Virus Research 203 (2015) 66–71

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Short communication

Analysis of new isolates reveals new genome organization and a hypervariable region in infectious myonecrosis virus (IMNV)

Márcia Danielle A. Dantas[a,b], Suely F. Chavante[b], Dárlio Inácio A. Teixeira[c], João Paulo M.S. Lima[b,d], Daniel C.F. Lanza[a,b,*]

[a] Laboratório de Biologia Molecular Aplicada, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[b] Programa de Pós-Graduação em Bioquímica, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[c] Escola Agrícola de Jundiaí, Universidade Federal do Rio Grande do Norte, Macaíba, RN, Brazil
[c] Laboratório de Sistemas Metabólicos e Bioinformática, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil

ARTICLE INFO

Article history:
Received 17 February 2015
Received in revised form 25 March 2015
Accepted 27 March 2015
Available online 4 April 2015

Keywords:
IMNV Brazil
Hypervariable region
dsRNA virus
Shrimp diseases
IRES

ABSTRACT

Infectious myonecrosis virus (IMNV) has been the cause of many losses in shrimp farming since 2002, when the first myonecrosis outbreak was reported at Brazilian's northeast coast. Two additional genomes of Brazilian IMNV isolates collected in 2009 and 2013 were sequenced and analyzed in the present study. The sequencing revealed extra 643 bp and 22 bp, at 5′ and 3′ ends of IMNV genome respectively, confirming that its actual size is at least 8226 bp long. Considering these additional sequences in genome extremities, ORF1 can starts at nt 470, encoding a 1708 aa polyprotein. Computational predictions reveal two stem loops and two pseudoknots in the 5′ end and a putative stem loop and a slippery motif located at 3′ end, indicating that these regions can be involved in the start and termination of translation. Through a careful phylogenetic analysis, a higher genetic variability among Brazilian isolates could be observed, comparing with Indonesian IMNV isolates. It was also observed that the most variable region of IMNV genome is located in the first half of ORF1, coinciding with a region which probably encodes the capsid protrusions. The results presented here are a starting point to elucidate the viral's translational regulation and the mechanisms involved in virulence.

Infectious myonecrosis (IMN) affects global shrimp farming causing significant production losses in countries such as Brazil and Indonesia (Nunes et al., 2004; Andrade et al., 2007; Senapin et al., 2007; Lightner, 2011). Myonecrosis, the main disease symptom, affects shrimp cephalothorax and abdominal muscles, evolving to an opaque appearance. In advanced stages, muscles and appendages appear reddish, resembling a "roasted" aspect. Infectious myonecrosis virus (IMNV), the causative agent of IMN, possess a dsRNA monopatite genome, packaged in an icosahedral capsid (Lightner et al., 2004; Poulos et al., 2006; Tang et al., 2008).

Previous studies reported that the IMNV genome comprises a 7561 bp long sequence, including the following annotations: a 5′ UTR (the first 135 nucleotides of genome); the putative RNA binding protein coding region (nt positions 136-414); a region encoding two small proteins, possible candidates to form the viral protrusions (nt 415–2247); a major capsid protein (MCP) coding region (nt 2248–4953); a RNA dependent RNA polymerase (RdRp) coding region (nt 4752–7490); and the 3′ UTR region comprising the last 71 nucleotides of genome (Poulos et al., 2006; Nibert, 2007; Tang et al., 2008). Based on the RdRp sequence, genomic organization and capsid characteristics, IMNV was classified as a member of Totiviridae family (Wang et al., 1993; Poulos et al., 2006; Tang et al., 2008). Recently, two phylogenetic studies based on RdRp sequence comparisons introduced IMNV as a member of a new monophyletic group called IMNV-like, closely related to the *Giardia virus* clade (Liu et al., 2012; Oliveira et al., 2014).

In the present study, the genomes of two Brazilian IMNV isolates collected in 2009 (Brazil_2009) and 2013 (Brazil_2013), in Pernambuco and Rio Grande do Norte states, respectively, were sequenced. Amplification and sequencing was carried out using the primers described by Senapin et al. (2007), and genome was assembled using the Geneious software (v. 8.0.5). As shown in Table 1, Brazil_2009 and Brazil_2013 are very similar (99.42%) and shared high similarity (at least 98.16%) with other isolates available in GenBank. During analysis, a sequence of an Indonesian

* Corresponding author at: Laboratório de Biologia Molecular Aplicada – LAPLIC, Departamento de Bioquímica, Centro de Biociências, Universidade Federal do Rio Grande do Norte, CEP: 59072-970, Natal, RN, Brazil. Tel.: +55 84 3215 3416; fax: +55 84 32153415.
E-mail address: danielclanza@gmail.com (D.C.F. Lanza).

http://dx.doi.org/10.1016/j.virusres.2015.03.015
0168-1702/

Table 1
Comparison of IMNV partial genomes (7561 bp) available so far.

	Brazil_2009	Brazil_2013	AY570982.2	EF061744.1	KJ636782.1	KJ636783.1	KF836757.1
Brazil_2009	100.00	99.42	98.90	98.66	98.19	98.19	98.24
Brazil_2013	99.42	100.00	98.90	98.68	98.16	98.18	98.24
AY570982.2_Brazil	98.90	98.90	100.00	99.60	99.13	99.00	99.21
EF061744.1_Indonesia_2006	98.66	98.68	99.60	100.00	99.26	99.26	99.34
KJ636782.1_Indonesia_2011	98.19	98.16	99.13	99.26	100.00	98.65	99.84
KJ636783.1_Indonesia_2012	98.19	98.18	99.00	99.26	98.65	100.00	98.73
KF836757.1_Indonesia_2013	98.24	98.24	99.21	99.34	99.84	98.73	100.00

isolate (accession number KF836757.1) drew our attention. This sequence has 8226 bp, showing 643 and 22 additional nucleotides at the 5′ and 3′ ends, respectively, when compared to the previously reported IMNV genome (7561 bp long). ORF1 shows 342 additional nucleotides in frame, starting at an AUG codon located at nt positions 437–439 and its putative encoding product has an increment of 114 amino acids. Using the primers described in Table 2, we amplified and sequenced this additional 5′ sequence from a Brazil_2009 sample. The full amplification of the 5′ region of the Brazil_2013 was not effective, probably due to some polymorphisms in primer annealing regions. The last 22 nt of genome were sequenced from both Brazilian isolates. This result confirms that IMNV genomic sequence has being underestimated. As observed in the KF836757.1 sequence entry, the genome of Brazil_2009 has at least 8226 bp, and the genome of Brazil_2013 has at least 7767 bp (Supplementary Fig. 1).

Supplementary Fig. 1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2015.03.015

On August 28, 2014, the IMNV sequence deposited by Poulos et al. in 2006 (accession number AY570982.3) was rectified and the additional nucleotides in genome extremities were added. From an alignment of our new sample sequence, Brazil_2009, and GenBank sequences AY570982.3 and KF836757.1 it was possible to observe a polymorphic site at nt position 469, which replaces a guanine to adenine, generating a UGA stop codon. Thus, we considered that ORF1 comprises 5127 nt (encoding 1708 aa), starting from the AUG at nt positions 470–472 to the in-frame stop codon UAA at nt positions 5594–5596, leaving a 5′ untranslated region of 469 nt. Taking these observations into account, it can be considered that ORF1 has an increment of 342 nucleotides if compared to previous reported IMNV genome. In addition, ORF2 covers 2739 nt (encoding 912 aa) starting at nt positions 5395–5397 and ending at the stop codon located at nt positions 8131–8133, leaving a 3′ untranslated region of 93 nt. ORF2 starts 199 nt upstream the end of ORF1 (nt positions 5395–5593) indicating that ORF1 and OFR2 are overlapped (Fig. 1A and Supplementary Fig. 1).

In silico predictions using RNAfold, using no dangling end energies (Gruber et al., 2008; Lorenz et al., 2011), revealed four conserved stem loops with high probability of occurrence in the first 770 nt of the 5′ end, (nt positions 335–386, 544–574, 584–646 and 711–733) (Supplementary Fig. 2A). At least four pseudoknot regions with lowest free energy were predicted by both HPknotter (Huang et al., 2005) and DotKnot (Sperschneider and Datta, 2010). All structures predicted are shown in Supplementary Table 1. Two of pseudoknots (nt positions 327–356 and 705–732) coincided with the stem loops predicted by RNAfold. Thereafter, we consider that these regions comprising nt 327–386 and 705–733 are the most probable pseudoknots. The potential of these regions to form stable pseudoknots was also validated by IPknot (Sato et al., 2011) and Mfold (Zuker, 2003) (Supplementary Fig. 3A and B). In short, the computational predictions indicate the existence of at least four RNA secondary structures at 5′ end: two pseudoknots (nt positions 327–386 and 705–733), and two stem loops, called here H1 (nt 544–574) and H2 (nt 584–646) (Fig. 1B). We speculate that these regions can form an Internal Ribosome Entry Site (IRES), similar to that reported in *Giardia lamblia virus* (GLV) (Garlapati and Wang, 2004, 2005). The IRES are alternative sites to start the mechanism of translation adopted by many viruses and also observed in some cellular higher eukaryotes mRNAs (Hellen and Sarnow, 2001; Stoneley and Willis, 2004; Baird et al., 2006). In GLV, a mechanism involving a highly structured 5′ UTR and its combination to a 264 bases stretch of the downstream coding sequence is required to starting the translation process (Yu and Wang, 1996; Garlapati and Wang, 2004). As observed in GLV IRES (Garlapati and Wang, 2005, 2009), two pseudoknots, one before and another after the start codon, were predicted in IMNV 5′ end (Fig. 1B). Additionally, the analysis of the last 176 nt from the 3′ extremity of IMNV genome revealed a hairpin (nt 8151–8165), and a pseudoknot (nt 8184–8213), which was also confirmed by IPknot and Mfold (Fig. 1B and Supplementary Figs. 2B and 3C). Moreover, a slippery motif GGGUUUU (nt 8185–8191) was observed inside the pseudoknot located in 3′ region. A RNA pseudoknot located just downstream of the slippery sequence, as observed here, is an element required for high-efficiency frameshifting and can act to slow or stall the ribosome (Brierley et al., 1989; Somogyi et al., 1993).

Supplementary Figs. 2 and 3 and Table 1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2015.03.015

Even though the number of IMNV complete genomes available to date is not sufficient for a robust analysis, the phylogenetic relationship among seven 7561 bp genomes was investigated by Bayesian MCMC approach using BEAST package v.1.8 (Drummond et al., 2012). Nucleotide sequences were aligned using MAFFT (v.6.85; Katoh and Toh, 2010; Katoh and Frith, 2012) with the L-INS-I parameter, gap opening penalty 1.53 and offset value 0.1. The best-fit nucleotide substitution model for the dataset, estimated using MEGA 6.06 (Tamura et al., 2013), was Tamura-Nei (TN93) with gamma distribution rates. The MCMC was performed for a sufficient number of generations to ensure convergence of all parameters (*i.e.*, length of chain 10,000,000) and strict (constant) molecular clock and a coalescent constant size tree prior were adopted. A consensus tree was estimated discarding 10% of trees. The tree was edited using FigTree (Drummond and Rambaut, 2007). Two well-defined groups were observed in the obtained dendogram (Fig. 2). Brazil_2009 and Brazil_2013 genomes were clustered as an isolated group. The Brazilian isolate collected in 2003 (AY570982.3) appears at the most basal branch of a group that comprises all Indonesian isolates, as previously observed by

Table 2
Primers used in this study to amplify IMNV additional sequences.

Region	Primer name	Sequence 5′ → 3′	TM, °C	Expected product size (bp)
1	F1-8226	ATTTTCTACATCTGGCCAAG	53.23	621
	R1-8226	TCCTTCTTACGTGAAGGAC	53.84	
2	F2-8226	GTATAGCGAAAGCGGTTAGA	55.11	619
	R2-8226	GGTGGCAGCATACAATCA	54.97	
3	Fend-8226	GGAATGGATACAGAAAGTGC	54.06	625
	Rend-8226	GACTATAACCTAGGGCCAAC	54.63	