Shrimp parvovirus circular DNA fragments arise from both endogenous viral elements (EVE) and the infecting virus

Suparat Taengchaiyaphum[1], Phasini Buathongkam[2], Suchitraporn Sukthaworn[3], Prapatsorn Wongkhaluang[1], Kallaya Sritunyalucksana[1*], Timothy William Flegel[4,5#]

[1]Aquatic Animal Health Research Team (AQHT), Integrative Aquaculture Biotechnology Research Group, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Yothi office, Rama VI Rd., Ratchathewi, Bangkok 10400, Thailand
[2]Department of Biotechnology, Faculty of Science, Mahidol University, Rama VI Rd., Ratchathewi, Bangkok 10400, Thailand
[3]National Cancer Institute, Rama VI Rd., Ratchathewi, Bangkok, 10400 Thailand
[4]Center for Excellence in Shrimp Molecular Biology and Biotechnology (Centex Shrimp), Faculty of Science, Mahidol University, Rama VI Rd., Ratchathewi, Bangkok 10400, Thailand
[5]National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), 113 Paholyothin Rd., Khlong Luang, Pathum Thani, 12120 Thailand

[#]Corresponding author:
Prof. Timothy W. Flegel (Email: tim.flegel@gmail.com)

***Co-corresponding author**
Dr. Kallaya Sritunyalucksana (Email: kallaya@biotec.or.th)

List of abbreviations:
vcDNA; viral copy DNA(s), **cvcDNA;** circular viral copy DNA, **IHHNV;** Infectious hypodermal and hematopoietic necrosis disease virus, **PS- DNase;** Plasmid- safe DNase, **EVE;** Endogenous viral element(s), **siRNA**; small interfering RNA(s)

ABSTRACT

Some insects use endogenous reverse transcriptase (RT) to make variable linear and circular viral copy DNA (vcDNA) fragments from viral RNA. The vcDNA produces small interfering RNA (siRNA) variants that inhibit viral replication via the RNA interference (RNAi) pathway. The vcDNA is also autonomously inserted into the host genome as endogenous viral elements (EVE) that can also result in RNAi. We hypothesized that similar mechanisms occurred in shrimp. We used the insect methods to extract circular viral copy DNA (cvcDNA) from the giant tiger shrimp (*Penaeus monodon*) infected with a virus originally named infectious hypodermal and hematopoietic necrosis virus (IHHNV). Simultaneous injection of the extracted cvcDNA plus IHHNV into whiteleg shrimp (*Penaeus vannamei*) resulted in a significant reduction in IHHNV replication when compared to shrimp injected with IHHNV only. Next generation sequencing (NGS) revealed that the extract contained a mixture of two general IHHNV-cvcDNA types. One showed 98 to 99% sequence identity to GenBank record AF218266 from an extant type of infectious IHHNV. The other type showed 98% sequence identity to GenBank record DQ228358, an EVE formerly called non-infectious IHHNV. The startling discovery that EVE could also give rise to cvcDNA revealed that cvcDNA provided an easy means to identify and characterize EVE in shrimp and perhaps other organisms. These studies open the way for identification, characterization and use of protective cvcDNA as a potential shrimp vaccine and as a

tool to identify, characterize and select naturally protective EVE to improve shrimp tolerance to homologous viruses in breeding programs.

KEYWORDS: shrimp; immunity; viral accommodation; circular DNA; circular viral copy DNA (cvcDNA); endogenous viral elements; EVE; IHHNV; RNAi;

1. INTRODUCTION

In 2009 (Flegel, 2009), it was hypothesized that endogenous viral elements (EVE) with high sequence identity to extant viruses in shrimp and insects could arise via host recognition of viral messenger RNA followed by formation of variable cDNA fragments (here called viral copy DNA or vcDNA) from it by host reverse transcriptase (RT). Integration of those vcDNA fragments into the host genome is via host integrase (IN). The EVE would give rise to negative sense RNA that would result in degradation of viral RNA by the RNA interference (RNAi) pathway. It was proposed that this was the underlying natural mechanism that leads to balanced persistent infections in which one or more viruses are tolerated by shrimp and insects, sometimes for a lifetime, without signs of disease. This phenomenon of tolerance to persistent viral infections had been called viral accommodation (Flegel & Pasharawipas, 1998; Flegel, 2001; Flegel, 2007) but the underlying mechanisms involving EVE were not hypothesized until 2009 (Flegel, 2009; Flegel, 2020). Viral accommodation via EVE constitutes a process of autonomous genetic modification (AGMO) that gives rise to natural transgenic organisms (NTO), and accommodation is heritable if the EVE occur in germ cells. Predictions of the hypothesis have been supported by research on insects since 2013 (Goic, et al., 2013; Goic, et al., 2016; Poirier, et al., 2018; Tassetto, et al., 2017; Tassetto, et al., 2019; Whitfield, et al., 2017) and proof of a protective EVE against a virus in mosquitoes was published in 2020 (Suzuki, et al. 2020). An updated summary diagram of the current mechanisms related to viral accommodation is shown in **Fig. 1**.

Not predicted by the viral accommodation hypothesis of 2009 was the discovery that vcDNA occurs in both linear (lvcDNA) an circular (cvcDNA) forms that, in turn, produce small interfering RNA (siRNA) transcripts that result in an immediate and specific cellular and systemic RNAi response to invading viruses (Goic, et al., 2016; Poirier, et al., 2018; Tassetto, et al., 2017). Although all these discoveries were made using RNA virus models, we considered it possible that they might also occur in shrimp since they too have been reported to have EVE homologous to extant DNA viruses (Saksmerprome, et al., 2011; Utari, et al., 2017; Taengchaiyaphum, et al., 2019). We were particularly interested in cvcDNA and the possibility that shrimp would produce protective cvcDNA in a manner similar to that reported for insects (Poirier, et al., 2018). We hypothesized that use of the techniques devised for extraction of cvcDNA from insects would be successful when used with the giant tiger shrimp (*Penaeus monodon*) infected with *Penstylhamaparvovirus* 1 from the family *Parvoviridae* and sub-family *Hamaparvovirinae* (Pénzes, et al., 2020). This virus was previously called infectious hypodermal and hematopoietic necrosis virus (IHHNV) and we will use that acronym here to maintain easy links to previous literature. We also hypothesized that the extracted cvcDNA would significantly reduce IHHNV replication in the whiteleg shrimp *Penaeus vannamei* challenged with IHHNV.

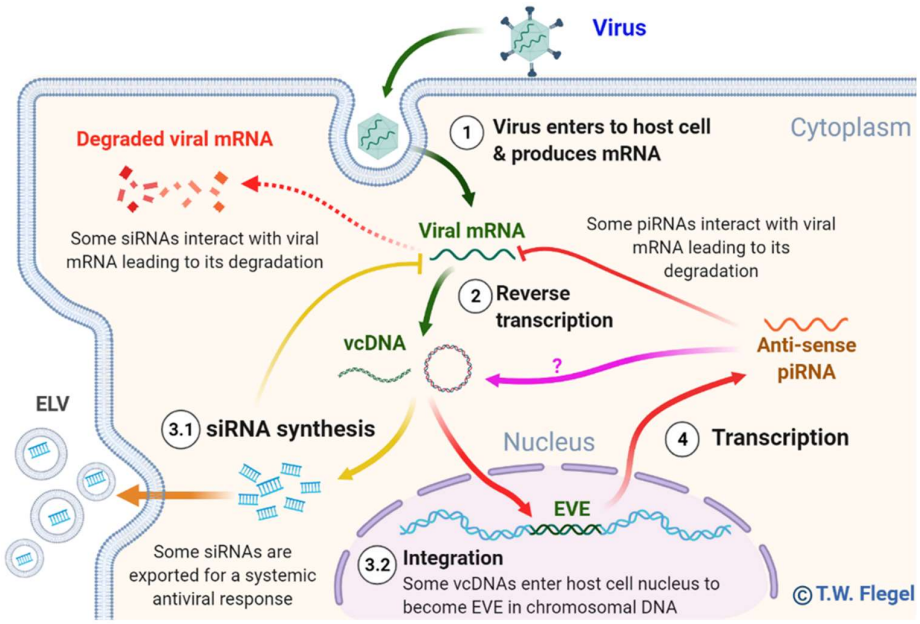

Figure 1. A simplified diagram of the mechanisms involved in viral accommodation as updated from Flegel (2020). The update includes additional pathways (indicated by yellow and orange arrows) that were not foreseen in the 2009 viral accommodation hypothesis (green and red arrows). Specifically, vcDNA was not predicted to occur also in a circular form. In addition, no immediate production of siRNA leading to an RNAi response was predicted. Nor was the occurrence of exosome like vesicles (ELV) for systemic dispersal of the RNAi response predicted. Nor was the discovery (this paper) that EVE could produce viral circular DNA (purple arrow). All these features are hypothesized to occur in shrimp. Schematic illustration was created by BioRender.com (https://app.biorender.com/).

2. MATERIALS AND METHODS

2.1. PCR methods and primers used in this study

The PCR primers used in this study are shown in **Table 1.** To determine the presence of infectious IHHNV and to test its replication level in challenged shrimp, a long-amp IHHNV detection method was used to detect a 3665 bp-region of IHHNV (approximately 92% of the whole genome and excluding its hairpin ends). A Long-Amp™ *Taq* PCR mix (New England Biolab, USA) was used with a total of 35 PCR cycles. The PCR reaction consisted of Long-Amp *Taq* PCR reaction mix, 0.4 µM of forward and reverse primers (98F/3762R), 1U Long-Amp™ *Taq* polymerase, and either 100 ng DNA before digestion or 2 ng DNA post enzyme digestion. The PCR cycle was started with initial denaturation at 94°C for 30 s then followed by 35 cycles of 94 °C for 20 s, 55 °C for 30 s, 72 °C for 2:30 min and final extension at 72 °C for 10 min.

The quantitative PCR by droplet digital PCR (IHHNV-ddPCR) was used to check the number of viral copies in the crude IHHNV stock and the number of IHHNV-cvcDNA in the circular DNA preparation. The ddPCR reaction was prepared by using EvaGreen™ ddPCR supermix (Bio-Rad, USA) which consisted of 1X ddPCR mix, 0.2 µM of forward and reverse primers (309NF/309NR), and 1 µl of diluted crude viral stock (at 10^{-7} dilution) or 1 ng of circular DNA. The ddPCR amplification cycle was set according to the manufacturer's protocol by adjusting the annealing temperature to 56 °C. After the complete PCR cycles, the reactions were analyzed by fluorescent signal using a ddPCR plate

reader. The absolute amount of target DNA copy per reaction was calculated based on Pearson's correlation method using QuantaSoft™ ddPCR analysis software (Bio-Rad, USA). PCR reactions for each individual sample were performed in duplicate.

The short-amp IHHNV-PCR method (Tang, et al., 2007) was used to check for infectious IHHNV sequences in DNA extracts and in infected shrimp. As an internal control gene for linear, chromosomal DNA, primers specific to shrimp elongation factor 1 alpha (EF-1α) gene were used to give an amplicon of 122 bp (Wongsurawat. T., et al. 2010). PCR amplicons were analyzed by 1.5% agarose gel electrophoresis followed by visualization of ethidium bromide staining by UV light.

Table 1. List of the primers used in this study.

PCR method with primer sequences	Expected amplicon size (bp)	Reference
Long-amp IHHNV-PCR		
IHHNV98F	3,665	AF218266
CCCAGTTTCTAACTGACGAGTGAAGAGA		This study
IHHNV3762R		
CCTGACTCTAAATGACTGACTGACGATAGGG		
IHHNV-ddPCR		
IHHNV309NF	157	AF218266
AAACAACTATGGACCCGTACC		This study
IHHNV309NR		
TCCACTGCATATTGTCGTAGTC		
Short-amp IHHNV-PCR		
IHHNV309F		AF218266
TCCAACACTTAGTCAAAACCAA	309	Tang KF, et al., (2006)
IHHNV309R		
TGTCTGCTACGATGATTATCCA		
IHHNV-cvcDNA joining amplification		
IHHNV3031F	800-1500	AF218266
CTAAGGAAGCCGACGTAACC		DQ228358
IHHNV3766F		This study
AGCTTGGATGCAAGCGATGTC		
IHHNV128R		
TGGACCTGGGGTGAGAAGGC		
***Pm*mtDNA-F**	150	This study
AAGAGATTTAGAGTAGGAGGAGCA		
***Pm*mtDNA-R**		
GCAGGAGGTCAACAACTACC		
***Pm*Ef-1α-F**	122	Wongsurawat. T., et al. (2010)
TTCCGACTCCAAGAACGACC		
***Pm*Ef-1α-R**		
GAGCAGTGTGGCAATCAAGC		

2.2 Preparation of crude IHHNV stock

The black tiger shrimp, *P. monodon* were checked for IHHNV-infection using the long amp-IHHNV detection method. The pleopods from the 5 infected shrimp were collected, homogenized and dissolved

in cold 1X PBS pH 7.4. The tissue homogenate was centrifuged at 8,000 rpm to remove cell debris before it was subjected to filtration through a 0.2 µm membrane filter. The filtrate was collected and aliquoted into small tubes and referred to as "crude IHHNV stock". The crude IHHNV stock was subsequently used in the challenge tests using *P. vannamei* where infection was confirmed by IHHNV presence and replication using PCR. The crude IHHNV stock was stored at -80°C for further experiments. To check the virus titer, crude IHHNV stock was serially diluted and subjected to quantification by the IHHNV-ddPCR.

2.3 Extraction of circular DNA from IHHNV-infected shrimp

Total DNA extract from pleopods of the IHHNV-infected *P. monodon* was subjected to circular DNA isolation as previously described (Poirier, et al., 2018). Briefly, 10 µg of total shrimp DNA was prepared using a DNA extraction kit (Qiagen, USA) and DNA concentration was determined by a NanoDrop spectrophotometer (Thermo Scientific, USA). Extraction of circular DNA from total DNA was performed by enzymatic digestion of linear DNA. The circular DNA extraction protocol is described in **Fig. 2**. During the extraction process, the *Xho*I (or *Not*I) enzyme in the protocol was used to cut shrimp chromosomal DNA into smaller fragments in order to accelerate DNA digestion by the Plasmid Safe-exonuclease (PS-DNase, Epicentre, UK). The enzymes *Xho*I (or *Not*I) were chosen because they have no cutting site in shrimp mitochondrial DNA or in the IHHNV genome. If there were IHHNV-cvcDNA entities that contained portions of host DNA, they might be cut by these enzymes and lost during circular DNA preparation. After digestion and extraction, the quantity of putative circular DNA was determined by Qubit fluorometer (Invitrogen, USA).

2.4 Confirmation and quality of circular DNA and circular viral copy IHHNV-DNA (IHHNV-cvcDNA)

To confirm the presence of circular DNA and lack of linear chromosomal DNA in the circular DNA preparation, PCR tests were carried out using 1) elongation factor 1 alpha (EF-1α) primers *Pm*Ef-1α-F/ *Pm*Ef-1α-R that yielded a 122 bp-amplicon as a representative of chromosomal linear DNA. 2) a shrimp mitochondrial DNA (mtDNA) PCR detection method using primers *Pm*mtDNA-F/*Pm*mtDNA-R that yielded a 150 bp amplicon representing circular DNA. The presence of amplicons for 2 targets in the pre-digested DNA extract and absence of EF-1α amplicons but presence of the mtDNA amplicons in the post-digestion extract would confirm the success of circular DNA preparation. The ddPCR reaction with 309NF/309NR primers was used to quantify the IHHNV fragments in the pre-digestion DNA extract (presumed to contain non-circular IHHNV genome + IHHNV-cvcDNA) and post-digestion DNA extract (presumed to contain only IHHNV-cvcDNA).

2.5 Amplification and sequencing of cvcDNA-IHHNV in the circular DNA extract

The concentration of purified circular DNA extract from IHHNV-infected shrimp was quantified by Qubit fluorometer (Invitrogen, USA). In order to obtain a sufficient concentration for sequencing, the circular DNA (40 ng) was subjected to rolling circle amplification (RCA) using Repli-G midi kit (Qiagen). RCA is a technique used for multiple amplification of circular DNA to obtain sufficient amplicons to quantity for next-generation nucleotide sequencing (NGS) (Nelson et al, 2002). The RCA amplified products were verified by gel electrophoresis and capillary electrophoresis to determine DNA quality before sending the DNA for sequencing using the 150 bp pair-end sequencing method (Ilumina sequencing). The random fragmentation of the RCA product was by sonication. The resulting DNA fragments were end polished, A-tailed and ligated with the full-length adapters of

Illumina sequencing. This was followed by further PCR amplification with P5 and indexed P7 oligos. The PCR products for the final construction of libraries were purified using the AMPure XP system. Then libraries were checked for size distribution by Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA), and quantified by real-time PCR (to meet the criteria of 3 nM). For data analysis, the raw reads of nucleotide sequences were *de novo* assembled and compared to IHHNV reference genomes in databases. The 21-nt clean reads was mapped with two IHHNV reference genome sequences i.e. DQ228358 and AF218266 by bowtie2 and then average counts for all reads from 21-nt position along the virus reference sequences were plotted. The plot of mean reads data was constructed by ggplot in R program (https://www.r-project.org/).

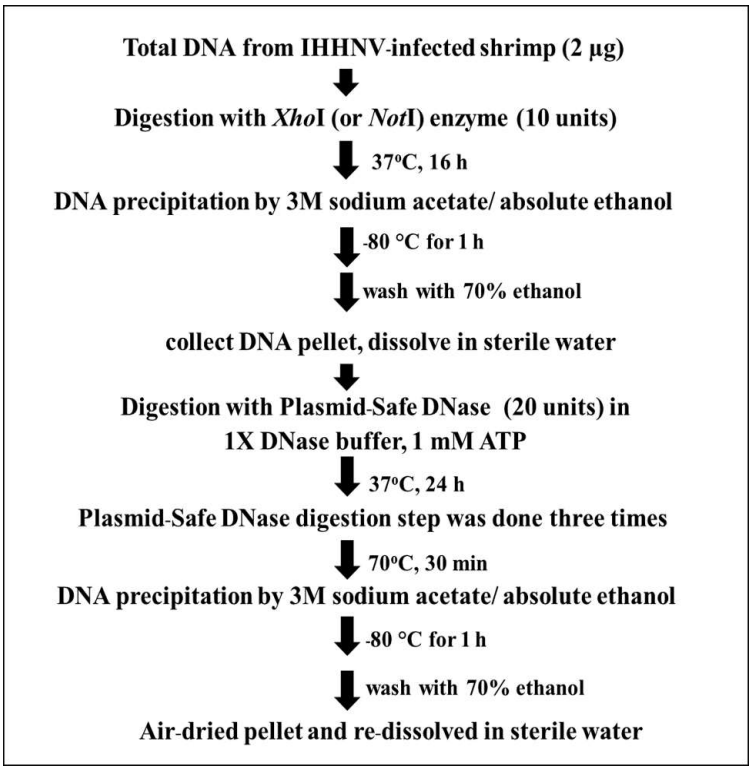

Figure 2: The protocol used to prepare circular DNA from the infected shrimp

2.6 Confirmation of the circular form by tail joining amplification

Primers were designed based on the cvcDNA sequencing result to prove that the annotated circular DNA sequences obtained were present in the original samples. Since the sequences obtained from NGS were reported as linear nucleotide chains, it was necessary to confirm the presence of circular forms in the original circular-DNA extract by using a forward primer designed specific for the tail (3') of the linear DNA fragment (IHHNV3031F or IHHNV3766F) together with a reverse primer designed from the 5' corresponding end of the same linear DNA fragment (IHHNV128R), i.e., back to back primers instead of forward and back primers (**Fig. 3**). Single step PCR was performed using a One-*Taq*™ PCR reaction kit (NEB, USA) with 30 amplification cycles and 2 ng of original circular-DNA extract as the DNA template. The general protocol for PCR was 94°C for 5 min followed by PCR for 30 cycles of 94°C for 15 s, 55°C for 15 s, 72°C for 30 s and then 72°C for 5 min. The PCR products were determined by 1.5% agarose gel electrophoresis and ethidium bromide staining. The amplicon band was cut and purified from the agarose gel before cloning into pGEM-T-vector. The plasmids

were transformed into *E. coli* cells. The plasmids were prepared from 8 clones (no. 1-8) were selected and subjected to digestion with *Eco*RI enzyme to check for the variation of inserted IHHNV fragments. Four of these products from clones no. 2, 4, 6, and 7 were sent for sequencing using T7/SP6 primer.

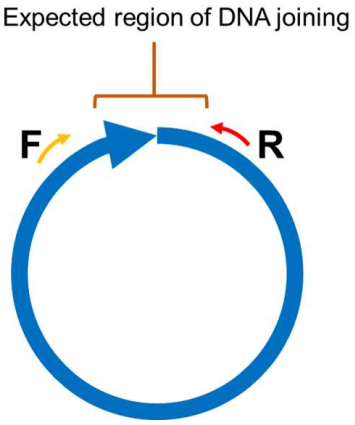

Figure 3. Diagram showing the method to determine circular DNA forms of IHHNV-cvcDNA by IHHNV-cvcDNA joining amplification. F and R indicate the designed primers regions base on assembled linear nucleotide sequences derived from the putative cvcDNA sequencing. Single step PCR amplification was carried out using purified cvcDNA as a template. The occurrence of positive PCR amplicons indicated closure of circular DNA including the nucleotide sequences for the end-joining region.

2.7 Inhibition of IHHNV replication by IHHNV-cvcDNA in shrimp challenge tests

A batch of juvenile *Penaeus vannamei* (2-3g body weight, n=50) were obtained from the shrimp demonstration farm in Chachoengsao province and maintained in the laboratory in a 500L tank containing artificial seawater (15 ppt salinity) for 2-days with continuous aeration and water temperature between 28-30°C. The shrimp were fed with commercial feed pellets at 5% body weight daily until starting experiments. Prior to experiments, a sub-sample of 3 arbitrarily selected shrimp was tested by PCR for the absence of IHHNV using the short-amp IHHNV detection method. This would later be compared with the negative test results expected from the negative control group shrimp at the end of the IHHNV challenge experiment. The shrimp were divided into 3 groups; the PBS injection group (negative control, n=5), the IHHNV injection group (positive control, n=5), and the test group injected with 100 ng circular DNA extract + crude IHHNV stock (n=10). The crude IHHNV stock was diluted with 1X cold PBS pH 7.4 to obtain 1 x 10^7 copies/50 µl to inject individual shrimp in the IHHNV-injected group positive control group and cvc-DNA test group. In the test group, the diluted virus was mixed with the circular DNA preparation containing putative IHHNV-cvcDNA before intramuscular injection into individual shrimp, while the negative control group was injected with 50 µl PBS only. At day 5 post injection, shrimp pleopods were collected from individual shrimp and then subjected to genomic DNA extraction followed by total DNA assessment by NanoDrop spectrophotometer (Thermo scientific, USA). Then, 100 ng was used as the template for PCR analysis for IHHNV replication level using the long-amp IHHNV method.

PCR intensities were determined using the Gel Doc™ EZ Gel Documentation System (Bio-Rad, USA). The relative ratio of virus level was calculated corresponding to the internal control gene expression (Ef-1α). Differences in IHHNV replication were determined by calculating the mean relative Ef-1α amplicon band intensity in the agarose gels followed by appropriate adjustment of the IHHNV band intensities before comparison of by One-Way ANOVA. Differences were considered significant at $p \leq 0.05$. Data analyses and graph preparations were carried out using using GraphPad Prism version 7.0 (https://www.graphpad.com/scientific-software/prism/).

3. Results

3.1 Extraction of circular DNA from the IHHNV-infected shrimp

After total DNA extracted from IHHNV-infected and non-infected shrimp (*P. monodon*) from the same source were exposed to PS-DNase digestion, the EF-1α amplicon linear DNA marker could no longer be detected, in contrast the untreated control. This indicated that all linear DNA had been digested. At the same time, the positive-control, circular-mtDNA could still be detected, indicating that circular DNA constructs could survive the exonuclease treatment (**Fig. 4)**. Starting from 2 µg of total shrimp DNA extract, there remained a putative circular DNA concentration in the range of 20-40 ng after digestion. Using digital droplet PCR to measure the quantity of IHHNV in the pre-digestion and post digestion preparations (**Fig. 5**) revealed that the average pre-digestion IHHNV quantity of 3.0 x 10^5 copies/ng DNA had dropped to 1.7 x 10^3 copies/ng DNA. This constituted a residual of approximately 0.6 % of the initial IHHNV-DNA quantity after linear DNA digestion. IHHNV could not be detected in the negative control digests obtained from IHHNV-negative shrimp. Taken together, the results from Figs. 4 and 5 suggested that the IHHNV copies detected in the DNase-digested preparation of the IHHNV-infected shrimp consisted of residual IHHNV in the form of cvcDNA.

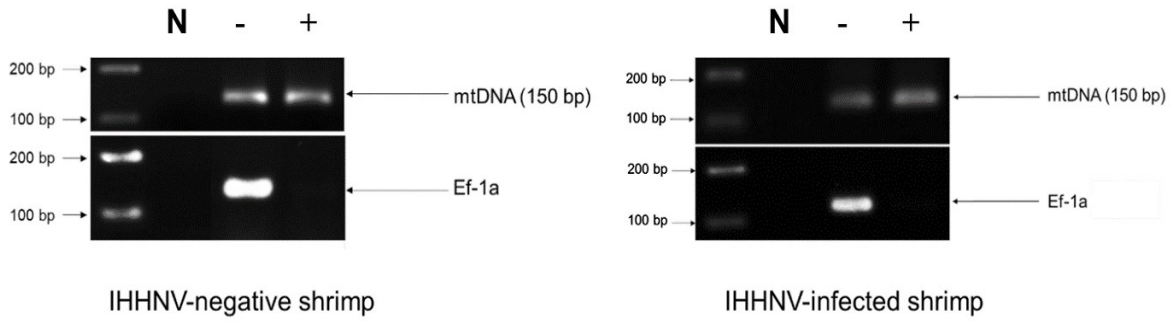

Figure 4. Photographs of agarose gels show amplicons from genomic DNA extracts of IHHNV-negative shrimp and IHHNV-infected samples before digestion and after digestion. - and + : without and with *Xho*I and PS-DNase digestion, respectively. PCR detection using mitochondrial DNA (mtDNA) primers and EF-1α primers as markers showed loss of an EF-1α amplicon after digestion but retained presence of the amplicon for circular mtDNA for both preparations.

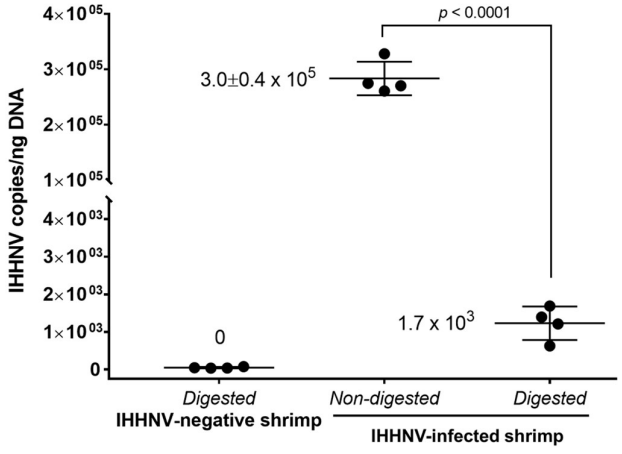

Figure 5. Detection of IHHNV by the digital droplet PCR method in non-digested and digested genomic DNA prepared from IHHNV-negative and IHHNV-infected shrimp.

3.2 Suppression of IHHNV replication by putative IHHNV-cvcDNA

To avoid the non-productive cost of NGS sequencing, it was necessary to test our putative IHHNV-cvcDNA preparation for protection against IHHNV in a shrimp challenge test. The details of IHHNV inoculum preparation from our *P. monodon* samples and testing of its infectivity in *P. vannamei* are given in **Supplementary Fig. S1**. On day 5 post injection in the protection test, shrimp from all 3 groups were collected and the DNA was extracted and checked by PCR using the long amp-IHHNV detection method with EF-1α as the internal control (**Supplementary Fig. S2**). The relative intensities of amplicon bands adjusted by the mean average of EF-1α intensities were compared (**Fig. 6**). The 5 shrimp in the negative control group injected with PBS gave no PCR amplicons for IHHNV, while the positive control group of 5 shrimp injected with IHHNV gave a mean band intensity of 1.2. In contrast, the group of 10 shrimp injected simultaneously with IHHNV and 100 ng of putative IHHNV-cvcDNA gave a mean amplicon intensity of 0.2 that was significantly lower intensity (p<0.01) than the positive control by a One-Way ANOVA test. We considered this sufficiently encouraging to proceed with NGS sequencing.

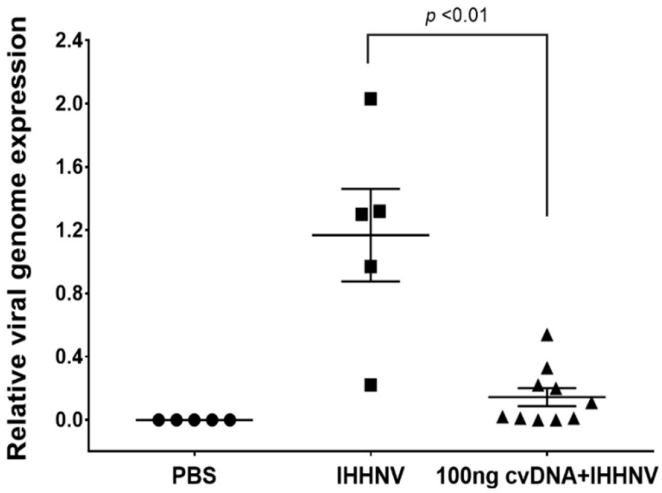

Figure 6. Graph of average band density for IHHNV- PCR amplicons among shrimp injected with IHHNV only or with IHHNV plus putative IHHNV- cvcDNA. PBS indicates IHHNV detection results for the naïve shrimp negative control.

3.3. RCA amplification of IHHNV-cvcDNA

The crude cvcDNA preparation (40 ng) was subjected to rolling circle amplification (RCA) using an REPLI-g mini kit (Qiagen) followed by quantification with a Qubit fluorometer revealing a total yield of 40 μg of RCA-amplified product. Subsequent agarose gel and capillary electrophoresis revealed that the majority of the RCA amplicons were of sizes larger than 11 kb (**Supplementary Fig. S3**). After sequencing of the RCA products, total reads of approximately 8 Mb nucleotides (8,363,236 bases) were obtained based on Illumina DNA sequencing with 99% effectiveness. A low sequencing error rate (0.03%) was found, as shown in **Table 2**. *De novo* assembly of the raw read sequences gave 81,026 contigs. A detailed analysis of the assembly is shown in **Table 3**.

Table 2. Sequencing information from the cvcDNA preparation obtained using the Illumina sequencing platform.

Sample	Raw reads (bases)	Raw data (Gb)	Effective (%)	Error (%)	Q20 (%)	GC (%)
Crude cvcDNA	8,363,236	1.3	99.73	0.03	96.79	47.1

Table 3. Statistics of the assembled contigs from NGS sequence analysis.

Assembly	Total number of assembly (contigs)
# contigs	81,026
# contigs (≤1,000 bp)	4,960
# contigs (1001- 5,000 bp)	307
# contigs (5001 to 10,000 bp)	132
GC (%)	42.18
Reference GC (%)	42.98
N50	553
L50	17,163

Comparison of the contigs in units of 21-nt bites against GenBank records for IHHNV revealed high homology to 2 accession numbers (**Fig. 7**). One was for accession no. AF218266 from an extant type of IHHNV. The other was for accession no. DQ228358, an endogenous viral element (EVE) of IHHNV formerly referred to as a non-infectious IHHNV (Tang & Lightner, 2006). These results suggested that the sequences homologous to AF218266 originated from the inoculated IHHNV virus extract and that they indicated the presence of IHHNV-cvcDNA in the putative cvcDNA extract. This was consistent with our prediction that IHHNV-cvcDNA would arise after IHHNV challenge. What was not predicted was the presence of putative cvcDNA originating from an IHHNV-EVE.

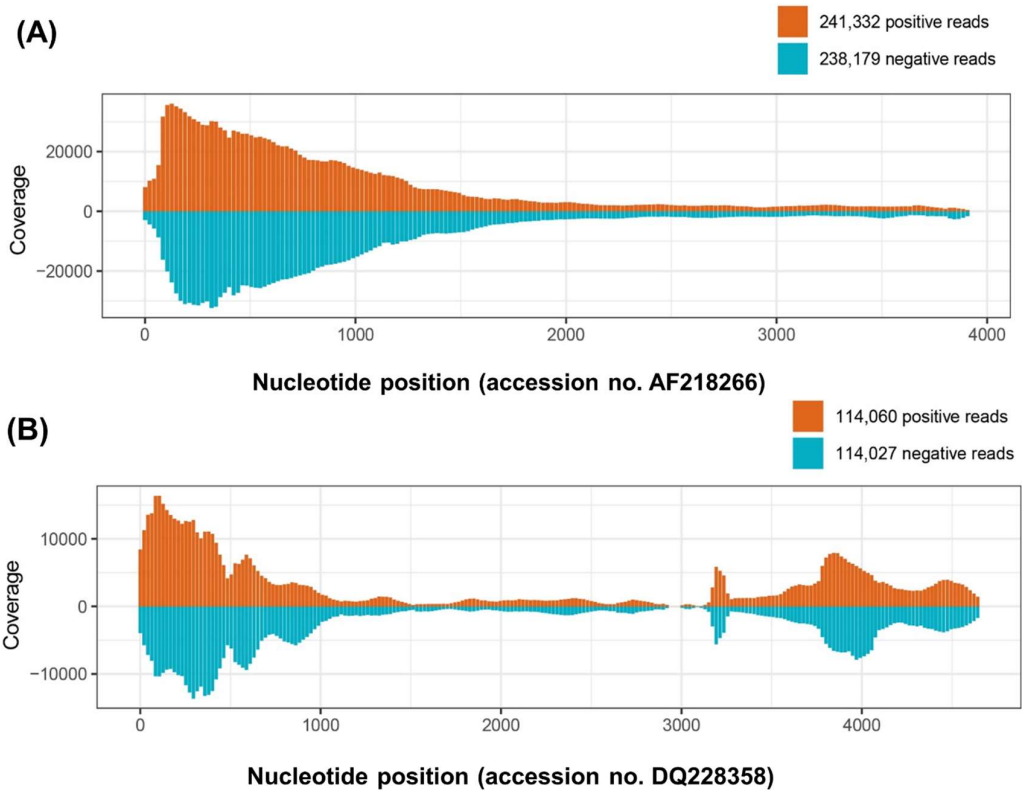

Figure 7. Diagrams of cvcDNA sequence reads distribution related to IHHNV reference sequences. The bar plots indicate distribution of the mean count of sequence reads obtained from DNA sequencing. The 21-nt sequence reads are shown as both plus and minus reads throughout the genome length. Distribution of 21-nt reads related to (A) GenBank accession no.AF218266 and (B) GenBank accession no. DQ228358. The shaded boxes above the graph represents the open reading frames of the related sequences corresponding to their nucleotide positions.

3.4 Determination of putative IHHNV-cvcDNA sequences

To predict IHHNV-cvcDNA size and sequence patterns, short read DNA assembly based on the reference genomes of IHHNV was carried out and the 4 longest contigs of interest are shown in **Table 4**. Three contigs of interest showed 98-99% similarity to IHHNV reference genome accession no. AF218266 (**Fig. 8A**) from an extant, infectious type of IHHNV. In contrast, the other longest contig showed 98% similarity to reference sequence accession no. DQ228358 which is known to be an inserted virus sequence (i.e., an endogenous viral element or EVE) (**Figure 8B**) that is sometimes called "non-infectious" IHHNV (Tang and Lightner, 2006). We chose to focus on the largest contigs that matched AF218266 (NODE_444 and NODE 1) and DQ228358 (NODE_439_3766) for further work to determine whether or not these contigs arose from cvcDNA.

Table 4. List of longest contigs with similarity to IHHNV genome references.

Sequence ID	Length (bp)	Backbone similarity
NODE_439_3766	3,766	DQ228358 (98%)
NODE_444_3736	3,736	AF218266 (99%)
NODE_1_3463	3,463	AF218266 (99%)
NODE_1_1548	1,548	AF218266 (99%)

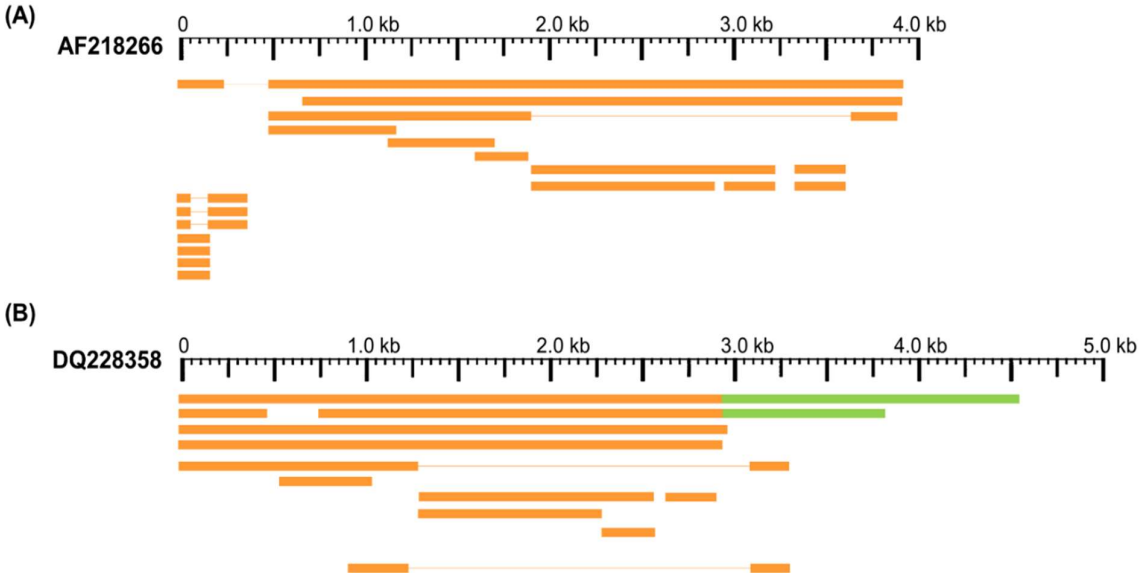

Figure 8. Schematic diagram showing the sequence similarity compared between putative cvcDNA contig sequences and Genbank records. The scale bar indicates nucleotide positions. Each box represents the sequence of an individual DNA contig. In the lowest cvcDNA in (A), there is a long deletion (indicated by a line) when compared to the GenBank record. (A) cvcDNA sequences with high similarity to the IHHNV virus GenBank record AF218266. (B) Sequences with high similarity to non-infectious IHHNV GenBank record DQ228358. The regions in green in B indicate the portion of the DQ228358 sequence that is part of a host shrimp transposable element and relates to the insertion point of the EVE in the shrimp genome.

3.5. PCR confirmed that contigs matching AF218266 arose from cvcDNA

To confirm that the linear contigs obtained by NGS sequencing from **Table 4** arose from cvcDNA, back-to-back primers were designed to match the ends of the linear contigs. Obtaining a product using these primers with the putative circular DNA mix as a template would fill in the missing regions between the ends and close the circles. Failure to obtain an amplicon would indicate that the target sequence was a linear DNA fragment. The primers, 3031F/128R were designed to amplify NODE_1_3463 and Node_444_3736 (**Fig 9A**) .and were tested with the putative circular DNA extract as template. The PCR result gave a single PCR amplicon of approximately 1,500 bp (**Fig. 9B**). This confirmed that the contigs ID NODE_1-3463 and/or NODE_444-3736 were derived from closed-circular DNA forms. The band of the 1,500 bp amplicon was purified and cloned. The plasmids from each clone were digested with *Eco*RI enzyme and the digestion results are shown in **Fig. 9C.** There were variations in the amplicon sizes among the 8 selected clones indicating a mixture of amplicons. Four selected clones were subjected for plasmid sequencing. Approximately 600-800 bp were read using forward and reverse primers and the assembled sequences (approximately 1,000-1,500 bp) matched sequences in the IHHNV reference genome AF218266. They corresponded to the expected amplified regions based on the GenBank reference genome. The ring in Fig. 9D represents a model cvcDNA of variable overall length with the orange portion indicating the contigs for NODE_1_3463 or NODE 444-3736 (or similar contigs that might contain the targets for primers IHHNV3031F/IHHNV128R at each end). In other words, the amplified sequences linked the two ends of each contig to form a circle and matched the expected corresponding sequences in the reference genome (**Fig. 9E**).

3.6 PCR confirmed that contigs matching DQ228358 arose from cvcDNA

Next we confirmed that contig no. NODE_439_3766 that matched the GenBank record DQ228358 also arose from cvcDNA. This contig must have arisen from an EVE in the genome of the experimental shrimp used, and it contained a portion of the host shrimp retrotransposon sequence. To close the cvcDNA circle, we followed the same protocol that was followed in the preceding section for ID NODE_444-3736 and NODE_1-3463.

Primers IHHNV3766F/IHHNV128R designed to match the ends of NODE_439_3766 were used (**Fig. 10A**). PCR results revealed an expected amplicon size of approximately 1,200 bp (**Fig. 10B**). After PCR cloning, 8 positive clones showed a variety of amplicon sizes indicating several cvcDNA types **(Fig. 10C)**, similar to the phenomenon previously observed for ID NODE_444-3736 and ID NODE_1-3463, described in section 3.7 above. DNA sequence reads of the 6 clones were of 2 general types, one showing sequence reads containing only IHHNV sequences (clone no. 1) and others (clones no.7,8) containing IHHNV sequences joined to shrimp retrotransposon sequences. Again, based on the publications linked to reference sequence accession no. DQ228358, these cvcDNA must have originated from the IHHNV-EVE called DQ228358. This was an unexpected finding providing evidence that cvcDNA can be generated from not only cognate viruses but also from EVE in the host genome itself.

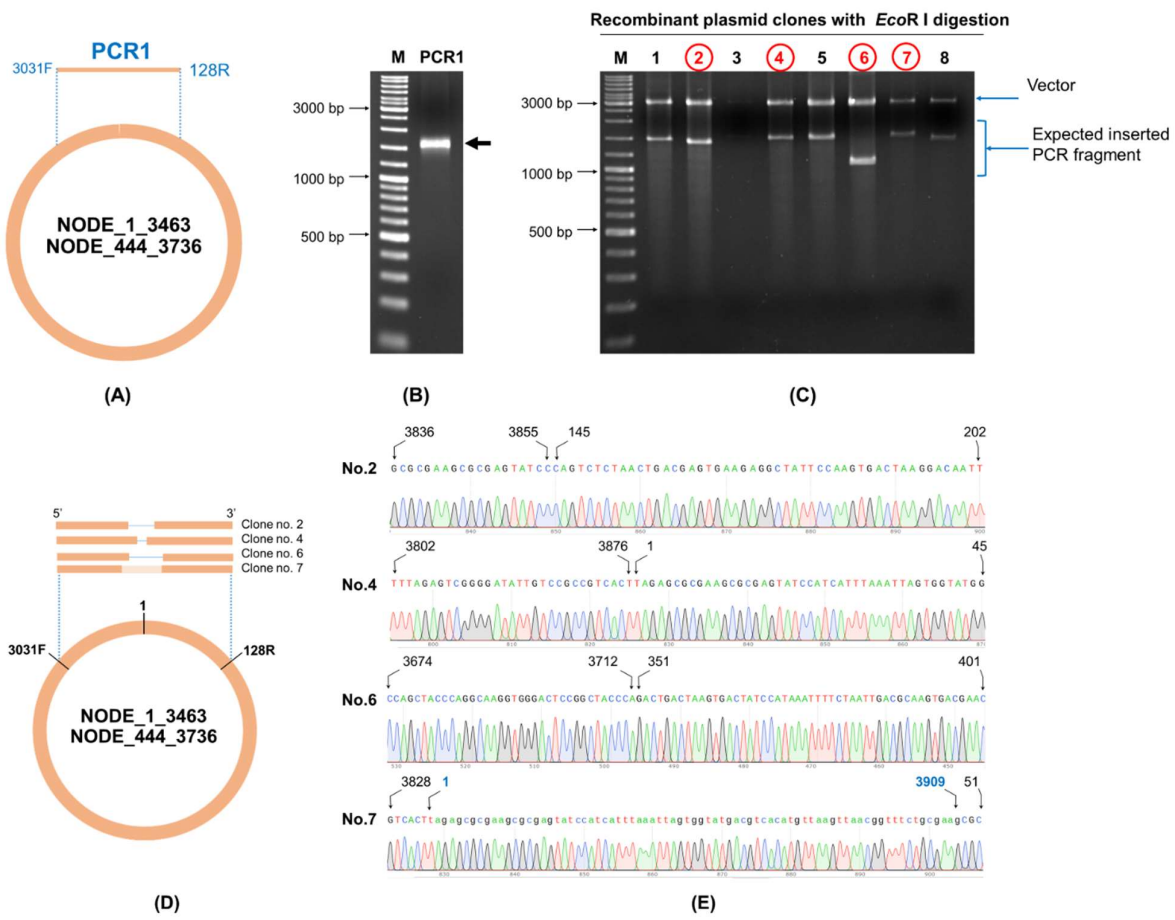

Figure 9. Confirmation of IHHNV-cvcDNA sequences by PCR and DNA sequencing. (A) A diagram using ID NODE_1- 3463 as a model with primer positions (IHHNV3032F/ IHHNV128R) and amplification directions indicated. (B) Photograph of an electrophoresis gel showing the amplified broad band of approximately 1,500 bp that was obtained from use of primers IHHNV3031F/ IHHNV128R. (C) *Eco*RI digestion of the plasmid preparations from 8 selected clones. The plasmid of clone no. 2, 4, 6, and 7 (in red) were subjected to sequencing using T7/ SP6 primers. (D) Diagram showing PCR sequence reads and alignments of the cvcDNA variants corresponding to the reference contigs no. NODE_1_3463 or NODE 444-3736. The ring represents a model cvcDNA of variable overall length with the orange portion indicating the contigs for NODE_1_3463 or NODE 444-3736. The orange boxes above the circle indicate variations in contiguous, amplified closure lengths that also share 99% identity to AF218266. The lines within some of the boxes simply indicate which missing part of the match to AF218266 has resulted in the smaller closure length. (E) Junction analysis based on PCR fragment sequencing in 9C showing a chromatogram for continuous DNA reads. None of the closed circles contained a match to a whole IHHNV genome sequence. Numbers indicate nucleotide position corresponding to reference genome accession no. AF218266.

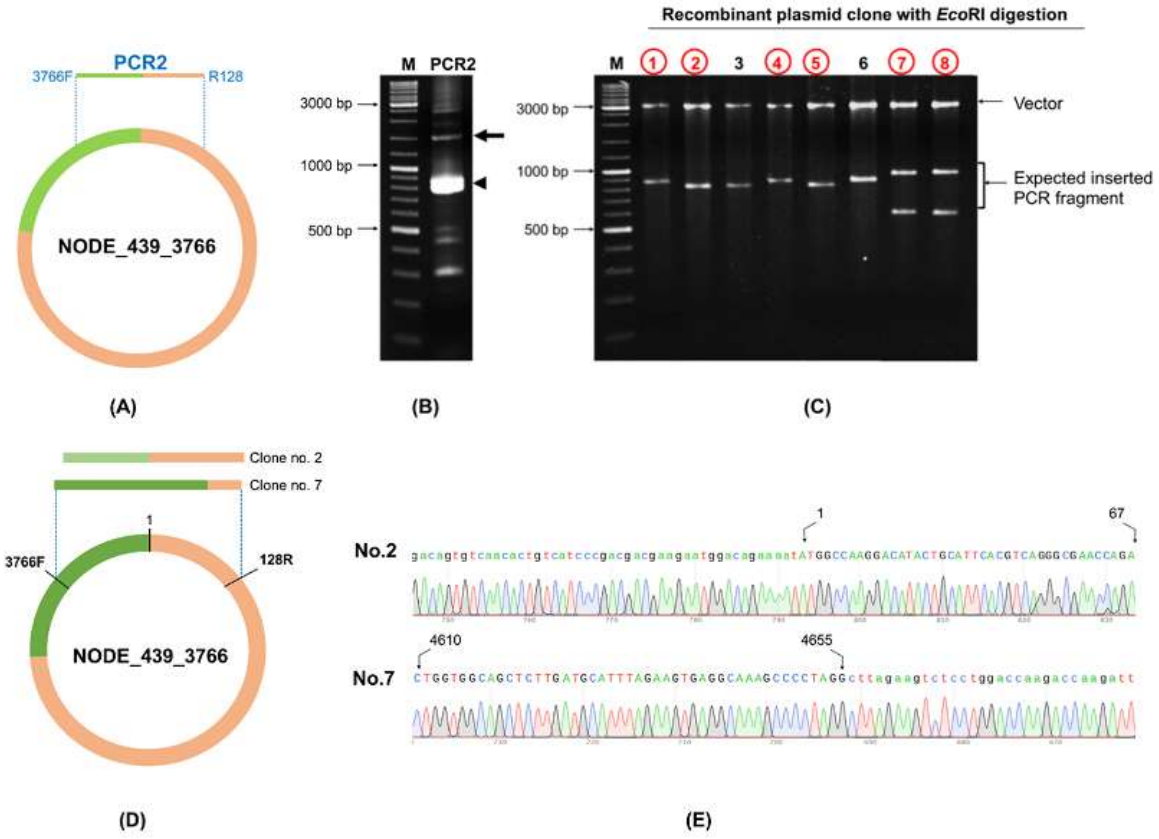

Figure 10. Confirmation of cvcDNA sequences of NODE_439_3766 by PCR and DNA sequencing homologous to GenBank record DQ228358. (A) Diagram showing the primer pairs designed for cvcDNA amplification. (B) The PCR amplicon was obtained from shrimp cvcDNA using single step PCR and confirming that the annotated cvcDNA sequences were closed circular forms. The expected PCR amplicon of approximately 1200 bp was observed (arrow). However, the majority of PCR amplicons were of smaller sizes (indicated by an arrowhead). (C) PCR fragment ligation and cloning to a plasmid vector revealed variation in the inserted PCR fragments. (D) Diagram showing the PCR sequence reads and alignments of the cvcDNA variants corresponding to the reference contigs no. NODE_439_3766. (E) Junction analysis of DNA sequencing obtained from PCR fragment sequences in 7C showing continuous DNA reads in the chromatogram linking the 3' and 5' ends of the reference sequence accession no. DQ228358. Extension of the shrimp retrotransposon was also observed (additional sequences indicated by lowercase letters).

4. DISCUSSION

According to our cvcDNA sequencing and PCR analysis, the cvcDNA forms produced in our IHHNV-infected shrimp could be classified into at least two major types (**Fig. 11**). The first type contained only variable-length fragments of infectious IHHNV genome sequences without containing any host nucleotide sequence. All had high identity to an infectious IHHNV sequence record. We do not currently know the process by which these cvcDNA constructs were generated. However, we hypothesize that they arose from the mRNA of infectious IHHNV as a result of endogenous (host) reverse transcriptase activity, as proposed by the viral accommodation hypothesis (Flegel, 2020) and as shown from research with RNA viruses in insects (Poirier, et al., 2018)(see Fig. 1).

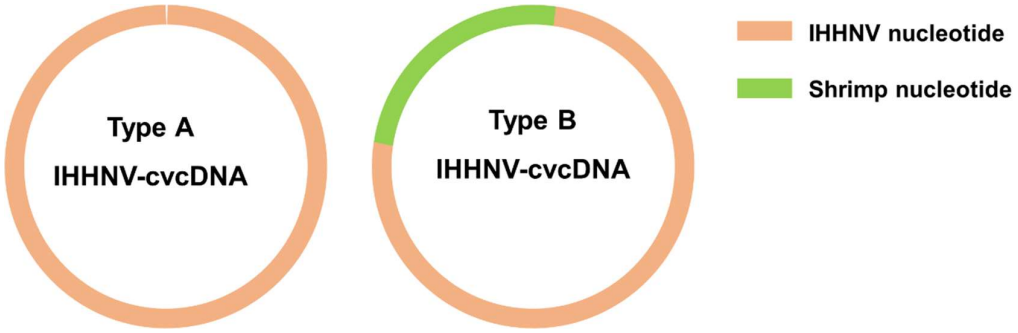

Figure 11. Diagram representing the two general types of cvcDNA produced in our IHHNV-infected *P. monodon.* Type A contained only DNA with sequences with 98-99% sequence identity to a currently extant species of IHHNV. Type B contained DNA with 98-99% sequence identity to an IHHNV-EVE and sometimes included a portion of the host genome.

The second type of cvcDNA we obtained contained sequences of the EVE represented by GenBank record DQ228358. This is known to be an integral part of the genome in some specimens of *P. monodon* from the Indo-Pacific region (Tang & Lightner, 2006). It was a surprising discovery, since it suggested that the cvcDNA may have arisen directly from the host shrimp genome via some kind of DNA polymerase (see Fig. 1). Alternatively, the EVE may have produced an RNA transcript that was subsequently used as a template by the host reverse transcriptase that normally produces cvcDNA entities in response to viral infection (see. Fig. 1). Given the information currently in hand, we know that EVE in shrimp (Flegel, 2020; Utari, et al., 2017) and in insects (Tassetto, et al., 2019) do produce RNA transcripts. We also know from the insect work that the RNA from infecting viruses can give rise to host generated cvcDNA via host RT activity. We hypothesize that EVE-DQ228358 can produce RNA transcripts that are processed in a manner similar to that used by insects, in which viral RNA serves as a template for production of lvcDNA and cvcDNA. An easy way to test this hypothesis would be to block host RT activity. The hypothesis predicts that doing so would prevent cvcDNA production from DQ228358. This pathway may be additional to the piRNA-like, cytoplasmic processing of long RNA transcripts into small RNA fragments leading to RNAi via interaction with specific PIWI binding proteins, as has been shown for insects {Tassetto, 2019 #9029;Whitfield, 2017 #8981}. Whatever the mechanisms for production of cvcDNA from EVE, the extraction of cvcDNA provides a very convenient method to identify EVE in normal, uninfected shrimp and screen them for possible antiviral protective activity.

It has been previously proven that shrimp and insects can accommodate both RNA and DNA viruses in tolerated, persistent infections and that they may also carry EVE for those tolerated viruses (Flegel, 2020). The occurrence of a protective EVE in insects has been recently proven for mosquitoes (Suzuki, et al., 2020). Now, we have proven that cvcDNA arises in shrimp from both an invading virus and from an EVE and that the extracted cvcDNA mix could interfere with IHHNV replication. These results support the hypothesis that there will be underling common mechanisms for viral accommodation (Flegel, 2020; Flegel, 2009) some of which have been described for RNA viruses in insects and now appear to occur also for a DNA virus in shrimp.

The discovery of cvcDNA arising from EVE was unexpected and is exciting, because it means that the process of isolation and characterization of cvcDNA appears to be a convenient way to screen for the presence of potentially protective EVE in shrimp and insects. Once the cvcDNA types have been characterized, primers can be designed to identify their presence in individual specimens of breeding stocks. They may also be produced *in vitro* and tested for protective capability by shrimp injection in laboratory trials. EVE mimics providing the most effective protection could be amplified *in vitro* by PCR and tested in shrimp and insects (e.g., silkworms and honeybees) as vaccines added to feeds. Such protective cvcDNA could also be specifically designed and injected into the ovaries of SPF broodstock shrimp for potential insertion in the shrimp genome. Then the offspring could be screened for possession of heritable, protective EVE. Such a model suggests that it might eventually allow for the improvement of current SPF shrimp stocks to increase their range of high tolerance to serious viral pathogens. The biggest advantage is that sourcing natural and protective EVE from shrimp and insects and using them for disease control should not elicit regulatory restrictions because the vaccines and reagents used would be from natural sources and because they are non-replicative and carry no antibiotic resistance genes.

We would like to proceed with this work by producing some of the IHHNV-cvcDNA entities we have discovered *in vitro* to test for their efficacy in controlling IHHNV infections in *P. vannamei* first by injection and then (if effective by injection) by addition to feed. One major question will be whether the cvcDNA with IHHNV sequences only or those also containing transposable element sequences will be more protective. If these trials prove to be successful, a practical vaccine might be produced for IHHNV and the model developed could also be used for other viruses such as white spot syndrome virus (WSSV).

6. Summary

A protocol for cvcDNA preparation was used and shown to be successful for extracting IHHNV-cvcDNA that matched the sequence of infective IHHNV in *P. monodon*. The extracted IHHNV-cvcDNA was shown to inhibit IHHNV replication when it was injected into *P. vannamei* challenged with IHHNV. Subsequent next generation sequencing (NGS) of the IHHNV-cvcDNA extract revealed a variety of IHHNV-cvcDNA types, one type that originated from the infecting IHHNV and another that originated from a host EVE. This unexpected discovery of cvcDNA arising from an EVE opens the way for relatively easy identification of natural and potentially protective EVE in shrimp via cvcDNA. This may lead to applications of EVE in shrimp and perhaps insects. The detailed mechanisms related to the production of cvcDNA from infecting viruses and from EVE in shrimp remain to be revealed, but its existence constitutes a new frontier for the discovery and potential application of cvcDNA for shrimp vaccination and for improvement of viral tolerance in shrimp breeding stocks.

We declare that our discovery of cvcDNA originating from EVE constitutes a revelation of a natural process that occurs in shrimp. As such, the process of using cvcDNA to detect and study EVE cannot be considered intellectual property eligible for patenting. Thus, anyone can use this knowledge freely to screen for protective EVE via the cvcDNA they may give rise to. It is possible that during this process some specific and highly protective, natural EVE may be discovered and used directly as vaccines or regents to genetically modify SPF shrimp, but again, such discoveries and applications

would not be patentable because of the natural occurrence of EVE and the cvcDNA they give rise to. It would be tantamount to trying to patent the shrimp themselves. On the other hand, it is possible, for example, that specific inventions of non-obvious vaccines and delivery methods may be suitable for patenting.

Acknowledgements

The authors would like to thank Haid Co., Ltd. and the Thai National Centre for Genetic Engineering and Biotechnology (BIOTEC) of the Thai National Science and Technology Development Agency (NSTDA) for the joint funding provided to carry out this work. We would also like to thank to Dr. Donghuo Jiang for encouraging innovation, for his helpful discussions and for his efforts that enabled us to explore this frontier area of shrimp immunology.

Ethic statement

Shrimp experiment and virus challenging tests were performed under the regulation of animal protocol hosted by national center for genetic engineering and biotechnology (BIOTEC), National for science and technology development agency (NSTDA), Thailand.

Conflict of interest

None.

7. References

Flegel, T., 2020. Research progress on viral accommodation 2009 to 2019. Dev. Comp. Immunol. 112, 103771.

Flegel, T.W., 2001. The shrimp response to viral pathogens. in: Browdy, C.L., Jory, D.E. (Eds.), The new wave. Proceedings of the special session on sustainable shrimp aquaculture, World Aquaculture 2001 , Orlando. World Aquaculture Society, Boca Raton, pp. 190-214.

Flegel, T.W., 2007. Update on viral accommodation, a model for host-viral interaction in shrimp and other arthropods. Dev. Comp. Immunol. 31, 217-231.

Flegel, T.W., 2009. Hypothesis for heritable, anti-viral immunity in crustaceans and insects. Biology Direct. 4, 32.

Flegel, T.W., Pasharawipas, T., 1998. Active viral accommodation: a new concept for crustacean response to viral pathogens. in: Flegel, T.W. (Ed.), Advances in Shrimp Biotechnology. National Center for Genetic Engineering and Biotechnology, Bangkok, pp. 245-250.

Goic, B., Vodovar, N., Mondotte, J.A., Monot, C., Frangeul, L., Blanc, H., Gausson, V., Vera-Otarola, J., Cristofari, G., Saleh, M.C., 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. Nat. Immunol. 14, 396-403.

Goic, B., Stapleford, K.A., Frangeul, L., Doucet, A.J., Gausson, V., Blanc, H., Schemmel-Jofre, N., Cristofari, G., Lambrechts, L., Vignuzzi, M., 2016. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nature Communications. 7, 12410.

Pénzes, J.J., Söderlund-Venermo, M., Canuti, M., Eis-Hübinger, A.M., Hughes, J., Cotmore, S.F., Harrach, B., 2020. Reorganizing the family *Parvoviridae*: a revised taxonomy independent of the canonical approach based on host association. Arch. Virol., 1-14.

Poirier, E.Z., Goic, B., Tomé-Poderti, L., Frangeul, L., Boussier, J., Gausson, V., Blanc, H., Vallet, T., Loyd, H., Levi, L.I., 2018. Dicer-2-dependent generation of viral DNA from defective genomes of RNA viruses modulates antiviral immunity in insects. Cell Host and Microbe. 23, 353-365.

686 Saksmerprome, V., Jitrakorn, S., Chayaburakul, K., Laiphrom, S., Boonsua, K., Flegel, T.W., 2011. Additional
687 random, single to multiple genome fragments of Penaeus stylirostris densovirus in the giant tiger shrimp
688 genome have implications for viral disease diagnosis. Virus Res. 160, 180-190.
689 Suzuki, Y., Baidaliuk, A., Miesen, P., Frangeul, L., Crist, A.B., Merkling, S.H., Fontaine, A., Lequime, S.,
690 Moltini-Conclois, I., Blanc, H., van Rij, R.P., Lambrechts, L., Saleh, M.-C., 2020. Non-retroviral
691 endogenous viral element limits cognate virus replication in *Aedes aegypti* ovaries. Curr. Biol. 30, 1–
692 12.
693 Taengchaiyaphum, S., Srisala, J., Bunphimpapha, P., Supungul, P., Tassanakajon, A., Chaiyapechara, S.,
694 Bowornpinyo, S., Sritunyalucksana, K., Flegel, T.W., 2019. Mendelian inheritance of endogenous viral
695 elements (EVE) of white spot syndrome virus (WSSV) in shrimp. Dev. Comp. Immunol. 96, 144-149.
696 Tang, K.F.J., Lightner, D.V., 2006. Infectious hypodermal and hematopoietic necrosis virus (IHHNV)-related
697 sequences in the genome of the black tiger prawn *Penaeus monodon* from Africa and Australia. Virus
698 Res. 118, 185-191.
699 Tang, K.F.J., Navarro, S.A., Lightner, D.V., 2007. PCR assay for discriminating between infectious hypodermal
700 and hematopoietic necrosis virus (IHHNV) and virus-related sequences in the genome of *Penaeus*
701 *monodon*. Dis. Aquat. Organ. 74, 165-170.
702 Tassetto, M., Kunitomi, M., Andino, R., 2017. Circulating immune cells mediate a systemic RNAi-based
703 adaptive antiviral response in *Drosophila*. Cell. 169, 314-325. e313.
704 Tassetto, M., Kunitomi, M., Whitfield, Z.J., Dolan, P.T., Sánchez-Vargas, I., Garcia-Knight, M., Ribiero, I.,
705 Chen, T., Olson, K.E., Andino, R., 2019. Control of RNA viruses in mosquito cells through the
706 acquisition of vDNA and endogenous viral elements. eLife. 8.
707 Utari, H.B., Soowannayan, C., Flegel, T.W., Whityachumnarnkul, B., Kruatrachue, M., 2017. Variable RNA
708 expression from recently acquired, endogenous viral elements (EVE) of white spot syndrome virus
709 (WSSV) in shrimp. Dev. Comp. Immunol. 76, 370-379.
710 Whitfield, Z.J., Dolan, P.T., Kunitomi, M., Tassetto, M., Seetin, M.G., Oh, S., Heiner, C., Paxinos, E., Andino,
711 R., 2017. The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes*
712 *aegypti* genome. Curr. Biol. 27, 3511-3519. e3517.

713 Wongsurawat T, Leelatanawit R, Thamniemdee N, Uawisetwathana U, Karoonuthaisiri N, Menasveta P,
714 Klinbunga S. Identification of testis-relevant genes using in silico analysis from testis ESTs and cDNA
715 microarray in the black tiger shrimp (*Penaeus monodon*). BMC Mol Biol. 2010 Aug 9;11:55. doi:
716 10.1186/1471-2199-11-55. PMID: 20696033; PMCID: PMC2928233.