Genomic and phylogenetic features of the *Picobirnaviridae* suggest microbial rather than animal hosts

Sabrina Sadiq[1], Edward C. Holmes[1,2]*, Jackie E. Mahar[1]

[1]Sydney Institute for Infectious Diseases, School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.
[2]Laboratory of Data Discovery for Health Limited, Hong Kong SAR, China.

*Corresponding Author:
Prof. Edward C. Holmes,
Sydney Institute for Infectious Diseases, School of Medical Sciences,
The University of Sydney,
Sydney, NSW 2006,
Australia.
Email: edward.holmes@sydney.edu.au

Keywords: *Picobirnaviridae*, phylogeny, cross-species transmission, bacteriophage, classification, taxonomy

ABSTRACT

The RNA virus family *Picobirnaviridae* has traditionally been associated with the gastrointestinal systems of terrestrial mammals and birds, with the majority of viruses detected in animal stool samples. Metatranscriptomic studies of vertebrates, invertebrates, microbial communities, and environmental samples have resulted in an enormous expansion of the genomic and phylogenetic diversity of this family. Yet picobirnaviruses remain poorly classified, with only one genus and three species formally ratified by the International Committee of Virus Taxonomy. Additionally, an inability to culture picobirnaviruses in a laboratory setting or isolate them in animal tissue samples, combined with the presence of bacterial genetic motifs in their genomes, suggests these viruses may represent RNA bacteriophage rather than being associated with animal infection. Utilising a data set of 2,286 picobirnaviruses sourced from mammals, birds, reptiles, fish, invertebrates, microbial communities, and environmental samples, we identified seven consistent phylogenetic clusters likely representing picobirnavirus genera that we tentatively name *Alpha-, Beta-, Gamma-, Delta-, Epsilon-, Zeta-,* and *Etapicobirnavirus*. A statistical analysis of topological congruence between virus-host phylogenies revealed more frequent cross-species transmission than any other RNA virus family. In addition, bacterial ribosomal binding site motifs were more enriched in picobirnavirus genomes than in the two groups of established RNA bacteriophage – the *Leviviricetes* and *Cystoviridae*. Overall, our findings support the hypothesis that the *Picobirnaviridae* have bacterial hosts and provide a lower-level taxonomic classification for this highly diverse and ubiquitous family of RNA viruses.

INTRODUCTION

A combination of metatranscriptomic sequencing and data mining of the Sequence Read Archive (SRA) has greatly enhanced our knowledge of the scale, diversity, and composition of the RNA virosphere. As well as documenting huge numbers of novel viruses, this research program has provided new insights into previously known families of RNA viruses. For example, major changes to taxonomic organisations have been proposed based on updated phylogenetic trees of RNA viruses (Shi *et al.*, 2016; Wolf *et al.*, 2020; Sadiq *et al.*, 2022), and host ranges have been extended when viruses are detected in organisms with which they are not typically associated (Urayama *et al.*, 2016; Geoghegan *et al.*, 2021; Bonny *et al.*, 2021).

Members of the double-stranded RNA virus family *Picobirnaviridae* (order *Durnavirales*, class *Duplornaviricetes*, phylum *Pisuviricota*) have traditionally been considered opportunistic enteric pathogens of mammals and birds, typically detected in animal faecal samples (Malik *et al.*, 2014), and were recently reported in invertebrate hosts (Delmas *et al.*, 2019). In mammalian and avian species, picobirnavirus infection has been associated with disease presenting with diarrhoea or gastroenteritis (Malik *et al.*, 2014; Ganesh *et al.*, 2014). However, as picobirnaviruses have also been detected in the stool of healthy or asymptomatic animals (Ganesh *et al.*, 2014), the presence of a picobirnavirus in the animal gastrointestinal system is not necessarily associated with overt disease. The genomes of these picobirnaviruses are typically bi-segmented, with segment 1 (2.4-2.7kb in length) comprising three open reading frames (ORF), although only the function of ORF3 is known, encoding a capsid protein precursor. Segment 2 (1.7-1.9kb) encodes the viral RNA-dependent RNA polymerase (RdRp) that is universal to RNA viruses and hence a powerful phylogenetic marker (Delmas *et al.*, 2019). Picobirnaviruses with non-segmented genomes have also been reported, in which all ORFs are present in a single, monopartite genome (Giannitti *et al.*, 2015; Shi *et al.*, 2016).

The increasing accessibility of metagenomic sequencing has resulted in the identification of huge numbers of picobirnavirus sequences from both aquatic and terrestrial environmental samples such as wastewater, sewage, permafrost, farmland soils, and forest soils (Adriaenssens *et al.*, 2018; Bell *et al.*, 2020; Guajardo-Leiva *et al.*, 2020; Chen *et al.*, 2022; Neri *et al.*, 2022). As metatranscriptomic sequencing necessarily produces a snapshot of all the RNA expressed in a sample, the host organisms for the majority of recently identified

80 picobirnaviruses cannot be easily determined and characterisation of novel virus species
81 largely relies on phylogenetic placement, often without the presence of complete virus
82 genomes. While the number of picobirnavirus and picobirna-like sequences available on
83 NCBI/GenBank is now in the thousands and comprises considerable genetic diversity, only a
84 single genus and three species have been formally ratified in the most recent report of the
85 International Committee on Taxonomy of Viruses (ICTV) (Delmas *et al.*, 2019). The
86 *Picobirnaviridae* are therefore in clear need of further taxonomic assessment and
87 classification. An additional complicating factor is the large number of picobirnaviruses
88 found within individual animal species has meant that multiple picobirnaviruses have been
89 assigned the same name. For example, there are 12 viral RdRp sequences designated as
90 'bovine picobirnavirus', 13 as 'dromedary picobirnavirus', 16 as 'human picobirnavirus', and
91 over 500 as 'porcine picobirnavirus' on NCBI/GenBank. Despite their identical names, these
92 sequences are distinct and do not group together in phylogenetic trees (Delmas *et al.*, 2019;
93 Ghosh and Malik, 2021).
94
95 The remarkably wide range of both animal and non-animal sources in which picobirnaviruses
96 have been detected, combined with an ongoing inability to propagate these viruses in
97 eukaryotic cell lines or detect them in solid tissue samples, has led to the proposal that the
98 *Picobirnaviridae* are not exclusively animal-associated, or that animals may not be the true
99 hosts (Ghosh and Malik, 2021). The closest relative to the *Picobirnaviridae* is the plant-,
100 fungi-, and protist-infecting family *Partitiviridae* (Vainio *et al.*, 2018), members of which
101 have also been detected in animal faeces (Chen *et al.*, 2021) and invertebrate samples (Shi *et*
102 *al.*, 2016; Le Lay *et al.*, 2020), suggesting they infect components of animal diet rather than
103 the animal themselves. Hence, it is reasonable to propose that the *Picobirnaviridae* are
104 similarly associated with animal diet and/or commensal gut microflora. Aspects of
105 picobirnavirus genomes also support the hypothesis that these viruses do not infect animals
106 (at least not exclusively) and may instead be associated with microbial organisms. In
107 particular, the prokaryotic ribosomal binding site (RBS) motif, necessary for the initiation of
108 translation in prokaryotes, is present in the genomes of other RNA bacteriophage such as in
109 the *Cystoviridae* (Ghosh and Malik, 2021). A high prevalence of RBS motif sequences has
110 been identified upstream of ORFs in both picobirnavirus genome segments (Krishnamurthy
111 and Wang, 2018), suggesting that the picobirnaviruses in fact predominantly represent a
112 family of bacteriophage. In addition, several picobirnavirus species identified from mammals
113 and invertebrates (Yinda *et al.*, 2019; Kleymann *et al.*, 2020) utilise the mitochondrial genetic

code, a characteristic shared by members of the RNA virus family *Mitoviridae* (phylum *Lenarviricota*) that replicate in the fungal mitochondria (Cole *et al.*, 2000; Hillman and Cai, 2013).

Using a set of 2,286 publicly available *Picobirnaviridae* RdRp sequences, the majority of which were generated through metatranscriptomic sequencing, we aimed to provide a realistic classification of this important group of viruses, identifying distinct genera, and to reassess their host range. We based our analysis on the extent of congruence between virus and host phylogenetic trees, as well as on features of *Picobirnaviridae* genomes that may be indicative of bacterial hosts.

METHODS

Data set collection and processing for phylogenetic analysis

For all phylogenies generated in this work, the number of sequences included in the data set, the outgroups used, and the lengths (trimmed and untrimmed) of each alignment generated are detailed in Table 1. The labels T1 through T21, where 'T' simply stands for 'tree', refer to an alignment or phylogeny generated on a particular combination of *Picobirnaviridae* and outgroup family sequences. The exception is T21, which is a host animal cladogram artificially generated in Newick format and used to assess host associations. The data sets assembled for each tree are detailed below.

As a base *Picornaviridae* data set, curated in June 2022, we collated all non-redundant picobirnavirus RdRp sequences available on NCBI/GenBank by searching the protein database for "picobirna". Only sequences with associated hosts or sampling environment metadata were included. Picobirnavirus and picobirna-like sequences described by Neri *et al*. (2022) were also added to this data set. The majority (89%) of sequences collected were faecal or environmentally sourced picobirnaviruses described in Chen *et al*. (2022) and Neri *et al*. (2022). A total of 3,644 sequences clustered at 90-99% amino acid identity using CD-HIT (Fu *et al.*, 2012). One representative sequence from each cluster was retained. The total number of picobirnavirus sequences collated in this manner for further analysis was 2,286.

RdRp sequences from families related to the *Picobirnaviridae* within the order *Durnavirales* were included in additional alignments as outgroups to provide directionality to the resulting phylogenies: the *Amalgaviridae* (14 sequences; T2, T6 in Table 1), *Curvulaviridae* (8

148 sequences; T3, T7), *Fusariviridae* (32 sequences; T4, T8), and the *Partitiviridae* (40, 394, or
149 2,536 sequences; T9, T10, and T5-8, respectively). A phylogeny of the order *Durnavirales*
150 was estimated on 4,717 sequences from the families *Amalgaviridae, Curvulaviridae,*
151 *Fusariviridae, Picobirnaviridae*, and *Partitiviridae* (T11). Phylogenetic trees were also
152 estimated based on sequence alignments of each of the proposed *Picobirnaviridae* genera
153 (T12-18). Finally, sequence alignments were constructed on a subset of 1,269 picobirnavirus
154 sequences detected in animal (typically faecal) samples, with (T19) and without (T20) an
155 outgroup of 40 partitivirus sequences.
156
157 **Sequence alignment and phylogenetic analysis**
158 All sequences were aligned using MAFFT (v7.487) (Katoh and Standley, 2013) and trimmed
159 using trimAL (v1.4.1) (Capella-Gutierrez *et al.*, 2009) to retain the most conserved amino
160 acid positions. Trimmed alignments were manually inspected in Geneious Prime
161 (v.11.0.14.1) to identify and remove any ambiguously aligned regions, resulting in trimmed
162 alignments ranging between 691-757 positions in length used to generate phylogenetic trees.
163
164 The best-fit amino acid substitution model for all data sets was found to be the Le-Gascuel
165 (LG) model, established using ModelFinder within IQ-TREE (v1.6.2). Due to the
166 consistently very high number of amino acid changes per position, a gamma distribution of
167 among-site rate variation was not used to estimate the full *Durnavirales* phylogeny (T11 in
168 Table 1). Maximum likelihood phylogenetic trees were then estimated on each data set using
169 IQ-TREE (v1.6.2) (Nguyen *et al.*, 2015), with node support evaluated using the SH-like
170 approximate likelihood ratio test (SH-aLRT), with 1,000 replicates. Trees were visualised in
171 FigTree (v1.4.4) (Rambaut, 2018), as well as in R (v4.1.0) using the packages 'ape' (v5.5)
172 (Paradis and Schliep, 2019) and ggtree (v3.0.2) (Yu *et al.*, 2017).
173
174 **Table 1. Details of sequence data sets, alignments and rooting schemes used to generate**
175 **phylogenetic trees**

Tree	**Sequences analysed (total number)**	**Alignment length (amino acid residues)**		**Rooting scheme**	**Figure(s)**
		Full length	**Trimmed**		
T1	*Picobirnaviridae* (2,286)	13,540	704	Midpoint	Fig. 1A

T2	*Picobirnaviridae* (2,286), *Amalgaviridae* (14)	13,430	712	*Amalgaviridae*	Fig. 1B
T3	*Picobirnaviridae* (2,286), *Curvulaviridae* (8)	13,744	701	*Curvulaviridae*	Fig. 1C
T4	*Picobirnaviridae* (2,286), *Fusariviridae* (32)	13,738	701	*Fusariviridae*	Fig. 1D
T5	*Picobirnaviridae* (2,286), *Partitiviridae* (2,536)	18,017	757	*Partitiviridae*	Fig. 1E
T6	*Picobirnaviridae* (2,286), *Partitiviridae* (2,536), *Amalgaviridae* (14)	17,740	710	*Amalgaviridae*	Fig. 1F
T7	*Picobirnaviridae* (2,286), *Partitiviridae* (2,536), *Curvulaviridae* (8)	17,742	710	*Curvulaviridae*	Fig. 1G
T8	*Picobirnaviridae* (2,286), *Partitiviridae* (2,536), *Fusariviridae* (32)	17,231	706	*Fusariviridae*	Fig. 1H
T9	*Picobirnaviridae* (2,286), *Partitiviridae* (40)	13,906	695	*Partitiviridae*	Fig. 1I, Fig. 4B, D, Supp. Fig. 9, 11
T10	*Picobirnaviridae* (2,286), *Partitiviridae* (394)	14,063	703	*Partitiviridae*	Fig. 1J
T11	*Durnavirales* excluding family *Hypoviridae* (4,717)	14,971	808	Unrooted	Fig. 2
T12	*Alphapicobirnavirus* (421)	3,639	728	Midpoint	Fig. 3A, Supp. Fig. 1
T13	*Betapicobirnavirus* (802)	2,844	711	Midpoint	Fig. 3B, Supp. Fig. 2
T14	*Gammapicobirnairus* (418)	3,856	733	Midpoint	Fig. 3E, Supp. Fig. 5
T15	*Deltapicobirnavirus* (311)	5,759	749	Midpoint	Fig. 3F, Supp. Fig. 6
T16	*Epsilonpicobirnavirus* (36)	853	691	Midpoint	Fig. 3C, Supp. Fig. 3
T17	*Zetapicobirnavirus* (67)	824	700	Midpoint	Fig. 3D, Supp. Fig. 4

T18	*Etapicobirnavirus* (70)	881	705	Midpoint	Fig. 3G, Supp. Fig. 7
T19	Animal-associated *Picobirnaviridae* (1,269), *Partitiviridae* (40)	5,734	745	*Partitiviridae*	Fig. 4A, C, Supp. Figs. 8, 10
T20	Animal-associated *Picobirnaviridae* (1,269)	5,591	727	Midpoint	Input for Fig. 5 (not shown)
T21	NA – artificially generated host cladogram	NA	NA	NA	Input for Fig. 5, Supp. Fig. 12

Classification of the *Picobirnaviridae*

Ten phylogenies (T1-T10; see Table 1) were used for determining potential genera within the *Picobirnaviridae*. Four large and three small defined clusters were identified in the midpoint-rooted tree (T1 in Table 1), defined by relatively long branch lengths to their respective common ancestor node and ≥80% SH-aLRT support. Sequences within each cluster were annotated as ‘clade 1’ through 'clade 7'. These clusters, or “draft genera" were then annotated in all other trees. Any sequences not clustering within their assigned “draft genus” with ≥80% SH-aLRT node support in any of the ten trees were excluded. The remaining sequences that consistently grouped within the same clade in every tree were considered the “core” sequences of each proposed genus. Sequence alignments were constructed on each set of core sequences and subsequent genus-level phylogenetic trees were estimated as described above.

Analysis of phylogenetic incongruence

Prior to analysis, *Picobirnaviridae* sequences were grouped by assigned host (animals) or sampling source (environments and microbial communities). Hosts and sampling sources were categorised in two ways (Table 2). The first focused on specific animal hosts, with non-animal sources simply reflected in a joint ‘microbial/environmental’ category, while mammals were further categorised into lower taxonomic levels. The second grouped animal hosts more broadly as ‘mammalian’, ‘avian’, ‘fish’, ‘reptile’, and ‘invertebrate’, with non-animal sources similarly expanded into ‘aquatic’, ‘terrestrial’, and ‘engineered’ environments and microbial communities. Aquatic sources included bodies of water and sediments, while terrestrial environments included various soils and plant matter (Chen *et al*., 2022; Neri *et al*.,

199 2022). Engineered sources included wastewater, food fermentation processes, and laboratory
200 cultures (Neri *et al*., 2022).
201
202 A cladogram of relevant 'host' animals (as listed in Table 2) was constructed based on the
203 evolutionary relationships demonstrated in the current literature on vertebrate and
204 invertebrate evolution (Lindblad-Toh *et al.*, 2005; Song *et al.*, 2012; Fong *et al.*, 2012; Shen
205 *et al.*, 2014; Prum *et al.*, 2015; Zhou *et al.*, 2017; Zurano *et al.*, 2019). The eMPRess program
206 (v1.2.1) (Santichaivekin *et al.*, 2021) was used to reconcile topologies of the animal-
207 associated picobirnavirus (T20 in Table 1) and host animal (T21 in Table 1) phylogenies to
208 determine the event likelihoods of co-divergence, cross-species transmission, duplication, and
209 extinction. Event costs were defined as 0 for co-divergence and 1 for cross-species
210 transmission, duplication, and extinction. The R package 'NELSI' (v0.21) (Ho *et al.*, 2015)
211 was used to calculate the normalised PH85 (nPH85) distance for the *Picobirnaviridae*. The
212 nPH85 distance is based on Penny and Hendy distance metric and describes the topological
213 distance between the virus and host phylogenies (Geoghegan *et al.*, 2017). An nPH85
214 distance of 0 indicates complete co-divergence between virus and host, whereas an nPH85 of
215 1 indicates complete cross-species transmission.
216
217 **Table 2. Categories of assumed host animal or environmental/microbial source of the**
218 ***Picobirnaviridae* sequences analysed in this study**

Animal-host focused list	**Environmental/microbial source-focused list**
Primate	Mammalian
Bovidae	
Cervidae	
Equidae	
Suidae (porcine)	
Dromedary	
Feline	
Canine	
Tasmanian devil	
Rodent	
Lagomorph	
Bat	

Avian	Avian
Reptile	Reptile
Fish	Fish
Arthropod	Invertebrate
Other invertebrate	
Microbial/environmental	Microbial - aquatic
	Microbial - engineered
	Microbial - terrestrial
	Environmental - aquatic
	Environmental - engineered
	Environmental - terrestrial

219

220 **Presence of bacterial ribosomal binding sites**

221 A sequence alignment was generated using MAFFT (v7.487) comprising all available non-
222 redundant and annotated nucleotide sequences of segment 1 of the picobirnavirus genome
223 that contained at least 30 nucleotides of a 5' untranslated region (UTR), totalling 360
224 sequences. Sequences were aligned to assist in locating the start codons of ORFs, upstream of
225 which the ribosomal binding site motifs are potentially located. Similarly, an alignment was
226 constructed comprising all available segment 2 *Picobirnaviridae* sequences using the same
227 criteria, for a total of 984 sequences. Occurrences of the 'AGGAGG' hexamer within 14-24
228 nucleotides upstream of each ORF's start codon were identified and annotated using
229 Geneious Prime (v11.0.14.1), as these likely represented a bacterial RBS motif and hence are
230 evidence of a bacterial host. The frequency of RBS occurrence for each segment was
231 calculated as a proportion of the total number of annotated ORFs in the set of sequences for
232 that segment. To compare the frequencies of RBS presence between the *Picobirnaviridae* and
233 established RNA bacteriophage families this process was repeated on 1,040 *Leviviricetes*
234 genomes and 17, 11, and 38 *Cystoviridae* S, M, and L segments, respectively.

235

236 **RESULTS**

237 **Classification of the *Picobirnaviridae***

238 We observed seven major well-supported clusters of RdRp sequences within the
239 *Picobirnaviridae* that were consistently present across ten phylogenies inferred using
240 different outgroups and rooting methods (Figure 1, Table 1). Each of the seven clades were

241 defined by relatively long branches to their respective common ancestral node with >80%
242 SH-aLRT node support in each of the ten phylogenies (Figure 1). We propose that these
243 groups are candidates for novel genera with the *Picobirnaviridae*, tentatively named:
244 *Alphapicobirnavirus* (421 sequences, containing *Orthopicobirnavirus equi*),
245 *Betapicobirnavirus* (802 sequences, containing *Orthopicobirnavirus hominis*),
246 *Gammapicobirnavirus* (418 sequences, containing *Orthopicobirnavirus beihaiense*),
247 *Deltapicobirnavirus* (311 sequences), *Epsilonpicobirnavirus* (36 sequences),
248 *Zetapicobirnavirus* (67 sequences), and *Etapicobirnavirus* (70 sequences), in line with the
249 nomenclature for other multi-genus families within the *Durnavirales*. Sequences consistently
250 clustering within one of these clades in each of the ten trees were considered the "core"
251 members comprising each genus. Those sequences whose phylogenetic position was
252 inconsistent remained as “unclassified” *Picobirnaviridae*. Using this approach, we were able
253 to assign 2,125 non-redundant picobirnaviruses to a proposed genus (Supplementary Table
254 1). Genus-level phylogenies of each proposed genus are shown in Supplementary Figures 1-
255 7.

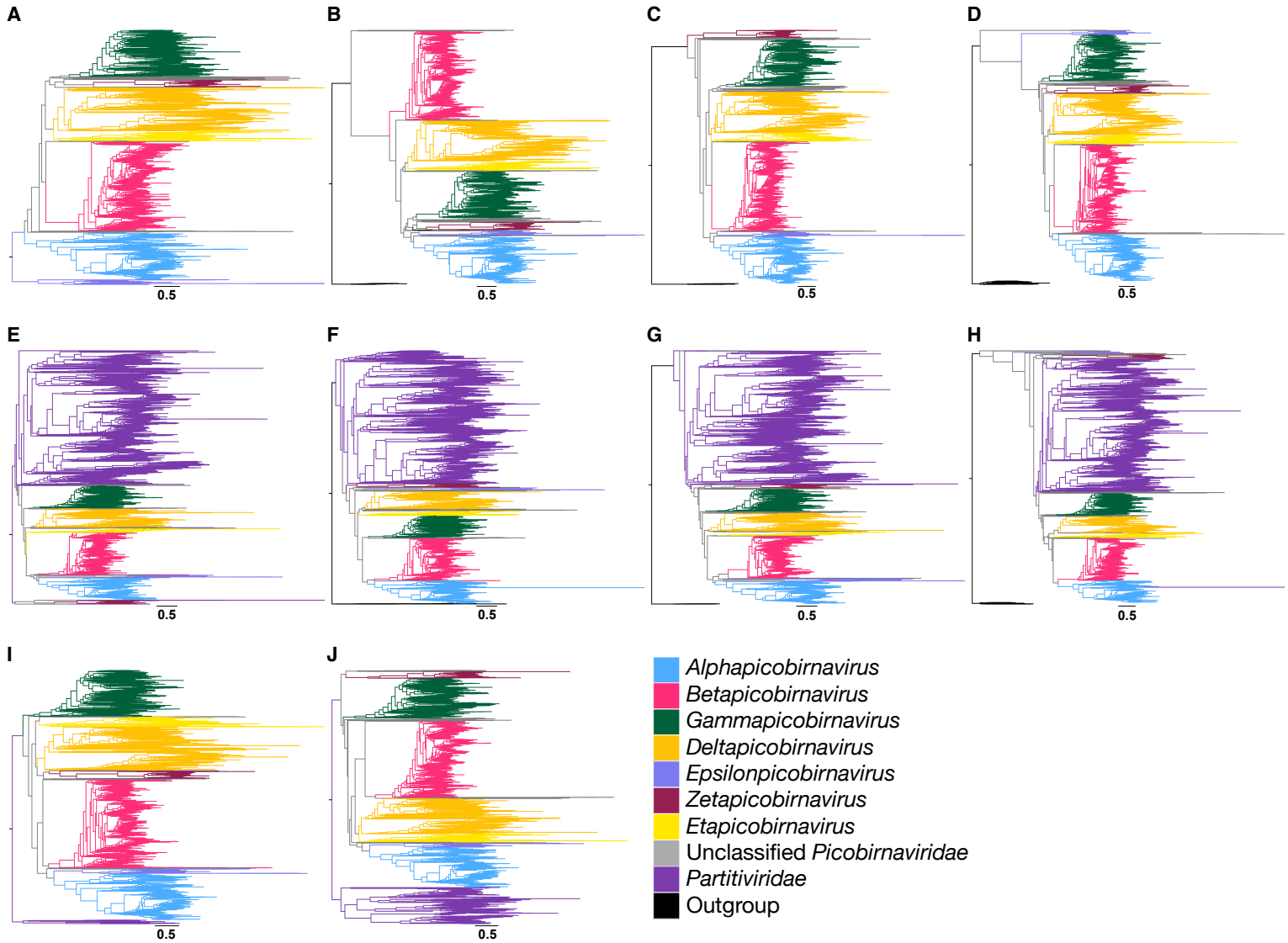
A
B
C
D
E
F
G
H
I
J
0.5
Alphapicobirnavirus
Betapicobirnavirus
Gammapicobirnavirus
Deltapicobirnavirus
Epsilonpicobirnavirus
Zetapicobirnavirus
Etapicobirnavirus
Unclassified Picobirnaviridae
Partitiviridae
Outgroup

Figure 1

258 **Figure 1.** Maximum likelihood phylogenies of the family *Picobirnaviridae* estimated on the
259 RdRp and using different families of the order *Durnavirales* as outgroups. Row one contains
260 the *Picobirnaviridae*: (A) midpoint-rooted or utilising the (B) *Amalgaviridae*, (C)
261 *Curvulaviridae*, or (D) *Fusariviridae* as outgroups. Row two contains the multi-family
262 *Partitivirdae-Picobirnaviridae* clade: (E) midpoint-rooted or utilising the (F) *Amalgaviridae*,
263 (G) *Curvulaviridae*, or (H) *Fusariviridae* as outgroups. Row three contains the
264 *Picobirnaviridae* with a subset of (I) 40 or (J) 394 *Partitiviridae* sequences as outgroups. In
265 all trees, the *Partitiviridae* are marked by purple branches and the outgroup is in black. The
266 proposed *Picobirnaviridae* genera – *Alpha*-, *Beta*-, *Gamma*-, *Delta*-, *Epsilon*-, *Zeta*-, and
267 *Etapicobirnavirus* – are coloured by blue, pink, green, orange, light purple, maroon, and
268 yellow branches, respectively, while sequences unable to be assigned to a genus are
269 represented by grey branches. Horizontal branches are drawn to scale, and the scale bars
270 below each tree represent 0.5 amino acid substitutions per site.
271
272
273 The only currently accepted genus in the *Picobirnaviridae* – *Orthopicobirnavirus* – contains
274 three species: *Orthopicobirnavirus hominis, Orthopicobirnavirus equi,* and
275 *Orthopicobirnavirus beihaiense*, isolated from human (Wakuda *et al.*, 2005), horse (Giannitti
276 *et al.*, 2015), and peanut worms (Shi *et al.*, 2016), respectively. Each of these species appear
277 in a different proposed genus described here (Supplementary Figures 1-3), limiting our ability
278 to clearly determine which most closely resembles the genus *Orthopicobirnavirus*.
279
280 There were no common topologies for the proposed *Picobirnaviridae* genera across all ten
281 phylogenetic trees. As a consequence, exact evolutionary relationships among genera cannot
282 be safely determined. However, the genera *Deltapicobirnavirus* and *Etapicobirnavirus*
283 formed sister clades in seven of the ten phylogenies (Figure 1A-D, F, H, J). In the remaining
284 three trees (Figure 1E, G, I), one genus fell directly basal to the other, but which genus was
285 basal to the other was inconsistent among trees. In six phylogenies, *Alphapicobirnavirus* and
286 *Epsilonpicobirnavirus* formed sister clades (Figure 1B, C, E, G, I, J), suggesting that these
287 genera are more closely related to each other. In eight trees, the two largest genera,
288 *Alphapicobirnavirus* and *Betapicobirnavirus,* either grouped together (Figure 1D-I) or one
289 fell directly basal to the other (Figure 1A, C). Furthermore, in four of the trees where
290 *Alphapicobirnavirus* and *Betapicobirnavirus* genera clustered together,

291 *Epsilonpicobirnavirus* was also part of the group as a sister clade to *Alphapicobirnavirus*
292 (Figure 1C, E, G, I).
293
294 We next sought to give these proposed genera evolutionary context by placing them in a tree
295 with multiple related families of RNA viruses. An unrooted phylogeny was estimated on
296 4,717 RdRp sequences from the order *Durnavirales*, including the *Picobirnaviridae*,
297 *Partitiviridae*, *Amalgaviridae*, *Curvulaviridae*, and *Fusariviridae* (Figure 2). Sequences from
298 the family *Hypoviridae* were too divergent to reliably align and were therefore excluded.
299 Importantly, the *Picobirnaviridae* formed a monophyletic group that was phylogenetically
300 distinct from the other families in the order. Certain trends in tree topology frequently
301 observed in Figure 1 were repeated here, such as *Deltapicobirnavirus* and *Etapicobirnavirus*
302 forming sister clades, as well as *Alphapicobirnavirus* and *Epsilonpicobirnavirus* forming
303 sister clades that clustered with *Betapicobirnavirus* (Figure 2). The proposed genera also
304 remained mostly intact despite the high level of diversity in the alignment. The only
305 exception was a group of seven divergent deltapicobirnaviruses that fell as a small sister
306 clade to the zetapicobirnaviruses (Figure 2, denoted by a star).

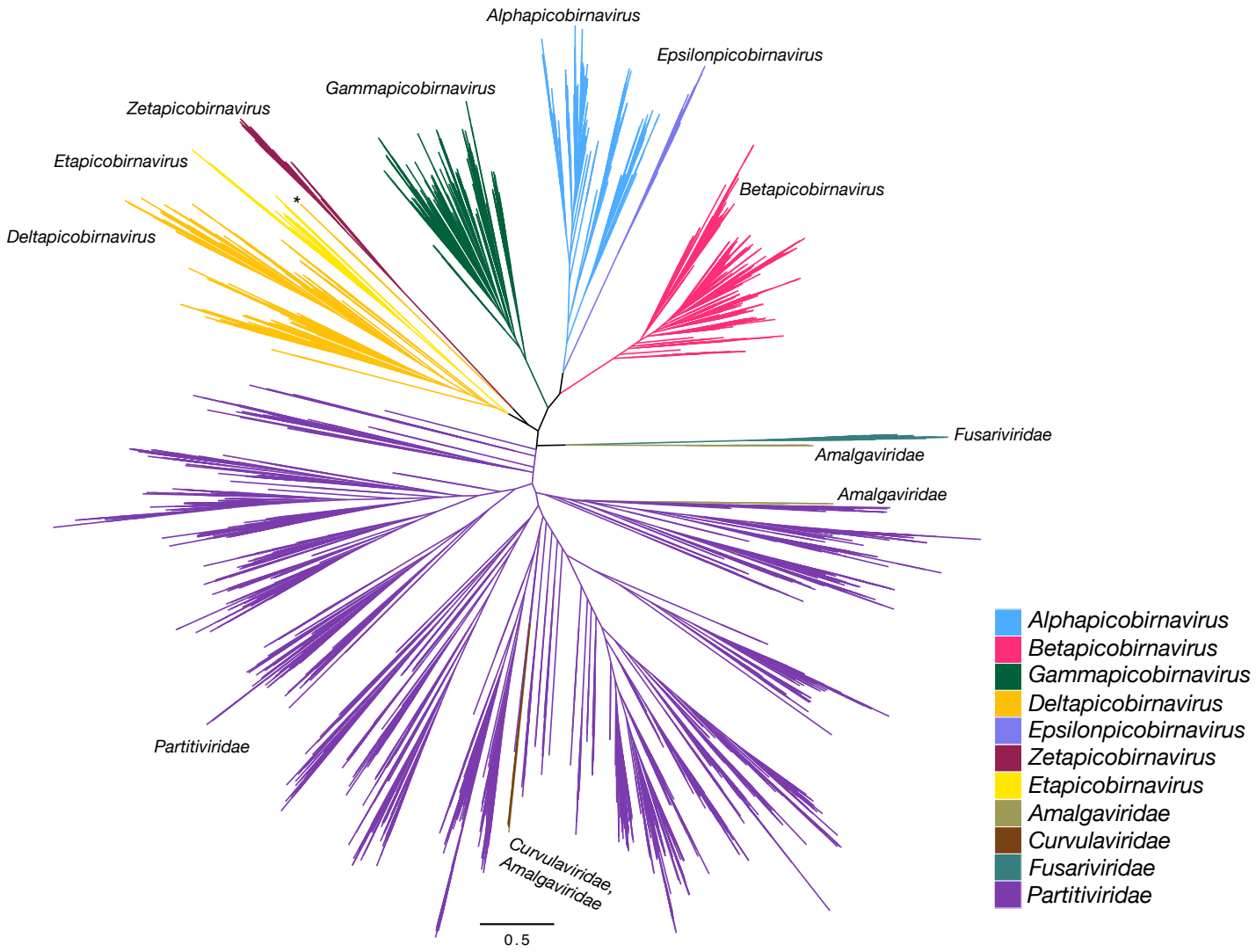
Alphapicobirnavirus
Epsilonpicobirnavirus
Gammapicobirnavirus
Zetapicobirnavirus
Etapicobirnavirus
Betapicobirnavirus
Deltapicobirnavirus
*
Fusariviridae
Amalgaviridae
Amalgaviridae
Partitiviridae
Curvulaviridae,
Amalgaviridae
0.5
Alphapicobirnavirus
Betapicobirnavirus
Gammapicobirnavirus
Deltapicobirnavirus
Epsilonpicobirnavirus
Zetapicobirnavirus
Etapicobirnavirus
Amalgaviridae
Curvulaviridae
Fusariviridae
Partitiviridae

Figure 2

309 **Figure 2.** Unrooted maximum likelihood tree of 4,717 *Durnavirales* RdRp sequences, with
310 branches coloured by family (*Amalgaviridae*, *Curvulaviridae*, *Fusariviridae*, *Partitiviridae*)
311 or proposed genus (*Picobirnaviridae*). Families *Amalgaviridae*, *Curvulaviridae*,
312 *Fusariviridae*, and *Partitiviridae* are marked by olive-green, brown, teal, and purple
313 branches, respectively. The proposed *Picobirnaviridae* genera – *Alpha*-, *Beta*-, *Gamma*-,
314 *Delta*-, *Epsilon*-, *Zeta*-, and *Etapicobirnavirus* – are coloured by blue, pink, green, orange,
315 light purple, maroon, and yellow branches, respectively, while sequences unable to be
316 assigned to a *Picobirnaviridae* genus are represented by grey branches. Branch lengths
317 indicate the number of amino acid substitutions per site, as represented by the scale bar.
318
319
320 **Distribution of apparent hosts or environmental source of picobirnaviruses in each**
321 **proposed genus**
322 The genus *Alphapicobirnavirus* (Figure 3A, Supplementary Figure 1) predominantly
323 comprised mammalian and avian picobirnaviruses, as well as those sampled from some
324 invertebrate species. In the most basal clade of this genus the majority of sequences were
325 identified in microbial communities, although a small number of mammalian, avian,
326 invertebrate, and environmentally-sourced picobirnaviruses also fell in this group. The genus
327 *Betapicobirnavirus* (Figure 3B, Supplementary Figure 2) was similarly dominated by
328 picobirnaviruses from mammalian and avian sources. This genus also contained the divergent
329 species identified in fish (AVM87403 Beihai goldsaddle goldfish picobirnavirus) and lizards
330 (AVM87436 Guangdong Chinese water skink picobirnavirus and UCS96434
331 Picobirnaviridae sp.). There were fewer microbial/environmental-sourced viruses in the
332 betapicobirnaviruses and none formed large, monophyletic clusters as seen in the basal
333 *Alphapicobirnavirus* group.
334
335 In contrast to the predominantly animal-associated genera described above, the
336 *Gammapicobirnavirus* (Figure 3E, Supplementary Figure 3) and *Deltapicobirnavirus* (Figure
337 3F, Supplementary Figure 4) genera were almost entirely composed of sequences detected in
338 microbial communities and environmental samples. Some invertebrate-associated
339 picobirnaviruses (namely, from Cnidara, Porifera, crustaceans, and molluscs) clustered within
340 the *Gammapicobirnaviridae* (Supplementary Figure 3), as well as two picobirnaviruses from
341 vertebrate faecal samples (USE08169 Picobirnavirus sp. from pig faeces and QUS52969

342 mute swan feces associated picobirnavirus 3). However, most viruses from this genus were
343 sampled in environmental or microbial samples, particularly aquatic sources.
344
345 Viruses in the genus *Deltapicobirnavirus* were predominantly sourced from terrestrial
346 microbe communities and environments (Figure 3F, Supplementary Figure 4). The most
347 recently diverged clade had a more even distribution of viruses sampled from aquatic and
348 terrestrial environments and microbes (Supplementary Figure 4). Again, two picobirnaviruses
349 sourced from animals were present in this genus - ND_241614 and ND_192065 - from cattle
350 and sheep, respectively. However, these sequences were mined from animal rumen
351 microbiome sequence data (Neri *et al.*, 2022), thereby making *Deltapicobirnavirus* the only
352 genus containing entirely microbe-associated or environmentally sourced viruses.
353
354 The remaining genera were considerably smaller. All but one of the 36 viruses falling into the
355 genus *Epsilonpicobirnavirus* were sourced from vertebrate faeces, the majority of which
356 were mammalian (cattle and porcine) (Figure 3C, Supplementary Figure 5). The genus
357 *Zetapicobirnavirus* also featured a similar range of sampling sources, with the majority being
358 cattle and porcine samples, along with one avian- and one primate-sourced picobirnavirus.
359 This phylogeny included a notable divergent clade of five arthropod-associated
360 zetapicobirnaviruses (Figure 3D, Supplementary Figure 6). Finally, despite only comprising
361 70 sequences, the genus *Etapicobirnavirus* had a broader "host" range across the two sister
362 clades observed within this group. One predominantly comprised vertebrate
363 etapicobirnaviruses, as well as one sequence identified in a termite that is likely associated
364 with its bacterial symbionts (Figure 3G, Supplementary Figure 7). The other clade included
365 another 15 etapicobirnaviruses identified in termite samples, with the basal lineages largely
366 viruses sourced from microbial and aquatic environmental samples (Supplementary Figure
367 7).

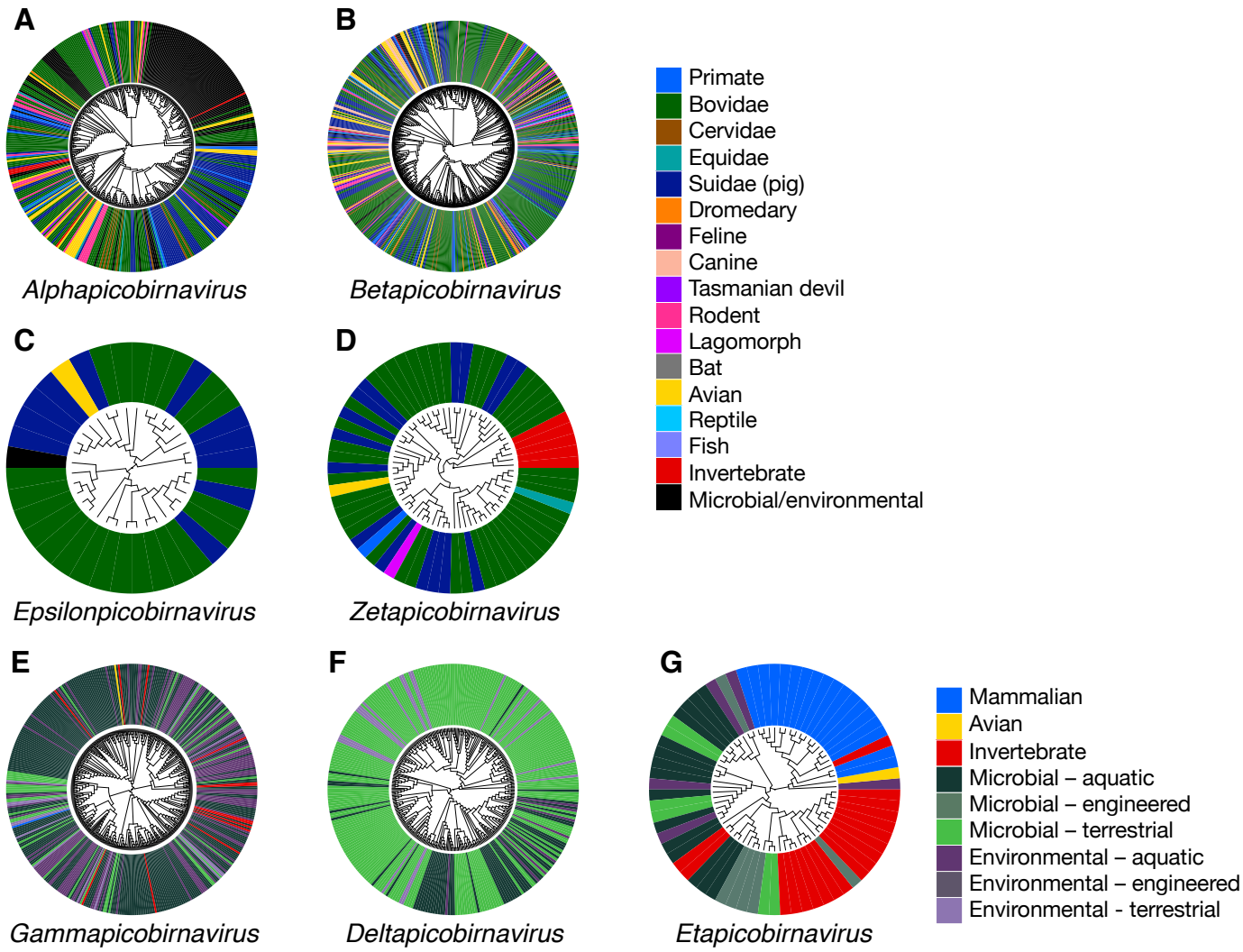
A
Alphapicobirnavirus
B
Betapicobirnavirus
C
Epsilonpicobirnavirus
D
Zetapicobirnavirus
E
Gammapicobirnavirus
F
Deltapicobirnavirus
G
Etapicobirnavirus
Primate
Bovidae
Cervidae
Equidae
Suidae (pig)
Dromedary
Feline
Canine
Tasmanian devil
Rodent
Lagomorph
Bat
Avian
Reptile
Fish
Invertebrate
Microbial/environmental
Mammalian
Avian
Invertebrate
Microbial – aquatic
Microbial – engineered
Microbial – terrestrial
Environmental – aquatic
Environmental – engineered
Environmental - terrestrial

Figure 3

371 **Figure 3.** Maximum likelihood phylogenetic trees of the RdRp of proposed
372 *Picorbirnaviridae* genera coloured by apparent host or sampling environment. All
373 phylogenies are midpoint-rooted. Tip labels are coloured to represent assigned host/sampling
374 environment. *Alphapicobirnavirus* (A), *Betapicobirnavirus* (B), *Epsilonpicobirnavirus* (C)
375 and *Zetapicobirnavirus* (D) are coloured by specific assigned animal host.
376 *Gammapicobirnavirus* (E), *Deltapicobirnavirus* (F), and *Etapicobirnavirus* (G) categorise
377 animal hosts more broadly as 'mammalian', 'avian', and 'invertebrate', and non-animal
378 associated sequences are grouped into 'environmental' or 'microbial' samples, both further
379 classed into 'terrestrial', 'aquatic', and 'engineered' sampling sources. Branch lengths
380 indicate the number of amino acid substitutions per site. Scale bars for branch lengths are
381 shown in rectangular versions of trees (Supplementary Figures 1-7).
382
383
384 **Accuracy of host assignments based on phylogenetic position**
385 To visualise the extent of apparent cross-species transmission within the *Picobirnaviridae*
386 based on assigned hosts/sample source, we estimated a phylogeny of only the animal-sourced
387 picobirnaviruses (Figure 4A, T19 in Table 1) and a complete picobirnavirus phylogeny
388 (Figure 4B, T9 in Table 1). Both trees were rooted by a set of 40 partitiviruses as they
389 represent the most closely related family to the *Picobirnaviridae*. The diverse distribution of
390 assigned host species across the phylogeny, as demonstrated in Figures 4A and 4B where tips
391 are coloured by host, revealed a lack of consistent clustering by assigned host. Even at the
392 order level within the *Mammalia*, closely related virus sequences had assigned hosts spanning
393 diverse groups of mammals. Picobirnaviruses from birds did not form any large
394 monophyletic groups, and instead clustered with species from cattle, pigs, felines, canines,
395 primates, and marsupials (Figure 4A, C). This was also clearly observable within the
396 predominantly animal-associated genera *Alphapicobirnavirus*, *Betapicobirnavirus*,
397 *Epsilonpicobirnavirus*, and *Zetapicobirnavirus* (Figure 3A-D, Supplementary Figures 1-4).
398 Although microbial/environmentally sourced viruses tended to group together in clades
399 distinct from the animal-sourced picobirnaviruses, there were several invertebrate-, primate-,
400 cattle-, pig-, and avian-associated species present in the environmental/microbial clades
401 (Figure 4B, D). Notably, representatives from all sub-categories of microbial and
402 environmental picobirnaviruses were present in animal-associated clades with the exception
403 of those sourced from terrestrial microbes (Figure 4D). Rectangular, midpoint-rooted

404 versions of the trees shown in Figure 4 with tip labels including accession numbers and virus
405 names are shown in Supplementary Figures 8-11.

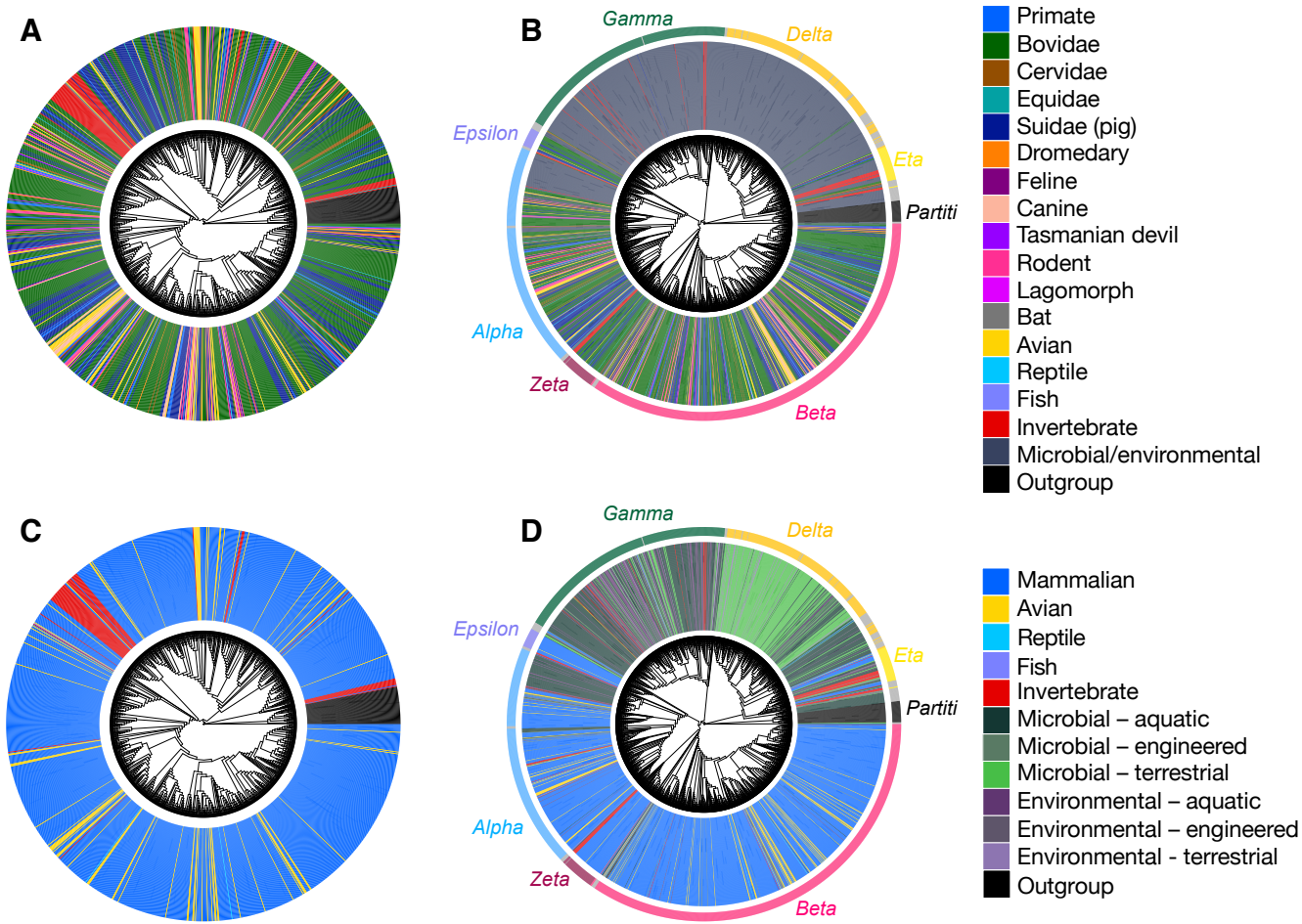
A
B
Gamma
Delta
Eta
Partiti
Beta
Zeta
Alpha
Epsilon
Primate
Bovidae
Cervidae
Equidae
Suidae (pig)
Dromedary
Feline
Canine
Tasmanian devil
Rodent
Lagomorph
Bat
Avian
Reptile
Fish
Invertebrate
Microbial/environmental
Outgroup
C
D
Gamma
Delta
Eta
Partiti
Beta
Zeta
Alpha
Epsilon
Mammalian
Avian
Reptile
Fish
Invertebrate
Microbial – aquatic
Microbial – engineered
Microbial – terrestrial
Environmental – aquatic
Environmental – engineered
Environmental - terrestrial
Outgroup

Figure 4

410 **Figure 4.** Maximum likelihood trees of *Picobirnaviridae* RdRp with tips coloured to
411 represent assigned host/sampling environment. *Partitiviridae* RdRp sequences (n=40) were
412 used to root all trees and are shown in black. The top two trees are coloured by animal host
413 groups, including lower level Mammalia groupings, and are estimated on (A)
414 *Picobirnaviridae* sequences with assumed animal associations and (B) all *Picobirnaviridae*
415 sequences. The bottom two trees categorise animal hosts more broadly as 'mammalian',
416 'avian', 'reptile', 'fish', and 'invertebrate' and are estimated using (C) *Picobirnaviridae*
417 sequences with assumed animal associations and (D) all *Picobirnaviridae* sequences with
418 non-animal associated sequences grouped into 'environmental' or 'microbial' samples, both
419 further classed into 'terrestrial', 'aquatic', and 'engineered' sampling sources. Scale bars for
420 branch lengths are shown in rectangular versions of trees (Supplementary Figures 8-11).
421
422
423 A phylogeny of animal-associated picobirnaviruses (T20 in Table 1) was compared to a host
424 phylogeny (Supplementary Figure 12, T21 in Table 1) using NELSI (Ho *et al.*, 2015) to
425 estimate the frequency of cross-species transmission (assuming assigned hosts are the true
426 hosts) by calculating the nPH85 metric that describes the topological distance between the
427 two phylogenies. The nPH85 distance between the unrooted animal-associated
428 *Picobirnaviridae* tree and host taxa tree was 0.998 (Figure 5A), suggesting almost complete
429 cross-species transmission and very little occurrence of co-divergence among the animal-
430 associated members of this family. Similarly, the nPH85 distance for the genus
431 *Alphapicobirnavirus* was 0.994, and 1 for both the *Betapicobirnavirus*, and
432 *Epsilonpicobirnavirus.* The genus *Zetapicobirnavirus* had a comparatively higher, yet still
433 low, level of co-divergence with an nPH85 distance of 0.921 (Figure 5A). eMPRess
434 (Santichaivekin *et al.*, 2021) was utilised to determine the respective likelihoods of co-
435 divergence, cross-species transmission, duplication, or extinction in the reconciled virus-host
436 co-phylogeny. The event likelihoods are shown in Figure 5B-F and reveal that cross-species
437 transmission was far more frequent than any other event (68% of events, Figure 5B) for the
438 entire family as well as for each individual genus (45-68% of events, Figure 5C-F). In
439 contrast, only low levels of apparent virus-host co-divergence were observed, especially at
440 the level of the entire *Picobirnaviriade* (1-23% of events, Figure 5B-F).
441

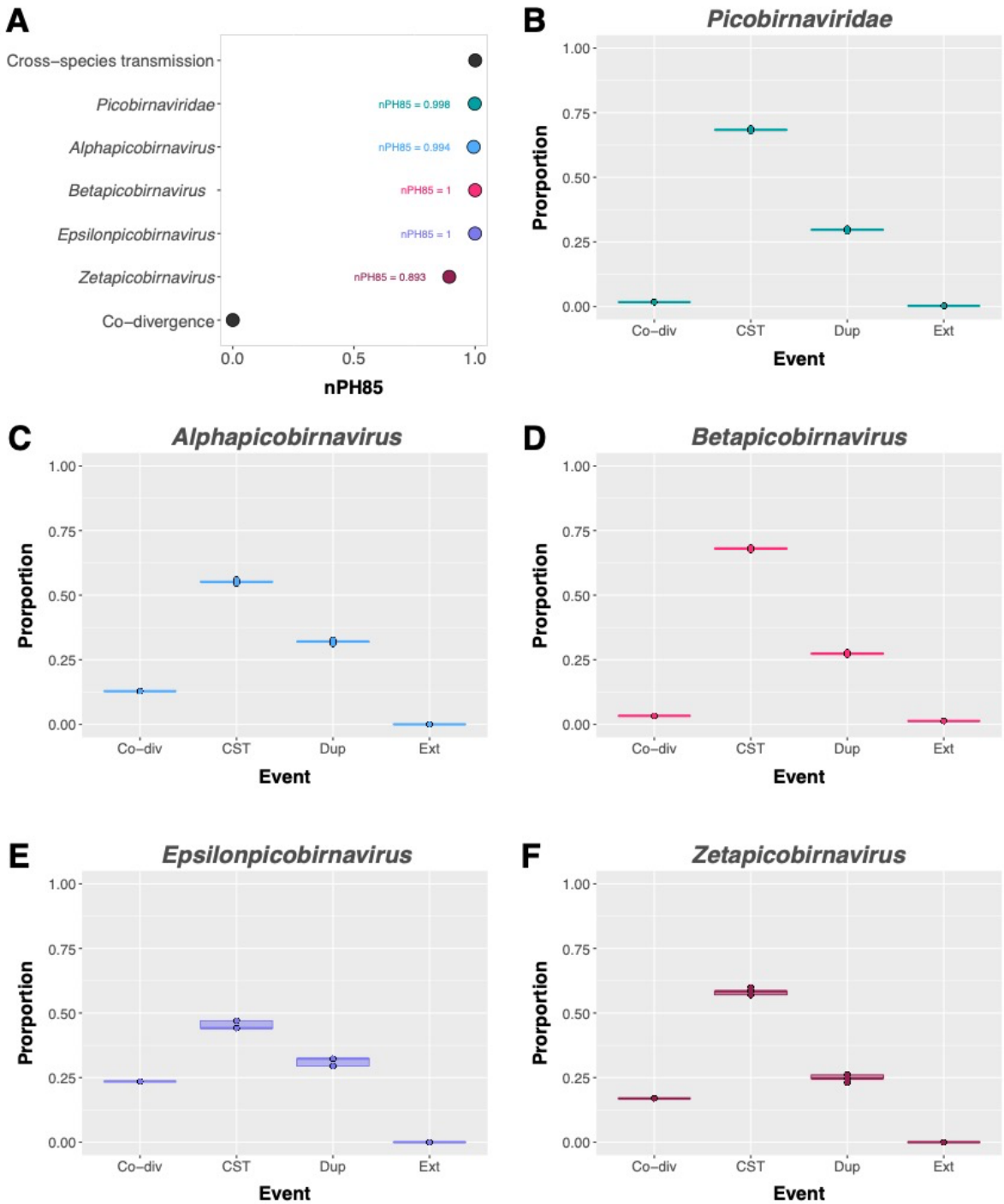

Figure 5

Figure 5. Levels of co-divergence of the *Picobirnaviridae* with associated animal hosts. (A) nPH85 metric calculated for the *Picobirnaviridae*, *Alphapicobirnavirus*, *Betapicobirnavirus, Epsilonpicobirnavirus,* and *Zetapicobirnavirus* using NELSI, representing the topological distance between unrooted virus and host trees. nPH85 = 0 indicates complete co-divergence and nPH85 = 1 indicates complete cross-species transmission. Event likelihoods of co-

450 divergence (Co-div), cross-species transmission (CST), duplication (Dup), and extinction
451 (Ext) as a proportion of all possible events in the reconciled co-phylogenies of animal-
452 associated picobirnavirus sequences and their assigned hosts for (B) the family
453 *Picobirnaviridae*, and the genera (C) *Alphapicobirnavirus*, (D) *Betapicobirnavirus*, (E)
454 *Epsilonpicobirnavirus*, and (F) *Zetapicobirnavirus*. Event likelihoods were calculated in
455 eMPRess with 100 replicates.
456
457
458 **Genetic evidence of *Picobirnaviridae* in non-animal hosts**
459 We next searched for bacterial RBS motifs in a data set of 1,344 picobirnavirus, 1,040
460 levivirus, and 66 cystovirus genome segments. The detection of an 'AGGAGG' hexamer 14-
461 24 nucleotides upstream of the start codon of an annotated ORF was considered to indicate
462 the presence of an RBS motif. There were 722 annotated ORFs across the 360 picobirnavirus
463 segment 1 nucleotide sequences and 993 annotated ORFs across the 984 picobirnavirus
464 segment 2 sequences. The number of segment 1 and segment 2 ORFs preceded by an RBS
465 motif was 431 (60%) and 665 (67%), respectively (Table 3). Strikingly, this was a higher
466 proportion of RBS motif enrichment than observed in the confirmed RNA bacteriophage
467 groups the *Leviviricetes* (210 of 3306 ORFs; 6.4%) and *Cystoviridae* (7 of 305 ORFs; 2.3%)
468 (Table 3).
469
470 **Table 3. Genome segments and calculated proportions of annotated ORFs with the**
471 **'AGGAGG' RBS motif 14-24 nucleotides upstream of the start codon.**

Genome segment	**Number of nucleotide sequences**	**Number of annotated ORFs**	**Number of RBS motifs[1]**
Picobirnavirus segment 1	360	722	431 (59.7%)
Picobirnavirus segment 2	984	993	665 (66.9%)
Levivirus genome	1040	3306	210 (6.3%)
Cystovirus S segment	17	85	0 (0%)
Cystovirus M segment	11	52	2 (3.8%)
Cystovirus L segment	38	167	5 (3%)

472 [1]Values in parentheses indicate the percentage of annotated ORFs precede by an RBS motif.
473
474 **DISCUSSION**
475 The *Picobirnaviridae* are a family of RNA viruses traditionally associated with opportunistic
476 gastroenteritis in humans and other animals that have dramatically increased in size and
477 diversity in recent years due to metatranscriptomic sequencing. Despite this marked increase
478 in diversity, picobirnaviruses remain largely unclassified beyond the family level, with only
479 three divergent species comprising a single genus having been formally accepted by the
480 ICTV (Delmas *et al.*, 2019). The central aim of this study was to provide a taxonomic
481 structure for this family by identifying distinct clades within the *Picobirnaviridae* that may
482 represent genera (while noting that all such definitions are arbitrary). To this end, multiple
483 outgroups and rooting schemes were utilised in an expansive phylogenetic analysis. From
484 this, we identified seven clusters of sequences that consistently appeared across ten different
485 phylogenies, which we propose be considered as distinct genera: *Alphapicobirnavirus*,
486 *Betapicobirnavirus*, *Gammapicobirnavirus*, *Deltapicobirnavirus*, *Epsilonpicobirnavirus*,
487 *Zetapicobirnavirus*, and *Etapicobirnavirus*. In addition, these seven proposed genera
488 remained mostly intact and together formed a monophyletic, phylogenetically distinct group
489 within the order *Durnavirales*. No genera shared greater than an average of 19% amino acid
490 identity in the RdRp with any other genera, which is in line with the <24% amino acid
491 identity shared between genera of the related *Partitiviridae* (Vainio *et al.*, 2018).
492
493 Correctly assigning host organisms to novel virus species is a major limitation of the meta-
494 transcriptomic approach to virus discovery. While the *Picobirnaviridae* have historically
495 been associated with humans and other animals, there is mounting evidence that animals may
496 not be the true hosts (Ghosh and Malik, 2021). As a consequence, we asked whether the
497 *Picobirnaviridae* were a true animal-infecting family of RNA viruses, or one that infect
498 microbiota themselves associated with animal hosts: i.e., that they are bacteriophage or other
499 microbe-infecting viruses. The marked lack of topological congruence between phylogenies
500 of the animal-associated picobirnaviruses and their apparent animal hosts indicates that the
501 evolutionary history of the *Picobirnaviridae* has been characterised by extensive host
502 jumping, with statistical tests revealing near complete cross-species transmission. Notably,
503 the nPH85 distance for animal associated picobirnaviruses (0.998) is higher than the average
504 for RNA virus families (0.95), and indeed higher than *Rhabdoviridae* (0.989), the family with
505 the highest nPH85 distance of those analysed in Geoghegan *et al*. (2017). This supports

506 previous suggestions that animal-associated picobirnaviruses do not cluster according to their
507 "host" of sampling (Duraisamy *et al.*, 2018; Woo *et al.*, 2019; Mahar *et al.*, 2020), although
508 the current study uses a much larger data set. The complete lack of apparent host and viral
509 phylogenetic congruence points toward an error in host assignment, rather than the less
510 biologically feasible alternative of near complete cross-species transmission at a rate not
511 seem in any other RNA virus family.
512
513 Picobirnaviruses have been detected in a large range of animal species, including those from
514 terrestrial and aquatic habitats, invertebrates and vertebrates, and those with herbivorous,
515 carnivorous, and omnivorous diets. This raises the question of how picobirnaviruses are able
516 to overcome the geographical and biological barriers between animals with remarkably
517 different habitats, lifestyles, and gastrointestinal systems, if they truly infect the hosts being
518 sampled. Broad host ranges spanning vertebrates and invertebrates have been observed in
519 several RNA virus orders, such as the *Articulavirales*, *Nidovirales*, *Reovirales*, and
520 *Mononegevirales*, with the latter two also carrying fungi-infecting species. At a family level,
521 the *Orthomyxoviridae*, *Rhabdoviridae*, and *Spinareoviridae* also infect a wide variety of
522 hosts, including mammals, birds, insects and other arthropods, along with fish in the case of
523 the *Orthomyxoviridae*, plants in the *Rhabdoviridae*, and fungi in the *Spinareoviridae*. There
524 are, however, clear phylogenetic distinctions between genera infecting vastly different hosts,
525 and host ranges are generally limited at a genus level. Genera with cross-phylum or cross-
526 kingdom host ranges are typically characterised by well-established transmission routes
527 between arthropod vectors and animal or plant hosts. For example, arboviruses in the genera
528 *Orthoflavivirus*, *Alphavirus*, *Coltivirus*, and *Orbivirus* replicate in their vertebrate hosts as
529 well as the arthropod vectors that transmit them (Huang *et al*., 2023). Similarly, plant-
530 infecting genera in the families *Tospoviridae*, *Tymoviridae* (*Marafivirus*), *Rhabdoviridae*,
531 and order *Reovirales* also replicate in their arthropod vectors (Gray and Banerjee, 1999).
532 While it is plausible that members of the *Picobirnaviridae* are capable of extensive host-
533 jumping, their remarkably wide host range raises the possibility that the picobirnaviruses are
534 in fact associated with gut microflora or dietary components present in and excreted by these
535 animals (Krishnamurthy and Wang, 2018). This would also explain why non-environmental
536 picobirnaviruses are detected almost exclusively in animal faeces and are still unable to be
537 cultured in any eukaryotic cell lines (Ghosh and Malik, 2021).
538

539 The presence of a bacterial ribosomal binding site motif is well-established in bacteriophage
540 genomes. Krishnamurthy and Wang (2018) noted that the 'AGGAGG' RBS hexamer is
541 enriched in the genomes of both RNA and DNA bacteriophage, and appears at much lower
542 frequency in eukaryote-infecting virus genomes. In our expansive data set, we detected an
543 RBS motif in 59-67% of ORFs in picobirnavirus genomes, such that this marker is
544 substantially more enriched than in confirmed RNA bacteriophages *Leviviricetes* and
545 *Cystoviridae*. This constitutes further evidence that the *Picobirnaviridae* indeed have
546 bacterial hosts.
547
548 The host range of virus species within each of the seven proposed *Picobirnaviridae* genera
549 appeared to comprise either predominantly animal-associated picobirnaviruses
550 (*Alphapicobirnavirus*, *Betapicobirnavirus, Epsilonpicobirnavirus,* and *Zetapicobirnavirus*) or
551 those identified in microbial or environmental samples (*Gammapicobirnavirus*,
552 *Deltapicobirnavirus*, and *Etapicobirnavirus*). However, the four genera comprising
553 predominantly animal-associated viruses did not group together to the exclusion of the three
554 genera comprising mainly microbe/environmentally sourced viruses. As such, this broad
555 host-associated clustering may reflect the different microbial community compositions in
556 terrestrial, aquatic, and engineered environments compared to those of mammals, birds, and
557 invertebrates. Picobirnaviruses may infect different microbial organisms that play roles in
558 food webs linking vertebrates, invertebrates, and their diets and habitats. This would facilitate
559 the extensive "host-jumping" observed within the animal-dominated genera. Hence, future
560 metagenomic studies on the viromes of animal (particularly faecal) or environmental samples
561 should also include community composition analyses of the microorganisms present.
562 Determining if the microbial community composition of a sampled animal or environment is
563 indeed driving the phylogenetic patterns of the *Picobirnaviridae* may elucidate the true host
564 range of picobirnavirus genera.
565
566
567 **ACKNOWLEDGMENTS**
568 This work was funded by a National Health and Medical Research Council (NHMRC)
569 Investigator award (GNT201719) to ECH and by AIR@InnoHK administered by the
570 Innovation and Technology Commission, Hong Kong Special Administrative Region, China.
571
572 **AUTHOR CONTRIBUTIONS**

SS, ECH and JEM conceptualised this study. SS collated and analysed the sequence data. SS wrote the original manuscript draft. SS, ECH and JEM edited and revised the manuscript. ECH and JEM supervised the project.

576 **REFERENCES**

577 Adriaenssens, E.M., Farkas, K., Harrison, C., Jones, D.L., Allison, H.E., and McCarthy, A.J.
578 (2018) Viromic analysis of wastewater input to a river catchment reveals a diverse
579 assemblage of RNA viruses. *mSystems* **3**: e00025-18.

580 Bell, N., Khamrin, P., Kumthip, K., Rojjanadumrongkul, K., Nantachit, N., and Maneekarn,
581 N. (2020) Molecular detection and characterization of *Picobirnavirus* in environmental water
582 in Thailand. *Clin Lab* **66**: doi: 10.7754/Clin.Lab.2019.191013.

583 Bonny, P., Schaeffer, J., Besnard, A., Desdouits, M., Ngang, J.J.E., and Le Guyader, F.S.
584 (2021) Human and animal RNA virus diversity detected by metagenomics in Cameroonian
585 clams. *Front Microbiol* **12**: 770385.

586 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for
587 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:
588 1972–1973.

589 Chen, Y.-M., Sadiq, S., Tian, J.-H., Chen, X., Lin, X.-D., Shen, J.-J., *et al.* (2022) RNA
590 viromes from terrestrial sites across China expand environmental viral diversity. *Nat*
591 *Microbiol* **7**: 1312–1323.

592 Cole, T.E., Hong, Y., Brasier, C.M., and Buck, K.W. (2000) Detection of an RNA-dependent
593 RNA polymerase in mitochondria from a mitovirus-infected isolate of the Dutch elm disease
594 fungus, *Ophiostoma novo-ulmi*. *Virology* **268**: 239–243.

595 Delmas, B., Attoui, H., Ghosh, S., Malik, Y.S., Mundt, E., Vakharia, V.N., and ICTV Report
596 Consortium (2019) ICTV virus taxonomy profile: *Picobirnaviridae*. *J Gen Virol* **100**: 133–
597 134.

598 Duraisamy, R., Akiana, J., Davoust, B., Mediannikov, O., Michelle, C., Robert, C., *et al.*
599 (2018) Detection of novel RNA viruses from free-living gorillas, Republic of the Congo:
600 genetic diversity of picobirnaviruses. *Virus Genes* **54**: 256–271.

601 Fong, J.J., Brown, J.M., Fujita, M.K., and Boussau, B. (2012) A phylogenomic approach to
602 vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic
603 Lissamphibia. *PLoS ONE* **7**: e48990.

604 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the
605 next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

606 Ganesh, B., Masachessi, G., and Mladenova, Z. (2014) Animal picobirnavirus. *Virus Dis* **25**:
607 223–238.

608 Geoghegan, J.L., Di Giallonardo, F., Wille, M., Ortiz-Baez, A.S., Costa, V.A., Ghaly, T., *et*
609 *al.* (2021) Virome composition in marine fish revealed by meta-transcriptomics. *Virus Evol*
610 **7**: veab005.

611 Geoghegan, J.L., Duchêne, S., and Holmes, E.C. (2017) Comparative analysis estimates the
612 relative frequencies of co-divergence and cross-species transmission within viral families.
613 *PLoS Pathog* **13**: e1006215.

614 Ghosh, S., and Malik, Y.S. (2021) The true host/s of picobirnaviruses. *Front Vet Sci* **7**:
615 615293.

616 Giannitti, F., Ullmann, L.S., Deng, X., Pesavento, P.A., Delwart, E., Pusterla, N., *et al.*
617 (2015) Exploring the virome of diseased horses. *J Gen Virol* **96**: 2721–2733.

618 Gray, S.M., and Banerjee, N. (1999) Mechanisms of arthropod transmission of plant and
619 animal viruses. *Microbiol Mol Biol Rev* **63**: 128-148.

620 Guajardo-Leiva, S., Chnaiderman, J., Gaggero, A., and Díez, B. (2020) Metagenomic
621 insights into the sewage RNA virosphere of a large city. *Viruses* **12**: 1050.

622 Hillman, B.I., and Cai, G. (2013) The family *Narnaviridae*. *Adv Virus Res* **86**: 149–176.

623 Ho, S.Y.W., Duchêne, S., and Duchêne, D. (2015) Simulating and detecting autocorrelation
624 of molecular evolutionary rates among lineages. *Mol Ecol Resour* **15**: 688–696.

625 Huang, Y., Wang, S., Liu, H., Atoni, E., Wang, F., Chen, W., *et al*. (2023) A global dataset of
626 sequence, diversity and biosafety recommendation of arbovirus and arthropod-specific virus.
627 *Sci Data* **10**: 305.

628 Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version
629 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

630 Kleymann, A., Becker, A.A.M.J., Malik, Y.S., Kobayashi, N., and Ghosh, S. (2020)
631 Detection and molecular characterization of picobirnaviruses (PBVs) in the mongoose:
632 identification of a novel PBV using an alternative genetic code. *Viruses* **12**: 99.

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

635 Le Lay, C., Shi, M., Buček, A., Bourguignon, T., Lo, N., and Holmes, E. (2020) Unmapped
636 RNA virus diversity in termites and their symbionts. *Viruses* **12**: 1145.

637 Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., *et*
638 *al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic
639 dog. *Nature* **438**: 803–819.

640 Mahar, J.E., Shi, M., Hall, R.N., Strive, T., and Holmes, E.C. (2020) Comparative analysis of
641 RNA virome composition in rabbits and associated ectoparasites. *J Virol* **94**: e02119-19.

642 Malik, Y.S., Kumar, N., Sharma, K., Dhama, K., Shabbir, M.Z., Ganesh, B., *et al.* (2014)
643 Epidemiology, phylogeny, and evolution of emerging enteric picobirnaviruses of animal
644 origin and their relationship to human strains. *BioMed Res Int* **2014**: 1–13.

645 Neri, U., Wolf, Y.I., Roux, S., Camargo, A.P., Lee, B., Kazlauskas, D., *et al.* (2022)
646 Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell* **185**:
647 4023-4037.e18.

648 Nguyen, L.-T., Schmidt, H.A., Haeseler, A. von, and Minh, B.Q. (2015) IQ-TREE: A fast
649 and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol*
650 *Evol* **32**: 268–274.

651 Paradis, E., and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and
652 evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

653 Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., and
654 Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-
655 generation DNA sequencing. *Nature* **526**: 569–573.

656 Rambaut, A. (2018) FigTree. https://github.com/rambaut/figtree/.

657 Sadiq, S., Chen, Y.-M., Zhang, Y.-Z., and Holmes, E.C. (2022) Resolving deep evolutionary
658 relationships within the RNA virus phylum *Lenarviricota*. *Virus Evol* **8**: veac055.

659 Santichaivekin, S., Yang, Q., Liu, J., Mawhorter, R., Jiang, J., Wesley, T., *et al.* (2021)
660 eMPRess: a systematic cophylogeny reconciliation tool. *Bioinformatics* **37**: 2481–2482.

661 Shen, Y.-Y., Dai, K., Cao, X., Murphy, R.W., Shen, X.-J., and Zhang, Y.-P. (2014) The
662 updated phylogenies of the *Phasianidae* based on combined data of nuclear and
663 mitochondrial DNA. *PLoS ONE* **9**: e95786.

664 Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., *et al.* (2016) Redefining the
665 invertebrate RNA virosphere. *Nature* **540**: 539–543.

666 Song, S., Liu, L., Edwards, S.V., and Wu, S. (2012) Resolving conflict in eutherian mammal
667 phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci*
668 *USA* **109**: 14942–14947.

669 Urayama, S., Takaki, Y., and Nunoura, T. (2016) FLDS: a comprehensive dsRNA
670 sequencing method for intracellular RNA virus surveillance. *Microbe Environ* **31**: 33–40.

671 Vainio, E.J., Chiba, S., Ghabrial, S.A., Maiss, E., Roossinck, M., Sabanadzovic, S., *et al.*
672 (2018) ICTV virus taxonomy profile: *Partitiviridae*. *J Gen Virol* **99**: 17–18.

673 Wakuda, M., Pongsuwanna, Y., and Taniguchi, K. (2005) Complete nucleotide sequences of
674 two RNA segments of human picobirnavirus. *J Virol Meth* **126**: 165–169.

675 Wolf, Y.I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., *et al.* (2020) Doubling of
676 the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol*
677 **5**: 1262–1270.

678 Woo, P.C.Y., Teng, J.L.L., Bai, R., Tang, Y., Wong, A.Y.P., Li, K.S.M., *et al.* (2019) Novel
679 picobirnaviruses in respiratory and alimentary tracts of cattle and monkeys with large intra-
680 and inter-host diversity. *Viruses* **11**: 574.

681 Yinda, C.K., Vanhulle, E., Conceição-Neto, N., Beller, L., Deboutte, W., Shi, C., *et al.*
682 (2019) Gut virome analysis of Cameroonians reveals high diversity of enteric viruses,
683 including potential interspecies transmitted viruses. *mSphere* **4**: e00585-18.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T. (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36.

Zhou, Y., Wang, S.-R., and Ma, J.-Z. (2017) Comprehensive species set revealing the phylogeny and biogeography of Feliformia (Mammalia, Carnivora) based on mitochondrial DNA. *PLoS ONE* **12**: e0174902.

Zurano, J.P., Magalhães, F.M., Asato, A.E., Silva, G., Bidau, C.J., Mesquita, D.O., and Costa, G.C. (2019) Cetartiodactyla: Updating a time-calibrated molecular phylogeny. *Mol Phyl Evol* **133**: 256–262.