A novel RNA virus, *Macrobrachium rosenbergii* Golda virus (MrGV), linked to mass mortalities of the larval giant freshwater prawn in Bangladesh

Hooper, Chantelle[1]; Debnath, Partho P. [2]; Biswas, Sukumar[3]; van Aerle, Ronny[1]; Bateman, Kelly S.[1]; Basak, Siddhawartha K.[2]; Rahman, Muhammad M.[2]; Mohan, Chadag V.[6]; Rakibul Islam, H.M.[4]; Ross, Stuart[1]; Stentiford, Grant D.[1]; Currie, David[3] and Bass, David[1,5]

[1] International Centre of Excellence for Aquatic Animal Health, Centre for Environment, Fisheries and Aquaculture Sciences (Cefas), Weymouth, Dorset, United Kingdom.

[2] WorldFish Bangladesh, Dhaka, Bangladesh.

[3] Winrock Bangladesh, Dhaka, Bangladesh.

[4] Bangladesh Fisheries Research Institute, Bagerhat, Bangladesh.

[5] Department of Life Sciences, The Natural History Museum, London, UK.

[6] WorldFish, Penang, Malaysia

Corresponding authors: Chantelle Hooper (chantelle.hooper@cefas.co.uk), Partho P. Debnath (P.Debnath@cgiar.org).

1 Abstract

Mass mortalities of the larval stage of the giant freshwater prawn, *Macrobrachium rosenbergii*, have been occurring in Bangladesh since 2011. Mortalities can reach 100% and have resulted in an 80% decline in the number of hatcheries actively producing *M. rosenbergii*. To investigate a causative agent for the mortalities, a disease challenge was carried out using infected material from a hatchery experiencing mortalities. Moribund larvae from the challenge were prepared for metatranscriptomic sequencing. *De novo* virus assembly revealed a 29 kb single-stranded positive-sense RNA virus with similarities in key protein motif sequences to yellow head virus (YHV), an RNA virus that causes mass mortalities in marine shrimp aquaculture, and other viruses in the *Nidovirales* order. Primers were designed against the novel virus and used to screen cDNA from larvae sampled from hatcheries in the South of Bangladesh from two consecutive years. Larvae from all hatcheries screened from both years were positive by PCR for the novel virus, including larvae from a hatchery that at the point of sampling appeared healthy, but later experienced mortalities. These screens suggest that the virus is widespread in *M. rosenbergii* hatchery culture in southern Bangladesh, and that early detection of the virus can be achieved by PCR. The hypothesised protein motifs of MrGV suggest that it is likely to be a new species within the *Nidovirales* order. Biosecurity measures should be taken in order to mitigate global spread through the movement of post-larvae within and between countries, which has previously been linked to other virus outbreaks in crustacean aquaculture.

2 Introduction

The giant river prawn *(Macrobrachium rosenbergii)* is a species of global aquaculture importance, with culture rapidly expanding from <7,000 tonnes in 1980 to >287,000 tonnes worldwide in 2017 (FAO, 2019). Production, particularly in developing parts of Asia, provides direct and indirect employment, a food source to alleviate poverty, and export income in a high value international market (New and Nair, 2012). Despite the exponential growth of the prawn industry, disease has been a significant problem, and when mortalities occur they are usually high, causing widespread impacts to the socioeconomic balance in these countries (Pillai and Bonami, 2012).

Bangladesh is considered to be a favourable environment for farming *M. rosenbergii,* named 'Golda' by Bangladeshi farmers, due to its vast inland freshwater river systems with adjacent brackish areas and a suitable climate for culturing tropical and sub-tropical aquaculture species (Wahab *et al.*, 2012). In 2017, Bangladesh was the world's second largest producer of *M. rosenbergii*, producing 48,574 tonnes (FAO, 2019). Despite this promising production figure, *M. rosenbergii* hatcheries have been experiencing mass mortalities of larvae, resulting in the number of hatcheries actively producing larvae declining by almost 80% over the past decade as they are unable to complete the production cycle and it becomes uneconomical to continue. Between 2009 and 2010, over 50 hatcheries were operating in Bangladesh, producing up to two hundred million post-larvae (PL) per year, however in 2019 production had decreased to 27.75 million PL/year from only 12 hatcheries. Hatcheries experiencing mortalities reported that larvae had abnormal shapes, reduced appetite, issues with moulting and a gradual whitening of the body; finally resulting in either disappearance from the culture system, or mortality (SAFETI, Winrock and Worldfish, *personal communication*; Alam *et al.*, 2019).

To compensate for the loss of hatchery-produced post-larvae, wild post-larvae are being caught from the vast river systems in the south of Bangladesh, violating a government ban on the capture of wild PL (Wahab *et al.*, 2012). Ahmed and Troell (2010) reviewed the potential for negative environmental impact associated with fishing for wild PL and identified two main issues; high levels of invertebrate and fish bycatch associated with fishing for larvae (and a subsequent decline in biodiversity) and, a reduction in numbers of larvae returning to the freshwater environment as adults to spawn (reducing natural *M. rosenbergii* abundance). As fishing for wild post-larvae increases due to the hatcheries being unable to complete production cycles, it is likely that the negative environmental impact caused by fishing will become more pronounced.

A number of surveys have been carried out by several organisations to determine possible causes for the mortalities observed in hatcheries in Bangladesh. The surveys identified many shortcomings, including deficiencies in water quality, water filtration systems, temperature fluctuations due to weather conditions, feeding practices, probiotic use, and poor quality inputs that had been exposed to formalin and bleaching (SAFETI, Winrock, Worldfish and BFRI, *personal communication*). However, despite these constraints, hatcheries were successfully completing production cycles until 2011, suggesting that a new factor has arisen, which could be disease related. Biosecurity has been highlighted as a major issue, in relation to the spread of mortalities throughout the hatcheries of Bangladesh.

As the culture of *M. rosenbergii* has increased globally, the incidence of diseases and emergence of novel pathogens has increased in parallel (Pillai and Bonami, 2012). Diseases known to affect different life stages of the giant freshwater prawn include numerous pathogenic bacteria, viruses and fungi. Opportunistic bacteria such as *Aeromonas* spp., *Pseudomonas* spp. and *Vibrio* spp. can cause infections in all life stages of *M. rosenbergii* culture (reviewed in Pillai and Bonami, 2012). *Spiroplasma eriocheiris* (Srisala *et al.*, 2018) has been shown to cause mortalities in later life stages of *M. rosenbergii.* Macrobrachium rosenbergii Nodavirus (MrNV) (Arcier *et al.*, 1999), in association with extra small virus (XSV) (Qian *et al.*, 2003), causes white tail disease (WTD), an OIE-listed disease. The viruses have caused devastation in the prawn hatchery industries of Thailand (Yoganandhan, Leartvibhas and Sriwongpuk, 2006), China (Qian *et al.*, 2003), India (Hameed, Yoganandhan and Widada, 2004), Taiwan (Wang *et al.*, 2007) and Indonesia (Murwantoko *et al.*, 2016). Other viruses causing mortalities include Macrobrachium rosenbergii Taihu Virus (MrTV), a novel dicistrovirus associated with larval mortalities in China (Pan *et al.*, 2016), Infectious Hypodermal and Haematopoietic Necrosis Virus (IHHNV) (Hsieh *et al.*, 2006), and Decapod Iridescent Virus (DIV1) (Qiu *et al.*, 2019). Other pathogens have been identified to infect *M. rosenbergii* without large mortalities including: *Macrobrachium* Parvo-like Virus (MPV) (Gangnonngiw *et al.*, 2009) and White Spot Syndrome Virus (WSSV) (Sahul Hameed, Charles and Anilkumar, 2000; Sarathi *et al.*, 2008). With the lack of immortal shrimp cell lines, disease challenges can be used to identify pathogenic agents infecting *M. rosenbergii*. This technique was recently used to determine that MrNV alone can cause mortalities in *M. rosenbergii* in the absence of XSV (Gangnonngiw *et al.*, 2020).

Alam *et al.* (2019) attempted to determine the causative agent(s) associated with the larval mortalities in Bangladesh hatcheries using a screen for *Vibrio* spp., which are known to cause disease in larval stages of prawn culture (Khuntia *et al.*, 2008). However, this study was unable to find a strong correlation between the *Vibrio* species present and the levels of mortalities seen in larvae. Larvae were also screened for the presence of MrNV and XSV, but neither of

these were detected. Therefore, no significant pathogens have yet been associated with the large-scale larval mortalities in Bangladesh. The clinical signs of disease observed in these larvae associated with the current large-scale mortalities do not precisely fit those described for infection with any of the known larval prawn pathogens, therefore an approach to detect novel pathogens was needed to determine if there was a pathogenic cause for the mortalities.

In this study we (1) carried out an experimental challenge trial using larvae sourced from a hatchery displaying disease, (2) generated and analysed metatranscriptomic libraries obtained from pools of larvae from experimental trials to identify any potential pathogens present in moribund animals, and (3) screened for pathogens identified as a risk in other hatcheries that had been experiencing mortalities to determine whether these agents were present.

3 Methods

3.1 Collection of hatchery cultured larvae and wild broodstock

Larvae were collected from five hatcheries in southern Bangladesh in production seasons of 2018 and 2019. Batches of approximately 20-30 whole larvae were fixed in RNAlater. In the same years, adult *M. rosenbergii* were collected from seven rivers also in southern Bangladesh (Figure 1).

3.2 Preparation of infectious study material

To produce materials for disease challenge, moribund larvae were collected from a hatchery that was experiencing mortalities at the time of the study. Larvae were ground using a sterile pestle and mortar, suspended in phosphate buffered saline (PBS) and filtered through a 0.45 µm syringe filter. Larvae from a healthy hatchery that had not experienced mortalities in past production cycles were also collected and prepared in the same manner as a control treatment.

3.3 *In vivo* tank experiments

Experimental larvae were obtained from Bangladesh Fisheries Research Institute's (BRFI) domesticated F1 broodstock. After hatching, larvae were reared up to stage three, before splitting into three groups for different experimental tank exposures.

(1) Experimental larvae were directly immersed into filtered challenge medium from moribund larvae for 10 minutes prior to being transferred to a fibreglass tank.
(2) Experimental larvae were fed artemia that had been immersed in challenge medium from moribund larvae for 10 minutes.
(3) Experimental larvae were immersed into filtered challenge material from healthy larvae for 10 minutes prior to being transferred to a fibreglass tank.

For all experiments 4000±200 larvae were used, and experiments were carried out in 40 l fibreglass tanks heated to a constant temperature of 30°C with a submerged thermostat heater. Twenty to thirty larvae were collected every day for the first five days of the experiment and on days seven and ten.

3.4 RNA extraction

RNA was extracted using RIBOZOL™ (VWR, Pennsylvania, U.S.). For larvae, approximately 20 larvae were added to 1 ml RIBOZOL, and for adults, approximately 5 mg of each tissue type (pleopod, hepatopancreas, gill, gut) was pooled and added to 1 ml of RIBOZOL. RNA extraction was carried out following the manufacturer's protocol, with a final resuspension in 50 µl molecular grade water (Fisher Scientific, Massachusetts, U.S.). To remove any DNA co-

extracted with RNA, a DNAse step was carried out using DNAse1 (Sigma Aldrich, Missouri, U.S.), following the manufacturer's protocol.

3.5 Double stranded cDNA synthesis

First strand synthesis was performed in 25 µl reaction volumes using 5 µl of MMLV-RT 5X buffer (Promega, Wisconsin, UK), 1.25 µl of dNTPs (10mM) (Promega), 0.625 µl of recombinant RNasin (Promega), 1 µl of random hexamer primers (Promega) and 1 µg of RNA in 16.125 µl. Prior to the addition of 1 µl of MMLV-RT enzyme (Promega), the reaction was incubated at 65°C in order to denature double stranded RNA. Following the addition of the MMLV-RT, the reaction was incubated at 37°C for 1 hour. Second strand synthesis was performed immediately after first strand synthesis. Single stranded cDNA was incubated at 94°C for 2 minutes, followed by 10°C for 5 minutes. Following incubation, 2 µl Sequenase buffer (Thermo Fisher, Massachusetts, U.S.), 0.3 µl Sequenase (Thermo Fisher) and 7.7 µl molecular grade water (Thermo Fisher) were added. Following this, reactions were incubated at 37°C for 8 minutes, 94°C for 2 minutes and 10°C for 5 minutes. During incubation at 10°C, 1.2 µl 1:4 (Sequenase:Sequenase dilution buffer (Thermo Fisher)) diluted Sequenase was added to reactions and incubated at 37°C for 8 minutes and 94°C for 8 minutes prior to transfer to ice.

3.6 Library Preparation

Three pools of cDNA were constructed from *in vivo* tank experiments: (1) Challenge days 1-4 larvae from immersion and feeding challenges using filtered material from moribund larvae (2) Challenge days 5-10 larvae from immersion and feeding challenges using filtered material from moribund larvae and (3) Larvae from a hatchery that had been experiencing mass mortalities. These pools were prepared for metatranscriptomic sequencing using the Illumina compatible Nextera XT library preparation kit (Illumina, California, U.S.) and sequenced on an Illumina MiSeq using v3 chemistry (Illumina).

3.7 Sequence analysis

The generated raw Illumina paired-end sequence reads were trimmed to remove adaptor and low quality sequences using Trim galore! v0.5.0 (Krueger, 2015). Trimmed sequences were quality-checked using FastQC v0.11.8 (Andrews, 2010) before reads from individual pools were assembled using both rnaSPAdes v3.13.0 (Nurk *et al.*, 2013) and the Iterative Virus Assembler (IVA) v1.0.8 (Hunt *et al.*, 2015). Assembled contigs were subsequently annotated using the BLASTp algorithm of Diamond v3.13.0 (Buchfink et al 2015) and the full NCBI non-redundant (nr) protein database (downloaded November 2018), and the results were visualised using MEGAN6 Community Edition v6.12.3 (Huson *et al.*, 2016). Paired reads from all samples were then mapped to the assembled contig representing the viral genome

sequence using BWA-MEM v0.7.17 and SAMtools v1.9 with default parameters (Li and Durbin, 2009; Li *et al.*, 2009). The output from BWA-MEM was visualised with Integrative Genomics Viewer (IGV) v2.5.2 (Robinson *et al.*, 2011). Assembly quality and accuracy were assessed with QualiMap v2.2.2 (García-Alcalde *et al.*, 2012).

Predicted open reading frames (ORFs) were identified using four different tools: Prokka v1.11 (Seemann, 2014) FgenesV0 (SoftBerry.com), GeneMarkS v4.28 (Besemer, Lomsadze and Borodovsky, 2001) and Vgas (Zhang *et al.*, 2019). ORFs that were supported by two or more programs were analysed further. Supported ORFs were annotated using NCBI BLASTp and the full NCBI protein sequence database (downloaded December 2019) and protein motifs were identified by HHpred (Hildebrand *et al.*, 2009; default parameters) and InterProScan 5 (Jones *et al.*, 2014). Predicted protein motifs were aligned against known nidovirus protein motifs using MAFFT (Katoh, Rozewicki and Yamada, 2017) and multiple sequence alignments (MSAs) were visualised in ESPRICT3 v3.0 (Robert and Gouet, 2014; default parameters). Transmembrane regions were identified using TMHMM v2.0 (Krogh *et al.*, 2001; default parameters). Ribosomal frameshift identifiers were found using FSFinder2 using the virus genome settings (Byun et al 2007). Secondary structure of the 3' UTR was predicted using MFOLD (Zuker, 2003; default parameters).

3.8 Phylogenetic analyses

Phylogenetic analyses was performed using the RdRp protein domain as in Saberi *et al.* (2018), conserved in known and proposed *Nidovirlaes,* on 15 *Nidovirales* with representatives from *Arteriviradae*, *Roniviridae, Mesoniviridae, Mononiviridae, Euroniviridae*, *Coronavirinae* and *Torovirinae* families and subfamilies. Two representatives from the *Astroviradae* order of viruses were also aligned as an outgroup. MSAs were performed using default parameters in MAFFT (Katoh, Rozewicki and Yamada, 2017). Maximum Likelihood phylogenetic analyses were carried out using RaxML BlackBox v.8 (Stamatakis, 2014) (GTR evolutionary model; all parameters estimated from the data). A Bayesian consensus tree was constructed using MrBayes v.3.2.6 (Ronquist *et al.*, 2012). Two separate MC3 runs with randomly generated starting trees were carried out for 2 million generations each with 1 cold and 3 heated chains using a GTR model. All parameters were estimated from the data. The trees were sampled every 1000 generations and the first 500,000 generations discarded as burn-in. All phylogenetic analyses were carried out on the Cipres server (Miller, Pfeiffer and Schwartz, 2010).

3.9 Primer design and specific PCR

Primers specific to the predicted enveloping protein ORF of MrGV were designed using Primer3 (version 4.1.0) (Untergasser *et al.* 2012) using default settings and an amplicon length

of 250-400 bp): MrGV_F1: 5'-TTTGCCCAGGTTAATTGCCC-3' and MrGV_F2: 5'-ACAAGTGCCAGTGAGACGTA-3', producing an amplicon of 319 bp. PCR amplification was performed in 25 µl reactions using 5 µl X Green GoTaq Flexi Buffer (Promega), 1.5 µl $MgCl_2$ (25mM) (Promega), 0.2 µl of each primer (10 µM), 0.5 µl dNTPs (25 mM) (Promega), 0.125 µl GoTaq polymerase (Promega), 15.9 µl molecular grade water and 1.25 µl of template cDNA. Initial denaturation was carried out at 94°C for 5 minutes, followed by 35 cycles of 94°C for 30 seconds, 63°C for 30 seconds and 72°C for 90 seconds. This was followed by a final extension at 72°C for 10 minutes. Amplicons were purified with Wizard® SV Gel and PCR Clean-Up System (Promega) and sequenced via the Eurofins TubeSeq service. Primers were tested on larval samples from 5 hatcheries and wild adult *M. rosenbergii* sampled from the river networks surrounding the hatcheries. *Penaeus monodon* tissues infected with yellow head virus (YHV) were tested as a negative control.

3.10 Specific PCR Screens

PCR screens for specific pathogens were carried out with the following primers: MrNV using MrNV2aF and Mr*N*V2aR primers (Sahul Hameed *et al.*, 2004); XSV using XSV-external forward and reverse primers (Sudhakaran, Yoganandhan and Ahmed, 2006); MrTV using MrTV472F and MrTV472R primers (Pan *et al.*, 2016); WSSV using 146F1, 146R1, 146F2 and 146R2 primers (Lo et al. 1996); PmNV using 261F and 261R primers (Surachetpong *et al.*, 2005); *Spiroplasma eriocheiris* using F28 and R5 primers (Ding *et al.*, 2007); and YHV using YC-F1a, YC-F1b, YC-R1a, YC-R1b, YC-F2a, YC-F2b, YC-R2a and YC-R2b primers (Mohr *et al.*, 2015). All PCR reactions were performed as above using cycling conditions specified in the original publications.

4 Results

4.1 In vivo tank experiments

At the end of the 10-day disease challenge, mortalities exceeding 80% had occurred in all three experimental groups, including larvae exposed to an extract produced from a hatchery that at the time had not experienced mortalities (Fig. 2). However, it was later discovered that the healthy hatchery experienced mass mortalities to a similar level of other hatcheries in the south of Bangladesh shortly after they were sampled for the experimental material. Larvae in the control treatment group showed a slightly lower mortality rate, with survivors up until day 15 of the challenge (data not shown). As the challenge progressed, reduction in swimming ability, feeding, and growth were observed in all treatments. Moribund larvae appeared white in colour compared to healthy larvae.

4.2 Complete genome assembly of virus

A total of 3,466,39, 3,556,054 and 5,888,915 Illumina read-pairs were generated for libraries 1, 2 and 3 respectively and after quality-trimming and filtering 3,453,843, 3,545,471 and 5,867,622 read-pairs remained. Contigs assembled using rnaSPAdes, both separately and by combining all three libraries, were annotated using Diamond in BLASTx mode. rnaSPAdes assembly of combined libraries produced 38,826 contigs; 23 contigs, of average length 4,560 bp, had similarity in protein sequence to yellow head virus (YHV) or gill-associated virus (GAV), but when the trimmed reads were aligned against the YHV genome (accession number GCA_003972805.1) no alignment was seen. *De novo* virus assembly using IVA from library three produced twelve contigs, and two non-overlapping contigs, of lengths 14,858 and 13,463 had BLASTx similarity to YHV. IVA assembly using a pool of all three libraries produced a full genome consensus sequence of 29.11 kb, with an average coverage of 1116x. The two contigs produced by IVA assembly of library three mapped to the 29.11 kb consensus sequence. The full genome sequence is deposited on in supplementary materials. Henceforth, we refer to this novel virus as Macrobrachium rosenbergii Golda Virus (MrGV). Mapping trimmed and quality-filtered reads from each individual library back to the consensus genome gave coverage of 5.37x for library one (pooled challenged larvae from days 1-4), 269.28x for library two (pooled challenged larvae from days 5-10) and 1462.64x coverage for library three (pooled larvae from a hatchery experiencing mortalities). To compare the efficiency of the rnaSPAdes assembly, assembled contigs from individual and pooled libraries were mapped back to the consensus genome. Eleven contigs from library one mapped to MrGV with 0.25x coverage, thirty-two contigs from library two mapped with 2.53x coverage, fifty contigs from library three mapped with 5.04x coverage and 81 contigs from the pooled libraries mapped with 7.25x coverage. rnaSPAdes contigs from libraries two and three gave good coverage

over the whole genome but failed to assemble the 5' and 3' ends. A comprehensive table of assembly statistics is in Table S1.

4.3 Open reading frame (ORF) prediction, genome prediction and motif identification

FgenesV0 predicted five protein-coding open reading frames (ORFs) in the positive sense direction, four of which were supported by GeneMarkS, Prokka and Vgas (Fig. 3), which were investigated in more detail. The two longest ORFs, ORF1a and ORF1b, showed homology to the replicase polyproteins of yellow head virus (Evalues of $1x10^{-29}$ and 0), whereas ORF3 showed homology to glycoproteins associated with species in the *Negarnaviricota* phylum (realm *Riboviria*), including a number of species in the genus *Orthobunyavirus* (Evalue of 1e-$1x10^{-14}$). ORF2 had no significant similarity to any known proteins. HHpred produced confident predictions (≥99 %) for a picornain-like protease and endopeptidase enzyme in ORF1a, a coronavirus-like RNA-dependent RNA polymerase (RpRp), a metal-binding helicase and a 3' – 5' exoribonuclease (ExoN) in ORF1b. HHpred did not detect any zinc-binding domains (ZBDs), however when ORF1b was run against the InterPro database using InterProScan5, a coronavirus-specific ZBD was identified. In ORF3, HHpred identified an enveloping glycoprotein associated primarily with viruses of the order *Bunyavirales.* Both HHpred and InterProScan5 were unable to identify any other protein domains, therefore those of MrGV were aligned against the nidovirus domains identified in Saberi *et al.* (2018). MSA identified a further three protein motifs: A nidovirus RdRp-associated nucleotidyltransferase (NiRAN) and S-adenosylmethionine (SAM)-dependent N7- and 2'-O-methyltransferases (N-MT and O-MT, respectively).

4.4 Identification of frameshift motifs

As *Nidovirales* typically translate ORF1a and ORF1b consecutively to produce pp1ab by -1 ribosomal frame shift (King *et al.*, 2012), FSFinder2 (Byun, Moon and Han, 2007) was used to identify -1 frameshift sites in the overlap between ORF1a and ORF1b. Both major elements of -1 frameshifting were identified in the overlap: A slippery site at position 15090 nt with the sequence "GGGTTTT", proceeded by a stem-loop stimulatory structure located a few nucleotides downstream between positions 15107 and 15165 nt.

4.5 Prediction of 3' UTR secondary structure

Analysis of the 3' UTR secondary structure following the final stop codon of the final 3' ORF revealed a thermodynamically stable RNA hairpin secondary structure (Fig. 4). The predicted hairpin structure of MrGV is stabilised by three helices with a 12 nt hairpin loop (ΔG=-40.5), in comparison to the GAV 3'-UTR structure, which forms a 8 nt hairpin loop stabilised by four helices (ΔG=-32.40), previously identified in Wijegoonawardane *et al.* (2008).

4.6 Phylogenetic characterisation of MrGV

A Bayesian consensus tree was produced based on the RdRp protein domain universally conserved in nidoviruses, including the RdRp sequence generated in this study for MrGV, and the only nidovirus protein domain able to align to the RdRp of *Astrovirales* (outgroup). All families within *Nidovirales* were monophyletic in our analyses. *Mr*GV branched as sister to GAV and YHV within the *Roniviridae* clade (Fig. 5). Other invertebrate-infecting *Nidovirales* shown in Fig. 5 include include Charynivirus-1 infecting the crab *Charybdis*, and Paguronivirus-1 infecting hermit crabs (Shi *et al.*, 2016), all branching within the *Euronivirdivae* clade. All invertebrate-infecting nidoviruses branched together with maximal Bayesian Posterior Probability.

4.7 MSA of 3' ORFs

As MrGV branches within *Roniviridae*, which have a distinct 3' genome organisation to other nidoviruses (Cowley *et al.*, 2004), MrGV ORF2 and ORF3 were aligned against the corresponding ORFs of GAV. This alignment (Fig. 6) showed similarities in protein sequences of the three 3' transmembrane helices of ORF3 of GAV to the predicted transmembrane helices of MrGV. Sequence similarities were also seen in ORF2 of GAV to ORF2 of MrGV (18.24% residue similarity), as well as gp116 (15.40% residue similarity) and gp64 (17.03 % residue similarity) protein coding regions of ORF3 in comparison to the aligned regions in MrGV ORF3.

4.8 Infection prevalence in hatchery-reared larvae and wild river-based adult populations

The virus was not detected by specific PCR in any adult tissues from wild river populations across the two-year sampling period (119 animals) and *S. eriocheiris* was the only pathogen detected in adult *M. rosenbergii* (10/119). However, MrGV was detected in larvae from all three hatcheries sampled in 2018 production cycles and all five hatcheries sampled in 2019 production cycles with no other pathogens detected by PCR in hatchery larvae. No histopathological signs of infection were seen in the larvae (data not shown).

5 Discussion

We report and analyse the complete genome of a novel single-stranded positive-sense RNA virus infecting cultured *M. rosenbergii* from hatcheries in southern Bangladesh. The novel virus, Macrobrachium rosenbergii Golda virus (MrGV), has a similar genome arrangement to viruses of the order *Nidovirales*, and appears to be most closely phylogenetically related to yellow head virus (YHV) and gill-associated virus (GAV), both infecting penaeid shrimp.

MrGV was detected by specific PCR in larvae from three hatcheries in 2018, two of which were experiencing mass mortality events when sampled, and one which was sampled just prior to a mass mortality event. In 2019 three of the hatcheries that had experienced mass mortalities the previous year were re-sampled during mass mortality events and were again positive for MrGV; two of these hatcheries underwent two production cycles, both suffering mass mortality events and were both PCR positive for MrGV. In addition to the three hatcheries that were re-sampled in 2019, a further two hatcheries, none of which had been sampled in 2018, were sampled in 2019 during mass mortality events. These further two hatcheries were also PCR positive for MrGV. Given the number of hatcheries affected by mass mortalities linked to MrGV, its temporal prevalence and spatial spread, with one of the hatcheries over 40 km in distance to the nearest hatchery that was PCR positive for MrGV, we suggest that this novel virus represents a very significant threat to *M. rosenbergii* aquaculture within Bangladesh, and may be a significant factor in the collapse of larval production in the industry of Bangladesh since 2010.

All PCR screens of larvae collected from hatcheries experiencing mass mortalities for pathogens known to infect the larval stage of *M. rosenbergii* were negative: MrNV and XSV, the causative agents of white tail disease (Arcier *et al.*, 1999; Qian *et al.*, 2003); MrTV, a virus associated with mass larval mortalities in China (Pan *et al.*, 2016), *Spiroplasma eriocheiris* (Srisala *et al.*, 2018), and WSSV - shown to be able to infect *M. rosenbergii* experimentally (Sahul Hameed, Charles and Anilkumar, 2000). Larvae were also screened for YHV and PmNV, viruses known to infect all life stages of marine shrimp species (Lightner and Redman, 1981; Boonyaratpalin *et al.*, 1993); these screens were all negative. Further to this, no sequences were assignable to any other known pathogens of *M. rosenbergii* in our metatranscriptomic data. Absence of these pathogens and no obvious bacterial cause observed by Alam *et al.* (2019) strongly suggests that the survival problems in Bangladesh hatcheries are not due to a currently known pathogen and that mass mortalities were linked either to hatchery practice factors or/and the emergence of a novel pathogenic agent. Despite numerous problems identified in hatchery practices (SAFETI, Winrock and Worldfish, *personal communication*), hatcheries were successfully producing until 2011, and since then, hatchery practices have not changed, suggesting that this is not the main source of mortality events.

PCR screens of adult *M. rosenbergii* from rivers used to collect berried females as the supply of broodstock for the hatcheries were also all negative for MrGV, suggesting that broodstock may not be the entry route of MrGV into the hatcheries. Adults were also negative for all other pathogens as above, except for *S. eriocheiris*. All larvae sampled from hatcheries experiencing mortalities were negative for *S. eriocheiris* and no reads for any *Spiroplasma* species were detected in metatranscriptomic analysis of the moribund larvae, suggesting that *S. eriocheiris* is also not causing the mortalities in hatcheries.

Nidoviruses (order *Nidovirales*) are enveloped positive-sense RNA viruses infecting a range of hosts including both vertebrates and invertebrates (King *et al.*, 2012). The invertebrate nidoviruses are composed of families: *Mesoniviridae, Roniviridae, Mononiviridae* and *Euroniviridae*, with the latter two families discovered within the last four years (King *et al.*, 2012; Shi *et al.*, 2016; Saberi *et al.*, 2018). *Roniviridae* comprises one genus, *Okavirus*, composed of two closely related crustacean-infecting viruses: Gill-associated virus (GAV) and yellow head virus (YHV) (King *et al.*, 2012). Both GAV and YHV are associated with disease in marine shrimp farming, with the former initially associated with mid-crop mortality syndrome (MCMS) in *Penaeus monodon* in Australia (Spann, Vickers and Lester, 1995) and the latter first associated with yellow head disease (YHD) in *P, monodon* in Thailand (Limsuwan, 1991). Prior to this study, no nidoviruses had been discovered in *Macrobrachium rosenbergii.*

When amino acid sequences of MrGV were screened against the NCBI non-redundant protein database, there was weak similarity to YHV and even weaker similarity to other nidoviruses. However, the only significant matches in pp1a and pp1ab were to this order, therefore we searched for protein motifs shared among other nidoviruses. Through database searches we identified five protein motifs present in all known nidoviruses: A protease, an RNA-dependent RNA polymerase, a zinc-binding domain, a helicase and a methyltransferase-exoribonuclease complex. A further three domains were found by sequence alignment of MrGV to the protein motifs of nidovirus-specific RdRp-associated nucleotidyltransferase (NiRAN) and S-adenosylmethionine (SAM)-dependent N7- and 2'-O-methyltransferases (N-MT and O-MT, respectively). We also identified nucleotide sequences within the overlap of ORF1a and ORF1b suggestive of a -1 ribosomal frameshift. The slippery sequence "GGGTTTT" proceeded downstream by a putative stem-loop stimulatory RNA structure suggests that pp1a and pp1ab are translated continuously by programmed ribosomal frameshift, a characteristic of nidoviruses (Fields, Knipe and Howley, 2007). The genome of *Okavirus*, the genus to which GAV and YHV belong, have a unique genome architecture to that of vertebrate nidoviruses – ORF2, encoding a nucleocapsid protein, is located upstream of the glycoprotein gene (ORF3) (Cowley *et al.*, 2004). This structure is also seen in MrGV, as well as sequence similarity of ORF2 and ORF3 to the corresponding ORFs in GAV.

A characteristic of positive-sense RNA viruses is the presence of a secondary RNA structure at the 3'-UTR; this is present in other members of the nidoviruses including *Coronavirdae* and *Arteriviradae* (Buck, 1996). These 3'-UTR structures and/or specific sequences have been shown to be critical to polymerase recognition and minus-strand genomic RNA synthesis (Dreher, 1999). A 3'-UTR secondary structure has been identified in GAV and YHV, and appears to be well conserved, with complementary base-changes implemented in YHV genotypes to retain a conserved stem-loop structure stabilised by four helices (Wijegoonawardane *et al.*, 2008). In this study we used MFOLD to predict the secondary RNA structure of *Mr*GV in the 3'UTR of the genome following the final stop codon of the 3' ORFs. MFOLD predicted a stem loop structure of a similar size to that of GAV and YHV, with three helices stabilising a 12 nt hairpin loop. It is hypothesised that the 3'-UTR RNA structure in GAV and YHV may act as a polymerase recognition site for minus-strand RNA synthesis (Wijegoonawardane *et al.*, 2008), and given the similarities in the 3'-UTR RNA structure in MrGV to GAV and YHV, this structure in MrGV may have a similar function.

Based on multiple sequence alignment of the RdRp protein motifs of a representation of families within *Nidovirales*, MrGV groups phylogenetically with family *Roniviridae*, which so far, exclusively comprises GAV and YHV. We hypothesise that MrGV also belongs to *Roniviridae*, however further sampling is needed to obtainadditional material for electron microscopy to visualise the virus *in situ* in order to confirm that virion morphology conforms with the rod-shaped characteristics of the family (King *et al.*, 2012).

Our study has also highlighted that choice of assembly tool is important in assembling viruses. rnaSPAdes was not able to assemble the complete MrGV genome, even when libraries were pooled to give higher coverage than single libraries alone; whereas IVA was able to assemble the full genome from the same pool of libraries. It is likely that IVA was able to assemble the full genome as it is capable of assembling RNA sequences at highly variable depths (Hunt *et al.*, 2015), a useful tool in the case of the nidoviruses, which are prone to variability in sequencing depth due to the production of sub-genomic mRNAs (sgmRNAs). The presence of sgmRNAs were not assessed in this study but given that rnaSPAdes could not assemble the full genome, it is likely that these sgmRNAs do exist. Future work includes identifying the presence or absence of sgmRNAs and potential transcription regulation sites (TRSs).

Samples for histology or electron microscopy matching those used for molecular analyses were not available from the disease challenge studies carried out here. Therefore, there is currently no information about histopathological signs of disease caused by MrGV. Histological samples were taken from the larvae sampled from hatcheries in 2019, however these animals were likely not to be moribund and were probably in the early stages of infection

by MrGV and thus, though infected, no pathology was seen. Furthermore, RNA viruses infections are often difficult to visualise in histology sections. Further sampling is ongoing in order to visualise the virus infecting larval tissue(s) using transmission electron microscopy.

The challenge experiment data from this study were unable to determine the route of entry of MrGV into the hatchery system as it was later determined that control animals were sourced from a hatchery assumed to be free of MrGV, but which after sampling larvae for the challenge experienced mass mortalities to a similar level as seen in other hatcheries in southern Bangladesh. Larvae were sampled from the apparently unaffected hatchery and fixed before challenging; these animals were screened by PCR for all pathogens as above and were only positive for MrGV, thus suggesting that animals were infected with MrGV prior to the challenge, explaining the high levels of mortality in 'control' animals. This confounding situation nonetheless further suggests the role and ubiquity of MrGV in mass mortality events across multiple hatcheries in Bangladesh. Future work to identify how the virus is entering the hatcheries is ongoing to suggest preventative methods that could be implemented to ensure biosecurity.

This is the first study, to our knowledge, to use a metatransciptomic approach to investigate the mass mortalities experienced in hatchery-reared prawn larvae in Bangladesh. Previous studies have investigated possible agents involved in mortality events using PCR screens and microbial culture, but required the agent to have gene sequences sufficiently similar to known pathogens to amplify with PCR primers or culturable on media respectively, thus limiting the detection of more genetically divergent pathogens, possibly including many viruses. The metatranscriptomic approach used in this study to discover the agent involved in the Bangladesh mortalities was able to identify and characterise a novel pathogen that would likely have not been identified by most commonly employed methods. Furthermore, this method can be used to rapidly detect and comprehensively analyse new and cryptic viral linages, of relevance to both animal and human health.

6 Acknowledgements

This research was funded by the Newton Fund (delivered by the Biotechnology and Biological Sciences Research Council and the Economic & Social Research Council), the Department for International Development, and the Indian Department of Biotechnology under the Newton Fund Global Research Partnership in Aquaculture programme (BB/N00504X/1), and by Defra contract FX001 (OIE Collaborating Centre: Emerging Aquatic Animal Diseases). PPD, MMR and VCM were supported by the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish

7 References

Ahmed, N. and Troell, M. (2010) 'Fishing for prawn larvae in Bangladesh: An important coastal livelihood causing negative effects on the environment', *Ambio*, 39(1), pp. 20–29. doi: 10.1007/s13280-009-0002-y.

Alam, S. *et al.* (2019) 'Mass larval mortality in a giant freshwater prawn Macrobrachium rosenbergii hatchery: an attempt to detect microbes in the berried and larvae', *Bangladesh Journal of Fisheries*, 31(1).

Andrews, S. (2010) 'FastQC: A quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/'.

Arcier, J. M. *et al.* (1999) 'A viral disease associated with mortalities in hatchery-reared postlarvae of the giant freshwater prawn Macrobrachium rosenbergii ', *Diseases of Aquatic Organisms* , pp. 177–181. Available at: http://www.scopus.com/inward/record.url?scp=0033619639&partnerID=8YFLogxK.

Besemer, J., Lomsadze, A. and Borodovsky, M. (2001) 'GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions', *Nucleic acids research*. Oxford University Press, 29(12), pp. 2607–2618. doi: 10.1093/nar/29.12.2607.

Boonyaratpalin, S. *et al.* (1993) 'Non-occluded baculolike virus, the causative agent of yellow-head disease in the black tiger shrimp (Penaeus monodon)', *Fish Pathology*, 28, pp. 103–109.

Buchfink, B., Xie, C. and Huson, D. H. (2015) 'Fast and sensitive protein alignment using DIAMOND', *Nature Methods*, 12(1), pp. 59–60. doi: 10.1038/nmeth.3176.

Buck, K. W. (1996) 'Comparison of The Replication of Positive-Stranded Rna Viruses of Plants and Animals', *Advances in Virus Research*. Academic Press, 47, pp. 159–251. doi: 10.1016/S0065-3527(08)60736-8.

Byun, Y., Moon, S. and Han, K. (2007) 'A general computational model for predicting ribosomal frameshifts in genome sequences', *Computers in Biology and Medicine*. Pergamon, 37(12), pp. 1796–1801. doi: 10.1016/J.COMPBIOMED.2007.06.001.

Cowley, J. A. *et al.* (2004) 'The gene encoding the nucleocapsid protein of Gill-associated nidovirus of Penaeus monodon prawns is located upstream of the glycoprotein gene', *Journal of virology*. American Society for Microbiology, 78(16), pp. 8935–8941. doi: 10.1128/JVI.78.16.8935-8941.2004.

Ding, Z. *et al.* (2007) 'A simple PCR method for the detection of pathogenic spiroplasmas in crustaceans and environmental samples', *Aquaculture*. Elsevier, 265(1–4), pp. 49–54. doi: 10.1016/J.AQUACULTURE.2007.01.032.

Dreher, T. W. (1999) 'FUNCTIONS OF THE 3′-UNTRANSLATED REGIONS OF POSITIVE STRAND RNA VIRAL GENOMES', *Annual Review of Phytopathology*. Annual Reviews, 37(1), pp. 151–174. doi: 10.1146/annurev.phyto.37.1.151.

FAO (2019) *FAO yearbook*, *Fishery and Aquaculture Statistics 2017*. Rome.

Fields, B. N., Knipe, D. M. and Howley, P. M. (2007) *Fields virology*. 5th edn. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Gangnonngiw, W. *et al.* (2009) 'Parvo-like virus in the hepatopancreas of freshwater prawns Macrobrachium rosenbergii cultivated in Thailand', *Diseases of Aquatic Organisms*, 85(3), pp. 167–173. Available at: http://www.int-res.com/abstracts/dao/v85/n3/p167-173/.

Gangnonngiw, W. *et al.* (2020) 'In experimental challenge with infectious clones of

Macrobrachium rosenbergii nodavirus (MrNV) and extra small virus (XSV), MrNV alone can cause mortality in freshwater prawn (Macrobrachium rosenbergii)', *Virology*. Academic Press, 540, pp. 30–37. doi: 10.1016/J.VIROL.2019.11.004.

García-Alcalde, F. *et al.* (2012) 'Qualimap: evaluating next-generation sequencing alignment data', *Bioinformatics*, 28(20), pp. 2678–2679. doi: 10.1093/bioinformatics/bts503.

Hameed, A. S. S., Yoganandhan, K. and Widada, J. S. (2004) 'Experimental transmission and tissue tropism of Macrobrachium rosenbergii nodavirus (MrNV) and its associated extra small virus (XSV) ', *Diseases of Aquatic Organisms*, 62(3), pp. 191–196. Available at: http://www.int-res.com/abstracts/dao/v62/n3/p191-196/.

Hildebrand, A. *et al.* (2009) 'Fast and accurate automatic structure prediction with HHpred', *Proteins: Structure, Function, and Bioinformatics*. John Wiley & Sons, Ltd, 77(S9), pp. 128–132. doi: 10.1002/prot.22499.

Hsieh, C. Y. *et al.* (2006) 'Infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in giant freshwater prawn, Macrobrachium rosenbergii', *Aquaculture*. Elsevier, 258(1–4), pp. 73–79. doi: 10.1016/J.AQUACULTURE.2006.04.007.

Hunt, M. *et al.* (2015) 'IVA: accurate de novo assembly of RNA virus genomes', *Bioinformatics*, 31(14), pp. 2374–2376. doi: 10.1093/bioinformatics/btv120.

Huson, D. H. *et al.* (2016) 'MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data', *PLoS computational biology*. Public Library of Science, 12(6), pp. e1004957–e1004957. doi: 10.1371/journal.pcbi.1004957.

Jones, P. *et al.* (2014) 'InterProScan 5: genome-scale protein function classification', *Bioinformatics (Oxford, England)*. 2014/01/21. Oxford University Press, 30(9), pp. 1236–1240. doi: 10.1093/bioinformatics/btu031.

Katoh, K., Rozewicki, J. and Yamada, K. D. (2017) 'MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization', *Briefings in Bioinformatics*, 20(4), pp. 1160–1166. doi: 10.1093/bib/bbx108.

Khuntia, C. P. *et al.* (2008) 'Characterization and pathogenicity studies of Vibrio parahaemolyticus isolated from diseased freshwater prawn, Macrobrachium rosenbergii (de Man)', *Aquaculture Research*. John Wiley & Sons, Ltd, 39(3), pp. 301–310. doi: 10.1111/j.1365-2109.2007.01888.x.

King, A. M. Q. *et al.* (2012) *Virus Taxonomy - Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier. doi: 10.1016/B978-0-12-384684-6.00066-5.

Krogh, A. *et al.* (2001) 'Predicting transmembrane protein topology with a hidden markov model: application to complete genomes', *Journal of Molecular Biology*. Academic Press, 305(3), pp. 567–580. doi: 10.1006/JMBI.2000.4315.

Krueger, F. (2015) 'A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available online at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/'.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics (Oxford, England)*. 2009/06/08. Oxford University Press, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics (Oxford, England)*. 2009/05/18. Oxford University Press, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Lightner, D. V. and Redman, R. M. (1981) 'A baculovirus-caused disease of the penaeid

shrimp, Penaeus monodon', *Journal of Invertebrate Pathology*. Academic Press, 38(2), pp. 299–302. doi: 10.1016/0022-2011(81)90137-3.

Limsuwan, C. (1991) *Handbook for Cultivation of Black Tiger Prawns*. Bangkok: Tamsetakit Co. Ltd.

Lo, C. *et al.* (1996) 'Detection of baculovirus associated with white spot syndrome (WSBV) in penaeid shrimps using polymerase chain reaction', *Diseases of Aquatic Organisms*, 25(1–2), pp. 133–141. Available at: https://www.int-res.com/abstracts/dao/v25/n1-2/p133-141/.

Miller, M. A., Pfeiffer, W. and Schwartz, T. (2010) 'Creating the CIPRES Science Gateway for inference of large phylogenetic trees', in *2010 Gateway Computing Environments Workshop (GCE)*, pp. 1–8. doi: 10.1109/GCE.2010.5676129.

Mohr, P. G. *et al.* (2015) 'New yellow head virus genotype (YHV7) in giant tiger shrimp Penaeus monodon indigenous to northern Australia', *Diseases of aquatic organisms*, 115(3), pp. 263–268.

Murwantoko, M. *et al.* (2016) 'Macrobrachium rosenbergii nodavirus infection in a giant freshwater prawn hatchery in Indonesia', *SpringerPlus*. Springer International Publishing, 5(1), p. 1729. doi: 10.1186/s40064-016-3127-z.

New, M. B. and Nair, C. M. (2012) 'Global scale of freshwater prawn farming', *Aquaculture Research*. John Wiley & Sons, Ltd, 43(7), pp. 960–969. doi: 10.1111/j.1365-2109.2011.03008.x.

Nurk, S. *et al.* (2013) 'Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads', in Deng, M. et al. (eds) *Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 158–170.

Pan, X. *et al.* (2016) 'Isolation and Characterization of a Novel Dicistrovirus Associated with Moralities of the Great Freshwater Prawn, Macrobrachium rosenbergii', *International journal of molecular sciences*. MDPI, 17(2), p. 204. doi: 10.3390/ijms17020204.

Pillai, D. and Bonami, J. R. (2012) 'A review on the diseases of freshwater prawns with special focus on white tail disease of Macrobrachium rosenbergii', *Aquaculture Research*. John Wiley & Sons, Ltd, 43(7), pp. 1029–1037. doi: 10.1111/j.1365-2109.2011.03061.x.

Qian, D. *et al.* (2003) 'Extra small virus-like particles (XSV) and nodavirus associated with whitish muscle disease in the giant freshwater prawn, Macrobrachium rosenbergii', *Journal of Fish Diseases*. John Wiley & Sons, Ltd, 26(9), pp. 521–527. doi: 10.1046/j.1365-2761.2003.00486.x.

Qiu, L. *et al.* (2019) 'Description of a Natural Infection with Decapod Iridescent Virus 1 in Farmed Giant Freshwater Prawn, Macrobrachium rosenbergii', *Viruses*. MDPI, 11(4), p. 354. doi: 10.3390/v11040354.

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42(W1), pp. W320–W324. doi: 10.1093/nar/gku316.

Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*, 29(1), pp. 24–26. doi: 10.1038/nbt.1754.

Ronquist, F. *et al.* (2012) 'MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space', *Systematic biology*. 2012/02/22. Oxford University Press, 61(3), pp. 539–542. doi: 10.1093/sysbio/sys029.

Saberi, A. *et al.* (2018) 'A planarian nidovirus expands the limits of RNA genome size', *PLOS Pathogens*. Public Library of Science, 14(11), p. e1007314. Available at:

https://doi.org/10.1371/journal.ppat.1007314.

SAFETI, Winrock and Worldfish (no date) ‘No Title’.

Sahul Hameed, A. ., Charles, M. X. and Anilkumar, M. (2000) ‘Tolerance of Macrobrachium rosenbergii to white spot syndrome virus’, *Aquaculture*. Elsevier, 183(3–4), pp. 207–213. doi: 10.1016/S0044-8486(99)00305-1.

Sahul Hameed, A. S. *et al.* (2004) ‘Studies on the occurrence of Macrobrachium rosenbergii nodavirus and extra small virus-like particles associated with white tail disease of M. rosenbergii in India by RT-PCR detection’, *Aquaculture*. Elsevier, 238(1–4), pp. 127–133. doi: 10.1016/J.AQUACULTURE.2004.06.009.

Sarathi, M. *et al.* (2008) ‘Clearance of white spot syndrome virus (WSSV) and immunological changes in experimentally WSSV-injected Macrobrachium rosenbergii’, *Fish & Shellfish Immunology*. Academic Press, 25(3), pp. 222–230. doi: 10.1016/J.FSI.2008.04.011.

Seemann, T. (2014) ‘Prokka: rapid prokaryotic genome annotation’, *Bioinformatics*, 30(14), pp. 2068–2069. doi: 10.1093/bioinformatics/btu153.

Shi, M. *et al.* (2016) ‘Redefining the invertebrate RNA virosphere’, *Nature*, 540(7634), pp. 539–543. doi: 10.1038/nature20167.

Spann, K. M., Vickers, J. E. and Lester, R. J. G. (1995) ‘Lymphoid organ virus of Penaeus monodon from Australia’, *Diseases of Aquatic Organisms*, 23, pp. 127–134.

Srisala, J. *et al.* (2018) ‘Distinctive histopathology of Spiroplasma eriocheiris infection in the giant river prawn Macrobrachium rosenbergii’, *Aquaculture*. Elsevier, 493, pp. 93–99. doi: 10.1016/J.AQUACULTURE.2018.04.057.

Stamatakis, A. (2014) ‘RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies’, *Bioinformatics (Oxford, England)*. 2014/01/21. Oxford University Press, 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.

Sudhakaran, R., Yoganandhan, K. and Ahmed, V. P. I. (2006) ‘Artemia as a possible vector for Macrobrachium rosenbergii nodavirus (MrNV) and extra small virus transmission (XSV) to Macrobrachium rosenbergii post-larvae ’, *Diseases of Aquatic Organisms*, 70(1–2), pp. 161–166. Available at: http://www.int-res.com/abstracts/dao/v70/n1-2/p161-166/.

Surachetpong, W. *et al.* (2005) ‘Improvement of PCR method for the detection of monodon baculovirus (MBV) in penaeid shrimp’, *Aquaculture*, 249(1), pp. 69–75. doi: https://doi.org/10.1016/j.aquaculture.2005.04.051.

Wahab, M. A. *et al.* (2012) ‘Current status and prospects of farming the giant river prawn Macrobrachium rosenbergii (De Man) in Bangladesh’, *Aquaculture Research*. John Wiley & Sons, Ltd, 43(7), pp. 970–983. doi: 10.1111/j.1365-2109.2012.03137.x.

Wang, C. S. *et al.* (2007) ‘RT-PCR amplification and sequence analysis of extra small virus associated with white tail disease of Macrobrachium rosenbergii (de Man) cultured in Taiwan’, *Journal of Fish Diseases*. John Wiley & Sons, Ltd, 30(3), pp. 127–132. doi: 10.1111/j.1365-2761.2007.00793.x.

Wijegoonawardane, P. K. M. *et al.* (2008) ‘Genetic diversity in the yellow head nidovirus complex’, *Virology*. Academic Press, 380(2), pp. 213–225. doi: 10.1016/J.VIROL.2008.07.005.

Yoganandhan, K., Leartvibhas, M. and Sriwongpuk, S. (2006) ‘White tail disease of the giant freshwater prawn Macrobrachium rosenbergii in Thailand ’, *Diseases of Aquatic Organisms*, 69(2–3), pp. 255–258. Available at: http://www.int-res.com/abstracts/dao/v69/n2-3/p255-258/.

Zhang, K.-Y. *et al.* (2019) ‘Vgas: A Viral Genome Annotation System’, *Frontiers in microbiology*. Frontiers Media S.A., 10, p. 184. doi: 10.3389/fmicb.2019.00184.

Zuker, M. (2003) ‘Mfold web server for nucleic acid folding and hybridization prediction’, *Nucleic acids research*. Oxford University Press, 31(13), pp. 3406–3415. doi: 10.1093/nar/gkg595.

8 Figures

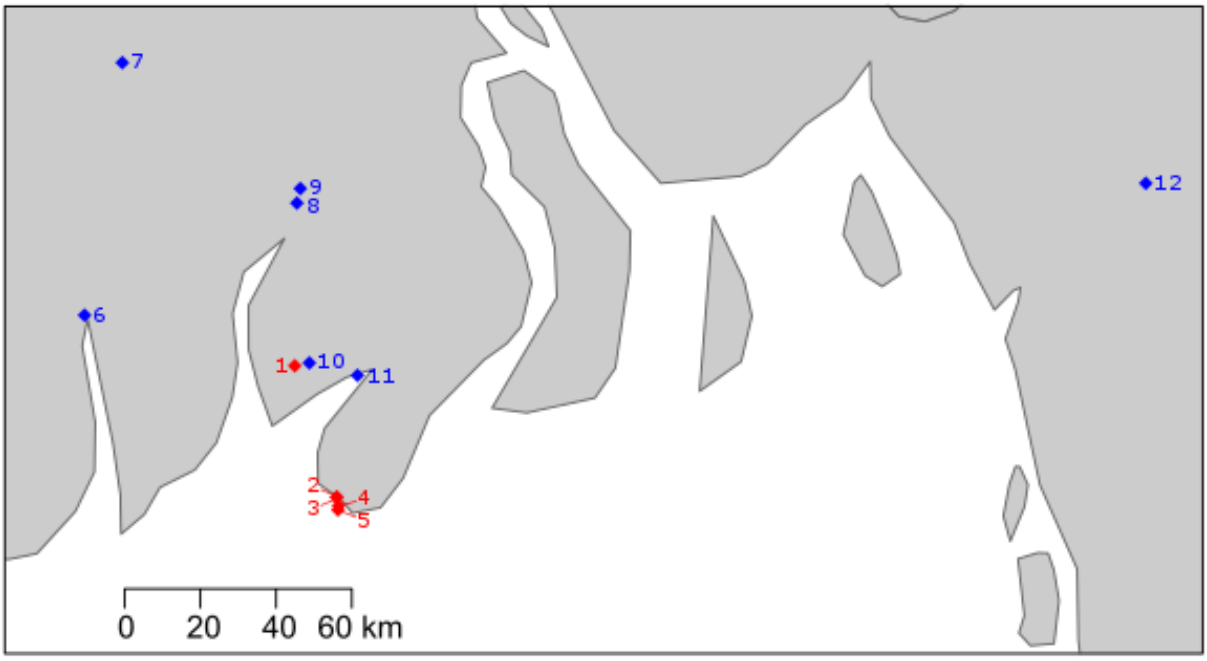

Figure 1: Map of South Bangladesh with sampling sites. Hatcheries denoted with a red diamond (1-5) and river sampling sites denoted with blue diamond (6: Shibsa River, 7: Rupsa River, 8: Katcha River, 9: Kocha River, 10: Biskhali River, 11: Payra River, 12: Karnaphuli River).

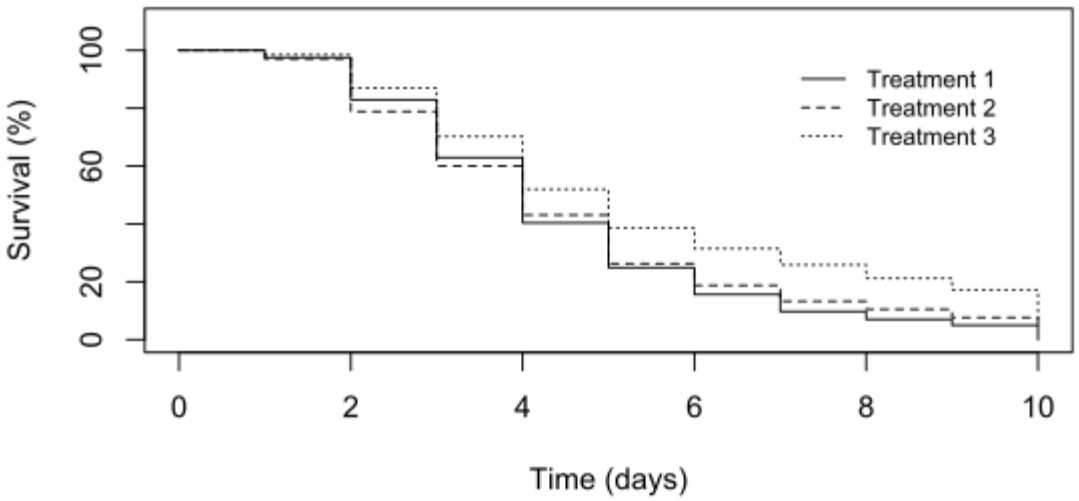

Figure 2: Larval mortalities observed in tank experiments. Treatment 1 = Larvae directly immersed into filtered material from moribund larvae. Treatment 2 = Larvae fed artemia that had been immersed into the filtered material from moribund larvae. Treatment 3 = Experimental larvae were directly immersed into filtered material from healthy larvae.

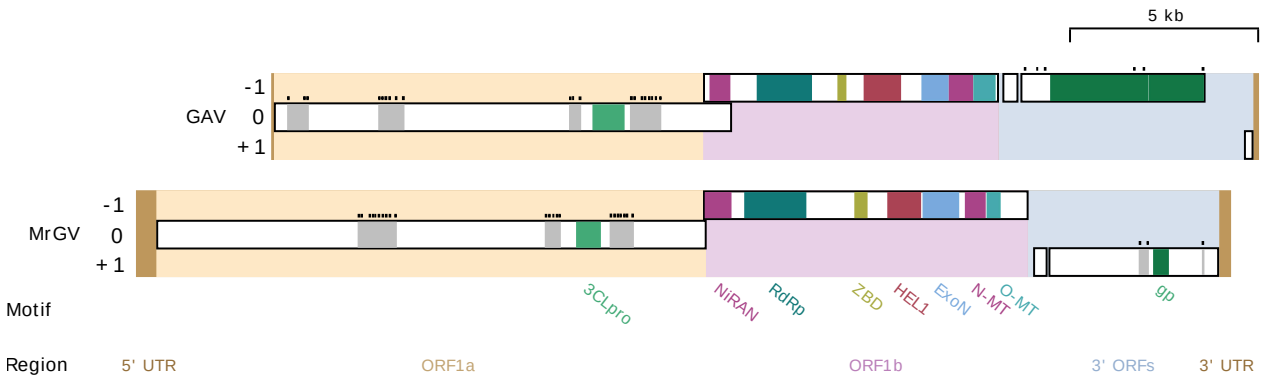

Figure 3: Schematic illustration of linear MrGV and gill-associated virus (GAV) genomes and proteomes. ORF1a is set as reading frame zero and genomes are spit into five sections: 5' UTR, ORF1a, ORF 1b, 3' ORFs and 3'UTR. Transmembrane (TM) regions are shown in grey with predicted TM helices shown as black bars above these regions. Predicted protein motifs are a 3C-like protease (3CLpro), nidovirus RdRp-associated nucleotidyltransferase (NiRAN), RNA-dependent RNA polymerase (RdRp), zinc-binding domain (ZBD), superfamily 1 helicase (HEL1), 3' – 5' exoribonuclease (ExoN), SAM-dependent N7- and 2'-O-methyltransferases (N-MT and O-MT, respectively) and glycoproteins (gp).

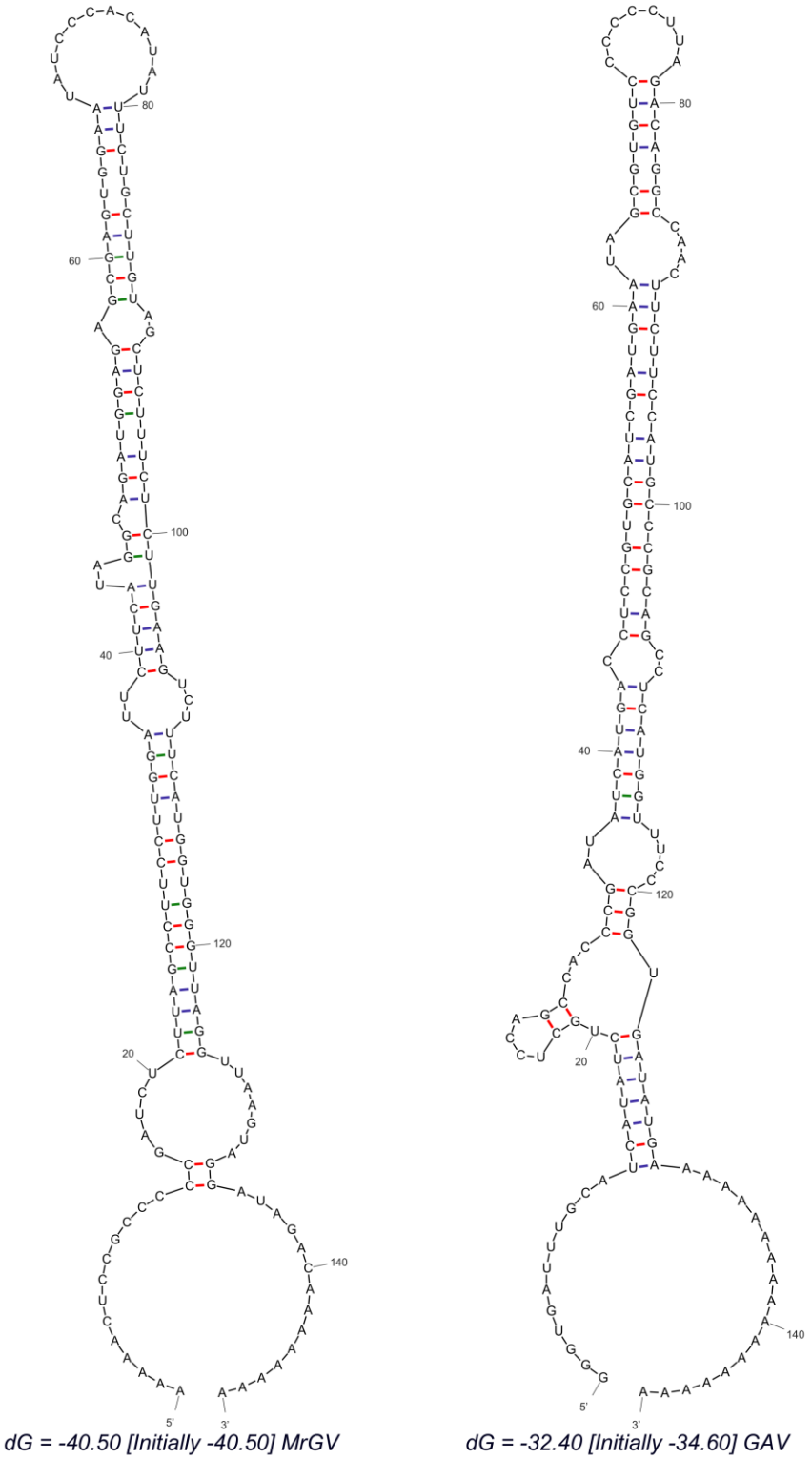

Figure 4: RNA secondary structure predicted using MFOLD of 3'-UTR sequences of MrGV and gill-associated virus (GAV) downstream of the stop codon of the final 3' ORF.

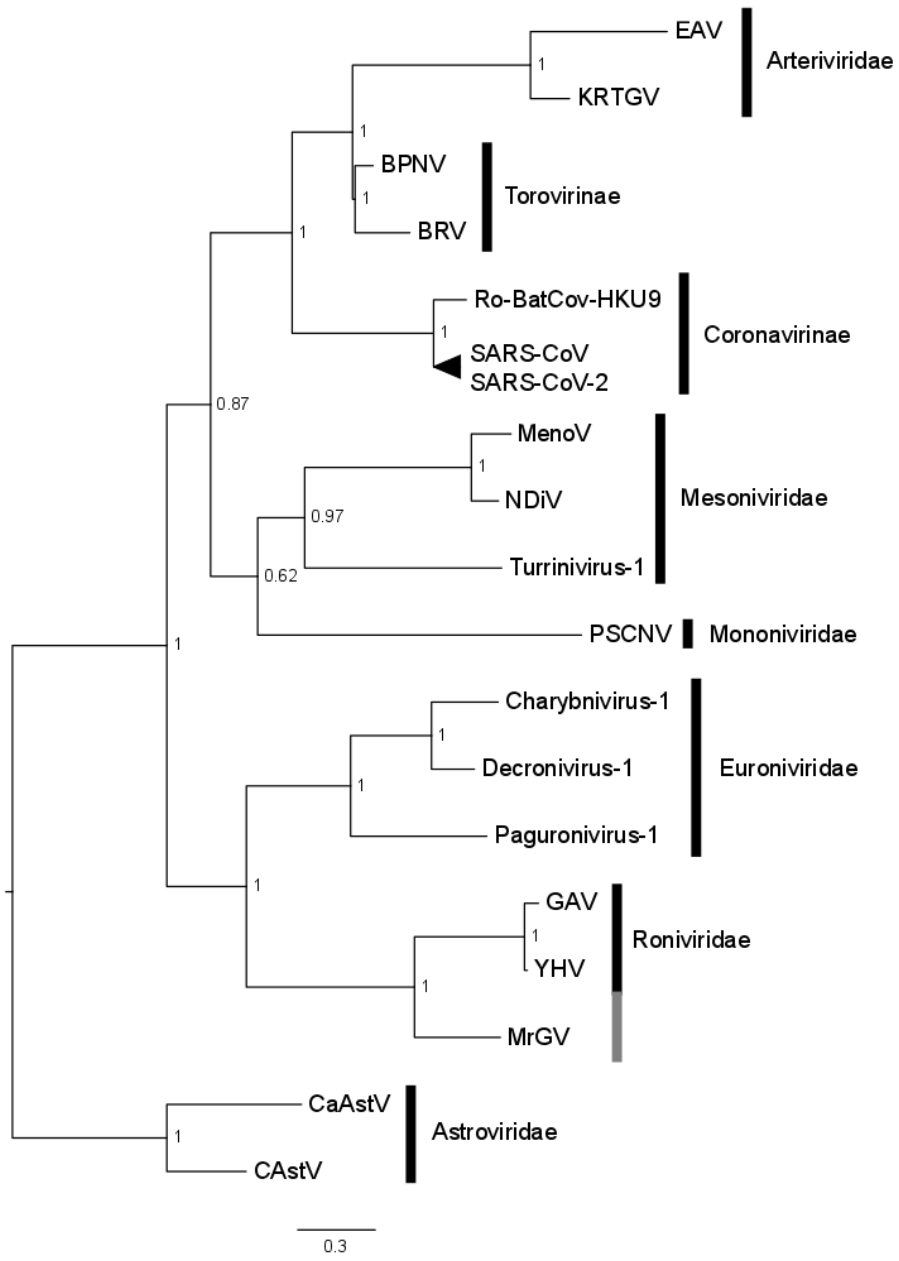

Figure 5: Bayesian consensus tree based on the RdRp of a representative set of nidoviruses, MrGV and astroviruses (outgroup). Accession numbers and names in Table S2.

Figure 6: Sequence alignment of MrGV and GAV 3'UTRs. Residue numbering based on the start codon of MrGV ORFs. ORF2 of GAV encodes the p20 nucleoprotein (purple bar). ORF3 of GAV encodes the gp116 glycoprotein (green bar) and gp64 glycoprotein (yellow bar). Transmembrane regions of GAV are shown as light grey bars and TMHMM-predicted transmembrane regions of MrGV are shown are dark grey bars. Residues in red boxes are conserved and residues in blue boxes have >0.7 Global Similarity Score based on physio-chemical similarities in ESPript3.