

RESEARCH

Open Access



Shrimp genome sequence contains independent clusters of ancient and current Endogenous Viral Elements (EVE) of the parvovirus IHNV

Suparat Taengchaiyaphum¹, Prapatsorn Wongkhaluang¹, Kanchana Sittikankaew², Nitsara Karoonuthaisiri^{2,3,4}, Timothy W. Flegel^{5,6} and Kallaya Sritunyalucksana^{1*}

Abstract

Background: Shrimp have the ability to accommodate viruses in long term, persistent infections without signs of disease. Endogenous viral elements (EVE) play a role in this process probably via production of negative-sense Piwi-interacting RNA (piRNA)-like fragments. These bind with Piwi proteins to dampen viral replication via the RNA interference (RNAi) pathway. We searched a genome sequence (GenBank record JABERT000000000) of the giant tiger shrimp (*Penaeus monodon*) for the presence of EVE related to a shrimp parvovirus originally named infectious hypodermal and hematopoietic necrosis virus (IHNV).

Results: The shrimp genome sequence contained three piRNA-like gene clusters containing scrambled IHNV EVE. Two clusters were located distant from one another in pseudochromosome 35 (PC35). Both PC35 clusters contained multiple sequences with high homology (99%) to GenBank records DQ228358 and EU675312 that were both called “non-infectious IHNV Type A” (IHNV-A) when originally discovered. However, our results and those from a recent Australian *P. monodon* genome assembly indicate that the relevant GenBank records for IHNV-A are sequence-assembly artifacts derived from scrambled and fragmental IHNV-EVE. Although the EVE in the two PC35 clusters showed high homology only to IHNV-A, the clusters were separate and distinct with respect to the arrangement (i.e., order and reading direction) and proportional content of the IHNV-A GenBank records. We conjecture that these 2 clusters may constitute independent allele-like clusters on a pair of homologous chromosomes. The third EVE cluster was found in pseudochromosome 7 (PC7). It contained EVE with high homology (99%) only to GenBank record AF218266 with the potential to protect shrimp against current types of infectious IHNV. One disadvantage was that some EVE in PC7 can give false positive PCR test results for infectious IHNV.

Conclusions: Our results suggested the possibility of viral-type specificity in EVE clusters. Specificity is important because whole EVE clusters for one viral type would be transmitted to offspring as collective hereditary units. This

*Correspondence: kallaya@biotec.or.th

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 Road, Ratchathewi, Bangkok 10400, Thailand

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

would be advantageous if one or more of the EVE within the cluster were protective against the disease caused by the cognate virus. It would also facilitate gene editing for removal of non-protective EVE clusters or for transfer of protective EVE clusters to genetically improve existing shrimp breeding stocks that might lack them.

Keywords: Viral accommodation, Endogenous viral element(s) (EVE), Infectious hypodermal and hematopoietic necrosis virus (IHHNV), *Penaeus monodon* whole genome sequence

Background

Non-retroviral viral gene sequences inserted into host genomes during the course of evolution began to be called endogenous viral elements (EVE) around 2012 upon their discovery in vertebrate genomes [1]. However, evidence for EVE in animals was first reported for insects in 1999 [2], but the significance of the paper was not fully appreciated at the time of its publication. For shrimp, an EVE was described much later in 2007 [3] and was named genomic “non-infectious IHHNV” also before the term EVE was proposed.

In 2009, it was hypothesized that EVE (at the time called non-retroviral viral genome inserts) resulted from a mechanism that provided for heritable immunity in shrimp and insects [4, 5]. In brief, the mechanism was proposed to involve host recognition of invading viral mRNA followed by its use as a substrate for host reverse transcriptase (RT) to produce variable fragments of viral copy DNA (vcDNA). The vcDNA fragments would subsequently be inserted into the shrimp genome where they would produce antisense RNA that could induce an RNA interference (RNAi) response. This would lead to inhibition of viral replication and allow the host to accommodate one or more viruses (i.e., allow viral accommodation) in persistent infections without signs of disease. Tests for predictions of the viral accommodation hypothesis proceeded much faster with insects than shrimp. By 2020, the basic predictions regarding EVE were proven for mosquitoes [6–8]. The studies in both *Drosophila* and mosquitoes revealed additional detailed mechanisms for an unpredicted, specific adaptive response involving the ability of vcDNA to produce siRNA resulting in an immediate, adaptive cellular and systemic antiviral RNAi response [7, 9, 10]. For details see a recent review [5].

One of the insect publications [10] revealed that the vcDNA produced by host RT in response to RNA virus infection came in two forms, one linear (lvcDNA) and

one circular (cvcDNA). The latter could be specifically isolated relatively easily and was shown, when injected, to protect insect hosts against the homologous virus. Following the protocols described for cvcDNA isolation in insects, it was shown [11] that IHHNV-cvcDNA could be extracted from IHHNV-infected *P. monodon* and that it could inhibit IHHNV replication in whiteleg shrimp *P. vannamei* challenged with IHHNV prepared from the infected *P. monodon*. During sequencing of the cvcDNA extract, it was also found that some of the cvcDNA constructs had high homology (98–99%) to the GenBank records DQ228358 and EU675312 known to arise from the host *P. monodon* genome (i.e., from EVE) originally called “non-infectious IHHNV Type-A” (IHHNV-A) [3, 12]. Thus, the 2021 report [11] revealed that EVE can also give rise to cvcDNA, and it was hypothesized that the cvcDNA arose via EVE-produced RNA interacting with host.

During the time the study with IHHNV cvcDNA [11], a draft whole genome sequence (WGS) of a Thai *P. monodon* specimen was published [13], and we became interested to determine whether the WGS might contain EVE related to IHHNV. We considered this possible because the specimens we used for our cvcDNA work and the specimen used for the WGS project originated from the same domesticated shrimp breeding stock.

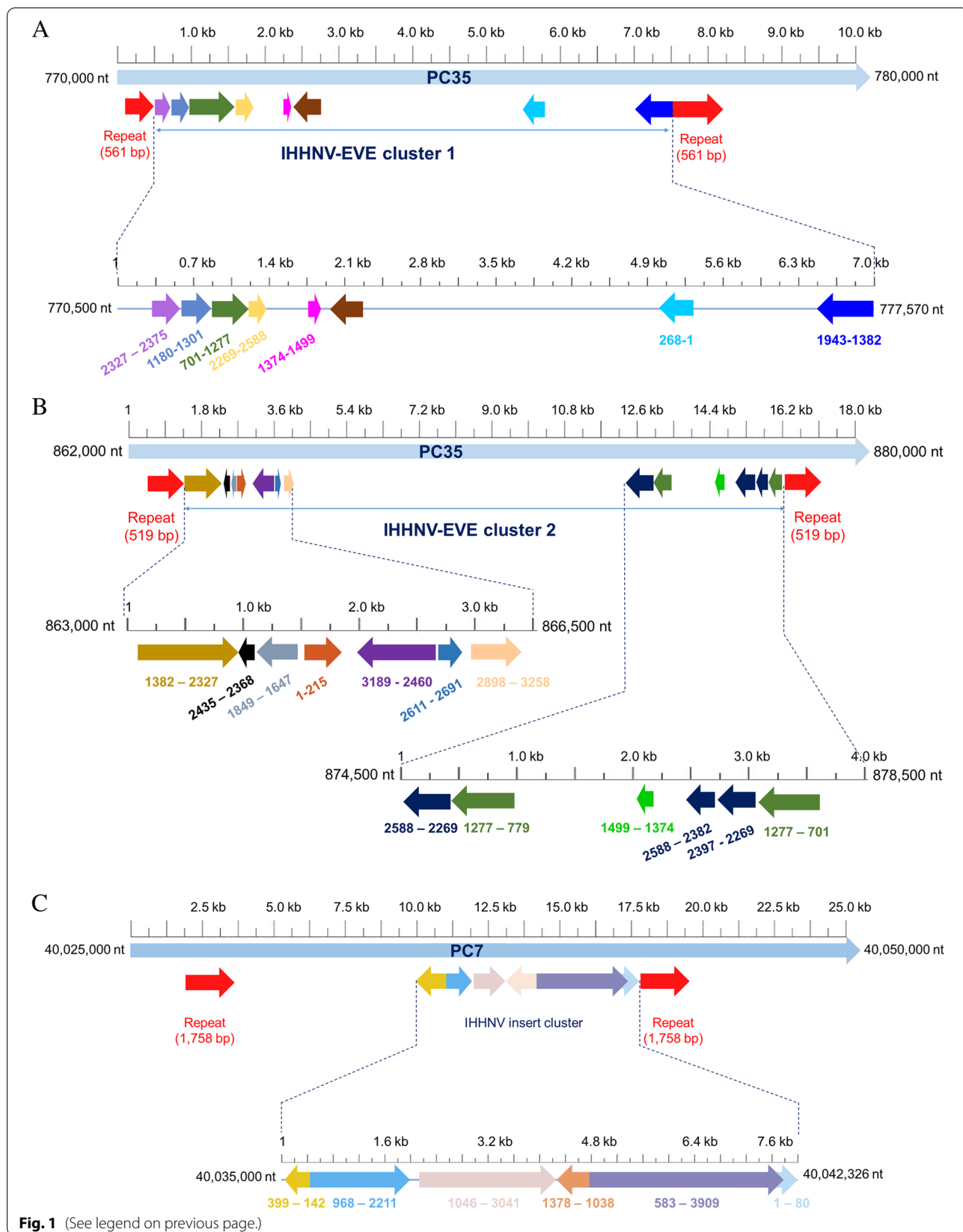
Results and discussion

Clusters of IHHNV-EVE were found in pseudochromosomes 35 and 7

A general BlastN search of the Thai *P. monodon* WGS project (GenBank accession no. JABERT000000000) [13] using the queries GenBank record DQ228358 for non-infectious IHHNV-A and GenBank record AF218266 for infectious IHHNV confirmed the presence of three clusters of EVE derived from IHHNV (Fig. 1). Two of these IHHNV clusters (Fig. 1A and B) were located

(See figure on next page.)

Fig. 1 Schematic diagrams of IHHNV-EVE clusters in the draft *Penaeus monodon* WGS. **A** and **B** Sequence diagrams for Cluster 1 and 2, respectively, in PC35 with high homology to the non-infectious IHHNV-A query portion (1–3025 bp) of the GenBank record DQ228358. Some of the EVE sequences in the 2 clusters correspond to the same region of DQ228358 (i.e., the same color) but may differ in length and in reading direction indicated by arrowheads. Others are unique to each cluster. A zoom-in expansion of each EVE cluster is shown beneath. Colored numbers below indicate the nucleotide positions corresponding to GenBank record DQ228358. The portion of the record related to the host transposable-element portion of DQ228358 is indicated by a dark brown arrow. **C** Diagram of the IHHNV-EVE cluster in PC7 (GenBank accession no. JABERT010000007.1) with EVE showing high sequence identity (99%) to GenBank record AF218266 for infectious IHHNV. The numbers below the arrows represent the matching positions AF218266



in pseudochromosome 35 (PC35) and showed high sequence homology (98–99%) to IHHNV in GenBank accession numbers DQ228358 and EU675312. The other cluster (Fig. 1C) was located in pseudochromosome 7 (PC7) and showed high sequence homology (98–99%) to GenBank accession number AF218266 for an extant type of infectious IHHNV. All three EVE clusters were clearly demarked by bracketing, direct host repeat sequences marked by red arrows in Fig. 1.

Characteristics of EVE homologous to non-infectious IHHNV-A in PC35

The GenBank records for DQ228358 (in *P. monodon* from East Africa) and EU675312 (in *P. monodon* from Australia) were initially referred to as “non-infectious IHHNV Type A” [3, 12]. The IHHNV-related portions of the records for DQ228358 and EU675312 share 99% sequence identity, indicating that they arose from a similar, ancient type(s) of IHHNV that is distantly related to the types of IHHNV currently reported to cause disease in cultivated shrimp [14]. We now refer to the two GenBank records DQ228358 and EU675312 as EVE of non-infectious IHHNV-A. Here we use the DQ228358 sequence for comparative analysis but focus primarily on the portion of the sequence (1 to 3025bp) that has homology to IHHNV. The remainder of the record (3026–4655 bp) constitutes shrimp host repeat and transposable element sequences.

We suspected that EVE of IHHNV-A might be present in the Thai *P. monodon* WGS data. This was because we had already reported the occurrence of a variety of cvcDNA sequences with high sequence identity (98–99%) to non-infectious IHHNV-A in *P. monodon* specimens obtained from a different generation of the same breeding stock from which the Thai genome specimen was obtained [11]. Those cvcDNA sequences covered 68% (779–3930 nt) of the matching portion of the DQ228358 record (including some transposable element portion) and could be assembled into a single linear construct. However, at the time of that publication, we did not know the actual composition of the genome region(s) from which the cvcDNA sequences arose. All we could conclude at the time was that a variety of IHHNV-cvcDNA types were produced, some linked with host transposable element sequences and some not.

Our search of the Thai WGS revealed that PC35 did not contain a single continuous sequence with high sequence identity to IHHNV-A. Instead, it contained many high-identity EVE fragments of variable length that were scrambled with respect to proportion, position and reading direction in the non-infectious IHHNV-A reference sequences. These scrambled EVE were arranged in

PC35 in two distinct clusters separated from one another by more than 80,000 bp (Fig. 1A and B).

The individual EVE are shown in different colors to allow easy visual comparison of the EVE to matching regions of the IHHNV portion of the DQ228358 reference sequence. The positions of the EVE in PC35 are shown below each cluster diagram in black typeface while the positions relative to the DQ228358 record are shown above in colors to match those of each EVE. The colored vertical lines indicate the boundaries of the individual EVE. The number and lengths of the individual EVE and the gaps between them in the two clusters in PC35 are shown in Table 1. The EVE in the two PC35 clusters contain some common DQ228358 coverage, but they are distinct from one another in terms of the EVE types, lengths and reading directions when compared to the DQ228358 reference sequence (Fig. 1A and B). The lengths of the EVE clusters bracketed by their respective repeat sequences were 6744 bp for Cluster 1 and 14,993 bp for Cluster 2.

The unexpanded diagrams for Clusters 1 and 2 in Fig. 1A and B shared some structural features. For example, both of the two EVE clusters (Fig. 1A and B) were bracketed by a pair of host, direct-repeat sequences (red arrows) of 2×561 bp in Cluster 1 (Fig. 1A) and 2×519 bp in Cluster 2 (Fig. 1B). In Cluster 1, the two repeats shared 98% sequence identity (547/561 bp) (Supplementary Fig. S1) while those in Cluster 2 shared only 77% sequence identity (400/519 bp) (Supplementary Fig. S2). These bracketing direct repeat sequences (red arrows) demarked the boundaries of the non-infectious IHHNV-A EVE clusters. They matched many other regions in several pseudochromosomes of the Thai WGS database but at relatively low sequence identities (around 80% or less).

Within the boundaries of Cluster 1 there was a host retrotransposon element (dark brown arrow) of 561 bp (Fig. 1A) that had 98% sequence identity to a matching portion of the 1395 bp host retrotransposon sequence that is part of GenBank record DQ228358. However, there was also a hit for 1354/1395 bases (98% identity) for the same retrotransposon portion of the DQ228358 sequence beginning very distantly from Cluster 1 at position 23,321,662 in PC35. All other spaces that separated the EVE fragments in Cluster 1 showed homology to various other shrimp host retrotransposon elements or host repeat sequences (not shown).

In contrast to Cluster 1, the pattern for Cluster 2 in PC35 (Fig. 1B) did not include any retrotransposon sequence with high homology to that in GenBank record DQ228358 within the bracketing 519 bp direct repeat sequences (red arrows). Like Cluster 1, all sequences separating the EVE showed homology to a variety of other shrimp host retrotransposon and repeat sequences (not

Table 1 Lengths of the EVE and intervening sequences in the 3 clusters found in the draft WGS of Thai *P. monodon*

EVE no.	Size (bp)	Identity	Gap (bp)	Labeled color (related to Fig. 1)	Nucleotide position on PC35	Nucleotide position on DQ228358/sequence direction
PC35 Cluster 1 related to DQ228358 (shown in Fig. 1A)						
1	49	100%	8	Purple	770,826–770,874	2327–2375/plus
2	122	99%	3	Navy blue	770,882–771,003	1180–1301/plus
3	577	98%	0	Dark green	771,000–771,576	701–1277/plus
4	320	99%	348	Light yellow	771,576–771,895	2269–2588/plus
5	126	98%	128	Pink	772,243–772,368	1374–1499/plus
6	562	95%	2923	Blue	777,570–777,009	1382–1943/minus
7	268	97%		Light blue	775,890–775,623	1–268/minus
Mean	237		568			
SD	176		1161			
PC35 Cluster 2 related to DQ228358 (shown in Fig. 1B)						
1	946	97%	2	Light brown	863,149–864,085	1382–2327/plus
2	68	93%	8	Black	864,145–864,083	2368–2435/minus
3	203	89%	234	Gray	864,344–864,153	1647–1849/minus
4	215	91%	14	Brown	864,578–864,781	1–215/plus
5	730	91%	189	Purple	864,970–865,679	2460–3189/minus
6	81	99%	194	Cyan	865,693–865,773	2611–2691/plus
7	361	88%	8125	Light orange	865,967–866,315	2898–3258/plus
8	321	97%	1169	Dark blue	874,530–874,848	2269–2588/minus
9	499	98%	0	Dark green	874,848–875,346	779–1277/minus
10	126	98%	521	Light green	876,515–876,643	1374–1499/minus
11	207	87%	15	Dark blue	877,164–877,379	2382–2588/minus
12	129	89%	43	Dark blue	877,422–877,551	2269–2397/minus
13	577	88%		Dark green	878,142–877,551	701–1277/minus
Mean	342		876			
SD	272		2308			
PC7 Cluster 3 related to AF218266 (shown in Fig. 1C)						
1	258	97%	0	Yellow	40,035,318–40,035,062	142–399/minus
2	1244	96%	101	Blue	40,035,318–40,036,561	968–2211/plus
3	1996	96%	2	Light brown	40,026,662–40,038,657	1046–3041/plus
4	341	96%	0	Brown	40,038,999–40,038,659	1038–1378/minus
5	3327	96%	0	Purple	40,038,998–40,042,323	583–3909/plus
6	80	95%		Light blue	40,042,247–40,042,326	1–80/plus
Mean	1207		21			
SD	1268		45			

shown). The bracketing sequences for EVE Clusters 1 and 2, matched many other regions in several pseudochromosomes of the Thai WGS database but at lower sequence identities of around 80% or less (not shown).

It is not known whether these two clusters in PC35 are linked on one chromosome of a diploid pair or whether they are matching “allele-like clusters” (“cluster-alleles”) located one each in the chromosome pair represented by PC35. Now that we have the sequences of the two clusters in PC35, it will be possible to design specific primers to screen for each type in the current generation of the

original *P. monodon* stock. By identifying one specimen positive for both clusters and mating it with one negative for both, obtaining about half the offspring each carrying only one or the other cluster would confirm whether or not they are cluster-alleles. Finally, the forms of these two clusters resemble those described from mosquitoes as piRNA-like gene clusters containing fragmented sequences from mosquito RNA viruses [6, 7].

A recent publication [15] using DNA extracted from an Australian *P. monodon* specimen for genome analysis revealed one EVE cluster in their Scaffold group 97

(SG97) with high identity to GenBank EU675312 (i.e., the Australian version of IHHNV-A with 99% sequence identity to DQ228358). Given the information from our earlier publication and from the work reported herein and from the Australian publication, it is likely that the continuous sequence records for DQ228358 and EU675312 at GenBank are assembly artifacts, and that they were obtained from fragmented and scrambled target sequences that had sufficient overlap to result in their assembly into single linear sequences when using the sequence of infectious IHHNV (GenBank record AF218266) as a reference [3, 12].

Huerlimann et al. [15] give a detailed analyses of the EVE cluster found on SG97 of their Australian *P. monodon* specimen plus a detailed comparison of it to the two EVE clusters that they (like us) discovered in PC35 of the Thai *P. monodon* WGS. Thus, those interested the detailed comparison should consult the Australian publication.

The IHHNV-EVE in PC7 showed high sequence homology to infectious IHHNV

Of greatest interest to us was the EVE cluster located in pseudochromosome 7 (PC7) (Fig. 1C) that showed high sequence identity (95–99%) only to GenBank accession number AF218266 from an extant form of infectious IHHNV [16]. In other words, this cluster contained no EVE with high sequence identity to GenBank records DQ228358 and EU675312 (i.e., non-infectious IHHNV Type A) that were found in EVE Clusters 1 and 2 in PC35. However, the general overall pattern for EVE Cluster 3 was similar to those of the two for IHHNV-A EVE clusters located on PC35. For example, the PC7 EVE were scrambled in terms of fragment location, portion and reading direction with respect to the AF218266 reference genome. They were also bracketed by two host direct-repeat sequences (red arrows) of 1758 bp (98% sequence identity, Supplementary Fig. S3). The EVE were either contiguous or separated by host retrotransposon or repeat sequences. Again, the overall arrangement resembled the Pi-RNA-like gene clusters reported for EVE of RNA viruses in mosquitoes [6, 7].

Our work with IHHNV-cvcdNA [11] and the work with mosquitoes [7, 8] suggest that negative sense RNA transcripts of the IHHNV-EVE in this cluster would have the potential capability of inducing a host RNAi response against infection by homologous types of infectious IHHNV. If our conjecture above that the 2 IHHNV-EVE in PC35 are alleles turns out to be correct, the fact that there is only one IHHNV-EVE cluster in PC7 suggests that it might be a single cluster/allele with no matching IHHNV-EVE on the homologous chromosome. If this is so, it would provide an opportunity to prove whether or

not the IHHNV-EVE in PC7 is protective against infectious IHHNV.

For example, primers could be designed to screen for the IHHNV-EVE cluster in PC7 in the current generation of the *P. monodon* stock. Several cluster-positive individuals could be mated with several cluster-negative individuals. If the cluster behaves in an allele-like manner, subsequent screening of the offspring from these crosses would reveal that at least one of the positive parents carried a single copy of the PC7 cluster/allele and would yield half the offspring positive for the allele and half negative for it. The offspring could then be challenged with infectious IHHNV followed by qPCR to determine infectious IHHNV loads in shrimp with and without the EVE cluster. A significantly lower mean IHHNV load in the cluster-positive offspring compared to the cluster-negative offspring would indicate that the cluster was protective. If so, the stock owner would then be able to use PCR to select crosses to maintain the protective cluster in subsequent stock generations.

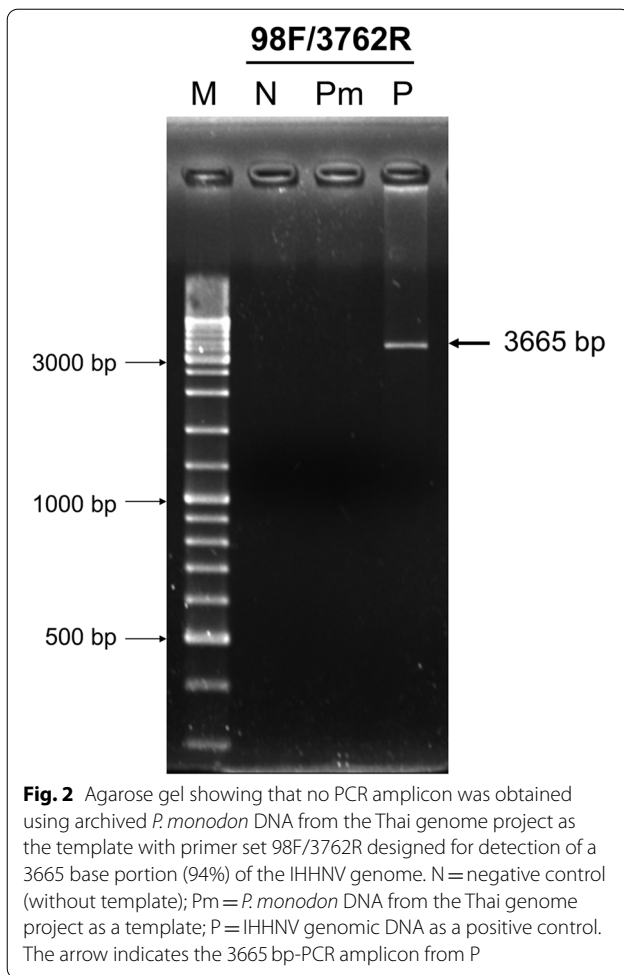
Detailed analysis of the IHHNV-EVE in PC7

Altogether, the IHHNV-EVE cluster in PC7 (Fig. 1C) is 7267 bp in length and contains six EVE having high sequence identity with a current type of infectious IHHNV (GenBank record AF218266). The EVE range from 80 to 3326 bp in length (Table 1) and they are scrambled with respect to portion, position and reading direction when compared to the IHHNV reference genome AF218266. In comparison to the two EVE clusters for non-infectious IHHNV in PC35, three of the EVE in Cluster 3 are much longer and the gaps between the EVE are shorter or do not exist. The significance of these differences is currently unknown. However, the general pattern of EVE scrambling, bracketing by host repeat sequences and separation or not by host transposable element or repeat sequences is similar to that in the other two EVE clusters in PC35.

To eliminate the possibility that the IHHNV-EVE cluster in PC7 with high identity to GenBank record AF218266 was the result of IHHNV genomic DNA contamination, we obtained an archived DNA sample from or the Thai shrimp genome project and used PCR primers 98F/3762R designed to amplify approximately 94% the whole IHHNV genome sequence [11]. This failed to give the expected amplicon, indicating the absence of contaminating, full genomic DNA of IHHNV in the Thai genome project DNA (Fig. 2).

Potential for false positive PCR detection of infectious IHHNV

A BlastN analysis of the IHHNV-EVE cluster in PC7 was carried out against the target sequences for the PCR



detection method (“309 method”) recommended for specific detection of infectious IHHNV in the OIE diagnostic manual [17]. The 309 method was designed to be specific for detection of infectious IHHNV [12]. Our analysis of the EVE cluster in PC7 revealed that 3 of the EVE (2, 3 and 5) contained sequences with 95% identity and 100% coverage for the target of the 309 method based on the AF218266 reference sequence (Table 2). However, despite that less than 100% overall sequence identities of the potential EVE targets, the primer sequences for the 309-method matched 100% with the relevant target sequences in all 3 of the potential EVE targets (Supplementary Fig. S4). Thus, the potential amplicons would be equal length (indistinguishable) from those that would arise from infectious IHHNV. Thus, the result for the specimen used for the genome project would constitute false positive test results for infectious IHHNV.

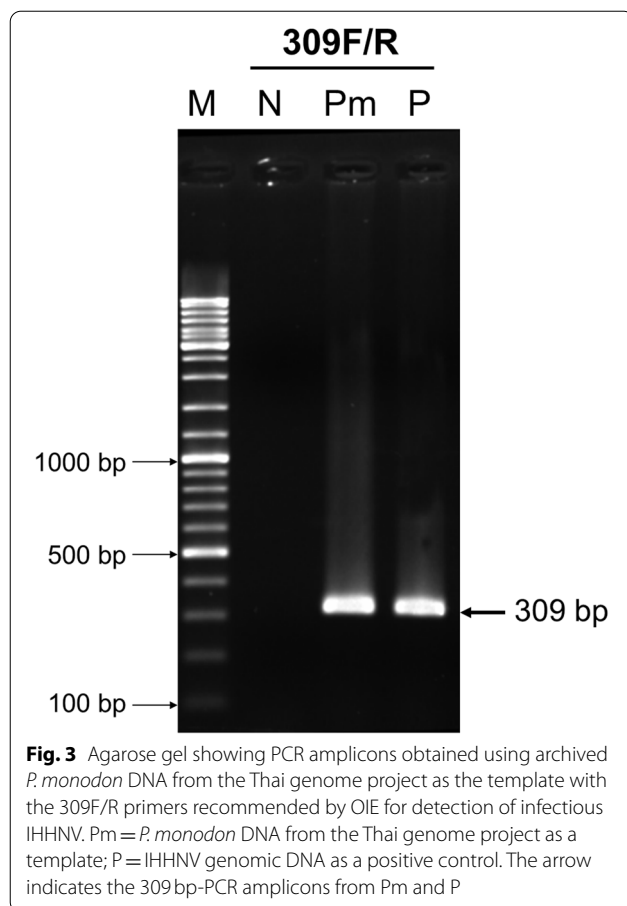
To test for this possibility of false positive test results, a sample of the same archived genomic DNA from the Thai *P. monodon* genome project (above) was subjected to PCR testing using the 309F/R primers recommended by OIE for detection of infectious IHHNV. This test gave an amplicon indistinguishable from that of the positive control containing DNA of infectious IHHNV and constituted a false positive test result for the presence of infectious IHHNV (Fig. 3).

Confirmation of the IHHNV-EVE cluster in PC7 by PCR and amplicon sequencing

We further confirmed that the IHHNV-EVE cluster in PC7 was real and not the result of assembly errors that somehow included contaminating fragments of the

Table 2 Potential targets in the IHHNV-EVE cluster of PC7. See the supplementary information (Supplementary Figure S1) for detailed information on the BlastN results

Detection Primer	Primer sequences	Hit positions on PC7	Target size/identity	Primer mismatch
OIE [17]	IHHNV309F/R	Location 1: 40,036,120 - 40,036,428 nt	309/95%	0/22
		Location 2: 40,037,386 - 40,037,694 nt	309/95%	0/22
		Location 3: 40,040,184 - 40,040,492 nt	309/95%	0/22
Jaroenram and Owens, 2014 [18]	RPA1F	Location 1: 40,041,559 - 40,041,588 nt	121/97%	1/30
	RPA1R	Location 1: 40,041,679 - 40,041,650 nt	121/97%	1/30
Cowley et al., 2018 [19]	IHHNV-q309F1	Location 1: 40,036,089 - 40,036,115 nt	98/100%	0/27
		Location 2: 40,037,355 - 40,037,381 nt	98/100%	0/27
		Location 3: 40,040,153 - 40,040,179 nt	98/100%	0/27
	IHHNV-q309R1	Location 1: 40,036,186 - 40,036,161 nt	98/100%	0/26
		Location 2: 40,037,452 - 40,037,427 nt	98/100%	0/26
		Location 3: 40,040,250 - 40,040,225 nt	98/100%	0/26
	IHHNV-q309Pr1 (probe)	Location 1: 40,036,119 - 40,036,136 nt	NA	0/18
		Location 2: 40,037,385 - 40,037,402 nt	NA	0/18
		Location 3: 40,040,183 - 40,040,200 nt	NA	0/18



IHHNV genome. The primers were designed to cover a region of the EVE cluster on PC7 that contained joined, unnatural, disjunctive fragments with high identity to extant IHHNV (GenBank record AF218266) (Table 1 and the associated diagram in Fig. 1). Both gave amplicons of the predicted size based on the WGS of the IHHNV-EVE cluster (Fig. 4). Furthermore, sequencing of the amplicon from R1 revealed 99% identity to the matching region of the whole shrimp genome record (Supplementary Fig. S5). From the result, the unnatural, disjunctive IHHNV sequence in the amplicon could not have arisen from contaminating IHHNV DNA and must have originated from its target sequence in the *P. monodon* DNA template.

Trade consequences from shrimp stocks carrying the PC7 IHHNV-EVE

In addition to the recommended method in the OIE manual, there are two published methods proposed to aid in the detection of infectious IHHNV while avoiding false positives arising from IHHNV-EVE. One of these [18] is an isothermal PCR detection method focused on a part of the infectious IHHNV genome that appeared to be of

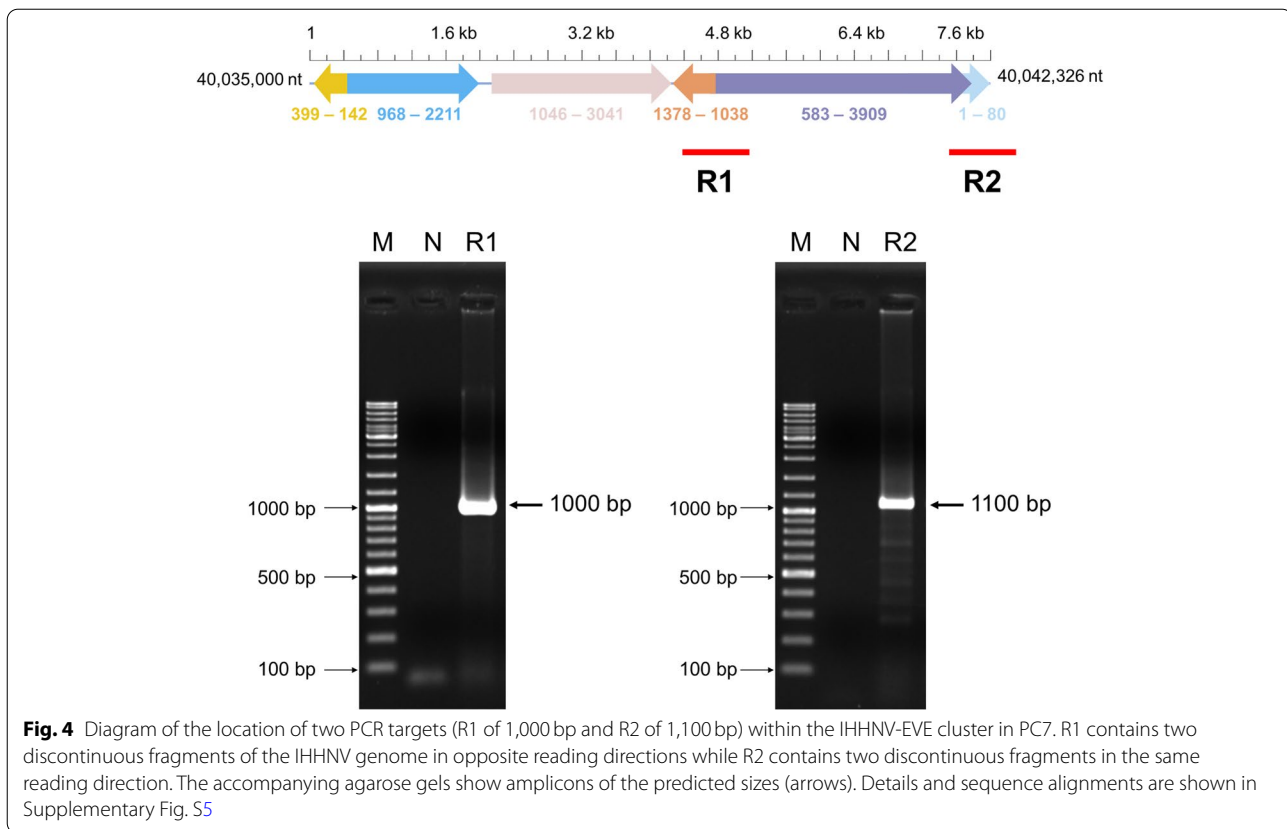
low prevalence or absent from IHHNV-EVE revealed by screening *P. monodon* from Australia. Our analysis of the target sequence for the primers from this method in the PC7 EVE for infections IHHNV (Table 2) revealed only one target sequence in the cluster in which there was a single base mismatch for each primer. There was a low calculated effect on the binding efficiency of the primers, so it is hard to predict whether or not the method would give a false positive test for infectious IHHNV with the potential PC7 target. In contrast, a more recent qPCR method designed to avoid false positive test results for infectious IHHNV that arise from IHHNV-EVE [19] would not be effective with the PC7 EVE cluster because there are three target sequences, each with 100% match to the primer and probe sequences.

In the past, without knowledge of EVE, shrimp specimens giving positive PCR test results with the methods above would have been considered IHHNV-infected and the positive shrimp specimens would have been discarded during the screening process to develop specific pathogen free (SPF) breeding stocks. In this way, it is possible that potentially protective EVE were discarded during stock development. This problem of false positive test results for infectious IHHNV arising from EVE has been raised previously [14, 20].

In addition, it is likely that shrimp products carrying the PC7 cluster would give false positive test results for the presence of infectious IHHNV, and this might result in the rejection of exported shrimp broodstock, shrimp PL and frozen shrimp products by countries that use the target of the OIE-recommended PCR method to screen imports for the presence of infectious IHHNV. It might be argued that use of a PCR method to detect such EVE followed by breeding selection to eliminate them from breeding stocks would be the simplest way to avoid this problem. However, such an approach might also remove naturally evolved, heritable resistance to IHHNV infection and lead to production problems for shrimp farmers.

In addition to the above potential for false positive results, another possibility has previously been raised [21]. This might occur in an SPF stock that had been developed by use of an internationally standardized PCR method to confirm absence of a particular virus based on a mutually agreed, small fragment of the target virus genome. This process would result not only in the discard of infected shrimp but also shrimp that carry EVE containing the PCR target sequence. Since some EVE may be protective and others not, it would be wise to choose the target region of any standard method to be in a genome sequence known to have no or low potential as a protective EVE.

However, this may not solve all the problems because individual shrimp in the stock population would carry



a variety of difference EVE cluster/alleles for a particular virus. Thus, it is possible that 2 individuals in a stock population might each carry one of 2 EVE in different cluster/alleles but in the same reading direction and have the potential to re-establish the PCR detection target by crossover. Specifically, the separated EVE fragments in the two shrimp would have to have sequence overlap but with the EVE in one individual containing the forward 5' primer target sequence for the chosen standard method but lacking the 3' primer sequence, while the other individual would contain the 3' primer sequence but lack the 5' sequence. These EVE would escape the screening process to remove stock individuals with EVE that carry the target for the standard PCR detection method chosen. However, at some unpredictable time in many crosses and generations, these to sequences might end up as cluster/alleles in a single individual by random assortment of chromosomes from its parents. If so, crossover events might occur in the overlap region to re-establish the PCR target sequence and give rise to a small portion of PCR positive individuals (i.e., "pop-up" positive individuals) in the offspring of such individuals. This could happen even in a population of shrimp with a good history

of freedom from the target virus [20]. For these reasons, it is essential that development of standard PCR detection methods involve a process of investigation and consultation among regulatory agencies and the companies or agencies that develop and maintain SPF shrimp breeding stocks for shrimp farmers.

The evolutionary advantage of jumbled contents of EVE clusters

We believe that the characteristic scrambling of EVE in piRNA-like clusters when compared to their arrangements in the originating genome in both insects and shrimp is worthy of some contemplation. We propose that this phenomenon may have evolved because it prevents the easy re-establishment of complete, infectious viral genome sequences in the host genome by the process of recombination between EVE-cluster alleles. On the other hand, the ability of the EVE to produce RNA transcripts and give rise to vcDNA [11] would seem to open the possibility that rare recombination events might occur between infecting viruses and EVE products (RNA or DNA) and be an additional potential source viral variation. As far as we know, this possibility has not been explored in shrimp or insects.

Conclusions

Revelation of piRNA-gene-like clusters of EVE from specific viral types is worth further investigation. This may exemplify an advantageous evolutionary development that arose because collections of EVE from specific viral types into single linkage clusters would assure their transmission to offspring as potential “protective antiviral EVE packages” (PEVEP). It would also be an evolutionary advantage to have PEVEP for different viruses located on different chromosomes to assure that maximum variation in PEVEP combinations would occur simply by random assortment of chromosomes during the production of gametes. If such PEVEP operated like alleles, additional variation would be possible via crossover during meiosis. Without viral genome fragmentation and scrambling (including reading direction) during the formation of these piRNA gene-like clusters, it might be possible that crossover between two suitable APP would occasionally re-establish a full, infectious viral genome or even give rise to new virulent types. Fracture and jumbling may have evolved to circumvent this possibility. However, it does not eliminate the possibility of pop-up, false-positive PCR amplicons that might arise from crossover between EVE that re-establish the sequence of a PCR target for a standard viral detection method. Nor does it eliminate the possibility of RNA and vDNA arising from EVE might on very rare occasions contribute to the evolution of viruses via recombination events.

Materials and methods

Bioinformatics analysis of a draft *P. monodon* genome for IHHNV

The sequences of non-infectious IHHNV (GenBank record DQ228358) and infectious IHHNV (GenBank record AF218266) were used as the subject reference sequences for BlastN analysis (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) of a recently released whole draft genome

(WDG) of the giant tiger shrimp *P. monodon* (GenBank accession no. JABERT000000000) [13]. Prediction of conserved DNA repeat sequences was determined using the Dfam database of repetitive DNA families (<https://www.dfam.org/home>) [21]. The MS PowerPoint program version 2019 was used to draw Fig. 1.

PCR detection and sequencing

The primer sequences used in this study were shown in Table 3. PCR and amplicons sequencing was used to confirm the IHHNV-EVE cluster in PC7. The PCR for IHHNV-EVE detection was performed in 12.5 µl consisting of 1X One-Taq PCR master mix (NEB, USA), 0.2 µl of each forward and reverse primer (PC7-R1F/PC7-R1R or PC7-R2F/PC7-R2R) and 100 ng of shrimp DNA extract. The PCR cycle was performed using PCR Thermocycler (Bio-Rad, USA) in total 35 cycles which consisted of initial denaturation at 94 °C for 5 min then followed by 35 cycles of 95 °C for 30 sec, 55 °C for 30 sec, 72 °C for 30 sec. Final extension at 72 °C for 5 min was applied. The PCR protocol for the 309 method was similar to those of the IHHNV-EVE detection. The long-amp PCR using 98F/3762R primers was performed according to previous described [11]. Briefly, the PCR reaction was performed in 12.5 µl by using Long-Amp™ Taq PCR mix (New England Biolab, USA). The PCR reaction consisted of Long-Amp Taq PCR reaction mix, 0.4 µM of forward and reverse primers (98F/3762R), 1 U Long-Amp™ Taq polymerase, and 100 ng of shrimp DNA extract. 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 2.5 min and final extension at 72 °C for 10 min. The PCR amplicons were determined by 1.5% agarose gel electrophoresis and ethidium bromide staining before subjected to extract from agarose gel by gel/PCR extraction kit (GeneAid Biotech, Taiwan). The gel purified PCR amplicons were submitted for sequencing by their specific primers (ATGC, Thailand).

Table 3 List of primers used in this study

Primer name	Sequence (5'-3')	Expected amplicon size	Reference
Long-amp PCR			
IHHNV-98F	CCCAGTTTCTAACTGACGAGTGAAGAGA	3665 bp	This study
IHHNV-3762R	CCTGACTCTAAATGACTGACTGACGATAGGG		
309 method			
IHHNV-309F	TCCAACACTTAGTCAAACCAA	309 bp	[3]
IHHNV-309R	TGTCTGCTACGATGATTATCCA		
IHHNV-EVE detection			
PC7-R1F	GGTGTGGAGTTTCTGTCT	1000 bp	This study
PC7-R1R	AATAGTAGCGGAACACAACCC		
PC7-R2F	ACCTACCCTCTACACATACCAG	1100 bp	
PC7-R2R	AGAGCGAGACTCAGACATAGA		

Abbreviations

IHHNV: Infectious hypodermal and haematopoietic necrosis virus; IHNNV-A: Non-infectious IHNNV Type A; cvcDNA: Circular viral copy DNA; RT-PCR: Reverse-transcriptase polymerase chain reaction; EVE: Endogenous viral element(s); piRNA: PIWI-interacting RNA(s); RNAi: RNA interference.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08802-3>.

Additional file 1: Supplementary Figure S1. Alignment of the 561 bp repetitive sequences flanking the non-infectious IHNNV-EVE Cluster 1 located in PC35. The sequences are aligned in the same direction and share 98% identity (547/561 bp with 9/561 nucleotide gaps). **Supplementary Figure S2.** Alignment of the two 519 bp repetitive sequences flanking IHNNV-EVE Cluster 2 located on PC35. The sequences are aligned in the same direction and share 77% identity (400/519 bp with 66/519 nucleotide gaps). **Supplementary Figure S3.** Alignment of the direct repetitive sequences flanking the infectious IHNNV-EVE Cluster 3 located on PC7. A total 1758 bp were aligned and share 98% identity (1722/1758 bp with 21/1758 nucleotide gaps). **Supplementary Figure S4.** Multiple sequence alignment of potential 309 bp targets for the OIE-recommended IHNNV detection method from three repeats in IHNNV-EVE observed in PC7 (R1-R3). The 309F/R primer positions are indicated by red rectangles. The primer sequences and potential amplicon sizes match 100% to those expected for infectious IHNNV. **Supplementary Figure S5.** Multiple nucleotide sequences alignment of PCR amplicons of R1 fragment in Fig. 4. Both forward and reverse sequence read directions were analyzed and the sequence similarity was 99% identity when aligned to original PC7 sequence portion. The nucleotide portion in rectangular represent anti-sense direction of IHNNV-EVE corresponding to the PC7 sequence assembly.

Acknowledgements

We thank the support from the National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Thailand.

Authors' contributions

Conceived and designed the experiments: ST, and TWF. Performed the experiments: PW, KanS, NK and ST. Analyzed the data: ST, KS and TWF. Wrote the paper: ST, KS and TWF. Revised the paper: ST, NK, KS and TWF. All authors read and approved the final manuscript.

Funding

This research project is supported by Mahidol University (Fundamental Fund: Basic Research Fund: fiscal year 2022) (Grant no. BRF1-054/2565).

Availability of data and materials

The datasets generated and/or analysed during the current study are available for whole genome sequence of *Penaeus monodon* in the GenBank accession number JABERT000000000, for the non-infectious IHNNV-A in the GenBank accession number DQ228358 and EU675312, and infectious IHNNV in the GenBank accession number AF218266.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Aquatic 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 Road, Ratchathewi, Bangkok 10400, Thailand. ²Biosensing and Bioprospecting Technology Research Group, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, 113 Thailand Science Park, Phahonyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand. ³International Joint Research Center on Food Security, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, 113 Thailand Science Park, Phahonyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand. ⁴Institute for Global Food Security, Queen's University, Belfast, Biological Sciences Building, 19 Chlorine Gardens, Belfast BT9 5DL, UK. ⁵National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), 113 Phahonyothin Road, Khlong Luang, Pathum Thani 12120, Thailand. ⁶Center for Excellence in Shrimp Molecular Biology and Biotechnology (Centex Shrimp), Faculty of Science, Mahidol University, Rama VI Road, Ratchathewi, Bangkok 10400, Thailand.

Received: 19 January 2022 Accepted: 25 July 2022

Published online: 06 August 2022

References

- Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet.* 2012;13(4):283–96.
- Lin CL, Lee JC, Chen SS, Wood HA, Li ML, Li CF, et al. Persistent Hz-1 virus infection in insect cells: evidence for insertion of viral DNA into host chromosomes and viral infection in a latent status. *J Virol.* 1999;73(1):128–39.
- Tang KF, Lightner DV. Infectious hypodermal and hematopoietic necrosis virus (IHHNV)-related sequences in the genome of the black tiger prawn *Penaeus monodon* from Africa and Australia. *Virus Res.* 2006;118(1–2):185–91.
- Flegel TW. Hypothesis for heritable, anti-viral immunity in crustaceans and insects. *Biol Direct.* 2009;4(1):32.
- Flegel TW. Research progress on viral accommodation 2009 to 2019. *Dev Comp Immunol.* 2020;112:103771.
- Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, Oh S, et al. The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Curr Biol.* 2017;27(22):3511–9 e3517.
- Tassetto M, Kunitomi M, Whitfield ZJ, Dolan PT, Sanchez-Vargas I, Garcia-Knight M, et al. Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. *Elife.* 2019;8:e41244.
- Suzuki Y, Baidaliuk A, Miesen P, Frangeul L, Crist AB, Merklings SH, et al. Non-retroviral endogenous viral element limits cognate virus replication in *Aedes aegypti* ovaries. *Curr Biol.* 2020;30(18):3495–3506 e3496.
- Saleh MC, Tassetto M, van Rij RP, Goic B, Gausson V, Berry B, et al. Antiviral immunity in *Drosophila* requires systemic RNA interference spread. *Nature.* 2009;458(7236):346–50.
- Poirier EZ, Goic B, Tome-Poderti L, Frangeul L, Boussier J, Gausson V, et al. Dicer-2-dependent generation of viral DNA from defective genomes of RNA viruses modulates antiviral immunity in insects. *Cell Host Microbe.* 2018;23(3):353–365 e358.
- Taengchaiyaphum S, Buathongkam P, Sukthaworn S, Wongkhaluang P, Sritunyalucksana K, Flegel TW. Shrimp parvovirus circular DNA fragments Arise from both endogenous viral elements and the infecting virus. *Front Immunol.* 2021;12(3979):729528.
- Tang KF, Navarro SA, Lightner DV. PCR assay for discriminating between infectious hypodermal and hematopoietic necrosis virus (IHHNV) and virus-related sequences in the genome of *Penaeus monodon*. *Dis Aquat Org.* 2007;74(2):165–70.
- Uengwetwanit T, Pootakham W, Nookaew I, Sonthirod C, Anghong P, Sittikankaew K, et al. A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon*) genome facilitates the identification of growth-associated genes. *Mol Ecol Resour.* 2021;21(5):1620–40.
- Saksmerprom V, Jitrakorn S, Chayaburakul K, Laiphrom S, Boonsua K, Flegel TW. Additional random, single to multiple genome fragments of *Penaeus stylirostris* densovirus in the giant tiger shrimp genome have implications for viral disease diagnosis. *Virus Res.* 2011;160(1–2):180–90.

15. Huerlimann R, Cowley JA, Wade NM, Wang Y, Kasinadhuni N, Chan C-KK, et al. Genome assembly of the Australian black tiger shrimp (*Penaeus monodon*) reveals a novel fragmented IHNV EVE sequence. G3 (Bethesda). 2022;12(4):jkac034.
16. Bonami JR, Trumper B, Mari J, Brehelin M, Lightner DV. Purification and characterization of the infectious hypodermal and haematopoietic necrosis virus of penaeid shrimps. J Gen Virol. 1990;71(Pt 11):2657–64.
17. OIE Manual of Diagnostic Tests for Aquatic Animals. 2017. Chapter 2.2.4. Infection with Infectious Hypodermal and Haematopoietic Necrosis Virus. <http://www.oie.int/international-standard-setting/aquatic-manual/access-online>.
18. Jaroenram W, Owens L. Recombinase polymerase amplification combined with a lateral flow dipstick for discriminating between infectious *Penaeus stylirostris* densovirus and virus-related sequences in shrimp genome. J Virol Methods. 2014;208:144–51.
19. Cowley JA, Rao M, Coman GJ. Real-time PCR tests to specifically detect IHNV lineages and an IHNV EVE integrated in the genome of *Penaeus monodon*. Dis Aquat Org. 2018;129:145–58.
20. Alday-Sanz V, Brock J, Flegel TW, McIntosh R, Bondad-Reantaso MG, Salazar M, et al. Facts, truths and myths about SPF shrimp in aquaculture. Rev Aquacult. 2020;12(1):76–84.
21. Hublely R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016;44(D1):D81–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

