The diverse viromes of Australian lizards are shaped by host taxonomy and habitat

Jackie E. Mahar[1]*, Michelle Wille[2], Erin Harvey[1], Craig C. Moritz[3], Edward C. Holmes[1]

[1]Sydney Institute for Infectious Diseases, School of Medical Sciences, The University of Sydney, Sydney, New South Wales 2006, Australia

[2]Centre for Pathogen Genomics, Department of Microbiology and Immunology, at the Peter Doherty Institute for Infection and Immunity. The University of Melbourne, Melbourne, Victoria 3000, Australia.

[3]Research School of Biology & Centre for Biodiversity Analysis, The Australian National University, Canberra, ACT 2600, Australia.

*Current address: Australian Centre for Disease Control, Australian Animal Health Laboratory and Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Geelong, VIC 3219, Australia

#Correspondence to:

Edward C. Holmes

edward.holmes@sydney.edu.au

Tel. +61 2 9351 5591

Text word count: 6443

Abstract

Lizards inhabit diverse ecologies and evolutionary histories and hence represent a promising group to explore how hosts shape virome structure and virus evolution. Yet little is known about the viromes of these animals. In Australia, squamates (lizards and snakes) comprise the most diverse order of vertebrates, and Australia hosts the highest diversity of lizards globally, with the greatest breadth of habitat use. We used meta-transcriptomic sequencing to determine the virome of nine co-distributed, tropical lizard species from three taxonomic families in Australia and analyzed these data to identify host traits associated with viral abundance and diversity. We show that lizards carry a large diversity of viruses, identifying more than 30 novel, highly divergent vertebrate-associated viruses. These viruses were from nine viral families, including several that contain well known pathogens, such as the *Flaviviridae, Picornaviridae, Bornaviridae, Iridoviridae* and *Rhabdoviridae*. Members of the *Flaviviridae* were particularly abundant across species sampled here, largely belonging to the genus *Hepacivirus*: 14 novel *Hepaciviruses* were identified, broadening the known diversity of this group and better defining its evolution by uncovering new reptilian clades. The evolutionary histories of the viruses studied here frequently aligned with the biogeographic and phylogenetic histories of the hosts, indicating that exogenous viruses may help infer host evolutionary history if sampling is strategic and sampling density high enough. Notably, analysis of alpha and beta diversity revealed that virome composition and richness was shaped by host taxonomy, habitat and range size. In sum, we identified a diverse range of reptile viruses that broadly contributes to our understanding of virus-host ecology and evolution.

Keywords: evolution, metagenomics, meta-transcriptomics, viral ecology, one health, *Hepacivirus*

1. Introduction

Animal virology has traditionally focused on viruses of mammals and birds as they are the most likely natural reservoir hosts for viruses that may emerge in humans and animals of economic importance (Mollentze and Streicker 2020; Zhang et al. 2018). While this has provided major insights, it has necessarily resulted in a skewed view of viral diversity that limits our understanding of virus evolution and ecology, including the frequency and determinants of cross-species transmission and host range. Indeed, many viral families traditionally associated with mammals and birds have now been found in other vertebrate classes such as amphibians, reptiles and fish (Harding et al. 2022; Shi et al. 2018). Since far greater biological diversity exists within these other vertebrates – that comprise ~33,000 documented species – it is reasonable to assume that they also harbour a substantial diversity of viruses (Zhang et al. 2018).

A variety of factors make lizards particularly informative for the study of viral ecology and evolution. Squamates (lizards and snakes) are a highly diverse and successful group of vertebrates, comprising 96.3% of non-avian reptile diversity (Pincheira-Donoso et al. 2013), with over 10,000 extant species globally (Herrera-Flores et al. 2021). Squamates exhibit a range of morphologies, life history traits, and diverse ecologies, inhabiting every continent except Antarctica (Pincheira-Donoso et al. 2013; Pyron, Burbrink and Wiens 2013). Australia is home to the largest number of reptile species, comprising ~10% of the world's reptiles (Geyle et al. 2021), of which squamates are the most diverse vertebrate order (Wilson and Swan 2017). These animals are well adapted to the Australian landscape, colonising and radiating across a diverse range of environments and ecological niches (Pianka 1969; Pianka 1973; Morton and James 1988;). Australian lizards comprise old endemic Gondwanan lineages that pre-date the isolation of Australia and Antarctica, as well as more recent immigrant lineages from the North (Brennan and Oliver 2017). For example, the Gondwanan group of Pygopodoidea have a crown age of greater than 50 million years, while the genus *Gehyra* have a crown age in the mid-Miocene ~20 million years ago, having colonized from Asia (Oliver and Hugall 2017). Several squamate species are at high risk of extinction within the next twenty years (Geyle et al. 2021) and infectious disease emergence can severely threaten reptile populations (Zhang et al. 2018).

83   A key, yet rarely addressed, aspect of virus evolution and emergence is understanding the

84   traits that determine the diversity and abundance of viruses carried by a host (Wille et al.

85   2019). A number of studies have linked viral diversity and abundance to specific host traits,

86   including phylogenetic history, habitat, body mass, geographic range, community diversity,

87   biome, location, and infection status with particular pathogens (Olival et al. 2017; Wille et al.

88   2018; Geoghegan et al. 2021). However, the effects of host ecology on viral diversity have

89   only been explored in a handful of systems with varying results. To better understand the

90   diversity of viruses carried in lizards and how host traits affect viral abundance and diversity

91   we used meta-transcriptomic sequencing to explore the virome of nine lizard species residing

92   in various environments and habitats across the biologically diverse monsoonal tropics of

93   northern Australia.

94

2. Methods

95

2.1 Ethics statement

96

97   All work was carried out according to the Australian Code for the Care and Use of Animals

98   for Scientific Purposes with approval from the institutional animal ethics committee (Permit

99   ANU animal ethics approval A2016-07) and State authorities (collecting permits NT 58454

100   and WA SF010911).

101

2.2 Animal Sampling

102

103   Liver samples were collected from apparently healthy *Carlia amax*, *Carlia munda*, *Carlia*

104   *sexdentata*, *Cryptoblepharus metallicus*, *Heteronotia planiceps*, *Heteronotia binoei*, *Gehyra*

105   *nana*, *Gehyra arnhemica*, and *Oedura marmorata*. Lizards were collected in autumn and

106   winter months between April 2016 and June 2017, in arid regions of the eastern Kimberley

107   and mesic regions of the "Top End", Australia (Table 1, Table S1). Samples were collected in

108   Australian bioregions Arnhem Coast (ARC), Victoria Bonaparte (VIB), Central Arnhem

109   (CEA), Daly Basin (DAB), or Darwin Coastal (DAC) (Table S1), as defined by Interim

110   Biogeographic Regionalisation for Australia (IBRA), version 7. Excised liver was stored in

111   RNAlater™ Stabilization solution (Thermofisher Scientific, MA, USA), at room temperature

112   while in the field and then at 4ºC for longer term storage. Sampled animals displayed no

113   obvious signs of serious pathology.

114

2.3 RNA extraction

Liver tissue was homogenized using the Qiagen Tissue Lyser II with 3mm stainless steel beads in Qiagen buffer RLT (Hilden, Germany); and RNA extracted using the Qiagen RNeasy Plus minikit (Hilden, Germany) according to the manufacturer's protocol. Purified RNA was pooled in equimolar ratios into eleven pools, grouping RNA samples from the same lizard species and collection location, with six to twelve individuals per pool (Table 1, Table S1). Pooled RNA was further purified using the RNeasy MinElute clean-up kit (Qiagen, Hilden, Germany) and quantified using the Qubit RNA Broad-range Assay with the Qubit Fluorometer v3.0 (Thermofisher Scientific).

2.4 Meta-transcriptomic sequencing

RNA pools were assessed for quality using the Agilent 2100 Bioanalyzer with the Agilent RNA 6000 Nano Assay (Agilent Technologies, CA, USA). Library construction and sequencing was performed at the Australian Genomic Research Facility (Victoria, Australia). Libraries were constructed using the TruSeq Total RNA Library Preparation protocol (Illumina, CA, USA) following rRNA removal using the Illumina Ribo-zero Gold epidemiology kit. Paired-end (100 bp) sequencing of each RNA library was performed on a HiSeq 2500 sequencing platform (Illumina, CA, USA).

2.5 Genome/transcript assembly, annotation and abundance calculation

After trimming with Trimmomatic v0.38 (Bolger, Lohse and Usadel 2014), reads were assembled *de novo* using two separate assemblers – Trinity v2.5.1 (Grabherr et al. 2011) and Megahit v1.2.9 (Li et al. 2015) – to increase the chances of correctly assembling all viral contigs. Contigs from both methods were combined and duplicates removed (retaining the longest version) using CD-HIT-EST v4.8.1 (Fu et al. 2012). Viral contigs were identified using BLASTn (Altschul et al. 1990) and DIAMOND BLASTx (Buchfink, Reuter and Drost 2021) tools through alignment with the NCBI nucleotide (nt) database (*e*-value cut-off 1×10^{-10}) and non-redundant protein (nr) database (*e*-value cut-off 1×10^{-5}), respectively. The Geneious assembler (available in Geneious Prime ® 2022.2.2) was used to extend viral contigs where possible. Open reading frames were identified using the Find ORFs tool within Geneious Prime, and conserved domains were identified using RSP-TBLASTN v2.12.0+ (Altschul et al. 1997) against the NCBI Conserved Domains database. The abundance of each viral contig was calculated as expected counts (from mapped trimmed reads) using the RSEM

148   tool in Trinity v2.5.2 (Li et al. 2010). Overall abundance was calculated as expected

149   count/total number of trimmed reads in library x 100. Novel viruses that shared greater than

150   90% amino acid identity in the RdRp were considered to represent the same species. Likely

151   vertebrate-infecting viruses were defined as those that belong to classically vertebrate-

152   infecting viral families and/or those that clustered with viruses known to infect vertebrate

153   species in phylogenetic trees.

154

155   **2.6 Alpha and beta diversity analyses**

156   Diversity statistics were only obtained for viruses considered likely to infect vertebrate

157   species (i.e., those likely to infect the sampled host) and viruses likely to be exogenous.

158   *Retroviruses* were excluded due to the difficulty in determining whether they are endogenous

159   or exogenous, and viral groups with disrupted ORFs were considered likely to be

160   endogenous. We performed generalized linear models and selected the best-fit model at both

161   the infraorder and family taxonomic level (among a set of possible models describing the

162   relationship between taxonomy, habitat, environment, range size and the number of

163   individuals in the library) based on the lowest AIC, as done previously (Geoghegan et al.

164   2021; Chen et al. 2022). Host genus and species level were not considered due to small

165   sample sizes. The Csex_M library was excluded from all analyses because it contained no

166   exogenous, biologically relevant viruses and as the only riparian library, may have skewed

167   results. All other libraries were included for comparisons considering host taxonomy at the

168   level of infraorder. Omar_M was similarly removed for comparisons at the host family level

169   as it was the only library representing Diplodactylidae. Alpha diversity, including richness,

170   Shannon index, Simpson Index, Shannon effective and Simpson effective were calculated for

171   each library (Lagkouvardos et al. 2017). Beta diversity was calculated using the Bray–Curtis

172   dissimilarity matrix and virome structure was plotted as a function of nonmetric

173   multidimensional scaling (NMDS) ordination and tested using Adonis tests (PERMANOVA)

174   using the *vegan* and *phyloseq* packages (McMurdie and Holmes 2013; Oksanen et al. 2022).

175   Analyses were performed in R version 4.2.3 in R Studio 2022.07.1.

176

177   **2.7 Phylogenetic analysis**

178   Viral amino acid sequences were aligned with representative sequences from the same viral

179   family obtained from NCBI, using Clustal Omega v1.2.3 available in Geneious Prime. Where

180   necessary, large data sets were condensed to a more manageable size using CD-HIT version

181   4.8.1 (Fu et al. 2012). TrimAl v1.4.1 (Capella-Gutiérrez, Silla-Martínez and Gabaldón 2009)

182   was used to remove ambiguously aligned regions (using the gappyout setting in all cases

183   except for the *Flaviviridae* alignment which required stricter manually applied settings to

184   remove uninformative columns). Alignments were visualized in Geneious Prime. Maximum

185   likelihood trees were inferred using IQTree v2.1.3 (Nguyen et al. 2015) with model selection

186   estimated using ModelFinder within IQTree (Kalyaanamoorthy et al. 2017). Branch supports

187   were estimated with the Shimodaira-Hasegawa (SH)-like approximate likelihood ratio test

188   (Guindon et al. 2010). The size and length of each alignment is provided in Table S2 and

189   details of assembled viral nucleotide consensus sequences that were translated for inclusion

190   in phylogenies are provided in Table S3.

191

192   To reduce the reporting of false positives due to index-hopping during sequencing for each

193   viral contig present in more than one library, a viral contig was presumed to be a contaminant

194   from another library if it met the following criteria: contig abundance was less than 0.1% of

195   the abundance of that contig in the library where that contig was most abundant. This is based

196   on the index-hopping rate of about 0.1-2% as listed by Illumina

197   (https://sapac.illumina.com/techniques/sequencing/ngs-library-prep/multiplexing/index-

198   hopping.html).

199

200   **2.8 Virus nomenclature**

201   Novel viral species were determined based on demarcation criteria assigned by the ICTV

202   (International Committee on Taxonomy of Viruses) for the relevant families/genera

203   (https://ictv.global/report/genome) and were randomly assigned provisional names. Note that

204   because these virus names are provisional, we have only assigned common names rather than

205   a full binomial nomenclature.

206

207   # 3. Results

208   ## 3.1 Viruses in Australian lizard species

209   We sequenced 11 pooled RNA libraries with an average read count of 35,041,464 (Table 1).

210   Each library was generated from liver samples of 6 to 12 apparently healthy lizards of a

211   single species sampled from either the mesic Top End or from the relatively arid east

212   Kimberley regions of Australia (Victoria-Bonaparte region; Table 1). Two widespread gecko

213   species – *Heteronotia binoei* and *Gehyra nana* – were sampled from both regions. A high

214 diversity of vertebrate-associated viruses was detected, comprising nine different families and

215 12 genera (Figure 1). While for most viral families only a single genus and species were

216 documented, we identified viruses from at least two genera in each of the *Flaviviridae*,

217 *Picornaviridae*, *Rhabdoviridae* and *Amnoonviridae,* and two species of the same genera in

218 the *Astroviridae* (Figure 1B). Despite being present in only one library, the *Arenaviridae*

219 were by far the most abundant viral family (Figure 1), with the *Flaviviridae* the second most

220 abundant and detected in all three lizard families (5/11 libraries) (Figure 1A). Members of the

221 *Picornaviridae* were also abundant and detected in five of 11 libraries, although were

222 restricted to the Gekkota species (Gekkonidae and Diplodactylidae). No vertebrate-infecting

223 viruses were shared between different lizard species (Figure 1B). At higher taxonomic scales,

224 the genera *Hepacivirus* (*Flaviviridae*), *Shanbavirus* (*Picornaviridae),* and an unclassified

225 genus of *Astroviridae*, were found in multiple lizard species (Figure 1). Five viral families

226 were found across multiple host species: the *Flaviviridae*, *Picornaviridae*, *Astroviridae*,

227 *Rhabdoviridae* and *Amnoonviridae*.

228 **Table 1. Details of lizard livers sampled and sequencing output for each library (see Table S1 for additional details).**

Library details		Taxonomy of sampled species			Sampling details and species characteristics			Output
Library Name	No. of samples[a]	Infra-order	Family	Genus and species	Sampled region[b]	Typical habitat	Range size[c]	No. of paired reads
Cam_M	8	Scinciformata	Scincidae	*Carlia amax*	Top End (mesic)	rocks	small	27,151,104
Cmun_M	6	Scinciformata	Scincidae	*Carlia munda*	Top End (mesic)	general	medium	26,737,353
CSex_M	8	Scinciformata	Scincidae	*Carlia sexdentata*	Top End (mesic)	riparian	very small	40,670,859
Cmet_M	9	Scinciformata	Scincidae	*Cryptoblepharus metallicus*	Top End (mesic)	trees	very small	29,719,339
Hplan_A	6	Gekkota	Gekkonidae	*Heteronotia planiceps*	Kimberley (arid)	rocks	very small	36,902,535
Hbin_A	8	Gekkota	Gekkonidae	*Heteronotia binoei*	Kimberley (arid)	general	very large	40,480,482
Hbin_M	12	Gekkota	Gekkonidae	*Heteronotia binoei*	Top End (mesic)	general	very large	52,580,948
Gnan_A	7	Gekkota	Gekkonidae	*Gehyra nana*	Kimberley (arid)	rocks	small	26,343,593
Gnan_M	9	Gekkota	Gekkonidae	*Gehyra nana*	Top End (mesic)	rocks	small	32,765,848
Garn_M	9	Gekkota	Gekkonidae	*Gehyra arnhemica*	Top End (mesic)	trees	small	40,865,843

9

Omar_M	10	Gekkota	Diplodactylidae	*Oedura marmorata*	Top End (mesic)	rocks	small	31,238,197

ᵃNo. of samples: Number of individuals pooled in library.

ᵇSampled region: Top End: Top End of Australia (Northern Territory), mesic environment; Kimberley: Kimberley region (Western Australia and Northern Territory), arid environment.

ᶜRange size: Broadly summarised based on range listed on https://arod.com.au.

Figure 1. Biologically relevant viruses (i.e., viruses that infect vertebrates) in the sampled lizard species. (A) Relative abundance (left plot) and overall abundance (right) of the vertebrate-infecting virus families present in each library. Libraries are plotted in taxonomic sequence, with host relationships indicated by a cladogram and host infraorder indicated in grey bars. Library names are as follows: Cam_M, *Carlia amax (collected from a mesic environment)*; Cmun_M, *Carlia munda* (mesic), Csex_M, *Carlia sexdentata* (mesic);

Cmet_M, *Cryptoblepharus metallicus* (mesic); Gnan_A, *Gehyra nana* (arid); Gnan_M *Gehyra nana* (mesic); Garn_M, *Gehyra arnhemica* (mesic); Hbin_A, *Heteronotia binoei* (arid); Hbin_M, *Heteronotia binoei* (mesic); Hplan_A, *Heteronotia planiceps* (arid); Omar_M, *Oedura marmorata* (mesic). (B) Viruses found in each library are represented by circles coloured by viral genus (UC=unclassified genus), with lines connecting them to the libraries in which they were found. Libraries are grouped by host family (indicated by grey bars above; Diplod.=Diplodactylidae). Numbers in parentheses beside virus names indicates the number of variants of that virus detected (where >1). Note that all the *Amnoonviridae* belong to currently unclassified genera, but may represent more than one genus.
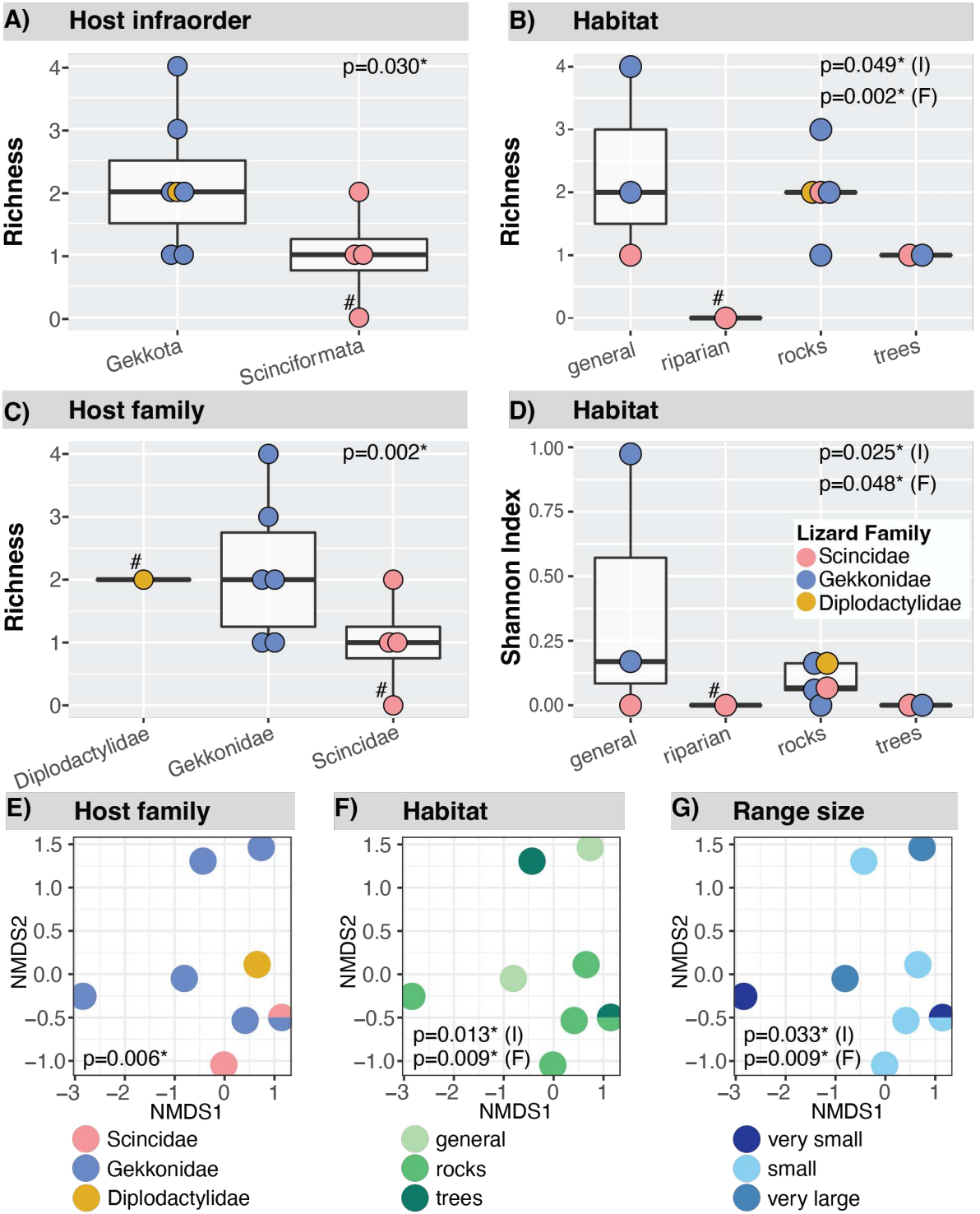
A number of other vertebrate-associated viral families were detected, but were determined as likely endogenous virus elements (EVEs) because longer contigs had disrupted open reading frames. These included members of the *Adintoviridae*, *Hepadnaviridae*, *Retroviridae*, and *Circoviridae,* and were excluded from abundance and alpha diversity analyses. Of note, library CSex_M (comprising *Carlia sexdentata*) did not contain any biologically relevant viruses, aside from some likely EVEs, but did contain some viruses that are not associated with vertebrates (Figure S1). Across all libraries, we detected several viral groups that were unlikely to be associated with vertebrate infection, representing the *Permutotetraviridae*, *Chuviridae*, *Tectiviridae*, *Mimiviridae*, *Autolykiviridae*, *Baculoviridae*, unclassified Ortervirales, unclassified Riboviria, and tombus-like, solemo-like, narna-like, partiti-like, and toti-like viruses. None of these viruses were abundant (Figure S1) and were considered likely to be viruses of commensal organisms in the lizard livers (i.e., narna-like) or contaminants. Only vertebrate-associated viral families are included in the analyses described below.

3.2 Abundance and diversity of lizard viruses

To help determine the factors influencing virome diversity, we compared the abundance and alpha diversity in biologically relevant viruses between libraries from different lizard species and sampling regions. Specifically, we considered host taxonomy (at the level of infraorder and family) as well as a range of ecological variables, including habitat, environment and range size (Csex_M was excluded from these analyses as it was the only riparian library and no biologically relevant viruses were found in this library). Given variability in the number of individuals per pool, we first assessed the effect of the number of samples per pool/library to account for any bias introduced by sampling strategy. The number of samples per pool correlated with the Richness ($p=0.037$) and Shannon diversity ($p=0.015$), but not the abundance ($p= 0.140$) of vertebrate viruses, or the Shannon Effective ($p=0.052$), Simpson ($p=0.434$) or Simpson Effective diversity ($p=0.077$) (Figure S2). As such, the number of

278   samples per pool was considered in all statistical models, and included as cofactors in final

279   models addressing Richness and Shannon diversity.

280

281   Host taxonomy and habitat were consistently important parameters in the best-fit models for

282   all diversity measures regardless of the level of taxonomy considered, indicating that these

283   factors are important modulators of viral diversity in this sample. Taxonomy had a

284   statistically significant effect on viral richness regardless of taxonomic level (host infraorder

285   p=0.030, family p=0.002; Figure 2A and 2C), but was not significant in modulating

286   abundance or Shannon and Simpson indexes (Figure S3). Habitat significantly modulated

287   richness and Shannon index diversity when considering host infraorder (p=0.049 and

288   p=0.025, respectively) and host family (p=0.002 and p=0.048; Figure 2B and 2D), and also

289   had a significant effect on the Simpson index of diversity, but only when considering host

290   infraorder (p=0.037; Figure S4). Range size was significant in the best-fit model for richness,

291   regardless of the taxonomy level explored (host infraorder p=0.049, host family p=0.002).

292

Figure 2. Diversity of biologically relevant viruses by host taxonomic and ecological variables. (A—D) viral alpha diversity plots: (A) viral richness according to host infraorder; (B) viral richness according to lizard habitat type; (C) viral richness according to host family; (D) Shannon index according to host habitat. Points are coloured by host family. Host taxonomic levels comprising only a single library were included in plots for visualisation only and excluded from the final statistical models (denoted by a # in host taxonomy alpha diversity plots). Csex_M was not included in statistical models or beta diversity plots as no biologically relevant viruses were found in this library and it was the only riparian library.

14

303 (E—G) viral beta diversity NMDS plot coloured by: (E) host family, (F) host habitat, and (G)
304 host range size. Cmet_M and Gnan_A overlap, depicted by a half circle. In all plots, P-values
305 for habitat or range size when considered with taxonomy at the level of host infraorder are
306 indicated with an "I", while those considered with taxonomy at the level of host family are
307 indicated with an "F". P-values <0.05 were considered significant and indicated with an
308 asterisk.
309

310 We also explored beta-diversity in relation to host traits (Figure 2E-G and Figure S5). A
311 similar trend to alpha diversity was observed, in which taxonomy, habitat and range size were
312 statistically significant in models considering host family (host family p=0.006, habitat
313 p=0.009, range size p=0.009), as was habitat and range size in models considering host
314 infraorder (p=0.013 and 0.033, respectively). Overall, we documented a general trend of host
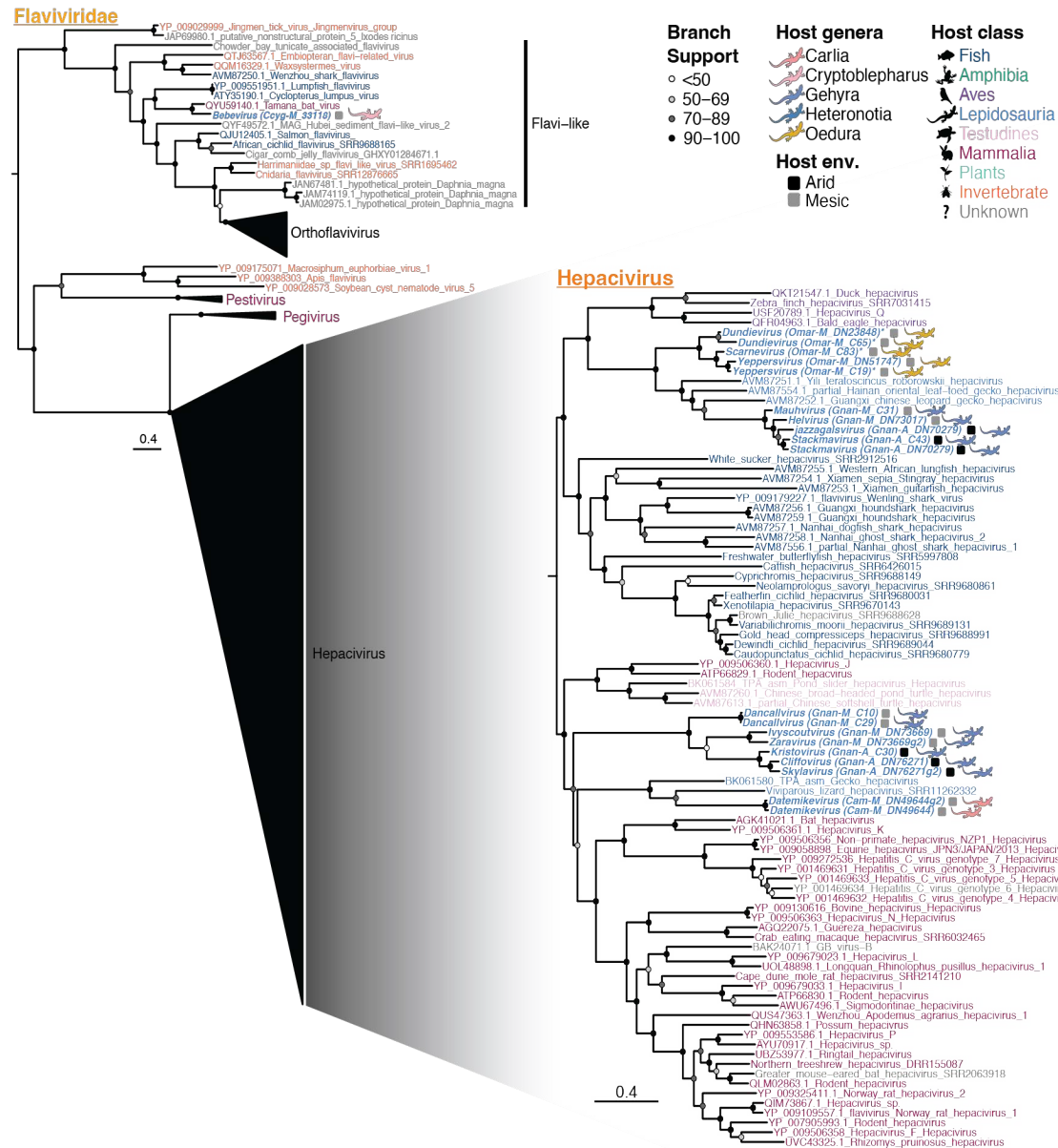315 taxonomy, habitat and range size as potential modulators of viral diversity.
316

317 **3.3 Evolutionary relationships of novel viruses**
318 *3.3.1 Flaviviridae*
319 Strikingly, most of the *Flaviviridae* identified in the lizards studied here fell within the genus
320 *Hepacivirus*. The Australian lizard hepaciviruses were all novel and fell into three clades in a
321 host specific manner (Figure 4). Hepaciviruses from *Carlia amax* formed a monophyletic
322 group with a viviparous lizard hepacivirus from *Zootoca vivipar*a (Lacertidae) sampled in the
323 UK (Scotland), and a Gecko hepacivirus previously sampled from *Oedura* in Australia
324 (although the RdRp sequence from the Gecko hepacivirus was only 92 amino acid residues,
325 necessarily impacting phylogenetic accuracy). In addition, the long branch lengths suggest
326 substantial unsampled genetic diversity within this clade. While this clade has strong support,
327 its specific phylogenetic placement within the hepaciviruses is poorly resolved such that its
328 closest relatives are unclear (Figure 4). *Gehyra nana* had a high diversity of hepaciviruses,
329 with viral sequences clustering into two groups: one distinct lineage that broadly clustered
330 with mammalian viruses and a second that grouped with reptile and bird viruses (Figure 4).
331 Viruses in the avian/reptile lineage grouped with viruses from other Gekkota species sampled
332 in China, including a *Goniurosaurus luii* (Gekkota; Eublepharidae, AVM87252.1),
333 *Hemidactylus bowringii* (Gekkota; Eublepharidae, AVM87554.1) and a *Teratoscincus*
334 *roborowskii* (Gekkota; Sphaerodactylidae, AVM87251.1). This is consistent with the Asian
335 origins of *Gehyra* (Oliver and Hugall 2017). Viruses from the *Oedura marmorata* sampled in
336 this study formed their own clade within the avian/reptile host lineage, basal to those from
337 China and *G. nana*.

15

338

339 Interestingly, the *Flaviviridae* identified in the *Cryptoblepharus metallicus* library did not

340 cluster with the other Australian lizard *Flaviviridae*, but rather with Tanama bat virus and

341 Lumpfish flavivirus (Figure 4), which are part of a broad "flavi-like" group that are

342 phylogenetically distinct from members of the genus *Orthoflavivirus*. The long branch

343 lengths between these three viruses suggest that there is considerable unsampled viral

344 diversity in this clade, and that this constitutes a new virus that we have tentatively named

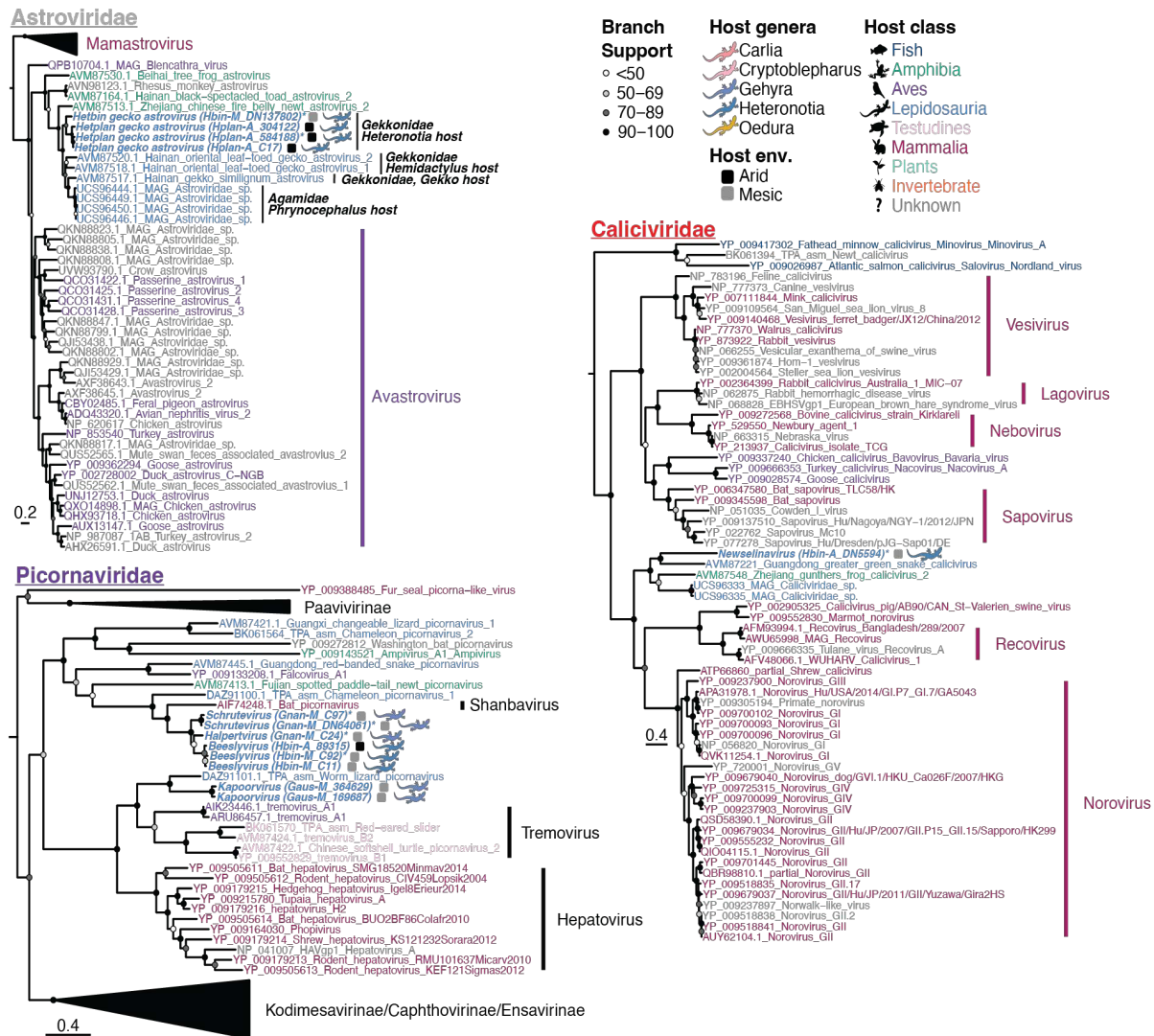345 Bebevirus.

346

347

348

349 **Figure 3. Maximum likelihood phylogeny of the RdRp of *Flaviviridae* in Australian**

350 **lizards.** Taxon names are coloured according to apparent host. Viruses discovered in this

16

351 study are indicated by bold and italicized taxa names and lizard silhouettes beside taxa
352 names, coloured by host genus. Squares next to the taxa names indicate the sampling
353 location/environment (env.) for viruses discovered in this study. An asterisk beside the taxa
354 name for viruses detected here indicates that the sequence is not the complete length of the
355 alignment. The accession number is indicated in the taxon name for sequences downloaded
356 from NCBI. Circles at the nodes represent the branch support as estimated using the SH-like
357 approximate likelihood ratio test. Trees are mid-point rooted for clarity. Scale bars indicate
358 the number of amino acid substitutions per site. To achieve greater resolution, the
359 *Hepacivirus* phylogeny was estimated from a sequence alignment of this genus only.
360

361 *3.3.2 Picornaviridae*

362 Members of the *Picornaviridae* were found in all Gekkota species analyzed, with the

363 exception of *H. planiceps* (although a picornavirus contig from *O. marmorata* was excluded

364 from the phylogeny as there were no RdRp contigs of sufficient length). The picornaviruses

365 isolated from lizards in this study fell into two main groups. Viruses from *H. binoei* and *G.*

366 *nana* grouped together and shared a common ancestor with bat picornavirus (AIF74248.1;

367 53.5-62% amino acid [aa] identity) and Chameleon picornavirus 1 (DAZ91100; 50-58.7% aa

368 identity) (Figure 4), a virus found in multiple lizard species including *Kinyongia boehmei*

369 (Chamaeleonidae), *Podarcis muralis* (Lacertidae), and *Timon pater* (Lacertidae). Bat

370 picornavirus is a member of the genus *Shanbavirus*, and (based on identity in the RdRp) the

371 *H. binoei* and *G. nana* viruses, as well as Chameleon picornavirus 1, should also belong to

372 the shanbaviruses (https://ictv.global/report/chapter/picornaviridae/picornaviridae). Three

373 distinct *Shanbavirus* species were identified in this study, provisionally named Halpertvirus

374 (*G. nana* host), Schrutevirus (*G. nana* host), and Beeslyvirus (*H. binoei* host). Notably,

375 Beeslyvirus was found in *H. binoei* from both mesic and arid environments and was the only

376 virus found in more than one library.

377

378 The second group contained the *G. arnhemica* picornavirus sequences, which were closely

379 related and hence assigned as the same virus species, here provisionally named Kapoorvirus.

380 This virus was most closely related to worm lizard picornavirus (57.6% aa identity in the

381 RdRp) from *Blanus cinereus* (Blanidae) and more broadly related to members of the genus

382 *Tremovirus* (Figure 4), associated with reptiles and birds, and known to cause

383 encephalomyelitis in avian hosts. The Kapoorviruses shared 34.3-41.6% amino acid identity

384 in the RdRp with the *Tremoviruses* included in the phylogeny, which placed them on the

385 border of the level of divergence required for a new genus (64% divergence). As such, we

386 have tentatively placed them within the genus *Tremovirus*.

387

Figure 4. Maximum likelihood phylogenies of the RdRp of positive-sense RNA viruses in Australian lizards (with the exception of the *Flaviviridae*, shown in Figure 3). Taxon names are coloured according to apparent host. Viruses discovered in this study are indicated by bold and italicized taxa names and lizard silhouettes beside taxa names, which are coloured by host genus. Squares next to the taxa names indicate the sampling location/environment (env.) for viruses discovered in this study. An asterisk beside the taxa name for viruses detected here indicates that the sequence is not the complete length of the alignment. The accession number is indicated in the taxon name for sequences downloaded from NCBI. Circles at the nodes represent the branch support as estimated using the SH-like approximate likelihood ratio test. Trees are mid-point rooted for clarity. Scale bars indicate the number of amino acid substitutions per site. Host family and genus names are indicated for the Lepidosauria in the *Astroviridae* phylogeny to demonstrate virus-host co-divergence within the Lepidosauria.

3.3.3 Astroviridae

Members of the *Astroviridae* were detected in two *Heteronotia* libraries from two different

species – *H. binoei* and *H. planiceps*. A distinct astrovirus was identified in each library,
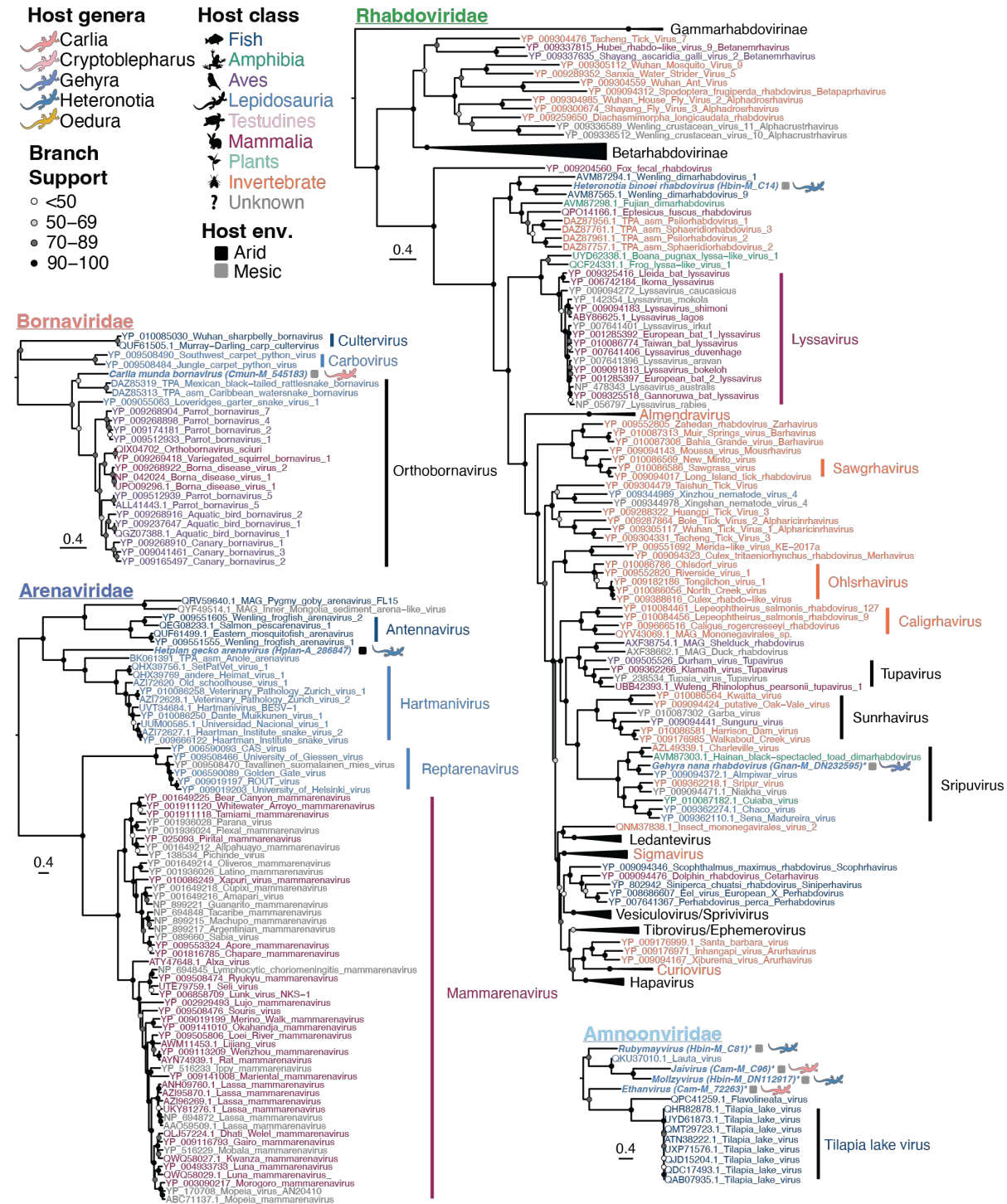
18

407 designated as Hetplan gecko astrovirus (from *H. planiceps*) and Hetbin gecko astrovirus

408 (from *H.* binoei), and the two viruses grouped together in the RdRp phylogeny (Figure 4).

409 These viruses then grouped more broadly with a larger well-supported clade of viruses from

410 lizards (Figure 4), clustering in a fashion reflective of host family and genus. Specifically, the

411 viruses from *Heteronotia* grouped closely with viruses from other members of the

412 Gekkonidae, including *Hemidactylus bowringii* and *Gekko similignum,* with each also

413 clustering by host genus, all of which fell basal to a clade of viruses from Agamidae

414 (*Phrynocephalus erythrurus* and *Phrynocephalus theobaldi*), mirroring the host phylogeny

415 (Figure 4). Genus demarcation criteria for the *Astroviridae* is unclear and is most strongly

416 associated with host species. As a consequence, the viruses found here, along with the

417 published lizard astroviruses, would most likely form their own genus. It is also notable that

418 the lizard astrovirus cluster grouped more broadly with viruses found in amphibians and

419 birds.

420

421 *3.3.4 Rhabdoviridae*

422 Viral contigs related to the *Rhabdoviridae* were identified in two libraries (Figure 1). The *H.*

423 *binoei* rhabdovirus fell with a group of viruses not classified to any genus, including those

424 isolated from ray-finned fishes (*Actinopterygii*; from liver, gut and gills), and a spotted

425 paddle-tail newt (*Pachytriton brevipes*; from gut) (Figure 5). This virus is highly divergent

426 (49% aa identity in the RdRp to the closest relative, Wenling dimarhabdovirus 9) and would

427 therefore represent a new species, that we have named Heteronotia binoei rhabdovirus. This

428 virus, together with Wenling dimarhabdovirus 9, Fujian dimarhabdovirus, and Wenling

429 dimarhabdovirus 1, should form at least one new genus based on the observed level of

430 sequence diversity. This group of unclassified viruses then fell as a sister-group to a large

431 clade of rhabdoviruses including lyssavirus and many other pathogenic viruses (Figure 5).

432 Only short *Rhabdoviridae* contigs could be obtained for the RdRp region from the *G. nana*

433 library (Gnan_M), the longest being 124 residues. A phylogeny was estimated using a

434 smaller region of the RdRp (230 aa after trimAl trimming) where this contig aligned to

435 confirm the phylogenic groupings observed in the larger alignment (774 aa after trimAl

436 trimming). Regardless of the RdRp alignment used, the *G. nana* rhabdovirus grouped with

437 the *Sripuviruses* (Figure 5). It shared 75% identity in the RdRp protein to its closest relative –

438 Almpiwar virus, isolated from skinks in Northern Queensland Australia. As all RdRp contigs

439 from Gnan_M (including the one in the phylogeny) exhibited less than 90% amino acid

440   identity to the closest relative, this virus likely represents a new species of *Sripuvirus,*

441   tentatively named Gehyra nana rhabdovirus.

Figure 5. Maximum likelihood phylogenies of the RdRp of negative-sense RNA viruses in Australian lizards. Taxon names are coloured according to apparent host. Viruses discovered in this study are indicated by bold and italicized taxa names and lizard silhouettes beside taxa names, which are coloured by host genus. Squares next to the taxa names indicate the sampling location/environment (env.) for viruses discovered in this study. An asterisk beside the taxa name for viruses detected here indicates that the sequence is not the complete length of the alignment. Accession numbers are indicated in the taxon name for sequences

20

451 downloaded from NCBI. Circles at the nodes represent the branch support as estimated using
452 the SH-like approximate likelihood ratio test. Trees are mid-point rooted for clarity. Scale
453 bars indicate the number of amino acid substitutions per site.
454

455 *3.3.5 Amnoonviridae*

456 Several contigs related to Lauta virus (MT386081), a member of the *Amnoonviridae*, were

457 identified in *C. amax* and *H. binoei* sampled from the same area of the north-east Top End.

458 However, the contigs obtained were short and hence difficult to classify. The amino acid

459 sequence identity between the longest contigs (270 aa and 238 aa) from the two libraries is

460 only 26.7%, while the identity of the longest contigs from *C. amax* and *H. binoei* with the

461 nearest published relative (MT386081, Lauta virus) were 31.3% and 35.8%, respectively.

462 The nearest known relative, found in an Australian gecko (*Gehyra lauta*) liver in 2013 (Ortiz-

463 Baez et al. 2020), is itself highly divergent from other published viruses, and was only

464 detected using a protein structure prediction approach. All *Amnoonviridae* contigs detected in

465 this study were from the RdRp segment as these viruses are so distinct from known viruses

466 that it was challenging to identify other genomic regions, particularly as these viruses are

467 likely to be segmented. Phylogenetic analysis of the RdRp of contigs longer than 100 aa from

468 each library suggested that there were four distinct species of lizard virus within

469 *Amnoonviridae*, two in each library (Figure 5). Three of these species clustered with Lauta

470 virus, creating a lizard-infecting clade, while the fourth was basal (yet quite distant) to a clade

471 of fish viruses, including Tilapia lake virus (Figure 5). However, the clustering of these

472 contigs may be unreliable since they are short (160 – 270 aa) and not the complete length of

473 the RdRp. Additionally, contigs from the same library did not overlap (or only overlapped

474 minimally) in the alignment and therefore their relationship is difficult to determine.

475

476 **3.4 Viral families in which a single species was identified**

477 A single virus species was detected in the lizard metatranscriptomes from the *Caliciviridae,*

478 *Arenaviridae*, *Bornaviridae*, and *Iridoviridae*. A member of the *Caliciviridae* was detected in

479 the *H. binoei* (Hbin_A) (Figure 1), but only partial contigs were obtained which aligned to

480 most of the capsid, parts of the RdRp, 2C-like protein and the proteinase. Alignment of a 449

481 aa region of the RdRp with published sequences grouped this virus with caliciviruses from

482 other squamates; *Cyclophiops major* (Colubridae), *Phrynocephalus theobaldi* (Agamidae),

483 and a frog, *Sylvirana guentheri* (Ranidae) (Figure 4). The *H. binoei* calicivirus was highly

484 divergent from its closest relative – Guangdong greater green snake calicivirus (AVM87221)

485 – sharing only 35.4% identity in the aligned region of the RdRp. We therefore suggest that it

486   comprises a new species that we have tentatively named Newselinavirus. Based on alignment

487   of the near complete major capsid protein (527 aa), this virus has >60% amino acid sequence

488   difference to other caliciviruses and therefore would also constitute a different genus

489   (https://ictv.global/report/chapter/caliciviridae/caliciviridae).
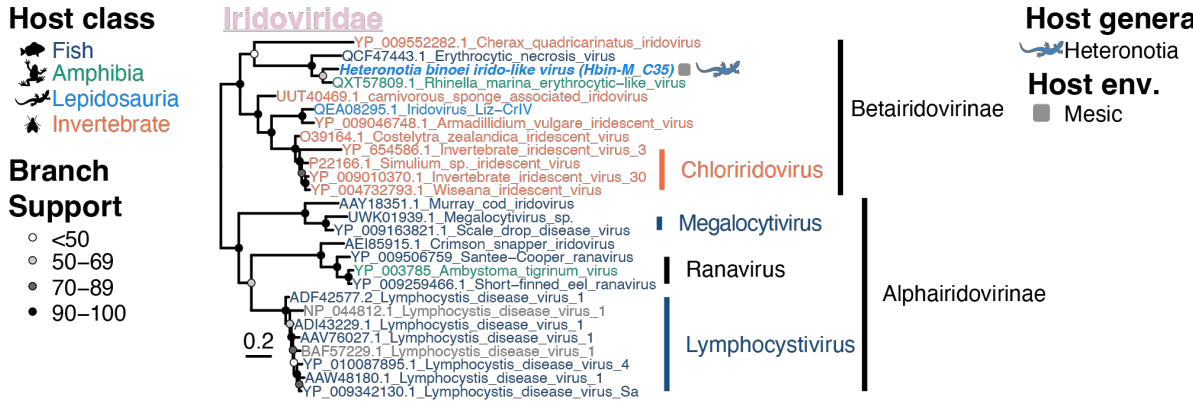
490

491   A member of the *Arenaviridae* was detected in *H. planiceps* (Hplan_A) with the highest

492   abundance of any virus detected in this study (>0.2% of total reads for that library, Figure 1)

493   with a near complete genome obtained. This virus clusters with the *Hartmaniviruses* (Figure

494   5), a genus known to infect snakes with unknown pathogenicity, although with only 23.6% aa

495   identity to its closest relative in the RdRp region. However, based on the genus demarcation

496   criteria for *Arenaviridae* (i.e., members of the same genus share >35% nucleotide identity in

497   the L gene; https://ictv.global/report/chapter/arenaviridae/arenaviridae), this virus falls just

498   within the genus *Hartmanivirus*. Like other members of this genus, the virus detected here,

499   tentatively named Hetplan gecko arenavirus, lacked the gene encoding the zinc-binding

500   matrix protein (Hepojoki et al. 2015).

501   A near-complete genome of a member of the *Bornaviridae* was found in the *C. munda* library

502   (Cmun_M). Phylogenetic analysis of the RdRp showed that it formed a clade with two snake

503   viruses from the genus *Orthobornavirus* which also contained Loveridges garter snake virus

504   1 (YP_009055063) and viruses with mammalian and avian hosts (Figure 5). As it shared only

505   53.5% nucleotide identity across the entire genome with its nearest relative, the virus detected

506   here would constitute its own species (Kuhn et al. 2015), tentatively named Carlia munda

507   bornavirus, and would likely belong to the genus *Orthobornavirus*. It is not clear whether the

508   two snake viruses that grouped with Carlia munda bornavirus are pathogenic (Pfaff and

509   Rubbenstroth 2021), but other members of the *Orthobornavirus* clade cause neurotropic

510   disease in mammals and avian hosts, while members of the *Carbovirus* genus cause

511   neurological disease in snakes (Hyndman et al. 2018).

512   Transcripts from multiple genes related to the *Iridoviridae* (double-strand DNA viruses) were

513   detected in *H. binoei* (Hbin_M) (Figure 1), including the major capsid, myristylated

514   membrane protein, phosphotransferase, ATPase, DNA helicase, dUTP, and NTPase I. This

515   suggested that an *Iridoviridae* was actively replicating in one or more lizards sampled in this

516   library. Phylogenetic analysis of the major capsid protein revealed that this virus clustered

517   within the subfamily *Betairidovirinae* that was originally considered invertebrate-specific but

518   which is increasingly being associated with vertebrate infection (Russo et al. 2021) (Figure

519   6). Interestingly, this virus grouped with erythrocytic necrosis virus, a pathogen of fish

520   (72.4% aa identity in the capsid), and Rhinella marina erythrocytic-like virus found in cane

521 toads (85.4% aa identity) and would likely fall into the same genus

522 (https://ictv.global/report/chapter/iridoviridae/iridoviridae), but as a novel species that we

523 have named Heteronotia binoei irido-like virus (Figure 6). A new genus containing these and

524 erythrocytic viruses of ectothermic hosts has been putatively proposed, but not yet formally

525 ratified or named (Emmenegger et al. 2014). Three other erythrocytic viruses from Squamata

526 hosts – Thamnophis sauritus erythrocytic virus (EV), Lacerta monticola EV, and Pogona

527 vitticeps EV – fell within this erythrocytic virus lineage based on sequence of a short PCR

528 product of a homolog to a region of the DNA dependent DNA polymerase (Grosset et al.

529 2014; Russo et al. 2018), although the transcript encoding this particular protein was not

530 found in our data, precluding a phylogenetic analysis. However, it is likely the virus

531 discovered here is a relative to the other Squamata erythrocytic viruses and supports the

532 presence of Squamata hosts for this lineage.

533

534
535
536 **Figure 6. Maximum likelihood phylogenetic tree of the major capsid protein of the**
537 *Iridoviridae* **in Australian lizards.** Taxon names are coloured according to apparent host.
538 The virus discovered in this study is indicated by a bold and italicized taxon name and a
539 lizard silhouette beside the taxon name, which is coloured by host genus. The square next to
540 the taxon name indicates the sampling location/environment (env.) for the virus discovered in
541 this study. The accession number is indicated in the taxon name for sequences downloaded
542 from NCBI. Circles at the nodes represent the branch support as estimated using the SH-like
543 approximate likelihood ratio test. The tree is mid-point rooted for clarity, and the scale bar
544 indicates the number of amino acid substitutions per site.

545

4. Discussion

547 We used Australian lizards as model organisms to better understand the host determinants of

548 viral abundance and diversity, while simultaneously expanding our knowledge of the lizard

549 virome and virus evolution. Indeed, to date relatively little attention has been paid to the

550 viromes of squamates (lizards and snakes) (Shi et al. 2018; Harding et al. 2022). In Australia,

551 the Squamata are the most species-rich vertebrate assemblage (Wilson and Swan 2017) and

552 the varied ages of Squamate radiations (Brennan and Oliver 2017), alongside their varied

553   population structures and their adaption to multiple habitats and environments (Pianka 1969;

554   Pianka 1973), makes them an informative group to study in the context of viral evolution and

555   ecology. Using a meta-transcriptomics approach we characterized the virome of nine lizard

556   species from three families and five genera, finding a wide diversity of novel viruses,

557   including those from families/genera that commonly cause disease in humans and other

558   animals, including the *Flaviviridae* (*Hepacivirus*), *Bornaviridae*, *Rhabdoviridae*, and

559   *Picornaviridae*. Our findings expand many known viral families, adding entire new clades.

560   This includes a new lizard-specific clade of the *Amnoonviridae*, a family previously

561   containing only fish viruses and a single lizard-infecting virus (Ortiz-Baez et al. 2020).

562

563   Viral diversity and abundance differed markedly between hosts, and consistent with other

564   studies on virome and pathogen ecology, environmental variables had some effect on virome

565   diversity across our sampled hosts – particularly host habitat. The effect of various habitats

566   on alpha diversity in viruses has previously been demonstrated in fish (Geoghegan et al.

567   2021), rodents (Tirera et al. 2021), and bats (Bergner et al. 2020), while virome composition

568   has also been shown to differ significantly in different habitats for bats, shrews and rats

569   (Chen et al. 2023). As different habitats have a range of environmental factors that modulate

570   host behavior and ecology, it is unsurprising that host habitat may also impact virome

571   composition. For example, foraging behavior is likely to play a role in the transmission of

572   low pathogenic avian influenza, where avian species that forage in shallow water are more

573   likely to influence avian influenza ecology (Wille et al. 2023).

574

575   Host taxonomy also impacted measures of virus richness and beta diversity, consistent with

576   some degree of host-specific virus evolution, including co-divergence. The effect of host

577   taxonomy on viral diversity is similar to that seen in fish (Geoghegan et al. 2021) and

578   mammalian viromes (Olival et al. 2017), although a different pattern was observed in some

579   bird viromes (Wille et al. 2018, 2019), suggesting that it is dependent on the taxa in question.

580   Over evolutionary time the likelihood of successful cross-species virus transmission is

581   predicted to decline because of host genetic differences in viral receptors, the cellular

582   machinery required for replication, immune response and other factors that affect viral

583   infection and spread (Ferris, Heise and Baric 2016). This may in part explain the link

584   between host taxonomy and viral richness and composition. Alternatively, this association

585   could simply reflect virome-level virus-host co-divergence over longer evolutionary

586   timescales. Indeed, in most cases, the lizard viruses discovered here grouped together and/or

587 with viruses found in other members of the reptilian class Lepidosauria, and phylogenies

588 generally followed a broad pattern of host-virus co-divergence. For example, in the case of

589 the *Astroviridae*, virus sequences from Squamata clearly group by host genera, and viruses

590 from the Gekkonidae fall basal to those from the Agamidae, matching the host phylogeny

591 (Oliver and Hugall 2017). Similarly, in the case of the relatively well-sampled hepaciviruses

592 and picornaviruses, more closely related Squamata hosts tended to carry more closely related

593 viruses, with viruses from the same host species forming a monophyletic group and viruses

594 from the same family and infraorder similarly tending to do the same.

595

596 As well as providing some evidence for virus-host co-divergence, we documented many

597 instances of virus-host jumping or multi-host viruses at higher taxonomic scales. Indeed, in

598 some cases, lizard viruses clustered more closely with amphibian, mammalian or fish viruses,

599 rather than with other reptile or bird viruses. Three distinct clades of lizard hepaciviruses

600 were identified, all of which group separately to a clade of turtle hepaciviruses. Although the

601 hepacivirus phylogeny is difficult to resolve, the presence of four clades of reptile

602 hepaciviruses and the placement of two reptile clades within the mammalian hepacivirus

603 clade is suggestive of host jumping between reptiles and mammals at some time in the distant

604 evolutionary past. Similarly, the *Iridoviridae* and *Rhabdoviridae* phylogenies are highly

605 incongruent with their host phylogenies, suggesting frequent host jumping. In the case of the

606 two lizard *Rhabdoviridae*, one groups with fish viruses while the other clusters in the

607 *Sripuvirus* clade with viruses from reptiles, amphibians and invertebrates. The closest relative

608 of the *Sripuvirus* detected here, Almpiwar virus, was isolated from skinks in Northern

609 Queensland Australia, and neutralizing antibodies against Almpiwar virus were also found in

610 the sera of crocodiles, a wild bird and multiple mammalian hosts, including a human

611 (McAllister et al. 2014). Although, the presence of neutralizing antibodies against Almpiwar

612 virus in mammals was relatively rare, it does suggest that this group of viruses could have a

613 broader host range than reptiles alone. In addition, there is evidence that this virus, along with

614 other members of this genus – including Charleville virus, which was also isolated in

615 Australia from *Gehyra* (the same host genus as the virus detected here) and sandflies – are

616 arthropod-borne (McAllister et al. 2014; Vasilakis et al. 2019). The lizard *Iridoviridae* found

617 was most closely related to an amphibian and fish virus, with a relatively short genetic

618 distance between them. This is perhaps unsurprising as recent host jumping seems to be

619 commonplace within the *Iridoviridae*; some members of this family can infect fish, reptiles

620 and amphibians (Brenes et al. 2014), and some invertebrate-infecting *Iridoviridae* can also

25

621 apparently infect vertebrates (Papp and Marschang 2019). In sum, this work supports a model

622 of frequent host-jumping throughout the evolutionary history of most virus families on a

623 backdrop of virus-host co-divergence (Geoghegan, Duchêne and Holmes 2017; Geoghegan

624 and Holmes 2017; Shi et al. 2018).

625

626 Our phylogenetic analyses also revealed some geospatial clustering. For virus families with

627 multiple detections, there is little overlap of virus species from different environments. This

628 suggests that lizard viruses are evolving within their host populations in a relatively isolated

629 manner, as might be expected given the generally strong phylogeographic structuring across

630 species ranges (Moritz et al. 2016; Moritz et al. 2018; Potter et al. 2018). It is also of interest

631 that phylogenetic groups that contain viruses sampled from both mesic and arid locations

632 tend to have mesic associated viruses in the basal positions. The hepacivirus phylogeny is

633 particularly enlightening in this respect and it suggests the movement of viruses from one

634 environment to another: both clades of *G. nana* hepaciviruses have viruses collected from

635 mesic environments located in basal phylogenetic positions to those viruses sampled from

636 animals in arid environments. This is consistent with the theory that Australia's mesic

637 terrestrial biota is mostly ancestral (Byrne et al. 2011). Additionally, more consistent mesic

638 conditions may favor the retention of viral diversity, such that younger taxa dominate in the

639 more climatically variable arid regions of the Kimberley. This could also be associated with

640 the more variable demographic histories of the drier Kimberley compared to the mesic Top

641 End populations (Potter et al. 2018).

642

643 The hepacivirus phylogeny also reflects host historical biogeography. As >90% of Australian

644 reptiles are endemic (Chapman 2009), it is expected that viruses infecting Australian reptiles

645 would have minimal opportunities to spread between countries. Indeed, even the currently

646 limited sampling of lizard viruses demonstrates little movement between countries. However,

647 in one clade of hepaciviruses, viruses from Australian *O. marmorata* hosts were separated

648 from viruses from Australian *G. nana* by viruses sampled from China, with the Chinese

649 sampled viruses falling as sister taxa to the Australian *G. nana* viruses. Interestingly, *O.*

650 *marmorata* are thought to have Gondwanan origins, while the *G. nana* are believed to have

651 immigrated from Asia around the Eocene-Oligocene transition (Oliver and Hugall 2017).

652 Thus, this clade of hepaciviruses could reflect the historical biogeography of their hosts, with

653 evidence of a lineage of viruses introduced to Australia within immigrant ancestors. As the

654 evolutionary histories of the viruses studied here frequently aligned with the biogeographic

26

655 and paleoclimatic histories of the hosts, this supports the role of virus evolution in informing

656 animal host ecological histories where sampling is strategic and sufficiently dense (Wilfert

657 and Jiggins 2014). Clearly, however, increased sampling across taxa and biogeographic

658 regions would be useful in confirming this hypothesis.

659

660 Although the disease association, if any, of the viruses collected here is unknown, we did

661 identify viruses related to known pathogens, including an *Orthobornavirus*, multiple

662 *Hepaciviruses*, and a member of the *Iridoviridae*. It is noteworthy that seemingly healthy

663 lizards can carry a very high abundance and richness of viruses in the liver, particularly the

664 *Gehyra* species (high viral richness) and *H. planiceps* (high abundance). Viral emergence can

665 have devastating effects on reptile species, such as the Bellinger River snapping turtle that is

666 now critically endangered following mass mortalities due to a novel nidovirus outbreak

667 (Zhang et al. 2018). Although it is unclear where the novel nidovirus (with kidney tropism)

668 originated, the closest known viral relatives were observed in pythons and lizards (usually

669 associated with respiratory disease) (Zhang et al. 2018). Given that lizards make up 59% of

670 non-avian reptiles (Pincheira-Donoso et al. 2013), and one fifth of reptile species are

671 threatened (Cox et al. 2022), this expansion of the lizard virosphere and elucidation of viral

672 genomic sequences may help inform and prepare for potential threats to reptile species.

673

674 While this work is the first structured examination of the lizard virome, the sample size is

675 relatively small and there are several caveats. Previous studies have found variation in

676 viromes depending on the age of the individuals sampled (Bergner et al. 2020) and the time

677 of year that sampling occurred (Raghwani et al. 2022). These variables were not controlled in

678 this study. In addition, a larger sample size would enable better comparison between matched

679 species sampled from different environments and a greater sample across the squamate

680 phylogeny would expand these results. The viral families detected here are reflective of the

681 fact that liver samples were sequenced, and as viromes commonly differ by the tissue of

682 sampling, studying a wider range of tissues would be beneficial. Despite these limitations,

683 this study demonstrates that lizards carry a large diversity of viruses, often in high abundance

684 and potentially species-specific. As such they are not only interesting models of vertebrate

685 evolution and ecology, but also serve as good host models by which to study virus ecology

686 and evolution. This work further demonstrates that virus evolution may be a useful tool for

687 understanding or corroborating host biogeographic and paleoclimatic histories.

688

Data availability

All raw data (fastq files) generated for this study are available in the NCBI SRA database under BioProject XXXX, BioSample accessions XXXX. Consensus sequences of assembled viral contigs presented in phylogenies are available in GenBank under accession numbers XXXXX-XXXX.

Acknowledgements

Funding

Conflict of interest: none declared.

References

Altschul, S. F. et al. (1990). 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.

Altschul, S. F. et al. (1997). 'Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs.', *Nucleic Acids Research*, 25: 3389–402.

Bergner, L. M. et al. (2020). 'Demographic and Environmental Drivers of Metagenomic Viral Diversity in Vampire Bats', *Molecular Ecology*, 29: 26–39.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). 'Trimmomatic: a Flexible Trimmer For Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.

Brenes, R. et al. (2014). 'Transmission of *Ranavirus* between Ectothermic Vertebrate Hosts', *PLoS One*, 9: e92476.

Brennan, I. G., and Oliver, P. M. (2017). 'Mass Turnover and Recovery Dynamics of a Diverse Australian Continental Radiation', *Evolution*, 71: 1352–65.

Buchfink, B., Reuter, K., and Drost, H.-G. (2021). 'Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND', *Nature Methods*, 18: 366–8.

Byrne, M. et al. (2011). 'Decline of a Biome: Evolution, Contraction, Fragmentation, Extinction and Invasion of the Australian Mesic Zone Biota', *Journal of Biogeography*, 38: 1635–56.

726   Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). 'trimAl: a Tool for
727   Automated Alignment Trimming in Large-Scale Phylogenetic Analyses',
728   *Bioinformatics*, 25: 1972–3.
729   Chapman, A. D. (2009). *Numbers of Living Species in Australia and the World*. Australian
730   Government, Department of the Environment, Water, Heritage, and the Arts.
731   Chen, Y.-M. et al. (2023). 'Host Traits Shape Virome Composition and Virus Transmission
732   in Wild Small Mammals', *Cell*, 186: 4662-75.
733   Chen, Y.-M. et al. (2022). 'RNA Viromes from Terrestrial Sites Across China Expand
734   Environmental Viral Diversity', *Nature Microbiology*, 7: 1312–23.
735   Cox, N. et al. (2022). 'A Global Reptile Assessment Highlights Shared Conservation Needs
736   of Tetrapods', *Nature*, 605: 285–90.
737   Emmenegger, E. J. et al. (2014). 'Molecular Identification of Erythrocytic Necrosis Virus
738   (ENV) From the Blood of Pacific herring (*Clupea pallasii*)', *Veterinary*
739   *Microbiology*, 174: 16–26.
740   Ferris, M. T., Heise, M. T., and Baric, R. S. (2016). 'Host Genetics: It Is Not Just the Virus,
741   Stupid'. In Katze M. G., Korth M. J., Law G. L., and Nathanson N. (eds) *Viral*
742   *Pathogenesis (Third Edition)*, pp. 169–79. Academic Press: Boston.
743   Fu, L., et al. (2012). 'CD-HIT: Accelerated for Clustering the Next-Generation Sequencing
744   Data', *Bioinformatics*, 28: 3150–2.
745   Geoghegan, J. L. et al. (2021). 'Virome Composition in Marine Fish Revealed by Meta-
746   Transcriptomics', *Virus Evolution*, 7: veab005.
747   Geoghegan, J. L., Duchêne, S., and Holmes, E. C. (2017). 'Comparative Analysis Estimates
748   the Relative Frequencies of Co-Divergence and Cross-Species Transmission Within
749   Viral Families', *PLoS Pathogens*, 13: e1006215.
750   Geoghegan, J. L., and Holmes, E. C. (2017). 'Predicting Virus emergence Amid Evolutionary
751   Noise', *Open Biology*, 7: 170189.
752   Geyle, H. M. et al. (2021). 'Reptiles on the Brink: Identifying the Australian Terrestrial
753   Snake and Lizard Species Most at Risk of Extinction', *Pacific Conservation Biology*,
754   27: 3-12.
755   Grabherr, M. G. et al. (2011). 'Full-Length Transcriptome Assembly from RNA-Seq Data
756   Without a Reference Genome', *Nature Biotechnology*, 29: 644–52.
757   Grosset, C. et al. (2014). 'Intraerythrocytic Iridovirus in Central Bearded Dragons (*Pogona*
758   *vitticeps*)', *Journal of Veterinary Diagnostic Investigation*, 26: 354–64.
759   Guindon, S. et al. (2010). 'New Algorithms and Methods to Estimate Maximum-Likelihood
760   Phylogenies: Assessing the Performance of PhyML 3.0', *Systematic Biology*, 59:
761   307–21.
762   Harding, E. F. et al. (2022). 'Revealing the Uncharacterised Diversity of Amphibian and
763   Reptile Viruses', *ISME Communications*, 2: 95.
764   Hepojoki, J. et al. (2015). 'Arenavirus Coinfections Are Common in Snakes with Boid
765   Inclusion Body Disease', *Journal of Virology*, 89: 8657–60.
766   Herrera-Flores, J. A., Stubbs, T. L., and Benton, M. J. (2021). 'Ecomorphological
767   Diversification of Squamates in the Cretaceous', *Royal Society Open Science*, 8:
768   201961.
769   Hyndman, T. H. et al. (2018). 'Divergent Bornaviruses from Australian Carpet Pythons with
770   Neurological Disease Date the Origin of Extant *Bornaviridae* Prior to the End-
771   Cretaceous Extinction', *PLoS Pathogens*, 14: e1006881.
772   Kalyaanamoorthy, S. et al. (2017). 'ModelFinder: Fast Model Selection for Accurate
773   Phylogenetic Estimates', *Nature Methods*, 14: 587–9. 5
774   Kuhn, J. H. et al. (2015). 'Taxonomic Reorganization of the Family *Bornaviridae*', *Archives*
775   *of Virology*, 160: 621–32.

29

776 Lagkouvardos, I. et al. (2017). 'Rhea: a Transparent and Modular R Pipeline for Microbial
777   Profiling Based on 16S rRNA Gene Amplicons', *PeerJ*, 5: e2836.
778 Li, B. et al. (2010). 'RNA-Seq Gene Expression Estimation with Read Mapping Uncertainty',
779   *Bioinformatics*, 26: 493–500.
780 Li, D. et al. (2015). 'MEGAHIT: an Ultra-Fast Single-Node Solution for Large and Complex
781   Metagenomics Assembly Via Succinct de Bruijn graph', *Bioinformatics*, 31: 1674–6.
782 McAllister, J. et al. (2014). 'Genomic Characterisation of Almpiwar virus, Harrison Dam
783   virus and Walkabout Creek virus; Three Novel Rhabdoviruses from Northern
784   Australia', *Virology Reports*, 3–4: 1–17.
785 McMurdie, P. J., and Holmes, S. (2013). 'phyloseq: an R Package for Reproducible
786   Interactive Analysis and Graphics of Microbiome Census Data', *PLoS One*, 8:
787   e61217.
788 Mollentze, N., and Streicker, D. G. (2020). 'Viral Zoonotic Risk is Homogenous Among
789   Taxonomic Orders of Mammalian and Avian Reservoir Hosts', *Proceedings of the
790   National Academy of Sciences*, 117: 9423–30.
791 Moritz, C. C. et al. (2018). 'Cryptic Lineage Diversity, Body Size Divergence, and Sympatry
792   in a Species Complex of Australian Lizards (*Gehyra*)', *Evolution*, 72: 54–66.
793 Moritz, C. et al. (2016). 'Multilocus Phylogeography Reveals Nested Endemism in a Gecko
794   Across the Monsoonal Tropics of Australia', *Molecular Ecology*, 25: 1354–66.
795 Morton, S. R., and James, C. D. (1988). 'The Diversity and Abundance of Lizards in Arid
796   Australia: A New Hypothesis', *The American Naturalist*, 132: 237–56.
797 Nguyen, L.-T. et al. (2015). 'IQ-TREE: A Fast and Effective Stochastic Algorithm for
798   Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*,
799   32: 268–74.
800 Oksanen, J. et al. (2022). 'Vegan: Community Ecology Package'. http://CRAN.R-
801   project.org/package=vegan.
802 Olival, K. J. et al. (2017). 'Host and Viral Traits Predict Zoonotic Spillover From Mammals',
803   *Nature*, 546: 646–50.
804 Oliver, P. M., and Hugall, A. F. (2017). 'Phylogenetic Evidence for Mid-Cenozoic Turnover
805   of a Diverse Continental biota', *Nature Ecology & Evolution*, 1: 1896–902.
806 Ortiz-Baez, A. S. et al. (2020). 'A Divergent *Articulavirus* in an Australian Gecko Identified
807   Using Meta-Transcriptomics and Protein Structure Comparisons', *Viruses*, 12: 613.
808 Papp, T., and Marschang, R. E. (2019). 'Detection and Characterization of Invertebrate
809   Iridoviruses Found in Reptiles and Prey Insects in Europe over the Past Two
810   Decades', *Viruses*, 11: 600.
811 Pfaff, F., and Rubbenstroth, D. (2021). 'Two Novel Bornaviruses Identified in Colubrid and
812   Viperid Snakes', *Archives of Virology*, 166: 2611–4.
813 Pianka, E. R. (1969). 'Habitat Specificity, Speciation, and Species Density in Australian
814   Desert Lizards', *Ecology*, 50: 498–502.
815 Pianka, E. R. (1973). 'The Structure of Lizard Communities', *Annual Review of Ecology and
816   Systematics*, 4: 53–74.
817 Pincheira-Donoso, D. et al. (2013). 'Global Taxonomic Diversity of Lving Reptiles', *PLoS
818   One*, 8: e59741.
819 Potter, S. et al. (2018). 'Pleistocene Climatic Changes Drive Diversification Across a
820   Tropical Savanna', *Molecular Ecology*, 27: 520–32.
821 Pyron, R. A., Burbrink, F. T., and Wiens, J. J. (2013). 'A Phylogeny and Revised
822   Classification of Squamata, Including 4161 Species of Lizards and Snakes', *BMC
823   Evolutionary Biology*, 13: 93.
824 Raghwani, J. et al. (2022). 'Seasonal Dynamics of the Wild Rodent Faecal Virome',
825   *Molecular Ecology*, 32: 4763-76.

Russo, A. G. et al. (2018). 'Viral Discovery in the Invasive Australian Cane Toad (*Rhinella marina*) Using Metatranscriptomic and Genomic Approaches', *Journal of Virology*, 92/17: e00768-18.

Russo, A. G. et al. (2021). 'Discovery of Novel Viruses Associated with the Invasive Cane Toad (*Rhinella marina*) in Its Native and Introduced Ranges', *Frontiers in Microbiology*, 12:733631.

Shi, M. et al. (2018). 'The Evolutionary History of Vertebrate RNA Viruses', *Nature*, 556: 197–202.

Tirera, S. et al. (2021). 'The Influence of Habitat on Viral Diversity in Neotropical Rodent Hosts', *Viruses*, 1: 1690.

Vasilakis, N. et al. (2019). 'Genomic Characterisation of Cuiaba and Charleville viruses: Arboviruses (family *Rhabdoviridae*, genus *Sripuvirus*) Infecting Reptiles and Amphibians', *Virus Genes*, 55: 87–94.

Wilfert, L., and Jiggins, F. M. (2014). 'Flies on the Move: An Inherited Virus Mirrors *Drosophila melanogaster's* Elusive Ecology and Demography', *Molecular Ecology*, 23: 2093–104.

Wille, M. et al. (2018). 'Virus–Virus Interactions and Host Ecology are Associated with RNA Virome Structure in Wild Birds', *Molecular Ecology*, 27: 5263–78.

Wille, M. et al. (2023). 'Strong Host Phylogenetic and Ecological Effects on Host Competency for Avian Influenza in Australian Wild Birds', *Proceedings of the Royal Society B: Biological Sciences*, 290: 20222237.

Wille, M. et al. (2019). 'Virome Heterogeneity and Connectivity in Waterfowl and Shorebird Communities', *The ISME Journal*, 13: 2603–16.

Wilson, S., and Swan, G. (2017). *A Complete Guide to Reptiles of Australia*. Reed New Holland.

Zhang, J. et al. (2018). 'Identification of a Novel Nidovirus as a Potential Cause of Large Scale Mortalities in the Endangered Bellinger River Snapping Turtle (*Myuchelys georgesi*)', *PLoS One*, 13/10: e0205209.

Zhang, Y.-Z. et al. (2018). 'The Diversity, Evolution and Origins of Vertebrate RNA Viruses', *Current Opinion in Virology*, 31: 9–16.