Divergent influenza-like viruses of amphibians and fish support an ancient evolutionary association

Rhys Parry[1, #, *], Michelle Wille[2, #], Olivia M. H. Turnbull[3], Jemma L. Geoghegan[4,5], Edward C. Holmes[2,*]

[1] Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4067, Australia; r.parry@uq.edu.au (R.P)

[2] Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life & Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia; michelle.wille@sydney.edu.au (M.W.); edward.holmes@sydney.edu.au (E.C.H)

[3] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; olivia.turnbull@hdr.mq.edu.au (O.M.H.T)

[4] Department of Microbiology and Immunology, University of Otago, Dunedin 9016, New Zealand; jemma.geoghegan@otago.ac.nz (J.L.G)

[5] Institute of Environmental Science and Research, Wellington 5018, New Zealand

# These authors contributed equally

* Correspondence: r.parry@uq.edu.au Tel +617 3365 3302 (R.P); edward.holmes@sydney.edu.au Tel +61 2 9351 5591 (E.C.H)

Abstract

Influenza viruses (family *Orthomyxoviridae*) infect a variety of vertebrates, including birds, humans, and other mammals. Recent metatranscriptomic studies have uncovered divergent influenza viruses in amphibians, fish and jawless vertebrates, suggesting that these viruses may be widely distributed. We sought to identify additional vertebrate influenza-like viruses through the analysis of publicly available RNA sequencing data. Accordingly, by data mining, we identified the complete coding segments of five divergent vertebrate influenza-like viruses. Three fell as sister lineages to influenza B virus: salamander influenza-like virus in Mexican walking fish (*Ambystoma mexicanum)* and plateau tiger salamander (*Ambystoma velasci*), siamese algae-eater influenza-like virus in siamese algae-eater fish (*Gyrinocheilus aymonieri*) and chum salmon influenza-like virus in chum salmon (*Oncorhynchus keta*). Similarly, we identified two influenza-like viruses of amphibians that fell as sister lineages to influenza D virus: cane toad influenza-like virus and the ornate chorus frog influenza-like virus, in the cane toad (*Rhinella marina)* and ornate chorus frog (*Microhyla fissipes)*, respectively. Despite their divergent phylogenetic positions, these viruses retained segment conservation and splicing consistent with transcriptional regulation in influenza B and influenza D viruses, and were detected in respiratory tissues. These data suggest that influenza viruses have been associated with vertebrates for their entire evolutionary history.

Keywords: *Orthomyxoviridae*; influenza; metatranscriptomics; fish; amphibians; evolution; phylogeny

1. Introduction

Influenza viruses are multi-segmented, negative-sense RNA viruses of the family *Orthomyxoviridae*, which also includes the genera *Quaranjavirus*, *Isavirus* and *Thogotovirus*, as well as a wide diversity of orthomyxo-like viruses found in invertebrates [1-3]. Until recently, influenza viruses had been limited to a small group of pathogens with public health importance for humans, particularly those of pandemic and epidemic potential – influenza A virus (IAV, *Alphainfluenzavirus*) and influenza B virus (IBV, *Betainfluenzavirus*). Influenza A virus is an important, multi-host virus which is best described in humans, pigs, horses and birds [4-7]. These viruses are epidemic in humans and are closely monitored for pandemic potential. Influenza B viruses similarly cause yearly epidemics in humans [8], and their detection in both seals [9] and pigs [10] highlights the potential for non-human IBV reservoirs.

While IAV and IBV are subject to extensive research, far less is known about two additional influenza viruses: Influenza C virus (ICV, *Gammainfluenzavirus*) and Influenza D virus (IDV, *Deltainfluenzavirus*). Influenza C virus has been found in humans worldwide and is associated with mild clinical outcomes [11, 12]. Based on its detection in pigs, dogs and camels, it is likely that this virus has a zoonotic origin [13-17]. Influenza D virus, first identified in 2015 [18], is currently thought to have a limited host range and is primarily associated with cattle and small ruminants [18-20], pigs [18] as well as dromedary camels [17]. Although there is no record of active infection in humans, there are indications of past infection based on serological studies [12, 21, 22].

Despite intensive research on the characterisation of influenza viruses, our understanding of their true diversity and their evolutionary history is undoubtedly limited. Of critical importance are recent meta-transcriptomic (i.e. bulk RNA sequencing) studies that report the discovery of novel "influenza-like" viruses from both amphibian and fish hosts – Wuhan spiny eel influenza virus, Wuhan asiatic toad influenza virus and Wenling hagfish influenza virus [3]. Such a phylogenetic pattern, particularly the position of the hagfish (jawless vertebrate) influenza virus as the sister-group to the other vertebrate influenza viruses, is compatible with virus-host co-divergence over millions of years. This, combined with relatively frequent cross-species transmission, suggests that a very large number of animal influenza-like viruses remain to be discovered.

Given the previous success of data-mining to identify a range of novel viruses, including rhabdoviruses [23], flaviviruses [24] and parvoviruses [25], we hypothesised we could mine publicly

available transcriptome databases for evidence of undescribed vertebrate influenza viruses and that this would provide further insights into their evolutionary origins. Given the highly under-sampled nature of "lower" vertebrate hosts, such as fish and amphibians, we focused on data from these taxa. From this, we identified five novel influenza-like viruses that provide new insights into the evolution and host range of this important group of animal viruses.

2. Materials and Methods

2.1 Identification of divergent influenza-like viruses from vertebrate de novo transcriptome assemblies

To identify novel and potentially divergent vertebrate influenza viruses we screened *de novo* transcriptome assemblies available at the National Center for Biotechnology Information (NCBI) Transcriptome Shotgun Assembly (TSA) Database (https://www.ncbi.nlm.nih.gov/genbank/tsa/) and the China National GeneBank (CNGB) Fish-T1K (Transcriptomes of 1,000 Fishes) Project database (https://db.cngb.org) [26]. Amino acid sequences of the influenza virus reference strains - IAV A/Puerto Rico/8/34 (H1N1), IBV (B/Lee/1940), ICV (C/Ann Arbor/1/50) and IDV (D/bovine/France/2986/2012) - were queried against the assemblies using the translated Basic Local Alignment Search Tool (tBLASTn) algorithm under default scoring parameters and the BLOSUM45 matrix. For the TSA, we restricted the search to the Vertebrata (taxonomic identifier [taxid] 7742). Putative influenza-like virus contigs were subsequently queried using BLASTx against the non-redundant virus database. Two fish transcriptomes (BioProjects PRJNA329073 and PRJNA359138) were excluded because of the presence of reads that were near identical to IAV, strongly suggestive of contamination.

2.2 Recovery of additional influenza-like virus segments through de novo assembly and coverage statistics

Putative influenza-like viruses identified in the transcriptome data were queried against genus-wide RNA-Seq samples deposited in the Sequence Read Archive (SRA) using the BLASTn tool. Raw fastq files originating from transcriptome sequencing libraries were downloaded and imported to the Galaxy Australia web server (https://usegalaxy.org.au/, v. 19.05). Sequencing adapters were identified using FastQC**,** and reads were quality trimmed using Trimmomatic (Galaxy v. 0.36.4) under the following conditions: sliding window = 4 and average quality = 20 [27]. For the recovery of full-length transcripts corresponding to putative influenza-like viruses, clean reads were then *de novo* assembled using Trinity (Galaxy Version 2.9.1) [28]. For virus genome statistics, clean reads were re-mapped

against virus segments using the Burrow-Wheeler Aligner (BWA-MEM Galaxy Version 0.7.17.1) under default conditions, and the resultant binary alignment file was analysed with the bedtools (v2.27.1) genome coverage tool [29]. For the production of the salamander influenza-virus infection heat map of *Ambystoma velasci* tissues, quantitation of abundance for each library was calculated using the proportion of total mapped salamander influenza-like reads for each library and heatmaps were produced using R v3.5.3 integrated in RStudio v1.1.463 and *ggplot2*.

2.3 Influenza virus genome annotation

Viral open reading frames were predicted using ORFFinder (https://www.ncbi.nlm.nih.gov/orffinder/). To characterise functional domains, predicted protein sequences were subjected to a domain-based search using the Conserved Domain Database version 3.16 (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and cross-referenced with the PFam database (version 32.0) hosted at (http://pfam.xfam.org/). Transmembrane topology prediction using the TMHMM Server version 2.0 (www.cbs.dtu.dk/services/TMHMM/). For the identification of potential splice sites in influenza-like viruses we used alternative isoforms assembled using Trinity and then manually validated for donor and acceptor sites using the Alternative Splice Site Predictor (ASSP) (http://wangcomputing.com/assp/index.html) [30] for NS and M segments guided by experimentally validated positions in IAV/IBV/ICV (Reviewed here [31]). Viral genome sequences have been deposited in GenBank under the accession numbers: MT926372-MT926409

2.4 Phylogenetic analysis

Multiple sequence alignments of predicted influenza virus protein sequences were performed using MAFFT-L-INS-i (version 7.471) [32]. Ambiguously aligned regions were removed using TrimAl (v. 1.3) under the automated 1 method [33]. Individual protein alignments were then analysed to determine the best-fit model of amino acid substitution according to the Bayesian Information Criterion using ModelFinder [34] incorporated in IQ-TREE (Version 2.1.1) [35] and excluding the invariant sites parameter. For the PB1, PA, NP segments the Le-Gascuel (LG) model [36] with discrete gamma model with 4 rate categories (+□$_4$) and empirical amino acid frequencies (+F) was selected as the most suitable. For the HA/HEF alignments, the Whelan and Goldman (WAG) model [37] with a discrete gamma model with 4 rate categories (+G_4) was selected. For the NA alignment, the FLU+G_4 model was identified [38], and finally, PB2 and NS (LG+G_4). For the M1/M2 alignment, LG + FreeRate model (+R3) [39] was selected. Maximum likelihood trees were then inferred using IQ-TREE (Version

2.1.1) with Ultrafast bootstrap approximation (n=40,000) [40]. Resultant consensus tree from combined bootstrap trees were visualised using FigTree version 1.4 (http://tree.bio.ed.ac.uk/software/figtree/).

3. Results

3.1 Discovery and annotation of novel influenza-like viruses in vertebrates

We initially screened 776 vertebrate *de novo* assembled transcriptomes deposited in the Transcriptome Shotgun Assembly Sequence Database (TSA) and a further 158 transcriptomes available on the Fish-T1K Project. Within the TSA entries, the most abundant assemblies were from the bony fish (*Actinopterygii)*, accounting for 346 transcriptomes. We also screened the amphibian (n=57) and cartilaginous fish (n=9) transcriptomes. From these libraries, we detected influenza-like viruses in 0.6% of the bony fish libraries (n=2) and in 7% (n=4) of the amphibian libraries.

Our transcriptome mining identified fragments of divergent IBV-like Polymerase Basic 1 and 2 (PB1, PB2) segments from an unpublished transcriptome of ear and neuromast samples from an *Ambystoma mexicanum* (Mexican walking fish or Axolotl) laboratory colony housed by The University of Iowa (BioProject Accession: PRJNA480225). BLASTx analysis of these IBV-like fragments revealed 76.32% amino acid identity to the PB1 gene of IBV (Top hit Genbank accession ANW79211.1; E-value: 0; query cover 100%) and 60.94% to the IBV PB2 (Top hit Genbank accession: AGX15674.1; E-value: 2e-49; query cover: 83%) (Table 1). To recover additional segments and full-length PB1 and PB2 genes, we used BLASTn to screen 2,037 *Ambystoma* RNA-Seq libraries from several *Ambystoma* species available on the SRA. Accordingly, we were able to identify this tentatively named salamander influenza-like virus in 139/2037 (6.8%) of screened libraries. BLASTn positive libraries were subsequently *de novo* assembled, allowing the recovery of all eight segments of this virus (Figure 1; Table 1). In addition to *A. mexicanum* samples infected with salamander influenza-like virus we identified a smaller number of reads from Anderson's salamander (*Ambystoma andersoni*) and spotted salamander (*Ambystoma maculatum*) libraries [42]. Finally, we identified a variant of salamander influenza virus from RNA-Seq data generated from laboratory colonies of another neotenic salamander, plateau tiger salamander (*Ambystoma velasci*) [41] with nucleotide sequence identity between both variants of all segments between 89-94% at the nucleotide level (PB1 90%, PB2 91%, PA 90%, NP 91%, NA 90%, HA 89%, NS 94%, M 93%) (Table 1). The presence of infection in up to four species of *Ambystoma* colonies indicates the potential for a wide host range for this virus across this host genus. Additionally, as the majority of *A. mexicanum* colonies originate from

the *Ambystoma* Genetic Stock Center (AGSC, University of Kentucky, KY) [43-46] and in some cases co-housed together with other species positive for this virus [42], this suggests that this virus may circulate between numerous laboratory-reared salamander colonies.

We similarly identified an influenza virus in the Siamese algae-eater (*Gyrinocheilus aymonieri*) (SRA accession: SRR5997773), tentatively termed siamese algae-eater influenza-like virus, sampled as part of the Fish-T1K Project (Table 1) [26]. While there is limited meta-data available, the library was prepared from gill tissues. There was no evidence of this virus in any other siamese algae-eater samples available on the SRA (n=2) through BLASTn analysis.

Finally, we identified another IBV-like virus in chum salmon (*Oncorhynchus keta)*, provisionally termed chum salmon influenza-like virus (Table 1). While the geographic origin of the samples is unknown the meta-data from this library indicates the RNA-Seq originates from gill tissues of alevin stage chum salmon, transferred to the Korea Fisheries Resources Agency (FIRA) laboratory and reared in tanks with recirculating freshwater [47]. We identified chum salmon influenza-like virus sequences in one of the three libraries from this project (SRA Accession: SRR6998471). However, all segments were convincing *de novo* assembled and covered through re-mapping with reads corresponding to 0.045% of the library (45,151/100,244,990 mapped reads).

In addition to novel viruses related to IBV, we identified two influenza viruses of amphibians that exhibited high pairwise amino acid identity to segments from IDV and ICV (Table 1). This result is particularly noteworthy as the previously identified amphibian influenza virus, wuhan asiatic toad influenza virus, was more closely related to IBV [3]. Cane toad influenza-like virus was identified in four larval samples of cane toad (*Rhinella marina*), from two geographic locations in Australia, Innisfail, Queensland, Australia (SRA accessions: SRR5446725, SRR5446726) and Oombulgurri, Western Australia, Australia (SRA accessions: SRR5446727, SRR5446728) [48]. We did not identify any virus reads in any adult tissue samples taken in Australia nor from libraries constructed from cane toad samples from Macapa city, Brazil, from the same study [48]. Similarly, we did not recover any cane toad influenza-like virus reads in any of the 16 liver tissue samples from Australia reported in another metatranscriptomic study, which aimed to reveal the virome of the cane toad [49].

The final IDV-related virus identified was tentatively named ornate chorus frog influenza-like virus detected in ornate chorus frog (*Microhyla fissipes*) (Table 1). This virus was identified in two

sequencing projects, both of which were from samples of ornate chorus frog collected from paddy fields in Chengdu, China, albeit at different times; May 2014 (Bioproject accession: PRJNA295354) [50] and June 2016 (Bioproject accession: PRJNA386601)[51]. Between both data sets, the largest number of ornate chorus frog influenza-like virus reads originated from the lung tissue of the stage 28 tadpole (SRA accession: SRR5557878) in the second study (June 2016 [51]), representing 0.13% of all reads (81,860/ 61,952,548). While there were fewer reads mapping to the ornate chorus frog influenza-like virus in whole tissues of three developmental stages in the study carried out in 2014 [50]. In the whole organism samples the highest viral abundance was in the metamorphic climax (1,462/59,993,514 reads) (SRA accession: SRR2418623), and pre-metamorphosis developmental phases (1,113/63,073,925 reads) (SRA accession: SRR2418554) with only 7 reads were identified in the complete metamorphosis library (7/56,000,000 reads) (SRA accession: SRR2418812). The nucleotide identity of ornate chorus frog influenza-like virus variants between the 2014 [50] and June 2016 [51] data sets was between 93.36%-97.10% for all segments (PB1 96.98% PB2 93.95% P3 93.36% NP 93.83% HEF 91.87% NS 97.10% M 96.29%).

3.2 The genome organisation and transcription of novel influenza-like virus genes are highly conserved

Through a combination of *de novo* assembly and tBLASTn analysis we were able to identify and assemble the complete coding segments of all novel influenza-like genomes described in this study (Figure 1, Figure S1-S5, Table S1-S6). For the salamander influenza-like virus, siamese algae eater influenza-like virus and chum salmon influenza-like viruses, all eight segments with genome arrangements similar to that of IAV and IBV were identified (Figure 1). Prediction of the ORFs from all segments and protein domain analyses suggested homology between these putative virus proteins (Table S1-S4, Figure S1-S3). Additionally, we identified the glycoprotein NB, which is encoded through a polycistronic mRNA of the NA segment of IBVs (Figure 1). The M2 domain of chum salmon influenza-like was not identified through BLASTp, likely due to high levels of sequence divergence. However, we did identify the N terminus using a domain search (Table S3). In contrast, we identified seven segments for each of cane toad influenza-like virus and ornate chorus frog influenza-like viruses, which are characteristic of ICV/IDV encoding 7 segments (Figure 1). We recovered all ORFs expected of viruses similar to ICV and IDV (Table S5-S6, Figure S4-S5).

There were a number of important differences in the composition of segments among IAV/IBV and ICV/IDV. For example, segment 3 is differentiated into “PA” or “P3” by isoelectric points. The

isoelectric point of ornate chorus frog influenza-like virus (~6.3) and cane toad influenza-like virus (~6) are similar to those of the P3 segment of ICV/IDV, while those of salamander influenza-like virus (~5.6), Siamese algae eater influenza-like virus (~5.3) and chum salmon influenza-like viruses (~5.4) had isoelectric points more consistent with PA of IAV (~5.4) and IBV (~5.5). With the exception of chum salmon influenza-like virus, we were able to assemble two discrete isoforms of the NS gene in all our assembles and manually validate the splice junctions as *bona fide* through the identification of splicing donor and acceptor sites (Table S2-6). The M segment in ICV/IDV related viruses is also post-transcriptionally spliced to encode the M1 protein, and we were able to assemble two isoforms of the M segment from cane toad influenza-like virus and ornate chorus frog influenza-like virus.

3.3 Phylogenetic analysis of novel influenza-like viruses suggests a history of genomic reassortment

One of the most striking observations of our study was that three novel influenza viruses from amphibians and fish fell basal to IBV in the phylogenetic analysis, although none of the relevant bootstrap values were exceptionally high (i.e. <75%) (Figure 2), and with pairwise amino acid sequences similarities to the PB1 of IBV ranging from 67- 75%. The placement of these viruses, in addition to Wuhan spiny eel influenza virus, at the base of the IBV lineage, strongly suggests that these viruses have a long evolutionary history in vertebrates. Indeed, it is notable that salamander influenza-like virus is the now the closest animal relative to IBV. Similarly, two other influenza-like viruses from amphibians fell basal to IDV, this time with strong (>90% bootstrap support) and exhibited 77-80% amino acid similarity in the PB1 segment.

These novel influenza viruses largely occupy consistent phylogenetic placements across segments (Figure 3). However, chum salmon influenza-like virus is the sister-group to IBV in the polymerase (PB2, PB1, PA) and NA segments, but is the sister-group to both IAV/IBV in the NP segment. Also of note was that the two amphibian influenza-like genomes are the sister lineages to IDV in all segments with the exception of the HEF, in which they are the sister lineages to the clade containing ICV/IDV.

3.4 Multi-tissue RNA-sequencing of developmental Plateau tiger salamander libraries suggests conserved influenza virus tropism in vertebrates.

To gain further insight into the tissue tropism of these divergent influenza viruses, we screened libraries from a multi-tissue and developmental library of the plateau tiger salamander. This data set comprised of 86 individual samples from gills, hearts and lungs at different stages of

premetamorphosis, prometamorphosis, metamorphosis and metamorphosis late [41]. We identified RNA originating from salamander influenza-like virus abundantly in the libraries originating from gills at the metamorphosis late stage with an average number of the reads in all 6 libraries corresponding to approximately 0.22% of all reads (Figure 4). The second most abundant tissue and developmental stage was the postmetamorphosis stage lung tissue which had, on average, 0.013% of all reads correspond to salamander influenza-like virus. The only other tissue and developmental stage with all segments represented were the gills from the premetamorphosis stage with an average ~0.003% of all reads originating from Salamander influenza-like virus.

4. Discussion

We have identified five novel influenza viruses in fish and amphibians through mining publicly available meta-transcriptomic data. Despite intensive research on influenza viruses for almost a century, it is only recently that these viruses have been identified in hosts other than birds and mammals. In particular, the recent identification of divergent influenza-like viruses in fish and amphibians [3], as well as even more divergent orthomyxoviruses in invertebrates [2], suggests that divergent members of this virus family may infect a wide range of animal hosts.

These data provide valuable insights into the evolution of influenza viruses across the vertebrates. First, that the phylogeny of the IBV-like viruses in part follows the phylogeny of the hosts from which they are sampled, with the Salamander influenza-like viruses falling as the sister-group to the mammalian viruses and the fish viruses occupying basal positions in that group as a whole. This is compatible with a process of virus-host co-divergence that likely extends hundreds of millions of years, albeit with relatively frequent host-jumping. Under these circumstances, it must also be the case that there are many more vertebrates carry IBV-like viruses that have yet to be investigated. Similarly, we identified two amphibian viruses that fall as sister lineages to IDV, again suggestive of both long-term co-divergence and highly limited sampling to date. Whether the same pattern will also be true of the influenza A virus group is unclear and will only be resolved with additional sampling. Finally, it is notable that Wenling hagfish virus, from a jawless vertebrate (a hagfish), occupies the basal position in all of the vertebrate influenza virus, again suggesting that these viruses have been in existence for hundreds of millions of years.

Despite detecting divergent influenza viruses, the stability and conservation of genome segments and transcriptional regulation of viral genes through splicing is conserved. For example, salamander influenza-like virus, siamese algae-eater influenza-like and chum salmon influenza-like virus, which all fall basal to IBV, possess eight segments encoding all the predicted/required ORFs including NB, which is a glycoprotein encoded on the NA gene of IBV only [52]. Similarly, cane toad influenza and ornate chorus frog influenza had only seven segments and all ORFs consistent with ICV and IDV. Overall, despite finding these viruses in non-mammalian hosts, the viral structure is highly consistent with other influenza viruses strongly supporting the inclusion into these viral genera.

These new data also raise questions of whether influenza and influenza-like viruses have the same tissue tropism among all vertebrate hosts. As with other influenza viruses, we detected all novel influenza viruses in respiratory tissues (i.e. gills and lungs) of their hosts: the only exception was cane toad influenza-like for which we don't have tissue-associated metadata. The most robust support for the conservation of influenza-like virus infection in respiratory tissues was in the analysis of the multi-tissue transcriptome of the plateau tiger salamander [41]. Before and during metamorphosis, salamanders rely on gills for respiration. During metamorphosis, there is a significant reduction in gill length, and in post-metamorphosis they rely on lungs for respiration [41, 53]. Based on the expectation that we would find these viruses in respiratory tissues, we found salamander influenza-like virus in the gills of metamorphosising salamanders and the lungs following metamorphosis. This suggests that all influenza viruses primarily infect the respiratory tissues, consistent with other influenza viruses [54]. To date, the only exception are IAV infections which infect the surface epithelium of the gastrointestinal tract in wild birds [55], and gut-associated lymphoid tissue and the squamous epithelium of the palatine tonsils in bats [56].

Finally, it is interesting to speculate on whether these influenza-like viruses are host specialist or generalist viruses. Influenza A-D viruses have extensive host ranges, including an array of mammalian species, particularly those associated with food production (*i.e.* cattle and swine), and in the case of IAV, wild and food production associated birds (*i.e.* poultry) [4-7, 18-20]. Salamander influenza-like virus also appears to be a multi-host virus, being found in four salamander species raised in laboratory colonies, with the greatest abundance in the Mexican walking fish and plateau tiger salamander. While these species have very different life-history strategies, and their ranges do not overlap in nature [57, 58], both are abundant in the pet trade and used in laboratory research. Thus, breeding facilities may be an important source of these viruses. Whether this multi-host trait is correlated with the use of breeding

facilities in salamanders, a parallel for animal production systems and influenza A-D viruses is unclear. Further research is certainly warranted to determine if a broad host range is a feature of all influenza viruses, including those found in "lower" vertebrates.

5. Conclusions

By mining transcriptome data we identified five novel influenza-like viruses in fish and amphibians. These data strongly suggest that influenza-like viruses can infect diverse classes of vertebrates and that influenza viruses have been associated with vertebrates for perhaps their entire evolutionary history. While the public health and economic ramifications are undoubtedly more significant for mammalian influenza viruses, only by looking at other vertebrate classes and other animal lineages will we fully understand the origins and evolution of this hugely important group of viruses.

Funding: R.P. is supported by a University of Queensland scholarship. E.C.H. is funded by an Australian Research Council Australian Laureate Fellowship (FL170100022). M.W. is supported by an Australian Research Council Discovery Early Career Researcher Award (DE200100977). J.G. and E.C.H. are also funded by an Australian Research Council Discovery Project (DP200102351).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Wille, M.; Holmes, E. C., The ecology and evolution of influenza viruses. *Cold Spring Harb Perspect Med* 2019. 10(7), a038489. doi: 10.1101/cshperspect.a038489
2. Shi, M.; Lin, X. D.; Tian, J. H.; Chen, L. J.; Chen, X.; Li, C. X.; Qin, X. C.; Li, J.; Cao, J. P.; Eden, J. S.; Buchmann, J.; Wang, W.; Xu, J.; Holmes, E. C.; Zhang, Y. Z., Redefining the invertebrate RNA virosphere. *Nature* 2016, 540, (7634), 539-543.

3. Shi, M.; Lin, X. D.; Chen, X.; Tian, J. H.; Chen, L. J.; Li, K.; Wang, W.; Eden, J. S.; Shen, J. J.; Liu, L.; Holmes, E. C.; Zhang, Y. Z., The evolutionary history of vertebrate RNA viruses. *Nature* 2018, 556, (7700), 197-202.
4. Olsen, B.; Munster, V. J.; Wallensten, A.; Waldenstrom, J.; Osterhaus, A. D.; Fouchier, R. A., Global patterns of influenza a virus in wild birds. *Science* 2006, 312, (5772), 384-8.
5. Lewis, N. S.; Russell, C. A.; Langat, P.; Anderson, T. K.; Berger, K.; Bielejec, F.; Burke, D. F.; Dudas, G.; Fonville, J. M.; Fouchier, R. A.; Kellam, P.; Koel, B. F.; Lemey, P.; Nguyen, T.; Nuansrichy, B.; Peiris, J. M.; Saito, T.; Simon, G.; Skepner, E.; Takemae, N.; consortium, E.; Webby, R. J.; Van Reeth, K.; Brookes, S. M.; Larsen, L.; Watson, S. J.; Brown, I. H.; Vincent, A. L., The global antigenic diversity of swine influenza A viruses. *eLife* 2016, 5, e12217. doi: 10.7554/eLife.12217
6. Sack, A.; Cullinane, A.; Daramragchaa, U.; Chuluunbaatar, M.; Gonchigoo, B.; Gray, G. C., Equine influenza virus - a neglected, reemergent disease threat. *Emerging Infectious Diseases* 2019, 25, (6), 1185-1191.
7. Smith, G. J.; Bahl, J.; Vijaykrishna, D.; Zhang, J.; Poon, L. L.; Chen, H.; Webster, R. G.; Peiris, J. S.; Guan, Y., Dating the emergence of pandemic influenza viruses. *Proc Natl Acad Sci U S A* 2009, 106, (28), 11709-12.
8. Vijaykrishna, D.; Holmes, E. C.; Joseph, U.; Fourment, M.; Su, Y. C.; Halpin, R.; Lee, R. T.; Deng, Y. M.; Gunalan, V.; Lin, X.; Stockwell, T. B.; Fedorova, N. B.; Zhou, B.; Spirason, N.; Kuhnert, D.; Boskova, V.; Stadler, T.; Costa, A. M.; Dwyer, D. E.; Huang, Q. S.; Jennings, L. C.; Rawlinson, W.; Sullivan, S. G.; Hurt, A. C.; Maurer-Stroh, S.; Wentworth, D. E.; Smith, G. J.; Barr, I. G., The contrasting phylodynamics of human influenza B viruses. *eLife* 2015, 4, e05055. doi: https://doi.org/10.7554/eLife.05055.001
9. Bodewes, R.; Morick, D.; de Mutsert, G.; Osinga, N.; Bestebroer, T.; van der Vliet, S.; Smits, S. L.; Kuiken, T.; Rimmelzwaan, G. F.; Fouchier, R. A.; Osterhaus, A. D., Recurring influenza B virus infections in seals. *Emerg Infect Dis* 2013, 19, (3), 511-2.
10. Ran, Z.; Shen, H.; Lang, Y.; Kolb, E. A.; Turan, N.; Zhu, L.; Ma, J.; Bawa, B.; Liu, Q.; Liu, H.; Quast, M.; Sexton, G.; Krammer, F.; Hause, B. M.; Christopher-Hennings, J.; Nelson, E. A.; Richt, J.; Li, F.; Ma, W., Domestic pigs are susceptible to infection with influenza B viruses. *J Virol* 2015, 89, (9), 4818-26.
11. Matsuzaki, Y.; Katsushima, N.; Nagai, Y.; Shoji, M.; Itagaki, T.; Sakamoto, M.; Kitaoka, S.; Mizuta, K.; Nishimura, H., Clinical features of influenza C virus infection in children. *J Infect Dis* 2006, 193, (9), 1229-35.

12. Smith, D. B.; Gaunt, E. R.; Digard, P.; Templeton, K.; Simmonds, P., Detection of influenza C virus but not influenza D virus in Scottish respiratory samples. *J Clin Virol* 2016, 74, 50-3.
13. Guo, Y. J.; Jin, F. G.; Wang, P.; Wang, M.; Zhu, J. M., Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. *J Gen Virol* 1983, 64 (Pt 1), 177-82.
14. Kimura, H.; Abiko, C.; Peng, G.; Muraki, Y.; Sugawara, K.; Hongo, S.; Kitame, F.; Mizuta, K.; Numazaki, Y.; Suzuki, H.; Nakamura, K., Interspecies transmission of influenza C virus between humans and pigs. *Virus Res* 1997, 48, (1), 71-79.
15. Manuguerra, J. C.; Hannoun, C., Natural infection of dogs by influenza C virus. *Res Virol* 1992, 143, (3), 199-204.
16. Manuguerra, J. C.; Hannoun, C.; Simon, F.; Villar, E.; Cabezas, J. A., Natural infection of dogs by influenza C virus: a serological survey in Spain. *New Microbiol* 1993, 16, (4), 367-71.
17. Salem, E.; Cook, E. A. J.; Lbacha, H. A.; Oliva, J.; Awoume, F.; Aplogan, G. L.; Hymann, E. C.; Muloi, D.; Deem, S. L.; Alali, S.; Zouagui, Z.; Fevre, E. M.; Meyer, G.; Ducatez, M. F., Serologic evidence for influenza C and D virus among ruminants and camelids, Africa, 1991-2015. *Emerging Infectious Diseases* 2017, 23, (9), 1556-1559.
18. Hause, B. M.; Ducatez, M.; Collin, E. A.; Ran, Z. G.; Liu, R. X.; Sheng, Z. Z.; Armien, A.; Kaplan, B.; Chakravarty, S.; Hoppe, A. D.; Webby, R. J.; Simonson, R. R.; Li, F., Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *Plos Pathog* 2013, 9, (2). e1003176. https://doi.org/10.1371/journal.ppat.1003176
19. Flynn, O.; Gallagher, C.; Mooney, J.; Irvine, C.; Ducatez, M.; Hause, B.; McGrath, G.; Ryan, E., Influenza D Virus in Cattle, Ireland. Emerg Infect Dis 2018, 24, (2), 389-391.
20. Collin, E. A.; Sheng, Z.; Lang, Y.; Ma, W.; Hause, B. M.; Li, F., Cocirculation of two distinct genetic and antigenic lineages of proposed influenza D virus in cattle. *J Virol* 2015, 89, (2), 1036-42.
21. Asha, K.; Kumar, B., Emerging influenza D virus threat: what we know so far! *J Clin Med* 2019, 8, (2). 192. doi: 10.3390/jcm8020192.
22. White, S. K.; Ma, W. J.; McDaniel, C. J.; Gray, G. C.; Lednicky, J. A., Serologic evidence of exposure to influenza D virus among persons with occupational contact with cattle. Journal of Clinical Virology 2016, 81, 31-33.
23. Longdon, B.; Murray, G. G.; Palmer, W. J.; Day, J. P.; Parker, D. J.; Welch, J. J.; Obbard, D. J.; Jiggins, F. M., The evolution, diversity, and host associations of rhabdoviruses. Virus Evol 2015, 1, (1), vev014. doi: 10.1093/ve/vev014

24. Parry, R.; Asgari, S., Discovery of novel crustacean and cephalopod faviviruses: insights into the evolution and circulation of flaviviruses between marine invertebrate and vertebrate hosts. *J Virol* 2019, 93, (14), e00432-19 doi: 10.1128/JVI.00432-19
25. Francois, S.; Filloux, D.; Roumagnac, P.; Bigot, D.; Gayral, P.; Martin, D. P.; Froissart, R.; Ogliastro, M., Discovery of parvovirus-related sequences in an unexpected broad range of animals. Sci Rep 2016, 6, 30880. doi: 10.1038/srep30880.
26. Sun, Y.; Huang, Y.; Li, X.; Baldwin, C. C.; Zhou, Z.; Yan, Z.; Crandall, K. A.; Zhang, Y.; Zhao, X.; Wang, M.; Wong, A.; Fang, C.; Zhang, X.; Huang, H.; Lopez, J. V.; Kilfoyle, K.; Zhang, Y.; Orti, G.; Venkatesh, B.; Shi, Q., Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies. *Gigascience* 2016, 5, 18. doi: 10.1186/s13742-016-0124-7
27. Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30, (15), 2114-2120.
28. Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. D.; Chen, Z. H.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A., Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, 29, (7), 644-U130.
29. Quinlan, A. R.; Hall, I. M., BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26, (6), 841-842.
30. Wang, M.; Marin, A., Characterization and prediction of alternative splice sites. *Gene* 2006, 366, (2), 219-27.
31. Dubois, J.; Terrier, O.; Rosa-Calatrava, M., Influenza viruses and mRNA splicing: doing more with less. *mBio* 2014, 5, (3), e00070-14.
32. Katoh, K.; Rozewicki, J.; Yamada, K. D., MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019, 20, (4), 1160-1166.
33. Capella-Gutierrez, S.; Silla-Martinez, J. M.; Gabaldon, T., trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009, 25, (15), 1972-3.
34. Kalyaanamoorthy, S.; Minh, B. Q.; Wong, T. K. F.; von Haeseler, A.; Jermiin, L. S., ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017, 14, (6), 587-589.

35. Nguyen, L. T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q., IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015, 32, (1), 268-274.
36. Le, S. Q.; Gascuel, O., An improved general amino acid replacement matrix. *Mol Bio Evol* 2008, 25, (7), 1307-1320.
37. Whelan, S.; Goldman, N., A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001, 18, (5), 691-9.
38. Dang, C. C.; Le, Q. S.; Gascuel, O.; Le, V. S., FLU, an amino acid substitution model for influenza proteins. *BMC Evol Biol* 2010, 10, 99.
39. Soubrier, J.; Steel, M.; Lee, M. S.; Der Sarkissian, C.; Guindon, S.; Ho, S. Y.; Cooper, A., The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol* 2012, 29, (11), 3345-58.
40. Minh, B. Q.; Nguyen, M. A. T.; von Haeseler, A., Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013, 30, (5), 1188-1195.
41. Janet, P.-M.; Juan, C.-P.; Annie, E.-C.; Gilberto, M.-C.; Hilda, L.; Enrique, S.-V.; Denhi, S.; Jesus, C.-M.; Alfredo, C.-R., Multi-organ transcriptomic landscape of *Ambystoma velasci* metamorphosis. bioRxiv 2020, 2020.02.06.937896.
42. Dwaraka, V. B.; Smith, J. J.; Woodcock, M. R.; Voss, S. R., Comparative transcriptomics of limb regeneration: Identification of conserved expression changes among three species of Ambystoma. *Genomics* 2019, 111, (6), 1216-1225.
43. Sabin, K. Z.; Jiang, P.; Gearhart, M. D.; Stewart, R.; Echeverri, K., AP-1(cFos/JunB)/miR-200a regulate the pro-regenerative glial cell response during axolotl spinal cord regeneration. *Commun Biol* 2019, 2, 91. doi: https://doi.org/10.1038/s42003-019-0335-4
44. Tsai, S. L.; Baselga-Garriga, C.; Melton, D. A., Midkine is a dual regulator of wound epidermis development and inflammation during the initiation of limb regeneration. *eLife* 2020, 9. e50765. doi: 10.7554/eLife.50765
45. Huggins, P.; Johnson, C. K.; Schoergendorfer, A.; Putta, S.; Bathke, A. C.; Stromberg, A. J.; Voss, S. R., Identification of differentially expressed thyroid hormone responsive genes from the brain of the Mexican Axolotl (*Ambystoma mexicanum*). *Comp Biochem Physiol C Toxicol Pharmacol* 2012, 155, (1), 128-35.

46. King, B. L.; Yin, V. P., A conserved microRNA regulatory circuit is differentially controlled during limb/appendage regeneration. *PLoS ONE* 2016, 11, (6). e0157106. doi: https://doi.org/10.1371/journal.pone.0157106

47. Lee, S. Y.; Lee, H. J.; Kim, Y. K., Comparative transcriptome profiling of selected osmotic regulatory proteins in the gill during seawater acclimation of chum salmon (*Oncorhynchus keta*) fry. *Sci Rep* 2020, 10, (1), 1987. doi: https://doi.org/10.1038/s41598-020-58915-6

48. Edwards, R. J.; Tuipulotu, D. E.; Amos, T. G.; O'Meally, D.; Richardson, M. F.; Russell, T. L.; Vallinoto, M.; Carneiro, M.; Ferrand, N.; Wilkins, M. R.; Sequeira, F.; Rollins, L. A.; Holmes, E. C.; Shine, R.; White, P. A., Draft genome assembly of the invasive cane toad, *Rhinella marina*. *Gigascience* 2018, 7, (9). giy095 doi: https://doi.org/10.1093/gigascience/giy095

49. Russo, A. G.; Eden, J. S.; Enosi Tuipulotu, D.; Shi, M.; Selechnik, D.; Shine, R.; Rollins, L. A.; Holmes, E. C.; White, P. A., Viral discovery in the invasive australian Cane Toad (*Rhinella marina*) using metatranscriptomic and genomic approaches. *J Virol* 2018, 92, (17). e00768-18. doi: 10.1128/JVI.00768-18

50. Zhao, L.; Liu, L.; Wang, S.; Wang, H.; Jiang, J., Transcriptome profiles of metamorphosis in the ornamented pygmy frog Microhyla fissipes clarify the functions of thyroid hormone receptors in metamorphosis. *Sci Rep* 2016, 6, 27310. doi: 10.1038/srep27310

51. Liu, L.; Wang, S.; Zhao, L.; Jiang, J., De novo transcriptome assembly for the lung of the ornamented pygmy frog (*Microhyla fissipes*). *Genom Data* 2017, 13, 44-45.

52. Betakova, T.; Nermut, M. V.; Hay, A. J., The NB protein is an integral component of the membrane of influenza B virus. *J Gen Virol* 1996, 77 (Pt 11), 2689-94.

53. Demircan, T.; Ilhan, A. E.; Ayturk, N.; Yildirim, B.; Ozturk, G.; Keskin, I., A histological atlas of the tissues and organs of neotenic and metamorphosed axolotl. *Acta Histochem* 2016, 118, (7), 746-759.

54. Ibricevic, A.; Pekosz, A.; Walter, M. J.; Newby, C.; Battaile, J. T.; Brown, E. G.; Holtzman, M. J.; Brody, S. L., Influenza virus receptor specificity and cell tropism in mouse and human airway epithelial cells. *J Virol* 2006, 80, (15), 7469-80.

55. Wille, M.; van Run, P.; Waldenstrom, J.; Kuiken, T., Infected or not: are PCR-positive oropharyngeal swabs indicative of low pathogenic influenza A virus infection in the respiratory tract of Mallard *Anas platyrhynchos*? Vet Res 2014, 45, 53.

56. Ciminski, K.; Schwemmle, M., Bat-borne influenza A viruses: an awakening. *Cold Spring Harb Perspect Med* 2019 a038612. doi: 10.1101/cshperspect.a038612.

509 57. Percino-Daniel, R., New record of *Ambystoma velasci* (Dugès, 1888) from Western Mexico.
510 *Herpetology Notes* 2019, 12, 351-352.
511 58. Contreras, V.; Martinez-Meyer, E.; Valiente, E.; Zambrano, L., Recent decline and potential
512 distribution in the last remnant area of the microendemic Mexican axolotl (*Ambystoma*
513 *mexicanum*). *Biol Conserv* 2009, 142, (12), 2881-2885.
514

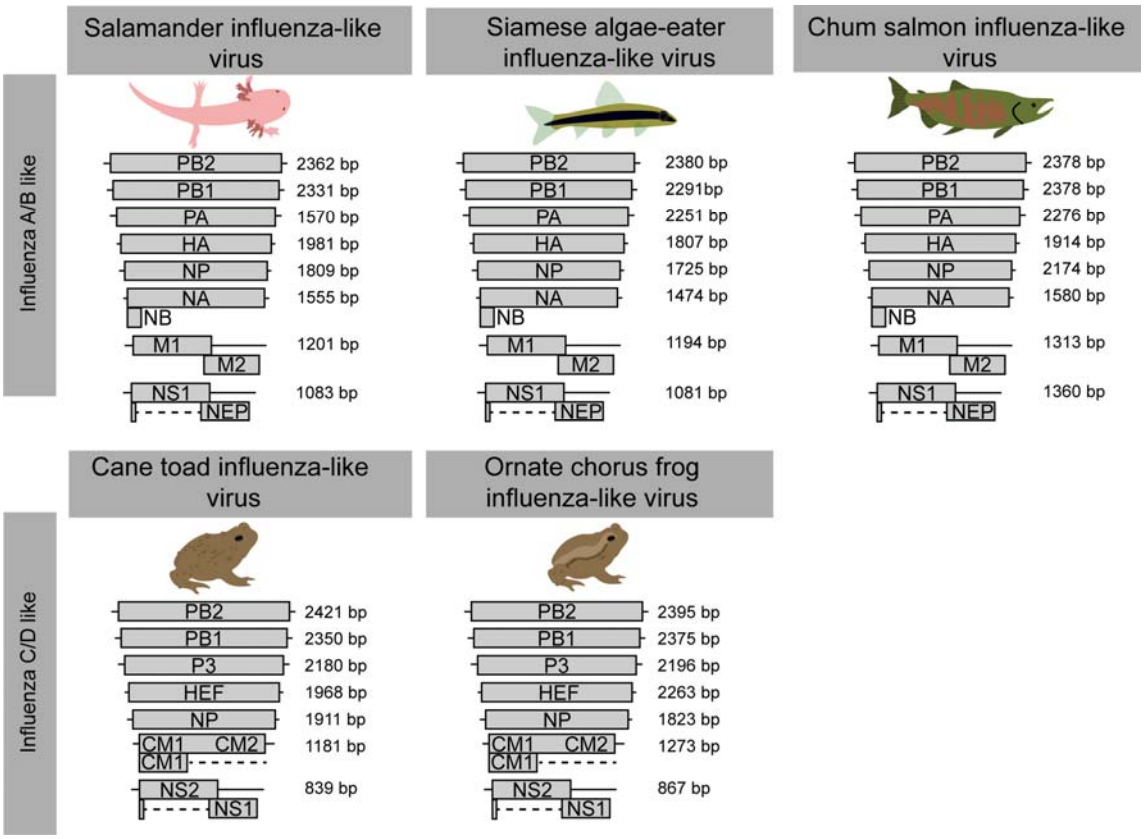

Figure 1. Genome architecture of the novel influenza-like viruses identified here. Genome architecture of IAV/IBV and ICV/IDV is shown on the left. For each novel virus, we provide the segment name and size. PB1, RNA-dependent RNA polymerase basic subunit 1; PB2, RNA-dependent RNA polymerase basic subunit 2; PA, RNA-dependent RNA polymerase acidic subunit; NP, nucleoprotein; HA, hemagglutinin; HEF, Hemagglutinin esterase; NA, neuraminidase; M, matrix; NS, non-structural protein. Detailed annotations for each virus are presented in Table S1-S6, Figure S1-S5.

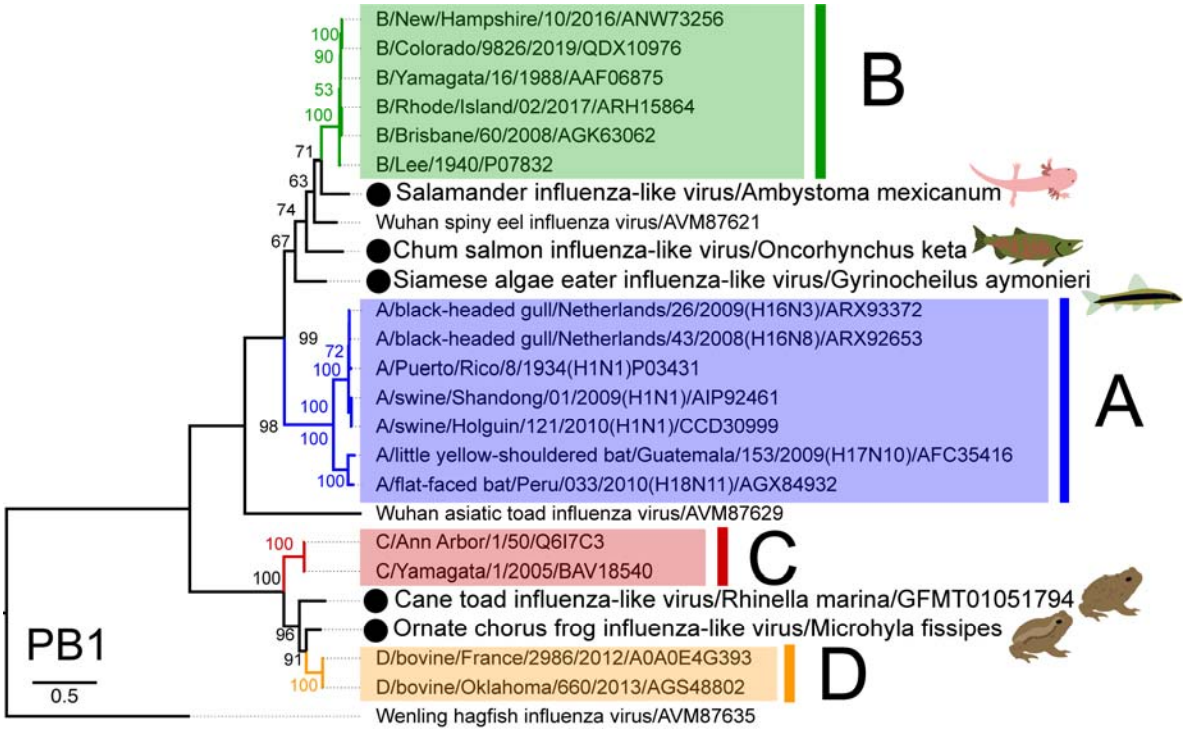

Figure 2. The evolutionary history of vertebrate influenza-like viruses. Maximum likelihood tree of the PB1 segment, which encodes the RNA-dependent RNA polymerase, of various influenza-like viruses. Lineages corresponding to IAV, IBV, ICV, IDV are coloured blue, green, red and orange, respectively. Viruses identified in this study are denoted by a black circle and pictogram of the host species of the library. Wenling hagfish influenza virus is set as the outgroup as per Shi et al. 2018 [3]. The scale bar indicates the number of amino acid substitutions per site.

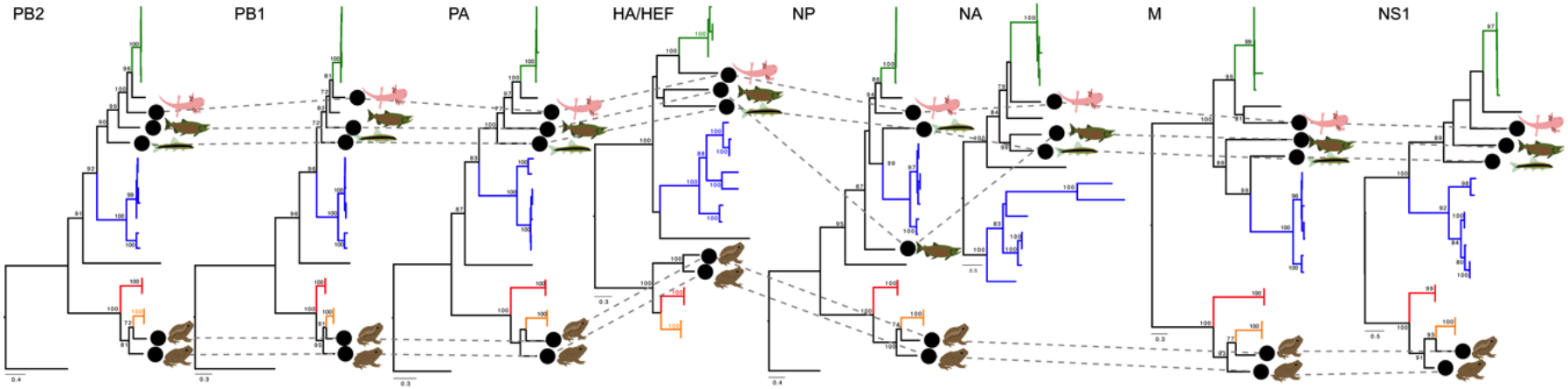

Figure 3. Phylogenetic trees for segments in cases where complete coding genomes could be recovered. Maximum likelihood trees for each segment ordered by (segment) size. Lineages are coloured by influenza virus, in which IAV, IBV, ICV, IDV are coloured blue, green, red and orange, respectively. Viruses identified in this study are denoted by a black circle and pictogram of the host species of the library. The phylogenetic position of each virus is traced across the trees with grey dashed lines. ICV, IDV, cane toad and chorus frog influenza-like viruses do not have an NA segment. The scale bar for each tree indicates the number of amino acid substitutions per site. Where possible the trees are rooted using the hagfish influenza virus (PB2, PB1, PA/P3, NP). The HA/HEF, M and NS trees were midpoint rooted.

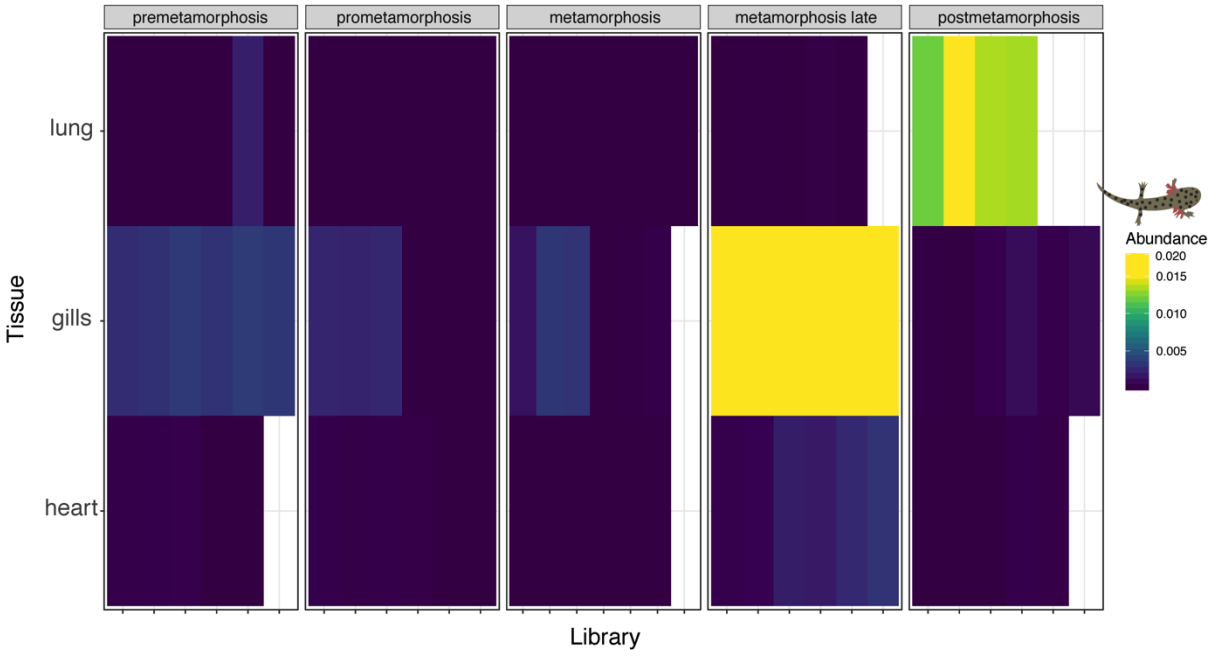

Figure 4. Heat map of the abundance of Salamander influenza-like virus reads in the multi-tissue transcriptome library of developmental stages of Plateau tiger salamander [41]. Read abundance for this virus was highest in the gills of late metamorphosis salamanders, and lungs of postmetamorphasis salamander, strongly suggesting tropism for respiratory epithelium.

547 **Table 1.** Metadata for the novel influenza viruses identified in this study.

Virus name	Host	Tissue sampled	Reference	BLASTp output of predicted amino acid identity of polymerase genes			
				Segment	GenBank Accession	Coverage (%); Identity (%)	E-value
Salamander influenza-like virus	Mexican walking fish (*Ambystoma mexicanum)*	Various	[42-46]	PB1 Influenza B virus (B/California/24/2016)	QHI05420	100%; 75.70%	0.0
				PB2 Influenza B virus (B/Sydney/19/2011)	AZY32600	99%; 61.83%	0.0
	Plateau tiger salamander (*Ambystoma velasci*)	Various	[41]	PA Influenza B virus (B/Indiana/07/2016)	ANW74127	97%; 59.64%	0.0
Siamese algae-eater influenza-like virus	Siamese algae-eater *(Gyrinocheilus aymonieri)*	Gills	[26]	PB1 Influenza B virus (B/California/24/2016)	ANW79211	99%; 69.46%	0.0
				PB2 Influenza B virus (B/New York/1121/2007)	AHL92298	99%; 48.58%	0.0
				PA Wuhan spiny eel influenza virus	AVM87622	96%; 53.89%	0.0
Chum salmon influenza-like virus	Chum salmon *(Oncorhynchus keta)*	Gills	[47]	PB1 Influenza B virus (B/Iowa/14/2017)	QHI05420	100%; 69.50%	0.0
				PB2 Influenza B virus (B/Memphis/5/93)	AAU94860	99%; 57.05%	0.0
				PA Influenza B virus (B/Taiwan/45/2007)	ACO06009	98%; 51.88%	0.0
Cane toad influenza-like virus	Cane Toad *(Rhinella marina)*	Tissue unknown, Larval	[48]	PB1 Influenza D virus (bovine/Mexico/S7/2015)	AMN87903	99% ; 77.03%	0.0
				PB2 Influenza D virus (bovine/Kansas/14-22/2012)	AIO11621	995; 62.87%	0.0
				P3 Influenza D virus (bovine/Yamagata/10710/2016)	BBC14929	100%; 62.66%	0.0
Ornate chorus frog influenza-like virus	Ornate chorus frog *(Microhyla fissipes)*	Larval, Lung	[50, 51]	PB1 Influenza D virus (bovine/Mississippi/C00046N/2014)	ALE66333	98%; 81.91%	0.0
				PB2 Influenza D virus (swine/Italy/173287-4/2016)	AON76692	99%; 68.22%	0.0
				P3 Influenza D virus (D/bovine/Shandong/Y217/2014)	AIE52099	99%; 63.78%	0.0

548