Transcription Regulation and Genomics of White Spot Syndrome Virus

HENDRIK MARKS[1], XINYING REN[1,2], JEROEN WITTEVELDT[1], HANS SANDBRINK[2,†], JUST M. VLAK1 AND MARIëLLE C.W. VAN HULTEN[1]

[1]Laboratory of Virology, Wageningen University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands

[2]Business Unit Genomics, Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands

ABSTRACT

White spot syndrome virus (WSSV) is a major disease agent of penaeid shrimp and other crustaceans. The enveloped virus particle has an ovoid-to-bacilliform shape with a tail-like appendage. Its rod-shaped striated nucleocapsid contains a large circular double-stranded DNA molecule with a size of about 300 kb. The virus replicates and assembles in the nucleus of infected cells. The WSSV genome is characterized by the presence of nine homologous regions (*hrs*), dispersed along the genome, each containing a variable number of 250-bp tandem repeats. The viral DNA encodes 184 open reading frames (ORFs) of which only 12 have putative homologues in databases, mainly representing genes encoding enzymes for nucleotide metabolism, DNA replication and protein modification. The remaining ORFs are mostly unassigned except for 19, which encode virion proteins. On the basis of its unique morphological and genetic features, WSSV has been accommodated in a new virus family (*Nimaviridae*) in the genus *Whispovirus*. WSSV isolates from different geographic origins differ primarily in a single region where extensive deletions can occur. Other variable regions of the WSSV genome include the *hrs* and short repeats, which can be used to study WSSV epidemiology and evolution. The analysis of the different WSSV isolates shows that not all WSSV genome regions are suitable for use in diagnostic tests. The WSSV sequence was further analyzed *in silico*, using a set of computer scripts, for the presence of conserved promoter motifs. However, besides the common TATA motif, no other conserved motifs were found. This conclusion is confirmed by experimental data performed on transcription initiation of WSSV genes.

Marks, H., X. Ren, J. Witteveldt, H. Sandbrink, J.M. Vlak and M.C.W. van Hulten. 2005. Transcription regulation and genomics of White Spot Syndrome Virus. *In* P. Walker, R. Lester and M.G. Bondad-Reantaso (eds). Diseases in Asian Aquaculture V, pp. 363-377. Fish Health Section, Asian Fisheries Society, Manila.

INTRODUCTION

White spot syndrome virus (WSSV) is a pathogen of major economic importance in cultured penaeid shrimp. The virus is not only present in shrimp, but has also been found in other freshwater and marine crustaceans including crabs and crayfish (Lo *et al*., 1996). In cultured shrimp, WSSV infection can reach a cumulative mortality of up to 100% within 3-10 days (Lightner, 1996) and can cause large economic losses to the shrimp culture industry. Thevirus was first discovered in China (Fujian) (1991/1992), from where it quickly spread to other shrimp farming areas in Southeast Asia (Cai *et al*., 1995). Initially WSSV appeared to be limited to Asia until it was found in Texas and South Carolina in November 1995 (Rosenberry, 1996). In early 1999, WSSV was also reported from Central- and South-America (Rosenberry, 2000). Intensive shrimp cultivation, inadequate sanitation and worldwide trade has aggravated the disease incidence in crustaceans and enhanced disease dissemination. As such WSSV has become an epizootic disease and is not only a major threat to shrimp culture (mariculture) but also to marine ecology (Flegel, 1997).

The nature of the virus has been elusive for some time. The virus was/is known under various names such as "Hypodermal and hematopoietic necrosis baculovirus" (HHNBV), "Rod-shaped nuclear virus of *Penaeus japonicus*" (RV-PJ), "Chinese baculovirus" (CBV), "Systemic ectodermal and mesodermal baculovirus" (SEMBV), "Penaid rod-shaped DNA virus" (PRDV) and "White spot baculovirus" (WSBV), but the virus is now best known as WSSV. Genetic information on the virus was lacking until recently (van Hulten *et al*., 2000c). Since then a burst of genetic information has become available including the entire sequence of WSSV (van Hulten *et al*., 2001a; Yang *et al*., 2001). In this paper we will focus on the molecular genetics and transcription regulation of WSSV including the latest results from the Wageningen group.

The WSSV genome

The complete double-stranded (ds) DNA genome from a WSSV isolate originating from Thailand (WSSV-Th) consists of a circular sequence of 292,967 base pairs (bp) in size (van Hulten *et al*., 2001a). The adenine residue at the translation initiation codon of the major structural virion envelope protein VP28, of which the coding capacity has been confirmed by amino acid sequencing (van Hulten *et al*., 2000a), was designated as the starting point of the circular physical map of the WSSV genome (Fig. 1). Hundred and eighty four (184) ORFs varying from 51 to 6077 amino acids in size have been located on the WSSV genome (Fig. 1). Based on homologies with other viral or cellular genes in GenBank only 12 of the 184 WSSV ORFs have been assigned a putative function or have similarities with known genes (Table 1).

The majority of known genes are involved in DNA replication or nucleotide metabolism such as DNA polymerase (Chen *et al*., 2002a; van Hulten *et al*., 2001a), ribonucleotide reductase (Tsai *et al*., 2000b; van Hulten *et al*., 2000c), Chimeric thymidine-thymidylate kinase (Tsai *et al*., 2000a; Tzeng *et al*., 2002), thymidylate synthase, dUTPase and a non-specific endonuclease (Witteveldt *et al*., 2002). Besides these ORFs the genes for five major and fourteen minor structural virion proteins have been analyzed and their genes identified on the genome (Fig. 1, Table 1) (Huang *et al*., 2002; van Hulten *et al*., 2000a, 2000b, 2002).

Table 1. Overview of WSSV ORFs with known functions (ORFs according to van Hulten *et al*., 2001a).

Homology to GenBank		Virion proteins	
		Major proteins	Minor proteins
DNA polymerase	(ORF27)	VP28 (ORF1)	ORF6
Collagen	(ORF30)	VP24 (ORF31)	ORF29
Thymidylate synthase	(ORF54)	VP15 (ORF109)	ORF30
Protein Kinase (2x)	(ORF2 and ORF61)	VP26 (ORF153)	ORF34
dUTPase	(ORF71)	VP19 (ORF182)	ORF75
Ribonucleotide reductase I	(ORF92)		ORF112
Ribonucleotide reductase II	(ORF98)		ORF118
Endonuclease	(ORF99)		ORF120
Class I cytokine receptor	(ORF112)		ORF127
TATA box binding protein	(ORF149)		ORF128
Chimeric Thymidine-thymidylate kinase	(ORF171)		ORF149
			ORF151
			ORF168
			ORF183

Some of these (VP28, VP19) have potential in intervention strategies to prevent disease (van Hulten *et al*., 2001b; Witteveldt *et al*., 2004a, 2004b, 2005). As till the end of 2002 WSSV is the largest animal virus entirely sequenced.

Nine direct repeat regions with different sizes were found dispersed throughout the genome (Fig. 1). Similar repeat regions have also been identified in baculovirus genomes and these regions were, in accordance to baculovirus terminology (Cochran and Faulkner, 1983), designated homologous region (*hr*) 1 to *hr*9 (van Hulten *et al*., 2001a). Analysis of these regions in WSSV revealed that they all consisted of identical repeat units of 250 bp or parts thereof. The hrs are largely located in intergenic regions, although several short ORFs are encoded within the *hrs* (Fig. 1; van Hulten *et al*., 2001a). Recently, a number of proteins that interact with WSSV *hr*3 were identified, including a putative shrimp arginine kinase and WSSV protein ORF59 (Li *et al*., 2003). Besides WSSV and baculoviruses, repeat regions dispersed throughout the genome have also been found in ascoviruses (Bigot *et al*., 2000). In baculoviruses it was demonstrated that these *hrs* function as enhancers of transcription (Guarino and Summers, 1986) and putative origins of replication (Kool *et al*., 1993; Pearson *et al*., 1992). Based on the observation that all large DNA viruses with a circular genome configuration (WSSV, baculoviruses and ascoviruses) contain *hrs* dispersed along their genomes, it is very well possible that they have a function in WSSV replication and transcription regulation. Their presence may imply that these viruses share a similar replication strategy and this may cluster WSSV, the baculoviruses and the ascoviruses into a superfamily of large circular ds DNA viruses. However, the function of the WSSV *hrs* in either replication or transcription remains to be investigated.

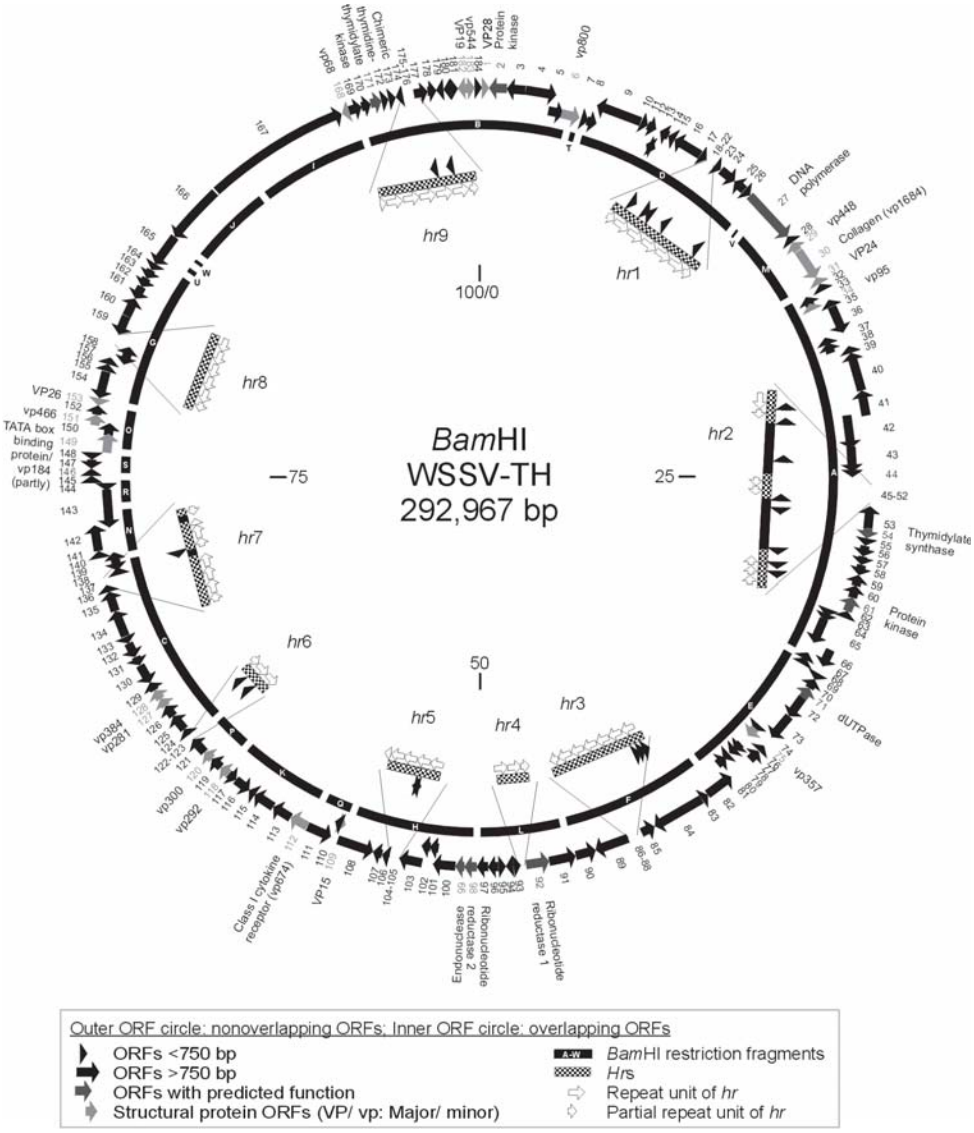

Figure 1. Map of the circular ds DNA WSSV-Th genome showing the genomic organization. Sites for BamH1 are shown in the inner circle. Fragments are numbered from A to W according to size from the largest (A) to the smallest (W) (van Hulten *et al*., 2001a). The positions of the ORFs are indicated by arrows, which also represent the direction of transcription. Dark gray arrows indicate ORFs with homologues in public databases. Light gray arrows represent the structural protein genes. The *hr* sequences are enlarged within the inner circle. The scale on the inner circle is in map units.

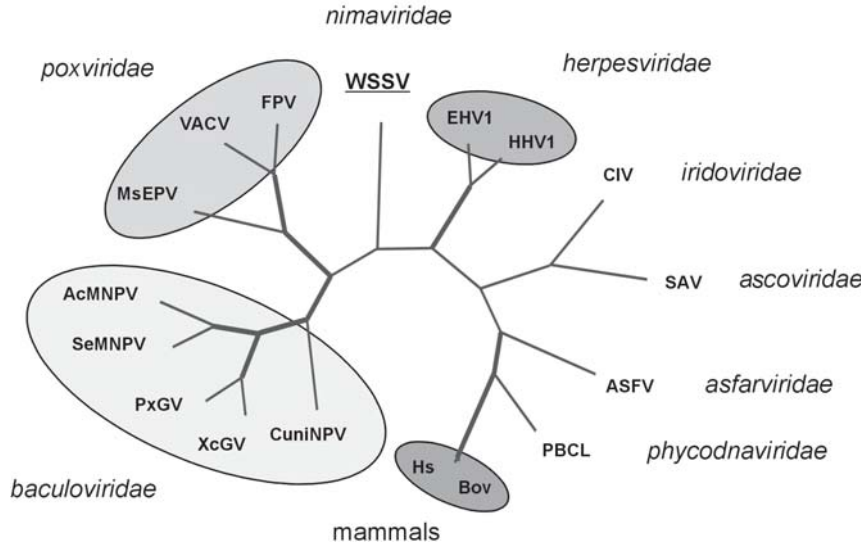

Figure 2. Bootstrap analysis (100 replicates) of an unrooted phylogenetic tree of DNA polymerase proteins constructed with the PAUP heuristic search algorithm. Thick lines indicate frequency of clusters over 70%. DNA Polymerase genes used and their accession numbers between brackets: EHV1: Equine herpesvirus 1 (NP_041039), HHV1: Human herpesvirus 1 (NP_044632), CIV: *Chilo* iridescent virus (AAD48150), SAV: *Spodoptera frugiperda* ascovirus 1 (CAC19170), ASFV: African swine fever virus (NP_042783), PbCl: *Paramecium bursaria Chlorella* virus 1 (NP_048532), Bov: *Bos taurus* (P28339), Hs: *Homo sapiens* (S35455), CuniNPV: *Culex nigripalpus* nucleopolyhedro baculovirus (AF274291), XcGV: *Xestia c-nigrum* granulo baculovirus (NP_059280), PxGV: *Plutella xylostella* GV (NP_068312), SeMNPV: *Spodoptera exigua* multiple NPV (NP_037853), AcMNPV: *Autographa californica* MNPV (NP_054095), MsEPV: *Melanoplus sanguinipes* entomopoxvirus (NP_048107), VACV: Vaccinia virus (NP_063712), FPV: Fowlpox virus (NP_039057).

WSSV TAXONOMY

The taxonomic position of WSSV has been evaluated using gene phylogeny on several genes identified on the WSSV genome, including the ribonucleotide reductase large (*rr1*) and small (*rr2*) subunits (van Hulten *et al*., 2000c), the protein kinase (pk) genes (Liu *et al*., 2001; van Hulten and Vlak, 2001) and the thymidine kinase and thymidylate kinase (tk-tmk) (Tsai *et al*., 2000a). The putative DNA polymerase gene (ORF27) was also used in an alignment with 14 other viral and 2 eukaryotic polymerases (Chen *et al*., 2002a; van Hulten *et al*., 2001a). All seven conserved DNA polymerase sequence motifs and the three conserved regions implicated in DNA polymerase 3'-5' exonuclease activity (Bernad *et al*., 1987; Bernad *et al*., 1989) were identified. The WSSV DNA polymerase gene is larger than other DNA polymerase genes, as it has a longer N-terminal region and extra amino acid stretches between the conserved domains. The function of these additional amino acid regions is unknown. The phylogenetic analysis was performed using the regions containing the conserved DNA polymerase motifs. In the DNA polymerase tree (Fig. 2) the different virus families are all present in clades which are well bootstrap-supported (van Hulten *et al*., 2001a). Together with the phylogenetic analysis of both RRs, the PK and the chimeric TK

and TMK (Liu *et al*., 2001; Tsai *et al*., 2000a; van Hulten *et al*., 2000c; van Hulten and Vlak, 2001) this tree confirms the unique taxonomic position of WSSV. Based on the primary genome structure, the extensive phylogenetic analysis and the distinct morphology of its virion (Durand *et al*., 1997; Nadala *et al*., 1998; Wongteerasupaya *et al*., 1995), WSSV has a unique position among the large ds DNA viruses and it is therefore justified to consider this virus to be a member of a new virus family. A proposal to this end has been submitted to the International Committee on Taxonomy of Viruses (ICTV) (van Hulten and Vlak, 2002; Vlak *et al*., 2003) and has been approved by the ICTV (ICTV Plenary session, Paris, 2002; see also http://www.ncbi.nlm.nih.gov/ICTV/db). In this proposal WSSV belongs to the family Nimaviridae (Nima = thread, appendage) and is the type member of the genus *Whispovirus*. In the new nomenclature proposed by the ICTV (pending a decision) WSSV as a species would be named *Penaeus monodon* whispovirus type 1 (Vlak *et al*., 2003).

VARIATION WITHIN THE SPECIES WSSV

Three WSSV isolates from different geographic origins have now been completely sequenced. One isolate was collected from a WSSV-infected pond in Suratthani, Thailand in 1996 (van Hulten *et al*., 2001a; AF369029), the second isolate originated from shrimp ponds in Tongan, Xiamen, East China in 1996 (Yang *et al*., 2001; AF332093), and the third from shrimp farms located in Southern Taiwan in 1994 (Wang *et al*., 1995; AF440570). The isolates will be referred to as WSSV-Th, WSSV-Cn and WSSV-Tw, respectively, according to their country of origin. The main characteristics for each isolate are shown in Table 2.

To analyze the genetic differences between the three WSSV isolates, the sequences were compared using dotplot analysis (Fig. 3; Marks *et al*., 2004). This analysis revealed that the three genome sequences were highly similar (over 99% nucleotide identity) and that the size differences could largely be explained by a ~13 kb segment present in the WSSV-Tw genome, partially present in WSSV-Cn genome (~12 kb) and absent in the WSSV-Th genome (Fig. 4). The 12,049 bp insertion in the WSSV-Cn genome was located at position 31135 in the intergenic region of ORF23 and ORF24 in the WSSV-Th genome, and contains 14 ORFs, of which two ORFs belong to gene family 4 of the WSSV-Th genome (Fig. 4; van Hulten *et al*., 2001a). Another ORF showed homology with WSSV-Th ORF23, which is flanking the deletion in WSSV-Th. For the other ORFs no homologues could be identified in GenBank. The presence of ORFs belonging to WSSV gene families in the 12,049 bp WSSV-Cn specific segment, strongly suggests that this is an authentic part of the WSSV genome and that the shorter size of the WSSV-Th genome is due to a deletion event. ORF I is thought to encode a nucleocapsid protein (VP35; Fig. 4; Chen *et al*., 2002b) and this protein should therefore be absent in virions of WSSV-Th. The function of the remaining genes present on the 12,049 bp segment is not known. However, all genes present on this segment are apparently dispensable for infection and replication in *Penaeus monodon* and *Procambarus clarkii* (van Hulten *et al*., 2001a). A part of the deleted fragment (12756-14204; WSSV-Tw coordinates) shows 68% nucleotide identity with the genomic fragment located at position 18320-19787 (Tw coordinates; located in ORF25), most likely as a consequence of sequence duplication. As the virus contained two similar sequences after this duplication, maybe the deletions can be explained by the genomic pressure on the virus to discard abundant sequences. The mechanism by which the deletions occur is at present unclear, although it was suggested that it occurred by homologous recombination (Marks *et al*., 2004).

Table 2. Characteristics of the isolates WSSV-Th, WSSV-Cn and WSSV-Tw.

	WSSV-Th	**WSSV-Cn**	**WSSV-Tw**
Genome size	292,967	305,107	307,287
G + C content (mol %)	41.1	41.0	41.0
Collection date of infected shrimp	05-1996	10-1996	11-1994
Species of origin	*Penaeus monodon*	*Penaeus japonicus*	*Penaeus monodon*

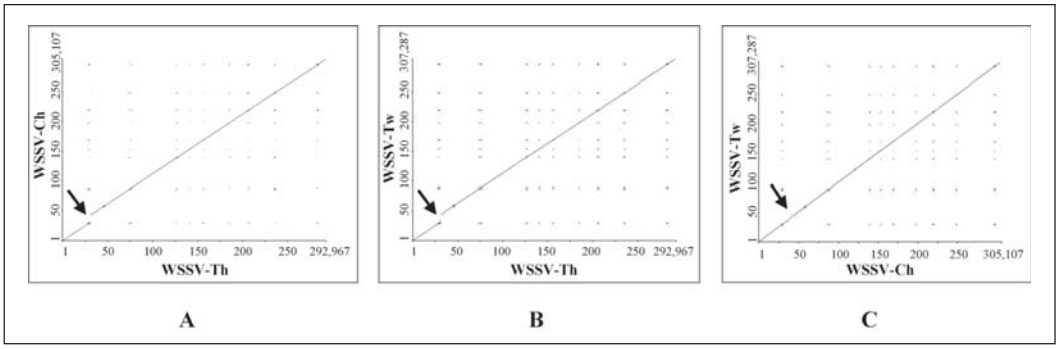

Figure 3. Dotplot comparison of the nucleotide sequences of WSSV-Th and WSSV-Cn (A), of WSSV-Th and WSSV-Tw (B), and of WSSV-Cn and WSSV-Tw (C).

Further comparison of the three genomes revealed several less prominent differences, which roughly fall into two categories: (i) single nucleotide polymorphisms (SNPs) and (ii) polymorphisms in repeats. In the WSSV-Th genome nine *hrs* have been identified (van Hulten *et al*., 2001a), which were also present in WSSV-Cn and WSSV-Tw. The location and sequence of these *hrs* were conserved, whereas the number of repeat units within four of the *hrs* varied between the three isolates. Several ORFs containing small repeated domains e.g. ORF75, ORF94 (ORF4 in van Hulten *et al*., 2000c) and ORF125, were also different between these isolates, as the number of repeated domains varied. The SNPs, variations in repeat regions, and the large deletion may be relevant markers to study WSSV evolution, as well as to assist in understanding WSSV epidemiology (Marks *et al*., 2004; Hoa *et al*., 2005; Wongteerasupaya *et al*., 2003).

It can be concluded that even though the WSSV-Th and WSSV-Cn genomes are very similar, several genomic rearrangements have occurred, of which the ~12 kb deletion is the most drastic one. It should be noted that two WSSV fragments (C42 and LN4) described by Wang *et al*., (2000) for WSSV detection are lacking in the WSSV-Th genome sequence (van Hulten *et al*., 2001a), as they are part of the 12,049 bp insertion fragment of the WSSV-Cn genome. Therefore, these fragments are not suitable for reliable detection of WSSV, as they are present in a variable part of the genome, not essential for WSSV infection of shrimp, and absent in the infectious WSSV-Th isolate. This genome comparison shows that care should be taken when using WSSV detection tests, based on genomic fragments. Fragments located at multiple positions of the WSSV genome should be included in a DNA probe or for detection using PCR to improve reliable WSSV detection. Various PCR-based WSSV detection kits are available and very sensitive (2-step PCR; e.g. DiagXotics Inc.,

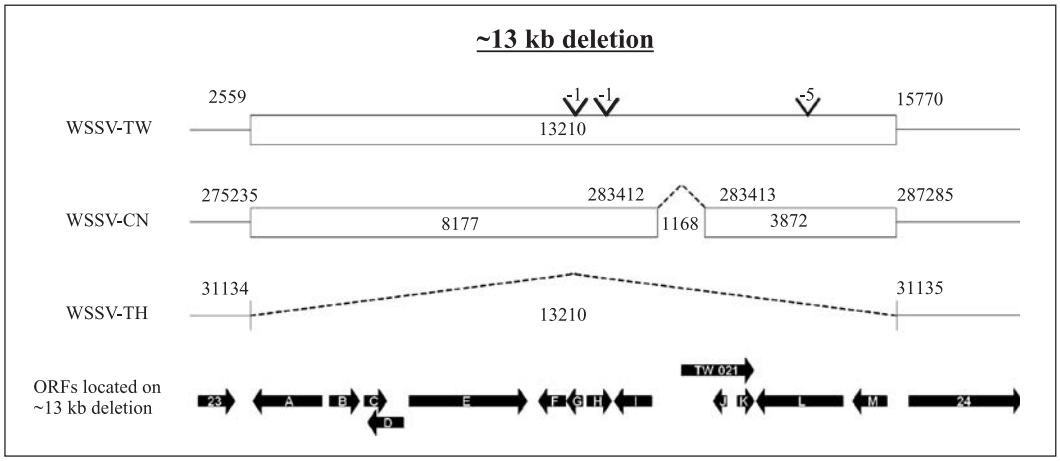

Figure 4. Schematic representation of the ~13 kb deletion of WSSV-Tw, WSSV-Cn and WSSV-Th. The map numbers, indicated above each isolate, are in accordance with the numbers in the NCBI databank for the genomic sequence of each isolate (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide). The length of the fragments is indicated within boxes or sequences. The positions of the ORFs located in this region are indicated by arrows, which also represent the direction of transcription. ORFs are numbered in accordance with the numbering used by Marks *et al*. (2004).

USA; IQ2000, Farm IntelliGene Tech. Corp., Taiwan; PWSBV Detection Kit, Xiamen Talent Biotech Co., Ltd., China; Fast Target(tm) White Spot Virus Detection Kit, Aquatic Health Unit, University Putra Malaysia, Serdang, Malaysia), but they are all relying on a single segment and thus prone to miss new variants of the virus due to deletions in the pertinent PCR sequences.

TRANSCRIPTION AND PROMOTER ANALYSIS

Genes of most large ds DNA viruses infecting (in)vertebrates are expressed in a cascaded fashion. Immediate early (IE) and early (E) genes are expressed before viral DNA replication, while expression of late (L) genes occurs after replication of the viral genome. Nucleotide motifs in the upstream region of viral genes are often involved in gene regulation and the kinetics of expression. Late genes usually encode structural virion proteins and are often expressed in a coordinate fashion. For proper late gene expression, the motif that contains the transcription initiation site (TIS) often plays a prominent role (Davison and Moss, 1989; Garcia-Escudero and Viñuela, 2000; Morris and Miller, 1994; Weir, 2001). Despite the available information about pathogenicity, epidemiology, virion structure and the genomic sequence of WSSV, insight in gene regulation is limited. To identify putative regulatory elements in WSSV promoter regions, a statistical method, based on a model proposed by Brazma *et al*. (1998), was used.

A set of computer scripts, which were made in the programming language Perl, was used to calculate the statistical difference in 4, 5, 6, 7 or 8 nucleotide motif frequencies in the upstream sequences of all genes compared to the total WSSV genome. Promoter regions of genes that are totally or partly located in the homologous regions (*hrs*) and the *hrs* itself were excluded from the analysis to avoid the possibility of finding false motifs due to the high homology of the sequences in the *hrs*. As nucleotide motifs with a regulatory function

are expected to be present in multiple promoter regions, only motifs that were present in upstream sequences of at least 5 ORFs were analyzed after running the scripts.

As a control, the scripts were tested on the genome sequence of AcMNPV (Ayres *et al*., 1994). Baculoviruses share the CAGT and a/g/tTAAG motif as transcription initiation site of early and late genes, respectively (Blissard and Rohrmann, 1989). The TAAG sequence occurs less frequently than would be expected for a random sequence in the whole AcMNPV genome, and is relatively enriched in the 80 bp promoter sequences of the AcMNPV ORFs (Ayres *et al*., 1994). The results of the 4-mer motif frequency in the 80 bp upstream of all AcMNPV ORFs analyzed with our computer scripts (Table 3a) not only shows that the TAAG motif frequency is 28% of the expected occurrence in the whole genome, it also shows this motif has the highest relative enrichment in promoter regions (4.6 times) of all possible 4-mer motifs. Also the baculovirus early promoter motif CAGT is relatively more present in promoter regions (1.6 times), although not as prominently as TAAG. Parts of the TATA-box (TATA, ATAA) as well as the sequence GATA, part of the GATA motif, which has a regulatory effect in expression of the early baculovirus gene gp64 and probably on the late gene *polh* (Kogan and Blissard, 1994; Ooi *et al*., 1989), occur relatively often in the promoter regions of the ORFs. The analysis using 160 bp upstream of the ORFs shows a similar outcome (data not shown), although the enrichment of TAAG is less pronounced (3.2 times), supporting the observation that in baculovirus the TAAG promoter motif is often located within 80 bp upstream of the ATG. Analysis of 5-mer motifs revealed that a/g/tTAAG was enriched in the promoter region, but not CTAAG (using 80 bp upstream of ORFs). Analysis of 6-mer motifs, using 160 bp upstream of ORFs, showed a relative enrichment of 4.6 of the consensus TATA-box TATAAA in promoter regions, and showed a specific enrichment (2.1 times) for the consensus GATA motif ^{a}/tGATA^{a}/t. Except for the baculovirus AcMNPV, the scripts also showed an enrichment of motifs important for immediate early and late gene expression for the large ds DNA virus *Chilo* iridescent virus (CIV; Nalçacioglu *et al*., 2003).

The same scripts were used to analyze the upstream sequences of the WSSV ORFs. For the frequency of a 4-mer motif 100 nts upstream of the ORFs, the TATA motif (TATA, ATAA, TAAA) appears relatively often in promoter sequences (Table 3b). The enrichment of these sequences is similar to what was observed for AcMNPV (Table 3a). No 4-mer motif is enriched in the promoter regions in the same order of magnitude as the TAAG sequence for AcMNPV. The results for the analysis of 200 bp upstream of the translational starts were only slightly different (data not shown). The analyses of 5, 6, 7 or 8 bp motifs in 100 or 200 bp show a relative enrichment of 4.7 of the consensus TATA-box TATAAA in promoter regions using 100 bp upstream of ORFs. The putative early WSSV promoter motif TCA^{c}/tTC (Tsai *et al*., 2000b) occurs almost as often (265 times) as expected in the total genome (300 times), but is 2.1 times enriched in promoter regions (100 bp upstream of ORFs). Furthermore, some motifs consisting of G and C, such as CCCG and CCGG (Table 3b), and G/C-rich sequences, have a relative high frequency in promoter regions and a lower frequency in the total WSSV genome than expected (data not shown). Because none of the 4, 5, 6, 7 or 8 nucleotide motifs in the WSSV promoter regions were as pronounced as the occurrences of the AcMNPV TAAG motif, a putative (late) WSSV promoter motif could not be extracted from the computer analysis.

Table 3. Frequency of 4-nucleotide motifs in the 5' upstream regions of the ORFs as compared to the whole genome for the baculovirus AcMNPV (A) and WSSV (B).

A: AcMNPV (the promoter region included 80 bp 5' upstream of the ORFs)

4-mer (motif)	Occurrence in genome[a] (% of expected occurrenceb)		Occurrence in promoter regions (% of expected occurrence[b])		Relative enrichment in promoter regions[c]
taag	393	(28)	83	(128)	4.6
tata	1314	(64)	145	(152)	2.4
ataa	1973	(98)	188	(203)	2.1
agta	671	(48)	62	(96)	2.0
atat	1616	(78)	143	(150)	1.9
aagg	473	(49)	39	(89)	1.8
ctag	256	(26)	20	(45)	1.7
tagt	737	(51)	53	(80)	1.6
aata	2230	(111)	160	(172)	1.6
cagt	698	(72)	50	(111)	1.6
atta	1957	(95)	138	(145)	1.5
taaa	2716	(135)	190	(205)	1.5
cact	612	(63)	42	(94)	1.5
aagc	909	(96)	62	(142)	1.5
gata	867	(62)	59	(91)	1.5

B: WSSV (the promoter region included 100 bp 5' upstream of the ORFs)

4-mer (motif)	Occurrence in genome[a] (% of expected occurrenceb)		Occurrence in promoter regions (% of expected occurrence[b])		Relative enrichment in promoter regions[c]
tata	2630	(60)	164	(128)	2.2
ataa	3431	(74)	201	(150)	2.0
taaa	3538	(77)	195	(146)	1.9
aacc	1774	(79)	92	(142)	1.8
aaaa	6450	(134)	330	(236)	1.8
accc	1333	(88)	68	(154)	1.8
accg	699	(46)	33	(75)	1.6
tacc	1424	(67)	64	(103)	1.5
caac	2792	(125)	123	(190)	1.5
aata	4104	(89)	179	(134)	1.5
tttt	6450	(160)	281	(40)	1.5
taac	1888	(60)	82	(90)	1.5
cccg	583	(56)	25	(83)	1.5
ccgg	374	(36)	16	(53)	1.5
ccgt	940	(64)	40	(94)	1.5

[a] both strands, excluding hrs.
[b] expected occurrence is the occurrence of a 4-mer based on random distribution of nucleotides in the total genome.
[c] only the 15 motifs with the highest relative enrichment are shown

Table 4. (Conserved) nucleotide sequences at the transcription initiation sites of WSSV mRNAs.

Gene(s)	Motif	Reference
vp28	TAACCAA	Marks *et al*., 2003
vp26	TGCTACA	Marks *et al*., 2003
vp19	CAAAAAC	Marks *et al*., 2003
vp24/vp15	TCATGAC	Marks *et al*., 2003
rr1/rr2	TCAc/tTC	Tsai *et al*., 2000b
DNA-pol	CACAGTC	Chen *et al*., 2002a
pk	GAGGGTG	Liu *et al*., 2001

The transcription initiation site has been determined experimentally for the putative early genes encoding the large and small subunit of ribonucleotide reductase (*rr1* and *rr2*, respectively), protein kinase (*pk*) and DNA polymerase (*DNA-pol*) (Chen *et al*., 2002a; Liu *et al*., 2001; Tsai *et al*., 2000b). Recently, transcriptional analysis has also been performed on the five major structural virion protein genes *vp28, vp26, vp24, vp19* and *vp15*, putative late genes (Marks *et al*., 2003) (Table 4). The 5' RACE studies only revealed a consensus transcription initiation motif for the two ribonucleotide reductase genes (*rr1/rr2*) and for only two of the five major virion protein genes (*vp15/vp24*). No other conserved sequence motifs could be detected in the sequences surrounding the transcriptional start sites of the studied WSSV genes. This is most striking for the five major structural protein genes as their expression most likely is co-regulated to secure correct assembly of the virion. Consensus TATA box sequences have been found at a functional distance upstream of TISs of the early genes. Late transcripts seem to start about 25 nucleotides downstream of an A/T rich region. No evidence has been obtained for the occurrence of spliced transcripts.

CONCLUSIONS

The complete nucleotide sequence of WSSV has now been determined and it appeared to be the largest animal virus sequenced as till the end of 2002. However, despite the ever increasing number of ORFs and genes in databases such as GenBank, only very few ORFs of WSSV could be assigned certain functions. These known functions relate to DNA replication, nucleotide metabolism and protein modification. Another 19 ORFs are assigned to structural virion proteins, but very few others out of the 184 ORFs have been analyzed suggesting that WSSV holds a lot of novel features. Therefore, the functional analysis of all these unknown ORFs will be the challenge ahead. It is well possible that many of these are related to the aquatic environment of this virus.

Multiple sequences of WSSV genomes are now available and extensive phylogenetic analysis has been performed on several of the conserved genes giving further credence to the unique position of this virus. This has been acknowledged by the establishment of a new virus family by the ICTV to accommodate WSSV (*Nimaviridae*). However, only a single species and a single genus have been established. It would be important to identify other unassigned virus isolates and to see if they belong to the same family.

When the three WSSV complete genome sequences available in GenBank are compared, the homology of the sequences is very high, which suggests that the virus has started to

spread only recently from its original host. Repeats and one specific genomic sequence seem to be the only genomic regions, which are readily subject to major changes. These regions are therefore valuable for the study of the spread of WSSV and hence to design strategies to control spread.

As a member of a new virus family the study on gene regulation and gene function of WSSV is of major interest. However, the research has been hampered by the lack of a suitable cell system for this virus, despite numerous attempts. Primary cultures from lymphoid tissue are only a limited alternative. Computational analysis of the sequences upstream of the translational start codons shows that gene regulation may be different from other large dsDNA viruses as no specific sequence motifs were identified. The absence of these consensus suggests a unique regulation of WSSV transcription, in line with its unique taxonomic position.

ACKNOWLEDGEMENTS

This paper is dedicated to Dr. Hans Sandbrink, who died May 10th, 2002. The research was supported by Intervet International BV, Boxmeer, The Netherlands.

REFERENCES

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez Ferber, M. and Possee, R.D. 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology 202, 586-605.

Bernad, A., Zaballos, Z., Salas, M. and Blanco, L. 1987. Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. EMBO Journal 6, 4219-4225.

Bernad, A., Blanco, L., Lazaro, J.M., Marin, G. and Salas, M. 1989. A conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. Cell 59, 219-228.

Bigot, Y., Stasiak, K., Rouleux-Bonnin, F. and Federici, B.A. 2000. Characterization of repetitive DNA regions and methylated DNA in ascovirus genomes. Journal of General Virology 81, 3073-3082.

Blissard, G.W. and Rohrmann, G.F. 1989. Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. Virology 170, 537-555.

Brazma, A., Jonassen, I., Vilo, J. and Ukkonen, E. 1998. Predicting gene regulatory elements in silico on a genomic scale. Genome Research 8, 1202-1215.

Cai, S.L., Huang, J., Wang, C.M., Song, X.L., Sun, X., Yu, J., Zhang, Y. and Yang, C.H. 1995. Epidemiological studies on the explosive epidemic disease of prawn in 1993-1994. Journal of Fisheries of China 19, 112-117.

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, G.H. 2002a. Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus. Virology 301, 136-147.

Chen, L.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.

Cochran, M.A. and Faulkner, P. 1983. Localization of homologous DNA sequences interspersed at five regions in the baculovirus AcMNPV genome. Journal of Virology 45, 961-970.

Davison, A.J. and Moss, B. 1989. Structure of vaccinia virus late promoters. Journal of Molecular Biology 210, 771-784.

Durand, S., Lightner, D.V., Redman, R.M. and Bonami, J.R. 1997. Ultrastructure and morphogenesis of white spot syndrome baculovirus (WSSV). Diseases of Aquatic Organisms 29, 205-211.

Flegel, T.W. 1997. Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. World Journal of Microbiology and Biotechnology 13, 433-442.

Garcia-Escudero, R. and Vi_uela, E. 2000. Structure of African swine fever virus late promoters: requirement of a TATA sequence at the initiation region. Journal of Virology 74, 8176-8182.

Guarino, L.A. and Summers, M.D. 1986. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed early gene expression. Journal of Virology 60, 215-223.

Hoa, T.T.T., Hodgson, R.A.J., Oanh, D.T.H., Phuong, N.T., Preston N.J. and Walker P.J. (2005). Variations in tandem repeat DNA segments in the ribonucleotide reductase gene of white spot syndrome virus (WSSV) isolates from Vietnam. *In* Walker, P.J., Reantaso, M.B. and Lester, R.J. (eds.). Diseases in Asian Aquaculture V, Proceedings of The Fifth Symposium on Diseases in Asian Aquaculture, Gold Coast, Australia. Asian Fisheries Society, Manila.

Huang, C., Zhang, X., Lin, Q., Xu, X., Hu, Z.H. and Hew, C.L. 2002. Proteomic analysis of shrimp white spot syndrome viral proteins and characterization of a novel envelope protein VP466. Molecular and Cellular Proteomics 1, 223-231.

Kogan, P.H. and Blissard, G.W. 1994. A baculovirus gp64 early promoter is activated by host transcription factor binding to CACGTG and GATA elements. Journal of Virology 68, 813-822.

Kool, M., Voeten, J.T.M., Goldbach, R.W., Tramper, J. and Vlak, J.M. 1993. Identification of seven putative origins of *Autographa californica* multiple nucleocapsid nuclear polyhedrosis virus DNA replication. Journal of General Virology 74, 2661-2668.

Li, Q., Yang, F., Zhang, J. and Chen, Y. 2003. Proteomic analysis of proteins that binds specifically to the homologous repeat regions of white spot syndrome virus. Biological and Pharmaceutical Bulletin 26, 1517-22.

Lightner, D.V. 1996. A handbook of pathology and diagnostic procedures for diseases of penaeid shrimp. Special publication of the World Aquaculture Society. Baton Rouge, LA.

Liu, W.J., Yu, H.T., Peng, S.E., Chang, Y.S., Pien, H.W., Lin, C.J., Huang, C.J., Tsai, M.F., Wang, C.H., Lin, J.Y., Lo, C.F. and Kou, G.H. 2001. Cloning, characterization, and phylogenetic analysis of a shrimp white spot syndrome virus gene that encodes a protein kinase. Virology 289, 362-377.

Lo, C.F., Ho, C.H., Peng, S.E., Chen, C.H., Hsu, H.C., Chiu, Y.L., Chang, C.F., Liu, K.F., Su, M.S., Wang, C.H. and Kou, G.H. 1996. White spot syndrome baculovirus (WSBV) detected in cultured and captured shrimp, crabs and other arthropods. Diseases of Aquatic Organisms 27, 215-225.

Marks, H., Mennens, M., Vlak, J.M. and van Hulten, M.C.W. 2 003. Transcriptional analysis of the white spot syndrome virus major virion protein genes. Journal of General Virology 84, 1517-1523.

Marks, H., Goldbach, R.W., Vlak, J.M. and van Hulten, M.C.W. 2004. Genetic variation among isolates of white spot syndrome virus. Archives of Virology 149 (in press).

Morris, T.D. and Miller, L.K. 1994. Mutational analysis of a baculovirus major late promoter. Gene 140, 147-153.

Nadala, E.C.B., Jr., Tapay, L.M. and Loh, P.C. 1998. Characterization of a non-occluded baculovirus-like agent pathogenic to penaeid shrimp. Diseases of Aquatic Organisms 33, 221-229.

Nalçaciog˘lu, R., Marks, H., Vlak, J.M., Demirbag, Z. and van Oers, M.M. 2003. Promoter analysis of the *Chilo* iridescent virus DNA polymerase and major capsid protein genes. Virology 317, 321-329.

Ooi, B.G., Rankin, C. and Miller, L.K. 1989. Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. Journal of Molecular Biology 210, 721-736.

Pearson, M., Bjornson, R., Pearson, G. and Rohrmann, G. 1992. The *Autographa californica* baculovirus genome: evidence for multiple replication origins. Science 257, 1382-1384.

Rosenberry, B. 1996. World Shrimp Farming 1996. Shrimp News International. San Diego, California. pp. 164.

Rosenberry, B. 2000. World Shrimp Farming 2000. Shrimp News International. San Diego, California. pp. 324.

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. 2000a. 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.

Tsai, M.F., Lo, C.F., van Hulten, M.C.W., Tzeng, H.F., Chou, C.M., Huang, C.J., Wang, C.H., Lin, J.Y., Vlak, J.M. and Kou, G.H. 2000b. Transcriptional analysis of the ribonucleotide reductase genes of shrimp white spot syndrome virus. Virology 277, 92-99.

Tzeng, H.F., Chang, Z.F., Peng, S.E., Wang, C.H., Lin, J.Y., Kou, G.H. and Lo, C.F. 2002. Chimeric polypeptide of thymidine kinase and thymidylate kinase of shrimp white spot syndrome virus: Thymidine kinase activity of the recombinant protein expressed in a baculovirus/insect cell system. Virology 299, 248-255.

van Hulten, M.C.W., Westenberg, M., Goodall, S.D. and Vlak, J.M. 2000a. Identification of two major virion protein genes of white spot syndrome virus of shrimp. Virology 266, 227-236.

van Hulten, M.C.W., Goldbach, R.W. and Vlak, J.M. 2000b. Three functionally diverged major structural proteins of white spot syndrome virus evolved by gene duplication. Journal of General Virology 81, 2525-2529.

van Hulten, M.C.W., Tsai, M.F., Schipper, C.A., Lo, C.F., Kou, G.H. and Vlak, J.M. 2000c. Analysis of a genomic segment of white spot syndrome virus of shrimp containing ribonucleotide reductase genes and repeat regions. Journal of General Virology 81, 307-316.

van Hulten, M.C.W. and Vlak, J.M. 2001. Identification and phylogeny of a protein kinase gene of white spot syndrome virus. Virus Genes 22, 201-207.

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

van Hulten, M.C.W., Witteveldt, J., Snippe, M. and Vlak, J.M. 2001b. White spot syndrome virus envelope protein VP28 is involved in the systemic infection of shrimp. Virology 285, 228-233.

van Hulten, M.C.W., Reijns, M., Vermeesch, A.M.G., Zandbergen, F. and Vlak, J.M. 2002. Identification of VP19 and VP15 of white spot syndrome virus (WSSV) and glycosylation of

the WSSV major structural proteins. Journal of General Virology 83, 257-265.

van Hulten, M.C.W. and Vlak, J.M. 2002. Genetic evidence for a unique taxonomic position of white spot syndrome virus of shrimp: genus whispovirus. *In* Lavilla-Pitogo, C.R. and Cruz-Lacierda, E.R. (eds.). Diseases in Asian Aquaculture IV, Fish Health Section, Asian Fisheries Society, Manila. pp. 25-35.

Vlak, J.M., Bonami, J.R., Flegel, T.W., Kou, G.H., Lightner, D.V., Lo, C.F., Loh, P.C. and Walker, P.J. 2004. *Nimaviridae*. In Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U. and Ball, L.A. (eds.). Virus Taxonomy. Eigth Report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press, London. pp. 187-192.

Wang, C.H., Lo, C.F., Leu, J.H., Chou, C.M., Yeh, P.Y., Chou, H.Y., Tung, M.C., Chang, C.F., Su, M.S. and Kou, G.H. 1995. Purification and genomic analysis of baculovirus associated with white spot syndrome (WSBV) of *Penaeus monodon*. Diseases of Aquatic Organisms 23, 239-242.

Wang, Q., Nunan, L.M. and Lightner, D.V. 2000. Identification of genomic variations among geographic isolates of white spot syndrome virus using restriction analysis and Southern blot hybridization. Diseases of Aquatic Organisms 43, 175-181.

Weir, J.P. 2001. Regulation of herpes simplex virus gene expression. Gene 271, 117-130.

Witteveldt, J., van Hulten, M.C.W. and Vlak, J.M. 2001. Identification and phylogeny of a non-specific endonuclease gene of white spot syndrome virus of shrimp. Virus Genes 23, 331-337.

Witteveldt, J., Jolink, M., Cifuentes, C.E., Vlak, J.M. and van Hulten, M.C.W. 2005. Vaccination of *Penaeus monodon* against white spot syndrome virus using structural virion proteins. *In* Walker, P.J., Reantaso, M.B. and Lester, R.J. (eds.). Diseases in Asian Aquaculture V, Proceedings of the Fifth Symposium on Diseases in Asian Aquaculture, Gold Coast, Australia. Asian Fisheries Society, Manila.

Witteveldt, J., Vlak, J.M. and van Hulten, M.C. 2004a. Protection of *Penaeus monodon* against white spot syndrome virus using a WSSV subunit vaccine. Fish & Shellfish Immunology 16: in press.

Witteveldt, J., Cifuentes, C.C., Vlak, J.M. and van Hulten, M.C. 2004b. Protection of *Penaeus monodon* against white spot syndrome virus by oral vaccination. Journal of Virology 78, 2057-2061.

Wongteerasupaya, C., Vickers, J.E., Sriurairatana, S., Nash, G.L., Akarajamorn, A., Boonsaeng, V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B. and Flegel, T.W. 1995. A non-occluded, systemic baculovirus that occurs in cells of ectodermal and mesodermal origin and causes high mortality in the black tiger prawn *Penaeus monodon*. Diseases of Aquatic Organisms 21, 69-77.

Wongteerasupaya, C., Pungchai, P., Withyachumnarnkul, B., Boonsaeng, V., Panyim, S., Flegel, T.W. and Walker, P.J. 2003. High variation in repetitive DNA fragment length for white spot syndrome virus (WSSV) isolates in Thailand. Diseases of Aquatic Organisms 54, 253-257.

Yang, F., He, J., Lin, X., Li, Q., Pan, D., Zhang, X. and Xu, X. 2001. Complete genome sequence of the shrimp white spot bacilliform virus. Journal of Virology 75, 11811-11820.