Metagenomic identification of diverse animal hepaciviruses and pegiviruses

Ashleigh F. Porter[1], John H.-O. Pettersson[1,2], Wei-Shan Chang[1], Erin Harvey[1], Karrie Rose[3], Mang Shi[5], John-Sebastian Eden[1,4], Jan Buchmann[1], Craig Moritz[6], Edward C. Holmes[1]

1. Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, Australia.
2. Zoonosis Science Center, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden.
3. Australian Registry of Wildlife Health, Taronga Conservation Society Australia, Mosman, Australia.
4. Centre for Virus Research, Westmead Institute for Medical Research, Westmead, Australia.
5. School of Medicine, Sun Yat-sen University, Guangzhou, China.
6. Research School of Biology, and Centre for Biodiversity Analysis, Australian National University, Acton, ACT, Australia.

Address correspondence to:
Edward C. Holmes,
Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, Australia
edward.holmes@sydney.edu.au

Abstract

The RNA virus family *Flaviviridae* harbours several important pathogens of humans and other animals, including Zika virus, dengue virus and hepatitis C virus. The *Flaviviridae* are currently divided into four genera - *Hepacivirus*, *Pegivirus*, *Pestivirus* and *Flavivirus* – each of which have a diverse host range. Members of the genus *Hepacivirus* are associated with a diverse array of animal species, including humans and non-human primates, other mammalian species, as well as birds and fish, while the closely related pegiviruses have been identified in a variety of mammalian taxa including humans. Using a combination of meta-transcriptomic and whole genome sequencing we identified four novel hepaciviruses and one novel variant of a known virus, in five species of native Australian wildlife, expanding our knowledge of the diversity in this important group of RNA viruses. The infected hosts comprised native Australian marsupials and birds, as well as a native gecko (*Gehyra lauta*). The addition of these novel viruses led to the identification of a distinct marsupial clade within the hepacivirus phylogeny that also included an engorged *Ixodes holocyclus* tick collected while feeding on Australian long-nosed bandicoots (*Perameles nasuta*). Gecko and avian associated hepacivirus lineages were also identified. In addition, by mining the short-read archive (SRA) database we identified another five novel members of *Flaviviridae,* comprising three new hepaciviruses from avian and primate hosts, as well as two primate-associated pegiviruses. The large-scale phylogenetic analysis of these novel hepacivirus and pegivirus genomes provides additional support for virus-host co-divergence over evolutionary time-scales.

Introduction

As the vast majority of emerging infectious disease in humans are caused by viral zoonoses [1], the identification and characterization of animal viruses is critical for identifying potential disease reservoirs and providing models for the study of human viruses [2,3]. Two related groups of viruses that have recently received considerable attention are the genera *Hepacivirus* and *Pegivirus* from the family *Flaviviridae* of single-strand positive-sense RNA viruses. Hepaciviruses infect a broad range of vertebrate hosts, including humans, non-human primates [4,5], and a variety of other mammals including rodents [6-10], horses [11], bats [12], and cows [13,14]. Hepaciviruses have also been detected in birds [15, 16], fish and a variety of other vertebrates [17, 18]. Additionally, two hepacivirus-like sequences have been identified in arthropods (a mosquito and tick), although their true host is uncertain [19, 20]. Despite such a wide diversity of animal hosts, hepaciviruses remain synonymous with liver infection, with the most notable example being human hepatitis C virus (HCV). While some non-primate hepaciviruses have been well characterized, particularly equine hepacivirus (also known as Hepacivirus AK, or non-primate hepacivirus, NPHV) and canine hepacivirus (CHV), it seems likely that HCV, along with NPHV and CHV, arose from a currently unknown zoonotic source, with rodents and bats suggested as possible reservoirs [21-24].

The genus *Pegivirus* contains 11 defined virus species known to infect humans and variety of other mammals. In addition, a novel avian pegivirus was recently identified in a Common myna bird (*Acridotheres tristis*) [25]. Human pegivirus (HPgV), previously known as GB virus C, is known to infect humans but with an unproven link with clinical illness [26], despite being identified in the brain tissue of patients with encephalitis [27]. Non-human primate pegiviruses have been identified in both New World monkeys [28, 29] and Old World apes, including chimpanzees [30-32], the latter of which ($SPgV_{cpz}$) are closely related to HPgV [33]. Other pegiviruses have been found in horses [34], bats [12, 35] and rodents [7,8].

Despite their broad host range, all hepaciviruses and pegiviruses described to date possess a single-strand positive-sense RNA genome and a large open reading frame encoding a single polyprotein flanked by untranslated regions. Although this structure is common among the *Flaviviridae* [36, 37], a more diverse set of genome structures, including segmented forms, have been identified in invertebrate flaviviruses [18]. The multifunctional polyprotein is cleaved by proteases to create ten proteins: three structural (core, E1, E2) and seven non-structural (NS1, NS2, NS3, NS4a, NS4b, NS5a, NS5b) [24]. It is important to

note that while many members of the genus *Flavivirus* are transmitted via arthropod vectors, and it has been hypothesised that biting arthropod vectors may transmit some hepaciviruses [37], there is currently no evidence of vector-borne transmission in either the *Hepacivirus* or *Pegivirus* genera. Recently, however, a novel hepacivirus, Brushtail possum hepacivirus, was discovered in brush-tailed possums (*Trichosurus vulpecula*), a native Australian marsupial [38]. Similarly, Collins beach hepacivirus was identified in a tick (*Ixodes holocyclus*) feeding on another Australian marsupial (long-nosed bandicoot, *Perameles nasuta*) [19]. These data suggest that Australian native wildlife harbour a diversity of hepaciviruses worthy of further exploration.

As hepaciviruses have traditionally been difficult to culture, the recent expansion of animal hepaciviruses is largely due to the advent of unbiased high-throughput sequencing. This expansion has greatly impacted our understanding of the diversity and evolution of this important genus. Herein, we used a bulk RNA-sequencing (meta-transcriptomic) approach to identify additional novel hepaciviruses in Australian wildlife and determine their abundance. This approach has previously proven successful in an Australian context, identifying a variety of viruses in Australian wildlife, including native Australian and invasive species [19, 38-43]. To supplement this analysis, we mined the sequence read archive (SRA) database for hepacivirus-like sequences. Previously generated RNA sequencing data, such as those found in the SRA, are a relatively untapped resource for novel virus discovery [44]. Through SRA-mining we identified three novel hepaciviruses and two novel pegiviruses, again highlighting the power of genomics to identify novel viral sequences.

Materials and Methods

Animal ethics

The magpie lark and pelican were handled under a series of NSW Office of Environment and Heritage Licences to Rehabilitate Injured, Sick or Orphaned Protected Wildlife (#MWL000100542). Samples were collected under the Opportunistic Sample Collection Program of the Taronga Animal Ethics Committee, and scientific licence SL100104 issued by the NSW Office of Environment and Heritage. Ticks were removed from the long-nosed bandicoot under the approval of the NSW Office of Environment and Heritage Animal Ethics Committee (#000214/05) and scientific license SL100104.

Sample collection

The samples analysed here were collected from a variety of sources (Table 1). Gecko hepacivirus (RNASeq library VERT7) was obtained from a liver sample (#CCM0247) of a seemingly healthy gecko (*Gehyra lauta*) collected at West Leichardt Station, Queensland in 2013 [45]. Pelican hepacivirus (RNASeq library VERT5) was obtained from the liver (Australian Registry of Wildlife Health #9381.1) of an Australian pelican (*Pelecanus conspicillatus*) collected at Blackwall Bay, New South Wales (NSW) in 2013. The pelican presented with an ongoing syndrome of profound weakness, diarrhoea, dyspnoea, and death, consistently associated with myocardial degeneration to necrosis. Magpie lark hepacivirus (Australian Registry of Wildlife Health #9585.8) was obtained from brain (RNASeq library VERT14) and heart samples (RNASeq library VERT 15) of an injured juvenile magpie lark (*Grallina cyanoleuca*) collected at Warigee, NSW in 2013 [46]. Collins Beach virus 1 was identified from three engorged female ticks (*Ixodes holocylus)* collected while feeding on a healthy long-nose bandicoot (*Perameles nasuta*) at North Head, NSW in 2016. These individual tick samples were part of a larger pool of ticks (RNASeq library TICK08, SRA projects SRS3932533 and SRS3932534) used to identify the partial genome sequence of Collins Beach virus [19]. The Koala hepacivirus was identified during our initial SRA mining of marsupial transcriptomes (Supplementary Table 1). In this case, the mRNA library (SRX501262) was prepared from the liver RNA of a deceased Australian koala (*Phascolarctos cinereus)* "Pacific Chocolate" that was known to be infected with chlamydia [46]. As viral coverage was incomplete due to the poly-A selection method, the original tissue sample (#M.45022.004) was kindly provided by the Australian Museum and subjected to total RNA sequencing (i.e. rRNA-depletion only) along with our other cases.

Table 1. Sample information for the novel hepaciviruses identified here, including the host species, library, region isolated, host pathology, and assembly data.

Virus name	*Acronym*	*Host (common)*	*Host (scientific)*	*RNA Library (Tissue)*	*Region*	*Pathology*	*Viral reads per library (abundance %)*	*Host marker reads per library (abundance %)**
Magpie lark hepacivirus	MaHV	Magpie lark	*Grallina cyanoleuca*	VERT14 (Brain); VERT15 (Heart)	Nowra, NSW	Hepatic rupture, urate nephrosis, ventricular haemorrhage, thin	VERT14 = 26 (0.00005%) VERT15 = 172 (0.00045%)	VERT14 = 3,095 (0.0063%) VERT15 = 5,140 (0.013%)
Collins beach virus 1	BaHV	Long-nosed bandicoot	*Perameles nasuta*	TICK07 (Tick); TICK08 (Tick); TICK10 (Tick); INVERT13 (Tick)	Sydney, NSW	Healthy	TICK07 = 6 (0.00001%) TICK08 = 430 (0.00079%) TICK10 = 6 (0.00001%) INVERT13 = 43 (0.00014%)	TICK07 = 19,752 (0.035%) TICK08 = 33,706 (0.062%) TICK10 = 6,274 (0.012%) INVERT13 = 12,207 (0.041%)
Koala hepacivirus	KHV	Koala	*Phascolarctos cinereus*	VERT20 (Liver)	Port Macquarie, NSW	Severe chlamydiosis	VERT20 = 29,391 (0.10%)	VERT20 = 3,205 (0.011%)
Pelican hepacivirus	PeHV	Australian pelican	*Pelecanus conspicillatus*	VERT5 (Liver); VERT44 (Liver)	Blackwall Bay, NSW	Profound weakness, diarrhoea, dyspnoea, and death, associated with myocardial degeneration to necrosis	VERT5 = 172 (0.00034%) VERT44 = 362 (0.0014%)	VERT5 = 8,642 (0.017%) VERT44 = 2,978 (0.011%)
Gecko hepacivirus	GHV	House gecko	*Gehyra lauta*	VERT7 (Liver)	West Leichardt Station, QLD	Healthy	VERT7 = 762 (0.0018%)	VERT7 = 6,651 (0.015%)

*Abundance of host determined using Ribosomal Protein L13a (RPL13A).

149
150 *RNA extraction, library preparation and sequencing*
151 Viral RNA was extracted from individual tissues of animal samples using the RNeasy Plus
152 Mini Kit (Qiagen, Germany). RNA concentration and integrity were determined using a
153 NanoDrop spectrophotometer (ThermoFisher) and TapeStation (Agilent). RNA samples were
154 pooled in equal proportions based on animal tissue types and syndrome (maximum eight
155 individuals per library). Illumina TruSeq stranded RNA libraries were prepared on the pooled
156 samples following rRNA depletion using a RiboZero Gold kit (Epidemiology) at the
157 Australian Genome Research Facility (AGRF), Melbourne. The rRNA depleted libraries were
158 then sequenced on an Illumina HiSeq 2500 system (Paired end 100 bp or 125 bp reads) to
159 depths of between 13-28 million paired reads.
160
161 *Viral discovery pipeline*
162 We employed an established meta-transcriptomic pipeline developed for RNA virus
163 discovery [18, 19, 47]. RNA sequences were trimmed of low-quality bases and adapter
164 sequences with Trimmomatic v0.36 [48] before *de novo* assembly using Trinity 2.5.1 [49].
165 The assembled contigs were then compared against the NCBI nucleotide (nt) and non-
166 redundant (nr) protein databases with Blastn v2.7.1+ [50] and Diamond v0.9.18 [51],
167 respectively, with an e-value threshold of 1E-5. Contig abundance was estimated by
168 calculating the number of reads in each library that mapped to the hepacivirus genome
169 divided by the number of total reads (Supplementary Table 2). Similarly, host abundance
170 was compared by mapping reads to a reference gene - ribosomal protein L13a (RPL13A).
171 Low abundance viruses (i.e. those that could not be assembled) were identified by aligning
172 the trimmed reads against either NCBI viral RefSeq or a curated hepacivirus/pegivirus
173 protein database using Diamond v.0.9.18, with an e-value threshold of 1E-4. Curated
174 hepacivirus/pegivirus databases were regularly updated with novel viruses to increase read
175 alignment for divergent viral species.
176
177 *PCR confirmation and whole-genome sequencing*
178 To confirm the presence of novel viruses in individual samples, RT-PCR was performed
179 using primers designed to amplify a short region of the viral genome (<300bp). Briefly,
180 Superscript IV VILO master mix was used to generate cDNA from individual RNA samples
181 before RT-PCR screening with Platinum SuperFi. Following the confirmation of viruses to
182 individual samples, long (2-4kb) overlapping PCRs were designed to amplify the available
183 viral genome (i.e. complete or partial) for the hepaciviruses identified and sequenced using

the Nextera XT library prep kit and Illumina MiSeq sequencing (150nt paired reads). Viral genomes were then assembled using MegaHit v1.1.3 [52].

SRA mining

To identify hepacivirus- and pegivirus-like viruses present in the SRA, we screened primates (excluding *Homo sapiens)*, birds (excluding *Gallus gallus)* and marsupials. We focused on these taxonomic groups because (i) although HCV is clearly a virus of humans, it is unclear whether it is present in other non-human primates (which may be indicative of virus-host co-divergence), (ii) a number of novel hepacivirus- and pegivirus-like viruses have recently been identified in avian hosts suggesting that these might be a rich source of novel viruses, and (iii) to confirm the presence of a marsupial specific-lineage as suggested by our previous studies. Accordingly, three different sets of single- and paired-end RNA sequencing SRAs were analysed: one included 2,312 avian SRAs (Supplementary Table 3), a second that included 3,902 primate data sets (Supplementary Table 4) and finally a marsupial data set of 330 SRAs (Supplementary Table 1). SRAs were fetched with the NCBI SRA toolkit using fastq-dump and analysed using our virus discovery pipeline, with the exception that the initial viral screening was performed by aligning the downloaded reads to our curated hepacivirus/pegivirus protein database with Diamond v.0.9.18. SRAs that contained hepacivirus-like and/or pegivirus-like reads were then assembled and annotated as above.

Three virus fragments were identified in three runs (SRR1325073, SRR1325072, SRR1325074) from SRA project SRX565268, isolated from Eurasian blue tit (*Cyanistes caeruleus)* from Germany. These three fragments were concatenated into a single genome, Blue tit hepacivirus (BtHV). The three fragments, and their respective coding regions, are shown in Supplementary Figure 1.

Genome annotation

The hepaciviruses identified from our samples and via SRA mining were subjected to an online sequence similarity search using NCBI blastx as well as a conserved domain (CDD) search. ORFs of verified contigs were included in a phylogenetic analysis as described below. Genome annotation of the 10 novel virus sequences was performed using Geneious Prime (version 2019.2.1) [53]. Specifically, using the Live Annotate and Predict tool, each novel hepacivirus genome was screened for hepacivirus-specific gene annotations utilising previously identified annotations in hepacivirus genomes downloaded from NCBI (n = 145).

To maximise the identification of hypothetical genes, even in divergent genomes, the gene similarity cut-off was set to 25%, with predicted genes then manually annotated.

Phylogenetic analysis

To investigate the evolutionary relationships among the hepaciviruses and pegiviruses we collated the amino acid sequences for the complete polyprotein of each genome from the ten novel viruses identified here (see Results), and from hepaciviruses and pegiviruses taken from NCBI Refseq database including Brushtail possum hepacivirus and Collins Beach hepacivirus (n =110, accession numbers available in Supplementary Table 5). These sequences were aligned using the E-INS-i algorithm in MAFFT v7 [54], with ambiguous regions removed using GBlocks [55]. The final sequence alignment comprised 110 sequences of 1345 amino acid residues in length. ProtTest v3.4 [56] was used to determine the most appropriate model of amino acid substitution for these data. From this, we estimated a phylogenetic tree using the maximum likelihood method in IQ-TREE, version 1.6.12 [57], employing the LG model of amino acid substitution with invariable sites and gamma model (4 categories), along with a bootstrap resampling analysis using 1000 replicates. Two additional phylogenetic trees were estimated using the NS2/3 (549 amino acid residues) and NS5 (407 amino acid residues) regions using the same procedure as described above (although with 100 bootstrap replicates).

Results

Novel hepaciviruses discovered in Australian wildlife

We identified four novel hepaciviruses and one new hepacivirus variant from five meta-transcriptomic libraries obtained from a variety of Australian wildlife samples (Table 1). The four novel hepaci-like genomes comprised: (i) Koala hepacivirus (denoted KHV); (ii) Pelican hepacivirus (PeHV), (ii) Magpie lark hepacivirus (MaHV); and (iv) Gecko hepacivirus (GHV). In addition, we identified a variant of the previously identified Collins beach virus isolated from ticks feeding on long-nosed bandicoots [19]. In accordance with the naming convention for hepacivirus variants [58] we term this Collins beach virus 1 (CBV1). The relative abundance of hepacivirus-like sequences in each library was low (Figure 1 and Table 1), with the abundance of the host RPL13A gene shown for comparison (Figure 1). The Koala hepacivirus was the most abundant, comprising 0.1% of the total reads and a complete genome could be assembled at a mean coverage depth of 386X. Reads associated with CBV1 were at an abundance <1% in all tick libraries. The abundance levels for the remaining novel hepaciviruses ranged from 0.0018-0.00001% of the reads in their

respective libraries, and most of the genomes were in-complete. Therefore, for these remaining viruses, we used a long, overlapping amplicon-based sequencing approach to fill gaps in the RNASeq data and from this produce complete or near-complete genomes for further analysis.

Novel hepaciviruses and pegiviruses discovered from SRA data mining

Three hepacivirus sequences and two pegivirus sequences were identified by mining the SRA (Table 2). The hepacivirus genomes identified were isolated from (i) a Diademed sifaka (*Propithecus diadema*) sampled in Toamasina province, Madagascar, and termed Sifaka hepacivirus (SfHV-mad), (ii) a Senegal bushbaby (*Galago senegalensis*) virus termed Bushbaby hepacivirus (BbHV), and (iii) three hepacivirus-like fragments recovered from a Eurasian blue tit (*Cyanistes caeruleus*) in Montpellier, France, that were compiled into a viral single genome termed Blue tit hepacivirus (BtHV). The two pegivirus sequences were isolated from a (i) Common marmoset (*Callithrix jacchus*) termed Marmoset pegivirus (MHV), and (ii) two pegivirus-like fragments recovered from a South African Vervet monkey (*Chlorocebus pygerythrus*) that were assembled into a single viral genome termed Simian pegivirus (SPV-saf). The only marsupial hepacivirus-like sequence identified in the SRA screen was the partial fragment of the Koala hepacivirus identified in library SRX501262 that we re-sequenced to obtain a complete viral genome (VERT20). All individual sequence fragments are described in the Supplementary materials.

Table 2. The five novel hepaciviruses and pegiviruses identified from SRA-mining, including host information, project runs (SRR) and SRA project.

Virus name	*Acronym*	*Host (common)*	*Host (scientific)*	*Runs (SRR)*	*SRA project*
Blue tit hepacivirus	BtHV	Eurasian blue tit	*Cyanistes caeruleus*	SRR1325073, SRR1325072, SRR1325074	SRX565268
Bushbaby hepacivirus	BbHV	Senegal bushbaby	*Galago senegalensis*	SRR361358	SRX104357
Sifaka hepacivirus	SfHV-mad	Diademed sifaka	*Propithecus diadema*	SRR3131110	SRX1550742
Marmoset pegivirus	MPV	Common marmoset	*Callithrix jacchus*	SRR1758976	SRX843207
Simian pegivirus	SPV-saf	Vervet monkey	Chlorocebus pygerythrus	SRR1046735, SRR1046733	SRX389659

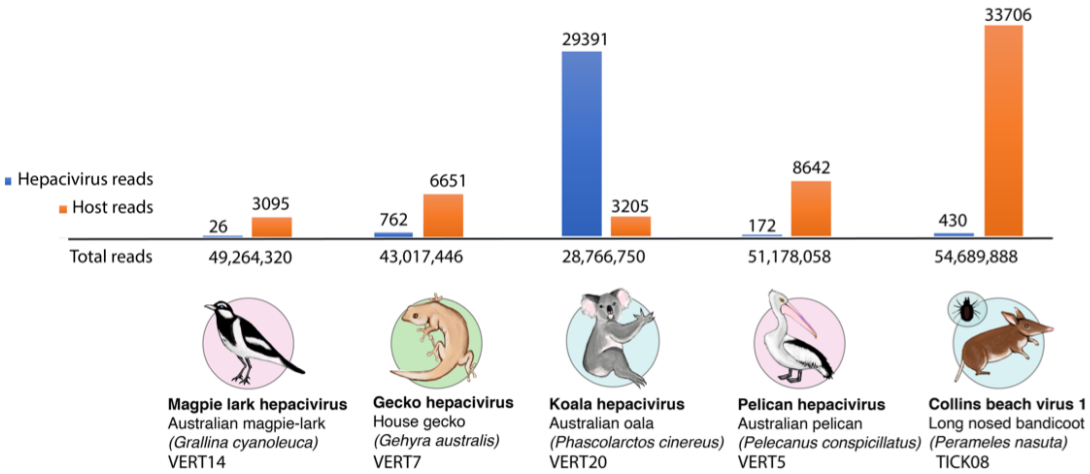
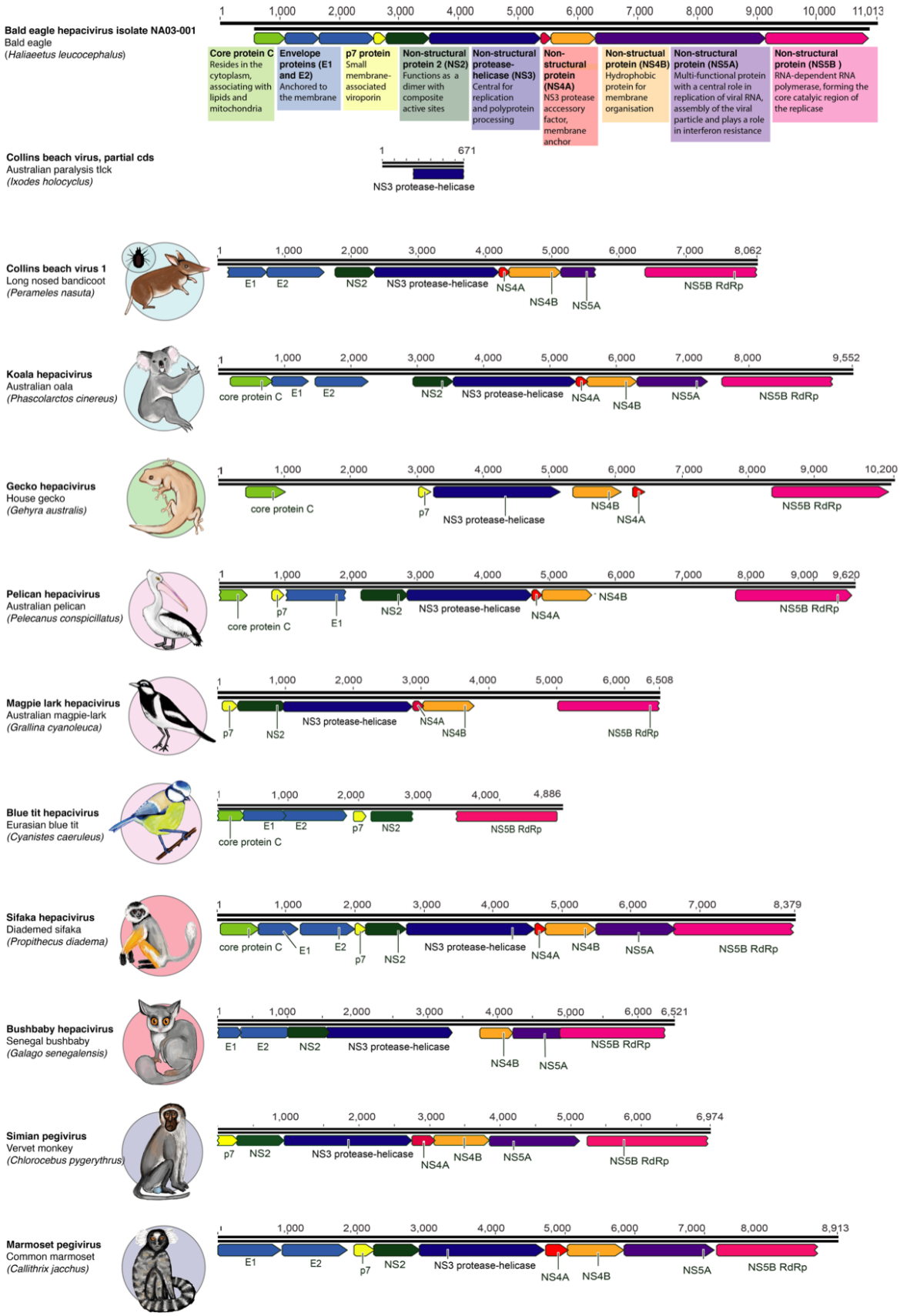


Figure 1. Abundance of hepacivirus-like contigs in each RNA-Seq library. The relative abundance of hepacivirus-like contigs (shown in blue) is presented as the number of hepacivirus-like reads compared to the number of host reads (based on the RPL13A gene, shown in orange) and the total number of reads each library (shown on the horizontal axis).

Genome annotation of novel viruses

Each of the novel hepaciviruses identified here, as well as the variant of Collins Beach hepacivirus, underwent genome annotation and were compared (Figure 2) to the fully annotated Bald eagle hepacivirus as a reference [16]. All of the novel viruses encoded near-complete polyproteins (described in Figure 2, highlighted in boxes). Two relatively well conserved *Flaviviridae* proteins, the NS5B protein that contains RNA-dependent RNA polymerase (RdRp) function and the NS3 protease-helicase protein (Figure 2, highlighted in pink and blue, respectively), were present in all the novel flavivirus polyproteins. The remaining flavivirus proteins, comprising the core protein C, protein p7, envelope protein E1, envelope protein E2, NS2, NS4A, NS4b, NS5A were identified in most novel virus genomes, although were unable to successfully identify and annotate NS5A (highlighted in purple, Figure 2) in any of the novel bird or reptile hepaciviruses at our threshold levels (25% sequence similarity). Two of the SRA derived viral sequences - Blue tit hepacivirus and Simian pegivirus - comprised partial genomes only and hence do not include all annotations. The more divergent Gecko hepacivirus was similarly not fully annotated. The genome regions missing from the annotation were the envelope proteins (E1 and E2) and the non-structural proteins NS2 and NS5A (Fig 2). Despite these seemingly incomplete annotations, all of the novel viruses have near-complete polyprotein encoding regions, suggesting that successful annotation has simply been prevented by high levels of sequence divergence in some genes.



306

307 **Figure 2. Genome annotation of the novel viruses identified in this study.** The genome
308 annotation of each virus is shown in comparison to Bald eagle hepacivirus (isolate NA03-
309 001, accession MN062427) as a reference. The conserved *Flaviviridae* polyprotein is

cleaved into several distinct protein products, represented in the diagram as coloured arrows. A short description of the function of each protein [59, 60] is shown underneath the Bald eagle hepacivirus genome, highlighted in coloured boxes that refer to each protein. Core protein C (light green), envelope proteins (light blue), p7 (yellow), NS2 (dark green), NS3 (purple), NS4A (red), NS4B (orange), NS5A (purple) and NS5B (pink).

Phylogenetic relationships of the novel viruses

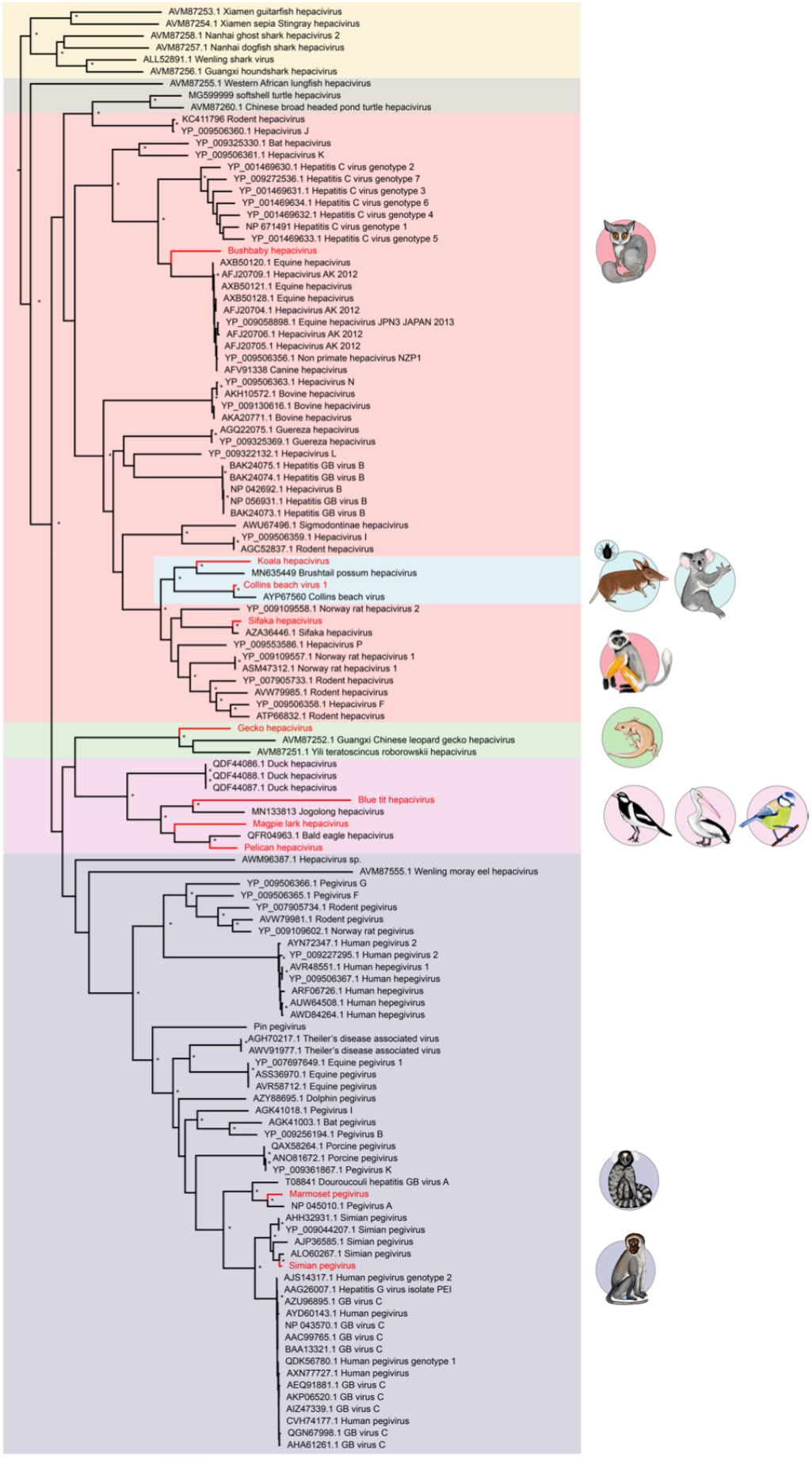
To determine the evolutionary relationships of the novel hepaci- and pegiviruses identified here we utilised the amino acid sequences of complete polypeptides from each genome in an alignment with other known pegiviruses and hepaciviruses (n = 110) (Figure 3). Maximum likelihood phylogenetic analysis revealed that the Australian Magpie lark hepacivirus (MaHV) and Pelican hepacivirus (PeHV) fall with known bird hepaciviruses, as does Blue tit hepacivirus (BtHV) (highlighted in the pink box, Figure 3). The three variants of previously identified duck hepacivirus (DuHV) [15] fall into a distinct, closely-related cluster within this clade, as do Bald eagle hepacivirus [15,16], MaHV, PeHV and BtHV, and Jogolong hepacivirus, recently isolated from a mosquito feeding on a bird host (see below) in Western Australia [20]. The existence of such an avian-specific clade is compatible with the notion that there has been some virus-host co-divergence across the hepaciviruses as a whole, particularly if the tree is rooted (as here) using the fish-associated hepacivirus that represent the most phylogenetically divergent host species [17] (Figure 3). For example, the lungfish hepacivirus falls as a sister-group to the tetrapod hepaciviruses as expected under virus-host co-divergence, and two turtle-infecting hepaciviruses, Softshell turtle hepacivirus and Chinese broad-headed pond turtle hepacivirus, are the sister-group to the mammalian hepaciviruses (highlighted in the brown box, Figure 3), although it is notable that two rodent hepaciviruses cluster anomalously with them. In this respect, it is noteworthy that the newly identified Gecko hepacivirus falls alongside other known gecko hepaciviruses - Yili teratoscincus roborowskii hepacivirus and Guangzi Chinese leopard gecko hepacivirus, both sampled in China - in a distinct gecko clade (highlighted in the green box, Figure 3).

The mammalian hepacivirus clade is the largest and best characterised and contains four of the novel viruses identified in this study. First, the two novel primate hepaciviruses identified via SRA mining fall within this clade, although they occupy different phylogenetic positions (Figure 3, red box). Sifaka hepacivirus (SfHV-mad) groups with a previously identified SfHV strain [61], and both viruses were isolated from different samples of *Propithecus diadema* and share 91% amino acid identity. These two viruses then cluster

with a group of rodent hepaciviruses. In contrast, Bushbaby hepacivirus (BbHV), sampled from a Senegal bushbaby, groups with the equine hepaciviruses (termed NPHV, equine hepaciviruses or Hepacivirus AK) and canine hepaciviruses (CHV). Notably, sequences of the human pathogen HCV fall as a sister-group to this clade. Importantly, however, as the tree is poorly supported on the Bushbaby hepacivirus branch, it is still uncertain whether Bushbaby hepacivirus is more closely related to the equine/canine hepaciviruses or to HCV.

The remaining novel hepaciviruses identified in the mammalian clade were obtained from marsupial hosts, the Australian koala and long-nosed bandicoot, although via tick bloodmeal in the case of Collins Beach virus 1. These viruses fall in a distinct marsupial cluster that is most closely related to the hepaciviruses identified in rodents and the sifaka. In addition to these two novel marsupial hepaciviruses, the previously identified Brushtail possum hepacivirus and Collins beach hepacivirus fall into this clade. Notably, the partial sequence of Collins beach hepacivirus was found in the same pooled tick libraries as the Collins beach virus 1 described here and share relatively high sequence similarity (74% nucleotide identity, 89% amino acid identity). The ticks that made up these libraries were sampled from long-nosed bandicoots in New South Wales, Australia [19].

Two novel pegiviruses were identified from the SRA - Simian pegivirus (SPV-saf) and Marmoset pegivirus (Table 2). These fall in distinct positions within the pegivirus clade of the phylogenetic tree in a manner that seems to follow the evolutionary relationships of their hosts (Figure 3). Specifically, Simian pegivirus, identified in a Vervet monkey sample, falls close to other Old World primate-associated pegiviruses, including Simian pegiviruses isolated from Yellow baboons (*Papio cynocephalus*) from Tanzania [62], African green monkey (*Chlorocebus sabaeus*) from Gambia, and Red colobus monkey (*Piliocolobus tephrosceles*) from Uganda [63]. In contrast, Marmoset pegivirus, isolated from a common marmoset, falls with New World primate-associated pegiviruses (Pegivirus A, Douroucouli hepatitis GB virus A), identified in tamarin monkeys (*Saguinus labiatus*), mystax monkeys (*Saguinus mystax*) and owl monkeys (*Aotus trivirgatus*) [64].



374

Figure 3. Phylogenetic analysis of the novel viruses identified in this study, identified by animal symbols, along with 106 known hepaciviruses and pegiviruses. Novel Blue tit hepacivirus, Magpie lark hepacivirus and Pelican hepacivirus, and other avian hepaciviruses are shown in the pink box. The novel Gecko hepacivirus is highlighted in the green box. Fish- and shark-associated hepaciviruses are shown in the yellow box. Turtle-associated hepaciviruses are shown in the brown box. Marsupial hepaciviruses, including those identified in bandicoots and koalas, are shown in the blue box. Bushbaby hepacivirus and Simian hepacivirus fall in the mammalian clade shown in the red box. Two novel pegiviruses, Marmoset pegivirus and Simian pegivirus, are shown in the purple box. Bootstrap values greater than 85% are shown next to the relevant nodes, represented by an asterisk. All horizontal branch lengths are scaled according to the number of amino acid substitutions per site.

Finally, it is notable two previously identified hepaciviruses, Hepacivirus sp., identified in a Red-eared slider turtle (*Trachemys scripta elegans*) and Wenling moray eel hepacivirus, identified from a moray eel (*Gymnothorax reticularis*), fall basal to all known pegiviruses, even though other fish and turtle viruses group with the hepaciviruses. To investigate this issue in more detail we performed an additional phylogenetic analysis of amino acid sequences of the NS2/3 and NS5 regions separately (Figure 4). While the phylogenetic position of the Red-eared slider turtle remains unchanged, the moray eel virus fell in markedly different positions in these two phylogenies, falling closer to the other fish hepaciviruses in the NS2/3 phylogeny (as expected with virus-host co-divergence) yet in a highly divergent position in the NS5 phylogeny. While the underlying cause of these disparate phylogenetic positions is uncertain and could plausibly have resulted from with recombination (with unknown parental sequences) or extreme rate variation, it does mean that the position of the moray eel virus as a sister-group to the pegiviruses in the polyprotein phylogeny is artifactual. While a number of other clades change position between the NS2/3 and NS5 phylogenies, particularly the fish and lungfish viruses that occupy basal positions in the polyprotein phylogeny, the deeper nodes on the NS2/3 and NS5 trees receive only weak bootstrap support such that these movements may simply reflect a lack of phylogenetic resolution. In addition, the position of the viruses newly identified here relative to those described previously remained unchanged between the NS2/3 and NS5 phylogenies.

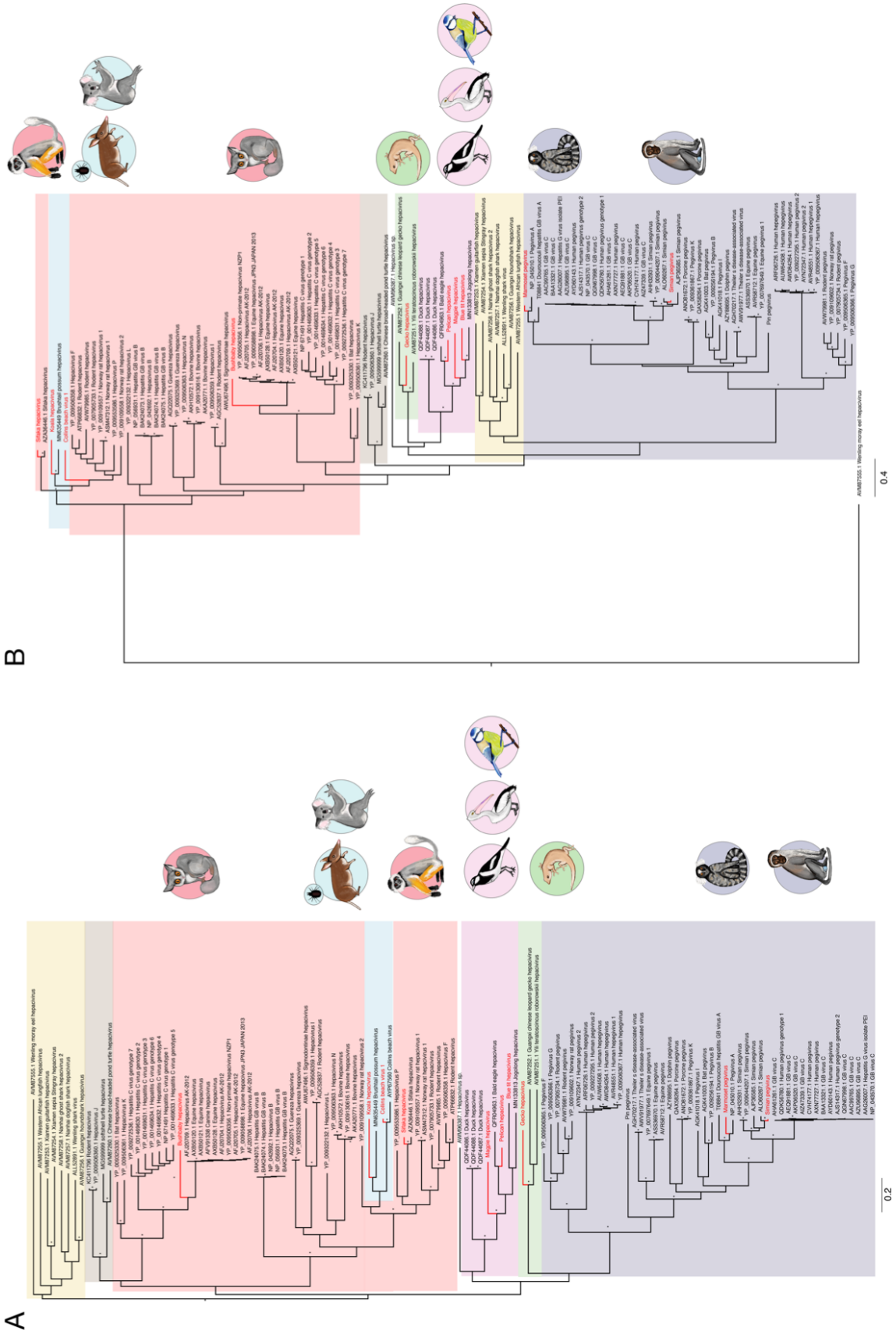


Figure 4. Phylogenetic analysis of the (A) NS2/NS3 and (B) NS5 amino acid sequences from the novel hepaciviruses and pegiviruses identified in this study, marked by animal

symbols, combined with 106 known hepaciviruses and pegiviruses. Animal groups are shaded as described in Figure 3. Bootstrap values greater than 85% (from a total of 100 bootstrap replicates in each case) are shown next to the relevant nodes, represented by an asterisk. All horizontal branch lengths are scaled to the number of amino acid substitutions per site. Note that Collins beach virus is necessarily excluded from the NS5 phylogeny as it lacks this gene region.

Discussion

Using a meta-transcriptomic approach we identified four novel hepaciviruses and a single variant of a previously described hepacivirus in Australian marsupials, birds, and a reptile. Additionally, through SRA-mining, we were able to discover three novel hepaciviruses and two novel pegiviruses in primate hosts. Accordingly, this work broadens our knowledge of two viral genera within the *Flaviviridae* and suggests that, with increased sampling, hepaciviruses and pegiviruses will likely be identified in many other animal hosts.

Hepaciviruses encode a single polyprotein that is commonly observed among vertebrate members of the *Flaviviridae.* The hepaciviruses identified in this study were characterised by high conservation in the NS5B, RdRp and NS3 genes. Other proteins were present in most of the viral genomes identified here, although the NS5A gene was not readily identified in the avian and gecko hepaciviruses, while the two envelope genes could not be identified in the Gecko hepacivirus. Based on the genome structure of the novel hepaciviruses, and that complete encoding region of the polyprotein is likely present, we anticipate that these genes are present in these viruses but are highly divergent in sequence and hence cannot be easily be identified using approaches that only assess primary sequence homology.

As hepaciviruses and pegiviruses are commonly thought to have co-diverged with their hosts over time-scales of many millions of years, we expected that these viruses would follow an evolutionary pattern similar to that of their hosts [17]. Indeed, our phylogenetic analysis demonstrates that this pattern generally holds true. In particular, the avian, reptile, mammal and marsupial viruses fall into distinct clades and the virus phylogeny broadly follows that of the hosts, although a number of the key nodes are only poorly supported. There are, however, some notable exceptions, such as rodent hepacivirus and Hepacivirus J, identified in bank voles (*Myodes glareolus*) from Germany [8], that clustered with turtle-

associated hepaciviruses. Similarly, rather than falling as a sister-group to those hepaciviruses sampled from placental mammals as expected under virus-host co-divergence, it is notable that the marsupial hepaciviruses fall within the mammalian (eutherian) phylogenetic diversity. Finally, novel Pin pegivirus [25], isolated from a Common myna bird (*Acridotheres tristis*), falls within the diversity of mammalian pegiviruses. As this is the first avian pegivirus identified, it is possible that a distinct avian clade will be identified with increased sampling.

The avian hepacivirus clade characterised here contains Blue tit hepacivirus (BtHV), recovered from the SRA, as well the novel Australian pelican and magpie lark hepaciviruses (PeHV and MaHV, respectively), three variants of duck hepacivirus [15], and the recently described Jogalong virus, identified from a *Culex annulirostris* mosquito in northern Western Australia. However, as discussed by the authors, the mosquito from which Jogalong virus was identified is hypothesised to have recently fed upon a tawny frogmouth (*Podargus strigoides*) [20]. Similarly, a distinct clade of gecko hepaciviruses was identified, comprising the novel Gecko hepacivirus from an Australian gecko (*Gehyra lauta*)[45], and the two previously identified gecko hepaciviruses from China. The gecko clade itself clusters with the avian hepaciviruses, which in turn group with the pegiviruses. Importantly, however, there is relatively weak support for the nodes in question, and if this reptile-avian cluster in fact grouped with the hepaciviruses then the overall phylogeny would offer better support to virus-host co-divergence.

The novel Koala hepacivirus and Collins beach virus 1 (identified in ticks feeding on a long-nosed bandicoot) group closely together, along with the single marsupial-infecting hepacivirus identified previously, Brushtail possum hepacivirus [38]. Together, these viruses establish a distinct marsupial clade of hepaciviruses. Both Collins beach virus and Collins beach virus 1 also fall in this clade and we suggest that they represent variants of the same virus. Interestingly, two marsupial-associated gammaherpesviruses (double-strand DNA viruses) from Australian koalas (*Phascolarctos cinereus*) and wombats (*Vombatus ursinus*), similarly formed a distinct clade [65]. Hence, marsupial-specific clades may be present in other vertebrate-infecting viruses and undergo co-divergence with their hosts. Additionally, two primate hepaciviruses were identified from the SRA, including a novel variant of SfHV (Sifaka hepacivirus) closely related to the previously identified SfHV, both of which were isolated from Diademed sifakas. In contrast, Bushbaby hepacivirus isolated from a Senegal bushbaby, instead falls as a sister-group to a cluster of equine and canine hepaciviruses

that are in turn closely related to HCV. Despite the obvious clinical importance of HCV, its ultimate animal reservoir is unknown [22-25]. The placement of another primate (i.e. bushbaby) virus in this clade is notable as it suggests that there are additional primate viruses within this group that are yet to be discovered and that might shed light on the origin of HCV.

Primate pegiviruses fall into distinct clades – in Old World versus New World primates - and the two novel sequences identified here follow this pattern. Several pegiviruses have been detected in a variety of Old World monkeys, including red colobus monkeys (*Procolobus ephroscles*), red-tailed guenons *(Cercopithecus ascanius)* and an olive baboon (*Papio Anubis)* [62, 63]*,* as well as the novel Simian pegivirus genome identified here in a South African vervet monkey. Similarly, the New World pegiviruses form a distinct clade that includes Pegivirus A (also termed GB virus A), isolated from tamarins [64, 66, 67], and newly identified Marmoset pegivirus, isolated from a Common marmoset.

Arthropod vectors are known to transmit a wide variety of infectious disease agents, and are responsible for more than 17% of total infectious diseases and cause upwards of 700,000 deaths per year [68]. Notably, the *Flaviviridae* contain many arthropod-borne viruses that impose a serious burden on human populations, including yellow fever virus, Zika virus, and dengue virus. However, all the vector-borne flaviviruses described to date are members of the genus *Flavivirus*, with no evidence of vector-borne transmission within the hepaciviruses or pegiviruses. Although the data presented here and previously [19] tentatively suggest that *I. holocyclus* ticks may act as a vector of Collins beach virus, the biology of ticks prevents us from clearly establishing this link [69]. In particular, as the ticks sampled for RNA extraction were engorged adult females it is impossible to determine if Collins beach virus infected the tick itself or was merely held within their considerable blood meal. As ticks take only one blood meal in each lifecycle stage (of which adulthood is the final stage) they do not act as 'biological syringes' as other arthropods such as mosquitos and biting flies do [70]. As Collins beach virus and Collins beach virus 1 were detected in very low abundance it does not seem likely that this virus is actively infecting the tick itself. Future meta-transcriptomic studies on unfed questing ticks are therefore a priority.

In sum, we have expanded the diversity of known hepaciviruses and hypothesise that a wide range of hosts that are yet to be identified globally. Unsurprisingly Australia's unique fauna host an equally diverse virome. The meta-transcriptomic tools described here can

517 help us to explore the breadth and depth of viral assemblages, investigate the role of
518 putative pathogens in wildlife disease syndromes, and contribute to the rapid and accurate
519 diagnosis of emergent disease syndromes.

520 **Supplementary Materials**

521

522 **Supplementary Figure 1**. Individual fragments used in compiling the Blue tit hepacivirus
523 and Simian pegivirus genomes. Each fragment has a coding region with draft annotations
524 of key viral proteins. (A) The three fragments from SRA project SRX565268 (SRR1325073,
525 SRR1325072, SRR1325074) from the Eurasian blue tit *(Cyanistes caeruleus)* from Germany.
526 (B) The two fragments (SRR1046735, SRR1046733) from SRA project SRX389659, taken
527 from a Vervet monkey (*Chlorocebus pygerythrus*) sampled in South Africa. These two
528 fragments were concatenated into a single genome, Simian pegivirus (SPV-saf).

529

530 **Supplementary Table 1.** Results of the SRA mining of marsupial transcriptomes.

531

532 **Supplementary Table 2.** Abundance of hepacivirus reads in the total read count of the
533 sequencing libraries used in this analysis.

534

535 **Supplementary Table 3.** Results of the SRA mining of avian transcriptomes (excluding
536 *Gallus gallus*).

537

538 **Supplementary Table 4.** Results of the SRA mining of primate transcriptomes (excluding
539 *Homo sapiens*).

540

541 **Supplementary Table 5.** GenBank accession numbers and description of the hepacivirus
542 and pegivirus amino acid sequences used in the phylogenetic analysis.

543

544 **Author Contributions:** Conceptualization, E.C.H; Methodology, E.C.H, J.H.O.P., J-S.E,
545 J.P.B, C.M., K.R, A.F.P; Formal Analysis, A.F.P, J.H.O.P., J-S.E; Writing – Original Draft
546 Preparation, A.F.P, E.C.H; Writing – Review & Editing, A.F.P, E.C.H, E.H.,J.H.O.P., W-S.C,
547 J-S.E; Funding Acquisition, E.C.H.

548

549 **Funding:** This work was supported by an ARC Australian Laureate Fellowship
550 (FL170100022) to E.C.H. A.F.P. is supported by the Australian Research Training Program
551 and University of Sydney Top-up and Merit grants. J.H.O.P is funded by the Swedish
552 research council FORMAS (grant nr: 2015-710).

553

554 **Data Availability:** The sequences have been submitted to GenBank with accession
555 numbers MT371434- MT371443.
556
557 **Acknowledgments:** Analysis was performed using the University of Sydney Artemis HPC.
558 We thank the Australian Museum for their generous provision of sample material used in
559 this study.
560
561 **Conflicts of Interest:** The authors declare no conflict of interest.

562 **References**

563 1. Jones, K.E.; Patel, N.G.; Levy, M.A.; Storeygard, A.; Balk, D.; Gittleman, J.L.; Daszak,
564 P. Global trends in emerging infectious diseases. *Nature* 2008, 451, 990-993.
565 2. Pfaender, S.; Brown, R.J.P.; Pietschmann, T.; Steinmann, E. Natural reservoirs for
566 homologs of hepatitis C virus. *Emerg. Microbes Infect*. 2014, 3, 1-9.
567 3. Pfaender, S.; Cavalleri, J.M.V.; Walter, S.; Doerrbecker, J.; Campana, B.; Brown, R.J.P.;
568 Burbelo, P.D.; Postel, A.; Hahn, K.; Anggakusuma, et al. Clinical course of infection and
569 viral tissue tropism of hepatitis C virus-like nonprimate hepaciviruses in Horses.
570 *Hepatol.* 2015, 61, 447-459.
571 4. Lauck, M.; Sibley, S.D.; Lara, J.; Purdy, M.A.; Khudyakov, Y.; Hyeroba, D.; Tumukunde,
572 A.; Weny, G.; Switzer, W.M.; Chapman, C.A., et al. A novel hepacivirus with an
573 unusually long and intrinsically disordered NS5A protein in a wild Old World primate. *J.*
574 *Virol*. 2013, 87, 8971-8981.
575 5. Canuti, M.; Williams, C.V.; Sagan, S.M.; Oude Munnink, B.B.; Gadi, S.; Verhoeven,
576 J.T.P.; Kellam, P.; Cotten, M.; Lang, A.S.; Junge, R.E., et al. Virus discovery reveals
577 frequent infection by diverse novel members of the *Flaviviridae* in wild lemurs. *Arch.*
578 *Virol*. 2019, 164, 509-522.
579 6. Firth, C.; Bhat, M.; Firth, M.A.; Williams, S.H.; Frye, M.J.; Simmonds, P.; Conte, J.M.;
580 Ng, J.; Garcia, J.; Bhuva, N.P., et al. Detection of zoonotic pathogens and
581 characterization of novel viruses carried by commensal *Rattus norvegicus* in New York
582 City. *mBio* 2014, 5, e01933-01914.
583 7. Kapoor, A.; Simmonds, P.; Scheel, T.K.; Hjelle, B.; Cullen, J.M.; Burbelo, P.D.;
584 Chauhan, L.V.; Duraisamy, R.; Sanchez Leon, M.; Jain, K., et al. Identification of rodent
585 homologs of hepatitis C virus and pegiviruses. *mBio* 2013, 4, e00216-00213.
586 8. Drexler, J.F.; Corman, V.M.; Muller, M.A.; Lukashev, A.N.; Gmyl, A.; Coutard, B.; Adam,
587 A.; Ritz, D.; Leijten, L.M.; van Riel, D., et al. Evidence for novel hepaciviruses in rodents.
588 *PLoS Pathog*. 2013, 9, e1003438.
589 9. Wu, Z.; Lu, L.; Du, J.; Yang, L.; Ren, X.; Liu, B.; Jiang, J.; Yang, J.; Dong, J.; Sun, L., et
590 al. Comparative analysis of rodent and small mammal viromes to better understand the
591 wildlife origin of emerging infectious diseases. *Microbiome* 2018, 6, 178.
592 10. de Souza, W.M.; Fumagalli, M.J.; Sabino-Santos, G., Jr.; Motta Maia, F.G.; Modha, S.;
593 Teixeira Nunes, M.R.; Murcia, P.R.; Moraes Figueiredo, L.T. A novel hepacivirus in wild
594 rodents from South America. *Viruses* 2019, 11, 297.

11. Burbelo, P.D.; Dubovi, E.J.; Simmonds, P.; Medina, J.L.; Henriquez, J.A.; Mishra, N.; Wagner, J.; Tokarz, R.; Cullen, J.M.; Iadarola, M.J., et al. Serology-enabled discovery of genetically diverse hepaciviruses in a new host. *J. Virol*. 2012, 86, 6171-6178.
12. Quan, P.L.; Firth, C.; Conte, J.M.; Williams, S.H.; Zambrana-Torrelio, C.M.; Anthony, S.J.; Ellison, J.A.; Gilbert, A.T.; Kuzmin, I.V.; Niezgoda, M., et al. Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proc. Natl. Acad. Sci. USA* 2013, 110, 8194-8199.
13. Baechlein, C.; Fischer, N.; Grundhoff, A.; Alawi, M.; Indenbirken, D.; Postel, A.; Baron, A.L.; Offinger, J.; Becker, K.; Beineke, A., et al. Identification of a novel hepacivirus in domestic cattle from Germany. *J. Virol*. 2015, 89, 7007-7015.
14. Corman, V.M.; Grundhoff, A.; Baechlein, C.; Fischer, N.; Gmyl, A.; Wollny, R.; Dei, D.; Ritz, D.; Binger, T.; Adankwah, E., et al. Highly divergent hepaciviruses from African cattle. *J. Virol*. 2015, 89, 5876-5882.
15. Chu, L.; Jin, M.; Feng, C.; Wang, X.; Zhang, D. A highly divergent hepacivirus-like flavivirus in domestic ducks. *J. Gen. Virol*. 2019, 100, 1234-1240.
16. Goldberg, T.L.; Sibley, S.D.; Pinkerton, M.E.; Dunn, C.D.; Long, L.J.; White, L.C.; Strom, S.M. Multidecade mortality and a homolog of hepatitis C virus in Bald Eagles (*Haliaeetus leucocephalus*), the national bird of the USA. *Sci. Rep*. 2019, 9, 14953-14953.
17. Shi, M.; Lin, X.D.; Chen, X.; Tian, J.H.; Chen, L.J.; Li, K.; Wang, W.; Eden, J.S.; Shen, J.J.; Liu, L., et al. The evolutionary history of vertebrate RNA viruses. *Nature* 2018, 556, 197-202.
18. Shi, M.; Lin, X.D.; Vasilakis, N.; Tian, J.H.; Li, C.X.; Chen, L.J.; Eastwood, G.; Diao, X.N.; Chen, M.H.; Chen, X., et al. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *J. Virol*. 2016, 90, 659-669.
19. Harvey, E.; Rose, K.; Eden, J.S.; Lo, N.; Abeyasuriya, T.; Shi, M.; Doggett, S.L.; Holmes, E.C. Extensive diversity of RNA viruses in Australian ticks. *J. Virol*. 2019, 93, e01358-18.
20. Williams, S.H.; Levy, A.; Yates, R.A.; Somaweera, N.; Neville, P.J.; Nicholson, J.; Lindsay, M.D.A.; Mackenzie, J.S.; Jain, K.; Imrie, A., et al. Discovery of Jogalong virus, a novel hepacivirus identified in a *Culex annulirostris* (Skuse) mosquito from the Kimberley region of Western Australia. *PLoS One* 2020, 15, e0227114.
21. Theze, J.; Lowes, S.; Parker, J.; Pybus, O.G. Evolutionary and phylogenetic analysis of the hepaciviruses and pegiviruses. *Genome Biol. Evol*. 2015, 7, 2996-3008.

22. Lyons, S.; Kapoor, A.; Schneider, B.S.; Wolfe, N.D.; Culshaw, G.; Corcoran, B.; Durham, A.E.; Burden, F.; McGorum, B.C.; Simmonds, P. Viraemic frequencies and seroprevalence of non-primate hepacivirus and equine pegiviruses in horses and other mammalian species. *J. Gen. Virol*. 2014, 95, 1701-1711.
23. Hartlage, A.S.; Cullen, J.M.; Kapoor, A. The strange, expanding world of animal hepaciviruses. *Annu. Rev. Virol*. 2016, 3, 53-75.
24. Pybus, O.G.; Gray, R.R. The virus whose family expanded. *Nature* 2013, 498, 310-311.
25. Chang, W.-S.; Eden, J.-S.; Hall, J.; Shi, M.; Rose, K.; Holmes, E.C. Meta-transcriptomic analysis of virus diversity in urban wild birds with paretic disease. *bioRxiv* 2020, https://www.biorxiv.org/content/10.1101/2020.03.07.982207v1.full.
26. Stapleton, J.T.; Foung, S.; Muerhoff, A.S.; Bukh, J.; Simmonds, P. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family *Flaviviridae*. *J. Gen. Virol*. 2011, 92, 233-246.
27. Tuddenham, R.; Eden, J.-S.; Gilbey, T.; Dwyer, D.E.; Jennings, Z.; Holmes, E.C.; Branley, J.M. Human pegivirus in brain tissue of a patient with encephalitis. *Diag .Micro. Infect. Dis*. 2020, 96, 114898.
28. Bukh, J.; Apgar, C.L. Five new or recently discovered (GBV-A) virus species are indigenous to New World monkeys and may constitute a separate genus of the *Flaviviridae*. *Virology* 1997, 229, 429-436.
29. Schaluder, G.G.; Dawson, G.J.; Simons, J.N.; Pilot-Matias, T.J.; Gutierrez, R.A.; Heynen, C.A.; Knigge, M.F.; Kurpiewski, G.S.; Buijk, S.L.; Leary, T.P., et al. Molecular and serologic analysis in the transmission of the GB hepatitis agents. *J. Med. Virol*. 1995, 46, 81-9.
30. Adams, N.J.; Prescott, L.E.; Jarvis, L.M.; Lewis, J.C.; McClure, M.O.; Smith, D.B.; Simmonds, P. Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. *J. Gen. Virol*. 1998, 79, 1871-1877.
31. Birkenmeyer, L.G.; Desai, S.M.; Muerhoff, A.S.; Leary, T.P.; Simons, J.N.; Montes, C.C.; Mushahwar, I.K. Isolation of a GB virus-related genome from a chimpanzee. *J. Med. Virol*. 1998, 56, 44-51.
32. Simons, J.N.; Leary, T.P.; Dawson, G.J.; Pilot-Matias, T.J.; Muerhoff, A.S.; Schlauder, G.G.; Desai, S.M.; Mushahwar, I.K. Isolation of novel virus-like sequences associated with human hepatitis. *Nat. Med*. 1995, 1, 564-569.
33. Sibley, S.D.; Lauck, M.; Bailey, A.L.; Hyeroba, D.; Tumukunde, A.; Weny, G.; Chapman, C.A.; O'Connor, D.H.; Goldberg, T.L.; Friedrich, T.C. Discovery and characterization of

distinct simian pegiviruses in three wild African Old World monkey species. *PLoS One* 2014, 9, e98569-e98569.

34. Chandriani, S.; Skewes-Cox, P.; Zhong, W.; Ganem, D.E.; Divers, T.J.; Van Blaricum, A.J.; Tennant, B.C.; Kistler, A.L. Identification of a previously undescribed divergent virus from the *Flaviviridae* family in an outbreak of equine serum hepatitis. *Proc. Natl. Acad. Sci. USA* 2013, 110, E1407-1415.
35. Epstein, J.H.; Quan, P.L.; Briese, T.; Street, C.; Jabado, O.; Conlan, S.; Ali Khan, S.; Verdugo, D.; Hossain, M.J.; Hutchison, S.K., et al. Identification of GBV-D, a novel GB-like flavivirus from old world frugivorous bats (*Pteropus giganteus*) in Bangladesh. *PLoS Pathog*. 2010, 6, e1000972.
36. Stapleton, J.T.; Foung, S.; Muerhoff, A.S.; Bukh, J.; Simmonds, P. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus *Pegivirus* within the family *Flaviviridae*. *J. Gen. Virol*. 2011, 92, 233-246.
37. Pybus, O.G.; Thézé, J. Hepacivirus cross-species transmission and the origins of the hepatitis C virus. *Curr. Opin. Virol*. 2016, 16, 1-7.
38. Chang, W.-S.; Eden, J.-S.; Hartley, W.J.; Shi, M.; Rose, K.; Holmes, E.C. Metagenomic discovery and co-infection of diverse wobbly possum disease viruses and a novel hepacivirus in Australian brushtail possums. *One Health Outlook* 2019, 1, 5.
39. Harvey, E.; Rose, K.; Eden, J.S.; Lawrence, A.; Doggett, S.L.; Holmes, E.C. Identification of diverse arthropod associated viruses in native Australian fleas. *Virology* 2019, 535, 189-199.
40. Mahar, J.E.; Read, A.J.; Gu, X.; Urakova, N.; Mourant, R.; Piper, M.; Haboury, S.; Holmes, E.C.; Strive, T.; Hall, R.N. Detection and circulation of a novel rabbit hemorrhagic disease virus in Australia. *Emerg. Infect. Dis*. 2018, 24, 22-31.
41. Harvey, E.; Madden, D.; Polkinghorne, A.; Holmes, E.C. Identification of a novel picorna-like virus, Burpengary virus, that is negatively associated with chlamydial disease in the Koala. *Viruses* 2019, 11, 211.
42. Wille, M.; Eden, J.-S.; Shi, M.; Klaassen, M.; Hurt, A.C.; Holmes, E.C. Virus–virus interactions and host ecology are associated with RNA virome structure in wild birds. *Mol. Ecol*. 2018, 27, 5263-5278.
43. Shi, M.; Neville, P.; Nicholson, J.; Eden, J.-S.; Imrie, A.; Holmes, E.C. High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J. Virol*. 2017, 91, e00680-00617.

44. Cantalupo, P.G.; Buck, C.B.; Pipas, J.M. Complete genome sequence of a polyomavirus recovered from a pomona leaf-nosed bat (*Hipposideros pomona*) metagenome data set. *Genome Announc*. 2017, 5, e01053-01016.
45. Oliver, P.M.; Prasetya, A.M.; Tedeschi, L.G.; Fenker, J.; Ellis, R.J.; Doughty, P.; Moritz, C. Crypsis and convergence: integrative taxonomic revision of the *Gehyra australis* group (Squamata: Gekkonidae) from northern Australia. *Peer J*. 2020, 8, e7971.
46. Johnson, R.N.; O'Meally, D.; Chen, Z.; Etherington, G.J.; Ho, S.Y.W.; Nash, W.J.; Grueber, C.E.; Cheng, Y.; Whittington, C.M.; Dennison, S., et al. Adaptation and conservation insights from the koala genome. *Nat. Genet*. 2018, 50, 1102-1111.
47. Shi, M.; Lin, X.D.; Tian, J.H.; Chen, L.J.; Chen, X.; Li, C.X.; Qin, X.C.; Li, J.; Cao, J.P.; Eden, J.S., et al. Redefining the invertebrate RNA virosphere. Nature 2016, 540, 539-543.
48. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30, 2114-2120.
49. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q., et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech*. 2011, 29, 644-652.
50. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* 2009, 10, 421.
51. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Meth*. 2015, 12, 59-60.
52. Li, D.; Luo, R.; Liu, C.M.; Leung, C.M.; Ting, H.F.; Sadakane, K.; Yamashita, H.; Lam, T.W. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 2016, 102, 3-11.
53. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C., et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, 28, 1647-1649.
54. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*. 2013, 30, 772-780.
55. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol*. 2007, 56, 564-577.

56. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 2011, 27, 1164-1165.
57. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol*. 2015, 32, 268-274.
58. Zein, N.N. Clinical significance of hepatitis C virus genotypes. *Clin. Microbiol. Rev*. 2000, 13, 223-235.
59. Moradpour, D.; Penin, F.; Rice, C.M. Replication of hepatitis C virus. *Nat. Rev. Micro*. 2007, 5, 453-463.
60. Major, M.E.; Feinstone, S.M. The molecular virology of hepatitis C. *Hepatol*. 1997, 25, 1527-1538.
61. Canuti, M.; Williams, C.V.; Sagan, S.M.; Oude Munnink, B.B.; Gadi, S.; Verhoeven, J.T.P.; Kellam, P.; Cotten, M.; Lang, A.S.; Junge, R.E., et al. Virus discovery reveals frequent infection by diverse novel members of the *Flaviviridae* in wild lemurs. *Arch. Virol*. 2019, 164, 509-522.
62. Bailey, A.L.; Lauck, M.; Mohns, M.; Peterson, E.J.; Beheler, K.; Brunner, K.G.; Crosno, K.; Mejia, A.; Mutschler, J.; Gehrke, M., et al. Durable sequence stability and bone marrow tropism in a macaque model of human pegivirus infection. *Sci. Transl. Med*. 2015, 7, 305ra144.
63. Sibley, S.D.; Lauck, M.; Bailey, A.L.; Hyeroba, D.; Tumukunde, A.; Weny, G.; Chapman, C.A.; O'Connor, D.H.; Goldberg, T.L.; Friedrich, T.C. Discovery and characterization of distinct simian pegiviruses in three wild African Old World monkey species. *PLoS One* 2014, 9, e985699.
64. Leary, T.P.; Desai, S.M.; Yamaguchi, J.; Chalmers, M.L.; Schlauder, G.G.; Dawson, G.J.; Mushahwar, I.K. Species-specific variants of GB virus A in captive monkeys. *J. Virol*. 1996, 70, 9028-9030.
65. Vaz, P.K.; Hartley, C.A.; Lee, S.-Y.; Sansom, F.M.; Adams, T.E.; Stalder, K.; Pearce, L.; Lovrecz, G.; Browning, G.F.; Müller, C.E., et al. Koala and wombat gammaherpesviruses encode the first known viral NTPDase homologs and are phylogenetically divergent from all known gammaherpesviruses. *J. Virol*. 2019, 93, e01404-01418.
66. Leary, T.P.; Desai, S.M.; Erker, J.C.; Mushahwar, I.K. The sequence and genomic organization of a GB virus A variant isolated from captive tamarins. *J. Gen. Virol*. 1997, 78, 2307-2313.

766 67. Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
767 Genomic analysis of two GB virus A variants isolated from captive monkeys. *J. Gen.*
768 *Virol*. 1998, 79, 41-45.
769 68. WHO, W.H.O. Global Hepatitis Report 2017. Geneva, 2017.
770 69. Sojka, D. Book review: Sonenshine D.E. & Roe R.M. 2013: *Biology of Ticks*. 2nd ed.
771 *European Journal of Entomology* 2015, 11.
772 70. Nuttall, P.A.; Jones, L.D.; Labuda, M.; Kaufman, W.R. Adaptations of arboviruses to
773 ticks. *J. Med. Entomol*. 1994, 31, 1-9.