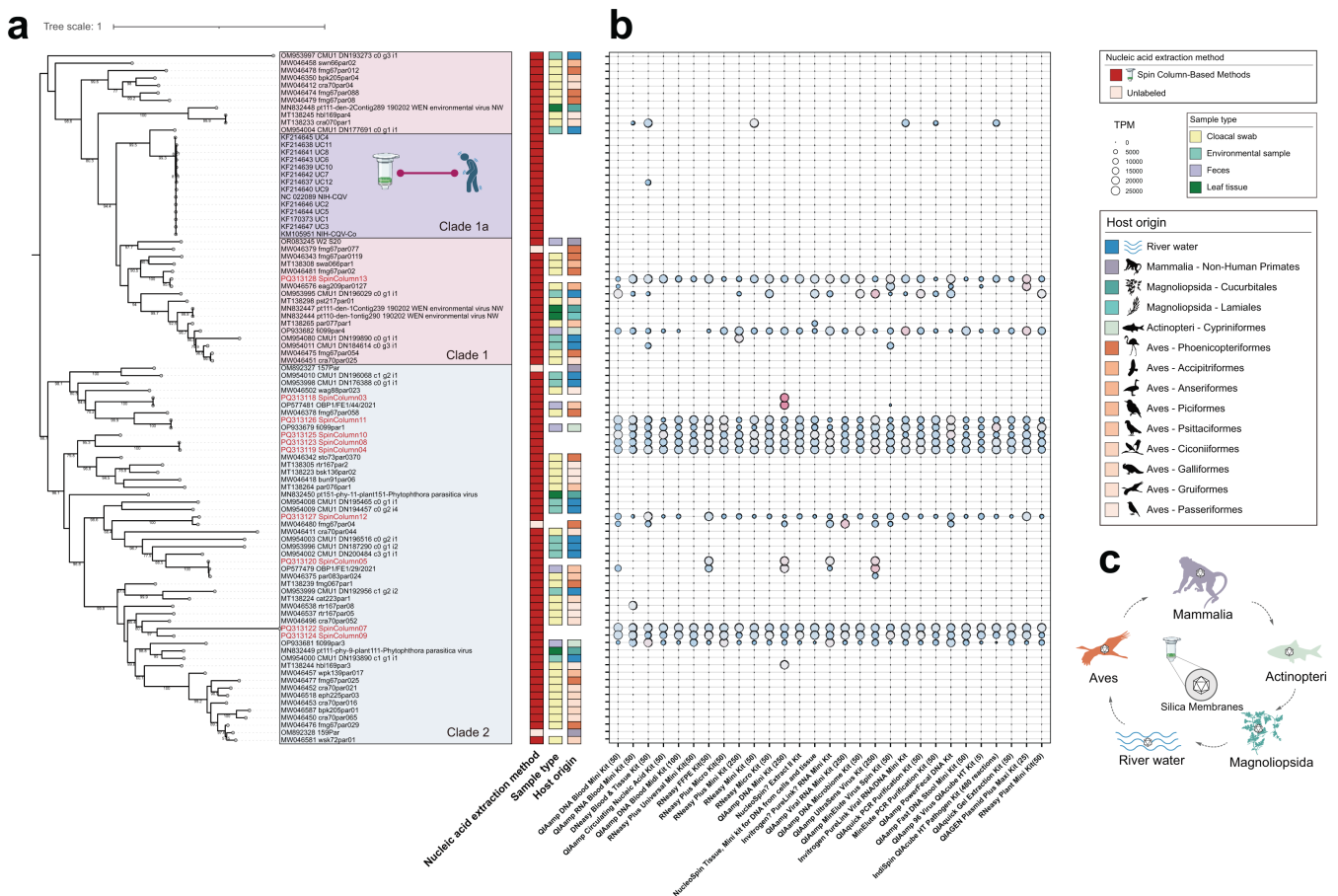


FIG 1 Evidence that parvo-like hybrid viral sequences are reagent-derived contaminants. (a) Maximum-likelihood phylogeny of parvo-like hybrid viral (PHV) sequences inferred from NS amino acid alignments using the LG+R5 substitution model with midpoint rooting. PHV sequences newly identified in this study are highlighted in red. Colored strips indicate the nucleic acid extraction method, biological sample type, and reported host or source associated with each sequence. (b) Bubble plot showing the abundance of PHVs detected in metagenomic data sets generated from silica membranes of 29 commercial spin column-based nucleic acid extraction kits. (c) Conceptual illustration highlighting the ecological implausibility of interpreting PHV sequences as evidence of cross-species viral infections across highly divergent host taxa.

Consistent with this interpretation, metagenomic profiling of silica membranes from 29 commercial spin column-based nucleic acid extraction kits revealed widespread and recurrent detection of PHV sequences independent of biological sample input (Fig. 1b). Several PHVs were detected in the majority of kits tested, indicating propagation at industrial scale through reagent supply chains. Detection of PHVs in sequencing blanks further confirms that laboratory consumables alone are sufficient to generate these signals (7, 9, 10).

Taken together, the absence of ecological coherence, the implausibly high sequence identity across highly divergent hosts, and the pervasive association with nucleic acid extraction reagents demonstrate that the vast majority of PHV sequences reported to date represent laboratory-derived artifacts rather than bona fide infectious viral entities. While we cannot completely exclude the possibility that rare PHV-related sequences may have a biological origin, our data indicate that the signals currently dominating public databases and metagenomic studies arise from reagent contamination. Continued interpretation of these sequences as infectious or zoonotic viruses risks propagating non-biological records in virology databases and confounding metagenomic-based pathogen discovery and emerging infection surveillance. We therefore recommend that PHV sequences be interpreted with caution and flagged as probable reagent-derived artifacts until definitive biological hosts are identified.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Project no. 825B2064), the Science & Technology Fundamental Resources Investigation Program of China (Grant no. 2024FY1009001), and the CAMS Innovation Fund for Medical Sciences (CIFMS; Grant no. 2021-I2M-1-039).

AUTHOR AFFILIATIONS

¹NHC Key Laboratory of Systems Biology of Pathogens, National Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

²Shandong Provincial Research Center for Bioinformatic Engineering and Technique, School of Life Sciences and Medicine, Shandong University of Technology, Zibo, China

³Key Laboratory of Respiratory Disease Pathogenomics, Chinese Academy of Medical Sciences, Beijing, China

⁴State Key Laboratory of Respiratory Health and Multimorbidity, Beijing, China

AUTHOR ORCID*s*

Peng Zhao  <http://orcid.org/0009-0007-5578-5262>

Hong Liu  <http://orcid.org/0000-0002-5182-4750>

Haoxiang Su  <http://orcid.org/0009-0004-0317-7933>

Qi Jin  <http://orcid.org/0000-0002-3586-6344>

Fan Yang  <http://orcid.org/0009-0000-0138-9271>

FUNDING

Funder	Grant(s)	Author(s)
CAMS Innovation Fund for Medical Sciences	2021-I2M-1-039	Fan Yang
Science & Technology Fundamental Resources Investigation Program of China	2024FY1009001	Fan Yang
National Natural Science Foundation of China	825B2064	Peng Zhao

AUTHOR CONTRIBUTIONS

Peng Zhao, Formal analysis, Investigation, Resources, Software, Visualization | Hong Liu, Investigation, Methodology, Validation | Jie Dong, Investigation, Methodology, Validation | Haoxiang Su, Investigation, Methodology, Resources, Validation | Qi Jin, Conceptualization, Resources, Supervision | Fan Yang, Conceptualization

DATA AVAILABILITY

The metagenomic sequencing data from the silica membrane virome have been deposited in the NCBI Sequence Read Archive under BioProject accession number [PRJNA1158863](#). All sequences generated in this study have been submitted to GenBank under accession numbers [PQ313118–PQ313128](#). Details of library construction are described in our previous study (6), and the phylogenetic analysis and metagenomic data processing are provided in the Supplemental material.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material ([Spectrum00157-26-s0001.docx](#)). Supplemental methods related to data analysis.

Table S1 (Spectrum00157-26-s0002.xlsx). Overview of the 29 commercial nucleic acid extraction kits analyzed in this study.

REFERENCES

1. Harvey E, Holmes EC. 2022. Diversity and evolution of the animal virome. *Nat Rev Microbiol* 20:321–334. <https://doi.org/10.1038/s41579-021-00665-x>
2. Liang G, Bushman FD. 2021. The human virome: assembly, composition and host interactions. *Nat Rev Microbiol* 19:514–527. <https://doi.org/10.1038/s41579-021-00536-5>
3. He T, Jin M, Cui P, Sun X, He X, Huang Y, Xiao X, Zhang T, Zhang X. 2024. Environmental viromes reveal the global distribution signatures of deep-sea DNA viruses. *J Adv Res* 57:107–117. <https://doi.org/10.1016/j.jare.2023.04.009>
4. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>
5. Zhang XA, Ma YD, Zhang YF, Hu ZY, Zhang JT, Han S, Wang G, Li S, Wang X, Tang F, Liang WJ, Yuan HX, Zhao JQ, Jiang LF, Zhang L, Si GQ, Peng C, Wang R, Ge HH, Li N, Jiang BG, Li C, Li H, Liu W. 2024. A new orthonairovirus associated with human febrile illness. *N Engl J Med* 391:821–831. <https://doi.org/10.1056/NEJMoa2313722>
6. Zhao P, Dong J, Liu H, Xue Y, Zhang T, Sui H, Hu Y, Du J, Su H, Jin Q, Yang F, Data Research and Curation Team. 2025. Unveiling origin pitfalls in metagenomic surveillance for emerging infectious diseases: parvoviruses as a model. *iMetaMed* 1:e70004. <https://doi.org/10.1002/imm3.7004>
7. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Delwart EL, Chiu CY. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol* 87:11966–11977. <https://doi.org/10.1128/JVI.02323-13>
8. Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, Kajigaya S, Zhao K, Mao Q, Young NS. 2013. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. *Proc Natl Acad Sci USA* 110:10264–10269. <https://doi.org/10.1073/pnas.1303744110>
9. Naccache SN, Hackett J, Delwart EL, Chiu CY. 2014. Concerns over the origin of NIH-CQV, a novel virus discovered in Chinese patients with seronegative hepatitis. *Proc Natl Acad Sci USA* 111:E976. <https://doi.org/10.1073/pnas.1317064111>
10. Smuts H, Kew M, Khan A, Korsman S. 2014. Novel hybrid parvovirus-like virus, NIH-CQV/PHV, contaminants *in silico* column-based nucleic acid extraction kits. *J Virol* 88:1398–1398. <https://doi.org/10.1128/JVI.03206-13>
11. Dai Z, Wang H, Wu H, Zhang Q, Ji L, Wang X, Shen Q, Yang S, Ma X, Shan T, Zhang W. 2022. Parvovirus dark matter in the cloaca of wild birds. *Gigascience* 12:giad001. <https://doi.org/10.1093/gigascience/giad001>