

***In silico* detection of highly diverse RNA
viruses in insects and mammals**

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

By

Sofia Paraskevopoulou

Berlin, 2020

Sofia Paraskevopoulou

Inaugural-Dissertation

In silico detection of highly diverse RNA viruses in insects and mammals

Research presented in this dissertation was carried out under the supervision of Prof. Dr. Christian Drosten at the Institute of Virology, Universitätsklinikum Bonn from 2016 to 2017, and at the Institute of Virology, Charité-Universitätsmedizin Berlin from 2017 to 2020.

1st Reviewer: Prof. Dr. Christian Drosten

Institute of Virology, Charité

Berlin, Germany

2nd Reviewer: Prof. Dr. Dino McMahon

Institute of Biology - Zoology, Freie Universität Berlin

Berlin, Germany

Date of defense: 20.05.2021

Contents

Summary	3
Zusammenfassung	5
1 General Introduction	7
1.1 An introduction to virus discovery	7
1.2 Insect virus diversity and evolution	9
1.3 Viral groups studied in this thesis	11
1.3.1 Single-stranded negative sense RNA viruses	12
1.3.1.1 <i>Bunyavirales</i>	12
1.3.1.2 <i>Articulavirales</i>	13
1.3.1.3 <i>Mononegavirales</i>	14
1.3.1.4 <i>Jingchuvirales, Muvirales, and Goujianvirales</i>	14
1.3.2 Single-stranded positive sense RNA viruses	15
1.3.2.1 <i>Amarillovirales</i>	15
1.3.2.2 <i>Tolivirales</i>	16
1.3.3 Hepatitis delta virus and other deltaviruses	18
1.3.3.1 Genome organization	18
1.3.3.2 Replication	20
1.3.3.3 Packaging	21
1.3.3.4 Theories on the evolutionary origins	21
1.3.3.5 Epidemiology and pathogenesis	23
1.3.3.6 Other deltaviruses	25
1.4 <i>In silico</i> virus detection	28
1.4.1 Published pipelines & common issues in bioinformatic analyses	30
1.4.2 Computational approach and methodology applied herein	31
1.5 Virus taxonomic classification and the ICTV	35
1.6 Aim of this thesis	38
2 Re-assessing the diversity of negative strand RNA viruses in insects	40
3 Mammalian deltavirus without hepadnavirus coinfection in the neotropical	

rodent <i>Proechimys semispinosus</i>	73
4 Viromics of extant insect orders unveil the evolution of the flavi-like superfamily	81
5 General Discussion	106
5.1 On virus taxonomy and discovery	106
5.2 On the discoveries of RNA viruses in insects	107
5.3 On deltaviruses	112
Bibliography	114
Appendix	128
Table A1	128
List of abbreviations	130
Appendix of Chapter 2	131
S1 Text	132
S1 Table	150
S2 Table	156
S3 Table	158
S4 Table	159
S1-S34 Figures	160
Appendix of Chapter 3	192
Supporting Information Appendix	193
Appendix of Chapter 4	224
Fig. S1	224
Table S1	225
Acknowledgements	227
Declaration of authorship	228
List of publications	229

Summary

RNA viruses are among the fastest-evolving biological entities and infect a broad range of organisms. They not only cause some of the most notorious and deadly human diseases, but they are also diverse in a number of characteristics, such as the presence/absence of an envelope, the genomic orientation, and the number of genomic segments, to name a few. As virus research is mainly focused on human, livestock, and crop pathogens, a large proportion of viruses is generally neglected and especially those that occur in hosts that are not directly significant to humans. Yet, viruses remain ubiquitous and the field of virus discovery provides a fundamental basis for understanding virus emergence and evolution. Also, the detection of closely-related viruses in distantly-related hosts offers insight and preparedness on emerging diseases.

Insects are an important disease vector and transmit a large variety of pathogenic viruses. Thus, viruses in non-blood-feeding insects are not studied in the same extent as viruses that occur in blood-feeding insects. However, blood-feeding insects represent only a tiny fraction of the enormous insect diversity. Computational analyses of the transcriptome of 1,243 insect species revealed viruses associated with a multitude of single-stranded RNA virus families of both positive- and negative-sense RNA orientation. This insect dataset is designed to have a worldwide sampling that represents all extant orders. The here-presented virus findings provide insights into insect virus diversity which is largely unexplored. Host associations contribute to broadening the general understanding of virus evolution within viral taxonomic units. With regards to viruses with a segmented genome, co-segregation phylogenetic analysis offers a useful tool to study segment reassortment. Consulting current virus taxonomic classification criteria, many of the identified viruses signify new viral species and genera, and, in some cases, phylogenetic topologies suggest taxonomic reforms of existing ranks.

The only RNA virus with a completely unknown origin and with no other known counterpart until recently was the human Hepatitis delta virus (HDV). The here-discovered rodent deltavirus is the first non-human mammalian deltavirus reported, shedding light into the evolutionary history of HDV. Rodent deltavirus occurs in rodents of the species *Proechimys semispinosus*, mainly in adult males. Genome assembly, annotation, and experimental virus replication studies show that this viral agent successfully replicates *in vitro* and expresses one

protein, the small delta antigen, which does not carry the necessary C-terminal motif for assembly into HBV envelope proteins. Although the debate of whether deltaviruses fulfill the virus definition criteria still holds, the non-human deltavirus findings resolve fundamental questions on the origin of HDV.

Zusammenfassung

RNA-Viren gehören zu den sich am schnellsten entwickelnden biologischen Einheiten und infizieren ein breites Spektrum von Organismen. Sie verursachen nicht nur einige der berüchtigtsten und tödlichsten menschlichen Krankheiten, sondern sind auch vielfältig in einer Reihe von Merkmalen, wie z.B. das Vorhandensein/Fehlen einer Hülle, die genomische Ausrichtung und die Anzahl der genomischen Segmente, um nur einige zu nennen. Da sich die Virusforschung hauptsächlich auf Erreger von Menschen, Nutztieren und Nutzpflanzen konzentriert, wird ein großer Teil der Viren im Allgemeinen vernachlässigt, insbesondere diejenigen, die in Wirten vorkommen, die für den Menschen nicht direkt von Bedeutung sind. Dennoch sind Viren nach wie vor allgegenwärtig, und das Feld der Virusentdeckung bietet die Grundlage für das Verständnis der Entstehung und Evolution von Viren. Außerdem bietet der Nachweis eng verwandter Viren in entfernt verwandten Wirten Einblicke und Vorsorge für neu auftretende Krankheiten.

Insekten sind ein wichtiger Krankheitsüberträger und übertragen eine große Vielfalt an pathogenen Viren. Daher werden Viren in nicht blutfressenden Insekten nicht in demselben Umfang untersucht wie Viren, die in blutfressenden Insekten vorkommen. Blutfressende Insekten stellen jedoch nur einen winzigen Bruchteil der enormen Insektenvielfalt dar. Computergestützte Analysen des Transkriptoms von 1.243 Insektenarten ergaben Viren, die zu einer Vielzahl von einzelsträngigen RNA-Virusfamilien mit positiver und negativer RNA-Orientierung gehören. Dieser Insektendatensatz ist so angelegt, dass eine weltweite Stichprobe vorliegt, die alle existierenden Ordnungen repräsentiert. Die hier vorgestellten Virusfunde geben Einblicke in die weitgehend unerforschte Virusvielfalt der Insekten. Wirtsassoziationen tragen dazu bei, das allgemeine Verständnis der Virusevolution innerhalb viraler taxonomischer Einheiten zu erweitern. Im Hinblick auf Viren mit einem segmentierten Genom, bietet die phylogenetische Co-Segregationsanalyse ein nützliches Werkzeug zur Untersuchung des Segment-Reassortiments. Unter Berücksichtigung der aktuellen taxonomischen Klassifizierungskriterien für Viren, stellen viele der identifizierten Viren neue virale Spezies und Gattungen dar und in einigen Fällen legen die phylogenetischen Topologien taxonomische Reformen bestehender Ränge vor.

Das einzige RNA-Virus mit völlig unbekanntem Ursprung, zu dem bis vor kurzem kein

Gegenstück bekannt war, war das menschliche Hepatitis-Delta-Virus (HDV). Das hier entdeckte Nagetier-Deltavirus ist das erste nicht-menschliche Säugetier-Deltavirus, über das berichtet wurde, und wirft ein Licht auf die Evolutionsgeschichte von HDV. Das Nagetier-Deltavirus kommt bei Nagetieren der Art *Proechimys semispinosus* vor, hauptsächlich bei erwachsenen Männchen. Genomassemblierung, Annotation und experimentelle Virusreplikationsstudien zeigen, dass dieser virale Erreger sich erfolgreich *in vitro* repliziert und ein Protein, das kleine Delta-Antigen, exprimiert, das das notwendige C-terminale Motiv für die Assemblierung zu HBV-Hüllproteinen nicht trägt. Obwohl die Debatte darüber, ob Deltaviren die Kriterien der Virusdefinition erfüllen, immer noch andauert, beantworten die Erkenntnisse über nicht-humane Deltaviren grundlegende Fragen über den Ursprung von HDV.

1 General Introduction

1.1 An introduction to virus discovery

At the very early ages of virology, end of the 19th century, plant tissue filtration was the first technique to ever demonstrate the existence of a virus, i.e. the discovery of tobacco mosaic virus (summarized by Horzinek, 1997). Since then, techniques such as tissue culture, electron microscopy, serology, or inoculation were applied for the purpose of virus identification as technology was advancing. With the advent of molecular techniques, methods like polymerase chain reaction (PCR), microarrays, or hybridization have come to aid in virus discovery (reviewed by Mokili et al., 2012). Essentially though, the occurrence of a virus has traditionally required verification of virus propagation in a cell culture system, where a cytopathic effect (CPE) could be directly observed on cells. However, not all viruses can be easily and effectively propagated in cell culture. Also, often the type of cells available for culture is itself a constraining factor and commercial cell lines are not readily available for every organismic species or tissue. In addition, unlike cellular organisms which share the evolutionary homologous 16S or 18S rRNA genes, a universally conserved ubiquitous viral gene that would serve as a global target for PCR-based virus detection does not exist. The development of next-generation sequencing (NGS) techniques has literally transformed the field of virus discovery and launched an era of viral metagenomics. NGS technologies offer massive parallel sequencing for large amounts of nucleic acids, in contrast to Sanger sequencing where only a limited fragment size can be sequenced. Independent of cell culture and prior sequence knowledge, this technology allows virus detection irrespective of the presence of a particular genetic marker.

Less than twenty years ago, the first viral metagenomic study in 2002 examined the abundance of viral communities in a marine environmental sample (Breitbart et al., 2002). Ever since, the field of viral metagenomics has continuously flourished and yet, analyses of metagenomic environmental samples suggest that less than 1% of the existing viral diversity has been explored (Mokili et al., 2012). Indicatively, within just one decade the number of reported uncultured viral metagenomic sequences increased by 10,000-fold, reaching almost one million sequences by 2018 (Roux et al., 2019). By comparison, in 2002 when the first viral metagenomic study was carried out, the number of publicly available complete and

annotated viral genomes in the database of the National Center for Biotechnology Information (NCBI) was about 2,200 genomes when today's corresponding number almost reaches 40,000 viral genomes. Exploring viral diversity with methods independent of cell culture, serology, or PCR-dependent methods has truly revolutionized the field of virus discovery.

1.2 Insect virus diversity and evolution

Among all metazoan organismic groups, insects harbor the most species-rich animal diversity counting over 2,000 recognised species (Berenbaum, 2017). Their enormous diversity inhabits nearly every ecosystem for approximately the last 479 million years (Misof et al., 2014). Insects feeding on host blood (*hematophagous*) are known vectors of pathogenic viruses, such as Dengue virus, Yellow fever virus, Chikungunya virus, and many others that are largely transmitted by mosquito or tick vectors and impose global health burdens (Jánová, 2019). Viruses that infect vertebrates and require an intermediate replicating step within an insect/arthropod host through which the virus is transmitted to the vertebrate host are called *arthropod-borne viruses* or simply *arboviruses*. These differ from the *insect-specific viruses* (ISVs) that solely infect insects and are not transmitted to other hosts; the latter additionally do not appear to replicate in vertebrate cells. Although insect distribution is global in even the most extreme habitats, the focus on studying insects viruses has traditionally and mainly involved arboviruses. Hematophagous insects are nevertheless only a tiny fraction of the entire insect biodiversity and scientists have often voiced the need for research on insect-specific viruses (Junglen & Drosten, 2013; Bolling et al., 2015; Vasilakis & Tesh, 2015; Nouri et al., 2018; Gurung et al., 2019).

Virtually following a matryoshka principle, insect virus diversity is inevitably expected to present high levels, as it is hosted within a highly-diversified organismic group. This has been manifested by large-scale metagenomic studies in recent years (Rosario et al., 2012; Li et al., 2015; Shi et al., 2016a,b; Käfer et al., 2019; Porter et al., 2019). Insect virus diversity has enormously expanded and even percolated through virus families that did not report any insect host previously (e.g. *Tombusviridae*, see also **Chapter 4**). At the same time, host associations in certain virus families remain stable and related insect viruses are not reported despite broad insect sampling in virus discovery studies (e.g. *Paramyxoviridae*, more details in **Chapter 2**). Ecological factors play a role in such cases, for example arbovirus taxa with the necessity of replication in mosquito salivary glands or taxa including viruses transmitted only between mammalian hosts.

The discovery of new insect viruses has also enriched the diversity of genomic architecture as new patterns of genome organization have emerged. For example, the newly discovered

virus family of *Chuviridae* includes viruses with genomic organization as diverse as linear/circular and segmented circular genomes (Li et al., 2015). Moreover, a whole clade of segmented insect viruses has emerged within the traditionally non-segmented *Flaviviridae* family. This clade, yet unclassified but known as "jingmenviruses", gathers a diversity of both arboviruses and insect viruses discovered less than a decade ago (detailed description in **Chapter 4**). Such findings - or the absence of them in the largely diversified group of insects - provide new insights in virus evolution (Calisher & Higgs, 2018).

1.3 Viral groups studied in this thesis

Viruses are principally grouped based on the structure of their genetic material, as proposed by David Baltimore (Baltimore, 1971). The Baltimore classification scheme distinguishes viruses in seven different groups:

- I: double-stranded DNA viruses
- II: single-stranded DNA viruses
- III: double-stranded RNA viruses
- IV: single-stranded positive sense RNA viruses
- V: single-stranded negative sense RNA viruses
- VI: single-stranded RNA-RT viruses, with DNA intermediate
- VII: double-stranded DNA-RT viruses, with RNA intermediate

Among these groups there are *hallmark* genes that encode key proteins involved in genome replication and virion formation, e.g. the RNA-dependent RNA polymerase (RdRp) or the reverse transcriptase (RT). Hallmark genes are shared among Baltimore groups unifying the virus world and ensuring the connectivity of its evolutionary relationships (Koonin et al., 2015, 2020). For example, viruses in Baltimore groups III, IV, and V share the RdRp gene; groups VI and VII share the RT gene; and groups I, II, and IV share the superfamily 3 helicase gene (a summarizing scheme is found in Koonin et al., 2020). Phylogenetic reconstructions based on hallmark genes offer a basis to comprehend the evolutionary relationships among Baltimore groups and generally within the virus world.

Viruses in Baltimore groups III, IV, and V are the fastest evolving biological entities with substitution rates between 10^{-3} to 10^{-5} per nucleotide per generation (Drake et al., 1998) or between 10^{-3} to 10^{-6} per nucleotide per cell infection (Sanjuán et al., 2010). These viruses share the RdRp hallmark gene which encodes the polymerase, necessary for viral replication. High mutation rates in RNA viruses occur because of the lack of proofreading activity of the RdRp. Among virus families there are different translation strategies and numbers of proteins translated, but the viral RdRp is globally expressed. As the animal kingdom is largely populated by RNA viruses of the Baltimore classes IV and V (aside from DNA viruses) (Koonin et al., 2020), the objective of this thesis is focused on the computational discovery of such viruses.

1.3.1 Single-stranded negative sense RNA viruses

A negative sense molecule in virology corresponds to a molecule that needs to undergo transcription for the production of an mRNA that will undergo downstream translation. Also, because RNA viruses in general utilize an own polymerase for replication/transcription and because their genome is not a "ready-for-translation" molecule, the RdRp is always packaged inside the virion of these viruses. Genome sizes in this group range between 10–25 kb and genome architecture is mainly linear (exceptions: *Tenuivirus* and some *Chuviridae* with circular genome), and configurations are encountered in both segmented and non-segmented forms. Recent taxonomic assignments have placed all viruses of this group under the phylum *Negarnaviricota*. The phylum is further divided into the subphyla *Haploviricotina* and *Polyploviricotina* which reflect the encoding of a 5' mRNA cap in the former and a host *cap snatching* in the latter. Apart from the family *Aspiviridae* and the metagenomic-derived families for which no virion morphology is known, all other viruses are enveloped in a lipid membrane. Viruses within the *Negarnaviricota* phylum infect a variety of organismic hosts and among members of this phylum are known causative agents of infectious disease in humans: Zaire ebolavirus, Marburg marburgvirus, Influenza A virus, Rabies lyssavirus, and many others. During the last years, a large amount of new viruses that are evolutionary related to viruses within *Negarnaviricota* has been discovered in metagenomic datasets of insects and invertebrates (Li et al., 2015; Shi et al., 2016a; Käfer et al., 2019). A detailed presentation and discussion on this topic is provided in **Chapter 2**.

1.3.1.1 *Bunyavirales*

The *Bunyavirales* order is the only member of the viral class *Ellioviricetes* and the most species-rich order among all single-stranded negative sense RNA viral orders. The currently 383 recognised species are divided into 12 families: *Arenaviridae*, *Cruliviridae*, *Fimoviridae*, *Hantaviridae*, *Leishbuviridae*, *Myopoviridae*, *Nairoviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Phenuiviridae*, *Tospoviridae*, and *Wupedeviridae*. More than half of the species are shared between two families, namely *Phenuiviridae* and *Peribunyaviridae*. The family *Phenuiviridae* is populated by 117 species shared in 19 genera, while *Peribunyaviridae* hosts 97 species shared in only four genera. *Arenaviridae* and *Hantaviridae* comprise 50 and 48 species respectively, and *Fimoviridae*, *Nairoviridae*, *Phasmaviridae*, and *Tospoviridae* are smaller families with 11, 17, 13, and 26 species, respectively. The remaining fami-

lies *Cruliviridae*, *Leishbuviridae*, *Mypoviridae*, and *Wupedeviridae* are recently-established monospecific families with viruses derived from metagenomic datasets (Li et al., 2015; Akopyants et al., 2016).

With the exception of *Leptomonas shilevirus* of the *Leishbuviridae* family that occurs in protists, viruses within *Bunyavirales* are mainly found in insects, mammals, and, plants to a smaller extent. Genomes are generally segmented, with tripartite genome arrangements being the most common and the three segments are named to represent segment size: large (L), medium (M), and small (S). Yet, higher degrees of segmentation are found in plant viruses, such as the genus *Tenuivirus* of *Phenuiviridae* or the monogeneric family *Fimoviridae*, where a complete genome configuration can reach up to six segments. Bi- and monopartite genomes are also commonly encountered throughout *Bunyavirales*. Widely-known infectious viruses of this group are Crimean-Congo hemorrhagic fever orthonairovirus, Rift Valley fever phlebovirus, and Schmallerberg orthobunyavirus.

1.3.1.2 *Articulavirales*

This order of segmented viruses, as reflected by its name ("articula" meaning "segmented" in latin), includes two families, namely *Amnoonviridae* and *Orthomyxoviridae*. The recently-established family of *Amnoonviridae* is a monospecific family that consists of the species *Tilapia tilapinevirus*. The genome of this virus comprises 10 segments that add up to ca. 10 kb and was initially identified in tilapia fish cultures by Eynogor et al., 2014, after observing massive fish mortality in Israel.

The *Orthomyxoviridae* family is a well-studied family and includes all influenza viruses distributed in four genera: *Alpha-*, *Beta-*, *Gamma-*, and *Deltainfluenzavirus*. Influenzaviruses are a global major burden that have caused four major pandemics over the last century, counting hundreds of thousands of deaths each. Apart from humans, hosts of influenza-viruses include other mammals, such as cattle, pigs, horses, and even seals, as well as wild and domestic birds. Viruses that belong to the genera *Alpha-* and *Beta**influenzavirus* have eight genomic segments that altogether reach about 13.5 kb in length, whereas viruses of the *Gamma-* and *Deltainfluenzavirus* have seven genomic segments that sum up to ca. 10 kb in length. The genus *Isavirus* is monospecific with *Salmon isavirus* being an infectious agent that causes severe anemia to atlantic salmon. The *Isavirus* genome size is about 13.5 kb and

is divided to eight segments. *Quarantavirus* and *Thogotovirus* are arboviruses transmitted mainly by ticks to vertebrate hosts, such as birds, domestic mammals, and humans. Genomes are segmented with six or seven segments that reach sizes of about 10 kb for *Thogotovirus* and 11.5 kb for *Quarantavirus*.

1.3.1.3 *Mononegavirales*

The *Mononegavirales* order is the second species-rich order within *Negarnaviricota*, counting currently 71 genera and 339 species. Many known pathogens belong to this order, such as Mumps orthorubulavirus, Zaire ebolavirus, and Measles morbillivirus. As a member of the subphylum *Haploviricotina*, genomes in this order are monopartite with the exception of *Dichoravirus* and *Varicosavirus* of *Rhabdoviridae* that have bi-segmented genomes. Genome sizes range from 6 to 19 kb. The most populated family within *Mononegavirales* is *Rhabdoviridae* with 30 genera and 191 species infecting vertebrate, invertebrate, and plant hosts. A variety of viruses related to this family has been identified in invertebrate and insect hosts a few years ago (Li et al., 2015). Within the scope of this thesis, a large number of insect viruses related to *Rhabdoviridae* has also been identified (**Chapter 2** - Käfer et al., 2019), enriching the viral diversity of the family and confirming many host associations of previously-discovered viruses from metagenomic datasets.

The families *Artoviridae*, *Lipsiviridae*, and *Xinmoviridae* were recently established and include mainly insect and other crustacean viruses but also invertebrate viruses that occur in annelid worms. Members of *Nyamiviridae* are found in birds, invertebrates (such as tapeworms and nematodes), and insects. The families *Bornaviridae*, *Filoviridae*, *Paramyxoviridae*, *Pneumoniviridae*, and *Sunviridae* include vertebrate-pathogenic viruses that can cause hemorrhagic fever, respiratory disease, and neurological disease (e.g., encephalitis). Viruses of the *Mymonaviridae* family are divided into two genera and nine species in total, and occur in fungal hosts.

1.3.1.4 *Jingchuvirales*, *Muvirales*, and *Goujianvirales*

These three viral orders were established in 2018 (Wolf et al., 2018a), after the discovery of viruses in arthropod metagenomic data (Li et al., 2015). All orders include single monogeneric families but belong to separate classes (immediate higher rank):

- *Monjiviricetes* - ***Jingchuvirales*** - *Chuviridae* - *Mivirus*

- *Chunqiuviricetes* - **Muvirales** - *Qinviridae* - *Yingvirus*
- *Yunchangviricetes* - **Goujianvirales** - *Yueviridae* - *Yuyuevirus*

In *Jingchuvirales*, genomes are encountered both in linear/circular and segmented circular configurations, but phylogenetic associations do not show a specific pattern of evolution for genome architecture. Thus, genome segmentation and circularity have been suggested to have evolved multiple times (Li et al., 2015). The genomes of viruses within *Muvirales* are linear and bi-segmented, and genome size ranges between 8–9 kb. In *Goujianvirales* genomes are linear bi-segmented and around 7.5 kb in size. In all three orders, viruses were found within invertebrate and insects hosts, but no virus isolate exists yet. Apart from genomic characteristics and, in some cases, host associations, information on these viruses is limited as they stem from metagenomic data that are in their majority pooled samples from different organisms. As with all viral orders of *Negarnaviricota*, a detailed discussion and novel virus findings are presented in **Chapter 2**.

1.3.2 Single-stranded positive sense RNA viruses

The single-stranded positive sense viral RNA resembles an mRNA molecule owing to its 5' to 3' direction, making it a "ready-for-translation" molecule by the host ribosomes. Genome sizes vary from 2 kb in the *Narnaviridae* family to over 30 kb in *Coronaviridae* and although showing both segmented and non-segmented architectures, only linear configurations are known so far. Viruses with a single-stranded positive sense RNA genome taxonomically belong to the phyla *Kitrinoviricota*, *Lenarviricota*, and *Pisuviricota* (but not within the class *Duplopiviricetes*). **Chapter 4** discusses new findings related to the families *Flaviviridae* and *Tombusviridae* (orders *Amarillovirales* and *Tolivirales*, respectively). These two viral families are often phylogenetically grouped together in the "flavivirus-like superfamily", owing to their RdRp amino acid similarities. In the latest taxonomy classification these two families were placed in the *Kitrinoviricota* viral phylum, yet in different orders after consulting genetic distance criteria (ICTV Executive Committee and others, 2020).

1.3.2.1 *Amarillovirales*

This order includes a single family, *Flaviviridae*, that is further divided into four genera: *Flavivirus*, *Hepacivirus*, *Pegivirus*, and *Pestivirus*. Viruses that belong to *Flaviviridae* are enveloped with a linear monopartite genome of about 9–13 kb. The 5' end of the genome of

viruses that belong to the genus *Flavivirus* has a methylated type 1 nucleotide cap structure (Bisaillon & Lemay, 1997). Genomes of the other three genera have an internal ribosome entry site structure (IRES) on their 5' genome end (Tsukiyama-Kohara et al., 1992; Poole et al., 1995). Both of the 5' end structures assist in the initiation of translation (Turner et al., 2004). None of the *Flaviviridae* members has a polyadenylated 3' end tail, but the 3' end forms a loop structure which is highly structured in members of genus *Flavivirus* and less structured in the rest three genera (Turner et al., 2004). The genome is translated into a single polyprotein that is subsequently cleaved by host and viral proteases to produce three structural proteins (capsid, premembrane/membrane, and envelope) and seven non-structural (NS) proteins.

With currently 89 viral species, most of the members of *Flaviviridae* are known human and animal pathogens, such as Dengue virus, Yellow fever virus, Zika virus, West Nile virus, and many others. Viruses within the genera *Hepacivirus*, *Pegivirus*, and *Pestivirus* are vertebrate-specific. The genus *Flavivirus* is the most populated genus, with 53 species, and divided further according to host associations into tick- and mosquito-borne flaviviruses (arboviruses), no-known-vector flaviviruses (vertebrate-specific) and insect-specific flaviviruses (Blitvich & Firth, 2015). In the last decade, a new group of viruses with similar RdRp and NS3 amino acid sequences to flaviviruses has been discovered in insects and tentatively named "jingmenviruses". Potentially pathogenic for humans (Kuivanen et al., 2019; Wang et al., 2019), these viruses have a segmented linear genome and some present a polyadenylated tail at the 3' genomic end.

1.3.2.2 *Tolivirales*

The viral families *Carmotetraviridae*, *Luteoviridae*, and *Tombusviridae* compose the newly established order of *Tolivirales*. Viruses within this order are non-enveloped and infect predominantly plants. The genomes of these viruses are between 4–5.7 kb long, linear and in principle monopartite, with the exception of *Dianthovirus* (*Tombusviridae*) which is bipartite. In all families, neither a polyadenylated tail in the 3' end nor a 5' end cap structure are present, but the genera *Enamovirus* and *Polerovirus* of *Luteoviridae* have a viral protein attached to their genomes at the 5' end. Suppression of translation termination and ribosomal readthrough are observed in all members of *Tolivirales*, and some experience also leaky scanning.

Previously, these three families were not classified in any higher rank. In detail, *Carmotetraviride* is a monogeneric and monospecific family with a short history, since it was established in 2011 by *Providence virus*, a former virus of the former *Tetraviridae* family that showed more similarity to *Tombusviridae* than to *Tetraviridae* (Walter et al., 2010). The decision to classify *Providence virus* as a monogeneric and monospecific family came after the observation that it does not retain genome terminal non-coding regions in the same secondary structures observed in other tetraviruses (Gorbalenya et al., 2011). Also, its genomic organization and phylogenetic relationship as a distant sibling of tombusviruses and umbraviruses (currently classified as *Umbravirus* of *Tombusviridae*), denoted its relatedness to these viral groups (Gorbalenya et al., 2011).

The family *Luteoviridae* was previously often associated with viruses of the unclassified genera *Sobemovirus* and *Polemovirus*. Owing to similarities in their expressed polyprotein, the now-established family *Solemoviridae* (consisting of *Solemovirus* and *Polemovirus*) are mainly similar to the *Polerovirus* and *Enamovirus* genera of the *Luteoviridae* family. However, while *Luteoviridae* are under the umbrella of the *Kitrinoviricota* phylum, *Solemoviridae* belong to a different viral phylum: *Pisuviricota*. Also, interestingly the genus *Sobemovirus* shows similarity in the capsid protein to that of *Tombusviridae*. Nevertheless, the genus *Sobemovirus* could not be classified to either of the two families, since it does not share homology in the capsid protein of *Luteoviridae* or the RdRp of *Tombusviridae*. After all, the latest taxonomic rearrangement of the whole virus world relied on RdRp homology (at least for the Baltimore groups III, IV, and V), thus *Luteoviridae* were placed in the same viral order as *Tombusviridae* (ICTV Executive Committee and others, 2020).

Tombusviridae is a diverse virus family of infectious plant viruses that have monopartite genomic organization, with the exception of *Dianthovirus* whose genome is bipartite. Recently, similar viruses were discovered in invertebrates and terrestrial arthropods that show both mono- and bipartite genomic organizations (Shi et al., 2016a). The phylogenetic placement of tombusviruses deep within the invertebrate viruses has led to the formulation of the hypothesis that events of horizontal virus transfer (HVT) between plants and invertebrates have given rise to tombusviruses (Dolja & Koonin, 2018). The invertebrate tombus-related

viral diversity has been recently enriched by the discovery of insect tombus-related viruses (Paraskevopoulou et al., unpublished). **Chapter 4** contains a detailed description and discussion of these findings.

1.3.3 Hepatitis delta virus and other deltaviruses

The genus *Deltavirus* remains an unclassified viral taxon which counts so far only one species, Hepatitis delta virus (HDV). It occurs as a satellite virus to Hepatitis B virus (HBV) and is of great medical importance since upon superinfection of HBV-infected individuals worsens the severity of the disease (Rizzetto et al., 1983). Folded in an unbranched rod-like structure, the circular single stranded RNA genome of HDV is of negative-sense orientation and approximately 1,700 nucleotides long. Hepatitis delta virus was detected initially in 1977 (Rizzetto et al., 1977), it is transmitted mainly by parenteral routes, and occurs at a prevalence of about 4.5–10% in chronically HBV-infected humans (Stockdale et al., 2020; Chen et al., 2019). As a satellite to HBV, Hepatitis delta virus requires the HBV envelope for transmission since it does not encode for its own envelope. The single stranded RNA genome encodes two protein isoforms, the small and large hepatitis delta antigens, S-HDAg and L-HDAg, respectively.

1.3.3.1 Genome organization

The genome of HDV is a circular negative sense single-stranded RNA molecule. Apart from the non-coding region, two distinct domains are found on the HDV genome, the protein-coding region and the viroid-like self-cleaving ribozyme. The two viral proteins L-HDAg and S-HDAg are encoded by the same open reading frame (ORF) when an RNA-editing event takes place on the antigenomic RNA strand (Wong & Lazinski, 2002). The amber stop codon of S-HDAg is edited at the second position by the cellular host enzyme *adenosine deaminase acting on RNA 1* (ADAR1), converting the UAG to a UIG codon. In the next replication round, the UIG codon of the antigenomic strand is matched to ACC and further transcribed to UGG during mRNA synthesis. Consequently, the resulting mRNA encoding for L-HDAg is longer in comparison to that of S-HDAg by 60 nucleotides (19 amino acids).

Ribozyme regions are found on both genomic and antigenomic strands and self-cleavage occurs during genome replication resulting in single RNA copies cleaved from a multimeric

RNA molecule. These regions are approximately 80 nucleotides long and each folds into a nested double pseudoknot structure, cleaving the RNA at nucleotide positions 685/686 and 900/901 respectively (Kuo et al., 1988). The double pseudoknot structures consist of two coaxial helical stacks held together on one end by two single-stranded joining strands and on the other end by two base pairs of one of the helical strands (see **Fig. S4** in **Appendix of Chapter 3** and Webb & Lupták, 2011). Nevertheless, it is yet unclear whether the ribozyme regions recruit cellular ligases or whether the ribozymes themselves possess ligase activity (Sharmeen et al., 1989; Reid & Lazinski, 2000).

HDAg contains a nuclear localization signal which is responsible for introducing the HDV-RNA into the nucleus. The NLS is located between residues 66–75 on HDAg (Alves et al., 2008) and it resembles the NLS of other nuclear proteins (Chao et al., 1991). The conserved basic amino acid domain is 66-EGAPPAKRAR-75, with the glutamic acid residue at position 66 playing an essential role for the nuclear import (Alves et al., 2008).

The exact nuclear export mechanism for the human HDV-RNA molecules has not been clarified yet, but experimental evidence points to a proline-rich region (amino acids 198–210 on HDAg) as the primary nuclear export signal (NES) (Lee et al., 2001; Macnaughton & Lai, 2002). Pro-205 of L-HDAg is critical for the NES function which directs L-HDAgs to the cytoplasm via a chromosome region maintenance 1 (CRM1)-independent pathway, a transport factor at nuclear pores (Lee et al., 2001). However, nuclear export of genomic HDV-RNA is independent of L-HDAg and because genomic HDV-RNA is synthesized by Pol-II, there is a hypothesis that it may be exported by the same mechanism as for splicing-dependent cellular mRNAs (Macnaughton & Lai, 2002).

HDAgs form oligomers in the cytoplasm and these formations have initially been suggested to facilitate nuclear transportation via the leucine zipper (Xia et al., 1992). Leucine positions correspond to amino acid positions 30, 36, 44, 51, 108, 115, and 122 on HDAg (Chen et al., 1992), but may slightly vary among HDV-RNAs. However, it was concluded that it is rather the N-terminus coiled-coil domain (CCD) and not the leucine zipper which is essential for forming protein complexes *in vivo*, as Leu-mutants were possible to transport into the nucleus (Xia et al., 1992; Chang et al., 1992).

The CCD at the N-terminal of HDAG is responsible for the formation of dimers/multimers, as well as for activating/promoting and inhibiting RNA replication (Chang et al., 1992; Lazinski & Taylor, 1993). It is located at residues 13–48 on HDAG and forms heptad repeats with hydrophobic residues at the second and sixth heptad position (Chang et al., 1992). In fact, L-HDAG does not require an RNA-binding domain to access the HDV-RNP, but can act via a coiled-coil interaction with S-HDAG and move the RNP from the nucleus to the cytoplasm (Chang et al., 1992).

Two arginine-rich motifs (ARMs) are present at amino acid positions 97-107 and 136-146, and have been claimed to be required for HDAG binding activity (Lee et al., 1993). However, experimental evidence has shown that mutants of the core Arg of either ARM1 or ARM2, did not lessen binding to HDV-RNA (Daigh et al., 2013). The HDAG displays three properties which likely contribute to non-specific nucleic acid binding: i) high positive charge, ii) oligomeric structure, iii) high disorder degree in more than half of the protein. Therefore, there are challenges in identifying the protein regions involved in HDAG binding HDV-RNA (Daigh et al., 2013).

HDV-RNA has the ability to fold itself into an unbranched rod-like structure due to its > 70% base pairing (Wang et al., 1986), a property also observed in some plant viroids (Taylor & Pelchat, 2010).

1.3.3.2 Replication

No polymerase is encoded by the HDV genome and thus HDV depends on cellular host polymerases for replication (MacNaughton et al., 1991; Filipovska & Konarska, 2000; Modahl et al., 2000; Li et al., 2006). Upon entering the cell, the HDV-RNA genome is packaged together with S-HDAG and L-HDAG in a ribonucleoprotein (RNP) formation (Ryu et al., 1993). The RNP complex gets transported to the nucleus using probably the NLS on S-HDAG (Chou et al., 1998; Tavanez et al., 2002). Both genomic (G) and antigenomic (AG) RNA replication cycles take place in the nucleus under a double rolling replication scheme (Branch & Robertson, 1984). During the replication cycle, both G- and AG-RNAs form multimers that are further self-cleaved by the HDV ribozyme structures (Kuo et al., 1988). The exact mechanism of subsequent ligation and circularization of both G- and AG-RNAs

remains undetermined.

G-RNA and mRNA are synthesized in the nuclear bodies, whereas AG-RNA is replicated in the nucleolus (Hong & Chen, 2010; Huang et al., 2008). Interestingly, evidence supports that S-HDAg recruits Pol-II for G-RNA replication by functionally interacting with its clamp formation to receive RNA instead of DNA template (Yamaguchi et al., 2007). Very recently Pol-II co-activation has been attributed to a possible histone mimic feature of S-HDAg (Abeywickrama-Samarakoon et al., 2020). During transcription, S-HDAg probably displaces the negative elongation factor (NELF), thus promoting mRNA-synthesis elongation (Yamaguchi et al., 2001). It is yet unknown whether it is Pol-I or -III that transcribes AG-RNA, however, experimental evidence points at a different cellular machinery to that used for processing mRNA and G-RNA (Lucifora & Delphin, 2020).

1.3.3.3 Packaging

After exiting the nucleus, G-RNA clusters with both antigens to a newly synthesized RNP. The complex travels to the endoplasmic reticulum (ER) where it finds and gets packaged into an HBV envelope. This process is facilitated by a post-translational modification on the carboxyl terminal cysteine-211 of L-HDAg (Hwang & Lai, 1993). Addition of an isoprenyl group to the terminal cysteine residue by the cellular *farnesyltransferase* (FTase), enables the RNP to get anchored to ER and bind the HBsAg envelope proteins (Otto & Casey, 1996). Cysteine is a universally conserved residue across all HDV genotypes, followed by another conserved glycine residue two positions downstream, constituting the L-HDAg conserved C-terminal -CXXQ motif. Recent experimental evidence shows that viruses other than HBV can provide envelope proteins which successfully produce infectious HDV particles in cell culture (Perez-Vargas et al., 2019). Nevertheless, farnesylation of the C-terminal -CXXQ motif of L-HDAg remains a crucial requirement for virion packaging, since particle production was prevented after experimental inhibition of this pathway (Perez-Vargas et al., 2019).

1.3.3.4 Theories on the evolutionary origins

To date, neither a theoretical nor an experimental frame have been able to delineate how HDV originated, making it the only RNA virus with an unknown origin. Nevertheless, several theories have made suggestions on the potential origins of HDV (Littlejohn et al., 2016).

One theory proposes that a viroid-like RNA originated in a cell infected with HBV, owing to the HBV coinfection requirement for packaging and release (Brazas & Ganem, 1996). A cellular protein, *delta-interacting protein A* (DIPA), has been found to interact with HDVAg and thus sparked the assumption that a viroid-like RNA captured a host-derived DIPA-encoding mRNA or an ancestral form of DIPA (Brazas & Ganem, 1996). Even though plant viroids have tiny genomes of 200–400 nucleotides, they exhibit some similarity to HDV on the basis that, like HDV, they also recruit host RNA polymerase II (Pol-II) and interact with some cellular proteins (Elena et al., 1991; Sikora et al., 2009). Experimental findings have provided some results on the viroid-precursor hypothesis as several genetic engineering experiments in plants show cytopathic effects when HDV multimer RNAs are inoculated into plant leaves. Also, plant viroids are shown to successfully replicate in animal cells, but only when complemented by HDVAg expression (Taylor & Pelchat, 2010). The similarity of plant viroids to HDV further lies on the circularity of their genomes and the self-cleaving ribozyme activity of some viroids (Taylor, 1999; Diener, 2001). Nevertheless, the origin of viroids and the evolutionary relationships among them are obscure (Flores et al., 2009). Therefore only independent convergent evolutionary scenarios can be proposed on this basis (Taylor & Pelchat, 2010).

The discovery of two self-cleaving RNAs in mammals (CLEC2 and CPEB3) triggered a different evolutionary theory for the origin of HDV, suggesting that HDV arose directly from the human transcriptome (Martick et al., 2008; Salehi-Ashtiani et al., 2006). Before discovering those two self-cleaving RNAs, HDV was thought to be the only self-cleaving RNA associated with humans. Even though one of the ribozymes, CPEB3, is structurally and biochemically related to the HDV ribozyme, none of the two mammalian ribozymes show any sequence similarity to that of HDV (Littlejohn et al., 2016).

Another proposal for the origin of HDV stems from the identification of circular host-derived mRNAs in cells (Taylor & Pelchat, 2010; Taylor, 2014). Based on this theory, a rare RNA circle in an HBV-infected hepatocyte might have been selected due to its ability to replicate using host Pol-II (Littlejohn et al., 2016). To a certain extent, the HDV genome organization corroborates the latter theory, since the HDV ribozyme is preceded by a poly(A) signal, an arrangement considered to be typical for host mRNAs as well (Taylor & Pelchat, 2010;

Littlejohn et al., 2016; Hsieh & Taylor, 1991). Candidates for this theory are suggested to be either HBV-spliced mRNAs or HBV-transcribed DNA sequences (Littlejohn et al., 2016).

All the above theories come to weak ground after the recent discovery of several delta-like viral agents in mammals, snakes, and other vertebrates and invertebrates (Wille et al., 2018; Hetzel et al., 2019; Chang et al., 2019; Paraskevopoulou et al., 2020; Edgar et al., 2020; Bergner et al., 2020). **Chapter 3** unfolds the discovery of rodent deltavirus (RDeV), the only non-human mammalian deltavirus known until now. This virus together with snake deltavirus (SDeV) are the only non-human deltaviruses for which there is demonstrated proof of replication in a cell culture set up. Other non-human deltaviruses have so far been reported in NGS metagenomic studies. Nevertheless, all those findings bring a whole new perspective to the theories of HDV origin.

1.3.3.5 Epidemiology and pathogenesis

Eight distinct clades are assigned to HDV which has a worldwide distribution (Deny, 2006). Interestingly, HBV is also composed of eight genotypes but the two geographical distributions are only slightly correlated (Taylor & Pelchat, 2010). The spatial distribution of the eight different HDV genotypes displays a moderate structure. **Fig. 1** shows the worldwide distribution of HDV genotypes and their sequence prevalence.

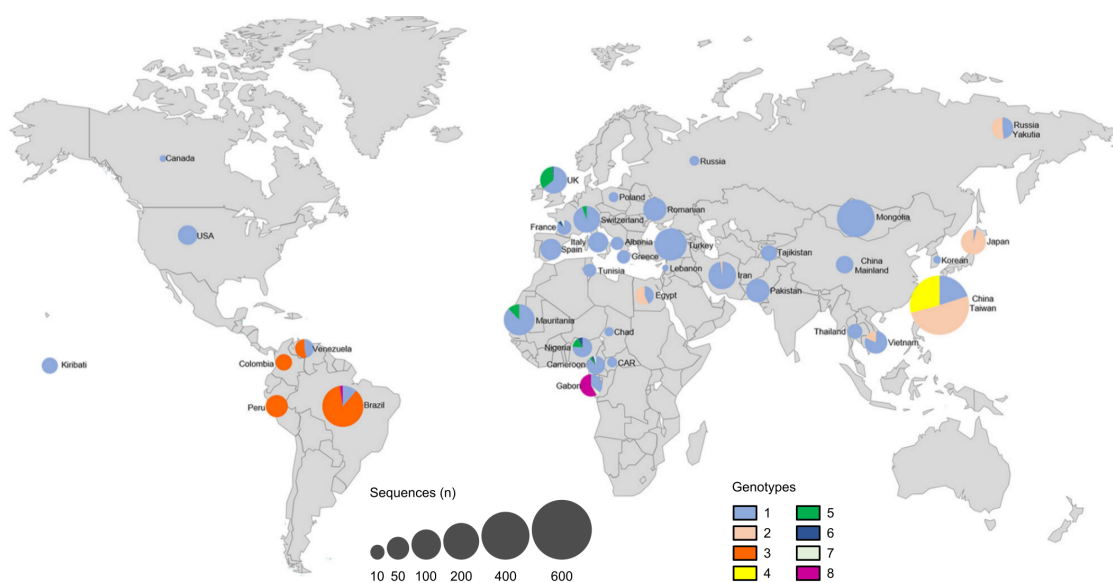


Figure 1: Global distribution of HDV genotypes and sequence prevalence. Figure modified from Chen et al. (2019).

Genotype 1 is found worldwide, but the other genotypes have more localized ranges (Stockdale et al., 2020). Genotype 2 occurs in Australia, the Yakutia region (Russia), Taiwan and Japan, together with genotype 4 which is also found in Japan and Taiwan. However, genotype 2 has been recently reported from Egypt and Iran (Chen et al., 2019). Genotype 3 is the most divergent and found in Latin America, mainly the Amazon Basin. Genotypes 5, 6, 7, and 8 are predominantly found in Africa, as well as in individuals in the UK, France, and Switzerland (Chen et al., 2019). Recently, HDV sequences from Australia were assigned to genotype 5 as well (Jackson et al., 2018).

Among the HBsAg-positive people, more than 12 million are exposed to HDV as the lowest estimated prevalence is 4.5% (Stockdale et al., 2020). HBV-HDV coinfection is considered a more severe form of viral hepatitis as a simultaneous infection can cause extensive hepatic necrosis or even fulminant hepatitis (Smedile et al., 1982; Vieira et al., 2017). When encountered as a superinfection in chronic HBV patients, HDV has been reported to persist, causing cirrhosis progression and presenting a higher risk of hepatocellular carcinoma relative to chronic HBV alone (Fattovich et al., 2000). A large proportion of HDV patients (70–90%) are HBeAg negative and most have reduced HBV-DNA serum (Hughes et al., 2011; Lutterkort et al., 2018). Measurements of the HDV viral load show that it fluctuates according to the stage of infection, yet it is not associated with liver injury (Zachou et al., 2010). Coinfection and superinfection of HBV-infected individuals with HDV has been found to generally downregulate HBV replication, as both S- and L-HDAgs suppress the activity of the two HBV enhancer regions (Gordien et al., 2001; Williams et al., 2009). Additionally, HDAgs transactivate the interferon-inducible MxA gene, thereby inhibiting HBV replication due to the reduction of viral mRNA nuclear export (Gordien et al., 2001; Williams et al., 2009). HBV can only start replicating once HDV infection is cleared. Nevertheless, a detailed HDV pathogenesis mechanism remains still unclear and the adaptive immune response is poorly defined (Hughes et al., 2011).

Woodchucks with actively replicating Woodchuck hepatitis virus (WHV) can also get infected with HDV (Negro et al., 1989; Freitas et al., 2012), but this remains yet to be confirmed in the wildlife.

The relative low awareness for HDV infection designates hepatitis D as a neglected disease despite its worldwide distribution (Stockdale et al., 2020). Awareness around hepatitis D varies widely owing to various factors, such as standardization issues in molecular diagnostics, incomplete testing among people positive for HBsAg, and lack of effective treatments (Stockdale et al., 2020).

1.3.3.6 Other deltaviruses

The discovery of non-human deltaviruses began about two years ago, with the first findings communicated in September 2018 on the pre-print server *bioRxiv*. Delta-like viral agents have so far been found in blood, tissues, and fecal samples of a rodent (*Proechimys semispinosus*) (Paraskevopoulou et al., 2020 - **Chapter 3**), tissues of boas (*Boa constrictor*) and a water python (*Liasis mackloti*) (Hetzl et al., 2019), cloacal and oropharyngeal transcriptome-derived data of teals and ducks (Wille et al., 2018), in the transcriptomes of various vertebrates and invertebrates (Chang et al., 2019; Edgar et al., 2020), and very recently in saliva and blood of bats (Bergner et al., 2020). These viral agents will be referred with their tentative names: rodent deltavirus (RDeV), snake deltavirus (SDeV), and *animal-Name* deltavirus (DeV) for the set: {duck-associated, fish, newt, termite, toad, bat, deer, marmot}. A common characteristic among all non-human deltaviruses is the lack of a C-terminal -CXXQ motif in the expressed protein. Whether RNA editing on the antigenomic strand takes place at all, resulting in the production of two protein isoforms, remains to be confirmed for some of these viral agents. However, RDeV, SDeV, and bat deltavirus present amino acid tails before the appearance of the next stop codon, of 19, 22, and 28 residues respectively.

Rodent deltavirus was discovered in NGS data of blood samples of *Proechimys semispinosus*, a rodent species inhabiting the neotropical regions in central America (Paraskevopoulou et al., 2020). The presence of RDeV was further confirmed by real-time RT-PCR in all available tissue samples from five organs (heart, liver, lung, kidney, small intestine), and in some fecal samples. The large animal sample size allowed for statistical analyses of correlation to organismic, as well as environmental factors, demonstrating pronounced RDeV detection in reproductive male rodents. RNA structure prediction using *mfold* Zuker (2003) revealed a high sequence complementarity and a rod-like secondary structure for RDeV shown in **Fig. 2**. Antibody detection against rodent delta antigen in the rodents' serum was confirmed

by an indirect immunofluorescence assay (IF) and protein reactivity on a Western blot. Viral replication and protein production have been proved in a cell culture system where clonal expansion of the transfected cells was also observed. Antigenomic RNA editing and subsequent large delta antigen production have not been experimentally documented *in vitro* or *in vivo*. No hepadnavirus-coinfection was reported and hepacivirus was excluded as a potential viral cofactor.

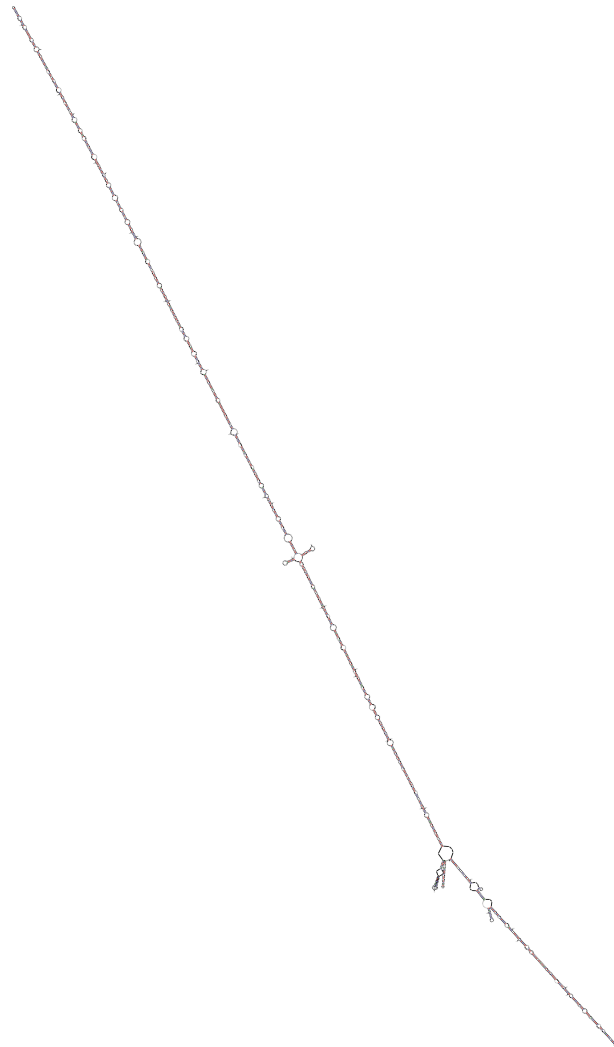


Figure 2: Predicted rod-like secondary structure of RDeV RNA; free energy is $dG = -839.02$.

The discovery of snake deltavirus occurred also during an NGS study of brain samples of *Boa constrictor* snakes which showed signs of a nervous disease (Hetzl et al., 2019). Subsequent RT-PCR and tissue immunohistology revealed presence of SDeV and snake delta antigen in brain, liver, lung, kidney, and spleen. In addition, SDeV RNA was detected in four boa constrictor offspring and a water python (*Liasis mackloti*) that was housed in the

same room for several years. SDeV was isolated from brain tissue and proved to replicate in cell culture. SDeV has been additionally shown to produce infectious particles with envelope proteins of reptarenaviruses and hartmaniviruses (Szirovicza et al., 2020).

NGS data (meta-transcriptomic data in particular), were the initiation points for all other deltavirus-like discoveries. However, those studies described sequence-derived characteristics and until now have not involved further experimental work. The duck-associated deltavirus was identified in combined cloacal and oropharyngeal samples of three different species of teals and ducks (genus *Anas*) (Wille et al., 2018). Termite deltavirus stems from whole body extractions of individuals that belong to the species *Schedorhinotermes intermedius*, newt deltavirus was assembled from gut and liver samples of *Cynops orientalis*, and NGS reads of fish and toad deltaviruses were found in pooled samples of jawless fish gills and amphibian lung tissues respectively (Chang et al., 2019). Bat, deer, and marmot deltaviruses were very recently reported in a pre-print article that presents a bioinformatic scheme for virus discovery in metagenomic data (Edgar et al., 2020). Deltaviruses in a frog and a bird are also reported in the same article, however sequence information remains currently unavailable. In parallel, the pipeline presented by Edgar et al. (2020) was utilized by Bergner et al. (2020) who presented an analysis on mammalian deltavirus origins and diversification. The bat, deer, and marmot deltaviruses were presented in the second article where genome features are described in detail. Also, bat deltavirus was detected within own lab samples of bat saliva and blood. HBV-coinfecting agents are absent, but the authors suggest hepaciviruses and poxviruses as potential sources of helper viral envelopes (Bergner et al., 2020). Yet sequence information is to be expected in the near future. Among all metagenomically-derived deltaviruses, an ORF that encodes a small delta antigen equivalent and predicted ribozyme structures were reported, with additional predictions of high degrees of genome base-pairing.

1.4 *In silico* virus detection

Commonly used NGS technologies generate short read fragments (150–300 bases) by sequencing a nucleic acid sample that has previously been fragmented and tagged with sequence tags (*adapters*) attached to the fragments during library preparation (Head et al., 2014). Processing the raw reads requires the following steps: demultiplexing, adapter removal, and quality check of the reads based on Phred quality scores. The processes of demultiplexing (i.e. sorting the reads by adapter group) and adapter removal are often offered by the sequencing facility. Phred quality scores are logarithmically linked to error probabilities of incorrect base calling during sequencing and are utilized in the read assembly step to assess the accuracy of consensus sequences (Ewing et al., 1998). For example, a base with Phred score 30 has a probability of 1 in 1000 of having been inaccurately called. Software for initial read quality control includes among others: FastQC (Andrew, 2010), Skewer (Jiang et al., 2014), Lighter (Song et al., 2014), and Trimmomatic (Bolger et al., 2014).

Together with the viral reads, host background sequence information inherently exists in any given sample. Read-mapping software, such as BWA (Li & Durbin, 2009) or bowtie2 (Langmead & Salzberg, 2012), is employed to find the reads that belong to the host using matching algorithms to map all raw reads on the host genome and filter out the matching ones. In this step, the availability of host genomes is a considerable bottleneck because a sequenced genome exists only for some thousands of eukaryotic organism species. Using the genome of a closely-related host organism can circumvent this issue, but a right selection might be several taxonomic ranks away. Nevertheless, host background sequence removal is a crucial step to avoid erroneously assembled chimeric viral sequences.

When a known viral genome exists, genome assembly follows a read-mapping strategy, similar to the one described above for removing host background reads. On the other hand, *de novo* assembly software is employed for *de novo* (i.e. without prior sequence knowledge) genome assembly, merging reads into a contiguous sequence (*contig*). Software for *de novo* assembly relies on two kinds of algorithms: graph assembly and greedy algorithms. Graph assembly algorithms, like de Bruijn graphs (DBG), use a k -mer approach where a read is split in shorter k -mer fragments which serve as nodes in the graph assembly. Read overlap is gradually achieved when two nodes overlap by at least a $(k-1)$ -mer and thus connected

to each other with edges. Software employing DBG algorithms includes Trinity (Grabherr et al., 2011), SOAPdenovo (Luo et al., 2012), Velvet (Zerbino & Birney, 2008), and SPAdes (Bankevich et al., 2012). Greedy algorithms, such as overlap layout consensus (OLC), follow a gradual assembly approach by clustering reads that overlap based on their pairwise distance and iteratively assembling them in contigs. An important metric in assessing the quality of results is *coverage* (or *depth*) which corresponds to the number of unique reads representing a given nucleotide. The OLC approach is more effective when applied in long reads with low coverage as opposed to the DBG algorithm that is more suitable for short reads with high coverage (Li et al., 2012).

Usually, functional genome annotation and taxonomic classification follow after viral genome assembly. Both of these processes rely on sequence comparison methods using matching algorithms. The most widely-used tool for this purpose is the basic local alignment search tool (BLAST, Altschul et al., 1990) which is incorporated in the NCBI database, the most updated online sequence database. BLAST uses a heuristic algorithm that breaks the query sequence into k -mers and searches for sequence matches in its database by gradual alignment. High-scoring pairs are reported and this process is iterated increasing the k -mer size step-wise. Statistic metrics, such as E-value and bit score, are advised to evaluate the significance of the results.

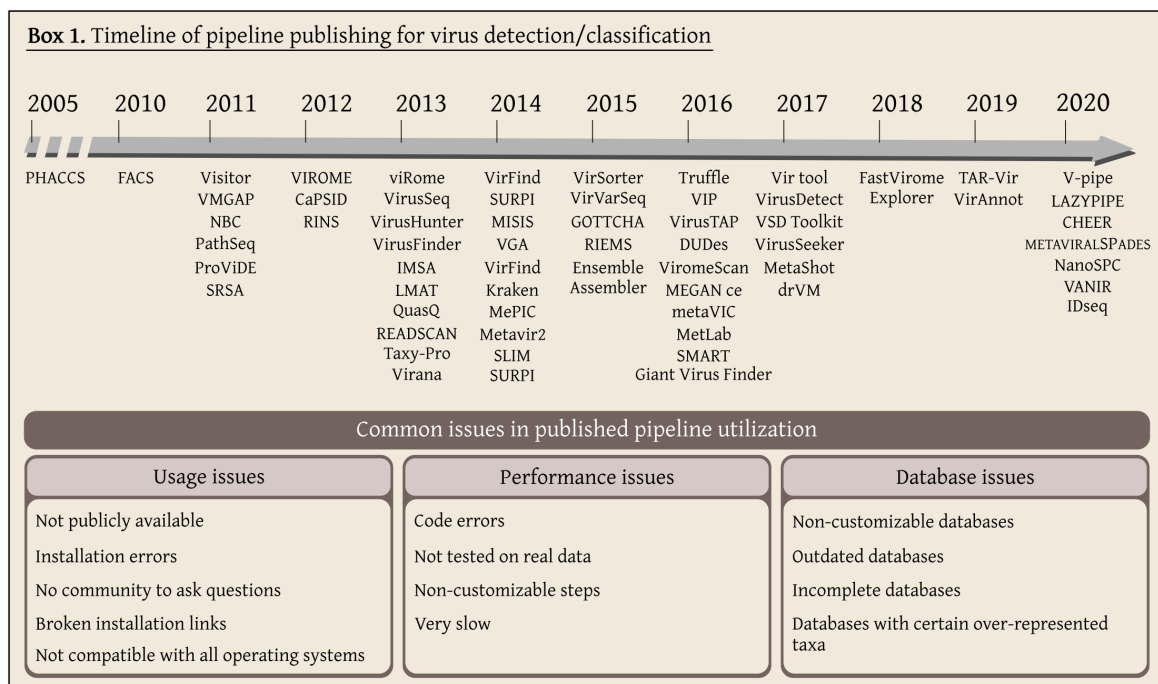
Combining the BLAST search results with a protein prediction tool provides a powerful framework for protein and protein domain annotation. Reliable annotations of not just protein families, but also sites and motifs, are facilitated by InterProScan (Jones et al., 2014). This tool queries the InterPro database, which is an integrative collection of the largest protein databases, such as Pfam (El-Gebali et al., 2019), PROSITE (Sigrist et al., 2012), and PHOBIUS (Käll et al., 2004). For instance, Pfam is the largest database of annotated protein families that currently hosts over 18,000 protein families.

Whereas functional genome annotation can rely on BLAST and InterProScan to a great extent, decisions on taxonomic classification are a more complicated issue. On this matter, the practice of virus identification and classification in environmental samples differs substantially from that followed when biological material is sampled from individual organ-

isms. In the first case, the goal usually is to characterize the underlying viral diversity in e.g. a given soil or marine sample, rather than assemble and annotate novel viral genomes. This type of analysis, aims at taxonomic classification at the read level (termed *taxonomic profiling*) without assembling the reads in contigs/genomes, and employs binning methods for grouping the reads into bins (methods reviewed by Mande et al., 2012). On the other hand, taxonomic classification of assembled novel viral genomes requires careful assessment that has to involve at least comparisons of genome organization to known viruses and phylogenetic reconstruction processes. A more detailed introduction on this topic is found in **Chapter 1.5**.

1.4.1 Published pipelines & common issues in bioinformatic analyses

During the last ten years over 60 bioinformatic pipelines have been published, offering potentially unified working frameworks for virus discovery (summarized in **Box 1**, and reviewed by Nooij et al. 2018). Despite this large number, the computational software employed by published pipelines is narrowed to a few tools for sequence analyses. The bioinformatic pipeline diversity for virus discovery and taxonomic classification in fact reflects the specificity of the respective research goals in question. As a consequence, construction of own pipelines and customization of the bioinformatic steps is a meaningful process worth investing.



1.4.2 Computational approach and methodology applied herein

The most challenging aspects in the process of computational virus identification involve processing the raw NGS data and, mainly, interpreting the results. Both are particularly challenging in cases of novel virus discovery where prior sequence information is not available. Normally, raw NGS reads would be mapped on a known viral genome which would serve as the basis to assemble contigs of the viral genome at question. Afterwards, sequence and ORF organization comparisons to related viruses would assist in annotating the assembled viral genome (assembly) and assigning it to a taxonomic group. In this thesis, two different approaches were followed for processing the NGS data: identifying novel viruses a) within raw NGS data (**Fig. 3A**) and b) within NGS data that were already assembled into contigs (**Fig. 3B**).

In the first approach, bioinformatic software for sequence analysis was assembled in a pipeline scheme as depicted in **Fig. 3A**. The nucleic acid extraction method of the samples examined with this approach followed a protocol of total RNA extraction (for details consult the **Appendix of Chapter 3**). Processing the NGS data required initially a read quality control step, necessary to remove low quality reads with incorrect base callings. For this purpose, the software Lighter was used (Song et al., 2014). Raw NGS data consisted of paired-end reads, thus a pairing step was applied using FLASH (Magoč & Salzberg, 2011). FLASH yields three sets of reads: the paired, the unpaired forward, and the unpaired reverse. Unpaired reads may result due to the initial quality check step, as their pair may have been removed because of low quality Phred scores. All the three sets of reads were combined before moving to the next step. Filtering the data for host-derived reads was applied with a subsequent step of read-mapping on the respective host genome with the BWA algorithm (Li & Durbin, 2009). Host mapped and unmapped reads were separated from each other using samtools (Li et al., 2009) and the unmapped set was used in downstream analyses. Afterwards followed a step of *de novo* assembly of the reads into contigs, based on a DBG method with the software SPAdes (Bankevich et al., 2012). The resulting contigs were subjected to additional elongation by using a second assembly algorithm: CAP3 (Huang & Madan, 1999). The last step included sequence comparison against viral databases with the programs BLAST+ (Camacho et al., 2009) and diamond (Buchfink et al., 2015). The RVDB database was used as a reference database; RVDB is a curated and regularly updated database for enhancing

virus detection using NGS technologies (Bigot et al., 2019). In some cases, the unmapped reads were included in the last sequence comparison step in order to spot reads that were potentially of viral origin, but were too few to make a contig in the *de novo* assembly step (shaded step in **Fig. 3A**). This pipeline was built for the purpose of novel virus identification within the project described in **Chapter 3**.

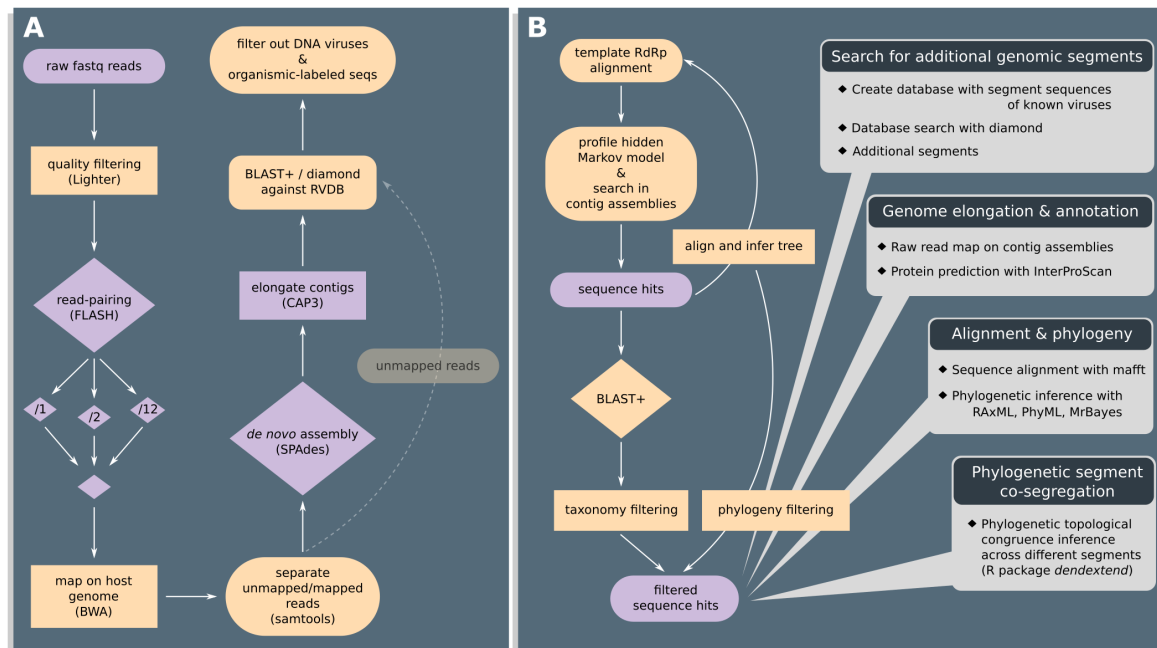


Figure 3: Pipelines for approaches (A) and (B). Steps with data generation are colored in purple and steps with data analysis in yellow. Optional steps appear shadowed. Word bubbles in B indicate downstream analyses.

The second approach (shown in **Fig. 3B**) follows an entirely different scheme, as raw NGS reads were already assembled into contigs beforehand by the 1000 Insect Transcriptome Evolution Project (1KITE, <https://1kite.org/>). Also, the sequence data generation used an RNA isolation protocol targeting the mRNA and therefore extracting all sequences undergoing transcription, i.e., the organism's *transcriptome*. Raw NGS reads were generated, checked for contamination/quality control, and were *de novo* assembled into contigs (consult the Supplementary Material by Misof et al. (2014) for details).

The goal of this approach was different to the one described in **Fig. 3A**. While novel virus identification without prior knowledge was sought in the case described in **Fig. 3A**, the pipeline described in **Fig. 3B** was built to detect distant homologs of known viruses. For this purpose, profile hidden Markov models (pHMMs), focused on the RNA-dependent RNA

polymerase (RdRp) region, were initially employed and built for each respective virus family of interest (see **Chapter 1.3**). Briefly, a pHMM receives a sequence (*template*) alignment as input and creates a "position-specific scoring table" that represents frequencies of elements (nucleotides/amino acids) for each column of the alignment (Gribskov et al., 1987). Using this probabilistic model as a guide, database sequence matching is performed, accounting for conserved and variable regions in the alignment at the same time (Karplus et al., 1998). The delivered sequence results (*hits*) represent both closely-related sequences and distant homologs. For this reason a filtering step followed, in order to sort out hits that are too distant and might in fact belong to a different virus family. This step appears as "phylogeny filtering" in **Fig. 3B** and included a) aligning the hits to the initial template alignment and removing hits that were too short and did not cover the full alignment length, and b) inferring a phylogeny to spot "outliers" that caused long branch attraction phenomena. At the same time, all hits are compared to the viral database with the BLAST tool to find their closest taxonomic relatives, a step called "taxonomy filtering". The combination of both filtering steps resulted in the final hit library that was utilized in downstream analyses.

Because some viral genomes consist of multiple segments (see **Chapter 1.3**), non-RdRp segments would not be captured by RdRp-pHMMs. To identify non-RdRp segments, databases that consisted solely of non-RdRp segments of known viral sequences were constructed. A diamond search within these databases identified contigs that constituted the missing additional non-RdRp segments for the respective viral hits. Diamond searches are exclusively amino acid searches and they have the advantage of bypassing the degenerated genetic code's drawbacks by relying on amino acid identity. Also, it has been shown that diamond operates on a significantly lower computation time compared to its blastx counterpart (Buchfink et al., 2015).

In some cases the delivered RdRp hits were smaller than coding-complete, therefore 5'- and 3'-end sequence elongation was applied by mapping raw reads on the initial contig assemblies. With this process, potential assembly errors were corrected too, as raw reads were mapped on the complete contig length. Genome annotations were carried with InterProScan for protein domain and motif identification (Jones et al., 2014).

Amino acid alignments of the main viral proteins (i.e., RdRp, nucleocapsid, glycoprotein) were calculated separately for every virus family with the software mafft (Katoh & Standley, 2013). The E-INS-i algorithm was selected because it operates iteratively and tries to align regions with conserved motifs in long and unalignable regions, such as the RdRp. For phylogenetic tree inference, three different programs were utilized: RAxML (Kozlov et al., 2019), PhyML (Guindon et al., 2010), and MrBayes (Ronquist et al., 2012). RAxML and PhyML are among the most widely-used maximum likelihood methods that however differ in their topology search algorithms (detailed performance comparison by Zhou et al. 2018). On the other hand, MrBayes is a bayesian method for tree inference. Phylogenies inferred with all three methods are provided in the **Appendix of Chapter 2**, where differences in topology, clade support, and branch lengths can be observed.

For segmented viruses, reassortment is a major means for diversification (reviewed in Lowen 2018) that results in recombined genomic segments or even exchanges of entire segments. In non-segmented viruses, genes experience different selection pressures that result in different evolutionary rates; recombination events also take place. To grasp signatures of divergent evolution, comparisons of topologies between different protein phylogenies were performed. Topological incongruence was visualized with the R package *dendextend* (Galili, 2015), and examples are available in the **Appendix of Chapter 2** and the **Appendix of Chapter 4**.

1.5 Virus taxonomic classification and the ICTV

Biological classification as we know it today received its foundation in the 18th century when Carl Linnaeus published his book "*Systema Naturae*" (von Linné, 1735), where he introduced a taxonomic system based on binomial nomenclature. Organisms are classified based on similarities that they share in a scheme of ranked hierarchy that includes the following major ranks (from higher to lower): domain, kingdom, phylum, class, order, family, genus, and species. The darwinian theory of evolution introduced the concept of similarity by descent, which subsequently contributed the recognition of fossils as members of the classification. With today's technological advances in molecular biology, taxonomic classification additionally consults molecular phylogenetics for taxonomic assignments. This has often introduced conflict and led to rearrangements in the taxonomy because of the discovery of *cryptic species* (i.e. morphologically similar species that are nonetheless reproductively isolated). However, because viruses are not regarded as "living organisms", they have historically not been included in the above effort. In the virus world, the Baltimore classification (described in **Chapter 1.3**) is a widely-accepted scheme. Nevertheless, as it is based on the structure of the virus genetic material, it does not necessarily reflect the evolutionary relationships among viruses.

A taxonomic scheme similar to the linnaean system was established in the '60s by the International Committee on Taxonomy of Viruses (ICTV). The ICTV is the official virus nomenclature body that develops, establishes, and maintains the official index of virus taxonomy. A brief history and summarized related information has been recently published by Kuhn (2020). According to ICTV, "*a species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria*". Such criteria have historically relied on virus antigenic relationships, virion structure, pathogenicity, epidemiology, and host range. With the advent of virus metagenomics and the discovery of numerous novel viruses that are evolutionary distinct from established species, the ICTV has defined criteria for the inclusion of metagenomic viral sequences in the taxonomy (Simmonds et al., 2017).

This decision came inevitably, as the number and diversity of viral sequences identified in metagenomic data already exceeds the number of experimentally-characterized virus isolates

(Simmonds et al., 2017). As long as metagenomic sequences are phylogenetically related to established viral ranks and present a set of characteristics, such as genome organization and homologous genes to the respective viral rank, they can be considered for official classification (Simmonds et al., 2017). Adding to that, a number of information and characteristics that should accompany any metagenomically-derived uncultivated virus was published as the "minimum information about an uncultivated virus genome" (MIUViG) (Roux et al., 2019). These refer to technical characteristics on the sequence generation methods and include among others the method/software used for sequence assembly and the quality, number of contigs, and annotation status of the assembly (Roux et al., 2019).

Also, until very recently mostly lower taxonomic ranks were common in virus taxonomy, such as species, families, and orders. Since 2019, the realm *Riboviria* was established as a unifying rank that includes all RNA viruses (Walker et al., 2019), yet lower ranks remained overall disconnected. Often, for large virus groups the suffix "supergroup" was utilized, e.g. alphavirus-supergroup or flavivirus-supergroup, instead of taxonomic assignment to higher ranks. Efforts to process the large flow of metagenomic virus sequence information in the frame of virus taxonomy resulted in the construction of computational pipelines, such as GRAViTy for eukaryotic viruses (Aiewsakun & Simmonds, 2018) and vConTACT v.2.0 for prokaryotic viruses (Jang et al., 2019). While vConTACT v.2.0 was built for taxonomic assignment of uncultivated prokaryotic virus genomes at the genus level, GRAViTy examines the phylogenetic relationships of family and above ranks within the Baltimore groups.

A detailed and complete taxonomy rearrangement for the entire virus world with the suggestions of linnaean-like high rank establishment was formulated by Koonin et al. (2020). This proposal was approved by the ICTV which subsequently published the taxonomy rearrangement (ICTV Executive Committee and others, 2020). The new taxonomic classification consists of four realms, that is the equivalent level of the "domain" rank in cellular organism taxonomy:

- *Duplodnaviria* - viruses with double-stranded DNA genome
- *Monodnaviria* - viruses with single-stranded DNA genome
- *Riboviria* - viruses that encode an RdRp or an RT
- *Varidnaviria* - viruses with DNA genome encoding a major capsid protein of vertical

jelly-roll fold and pseudohexameric capsomeres; also those viruses that have secondarily lost the capsid protein

In contrast to the Baltimore classification scheme, the ICTV taxonomy reflects the evolutionary relationships among viruses. Interestingly, only few Baltimore classes maintained their integrity within the current ICTV taxonomic re-arrangement: the phylum *Negarnaviricota* that hosts all group V viruses (single-stranded negative sense RNA viruses) and the order *Blubervirales* which includes all group VII viruses (double-stranded DNA-RT viruses with RNA intermediate), both classified within *Riboviria*. Viruses of Baltimore group IV (single-stranded positive sense RNA) although monophyletic when examined alone, became paraphyletic to double-stranded RNA and single-stranded negative sense RNA viruses after the inclusion of the last two in the phylogeny (Wolf et al., 2018b).

1.6 Aim of this thesis

The objective of this thesis is to use computational tools for the discovery of new RNA viruses in metagenomic data, with a focus on insect and mammalian viruses. Special attention is given on detecting evolutionary homologs that are distantly related to viruses known so far.

The detection of diverse viruses will be achieved by combining computational tools in a unified bioinformatics pipeline to analyze next-generation sequencing data (presented in **Chapter 1.4.2**). With the resulting sequences in hand, computational virus identification will be possible, enabling unknown virus characterization. Given the size of the insect collection to be screened and the number of samples for mammalian virus identification, this analysis can only be performed on a substantial computer cluster. For this purpose, the HPC facility of the Charité partner Berlin Institute of Health will be utilized.

The most complete collection of insect transcriptomes to date will be screened computationally to identify new RNA viruses. This data collection is known as "The 1KITE project: evolution of insects" and comprises 1,243 transcriptomes of individual insect species. Transcriptome data capture not only those cellular genes that are expressed at the given time point of the RNA extraction, but also all other non-host genes that undergo expression in the cell, therefore enabling the detection of reproducing viral genes. Using hidden Markov models of known viruses as a template for the search, evolutionary distant forms of insect viruses will be identified, thus contributing to a more integrated understanding of insect virus evolution and their relationship to viruses of other organismic groups (results presented in **Chapter 2** and **Chapter 4** and discussed in **Chapter 2**, **Chapter 4**, and **Chapter 5.2**).

Virus discoveries in non-human mammalian samples increase our understanding of the evolutionary processes that human viruses have followed. The discovery of a Hepatitis delta virus-like agent in non-human mammals and its *in vitro* proof of replication provides the opportunity to study for the first time the origins and evolution of a human pathogen with an unknown origin (results presented in **Chapter 3**). The many theories on the origin of this viral agent (presented in **Chapter 1.3.3.4**) are now placed on a different perspective, weakening some of them while favouring others (discussed in **Chapter 3** and **Chapter 5.3**).

At the same time, the discussion on whether Hepatitis delta virus and its newly-discovered evolutionary counterparts fit the definition of a virus remains unresolved and highly-debated (discussed in **Chapter 3** and **Chapter 5.3**).

Virus discovery is inevitably followed by taxonomic assessment and assignment of the new findings. The International Committee for Virus Taxonomy evaluates taxonomy proposals prepared by groups of scientists currently studying a given virus group. Newly-identified viruses that are divergent from already-known viruses and fulfill the criteria for taxonomic classification will be included in upcoming taxonomy proposals by the ICTV study groups (discussed in **Chapter 5** and summarized in **Table A1**).

2 Re-assessing the diversity of negative strand RNA viruses in insects

This publication is a collaborative work with the Zoological Research Museum Alexander Koenig and the group of Prof. Dr. Bernhard Misof, which dates back to the time when Prof. Dr. Christian Drosten was the director of the Institute of Virology in Universitätsklinikum Bonn. Data analysis, methodology, and investigation was carried out in close collaboration with Dr. Simon Käfer, with whom I share the first authorship in this publication.

Published as: Käfer, S., Paraskevopoulou, S., Zirkel, F., Wieseke, N., Donath, A., Petersen, M., Jones, T.C., Liu, S., Zhou, X., Middendorf, M., Junglen, S., Misof, B., and Drosten C. (2019). Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathogens*, 15(12):e1008224.

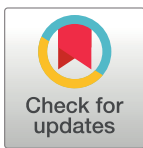
The publication is also available at: <https://doi.org/10.1371/journal.ppat.1008224>.


RESEARCH ARTICLE

Re-assessing the diversity of negative strand RNA viruses in insects

Simon Käfer^{1,2} , Sofia Paraskevopoulou¹ , Florian Zirkel^{3a}, Nicolas Wieseke⁴, Alexander Donath² , Malte Petersen⁵ , Terry C. Jones^{1,6} , Shanlin Liu⁷ , Xin Zhou^{8,9} , Martin Middendorf⁴ , Sandra Junglen^{1,10*} , Bernhard Misof^{2*}, Christian Drosten^{1,10*} 

1 Institute of Virology, Charité-Universitätsmedizin Berlin, Corporate Member of Free University, Humboldt-University and Berlin Institute of Health, Berlin, Germany, 2 Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Bonn, Germany, 3 Institute of Virology, University of Bonn Medical Centre, Bonn, Germany, 4 Swarm Intelligence and Complex Systems Group, Department of Computer Science, Leipzig University, Leipzig, Germany, 5 Senckenberg Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany, 6 Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge, United Kingdom, 7 BGI-Shenzhen, China Beishan Industrial Zone, Shenzhen, Guangdong Province, China, 8 College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, China, 9 Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China, 10 German Center for Infection Research (DZIF), associated partner site Charité, Berlin, Germany



 These authors contributed equally to this work.

 Current address: Biotest AG, Dreieich, Germany

* sandra.junglen@charite.de (SJ); bmisof@zfmk.uni-bonn.de (BM); christian.drosten@charite.de (CD)

 OPEN ACCESS

Citation: Käfer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A, Petersen M, et al. (2019) Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathog* 15(12): e1008224. <https://doi.org/10.1371/journal.ppat.1008224>

Editor: Francis Michael Jiggins, University of Cambridge, UNITED KINGDOM

Received: March 5, 2019

Accepted: November 19, 2019

Published: December 12, 2019

Copyright: © 2019 Käfer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data are available within the Supporting Information, except for nucleotide sequences of the described viral genomes. As these viral transcript sequences have been extracted from transcriptomes of the 1KITE consortium (Umbrella BioProject accession number NCBI: PRJNA183205 (“The 1KITE project: evolution of insects”)), they do not receive separate NCBI accession numbers. Until all 1KITE transcriptomes are publicly available, we deposit these viral sequences with the permission of the 1KITE consortium on Dryad (<https://doi.org/10.1371/journal.ppat.1008224>).

Abstract

The spectrum of viruses in insects is important for subjects as diverse as public health, veterinary medicine, food production, and biodiversity conservation. The traditional interest in vector-borne diseases of humans and livestock has drawn the attention of virus studies to hematophagous insect species. However, these represent only a tiny fraction of the broad diversity of Hexapoda, the most speciose group of animals. Here, we systematically probed the diversity of negative strand RNA viruses in the largest and most representative collection of insect transcriptomes from samples representing all 34 extant orders of Hexapoda and 3 orders of Entognatha, as well as outgroups, altogether representing 1243 species. Based on profile hidden Markov models we detected 488 viral RNA-directed RNA polymerase (RdRp) sequences with similarity to negative strand RNA viruses. These were identified in members of 324 arthropod species. Selection for length, quality, and uniqueness left 234 sequences for analyses, showing similarity to genomes of viruses classified in *Bunyavirales* (n = 86), *Articulavirales* (n = 54), and several orders within *Haploviricotina* (n = 94). Coding-complete genomes or nearly-complete subgenomic assemblies were obtained in 61 cases. Based on phylogenetic topology and the availability of coding-complete genomes we estimate that at least 20 novel viral genera in seven families need to be defined, only two of them monospecific. Seven additional viral clades emerge when adding sequences from the present study to formerly monospecific lineages, potentially requiring up to seven additional genera. One long sequence may indicate a novel family. For segmented viruses, cophylogenies between genome segments were generally improved by the inclusion of viruses from the present study, suggesting that *in silico* misassembly of segmented genomes is rare or absent. Contrary to previous assessments, significant virus-host codivergence was

5061/dryad.87vt6hm) to make them publicly available.

Funding: This work was supported by grants awarded to CD by the Deutsche Forschungsgemeinschaft (DR 772/7-2, www.dfg.de) and the German Ministry of Research (DZIF TTU 01.801, www.bmbf.de). Work of BM has been funded by a Leibniz competition grant "Graduate School on Biodiversity Genomics." The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

identified in major phylogenetic lineages based on two different approaches of codivergence analysis in a hypotheses testing framework. In spite of these additions to the known spectrum of viruses in insects, we caution that basing taxonomic decisions on genome information alone is challenging due to technical uncertainties, such as the inability to prove integrity of complete genome assemblies of segmented viruses.

Author summary

The diversity of insect viruses is relevant to medical, environmental, and food sciences. Our knowledge of insect viruses is highly biased because medical research has focused on mosquitoes and a few other blood-feeding species. While insects are the most diversified group of animals on the planet, the great majority of all insect species remain completely unexamined for viruses. Here we searched the most comprehensive and most evenly composed collection of insects for negative strand RNA viruses based on full transcriptomes. In 1243 insect species of all orders, we found 488 independent viral sequences encoding an RNA-directed RNA polymerase, a signature gene for RNA viruses. These data add considerably to our knowledge on viral diversity, and reveal that viruses have coevolved with insect hosts. However, our results also provide a reminder of the pitfalls associated with virus discovery and taxonomic classification in the age of metagenomics.

Introduction

Negative strand RNA viruses contain major groups of pathogenic viruses that cause rabies, hemorrhagic fevers, respiratory infections, measles, as well as a large range of important diseases and economically important conditions in livestock and plants [1–4]. Our current knowledge of negative strand RNA viruses is biased by the interest in medical disciplines and provides an incomplete image when it comes to more fundamental questions in viral evolution, such as the contribution of codivergence in the formation of major viral genetic lineages. These questions can only be addressed by systematic studies of larger taxonomic units of viral hosts, corresponding to whole orders or classes of animals, which is complicated by the difficulty to establish representative sample collections. Samples utilized for viral diversity studies are often collected on an opportunistic basis or repurposed from other studies, resulting in imbalance in host species representation, uncertainty in host classification, and uncertain assignment of samples. This is especially true for studies of insects that show an enormous genetic and morphological diversity.

Insects are the most speciose group of animals. Their origin has been dated to the early Ordovician, 479 million years ago, a time that predates the formation of terrestrial ecosystems [5]. Insects engage in symbiotic and parasitic relationships with a multitude of plants and animals, and are a vital component of the diet of animals, potentially facilitating virus transmission. Nevertheless, research on insect viruses has been mainly driven by interest in vector-borne diseases, resulting in virological studies that have focused on blood-feeding species, with rare exceptions [4, 6, 7]. However, blood-feeding insects represent only a minute fraction of the biological diversity of insects. Studies using massively parallel sequencing of collections of invertebrates have yielded an unprecedented diversity of novel RNA viruses [4, 6, 8]. However, the samples used in these studies only covered a limited range of insect species, contained many other groups of invertebrates such as spiders, worms, and molluscs, and were generated

by sample pooling. Uncertain knowledge of host associations in these and other studies have caused a tendency to abandon host association as an important auxiliary criterion for taxonomic classification [9].

Here, we systematically probed the diversity of negative strand RNA viruses in the largest and most representative collection of full transcriptome datasets of arthropods. The collection is designed to represent all extant lineages of Hexapoda without representational bias. This transcriptome database was first utilized for a phylogenomic re-assessment of the Hexapoda phylogeny in 2014, based on 103 full transcriptomes [9]. Since that time, the collection has been significantly extended to now cover 1243 full transcriptome datasets. All datasets including their corresponding unassigned contigs and scaffolds were screened for negative strand RNA viruses. The collection represents all orders of Insecta (insects, $n = 1178$), the orders Collembola (springtails, $n = 23$), Protura (coneheads, $n = 4$), and Diplura ($n = 14$) of Entognatha, as well as 24 outgroup species pertaining to Crustacea ($n = 10$), Myriapoda ($n = 11$), and Chelicerata ($n = 3$).

Results

We based our search on conserved sequence motifs within the RdRp gene that is present in the genomes of all replicating RNA viruses without a DNA stage except deltaviruses, and is not present in the genome of the eukaryotic or prokaryotic cell. We utilized profile hidden Markov models (pHMMs) to search for candidate viral RdRp motifs within 42,618,061 contigs and scaffolds which were 66 to 20,314 amino acids long. pHMMs were trained on template amino acid alignments covering the core conserved RdRp regions of representative viruses assigned to the families *Rhabdoviridae*, *Paramyxoviridae*, *Filoviridae*, *Nyamiviridae*, and *Orthomyxoviridae*, as well as the genera *Orthonairovirus*, *Mammarenavirus*, *Jonvirus*, *Orthohantavirus*, *Orthobunyavirus*, *Tospovirus*, *Herbevirus*, *Phlebovirus*, and *Goukovirus*.

According to the results of contig assembly, we initially detected 488 viral RdRp sequences. These were identified in 324 arthropod species belonging to all insect orders and several outgroup taxa. The host associations, exact taxonomic classification, as well as sampling sites of hosts for the viral genomes that appear in the phylogenetic trees in Fig 1 are summarized in S1 Table.

A large proportion of the viral sequences were co-detected with different pHMMs, owing to the nature of the search algorithm which makes the detection of distant homologs possible. This is of particular relevance since the template alignments, as well as the pHMM searches were done before the release of any of the sequences described in [4] and [8]. All data were later re-examined using BLASTp, with the inclusion of the data of [4] and [8], but no additional matches were retrieved. This confirms the sensitivity of the pHMM search approach and demonstrates that the search strategy is not biased by a virus reference library that stems from a fragmentary sample of insect species.

From the obtained contigs, 234 large sequences were selected for further analysis based on length, quality, and dissimilarity toward other sequences in the dataset. These sequences were later found to have highest similarity to members of *Bunyavirales* ($n = 86$), *Articulavirales* ($n = 54$), or *Haploviricotina* ($n = 94$), respectively. For non-segmented viruses, full genome assembly was often successful. For viruses with segmented genomes, assembly focused on the RdRp-encoding segment was later on complemented by BLAST-based searches for other genome segments expected. Thereby, 218 coding sequences from genes that are not encoded on the same segment as the RdRp gene, such as glycoproteins, nucleoproteins, polymerase subunits, and proteins with unknown function were identified. Complete or coding-complete genomes were assembled in 61 cases. Many additional large but incomplete genome assemblies

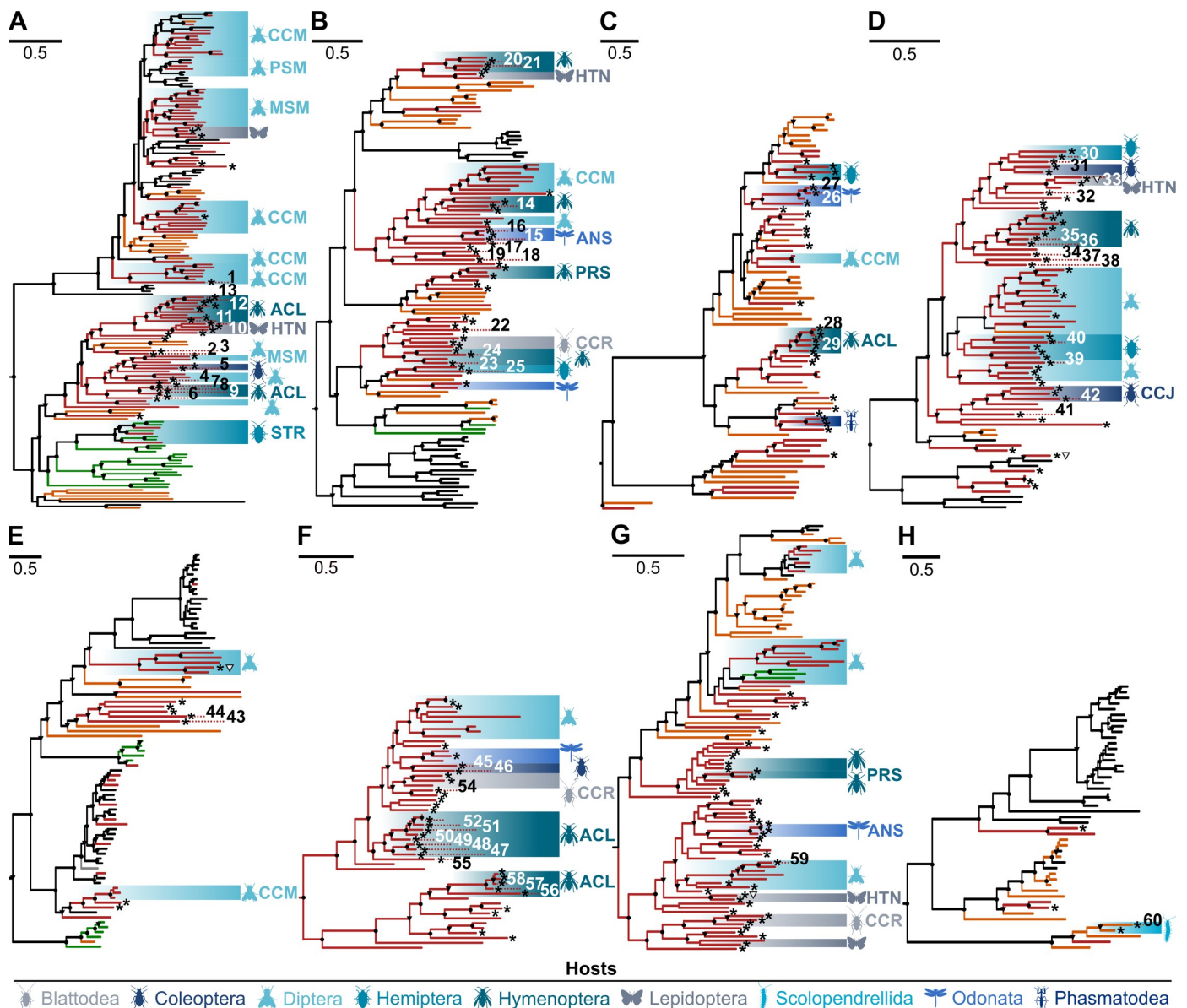


Fig 1. Maximum likelihood phylogenies of viruses found in the present study, viruses defined by ICTV as species, as well as selected unclassified viruses. Novel viruses from the present study are identified by an asterisk. Novel viruses to be considered for taxonomic proposals because of unique phylogenetic position and availability of a coding-complete genome are identified by numbers corresponding to Table 1. Trees were inferred in RAxML based on alignments of viral sequences pertaining to: A: *Rhabdoviridae*; B: *Ximoviridae*, *Bornaviridae*, *Artoviridae*, *Lispiviridae*, *Paramyxoviridae*, *Sunviridae*, *Filoviridae*, and *Pneumonoviridae*; C: *Chuviridae*, *Qinviridae*, and *Yueviridae*; D: *Orthomyxoviridae*; E: *Hantaviridae*, *Cruliviridae*, *Peribunyaviridae*, and *Fimoviridae*; F: *Phasmaviridae*; G: *Pheniviridae*; H: *Arenaviridae*, *Mypoviridae*, *Nairoviridae*, and *Wupedeviridae*. Branch colors show host associations: Black, viruses associated with vertebrates; red, insects; orange, invertebrates other than insects; green, plants. Bootstrap supports on tree nodes are shown by circles (>= 95) and triangles (>70). Designated host infraorders for Blattodea: CCR: Cockroaches; for Coleoptera: CCJ: Cucujiformia; for Diptera: CCM: Culicomorpha, MSM: Muscomorpha, PSM: Psychodomorpha; for Hemiptera: STR: Stenorrhyncha; for Hymenoptera: ACL: Aculeata, PRS: Parasitica; for Lepidoptera: HTN: Heteroneura; for Odonata: ANS: Anisoptera.

<https://doi.org/10.1371/journal.ppat.1008224.g001>

were obtained that contain open reading frames of unknown function and represent unknown genome architectures.

Of note, poly-A purification was applied on all samples due to the intended use for transcriptome analysis [5]. We expected a general loss of sensitivity in detecting viral sequences but did not expect a huge bias against viral genomes as opposed to viral mRNAs. For instance,

even during single-cycle infection, more viral genomic RNA than viral mRNA is obtained from poly-A preparation in mononegaviruses, in spite their mRNAs but not genomes contain poly-A tails [10]. In natural infections an even larger excess of genomes over mRNA is expected, because genomic RNA accumulates once packed, while viral mRNA is degraded at a rate that weighs against new synthesis. Also in members of *Bunyavirales* a poly-A-related bias is not expected because neither mRNAs nor genomes of these viruses have poly-A tails.

Novel viruses found in this study are named after their host order, related viral family, and the designation “OKIAV” (for 1KITE insect-associated virus), followed by a number (e.g., Hemipteran orthomyxo-related virus OKIAV183). All sequences, including genome annotations, host associations, and sampling sites, are available on the Dryad Digital Repository under <https://doi.org/10.5061/dryad.87vt6hm>. All sequences will be released in GenBank along with the release of transcriptomes from the 1KITE project (BioProject PRJNA183205: ‘The 1KITE project: evolution of insects’).

Phylogeny and implications on taxonomy

To enable phylogenetic analysis, contigs were translated and grouped into eight alignments based on preliminary sequence matching and phylogenies. The eight resulting trees as shown in Fig 1 were generated following intense optimization of alignments by trimming and focusing on conserved domains, with the aim to leave sufficient information in alignments while allowing large taxonomic groups of viruses to be compared. The trees A-H in Fig 1 cover the *Rhabdoviridae* (Fig 1A), all other *Mononegavirales* (Fig 1B), *Chu-*, *Qin-*, and *Yueviridae* (Fig 1C), *Orthomyxoviridae* (Fig 1D), *Hanta-*, *Cruli-*, *Peribunya-*, and *Fimoviridae* (Fig 1E), *Phasmaviridae* (Fig 1G), *Phenuiviridae* (Fig 1G), as well as *Arena-*, *Mypo-*, *Nairo-*, and *Wupedeviridae* (Fig 1H). Detailed phylogenies including host associations down to the host species level, as well as viral taxonomy information are shown in **Supporting Information S1–S3** and **S5–S33 Figs**. A detailed description of current taxonomy including the novel virus findings, as well as classification suggestions resulting from the present data are provided in **Supporting Information S1 Text**.

The tree structures in Fig 1 suggest a remarkable separation between vertebrate and insect viruses, as noted already on the basis of a less inclusive sample of insect diversity [4, 8]. With the exception of the subtree of non-plant-associated rhabdoviruses that still remains star-like and may thus be undersampled, many insect-associated clades now appear well-differentiated with a balanced proportion of intermediate versus terminal branches. In spite of the more inclusive insect sampling contributed by the present study, novel insect viruses remain absent in well-known clades of pathogenic vertebrate viruses, such as the genus *Lyssavirus*, the families *Paramyxoviridae*, *Bornaviridae*, *Filoviridae*, *Hantaviridae*, and *Arenaviridae*. Also, some major groups of pathogenic arboviruses do not show an expansion of host associations following our search. For instance, the phleboviruses and orthobunyaviruses that are known to be mosquito-, sandfly-, midge-, or tick-borne, do not yield any novel insect-associated viruses in our sample in spite of its enormous genetic diversity (note that there are no sandfly, no mosquito, and only two midge species in our sample). This absence is remarkable as also the studies of [4] and [8] did not find any novel phleboviruses in the insects they sampled, while they did find novel phleboviruses in ticks. There may exist an ecologically-driven association of these viruses with blood-feeding insects. The additional association with ticks makes it possible that viruses could be exchanged between insects and ticks based on common bloodmeal sources.

Classification criteria exist only for a minority of viral genera. For instance, amino acid sequence distances of 4% and 5% have been proposed for species demarcation in

orthobunyaviruses and phleboviruses, respectively. Our distance-based selection of sequences for inclusion in trees exceeds this distance criterion, making it likely that all novel sequences identified by an asterisk in Fig 1 could be classified as novel species. While many recently-described viruses still form solitary lineages in trees, a deep topological separation and structured host association emerges after inclusion of data from the present study. Large subclades within viral families are often associated with insect orders or suborders, indicating an important auxiliary criterion for subdivision of these viral families into genera. Structured host associations become particularly obvious in the families *Rhabdoviridae*, *Xinmoviridae*, *Nyamiviridae*, *Artoviridae*, *Lispiviridae*, *Chuviridae*, *Phasmaviridae*, and *Feraviridae* (**Supporting Information S1 Text**). Table 1 lists those virus clades that we propose to be considered for classification on the genus level based on sequence distance and host associations, while taking into account the completeness of available genome sequences. All in all, the analyzed sequences suggest a potential for classifying at least 27 novel genera based on coding-complete virus genomes, 20 of them without any previously known representative, and identify deep-branching viral lineages that in the future may be classified as three novel families or subfamilies. Host associations are especially informative for subclassification of rhabdo- and xinmoviruses, chuviruses, orthomyxoviruses, and phasmaviruses. Furthermore, we add the first independent description of a full qinivirus genome (Collembolan qin-related virus OKIAV112), detected in the entognath *Anurida maritima* (seashore springtail, class Collembola) (Fig 1C). Detailed taxonomical considerations are provided in the **Supporting Information S1 Text**.

The discovery of a large diversity of novel lineages warrants a re-assessment of the overall topology of negative strand viruses. Based on a manually curated alignment, we inferred a tree as shown in Fig 2 using Bayesian phylogeny. Only few of the topological relationships differ from those in Fig 1, which incorporates more alignment information specific for the smaller units of diversity covered therein. For instance, there is weak support for the branching point of rhabdoviruses, as also observed by [13]. As in Shi *et al.* [8] and as implied by current taxonomy, but unlike the results by Wolf *et al.* [13], the members of the order *Articulovirales* branch to the exclusion of all members of *Bunyavirales*, while the deep topology of *Bunyavirales* is well supported. A noteworthy finding is Collembolan phasma-related virus OKIAV223, a large sequence of 8154 nucleotides extending beyond the L-gene ORF, albeit not covering segment termini. It clusters with *Phasmaviridae* and branches from the phasmavirus lineage short after the split from the last common ancestor of *Peribunyaviridae* and *Phasmaviridae* (Fig 2). It is therefore the most appropriate outgroup for the peribunyavirus tree (Fig 1E), which has been considered for rooting that tree. It is interesting to note that this topology suggests acquisitions of tospo- and emaraviruses by plants from invertebrates, rather than an evolution of peribunyaviruses from plant viruses as suggested by alternative tree topologies. A number of other findings that indicate deeply diverged and novel virus groups are described in **Supporting Information S1 Text**.

Genome composition

Tentative structures of complete or nearly-complete genomes are summarized in Fig 3. It is noteworthy that genomes of chuviruses were found to appear in linear, circular, and segmented circular forms [4]. Fig 1C includes an additional 25 exemplary chuviruses from the present study, including seven with at least one complete segment and one with two complete segments. According to the mapping of raw RNAseq reads, all genomes or genome segments of these viruses are linear. Gene order is L-G-N, or N-G-L (Fig 3), confirming the two principal gene orders described in Li *et al.* [4]. Genomes with a missing glycoprotein gene or over-

Table 1. List of phylogenetic groups to be considered for taxonomic proposals.

Superordinate taxon	Putative taxonomic level	Clade annotation	Included tentative species (full genomes)	Source	Remarks	No. in Fig 1
<i>Rhabdoviridae</i>	Genus	ARR	Hymenopteran almendra-related virus OKIAV1	This study	Monospecific	1
	Genus	MBAR	Blattodean rhabdo-related virus OKIAV14, Mantodean rhabdo-related virus OKIAV15	This study		2 3
	Genus	DHCR	Dipteran rhabdo-related virus OKIAV19, Coleopteran rhabdo-related virus OKIAV28, Wuhan mosquito virus 9*	This study and Li <i>et al.</i> [4]		4 5
	Genus	CAR	Coleopteran rhabdo-related virus OKIAV20	This study	Monospecific	6
	Genus	HAR2	Hymenopteran rhabdo-related virus OKIAV22, -OKIAV23, -OKIAV24	This study		7 8 9
	Genus	LAR	Lepidopteran rhabdo-related virus OKIAV34	This study		10
	Genus	HAR1	Hymenopteran rhabdo-related virus OKIAV38, -OKIAV46, -OKIAV109, Hubei rhabdo-like virus 1	This study and Shi <i>et al.</i> [8]		11 12 13
<i>Xinmoviridae</i>	Genus	<i>Anphevirus</i> lineage I	<i>Xincheng anphevirus</i> ** <i>Aedes aegypti anphevirus</i> **, Hymenopteran anphe-related virus OKIAV71	This study, Shi <i>et al.</i> [8], and Di Giallonardo <i>et al.</i> [11]		14
	Genus	<i>Anphevirus</i> lineage II	Odonatan anphe-related virus OKIAV57, -OKIAV59	This study		15 16
	Genus	<i>Anphevirus</i> lineage III	Coleopteran anphe-related virus OKIAV54	This study	Subcomplete genome	17
	Genus	<i>Anphevirus</i> lineage V	Odonatan anphe-related virus OKIAV90, Mantodean anphe-related virus OKIAV92, <i>Orthopteran anphevirus</i>	This study and Shi <i>et al.</i> [8]		18 19
<i>Nyamiviridae</i>	Genus	<i>Orinovirus</i> lineage I	Hymenopteran orino-related virus OKIAV85, -OKIAV87	This study		20 21
<i>Lispiviridae</i>	Genus	<i>Arlivirus</i> lineage I	Strepsipteran aril-related virus OKIAV104, Hubei arlivirus	This study and Shi <i>et al.</i> [8]		22
	Genus	<i>Arlivirus</i> lineage III	Hymenopteran arli-related virus OKIAV98, -OKIAV99	This study		23 24
	Genus	<i>Arlivirus</i> lineage IV	Hemipteran aril-related virus OKIAV94	This study		25
<i>Chuviridae</i>	Genus	OAM	Odonatan chu-related virus OKIAV136, -OKIAV137, <i>Odonate mivirus</i>	This study and Shi <i>et al.</i> [8]		26 27
	Genus	HyAM	Hymenopteran chu-related virus OKIAV123, -OKIAV124	This study		28 29

(Continued)

Table 1. (Continued)

Superordinate taxon	Putative taxonomic level	Clade annotation	Included tentative species (full genomes)	Source	Remarks	No. in Fig 1
<i>Orthomyxoviridae</i>	Genus	O1	Hemipteran orthomyxo-related virus OKIAV183, Coleopteran orthomyxo-related virus OKIAV184	This study		30
						31
	Genus	O2	Blattodean orthomyxo-related virus OKIAV181, Lepidopteran orthomyxo-related virus OKIAV178	This study		32
						33
	Genus	O3	Dermapteran orthomyxo-related virus OKIAV162, Hymenopteran orthomyxo-related virus OKIAV171, Phasmatodean orthomyxo-related virus OKIAV172	This study		34
						35
						36
Genus	O4	Siphonapteran orthomyxo-related virus OKIAV157, Coleopteran orthomyxo-related virus OKIAV158	This study	Only 4 segments	37	
						38
Genus	O5	Dipteran orthomyxo-related virus OKIAV164	This study		39	
Genus	O6	Hemipteran orthomyxo-related virus OKIAV188	This study		40	
Genus	O7	Dipteran orthomyxo-related virus OKIAV199, Coleopteran orthomyxo-related virus OKIAV200, Hubei orthomyxo-like virus 2	This and Shi <i>et al.</i> [8]		41	
					42	
<i>Bunyavirales</i>	Family	Novel group	Dipluran hanta-related virus OKIAV217, -OKIAV218	This study	Only 2 segments	43 44
<i>Bunyavirales</i>	Family	Not annotated	Collembolan phasma-related virus OKIAV223	This study	Only L-gene	Fig 2
<i>Phasmaviridae</i>	Genus	CAP	Coleopteran phasma-related virus OKIAV235, -OKIAV236	This study		45
						46
	Genus	HAP	Hymenopteran phasma-related virus OKIAV227, -OKIAV229, -OKIAV230, -OKIAV228, -OKIAV233, -OKIAV234, -OKIAV232, Ganda bee virus	This study and Schoonvaere <i>et al.</i> [12]		47
						48
						49
						50
						51
52						
53						
Genus	MAP1	Coleopteran phasma-related virus OKIAV243	This study		54	
Genus	DAP2	Dipteran phasma-related virus OKIAV226	This study		55	
Genus	HAF	Hymenopteran phasma-related virus OKIAV244, -OKIAV250, -OKIAV252	This study		56	
					57	
					58	

(Continued)

Table 1. (Continued)

Superordinate taxon	Putative taxonomic level	Clade annotation	Included tentative species (full genomes)	Source	Remarks	No. in Fig 1
<i>Phenuiviridae</i>	Subfamily	Putative subfamily	Dipteran phenui-related virus OKIAV273, Salarivirus, Shuangao insect virus 3	This study and Li <i>et al.</i> [4]		59
<i>Bunyavirales</i>	Family	Not annotated	Myriapodan Negavirus OKIAV320, Jiangxia mosquito virus 1	This study and Li <i>et al.</i> [4]	Genome status uncertain	60

*Wuhan mosquito virus 9, but none of the other members of the clade, is an endogenous viral element

***Xincheng anphevirus* and *Aedes aegypti anphevirus*, but none of the other members of the clade, are likely to be endogenous viral elements.

ARR: Almendra-related rhabdovirus; DHCR: Diptera-, Hemiptera-, Coleoptera-related rhabdovirus; HAR: Hymenoptera-associated rhabdovirus; LAR: Lepidoptera-associated rhabdovirus; MBAR: Mantodea-/Blattodea-associated rhabdovirus; CAR: Coleoptera-associated rhabdovirus; OAM: Odonata-associated Mivirus; HyAM: Hymenoptera-associated Mivirus; O1-O7: Orthomyxovirus clades 1–7; CAP: Coleoptera-associated phasmaviruses; HAP: Hymenoptera-associated phasmaviruses; MAP1: Multiple host-associated phasmaviruses clade 1; DAP2: Diptera-associated phasmaviruses clade 2; HAF: Hymenoptera-associated feraviruses.

<https://doi.org/10.1371/journal.ppat.1008224.t001>

assembled contigs, as described in the same work, are not observed in the present study. To check for circular genome organization, we have re-mapped all raw RNAseq reads to consensus alignments of chuviral sequences joined head to tail. This approach did not find any reads crossing the potential head/tail sequence boundaries, as would be expected in the case of genome segments that are circular. While we do not claim to refute circular genomes in chuviruses, we cannot confirm this genome conformation based on our data (**Supporting Information S34 Fig**) and recommend further experimental validation.

The genome segment termini in members of *Bunyavirales* form complementary panhandle structures [15, 16]. These short sequences are identical within, and similar between genome segments of a given viral genome, and are usually conserved in viruses that belong to one same genus. Because segment termini in most of the members of recently-defined novel viral genera have not been analyzed (including in the present study; refer to **Supporting Information S1 Text**), we determined segment co-segregation as an indicator of grouping congruence of genome elements. Tanglegrams are shown in **Supporting Information S15, S19, S24, S28, S29, and S33 Figs**. In most major clades there is congruence among segments. Some clades, such as clade C of the orthomyxoviruses (**Supporting Information S15 Fig**) or the clades that define *Shanga-* and *Herbevirus* in the peribunyaviruses, show signs of reassortment in lineage precursors, as topological incongruence is observed for all members of the respective clades. In cases where individual incongruences are seen, such as in Dipteran phasma-related virus OKIAV224, Zorapteran phasma-related virus OKIAV242, or Coleopteran phasma-related virus OKIAV243, we cannot discriminate between *in silico* misassembly of genomes and actual reassortment based on the present data. Confirmation by virus isolation or re-sequencing including genome ends will be necessary.

To obtain an overall impression of segment co-segregation in newly-discovered segmented RNA viruses, we analyzed cophylogenies of RdRp-encoding segments and other segments from the same putative viral genomes using Jane [17]. We compared cophylogeny costs against that of datasets with randomized segment associations. As summarized in **Fig 4**, addition of the present findings rather improved the cophylogeny costs except in cophylogenies between L- and M-segments (RdRp- and glycoprotein-encoding) of phasma- and pheniviruses where there was no relevant change (**Fig 4**). Also, in some viral trees the addition of the present data reveal segment cosegregation where this was not evident from the genomes of previously

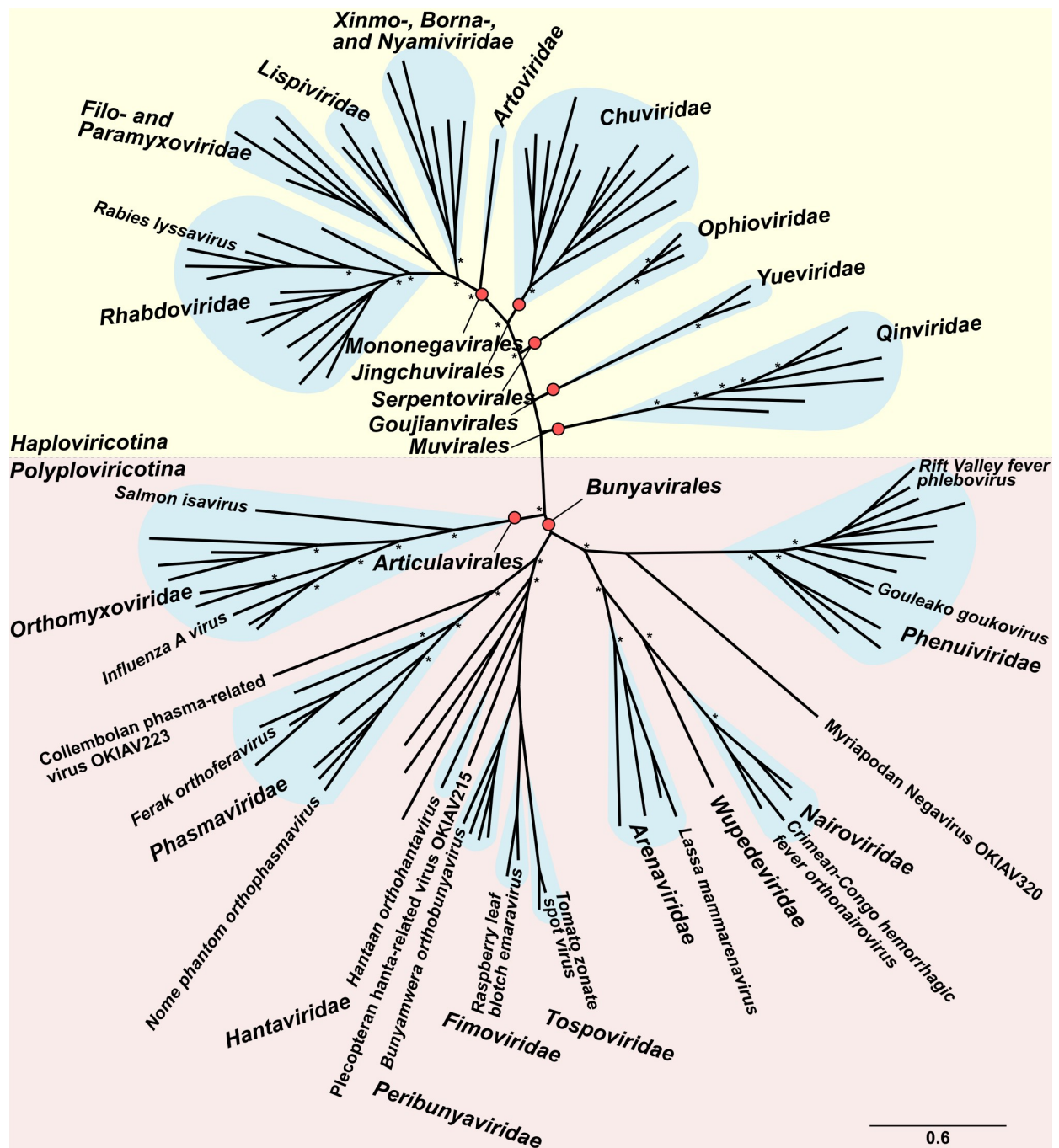


Fig 2. Bayesian phylogeny of negative strand RNA viruses based on MrBayes. Red dots identify virus orders according to current taxonomy. Blue clouds show virus families. Selected viral species are identified for orientation. Asterisks indicate posterior probabilities > 0.9.

<https://doi.org/10.1371/journal.ppat.1008224.g002>

described viruses. We thus assume that our segmented genome findings overall do not suffer from *in silico* misassembly or other artifacts. However, the genomes of exemplary strains defining novel genera should be confirmed experimentally.

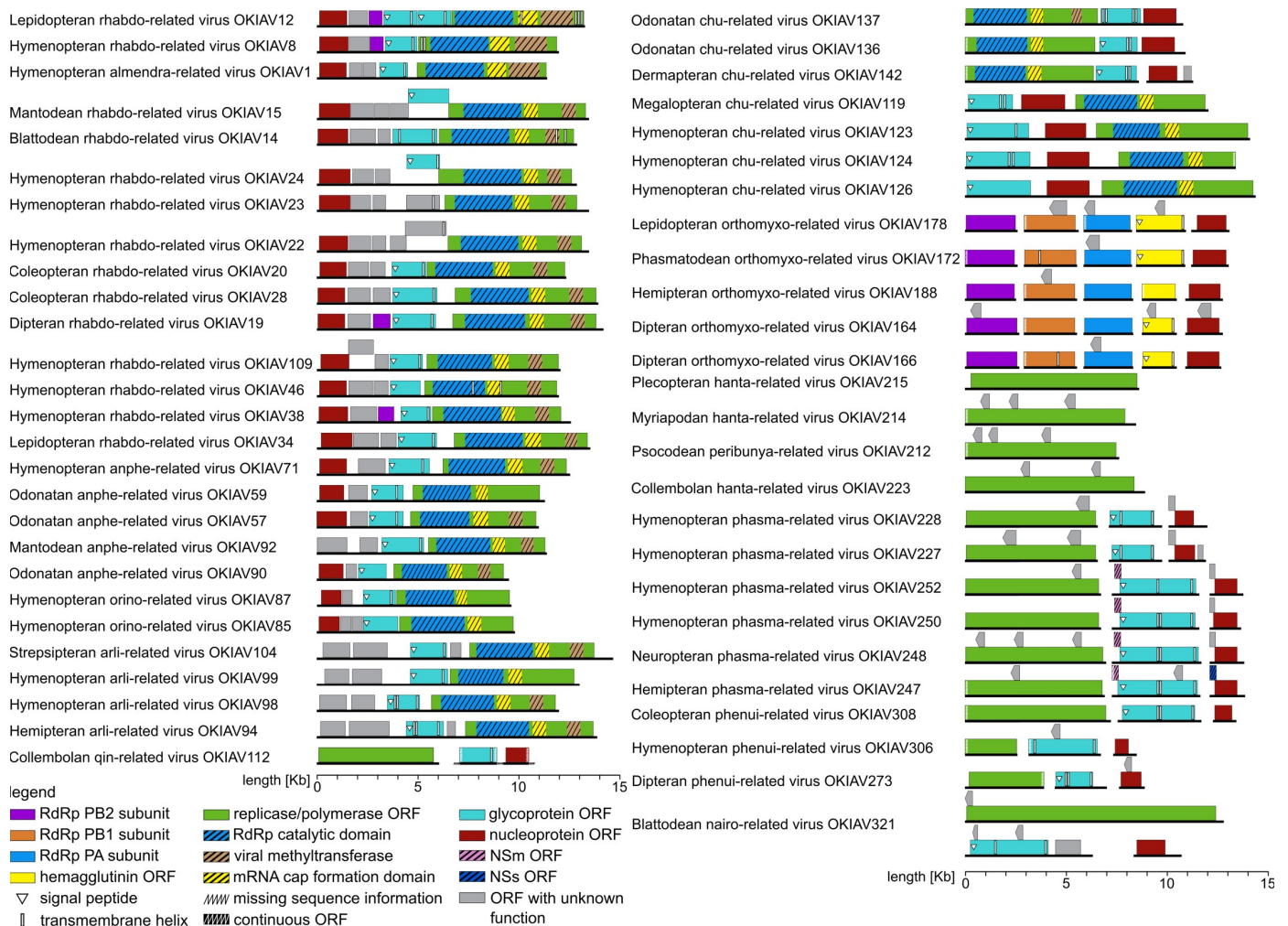


Fig 3. Annotations of full or nearly-full genomes and selected full bunyaviral L-segments found in the present study. Annotation was done using InterProScan [14]. Bunyaviral putative nonstructural genes, as well as other significant subgenomic reading frames were annotated manually.

<https://doi.org/10.1371/journal.ppat.1008224.g003>

Host-virus co-segregation

Recent studies on invertebrate-associated RNA viruses found no evidence of host-virus co-segregation, and proposed frequent cross-host transmission of viruses between insect hosts that co-occupy the same ecological niches [4]. However, these results were based on a limited and spatially restricted sample of insects and other invertebrates. For the present study we have subjected all viral phylogenies to formal cophylogenetic comparisons on the basis of resolved and updated phylogenies of insects as in [5].

The phylogenies shown in Fig 1 were subjected to tests of breaches of cophylogeny using Jane [17] (refer to **Supporting Information S1 Table** for host associations). To determine the contribution of the novel sequences, separate analyses were done without the OKIAV sequences but incorporating all known and novel viruses as per ICTV taxonomy update end of 2018. The limited knowledge of host associations in most studies restricted the resolution of these analyses to the level of insect orders. Significant virus-host co-segregation was identified in both analyses (with and without OKIAV findings) for the majority of trees (Fig 5A). For the

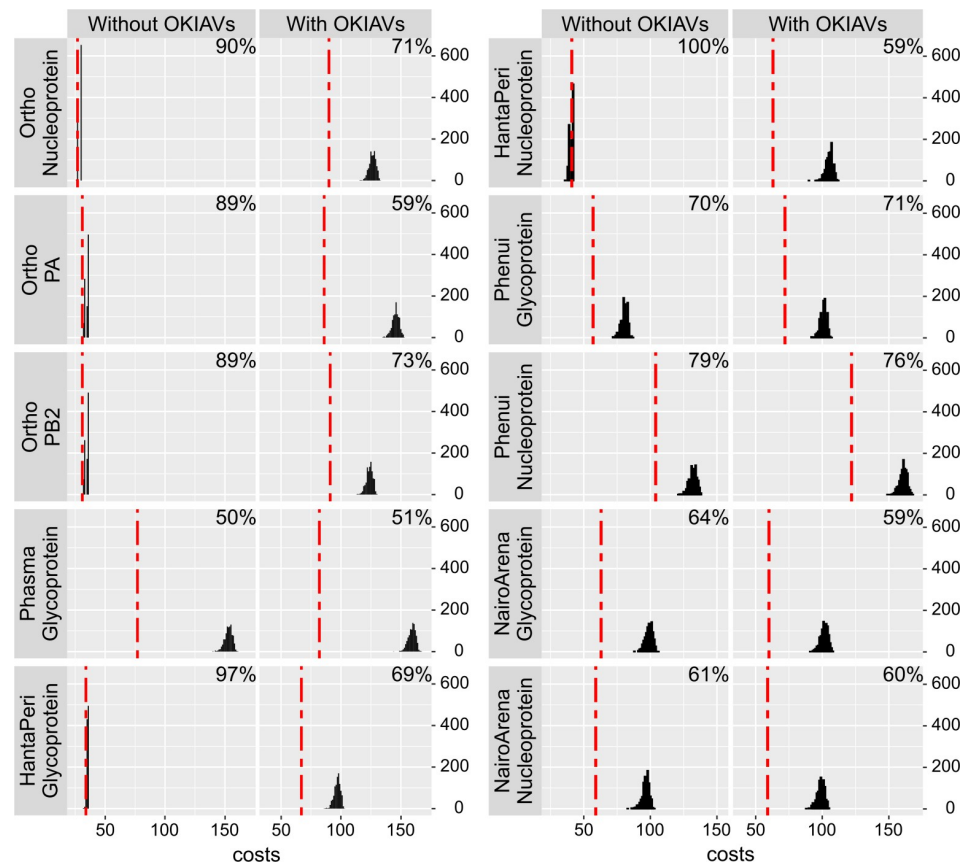


Fig 4. Genome segment cophylogeny costs. Red dotted lines show lowest possible costs for breaches of cophylogeny when using Jane [17] to optimize the tip opposition of RdRp-encoding genome segment trees with the corresponding trees for genome segments indicated in each figure panel name. Black curves show the same costs accrued during each of 1000 different randomizations of the dataset. Percent values in panels show the cophylogeny costs of the real cophylogeny dataset divided by the costs with randomization (medians from 1000 randomizations). Lower percent values indicate better agreement between the RdRp and the respective genome segment cophylogeny.

<https://doi.org/10.1371/journal.ppat.1008224.g004>

trees including *Chuviridae* and *Hanta-/Peribunyaviridae*, significant co-segregation could only be detected when including findings from the present study. Only for the tree summarizing members of *Phasmaviridae*, no host-virus co-segregation could be detected.

Because analysis in Jane does not take branch lengths into account, does not incorporate topological uncertainty, and mainly penalizes breaches of phylogenetic congruence between virus and host tree, an alternative algorithm, CoRe-PA [18], was applied to further examine host-virus co-segregation. This more sophisticated algorithm takes into account the reconstruction of co-speciation and host-switching separately, and also discriminates these from duplication events. To exclude potential bias introduced by the analysis of small phylogenetic trees, we utilized trees that represent the complete tree topologies of *Haploviricotina* and *Polyploviricotina*, including topological uncertainty based on 1000 bootstrap replicates. All possible rootings (i.e., tree versions rooted on every branch) of all replicate trees were modeled in order to exclude the impact of false co-segregation signals among outgroup branches in insect and host trees. Randomization of host associations was performed as previously and used as null hypotheses in separate formal tests of co-segregation and host switching, respectively. As summarized in Fig 5B and 5C, trees of both virus subphyla showed significantly more co-

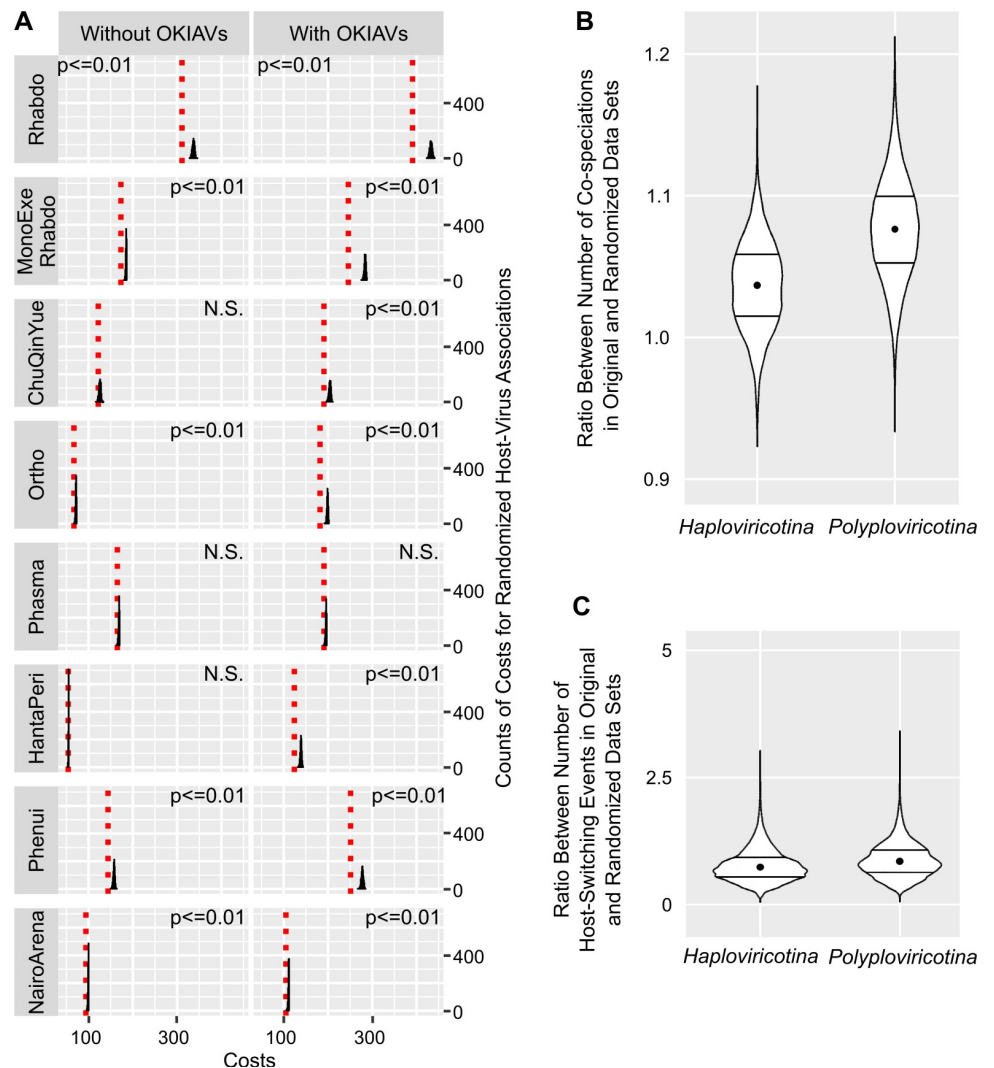


Fig 5. Analysis of host-virus phylogenetic co-segregation. A: Histograms of costs for co-segregation of randomized host associations (1000 iterations) compared to the original host associations (red dotted lines) for all phylogenies, without (left) and with (right) OKIAV sequence inclusion, calculated with Jane [17]. The p-values of all z-tests indicate an increase in costs of over 5% above the original costs. N.S. indicates non-significant cost differences. B: Ratio between number of co-speciations in original and randomized data sets calculated for *Haploviricotina* and *Polyploviricotina* with CoRe-PA [18]. C: Ratio between number of host-switching events in original and randomized data sets calculated for *Haploviricotina* and *Polyploviricotina* with CoRe-PA.

<https://doi.org/10.1371/journal.ppat.1008224.g005>

segregation events with real host associations as opposed to randomized host associations. In contrast, tests for host-switching events were not different when analyzing real versus randomized host associations. This suggests that earlier indications of predominance of host-switching were likely to have been caused by uneven sampling, incorrect host attributions, or issues with viral phylogenies.

Discussion

Our study summarizes curated results from the most comprehensive metagenomic virus screen of a whole class of animals. The host dataset is resolved to all known orders of insects

and involves an evenly selected representation of insect families. Because every sample has been genetically classified, the present work provides exact knowledge of host associations for a large range of insect-associated viruses for the first time. This knowledge is essential to study viruses as potential agents of vector-borne diseases in animals or plants, as well as to understand insect viruses in the context of biodiversity conservation and food production [19]. It also offers important criteria to define novel viral species and higher taxa. In this and other regards, our work complements the recent work by Li *et al.* [4] and Shi *et al.* [8] that was mainly focused on the description of viral diversity. For several taxa described in these studies, host associations are specified for the first time. Furthermore, some of the recent taxonomic proposals or re-classifications by ICTV are reconciled by the present data.

While our search has revealed an even greater diversity of positive strand RNA viruses, it has been challenging to curate large numbers of viral sequences while keeping track with taxonomy as it is being changed. The present frequency of taxonomy revisions makes it difficult to design basic analyses, such as phylogenies, so that they incorporate the latest viral taxonomy. The process of taxonomic classification owes to be more conservative and undergo slower cycles of renewal. It should allow time for independent reproduction of sequence findings before these go into taxonomic classification. Independent reproduction is becoming more important as ICTV has recently decided to allow classification on the sole basis of genome sequence data. As our study exemplifies, classification may be uncertain when full genomes cannot reliably be obtained, *e.g.*, in segmented viruses. Genome segment termini should be tested for complementarity among segments, but are often not covered in sequencing studies [4, 8, 20]. Even in the few cases where genome ends are known and identical, this does not prove that detected segments stem from one viral genome, because the ends of *e.g.*, bunyaviral segments are identical between viral species and can agree across genera and even families [21]. The study of viral sequences derived from mixtures of insects is particularly difficult, as they may be assembled into chimeric genomes whose fragments may, more likely, stem from different viruses than one virus. The ICTV executive board has stated, with regards to the danger of assembling chimeric genomes, that “These are all caveats that must be addressed experimentally for MG [metagenomic] sequence data to be used for classification purposes” [9]. From a virological view, segmented viruses that are to be classified as species based only on sequence data should have full genome coverage including genome ends, and should stem from individual insects. They should also be examined for phylogenetic congruence between genome segments. Multiple infections should be considered as a source of potential misclassification. In the present study we were able to infer phylogenies of several segments other than RdRp-encoding segments and perform analyses on co-segregation between the RdRp and the proteins encoded by those segments. While many clades of accepted genera formed monophyletic groups in segment trees, the topology between trees was not always congruent (**Supporting Information S15, S19, S24, S28, S29, and S33 Figs**). This suggests that reassortment was involved in the formation of major viral taxa. Topological conflicts in individual virus genomes appeared to occur more often in clades consisting of recently discovered viruses ([4, 8, 20], and our study), than in known and functionally-characterized viruses. Even if the present data were not found to disturb topological congruence in phylogenetic analyses of segment association, this emphasizes the general concern regarding hasty classification of viruses identified in metagenomic studies in the absence of experimental evidence.

In many instances, the comprehensiveness of our dataset allows us to estimate the extent of genetic space occupied by taxonomic entities below the family level. Issues such as long-branch attraction and loss of information by reduction of alignments [22] are alleviated by the addition of novel members to formerly solitary lineages, even if their genome sequences are incomplete [23]. The delineation of taxon-specific genetic space is further assisted by reliance on the

assumption of long-term co-segregation. In the absence of any other criteria, the knowledge of host associations can provide valuable information to define taxonomic units when newly discovered viruses are added to formerly monogeneric and monospecific families. In light of the obvious under-sampling within viral phylogenies, we regard it as an omission to not use host associations as assistance for virus classification decisions. Host associations allow the intuitive use of the principle of co-segregation that is confirmed for distinct, but related, insect-associated viruses in the present study. Several monophyletic clades in a number of viral families were found to be associated with defined orders of insects, yet all these associations were previously undiscovered [4, 8].

Our data suggest a potential to classify at least 27 novel genera (20 of them without any previously known species), and probably three novel families. Regarding viruses for which complete genomes or live isolates could not be obtained, the knowledge of host associations will orient future efforts to identify or isolate viruses in a more targeted manner. This will be of particular importance in insect groups that have relevance in food production or act as vectors of disease, as well as insect species that change their distribution and abundance due to environmental change.

The utilization of host associations for species classification is an accepted approach in virology, even if, in rare cases, spillovers or dual host associations will have to be taken into account and corrected for [24]. We expect cross-host transmission to be comparatively rare based on our analyses, and challenge the notion of frequent occurrence of cross host-virus transmission in the context of ecological and geographic proximity. The available studies are based on under-sampled viral trees and may suffer from other issues related to host assignment and topological correctness, in particular of viral phylogenies [4]. Tests of co-segregation need to take topological uncertainty into account and should be contrasted to null hypotheses.

We furthermore have concerns regarding the sole reliance on distance-based classification tools in the latest taxonomy proposals for negative strand RNA viruses [25, 26]. The utilized tool, DEmARC [27], analyzes concatenated sets of homologous protein domains that are conserved across the family of interest. It detects local regions of discontinuity in the pairwise distance spectrum within protein primary sequence alignments. Regions of discontinuity define putative limits of taxonomic units below the family level. It has only been evaluated for three families of vertebrate-infecting viruses (*Picornaviridae*, *Coronaviridae*, *Filoviridae*) and not for taxa that contain more than one family. Results for some of the larger virus groups, such as the genera *Enterovirus* versus *Rhinovirus*, have been controversial [28]. If applied to taxonomic units that exceed the family level, changes in genome architecture are common, and it becomes unclear whether protein-encoding genes other than the RdRp are homologous or have instead been acquired by lateral gene transfer [4, 8, 29]. The rate of recombination in newly-discovered viruses is unknown, and in segmented viruses this problem is aggravated by genome segment reassortment. Notably, DEmARC has not been validated at all for segmented viruses. On this basis, the analysis by distance-based tools has to be restricted to the RdRp gene and thus is hardly different from a phylogeny based only on RdRp genes. Other tools, such as the recently-proposed GRAViTy algorithm that incorporates genome composition, should be used and evaluated on the same problem [30].

As demonstrated in the present study and elsewhere [4, 8, 13], there is considerable topological uncertainty in large RNA virus phylogenies that comprise units of genetic diversity corresponding to viral orders. While genetic distance can be estimated in trees, its estimation is based on evolutionary models that cannot accommodate all biological factors of sequence evolution. For instance, viral population sizes, host generation times and infection or coinfection rates are expected to vary considerably between major lineages of viruses carried by hosts as different as mammals, birds, reptiles, fish, insects, spiders, crustaceans, worms, and protists.

As these host associations may have been in place since millions of years, effects such as substitution saturation or footprints of recombination are expected to influence the inference of deep phylogenetic relationships. Considering the shortcomings of evolutionary distance as a single classification criterion, additional biological criteria including host associations should be included in taxonomic considerations, if only as a test of plausibility.

Like all other approaches to detect viruses based on sequence information, our work has several additional limitations. For instance, it remains difficult to differentiate endogenous viral elements from replicating viruses. Our approach to extract virus sequence data from insect transcriptomes makes it highly unlikely that integrated, non-transcribed viruses are falsely assigned as viral matches. Earlier studies have relied on abundance of viral transcripts in comparison to total RNA content in RNAseq datasets, without the knowledge of complete transcriptome data. However, none of these approaches can exclude endogenous viral elements with certainty. For this reason, independent confirmation of findings is necessary before taxonomic conclusions are drawn. Our exact knowledge of host associations will enable host genome, as well as virus isolation studies that will ultimately exclude endogenization of viruses in the host germ line and confirm viral replication.

Materials and methods

Insect transcriptome data

We screened 1243 insect transcriptomes sequenced within the 1KITE project [5] including species from all recognized extant insect orders and additional arthropod orders. Samples were collected worldwide and RNAseq data were sequenced in an Illumina HiSeq2000 platform. Raw RNAseq data were assembled using SOAPdenovo-Trans-31kmer (version 1.01) [31], and checked for quality and cross contaminations with VecScreen (www.ncbi.nlm.nih.gov/tools/vecscreen/), and UniVec database build 7.0 (www.ncbi.nlm.nih.gov/tools/vecscreen/univec/).

Viral sequence generation and sorting

Template alignments for building profile hidden Markov models (pHMMs) were created using characterized RdRp amino acid sequences of negative strand RNA virus families. Due to the high sequence divergence of viral genes, even for closely related species, the sequence search was conducted at the amino acid level. Sequences were downloaded from the NCBI database, in October 2014, and aligned with the web-based alignment tool T-coffee in Expresso mode [32]. Transcriptome assemblies were translated in all six ORFs with the *fastatranslate* program within the package EXONERATE (version 2.2.0) [33]. This ORF library was scanned using HMMER version 3 [34] and only sequences with contiguous ORFs were regarded as viral matches. HMMER builds a pHMM from a template alignment and uses it to extract sequences that match the underlying probabilities of the model. This allows for detection of evolutionary distantly related sequences, with the advantage of remote RdRp homolog detection, but also the disadvantage of inflating the results with redundant duplicate sequences. Viral amino acid matches were checked for redundancy with a twofold approach: a) matches from each RdRp-pHMM were aligned to the original template alignments with MAFFT version 7.1.23, *E-INS-i* [35]. Poorly aligned regions and sequences that were too short and did not overlap with the selected alignment region were removed using trimAl [36] or manually upon inspection of the alignment in Geneious (Geneious v.9.1.8, Biomatters, Auckland, New Zealand, <https://www.geneious.com>), always complying with the preservation of known RdRp motifs. Trees were inferred with PhyML v.3.2.0 [37], using 1000 bootstrap replicates and Blossum62 amino acid substitution matrix. b) viral hits were compared with BLAST+ v.2.2.28+

[38] against the non-redundant NCBI protein database which had previously been filtered for viral sequences. This twofold approach enabled sorting the viral hits, thus removing any redundancy introduced by the HMM-search among different virus families.

Inference of alignments and phylogenies

To compose alignments for phylogeny, viral sequence hits (OKIAVs) from the present study were compared to GenBank as of August 2018 using BLASTP with an e-value cutoff of 10^{-6} . Sequences over 30% similarity to any OKIAV were selected. All species and genera listed in the ICTV taxonomy table as of end of 2018 were added. During the revision, selected additional species from the ICTV species update released in February 2019 were added. The following literature contributions were additionally consulted and relevant unclassified viruses were added: [39–57].

In total, 234 of the RdRp sequences found in insects in the present study were used for phylogenetic analyses (**Supporting Information S1 Table**). Alignments were calculated anew and refined with trimAl as described above. Model testing in MrBayes identified Blosum62 to be the amino acid substitution matrix compatible with all alignments. Trees were inferred in RAxML-NG version 0.7.0 BETA [58] plotting the transfer bootstrap expectation values [59]. Confirmatory phylogenetic analyses were done in PhyML v.3.2.0 [37] and MrBayes v3.2.6 [60], using the same substitution model and four different substitution rate categories with gamma distribution. For RAxML and PhyML, 1000 bootstrap replicates were computed, and for MrBayes chains were run until fully converged. All trees were plotted and annotated using the R package *ggtree* [61].

Virus genome organization

All ORFs of the full-length viral hits were annotated after comparing them against our customized viral database as well as with the InterProScan protein domain search tool [14].

Phylogenetic co-segregation of virus segments

Considering that orthomyxo- and bunyaviruses have segmented genomes, we additionally searched for proteins encoded by other segments (nucleoprotein, glycoprotein, PB2, and PA). For this search, we used the available protein sequences of the respective genera (NCBI) for a BLASTp search within transcriptomes we had detected the RdRp segments in already. Trees shown in **Supporting Information S15, S19, S24, S28, S29, and S33 Figs** include only those taxa for which additional protein genes were found. The R package *dendextend* [62], was used to create tanglegram figures, that allow examination of topological consistency among the trees. Jane [17] was used to match trees of the RdRp-encoding genome segments to trees of the other segments, based on costs for breaches of cophylogeny (best match = lowest costs). Costs were also determined when segment-segment associations were randomized and these pairs of trees were then subjected to cophylogeny optimization in Jane. To obtain a quantitative measure of topological congruence, the costs associated with the real datasets were divided by the costs with randomization (median from 1000 randomizations). The resulting value is a percentage that expresses the cophylogeny cost relative to a randomly-associated cophylogeny of same tree size and structure (resulting in 0 for perfect cophylogeny and 1 for absence of cophylogeny). These relative costs are expressed as percentages in **Fig 4**. Because adding branches to cophylogenies is expected to increase cophylogeny costs in randomized datasets, this was tested by Wilcoxon's paired samples test. In all comparisons, the differences were highly significant.

Analysis of co-segregation of viruses with their hosts

Host-virus associations for each phylogenetic tree were examined to assess concurrent phylogenetic relationships using Jane [17]. As a basis for the host tree we used a modified version of the arthropod phylogeny from Misof *et al.* [5]. Since Jane needs a host for each taxon of the virus tree, unknown or undefined hosts cannot be assessed. Therefore, we added outlier branches of unidentified insect, unidentified arthropod, and non-arthropod hosts, to enable mapping to non-arthropod and pooled insect/arthropod hosts. The co-evolution costs of the original phylogenies were compared to 1000 iterations of randomized host-virus associations. A one-sided z-test (implemented in the R package *BSDA* [63]) was used to test whether the randomized costs are at least 5% higher than the original costs. This threshold was set to ensure that miniscule cost changes do not lead to false interpretations.

CoRe-PA [18] was used to evaluate the co-evolutionary dependencies of the two major virus subphyla of *Haploviricotina* and *Polyploviricotina* with their corresponding insect hosts. Given a co-segregating scenario, CoRe-PA aims to find the most parsimonious reconciliation between host and virus trees by evaluating four co-evolutionary events: co-speciation, sorting, duplication, and host-switching. Each type of event is assigned a certain cost and the co-phylogenetic assessment that minimizes the total cost of events is accepted. Since both *Haploviricotina* and *Polyploviricotina* trees were unrooted, all possible rooted versions were evaluated, meaning that for every edge a rooted tree was created. Insect phylogenies were rooted to the arthropod order Chelicerata. For each of the previously 1000 unrooted RAxML bootstrap trees, 229 and 266 rooted trees were created for *Haploviricotina* and *Polyploviricotina* respectively. A reconciliation for each of these trees to the corresponding insect phylogeny was computed. To determine the strength and significance of host-virus co-evolution, each reconciliation was compared against a randomized association of each co-phylogenetic scenario, keeping the tree topologies unchanged. 100 randomized scenarios were computed by randomly renaming the host tree tips. This preserved the structure of host-virus associations, while avoiding bias introduction from sampling random trees. To estimate the fit of each randomized scenario, reconciliations were computed with the following costs: 0 for sorting and duplication, -1000 for co-speciation, and -0.001 for a host-switching event.

Completeness of genome segments

The completeness of viral segments was assessed for all segmented-related findings. Segments with size similar to known relative viruses were regarded as at least coding-complete regions, if the segment ORF was terminated by a stop codon within the segment. Bunyaviruses have segments that form panhandles, with conserved, species-specific termini [15, 16]. We examined the genome termini for complementarity, and also evaluated whether the termini of one segment match those of the other segments.

Cytochrome oxidase subunit 1 (COI) barcode analysis

To investigate the possibility that other organisms were accidentally collected and therefore whether the hosts of the OKIAV viruses might not actually be the intended sampled organisms, a barcode search was conducted based on two databases. First, 2,534,455 cytochrome c oxidase subunit 1 (COI) gene sequences from GenBank sequences were retrieved on October 10, 2019 with the query “txid2759[Organism:exp] AND cytochrome oxidase subunit 1[All fields]”. Second, the German Barcode of Life (GBOL) database, which contains barcode sequences from species recovered in Germany (Animalia: 287,377 barcodes including 261,015 hexapods; Plantae 7,884 barcodes, Fungi 1,038 barcodes). Contigs assembled from the 1243 insect transcriptomes were matched against these databases using BLAST+ (version 2.6.0).

The BLAST results were filtered for matches of length at least 500 nucleotides, with a nucleotide identity of at least 98%.

Of the 1243 insect transcriptomes, 34 (2.73%) have at least one contig that matches a non-Hexapoda barcode. The non-Hexapoda barcodes fall into 20 phylum/class categories, as shown in **Supporting Information S3 Table**.

Four of these 34 transcriptomes contained one or more negative strand RNA OKIAV, comprising a total of nine negative strand RNA viruses out of a total of 488 (1.8%) negative strand RNA viruses identified overall. Details of these four assemblies, their nine OKIAV viruses, and the matched non-Hexapoda barcodes are shown in **Supporting Information S4 Table**. Of the nine viruses, four are shown in the phylogenies in **Fig 1**, marked with an empty triangle to indicate the presence of a non-Hexapoda barcode in the associated assembly.

Supporting information

S1 Text. Supplementary results and discussion text.
(DOCX)

S1 Table. Additional data for the identified viral genomes. Information such as the host taxonomy, insect sample location, and collection date is provided.
(XLSX)

S2 Table. Read count and insect transcriptome library size for the full and nearly-full OKIAV genomes.
(XLSX)

S3 Table. Non-Hexapoda COI barcodes found across all insect transcriptome assemblies. Twenty combinations of non-Hexapoda phylum/class were found across all 1243 assemblies. Matches were required to be of at least 500 nucleotides, with at least a 98% nucleotide identity level.
(XLSX)

S4 Table. Detection of non-Hexapoda COI barcodes. Four assemblies (from which a total of nine OKIAV viruses were recovered) contained contigs that matched non-Hexapoda barcode sequences from the GBOL and NCBI databases (see **Materials and Methods**). Four of those nine viruses appear in phylogenies in **Fig 1**, where they are marked with an empty triangle. The five other viruses, indicated by asterisks, do not appear in **Fig 1** because they were not included in trees due to criteria of length, quality, and uniqueness. The table columns show, in order: the assembly identifier, the list of OKIAV viruses found in the assembly, the phylum and class of the non-Hexapoda organism(s) whose barcode was matched, and then for both GBOL and NCBI databases (when matches were found) the percentage nucleotide identity of the match and the length of the match.
(XLSX)

S1 Fig. Viruses pertaining to *Rhabdoviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, taxonomic grouping of hosts, and viral genus and family. The outgroup taxon (not shown) is *Mammalian 1 orthobornavirus* (*Bornaviridae*). Analyses based on PhyML and MrBayes can be found in **S2** and **S3 Figs**.
(TIF)

S2 Fig. Maximum likelihood phylogenies with PhyML of viruses pertaining to *Rhabdoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, taxonomic grouping of hosts, and viral genus and family. The outgroup taxon not shown) is *Mammalian 1 orthobornavirus (Bornaviridae)*.
(TIF)

S3 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Rhabdoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, taxonomic grouping of hosts, and viral genus and family. The outgroup taxon not shown) is *Mammalian 1 orthobornavirus (Bornaviridae)*.
(TIF)

S4 Fig. Alignment of *Almendravirus viroporins* and the potential precursor ORF of Hymenopteran almendra-related virus OKIAV1. Reference sequences belong to *Arboretum almendravirus (ABTV)*, *Puerto Almedras almendravirus (PTAMV)*, *Coot Bay almendravirus (CBV)*, *Balsa almendravirus (BALV)*, and *Rio Chico almendravirus (RCHV)*. The hydrophobic stretches with multiple leucins (L) and isoleucins (I) interact with the cell membrane to facilitate cell entry [1].
(TIF)

S5 Fig. Subtree of the sister clade of *Cyto-*, *Nucleo-*, *Dichorhabdo-*, and *Varicosavirus*. Maximum likelihood phylogenies based on RaxML (A) and Bayesian inference of phylogeny with MrBayes (B). Grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, and taxonomic grouping of hosts.
(TIF)

S6 Fig. Viruses pertaining to *Xinmoviridae*, *Nyamiviridae*, *Bornaviridae*, *Artoviridae*, *Lispiviridae*, *Paramyxoviridae*, *Sunviridae*, *Filoviridae*, and *Pneumoviridae*. Maximum likelihood phylogenies based on RAXML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, taxonomic grouping of hosts, and viral genus and family. The outgroup taxon (not shown) is *Salmonid rhabdovirus (Rhabdoviridae)*. Analyses based on PhyML and MrBayes can be found in S7 and S8 Figs.
(TIF)

S7 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Xinmoviridae*, *Nyamiviridae*, *Bornaviridae*, *Artoviridae*, *Lispiviridae*, *Paramyxoviridae*, *Sunviridae*, *Filoviridae*, and *Pneumoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, taxonomic grouping of hosts, and viral genus and family. The outgroup taxon (not shown) is *Salmonid rhabdovirus (Rhabdoviridae)*.
(TIF)

S8 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Xinmoviridae*, *Nyamiviridae*, *Bornaviridae*, *Artoviridae*, *Lispiviridae*, *Paramyxoviridae*, *Sunviridae*, *Filoviridae*, and *Pneumoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, taxonomic grouping of hosts, and viral genus and family.

The outgroup taxon (not shown) is *Salmonid rhabdovirus* (*Rhabdoviridae*).
(TIF)

S9 Fig. Viruses pertaining to *Chuviridae*, *Qinviridae*, and *Yueviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment length and organization are shown in parentheses: linear (L) or circular (C). For linear segments, information on the segment ends is given as: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to *Yuevirus* (*Yueviridae*). Analyses based on PhyML and MrBayes can be found in **S10** and **S11 Figs**.
(TIF)

S10 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Chuviridae*, *Qinviridae*, and *Yueviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment length and organization are shown in parentheses: linear (L) or circular (C). For linear segments, information on the sequence similarity of segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to *Yuevirus* (*Yueviridae*).
(TIF)

S11 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Chuviridae*, *Qinviridae*, and *Yueviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment length and organization are shown in parentheses: linear (L) or circular (C). For linear segments, information on the sequence similarity of segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to *Yuevirus* (*Yueviridae*).
(TIF)

S12 Fig. Viruses pertaining to *Orthomyxoviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: PB1 = polymerase subunit PB1, PB2 = polymerase subunit PB2, PA = polymerase subunit PA, G = glycoprotein, N = nucleoprotein, H = hemagglutinin, NA = neuraminidase, M = matrix protein, NS = non-structural protein, and Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on the segment ends is indicated by: (n) = segment ends not matching the ends of

the PB1 segment, (y) = segment ends matching the ends of the PB1 segment, (p) = segment ends partially matching the ends of the PB1 segment. The outgroup taxon (not shown) is *Salmon isavirus* (*Isavirus*, *Orthomyxoviridae*). Analyses based on PhyML and MrBayes can be found in **S13** and **S14 Figs**.

(TIF)

S13 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Orthomyxoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: PB1 = polymerase subunit PB1, PB2 = polymerase subunit PB2, PA = polymerase subunit PA, G = glycoprotein, N = nucleoprotein, H = hemagglutinin, NA = neuraminidase, M = matrix protein, NS = non-structural protein, and Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on the sequence similarity of segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the PB1 segment, (y) = segment ends matching the ends of the PB1 segment, (p) = segment ends partially matching the ends of the PB1 segment. The outgroup taxon (not shown) is *Salmon isavirus* (*Isavirus*, *Orthomyxoviridae*).

(TIF)

S14 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Orthomyxoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: PB1 = polymerase subunit PB1, PB2 = polymerase subunit PB2, PA = polymerase subunit PA, G = glycoprotein, N = nucleoprotein, H = hemagglutinin, NA = neuraminidase, M = matrix protein, NS = non-structural protein, and Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on the sequence similarity of segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the PB1 segment, (y) = segment ends matching the ends of the PB1 segment, (p) = segment ends partially matching the ends of the PB1 segment. The outgroup taxon (not shown) is *Salmon isavirus* (*Isavirus*, *Orthomyxoviridae*).

(TIF)

S15 Fig. Phylogenetic co-segregation between PB1 and PB2, PB1 and PA, and PB1 and nucleoprotein of the viruses pertaining to *Orthomyxoviridae*. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. Not all genomic segments have been identified for all viral taxa; therefore, the number of taxa varies among the tree pairs. Clade H, mainly composed of the *Orthomyxoviridae* genera, is not congruent between the PB1 and the nucleoprotein trees and shows two different topologies within *Thogotovirus*. The PB1 segments of Coleopteran orthomyxo-related virus OKIAV196 and -200 are both found within the same insect transcriptome (*Ips typographus*). Within this transcriptome we have also identified one PB2, one PA, and one nucleoprotein segments. The co-segregation analysis allowed us to assign all the latter segments to Coleopteran orthomyxo-related virus OKIAV200 rather than -196.

In the phylogenies of PB1 and PA, 52 of 54 taxa are distributed in eight monophyletic clades. Clade B consists of three distinct inner clades that are topology-wise stable and share

similarities within their inner parts. Whereas clades D and E are direct sisters in the PB1 tree, clade D is not directly linked to clade E in the PA tree. However, the PA tree has a higher support in this region of the tree. The position of clade G is maintained across all trees, except of Coleopteran orthomyxo-related virus OKIAV196 in the PA tree, which actually belongs next to Zygentoman orthomyxo-related virus OKIAV204 (clade A). Clade H consists exclusively of ICTV-accepted genera of *Orthomyxoviridae*, except *Hubei orthoptera virus 6*. These clades (G and H) show identical topology in both trees.

In the phylogenies of PB1 and PB2, 43 of 46 taxa are distributed in six monophyletic clades. Clades D and E are merely represented by Archaeognathan orthomyxo-related virus OKIAV189, Hemipteran orthomyxo-related virus OKIAV188, and Neuropteran orthomyxo-related virus OKIAV190. Clade B shows a similar topology as in PA and PB1, and Clade G maintains its phylogenetic position, except of Coleopteran orthomyxo-related virus OKIAV196 in the PB2 tree, because it belongs next to Zygentoman orthomyxo-related virus OKIAV204 of clade A.

For the nucleoprotein, 45 of 57 taxa were distributed in eight monophyletic clades. Clade G maintains its topology in this phylogeny as well, showing additionally that the identified N-segment is indeed part of Zygentoman orthomyxo-related virus OKIAV204. Clade E is a sister to clade G on the nucleoprotein tree, in contrast to the PB1 tree where it is sister to clade D. (TIF)

S16 Fig. Viruses pertaining to *Hantaviridae*, *Cruliviridae*, *Peribunyaviridae*, and *Fimoviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. The outgroup taxon (not shown) is Collembolan hanta-related virus OKIAV223 (shown in S20 Fig). Analyses based on PhyML and MrBayes can be found in S17 and S18 Figs. (TIF)

S17 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Hantaviridae*, *Cruliviridae*, *Peribunyaviridae*, and *Fimoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The outgroup taxon (not shown) is Collembolan hanta-related virus OKIAV223 (shown in S20 Fig). (TIF)

S18 Fig. Maximum likelihood phylogeny based on RAxML (A) and Bayesian phylogeny inference with MrBayes (B) of viruses pertaining to *Hantaviridae*, *Cruliviridae*, *Peribunyaviridae*, and *Fimoviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends

partially matching the ends of the RdRp segment. The outgroup taxon (not shown) is Collem-bolan hanta-related virus OKIAV223 (shown in [S20 Fig](#)).
(TIF)

S19 Fig. Phylogenetic co-segregation between RdRp and glycoprotein, and RdRp and nucleoprotein of the viruses pertaining to *Hantaviridae*, *Cruliviridae*, *Peribunyaviridae*, and *Fimoviridae*. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. In the phylogenies of RdRp and glycoprotein, 55 of 59 taxa were distributed in five monophyletic clades represented by the genera *Orthobunyavirus*, *Herbevirus*, *Tospovirus*, *Orthohantavirus*, and *Emaravirus*. Despite spanning four different families, the phylogenies are well-supported and some of the (sub)topologies can be confirmed. *Tospovirus* and *Emaravirus* have a completely congruent topology, while *Orthobunyavirus* and *Orthohantavirus* have some topologically stable subclades. Noteworthy are not only the congruent topologies, but also the very similar branch lengths of *Tospovirus*, *Orthohantavirus*, and *Emaravirus*. Based only on the tree topology, an assignment of the M-segment to either Dipluran hanta-related virus OKIAV217 or -218 cannot be assessed. In the phylogenies of RdRp and nucleoprotein, 44 of 49 taxa distributed in the genera *Orthobunyavirus*, *Herbevirus*, *Tospovirus*, *Orthohantavirus*, and *Emaravirus*. *Tospovirus* is the only genus that retains its inner topological structure among all trees, except of the position of *Bean necrotic mosaic virus*, that in the nucleoprotein tree is sister to all other tospoviruses. However, the position of *Tospovirus* within *Peribunyaviridae* is not supported by the nucleoprotein phylogeny. Apart from that, all genera are still monophyletic. The position of *Khurdun virus* as the first split from *Herbevirus* can be confirmed by both nucleoprotein and glycoprotein phylogenies.
(TIF)

S20 Fig. Representative viruses of *Bunyvirales*. Bayesian inference of phylogeny based on MrBayes. Black branches show selected reference taxa, and red branches show some OKIAVs.
(TIF)

S21 Fig. Viruses pertaining to *Phasmaviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on the segment ends is indicated by: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree was rooted using a hantavirus outgroup. The outgroup was then removed, the tree recalculated, and the rooting between *Orthophasmavirus* vs. (*HAF*, *Feravirus*, *Wuhivirus*, and *Jonvirus*) was maintained. Analyses based on PhyML and MrBayes can be found in [S22](#) and [S23 Figs](#).
(TIF)

S22 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Phasmaviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well:

(n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to the exclusion of the lower clade of *Phasmaviridae* (HAF, *Feravirus*, *Wuhivirus*, *Jonvirus*) following preliminary analyses using a hantavirus outgroup. (TIF)

S23 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Phasmaviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to the lower clade of *Phasmaviridae* (HAF, *Feravirus*, *Wuhivirus*, *Jonvirus*) following preliminary analyses using a hantavirus outgroup. (TIF)

S24 Fig. Phylogenetic co-segregation between RdRp and glycoprotein, and RdRp and nucleoprotein of the viruses pertaining to *Phasmaviridae*. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. In the phylogenies of RdRp and nucleoprotein, 34 of 40 taxa were distributed in three monophyletic clades that consist of the ICTV-accepted genera *Feravirus*, *Wuhivirus*, and *Orthophasmavirus*. In the nucleoprotein phylogeny, clades DAP and HAP form inner congruent monophyletic groups within *Orthophasmavirus*. The same pattern applies to 30 of 36 taxa in the glycoprotein phylogeny. Within *Orthophasmavirus*, HAP is the most stable clade, with a subclade of five taxa that have identical topology among the phylogenies. However, the bootstrap support of the *Orthophasmavirus* subclades is below 80%. The clear subdivision into clades A and B on the RdRp and nucleoprotein trees is not verified in the glycoprotein phylogeny, yet *Feravirus* and HAF group together with a high support (99%), and maintain their inner topological structures. Most of the *Orthophasmavirus* taxa have not been subjected to laboratory studies. Additionally, the OKIAV sequences are only fragmentarily assembled. Obtaining stable and congruent phylogenies among genomic segments can thus not be expected. (TIF)

S25 Fig. Viruses pertaining to *Phenuiviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on the segment ends is indicated by: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to the putative subfamily. Analyses based on PhyML and MrBayes can be found in **S26** and **S27 Figs**. (TIF)

S26 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Phenuiviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red

branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to the putative subfamily. (TIF)

S27 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Phenuiviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The tree is rooted to the putative subfamily. (TIF)

S28 Fig. Phylogenetic co-segregation between RdRp and glycoprotein, and RdRp and nucleoprotein of the viruses pertaining to *Phenuiviridae*. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. The lack of complete genomes for most of the taxa that appear on the trees causes high topological conflict between the phylogenies of the different segments. The glycoprotein- and nucleoprotein-segments have not been identified for most of the OKIAV. The bootstrap support on the clades of the single-species genera *Hudovirus*, *Pidchovirus*, *Hudivirus*, *Beidivirus*, and *Horwuvirus* is low in comparison to the rest of the tree. Thus, the phylogenetic signal is probably not sufficient to draw meaningful conclusions on co-segregations for these genera. Additionally, within the putative subfamily, the lack of genomic segments for co-segregation analysis does not allow us drawing conclusions either. In the phylogenies of RdRp and nucleoprotein, 45 of 64 taxa are distributed in topologically stable monophyletic clades within the genera *Phasivirus*, *Wubeivirus*, *Tenuivirus*, *Phlebovirus* (with the exception of clade B), *Banyangvirus*, *Goukovirus*, and additionally the unclassified clades A, E, and F. Within *Phlebovirus*, clade B is sister to clade A and includes the genus *Tenuivirus*. Clade C is topologically congruent among all three phylogenies. Both clades C and D maintain their taxa composition across all three trees as well as their relation to *Banyangvirus*. The topological stability of the *Banyangvirus* clade within the *Phlebovirus* clade, suggests that *Banyangvirus* should rather be classified as a sub-genus of *Phlebovirus*. In the phylogenies of RdRp and glycoprotein, 35 of 41 taxa are distributed in topologically stable monophyletic clades that are accepted genera and the unclassified clades D, E, and F. *Hudivirus* and *Beidivirus* group together in this case, indicating a close relationship. A doubtful classification is the one of *Wubeivirus*: it is monophyletic only in the glycoprotein phylogeny, but groups consistently with *Phasivirus* in all phylogenies. (TIF)

S29 Fig. Phylogenetic co-segregation between RdRp and glycoprotein, and RdRp and nucleoprotein of the viruses pertaining to *Phlebovirus* and *Banyangvirus*. Topologically

congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. *Banyangvirus* is sister to the main *Phlebovirus* clades in the RdRp phylogeny, indicating that *Banyangvirus* should not be regarded as an independent genus.

(TIF)

S30 Fig. Viruses pertaining to *Arenaviridae*, *Myxoviridae*, *Nairoviridae*, and *Wupedeviridae*. Maximum likelihood phylogenies based on RAxML. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on the segment ends is indicated by: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The outgroup taxon (not shown) is *Rift Valley fever phlebovirus (Phenuiviridae)*. Analyses based on PhyML and MrBayes can be found in **S31** and **S32 Figs**.

(TIF)

S31 Fig. Maximum likelihood phylogeny with PhyML of viruses pertaining to *Arenaviridae*, *Myxoviridae*, *Nairoviridae*, and *Wupedeviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The outgroup taxon (not shown) is *Rift Valley fever phlebovirus (Phenuiviridae)*.

(TIF)

S32 Fig. Bayesian phylogeny inference with MrBayes of viruses pertaining to *Arenaviridae*, *Myxoviridae*, *Nairoviridae*, and *Wupedeviridae*. Black branches show ICTV-accepted taxa, grey branches show unclassified taxa, and red branches show OKIAVs. Columns on the right side summarize contig length, genome completeness, number of segments, taxonomic grouping of hosts, and viral genus and family. Genomic protein-coding regions are: R = RdRp, G = glycoprotein, N = nucleoprotein, Hy = hypothetical protein with unknown function. Segment lengths are shown in parentheses. Information on sequence similarity of the segment ends among different genome segments is given as well: (n) = segment ends not matching the ends of the RdRp segment, (y) = segment ends matching the ends of the RdRp segment, (p) = segment ends partially matching the ends of the RdRp segment. The outgroup taxon (not shown) is *Rift Valley fever phlebovirus (Phenuiviridae)*.

(TIF)

S33 Fig. Phylogenetic co-segregation between RdRp and glycoprotein, and RdRp and nucleoprotein of the viruses pertaining to *Arenaviridae*, *Myxoviridae*, *Nairoviridae*, and *Wupedeviridae*. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. In the phylogenies of RdRp and nucleoprotein, 31 of 38 taxa are distributed in the monophyletic genera *Orthonairovirus*, *Reptarenavirus*, and *Mammarenavirus*. *Reptarenavirus* is the only

genus that consistently has a congruent topology among all trees. There are very few viruses that are not formally accepted by ICTV in this tree, and the ones that are new fit well in between the established genera, resulting in a stable backbone of the phylogeny. In the phylogenies of RdRp and glycoprotein, 30 of 39 taxa are distributed in *Orthonairovirus*, *Reptarenavirus*, and *Mammarenavirus*. The glycoprotein phylogeny shows similar topology to the nucleoprotein one, but only *Reptarenavirus* maintains its topological position. The biggest disagreement is the positioning of *Reptarenavirus* within the *Nairoviridae* clade. The backbones of the trees are not in agreement, therefore the positions of the single-species genera *Shaspi-virus*, *Wumivirus*, and *Hubevirus* cannot be confirmed. However, in both phylogenies, the positions of *Striwavivirus*, Blattodean nairo-related virus OKIAV321, and *Xinzhou spider virus* are stable.

(TIF)

S34 Fig. Read mapping on the genome of Odonatan chu-related virus OKIAV137. The sequence is joined head-to-tail, genome start and end are colored and indicated by arrows. The end-to-start gap is solely bridged by two flanking nucleotides of four reads (marked in red). The ORFs encoding for glycoprotein (G), nucleoprotein (N), and RdRp, as well as the read coverage are shown.

(TIF)

Acknowledgments

We thank Dr. Karen Meusemann, Dr. Jeanne Wilbrandt, and Dr. Tanja Ziesmann for providing the meta data of the 1KITE transcriptomes. We thank Dr. Jens Kuhn for his helpful comments on an earlier manuscript version.

Author Contributions

Conceptualization: Simon Käfer, Sofia Paraskevopoulou, Alexander Donath, Sandra Junglen, Bernhard Misof, Christian Drosten.

Data curation: Simon Käfer, Sofia Paraskevopoulou, Florian Zirkel, Shanlin Liu, Xin Zhou.

Formal analysis: Simon Käfer, Sofia Paraskevopoulou, Florian Zirkel, Nicolas Wieseke, Martin Middendorf, Sandra Junglen.

Funding acquisition: Christian Drosten.

Investigation: Simon Käfer, Sofia Paraskevopoulou, Florian Zirkel, Sandra Junglen, Christian Drosten.

Methodology: Simon Käfer, Sofia Paraskevopoulou, Florian Zirkel, Nicolas Wieseke, Alexander Donath, Shanlin Liu, Xin Zhou, Martin Middendorf, Sandra Junglen.

Project administration: Christian Drosten.

Resources: Alexander Donath, Terry C. Jones, Martin Middendorf.

Software: Simon Käfer, Sofia Paraskevopoulou, Florian Zirkel, Nicolas Wieseke, Alexander Donath, Malte Petersen, Terry C. Jones, Martin Middendorf.

Supervision: Sandra Junglen, Bernhard Misof, Christian Drosten.

Validation: Simon Käfer, Sofia Paraskevopoulou.

Visualization: Simon Käfer, Sofia Paraskevopoulou, Florian Zirkel, Nicolas Wieseke.

Writing – original draft: Simon Käfer, Sofia Paraskevopoulou, Sandra Junglen, Bernhard Misof, Christian Drosten.

Writing – review & editing: Simon Käfer, Sofia Paraskevopoulou, Terry C. Jones, Bernhard Misof, Christian Drosten.

References

1. Claas ECJ, Osterhaus AD, van Beek R, De Jong JC, Rimmelzwaan GF, Senne DA, et al. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*. 1998; 351(9101):472–7. [https://doi.org/10.1016/S0140-6736\(97\)11212-0](https://doi.org/10.1016/S0140-6736(97)11212-0) PMID: 9482438
2. Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, Binger T, et al. Bats host major mammalian paramyxoviruses. *Nat Commun*. 2012; 3:796. <https://doi.org/10.1038/ncomms1796> PMID: 22531181
3. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*. 2008; 451:990–3. <https://doi.org/10.1038/nature06536> PMID: 18288193
4. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife*. 2015; 4. <https://doi.org/10.7554/eLife.05378> PMID: 25633976
5. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014; 346(6210):763–7. <https://doi.org/10.1126/science.1257570> PMID: 25378627
6. Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, et al. The evolutionary history of vertebrate RNA viruses. *Nature*. 2018; 556(7700):197–202. <https://doi.org/10.1038/s41586-018-0012-7> PMID: 29618816
7. Junglen S, Drosten C. Virus discovery and recent insights into virus diversity in arthropods. *Curr Opin Microbiol*. 2013; 16(4):507–13. <https://doi.org/10.1016/j.mib.2013.06.005> PMID: 23850098
8. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016; 540(7634):539–43. <https://doi.org/10.1038/nature20167> PMID: 27880757
9. Simmonds P, Adams MJ, Benkó M, Breitbart M, Brister JR, Carstens EB, et al. Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*. 2017; 15:161. <https://doi.org/10.1038/nrmicro.2016.177>
10. Wignall-Fleming EB, Hughes DJ, Vattipally S, Modha S, Goodbourn S, Davison AJ, et al. Analysis of Paramyxovirus Transcription and Replication by High-Throughput Sequencing. *J Virol*. 2019; 93(17). Epub 2019/06/14. <https://doi.org/10.1128/jvi.00571-19> PMID: 31189700; PubMed Central PMCID: PMC6694822.
11. Di Giallonardo F, Audsley MD, Shi M, Young PR, McGraw EA, Holmes EC. Complete genome of *Aedes aegypti* anphevirus in the Aag2 mosquito cell line. *Journal of General Virology*. 2018; 99(6):832–6. <https://doi.org/10.1099/jgv.0.001079> PMID: 29741476
12. Schoonvaere K, De Smet L, Smaghe G, Vierstraete A, Braeckman BP, de Graaf DC. Unbiased RNA Shotgun Metagenomics in Social and Solitary Wild Bees Detects Associations with Eukaryote Parasites and New Viruses. *PLoS One*. 2016; 11(12):e0168456. <https://doi.org/10.1371/journal.pone.0168456> PMID: 28006002
13. Wolf YI, Kazlauskas D, Iranzo J, Lucia-Sanz A, Kuhn JH, Krupovic M, et al. Origins and Evolution of the Global RNA Virome. *MBio*. 2018; 9(6). Epub 2018/11/30. <https://doi.org/10.1128/mBio.02329-18> PMID: 30482837; PubMed Central PMCID: PMC6282212.
14. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014; 30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
15. Elliott RM. Molecular biology of the Bunyviridae. *J Gen Virol*. 1990; 71 (Pt 3):501–22. <https://doi.org/10.1099/0022-1317-71-3-501> PMID: 2179464
16. Mir MA, Brown B, Hjelle B, Duran WA, Panganiban AT. Hantavirus N protein exhibits genus-specific recognition of the viral RNA panhandle. *J Virol*. 2006; 80(22):11283–92. <https://doi.org/10.1128/JVI.00820-06> PMID: 16971445
17. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms Mol Biol*. 2010; 5:16. <https://doi.org/10.1186/1748-7188-5-16> PMID: 20181081
18. Merkle D, Middendorf M, Wieseke N. A parameter-adaptive dynamic programming approach for inferring cophylogenies. *BMC Bioinformatics*. 2010; 11 Suppl 1:S60. <https://doi.org/10.1186/1471-2105-11-S1-S60> PMID: 20122236

19. Halloran A, Vantomme P, Hanboonsong Y, Ekesi S. Regulating edible insects: the challenge of addressing food security, nature conservation, and the erosion of traditional food culture. *Food Security*. 2015; 7(3):739–46. <https://doi.org/10.1007/s12571-015-0463-8>
20. Makhssous N, Shean RC, Droppers D, Guan J, Jerome KR, Greninger AL. Genome Sequences of Three Novel Bunyaviruses, Two Novel Rhabdoviruses, and One Novel Nyamivirus from Washington State Moths. *Genome Announc*. 2017; 5(7). <https://doi.org/10.1128/genomeA.01668-16> PMID: 28209840
21. Marklewitz M, Zirkel F, Rwego IB, Heidemann H, Trippner P, Kurth A, et al. Discovery of a unique novel clade of mosquito-associated bunyaviruses. *J Virol*. 2013; 87(23):12850–65. <https://doi.org/10.1128/JVI.01862-13> PMID: 24067954
22. Bergsten J. A review of long-branch attraction. *Cladistics*. 2005; 21(2):163–93.
23. Wiens JJ. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst Biol*. 2005; 54(5):731–42. <https://doi.org/10.1080/10635150500234583> PMID: 16243761
24. Obbard DJ. Expansion of the metazoan virosphere: Progress, pitfalls, and prospects. *Current opinion in virology*. 2018; 31:17–23. <https://doi.org/10.1016/j.coviro.2018.08.008> PMID: 30237139
25. Maes P, Song T, Mark S, Paweska J, Song Q, Ye G, et al. ICTV taxonomic report 2017.016M.R.: Taxonomic expansion and reorganization of the order Mononegavirales.
26. Maes P, Alkhovsky S, Beer M, Briese T, Buchmeier MJ, Calisher CH, et al. ICTV taxonomic proposal 2017.012M: Taxonomic expansion and reorganization of the order Bunyavirales.
27. Lauber C, Gorbalenya AE. Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. *J Virol*. 2012; 86(7):3890–904. <https://doi.org/10.1128/JVI.07173-11> PMID: 22278230
28. Simmonds P. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J Gen Virol*. 2015; 96(Pt 6):1193–206. <https://doi.org/10.1099/jgv.0.000016> PMID: 26068186
29. Schuster S, Zirkel F, Kurth A, van Cleef KWR, Drosten C, van Rij RP, et al. A Unique Nodavirus with Novel Features: Mosinovirus Expresses Two Subgenomic RNAs, a Capsid Gene of Unknown Origin, and a Suppressor of the Antiviral RNA Interference Pathway. *Journal of Virology*. 2014; 88(22):13447–59. <https://doi.org/10.1128/JVI.02144-14> PMID: 25210176
30. Aiewsakun P, Adriaenssens EM, Lavigne R, Kropinski AM, Simmonds P. Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. *J Gen Virol*. 2018; 99:1331–43. <https://doi.org/10.1099/jgv.0.001110> PMID: 30016225
31. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010; 20(2):265–72. <https://doi.org/10.1101/gr.097261.109> PMID: 20019144
32. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 2000; 302(1):205–17. <https://doi.org/10.1006/jmbi.2000.4042> PMID: 10964570
33. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005; 6:31. <https://doi.org/10.1186/1471-2105-6-31> PMID: 15713233
34. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol*. 2011; 7(10):e1002195. <https://doi.org/10.1371/journal.pcbi.1002195> PMID: 22039361
35. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002; 30(14):3059–66. <https://doi.org/10.1093/nar/gkf436> PMID: 12136088
36. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945
37. Guindon S, Dufayard J-F, 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(3):307–21. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638
38. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
39. Contreras MA, Eastwood G, Guzman H, Popov V, Savit C, Uribe S, et al. Almendravirus: A Proposed New Genus of Rhabdoviruses Isolated from Mosquitoes in Tropical Regions of the Americas. *Am J Trop Med Hyg*. 2017; 96:100–9. <https://doi.org/10.4269/ajtmh.16-0403> PMID: 27799634

40. Reuter G, Boros Á, Pál J, Kapusinszky B, Delwart E, Pankovics P. Detection and genome analysis of a novel (dima)rhabdovirus (Riverside virus) from *Ochlerotatus* sp. mosquitoes in Central Europe. *Infect Genet Evol.* 2016; 39:336–41. <https://doi.org/10.1016/j.meegid.2016.02.016> PMID: 26883377
41. Shahhosseini N, Lühken R, Jöst H, Jansen S, Börstler J, Rieger T, et al. Detection and characterization of a novel rhabdovirus in *Aedes cantans* mosquitoes and evidence for a mosquito-associated new genus in the family Rhabdoviridae. *Infect Genet Evol.* 2017; 55:260–8. <https://doi.org/10.1016/j.meegid.2017.09.026> PMID: 28943405
42. Sun Q, Zhao Q, An X, Guo X, Zuo S, Zhang X, et al. Complete genome sequence of Menghai rhabdovirus, a novel mosquito-borne rhabdovirus from China. *Arch Virol.* 2017; 162:1103–6. <https://doi.org/10.1007/s00705-016-3188-x> PMID: 28000049
43. Tokarz R, Sameroff S, Tagliaferro T, Jain K, Williams SH, Cucura DM, et al. Identification of Novel Viruses in *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* Ticks. *mSphere.* 2018; 3. <https://doi.org/10.1128/mSphere.00614-17> PMID: 29564401
44. Xu CL, Cantalupo PG, Sáenz-Robles MT, Baldwin A, Fitzpatrick D, Norris DE, et al. Draft Genome Sequence of a Novel Rhabdovirus Isolated from *Deinocerites* Mosquitoes. *Genome Announc.* 2018; 6. <https://doi.org/10.1128/genomeA.01438-17> PMID: 29748415
45. Axén C, Hakhverdyan M, Boutrup TS, Blomkvist E, Ljunghager F, Alfjorden A, et al. Emergence of a new rhabdovirus associated with mass mortalities in eelpout (*Zoarces viviparus*) in the Baltic Sea. *J Fish Dis.* 2017; 40:219–29. <https://doi.org/10.1111/jfd.12506> PMID: 27416895
46. Charles J, Firth AE, Loroño-Pino MA, Garcia-Rejon JE, Farfan-Ale JA, Lipkin WI, et al. Merida virus, a putative novel rhabdovirus discovered in *Culex* and *Ochlerotatus* spp. mosquitoes in the Yucatan Peninsula of Mexico. *J Gen Virol.* 2016; 97:977–87. <https://doi.org/10.1099/jgv.0.000424> PMID: 26868915
47. Cholleti H, Hayer J, Abilio AP, Mulandane FC, Verner-Carlsson J, Falk KI, et al. Discovery of Novel Viruses in Mosquitoes from the Zambezi Valley of Mozambique. *PLoS One.* 2016; 11:e0162751. <https://doi.org/10.1371/journal.pone.0162751> PMID: 27682810
48. Lara Pinto AZd, Santos de Carvalho M, de Melo FL, Ribeiro ALM, Moraes Ribeiro B, Dezengrini Silhesarenko R. Novel viruses in salivary glands of mosquitoes from sylvatic Cerrado, Midwestern Brazil. *PLoS One.* 2017; 12:e0187429. <https://doi.org/10.1371/journal.pone.0187429> PMID: 29117239
49. Økland AL, Nylund A, Øvergård A-C, Blindheim S, Watanabe K, Grotmol S, et al. Genomic characterization and phylogenetic position of two new species in Rhabdoviridae infecting the parasitic copepod, salmon louse (*Lepeophtheirus salmonis*). *PLoS One.* 2014; 9:e112517. <https://doi.org/10.1371/journal.pone.0112517> PMID: 25402203
50. Sabbadin F, Glover R, Stafford R, Rozado-Aguirre Z, Boonham N, Adams I, et al. Transcriptome sequencing identifies novel persistent viruses in herbicide resistant wild-grasses. *Sci Rep.* 2017; 7:41987. <https://doi.org/10.1038/srep41987> PMID: 28165016
51. Shi M, Neville P, Nicholson J, Eden J-S, Imrie A, Holmes EC. High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia. *J Virol.* 2017; 91. <https://doi.org/10.1128/JVI.00680-17> PMID: 28637756
52. Simo Tchétgna HD, Nakoune E, Selekon B, Gessain A, Manuguerra J-C, Kazanji M, et al. Molecular Characterization of the Kamese Virus, an Unassigned Rhabdovirus, Isolated from *Culex pruinia* in the Central African Republic. *Vector Borne Zoonotic Dis.* 2017; 17:447–51. <https://doi.org/10.1089/vbz.2016.2068> PMID: 28350284
53. Coffey LL, Page BL, Greninger AL, Herring BL, Russell RC, Doggett SL, et al. Enhanced arbovirus surveillance with deep sequencing: Identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes. *Virology.* 2014; 448:146–58. <https://doi.org/10.1016/j.virol.2013.09.026> PMID: 24314645
54. Contreras-Gutiérrez MA, Nunes MRT, Guzman H, Uribe S, Suaza Vasco JD, Cardoso JF, et al. Sinu virus, a novel and divergent orthomyxovirus related to members of the genus *Thogotovirus* isolated from mosquitoes in Colombia. *Virology.* 2017; 501:166–75. <https://doi.org/10.1016/j.virol.2016.11.014> PMID: 27936462
55. Ejiri H, Lim C-K, Isawa H, Fujita R, Murota K, Sato T, et al. Characterization of a novel thogotovirus isolated from *Amblyomma testudinarium* ticks in Ehime, Japan: A significant phylogenetic relationship to Bourbon virus. *Virus Res.* 2018; 249:57–65. <https://doi.org/10.1016/j.virusres.2018.03.004> PMID: 29548745
56. Medd NC, Fellous S, Waldron FM, Xuéreb A, Nakai M, Cross JV, et al. The virome of *Drosophila suzukii*, an invasive pest of soft fruit. *Virus Evol.* 2018; 4:vey009. <https://doi.org/10.1093/ve/vey009> PMID: 29644097
57. Rodrigues DSG, Medeiros DBdA, Rodrigues SG, Martins LC, de Lima CPS, de Oliveira LF, et al. Pacui Virus, Rio Preto da Eva Virus, and Tapirape Virus, Three Distinct Viruses within the Family Bunyaviridae. *Genome Announc.* 2014; 2. <https://doi.org/10.1128/genomeA.00923-14> PMID: 25395627

58. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAXML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019; 35(21):4453–5. <https://doi.org/10.1093/bioinformatics/btz305> PMID: 31070718
59. Lemoine F, B. Domelevo Entfellner J, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*. 2018; 556(7702):452–6. <https://doi.org/10.1038/s41586-018-0043-0> PMID: 29670290
60. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012; 61(3):539–42. <https://doi.org/10.1093/sysbio/sys029> PMID: 22357727
61. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2016; 8(1):28–36. <https://doi.org/10.1111/2041-210x.12628>
62. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*. 2015; 31(22):3718–20. <https://doi.org/10.1093/bioinformatics/btv428> PMID: 26209431
63. Arnholt AT, Evans B. BSDA: Basic Statistics and Data Analysis. 2017.

3 Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*

This publication is the result of work conducted at the Institute of Virology in Charité-Universitätsmedizin Berlin, additionally carried out in collaboration with Prof. Dr. Dieter Glebe from the Institute of Medical Virology, at Justus Liebig University in Giessen. The original material was sampled and provided by the group of Prof. Dr. Simone Sommer at the Institute of Evolutionary Ecology and Conservation Genomics at the University of Ulm. While performing data analysis on the original samples in the frame of my virus discovery pipeline, the presence of a deltavirus-like sequence initiated this project. Extensive wet lab experimental work was mainly carried out by Fabian Pirzer and Nora Goldmann, with whom I share the first authorship in this publication.

Published as: Paraskevopoulou S., Pirzer F., Goldmann N., Schmid J., Corman V.M., Gottula L.T., Schroeder S., Rasche A., Muth D., Drexler J.F., Heni A.C., Eibner G.J., Page R.A., Jones T.C., Müller M.A., Sommer S., Glebe D., and Drosten C. (2020). Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*. *Proceedings of the National Academy of Sciences*, 117(30):17977–17983.

This publication is also available at: <https://doi.org/10.1073/pnas.2006750117>.



Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*

Sofia Paraskevopoulou^{a,1}, Fabian Pirzer^{a,1}, Nora Goldmann^{b,c,d,1}, Julian Schmid^{e,f}, Victor Max Corman^{a,g}, Lina Theresa Gottula^a, Simon Schroeder^a, Andrea Rasche^{a,g}, Doreen Muth^{a,2}, Jan Felix Drexler^{a,g,h}, Alexander Christoph Heni^{e,f}, Georg Joachim Eibner^{a,e,f}, Rachel A. Page^f, Terry C. Jones^{a,i}, Marcel A. Müller^{a,g,h}, Simone Sommer^{e,3}, Dieter Glebe^{b,c,d,3}, and Christian Drosten^{a,g,3}

^aInstitute of Virology, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany; ^bInstitute of Medical Virology, Justus Liebig University Giessen, 35392 Giessen, Germany; ^cNational Reference Centre for Hepatitis B Viruses and Hepatitis D Viruses, 35392 Giessen, Germany; ^dGerman Centre for Infection Research (DZIF), partner site Giessen, 35392 Giessen, Germany; ^eInstitute of Evolutionary Ecology and Conservation Genomics, University of Ulm, 89069 Ulm, Germany; ^fSmithsonian Tropical Research Institute, Apartado Postal 0843-03092, Balboa, Republic of Panama; ^gGerman Centre for Infection Research (DZIF), partner site Charité, 10117 Berlin, Germany; ^hMartsinovskiy Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov University, 119991 Moscow, Russia; and ⁱCentre for Pathogen Evolution, Department of Zoology, University of Cambridge, CB2 3EJ Cambridge, United Kingdom

Edited by Francis V. Chisari, The Scripps Research Institute, La Jolla, CA, and approved June 3, 2020 (received for review April 14, 2020)

Hepatitis delta virus (HDV) is a human hepatitis-causing RNA virus, unrelated to any other taxonomic group of RNA viruses. Its occurrence as a satellite virus of hepatitis B virus (HBV) is a singular case in animal virology for which no consensus evolutionary explanation exists. Here we present a mammalian deltavirus that does not occur in humans, identified in the neotropical rodent species *Proechimys semispinosus*. The rodent deltavirus is highly distinct, showing a common ancestor with a recently described deltavirus in snakes. Reverse genetics based on a tandem minus-strand complementary DNA genome copy under the control of a cytomegalovirus (CMV) promoter confirms autonomous genome replication in transfected cells, with initiation of replication from the upstream genome copy. In contrast to HDV, a large delta antigen is not expressed and the farnesylation motif critical for HBV interaction is absent from a genome region that might correspond to a hypothetical rodent large delta antigen. Correspondingly, there is no evidence for coinfection with an HBV-related hepadnavirus based on virus detection and serology in any deltavirus-positive animal. No other coinfecting viruses were detected by RNA sequencing studies of 120 wild-caught animals that could serve as a potential helper virus. The presence of virus in blood and pronounced detection in reproductively active males suggest horizontal transmission linked to competitive behavior. Our study establishes a nonhuman, mammalian deltavirus that occurs as a horizontally transmitted infection, is potentially cleared by immune response, is not focused in the liver, and possibly does not require helper virus coinfection.

deltavirus | *Proechimys semispinosus* | hepadnavirus | coinfection | neotropical rodent

The genus *Deltavirus* is an unclassified RNA virus taxon that currently includes only one species, the hepatitis delta virus (HDV). This virus is an important human pathogen and the only satellite virus known in animals. For transmission, HDV must acquire an envelope from a coinfecting hepatitis B virus (HBV; family *Hepadnaviridae*, genus *Orthohepadnavirus*). Following superinfection of HBV-infected individuals, HDV increases the severity of liver disease (1). Initially detected in 1977 (2), HDV is present in about 10% of chronically HBV-infected patients (3). The circular single-stranded RNA genome is ~1,700 nt long and is packaged together with the small and large hepatitis delta antigens (S-HDAg and L-HDAg) into an HDV ribonucleoprotein particle that buds with the help of HBV envelope proteins from HDV/HBV-coinfecting hepatocytes. During replication, a double rolling-circle scheme produces multimeric antigenomic RNAs that are self-cleaved into monomers by a viroid-like ribozyme structure. The ligated circular RNA (circRNA) is further

transcribed into multimeric genomic RNAs which also undergo self-cleavage and ligation (4). Remarkably, HDV RNA replication is performed by cellular RNA polymerase II (Pol-II), with S-HDAg being essential in this process, as it coactivates Pol-II by possibly acting as a histone mimic (5). The RNA genome of HDV folds into an unbranched rod-like structure due to its high degree of self-complementarity (6). Budding and interaction with HBV envelope proteins require expression of L-HDAg, which is identical to S-HDAg except for a 19-amino acid extension at its C terminus, directed by cellular adenosine deaminase activity that edits the

Significance

Hepatitis delta virus (HDV) aggravates hepatitis B virus (HBV) infection of liver cells. Although the viruses are evolutionarily unrelated, HDV depends on HBV because it requires the HBV envelope protein for its transmission. HDV is only described in humans, which has triggered diverse hypotheses regarding its evolution and origins. Here we show that spiny rats (*Proechimys semispinosus*) carry a counterpart to HDV that surprisingly does not cause hepatitis and is not linked to HBV. The rodent deltavirus finding alone, but also taken together with the recent deltavirus findings in snakes and other vertebrates and invertebrates, suggests that a deltavirus precursor may have infected mammals before it acquired dependence on HBV as seen in humans.

Author contributions: S. Sommer and C.D. designed research; S.P., F.P., N.G., J.S., V.M.C., L.T.G., S. Schroeder, A.R., D.M., J.F.D., A.C.H., G.J.E., R.A.P., T.C.J., M.A.M., S. Sommer, D.G., and C.D. performed research; J.S., A.C.H., G.J.E., R.A.P., and S. Sommer designed and performed fieldwork; S.P., F.P., N.G., L.T.G., and S. Schroeder contributed new reagents/analytic tools; S.P., F.P., N.G., J.S., V.M.C., L.T.G., S. Schroeder, A.R., D.M., J.F.D., A.C.H., G.J.E., R.A.P., T.C.J., M.A.M., S. Sommer, D.G., and C.D. analyzed data; and S.P., D.G., and C.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: Sequences reported in this paper have been deposited in GenBank (accession numbers [MK598003-MK598012](https://www.ncbi.nlm.nih.gov/nuclink/MK598003-MK598012)).

¹S.P., F.P., and N.G. contributed equally to this work.

²Present address: Gentechnik, State Office for Health and Social Affairs (LAGeSo), 10559 Berlin, Germany.

³To whom correspondence may be addressed. Email: simone.sommer@uni-ulm.de, dieter.glebe@viro.med.uni-giessen.de, or christian.drosten@charite.de.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2006750117/-/DCSupplemental>.

amber codon on the antigenomic RNA (7). Posttranslational farnesylation of the C terminus of L-HDAg is essential for successful interaction of the HDV ribonucleoprotein particle with HBV envelope proteins (8). HDV has one serotype and is divided into eight distinct genotypes which show only a moderately structured geographic distribution (9, 10). The distribution of HBV and HDV genotypes is also only slightly correlated (11). HDV can be experimentally transmitted to woodchucks that carry replicating woodchuck hepatitis virus, but has not been observed in feral animals (12, 13). HDV remains the only human RNA virus for which there is no known replicating counterpart in other animals.

Recently, deltavirus-like sequences have been found in transcriptome data derived from pooled oropharyngeal and cloacal samples of teals and ducks (14), tissues of boas (*Boa constrictor*) and a water python (*Liasis mackloti*) (15), and various vertebrate and invertebrate transcriptomes (16). The snake- and duck-associated sequences, although highly distinct from HDV, show predicted similarity in secondary structure to the HDV ribozyme element. Amino acid motifs known to have functional importance for HDV replication occur in both sequences. However, the sequences described by Chang et al. (16) lack the two ribozyme structural elements and some of the functional motifs, as well as experimental proof of the viral replication mechanism. Also, the natural history of disease, if transmissible, is not known for these putative viruses.

Here we describe a unique mammalian deltavirus from the neotropical rodent species *Proechimys semispinosus*, tentatively named rodent deltavirus (RDeV). We initially discovered the viral RNA following undirected next-generation sequencing of blood samples. Our study of 763 animals provides epidemiological and virological evidence for a transmissible and prevalent infection with a delta-like agent in a specific host. Surprisingly, and in striking contrast to human HDV, we neither find a liver-specific tropism of RDeV nor epidemiological or functional evidence of coinfection with a hepadnavirus in the animals studied.

Results

In an ongoing study in the Panama Canal area, focused on the effects of habitat disturbance on the infection rate with rodent *Hepacivirus* (*SI Appendix, Fig. S1*) (17), we subjected blood samples of 120 individuals of *P. semispinosus* to total RNA sequencing (RNA-seq) using the Illumina MiSeq platform. High bit-score matches to human HDV were obtained after comparing de novo assembled sequences with a viral reference library (18). Based on mapping against a reference alignment of all known HDV genotypes, we compiled HDV-related sequences covering the whole deltavirus genome from two individual animals. After aligning these initial genome assemblies with HDV reference genotypes, we designed an RT-PCR assay targeting a fragment in the coding region of the rodent deltavirus antigen (RDeAg). The assay was applied to pooled blood samples from 763 *P. semispinosus* individuals. Resolution of positive-testing pools resulted in 30 positive individuals (overall detection rate 3.9%; 95% CI 2.6 to 5.3%). Only adult animals were found positive. In addition to *P. semispinosus*, we examined blood samples from individuals ($n = 183$) from 11 other rodent and marsupial species. No species other than *P. semispinosus* yielded evidence of RDeV presence (*SI Appendix, Table S1*).

Because HDV has a circular genome, we used primer pairs whose 5' ends are adjacent on reverse-complementary template strands, with 3' ends facing in opposite directions (Fig. 1B and *SI Appendix, Table S2*). These primers can generate RT-PCR amplicons covering the near-full deltavirus genome if the template RNA molecule is circular. This was successful in three samples with high viral load (Fig. 1C). Genome circularity was further confirmed by mapping RNA-seq reads from the respective samples to the initial amplicon sequences, resulting in assemblies with protruding ends that are identical, matching both ends of the amplicon, and therefore derived from circular templates. Ten full circular RDeV

genomes and 10 partial genome assemblies were subsequently obtained from the 30 initially RNA-positive samples by mapping reads from individual RNA-seq datasets. PCR products for the other 10 samples produced small fragments of less than 200 bases. The nucleotide sequences of all 10 full genomes are 97.5 to 99.6% identical. The sequences are available in GenBank under accession numbers MK598003 to MK598012.

Major structural and functional HDV domains are present in all recovered RDeV sequences (Fig. 1A and *SI Appendix, Fig. S2 and Table S10*), including a single open reading frame (ORF) representing the delta antigen. A 19-amino acid tail that differentiates the S-HDAg from the L-HDAg in human HDV is present in RDeV, though with a different amino acid composition. This tail is the result of antigenomic editing at a site corresponding to the amber stop codon in the S-HDAg ORF of human HDV (*SI Appendix, Fig. S3*). There is no experimental evidence of RNA editing or large-antigen expression in RDeV, as discussed below. In addition, the farnesylation motif CXXQ in the HDV L-HDAg C terminus, required to acquire the HBV-derived envelope (8), is absent in all *P. semispinosus* deltavirus sequences.

Predicted secondary structures of genomic and antigenomic ribozymes are highly similar between human and rodent deltavirus (*SI Appendix, Fig. S4*). The ribozyme active site is identical between rodent and human virus on both antigenomic and genomic strands. The nuclear export sequence motif and the nuclear localization signal of S-HDAg are also present in the S-RDeAg. Evidence for the existence and functionality of the nuclear localization signal is provided by the occurrence of the small delta antigen in the cell nuclei of HuH7 cells overexpressing S-RDeAg (Fig. 2A).

Deltavirus Phylogeny. Nucleotide and amino acid sequences were compared as summarized in *SI Appendix, Fig. S2 and Table S3*. Phylogenies were calculated based on full-genome nucleotide sequence alignments, as well as translated small delta antigen-coding sequences. Both tree topologies are equivalent and the distinction of RDeV from human HDV is clear (the full-genome phylogeny is shown in Fig. 3). The nonhuman deltaviruses are highly diverged and therefore chosen to root the phylogeny. The snake deltavirus (SDeV) clusters with the rodent deltavirus, and both form a highly diversified and well-supported sister group to human HDV, with HDV genotype 3 being the most divergent among HDV genotypes. All other nonhuman deltaviruses group together consistently, although without significant bootstrap support among subclades, owing to their high sequence divergence. RDeV is most closely related to human HDV, as averaged over all 1,000 bootstrap replicate trees (Fig. 3).

Organ Distribution of RDeV RNA and Protein Detection. To obtain insight into the infection pattern caused by the rodent deltavirus, we tested for viral RNA in organs of 18 animals that were found dead during sampling (note that we followed a noninvasive sampling approach and animals could not be purposefully euthanized during our study). One of these animals tested RDeV RNA-positive in all available organs (liver, kidney, lung, heart, and small intestine), but its organ-specific RNA concentrations did not suggest specific virus replication in the liver (*SI Appendix, Table S4*). Also, a commercial HDV antigen enzyme-linked immunosorbent assay that we found to cross-detect with RDeV did not identify organ-specific protein expression in the liver, lung, small intestine, heart, or kidney of this animal (*SI Appendix, Fig. S5A*). To further understand potential virus excretion, fecal samples from 822 individuals were tested by RT-PCR, 10 of which were positive for RDeV RNA (the numbers of individuals tested for each type of material are summarized in *SI Appendix, Table S5*). This 1.2% detection rate was significantly lower than that in blood samples (χ^2 test, $P < 10^{-4}$). The average virus

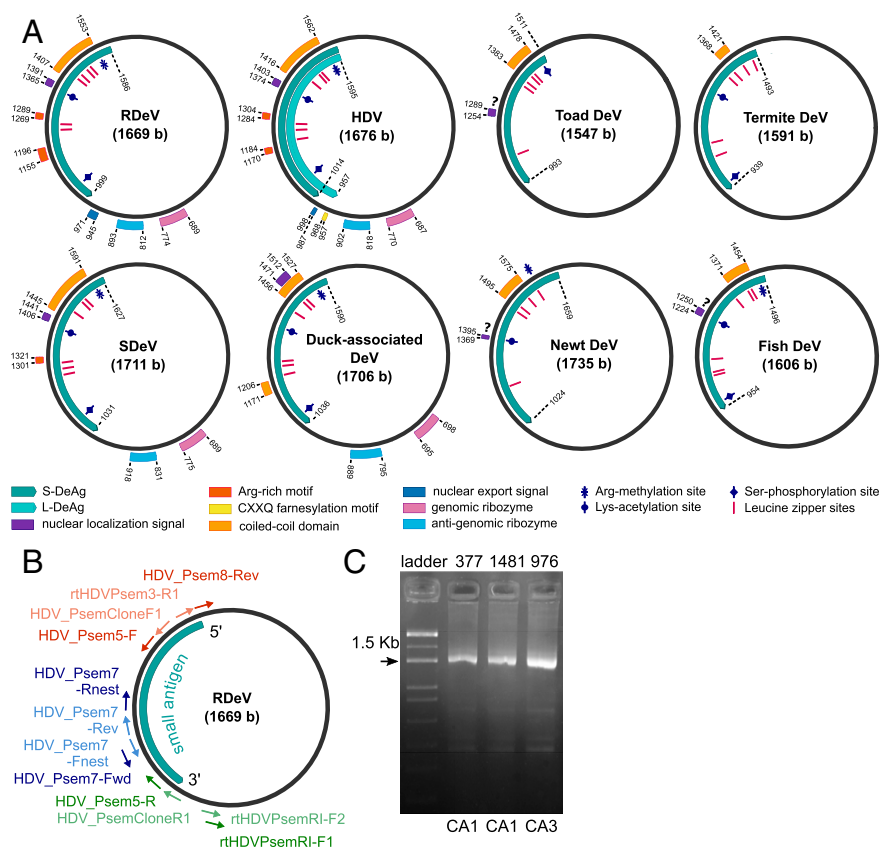


Fig. 1. (A) Genome organization of human HDV, RDeV, SDeV, duck-associated DeV, newt DeV, toad DeV, fish DeV, and termite DeV. Delta antigen-coding regions, functional and structural motifs, as well as post-translational modification sites. (B) Primer-binding positions for the circularization assays shown in their locations on the RDeV genome. Primer names and sequences are listed in *SI Appendix, Table S2*. Blue, circularization assay 1 (CA1); red, circularization assay 2 (CA2); green, circularization assay 3 (CA3). External primers are shown in dark color, and nested primers in light color. The S-RDeAg ORF is shown for orientation. (C) Exemplary results from CA1 and CA3 obtained from total RNA blood extracts of three RDeV RNA-positive *P. semispinosus*. Fragment sizes are 1,608 bp for CA1 and 1,532 bp for CA3; label correspondence to RDeV GenBank accession numbers: 377, MK598006; 1481, MK598003; 976, MK598009.

concentration in fecal samples was 2.9×10^9 RNA copies per gram, with a maximum concentration of 2.2×10^{10} copies per gram. Average and maximum concentrations in blood were 1.8×10^8 and 2.1×10^9 copies per milliliter.

RDeV Genome Replication. Rodent deltavirus genome replication was confirmed by Northern blot analysis, as well as by clonal expansion of RDeV-transfected cells. The full RDeV genome was cloned as a tandem head-to-tail fusion dimeric construct in genomic (minus-strand) orientation downstream of a cytomegalovirus (CMV) promoter (see the construct map in *SI Appendix, Fig. S6*) and transfected into HuH7 cells. Antigenomic RDeV RNA as a marker of ongoing RDeV replication was detected in transfected cells at all time points (Fig. 2F). The expression of antigenomic RDeV RNA was accompanied by a robust synthesis of genomic RDeV RNA (*SI Appendix, Fig. S5D*). Massive clonal expansion of RDeV-transfected cells verifies the autonomous replicating nature of RDeV, even exceeding HDV clonal expansion as determined by immunofluorescence staining (Fig. 2G).

Deltavirus Antigens and Editing of the Amber Codon. One key feature of HDV is the expression of two viral proteins (S-HDAg and L-HDAg) from a single ORF. A 19-amino acid carboxyl-terminal extension differentiates the S-HDAg (195 codons) from the L-HDAg (214 codons) in human HDV. This extension is the result of antigenomic editing by cellular adenosine deaminase acting on RNA 1 (ADAR1) at a site corresponding to the amber stop codon in the S-HDAg ORF of the circular antigenomic

HDV RNA, edited from a UAG to a UIG codon. Following HDV RNA replication, the UIG codon is converted into a tryptophan codon (UGG) (*SI Appendix, Fig. S3*) during subsequent HDV messenger RNA (mRNA) synthesis, allowing expression of the L-HDAg (20). Interestingly, a 19-amino acid carboxyl-terminal extension is present in RDeV, though different in 15 of the 19 amino acids. To investigate potential RDeV large antigen (L-RDeAg) expression, the full RDeV genome was cloned as a tandem head-to-tail fusion construct in genomic (minus-strand) orientation downstream of a CMV promoter (see the construct map in *SI Appendix, Fig. S6*) and transfected into HuH7 cells. As positive controls, S- or (hypothetical) L-RDeAg was expressed under the control of a CMV promoter. Specific S- or L-RDeAg expression was examined with antibodies from rabbits immunized with synthetic peptides against S-RDeAg (Fig. 2A) or the putative 19-amino acid extension of L-RDeAg (Fig. 2B). The tandem genome construct expressed S-RDeAg but not L-RDeAg (Fig. 2A and B).

Corresponding results were obtained by Western blot analysis of RDeV dimer transfection experiments using a broadly cross-reactive human anti-HDV serum (Fig. 2H, Upper). Evidence for the absence of RNA editing on the RDeAg ORF was also obtained from RNA-seq data. Mapping of reads from total RNA blood and organ extracts of infected animals and from transfected RDeV dimeric genome constructs in cell culture yielded only unedited reads, whereas both edited and unedited reads were detected for human HDV in a control experiment (*SI Appendix, Fig. S3*). In summary, there is no experimental evidence of

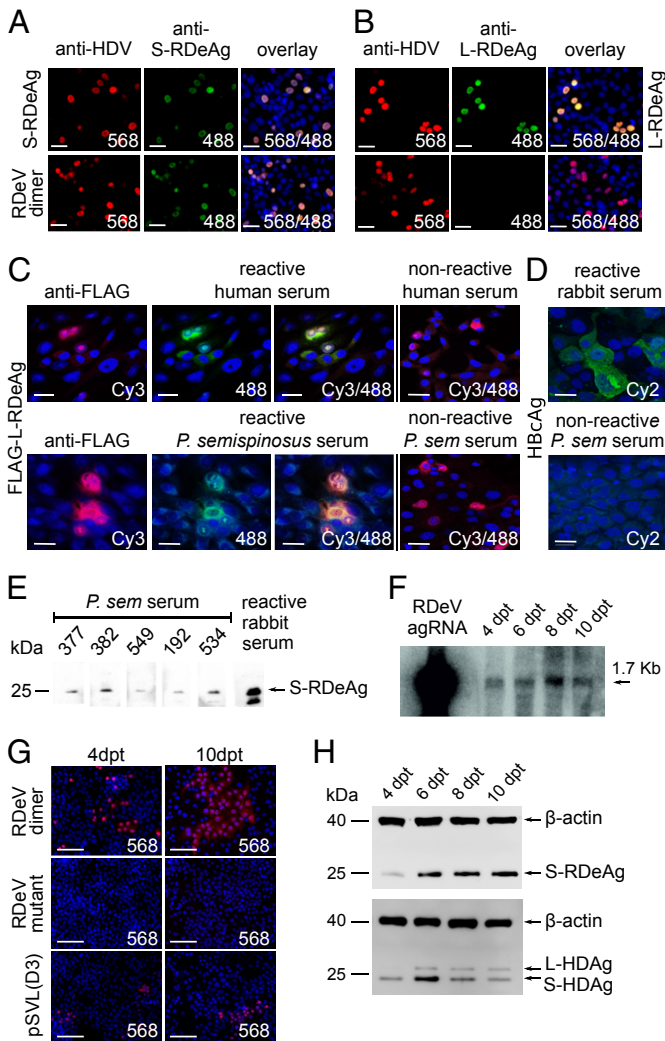


Fig. 2. (A and B) Immunofluorescence staining of Hu7 cells transfected with plasmids expressing S-RDeAg (A), a stop codon mutant expressing a hypothetical L-RDeAg (B), as well as an intact dimer of the RDeV genome (both panels). Cell nuclei were stained with DAPI; delta antigens were detected with anti-HDAG immunoglobulin G (IgG) from an HDV/HBV-coinfected patient followed by an Alexa 568-labeled goat anti-human antibody (red); S- or L-RDeAg was identified using S- or L-RDeAg-specific peptide antisera from immunized rabbits, followed by an Alexa 488-labeled goat anti-rabbit antibody (green). Channels are overlaid for both panels. (C) Immunofluorescence staining of L-RDeAg expression in Vero B4 cells with HDV-reactive human and *P. semispinosus* serum samples. Cell nuclei were stained with DAPI (blue); FLAG-tagged L-RDeAg was detected using a mouse anti-FLAG antibody followed by a goat anti-mouse Cy3-labeled antibody (red); reactivity of human and *P. semispinosus* serum against L-RDeAg is visualized by an Alexa 488-labeled goat anti-human and goat anti-guinea pig antibody, respectively (green). Nonreactive human and *P. semispinosus* serum samples are shown for comparison. (D) Immunofluorescence staining of HBcAg expression in Hu7 cells with anti-HBc-specific rabbit antiserum (Upper) and *P. semispinosus* serum (Lower) samples. Cell nuclei are stained with DAPI (blue); expression of HBcAg is visualized using a Cy2-labeled secondary antibody (green). Channels are merged for both panels. (Scale bars, 30 μm .) (E) Specific reactivity of *P. semispinosus* sera against S-RDeAg in a Western blot. Total protein extracts from Hu7 cells overexpressing S-RDeAg were separated by sodium dodecyl sulfate/polyacrylamide gel electrophoresis and blotted. The membranes were incubated with five *P. semispinosus* serum samples (IDs shown in figure) followed by a goat anti-guinea pig horseradish peroxidase (HRP) antibody. Anti-S-RDeAg-specific peptide antiserum from an immunized rabbit was used as positive control, with goat anti-rabbit HRP as secondary antibody. (F) Northern blot of anti-genomic RDeV RNA from transfected Hu7 cells. The cells were transfected with an expression plasmid containing an intact dimer of the RDeV genome.

RNA editing on the RDeV antigen ORF or of large-antigen expression during RDeV replication in vivo or in vitro.

Search for Signs of Hepadnaviral Coinfection. Because human HDV transmission requires an envelope protein provided by an active coinfection with HBV, we subjected all 763 *P. semispinosus* blood samples to PCR screening for hepadnaviruses. The assay was designed to detect all mammalian orthohepadnaviruses, including those from rodents, bats, and primates (21). None of the samples tested positive. Specific reanalysis of all available transcriptome datasets from the present study did not yield any evidence for orthohepadnaviral genomes in a basic local alignment search tool (BLAST) analysis. All blood samples with sufficient volume, whether RNA-positive or -negative for RDeV ($n = 68$), were tested for antibodies against woodchuck HBV core antigen (WHcAg). This antigen was chosen because the woodchuck is most closely related to *Proechimys* among known HBV hosts. In previous studies, we demonstrated that anti-HBc antibodies broadly cross-react among HBVs from different hosts, even across mammalian orders (21), so it is likely that WHcAg would be bound by antibodies against a possible HBV in *Proechimys*. However, no such antibodies were found in any of the tested samples (Fig. 2D).

Investigation of Hepacivirus as Potential Cofactor. Experimental evidence shows that apart from HBV, other viruses, including hepatitis C virus (genus *Hepacivirus*), can provide envelope proteins for human HDV (22). As we previously detected a high prevalence of hepacivirus in the rodents studied (17), this possibility was investigated. Overall, we found 4 out of 30 RDeV RNA-positive rodents in which hepacivirus was not detected by the tests initially applied (SI Appendix, Table S6). To exclude having missed hepacivirus detection, we applied RT-PCR assays specifically designed for the E1, NS3, and NS5B genes of *P. semispinosus* hepacivirus to those samples. All four animals were confirmed to be hepacivirus RNA-negative by these assays as well as by RNA-seq read mapping against *P. semispinosus* hepacivirus. Statistical analysis of the degree of dependency between deltavirus and hepacivirus infection in individuals was conducted, providing no support for a dependency between the two infections (SI Appendix, Table S6; χ^2 test, $P = 0.24$). Also, analyses of sampling site-specific deltavirus and hepacivirus detection rates (SI Appendix, Table S7) did not reveal any correlation between deltavirus and hepacivirus (Pearson's correlation coefficient, $\rho = 0.15$), arguing against a linear correlation. Logarithmic and linear regression fits did not yield significant associations either, suggesting that the two viruses follow different patterns of distribution. We conclude that

Total RNA was isolated 4, 6, 8, and 10 d posttransfection (dpt) and subjected to a 1% denaturing formaldehyde agarose gel. After blotting, the RNA was hybridized with a single-stranded oligonucleotide probe labeled with [γ - ^{32}P] ATP and T4 polynucleotide kinase at the 5' end. In vitro-transcribed antigenomic RDeV RNA was used as positive control. (G) Immunofluorescence staining of Hu7 cells transfected with expression plasmids containing an intact dimer of the RDeV genome, a mutant dimer of the RDeV genome with abrogated S-RDeAg expression, and a trimeric genome of human HDV genotype 1 [pSVL(D3)]. Cells (Left) were fixed on day 4 posttransfection. A 1:64 dilution of the cells (Right) were subjected to clonal expansion and fixed on day 10 posttransfection. Cell nuclei were stained with DAPI (blue) and delta antigens were detected with anti-HDV IgGs followed by an Alexa 568-labeled goat anti-human antibody (red). (Scale bars, 100 μm .) (H) Detection of deltavirus antigens from HDV- and RDeV-transfected Hu7 cells by Western blot. Hu7 cells were transfected with a dimeric RDeV genome and a trimeric genome of human HDV genotype 1 [pSVL(D3)], and total protein was isolated on 4, 6, 8, and 10 d posttransfection. S-RDeAg (Upper) and S- and L-HDAG (Lower) were detected using cross-reactive human anti-HDV serum; beta-actin was detected by an anti-beta-actin antibody.

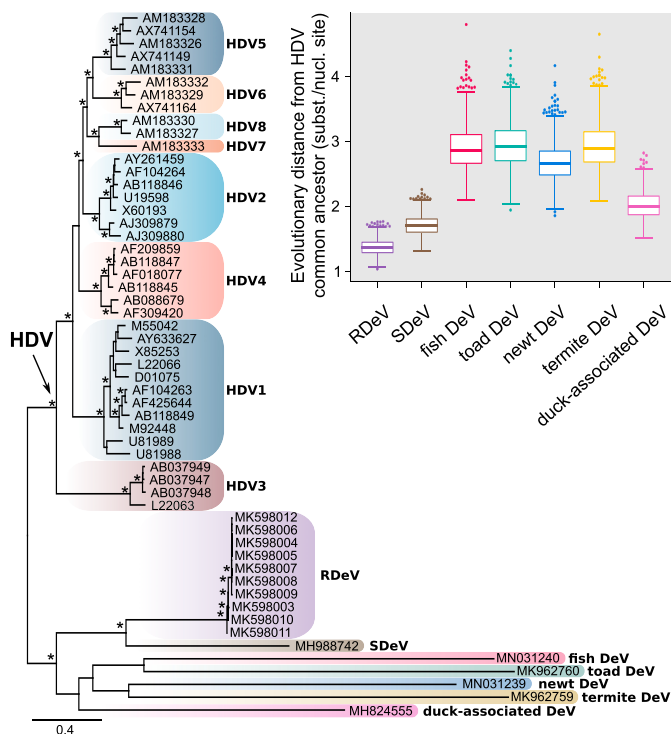


Fig. 3. Maximum-likelihood phylogeny based on a full-genome alignment of HDV representatives from the eight different genotype groups and RDeV, SDeV, duck-associated DeV, newt DeV, toad DeV, fish DeV, and termite DeV. The phylogenetic inference was done using RAXML-NG version 0.7.0 BETA (19). Asterisks indicate bootstrap support >80 and tree tips show GenBank accession numbers. The tree is rooted to the branch leading to the nonhuman DeVs. Distribution of evolutionary distances (branch lengths, extracted from 1,000 bootstraps) for all nonhuman DeVs relative to the common ancestor of HDV (arrow on tree) is shown in the boxplots. The boxplots show median, interquartile range, Tukey's minimum and maximum (whiskers), and extreme outlier values (dotted).

RDeV infection in *P. semispinosus* does not depend on active hepatitis infection.

Immune Reaction in RDeV-Infected Animals. To determine whether RDeV causes an immune reaction in *P. semispinosus*, serum samples were subjected to an indirect immunofluorescence assay. L-RDeAg was cloned and transiently expressed in Vero B4 cells, as described (23). Fig. 2C shows the reactivity of antibodies from human sera against L-RDeAg, along with examples of reactive and nonreactive *P. semispinosus* sera. At a dilution of 1:100, 17 of 30 RDeV RNA-positive and 3 of 115 RDeV RNA-negative animals tested positive for anti-RDeAg antibodies ($P < 10^{-6}$, χ^2 test). To exclude RNA degradation in the case of antibody-positive samples for which RDeV RNA was not detected, we applied a real-time RT-PCR assay to detect RNA transcripts of a host housekeeping gene (TATA-binding protein) (SI Appendix, Table S8). The 20 positive sera were additionally tested at 1:1,000 dilution, testing positive in 5 cases. Western blots of these sera against expressed S-RDeAg confirmed the specificity of serum antibodies (Fig. 2E). The average RDeV RNA concentration in all anti-RDeAg-positive sera was 262,400 RNA copies per microliter of blood. The average RDeV RNA concentration in only those samples that tested antibody-positive up to a dilution of 1:1,000 was 8,500 RNA copies per microliter of blood (SI Appendix, Fig. S5C). These concentrations were significantly different ($P < 0.05$, Mann-Whitney U test), which suggests that an adaptive immune response can potentially limit or eliminate viral replication. Elimination is also suggested by the occurrence of

three anti-RDeAg antibody-positive but RDeV RNA-negative animals.

Ecological Factors That Influence Rodent Deltavirus Infection. The study area in the Barro Colorado Nature Monument consists of islands and peninsulas with different densities of populations of *P. semispinosus*, a generalist species that can adapt to environmental changes (24) and thus (relatively) competitively benefit from anthropogenic habitat disturbance. We have previously noted a positive correlation between site-specific *P. semispinosus* population density and the detection rate of hepatitis delta virus (SI Appendix, Table S7) (17). For rodent deltavirus, a different site-specific distribution was observed (SI Appendix, Table S7). Highest detection rates were identified in continuous forest sites that have largely preserved the primordial habitat composition that existed before the flooding of the Gatún Lake area during the construction of the Panama Canal. Lower detection rates were found on forested islands, and minimal rates in forest fragments surrounded by agriculture. To identify environmental factors that might correlate with rodent deltavirus transmission, host- and habitat-specific properties were examined under two different logistic regression models (SI Appendix, Table S9). RDeV infection was found to be positively correlated with male sex and reproductive activity, as well as with inhabiting continuous forest sites with undisturbed habitat. In addition to the variables shown in SI Appendix, Table S9, other habitat-specific factors such as population density, canopy height and coverage, and understory density did not add explanatory power to the analysis. The inclusion of site-specific data on mosquito diversity narrowed the sample size to $n = 686$ individuals for sampling sites where these data were available. There was no detectable correlation between mosquito diversity and RDeV detection rate.

Discussion

The present results enhance our understanding of the evolution of hepatitis delta virus, a satellite virus to HBV and until recently the only human RNA virus with no known counterpart in other animals. The first question that must follow the discovery of novel deltavirus-like sequences in rodent, reptile (15), and putatively avian (14) and other vertebrate and invertebrate hosts (16) is whether these represent transmissible viruses as opposed to, for example, endogenous viral elements. Hetzel et al. (15) demonstrated protein expression in organs of those snakes in which viral sequences were detected, and Szivovics et al. (25) showed that snake deltavirus can replicate in several cell lines. Our reverse genetic and clonal expansion experiments now prove the existence of mammalian nonhuman deltavirus genomes as self-replicating viral entities, in which genomic replication involves an internal initiation of replication. Moreover, correlations of serology and virus detection in a natural animal population of considerable size demonstrate that rodent deltavirus infection is acquired during the course of life and is likely to be cleared by an active immune response. Previous observations on the snake deltavirus corroborate this infection pattern, showing that although snake deltavirus infection was found in both maternal and offspring boas (15), not all offspring were infected, and snakes with no kinship to the mating pairs exhibited both the presence and absence of the virus. Horizontal transmission seems likely in both rodent and reptile deltaviruses. The significantly lower detection rate of viral RNA in stool samples compared with blood suggest blood-borne rather than fecal-oral transmission. The observed predominant infection of adult males is compatible with transmission through competitive behavior.

The novel deltaviruses associated with diverse hosts warrant considerations regarding evolutionary origins. Phylogeny suggests that fish- and amphibian-associated viruses are most distant from other known deltaviruses, while snake and rodent viruses are significantly closer to the HDV common ancestor. Considering the relationship of these hosts, the tree topology seems to disfavor

virus-host codivergence. Besides, the duck-associated virus was not detected in an internal organ but in cloacal and oropharyngeal swabs, leaving open the possibility of contamination from alimentary sources (14). The phylogenetic branch pattern supports previous work on human HDV phylogeny, suggesting that HDV genotype 3 stems from the oldest common ancestor of human HDV genotypes (26). As this ancestor is more closely related to the rodent virus than to any of the nonmammalian deltaviruses, this would be compatible with the hypothesis that human and rodent deltaviruses have codiverged with mammalian lineages. Deltaviruses in other mammals may remain to be discovered, but equally may now be extinct. A hypothesis of codivergence would be corroborated by the discovery of at least one more mammalian virus whose phylogenetic position reflects that of its host in relation to humans and rodents. The clustering of rodent and snake viruses might then be interpreted as a sign of cross-species acquisition from mammals. It is noteworthy that the parental boas carrying the snake deltavirus stemmed from Panama, like the rodents studied here (15). Human HDV genotype 3 is also found in South and Middle America, but this alone should not be taken as support to claim cross-host transmission.

The dependence of HDV on HBV coinfection for transmission may have evolved only in a mammalian viral lineage leading to human HDV, given that hepadnaviruses occur in diverse mammalian species including rodents, bats, and particularly New World primates (21, 27, 28). Based on the lack of large-antigen expression and a farnesylation motif, both necessary for interaction with hepadnaviral envelope proteins, as well as on evidence from epidemiological and serological data, we can exclude the presence of HBV coinfection in deltavirus-infected *P. semispinosus*. Neither HBV DNA nor anti-HBc antibodies were detected in any animals infected with the deltavirus. In all other deltavirus studies, no sequences of coinfecting hepadnaviruses were found in transcriptome data. Other coinfecting viruses were identified in each of these studies, designating their envelope proteins as a potential source for an RDeV envelope (14–16). As very recently shown by Szivovics et al. (25), snake deltavirus can produce infectious viral particles *in vitro* by utilization of envelope proteins from arenaviruses and orthohantaviruses.

In the present study, we investigated the potential influence of hepacivirus infection. Despite a high hepacivirus prevalence in the studied rodents [ca. 80% (17)], we found four deltavirus RNA-positive individuals lacking a hepacivirus coinfection. Human HDV has been found to produce infectious particles *in vitro* with the help of an envelope protein from hepatitis C virus, with the required presence of an L-HDAg farnesylation motif (22). Given the absence of this motif in the rodent delta antigen and the lack of a hepacivirus coinfection in some individuals carrying deltavirus RNA, we do not consider deltavirus infection to depend on hepacivirus infection in these rodents. In addition, we screened individual RNA-seq datasets from 120 animals and did not identify any other infectious viral agent in any deltavirus RNA-positive *P. semispinosus*. There are multiple alternatives that may have resulted in rodent deltavirus transmission. One possibility is the presence of a currently unknown virus, bacterium, or integrated sequence that provides an envelope protein to rodent deltavirus. Furthermore, recent advances in cellular cross-communication have revealed the exchange of extracellular vesicles containing mRNA, circRNA, and other noncoding RNA molecules (29). It should be investigated whether rodent deltavirus might employ such pathways for transmission.

In their study of snake organ samples, Hetzel et al. (15) found two bands of ca. 20 and 27 kDa in Western blots using an anti-S-SDeAg serum. While the larger band might correspond to a larger protein variant, expression of an actual L-SDeAg was not confirmed by specific anti-L-SDeAg immunoblots, analyses of RNA editing, or experimental studies of gene expression. The present data, based on all these approaches, show that the

related rodent virus does not express a large delta antigen, suggesting that the use of L-HDAg may have evolved only in connection with the utilization of the HBV envelope. In this scenario, the encounter with hepadnaviruses may have selected for viral variants with a farnesylation signal in an L-HDAg tail that is expressed by an initially low-level editing activity or stop codon readthrough, allowing these viruses to exploit a novel transmission opportunity by acquisition of an HBV envelope. The sequence information from the present study can now be used in directed searches for deltaviruses in other mammalian species, even in the absence of HBV infection.

Alternative hypotheses of HDV origin have suggested either a viroid origin (perhaps via an insect vector) or derivation from a host DIPA gene (reviewed in ref. 11). A viroid origin is supported by the structural similarity of the HDV ribozyme to plant viroid ribozymes, the genome circularity of both HDV and viroids, and the interaction with common cellular proteins (30). Derivation from the human genome has been suggested based on the incomplete set of genes essential for replication, the occurrence of HDV-like ribozyme elements in the human genome (31), and the interaction of HDV with a human-derived protein (DIPA) (32). The deltavirus findings do not increase or decrease support for the viroid origin hypothesis but force us to consider the DIPA hypothesis from a wider perspective. The rodent deltavirus ORF shows 30% nucleotide and 22% amino acid identity when compared with the DIPA ORF in mice and rats, at which levels similarity due to chance must be considered as a likely explanation (33). In addition to these caveats, the present results suggest that other mammalian deltaviruses existed before the origin of human HDV, and that precursors to these mammalian viruses existed in other vertebrates (14–16). Alternate scenarios, specifically the derivation of nonhuman deltaviruses from human HDV precursors, seem unlikely, considering the genetic distances between these viruses, the relatively short time span available for the evolution of such diversity in the case of a human origin, and the requirement that the human virus infect several nonhuman hosts. It appears instead that deltavirus originated as an exogenous infectious agent that established acute infection in mammals. The present study provides a basis for the evolution of HDV, confirming the general understanding that RNA virus diversity evolved along the major evolutionary lineages of their hosts (34).

Methods

Animals were captured using Sherman traps in the Barro Colorado Nature Monument, Panama (17). RNA-seq experiments were done on an Illumina MiSeq instrument. RT-PCR followed standard protocols, including for circular genome amplification. Immunofluorescence assay (IFA) serology was based on cells overexpressing the L-RDeAg or woodchuck WHc antigen under the control of a CMV promoter. Viral full RDeV genome complementary DNA was cloned as a monomer or head-to-tail dimer under the control of a CMV promoter. A full description of methods is provided in *SI Appendix, Materials and Methods*.

All fieldwork was carried out with full ethical approval (Smithsonian Institutional Animal Care and Use Committee protocols 2013-0401-2016-A1-A7 and 2016-0627-2019-A2) and samples were exported to Germany with permission from the Panamanian government (SEX/A-21-14, SE/A-69-14, SEX/A-22-15, SEX/A-24-17, SEX/A-120-16, and SEX/A-52-17).

Data Availability. The rodent deltavirus complete genome sequences described here are available in GenBank under accession numbers MK598003 to MK598012.

ACKNOWLEDGMENTS. We thank Stefan Brändel for arranging research permits and Tobias Bleicker for technical assistance. This work was supported by grants from the German Research Foundation (DFG), B08/SFB 1021/2 (to D.G.), SPP 1596 Grants SO 428/9-1 and 2 (to S. Sommer), and DR772/8-22 (to C.D.). The National Reference Centre for Hepatitis B Viruses and Hepatitis D Viruses is supported by the German Ministry of Health via the Robert Koch Institute. Additional support for sequencing work was provided by the European Commission's COMPARE project.

1. M. Rizzetto *et al.*, Chronic hepatitis in carriers of hepatitis B surface antigen, with intrahepatic expression of the delta antigen. An active and progressive disease unresponsive to immunosuppressive treatment. *Ann. Intern. Med.* **98**, 437–441 (1983).
2. M. Rizzetto *et al.*, Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut* **18**, 997–1003 (1977).
3. H.-Y. Chen *et al.*, Prevalence and burden of hepatitis D virus infection in the global population: A systematic review and meta-analysis. *Gut* **68**, 512–521 (2019).
4. A. D. Branch, H. D. Robertson, A replication cycle for viroids and other small infectious RNA's. *Science* **223**, 450–455 (1984).
5. N. Abeywickrama-Samarakoon *et al.*, Hepatitis delta virus histone mimicry drives the recruitment of chromatin remodelers for viral RNA replication. *Nat. Commun.* **11**, 419 (2020).
6. K.-S. Wang *et al.*, Structure, sequence and expression of the hepatitis delta (δ) viral genome. *Nature* **323**, 508–514 (1986).
7. S. K. Wong, D. W. Lazinski, Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15118–15123 (2002).
8. J. C. Otto, P. J. Casey, The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells. *J. Biol. Chem.* **271**, 4569–4572 (1996).
9. K. Jackson *et al.*, Epidemiology and phylogenetic analysis of hepatitis D virus infection in Australia. *Intern. Med. J.* **48**, 1308–1317 (2018).
10. F. Le Gal *et al.*, Eighth Major Clade for Hepatitis Delta Virus. *Emerging Infectious Diseases* **12**, 1447–1450 (2006).
11. J. Taylor, M. Pelchat, Origin of hepatitis δ virus. *Future Microbiol.* **5**, 393–402 (2010).
12. F. Negro *et al.*, Hepatitis delta virus (HDV) and woodchuck hepatitis virus (WHV) nucleic acids in tissues of HDV-infected chronic WHV carrier woodchucks. *J. Virol.* **63**, 1612–1618 (1989).
13. N. Freitas *et al.*, Hepatitis delta virus infects the cells of hepadnavirus-induced hepatocellular carcinoma in woodchucks. *Hepatology* **56**, 76–85 (2012).
14. M. Wille *et al.*, A divergent hepatitis D-like agent in birds. *Viruses* **10**, 720 (2018).
15. U. Hetzel *et al.*, Identification of a novel deltavirus in boa constrictors. *MBio* **10**, e00014-19 (2019).
16. W.-S. Chang *et al.*, Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol.* **5**, vez021 (2019).
17. J. Schmid *et al.*, Ecological drivers of *Hepacivirus* infection in a neotropical rodent inhabiting landscapes with various degrees of human environmental change. *Oecologia* **188**, 289–302 (2018).
18. N. Goodacre, A. Aljanahi, S. Nandakumar, M. Mikailov, A. S. Khan, A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *MSphere* **3**, e00069-18 (2018).
19. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
20. G. X. Luo *et al.*, A specific base transition occurs on replicating hepatitis delta virus RNA. *J. Virol.* **64**, 1021–1027 (1990).
21. J. F. Drexler *et al.*, Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16151–16156 (2013).
22. J. Perez-Vargas *et al.*, Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat. Commun.* **10**, 2098 (2019).
23. M. A. Müller *et al.*, Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: A nationwide, cross-sectional, serological study. *Lancet Infect. Dis.* **15**, 629 (2015).
24. G. H. Adler, The island syndrome in isolated populations of a tropical forest rodent. *Oecologia* **108**, 694–700 (1996).
25. L. Szivovicza *et al.*, Snake deltavirus utilizes envelope proteins of different viruses to generate infectious particles. *MBio* **11**, e03250-19 (2020).
26. F. Le Gal *et al.*, Eighth major clade for hepatitis delta virus. *Emerg. Infect. Dis.* **12**, 1447–1450 (2006).
27. B. F. de Carvalho Dominguez Souza *et al.*, A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. *J. Hepatol.* **68**, 1114–1122 (2018).
28. A. Rasche, A.-L. Sander, V. M. Corman, J. F. Drexler, Evolutionary biology of human hepatitis viruses. *J. Hepatol.* **70**, 501–520 (2019).
29. K. M. Kim, K. Abdelmohsen, M. Mustapic, D. Kapogiannis, M. Gorospe, RNA in extracellular vesicles. *Wiley Interdiscip. Rev. RNA* **8**, e1413 (2017).
30. R. Flores, S. Ruiz-Ruiz, P. Serra, Viroids and hepatitis delta virus. *Semin. Liver Dis.* **32**, 201–210 (2012).
31. C.-H. T. Webb, A. Lupták, HDV-like self-cleaving ribozymes. *RNA Biol.* **8**, 719–727 (2011).
32. R. Brazas, D. Ganem, A cellular homolog of hepatitis delta antigen: Implications for viral replication and evolution. *Science* **274**, 90–94 (1996).
33. M. Long, S. J. de Souza, W. Gilbert, Delta-interacting protein A and the origin of hepatitis delta antigen. *Science* **276**, 824–825 (1997).
34. Y.-Z. Zhang, W.-C. Wu, M. Shi, E. C. Holmes, The diversity, evolution and origins of vertebrate RNA viruses. *Curr. Opin. Virol.* **31**, 9–16 (2018).

4 Viromics of extant insect orders unveil the evolution of the flavi-like superfamily

This manuscript is a collaborative work with the Zoological Research Museum Alexander Koenig and the group of Prof. Dr. Bernhard Misof, which dates back to the time when Prof. Dr. Christian Drosten was the director of the Institute of Virology in Universitätsklinikum Bonn. A preliminary search for flaviviruses and tombusviruses was held during my MSc study at Rheinische Friedrich-Wilhelms-Universität in Bonn, 2015, and is delineated in my MSc thesis. The data of that search was taken, revisited, and newly analyzed for this work.

Submitted to the journal *Virus Evolution* as: Paraskevopoulou S., Käfer S., Zirkel F., Donath, A., Liu, S., Zhou, X., Drosten C., Misof B., and Junglen S. Viromics of extant insect orders unveil the evolution of the flavi-like superfamily. (unpublished).

Viromics of extant insect orders unveil the evolution of the flavi-like superfamily

Paraskevopoulou Sofia^a, Käfer Simon^{a,*}, Zirkel Florian^{b,§}, Donath Alexander^c, Petersen Malte^{c,+}, Liu Shanlin^d, Zhou Xin^d, Drosten Christian^{a,e}, Misof Bernhard^{c,#}, Junglen Sandra^{a,e,#}

a. Institute of Virology, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

b. Institute of Virology, University of Bonn Medical Center, Bonn, Germany

c. Centre for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Bonn, Germany

d. Department of Entomology, China Agricultural University, Beijing, China

e. German Center for Infection Research (DZIF), partner site Charité, Berlin, Germany

#Corresponding authors: Junglen Sandra, sandra.junglen@charite.de, Misof Bernhard, B.Misof@leibniz-zfmk.de

*Present address: Hannover Medical School, Hannover, Germany

§Present address: Biotest AG, Dreieich, Germany

+Present address: Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany

Keywords: flavi-like virus, insect virus, phylogeny, positive-sense RNA, tombusvirus.

Abstract

Insects are the most diversified and species-rich group of animals and harbor an immense diversity of viruses. Several taxa in the flavi-like superfamily, such as the genus *Flavivirus*, are associated with insects; however, systematic studies on insect virus genetic diversity are lacking, limiting our understanding of the evolution of the flavi-like superfamily. Here, we examined the diversity of flavi-like viruses within the most complete and up-to-date insect transcriptome collection comprising 1,243 insect species by employing a *Flaviviridae* RdRp profile hidden Markov model search. We identified 76 viral sequences in 63 insect species belonging to 18 insect orders. Phylogenetic analyses revealed that 27 sequences fell within the *Flaviviridae* phylogeny but did not group with established genera. Despite the large diversity of insect hosts studied, we only detected one virus in a blood-feeding insect which branched within the genus *Flavivirus*, indicating that this genus diversified only in hematophagous arthropods. Nine new jingmenviruses with novel host associations were identified. One of the jingmenviruses established a deep rooting lineage in addition to the insect- and tick-associated clades. Segment co-segregation phylogenies support the separation of tick- and insect-associated groups within jingmenviruses, with evidence for segment reassortment. In addition, 14 viruses grouped with unclassified flaviviruses encompassing genome length of up to 20 kb. Species-specific clades for

Hymenopteran- and Orthopteran-associated viruses were identified. Forty nine viruses populated three highly diversified clades in distant relationship to *Tombusviridae*, a family of plant-infecting viruses, suggesting the detection of three previously unknown insect-associated families that contributed to the evolution of tombusviruses.

Introduction

During the last decade, the scientific interest in arthropod-borne viruses (arboviruses) has broadened to include arthropod- and, in particular, insect-specific viruses (1). Whereas the arthropod vector in the arboviral transmission cycle was originally noted for its function to transmit the virus to vertebrates, it is becoming more apparent that the evolutionary origins of arboviruses may lie in arthropods. Arthropod-specific viruses could thus be regarded as precursors (but not ancestors) of arboviruses (2). For instance, within bunya- and rhabdoviruses, known animal pathogenic viruses are embedded in a diversity of novel arthropod-specific viruses discovered only in the past years (3, 4). The great majority in the phylum Arthropoda are insects. Insects are the most abundant and diversified animal group with an estimated number of ~5.5 million species representing about 80 percent of the world's species (5). Metagenomic studies revealed an enormous virus diversity in insects (3, 4, 6, 7), yet, only a fraction of the tremendous diversity of insect species has been analyzed.

The genus *Flavivirus* (*sensu stricto*, relates to ICTV-classified species) within the *Flaviviridae* family hosts an extensive list of human-pathogenic arboviruses, transmitted by mosquitoes and ticks. On the other hand, all other established *Flaviviridae* genera (*Pegivirus*, *Pestivirus*, and *Hepacivirus*) contain only vertebrate-infecting viruses that are not associated with arthropod vectors and show a lower intra-genetic diversity in comparison to the genus *Flavivirus*. New hepaci- and pegi-like viruses within virus discovery studies have only been reported to occur in primates and mammals (8). Recent findings of viruses that have different genome organizations from classical flaviviruses but group within the family *Flaviviridae* in phylogenetic analyses based on the RNA-dependent RNA polymerase (RdRp) gene, indicate that the evolution of this family is complex and not well understood. One of these groups contains viruses that have been found in bees, flies, and aphids, as well as in plant and nematode hosts (6, 9–15). This group contains viruses with genomic length of 16–23 kb as opposed to 10 kb in classical flaviviruses, albeit with a similar organization. Another unclassified virus group contains segmented viruses, tentatively named “jingmenviruses”, which were recently discovered in ticks and several insect species, as well as in humans with reported febrile illness (16, 17). This data suggest that non-hematophagous insect hosts played an important role in flavivirus evolution. However, systematic studies in insects have been neglected. The *Flaviviridae* and *Tombusviridae* families are grouped in the so-called flavi-like superfamily (18), however, the evolution of host associations within and between both families remains elusive. Tombusviruses are known to infect angiosperm plants in more than 15 different orders, mainly causing leaf mottling and deformations, and stunting. Virus spread occurs by contact between plants, pollen, or seeds, or by contact to infected soil or water (19, 20). There is little knowledge if virions can also be

transmitted by arthropods perhaps acting as carriers for seeds or pollen.

Current knowledge on flavi-related viruses is mainly driven by interests in human or domestic animal health, and strategies to combat their associated repercussions. The same pattern occurs similarly in the tombus-related viruses with research lines mostly drawn across mitigation strategies for plant viral disease. Despite the need for a broad understanding of flavivirus evolution (21) and RNA virus evolution in general (22), only few studies have previously focused on a systematic and comprehensive search in non-typical tombusvirus hosts, i.e. plant hosts, or non-hematophagous flavivirus hosts, other than mosquitoes or ticks (7, 9–14, 23). Crucially too, understanding the evolution of the flavi-like superfamily requires a unified sampling strategy of large organismic host groups in higher taxonomic ranks, such as orders or classes, and in a variety of geographic locations. However, sampling is often performed on a basis of limited geographical sites or organismic groups for reasons of capacity. Also, the practice of sample pooling, though assisting in the era of next-generation sequencing, introduces doubts in ascertaining species-specific host associations as well as in genome assembly (especially for segmented viruses) when individual samples are not retained. Here, we explored the diversity of flavi-related viruses in the largest insect collection of transcriptomes sampled worldwide representing all extant insect orders (24). Our findings unveil the evolution of the flavi-like superfamily within insects, contributing new coding-complete viral genomes and previously unknown insect host associations, even at the order level.

Results

Our search, based on an amino acid profile hidden Markov model (pHMM) of the *Flaviviridae* RdRp gene, yielded a broad diversity of previously unknown viruses. In total, 162 putative viral sequences showing 18–61% amino acid identity to flaviviruses and 22–68% amino acid identity to tombusviruses were identified. The pHMM search, while based on a *Flaviviridae* RdRp model, delivered hits for both *Flaviviridae* and *Tombusviridae*, confirming the grouping of the two families in the flavi-like superfamily and their phylogenetic clustering within the *Kitrinoviricota* phylum (25). An overview unrooted phylogeny is shown in **Fig. 1** that includes *Amarillovirales* and *Tolivirales*, the two virus orders of *Flaviviridae* and *Tombusviridae*, respectively. Extensive filtering and verification of the pHMM hits based on sequence length (>300 nucleotides), taxonomic grouping, and presence of a continuous open reading frame containing the conserved canonical glycine–aspartate–aspartate (GDD) motif resulted in 76 unique viral sequences (n = 27 flavi-related and n = 49 tombus-related) that were included in downstream phylogenetic and genome organization analyses. These viral sequences were identified in 63 insect species belonging to 18 orders, such as Trichoptera, Embioptera, and Odonata, with nearly each detected virus associated with a different insect species. Details of host associations, taxonomic classification, and sampling location and date are given in **Table S1**. Novel viruses were named after the host order, viral family, and the designation “OKIAV” (for “1KITE insect-associated virus”), followed by the sample ID, e.g. Hymenopteran flavi-related virus OKIAV350, in conformity with viruses previously identified in the same sample set (3).

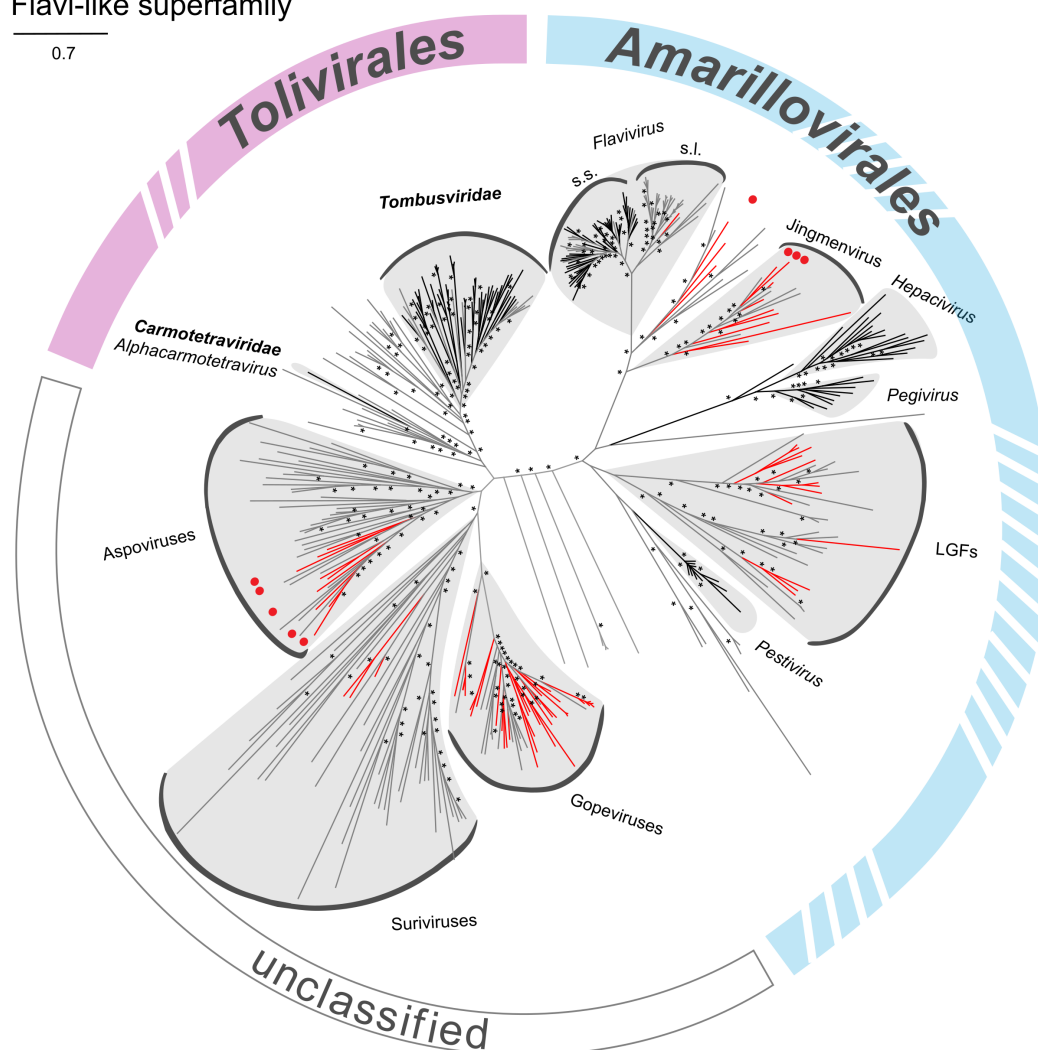


Fig. 1. Maximum likelihood phylogeny of *Flaviviridae* (order: *Amarillovirales*), and *Tombusviridae* and *Carmotetraviridae* (order: *Tolivirales*). The phylogenetic inference was based on an alignment of the RdRp region using RAxML-NG version 0.7.0 BETA (52). ICTV-classified viruses are shown in black, unclassified viruses in grey, and new viruses described in this study are marked in red. Red dots represent complete genomes of new viruses described in the present study. Bootstrap values above 70 are marked with an asterisk.

Legend of the following figure:

Fig. 2. Maximum likelihood phylogeny of flaviviruses. The phylogenetic inference was based on an alignment of the RdRp region with 1,000 bootstrap replicates, using RAxML-NG version 0.7.0 BETA (52). ICTV-classified viruses that belong to the *Flaviviridae* family are shown in black, unclassified flavi-related viruses in grey, and new flavi-related viral taxa described in this study are marked in red. Genomic sequence length is noted for every viral taxon and for segmented viruses sequence length corresponds to the RdRp-encoding segment. Viruses with a coding-complete genome described in this study are marked by a black dot. The tree is rooted to the branch leading to *Hepacivirus*. Bootstrap values below 70 are not shown.

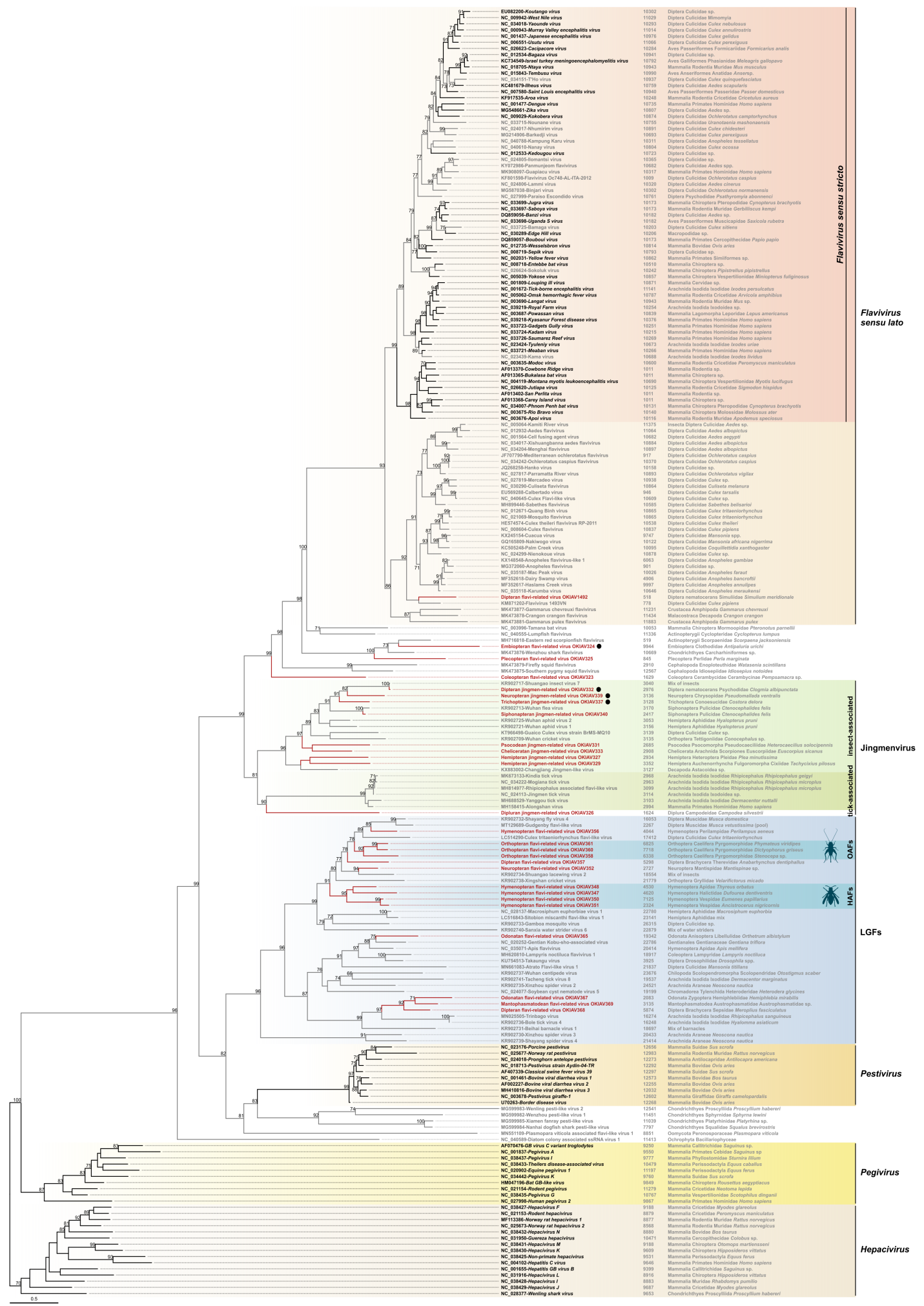


Fig. 2

Flavivirus sensu stricto

Flavivirus sensu lato

insect-associated

Jingmenvirus

tick-associated

DAEs

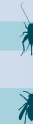
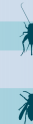
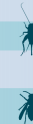
HAFs

LGFs

Pestivirus

Pegivirus

Hepacivirus



Genome and phylogenetic analyses of novel flavi-related viruses

The four viral sequences Dipteran flavi-related virus OKIAV1492, Embiopteran flavi-related virus OKIAV324, Plecopteran flavi-related virus OKIAV325, and Coleopteran flavi-related virus OKIAV323 established four lineages in phylogenetic sister relationship to members of the genus *Flavivirus* (**Fig. 2**). Coleopteran flavi-related virus OKIAV323 branched basal to the genus *Flavivirus* and to the unclassified flavi-related viruses, and was identified in a longhorn beetle (*Pempsamacra* sp.). Dipteran flavi-related virus OKIAV1492 was found in a black fly (*Simulium meridionale*) and shared a most recent common ancestor with the classical insect-specific flaviviruses. Black flies feed on vertebrates, e.g. birds and humans (26), supporting the notion that the genus *Flavivirus* (*sensu lato*) diversified within blood-feeding arthropods. Embiopteran flavi-related virus OKIAV324 was found in a webspinner and grouped with Plecopteran flavi-related virus OKIAV325, detected in a stonefly, as well as with viruses from marine hosts, such as sharks and molluscs, and a bat host. Embiopteran flavi-related virus OKIAV324 comprised a coding-complete flavivirus genome with a typical genome organization as shown in **Fig. 3**. Putative protein cleavage sites were similar as in other flaviviruses and are listed in **Table 1**. Exact size of corresponding proteins may differ from classical flaviviruses.

Genome and phylogenetic analyses of novel jingmenviruses

We have identified nine jingmenviruses, of which eight grouped with the insect-associated clade and one was placed on a long branch basal to the tick-associated clade based on phylogenetic analyses of RdRp proteins (**Fig. 2**). A search for non-RdRp genome segments yielded in total 14 additional segments. Coding-complete genomes, each with four segments, were assembled for the following three viruses: Dipteran jingmen-related virus OKIAV332, Trichopteran jingmen-related virus OKIAV337, and Neuropteran jingmen-related virus OKIAV339. As shown in **Fig. 4**, these genomes have a typical jingmenvirus-like genome organization. However, segment 2 and segment 4 of Neuropteran jingmen-related virus OKIAV339 differed from those of other insect-associated jingmenviruses as the respective ORFs did not overlap (27). Overlapping ORFs are generally encountered in insect-associated jingmenviruses but have also been observed in some tick-associated jingmenviruses (28).

Our data on a novel jingmenvirus in bristletails (*Campodea silvestrii*), which established a new deep branch (Dipluran jingmen-related virus OKIAV326) in addition to the insect- and tick-associated clades, indicated that the group of jingmenviruses is more diversified than previously shown (29) (**Fig. 2**). Dipluran jingmen-related virus OKIAV326 shared a most recent common ancestor with the tick-associated jingmenviruses and not as expected with the insect-associated jingmenviruses. This topology was confirmed by additional phylogenetic analyses based on an alignment containing only jingmenviruses (**Fig. S1**).

Genetic markers which differ between insect- and tick-associated jingmenviruses and which are present on genomic segments other than the RdRp could not be analyzed for Dipluran jingmen-related virus OKIAV326 as only the RdRp-encoding segment 1 was identified. Siphonapteran jingmen-related virus OKIAV340 grouped with Wuhan flea virus and both viruses were found in the same host species suggesting that this cluster of jingmenviruses is associated with fleas (order Siphonaptera). Of note, Shuangao insect virus 7 was identified in a mix of two dipteran

and one neuropteran insects that belong to the Psychodidae and Chrysopidae families (6), the same families of the hosts of its sister taxa, Dipteran jingmen-related virus OKIAV332 and Neuropteran jingmen-related virus OKIAV339. However, Dipteran jingmen-related virus OKIAV332 and Shuangao insect virus 7 were closer associated (**Fig. 2**), suggesting that Psychodidae species are the likely hosts for Shuangao insect virus 7.

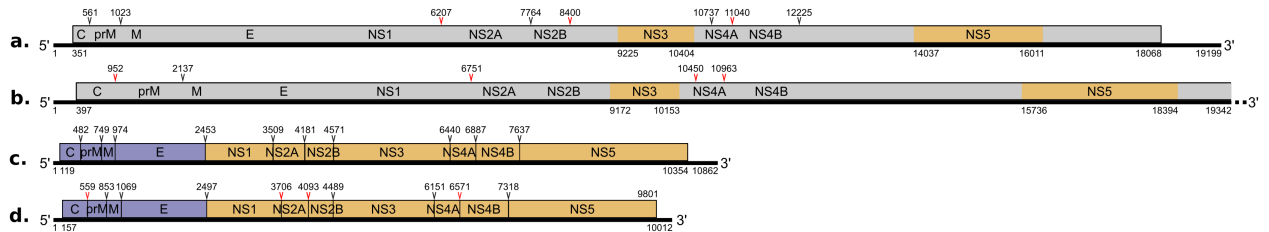


Fig. 3. Genome organization of Odonatan flavivirus-related virus OKIAV365 (**B**) and Embiopteran flavivirus-related virus OKIAV324 (**D**). Reference genomes of Soybean cyst nematode virus 5 (**A**) and Yellow fever virus (**C**) are shown for comparison. Nucleotide positions of ORF start/end and cleavage sites are indicated. Arrows point to cleavage sites: black arrows for sites similar to reference sequences, red arrows for sequence length/position variation in comparison to reference sequences. A complete list of the sequence stretches around the cleavage sites can be found in **Table 1**.

In addition to the establishment of host-specific clusters, the geographical distribution of jingmenviruses expanded significantly with the inclusion of data from this study, adding Australia, USA, Japan, and countries of Central Europe to the areas where jingmenviruses have been detected in previous studies (see **Table 2** for detailed host associations and sampling sites).

Segment co-segregation phylogenetic analyses shown in **Fig. 5** and **Fig. 6** reveal a consistent phylogenetic grouping among segment proteins for viruses that occur in the same host group, thus confirming the observation of intra-host adaptation for the tick-associated jingmenviruses (28) and extending the phylogenetic congruence to some of the insect-associated jingmenviruses. All phylogenetic co-segregation topologies were identical for Wuhan aphid virus 1 and Wuhan aphid virus 2, two jingmenviruses that were both encountered in aphids. The RdRp, capsid, and NS3 phylogenies were also topologically consistent for the grouping of Shuangao insect virus 7 and Dipteran jingmen-related virus OKIAV332, supporting the hypothesis of a dipteran host of Shuangao insect virus 7. Phylogenetic incongruence among the remaining jingmenviruses suggest an independent evolutionary scheme of reassortment events of the segments coding for the different proteins.

Genome and phylogenetic analyses of a novel clade of unclassified flavivirus-related viruses

We have identified 14 viruses within the clade of unclassified flaviviruses with a genome length ranging from 16–26 kb, marked as LGFs for “large genome flaviviruses” in **Fig. 2**. This clade is highly diversified, especially in comparison to pestiviruses branching as sister clade and associated with a wide diversity of arthropod hosts. Two novel host-specific subgroups were established based on four viruses detected in hymenopterans and three viruses detected in orthopterans, marked as HAFs and OAFs for Hymenoptera- and Oртоptera-associated flaviviruses, respectively. Hymenopteran flavivirus-related virus OKIAV356 did not group within the

HAF clade as confirmed by tree inference analyses based on a clade-specific alignment. Whereas HAFs and OAFs formed host-specific subclades, LGFs detected in other insects did not form such groups, e.g. viruses detected in Odonata or Diptera did not group together. The nearly coding-complete genome of Odonatan flavi-related virus OKIAV365 is comparable to the genome of Soybean cyst nematode virus 5 in **Fig. 3**. However, as LGFs show very limited similarity to flaviviruses and as no complete genome annotation for any LGF is available, annotations were only possible for the NS3 and NS5 genes.

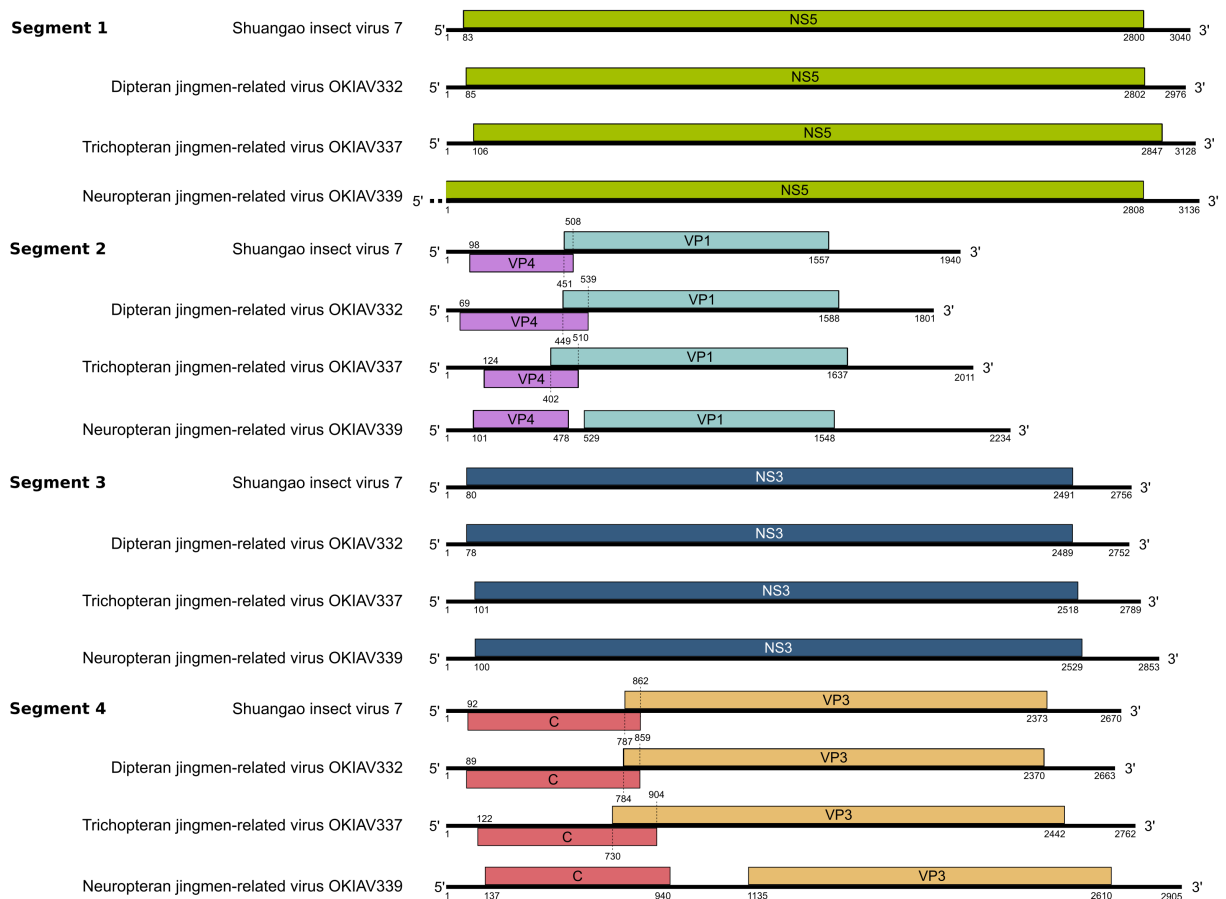


Fig. 4. Genome organization of Dipteran jingmen-related virus OKIAV332, Trichopteran jingmen-related virus OKIAV337, and Neuropteran jingmen-related virus OKIAV339. Nucleotide positions of ORF start/end and cleavage sites are indicated. The genome of Shuangao insect virus 7 is shown for comparison.

Legends of the following figures:

Fig. 5. Phylogenetic co-segregation of jingmenviruses. Analyses have been performed between RdRp and NS3, capsid, and VP1 genes. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. Bootstrap values below 70 are not shown.

Fig. 6. Phylogenetic co-segregation of jingmenviruses. Analyses have been performed between NS3, capsid, and VP1 genes. Topologically congruent clades are highlighted in color. Branches in black indicate taxa that do not share a common topological pattern in the respective tree pairs. Bootstrap values below 70 are not shown.

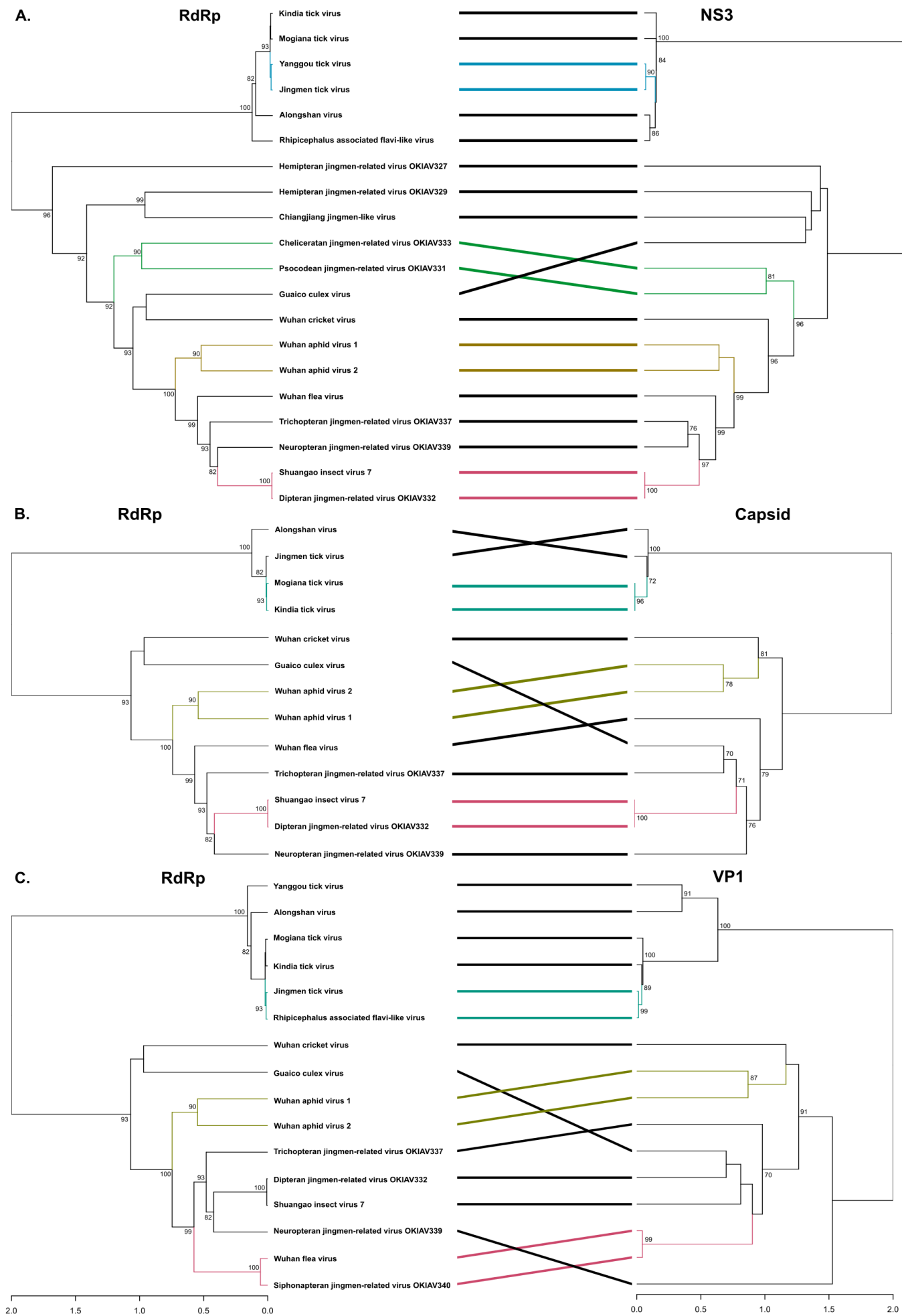


Fig. 5

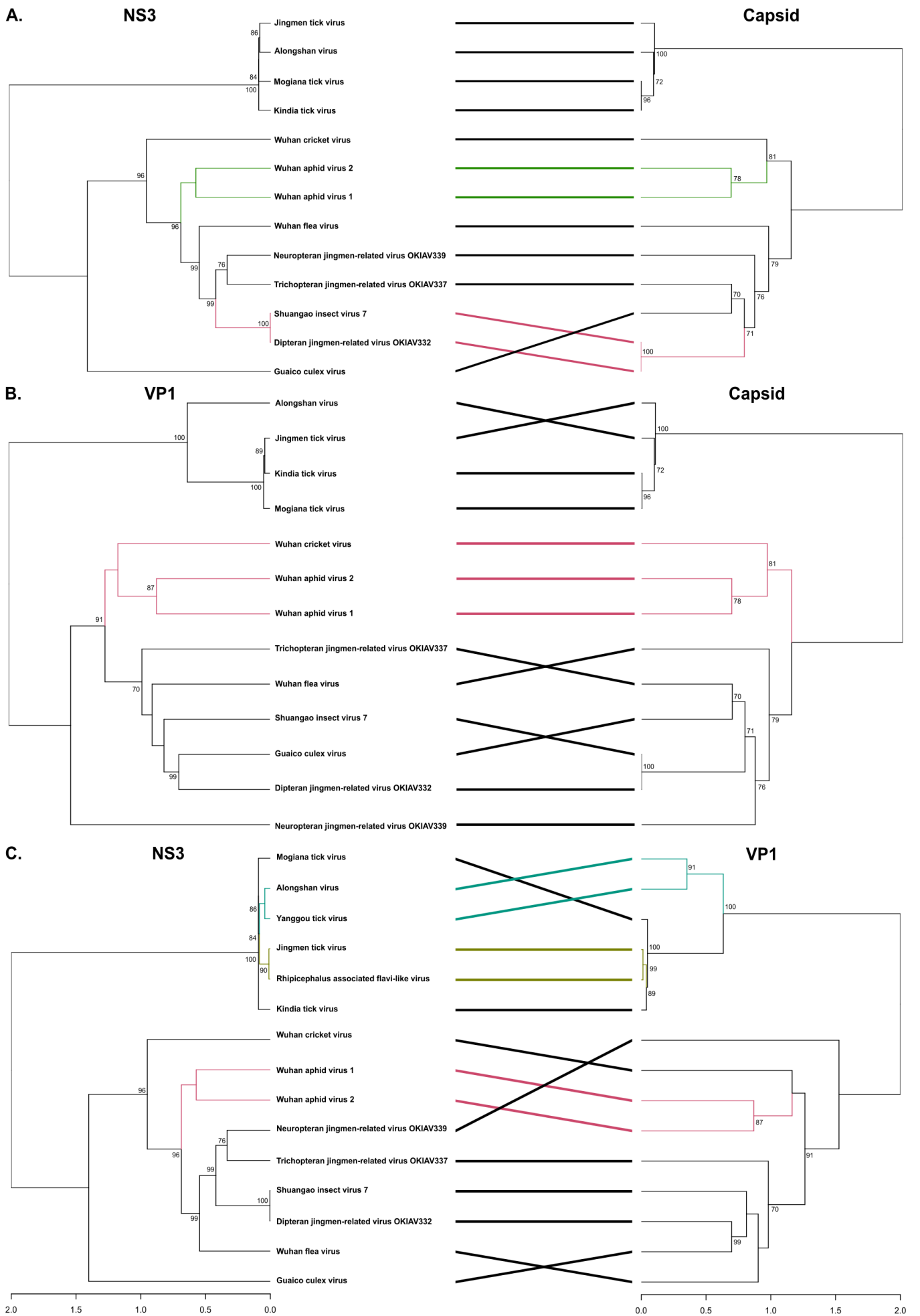


Fig. 6

Genome and phylogenetic analyses of novel tombus-related viruses

We detected forty-nine tombus-related viruses, which, together with other invertebrate viruses, established three distinct phylogenetic clades in distant relationship to the family *Tombusviridae* (**Fig. 7**). The three clades were provisionally named Gopeviruses, Suriviruses, and Aspoviruses. The clade of Gopeviruses was mostly populated by viruses identified within this study and showed several host-specific subclades. The Dipteran tombus-related viruses OKIAV386, -OKIAV387, and -OKIAV388 grouped together, and the subclade was named flower-feeding Diptera-associated tombus-related viruses (FF-DATs), stemming from three different Brachycera species that feed on nectar and pollen. The larvae of *Pterodontia mellii* are parasitic to spiders, which could explain the close relationship between Dipteran tombus-related virus OKIAV386 and Hubei tombus-like virus 30 that was found in spiders (7). Six viruses were identified in a wingless ectoparasite of bees, *Braula coeca*, and group in a monophyletic subclade of wingless Diptera-associated tombus-related viruses (W-DATs). The sister phylogenetic relationship of subclade W-DATs to Fig wasp tombus-like virus 1 and to the subclade of Hymenoptera-associated tombus-related viruses (HATs) indicated that the HATs subclade probably harbors a larger diversity of hymenopteran viruses, yet to be discovered. Both the W-DATs and HATs clades consist of viruses identified in monospecific insect hosts. The short genetic distance within the W-DATs and HATs subclades suggested a co-evolutionary relationship between hosts and viruses. Brachycera species host the viruses identified within the Diptera-associated tombus-related subclade (DATs), with the exception of Dipteran tombus-related virus OKIAV407 that stems from a Nematocera species.

Within the clade of Suriviruses, the insect-associated virus diversity dropped significantly, as this clade contains mostly viruses with non-insect arthropod hosts, such as Crustacea and Myriapoda. Only three viruses from this study fell within this clade, each stemming from a host of a different insect order (**Fig. 7**). Interestingly, *Leptomonas pyrrocoris* tombus-like virus 1 has a protozoan origin, found in the trypanosoma species *Leptomonas pyrrocoris* (30). It has been suggested that a possible transmission route for this parasite to acquire *Leptomonas pyrrocoris* tombus-like virus 1 is through the parasite's host, firebugs (order: Hemiptera, species: *Pyrrocoris apterus*) (30). Yet, no closely related virus of hemipteran origin was detected in our sample set.

Nine viruses from the present study were found within the clade of Aspoviruses (**Fig. 7**). Six of them are potentially coding-complete viral genomes (genome organization shown in **Fig. 8A**) and each belongs to a host of a different insect order. The genomes of these viruses showed similar organizations to the ones of tombusviruses, though with longer regions of overlapping ORFs. All tombusviruses share the common characteristic of amber codon readthrough in ORF1 which results in the expression of a larger protein (reviewed in (31)). Amber codons (UAG triplets) appeared in the predicted protein 1 of Hemipteran tombus-related virus OKIAV417 and in the predicted protein 2 of Odonatan tombus-related virus OKIAV411. We could not verify whether those two codons undergo readthrough, but it seems a likely outcome in both cases as this is a common observation for ORF1 across all tombusviruses. Viruses of the genus *Machlomovirus* involve an additional readthrough event in ORF3 (32). **Fig. 8B** shows five different types of genome organizations of incomplete tombus-related sequences listed in **Table 3**. Because some of the tombus-related viruses (Hubei tombus-like virus 28 and Wuhan insect

virus 35 and -21), as well as the genus *Dianthovirus* within the *Tombusviridae* family have a segmented genome, we searched for the presence of additional genome segments in our data. However, we only identified monopartite tombus-related viral sequences.

Legend of the following figure:

Fig. 7. Maximum likelihood phylogeny of the families *Tombusviridae* and *Carmotetraviridae* and related viruses. The phylogenetic inference was based on an alignment of the RdRp region with 1,000 bootstrap replicates, using RAxML-NG version 0.7.0 BETA (52). ICTV-classified viruses of the families *Tombusviridae* and *Carmotetraviridae* are shown in black, unclassified tombus-related viruses in grey, and tombus-related viral taxa described in this study in red. Genomic sequence length is noted for every viral taxon and for segmented viruses sequence length corresponds to the RdRp-encoding segment. Viruses with a coding-complete genome described in this study are marked by a black dot. The tree is rooted to the branch leading to the lower part of the tree which includes *Tombusviridae* and *Carmotetraviridae*. Bootstrap values below 70 are not shown.

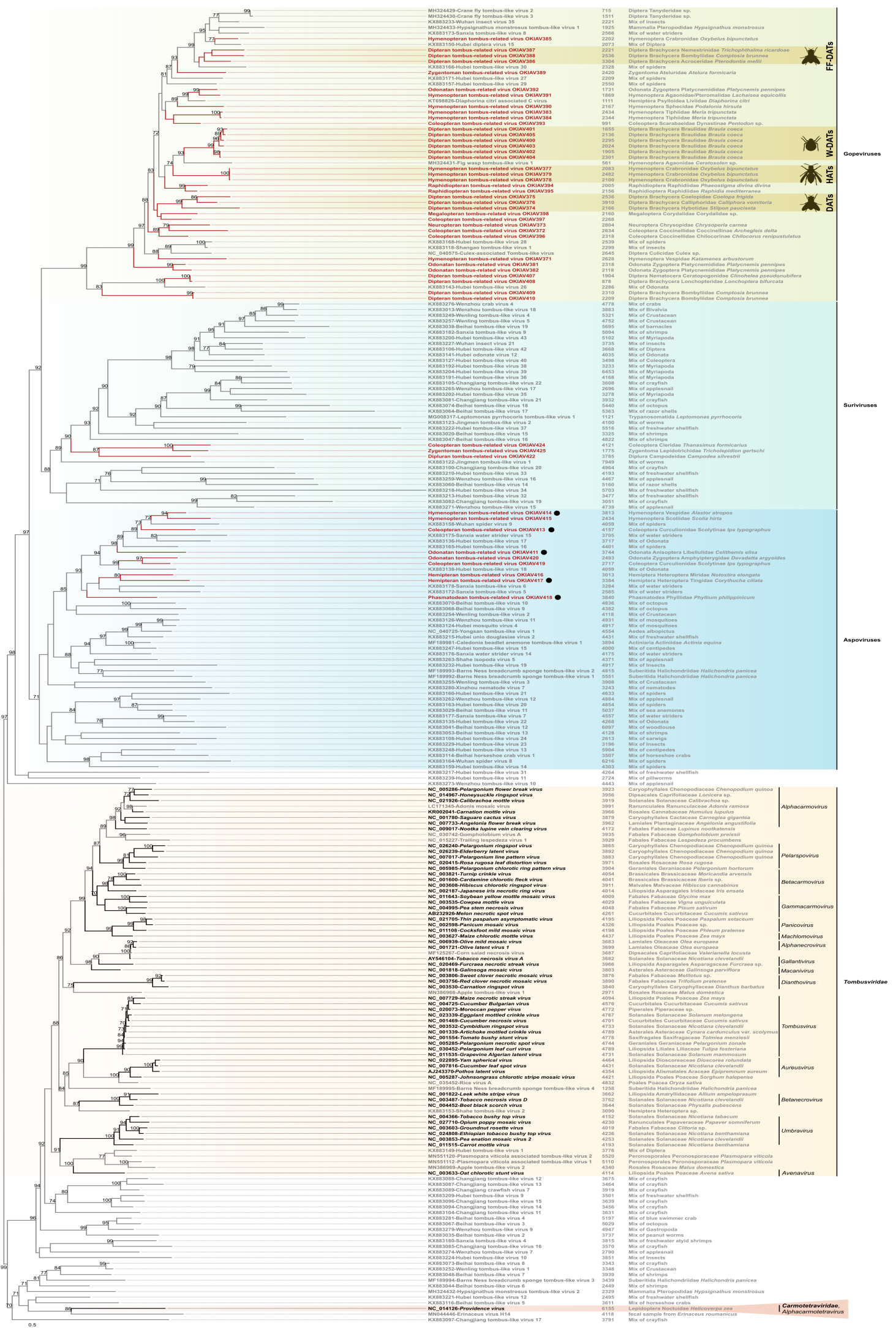


Fig. 7

FF-DVIRUS
W-DVIRUS
HAVS
DAVS
Survivivrus
Aspovirus
Alphacarnovirus
Peiarovirus
Betacarnovirus
Gammacarnovirus
Panovirus
Machonovirus
Alphacoronavirus
Mecanovirus
Dianthovirus
Tomoviridae
Aureovirus
Betanecrovirus
Umbravirus
Avenavirus
Carmotomoviridae
Alphacarnoviridae

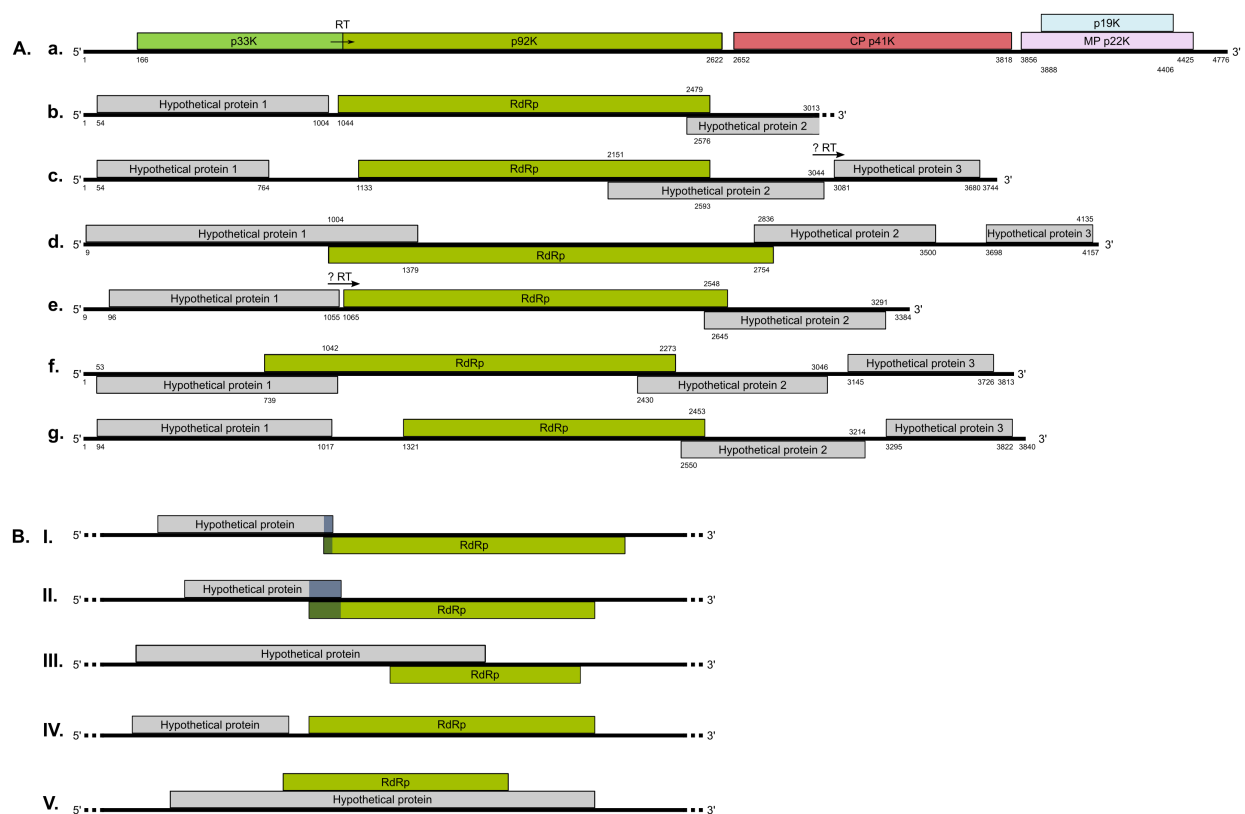


Fig. 8. **A.** Genome organization of tomosus-related viral sequences. The genome of *Tomato bushy stunt virus* is shown for comparison. Potential amber codon readthroughs are indicated with “RT”. Nucleotide positions of ORF start/end are indicated. Genomes correspond to **a.** Tomato bushy stunt virus, **b.** Hemipteran tomosus-related virus OKIAV416, **c.** Odonatan tomosus-related virus OKIAV411, **d.** Coleopteran tomosus-related virus OKIAV413, **e.** Hemipteran tomosus-related virus OKIAV417, **f.** Hymenopteran tomosus-related virus OKIAV414, **g.** Phasmatodean tomosus-related virus OKIAV418. **B.** Five scheme types corresponding to the genome organization for smaller-than-genome viral sequences: **I.** ORF1 and RdRp ORF overlap for less than 100 bases; **II.** ORF1 and RdRp ORF overlap for more than 100 bases; **III.** ORF1 is longer than RdRp ORF; **IV.** ORF1 and RdRp ORF don't overlap; **V.** ORF1 and RdRp ORF are nested. Correspondence of genome scheme to viral sequence is listed in **Table 3**.

Discussion

In this study, we discovered 76 novel insect-associated viruses, of which ten are coding-complete and belong to nine insect host orders, related to two of the largest families within the *Kitrinoviricota* phylum: *Flaviviridae* and *Tombusviridae*. Our data contribute to the known flavivirus-related insect-specific viral diversity introducing new insect host associations of non-typical flavivirus hosts, such as the Mantophasmatodea, Trichoptera, and Neuroptera orders. Phylogenetic relationships and distances suggest the establishment of two new taxonomic genera within the *Flaviviridae* family (**Fig. 2**). In addition, our sequences populate three highly diversified phylogenetic clades grouping as sister clades to the *Tombusviridae* (**Fig. 7**), each as diverse as to signify taxonomic assignment of a new family.

Phylogenetic inference of identified sequences in relationship to flaviviruses revealed the new sequences were widely distributed across the phylogeny and did not group with the established

genera of the family *Flaviviridae* (**Fig. 2**). Interestingly, no novel virus from the present study was identified within the genera *Flavivirus* (*sensu stricto*), *Pestivirus*, *Pegivirus*, or *Hepacivirus*. Despite the large number of diverse Diptera species (n = 81) that were sampled in the course of the 1KITE project (24), the sampling did not include any members of Culicidae, the family that includes all species of mosquitoes. Ticks were also not included in the sampling, which explains the absence of viral sequences within the *Flavivirus* genus in traditional arthropod hosts altogether. However, the absence of viruses from this genus (*s.s.*) in such a comprehensive sample of insects covering all extant orders is remarkable. Taken together with the occurrence of Dipteran flavi-like virus OKIAV1492 in a blood-feeding black fly (Simuliidae, *Simulium meridionale*), these findings provide evidence that the evolution of the genus *Flavivirus* (*s.l.*) is strongly associated with blood-feeding arthropods. Viruses of this genus mainly diversified in Culicidae and Ixodida.

Until now, jingmenviruses have been described in a set of hosts that, apart from mammals, includes ticks, mosquitoes, fleas, and non-bloodfeeding insects (6, 7, 16, 17, 28, 33–41). This group of viruses has not been taxonomically classified yet but is phylogenetically related to the genus *Flavivirus* based on NS3 and NS5 protein similarities. Yet, their genomes, unlike flaviviruses, are segmented and generally consist of four segments (with the exception of Guaiaco Culex virus that has an additional fifth segment, however experimentally shown to be dispensable for genome replication (35)). We identified jingmenviruses in hosts with a variety of ecological lifestyles such as drain flies, booklice, aquatic insects, scorpions, and soil-dwelling bristletails, thereby enriching the current knowledge of jingmenvirus host range.

Previous phylogenetic reconstructions separated jingmenviruses into two clades according to their host associations: the insect- and the tick-associated clades (28). In addition to the distinct host range, the insect- and tick-associated clades were shown to differ by a number of genetic features. For example, conserved nucleotide stretches in the 5' and 3' UTRs, and a 3'-end poly-A tail were present in all four genomic segments in tick-associated but not in insect-associated jingmenviruses underlining the differentiation of these two groups. Of note, for Dipluran jingmen-related virus OKIAV326 only a partial sequence of the RdRp genomic segment was retrieved. The 5'-end of the segment is missing but the 3'-end is available and lacks a poly-A tail. Dipluran jingmen-related virus OKIAV326 shared a most recent common ancestor with the tick-associated jingmenviruses. These findings imply that the ancestor of jingmenviruses may have existed in insects with genomic segments not containing 3'-end poly-A tails. Recently, class II viral fusion protein domains (VFPs) were identified in envelope glycoproteins of tick-associated jingmenviruses, whereas the insect-associated viruses did not contain class II VFPs. Class II VFPs in tick-associated jingmenviruses stem likely from flavivirus class II VFPs and have probably been secondarily lost in the insect-associated clade when this clade separated from the tick-associated clade via divergent evolution, segment reassortment or recombination (27). The presence of a mucin-like domain, an ectodomain of beta-sheets and alpha-helices in the tick-associated jingmenvirus glycoproteins further differentiates them from the insect-associated jingmenvirus glycoproteins. None of the four insect-associated jingmenvirus glycoprotein segments reported here contained the fusion peptide or the mucin-like domain present in the tick-associated jingmenviruses. The phylogenetic separation of the tick-associated from the insect-

associated jingmenviruses was apparent and maintained in all phylogenetic segment co-segregation analyses (**Fig. 5** and **Fig. 6**), supporting previous suggestions on divergent evolution of the envelope glycoprotein structure in tick- and insect-associated jingmenviruses (27).

Segmentation within the *Flaviviridae*-related viruses is suggested to have evolved only once because of the monophyletic nature of the jingmenvirus clade (6). Although the non-structural proteins NS3 and NS5 share sequence similarities with flaviviruses, homologs of the jingmenvirus structural proteins, glycoprotein and capsid, are yet unknown (34). Explanations for the unrelatedness to the flavivirus structural proteins may lie within the hypothesis that two segments of an ancestral flavivirus genome were perhaps co-packaged or captured by structural “orphan” proteins of a co-infecting virus (34).

A series of previously described viruses with distinct genome organization (similar to flaviviruses, but with substantially larger genome length: 16–23 kb) (6, 9–15), share homologous sequence regions with flaviviruses mainly in the protease, helicase, and RNA-dependent RNA polymerase (RdRp) genes. We identified 14 viruses that together with the previously described viruses formed the phylogenetic clade of LGFs within the *Flaviviridae* (**Fig. 2**). The distinct genomic length and organization (**Fig. 3**), as well as the phylogenetic distances of LGFs (**Fig. 2**) signify their classification as a separate genus within the *Flaviviridae* family. LGFs show a remarkably broad spectrum of host associations, ranging from gentian plants to nematodes and other arthropods. LGFs grouped as sister clade to the genus *Pestivirus* which contains livestock pathogenic viruses with genomes of approximately 12 kb. The observed genetic divergence of LGFs was much greater than that of pestiviruses suggesting that the LGF viruses are ancestral to the mammalian pestiviruses.

Numerous tombus-related viruses have recently been described in non-plant hosts, such as marine invertebrates and terrestrial arthropods, demonstrating both mono- and bipartite genome organizations (7). These viruses show a large genetic diversity and branch basal to the plant-associated *Tombusviridae* suggesting that the RNA virome of angiosperm plants evolved via horizontal virus transfer from invertebrates (7, 25, 42). Phylogenetic analyses of the 49 newly identified tombus-related viruses in this study showed that the sequences fell within three highly diversified major clades, each being as diverse as the family *Tombusviridae*, and shared a most recent common ancestor with the family *Tombusviridae* (**Fig. 7**). Our findings corroborate the horizontal virus transfer theory of plant virus evolution via multiple events of horizontal virus transfer from different invertebrate hosts (42). As the origin of invertebrates precedes that of angiosperm plants, the members of *Tombusviridae* possibly evolved after being passed on to angiosperm plant hosts from invertebrates due to the tight biological associations formed between these two host groups (42). Also, the absence of tombus-related viruses in unicellular eukaryotic hosts indicates the absence of a common viral ancestry that preceded the last eukaryotic common ancestor (43). Despite the position of *Leptomonas pyrrocoris* tombus-like virus 1 within the clade of Suriviruses (**Fig. 7**), the transmission route of this virus likely involves the trypanosomatid firebug host (30). These findings together with the novel viruses identified in insect hosts within the present study suggest that insect-associated viruses played an ancestral role in plant-insect virus exchange but are not involved in current virus transmission

(42, 43).

With the given ease of generating environmental and organismic metagenomic data, large projects of virus sequence discoveries will accompany us in the future years. The number of uncultivated virus genomes has surpassed that of viral genomes stemming from cultivated isolates (44). An important effort has been made to define the minimum information that should accompany uncultured viral genomes with their sequence publication (44). Crucially also, the ICTV has made an important step to align with the current flow of metagenomic sequence information, allowing the classification of viruses that stem from metagenomic sources (45). Nevertheless, efforts towards elaborate examinations of sequence information should extend to protect the quality of published sequence information.

Materials and Methods

Virus identification

Screening of insect transcriptome data to identify and filter viral sequences was done as described previously (3). In brief, nucleotide assemblies from 1,243 insect and other arthropod transcriptomes were translated in all six open reading frames (ORFs). These were computationally screened for viral infections utilizing an amino acid profile hidden Markov model (pHMM) of the core region of the flavivirus RdRp gene. The transcriptome data stem from the 1KITE consortium (NCBI Umbrella BioProject accession number: PRJNA183205, “The 1KITE project: evolution of insects”). The detection of distant evolutionary homologs by utilizing pHMMs comes with the downside of inflating the results with redundant sequences. Filtering out redundant matches included a two-fold approach as previously described (3): a) by aligning all hits to the initial flavivirus template RdRp alignment and subsequently removing sequences that were too short and/or did not overlap with the core RdRp motifs. Phylogenetic tree inference with PhyML v.3.2.0 (46) assisted further in identifying erroneously matched sequences by inspecting the topologies for instances of long branch attraction; b) all sequences were compared with blastp of the BLAST+ suite v.2.2.28 (47) against the non-redundant NCBI protein database, which had been filtered for viral sequences beforehand. The combination of results from both approaches and a cut-off of 300 nucleotides that served as the minimum sequence length yielded the final viral sequences.

Genome assembly and annotation

Search for further viral sequence information was performed using full genome sequences of related viruses downloaded from GenBank (**Table S1**). In particular, for segmented viruses (jingmenviruses, the genus *Dianthovirus* of *Tombusviridae*, Hubei tombus-like virus 28, and Wuhan insect virus 35 and -21), the corresponding sequences for non-RdRp segments were downloaded from GenBank and served as the basis to create per-segment sequence databases. Insect transcriptome assemblies were matched to amino acid libraries per genomic segment using diamond version 0.9.26 (48). A read-mapping step using Geneious (Geneious v.9.1.8, Biomatters, 581 Auckland, New Zealand, <https://www.geneious.com>) was applied to all viral sequences (flavi-, tombus-, and jingmen-associated) in order to verify the assemblies and

potentially elongate them at their 5' or 3' ends. Genome annotations were carried out using the webservice mode of InterProScan software (49) for protein domain predictions. Cleavage sites were predicted by aligning obtained sequences to known flaviviruses as well as by the webservice mode of SignalP-5.0 (50) for host signalase site prediction.

Phylogenetic analyses

Amino acid multiple sequence alignments of the RdRp gene for *Flaviviridae*- and *Tombusviridae*-related sequences were performed using the E-INS-i mode of mafft v7.407 (51). Unclassified viral sequences that have previously been reported as flavi-related, as well as unclassified flaviviruses and tombusviruses were included in the alignments. The LG amino acid substitution model was selected based on the Bayesian Information Criterion (BIC), with the options that count empirically amino acid frequencies from the data and allow a proportion of invariable sites. Maximum likelihood phylogenies were inferred in RAxML-NG version 0.7.0 BETA (52) using 1,000 bootstrap replicates and the transfer bootstrap expectation metric for clade credibility. Trees were visualized using the R package *ggtree* (53). Phylogenetic co-segregation of different jingmenvirus segments was based on maximum likelihood phylogenies of the corresponding viral proteins, following the alignment and tree inference as described above. Tanglegrams were visualized using the R package *dendextend* (54).

Data availability

Sequences are available under GenBank accession numbers MW208755-MW208806 and MW314679-MW314716 (see also **Table S1**).

Acknowledgements

We thank Dr. Karen Meusemann, Dr. Jeanne Wilbrandt, and Dr. Tanja Ziesmann for providing the metadata of the 1KITE transcriptomes. Computation has been performed on the HPC for Research cluster of the Berlin Institute of Health. This work was supported by grants from the Deutsche Forschungsgemeinschaft DR 772/7-2 (grant to CD), the German Ministry of Research DZIF TTU 01.801 (grant to CD), and the Federal Ministry of Education and Research BMBF 01KI1716 (grant to SJ) as part of the Research Network Zoonotic Infectious Diseases. Work of BM has been funded by a Leibniz competition grant “Graduate School on Biodiversity Genomics”.

References

1. Bolling BG, Weaver SC, Tesh RB, et al. Insect-Specific Virus Discovery: Significance for the Arbovirus Community. *Viruses*. 2015;7(9):4911–4928.
2. Marklewitz M, Zirkel F, Kurth A, et al. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proc. Natl. Acad. Sci. U. S. A.* 2015;112(24):7536–7541.
3. Käfer S, Paraskevopoulou S, Zirkel F, et al. Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathog.* 2019;15(12):e1008224.

4. Li C-X, Shi M, Tian J-H, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife*. 2015;4:e05378.
5. Stork NE. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annu. Rev. Entomol.* 2018;63:31–45.
6. Shi M, Lin X-D, Vasilakis N, et al. Divergent Viruses Discovered in Arthropods and Vertebrates Revise the Evolutionary History of the Flaviviridae and Related Viruses. *J. Virol.* 2016;90(2):659–669.
7. Shi M, Lin X-D, Tian J-H, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540(7634):539–543.
8. Porter AF, Pettersson JH-O, Chang W-S, et al. Novel hepac- and pegi-like viruses in native Australian wildlife and non-human primates. *Virus Evol.* 2020;6(2):veaa064.
9. Kobayashi K, Atsumi G, Iwadate Y, et al. Gentian Koku-sho-associated virus: a tentative, novel double-stranded RNA virus that is relevant to gentian Koku-sho syndrome. *J. Gen. Plant Pathol.* 2013;79(1):56–63.
10. Bekal S, Domier LL, Gonfa B, et al. A novel flavivirus in the soybean cyst nematode. *J. Gen. Virol.* 2014;95(Pt 6):1272–1280.
11. Teixeira M, Sela N, Ng J, et al. A novel virus from *Macrosiphum euphorbiae* with similarities to members of the family Flaviviridae. *J. Gen. Virol.* 2016;97(5):1261–1271.
12. Webster CL, Longdon B, Lewis SH, et al. Twenty-Five New Viruses Associated with the Drosophilidae (Diptera). *Evol. Bioinform. Online*. 2016;12(Suppl 2):13–25.
13. Remnant EJ, Shi M, Buchmann G, et al. A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations. *J. Virol.* 2017;91(16):e00158–17.
14. Kondo H, Fujita M, Hisano H, et al. Virome Analysis of Aphid Populations That Infest the Barley Field: The Discovery of Two Novel Groups of Nege/Kita-Like Viruses and Other Novel RNA Viruses. *Front. Microbiol.* 2020;11:509.
15. Faizah AN, Kobayashi D, Isawa H, et al. Deciphering the Virome of *Culex vishnui* Subgroup Mosquitoes, the Major Vectors of Japanese Encephalitis, in Japan. *Viruses*. 2020;12(3):264.
16. Kuivanen S, Levanov L, Kareinen L, et al. Detection of novel tick-borne pathogen, Alongshan virus, in *Ixodes ricinus* ticks, south-eastern Finland, 2019. *Euro Surveill.* 2019;24(27):e1900394.
17. Wang Z-D, Wang B, Wei F, et al. A New Segmented Virus Associated with Human Febrile Illness in China. *N. Engl. J. Med.* 2019;380(22):2116–2125.
18. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*. 2015;479-480:2–25.
19. Herrera-Vásquez JA, Córdoba-Sellés MC, Cebrián MC, et al. Seed transmission of Melon necrotic spot virus and efficacy of seed-disinfection treatments. *Plant Pathol.* 2009;58(3):436–442.
20. Mehle N, Ravnikar M. Plant viruses in aqueous environment—survival, water mediated transmission and detection. *Water Res.* 2012;46(16):4902–4917.
21. Blitvich BJ, Firth AE. A Review of Flaviviruses that Have No Known Arthropod Vector. *Viruses*. 2017;9(6):154.
22. Koonin EV, Dolja VV. Metaviromics: a tectonic shift in understanding virus evolution. *Virus Res.* 2018;246:A1–A3.
23. Parry R, Asgari S. Discovery of Novel Crustacean and Cephalopod Flaviviruses: Insights into the Evolution and Circulation of Flaviviruses between Marine Invertebrate and Vertebrate Hosts. *J. Virol.*

2019;93(14):e00432–19.

24. Misof B, Liu S, Meusemann K, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014;346(6210):763–767.
25. Wolf YI, Kazlauskas D, Iranzo J, et al. Origins and Evolution of the Global RNA Virome. *MBio*. 2018;9(6):e02329–18.
26. DeFoliart GR, Rao MR. The Ornithophilic Black Fly *Simulium Meridionale* Riley (Diptera: Simuliidae) Feeding On Man During Autumn. *J. Med. Entomol.* 1965;2(1):84–85.
27. Garry CE, Garry RF. Proteomics Computational Analyses Suggest that the Envelope Glycoproteins of Segmented Jingmen Flavi-Like Viruses Are Class II Viral Fusion Proteins (β -Penetrenes) with Mucin-Like Domains. *Viruses*. 2020;12(3):260.
28. Temmam S, Bigot T, Chrétien D, et al. Insights into the Host Range, Genetic Diversity, and Geographical Distribution of Jingmenviruses. *mSphere*. 2019;4(6):e00645–19.
29. Zhang X, Wang N, Wang Z, et al. The discovery of segmented flaviviruses: implications for viral emergence. *Curr. Opin. Virol.* 2020;40:11–18.
30. Grybchuk D, Akopyants NS, Kostygov AY, et al. Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite *Leishmania*. *Proc. Natl. Acad. Sci. U. S. A.* 2018;115(3):E506–E515.
31. Sit TL, Lommel SA. Tombusviridae. *Elsevier Oceanogr. Ser.* 2015;1–9.
32. Scheets K. Analysis of gene functions in Maize chlorotic mottle virus. *Virus Res.* 2016;222:71–79.
33. Maruyama SR, Castro-Jorge LA, Ribeiro JMC, et al. Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in *Rhipicephalus microplus* ticks from Brazil. *Mem. Inst. Oswaldo Cruz*. 2014;109(1):38–50.
34. Qin X-C, Shi M, Tian J-H, et al. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proc. Natl. Acad. Sci. U. S. A.* 2014;111(18):6744–6749.
35. Ladner JT, Wiley MR, Beitzel B, et al. A Multicomponent Animal Virus Isolated from Mosquitoes. *Cell Host Microbe*. 2016;20(3):357–367.
36. Pauvolid-Corrêa A, Solberg O, Couto-Lima D, et al. Novel Viruses Isolated from Mosquitoes in Pantanal, Brazil. *Genome Announc.* 2016;4(6):e01195–16.
37. Villa EC, Maruyama SR, de Miranda-Santos IKF, et al. Complete Coding Genome Sequence for Mogiana Tick Virus, a Jingmenvirus Isolated from Ticks in Brazil. *Genome Announc.* 2017;5(18):e00232–17.
38. Emmerich P, Jakupi X, von Pössel R, et al. Viral metagenomics, genetic and evolutionary characteristics of Crimean-Congo hemorrhagic fever orthonairovirus in humans, Kosovo. *Infect. Genet. Evol.* 2018;65:6–11.
39. Kholodilov IS, Litov AG, Klimentov AS, et al. Isolation and Characterisation of Alongshan Virus in Russia. *Viruses*. 2020;12(4):362.
40. Ternovoi VA, Protopopova EV, Shvalov AN, et al. Complete Coding Genome Sequence for Novel Multicomponent Kindia Tick Virus Isolated from Ticks Collected in Guinea. *submitted to Microbiol. Resour. Announc.*
41. Shen S, Moming A, Luo T, et al. Viromes of *Hyalomma asiaticum*, *Hyalomma detritum* and *Dermacentor nuttalli* ticks from Xinjiang Uygur Autonomous Region, China. GenBank 2018. *unpublished*.
42. Dolja VV, Koonin EV. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* 2018;244:36–52.

43. Dolja VV, Koonin EV. Common origins and host-dependent diversity of plant and animal viromes. *Curr. Opin. Virol.* 2011;1(5):322–331.
44. Roux S, Adriaenssens EM, Dutilh BE, et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* 2019;37(1):29–37.
45. Simmonds P, Adams MJ, Benkő M, et al. Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 2017;15(3):161–168.
46. Guindon S, Dufayard J-F, Lefort V, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 2010;59(3):307–321.
47. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
48. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods.* 2015;12(1):59–60.
49. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9):1236–1240.
50. Armenteros JJA, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 2019;37(4):420–423.
51. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013;30(4):772–780.
52. Kozlov AM, Darriba D, Flouri T, et al. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* 2019;35(21):4453–4455.
53. Yu G, Smith DK, Zhu H, et al. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 2017;8(1):28–36.
54. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics.* 2015;31(22):3718–3720.
55. Souza WM de, Fumagalli MJ, Torres Carrasco A de O, et al. Viral diversity of Rhipicephalus microplus parasitizing cattle in southern Brazil. *Sci. Rep.* 2018;8(1):16315.
56. Pascoal J de O, Siqueira SM de, Maia R da C, et al. Detection and molecular characterization of Mogiana tick virus (MGTV) in Rhipicephalus microplus collected from cattle in a savannah area, Uberlândia, Brazil. *Ticks Tick Borne Dis.* 2019;10(1):162–165.
57. Sameroff S, Tokarz R, Charles RA, et al. Viral Diversity of Tick Species Parasitizing Cattle and Dogs in Trinidad and Tobago. *Sci. Rep.* 2019;9(1):10421.
58. Xia H, Hu C, Zhang D, et al. Metagenomic profile of the viral communities in Rhipicephalus spp. ticks from Yunnan, China. *PLoS One.* 2015;10(3):e0121609.
59. Jia N, Liu H-B, Ni X-B, et al. Emergence of human infection with Jingmen tick virus in China: A retrospective study. *EBioMedicine.* 2019;43:317–324.
60. Meng F, Ding M, Tan Z, et al. Virome analysis of tick-borne viruses in Heilongjiang Province, China. *Ticks Tick Borne Dis.* 2019;10(2):412–420.
61. Dinçer E, Hacıoğlu S, Kar S, et al. Survey and Characterization of Jingmen Tick Virus Variants. *Viruses.* 2019;11(11):1071.
62. Wang Z-D, Wang W, Wang N-N, et al. Prevalence of the emerging novel Alongshan virus infection in sheep and cattle in Inner Mongolia, northeastern China. *Parasit. Vectors.* 2019;12(1):450.

Table 1. Putative polyprotein cleavage sites of Embiopteran flavi-related virus OKIAV324, Odonatan flavi-related virus OKIAV365, and closely related flaviviruses.

Cleavage site	JEV	WNV	YFV	Embiopteran flavi-related virus OKIAV324	SbCNV-5	Odonatan flavi-related virus OKIAV365
C/prM	IAYAGA/MKLSNF	IASVGA/VTLN	LLMTGG/VTLVRK	LVMVAA/AQFSAD	ILLGGG/ARFVRK	FIGKET/VKSAA
pr/M	SKRSRR/SVSVQT	SRRSRR/SLTVQT	SRRSRR/AIDLPT	KTRLR/VAISIP	STRGKR/AAAKSS	DRSARP/AHAGRK
prM/E	VAPAYS/FNCLGM	VAPAYS/FNCLGM	VGPAYS/AHCIGI	YLVGS/KACHQV	?	?
E/NS1	TNVHA/DTGCAI	VNVHA/DTGCAI	SLGVGA/DOGCAI	FWGVKG/DEMVLS	?	?
NS1/NS2A	QVDAF/NGEMV	QVNAY/NADMID	RSWVTA/GEIHAY	KPVYTS/GYIHDL	DSVDTA/SLRHRL	KGIDDV/YNETNK
NS2A/NS2B	PNKKR/GWPATE	PNRKR/GWPATE	RIFGRR/SIPVNE	IYRRKR/PKHDDP	YFPKR/SSGWNE	?
NS2B/NS3	LKTTKR/GGVFWD	LQYTKR/GGVLWD	VRGARR/SGDVLW	YGQWQG/RGTIMD	SGTERR/VSVAEG	?
NS3/NS4A	AAGKR/SAISFI	ASGKR/SQIGLI	FAEGRR/GAAEVL	KYARLR/GKHASF	LSTGRF/GLFKTQ	HANFKR/DNVKKA
NS4A/NS2B	GVVAA/NEYGM	SAVAA/NEMGW	VSAVAA/NELGML	NPQIIS/ALIEVK	RSAPKE/LEGMDE	GKLEK/LAGLKN
NS2B/NS5	PSLKR/GRPGG	PGLKR/GGAKG	MKTGRR/GSANG	FETPRT/GSSHAE	SAHAKK/EGKDKA	?

Table 2. Host associations down to the species level (whenever available) of jingmenviruses shown in **Fig. 2**.

Virus	Order	Family	Species	Geographic location [reference]
Shuangao insect virus 7	Diptera Neuroptera	Psychodidae Chrysopidae	<i>Chrysopidae sp.</i> <i>Psychoda alternata</i> <i>Diptera sp.</i>	China: Zhejiang ⁽⁶⁾
Dipteran jingmen-related virus OKIAV332	Diptera	Psychodidae	<i>Clogmia albipunctata</i>	USA: North Carolina [this study]
Neuropteran jingmen-related virus OKIAV339	Neuroptera	Chrysopidae	<i>Pseudomallada ventralis</i>	Austria: near Vienna [this study]
Trichopteran jingmen-related virus OKIAV337	Trichoptera	Conoesucidae	<i>Costora delora</i>	Australia: Victoria [this study]
Siphonapteran jingmen-related virus OKIAV340	Siphonaptera	Pulicidae	<i>Ctenocephalides felis</i>	USA [this study]
Wuhan flea virus	Siphonaptera	Pulicidae	<i>Ctenocephalides felis</i>	China: Hubei ⁽⁶⁾
Wuhan aphid virus 1	Hemiptera	Aphididae	<i>Hyalopterus pruni</i>	China: Hubei ⁽⁶⁾
Wuhan aphid virus 2	Hemiptera	Aphididae	Mix of <i>Hyalopterus pruni</i> and <i>Aulacorthum magnoliae</i>	China: Hubei ⁽⁶⁾
Cheliceratan jingmen-related virus OKIAV333	Scorpiones	Euscorpiidae	<i>Euscorpius sicanus</i>	Italy: Sicily [this study]
Psocodean jingmen-related virus OKIAV331	Psocoptera	Pseudocaeciliidae	<i>Heterocaecilius solocipennis</i>	Japan: Hokkaido [this study]
Guaico Culex virus	Diptera	Culicidae	<i>Culex coronator</i> , <i>Culex interrogator</i> , <i>Culex declarator</i>	Brazil: Nhecolandia ⁽³⁶⁾ Trinidad: Aripo ⁽³⁵⁾ Peru: Loreto ⁽³⁵⁾ Panama: Soberania, Achiotos ⁽³⁵⁾
Wuhan cricket virus	Orthoptera	Tettigoniidae	<i>Conocephalus sp.</i>	China: Hubei ⁽⁶⁾
Hemipteran jingmen-related virus OKIAV329	Hemipteran	Cixiidae	<i>Tachycixius pilosus</i>	Germany: Thuringia [this study]
Hemipteran jingmen-related virus OKIAV327	Hemipteran	Pleidae	<i>Plea minutissima</i>	Germany: Lower Saxony [this study]
Changjiang Jingmen-like virus	Decapoda	Cambaridae	<i>Procambarus clarkii</i>	China: Hubei ⁽⁷⁾
Mogiana tick virus	Ixodida Artiodactyla	Ixodidae Bovidae	<i>Rhipicephalus microplus</i> <i>Bos sp.</i>	Brazil: Uberlandia ^(33, 37, 55, 56) Trinidad and Tobago ⁽⁵⁷⁾
Kindia tick virus	Ixodida	Ixodidae	<i>Rhipicephalus geigy</i>	Guinea ⁽⁴⁰⁾
Rhipicephalus associated flavivirus-like virus	Ixodida	Ixodidae	<i>Rhipicephalus microplus</i>	China: Yunnan ⁽⁵⁸⁾
Jingmen tick virus	Chiroptera Primates Primates Diptera Ixodida	Pteropodidae Hominidae Cercopithecidae Culicidae Ixodidae	<i>Pteropus lylei</i> <i>Homo sapiens</i> <i>Ptilocolobus rufomitrat</i> <i>Armigeres sp.</i> , <i>Anopheles sp.</i> , <i>Culex sp.</i> <i>Amblyomma testudinarium</i> ,	China: Hubei, Heilongjiang ^(34, 59, 60) Lao PDR ⁽²⁸⁾ Uganda: Kibale National Park ⁽³⁵⁾ Kosovo ⁽³⁸⁾ French Antilles ⁽²⁸⁾ Cambodia ⁽²⁸⁾ France ⁽²⁸⁾

			<i>Dermacentor nuttalli</i> , <i>Haemaphysalis longicornis</i> , <i>Haemaphysalis campanulata</i> , <i>Haemaphysalis flava</i> , <i>Hyalomma marginatum</i> , <i>Ixodes sinensis</i> , <i>Ixodes granulatus</i> , <i>Ixodes ricinus</i> , <i>Rhipicephalus microplus</i> , <i>Rhipicephalus sanguineus</i>	Turkey ⁽⁶¹⁾
Yanggou tick virus	Ixodida	Ixodidae	<i>Dermacentor nuttalli</i>	China: Xinjiang ⁽⁴¹⁾
Alongshan virus	Ixodida Artiodactyla Primates Diptera	Ixodidae Bovidae Hominidae Culicidae	<i>Ixodes ricinus</i> , <i>Ixodes persulcatus</i> <i>Bos sp.</i> , <i>Ovis aries</i> <i>Homo sapiens</i> <i>Culex pipiens</i> , <i>Culex tritaeniorhynchus</i> , <i>Anopheles yatsushiroensis</i> , <i>Aedes vexans</i>	Finland: south-eastern Finland ⁽¹⁶⁾ Russia: Chelyabinsk, Republic of Karelia ⁽³⁹⁾ China: Hulunbuir, Hinggan, Qiqihar, Greater Khingan, Jilin ^(17, 62)
Dipluran jingmen-related virus OKIAV326	Diplura	Campodeidae	<i>Campodea silvestrii</i>	Germany: North Rhine Westphalia [this study]

Table 3. Genome scheme type to viral sequence correspondence from **Fig. 8B**. Sequence length for all sequences below is in the range of 2–2.8 kb.

Type I	Coleopteran tombus-related virus OKIAV396, Hymenopteran tombus-related virus OKIAV377, -OKIAV378, -OKIAV379, -OKIAV390, -OKIAV415, Megalopteran tombus-related virus OKIAV398, Neuropteran tombus-related virus OKIAV373, Odonatan tombus-related virus OKIAV382
Type II	Coleopteran tombus-related virus OKIAV372, -OKIAV397, Dipteran tombus-related virus OKIAV374, -OKIAV400, -OKIAV404, -OKIAV405, -OKIAV409, -OKIAV410, Hymenopteran tombus-related virus OKIAV383, -OKIAV384, Odonatan tombus-related virus OKIAV381, Zygentoman tombus-related virus OKIAV389
Type III	Dipteran tombus-related virus OKIAV374, -OKIAV375, -OKIAV387, -OKIAV388, Hymenopteran tombus-related virus OKIAV385, Raphidiopteran tombus-related virus OKIAV395
Type IV	Coleopteran tombus-related virus OKIAV419, Hymenopteran tombus-related virus OKIAV371, Odonatan tombus-related virus OKIAV420
Type V	Dipluran tombus-related virus OKIAV422

5 General Discussion

5.1 On virus taxonomy and discovery

Virus discovery is undoubtedly a fascinating field of research both in identifying evolutionary links to known viruses and in predicting events of new virus spread. Over the course of time, the impact of viruses in living forms and life in the broad sense has been noticeable in a wide spectrum of events that do not necessarily involve disease (Horzinek, 1997; Durruthy-Durruthy et al., 2016). The Digital Revolution of the late 20th century has launched the Information Era, facilitating a data-based approach to study genetic information and rethink the ways we understand associations among organisms. Often, voices of concern were raised in an attempt to cultivate awareness and cautiousness for the shortcomings of an exclusively technology-based standpoint in studying the various aspects of life (Levins & Lewontin, 1985).

The field of virology has indisputably and largely benefited from the technological advancements. The potential offered by the next-generation sequencing technologies in processing sequence information massively and in parallel, has had a huge impact in virus research. Manifestations of this impact include the tremendous amount of viral sequence information derived from metagenomic analyses, stemming from environmental or organismic samples (Roux et al., 2019). Adding to that, the magnitude of computationally-derived viral sequences and genomes has also impacted the prerequisites needed to be fulfilled for a virus to be officially classified as such (Simmonds et al., 2017).

The official body for virus taxonomic classification, the ICTV, publishes new reports every year where newly-discovered viruses that fulfill the classification criteria are reported and are officially classified as viruses (Adams et al., 2017). Nevertheless, as the criteria are not universal across all viral taxa (and cannot be due to the variety of characteristics of viruses), question marks and objections arise often in terms of virus species classifications (reviewed in Simmonds, 2018). Also, the flood of new viral sequences in the taxonomy has possibly served as a trigger for the recent discussion on whether virus species nomenclature should follow a binomial scheme, similar to that of cellular organisms (Siddell et al., 2020; Hull & Rima, 2020).

Another topic worth mentioning is the critique that has been voiced on the frequency of taxonomic proposals and assignments (Käfer et al., 2019). Given the flood of viral sequence information produced almost at weekly pace, it is important to enable independent verification of these discoveries by the scientific community. Taken together with the time required between the submission of a manuscript and its final publication, it is not to be unexpected that virus taxa names will differ between a published paper and current taxonomy, as classification updates may have meanwhile been released.

On a different note, the field of virus metagenomics manifests its necessity in clinical diagnostics whenever a wet lab approach proves to be limited (Gu et al., 2019; Chiu & Miller, 2019). The number of available computational pipelines for virus discovery is continuously growing (as mentioned in **Chapter 1.4.1**) in an attempt to address the needs for more accurate results. Recently, a proficiency testing for validation of results delivered by various pipelines was published (Brinkmann et al., 2019). The here-presented pipeline (**Chapter 1.4.2**) has been tested against a subset of the dataset of Brinkmann et al. (2019) and successfully identified all 500 reads of the (novel) avian bornavirus within a two-hour time frame. Despite the long list of published pipelines, many labs prefer to develop their own algorithms for processing metagenomic data for virus discovery since the main tools for data processing are few. Therefore, basic computational skills are indispensable with the ever-increasing need for bioinformatic analyses in the medical and natural sciences.

5.2 On the discoveries of RNA viruses in insects

With viruses being ubiquitous across all life domains, the organismic group of insects could not have been an exception. Apart from numerous known pathogenic viruses that occur in hematophagous insects and impose public and global health burdens, insects harbour a vast diversity of viruses. A plethora of insect viruses are continuously being discovered since the advent of metagenomics, increasing our understanding of virus evolution. Phenomena such as horizontal virus transfer between insect and plant hosts, virus genome segmentation, or arboviruses with evolutionary origins in insect viruses gain insights with each virus discovery study.

Numerous single-stranded negative sense RNA viruses were discovered in insects within the scope of research presented in **Chapter 2** and published by Käfer et al. (2019). These viruses showed similarity to families of the phylum *Negarnaviricota*. With regards to non-segmented viruses of the *Haploviricotina* order, many of the discovered insect viruses were related to *Rhabdoviridae*. Interestingly, genera that are mammal-specific (such as *Lyssavirus*) or arbovirus-specific (such as *Ephemerovirus* and *Tibrovirus*) did not show insect virus relatives. Most of the *Rhabdoviridae*-related viruses identified by Käfer et al. (2019) branched together with other insect viruses described by Li et al. (2015). Many clades with clear insect host associations to the order level appeared after the addition of the viruses identified here. Also, the majority of viruses identified by Li et al. (2015) and Käfer et al. (2019) formed a large clade sister to plant viruses, suggesting an ancestral horizontal virus transfer event. *Rhabdoviridae*-related viruses derived from the study by Käfer et al. (2019) that have been officially proposed for taxonomic classification appear in the ICTV taxonomic proposal by Walker et al. (2020) and can be found in **Table A1**.

An exception to the non-segmented genome architecture of viruses within *Haploviricotina* are some viruses of the *Chuviridae*, *Qinviridae*, and *Yueviridae* families (Li et al., 2015; Käfer et al., 2019). These families host viruses identified within insects, crustaceans, and nematodes with mono- or bi-segmented genome that appears in circular conformations in some cases. Phylogenetic addition of the viruses identified by Käfer et al. (2019) revealed several host-specific clades, and Odonatan chu-related virus OKIAV136 and -137 confirmed the host association with Odonata of *Odonate mivirus*. The close phylogenetic relationship of these viruses is reflected in the recent taxonomy proposal by the *Jingchuvirales* ICTV study group where these viruses were grouped in one genus, *Odonatavirus* (Di Paola et al. (2020) and **Table A1**). This proposal suggested a general reformation of the whole *Jingchuvirales* order to include new families, genera, and species with the minimum taxonomy inclusion criterion of a coding-complete L protein (Di Paola et al., 2020).

In the *Mononegavirales* order, the pattern of host integrity maintenance was encountered in families with mammal-specific viruses, such as *Paramyxoviridae* and *Bornaviridae*. On the contrary, insect virus families were enriched by the findings of Käfer et al. (2019) that suggested the establishment of new genera within the *Nyamiviridae*, *Xinmoviridae*, and *Lispiviri-*

dae families. The most recent ICTV taxonomic proposal for *Nyamiviridae* includes viruses identified here (Dietzgen et al. (2020) and **Table A1**). Interestingly, the *Lispiviridae* family is in sister relationship to the *Mymonaviridae* family that includes viruses identified primarily in plant fungal hosts, but also plants and arthropods, with implications for horizontal virus transfer events between these hosts.

The viral subphylum *Polyploviricotina* is characterized by segmented genomes of its members. The variety of viruses related to the *Orthomyxoviridae* family discovered within arthropod and invertebrate hosts by Li et al. (2015) and Shi et al. (2016a) is not phylogenetically related to influenza viruses within the family. Viruses discovered in insects by Käfer et al. (2019) provide a major enrichment in virus diversity of this family. The genus *Thogotovirus*, associated until now only with mammalian hosts, encountered a host expansion as viruses that group with this genus were found in aquatic hemipterans and green lacewings. Also, the genus *Quaranjavirus* became paraphyletic to the other genera of *Orthomyxoviridae* and, at the same time, appeared in sister relationship to the majority of the viruses identified by Käfer et al. (2019). As observed in other families, insect host-specific clades arose in the *Orthomyxoviridae* phylogeny as well, suggesting the classification of many new virus taxa.

Several new viruses grouped within or in relation to families of the *Bunyavirales* order, with most of them belonging to *Phasmaviridae* and *Phenuiviridae*. Both of these families are rich in insect viruses, especially *Phasmaviridae* which is exclusively populated by insect hosts. Host-specific clades emerged in *Phasmaviridae* with the addition of viruses discovered by Käfer et al. (2019), in particular within the genus *Orthophasmavirus*. Most of the new viruses were found in insects of the Hymenoptera order, introducing this new host association in the family for the first time. Apart from the taxonomic classification of viruses identified by Käfer et al. (2019) within known *Phasmaviridae* genera, the genus *Hymovirus* was newly established in the latest ICTV taxonomic proposal (Ballinger et al. (2020) and **Table A1**).

Within already-established *Phenuiviridae* genera, new viruses grouped mostly with *Goukovirus* and *Pidchovirus*. One of those viruses has been proposed for official taxonomic classification within *Pidchovirus* in the latest *Phenuiviridae* ICTV taxonomic proposal (Marklewitz

et al. (2020b) and **Table A1**). Also, with the addition of new insect viruses in the *Phenuiviridae* phylogeny, a large insect virus diversity was unravelled, forming a large and distant clade to the rest of the genera within the family, but including *Mobuvirus*. Interesting findings related to the rest of the families in *Bunyavirales* included an arthropod-associated clade of viruses sister to *Hantaviridae* and the Blattodean nairo-related virus OKIAV321 which appeared sister to the genus *Striavivirus* of *Nairoviridae*. The latest virus is now proposed to constitute an own genus, *Ocetevirus*, according to the latest *Nairoviridae* taxonomic proposal of the ICTV (Marklewitz et al. (2020a) and **Table A1**).

~

New viruses associated with positive-sense single stranded RNA virus families were identified in insect hosts (presented in **Chapter 4** and submitted for publication by Paraskevopoulou et al. (unpublished)). Viruses in the *Tombusviridae* family occur in plants, yet a major diversity of insect, other arthropod, and invertebrate viruses were associated with tombusviruses (Shi et al., 2016b; Paraskevopoulou et al., unpublished). The phylogenetic scheme presented by Shi et al. (2016b) was maintained and enriched by current virus findings in insects (Paraskevopoulou et al., unpublished). A whole new clade of insect viruses and several major branches were added to the phylogeny. Interestingly, *Tombusviridae* viruses appeared to derive from arthropod and invertebrate viruses, as they form an internal clade in the phylogeny, branching from the non-plant viruses (Paraskevopoulou et al., unpublished). This pointed at the occurrence of an ancestral horizontal virus transfer event that probably gave rise to the today's *Tombusviridae* diversity.

Viruses associated with *Flaviviridae* were also discovered by Paraskevopoulou et al. (unpublished). Three of the four classified genera of this family (*Hepacivirus*, *Pegivirus*, *Pestivirus*) are associated only with mammalian hosts. Members of the genus *Flavivirus* include renown arboviruses but also insect-specific viruses. Although the term "insect-specific" implies a rather broad insect host spectrum, these viruses occur only within mosquitoes and ticks (Blitvich & Firth, 2015). Despite the broad spectrum of sampling to represent all extant orders of insects, newly-discovered viruses did not cluster within any of the four *Flaviviridae* genera. However, Dipteran flavi-related virus OKIAV1492, which appeared in sister relationship and distant to the members of *Flavivirus*, was identified within a black fly, a

host known to feed on vertebrates (DeFoliart & Rao, 1965). It seems plausible that a flavivirus-related ancestral virus jumped from a non-arbovirus host to a blood-feeding insect that gave rise to the great arbovirus diversity of *Flavivirus*.

A novel unclassified virus group provisionally named "jingmenviruses" has been discovered during the last years (Temmam et al., 2019). Presenting an amino acid identity to RdRp and NS3 proteins, jingmenviruses are related to *Flaviviridae*. Jingmenviruses are found within two host groups, namely ticks and various insects, and the discoveries by Paraskevopoulou et al. (unpublished) enriched the existing jingmenvirus insect host diversity. Jingmenviruses are phylogenetically divided into two clades, in agreement with a tick or an insect host association. Of note, all jingmenvirus genomes from the tick-associated clade present a 3' polyadenylated tail, a unique characteristic not only of this particular clade, but of the *Flaviviridae* family in general. Additionally, the segmented jingmenvirus genome configuration is encountered for the first time in *Flaviviridae*-associated viruses. These characteristics are not only valuable in an evolutionary frame, but can be useful criteria for the taxonomic establishment of jingmenviruses.

Importantly, it should be highlighted that as metagenomic studies rely mainly on environmental or pooled organismic samples, it is often difficult (as much as impossible) to connect virus findings with specific hosts. In studies where a single-host extraction protocol is utilized, a direct host association is very likely to be in place but might as well be misleading. Explanations for the later relate to the numerous underlying bacterial, fungal, etc microorganisms that occur within an organism, or to partly ingested food fragments. Therefore, *in vitro* studies of virus isolation and characterization can assist in deciphering clear host associations. Subsequent virus discoveries with a known host association are necessary and provide independent confirmation of previously-reported host associations.

As a concluding remark to the above, it should be mentioned that the here-presented findings could not be verified using an *in vitro* approach due to sample aliquot limitations. Although the virus findings were derived from transcriptomic data, the ones that are smaller-than-genome size could also represent integrated elements in the host genome. Protein-coding regions of viruses that belong to single-stranded RNA virus families have been found to be

integrated into mammalian and *Aedes* genomes and also actively expressed (Horie et al., 2016; Horie, 2019; Pischedda et al., 2019). Nevertheless, viruses with a coding-complete genome are unlikely to represent cases of host integration since integration is not known to occur in this pattern.

5.3 On deltaviruses

Hepatitis delta virus (HDV) is a matter of health concern globally, responsible for severe courses of hepatitis. Humans can get infected with HDV provided that they have previously been infected with Hepatitis B virus (HBV), or when the two viruses occur simultaneously as a coinfection. As HDV hijacks the viral envelope of HBV to produce infectious viral particles and as we do not know a case with an HDV-alone infection, HDV has been characterized as a *satellite virus*. Often, HDV is referred to as "viroid" since its genome is a short single-stranded RNA molecule with ribozyme structures. However, this is an incorrect characterization because although viroids are able to use the host's translation machinery for replication, they do not encode any protein. HDV encodes two protein isoforms, therefore it cannot be placed in this category. Also, technically, because the HDV genome is "encapsidated" together with the two delta antigens in a ribonucleoprotein complex, it could be accounted as a virus, but this is still a matter of ongoing debate.

The first mammalian non-human deltavirus reported by Paraskevopoulou et al., 2020 (presented also in **Chapter 3**) sheds light on the evolution of HDV. This novel viral agent occurred in *Proechimys semispinosus* rodents without a hepadnavirus coinfection and was detected in blood, stool, and various organs. Together with predominant identification in adult male rodents and absence of significant ecological co-factor influence, a horizontal route of transmission linked to competitive male behaviour seems highly likely. From an evolutionary perspective, RDeV is the closest known relative to HDV. However, unlike HDV, RDeV does not encode two protein isoforms, i.e. the small and large delta antigens. Only the small delta antigen was identified in rodent serum and was expressed by an RDeV construct *in vitro*. Until now, no viral coinfecting agent was found in the case of RDeV, despite the high hepacivirus prevalence in the studied rodents (Schmid et al., 2018).

Not long before, no equivalent form of HDV was known, neither as an entity, nor in hosts

other than humans. The recent discoveries of delta-like viruses in a variety of non-human organisms challenge the evolutionary theories on the origin of Hepatitis delta virus (Wille et al., 2018; Hetzel et al., 2019; Chang et al., 2019; Paraskevopoulou et al., 2020; Edgar et al., 2020; Bergner et al., 2020). Before these discoveries, Hepatitis delta virus was suggested to have evolved from either a viroid-like RNA or a host gene (Taylor & Pelchat, 2010). The proof of viral replication for two non-human deltaviruses, RDeV and SDeV (Paraskevopoulou et al., 2020; Szirovicza et al., 2020), substantially weakens the host derivation theory, at least with regards to a human host derivation. The lack of a coinfecting HBV together with the lack of a C-terminal -CXXQ motif in all non-human deltaviruses, rather suggest a potential adaptation of a deltavirus precursor to human hosts via utilization of the HBV envelope proteins. The experimentally proven ability of HDV to utilize a viral envelope which does not stem from HBV creates a new understanding for deltavirus transmission (Perez-Vargas et al., 2019; Szirovicza et al., 2020). Host jumps that enabled deltaviruses to exploit different helpers are also suggested within two recent pre-prints reporting additional mammalian deltaviruses in bats, marmots, and deer (Edgar et al., 2020; Bergner et al., 2020). Also, an *in vitro* cytokine response in human cells was recently reported to be triggered by HDV-derived extracellular vesicles from cells infected with HDV alone (Jung et al., 2020). This study provides important insights in the use of extracellular vesicles for potential transportation of genomic or sub-genomic HDV-RNA to non-infected cells. Therefore, experimental *in vitro* infection of cells with non-human deltaviruses and analyses of the produced extracellular vesicles will have important implications on understanding deltavirus evolution and the necessity of a helper coinfecting viral envelope for deltavirus transmission.

A viroid origin for deltaviruses collectively could still be considered on the basis of similar genome properties (circularity, self-cleaving activity). However, the discovery of a variety of non-human deltaviruses in the latest years should be followed by an extensive search for deltavirus-like agents in ancestral lineages and related organisms to decipher the evolutionary origins of deltaviruses. Functional genome studies to understand the transmission, replication, and packaging mechanisms of the novel non-human deltaviruses will assist in unravelling the evolutionary history of this infectious agent.

Bibliography

- Abeywickrama-Samarakoon, N., Cortay, J.-C., Sureau, C., Müller, S., Alfaiate, D., Guerrieri, F., Chaikuad, A., Schröder, M., Merle, P., Levrero, M., et al. (2020). Hepatitis Delta Virus histone mimicry drives the recruitment of chromatin remodelers for viral RNA replication. *Nature communications*, *11*(1), 1–13.
- Adams, M. J., Lefkowitz, E. J., King, A. M., Harrach, B., Harrison, R. L., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Mushegian, A. R., et al. (2017). Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Archives of virology*, *162*(8), 2505–2538.
- Aiewsakun, P., & Simmonds, P. (2018). The genomic underpinnings of eukaryotic virus taxonomy: creating a sequence-based framework for family-level virus classification. *Microbiome*, *6*(1), 38.
- Akopyants, N. S., Lye, L.-F., Dobson, D. E., Lukeš, J., & Beverley, S. M. (2016). A novel bunyavirus-like virus of trypanosomatid protist parasites. *Genome announcements*, *4*(4).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403–410.
- Alves, C., Freitas, N., & Cunha, C. (2008). Characterization of the nuclear localization signal of the hepatitis delta virus antigen. *Virology*, *370*(1), 12–21.
- Andrew, S. (2010). A quality control tool for high throughput sequence data.
URL <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Ballinger, M. J., Hall, R. A., Langevin, S. A., Pauvolid-Correa, A., Paraskevopoulou, S., Drosten, C., & Junglen, S. (2020). Create four new species in the genus *Orthophasmavirus*, create two new species in the genus *Feravirus*, and create one new genus (*Hymovirus*) including two new species (*Bunyavirales: Phasmaviridae*). Tech. rep., ICTV taxonomic report 2020.020M.
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological reviews*, *35*(3), 235.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, *19*(5), 455–477.
- Berenbaum, M. (2017). Insect biodiversity—millions and millions. *Insect biodiversity: science and society*, (pp. 783–792).
- Bergner, L., Orton, R., Broos, A., Tello, C., Becker, D., Carrera, J., Patel, A., Biek, R., & Streicker, D. (2020). Diversification of mammalian deltaviruses by host shifting. *bioRxiv*.
- Bigot, T., Temmam, S., Pérot, P., & Eloit, M. (2019). RVDB-prot, a reference viral protein database and its HMM profiles. *F1000Research*, *8*(530), 530.

- Bisaillon, M., & Lemay, G. (1997). Viral and cellular enzymes involved in synthesis of mRNA cap structure. *Virology*, *236*(1), 1–7.
- Blitvich, B. J., & Firth, A. E. (2015). Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, *7*(4), 1927–1959.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
- Bolling, B. G., Weaver, S. C., Tesh, R. B., & Vasilakis, N. (2015). Insect-specific virus discovery: significance for the arbovirus community. *Viruses*, *7*(9), 4911–4928.
- Branch, A. D., & Robertson, H. D. (1984). A replication cycle for viroids and other small infectious RNA's. *Science*, *223*(4635), 450–455.
- Brazas, R., & Ganem, D. (1996). A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. *Science*, *274*(5284), 90–94.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, *99*(22), 14250–14255.
- Brinkmann, A., Andrusch, A., Belka, A., Wylezich, C., Höper, D., Pohlmann, A., Petersen, T. N., Lucas, P., Blanchard, Y., Papa, A., et al. (2019). Proficiency testing of virus diagnostics based on bioinformatics analysis of simulated *in silico* high-throughput sequencing data sets. *Journal of clinical microbiology*, *57*(8), e00466–19.
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature methods*, *12*(1), 59–60.
- Calisher, C. H., & Higgs, S. (2018). The discovery of arthropod-specific viruses in hematophagous arthropods: an open door to understanding the mechanisms of arbovirus and arthropod evolution? *Annual review of entomology*, *63*, 87–103.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, *10*(1), 421.
- Chang, M.-F., Chang, S., Chang, C.-I., Wu, K., & Kang, H. (1992). Nuclear localization signals, but not putative leucine zipper motifs, are essential for nuclear transport of hepatitis delta antigen. *Journal of virology*, *66*(10), 6019–6027.
- Chang, W.-S., Pettersson, J. H., Le Lay, C., Shi, M., Lo, N., Wille, M., Eden, J.-S., & Holmes, E. C. (2019). Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evolution*, *5*(2), vez021.

- Chao, M., Hsieh, S., & Taylor, J. (1991). The antigen of hepatitis delta virus: examination of in vitro RNA-binding specificity. *Journal of virology*, *65*(8), 4057–4062.
- Chen, H.-Y., Shen, D.-T., Ji, D.-Z., Han, P.-C., Zhang, W.-M., Ma, J.-F., Chen, W.-S., Goyal, H., Pan, S., & Xu, H.-G. (2019). Prevalence and burden of hepatitis D virus infection in the global population: a systematic review and meta-analysis. *Gut*, *68*(3), 512–521.
- Chen, P. J., Chang, F. L., Wang, C. J., Lin, C. J., Sung, S. Y., & Chen, D. S. (1992). Functional study of hepatitis delta virus large antigen in packaging and replication inhibition: role of the amino-terminal leucine zipper. *Journal of virology*, *66*(5), 2853–2859.
- Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. *Nature reviews. Genetics*, *20*(6), 341.
- Chou, H.-C., Hsieh, T.-Y., Sheu, G.-T., & Lai, M. M. (1998). Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. *Journal of virology*, *72*(5), 3684–3690.
- Daigh, L. H., Griffin, B. L., Soroush, A., Mamedov, M. R., & Casey, J. L. (2013). Arginine-rich motifs are not required for hepatitis delta virus RNA binding activity of the hepatitis delta antigen. *Journal of virology*, *87*(15), 8665–8674.
- DeFoliart, G., & Rao, M. R. (1965). The ornithophilic black fly *Simulium meridionale* Riley (Diptera: Simuliidae) feeding on man during autumn. *Journal of medical entomology*, *2*(1), 84–85.
- Deny, P. (2006). Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? In *Hepatitis Delta Virus*, (pp. 151–171). Springer.
- Di Paola, N., Dheilly, N. M., Kuhn, J. H., Junglen, S., Paraskevopoulou, S., Postler, T. S., & Shi, M. (2020). Reorganize the order to include four new families, 18 new genera, and 22 new species (*Jingchuvirales*). Tech. rep., ICTV taxonomic report 2020.026M.
- Diener, T. O. (2001). The viroid: biological oddity or evolutionary fossil? *Advances in virus research*, *57*, 137–184.
- Dietzgen, R. G., Kondo, H., Kuhn, J. H., Vasilakis, N., Jiang, D., & Junglen, S. (2020). Create one new species in the genus *Nyavirus* and four new species in one new genus *Formivirus* (*Mononegavirales: Nyamiviridae*). Tech. rep., ICTV taxonomic report 2020.024M.
- Dolja, V. V., & Koonin, E. V. (2018). Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus research*, *244*, 36–52.
- Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, *148*(4), 1667–1686.
- Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E. J., Davila, J., Mall, M., Wong, W. H., Wysocka, J., et al. (2016). The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nature genetics*, *48*(1), 44–52.

- Edgar, R. C., Taylor, J., Altman, T., Barbera, P., Meleshko, D., Lin, V., Lohr, D., Novakovsky, G., Al-Shayeb, B., Banfield, J. F., et al. (2020). Petabase-scale sequence alignment catalyses viral discovery. *bioRxiv*.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic acids research*, 47(D1), D427–D432.
- Elena, S. F., Dopazo, J., Flores, R., Diener, T. O., & Moya, A. (1991). Phylogeny of viroids, viroidlike satellite RNAs, and the viroidlike domain of hepatitis delta virus RNA. *Proceedings of the National Academy of Sciences*, 88(13), 5631–5634.
- Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. i. accuracy assessment. *Genome research*, 8(3), 175–185.
- Eyngor, M., Zamostiano, R., Tsofack, J. E. K., Berkowitz, A., Bercovier, H., Tinman, S., Lev, M., Hurvitz, A., Galeotti, M., Bacharach, E., et al. (2014). Identification of a novel RNA virus lethal to tilapia. *Journal of clinical microbiology*, 52(12), 4137–4146.
- Fattovich, G., Giustina, G., Christensen, E., Pantalena, M., Zagni, I., Realdi, G., Schalm, S., on Viral Hepatitis (Eurohep, E. C. A., et al. (2000). Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type B. *Gut*, 46(3), 420–426.
- Filipovska, J., & Konarska, M. M. (2000). Specific HDV RNA-templated transcription by pol II in vitro. *Rna*, 6(1), 41–54.
- Flores, R., Gas, M.-E., Molina-Serrano, D., Nohales, M.-Á., Carbonell, A., Gago, S., De la Peña, M., & Daròs, J.-A. (2009). Viroid replication: rolling-circles, enzymes and ribozymes. *Viruses*, 1(2), 317–334.
- Freitas, N., Salisse, J., Cunha, C., Toshkov, I., Menne, S., & Gudima, S. O. (2012). Hepatitis delta virus infects the cells of hepadnavirus-induced hepatocellular carcinoma in woodchucks. *Hepatology*, 56(1), 76–85.
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22), 3718–3720.
- Gorbalenya, A. E., Gordo, K. H. J., Lauber, C., Ward, V. K., & Dorrington, R. A. (2011). ICTV taxonomic report 2010.001a-qI: Revision of the family *Tetraviridae*. *International Committee on Taxonomy of Viruses*.
- Gordien, E., Rosmorduc, O., Peltekian, C., Garreau, F., Bréchet, C., & Kremsdorf, D. (2001). Inhibition of hepatitis B virus replication by the interferon-inducible MxA protein. *Journal of virology*, 75(6), 2684–2691.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, 29(7), 644.

- Gribskov, M., McLachlan, A. D., & Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proceedings of the National Academy of Sciences*, *84*(13), 4355–4358.
- Gu, W., Miller, S., & Chiu, C. Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. *Annual Review of Pathology: Mechanisms of Disease*, *14*, 319–338.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, *59*(3), 307–321.
- Gurung, K., Wertheim, B., & Falcao Salles, J. (2019). The microbiome of pest insects: it is not just bacteria. *Entomologia Experimentalis et Applicata*, *167*(3), 156–170.
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*, *56*(2), 61–77.
- Hepojoki, J., Hetzel, U., Paraskevopoulou, S., Drosten, C., Harrach, B., Zerbini, M., Koonin, E. V., Krupovic, M., Dolja, V., & Kuhn, J. H. (2020). Create one new realm (*Ribozyviria*) including one new family (*Kolmioviridae*) including genus *Deltavirus* and seven new genera for a total of 15 species. Tech. rep., ICTV taxonomic report 2020.012D.
- Hetzel, U., Szirovicza, L., Smura, T., Prähauser, B., Vapalahti, O., Kipar, A., & Hepojoki, J. (2019). Identification of a novel deltavirus in Boa Constrictors. *MBio*, *10*(2), e00014–19.
- Hong, S.-Y., & Chen, P.-J. (2010). Phosphorylation of serine 177 of the small hepatitis delta antigen regulates viral antigenomic RNA replication by interacting with the processive RNA polymerase II. *Journal of virology*, *84*(3), 1430–1438.
- Horie, M. (2019). Interactions among eukaryotes, retrotransposons and riboviruses: endogenous riboviral elements in eukaryotic genomes. *Genes & Genetic Systems*, *94*(6), 253–267.
- Horie, M., Kobayashi, Y., Honda, T., Fujino, K., Akasaka, T., Kohl, C., Wibbelt, G., Mühlendorfer, K., Kurth, A., Müller, M. A., et al. (2016). An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. *Scientific reports*, *6*, 25873.
- Horzinek, M. C. (1997). The birth of virology. *Antonie van Leeuwenhoek*, *71*(1-2), 15–20.
- Hsieh, S., & Taylor, J. (1991). Regulation of polyadenylation of hepatitis delta virus antigenomic RNA. *Journal of virology*, *65*(12), 6438–6446.
- Huang, W.-H., Chen, Y.-S., & Chen, P.-J. (2008). Nucleolar targeting of hepatitis delta antigen abolishes its ability to initiate viral antigenomic RNA replication. *Journal of virology*, *82*(2), 692–699.

- Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome research*, 9(9), 868–877.
- Hughes, S. A., Wedemeyer, H., & Harrison, P. M. (2011). Hepatitis delta virus. *The Lancet*, 378(9785), 73–85.
- Hull, R., & Rima, B. (2020). Virus taxonomy and classification: naming of virus species. *Archives of Virology*, (pp. 1–4).
- Hwang, S., & Lai, M. (1993). Isoprenylation mediates direct protein-protein interactions between hepatitis large delta antigen and hepatitis B virus surface antigen. *Journal of virology*, 67(12), 7659–7662.
- ICTV Executive Committee and others (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology*, 5(5), 668.
- Jackson, K., MacLachlan, J., Cowie, B., Locarnini, S., Bowden, S., Higgins, N., Karapanagiotidis, T., Nicholson, S., & Littlejohn, M. (2018). Epidemiology and phylogenetic analysis of hepatitis D virus infection in Australia. *Internal medicine journal*, 48(11), 1308–1317.
- Jang, H. B., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., Brister, J. R., Kropinski, A. M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature biotechnology*, 37(6), 632–639.
- Jánová, E. (2019). Emerging and threatening vector-borne zoonoses in the world and in Europe: a brief update. *Pathogens and global health*, 113(2), 49–57.
- Jiang, H., Lei, R., Ding, S.-W., & Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics*, 15(1), 182.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240.
- Jung, S., Altstetter, S. M., Wilsch, F., Shein, M., Schütz, A. K., & Protzer, U. (2020). Extracellular vesicles derived from Hepatitis-D Virus infected cells induce a proinflammatory cytokine response in human peripheral blood mononuclear cells and macrophages. *Matters*, 6(2), e202002000001.
- Junglen, S., & Drosten, C. (2013). Virus discovery and recent insights into virus diversity in arthropods. *Current Opinion in Microbiology*, 16(4), 507–513.
- Käfer, S., Paraskevopoulou, S., Zirkel, F., Wieseke, N., Donath, A., Petersen, M., Jones, T. C., Liu, S., Zhou, X., Middendorf, M., Junglen, S., Misof, B., & Drosten, C. (2019). Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS pathogens*, 15(12), e1008224.
- Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology*, 338(5), 1027–1036.

- Karplus, K., Barrett, C., & Hughey, R. (1998). Hidden Markov models for detecting remote protein homologies. *Bioinformatics (Oxford, England)*, *14*(10), 846–856.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772–780.
- Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology*, *479*, 2–25.
- Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Kuhn, J. H. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiology and Molecular Biology Reviews*, *84*(2).
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455.
- Kuhn, J. H. (2020). Virus taxonomy. *Reference Module in Life Sciences*.
- Kuivanen, S., Levanov, L., Kareinen, L., Sironen, T., Jääskeläinen, A. J., Plyusnin, I., Zakham, F., Emmerich, P., Schmidt-Chanasit, J., Hepojoki, J., et al. (2019). Detection of novel tick-borne pathogen, Alongshan virus, in *Ixodes ricinus* ticks, south-eastern Finland, 2019. *Eurosurveillance*, *24*(27), 1900394.
- Kuo, M., Sharmeen, L., Dinter-Gottlieb, G., & Taylor, J. (1988). Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *Journal of virology*, *62*(12), 4439–4444.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.
- Lazinski, D. W., & Taylor, J. M. (1993). Relating structure to function in the hepatitis delta virus antigen. *Journal of virology*, *67*(5), 2672–2680.
- Lee, C.-H., Chang, S. C., Wu, C. H., & Chang, M.-F. (2001). A novel chromosome region maintenance 1-independent nuclear export signal of the large form of hepatitis delta antigen that is required for the viral assembly. *Journal of Biological Chemistry*, *276*(11), 8142–8148.
- Lee, C.-Z., Lin, J., Chao, M., McKnight, K., & Lai, M. (1993). RNA-binding activity of hepatitis delta antigen involves two arginine-rich motifs and is required for hepatitis delta virus RNA replication. *Journal of virology*, *67*(4), 2221–2227.
- Levins, R., & Lewontin, R. C. (1985). *The dialectical biologist*. Harvard University Press.
- Li, C.-X., Shi, M., Tian, J.-H., Lin, X.-D., Kang, Y.-J., Chen, L.-J., Qin, X.-C., Xu, J., Holmes, E. C., & Zhang, Y.-Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *elife*, *4*, e05378.

- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Li, Y.-J., Macnaughton, T., Gao, L., & Lai, M. M. (2006). RNA-templated replication of hepatitis delta virus: genomic and antigenomic RNAs associate with different nuclear bodies. *Journal of virology*, 80(13), 6478–6486.
- Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B., et al. (2012). Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-Bruijn-graph. *Briefings in functional genomics*, 11(1), 25–37.
- Littlejohn, M., Locarnini, S., & Yuen, L. (2016). Origins and evolution of hepatitis B virus and hepatitis D virus. *Cold Spring Harbor perspectives in medicine*, 6(1), a021360.
- Lowen, A. C. (2018). It's in the mix: Reassortment of segmented viral genomes. *PLoS pathogens*, 14(9), e1007200.
- Lucifora, J., & Delphin, M. (2020). Current knowledge on Hepatitis Delta Virus replication. *Antiviral Research*, (p. 104812).
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience*, 1(1), 2047–217X.
- Lutterkort, G., Wranke, A., Hengst, J., Yurdaydin, C., Stift, J., Bremer, B., Hardtke, S., Keskin, O., Idilman, R., Manns, M., et al. (2018). Viral dominance patterns in chronic hepatitis delta determine early response to interferon alpha therapy. *Journal of Viral Hepatitis*, 25(11), 1384–1394.
- MacNaughton, T. B., Gowans, E. J., McNamara, S. P., & Burrell, C. J. (1991). Hepatitis δ antigen is necessary for access of hepatitis s virus RNA to the cell transcriptional machinery but is not part of the transcriptional complex. *Virology*, 184(1), 387–390.
- Macnaughton, T. B., & Lai, M. M. (2002). Genomic but not antigenomic hepatitis delta virus RNA is preferentially exported from the nucleus immediately after synthesis and processing. *Journal of virology*, 76(8), 3928–3935.
- Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963.
- Mande, S. S., Mohammed, M. H., & Ghosh, T. S. (2012). Classification of metagenomic sequences: methods and challenges. *Briefings in bioinformatics*, 13(6), 669–681.

- Marklewitz, M., Paraskevopoulou, S., Alkhovskiy, S. V., Avsic-Zupanc, T., Bente, D. A., Bergeron, E., Burt, F. J., Ergunay, K., Garrison, A. R., Hewson, R., Mirazimi, A., Palacios, G., Papa, A., Paweska, J. T., Sall, A. A., Sprengler, J. R., Di Paola, N., & Kuhn, J. H. (2020a). Create four new genera and 30 new species (*Bunyavirales: Nairoviridae*). Tech. rep., ICTV taxonomic report 2020.027M.
- Marklewitz, M., Paraskevopoulou, S., Briese, T., Charrel, R., Choi, I.-R., de Lamballerie, X., Ebihara, H., Fu Gao, G., Groschup, M. H., Johnson, G., Nunes, M., Palacios, G., Sasaya, T., Shirako, Y., Song, J.-W., Wei, T., Zerbini, F. M., Zhou, X., & Kuhn, J. H. (2020b). Create one new genus and 16 new species (*Bunyavirales: Phenuiviridae*). Tech. rep., ICTV taxonomic report 2020.029M.
- Martick, M., Horan, L. H., Noller, H. F., & Scott, W. G. (2008). A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*, *454*(7206), 899–902.
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*(6210), 763–767.
- Modahl, L. E., Macnaughton, T. B., Zhu, N., Johnson, D. L., & Lai, M. M. (2000). RNA-dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Molecular and cellular biology*, *20*(16), 6030–6039.
- Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current opinion in virology*, *2*(1), 63–77.
- Negro, F., Korba, B., Forzani, B., Baroudy, B., Brown, T., Gerin, J., & Ponzetto, A. (1989). Hepatitis delta virus (HDV) and woodchuck hepatitis virus (WHV) nucleic acids in tissues of HDV-infected chronic WHV carrier woodchucks. *Journal of virology*, *63*(4), 1612–1618.
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., & Koopmans, M. P. (2018). Overview of virus metagenomic classification methods and their biological applications. *Frontiers in microbiology*, *9*, 749.
- Nouri, S., Matsumura, E. E., Kuo, Y.-W., & Falk, B. W. (2018). Insect-specific viruses: from discovery to potential translational applications. *Current opinion in virology*, *33*, 33–41.
- Otto, J. C., & Casey, P. J. (1996). The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells. *Journal of Biological Chemistry*, *271*(9), 4569–4572.
- Paraskevopoulou, S., Käfer, S., Zirkel, F., Donath, A., Petersen, M., Liu, S., Zhou, X., Drosten, C., Misof, B., & Junglen, S. (unpublished). Viromics of extant insect orders unveil the evolution of the flavi-like superfamily. *Submitted to Virus Evolution*.
- Paraskevopoulou, S., Pirzer, F., Goldmann, N., Schmid, J., Corman, V. M., Gottula, L. T., Schroeder, S., Rasche, A., Muth, D., Drexler, J. F., et al. (2020). Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*. *Proceedings of the National Academy of Sciences*, *117*(30), 17977–17983.

- Perez-Vargas, J., Amirache, F., Boson, B., Mialon, C., Freitas, N., Sureau, C., Fusil, F., & Cosset, F.-L. (2019). Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nature communications*, *10*(1), 1–15.
- Pischedda, E., Scolari, F., Valerio, F., Carballar-Lejarazú, R., Catapano, P. L., Waterhouse, R. M., & Bonizzoni, M. (2019). Insights into an unexplored component of the mosquito repeatome: distribution and variability of viral sequences integrated into the genome of the arboviral vector *Aedes albopictus*. *Frontiers in genetics*, *10*, 93.
- Poole, T. L., Wang, C., Popp, R., Potgieter, L., Siddiqui, A., & Collett, M. S. (1995). Pestivirus translation initiation occurs by internal ribosome entry. *Virology*, *206*(1), 750–754.
- Porter, A. F., Shi, M., Eden, J.-S., Zhang, Y.-Z., & Holmes, E. C. (2019). Diversity and evolution of novel invertebrate DNA viruses revealed by meta-transcriptomics. *Viruses*, *11*(12), 1092.
- Reid, C. E., & Lazinski, D. W. (2000). A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proceedings of the National Academy of Sciences*, *97*(1), 424–429.
- Rizzetto, M., Canese, M. G., Arico, S., Crivelli, O., Trepo, C., Bonino, F., & Verme, G. (1977). Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut*, *18*(12), 997–1003.
- Rizzetto, M., Verme, G., Recchia, S., Bonino, F., Farci, P., Aricò, S., Calzia, R., Picciotto, A., Colombo, M., & Popper, H. (1983). Chronic hepatitis in carriers of hepatitis B surface antigen, with intrahepatic expression of the delta antigen: an active and progressive disease unresponsive to immunosuppressive treatment. *Annals of Internal Medicine*, *98*(4), 437–441.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, *61*(3), 539–542.
- Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M., & Varsani, A. (2012). Diverse circular ssDNA viruses discovered in dragonflies (odonata: Eiprocta). *Journal of General Virology*, *93*(12), 2668–2681.
- Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Lavigne, R., Brister, J. R., Varsani, A., et al. (2019). Minimum information about an uncultivated virus genome (MIUViG). *Nature biotechnology*, *37*(1), 29–37.
- Ryu, W.-S., Netter, H. J., Bayer, M., & Taylor, J. (1993). Ribonucleoprotein complexes of hepatitis delta virus. *Journal of virology*, *67*(6), 3281–3287.
- Salehi-Ashtiani, K., Lupták, A., Litovchick, A., & Szostak, J. W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, *313*(5794), 1788–1792.

- Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *Journal of virology*, *84*(19), 9733–9748.
- Schmid, J., Rasche, A., Eibner, G., Jeworowski, L., Page, R. A., Corman, V. M., Drosten, C., & Sommer, S. (2018). Ecological drivers of *Hepacivirus* infection in a neotropical rodent inhabiting landscapes with various degrees of human environmental change. *Oecologia*, *188*(1), 289–302.
- Sharmeen, L., Kuo, M., & Taylor, J. (1989). Self-ligating RNA sequences on the antigenome of human hepatitis delta virus. *Journal of virology*, *63*(3), 1428–1430.
- Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., et al. (2016a). Redefining the invertebrate RNA virosphere. *Nature*, *540*(7634), 539–543.
- Shi, M., Lin, X.-D., Vasilakis, N., Tian, J.-H., Li, C.-X., Chen, L.-J., Eastwood, G., Diao, X.-N., Chen, M.-H., Chen, X., et al. (2016b). Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *Journal of virology*, *90*(2), 659–669.
- Siddell, S. G., Walker, P. J., Lefkowitz, E. J., Mushegian, A. R., Dutilh, B. E., Harrach, B., Harrison, R. L., Junglen, S., Knowles, N. J., Kropinski, A. M., et al. (2020). Binomial nomenclature for virus species: a consultation. *Archives of virology*, *165*(2), 519–525.
- Sigrist, C. J., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., & Xenarios, I. (2012). New and continuing developments at PROSITE. *Nucleic acids research*, *41*(D1), D344–D347.
- Sikora, D., Greco-Stewart, V. S., Miron, P., & Pelchat, M. (2009). The hepatitis delta virus RNA genome interacts with eEF1A1, p54nrb, hnRNP-L, GAPDH and ASF/SF2. *Virology*, *390*(1), 71–78.
- Simmonds, P. (2018). A clash of ideas—the varying uses of the ‘species’ term in virology and their utility for classifying viruses in metagenomic datasets. *Journal of General Virology*, *99*(3), 277–287.
- Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, *15*(3), 161–168.
- Smedile, A., Verme, G., Cargnel, A., Dentico, P., Opolon, P., Vergani, D., Farci, P., Caredda, F., Caporaso, N., Trepo, C., et al. (1982). Influence of delta infection on severity of hepatitis B. *The Lancet*, *320*(8305), 945–947.
- Song, L., Florea, L., & Langmead, B. (2014). Lighter: fast and memory-efficient sequencing error correction without counting. *Genome biology*, *15*(11), 509.
- Stockdale, A. J., Kreuels, B., Henrion, M. Y., Giorgi, E., Kyomuhangi, I., de Martel, C., Hutin, Y., & Geretti, A. M. (2020). The global prevalence of hepatitis D virus infection: systematic review and meta-analysis. *Journal of Hepatology*.

- Szirovicza, L., Hetzel, U., Kipar, A., Martinez-Sobrido, L., Vapalahti, O., & Hepojoki, J. (2020). Snake deltavirus utilizes envelope proteins of different viruses to generate infectious particles. *Mbio*, *11*(2).
- Tavanez, J. P., Cunha, C., Silva, M. C., David, E., Monjardino, J., & Carmo-Fonseca, M. (2002). Hepatitis delta virus ribonucleoproteins shuttle between the nucleus and the cytoplasm. *Rna*, *8*(5), 637–646.
- Taylor, J., & Pelchat, M. (2010). Origin of hepatitis δ virus. *Future microbiology*, *5*(3), 393–402.
- Taylor, J. M. (1999). Replication of human hepatitis delta virus: influence of studies on subviral plant pathogens. *Advances in virus research*, *54*, 45–60.
- Taylor, J. M. (2014). Host RNA circles and the origin of hepatitis delta virus. *World Journal of Gastroenterology: WJG*, *20*(11), 2971.
- Temmam, S., Bigot, T., Chrétien, D., Gondard, M., Pérot, P., Pommelet, V., Dufour, E., Petres, S., Devillers, E., Hoem, T., et al. (2019). Insights into the host range, genetic diversity, and geographical distribution of jingmenviruses. *MSphere*, *4*(6).
- Thurner, C., Witwer, C., Hofacker, I. L., & Stadler, P. F. (2004). Conserved RNA secondary structures in *Flaviviridae* genomes. *Journal of General Virology*, *85*(5), 1113–1124.
- Tsukiyama-Kohara, K., Iizuka, N., Kohara, M., & Nomoto, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. *Journal of virology*, *66*(3), 1476–1483.
- Vasilakis, N., & Tesh, R. B. (2015). Insect-specific viruses and their potential impact on arbovirus transmission. *Current opinion in virology*, *15*, 69–74.
- Vieira, D. S., Souza, L. F. B., Vasconcelos, M. P. A., Santos, A. d. O. d., Salcedo, J. M. V., et al. (2017). Hepatitis delta: virological and clinical aspects. *Virology Journal*, *14*(177).
- von Linné, C. (1735). *Systema naturae; sive, Regna tria naturae: systematice proposita per classes, ordines, genera & species*. Haak.
- Walker, P., Blasdel, K. R., Dietzgen, R. G., Freitas-Astúa, J., Kondo, H., Kurath, G., Kuzmin, I. V., Tesh, R. B., Tordo, N., Vasilakis, N., & Whitfield, A. E. (2020). Create seven new genera (*Alphacrustrhavirus*, *Alphadrosrhavirus*, *Alphahymrhavirus*, *Betahymrhavirus*, *Betanemrhavirus*, *Betapaprhavirus*, and *Betaricinhavirus*), including 16 new species (*Mononegavirales: Rhabdoviridae*). Tech. rep., ICTV taxonomic report 2020.023M.
- Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Dempsey, D. M., Dutilh, B. E., Harrach, B., Harrison, R. L., Hendrickson, R. C., Junglen, S., et al. (2019). Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Archives of virology*, *164*(9), 2417–2429.

- Walter, C. T., Pringle, F. M., Nakayinga, R., de Felipe, P., Ryan, M. D., Ball, L. A., & Dorrington, R. A. (2010). Genome organization and translation products of Providence virus: insight into a unique tetravirus. *Journal of general virology*, *91*(11), 2826–2835.
- Wang, K.-S., Choo, Q.-L., Weiner, A. J., Ou, J.-H., Najarian, R. C., Thayer, R. M., Mullenbach, G. T., Denniston, K. J., Gerin, J. L., & Houghton, M. (1986). Structure, sequence and expression of the hepatitis delta (δ) viral genome. *Nature*, *323*(6088), 508–514.
- Wang, Z.-D., Wang, B., Wei, F., Han, S.-Z., Zhang, L., Yang, Z.-T., Yan, Y., Lv, X.-L., Li, L., Wang, S.-C., et al. (2019). A new segmented virus associated with human febrile illness in China. *New England Journal of Medicine*, *380*(22), 2116–2125.
- Webb, C.-H. T., & Lupták, A. (2011). HDV-like self-cleaving ribozymes. *RNA biology*, *8*(5), 719–727.
- Wille, M., Netter, H. J., Littlejohn, M., Yuen, L., Shi, M., Eden, J.-S., Klaassen, M., Holmes, E. C., & Hurt, A. C. (2018). A divergent hepatitis D-like agent in birds. *Viruses*, *10*(12), 720.
- Williams, V., Brichtler, S., Radjef, N., Lebon, P., Goffard, A., Hober, D., Fagard, R., Kremsdorf, D., Deny, P., & Gordien, E. (2009). Hepatitis delta virus proteins repress hepatitis B virus enhancers and activate the alpha/beta interferon-inducible MxA gene. *Journal of general virology*, *90*(11), 2759–2767.
- Wolf, Y., Krupovic, M., Zhang, Y. Z., Maes, P., Dolja, V., Koonin, E. V., & Kuhn, J. H. (2018a). ICTV taxonomic report 2017.006M: Megataxonomy of negative-sense RNA viruses. *International Committee on Taxonomy of Viruses*.
- Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., & Koonin, E. V. (2018b). Origins and evolution of the global RNA virome. *MBio*, *9*(6).
- Wong, S. K., & Lazinski, D. W. (2002). Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proceedings of the National Academy of Sciences*, *99*(23), 15118–15123.
- Xia, Y., Yeh, C.-T., Ou, J.-H., & Lai, M. (1992). Characterization of nuclear targeting signal of hepatitis delta antigen: nuclear transport as a protein complex. *Journal of virology*, *66*(2), 914–921.
- Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M. M., & Handa, H. (2001). Stimulation of RNA polymerase II elongation by hepatitis delta antigen. *Science*, *293*(5527), 124–127.
- Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S., & Handa, H. (2007). Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. *Genes to Cells*, *12*(7), 863–875.
- Zachou, K., Yurdaydin, C., Drebber, U., Dalekos, G. N., Erhardt, A., Cakaloglu, Y., Degertekin, H., Gurel, S., Zeuzem, S., Bozkaya, H., et al. (2010). Quantitative HBsAg and HDV-RNA levels in chronic delta hepatitis. *Liver International*, *30*(3), 430–437.

- Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de bruijn graphs. *Genome research*, *18*(5), 821–829.
- Zhou, X., Shen, X.-X., Hittinger, C. T., & Rokas, A. (2018). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Molecular biology and evolution*, *35*(2), 486–503.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), 3406–3415.

Appendix

Table A1. List of viruses reported in Käfer et al. (2019) and Paraskevopoulou et al. (2020) proposed to be officially classified according to the latest ICTV taxonomic proposals (Walker et al., 2020; Di Paola et al., 2020; Dietzgen et al., 2020; Ballinger et al., 2020; Marklewitz et al., 2020a,b; Hepojoki et al., 2020). Asterisks mark species that were assigned as type species; table continues in next page.

Order	Family	Genus	Species	Virus name as reported in Käfer et al. (2019)	
<i>Mononegavirales</i>	<i>Rhabdoviridae</i>	<i>Alphahymrhavirus</i>	<i>Cinereus alphahymrhavirus</i>	Hymenopteran rhabdo-related virus OKIAV38	
			<i>Radians alphahymrhavirus</i>	Hymenopteran rhabdo-related virus OKIAV46	
			<i>Hirtum alphahymrhavirus</i>	Hymenopteran rhabdo-related virus OKIAV109	
		<i>Betahymrhavirus</i>	<i>Austriaca betahymrhavirus</i> *	Hymenopteran rhabdo-related virus OKIAV23	
			<i>Heterodontonyx betahymrhavirus</i>	Hymenopteran rhabdo-related virus OKIAV24	
		<i>Betapapravirus</i>	<i>Sylvina betapapravirus</i>	Lepidopteran rhabdo-related virus OKIAV34	
<i>Jingchuvirales</i>	<i>Crepuscuviridae</i>	<i>Aqualaruvirus</i>	<i>Aqualaruvirus sialis</i> *	Megalopteran chu-related virus OKIAV119	
			<i>Chuviridae</i>	<i>Dermapteravirus</i>	<i>Dermapteravirus dermapteri</i> *
	<i>Odonatavirus</i>	<i>Odonatavirus odontis</i>			Odonatan chu-related virus OKIAV136
		<i>Odonatavirus draconis</i>			Odonatan chu-related virus OKIAV137
	<i>Pterovirus</i>	<i>Pterovirus chulinense</i> *	Hymenopteran chu-related virus OKIAV147		
	<i>Alliusviridae</i>	<i>Ollusvirus</i>	<i>Ollusvirus hymenopteri</i>	Hymenopteran chu-related virus OKIAV123	
			<i>Ollusvirus insectii</i>	Hymenopteran chu-related virus OKIAV126	
<i>Bunyavirales</i>	<i>Nyamiviridae</i>	<i>Formivirus</i>	<i>Gorytes formivirus</i>	Hymenopteran orino-related virus OKIAV85	
			<i>Chalybion formivirus</i>	Hymenopteran orino-related virus OKIAV87	

Table A1: Continues from previous page.

Order	Family	Genus	Species	Virus name as reported in Käfer et al. (2019)
<i>Bunyvirales</i>	<i>Phasmaviridae</i>	<i>Hymovirus</i>	<i>Hymenopteran hymovirus 1</i> *	Hymenopteran phasma-related virus OKIAV252
			<i>Hymenopteran hymovirus 2</i>	Hymenopteran phasma-related virus OKIAV250
		<i>Feravirus</i>	<i>Hemipteran feravirus</i>	Hemipteran phasma-related virus OKIAV247
		<i>Orthophasmavirus</i>	<i>Coleopteran orthophasmavirus</i>	Coleopteran phasma-related virus OKIAV235
	<i>Hymenopteran orthophasmavirus 1</i>		Hymenopteran phasma-related virus OKIAV228	
	<i>Phasmaviridae</i>	<i>Orthophasmavirus</i>	<i>Hymenopteran orthophasmavirus 2</i>	Hymenopteran phasma-related virus OKIAV227
	<i>Phenuiviridae</i>	<i>Pidchovirus</i>	<i>Coleopteran pidchovirus</i>	Coleopteran phenui-related virus OKIAV308
<i>Nairoviridae</i>	<i>Ocetevirus</i>	<i>Blattodean ocetevirus</i> *	Blattodean nario-related virus OKIAV321	
<hr/>				
Realm	Family	Genus	Species	Virus name as reported in Paraskevopoulou et al. (2020)
<i>Ribozyviria</i>	<i>Kolmioviridae</i>	<i>Thuriszavirus</i>	<i>Thuriszavirus myis</i> *	Rodent deltavirus

List of abbreviations

A		M	
ADAR1	<i>adenosine deaminase acting on RNA 1</i>		
AG-RNA	antigenomic RNA	N	
ARM	arginine-rich motif	NCBI	National Center for Biotechnology Information
B		NELF	negative elongation factor
BLAST	basic local alignment search tool	NES	nuclear export signal
C		NGS	next generation sequencing
CCD	coiled-coil domain	NLS	nuclear localization signal
CLEC2	C-type lectin type II	O	
CPE	cytopathic effect	OLC	overlap layout consensus
CPEB3	cytoplasmic polyadenylation element-binding protein 3	ORF	open reading frame
CRM1	chromosome region maintainance 1	P	
D		PCR	polymerase chain reaction
DBG	de Bruijn graph	pHMM	profile hidden Markov model
DIPA	<i>delta interacting protein A</i>	Pol-I	DNA polymerase I
E		Pol-II	RNA polymerase II
F		Pol-III	RNA polymerase III
FTase	<i>farnesyltransferase</i>	Q	
G		R	
G-RNA	genomic RNA	RDeV	rodent deltavirus
H		RdRp	RNA-dependent RNA polymerase
HBV	Hepatitis B virus	RNP	ribonucleoprotein
HBeAg	hepatitis B e-antigen	RT	reverse-transcriptase
HBsAg	hepatitis B surface antigen	RT-PCR	reverse transcription polymerase chain reaction
HDAg	hepatitis delta antigen	S	
HDV	Hepatitis delta virus	S-HDAg	small hepatitis delta antigen
HVT	horizontal virus transfer	S-RDeAg	small rodent deltavirus antigen
I		S-SDeAg	small snake deltavirus antigen
ICTV	International Committee on Taxonomy of Viruses	T	
IF	immunofluorescence assay	U	
IRES	internal ribosome entry site	V	
ISV	insect-specific virus	W	
J		WHV	<i>Woodchuck hepatitis virus</i>
K		X	
L		Y	
L-HDAg	large hepatitis delta antigen	Z	

Appendix of Chapter 2

Supplementary Material for the above publication published as: Käfer, S., Paraskevopoulou, S., Zirkel, F., Wieseke, N., Donath, A., Petersen, M., Jones, T.C., Liu, S., Zhou, X., Middendorf, M., Junglen, S., Misof, B., and Drosten C. (2019). **Re-assessing the diversity of negative strand RNA viruses in insects.** *PLoS Pathogens*, 15(12):e1008224.

The Supplementary Material is also available at: <https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008224#sec016>.

Supplementary Material includes:

- **S1 Text:** Supplementary results and discussion text.
- **S1 - S4 Tables.** Table headers are found in **Chapter 2**, page 19.
- **S1 - S34 Figures.** Figure legends are found in **Chapter 2**, pages 19-28.

Note: **S1 Table** follows directly after **S1 Text**.

1 Supporting Information

2 3 S1 Text: Supplementary results and discussion text.

4 5 *Rhabdoviridae*

6 This family includes 18 genera (**S1 Fig**). The tree has a bifurcation from which mainly plant-
7 associated viruses, such as those assigned to the genera *Cytorhabdovirus* and
8 *Nucleorhabdovirus*, split to the exclusion of animal viruses, such as those assigned to the
9 genera *Sigmavirus* and *Lyssavirus*. The genus *Novirhabdovirus* is a phylogenetic outlier.

10
11 **S1 Fig** shows 33 selected novel rhabdovirus sequences from the present study that do not
12 cluster with members of the established genera. A comparatively small number of viruses are
13 grouped with dimarhabdoviruses, which is a historical classification for a now paraphyletic
14 group of viruses found in diptera and mammalia. Full genomes of novel dimarhabdovirus-
15 related viruses (Lepidopteran rhabdo-related virus OKIAV12 and Hymenopteran rhabdo-
16 related virus OKIAV8) suggest a typical rhabdovirus gene order comprising nucleoprotein -
17 phosphoprotein - matrix protein - glycoprotein - large protein (including RdRp, **Fig 3**).

18 A particularly interesting finding is Hymenopteran almendra-related virus OKIAV1, a full
19 genome with typical rhabdovirus organization (**Fig 3**). It is positioned in sister relationship to
20 all members of *Almendravirus*, a recently established genus [1]. The type isolates stem from
21 mosquitoes (order Diptera), while Hymenopteran almendra-related virus OKIAV1 is from a
22 hymenopteran species, thus pertaining to a sister order to Diptera. Almendraviruses have a
23 viroporin protein encoded between the G and L genes [1]; Hymenopteran almendra-related
24 virus OKIAV1 has a region of hydrophobic residues, as in viroporins, but none of the
25 remaining salient features and no start codon (**S4 Fig**). Degeneration of a viroporin gene
26 after divergence from a common ancestor with almendraviruses seems possible.

27 Hymenopteran almendra-related virus OKIAV1 and almendraviruses provide one of the
28 many examples of putative virus-host co-segregation in negavirus evolution.

29
30 A multitude of novel rhabdovirus sequences branch in sister relationship to a clade formed by
31 cyto-, nucleo-, dichorhabdo-, as well as varicosaviruses (**S1 Fig**). The clades shown are well-
32 supported but different phylogenetic algorithms yield alternative topologies. We therefore
33 conducted additional phylogenetic studies focused only on this clade, as shown in **S5 Fig**.
34 The resulting tree suggests the need for at least eight distinct taxa comprising 14 full or
35 partial sequences recently identified in metagenomic studies [2, 3], as well as 12 full and 14
36 partial genome sequences from the present study. All have typical rhabdovirus genome
37 architectures. Five of these tentative genera are defined by the present study (**Table 1**).

38

39 ***Xinmoviridae***

40 The novel family *Xinmoviridae* comprises only the formerly unassigned genus *Anphevirus*.
41 We find 13 OKIAVs related to this family (**S6 Fig**). Five sequences represent complete
42 genomes (**Fig 3**). These sequences stem from a broad range of hosts, belonging to 13
43 orders of insects (Blattodea, Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera,
44 Mantodea, Neuroptera, Odonata, Orthoptera, Psocodea, Trichoptera, Zygentoma; consult **S1**
45 **Table** for geographic origins). While the family is not formally subdivided into genera, a deep
46 topological separation and structured host association emerges after addition of data from
47 the present study (**S6 Fig**). To orient future taxonomic classification, we have assigned six
48 different *Anphevirus* lineages, two of which contain the sequences previously detected in a
49 mosquito and a fly [2], as well as in pooled insects [3]. Two clades can be considered to
50 define potential novel genera based on full genome data from the present study (**Table 1**).
51 Lineage I appears to be associated with Hymenoptera, Hemiptera, and Diptera, while lineage
52 III is specific for Odonata. Lineage I contains endogenous viral elements of mosquitos but is
53 represented by full genomes from other insect orders. Together with the recent reconciliation
54 of insect phylogeny that incorporates molecular data as well as the fossil record [4], this
55 provides an interesting scenario for studies of viral germline integration.

56

57 ***Nyamiviridae*** is a newly-defined family that was recently extended to contain six genera [5].
58 In the genus *Nyavirus*, the type species *Nyamanini-* and *Midway nyavirus* are both known to
59 infect birds and ticks, causing phenotypic changes typical of arboviruses [6, 7]. *Sierra*
60 *Nevada nyavirus* is another virus that was recently assigned to *Nyavirus*, and is known to
61 infect ticks [8]. In the genus *Socyvirus*, the only known member is *Soybean cyst nematode*
62 *socyvirus* that replicates and is detected in at least two developmental stages of nematodes
63 [9]. No biological knowledge is available for any member of the genera *Tapwovirus*,
64 *Orinovirus*, *Crustavirus*, and *Berhavirus*, which are derived from metagenomic studies [3, 10].
65 As shown in **S6 Fig**, our data contribute four novel viruses related to the monospecific genus
66 *Orinovirus*, two of them with full genomes. Lepidopteran orino-related virus OKIAV84 is the
67 closest relative to *Orinoco orinovirus* and stems from an insect of the same order
68 (Lepidoptera, although from a different family) [4]. Hymenopteran orino-related virus
69 OKIAV85, -87, and -88 form a monophyletic clade and are associated with Hymenoptera. A
70 very recent sequence description from ants (*Formica fusca virus 1*) also falls in this clade,
71 providing independent confirmation of association with Hymenoptera [10]. The deep
72 divergence and consistent association with a specific insect order suggests this clade may
73 constitute a separate genus (termed “*Orinovirus* lineage 1” in **S6 Fig**). In that case, only

74 Lepidopteran orino-related virus OKIAV84 would group with the genus *Orinovirus* (termed
75 “*Orinovirus* lineage II” in **S6 Fig**).

76

77 The family ***Bornaviridae*** has three established genera (*Orthobornavirus*, *Carbovirus*, and
78 *Cultervirus*) that occur in mammals, birds, fish, and reptiles. The absence of bornaviruses in
79 our extensive insect dataset is compatible with a divergence of these viruses in primordial
80 vertebrates that pre-dated extant reptiles and mammals.

81

82 The family ***Artoviridae*** only contains one genus, *Peropuvirus*. The type species is
83 *Pteromalus puparum peropuvirus*, which is vertically transmitted in parasitoid wasps and
84 regulates the sex ratio by decreasing the number of female offspring [11]. Our data add three
85 novel viruses associated with three orders of insects (Diptera, Hymenoptera, Trichoptera).
86 The closest relative to *Pteromalus puparum peropuvirus*, Hymenopteran arto-related virus
87 OKIAV1338, also stems from a parasitoid wasp. The other host associations suggest that
88 further subdivision of the family *Artoviridae* should be considered. Based on host order
89 associations, for instance, phylogeny suggests the existence of three additional genera in
90 insects. However, classification is impossible at this point because of the lack of full genome
91 sequences.

92

93 The family ***Lispiviridae*** contains the formerly unassigned genera *Arlivirus*, *Chengtivirus*, and
94 *Wastrivirus*, detected in a tick, a spider, and a water strider in China [2]. The International
95 Committee on Taxonomy of Viruses (ICTV) established and subsequently withdrew a part of
96 these genera within the last two years [5, 12]. The novel family *Lispiviridae* is now
97 monogeneric and defined by six type species that have all been associated with the genus
98 *Arlivirus*. The present study adds viruses from 11 orders of insects sampled worldwide
99 (Archaeognatha, Blattodea, Coleoptera, Diptera, Hemiptera, Hymenoptera, Isoptera,
100 Megaloptera, Neuroptera, Odonata, and Strepsiptera). As shown in the phylogenetic tree in
101 **S6 Fig**, we have identified 13 viral sequences, four of which have complete genomes (**Fig 3**).
102 The host associations of insect viruses in the present data suggest a criterion for separating
103 this family into nine distinct genetic lineages at the genus level. Lineage III is specific for
104 Hymenoptera, lineage IV for Hemiptera, and lineage VII for Odonata. Four full genomes of
105 lispiviruses were obtained. Due to the phylogenetic distance, genome annotation is
106 challenging, and it remains unknown which proteins other than G and L are encoded [10, 13].
107 Availability of full genomes justifies formal consideration of two of these lineages as newly
108 discovered putative genera (lineages III, IV in **S6 Fig, Table 1**). Interestingly, the
109 hymenopteran association of lineage III was very recently confirmed by a sequence termed
110 *Linepithema humile rhabdo-likevirus 1*, found in ants [13].

111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147

Mymonaviridae

The fungal mononegavirus *Sclerotinia sclerotimonavirus* is the type species of the monogeneric family *Mymonaviridae* (genus *Sclerotimonavirus* [14]). With the exception of *Hubei sclerotimonavirus*, found in a mix of arthropods from the Hubei province in China [3], all other mymonaviruses are fungal viruses. Our study does not identify any mymonavirus in insects.

Filo-, Paramyxo-, and Pneumoviridae

The families *Filoviridae* (genera *Ebolavirus*, *Marburgvirus*, *Cuevavirus*, *Striavirus*, *Thamnovirus*), *Paramyxoviridae* (subfamilies *Avula-*, *Metaparamyxo-*, *Orthparamyxo-*, and *Rubulavirinae*), *Pneumoviridae* (genera *Metapneumovirus* and *Orthopneumovirus*), and *Sunviridae* (genus *Sunshinevirus*) form a monophyletic clade of vertebrate-pathogenic virus families. While earlier studies in mammals and other vertebrates have contributed a significant extension of the diversity of *Paramyxo-* and *Pneumoviridae* [15, 16], the present study, covering an extremely broad host genetic space, does not yield any insect-associated members of these families.

Chuviridae

The family *Chuviridae* is a novel family of arthropod-associated viruses that has been ascribed as sole element of the novel class *Monjiviricetes* and the novel order *Jingchuvirales*. The only genus in the family, termed *Mivirus*, has 29 species. The genomes of chuviruses were found to appear in linear, circular, and segmented circular forms [2]. **S9 Fig** shows an additional 25 exemplary sequences from the present study, including seven with at least one full segment and one with two full segments. We have encountered 10 chuviruses with segmented genomes.

Most, or all, of the chuviral sequences discovered in the present study are diversified enough to define independent species. The addition of our data reveals a structure of three major clades within *Chuviridae* (A, B, and C in **S9 Fig**), each divided into monophyla associated with insects and lineages in other non-vertebrate hosts. The long-branching position occupied by viruses from non-insect hosts may be a result of under-sampling of these hosts. Four subclades show associations with members of the insect orders Odonata (OAM - Odonata-associated *Mivirus*), Hemiptera (HeAM - Hemiptera-associated *Mivirus*), Hymenoptera (HyAM - Hymenoptera-associated *Mivirus*), and Phasmatodea (PAM - Phasmatodea-associated *Mivirus*). These clades provide an assessment of the depth of a genetic unit of chuviruses that could be classified as genera (**Fig 3**).

148

149 Within the OAM clade, Odonatan chu-related virus OKIAV136 and -137 have coding-
150 complete genomes and group with *Odonate mivirus*. As well as the predicted proteins of
151 *Odonate mivirus*, the two OKIAV genomes show nucleoprotein-coding predictions for the
152 third ORF (**Fig 3**). Beneath the OAM clade, the sister taxon of Wenling chuvirus-like virus 1,
153 Dermapteran chu-related virus OKIAV142, holds an extra viral segment which is predicted to
154 encode a nucleoprotein. Dermapteran chu-related virus OKIAV142 encodes an RdRp and
155 glycoprotein on its other genomic segment. Other miviruses of the same clade have two
156 genomic segments but only encode RdRp and nucleoprotein. At least OAM (Odonatan chu-
157 related virus OKIAV136 and -137 with full genomes) and HyAM (Hymenopteran chu-related
158 virus OKIAV123, -124, and -126 with full genomes) should be considered as representatives
159 of novel genera.

160

161 ***Qinviridae***

162 The family *Qinviridae* is a group of novel arthropod-associated viruses that has been
163 ascribed as sole member of the novel order *Muvirales* and to the novel class
164 *Chunqiuviricetes*. The only genus in the family, termed *Yingvirus*, has eight species. Two
165 species, *Wuhan yingvirus* and *Hubei yingvirus*, occur in insects while other yingviruses occur
166 in crustaceans, helminths, and lophotrochozoa. Our study provides the first independent
167 confirmation of the existence of *Qinviridae* by adding the longest so-far known qinivirus
168 sequence (Collembolan qin-related virus OKIAV112) that stems from the hexapod *Anurida*
169 *maritima* (seashore springtail). It is equidistant from the aforementioned taxa and therefore
170 likely constitutes a separate genus. Its genome organization agrees with the proposed bi-
171 segmented genome organization of the non-insect *Qinviridae* (L and N-G, see **Fig 3**), and it
172 adds a protein prediction for the second segment of insect qinviruses.

173 The topology of the *Qinviridae* tree may reflect that of the corresponding hosts, as viruses in
174 sister relationship to insect-associated viruses are found in invertebrates (crustaceans,
175 helminths, and lophotrochozoans) and insect-associated viruses show a branching pattern
176 reflecting that of their insect hosts. This is also confirmed by the very recent description of
177 *Linepithema humile* qinivirus-like virus 1 [13], an incomplete qinivirus sequence found in ants.

178

179 ***Orthomyxoviridae***

180 The *Orthomyxoviridae* family contains seven genera: *Alpha-*, *Beta-*, *Gamma-*, and
181 *Deltainfluenzavirus*, *Isavirus*, *Thogotovirus*, and *Quaranjavirus*, which comprise nine type
182 species altogether. With the exception of the influenza viruses, orthomyxoviruses are
183 transmitted by arthropod vectors [17, 18]. Isaviruses infect fish and are considered to be
184 transmitted by sea lice [19]. Thogotoviruses and quaranjaviruses are transmitted to

185 mammals by ticks [20]. The genus *Quaranjavirus* contains viruses that appear to cycle
186 between soft ticks (family Argasidae) and aquatic birds [17]. All genera are summarized in
187 **S12 Fig** and genome organizations are shown in **Fig 3**. Not included in the figure is the
188 Tilapia virus (species *Tilapia tilapinevirus*), recently detected in tilapia aquacultures
189 (*Oreochromis sp.*) [21]. This virus was classified into the novel viral order *Articulavirales* and
190 the solitary monogeneric family *Amnoonviridae* (genus *Tilapinevirus*).

191
192 The segmentation of orthomyxovirus genomes makes it difficult to assemble full genome
193 sequences based on RNAseq studies. The orthomyxovirus RdRp is encoded on three
194 segments, PB1, PB2, and PA, of which PB1 contains the conserved motifs for our pHMM
195 search as well as phylogeny. Our study contributes 46 viruses with up to five segments.
196 Among those, 31 PB2 and 41 PA segments are found, thus providing a complete RdRp for
197 31 viral sequences. Only a minority of the viral sequences that we identified group together
198 with members of established genera. Most of the sequences from the present study fall in
199 previously undescribed clades. Long-branching viral taxa are mostly derived from members
200 of basal insect groups, or outgroup taxa including Zygentoma (silverfish), Myriapoda, and
201 Archaeognatha (jumping bristletails). Based on the criteria that genomes should at least
202 comprise five segments, topology of several phylogenetic analyses should agree, and
203 diversity should be comparable to that within established genera, we can define at least
204 seven clades that may determine novel putative genera (designated O1-O7 in **S12 Fig**). Five
205 of these are newly described in the present study and have five genome segments (**Table 1**).
206 The diversity and the number of assigned members of the genera *Thogotovirus* and
207 *Quaranjavirus* are also expanded by the addition of sequences from the present study. The
208 genome termini in seven novel viruses are partially sequenced and confirmed to be reverse
209 complementary. However, in the absence of live virus isolates, the completeness of
210 segmented genomes cannot be confirmed. A recently described virus (Sinu virus, [22]) is
211 related to one of the novel viruses, Neuropteran orthomyxo-related virus OKIAV210, but not
212 shown in the tree figure.

213
214 Phylogenies of each RdRp subunit were inferred separately and then compared to assess
215 topological congruence (**S15 Fig**). Overall, there is congruence between segments in major
216 clades. In some clades, such as clade C of **S15 Fig**, reassortment of at least some of the
217 segments seems to have occurred in lineage precursors, as topological incongruence is
218 observed for all members of the clade. In cases where individual incongruences are
219 suggested, we cannot discriminate between *in silico* misassembly of genomes and actual
220 reassortment based on the present data.

221

222 **Novel *Orthohantavirus*-related lineages**

223 Novel unclassified sequences related to hantaviruses have recently been detected in diverse
224 insects [3]. The unclassified clades in sister relationship to *Hantaviridae* are paraphyletic in
225 some of our analyses (**S16 Fig**) but monophyletic in others (**S17 Fig, S18 Fig**). Of all these
226 viruses, only one virus, termed Hubei myriapoda virus 6, has more than one genome
227 segment and compatible segment termini. None of the viruses have the three segments that
228 would be expected in hantaviruses and related bunyaviruses. Our data contribute the first full
229 L-gene sequence in the subclade that is most closely related to hantaviruses (Coleopteran
230 hanta-related virus OKIAV221). Comparison of genome termini identifies only minimal
231 agreement with genome termini of hantaviruses. This makes it likely that these viruses can
232 be classified in a new genus, once the first full genomes with overlapping segment ends
233 become available.

234

235 **Novel group in the order *Bunyvirales***

236 A diversified clade of bunyaviruses branches as outgroup to members of *Hantavirus* and
237 *Hantavirus*-related lineages, containing viruses associated with basal insect sequences and
238 insect outgroups (Plecoptera, Odonata, Diplura, Campodea, and Myriapoda, **S16 Fig**). Some
239 algorithms place it in sister relationship to *Peribunyaviridae*. Given the phylogenetic position
240 of this monophyletic clade, and its distance to the neighbouring clades, it could be
241 considered a new genus in the family *Peribunyaviridae* or a new family within the order
242 *Bunyvirales* (**Table 1**). The lengths of Myriapodan hanta-related virus OKIAV214 and
243 Plecopteran hanta-related virus OKIAV215 suggest detection of a full viral L-segment. An M-
244 segment was detected for either Dipluran hanta-related virus OKIAV217 or -218. However,
245 as these sequences stem from the same insect, it is unclear to which L-segment this M-
246 segment belongs. The possibility of double infection with different viruses is another reason
247 why segmented virus sequences should not be ascribed to genomes without experimental
248 proof of viral replication.

249

250 The two recently classified genera, *Actinovirus* and *Agnathovirus*, found in fish, were not
251 represented in any insect transcriptome in the present study [16, 23].

252

253 ***Cruliviridae***

254 A solitary virus genome from a mix of crustaceans has been the basis for the designation of
255 a novel viral family, *Cruliviridae* [3, 24]. The genome termini of the type species, *Crustacean*
256 *lincrivirus* (Wenling crustacean virus 9), are virtually identical to those of the genus
257 *Emaravirus* (family *Fimoviridae*). This suggests that bunyavirus genome termini can be
258 conserved even across families (AGUAGA(A/G)), therefore this is an insufficient criterion for

259 the definitive delineation of putatively novel genera [25, 26]. No cruliviruses are found in our
260 large sample of insects.

261

262 ***Peribunyaviridae***

263 The family *Peribunyaviridae* contains the genera *Orthobunyavirus*, *Herbevirus*, and
264 *Shangavirus* (see below). A recently-added genus *Pacuvirus* is very closely related to
265 *Orthobunyavirus* [16]. Orthobunyaviruses often infect vertebrates and are transmitted by
266 mosquitoes, midges, phlebotomines, and ticks. Despite the comprehensive sample of insects
267 tested, the present study identifies only two viruses found in Diptera (not included in **S5 Fig**),
268 and thus does not add any new hosts, in terms of insect orders, for these viruses.

269 Pacuviruses were also absent. The genus *Herbevirus* is a sister group to *Orthobunyavirus*.
270 Herbeviruses have been fully characterized in cell culture, and their replication properties
271 suggest a restriction to invertebrates [27]. The present data add novel sequences that clearly
272 belong to *Herbevirus* and are derived from hosts as distant as lacewings and bark lice. This
273 suggests that the dipteran-associated orthobunyaviruses may have evolved from a wider
274 peribunyaviral diversity distributed in precursors to a range of insect orders. According to the
275 present data, this range includes two groups basal to Diptera: Psocodea and Neuroptera. In
276 view of the absence of herbe-/orthobunyaviruses in other major groups of insects it should be
277 considered that tick-associated orthobunyaviruses may have been secondarily acquired from
278 viremic vertebrate hosts.

279

280 It is noteworthy that Khurdun virus, isolated in Russia from coots at high prevalence and
281 suggested to be an orthobunyavirus [28, 29], reliably groups with herbeviruses. The genome
282 termini do not provide additional information for classification, as they are highly similar
283 across orthobunya- and herbeviruses, including Khurdun virus. According to the present
284 analysis, Khurdun virus is closer to herbeviruses, and in particular Psocodean peribunya-
285 related virus OKIAV212 and Neuropteran peribunya-related virus OKIAV213, than to
286 orthobunyaviruses [28]. Potential vertebrate tropism of Psocodean peribunya-related virus
287 OKIAV212 and Neuropteran peribunya-related virus OKIAV213, or related viruses should be
288 studied. Khurdun virus may in fact belong to herbeviruses, which would then suggest that
289 herbeviruses may be able to infect vertebrates.

290

291 A recent ICTV report has designated a novel genus named *Shangavirus*, based on the
292 solitary genome sequence of *Insect shangavirus* [24]. The phylogenetic position of
293 *Shangavirus* is ambiguous, as shown by RdRp-based phylogenies. Some analyses place it
294 in sister relationship to *Herbevirus* including Khurdun virus (**S16 Fig**), while others place it in
295 sister relationship to *Herbe-* and *Orthobunyavirus* (**S16 Fig**, **S17 Fig**, **S18 Fig**). It should be

296 noted that full genome analysis, as shown in the latest ICTV taxonomy report (2017.012M,
297 [24]), is difficult to utilize as the combination of genome segments has not been confirmed by
298 a live virus isolate. Segments of orthobunya- and herbeviruses normally have identical or
299 highly similar reverse complementary ends within a viral genome, but this is not the case for
300 *Insect shangavirus*. The M-segment tree topology of *Shangavirus* is congruent with that of
301 the L-segment while congruence with the S-segment sequence (N-gene) is not supported
302 (**S19 Fig**). As the related *Herbevirus* is also congruent between L- and M-, but not between
303 L- and N-genes, this suggests that reassortment in lineage precursors to *Herbe-* and
304 *Shangavirus* may have been involved in initial speciation and separation of these genera.

305

306 **Relationships among Hanta-, Peribunya-, Tospo-, and Fimoviridae**

307 Collembolan phasma-related virus OKIAV223 is a large sequence of 8154 nucleotides,
308 extending beyond the L-gene-ORF, but its genome termini are not convincingly shown to
309 induce a reverse complementary panhandle structure. Further research is necessary to
310 define whether a novel bunyaviral family should be established based on this deeply-
311 branching sequence. Based on conserved motifs, the RdRp of Collembolan phasma-related
312 virus OKIAV223 clusters with members of *Phasmaviridae* and branches from close to the last
313 common ancestor of all members of *Peribunyaviridae* and *Phasmaviridae* (**S20 Fig**). If used
314 as an outgroup for the peribunyavirus tree (**S16 Fig**), this outgroup re-defines the internal
315 structure of the tree, which is also confirmed by analysis in MrBayes (alternative topologies
316 are shown in **S17 Fig** and **S18 Fig A, B**). Since tospoviruses and emaraviruses are plant-
317 associated viruses, two independent transitions from plants into metazoa may then have to
318 be considered as compared to the simpler scenario of *Peribunyaviridae*, but not
319 *Hantaviridae*, having evolved from plant viruses (**S16 Fig** and **S20 Fig**).

320

321 **Phasmaviridae**

322 The type species of this family were initially discovered in transcriptome sequences of
323 phantom midges [30], establishing the current genus *Orthophasmavirus*. The family
324 *Phasmaviridae* has recently been extended to include the genera *Orthophasmavirus*,
325 *Inshuvirus*, *Feravirus*, *Wuhivirus*, and *Jonvirus* [24]. *Feravirus* and *Jonvirus* were designated
326 as separate families in 2016 [31], but this classification was revised in 2017 [24].

327

328 The isolation in cell culture of the type species of the genera *Feravirus* and *Jonvirus* (Ferah
329 and Jonchet virus, respectively) is still the only virological proof of phasmaviruses as
330 replicating viruses [32]. Evidence of the formation of 21 nt small RNAs, indicating infection,
331 was found in-vivo for a phasmavirus-like sequence in *Drosophila* [33]. The replication and
332 transcription strategy of phasmaviruses has been characterized [32]. Both *Jonchet jonvirus*

333 and *Ferak feravirus* have complementary genome segment ends that are identical between
334 L-, M-, and S-segments. This criterion is not fulfilled for any of the sequence-based viruses
335 classified as shown in **S21 Fig**, except for *Kigluaik phantom orthophasmavirus*. Experimental
336 proof of replication should be gathered, especially in putative viruses detected in pooled
337 samples [2].

338

339 Our data confirm many findings described in [3], and add a number of novel viruses that
340 enable proposals for genus classification based on distance and, often, host association.
341 These clades are named in **S21 Fig**. Four of them can be considered for the definition of
342 novel genera, based on distance and availability of three-segmented genomes (**Table 1**).
343 Dipteran-associated phasmaviruses (DAP) are defined by Wuhan mosquito
344 orthophasmavirus 1 as well as sequences from the present study. All DAP members show
345 tree congruence between L- and N-gene sequences but unclear topological resolution for the
346 M-segment (glycoprotein gene) sequences (**S21 Fig, S24 Fig**). Li *et al.* and Shi *et al.* have
347 found the viral genomes in this clade to only consist of L- and S-segments, with no M-
348 segment [2, 3]. However, our data contain members of the same clades that do have M-
349 segments, which raises the question of whether viruses with only two segments actually
350 exist. RNAseq studies can fail to identify the full set of genome segments.

351

352 Odonatan phasma-related virus OKIAV254 confirms the host association of two viruses
353 described in [3], Hubei odonate virus 8 and -9 (summarized in **S21 Fig**), here named
354 Odonate-associated phasmaviruses (OAP). Three segments are known. Coleopteran
355 phasma-related virus OKIAV236 and -235 stem from beetles, suggesting a Coleoptera-
356 associated phasmavirus clade (CAP, with 3 segments known and congruence between L-
357 and N-segment trees detected, **S24 Fig**). Blattodean phasma-related virus OKIAV239 and -
358 238 confirm the host association of Wuchang cockroach virus 1, defining the Blattodea-
359 associated phasmavirus clade (BAP). We find a large clade of OKIAV Hymenoptera-
360 associated phasmaviruses (HAP), whose only additional member is a virus found recently in
361 Hymenoptera, Ganda bee virus [34]. All HAP are topologically congruent between L-, N-, and
362 M-segment phylogenies (**S24 Fig**). The only exception is Hymenopteran phasma-related
363 virus OKIAV227 whose glycoprotein placement suggests reassortment or false assignment
364 due to coinfection. There are two clades with mixed host associations that are sufficiently
365 phylogenetically separate and have three known genome segments, altogether justifying
366 consideration as novel genera: MAP1 (mixed host-associated phasmavirus 1) and MAP2
367 (mixed host-associated phasmavirus 2). DAP2 is another dipteran-associated lineage
368 defined by a solitary virus (Dipteran phasma-related virus OKIAV226) with a fully-sequenced
369 three-partite genome, and may constitute a new genus (**S21 Fig**).

370

371 A novel sister clade to all members of the genus *Feravirus* is entirely defined by viruses from
372 the present study. For two viruses, Hymenopteran phasma-related virus OKIAV252 and -
373 250, complete L-segments are available. Hymenopteran phasma-related virus OKIAV252
374 has a matching S-segment sequence whose 5'-end aligns with that of the L-segment. We
375 use the clade designation HAF for these Hymenoptera-associated Feravirus-related viruses.
376 For almost all HAF, three genome segments have been identified, and one of them has
377 complete (and identical) genome segment termini (Hymenopteran phasma-related virus
378 OKIAV252). Classification of HAF as a genus may be formally possible on this basis. Of
379 note, a very recent description of a partial L-segment of Notori virus suggests this virus may
380 belong to HAF. However, in view of its evolutionary distance and association with Diptera
381 instead of Hymenoptera, it may as well represent another genus. Also this virus cannot be
382 classified because it lacks a full genome sequence.

383

384 Hemipteran phasma-related virus OKIAV247 and Neuropteran phasma-related virus
385 OKIAV248 are novel feraviruses with complete genome sequences and reverse
386 complementary genome termini. While feraviruses have only been found in mosquitoes
387 (Diptera) so far, the novel viruses stem from a lacewing (Neuroptera) and a shieldbug
388 (Hemiptera). Notably, whereas diversified clades around members of *Ferak feravirus*
389 emerge, *Jonchet jonvirus* remains the sole member of a deep-branching lineage (genus
390 *Jonvirus*). Considering that members of the species *Jonchet jonvirus* were repeatedly found
391 in mosquitoes by our group, this suggests tight host restriction (no mosquitos are
392 represented in the current dataset).

393

394 The recently classified genus *Wuhivirus* was identified in water striders and a mix of insects
395 in China [2, 24]. We find viral sequences which group together with *Wuhivirus*, thus
396 confirming the finding. The host associations of this clade are highly diverse since viral
397 sequences are derived from members of two different families of Hemiptera, Lepidoptera,
398 and mixes of water striders and Odonata. We did not find identical genome termini in either
399 the known or the novel members of this tentative genus. In summary, 10 novel genera may
400 be considered within the family *Phasmaviridae* based on the present data.

401

402 ***Phenuiviridae***

403 The genus *Phlebovirus* in the newly defined family *Phenuiviridae* classifies some of the most
404 medically-relevant bunyaviruses. Viruses belong to over 70 antigenically distinct serotypes
405 and include several human and livestock pathogens, such as *Rift Valley fever phlebovirus*,

406 and several sandfly fever viruses. Most phleboviruses are arboviruses, and most are
407 mosquito-, sandfly-, or tick-borne.

408

409 A novel genus termed *Banyangvirus* has recently been proposed and accepted in the most
410 recent ICTV taxonomy update (proposal 2017.012M, [24]). An additional proposal suggests
411 the expansion of *Banyangvirus* by adding further viral taxa (proposal 2018.013M, [35]).

412 However, our phylogenetic analysis finds that the members of *Banyangvirus* are in sister
413 relationship to members of the classical genus *Phlebovirus*. As it appears, mammalian and
414 dipteran members of *Phlebovirus* group together with *Banyangvirus*, to the exclusion of
415 *Uukuniemi*- and other tick-associated phleboviruses. The existence of *Banyangvirus* makes
416 *Phlebovirus* a paraphyletic genus (**S25 Fig**). This issue has not been addressed by the
417 proposal 2017.012M, in which the re-organisation of the order *Bunyavirales* has been
418 promoted. In this proposal, the phylogenetic inference includes only the type species of the
419 genera, therefore overlooking the within-genus relationships and inevitably rendering
420 *Phlebovirus* and *Banyangvirus* as sister clades.

421

422 The ICTV proposal 2018.013M is based on an analysis of nucleotide sequences, while
423 amino acid sequence analysis is more robust against errors caused by substitution
424 saturation. Indeed, the phylogenetic support for the L-gene analysis in 2018.013M is at the
425 borderline of significance. The alignment of structural gene portions might be better resolved,
426 but the topological incongruence implied by the overall phylogeny (in our and earlier
427 analyses compared to 2018.013M) suggests reassortment within the phlebo- and
428 banyangvirus stem lineages. Consequently, members of *Phlebovirus* and *Banyangvirus*
429 seem to have undergone recent speciation and still carry footprints of genetic exchange. This
430 argues against the classification of each as a separate genus, but rather for designating
431 *Banyangvirus* as a subgenus of *Phlebovirus*. Additional supports for this ranking are the
432 phylogenies derived from the nucleoprotein and glycoprotein of phlebo- and banyangviruses,
433 which show a very similar topology (**S28 Fig**).

434

435 Interestingly, we do not find any novel insect-associated phleboviruses. This is remarkable
436 as the studies of [2] and [3] also did not reveal any novel phleboviruses in insects, but did
437 find phleboviruses in ticks. Odonatan phenui-related virus OKIAV258 and Hymenopteran
438 phenui-related virus OKIAV306 fall between phlebo- and tenuiviruses, together with several
439 sequence findings from [2] and [3]. However, none of the genomes of these viruses is fully
440 sequenced.

441

442 Two newly classified genera, *Hudovirus* and *Pidchovirus*, have been defined in the family
443 *Phenuiviridae*. The genome of *Lepidopteran hudovirus*, type species of *Hudovirus*, has 3
444 segments whose ends are reverse-complementary and identical among all 3 segments. The
445 genome of *Pidgey pidchovirus*, type species of *Pidchovirus*, has segment termini that are
446 neither reverse-complementary nor identical for any segment. We find fragments of viral
447 sequences for all L-, M-, and S-segments, but none of them have known genome termini.
448 The sequences of *Hudo-* and *Pidchovirus* were both found in Lepidoptera, unlike our findings
449 which are all found in various insect orders.

450
451 Four additional novel genera have been assigned by ICTV to *Phenuiviridae*: *Phasi-*, *Hudi-*,
452 *Beidi-*, and *Wubeivirus*. We do not find viral sequences that cluster with any of those genera.
453 The genomes of *Hudi-*, *Beidi-*, and *Wubeivirus* do not have complementary genome segment
454 termini and were initially found in mixed samples of Diptera (**S25 Fig**). Of note, the two
455 viruses assigned to *Wubeivirus* appear paraphyletic in all our analysis, which argues against
456 their classification in the same genus. Phylogenetic analyses in now-accepted classification
457 proposals includes M- and S-segment sequences. However, the use of concatenated
458 alignments in this approach ignores the possibility that the segments evolved as independent
459 genomic units. Our analysis of segment co-segregation does not find topological congruence
460 between L-, M-, and S-segments in these viruses (**S28 Fig**). With the exception of *Badu*
461 *phasivirus*, none of the sequences pertaining to the members of these genera have
462 complementary genome segment termini. In addition, for all members, including *Badu*
463 *phasivirus*, the genome termini among segments are not identical. Because most of the
464 sequences were detected in mixed samples, it is possible that a mixture of segments from
465 different viruses was isolated. Therefore, none of these sequences should be regarded as
466 concrete viral genomes.

467
468 The genus *Goukovirus* is named after *Gouleako goukovirus*, a fully-characterized virus
469 based on cell culture isolates [36]. The genome termini of *Gouleako goukovirus* are reverse-
470 complementary, over ~14nt, and identical among S-, M-, and L-segments. Host associations
471 of the viral sequences from the present study suggest a predominant, but not exclusive,
472 representation in Hymenoptera while the type isolates stem from mosquitoes (Diptera).

473
474 Coleopteran phenui-related virus OKIAV293 represents a novel lineage branching as deep
475 as typical for a novel genus but cannot be classified as such because it does not constitute a
476 coding-complete genome.

477

478 At the bottom of the phylogenetic tree in **S25 Fig**, a large clade within the family
479 *Phenuiviridae* contains 10 sequences from [2] and [3], *Mothra mobuvirus*, and 26 novel
480 sequences from the present study. *Mothra mobuvirus* is the type species of the novel genus
481 *Mobuvirus* and has been detected in moths [4]. The phylogenetic diversity and the genetic
482 distance of this large monophyletic group suggest the definition of several genera that may
483 then form a subfamily within *Phenuiviridae*. The host associations of this clade's viruses are
484 diverse, spanning 10 different insect orders. Among all of these viruses, Hemipteran phenui-
485 related virus OKIAV285 and Blattodean phenui-related virus OKIAV266 are the only
486 sequences with complementary L-segment genome termini. Each of them is found in one of
487 two subclades of the new monophyletic group, differing considerably from each other and
488 from *Goukovirus* and *Phlebovirus*. Across all members of this large clade, the number of
489 genome segments ranges from one to three. Wherever two segments are found, those
490 represent the L- and M-, or L- and S-segments. This points to the presence of three
491 segments in all viruses. Given the tremendous evolutionary distance between S- and M-
492 segments among bunyaviruses, it is likely that virus identification based on RNAseq data will
493 not correctly, if at all, detect missing diverse segments. Of note, *Mothra mobuvirus* has only
494 two segments and should thus be subjected to experimental confirmation of replication to
495 confirm this unusual genome architecture.

496

497 ***Arenaviridae***

498 The family *Arenaviridae* consists of the genera *Mammarenavirus*, *Reptarenavirus*,
499 *Antennavirus*, and *Hartmanivirus*. The genus *Mammarenavirus* includes important human
500 pathogens such as *Lassa virus* and the highly pathogenic New World arenaviruses. All
501 species in the genera *Reptarenavirus* and *Hartmanivirus* occur only in snakes. As shown in
502 **S30 Fig**, Dipteran mypo-related virus OKIAV322, found in australian long-legged flies
503 (*Heteropsilopus ingenuus*), groups together with *Myriapod hubavirus*. The latter is the sole
504 member of the newly assigned genus *Hubavirus* and family *Mypoviridae*. The phylogenetic
505 distance in this newly-formed clade matches the intra-genus distance of *Mammarenavirus*,
506 thus supporting the assignment of *Hubavirus*.

507

508 ***Nairoviridae***

509 The family *Nairoviridae* includes tick-transmitted arboviruses such as the *Crimean-Congo*
510 *hemorrhagic fever orthonairovirus*, classified in the genus *Orthonairovirus*. Arthropod viruses
511 related to this genus are almost exclusively found in ticks [2, 3]. In the most recent
512 classification proposal (ICTV proposal 2018.017M, [23]), the sequence of *Strider striwavirus*
513 found in a mix of water striders, is assigned to the monospecific genus *Striwavirus*. Our data
514 confirm the existence of this clade by contributing another insect-associated viral sequence,

515 Blattodean nairo-related virus OKIAV321, which groups together with *Strider striwavirus*.
516 Blattodean nairo-related virus OKIAV321 has three genome segments, like *Strider*
517 *striwavirus*, which are however incomplete, and do not allow for genome termini examination.
518 Our finding does not stem from a water strider, but from the cockroach species
519 *Paratemnopterix coulouiana* that is widely used by pet keepers as a food source for
520 scorpions, spiders, and other insect-fed companion animals. The clustering of the
521 unclassified Xinzhou spider virus within this clade suggests that this virus may fall into this
522 tentative genus.

523

524 The absence of members of the genus *Orthonairovirus* in our vast sample of insects
525 suggests that orthonairoviruses may be exclusively associated with ticks, while the source of
526 nairovirus diversity spans a wide invertebrate host range.

527

528 Viral sequences found in myriapods in the present study, as well as in crabs from [3], group
529 together into a highly distinct sister clade to arena- and nairoviruses (**S30 Fig**). While
530 Jiangxia mosquito virus 1 could be considered a type species with two known segments,
531 none of the members of this clade have been formally assigned to a novel genus or family.
532 Its phylogenetic distance is similar to that of other families in the tree, suggesting that it may
533 define a novel family.

534

535 Overall, the co-segregation of segments shows some topologically congruent clades that are
536 larger compared to the other phylogenies in this study (**S33 Fig**). This might be explained by
537 the lack of novel sequences and the appropriate annotation and characterization of the
538 viruses backed by laboratory experiments.

539

540 **Results of cytochrome oxidase subunit 1 (COI) barcode analysis**

541 We conducted searches for contaminating non-hexapod sequences in all transcriptome
542 assemblies used in the study. The search was based on ~2.5M Eukaryote cytochrome c
543 oxidase subunit 1 gene sequences from GenBank and ~0.5M additional gene barcodes from
544 the German Barcode of Life project. All findings are summarized in **Supporting Information**
545 **S3 and S4 Tables**. Four OKIAV viruses assembled from transcriptomes in which non-
546 hexapod sequences were detected are included in phylogenies of **Fig 1**, marked with an
547 empty triangle. Based on the distribution of contaminated assemblies matching non-hexapod
548 organisms, we regard the possibility of having detected a clade of viruses that actually
549 belongs to contaminant organisms instead of hexapods as very unlikely.

550

551 **References**

- 552 1. Contreras MA, Eastwood G, Guzman H, Popov V, Savit C, Uribe S, et al.
553 Almendravirus: A Proposed New Genus of Rhabdoviruses Isolated from Mosquitoes in
554 Tropical Regions of the Americas. *Am J Trop Med Hyg.* 2017;96(1):100-9. doi:
555 10.4269/ajtmh.16-0403.
- 556 2. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, et al. Unprecedented genomic
557 diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses.
558 *Elife.* 2015;4. doi: 10.7554/eLife.05378.
- 559 3. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate
560 RNA virosphere. *Nature.* 2016. Epub 2016/11/24. doi: 10.1038/nature20167.
- 561 4. Makhsous N, Shean RC, Droppers D, Guan J, Jerome KR, Greninger AL. Genome
562 Sequences of Three Novel Bunyaviruses, Two Novel Rhabdoviruses, and One Novel
563 Nyamivirus from Washington State Moths. *Genome Announc.* 2017;5(7). doi:
564 10.1128/genomeA.01668-16.
- 565 5. Maes P, Song T, Mark S, Paweska J, Song Q, Ye G, et al. ICTV taxonomic report
566 2017.016M.R.: Taxonomic expansion and reorganization of the order Mononegavirales.
- 567 6. Kuhn JH, Bekal S, Cai Y, Clawson AN, Domier LL, Herrel M, et al. Nyamiviridae:
568 proposal for a new family in the order Mononegavirales. *Arch Virol.* 2013;158(10):2209-26.
569 doi: 10.1007/s00705-013-1674-y.
- 570 7. Mihindukulasuriya KA, Nguyen NL, Wu G, Huang HV, da Rosa APAT, Popov VL, et
571 al. Nyamanini and midway viruses define a novel taxon of RNA viruses in the order
572 Mononegavirales. *J Virol.* 2009;83(10):5109-16. doi: 10.1128/JVI.02667-08.
- 573 8. Rogers MB, Cui L, Fitch A, Popov V, Travassos da Rosa APA, Vasilakis N, et al.
574 Whole genome analysis of sierra nevada virus, a novel mononegavirus in the family
575 nyamiviridae. *Am J Trop Med Hyg.* 2014;91(1):159-64. doi: 10.4269/ajtmh.14-0076.
- 576 9. Bekal S, Domier LL, Niblack TL, Lambert KN. Discovery and initial analysis of novel
577 viral genomes in the soybean cyst nematode. *J Gen Virol.* 2011;92(Pt 8):1870-9. doi:
578 10.1099/vir.0.030585-0.
- 579 10. Kleanthous E, Olendraite I, Lukhovitskaya NI, Firth AE. Discovery of three RNA
580 viruses using ant transcriptomic datasets. *Arch Virol.* 2019;164(2):643-7. doi:
581 10.1007/s00705-018-4093-2.
- 582 11. Wang F, Fang Q, Wang B, Yan Z, Hong J, Bao Y, et al. A novel negative-stranded
583 RNA virus mediates sex ratio in its parasitoid host. *PLoS Pathog.* 2017;13(3):e1006201. doi:
584 10.1371/journal.ppat.1006201.
- 585 12. Kuhn JH, Dietzgen RG, Easton AJ, Kurath G, Nowotny N, Rima BK, et al. ICTV
586 taxonomic proposal 2015.010A: Five (5) new unassigned genera in the order
587 Mononegavirales.

- 588 13. Viljakainen L, Holmberg I, Abril S, Juvansuu J. Viruses of invasive Argentine ants
589 from the European Main supercolony: characterization, interactions and evolution. *J Gen*
590 *Virool.* 2018;99(8):1129-40. doi: 10.1099/jgv.0.001104.
- 591 14. Liu L, Xie J, Cheng J, Fu Y, Li G, Yi X, et al. Fungal negative-stranded RNA virus that
592 is related to bornaviruses and nyaviruses. *Proc Natl Acad Sci U S A.* 2014;111(33):12205-
593 10. doi: 10.1073/pnas.1401786111.
- 594 15. Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, Binger T, et al. Bats host
595 major mammalian paramyxoviruses. *Nat Commun.* 2012;3:796. doi: 10.1038/ncomms1796.
- 596 16. Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, et al. The evolutionary history of
597 vertebrate RNA viruses. *Nature.* 2018;556(7700):197-202. doi: 10.1038/s41586-018-0012-7.
- 598 17. Presti RM, Zhao G, Beatty WL, Mihindikulasuriya KA, da Rosa APAT, Popov VL, et
599 al. Quarantfil, Johnston Atoll, and Lake Chad viruses are novel members of the family
600 Orthomyxoviridae. *J Virol.* 2009;83(22):11599-606. doi: 10.1128/JVI.00677-09.
- 601 18. Allison AB, Ballard JR, Tesh RB, Brown JD, Ruder MG, Keel MK, et al. Cyclic avian
602 mass mortality in the northeastern United States is associated with a novel orthomyxovirus. *J*
603 *Virool.* 2015;89(2):1389-403. doi: 10.1128/JVI.02019-14.
- 604 19. Dannevig BH, Falk K, Namork E. Isolation of the causal virus of infectious salmon
605 anaemia (ISA) in a long-term cell line from Atlantic salmon head kidney. *J Gen Virool.* 1995;76
606 (Pt 6):1353-9. doi: 10.1099/0022-1317-76-6-1353.
- 607 20. Jones LD, Matthewson M, Nuttall PA. Saliva-activated transmission (SAT) of Thogoto
608 virus: dynamics of SAT factor activity in the salivary glands of *Rhipicephalus appendiculatus*,
609 *Amblyomma variegatum*, and *Boophilus microplus* ticks. *Exp Appl Acarol.* 1992;13(4):241-8.
610 doi: 10.1007/BF01195081
- 611 21. Eyngor M, Zamostiano R, Kembou Tsofack JE, Berkowitz A, Bercovier H, Tinman S,
612 et al. Identification of a novel RNA virus lethal to tilapia. *J Clin Microbiol.* 2014;52(12):4137-
613 46. doi: 10.1128/JCM.00827-14.
- 614 22. Contreras-Gutiérrez MA, Nunes MRT, Guzman H, Uribe S, Suaza Vasco JD,
615 Cardoso JF, et al. Sinu virus, a novel and divergent orthomyxovirus related to members of
616 the genus Thogotovirus isolated from mosquitoes in Colombia. *Virology.* 2017;501:166-75.
617 doi: 10.1016/j.virol.2016.11.014.
- 618 23. Maes P, Kuhn JH. ICTV taxonomic proposal 2018.017M: Expansion of the order
619 Bunyavirales.
- 620 24. Maes P, Alkhovsky S, Beer M, Briese T, Buchmeier MJ, Calisher CH, et al. ICTV
621 taxonomic proposal 2017.012M: Taxonomic expansion and reorganization of the order
622 Bunyavirales.
- 623 25. Elliott RM. Molecular biology of the Bunyaviridae. *J Gen Virool.* 1990;71 (Pt 3):501-22.
624 doi: 10.1099/0022-1317-71-3-501.

- 625 26. Mir MA, Brown B, Hjelle B, Duran WA, Panganiban AT. Hantavirus N protein exhibits
626 genus-specific recognition of the viral RNA panhandle. *J Virol.* 2006;80(22):11283-92. doi:
627 10.1128/JVI.00820-06.
- 628 27. Marklewitz M, Zirkel F, Rwego IB, Heidemann H, Trippner P, Kurth A, et al. Discovery
629 of a unique novel clade of mosquito-associated bunyaviruses. *J Virol.* 2013;87(23):12850-65.
630 doi: 10.1128/JVI.01862-13.
- 631 28. Al'kovskhovskii SV, Shchetinin AM, L'Vov DK, Shchelkanov MI, Deriabin PG, L'Vov
632 DN, et al. The Khurdun virus (KHURV): a new representative of the orthobunyavirus
633 (Bunyaviridae). *Voprosy virusologii.* 2013;58(4):10-3.
- 634 29. Galkina IV, L'Vov LN, Gromashevskii VL, others. Khurdun virus, a presumably new
635 RNA-containing virus associated with coots (*Fulica atra*), isolated in the Volga river delta.
636 *Voprosy virusologii.* 2005;50(4):29-31.
- 637 30. Ballinger MJ, Bruenn JA, Hay J, Czechowski D, Taylor DJ. Discovery and evolution of
638 bunyavirids in arctic phantom midges and ancient bunyavirid-like sequences in insect
639 genomes. *J Virol.* 2014;88(16):8783-94. doi: 10.1128/JVI.00531-14.
- 640 31. Briese T, Alkhovsky S, Beer M, Calisher CH, Charrel R, Ebihara H, et al. ICTV
641 taxonomic proposal 2016.030a-vM: Create a new order, Bunyvirales, to accommodate nine
642 families (eight new, one renamed) comprising thirteen genera.
- 643 32. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. Evolutionary and phenotypic
644 analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family.
645 *Proc Natl Acad Sci U S A.* 2015;112(24):7536-41. doi: 10.1073/pnas.1502036112.
- 646 33. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, et al.
647 The discovery, distribution, and evolution of viruses associated with *Drosophila*
648 *melanogaster*. *PLoS biology.* 2015;13(7):e1002210.
- 649 34. Schoonvaere K, De Smet L, Smagghe G, Vierstraete A, Braeckman BP, de Graaf
650 DC. Unbiased RNA Shotgun Metagenomics in Social and Solitary Wild Bees Detects
651 Associations with Eukaryote Parasites and New Viruses. *PLoS One.* 2016;11(12):e0168456.
652 doi: 10.1371/journal.pone.0168456.
- 653 35. Shen S, Duan X, Wang B, Zhu L, Zhang Y, Zhang J, et al. ICTV taxonomic proposal
654 2018.013M: One (1) new species in the genus Banyangvirus (Bunyvirales: Phenuiviridae).
- 655 36. Marklewitz M, Handrick S, Grasse W, Kurth A, Lukashev A, Drosten C, et al.
656 Gouleako virus isolated from West African mosquitoes constitutes a proposed novel genus in
657 the family Bunyviridae. *J Virol.* 2011;85(17):9227-34. doi: 10.1128/JVI.00230-11.
- 658

Virus_name	Figure	Virus_family	Host_subphylum	Host_class	Host_superorder	Host_order	Host_order_group	Host_family	Host_species	Sample_location	Sample_date
Blattodean rhabdo-related virus OKIAV14	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Blattidae	<i>Deropeltis erythrocephala</i>	Germany: Lab culture with samples originating from Germany; private breeder	3/13/2011
Coleopteran rhabdo-related virus OKIAV10	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Noteridae	<i>Noterus clavicornis</i>	Germany: Thuringia; Jena - Closewitz	4/10/2011
Coleopteran rhabdo-related virus OKIAV20	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Erotylidae	<i>Pharaxonotha floridana</i>	USA: Florida; Placida near Don Pedro Island State Park	12/1/2013
Coleopteran rhabdo-related virus OKIAV28	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Scarabaeidae	<i>Onthophagus similis</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	3/28/2014
Coleopteran rhabdo-related virus OKIAV29	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Cryptophagidae	<i>Atomaria fuscata</i>	Germany: Thuringia; Jena - Loebstedt	4/11/2011
Dipteran rhabdo-related virus OKIAV19	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Perissommatidae	<i>Perissomma mcalpinei</i>	Australia: New South Wales; Tallaganda National Park	6/9/2013
Dipteran rhabdo-related virus OKIAV27	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Agromyzidae	<i>Phytomyza hellebori</i>	Germany: Lab culture - Susanne Dobler - University of Hamburg	2/12/2013
Dipteran rhabdo-related virus OKIAV36	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Acroceridae	<i>Pterodontia mellii</i>	Australia: Australian Capital Territory; Bonython	11/21/2012
Dipteran rhabdo-related virus OKIAV5	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Chaoboridae	<i>Chaoborus flavidulus</i>	Singapore: Upper Seletar Reservoir Park	7/1/2013
Hemipteran rhabdo-related virus OKIAV26	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Membracidae	<i>Centrotus cornutus</i>	Croatia: Zadar Donji Brig	5/23/2011
Hemipteran rhabdo-related virus OKIAV30	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Aphrophoridae	<i>Aphrophora alni</i>	Germany: Thuringia; Jena	7/30/2012
Hemipteran rhabdo-related virus OKIAV47	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Pseudococcidae	<i>Planococcus citri</i>	Germany: commercial lab culture KATZ Biotech AG Baruth Brandenburg	11/10/2011
Hymenopteran almendra-related virus OKIAV1	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Eulophidae	<i>Diglyphus isaea</i>	Lab culture of unknown geographical origin	5/12/2011
Hymenopteran rhabdo-related virus OKIAV109	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Sphécidae	<i>Chlorion hirtum</i>	Israel: W Ein Bokek	4/24/2013
Hymenopteran rhabdo-related virus OKIAV22	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysis sp.</i>	Israel: Revivim	6/3/2012
Hymenopteran rhabdo-related virus OKIAV23	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysura austriaca</i>	Germany: Rhineland-Palatinate; Battenberg	7/3/2010
Hymenopteran rhabdo-related virus OKIAV24	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Pompilidae	<i>Heterodontonyx sp.</i>	Australia: Western Australia; 118 km N Esperance	11/7/2011
Hymenopteran rhabdo-related virus OKIAV25	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Pompilidae	<i>Heterodontonyx sp.</i>	Australia: Western Australia; 118 km N Esperance	11/7/2011
Hymenopteran rhabdo-related virus OKIAV38	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Pompilidae	<i>Pompilus cinereus</i>	Germany: Rhineland-Palatinate; Birkenheide	5/22/2011
Hymenopteran rhabdo-related virus OKIAV40	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Sphécidae	<i>Ammophila sabulosa</i>	Germany: Lower Saxony; Pevestorf - Elbauenstation	8/28/2012
Hymenopteran rhabdo-related virus OKIAV45	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Sapygidae	<i>Monosapyga clavicornis</i>	Germany: Rhineland-Palatinate; Albersweiler	4/1/2011
Hymenopteran rhabdo-related virus OKIAV46	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysura radians</i>	Italy: Valle d' Aosta; Turlin	6/18/2012
Hymenopteran rhabdo-related virus OKIAV8	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Mymaridae	<i>Gonatocerus morilli</i>	USA: Lab culture with samples originating from Texas - Weslaco	2013
Lepidopteran rhabdo-related virus OKIAV11	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Epipyropidae	<i>Epipomponia nawai</i>	South Korea: Ulsan City - Ulju Mount Ganweolsan	8/28/2012
Lepidopteran rhabdo-related virus OKIAV12	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Nymphalidae	<i>Pararge aegeria</i>	Germany: North Rhine-Westphalia; Rheinland Nature Park Kottenforst	5/10/2011
Lepidopteran rhabdo-related virus OKIAV3	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Mimallonidae	<i>Menevia lucara</i>	French Guiana: Nouragues Field Station; Parare Camp	10/1/2012
Lepidopteran rhabdo-related virus OKIAV32	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Adelidae	<i>Nemophora degeerella</i>	Austria: Lower Austria; Guntramsdorf-Eichkogel	6/2/2011
Lepidopteran rhabdo-related virus OKIAV33	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Nymphalidae	<i>Pararge aegeria</i>	Germany: North Rhine-Westphalia; Rheinland Nature Park Kottenforst	5/10/2011
Lepidopteran rhabdo-related virus OKIAV34	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Hepialidae	<i>Triodia sylvina</i>	Germany: Lower Saxony; Hoehbeck - Pevestorf - Station	8/27/2011
Lepidopteran rhabdo-related virus OKIAV35	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Lasiocampidae	<i>Dendrolimus pini</i>	Germany: Lower Saxony; Luechow-Dannenberg - Hoehbeck - Pevestorf	8/9/2012
Mantodean rhabdo-related virus OKIAV15	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Dictyoptera	Mantodea	Polyneoptera	Mantidae	<i>Nilomantis floweri</i>	Germany: Lab culture with samples originating from Oman	2012
Mecopteran rhabdo-related virus OKIAV42	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Panorpida	Mecoptera	Mecoptera	Boreidae	<i>Caurinus tlagu</i>	USA: Alaska; Prince of Wales Island - Tongass National Forest Stanley Creek	5/16/2013
Neuropteran rhabdo-related virus OKIAV31	S1 Fig	<i>Rhabdoviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Myrmeleontidae	<i>Myrmeleon formicarius</i>	Japan: Ibaraki Prefecture; Tsukuba	6/25/2012
Blattodean arli-related virus OKIAV101	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Corydiidae	<i>Polyphaga aegyptiaca</i>	Germany: Lab culture with samples originating from Germany; Helbigsdorf	3/20/2013
Blattodean arli-related virus OKIAV102	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Ectobiidae	<i>Supella longipalpa</i>	Japan: Lab culture with samples originating from Thailand	12/28/2011
Blattodean arli-related virus OKIAV103	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Termitidae	<i>Coptotermes sp.</i>	Gabon: Ivingo National Park	9/16/2012
Coleopteran anphe-related virus OKIAV54	S6 Fig	<i>Ximoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Coccinellidae	<i>Chilocorus renipustulatus</i>	Germany: Thuringia; Jena - Tautenburger Forest	3/23/2011
Coleopteran arli-related virus OKIAV107	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Dytiscidae	<i>Cybister lateralmarginalis</i>	Germany: Lower Saxony; Luechow-Dannenberg - Hoehbeck - Pevestorf	8/27/2011
Dipteran anphe-related virus OKIAV69	S6 Fig	<i>Ximoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Vermileonidae	<i>Vermileo vermileo</i>	Italy: Umbria; Perugia Panicale	9/1/2012
Dipteran arto-related virus OKIAV77	S6 Fig	<i>Artoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Stratiomyidae	<i>Boreoides subulatus</i>	Australia: Australian Capital Territory; Canberra	2013
Hemipteran anphe-related virus OKIAV63	S6 Fig	<i>Ximoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Psyllidae	<i>Glycaspis brimblecombei</i>	Australia: South Australia; Adelaide River Torrens	2/20/2012
Hemipteran arli-related virus OKIAV94	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Aleyrodidae	<i>Trialeurodes vaporariorum</i>	Germany: commercial lab culture KATZ Biotech AG Baruth Brandenburg	9/29/2011
Hemipteran arli-related virus OKIAV95	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Peloriidae	<i>Xenophyes metoponcus</i>	New Zealand: South Island Westland District; Lake Matheson	9/1/2011

Virus_name	Figure	Virus_family	Host_subphylum	Host_class	Host_superorder	Host_order	Host_order_group	Host_family	Host_species	Sample_location	Sample_date
Hymenopteran anphe-related virus OKIAV71	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Ichneumonidae	<i>Heteropelma amictum</i>	Switzerland: Bern Eymatt	9/17/2012
Hymenopteran anphe-related virus OKIAV72	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Scoliidae	<i>Colpa sexmaculata</i>	Italy: Sardinia; Costa Verde 13 km E Arbus	9/9/2011
Hymenopteran anphe-related virus OKIAV73	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Evanidae	<i>Brachygaster minutus</i>	Croatia: Zadar Donji Brig	5/23/2011
Hymenopteran arli-related virus OKIAV100	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Orussidae	<i>Orussus unicolor</i>	Germany: Hessen; Darmstadt	5/10/2012
Hymenopteran arli-related virus OKIAV98	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Cynipidae	<i>Synergus umbraculus</i>	Germany: North Rhine-Westphalia; Bonn	2012
Hymenopteran arli-related virus OKIAV99	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Aphelinidae	<i>Aphelinus abdominalis</i>	Lab culture of unknown geographical origin	5/12/2011
Hymenopteran arto-related virus OKIAV1338	S6 Fig	<i>Artoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Torymidae	<i>Podagrion pachymerum</i>	Slovakia: Rudno nad Hronom	1/1/2012
Hymenopteran arto-related virus OKIAV79	S6 Fig	<i>Artoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Eucharitidae	<i>Eucharis adscendens</i>	Germany: Baden-Wuerttemberg; Landkreis Boeblingen - Grafenberg	7/19/2013
Hymenopteran orino-related virus OKIAV85	S6 Fig	<i>Nyamiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Crabronidae	<i>Gorytes laticinctus</i>	Germany: Rhineland-Palatinate; Albersweiler	6/6/2011
Hymenopteran orino-related virus OKIAV87	S6 Fig	<i>Nyamiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Sphexidae	<i>Chalybion californicum</i>	USA: Tennessee; ca. 12 km ssE Lebanon	6/20/2011
Hymenopteran orino-related virus OKIAV88	S6 Fig	<i>Nyamiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Cephidae	<i>Cephus spinipes</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	5/31/2012
Lepidopteran anphe-related virus OKIAV50	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Pieridae	<i>Colias croceus</i>	Greece: East Macedonia and Thrace Thassos near Theologos	5/9/2012
Lepidopteran orino-related virus OKIAV84	S6 Fig	<i>Nyamiviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Nymphalidae	<i>Pararge aegeria</i>	Germany: North Rhine-Westphalia; Rheinland Nature Park Kottenforst	5/10/2011
Mantodean anphe-related virus OKIAV92	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Dictyoptera	Mantodea	Polyneoptera	Empusidae	<i>Idolomantis diabolica</i>	Germany: Lab culture with samples originating from Tanzania	2012
Megalopteran arli-related virus OKIAV106	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Endopterygota	Megaloptera	Neuropterida	Corydalidae	<i>Corydalus cornutus</i>	USA: Tennessee; Smokey Mountain National Park	6/9/2011
Neuropteran arli-related virus OKIAV105	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Hemerobiidae	<i>Hemerobius nitidulus</i>	Austria: Lower Austria; Krems-Land Duerstein	9/1/2012
Odonatan anphe-related virus OKIAV57	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Libellulidae	<i>Libellula quadrimaculata</i>	Germany: Rhineland-Palatinate; Germersheim Berg	4/23/2011
Odonatan anphe-related virus OKIAV59	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Cordulegastriidae	<i>Cordulegaster boltonii</i>	Germany: Rhineland-Palatinate; Germersheim Freckenfeld	7/16/2011
Odonatan anphe-related virus OKIAV60	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Petaluridae	<i>Tanypteryx pryeri</i>	Japan: Ibaraki Prefecture; Hitachi-Omiya	5/29/2012
Odonatan anphe-related virus OKIAV90	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Coenagrionidae	<i>Paracercion plagiosum</i>	Japan: Saitama Prefecture; Sugito	6/3/2012
Odonatan arli-related virus OKIAV93	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Pseudostigmatidae	<i>Mecistogaster ornata</i>	Panama: Barro Colorado Island	2011
Strepsipteran arli-related virus OKIAV104	S6 Fig	<i>Lispiviridae</i>	Hexapoda	Insecta	NA	Strepsiptera	Strepsiptera	Stylopidae	<i>Stylops melittae</i>	Germany: Lower Saxony; Osnabrück-Sandgrube-Niedrighaussee	2012
Trichopteran arto-related virus OKIAV76	S6 Fig	<i>Artoviridae</i>	Hexapoda	Insecta	Amphiesmenoptera	Trichoptera	NA	NA	NA	NA	NA
Zygentoman anphe-related virus OKIAV89	S6 Fig	<i>Xinmoviridae</i>	Hexapoda	Insecta	NA	Zygentoma	Zygentoma	Lepismatidae	<i>Ctenolepisma longicaudata</i>	Germany: North Rhine-Westphalia; Bonn	2011
Blattodean chu-related virus OKIAV148	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Blattidae	<i>Periplaneta americana</i>	Germany: lab culture Kai Schuette Zoologisches Museum Hamburg	11/1/2011
Coleopteran chu-related virus OKIAV127	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Chrysomelidae	<i>Oreina cacaliae</i>	Germany: Lab culture with samples from Baden-Wuerttemberg	2/12/2012
Coleopteran chu-related virus OKIAV151	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Larinus minutus</i>	USA: Arkansas; Fayetteville University of Arkansas	11/1/2013
Collembolan qin-related virus OKIAV112	S9 Fig	<i>Qinviridae</i>	Hexapoda	Collembola	NA	Poduromorpha	Ellipura	Neanuridae	<i>Anurida maritima</i>	Netherlands: North Holland; Texel Ferry Bay	9/1/2011
Dermapteran chu-related virus OKIAV142	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	NA	Dermaptera	Polyneoptera	Anisolabididae	<i>Gonolabis marginalis</i>	Japan: Ibaraki Prefecture; Tsukuba - Tennodai	10/4/2011
Hemipteran chu-related virus OKIAV138	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Aleyrodidae	<i>Trialeurodes vaporariorum</i>	Germany: commercial lab culture KATZ Biotech AG	9/29/2011
Hemipteran chu-related virus OKIAV139	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Triozidae	<i>Acanthocasuaria muelleriana</i>	Australia: South Australia; Kangaroo Island Sedden Conservation Park	2/9/2012
Hemipteran chu-related virus OKIAV140	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Belostomatidae	<i>Diplonychus rusticus</i>	China: Guangdong Province	8/28/2012
Hymenopteran chu-related virus OKIAV122	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Megachilidae	<i>Chelostoma florisomme</i>	Germany: Rhineland-Palatinate; Albersweiler	4/22/2011
Hymenopteran chu-related virus OKIAV123	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Megachilidae	<i>Dioxys cincta</i>	France: Alpes-Maritimes Millefont	7/11/2011
Hymenopteran chu-related virus OKIAV124	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Apidae	<i>Nomada lathburiana</i>	Germany: Rhineland-Palatinate; Hochstadt - Pfalz	4/2/2011
Hymenopteran chu-related virus OKIAV125	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Megachilidae	<i>Megachile willughbiella</i>	Germany: Rhineland-Palatinate; Birkenheide	7/5/2011
Hymenopteran chu-related virus OKIAV126	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Crabronidae	<i>Oxybelus bipunctatus</i>	Germany: Rhineland-Palatinate; Battenberg	7/5/2011
Hymenopteran chu-related virus OKIAV146	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Diapriidae	<i>Trichopria drosophilae</i>	Lab culture with samples from France: 60 km south of Lyon	2012
Hymenopteran chu-related virus OKIAV147	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Aphelinidae	<i>Aphelinus abdominalis</i>	Lab culture of unknown geographical origin	5/12/2011
Megalopteran chu-related virus OKIAV119	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Endopterygota	Megaloptera	Neuropterida	Corydalidae	<i>Corydalidae sp.</i>	Venezuela: Carabobo Bejuma	11/9/2011
Megalopteran chu-related virus OKIAV130	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Endopterygota	Megaloptera	Neuropterida	Sialidae	<i>Sialis lutaria</i>	Austria: Vienna surroundings	5/1/2012
Neuropteran chu-related virus OKIAV150	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Mantispidae	<i>Eumantispia harmandi</i>	Japan: Nagano Prefecture	8/8/2012
Notopteran chu-related virus OKIAV120	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Exopterygota	Notoptera	Polyneoptera	Grylloblattidae	<i>Grylloblatta bifurcata</i>	USA: California; Tuolumne County Stanislaus National Forest	8/11/2011

Virus_name	Figure	Virus_family	Host_subphylum	Host_class	Host_superorder	Host_order	Host_order_group	Host_family	Host_species	Sample_location	Sample_date
Odonatan chu-related virus OKIAV136	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Odonoptera	Odonata	Odonata	Lestoideidae	<i>Diphlebia lestoides</i>	Australia: Australian Capital Territory; Gibraltar Creek	12/7/2011
Odonatan chu-related virus OKIAV137	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Odonoptera	Odonata	Odonata	Argiolestidae	<i>Austroargiolestes icteromelas</i>	Australia: Australian Capital Territory; Uriarra State Forest	12/6/2011
Orthopteran chu-related virus OKIAV116	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Orthoptera	Orthoptera	Polynoptera	Tetrigidae	<i>Tetrix subulata</i>	Germany: North Rhine-Westphalia; Rhein-Sieg-Kreis	5/17/2011
Orthopteran chu-related virus OKIAV152	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Orthoptera	Orthoptera	Polynoptera	Pyrgomorphae	<i>Phymateus viridipes</i>	Germany: Lab culture with samples from South Africa	1/1/2013
Phasmatodean chu-related virus OKIAV118	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Exopterygota	Phasmatodea	Polynoptera	Phasmatidae	<i>Abrosoma johorensis</i>	Germany: Lab culture with samples from West Malaysia	10/1/2012
Phasmatodean chu-related virus OKIAV1344	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Exopterygota	Phasmatodea	Polynoptera	Aschiphasmatidae	<i>Orthomeria sp.</i>	Germany: lab culture	10/16/2012
Trichopteran chu-related virus OKIAV114	S9 Fig	<i>Chuviridae</i>	Hexapoda	Insecta	Amphiesmenoptera	Trichoptera	Amphiesmenoptera	Limnephilidae	<i>Dicosmoecus jozankeanus</i>	Russia: Primorsky	7/20/2011
Archaeognathan orthomyxo-related virus OKIAV189	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	NA	Archaeognatha	Archaeognatha	Machilidae	<i>Pedetontus okajimae</i>	Japan: Shizuoka Prefecture; Shimoda - Touji	2/29/2012
Blattodean orthomyxo-related virus OKIAV181	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polynoptera	Cryptoceridae	<i>Cryptocercus wrighti</i>	USA: Oregon	2/1/2012
Coleopteran orthomyxo-related virus OKIAV158	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Sitophilus zeamais</i>	USA: Tennessee; Tipton County	6/1/2013
Coleopteran orthomyxo-related virus OKIAV168	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Lymexyelidae	<i>Melittomma sp.</i>	Australia: Queensland; Cape Tribulation	12/8/2012
Coleopteran orthomyxo-related virus OKIAV179	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Dytiscidae	<i>Cybister lateralmarginalis</i>	Germany: Lower Saxony	8/27/2011
Coleopteran orthomyxo-related virus OKIAV184	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Geotrupidae	<i>Anaplotrupes stercorosus</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	3/28/2014
Coleopteran orthomyxo-related virus OKIAV185	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Glaresidae	<i>Glareis sp.</i>	South Africa: Western Cape; Vrolijkheid Nature Reserve	12/9/2013
Coleopteran orthomyxo-related virus OKIAV186	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Gyrinidae	<i>Dineutes sp.</i>	USA: New Jersey; Sussex County - Stokes State Forest	5/5/2011
Coleopteran orthomyxo-related virus OKIAV196	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Ips typographus</i>	Germany: Rhineland-Palatinate; Cochem-Cell - Brohl	5/13/2012
Coleopteran orthomyxo-related virus OKIAV200	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Ips typographus</i>	Germany: Rhineland-Palatinate; Cochem-Cell - Brohl	5/13/2012
Coleopteran orthomyxo-related virus OKIAV201	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Coccinellidae	<i>Diomus ementitor</i>	Australia: Queensland; Mount Tamborine	10/31/2013
Coleopteran orthomyxo-related virus OKIAV202	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Coccinellidae	<i>Chilocorus renipustulatus</i>	Germany: Thuringia; Jena - Tautenburger Forest	3/23/2011
Dermapteran orthomyxo-related virus OKIAV161	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	NA	Dermaptera	Polynoptera	Labiduridae	<i>Allostethus burri</i>	Malaysia: Pahang; Cameron Highland Tanah Rata	4/9/2012
Dermapteran orthomyxo-related virus OKIAV162	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	NA	Dermaptera	Polynoptera	Pygidicranidae	<i>Anataelia canariensis</i>	Spain: Tenerife Punta de Teno	1/4/2011
Dermapteran orthomyxo-related virus OKIAV170	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	NA	Dermaptera	Polynoptera	Forficulidae	<i>Forficula auricularia</i>	Germany: North Rhine-Westphalia; Bonn; Poppelsdorf	6/7/2011
Dipteran orthomyxo-related virus OKIAV159	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Tachinidae	<i>Ocytata pallipes</i>	Germany: Hamburg; Sternschanzenpark	8/1/2011
Dipteran orthomyxo-related virus OKIAV163	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Stratiomyidae	<i>Inopus rubriceps</i>	Australia: Australian Capital Territory; Canberra	11/1/2012
Dipteran orthomyxo-related virus OKIAV164	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Stratiomyidae	<i>Exaireta spinigera</i>	Australia: Australian Capital Territory; Canberra O'Connor	1/14/2012
Dipteran orthomyxo-related virus OKIAV165	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Scenopinidae	<i>Scenopinus velutinus</i>	Costa Rica: Talamanca Nature Reserve	4/11/2014
Dipteran orthomyxo-related virus OKIAV166	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Apioceridae	<i>Apiocera maritima</i>	Australia: New South Wales; Mystery Bay Beach	1/21/2014
Dipteran orthomyxo-related virus OKIAV192	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Nemestrinidae	<i>Trichophthalma ricardoae</i>	Australia: Australian Capital Territory; Canberra	11/1/2012
Dipteran orthomyxo-related virus OKIAV193	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Trichoceridae	<i>Trichocera saltator</i>	Germany: Thuringia; Jena	11/14/2011
Dipteran orthomyxo-related virus OKIAV194	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Hybotidae	<i>Leptozebra sp.</i>	USA: North Carolina; Schenck Forest	5/15/2013
Dipteran orthomyxo-related virus OKIAV195	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Oestridae	<i>Cuterebra austeni</i>	USA: New Mexico Grant County; Silver City Gomez Park	5/27/2013
Dipteran orthomyxo-related virus OKIAV199	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Ulidiidae	<i>Diacritha costalis</i>	USA: Arizona; Pima County Tucson	6/6/2012
Hemipteran orthomyxo-related virus OKIAV182	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Delphacidae	<i>Nilaparvata lugens</i>	Germany: lab culture with samples from private breeder	10/1/2011
Hemipteran orthomyxo-related virus OKIAV183	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Peloriidae	<i>Xenophysella greensladeae</i>	New Zealand: South Island Westland District; Lake Matheson	9/20/2011
Hemipteran orthomyxo-related virus OKIAV187	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Psyllidae	<i>Glycaspis brimblecombei</i>	Australia: South Australia; Adelaide River Torrens	2/20/2012
Hemipteran orthomyxo-related virus OKIAV188	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Triozidae	<i>Acanthosacarina muellerianae</i>	Australia: South Australia; Seddon Conservation Park	2/9/2012
Hemipteran orthomyxo-related virus OKIAV191	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Acanthosomatidae	<i>Acanthosoma haemorrhoidale</i>	Germany: North Rhine-Westphalia; Rhein-Sieg-Kreis	5/17/2011
Hemipteran orthomyxo-related virus OKIAV205	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Peloriidae	<i>Xenophysella greensladeae</i>	New Zealand: South Island Westland District; Lake Matheson	9/20/2011
Hemipteran orthomyxo-related virus OKIAV211	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Pleidae	<i>Plea minutissima</i>	Germany: Lower Saxony	8/1/2011
Hymenopteran orthomyxo-related virus OKIAV171	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Pelecniidae	<i>Pelecinus polyurator</i>	Honduras: Department Olancho Municipio Orica Batideros	8/10/2012
Hymenopteran orthomyxo-related virus OKIAV173	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Cephalidae	<i>Cephus spinipes</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	5/31/2012
Hymenopteran orthomyxo-related virus OKIAV174	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Siricidae	<i>Tremex magus</i>	Germany: Hessen; Darmstadt	6/7/2013
Hymenopteran orthomyxo-related virus OKIAV175	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Crabronidae	<i>Spilomena beata</i>	Germany: Rhineland-Palatinate; Albersweiler	2012

Virus_name	Figure	Virus_family	Host_subphylum	Host_class	Host_superorder	Host_order	Host_order_group	Host_family	Host_species	Sample_location	Sample_date
Hymenopteran orthomyxo-related virus OKIAV176	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Stilbum cyanurum</i>	Italy: Lombardy Varese Inarzo	7/17/2012
Hymenopteran orthomyxo-related virus OKIAV177	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysis fasciata</i>	Germany: Rhineland-Palatinate; Bellheim	5/20/2012
Lepidopteran orthomyxo-related virus OKIAV1731	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera		Erebidae	<i>Amata phegea</i>	Austria: Lower Austria; Moedling Guntramsdorf Eichkogel	6/2/2011
Lepidopteran orthomyxo-related virus OKIAV178	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphimesmenoptera	Crambidae	<i>Crambus sp.</i>	Germany: Lower Saxony	8/11/2012
Mecopteran orthomyxo-related virus OKIAV197	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Panorpida	Mecoptera	Mecoptera	Nannochoristidae	<i>Nannochorista dipteroides</i>	Australia: Victoria; Bogong High Plains	10/16/2013
Myriapodan orthomyxo-related virus OKIAV203	S12 Fig	<i>Orthomyxoviridae</i>	Myriapoda	Chilopoda	NA	Geophilomorpha	Myriapoda	Geophilidae	<i>Clinopodes flavidus</i>	Austria: Vienna; Donaustadt Lobau	5/2/2013
Neuropteran orthomyxo-related virus OKIAV190	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Mantispidae	<i>Eumantispa harmandi</i>	Japan: Nagano Prefecture; Ueda - Sugadaira Kogen	8/8/2012
Neuropteran orthomyxo-related virus OKIAV210	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Chrysopidae	<i>Peyerimhoffina gracilis</i>	Austria: Lower Austria; Krems-Land Duernstein	6/19/2012
Odonatan orthomyxo-related virus OKIAV208	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Coenagrionidae	<i>Agriocnemis femina oryzae</i>	Japan: Kochi Prefecture; Shimanto	6/7/2012
Odonatan orthomyxo-related virus OKIAV209	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Aeshnidae	<i>Boyeria irene</i>	Italy: Sardinia	9/1/2011
Orthopteran orthomyxo-related virus OKIAV160	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Orthoptera	Orthoptera	Polyneoptera	Gryllacrididae	<i>Nippancistroger testaceus</i>	Japan: Ehime Prefecture; Matsuyama - Kamijiro	6/6/2012
Orthopteran orthomyxo-related virus OKIAV207	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Orthoptera	Orthoptera	Polyneoptera	Pyrgomorphae	<i>Atractomorpha sp.</i>	Germany: Lab culture with samples from Malaysia	1/1/2013
Phasmatodean orthomyxo-related virus OKIAV167	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Exopterygota	Phasmatodea	Polyneoptera	Pseudophasmatidae	<i>Peruphasma schultei</i>	Germany: North Rhine-Westphalia; Duesseldorf Aquazoo	9/15/2011
Phasmatodean orthomyxo-related virus OKIAV172	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Exopterygota	Phasmatodea	Polyneoptera	Phasmatidae	<i>Spathomorpha lancettifer</i>	Madagascar	4/24/2013
Phasmatodean orthomyxo-related virus OKIAV198	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Exopterygota	Phasmatodea	Polyneoptera	Phasmatidae	<i>Megacrania phelaus</i>	Lab culture with samples from Germany; private breeder	4/24/2013
Raphidiopteran orthomyxo-related virus OKIAV180	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Raphidioptera	Neuropterida	Raphidiidae	<i>Puncha ratzeburgi</i>	Austria: Lower Austria; Moedling Guntramsdorf Eichkogel	5/20/2012
Siphonapteran orthomyxo-related virus OKIAV157	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	Endopterygota	Siphonaptera	Siphonaptera	Ceratophyllidae	<i>Ceratophyllus gallinae</i>	Germany: Hamburg; Altona Sternschanzenpark	8/9/2011
Zygentoman orthomyxo-related virus OKIAV204	S12 Fig	<i>Orthomyxoviridae</i>	Hexapoda	Insecta	NA	Zygentoma	Zygentoma	Ateluridae	<i>Atelura formicaria</i>	Austria: Vienna	2011
Coleopteran hanta-related virus OKIAV221	S16 Fig	<i>Hantaviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Diaprepes abbreviatus</i>	USA: Lab culture with samples from unknown location	12/1/2013
Dipluran hanta-related virus OKIAV217	S16 Fig	<i>Hantaviridae</i>	Hexapoda	Entognatha	NA	Diplura	Diplura	Campodeidae	<i>Remycampa launeyi</i>	Spain: Canary Islands; Tenerife	6/14/2012
Dipluran hanta-related virus OKIAV218	S16 Fig	<i>Hantaviridae</i>	Hexapoda	Entognatha	NA	Diplura	Diplura	Campodeidae	<i>Remycampa launeyi</i>	Spain: Canary Islands; Tenerife	6/14/2012
Myriapodan hanta-related virus OKIAV214	S16 Fig	<i>Hantaviridae</i>	Myriapoda	Chilopoda	NA	Geophilomorpha	Myriapoda	Himantariidae	<i>Himantarium gabrielis</i>	Italy: Tuscany Siena; Co. Montalbucco near castle	10/23/2012
Neuropteran peribunya-related virus OKIAV213	S16 Fig	<i>Peribunyaviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Chrysopidae	<i>Peyerimhoffina gracilis</i>	Austria: Lower Austria; Krems-Land Duernstein	6/19/2012
Odonatan hanta-related virus OKIAV206	S16 Fig	<i>Hantaviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Calopterygidae	<i>Mnais pruinosa</i>	Japan: Nagano Prefecture; Chiisagata Aoki	5/16/2012
Plecopteran hanta-related virus OKIAV215	S16 Fig	<i>Hantaviridae</i>	Hexapoda	Insecta	Exopterygota	Plecoptera	Polyneoptera	Perlidae	<i>Perla marginata</i>	Germany: Baden-Wuerttemberg; Black Forest	5/16/2011
Psocodean peribunya-related virus OKIAV212	S16 Fig	<i>Peribunyaviridae</i>	Hexapoda	Insecta	Psocodea	Psocoptera	Psocodea	Amphipsocidae	<i>Amphipsocus japonicus</i>	Japan: Hyogo Prefecture; Kobe	6/25/2012
Blattodean phasma-related virus OKIAV238	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Ectobiidae	<i>Ischnoptera deropeltiformis</i>	Germany: lab culture origin unclear	2/1/2013
Blattodean phasma-related virus OKIAV239	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Ectobiidae	<i>Supella longipalpa</i>	Japan: Lab culture with samples from Thailand	12/28/2011
Coleopteran phasma-related virus OKIAV235	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Coccinellidae	<i>Harmonia axyridis</i>	Germany: Thuringia; Jena - Jena Forest	8/25/2011
Coleopteran phasma-related virus OKIAV236	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Rhipiceridae	<i>Sandalus niger</i>	USA: Tennessee; Tipton County	10/1/2013
Coleopteran phasma-related virus OKIAV241	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Silvanidae	<i>Uleiota planata</i>	Germany: Brandenburg; Rheinsberg Lake	10/15/2011
Coleopteran phasma-related virus OKIAV243	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Drilidae	<i>Drilus concolor</i>	Germany: Thuringia; Jena - Alt-Winzerla	5/14/2011
Collembolan phasma-related virus OKIAV223	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Collembola	NA	Poduromorpha	Ellipura	Onychiuridae	<i>Tetradontophora bielanensis</i>	Germany: Saxony; Goerlitz - Ostritz Neissetal	11/11/2011
Dermapteran phasma-related virus OKIAV237	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	NA	Dermaptera	Polyneoptera	Forficulidae	<i>Apterygida media</i>	Germany: Lower Saxony; Hoehbeck	8/11/2012
Dermapteran phasma-related virus OKIAV240	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	NA	Dermaptera	Polyneoptera	Diplatyidae	<i>Diplatys flavicollis</i>	Japan: Okinawa Prefecture; Ishigaki Island - Miyara	7/13/2012
Dipteran phasma-related virus OKIAV1754	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Mydidae	<i>Miltinus viduatus</i>	Australia: New South Wales; Palerang Nerriga	1/1/2013
Dipteran phasma-related virus OKIAV224	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Syrphidae	<i>Eristalis pertinax</i>	Germany: Rhineland-Palatinate; Ahrschleife bei Ahrweiler	4/25/2013
Dipteran phasma-related virus OKIAV225	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Syrphidae	<i>Ferdinandea cuprea</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	6/6/2013
Dipteran phasma-related virus OKIAV226	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Stratiomyidae	<i>Boreoides subulatus</i>	Australia: Australian Capital Territory; Canberra	2013
Hemipteran phasma-related virus OKIAV245	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Psyllidae	<i>Glycaspis brimblecombei</i>	Australia: South Australia; Adelaide River Torrens	2/20/2012
Hemipteran phasma-related virus OKIAV247	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Acanthosomatidae	<i>Acanthosoma haemorrhoidale</i>	Germany: North Rhine-Westphalia; Rhein-Sieg-Kreis	5/17/2011
Hymenopteran phasma-related virus OKIAV227	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysis fasciata</i>	Germany: Rhineland-Palatinate; Bellheim	5/20/2012
Hymenopteran phasma-related virus OKIAV228	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Philoctetes bogdanovii</i>	Italy: Valle d'Aosta; Pongel	6/17/2012

Virus_name	Figure	Virus_family	Host_subphylum	Host_class	Host_superorder	Host_order	Host_order_group	Host_family	Host_species	Sample_location	Sample_date
Hymenopteran phasma-related virus OKIAV229	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Vespidae	<i>Celonites abbreviatus</i>	Italy: Valle de Cogne; Lillaz	7/16/2011
Hymenopteran phasma-related virus OKIAV230	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Vespidae	<i>Quartinia thebaica</i>	Israel: Wadi Ze' elim	5/25/2012
Hymenopteran phasma-related virus OKIAV231	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Halicidae	<i>Sphecodes albilabris</i>	Germany: Rhineland-Palatinate; Hochstadt -Pfalz	4/2/2011
Hymenopteran phasma-related virus OKIAV232	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Megachilidae	<i>Osmia cornuta</i>	Germany: Rhineland-Palatinate; Albersweiler	4/1/2011
Hymenopteran phasma-related virus OKIAV233	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Apidae	<i>Bombus rupestris</i>	France: Alpes-Maritimes Millefontes	7/12/2011
Hymenopteran phasma-related virus OKIAV234	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Megachilidae	<i>Coelioxys conoidea</i>	Germany: Rhineland-Palatinate; Birkenheide	7/5/2011
Hymenopteran phasma-related virus OKIAV244	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Torymidae	<i>Podagrion pachymerum</i>	Slovakia: Rudno nad Hronom	1/1/2012
Hymenopteran phasma-related virus OKIAV249	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysis succinta (succintula)</i>	Italy: Valle d' Aosta; Ozein	6/21/2012
Hymenopteran phasma-related virus OKIAV250	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysis gracillima</i>	Germany: Rhineland-Palatinate; Buechberg	7/1/2012
Hymenopteran phasma-related virus OKIAV251	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysura austriaca</i>	Germany: Rhineland-Palatinate; Battenberg	7/3/2010
Hymenopteran phasma-related virus OKIAV252	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysura cuprea</i>	Italy: Valle d' Aosta; Pondel	6/19/2012
Lepidopteran phasma-related virus OKIAV246	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Hesperiidae	<i>Urbanus proteus</i>	USA: Florida; Gainesville	11/2/2012
Neuropteran phasma-related virus OKIAV248	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Chrysopidae	<i>Peyerimhoffina gracilis</i>	Austria: Lower Austria; Krems-Land Duernstein	6/19/2012
Odonatan phasma-related virus OKIAV254	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Coenagrionidae	<i>Enallagma aspersum</i>	USA: Tennessee; Campbell County - Caryville	6/22/2011
Zorapteran phasma-related virus OKIAV242	S21 Fig	<i>Phasmaviridae</i>	Hexapoda	Insecta	NA	Zoraptera	Polyneoptera	Zorotypidae	<i>Zorotypus caudelli</i>	Malaysia: Selangor; Gombak	4/9/2011
Archaeognathan phenui-related virus OKIAV269	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	NA	Archaeognatha	Archaeognatha	Machilidae	<i>Trigoniphthalmus alternatus</i>	USA: California; Angelo Coast Range Reserve	8/4/2011
Blattodean phenui-related virus OKIAV261	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Cryptoceridae	<i>Cryptocercus wrighti</i>	USA: Oregon	2/1/2012
Blattodean phenui-related virus OKIAV266	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Ectobiidae	<i>Ectobius sylvestris</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	6/14/2013
Blattodean phenui-related virus OKIAV267	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Ectobiidae	<i>Balta vilis</i>	Germany: Lab culture with samples originating from Germany	3/20/2013
Blattodean phenui-related virus OKIAV268	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Blattidae	<i>Deropeltis erythrocephala</i>	Germany: Lab culture with samples originating from Germany	3/13/2011
Blattodean phenui-related virus OKIAV294	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Tryonicidae	<i>Tryonicus parvus</i>	Australia: Queensland; Tullawallal Lamington National Park	3/14/2013
Coleopteran phenui-related virus OKIAV264	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Haliplidae	<i>Haliphys fluviatilis</i>	Germany: Mecklenburg-Hither Pomerania; Fuerstenhagen	5/7/2011
Coleopteran phenui-related virus OKIAV287	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Sitophilus zeamais</i>	USA: Tennessee; Tipton County	6/1/2013
Coleopteran phenui-related virus OKIAV293	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Curculionidae	<i>Larinus minutus</i>	USA: Arkansas; Fayetteville University of Arkansas	11/1/2013
Coleopteran phenui-related virus OKIAV308	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Endopterygota	Coleoptera	Coleoptera	Coccinellidae	<i>Stethorus sp.</i>	Australia: Queensland; Mount Tamborine	10/31/2013
Crustacean phenui-related virus OKIAV307	S25 Fig	<i>Phenuiviridae</i>	Crustacea	Maxillopoda	Thoracica	Sessilia	Crustacea	Archaeobalanidae	<i>Semibalanus balanoides</i>	Norway: Hordaland; Store Lungegardsvannet	1/20/2013
Dipteran phenui-related virus OKIAV273	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Chaoboridae	<i>Chaoborus flavidulus</i>	Singapore: Upper Seletar Reservoir Park	7/1/2013
Dipteran phenui-related virus OKIAV274	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Bibionidae	<i>Bibio marci</i>	Germany: North Rhine-Westphalia; Bonn ZFMK	4/5/2011
Dipteran phenui-related virus OKIAV281	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Trichoceridae	<i>Trichocera saltator</i>	Germany: Thuringia; Jena	11/14/2011
Hemipteran phenui-related virus OKIAV272	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Cicadidae	<i>Okanagana villosa</i>	USA: California; Tuolumne-Stanislaus National Forest	8/12/2011
Hemipteran phenui-related virus OKIAV285	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Paraneoptera	Hemiptera	Hemiptera	Cicadellidae	<i>Graphocephala fennahi</i>	Germany: Hamburg; Hamburg-Harburg	8/10/2011
Hymenopteran phenui-related virus OKIAV255	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Agaonidae	<i>Elisabethiella stueckenbergi</i>	South Africa: Western Cape; Westlake	3/13/2013
Hymenopteran phenui-related virus OKIAV275	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Cynipidae	<i>Synergus umbraculus</i>	Germany: North Rhine-Westphalia; Bonn	2012
Hymenopteran phenui-related virus OKIAV282	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Vespidae	<i>Symmorphus murarius</i>	Germany: Rhineland-Palatinate; Fischbach near Dahn	6/22/2012
Hymenopteran phenui-related virus OKIAV296	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Cephidae	<i>Cephus spinipes</i>	Germany: North Rhine-Westphalia; Bonn Kottenforst	5/31/2012
Hymenopteran phenui-related virus OKIAV297	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Chrysididae	<i>Chrysis ehrenbergi</i>	Israel: Negev Revivim	6/3/2012
Hymenopteran phenui-related virus OKIAV298	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Braconidae	<i>Macrocentrus marginator</i>	Sweden: Uppland Hultmanstorp	8/25/2012
Hymenopteran phenui-related virus OKIAV299	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Braconidae	<i>Macrocentrus marginator</i>	Sweden: Uppland Hultmanstorp	8/25/2012
Hymenopteran phenui-related virus OKIAV302	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Ceraphronidae	<i>Ceraphron sp.</i>	Germany: North Rhine-Westphalia; Bonn ZFMK	6/1/2013
Hymenopteran phenui-related virus OKIAV305	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Pteromalidae	<i>Spalangia cameroni</i>	Lab culture of unknown geographical origin	8/10/2012
Hymenopteran phenui-related virus OKIAV306	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Hymenoptera	Bradynobaenidae	<i>Chyphotes sp.</i>	USA: Arizona; Willcox	8/11/2011
Lepidopteran phenui-related virus OKIAV262	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Hepialidae	<i>Triodia sylvina</i>	Germany: Lower Saxony; Hoehbeck - Pevestorf - Station	8/27/2011
Lepidopteran phenui-related virus OKIAV270	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Crambidae	<i>Crambus sp.</i>	Germany: Lower Saxony	8/11/2012

Virus_name	Figure	Virus_family	Host_subphylum	Host_class	Host_superorder	Host_order	Host_order_group	Host_family	Host_species	Sample_location	Sample_date
Lepidopteran phenui-related virus OKIAV271	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Panorpida	Lepidoptera	Amphiesmenoptera	Nymphalidae	<i>Pararge aegeria</i>	Germany: North Rhine-Westphalia; Rheinland Kottenforst	5/10/2011
Mantodean phenui-related virus OKIAV283	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Dictyoptera	Mantodea	Polyneoptera	Mantidae	<i>Mantis religiosa</i>	Germany: Rhineland-Palatinate; Albersweiler	8/20/2011
Megalopteran phenui-related virus OKIAV286	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Endopterygota	Megaloptera	Neuropterida	Corydalidae	<i>Chauliodes sp.</i>	USA: Missouri; Clinton County - Wallace State Park	6/22/2011
Neuropteran phenui-related virus OKIAV300	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Chrysopidae	<i>Peyerimhoffina gracilis</i>	Austria: Lower Austria; Krems-Land Duernstein	6/19/2012
Neuropteran phenui-related virus OKIAV304	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Hemerobiidae	<i>Hemerobius handschini</i>	Austria: Lower Austria; Moedling Guntramsdorf Eichkogel	5/18/2013
Neuropteran phenui-related virus OKIAV309	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Neuropterida	Neuroptera	Neuropterida	Chrysopidae	<i>Pseudomallada flavifrons flavifrons</i>	Austria: Lower Austria; Krems-Land Duernstein	6/19/2012
Odonatan phenui-related virus OKIAV258	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Coenagrionidae	<i>Coenagrion puella</i>	Germany: North Rhine-Westphalia; Rheinland Nature Park	5/10/2011
Odonatan phenui-related virus OKIAV265	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Gomphidae	<i>Paragomphus genei</i>	Israel	5/24/2011
Odonatan phenui-related virus OKIAV277	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Petaluridae	<i>Tanypteryx pryeri</i>	Japan: Ibaraki Prefecture; Hitachi-Omiya	5/29/2012
Odonatan phenui-related virus OKIAV278	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Gomphidae	<i>Sieboldius albardae</i>	Japan: Kyoto	7/23/2012
Odonatan phenui-related virus OKIAV279	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Odonatoptera	Odonata	Odonata	Gomphidae	<i>Nihonogomphus viridis</i>	Japan: Kochi Prefecture; Shimanto	5/23/2012
Plecopteran phenui-related virus OKIAV280	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Exopterygota	Plecoptera	Polyneoptera	Perlidae	<i>Perla marginata</i>	Germany: Baden-Wuerttemberg; Black Forest	5/16/2011
Psocodean phenui-related virus OKIAV301	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Psocodea	Psocoptera	Psocodea	Peripsocidae	<i>Peripsocus phaeopterus</i>	Japan: Hokkaido; Sapporo	6/28/2012
Trichopteran phenui-related virus OKIAV263	S25 Fig	<i>Phenuiviridae</i>	Hexapoda	Insecta	Amphiesmenoptera	Trichoptera	Amphiesmenoptera	Rhyacophilidae	<i>Himalopsyche phryganea</i>	USA: Oregon; Benton County - Parker Creek Falls	6/23/2013
Blattodean nairo-related virus OKIAV321	S30 Fig	<i>Nairoviridae</i>	Hexapoda	Insecta	Dictyoptera	Blattodea	Polyneoptera	Ectobiidae	<i>Paratemnopteryx coulöniana</i>	Germany: Lab culture with samples from Germany; private breeder	3/13/2011
Dipteran mypo-related virus OKIAV322	S30 Fig	<i>Mypoviridae</i>	Hexapoda	Insecta	Panorpida	Diptera	Diptera	Dolichopodidae	<i>Heteropsilopus ingenuus</i>	Australia: Canberra; Australian National Botanic Gardens	1/18/2012
Myriapodan Negavirus OKIAV319	S30 Fig	unclassified	Myriapoda	Symphyla	NA	NA	Myriapoda	Scolopendrellidae	<i>Symphylella sp.</i>	Japan: Nagano Prefecture; Ueda-city - Shinko-ji	7/1/2012
Myriapodan Negavirus OKIAV320	S30 Fig	unclassified	Myriapoda	Symphyla	NA	NA	Myriapoda	Scolopendrellidae	<i>Symphylella sp.</i>	Japan: Nagano Prefecture; Ueda-city - Shinko-ji	7/1/2012

S2 Table. Continues in next page.

Virus name	Total_number of_reads	Number_of reads_mapped	%_of_reads mapped
Hymenopteran almendra-related virus OKIAV1 segment NA	19473772	4229	0.0217
Hymenopteran rhabdo-related virus OKIAV8 segment NA	21466468	69766	0.325
Lepidopteran rhabdo-related virus OKIAV12 segment NA	19953700	7099	0.0356
Blattodean rhabdo-related virus OKIAV14 segment NA	21695908	7219	0.0333
Mantodean rhabdo-related virus OKIAV15 segment NA	16564411	9668	0.0584
Dipteran rhabdo-related virus OKIAV19 segment NA	38732882	17554	0.0453
Coleopteran rhabdo-related virus OKIAV20 segment NA	50968121	46473	0.0912
Hymenopteran rhabdo-related virus OKIAV22 segment NA	22113490	55177	0.2495
Hymenopteran rhabdo-related virus OKIAV23 segment NA	14527002	9823	0.0676
Hymenopteran rhabdo-related virus OKIAV24 segment NA	14715518	15903	0.1081
Coleopteran rhabdo-related virus OKIAV28 segment NA	21130680	20208	0.0956
Lepidopteran rhabdo-related virus OKIAV34 segment NA	21879113	15023	0.0687
Hymenopteran rhabdo-related virus OKIAV38 segment NA	26293950	10983	0.0418
Hymenopteran rhabdo-related virus OKIAV46 segment NA	16344486	2348	0.0144
Odonatan anphe-related virus OKIAV57 segment NA	19312668	6554	0.0339
Odonatan anphe-related virus OKIAV59 segment NA	21159020	4013	0.019
Hymenopteran anphe-related virus OKIAV71 segment NA	18516658	11013	0.0595
Hymenopteran orino-related virus OKIAV85 segment NA	22040446	15243	0.0692
Hymenopteran orino-related virus OKIAV87 segment NA	17630247	9137	0.0518
Odonatan anphe-related virus OKIAV90 segment NA	17749949	19863	0.1119
Mantodean anphe-related virus OKIAV92 segment NA	27783925	4010	0.0144
Hemipteran arli-related virus OKIAV94 segment NA	19436766	5945	0.0306
Hymenopteran arli-related virus OKIAV98 segment NA	20462523	17320	0.0846
Hymenopteran arli-related virus OKIAV99 segment NA	21149067	9150	0.0433
Strepsipteran arli-related virus OKIAV104 segment NA	21939006	11346	0.0517
Hymenopteran rhabdo-related virus OKIAV109 segment NA	21585225	18837	0.0873
Collembolan qin-related virus OKIAV112 segment L	24543499	3566	0.0145
Collembolan qin-related virus OKIAV112 segment N	24543499	103	0.0004
Collembolan qin-related virus OKIAV112 segment G	24543499	129	0.0005
Megalopteran chu-related virus OKIAV119 segment NA	4668284	3980	0.0853
Hymenopteran chu-related virus OKIAV123 segment NA	23744192	8017	0.0338
Hymenopteran chu-related virus OKIAV124 segment NA	16339414	6672	0.0408
Hymenopteran chu-related virus OKIAV126 segment NA	23948314	6373	0.0266
Odonatan chu-related virus OKIAV136 segment NA	22275367	4153	0.0186
Odonatan chu-related virus OKIAV137 segment NA	26273745	2654	0.0101
Dermapteran chu-related virus OKIAV142 segment L-G	17954649	1832	0.0102
Dermapteran chu-related virus OKIAV142 segment N	17954649	1200	0.0067
Dipteran orthomyxo-related virus OKIAV164 segment 1	18211096	590	0.0032
Dipteran orthomyxo-related virus OKIAV164 segment 2	18211096	329	0.0018
Dipteran orthomyxo-related virus OKIAV164 segment 3	18211096	565	0.0031
Dipteran orthomyxo-related virus OKIAV164 segment 4	18211096	2817	0.0155
Dipteran orthomyxo-related virus OKIAV164 segment 5	18211096	862	0.0047
Dipteran orthomyxo-related virus OKIAV166 segment 1	29706480	2351	0.0079
Dipteran orthomyxo-related virus OKIAV166 segment 2	29706480	3758	0.0127
Dipteran orthomyxo-related virus OKIAV166 segment 3	29706480	2860	0.0096
Dipteran orthomyxo-related virus OKIAV166 segment 4	29706480	2972	0.01
Dipteran orthomyxo-related virus OKIAV166 segment 5	29706480	3774	0.0127

S2 Table: Continues from previous page.

Virus name	Total_number _of_reads	Number_of _reads_mapped	%_of_reads _mapped
Phasmatodean orthomyxo-related virus OKIAV172 segment 1	15242342	7732	0.0507
Phasmatodean orthomyxo-related virus OKIAV172 segment 2	15242342	1196	0.0078
Phasmatodean orthomyxo-related virus OKIAV172 segment 3	15242342	1177	0.0077
Phasmatodean orthomyxo-related virus OKIAV172 segment 4	15242342	1000	0.0066
Phasmatodean orthomyxo-related virus OKIAV172 segment 5	15242342	302	0.002
Lepidopteran orthomyxo-related virus OKIAV178 segment 1	17344584	1629	0.0094
Lepidopteran orthomyxo-related virus OKIAV178 segment 2	17344584	2238	0.0129
Lepidopteran orthomyxo-related virus OKIAV178 segment 3	17344584	3390	0.0195
Lepidopteran orthomyxo-related virus OKIAV178 segment 4	17344584	1984	0.0114
Lepidopteran orthomyxo-related virus OKIAV178 segment 5	17344584	2425	0.014
Hemipteran orthomyxo-related virus OKIAV188 segment 1	22049956	333	0.0015
Hemipteran orthomyxo-related virus OKIAV188 segment 2	22049956	228	0.001
Hemipteran orthomyxo-related virus OKIAV188 segment 3	22049956	588	0.0027
Hemipteran orthomyxo-related virus OKIAV188 segment 4	22049956	624	0.0028
Hemipteran orthomyxo-related virus OKIAV188 segment 5	22049956	708	0.0032
Psocodean peribunya-related virus OKIAV212 segment L	19173232	976	0.0051
Myriapodan hanta-related virus OKIAV214 segment L	24095723	1637	0.0068
Plecopteran hanta-related virus OKIAV215 segment L	21920915	3512	0.016
Collembolan phasma-related virus OKIAV223 segment L	22441125	1614	0.0072
Hymenopteran phasma-related virus OKIAV227 segment L	24738193	1414	0.0057
Hymenopteran phasma-related virus OKIAV227 segment S	24738193	506	0.002
Hymenopteran phasma-related virus OKIAV227 segment M	24738193	1486	0.006
Hymenopteran phasma-related virus OKIAV228 segment L	42542851	2534	0.006
Hymenopteran phasma-related virus OKIAV228 segment S	42542851	12243	0.0288
Hymenopteran phasma-related virus OKIAV228 segment M	42542851	942	0.0022
Hemipteran phasma-related virus OKIAV247 segment L	16027600	9503	0.0593
Hemipteran phasma-related virus OKIAV247 segment S	16027600	1702	0.0106
Hemipteran phasma-related virus OKIAV247 segment M	16027600	4853	0.0303
Neuropteran phasma-related virus OKIAV248 segment L	26381665	6034	0.0229
Neuropteran phasma-related virus OKIAV248 segment M	26381665	3421	0.013
Neuropteran phasma-related virus OKIAV248 segment S	26381665	2663	0.0101
Hymenopteran phasma-related virus OKIAV250 segment L	29612827	12661	0.0428
Hymenopteran phasma-related virus OKIAV250 segment S	29612827	1471	0.005
Hymenopteran phasma-related virus OKIAV250 segment M	29612827	8748	0.0295
Hymenopteran phasma-related virus OKIAV252 segment L	15371479	2278	0.0148
Hymenopteran phasma-related virus OKIAV252 segment S	15371479	1045	0.0068
Hymenopteran phasma-related virus OKIAV252 segment M	15371479	2194	0.0143
Dipteran phenui-related virus OKIAV273 segment L	38842361	790	0.002
Dipteran phenui-related virus OKIAV273 segment M	38842361	2612	0.0067
Dipteran phenui-related virus OKIAV273 segment S	38842361	3352	0.0086
Hymenopteran phenui-related virus OKIAV306 segment L	14977881	110	0.0007
Hymenopteran phenui-related virus OKIAV306 segment M	14977881	959	0.0064
Hymenopteran phenui-related virus OKIAV306 segment S	14977881	6869	0.0459
Coleopteran phenui-related virus OKIAV308 segment L	41189168	3878	0.0094
Coleopteran phenui-related virus OKIAV308 segment M	41189168	12233	0.0297
Coleopteran phenui-related virus OKIAV308 segment S	41189168	1564	0.0038
Blattodean nairo-related virus OKIAV321 segment L	18685950	18423	0.0986
Blattodean nairo-related virus OKIAV321 segment M	18685950	3731	0.02

S2 Table: Continues from previous page.

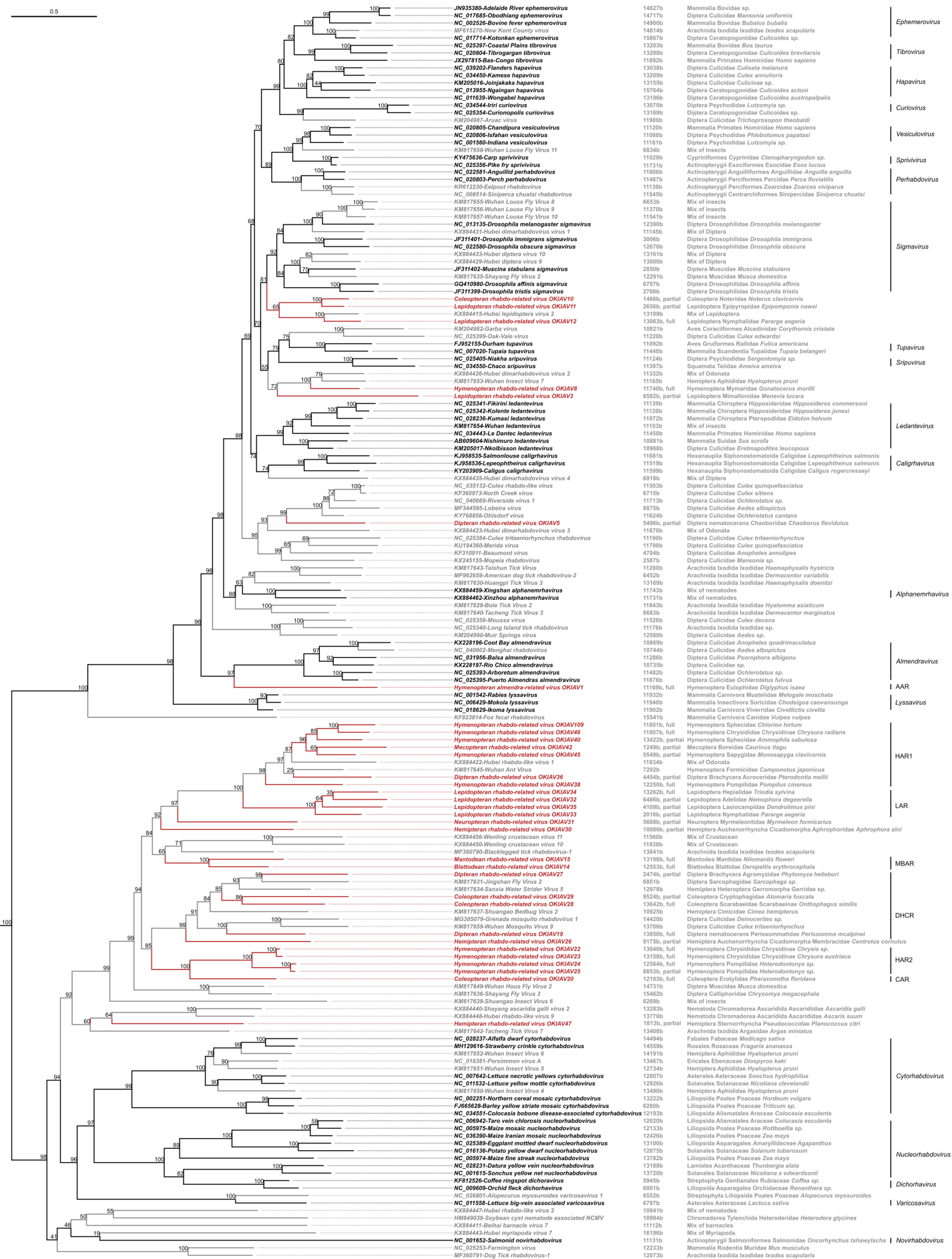
Virus name	Total_number_of_reads	Number_of_reads_mapped	%_of_reads_mapped
Blattodean nairo-related virus OKIAV321 segment S	18685950	4326	0.0232
min	4668284	103	0.0004197
max	50968121	69766	0.325
median	21525846.5	3405.5	0.0144
average	22988826.3	6784.4	0.0312

S3 Table.

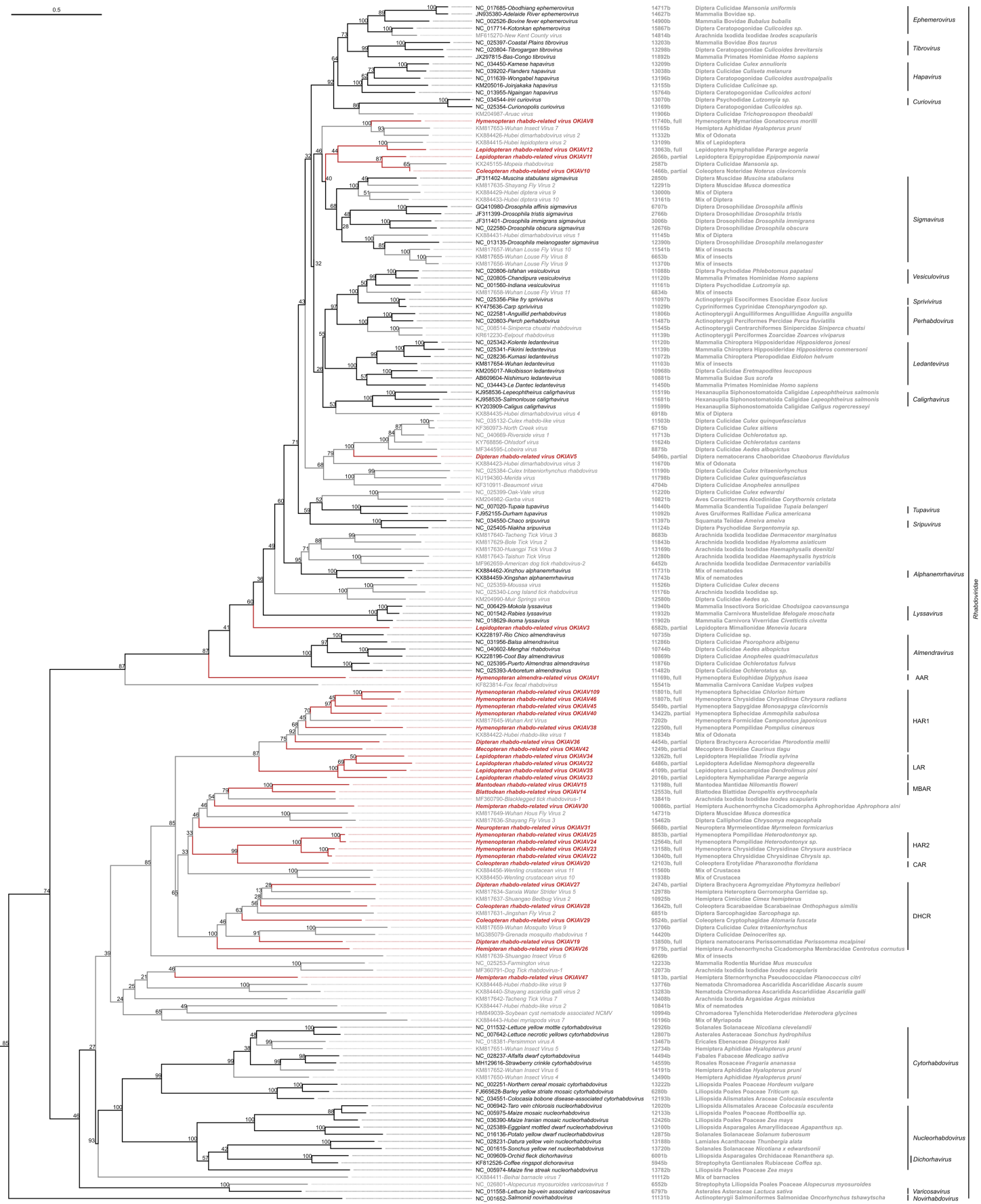
Phylum	Class
Annelida	Clitellata
Annelida	Polychaeta
Arthropoda	Arachnida
Arthropoda	Branchiopoda
Arthropoda	Chilopoda
Arthropoda	Diplopoda
Arthropoda	Hexanauplia
Arthropoda	Malacostraca
Arthropoda	Ostracoda
Ascomycota	Dothideomycetes
Ascomycota	Eurotiomycetes
Ascomycota	Saccharomycetes
Chordata	Actinopteri
Chordata	Aves
Chordata	Mammalia
Echinodermata	Echinoidea
Mollusca	Bivalvia
Mollusca	Gastropoda
Platyhelminthes	Rhabditophora
Rotifera	Monogononta

S4 Table.

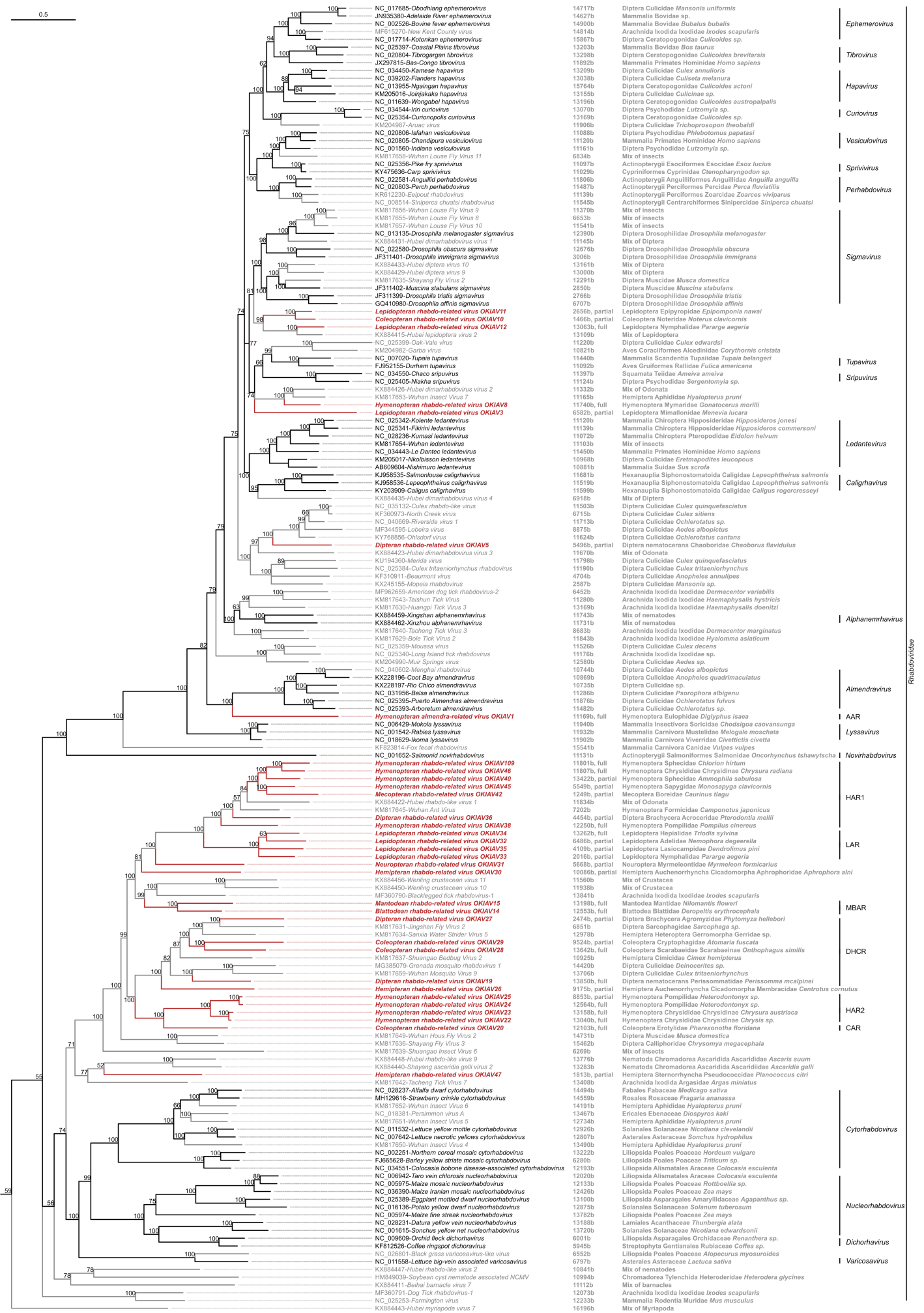
Assembly ID	OKIAV(s)	Phylum	Class	GBOL		NCBI	
				Identity	Length	Identity	Length
INSlupTAXRAAPEI-11	Lepidopteran phenui-related virus OKIAV270						
	Lepidopteran orthomyxo-related virus OKIAV178	Mollusca	Bivalvia			99.82	566
	Lepidopteran Negavirus OKIAV1789*						
INSnfrTABRAAPEI-14	Hemipteran orthomyxo-related virus OKIAV211	Mollusca	Gastropoda	99.54	655	99.53	633
	Hemipteran Negavirus OKIAV1377*	Annelida	Clitellata	99.7	658	98.94	658
	Hemipteran Negavirus OKIAV1782*	Arthropoda	Branchiopoda			99.26	672
	Hemipteran Negavirus OKIAV312*						
WHANIsrmTMCNRAAPEI-13	Coleopteran hanta-related virus OKIAV221	Chordata	Mammalia			100	716
INSofmTBCRAAPEI-37	Psocodean Negavirus OKIAV1356*	Ascomycota	Eurotiomycetes			100	1311



S1 Fig.



S2 Fig.

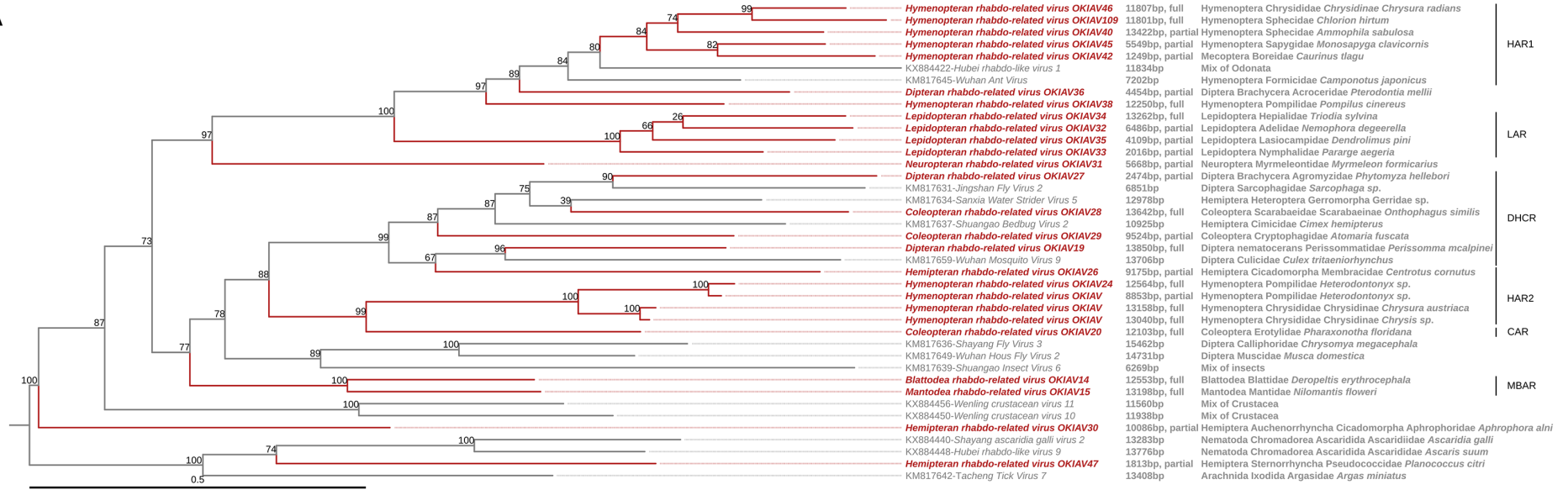


S3 Fig.

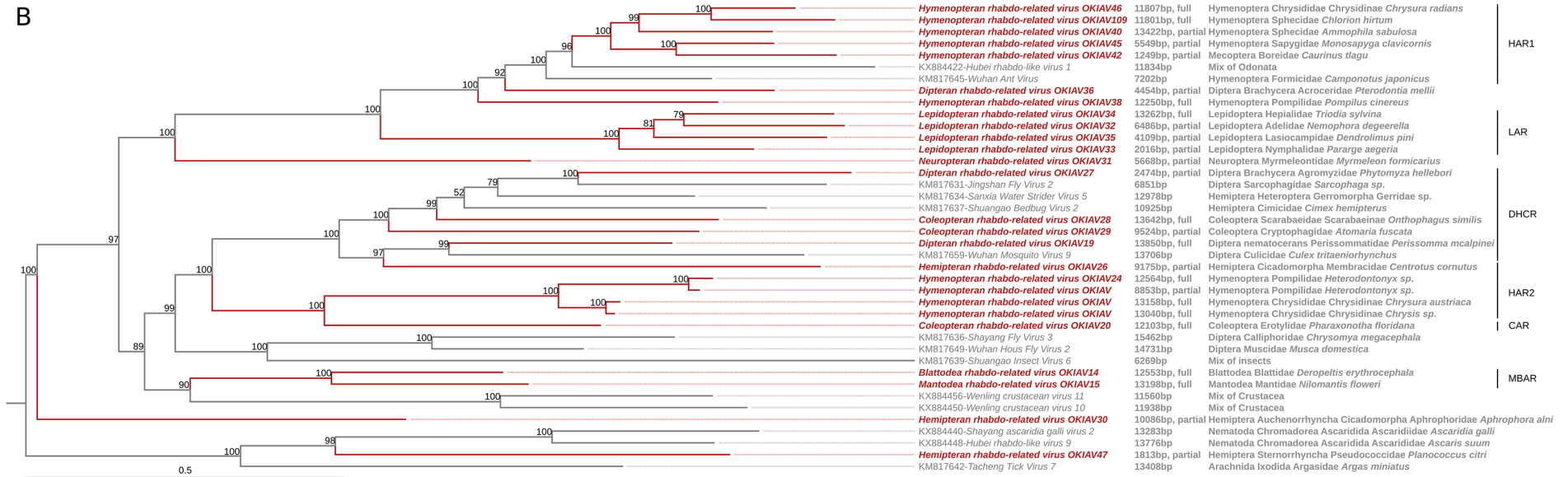
NC_025393-ABTV
 NC_025395-PTAMV
 KX228196-CBV
 NC_031956-BALV
 KX228197-RCHV
 HARV OKIAV1

S4 Fig.

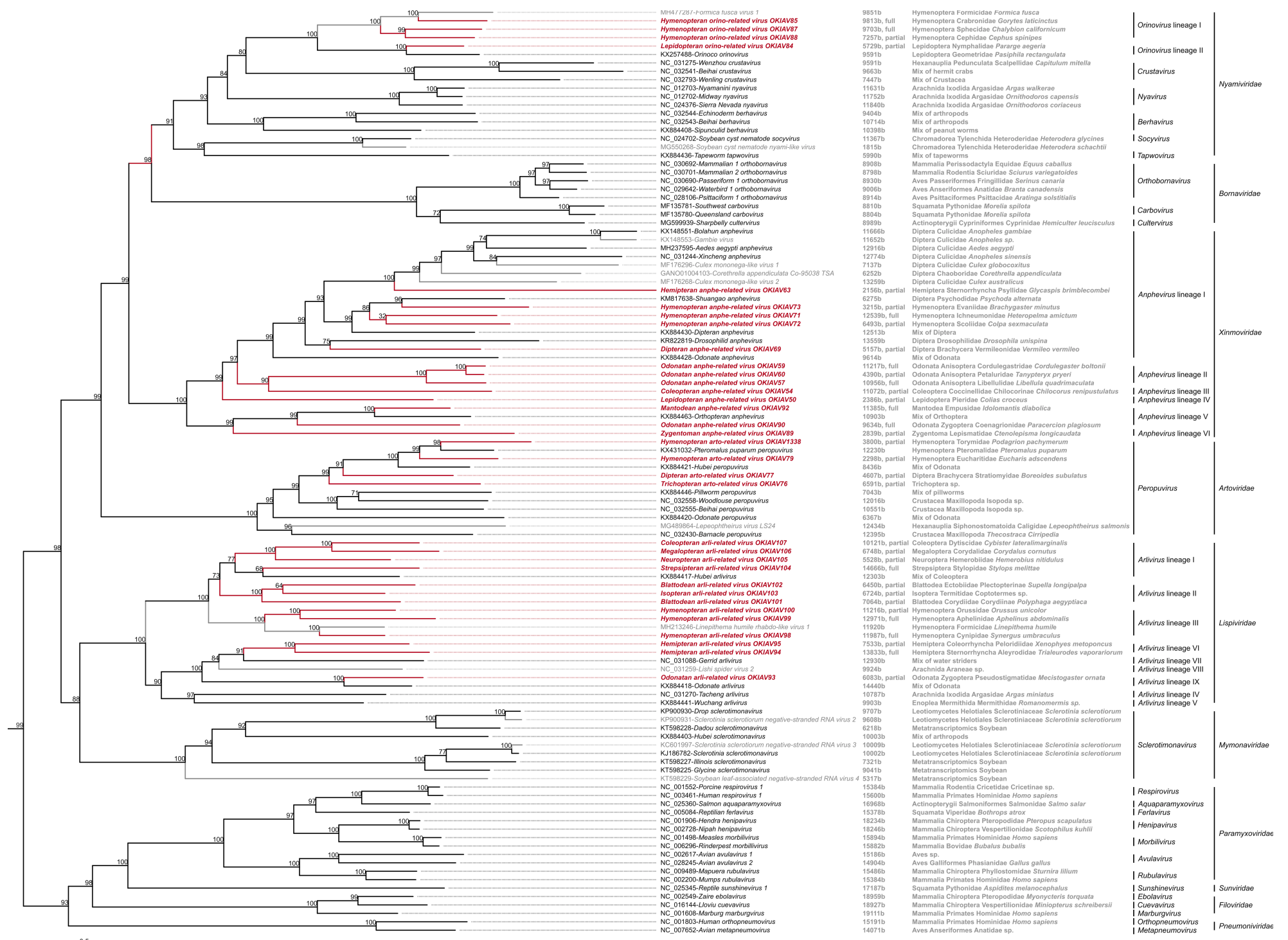
A



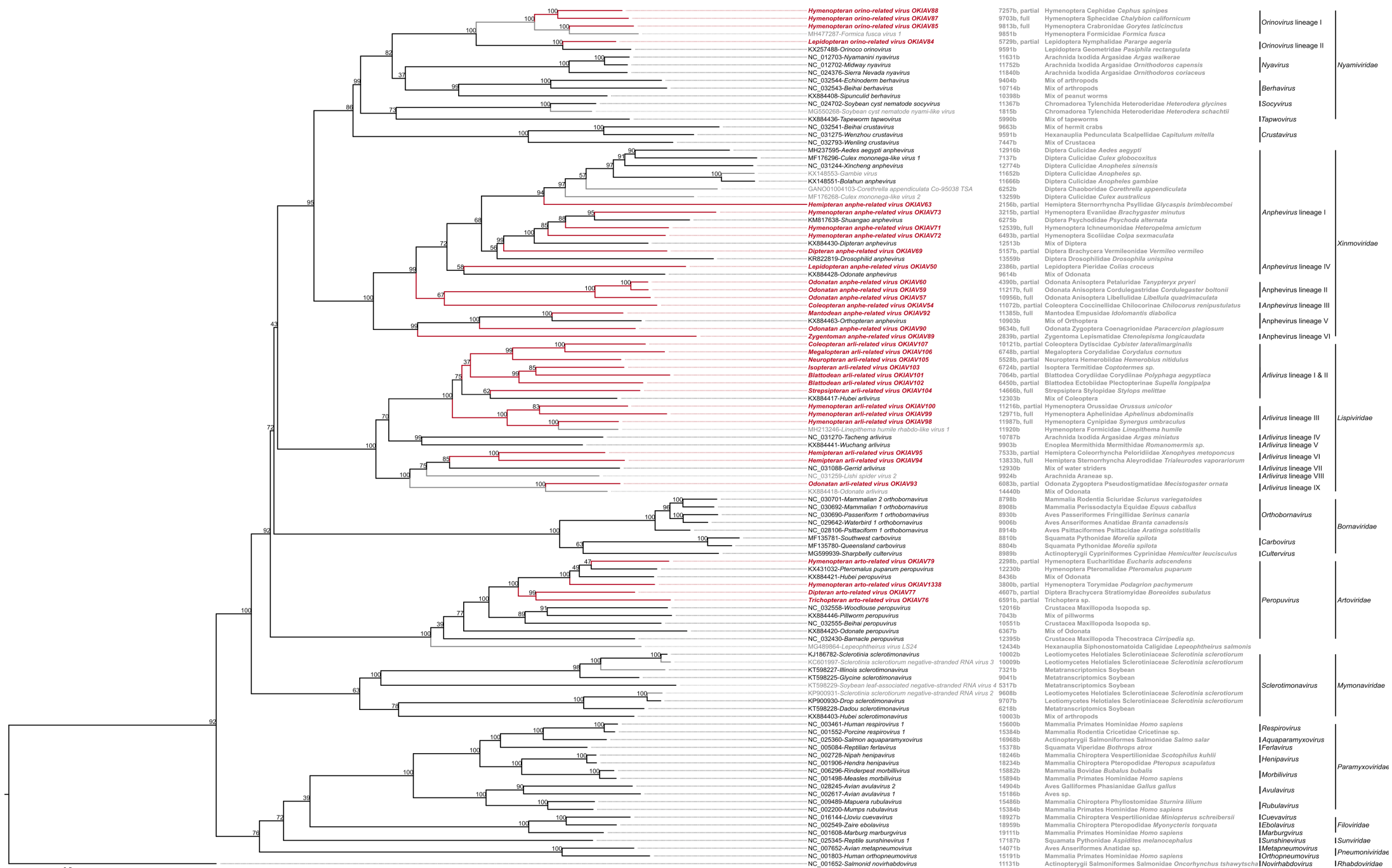
B



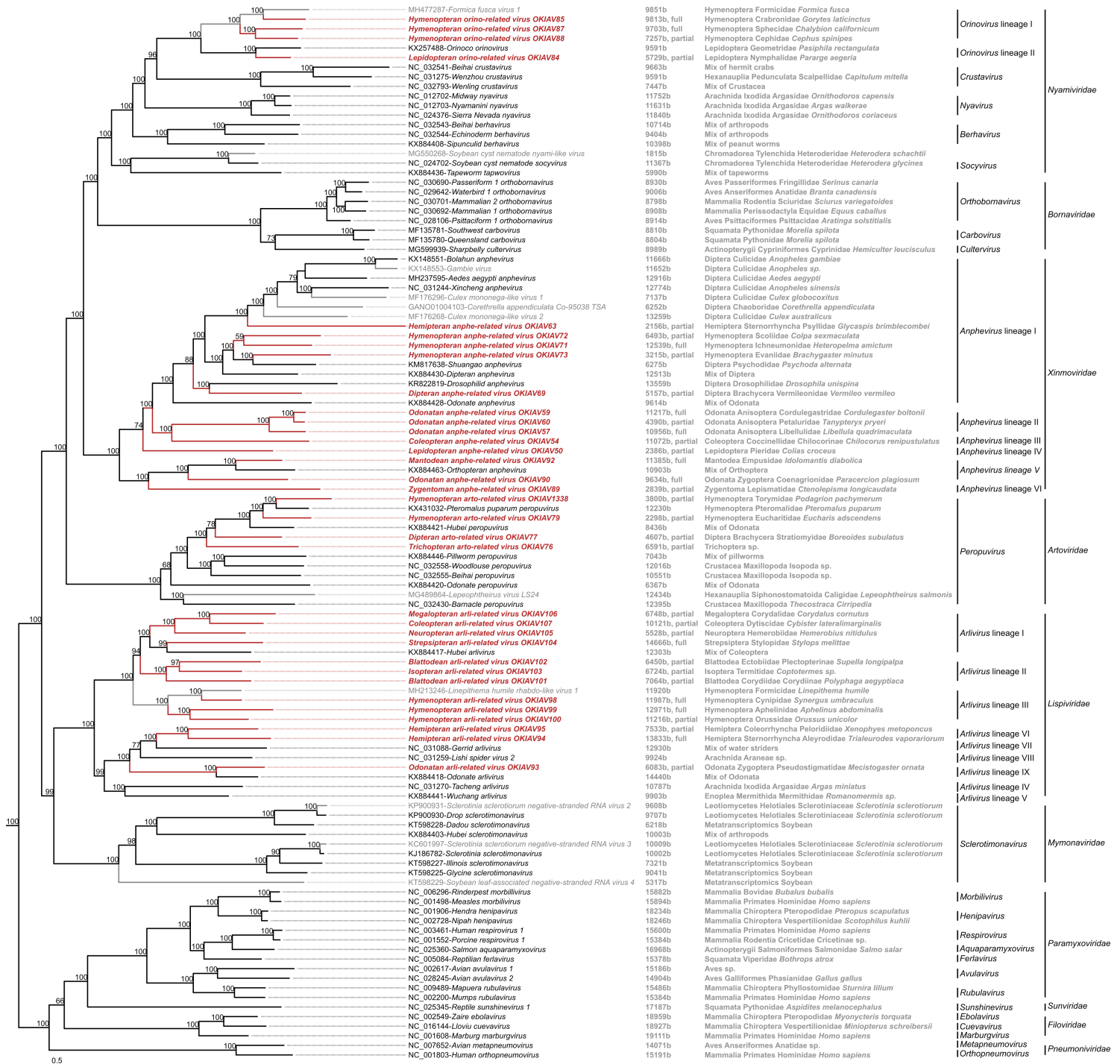
S5 Fig.



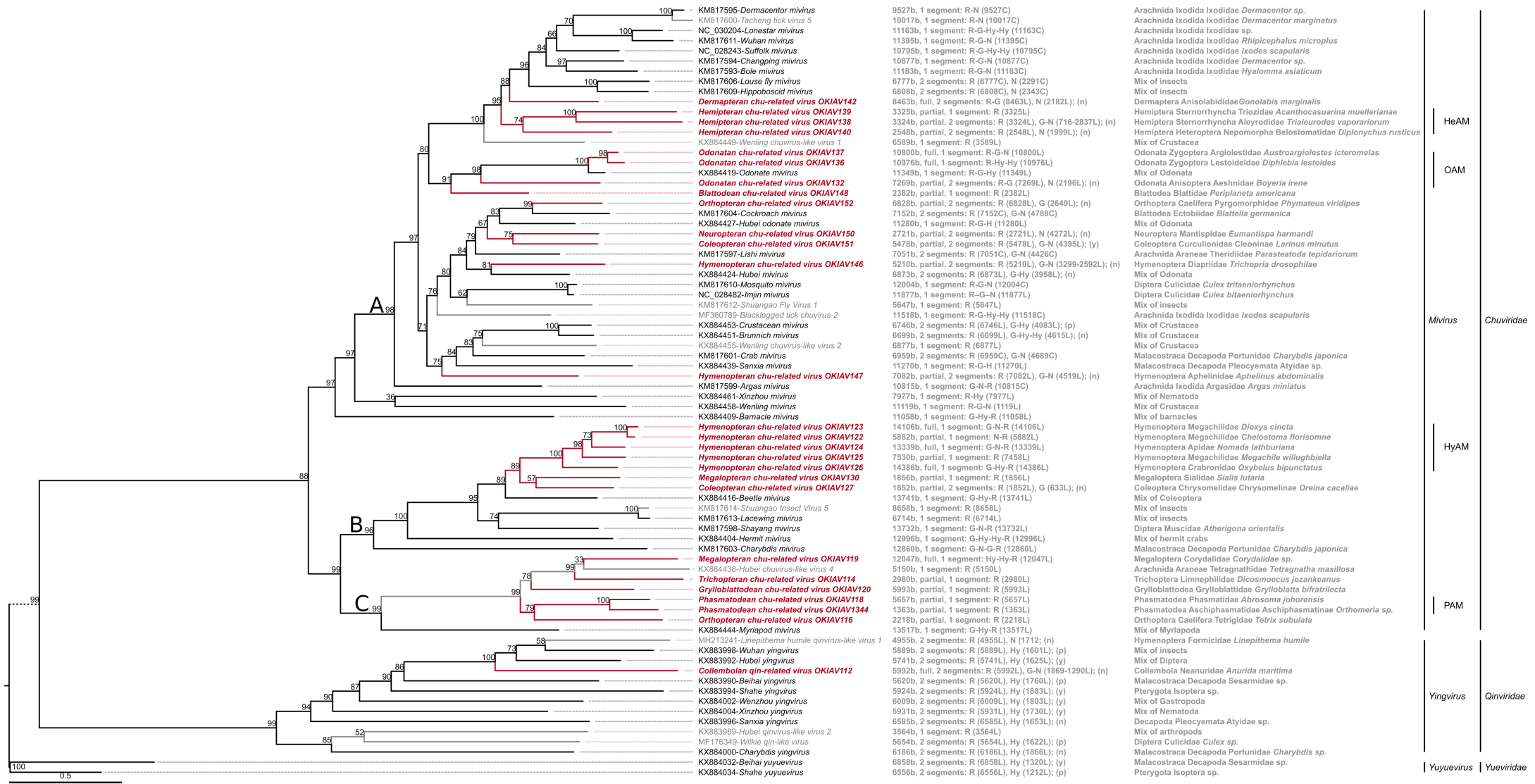
S6 Fig.



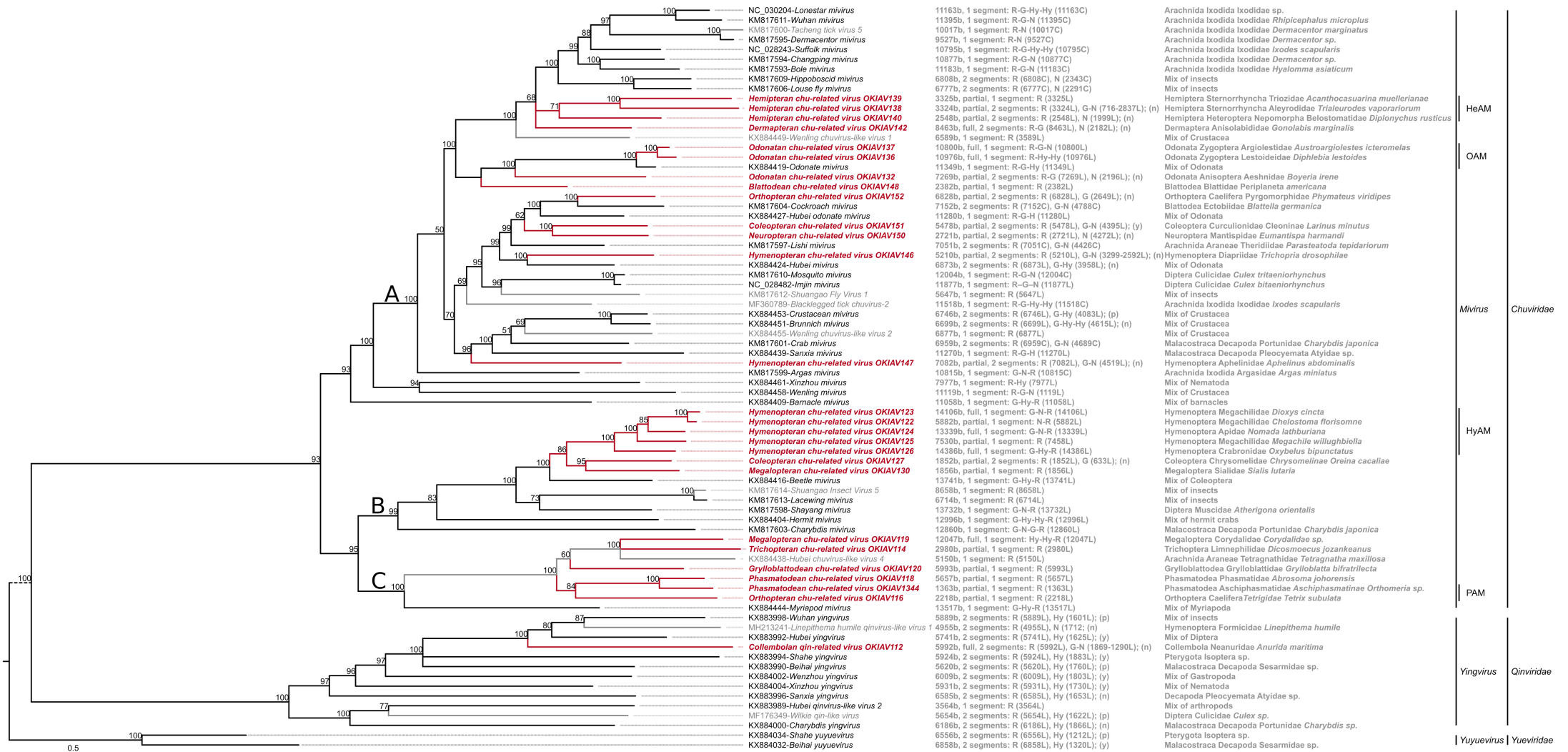
S7 Fig.



S8 Fig.



S9 Fig.



S11 Fig.



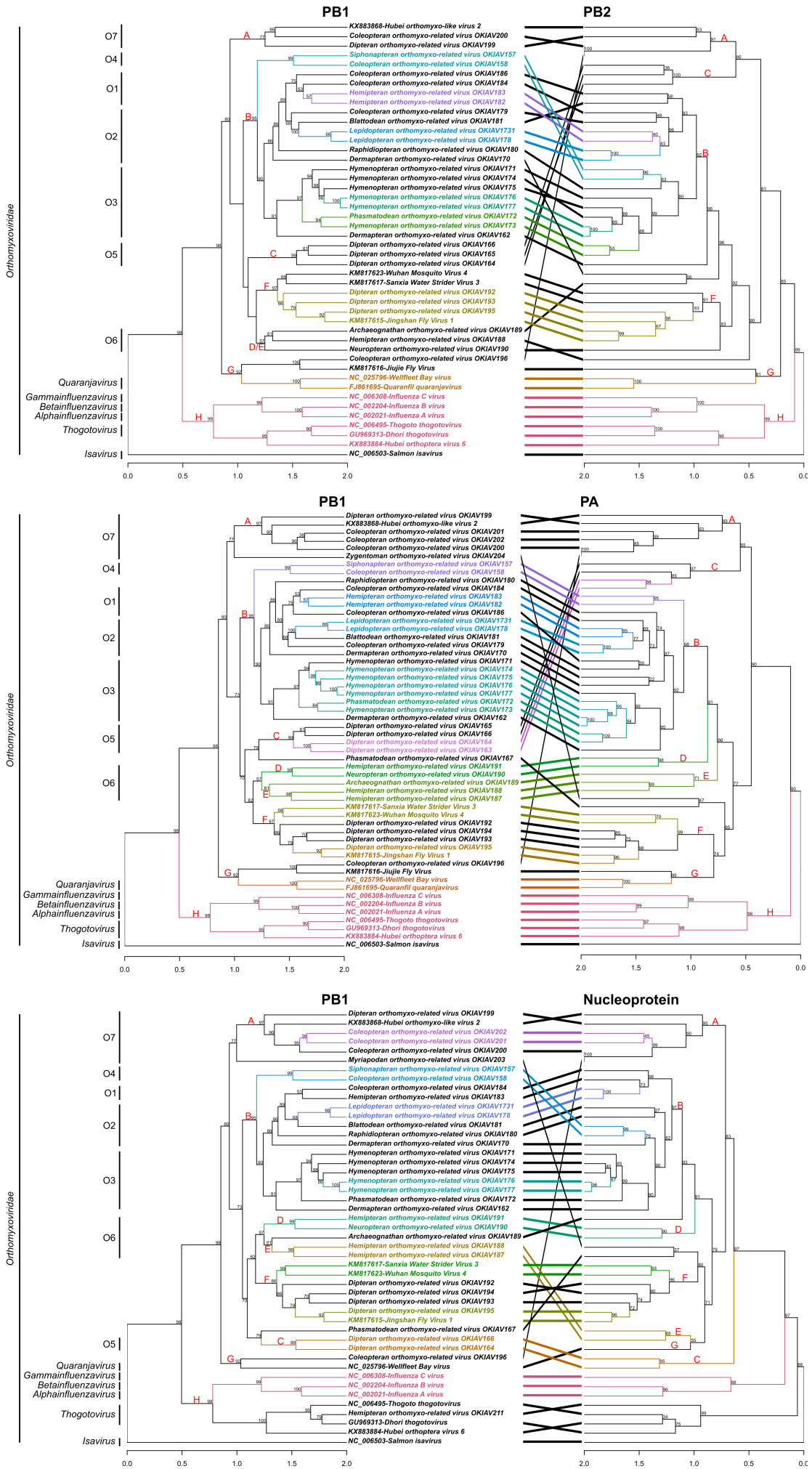
S12 Fig.



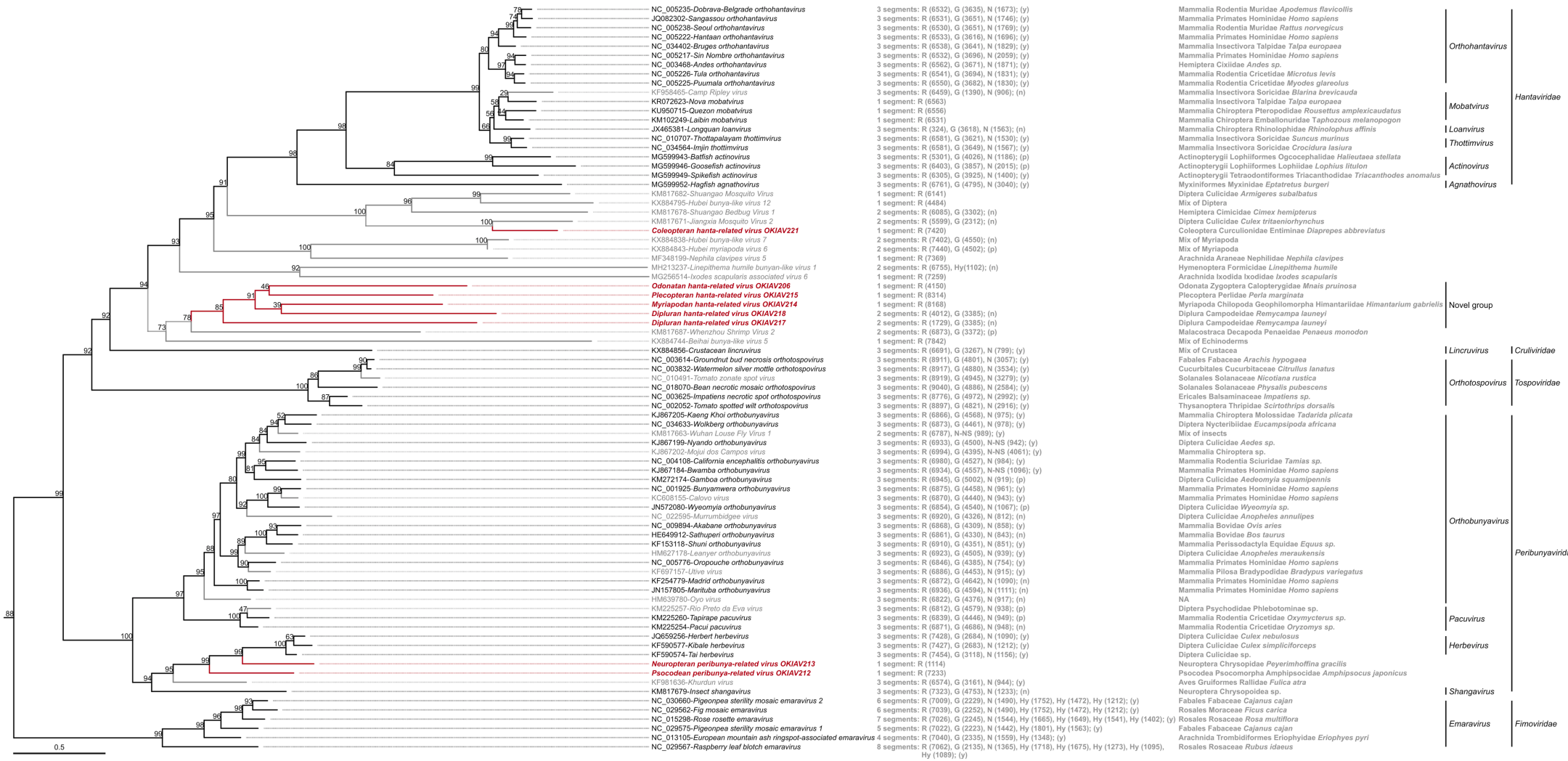
S13 Fig.



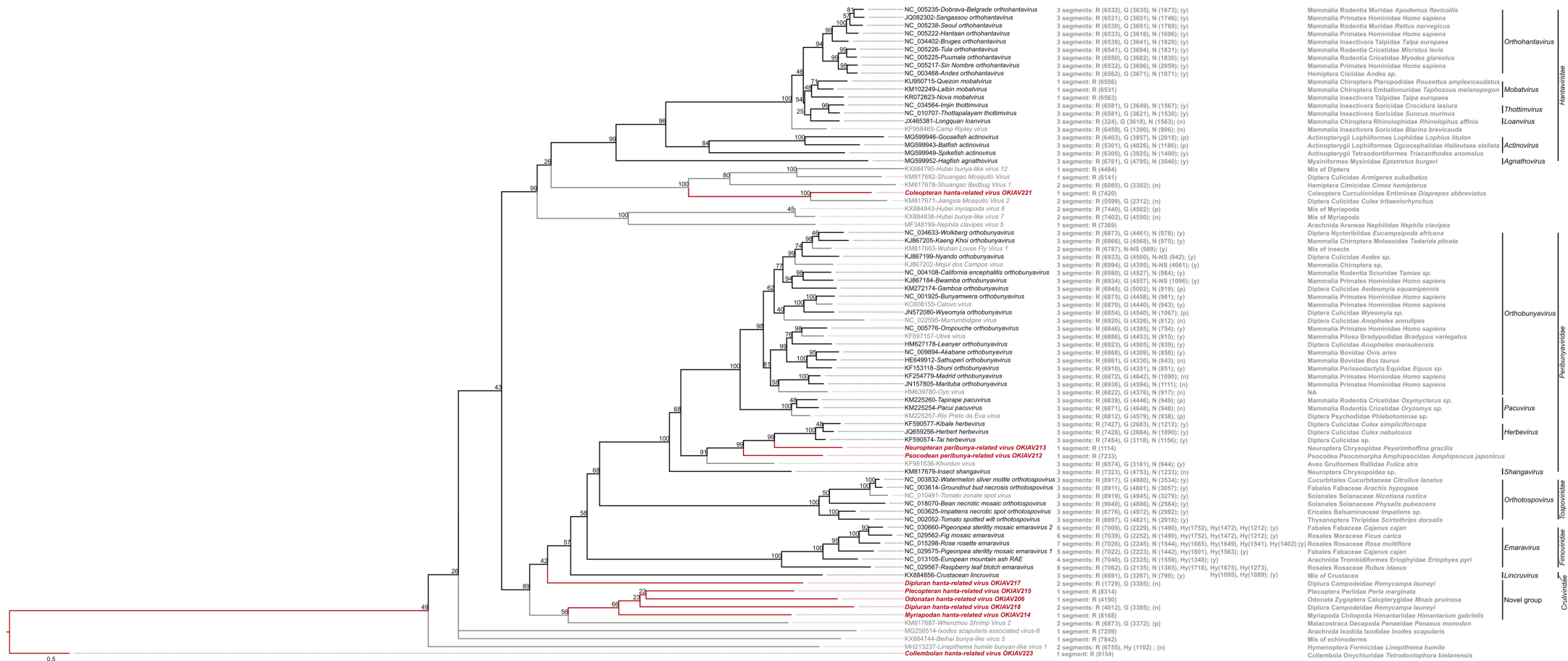
S14 Fig.



S15 Fig.

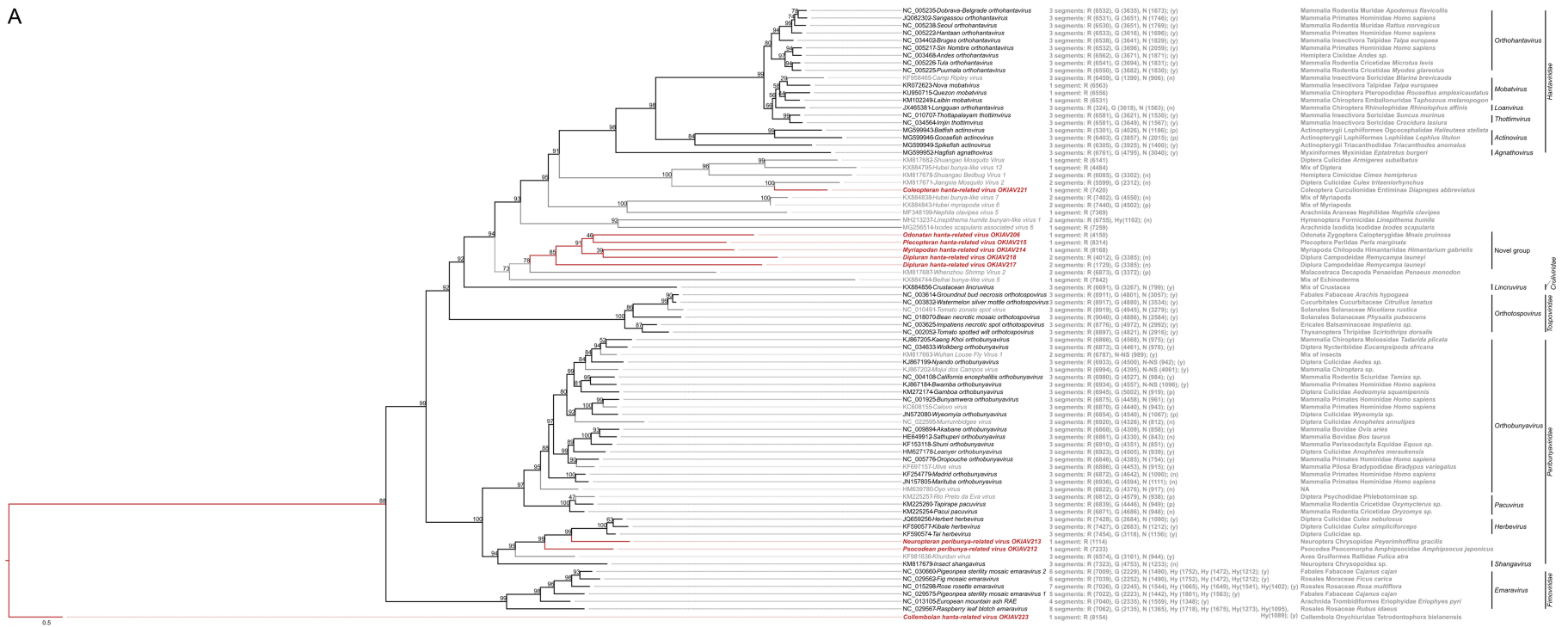


S16 Fig.

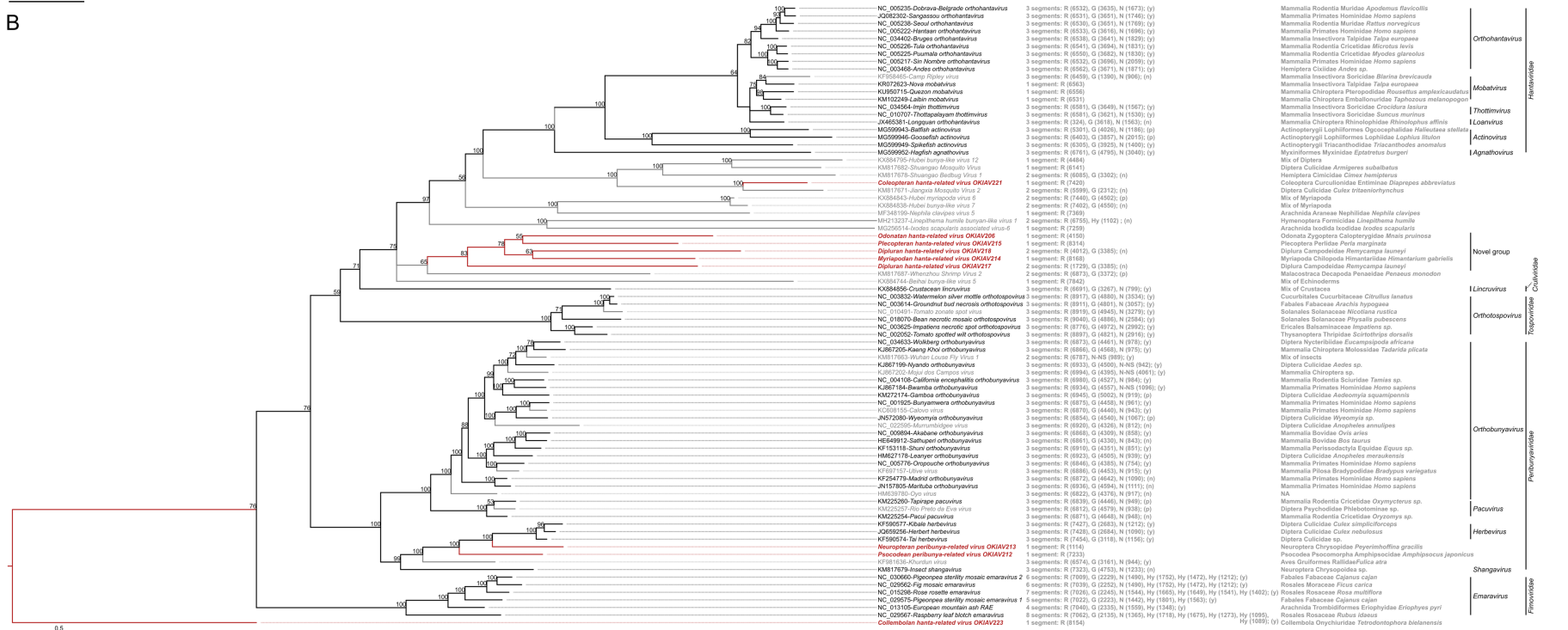


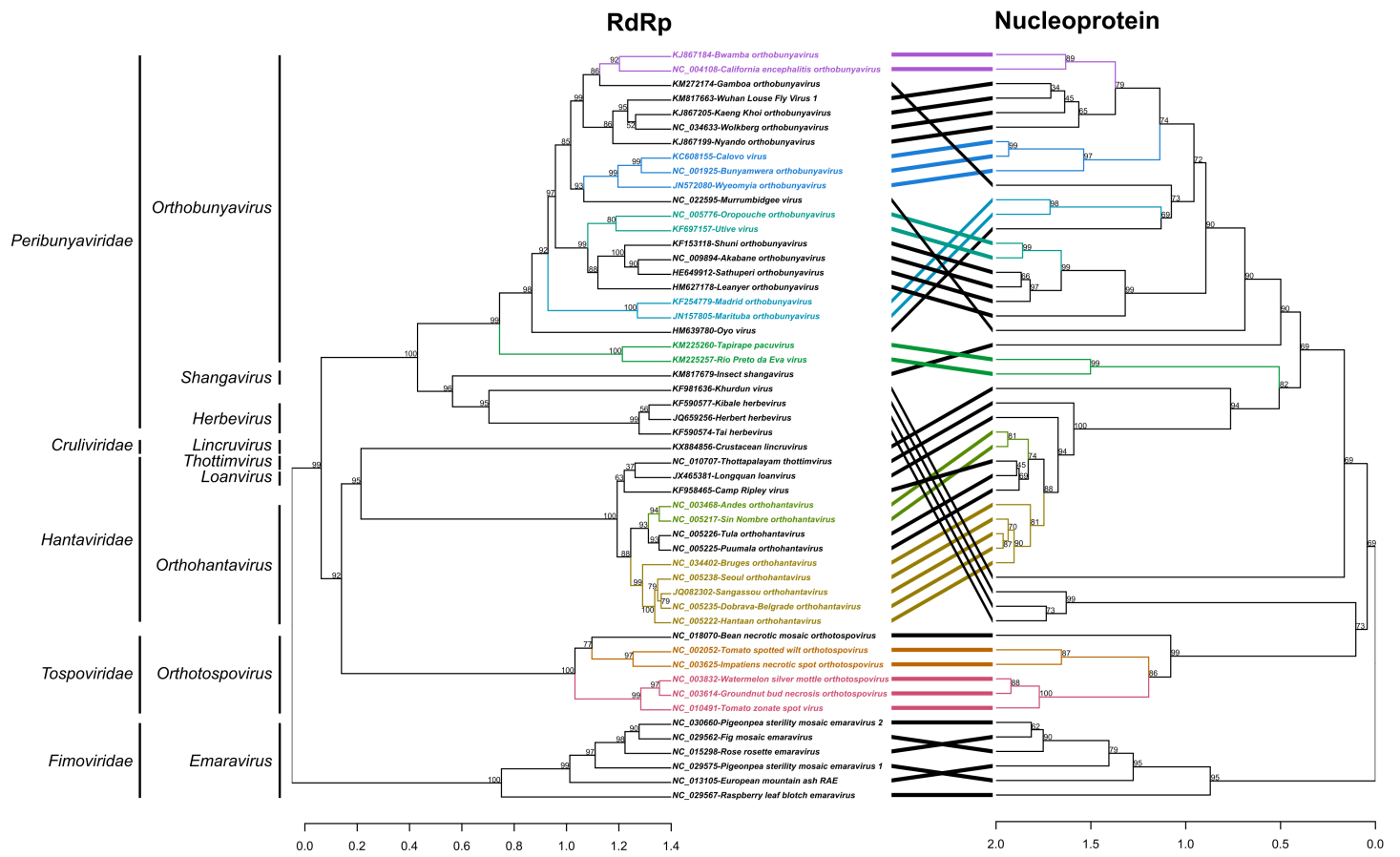
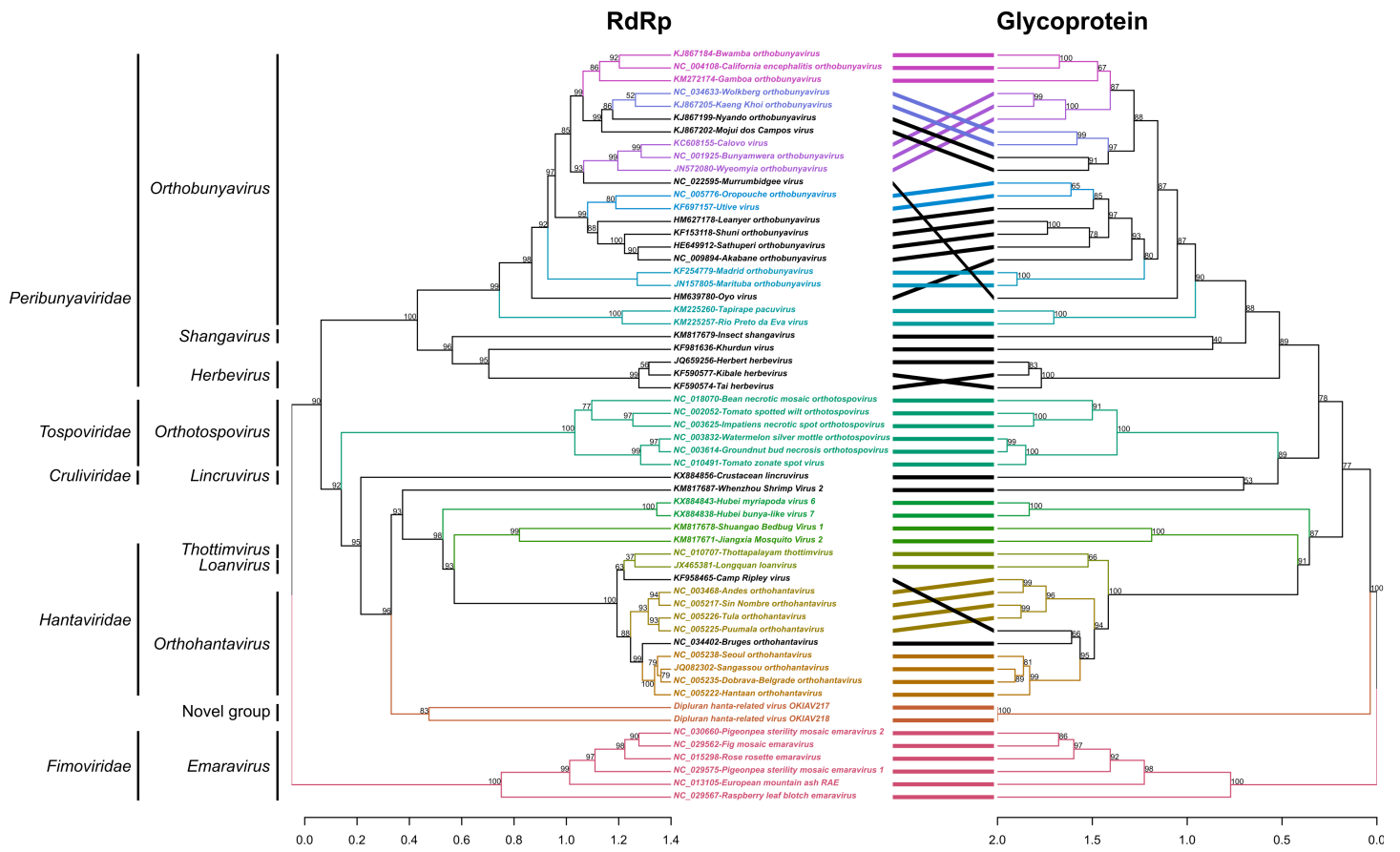
S17 Fig.

A

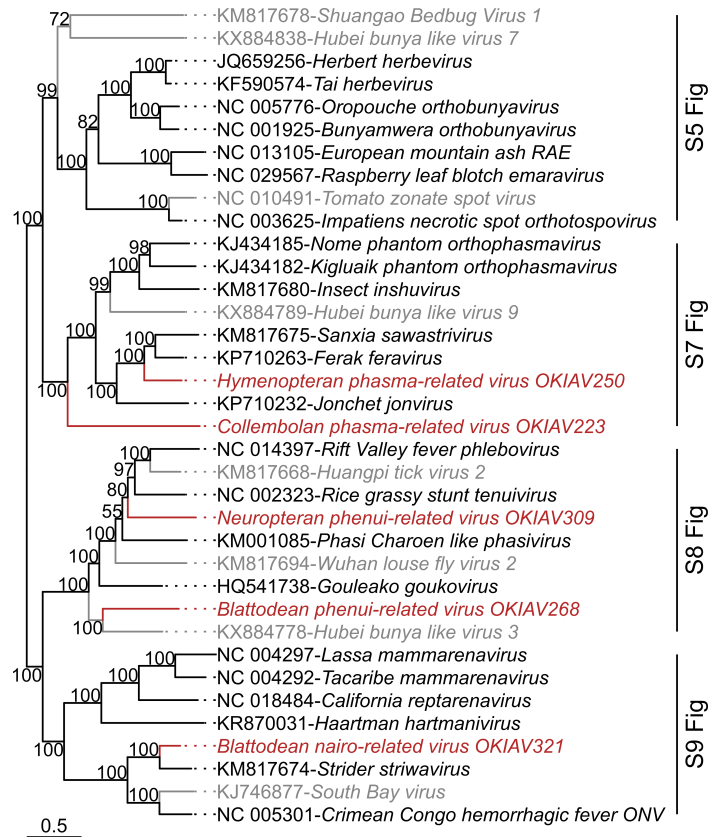


B

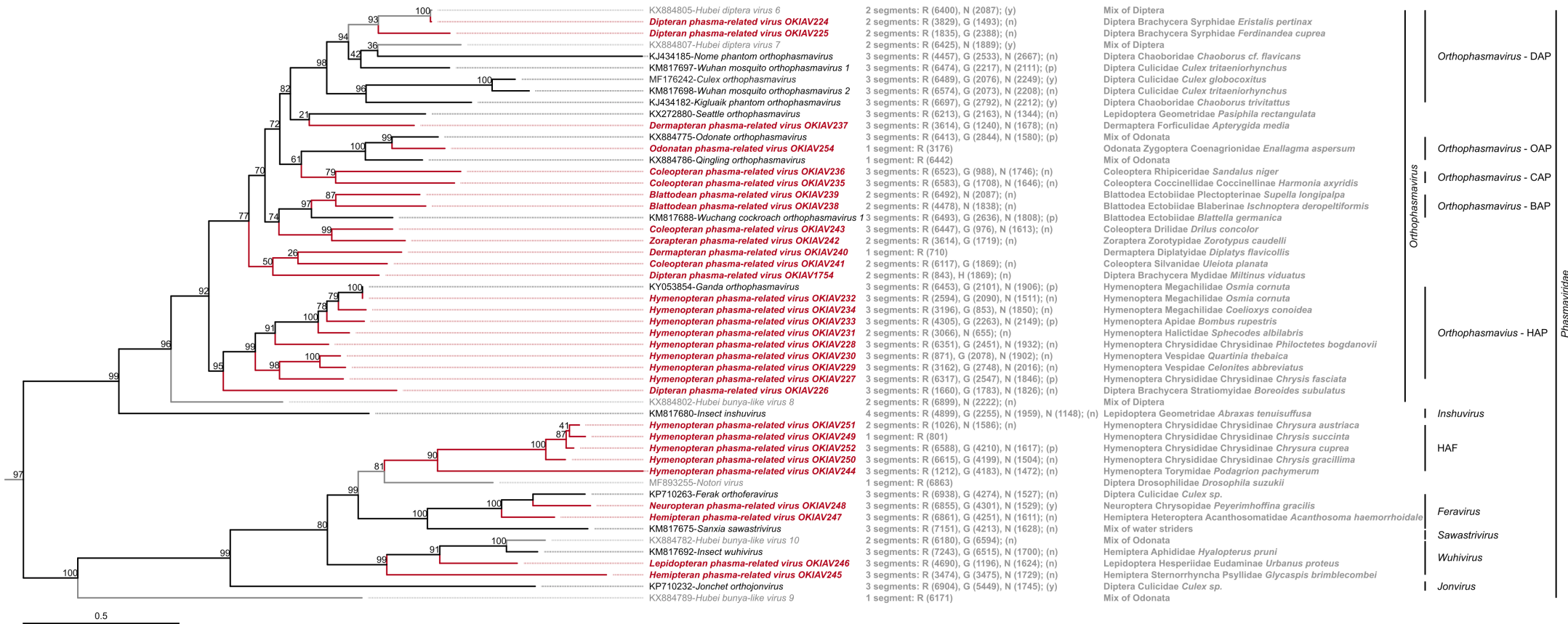




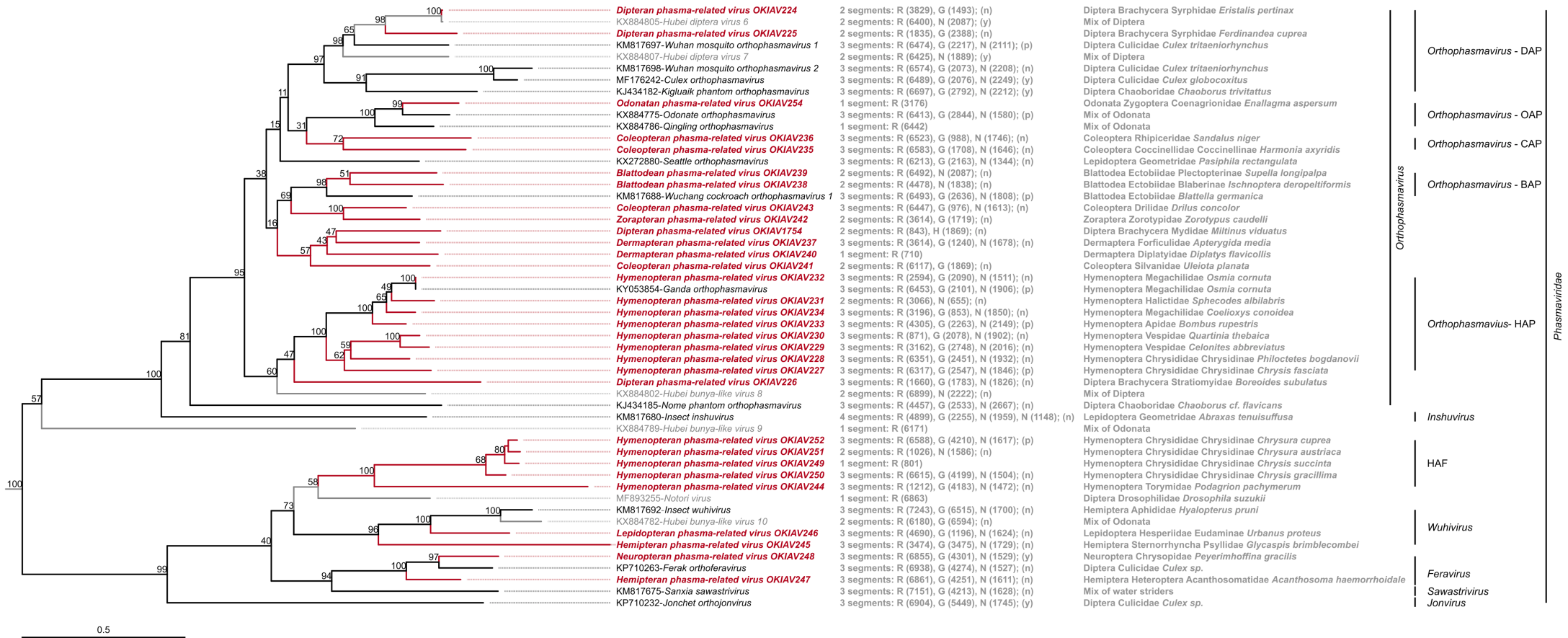
S19 Fig.



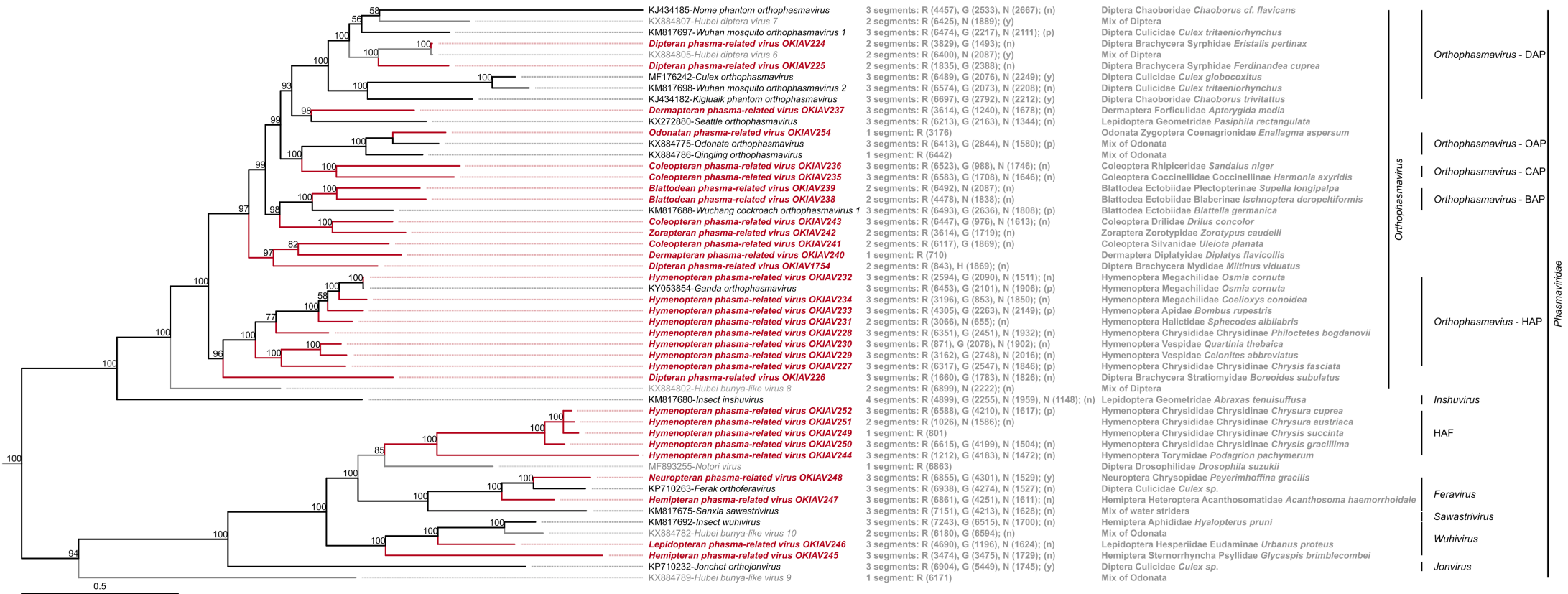
S20 Fig.



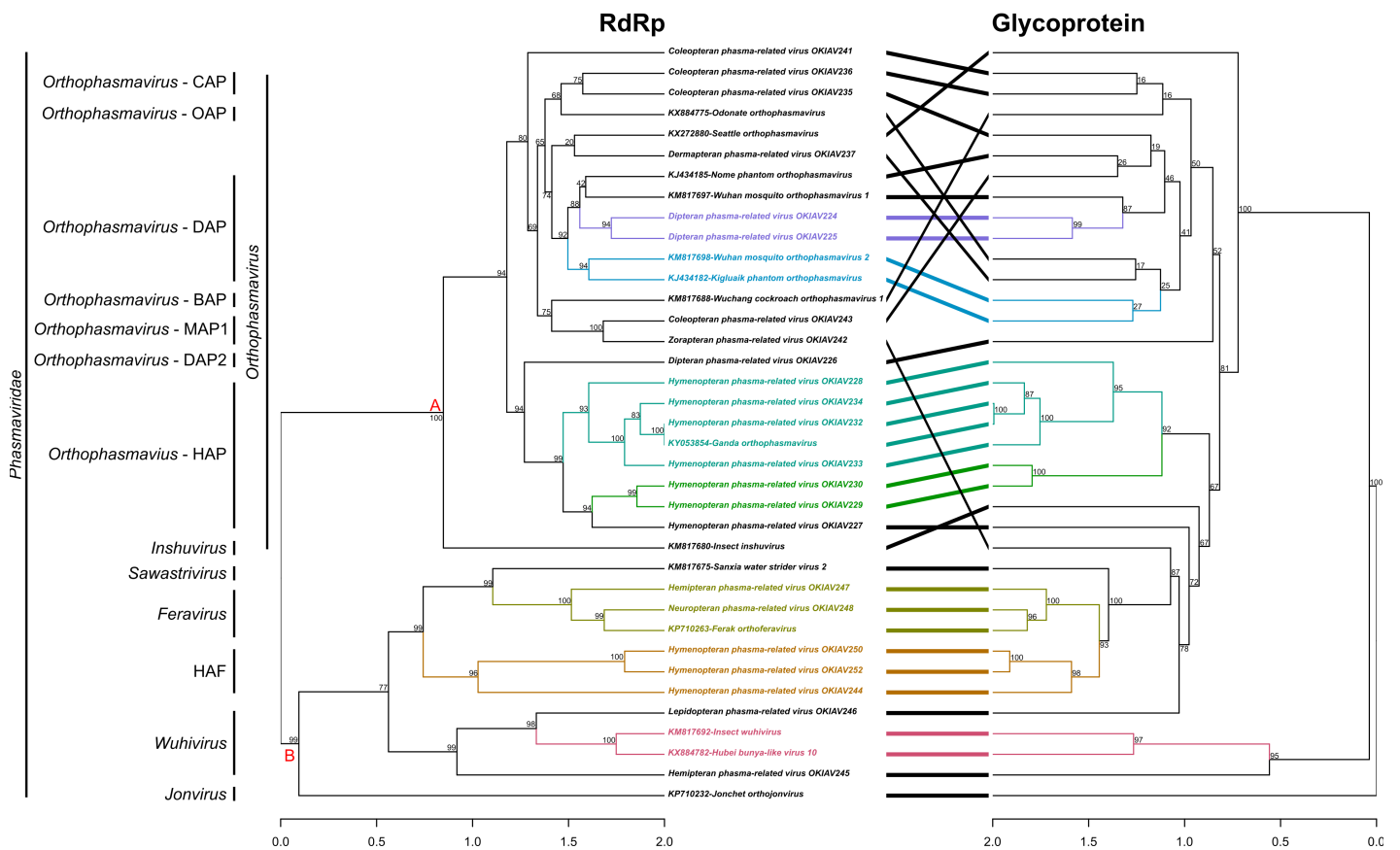
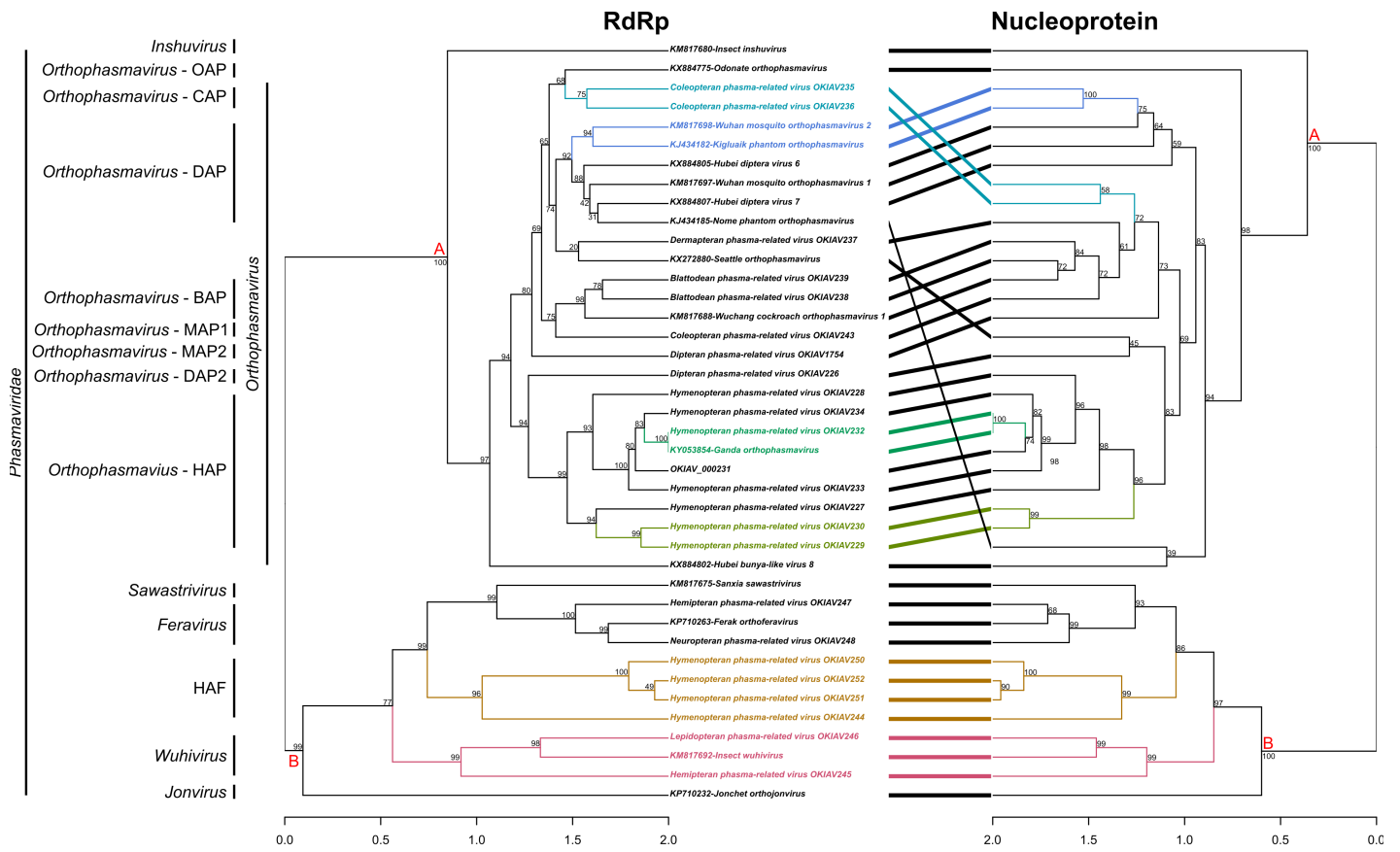
S21 Fig.



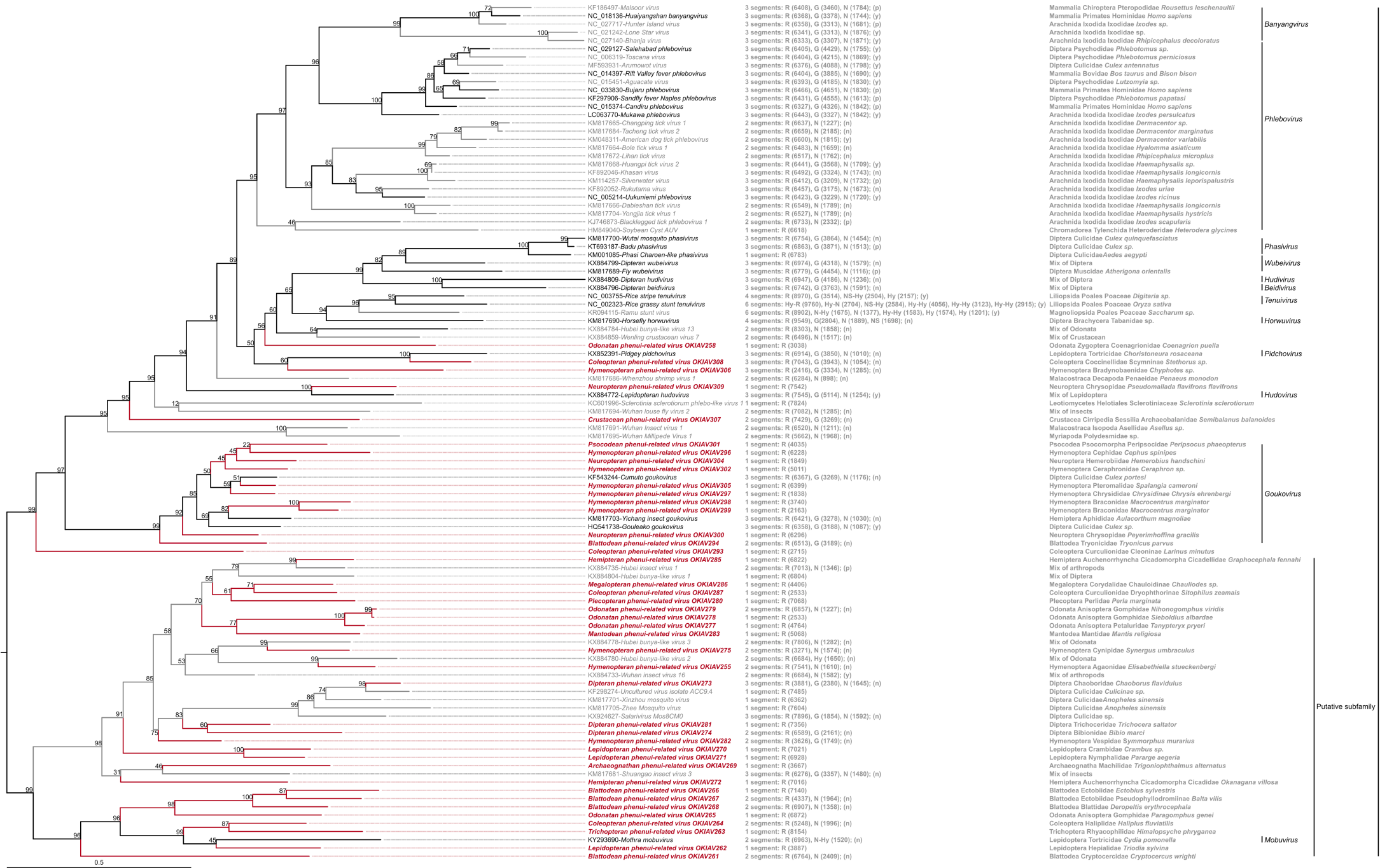
S22 Fig.



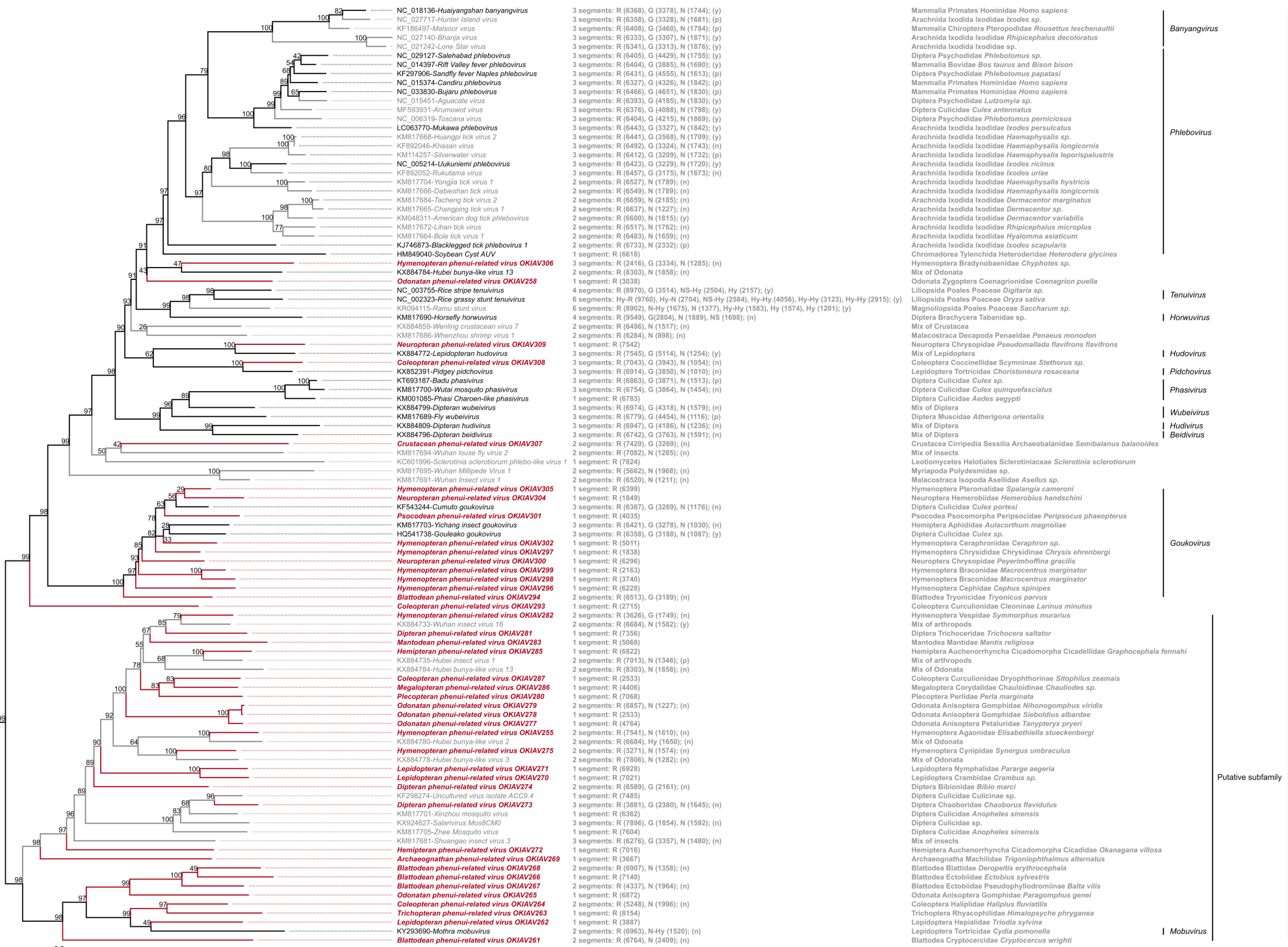
S23 Fig.



S24 Fig.



S25 Fig.



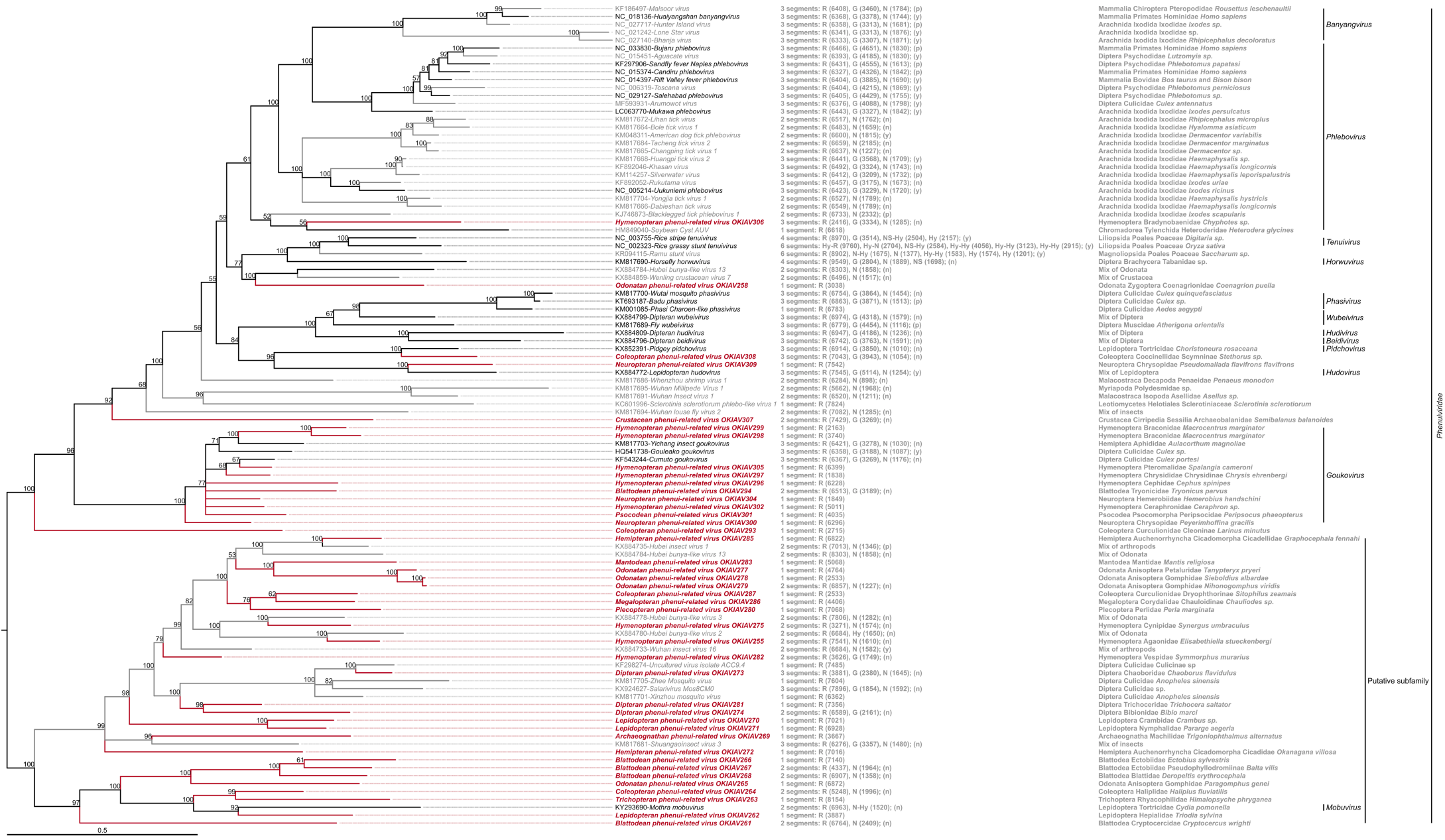
S26 Fig.

- Mammalia Primates Hominiidae *Homo sapiens*
- Arachnida Ixodida Ixodidae *Ixodes sp.*
- Mammalia Chiroptera Pteropodidae *Rousettus leschenaultii*
- Arachnida Ixodida Ixodidae *Rhipicephalus decoloratus*
- Arachnida Ixodida Ixodidae sp.
- Diptera Psychodidae *Phlebotomus* sp.
- Mammalia Bovidae *Bos taurus* and *Bison bison*
- Diptera Psychodidae *Phlebotomus papatasi*
- Mammalia Primates Hominiidae *Homo sapiens*
- Mammalia Primates Hominiidae *Homo sapiens*
- Diptera Psychodidae *Lutzomyia* sp.
- Diptera Culicidae *Culex antennatus*
- Diptera Psychodidae *Phlebotomus perniciosus*
- Arachnida Ixodida Ixodidae *Ixodes persulcatus*
- Arachnida Ixodida Ixodidae *Haemaphysalis* sp.
- Arachnida Ixodida Ixodidae *Haemaphysalis longicornis*
- Arachnida Ixodida Ixodidae *Haemaphysalis leporispalustris*
- Arachnida Ixodida Ixodidae *Ixodes ricinus*
- Arachnida Ixodida Ixodidae *Ixodes uriae*
- Arachnida Ixodida Ixodidae *Haemaphysalis hystricis*
- Arachnida Ixodida Ixodidae *Haemaphysalis longicornis*
- Arachnida Ixodida Ixodidae *Dermacentor marginatus*
- Arachnida Ixodida Ixodidae *Dermacentor* sp.
- Arachnida Ixodida Ixodidae *Dermacentor variabilis*
- Arachnida Ixodida Ixodidae *Rhipicephalus microplus*
- Arachnida Ixodida Ixodidae *Hyalomma asiaticum*
- Arachnida Ixodida Ixodidae *Ixodes scapularis*
- Chromadorea Tylenchida Heteroderidae *Heterodera glycines*
- Hymenoptera Bradynobaenidae *Clypeothys* sp.
- Mix of Odonata
- Odonata Zygoptera Coenagrionidae *Coenagrion puella*
- Liliopsida Poales Poaceae *Digitaria* sp.
- Liliopsida Poales Poaceae *Oryza sativa*
- Magnoliopsida Poales Poaceae *Saccharum* sp.
- Diptera Brachycera Tabanidae sp.
- Mix of Crustacea
- Malacostraca Decapoda Penaeidae *Penaeus monodon*
- Neuroptera Chrysopidae *Spalangia flavifrons flavifrons*
- Mix of Lepidoptera
- Coleoptera Coccinellidae *Scymninae Stethorus* sp.
- Lepidoptera Tortricidae *Choristoneura rosaceana*
- Diptera Culicidae *Culex* sp.
- Diptera Culicidae *Culex quinquefasciatus*
- Diptera Culicidae *Aedes aegypti*
- Mix of Diptera
- Diptera Muscidae *Atherigona orientalis*
- Mix of Diptera
- Mix of Diptera
- Crustacea Cirripedia Sessilia Archaeobalanidae *Semibalanus balanoides*
- Mix of insects
- Leotiomycetes Helotiales Sclerotiniaceae *Sclerotinia sclerotiorum*
- Myriapoda Polydesmidae sp.
- Malacostraca Isopoda Asellidae *Asellus* sp.
- Hymenoptera Pteromalidae *Spalangia cameroni*
- Neuroptera Hemerobidae *Hemerobius handschini*
- Diptera Culicidae *Culex portesi*
- Psococlea Psocomorpha Peripsocidae *Peripsocus phaeopterus*
- Hemiptera Aphididae *Aulacorthum magnoliae*
- Diptera Culicidae *Culex* sp.
- Hymenoptera Ceraphronidae *Ceraphron* sp.
- Hymenoptera Chrysididae *Chrysidinae Chrysis ehrenbergi*
- Neuroptera Chrysopidae *Peyerimhoffia gracilis*
- Hymenoptera Braconidae *Macrocentrus marginator*
- Hymenoptera Braconidae *Macrocentrus marginator*
- Hymenoptera Cephidae *Cephus spinipes*
- Blattodea Trycnicidae *Tryonicus parvus*
- Coleoptera Curculionidae *Cleoneinae Larinus minutus*
- Hymenoptera Vespididae *Symmorphus murarius*
- Mix of arthropods
- Diptera Trichoceridae *Trichocera saltator*
- Mantodea Mantidae *Mantis religiosa*
- Hemiptera Auchenorrhyncha Cicadomorpha Cicadellidae *Graphocephala fennahi*
- Mix of arthropods
- Mix of Odonata
- Coleoptera Curculionidae Dryophthorinae *Sitophilus zeamais*
- Megaloptera Corydalidae Chaulioidinae *Chauliodes* sp.
- Plecoptera Perlidae *Perla marginata*
- Odonata Anisoptera Gomphidae *Nihonogomphus viridis*
- Odonata Anisoptera Gomphidae *Sieboldia albidaria*
- Odonata Anisoptera Petaluridae *Tanypteryx pryeri*
- Hymenoptera Agaonidae *Elisabethiella stueckenbergi*
- Mix of Odonata
- Hymenoptera Cynipidae *Synergus umbraculus*
- Mix of Odonata
- Lepidoptera Nymphalidae *Pararge aegeria*
- Lepidoptera Crambidae *Crambus* sp.
- Diptera Bibionidae *Biblio marci*
- Diptera Culicidae Culicinae sp.
- Diptera Chaoboridae *Chaoborus flavidulus*
- Diptera Culicidae *Anopheles sinensis*
- Diptera Culicidae sp.
- Diptera Culicidae *Anopheles sinensis*
- Mix of insects
- Hemiptera Auchenorrhyncha Cicadomorpha Cicadidae *Okanagana villosa*
- Archaeognatha Machilidae *Trigonophthalmus alternatus*
- Blattodea Blattellidae *Deropeltis erythrocephala*
- Blattodea Ectobiidae *Ectobius sylvestris*
- Blattodea Ectobiidae Pseudophyllodrominae *Balta viilis*
- Odonata Anisoptera Gomphidae *Paragomphus geni*
- Coleoptera Halipidae *Halipus fluvialis*
- Trichoptera Rhyacophilidae *Himalopsycha phryganea*
- Lepidoptera Hepialidae *Triodia sylvina*
- Lepidoptera Tortricidae *Cydia pomonella*
- Blattodea Cryptoceridae *Cryptocercus wrighti*

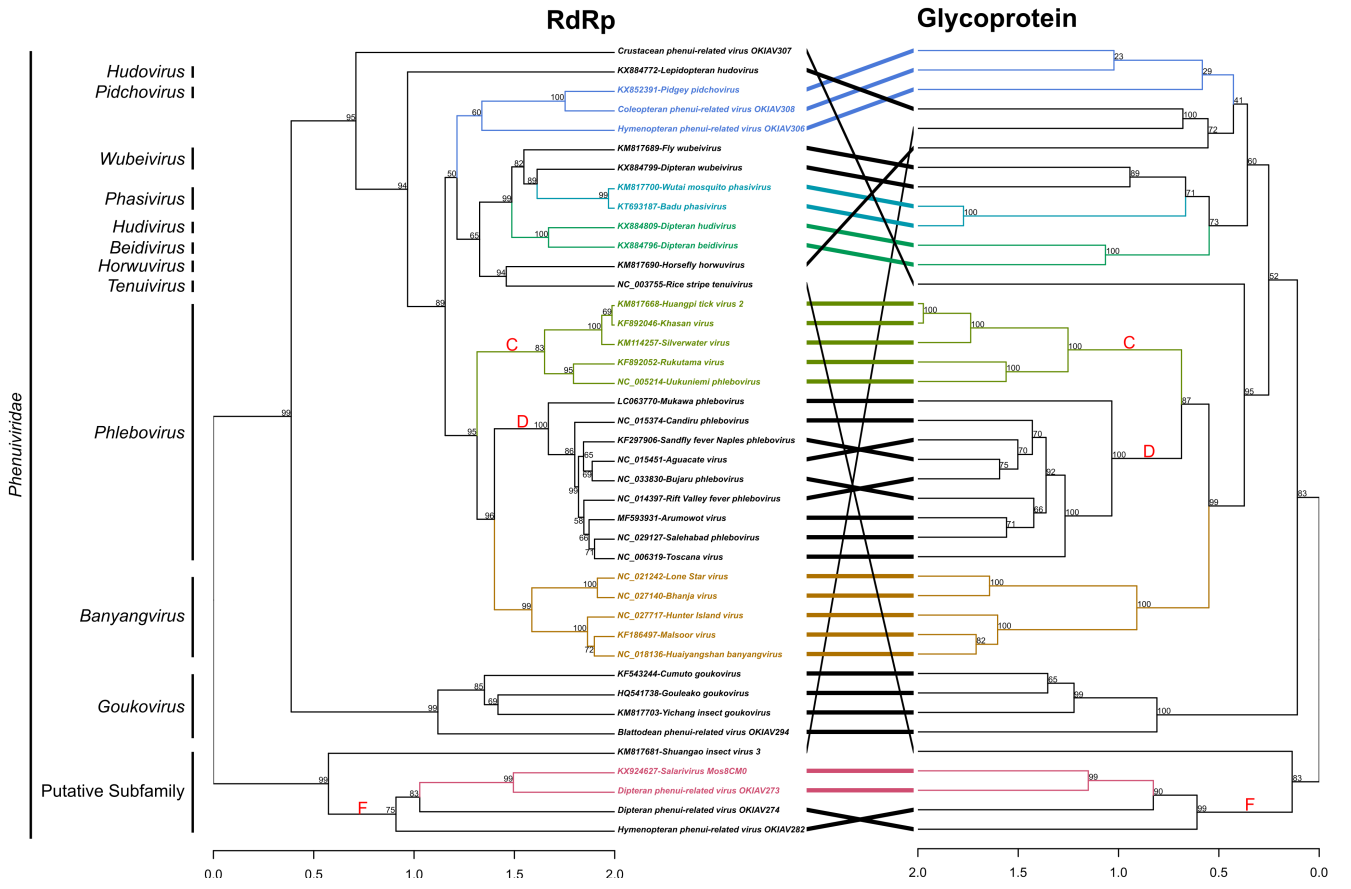
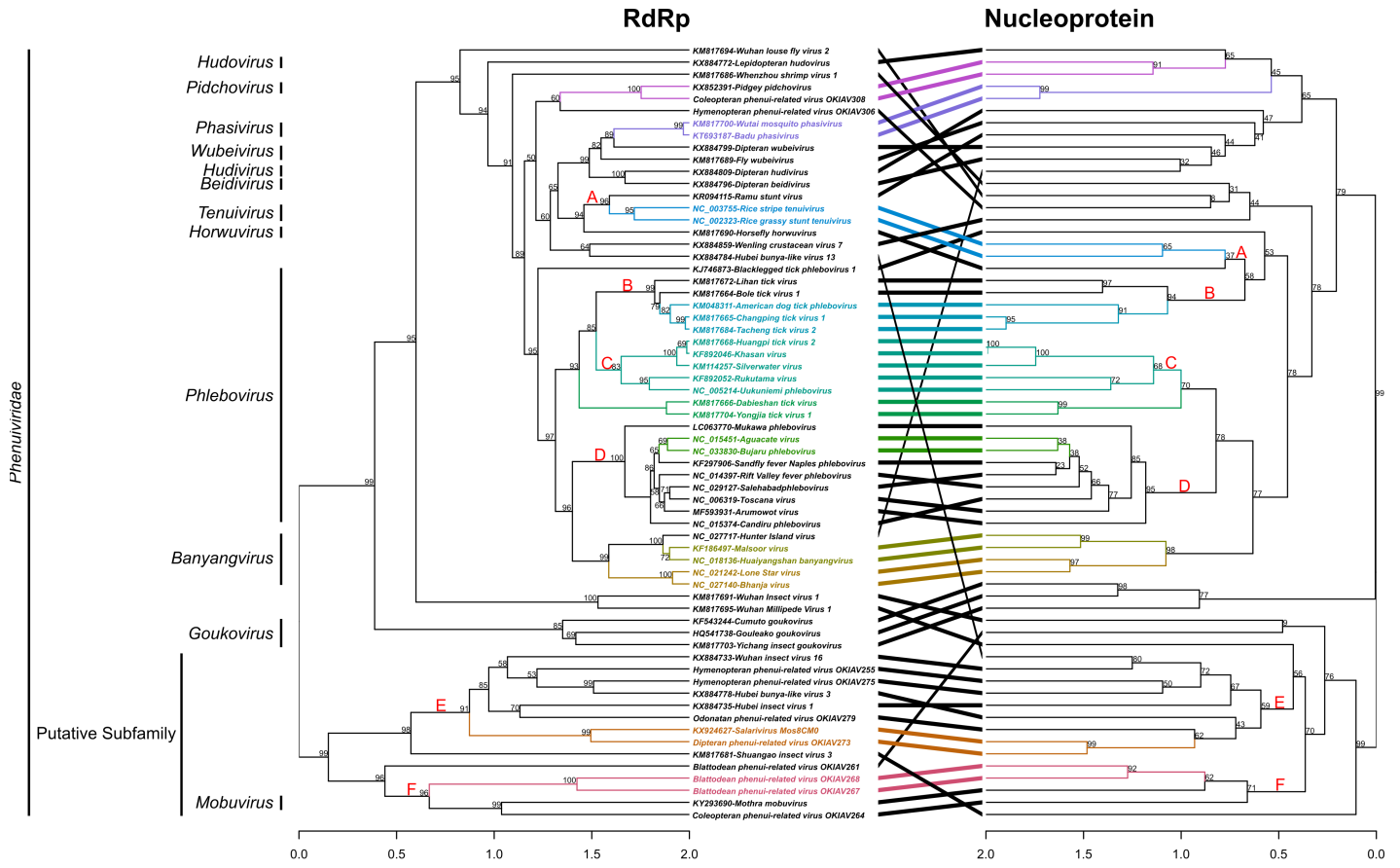
Phleboviridae

Putative subfamily

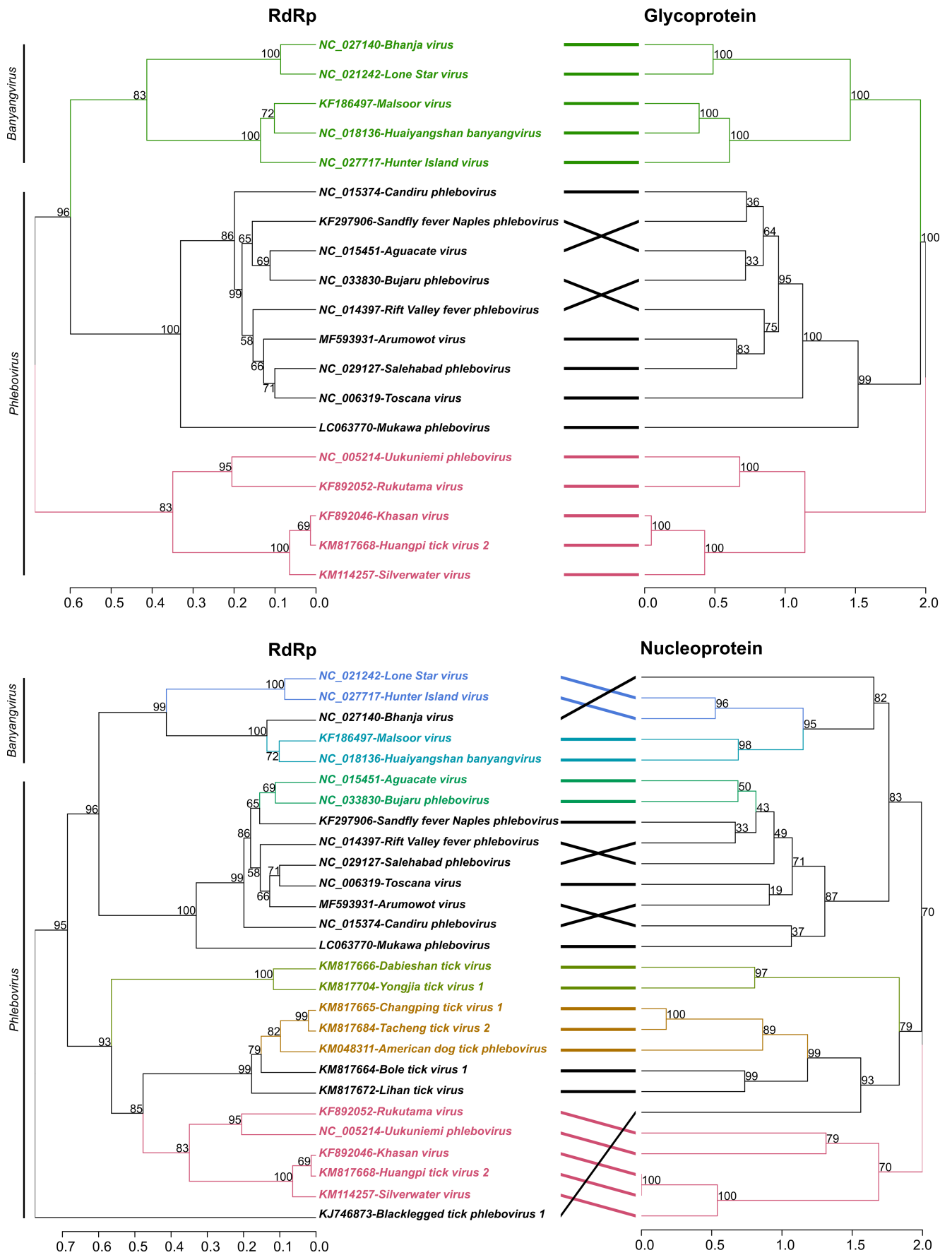
Mobovirus



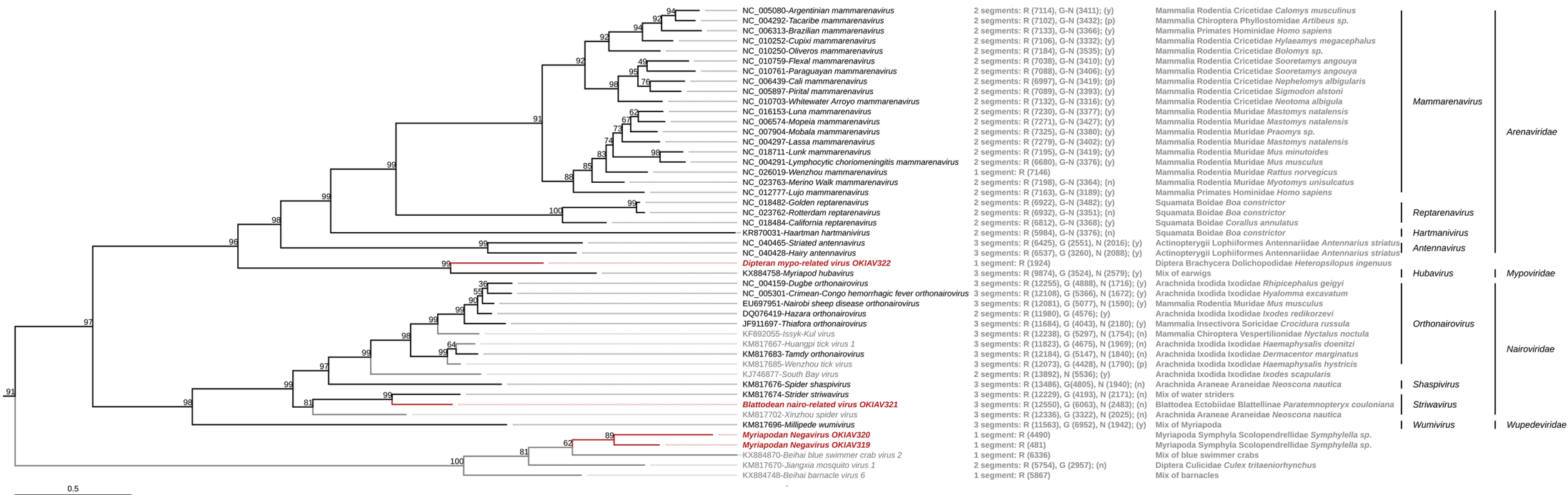
S27 Fig.



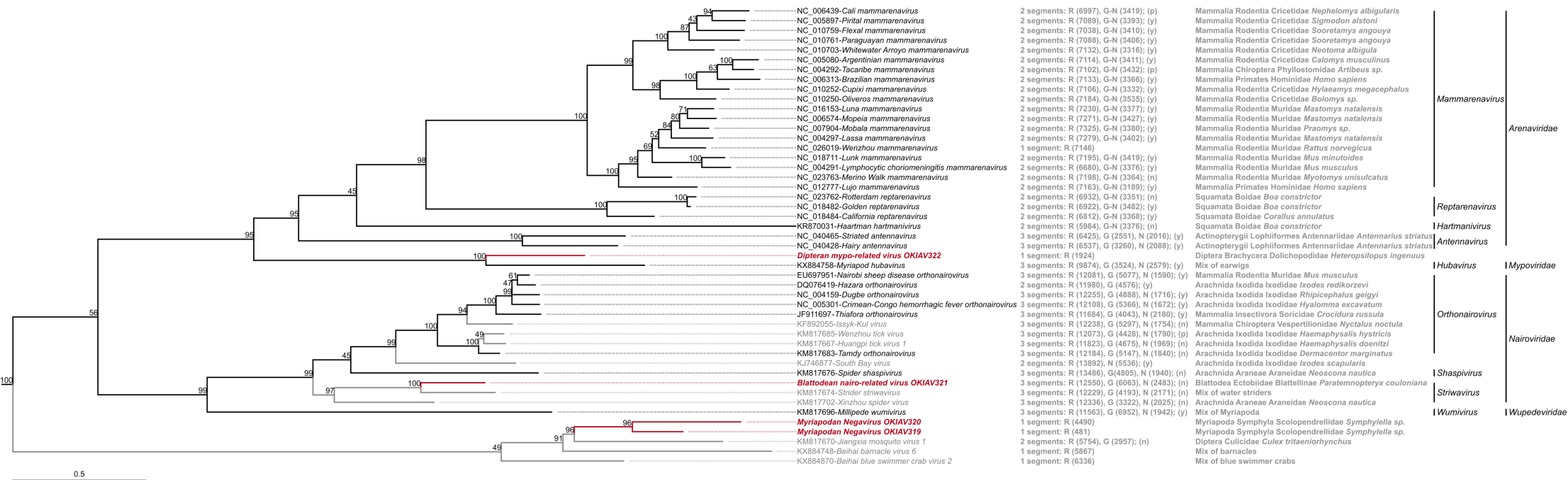
S28 Fig.



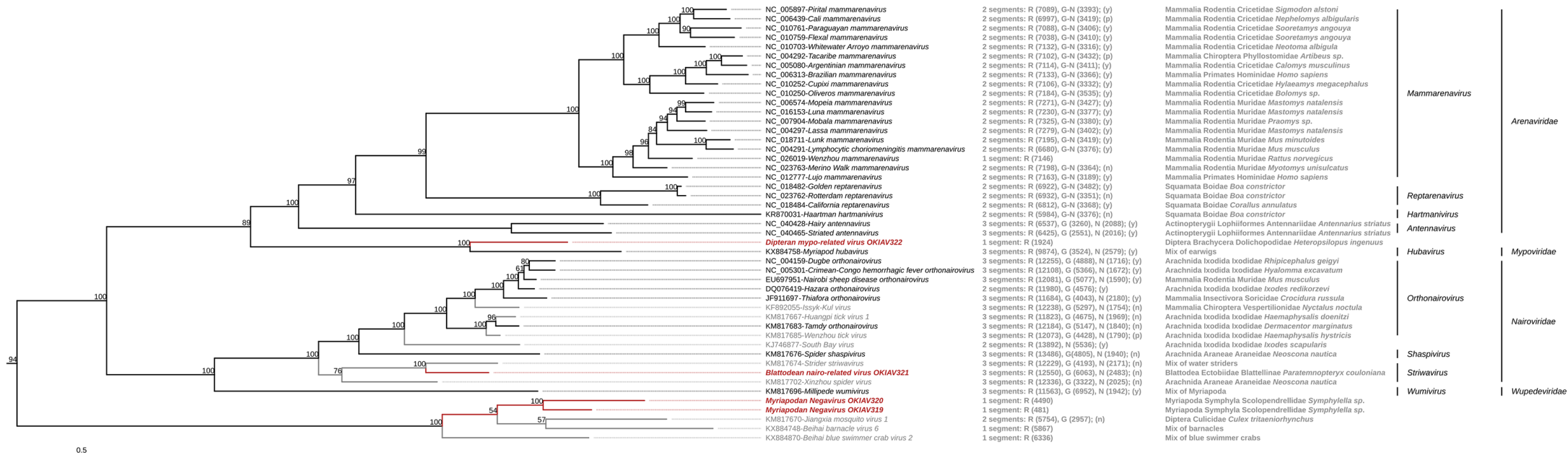
S29 Fig.



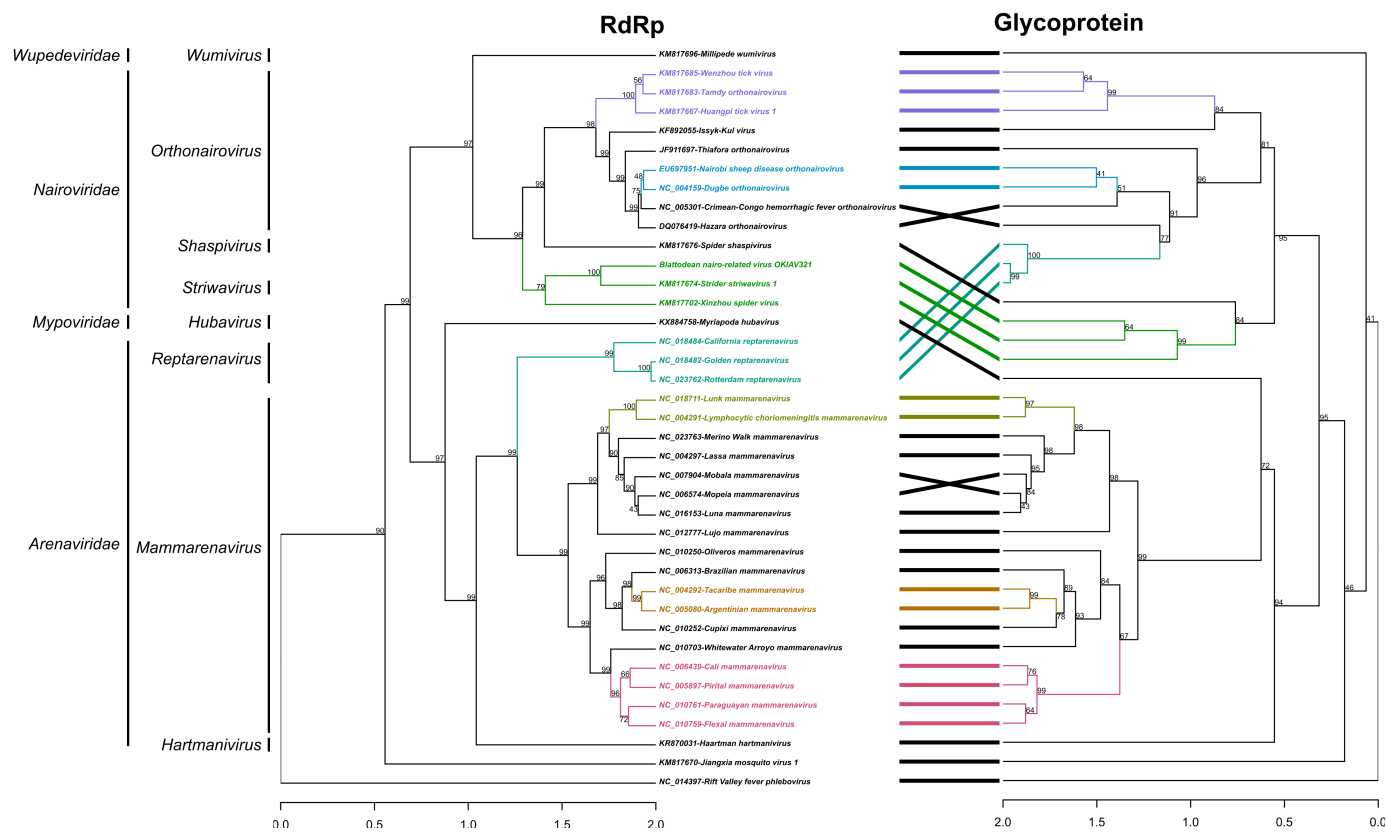
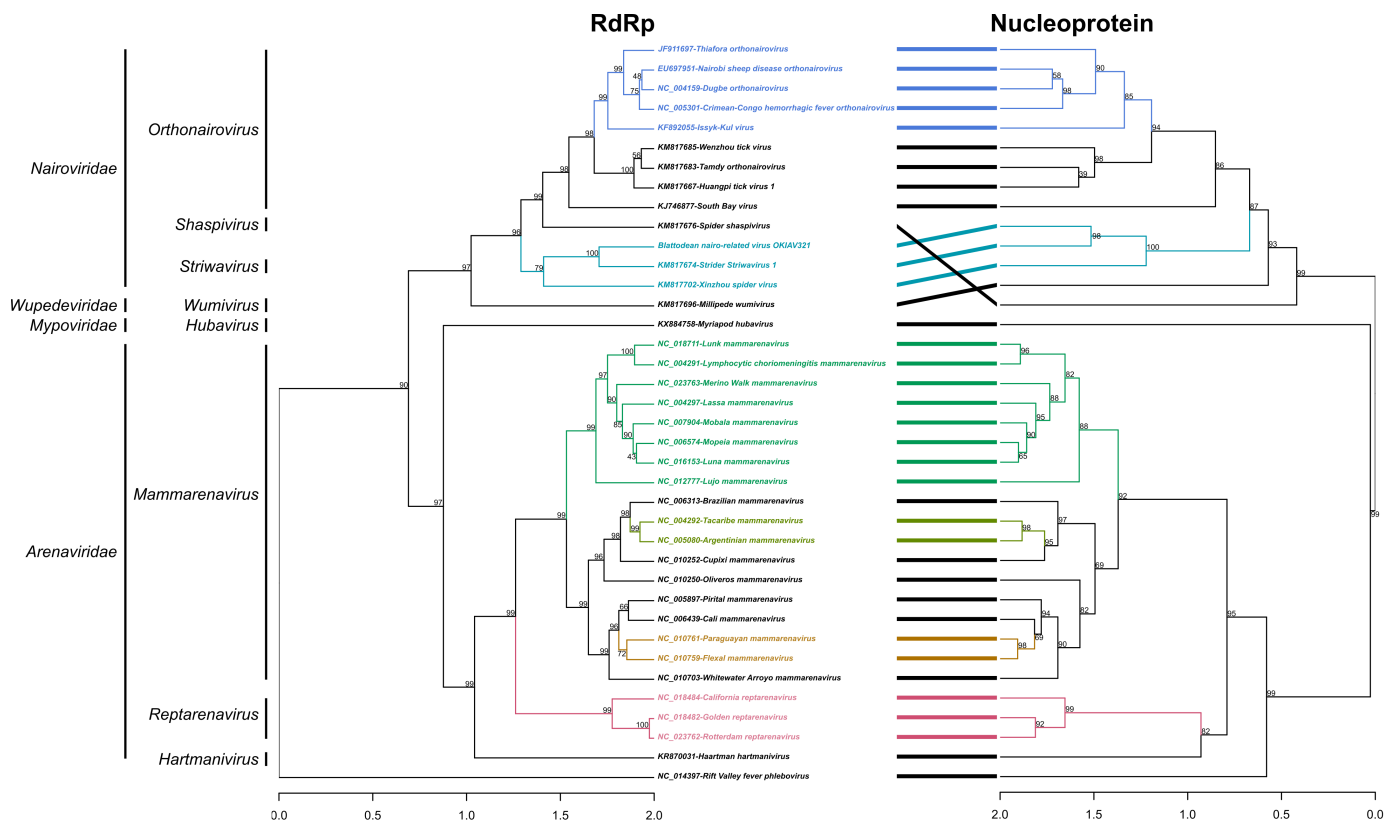
S30 Fig.



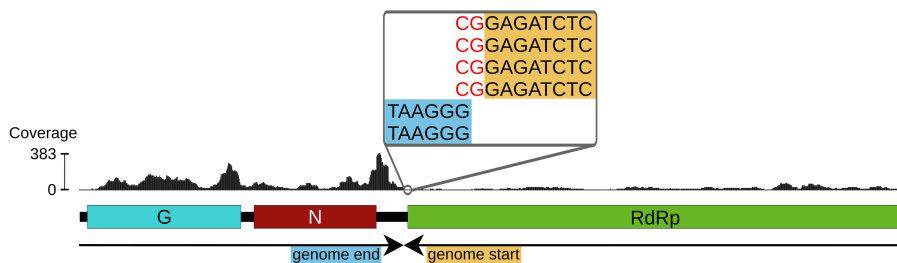
S31 Fig.



S32 Fig.



S33 Fig.



S34 Fig.

Appendix of Chapter 3

Supplementary Material for the above publication published as: Paraskevopoulou S., Pirzer F., Goldmann N., Schmid J., Corman V.M., Gottula L.T., Schroeder S., Rasche A., Muth D., Drexler J.F., Heni A.C., Eibner G.J., Page R.A., Jones T.C., Müller M.A., Sommer S., Glebe D., and Drosten C. (2020). **Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus***. *Proceedings of the National Academy of Sciences*, 117(30):17977–17983.

The Supplementary Material is also available at: <https://www.pnas.org/content/pnas/suppl/2020/07/10/2006750117.DCSupplemental/pnas.2006750117.sapp.pdf>



Supporting Information Appendix for

Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*

Sofia Paraskevopoulou, Fabian Pirzer, Nora Goldmann, Julian Schmid, Victor Max Corman, Lina Theresa Gottula, Simon Schroeder, Andrea Rasche, Doreen Muth, Jan Felix Drexler, Alexander Christoph Heni, Georg Joachim Eibner, Rachel A. Page, Terry C. Jones, Marcel Alexander Müller, Simone Sommer, Dieter Glebe, Christian Drosten

Corresponding authors: Christian Drosten, Dieter Glebe, Simone Sommer
Emails: christian.drosten@charite.de, dieter.glebe@viro.med.uni-giessen.de, simone.sommer@uni-ulm.de

This PDF file includes:

- SI Appendix Text
- SI Appendix Materials and Methods
- Figures S1 to S6
- Tables S1 to S10
- References for SI Appendix citations

SI Appendix Text

Genome organisation, motifs, and post-translational modifications

Structural HDV domains, amino acid motifs, and post-translational modifications of HDAG regulate the processes of HDV transcription, nuclear localization and export, and replication. RNA-editing of the amber stop codon of the antigenomic strand leads to L-HDAG mRNA transcription. L-HDAG acts mainly as an inhibitor of certain stages of HDV RNA replication and is required for virion assembly and attachment to HBV envelope proteins (1). Farnesylation of the C-terminus Cys-211 is a crucial modification required for incorporating HDV particles into HBV envelope proteins (2). Recent experimental evidence demonstrates that HDV infectious particles can be produced using envelope proteins from viruses other than HBV (3). However, farnesylation of the L-HDAG C-terminus is a requirement. As with the recently described snake deltavirus, duck-associated, and other vertebrate and invertebrate deltavirus sequences (4–6), the rodent deltavirus (RDeV) ORF does not encode the CXXQ C-terminal motif required for packaging. Genomic and antigenomic HDV RNA replication takes place in the nucleus, with genomic HDV RNA and mRNA synthesized in nuclear bodies, and antigenomic HDV RNA replicated in the nucleolus (7, 8). S-HDAG employs DNA-dependent RNA polymerase II (Pol-II) for genomic HDV RNA replication by functionally interacting with its clamp formation, forcing it to use an RNA instead of a DNA template (9, 10). HDV RNA genomes are replicated under a double rolling-circle replication scheme, forming multimers that are later self-cleaved by ribozyme structures (11–13). Two ribozyme regions self-cleave the genomic and the antigenomic human HDV RNA respectively (14). SI Appendix **Fig. S4** shows the HDV ribozyme structures together with predicted ribozyme structures for the rodent deltavirus. The coiled-coil domain at the N-terminal HDAG is responsible for the formation of dimers/multimers as well as for activating/promoting and inhibiting RNA replication (15, 16). The HDV nuclear localization signal (NLS) is located between residues 66-75 on the S-HDAG with the conserved amino acid domains 66-EGAPPAKRAR-75 (17). Essential for the nuclear import are the acidic glutamic acid residue at position 66 and the basic arginine residue at position 75 (17). S-RDeAg shows a slightly different NLS composition, although presenting an acidic aspartic acid residue at position 66 and a basic arginine residue at position 74. Pro-205 of the L-HDAG is critical for the nuclear export signal's (NES) functionality, which directs L-HDAG to the cytoplasm via a chromosome region maintenance 1 (CRM1)-independent pathway, a transport factor found at nuclear pores (18, 19). However, nuclear export of genomic HDV RNA is independent of L-HDAG and because genomic HDV RNA is synthesized by Pol-II, it has been speculated that it may be exported by the same mechanism as for splicing-dependent cellular mRNAs (18).

S-HDAG is an essential activator for HDV RNA replication and undergoes the following modifications: i) Arg-13 methylation, ii) Lys-72 acetylation, iii) Ser-177 phosphorylation, iv) sumoylation of several Lysine residues. Methylation at Arg-13 is essential for translocating S-HDAG to the nucleus and for antigenomic RNA replication (20, 21). Experimental evidence shows this site to be involved in transporting the antigenomic RNA into the nucleus (21). Non-methylated S-HDAG mediates the nuclear transport of genomic HDV RNA (21). The nuclear transport of HDAG is also mediated by Lys-72 acetylation (22). Lys-72 acetylation additionally regulates the synthesis of the different HDV RNA species, enhancing genomic HDV RNA and mRNA replication (23). Very recently, experimental evidence showed that acetylation

of Lys-72 (K72acXXR motif) is responsible for histone mimicry to recruit RNA Pol-II on the HDV RNP and sustain HDV replication (10). On the contrary, unmodified S-HDAg is advantageous for antigenomic HDV RNA replication, proving to be a molecular switch for the replication of the different HDV RNA species (23). Phosphorylation of Serine residues (most importantly of Ser-177) enhances antigenomic RNA replication for the production of genomic RNA by interacting with RNA Pol-II (7, 24). S-HDAg and RNA Pol-II are both phosphoproteins, and phosphorylation has been shown to act as a molecular switch for activating/deactivating protein functions (25). The targets of SUMO proteins are cellular and viral proteins that function in the nucleus. Protein SUMO1 targets Lysine sites of the S-HDAg, thereby selectively enhancing HDV genomic RNA and mRNA synthesis, but not antigenomic HDV RNA synthesis (26). **Fig. 1 A** shows the location of all post-translational modifications on HDV, RDeV, SDeV, duck-associated DeV, fish DeV, newt DeV, termite DeV, and toad DeV and SI Appendix **Table S10** summarizes their functions.

Absence of HBV in transcriptome data of all samples that tested RDeV RNA-positive by qRT-PCR and had a transcriptome available

Because none of the *P. semispinosus* individuals were found to be positive for hepadnaviruses, we also searched the NGS sequencing reads that were available for 120 samples for the presence of additional viruses. For each sample, all reads were matched against the RVDB protein database (version 13.0 (27)) using DIAMOND (version 0.9.22.123 (27, 28)). Matches against individual proteins were collected and the coverage of each protein was calculated. Across all 120 samples, only rodent hepacivirus polyproteins (GenBank accession numbers KC411780, KC411783, NC_021153, and KC815312) were found. This was as expected because all samples had been pre-selected for the presence of hepacivirus.

SI Appendix Materials and Methods

Virus detection and screening for viral infection

Total RNA was extracted from blood using the Viral NA Small Volume Kit (Roche Molecular Diagnostics) as described previously (29). RNA extracts were pooled (up to 10 samples per pool). For organ RNA, extraction tissues were homogenized in 350 μ L PBS. Available organs from 21 individuals were: liver, lung, heart, small intestine, kidney, colon, and spleen. 250 μ L of homogenized tissue, together with 250 μ L tissue lysis buffer were extracted using the DNA/Viral NA Large Volume 2.0 Kit (Roche Molecular Diagnostics) according to manufacturer instructions and eluted in 100 μ L elution buffer. All RNA extracts were tested with a real-time reverse transcription polymerase chain reaction (qRT-PCR) specifically designed for *P. semispinosus* deltavirus, targeting a 94-nucleotide fragment in the S-HDAg coding region (primers listed in SI Appendix **Table S2**). The SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Thermo Fisher Scientific) was used in a 12.5 μ L reaction volume. Primer concentrations were 400 nM for forward and reverse primers, 200 nM TaqMan probe, 6.25 μ L 2x kit reaction buffer (containing 400 nM of each dNTP and 3.2 mM magnesium sulfate), 10 mM magnesium sulfate solution (Thermo Fisher Scientific), 0.5 μ L enzyme mix, 1.8 μ L of RNase-free water, and 2.5 μ L RNA extract. The thermal cycling protocol on a Roche LightCycler 480 started with 20 min at 55°C for reverse transcription, followed by 3 min at 94°C, then 45 cycles of 15 sec at 94°C and 20 sec at 58°C, with fluorescence signal detection at

the end of the final 58°C step.

RNA extracts were additionally tested with a real-time RT-PCR assay for the presence of transcripts from a host housekeeping gene (TBP: TATA-binding protein, primers are listed in SI Appendix **Table S2**), to exclude potential RNA degradation (SI Appendix **Table S8**). The real-time RT-PCR protocol was followed as described above, on a Roche LightCycler 480 starting with 20 min at 55°C for reverse transcription, followed by 3 min at 95°C, then 45 cycles of 15 sec at 95°C and 45 sec at 57°C, with fluorescence signal detection at the end of the final 57°C step.

Circularization assay

To confirm that the rodent deltavirus genome is circular, all deltavirus-positive blood RNA extracts were subjected to cDNA synthesis using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with specific forward primers. Two mixes were prepared for the cDNA synthesis. Mix 1 contained 8 µL RNase-free water, 2 µL First strand buffer (5x, SS III Kit, Invitrogen), 1 µL dNTP mix (10 mM each), 1 µL forward primer (10 µM), and 2.5 µL RNA extract. Mix 2 contained 0.5 µL RNase-free water, 2 µL First strand buffer (5x, SS III Kit, Invitrogen), 1 µL dithiothreitol (100 mM), 1 µL BSA (1 mg/mL), and 1 µL SuperScript III Reverse Transcriptase. Mix 1 was incubated for 5 min at 65 °C and put on ice for 1 min. In parallel with the cooling step, mix 2 was heated to 55 °C. After cooling on ice, 5.5 µL of mix 2 was added into mix 1 and heated to 55°C for 60 min followed by 70 °C for 15 min. Afterwards, 1 µL RNase H was added and incubated for 20 min at 37°C. This cDNA was used as a template in 3 different fully-nested Taq-Polymerase PCRs with a 25 µL and 50 µL reaction volume for first and second round, respectively. Primer concentrations were 400 nM for each forward and reverse primer, 10x buffer magnesium chloride free, nucleotide mix (containing 200 nM of each dNTP), 2 mM magnesium chloride solution (Thermo Fisher Scientific), 2 U of Platinum Taq DNA Polymerase, and 1 µL of cDNA (for the first round) and 2 µL PCR product (of the first round) for the second round. The thermal cycling protocol started with 3 min at 95°C for denaturation, followed by 45 cycles of 15 sec at 95°C, 20 sec at 58°C, and 90 sec at 72°C, then a final elongation step for 3 min at 72°C. The amplicons were sequenced by Sanger sequencing (primers listed in SI Appendix **Table S2**). As a positive control, a circularization assay was applied on HDV (HBV/HDV positive human serum, SI Appendix **Fig. S5 B**) under the same PCR conditions described above, using the primers listed in SI Appendix **Table S2**.

RT-PCR amplicon quantification

For quantification, a serially diluted and photometrically quantified *in vitro* transcript (IVT) was used in all real-time RT-PCR runs. Using the initial deltavirus-positive *P. semispinosus* blood extract as a template, a reverse transcribed PCR fragment encompassing a part of the S-RDeAg coding region was amplified. SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Thermo Fisher Scientific) was used for PCR in a 25 µL reaction volume. Primer concentrations were 600 nM for forward and reverse primers, 12.5 µL 2x kit reaction buffer (containing 400 nM of each dNTP and 3.2 mM magnesium sulfate), 1 µL bovine serum albumin (1 mg/mL), 0.4 µL of a 50 mM magnesium sulfate solution (Thermo Fisher Scientific), 1 µL enzyme mix, 2.1 µL of RNase-free water, and 5 µL RNA extract. The thermal cycling protocol started with 20 min at 55°C for reverse transcription, followed by 3 min at 94°C, 45 cycles of 15 sec at 94°C, 20 sec at 58°C, and 40 sec at 72°C, ending with a final elongation for 2 min at 72°C.

The amplicon was cloned using TOPO TA Cloning Kit (Thermo Fisher Scientific) and chemically competent *E. coli* cells. The ligation mix contained 0.5 μ L salt solution from the cloning kit, 2 μ L PCR product and 0.5 μ L pCR4 TOPO vector. This mix was incubated for 30 min at room temperature and 2 min on ice. Two μ L of the ligation mix were blended into the competent cells, incubated on ice for 25 min, followed by a heat shock at 42°C for 30 sec and finally incubated on ice for 1 min. 250 μ L SOC medium were added to the transformation mix, shaken for 1 hour at 37°C, plated on LB-plates with kanamycin, and incubated at 37°C. Clones were subjected to PCR and Sanger sequencing and those with the correct insert orientation were incubated overnight in 2 mL SOC medium. Plasmid colonies were extracted from the overnight culture by QIAprep Spin Miniprep Kit with the standard protocol and diluted in RNase-free water with a factor up to 10^{-10} . A dilution row ranging from dilution factor 10^{-6} to 10^{-10} , in 10^{-1} steps, was tested in a PCR with M13 primers. The Taq-Polymerase PCR contained a 50 μ L reaction volume. Primer concentrations were 200 nM (1 μ L of 10 μ M) for forward and reverse primers, 10x buffer magnesium chloride free, nucleotide mix (with 200 nM end concentration of each dNTP), 2 mM magnesium chloride solution (Life Technologies), 2 U of Platinum Taq DNA Polymerase, and 1 μ L of plasmid dilution. The thermal cycling protocol started with 3 min at 95°C for denaturation, followed by 45 cycles of 15 sec at 95°C, 15 sec at 58°C, and 30 sec at 72°C, ending with a final elongation for 1 min at 72°C. To reduce the plasmid background, we used the dilution with the last clearly visible band on a 2% agarose gel before fading (dilution factor 10^{-6}). The PCR product was purified using QiaQuick PCR Purification Kit (Qiagen) according to manufacturer's instructions. The purified PCR product was transcribed using MEGAscript T7 Kit (Thermo Fisher Scientific) and purified again using RNeasy Mini kit (Qiagen), all according to manufacturer instructions. The RNA concentration was determined by Nanodrop (Thermo Fisher Scientific) from which the amount of IVT was calculated. The IVT was diluted in nuclease-free water containing 10 μ g/mL carrier RNA (Qiagen).

Next-generation sequencing

RNA was quantified using the Qubit RNA HS Assay kit (Thermo Fisher Scientific). Since all sample concentrations were too low, 5 μ L of extracted RNA were subjected to library preparation using the KAPA RNA Hyper Prep kit (Roche Molecular Diagnostics) with half volume reactions. The only changes to a standard procedure was that the fragmentation was performed at 85 °C for 5 minutes and the amount of adapter in the adapter ligation was lowered to 0.75 μ M. To purify the library, a 0.8x bead clean-up was performed (55 μ L library and 44 μ L beads) with KAPA Pure Beads and eluted with EB buffer (Qiagen). Library amplification started with a denaturation at 98°C for 45 sec followed by 12 cycles of 15 sec at 98°C, 30 sec at 60°C, and final extension at 72°C for 1 min, followed by a 0.7x bead clean-up. Indexed DNA libraries were then normalized to 1.5 nM using RNase-free water and were pooled. The RNA concentration of the pool was measured by Qubit dsDNA HS Assay kit (Thermo Fisher Scientific). Pooled libraries were denatured using 0.2 M NaOH and neutralized with Tris-HCL 0.2 M/pH 7. The pools were diluted with Hybridization buffer (HT1, Illumina) to a final concentration of 13 pM. Sequencing was performed using the 600-cycle MiSeq reagent v3 cartridge (Illumina). The preparation of the flow cell and the set-up of the instrument were performed according to manufacturer instructions.

Construction of a rodent deltavirus genome dimer, transfection in HuH7 cells, and clonal expansion

A double-stranded DNA fragment of the RDeV genome (GenBank accession number MK598004) was synthesized (IDT, Leuven, Belgium) and inserted in genomic orientation (non-coding) into linearized pcDNA3.1+ using the In-Fusion HD Cloning Kit according to manufacturer's protocol (Takara). The plasmid containing the monomeric RDeV genome was amplified with the non-phosphorylated primers pcDNA3.1-Hind-rev and rodHDV-for, while the RDeV genome was amplified using the phosphorylated primers rodHDV-for-P and rodHDV-rev-P (primers listed in SI Appendix **Table S2**). The amplicons were ligated using T4 DNA Ligase (Roche) according to manufacturer's instructions, resulting in an expression plasmid containing a head-to-tail dimer of the RDeV genome (construct map shown in SI Appendix **Fig. S6**). Insertion and orientation was assessed by colony PCR using pSEM-117F and pSEM-1540R primers (SI Appendix **Table S2**).

To construct the mutant RDeV dimer and abrogate the expression of RDeAg, the start codon of the RDeAg ORF was mutated from AUG to AUU in the monomeric RDeV construct. A dimer with the mutated start codon on the RDeAg ORF in both copies of the genome was constructed as described above.

Human hepatoma cells (HuH7) were cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and antibiotics. Transfection was performed using Fugene HD (Promega) according to the manufacturer's protocol. Two days post transfection, media were replaced by William's E Medium (Thermo Fisher Scientific) supplemented with 2% fetal bovine serum and antibiotics.

If replication of the RDeV genome occurs, then cells containing this genome will pass it onto the next generation cells. Therefore, a clonal expansion experiment was performed by transfecting two wells with the dimeric RDeV, the mutant RDeV dimer, and a trimer of the human HDV genotype 1 (pSVL(D3)) as already described. Four days post transfection, one well was fixed while another well was trypsinized and split in a 1:64 dilution. Cells were further cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and antibiotics. On day 6 post-splitting, the cells were fixed and stained with anti-HDAg IgGs from an HBV/HDV-co-infected patient for RDeAg and HDAg expression as described below.

Immunofluorescence

To detect antibodies against RDeV in rodent serum samples, Vero B4 cells were transfected with a plasmid overexpressing FLAG-tagged L-RDeAg following laboratory procedures as described in (30, 31). Cells were fixed with 4% Roti-Histofix (Roth) and permeabilized as described in (28). FLAG-L-RDeAg expression was confirmed by applying a 1:100 dilution of a mouse anti-FLAG antibody. *P. semispinosus* and human serum samples were diluted 1:100 in sample buffers (Euroimmun, Germany). The slides were then incubated for 1 hour at 37°C and washed with PBS containing 0.1% Tween (PBS-T). Secondary detection was done by applying a Cy3-labeled goat anti-mouse IgG antibody (1:200 dilution in PBS, Dianova) and an Alexa 488-labeled anti-human or anti-guinea pig IgG antibody (1:200 dilution in PBS, Dianova), incubating for 1 hour at 37°C. Slides were washed three times with PBS-T and rinsed once in

water. Nuclei were stained with DAPI Gold Mounting Medium (Thermo Fisher Scientific). For anti-HBc antibody detection, HuH7 cells were transfected with a 1.1 over-length expression plasmid of HBV as described in (25, 26). The HBcAg expression was controlled by applying a 1:200 dilution of anti-HBc reactive serum produced in rabbits (30, 31) and a Cy2-labeled anti-rabbit antibody (1:200 dilution in PBS, goat anti-rabbit antibody, Dianova).

For detection of S- and L-RDeAg, HuH7 cells were transfected with plasmids expressing S-RDeAg, a stop codon mutant expressing a hypothetical L-RDeAg, as well as an intact dimer of the RDeV genome with Fugene HD according to the manufacturer's protocol (Promega). After fixation and permeabilization, cells were stained with antigen-specific IgGs from immunized rabbits followed by Alexa 488-labeled goat anti-rabbit IgG (H+L) secondary antibody (1:400 in PBS, Thermo Fisher Scientific) as described above. Immunofluorescence signals were analyzed with an inverted fluorescence microscope (Leica DMI6000B, Leica).

Generation of specific antisera against S- and L-RDeAg

For the generation of antiserum against the putative L-RDeAg, two rabbits were immunized with a synthetic peptide (NH₂-C+GNLEAKGEPPTSRKIPE-COHN₂) conjugated to the carrier protein keyhole limpet hemocyanin (KLH). For the generation of antiserum against S-RDeAg, a mixture of two peptides (NH₂-C+RGDGLSLRGEGEYPW-COOH and NH₂-C+GGDGDVNPPEGTPRG-CONH₂) conjugated to KLH was used for immunization. Peptide synthesis, immunization and bleeding of immunized rabbits was performed by Eurogentec (Eurogentec S.A., Seraing, Belgium).

Northern blot analysis

Total RNA was extracted from transfected cells using TRIzol reagent (Thermo Fisher Scientific) and 10 µg RNA was separated using a 1% denaturing formaldehyde agarose gel. RNA was vacuum-blotted onto a positively charged nylon membrane (GE Healthcare, Amersham). After blotting, the membranes were stained with 0.04% methylene blue/ 0.5 M sodium acetate (pH 5.2) for 10 minutes. After destaining with DEPC ddH₂O, pre-hybridization was performed at 65°C in hybridization buffer (5x SSC, 5x Denhardt's solution, 1% [w/v] SDS and 100 µg/ml salmon sperm DNA) for 18 hours. For detection of antigenomic RDeV RNA, a genomic probe (5'-TGAGGGGCCCGACCTCCATCAACCTCCATTTCTCGCCAAGGCGTCTTTTCTTG GCCTG-3') was 5'-labeled with T4 Polynucleotide Kinase (Thermo Fisher Scientific) and [γ -32P] ATP. For detection of genomic RDeV RNA, the reverse complementary sequence was used. After labeling, free nucleotides were removed by MicroSpin G-25 columns (GE Healthcare) according to the manufacturer's instructions. For hybridization, the corresponding labeled oligonucleotide probe was added and incubated at 65°C overnight. Low-stringency washing was performed with 2x SSC and 0.1% [v/v] SDS) at room temperature for 5 minutes. After two low-stringency washing steps, washing was completed at room temperature for 15 minutes with a high-stringency buffer containing 0.2% SSC and 0.1% [v/v] SDS. Radioactivity was visualized with a Fujifilm FLA-7000 phosphorimager (Typhoon FLA, GE Healthcare, Amersham). Genomic and antigenomic RDeV RNA was generated by *in vitro* transcription of PCR products from the monomeric RDeV construct, containing either the SP6 or T7 promoter sequence.

Western blot and ELISA analyses

Total protein from transfected cells and *P. semispinosus* tissues were isolated using TRIzol Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. Proteins were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis in a 14% separation gel. After blotting onto a 0.45 µm pore-size nitrocellulose membrane (Millipore), TBS-T (TBS containing 0.1% Tween-20) supplemented with 5% nonfat dry milk was used for blocking at 4°C overnight. The membrane was incubated with anti-HDAg IgGs from an HBV/HDV-co-infected patient (1:10,000) and a mouse monoclonal to β-actin antibody (1:10,000, Abcam) diluted in blocking solution for 1 hour at room temperature. After extensive washing with TBS-T, the membrane was incubated with goat anti-human IgG (H+L)-HRPO (1:10,000, Dianova) and goat anti-mouse IgG-HRP (1:10,000, Santa Cruz) for 30 minutes at room temperature. After three more washing steps, SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) was applied to the membrane and proteins were visualized using a ChemoCam Imager (Intas).

For the detection of specific anti-S-RDeAg antibodies, total protein was extracted as described above from HuH7 cells transfected with a plasmid overexpressing S-RDeAg. Protein separation and blotting onto the nitrocellulose membrane was performed as above. The membrane was divided and incubated with five different *P. semispinosus* sera (1:1000) diluted in TBS-T supplemented with 5% nonfat dry milk for 1 h at room temperature. After extensive washing with TBS-T, the membranes were incubated with goat anti-guinea pig (H+L)- HRP (1:100, Abcam) for 1 h at room temperature. The anti-S-RDeAg serum from immunized rabbits with S-RDeAg (described above) was used as positive control, applying goat anti-rabbit (H+L)- HRP (1:2000, Cell Signaling Technology) as secondary antibody.

A commercial HDAg ELISA was used according to the manufacturer's protocol (ETI-Deltak-2, DiaSorin) using 30 µg of total protein.

Read mapping and *de novo* assembly

Next-generation sequencing reads were mapped to the reference host genome sequence of the rodent species *Tympanoctomys barrerae* using bwa version 0.7.17-r1188 (32). *T. barrerae*, which belongs to the Hystricomorpha suborder with *P. semispinosus*, was used because no *Proechimys* genome is available and it is the closest species with an available genome. Reads that did not match the host genome were assembled *de novo* using SPAdes version 3.11.1 (33). All sequences from the resulting contig library were translated into all six reading frames and matched against a viral reference library (27) using BLAST (version 2.6.0+ (34)). This revealed several high bit-score matches against human HDV. Cross-contamination of samples was excluded by comparing hepacivirus sequences found in the same samples (29). In no case were identical hepacivirus sequences obtained. Also, RNAseq runs were done on total RNA extracted from the sera of the four individuals that tested RNA-negative for hepacivirus, but that were carrying RDeV RNA, and returned no matching reads on *P. semispinosus* hepacivirus (matching done with the Geneious mapper for RNAseq, Geneious version 9.1.8, Biomatters, Auckland, New Zealand, <https://www.geneious.com>).

Antigenomic editing search in NGS reads

Total RNA from five organs of the rodent tested RNA-positive for RDeV was extracted and sequenced on the Illumina MiSeq platform (as described above). Next-generation sequencing

reads from the organs and all blood samples from which a full RDeV was generated were mapped anew onto their corresponding genomes with bwa version 0.7.17-r1188 (32), bowtie2 version 2.3.0 (35), and Geneious (version 9.1.8, Biomatters, Auckland, New Zealand, <https://www.geneious.com>). If the antigenomic strand of deltavirus in rodents undergoes RNA editing at the second amber codon position of S-RDeAg during replication, then two species of NGS reads should be detected at that position. Reads were also matched against a custom BLAST database of all RDeV full genomes, however, none of the matching/mapping strategies revealed two species of NGS reads for RDeVs (SI Appendix **Fig. S3**). As a control, the same strategies were applied to the transfected trimer of the human HDV genotype 1, pSVL(D3) (GenBank accession number M21012), for which two species of NGS reads were found in all cases.

Alignment and phylogeny generation, statistics, and structure prediction tools

Full genome representative sequences of all eight HDV genotype groups along with RDeV, SDeV, duck-associated DeV, fish DeV, newt DeV, termite DeV, and toad DeV were aligned using MAFFT version 7.309 (36) on the basis of a nucleotide alignment. Maximum Likelihood inference of phylogeny was done using RAxML-NG version 0.7.0 BETA (37).

All statistical analyses were carried out using R version 3.4.4 (38). Evolutionary phylogenetic distances plotted in **Fig. 3** were extracted using the *ape* R package. For the logistic regression model presented in SI Appendix **Table S9**, the presence/absence of RDeV was the binomial response variable against which all other variables were compared. Explanatory variables (and their reference groups in parentheses) are: sex (males), reproductive status (reproductive), habitat (forest fragment surrounded by agriculture), and capture season type (wet). Confounding effects were excluded using the *vif* function from the R library *car* version 3.0-2 (VIF < 1.2 for both estimated models in SI Appendix **Table S9**).

GenBank accession numbers of the *deltavirus-interacting proteins a* (DIPA) used for nucleotide and amino acid comparison respectively are: NM_198616 and NP_941018 (mouse), and XM_006230939 and XP_006231001 (rat).

Predictions for the secondary structures of the RDeV genomic and antigenomic ribozymes were carried out with the TT2NE algorithm (39).

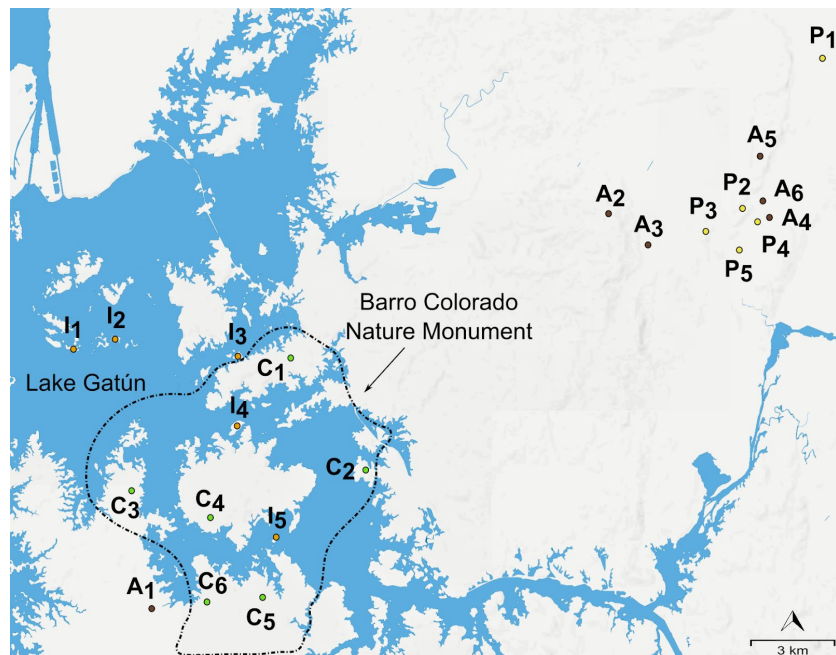


Fig. S1. Geographical sampling locations of *P. semispinosus*. Abbreviations are: C: continuous forest (green), I: forested island (orange), A: forest fragment surrounded by agriculture (brown), P: teak plantation (yellow). SI Appendix **Table S7** lists the number of individuals collected and tested positive or negative for RDeV RNA at each site.

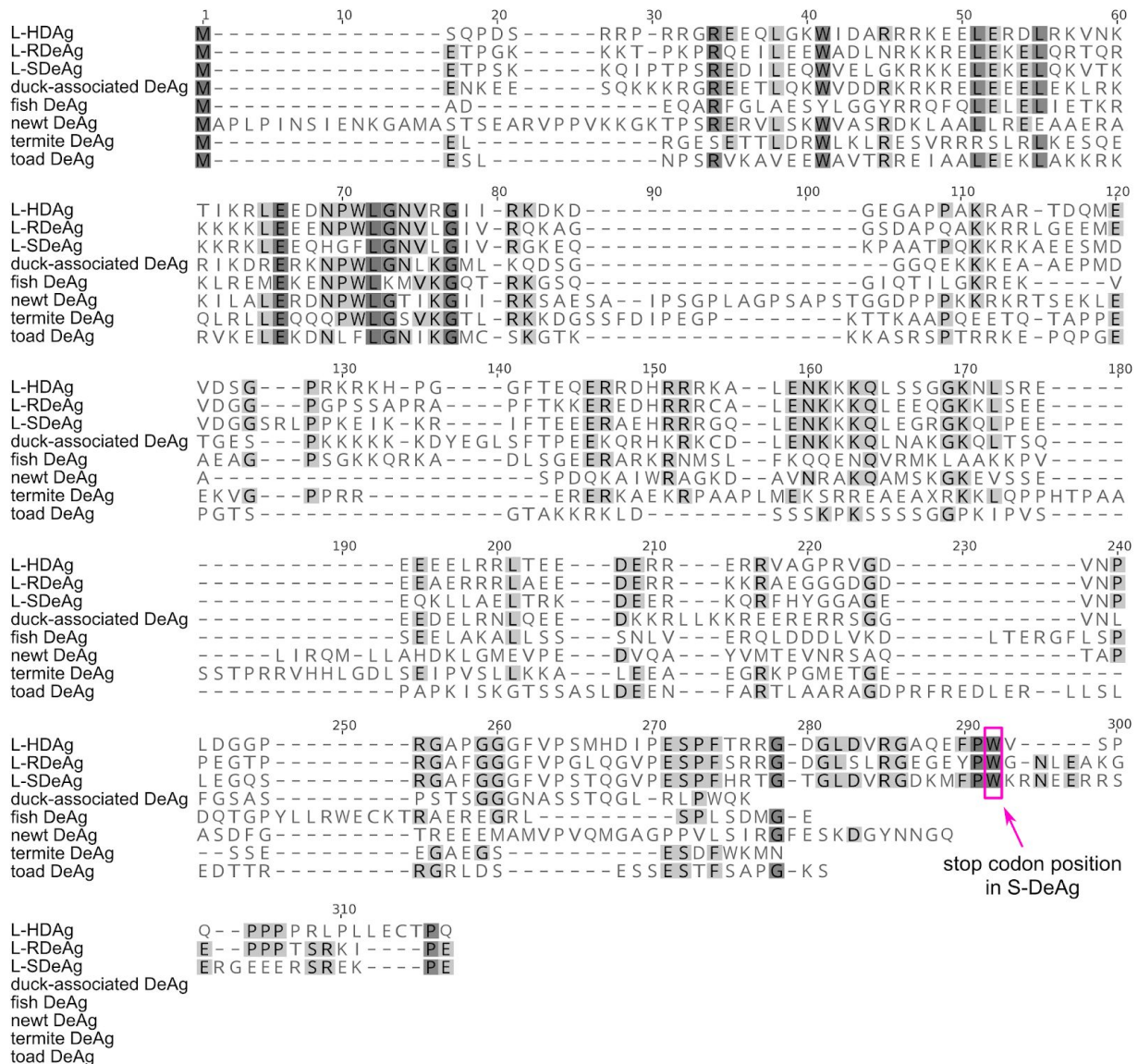


Fig. S2. Amino acid alignment of human L-HDAg (GenBank accession number AF018077), (putative) L-RDeAg (GenBank accession number MK598004), (putative) L-SDeAg (GenBank accession number AYP55701), duck-associated DeAg (GenBank accession number AYC81245), fish DeAg (GenBank accession number MN031240), newt DeAg (GenBank accession number MN031239), termite DeAg (GenBank accession number MK962759), and toad DeAg (GenBank accession number MK962760). Amino acid similarity is shown by grey shading (dark grey indicates 100% identity). While the occurrence of L-RDeAg and L-SDeAg has not been confirmed, we include their 19- and 22 amino acid tails before the appearance of the next stop codon for comparison to L-HDAg.



Fig. S3. Search for antigenomic editing of NGS reads. The upper panel shows pSVL(D3) reads mapped onto the respective reference genome (trimer of human HDV genotype 1, GenBank accession number M21012). The lower panel shows RDeV reads mapped onto the respective reference genome (RDeV genome, GenBank accession number MK598004). Both samples correspond to HuH7 cells transfected with the pSVL(D3) and the dimeric RDeV constructs respectively, and harvested 6 days post-transfection. The nucleotide position where editing takes place is highlighted, showing two species of NGS reads for the human HDV in contrast to RDeV where antigenomic editing does not occur.

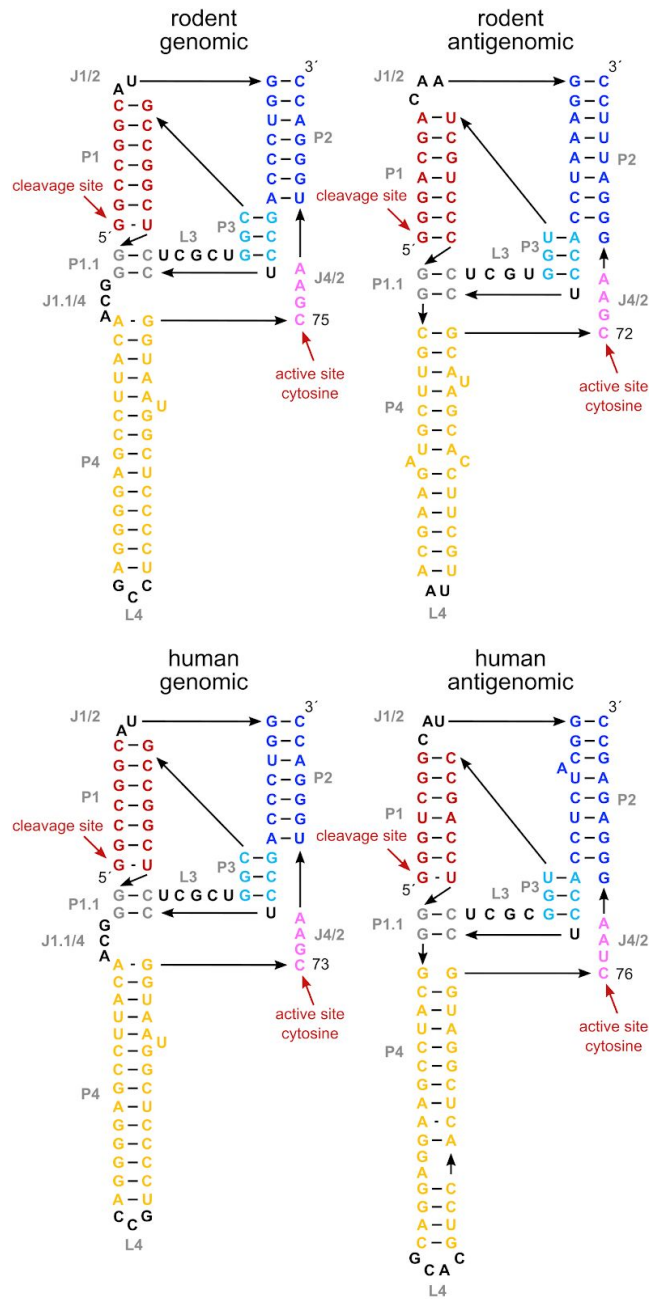


Fig. S4. Secondary structure predictions of genomic and antigenomic ribozymes of rodent deltavirus. Predictions were carried out using the TT2NE algorithm (39). Human HDV genomic and antigenomic ribozyme structures are shown for comparison.

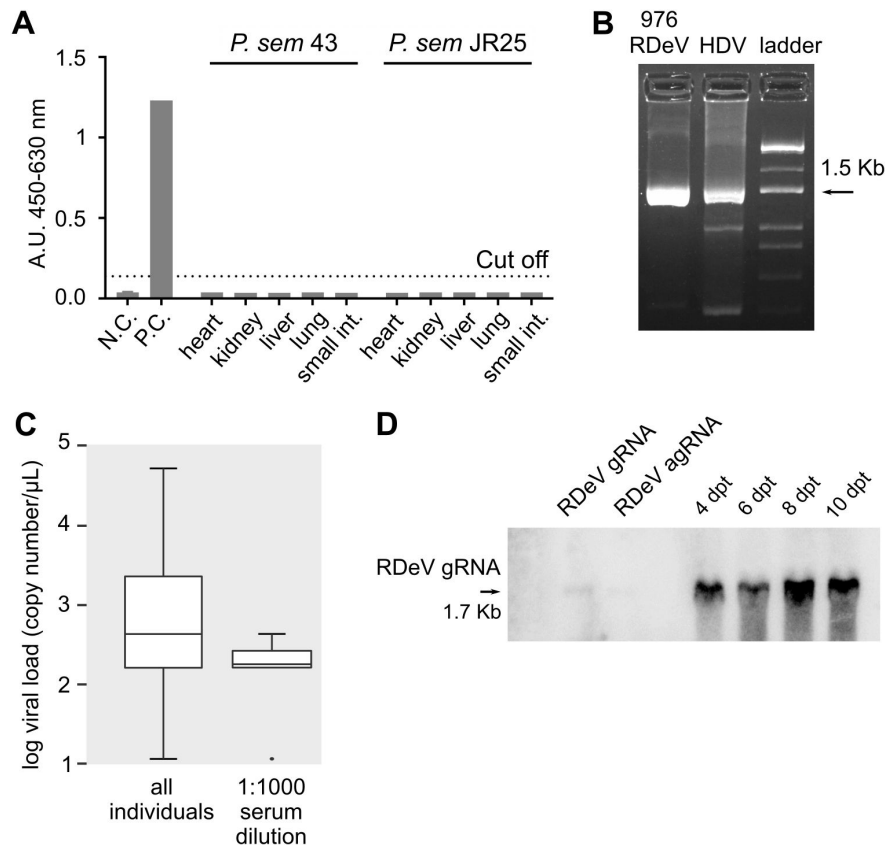


Fig. S5. A. Detection of RDeAg from *P. semispinosus* (*P. sem*) organs via a commercial HDAg ELISA assay. HuH7 cells transfected with an expression plasmid containing the RDeV dimer were used as positive control (P.C.), while mock transfected cells served as negative control (N.C.). Cells were lysed six days post-transfection. Lysates from organs of RDeV RNA-positive (*P. sem* 43) and RNA-negative (*P. sem* JR25) animals were isolated with TRIzol and 30 μ g total protein was used. **B.** Products of deltavirus circularization assay 3 on *P. semispinosus* blood (976 RDeV product contained in GenBank accession number MK598009) and circularization assay on human HDV (serum of HBV/HDV positive patient). Fragment sizes are 1532 bp for RDeV and 1511 bp for HDV. **C.** Distribution of viral load (log copy number per μ L) in all S-RDeAg-positive *P. semispinosus* individuals tested in 1:100 and 1:1000 serum dilution (left), and individuals tested positive at 1:1000 serum dilution (right). The box plots show median, interquartile range, Tukey's minimum and maximum (whiskers), and extreme outlier values (dotted). **D.** Northern blot of genomic RDeV RNA from transfected HuH7 cells. Cells were transfected with an expression plasmid containing a dimer of the RDeV genome. Total RNA was isolated 4, 6, 8, and 10 days post-transfection and subjected to a 1% denaturing formaldehyde agarose gel. After blotting, the RNA was hybridized with a single-stranded oligonucleotide probe labeled with [γ - 32 P]ATP and T4 polynucleotide kinase at the 5' end. Two nanograms of *in vitro*-transcribed (anti-)genomic RDeV RNA were used as a control.

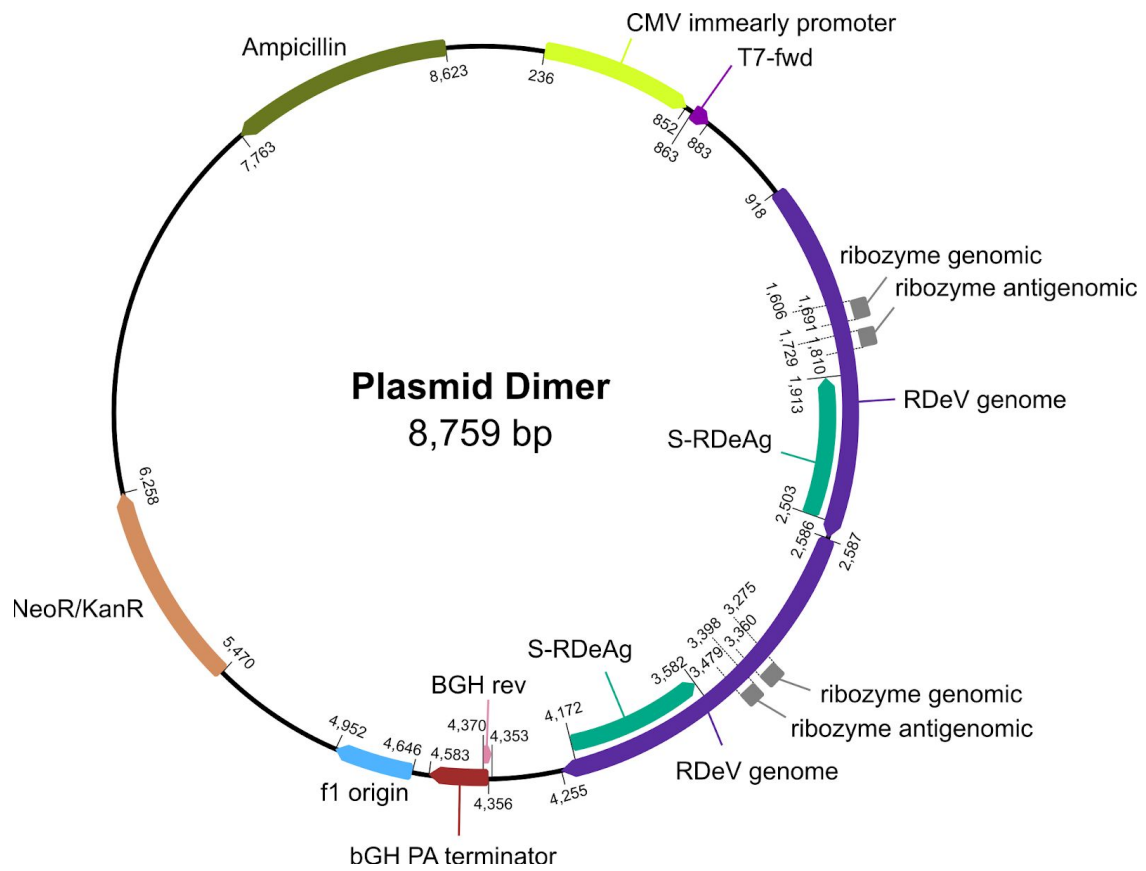


Fig. S6. Expression construct of the RDeV genome dimer. RDeV is cloned as a tandem head-to-tail fusion construct in genomic (minus strand) orientation, under a CMV promoter. The coding region of the S-RDeAg and both ribozyme structures are annotated on both RDeV genomes. The ORFs are indicated by arrows showing the direction of transcription. Primer sequence correspondence can be found in SI Appendix **Table S2**.

Table S1. Number of individuals for eight rodent- (Rodentia) and four marsupial- (Didelphimorphia) species tested for rodent deltavirus RNA with qRT-PCR.

Order, family, and species	Number of blood samples	RDeV RNA-positive
Rodentia, Echimyidae		
<i>Proechimys semispinosus</i>	763	30
<i>Hoplomys gymnurus</i>	21	-
Rodentia, Heteromyidae		
<i>Heteromys desmarestianus</i>	3	-
<i>Liomys adpersus</i>	1	-
Rodentia, Cricetidae		
<i>Zygodontomys brevicauda</i>	3	-
<i>Sigmodon hirsutus</i>	2	-
<i>Oecomys bicolor trinitatis</i>	1	-
<i>Transandinomys talamancae</i>	1	-
Didelphimorphia, Didelphidae		
<i>Didelphis marsupialis</i>	90	-
<i>Philander opossum</i>	51	-
<i>Marmosa spp</i>	16	-
<i>Metachirus nudicaudatus</i>	4	-

Table S2. Primers used for PCR screening assays, virus quantification, and cloning. An asterisk indicates identical primers, used in multiple assays.

Name	Sequence (5'-3')	Primer direction	Assay used
rtHDVPsemAG-F1	AGG AAA GGG AGG ACC ATC GC	forward	Screening qRT-PCR targeting HDAg
rtHDVPsemAG-R1	GCC TCT TCC TCC TCG CTC A	reverse	Screening qRT-PCR targeting HDAg
rtHDVPsemAG-P (probe)	FAM-AGA AGC AGC-ZEN-TGG AGG AGC AAG GA-IBFQ	forward	Screening qRT-PCR targeting HDAg
Tbp_f	AA Y CTT GGT TGT AAA CTT GAC CT	forward	TBP qRT-PCR
Tbp_r	GGG CTC CCT TAT TCT CAT GA	reverse	TBP qRT-PCR
Tbp_p (probe)	FAM-CCG AAA TGC-ZEN-TGA ATA TAA TCC CAA GC-IBFQ	forward	TBP qRT-PCR
HDV_Psem7-Fwd	CGA CGG CTC GCC GAG GA	forward	Circularization assay 1 first round
HDV_Psem7-Rev	CTT CTT TCC TTG CTC CTC CAG C	reverse	Circularization assay 1 first round
HDV_Psem7-Fnest	TCG CCG AGG AGG ACG AAC GTC	forward	Circularization assay 1 second round
HDV_Psem7-Rnest	TTT GTT CTC GAG GGC ACA CCT TCG	reverse	Circularization assay 1 second round
HDV_PsemCloneF1*	GAG GAA GAA GAA GAA GCT TGA GG	forward	Circularization assay 2 first round
rtHDVPsem3-R1	TCC TGT CTG GGC TTG GGA GT	reverse	Circularization assay 2 first round
HDV_Psem5-F	CAT GGC TGG GTA ACG TTC TTG GAA	forward	Circularization assay 2 second round
HDV_Psem8-Rev	ACC GGG AGT CTC CAT CCT GGA GT	reverse	Circularization assay 2 second round
rtHDVPsemRI-F1	TCC TCT TCG GGA CGA CAA GG	forward	Circularization assay 3 first round
HDV_PsemCloneR1**	CTA GGG ATA TTC CCC TTC TCC TCT	reverse	Circularization assay 3 first round
rtHDVPsemRI-F2	CTT CGG GAC GAC AAG GAA ATC C	forward	Circularization assay 3 second round
HDV_Psem5-R	CGT CTC CTC TTC TGG AGA ATG G	reverse	Circularization assay 3 second round

hHDV1_2-Fwd	TCC TCT TCG GGT CGG CAT G	forward	Human HDV circularization assay first round
hHDV1_2-Rev	ATC GGA TGG AAA GAG TAT ATC C	reverse	Human HDV circularization assay first round
hHDV1_2-Fnest	GGT CGG CAT GGC ATC TCC	forward	Human HDV circularization assay second round
hHDV1_2-Rnest	ACT CCG GAA CTC CTT GCA T	reverse	Human HDV circularization assay second round
HDV_PsemCloneF1*	GAG GAA GAA GAA GAG GCT TGA GG	forward	Generation of cloning fragment
HDV_PsemCloneR1**	CTA GGG ATA TTC CCC TTC TCC TCT	reverse	Generation of cloning fragment
T7-fwd	TAA TAC GAC TCA CTA TAG GG	forward	Sequencing cloning fragment
BGH-rev	TAG AAG GCA CAG TCG AGG	reverse	Sequencing cloning fragment
Seq-Dimer-F	TCG CTC TCC GGA TGG	forward	Sequencing cloning fragment
pcDNA3.1-Hind-rev	CAA GCT TAA GTT TAA ACG CTA GC	reverse	PsemD1 amplification
rodHDV-for	ATG GAA ACA CCT CCT GGA GAA G	forward	PsemD1 amplification
rodHDV-for-P	P-ATG GAA ACA CCT CCT GGA GAA G	forward	cDNA genome amplification
rodHDV-rev-P	P-CCG GAG AGC GAG GAC CG	reverse	cDNA genome amplification
pSEM-117F	GCT TGG GAG TTT TCT TCT TAC C	forward	Colony PCR, sequencing cloning fragment
pSEM-1540R	CTG CGT TTC CAG CAG CTA G	reverse	Colony PCR, sequencing cloning fragment

Table S3. Average nucleotide and amino acid percent identities among HDV, RDeV, SDeV, duck-associated DeV, fish DeV, newt DeV, termite DeV, and toad DeV genomes and small delta antigens. The three percentage identities in each cell are, in order: full-genome nucleotide identity, small delta antigen nucleotide identity, and small delta antigen amino acid identity. The HDV-to-HDV and the RDeV-to-RDeV identities are the average identities among all HDV and RDeV sequences used for phylogenetic inference.

	HDV	RDeV	SDeV	Duck-associated DeV	Fish DeV	Newt DeV	Termite DeV	Toad DeV
HDV	74/80/74	46/59/54	41/54/47	37/48/38	32/35/21	32/32/23	30/35/21	29/36/23
RDeV	-	97/99/99	42/59/54	36/47/35	29/34/20	31/34/20	29/31/20	24/32/20
SDeV	-	-	-	35/42/35	28/31/17	28/29/17	28/29/20	24/29/19
Duck-associated DeV	-	-	-	-	25/33/15	27/28/19	25/29/15	25/30/16
Fish DeV	-	-	-	-	-	28/27/12	26/26/13	21/24/14
Newt DeV	-	-	-	-	-	-	26/26/11	28/28/15
Termite DeV	-	-	-	-	-	-	-	21/26/13
Toad DeV	-	-	-	-	-	-	-	-

Table S4. Number of RDeV RNA copies in five organs of a *P. semispinosus* individual.

Organ	RDeV RNA copies per gram of tissue
Heart	1,045,688
Kidney	1,855,127
Liver	955,878
Lung	311,926
Small intestine	1,589,207

Table S5. Number of *P. semispinosus* individuals tested in different types of material by qRT-PCR for the presence of RDeV RNA.

	Individuals tested	RNA-positive individuals per material	Blood-only RNA-positive individuals	Feces-only RNA-positive individuals	Organs-only RNA-positive individuals
Blood	763	30	25	-	-
Feces	822	10	-	5	-
Organs	18	1	-	-	1

Table S6. Outcome counts of qRT-PCR-tested *P. semispinosus* for the presence of RDeV and hepacivirus RNA.

	RDeV RNA-positive	RDeV RNA-negative
Hepacivirus RNA-positive	26	568
Hepacivirus RNA-negative	4	164

Table S7. Sampling sites of *P. semispinosus* blood samples collected and tested for presence of RDeV and hepacivirus RNA. Site names correspond to marked locations on the map (SI Appendix Fig. S1). Total number and percent of RNA-positive individuals for both viruses are shown.

Sampling site	Number of <i>P.semispinosus</i> tested	RDeV RNA-positive (%)	Hepacivirus RNA-positive (%)
I ₁	113	9 (7.9)	108 (95.6)
I ₂	120	0 (0)	93 (77.5)
I ₃	65	0 (0)	61 (93.8)
I ₄	23	1 (4.3)	21 (91.3)
I ₅	29	0 (0)	26 (89.6)
C ₁	27	0 (0)	26 (96.3)
C ₂	93	1 (1.1)	84 (91.3)
C ₃	30	7 (23.3)	24 (80)
C ₄	22	2 (9.1)	16 (72.7)
C ₅	18	4 (22.2)	15 (83.3)
C ₆	25	3 (12)	20 (80)
A ₁	9	1 (11.1)	9 (100)
A ₂	15	0 (0)	9 (60)
A ₃	5	0 (0)	3 (60)
A ₄	32	1 (3.1)	9 (28.1)
A ₅	57	0 (0)	27 (47.4)
A ₆	40	0 (0)	32 (80)
P ₁	4	0 (0)	3 (75)
P ₂	6	0 (0)	4 (66.7)
P ₃	10	0 (0)	2 (20)
P ₄	10	0 (0)	2 (20)
P ₅	10	1 (10)	0 (0)

Table S8. Summary table of *P. semispinosus* blood samples tested with a real-time RT-PCR assay to detect RNA transcripts of the host TBP (TATA-binding protein) gene. CT (cycle threshold) values for the TBP real-time RT-PCR assay are shown, as well as the presence (+) or absence (-) of RDeV RNA and anti-RDeAg antibodies in the corresponding samples.

Sample ID	CT value of TBP real-time RT-PCR	RDeV RNA	Anti-RDeAg antibodies
34	36.49	+	-
39	35.06	-	+
54	34.51	-	-
192	34.66	+	+
959	30.54	-	+

Table S9. Logistic regression models examining the effect of different variables on rodent deltavirus acquisition. Nominal variables and their states are abbreviated in parentheses. Sex: male (M), female (F); reproductive (R), non-reproductive (NR); habitat: continuous forest (C), forested island (I), forest fragment surrounded by agriculture (A), teak plantation (P); capture season type: wet (Oct-Dec), dry (Jan-Apr). Sampling size n=686 individuals for all models. Values shown represent the correlation coefficients with standard errors in parentheses. Chi-squared p-values for model comparison are shown at bottom. Asterisks indicate the level of significance of the corresponding variable for each model evaluation.

Variables	Model 1	Model 2
Sex (M)	0.94** (0.41)	0.87** (0.41)
Reproductive status (R)	1.82*** (0.62)	1.8*** (0.62)
(C)		1.83** (0.77)
Habitat (I)		0.99 (0.79)
(P)		2.15 (1.32)
Capture season type (wet)		0.09 (0.46)
AIC	222.61	221.16
R ²	0.12	0.16
degrees of freedom	3	7
x ² model comparison	Model 2 vs. Model 1 0.05**	

*** p < 0.01, ** p < 0.05, p* < 0.1

Table S10. HDAG motifs and their functions. Presence/absence designation refers to corresponding residues found/not found in RDeAg and not to post-translational modifications.

Motifs and post-translational modifications	Description of function	Presence / absence in RDeAg
NLS 66-EGAPPAKRAR-75	Nuclear Localization Signal. Introducing the HDV RNA to the nucleus	present
Leucine zipper	HDAG oligomerization that facilitates nuclear transportation: found NOT to have this function (16, 40)	present
Coiled-coil domain (CCD)	Multimer formation and activation/inhibition of HDV RNA replication	present
Arginine-rich motifs (ARMs)	HDAG binding activity: found NOT to have this function (41)	present
NES (Proline-rich region)	Nuclear export of L-HDAG	present
N-terminal Cys-farnesylation CXXQ	HDV particle assembly: packaging with HBV envelope proteins	absent
Arg-13 methylation	Nuclear transportation of S-HDAG Antigenomic RNA replication Nuclear transport of antigenomic RNA Absence of methylation facilitates genomic RNA nuclear transport	present
Lys-72 acetylation	Nuclear transport of L-HDAG Enhancing synthesis of genomic RNA and mRNA Acetylation absence enhances antigenomic RNA replication	present

Ser-177 phosphorylation	Enhancing genomic RNA replication by interacting with RNA Pol-II	present
Lys-sumoylation	Selectively enhancing genomic and mRNA synthesis, but NOT antigenomic	present
Ribozymes	Self-cleaving HDV RNA	present

References

1. C. Sureau, F. Negro, The hepatitis delta virus: Replication and pathogenesis. *J. Hepatol.* **64**, S102–S116 (2016).
2. J. C. Otto, P. J. Casey, The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells. *J. Biol. Chem.* **271**, 4569–4572 (1996).
3. J. Perez-Vargas, *et al.*, Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat. Commun.* **10**, 2098 (2019).
4. M. Wille, *et al.*, A Divergent Hepatitis D-Like Agent in Birds. *Viruses* **10**, 720 (2018).
5. U. Hetzel, *et al.*, Identification of a Novel Deltavirus in Boa Constrictors. *MBio* **10**, e00014–19 (2019).
6. W.-S. Chang, *et al.*, Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol* **5**, vez021 (2019).
7. S.-Y. Hong, P.-J. Chen, Phosphorylation of serine 177 of the small hepatitis delta antigen regulates viral antigenomic RNA replication by interacting with the processive RNA polymerase II. *J. Virol.* **84**, 1430–1438 (2010).
8. W.-H. Huang, Y.-S. Chen, P.-J. Chen, Nucleolar targeting of hepatitis delta antigen abolishes its ability to initiate viral antigenomic RNA replication. *J. Virol.* **82**, 692–699 (2008).
9. Y. Yamaguchi, T. Mura, S. Chanarat, S. Okamoto, H. Handa, Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. *Genes Cells* **12**, 863–875 (2007).
10. N. Abeywickrama-Samarakoon, *et al.*, Hepatitis Delta Virus histone mimicry drives the recruitment of chromatin remodelers for viral RNA replication. *Nat. Commun.* **11**, 419 (2020).
11. A. D. Branch, H. D. Robertson, A replication cycle for viroids and other small infectious RNA's. *Science* **223**, 450–455 (1984).
12. A. T. Perrotta, M. D. Been, A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* **350**, 434 (1991).
13. T. B. Macnaughton, S. T. Shi, L. E. Modahl, M. M. C. Lai, Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases. *J. Virol.* **76**, 3920–3927 (2002).
14. M. Y. Kuo, L. Sharmeen, G. Dinter-Gottlieb, J. Taylor, Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J. Virol.* **62**,

- 4439–4444 (1988).
15. D. W. Lazinski, J. M. Taylor, Relating structure to function in the hepatitis delta virus antigen. *J. Virol.* **67**, 2672–2680 (1993).
 16. M.-F. Chang, S. C. Chang, C.-I. Chang, K. Wu, H. Y. Kang, Nuclear localization signals, but not putative leucine zipper motifs, are essential for nuclear transport of hepatitis delta antigen. *J. Virol.* **66**, 6019–6027 (1992).
 17. C. Alves, N. Freitas, C. Cunha, Characterization of the nuclear localization signal of the hepatitis delta virus antigen. *Virology* **370**, 12–21 (2008).
 18. T. B. Macnaughton, M. M. C. Lai, Genomic but not antigenomic hepatitis delta virus RNA is preferentially exported from the nucleus immediately after synthesis and processing. *J. Virol.* **76**, 3928–3935 (2002).
 19. C.-H. Lee, S. C. Chang, C. H. H. Wu, M.-F. Chang, A novel chromosome region maintenance 1-independent nuclear export signal of the large form of hepatitis delta antigen that is required for the viral assembly. *J. Biol. Chem.* **276**, 8142–8148 (2001).
 20. S. A. Hughes, H. Wedemeyer, P. M. Harrison, Hepatitis delta virus. *Lancet* **378**, 73–85 (2011).
 21. Y.-J. Li, M. R. Stallcup, M. M. C. Lai, Hepatitis delta virus antigen is methylated at arginine residues, and methylation regulates subcellular localization and RNA replication. *J. Virol.* **78**, 13325–13334 (2004).
 22. J.-J. Mu, *et al.*, The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine 72 may influence its cellular localization and viral RNA synthesis. *Virology* **319**, 60–70 (2004).
 23. C.-H. Tseng, K.-S. Jeng, M. M. C. Lai, Transcription of subgenomic mRNA of hepatitis delta virus requires a modified hepatitis delta antigen that is distinct from antigenomic RNA synthesis. *J. Virol.* **82**, 9409–9416 (2008).
 24. Y.-S. Chen, W.-H. Huang, S.-Y. Hong, Y.-G. Tsay, P.-J. Chen, ERK1/2-mediated phosphorylation of small hepatitis delta antigen at serine 177 enhances hepatitis delta virus antigenomic RNA replication. *J. Virol.* **82**, 9345–9358 (2008).
 25. D. Barford, A. K. Das, M.-P. Egloff, The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 133–164 (1998).
 26. C.-H. Tseng, T.-S. Cheng, C.-Y. Shu, K.-S. Jeng, M. M. C. Lai, Modification of small hepatitis delta virus antigen by SUMO protein. *J. Virol.* **84**, 918–927 (2010).
 27. N. Goodacre, A. Aljanahi, S. Nandakumar, M. Mikailov, A. S. Khan, A Reference Viral

- Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection. *mSphere* **3**, e00069–18 (2018).
28. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
 29. J. Schmid, *et al.*, Ecological drivers of Hepacivirus infection in a neotropical rodent inhabiting landscapes with various degrees of human environmental change. *Oecologia* **188**, 289–302 (2018).
 30. J. F. Drexler, *et al.*, Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proceedings of the National Academy of Sciences* **110**, 16151–16156 (2013).
 31. B. F. de Carvalho Dominguez Souza, *et al.*, A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. *J. Hepatol.* **68**, 1114–1122 (2018).
 32. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
 33. A. Bankevich, *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
 34. C. Camacho, *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
 35. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
 36. K. Katoh, MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
 37. A. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2018).
 38. R. C. Team, R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016 (2017).
 39. M. Bon, H. Orland, TT2NE: a novel algorithm to predict RNA secondary structures with pseudoknots. *Nucleic Acids Res.* **39**, e93 (2011).
 40. Y. P. Xia, C. T. Yeh, J. H. Ou, M. M. Lai, Characterization of nuclear targeting signal of hepatitis delta antigen: nuclear transport as a protein complex. *J. Virol.* **66**, 914–921 (1992).
 41. L. H. Daigh, B. L. Griffin, A. Soroush, M. R. Mamedov, J. L. Casey, Arginine-rich motifs

are not required for HDV RNA binding activity of hepatitis delta antigen. *J. Virol.*, JVI-00929 (2013).

Appendix of Chapter 4

Supplementary Material for the manuscript submitted to the journal *Virus Evolution* as: Paraskevopoulou S., Käfer S., Zirkel F., Donath, A., Liu, S., Zhou, X., Drosten C., Misof B., and Junglen S. **Viromics of extant insect orders unveil the evolution of the flavi-like superfamily.** (unpublished).

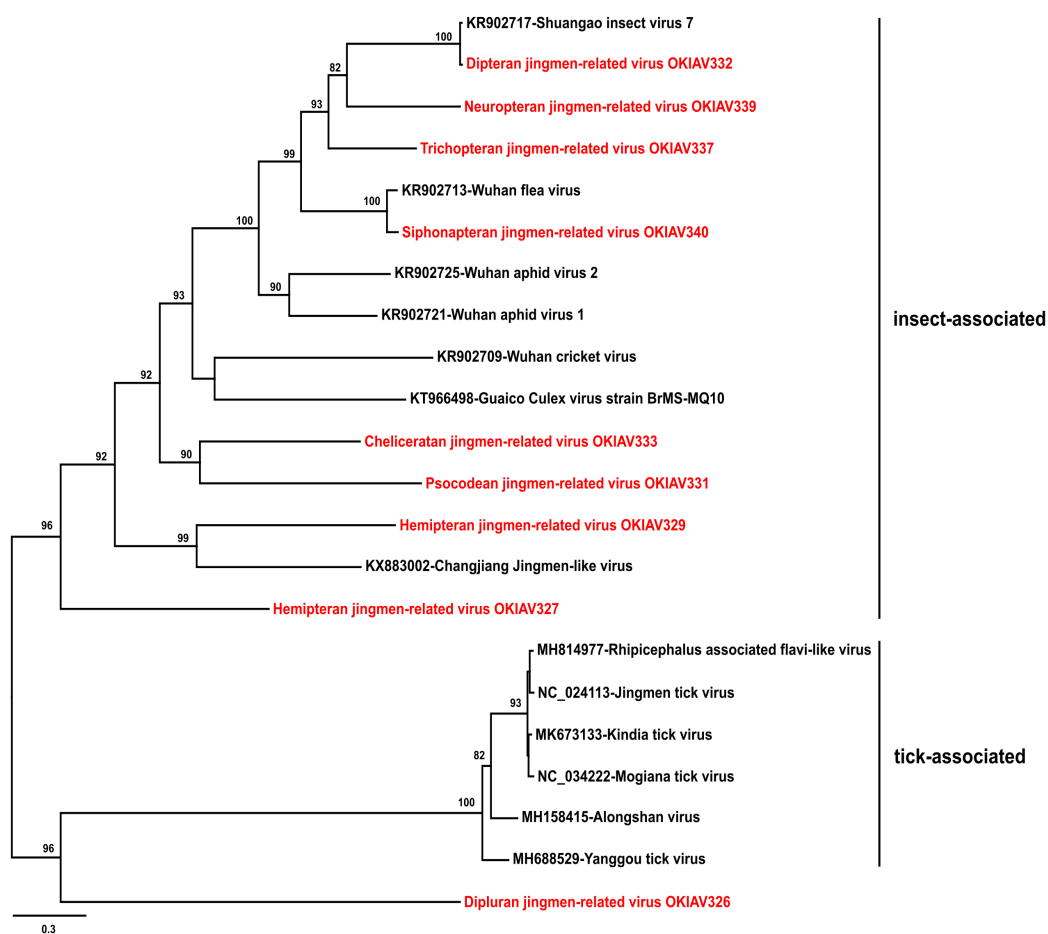


Fig. S1. Maximum likelihood phylogeny of jingmenviruses. The phylogenetic inference was based on an alignment of the RdRp region with 1,000 bootstrap replicates, using RAxML-NG version 0.7.0 BETA (52). Bootstrap values below 70 are not shown.

Legend of the following table:

Table S1. Additional data for the identified viral genomes. Information such as the GenBank accession numbers, host taxonomy, insect sample location, and collection date is provided.

Virus_name	GenBank_accnum	Host_subphylum	Host_class	Host_superorder	Host_order	Host_family	Host_species	Sample_location	Sample_date
Coleopteran flavi-related virus OKIAV323	MW208755	Hexapoda	Insecta	Endopterygota	Coleoptera	Cerambycidae	<i>Pempsamacra sp.</i>	Australia: New South Wales; 23 km northwest of Batemans Bay	10/13/2013
Hymenopteran flavi-related virus OKIAV347	MW208756	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Halicitidae	<i>Dufourea dentiventris</i>	Germany: Hessen; Gersfeld Gemarkung Hettenhausen - Trei Weyherer Bergfeld	7/26/2012
Neuropteran flavi-related virus OKIAV352	MW208757	Hexapoda	Insecta	Neuropterida	Neuroptera	Mantispidae	<i>Mantispinia sp.</i>	Venezuela: Cojedes Girardot Hato Pinero - Farm	10/26/2011
Dipteran flavi-related virus OKIAV357	MW314679	Hexapoda	Insecta	Panorpida	Diptera	Therevidae	<i>Anabarhynchus dentiphallus</i>	Australia: Australian Capital Territory; Cotter River	2011
Orthopteran flavi-related virus OKIAV358	MW208758	Hexapoda	Insecta	Orthopterida	Orthoptera	Pyrgomorphidae	<i>Stenocepa sp.</i>	Germany: Lab culture with samples originating from Germany; private breeder	2013
Embiopteran flavi-related virus OKIAV324	MW208759	Hexapoda	Insecta	NA	Embioptera	Clothodidae	<i>Antipaluria urichi</i>	USA: lab culture	11/08/2012
Plecopteran flavi-related virus OKIAV325	MW208760	Hexapoda	Insecta	Exopterygota	Plecoptera	Perlidae	<i>Perla marginata</i>	Germany: Baden-Wuerttemberg; Black Forest - Wutachschlucht	5/16/2011
Hymenopteran flavi-related virus OKIAV348	MW208761	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Apidae	<i>Thyreus orbatus</i>	Italy: Valle de Cogne; Lillaz	7/16/2011
Hymenopteran flavi-related virus OKIAV350	MW208762	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Vespidae	<i>Eumenes papillarius</i>	Germany: Baden-Wuerttemberg; Rheinau	6/23/2012
Hymenopteran flavi-related virus OKIAV351	MW208763	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Vespidae	<i>Ancistrocerus nigricornis</i>	Germany: Rhineland-Palatinate; Albersweiler	3/23/2012
Hymenopteran flavi-related virus OKIAV356	MW314680	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Perilampidae	<i>Perilampus aeneus</i>	Germany: Rhineland-Palatinate; S of Guntersblum	06/02/2011
Orthopteran flavi-related virus OKIAV360	MW208764	Hexapoda	Insecta	Orthopterida	Orthoptera	Pyrgomorphidae	<i>Dictyophorus griseus</i>	Germany: Lab culture with samples originating from Tanzania	05/01/2013
Orthopteran flavi-related virus OKIAV361	MW208765	Hexapoda	Insecta	Orthopterida	Orthoptera	Pyrgomorphidae	<i>Phymateus viridipes</i>	Germany: Lab culture with samples originating from South Africa	01/01/2013
Odonatan flavi-related virus OKIAV365	MW314681	Hexapoda	Insecta	Odonatoptera	Odonata	Libellulidae	<i>Orthetrum albistylum</i>	Greece: Thessaloniki; Regional Unit Lake Volvi	05/08/2012
Odonatan flavi-related virus OKIAV367	MW314682	Hexapoda	Insecta	Odonatoptera	Odonata	Hemiphlebidae	<i>Hemiphlebia mirabilis</i>	Australia: Victoria; Swamp south of Boreung Camp Grampians National Park	12/31/2013
Dipteran flavi-related virus OKIAV368	MW208766	Hexapoda	Insecta	Panorpida	Diptera	Sepsidae	<i>Meropterus fasciculatus</i>	Singapore: Lab culture with samples originating from Singapore	2013
Notopteran flavi-related virus OKIAV369	MW208767	Hexapoda	Insecta	Exopterygota	Notoptera	Austrophasmatidae	<i>Austrophasmatidae sp.</i>	South Africa: Western Cape; province West Coast District 5 km W of Vanrhynsdorp	8/13/2004
Dipteran flavi-related virus OKIAV1492	MW314683	Hexapoda	Insecta	Panorpida	Diptera	Simuliidae	<i>Simulium meridionale</i>	USA: Indiana; Tippecanoe County West Lafayette	5/22/2013
Siphonapteran jingmen-related virus OKIAV340	MW208795; MW208800	Hexapoda	Insecta	Endopterygota	Siphonaptera	Pulicidae	<i>Ctenocephalides felis</i>	USA: Labstock Kansas State University Manhattan Kansas - origin unclear	12/16/2012
Arachnidan jingmen-related virus OKIAV333	MW314684; MW314685 MW314686;	Chelicerata	Arachnida	NA	Scorpiones	Euscorpidae	<i>Euscorpius sicanus</i>	Italy: Sicily; Catania - Co. oak forest near Castanea 933 m	11/03/2012
Dipteran jingmen-related virus OKIAV332	MW314687; MW314688; MW314689	Hexapoda	Insecta	Panorpida	Diptera	Psychodidae	<i>Clogmia albipunctata</i>	USA: North Carolina; Wake County - Raleigh	06/01/2012
Hemipteran jingmen-related virus OKIAV329	MW208796; MW208802	Hexapoda	Insecta	Paraneoptera	Hemiptera	Cixiidae	<i>Tachycixius pilosus</i>	Germany: Thuringia; Jena	05/01/2012
Hemipteran jingmen-related virus OKIAV327	MW208797; MW208803 MW314690;	Hexapoda	Insecta	Paraneoptera	Hemiptera	Pleidae	<i>Plea minutissima</i>	Germany: Lower Saxony; Luechow-Dannenberg - Hoehbeck - Pevestorf	08/01/2011
Trichopteran jingmen-related virus OKIAV337	MW314691; MW314692; MW314693	Hexapoda	Insecta	Amphiesmenoptera	Trichoptera	Conoesucidae	<i>Costora delora</i>	Australia: Victoria; near Eskdale	4/14/2013
Psocopteran jingmen-related virus OKIAV331	MW208798; MW208804	Hexapoda	Insecta	Psocodea	Psocoptera	Pseudocaeiliidae	<i>Heterocaeilius solocipennis</i>	Japan: Hokkaido; Sapporo	6/30/2012
Dipluran jingmen-related virus OKIAV326	MW314694 MW208799;	Hexapoda	Entognatha	NA	Diplura	Campodeidae	<i>Campodea silvestrii</i>	Germany: North Rhine-Westphalia; Bonn Friesdorf Klufertbachtal in private garden	07/05/2012
Neuropteran jingmen-related virus OKIAV339	MW208801; MW208805; MW208806	Hexapoda	Insecta	Neuropterida	Neuroptera	Chrysopidae	<i>Pseudomallada ventralis</i>	Austria: Lower Austria; Vienna surroundings - Klosterneuburg	06/01/2012
Hymenopteran tomabus-related virus OKIAV371	MW208768	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Vespidae	<i>Katamenes arbustorum</i>	Italy: Valle de Cogne; Lillaz	7/16/2011
Coleopteran tomabus-related virus OKIAV372	MW314695	Hexapoda	Insecta	Endopterygota	Coleoptera	Coccinellidae	<i>Archeleis delta</i>	Australia: Queensland; Mount Tamborine	10/31/2013
Neuropteran tomabus-related virus OKIAV373	MW208769	Hexapoda	Insecta	Neuropterida	Neuroptera	Chrysopidae	<i>Chrysoperla carnea s.l.</i>	Austria: Lower Austria; Vienna surroundings - Klosterneuburg	5/29/2012
Dipteran tomabus-related virus OKIAV374	MW314696	Hexapoda	Insecta	Panorpida	Diptera	Hybotidae	<i>Stilpon pauciseta</i>	USA: North Carolina; Wake County - Raleigh - suburban yard	06/01/2013
Dipteran tomabus-related virus OKIAV375	MW314697	Hexapoda	Insecta	Panorpida	Diptera	Coelopidae	<i>Coelopa frigida</i>	Singapore: Lab culture with samples originating from Lund Skane Sweden	2013
Dipteran tomabus-related virus OKIAV376	MW208770	Hexapoda	Insecta	Panorpida	Diptera	Calliphoridae	<i>Calliphora vomitoria</i>	Australia: Australian Capital Territory; Canberra O'Connor	1/16/2012
Hymenopteran tomabus-related virus OKIAV377	MW208771	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Crabronidae	<i>Oxybelus bipunctatus</i>	Germany: Rhineland-Palatinate; Battenberg	07/05/2011
Hymenopteran tomabus-related virus OKIAV378	MW208772	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Crabronidae	<i>Oxybelus bipunctatus</i>	Germany: Rhineland-Palatinate; Battenberg	07/05/2011
Hymenopteran tomabus-related virus OKIAV379	MW208773	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Crabronidae	<i>Oxybelus bipunctatus</i>	Germany: Rhineland-Palatinate; Battenberg	07/05/2011
Odonatan tomabus-related virus OKIAV381	MW314698	Hexapoda	Insecta	Odonatoptera	Odonata	Platycnemididae	<i>Platycnemis pennipes</i>	Germany: Lower Saxony; Luechow-Dannenberg - Hoehbeck - Pevestorf	08/11/2011
Odonatan tomabus-related virus OKIAV382	MW314699	Hexapoda	Insecta	Odonatoptera	Odonata	Platycnemididae	<i>Platycnemis pennipes</i>	Germany: Lower Saxony; Luechow-Dannenberg - Hoehbeck - Pevestorf	08/11/2011
Hymenopteran tomabus-related virus OKIAV383	MW208774	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Tiphiidae	<i>Meria tripunctata</i>	France: Var Frejus	7/13/2011

Virus_name	GenBank_accnum	Host_subphylum	Host_class	Host_superorder	Host_order	Host_family	Host_species	Sample_location	Sample_date
Hymenopteran tombus-related virus OKIAV384	MW208775	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Tiphiidae	<i>Meria tripunctata</i>	France: Var Frejus	7/13/2011
Hymenopteran tombus-related virus OKIAV385	MW208776	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Crabronidae	<i>Oxybelus bipunctatus</i>	Germany: Rhineland-Palatinate; Battenberg	07/05/2011
Dipteran tombus-related virus OKIAV386	MW314700	Hexapoda	Insecta	Panorpida	Diptera	Acroceridae	<i>Pterodontia mellii</i>	Australia: Australian Capital Territory; Bonython	11/21/2012
Dipteran tombus-related virus OKIAV387	MW314701	Hexapoda	Insecta	Panorpida	Diptera	Nemestrinidae	<i>Trichophthalma ricardoae</i>	Australia: Australian Capital Territory; Canberra	11/01/2012
Dipteran tombus-related virus OKIAV388	MW314702	Hexapoda	Insecta	Panorpida	Diptera	Bombyliidae	<i>Comptosia brunnea</i>	Australia: Australian Capital Territory; Canberra O'Connor	01/12/2012
Zygentoman tombus-related virus OKIAV389	MW208777	Hexapoda	Insecta	NA	Zygentoma	Ateluridae	<i>Atelura formicaria</i>	Austria: Vienna	03/07/1905
Hymenopteran tombus-related virus OKIAV390	MW208778	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Sphecidae	<i>Podalonia hirsuta</i>	Germany: Rhineland-Palatinate; Niederzissen	5/18/2011
Hymenopteran tombus-related virus OKIAV391	MW208779	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Agaonidae/ Pteromalidae	<i>Lachaisea equicollis</i>	South Africa: Western Cape; Westlake	3/13/2013
Odonatan tombus-related virus OKIAV392	MW314703	Hexapoda	Insecta	Odonoptera	Odonata	Platycnemididae	<i>Platycnemis pennipes</i>	Germany: Lower Saxony; Luechow-Dannenberg - Hoehbeck - Pevestorf	08/11/2011
Coleopteran tombus-related virus OKIAV393	MW208780	Hexapoda	Insecta	Endopterygota	Coleoptera	Scarabaeidae	<i>Pentodon sp.</i>	South Africa: Western Cape; Koka Tsara Bush Camp near Karoo National Park	11/23/2013
Raphidiopteran tombus-related virus OKIAV394	MW208781	Hexapoda	Insecta	Endopterygota	Raphidioptera	Raphidiidae	<i>Phaeostigma divina divina</i>	Austria: Lab culture with samples originating from Greece Parnassos	5/20/2012
Raphidiopteran tombus-related virus OKIAV395	MW208782	Hexapoda	Insecta	Endopterygota	Raphidioptera	Raphidiidae	<i>Raphidia mediterranea</i>	Austria: Lab culture with samples originating from Greece; Achaia Kalavrita	05/01/2011
Coleopteran tombus-related virus OKIAV396	MW314704	Hexapoda	Insecta	Endopterygota	Coleoptera	Coccinellidae	<i>Chilocorus renipustulatus</i>	Germany: Thuringia; Jena - Tautenburger Forest	3/23/2011
Coleopteran tombus-related virus OKIAV397	MW208783	Hexapoda	Insecta	Endopterygota	Coleoptera	Curculionidae	<i>Oxoplatypus quadridentatus</i>	USA: Florida; Gainesville	10/01/2013
Megalopteran tombus-related virus OKIAV398	MW208784	Hexapoda	Insecta	Endopterygota	Megaloptera	Corydalidae	<i>Corydalidae sp.</i>	Venezuela: Carabobo Bejuma	11/09/2011
Dipteran tombus-related virus OKIAV400	MW314705	Hexapoda	Insecta	Panorpida	Diptera	Braulidae	<i>Braula coeca</i>	South Africa: Mpumalanga	3/24/2014
Dipteran tombus-related virus OKIAV401	MW314706	Hexapoda	Insecta	Panorpida	Diptera	Braulidae	<i>Braula coeca</i>	South Africa: Mpumalanga	3/24/2014
Dipteran tombus-related virus OKIAV402	MW314707	Hexapoda	Insecta	Panorpida	Diptera	Braulidae	<i>Braula coeca</i>	South Africa: Mpumalanga	3/24/2014
Dipteran tombus-related virus OKIAV403	MW314708	Hexapoda	Insecta	Panorpida	Diptera	Braulidae	<i>Braula coeca</i>	South Africa: Mpumalanga	3/24/2014
Dipteran tombus-related virus OKIAV404	MW314709	Hexapoda	Insecta	Panorpida	Diptera	Braulidae	<i>Braula coeca</i>	South Africa: Mpumalanga	3/24/2014
Dipteran tombus-related virus OKIAV405	MW314710	Hexapoda	Insecta	Panorpida	Diptera	Braulidae	<i>Braula coeca</i>	South Africa: Mpumalanga	3/24/2014
Dipteran tombus-related virus OKIAV407	MW314711	Hexapoda	Insecta	Panorpida	Diptera	Ceratopogonidae	<i>Clinohoele pseudonubifera</i>	USA: North Carolina; Wake County-Raleigh-North Carolina State University campus	5/30/2013
Dipteran tombus-related virus OKIAV408	MW208785	Hexapoda	Insecta	Panorpida	Diptera	Lonchopteridae	<i>Lonchoptera bifurcata</i>	USA: North Carolina; Wake County - Raleigh - Schenck forest	5/25/2013
Dipteran tombus-related virus OKIAV409	MW314712	Hexapoda	Insecta	Panorpida	Diptera	Bombyliidae	<i>Comptosia brunnea</i>	Australia: Australian Capital Territory; Canberra O'Connor	01/12/2012
Dipteran tombus-related virus OKIAV410	MW314713	Hexapoda	Insecta	Panorpida	Diptera	Bombyliidae	<i>Comptosia brunnea</i>	Australia: Australian Capital Territory; Canberra O'Connor	01/12/2012
Odonatan tombus-related virus OKIAV411	MW314714	Hexapoda	Insecta	Odonoptera	Odonata	Libellulidae	<i>Celithemis elisa</i>	USA: Tennessee; Monroe County - Cherokee National Forest	6/21/2011
Odonatan tombus-related virus OKIAV420	MW314715	Hexapoda	Insecta	Odonoptera	Odonata	Amphipterygidae	<i>Devadatta argyroides</i>	Singapore: Dairy Farm	7/23/2013
Hemipteran tombus-related virus OKIAV416	MW208786	Hexapoda	Insecta	Paraneoptera	Hemiptera	Miridae	<i>Notostira elongata</i>	Germany: North Rhine-Westphalia; Rhein-Sieg-Kreis Donrath Agger	5/17/2011
Hemipteran tombus-related virus OKIAV417	MW208787	Hexapoda	Insecta	Paraneoptera	Hemiptera	Tingidae	<i>Corythucha ciliata</i>	Germany: North Rhine-Westphalia; Bonn	04/10/2011
Phasmatodean tombus-related virus OKIAV418	MW208788	Hexapoda	Insecta	Exopterygota	Phasmatodea	Phylliidae	<i>Phyllium philippinicum</i>	Germany: lab culture	11/13/2012
Coleopteran tombus-related virus OKIAV419	MW208789	Hexapoda	Insecta	Endopterygota	Coleoptera	Curculionidae	<i>Ips typographus</i>	Germany: Rhineland-Palatinate; Cochem-Cell - Brohl	5/13/2012
Coleopteran tombus-related virus OKIAV413	MW208790	Hexapoda	Insecta	Endopterygota	Coleoptera	Curculionidae	<i>Ips typographus</i>	Germany: Rhineland-Palatinate; Cochem-Cell - Brohl	5/13/2012
Coleopteran tombus-related virus OKIAV424	MW208791	Hexapoda	Insecta	Endopterygota	Coleoptera	Cleridae	<i>Thanasimus formicarius</i>	Germany: Mecklenburg-Hither Pomerania; Mecklenburg Lake District Fuerstenhagen	05/07/2011
Zygentoman tombus-related virus OKIAV425	MW208792	Hexapoda	Insecta	NA	Zygentoma	Lepidotrichidae	<i>Tricholepidion gertschi</i>	USA: California; Mendocino County Angelo Coast Range Reserve	08/04/2011
Dipluran tombus-related virus OKIAV422	MW314716	Hexapoda	Entognatha	NA	Diplura	Campodeidae	<i>Campodea silvestrii</i>	Germany: North Rhine-Westphalia; Bonn Friesdorf Klufterbachtal in private garden	07/05/2012
Hymenopteran tombus-related virus OKIAV414	MW208794	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Vespidae	<i>Alastor atropos</i>	Germany: Baden-Wuerttemberg; Kaiserstuhl - Ihringen	7/25/2012
Hymenopteran tombus-related virus OKIAV415	MW208793	Hexapoda	Insecta	Hymenoptera	Hymenoptera	Scoliidae	<i>Scolia hirta</i>	Austria: Vienna; Arztgasse 73-garden	07/10/2012

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Christian Drosten for the opportunity to continue my academic development in virus research pursuing this doctoral degree. I am grateful for having had the chance to work closely with him in fascinating projects and receive valuable mentoring and critical feedback on the progress of my work. Crucially too, I owe many thanks to my second reviewer Prof. Dr. Dino McMahon for enabling the realization of this thesis at the Freie Universität Berlin.

I would also like to thank PD Dr. Sandra Junglen and Dr. Terry C. Jones for valuable discussions, teaching, and mentoring since the beginning of my time as a doctoral student.

Special thanks goes to Dr. Simon Käfer and MSc. Fabian Pirzer for excellent collaboration, mutual teaching, and lots of fun at work.

The realization of the insect virus discovery project wouldn't have been possible without the contributions from Prof. Dr. Bernhard Misof of the Zoological Research Institute Museum Koenig in Bonn and the people that have worked on the 1000 Insect Transcriptome Evolution Project, 1KITE. Similarly, Prof. Dr. Dieter Glebe of the Justus-Liebig-Universität in Gießen and Prof. Dr. Simone Sommer of the University of Ulm have crucially contributed in the rodent deltavirus project. Also, Dr. Florian Zirkel, MSc. Nora Goldmann, and MSc. Lina Theresa Gottula deserve significant thanks for excellent collaborations. I would also like to take the opportunity and thank the Berlin Institute of Health for providing the HPC research cluster facility.

Beloved thanks goes to my dear friends Dimitra, Savvas, Pavlos, Voula, Anna, and Giorgos for being supportive and being there in both great and difficult times.

Finally, I am extremely grateful to my parents, brother, and grandparents for doing their very best for me and for all their love and support throughout my life.

Declaration of authorship

I herewith declare, that I have written this thesis independently and myself. I have used no other sources than those listed. I have indicated all places where the exact words or analogous text were taken from sources. I assure that this thesis has not been submitted for examination elsewhere.

Sofia Paraskevopoulou

Berlin, 28.12.2020

List of publications

Parts of the data presented in this thesis are published in peer-reviewed journals as the following papers (#authors contributed equally):

1. Re-assessing the diversity of negative strand RNA viruses in insects

Käfer, S.[#], **Paraskevopoulou, S.[#]**, Zirkel, F., Wieseke, N., Donath, A., Petersen, M., Jones, T.C., Liu, S., Zhou, X., Middendorf, M., Junglen, S., Misof, B., and Drosten C. (2019).

PLoS Pathogens, 15(12):e1008224, doi:10.1371/journal.ppat.1008224.

Author contributions: Conceptualization: S.K., S.P., A.D., S.J., B.M., and C.D.; Data curation: S.K., S.P., F.Z., S.L., and X.Z.; Formal analysis: S.K., S.P., F.Z., N.W., M.M., and S.J.; Funding acquisition: C.D.; Investigation: S.K., S.P., F.Z., S.J., and C.D.; Methodology: S.K., S.P., F.Z., N.W., A.D., S.L., X.Z., M.M., and S.J.; Project administration: C.D.; Resources: A.D., T.C.J., and M.M.; Software: S.K., S.P., F.Z., N.W., A.D., M.P., T.C.J., and M.M.; Supervision: S.J., B.M., and C.D.; Validation: S.K. and S.P.; Visualization: S.K., S.P., F.Z., and N.W.; Writing – original draft: S.K., S.P., S.J., B.M., and C.D.; Writing – review & editing: S.K., S.P., T.C.J., B.M., and C.D..

2. Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent

Proechimys semispinosus

Paraskevopoulou S.[#], Pirzer F.[#], Goldmann N.[#], Schmid J., Corman V.M., Gottula L.T., Schroeder S., Rasche A., Muth D., Drexler J.F., Heni A.C., Eibner G.J., Page R.A., Jones T.C., Müller M.A., Sommer S., Glebe D., and Drosten C. (2020).

Proceedings of the National Academy of Sciences, 117(30):17977–17983, doi:10.1073/pnas.2006750117.

Author contributions: S. Sommer and C.D. designed research; S.P., F.P., N.G., J.S., V.M.C., L.T.G., S. Schroeder, A.R., D.M., J.F.D., A.C.H., G.J.E., R.A.P., T.C.J., M.A.M., S. Sommer, D.G., and C.D. performed research; J.S., A.C.H., G.J.E., R.A.P., and S. Sommer designed and performed fieldwork; S.P., F.P., N.G., L.T.G., and S. Schroeder contributed new reagents/analytic tools; S.P., F.P., N.G., J.S., V.M.C., L.T.G., S. Schroeder, A.R., D.M., J.F.D., A.C.H., G.J.E., R.A.P., T.C.J., M.A.M., S. Sommer, D.G., and C.D. analyzed data; S.P., D.G., and C.D. wrote the paper.

3. Viromics of extant insect orders unveil the evolution of the flavi-like superfamily

Paraskevopoulou S., Käfer S., Zirkel F., Donath, A., Liu, S., Zhou, X., Drosten C., Misof B., and Junglen S. (unpublished). Submitted to the journal *Virus Evolution*.

Parts of the data presented in this thesis are included in the following ICTV taxonomic reports:

- **2020.020M** - Create four new species in the genus *Orthophasmavirus*, create two new species in the genus *Feravirus*, and create one new genus (*Hymovirus*) including two new species (*Bunyavirales: Phasmaviridae*).
Ballinger M.J., Hall R.A., Langevin S.A., Pauvolid-Correa A., **Paraskevopoulou S.**, Drosten C., & Junglen S. (2020).
- **2020.023M** - Create seven new genera (*Alphacrustrhavirus*, *Alphadrosrhavirus*, *Alphahymrhavirus*, *Betahymrhavirus*, *Betanemrhavirus*, *Betapaprhavirus*, and *Betaricinrhavirus*), including 16 new species (*Mononegavirales: Rhabdoviridae*).
Walker P., Blasdell K.R., Dietzgen R.G., Freitas-Astúa J., Kondo H., Kurath G., Kuzmin I.V., Tesh R.B., Tordo N., Vasilakis N., & Whitfield A.E. (2020).
- **2020.024M** - Create one new species in the genus *Nyavirus* and four new species in one new genus *Formivirus* (*Mononegavirales: Nyamiviridae*).
Dietzgen R.G., Kondo H., Kuhn J.H., Vasilakis N., Jiang D., & Junglen S. (2020).
- **2020.026M** - Reorganize the order to include four new families, 18 new genera, and 22 new species (*Jingchuvirales*).
Di Paola N., Dheilly N.M., Kuhn J.H., Junglen S., **Paraskevopoulou S.**, Postler T.S., & Shi M. (2020).
- **2020.027M** - Create four new genera and 30 new species (*Bunyavirales: Nairoviridae*).
Marklewitz M., **Paraskevopoulou S.**, Alkhovskiy S.V., Avsic-Zupanc T., Bente D.A., Bergeron E., Burt F.J., Ergunay K., Garrison A.R., Hewson R., Mirazimi A., Palacios G., Papa A., Paweska J.T., Sall A.A., Sprengler J.R., Di Paola N., & Kuhn J.H. (2020).
- **2020.029M** - Create one new genus and 16 new species (*Bunyavirales: Phenuiviridae*).
Marklewitz M., **Paraskevopoulou S.**, Briese T., Charrel R., Choi I.-R., de Lamballerie X., Ebihara H., Fu Gao G., Groschup M.H., Johnson G., Nunes M., Palacios G., Sasaya T., Shirako Y., Song J.-W., Wei T., Zerbini F.M., Zhou X., & Kuhn J.H. (2020).

-
- **2020.012D** - Create one new realm (*Ribozyviria*) including one new family (*Kolmioviridae*) including genus *Deltavirus* and seven new genera for a total of 15 species.

Hepojoki J., Hetzel U., **Paraskevopoulou S.**, Drosten C., Harrach B., Zerbini M., Koonin E.V., Krupovic M., Dolja V., & Kuhn J.H. (2020).