Discovery of Novel Crustacean and Cephalopod Flaviviruses: Insights into the Evolution and Circulation of Flaviviruses between Marine Invertebrate and Vertebrate Hosts

Rhys Parry,^a Sassan Asgari^a

^aAustralian Infectious Disease Research Centre, School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia

ABSTRACT Most described flaviviruses (family *Flaviviridae*) are disease-causing pathogens of vertebrates maintained in zoonotic cycles between mosquitoes or ticks and vertebrate hosts. Poor sampling of flaviviruses outside vector-borne flaviviruses such as Zika virus and dengue virus has presented a narrow understanding of flavivirus diversity and evolution. In this study, we discovered three crustacean flaviviruses (Gammarus chevreuxi flavivirus, Gammarus pulex flavivirus, and Crangon crangon flavivirus) and two cephalopod flaviviruses (Southern Pygmy squid flavivirus and Firefly squid flavivirus). Bayesian and maximum likelihood phylogenetic methods demonstrate that crustacean flaviviruses form a well-supported clade and share a more closely related ancestor with terrestrial vector-borne flaviviruses than with classical insect-specific flaviviruses. In addition, we identify variants of Wenzhou shark flavivirus in multiple gazami crab (*Portunus trituberculatus*) populations, with active replication supported by evidence of an active RNA interference response. This suggests that Wenzhou shark flavivirus moves horizontally between sharks and gazami crabs in ocean ecosystems. Analyses of the mono- and dinucleotide composition of marine flaviviruses compared to that of flaviviruses with known host status suggest that some marine flaviviruses share a nucleotide bias similar to that of vector-borne flaviviruses. Furthermore, we identify crustacean flavivirus endogenous viral elements that are closely related to elements of terrestrial vector-borne flaviviruses. Taken together, these data provide evidence of flaviviruses circulating between marine vertebrates and invertebrates, expand our understanding of flavivirus host range, and offer potential insights into the evolution and emergence of terrestrial vector-borne flaviviruses.

IMPORTANCE Some flaviviruses are known to cause disease in vertebrates and are typically transmitted by blood-feeding arthropods such as ticks and mosquitoes. While an ever-increasing number of insect-specific flaviviruses have been described, we have a narrow understanding of flavivirus incidence and evolution. To expand this understanding, we discovered a number of novel flaviviruses that infect a range of crustaceans and cephalopod hosts. Phylogenetic analyses of these novel marine flaviviruses suggest that crustacean flaviviruses share a close ancestor to all terrestrial vector-borne flaviviruses, and squid flaviviruses are the most divergent of all known flaviviruses to date. Additionally, our results indicate horizontal transmission of a marine flavivirus between crabs and sharks. Taken together, these data suggest that flaviviruses move horizontally between invertebrates and vertebrates in ocean ecosystems. This study demonstrates that flavivirus invertebrate-vertebrate host associations have arisen in flaviviruses at least twice and may potentially provide insights into the emergence or origin of terrestrial vector-borne flaviviruses.

KEYWORDS RNA interference, cephalopod, Crustacea, evolutionary biology, flavivirus

Citation Parry R, Asgari S. 2019. Discovery of novel crustacean and cephalopod flaviviruses: insights into the evolution and circulation of flaviviruses between marine invertebrate and vertebrate hosts. *J Virol* 93:e00432-19. <https://doi.org/10.1128/JVI.00432-19>.

Editor Bryan R. G. Williams, Hudson Institute of Medical Research

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Sassan Asgari, s.asgari@uq.edu.au.

Received 12 March 2019

Accepted 23 April 2019

Accepted manuscript posted online 8 May 2019

Published 28 June 2019

Classical flaviviruses (family *Flaviviridae*) are monopartite, single-stranded, positive-sense RNA viruses between 9 and 13 kb in size that encode a large polyprotein (1, 2). For maturation of flavivirus particles, the translated polyprotein requires posttranslational cleavage into conserved structural proteins (capsid protein [C], premembrane protein [prM], and envelope protein [E]), and also nonstructural (NS) proteins NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5, by the host proteases (3) and virus-encoded chymotrypsin-like protease (4, 5). The flavivirus type species *Yellow fever virus* (YFV), along with dengue virus (DENV) and Zika virus (ZIKV), is vectored by the *Aedes aegypti* mosquito, and these species are collectively responsible for millions of global infections per year that result in a range of disease outcomes in humans (6, 7).

As mosquitoes and ticks are potential reservoirs of vertebrate-infecting flaviviruses (VIFs), biosecurity and surveillance programs have been established to characterize the ecology and diversity of potential VIFs in these hosts (8, 9). As a result of these efforts, a number of insect-specific flaviviruses (ISFs), so named for an inability to replicate within vertebrate cells (reviewed by Blitvich and Firth [10]), have been fully sequenced and characterized. As the number of ISFs identified in the literature expands, there is evidence that ISFs exist both in a classical ISF (cISF) clade, in which all members form a monophyletic lineage, and a dual-host-associated ISF (dISF) clade. These dISFs are similar in phenotype to cISFs and show an inability to replicate in mammalian cells, but unlike cISF, dISFs are paraphyletic and phylogenetically related to arboviruses (10). This suggests that vertebrate-infecting and insect-specific flaviviruses may have arisen and been lost multiple times throughout the evolutionary history of flaviviruses. In addition, the International Committee on Taxonomy of Viruses also recognizes a number of flaviviruses isolated from vertebrates, such as rodents or birds, that are classified as being flaviviruses with no known vector (NKV) (reviewed by Blitvich and Firth [11]). These viruses are genetically related to vector-borne flaviviruses, with some experimentally demonstrated to replicate in both mosquito and vertebrate cells (12, 13). These are presumed to be vectored by as yet unidentified invertebrate hosts. Not all NKV viruses replicate in mosquito cells; for example, Rio Bravo virus (RBV), Montana myotis leukoencephalitis virus (MMLV), Apoi virus (APOV), and Modoc virus (MODV) do not (14, 15). Transmission of these NKV flaviviruses is assumed to be horizontally, through salivary glands or aerosol droplets between mammals. The aforementioned NKV, dISFs, and VIFs are all grouped within one well-supported monophyletic lineage, however; one notable exception to this is Tamana bat virus (TABV), isolated from the Parnell's mustached bat, *Pteronotus parnellii* (16). Tamana bat virus is not phylogenetically related to other NKV flaviviruses, has never been isolated from other vertebrate or invertebrate species, and groups basally to all flaviviruses. It is sometimes referred to as a vertebrate-only flavivirus.

Metatranscriptomic analyses of invertebrate and vertebrate pools, not typically screened in biosurveillance programs, have resulted in uncovering a remarkable diversity of viruses infecting eukaryotic species (17, 18). As a result of these efforts, flaviviruses have been identified outside terrestrial vertebrates and invertebrates for the first time, in two groups of fish. The first of these marine vertebrate flaviviruses, Cyclopterus lumpus virus (CLuV), was identified in tissues of diseased lumpfish (*Cyclopterus lumpus*) (19). The second marine vertebrate flavivirus is Wenzhou shark flavivirus, identified in a metagenomic analysis of the Pacific spadenose shark, *Scoliodon macrorhynchos* (17). Spadenose sharks are members of the cartilaginous fish group, and while it is currently unknown if Wenzhou shark flavivirus is responsible for any pathology within the shark host, the flavivirus was reported to be abundant in all tissues (17). Phylogenetic analysis based on the polyprotein of these marine vertebrate flaviviruses suggests that they form a basal and genetically divergent group of flaviviruses along with TABV (20). Additionally, flavivirus fragments have been identified in a metatranscriptomic analysis of the Eastern red scorpionfish, *Scorpaena jacksoniensis* (20), and the sea spider (*Endeis spinosa*) (21).

Attempts to evaluate the temporal evolution of flaviviruses using time to most recent common ancestor (tMRCA) analyses have remained controversial (22). In addi-

tion, poor sampling of flaviviruses outside vector-borne flaviviruses has hindered our potential for exploring the origins or emergence of terrestrial vector-borne flaviviruses. Flaviviruses that have recently been discovered infecting two large groups of fish present an avenue to explore marine flavivirus evolution and diversity.

Previously, it has been suggested that alphaviruses, single-stranded RNA arboviruses belonging to the family *Togaviridae*, are likely to have a marine origin (23). Most alphaviruses have a similar vector-borne host range, with mosquitoes as the principle vectors. The hypothesis of emergence of alphaviruses from a marine origin is supported both by phylogenetic evidence of the basal origin of alphaviruses infecting marine mammals, and also by the fact that Southern elephant seal virus (SESV) and salmon pancreatic disease virus (SPDV) have also been isolated from blood- and mucus-eating *Lepidohthirus* ectoparasites (24). The evidence of flaviviruses infecting sharks suggests that flaviviruses may have circulated long before the emergence of hematophagous mosquitoes, which are part of the infraorder Culicomorpha in the lower Diptera, which emerged ~220 million years ago (25). In comparison, vertebrate and arthropod lineages diverged between 573 and 656 million years ago (26). Previous approaches in mining sequencing data have yielded insights into the evolutionary history and host jumps of rhabdoviruses (27) and parvoviruses (28). We hypothesized that there may be potential dual invertebrate-vertebrate flaviviruses missing from the virus records. Therefore, we set out to explore and supplement the flavivirus record with additional metazoans, with a view to gaining insight into the evolution and emergence of flaviviruses.

RESULTS

Discovery and annotation of novel marine invertebrate flaviviruses. To discover divergent flaviviruses, we queried the assembled transcriptomes deposited from all Animalia in the Transcriptome Shotgun Assembly Sequence Database (TSA) hosted by the National Center for Biotechnology Information (NCBI) using the tblastn algorithm flavivirus polyprotein sequences (29). A number of 10- to 12-kb transcripts were identified as having 29% to 27% identity over 28% to 59% of the query length of flavivirus polyproteins, suggesting potential divergent hits. We excluded fragments that were unlikely to form complete single-strand flaviviruses and multisegmented members of the unranked Jingmenvirus taxon (30). Additional viruses were identified in unassembled cephalopod and crustacean RNA sequencing (RNA-Seq) data and subsequently *de novo* assembled. Genome features and statistics are summarized in Table 1, and metadata of the samples used are available in Table S1.

Six flaviviruses were identified in this study; five novel and one strain of a previously published virus. One 10,669-nucleotide (nt) contig from the transcriptome of the gazami crab or Japanese blue crab, *Portunus trituberculatus*, was identified with 95% nucleotide sequence identity to Wenzhou shark virus. This transcriptome was produced from eyestalk, gill, heart, hepatopancreas, and muscle tissue from 90 healthy swimming crabs from Shandong, China (31). Reanalysis of raw data of this RNA-Seq library suggested that the flavivirus had an average coverage of 33.50×. Further analysis of the incidence of Wenzhou shark flaviviruses in other *P. trituberculatus* data sets is covered in later sections of the results.

Three ~10- to 11-kb putative flaviviruses were identified from wild-caught malacostracan crustaceans. A flavivirus was identified in the brown shrimp, *Crangon crangon*, from midgut samples originating from Weser estuary, Germany (32). Prediction of the complete polyprotein sequence of Crangon crangon flavivirus (CcFV) suggested that the closest known flavivirus in amino acid identity is the arbovirus Usutu virus (33). Two flaviviruses were identified in amphipod species. Gammarus chevreuxi flavivirus (GcFV) was identified in transcriptomes from two publications on *Gammarus chevreuxi* in both embryonic and adult samples originating from the Plym estuary, Plymouth, United Kingdom (34, 35) and in Gammarus pulex flavivirus (GpFV) identified from a male *Gammarus pulex* wild-caught from the Bourbre River, France (36). The polyproteins of these viruses were more closely related in amino acid identity to those of the

TABLE 1 Genomic features and metadata of novel marine flaviviruses identified in this study

Name of virus	Host (class; order; genus and species)	GenBank accession no.	Genome length (nt); polypeptide length (aa)	Coverage (x)	Closest blastp hit(s) (GenBank accession no.), aa identity (%), query coverage (%) ^a	GC %	Reference(s)
Wenzhou shark flavivirus (<i>P. trituberculatus</i> strain)	Malacostraca; Decapoda; <i>Portunus trituberculatus</i>	MK473876	10,690; 3,420	~33.50	Wenzhou shark flavivirus (AVM87250.1), 98, 100	49.1	31
Southern Pygmy squid flavivirus (SpsFV)	Cephalopoda; Idiosepiida; <i>Idiosepius notoides</i>	MK473875	12,567; 3,843	~834.54	Tamana bat virus (NP_658908.1), 24, 64; Firefly squid flavivirus, 34, 68	46.5	40
Firefly squid flavivirus (FsFV)	Cephalopoda; Teuthida; <i>Watasenia scintillans</i>	MK473880 (fragment 1); MK473879 (fragment 2)	7,496; 2,910	~10.5	Cyclopterus lumpus virus (ATY35190.1), 30, 53; Southern pygmy squid flavivirus, 34, 80	52.3	39
Crangon crangon flavivirus (CcFV)	Malacostraca; Decapoda; <i>Crangon crangon</i>	MK473878	11,434; 3,626	~164.89	Usutu virus (AID60242.1), 33, 80; Gammarus chevreuxi flavivirus, 35, 85	40.9	32
Gammarus chevreuxi flavivirus (GcFV)	Malacostraca; Amphipoda; <i>Gammarus chevreuxi</i>	MK473877	11,231; 3,558	~40	New Mapoon virus (YP_009328360.1), 37, 79; Crangon crangon flavivirus, 35, 84	52.1	34, 35
Gammarus pulex flavivirus	Malacostraca; Amphipoda; <i>Gammarus pulex</i>	MK473881	11,883; 3,767	~22	Apoi virus (NP_620045.1), 39, 46; Gammarus chevreuxi flavivirus, 35, 85	53.5	36

^aClosest blastp hit in nonredundant NCBI database and closest blastp hit compared to flaviviruses reported in this study.

FIG 1 Genome architecture of novel marine invertebrate flaviviruses compared to that of the flavivirus type species *Yellow fever virus*. Predicted flavivirus Pfam domains are indicated by colored boxes.

mosquito-borne flavivirus New Mapoon virus (37) and also the NKV Apoi flavivirus, the latter of which has been isolated from three different bat species (38).

Flaviviruses were also identified from transcriptomes of two different squid species (class Cephalopoda). Fragments of flavivirus-like contigs from the TSA record of the firefly squid, *Watasenia scintillans* (39), were identified. These contigs originated from the RNA-Seq data of a single arm tip sample, which was downloaded and reassembled into two larger 7.4- and 3-kb fragments. The final novel flavivirus identified was a 12,567-nt contig from the southern pygmy squid, *Idiosepius notoides* (40). Both polyproteins from these squid flaviviruses showed higher amino acid identity to those of Tamana bat virus and *Cyclopterus lumpus* virus than to those of the crustacean flaviviruses reported here. We have tentatively named these Firefly squid flavivirus (FsFV) and Southern pygmy squid flavivirus (SpsFV). For clarity within the document, we will refer to the collective grouping of the cephalopod and crustacean flaviviruses as marine invertebrate flaviviruses (MIFs).

With the exception of FsFV, all MIFs appeared to be coding complete. Prediction of transmembrane domains and conserved flavivirus protein domains in the polyprotein of all MIFs suggest remarkable conservation of genome orientation between these putative flaviviruses and the flavivirus type species (Fig. 1). All MIFs were identified as encoding the conserved flavivirus RNA-dependent RNA polymerase NS5 (pfam00972; E value, $\leq 3E-43$) and also the flavivirus DEAD domain of the NS3 helicase (pfam07652; E value, $\leq 6E-20$). All but one were identified as having flavivirus glycoprotein, central and dimerization domains (pfam00869; E value, $\leq 7E-10$), or flavivirus envelope gly-

FIG 2 Translation of a complete polyprotein depends on highly structured programmed -1 ribosomal frameshift in four marine flaviviruses. (A) Genome orientation of discrete open reading frames predicted for marine flaviviruses. (B) Comparison of predicted structured programmed ribosomal frameshift (-1 PRF) regions identified. Slippery heptanucleotide regions are highlighted in pink. Predicted translated protein regions are indicated (yellow), as are the translation frame and the RNA structure dot plot. (C) Predicted minimum free energy structure of the stem-loop structure immediately downstream of the slippery heptanucleotide motif. Nucleotides are colored as base pair probabilities. Green nucleotide pairings indicate a 50% probability, whereas strong complementarity between paired RNAs is indicated in red.

coprotein E, stem/anchor domain (TIGR04240; E value: $\leq 2E-10$) (Fig. 1). Crustacean flaviviruses were all predicted to have a NS1 protein domain (pfam00948; E value, $\leq 7E-39$), FtsJ-like methyltransferases (pfam01728; E value, $6E-15$), and the S7 peptidase domain of the NS3-protease protein (pfam00949; E value, $4E-10$). Remarkably, CcFV was predicted to not only have all of these features, but also similarity to the flavivirus nonstructural protein NS4A domain (PF01350; E value, 0.15) and the immunoglobulin-like domain III of flavivirus envelope glycoprotein (cd12149; E value: $3E-08$). In addition to the conserved protein domain architecture, flaviviruses are known to have conserved N-linked glycosylation sites on the envelope (41, 42) and the NS1 protein (43). The predicted N-linked glycosylation sites were similar among the assembled genomes (Fig. 1).

We propose that the viruses presented here bear all the prototypical hallmarks of conventional flaviviruses and, based on the current species demarcation guides for the genus *Flavivirus* (44), are sufficiently divergent from other known flaviviruses to be considered novel species of the genus.

Translation of the CcFV, GcFV, and GpFV polyprotein is predicted to depend on programmed -1 ribosomal frameshift. Initial prediction of the open reading frames of CcFV, GcFV, and GpFV suggests that the genomes of these flaviviruses have two discrete open reading frames (Fig. 2A) with a stop codon near the C terminus of the NS1 region. If the flaviviruses presented here encoded proteins on two discrete open

reading frames, a number of transmembrane domains would not be produced between the predicted NS1 and the NS2 protein region. A similar genome structure was observed in Cyclopterus lumpus virus (CLuV), in which the production of a complete polyprotein is predicted to depend on a programmed -1 ribosomal frameshift (-1 PRF) on a “slippery” heptanucleotide sequence proximal to the stop codon (19). Programmed -1 ribosomal frameshifting has been documented to be used by many flaviviruses for the synthesis of additional proteins (45–47); however, only in CLuV has a -1 PRF been predicted for the production of a complete polyprotein.

We screened the genomes of CcFV, GcFV, and GpFV for evidence of -1 PRF. Three criteria have been suggested that implicate the presence of a potential -1 PRF (48). First, there is the canonical “slippery” heptanucleotide sequence X₃XXY₃YYZ, where XXX and YYY represent three identical nucleotides, with Y being A/U, and Z being A/C/U, although there are known exceptions within this sequence (48). This slippery heptanucleotide is followed by a 5- to 9-nt spacer region and then by a downstream stem-loop or pseudoknot structure, which serves as a stimulatory RNA signal for slippage of the ribosome. We were able to identify all these features in these flaviviruses (Fig. 2B and C) suggesting that each produces a bona fide -1 PRF. Subsequent slippage of the ribosome at all of these sites would result in the production of a complete and single polyprotein.

Comparisons of the -1 PRF of CluV and the three novel flaviviruses presented here suggests that -1 PRF may be a common strategy of marine flaviviruses to produce a single polyprotein, and that the fairly interesting flavivirus open reading frame (*ffo*) and the other ribosomal slippage proteins produced by VIFs and cISFs may have arisen from a vestige of this protein expression strategy (45, 46).

Putative polyprotein cleavage sites. The polyprotein of flaviviruses becomes embedded in multiple positions in the membranes of the endoplasmic reticulum, and for efficient flavivirus budding and maturation, it is proteolytically processed by the catalytically active virus encoded NS2B-NS3 proteinase (5) and host proteases (49). The premembrane (pr) and membrane (M) proteins in the polyprotein localized in the *trans*-Golgi network are cleaved by the host convertase furin, which cleaves at the highly conserved motif RXR/KR (3). The embedded polyprotein exposed on the luminal side of the membrane is processed by the enzyme signalase (50). Protein cleavage motifs are highly conserved among some flaviviruses, suggesting that co- and posttranslational processing of the polyprotein is similar even in diverse species. We predicted potential host signalase, host furin, and putative virus NS2B-NS3-protease sites for processing of the polyprotein (Table 2). Additionally, we identified weak C-terminal NS1 octapeptide sequences (L/M-V-X-S-X-V-X-A) in four of these flaviviruses. This octapeptide sequence is conserved in flaviviruses and has been demonstrated as important for the cleavage between NS1/NS2A by an unknown host protease (49).

For potential virus NS2B-NS3-protease cleavage sites, we conducted a Multiple Sequence Comparison by Log-Expectation (MUSCLE) alignment and used well-elucidated YFV and DENV processing sites, proximity to transmembrane domains, and protein domain homology to guide our search. Prediction of many NS2B-NS3-protease cleavage sites for the MIFs described here was challenging, as it appears that residues that are well established for flaviviruses are weakly conserved in basal group members, a finding that has previously been described for the divergent Tamana bat virus (16).

The NS3-Pro protein of many vector-borne flaviviruses and cISF species has been experimentally demonstrated to cleave after two basic amino acid residues (RR/RK/KR or, rarely, QR) before a small amino acid (G/A/S) (51). Some of the predicted sites are identical to canonical NS3-Pro motifs (Table 2), whereas many of the predicted NS3-Pro sites, especially in cephalopod flaviviruses, are at sites never previously been indicated for flaviviruses.

Analysis of the putative protease domains of the novel flaviviruses presented here (Fig. 3) suggests that they all share the trypsin-like serine protease catalytic triad (His, Asp, and Ser) known to all flaviviruses (52). However, boxes 3 and 4 of this domain,

TABLE 2 Predicted cleavage residues of marine flaviviruses

	Flavivirus					
Protein cleavage	Southern pygmy squid flavivirus (SpsFV)	Firefly squid flavivirus (FsFV)	Wenzhou shark flavivirus (<i>P. trituberculatus</i> strain)	Crangon crangon flavivirus (CcFV)	Gammarus chevreuxi flavivirus (GcFV)	Gammarus pulex flavivirus (GpFV)
<i>C/AnchC</i> (VSP)	106 ¹ BSVR ↓ SFV ₅ ¹¹³	86 ⁵ SVRR ↓ SVSS ₉₃	MSRR ↓ AGLG ₉₇ 97 ¹ GSGR ↓ MTMP ₁₀₄	121 ¹ SKRK ↓ SAGM ₁₂₈	110 ¹ VGRR ↓ NFPW ₁₁₄	101 ¹ SGQQ ↓ WKLS ₁₀₈
AnchC/pr-M (SP)	143 ¹ LAVA ↓ SLPA ₁₅₀	127 ¹ SAET ↓ FEHL ₁₃₄	132 ¹ ACAG ↓ MYCK ₁₃₉	141 ¹ CVLG ↓ INVS ₁₄₈	132 ¹ PTVG ↓ QATT ₁₃₉	Not identified
Pr-M/M-M (Furin)	179 ¹ RLAR ↓ NSGQ ₁₈₆ ¹	196 ¹ RLTR ↓ HYAL ₂₀₃ ¹	178 ¹ KTTR ↓ EVEK ₁₈₅ ¹	231 ¹ RRKR ↓ SIVEH ₂₃₉ ¹	271 ¹ RRKR ↓ SIPE ₂₇₈ ¹	Not identified
	<i>bona fide</i> furin motif	minimal RXXR motif	poor furin motif	<i>bona fide</i> furin motif	<i>bona fide</i> furin motif	
M/E (SP)	208 ¹ LVSS ↓ NTDV ₂₁₅	Not identified	427 ¹ KAEA ↓ SCHY ₄₃₄	284 ¹ TTRS ↓ KVKT ₂₉₁	345 ¹ FGSA ↓ YGD ₃₅₂	441 ¹ TTAL ↓ TPIR ₄₄₈
E/NS1 (SP)	717 ¹ LAEA ↓ APNT ₇₂₄	651 ¹ FATG ↓ DLVE ₆₅₈	878 ¹ VGLQ ↓ QKND ₈₈₅	792 ¹ VLGN ↓ SVGC ₇₉₉	839 ¹ GVHA ↓ NSLG ₈₄₆	561 ¹ TAWG ↓ SFQE ₅₆₈
NS1/NS2A (Host/VSP)	1438 ¹ RVA ↓ YTVT ₁₄₄₅	1186 ¹ ALRK ↓ FPRR ₁₁₉₃	1223 ¹ EVDA ↓ HCDL ₁₂₃₀	Not identified	1219 ¹ TVEA ↓ YSRW ₁₂₂₆	1315 ¹ IWA ↓ TLM ₁₃₂₃
NS2A/NS2B NS2B/NS2B (VSP)	1571 ¹ TLSR ↓ TRHA ₁₅₇₈	1460 ¹ FTRR ↓ VPLL ₁₄₆₇	Not identified	1389 ¹ SPRRR ↓ FIGS ₁₃₉₇	1508 ¹ NQQR ↓ EGKR ₁₅₁₅	1553 ¹ AFKQ ↓ VSTM ₁₅₆₀
	1737 ¹ PGRR ↓ IFSL ₁₇₄₄			1611 ¹ PRRR ↓ SGGT ₁₆₁₈		1592 ¹ FNTR ↓ SDLP ₁₅₉₉
NS2B/NS3 (VSP)	1923 ¹ FLPR ↓ SVTS ₁₉₃₀	1689 ¹ DHIK ↓ PVTE ₁₆₉₆	1566 ¹ QTSK ↓ SGLQ ₁₅₇₃	1745 ¹ SNQR ↓ TEPL ₁₇₅₂	1647 ¹ TTKK ↓ ANVV ₁₆₅₄	1739 ¹ SSK ↓ HSYN ₁₇₄₆
					1756 ¹ FYDK ↓ DYED ₁₇₆₅	
NS3/NS4A (VSP)	2538 ¹ FCPK ↓ NIIT ₂₅₆₀	2251 ¹ QCQR ↓ SRFV ₂₂₅₈	Not identified	2332 ¹ TGYR ↓ GGIS ₂₃₃₉	2259 ¹ LCYK ↓ SWNY ₂₂₆₄	2370 ¹ GPKR ↓ ATAQ ₂₃₇₇
NS4A/2K (VSP)	2687 ¹ TMQR ↓ NSGV ₂₆₉₅	~2434 ¹ LRQN ↓ SSTP ₂₄₄₃	2296 ¹ EGKK ↓ NKYE ₂₃₀₃	2444 ¹ SSFR ↓ STWD ₂₄₅₄	2381 ¹ GTRR ↓ STPE ₂₃₈₇	2555 ¹ NGTR ↓ SMVT ₂₅₆₂
2K/NS4B (SP)	2618 ¹ FSRS ↓ SFRK ₂₆₂₅	~2451 ¹ CIDA ↓ WVDP ₂₄₅₈	2317 ¹ VMLA ↓ GVTL ₂₃₂₄	2471 ¹ IAF ↓ ELD ₂₄₇₈	2402 ¹ LVA ↓ ANEA ₂₄₀₉	2581 ¹ FEA ↓ VFLP ₂₅₈₈
NS4B/NS5 (VSP)	3007 ¹ ITNQ ↓ SDDS ₃₀₁₄	~2733 ¹ EFKR ↓ AGVH ₂₇₄₀	2590 ¹ YFSK ↓ SGDP ₂₅₉₇	2743 ¹ NQKK ↓ GYRG ₂₇₅₀	2665 ¹ LSRR ↓ SDGA ₂₆₇₂	2861 ¹ DGVR ↓ SSFK ₂₈₆₈

FIG 3 Alignment of the four conserved trypsin-like serine motif boxes of the NS3-Pro region of novel flaviviruses compared to that of other flaviviruses, including the type species *Yellow fever virus*, and *Tamana bat virus*. Arrows indicate putative substrate-binding sites, and trypsin-like serine protease catalytic triad residues are indicated with an asterisk (*).

which contain residues that are involved in substrate binding and recognition, show reasonable flexibility. This may partly explain the imperfect prediction of the cleavage sites and suggests that the sensitivity and specificity for the canonical two basic amino acid residues may be inaccurate.

Phylogenetic placement of the novel marine invertebrate flaviviruses. While there is an ever-growing collection of ISFs, dISFs, and VIFs identified in the literature, most of these viruses are placed within well-supported lineages. Previous genus-wide phylogenetic analyses suggested three large well-supported clades of cISF, VIF, and a basal clade with the divergent flaviviruses *Tamana bat flavivirus* and *Wenzhou shark flavivirus* (17). We explored genus-wide phylogenetic relationships between the novel flaviviruses discovered here and representatives of all flaviviruses. We aligned 78 flaviviruses representing a comprehensive set of published species and employed two different phylogenetic inferences, the first being a maximum likelihood (ML) phylogeny using IQ-TREE, and the second being a maximum clade credibility (MCC) tree using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in MrBayes (53). The resultant trees were visualized using FigTree and midpoint rooted for clarity only. Examination of topologies of phylogenetic trees produced the same topology and strong support, indicated by posterior probabilities in the Bayesian tree and high bootstrap values for the ML tree. Importantly, the topologies of these trees are congruent with those of previously produced genus-wide phylogenies (11, 13, 17, 54). For simplicity, we have depicted the Bayesian tree (Fig. 4). Analysis of the posterior probability of the MCC tree suggests that the cephalopod flaviviruses are the most divergent of all known flaviviruses and cluster basally to a clade that encompasses *Cyclopterus lumpus virus*, *Tamana bat flavivirus*, and both strains of the *Wenzhou shark flavivirus*. This implies that cephalopod and marine vertebrate flaviviruses share a more closely related ancestor than any other flaviviruses. The three novel crustacean flaviviruses GcFV, GpFV, and CcFV form a well-supported clade that falls between cISFs and VIFs. This suggests that crustacean flaviviruses and VIFs share a more closely related ancestor with each other than cISFs do with VIFs.

Wenzhou shark flavivirus is abundant in swimming crabs (*Portunus trituberculatus*), and active replication is supported by a functional RNA interference response. The gazami crab, also known as the swimming crab, *P. trituberculatus*, is a commercially important crustacean widely distributed in Indian and West Pacific oceans, and it is ubiquitous in southeast/east Asian countries and the north and eastern coastal waters of Australia (55). It is the most widely fished crab in the world, with annual catches exceeding ~600,000 metric tons (56). Due to its ubiquity in shallow waters and its commercial fishing value, *P. trituberculatus* is a widely studied crustacean model. As such, there exists a wealth of RNA-Seq data from *P. trituberculatus* from a number of different catch locations. Hence, we explored the incidence of *Wenzhou shark flavivirus* within these crab populations, in addition to *P. trituberculatus* RNA-Seq data deposited in the short-read archive. RNA-Seq libraries were downloaded and

FIG 4 Maximum clade credibility tree of phylogenetic relationship of the novel flaviviruses discovered in this study within the *Flavivirus* genus under the protein substitution model (LG+G4). Classical insect-specific flaviviruses (cISF) and vertebrate-infecting flaviviruses (VIFs) are grouped by blue and red lines, respectively. Novel crustacean and cephalopod flaviviruses are indicated by orange and green boxes, respectively. Labels of two dual-host-affiliated insect-specific flaviviruses are labeled as blue text. Numbers on internal nodes represent posterior probabilities. Branch length represents amino acid substitutions per site. The GenBank identifier of each protein is indicated in the label.

TABLE 3 Incidence of Wenzhou shark flavivirus in published *P. trituberculatus* RNA-Seq data

Location of <i>P. trituberculatus</i> samples	Date of catch	Tissue type	Coverage (×) (no. of mapped reads/total)	Reference(s)
Shandong, China and Qingdao, China	July 2013 and October 2013	Combined eyestalk, gill, heart, hepatopancreas, and muscle from 90 crabs	33.50 (10,802/184,787,733)	31, 94
Weifang, China	June 2016	Gill tissue	86.48 (18484/95767617)	No associated publication
Weifang, China	November 2014	Ovary and testes tissues from five pooled crabs	139.20 (15,242/73,775,082)	95
Xiangshan, China	June 2015	Muscle tissue from 18 crabs	10,665.98 (759,435/334,037,716)	96
Xiangshan, China	May 2012	Ovary tissue from six crabs	550.77 (59060/191,845,918)	97
Weifang, China	March 2014	Pooled testes from five crabs	23.57 (11,726/3,302,979)	59

mapped to Wenzhou shark flavivirus, revealing evidence that five additional *P. trituberculatus* sequencing projects harbored Wenzhou shark flavivirus (Table 3). These *P. trituberculatus* sequencing projects derive from a range of eastern coastal China geographic origins and share some overlap with the location of the Pacific spade-nose shark, host of Wenzhou shark flavivirus, which is listed as Heilongjiang, Eastern China Sea.

One of these libraries was for small RNAs (sRNAs) originating from the testis tissue of *P. trituberculatus* wild-caught from Weifang, China. During virus infection in diverse arthropods, the ribonuclease III enzyme Dicer-2 processes double-stranded RNA (dsRNA) produced by viruses during replication into viral-derived short interfering RNAs (vsiRNAs) between 20 and 25 nt in length (57, 58). To exclude the likelihood of contamination and also to consider the strandedness and composition of these putative vsiRNA reads, we mapped the small RNAs to the representative genome. Among reads with a 5' adapter and a length of 18 to 30 nt, 11,726 out of 3,302,979 reads (0.3%) mapped to both the sense and antisense strands of the Wenzhou shark flavivirus contig along its entire length (Fig. 5A) (59). Visualizing the size distribution of these mapped reads indicated that 21- to 22-nt reads represented 31% of the total mapped reads, with 22-nt reads as the most abundant size of vsiRNAs (16.8%) (Fig. 5B). The presence of vsiRNA reads originating from both the sense and antisense strands of Wenzhou shark flavivirus in the gazami crab, and also the 21- to 22-nt length bias in the mapping of this library, indicate not only that the virus is present in these crab populations but also that an active RNA interference (RNAi) response is produced against the virus.

Flavivirus endogenous viral elements in crustacean genomes are more closely related to vector-borne flaviviruses than to cISF. Flavivirus endogenous viral elements (EVEs) are fragments of flaviviruses that are known to be integrated within the genomic DNA of a previously infected host (60). Flaviviruses do not encode reverse transcriptase or integrase domains, but flavivirus EVEs are reported in the genomes of two well-known flavivirus vectors, *Aedes aegypti* and *Aedes albopictus* (60, 61), as well as those of *Anopheles* mosquitoes (62). We sought to identify previous flavivirus infections in crustacean or cephalopod genomes by identifying these EVEs. In a screen of genomes deposited in the NCBI whole-genome contig database, we identified numerous flavivirus EVEs within crustacean genomes but were unable to identify any in deposited cephalopod genomes. Two representative flavivirus EVEs from the tadpole shrimp, *Lepidurus arcticus* (63), and the planktonic crustacean *Daphnia magna* are presented (Fig. 6). Both genomic regions have fragmented NS5 protein homology (pfam00972; E value, $<5E-48$) and association with retrotransposable (RT) elements, such as RT peptidases or RT domains. The *D. magna* flavivirus EVE also has a FstJ methyltransferase domain (pfam01728; E value, $<2E-11$). BLASTX and BLASTN analyses of these regions suggested that they share the highest nucleotide and protein identity to vector-borne and vertebrate-infecting flaviviruses. Attempts to phylogenetically place these EVEs proved difficult, as phylogenies produced using the flaviviral EVEs discovered here indicated low bootstrap support. The existence of highly fragmented

FIG 5 RNAi response targets Wenzhou shark flavivirus in *P. trituberculatus*. (A) Mapping profile of pooled small RNA fraction in the testes of the gazami crab mapped to the Wenzhou shark flavivirus genome (red) and antigenome (blue). (B) Profile of the distribution of 18- to 30-nt sRNA reads mapping to Wenzhou shark flavivirus genome (red) and antigenome (blue).

flavivirus EVEs in crustaceans suggests a long evolutionary history of previous challenge by flaviviruses. These flavivirus EVEs in two Branchiopoda crustaceans indicate that infection of crustaceans extends beyond the hosts we have assembled whole flavivirus genomes from.

FIG 6 Flavivirus endogenous viral elements exist in crustaceans and are closely related to those of vertebrate-infecting flaviviruses. Genome regions (5 kb) of *Lepidurus arcticus* (A) and *Daphnia magna* (B). First line indicates evidence of Pfam protein motifs. Flavivirus NS5 domain is indicated in green, retrotransposable (RT) element domain in red, and FtsJ methyltransferase in gray. Second and third lines show the highest total scores given from BLASTX and BLASTN hits against the nonredundant virus database hosted by NCBI.

FIG 7 Odds ratio of (A) CpG dinucleotides and (B) UpA dinucleotides of marine flaviviruses compared to members of the *Flavivirus* genus. Analysis included vertebrate flaviviruses ($n = 62$), representing pooled vector-borne and NKV flaviviruses, as well as insect-specific flaviviruses (cISF $n = 19$ and dISF $n = 10$). Statistically underrepresented dinucleotide motifs (≤ 0.78) are indicated as blue points, whereas points that show no difference are black.

Analysis of the mononucleotide and dinucleotide composition of novel marine flaviviruses to predict host range.

Classic CpG underrepresentation has been demonstrated in all flaviviruses known to infect vertebrates. In comparison, there is no clear selection against or for CpG motifs in cISFs, and they are weakly selected against in dISFs (64). We initially explored the odds ratios of CpG and also UpA, which are typically underrepresented in arthropod mRNAs of marine flaviviruses, and compared these in a genus-wide flavivirus context (Fig. 7). Odds ratios are calculated as the observed ratio of this dinucleotide motif over the expected ratios of individual mononucleotides. When there is no selection against a motif, the odds ratio should approach 1, whereas an odds ratio of a dinucleotide motif that is ≤ 0.78 or ≥ 1.23 indicates a statistically significant under or overrepresentation of the motif (65). We calculated and grouped flaviviruses that are known to replicate in vertebrates ($n = 62$) against both groups of cISF ($n = 19$) and dISF ($n = 10$). Comparing the odds ratios of the CpG motif in marine flaviviruses suggested that for most of these viruses CpG is underrepresented, with only *Gammarus chevreuxi* flavivirus and Firefly squid flavivirus not having statistically underrepresented CpG motifs (Fig. 7A) (66, 67). In comparison, there is a genus-wide selection against UpA for all flaviviruses irrespective of host range (Fig. 7B).

Using odds ratios of only two dinucleotide motifs to predict host range is too simplistic; therefore, we sought to use a hierarchical clustering method to assess the natural groupings of all flaviviruses based on mono- and dinucleotide composition. Hierarchical clustering is a statistical clustering analysis that builds clusters out of associations between groups using predictive variables. For this, we calculated the odds ratios of each mononucleotide ($n = 4$) and dinucleotide ($n = 16$) bias of the polyprotein of 96 complete or coding-complete flavivirus genomes deposited in GenBank, as well as those of the six flaviviruses from this study. We only considered the open reading frame.

Using the frequencies of each mononucleotide and dinucleotide (20 parameters) as predictive factors, we used the pvclust package, which not only performs the agglomerative hierarchical clustering (HC) analysis but also assesses the certainty of different groupings by calculating P values via multiscale bootstrap resampling (68). Analysis of the resultant dendrogram (Fig. S1) suggests there are two large supergroupings of similarity with composition of nucleotides within the *Flavivirus* genus, an invertebrate flavivirus supergroup and a vertebrate-associated supergroup. Hierarchical clustering analysis assigned marine flaviviruses into two distinct groups; the first was a vertebrate-associated and vector-borne group with both strains of Wenzhou shark flavivirus and *Cyclopterus lumpus* virus and *Crangon crangon* flavivirus. This first grouping was most

FIG 8 Novel marine flaviviruses associate with both invertebrate- and vertebrate-infecting flaviviruses. Linear discriminant analysis of the 6 groups as resolved by hierarchical clustering analysis (Fig. S1) using mononucleotide and dinucleotide odds ratios as predictive variables.

closely associated with reasonable probabilistic support (85%) to a group of vector-borne flaviviruses that includes ZIKV and DENV. The second marine flavivirus group contained the two *Gammarus* and two cephalopod flaviviruses. These flaviviruses shared mononucleotide and dinucleotide composition more closely with the cISF grouping. What is important to note is that this HC method was successfully able to assign all dISFs together as one well-supported group, suggesting that HC analysis can sensitively discriminate between dISFs and VIFs.

As the HC analysis provided some degree of confidence in the natural structure of groups, we used the HC results to inform the linear discriminant analyses (LDA). Flaviviruses were then assigned into groups of vertebrate-infecting flaviviruses (VIF-1 and VIF-2), marine vertebrate-associated flaviviruses (MF1), cISF, dISF, and marine invertebrate flaviviruses (MF2) (Fig. 8). Both the HC analysis and the LDA shared complete consensus in assignments to these groups, with the exception of one outlier, New Mapoon virus. These results provide evidence for two different dinucleotide compositions for the marine flaviviruses, a marine invertebrate flavivirus group and a marine vertebrate-associated group.

DISCUSSION

In this study, we have undertaken a metagenomic and phylogenetic approach to further uncover the diversity of flaviviruses. To this end, we have discovered and phylogenetically placed five novel flaviviruses of marine invertebrates. The grouping of crustacean flaviviruses suggests that terrestrial vector-borne flaviviruses and crustacean flaviviruses share a closer common ancestor than any reported and sequenced flaviviruses and also that the flavivirus endogenous elements of crustaceans are closer in nucleotide and protein identity to vector-borne flaviviruses than to cISFs. Furthermore, we provide four lines of evidence for horizontal transmission of Wenzhou shark flavivirus between crabs and Pacific spadenose sharks, i.e., strong nucleotide identity between both strains of Wenzhou shark flavivirus, the geographical overlap of the catch locations of both crab and shark samples, proof of replication within *P. trituberculatus* through evidence of an RNAi response, and mono- and dinucleotide composition that most closely aligns with that of vector-borne terrestrial flaviviruses.

Flaviviruses have gained and lost the ability to infect vertebrates a number of times through their evolutionary history (17). As we have demonstrated that Wenzhou shark flavivirus has jumped from or jumped to *P. trituberculatus*, this provides evidence of additional dual vertebrate-invertebrate host associations in flavivirus evolution. Previously, it has been demonstrated that rhabdoviruses have occasionally moved between distantly related hosts and then established in these environments by spreading into closely related hosts (27). It is difficult to extrapolate the evolution and emergence of terrestrial vector-borne flaviviruses in relation to the crustacean flaviviruses and cISF and VIF clades. We have limited data to conclude the evolutionary direction of flaviviruses with any certainty. While it is possible to date the crustacean node of EVEs from other closely related *Daphnia* and crustacean species, no additional genomes for these species are available. As such, we cannot give an accurate estimate as to when these endogenization events may have taken place. As we have discovered a new clade of flaviviruses that are closely related to VIF, the following two possible scenarios could be speculated: (i) terrestrial vector-borne flaviviruses evolved from a crustacean flavivirus ancestor, or (ii) insect-specific flaviviruses and crustacean flaviviruses codiverged, and VIFs subsequently gained an ability to infect vertebrates. However, additional data are required to resolve the scenarios and the direction of evolution.

In invertebrates, upon RNAi response, short interfering RNAs (siRNAs) are produced in response to dsRNA triggers (reviewed in Liu et al. [69] and Asgari [70]). During viral infection, the ribonuclease III enzyme Dicer-2 processes dsRNA, produced by viruses during replication, into siRNAs between 20 and 25 nt in length (57, 58). These virus-derived siRNAs (vsiRNAs) are then loaded into the RNA-induced silencing complex (RISC), and they have been shown to modulate and control virus accumulation and tolerance in the host (71). In invertebrates, the RNAi response against flaviviruses has been well established in mosquitoes infected with both VIF (72–74) and cISFs (75, 76). In crustaceans, delivery of siRNA and dsRNA have been experimentally demonstrated to control white spot syndrome virus (WSSV) (family *Nimaviridae*) replication in both the *Penaeus japonicus* shrimp and the whiteleg shrimp, *Litopenaeus vannamei*, challenged with WSSV (77, 78). While these studies have experimentally demonstrated that crustaceans may share the RNAi machinery that targets viral replication, the first report of exogenously produced vsiRNAs was only recently described from small RNA sequencing of the penaeid shrimp *Fenneropenaeus chinensis* challenged with WSSV (79). In *F. chinensis*, it was shown that processing of viral dsRNA resulted in 21- to 22-nt vsiRNAs, with the most abundant size being 22 nt. Consistently, we found large numbers of 21- to 22-nt vsiRNAs (with a slight bias toward 22 nt) in *P. trituberculatus* that mapped to both positive-sense and negative-sense RNA strands of Wenzhou shark flavivirus. The mapped reads indicated that vsiRNAs are derived from along the entire length of the virus, suggesting active replication in the invertebrate host. This is the second report of vsiRNAs in crustaceans and indicates that 22-nt sRNA fragments are produced preferentially by the RNAi pathway in crustaceans against viral pathogens.

It appears that some marine flaviviruses contain a -1 PRF motif to produce a singular polyprotein. Furthermore, it seems possible that the two discrete open reading frames observed in these viruses may produce functional virions with only transmembrane proteins lost within the frameshifted region. Numerous monopartite RNA viruses have structural and nonstructural genes encoded on discrete open reading frames, notably alphaviruses, which encode two discrete open reading frames as a means to control translation of these proteins.

Within the *Flavivirus* genus, -1 PRF motifs are accepted and reported in both VIF and cISF species. In cISF members, the alternative reading frame is located near the NS2A/NS2B-coding region, the fairly interesting flavivirus open reading frame (*fifo*) (45). It has been demonstrated that cISFs encode a downstream *fifo* between 221 and 293 additional amino acids (aa) in length. This is much larger than the -1 PRF region that exists in a number of VIFs, which extends the NS1 by additional residues (delineated NS1') (46, 47). Experimentally validated NS1' translation efficiencies in West Nile virus suggests that slippage happens between $\sim 20\%$ and 50% of the time, depending on

the cell line (47). The existence of this -1 PRF within the marine flaviviruses reported here occurs within the same regions as those two -1 PRF regions of cISFs and VIFs. Not all VIFs encode this -1 PRF, and it appears that the -1 PRF region has evolved independently or been lost multiple times in the *Flavivirus* genus. The reasons for this location have been discussed extensively (80). However, identification of this -1 PRF within the crustacean flaviviruses also presents an alternative to the origin of the NS1' extension and *fifo* protein as a vestigial or retained genomic feature arisen from a mechanism that was required to produce a complete polyprotein.

Ongoing coevolution of flaviviruses, either with their invertebrate or dual invertebrate-vertebrate hosts, places certain constraints on the coding region of the viral polyprotein that limit the nucleotide and codon usage (64). It is well established that vertebrate mRNAs show an underrepresentation of the UpA and CpG dinucleotides, whereas typical arthropod genes display an underrepresentation of the UpA dinucleotide (66, 67). It is also known that the dinucleotide composition is conserved among VIFs and cISFs. Considering that in mammals, CpG motifs modulate the innate and adaptive immune response (81), it has been suggested that this is the reason for CpG motifs being selected against in VIFs, whereas CpG is not selected against in cISFs (82).

Previous studies have shown that mono- and dinucleotide compositions are reasonable predictors of the host range of flaviviruses. In a family-wide analysis of mono- and dinucleotide compositions for *Flaviviridae*, 76% of the virus hosts could be accurately predicted using odds ratios of all possible 16 dinucleotides and the four mononucleotides. (83). Mono- and dinucleotide compositions have also been used in linear discriminant analyses using a training set of RNA viruses with established host ranges (vertebrate-only, invertebrate-only [cISF], vector-borne, plant, and bacterial). Using this approach, it was possible to sensitively assign *Anopheles* cISFs as an invertebrate-only grouping (84). Importantly, our analyses indicate that dinucleotide compositions of the marine flaviviruses discovered here associate these viruses into tentative marine invertebrate and marine vertebrate host-associated groupings.

CONCLUSIONS

This work presents evidence that flaviviruses infect a range of marine invertebrates and improves our understanding of the potential origins and emergence of invertebrate-vertebrate flaviviruses. The identified marine invertebrate flaviviruses provide insight into flavivirus genome organization, and we demonstrate a clear example of dual host invertebrate-vertebrate flaviviruses in sharks and crabs. This study presents interesting future avenues to explore the emergence of terrestrial vector-borne flaviviruses and scenarios for a jump to terrestrial arthropods and terrestrial vertebrates.

MATERIALS AND METHODS

Identification of divergent flaviviruses and endogenous viral elements from published transcriptomes and genome assemblies. To uncover unknown and divergent flaviviruses infecting other eukaryotic organisms, we queried the polyprotein sequence of Wenzhou shark flavivirus (GenBank accession number [AVM87250.1](#)) using the translated BLAST: tblastn algorithm against the Transcriptome Shotgun Assembly Sequence Database and the whole-genome shotgun contigs database hosted by the National Center for Biotechnology Information (NCBI). We restricted the search to Animalia (taxonomic identifier [taxid] 33208) under default parameters. A number of 2- to 3-kb contigs encoding a flavivirus NS5 and NS3 Pfam domain were identified in terrestrial arthropods and excluded, as they appeared to be related to the multisegmented Jingmenvirus group (30). Contigs were then subsequently queried using BLASTx against the nonredundant database. After positive identification, these flavivirus strains were then used to query additional wild-caught published Cephalopoda (taxid 6605) and Crustacea (taxid 6657) RNA-Seq data using the BLASTn search on SRA.

De novo assembly of Gammarus pulex flavivirus and Firefly squid flavivirus, and bioinformatic validation of previously assembled data. For assembly of Firefly squid flavivirus, RNA-Seq data (SRA accession number [SRR2960129](#)) was imported into the Galaxy Australia Webserver (<https://usegalaxy.org.au/>), and quality and adapter trimmed using Trimmomatic (85). Trimmed reads were then *de novo* assembled using Trinity (Galaxy version 2.4.0.2) (86). *Gammarus pulex* flavivirus (GPFV) was assembled using the CLC Genomics Workbench (version 10.1.1) with default settings. To obtain mapping statistics and also validate putative flavivirus contigs, original total RNA-Seq data were downloaded with CLC Genomics Workbench, adapter and quality read trimmed (quality score, <0.05 ; ambiguous nucleotides, 2), and then remapped to flavivirus contigs using identity (0.9) and length (0.9) mapping criteria.

RNAi analysis. For analysis of the mapping and profile of virus-derived small RNAs from the testes of *P. trituberculatus*, we used a previously established workflow (87). Briefly, raw sRNA data were downloaded and trimmed with the same quality and ambiguous nucleotide score. Reads that did not have adapters, and those that were less than 16 nt, were discarded. Clean reads were then mapped to genome and antigenome with the RNA-Seq analysis tool, using strict mapping criteria (mismatch, insertion, and deletion costs: 2:3:3, respectively). Profiles of mapped read sizes were then graphed using Excel.

Flavivirus genome annotation. Viral open reading frames were predicted with ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) using the invertebrate genetic code. To characterize the functional domains, predicted polyprotein sequences were subjected to a domain-based search using the Conserved Domain Database (CDD) version 3.16 (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) and cross-referenced with the Pfam database (version 32.0) hosted at <http://pfam.xfam.org/>. For predicted cleavage residues of the polyprotein, we used a transmembrane topology prediction using the TMHMM Server version 2.0 (www.cbs.dtu.dk/services/TMHMM/), guided by previously experimentally validated motifs from YFV (4, 5). To identify signal peptides, a 40- to 60-amino acid (aa) sliding window of the polyprotein was assessed by the SignalP version 4.1 webserver (<http://www.cbs.dtu.dk/services/SignalP/>). Putative furin cleavage sites were identified using the ProP webserver (<http://www.cbs.dtu.dk/services/ProP/>). Putative N-glycosylation sites were predicted using the NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). The secondary structure of RNA was predicted using the RNAfold webserver (<http://ma.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) under default conditions.

Phylogenetic analysis. Protein alignment of 81 flavivirus sequences were performed using the MUSCLE algorithm (88), which resulted in an alignment of 81 sequences with 4,717 alignment positions. Ambiguous or problematic alignment blocks were removed using the Gblocks version 0.91 webserver, which resulted in 509 sites of alignment (http://molevol.cmima.csic.es/castresana/Gblocks_server.html). The Gblocks alignment was then analyzed for best amino acid substitution model using ModelFinder (89) incorporated in IQ-TREE (90), which explored 546 protein substitution models and predicted the general matrix (LG) model (91) with discrete gamma model with 4 rate categories (+G4) as the most suitable.

Maximum likelihood tree inference was made using IQ-TREE with Ultrafast bootstrap approximation (92), using computational resources from the Cyberinfrastructure for Phylogenetic Research Science Gateway (93). A maximum clade credibility tree was estimated using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in MrBayes version 3.2.3, using the same protein substitution model run for 50,000,000 generations (six runs, four chains) sampled every 1,000 generations, with 25% discarded as burn in. Convergence diagnostics were assessed using the information logged, such as the potential scale reduction factor (PSRF; equal to 1). The resultant tree was visualized using FigTree version 1.4 (A. Rambaut; <http://tree.bio.ed.ac.uk/software/figtree/>).

Analysis of mononucleotide and dinucleotide composition. For analysis of mononucleotide and dinucleotide motifs, we used a total of 103 flavivirus genomes deposited in GenBank. Open reading frames of the flaviviruses were predicted using CLC Genomics Workbench, and frequencies were counted using the length (LEN) function in Excel. Odds ratio of mono- and dinucleotide motifs were calculated as previously described (65), and hierarchical clustering was performed on 16 dinucleotide ratios and 4 mononucleotide ratios for 103 flaviviruses analyzed using the pvclust package, which assesses the uncertainty of hierarchical cluster analysis by calculating *P* values via multiscale bootstrap resampling in R (68). Prior to using pvclust, data were transposed, then scaled in R. pvclust was run using correlation and Ward's clustering methods with 10,000 bootstraps (method.dist="cor," method.hclust="ward.D2," nboot = 10000). The groupings ascertained using the HC method were then used as the categories for a linear discriminant analysis using the lda function of the MASS package suite in R studio. All raw data used for this analysis and the Excel calculator are available in Table S2.

Data availability. All data used within this publication are available within text and supplementary files. Sequencing data used to assemble the genome are available in Table S1. Accession numbers of annotated flaviviruses have been deposited in GenBank under the accession numbers MK473875 to MK473881.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.00432-19>.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We acknowledge the efforts of the original producers of RNA-Seq data used in this study, without the open availability of which this project would not have been possible. We thank Karyn Johnson for suggestions on preliminary data and members of the Asgari lab for feedback. This work utilized the resources of the Galaxy Australia server (<https://usegalaxy.org.au/>). The picture of Gammaridea used in the phylogeny is by Hans Hillewaert, the graphic of Portunidae is by Hans Hillewaert (vectorized by T. Michael Keesey), and the tick picture is by Henry Lydecker. All pictures were reused under the Creative Commons Attribution-ShareAlike 3.0 Unported license (<http://creativecommons.org/licenses/by-sa/3.0/>) and were downloaded from Phylopic (<http://www.phylopic.org>).

This project was funded by an Australian Research Council grant (DP150101782) to S.A. and by a University of Queensland scholarship to R.P.

REFERENCES

- Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, Muerhoff S, Pletnev A, Rico-Hesse R, Smith DB, Stapleton JT, Ictv Report C. 2017. ICTV virus taxonomy profile: *Flaviviridae*. *J Gen Virol* 98:2–3. <https://doi.org/10.1099/jgv.0.000672>.
- Chambers TJ, Hahn CS, Galler R, Rice CM. 1990. Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* 44:649–688. <https://doi.org/10.1146/annurev.mi.44.100190.003245>.
- Stadler K, Allison SL, Schlich J, Heinz FX. 1997. Proteolytic activation of tick-borne encephalitis virus by furin. *J Virol* 71:8475–8481.
- Chambers TJ, Grakoui A, Rice CM. 1991. Processing of the yellow fever virus nonstructural polyprotein—a catalytically active NS3-proteinase domain and NS2b are required for cleavages at dibasic sites. *J Virol* 65:6042–6050.
- Chambers TJ, Weir RC, Grakoui A, Mccourt DW, Bazan JF, Fletterick RJ, Rice CM. 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. *Proc Natl Acad Sci U S A* 87:8898–8902. <https://doi.org/10.1073/pnas.87.22.8898>.
- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI. 2013. The global distribution and burden of dengue. *Nature* 496:504–507. <https://doi.org/10.1038/nature12060>.
- Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, West K, Qu J, Baniecki ML, Gladden-Young A, Lin AE, Tomkins-Tinch CH, Ye SH, Park DJ, Luo CY, Barnes KG, Shah RR, Chak B, Barbosa-Lima G, Delatorre E, Vieira YR, Paul LM, Tan AL, Barcellona CM, Porcelli MC, Vasquez C, Cannons AC, Cone MR, Hogan KN, Kopp EW, Anzinger JJ, Garcia KF, Parham LA, Ramirez RM, Montoya MCM, Rojas DP, Brown CM, Hennigan S, Sabina B, Scotland S, Gangavarapu K, Grubaugh ND, Oliveira G, Robles-Sikisaka R, Rambaut A, Gehrke L, Smole S, Halloran ME, Villar L, Mattar S, Lorenzana I, Cerbino-Neto J, Valim C, Degraeve W, Bozza PT, Gnirke A, Andersen CG, Isern S, Michael SF, Bozza FA, Souza TML, Bosch I, Yozwiak NL, MacInnis BL, Sabeti PC. 2017. Zika virus evolution and spread in the Americas. *Nature* 546:411–415. <https://doi.org/10.1038/nature22402>.
- Huang BX, Firth C, Watterson D, Allcock R, Colmant AMG, Hobson-Peters J, Kirkland P, Hewitson G, McMahon J, Hall-Mendelin S, van den Hurk AF, Warrilow D. 2016. Genetic characterization of archived bunyaviruses and their potential for emergence in Australia. *Emerg Infect Dis* 22:833–840. <https://doi.org/10.3201/eid2205.151566>.
- Shi M, Neville P, Nicholson J, Eden JS, Imrie A, Holmes EC. 2017. High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J Virol* 91:e00680–17. <https://doi.org/10.1128/JVI.00680-17>.
- Blitvich BJ, Firth AE. 2015. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses* 7:1927–1959. <https://doi.org/10.3390/v7041927>.
- Blitvich BJ, Firth AE. 2017. A review of flaviviruses that have no known arthropod vector. *Viruses* 9:154. <https://doi.org/10.3390/v9060154>.
- Varelas-Wesley I, Calisher CH. 1982. Antigenic relationships of flaviviruses with undetermined arthropod-borne status. *Am J Trop Med Hyg* 31:1273–1284. <https://doi.org/10.4269/ajtmh.1982.31.1273>.
- Moureaux G, Cook S, Lemey P, Nougaiere A, Forrester NL, Khasnatinov M, Charrel RN, Firth AE, Gould EA, de Lamballerie X. 2015. New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. *PLoS One* 10:e0117849. <https://doi.org/10.1371/journal.pone.0117849>.
- Lawrie CH, Uzategui NY, Armesto M, Bell-Sakyi L, Gould EA. 2004. Susceptibility of mosquito and tick cell lines to infection with various flaviviruses. *Med Vet Entomol* 18:268–274. <https://doi.org/10.1111/j.0269-283X.2004.00505.x>.
- Kuno G. 2007. Host range specificity of flaviviruses: correlation with *in vitro* replication. *J Med Entomol* 44:93–101. <https://doi.org/10.1093/jmedent/41.5.93>.
- de Lamballerie X, Crochu S, Billoir F, Neyts J, de Micco P, Holmes EC, Gould EA. 2002. Genome sequence analysis of Tamana bat virus and its relationship with the genus *Flavivirus*. *J Gen Virol* 83:2443–2454. <https://doi.org/10.1099/0022-1317-83-10-2443>.
- Shi M, Lin XD, Chen X, Tian JH, Chen LJ, Li K, Wang W, Eden JS, Shen JJ, Liu L, Holmes EC, Zhang YZ. 2018. The evolutionary history of vertebrate RNA viruses. *Nature* 556:197–202. <https://doi.org/10.1038/s41586-018-0012-7>.
- Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu JG, Holmes EC, Zhang YZ. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* 4:e05378. <https://doi.org/10.7554/eLife.05378>.
- Skoge RH, Brattespe J, Okland AL, Plarre H, Nylund A. 2018. New virus of the family *Flaviviridae* detected in lumpfish (*Cyclopterus lumpus*). *Arch Virol* 163:679–685. <https://doi.org/10.1007/s00705-017-3643-3>.
- Geoghegan JL, Di Giallonardo F, Cousins K, Shi M, Williamson JE, Holmes EC. 2018. Hidden diversity and evolution of viruses in market fish. *Virus Evol* 4:vey031. <https://doi.org/10.1093/ve/vey031>.
- Conway MJ. 2015. Identification of a flavivirus sequence in a marine arthropod. *PLoS One* 10:e0146037. <https://doi.org/10.1371/journal.pone.0146037>.
- Duchene S, Di Giallonardo F, Holmes EC. 2016. Substitution model adequacy and assessing the reliability of estimates of virus evolutionary rates and time scales. *Mol Biol Evol* 33:255–267. <https://doi.org/10.1093/molbev/msv207>.
- Forrester NL, Palacios G, Tesh RB, Savji N, Guzman H, Sherman M, Weaver SC, Lipkin WI. 2012. Genome-scale phylogeny of the *Alphavirus* genus suggests a marine origin. *J Virol* 86:2729–2738. <https://doi.org/10.1128/JVI.05591-11>.
- La Linn M, Gardner J, Warrilow D, Darnell GA, McMahon CR, Field I, Hyatt AD, Slade RW, Suhrbier A. 2001. Arbovirus of marine mammals: a new alphavirus isolated from the elephant seal louse, *Lepidophthirus macrorhini*. *J Virol* 75:4103–4109. <https://doi.org/10.1128/JVI.75.9.4103-4109.2001>.
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GW, Friedrich M, Meier R, Yeates DK. 2011. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A* 108:5690–5695. <https://doi.org/10.1073/pnas.1012675108>.
- Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeck MA. 2004. Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci U S A* 101:6536–6541. <https://doi.org/10.1073/pnas.0401670101>.
- Longdon B, Murray GG, Palmer WJ, Day JP, Parker DJ, Welch JJ, Obbard D, Jiggins FM. 2015. The evolution, diversity, and host associations of rhabdoviruses. *Virus Evol* 1:vev014. <https://doi.org/10.1093/ve/vev014>.
- Francois S, Filloux D, Roumagnac P, Bigot D, Gayral P, Martin DP, Froissart R, Ogliastro M. 2016. Discovery of parvovirus-related sequences in an unexpected broad range of animals. *Sci Rep* 6:30880. <https://doi.org/10.1038/srep30880>.
- Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res* 36:W5–9. <https://doi.org/10.1093/nar/gkn201>.
- Shi M, Lin XD, Vasilakis N, Tian JH, Li CX, Chen LJ, Eastwood G, Diao XN, Chen MH, Chen X, Qin XC, Widen SG, Wood TG, Tesh RB, Xu J, Holmes EC, Zhang YZ. 2016. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *J Virol* 90:659–669. <https://doi.org/10.1128/JVI.02036-15>.
- Lv J, Zhang L, Liu P, Li J. 2017. Transcriptomic variation of eyestalk reveals the genes and biological processes associated with molting in *Portunus trituberculatus*. *PLoS One* 12:e0175315. <https://doi.org/10.1371/journal.pone.0175315>.
- Martinez-Alarcon D, Harms L, Hagen W, Saborowski R. 2018. Transcriptome analysis of the midgut gland of the brown shrimp *Crangon crangon* indicates high polymorphism in digestive enzymes. *Mar Genom* 43:1–8. <https://doi.org/10.1016/j.margen.2018.09.006>.

33. Gaibani P, Rossini G. 2017. An overview of Usutu virus. *Microbes Infect* 19:382–387. <https://doi.org/10.1016/j.micinf.2017.05.003>.
34. Truebano M, Tills O, Spicer JI. 2016. Embryonic transcriptome of the brackishwater amphipod *Gammarus chevreuxi*. *Mar Genom* 28:5–6. <https://doi.org/10.1016/j.margen.2016.02.002>.
35. Collins M, Tills O, Spicer JI, Truebano M. 2017. *De novo* transcriptome assembly of the amphipod *Gammarus chevreuxi* exposed to chronic hypoxia. *Mar Genom* 33:17–19. <https://doi.org/10.1016/j.margen.2017.01.006>.
36. Trapp J, Geffard O, Imbert G, Gaillard JC, Davin AH, Chaumot A, Armen-gaud J. 2014. Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods. *Mol Cell Proteomics* 13:3612–3625. <https://doi.org/10.1074/mcp.M114.038851>.
37. May FJ, Clark DC, Pham K, Diviney SM, Williams DT, Field EJ, Kuno G, Chang GJ, Cheah WY, Setoh YX, Prow NA, Hobson-Peters J, Hall RA. 2013. Genetic divergence among members of the Kokobera group of flaviviruses supports their separation into distinct species. *J Gen Virol* 94: 1462–1467. <https://doi.org/10.1099/vir.0.049940-0>.
38. Billoir F, de Chesse R, Tolou H, de Micco P, Gould EA, de Lamballerie X. 2000. Phylogeny of the genus *Flavivirus* using complete coding sequences of arthropod-borne viruses and viruses with no known vector. *J Gen Virol* 81:781–790. <https://doi.org/10.1099/0022-1317-81-3-781>.
39. Gimenez G, Metcalf P, Paterson NG, Sharpe ML. 2016. Mass spectrometry analysis and transcriptome sequencing reveal glowing squid crystal proteins are in the same superfamily as firefly luciferase. *Sci Rep* 6:27638. <https://doi.org/10.1038/srep27638>.
40. Caruana NJ, Cooke IR, Faou P, Finn J, Hall NE, Norman M, Pineda SS, Strugnell JM. 2016. A combined proteomic and transcriptomic analysis of slime secreted by the southern bottletail squid, *Sepiadiarium austrinum* (Cephalopoda). *J Proteomics* 148:170–182. <https://doi.org/10.1016/j.jpropt.2016.07.026>.
41. Fontes-Garfias CR, Shan C, Luo H, Muruato AE, Medeiros DBA, Mays E, Xie X, Zou J, Roundy CM, Wakamiya M, Rossi SL, Wang T, Weaver SC, Shi PY. 2017. Functional analysis of glycosylation of Zika virus envelope protein. *Cell Rep* 21:1180–1190. <https://doi.org/10.1016/j.celrep.2017.10.016>.
42. Winkler G, Heinz FX, Kunz C. 1987. Studies on the glycosylation of flavivirus E proteins and the role of carbohydrate in antigenic structure. *Virology* 159:237–243. [https://doi.org/10.1016/0042-6822\(87\)90460-0](https://doi.org/10.1016/0042-6822(87)90460-0).
43. Flamand M, Megret F, Mathieu M, Lepault J, Rey FA, Deubel V. 1999. Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion. *J Virol* 73:6104–6110.
44. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. 1998. Phylogeny of the genus *Flavivirus*. *J Virol* 72:73–83.
45. Blitvich BJ, Firth AE, Wills NM, Brault AC, Miller CL, Atkins JF. 2010. Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. *Virology* 399:153–166. <https://doi.org/10.1016/j.virol.2009.12.033>.
46. Firth AE, Atkins JF. 2009. A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Virol J* 6:14. <https://doi.org/10.1186/1743-422X-6-14>.
47. Melian EB, Hinzman E, Nagasaki T, Firth AE, Wills NM, Nouwens AS, Blitvich BJ, Leung J, Funk A, Atkins JF, Hall R, Khromykh AA. 2010. NS1' of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *J Virol* 84:1641–1647. <https://doi.org/10.1128/JVI.01979-09>.
48. Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV. 2016. Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res* 44:7007–7078. <https://doi.org/10.1093/nar/gkw530>.
49. Hori H, Lai CJ. 1990. Cleavage of dengue virus NS1-NS2a requires an octapeptide sequence at the C-terminus of NS1. *J Virol* 64:4573–4577.
50. Lobigs M, Lee E. 2004. Inefficient signalase cleavage promotes efficient nucleocapsid incorporation into budding flavivirus membranes. *J Virol* 78:178–186. <https://doi.org/10.1128/JVI.78.1.178-186.2004>.
51. Cammisa-Parks H, Cisar LA, Kane A, Stollar V. 1992. The complete nucleotide sequence of cell fusing agent (CFA): homology between the nonstructural proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. *Virology* 189:511–524. [https://doi.org/10.1016/0042-6822\(92\)90575-A](https://doi.org/10.1016/0042-6822(92)90575-A).
52. Bazan JF, Fletterick RJ. 1989. Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology* 171:637–639. [https://doi.org/10.1016/0042-6822\(89\)90639-9](https://doi.org/10.1016/0042-6822(89)90639-9).
53. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>.
54. Pettersson JHO, Fiz-Palacios O. 2014. Dating the origin of the genus *Flavivirus* in the light of Beringian biogeography. *J Gen Virol* 95: 1969–1982. <https://doi.org/10.1099/vir.0.065227-0>.
55. Carpenter KE, Niem VH, Norsk utviklingshjelp, South Pacific Forum Fisheries Agency, Food and Agriculture Organization of the United Nations. 1998. The living marine resources of the Western Central Pacific, vol 1. Food and Agriculture Organization of the United Nations, Rome, Italy.
56. FAO. 2010. FishStat Plus: universal software for fishery statistical time series. FAO, Rome, Italy.
57. Sabin LR, Zheng Q, Thekkat P, Yang J, Hannon GJ, Gregory BD, Tudor M, Cherry S. 2013. Dicer-2 processes diverse viral RNA species. *PLoS One* 8:e55458. <https://doi.org/10.1371/journal.pone.0055458>.
58. Jayachandran B, Hussain M, Asgari S. 2012. RNA interference as a cellular defense mechanism against the DNA virus baculovirus. *J Virol* 86: 13729–13734. <https://doi.org/10.1128/JVI.02041-12>.
59. Meng X, Zhang X, Li J, Liu P. 2018. Identification and comparative profiling of ovarian and testicular microRNAs in the swimming crab *Portunus trituberculatus*. *Gene* 640:6–13. <https://doi.org/10.1016/j.gene.2017.10.026>.
60. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLoS Genet* 6:e1001191. <https://doi.org/10.1371/journal.pgen.1001191>.
61. Suzuki Y, Frangeul L, Dickson LB, Blanc H, Verdier Y, Vinh J, Lambrechts L, Saleh MC. 2017. Uncovering the repertoire of endogenous flaviviral elements in *Aedes* mosquito genomes. *J Virol* 91:e00571-17. <https://doi.org/10.1128/JVI.00571-17>.
62. Lequime S, Lambrechts L. 2017. Discovery of flavivirus-derived endogenous viral elements in *Anopheles* mosquito genomes supports the existence of *Anopheles*-associated insect-specific flaviviruses. *Virus Evol* 3:vev035. <https://doi.org/10.1093/ve/vev035>.
63. Savojardo C, Luchetti A, Martelli PL, Casadio R, Mantovani B. 2019. Draft genomes and genomic divergence of two *Lepidurus* tadpole shrimp species (Crustacea, Branchiopoda, Notostraca). *Mol Ecol Resour* 19: 235–244. <https://doi.org/10.1111/1755-0998.12952>.
64. Lobo FP, Mota BEF, Pena SDJ, Azevedo V, Macedo AM, Tauch A, Machado CR, Franco GR. 2009. Virus-host coevolution: common patterns of nucleotide motif usage in *Flaviviridae* and their hosts. *PLoS One* 4:e6282. <https://doi.org/10.1371/journal.pone.0006282>.
65. Karlin S, Mrzek J. 1997. Compositional differences within and between eukaryotic genomes. *Proc Natl Acad Sci U S A* 94:10227–10232. <https://doi.org/10.1073/pnas.94.19.10227>.
66. Beutler E, Gelbart T, Han JH, Koziol JA, Beutler B. 1989. Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. *Proc Natl Acad Sci U S A* 86:192–196. <https://doi.org/10.1073/pnas.86.1.192>.
67. Duan J, Antezana MA. 2003. Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. *J Mol Evol* 57:694–701. <https://doi.org/10.1007/s00239-003-2519-1>.
68. Suzuki R, Shimodaira H. 2006. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22:1540–1542. <https://doi.org/10.1093/bioinformatics/btl117>.
69. Liu H, Soderhall K, Jiravanichpaisal P. 2009. Antiviral immunity in crustaceans. *Fish Shellfish Immunol* 27:79–88. <https://doi.org/10.1016/j.fsi.2009.02.009>.
70. Asgari S. 2018. microRNAs as regulators of insect host-pathogen interactions and immunity. *Adv Insect Physiol* 55:19–45. <https://doi.org/10.1016/bs.aiip.2018.07.004>.
71. Barnard AC, Nijhof AM, Fick W, Stutzer C, Maritz-Olivier C. 2012. RNAi in arthropods: insight into the machinery and applications for understanding the pathogen-vector interface. *Genes (Basel)* 3:702–741. <https://doi.org/10.3390/genes3040702>.
72. Sanchez-Vargas I, Scott J, Poole-Smith B, Franz A, Barbosa-Solomieu V, Wilusz J, Olson K, Blair C. 2009. Dengue virus type 2 infections of *Aedes aegypti* are modulated by the mosquito's RNA interference pathway. *PLoS Pathog* 5:e1000299. <https://doi.org/10.1371/journal.ppat.1000299>.
73. Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, Olson KE, Blair CD. 2010. Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells. *PLoS Negl Trop Dis* 4:e848. <https://doi.org/10.1371/journal.pntd.0000848>.
74. Hess AM, Prasad AN, Pitsyn A, Ebel GD, Olson KE, Barbacioru C, Mo-

- nighetti C, Campbell CL. 2011. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. *BMC Microbiol* 11:45. <https://doi.org/10.1186/1471-2180-11-45>.
75. Zhang G, Etebari K, Asgari S. 2016. *Wolbachia* suppresses cell fusing agent virus in mosquito cells. *J Gen Virol* 97:3427–3432. <https://doi.org/10.1099/jgv.0.000653>.
 76. Lee M, Etebari K, Hall-Mendelin S, van den Hurk AF, Hobson-Peters J, Vutipally S, Schnettler E, Hall R, Asgari S. 2017. Understanding the role of microRNAs in the interaction of *Aedes aegypti* mosquitoes with an insect-specific flavivirus. *J Gen Virol* 98:892–1903. <https://doi.org/10.1099/jgv.0.000832>.
 77. Mejía-Ruiz CH, Vega-Peña S, Alvarez-Ruiz P, Escobedo-Bonilla CM. 2011. Double-stranded RNA against white spot syndrome virus (WSSV) vp28 or vp26 reduced susceptibility of *Litopenaeus vannamei* to WSSV, and survivors exhibited decreased susceptibility in subsequent re-infections. *J Invertebr Pathol* 107:65–68. <https://doi.org/10.1016/j.jip.2011.02.002>.
 78. Xu JY, Han F, Zhang XB. 2007. Silencing shrimp white spot syndrome virus (WSSV) genes by siRNA. *Antiviral Res* 73:126–131. <https://doi.org/10.1016/j.antiviral.2006.08.007>.
 79. Liu C, Li F, Sun Y, Zhang X, Yuan J, Yang H, Xiang J. 2016. Virus-derived small RNAs in the penaeid shrimp *Fenneropenaeus chinensis* during acute infection of the DNA virus WSSV. *Sci Rep* 6:28678. <https://doi.org/10.1038/srep28678>.
 80. Firth AE, Brierley I. 2012. Non-canonical translation in RNA viruses. *J Gen Virol* 93:1385–1409. <https://doi.org/10.1099/vir.0.042499-0>.
 81. Vollmer J. 2006. CpG motifs to modulate innate and adaptive immune responses. *Int Rev Immunol* 25:125–134. <https://doi.org/10.1080/08830180600743115>.
 82. Cheng XF, Virk N, Chen W, Ji SQ, Ji SX, Sun YQ, Wu XY. 2013. CpG usage in RNA viruses: data and hypotheses. *PLoS One* 8:e74109. <https://doi.org/10.1371/journal.pone.0074109>.
 83. Di Giallonardo F, Schlub TE, Shi M, Holmes EC. 2017. Dinucleotide composition in animal RNA viruses is shaped more by virus family than by host species. *J Virol* 91:e02381. <https://doi.org/10.1128/JVI.02381-16>.
 84. Colmant AMG, Hobson-Peters J, Bielefeldt-Ohmann H, van den Hurk AF, Hall-Mendelin S, Chow WK, Johansen CA, Fros J, Simmonds P, Watterson D, Cazier C, Etebari K, Asgari S, Schulz BL, Beebe N, Vet LJ, Piyasena TBH, Nguyen HD, Barnard RT, Hall RA. 2017. A new clade of insect-specific flaviviruses from Australian *Anopheles* mosquitoes displays species-specific host restriction. *mSphere* 2:e00262. <https://doi.org/10.1128/mSphere.00262-17>.
 85. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 86. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–U130. <https://doi.org/10.1038/nbt.1883>.
 87. Parry R, Asgari S. 2018. *Aedes anphevirus*: an insect-specific virus distributed worldwide in *Aedes aegypti* mosquitoes that has complex interplays with *Wolbachia* and dengue virus infection in cells. *J Virol* 92:e00224-18. <https://doi.org/10.1128/JVI.00224-18>.
 88. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
 89. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.
 90. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
 91. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320. <https://doi.org/10.1093/molbev/msn067>.
 92. Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195. <https://doi.org/10.1093/molbev/mst024>.
 93. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p 1–8. *In* 2010 Gateway Computing Environments Workshop (GCE). IEEE, New Orleans, LA.
 94. Lv JJ, Liu P, Wang Y, Gao BQ, Chen P, Li J. 2013. Transcriptome analysis of *Portunus trituberculatus* in response to salinity stress provides insights into the molecular basis of osmoregulation. *PLoS One* 8:e82155. <https://doi.org/10.1371/journal.pone.0082155>.
 95. Meng XL, Liu P, Jia FL, Li J, Gao BQ. 2015. *De novo* transcriptome analysis of *Portunus trituberculatus* ovary and testis by RNA-Seq: identification of genes involved in gonadal development. *PLoS One* 10:e0133659. <https://doi.org/10.1371/journal.pone.0128659>.
 96. Liu L, Fu Y, Zhu F, Mu C, Li R, Song W, Shi C, Ye Y, Wang C. 2018. Transcriptomic analysis of *Portunus trituberculatus* reveals a critical role for WNT4 and WNT signalling in limb regeneration. *Gene* 658:113–122. <https://doi.org/10.1016/j.gene.2018.03.015>.
 97. Yang YX, Wang JT, Han T, Liu T, Wang CL, Xiao J, Mu CK, Li RH, Yu FP, Shi HL. 2015. Ovarian transcriptome analysis of *Portunus trituberculatus* provides insights into genes expressed during phase III and IV development. *PLoS One* 10:e0138862. <https://doi.org/10.1371/journal.pone.0138862>.