2015

Characterization of Newly Revealed Sequences in the Infectious Myonecrosis Virus Genome in *Litopenaeus vannamei*

Duan S. Loy
Iowa State University, dloy2@unl.edu

Sijun Liu
Iowa State University, sliu@iastate.edu

Mark A. Mogler
Harrisvaccines, Inc.

J. Dustin Loy
Iowa State University, jdloy@unl.edu

Bradley J. Blitvich
Iowa State University, blitvich@iastate.edu

Authors

Duan S. Loy, Sijun Liu, Mark A. Mogler, J. Dustin Loy, Bradley J. Blitvich, and Lyric C. Bartholomay

Characterization of Newly Revealed Sequences in the Infectious Myonecrosis Virus Genome in *Litopenaeus vannamei*

Duan S. Loy,[1] Sijun Liu,[2] Mark A. Mogler,[3] J. Dustin Loy,[1]*

Bradley J. Blitvich,[1] and Lyric C. Bartholomay[1,2]**

1. Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, Iowa, USA
2. Department of Entomology, Iowa State University, Ames, Iowa, USA
3. Harrisvaccines, Inc., Ames, Iowa, USA

* Present address: School of Veterinary Medicine and Biomedical Sciences, University of Nebraska–Lincoln, Lincoln, Nebraska, USA
** Present address: School of Veterinary Medicine, University of Wisconsin–Madison, Madison, Wisconsin, USA

Corresponding author – Lyric C. Bartholomay, email lbartholomay@wisc.edu

Abstract

Infectious myonecrosis virus (IMNV) causes significant economic losses in farmed shrimp, where associated mortality in ponds can reach 70%. To explore host/pathogen interactions, a next-generation sequencing approach using lymphoid organ tissue from IMNV-infected *Litopenaeus vannamei* shrimp was conducted. Preliminary sequence assembly of just the virus showed that there were at least an additional 639 bp at the 5′ terminus and 23 nt at the 3′ terminus as compared with the original description of the IMNV genome (7561 nt). Northern blot and reverse transcription-PCR analysis confirmed the presence of novel sequence at both ends of the genome. Using 5′ RACE, an additional 4 nt were discovered; 3′ RACE confirmed the presence of 22 bp rather than 23 bp of sequence. Based on these data, the IMNV genome is 8226 bp in length. dsRNA was used to trigger RNA interference

(RNAi) and suppress expression of the newly revealed genome sections at the 5′ end of the IMNV genome in IMNV-infected *L. vannamei*. An RNAi trigger targeting a 376 bp length of the 5′ UTR did not improve survival of infected shrimp. In contrast, an RNAi trigger targeting a 381 bp sequence in ORF1 improved survival to 82.2% as compared with 2.2% survival in positive control animals. These studies revealed the importance of the new genome sections to produce high-titre infection, and associated disease and mortality, in infected shrimp.

Introduction

A large number of pathogens threaten the shrimp aquaculture industry with a majority of these being viral in aetiology. The total economic impact is significant, with losses in the billions of dollars (Lightner et al., 2012). The emergence of novel pathogens is inevitable because of the combined effects of high-intensity farming, and a globalized aquaculture industry that exports and imports either live or unprocessed products, and thereby rapidly disseminates pathogens worldwide. The emergence of infectious myonecrosis virus (IMNV) is thought to have occurred via international transport of live animals (Senapin et al., 2007). IMNV was first reported in Brazil in 2003 after a disease outbreak occurred in 2002 (Poulos et al., 2006) and then an outbreak occurred in Indonesia in 2006 where it caused devastating economic losses (Senapin et al., 2007). Clinically, IMNV-infected animals demonstrate lethargy and the appearance of white muscle that spreads from the sixth abdominal segment; this pathology in individuals is an indicator for IMNV-induced mortality in ponds that can reach 70% (Senapin et al., 2007).

IMNV is an unenveloped, nonsegmented, dsRNA virus with icosahedral symmetry and complex protrusions at the fivefold symmetry axis that are thought to be involved in virus entry (Tang et al., 2008). IMNV is classified in the family Totiviridae based on its genomic organization and significant amino acid identity to the genome of a strain of *Giardia lamblia* virus, specifically in the RNA-dependent RNA polymerase (RdRp) gene (Poulos et al., 2006). The IMNV genome encodes two ORFs. ORF1 is translated into a polyprotein that is processed into three smaller proteins, the functions of which remain undetermined, via "2A-like" cleavage sites (Nibert, 2007; Tang et al., 2008) and also encodes the major capsid protein (MCP). ORF2 encodes the RdRp. Another protein (the MCP–RdRp fusion protein) is translated by virtue of a –1 frameshift within ORF1 (Nibert, 2007; Poulos et al., 2006).

The IMNV genome was initially reported to be 7561 bp in length (GenBank accession number EF061744.2) (Poulos et al., 2006). Although the reported IMNV genome size was approximately in the size range of members of the family Totiviridae (5.0–7.0 kbp), the published IMNV sequence did not appear to have a predicted internal ribosomal entry site (IRES) in its 5′ UTR that is characteristic of viruses in this family (Garlapati & Wang, 2005). In addition, some protein bands observed from purified IMNV virions did not match the sizes of the predicted cleavage products from the ORF1 polypeptide (Tang et al. 2008).

To explore shrimp/virus interactions, we undertook a next-generation sequencing (NGS) approach using lymphoid organ tissues. The penaeid shrimp lymphoid organ is located at the anterior end of the shrimp on the rostral/ventral side of hepatopancreas. The function of the lymphoid organ is, in part, removal of pathogens from haemolymph during circulation. Furthermore, the lymphoid organ is a common site for virus replication and

associated pathology (e.g., spheroids) to occur (Assavalapsakul et al., 2006; Escobedo-Bonilla et al., 2007). Lymphoid organ tissues express a significant number of immune-related genes in normal and infected shrimp (Pongsomboon et al., 2008). It follows that this is a lead tissue in which to explore IMNV RNA production and breakdown during an active IMNV infection. Using an NGS approach to study mRNA production in the lymphoid organ, we identified a substantial amount of virus genome sequence that extended beyond the original genome description. This additional sequence appeared to include sequence that is key to virus replication.

Results and Discussion

IMNV sequencing analysis

A total of 43 567 120 trimmed 50 bp reads were obtained from IMNV-infected *L. vannamei* and used for sequence assembly. To identify IMNV sequence, each of the contig sets was aligned to the existing IMNV sequences. Two IMNV contigs were identified: one 8044 bp in length and another of 8223 bp. Both contigs were significantly longer in size than that of the originally published IMNV sequence (7561 nt, Genbank accession number EF061744.1). A sequence alignment revealed that 662 bp of new, putative IMNV sequence was not within the limits of the known sequence but fell primarily upstream of the 5′ end (639 bp) and, to a more limited extent, at the 3′ end (23 bp) of the genome (Fig. S1).

New 5′ end and known IMNV sequences have an equal depth of sequencing

RNA sequencing reads that represented IMNV sequences were subjected to de novo assembly and mapped across the length of the genome (Fig. 1). Sequence coverage of the new sequences and the previously published sequences is listed in Table 1. Although there were significant variations in the sequencing depth at specific nucleotide locations in the IMNV genome (Fig. 1), the mean sequencing coverage between the new and known sequences (Table 1) was similar, suggesting that the new sequences were equally represented in the cDNA library.

Table 1. Comparison of sequencing coverage between the new (current) IMNV genome sequence produced by Illumina sequencing and previously published IMNV sequences (see also Fig. 1)

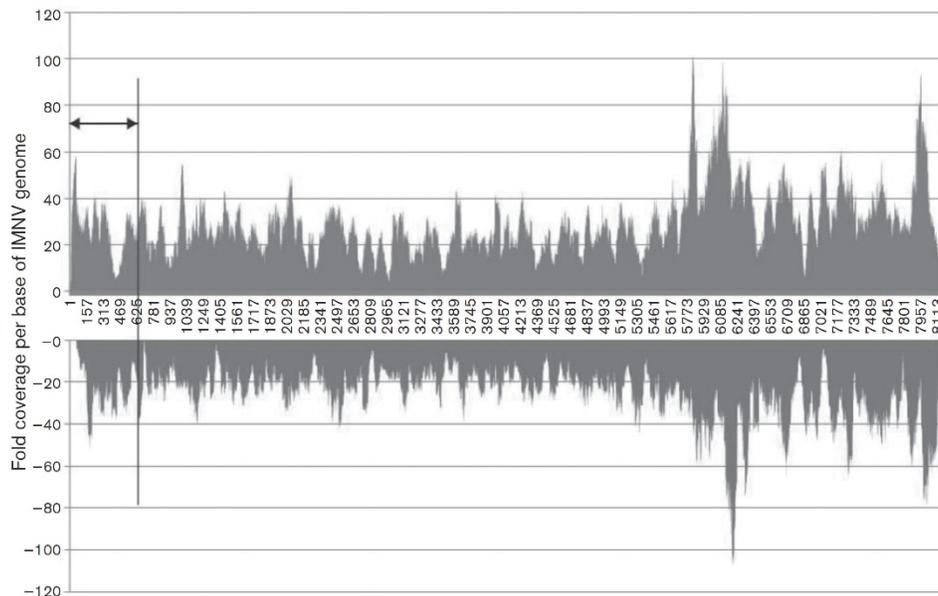| IMNV sequence	Sequencing coverage (fold)		
	Positive strand	Negative strand	Overall
Current IMNV genome (8223 bp)	29.39	26.21	55.60
5′ Current IMNV (639 bp)	26.37	21.48	47.85
Original IMNV genome (7561 bp)	29.73	26.67	56.40

Figure 1. Sequencing coverage of IMNV genomic sequences. Illumina reads from IMNV-infected shrimp that represent IMNV sequence were mapped to the assembled IMNV sequence using a custom-designed Perl program. Sequencing coverage was calculated according to the numbers of times each base mapped to the IMNV genome. New sequence is delineated with a vertical line at the position where the original 7561 bp sequence started or stopped at the 5′ and 3′ ends, respectively (see also Figs. S1 and S4).

Confirmation of presence of sequence in IMNV Indonesia and Brazil

Deep-sequencing results indicated that there were at least an additional 639 nt at the 5′ end and 23 nt at the 3′ end when compared with the previously published IMNV genome sequences. Several sets of primers were designed to confirm these novel IMNV sequences in isolates from Indonesia and Brazil. Four sets of primers were used to confirm the 5′ end by reverse transcription (RT)-PCR, including IMNV26F, IMNV312F, IMNV643R, and IMNV1016R (Table S1). Previously published primers (95F and 475R) that amplified the known genome sequence were used as positive controls (Table S1). RTPCR amplicons were subjected to gel electrophoresis and sequence length was confirmed (Fig. S2, lane 1–4). The largest amplicon (990 bp, Fig. S2, lane 4) was cloned and sequenced, and had 95 % nucleotide identity compared with published sequences. These primer pairs also were tested with IMNV Brazil, and RT-PCR results were identical to those for IMNV Indonesia (Fig. S2, lanes 5–8), and sequenced PCR products had 100 and 99% nucleotide identity to IMNV Brazil and Indonesia, respectively.

Two reverse primers (IMNV8204R1 and IMNV8204R2) were designed to anneal the newly discovered sequence, in combination with forward primers designed to previously known sequence at the 3′ terminus (Table S1). The novel sequence was confirmed in both IMNV strains using these primers (Fig. S3). These data did not prove that the sequence at

the 3′ end of the genome strains is identical in the two strains, but amplification did indicate that additional sequence at the 3′ terminus is indeed incorporated into IMNV virions.

IMNV genome size determination and Northern blot analysis

Gel electrophoresis was used to further confirm the size of the IMNV genome. A distinct band was evident at ~8–9 kb (Fig. 2). Total RNA from purified IMNV virus and noninfected shrimp was subjected to Northern analysis with RNA probes that were generated to anneal to the old sequence and new sequence. Again, a band of ~8–9 kb was detected in extracted RNA from IMNV-infected shrimp using probes designed to the original and new IMNV sequence (Fig. 3). These results further confirmed that the novel sequence discovered by deep sequencing existed in IMNV viral genome.
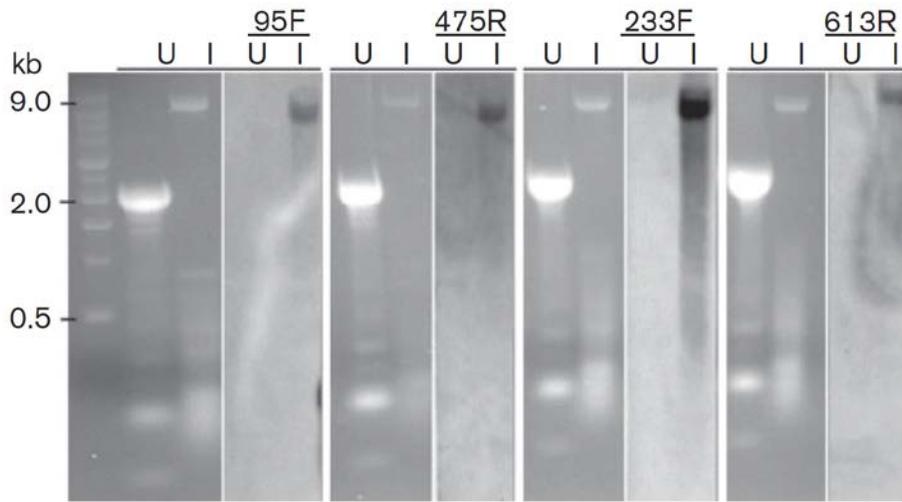
Figure 2. Northern analysis of IMNV novel and previously known sequence. RNA from *L. vannamei* was subjected to electrophoresis on a denaturing 1% agarose glyoxal gel. In each lane, 5 mg total RNA was loaded from uninfected (U) shrimp tissues or from clarified IMNV-infected (I) shrimp tissues. RNA loading is shown in the ethidium bromide-labeled gel image to the left of each hybridized blot. To detect the presence of IMNV, RNA was subjected to hybridization with probes to specific genome locations of interest, including the original genome with 95 (sense, 95F) and 475 (antisense, 475R) probes. Novel genome sequence was detected using 233 (sense, 233F) and 613 (antisense, 613R) probes, as labeled. Lane 1 was loaded with 0.5–9 kb (RNA) Millennium Marker (Ambion).

Figure 3. Sequence length analysis of the IMNV genome. IMNV RNA purified from shrimp tissues (IMNV, lane 1), and RNA from uninfected *L. vannamei* (*L.v.*, lane 2) were subjected to electrophoresis on a denaturing 0.8% agarose glyoxal gel containing ethidium bromide. Markers to determine the molecular weight of RNA include the 0.5–9 kb (RNA) (Millennium Marker Ambion) (left) and 0.5–10 kb RNA ladder (Invitrogen) (right).
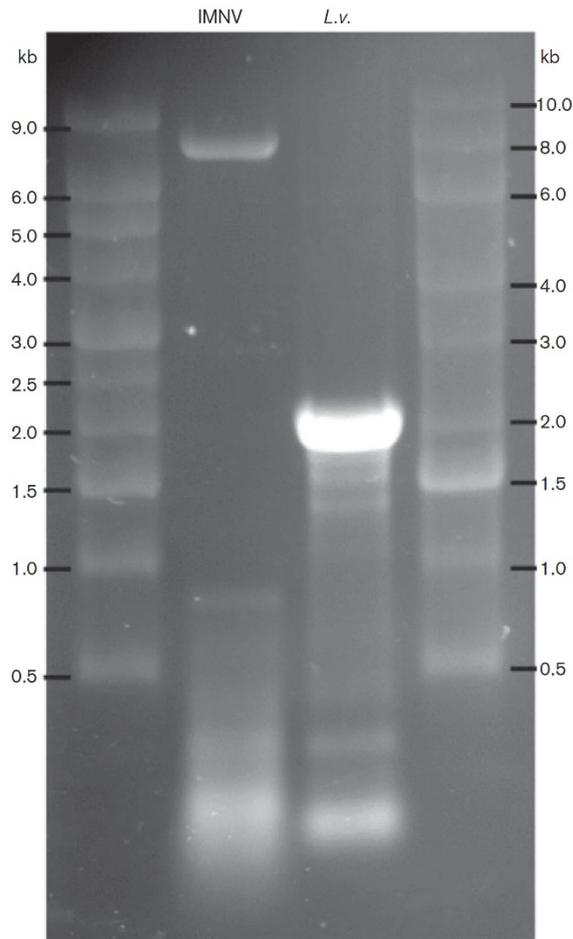
5′ and 3′ RACE PCRs to confirm the presence of additional nucleotides in the IMNV genome
To validate the new IMNV sequence identified by NGS, 5′ and 3′ RACE PCRs were performed. Fourteen PCR products generated from 5′ RACE of cDNA from IMNV Indonesia were sequenced; sequence analysis confirmed the presence of 639 bp identified by NGS and added an additional 4 bp upstream. These additional 4 bp were present both in IMNV Indonesia and Brazil. All told, 5′ RACE PCR confirmed that the IMNV sequences previously published and submitted to GenBank (accession numbers NC_007915.2, AY570982.2, and EF061744.1; 7561 bp) were truncated at the 5′ terminus by 643 bp.

As the IMNV genome is not polyadenylated, a synthetic poly(A) tail was added to the 3′ end and then 3′ RACE experiments were performed to confirm the presence of additional

sequence at the 3′ terminus of the genome. Ten PCR products were generated and sequenced using 3′ RACE for IMNV Brazil. These experiments confirmed that an additional 22 bp were present at the distal 3′ end, supporting the findings from the NGS analysis. We were unable to generate any positive clones in the 3′ RACE experiments performed with IMNV Indonesia. However, we were able to amplify and sequence the 3′ end of IMNV Indonesia using IMNV7852F as the forward primer and IMNV8210R2 as the reverse primer; the sequence of the latter primer corresponded to the distal 22 bp of the 39 UTR of IMNV Brazil (Table S1). These data indicated that an additional sequence was present at the 3′ end of the genome in both virus strains. Therefore, using 5′ and 3′ RACE, we confirmed the presence of the nucleotide sequences identified by NGS at the 5′ and 3′ ends of the IMNV genome, and revealed the presence of an additional 4 nt at the distal 5′ end. The full-length 8226 bp genomic sequence of IMNV was submitted to GenBank (accession numbers KF836757 and KJ556923).

Analysis of the newly identified sequence at the 59 end of the IMNV genome

The IMNV sequence is nonsegmented and encodes two predicted ORFs. To clarify whether the newly revealed 639 nt at the 5′ end of the genome encoded protein sequences, we translated the total length of the IMNV genome in silico in all six frames (Fig. S4). The predicted translation analysis showed that the longer IMNV genome encoded two predicted ORFs. The longer genome produced a predicted ORF1 in frame 1 that comprised 1719 aa, which was 114 aa larger than the previously published predicted product of ORF1 (1605 aa). The predicted start codon of the new ORF1 was 342 nt upstream of the documented start codon. Fig. S1 shows the new N-terminal sequence and alignment of the new N-terminal sequences and the published sequences of the IMNV ORF1. The ORF2 (frame 3) is almost identical to the previously predicted ORF2 protein sequence. With a new understanding of the coding region, the secondary RNA structure of the 5′ UTR sequence was predicted using RNAfold. The resulting predicted 5′ UTR RNA structure was then compared with known IRES structures (using IRESite and IRSS). There was no significant homology with known viral IRES elements; however, the IRES elements of viruses in the family *Totiviridae* do not always fall strictly in the 5′ UTR, and are characterized by multiple complex secondary RNA structures and multiple, noninitiating AUG codons (Garlapati & Wang, 2005). The predicted 5′ UTR structure of IMNV contains several stem-loop structures and multiple AUG codons.

In a previously published analysis of the translation productions of IMNV ORF1, two "2A-like" peptide cleavage motifs were described (Nibert, 2007), along with a cleavage site that cleaves the C terminus of ORF1 to release the MCP. These three cleavage motifs hydrolyze the ORF1 polyprotein into four predicted products: the N-terminal peptide 1 (10 kDa), peptide 2 (31 kDa), predicted peptide 3 (36 kDa), and the C-terminal 99 kDa MCP (Fig. S4). ORF2 (frame 3) is translated by −1 frameshifting of ORF1 and produces a 196 kDa MCP-RdRp fusion (Nibert, 2007; Poulos et al., 2006; Tang et al., 2008). Indeed, Tang et al. (2008) showed that when purified IMNV virions were separated by SDS-PAGE and stained, five clear protein bands were evident. Two bands at 99 and 196 kDa corresponded to MCP and MCP-RdRp, respectively. The remaining three bands were between 25 and 40 kDa (Tang et al., 2008). A 10 kDa peptide (the predicted size of peptide 1 of ORF1) was

never observed. The newly revealed genome sequence explained this discrepancy because it extended the N terminus of peptide 1 to 207 aa, with a predicted molecular mass of 23 kDa. Therefore, it is likely that the 25 kDa protein observed previously was the full-length peptide 1, and the other two bands represented peptides 2 (31 kDa) and peptide 3 (36 kDa) (Tang et al., 2008).

dsRNA targeting novel IMNV genome sections provides protection against infectious myonecrosis disease

In order to begin to test the biological relevance of the novel sequences at the 5′ terminus of IMNV, two dsRNAs (designated dsRNA26-401 and dsRNA233-613; Fig. S4) were designed and compared with dsRNA95-475, which targeted the previously known coding region of peptide 1. dsRNA95-475 provided significant disease protection [81.67 % survival against IMNV at 30 days post-infection (p.i.)] when animals were subjected to the dsRNA prior to infection (Loy et al., 2012). For comparative purposes, dsRNAs to the novel IMNV sequence were designed to have a similar length to dsRNA95-475. Shrimp were injected with 5 mg per shrimp (100 ml) of either dsRNA26-401, dsRNA233-613, dsRNA95-475, dsEGFP or 2% NaCl (sham inoculation), held for 48 h and then infected with IMNV (1:100 dilution) by injection at 48 h. Negative control shrimp were injected with 2% NaCl (virus diluent) rather than virus. In three separate trials, significant mortality was observed in shrimp exposed to dsRNA26-401, dsEGFP, or sham inoculation with 3.3, 2.2, and 2.2% mean survival, respectively. In contrast, shrimp that received dsRNA233-613 and dsRNA95-475 showed similar protection against IMNV with mean survival of 82.2 and 97.8%, respectively, at 14 days p.i. (Fig. 4). Mean survival for dsRNA233-613 and 95-475 was significantly different from the other groups (dsEGFP, dsRNA26-401, and positive control) (Pv0.001) using one-way ANOVA followed by Tukey's multiple comparison test. All of the negative control animals (nonchallenged) survived during the experiments. Therefore, only dsRNA233-613 showed protection against IMNV, and the protection efficacy was comparable with dsRNA95-475. This differential effect of RNAi triggers on IMNV virus load and virus-induced pathology is in keeping with a previous study wherein we showed that dsRNA targeting in ORF1 within the peptide-1-encoding region provided maximum protection from morbidity and mortality (Loy et al., 2012). dsRNA233-613 does encompass a region of the predicted ORF1 and dsRNA26-401 falls exclusively in the 5′ UTR (Fig. S4). ORF1 contains a dsRNA-binding protein domain; in a previous study, a dsRNA (dsRNA95-475) that also targeted the peptide-1-coding region was the most effective antiviral dsRNA tested, as compared with dsRNAs that targeted the MCP and RdRP (Loy et al., 2012).
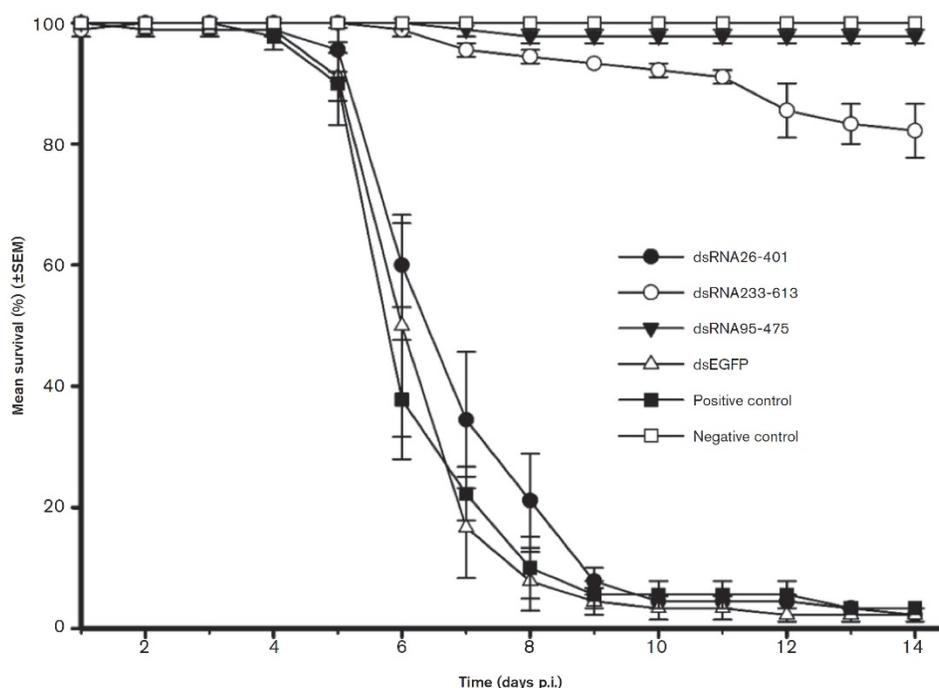
Figure 4. Shrimp survival post-administration with novel dsRNAs (dsRNA26-401 and dsRNA233-613) compared with previously reported dsRNA (dsRNA95-475) and heterologous dsRNA (dsEGFP) post 48 h IMNV infection. Data represent mean ± SEM; $n = 90$ animals per treatment except for the negative control group, $n = 30$ animals.

Novel IMNV dsRNA inhibition of IMNV replication

Virus load was analyzed by quantitative real-time (qRT)-PCR on muscle tissues from dead animals in dsRNA26-401, dsEGFP, and positive control groups at 9–14 days p.i., and the surviving animals from the dsRNA233-613 and dsRNA95-475 groups at 14 days p.i. The results showed that dsRNA95-475 and dsRNA233-613 inhibited virus replication [mean 4.47×10^2 and 2.90×10^3 virus copies $(\mu l \ RNA)^{-1}$, respectively] and virus load in these animals was significantly different from that in the positive control group (mean 1.1×10^5 virus copies $(\mu l \ RNA)^{-1}$) ($P = 0.013$ and $P = 0.016$, respectively). By contrast, dsRNA26-401 and dsEGFP-exposed shrimp did not show reduction of virus (mean 4.77×10^4 and 7.72×10^4 virus copies $(\mu l \ RNA)^{-1}$, respectively) (Fig. S5). This was further evidence that dsRNA233-613 suppressed virus replication during IMNV infection.

In conclusion, new genome sections of IMNV were discovered using deep sequencing of the lymphoid organ from infected *L. vannamei*. Based on these data, and 5′ and 3′ RACE, the genome sequence of IMNV is 8226 bp in length, as compared with the original understanding of a 7561 bp genome. This new sequence represents sections that expand the predicted coding region of ORF1 by 114 aa and add an additional 23 nt at the 3′ end. The presence of these genome sections was confirmed by RT-PCR, RACE PCR, and Northern blotting. RNAi triggers were developed to specifically target the new 5′-most sequence, including one trigger that falls on sequence that represents just the 5′ UTR, and one that

encompasses some of the 5′ UTR and some of the predicted coding region of ORF1, peptide 1. The latter of these provided significant protection from IMNV challenge manifest as significantly higher survival, decreased pathology, and a reduction in viral replication compared with the RNAi trigger that targets just the 5′ UTR. This discovery provides an opportunity to further understand the function of the peptides encoded at the N terminus of ORF1, as well as novel targets for RNA-based vaccines and antivirals.

Methods

Shrimp culture
Specific-pathogen-free (SPF) *L. vannamei* were acquired from Shrimp Improvement Systems and maintained in a SPF, biosecure facility at Iowa State University. Post-larvae were placed in 1,000 liter fiberglass tanks filled with artificial sea water. Water conditions were maintained at 25–27°C, 30–35 parts per thousand salinity, with constant airstone aeration. Each tank was maintained with a carbon filter and an oyster shell airlift biofilter. Water quality was measured according to ammonia and nitrite levels (Nitriver3; Hach); temperature and salinity were measured daily.

IMNV Indonesia-infected lymphoid organs
IMNV-infected tissues used for this study were collected during an IMNV outbreak in Indonesia. The virus inoculum was prepared as described previously (Loy et al. 2012). Six shrimp (15–20 g) were infected by intramuscular injection with IMNV (strain Indonesia). Virus stock was prepared and diluted in 1:100 2% NaCl [~2.11×10^3 viral copies (μl RNA)$^{-1}$] and filter-sterilized. Control shrimp were injected with 2% NaCl (viral diluent), as reported by Loy et al. (2012).

Lymphoid organ tissues were collected from IMNV-infected animals at 5 days p.i. or from normal control shrimp at the same time. Gross organ characteristics were confirmed via light microscopy on a dissecting microscope. All lymphoid organs were preserved in RNAlater RNA Stabilization Reagent (QIAGEN) and stored at –80°C. Samples were then transferred to 750 μl TRIzol LS (Invitrogen) and processed for RNA extraction. Residual DNA was removed from the extracted RNA using a TURBO DNA-free kit (Ambion). RNA integrity was determined using an Agilent 2100 Bioanalyzer. Total RNA from five or six individual animals was pooled and diluted in nuclease-free water to achieve a concentration of 0.1 μg ml^{-1}. cDNA libraries were prepared from pooled RNA samples using a TruSeq RNA Sample Prep kit (Illumina) following the manufacturer's instructions.

Illumina sequencing, sequence assembly and sequencing data processing
cDNA libraries were subjected to 50 nt (single-end) sequencing using the HiSeq 2000 platform (Illumina). Raw sequencing reads were obtained and trimmed to remove adaptor sequences, empty reads and low-quality sequences using the FAST-X toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). These reads were assembled using Velvet (version 1.2.08) (Zerbino & Birney, 2008). The performance of Velvet assemblies is determined by hash length (k) and coverage cut-off (c). An iterative process was used to select the optimum k

and *c* values to produce the best assembly (Liu et al., 2011). The parameters used for assembly of contigs were *k* = 31 and multiple *c* (*c* = 2–10). To identify contigs derived from IMNV sequences, the resulting contig sets (≥ 100 nt) were aligned against IMNV genomic sequences by running BLAST (Altschul et al., 1990). To determine IMNV sequencing coverage, transcriptomic reads were clustered using the FAST-X toolkit, and clustered reads were subsequently mapped to the IMNV genome using a custom-designed Perl program. The Perl script was designed to allow mapping only once to the reference sequence, with no mismatched residues allowed. The number of reads that mapped to each nucleotide position was used to calculate coverage (fold) of each nucleotide.

A biological sequence alignment editor (BioEdit version 7.1.11; (http://www.mbio.ncsu.edu/bioedit/page2.html) was used as a platform for basic sequence manipulations of IMNV (e.g., sequence alignment, protein translation, ORF prediction, etc.). ClustalW2 was used for multiple sequence alignment (Larkin et al., 2007).

Identification of sequences at the 5' and 3' ends of the IMNV genome by RT-PCR

Oligonucleotide primers were designed to amplify novel IMNV sequence at both the 5' and 3' end of the genome (Table S1). Total RNA was extracted from shrimp inoculated with the Indonesia or Brazil strains of IMNV using TRIzol LS (Invitrogen). A one-step RT-PCR kit (QIAGEN) was used with the following amplification parameters: 50°C for 30 min, 95°C for 15 min and 35 cycles of 95°C for 15 s, 55°C for 30 s and 72°C for 45 s, followed by a final elongation at 72°C for 5 min. RT-PCR products were cloned into the pCR 2.1 vector (TOPO TA Cloning kit; Invitrogen) and transformed into *Escherichia coli* (One Shot TOP10; Invitrogen). Cells were grown on Luria-Bertani agar containing kanamycin (50 µg ml^{-1}). A subset of colonies was screened for the presence of the desired insert by PCR, and these inserts were then purified and sequenced with vector-specific primers (M13F and M13R) using a 3730xl DNA sequencer (Applied Biosystems) at the Iowa State University DNA facility.

RACE PCR

5' RACE PCR

To identify the nucleotide sequences at the 5' end of the IMNV genome sequence, 5' RACE was conducted. Total RNA (1 µg) was reverse transcribed using the appropriate primer (452R for IMNV Indonesia and 163R for IMNV Brazil) and the resulting cDNAs were purified using a Purelink Quick PCR Purification kit (Invitrogen). Oligo(dC) tails were added to the 3' end of 2 µg purified cDNA using 15 U terminal deoxynucleotidyltransferase (Invitrogen), 1 mM dCTP and 50 µg BSA ml^{-1} in 1× tailing buffer (100 mM potassium cacodylate, pH 7.2, 2 mM CoCl$_2$ and 0.2 mM DTT). Tailing reactions were performed at 37°C for 30 min and heat inactivated at 65°C for 10 min. C-tailed products were purified then PCR amplified using a consensus forward primer specific to the C-tailed termini (C tail F) and one of several reverse primers specific to the IMNV cDNA sequences (428R for IMNV Indonesia and 243R for IMNV Brazil). The resulting products were cloned and transformed into competent *E. coli* cells. Selected colonies were screened by PCR using both vector-

specific primers (M13F and M13R) and IMNV-specific primers (273F, 428R for IMNV Indonesia and 51F, 243R for IMNV Brazil). Select amplicons were purified (QIAquick PCR purification kit; QIAGEN) and sequenced with an IMNV-specific reverse primer (374R for IMNV Indonesia and 304R for IMNV Brazil).

3' RACE PCR

To identify the nucleotide sequences at the 3' end of the IMNV genome, synthetic poly(A) tails were added to 10 μg IMNV RNA using 6 U poly(A) polymerase A (Ambion) in 1× reaction buffer (40 mM Tris/HCl, pH 8.0, 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 250 mM NaCl, 50 mg BSA ml^{-1} and 1 mM ATP). Tailing reactions were incubated at 37°C for 1 h and then placed at –20°C for 2 min to inactivate the enzyme. Reaction products were purified by ethanol precipitation and poly(A)-tail-enriched RNAs were reverse transcribed using a reverse primer specific to the poly(A) tail (3' RT). Resulting cDNAs were then PCR amplified using the appropriate primers (IMNV7852F and 3'Rev). PCR products from the 3' RACE were also cloned and screened with gene specific primers (IMNV7852F and 3'Rev) and plasmid primers (M13F and M13R), and then sequenced with 7945F for both IMNV strains.

dsRNA synthesis

Two dsRNAs, each ~381 bp in length, were generated in the 639 bp of newly discovered sequence at the 5' end of the IMNV genome. Primers for dsRNA synthesis were deigned to amplify from positions 26 to 401 (IMNV26T7F and IMNV401T7R) and 233 to 613 (IMNV233T7F and IMNV613T7R) (Table S1, Fig. S4). The aforementioned dsRNAs were nearly identical in size to dsRNA95-475 (with the nucleotide range dictated by the position on the original IMNV genome sequence), which had proven protective effect in suppressing IMNV replication. dsRNA95-475 of IMNV original sequences was generated according to a previous study (Loy et al., 2012). dsRNA was synthesized using the Ambion Megascript RNAi kit according to the manufacturer's instructions. Briefly, a PCR was run containing primers with T7 sequences added to the 5' terminus of primers (Table S1). Complementary DNA was generated following the manufacturer's instructions for ThermoScript RT (Invitrogen) using total RNA extracted from IMNV-infected shrimp and forward primer IMNV26F. PCR was performed using PCR master mix (PuReTaq Ready-To-Go PCR Beads; GE Healthcare), cDNA template and the primers containing T7 extensions that were designed to specific regions of the new IMNV genome. PCR products were purified using a QIAquick PCR purification kit (QIAGEN), concentrated using a vacuum concentrator and 1 μg was added to an in vitro transcription reaction. The synthesis reaction was incubated at 37°C for 12–16 h, and then treated with RNase and DNase I. The formation of dsRNA was confirmed by 1% agarose gel electrophoresis and quantified by spectrophotometry (NanoDrop 2000; Thermo Scientific). The dsRNAs were stored in –20°C until use and were then diluted in RNase-free water to produce the desired concentration.

IMNV Brazil viral isolation

An IMNV isolate from Brazil was kindly provided by the Aquaculture Pathology Laboratory at the University of Arizona. Virus inoculum was clarified directly from frozen infected samples as described previously (Loy et al., 2012). Shrimp (7–10 g) were injected

intramuscularly (100 µl) with 1:100 dilution of IMNV Brazil [~7.2 × 10² virus copies (µl RNA)⁻¹] (Loy et al., 2012). Clinical signs of lethargy and white muscle were observed at 3 days p.i. Moribund shrimp with gross lesions were collected for RNA extraction using a QIAGEN RNeasy Mini kit according to the manufacturer's instructions. The RNA from shrimp infected with IMNV Brazil was tested by RT-PCR and sequence length was confirmed on a 1% agarose gel.

Determination of novel dsRNA disease protection efficacy in vivo

Shrimp (3–5 g) were divided into five groups of 10 shrimp with three replicates in each group and 10 shrimp for negative controls. Tanks were maintained at 29–30°C. Shrimp were injected intramuscularly with 5 µg per shrimp (100 µl) of dsRNA95-475, dsRNA26-401, dsRNA233-613, and dsEGFP or 2% NaCl, and shrimp were injected at 48 h p.i. with 2% NaCl (negative control group) or IMNV 1:100 dilution of IMNV Indonesia that reached 95–100% mortality within 2 weeks [~2.11 × 10³ virus copies (µl RNA)⁻¹] (Loy et al., 2012). Shrimp were fed and mortality observed twice per day. Tissue samples from moribund or dead shrimp were stored in –80°C for further quantification of virus copy number.

qRT-PCR

qRT-PCR for assay virus load was adapted from a previously described method (Andrade et al., 2007). RNA was extracted using a RNeasy Mini kit (QIAGEN) and qRT-PCR was performed with conditions from a previous report (Loy et al., 2012). Each reaction was run in duplicate on a Bio-Rad CFX96 Real-Time PCR Detection System using a standard as described previously (Andrade et al., 2007; Loy et al., 2012).

IMNV purification

IMNV was purified using a combination of methods adapted from Mello et al. (2011) and Poulos & Lightner (2006). Briefly, 100 shrimp (7–10 g) were infected with a 1:100 dilution of IMNV Indonesia (Loy et al., 2012). Moribund shrimp were collected during a 2-week course of infection to obtain ~500 g abdominal muscle tissue from acutely infected animals. Heads were removed to reduce the amount of lipid input into the virus purification process. Infected tissues were placed in 0.1 M PBS buffer, pH 7.5 plus 0.5 % of Na₂SO₄ (w/v) at a ratio of 1:3 and homogenized in a blender. Additionally, 10% chloroform (v/v) was added into the homogenate and was then centrifuged at 8000 g for 15 min to clarify the suspension. Following the first centrifugation, the supernatant was removed and centrifuged again to remove the tissue debris. The supernatant was collected and precipitated by adding 10% polyethylene glycol (w/v) and 4% NaCl (w/v), stirred at 4°C for 1 h, and centrifuged at 8000 g for 20 min at 4°C (Mello et al., 2011). The pellet was then resuspended and loaded onto 10–40% sucrose gradient and centrifuged at 70 000 g for 2 h at 4°C. A pellet was observed in the bottom of the ultracentrifuge tube at the 40% sucrose gradient. The virus pellet was removed from the ultracentrifuge tube and diluted in the same buffer to be stored at –80°C. The presence of virus was confirmed by RNA extraction and RT-PCR. Purified virus also was injected intramuscularly into 7–10 g shrimp to confirm the virulence and presence of virus in the preparation.

Northern blot analysis

Northern blot analysis was performed using methods adapted from Brown et al. (2001) and the manufacturer's instructions (NorthernMax-Gly; Life Technologies). Total RNA (5 mg) was resolved on a 1% agarose glyoxal gel for 1 h at 100 V, alongside a biotinylated RNA molecular mass marker (Millennium Marker; Ambion). The gel was observed for the presence of RNA via ethidium bromide staining in a UV light box prior to transfer. Then the gel was nicked with 500 kJ UV light and transferred to a positively charged membrane (Whatman Nytran Supercharge TurboBlotter; Sigma Aldrich) using a passive downhill transfer technique for 1 h. Membranes were UV cross-linked at 1200 kJ and hybridized overnight in ULTRAhyb Ultrasensitive Hybridization Buffer (Ambion) at 68°C with bioti-nylated RNA probes. Biotinylated RNA probes, including a probe to positions 95 (sense) and 475 (antisense), and to the new sequences at positions 233 (sense) and 613 (antisense) (Table S1), were generated using a BrightStar Psoralen-Biotin Nonisotopic Labelling kit (Ambion). PCR templates with T7 sequence at the 5' ends were used to generate in vitro transcripts (T7 RNA polymerase; Promega). RNA transcripts were then labeled with biotin according to the manufacturer's instruction (BrightStar Biotin-Psoralen kit). The hybrid-ized membranes were developed following the manufacturer's instruction and the signal was detected by chemiluminescence using a BrightStar BioDetect kit (Ambion).

Statistical analysis

Survival and qRT-PCR data were analyzed using one-way ANOVA followed by Tukey's multiple comparison test ($P < 0.05$) using JMP Pro 10.0.0 software (SAS Institute). Survival data and viral copy number were expressed as the mean ± SEM. Mean survival data were tested at the termination of experiment on day 14 p.i. and the mean log viral copy number was compared between groups.

Note – The GenBank/EMBL/DDBJ accession numbers for the complete genome sequences of IMNV Indonesia and Brazil strains are KF836757 and KJ556923, respectively.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.

Andrade, T. P. D., Srisuvan, T., Tang, K. F. J. & Lightner, D. V. (2007). Real-time reverse transcription polymerase chain reaction assay using TaqMan probe for detection and quantification of infec-tious myonecrosis virus (IMNV). *Aquaculture* 264, 9–15.

Assavalapsakul, W., Smith, D. R. & Panyim, S. (2006). Identification and characterization of a *Penaeus monodon* lymphoid cell-expressed receptor for the yellow head virus. *J Virol* 80, 262–269.

Brown, T., Mackey, K. & Du, T. (2001). Analysis of RNA by Northern and slot-blot hybridization. *Curr Protoc Mol Biol* 10, 12.1–12.14.

Escobedo-Bonilla, C. M., Wille, M., Alday Sanz, V., Sorgeloos, P., Pensaert, M. B. & Nauwynck, H. J. (2007). Pathogenesis of a Thai strain of white spot syndrome virus (WSSV) in juvenile, specific pathogen-free *Litopenaeus vannamei*. *Dis Aquat Organ* 74, 85–94.

Garlapati, S. & Wang, C. C. (2005). Structural elements in the 5′-untranslated region of giardiavirus transcript essential for internal ribosome entry site-mediated translation initiation. *Eukaryot Cell* 4, 742–754.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.

Lightner, D. V., Redman, R. M., Pantoja, C. R., Tang, K. F., Noble, B. L., Schofield, P., Mohney, L. L., Nunan, L. M. & Navarro, S. A. (2012). Historic emergence, impact and current status of shrimp pathogens in the Americas. *J Invertebr Pathol* 110, 174–183.

Liu, S., Vijayendran, D. & Bonning, B. C. (2011). Next generation sequencing technologies for insect virus discovery. *Viruses* 3, 1849–1869.

Loy, J. D., Mogler, M. A., Loy, D. S., Janke, B., Kamrud, K., Scura, E. D., Harris, D. L. & Bartholomay, L. C. (2012). dsRNA provides sequence-dependent protection against infectious myonecrosis virus in *Litopenaeus vannamei*. *J Gen Virol* 93, 880–888.

Mello, M. V., Aragão, M. E., Torres-Franklin, M. L., Neto, J. M. & Guedes, M. I. (2011). Purification of infectious myonecrosis virus (IMNV) in species of marine shrimp *Litopenaeus vannamei* in the State of Ceará. *J Virol Methods* 177, 10–14.

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, 1315–1318.

Pongsomboon, S., Wongpanya, R., Tang, S., Chalorsrikul, A. & Tassanakajon, A. (2008). Abundantly expressed transcripts in the lymphoid organ of the black tiger shrimp, *Penaeus monodon*, and their implication in immune function. *Fish Shellfish Immunol* 25, 485–493.

Poulos, B. T. & Lightner, D. V. (2006). Detection of infectious myonecrosis virus (IMNV) of penaeid shrimp by reverse transcriptase polymerase chain reaction (RT-PCR). *Dis Aquat Organ* 73, 69–72.

Poulos, B. T., Tang, K. F., Pantoja, C. R., Bonami, J. R. & Lightner, D. V. (2006). Purification and characterization of infectious myonecrosis virus of penaeid shrimp. *J Gen Virol* 87, 987–996.

Senapin, S., Phewsaiya, K., Briggs, M. & Flegel, T. W. (2007). Outbreaks of infectious myonecrosis virus (IMNV) in Indonesia confirmed by genome sequencing and use of an alternative RT-PCR detection method. *Aquaculture* 266, 32–38.

Tang, J., Ochoa, W. F., Sinkovits, R. S., Poulos, B. T., Ghabrial, S. A., Lightner, D. V., Baker, T. S. & Nibert, M. L. (2008). Infectious myonecrosis virus has a totivirus-like, 120-subunit capsid, but with fiber complexes at the fivefold axes. *Proc Natl Acad Sci USA* 105, 17526–17531.

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18, 821–829.

Supplemental Table and Figures:

S1. Sequence alignment of the originally published, 7,561 bp IMNV (Indonesia) genome sequence [EF061744.1] with full-length, 8,226 bp, genome sequence of IMNV (Indonesia) [KF836757.1] and IMNV (Brazil) [KJ556923.1]. Full-length sequence was generated based on RNA-seq and 5' and 3'RACE using IMNV-infected shrimp cDNA.   Top) alignment of the 5'- end of the sequence shows 643 new bases of sequence preceding the originally published genome start site. Bottom) alignment of the 3'- end of the sequence shows 23 new bases of sequence succeeding the originally published genome end site. Sequences were subjected to Clustal Omega alignment (Sievers *et al.*, 2011). *denotes strictly conserved nucleotide sequences.

S2. RT-PCR confirmation of the existence of novel IMNV sequences in both IMNV (Indonesia) and (Brazil) strain using four different sets of primers targeting the 5'-end (see Table S1). Lanes 1-4 and 5-8 show PCR products from template generated from *L. vannamei* infected with IMNV (Indonesia) and (Brazil), respectively. RT-PCR products from lanes 1-4 and 5-8 are 331, 704, 643 and 990 kb in length, respectively. Lane 9 is amplicon from primers designed to the original IMNV sequence (95F, 475R, Table S1); M = Marker. These primers also were tested on RNA from uninfected, specific pathogen-free (SPF) shrimp; RT-PCR results were negative at a variety of cycling conditions for all primers using template from uninfected shrimp (data not shown).

S3. RT-PCR confirmation of novel sequences at 3'-end of IMNV genome. Primers were designed to specifically amplify to IMNV (Indonesia) [lanes 1 and 2] and (Brazil) [lanes 3 and 4]. Amplicons are 352 kb (lanes 1 and 3) and 358 kb (lanes 2 and 4) in size. These primers were

tested with SPF shrimp cDNA and no amplification of virus was detected, indicating that sequences are present only in virus-infected tissues (data not shown).

S4. Annotated nucleotide and predicted translation of additional sequence at the 5'- end of the IMNV genome. The start site for the longest predicted ORF, in frame +1, is shown. The predicted ORF1 has additional 114 aa at the N-terminus. The nt sequence start from the original genome sequence, and originally predicted translation start site are denoted as 7561. Arrows also denote primers used to generate dsRNA for bioassays (see Table S1).

S5. Confirmation of suppression of IMNV replication in muscle tissue by qRT-PCR analysis. Virus load in animals subjected to dsRNA vaccines for novel and known regions of the IMNV genome. qRT-PCR results are shown for log transformed total genome copy numbers as calculated from a standard curve. X-axis is treatment and Y-axis is $\log_{10}$ viral genome copy number. Bars represent standard error within the sample. n=18 samples/treatment.

Supplemental Table 1. Oligonucleotide primer sequences for RT-PCR, dsRNA synthesis and RACE PCR

Primers	Sequence (5'- 3')
RT-PCR to confirm new IMNV genome sequence	
IMNV26F	TTCGTAGGCAAGAACGAAGTGGCT
IMNV312F	GTAGCATTACTTGGCGCATGCAGT
IMNV643R	ACTATGTGGACATGACACAGGGCA
IMNV1016R	TGATTGTATGCTGCCACCTCCAGA
IMNV8204R1	CGACTATAACCTAGGGCCAACCGC
IMNV8210R2	GACTATAACCTAGGGCC
IMNV7852F	TTGGACAAATGGTTCCCGGTCCAA
dsRNA synthesis*	
IMNV26T7F	<u>TAATACGACTCACTATAGGG</u>TTCGTAGGCAAGAACGAAGTGGCT
IMNV401T7R	<u>TAATACGACTCACTATAGGG</u>TTCGCTATACGCCGAGGTTACAGT
IMNV233T7F	<u>TAATACGACTCACTATAGGG</u>ACGAGTCGCATCCTTACGAGAAGT
IMNV613T7R	<u>TAATACGACTCACTATAGGG</u>TCTTACGTGAAGGACGAACGGGAA
5' RACE, IMNV (Indonesia)	
273F	ACGAGTCGCATCCTTACGAGAAG
452R	AGACATCCAATGGTATATGGACAG
428R	TCCAGTTTCTAGCATACAAGCCGG
C tail F	GACATCGAAAGGGGGGGGGGGG
374R	GCCGAGGTTACAGTCGGACTAG
5' RACE, IMNV (Brazil)	
51F	GCTATGGGGTTTGAGGTTTAGCCG
163R	GAGAACAACAGCTAATCGAGGACG
243R	TACTGCACTTCTCGTAAGGATGCG
304R	AAGTAATGCTACCCGTAACGACGG
3' RACE , IMNV (Brazil) and (Indonesia)	
3' RT	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTT
3' Rev	GGCCACGCGTCGACTAGTA
7945F	GAAATTCCAACTGAACCAGCAGTT
M13F	GTAAAACGACGGCCAG
M13R	CAGGAAACAGCTATGAC

* Underlined sequence corresponds to T7 primer added for *in vitro* transcription.

Figure S1.

```
Indonesia (7561)|EF061744.1|   ------------------------------------------------------------   0
Indonesia (8226)|KF836757.1|   ATTTTCTACATCTGGCCAAGGAAAATCTGTTCGTAGGCAAGAACGAAGTGGCTATGGGGT   60
Brazil (8226)   |KJ556923.1|   ATTTTCTACATCTGGCCAAGGAAAATCTCTTCGTAGGCAAGAACGAAGTGGCTATGGGGT   60

                    -----------------SKIP------------------

Indonesia (7561)|EF061744.1|   ----------------------------------------------GGCAATTTCAACCTAAT   17
Indonesia (8226)|KF836757.1|   TCGTCCTTCACGTAAGAAGGATGACTATGTGGACATGACACAGGGCAATTTCAACCTAAT   660
Brazil (8226)|KJ556923.1|      TCGTCCTTCACGTAAGAAGGATGACTATGTGGACATGACACAGGGCAATTTCAACCTAAT   660
                                                                             *****************

Indonesia (7561)|EF061744.1|   TCTAAAACTGTTACCAACAATTTCTGAACTACAGAAGAGGCATATTCTTCATCTACTTCG   77
Indonesia (8226)|KF836757.1|   TCTAAAACTGTTACCAACAATTTCTGAACTACAGAAGAGGCATATTCTTCATCTACTTCG   720
Brazil (8226)|KJ556923.1|      TCTAAAACTGTTACCAACAATTTCTGAACTACAGAAGAGGCATATTCTTCATCTACTTCG   720
                               ************************************************************

                    -----------------SKIP------------------

Indonesia (7561)|EF061744.1|   CAAAGTAGTTTACAACAAATGTATAGGTACTAAATTAAGGGACCAATAAAGAAACTTCGA   7517
Indonesia (8226)|KF836757.1|   CAAAGTAGTTTACAACAAATGTATAGGTACTAAATTAAGGGACCAATAAAGAAACTTCGA   8160
Brazil (8226)|KJ556923.1|      CAAAGTAGTTTACAACAAATGTATAGGTACTAAATTAAGGGACCAATAAAGAAACTTCGA   8160
                               ************************************************************

Indonesia (7561)|EF061744.1|   GTTTCCTATAACACATTCCCAGTTGGGTTTTGTGGCCAGCCATG----------------   7561
Indonesia (8226)|KF836757.1|   GTTTCCTATAACACATTCCCAGTTGGGTTTTGTGGCCAGCCATGCGGTTGGCCCTAGGTT   8220
Brazil (8226)|KJ556923.1|      GTTTCCTATAACACATTCCCAGTTGGGTTTTGTGGCCAGCCATGCGGTTGGCCCTAGGTT   8220
                               ********************************************

Indonesia (7561)|EF061744.1|   ------
Indonesia (8226)|KF836757.1|   ATAGTC      8226
Brazil (8226)|KJ556923.1|      ATAGTC      8226
```
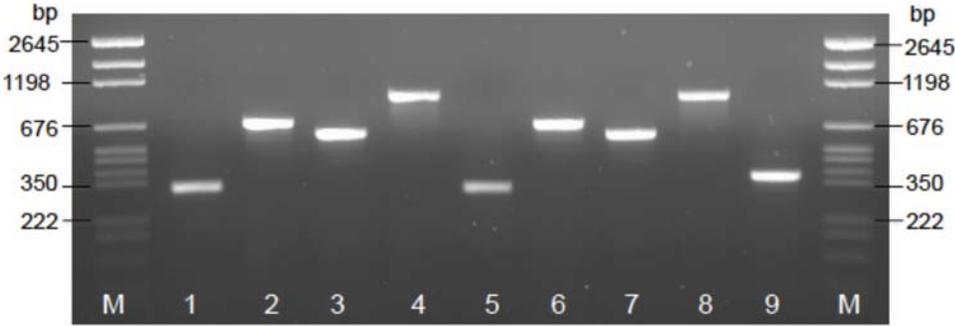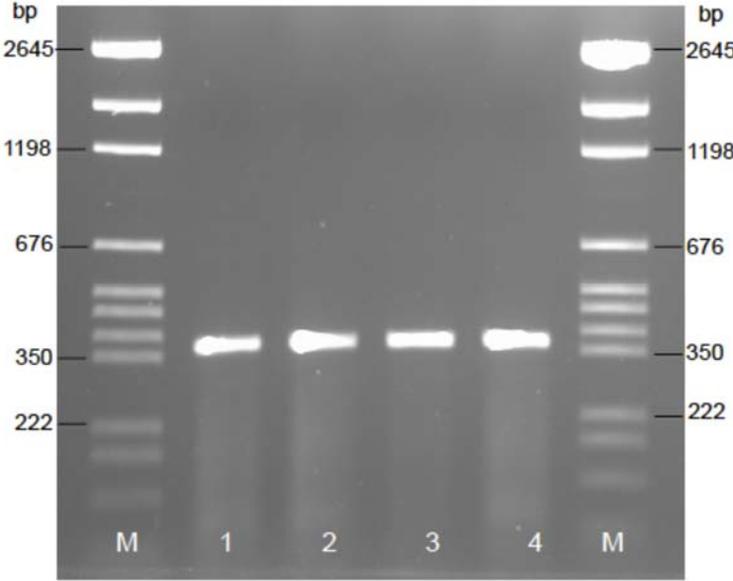
Figure S2.

Figure S3.

Figure S4.

```
                           ▼26F
TCTACATCTGGCCAAGGAAAATCTGTTCGTAGGCAAGAACGAAGTGGCTATGGGGTTTGAGGTTTAGCCG

AAAGAGATCTCGACAGTCTTTAGTAGGATACACTTCTTACACTCCTGGGAAAGAGAAACATTATCTGTCC

AGCAATGATACACGGGTACGTCCTCCATTAGCTGTTGTTTTCTTTTCCTCGGCAGGTCCTACAATGTCGG
                        ▼233F
AGCGCAGGAACAGGACGAACCCACGAGTCGCATCCTTACGAGAAGTGCAGTAGGTTGGCGCTGTTGAGAG

CGGGTGATCTGGGTAATGTCCGTCGTTACGGGTAGCATTACTTGGCGCATGCAGTGGGAAGTCATGGCCC
                              ▼401R
ACGTCACCTTAACAAAAGACTAGTCCGACTGTAACCTCGGCGTATAGCGAAAGCGGTTAGATAACATTAA
                   ★
ATACCGGCTTGT ATG CTA GAA ACT GGA CTG TCC ATA TAC CAT TGG ATG TCT CAA ACA AAG
             M   L   E   T   G   L   S   I   Y   H   W   M   S   Q   T   K

AAT AAA GAA CCT GTT GTT GAA CAG CAA AAT CAA CAA ACC TTC AAC CAA AAA GAC CAC A
 N   K   E   P   V   V   E   Q   Q   N   Q   Q   T   F   N   Q   K   D   H
                                                          ▼613R
AT GCG CTA GTG ATT GAA CAC AAA ACT AGT GCA ACA ACG AGC GCC CAA AGC ATT CCC GT
 N   A   L   V   I   E   H   K   T   S   A   T   T   S   A   Q   S   I   P   V
                              ▼643R                    ▼7561
T CGT CCT TCA CGT AAG AAG GAT GAC TAT GTG GAC ATG ACA CAG GGC AAT TTC AAC CTA
   R   P   S   R   K   K   D   D   Y   V   D   M   T   Q   G   N   F   N   L

ATT CTA AAA CTG TTA CCA ACA ATT TCT GAA CTA CAG AAG AGG CAT ATT CTT CAT CTA C
 I   L   K   L   L   P   T   I   S   E   L   Q   K   R   H   I   L   H   L
               ▼95F
TT CGA GAA GAA GTA GAA GGA AAG AAA GTT TGT TTC GTA AAG CGA GAA AAA CAG AAT CC
 L   R   E   E   V   E   G   K   K   V   C   F   V   K   R   E   K   Q   N   P
         ★7561
A CTT ATG GCT ATA AAC GAA CTA GCC GTT AAA GTT GGA GAG AAG CCT AAA TAC ACA TCG
 L   M   A   I   N   E   L   A   V   K   V   G   E   K   P   K   Y   T   S

ACGAAAACCGGAGCTGACCACATTCCAAGCTGGACTGTATTGGTTGAGTTCGCAGGTTTTAGCGAAGCAGCGACATG

TGACACAGTTAAAAACGCAAAAATGATTGCTGCTTACAAATTAGTTAAAAGATTTTGTAAATGGGACCCAACCTACA

TTGAAATTTCTGATTGTATGCTGCCACCTCCAGACCTTACATCGTGCGGGGACGTTGAGAGTAATCCTGGACCTATC
                       ▼475R
ATACATAGCGTTGCATTTGCAAGAACTGGTTCAGTATGGACACCTGCCACCTTTACTTTCAATACTACATCATCCCC
```

Figure S5.
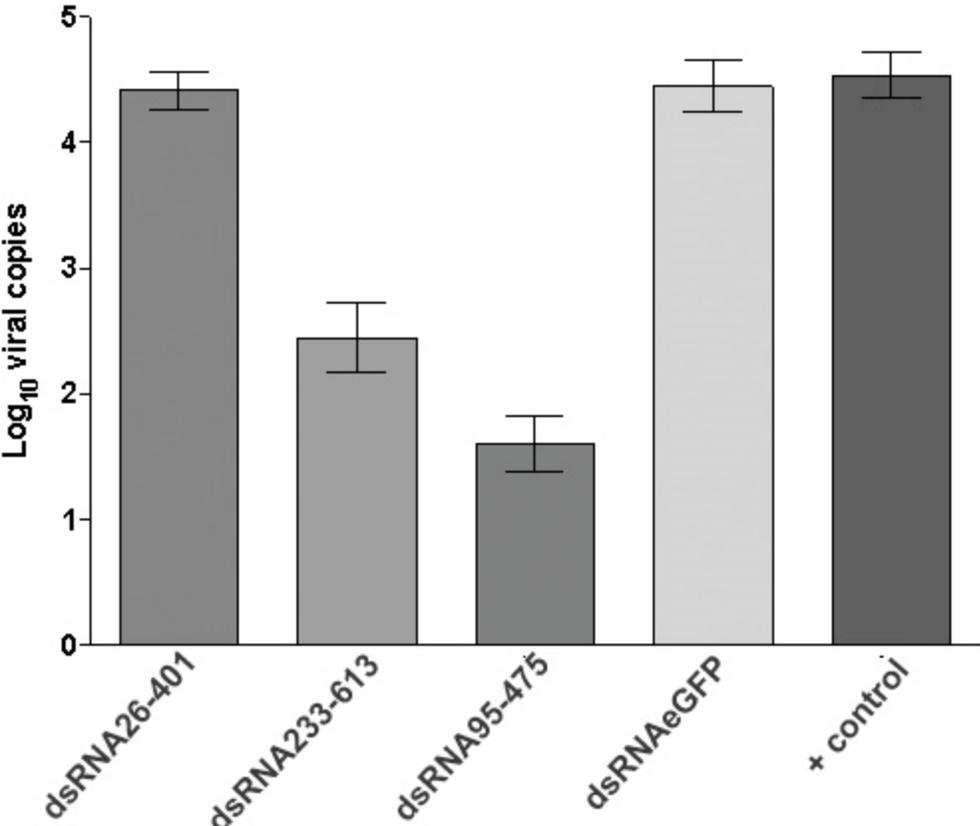
References

Sievers, F, Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**:539.