Infection, Genetics and Evolution 112 (2023) 105456

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Abundant dsRNA picobirnaviruses show little geographic or host association in terrestrial systems

Matthew A. Knox [a,*,1], Janelle Wierenga [a,1], Patrick J. Biggs [a,b], Kristene Gedye [a], Valter Almeida [a], Richard Hall [c], Gladys Kalema-Zikusoka [d], Stephen Rubanga [d], Alex Ngabirano [e], Willy Valdivia-Granda [f], David T.S. Hayman [a]

[a] School of Veterinary Science, Massey University, New Zealand
[b] School of Natural Sciences, Massey University, New Zealand
[c] Ministry for Primary Industries, New Zealand
[d] Conservation Through Public Health, Uganda
[e] Bwindi Development Network, Uganda
[f] Orion Integrated Biosciences, USA

ARTICLE INFO

Keywords:
Bioinformatics
Gorilla
Metagenomics
Picobirnaviridae
Uganda

ABSTRACT

Picobirnaviruses are double-stranded RNA viruses known from a wide range of host species and locations but with unknown pathogenicity and host relationships. Here, we examined the diversity of picobirnaviruses from cattle and gorillas within and around Bwindi Impenetrable Forest National Park (BIFNP), Uganda, where wild and domesticated animals and humans live in relatively close contact. We use metagenomic sequencing with bioinformatic analyses to examine genetic diversity. We compared our findings to global *Picobirnavirus* diversity using clustering-based analyses. *Picobirnavirus* diversity at Bwindi was high, with 14 near-complete RdRp and 15 capsid protein sequences, and 497 new partial viral sequences recovered from 44 gorilla samples and 664 from 16 cattle samples. Sequences were distributed throughout a phylogenetic tree of globally derived picobirnaviruses. The relationship with *Picobirnavirus* diversity and host taxonomy follows a similar pattern to the global dataset, generally lacking pattern with either host or geography.

1. Introduction

Viruses are ubiquitous on earth, infecting all cellular organisms (Paez-Espino et al., 2016) and possibly, even the majority of cells (Munson-McGee et al., 2018). Global virus numbers are a staggering 1×10^{31} (Editorial, 2011), and diversity is also extraordinarily high. Between 40,000 (Carlson et al., 2019) and 320,000 (Anthony et al., 2013) virus species infect mammals alone, while bacteriophages or phages, viruses that infect bacteria and archaea, also exhibit very high species diversity (Camarillo-Guerrero et al., 2021; Keen, 2015). Difficulties in estimating global viral diversity arise because the vast majority of viruses are unknown. Complete genomic sequence information is limited for all domains of life (Rhie et al., 2021; Zhang et al., 2020) but is particularly lacking for viruses. Currently, 9110 viral taxa are recognised by the International Committee on Taxonomy of Viruses (ICTV) (Walker et al., 2021), a tiny fraction of total diversity (Delwart, 2007), with a bias towards those that cause disease in humans and domesticated animals and plants. However, recent developments in sequencing technology, i.e. shotgun metagenomics, which allow 'unbiased' amplification of viral nucleic acids, have improved detection rates and accelerated the discovery of new species (Roux et al., 2021) and recent studies using metagenomic sequencing routinely identify a high proportion of unknown viral taxa (Camarillo-Guerrero et al., 2021; Cebriá-Mendoza et al., 2021; Guajardo-Leiva et al., 2020; Zheng et al., 2021). Virus species vary in host range, with some known from a single host species e.g., smallpox, while others can infect multiple taxa e.g., rabies virus. Increasing knowledge of viromes from different environments and hosts enhances understanding of the range of habitats and hosts where virus species exist and may identify previously unknown reservoirs or transmission pathways. Ultimately, this will allow research into disease emergence and the ecological and evolutionary processes that lead to cross-species transmission events (Lloyd-Smith et al., 2009; Pepin et al.,

* Corresponding author.
E-mail address: m.knox@massey.ac.nz (M.A. Knox).
[1] These authors contributed equally to this article.

https://doi.org/10.1016/j.meegid.2023.105456
Received 3 April 2023; Received in revised form 8 May 2023; Accepted 25 May 2023
Available online 29 May 2023
1567-1348/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2010). A suitable virus group to study these processes would be abundant, widespread, diverse, and found across a wide range of host taxa, all such attributes are evident in the family *Picobirnaviridae* (Perez et al., 2021).

Picobirnaviruses were first discovered in 1988 and possess a bisegmented double-stranded RNA (dsRNA) genome (Pereira et al., 1988). Picobirnaviruses are dsRNA viruses with short, but variable segmented genomes with either a large or small genome profile. The size of the genomic segments for the large genome profile is 2.3–2.6 kb for segment 1 and 1.5–1.9 kb for segment 2, respectively (Rosen et al., 2000). Segment 2 contains the RNA-dependent RNA polymerase (RdRp). The genus is divided into at least two genogroups based on nucleotide sequence similarities in the RdRp (segment 2). The small genome profile is 1.75 and 1.55 kb for segment 1 and 2, respectively (Bhattacharya et al., 2007). Overall, most studies have found higher quantities of genogroup I as compared to genogroup II of segment 2 (Perez et al., 2021).

Picobirnaviruses are typically found in faecal or wastewater transcriptome studies (Buzinaro et al., 2003; Cascio et al., 1996; Symonds et al., 2009; Zhang et al., 2015), but also in respiratory and alimentary tract samples (Smits et al., 2012; Woo et al., 2019). It is, however, unclear if they are associated with disease in animal hosts (Smits et al., 2016) or indeed whether they are even truly mammalian viruses with recent evidence suggesting a bacterial or fungal host (Ghosh and Malik, 2021; Krishnamurthy and Wang, 2018). The high degree of genetic diversity between *Picobirnavirus* sequences from different taxa has hampered alignment-dependent phylogenetic analyses (Knox et al., 2018; Perez et al., 2021). However, it is clear they are abundant and ubiquitous and have been found in many species and locations (Kashnikov et al., 2020; Woo et al., 2016). Despite this, any phylogenetic association with host is presently unclear (Ghosh and Malik, 2021).

Picobirnaviruses are frequently detected in metagenomic studies (Chong et al., 2019; Guajardo-Leiva et al., 2020; Hull et al., 2020; Ramesh et al., 2021; Woo et al., 2014) and their abundance and ubiquitous nature suggest they must play a major role in ecosystems, but given the lack of any clear biological or geographical associations within or between related *Picobirnavirus* taxa, it is unclear what this role may be. Therefore, such abundant organisms might allow the following questions to be asked: Is there biogeographic structure in *Picobirnavirus* populations? Is there host association among *Picobirnavirus*? Can ubiquitous and diverse dsRNAs be used as markers of host switching to address questions about pathogen emergence? Here, we aim to address those questions using published and unpublished *Picobirnavirus* viral fragments from 35 countries, including new data from people, cattle and mountain gorillas in a location with high potential for cross-species transmission among mammals in Uganda (Muylaert et al., 2021).

2. Methods and results

2.1. Ethics

Ethical approval was obtained through the Massey University Human Ethics Committee, application number Southern A Application 14–44.

2.2. Sample collection

Sample collection was undertaken by collaborators in Uganda, Conservation through Public Health (CTPH) between 2014 and 2015. Faecal samples were collected from mountain gorillas (*Gorilla beringei beringei*), domesticated cattle (*Bos taurus africanus*) and people from the region of Bwindi Impenetrable Forest National Park (BIFNP) in southwestern Uganda (hereafter, Bwindi). Habituated gorilla groups are tracked daily by CTPH and the faecal samples were collected from the night nests, cattle faeces were from the cattle in and around Buhoma village, and anonymous diagnostic human faecal samples were from the Bwindi Community Hospital, the healthcare clinic serving Buhoma. The health status of the gorillas and cattle were unknown, but no signs of illness observed. The human samples were from clinical diagnostic samples, but were anonymous and the symptoms and signs were not provided. The study site and species were selected to investigate cross-species transmission as part of a wider research project. In Bwindi, wildlife (including gorilla) is found in proximity to humans and livestock (e.g., cattle) and research has demonstrated contact between these groups (Muylaert et al., 2021). Faeces from all species were collected and placed into 5 mL cryovials containing RNAlater® (Sigma, MO, USA) and stored in a −20 °C freezer until shipment and shipped on ice. In total, three anonymous faecal samples from clinically unwell humans with unknown symptoms, 16 cattle faecal samples and 44 mountain gorilla faecal samples were imported into New Zealand and were stored until analysis at −80 °C in a PC2 containment facility at the Hopkirk Research Institute, Massey University, New Zealand.

2.3. Laboratory

The extraction of nucleic acids from the collected faeces was performed following the protocol as described by Hall et al. (2014). Approximately 200 mg of faeces was placed into a 2 mL Safelock tube (Eppendorf, Hamburg, Germany) with 1 mL of 0.01 M PBS pH 7.3 and incubated for one hour at 4 °C. The suspended faeces were centrifuged at 6000 xg for 5 min and the supernatant passed through a sterile 0.45 μm filter (Macherey-Nagel GmbH & Co. KG, Düren, Germany) and 200 μL was used for nucleic acid extraction with a Roche High Pure Viral Nucleic Acid Kit (Roche, Basel, Switzerland). DNA was removed with Ambion DNA-free™ (ThermoFisher, Massachusetts, USA) following the manufacturer's instructions. First-strand cDNA synthesis was performed using Invitrogen SuperScript™ III (ThermoFisher, Massachusetts, USA). The resulting single-stranded DNA was used as the template for whole transcriptome amplification using the QuantiTect Whole Transcriptome kit with the substitution of Superscript III for the reverse transcriptase (RT) enzyme (Qiagen, Hilden, Germany; ThermoFisher, Massachusetts, USA). Quantification of cDNA was performed using the DNA Qubit kit (Thermofisher, Massachusetts, USA). Aliquots of 5 μg of whole transcriptome and whole genome amplified nucleic acids were dried down in a hooded cabinet under negative pressure at room temperature overnight in a DNAstable® plate (Biomatrica, California, USA) and sent overnight in room temperature conditions via courier to Orion Integrated Biosciences Inc. (Manhattan, Kansas, USA) for next generation sequencing (NGS) on an Illumina® MiSeq with Nextera XT DNA Library Prep (Illumina®, California, USA). Raw sequencing data was deposited in NCBI (BioProject PRJNA949793, SAMN33959666-SAMN33959728).

2.4. Bioinformatics

Following quality control checks and trimming, sequences were assembled with Megahit (1.1.4) (Li et al., 2015), using the default settings. Diamond BLASTX (Diamond 2.0.9) was used to identify picobirnaviruses by comparing the assembled contigs to the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database, downloaded in June 2022. False positive results were avoided by using sequence similarity cut-off e-values of 1E-5 for the database (Buchfink et al., 2021). Then, the contigs of interest were saved for further analysis in Geneious 10.2.6. First ORFs were identified (minimum ORF size 300 nt, standard genetic code, assuming start and stop codons outside sequence) and translated to amino acids. Spurious ORF translations were removed by running blastp (protein-protein BLAST) sequences against *Picobirnavirus* (taxid:104394) and *Picobirnaviridae* (taxid:585893) and non-hits were removed.

2.5. Analysis 1. Diversity of study picobirnavirus sequences

Due to the high levels of sequence diversity within *Picobirnavirus* and

difficulties with alignment-based approaches, especially with partial sequences, we chose to investigate diversity using a clustering approach. To approximate species-level diversity, amino acid sequences were grouped at 90% similarity using the CD-HIT webserver (Huang et al., 2010) using the default algorithm and alignment coverage parameters (see supplementary fig. 1) and clusters were treated as biological units in subsequent analyses.

2.6. Analysis 2. Comparison with NCBI data and global patterns of geographic and host diversity

To investigate linkages with previously described picobirnaviruses we conducted a NCBI search for *Picobirnavirus* protein sequences (n = 4634, range 18–752 aa). Sequence metadata (host, country) for all accessions was recorded for later analyses. The combined list of sequences (n = 5795) was run through CD-HIT at 90% similarity threshold to generate the clusters of highly similar sequences which we treat as units of *Picobirnavirus* diversity in our analyses. We were able to identify sequences from other studies which clustered with our sequences. Based on CD-HIT clusters, we summarize the current extent of knowledge on *Picobirnavirus* diversity and explore patterns in host and location within *Picobirnavirus* clusters with high sequence similarity. We visualised these relationships with chord diagrams created in R using igraph (Csardi and Nepusz, 2006). Initially, we examined the relationship between host (family) and continent among all *Picobirnavirus* clusters in our dataset excluding those from host families with $n < 5$ sequences (31 sequences removed). Then, to visualise relationships within *Picobirnavirus* clusters identified with CD-HIT, we also created chord diagrams using data from clusters with 10 or more sequences (n = 999 sequences, 44 clusters) using clusters and host (Family level classification) and clusters and continent.

2.7. Analysis 3: Phylogenetic analysis on near complete RdRp sequences

Several near full-length polymerase (RdRp) and capsid protein sequences (>500 aa) from our data were recovered and *RdRp* sequences were compared with representatives from the NCBI nr database. We chose the same representatives used in a recent analysis of viral (including *Picobirnavirus*) diversity (Chen et al., 2022) as well as some from a recent metagenomic study (Ramesh et al., 2021) which had sequences closely matching sequences in our study. We chose a partitivirus (AVM41706.1, Heterobasidion partitivirus 20, *Partitiviridae*) as an outgroup for the maximum likelihood analyses. The sequences were aligned using MAFFT (v7) employing the *E*-INS-i algorithm (Katoh and Standley, 2013). Sequence alignments were then trimmed using trimAL with the gappyout setting to remove any ambiguously aligned regions (Capella-Gutiérrez et al., 2009). The best-fit model of amino acid substitution was determined using smart model selection in PhyML (Lefort et al., 2017). Phylogenetic trees were subsequently inferred using the maximum likelihood approach (ML) implemented in PhyML version 3.0 (Guindon et al., 2010), employing Subtree Pruning and Regrafting (SPR) branch-swapping. Branch support was estimated using an approximate likelihood ratio test (aLRT) with the Shimodaira–Hasegawa-like procedure implemented in PhyML.

3. Results

Our initial analyses returned 1444 nucleotide sequences (200–3049 bp, n = 239 > 1000 bp) identified from the metagenomic assembled contigs using Diamond. Following translation based on ORFs, we were left with 1161 sequences between 98 and 627 aa which have been deposited in NCBI (Accessions OP904296- OP905455). Of these, 497 were recovered from 44 gorilla samples, 664 from the 16 cattle samples and none from the three human samples. At least one PBV was present in 35/44 (79%, 65–91% 95% confidence interval, CI) gorilla, 16/16 (100%, 79–100% 95% CI) cattle and 0/3 (0%, 0–71% 95% CI) human samples. The presence of PBV was not significantly different between the cattle and gorilla ($\chi^2 = 2.4$, p-value = 0.12), but was among all three with PBV less in people ($\chi^2 = 16.6$, p-value <0.001). We recovered 29 sequences >500 aa (near complete). Of these, 14 were RdRp (3 cattle, 11 gorilla) and 15 capsid (13 cattle, 2 gorilla), and overall 16 were from cattle, and 13 from gorilla samples.

3.1. Analysis 1. Diversity of study picobirnavirus sequences

We used a clustering analysis to examine similarities within the set of sequences. At 90% similarity, 910 clusters were formed from the 1161 sequences. 830 sequences did not cluster with anything else. The remaining 331 sequences formed 80 clusters made up of between 2 and 41 sequences. We observed no evidence of cross species transmission. Of the clusters, 65 were from gorilla only, and 15 from cattle only. Picobirnaviruses were shared within species. Based on our clustering analysis, 90% similar *Picobirnavirus* sequences were present in up to 20/44 gorilla individuals. Sharing appeared to be more common among gorilla samples. For example, 37 *Picobirnavirus* clusters were shared by 3 or more individual gorillas and only one cluster was shared among 3 or more cattle.

3.2. Analysis 2. Comparison with NCBI data and global patterns of geographic and host diversity

CD-HIT analyses formed 3428 clusters from the 5795 sequences. Similar to the previous analysis, many sequences (2727) did not cluster with anything else. The remaining 3066 sequences formed 702 clusters of between 2 and 98 sequences. We observed a high number of clusters with intercontinental distributions. Of 702 clusters identified, 202 were made up of sequences from between two and five different continents. Hosts were also commonly shared by *Picobirnavirus* clusters. Samples with unknown host information or from wastewater or environmental samples were excluded and 152/702 clusters had between two and six different host species. *Picobirnavirus* sharing did not appear to be related to host taxonomy, 138 clusters were made up of hosts from different families, 109 with different suborders, 85 with different orders.

Picobirnaviruses have been identified from every continent except Antarctica, with the bulk of sequences coming from Asia, Africa and North America (Fig. 1). In terms of host taxa, *Bovidae*, *Suidae* and *Hominidae* are well represented, along with the *Cercopithidae* (Old World monkeys), *Sciuridae* (squirrels) and wastewater (Fig. 1). Among the most well represented clusters (CD-HIT clusters with 10 or more sequences) the pattern is similar. Host sharing among different families is common (supplementary fig. 2). Interestingly, clusters from *Bovidae* (cloven-hooved ruminants, including cattle) and *Sciuridae* are less well represented, suggesting under sampling in these families. Asia hosts the majority of well-represented clusters, and clusters are frequently distributed across multiple continents (supplementary fig. 3).

Our sequences were also found to be similar to sequences from a wide range of hosts and locations. In all, 17 clusters from our analyses were grouped with NCBI sequences at >90% similarity. These NCBI sequences included 12 countries across six continents and 12 host species (supplementary fig. 4; supplementary Table 1). In some instances, the clusters shared a single host species in multiple countries, though more commonly the clusters were comprised of multiple host taxa (see supplementary fig. 3).

3.3. Analysis 3: Phylogenetic analysis on near complete RdRp sequences

The phylogenetic analysis of *Picobirnavirus* RdRp sequences >500 aa length showed that our sequences clustered with others throughout the *Picobirnavirus* phylogenetic tree (Fig. 2). Genogroups I and II were both represented within our set of sequences. Interestingly, many of our sequences cluster with other non-genogroup 1/2 *Picobirnavirus* RdRp sequences. Branch support for the main clades was poor, but strong within

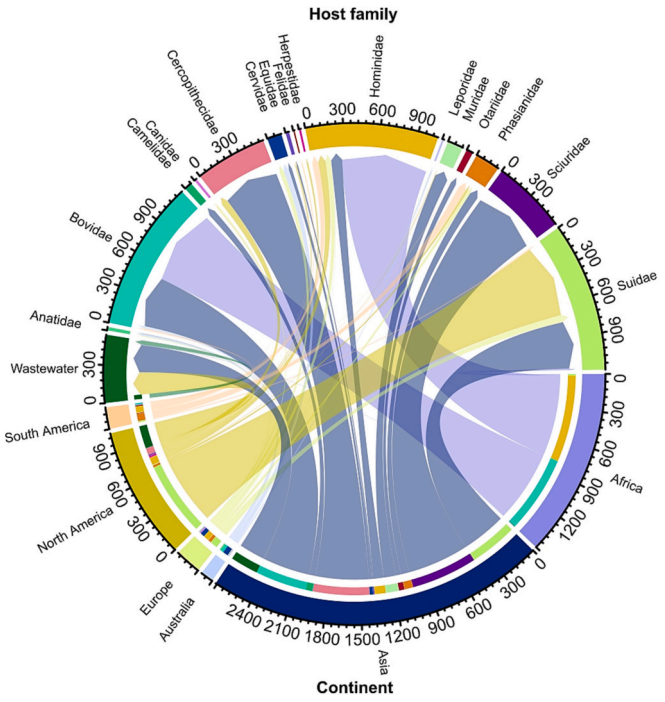

Fig. 1. Summary of global *Picobirnavirus* host family diversity and continent of sampling. Inner circle colours on lower part of the diagram correspond to host family colours.

clusters.

4. Discussion

Picobirnaviruses are frequently detected in metagenomic studies throughout the world, with seemingly little pattern in diversity among highly similar sequences in terms of host or geography (Ramesh et al., 2021; Woo et al., 2016). This, as well as other evidence (Krishnamurthy and Wang, 2018) has led several authors to question whether picobirnaviruses are truly pathogens of eukaryotes, a debate which has yet to be resolved (Ghosh and Malik, 2021; Wang, 2022). We report a high diversity of *Picobirnavirus* sequences in cattle and gorilla faecal samples from Bwindi Impenetrable Forest National Park in Uganda, a system where humans, domestic and wild animals intersect, and a potential hotspot for cross species transmission events (Muylaert et al., 2021; Wilkinson et al., 2018).

Despite the high diversity, our sampling did not detect any evidence for cross species transmission among the animals in our study system. Infection prevalence in our Ugandan samples was not significantly different between the cattle ($n = 16$) and gorilla ($n = 44$), but the absence of infection in the three people was significant (albeit from a very small sample size ($n = 3$)) and future studies could aim to better estimate the prevalence in humans. Sharing within a host species was detected, indicating that the host species plays at least some role in the biology of picobirnaviruses, as has been observed previously (Du et al., 2022; Woo et al., 2016). Virus sharing appeared more common among gorilla hosts, though this may be due to the larger sample size, compared with cattle (and human).

The high levels of sequence diversity and often fragmented nature of metagenomic data creates major difficulties in analysing sequence data from *Picobirnavirus* (Knox et al., 2018; Perez et al., 2021). In addition, since most sequence data from *Picobirnavirus* is comprised of incomplete sequences from different genes (e.g. RdRp, capsid), directly comparing all sequences using phylogenetic methods is impractical, resulting in large amounts of data that is not analysed. The use of clustering programs e.g., CD-HIT, allows for comparison of all sequences which have been classified as a particular taxon. Our chosen clustering threshold of 90% intended to provide an approximation of similarity at the species level, but due to the high number of partial *Picobirnavirus* sequences both in our study and in reference databases, the number of clusters is likely far higher than the number of species. In our study we were careful to use only amino acid sequences from ORFs >300 bp, thus screening out any non-coding or very short regions. This approach may be useful to cluster similar sequences and examine patterns of diversity in other diverse and under-sequenced taxa. In many ways, limitations of CD-HIT are similar to alignment-based phylogenetic methods. Partial sequences are not directly comparable, and the method is reliant on longer sequences to cluster them together. In addition, CD-HIT may not group identical sections of sequences if overlapping regions are not long enough relative to the total sequence length. However, since PBVs have short genomes, this seems to be an issue in a very small proportion of clustered sequences.

Few previous studies have examined picobirnaviruses from multiple species in the same location (though see (Du et al., 2022; Woo et al., 2016)). Therefore, the majority of sequence data in NCBI databases come from a small number of isolated studies on single host species from a single location. Nonetheless our study and analyses reveal that many *Picobirnavirus* sequences both from our study and elsewhere appear closely related to others, often from different studies, and from different hosts and continents. These analyses show that *Picobirnavirus* diversity is not well characterised and that more data is needed to understand any patterns. The high levels of diversity observed in picobirnaviruses may be due to a combination of quickly-evolving site saturation, short RNA viral genomes, and other virus-host (e.g., non-vertebrate) relationships.

Our phylogenetic analyses showed sequences distributed within the currently recognised genogroups within picobirnaviruses based on comparison with a selection of representative sequences (Chen et al., 2022). The majority of near full length RdRp sequences recovered from our study come from outside genogroup I and II, which typically comprise the majority of picobirnaviruses found in metagenomic studies (Teng et al., 2021; Woo et al., 2014). Our divergent *Picobirnavirus* sequences are most closely related to others found previously in North American swine slurry (Rhie et al., 2021) and expand the currently recognised diversity of the family. These suggest that there is much greater *Picobirnaviridae* diversity to be discovered.

Our study revealed a high diversity of novel *Picobirnavirus* sequences present in Ugandan gorilla and cattle. Despite the potential for virus sharing due to interactions between these species in this area (Muylaert et al., 2021), we did not find evidence for cross-species transmission of picobirnaviruses. However, among the global dataset we identified sequence clusters with high similarity, including several of our study sequences. These findings suggest that many *Picobirnavirus* species have wide distributions and broad host ranges and that with further sampling, many more similar linkages are likely, further elucidating patterns of diversity in this under-studied RNA virus family.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2023.105456.

Funding

Royal Society Te Aparangi (Royal Society of New Zealand) [contracts: RDF-MAU1701; MAU1503], the Percival Carmine Chair in Epidemiology and Public Health, Massey University Research Fund and McGeorge Research Grant (all DTSH). WAVG was funded by the U.S. Department of Homeland Security under Grant Award Number #18STCBT00001 for the Cross-Border Threat Screening and Supply Chain Defense (CBTS) and the Department of Homeland Security Award W9-11-SR20F0094-OR-61.

CRediT authorship contribution statement

Matthew A. Knox: Writing – original draft, Formal analysis,

Fig. 2. Maximum likelihood phylogenetic tree of RdRp amino acid sequences from a representative selection of *Picobirnavirus* sequences and host taxa as well as study sequences coloured by host (blue- gorilla, red- cattle). *Picobirnavirus* genogroup information (GI and GII) is indicated with yellow bars. AMV41706.1 is from a *Partitivirus* RdRp gene and has been used as an outgroup for the analyses. Scale bar corresponds to 1 substitution per amino acid. Branch supports of >80% are shown at nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Visualization. **Janelle Wierenga:** Writing – original draft, Formal analysis, Investigation. **Patrick J. Biggs:** Formal analysis, Writing – review & editing, Supervision. **Kristene Gedye:** Investigation, Supervision. **Valter Almeida:** Software. **Richard Hall:** Writing – review & editing, Supervision. **Gladys Kalema-Zikusoka:** Methodology, Investigation. **Stephen Rubanga:** Methodology, Investigation. **Alex Ngabirano:** Methodology, Investigation. **Willy Valdivia-Granda:** Resources. **David T.S. Hayman:** Conceptualization, Writing – review & editing, Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data has been deposited in NCBI (accessions in paper). Additional data used in analyses were also accessed from NCBI and listed in supplementary.

Acknowledgements

Conservation Through Public Health staff for support with sampling. mEpiLab and AgResearch Hopkirk staff, Jamie Lloyd-Smith, Hayley MacGregor and Rebecca French for helpful discussion.

References

Anthony, S.J., Epstein, J.H., Murray, K.A., Navarrete-Macias, I., Zambrana-Torrelio, C. M., Solovyov, A., Ojeda-Flores, R., Arrigo, N.C., Islam, A., Ali Khan, S., Hosseini, P., Bogich, T.L., Olival, K.J., Sanchez-Leon, M.D., Karesh, W.B., Goldstein, T., Luby, S.P., Morse, S.S., Mazet, J.A., Daszak, P., Lipkin, W.I., 2013. A strategy to estimate unknown viral diversity in mammals. mBio 4 e00598–00513.

Bhattacharya, R., Sahoo, G.C., Nayak, M.K., Rajendran, K., Dutta, P., Mitra, U., Bhattacharya, M.K., Naik, T.N., Bhattacharya, S.K., Krishnan, T., 2007. Detection of genogroup I and II human picobirnaviruses showing small genomic RNA profile causing acute watery diarrhoea among children in Kolkata, India. Infect. Genet. Evol. 7, 229–238.

Buchfink, B., Reuter, K., Drost, H.G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368.

Buzinaro, M.G., Freitas, P.P.S., Kisiellius, J.J., Ueda, M., Jerez, J.A., 2003. Identification of a bisegmented double-stranded RNA virus (picobirnavirus) in calf faeces. Vet. J. 166, 185–187.

Camarillo-Guerrero, L.F., Almeida, A., Rangel-Pineros, G., Finn, R.D., Lawley, T.D., 2021. Massive expansion of human gut bacteriophage diversity. Cell 184, 1098–1109. e1099.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Carlson, C.J., Zipfel, C.M., Garnier, R., Bansal, S., 2019. Global estimates of mammalian viral diversity accounting for host sharing. Nat. Ecol. Evol. 3, 1070–1075.

Cascio, A., Bosco, M., Vizzi, E., Giammanco, A., Ferraro, D., Arista, S., 1996. Identification of picobirnavirus from faeces of Italian children suffering from acute diarrhea. Eur. J. Epidemiol. 12, 545–547.

Cebriá-Mendoza, M., Arbona, C., Larrea, L., Díaz, W., Arnau, V., Peña, C., Bou, J.V., Sanjuán, R., Cuevas, J.M., 2021. Deep viral blood metagenomics reveals extensive anellovirus diversity in healthy humans. Sci. Rep. 11, 6921.

Chen, Y.-M., Sadiq, S., Tian, J.-H., Chen, X., Lin, X.-D., Shen, J.-J., Chen, H., Hao, Z.-Y., Wille, M., Zhou, Z.-C., Wu, J., Li, F., Wang, H.-W., Yang, W.-D., Xu, Q.-Y., Wang, W., Gao, W.-H., Holmes, E.C., Zhang, Y.-Z., 2022. RNA viromes from terrestrial sites across China expand environmental viral diversity. Nat. Microbiol. 7, 1312–1323.

Chong, R., Shi, M., Grueber, C.E., Holmes, E.C., Hogg, C.J., Belov, K., Barrs, V.R., 2019. Fecal viral diversity of captive and wild Tasmanian devils characterized using Virion-enriched metagenomics and Metatranscriptomics. J. Virol. 93.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. Int. J. Complex Syst. 1695, 1–9.

Delwart, E.L., 2007. Viral metagenomics. Rev. Med. Virol. 17, 115–131.

Du, H., Zhang, L., Zhang, X., Yun, F., Chang, Y., Tuersun, A., Aisaiti, K., Ma, Z., 2022. Metagenome-assembled viral genomes analysis reveals diversity and infectivity of the RNA Virome of Gerbillinae species. Viruses 14.

Editorial, 2011. Microbiology by numbers. Nat. Rev. Microbiol. 9, 628–628.

Ghosh, S., Malik, Y.S., 2021. The true host/s of Picobirnaviruses. Front. Vet. Sci. 7.

Guajardo-Leiva, S., Chnaiderman, J., Gaggero, A., Diez, B., 2020. Metagenomic insights into the sewage RNA virosphere of a large City. Viruses-Basel 12.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X., Huang, Q.S., Carter, P.E., Peacey, M., 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J. Virol. Methods 195, 194–204.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682.

Hull, J.J.A., Qi, M.P., Montmayeur, A.M., Kumar, D., Velasquez, D.E., Moon, S.S., Magana, L.C., Betrapally, N., Ng, T.F.F., Jiang, B.M., Marthaler, D., 2020. Metagenomic sequencing generates the whole genomes of porcine rotavirus A, C, and H from the United States. PLoS One 15.

Kashnikov, A.Y., Epifanova, N.V., Novikova, N.A., 2020. Picobirnaviruses: prevalence, genetic diversity, detection methods. Vavilovskii Zhurnal Genetiki I Selektsii 24, 661–672.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Keen, E.C., 2015. A century of phage research: bacteriophages and the shaping of modern biology. Bioessays 37, 6–9.

Knox, M.A., Gedye, K.R., Hayman, D.T.S., 2018. The challenges of analysing highly diverse picobirnavirus sequence data. Viruses-Basel 10.

Krishnamurthy, S.R., Wang, D., 2018. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. Virology 516, 108–114.

Lefort, V., Longueville, J.-E., Gascuel, O., 2017. SMS: smart model selection in PhyML. Mol. Biol. Evol. 34, 2422–2424.

Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

Lloyd-Smith, J.O., George, D., Pepin, K.M., Pitzer, V.E., Pulliam, J.R.C., Dobson, A.P., Hudson, P.J., Grenfell, B.T., 2009. Epidemic dynamics at the human-animal Interface. Science 326, 1362–1367.

Munson-McGee, J.H., Peng, S., Dewerff, S., Stepanauskas, R., Whitaker, R.J., Weitz, J.S., Young, M.J., 2018. A virus or more in (nearly) every cell: ubiquitous networks of virus-host interactions in extreme environments. ISME J. 12, 1706–1714.

Muylaert, R.L., Davidson, B., Ngabirano, A., Kalema-Zikusoka, G., MacGregor, H., Lloyd-Smith, J.O., Fayaz, A., Knox, M.A., Hayman, D.T.S., 2021. Community health and human-animal contacts on the edges of Bwindi impenetrable National Park, Uganda. PLoS One 16, e0254467.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., Kyrpides, N.C., 2016. Uncovering Earth's virome. Nature 536, 425-+.

Pepin, K.M., Lass, S., Pulliam, J.R.C., Read, A.F., Lloyd-Smith, J.O., 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nat. Rev. Microbiol. 8, 802–813.

Pereira, H.G., Flewett, T.H., Candeias, J.A.N., Barth, O.M., 1988. A virus with a Bisegmented double-stranded RNA genome in rat (Oryzomys nigripes) intestines. J. Gen. Virol. 69, 2749–2754.

Perez, L.J., Cloherty, G.A., Berg, M.G., 2021. Understanding the Genetic Diversity of Picobirnavirus: A Classification Update Based on Phylogenetic and Pairwise Sequence Comparison Approaches (Viruses-Basel 13.).

Ramesh, A., Bailey, E.S., Ahyong, V., Langelier, C., Phelps, M., Neff, N., Sit, R., Tato, C., DeRisi, J.L., Greer, A.G., Gray, G.C., 2021. Metagenomic characterization of swine slurry in a north American swine farm operation. Sci. Rep. 11.

Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B.J., Chaisson, M., Gedman, G.L., Cantin, L.J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S.C., Lama, T. M., Grutzner, F., Warren, W.C., Balakrishnan, C.N., Burt, D., George, J.M., Biegler, M.T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A.F., Franchini, P., Detrich, H.W., Svardal, H., Wagner, M., Naylor, G.J.P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B.T., Pesout, T., Houck, M., Misuraca, A., Kingan, S.B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R.E., Putnam, N.H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R.E., Guan, D., London, S.E., Clayton, D.F., Mello, C.V., Friedrich, S.R., Lovell, P.V., Osipova, E., Al-Ajli, F.O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R.S., Makova, K.D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K., Martin, F., Howe, K., Flicek, P., Walenz, B.P., Kwak, W., Clawson, H., Diekhans, M., Nassar, L., Paten, B., Kraus, R.H.S., Crawford, A.J., Gilbert, M.T.P., Zhang, G., Venkatesh, B., Murphy, R.W., Koepfli, K.-P., Shapiro, B., Johnson, W.E., Di Palma, F., Marques-Bonet, T., Teeling, E.C., Warnow, T., Graves, J.M., Ryder, O.A., Haussler, D., O'Brien, S.J., Korlach, J., Lewin, H.A., Howe, K., Myers, E.W., Durbin, R., Phillippy, A.M., Jarvis, E.D., 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592, 737–746.

Rosen, B.I., Fang, Z.Y., Glass, R.I., Monroe, S.S., 2000. Cloning of human picobirnavirus genomic segments and development of an RT-PCR detection assay. Virology 277, 316–329.

Roux, S., Matthijnssens, J., Dutilh, B.E., 2021. Metagenomics in virology. Encycl. Virol. 133–140.

Smits, S.L., van Leeuwen, M., Schapendonk, C.M.E., Schurch, A.C., Bodewes, R., Haagmans, B.L., Osterhaus, A., 2012. Picobirnaviruses in the human respiratory tract. Emerg. Infect. Dis. 18, 1539–1540.

Smits, S.L., Osterhaus, A., Koopmans, M.P., 2016. Newly identified viruses in human gastroenteritis pathogens or not? Pediatr. Infect. Dis. J. 35, 104–107.

Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic viruses in wastewater samples from the United States. Appl. Environ. Microbiol. 75, 1402–1409.
Teng, J.L.L., Wernery, U., Wong, P.C., Chan, E., Lee, H.H., Joseph, S., Bai, R., Tang, Y., Wong, E.Y.M., Lau, S.K.P., Woo, P.C.Y., 2021. High prevalence of Genogroup I and Genogroup II Picobirnaviruses in dromedary camels. Viruses 13.
Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Alfenas-Zerbini, P., Davison, A.J., Dempsey, D.M., Dutilh, B.E., Garcia, M.L., Harrach, B., Harrison, R.L., Hendrickson, R.C., Junglen, S., Knowles, N.J., Krupovic, M., Kuhn, J. H., Lambert, A.J., Lobocka, M., Nibert, M.L., Oksanen, H.M., Orton, R.J., Robertson, D.L., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D.B., Suzuki, N., Van Dooerslaer, K., Vandamme, A.M., Varsani, A., Zerbini, F.M., 2021. Changes to virus taxonomy and to the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2021). Arch. Virol. 166, 2633–2648.
Wang, D., 2022. The enigma of picobirnaviruses: viruses of animals, fungi, or bacteria? Curr. Opin. Virol. 54, 101232.
Wilkinson, D.A., Marshall, J.C., French, N.P., Hayman, D.T.S., 2018. Habitat fragmentation, biodiversity loss and the risk of novel infectious disease emergence. J. R. Soc. Interface 15.
Woo, P.C.Y., Lau, S.K.P., Teng, J.L.L., Tsang, A.K.L., Joseph, M., Wong, E.Y.M., Tang, Y., Sivakumar, S., Bai, R., Wernery, R., Wernery, U., Yuen, K.Y., 2014. Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. Virology 471, 117–125.
Woo, P.C.Y., Teng, J.L.L., Bai, R., Wong, A.Y.P., Martelli, P., Hui, S.W., Tsang, A.K.L., Lau, C.C.Y., Ahmed, S.S., Yip, C.C.Y., Choi, G.K.Y., Li, K.S.M., Lam, C.S.F., Lau, S.K. P., Yuen, K.Y., 2016. High diversity of Genogroup I Picobirnaviruses in mammals. Front. Microbiol. 7.
Woo, P.C.Y., Teng, J.L.L., Bai, R., Tang, Y., Wong, A.Y.P., Li, K.S.M., Lam, C.S.F., Fan, R. Y.Y., Lau, S.K.P., Yuen, K.Y., 2019. Novel picobirnaviruses in respiratory and alimentary tracts of cattle and monkeys with large intra- and inter-host diversity. Viruses-Basel 11.
Zhang, S.W., Bai, R., Feng, R., Zhang, H.X., Liu, L.X., 2015. Detection and evolutionary analysis of picobirnaviruses in treated wastewater. Microb. Biotechnol. 8, 474–482.
Zhang, Z., Wang, J.N., Wang, J.L., Wang, J.J., Li, Y.Z., 2020. Estimate of the sequenced proportion of the global prokaryotic genome. Microbiome 8.
Zheng, X., Liu, W., Dai, X., Zhu, Y., Wang, J., Zhu, Y., Zheng, H., Huang, Y., Dong, Z., Du, W., Zhao, F., Huang, L., 2021. Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation. Environ. Microbiol. 23, 728–743.