Evaluation of the zoonotic potential of tick-borne viruses in public datasets

Yuting Lin[1,2] and David J Pascall[1]

1. MRC Biostatistics Unit, University of Cambridge, Cambridge
2. Royal Veterinary College, University of London, London

Abstract

Tick-borne viruses remain a substantial zoonotic risk worldwide. The advent of next-generation sequencing has significantly broadened our knowledge of viruses. However, large amounts of sequences in public datasets remain unannotated which could represent unrealised or undocumented viruses. This study aimed to assess whether these novel and other understudied viruses likely represent zoonotic risks and thus deserve further investigation. We ranked the human infection potential of 136 known tick-borne viruses using a genome composition-based machine learning model. We found five high-risk tick-borne viruses (Langat virus, Lonestar tick chuvirus 1, Grotenhout virus, Taggert virus, and Johnston Atoll virus) that have not been known to infect human and two viral families (*Nairoviridae* and *Phenuiviridae*) that contain a large proportion of potential emerging tick-borne viruses. Through bioinformatic analyses, we presented 83 unannotated contigs exhibiting high identity with known tick viruses in public meta-genomic and -transcriptomic datasets. These putative novel viral contigs were classified into three RNA viral families (*Alphatetraviridae*, *Orthomyxoviridae*, *Chuviridae*) and one DNA viral family (*Asfaviridae*). After manual checking, these 83 contigs were reduced to 5 putative novel Alphatetra-like viral contigs, 5 putative novel Orthomyxo-like viral contigs, and 1 Chu-like viral contig which clustered with known tick-borne viruses, forming a separate clade within the viral families. This study added to the knowledge of tick virus diversity and highlighted the importance of surveillance of newly emerging tick-borne diseases.

Importance

Ticks are important hosts of pathogens that cause diseases worldwide. Despite this, numerous tick-borne viruses are still unknown or poorly characterised. To overcome this, we re-examined currently known tick-borne viruses and identified putative novel viruses associated with ticks in public datasets. The study highlighted five high-risk tick-borne viruses that currently have no reports on human infection, and also put emphasis on two viral families which may pose potential public health threats. From meta-genomic and –transcriptomic datasets, we described 11 putative novel RNA viral contigs clustered with known tick-borne viruses and suggested novel viral contigs in the family *Orthomyxoviridae* and *Chuviridae* may cause public health concern. Overall, the study increased our knowledge of tick-borne viruses and demonstrated the power of metagenomic and –transcriptomic studies in exploring the diversity of viruses. The study also suggested that monitoring ticks and the viruses carried by them in the tick-endemic areas is still an important public health question.

Introduction

The role of ticks in the transmission of viruses has been known for over 100 years (1) with ticks being second only to mosquitoes as vectors of pathogens to humans and the primary vector of pathogens to livestock, wildlife, and companion animals (2). Nonetheless, new human pathogenic tick viruses are still being detected (2). Two recent examples are Yezo virus identified in Japan (3) and Songling virus identified in China (4), which have both been associated with acute human febrile disease.

Prior to the advent of molecular biology, tick-borne viruses were identified through microscopy, isolation in culture or serology (5-10). Over the last two decades, metagenomics and

metatranscriptomics have revolutionised virus discovery. These approaches employ high-throughput sequencing technologies to capture total viral communities in a relatively unbiased manner (11,12) and have facilitated studies of the tick virome that resulted in the discovery and characterisation of hundreds of new tick-borne viruses globally (13-22).

While detecting novel viruses is valuable from a basic ecology perspective, to make these studies useful for public health we need to be able to evaluate the human infection potential of the discovered viruses. Models of zoonotic risk assessment often rely on the existence of zoonotic viruses in the same viral family or evolutionary relatedness to known zoonoses (23). These models rely on phenotypic information commonly unavailable for novel viruses or of insufficient resolution to estimate different levels of zoonotic potential for closely related viruses (24-26). More recent models based on the genomic features of viruses which do not strictly rely on phylogeny or taxonomy avoid this issue. Such models use genomic features that may individually contain weak signals but can be combined and exploited via machine learning algorithms (27-30). These models, built from genomic data alone, have particular strength in evaluating newly discovered viruses that lack other information sources, but are limited by the need for full viral genomes for accurate assessment.

In this study we aimed detect novel viruses that may have been overlooked by previous tick meta-genomic and –transcriptomic studies, and assess the zoonotic risk posed by both these newly described viruses and other poorly studied tick viruses. Thus, we searched publicly available tick sequence data for signs of viruses misidentified as tick sequences and evaluated the zoonotic risk of these and other tick viruses to find those which deserve further investigation.

Results and discussion

Identification of putative novel viruses associated with ticks

We examined contigs that showed similarity (< 90% identity) to previously characterised tick virus sequences from an extensive collection of tick meta-genomic and –transcriptomic datasets (23,050 SRA, 3523 TSA, and 11,279 WGS sequences; collected on 3 March 2022). The pipeline of detecting novel tick-associated viral contigs can be found in **Figure 1**. For more information on the tick virus database and tick meta-genomic and –transcriptomic datasets, see Methods.

In total, we identified 83 putative novel viral contigs, including 53 double-stranded DNA viral contigs, 5 positive-sense and 25 negative-sense RNA viral contigs (**Table S1-3** in the supplementary material). The 53 DNA viral contigs belong to the family *Asfaviridae*, with homology to African swine fever virus. These African swine fever virus-like contigs were found in *Ornithodoros porcinus* and *Ornithodoros moubata* ticks (**Table S3**), and all but two of these contigs derived from a study that had previously reported the integration of African swine fever like elements into the genomes of these soft tick species (31), so they will not be discussed further. The 5 positive-sense RNA viral contigs belong to the family *Alphatetraviridae*. Of the negative-sense RNA viral contigs, 5 belong to the family *Orthomyxoviridae*, and 20 belong to the family *Chuviridae*, including 19 chuviruses believed to be replication defective from tick SRA projects. 11 putative RNA viral contigs ranging from 728 bp to 5496 bp were selected for further analysis based on quality and dissimilarity toward other sequences in the dataset (**Table 1**). The putative novel viruses found in this study were named after their host order and related viral family, followed by a number (e.g., Alphatetra-like tick virus 1). All sequences identified in the present study with can be found on GenBank (see accession numbers in **Table 1** and **Table S1-3**).

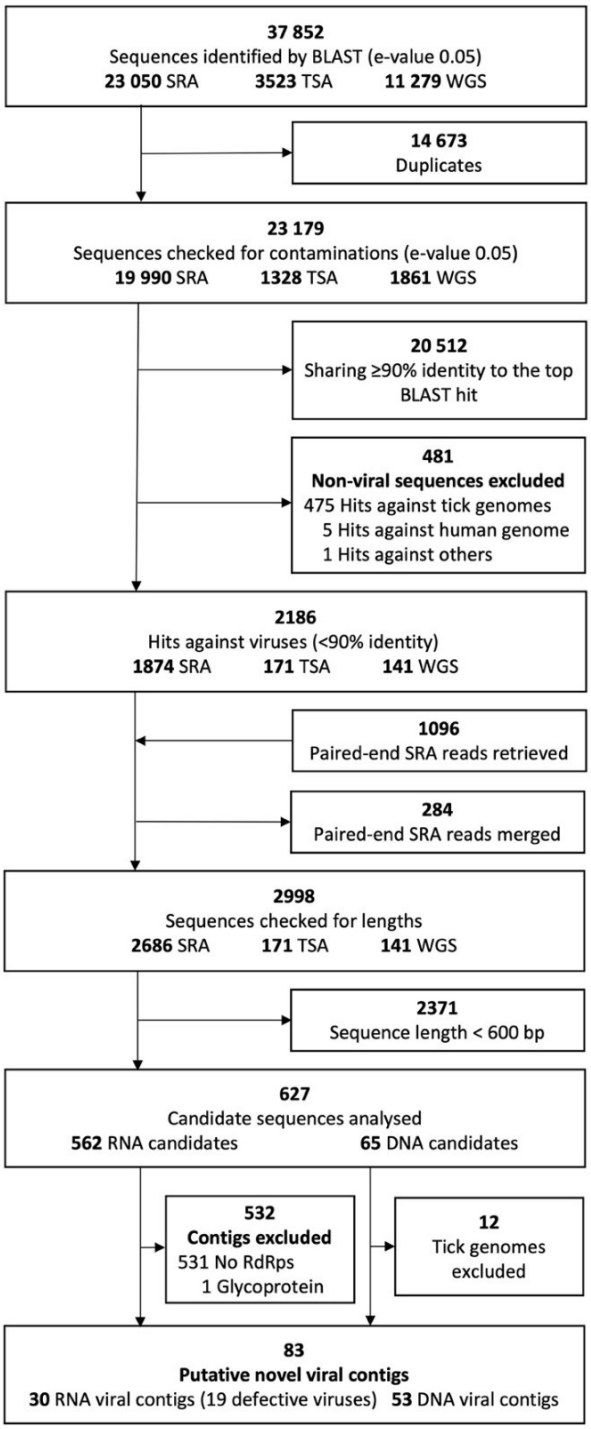

Figure 1 The workflow of identification of novel tick-associated viral contigs. Through BLAST search and curation, 53 DNA and 30 RNA sequences from a collection of tick meta-genomic and –transcriptomic datasets (3 March 2022) were identified as putative novel viral contigs. TSA = Transcriptome Shotgun Assembly, SRA = Sequence Read Archive, WGS = Whole Genome Shotgun, E-value = Expectation value.

Table 1 The putative novel RNA viral contigs found after bioinformatic checking.

Putative novel viral contig	Tick species	Geographical location	Provisional family	Genome structure	Sequence length (nt)	Closest BLASTx match (% amino acid identity)	Genbank ID
Orthomyxo-like tick virus 1	*Rhipicephalus sanguineus*	Italy	Orthomyxoviridae	ssRNA(-)	2136	Zambezi tick virus 1 (89)	HACP01027211.1 (HACW01024387.1)
Orthomyxo-like tick virus 2	*Rhipicephalus sanguineus*	Italy	Orthomyxoviridae	ssRNA(-)	1290	Zambezi tick virus 1 (73)	HACP01022575.1
Orthomyxo-like tick virus 3	*Rhipicephalus sanguineus*	Italy	Orthomyxoviridae	ssRNA(-)	1290	Zambezi tick virus 1 (88)	HACW01018819.1
Orthomyxo-like tick virus 4	*Rhipicephalus haemaphysaloides*	China	Orthomyxoviridae	ssRNA(-)	2357	Zambezi tick virus 1 (89)	GIJA01018702.1
Alphatetra-like tick virus 1	*Ixodes ricinus*	China	Alphatetraviridae	ssRNA(+)	5496	Bulatov virus (44)	GIXL01013119.1
Alphatetra-like tick virus 2	*Ixodes persulcatus*	USA	Alphatetraviridae	ssRNA(+)	1056	Vovk virus (60)	GBXQ01012426.1
Alphatetra-like tick virus 3	*Ornithodoros moubata*	Spain	Alphatetraviridae	ssRNA(+)	5452	Bulatov virus (40)	GIXP02005464.1
Alphatetra-like tick virus 4	*Ornithodoros moubata*	Spain	Alphatetraviridae	ssRNA(+)	3873	Bulatov virus (41)	GFJQ02007921.1
Alphatetra-like tick virus 5	*Rhipicephalus microplus*	USA	Alphatetraviridae	ssRNA(+)	1488	Hepelivirales sp. (80)	GEMR01004364.1
Chu-like tick virus 1	*Ixodes scapularis*	USA	Chuviridae	ssRNA(-)	728	Nuomin virus (89)	GGIX01201876.1

(i) Orthomyxoviridae-like contigs

The contigs we found contained homologs of Influenza RdRp subunit PB1, the most conserved of the orthomyxovirus genes (**Figure 2A**). Phylogenetic analysis revealed that the putative novel Orthomyxo-like viral contigs fell into the genus *Quaranjavirus* and formed a separate clade with other tick-borne quaranjaviruses (**Figure 3A**; **Table 1**). Within PB1, the contigs were phylogenetically closely related to Zambezi tick virus 1 (ZaTV-1) – a highly divergent virus identified in Rhipicephalus ticks from Mozambique (32) and shared approximately 77% amino-acids identity (**Figure 3A**; **Table 1**). Quaranjaviruses are enveloped negative-sense single-stranded segmented RNA viruses, generally with multiple segments (33). Quaranjaviruses have been known to be transmitted to birds mainly by argasid ticks, but a recent study found a quaranjavirus-like sequence in *Am. dissimile* in Romania (34). In the present study, we identified 4 putative novel Orthomyxo-like tick viruses from ixodid tick metagenomic sequences, 3 of which were sampled from *R. sanguineus* in Italy, and one from *R. haemaphysaloides* in China (**Table 1**). These findings added evidence to the quaranjaviruses diversity in ixodid ticks.

We note that four of the six the viral species that have complete genomes in this clade were either predicted to have very high (N = 1) or high (N = 3) zoonotic potential based on the GCB model (**Table 2**; the GenBank ID can be found in **S4 Table** in the supplementary material), with Quaranfil virus (QRFV), one of the viruses rated to have high zoonotic potential, being confirmed to lead to infections in humans (33). Taken together, these results suggest that the novel Orthomyxo-like viral contigs identified in the current study may represent a potential risk of human infection. However,

without direct sources of evidence, caution must be applied when predicting the zoonotic risks. We suggest fundamental studies including laboratory testing and sequencing of whole genomes of the putative novel Orthomyxo-like viruses would provide more evidence for public health surveillance.

(ii) Alphatetraviridae-like contigs

Members of the *Alphatetraviridae* are non-enveloped positive-sense single-stranded RNA viruses (35). The identified Alphatetra-like contigs ranged from 1488 nt to 5496 nt, detected via the presence of a putative RdRp 2 motif; some of the longer sequences also encoded putative nucleoside triphosphate hydrolase, viral RNA helicase, and viral methyltransferase three motifs (**Figure 2B**). Viruses in the family *Alphatetraviridae* had been previously believed to infect predominantly moths and butterflies (35), but the recent discovery of several Alphatetra-like viruses in ticks (22,36) implies that this may have been due to sampling biases. Consistent with this, we found the presence of potential Alphatetra-like viruses in four ixodid tick species (*I. ricinus*, *I. persulcatus*, *O. moubata* and *R. microplus*) sampled in three geographically diverse locations (China, Spain, and USA) (**Table 1**). Phylogenetic analysis suggested that the novel Alphatetra-like tick viruses formed a separate clade, being placed with the tick-borne viruses recently identified in ixodid ticks from Antarctica (Vovk virus and Bulatov virus from the seabird tick, *I. uriae*) (36) and China (Hubei tick hepe-like virus found in *R. microplus* and *H. longicornis*) (22), and Heilongjiang sediment betatetravirus originally sampled from river sediment in China (37) (**Figure 3B**; **Table 1**). These results indicated that the viruses detected in this study, along with other tick-borne alphatetraviruses and Heilongjiang sediment betatetravirus, probably form a new viral genus in the *Alphatetraviridae* family.

No members of the *Alphatetraviridae* are known to infect humans, and, unfortunately, only partial genomes were available for the previously identified viruses in the proposed new clade (Bulatov virus, Vovk virus and Heilongjiang sediment betatetravirus), thus they were inappropriate for zoonotic risk analysis based on the GCB model. The closest related viruses with full genomes were all isolated from moths. Of these viruses Nudaurelia capensis beta virus was predicted to have high zoonotic potential; Dendrolimus punctatus virus and Helicoverpa armigera stunt virus were predicted to have medium zoonotic potential (**Table 2**; the GenBank IDs can be found in **S4 Table** in the supplementary material). This result illustrates a deficiency of the currently available zoonotic risk assessment tools. For viruses discovered through meta-genomic or –transcriptomic methods, zoonotic risk estimation using genome information is challenging due to the incompleteness of viral genomes or uncertainty about segment linkage for segmented viruses, which impacts the ability to make fast decisions on virus research and surveillance at the earliest stage of virus discovery. Training existing models on features from partial viral genomes provides a potential avenue for future work to help alleviate this issue.

(iii) Chuviridae-like contigs

We identified 20 Chu-like viral contigs from *I. scapularis* and *Hy. asiaticum* sampled in the US and China (**Table 1**; **Table S2**). Members of the *Chuviridae* are negative-sense single-stranded RNA viruses, which display a wide variety of genome organisations, including unsegmented, bi-segmented, and circular forms (14). Only the contig labelled Chu-like tick virus 1 is consistent with a live virus, as all other contigs showed signs of mutational degradation with many stop codons present in the region with homology to the RdRp. All but one of the contigs exhibiting evidence of degradation were generated from a laboratory population of *Hy. asiaticum*, maintained under controlled conditions (38). As such, we speculate that they likely result from the presence of an

integrated Chu-like virus in the *Hy. asiaticum* genome present in this laboratory population, and potentially more broadly.

The viral sequence of Chu-like tick virus 1 was 728 nt in length, comprising one putative ORF that encodes a 240 aa protein containing the potential RdRp domain, nucleoside triphosphate hydrolase, viral RNA helicase, and viral methyltransferase 3 (**Figure 2C**). Chu-like tick virus 1 was clustered with other miviruses and was phylogenetically closely related to Suffolk virus, originally sampled in USA (19), with approximately 87% nucleotide identity (**Figure C**; **Table 1**). With regards to the human infection potential of Chu-like tick virus 1, it is notable that it showed a RdRp amino-acid identity of around 90% with Nuomin virus which has recently been associated with human febrile illness in China (39). The prediction results also classified Suffolk virus into the high zoonotic potential category, suggesting a zoonotic risk score of 0.309 (95% CI: 0.151-0.562) (**Table 2**; the GenBank IDs can be found in **S4 Table** in the supplementary material). These results suggest that Chu-like tick virus 1 may represent a zoonotic risk. However, given the small number of known species in the family *Chuviridae* when the GCB model was trained (30), caution must be applied, as the model might not be able to accurately predict the zoonotic potential viruses in this family. There would therefore seem to be a definite need for characterising viruses in the family *Chuviridae* in terms of their viability in human cell lines with these results feeding through to future zoonotic modelling efforts.

Table 2 Zoonotic risks of virus species closely related to the putative novel Alphatetra-like viral contigs. Zoonotic risks of sequences in GenBank format were predicted using the GCB model.

Putative novel viral contigs	Relative viral species	Current status	Predicted probability	95% Confidence interval	Zoonotic potential
Orthomyxo-like viral contigs	*Wellfleet Bay virus*	No known human infections	0.368	0.213-0.636	High
	Araguari virus	No known human infections	0.222	0.102-0.372	Medium
	Johnston Atoll quaranjavirus	No known human infections	0.513	0.311-0.802	Very high
	Lake Chad virus	No known human infections	0.184	0.081-0.334	Medium
	Tjuloc virus	No known human infections	0.352	0.141-0.607	High
	Quaranfil quaranjavirus	Human virus	0.561	0.128-0.863	High
Alphatetra-like viral contigs	*Nudaurelia capensis beta virus*	No known human infections	0.302	0.153-0.497	High
	Dendrolimus punctatus virus	No known human infections	0.234	0.128-0.352	Medium
	Helicoverpa armigera stunt virus	No known human infections	0.243	0.135-0.387	Medium
Chu-like viral contigs	*Mivirus suffolkense*	No known human infections	0.309	0.151-0.562	High

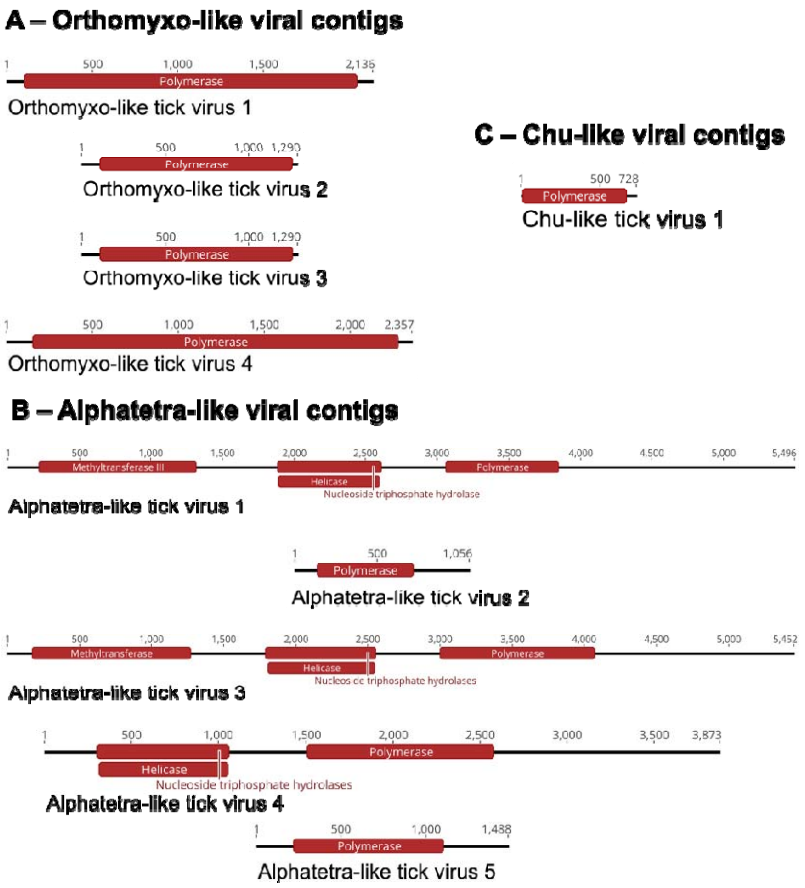

Figure 2 Genome structures of novel viral contigs.

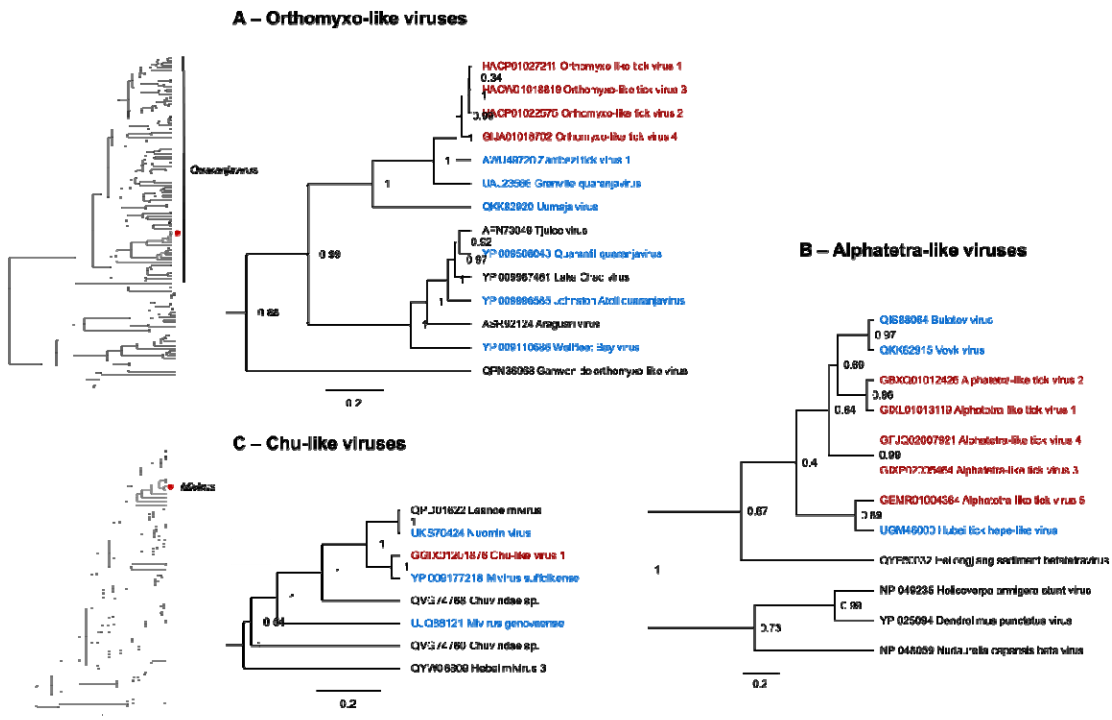

Figure 3 Phylogenetic relationship of the novel viral contigs. Maximum clade credibility trees were inferred from the Bayesian posterior sample. Posterior clade probabilities are shown at the nodes, and the scale is given in amino-acid substitutions per site. Putative novel viral contigs found in the present study are shown in red, and previously known tick-borne viruses are shown in blue. GenBank ID can be found at Figure S2-4 in the supplementary material. Alignments using ClustalW and maximum clade credibility trees are provided in at https://github.com/ytlin2021/TickVirus.git.

The findings of putative novel viruses from the current study demonstrated again the power of meta-genomic and –transcriptomic studies in exploring the diversity of viruses, but also reminded us of the challenges and limitations that require careful analysis workflows to ensure reliable results. The short lengths of viral contigs found through meta-genomic and –transcriptomic datasets limited the possible scope of our downstream analyses, including protein function prediction and zoonotic risk assessment. The prediction of zoonotic potential would have been more convincing if whole genome sequences had been obtained through (RT-)PCR and Sanger sequencing confirmation, combined with further confirmation tests such as sequencing virus-derived small RNAs or infectivity experiments on tick-derived cell lines. In addition, it is unavoidable that some putative novel tick-borne viruses are not detected, as no similar viruses have been identified, so there was no reference to search against.

Zoonotic ranking of known tick-borne viruses

We used the prediction framework illustrated by Mollentze and colleagues (30) to rank 136 known tick-borne viruses with complete genomes (**Table S4** in the supplementary material). This subset dataset contained representatives from 13 viral families (29 genera) and unassigned viruses, including 21 viruses that were known to infect humans by our criteria.

In total, 5.1% of tick-borne viruses were identified as having very high zoonotic potential (N = 7), 39.7% having high zoonotic potential (N = 54), 44.1% having medium zoonotic potential (N = 60) and

11.0% having low zoonotic potential (N = 15; **Figure 4A**). Among the 21 currently known tick-borne viruses, 85.7% were correctly identified as having either very high (N = 2) or high zoonotic potential (N = 16), and the remaining human-associated viruses were classified as medium zoonotic potential (N = 3; **Figure 4A**). Among the tick-borne viruses with unknown human infectivity that were sequenced from nonhuman animal or tick samples, 37.4% were predicted to have either very high (N = 5) or high zoonotic potential (N = 38; **Figure 4A**), including viruses in the family *Chuviridae*, *Nyamiviridae* and *Parvoviridae* that currently do not contain tick-borne viruses known to infect humans. Isfahan virus and Uukuniemi virus, which have previously been confirmed to infect humans, were the top 2 potential zoonotic viruses by mean probability. We also identified 5 tick viruses not previously known to infect humans which were predicted to have very high zoonotic risk (Langat virus, Lonestar tick chuvirus 1, Grotenhout virus, Taggert virus, and Johnston Atoll virus) (**Figure 4B**). We therefore recommend these viruses be a high priority for further virological research, in order to assess whether greater surveillance of them in wild tick populations would be warranted.

We noted that the high and very high zoonotic potential categories were dominated by *Nairoviridae* (31.0%) and *Phenuiviridae* (28.6%), consistent with the recent findings of emerging human-infecting tick viruses in these two viral families. For example, among members of the *Nairoviridae* family, two emerging orthonairoviruses called Songling virus (SGLV) and Yezo virus (YEZV) have been recently detected in *Hyalomma* spp. ticks and isolated in patients with the acute febrile disease (3,4). Hospitals in Inner Mongolia and Heilongjiang province, China also reported infections from a novel nairovirus named Beiji nairovirus (BJNV) which mainly circulates in the ixodid ticks *I. crenulatus* and *I. persulcatus* (40). Similarly, among the *Phenuiviridae*, a novel phlebovirus named Tacheng tick virus 2 (TcTV-2) was first sampled from *D. marginatus* ticks in China (14) and later identified in various countries and tick species, primarily circulating in *D. marginatus* and *H. marginatum* ticks (14,41-43). The risk for human infection from TcTV-2 was not known until 2021 when Dong and colleagues reported on TcTV-2 infection in a patient in China (44). As recently emerged viruses, the geographic ranges of these viruses are poorly understood, and only partial genomes are available. The wide distribution of these tick species suggests that the geographic limits of the emerging tick-borne viruses in the family *Nairoviridae* and *Phenuiviridae* may be larger than presently assumed.

The genome composition-based machine learning model also had strength in distinguishing the zoonotic potential of closely related viruses in a genus. For example, among miviruses in the family *Chuviridae*, Mivirus boleense was ranked in the low zoonotic potential category, while Mivirus suffolkense was considered high zoonotic potential. The prediction results demonstrated again the value of evaluating zoonotic potential based on genome composition, which can prioritise risks of viruses in a quick and low-cost way. Integrating this tool with other information from phylogenetic relationships, epidemiological studies, and further infection experiments, we can provide more informed suggestions on the surveillance of emerging or re-emerging tick-borne viruses.

From the results, we also found that a significant fraction of tick-borne viruses with high or very high zoonotic potential were sampled from ixodid ticks, which can potentially be explained by a large sampling bias towards viruses infecting hard ticks (family *Ixodidae*) in previous studies. Systematic studies with sampling of both hard and soft ticks would increase the knowledge of the diversity of tick-borne viruses and provide a basis for further studies about the interrelationship among human, ticks, and tick-borne viruses.

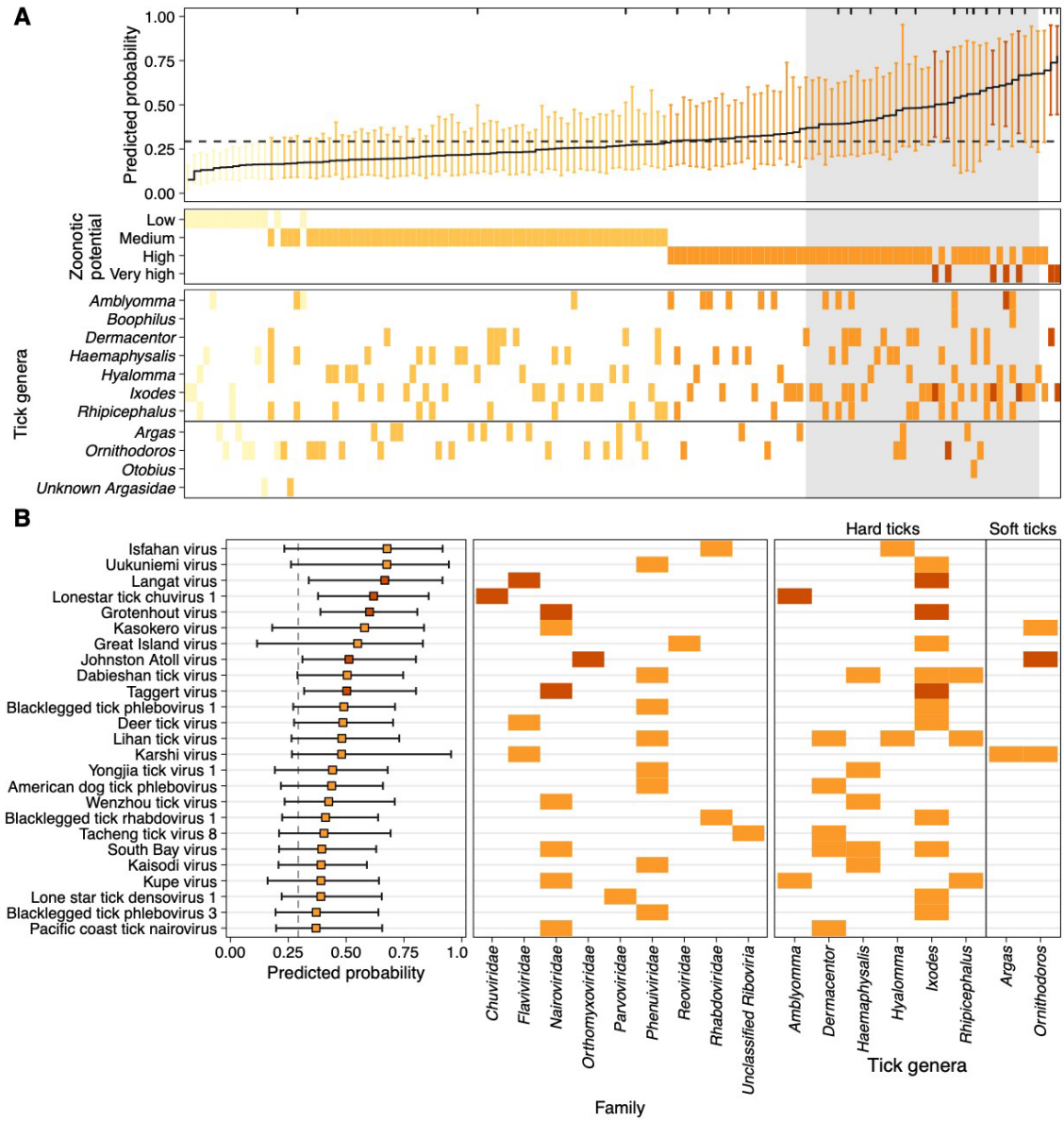

Figure 4 Predicted probability of human infection for tick viruses based on genome composition-based model. (**A**) Predicted probability of human infection for 136 tick viruses with complete genomes. Colour scale show the assigned zoonotic potential categories. Tick marks along the top edge of the first panel show the location of viruses known to infect humans, while a dashed line shows the cut-off 0.293 that balanced sensitivity and specificity according to Mollentze et al. (30). The top 25 viruses that have no known human infection (contained within the grey box) are illustrated in more detail in (**B**). Viruses that have no known human infection are shown in red. Points show the mean calibrated score, with lines indicating 95% confidence intervals. Figures were drawn in R v4.1.1. Numeric data underlying this figure can be found at https://github.com/ytlin2021/TickVirus.git.

Conclusion

While it is clear that ticks are important vectors of pathogens, the human infection risks of most tick-borne viruses remain uncertain (45). With the putative novel tick-borne viruses identified in this study, this study has investigated the human infection risk of undescribed tick-borne viruses using the genome composition-based machine learning model. The results support prior empirical evidence of tick-borne viruses and provide guidance for public health surveillance of high-risk species. However, our limited understanding of tick virus diversity and host species would likely impact the accuracy of classification, protein function annotation and zoonotic risk prediction of emerging tick-borne viruses. In conclusion, viruses present in ticks remain an important zoonotic risk worldwide. Apart from those high-risk viruses, potential emerging tick-borne viruses deserve further investigation.

Methods

Datasets. We collected tick virus data from published reports by 2021 and constructed a tick virus dataset that contained 261 tick-borne viruses from 23 viral families (34 genera) and 12 putative tick-borne viruses sampled from tick pools. Tick-borne viruses with unknown tick vectors (N = 12) or sequences unavailable (N = 14) were excluded from analyses. This resulted in 235 viruses in the final tick virus database, 3 of them were added later and not included in the original tick virus database used for virus detection. Among 235 tick-borne viruses, more than 92% of viruses were RNA viruses, while few other viral sequences were identified (2.2% DNA viruses and 5.6% unknown viruses). Among the RNA viruses, most sequences were assigned to negative-sense (-) single-stranded RNA (ssRNA) viruses (63.0%), followed by positive-sense (+) single-stranded RNA (ssRNA) (20.8%), double-stranded RNA (dsRNA) (8.5%), and unclassified RNA (5.6%) viruses, depending on version 2020 of the International Committee on Taxonomy of Viruses (ICTV) master species list (https://talk.ictvonline.org/files/master-species-lists/). We labelled each virus as being capable of infecting humans or not known to infect humans using published reports as ground truth. In all cases, only viruses detected in humans by either PCR or sequencing were considered to have proven ability to infect humans. See the summary table of currently known tick-borne viruses by 2021 in **Table S6** in the supplementary material.

We used representative genomic sequences of 232 tick-borne viruses in the original tick virus database as 'baits', with genome segments of segmented viruses concatenated into one. Candidate viral contigs were searched from Sequence Read Archive (SRA), Transcriptome Shotgun Assembly (TSA), and Whole Genome Shotgun (WGS). Tick metagenomic and meta-transcriptomic datasets were available at the National Centre for Biotechnology Information (NCBI) FTP server (https://ftp.ncbi.nlm.nih.gov/).

Bioinformatic analyses. Using the 'blastn_vdb' (corresponding to nucleotide BLAST) executable in the BLAST+ v2.13.0 (46-48), we searched the SRA, TSA, and WGS datasets that derived from the ticks and subordinate taxa directly with an expectation cut-off of 0.05. We also performed a 'tBLASTx' search for the TSA database to find distant relationships between nucleotide sequences; SRA and WGS datasets were not suitable for 'tBLASTx' search because it was computationally prohibitive. This resulted in 1328 TSA, 19,990 SRA, and 1861 WGS sequences after removing duplicates (as of 3 March 2022). The GenBank accession numbers of the candidate sequences were provided at https://github.com/ytlin2021/TickVirus.git.

Candidate sequences were filtered according to the pipeline modified from Webster et al. (49). The thresholds used in each step can be seen in **Figure 1**. To confirm final RNA candidates, we checked the presence of RNA-dependent RNA polymerase (RdRp), an essential protein encoded 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 adopted a combined analysis approach to predict RdRps by using a general protein function prediction software package InterProScan v5.55-88.0 (50) and a new open bioinformatics toolkit - RdRp-scan - that allowed the detection of divergent viral RdRp sequences (51). Candidate sequences less than 600 nucleotides (nt) in length were discarded, given the uncertainty we would have in relatedness and the lengths of viral RdRps (51). The translated candidate sequences were searched according to the RdRp-scan workflow (51) and were also predicted using the InterProScan search function. The conserved RdRp motifs were annotated manually based on the RdRp motif database in RdRp-scan (51), and final RNA viral candidates were confirmed. The workflow of detecting viral RdRp sequences and the thresholds used in each step can be seen in **Figure S1** in the supplementary material.

Protein functions of putative novel viral contigs were annotated using InterProScan and 'BLASTx'. Open Reading Frames (ORFs) were predicted using ORFfinder in NCBI with a 150 nt length cut-off. InterProScan searched for protein function annotations in its available databases, including the Gene Ontology and Pfam databases. At the same time, 'BLASTx' were used to search against the non-redundant protein database with an expectation cut-off of 1e-5 following Charon et al. (51).

Phylogenetics. We searched for close and distant relatives of the identified novel viral contigs in viruses (NCBI taxa ID: 10239) non-redundant nucleotide and protein databases by performing 'BLASTn' and 'BLASTx' separately with an expectation cut-off of 0.005. For each contig with BLAST hits, the phylogenetic hierarchy of the best hits (top 10 hits from 'BLASTx') was traversed upward to identify the lowest taxonomic classification displaying a 75% majority taxonomic agreement following Webster et al. (49). If the best hits were unclassified viruses, the upper order was identified, and the protein sequences were aligned in each family under the order to determine the classification. This allowed putative novel viral contigs in the study to be classified into three RNA viral families (*Alphatetraviridae*, *Orthomyxoviridae*, *Chuviridae*) and one DNA viral family (*Asfaviridae*). All virus species in the same family, as well as unassigned viruses, were aligned and used to construct phylogenetic trees.

Phylogeny of RNA viruses was inferred on the basis of RNA-dependent RNA polymerase protein sequences. We aligned viral RdRp protein sequences using the ClustalW algorithm (52) embedded in the GUIDANCE2 server v2.02 (53). GUIDANCE2 server removed any sequences with a score of less than 0.6 and columns with a score of less than 0.93, aiming to leave sufficient information in alignments while allowing large taxonomic groups of viruses to be compared.

We performed tree estimation for all trees in BEAST2, averaging over protein models using the OBAMA module (54,55). A lognormal relaxed clock model was applied (56-58), and a Yule tree prior was used (59,60). Models were fit using BEAST2's Metropolis-coupled MCMC routines with 4 chains and a target cross-chain move acceptance probability of 0.029 for the *Alphaviridae* and *Chuviridae* trees, and a fixed heating parameter of 0.3 for the *Orthomyxoviridae* tree (61,62), after convergence repeatedly failed under standard MCMC for the *Chuviridae* and *Orthomyxoviridae* trees. The Yule process birth rate was given a vague Exponential prior with mean 1000, all other parameters took default prior values. Trees were run for 100,000,000 iterations with 50,000,000 discarded as burn-in, and the posterior then being thinned to 1000 trees. A second run was used to assess convergence with the R package RWTY (63).

Zoonotic risk analyses. The probabilities of human infection for 136 tick-borne viruses with complete genome sequences were evaluated using the best performing model from Mollentze et al.

(30), referred to here as genome composition-based (GCB), which was trained on a range of viral genome features and human similarity features. A representative genome was selected for each virus, giving preference to sequences from NCBI Reference Sequence Database (RefSeq) (www.ncbi.nlm.nih.gov/refseq/) wherever possible. RefSeq sequences that had annotation issues or were not judged to be representative of the naturally circulating virus were replaced with alternative genomes. Following Mollentze et al. (30), the predicted probabilities from the GCB model were used to categorise viruses into four priority categories with the cut-off value of 0.293. Since the current GCB model was trained for viruses with complete sequences, 99 tick viruses with only partial genomes available were excluded from the analyses. To characterise the zoonotic potential of the identified novel viral contigs, we did qualitative analyses by evaluating the published complete genomes of known viruses most similar to these incomplete viral sequences as determined by a nucleotide BLAST against GenBank.

Data availability. All novel viral sequences identified in this study have been submitted to Third Party Annotation and will be deposited in the GenBank database along with the publication.

Acknowledgements.

We thank members of the Medical Research Council (MRC) Biostatistics Unit at the University of Cambridge (MRC BSU) for scientific input. We are grateful for the technical support of core facilities the Cambridge University High Performance Computing Service.

D.J.P. is funded by UKRI through the JUNIPER consortium (MR/V038613/1), and through the MRC Biostatistics Unit core award from the MRC (MC UU 00002/11).

We declare that no competing interests exist.

References

1. Stockman S. 1918. Louping-ill. J Comp Pathol Ther 31:137–193.
2. Mansfield KL, Jizhou L, Phipps LP, Johnson N. 2017. Emerging tick-borne viruses in the twenty-first century. Front Cell Infect Microbiol 7:298.
3. Kodama F, Yamaguchi H, Park E, Tatemoto K, Sashika M, Nakao R, Terauchi Y, Mizuma K, Orba Y, Kariwa H, Hagiwara K, Okazaki K, Goto A, Komagome R, Miyoshi M, Ito T, Yamano K, Yoshii K, Funaki C, Ishizuka M, Shigeno A, Itakura Y, Bell-Sakyi L, Edagawa S, Nagasaka A, Sadoka Y, Sawa H, Maeda K, Saijo M, Matsuno K. 2021. A novel nairovirus associated with acute febrile illness in Hokkaido, Japan. Nat Commun 12(1):5539.
4. Ma J, Lv XL, Zhang X, Han SZ, Wang ZD, Li L, Sun HT, Ma LX, Cheng ZL, Shoa JW, Chen C, Zhao YH, Sui L, Liu LN, Qian J, Wang W, Liu Q. 2021. Identification of a new orthonairovirus associated with human febrile illness in China. Nat Med 27(3):434–439.
5. Cowdry EV. 1925. A group of microorganisms transmitted hereditarily in ticks and apparently unassociated with disease. J Exp Med 41(6):817–830.
6. Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, Davis JP. 1982. Lyme disease—a tick-borne spirochetosis? Science 216(4552):1317–1319.
7. Benach JL, Bosler EM, Hanrahan JP, Coleman JL, Habicht GS, Bast TF, Cameron DJ, Ziegler JL, Barbour AG, Burgdorfer W, Edelman R, Kaslow RA. 1983. Spirochetes isolated from the blood of two patients with Lyme disease. N Engl J Med. 308(13):740–742.
8. La Scola B, Raoult D. 1997. Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. J Clin Microbiol 35(11):2715–2727.

9. Raoult D, Roux V. 1997. Rickettsioses as paradigms of new or emerging infectious diseases. Clin Microbiol Rev 10(4):694–719.
10. Mans BJ, Pienaar R, Latif AA. 2015. A review of Theileria diagnostics and epidemiology. Int J Parasitol Parasites Wildl 4(1):104–118.
11. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. Curr Opin Virol 2(1):63–77.
12. Tokarz R, Lipkin WI. 2021. Discovery and surveillance of tick-borne pathogens. J Med Entomol 58(4):1525–1535.
13. Garrison AR, Alkhovsky SV, Avšič-Županc T, Bente DA, Bergeron É, Burt F, Di Paola N, Ergünay K, Hewson R, Kuhn JH, Mirazimi A, Papa A, Sall AA, Spengler JR, Palacios G, ICTV Report Consortium. 2020. ICTV Virus Taxonomy Profile: Nairoviridae. J Gen Virol 101(8):798–799.
14. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. eLife 4:e05378.
15. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ. 2016. Redefining the invertebrate RNA virosphere. Nature 540(7634):539–543.
16. Bouquet J, Melgar M, Swei A, Delwart E, Lane RS, Chiu CY. 2017. Metagenomic-based surveillance of pacific coast tick *Dermacentor occidentalis* identifies two novel bunyaviruses and an emerging human ricksettsial pathogen. Sci Rep 7(1):12234.
17. Pettersson JHO, Shi M, Bohlin J, Eldholm V, Brynildsrud OB, Paulsen KM, Andreassen, Holmes EC. 2017. Characterizing the virome of *Ixodes ricinus* ticks from northern Europe. Sci Rep 7(1):10870.
18. Harvey E, Rose K, Eden JS, Lo N, Abeyasuriya T, Shi M, Doggett SL, Holmes EC. 2019. Extensive diversity of RNA viruses in Australian ticks. J Virol 93(3):e01358-18.
19. Tokarz R, Sameroff S, Tagliafierro T, Jain K, Williams SH, Cucura DM, Rochlin I, Monson J, Carpi G, Tufts D, Diuk-Wasser M, Brinkerhoff J, Lipkin WI. 2018. Identification of novel viruses in *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks. mSphere 3(2):e00614-17.
20. Chandra S, Harvey E, Emery D, Holmes EC, Šlapeta J. 2021. Unbiased characterization of the microbiome and virome of questing ticks. Front Microbiol 12;12:627327.
21. Shi J, Shen S, Wu H, Zhang Y, Deng F. 2021. Metagenomic profiling of viruses associated with *Rhipicephalus microplus* ticks in Yunnan Province, China. Virol Sin 36(4):623–635.
22. Xu L, Guo M, Hu B, Zhou H, Yang W, Hui L, Huang R, Zhan J, Shi W, Wu Y. 2021. Tick virome diversity in Hubei Province, China, and the influence of host ecology. Virus Evol 7(2):veab089.
23. Bergner LM, Mollentze N, Orton RJ, Tello C, Broos A, Biek R, Streicker DG. 2021. Characterizing and evaluating the zoonotic potential of novel viruses discovered in vampire bats. Viruses 13(2):252.
24. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. 2017. Host and viral traits predict zoonotic spillover from mammals. Nature 546(7660):646–650.
25. Pulliam JRC, Dushoff J. 2009. Ability to replicate in the cytoplasm predicts zoonotic transmission of livestock viruses. J Infect Dis 199(4):565–568.
26. Grange ZL, Goldstein T, Johnson CK, Anthony S, Gilardi K, Daszak P, Olival KJ, O'Rourke T, Murray S, Olson SH, Togami E, Vidal G, Expert Panel, the PREDICT Consortium, Mazet JAK. 2021. Ranking the risk of animal-to-human spillover for newly discovered viruses. Proc Natl Acad Sci 118(15):e2002324118.

27. Bartoszewicz JM, Seidel A, Renard BY. 2021. Interpretable detection of novel human viruses from genome sequencing data. NAR Genomics Bioinforma 3(1):lqab004.
28. Babayan SA, Orton RJ, Streicker DG. 2018. Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. Science 362(6414):577–580.
29. Zhang Z, Cai Z, Tan Z, Lu C, Jiang T, Zhang G, Peng Y. 2019. Rapid identification of human-infecting viruses. Transbound Emerg Dis 66(6):2517–2522.
30. Mollentze N, Babayan SA, Streicker DG. 2021. Identifying and prioritizing potential human-infecting viruses from their genome sequences. PLOS Biol 19(9):e3001390.
31. Forth JH, Forth LF, Lycett S, Bell-Sakyi L, Keil GM, Blome S, Calvigac-Spencer S, Wissgott A, Krause J, Höper D, Kampen H, Beer M. 2020. Identification of African swine fever virus-like elements in the soft tick genome provides insights into the virus' evolution. BMC Bio 18(1):136.
32. Cholleti H, Hayer J, Mulandane FC, Falk K, Fafetine J, Berg M, Blomström AL. 2018. Viral metagenomics reveals the presence of highly divergent quaranjavirus in Rhipicephalus ticks from Mozambique. Infect Ecol Epidemiol 8(1):1478585.
33. Mourya DT, Yadav PD, Nyayanit DA, Majumdar TD, Jain S, Sarkale P, Shete A. 2019. Characterization of a strain of quaranfil virus isolated from soft ticks in India. Is quaranfil virus an unrecognized cause of disease in human and animals? Heliyon 5(3):e01368.
34. Bratuleanu BE, Temmam S, Munier S, Chrétien D, Bigot T, van der Werf S, Savuta G, Eloit M. 2022. Detection of Phenuiviridae, Chuviridae members, and a novel quaranjavirus in hard ticks from Danube Delta. Front Vet Sci 9:863814.
35. Alphatetraviridae, ViralZone. [cited 2022 Jul 9]. Available from: https://viralzone.expasy.org/129?outline=all_by_species.
36. Wille M, Harvey E, Shi M, Gonzalez-Acuña D, Holmes EC, Hurt AC. 2020. Sustained RNA virome diversity in Antarctic penguins and their ticks. ISME J 14(7):1768–82.
37. Chen YM, Sadiq S, Tian JH, Chen X, Lin XD, Shen JJ, Chen H, Hoa ZY, Wille M, Zhou ZC, Wu J, Wang HW, Yang WD, Xu QY, Wang W, Goa WH, Holmes EC, Zhang YZ. 2022. RNA viromes from terrestrial sites across China expand environmental viral diversity. Nat Microbio 7:1312-1323.
38. Yuan C, Wu J, Peng Y, Li Y, Shen S, Deng F, Hu Z, Zhou J, Wang M, Zou Z. 2020. Transcriptome analysis of the innate immune system of *Hyalomma asiaticum*. J Invert Path 177;107481.
39. Quan L, Wang ZD, Gao Y, Lv X, Han S, Zhang X, Shao JW, Chen C, Li L, Hou SJ, Sui L, Zhao Y, Wang B, Wang W, Song M. 2020. Identification of a new chuvirus associated with febrile illness in China. Research Square https://doi.org/10.21203/rs.3.rs-104938/v1
40. Wang YC, Wei Z, Lv X, Han S, Wang Z, Fan C, Zhang X, Shao J, Zhao YH, Sui L, Chen C, Liao M, Wang B, Jin N, Li C, Ma J, Hou ZJ, Yang Z, Han Z, Zhang Y, Niu J, Wang W, Wang Y, Liu Q. 2021. A new nairo-like virus associated with human febrile illness in China. Emerg Microbes Infect 10(1):1200–1208.
41. Brinkmann A, Dinçer E, Polat C, Hekimoğlu O, Hacıoğlu S, Földes K, Özkul A, Öktem IMA, Nitsche A, Ergünay K. 2018. A metagenomic survey identifies Tamdy orthonairovirus as well as divergent phlebo-, rhabdo-, chu- and flavi-like viruses in Anatolia, Turkey. Ticks Tick-Borne Dis 9(5):1173–1183.
42. Temmam S, Chrétien D, Bigot T, Dufour E, Petres S, Desquesnes M, Devillers E, Dumarest M, Yousfi L, Jittapalapong S, Karnchanabanthoeng A, Chaisiri K, Gagnieur L, Cosson JF, Vayssier-Taussat M, Morand S, Moutailler S, Eloit M. 2019. Monitoring silent spillovers before emergence: A pilot study at the tick/human interface in Thailand. Front Microbiol 10;2315.

43. Bratuleanu BE, Temmam S, Chrétien D, Regnault B, Pérot P, Bouchier C, Bigot T, Savuța G, Eloit M. 2022. The virome of Rhipicephalus, Dermacentor and Haemaphysalis ticks from Eastern Romania includes novel viruses with potential relevance for public health. Transbound Emerg Dis 69(3):1387-1403.
44. Dong Z, Yang M, Wang Z, Zhao S, Xie S, Yang Y, Lui G, Zhao S, Xie J, Liu Q, Wang Y. 2021. Human Tacheng tick virus 2 infection, China, 2019. Emerg Infect Dis 27(2):594–8.
45. Shi J, Hu Z, Deng F, Shen S. 2018. Tick-borne viruses. Virol Sin 33(1):21–43.
46. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215(3):403–410.
48. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402.
49. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui JM, Bayne EH, Longdon B, Buck AH, Lazzaro BP, Akorli J, Haddrill PR, Obbard DJ. 2015. The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. PLOS Biol 13(7):e1002210.
50. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
51. Charon J, Buchmann JP, Sadiq S, Holmes EC. 2022. RdRp-scan: A bioinformatic resource to identify and annotate divergent RNA viruses in metagenomic sequence data. Virus Evol 8(2):veac082.
52. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680.
53. Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res 43(W1):W7–14.
54. Bouckaert RR, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C Stadler T, Drummond AJ. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 15(4):e1006650.
55. Bouckaert RR. 2020. OBAMA: OBAMA for Bayesian amino-acid model averaging. PeerJ 8:e9460
56. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol 4(5):e88.
57. Bouckaert RR. 2022. An efficient coalescent epoch model for Bayesian phylogenetic inference. Syst Biol 71(6):1549-1560.
58. Zhang R., Drummond A. 2020. Improving the performance of Bayesian phylogenetic inference under relaxed clock models. BMC Evol Biol 20:54.
59. Yule GV. 1924. A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, P.R.S. Phil Trans R Soc B 213:21-87.

60. Edwards AWF. 1970. Estimation of branch points of a branching diffusion process. J Roy Stat Soc B 32:155-174.
61. Müller NF, Bouckaert RR. 2020. Adaptive Metropolis-coupled MCMC for BEAST 2. PeerJ 8:e9473.
62. Altekar G, Dwarkadas S, Huelsenbeck J, Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20(3):407-415.
63. Warren D, Geneva A, Lanfear R. 2017. RWTY (R We There Yet): An R package for examining convergence of Bayesian phylogenetic analyses. Mol Biol Evol 34(4):1016-1020.