



ARTICLE

# Virome heterogeneity and connectivity in waterfowl and shorebird communities

Michelle Wille<sup>1</sup> · Mang Shi<sup>2</sup> · Marcel Klaassen<sup>3</sup> · Aeron C. Hurt<sup>1</sup> · Edward C. Holmes<sup>2</sup>

Received: 29 January 2019 / Revised: 2 April 2019 / Accepted: 27 May 2019  
© The Author(s) 2019. This article is published with open access

## Abstract

Models of host-microbe dynamics typically assume a single-host population infected by a single pathogen. In reality, many hosts form multi-species aggregations and may be infected with an assemblage of pathogens. We used a meta-transcriptomic approach to characterize the viromes of nine avian species in the Anseriformes (ducks) and Charadriiformes (shorebirds). This revealed the presence of 27 viral species, of which 24 were novel, including double-stranded RNA viruses (*Picobirnaviridae* and *Reoviridae*), single-stranded RNA viruses (*Astroviridae*, *Caliciviridae*, *Picornaviridae*), a retro-transcribing DNA virus (*Hepadnaviridae*), and a single-stranded DNA virus (*Parvoviridae*). These viruses comprise multi-host generalist viruses and those that are host-specific, indicative of both virome connectivity (host sharing) and heterogeneity (host specificity). Virome connectivity was apparent in two well described multi-host virus species -avian coronavirus and influenza A virus- and a novel *Rotavirus* species that were shared among some Anseriform species, while virome heterogeneity was reflected in the absence of viruses shared between Anseriformes and Charadriiformes, as well as differences in viral abundance and alpha diversity among species. Overall, we demonstrate complex virome structures across host species that co-exist in multi-species aggregations.

## Introduction

Many hosts are members of multi-species aggregations and may be infected by an assemblage of specialist and/or multi-host generalist infectious agents. Host community diversity is central to microbial dynamics [1, 2], and species richness,

relative abundance, specificity and intra- and inter-species interactions within assemblages likely have complex roles in modulating microbe levels within populations [1, 3–8]. A significant limitation in studying viral communities in hosts is that most viral species remained undescribed [9], such that viral ecology across multi-host systems has been limited to “single-virus” dynamics, particularly in vertebrate systems (for example, Influenza A virus [IAV] in avian populations). With the advent of unbiased, bulk ‘meta-transcriptomic’ RNA sequencing we can now explore, in more detail, how viral community structure may be shaped by host-species interactions.

Through long distance migration, wild birds connect the planet. Crucially, they are important reservoirs for viruses with negative consequences for wild birds (e.g. Wellfleet Bay virus; [10]), poultry (e.g. Avian avulavirus 1; [11]), and humans (e.g. IAV; [12]). Despite their importance, we know little of avian viral communities. Birds of the orders Anseriformes and Charadriiformes, the central reservoirs for notifiable avian viruses such as IAV, avian avulavirus, and avian coronavirus [13, 14], form multi-host flocks, in which many species may migrate, forage, or roost together [15], making these groups excellent models for studying virome ecology. Flocks may comprise species along a

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-019-0458-0>) contains supplementary material, which is available to authorized users.

- ✉ Michelle Wille  
michelle.wille@influenzacentre.org
- ✉ Edward C. Holmes  
edward.holmes@sydney.edu.au

- <sup>1</sup> WHO Collaborating Centre for Reference and Research on Influenza, at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia
- <sup>2</sup> Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, Australia
- <sup>3</sup> Centre for Integrative Ecology, Deakin University, Geelong, Australia

taxonomically related gradient and may utilize similar or different ecological niches in the same environment. For example, in Australia, taxonomically related dabbling Grey Teals (*Anas gracilis*) and Pacific Black Ducks (*Anas superciliosa*) may share the environment with the distantly related filter feeding Pink-eared Duck (*Malacorhynchus membranaceus*). These multi-host flocks form multi-host maintenance communities [6], with consequences for virus ecology, transmission, and virulence [1, 16, 17].

Studies of the ecology of IAV, the best studied multi-host virus in wild birds, have shown that not all hosts are equal [18, 19]. In particular, there are marked differences in susceptibility, pathology and the subsequent immune response in related species, or more diverse species within similar ecological niches. For example, dramatic differences in viral prevalence exist within the Charadriiformes, such that Ruddy Turnstones (*Arenaria interpres*) may have an IAV prevalence of ~15%, compared to the negligible IAV prevalence in co-sampled Sanderlings (*Calidris alba*) at Delaware Bay, USA [20]. There are also major differences in the pathology of highly pathogenic IAV in Anseriformes in field and experimental infections. Mallards (*Anas platyrhynchos*) infected with highly pathogenic IAV are thought to move the virus large distances and remain free of clinical signs, while Tufted Ducks (*Aythya fuligula*), in contrast, experience severe mortality [21–23]. Following IAV infection, dabbling ducks of the genus *Anas* are believed to develop poor immune memory [24], allowing life-long IAV re-infection, in contrast to swans that have long-term immune memory [25] and where re-infection probability is likely very low in adults. These differences are driven by factors encompassing both virus (e.g. virulence, transmission) and host (e.g. receptor availability, immune responses) biology [14].

The goal of this study was to use the analysis of comparative virome structures, particularly virome composition, diversity and abundance, as a means to describe host-virus interactions beyond the “one-host, one-virus” model. Given their role as hosts for multi-host viruses, we used apparently healthy members of the Anseriformes and Charadriiformes as model avian taxa in these comparisons. In particular, using samples collected from Australian birds, we aimed to (i) reveal viromes and describe novel viruses in the bird taxa sampled, (ii) determine whether viromes of different host orders have different abundance and viral diversity, (iii) determine whether closely taxonomically related and co-occurring avian hosts have viromes that are more homogenous, and (iv) identify the impact of host taxonomy, which we use as a proxy for differences in relevant host traits (such as host physiology, cell receptors), in shaping virome structure. Overall, we reveal a combination of virome heterogeneity and connectivity across species that are important reservoirs of avian viruses.

## Materials and methods

### Sample selection

Samples were collected as part of a long-term IAV surveillance study [26, 27]. Four species of shorebirds were captured using cannon nets at the Western Treatment Plant near Melbourne (37°59′11.62″S, 144°39′38.66″E) during the same sampling event in December 2012 ( $n = 434$ ): Red-necked Stint (*Calidris ruficollis*), Sharp-tailed Sandpiper (*Calidris acuminata*), Curlew Sandpiper (*Calidris ferruginea*) and Red-capped Plover (*Charadrius ruficapillus*). Waterfowl were sampled post-mortem (within 12 h) following harvest from lakes in south-west Victoria in March 2017 ( $n = 125$ ) (36°58′S, 141°05′E) (Table S1): Pacific Black Duck, Grey Teal, Australian Shelduck (*Tadorna tadornoides*), Australian Wood Duck (*Chenonetta jubata*), and Pink-eared Duck. No birds showed any signs of disease. Additional details can be found in the Supplemental Methods.

### RNA virus discovery

RNA was extracted and libraries constructed as per [28] (Table S1) and are described in the Supplemental Methods. Paired end sequencing (100 bp) of the RNA library was performed on an Illumina HiSeq 2500 platform at the Australian Genome Research Facility (AGRF, Melbourne). We used the bioinformatics pipeline reported in refs. [28–30]. All contigs were filtered to remove plant, invertebrate, fungal, bacterial, and host sequences. The virus list was further filtered to remove viruses with likely invertebrate [30], lower vertebrate [29], plant, or bacterial host associations using the Virus-Host database (<http://www.genome.jp/virushostdb/>). Hence, only those viruses that grouped within the previously defined vertebrate virus groups are identified as bird associated.

### Virus genome annotation and phylogenetic analysis

Contigs were annotated, and phylogenetic trees inferred as per ref. [28], and are described in the Supplemental Methods. Viruses with full-length genomes, or incomplete genomes but that possess the full-length RNA-dependant RNA polymerase (RdRp) gene, were used for phylogenetic analysis estimated using PhyML 3.0 [31]. Final alignment lengths for each data set are presented in Table S2. Novel viral species were identified as those that had <90% RdRp protein identity or <80% genome identity to previously described viruses. All reads have been deposited in the Short Read Archive (PRJNA505206) and viral sequences described have been deposited in GenBank (MK204384–MK20441, MK213322–MK213337).

## Diversity and abundance across libraries

Virus abundance was estimated as the proportion of the total viral reads in each library (excluding rRNA). All ecological measures were calculated using the data set comprising viruses associated with “higher” vertebrates (i.e. birds and mammals), albeit with all retroviruses and retro-transposons removed (hereafter, “avian virus data set”).

All analyses were performed using R v 3.4.0. Specifically, we calculated observed virome richness, Shannon and Shannon effective indices (i.e. alpha diversity; the diversity within each sample) for each library at the virus family and genus levels using the Rhea alpha diversity script set [32]. Observed virome richness is the number of viruses in each library, whereas Shannon effective is a weighted metric and measures how evenly the viruses are distributed in the sample. Alpha diversity indices were compared between avian orders or virus status using a linear model following a box–cox transformation aiming at achieving homoscedasticity and normality of data. In cases with significant results we also used a more conservative non-parametric test for comparison (Kruskal–Wallis rank sum test). Beta diversity (i.e. virus diversity between different libraries) was estimated as the Bray–Curtis dissimilarity matrix, which takes into account shared taxonomic composition and abundance of viromes, and was plotted as a function of nonmetric multidimensional scaling (NMDS) ordination [33]. We tested whether clustering was significant using PERMANOVA (Adonis tests). We also employed Mantel tests to assess the relationship between beta diversity and factors that may be relevant to virome structure. Beta diversity and all associated plots and statistics were calculated using the *vegan* [34] and *phyloseq* packages [35]. To determine whether differences in virome composition can be explained by host phylogeny [36], dendrograms of beta diversity were constructed using the Bray–Curtis dissimilarity matrix incorporated into the *hclust()* function. Dendrograms representing library abundance and composition at the viral family, genera, and species level were compared to host phylogeny. For the species level comparisons, we used only those viruses presented in Table S3. Two phylogenetic tree congruence metrics were then calculated to compare the match between the virome dendrogram and host phylogeny: the matching cluster Robinson–Foulds tree metric [37] using the *phangorn* package [38], and the normalized PH85 metric [39, 40] using the *ape* package [41]. For both metrics, a distance of 0 indicates complete congruence and 1, incomplete congruence. The phylogenetic relationships among the avian host species were inferred using a maximum likelihood tree of the cytochrome B mitochondrial sequences and accords with those determined previously [42–45]. The overall co-phylogenetic analysis was visualized using the *phytools* package [46].

Finally, the relative abundance of virus families between each avian host order (Charadriiformes versus Anseriformes) was assessed using the wilcox test. Subsequently, log2 relative abundance was calculated using *DESeq2* [47] implemented in *phyloseq* [35]. Given the large number of genera detected within the *Picornaviridae*, this analysis was repeated at the viral genus level in this case.

## Results

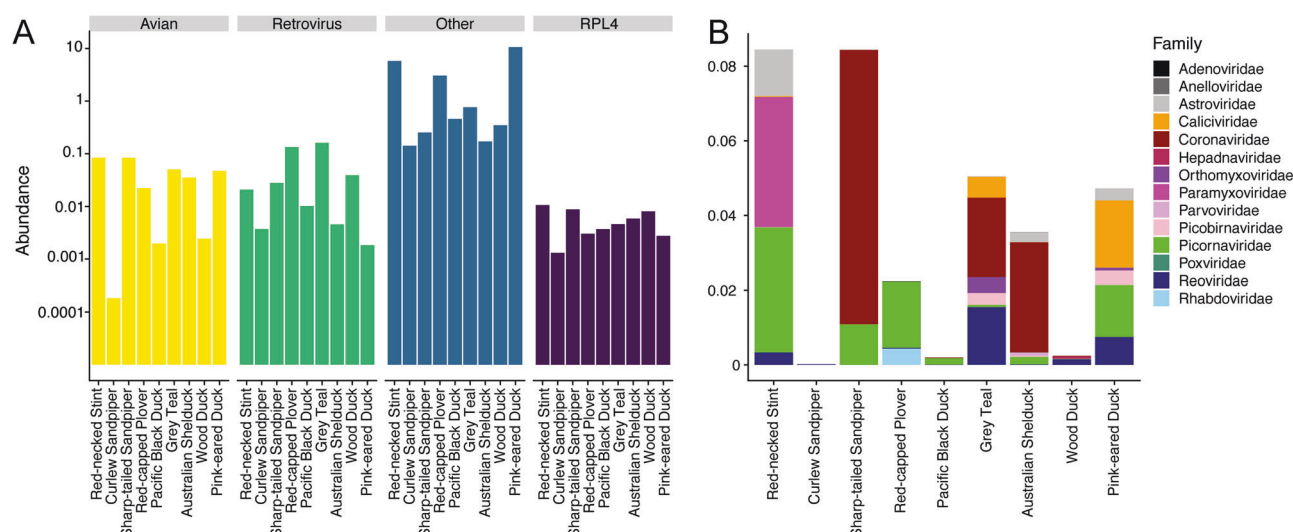
### Substantial undescribed variety of RNA viruses in wild birds

We characterized the total transcriptome of nine avian species from the Anseriformes and Charadriiformes (Table S1). Within each avian order, bird species were sampled across the same spatial and temporal scales, although members of the Anseriformes and Charadriiformes were sampled at different time points (i.e. years) and locations. Matching samples in space and time was crucial to understand the role of species within multi-host maintenance flocks with respect to virus sharing or host-specificity. This resulted in two bird communities comprising five and four species, respectively (Table S1). There was a large range in total viral abundance (0.14–10.67% viral reads) and putative avian viral abundance (0.00083–0.327% viral reads) in each library (Table S1, Fig. 1). There was no correlation between RNA concentration and avian viral abundance (Fig S1). In addition to avian viral reads, libraries had numerous reads matching arthropod viruses and retroviruses (Fig. 1). Although these retroviruses are likely bird associated, the challenge of differentiating between endogenous and exogenous retroviruses meant that they were excluded from the analysis, as were those viruses most likely associated with arthropods, plants, and bacteria.

A total of 24 of the 27 viruses identified in this study likely represent novel avian viral species (Table S3, Fig. 2, Fig S2). Novel species were identified in the double-stranded RNA viruses (*Picobirnaviridae* and *Reoviridae*, genus Rotavirus), positive-sense single-stranded RNA viruses (*Astroviridae*, *Caliciviridae*, *Picornaviridae* genus Megrivirus, Gallivirus, and unassigned genera), and both retro-transcribing (*Avihepadnaviridae*) and single-stranded DNA viruses (*Parvoviridae*).

### Novel ssRNA viruses

Two novel avastroviruses were identified. Red-necked Stint avastrovirus likely comprise a new virus “Group” as it falls basal to the Group 1 and 2 viruses in our phylogenetic analysis (Fig S3). Although no full genome Group 3 viruses



**Fig. 1** Overview of viral contigs identified in this study. **a** Host association of viral contigs identified in this study: avian, retroviruses and retrotransposons, all other hosts including lower vertebrate, invertebrate, plant, bacterial hosts, and host reference gene ribosomal protein L4 (RPL4). **b** Abundance and alpha diversity of avian viral

families identified in each library. Relative abundance and alpha diversity calculations are presented in Fig S16. Abundance and alpha diversity of viral genera and species is presented in Figs S17 and S18, respectively

exist, phylogenetic analysis of a short region of the RdRp demonstrated that this virus does not belong to Group 3 avastroviruses (Fig S4). Analysis of this short RdRp region also suggested that Red-necked Stint avastrovirus is sister to a virus detected in Swedish Mallards, indicating that this new group may be globally distributed (Fig S4). A Group 1 avastrovirus, Pink-eared Duck astrovirus was also identified, and was sister to Turkey astrovirus 2 (Fig S3, S4).

Our study further expanded an unassigned avian specific genus of the *Caliciviridae* with the identification of three new species, two from Pink-eared Ducks and one from Grey Teal. These viruses form a clade comprising Goose calicivirus, Turkey calicivirus and a calicivirus previously identified in Red-necked Avocets (*Recurvirostra novaehollandiae*) from Australia (Fig. 3, ref.[28]).

Members of the *Picornaviridae* were commonplace and complete or partial genomes were identified in almost every library sequenced (Fig. 4, Fig S5). Further, two different species of Picornaviruses were detected in the Red-necked Stint (Red-necked Stint Gallivirus and Red-necked Stint Picornavirus B-like) and Pink-eared Duck (Pink-eared Duck Megrivirus and Pink-eared Duck Picornavirus) libraries. To date, Galliviruses have only been isolated from Galliformes (chickens, turkeys, quails), so it was unexpected to identify a virus that was sister to this genus in the Red-necked Stint, although the long branch lengths involved may indicate a novel viral genus. A virus similar to Avian sapeloviruses was identified in an Australian Shelduck (Fig. 4, Fig S5). Three different Megrivirus were also identified, two from Anseriformes and one from Charadriiformes. Finally, a number of picornaviruses from

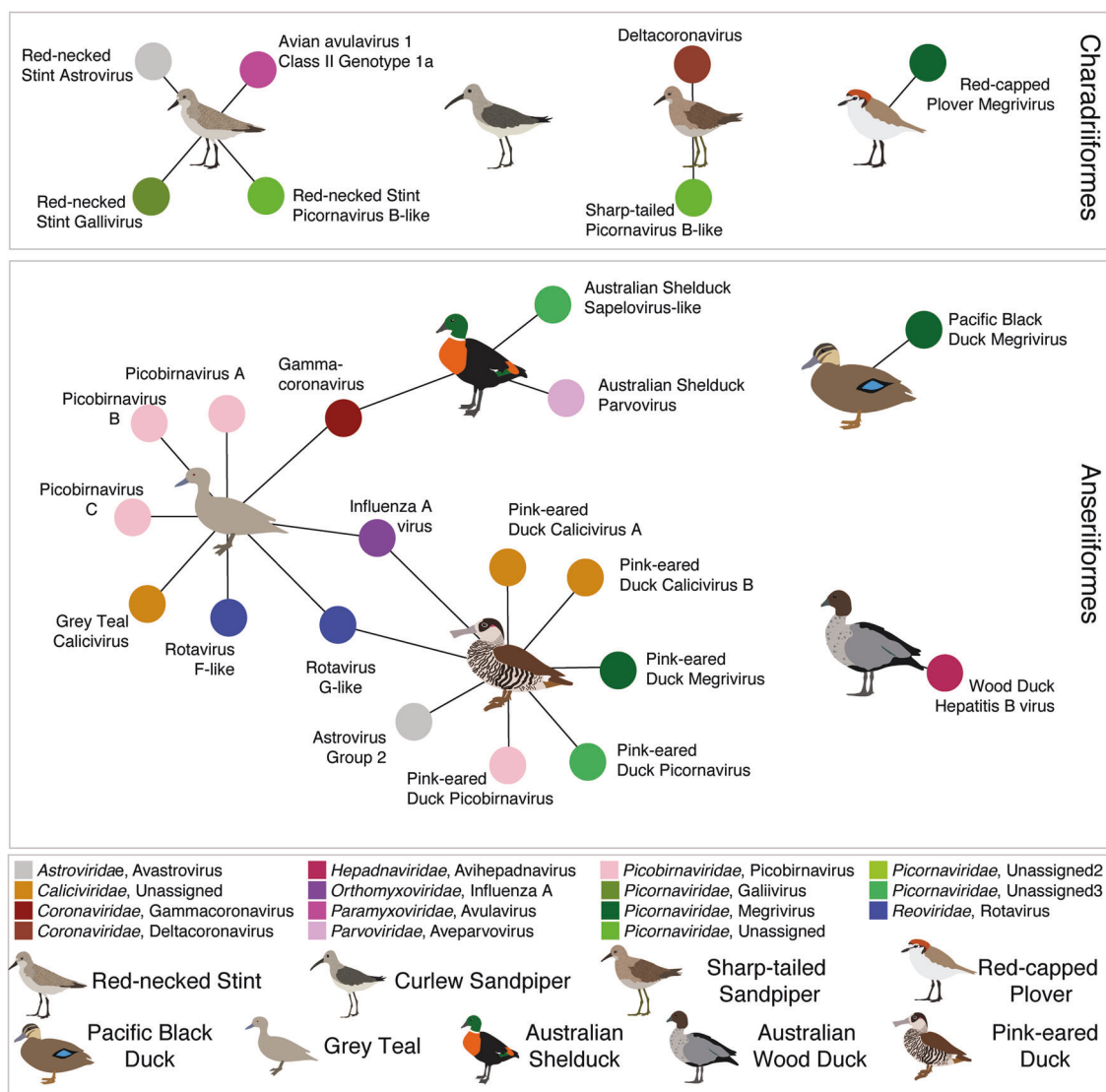
unassigned genera similar to Pigeon picornavirus B were identified in Charadriiformes (Fig. 4, Fig S5). These form a clade with a number of picornaviruses previously detected in Red-necked Avocets from Australia.

In addition to a previously described coronavirus, we identified a potentially novel species of deltacoronavirus. Specifically, Sharp-tailed Sandpiper deltacoronavirus was most closely related to deltaviruses in wild birds from the United Arab Emirates and gulls from Europe (Fig S11), although the limited number of deltacoronaviruses sequences available inhibits a detailed analysis of its geographic range.

### Novel dsRNA viruses

Three picobirnavirus species from two waterfowl species were newly identified here. Grey Teal picobirnavirus X and Pink-eared Duck picobirnavirus are members of a broad clade closely related to picobirnaviruses sampled in a number of species including domestic Turkeys (from which only very short sequences are available and hence not analysed here) (Fig. 5). Grey Teal picobirnavirus Y falls into a divergent clade largely comprised of human and porcine picobirnaviruses (and no turkey picobirnaviruses), potentially representing an interesting host-switching event (Fig. 5). However, due to limited sampling in wild birds, this virus could be related to other, currently unsampled, avian picobirnaviruses. Two novel rotavirus species were also revealed from these host species. Indeed, Grey Teals and Pink-eared Ducks shared a rotavirus species, distantly related to rotavirus G. This virus is one of three shared





**Fig. 2** Bipartite network illustrating the species for which complete viral genomes were found in each library. Each library is represented as a central node, with a pictogram of the avian species, surrounded by each viral species. Where no complete viral genomes were revealed, the pictogram is shown with no viruses. Where two libraries share a

virus species the networks between the two libraries are linked. Virus colour corresponds to virus taxonomy. A list of viruses from each library is presented in Table S3, and phylogenetic trees for each virus family and species can be found in (Figs. 3–6, Figs. S2–S14)

viruses in our entire data set. Grey Teals also carried a second rotavirus species, distantly related to rotavirus F (Fig S7).

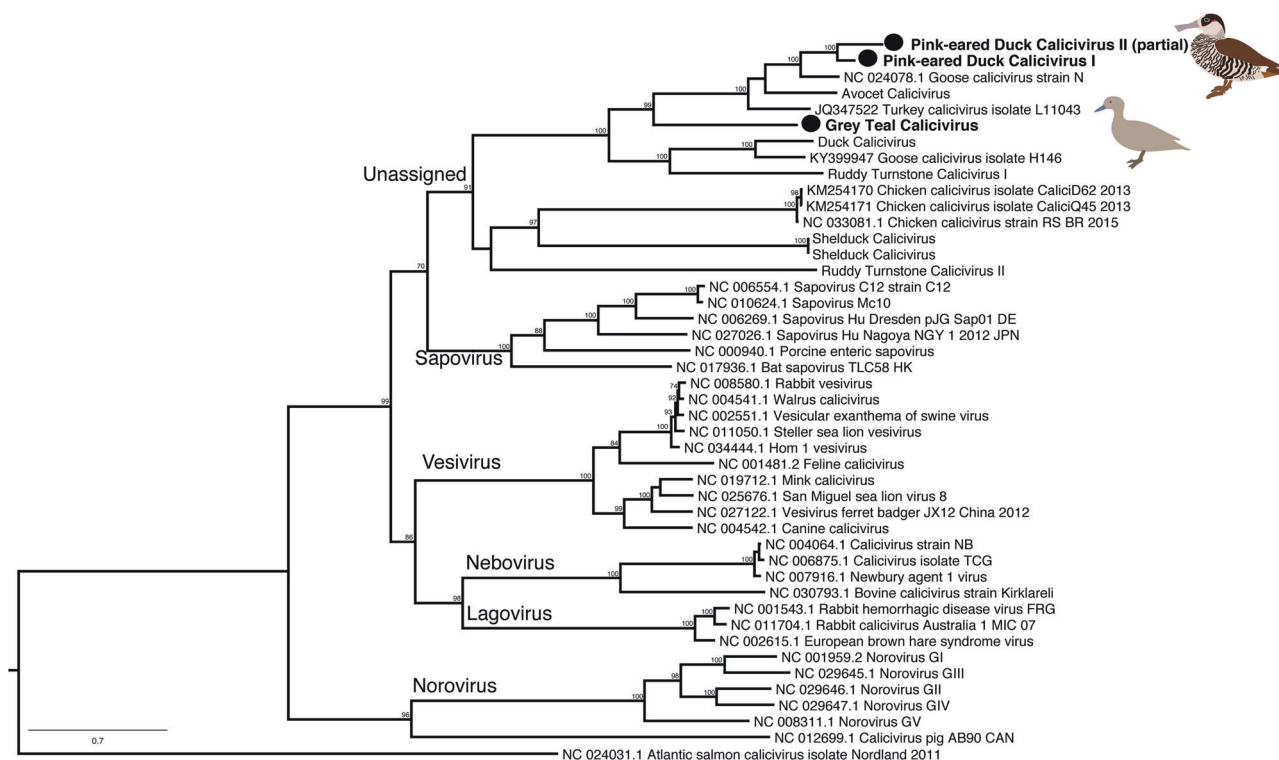
### Novel DNA viruses

These data also provided evidence for the presence of a novel single-stranded DNA virus and retro-transcribing DNA virus. An Australian Shelduck parvovirus (ssDNA) was revealed in Australian Shelducks that belongs to the highly divergent genus Chapparrivirus of the *Parvoviridae* (Fig. 6). Exogenous hepadnaviruses (retro-transcribing DNA viruses) from waterfowl are host specific, and the novel Wood Duck Hepatitis B virus identified

here is most closely related to Shelgoose Hepatitis B virus, Duck Hepatitis B virus and Snow Goose Hepatitis virus (Fig S8).

### Previously described avian RNA viruses

Given their frequency in avian populations as described in numerous surveillance schemes, we anticipated finding IAV, Avian avulavirus type 1 (formerly avian paramyxovirus type 1), and members of the *Coronaviridae*. Not only did we detect these viruses, but IAV and avian gammacoronavirus were shared across three different waterfowl libraries (Figs S9–S10). Phylogenetic analysis of a partial RdRp revealed that the avian gammacoronavirus identified



**Fig. 3** Phylogenetic tree of the virus polyprotein, including the RdRp, of representatives of the vertebrate RNA virus family *Caliciviridae*. Viruses identified in this study are denoted with a filled circle and in bold. The most divergent calicivirus, Atlantic Salmon calicivirus,

was used as outgroup to root the tree. Bootstrap values >70% are shown for key nodes. The scale bar represents the number of amino acid substitutions per site

here was most closely related to those already found in Australian wild birds (Fig S10).

We identified two subtypes of IAV—H9N1 and H3N1 (Figs S12, S13)—in Grey Teal and Pink-eared Duck, respectively. Both H9N1 and H3N1 are rarely detected subtype combinations in large waterfowl surveillance schemes [48, 49]. Segments of these two viruses generally fell into the geographically isolated “Eurasian” clade, with the exception of the NP segment that fell within the “North American” clade, thereby demonstrating intercontinental reassortment (Figs S12, S13). Finally, although avian avulavirus Type 1 Class II Genotype 1b are frequently isolated from wild birds globally, we detected a Class II Genotype 1a virus infrequently isolated in wild birds (Fig S14). This genotype has been previously isolated from Australian chickens, although the F gene cleavage site (GRQGR\*L) indicates this virus is of the low pathogenic phenotype.

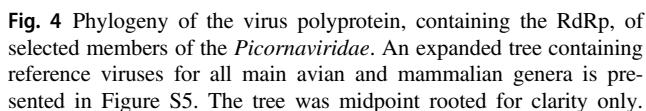
### Host heterogeneity and connectivity of avian viromes

There was a variable abundance of avian viral reads across the libraries. The highest abundance of avian viruses was in Red-necked Stint and Sharp-tailed Sandpiper, with 0.08% of reads in both cases; the lowest viral abundance was also

in this avian order (Curlew Sandpiper, 0.00018% of reads). Of the Anseriformes, Grey Teal, Australian Shelduck and Pink-eared Duck had high abundance (0.051%, 0.035%, and 0.047% of reads, respectively) while Wood Duck and Pacific Black Duck had very low abundance, albeit only one order of magnitude lower (0.0024% and 0.0019%, respectively) (Table S1).

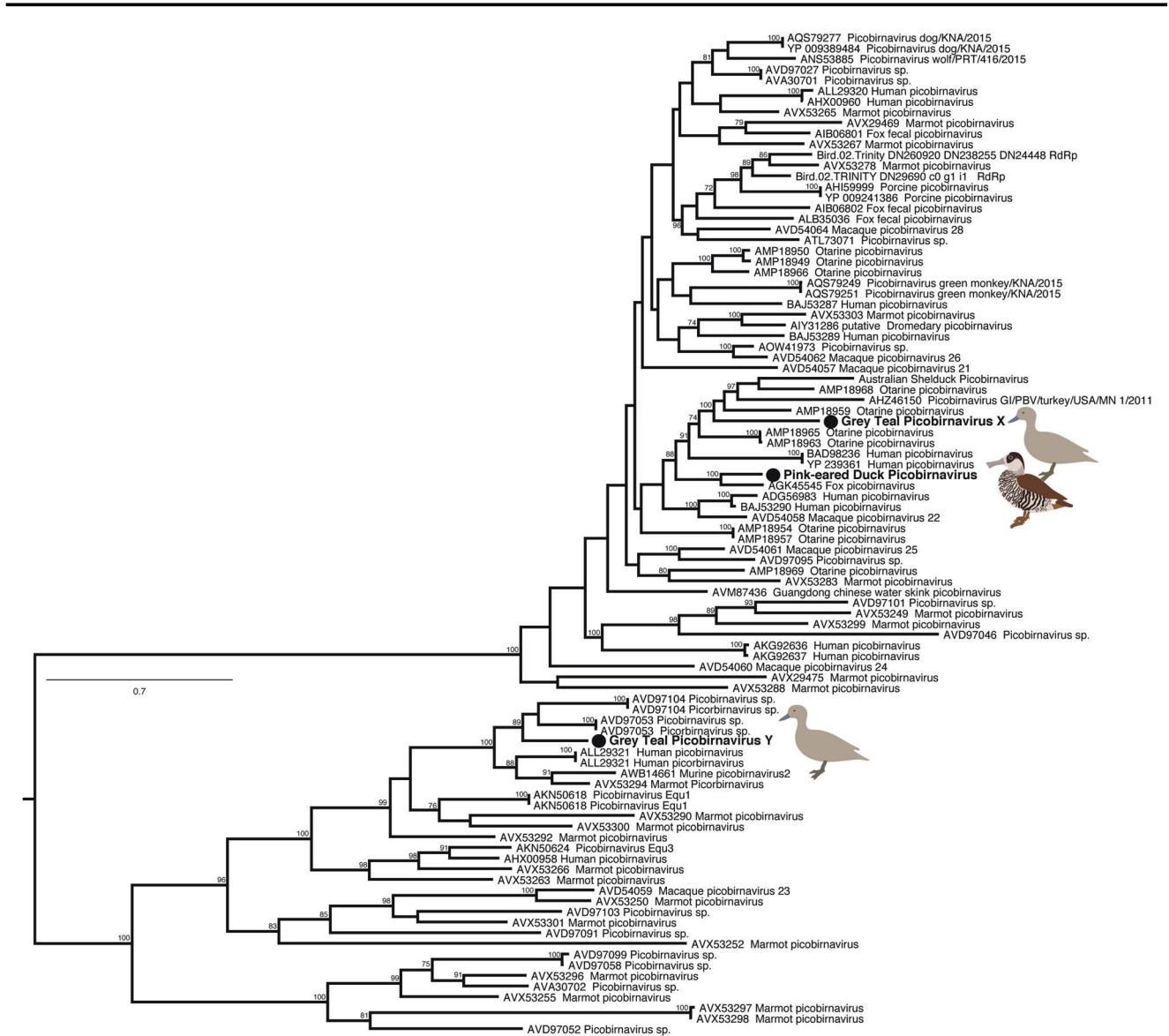
As with abundance, there was marked heterogeneity in alpha diversity indices (i.e. the diversity of viruses in each library) within the Anseriformes and Charadriiformes. Overall, while alpha diversity was very high in some Anseriiform libraries, there was no statistically significant difference between alpha diversity in Anseriiform and Charadriiform libraries at the viral family, genus, and species levels (Figs S16–S18). Of note was the surprisingly high alpha diversity in Red-necked Stint compared to other Charadriiformes, and the low alpha diversity in Pacific Black Ducks and Australian Wood Ducks in the Anseriformes (Fig. 7a, Figs S15–S18). Hence, despite sampling multispecies flocks, there can be a large variation in virome structure across species and potentially individuals.

There was also substantial variation in the viral genera and species composition within each library (Fig. 7a, Figs S15–S18). Members of the *Picornaviridae*, *Caliciviridae*, and *Reoviridae* (genus *Rotavirus*) were ubiquitous and

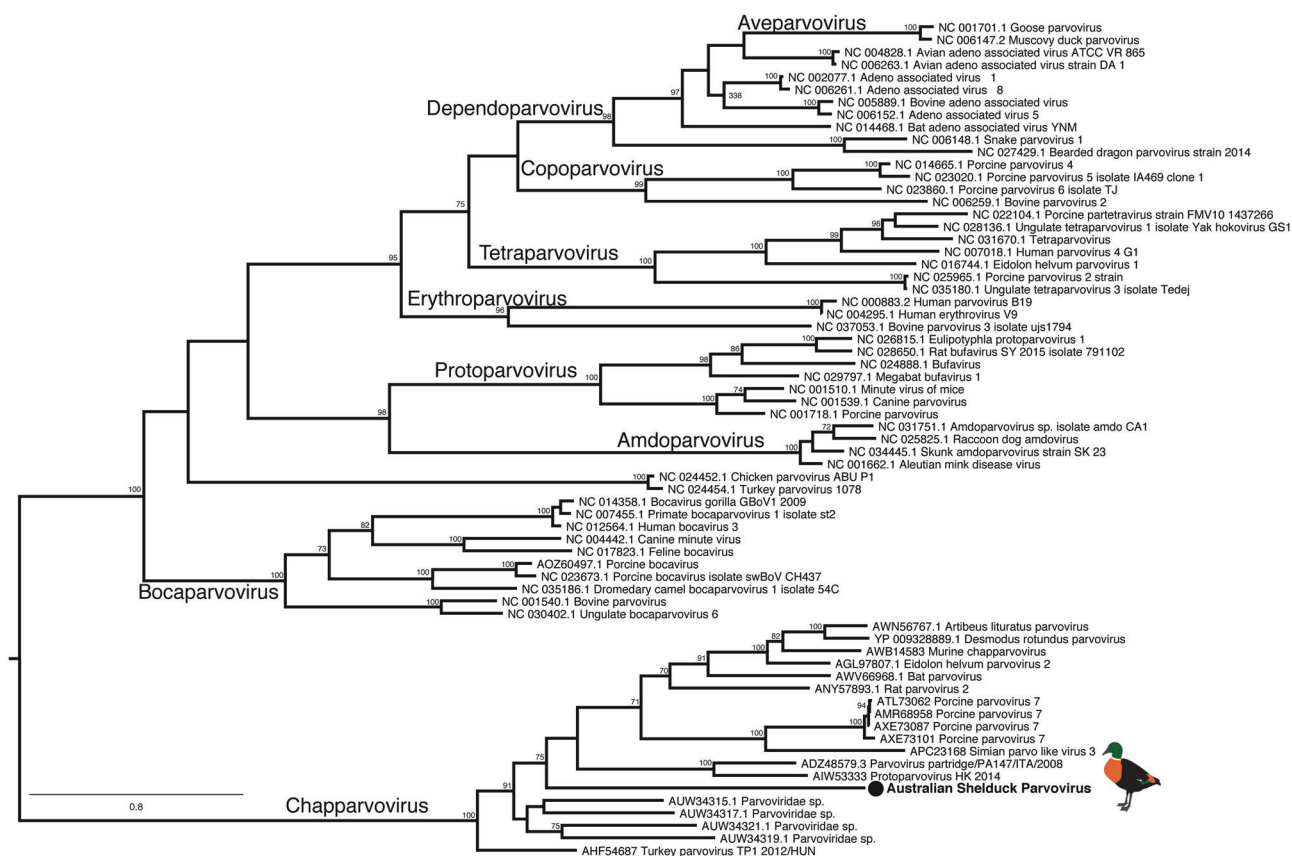


Viruses described in this study are marked in bold, adjacent to a filled circle. Bootstrap values >70% are shown for key nodes. The scale bar indicates the number of amino acid substitutions per site

Australian Shelduck shed avian gammacoronavirus at high abundance (0.021% and 0.029%, respectively) and Grey Teal and Pink-eared Duck shed IAV, although at lower abundance (0.0041% and 0.000728%, respectively) (Fig. 7a, Fig S17). Overall, there were no trends towards differential abundance of viral families between the Charadriiformes and Anseriformes (Fig S22-23). Viruses from the *Astroviridae*, *Calciviridae*, *Coronaviridae*, *Picobirnaviridae*, *Picornaviridae*, *Reoviridae*, and *Rhabdoviridae* were found in both Anseriformes and Charadriiformes.







**Fig. 6** Phylogenetic tree of the NS protein of representative *Parvovirinae* (ssDNA). The sequence generated in this study are indicated by a filled circle and are shown in bold. The tree was midpoint rooted

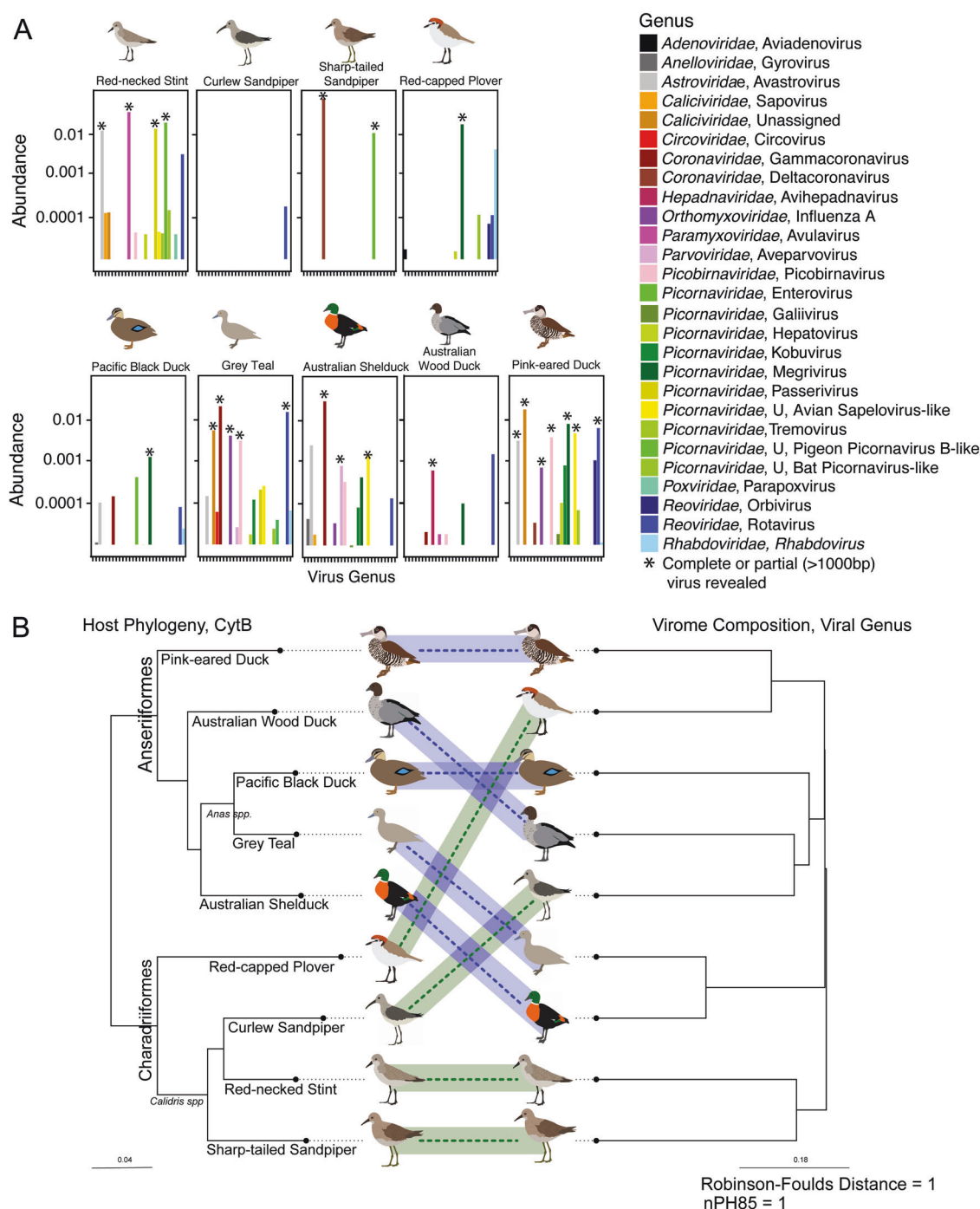
for clarity only. Bootstrap values >70% are shown for key nodes. The scale bar indicates the number of amino acid substitutions per site

connectivity of viral families and genera across species of both avian orders (Fig. 7b, Figs S25–S26). Two post hoc hypotheses of interest were whether there was clustering by feeding mechanism in the Anseriformes or migratory propensity in the Charadriiformes. There is, however, no apparent clustering on the NMDS plots, and due to small sample size we are unable to adequately assess this statistically (Fig S25).

Assessing comparative virome structure at the viral family and genera level is critical in demonstrating core viral families and genera in waterbirds. Species level analysis, albeit limited to viruses in which we were able to assemble >1000 bp, is a more accurate measure of connectivity and heterogeneity of avian viromes. First, while there was no marked division in virome composition at the level of viral family or genera between the Anseriformes and the Charadriiformes, there was such a distinction at the level of viral species. Specifically, no viral species were shared between the Anseriformes and Charadriiformes (Fig. 3), although they were sampled at different time points and at different locations. Within the Anseriformes, three viruses (IAV, avian gammacoronavirus, and duck rotavirus G-like) were shared between three libraries: Grey Teal,

Pink-eared Duck and Australian Shelduck (Fig. 3). These shared viruses were especially common in the viromes of the Grey Teal (80% of avian viral reads), Australian Shelduck (82% of avian viral reads), and a small proportion of Pink-eared Duck avian viral reads (17%). The two *Anas* ducks (Grey Teal and Pacific Black Duck), the most closely related Anseriformes, did not share any viral species; surprisingly, the virome of the Pacific-Black Duck was different from the three connected host species. Further, Grey Teal and Pink-eared Ducks, the most taxonomically distinct waterfowl, shared two viral species, demonstrating the limited impact of host phylogeny (Fig. 3). These viruses were also shared across different feeding strategies (dabbling and filter feeding), suggesting that co-occurrence was potentially responsible for their spread.

Within the Anseriformes we tested for the effect of virus–virus interactions on alpha diversity, specifically whether the presence of viruses shared across multiple libraries had an effect on virome composition. Despite a low abundance of IAV in the libraries, there was a significant difference whereby libraries containing IAV (Grey Teal, Pink-eared Duck) had a higher alpha diversity than libraries without IAV in Anseriiformes at the viral family, genus, and species



**Fig. 7** Heterogeneity and lack of taxonomic structure in avian viromes. **a** Abundance of avian viral genera identities in each library. Libraries are arranged taxonomically, with cladograms illustrating host species phylogenetic relationships within the Charadriiformes and Anseriformes. The taxonomic identification presented is that of the top blastx hit of all avian viral contigs. Asterisks indicate cases in which at least one complete or partial (>1000 bp) virus was obtained. Alpha diversity metrics are presented in Fig S17. **b** Co-phylogeny

demonstrating the discordance between host-taxonomic relationship and virome composition. Host (phylogenetic) taxonomic relationship was inferred using the mitochondrial cytochrome B gene. Virome composition dendrogram generated by clustering of bray-curtis dissimilarity matrix. The relationship between host taxonomy and virome composition was tested using two discordance metrics: Robinson-Foulds and nPH85, where 1 is discordance and 0 is agreement

levels (Richness:  $F_{1,3} = 0.65$ ,  $p = 0.479$ ;  $F_{1,3} = 12.8$ ,  $p = 0.03733$ ;  $F_{1,3} = 81.25$ ,  $p = 0.002883$ ) (Shannon Effective:  $F_{1,3} = 102.8$ ,  $p = 0.00204$ ;  $F_{1,3} = 19.537$ ,  $p = 0.04279$ ;

$F_{1,3} = 37.47$ ,  $p = 0.00763$ ) (Fig S20). Although this relationship was not significant using the more conservative non-parametric Kruskal–Wallace test (Richness:  $X^2 =$

2.6368,  $df = 1$ ,  $p$ -value = 0.1044;  $X^2 = 3.1579$ ,  $df = 1$ ,  $p$ -value = 0.07556;  $X^2 = 3$ ,  $df = 1$ ,  $p$ -value = 0.08326) (Shannon Effective:  $X^2 = 3$ ,  $df = 1$ ,  $p$ -value = 0.08326;  $X^2 = 3$ ,  $df = 1$ ,  $p$ -value = 0.08326;  $X^2 = 3$ ,  $df = 1$ ,  $p$ -value = 0.08326) (Fig S20); the significant effect of IAV on alpha diversity confirms a previous study [28]. There was, however, no significant clustering based on IAV infection on the NMDS plot (Viral Family  $R^2 = 0.14$ ,  $df = 8$ ,  $p = 0.307$ , Viral Genera,  $R^2 = 0.14$ ,  $df = 8$ ,  $p = 0.224$ ). There was no statistically significant difference in alpha diversity in libraries depending on whether gammacoronavirus was present or absent (Fig S21).

Finally, the phylogenetic analysis did not reveal a clear host-taxonomic gradient in viral species relationships. However, within the megaviruses (*Picornaviridae*), there appear to be large clades that may reflect avian order, with the viruses identified in the Anseriformes and Charadriiformes falling into two different clades. Furthermore, viruses from wild Anseriformes fall as sister taxa to previously described duck and goose megaviruses (Fig S6). In addition, this and our previous study [28] identified a number of picornaviruses from an unassigned genus only found in Charadriiformes, such that it might similarly represent a virus genus that is specific to a particular host order.

## Discussion

We identified 27 novel and previously described viral species from nine waterbirds falling into two avian orders. Anseriformes and Charadriiformes are important reservoirs for the best described avian virus, IAV, but are also central to the epidemiology of other multi-host viruses such as avian coronavirus and avian avulavirus type 1 [13, 14, 27, 50–54]. As such, these avian hosts are excellent model species for understanding the determinants of virome composition. Indeed, we detected all these previously described low pathogenic avian viruses in our sample set, and coronaviruses and IAV were shared across different Anseriform species. We also genomically described 24 novel viral species belonging to 10 viral families, including both RNA and DNA viruses. The largest number and diversity of viruses belonged to the *Picornaviridae*, although a number of rotaviruses and caliciviruses were also described.

Overall, the avian viruses identified in this study were most closely related to other avian viruses, or in genera containing avian viruses. The exception was Grey Teal picorbinavirus Y that occupies a clade dominated by viruses from human and porcine hosts. Whether this represents a host switch, or is due to lack of sampling in other hosts, will likely be revealed in additional meta-

transcriptomic studies. The Shelduck parvovirus described here is of particular interest as it is a member of the genus *Chapparrivirus*. Metagenomic analyses have recently identified members of this genus in a large number of vertebrates [55], and are known agents of severe disease [56].

Beyond viral discovery, our study revealed no predictable clustering of viromes according to host taxonomy in either the Anseriformes and/or Charadriiformes. Given the data on IAV, we might expect to see differences in virome structure due to a number of host factors [14], including differences in biology/physiology. For example, different host species have different cell receptors which in turn results in different cell and tissue tropisms and patterns of viral attachment [57]. Further, following infection, different species have differences in long-term immune memory [24, 25]. However, we saw no clear distinction between the viromes of Anseriformes or Charadriiformes based on host taxonomy, suggesting these host factors are not central to virome structuring. For example, within the Charadriiformes, the closely related *Calidris* sandpipers (Scolopacidae) did not have similar viromes and did not cluster as a group independently from Red-capped Plover, a member of a different avian family (Charadriidae). Alternatively, it is possible that aspects of host ecology, such as foraging ecology, may be more important in shaping virome composition than host taxonomy (a proxy for physiology). Specifically, differences in ecology may generate differences in virus exposure across closely related hosts [14, 58, 59]. The five Anseriform species studied here utilize three different feeding ecologies—dabbling, grazing, and filter feeding—while the four Charadriiform species have different bill lengths and forage in different layers of sediment [15]. Notably, however, there was no obvious clustering based upon feeding ecologies and this data set was too small to adequately test this hypothesis.

Central to our study was considering virome structure in the context of a multi-host and multi-virus model of virus-host interactions. Accordingly, the data generated here revealed large-scale heterogeneity in virus abundance, alpha diversity and species level composition in the nine avian species assessed in this study, and at the levels of virus family, genus, and even species. Despite this heterogeneity, there was also some connectivity (i.e. host sharing) among viromes at the levels of virus family, genus, and even species. In particular, some viral families and genera were ubiquitous in almost all avian libraries, including members of the *Picornaviridae* and *Reoviridae*. More striking was the connectivity between three avian species (Grey Teal, Australian Shelduck, and Pink-eared Duck) at the level of viral species: these hosts shared IAV, gammacoronavirus, and Duck G-like rotavirus. As these Anseriformes were sampled in the same temporal and spatial frames, such a similarity in

viromes was not unexpected despite the differences in taxonomy. However, at the level of viral species there was no host sharing of viruses between the Anseriform and Charadriiform libraries. This may be due to the physiological differences noted above or, more simply, that the ducks and wader viruses were sampled at different times and places. Despite the lack of connectivity between the Anseriformes and Charadriiformes at a viral species level, avian avulavirus 1 and deltacoronavirus detected in Red-necked Stint and Sharp-tailed Sandpiper, respectively, have been previously described in Anseriformes [13, 52], likely facilitated by the association of these birds to water. Specifically, viruses such as IAV are thought to be primarily transmitted by the faecal-oral route, in which viruses contaminate water through the faeces and birds ingest the viruses while feeding or preening [18]. Such water-borne transmission is critical to the dynamics of infection in bird communities [60]. Furthermore, aquatic habitats seemingly support a higher risk of infection as compared to terrestrial habitats [59, 61]. In support of this was the observation of lower viral diversity and abundance in the grazing Australian Wood duck, which has a more terrestrial dietary strategy compared to the other Anseriform species.

In sum, viral families and genera appeared to be readily shared among hosts, suggesting that waterbirds are key hosts for these families and genera. More importantly, our results indicate that avian viromes are largely comprised of seemingly multi-host generalist viruses (here, IAV, avian coronavirus, avian avulavirus type 1, duck rotavirus D-like) along with potential host-specific specialist viruses, which likely play a role in driving both heterogeneity and connectivity. While we found no evidence for viral species shared across avian orders, known multi-host virus species were revealed in both avian orders. Cases of clear host specificity were rare, but we speculate that Wood Duck Hepatitis B virus is likely host specific given high host specificity in this viral family [62]. In addition, the clade level structuring of Megriviruses (*Picornaviridae*), and previous report of an identical Megrivirus species found in the same avian species in very different locations [28], similarly suggests some level of host specificity in this viral genus. Of course, large-scale testing and experimental infection studies will be required to better understand the host range of these newly described species. Viral discovery efforts are imperative to better understanding factors that shape virome structure and the scope of host specificity in the avian reservoir. Importantly, we believe it is also imperative to consider multi-host, multi-virus systems in virus ecology.

**Acknowledgements** The sampling was supported by NIAID (HHSN266200700010C), ARC discovery grants (DP130101935 and DP160102146). The Melbourne WHO Collaborating Centre for Reference and Research on Influenza is supported by the Australian Department of Health. ECH is funded by an ARC Australian Laureate

Fellowship (FL170100022). We thank the Victorian Wader Study Group the logistic support from Melbourne Water, and duck hunters who kindly sampled duck carcasses. We also thank Simeon Lisovski and Marta Ferenczi for their support

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This research was conducted under approval of the Deakin University Animal Ethics Committee (permit numbers A113-2010 and B37-2013). Banding was performed under Australian Bird Banding Scheme permit (banding authority numbers 2915 and 2703). Research permits were approved by the Department of Environment, Land, Water and Planning Victoria (permit numbers 10006663 and 10005726).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ostfeld RS, Keesing F. Effects of host diversity on infectious disease. *Annu Rev Ecol Evol S.* 2012;43:157–82.
- Keesing F, Holt RD, Ostfeld RS. Effects of species diversity on disease risk. *Ecol Lett.* 2006;9:485–98.
- Johnson PT, Preston DL, Hoverman JT, LaFonte BE. Host and parasite diversity jointly control disease risk in complex communities. *Proc Natl Acad Sci USA.* 2013;110:16916–21.
- LoGiudice K, Ostfeld RS, Schmidt KA, Keesing F. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proc Natl Acad Sci USA.* 2003;100:567–71.
- Milholland MT, Castro-Arellano I, Suzan G, Garcia-Pena GE, Lee TE, Rohde RE, et al. Global diversity and distribution of hantaviruses and their hosts. *EcoHealth.* 2018;15:163–208.
- Haydon DT, Cleaveland S, Taylor LH, Laurenson MK. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerg Infect Dis.* 2002;8:1468–73.
- Altizer S, Nunn CL, Thrall PH, Gittleman JL, Antonovics J, Cunningham AA, et al. Social organization and parasite risk in mammals: Integrating theory and empirical studies. *Ann Rev Ecol Evol Syst.* 2003;34:517–47.
- Streicker DG, Fenton A, Pedersen AB. Differential sources of host species heterogeneity influence the transmission and control of multihost parasites. *Ecol Lett.* 2013;16:975–84.
- Geoghegan JL, Holmes EC. Predicting virus emergence amid evolutionary noise. *Open Biol.* 2017;7:10.



10. Ballard JR, Mickley R, Gibbs SEJ, Dwyer C, Soos C, Harms NJ, et al. Prevalence and distribution of Wellfleet Bay virus exposure in the Common Eider (*Somateria mollissima*). *J Wildl Dis*. 2017;53:81–90.
11. Alexander DJ. Newcastle disease and other avian paramyxoviruses. *Rev Sci Tech OIE*. 2000;19:443–62.
12. Gao R, Cao B, Hu Y, Feng Z, Wang D, Hu W, et al. Human infection with a novel avian-origin influenza A (H7N9) virus. *N Engl J Med*. 2013;368:1888–97.
13. Wille M, Avril A, Tolf C, Schager A, Larsson S, Borg O, et al. Temporal dynamics, diversity, and interplay in three components of the viriodiversity of a Mallard population: Influenza A virus, avian paramyxovirus and avian coronavirus. *Infect Genet Evol*. 2015;29:129–37.
14. van Dijk JGB, Verhagen JH, Wille M, Waldenström J. Host and virus ecology as determinants of influenza A virus transmission in wild birds. *Curr Opin Virol*. 2018;28:26–36.
15. Dann P. Foraging behaviour and diets of red-necked stints and curlew sandpipers in south-eastern Australia. *Wildl Res*. 2000;27:61–8.
16. Wikramaratna PS, Pybus OG, Gupta S. Contact between bird species of different lifespans can promote the emergence of highly pathogenic avian influenza strains. *Proc Natl Acad Sci USA*. 2014;111:10767–72.
17. Ren HG, Jin Y, Hu MD, Zhou J, Song T, Huang ZS, et al. Ecological dynamics of influenza A viruses: cross-species transmission and global migration. *Sci Rep*. 2016;6:36839.
18. Olsen B, Munster VJ, Wallensten A, Waldenström J, Osterhaus ADME, Fouchier RAM. Global patterns of influenza A virus in wild birds. *Science*. 2006;312:384–8.
19. Munster VJ, Baas C, Lexmond P, Waldenström J, Wallensten A, Fransson T, et al. Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog*. 2007;3:e61. <https://doi.org/10.1371/journal.ppat.0030061>.
20. Maxted AM, Luttrell MP, Goekjian VH, Brown JD, Niles LJ, Dey AD, et al. Avian influenza virus infection dynamics in shorebird hosts. *J Wildl Dis*. 2012;48:322–34.
21. Bröjer C, Ågren EO, Uhlhorn H, Bernodt K, Mörner T, Jansson DS, et al. Pathology of natural highly pathogenic avian influenza H5N1 infection in wild tufted ducks (*Aythya fuligula*). *J Vet Diagn Invest*. 2009;21:579–87.
22. Pantin-Jackwood MJ, Costa-Hurtado M, Shepherd E, DeJesus E, Smith D, Spackman E, et al. Pathogenicity and transmission of H5 and H7 highly pathogenic avian influenza viruses in Mallards. *J Virol*. 2016;90:9967–82.
23. Kim J-K, Negovetich NJ, Forrest HL, Webster RG. Ducks: the “trojan horses” of H5N1 influenza. *Influenza Other Respir Viruses*. 2009;3:121–8.
24. Magor KE. Immunoglobulin genetics and antibody responses to influenza in ducks. *Dev Comp Immunol*. 2011;35:1008–16.
25. Hill SC, Manvell RJ, Schulenburg B, Shell W, Wikramaratna PS, Perrins C, et al. Antibody responses to avian influenza viruses in wild birds broaden with age. *Proc Biol Sci*. 2016. <https://doi.org/10.1098/rspb.2016.159>.
26. Ferenczi M. Avian influenza virus dynamics in Australian wild birds. PhD Thesis: Deakin University; Geelong, Australia, 2016.
27. Ferenczi M, Beckmann C, Warner S, Loyn R, O’Riley K, Wang X, et al. Avian influenza infection dynamics under variable climatic conditions, viral prevalence is rainfall driven in waterfowl from temperate, south-east Australia. *Vet Res*. 2016;47:23.
28. Wille M, Eden JS, Shi M, Klaassen M, Hurt AC, Holmes EC. Virus-virus interactions and host ecology are associated with RNA virome structure in wild birds. *Mol Ecol*. 2018. <https://doi.org/10.1111/mec.14918>.
29. Shi M, Lin XD, Chen X, Tian JH, Chen LJ, Li K, et al. The evolutionary history of vertebrate RNA viruses. *Nature*. 2018;556:197–202.
30. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540:539–43.
31. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
32. Lagkouvardos I, Fischer S, Kumar N, Clavel T. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ*. 2017;5:e2836. <https://doi.org/10.7717/peerj.2836>.
33. Minchin PR. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*. 1987;69:89–107.
34. Oksanen J, Kindt R, Legendre P, O’Hara B, Stevens MHH, Oksanen MJ, et al. The vegan package. *Commun Ecol Package*. 2007;10:631–7.
35. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*. 2013;8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
36. Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. Phyllosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol*. 2016;14:e2000225.
37. Robinson DF, Foulds LR. Comparison of phylogenetic trees. *Math Biosci*. 1981;53:131–47.
38. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011;27:592–3.
39. Penny D, Hendy MD. The use of tree comparison metrics. *Syst Zool*. 1985;34:75–82.
40. Geoghegan JL, Duchene S, Holmes EC. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog*. 2017;13:2.
41. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20:289–90.
42. Barth JMI, Matschiner M, Robertson BC. Phylogenetic position and subspecies divergence of the Endangered New Zealand Dotterel (*Charadrius obscurus*). *Plos ONE*. 2013;8:e78068.
43. Baker AJ, Pereira SL, Paton TA. Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. *Biol Lett-Uk*. 2007;3:205–9.
44. Gibson R, Baker A. Multiple gene sequences resolve phylogenetic relationships in the shorebird suborder Scolopaci (Aves: Charadriiformes). *Mol Phylogenet Evol*. 2012;64:66–72.
45. Sun ZL, Pan T, Hu CC, Sun L, Ding HW, Wang H, et al. Rapid and recent diversification patterns in Anseriformes birds: Inferred from molecular phylogeny and diversification analyses. *Plos ONE*. 2017;12:e0184529.
46. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012;3:217–23.
47. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15:550.
48. Latorre-Margalef N, Tolf C, Grosbois V, Avril A, Bengtsson D, Wille M, et al. Long-term variation in influenza A virus prevalence and subtype diversity in a migratory Mallards in Northern Europe. *Proc Biol Sci*. 2014;281. <https://doi.org/10.1098/rspb.2014.0098>.
49. Wilcox BR, Knutsen GA, Berdeen J, Goekjian VH, Poulson R, Goyal S, et al. Influenza A viruses in ducks in Northwestern Minnesota: fine scale spatial and temporal variation in prevalence and subtype diversity. *PLoS ONE*. 2011;6:e24010. <https://doi.org/10.1371/journal.pone.0024010>.

50. Wille M, Lindqvist K, Muradrasoli S, Olsen B, Jarhult JD. Urbanization and the dynamics of RNA viruses in Mallards (*Anas platyrhynchos*). *Infect Genet Evol.* 2017;51:89–97.
51. Wille M, Muradrasoli S, Nilsson A, Jarhult JD. High prevalence and putative lineage maintenance of avian coronaviruses in Scandinavian waterfowl. *PLoS ONE.* 2016;11:e0150198. <https://doi.org/10.1371/journal.pone.0150198>.
52. Chamings A, Nelson TM, Vibin J, Wille M, Klaassen M, Alexandersen S. Detection and characterisation of coronaviruses in migratory and non-migratory Australian wild birds. *Sci Rep.* 2018;8:5980.
53. Hoque MA, Burgess GW, Cheam AL, Skerratt LF. Epidemiology of avian influenza in wild aquatic birds in a biosecurity hotspot, North Queensland, Australia. *Prev Vet Med.* 2015;118:169–81.
54. Hoque MA, Burgess GW, Karo-Karo D, Cheam AL, Skerratt LF. Monitoring of wild birds for Newcastle disease virus in north Queensland, Australia. *Prev Vet Med.* 2012;103:49–62.
55. de Souza WM, Romeiro MF, Fumagalli MJ, Modha S, de Araujo J, Queiroz LH, et al. Chapparraviruses occur in at least three vertebrate classes and have a broad biogeographic distribution. *J Gen Virol.* 2017;98:225–9.
56. Roediger B, Lee Q, Tikoo S, Cobbin JCA, Henderson JM, Jormakka M, et al. An atypical parvovirus drives chronic tubulointerstitial nephropathy and kidney fibrosis. *Cell.* 2018;175:530–43 e24.
57. Jourdain E, van Riel D, Munster V, Kuiken T, Waldenström J, Olsen B, et al. The pattern of influenza virus attachment varies among wild bird species. *PLoS ONE.* 2011. <https://doi.org/10.1371/journal.pone.0024155>.
58. Hall SR, Sivers-Becker L, Becker C, Duffy MA, Tessier AJ, Caceres CE. Eating yourself sick: transmission of disease as a function of foraging ecology. *Ecol Lett.* 2007;10:207–18.
59. Lozano GA. Optimal foraging theory—a possible role for parasites. *Oikos.* 1991;60:391–5.
60. Roche B, Lebarbenchon C, Gauthier-Clerc M, Chang CM, Thomas F, Renaud F, et al. Water-borne transmission drives avian influenza dynamics in wild birds: the case of the 2005–2006 epidemics in the Camargue area. *Infect Genet Evol.* 2009;9:800–5.
61. Hoyer BJ, Fouchier RAM, Klaassen M. Host behaviour and physiology underpin individual variation in avian influenza virus infection in migratory Bewick's Swans. *Proc Biol Sci.* 2012;279:529–34.
62. Guo H, Mason WS, Aldrich CE, Saputelli JR, Miller DS, Jilbert AR, et al. Identification and characterization of avihepadnaviruses isolated from exotic anseriformes maintained in captivity. *J Virol.* 2005;79:2729–42.