Co-infection of two novel toti-like viruses *in Callinectes Sapidus* is restricted to its northern host range

Mingli Zhao[1], Lan Xu[2], Holly Bowers[3], Eric J. Schott[4]*

[1] Institute of Marine and Environmental Technology, University of Maryland Baltimore County, Maryland, MD, United States

[2] Department of Marine Biotechnology & Institute of Marine and Environmental Technology, University of Maryland Baltimore County, Baltimore, MD 21202, USA

[3] Moss Landing Marine Laboratory, San Jose State University, California, CA, United States

[4] Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Maryland, MD, United States

***Correspondence:**
Corresponding Author
schott@umces.edu

Keywords: **dsRNA virus, Totivirus, Crustacean, Disease ecology, Host habitat expansion, Next generation sequencing, Taxonomy, Climate**

Abstract

The advancement of high throughput sequencing has greatly facilitated the exploration of viruses that infect marine hosts. For example, a number of putative virus genomes belonging to the *Totiviridae* family have been described in crustacean hosts. However, there has been no characterization of the most recently discovered putative viruses beyond description of their genomes. In this study, two novel dsRNA virus genomes were discovered in the Atlantic blue crab (*Callinectes sapidus*) and further investigated. Sequencing of both virus genomes revealed that they each encode RNA dependent RNA polymerase proteins (RdRps) with similarities to toti-like viruses. The viruses were tentatively named *Callinectes sapidus* toti-like virus 1 (CsTLV1) and *Callinectes sapidus* toti-like virus 2 (CsTLV2). Both genomes have typical elements required for -1 ribosomal frameshifting, which may induce the expression of an encoded ORF1-ORF2 (gag-pol) fusion protein. Phylogenetic analyses of CsTLV1 and CsTLV2 RdRp amino acid sequences suggested that they are members of two new genera in the family *Totiviridae*. The CsTLV1 and CsTLV2 genomes were detected in muscle, gill and hepatopancreas of blue crabs by real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR). The presence of ~40 nm totivirus-like viral particles in all three tissues was verified by transmission electron microscopy, and pathology associated with CsTLV1 and CsTLV2 infections were observed by histology. PCR assays showed the prevalence and geographic range of these viruses to be restricted to the northeast US sites sampled. The two virus genomes co-occurred in almost all cases, with the CsTLV2 genome being found on its own in 8.5% cases, and the CsTLV1 genome not yet found on its own. To our knowledge, this is the first report of toti-like viruses in *C. sapidus*. The information reported here provides the knowledge and tools to investigate transmission and potential pathogenicity of these viruses.

INTRODUCTION

With the wide application of next generation sequencing (NGS), a huge number of virus genomes have been described from studies of metagenomes and viromes. How to best use the massive amount of data produced by NGS remains a fundamental challenge. For instance, an increasing number of toti-like virus sequences have been revealed by metagenomic studies, but further characterization and investigation is lacking (Shi et al., 2016). Ideally, following NGS-based discovery, attention should also be paid to characterize the biological properties of putative viruses, especially their genetics, viral morphological characteristics, geographic range, and potential impacts on hosts.

The *Totiviridae* family of viruses have a non-segmented, double-stranded RNA (dsRNA) genome, with two open reading frames (ORFs) encoding the putative capsid protein (Cp) and the RNA dependent RNA polymerase (RdRp). Most virions in this family are isometric with no projections, and are ~40 nm in diameter (Wickner et al., 2011). At present, five genera are officially recognized by the International Committee on Taxonomy of Viruses (ICTV), including *Totivirus*, *Victorivirus*, *Giardiavirus*, *Leishmaniavirus* and *Trichomonasvirus* (Wickner et al., 2011; King et al., 2012). Viruses belonging to the genera *Totivirus* and *Victorivirus* mainly infect fungi, whereas those in the genera *Giardiavirus*, *Leishmaniavirus* and *Trichomonasvirus* are present in parasitic protozoa and do not appear to cause cytopathic effects (Ghabrial et al., 2009; Goodman et al., 2011). Recently, non-ICTV recognized totivirus species have been found in arthropod hosts such as mosquitoes, ants, flies, as well as crustaceans (Poulos et al., 2006; Wu et al., 2010; Koyama et al., 2015; Koyama et al., 2016). Novel toti-like viruses have also been found in fish and plant hosts (Haugland et al., 2011; Abreu et al., 2015; Chen et al., 2015). Two genera were proposed recently, including *Artivirus* which infect arthropod and fish hosts (Zhai et al., 2010), and *Insevirus* which infect insect hosts (Zhang et al., 2018).

Two totiviruses have been reported to cause crustacean disease: Cherax *Giardiavirus*-like virus (CGV) in freshwater crayfish (*Cherax quadricarinatus*) and infectious myonecrosis virus (IMNV) in the Pacific white shrimp (*Litopenaeus vannamei*). CGV was the first totivirus identified in crustaceans and caused high morbidity and mortality in infected juvenile crayfish (Edgerton et al., 1994). IMNV is the most well studied totivirus in crustaceans, which causes infectious myonecrosis (IMN) in the Pacific white shrimp in Brazil and Indonesia (Lightner et al., 2004; Poulos et al., 2006; Senapin et al., 2007; Naim et al., 2014). Additionally, metagenomics studies have reported totivirus-like dsRNA genome sequences in sesarmid and charybdis crabs (GenBank accessions NC_032566.1 and NC_032462.1) but these have not had further characterization or investigation (Shi et al., 2016).

The Atlantic blue crab, *Callinectes sapidus*, is an adaptable estuarine species that functions as both predator and prey in food webs and supports important fisheries from the US mid-Atlantic coast to southern Brazil (Millikin et al., 1984; NOAA, 2020). *C. sapidus* has greatly expanded its geographic habitat range since the Last Glacial Maximum when the seas became warmer (Macedo et al., 2019), and has been introduced to Asia and Europe waters as an invasive species since 1901 (Millikin et al., 1984; Mancinelli et al., 2021). Unique within the *Callinectes* genus, *C. sapidus* has the ability to inhabit high latitudes by becoming dormant in winter. As the climate and ocean temperatures have changed, the distribution of *C. sapidus* has shifted poleward and the abundance of *C. sapidus* has increased at high latitudes, as far north as Nova scotia, Canada and as far south as Argentina (Gosner, 1978; Johnson et al., 2015; Piers, 1920).

Within the mid-Atlantic coast and Gulf of Mexico, a range of viruses have been described in blue crabs, in the families *baculoviridae*, *herpesviridae*, *reoviridae*, *picornaviridae*, *roniviridae*,

rhabdoviridae and *bunyaviridae* (Johnson 1978, 1983, 1984; Bowers et al., 2010; Shields et al., 2015; Zhao et al., 2021a, 2021b). With one exception, the relationship of these viruses to the blue crab range, and climate factors is unknown. *Callinectes sapidus* reovirus 1 (CsRV1), which causes disease and mortality in *C. sapidus*, is more prevalent in blue crabs at higher latitudes (Zhao et al., 2020), which illustrates that host-pathogen interactions can be strongly affected by habitat and environmental changes. Therefore, investigations of the effects of climate-related range extension and variation on host-pathogen interactions of other viruses will advance the understanding of drivers for virus epizootiology and ecology. The feasibility of such studies has been dramatically accelerated by molecular technologies of qPCR and high throughput sequencing, enabling virus discovery and tracking (Maclot et al., 2020).

Here, we report the discovery and characterization of two novel toti-like virus genomes that co-infect *C. sapidus* along the northern Atlantic coast of the US. Virus morphologies were visualized by TEM and pathology caused by the viruses were revealed by histology. Additionally, probe-based RT-qPCR assays were developed to screen and quantify totivirus infections in large numbers of *C. sapidus* across a climatological gradient.

MATERIALS AND METHODS

Crab Sampling

Blue crabs were collected from coastal states of the US, including Massachusetts (MA), Rhode Island (RI), New York (NY), Maryland (MD), Delaware (DE), North Carolina (NC), Texas (TX) and Louisiana (LA) between the years 2009 and 2021. A portion of crabs collected prior to 2020 were also used in the analysis of CsRV1 prevalence (Zhao et al., 2020). Crab sex and carapace width (CW, measured laterally spine-to-spine), sampling date and locations were recorded during collection. Whole crab or two walking legs removed from each crab were chilled on ice at the time of harvest. For molecular analysis, frozen specimens were then shipped to the Institute for Marine and Environmental Technology (IMET) in Baltimore, MD and stored at -20°C until further analyses. Live crabs from RI and Shinnecock Bay, NY collected in the year 2021 were shipped chilled to IMET where tissues were collected for virus purification, histology and electron microscope observations.

RNA extraction

RNA extractions were performed with sterile wooden rods and single-use razor blades on a bench cleaned with ELIMINase™. After the external cuticle was cleaned with ELIMINase™, approximately 50 mg of muscle and hypodermis was dissected from a walking leg and homogenized in 1.0 mL of homemade Trizol (Rodriguez-Ezpeleta et al., 2009), with a Savant FastPrep™ FP120 homogenizer (MP Biomedicals, Santa Ana, CA, USA). RNA extraction followed protocols used by Spitznagel et al. (2019). After Trizol-chloroform separation of RNA and precipitation with isopropanol, two 12,000 g centrifuge washes with 500 µL 75% ethanol were carried out. Resulting RNA pellets were dissolved in 50 µl 1 mM EDTA and stored at -80°C. Process control samples (muscle from frozen smelt) were extracted before and after sets of tested crab samples to monitor for cross contamination between each sample. RNA quality and concentration were determined by NanoDrop™ spectrophotometry (Thermo Scientific, Waltham, MA, USA). The dsRNA content of RNA extractions were revealed by electrophoresis on 1.0% agarose TBE gels stained with ethidium bromide. The isolated dsRNA was agarose gel purified with the NucleoSpin® Gel and PCR Clean-Up Kit (Takara Bio, San Jose, CA, USA).

DNA Library Construction and Illumina Sequencing

Purified dsRNA was used for cDNA synthesis with barcoded octamers (5'-GGCGGAGCTCTGCAGATATC-NNNNNNNN-3') with the M-MLV Reverse Transcriptase 1st-Strand cDNA Synthesis Kit (Biosearch Technologies, Hoddesdon, UK). The resulting cDNA was amplified by PCR using the barcode primers (5'-GGCGGAGCTCTGCAGATATC-3'). Amplification was achieved through 40 cycles of 95 °C for 5 s (denaturation), and 60 °C for 30 s (annealing), followed by 72 °C for 30 s (elongation). PCR products of 250-500 bp were obtained by agarose gel purification with a NucleoSpin® Gel and PCR Clean-Up Kit (Takara Bio, San Jose, CA, USA). The DNA library was constructed using the NEBNext R Ultra™ DNA Library Prep Kit for Illumina (NEB, Ipswich, MA, United States) following manufacturer's instructions (NEB, Ipswich, MA, USA). The library was sequenced in a 2 × 250 paired-end configuration on the Illumina MiSeq platform with a MiSeq Reagent kit v3 (Illumina, San Diego, CA, USA).

Sequence analyses

Sequencing barcodes were trimmed, and low quality and short reads were removed with CLC Genomics Workbench 9.5.2 (Qiagen, Hilden, Germany). The clean reads were collected and used for de novo assembly (Grabherr et al., 2011) with default settings (word-size = 45, Minimum contig length >=500). A preliminary set of contigs coding proteins of at least 150 amino acids were identified with ORF finder in CLC Genomics Workbench. ORFs of de novo derived contigs were used to search using the NCBI web server for non-redundant database using the BLASTp program. The conserved domains and motifs in the ORF were searched by NCBI Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Dotknot (Huang et al., 2005) was used to search the H-type pseudoknots with estimated free energy (EFE). Predicted RNA secondary structures were visualized by Pseudoviewer 2.5 (Byun and Han, 2006).

Rapid Amplification of cDNA Ends (RACE) and sequence verification

Terminal sequences were determined using a SMARTer® RACE 5'/3' Kit (Takara Bio, San Jose, CA, USA) with purified dsRNA as the initial template. The 3′ poly(A) tailing of RNA was performed at 37°C for 30 min using *E. coli* Poly(A) Polymerase (NEB, Ipswich, MA, United States). Poly (A)-tailed dsRNA was used for RACE First-strand cDNA synthesis of each terminus using 5′- or 3′ -CDS primers as described by the manufacturer (Takara Bio, San Jose, CA, USA). Then, 5′-RACE and 3′-RACE PCR amplification was performed with viral gene specific primers (GSP) and universal primers (UMP), and then a nested PCR was performed with gene specific primers short (GSPS) and universal primer short (UPS) (**Supplementary Table 1**) using Advantage 2 polymerase Mix (Takara Bio, San Jose, CA, USA). The conditions for amplification were 30 cycles at 94°C for 30 s, annealing at 68°C for 30 s, and elongation at 72°C for 2 min, with a final extension at 72°C for 10 min. Amplicons were then purified from the gel using NucleoSpin® Gel and PCR Clean-Up Kit (Takara Bio, San Jose, CA, USA), cloned into pGEM®-T Vector Systems (Promega Corporation, Madison, WI, USA), and sequenced. Sequence verification and filling of gaps between contigs were achieved with Sanger sequencing with primers in **Supplementary Table 1**. PCR conditions were 30 cycles at 94°C for 30 s, annealing at 56°C for 30 s, and elongation at 72°C for 90 s, with a final extension at 72°C for 10 min.

Sequence alignment and phylogenetic analyses

RNA-dependent RNA polymerase proteins (RdRps) from all reference sequences, and closest homologues from NCBI were aligned with RdRps of newly identified viruses in this study using MAFFT 7.0 (Katoh and Standley, 2013) with an accurate option (L-INS-i). The alignment was used for constructing the Maximum Likelihood (ML) phylogenetic tree in CLC Genomics Workbench.

RdRp amino acid sequence of *Helminthosporium victoriae* 145S virus (HvV-145S) (GenBank accession no. YP-052858) was used for the outgroup of the ML tree. Branch support values greater than 0.5 were shown in the tree. The accession numbers of the proteins and the corresponding virus names and acronyms are shown in **Supplementary Table 2**.

Reverse Transcription quantitative PCR (RT-qPCR) development

To screen CsTLV1 and CsTLV2 infections in *C. sapidus*, a probe-based RT-qPCR assay was developed with primer pairs designed to detect a 193-bp region of CsTLV1 genome and a 183-bp region of CsTLV2 (**Table 1**) simultaneously. Probes designed for detecting CsTLV1 and CsTLV2 were also shown in **Table 1**. DsRNA standards were created by *in vitro* RNA synthesis. In brief, PCR products amplified by the primer pairs mentioned above were purified and cloned into pGEM®-T Vector Systems (Promega Corporation, Madison, WI, USA). Plasmids containing the targeted region were used as templates to synthesize each strand of the viral RNA standards by T7 or Sp6 RNA polymerase, respectively (Sigma-Aldrich, St. Louis, MO, USA). Viral RNAs were then quantified and annealed into dsRNA on ice, and serially diluted in 25 ng per µL yeast tRNA carrier. Standard curves were generated by RT-qPCR amplifications of a 10-fold dilution series of synthesized dsRNA containing 10 to 10e6 genome copies per µl. The qPCR cycling contained qScript® Virus 1-Step ToughMix® (Quantabio, Beverly, MA, USA) in 10 µl reactions with 0.25 µM each primer and 0. 25µM each probe for both genomes. To anneal PCR primers to dsRNA, primers and extracted RNA were combined, heated to 95°C for 5 min then cooled to 4°C prior to being added to the reverse transcriptase and Taq polymerase reaction mixture. Reverse transcription and amplification conditions were 50°C for 10 min (reverse transcription) followed by 1 min at 95°C (reverse transcriptase inactivation and template denaturation). Amplification was achieved through 40 cycles of 95°C for 10 s, and 61°C for 30 s. Gene target copies were then calculated as copies per mg of crab muscle, and samples with greater than 100 copies per mg were recorded as CsTLV1/2 positive, which was based on empirical observations of cross contamination in process control RNA extractions.

Statistical analyses

All statistical tests were conducted using RStudio 1.1.456 (R Core Team, 2019). Significant correlations were defined as those where $p \leq 0.05$. To determine whether CsTLV1 and CsTLV2 infections were correlated with sex, crab size or latitude, binomial (infected vs non-infected) generalized logistic regression models (GLM) were conducted (alpha = 0.05). Akaike's information criterion (AIC) was used to choose the best GLM to determine which factors best correlate with CsTLV1 and CsTLV2 prevalence and intensity (Aho et al., 2014). The Pearson correlation was used to test the correlation between the variables (Kirch, 2008).

Histology and Electron Microscopy

C. sapidus collected from RI and Long Island (NY) with CsTLV1 and CsTLV2 confirmed and quantified by Rt-qPCR were dissected with muscle, gill and hepatopancreas tissue removed for further examination for virus presence. For histological analyses, the tissues were fixed in Bouin's solution at 4 °C overnight, and then placed into 75% ethanol for storage. Preserved tissues were processed according to the standard operating procedures for embedding, sectioning and Hematoxylin and Eosin (H&E) staining. Slides were then observed with an Echo Revolve Microscope (San Diego, CA, USA).

For electron microscopy examination, crab tissues were immersion fixed in fixative buffer (2% paraformaldehyde, 2.5% glutaraldehyde, 2 mM CaCl2 in 0.1 M PIPES Buffer, pH 7.35) at 4 °C

overnight. Tissue fragments were then trimmed into ~1 mm^3 cubes, post-fixed with 1% osmium tetroxide, washed in water and stained *en bloc* with 1% (w/v) uranyl acetate for 1h. Specimens were then washed and dehydrated using 30%, 50%, 70%, 90% and 100% ethanol in series. After dehydration, specimens were embedded in Araldite-Epoxy resin (Araldite, Embed 812, Electron Microscopy Sciences, Hatfield, PA, USA). Ultrathin sections (~ 70 nm) were cut and examined in a Tecnai T12 transmission electron microscope (FEI, Hillsboro, OR) operated at 80KV. Digital images were acquired using an AMT bottom mount CCD camera and AMT600 software (Advanced Microscopy Techniques, Woburn, MA, USA). Crab samples infected with only CsTLV2 were not preserved well enough to be included in TEM and histology examinations.

RESULTS

Screening for the presence of putative viral dsRNA: RNA extraction

In a search for viral dsRNA in blue crabs, RNA extracted from leg muscle was analyzed on agarose gels. The RNA of 31% of sampled of crabs (9 of 29) harvested from the Agawam River, MA in the summer of 2008, showed two prominent dsRNA bands, termed dsRNA-S and dsRNA-L. The apparent molecular weights of the two dsRNA segments were ~6.5 kbp and ~7.5 kbp, respectively (**Figure 1**).

Sequence Analyses of dsRNA-S and dsRNA-L:

Total dsRNA, containing both dsRNA-S and dsRNA-L was sequenced. Illumina paired end reads (8377) were trimmed, quality filtered and assembled into contigs. The longest contig was 4016 nucleotides (nt) in length. ORFs were identified by BlastX and BlastN comparisons with GenBank using default parameters. Two totivirus-like genome sequences were identified from the post-assembly contigs. The 5' - and 3' untranslated regions (UTRs) of both genomes were obtained by RACE PCR and Sanger sequencing, to reveal two toti-like sequences of 6444 nt and 7421 nt, and designated as CsTLV1 and CsTLV2 respectively. Each genome contained two ORFs (ORF1 and ORF2) encoding Cp and RdRp proteins, respectively (**Figure 2**).

The predicted Cp and RdRp proteins of CsTLV1 and CsTLV2 showed limited amino acid sequence identity with each other (21% for Cp and 27% for RdRp, respectively). CsTLV1 RdRp amino acid sequences showed >33% identity with the corresponding predicted RdRp of Beihai barnacle virus 15, Ahus virus, Parry's Creek toti-like virus 1 and Diatom colony associated dsRNA virus 17 genome (Shi et al., 2016; Urayama et al., 2016; Pettersson et al., 2019; Williams et al., 2020). CsTLV2 RdRp amino acid sequence showed >40% identity with the RdRp encoded in *Plasmopara viticola* lesion associated totivirus-like 5, Hubei toti-like virus 5 and Beihai sesarmid crab virus 7 (Shi et al., 2016; Chiapello et al., 2020). A search of the conserved domain database (CDD) and multiple protein alignment confirmed that the predicted RdRp domains of CsTLV1/2 contain eight conserved motifs (I to VIII), including the GDD motif, which are the typical of functional virus RdRps **(Figure 3)**. Sequence analyses of CsLTV1/2 indicated that there is an overlap region between ORFs 1 and 2 (**Figure 2**), that in other viruses allows ORF2 to be translated as a fusion protein with ORF 1 through a -1 ribosomal frameshift motif 'GGAUUUU' at 3199 - 3205 nt positions in CsTLV1, and "AAGAAAA" at positions of 2685 - 2691 nt in CsTLV2. An H-type pseudoknot structure was detected in the downstream of each putative slippery site at positions 3227 - 3259 nt of CsTLV1 genome and 2697 - 2760 of CsTLV2 (**Figure 2**).

Phylogeny of CsTLV1 and CsTLV2

A maximum likelihood phylogenetic tree was used to show the relationships between CsTLV1/2 and other selected virus species in *Totiviridae*. As shown in **Figure 4**, RdRp amino acid sequence

multiple alignments of CsTLV1 and CsTLV2 and the corresponding toti and toti-like viral sequences revealed that CsTLVs is most closely related to, but distinct from, *Totivirus*, and *Artivirus* which is a proposed genus that includes IMNV and IMNV-like viruses (Zhai et al., 2010). CsTLV1 and the five toti-like viruses with the highest identity from GenBank formed a cluster in the tree, adjacent to but different from the cluster of CsTLV2 and its close toti-like virus species.

RT-qPCR assay performance

The probe-based RT-qPCR assay consistently detected as few as 10 copies of the target when tested on a dilution series of synthesized dsRNA standards (**Table 2**). Efficiency and sensitivity of the assay were evaluated by running10 RT-qPCR standard curves. The mean slope was 3.29 with a standard deviation of 0.06 for CsTLV1, and the mean slope was 3.25 with a standard deviation of 0.07 for CsTLV2. There is an average efficiency of 100.1% and 100.3% for CsTLV1 and CsTLV2 respectively, under typical use with the synthesized dsRNA standard.

Prevalence of infections and co-infections of CsTLV1 and CsTLV2

The prevalence of CsTLV1 and CsTLV2 was investigated by RT-qPCR in 875 crabs from the US Atlantic and Gulf of Mexico coasts. CsTLV1 and 2 infections were detected in the northern states (MA, RI, and NY) but not the lower latitudinal Atlantic states (DE, MD, and NC, n=299) and Gulf states (LA and TX, n=80). CsTLV1 and CsTLV2 infections in *C. sapidus* were detected in all three estuaries sampled in MA: Agawam River, Acushnet River and Falmouth (**Figure 5**; **Table 3**). CsTLV1 RNA prevalence in *C. sapidus* sampled from MA (n = 198) varied from 11.8% (Agawam River; n=127) to 30.6% (Acushnet River; n=49), and CsTLV2 RNA prevalence ranged from 12.6% to 36.7%. In crabs from RI, CsTLV1 prevalence was 27.6% and CsTLV2 prevalence was 32.8%. Viral RNA was detected in crab specimens collected from Moriches Bay, Shinnecock Bay and Napeague Bay in Long Island, NY (n=133) with an average prevalence of 42.1% and 43.6% for CsTLV1 and CsTLV2, respectively. CsTLV1 RNA was detected in Georgica Pond, NY, with low prevalence in 2012 (5.5%; n=18), but not in 2013, 2020 or 2021, and CsTLV2 RNA was detected in 2012 (5.5%; n=18) and 2013 (5.2%; n=19), but not in 2020 (n = 33) or 2021 (n=37).

Overall, CsTLV1 was never observed in the absence of CsTLV2, and co-infections of CsTLV1 were detected in 91.5% (107/117) of CsTLV2 positive specimens (Table 3). The dsRNA copy number per mg muscle ranged from 6.5×10e2 to 1.2 ×10e8 for CsTLV1, and 1.3×10e2 to 6.3×10e8 for CsTLV2.

Association of CsTLV1 and CsTLV2 genome presence with latitude, and crab traits

A binomial (infected vs. non-infected) generalized linear model (GLM) was used to test whether latitude, crab size or sex could predict CsTLV1 and CsTLV2 infection status (**Table 4**). The 415 male and 140 female specimens, from 20 mm to 196 mm in CW, were used for GLM analysis. Specimens that were PCR-positive for CsTLV1 and CsTLV2 ranged from 29 to 150 mm in CW. Prevalence for male and female crabs was 12.0% and 10.7%, respectively. Pearson correlation tests showed no significant correlation between latitude and crab size or sex. The full model analyzing the association between CsTLV1 and CsTLV2 prevalence and latitude, crab sex and carapace width (CW), differed significantly from null models ($P < 0.01$), in which latitude and crab size (CW) were significant fixed effects ($P < 0.01$). The reduced model, including only the association between CsTLV1 and CsTLV2 prevalence and latitude, sex or CW, reinforced that latitude and CW were the significant factors correlated with CsTLV1 and CsTLV2 prevalence ($P < 0.01$) (**Table 4**). CsTLV1 and CsTLV2 prevalence was positively related to latitude in the reduced model (slope is 2.33 for CsTLV1 and 2.38 for CsTLV2; $P < 0.01$), and CW showed a negative association with CsTLV1 and

CsTLV2 prevalence (slope = -1.00 for CsTLV1 and slope = -1.13 for CsTLV2; $P < 0.01$). In both full and reduced models, the association between CsTLV1/CsTLV2 prevalence and crab sex was not significant ($P > 0.1$).

Electron microscopy: Observation of viral particles

Crabs assessed to be infected with CsTLV1 and CsTLV2 by RT-qPCR were selected for transmission electron microscope (TEM) observation (**Supplementary Fig 1**). TEM revealed the presence of isometric virus particles, with a diameter of ~40 nm in *C. sapidus* muscle, gill and hepatopancreas tissues (**Figure 6**). Completed virions were present in the connective tissue and hemocytes of these tissues. We observed putative viroplasm in the gill of CsTLV1 and CsTLV2 infected crab and packed arrays of mature virions in the hepatopancreas of infected crab.

Histology of CsTLV1 and CsTLV2 infected blue crab tissues

Histological analysis of muscle, hepatopancreas and gills tissues of crabs naturally infected with CsTLV1 and CsTLV2 showed necrosis and hemocyte infiltration. Skeletal muscle in normal uninfected crabs is generally smooth, striated and with few circulating hemocytes (**Figure 7A**). Infected skeletal muscle had general necrosis and showed vacuolated areas with increased numbers of circulating hemocytes (**Figure 7B**). Hepatopancreas tubules in normal uninfected crabs have defined outer membranes and moderate numbers of circulating hemocytes circulating within hemal spaces between tubules (**Figure 7D**). Infected hepatopancreas often showed massive hemocytic infiltration (**Figure 7E**). Gills of normal uninfected crabs have moderate numbers of circulating hemocytes in hemal spaces (**Figure 7G**). Infected gills had considerably increased numbers of circulating hemocytes within necrotic areas (**Figure 7H**). At higher magnification, infected hemocytes in muscle, hepatopancreas and gills often had pyknotic or karyorrhectic nuclei (magenta arrows) as well as opaque, slightly eosinophilic intracytoplasmic inclusion bodies (blue arrows) (**Figure 7C, F and I**).

DISCUSSION

Molecular approaches for discovery of virus-like genomes have verified that viruses are an important and universal feature of the life history of marine organisms (Munn, 2006; Suttle, 2007). Beyond the discovery of new viruses, the characterization of these newly discovered viruses contributes to better understanding of their diversity, evolution, and ecology in marine environments. Partial toti-like virus sequences reported in some crabs by metagenomics (e.g., in Shi et al., 2016), have only detected the virus-like genome element, but not the actual presence of viral infections. In this study, we sequenced and characterized the genomes of two new putative *C. sapidus* totiviruses - CsTLV1 and CsTLV2, and showed that viral particles are present in tissues of infected crabs and cause pathology, signifying that these newly discovered viruses may affect the blue crab populations. This study is the first description of an endemic infection of totivirus in *C. sapidus* along the north Atlantic coast of the US.

Both CsTLV1 and CsTLV2 genomes contained two ORFs encoding the conserved domains of Cp and RdRp, respectively. Moreover, the two viruses contain a -1 ribosomal frameshifting in their genomes (**Figure 2**), which could facilitate the translation of ORF1 and ORF2 as a fusion polyprotein (Dinman et al., 1991). The predicted ORF2 coding strategy of CsTLV1/2 was consistent with other viruses in the family *Totiviridae*, such as *Saccharomyces cerevisiae* virus L-A (ScVL-A) (Dinman et al., 1991) and infectious myonecrosis virus (IMNV) (Nibert et al., 2007). CsTLV1 and CsTLV2 have all three elements that are required to accomplish -1 ribosomal frameshifting in RNA viruses: a slippery heptamer motif, an RNA pseudoknot shortly downstream of the site and a short spacer region between the slippery site and the pseudoknot (Rice et al., 1985; Dinman et al., 1991; Khalifa

et al., 2019). The classical slippery site sequence is 'XXXYYYZ' (where X is A/C/G/U, Y is A/U, and Z is A/C/U) within the overlapping region (Bekaert and Rousset, 2005). The slippery site of CsTLV1 (GGAUUUU) is the same to the slippery heptamer nucleotides found in other totiviruses, such as *Xanthophyllomyces dendrorhous* viruses (GGAUUUU) (Baeza et al., 2012), *Puccinia striiformis* totiviruses (PsTVs; GGG/AUUUU) (Zheng et al., 2017) and red clover powdery mildew-associated totiviruses (RPaTVs; GGG/AUUUU) (Kondo et al., 2016). Meanwhile, the slippery site is "AAGAAAA" in CsTLV2, which is the same as that used by plant associated astro-like virus (Lauber et al., 2019).

In the current ICTV scheme of totivirus taxonomy, 50% sequence identity of Cp/RdRp proteins is generally considered a threshold to define different species (Wickner et al., 2011). CsTLV1 and CsTLV2 share only 21% identity for Cp and 27% for RdRp, indicating they are distinct species in the family *Totiviridae*. Phylogenetic analyses of RdRp amino acid sequences showed that CsTLV1 and CsTLV2 formed a distinct branch from other genera in the family *Totiviridae* but clustered into two subgroups (**Figure 4**). CsTLV1, together with toti-like viruses identified from arthropod and crustacean hosts were classified into one group (Shi et al., 2016), and CsTLV2 formed another group with totiviruses sequenced from spirurian nematodes, sesarmid crab, and razor shell clam (Shi et al., 2016). Compared to other genera of the family *Totiviridae*, members of CsTLV1-like and CsTLV2-like groups have the highest similarity between each other. Taken together with their genome structure and phylogenetic position, CsTLV1 and CsTLV2 may represent two new viral species within two novel genera of the family *Totiviridae*.

Co-infection by two distinct viruses has been reported in *C. sapidus* such as reovirus and RhVA (Johnson, 1978, 1983), and bunya-like virus (Zhang et al., 2004). Co-infection of distinct totiviruses has also been commonly reported, such as in *Sphaeropsis sapinea* and *Chalara elegans* (Preisig et al., 1998; Park et al., 2005). Recently, co-infection of three dsRNA viruses *T. vaginalis* virus (TVV1, TVV2 and TVV3) were revealed (Bokharaei-Salim et al., 2020). Co-infection of CsTLV1 was detected in more than 90% CsTLV2-positive specimens (**Figure 6**; **Table 3**), suggesting that there is a significant relationship between these two totiviruses in *C. sapidus*. Interestingly, although independent infection of CsTLV2 was identified, no crab was ever found to contain the CsTLV1 genome alone. One possible explanation for this observation may be that the CsTLV1 genome or virus cannot replicate or be encapsulated in the absence of CsTLV2. The relationship between CsTLV1 and 2 does not have the characteristics of defective virus genomes (Vignuzzi and López, 2019); the CsTLV1 genome does not have obvious deletions or frame shifts, although the CsTLV2 genome is over 1000 nt longer than the CsTLV1 genome. A similar phenomenon has been revealed that HvV190S (*Totiviridae*) and HvV145S (*Chrysoviridae*) co-infect the pathogenic fungus *Helminthosporium victoriae.* HvV145S has never been found alone but is always associated with HvV190S virus. HvV145S was originally thought to be the cause of the diseases, however, a recent study suggested that HvV190S alone is the cause of diseases, and the co-infection is not required (Xie et al., 2016). In our study, TEM of co-infected blue crabs, revealed all virions had a diameter of ~40 nm, suggesting that either CsTLV1 is indistinguishable in size or appearance from CsTLV2, or that only one of the viruses produces virions.

Most members of *Totiviridae* infecting fungi and protozoans lack extracellular transmission; instead, they are transmitted vertically during cell division, sporogenesis and cell fusion (Ghabrial et al., 2009). However, some totiviruses with fiber-like protrusions on their surface, such as IMNV and Omono River virus (OmRV), are capable of extracellular transmission in their metazoan hosts (Poulos et al., 2006; Tang et al., 2008; Dantas et al., 2015; Shao at al., 2021). The transmission mechanism for CsTLV1 and CsTLV2 in the blue crab is yet unknown. Attempts to transmit the

viruses by injections (CsTLV1/2) to naïve crabs have been so far unsuccessful (Zhao and Schott, unpublished data). Necrosis and massive hemocyte infiltration in CsTLV1 and CsTLV2 infected muscle, gill and hepatopancreas indicated that the viruses are detrimental to the health of blue crabs. CsTLV1 and CsTV2 infections were negatively correlated with crab size in GLM analyses, which suggested that juveniles may be more susceptible to infection, or that older animals infected with CsTLV1 and CsTLV2 either die or clear the virus as they mature or age. All these results provide the fundamental knowledge for future studies to investigate how these viruses are transmitted and how they affect the ecology of blue crabs.

The significant correlation between CsTLV1 and CsTLV2 infections and latitude has also been identified in another blue crab dsRNA virus-CsRV1, which also showed significantly higher prevalence at higher latitudinal locations compared to lower latitudes (Flowers et al., 2016, 2018; Zhao et al., 2020). However, compared to the wide geographic range of CsRV1 infections in blue crabs, infections of CsTLV1 and CsTLV2 were restricted to the most northeastern estuaries we sampled in MA, RI and NY, but absent from the lower latitudinal estuaries of DE, MD, NC, LA and TX. Although factors driving the emergence of viruses and the gradient of virus prevalence at different geographic locations could be complex, two likely covariates in our study are water temperature and length of the active period for blue crabs, which have strong correlations to latitudes (Zhao et al., 2020). It is notable that the virus is present in crabs at the northern edge of their geographic range. Microbiome community changes and emergence of novel pathogens have been widely reported during the dispersal of host invasion and extension range (Engering et al., 2013; Dragicevic et al., 2021). The extensive poleward expansion of *C. sapidus* in its native range along the western Atlantic and its successful invasion to European waters (Johnson 2015; Mancinelli et al., 2021), make *C. sapidus* a well-suited model to study virus evolution, diversity, and viral ecology of marine animals during host habitat expansion and invasion. In Rhode Island, state managers are beginning to survey blue crab abundance in anticipation of a growing commercial and recreational fishery (K. Rodigue, personal communication). Therefore, further systematic and comprehensive studies on the virome of *C. sapidus,* including CsTLV1 and CsTLV2, at different geographical locations are urgently needed for a better understanding of the virus ecology and epidemiology with the host habitat expansion.

In conclusion, two putative viral dsRNA sequences in *C. sapidus* were characterized with NGS, and shown to be associated with virus particles and histopathology. Based on their genomic organizations, phylogenetic relationships, and conserved motifs, the viruses are tentatively named CsTLV1 and CsTLV2, and proposed to be members of two new genera in the family *Totiviridae*. This study demonstrated how discovery and investigation of putative virus genomes in *C. sapidus* can proceed to discovery of new viruses and will hopefully inspire additional multi-disciplinary studies on viral diversity and ecology in marine organisms.

Data Availability Statement

CsTLV1 and CsTLV2 genome sequences are available in GenBank under the accession numbers - OL456199 and OL456200, respectively.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

EJS and MZ designed the experiments; MZ, LX and HB performed the experiments; MZ and EJS analyzed the results and drafted the manuscript. EJS, LX, HB revised the paper.

Funding

This work was supported by NSF Division of Ocean Sciences - Biological Oceanography awards 1658466 (ES, LP), Maryland Sea Grant - Graduate student research support grant (ES, MZ). MZ and LX were supported by awards from the China Scholarship Council.

Acknowledgments

We thank many managers, fishermen and collaborators for assistance in collecting and processing samples, especially Bruce Estrella (MA Division of Marine Fisheries), Christopher Paparo (Stony Brook University, School of Marine and Atmospheric Sciences), Katie Rodrigue and Patrick Barrett (Rhode Island Department of Environmental Management). We thank Gretchen Messick (Cooperative Oxford Laboratory, NOAA) for assistance in diagnoses and description of histology specimens. We thank USDA scientists Janet Warg (APHIS) and John Neil (ARS) who generated preliminary NGS sequences of the mysterious Agawam River dsRNA; Olivia Pares (UMCES-Institute of Marine Environmental Technology) and Jennifer Akanoh (UMBC) for their help with the experiments. We are very grateful to Ten-Tsao Wong (UMBC-IMET) for experiment ideas on RACE. We also thank the Electron Microscopy Core Imaging Facility and Histology core at University of Maryland Baltimore (UMB) for helping with TEM and Histology.

Supplementary Material

The Supplementary Material for this article can be found online.

References

Abreu, E. F., Daltro, C. B., Nogueira, E. O., Andrade, E. C., and Aragao, F. J. (2015). Sequence and genome organization of papaya meleira virus infecting papaya in Brazil. *Arch. Virol.* 160(12), 3143-3147. doi: 10.1007/s00705-015-2605-x

Aho K, Derryberry D, Peterson T (2014) Model selection for ecologists: The worldviews of AIC and BIC. *Ecology* 95:631-636. doi: 10.1890/13-1452.1

Baeza, M., Bravo, N., Sanhueza, M., Flores, O., Villarreal, P., and Cifuentes, V. (2012). Molecular characterization of totiviruses in *Xanthophyllomyces dendrorhous*. *Virol. J.* 9:140. doi: 10.1186/1743-422X-9-140

Bekaert, M., and Rousset, J. P. (2005). An extended signal involved in eukaryotic− 1 frameshifting operates through modification of the E site tRNA. *Mol. Cell* 17(1), 61-68. doi: 10.1016/j.molcel.2004.12.009

Bokharaei-Salim, F., Esteghamati, A., Khanaliha, K., Esghaei, M., Donyavi, T., and Salemi, B. (2020). The First Detection of Co-Infection of Double-Stranded RNA Virus 1, 2 and 3 in Iranian Isolates of *Trichomonas vaginalis*. *Iran J. Parasitol.* 15(3), 357. doi: 10.18502/ijpa.v15i3.4200

Bowers, H. A., Messick, G. A., Hanif, A., Jagus, R., Carrion, L., Zmora, O., et al. (2010). Physicochemical properties of double-stranded RNA used to discover a reo-like virus from blue crab *Callinectes sapidus*. *Dis. Aquat. Org.* 93(1), 17-29. doi: 10.3354/dao02280

Byun, Y., and Han, K. (2006). PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures. *Nucleic Acids Res.* 34 (suppl_2), W416-W422. doi: 10.1093/nar/gkl210

Chen, S., Huang, Q., Wu, L., and Qian, Y. (2015). Identification and characterization of a maize-associated *mastrevirus* in China by deep sequencing small RNA populations. *Virol. J.* 12(1), 1-9. doi: 10.1186/s12985-015-0384-3

503 Chiapello, M., Rodríguez-Romero, J., Ayllón, M. A., and Turina, M. (2020). Analysis of the virome
504 associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus Evol.* 6(2),
505 veaa058. doi: 10.1093/ve/veaa058
506 Dantas, M. D. A., Chavante, S. F., Teixeira, D. I. A., Lima, J. P. M., and Lanza, D. C. (2015).
507 Analysis of new isolates reveals new genome organization and a hypervariable region in
508 infectious myonecrosis virus (IMNV). *Virus Res. 203*, 66-71. doi: 10.1016/j.virusres.2015.03.015
509 Dinman, J. D., Icho, T., and Wickner, R. B. (1991). A-1 ribosomal frameshift in a double-stranded
510 RNA virus of yeast forms a gag-pol fusion protein. *Proc. Natl. Acad. Sci.* 88(1), 174-178. doi:
511 10.1073/pnas.88.1.174
512 Edgerton, B., Owens, L., Giasson, B., and De Beer, S. (1994). Description of a small dsRNA virus
513 from freshwater crayfish *Cherax quadricarinatus*. *Dis. Aquat. Org.* 18, 63-63. doi:
514 10.3354/dao018063
515 Engering, A., Hogerwerf, L. and Slingenbergh, J., 2013. Pathogen–host–environment interplay and
516 disease emergence. *Emerg. microbes & infect*. 2(1), 1-7. doi: 10.1038/emi.2013.5
517 Flowers E.M., Simmonds K., Messick G.A., Sullivan L., and Schott E. J. (2016) PCR-based
518 prevalence of a fatal reovirus of the blue crab, *Callinectes sapidus* (Rathbun) along the northern
519 Atlantic coast of the USA. *J. Fish Dis.* 39, 705-714. doi: 10.1111/jfd.12403
520 Flowers, E.M., Johnson, A.F., Aguilar, R. and Schott, E. J. (2018). Prevalence of the pathogenic
521 crustacean virus *Callinectes sapidus* reovirus 1 near flow-through blue crab aquaculture in
522 Chesapeake Bay, USA. *Dis. Aquat. Organ*. 129(2), 135-144. doi: 10.3354/dao03232
523 Gao, Z., Cai, L., Liu, M., Wang, X., Yang, J., An, H., et al. (2021). A novel previously undescribed
524 fusarivirus from the phytopathogenic fungus *Setosphaeria turcica*. *Arch. Virol.* 166(2), 665-669.
525 doi: 10.1007/s00705-021-04954-x
526 Ghabrial, S. A., and Suzuki, N. (2009). Viruses of plant pathogenic fungi. *Annu. Rev. Phytopathol*.
527 47, 353–384. doi: 10.1146/annurev-phyto-080508-081932
528 Goodman, R. P., Ghabrial, S. A., Fichorova, R. N., and Nibert, M. L. (2011). Trichomonasvirus: a
529 new genus of protozoan viruses in the family *Totiviridae*. *Arch. Virol*. 156, 171–179. doi:
530 10.1007/s00705-010-0832-838
531 Gosner, K.L., 1978. A field guide to the Atlantic seashore. The Peterson field guide series.
532 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-
533 length transcriptome assembly from RNA-Seq data without a reference genome. *Nat.*
534 *Biotechnol.* 29(7), 644. doi: 10.1038/nbt.1883
535 Haugland, Ø., Mikalsen, A. B., Nilsen, P., Lindmo, K., Thu, B. J., Eliassen, T. M., et al. (2011).
536 Cardiomyopathy syndrome of Atlantic salmon (Salmo salar L.) is caused by a double-stranded
537 RNA virus of the *Totiviridae* family. *J. Virol.* 85(11), 5275. doi: 10.1128/JVI.02154-10
538 Huang, C. H., Lu, C. L., and Chiu, H. T. (2005). A heuristic approach for detecting RNA H-type
539 pseudoknots. *Bioinformatics* 21(17), 3501-3508. doi: 10.1093/bioinformatics/bti568
540 Johnson, D.S., 2015. The savory swimmer swims north: a northern range extension of the blue crab
541 *Callinectes sapidus*? *J. Crust. Biol.* 35(1), 105-110. doi: https://doi.org/10.1163/1937240X-
542 00002293
543 Johnson, P. T. (1978) Viral diseases of the blue crab, *Callinectes sapidus*. *Mar. Fish. Rev*., 40: 13-15.
544 Johnson, P. T. (1983) Diseases caused by viruses, rickettsiae, bacteria, and fungi. In: A. J.
545 Provenzano (ed.), The biology of Crustacea, 6, *Pathology*, 1-78. (Academic Press, New York.)
546 Johnson, P.T. (1984). Viral diseases of marine invertebrates. *Helgolander Meeresunters* 37(1), 65-98.
547 https://doi.org/10.1007/BF01989296
548 Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
549 improvements in performance and usability. *Mol. Biol. Evol.* 30(4), 772-780. doi:
550 10.1093/molbev/mst010

Khalifa, M. E., and MacDiarmid, R. M. (2019). A novel totivirus naturally occurring in two different fungal genera. *Front. microbiol.* 10, 2318. doi: 10.3389/fmicb.2019.02318

Kirch W (ed) (2008) Pearson's correlation coefficient, 1090–1091. Springer, Dordrecht (2008). https://doi.org/10.1007/978-1-4020-5614-7_2569

King, A. M. Q., Adams, M. J., Carstens, E. B., and Lefkowitz, E. J. (2011). Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier.

Kondo, H., Hisano, S., Chiba, S., Maruyama, K., Andika, I. B., Toyoda, K., et al. (2016). Sequence and phylogenetic analyses of novel totivirus-like double-stranded RNAs from field-collected powdery mildew fungi. *Virus Res.* 213, 353–364. doi: 10.1016/j.virusres.2015.11.015

Koyama, S., Sakai, C., Thomas, C. E., Nunoura, T., and Urayama, S. I. (2016). A new member of the family *Totiviridae* associated with arboreal ants (*Camponotus nipponicus*). *Arch. Virol.* 161(7), 2043-2045. doi: 10.1007/s00705-016-2876-x

Koyama, S., Urayama, S. I., Ohmatsu, T., Sassa, Y., Sakai, C., Takata, M., et al. (2015). Identification, characterization and full-length sequence analysis of a novel dsRNA virus isolated from the arboreal ant *Camponotus yamaokai*. *J. Gen. Virol. 96*(7), 1930-1937. doi: 10.1099/vir.0.000126

Lauber, C., Seifert, M., Bartenschlager, R., and Seitz, S. (2019). Discovery of highly divergent lineages of plant-associated astro-like viruses sheds light on the emergence of potyviruses. *Virus Res.* 260, 38-48. doi: 10.1016/j.virusres.2018.11.009

Lightner, D. V., Pantoja, C. R., Poulos, B. T., Tang, K. F. J., Redman, R. M., Andrade, T. P. D., et al. (2004). Infectious myonecrosis: new disease in Pacific white shrimp. *Glob. Aquac. Advocate 7*, 85.

Macedo, D., Caballero, I., Mateos, M., Leblois, R., McCay, S. and Hurtado, L.A., 2019. Population genetics and historical demographic inferences of the blue crab *Callinectes sapidus* in the US based on microsatellites. *PeerJ* 7, p.e7780.

Maclot, F., Candresse, T., Filloux, D., Malmstrom, C.M., Roumagnac, P., van der Vlugt, R., et al. 2020. Illuminating an ecological blackbox: using high throughput sequencing to characterize the plant virome across scales. *Front. Microbiol.* 11, p.2575. doi: 10.3389/fmicb.2020.578064

Mancinelli, G., Bardelli, R. and Zenetos, A., 2021. A global occurrence database of the Atlantic blue crab *Callinectes sapidus*. Scientific Data, 8(1), 1-10. doi: 10.1038/s41597-021-00888-w

Millikin, M.R., 1984. Synopsis of biological data on the blue crab, *Callinectes sapidus* Rathbun (No. 138). National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

Munn, C.B., 2006. Viruses as pathogens of marine organisms-from bacteria to whales. *J. Mar. Biolog.* 86(3), 453-467. doi: 10.1017/S002531540601335X

Naim, S., Brown, J. K., and Nibert, M. L. (2014). Genetic diversification of penaeid shrimp infectious myonecrosis virus between Indonesia and Brazil. *Virus Res.* 189, 97-105. doi: 10.1016/j.virusres.2014.05.013

Nibert, M. L. (2007). '2A-like'and 'shifty heptamer'motifs in penaeid shrimp infectious myonecrosis virus, a monosegmented double-stranded RNA virus. *J. Gen. Virol.* 88(4), 1315-1318. doi: 10.1099/vir.0.82681-0

NOAA (2020) NOAA Landings. https://foss.nmfs.noaa.gov [Accessed November 24, 2021].

Park, Y., James, D., and Punja, Z. K. (2005). Co-infection by two distinct totivirus-like double-stranded RNA elements in *Chalara elegans* (*Thielaviopsis basicola*). *Virus Res.* 109(1), 71-85. doi: 10.1016/j.virusres.2004.10.011

Pettersson, J. H. O., Shi, M., Eden, J. S., Holmes, E. C., and Hesson, J. C. (2019). Meta-transcriptomic comparison of the RNA viromes of the mosquito vectors *Culex pipiens* and *Culex torrentium* in northern Europe. *Viruses* 11(11), 1033. doi: 10.3390/v11111033

598 Poulos, B., Tang, K., Pantoja, C., Bonami, J. R., and Lightner, D. (2006). Purification and
599 characterization of infectious myonecrosis virus of penaeid shrimp. *J. Gen. Virol.* 87, 987-996.
600 doi: 10.1099/vir.0.81127-0
601 Preisig, O., Wingfield, B. D., and Wingfield, M. J. (1998). Coinfection of a fungal pathogen by two
602 distinct double-stranded RNA viruses. *Virology*, *252*(2), 399-406. doi: 10.1006/viro.1998.9480
603 Piers, H., 1920. The blue crab (*Callinectes sapidus* Rathbun): Extension of its range northward to
604 near Halifax, Nova Scotia. *Proceedings of the Nova Scotian Institute of Science*, 15, 1918-1922.
605 R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for
606 Statistical Computing, Vienna, Austria
607 Rice, N. R., Stephens, R. M., Burny, A., and Gilden, R. V. (1985). The gag and pol genes of bovine
608 leukemia virus: nucleotide sequence and analysis. *Virology* 142, 357–377. doi: 10.1016/0042-
609 6822(85)90344-90347
610 Rodriguez-Ezpeleta N, Teijeiro S, Forget L, Burger G, Lang BF (2009) Construction of cDNA
611 libraries: focus on protists and fungi. *Methods Mol. Biol.* 533, 33-47. doi: 10.1007/978-1-60327-
612 136-3_3
613 Senapin, S., Phewsaiya, K., Briggs, M., and Flegel, T. W. (2007). Outbreaks of infectious
614 myonecrosis virus (IMNV) in Indonesia confirmed by genome sequencing and use of an
615 alternative RT-PCR detection method. *Aquaculture* 266(1-4), 32-38. doi:
616 10.1016/j.aquaculture.2007.02.026
617 Shao Q, Jia X, Gao Y, Liu Z, Zhang H, Tan Q, et al. (2021) Cryo-EM reveals a previously
618 unrecognized structural protein of a dsRNA virus implicated in its extracellular transmission.
619 *PLoS Pathog*. 17(3): e1009396. doi: 10.1371/journal.ppat.1009396
620 Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., et al. (2016). Redefining the
621 invertebrate RNA virosphere. *Nature* 540(7634), 539-543. doi: 10.1038/nature20167
622 Shields, J. D., Williams, J. D., and Boyko, C. B. (2015). Parasites and diseases of Brachyura
623 in Treatise on Zoology-Anatomy, Taxonomy, Biology. The Crustacea (Brill), Volume 9 Part C (2
624 vols), 639-774.
625 Spitznagel, M.I., Small, H.J., Lively, J.A., Shields, J.D. and Schott, E. J. (2019). Investigating risk
626 factors for mortality and reovirus infection in aquaculture production of soft-shell blue crabs
627 (*Callinectes sapidus*). *Aquaculture* 502, 289-295. doi: 10.1016/j.aquaculture.2018.12.051
628 Suttle, C.A., 2007. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.*
629 5(10), 801-812. doi:10.1038/nrmicro1750
630 Tang, J., Ochoa, W. F., Sinkovits, R. S., Poulos, B. T., Ghabrial, S. A., Lightner, D. V., at al. (2008).
631 Infectious myonecrosis virus has a totivirus-like, 120-subunit capsid, but with fiber complexes at
632 the fivefold axes. *Proc. Natl. Acad. Sci.* 105(45), 17526-17531. doi: 10.1073/pnas.0806724105
633 Urayama, S. I., Takaki, Y., and Nunoura, T. (2016). FLDS: a comprehensive dsRNA sequencing
634 method for intracellular RNA virus surveillance. *Microbes Environ.* ME15171. doi:
635 10.1264/jsme2.ME15171
636 Vignuzzi, M. and López, C.B. (2019). Defective viral genomes are key drivers of the virus–host
637 interaction. *Nat. Microbiol.* 4(7), 1075-1087. Doi: https://doi.org/10.1038/s41564-019-0465-y
638 Wickner, R. B., Ghabrial, S. A., Nibert, M. L., Patterson, J. L., and Wang, C. C. (2011). "Family
639 Totiviridae," in Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of
640 the International Committee on Taxonomy of Viruses, eds A. M. Q. King, M. J. Adams, E. B.
641 Carstens, and E. J. Lefkowits (Tokyo: Elsevier Academic Press), 639–650.
642 Williams, S. H., Levy, A., Yates, R. A., Somaweera, N., Neville, P. J., Nicholson, J., et al. (2020).
643 The diversity and distribution of viruses associated with *Culex annulirostris* mosquitoes from the
644 Kimberley region of western Australia. *Viruses* 12(7), 717. doi: 10.3390/v12070717

Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E. C., Li, W. X., et al. (2010). Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci.* 107(4), 1606-1611. doi: 10.1073/pnas.0911353107

Xie, J., Havens, W.M., Lin, Y.H., Suzuki, N. and Ghabrial, S.A., 2016. The victorivirus *Helminthosporium victoriae* virus 190S is the primary cause of disease/hypovirulence in its natural host and a heterologous host. *Virus res.* 213, 238-245.

Zhai, Y., Attoui, H., Jaafar, F. M., Wang, H. Q., Cao, Y. X., Fan, S. P., et al. (2010). Isolation and full-length sequence analysis of *Armigeres subalbatus* totivirus, the first totivirus isolate from mosquitoes representing a proposed novel genus (*Artivirus*) of the family *Totiviridae*. *J. Gen. Virol.* 91(11), 2836-2845. doi: 10.1099/vir.0.024794-0

Zhang, P., Liu, W., Cao, M., Massart, S., and Wang, X. (2018). Two novel totiviruses in the white-backed planthopper, *Sogatella furcifera*. *J. Gen. Virol.* 99(5), 710-716. doi: 10.1099/jgv.0.001052

Zhao, M., Behringer, D. C., Bojko, J., Kough, A. S., Plough, L., dos Santos Tavares, C. P., et al. (2020). Climate and season are associated with prevalence and distribution of trans-hemispheric blue crab reovirus (*Callinectes sapidus* reovirus 1). *Mar. Ecol. Prog. Ser.* 647, 123-133. DOI: https://doi.org/10.3354/meps13405

Zhao, M., dos Santos Tavares, C. P., and Schott, E. J. (2021a). Diversity and classification of reoviruses in crustaceans: A proposal. *J. Invertebr. Pathol.* 182: 107568. doi: 10.1016/j.jip.2021.107568

Zhao, M., Flowers, E. M., and Schott, E. J. (2021b). Near-Complete Sequence of a Highly Divergent Reovirus Genome Recovered from *Callinectes sapidus. Microbiol. Resour. Announc.* 10(1). doi: 10.1128/MRA.01278-20

Zheng, L., Lu, X., Liang, X., Jiang, S., Zhao, J., Zhan, G., et al. (2017). Molecular characterization of novel totivirus-like double-stranded RNAs from *Puccinia striiformis* f. sp. tritici, the causal agent of wheat stripe rust. *Front. Microbiol*. 8:1960. doi: 10.3389/fmicb.2017.01960

Zhang, S., Shi, Z., Zhang, J. and Bonami, J.R., 2004. Purification and characterization of a new reovirus from the Chinese mitten crab, *Eriocheir sinensis*. *J. Fish Dis.*, 27(12), pp.687-692. doi: 10.1111/j.1365-2761.2004.00587.x

673 **Figure captions and Tables:**

674 **Figure 1. Agarose gel electrophoresis of apparent viral dsRNA bands of CsTLV1 and CsTLV2.**
675 M: marker; Lane 1: control RNA of uninfected blue crab leg muscle; Lane 2: RNA extracted from
676 infected blue crab leg muscle showing two dsRNA bands of dsRNA-S and dsRNA-L, respectively.
677
678 **Figure 2. Schematic representation of CsTLV1 and CsTLV2 genome.** The two overlapping ORFs
679 and the untranslated regions (UTRs) are shown by boxes and a single line, respectively. Nucleotide
680 positions of ORFs and the putative slippery site for -1 frameshifting, spacer and pseudoknot are
681 indicated too. EFE (kcal/mol) indicates the minimal free energy.
682
683 **Figure 3. Conserved motifs in RdRp of CsTLV1 and CsTLV2.** Amino acid sequences alignment
684 of CsTLV1 and CsTLV2 with closely related toti-like viruses from NCBI database. Horizontal lines
685 above the alignment indicate the eight motifs, numbers in brackets suggest the amino acid sequence
686 lengths between the motifs, asterisks indicate identical amino acid residues, color gradients indicate
687 the similarity level of amino acid residues. Virus notations are as in **Supplementary Table 2**.
688
689 **Figure 4. Phylogenetic relationships between putative RdRp amino acids of CsTLV1 and**
690 **CsTLV2 with other selected *Totiviridae* members.** A phylogenetic tree was generated using the
691 maximum likelihood (ML) method with 1,000 bootstrap replicates. The asterisks indicate CsTLV1
692 and CsTLV2 in the present study. Green shade indicates totiviruses identified in crustacean hosts and
693 orange shade indicates totiviruses in other arthropod hosts. Virus notations are as in **Supplementary**
694 **Table 2**.
695
696 **Figure 5. Sampling sites in the northeastern Atlantic coast of the United States.** Triangles 1-8
697 within the inset show sampling sites in MA, RI, and NY. Circles 9-13 indicate sampling sites in DE,
698 MD, NC, LA, and TX. Red color indicates the presence of CsTLV1 and CsTLV2; and green
699 indicates the absence of CsTLV1 and CsTLV2.
700
701 **Figure 6. Electron microscopy images of putative CsTLV1 and/or CsTLV2 viral particles in**
702 **muscle, gill, and hepatopancreas tissue of *C. sapidus*.** (A) muscle; (B)gills; (C) hepatopancreas.
703 White arrow: virions; Red asterisk: putative viroplasm in gill. White triangle: dense arrangement of
704 virions in hepatopancreas. Scar bar, 100 nm.
705
706 **Figure 7. Histology of CsTLV1 and CsTLV2 infections in muscle, gill, and hepatopancreas of *C.***
707 ***sapidus*.** (A, D, and G) Muscle, hepatopancreas, and gills of uninfected crabs; (B, E, and H)
708 Muscles, hepatopancreas, and gills of CsTLV1 and CsTLV2 infected blue crabs; (C, F, and I)
709 Magnified vision of boxed area in B, E and H, respectively. Infected hemocytes in muscle,
710 hepatopancreases and gills often had pyknotic or karyopyknotic nuclei (magenta arrows) as well as
711 opaque, slightly eosinophilic intracytoplasmic inclusion bodies (blue arrows). Scale bar, 25 μm.
712
713
714
715
716
717
718
719
720

Table 1 Primers and probes used in RT-qPCR.

Name		Sequences	Size (bp)
CsTLV1	forward	GCAAAGGAGTGAAGGAGTGG	193
	reverse	GCAAGACGCATAGCACGATA	
	Probe	5'6-FAM/TGCTTGCGG/ZEN/AGAAACTGAACGAGA/3'IABkFQ	
CsTLV2	forward	ACGGCGACTTTGTTGAGT TT	183
	reverse	ACGGTAACCCAGACCATTGA	
	probe	5'Cy5/AGTTGGGAG/TAO/GCAGAGATGTGTGTT/3'IAbRQSp	

Table 2 RT-qPCR efficiency with dsRNA standards. The threshold cycles for a log10 dilution series are used to assess efficiency relative to 100% theoretical efficiency for a slope of 3.32 based on 10 replicates.

Genome	Slope	Y-intercept	R^2	efficiency
CsTLV1	3.29	38.730	0.995	100.1
CsTLV2	3.25	37.624	0.995	100.3

Table 3 CsTLV1 and CsTLV2 prevalence in *C. sapidus.* Specimens were collected from locations along the US Atlantic coasts and Gulf coasts of the United States.

Location	Collection date (Month-Year)	Latitude	Longitude	Total *N*	CsTLV1		CsTLV2	
					Infected (N)	Prev (%)	Infected (N)	Prev (%)
Agawam River, MA	Aug-2009	41.7619° N	71.6773° W	29	11	37.9	11	37.9
	Aug-2012			47	4	8.5	4	8.5
	Sep-2018			51	0	0	1	2
Falmouth, MA	Sep-2018	41.5388° N	70.6266° W	22	4	18	4	18
Acushnet River, MA	Aug-2012	41.6617° N	70.9182° W	49	15	30.6	18	36.7
Ninigret Pond, RI	Aug-2021	41.3697° N	71.6426° W	58	16	27.6	19	32.8
Napeague Bay, NY	Jul-2021	40.9987° N	72.0972° W	10	6	60	6	60
Georgica Pond, NY	Aug-2012	40.9361° N	72.2138° W	18	1	5.5	1	5.5
	Jul-2013			19	0	0	1	5.2
	Jul-2020			33	0	0	0	0
	Jul-2021			37	0	0	0	0
Moriches Bay, NY	Jul-2018	40.7738° N	72.8052° W	32	7	21.8	8	36.7
	Jul-2021			25	4	16	5	20
Shinnecock Bay, NY	Jul-2021	40.8426° N	72.4762° W	28	18	64.3	18	64.3
	Sep-2021			21	10	47.6	10	47.6
				17	11	64.7	11	64.7
Delaware Bay, DE	Apr-2019	38.9108° N	75.5277° W	51	0	0	0	0
	Aug-2021			38	0	0	0	0
Rhode River, MD	Mar-2015	38.8795° N	76.5216° W	33	0	0	0	0
	Jul-2018			52	0	0	0	0
	Aug-2020			30	0	0	0	0
Albemarle Sound, NC	Oct-2019	33.8772° N	76.1248° W	95	0	0	0	0
Port Aransas, TX	Jan-2021	27.8339° N	97.0611° W	40	0	0	0	0
Lake Salvador, LA	Jan-2021	29.7192° N	90.2432° W	40	0	0	0	0

Prev: Prevalence

730 **Table 4. Generalized linear modelling (GLM) of potential factors affecting CsTLV1 and**
731 **CsTLV2 infection.** The model with the lowest Akaike's information criterion (AIC) of all
732 combinations of predictor variables includes size and latitude. *** denotes significance ($p < 0.001$).
733

model	Predictor variable	Estimate (slope)	Standard Error	P-value
A. Full model:				
CsTLV1 Infection~ Latitude + Size + Sex (AIC = 339.98; df = 551)	Latitude	2.54	0.64	6.64e-05 ***
	Size	-1.00	0.25	9.47e-05 ***
	Sex	-0.31	0.35	0.37
CsTLV2 Infection1~ Latitude + Size + Sex (AIC = 340.89; df = 551)	Latitude	2.68	0.66	4.77e-05 ***
	Size	-1.12	0.26	1.12e-05 ***
	Sex	-0.30	0.35	0.39
B. Reduced model:				
CsTLV1 Infection ~ Latitude (AIC = 352.24; df = 553)	Latitude	2.33	0.54	1.43e-05 ***
CsTLV1 Infection ~Size (AIC = 382.85; df = 553)	Size	-1.13	0.24	2.24e-06 ***
CsTLV1 Infection ~Sex (AIC = 404.68; df = 553)	Sex	-0.13	0.31	0.67
CsTLV2 Infection ~ Latitude (AIC = 357.72; df = 553)	Latitude	2.38	0.54	9.78e-06 ***
CsTLV2 Infection ~Size (AIC= 385.34; df= 553)	Size	-1.25	0.24	1.42e-07 ***
CsTLV2 Infection ~Sex (AIC= 412.8; df= 532)	Sex	-0.08	0.30	0.78

734
735