Journal of Microbiology (2013) Vol. 51, No. 5, pp. 695–699
DOI 10.1007/s12275-013-3171-0

NOTE

Analysis of the Complete Nucleotide Sequence of a White Spot Syndrome Virus Isolated from Pacific White Shrimp[§]

Choong Yee Chai[1], Jangmi Yoon[1], Yong Seok Lee[2], Young Bong Kim[3], and Tae-Jin Choi[1]*

[1]*Department of Microbiology, Pukyong National University, Busan 608-737, Republic of Korea*
[2]*Department of Life Science and Biotechnology, Soonchunhyang University, Asan 336-745, Republic of Korea*
[3]*Department Animal Biotechnology, Konkuk University, Seoul 143-701, Republic of Korea*

(Received March 26, 2013 / Accepted April 26, 2013)

The fourth complete genome sequence of the white spot syndrome virus (WSSV) from Pacific white shrimp (*Litopenaeus vannamei*) isolated from Korea (WSSV-KR) was determined. The genome is composed of 295,884 bp encompassing 515 open reading frames (ORFs) among which 90 showed no sequence homology with any known protein in BLAST searches. The remaining 425 ORFs encode functional proteins including enzymes for nucleic acid metabolism, DNA replication and transcription, and several major structural proteins. Dot plot and linear comparisons of WSSV Korean strain with other WSSV isolates showed overall similarity but with some areas of sequence difference and one large deletion area.

***Keywords*:** WSSV, genome, Pacific white shrimp

WSSV, a double-stranded DNA virus, is the causative agent of white spot syndrome (WSS) and can cause severe mortality in penaeid shrimp (Lightner, 1999; Lo *et al*., 1999; Huang *et al*., 2001). Mortalities of up to 100% can be reached within 3–10 days after infection in shrimp populations (Marks *et al*., 2004; Pradeep *et al*., 2008). WSSV infects many species of cultured and wild, marine and freshwater penaeid shrimps and a broad range of other decapod crustaceans, including crabs and crayfish (Lightner, 1999; Esparza-Leala *et al*., 2009; Dieu *et al*., 2010). WSSV outbreaks in farmed prawns are often characterized by high and rapid mortality (Crockford, 2008). Acutely affected prawns often show changes in behavior and pink-to-red discoloration of the body and appendages, with numerous white spots, 0.5 to 2.0 mm in diameter, on the inside of carapace cephalothorax with consequent attraction of shrimp-eating birds (Park *et al*., 1998; Crockford, 2008; Lightner and Lo, 2008; Pazir *et al*., 2012).

According to Jang *et al*. (2008), *Fenneropenaeus chinensis*, a native shrimp species in Korea, has been largely replaced with *L. vannamei* since 2002 because the latter is less susceptible to WSSV (Tan *et al*., 2009). It was later found that *L. vannamei* can also be infected by WSSV and WSSV has recently caused mass mortality in this new species in Korea. However, there is presently no reported genetic information on the WSSV that affects *L. vannamei*. Thus, we investigated the nucleotide sequence of the entire genome of a WSSV isolate from Pacific white shrimp in Korea, named WSSV-KR, and compared the variation in WSSV-KR to WSSV isolates from other geographical locations.

Currently, three complete genomes of WSSV isolates are known: WSSV isolates from China (Yang *et al*., 2001), Thailand (Van Hulten *et al*., 2001), and Taiwan (Tsai *et al.,* 2000). WSSV-KR from this study is the fourth complete genome sequence of a WSSV isolate. With these known sequences, we tried sequencing the WSSV-KR genome using next-generation sequencing (NGS) techniques using total DNA extracted from an infected shrimp, avoiding time-consuming virus-purification procedures.

WSSV-infected Pacific white shrimp (*L. vannamei*) were collected from a shrimp pond in Sinan-Gun, Jeollanam-Do, Korea. WSSV genomic DNA was extracted as a mixture with host DNA from the shrimp by proteinase K and cetyltrimethylammonium bromide (CTAB) treatments, followed by phenol-chloroform extraction and ethanol precipitation. The extracted total DNA was sent to NICEM, Seoul National University, for analysis of the whole genome by NGS. Comparisons of WSSV-KR with WSSV isolates from China (WSSV-CN; accession no. AF332093), Thailand (WSSV-TH; accession no. AF369029), and Taiwan (WSSV-TW; accession no. AF440570) were performed for the annotation of identified open reading frames (ORFs). The formats of ORFs were wssv_00010 (WSSV-KR), wsv001 (WSSV-CN), ORF1 (WSSV-TH), and WSSV001 (WSSV-TW).

NGS allows whole genome sequencing of WSSV without virus purification. The total number of reads was 31,302,524 and the total length was 3,161,554,924 bp. The WSSV-KR sequences were separated from the host genome sequence using the genomic sequences of the three known WSSV strains in 120,563 contigs, 55,686,319 bp in length. Then it

*For correspondence. E-mail: choitj@pknu.ac.kr; Tel.: +82-51-629-5617; Fax: +82-51-629-5619
[§]Supplemental material for this article may be found at http://www.springerlink.com/content/120956.

Table 1. General information on WSSS isolates and their genomes

Isolates	WSSV-KR	WSSV-CN	WSSV-TH	WSSV-TW
Originating host	*Litopenaeus vannamei*	*Penaeus japonicus*	*Penaeus monodon*	*Penaeus monodon*
Genome size (base pairs)	295,884	305,107	292,967	307,287
GenBank accession no.	JX515788	AF332093	AF369029	AF440570
Date of sampling	Aug 2011	Oct 1996	May 1996	Nov 1994
Sampling location	Jeollanam-Do, Korea	Xiamen, China	Suratthan, Thailand	Southern Taiwan

was assembled as a full-length sequence of 295,884 bp. The complete DNA sequence of WSSV-KR was deposited in GenBank under the Accession Number JX515788 (Table 1).

WSSV-KR has a double-stranded circular genome and a total G+C content of 41%, containing 89,416 A (30%), 60,452 C (20%), 60,702 G (21%), and 85,314 T (29%). It encodes 515 ORFs, and the ORFs are present on both strands in almost equal proportions: 271 are forwardly transcribed (53%) and 244 are reversely transcribed (47%). Figure 1 shows a linear map of the WSSV-KR genome.

Homology searches were performed with the BLAST program and 90 of the ORFs showed no homology to any known protein. The remaining 425 ORFs likely encode functional viral proteins (Supplementary data Table S1). A very long ORF of 18,222 nucleotides (wssv_03600) with unknown function, also present in other strains, was identified in WSSV-KR. WSSV-KR showed high homology (97–99%) with the previously described isolates WSSV-CN, WSSV-TH, and WSSV-TW.

Variation between WSSV-KR and other isolates was assessed by dot plot and linear comparison. The one diagonal line in the self comparison shown in Fig. 2A confirms that there was no duplication of specific genes or a genome fragment in WSSV-KR. WSSV-KR and WSSV-CN also showed a linear relationship with no transposition (Fig. 2B). In the comparisons with WSSV-TH and WSSV-TW, a similar linear pattern was observed with two diagonal lines (Figs. 2C and 2D) due to the circular characteristics of the WSSV genome. The first nucleotide of WSSV-KR in the GenBank (JX515788) corresponds to nucleotides 33,589 and 48,957 of WSSV-TW and WSSV-TH, respectively.

Chen *et al.* (2002b) showed that differences between three WSSV isolates (WSSV-CN, WSSV-TH, and WSSV-TW) were mostly due to several small insertions and one large (~12 kb) deletion. The same issues were observed when WSSV-KR was compared in detail with other WSSV isolates. WSSV-KR (295,884 bp) is shorter, by 9,223 bp, than WSSV-CN (305,107 bp) and a large area of deletion in WSSV-KR could be observed in the area indicated as region 7 in Fig. 3. In a pairwise comparison with WSSV-CN (305 kb), WSSV-KR (296 kb) showed a large deletion of 9.2 kb (Fig. 3A). Also, WSSV-KR showed a large deletion of 3.8 kb in the comparison with WSSV-TH (293 kb; Fig. 3B). On the other hand, WSSV-TW (307 kb) has one deletion region of 2.5 kb compared to WSSV-KR (Fig. 3C).

In region 1, WSSV-CN has ORFs 127, 128, and 129 of 287, 1,109, and 1,070 nt, respectively. In the case of WSSV-TH, one ORF of 1,322 nt (ORF75) is located in this region. However, WSSV-KR has one ORF of 848 nt, showing fusion of wsv127 and wsv129 of WSSV-CN (Supplementary

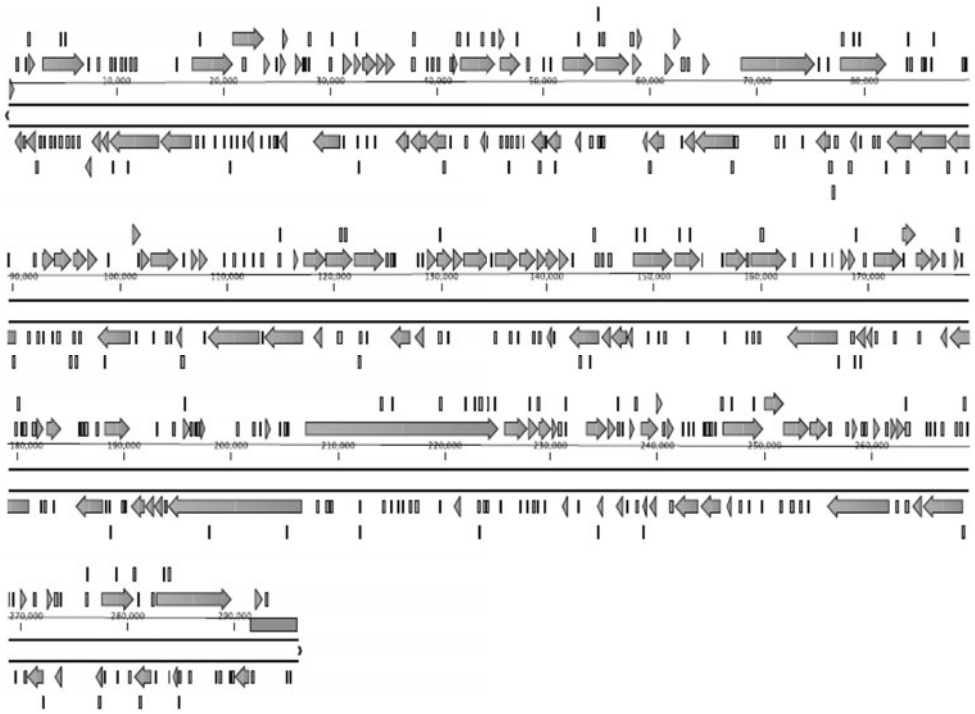

Fig. 1. Linear map of the WSSV-KR genome. ORFs transcribed in the forward direction are indicated above the genome and ORFs transcribed in the reverse orientation are indicated below.

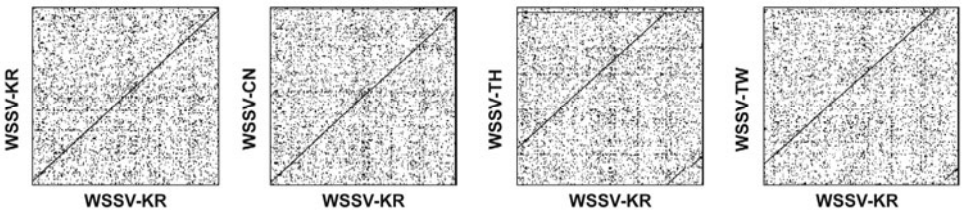

Fig. 2. Dot plot comparison of WSSV- KR with other WSSV isolates.

data Fig. S1). In region 2, ORF wsv150 of 914 nt which presents in the reverse orientation in WSSV-CN is missing in WSSV-KR. Instead, WSSV-KR has two short ORFs, wssv_01500 (347 nt) and wssv_01510 (413 nt), which showed the highest sequence similarity to ORF87 of WSSV-TH (Supplementary data Fig. S2). In region 3, wsv178 of WSSV-CN (909 nt long), WSSV234 of WSSV-TW and ORF94 of WSSV-TH (both 585 nt long) contain a repeated sequence of a "AAPPPEDEEEDDFYRKKR" motif, 12 times in WSSV-CN and 6 times in both WSSV-TH and WSSV-TW. In contrast, wssv_01790 of WSSV-KR is only 315 nt long and the "AAPPPEDEEEDDFYRKKR" motif is present only once. (Supplementary data Fig. S3). In region 4, two ORFs, wsv245 (230 nt) and wsv246 (230 nt), present in WSSV-CN, with no sequence homology to any known protein, are not present in WSSV-KR (Supplementary data Fig. S4). In region 5, there are some ORFs present in other strains that are not present in WSSV-KR. Compared to WSSV-CN, WSSV-KR lack three: wsv461, wsv462, and wsv463. The ORF corresponding to ORF14 of WSSV-TH is missing in WSSV-KR. Also, WSSV521 and WSSV522 of WSSV-TW are missing in WSSV-KR in this region (Supplementary data Fig. S5). Of the six ORFs that are not present in WSSV-KR, only four of them showed homology in the protein database search with BLAST. In region 6, WSSV-KR does not have sequences that correspond to wsv475 of WSSV-CN, ORF20 and ORF21 of WSSV-TH, or WSSV002 of WSSV-TW. All of these sequences are 186 nt long. However, these sequences do not identify protein homologs in BLAST database searches. Region 7 is the area where large deletions of sequences occurred in WSSV-KR (Fig. 3A). Compared to WSSV-CN, WSSV-KR lacks 10 ORFs (Supplementary data Figs. S6A and S6C). Half of these did not have homologs in BLAST database searches, while the others did (Supplementary data Table S2). Compared to WSSV-TW, WSSV-KR lacks nine ORFs, including one encoding the WSSV nucleocapsid protein vp35, in this region (Supplementary data Fig. S6B and S6C).

Among the proteins encoded by the 425 ORFs, enzymes involved in nucleotide metabolism, DNA replication and transcription, and some structural proteins are shown in Table 2. dUTPase, encoded by wssv_01120, is a ubiquitous, homotrimeric enzyme encoded by many large DNA viruses. The ORF wssv_0670 showed an identity of 68% over 289 amino acids with *Anolis carolinensis* thymidylate synthase-like predicted protein. ORF wssv_01730 (848 amino acids long) may encode ribonucleotide reductase large subunit (RR1) because it shows high homology to *Cecembia lonarensis* ribonucleoside diphosphate reductase 1 subunit alpha, with an identity of 49%. Ribonucleotide reductase, small subunit (RR2), may be encoded by wssv_01880 because the ORF showed high homology with *Aedes aegypti* ribonucleoside diphosphate reductase small chain, with an identity of 54% over 413 amino acids. The ORF wssv_03950 may encode chimeric thymidine/thymidylate kinase (TKTMK). It is 398 amino acids long and showed 43% identity with *Phytophthora infestans* thymidylate kinase.

WSSV-KR has three ORFs that may encode DNA polymerase. The ORFs are wssv_02300, which shows high homology with *Clostridium phytofermentans* DNA-directed DNA polymerase (37% identity over 82 amino acids); wssv_04830, which shows 34% homology with *Salimicrobium* sp.

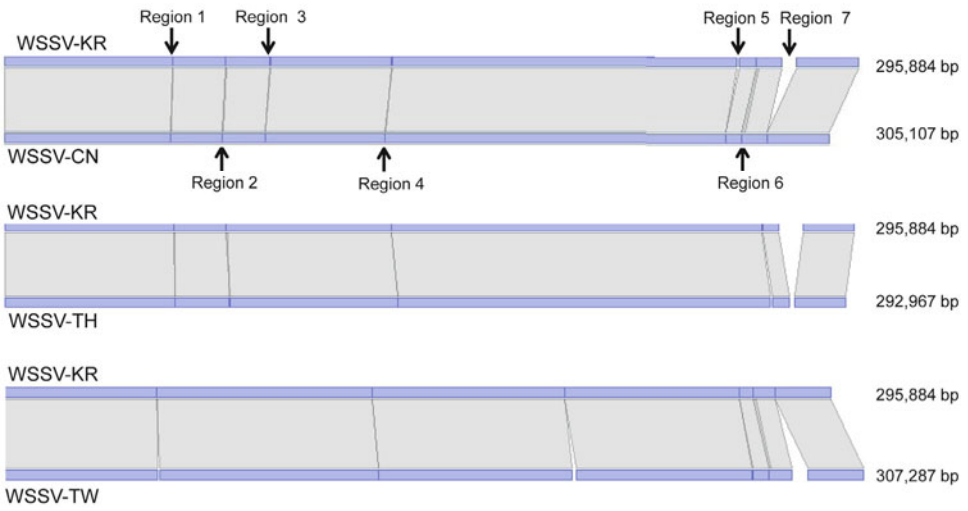

Fig. 3. Linear comparison of WSSV-KR with, WSSV-CN, WSSV-TH, and WSSV-TW. Arrows indicate variable regions between the isolates discussed in the text.

Table 2. Major ORFs encoded by WSSV-KR grouped by function

ORFs	Functions
Nucleotide metabolism	
wssv_01120	dUTP pyrophosphatase
wssv_00670	Thymidylate synthase
wssv_01730	Ribonucleotide reductase large unit
wssv_01880	Ribonucleotide reductase small unit
wssv_03950	Thymidine/thymidylate kinase
DNA replication and transcription	
wssv_02300, wssv_04830, wssv_04980	DNA polymerase
wssv_01910	Endonuclease
wssv_03030	TATA-box binding protein
wssv_00830, wssv_01860, wssv_04230	Protein kinase
Major structural virion proteins	
wssv_02140	vp15
wssv_04140	vp19
wssv_00020	vp24
wssv_03110	vp26
wssv_04210	vp28

DNA polymerase III (Gram-positive type) over 82 amino acids; and wssv_04980, which shows 22% identity over 2,351 amino acids with *Auricularia delicata* putative delta DNA polymerase. ORF wssv_04980 was putatively identified as DNA polymerase, based on the presence of three highly conserved motifs, YGDTDSVFC, KLGMNSMYG, and DMTSLYP. Chen *et al.* (2002a) reported that WSSV DNA polymerase was bigger than the DNA polymerases of other organisms (2,351 amino acid residues vs. 913–1,244 amino acid residue) due to extra-large spacer regions between consensus domains.

ORF wssv_01910 (311 amino acids) encodes a protein that shows 28% homology with *Pelobacter propionicus* DNA/RNA non-specific endonuclease. WSSV-KR may also encode TATA box-binding protein (TBP), because ORF wssv_03030 showed a homology of 27% over 891 amino acids with *Archaeoglobus profundus* TATA box-binding family protein. ORFs wssv_00830 (581 amino acids), wssv_01860 (94 amino acids), and wssv_04230 (730 amino acids) may encode protein kinases (PK) because they showed homology of 26% with *Cryptosporidium muris* protein kinase domain-containing protein, 35% homology with *Haliangium ochraceum* serine/threonine protein kinase, and 21% identity with *Sulfurovum* sp. dual serine/threonine-protein kinase/phosphatase, respectively.

Structural proteins of the virus are of major importance in characterization because these proteins are the first molecules to interact with the host. WSSV-KR encodes several envelope proteins such as vp19 (wssv_04140) and vp28 (wssv_04210), nucleocapsid proteins such as vp15 (wssv_02140) and vp24 (wssv_00020), and tegument protein vp26 (wssv_03110). The vp24, vp26, and vp28 proteins share sequence homology with each other. Vp24 showed 23% sequence similarity to vp28 and vp26 also showed 28% sequence similarity to vp28.

In region region 7, showing the large deletion of sequence in WSSV-KR, nucleocapsid protein vp35 was absent in the WSSV-KR sequence. According to Chen *et al.* (2002b), nucleocapsid protein vp35 is a structural protein that exhibits nuclear targeting behavior. WSSV-CN (305 kb) and WSSV-TW (307 kb) contain this protein but WSSV-KR (296 kb) and WSSV-TH (293 kb), which have smaller full genome sequences, lack this protein (Marks *et al.*, 2004). Pradeep *et al.* (2008) showed that the WSSV Indian strain also lacked this vp35 gene sequence. These reports support the theory that vp35 protein is not important for the virulence of WSSV.

WSSV-KR contains a very large ORF, wssv_03600, which is 18,222 nt long with no known function. WSSV-TH and WSSV-CN also contain extremely large ORFs, ORF167, and wsv360 (both 18,234 nt long), with no homology to sequences in GenBank. The ORF wssv_03600 is 13 nt shorter than ORF167 and wsv360. According to Van Hulten *et al.* (2001), ORF167 is the largest ORF known in viruses. In conclusion, through detailed study and analysis of the genome structure and possible functions of WSSV-KR genes, it can be concluded that WSSV-KR shows little variation compared to other WSSV isolates but further research is required to determine the function of genes and the encoded proteins in relation to its pathogenicity to *L. vannamei.*

This research was supported by Technology Development Program for Fisheries, Ministry for Food, Agriculture, Forestry and Fisheries, Republic of Korea.

References

Chen, L.L., Wang, H.C., Huang, C.J., Peng, S.E., Chen, Y.G., Lin, S.J., Chen, W.Y., Dai, C.F., Yu, H.T., Wang, C.H., Lo, C.F., and Kou, H.K. 2002a. Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus. *Virology* **301**, 136–147.

Chen, L., Leu, J.H., Huang, C.J., Chou, C.M., Chen, S.M., Wang, C.H., Lo, C.F., and Kou, G.H. 2002b. Identification of a nucleocapsid protein (VP35) gene of shrimp white spot syndrome virus and characterization of the motif important for targeting VP35 to the nuclei of transfected insect cells. *Virology* **293**, 44–53.

Crockford, M. 2008. White spot disease. Australia and New Zealand Standard Diagnostic Procedures, 1-13. Department of Fisheries c/o Department of Agriculture and Food, South Perth.

Dieu, B.T.M., Marks, H., Zwart, M.P., and Vlak, J.M. 2010. Evaluation of white spot syndrome virus variable DNA loci as molecular markers of virus spread at intermediate spatiotemporal scales. *J. Gen. Virol.* **91**, 1164–1172.

Esparza-Leala, H.M., Escobedo-Bonillac, C.M., Casillas-Hernándezb, R., Álvarez-Ruíza, P., Portillo-Clarkd, G., Valerio-Garcíac, R.C., Hernández-López, J., Méndez-Lozanoa, J., Vibanco-Péreze, N., and Magallón-Barajasd, F.J. 2009. Detection of white spot syndrome virus in filtered shrimp-farm water fractions and experimental evaluation of its infectivity in *Penaeus (Litopenaeus) vannamei*. *Aquaculture* **292**, 16–22.

Huang, C.H., Zhang, L.R., Zhang, J.H., Xiao, L.C., Wu, Q.J., Chen, D.H., and Li, J.K.K. 2001. Purification and characterization of White Spot Syndrome Virus (WSSV) produced in an alternate host: crayfish, *Cambarus clarkii*. *Virus Res.* **76**, 115–125.

Jang, I., Suriakala, K., Kim, B., and Meng, X. 2008. Real-Times PCR quantification of white spot syndrome virus (WSSV) and hepatopancreatic parvovirus (HPV) loads in shrimp and seawaters of shrimp ponds on the west coast of South Korea. *Fish. Aquat. Sci.* **11**, 195–204.

Lightner, D.V. 1999. The penaeid shrimp viruses TSV, IHHNV, WSSV, and YHV: Current Status in the Americas, Available Diagnostic Methods, and Management Strategies. *J. Appl. Aqua-*

cult. **9**, 27–52.

Lightner, D.V. and Lo, C.F. 2008. White spot disease. European Community Reference Laboratory for Crustacean Diseases leaflet 2008.

Lo, C.F., Hsu, H.C., Tsai, M.F., Ho, C.H., Peng, S.E., Kou, G.H., and Lightner, D.V. 1999. Specific genomic DNA fragment analysis of different geographical clinical samples of shrimp white spot syndrome virus. *Dis. Aquat. Organ.* **35**, 175–185.

Marks, H., Goldbach, R.W., Vlak, J.M., and Van Hulten, M.C.W. 2004. Genetic variation among isolates of *White spot syndrome virus*. *Arch. Virol.* **149**, 673–697.

Park, J.H., Lee, Y.S., and Lee, Y. 1998. An infectious viral disease of penaeid shrimp newly found in Korea. *Dis. Aquat. Organ.* **34**, 71–75.

Pazir, M.K., Afsharnasab, M., Niamaymandi, N., Khaden, H., Akbarpour, E., and Zendebudi, A.A. 2012. Histopathological observation of white spot syndrome virus and infectious hypodermal and hematopoiectic necrosis virus in shrimp farms, *Litopenaeus vannamei*, in Bushehr Province, Iran. *Asian J. Anim. Sci.* **6**, 209–219.

Pradeep, B., Shekar, M., Karunasagar, I., and Karunasagar, I. 2008. Characterization of variable genomic regions of Indian white spot syndrome virus. *Virology* **376**, 24–30.

Tan, Y., Xing, Y., Zhang, H., Feng, Y., Zhou, Y., and Shi, Z.L. 2009. Molecular detection of three shrimp viruses and genetic variation of white spot syndrome virus in Hainan Province, China, in 2007. *J. Fish. Dis.* **32**, 777–784.

Tsai, M.F., Yu, H.T., Tzeng, H.F., Leu, J.H., Chou, C.M., Huang, C.J., Wang, C.H., Lin, J.Y., Kou, G.H., and Lo, C.F. 2000. Identification and characterization of a shrimp white spot syndrome virus (WSSV) gene that encodes a novel chimeric polypeptide of cellular-type thymidine kinase and thymidylate kinase. *Virology* **277**, 100–110.

Van Hulten, M.C.W., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R.K., and Vlak, J.M. 2001. The white spot syndrome virus DNA genome sequence. *Virology* **286**, 7–22.

Yang, F., He, J., Lin, X.H., Li, Q., Pan, D., Zhang, X.B., and Xu, X. 2001. Complete genome sequence of the shrimp white spot bacilliform virus. *J. Virol.* **75**, 11811–11820.