Discovery and Evolution of Bunyavirids in Arctic Phantom Midges and Ancient Bunyavirid-Like Sequences in Insect Genomes

Matthew J. Ballinger,^a Jeremy A. Bruenn,^a John Hay,^b Donna Czechowski,^b Derek J. Taylor^a

Department of Biological Sciences^a and Department of Microbiology and Immunology,^b State University of New York at Buffalo, Buffalo, New York, USA

ABSTRACT

Bunyaviridae is a large family of RNA viruses chiefly comprised of vertebrate and plant pathogens. We discovered novel bunyavirids that are approximately equally divergent from each of the five known genera. We characterized novel genome sequences for two bunyavirids, namely, Kigluaik phantom virus (KIGV), from tundra-native phantom midges (*Chaoborus*), and Nome phantom virus (NOMV), from tundra-invading phantom midges, and demonstrated that these bunyavirid-like sequences belong to an infectious virus by passaging KIGV in mosquito cell culture, although the infection does not seem to be well sustained beyond a few passages. Virus and host gene sequences from individuals collected on opposite ends of North America, a region spanning 4,000 km, support a long-term, vertically transmitted infection of KIGV in *Chaoborus trivittatus*. KIGV-like sequences ranging from single genes to full genomes are present in transcriptomes and genomes of insects belonging to six taxonomic orders, suggesting an ancient association of this clade with insect hosts. In *Drosophila*, endogenous virus genes have been coopted, forming an orthologous tandem gene family that has been maintained by selection during the radiation of the host genus. Our findings indicate that bunyavirid-host interactions in nonbloodsucking arthropods have been much more extensive than previously thought.

IMPORTANCE

Very little is known about the viral diversity in polar freshwater ponds, and perhaps less is known about the effects that climate-induced habitat changes in these regions will have on virus-host interactions in the coming years. Our results show that at the tundra-boreal boundary, a hidden viral landscape is being altered as infected boreal phantom midges colonize tundra ponds. Likewise, relatively little is known of the deeper evolutionary history of bunyavirids that has led to the stark lifestyle contrasts between some genera. The discovery of this novel bunyavirid group suggests that ancient and highly divergent bunyavirid lineages remain undetected in nature and may offer fresh insight into host reservoirs, potential sources of emerging disease, and major lifestyle shifts in the evolutionary history of viruses in the family *Bunyaviridae*.

Little is known of viral diversity and virus-host interactions in polar freshwaters, one of the most threatened ecosystems. With recent warming and permafrost thawing, arctic ponds have suffered dramatic drainage in some areas and expansion in other areas (1–3), which is expected to dramatically affect freshwater insect populations in these regions and, by extension, any virus populations they may harbor (4–6). However, the effects of future global change on virus-host interactions will be difficult to understand without an understanding of virus-host interactions before radical ecosystem alterations. The gap in knowledge is perhaps largest for arthropod hosts in which viruses are either asymptomatic or beneficial to the host (7). Several authors have suggested publication and detection biases for pathogenic interactions over nonpathogenic interactions (7–9). Thus, for some groups of RNA viruses, the “silent” majority may remain undetected, because genetic probes, sequence and motif searches, and even host preservation methods favor either pathogenic or DNA-based taxa. A further focus on viruses from bloodsucking arthropods may bias discovery against host specificity.

While exploring arthropod transcriptomes for viruses, we discovered evidence of bunyavirid-like infections in arctic phantom midges (*Chaoborus*). These flies are commonly used as biological indicators of environmental change in freshwater ecosystems. The larvae of *Chaoborus* flies are also among the most important predators of freshwater zooplankton. One species, *Chaoborus trivittatus*, is adapted to both arctic and boreal habitats; others, such as

Chaoborus cf. flavicans (a *Chaoborus* sp. that looks like *Chaoborus flavicans*), show a strong preference for boreal habitats but have recently expanded from boreal to newly formed tundra ponds in Beringian Alaska (10; D. J. Taylor, M. J. Ballinger, A. S. Medeiros and A. A. Kotov, submitted for publication). Chaoborids share an ancestor with mosquitoes, but chaoborids lack the blood-feeding habit of other culicomorphs. With the evolutionary loss of hematophagy (11), viruses of chaoborids may be more amenable than those of mosquitoes to the evolution of host specificity.

The family *Bunyaviridae* is comprised of five evolutionarily divergent genera of RNA viruses: *Hantavirus*, *Nairovirus*, *Orthobunyavirus*, *Phlebovirus*, and *Tospovirus*. With the exception of the plant- and insect-infecting tospoviruses, bunyavirids infect both arthropods and vertebrates and are agents of human disease, including hemorrhagic fever, encephalitis, and influenza-like re-

Received 20 February 2014 Accepted 18 May 2014

Published ahead of print 21 May 2014

Editor: R. W. Doms

Address correspondence to Matthew J. Ballinger, ballinge@buffalo.edu.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JVI.00531-14>.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.00531-14

spiratory illness. The viral genome is divided among three single-stranded, negative-sense segments. These three segments are designated the large (L), medium (M), and small (S) segments, named for their lengths, in nucleotides, but they are characterized by the genes they carry. L encodes the viral L protein, an RNA-dependent RNA polymerase (RdRp); M encodes the glycoprotein precursor protein, which is cleaved into the Gn and Gc proteins; and S encodes the viral nucleocapsid protein (N). In some genera, the M and S segments encode additional nonstructural proteins (NSm and NSs) by ambisense, overlapping, or typical coding strategies. The termini of each genome segment possess short, noncoding nucleotide sequence motifs that complement the sequence of the opposite terminus and assist in the formation of a panhandle-like structure between the termini of the packaged genome. While this feature is conserved among all bunyavirids, the sequences themselves vary between genera. All bunyavirids have enveloped virions, and maturation takes place intracellularly, at the Golgi membrane (12).

In regard to the evolutionary history of bunyavirids, much remains unknown. A minimum age of the family can be inferred from that of the most ancient genus, which is thought to be *Hantavirus*. Long-term codivergence with rodents, insectivores, and bats (~100 million years [MY]) has been proposed by many authors, based on topological congruence between virus and host phylogenies (12–17), though this evidence has been challenged by others (18). Paleovirological evidence is a powerful complement to data from extant viruses because it has the potential to extend the time scales of virus analyses into the distant past. Paleoviruses, or endogenous viral elements (EVEs), are “fossilized” sequences of viral origin that have been preserved following their integration into host genomes (19). Endogenous retroviruses (ERVs) are a well-known class of EVEs, but perhaps less well known are the nonretroviral integrated RNA viruses (NIRVs), which are formed from RNA viruses lacking reverse transcriptase, including bunyavirids (20–23). Homology between specific EVEs in related host genomes has provided compelling evidence for the antiquity of virus families, which extends far beyond the threshold of positional mutation saturation (23), which constrains the reach of molecular clock-based methods for virus divergence date estimation (24).

Here we provide evidence of an insect-infecting group of bunyavirids that is genetically divergent from all known genera. We performed culture experiments with mosquito cells to obtain an isolate and demonstrate infectivity. For *C. trivittatus*, we present evidence of long-term virus-host codivergence, multiyear persistence, and an exceptionally high prevalence. The tundra invader, *Chaoborus cf. flavicans*, is also infected with a virus belonging to this novel group. Using the gene sequences from each of these divergent bunyavirids, we identified virus-like sequences in the genomes and transcriptomes of other insects, including species of *Drosophila* and *Anopheles*, and carried out bioinformatic analyses to better understand the evolution of these genes and of the novel viruses. The results of these analyses support an ancient age for this virus group (>20 MY), a remarkable case of cooption of a viral gene by a host, and a historical association with a wide range of insect hosts.

MATERIALS AND METHODS

Field collections and sample preservation. *Chaoborus* larvae were collected from freshwater ponds in late July or early August of 2011 to 2013 by multiple oblique tows, using a 200- to 250- μ m throw or dip net

(Wildco Scientific), and were stored in 100% ethanol at -20°C prior to cDNA library construction. Specimens were identified to the species level according to larval morphology (mandibles and labral blades) and COI DNA sequence information. A subsample of one population of *C. trivittatus* collected in 2013 was stored in a 15% trehalose solution at -80°C to preserve viral particles for cell culture experiments. Collection site GPS coordinates are listed in Table S1 in the supplemental material.

RNA-seq and RT-PCR. To produce transcriptome sequencing (RNA-seq) libraries, RNAs were purified from host tissues by use of an RNeasy minikit (Qiagen). rRNA was removed using a Ribo-Zero Gold rRNA removal kit (Epicentre), and library generation was carried out using a ScriptSeq v2 RNA Seq library preparation kit (Epicentre). Libraries were quantified with an Agilent 2100 Bioanalyzer RNA 6000 picochip, and RNA sequencing was carried out at the University at Buffalo next-generation sequencing facility. RNA library sequencing was done using 50-cycle paired-end runs on two flow cells of an Illumina HiSeq 2000 instrument. CLC Assembly Cell 4.06 (CLCbio) was used for *de novo* sequence assembly. To confirm that the CLC software had correctly assembled the ends of the genomic segments, we performed 5' rapid amplification of cDNA ends (RACE) by using a SMARTer RACE 5'/3' kit (Clontech Laboratories) according to the manufacturer's specifications. Individual RNA extractions were done by using QuickExtract (Epicentre) and RQ1 RNase-free DNase (Promega) according to the manufacturers' protocols. For reverse transcription-PCR (RT-PCR), cDNA was generated using GoScript reverse transcriptase (Promega). GoTaq master mix (Promega) was used for PCR amplification. PCR primer sequences and thermal cycling programs are listed in Table S1 in the supplemental material. RNP isolation was done by CsCl density gradient centrifugation as described previously for application to tospoviruses (25).

Bioinformatics. *In silico* filtering of *Anopheles sinensis* transcripts (Bioproject no. PRJNA186896) was performed to identify virus S segment candidates. *A. sinensis* transcripts of 2.0 to 2.5 kb were queried against the *A. sinensis* genome assembly (accession no. GCA_000472065.2) by using the MegaBLAST search algorithm, with an expect-value cutoff of $1e-05$. RNA transcripts lacking corresponding genomic coding matches by these criteria were then queried against the *C. trivittatus* transcriptome by using the tBLASTx search algorithm and the same expect-value cutoff.

Novel bunyavirid sequences were translated and aligned with established bunyavirid amino acid sequences by using the MAFFT Web server with the E-INS-i alignment algorithm (26). For Fig. 1, the included sequences were limited to the region from the endonuclease domain to motif E of the conserved polymerase domains. Following alignment of these sequences, alignment quality was assessed using the scoring program TCS (27), and only columns with scores of 5 or greater were retained; 325 amino acid positions were present in the final alignment. The optimal model of protein sequence evolution was identified using ProtTest 3 (28). Phylogenetic trees were constructed under the LG + I + G + F model of amino acid substitution. SeaView 4.3.5 (29) was used to run maximum likelihood analyses, MrBayes implemented on the CIPRES Web portal (30) was used to run Bayesian analyses, and FigTree 1.4 (31) was used to visualize trees. MrBayes does not support the LG model, so we used the second-best model determined by ProtTest for that analysis, which was rtREV + I + G + F. Paleoviruses were identified with tBLASTn searches querying phasmavirus amino acid sequences against the WGS and reference genome databases available on GenBank. Matches with expect values of $\leq 1e-05$ were retained as phasmavirus-like paleoviruses. Phylogenetic analyses were carried out as described above.

GenBank accession numbers for established members of the *Bunyaviridae* presented in Fig. 1 and Table 1 are as follows: impatiens necrotic spot virus, NC_003625; tomato spotted wilt virus, NC_002052; tomato zonate spot virus, NC_010491; capsicum chlorosis virus, NC_008302; watermelon silver mottle virus, NC_003832; Herbert virus, AFR34023; Kibale virus, KF590577; Tai virus, KF590574; Bunyamwera virus, X14383; La Crosse virus, GU596376; Oyo virus, HM639780; Douglas virus, HE795090; Leanyer virus, HM627178; Carrizal virus, AB620105; Pu-

TABLE 1 Similarities between complete polymerase amino acid sequences of novel, established, and unclassified bunyavirids

Virus	% Similarity ^a					
	<i>Hantavirus</i>	<i>Orthobunyavirus</i>	<i>Tospovirus</i>	<i>Phlebovirus</i>	<i>Nairovirus</i>	KIGV
Kigluaik phantom virus (KIGV)	15.4	13.9	13.3	12.7	9.3	
Nome phantom virus	14.9	12.8	10.9	10.9	8.7	29.8
Hantaan virus	73.2	13.5	13.4	12.8	10.0	17.5
Bunyamwera virus	16.0	50.2	15.4	12.4	9.3	16.5
Tomato spotted wilt virus	14.0	14.9	49.9	13.5	12.0	14.7
Rift Valley fever virus	15.2	13.5	11.7	44.8	9.8	16.8
Dugbe virus	10.3	9.8	10.7	9.4	59.2	9.7
Herbert virus	14.6	24.7	16.9	11.6	11.2	15.1
Gouleako virus	15.5	13.0	12.5	23.7	9.3	17.3

^a Genera are represented by the five viruses present for each genus in Fig. 1. See Materials and Methods for further information regarding these calculations.

umala virus, [BAJ14125](#); Tula virus, [NP_942124](#); Dobrava-Belgrade virus, [ADP21260](#); Hantaan virus, [BAK08372](#); Gouleako virus, [AEJ38175](#); rice grassy stunt virus, [NP_058528](#); rice stripe virus, [Q85431](#); Uukuniemi virus, [NC_005214](#); Heartland virus, [JX005847](#); Turuna virus, [HM119431](#); Rift Valley fever virus, [NC_014397](#); Toscana virus, [NC_006319](#); Erve virus, [JF911697](#); Dugbe virus, [U15018](#); Hazara virus, [DQ076419](#); Nairobi sheep disease virus, [EU697951](#); and Crimean-Congo hemorrhagic fever virus, [AY422209.2](#).

Amino acid identity calculations between subject sequences and the established genera of *Bunyaviridae* (Table 1) were performed as follows. For each genus, the five polymerase sequences presented in Fig. 1 were

aligned with one subject sequence by use of the MAFFT E-INS-I algorithm. Amino acid similarities between the subject and each of the five sequences present in each alignment were averaged to present the mean amino acid sequence similarity for each comparison. For the last column of the table, subject sequences were aligned individually with the Kigluaik phantom virus (KIGV) polymerase sequence.

Phasmavirus Gn-Gc domain and topology predictions (Fig. 2A) were done using SignalP 4.1 (32) for signal peptide and cleavage prediction, TMHMM 2.0 for transmembrane domain prediction (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), and NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) for glycosylation site prediction.

FIG 1 Evolutionary relationships between bunyavirid polymerase sequences. A phylogram of RdRp amino acid sequences belonging to representatives of the five established genera of *Bunyaviridae*, currently unclassified bunyavirids, and the novel bunyavirids described in this study supports the evolutionary position of the novel viruses as members of the family *Bunyaviridae*. Branches are labeled with Bayesian posterior probabilities/approximate likelihood ratio test scores. Taxonomic groups are labeled with their accepted or proposed names.

FIG 2 Sequence, genome structure, and PCR evidence of a novel, exogenous bunyavirid. (A) A graphical representation of the genomic structure for each of the three genome segments of KIGV is shown in positive-sense directionality. On the M segment, locations of predicted signal peptides and cleavage sites within the glycoprotein precursor protein are indicated by pink blocks and scissors, respectively. The predicted Gn-Gc cleavage site (pink asterisk) is shown to the right, as a sequence alignment between KIGV, Nome phantom virus (NOMV), and four phleboviruses (Jacunda virus [JANV] [included because this is the closest BLASTp match to KIGV Gn-Gc], sandfly fever Naples virus [SFNV], Rift Valley fever virus [RVFV], and Uukuniemi virus [UUKV]). Above the M segment is the result of transmembrane domain (TMD) prediction. Gray blocks represent predicted TMDs, dark blue blocks represent external domains, and light blue blocks represent internal domains. Y's represent predicted glycosylation sites. (B) Sequences and predicted secondary structures of the complementary terminal NCRs of the three KIGV segments and of the M and S segments of NOMV. Sequences underlined in green are those that NOMV shares with KIGV. (C) Agarose gel electrophoresis following RT-PCR experiments targeting the putative viral segments of KIGV demonstrates that products were amplified from RNA templates and not from DNA templates. Primer targets and treatment conditions are displayed above the appropriate lanes. (D) Agarose gel electrophoresis following RT-PCR experiments targeting the putative viral segments of KIGV, enriched by CsCl ribonucleoprotein particle isolation, shows that each putative viral segment was enriched relative to the host-encoded transcript.

BEAST v2.0.1 and the graphical user interface BEAUTi (33) were used to estimate the divergence date for the N-like NIRV family in *Drosophila*. For the substitution model, jModelTest (34) was used to determine the most likely model and parameters for nucleotide evolution. The clock model was set to relaxed log normal, and an exponential distribution prior offset at 20 MY calibrated the ancestral node of each NIRV clade. The analysis was run for 10 million generations and sampled at every thousand generations. The first 10,000 generations were discarded as burn-in. Tree-Annotator (33) was used to generate the averaged NIRV tree.

Cell culture. Virus culturing was done using biosafety level 2 (BSL-2) safety practices in a class II biosafety hood in a registered laboratory. KIGV was isolated from *C. trivittatus* larvae that had been collected in July 2013 and stored at -80°C . Ten larvae were homogenized for 60 s in cell culture medium containing a combination of 2-mm yellow zirconia beads and 4-mm black ceramic beads (MP Biomedicals). Cells were split into 9.62-cm² 6-well culture plates and grown to 80% confluence before inoculation. *Aedes albopictus* c6/36 cell cultures were inoculated with 250 μl of the supernatant or a 1:10 dilution and grown in 3 ml liquid medium at 28°C and 5% CO_2 in Eagle minimal essential medium (EMEM) with 10% fetal bovine serum (FBS) (ATCC) for 0, 1, 2, 4, and 8 days before harvesting. Samples from the zero time point were harvested following a 1-hour incubation under growing conditions. After harvesting, cell pellets were spun down at 3,000 rpm for 10 min, nucleic acids were purified from 200 μl of supernatant by using an RNeasy minikit (Qiagen), and each sample was screened for KIGV RNA by RT-PCR using specific primers targeting the RdRp. RT-PCR products were visualized by DNA electrophoresis on a 1% agarose gel. For passaging, 250 μl of supernatant was collected from day 9 samples and passaged along with 500 μl fresh medium to *A. albopictus* cells. Following a 1-hour incubation, all of the supernatant was aspirated and replaced with 3 ml fresh medium. Samples were collected and

tested as described above, at days 0, 5, 9, and 14 for passage 1 and days 0, 5, 10, and 14 for passage 2.

Nucleotide sequence accession numbers. Nucleotide sequences generated during the course of this project have been deposited in GenBank. Accession numbers for virus genome segments are [KJ434182](#) to [KJ434187](#), and those for *C. trivittatus* COI and KIGV RdRp PCR products are [KJ461793](#) to [KJ461811](#).

RESULTS

Genome segment identification. We identified sequences similar to bunyavirid polymerase queries by performing BLAST searches against RNA-seq transcriptomes of the phantom midges *C. trivittatus* (subgenus *Schadonophasma*) and *Chaoborus* cf. *flavicans* (subgenus *Chaoborus*) and performed phylogenetic analyses which placed these polymerases, with strong support, inside the family *Bunyaviridae* (Fig. 1). We performed RT-PCR assays to confirm that these sequences were derived from exogenous RNA rather than expressed copies of endogenous bunyavirid-like genome fragments carried by the host genome (Fig. 2C). Following this confirmation, we assigned the tentative names Kigluaik phantom virus (KIGV) and Nome phantom virus (NOMV), named for the locations of the *C. trivittatus* and *Chaoborus* cf. *flavicans* collection sites, respectively. The two phantom midge-infecting bunyavirids display only 30% amino acid similarity between the RdRp genes, suggesting that they are distantly related members of a diverse group, possibly a novel genus, which we informally refer to as phasmaviruses. We attempted to identify the remaining genome segments in our transcriptomes by tBLASTn searches, using

TABLE 2 Terminal sequences, genome size comparisons, and gene size comparisons between novel and established bunyavirids^d

Virus	Terminal sequence	Genome or fragment size (nt)				Mol wt						Reference
		Genome	L	M	S	RdRp	Gn ^a	Gc ^a	N	NSs ^b	S3 ^c	
Kigluaik phantom virus	UCGUCGUGCG	11,701	6,697	2,792	2,212	245.639	28.981	52.683	43.309	13.261	20.987	This study
Nome phantom virus	UCGUCGUUCG			2,533	2,667		34.395	48.208	45.517	16.804	33.483	This study
Hantaan virus	AUCAUCAUCU	11,845	6,533	3,616	1,696	246.499	63.040	52.904	48.142			51–53
Bunyamwera virus	UCAUCACAUG	12,294	6,875	4,458	961	258.670	97.864	32.241	26.664	11.024		54–56
Rift Valley fever virus	UGUGUUUC	11,979	6,404	3,885	1,690	237.977	58.433	55.332	27.360	29.931		57
Tomato spotted wilt virus	UCUCGUUA	16,634	8,897	4,821	2,916	331.503	48.202	75.118	28.845	52.448		58–60
Dugbe virus	AGAGUUUCU	18,859	12,255	4,888	1,716	459.392	35.129	73.198	53.917			61–63

^a Predicted from the complete glycoprotein precursor sequences.^b Putative ORF 1 of the S segment for the novel viruses.^c Putative ORF 3 of the S segment.^d Bioproject numbers for the genome sequences used were as follows: Hantaan virus, PRJNA14929; Bunyamwera virus, PRJNA14649; Rift Valley fever virus, PRJNA14631; tomato spotted wilt virus, PRJNA14997; and Dugbe virus, PRJNA14851.

Gn-Gc, NSm, N, and NSs amino acid sequences of known bunyavirids as queries. This approach yielded a low-scoring match to a phlebovirus Gc conserved domain (pfam07245) for a *C. trivittatus* RNA contig of 2.7 kb that encodes an open reading frame (ORF) for a 727-amino-acid protein. No prospective S segment candidates were found by BLAST searches. We optimized our search approach to identify more divergent sequences by decreasing the BLAST significance threshold and performing searches using domain-focused algorithms, such as HMMER (<http://hmm.janelia.org>), but we failed to identify a putative S segment. The identification of an S segment candidate came after we discovered matches to the KIGV L and M segments in the transcriptome of a malaria mosquito, *Anopheles sinensis* (Bioproject no. PRJNA186896), and confirmed that these L- and M-like sequences were not present in the *A. sinensis* reference genome assembly (GCA_000472065.2), suggesting that they are exogenous RNA rather than paleovirus sequences; therefore, we reasoned than an S segment could also be present. We used the *A. sinensis* genome as an *in silico* filter to identify any additional RNA sequences present in the transcriptome that lacked corresponding genomic coding regions, and we screened those for matches in the *C. trivittatus* transcriptome. This approach yielded two *C. trivittatus* contigs; the first was the M segment identified previously, and the second was a contig of 2.17 kb which encoded a putative protein of 396 amino acids (43.3 kDa) and returned significant matches ($\leq 1e-5$) to an uncharacterized protein found in many species of *Drosophila*. This candidate was initially thought to be a host gene; however, the lack of orthologous genes in many other dipteran genomes led us to wonder whether this could indeed be the viral S segment with paleoviral copies in *Drosophila*. We performed PCR and RT-PCR to determine whether this gene is present in the *C. trivittatus* genome. *Taq* polymerase-only PCRs consistently failed, while RT-PCR analysis of DNase-treated templates amplified the segment, indicating that the primer targets were RNA only (Fig. 2C). To provide further evidence that this S segment candidate was not a host-encoded transcript, we enriched the RNA virus genomes by purifying *C. trivittatus* RNP particles by CsCl density gradient centrifugation. We collected a fraction from a visible band at the 1.3- to 1.5-g/cm³ interface,

purified nucleic acids, and performed RT-PCRs targeting each of the putative virus genomic segments, which confirmed their enrichment (Fig. 2D) relative to host transcripts.

The RT-PCR and RNP isolation approaches provided evidence that the S candidate was an exogenous RNA virus sequence but not that it was part of the same bunyavirid genome as the L and M segments. We looked in the 3' and 5' noncoding regions (NCRs) of the S segment candidate for evidence that it was the missing bunyavirid segment. Terminal complementarity is characteristic of bunyavirids and other segmented RNA viruses. Within all five established genera of *Bunyaviridae*, these complementary sequences are conserved between genomic segments, and in some genera they have been shown to play important roles in RNA synthesis and genome packaging (35, 36). We looked for similar sequence characteristics in the phasmavirus genome segment candidates and identified complementary terminal sequences conserved among all three segments (Fig. 2B). We confirmed that our assembly had correctly determined the genome termini by performing 5' RACE on the KIGV L and S segments. For L, the RACE-determined terminus matched the assembled L segment from our transcriptome to the nucleotide. The S RACE result was nearly identical, save for the terminal nucleotide, a uracil in the assembly which was absent by RACE. The first seven of the terminal nucleotides may be genus specific, as they are identical between KIGV and NOMV segments but distinct from those of established bunyavirids (Table 2). Interestingly, the positions beyond the first seven show divergence between KIGV and NOMV in both sequence and structure. In KIGV, sequence identity is conserved and complementarity is uninterrupted for 10 positions in all three segments, while in NOMV, complementarity is interrupted once. For NOMV, terminal sequence comparison for the L segment is absent because we were unable to assemble the complete segment.

Due to the high level of sequence divergence between these phasmaviruses and their closest known bunyavirid relatives, we considered that undiscovered sequences of phasmaviruses could be present in transcriptome databases generated for other studies. We used the KIGV and NOMV genes to query arthropod RNA transcriptome sequence databases in GenBank and identified par-

tial matches to each segment in four other insects: *A. sinensis*, the mosquito used to help identify candidates for the viral segments; the Asian corn borer moth, *Ostrinia furnacalis*; the mud dauber wasp, *Sceliphron caementarium*; and the builder bee, *Osmia cornuta*. The RdRp sequence of each of these is included in Fig. 1. Interestingly, the two RdRp sequences from hymenopteran transcriptomes (both belonging to the superfamily Apoidea) were the most closely related of the putative phasmavirus sequences we identified, much more so than those from the two phantom midge-infecting members. We also identified shorter but still significant matches (expect values of $<1e-10$) to the glycoprotein and nucleoprotein in an *Aedes albopictus* oocyte-generated transcriptome that are not present in the *A. albopictus* genome.

Genome structure and domain conservation. As shown in Fig. 2 and Table 2, the phasmavirus genome structure is comparable to that of established bunyaviruses. L carries a single ORF, encoding the RdRp. M encodes a relatively short glycoprotein precursor, and there is no obvious NSm candidate ORF in overlapping or ambisense orientation. S carries three putative ORFs. The first is short (encodes 13 to 17 kDa) and overlaps the 5' coding region of the second, the second encodes a 43-kDa protein, which is similar in size to hantavirus N, and the third overlaps the 3' coding region of the second. We identified the central ORF as the N gene, as it is the only one that exhibits domain conservation between phasmaviruses. As for the two putative ORFs flanking the N gene, one of them may be homologous to the NSs gene of other bunyavirids, and the other could be a functional equivalent of NSm, or an entirely novel gene unique to this group. Further work will be needed to determine whether these predicted ORFs carry functional genes, but the conservation of this unique genomic structure between all three dipteran phasmaviruses, despite extensive sequence divergence, suggests that these putative ORFs have a functional role.

We aligned translated amino acid sequences to compare domain conservation between established bunyavirids and phasmaviruses. Each of these alignments is available in Fig. S1 in the supplemental material. All the highly conserved active domains of the nuclease and polymerase modules of the RdRp are present (37, 38). BLAST results indicated that the phasmavirus glycoprotein shows sequence similarity only to phleboviruses, so we restricted our domain alignments to this genus. The Gn-Gc cleavage site is visibly conserved (Fig. 2A), but the phasmavirus Gn is much shorter than that of phleboviruses and seems to bear little resemblance, as evident in the alignment; in fact, the NOMV M segment is shorter than its S segment (Table 2). The nucleoprotein returned no significant BLAST matches to known bunyavirids, but alignment to individual genera revealed putative conserved motifs with hantavirus nucleoprotein, as shown in Fig. S1. Incongruence between the viral segments could be indicative of ancient segment reassortment having led to the ancestral phasmavirus, but the extents of divergence between the phasmavirus RdRp and those of other genera are nearly equivalent (Table 1), suggesting that lineage sorting between segments may simply be a result of the extensive time and varied selective pressures between genera.

KIGV replication in mosquito cell culture. To confirm that KIGV is an infectious virus and to obtain an isolate, we inoculated *Aedes albopictus* c6/36 cells with KIGV-infected *C. trivittatus* tissue homogenate, harvested cells at the indicated time points, and screened the supernatants for the viral RdRp by RT-PCR. The amplified products showed an initial decrease in amplification

FIG 3 KIGV infection experiments in mosquito cell culture. (A) Agarose gel electrophoresis of RT-PCR products amplified with primers targeting the KIGV polymerase in an *Aedes albopictus* c6/36 cell culture inoculated with KIGV-infected *Chaoborus trivittatus* tissue homogenate shows increasing viral RNA detection with days postinoculation. (B) RT-PCR results of a 2-week passage of KIGV in *Aedes albopictus* cells. Numbers above lanes indicate the days on which samples were tested.

between hours 1 and 24, followed by an unambiguous increase after the first day (Fig. 3A). We considered the possibility that residual *Chaoborus* cells were supporting the viral growth seen in *A. albopictus* cells, but the use of filtration and centrifugation of extracts makes that seem unlikely. Also, we were able to use *A. albopictus* cell supernatants to successfully infect fresh *A. albopictus* cell cultures (Fig. 3B). However, failure to go beyond a second round of viral growth by this method is puzzling and suggests that some *Chaoborus* factor(s) may be important for sustainable viral replication.

Primers specific to the KIGV M and S segments also yielded amplified products from this cell culture supernatant (data not shown). Note that we did identify phasmavirus-like sequences (fragments of G and N) in *Aedes albopictus*, but this could not be the source of amplification, as demonstrated by their failure to be amplified from the mock treatment sample. In addition, while the *A. albopictus* nucleoprotein fragment is one of the closest phasmavirus relatives of KIGV (51% nucleotide identity) (see Fig. 5), it is too divergent for KIGV-specific primers to efficiently amplify, and the observed amplification pattern is consistent with an increase in viral RNA following inoculation, while an *A. albopictus*-specific virus would be expected to be amplified at each time point. We are not yet able to report on whether KIGV is able to infect vertebrate cells, but we consider this to be an important question to address and aim to do so in the future.

KIGV infection prevalence and geographic distribution. We determined the frequencies of KIGV infection in three Alaskan *C. trivittatus* populations by RT-PCR screening for each of the three segments. By screening 20 larvae per population, we found that 76% of *C. trivittatus* larvae were infected with KIGV. For one of these populations, we also screened for infection in each of three consecutive field seasons (larvae were collected in late July of 2011, 2012, and 2013). The prevalence of KIGV infection decreased throughout the time interval, from 70% in 2011 to 60 and 45% in 2012 and 2013, respectively. Further sampling and screening in the future will be required to determine whether this is reflective of overall declining rates of KIGV infection in the study area. *C. trivittatus* populations collected in Iqaluit, Baffin Island, and Salmon Arm, British Columbia, Canada (Fig. 4A), were also positive for L segment RNA. Eight individuals representing eight pop-

FIG 4 Evidence for long-term, vertically transmitted KIGV infection in *Chaoborus trivittatus*. (A) Map of the northern United States and Canada, with colored stars indicating locations from which KIGV-positive *Chaoborus trivittatus* flies were collected. Pink, Seward Peninsula of Alaska; green, Iqaluit, Baffin Island, Canada; blue, Salmon Arm, British Columbia, Canada. (B) Virus and host nucleotide sequence-based maximum likelihood phylogenies showing topological congruence between the viral polymerase and the host mitochondrial COI gene. Branches are labeled with approximate likelihood ratio test scores. Tips are colored to match the host collection site to the geographic locations shown in panel A. (C) Parsimony tanglegram built from the virus and host sequences used for panel B, shown to provide resolution to within-region relationships.

ulations from Baffin Island were screened, and seven were positive. One individual from British Columbia was screened and was positive. For all screens, controls to exclude genomic DNA as the source of amplification were performed. Reagent contamination and other concerns related to a secondary source for virus amplification are unlikely, as sequencing of the products revealed distinct virus genotypes both within and between populations. Sequences within Beringian populations were 99.1 to 100% similar for each of the three genomic segment fragments that were sequenced. Between Beringian populations, these values were 95.8 to 99.8% for the L fragment, 95.4 to 99.6% for the M fragment,

and 94.2 to 98.9% for the S fragment. One population contained two virus groups, with one being 95.8 to 96.2% similar to the other. These sequences form two distinct phylogenetic clades, neither of which groups with either of the other two populations we screened. It is possible that the source of one of these genotypes is migration from a nearby population that we have not sampled. The sequence similarity between the L fragments from the Seward Peninsula of Alaska and Baffin Island was 87.2 to 87.7%, that between Seward and British Columbia was 86.3 to 86.9%, and that between Baffin Island and British Columbia was 96.2 to 96.7%. We constructed phylogenies for both the host mitochondrial DNA (mtDNA) and virus polymerase sequences and found a clear pattern of topological congruence between virus and host (Fig. 4B), which makes a strong case for persistent, vertically transmitted infection in *C. trivittatus*. Data from additional chaoborids in the subgenus *Schadonophasma* may help to address how deep the cophylogeny reaches.

Paleovirology. As mentioned previously, our search for the viral S segment uncovered a nucleoprotein-like paleovirus in the genomes of many species of *Drosophila*, including members of each of the two major subgenera, *Sophophora* and *Drosophila*. The majority of these N-like paleoviruses are present in ORFs of the approximate length of the extant phasmavirus N gene, and the phylogenetic relationships between these NIRVs and the phasmavirus N amino acid sequences are shown in Fig. 5. Phasmavirus NIRVs are almost entirely absent from species of *Drosophila* belonging to the subgroup melanogaster of the subgenus *Drosophila*. The exceptions are *Drosophila yakuba*, which does possess one complete N-like paleovirus and a short fragment of the second, and *Drosophila erecta*, which possesses the short fragment of the second but has entirely lost the larger NIRV. In several species of *Drosophila* for which transcriptomes are available, expressed copies of the N-like NIRVs are present, which suggests that the viral N gene may have been coopted by the host following integration. To test for evidence of cooption, we performed tests for detection of natural selection by using the FUBAR method of the HyPhy package on the Datamonkey Web server (39–41), and we found that the N-like NIRVs evolved under strong purifying selection in the *Drosophila* genome. We compared the results of this analysis on a site-by-site basis between each NIRV clade and the exogenous phasmavirus nucleoprotein to better understand how the two clades have diverged since duplication, and how their sequence evolution compares to that of the viral gene (see Fig. S2 in the supplemental material). Overwhelmingly, purifying selection has been the driving force shaping evolution in the NIRV family, and where divergence is present between the two NIRV clades, it can be attributed to relaxation of purifying selection rather than to diversifying selection; no significant evidence for pervasive diversifying selection was detected across the gene family. However, we did detect support for episodic diversifying selection at a few codon positions by using the MEME method (42).

Comparison of the genomic DNA sequences (~70-kb window) flanking the NIRVs revealed that the integration sites are homologous throughout the genus *Drosophila* (Fig. 5), indicating that this bunyavirid association predates the divergence of the host genus. The minimum divergence date estimate for the genus *Drosophila* is 20 MY, based on molecular clock and fossil evidence (43–45). Since some genetic divergence clearly took place in the ancestral NIRVs between the time of duplication and the *Drosophila* species radiation, we used Bayesian phylogenetic inference

FIG 5 An ancient, orthologous gene family in *Drosophila* is derived from a phasmavirus nucleoprotein gene. (A) A maximum likelihood phylogram shows the evolutionary relationships between the two clades of phasmavirus nucleoprotein (N)-like paleoviruses and exogenous phasmavirus N amino acid sequences. Branches are labeled with filled and hollow circles to indicate approximate likelihood ratio test scores of >0.95 and >0.8 , respectively. Tips are labeled with species names. (B) Microsynteny map of the genomic features flanking the N-like paleoviruses of each species. Green arrows indicate the positions of clade 1 paleoviruses, and blue arrows indicate the positions of clade 2 paleoviruses. Arrow direction indicates gene directionality. The Greek letter ψ indicates a pseudogenized paleovirus. Pseudogenized paleoviruses shorter than 100 amino acids were excluded from the phylogenetic analysis. Regions of the genome maps that remain unshaded are those where the assembly is incomplete. Genes of host origin used to establish positional homology were as follows: DPEP, dipeptidase; CCNA, cyclin A; IAP-1, inhibitor of apoptosis 1; GnT-IV, *N*-acetylglucosaminyltransferase IV.

methods to assign an approximate date to that duplication event and to acquire a minimum age estimate for phasmaviruses. We used BEAST (33) to apply a relaxed-log-normal clock model and calibrated the age of each NIRV clade to 20 MY to coincide with the minimum age estimate for the genus *Drosophila*. After running this analysis for 10 million Markov chain Monte Carlo (MCMC) steps, BEAST estimated the minimum age of the two N-like NIRV clades to be 42 MY (95% highest posterior density [HPD], 32 to 52 MY) (see Fig. S3 in the supplemental material). This estimate does not take into account the time elapsed between integration of the original N-like NIRV and its duplication, and as such, it represents an approximate minimum age window for phasmaviruses. The identification of phasmavirus-like paleovi-

rus in a wide range of insect hosts (Fig. 6 and Table 3; accession numbers are provided in Table S2) is consistent with this estimate of the antiquity of insect-infecting phasmaviruses. Many of the paleoviruses we have identified are present in the transcriptomes of insects for which genomes have not yet been sequenced. As a result, it is not yet clear in some cases whether these sequences are derived from viruses or from paleoviruses.

DISCUSSION

In recent years, the discovery of several divergent bunyavirids has stretched the boundaries of the impressive diversity among the members of the *Bunyaviridae*. Phasmaviruses occupy a vital position in the effort to more completely understand bunyavirid di-

FIG 6 Phylogenetic relationships between phasmavirus genes and phasmavirus-like paleoviruses in insects. Maximum likelihood phylograms show predicted relationships between extant phasmavirus genes and paleoviruses identified by significant tBLASTn sequence matches in insect genomes and transcriptomes. Individual trees are labeled with gene names or abbreviations. Branches are labeled with aLRT support values of ≥ 0.70 . Branch tips are labeled with host species names, and GenBank accession numbers are provided in Table S2 in the supplemental material. Taxa for which virus-like sequences corresponding to each of the 3 genome segments were identified in transcriptomes are marked with pink asterisks. The nucleoprotein-like paleoviruses of *Drosophila* are represented by a subset of taxa in the interest of conserving space. See Fig. 5 for a complete phylogenetic analysis of this paleovirus family.

versity and evolution, as our phylogenetic analysis indicates that they constitute the most divergent group of bunyavirids to be described since the type specimen of the genus *Hantavirus* was isolated in 1976 (46). The genetic divergences between Beringian and Iqaluit populations of both virus and host are consistent with previous proposals of multiple glacial refugia for the host (47). The tundra-invading *Chaoborus* species that is a host to NOMV has a newly established and nearly ubiquitous distribution throughout the study area, which encompasses the entire southern half of the Seward Peninsula. While KIGV may be a stable part of both tundra and boreal pond communities, NOMV and *Chaoborus* cf. *flavicans* are newcomers to the tundra, and the effects of novel community interactions are unknown. *Chaoborus* has a worldwide habitat distribution, and we expect that the Seward Peninsula is not the only region where the tundra-boreal ecotone is experiencing host expansions. It is therefore important to continue to characterize phasmavirus-host associations in arctic and subarctic environments in both narrow and broad scopes by investigating the prevalence and mode of transmission of NOMV in *Chaoborus* cf. *flavicans*, as well as the extent of phasmavirus infection in additional species of the genus *Chaoborus*.

Our results suggest that KIGV has been transmitted vertically in North American *C. trivittatus* populations for at least thousands of years. We found a remarkably high viral prevalence and evidence for long-term coevolution, which together seem more reflective of an endosymbiont than a virus. We also present paleovirological evidence that the relationship between phasmaviruses and nonhematophagous insects reaches back tens of millions of years. Findings such as these tempt speculation that a symbiosis-like lifestyle could have played a role in the evolution of diverse virus families, including *Bunyaviridae*. Indeed, work in other vertically transmitted virus systems, e.g., the sigma viruses of *Drosophila* (9) and the totiviruses of fungi (48), has led authors to similar conclusions, yet it remains unclear whether cases such as these might represent a previously obscured rule of virus evolution or merely the exceptions to it. What is the evolutionary significance of divergent bunyavirid groups that seem to infect only insect hosts, e.g., Gouleako virus (49) and the Herbert virus clade (50), while other genera, e.g., *Hantavirus*, have made the jump to mammals? In a recent review of hantavirus evolution, Plyusnin and Sironen (14) proposed that the relationship between hantaviruses and their mammalian hosts may have emerged from an

TABLE 3 Significant tBLASTn matches to phasmavirus genes in insect sequence databases

Host	Host common name	Host order	Gene(s) present	DNA ^a
<i>Dendroctonus ponderosae</i>	Mountain pine beetle	Coleoptera	G, N	Y
<i>Chaoborus trivittatus</i>	Phantom midge	Diptera	RdRp, G, N	N
<i>Chaoborus</i> cf. <i>flavicans</i>	Phantom midge	Diptera	RdRp, G, N	N
<i>Drosophila</i> spp.	Fruit fly	Diptera	N	Y
<i>Aedes albopictus</i>	Tiger mosquito	Diptera	G, N	N
<i>Anopheles sinensis</i>	Malaria mosquito	Diptera	RdRp, G, N	N
<i>Anopheles epiroticus</i>	Malaria mosquito	Diptera	G, N	Y
<i>Anopheles minimus</i>	Malaria mosquito	Diptera	G, N	Y
<i>Anopheles stephensi</i>	Malaria mosquito	Diptera	N	Y
<i>Anopheles dirus</i>	Malaria mosquito	Diptera	G	Y
<i>Aedes aegypti</i>	Yellow fever mosquito	Diptera	G	Y
<i>Lutzomyia longipalpis</i>	Sand fly	Diptera	G, N	Y
<i>Teleopsis dalmanni</i>	Stalk-eyed fly	Diptera	N	—
<i>Teleopsis whitei</i>	Stalk-eyed fly	Diptera	N	—
<i>Culicoides sonorensis</i>	Biting midge	Diptera	G	—
<i>Rhodnius prolixus</i>	Assassin bug	Hemiptera	G	Y
<i>Pachypsylla venusta</i>	Petiole gall psyllid	Hemiptera	RdRp, N	—
<i>Diaphorina citri</i>	Asian citrus psyllid	Hemiptera	N	Y
<i>Osmia cornuta</i>	Builder bee	Hymenoptera	RdRp, G, N	—
<i>Megachile rotundata</i>	Leafcutter bee	Hymenoptera	N	Y
<i>Bombus terrestris</i>	Bumble bee	Hymenoptera	N	Y
<i>Exoneura robusta</i>	Allodapine bee	Hymenoptera	N	—
<i>Sceliphron caementarium</i>	Mud dauber wasp	Hymenoptera	RdRp, G, N	—
<i>Leptopilina heterotoma</i>	Parasitoid wasp	Hymenoptera	N	—
<i>Plutella xylostella</i>	Diamondback moth	Lepidoptera	N	Y
<i>Bombyx mori</i>	Silk moth	Lepidoptera	RdRp, G	Y
<i>Ostrinia furnacalis</i>	Asian corn borer	Lepidoptera	RdRp, G, N	—
<i>Mengenilla moldrzyki</i>	Twisted-wing parasite	Strepsiptera	N	Y

^a Y, the organism's genome has been sequenced, and the virus-like sequence is present; N, the genome has been sequenced, but the virus-like sequence is absent; —, the genome has not been sequenced.

ancient, insect-borne bunyavirid lineage; that is, a “prebunyavirus” made the jump from insects to mammals, and the extant members of this lineage comprise the genus *Hantavirus*. We have presented evidence that phasmaviruses are or have been associated with a range of insect hosts whose lifestyles range from plant associated to vertebrate associated and seem to be members of an ancient bunyavirid lineage grouping with *Hantavirus*, *Orthobunyavirus*, and *Tospovirus*. While it is perhaps too early to suggest that phasmaviruses provide clear support for this hypothesis, our results do coincide neatly with such a scenario.

ACKNOWLEDGMENTS

We thank Alexey Kotov and Andrew Medeiros for assistance in the field and A. Medeiros and Art Borkent for collecting and identifying specimens of *Chaoborus* from Baffin Island and British Columbia.

This work was supported by the National Science Foundation (grant ARC 1023334).

REFERENCES

- Jorgenson MT, Shur YL, Pullman ER. 2006. Abrupt increase in permafrost degradation in Arctic Alaska. *Geophys. Res. Lett.* 33:L02503. <http://dx.doi.org/10.1029/2005GL024960>.
- Prowse TD, Wrona FJ, Reist JD, Gibson JJ, Hobbie JE, Levesque LM, Vincent WF. 2006. Climate change effects on hydroecology of arctic freshwater ecosystems. *Ambio* 35:347–358. [http://dx.doi.org/10.1579/0044-7447\(2006\)35\[347:CCEOHO\]2.0.CO;2](http://dx.doi.org/10.1579/0044-7447(2006)35[347:CCEOHO]2.0.CO;2).
- Ruhland KM, Paterson AM, Keller W, Michelutti N, Smol JP. 2013. Global warming triggers the loss of a key Arctic refugium. *Proc. Biol. Sci.* 280:20131887. <http://dx.doi.org/10.1098/rspb.2013.1887>.
- Elliott RM. 2009. Bunyaviruses and climate change. *Clin. Microbiol. Infect.* 15:510–517. <http://dx.doi.org/10.1111/j.1469-0691.2009.02849.x>.
- Maltezos HC, Papa A. 2010. Crimean-Congo hemorrhagic fever: risk for emergence of new endemic foci in Europe? *Travel Med. Infect. Dis.* 8:139–143. <http://dx.doi.org/10.1016/j.tmaid.2010.04.008>.
- Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PP, Baylis M. 2005. Climate change and the recent emergence of bluetongue in Europe. *Nat. Rev. Microbiol.* 3:171–181. <http://dx.doi.org/10.1038/nrmicro1090>.
- Cook S, Chung BY, Bass D, Moureau G, Tang S, McAlister E, Culverwell CL, Glucksman E, Wang H, Brown TD, Gould EA, Harbach RE, de Lamballerie X, Firth AE. 2013. Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* 8:e80720. <http://dx.doi.org/10.1371/journal.pone.0080720>.
- Jaenike J. 2012. Population genetics of beneficial heritable symbionts. *Trends Ecol. Evol.* 27:226–232. <http://dx.doi.org/10.1016/j.tree.2011.10.005>.
- Longdon B, Jiggins FM. 2012. Vertically transmitted viral endosymbionts of insects: do sigma viruses walk alone? *Proc. Biol. Sci.* 279:3889–3898. <http://dx.doi.org/10.1098/rspb.2012.1208>.
- Dupuis DD, Svensson JE, Taylor DJ. 2008. The cryptic origins of environment-indicating phantom midges (*Chaoborus*). *Limnol. Oceanogr.* 53:236–243. <http://dx.doi.org/10.4319/lo.2008.53.1.0236>.
- Grimaldi DA, Engel MS. 2005. *Evolution of the insects*. Cambridge University Press, Cambridge, United Kingdom.
- Plyusnin A, Elliott RM. 2011. *Bunyaviridae: molecular and cellular biology*. Caister Academic Press, Norfolk, United Kingdom.
- Antic D, Kang CY, Spik K, Schmaljohn C, Vapalahti O, Vaheri A. 1992. Comparison of the deduced gene products of the L, M and S genome segments of hantaviruses. *Virus Res.* 24:35–46. [http://dx.doi.org/10.1016/0168-1702\(92\)90029-9](http://dx.doi.org/10.1016/0168-1702(92)90029-9).
- Plyusnin A, Sironen T. 21 January 2014. Evolution of hantaviruses: co-speciation with reservoir hosts for more than 100MYR. *Virus Res.* <http://dx.doi.org/10.1016/j.virusres.2014.01.008>.
- Plyusnin A, Vapalahti O, Lankinen H, Lehtvaslaihio H, Apekina N, Myasnikov Y, Kallio-Kokko H, Henttonen H, Lundkvist A, Brummer-

- Korvenkontio M. 1994. Tula virus: a newly detected hantavirus carried by European common voles. *J. Virol.* 68:7833–7839.
16. Spiropoulou CF, Morzunov S, Feldmann H, Sanchez A, Peters CJ, Nichol ST. 1994. Genome structure and variability of a virus causing hantavirus pulmonary syndrome. *Virology* 200:715–723. <http://dx.doi.org/10.1006/viro.1994.1235>.
 17. Xiao SY, Leduc JW, Chu YK, Schmaljohn CS. 1994. Phylogenetic analyses of virus isolates in the genus Hantavirus, family Bunyaviridae. *Virology* 198:205–217. <http://dx.doi.org/10.1006/viro.1994.1023>.
 18. Ramsden C, Holmes EC, Charleston MA. 2009. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol. Biol. Evol.* 26:143–153. <http://dx.doi.org/10.1093/molbev/msn234>.
 19. Patel MR, Emerman M, Malik HS. 2011. Paleovirology—ghosts and gifts of viruses past. *Curr. Opin. Virol.* 1:304–309. <http://dx.doi.org/10.1016/j.coviro.2011.06.007>.
 20. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, Gojohori T, Coffin JM, Tomonaga K. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* 463:84–87. <http://dx.doi.org/10.1038/nature08695>.
 21. Katourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLoS Genet.* 6:e1001191. <http://dx.doi.org/10.1371/journal.pgen.1001191>.
 22. Taylor DJ, Bruenn J. 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC Biol.* 7:88. <http://dx.doi.org/10.1186/1741-7007-7-88>.
 23. Taylor DJ, Leach RW, Bruenn J. 2010. Filoviruses are ancient and integrated into mammalian genomes. *BMC Evol. Biol.* 10:193. <http://dx.doi.org/10.1186/1471-2148-10-193>.
 24. Holmes EC. 2003. Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* 77:3893–3897. <http://dx.doi.org/10.1128/JVI.77.7.3893-3897.2003>.
 25. Roggero P, Ciuffo M, Vaira AM, Milne RG. 1998. Rapid purification of tospovirus nucleocapsids for antibody production and RNA analysis, p 25–28. *In* Peters D, Goldbach R (ed), Recent progress in tospovirus and thrips research. Wageningen University, Wageningen, The Netherlands.
 26. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066. <http://dx.doi.org/10.1093/nar/gkf436>.
 27. Chang JM, Tommaso PD, Notredame C. 2014. TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.* 31:1625–1637. <http://dx.doi.org/10.1093/molbev/msu117>.
 28. Abascal J, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105. <http://dx.doi.org/10.1093/bioinformatics/bti263>.
 29. Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27:221–224. <http://dx.doi.org/10.1093/molbev/msp259>.
 30. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p 1–8. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA. IEEE, New York, NY.
 31. Rambaut A. 2007. FigTree, v1.2.3 ed. <http://tree.bio.ed.ac.uk/software/figtree/>.
 32. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785–786. <http://dx.doi.org/10.1038/nmeth.1701>.
 33. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973. <http://dx.doi.org/10.1093/molbev/mss075>.
 34. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. <http://dx.doi.org/10.1038/nmeth.2109>.
 35. Barr JN, Elliott RM, Dunn EF, Wertz GW. 2003. Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication. *Virology* 311:326–338. [http://dx.doi.org/10.1016/S0042-6822\(03\)00130-2](http://dx.doi.org/10.1016/S0042-6822(03)00130-2).
 36. Gualiard N, Billecocq A, Flick R, Bouloy M. 2006. Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis. *Virology* 351:170–179. <http://dx.doi.org/10.1016/j.virol.2006.03.018>.
 37. Bruenn JA. 2003. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* 31:1821–1829. <http://dx.doi.org/10.1093/nar/kgk277>.
 38. Muller R, Poch O, Delarue M, Bishop DH, Bouloy M. 1994. Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J. Gen. Virol.* 75:1345–1352. <http://dx.doi.org/10.1099/0022-1317-75-6-1345>.
 39. Delpont W, Poon AFY, Frost SDW, Pond SLK. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457. <http://dx.doi.org/10.1093/bioinformatics/btq429>.
 40. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol. Biol. Evol.* 30:1196–1205. <http://dx.doi.org/10.1093/molbev/mst030>.
 41. Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. <http://dx.doi.org/10.1093/bioinformatics/bti079>.
 42. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764. <http://dx.doi.org/10.1371/journal.pgen.1002764>.
 43. Obbard DJ, MacLennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol. Biol. Evol.* 29:3459–3473. <http://dx.doi.org/10.1093/molbev/mss150>.
 44. Grimaldi D. 1987. Amber fossil Drosophilidae (Diptera), with particular reference to the Hispaniolan taxa. *Am. Mus. Novitates* 2880:1–23.
 45. Throckmorton LH. 1975. The phylogeny, ecology and geography of *Drosophila*, p 421–469. *In* King RC (ed), *Handbook of genetics*. Plenum, New York, NY.
 46. Lee HW, Lee PW. 1976. Korean hemorrhagic fever. I. Demonstration of causative antigen and antibodies. *Korean J. Intern. Med.* 19:371–383.
 47. Borkent A. 1979. Systematics and bionomics of the species of the subgenus *Schadonophasma* Dyar and Shannon (Chaoboridae, Diptera). *Quaest. Entomol.* 15:122–255.
 48. Bruenn J. 2012. Genes from double-stranded RNA viruses in the nuclear genomes of fungi, p 71–83. *In* Witzany G (ed), *Biocommunication of fungi*. Springer, Dordrecht, Netherlands.
 49. Marklewitz M, Handrick S, Grasse W, Kurth A, Lukashev A, Drosten C, Ellerbrok H, Leendertz FH, Pauli G, Junglen S. 2011. Gouleako virus isolated from West African mosquitoes constitutes a proposed novel genus in the family Bunyaviridae. *J. Virol.* 85:9227–9234. <http://dx.doi.org/10.1128/JVI.00230-11>.
 50. Marklewitz M, Zirkel F, Rwego IB, Heidemann H, Trippner P, Kurth A, Kallies R, Briese T, Lipkin WI, Drosten C, Gillespie TR, Junglen S. 2013. Discovery of a unique novel clade of mosquito-associated bunyaviruses. *J. Virol.* 87:12850–12865. <http://dx.doi.org/10.1128/JVI.01862-13>.
 51. Schmaljohn CS. 1990. Nucleotide sequence of the L genome segment of Hantaan virus. *Nucleic Acids Res.* 18:6728. <http://dx.doi.org/10.1093/nar/18.22.6728>.
 52. Schmaljohn CS, Jennings GB, Hay J, Dalrymple JM. 1986. Coding strategy of the S genome segment of Hantaan virus. *Virology* 155:633–643. [http://dx.doi.org/10.1016/0042-6822\(86\)90223-0](http://dx.doi.org/10.1016/0042-6822(86)90223-0).
 53. Schmaljohn CS, Schmaljohn AL, Dalrymple JM. 1987. Hantaan virus M RNA: coding strategy, nucleotide sequence, and gene order. *Virology* 157:31–39. [http://dx.doi.org/10.1016/0042-6822\(87\)90310-2](http://dx.doi.org/10.1016/0042-6822(87)90310-2).
 54. Elliott RM. 1989. Nucleotide sequence analysis of the large (L) genomic RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. *Virology* 173:426–436. [http://dx.doi.org/10.1016/0042-6822\(89\)90555-2](http://dx.doi.org/10.1016/0042-6822(89)90555-2).
 55. Elliott RM. 1989. Nucleotide sequence analysis of the small (S) RNA segment of Bunyamwera virus, the prototype of the family Bunyaviridae. *J. Gen. Virol.* 70:1281–1285. <http://dx.doi.org/10.1099/0022-1317-70-5-1281>.
 56. Lees JF, Pringle CR, Elliott RM. 1986. Nucleotide sequence of the Bunyamwera virus M RNA segment: conservation of structural features in the Bunyavirus glycoprotein gene product. *Virology* 148:1–14. [http://dx.doi.org/10.1016/0042-6822\(86\)90398-3](http://dx.doi.org/10.1016/0042-6822(86)90398-3).
 57. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST. 2007. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *J. Virol.* 81:2805–2816. <http://dx.doi.org/10.1128/JVI.02095-06>.
 58. de Haan P, Kormelink R, de Oliveira Resende R, van Poelwijk F, Peters D, Goldbach R. 1991. Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. *J. Gen. Virol.* 72:2207–2216. <http://dx.doi.org/10.1099/0022-1317-72-9-2207>.

59. de Haan P, Wagemakers L, Peters D, Goldbach R. 1990. The S-RNA segment of tomato spotted wilt virus has an ambisense character. *J. Gen. Virol.* 71:1001–1007. <http://dx.doi.org/10.1099/0022-1317-71-5-1001>.
60. Kormelink R, de Haan P, Meurs C, Peters D, Goldbach R. 1992. The nucleotide sequence of the M RNA segment of tomato spotted wilt virus, a bunyavirus with two ambisense RNA segments. *J. Gen. Virol.* 73:2795–2804. <http://dx.doi.org/10.1099/0022-1317-73-11-2795>.
61. Bridgen A, Dalrymple DA, Elliott RM. 2002. Dugbe nairovirus S segment: correction of published sequence and comparison of five isolates. *Virology* 294:364–371. <http://dx.doi.org/10.1006/viro.2001.1324>.
62. Marriott AC, el-Ghorr AA, Nuttall PA. 1992. Dugbe nairovirus M RNA: nucleotide sequence and coding strategy. *Virology* 190:606–615. [http://dx.doi.org/10.1016/0042-6822\(92\)90898-Y](http://dx.doi.org/10.1016/0042-6822(92)90898-Y).
63. Marriott AC, Nuttall PA. 1996. Large RNA segment of Dugbe nairovirus encodes the putative RNA polymerase. *J. Gen. Virol.* 77:1775–1780. <http://dx.doi.org/10.1099/0022-1317-77-8-1775>.