Genomics and transcriptomics of White spot syndrome virus

Hendrik Marks

Promotoren:

Prof. dr. J. M. Vlak

Persoonlijk hoogleraar bij de leerstoelgroep Virologie

Prof. dr. R. W. Goldbach

Hoogleraar in de Virologie

Co-promotor:

Dr. ir. M. C. W. van Hulten

Wetenschappelijk medewerker bij de Commonwealth Scientific and Industrial Research Organisation (Brisbane, Australië)

Samenstelling promotiecommissie:

Prof. Dr. H. J. Nauwynck (Universiteit Gent, België)

Prof. Dr. V. E. J. C. Schijns (Intervet International B.V., Boxmeer)

Dr. O. L. M. Haenen (Centraal Instituut voor DierziekteControle, Lelystad)

Prof. Dr. W. J. Stiekema (Plant Research International, Wageningen)

Genomics and transcriptomics of White spot syndrome virus

Hendrik Marks

Proefschrift
ter verkrijging van de graad van doctor
op het gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. M.J. Kropff,
in het openbaar te verdedigen
op maandag 28 november 2005
des namiddags te 13.30 uur in de Aula

"..., mijn moeder was verbonden aan een dagverblijf voor psychiatrische patiënten. Haar beroemdste patiënt was een man die tegen een muur praatte in de veronderstelling dat de geheime dienst hem kon verstaan. Toen de geheime dienst niet op zijn berichten inging, werd hij destructief. Zo belandde hij bij mijn moeder. Op deze man is ze gepromoveerd. Dankzij mijn moeder kwam hij erachter dat hij twintig jaar voor niets tegen een muur had gepraat. Deze wetenschap was te veel voor hem en hij stortte zich in een liftschacht. Dat wierp een smet op het promotiefeestje van mijn moeder, want het was de bedoeling dat hij daar de eregast zou zijn. Na dit incident heeft ze een paar artikelen geschreven waarin ze zich afvroeg of het wel verstandig is mensen volledig te genezen van hun waanideeën."
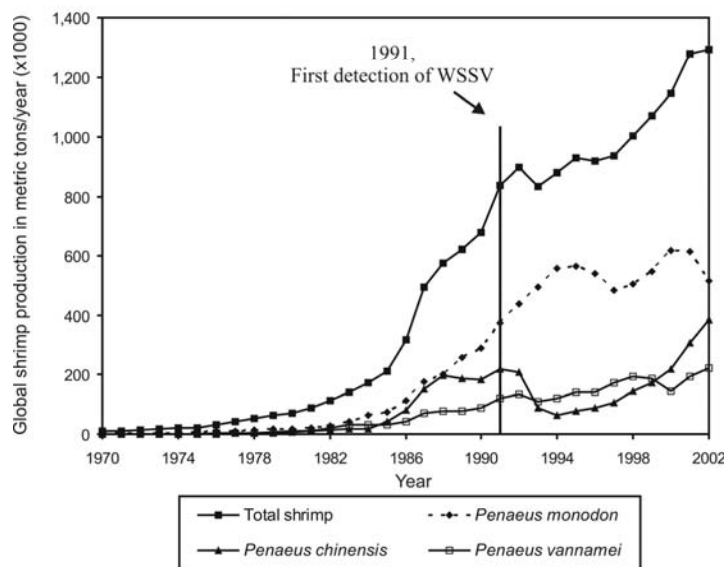
Arnon Grunberg in "Fantoompijn" (2000)

Contents

Chapter 1

General Introduction

Shrimp culture

Shrimp appear in many sea food dishes and are considered a delicacy in countries around the world. To meet consumers' demands, and stimulated by advances in shrimp reproduction and hatchery technologies, an enormous growth in shrimp production was observed from the

Figure 1.1. Shrimp production from 1970 to 2002 (source: Food and Agriculture Organization of the United Nations (FAO); FAO website: http://www.fao.org/fi/figis/).

1980s onwards (Fig. 1.1). Shrimp culture developed from extensive in the 1970s (50-500 kg ha^{-1} yr^{-1}) using tidal zones to super-intensive in the 2000s (20,000-100,000 kg ha^{-1} yr^{-1}) using

Figure 1.2. Drawing of *P. monodon* shrimp (source: FAO).

ponds more inland (Rosenberry, 2004). Traditionally, the majority of the shrimp farming takes place in Asia (especially in China, Thailand, Vietnam, Indonesia and India), where it is an important source of income for people in coastal areas. Nowadays also South- and Latin-America (predominantly Mexico, Ecuador and Brazil) are responsible for a considerable amount of the worldwide production of shrimp. Furthermore, the shrimp farming industry is increasing in other areas in the world such as Australia and the

Middle East (Saudi Arabia and Iran) (Rosenberry, 2004).

All farmed shrimp species belong to the *Penaeidae* family of decapod crustaceans (genus *Penaeus*). The largest, most expensive and commercially most important species is *Penaeus monodon* (giant black tiger shrimp) (Fig. 1.2), while other important cultured species include *P. chinensis* (Chinese white shrimp) and *P. vannamei* (Western white shrimp). *P. monodon* is named giant black tiger shrimp after its striped tail and its big size (maximum length of 363 millimeters). This species originates from the Indian Ocean and the southwestern Pacific Ocean (Japan to Australia) and is mainly cultured in Asia (Rosenberry, 2004). The species *P. vannamei* originates from the Pacific coast of Central- and South-America (from Mexico to Peru) and is the leading farm-raised species in South- and Latin-America. Because *P. vannamei* feeds on organisms which grow naturally in a pond, it is cheaper to feed than *P. monodon*. *P. vannamei* has a uniform growth rate and reaches a maximum length of 230 millimeters (Rosenberry, 2004). *P. chinensis* is native to the coast of China and the west coast of the Korean peninsula. The disadvantages of culturing *P. chinensis* for production for economy are their high protein requirement and their small size (maximum length of 183 millimeters). *P. vannamei* and *P. chinensis*, unlike *P. monodon*, readily mature and spawn in ponds, and are therefore easy to culture (Rosenberry, 2004).

Shrimp diseases

In the 1990s, the annual growth of shrimp production decelerated reaching a plateau level of around 900,000 tons per year (Fig. 1.1: 1991-1997; FAO), even though the total area used for shrimp culturing increased. This relative decline was due to mass mortalities in shrimp ponds, predominantly caused by viruses (Lotz, 1997). Whereas the various non-viral infectious agents, such as bacteria, fungi and protozoa, can be controlled by antibiotics and vaccination (Teunissen *et al.*, 1998), efficient control of viral pathogens in the field has not been achieved thus far. More than 20 viruses have been reported to infect shrimp (Lightner, 1996) and the list is growing. The economical most damaging shrimp viruses are White spot syndrome virus (WSSV), Yellow head virus (YHV) and Taura syndrome virus (TSV), which are discussed below. The diseases caused by these viruses are notifiable to the Office International des Epizooties (OIE; World Organisation for Animal Health).

The next paragraphs will focus on the large double-stranded DNA virus WSSV, the subject of this thesis. YHV is a positive single stranded RNA virus and it is, together with the closely related shrimp virus Gill-associated virus (GAV), classified in a new taxon (family *Roniviridae*, genus *Okavirus*) within the order *Nidovirales* (Cowley *et al.*, 1999; Jitrapakdee *et al.*, 2003). Signs associated with YHV infection include cessation of feeding, swimming near the water surface or near edges of a pond, and the development of a yellow coloration of the cephalothorax and gills (Chantanacookin *et al.*, 1993). TSV is a positive single stranded RNA virus belonging to the virusfamily *Dicistroviridae* (Mari *et al.*, 2002). Gross signs associated with TSV infection include the appearance of a distinct blue or red hue on the shell

and tail. Infected shrimp usually have empty digestive tracts (Lightner, 1996). For a recent review on YHV and TSV see Dhar *et al.* (2004).

As *P. monodon* is very susceptible to WSSV and YHV, the emergence of these two viruses in the 1990s was one of the main causes of the reduction in productivity. Therefore, farmers in Southeast-Asia often switched from *P. monodon* to culture *P. vannamei* or *P. chinensis*, as these species are considered to be less vulnerable for these viruses (Fig. 1.1) (Briggs *et al.*, 2004; Rosenberry, 2004; Wyban and Sweeney, 1991). Moreover, the breeding techniques for these species are further developed than for *P. monodon* (Rosenberry, 2004), enabling selection of virus resistant shrimp. However, especially *P. vannamei* is very sensitive to TSV (Lightner *et al.*, 1995).

White spot syndrome virus
Of all shrimp viruses, White spot syndrome virus (WSSV) has had the largest impact on shrimp culture and remains a major problem up to the present day (Rosenberry, 2004). In cultured shrimp, WSSV infection can reach a cumulative mortality of up to 100% within 3-10 days (Lightner, 1996). Infected animals show lethargic behavior, such as lack of appetite and slow movement, and a reddish to pink body discoloration. Characteristic for WSSV infected shrimp are white spots on the exoskeleton (Fig. 1.3). These spots are the result of calcified deposits that range in size from a few mm to 1 cm or more in diameter (Chou *et al.*, 1995). However, in case of acute (experimental) infections the only signs of WSSV infection observed are lethargy and lack of appetite.
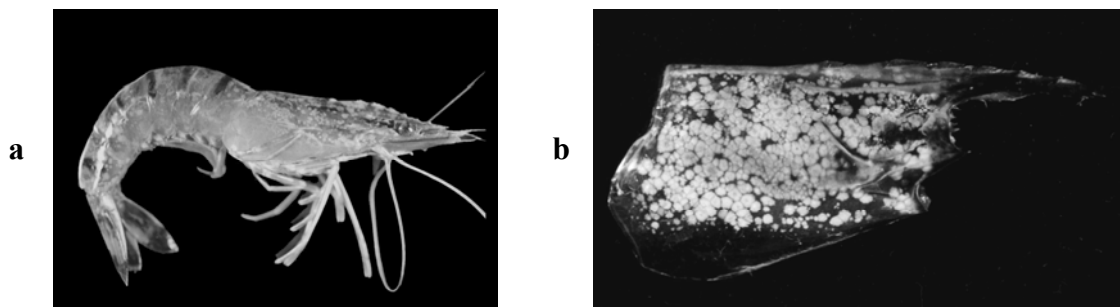
Figure 1.3. Photograph of a WSSV infected *P. monodon* showing white spots on the exoskeleton **(a)**. Close-up of the carapace of a WSSV infected *P. monodon*, emphasizing the white spots **(b)**. (Source: C.F. Lo, National Taiwan University, Taiwan)

WSSV can be transmitted horizontally either *per os* by predation on diseased individuals, or by virus particles in the water. Infection by the latter is thought to occur primarily through the gills, but may occur via other body surfaces as well (Chang *et al.*, 1996; Chou *et al.*, 1995; 1998). The virus is also transmitted from mother to offspring, although it is not clear whether the WSSV virions are present inside the shrimp eggs (Hsu *et al.*, 1999; Lo *et al.*, 1997; Peng *et al.*, 2001; Tsai *et al.*, 1999). No penaeid shrimp species is known to be resistant to WSSV infection (Lightner, 1996; Lotz, 1997).

WSSV was first reported in the Chinese province Fujian in 1991. Due to intensive shrimp cultivation, inadequate sanitation and (worldwide) trade, it quickly spread to other shrimp farming areas in Southeast-Asia (Cai *et al.*, 1995; Flegel, 1997; Lightner, 1996). Although the virus initially appeared to be limited to Asia, in November 1995 it was detected in the United States of America (Texas and South-Carolina) (Rosenberry, 1996). In early 1999, WSSV was reported in Central- and South-America (Rosenberry, 2000). In 2002, WSSV was also detected in Europe (France) and the Middle East (Iran) (Rosenberry, 2002). Although WSSV is the approved name for the virus nowadays (Vlak *et al.*, 2005), other names have been used in the past for the same virus, 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). The disease caused by WSSV is often indicated as "White spot disease" (WSD).

Besides shrimp, WSSV infects a wide range of other aquatic crustaceans such as salt, brackish and fresh water crayfishes, crabs and lobsters. In contrast to shrimp, the infection is often not lethal for these species, and therefore they may serve as reservoirs and carriers of the virus (Lo *et al.*, 1996a; Wang *et al.*, 1998). Because of its broad host range, WSSV is not only a major threat to shrimp farming, but also to the worldwide marine ecology (Flegel, 1997). Therefore, there is a strong demand for the development of WSSV control methods. Sanitation of shrimp cultures, preventive measures (PCR tests of broodstock, fry and postlarvae) and chain management are most important. Recent studies indicate the possibility of (passive) immunization of shrimp against WSSV using the WSSV structural proteins VP19 or VP28 (Witteveldt *et al.*, 2004a,b). Especially the protection of shrimp against WSSV after oral administration of the WSSV envelope protein VP28 may benefit the shrimp farming industry, as this may result in the development of a commercial vaccine (Witteveldt *et al.*, 2004a).

WSSV histopathology
Histopathological studies on WSSV infected shrimp show that the prime targets for viral replication are mainly tissues of ectodermal and mesodermal origin (Momoyama *et al.*, 1994; Lo *et al.*, 1997; Wang *et al.*, 1995; Wongteerasupaya *et al.*, 1995). Early in infection, the stomach, gills, cuticular epidermis and the connective tissue of the hepatopancreas are WSSV positive. At later stages of infection the lymphoid organ, antennal gland, muscle tissue, hematopoietic tissue, heart, hindgut and parts of the midgut also become positive for WSSV. The nervous system and the compound eyes are only infected very late in infection. The stomach, gills, cuticular epidermis, lymphoid organ, hematopoietic tissue and antennal gland are all heavily infected with WSSV at late stages of infection and become necrotic (Chang *et al.*, 1996; Lo *et al.*, 1997). Cells of the hepatopancreatocytes and epithelial cells of the midgut have never shown to be infected with WSSV.

A significant reduction in the total hemocyte count is observed after shrimp are infected with WSSV (Hennig *et al.*, 1998; Jiravanichpaisal *et al.*, 2001; van de Braak *et al.*, 2002). This is probably caused by infection of the hemocytes themselves as well as by apoptosis in the WSSV infected hematopoetic tissue, from which the hemocytes derive (Wongprasert *et al.*, 2003). WSSV infected hemocytes either undergo apoptosis or are removed from the circulation by attaching to host tissues. As a consequence abnormal low amount of hemocytes will weaken the shrimp defenses.

WSSV cytopathology

WSSV replication and virion assembly occur in the nucleus and early signs of infection are characterized by the appearance of homogeneous hypertrophied nuclei and chromatin margination (Fig. 1.4) (Lightner, 1996; Wang *et al.*, 1999b; Wongteerasupaya *et al.*, 1995).

Virus morphogenesis is initiated by the *de novo* formation of viral envelopes in the nucleoplasm. The formation of the nucleocapsids begins with extended, empty, long tubules, which break up into fragments of 12 to 14 rings to form empty nucleocapsid shells. Subsequently, the empty capsids are surrounded by the envelope leaving at one end an open extremity. The nucleoproteins, possibly together with the viral DNA, enter the empty capsid through its open end. Mature virions are obtained after narrowing of the open end and formation of a tail-like extension of the envelope (Durand *et al.*, 1997; Wang *et al.*, 1999b; 2000a). It is not clear how the virions are released from the nucleus of an infected cell, most likely by budding or by rupture of the nuclear envelope and/or the cell membrane.

Figure 1.4. Electron micrograph picture of a WSSV infected cell in the gill of the crayfish *Procambarus clarkii*, showing a hypertrophied nucleus containing WSSV virus particles (bar represents 1 μm).

WSSV morphology

Electron microscopical studies on thin sections and viral suspensions obtained from infected shrimp revealed that the virion of WSSV is a large, ovoidal particle of about 275 nm in length and 120 nm in width, with a tail-like appendage at one end (Fig. 1.5a) (Durand *et al.*, 1997). So far, the function nor the composition of this appendage is known.

The virion consists of a rod-shaped nucleocapsid with a tight-fitting capsid layer, surrounded by a loose-fitting trilaminar envelope, which consists mainly of the WSSV encoded proteins VP28 and VP19 (Fig. 1.5b; Fig. 1.6) (Durand *et al.*, 1997; Nadala *et al.*, 1998; van Hulten *et al.*, 2000a,c). VP28 is most likely located on the surface of the virus particle and plays a key role in WSSV infection (van Hulten *et al.*, 2001b). Isolated nucleocapsids have a cross-

Figure 1.5. Negative contrast electron micrographs of an enveloped WSSV virion with its characteristic tail-like extension **(a)**, a WSSV nucleocapsid with residual envelope **(b)** and WSSV nucleocapsids **(c)**.
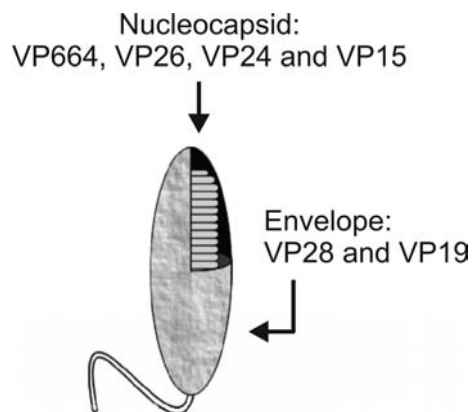
Figure 1.6. Simplified model of the WSSV virion.

hatched appearance and a size of about 300 x 70 nm (Fig. 1.5c; Fig. 1.6). The nucleocapsid is formed by stacks of rings (about 14 in total), which are in turn formed by regular spaced globular subunits of about 8 nm in diameter, arranged in two parallel rows (Durand *et al.*, 1997; Nadala *et al.*, 1998). The nucleocapsid contains the viral genome and mainly consists of the WSSV encoded proteins VP664, VP26, VP24 and VP15 (Leu *et al.*, 2005; van Hulten *et al.*, 2000a,c; 2002). VP664, a remarkable large protein of around 664 kDa, is thought to be the major core protein, responsible for the striated appearance of the nucleocapsids (Leu *et al.*, 2005). VP15, a highly basic protein with no hydrophobic regions, is a histon-like, double-stranded DNA-binding protein (Witteveldt *et al.*, 2005). The function of VP26 and VP24 in the nucleocapsid is unknown. Furthermore, about 40 WSSV encoded minor virion proteins were identified by protein sequencing of individual bands after applying purified WSSV virions on a SDS-PAGE gel (Huang *et al.*, 2002b; Leu *et al.*, 2005; Tsai *et al.*, 2004; van Hulten *et al.*, 2000a,c; 2002; Zhang *et al.*, 2004). The various geographic isolates of WSSV, which have been characterized, are very similar if not identical in morphology and proteome (Nadala and Loh, 1998; Wang *et al.*, 2000c).

WSSV genome

The virions of WSSV contain a circular, supercoiled, double-stranded (ds) DNA genome, originally estimated to be ~300 kilobase pairs (kb). Between different geographic isolates of WSSV a few restriction fragment length polymorphisms (RFLPs) were reported, indicating the presence of some genomic variation (Lo *et al.*, 1999; Nadala and Loh, 1998; Wang *et al.*, 2000b).

The genome of a WSSV isolate originating from Thailand (WSSV-TH) was the first to be completely sequenced and consisted of 292,967 base pairs (bp) (van Hulten *et al.*, 2001a). The adenine residue of the translational start codon of the major envelope protein VP28 was designated as starting point of the circular physical map of the WSSV genome (Fig. 1.7). Computer-assisted analysis of the WSSV-TH genome identified 184 putative ORFs
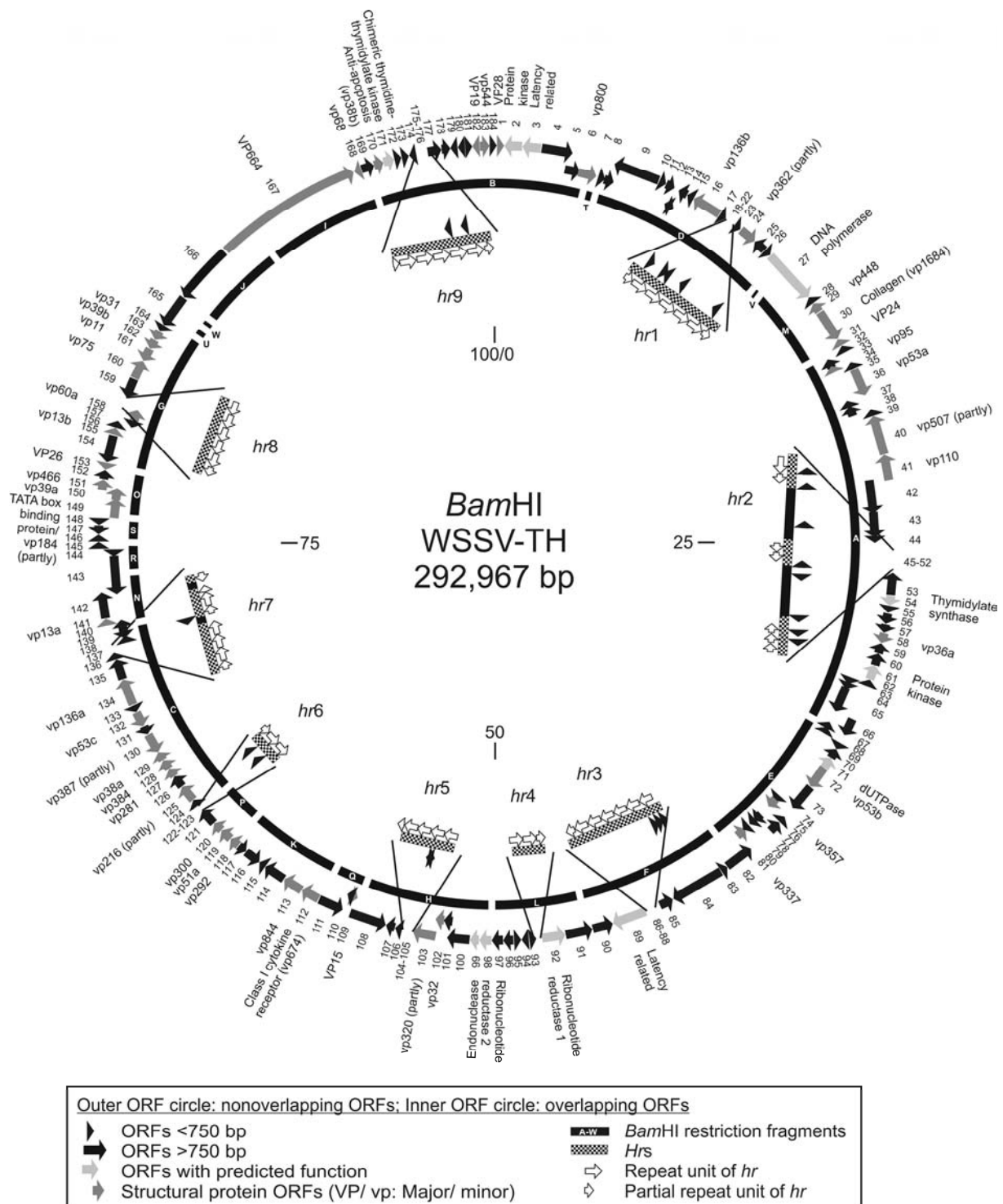
Figure 1.7. Map of the circular ds DNA WSSV-TH genome showing the genomic organization. Sites for *Bam*H1 are shown in the inner circle. Fragments are numbered from A to W according to size from the largest (A) to the smallest (W). The positions of the ORFs are indicated by arrows, which also represent the direction of transcription. Dark gray arrows represent the structural protein genes. Light gray arrows indicate ORFs with homologues in public databases. The *hr* sequences (dashed blocks) are enlarged within the inner circle, with their repeat units (open arrows) indicated. The scale on the inner circle is in map units.

Table 1.1. Overview of WSSV ORFs with known or predicted functions, ordered by WSSV-TH ORF numbering (WSSV-TH ORF numbering according to van Hulten *et al.* (2001a)).

WSSV-TH ORF	Homology to GenBank	References
ORF2	Protein Kinase *,†	van Hulten and Vlak, 2001; Liu *et al.*, 2001
ORF9	Helicase †	
ORF27	DNA polymerase *,†	Chen *et al.*, 2002b
ORF30	Collagen *,†	Li *et al.*, 2004a
ORF54	Thymidylate synthase *,†	Li *et al.*, 2004b
ORF61	Protein Kinase *,†	van Hulten and Vlak, 2001
ORF66	CREB-binding protein (CBP) †	
ORF71	dUTPase *,†	Liu and Fang, 2005
ORF92	Ribonucleotide reductase I (large subunit; rr1) *,†	van Hulten *et al.*, 2000b; Tsai *et al.*, 2000a; Lin *et al.*, 2002
ORF98	Ribonucleotide reductase II (small subunit; rr2) *,†	van Hulten *et al.*, 2000b; Tsai *et al.*, 2000a; Lin *et al.*, 2002
ORF99	Endonuclease *,†	Witteveldt *et al.*, 2001; Li *et al.*, 2005a
ORF112	Class I cytokine receptor *	
ORF143	Protein Kinase †	
ORF149	TATA box binding protein (TBP) *,†	
ORF171	Chimeric Thymidine kinase-thymidylate kinase *,†	Tsai *et al.*, 2000b; Tzeng *et al.*, 2002
ORF TW 410 (not present in WSSV-TH and WSSV-CN)	Transposase	Acc. no. AF440570

WSSV-TH ORF	Virion protein genes	References
Major virion protein genes		
ORF1	VP28 (p204) *,†,‡,§,¶	van Hulten *et al.*, 2000c; Zhang *et al.*, 2002a
ORF31	VP24 (vp208) *,†,‡,§,¶	van Hulten *et al.*, 2000a
ORF109	VP15 (p6.8) *	Zhang *et al.*, 2001; van Hulten *et al.*, 2002; Witteveldt *et al.*, 2005
ORF153	VP26 (p22) *,†,‡,§	van Hulten *et al.*, 2000c; Zhang *et al.*, 2002b
ORF167	VP664 §	Leu *et al.*, 2005
ORF182	VP19 (vp121) *,‡,§	van Hulten *et al.*, 2002
Minor virion protein genes		
ORF6	vp800 (vp95) ‡,§	
ORF16	vp136b §	
ORF24 #	vp362 ¶	
ORF29	vp448 ‡	
ORF30	vp1684 (collagen) ‡	
ORF34	vp95 ‡	
ORF36 #	vp53a §	
ORF40 #	vp507 ¶	
ORF41	vp110 §	
ORF58	vp36a §	
ORF72 #	vp53b §	
ORF75	vp357 ‡	
ORF81	vp337 ¶	
ORF102	vp32 §	
ORF103 #	vp320 ¶	
ORF112	vp674 (class I cytokine receptor) (vp73) ‡,§	Huang *et al.*, 2005
ORF113	vp844 ¶	
ORF118	vp292 (vp41a) ‡,§	Zhang *et al.*, 2004
ORF119	vp51a §	
ORF120	vp300 (vp41b) ‡,§	
ORF125 #	vp216 ¶	
ORF127	vp281 (vp36b) ‡,§,¶	Huang *et al.*, 2002a; Wu *et al.*, 2005

WSSV-TH ORF	Virion protein genes	References
ORF128	vp384 (vp51b) ‡,§	
ORF129	vp38a §	
ORF130 #	vp387	
ORF132	vp53c §	
ORF134	vp136a §	
ORF141	vp13a §	
ORF149 #	vp184 (TATA box binding protein) ‡	
ORF150	vp39a §	
ORF151	vp466 (vp51c) ‡,§	Huang *et al.*, 2002b; Wu *et al.*, 2005
ORF155	vp13b §	
ORF158	vp60a §	
ORF160	vp75 §	
ORF161 #	vp11 §	
ORF162	vp39b §	
ORF163	vp31 §	Li *et al.*, 2005b
ORF168	vp68 (vp12b)‡,§,¶	Zhang *et al.*, 2004; Wu *et al.*, 2005
ORF170	vp38b §	
ORF183	vp544 (vp60b) ‡,§,¶	
ORF CN 493; ORF TW 19 (not present in WSSV-TH)	vp35	Chen *et al.*, 2002a

WSSV-TH ORF	Empirical established gene functions/ properties	References
ORF 3 (ORF CN 427)	Latency related gene	Khadijah *et al.*, 2003; Lu and Kwang, 2004
ORF55 (ORF TW 126)	Immediate early gene 1 (ie1)	Liu *et al.*, 2005
ORF 89 (ORF CN 151)	Latency related gene	Khadijah *et al.*, 2003; Hossain *et al.*, 2004
ORF170 (ORF CN 390)	Anti-apoptosis gene	Wang *et al.*, 2004
ORF CN 366; ORF TW 425 (not annotated in WSSV-TH)	Latency related gene	Khadijah *et al.*, 2003
ORF TW 242, ORF CN 187 (not annotated in WSSV-TH)	Immediate early gene 2 (ie2)	Liu *et al.*, 2005
ORF TW 418, ORF CN 359 (not annotated in WSSV-TH)	Immediate early gene 3 (ie3)	Liu *et al.*, 2005

* annotated by van Hulten *et al.*, 2001a
† annotated by Yang *et al.*, 2001
‡ protein sequenced from virions by Huang *et al.*, 2002b
§ protein sequenced from virions by Tsai *et al.*, 2004
¶ protein sequenced from virions by Zhang *et al.*, 2004
# the size of the proteins as determined by SDS PAGE gels is considerably smaller than the size of the proteins encoded by the WSSV-TH ORFs

encoding proteins >50 amino acids (aa), these proteins ranging in size from 50 to 6077 aa (Fig. 1.7). Among these ORFs, ten gene families consisting of two to four ORFs with pairwise similarities of 40% or higher, were identified. Based on homologies with other genes in GenBank, only 12 of the 184 WSSV ORFs could be assigned a putative function (Table 1.1). The majority of these known genes encode enzymes involved in DNA replication, nucleotide metabolism and protein modification (Table 1.1). Besides these ORFs, the genes for the six major (VP664, VP28, VP26, VP24, VP19 and VP15) and about 40 minor structural virion proteins have been identified on the genome (Fig. 1.7; Table 1.1) (Huang *et al.*, 2002b; Leu *et al.*, 2005; Tsai *et al.*, 2004; van Hulten *et al.*, 2000a,c; 2002; Zhang *et al.*, 2004). Recently, three WSSV ORFs have been suggested to be involved in WSSV latency (Khadijah

et al., 2003) and ORF170 was shown to encode an anti-apoptosis protein (Wang *et al.*, 2004) (Table 1.1). However, most of the WSSV ORFs are still unassigned.

The WSSV genome is further characterized by the presence of nine direct repeat regions with different sizes, designated homologous region (*hr*) 1 to 9 (Fig. 1.7). These *hr*s are dispersed throughout the WSSV genome and consist of three to eight identical repeat units of 250 bp or parts thereof. The *hr*s are largely located in intergenic regions, although several short ORFs are annotated within the WSSV *hr*s (Fig. 1.7) (van Hulten *et al.*, 2001a). Similar but (slightly) smaller repeat regions have been identified in ascoviruses and baculoviruses (Bigot *et al.*, 2000; Cochran and Faulkner, 1983). For baculoviruses it was demonstrated that the *hr*s function as enhancers of transcription and origins of DNA replication (Guarino and Summers, 1986; Kool *et al.*, 1993; Pearson *et al.*, 1992). The *hr*s of WSSV could have a similar function. Based on the observation that most large DNA viruses with a circular genome configuration (i.e. baculoviruses, ascoviruses and WSSV) contain *hr*s, their presence may imply that these viruses share a similar replication strategy.

Except for WSSV-TH, the genomes of two WSSV isolates originating from Taiwan (WSSV-TW) and China (WSSV-CN) have also been completely sequenced (GenBank acc. no. AF440570; Yang *et al.*, 2001). Isolate WSSV-TW contained a genome with a size of 307,287 bp, while the genome size of WSSV-CN was 305,107 bp. A comparative analysis between the three completely sequenced WSSV genomes is presented in Chapter 2.

WSSV taxonomy

The taxonomic position of WSSV has been evaluated using phylogenetic analysis on several genes identified on the WSSV genome, including the ribonucleotide reductase large and small subunits (Tsai *et al.*, 2000a; van Hulten *et al.*, 2000b), the protein kinases (Liu *et al.*, 2001; van Hulten and Vlak, 2001), the endonuclease (Witteveldt *et al.*, 2001), the chimeric thymidine-thymidylate kinase (Tsai *et al.*, 2000b) and the DNA polymerase (Chen *et al.*, 2002b; van Hulten *et al.*, 2001a). DNA polymerases are most frequently used to study the phylogenetic relatedness of viruses, as these enzymes are commonly present and are one of the very few gene products which have sequence motifs conserved among all large DNA viruses and host organisms (Hannenhali *et al.*, 1995; Knopf, 1998; Tidona and Darai, 2000). The DNA polymerases of the different virus families used for construction of an unrooted tree (Fig. 1.8) were all present in clades which are well bootstrap supported. No recent relationship was revealed for WSSV with members of the established large ds DNA virus families (van Hulten *et al.*, 2001a). Together with the phylogenetic data of the other WSSV genes (data not shown), this tree confirms the unique taxonomic position of WSSV.

Based on these extensive phylogenetic analyses, but also on the primary genomic structure and composition as well as the distinct morphology of the virion, the International Committee on Taxonomy for Viruses (ICTV) approved a proposal in 2002 to accommodate WSSV in a new virus family called *Nimaviridae*, referring to the thread-like polar extension
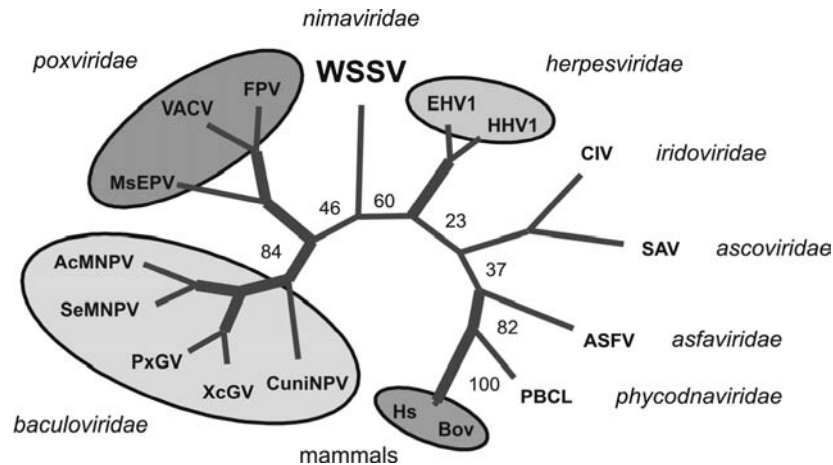
Figure 1.8. Bootstrap analysis (100 replicates) of an unrooted phylogenetic tree of DNA polymerase proteins constructed with the PAUP heuristic search algorithm (van Hulten *et al.*, 2001a). Thick lines indicate frequency of clusters over 70% (Bootstrap value >70). DNA polymerase genes used and their 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).
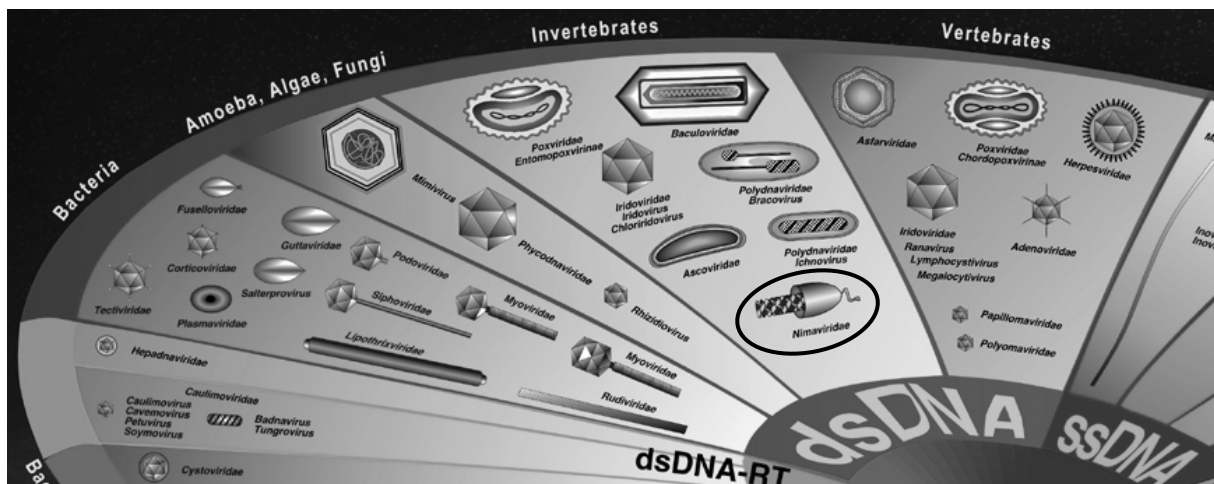
Figure 1.9. Part of the "Virosphere 2005" showing all dsDNA virus families currently approved by the International Committee on Taxonomy for Viruses (ICTV), including the virus family *Nimaviridae* (encircled) (Source: 8th report of the ICTV).

11

Table 1.2. Properties of large (>100 kb) ds DNA virus families infecting (in)vertebrates, algae or amoebae (8th report of the ICTV; data mostly obtained from the ICTV website: http://www.ncbi.nlm.nih.gov/ICTVdb/).

Virusfamily	Host	(one of the) Type species (genus)	Presence of an envelope*	Approximate size virion (nm)*	Approximate size nucleocapsid type species (nm)*	Virion morphology*	Nucleocapsid morphology*	Site of replication	Genome configuration	Complete genome size type species (~kb)	Range of genome sizes of family (~kb)
Nimaviridae	invertebrates (crustaceans)	WSSV (Whispovirus)	+	275 x 120	300 x 70	bacilliform with tail	rod-shaped	nucleus	1 circular	300	290-310
Baculoviridae	invertebrates (insects)	*AcMNPV* (Nucleopoly hedrovirus)	+	400 x 200	275 x 45	bacilliform	rod-shaped	nucleus	1 circular	135	80-180
Ascoviridae	invertebrates (insects)	SfAV-1a (Ascovirus)	+	350 x 125	350 x 125	bacilliform	bacilliform	nucleus	1 circular	140	115-185
Polydna-viridae	invertebrates (wasps/larvae)	CsIV (Ichnovirus)	+	330 x 85	330 x 85	drop-shaped	cylindrical	nucleus	multiple circular	250	130-570
Iridoviridae	invertebrates/vertebrates	CIV (or IIV-6) (Iridovirus)	-	see nucleocapsid	130 (diameter)	spherical	isometric	cytoplasm	1 linear	210	140-280
Poxviridae	invertebrates/vertebrates	VACV (Orthopox-virus)	+	350 x 200	300 x 150	pleomorphic	brick-shaped	cytoplasm	1 linear	190	130-375
Herpesviridae	vertebrates	HHV-1 (Simplexvirus)	+	160 (diameter)	105 (diameter)	spherical	icosahedral	nucleus	1 linear	150	125-240
Asfarviridae	vertebrates (transmitted by arthropods)	ASFV (Asfivirus)	+	250 (diameter)	80 (diameter)	spherical	isometric	cytoplasm	1 linear	170	170-190
Phycodna-viridae	algae	PBCV-1 (Chlorovirus)	-	see nucleocapsid	180 (diameter)	isometric	isometric	cytoplasm	1 linear	330	160-380
Mimiviridae	amoebae	*Acanthamoeba polyphaga* mimivirus	-	see nucleocapsid	400 (diameter)	isometric	isometric	cytoplasm	1 linear	1,200	1,200

on the virus particle (Nima: Latin for thread) (Fig. 1.9). This virus family consists of a single genus (*Whispovirus*) and contains White spot syndrome virus I (WSSV) as its sole species so far (Mayo, 2002; Vlak *et al.*, 2005). Probably all WSSV isolates identified thus far are variants of the same species.

Large ds DNA viruses

Over the last years, a vast amount of information has become available on other eukaryotic large ds DNA viruses (arbitrarily set at a genome size of >100 kb), predominantly on genomics and transcriptomics. This allows a comparative analysis useful for studying WSSV. To provide a brief overview on the characteristics of these viruses, some important features are summarized in Table 1.2. Eukaryotic large ds DNA viruses form a diverse group of pathogens with a wide variety in host range, virulence, virion morphology and genome complexity, infecting vertebrates, invertebrates and algae (Table 1.2). Their genomes generally contain tightly packed, predominantly non-overlapping ORFs, which usually lack introns. In general, there is no correlation between the transcription polarity or function of a gene and its position on the genome (Hannenhali *et al.*, 1995; Hayakawa *et al.*, 2000; IJkel *et al.*, 2001). This is in contrast to regulation in bacterial genomes, where functionally-related genes tend to cluster along the genome (Huynen *et al.*, 2000).

The viral ORFs are preceded by virus-specific promoters that temporally regulate transcription of an early and a late class of genes in a cascaded manner. Nucleotide motifs, generally present within 100 to 200 nucleotides upstream of the translation initiation codon (AUG) of the viral genes, are often involved in the cascaded gene regulation and the kinetics of expression (Broyles, 2003; Davison and Moss, 1989; Friesen *et al.*, 1997; Gaffney *et al.*, 1985; Garcia-Escudero and Viñuela, 2000; Kim *et al.*, 2002; Lu and Miller, 1997; Morris and Miller, 1994; Rajčáni *et al.*, 2004; Weir, 2001; Almazán *et al.*, 1992). The early genes are expressed prior to DNA replication and encode many non-structural proteins, including enzymes involved in genome replication, viral transcription initiation and host response modulation. Viruses replicating in the nucleus generally utilize the host's transcriptional apparatus (host RNA polymerase II) for early gene expression (Friesen *et al.*, 1997; Wagner *et al.*, 1995). However, viruses which are replicating in the cytoplasm (Table 1.2) probably all use their own transcriptional enzymes for early gene expression, as has been shown for at least pox- and asfarviruses. For these virus families, the enzymes involved in transcription are present in the virus particle to enable initiation of gene expression (Pena *et al.*, 1993; Broyles, 2003). Late genes are expressed after viral replication and mainly encode structural virion proteins. These genes are generally expressed at (very) high levels compared to cellular and early viral genes. For late gene expression, most viruses encode their own transcriptional enzymes, such as a viral RNA polymerase and transcription regulator genes (Broyles, 2003; Garcia-Escudero and Viñuela, 2000; Lu and Miller, 1997). However, herpesviruses continue exploiting the cellular RNA polymerase II system for late expression (Wagner *et al.*, 1995).

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). Information on transcription regulation of WSSV is very limited and the rationale for this thesis work.

Outline of the thesis

At the onset of this thesis (2001), complete genome sequences of WSSV isolates originating from Taiwan, China and Thailand (WSSV-TW, WSSV-CN and WSSV-TH, respectively) just became available (GenBank acc. no. AF440570; Yang *et al.*, 2001; van Hulten *et al.*, 2001a). The WSSV genome sequence was taken as a starting-point to address fundamental questions as to how WSSV evolved over time and space and how gene expression of this unique virus is regulated. This information may provide a (genetic) basis to protect shrimp from disease. Chapters 2 through 4 focus on the genomic variation and epidemiology of WSSV using molecular tools, which may enhance our understanding of the quick geographical spread of the virus. Furthermore, the genetic characterization of WSSV isolates may result in the identification of virulence-related genes on the WSSV genome. Because WSSV contains a number of unique genetic features (see above), in chapters 5 through 7 we studied WSSV transcription regulation, focusing on promoter motifs involved in gene expression. The expression profiling using DNA microarray technology (Chapter 5) may contribute to the identification of key genes in the regulation of WSSV transcription. A major difficulty in current WSSV research is the lack of relatively simple, well defined systems, such as permissive cell cultures, in which WSSV can be propagated and studied. Therefore, the studies described in this thesis were all performed *in vivo*.

Lo *et al.* (1999), Nadala and Loh (1998) and Wang *et al.* (2000b) had shown the presence of genetic variation among different geographic isolates of WSSV, but the genomic loci of these polymorphisms were unknown. The aim of **Chapter 2** is to identify the polymorphic loci of WSSV by computational comparison of the three completely sequenced isolates: WSSV-TW, WSSV-CN and WSSV-TH. This alignment revealed a very high degree of sequence identity (over 99%) among the WSSV isolates. However, various variable genomic loci were identified, leading to the identification of genetic markers which can be used in epidemiological and ecological studies of WSSV. Using these markers the genetic relationship between various Vietnamese (VN) WSSV isolates and the three completely sequenced WSSV isolates was studied in **Chapter 3**. The VN isolates were collected along the central and south coast of Vietnam during WSSV outbreaks in 2003 and 2004. Preliminary studies by Lan *et al.* (2002) and Wang *et al.* (1999a) suggested differences in virulence between various WSSV isolates. To further determine whether the genomic differences identified in chapters 2 & 3 correlate with differences in fitness and virulence, two WSSV isolates (TH96-II and WSSV-TH) were compared in **Chapter 4**. TH-96-II is a WSSV isolate with the largest genome size (~312 kb) reported at present, while WSSV-TH contains

the smallest WSSV genome known (~293 kb). Virulence and fitness were determined in shrimp by recording the mortality upon WSSV infection as well as by competition of both isolates in serial passage.

The remainder of the thesis concentrates on WSSV transcription. The main focus of **Chapter 5** is to determine the expression profiles of the 184 ORFs encoded by the WSSV genome using newly made WSSV DNA microarrays. Furthermore, this chapter aims to correlate the differences between TH-96-II and WSSV-TH observed in chapter 4 with differences in WSSV gene expression. During a WSSV-TH infection time course, transcription of all individual genes in *P. monodon* gill tissue was monitored over time. Clustering of these expression profiles enabled the classification of WSSV genes into a putative early or a putative late class. The transcription regulation of the late class genes is investigated in more detail in **Chapter 6**. Transcription of five major structural virion protein genes of WSSV (*vp28*, *vp26*, *vp24*, *vp19* and *vp15*), which clustered in the late class in chapter 5, was further studied using RT-PCR. To search for common (consensus) late promoter motifs, which have been identified in most other eukaryotic large ds DNA viruses, the transcription initiation sites of the five major structural virion protein genes were determined and aligned. Furthermore, polyadenylation of these genes is studied by determining the 3' ends of their transcripts. **Chapter 7** aims to identify promoter motifs of the early and late WSSV gene class *in silico*. The results are validated by alignment of various empirically determined 5' ends of WSSV mRNAs, either early or late. Using this analysis a simple model of WSSV transcription is presented. In **Chapter 8** the results communicated in the various chapters are discussed and evaluated. The (dis)advantages of the use of microarrays in viral gene expression profiling (chapter 5) are summarized. The classes of WSSV genes obtained by the microarray analysis and the promoter motifs identified in chapters 6 and 7 are discussed in the context of transcription regulation of other eukaryotic large ds DNA viruses. Furthermore, the origin of the genetic heterogeneity among WSSV isolates, as observed in chapters 2, 3 and 4, is discussed and put into an evolutionary context. Finally, we describe how the results obtained in this thesis may help in the design and development of novel WSSV intervention strategies.

Chapter 2

Genetic variation among isolates of White spot syndrome virus

Abstract

White spot syndrome virus (WSSV), member of a new virus family called *Nimaviridae*, is a major scourge in worldwide shrimp cultivation. Geographical isolates of WSSV identified so far are very similar in morphology and proteome, and show little difference in restriction fragment length polymorphism (RFLP) pattern. We have mapped the genomic differences between three completely sequenced WSSV isolates, originating from Thailand (WSSV-TH), China (WSSV-CN) and Taiwan (WSSV-TW). Alignment of the genomic sequences of these geographical isolates revealed an overall nucleotide identity of 99.32%. The major difference among the three isolates is a deletion of approximately 13kb (WSSV-TH) and 1kb (WSSV-CN), present in the same genomic region, relative to WSSV-TW. A second difference involves a genetically variable region of about 750bp. All other variations > 2bp between the three isolates are located in repeat regions along the genome. Except for the homologous regions (*hr*1, *hr*3, *hr*8 and *hr*9), these variable repeat regions are almost exclusively located in ORFs, of which the genomic repeat regions in ORF75, ORF94 and ORF125 can be used for PCR based classification of WSSV isolates in epidemiological studies. Furthermore, the comparison identified highly invariable genomic loci, which may be used for reliable monitoring of WSSV infections and for shrimp health certification.

Introduction

White spot syndrome virus (WSSV) is the sole species of a new monotypic family called *Nimaviridae* (genus *Whispoviru*s). WSSV infects a wide range of aquatic crustaceans, including salt and brackish water penaeids, crabs and fresh water crayfish (Lo *et al.*, 1996; Wang *et al.*, 1998). Characteristic for infected shrimps are the white spots on the exoskeleton. In cultured shrimp, WSSV infection can reach a mortality of up to 100% within 3-10 days (Lightner, 1996). The WSSV virion consists of a rod-shaped nucleocapsid surrounded by a trilaminar envelope and a unique thread-like polar extension at one end (Durand *et al.*, 1997; Wongteerasupaya *et al.*, 1995). Isolated nucleocapsids have a crosshatched appearance, and contain a circular double-stranded (ds) DNA molecule of approximately 300 kilobasepairs (van Hulten *et al.*, 2001a; Yang *et al.*, 2001). The WSSV genome is at present the largest animal virus that has been completely sequenced.

The genome of one isolate (WSSV-TH) (van Hulten *et al.*, 2001a) encompasses 184 putative ORFs, of which only 11 have homologues in public databases, mainly representing genes encoding enzymes for nucleotide metabolism, DNA replication and protein modification (van Hulten *et al.*, 2001a). Ten gene families, consisting of 2 to 4 ORFs with pairwise similarities of 40% or higher, were also identified on the genome. Unique features of the WSSV genome are further the presence of an extremely long ORF of 18,234 nt (ORF167) with unknown function, a collagen-like ORF (ORF30), and nine non-coding regions (*hr*), dispersed along the genome, each containing a variable number of about 250 bp-long homologous tandem repeats (van Hulten *et al.*, 2001a). Phylogenetic analysis based on the DNA polymerase and other conserved WSSV genes, confirms the unique taxonomic position of WSSV, with no or only distant relationships to other families of large dsDNA viruses (Liu *et al.*, 2001; Tsai *et al.*, 2000b; van Hulten *et al.*, 2000b; van Hulten and Vlak, 2001; van Hulten *et al.*, 2001a; Witteveldt *et al.*, 2001).

Since its initial discovery in East-Asia in China and Taiwan in 1992, WSSV has quickly spread to shrimp-farming areas in Southeast-Asia and later to the Indian continent and Central- and Latin-America, causing major economic damage to shrimp culture (Cai *et al.*, 1995; Rosenberry, 1996; Rosenberry, 2000). The geographical isolates of WSSV identified so far are very similar in morphology and proteome (Huang *et al.*, 2001; Nadala and Loh, 1998; Wang *et al.*, 2000c). Although direct comparisons were not made, preliminary studies indicate that there seems to be little difference in virulence between various WSSV isolates (Lan *et al.*, 2002; Wang *et al.*, 1999a). Several attempts to distinguish WSSV isolates by genetic markers revealed only minor or no differences in RFLP pattern (Lo *et al.*, 1999; Nadala and Loh, 1998; Wang *et al.*, 2000b), suggesting a high genetic similarity between WSSV isolates.

Recently, the complete genomes of three different WSSV isolates have been sequenced. One isolate was collected from a WSSV-infected pond in Suratthan, Thailand in 1996 (acc. no. AF369029) (van Hulten *et al.*, 2001a), the second isolate from shrimp ponds in Tongan, Xiamen, East-China in 1996 (acc. no. AF332093) (Yang *et al.*, 2001), and the third

from shrimp farms located in Southern Taiwan in 1994 (acc. no. AF440570) (Wang *et al.*, 1995). In this paper these three isolates will be referred to as WSSV-TH, WSSV-CN and WSSV-TW, respectively, according to their country of origin (Fig. 2.1). The main characteristics for each isolate are summarized in Table 2.1. No biological differences, in host range or in virulence, have been described for these geographical WSSV isolates so far. Restriction enzyme analysis of the three isolates only showed minor differences in RFLP pattern.

Figure 2.1. Map of Southeast-Asia, showing the geographical origin of the isolates WSSV-TH (Thailand), WSSV-CN (China) and WSSV-TW (Taiwan).

The availability of the three completely sequenced WSSV genomes now allows a computational analysis of the genetic variation within this novel viral taxon. We have mapped all genomic differences between these geographical WSSV isolates. By determining the genomic location of variable regions and polymorphic sites, new WSSV isolates may be classified by their variation within these regions. Furthermore, the comparison identifies highly invariable genomic loci, present in all three sequenced WSSV isolates. These loci are suitable for reliable monitoring of WSSV infections, useful for shrimp health certification and in forensics.

Table 2.1. 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
Date of collection infected shrimp	05-1996	10-1996	11-1994
Infected species	*Penaeus monodon*	*Penaeus japonicus*	*Penaeus monodon*

Results

Three WSSV isolates originating from different geographical locations throughout Southeast-Asia have been completely sequenced. To reveal the genetic variation between the different isolates, the full genomic sequences were aligned and compared. The three complete circular genomic sequences were taken from the NCBI database entry, in which each of the three sequences is present in the same orientation ("clockwise"). The sequences of WSSV-TW and WSSV-CN were aligned to WSSV-TH, using the startcodon of the major structural virion

protein VP28 as starting point of the alignment. The original nucleotide coordinates of each of the isolates were kept. Further analysis on the open reading frames (ORFs) was performed using the annotation of each of the individual sequences, retaining the original designation of the ORFs present in the annotation.

After alignment, five different categories of genetic variation were identified, each being dealt with separately: (i) a large deletion of about 13 kb present in the WSSV-TH genome relative to WSSV-CN and WSSV-TW, (ii) a variable region prone to recombination, (iii) a transposase sequence present only in WSSV-TW, (iv) variation in the number of repeat units within *hr*s and direct repeats, (v) single nucleotide mutations, including deletions/insertions and single nucleotide polymorphisms (SNPs).

(i) Occurrence of a large deletion

Compared to the genomic sequence of the TW isolate, having the largest genome, the WSSV-TH isolate contains a deletion of approximately 13 kb. This 13,210 bp sequence is present at an intergenic sequence of WSSV-TH with genomic location 31134-31135 (between the sequences coding for ORF23 and ORF24; Fig. 2.2). The main difference between WSSV-TW and WSSV-CN in this genomic region, except for some minor single nucleotide mutations, is a deletion of 1168 bp in WSSV-CN (Fig. 2.2).

Relative to WSSV-TH, WSSV-CN has 12,049 bp extra in this region (Fig. 2.3a). The 12,049 bp encompass thirteen ORFs (designated A-M, or according to the annotation of WSSV-CN CN479-CN500; Fig. 2.2). Due to the extra 1,168 bp present in WSSV-TW compared to WSSV-CN, ORF K (WSSV-TW ORF021) is 1,137 bp longer at the 5' end in WSSV-TW (Fig. 2.3a). Two of the ORFs (ORFs A and L) belong to gene family 4 described for the WSSV-TH genome (van Hulten *et al.*, 2001a) and have an average amino acid similarity with the other gene members of about 40% and 50%, respectively. One ORF (ORF M) shows 56% amino acid similarity to WSSV-TH ORF23, which is flanking the deletion in WSSV-TH, and might also be the result of an ancient gene duplication event. ORF I (WSSV-CN; CN493) is thought to encode a nucleocapsid protein (VP35) (Chen *et al.*, 2002a) and this protein should therefore be absent in virions of WSSV-TH. For the remaining ORFs, no homologues could be identified in GenBank. However, all ORFs present on this large deleted genomic sequence are apparently dispensable for infection and replication of WSSV in *Penaeus monodon* and *Orconectes limosus*, as both crustacean species are permissive host for WSSV-TH (van Hulten *et al.*, 2001a).

The presence of ORFs belonging to WSSV gene families strongly suggests that the 13 kb sequence is an authentic part of the WSSV genome and that the TH isolate lacks this fragment due to a deletion event. Both ends of the 13 kb deletion contain a homologous sequence (11 out of 16 nucleotides are identical), which occurs only once (TCCCCCCTCTCTAGTG) in the WSSV-TH genome at the deletion site (Fig. 2.3a). This could suggest looping and subsequent excision (intramolecular recombination) of this region

Figure 2.2. Circular map of the WSSV-TH DNA genome showing the three major differences (13 kb deletion, variable region and transposase) between the genetic sequences of WSSV-TH, WSSV-CN and WSSV-TW. Sites for *Bam*H1 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. Middle gray arrows represent the structural protein genes. Light gray arrows represent ORFs present in both WSSV-CN and WSSV-TW, but not in WSSV-TH. These ORFs are numbered A to M, but also CN479 to CN500, in accordance with the numbering used by Yang *et al.* (2001). Light-middle gray arrows are ORFs only present in WSSV-TW. TW021 is numbered according with the annotated sequence from accession number AF440570 (WSSV-TW). The *hr* sequences are enlarged within the inner circle. The scale on the inner circle is in map units. A coloured version of this figure can be obtained from http://www.dpw.wau.nl/viro/research/wssv-genome.html.

21

Figure 2.3. Schematic representation of the three major differences **(a, b and c)** between WSSV-TH, WSSV-CN and WSSV-TW. The map numbers, indicated above each isolate, are in accordance with the numbers in the NCBI databank (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) for the genomic sequence of each isolate. Length of fragments is indicated within boxes or sequences. ORF numbering in (A) is explained in the legend of Fig. 2.2.

from the WSSV genome in a recent ancestor. This process has been observed also in large dsDNA viruses like herpesviruses and baculoviruses (Croizier and Ribeiro, 1992; Delius and Clements, 1976). The absence of a 1,168 bp fragment in WSSV-CN compared to WSSV-TW did most likely not occur by intramolecular recombination, as there are no homologous sequences flanking the site of this fragment in WSSV-TW.

*(ii) A variable region prone to recombination*

A second major difference between the three isolates concerns a genetic variation located at WSSV-TH genomic location 22961-23619 (in the genome segment coding for ORF14 and ORF15; Fig. 2.2). In this segment, WSSV-TH and WSSV-TW contain different sequences of 657 and 834 bp, respectively, both with no homology to any nucleotide sequences available in public databases, nor elsewhere in the WSSV genome (Fig. 2.3b). Of these sequences, 257 bp of WSSV-TH are present at the 5' end and 585 bp of WSSV-TW at the 3' end of the WSSV-CN sequence in this region (Fig. 2.3b). Although other mechanisms cannot be excluded, this also seems to be the result of a recombinatorial event, in which the sequence of WSSV-CN in this region evolved from the sequences of WSSV-TH and WSSV-TW. However, no sequences that could be involved in a recombination event were identified within 300 bp of the putative recombination site in the genomic sequences of the three isolates, so the mechanism by which this recombination could have occurred remains unclear. The presence of this genomic region shared by three different virus isolates suggests that animals can be infected by multiple viruses enabling recombination between the viruses.

(iii) Occurrence of a transposase sequence

A third major difference between the three WSSV genomes is an insert of 1,337 bp in the TW isolate. This insert is located at the WSSV-TH genomic location 254028-254029, in the WSSV genome sequence coding for the putative ORF166 (Fig. 2.2). The 1,337 bp insertion has 100% homology with known transposable elements, both from prokaryotic as well as eukaryotic origin (transposon type IS2), and encompasses an ORF encoding a transposase (Kleckner, 1981). The upstream and downstream sequences are inverted repeats, a typical character of transposons. The viral sequence (TGCCTAACA) at the site of insertion in the WSSV-TW genome has been duplicated (Fig. 2.3c). This sequence is also present at position 11504 in the WSSV-TH genome and at the corresponding positions in the WSSV-TW and WSSV-CN genome. Both the inverted terminal repeat and the duplicated viral sequence are typically for the insertion of a transposon. Since the origin of the transposase sequence is unknown and transposons are easily excised and inserted in DNA, this sequence is not suitable as a genetic marker.

(iv) Variation in number of repeat units within hr*s and direct repeats*

Four of the nine homologous regions (*hr*s) show a difference in the number of repeat units between WSSV-TH, WSSV-CN and WSSV-TW. Compared to WSSV-CN and WSSV-TW, WSSV-TH has one additional repeat unit in *hr*1 and *hr*8, and two extra repeat units in *hr*3, while WSSV-CN has one additional repeat unit in *hr*9 compared to the other two isolates (Fig. 2.4). Other large circular dsDNA viruses like the baculoviruses and ascoviruses also
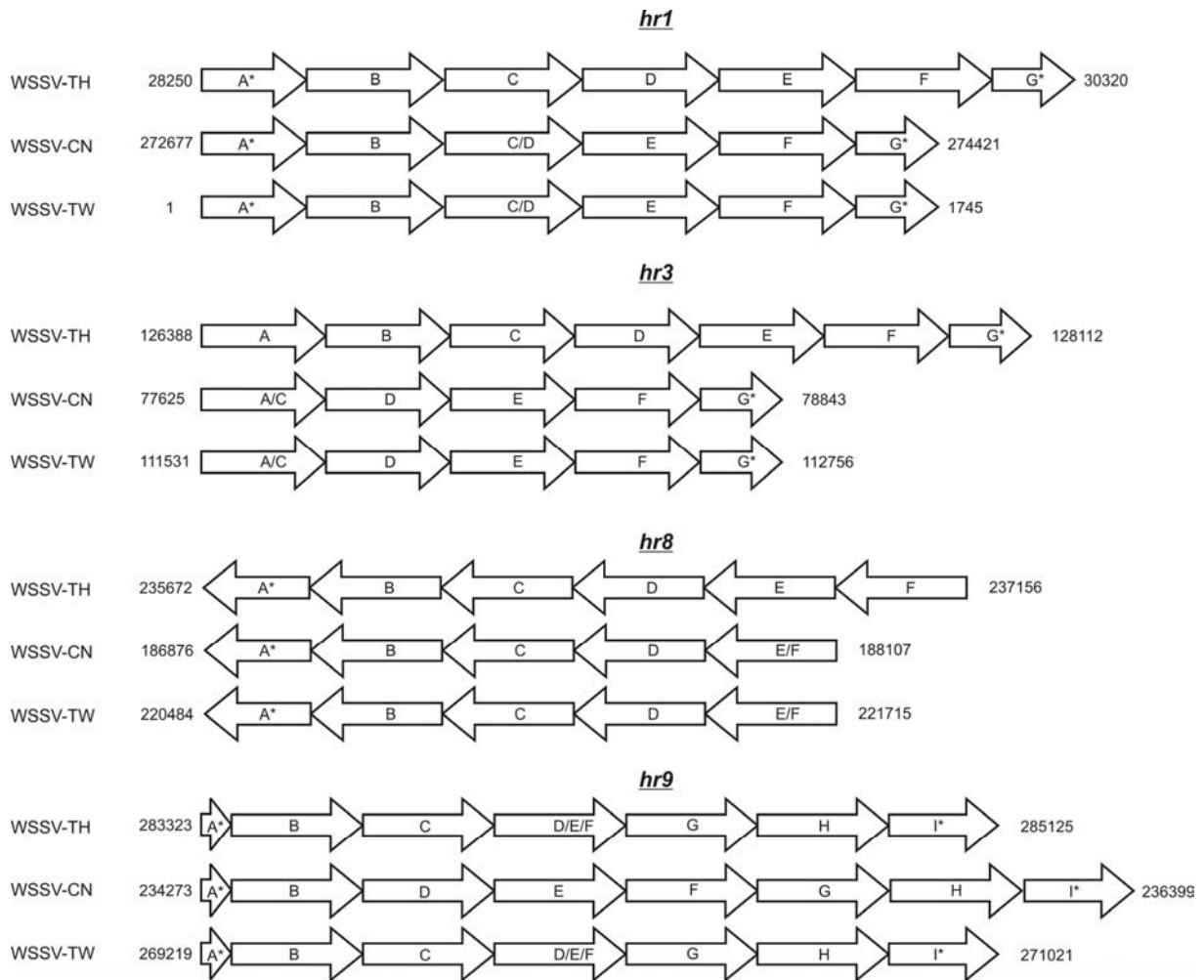
Figure 2.4. Schematic representation of the *hr*s different between WSSV-TH, WSSV-CN and WSSV-TW. The repeat units are depicted as arrows, indicating their respective orientation on the genome. Partial repeats are shown by a shorter arrow and an asterisk (*) following its letter. The map numbers are in accordance with the numbers in the NCBI databank for each isolate (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide).

contain *hr*s (Bigot *et al.*, 2000; Cochran and Faulkner, 1983). In baculoviruses, *hr*s play a role in DNA replication and enhancement of transcription (Guarino and Summers, 1986; Kool *et al.*, 1993). In baculoviruses as well as ascoviruses, the number of repeat units within one *hr* can be different between variants of the same virus species, most likely as a consequence of sequence duplication (Bigot *et al.*, 2000; Garcia-Maruniak *et al.*, 1996; Muñoz *et al.*, 1999). As also variants of WSSV show differences in repeat units within *hr*s (Fig. 2.4), this might be a general feature among large circular dsDNA viruses.

Comparison of WSSV-CN with WSSV-TH revealed 16 differences > 2 bp (Fig. 2.5a). All these differences, ranging in size from 3 to 324 bp, were identified only in direct repetitive sequences and not in other, non-repetitive genomic locations, indicating that these repeat regions are most prone to mutations in the WSSV genome. Only one of the variable direct repeats is present in an intergenic region while all others are located in ORFs (Fig. 2.5a). Six of the 16 differences were identified in non-*hr* unidirectional repeats, which were identified using CBS prediction servers (Fig. 2.5a). These repeats are sequences of 100 bp that are present in at least two copies on the same strand and unidirectional, with an identity between these copies of 80% or higher. Ten non-*hr* unidirectional repeats, all direct tandem repeats, are present all together in the WSSV-TH genome, one more is present in WSSV-TW in the part of the large deletion that is absent in WSSV-CN and WSSV-TH (Fig. 2.5b). Of the 10 non-*hr* unidirectional repeats present in the WSSV-TH genome (Table 2.2; Fig. 2.5a), 6 contain differences between the three isolates (bold in Table 2.2), of which 5 have differences in number of repeat units within these repeats (underlined in Table 2.2). The three largest changes are located in the genomic regions coding for the putative ORF75, ORF94 and ORF125.

Table 2.2. Comparison of repeat units of non-*hr* unidirectional repeats between WSSV isolates.

Genomic position (WSSV-TH coordinates)	Length of repeat units (bp)	Number of repeat units present (length of units)		
		WSSV-TH	WSSV-CN	WSSV-TW
11167-11454[a] (Only in WSSV-TW; TW 021)	9	-	-	32
44830-48294 (ORF30)	63	55.5	55.5	55.5
93118-93219 (ORF65)	3	34	33	33
97066-97400 (ORF67)	56/84	2.5 (84) and 2 (56)	2.5 (84) and 2 (56)	2.5 (84) and 2 (56)
107965-108675 (ORF75)	45/102	3 (102) and 9 (45)	4 (102) and 11 (45)	5 (102) and 16 (45)
119018-119311 (ORF84)	84	3.5[b]	3.5	3.5
142744-143067 (ORF94)	54	6	12	6
176987-177100 (ORF116)	42	2.75	2.75	2.75
180619-180834 (ORF119)	66/6	3 (66) and 4 (6)	3 (66) and 7 (6)	3 (66) and 7 (6)
187899-188312 (ORF125)	69	6	8	8
286717-286878 (ORF177)	54	3	3	3

[a] WSSV-TW coordinates
[b] One repeat unit has 3 bp extra in WSSV-TH
Bold: non-*hr* unidirectional repeats which contain differences between the three isolates
Underlined: non-*hr* unidirectional repeats with a different number of repeat units between the three isolates

Comparing WSSV-TW with WSSV-TH, 19 changes > 2 bp were identified, also all in direct repetitive sequences (Fig. 2.5b). From these sequences, ranging in size from 3 to 519 bp, 8 are identical to the WSSV-CN sequence and 3 others are located on the same genomic positions as differences between WSSV-TH and WSSV-CN (Fig. 2.5).
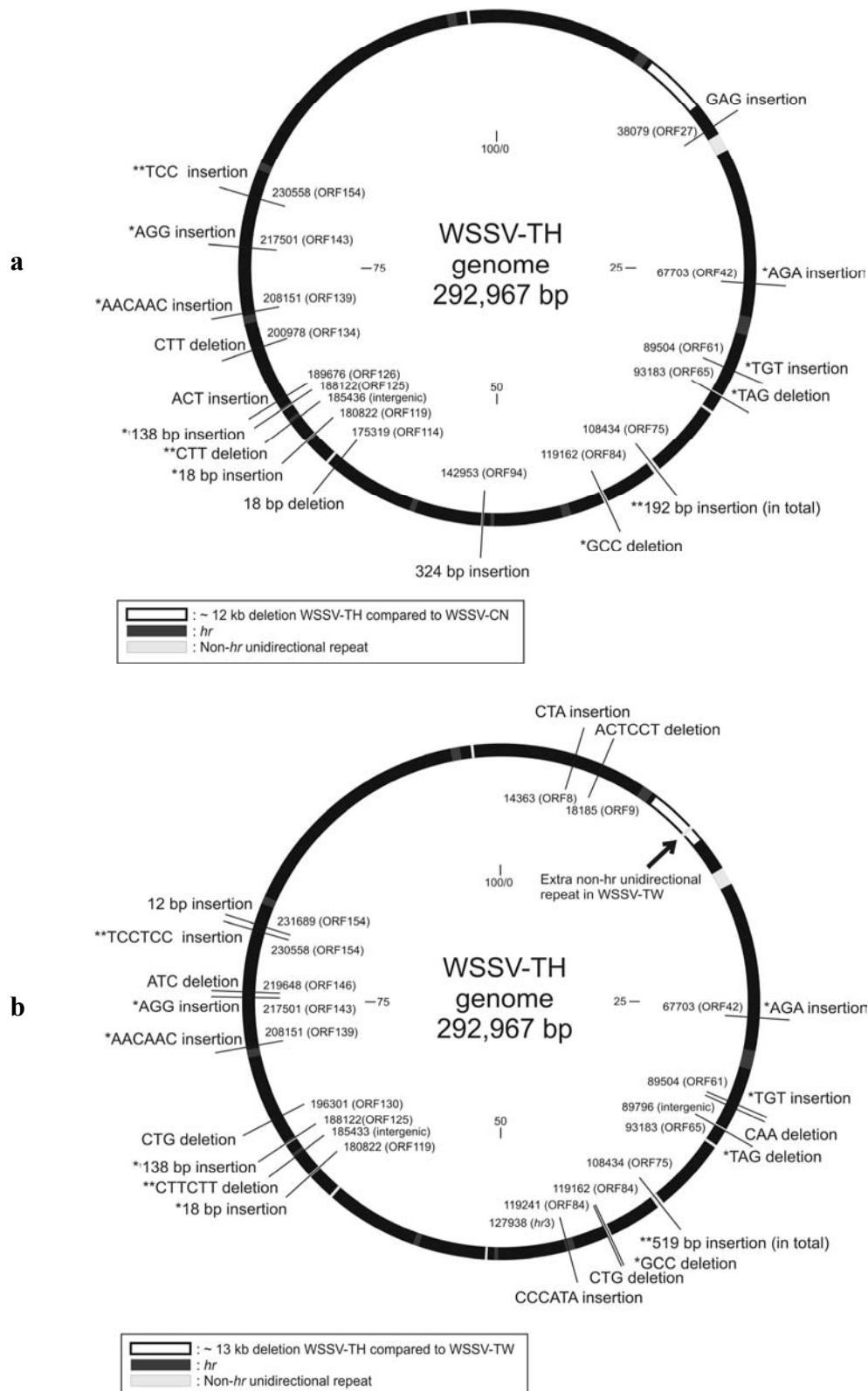
Figure 2.5. Comparison of WSSV-CN **(a)** and WSSV-TW **(b)** with WSSV-TH using a circular map, showing differences (>2 bp) in non-*hr* direct repeats. Differences in WSSV-CN (A) and WSSV-TW (B) compared to WSSV-TH are shown outside of the circle. Inside are the genomic positions (startpositions) of the changes in WSSV-TH coordinates. One asterisk (*) means the mutation is present in both WSSV-CN and WSSV-TW, two asterisks (**) mean a similar mutation at the same location is present in WSSV-CN and WSSV-TW. The cross (†) means 3 SNPs are present between WSSV-CN and WSSV-TW in the insertion in ORF125. The scale is in map units.
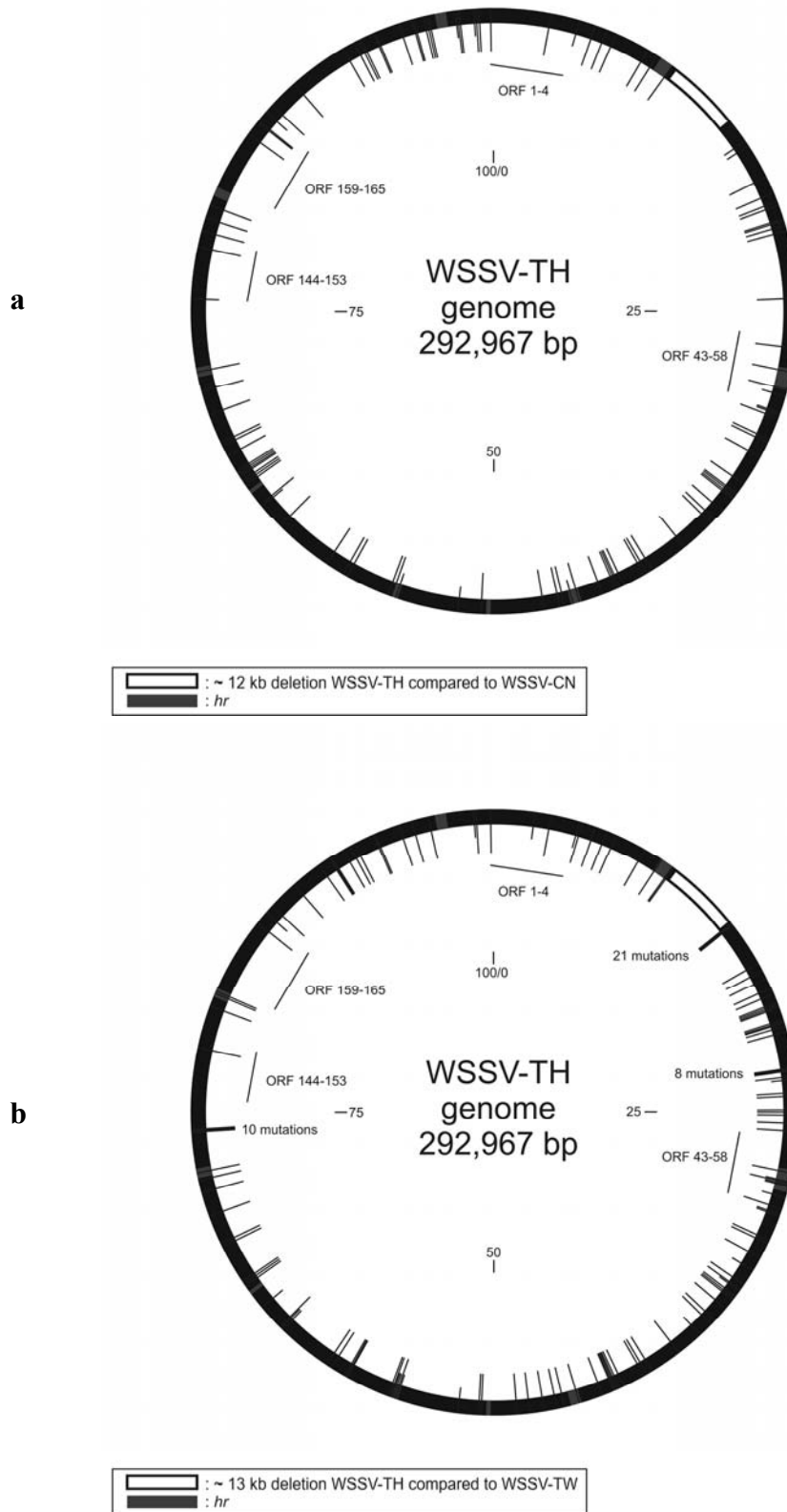
Figure 2.6. Comparison of WSSV-CN **(a)** and WSSV-TW **(b)** with WSSV-TH using a circular map, showing single nucleotide mutations. The short lines inside the map indicate a single nucleotide insertion or deletion, while the longer lines indicate SNPs. ORFs within regions with little mutations are indicated inside the genome. The scale is in map units.

(v) Single nucleotide mutations (Single nucleotide polymorphisms (SNPs) and deletions/insertions)

Alignment of the genomic sequences, without the three major differences (large deletion, variable region and the transposase sequence) and the variation in repeats, revealed a 99.65% pairwise nucleotide identity between WSSV-TH and WSSV-CN, a 99.45% pairwise nucleotide identity between WSSV-TH and WSSV-TW and a multiple nucleotide identity between the three isolates of 99.32%.

An overview of the single nucleotide mutations (SNPs and insertions/deletions) of WSSV-CN and WSSV-TW compared to WSSV-TH is shown in Table 2.3. These mutations are randomly distributed over the WSSV genome (Fig. 2.6a for WSSV-CN; Fig. 2.6b for WSSV-TW), except that for WSSV-TW approximately 25% occurred in the coding regions of ORF24, ORF25, ORF30, ORF38 and ORF84. Except for ORF30 (collagen-like ORF), these ORFs do not have homologues in public databases.

Table 2.3. Single nucleotide mutations (SNPs and insertions/deletions) of WSSV-CN and WSSV-TW compared to WSSV-TH.

	WSSV-CN		WSSV-TW	
Total of single nucleotide changes	105		163	
Similar changes WSSV-CN/ WSSV-TW	70		70	
	Coding	Intergenic	Coding	Intergenic
Insertions	2	5	3	5
Deletions	6	1	7	9
Total insertions/deletions	8 (0.030‰)	6 (0.26‰)	10 (0.037‰)	14 (0.60‰)
Single nucleotide polymorphisms (SNPs)	91 (0.31‰)		139 (0.54‰)	
Non coding SNPs (ncSNPs)	11 (0.46‰)		17 (0.73‰)	
Coding SNPs (cSNPs)	80 (0.29‰)		122 (0.45‰)	
Synonymous SNPs (sSNPs)	18		33	
Non-synonymous SNPs (nsSNPs)	62		89	

As the 184 ORFs account for 92% of the genetic information for WSSV-TH, there are about 10 times more deletions/insertions in intergenic regions than in ORFs (for WSSV-CN 0.26‰ and 0.030‰, respectively, and for WSSV-TW 0.60‰ and 0.037‰, respectively) (Table 2.3). The SNPs show a similar trend, as they occur about 1.5 times more frequently in non-coding sequences (Table 2.3).

Effect of differences on WSSV genes

The WSSV-TH ORFs and their corresponding homologues of WSSV-CN and WSSV-TW are listed in Table 2.4. Of the major structural virion protein genes *vp28*, *vp26*, *vp24*, *vp19* and *vp15*, only the genes *vp24* and *vp19* show differences between the three isolates. Compared to WSSV-TH, the WSSV-CN *vp19* has two non-synonymous SNPs, while WSSV-TW *vp19* only has one. WSSV-TW *vp24* has one non-synonymous SNP compared to the other two isolates. Previous studies have also shown that minimal sequence and immunological variation exists between the major structural proteins of WSSV (Anil *et al.*, 2002; Moon *et al.*, 2003; Poulos *et al.*, 2001; You *et al.*, 2002). Of the minor proteins that have been reported to be present in the virion (except for VP35, for which the gene is not present in WSSV-TH

Table 2.4. Comparison of the ORFs of WSSV-TH with the ORFs of WSSV-CN and WSSV-TW.

ORF WSSV-TH†	Putative function/ name$	ORF WSSV-CN*	ID (nt) %	ID (aa) %	Type of change (remark)	ORF WSSV-TW**	ID (nt) %	ID (aa) %	Type of change (remark)
1	in virion (VP28)	421	100	100		480	100	100	
2	Protein kinase	423	100	100		482	100	100	
3		427	100	100		486	100	100	
4		433	100	100			99.97	-	frameshift
5		440	99.95	99.84	M -> V		99.95	-	M -> V
6	in virion (vp800)	442	100	100		502	100	100	
7			99.85	-	frameshift (high homology with CN446)		99.70	-	frameshift, Stop -> W
8			100	-			99.56	-	M (start) -> V, 3 bp insertion
9		447	99.95	100	3 silent mutations (V -> V, S -> S, E -> E)	507	99.85	99.90	6 bp deletion, 3 silent mutations (V -> V, S -> S, E -> E)
10		455	100	100		515	100	100	
11		457	100	100		517	100	100	
12		459	100	100		519	100	100	
13		460 in databank	100	100		520	100	100	
14			-	-	disrupted (high homology with CN461)		-	-	disrupted
15			-	-	disrupted (high homology with CN464)		-	-	disrupted
16		465	99.95	99.92	silent mutation (I -> I), M -> V	524	99.97	99.92	L -> S
17		473 in databank	100	100			100	-	
hr1									
18 (*hr1*)			100	-			100	-	
19 (*hr1*)		474 in databank	100	100			100	-	
20 (*hr1*)			-	-	not in WSSV-CN		-	-	not in WSSV-TW
21 (*hr1*)			-	-	not in WSSV-CN		-	-	not in WSSV-TW
22 (*hr1*)		476 in databank	100	100		3	100	100	
23		477	100	100		4	100	100	
		479	-	-		6	-	-	
		482	-	-		9	-	-	
		483	-	-		10	-	-	
		484	-	-		11	-	-	
		486	-	-		13	-	-	
		489	-	-		16	-	-	
		492	-	-			-	-	
	in virion (vp35)	493	-	-		19	-	-	
		495	-	-		21	-	-	
		497	-	-			-	-	
		500	-	-		25	-	-	
24		502	100	100	CN502 is 288 aa longer N-terminal		99.54	-	C -> Stop, silent mutation (G ->G), V -> E, frameshift, E-> V, R -> Stop, R-> G, E-> V, K -> L (high homolgy with TW27 and TW30)
25		508	100	100			99.32	-	2* frameshift, L -> R, R -> A, 2 silent mutations (T -> T, T -> T), D -> V, T -> S
26			99.57	-	frameshift	36	100	100	
27	DNA polymerase	514	99.93	-	frameshift , 3bp insertion, F -> L (CN514 is 156 aa shorter N-terminal)	39	99.96	99.87	D -> G, N -> S, F -> L
28		525	100	100		50	100	100	
29	in virion (vp448)	526	100	100		51	99.93	99.77	A -> V
30	Collagen/ in virion (vp1684)	1	99.90	99.76	G -> E, Q -> H, silent mutation (D -> D), G -> R, L -> F	52	99.86	99.76	G -> E, Q -> H, 3 silent mutations (V -> V, G -> G, D -> D), G -> R, L -> F
31	in virion (VP24)	2	100	100		58	99.84	99.52	K -> R
32		4	99.71	100	C -> Stop (CN4 is 8 aa shorter N-terminal)	60	99.71	100	C -> Stop (TW60 is 8 aa shorter N-terminal)
33		6	99.88	99.66	M (start) -> R	62	99.88	99.66	M (start) -> R
34	in virion (vp95)	9	100	100		65	100	100	
35		10 in databank	100	100		66	100	100	
36		11	100	100		67	100	100	
37		21	100	100		77	100	100	
38		23	100	100		79	99.07	98.25	I -> V, 3 silent mutations (L -> L, V -> V, L -> L), Q -> K, D -> Y, N -> D, A -> V
39		25	100	100		81	100	100	
40		26	99.98	100	silent mutation (T -> T)		99.91	-	silent mutation (I -> I), frameshift, F -> S, T -> S

ORF WSSV-TH†	Putative function/name$	ORF WSSV-CN*	ID (nt) %	ID (aa) %	Type of change (remark)	ORF WSSV-TW**	ID (nt) %	ID (aa) %	Type of change (remark)
41		35	100	100		92	99.90	100	3 silent mutations (I -> I, S -> S, V -> V)
42		37	99.92	99.92	3 bp insertion	94	99.87	99.77	3 bp insertion, G -> R, S -> F
43		45	99.97	100	silent mutation (L -> L)	102	100	100	
44		51	100	100		108	100	100	
hr2									
45 (*hr2*)		53	100	100		110	100	100	
46 (*hr2*)		55	100	100		112	100	100	
47 (*hr2*)		56	100	100		113	100	100	
48 (*hr2*)			100	-			100	-	
49 (*hr2*)		59	100	100		116	100	100	
50 (*hr2*)			100	-			100	-	
51 (*hr2*)		60 in databank	100	100		117	100	100	
52 (*hr2*)		61 in databank	100	100		118	100	100	
53			99.96	-	frameshift (high homology with CN64)		99.96	-	frameshift
54	Thymidylate synthase	67	100	100		124	99.89	99.65	L -> S
55		69	100	100		126	100	100	
56		73	100	100		130	100	100	
57		76	100	100		133	100	100	
58		77	100	100		134	100	100	
59		78	99.83	99.50	S -> P, K -> E	135	99.83	99.50	S -> P, K -> E
60		79	100	100		136	100	100	
61	Protein kinase	83	99.77	99.66	V -> D, 3 bp insertion	140	99.77	99.66	V -> D, 3 bp insertion
62		87 in databank	100	100		144	100	100	
63		89 in databank	100	100		146	100	100	
64		90 in databank	100	100		147	100	100	
65		91	99.88	99.83	3 bp deletion, N -> D		99.85	-	frameshift, 3 bp deletion, N -> D
66		100	99.84	99.52	F -> S, G -> R, A -> V	156	99.89	99.68	G -> R, A -> V
67		106 in databank	100	100		162	100	100	
68			100	-			100	-	
69		107	100	100		163	100	100	
70		108	100	100		164	99.92	99.75	M -> V
71	dUTPase	112	100	100		168	100	100	
72		115	99.97	99.90	A -> T	171	99.97	99.90	A -> T
73		119	100	100		175	100	100	
74		126 in databank	100	100		182	100	100	
75	in virion (vp357)	129	78.61	85.22	C -> Stop, 45 bp insertion, 45 bp deletion, 192 bp insertion (CN129 is 147 aa shorter N-terminal)	183	60.70	60.70	45 bp insertion, 474 bp insertion, silent mutation (A -> A)
76		130	100	100		185	100	100	
77		131	100	100		186	100	100	
78		133	100	100		188	100	100	
79		134	100	100		189	100	100	
80		136	100	100		191	100	100	
81		137	100	100		192	100	100	
82		139	99.92	99.83	H -> Q, G -> D, silent mutation (I -> I)	194	99.92	99.83	H -> Q, G -> D, silent mutation (I -> I)
83		142	100	100		197	100	100	
84		143	99.88	99.78	silent mutation (A -> A), 3 bp deletion, R -> K, A -> V, K -> R, S -> T	198	99.81	99.74	3 silent mutations (A -> A, A -> A, A -> A), 3 bp deletion, R -> K, A -> V, 3 bp deletion, K -> R, S -> T
85		147	100	100		202	100	100	
hr3									
86 (*hr3*)			-	-	R -> L, disrupted in WSSV-CN, frameshift (high homology with CN150)		-	-	R -> L, disrupted in WSSV-TW
87 (*hr3*)			-	-	disrupted in WSSV-CN		-	-	disrupted in WSSV-TW
88 (*hr3*)			99.83	-	frameshift (high homology with CN150)	206	100	100	
89		151	99.93	99.86	T -> M, N -> I, silent mutation (T -> T)	207	99.93	99.86	T -> M, E -> G, silent mutation (T -> T)
90		161	100	100		217	100	100	
91		166	100	100		222	99.94	100	2 silent mutations (P -> P, A -> A)
92	Ribonucleotide reductase (large subunit)	172	100	100		228	100	100	
hr4									
93		177	100	100		233	100	100	

ORF WSSV-TH†	Putative function/name$	ORF WSSV-CN*	ID (nt) %	ID (aa) %	Type of change (remark)	ORF WSSV-TW**	ID (nt) %	ID (aa) %	Type of change (remark)
94		178	44.44	43.81	E -> D, 324 bp insertion	234	99.66	98.96	E -> D, E -> D
95		180 in databank	100	100		236	100	100	
96		181	100	100		237	100	100	
97		184	100	100		240	100	100	
98	Ribonucleotide reductase (small subunit)	188	100	100		243	100	100	
99	Endonuclease	191	100	100		246	100	100	
100		192	100	100		247	100	100	
101		195	100	100		250	100	100	
102		198	100	100		253	100	100	
103		199	100	100		254	100	100	
*hr*5									
104 (*hr*5)		204 in databank	100	100		259	100	100	
105 (*hr*5)		105 in databank	100	100		260	100	100	
106		206	99.84	99.51	G -> E	261	99.84	99.51	G -> E
107		207	100	100		262	100	100	
108		209	100	100		264	100	100	
109	in virion (VP15; p6.8)	214	100	100		269	100	100	
110		215 in databank	100	100		270	100	100	
111		216	99.94	99.83	E -> K, S -> T	271	99.92	99.83	E -> K, silent mutation (F -> F), P -> T
112	Class I cytokine receptor/ in virion (vp674)	220	99.95	99.85	V -> D	275	99.90	99.70	D -> V, V -> D
113		222	100	100		277	100	100	
114		226	99.36	99.36	18 bp deletion	281	100	100	
115		230	100	100		285	100	100	
116		231	100	100		286	100	100	
117		234	99.89	99.66	A -> V	290	99.89	99.66	A -> V
118	in virion (vp292)	237	100	100		293	100	100	
119		238	98.75	98.75	18 bp insertion	294	98.75	98.75	18 bp insertion
120	in virion (vp300)	242	100	100		298	100	100	
121		244	99.96	99.87	I -> L	299	100	100	
*hr*6									
122 (*hr*6)		245 in databank	100	100		300	100	100	
123 (*hr*6)		246 in databank	100	100		301	100	100	
124		247 in databank	100	100		302	100	100	
125		249	93.50	93.22	C -> R, 2 silent mutations (G -> G, L -> L),138 bp insertion, T -> I, K -> N, K ->E	304	93.63	93.62	C -> R, 2 silent mutations (G -> G, L -> L),138 bp insertion
126		252	99.73	99.80	silent mutation (T -> T), 3 bp insertion	307	100	100	
127	in virion (vp281)	254	100	100		309	100	100	
128	in virion (vp384)	256	100	100		311	100	100	
129		259	99.89	99.68	Y -> H	314	100	100	
130		260	99.93	99.89	Y -> N, silent mutation (V -> V)	315	99.82	99.78	Y -> N, silent mutation (V -> V), 3 bp deletion
131		267	100	100		321	100	100	
132		269	100	100		324	100	100	
133		270	100	100		325	100	100	
134		271	99.89	99.83	V -> D, 3 bp deletion	326	99.97	99.92	V -> D
135		277	99.96	100	silent mutation (G -> G)	332	99.96	100	silent mutation (G -> G)
136			100	-			100	100	
*hr*7									
137 (*hr*7)			100	-			100	-	
138		281 in databank	100	100		336	100	100	
139			96.42	-	6 bp insertion		96.42	-	6 bp insertion
140		282	100	100		337	100	100	
141		284	100	100		339	100	100	
142		285	100	100		340	100	100	
143		289	99.94	99.94	3 bp insertion	344	99.74	99.55	Y -> Stop, T -> I, 3 silent mutations (V -> V, V -> V, V-> V), D -> N, Q -> P, R -> K, Y -> F, 3 bp insertion (TW344 is 19 aa shorter C-terminal)
144			99.66	-	frameshift	349	100	100	

ORF WSSV-TH†	Putative function/ name$	ORF WSSV-CN*	ID (nt) %	ID (aa) %	Type of change (remark)	ORF WSSV-TW**	ID (nt) %	ID (aa) %	Type of change (remark)
145		294	100	100		350	100	100	
146		295	100	100		351	99.52	99.52	3 bp deletion
147		299	100	100		355	100	100	
148		302	100	100		358	100	100	
149	TATA box binding protein/ in virion (vp184)	303	100	100		359	100	100	
150		306	100	100		362	100	100	
151	in virion (vp466)	308	99.93	99.79	K -> N	364	99.93	99.79	K -> N
152		310	100	100		366	100	100	
153	in virion (VP26; p22)	311	100	100		367	100	100	
154		313	99.89	99.83	N -> S, 3 bp insertion, silent mutation (K -> K)	369	99.49	99.49	6 bp insertion, 12 bp insertion
155		321	100	100		377	100	100	
156		322	100	100		378	100	100	
157		324	99.60	98.78	A -> V	380	99.60	98.78	A -> V
158		325	100	100		381	100	100	
*hr*8									
159		327	100	100		383	100	100	
160		332	100	100		388	100	100	
161		338	100	100		394	100	100	
162		339	100	100		395	100	100	
163		340	100	100		396	100	100	
164		342	100	100		398	100	100	
165			99.94	-	Q -> K, G -> D, silent mutation (S -> S) (CN343 is 2475aa longer N-terminal)	399	99.98	99.94	G -> D (TW399 is 1026 aa longer N-terminal)
166			99.96	-	frameshift, G -> D, K -> Q (CN343 is 1850 aa longer C-terminal)		80.84	-	frameshift, G -> D, 1337 bp insertion, K -> Q
	Transposase		-	-		410	-	-	
167		360	99.97	99.92	L -> F, I -> T, R -> S, V -> A, I -> G	419	99.95	99.90	L -> P, E -> G, E -> G, 2 silent mutations (I -> I, V-> V), L -> F, I -> T, I -> G
168	in virion (vp68)	386	100	100		445	100	100	
169		387	99.95	99.86	I -> T	446	99.95	99.86	I -> T
170		390	99.90	99.69	K -> E	449	100	100	
171	Chimeric Thymidine kinase- Thymidylate kinase	395	99.91	100	silent mutation (Y -> Y) (CN395 is 10 aa longer N-terminal)	454	100	100	
172		397	99.61	98.83	D -> G, Y -> H	456	99.81	99.42	Y -> H
173		398	100	100		457	100	100	
174		399	100	100		458	100	100	
*hr*9									
175 (*hr*9)			100	-			100	-	
176 (*hr*9)		402 in databank	100	100		461	100	100	
177		403	99.90	99.69	M -> I, T -> A	462	100	100	
178		406	100	100		465	100	100	
179		407	100	100		466	100	100	
180		411 in databank	100	100		470	100	100	
181			99.52	-	frameshift		99.52	-	frameshift
182	in virion (VP19)	414	99.45	98.35	V -> D, S -> P	473	99.73	99.17	V -> D
183	in virion (vp544)	415	100	100		474	100	100	
184		420 in databank	100	100		479	100	100	

† The WSSV-TH ORFs are numbered in accordance with van Hulten *et al.* (2001a)

$ Virion proteins indicated to be present "in virion" have been published by van Hulten *et al.* (2000a, c; 2002), Chen *et al.* (2002a), and Huang *et al.* (2002b)

*The WSSV-CN ORFs are numbered in accordance with Yang *et al.* (2001) or with the annotated genomic sequence (referred to as "in databank") with accession number AF332093 (WSSV-CN)

** The WSSV-TW ORFs are numbered in accordance with the annotated genomic sequence with accession number AF440570 (WSSV-TW)

genome), only three (ORF30 (collagen-like ORF; vp1684), ORF112 (Class I cytokine receptor; vp674) and ORF151 (vp466)) contain SNP(s) between the three isolates, while one (ORF75; vp357) contains a different number of repeat units. Of the other ORFs which have (putative) functions, only the ORF27 (DNA polymerase) (Chen *et al.*, 2002b), ORF61 (protein kinase) and ORF171 (chimeric thymidine kinase-thymidylate kinase) have differences between the three isolates.

Discussion

The development of genetic markers for the identification of WSSV and the ability to distinguish WSSV isolates from different geographic origins is important for shrimp farming and control of WSSV. The main problem up to now has been the lack of genetic information that could be used to develop such markers. For example, the target of the probe used by Nunan & Lightner (1997) for *in situ* hybridizations was located within the 13 kb segment absent in WSSV-TH, and therefore this probe is not suitable for diagnostics. We have mapped all genomic differences between three geographical WSSV isolates to identify variable as well as conserved genomic regions. Because WSSV-TH and WSSV-TW were isolated from *Penaeus monodon*, while WSSV-CN was isolated from *Penaeus japonicus*, it is very well possible that some of the variations identified between the isolates are host dependent. However, part of the variations found have already been shown to be variable within one host species in other studies, including the studies presented by Lan *et al.* (2002) and Wongteerasupaya *et al.* (2003).

This study shows that there are several highly conserved loci on the WSSV genome that are eligible for reliable monitoring of WSSV infections in crustaceans and shrimp health certification using PCR-based tests. Especially the regions between WSSV-TH genomic positions 1 to 8970 (coding for ORF1 to ORF4), 69737 to 84919 (coding for ORF43 to ORF58), 218566 to 228835 (coding for ORF144 to ORF153), and 237222 to 250966 (coding for ORF159 to ORF165) contain very little genetic variability when the three sequenced isolates are compared (Fig. 2.6). The genomic locations used for PCR based WSSV detection proposed by Lo *et al.* (1996b) (pms146 primers; the genomic location (WSSV-TH) of the PCR product is 273311-274757, amplifying a 1,447 bp product located in the coding region of ORF167) and Tapay *et al.* (1999) (the genomic location (WSSV-TH) of the PCR product is 181309-181519, amplifying a 211 bp product located in the coding region of ORF119) also contain little sequence variation. In our laboratory, we successfully use the conserved genomic regions coding for VP28 (ORF1) and VP26 (ORF153) in a combined PCR assay for unequivocal identification of WSSV in diseased crustaceans including shrimp.

With the small genetic variation among WSSV isolates, it has been difficult to develop markers to distinguish WSSV isolates by selecting random sequences (Lo *et al.*, 1999; Nadala and Loh, 1998; Wang *et al.*, 2000b). The use of RFLP to distinguish between the existing genotypes of WSSV does not seem to be adequate due to the high homology between the

WSSV isolates. Moreover, the use of RFLP requires a high amount of virus DNA, which cannot always be easily obtained for WSSV. Using the genome comparison, we identified genetic loci in the WSSV genome that can be used as markers to distinguish the genotypic variants of WSSV. Sequences around and within the 13 kb fragment absent in the WSSV-TH genome may be used for this purpose, as not only WSSV-TH and WSSV-CN contain deletions compared to WSSV-TW (Fig. 2.3a), but deletions of variable size have been reported in this region in other isolates as well. Wang *et al.* (2000b) showed, using probes located in the 13 kb region (A6/ LN4), that a WSSV variant isolated from crayfish *Orconectes punctimanus* in the US National Zoological Park, Washington DC, lacks (a part of) the 13 kb sequence. The differences in restriction pattern of three WSSV isolates shown by Nadala and Loh (1998) are probably also caused by a deletion of (part of) the 13 kb sequence. Unfortunately, detailed sequence information of these isolates is not available. Furthermore, Lan *et al.* (2002) showed that deletions in this region varying in size from 4.6 to 8.1 kb exist in WSSV isolates derived from different shrimp species, suggesting that WSSV diversity extends into different hosts.

Variations present in the repeat numbers of the homologous regions *hr*1, *hr*3, *hr*8 and *hr*9 (Fig. 2.4) or the non-*hr* unidirectional repeats located in ORF75, ORF94 and ORF125 (Fig. 2.5) can be used for screening in PCR-based assays to distinguish isolates, making epidemiological studies possible. Differences in *hr*s are used to distinguish genotypic variants for the baculoviruses *Se*MNPV and *Ag*MNPV (Garcia-Maruniak *et al.*, 1996; Muñoz *et al.*, 1999) and were suggested to use for this purpose for ascoviruses (Bigot *et al.*, 2000). However, in the case of WSSV, the use of non-*hr* unidirectional repeats could be preferred above the use of *hr*s, as the *hr*s are significantly longer resulting in large amplification products, which complicate the design of a robust PCR assay. Moreover, PCR primer design for *hr*s will be difficult as the *hr*s are located at multiple positions in the WSSV genome. Wongteerasupaya *et al.* (2003) already showed that the tandem repeats in the coding region of ORF94 are useful as marker for WSSV, as the number of repeats were variable (from 6 to 20 repeat units) between isolates from different shrimp ponds in Thailand. The genomic region of ORF75, which has a different number of repeats between WSSV-TH, WSSV-CN and WSSV-TW, could be used in a similar manner. Despite the variability in ORF75 between the three isolates, it has been shown to encode a functional protein, present as minor protein (vp357; see Fig. 2.2) in the WSSV virion (Huang *et al.*, 2002b). For epidemiological and ecological studies on the spread of this virus, efficient discrimination between WSSV isolates is pivotal and therefore multiple markers should be targeted to genotype isolates. The repeats in the coding regions of ORF75, ORF94, ORF125, if necessary together with the sequence present in the 13 kb region and *hr*s that show variation, could be used for this purpose. However, the variable direct repeats are not useful in phylogenetic analysis of WSSV, because of the high mutation frequencies of insertions and deletions of repeat units (Wongteerasupaya *et al.*, 2003). This likewise holds for *hr*s.

The variable region that is located at position 22961 in the WSSV-TH genome (Fig. 2.3b) is also of interest for further study as it may shed light on the evolution of WSSV. The differences between the isolates in this region may have been introduced by one recombinational event (which is the most simple explanation concerning the variation between the three isolates shown in Fig. 2.3b), implying a (former) mixed occurrence of (ancestors of) the WSSV-TH and WSSV-TW. However, when considering the 13 kb region, which is (partially) absent in WSSV-CN and WSSV-TH (Fig. 2.3a) and completely present in WSSV-TW, the latter isolate may be the ancestor of WSSV-CN and WSSV-TH. Furthermore, when the single nucleotide changes between the three isolates are studied, WSSV-CN and WSSV-TW are most similar (Table 2.3), and therefore it could be concluded that WSSV-CN and WSSV-TW have the closest relation of the three isolates. Clearly more information of other isolates about both the variable region (Fig. 2.3b) as well as the 13 kb deletion (Fig. 2.3a) is needed to understand more about the relationship between the WSSV-TH, WSSV-CN and WSSV-TW isolates.

Materials and Methods

Computer analysis

Sequences of the three WSSV isolates were obtained from the NCBI databank (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) using the accession numbers for WSSV-TH (AF369029), WSSV-CN (AF332093) and WSSV-TW (AF440570) (also: http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/196937.html). WSSV-CN and WSSV-TH were sequenced using a shotgun based method, while WSSV-TW was sequenced using WSSV libraries. Upon alignment the nucleotide coordinates of the respective isolates are kept. Unidirectional non-*hr* repeats were identified on the WSSV genome using the GenomeAtlas (DNA Structural Analysis) of CBS prediction servers (http://www.cbs.dtu.dk). Sequences were analyzed using the software package DNASTAR4.2 and GeneDoc, version 2.6.000 (Nicholas *et al.*, 1997).

Chapter 3

Molecular epidemiology of White spot syndrome virus within Vietnam

Abstract

WSSV, sole member of the virus family *Nimaviridae*, is a large dsDNA virus infecting shrimp and other crustaceans. By alignment of three completely sequenced isolates originating from Taiwan (WSSV-TW), China (WSSV-CN) and Thailand (WSSV-TH), the variable loci in the genome were mapped. The variation suggests the spread of WSSV from a common ancestor originating from either site of the Taiwan Strait to Thailand, but support for this hypothesis through analysis of geographic intermediates is sought. RFLP analysis of eight Vietnamese WSSV isolates, of which six were collected along the central coast (VN-central) and two along the south coast (VN-south), showed apparent sequence variation in the variable loci identified previously. These loci were characterized in detail by PCR amplification, cloning and sequencing. Relative to WSSV-TW, all VN-central isolates showed a ~8.5 kb deletion in the major "variable region ORF23/24", whereas the VN-south isolates contain a deletion of ~11.5 kb and ~12.2 kb, compared to a ~1.2 kb and ~13.2 kb deletion in WSSV-CN and WSSV-TH, respectively. The minor "variable region ORF14/15" showed deletions of various sizes compared to WSSV-TH for all eight VN isolates. The data suggest that the VN isolates and WSSV-TH have a common lineage, which branched off from WSSV-TW and WSSV-CN early on, and that WSSV entered Vietnam by multiple introductions. We present a model for the spread of WSSV from either site of the Taiwan Strait into Vietnam based on the gradually increasing deletions of both "variable regions". The number and order of repeat units within ORF75 and ORF125 appeared to be suitable markers to study regional spread of WSSV.

Introduction

White spot syndrome virus (WSSV) is a large dsDNA virus belonging to the virus family *Nimaviridae* (Mayo, 2002). Since its discovery in China (Fujian) in 1991/1992, the virus quickly spread causing serious losses to commercial shrimp farming worldwide (Cai *et al.*, 1995; Flegel, 1997). The natural marine ecology is also threatened by WSSV as the virus has a wide host range, including salt and brackish water penaeids, crabs, spiny lobsters and fresh water shrimp and crayfish (Lo *et al.*, 1996b; Flegel, 1997; Wang *et al.*, 1998; Chen *et al.*, 2000; Maeda *et al.*, 2000; Hameed *et al.*, 2003). An overt clinical sign of infected shrimp is the presence of white spots on the exoskeleton (Chou *et al.*, 1995).

Electron microscopical analysis showed that WSSV consists of a rod-shaped nucleocapsid with a crosshatched appearance, surrounded by a trilaminar envelope with a unique tail-like appendix at one end (Wongteerasupaya *et al.*, 1995; Durand *et al.*, 1997; Nadala *et al.*, 1998). The circular double-stranded DNA genome of WSSV has a size of around 300 kb and is one of the largest animal virus genomes that has been entirely sequenced (van Hulten *et al.*, 2001a; Yang *et al.*, 2001). Only 6% of the putative 184 ORFs encoded by the viral genome have homologues in public databases, mainly representing genes encoding enzymes for nucleotide metabolism, DNA replication and protein modification (van Hulten *et al.*, 2001a).

Except for Southeast-Asia, WSSV also has been reported from the United States in 1995 (Rosenberry, 1996), and from Central- and South-America since early 1999 (Rosenberry, 2000). In 2002, WSSV was also detected in France and Iran (Rosenberry, 2002). The various geographical isolates of WSSV identified thus far are very similar in morphology and proteome. Limited differences in restriction fragment length polymorphism (RFLP) patterns have been reported, suggesting either a high degree of genomic stability or a recent emergence (Nadala & Loh, 1998; Lo *et al.*, 1999; Wang *et al.*, 2000b, c; Marks *et al.*, 2004). Preliminary studies indicated that there is also little difference in virulence between various WSSV isolates, although direct comparisons were not made (Wang *et al.*, 1999a; Lan *et al.*, 2002). After the complete sequencing of three different WSSV isolates originating from Taiwan (WSSV-TW; Wang *et al.*, 1995), China (WSSV-CN; Yang *et al.*, 2001) and Thailand (WSSV-TH; van Hulten *et al.*, 2001a), the major variable loci in the WSSV genome were mapped by alignment of these sequences (Marks *et al.*, 2004). Roughly, the variable loci can be divided into deletions, variable number of tandem repeats (VNTRs), single nucleotide indels and single nucleotide polymorphisms (SNPs). The variation within these loci, in particular in the large genomic deletions, suggested a geographical spread from a common ancestor from either site of the Taiwan Strait to Thailand (Marks *et al.*, 2004), but genetic intermediates were missing to support this hypothesis.

The present study focuses on WSSV isolates from Vietnam (VN), from eight different locations along the central and south coast. The variable loci, as identified by Marks *et al.* (2004), were subject of detailed analysis, including sequencing. Using these newly

characterized WSSV-VN genotypes, the value of each of the identified loci as genetic marker for strain identification as well as epidemiological and ecological studies is evaluated. Furthermore, molecular typing was used to analyze the relationship between the eight WSSV isolates from Vietnam and those from Taiwan, China and Thailand. The genetic changes could be correlated with spread of WSSV radiating out from either site of the Taiwan Strait to Thailand.

Results

Shrimp (*Penaeus monodon*) infected with WSSV and analyzed in this study were collected in 2003/ 2004 from eight shrimp culture ponds in Vietnam (Table 3.1). The ponds are distributed

Table 3.1. Origin of the Vietnamese WSSV isolates used in this study.

Name of the pond	Place (district)	Province	Date of collection	Origin of postlarvae*	Abbreviation
central WSSV-VN isolates					
Khanh	Hoi An	Da Nang	18/03/2003	Local	K
Thanh	Son Tinh	Quang Ngai	19/03/2003	Da Nang	T
Luong	Qui Nhon	Binh Dinh	20/03/2003	Local	L
Xu	Tuy Hoa	Phu Yen	21/03/2003	Local	X
Suu					S
Anh	Ninh Hai	Ninh Thuan	22/03/2003	Local	A
south WSSV-VN isolates					
C. Thanh A	Long Hoa	Tra Vinh	10/01/2004	Unknown	Tv
T. Sang	Kien Luong	Kien Giang	04/03/2003	Local	Kg
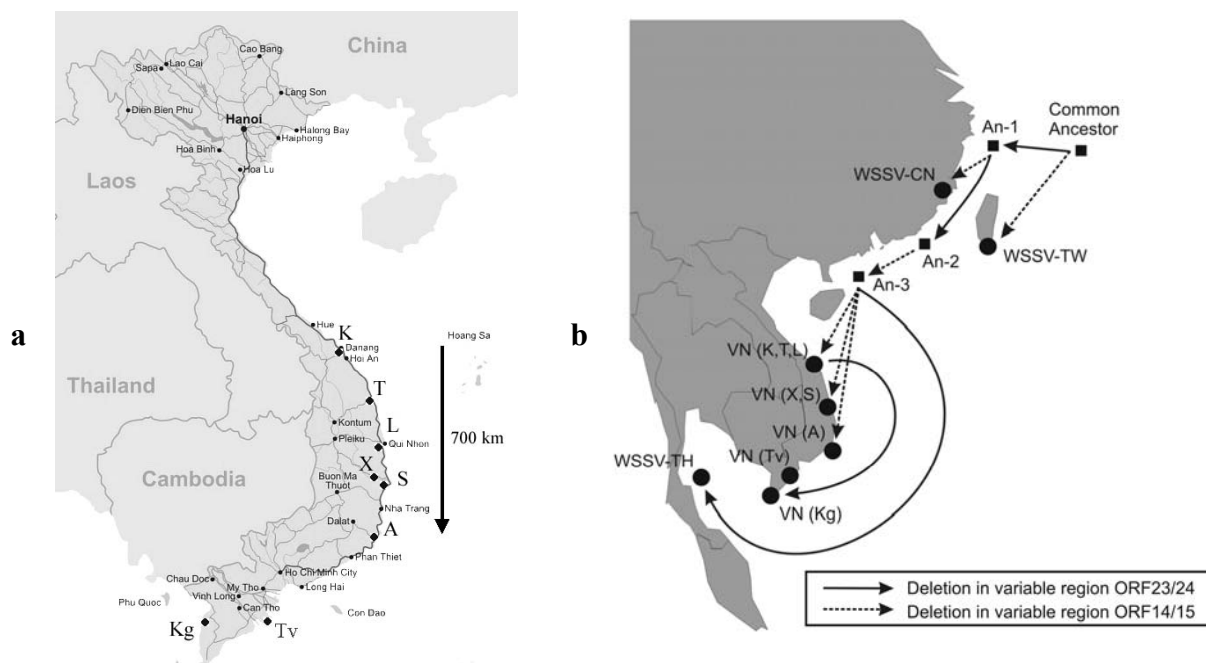
Figure 3.1. Map of Vietnam, showing the geographical origins of the isolates used for this study, indicated by K, T, L, X, S, A, Tv and Kg, respectively, according to Table 3.1 **(a)**. Model of spread of WSSV in Southeast-Asia from either site of the Taiwan Strait towards the west. Circular dots represent identified isolates, while square dots represent hypothetical isolates. Each line represents a single-step deletion. Ancestor is abbreviated as An, WSSV-VN isolates are abbreviated as VN **(b)**.

over seven different provinces, which are located along the coast from Central- to South-Vietnam (Fig. 3.1a: K till Kg). Three shrimp from each pond were chosen randomly from juvenile *P. monodon* showing gross signs of WSSV infection. All collected shrimp were tested positive for WSSV using a single-step PCR. Therefore, from each pond one shrimp was chosen at random as representative for that pond and used for further analysis. All WSSV-VN isolates, including the abbreviations used in this paper, are listed in Table 3.1.

The WSSV genomic loci, which were shown to be variable in their genetic make-up among different WSSV isolates (Marks *et al.*, 2004), were used for our analysis. These loci were studied in detail for each of the VN isolates by PCR amplification, cloning and sequence analysis. The variable loci screened for can be divided into (*i*) a genomic region prone to large deletions, referred to as "variable region ORF23/24" (this region is called "13 kb deletion" by Marks *et al.* (2004)), (*ii*) a genetic variable region, which will be referred to as "variable region ORF14/15", (*iii*) a genomic region encoding a putative transposase and (*iv*) the VNTRs located in ORF75, ORF94 and ORF125. Furthermore, we analyzed a conserved genomic fragment encoding part of WSSV DNA polymerase (*v*). The data for each of these loci will be dealt with separately. To reduce the possibility that the VN isolates have major genetic differences at loci which were not screened for, a detailed restriction enzyme analysis (RFLP) was performed for one of the VN isolates (WSSV-VN-T) and the result was compared to WSSV-TH.

(i) Variable region ORF23/24

Previously, this genomic region was shown to contain deletions of ~1.2 kb and ~13.2 kb in WSSV-CN and WSSV-TH, respectively, compared to WSSV-TW (Fig. 3.2a; Marks *et al.*, 2004). Three other unique deletions in this region were reported in Chinese isolates collected in Tong'an and Anhui in Southeast-China (2001) (Fig. 3.2a: WSSV-CN -A till -C; Lan *et al.*, 2002).

We mapped this locus initially in WSSV-VN isolate K by performing PCR reactions with primer pairs equally distributed over the genomic region 2332-15861 (WSSV-TW coordinates), which harbors the "variable region ORF23/24". Based on the primer pairs that failed to give a product, the flanking primer set "VR23/24-1" (Table 3.2b; Fig. 3.2a) was used to exactly pinpoint the coordinates of the deletion. Cloning and sequencing of the ~2.3 kb fragment obtained in the PCR revealed that a deletion of 8,539 bp existed in WSSV-VN isolate K compared to WSSV-TW (Fig. 3.2a). The flanking sequences present in the ~2.3 kb fragment were 100% homologous to the sequences of WSSV-TW and WSSV-CN. Based on this result, a new PCR reaction was performed with primer set "VR23/24-screen" (Table 3.2b; Fig. 3.2a) flanking this deletion to specifically detect this deletion in all VN isolates. The amplified fragment had a similar size of 548 bp for the six VN-central WSSV isolates, indicating that they have a deletion of about 8,539 bp in this locus compared to WSSV-TW (Fig. 3.2a). The VN-south isolates (Tv and Kg) failed to give a product in this PCR reaction.
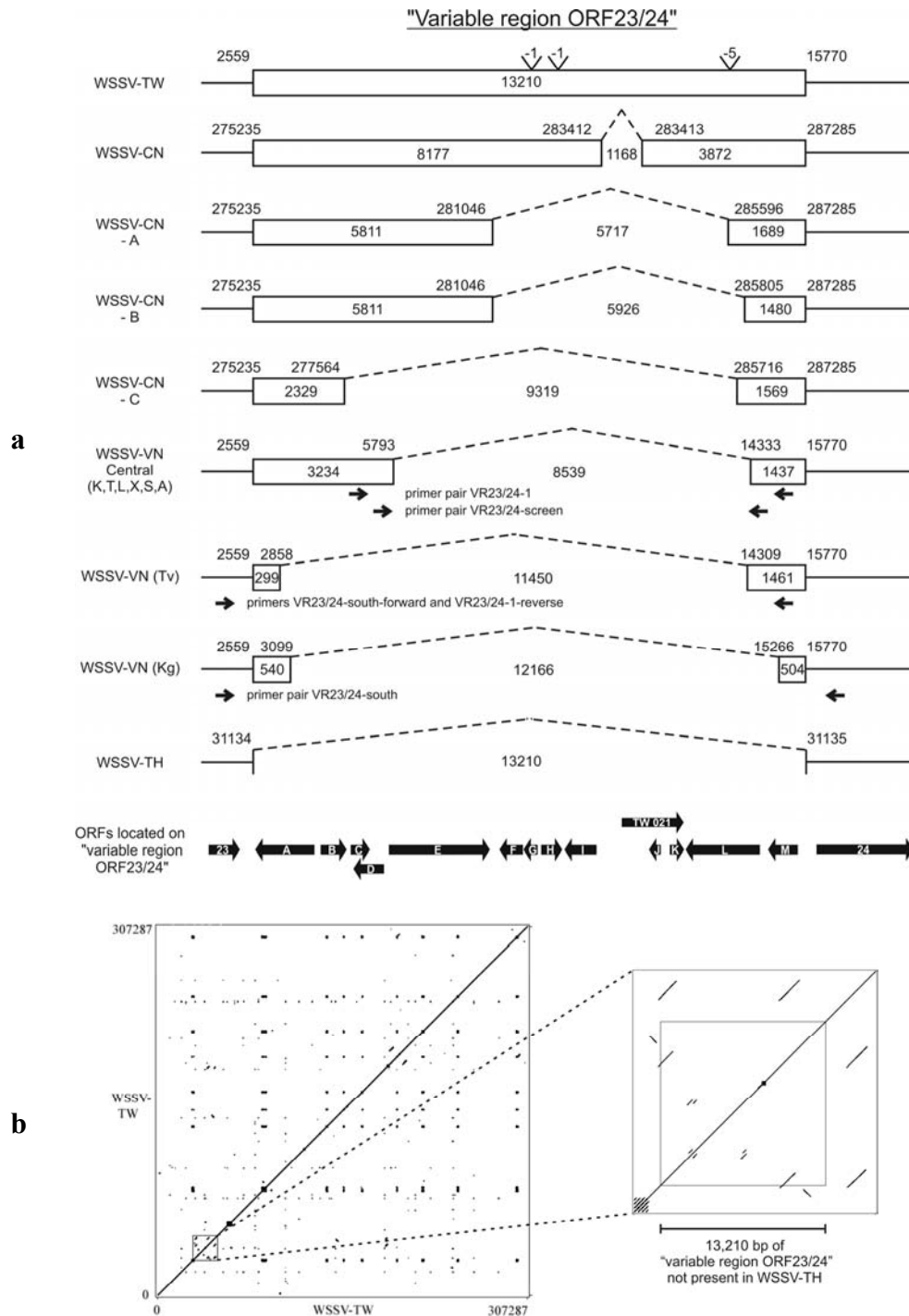
Figure 3.2. Schematic representation of the "variable region ORF23/24" of WSSV-TW, WSSV-CN, WSSV-TH, three different isolates from China 2001 (WSSV-CN - A, - B and - C, respectively: map numbers are in accordance with WSSV-CN; Lan *et al.*, 2002) and the VN isolates. The map numbers, indicated above each isolate, are in accordance with the numbers in the NCBI databank for the genomic sequence of each isolate. The coordinates of the WSSV-VN isolates are according to the WSSV-TW annotation. The length of the fragments is indicated within boxes or sequences. The positions of the ORFs located in this region are indicated by closed arrows, which also represent the direction of transcription. ORFs are numbered in accordance with the numbering used by Marks *et al.* (2004). Open arrows represent primers **(a)**. Dotplot comparison of the nucleotide sequences of WSSV-TW to itself (the adenine residue at the translation initiation codon of VP28 was designated as starting point for the numbering of WSSV-TW in this dotplot), including an enlargement (of original WSSV-TW coordinates 425-20425) **(b)**.

41

Table 3.2. Primers used during PCR analysis for WSSV screening **(a)** and for the variable loci of WSSV **(b)**

a. WSSV screening.

Primer pair name	Primer name	Sequence (5'-3')	Annealing temperature (^{0}C) / elongation time (s)	WSSV-TH sequence coordinates	Size (bp) of PCR product
16s rRNA	16S-FW 16S-RV	GTGCGAAGGTAGCATAATC CTGCTGCAACATAAGGATAC	52 / 50		414
VP26	VP26-FW VP26-RV	ATGGAATTTGGCAACCTAACAAACCTG GGGCTGTGACGGTAGAGATGAC	52 / 50	228835-228809 228532-228553	304

b. WSSV variable loci.

Primer pair name	Primer orientation	Sequence (5'-3')	Annealing temperature (^{0}C) / elongation time (s)	WSSV Sequence coordinates	Size (bp) of PCR product
VR23/24-1	Forward Reverse	ATGGGCTCTGCTAACTTG ATGATTGTATTCGTCGAAGG	50 / 360	4359-4376* 15191-15172*	10833*
VR23/24-screen	Forward Reverse	CACACTTGAAAAATACACCAG GTAAGTTTATTGCTGAGAAG	49 / 65	5503-5523* 14590-14571*	9088*
VR23/24-south	Forward Reverse	CTACAACGGCCAAGTCAT CGCAATTCTCCTCGCAGTT	49 / 100	30701-30718[†] 32255-32237[†]	1555[†]
VR14/15-screen	Forward Reverse	GAGATGCGAACCACTAAAAG ATGGAGGCGAGACTTGC	49 / 75	22904-22923[†] 24157-24141[†]	1254[†]
Transposase	Forward Reverse	GTGGATAATATTCGTCTTCAAC CTCAAAGACAACGACATTAG	55 / 120	253988-254009[†] 254138-254119[†]	1489* (151[†])
ORF75-flank	Forward Reverse	GAAGCAGTATCTCTAACAC CAACAGGTGCGTAAAAGAAG	49 / 80	107875-107893[†] 108742-108723[†]	868[†]
ORF94-flank	Forward Reverse	GTGCCGCAGGTCTACTC CATACGACTCTGCTTCTTG	51 / 80	142656-142672[†] 143337-143319[†]	682[†]
ORF125-flank	Forward Reverse	CGAAATCTTGATATGTTGTGC CCATATCCATTGCCCTTCTC	52 / 100	187791-187811[†] 188442-188423[†]	652[†]
Polymerase	Forward Reverse	CAATATTACACGCCCTTCAG GCTTGCATGATTTTTCTCC	49 / 60	35867-35886[†] 36370-36352[†]	504*,[†]

*WSSV-TW sequence coordinates or WSSV-TW sizes
[†]WSSV-TH sequence coordinates or WSSV-TH sizes

We mapped the "variable region ORF23/24" in isolates Tv and Kg using a similar method used to map the deletion for WSSV-VN isolate K. Cloning and sequencing of the ~1.6 kb PCR product obtained with primers VR23/24-south-forward and VR23/24-1-reverse (Table 3.2b; Fig. 3.2a) showed that isolate Tv has a deletion of 11,450 bp relative to the WSSV-TW genome (Fig. 3.2a). A PCR with the primer pair "VR23/24-south" (Table 3.2b; Fig. 3.2a) for isolate Kg resulted in a ~2.6 kb PCR product, which after cloning and sequencing showed that this isolate contains a deletion of 12,166 bp relative to the WSSV-TW genome (Fig. 3.2a). We previously mapped 5 SNPs and a 1 bp deletion within WSSV-TW coordinates 16447-16773 (flanking the deletion) compared to WSSV-CN and WSSV-TH (Marks *et al.*, 2004). With respect to these genetic differences, isolate Kg is identical to WSSV-CN and WSSV-TH, suggesting that this isolate is more closely related to these isolates than to WSSV-TW.

Dotplot analysis showed that, except for the *hr*s (van Hulten *et al.*, 2001a), the genomic region in WSSV-TW in which these deletions occur contains the most direct and inverted repeats of the entire WSSV genome (Fig. 3.2b). However, for the deletion in the VN-south isolates Tv and Kg as well as in the six VN-central isolates no direct repeats were identified within 300 bp flanking the putative recombination sites in WSSV-TW that could be involved in recombination (Fig. 3.2b; sequence data not shown).
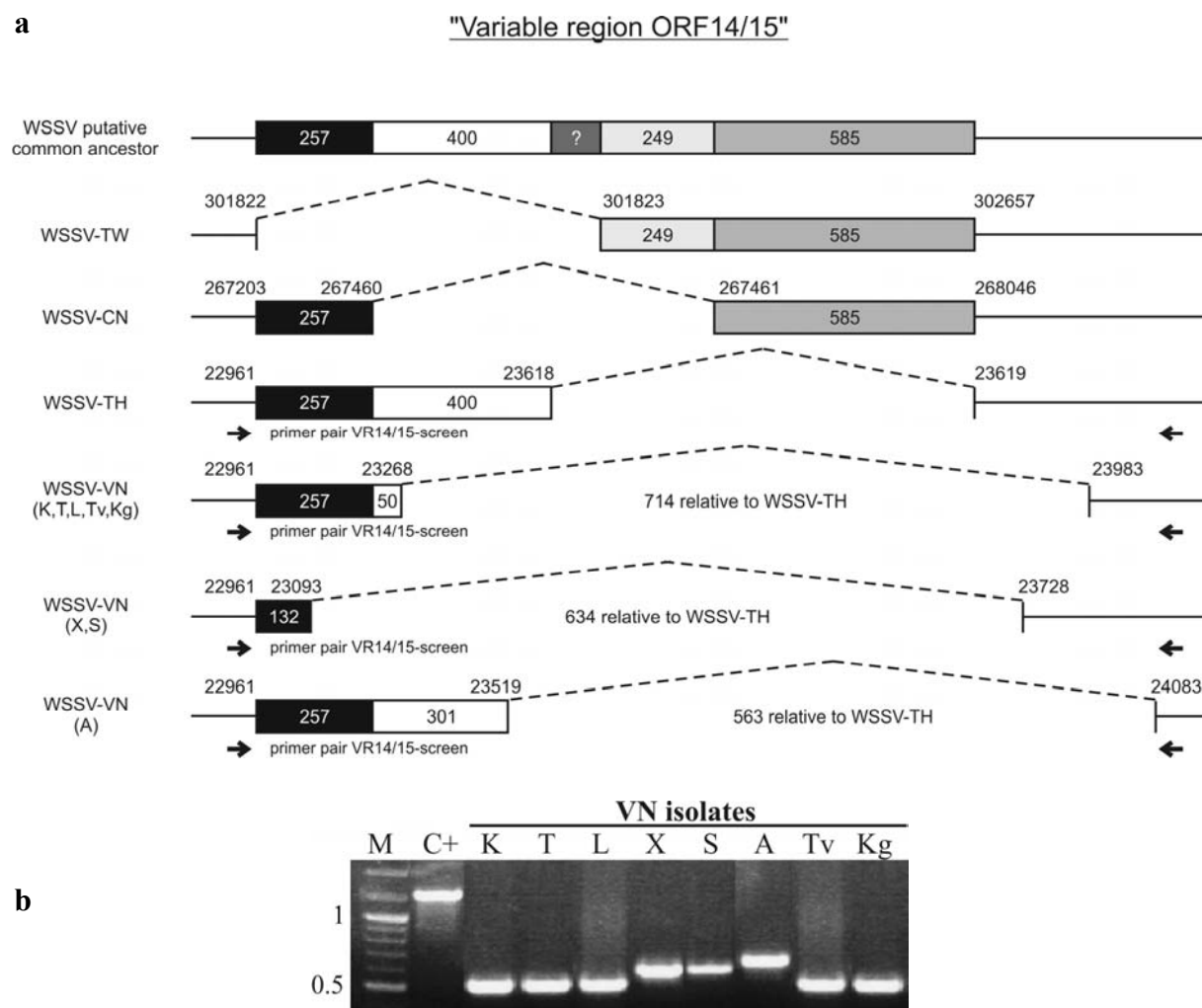
(ii) Variable region ORF14/15

The "variable region ORF14/15" is centered in a region of 842 bp in size in WSSV-CN, of which 257 bp of its 5' end is only present in WSSV-TH, while the remaining 585 bp of its 3' end is only present in WSSV-TW (Fig. 3.3a; Marks *et al.*, 2004). This locus was thought to be a variable region prone to recombination (Marks *et al.*, 2004). However, a partly characterized isolate recently studied by our laboratory, contains at least all unique sequences present in this locus, suggesting that WSSV-TW, WSSV-CN and WSSV-TH are derived from a common ancestor by deletions of various sizes (Fig. 3.3a). Because WSSV-TW, WSSV-CN and WSSV-TH each contain unique sequences, these isolates seem to be distinct and probably evolved separately. Using the same strategy as used for the "variable region ORF23/24", this locus was mapped for all VN isolates using primer set "VR14/15-screen" (Table 3.2b; Fig. 3.3a). WSSV-TH DNA, taken as positive control for the PCR, showed the expected fragment of 1,254 bp, whereas the VN isolates showed fragments of different sizes ranging from ~500 bp to ~700 bp (Fig. 3.3b). Cloning and sequencing of these fragments revealed that all VN isolates had deletions relative to WSSV-TH (Fig. 3.3a). The flanking sequences of the deletions present in the ~500 bp to ~700 bp fragments were identical to the sequences of WSSV-TH. The VN isolates K, T, L, Tv and Kg had the same deletion of 714 bp, VN isolates X and S had a deletion of 634 bp, while VN isolate A had the smallest deletion of 563 bp compared to WSSV-TH (Fig. 3.3a).

(iii) A genomic region coding for a putative transposase

The genome of WSSV-TW encodes a putative transposase, which is not present in WSSV-CN and WSSV-TH. Using primer pair "Transposase" (Table 3.2b) flanking the transposase gene in the WSSV-TW genome, we obtained a PCR fragment of ~150 bp for all VN isolates (data not shown). A fragment of similar size was obtained with WSSV-TH DNA, used as positive control. Therefore, we conclude that all VN isolates do not contain this particular transposase sequence.

(iv) Genetic variation in VNTR loci

Three non-*hr* unidirectional tandem repeats, in the region coding for ORF75, ORF94 and ORF125, have been shown to be variable in the number of repeat units (RUs) between the WSSV isolates identified thus far (Table 3.3a; Wongteerasupaya *et al.*, 2003; Marks *et al.*, 2004). The repeats are positioned in the middle of the ORFs, which have non-repeated 5' and 3' ends. For both ORF75 and ORF94 around 50% of the coding region consists of repeats, while for ORF125 around 20% of the coding region consists of repeats. Differences in the number of RUs do not cause frameshifts for the respective ORFs, since the length of these RUs is always a multimer of 3 bp. The protein encoded by ORF75 has been shown to be present in WSSV virions (Huang *et al.*, 2002b). ORF94 may have a similar function as ORF75, as the repeat units of both ORFs share a common motif at the protein level consisting

Figure 3.3. Schematic representation of the "variable region ORF14/15" of the WSSV putative common ancestor, WSSV-TW, WSSV-CN, WSSV-TH, and the VN isolates. Different gray tints represent unique sequences within the WSSV putative common ancestor, WSSV-TW, WSSV-CN, 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. The coordinates of the WSSV-VN isolates are according to the WSSV-TH annotation. The length of the fragments is indicated within boxes or sequences. Open arrows represent primers **(a)**. PCR on the "variable region ORF 14/15" using genomic DNA of the WSSV-VN isolates as template. The lanes indicate the respective VN isolate used. C+ is the same PCR on genomic DNA of WSSV-TH, used as positive control for the PCR. M represents a 100 bp DNA marker of which some of the sizes are indicated next to the gels (in kb) **(b)**.

of four basic amino acids (R or K) followed by two alanines, two or three prolines and a stretch of acidic amino acids (E or D).

To study the VN-central isolates for each of these loci, we performed a PCR reaction with a specific primer set (Table 3.2b) flanking the non-*hr* unidirectional tandem repeats. The results for ORF75, ORF94 and ORF125 are shown in Fig. 3.4. For all three loci a major band was observed for each isolate, often different in size among isolates. The PCR fragments of all VN isolates were cloned, sequenced and aligned. The sequenced regions flanking the tandem repeats (between the primers used and the actual repeats) on both the 5' and the 3' end showed 99.6-100% nucleotide identity with the corresponding sequences of WSSV-TW,

Table 3.3. Number of repeat units present within the non-*hr* unidirectional repeats of ORF75, ORF94 and ORF125.

a. Number of RUs in each ORFs.

WSSV isolate	ORF75 (45 bp and 102 bp*) / 107965-108675[†]	ORF94 (54 bp*) / 142744-143067[†]	ORF125 (69 bp*) / 187899-188312[†]
TW	21 (16 and 5[‡])	6	8
CN	15 (11 and 4[‡])	12	8
TH	12 (9 and 3[‡])	6	6
VN-central			
K	5 (3 and 2[‡])	10	6
T	5 (3 and 2[‡])	17	5
L	5 (3 and 2[‡])	10	6
X	5 (3 and 2[‡])	7	7
S	14 (10 and 4[‡])	7	7
A	6 (4 and 2[‡])	10	6

*Length of RUs
[†]WSSV-TH coordinates of total repeat
[‡]Number of 45 bp and 102 bp RUs, respectively

b. Number and position of the RUs located within the non-*hr* unidirectional repeat of ORF75.

WSSV isolate	Number of repeat units	Positioning of 45 bp and 102 bp RUs[†]										
		1	2	3	4	5	6	7	8	9	10	11
TW	21	45	102	4*45	102	3*45	102	2*45	102	4*45	102	2*45
CN	15	45	102	4*45	102	2*45	102	2*45	102	2*45		
TH	12	45	102	4*45	102	2*45	102	2*45				
VN-central												
K	5	102	45	102	2*45							
T	5	102	45	102	2*45							
L	5	102	45	102	2*45							
X	5	102	45	102	2*45							
S	14	45	102	4*45	102	45	102	2*45	102	2*45		
A	6	102	2*45	102	2*45							

[†]Number of successive tandem repeat units of 45 bp are summarized as x*45

c. Genotype of the SNP of each RU within the non-*hr* unidirectional repeat of ORF94.

WSSV isolate	Number of repeat units	Successive RU*																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
TW	6	T	T	T	G	T	T											
CN	12	T	T	G	G	G	G	G	G	T	T	T	T					
TH	6	T	G	G	G	T	T											
[†]**Thai**																		
Sur #1	9	T	T	T	T	G	T	T	G	T								
Sur #2	8	T	T	G	T	T	G	G	T									
Chu #2	7	T	T	T	G	T	G	T										
Chu #3	8	T	T	G	T	T	G	G	T									
Chu #4	6	T	T	T	G	G	G											
VN-central																		
K	10	G	G	G	T	T	T	G	G	T	T							
T	17	G	T	T	T	T	G	T	T	T	G	T	G	G	G	G	T	T
L	10	G	G	G	G	G	G	G	G	G	G							
X	7	T	T	T	T	G	T	T										
S	7	T	T	T	T	G	T	T										
A	10	G	G	G	T	G	G	T	T	T	T							

*Genotype of each of the successive RU at position 48 (guanine (G) or thymine (T)) is shown
[†]Data on the Thailand isolates are cited from Wongteerasupaya *et al.* (2003). These isolates originate from different ponds in Surat Thani or Chumporn (abbreviated as Sur and Chu, respectively), Thailand, 2000

d. Genotype of the SNPs of each RU within the non-*hr* unidirectional repeat of ORF125.

WSSV isolate	Number of repeat units	Successive RU*[†]						
		C	**D**	**E**	**F**	**G**	**H**	**I**
TW	8	TGGTC	–	TGGTC	TGGTC	–	TTGGT	CGAGT
CN	8	TGGTC	TTGGT	TGGTC	–	TTGGT	TTGGT	–
TH	6	TGGTC	–	–	–	–	TTGGT	CGAGT
VN-central								
K	6	TGGTC	–	TGGTC	–	–	TTGGT	–
T	5	TGGTC	–	–	–	–	TTGGT	–
L	6	TGGTC	–	–	–	–	TTGGT	CGAGT
X	7	TGGTC	–	TGGTC	–	–	TTGGT	CGAGT
S	7	TGGTC	–	TGGTC	–	–	TTGGT	CGAGT
A	6	TGGTC	–	–	–	–	TTGGT	CGAGT

* The order of the RUs is kept, but the RUs are categorized (C-I) by genotype, starting from the 3rd RU (C) to the penultimate RU (I). In case no RU is present, it is indicated by –
[†]Genotype of each successive RU at position 8, 18, 25, 66 and 69, respectively, is shown
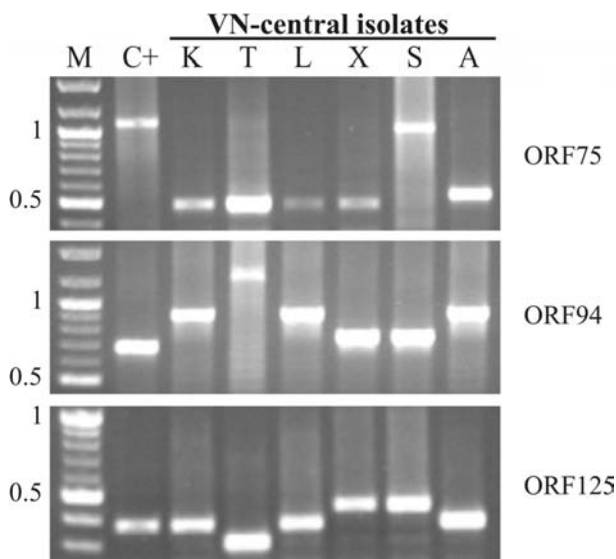
Figure 3.4. PCR on the non-*hr* unidirectional repeats of ORF75, ORF94 and ORF125, using genomic DNA of the WSSV VN-central isolates as template. The lanes indicate the respective VN isolate used. C+ is the same PCR on genomic DNA of WSSV-TH, used as positive control for the PCR. M represents a 100 bp DNA marker of which some of the sizes are indicated next to the gels (in kb).

WSSV-CN and WSSV-TH. This indicates that the correct fragment had been amplified for each of the three loci of the VN-central isolates, eliminating the possibility of false annealing of the primers.

ORF75: For all WSSV isolates characterized thus far, ORF75 has two types of RUs with a length of 102 bp and 45 bp, respectively (Table 3.3b). The first 45 nucleotides of the 102 bp RUs are identical to the RUs of 45 bp. Comparing all RUs within one isolate, they contained SNPs at position 3, 15, 30, 40, 42 and 44, the RUs of 102 bp have an extra SNP at position 83. Each of the RUs can be recognized by its specific SNPs.

The number of RUs present in ORF75 of the WSSV-VN-central isolates is summarized in Table 3.3a, while Table 3.3b shows the exact order of appearance of the 45 bp and 102 bp RUs. The number of RUs identified for each isolate corresponded to the respective sizes of their PCR fragments shown in Fig. 3.4. VN isolates K, T, L and X are identical at this point. VN isolate A has an extra RU of 45 bp, which is, based on the SNPs, located after the second repeat unit (sequence data not shown). The VN isolate S has a higher number of RUs and, based on the SNPs, more resembles the genotype of WSSV-CN (sequence data not shown).

ORF94: In all WSSV isolates characterized thus far, ORF94 has tandem RUs of 54 bp with a SNP at position 48 (either guanine or thymine) when comparing the RUs mutually within one

isolate (Table 3.3c). The number of RUs was highly variable between the various isolates for which this locus has been characterized: WSSV-TW, WSSV-CN, WSSV-TH, and 55 other isolates originating from Thailand. The number of RUs varied from 6 to 20 repeat units (van Hulten *et al.*, 2000b; Wongteerasupaya *et al.*, 2003; Marks *et al.*, 2004).

The WSSV-VN-central isolates contained between 7 and 17 RUs (Table 3.3a), corresponding to the respective sizes of their PCR fragments (Fig. 3.4). The identity of the nucleotide at position 48 of each of the VN isolates is shown in Table 3.3c. Isolates X and S are identical, while the other isolates, although some having the same number of RUs, all have a unique pattern of the nucleotides at position 48. The VN isolates K, T and L had a thymine deletion at position 143149 (WSSV-TH coordinates), located in the 3' end flanking the repeat. As this is outside the coding region, it will not cause a frameshift in ORF94.

ORF125: This ORF contains tandem RUs of 69 bp, of which the first two as well as the last can be recognized by their specific SNPs when comparing the RUs mutually within one isolate (Table 3.3d). The other RUs (the 3rd till the penultimate) contain SNPs at position 8, 18, 25, 66 and 69 (Marks *et al.*, 2004). The WSSV-VN-central isolates contained between 5 and 7 RUs (Table 3.3a), corresponding to the respective sizes of their PCR fragments (Fig. 3.4). VN isolates X and S, as well as VN isolates A and L, are identical in this locus (Table 3.3d). The genotype of the VN isolates A and L is identical to the genotype of WSSV-TH (Table 3.3d).

(v) Fragment encoding part of DNA polymerase

To further classify the WSSV-VN isolates, a PCR was performed on a conserved genomic fragment encoding part of WSSV DNA polymerase using primer set "Polymerase" (Table 3.2b). Within this genomic fragment, a single nucleotide deletion occurs in WSSV-CN (WSSV-TH coordinates 36030) compared to WSSV-TW and WSSV-TH, causing a frameshift in the polymerase gene (Chen *et al.*, 2002b; Marks *et al.*, 2004). The WSSV-VN isolates gave a PCR fragment of the similar size as the positive control WSSV-TH. Cloning and sequencing of the 8 PCR fragments of the central and south VN isolates failed to detect an adenine deletion as is present in WSSV-CN. The PCR fragments showed 100% nucleotide identity with the respective fragments of WSSV-TW and WSSV-TH.

Restriction enzyme analysis of VN isolate T

The RFLP analysis between WSSV-TH and WSSV-VN-T is shown in Fig. 3.5. The *Bam*HI restriction pattern of WSSV-TH exactly matches the expected pattern based on the complete nucleotide sequence (van Hulten *et al.*, 2001a), except for the 3 smallest fragments which are not visible due to their estimated size of <1 kb. Two clear polymorphisms (shifts) are visible between WSSV-TH and VN isolate T, indicated with A and B, respectively. Shift A, in which a fragment of ~27.5 kb for VN isolate T shifts to ~24.5 kb for WSSV-TH, can be explained
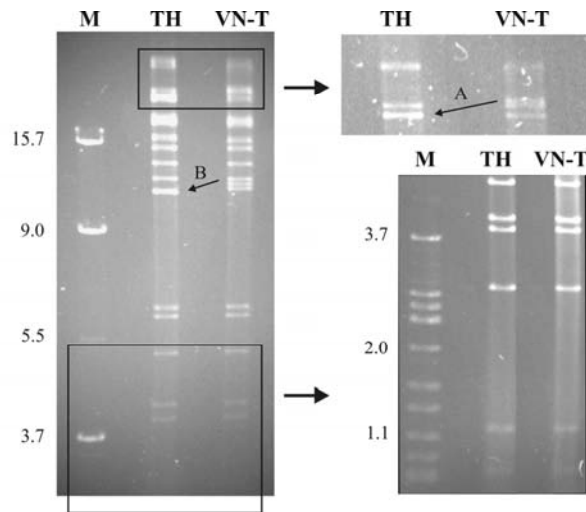
Figure 3.5. WSSV genomic DNA of WSSV-TH (TH) and WSSV-VN isolate T (VN-T) digested with *Bam*HI. M represents a molecular size standard (lambda digested with *Bam*HI-*Eco*RI-*Hin*dIII), of which some of the sizes are indicated next to the gels (in kb). The enlargements focus on the major (>20 kb) and minor (<4 kb) fragments on similar gels. The clear band shifts between the two isolates are indicated by A and B.

by the observed sequence diversity in "variable region ORF 14/15" and "variable region ORF 23/24", which are both located on this large fragment. The ~-3 kb discrepancy is the sum of the observed differences in PCR mapping of both "variable regions" of ~0.7 kb and ~-3.7 kb, respectively (Fig. 3.3a & 3.2a). Shift B, in which a corresponding fragment has a size of ~11.2 kb for WSSV-TH and of ~11.8 kb for VN isolate T, can be explained by the sequence variation of the repeat in ORF94 (Table 3.3a). The difference of 11 RUs of each 54 bp results in a shift of 594 bp. The differences in the repeats in ORF75 and ORF125 are not clearly visible. ORF75 is located on a large fragment (~20 kb) for which the 350 bp difference in size will only show a minor shift, whereas the difference in the repeats of ORF125 between WSSV-TH and VN isolate T is marginal (138 bp).

Discussion

Genomic analyses of WSSV showed that conserved genes, often used in molecular epidemiological studies to unravel evolutionary relationships by phylogenetic analysis, are too homologous to use for this purpose in case of WSSV (Marks *et al.*, 2004). For example, the complete DNA polymerase gene of WSSV only contains 3 SNPs and a 1 bp and 3 bp deletion when comparing this gene for the three completely sequenced WSSV isolates (Chen *et al.*, 2002b; Marks *et al.*, 2004). Similar high homologies were found for other conserved WSSV genes (Chang *et al.*, 2001; Marks *et al.*, 2004). Moreover, also the major structural protein genes, which for some virus families show a relative high number of mutations due to antigenic drift or adaptation to different hosts, show 99.5%-100% nucleotide identity between several geographical WSSV isolates (Moon *et al.*, 2003; Marks *et al.*, 2004). These data indicate that the isolates of WSSV identified thus far are very closely related and probably evolved recently from a common ancestor. The DNA polymerase sequences obtained from all VN isolates, showing 100% identity with WSSV-TW and WSSV-TH, further confirmed this observation. Therefore, we chose the most variable loci of WSSV to classify new WSSV isolates from Vietnam (Marks *et al.*, 2004). The RFLP analysis between WSSV-TH and WSSV-VN-T (Fig. 3.5) confirmed the high degree of homology among WSSV isolates, but indeed identified the major genomic insertions and deletions in WSSV-VN (isolate T).

Based on both "variable region ORF23/24" and "variable region ORF14/15", we propose a model to explain the genotypic changes of WSSV during its geographical spread from either site of the Taiwan Strait towards the west to Thailand between 1992 and 1995 (Fig. 3.1b). In this model, both loci evolved independently, and both deletions in the "variable regions" showed a progressive increase in length during the spread of WSSV. The WSSV common ancestor (Fig. 3.1b) contains a genotype similar to WSSV-TW in the "variable region ORF23/24" (Fig. 3.2a) and a genotype similar to the putative common ancestor in "variable region ORF14/15" (Fig. 3.3a). WSSV-TW evolved from this common ancestor by a deletion in "variable region ORF14/15", while WSSV-CN evolved by a deletion of ~1.2 kb in "variable region ORF23/24" (Fig. 3.1b: An-1) followed by a deletion in "variable region ORF14/15". Based on the observation that the genotypes of the VN isolates seem to have evolved from a genotype similar to WSSV-TH in "variable region ORF14/15" by separate unique deletions of different sizes, the VN isolates and WSSV-TH probably have a common lineage, which branched off at an early stage from WSSV-TW and WSSV-CN. However, the extra sequences in the "variable region ORF23/24" present in the VN isolates compared to WSSV-TH exclude the possibility that the WSSV-VN isolates are derived from WSSV-TH. Therefore, WSSV-TH and the WSSV-VN isolates probably have a common ancestor An-3 (Fig. 3.1b), which could contain the genotype of WSSV-TH in "variable region ORF14/15", but the ~8.5 kb deletion similar to the VN-central isolates in "variable region ORF23/24". Within the three different WSSV-VN genotypes in "variable region ORF14/15", each contains unique sequences, and thus probably evolved separately. Therefore, WSSV entered Vietnam by multiple introductions from the common ancestor An-3, from where it further spread within Vietnam (VN isolate Kg; Fig. 3.1b). WSSV isolates, collected along the further coast of Southeast-Asia (i.e. isolates from North-Vietnam, China (Hainan) and Cambodia), should be genotyped to confirm and further detail this model.

The mechanism(s) by which the changes or (gradual) deletions in both "variable regions" occur is unclear. For WSSV-TH, it was suggested that the deletions in "variable region ORF23/24" might have occurred by homologous recombination, as a direct repeat is present at both ends of the deletion in WSSV-TW (Marks *et al.*, 2004). However, no direct repeats that could be involved in recombination were identified for the deletion in the VN-south isolates Tv and Kg as well as in the six VN-central isolates (Fig. 3.2b). Maybe the deletions in the "variable region ORF23/24" can be explained by the genomic pressure on the virus to discard redundant sequences, as Fig. 3.2b show that WSSV-TW contains a lot of duplicated sequences and ORFs (especially genes of WSSV gene family 4; van Hulten *et al.*, 2001a) in this region. It is also possible that the host species or an intermediate host has an effect on the size of the deletion, as WSSV-CN -A (*Metapenaeus ensis*), -B (*P. japonicus*) and -C (*P. vannamei, P. monodon, P. chinensis*) were isolated from different host species (Fig. 3.2a; Lan *et al.*, 2002). However, within one host species, WSSV isolates can show different sizes of deletion, as WSSV-TW, WSSV-TH, WSSV-CN -C and the VN isolates

were all obtained from *P. monodon*, and WSSV-CN and WSSV-CN -B were both isolated from *P. japonicus*. To date, there seems to be no difference in host range between the characterized WSSV isolates (Wang *et al.*, 1998; Wang *et al.*, 1999a; Chen *et al.*, 2000; Lan *et al.*, 2002; Hameed *et al.*, 2003).

Based on the genetic make-up in both "variable regions" and the thymine deletion shared by the isolates K, T and L in the 3' flanking region of the repeat located in ORF94, three groups of VN-central isolates can be distinguished ((K, T, L) and (X, S) and (A); Fig. 3.1b). Within these groups, each of the non-*hr* unidirectional tandem repeats located in ORF75, ORF94 and ORF125 seem to have their own, independent genesis in terms of insertion or deletion of repeat units (Table 3.3). Possibly, insertion or deletion of repeat units are generated during homologous recombination or replication slippage, as is proposed for repeats such as the baculovirus homologous repeats (*hr*s) (Garcia-Maruniak *et al.*, 1996) and the herpesvirus direct repeats (DRs) (Umene, 1991).

Compared to the other two non-*hr* unidirectional tandem repeats (ORF94, ORF125), the repeats in ORF75 seem to be rather conserved within and between the three groups of VN-central isolates. The additional repeat unit in VN isolate A could be explained by a single insertion event. The large number of repeat units present in ORF75 for WSSV-VN-S is surprising. Especially because the VN isolates X and S, whose geographical origins are very close (~10 km) and maybe even originate from postlarvae from the same supplier, are in all other loci screened for completely identical. Analysis of more WSSV isolates at this locus from different infected shrimp from the same pond may provide clarification whether this is the common genotype of WSSV isolates derived from pond S or whether it is an irregularity. Also for the repeats in ORF125, the genotypic differences in VN isolates can be explained by a one step deletion or insertion of a single repeat unit (Table 3.3). Analysis of the genotypes present within the WSSV-VN group K, T, L suggests that this locus has a higher mutation frequency than ORF75.

The largest genomic variation among the VN-central isolates was observed for the non-*hr* unidirectional tandem repeats located in ORF94. The number of repeat units within ORF94, as well as the SNP located at position 48, already appeared highly variable for WSSV isolates within Thailand (Wongteerasupaya *et al.*, 2003). Also between the isolates characterized within Vietnam, a wide range of genotypic variation was found for this locus without any obvious correlation with its geographical location. It is interesting to note that the repeats of ORF94 are highly variable in number, whereas the repeat in ORF75 seems to be more stable, although both repeat regions share structural properties on the protein level. In conclusion, the repeats of ORF75 and ORF125, each having its own mutation dynamics different from both more stable "variable regions", seem suitable to study WSSV spread at a more local or regional scale.

This paper shows the potential to use genetic markers to study WSSV epidemiology and ecology. However, more information about the mode of spread of WSSV is necessary to

further understand the relationship between the VN isolates. Often, WSSV infection in a pond can be traced back to the broodstock supplier or the postlarvae producers. Therefore, on a regional scale, most likely the virus spreads in a myriad way during the turnover of shrimp. However, on a global scale, this study provides support for the contention that WSSV originated from either site of the Taiwan Strait and evolved concurrently with its geographical spread over time in Southeast-Asia.

Materials and Methods

Infected shrimp sampling
The origin of the collected WSSV infected shrimp (*Penaeus monodon*) analyzed in this study is shown in Table 3.1. The shrimp were cleaned with ethanol and transported in liquid nitrogen from the respective ponds to Can Tho University (Vietnam), where they were stored at -80 °C till further processing.

DNA extraction
DNA extracts of collected shrimp were obtained from muscle tissue. A small piece (approximately 50 mg) of the tail of dead shrimp was homogenized using a disposable rod and mixed with 200 µl 5% (w/v) Chelex X-100 resin (BioRad) and 16 µl proteinase-K (20 mg ml^{-1} stock). This mixture was incubated overnight at 56 °C, followed by 10 min at 95 °C to inactivate the proteinase-K and centrifugation for 1 min at 18,000 g to pellet cellular debris. One µl of the supernatant was used in PCR reactions.

PCR analysis of WSSV-infected shrimp
To screen for WSSV, we developed a standardized PCR-based WSSV detection protocol. One µl of DNA extract was tested in two similar single-step PCR reactions with a shrimp 16S rRNA or a WSSV VP26 primer pair (Table 3.2a), using *Taq* DNA polymerase (Promega). The 16S rRNA primer pair amplifies a shrimp mitochondrial DNA fragment coding for the 16S ribosomal RNA, and is used as a positive control for the presence of host DNA. The VP26 primer pair amplifies part of the WSSV VP26 ORF (van Hulten *et al.*, 2000a), and is used to screen for WSSV positive shrimp. PCR conditions used and sizes of the PCR products are shown in Table 3.2a.

PCR analysis for WSSV variable loci
PCR on the genomic variable loci of WSSV was performed with 1 µl DNA extracts, using *Taq* DNA polymerase (Promega). The specific primer sets, PCR conditions used and sizes of the PCR products are shown in Table 3.2b.

Cloning of PCR products

PCR products were purified from 1% agarose gels using a DNA extraction Kit (MBI Fermentas). These products were subsequently cloned into DH5α competent cells using the pGEM-T easy vector system I (Promega). Plasmids containing the correct insert, as screened by restriction enzyme analysis and/or by colony PCR, were prepared for sequencing by purification with the High Pure Plasmid Isolation Kit (Roche).

Virus production and purification

The virus isolate WSSV-TH used in this study originated from infected *P. monodon* imported from Thailand in 1996 and was obtained as described before (van Hulten *et al.*, 2000b). The virus WSSV-VN isolate T (Table 3.1) originated from a single infected *P. monodon*. Tissue of a WSSV-VN-T infected *P. monodon* was homogenized in 330mM NaCl. After centrifugation at 1,700 x g for 10 min the supernatant was filtered (0.45 μm filter; Schleicher & Schuell) to obtain the virus. Crayfish *Orconectes limosus* or *Astacus leptodactylus* were injected intramuscularly with a lethal dose of WSSV (WSSV-TH or WSSV-VN-T), using a 26-gauge needle (Microfine B&D). Virus was isolated and processed according to published procedures (van Hulten *et al.*, 2000b).

Purification of viral DNA and restriction enzyme analysis

Viral DNA was isolated from purified virions as described by van Hulten *et al.* (2000b). WSSV DNA was digested with *Bam*HI (Invitrogen) and fragments were separated by electrophoresis in a 0,6% agarose gel at 40 V (1.3 V cm^{-1}) for 20 h. After separation, the gels were stained with ethidiumbromide (0.5 μg ml^{-1} in Tris-Acetate-EDTA (TAE)).

Sequencing and computer analysis

Plasmid clones were sequenced using universal T7 and/or Sp6 primers, or by primer walking in case inserts were >1.5 kb (BaseClear, the Netherlands). Sequence data were analyzed using the software package DNASTAR 4.2 (DNASTAR Inc.), the output was edited in GeneDoc, version 2.6.000 (Nicholas *et al.*, 1997). Complete WSSV sequences were obtained from the NCBI databank (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) using the acc. no. for WSSV-TW (AF440570), WSSV-CN (AF332093) and WSSV-TH (AF369029). Dotplot analysis was performed using PIPmaker (http://bio.cse.psu.edu/pipmaker/).

Chapter 4

Fitness and virulence of an ancestral White spot syndrome virus isolate from shrimp

Abstract

White spot syndrome virus, the type species of the virus family *Nimaviridae*, is a large dsDNA virus infecting shrimp and other crustaceans. Genomic analysis of three completely sequenced WSSV isolates identified two major polymorphic loci, "variable region ORF14/15" and "variable region ORF23/24". Here, we characterize a WSSV isolate originating from shrimp collected in Thailand in 1996 (TH-96-II). This isolate contains the largest WSSV genome (~312 kb) identified so far, mainly because of its sequences in both major polymorphic loci. Analysis of "variable region ORF14/15" suggests that TH-96-II may be ancestral to the WSSV isolates described to date. A comparison for virulence was made between TH-96-II and WSSV-TH, a well characterized isolate containing the smallest genome (~293 kb) identified at present. After injection of the isolates into *Penaeus monodon* the mortality rates showed that the median lethal time (LT_{50}) of TH-96-II was approximately 14 days, compared to 3.5 days for WSSV-TH. When both isolates were mixed in equal amounts and serially passaged in shrimp, WSSV-TH outcompeted TH-96-II within four passages. These data suggest a higher virulence of WSSV-TH compared to TH-96-II. The molecular basis for the difference in virulence remains unclear, but a replication advantage of the 19 kb smaller WSSV-TH genome could play a role.

Introduction

White spot syndrome virus (WSSV) is a large dsDNA virus belonging to the virus family *Nimaviridae*, genus *Whispovirus* (Mayo, 2002). WSSV causes serious economic losses in shrimp culture, as 100% cumulative mortalities can be reached within 3-10 days under farming conditions (Lightner, 1996). The virus has a remarkably broad host range among crustaceans, including many species of shrimp, crayfish, crab and lobster (Lo *et al.*, 1996b; Flegel, 1997; Wang *et al.*, 1998; Chen *et al.*, 2000; Maeda *et al.*, 2000; Hameed *et al.*, 2003).

Sequence analysis showed that WSSV contained a circular dsDNA genome of approximately 300 kb in size. The WSSV genome encodes about 184 open reading frames (ORFs) the function of most of which is still enigmatic (van Hulten *et al.*, 2001a; Yang *et al.*, 2001). Large regions, each consisting of a variable number of 250 bp repeat units, were identified dispersed along the viral genome (*hr*s). Similar to baculoviruses, these regions could be involved in transcription enhancement and/or DNA replication (Guarino & Summers, 1986; Kool *et al.*, 1993).

WSSV was first discovered in the Chinese province of Fujian in 1992, from where it quickly spread to other areas in Southeast-Asia and subsequently to shrimp farming areas all over the world (Cai *et al.*, 1995; Flegel, 1997). WSSV was reported from the United States in 1995 (Rosenberry, 1996), and from Central- and South-America since early 1999 (Rosenberry, 2000). In 2002, WSSV was also detected in Europe (France) and the Middle East (Iran) (Rosenberry, 2002). Thus far, the various geographic isolates of WSSV characterized in detail are very similar in morphology and proteome and belong to the same species *Whispovirus*. Limited differences in restriction fragment length polymorphism (RFLP) patterns have been reported, suggesting either a high degree of genomic stability or a recent emergence (Nadala & Loh, 1998; Lo *et al.*, 1999; Wang *et al.*, 2000b, c; Marks *et al.*, 2004). There seems to be no difference in host range between various characterized WSSV isolates (Wang *et al.*, 1998; Wang *et al.*, 1999a; Chen *et al.*, 2000; Lan *et al.*, 2002; Hameed *et al.*, 2003). Preliminary studies suggest differences in virulence between various WSSV isolates, but direct comparisons were not made (Wang *et al.*, 1999a; Lan *et al.*, 2002).

A recent study in our laboratory focused on the genetic variation present among the various WSSV isolates by alignment of three completely sequenced WSSV isolates originating from Taiwan (WSSV-TW), China (WSSV-CN) and Thailand (WSSV-TH) (Marks *et al.*, 2004). Roughly, the polymorphic loci identified can be grouped into (i) two major variable loci, (ii) regions with a variable number of tandem repeats (VNTRs), (iii) single nucleotide insertions or deletions (indels) and (iv) single nucleotide polymorphisms (SNPs). One of the major variable loci ("variable region ORF23/24") is a genomic region where large deletions frequently occur. The other major variable locus ("variable region ORF14/15") may have evolved by recombination (Marks *et al.*, 2004). Genetic characterization of eight WSSV isolates from Vietnam confirmed that polymorphisms mainly occurred at the loci identified previously (Marks *et al.*, 2004; Dieu *et al.*, 2004).

The variation found within the polymorphic loci suggests a recent spread of WSSV from a common ancestor (Marks *et al.*, 2004; Dieu *et al.*, 2004). However, such a common ancestor was unavailable to support the hypotheses on these evolutionary relationships between the WSSV isolates. In this paper, we report the analysis of an archival WSSV isolate collected in Thailand in September 1996, referred to as TH-96-II. The genome of this isolate (~312 kb) is the largest WSSV genome reported to date and this isolate may be ancestral to the WSSV isolates described thus far. We also investigated the relative virulence of the TH-96-II isolate compared to the WSSV isolate containing the largest deletions identified at present (WSSV-TH with a genome size of ~293 kb; van Hulten *et al.*, 2001a) in *Penaeus monodon* shrimp. Virulence was determined by recording (i) shrimp mortality rates upon WSSV infection and (ii) competitive fitness of both isolates in serial passage using quantitative real-time PCR.

Results

Genotyping of TH-96-II
Restriction Fragment Length Polymorphism (RFLP) analysis using restriction enzymes *Bsu*36I, *Bam*HI and *Kpn*I was performed to compare TH-96-II and WSSV-TH (Fig. 4.1a). Analysis of this RFLP revealed two major polymorphic loci located between the *Kpn*I restriction sites (bold and underlined) at WSSV-TH coordinates 21,009 and 34,595 (Fig. 4.1b). This region includes the two major WSSV polymorphic sites identified previously, "variable region ORF14/15" and "variable region ORF23/24" (Marks *et al.*, 2004), and both regions will be dealt with separately below. To further characterize the TH-96-II isolate, the WSSV genomic loci which were previously shown to be variable in their genetic make-up among different WSSV isolates (Marks *et al.*, 2004) were studied in detail by PCR amplification, cloning and sequence analysis.

"Variable region ORF14/15": Between the two *Kpn*I-sites located at positions 21,009 and 30,348 (WSSV-TH coordinates; Fig. 4.1b), the additional sequences in TH-96-II could be mapped to the locus "variable region ORF14/15" by the presence of two extra restriction sites (*Kpn*I and *Bsu*36I) specific for TH-96-II. This region was previously identified as a putative WSSV recombination locus (Marks *et al.*, 2004). The sizes of the fragments obtained with the three enzymes *Bsu*36I, *Bam*HI and *Kpn*I used for the restriction analysis suggested an insert of about 5 kb in TH-96-II compared to WSSV-TH (Fig. 4.1b). To exactly map this extra sequence, we performed a PCR reaction with primer pair "14/15-complete" (Table 4.1; Fig. 4.2a). Cloning and sequencing of the 7.8 kb amplified fragment revealed that TH-96-II had an insertion of 5,316 bp compared to WSSV-TH (GenBank acc. no. AY753327). Alignment of this TH-96-II fragment with the completely sequenced isolates WSSV-CN and WSSV-TW showed that these have deletions of 5,131 bp and 5,138 bp at this locus, resp. (Fig. 4.2a).
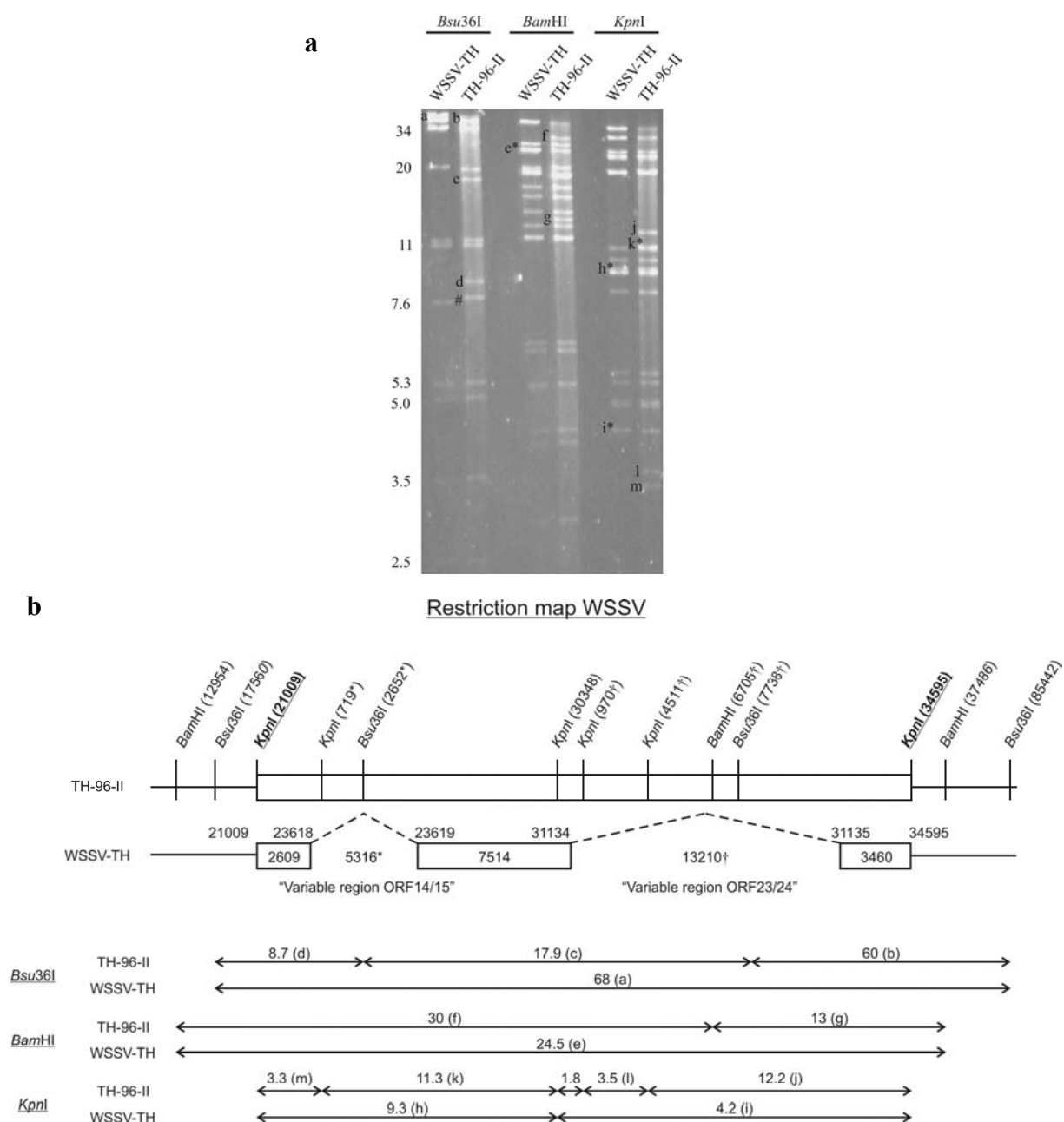
Figure 4.1. RFLP analysis of WSSV-TH and TH-96-II genomic DNA using *Bsu*36I, *Bam*HI and *Kpn*I. Some fragment sizes of WSSV-TH digested with *Bsu*36I are indicated on the left (in kb). Polymorphic DNA fragments are numbered (a-m) and marked with an asterisk in case the polymorphic fragment is one of a double band. The double bands h* and i* are not clearly visible on this gel, but were visible on other gels (data not shown). The numbers (a-m) correspond to fragments shown in Fig. 4.1b. The *Bsu*36I DNA fragments containing the ORF75 polymorphic locus are marked with a number sign **(a)**. Schematic representation of the genomic region of TH-96-II and WSSV-TH containing the two major polymorphic sites. The location of the restriction sites of *Bsu*36I, *Bam*HI and *Kpn*I are indicated within isolate TH-96-II. The restriction map numbers indicated above TH-96-II are WSSV-TH coordinates, while the TH-96-II unique sequences of 5,316 bp and 13,210 bp have their individual numbering (starting from 1 till 5,316 and from 1 till 13,210, respectively), indicated with * and †, respectively. The map numbers above WSSV-TH are in accordance with the annotation in the NCBI databank for the genomic sequence (van Hulten *et al.*, 2001a). The length of the fragments is indicated within boxes or sequences. Below, the fragments obtained with individual restriction enzymes are shown with double sided arrows for both TH-96-II and WSSV-TH. The sizes of the fragments (in kb) are shown above the arrows and the subsequent numbering (a-m) correspond to fragment numbers shown in Fig. 4.1a (the TH-96-II 1.8 kb *Kpn*I fragment is not visible in Fig. 4.1a) **(b)**.

Table 4.1. Primers used during PCR analysis for the variable loci of WSSV and for the competition assay.

| Primer pair name	Primer orientation	Sequence (5'-3')	Annealing temperature (°C) / elongation time (s)	WSSV sequence coordinates	Size (bp) of PCR product
WSSV variable loci					
14/15-complete	Forward	AATATGGAACGACGGGTG	55 / 420	22229-22246[a]	2485[a]
	Reverse	GACCAGCGCCTCTTCAG		24713-24697[a]	
23/24-flank-A	Forward	GAGTAGTCTTCAATGGCAATGT	50 / 60	2332-2353[b]	918[b]
	Reverse	GGAAATACCTAGAAGAAAAACT		3249-3228[b]	
23/24-flank-B	Forward	GGAGATCCTTCGACGAAT	50 / 60	15166-15183[b]	696[b]
	Reverse	GATGACTCGGTACGCTTTAG		15861-15842[b]	
23/24-seq	Forward	AGAAATGTACTTAGATAAAGACCCAG	60 / 120	12354-12379[b]	1490[b]
	Reverse	GAAAAGGCCCTTCTTAGTGACAAAAAC		13843-13817[b]	
23/24-TW-spec	Forward	GTTACTTGACGAGTCCTCTAC	52 / 120	10527-10547[b]	1411[b]
	Reverse	CAAACTAGTAACAAAGAAGTC		11937-11917[b]	
ORF75-flank	Forward	GAAGCAGTATCTCTAACAC	49 / 80	107875-107893[a]	868[a]
	Reverse	CAACAGGTGCGTAAAAGAAG		108742-108723[a]	
ORF94-flank	Forward	GTGCCGCAGGTCTACTC	51 / 80	142656-142672[a]	682[a]
	Reverse	CATACGACTCTGCTTCTTG		143337-143319[a]	
ORF125-flank	Forward	CGAAATCTTGATATGTTGTGC	52 / 100	187791-187811[a]	652[a]
	Reverse	CCATATCCATTGCCCTTCTC		188442-188423[a]	
Transposase	Forward	GTGGATAATATTCGTCTTCAAC	55 / 120	253988-254009[a]	1489[b]
	Reverse	CTCAAAGACAACGACATTAG		254138-254119[a]	(151[a])
*hr*9	Forward	GGTGGGAGAGGAAGTATTATG	58 / 180	283189-283209[a]	2181[a]
	Reverse	GTCCATATGATGATTCTCTACC		285369-285348[a]	
Competition assay					
VP26-QT	Forward	GGAATTTGGCAACCTAACAAACCTG	50 / 60	228833-228809[a]	306[a,b]
	Reverse	GGGCTGTGACGGTAGAGATGAC		228532-228553[a]	
TH-96-II-QT	Forward	CAGGAGATATCTCTATGGAAAC	50 / 60	8083-8104[b]	688[b]
	Reverse	CAACGAAATGACCATTACAGAC		8770-8749[b]	
WSSV-TH-QT	Forward	GAGTAGTCTTCAATGGCAATGT	50 / 60	30907-30928[a]	320[a]
	Reverse	GATGACTCGGTACGCTTTAG		31226-31207[a]	

[a] WSSV-TH sequence coordinates or WSSV-TH sizes
[b] WSSV-TW sequence coordinates or WSSV-TW sizes

Table 4.2. ORFs encoded by "variable region ORF14/15" of isolate TH-96-II.

| ORF	Position[a]		Size[b]		pI[c]	Characteristics
	Start	Stop	aa	Mr		
ORF-I	1034	42	330	36	4.4	Gene family 8; 47% similarity with TH-96-II ORF-V
ORF-II	1597	1043	184	21	10.2	Transmembrane domain (TM); repeat
Repeat	1311	1096				Direct tandem repeat (9*24 bp)
ORF-III	2515	1640	291	32	5.8	53% similarity with WSSV CN464
ORF-IV	3528	2614	304	34	7.5	64% similarity with WSSV ORF14
ORF-V	4925	3588	445	49	5.2	TM; Gene family 8; 47% similarity with TH-96-II ORF-I

[a] Position numbering in accordance to Fig. 4.1b & 4.2a
[b] Size of ORFs in amino acids (aa) and predicted molecular mass in kDa (Mr)
[c] Predicted isoelectric point (pI)

The identified TH-96-II fragment encompassed a unique sequence encoding five putative ORFs, designated ORF I to ORF V (Fig. 4.2a; Table 4.2). These five ORFs are flanked by ORF14 (which is not present in WSSV-TW and WSSV-CN) and WSSV-CN ORF CN464 (which is not present in WSSV-TH). Two ORFs (ORFs I and V) belong to gene family 8 described for the WSSV-TH genome (van Hulten *et al.*, 2001a). The proteins encoded by these ORFs have a mutual amino acid similarity of 47% and an average amino acid similarity of about 40% with the other members of gene family 8 in the WSSV-TH genome. ORF III encodes a protein with 53% similarity to CN464, which is flanking the deletion in WSSV-CN,
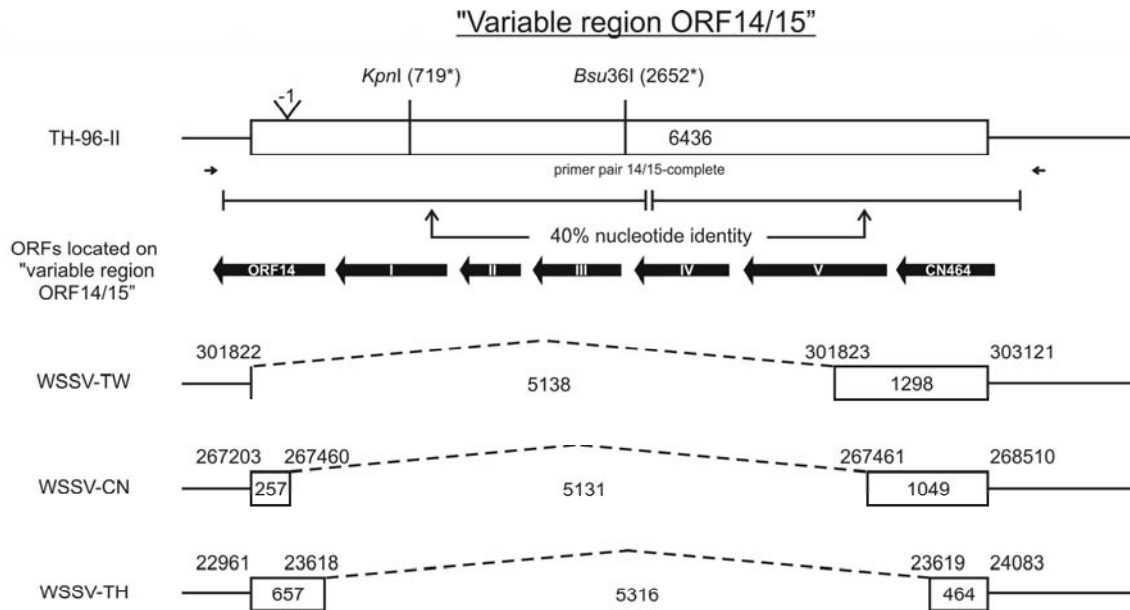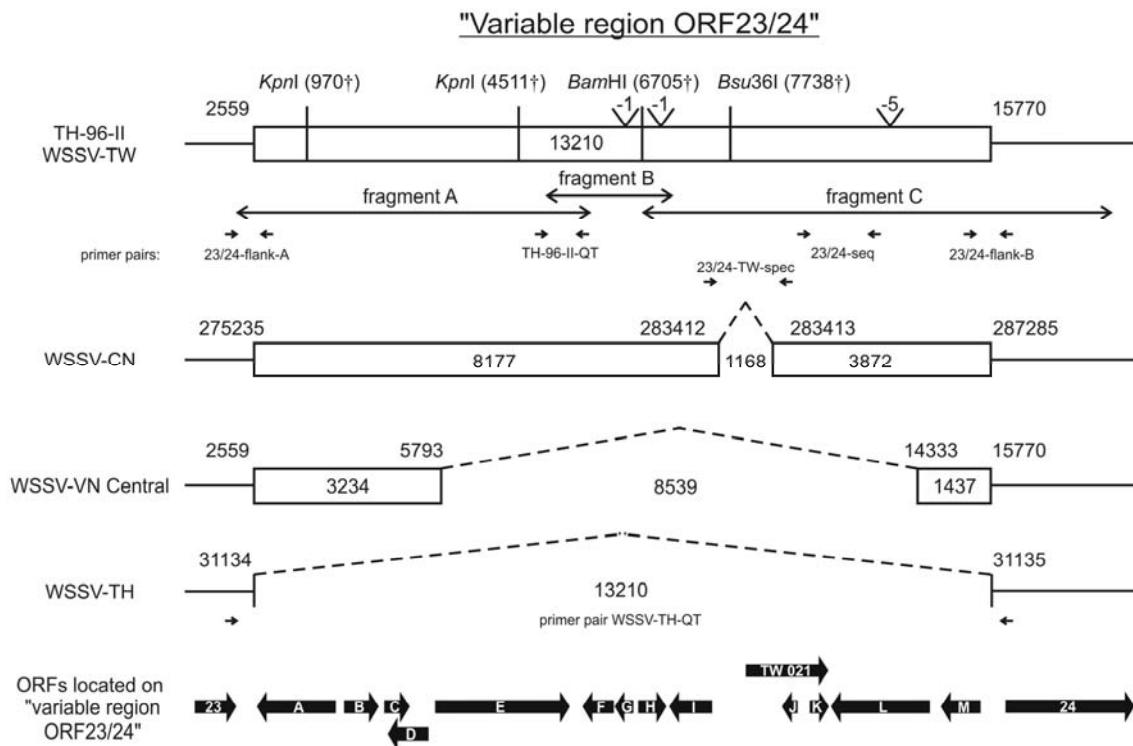
Figure 4.2. Schematic representation of the "variable region ORF14/15" of TH-96-II, WSSV-TW, WSSV-CN and WSSV-TH. The map numbers, indicated above each isolate, are in accordance with the numbers in the NCBI databank for the genomic sequence of each isolate. The length of the fragments is indicated within boxes or sequences. The positions of the ORFs are indicated by closed arrows, which also represent the direction of transcription. Open arrows represent primers. Restriction sites are similar to Fig. 4.1b **(a)**. Schematic representation of "variable region ORF23/24" of the isolates mentioned and WSSV-VN central. The coordinates of the WSSV-VN central isolates are according to the WSSV-TW annotation. ORFs of "variable region ORF23/24" are numbered in accordance with the numbering used by Marks *et al.* (2004). Characterized fragments are shown by double sided arrows. Further annotation is similar to Fig. 4.2a **(b)**.

and 39% similarity to WSSV ORF118 (gene family 10) located elsewhere on the genome. ORF IV encodes a protein with 64% similarity to ORF14, which flanks the deletion site in WSSV-TH. Possibly the insert initially evolved by a duplication event, as the left part of the sequence (the fragment encoding ORF14, ORFI and ORFII and ORFIII; coordinates 579-4,124 of GenBank acc. no. AY753327) showed 40% nucleotide identity to the right part of the insert (containing ORFIV, ORFV and CN464; coordinates 4,125-7,475 of acc. no. AY753327; Fig. 4.2a).

"Variable region ORF23/24": The second large insert of TH-96-II was identified between the two *Kpn*I-sites located at positions 30,348 and 34,595 (WSSV-TH coordinates; Fig. 4.1b). This genomic segment had already been shown to contain deletions of various sizes in multiple WSSV isolates, and was hence designated "variable region ORF23/24" (Fig. 4.2b) (Lan *et al.*, 2002; Marks *et al.*, 2004; Dieu *et al.*, 2004). Analysis of the RFLP shown in Fig. 4.1a (Fig. 4.1b) showed that TH-96-II has an additional insertion of about 13 kb compared to WSSV-TH. This region was further analyzed by restriction enzyme analysis with *Eco*RI and *Xho*I of the 13 kb *Bam*HI restriction fragment designated g in Fig. 4.1b (fragment C in Fig. 4.2b) and the PCR fragments A and B (Fig. 4.2b). The restriction patterns obtained were the same as the patterns deduced from the complete genome sequence of WSSV-TW. PCR fragments obtained on TH-96-II DNA with primer pairs "23/24-flank-A", "TH-96-II-QT", "23/24-seq" and "23/24-flank-B" (Table 4.1; Fig. 4.2b) were cloned and sequenced. The obtained sequences showed 99.8-100% nucleotide identity to WSSV-TW and WSSV-CN. A clear 1.4 kb PCR fragment was obtained with primer pair "23/24-TW-spec" using TH-96-II DNA as template. WSSV-TH DNA failed to give fragments in all these PCR reactions. These results confirm that isolate TH-96-II indeed contains genetic information very similar to WSSV-TW in "variable region ORF23/24".

Further polymorphisms: The number of repeat units (RUs) present in the non-*hr* unidirectional repeats in ORF75, ORF94 and ORF125 have been shown to be variable for the WSSV isolates identified thus far (Table 4.3; Wongteerasupaya *et al.*, 2003; Marks *et al.*, 2004). To determine the number of RUs in isolate TH-96-II, we performed PCR reactions with specific primer sets (Table 4.1) flanking the repeats of ORF75, ORF94 and ORF125. WSSV-TH was included as control. The PCR fragments obtained were cloned, sequenced and aligned, and the numbers of RUs were determined (Table 4.3). The numbers of RUs identified correlated to the respective sizes of the PCR fragments for each of the three VNTR loci for both WSSV-TH and TH-96-II. The sequenced regions flanking the tandem repeats (between the primers used and the actual repeats) on both the 5' and 3' end showed 100% nucleotide identity with the corresponding sequences of WSSV-TW, WSSV-CN and WSSV-TH. WSSV-TH was found to contain 15 repeat units in ORF75 (Table 4.3: WSSV-TH cloned), instead of 12 as previously published (van Hulten *et al.*, 2001a) and deposited in the database

Table 4.3. Number of repeat units present within the non-*hr* unidirectional repeats of ORF75, ORF94 and ORF125.

WSSV isolate	ORF75 (45 bp and 102 bp[a]) / 107965-108675[b]	ORF94 (54 bp[a]) / 142744-143067[b]	ORF125 (69 bp[a]) / 187899-188312[b]
WSSV-TW	21 (16 and 5[c])	6 (TTTGTT[d])	8
WSSV-CN	15 (11 and 4[c])	12 (TTGGGGGGTTTT[d])	8
WSSV-TH (database)	12 (9 and 3[c])	6 (TGGGTT[d])	6[e]
WSSV-TH (cloned)	15 (11 and 4[c])	6 (TGGGTT[d])	6[e]
TH-96-II	16 (12 and 4[c])	6 (GTGTTT[d])	6[e]

[a] Length of repeat units (RUs)
[b] WSSV-TH coordinates of total repeat
[c] Number of 45 bp and 102 bp RUs, respectively
[d] Genotype of each of the successive RUs of the SNP present at position 48 (guanine (G) / thymine (T)) is shown
[e] Identical

(acc. no. AF369029; Table 4.3: WSSV-TH database). The PCR product used for cloning was directly obtained from WSSV-TH DNA that was sequenced by van Hulten *et al.* (2001a). Due to the repetitive nature of this locus, this discrepancy has probably been introduced by an artifact in the automated assembly procedure of the complete WSSV-TH genome. This result adds 192 nucleotides to the size of WSSV-TH. For both ORF94 and ORF125, the sequences obtained for WSSV-TH were identical to the complete genome sequence of WSSV-TH present in the database (acc. no. AF369029; van Hulten *et al.*, 2001a).

WSSV-TH and TH-96-II can be distinguished by the number of RUs in ORF75 (15 vs 16 repeats) and by SNPs present in the RUs of ORF94, whereas the ORF125 region is identical (Table 4.3). The difference in the VNTR of ORF75 is also visible in Fig. 4.1a (in the lanes where DNA was digested with *Bsu*36I), where the 7.6 kb fragment of WSSV-TH which contains this polymorphic locus (indicated with #) is slightly larger for TH-96-II.

The genome of WSSV-TW encodes a putative transposase, which is not present in WSSV-CN and WSSV-TH (Marks *et al.*, 2004). Using primer pair "Transposase" (Table 4.1) flanking the transposase gene in the WSSV-TW genome, we obtained a PCR fragment of about 150 bp for TH-96-II. A fragment of similar size was obtained with WSSV-TH DNA, used as control. Therefore, we conclude that TH-96-II does not contain this particular transposase sequence. For *hr*9 it was shown that WSSV-CN (8 RUs) has one extra repeat unit compared to WSSV-TH and WSSV-TW (7 RUs; Marks *et al.*, 2004). Using primer pair "*hr*9" (Table 4.1) flanking *hr*9, a PCR fragment of 2.2 kb was obtained for WSSV-TH, used as control, but also for TH-96-II. Restriction enzyme analysis on both PCR fragments using *Sph*I and *Xba*I confirmed their similarity. Therefore, we conclude that TH-96-II *hr*9 contains 7 RUs, similar to WSSV-TH. These data show that the transposase and *hr*9 loci cannot be used to distinguish between TH-96-II and WSSV-TH.

*Virulence and competitive fitness*

So far, no differences in host range or tissue tropisms have been described for the various WSSV isolates characterized (Wang *et al.*, 1998; Wang *et al.*, 1999a; Chen *et al.*, 2000; Lan

et al., 2002; Hameed *et al.*, 2003). Both WSSV-TH and TH-96-II infected shrimp *Penaeus monodon* and crayfish *Orconectes limosus* and *Astacus leptodactylus* without any clear respective symptom differences. To evaluate potential differences in virulence between both isolates, we (i) recorded *in vivo* mortality rates of WSSV-infected *P. monodon* and (ii) determined the competitive fitness of both WSSV isolates in serial passaging in *P. monodon*.

In vivo *mortality rates*: Stocks of both WSSV isolates were purified using the same standard isolation procedure. Transmission electron microscopy, Bradford assays and SDS-PAGE showed that both stock solutions finally contained WSSV in the same quantity and quality. The *in vivo* mortality rates were compared for both virus stocks using a series of dilutions identical for both isolates (Fig. 4.3). For each of the dilutions injected, shrimp mortality commences at an earlier time p.i. for WSSV-TH than for TH-96-II. When a sublethal dose ($<LC_{100}$) was injected (10,000x and 100,000x dilutions), similar virus dilutions resulted in a lower cumulative mortality for TH-96-II relative to WSSV-TH. For most of the dilutions, the final cumulative mortality is reached at earlier time points for WSSV-TH than for TH-96-II. This can be further specified by estimating the LT_{50} (median lethal time) for the 1,000x times dilutions, the highest dilution which for both isolates result in 100% cumulative mortality (LC_{100}). The LT_{50} for WSSV-TH is around 3.5 days, while the LT_{50} for TH96-II is around 14 days (Fig. 4.3, double sided arrows).

Competition of TH-96-II versus WSSV-TH: To make a direct comparison in virulence between the two isolates, we performed a competition assay during serial passage in shrimp. Three experimental groups of *P. monodon* were set up: group A was infected with WSSV-TH, group B with isolate TH-96-II, and group C with a mixture of the two isolates with a DNA ratio of about 1:1 (approximately an equal amount of virus particles of both isolates in the mix; Table 4.4, passage 0). The moribund shrimps of each group were pooled and new virus stocks were isolated by collection of the hemolymph and used for a new infection (passage n+1) in shrimp. This cycle was continued for each of the three groups for five passages.

After the fifth passage, the virus preparations were analyzed for the presence of WSSV-TH and TH-96-II DNA for each passage by quantitative real-time PCR using primer pairs "WSSV-TH-96-QT" and "TH-96-II-QT", located in "variable region ORF23/24" (Table 4.1; Fig. 4.2b). Each primer pair was shown to be specific for the respective isolate and clear fragments were obtained in a regular PCR (cloning and sequencing of these bands revealed that they contained the anticipated sequences). For group A only WSSV-TH was detected and for group B only isolate TH-96-II. For group C, in the first two passages, DNA of both isolates was detected (Table 4.4). However, in the last three passages only WSSV-TH DNA was detected by this single step real-time PCR (Table 4.4). Fig. 4.4 shows the relative virus amounts of WSSV-TH and isolate TH-96-II during the five passages of group C.
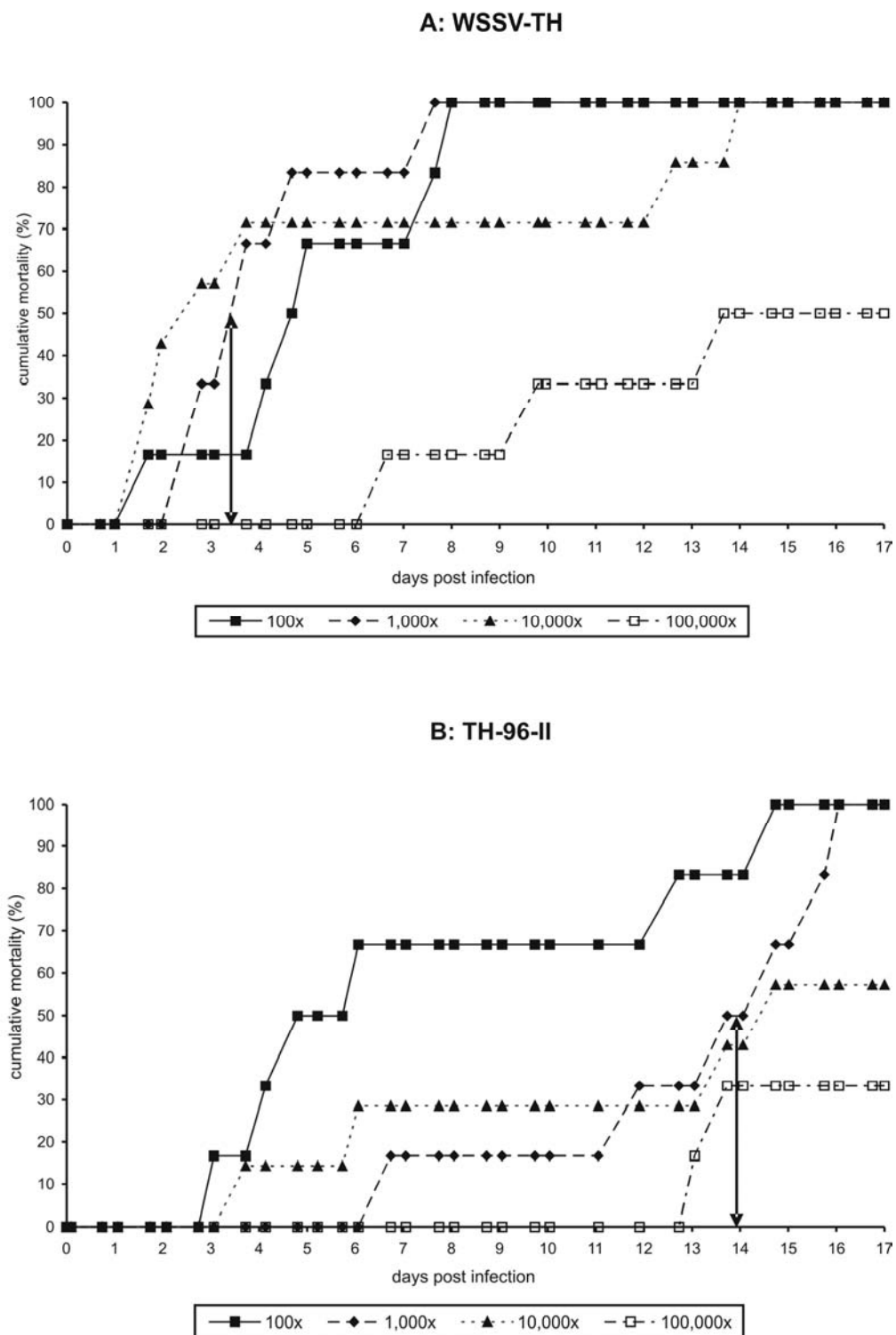
A: WSSV-TH

B: TH-96-II

Figure 4.3. Cumulative mortality rates of *P. monodon* by injection of WSSV using dilutions of WSSV-TH **(A)** and TH-96-II **(B)**. Cumulative mortality rates of shrimp from groups infected with different virus concentrations are plotted against the days after infection. The double sided arrows indicate the LT_{50} values for the 1,000x dilutions.

Table 4.4. (Relative) WSSV-TH and TH-96-II viral DNA concentration of group C as measured by real-time PCR.

Passage	Concentration WSSV-TH[a]	Concentration TH-96-II[a]	Total Concentration WSSV[a]	% WSSV-TH of total amount WSSV	% TH-96-II of total amount WSSV
0	2500	3500	6000	42	58
1	159400	50000	209400	76	24
2	70600	1100	71700	98	2
3	239000	-[b]	239000	100	0
4	512000	-[b]	512000	100	0
5	187200	-[b]	187200	100	0

[a] Concentration in DNA copy number μl^{-1}
[b] Not detectable by this single step real-time PCR
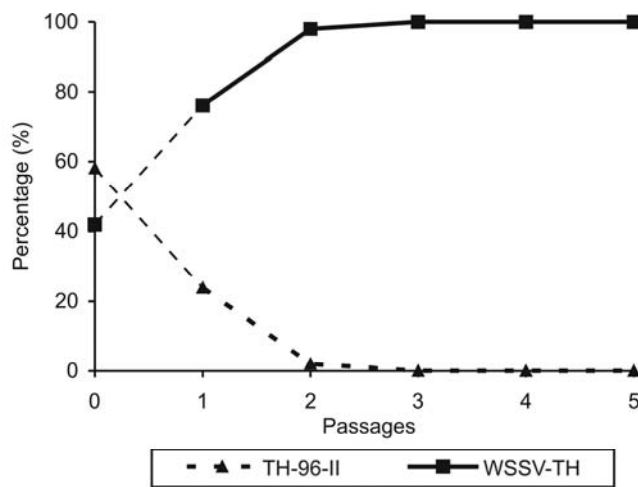
Figure 4.4. Relative amount (%) of WSSV-TH DNA and TH-96-II DNA present during five consecutive passages of *P. monodon* injected with a 1:1 mix of both isolates (group C).

To investigate the stability of the genomes in groups A & B during passaging, we performed PCR on the VNTRs of ORF75, ORF94 and ORF125 at passage 5 with primer pair "ORF75-flank", "ORF94-flank" and "ORF125-flank", respectively (Table 4.1). These loci belong to the most variable ones of the WSSV genome (Marks *et al.*, 2004; Wongteerasupaya *et al.*, 2003; Dieu *et al.*, 2004) and genomic changes could have occurred during passaging. Cloning and sequencing of the PCR products showed that the sequences were identical to the respective sequences presented in Table 4.3 and therefore it could be concluded that no changes in the major genotype occurred at these loci during the five passages of both isolates. The fragments obtained after performing PCR reactions for all five passages of group A & B with primer pairs spanning (for WSSV-TH; group A) or distributed within (for TH-96-II; group B) "variable region ORF23/24" and "variable region ORF14/15" revealed that also for these loci no major genotypic changes occurred during the five passages for both isolates (data not shown).

During mixed infections, homologous recombination between two genotypes of the same virus occurs with high frequency for large dsDNA viruses like baculoviruses and herpesviruses, both in cultured cells as well as *in vivo* (Possee & Rohrmann, 1997; Meignier, 1985). To examine whether recombination occurred between the two WSSV isolates during our competition assay, the ORF75 and ORF94 loci were analyzed for group C at passage 5 in the mixed infection (the ORF125 locus cannot be used to distinguish between both isolates). A 1.05 kb PCR product was obtained for ORF75, which is identical to the band obtained for WSSV-TH, and about 100 bp smaller than the PCR product expected for TH-96 II. The 1.05

kb PCR product was cloned and three clones were sequenced. All three clones contained 15 RUs, identical to WSSV-TH. For ORF94, TH-96-II and WSSV-TH can only be distinguished by genotype (SNPs), because the number of RUs is the same. The 0.7 kb PCR product obtained for group C at passage 5 was cloned and five clones were sequenced. All five clones contained 6 RUs with a genotype identical to WSSV-TH, different from TH-96-II. These data confirmed the results of the real-time PCR, viz. only WSSV-TH DNA was detectable in group C after passage 5, and show that in our setting we could not detect homologous recombination between the two WSSV isolates.

Discussion

Analysis of three completely sequenced WSSV isolates identified different genomic polymorphic loci (Marks *et al.*, 2004). The origin of one of the major polymorphic loci, "variable region ORF14/15", remained obscure. In the present study we characterized an archival WSSV isolate, TH-96-II. "Variable region ORF14/15" of this isolate harbored all of the sequences present in other isolates, thereby identifying "variable region ORF14/15" as a second WSSV locus where deletions frequently occur (Fig. 4.2a). The presence of extra ORFs belonging to WSSV gene family 8 in this region strongly suggests that the additional 5,316 bp sequence, specific for TH-96-II, is an authentic part of the WSSV genome and that the TW, CN and TH isolates lack (part of this) fragment due to a deletion event. The function of the five ORFs present on this large genomic sequence is not known, as they have no significant similarity with known genes present in public databases.

The second major polymorphic site, "variable region ORF23/24", had already been shown to be a region prone to (large) deletions (Lan *et al.*, 2002; Marks *et al.*, 2004; Dieu *et al.*, 2004). In this region, TH-96-II contains the complete 13 kb similar to WSSV-TW. Based on the two "variable regions", isolate TH-96-II is possibly ancestral to the WSSV isolates characterized thus far. Other WSSV genotypes could have evolved from this ancestor by deletions of various lengths in one or both "variable regions". To explain this WSSV isolate in Thailand in 1996 and to be compatible with the hypothesis that WSSV mutated away from its ancestral site on either site of the Taiwan Strait by gradually loosing sequences in both "variable regions" (Dieu *et al.*, 2004), we assume a direct introduction of this isolate TH96-II from China/Taiwan into Thailand. The mechanisms and circumstances under which the deletions in the WSSV genome occur are still unclear (Marks *et al.*, 2004; Dieu *et al.*, 2004). However, the genes encoded by the extra sequences of both "variable regions" in TH-96-II are apparently dispensable for infection and replication of WSSV in *P. monodon*, *O. limosus* and *A. leptodactylus*, as these crustacean species are permissive hosts for WSSV-TH (van Hulten *et al.*, 2001a; Dieu *et al.*, 2004).

To determine the effect of differences in genotype on virulence and competitive fitness of WSSV, we made a comparison between TH-96-II and WSSV-TH in *P. monodon*. Since no crustacean cell lines are available for culturing WSSV and primary cell cultures are

difficult to use, the concentration of infectious WSSV cannot easily be determined. Therefore, virulence was evaluated using *in vivo* cumulative mortality rates after injection of quantified virus stocks of both WSSV isolates (Fig. 4.3). In addition, we performed a direct comparison of the two isolates in a competitive situation, with the advantage that both isolates are compared in the same shrimp (Fig. 4.4). As we considered the possibility of homologous recombination between the two virus isolates present in group C of the competition assay, three different markers were used to detect both virus isolates. "Variable region ORF23/24" was measured with real-time PCR, while the VNTRs of ORF75 and ORF94 were detected by PCR based cloning and sequencing. Only the genotype of WSSV-TH was detected in the DNA isolated after passage 5 in group C for all three markers suggesting the presence of only isolate WSSV-TH, and not isolate TH-96-II or a recombinant WSSV genotype consisting of a mosaic of WSSV-TH and TH-96-II.

Both the cumulative mortality rates as well as the competition experiment consistently show a higher virulence for WSSV-TH compared to TH-96-II. The cumulative mortality rates further suggest that the increased virulence is mainly due to more rapid mortalities caused by WSSV-TH. This result implies that the ORFs in the extra sequences of both "variable regions" of TH-96-II do not encode proteins enhancing WSSV virulence in *P. monodon*. It has been suggested for baculoviruses that smaller genomes with an increased density of homologous regions (*hr*s), which function as origin of DNA replication, may have a replication advantage (Krell, 1996). WSSV-TH and TH-96-II contain the same number of *hr*s, but the WSSV-TH genome, containing the largest deletions for both "variable regions" identified at present, is around 19 kb smaller in size than TH-96-II. This smaller genomic size may give WSSV-TH a replication advantage. However, it is also possible that the observed differences in virulence are caused by other genomic polymorphisms between the two isolates, e.g. SNPs resulting in amino acid changes in virulence related WSSV proteins.

Wang *et al.* (1999a) observed slight differences in virulence between various geographical WSSV isolates by comparison of *in vivo* cumulative mortality rates of WSSV isolates in shrimp. However, as the shrimp were challenged orally with WSSV-infected tissue (i) the infection dose was undetermined in this study and hence could vary between the isolates used and (ii) infection of the shrimp most likely was less synchronous. A correlation between genetic composition of the WSSV isolates and virulence could not be made, as the genetic character of the isolates was not known. In a second study, Lan *et al.* (2002) compared virulence of two WSSV isolates in crayfish *Procambarus clarkii* using quantified virus stocks. Although no complete genetic comparison between the two isolates was made, the genetic structure of both WSSV isolates was mapped for "variable region ORF23/24". In their study the WSSV isolate with the largest genome (WSSV-CN, containing a deletion of 1,168 bp in "variable region ORF23/24" compared to WSSV-TW; Fig. 4.2b) caused mortalities more rapidly than a WSSV isolate containing a larger deletion (a partly characterized isolate containing a deletion of 5,826 bp compared to WSSV-TW in "variable region ORF23/24").

This is in contrast to our data correlating a deletion in this region with enhanced virulence. Although a differential virulence of different WSSV isolates in shrimp (used for our experiments) and crayfish (used by Lan *et al.* (2002)) cannot be excluded, their conclusion that one or more genes present in the "variable region ORF23/24" locus are involved in virulence seems premature. Comparison of the complete genome sequences available showed that there are many SNPs present (Marks *et al.*, 2004). Some of these SNPs result in amino acid changes in proteins with unknown functions. Putatively, some of these proteins may play a role in virulence and could cause the differences in virulence observed between the WSSV isolates.

By identifying and characterizing a putative common ancestor, TH-96-II, this study sheds further light on the genesis and evolution of WSSV. As TH-96-II contains the most intact WSSV genome characterized thus far, it can be used as a new tool and reference to study the origin of the deletions in both "variable regions". A deeper understanding of WSSV gene functions will be needed to explain the differences in virulence between WSSV isolates. Further experiments will be necessary to identify factors influencing WSSV (differential) virulence in shrimp or other host species. Together, these studies could provide new insights in the epidemiology and ecology of WSSV.

Materials and Methods

WSSV virus stock

The virus isolate WSSV-TH used in this study originated from an infected *P. monodon* shrimp imported from Thailand in May 1996 and was maintained as described before (van Hulten *et al.*, 2001b). The virus isolate TH-96-II originated from an infected *P. monodon* imported from Thailand in September 1996. Tissue of this TH-96-II infected *P. monodon* was homogenized in 330mM NaCl. After centrifugation at 1,700 x g for 10 min the supernatant was filtered (0.45 μm filter; produced by Schleicher & Schuell MicroScience) to obtain the virus in the filtrate. Crayfish *Orconectes limosus* or *Astacus leptodactylus* were injected intramuscularly with a lethal dose of WSSV (WSSV-TH or TH-96-II) using a 26-gauge needle (Microfine B&D) to propagate the virus. After approximately one week, virus was isolated from freshly extracted hemolymph as described by van Hulten *et al.* (2001b). Virus samples were examined in the transmission electron microscope for integrity and purity, and stored in aliquots at -80°C until further use.

Viral DNA

Viral DNA was isolated from purified virions as described by van Hulten *et al.* (2001a). WSSV DNA was digested with *Bam*HI, *Kpn*I (Invitrogen) and *Bsu*36I (NE Biolabs) and fragments were separated by electrophoresis in a 0.6% agarose gel at 40 V (1.3 V cm^{-1}) for 20

h. After separation, the gels were stained with ethidiumbromide (0.5 μg ml^{-1} in Tris-Acetate-EDTA (TAE)).

PCR analysis and cloning for WSSV variable loci

PCR on the genomic variable loci of WSSV was performed with 1 μl purified viral DNA, using *Taq* DNA polymerase (Promega) or the Expand Long Template PCR System (Roche) in case the expected size was >1.4 kb. The specific primer sets, PCR conditions and sizes of the PCR products are summarized in Table 4.1. PCR products were separated in 0.8% agarose gels (Sambrook *et al.*, 1989) and purified using a DNA extraction kit (MBI Fermentas). These products were subsequently cloned into DH5α competent cells using the pGEM-T easy vector system I (Promega). Plasmids containing the correct insert, as screened by restriction enzyme analysis and/or by colony PCR, were prepared for sequencing by purification with the High Pure Plasmid Isolation Kit (Roche).

Sequencing and computer analysis

Plasmid clones were sequenced using universal T7 and/or Sp6 primers, and by primer walking in case inserts were >1.5 kb (BaseClear, the Netherlands). Sequence data, genomic DNA composition and restriction enzyme patterns were analyzed using the software package DNASTAR 4.2 (DNASTAR Inc.). The output of the sequence analysis was edited in GeneDoc, version 2.6.000 (Nicholas *et al.*, 1997). Gene designation and gene annotation was according to van Hulten *et al.* (2001a). Complete WSSV sequences were retrieved from the NCBI databank (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) using the accession numbers for WSSV-TW (AF440570), WSSV-CN (AF332093) and WSSV-TH (AF369029). The "Variable region ORF14/15" of isolate TH-96-II has been deposited in GenBank under acc. no. AY753327. Other sequences of (variable) loci of TH-96-II have been deposited in GenBank under acc. no. AY864664-AY864670. The WSSV-TH VNTR ORF75 locus been deposited in GenBank under acc. no. AY864671.

Shrimp culture

Healthy *P. monodon* were imported as post-larvae from Malaysia and maintained in a recirculation system at the aquaculture facility "De Haar" at Wageningen University. Prior to each experiment, shrimp were transferred to an experimental system located at the Laboratory of Virology at Wageningen University and stocked in 180-liter aquariums. Each aquarium was fitted with an individual filter system (Eheim, Germany) containing pre-conditioned filter material, heating at 28 ± 1 °C (Schego, Germany) and continuous aeration. All experiments were performed in artificial seawater (Instant Ocean, Aquarium Systems) at a salinity of approximately 20 parts per thousand (ppt).

In vivo *shrimp mortality rates*

Shrimp of approximately 5-10 gram were injected intramuscularly with 10 µl of different virus dilutions in 330 mM NaCl in the fourth or fifth abdominal segment of the shrimp using a 29 gauge needle (Microfine B&D). Each experimental group consisted of 6-7 shrimp. After injection, the shrimp were maintained in individual housing to prevent horizontal transmission of WSSV by predation. The mortality was recorded twice a day and dead shrimp were tested for the presence of WSSV by a standardized PCR-based WSSV detection protocol developed in our laboratory (Dieu *et al.*, 2004).

Competition assay

For the first passage of the competition assay shrimp (5-10 gram) were infected, maintained and monitored as described above. Each experimental group of each passage consisted of three shrimp, of which the hemolymph was pooled before virus precipitation. Hemolymph was withdrawn from freshly deceased shrimp and mixed with modified salt Alsever solution (19.3 mM $C_6H_5Na_3O_7$, 182.5 mM glucose, 6.2 mM EDTA, 239.8 mM NaCl; pH 7.3) as an anticoagulant. The virus was precipitated by centrifugation at 80,000 x g for 1 h at 4 °C, and the pellet was resuspended in 400 µl ddH$_2$O. Virus passaging was achieved by dilution of the virus suspensions two times with 330mM NaCl. Shrimp (5-10 gram) were injected with 10 µl of this virus solution and maintained and monitored as described. Virus suspension or dilutions of these were directly used as template for (real-time) PCR.

Real-time PCR

The reaction mixture of the real-time PCR consisted of 31.5 µl ddH$_2$O, 5 µl 10x Buffer (500mM KCl, 100mM Tris-HCl; pH 8.5), 1.5 µl 100mM MgCl$_2$, 5 µl 2mM dNTPs, 2 µl forward primer (10 µM), 2 µl reverse primer (10 µM), 1 µl Supertaq (1U/µl), 0.5 µl 100x SYBR Green, 0.5 µl 100x fluorescine and 1 µl template DNA. SYBR Green chemistry was used as a fluorescent reporter for the detection of DNA. The following PCR-program was used: 94 ^{o}C for 3', 30 cycles (94 ^{o}C for 30", 50 ^{o}C for 30", 72 ^{o}C for 1') followed by 72 ^{o}C for 7'. To be able to calculate the copy numbers of the individual samples, plasmid samples with increasing, known concentrations (as measured by A$_{260}$), which contained the same fragment, were included as standards in the real-time PCR. With the standard curves obtained, the concentrations of the samples were calculated by comparing the Threshold cycle values (C$_T$) of the samples with the standard curves. Initial virus concentrations were measured using the "VP26-QT" primer pair (Table 4.1) on stocks of both isolates. Virus isolations of later passages of the competition assay were quantified by the specific "TH-96-II-QT" and "WSSV-TH-QT" primer pairs (Table 4.1). The real-time PCR was performed on an iCycler (Bio-Rad). The data were analyzed with iCycler iQ Real Time Detection System Software v3.0 for Windows (Bio-Rad).

Acknowledgements

Chapter 5

Gene expression profiling of White spot syndrome virus
in vivo

Abstract

White spot syndrome virus, type member of the virus family *Nimaviridae*, is a large dsDNA virus infecting crustaceans. The genome of the completely sequenced isolate WSSV-TH encodes 184 putative ORFs, the function of which is largely unknown. To study the transcription of these ORFs we constructed a DNA microarray containing probes corresponding to nearly all putative WSSV-TH ORFs. Transcripts of 79% of these ORFs could be detected in the gills of WSSV infected shrimp *Penaeus monodon*. Clustering of the transcription profiles of the individual genes during infection showed two major classes of genes, the first class reaching maximal expression at 20 hours post infection (putative early), the other class at 2 days p.i. (putative late). Nearly all major and minor structural virion protein genes clustered in the latter group. These data provide evidence that, similar to other large dsDNA viruses, the WSSV genes at large are expressed in a coordinated and cascaded fashion. Furthermore, we compared the transcriptome of the WSSV isolates WSSV-TH and TH-96-II, which have a differential virulence, at 2 d.p.i. The TH-96-II genome encodes 10 ORFs not present in WSSV-TH, of which at least 7 were expressed in *P. monodon* as well as in crayfish *Astacus leptodactylus*, suggesting a functional but not essential role for these genes during infection. Expression levels of most other ORFs shared by both isolates were similar. Evaluation of transcription profiles using a genome-wide approach provides a better understanding of WSSV transcription regulation and a new tool to study WSSV gene function.

*Figures 5.1 & 5.2 have been printed in grayscale. For color figures see the original publication.

Introduction

White spot syndrome virus (WSSV), member of the virus family *Nimaviridae* (genus *Whispovirus*), is a large enveloped virus infecting a broad range of crustacean species (Wang *et al.*, 1998; Mayo, 2002). In cultured shrimp WSSV infection can cause a cumulative mortality of up to 100% within 3-10 days (Lightner, 1996), leading to large economic losses to the shrimp culture industry. WSSV was first discovered in the Chinese province Fujian in 1992, from where it quickly spread (Cai *et al.*, 1995; Flegel, 1997). Nowadays, the virus has spread to almost all major shrimp farming areas of the world (Rosenberry, 2002).

Sequencing of three different WSSV isolates (WSSV-TH, WSSV-CN and WSSV-TW) showed that the size of the dsDNA genome varies from 293 to 307 kb (van Hulten *et al.*, 2001a; Yang *et al.*, 2001; acc. no. AF440570). The completely sequenced isolate WSSV-TH has a genome size of 292,967 bp. The genome encodes 184 putative ORFs, for which only 6% could be assigned a putative function based on homology with sequences in public databases. Of the remaining ORFs, 5 major and approximately 40 minor structural virion protein genes have been identified (van Hulten *et al.*, 2001a; Huang *et al.*, 2002b; Tsai *et al.*, 2004). Additionally, 3 ORFs (ORF3, ORF89 and CN-ORF366) have been suggested to be involved in WSSV latency (Khadijah *et al.*, 2003) and ORF170 was shown to encode an anti-apoptosis protein (Wang *et al.*, 2004). Large regions, each consisting of a variable number of 250 bp repeat units, were identified dispersed along the viral genome (homologous repeats: *hr*s). Similar to baculoviruses, these regions may be involved in transcription enhancement and/or DNA replication (Guarino & Summers, 1986; Kool *et al.*, 1993).

WSSV transcriptional analysis performed thus far mainly focused on WSSV genes which showed homology to known genes, such as the ribonucleotide reductases (Tsai *et al.*, 2000a), the chimeric thymidine kinase-thymidylate kinase (Tsai *et al.*, 2000b), the DNA polymerase (Chen *et al.*, 2002b), a protein kinase (ORF2; Liu *et al.*, 2001), the thymidylate synthase (Li *et al.*, 2004b) and the collagen-like ORF (Li *et al.*, 2004a). Also, the major and minor structural protein genes were subject to transcriptional analysis (Marks *et al.*, 2003; Tsai *et al.*, 2004). The analysis of these genes mostly included RT-PCRs of WSSV infection time courses and mapping of the 5' and 3' ends of the mRNAs.

To study WSSV gene expression on a genome wide scale we constructed a WSSV DNA microarray containing one or more probes for most putative WSSV ORFs. Microarrays have been successfully used to study gene expression of large dsDNA viruses such as herpesviruses (Stingley *et al.*, 2000; Paulose-Murphy *et al.*, 2001; Chambers *et al.*, 1999; Ebrahimi *et al.*, 2003), the myovirus Bacteriophage T4 (Luke *et al.*, 2002) and the baculoviruses AcMNPV and BmNPV (Yamagishi *et al.*, 2003; Iwanaga *et al.*, 2004). Using a WSSV infection time course *in vivo* in the shrimp *Penaeus monodon*, we could show expression of at least 79% of the WSSV ORFs included on the microarray, indicating that most WSSV computational ORFs are transcriptionally active. For these ORFs, transcription profiles and transcription levels (semi-quantitatively) are analyzed semi-quantitatively.

Recently, we made a genomic comparison between the WSSV isolates TH-96-II, containing the largest WSSV genome size identified thus far (around 312 kb), and WSSV-TH, containing the smallest genome size (Marks *et al.*, 2005a). The difference in genome size is mainly due to a major genomic polymorphism designated "variable region ORF 23/24", for which WSSV-TH contains a contiguous deletion of ~13.2 kb compared to TH-96-II. The ~13.2 kb fragment encompasses 10 ORFs (Marks *et al.*, 2005a). Using our microarray analysis, gene expression of these ORFs of TH-96-II is evaluated in two permissive crustacean hosts for both isolates, *P. monodon* and crayfish *Astacus leptodactylus*. Marks *et al.* (2005a) also demonstrated a higher virulence of WSSV-TH as compared to TH-96-II in *P. monodon*. To correlate this biological feature with differential WSSV gene expression, the complete WSSV transcriptome at 2 d.p.i. is compared between the isolates WSSV-TH and TH-96-II in *P. monodon*. The importance of genome wide transcription studies using microarrays in understanding the regulation of WSSV gene expression is discussed.
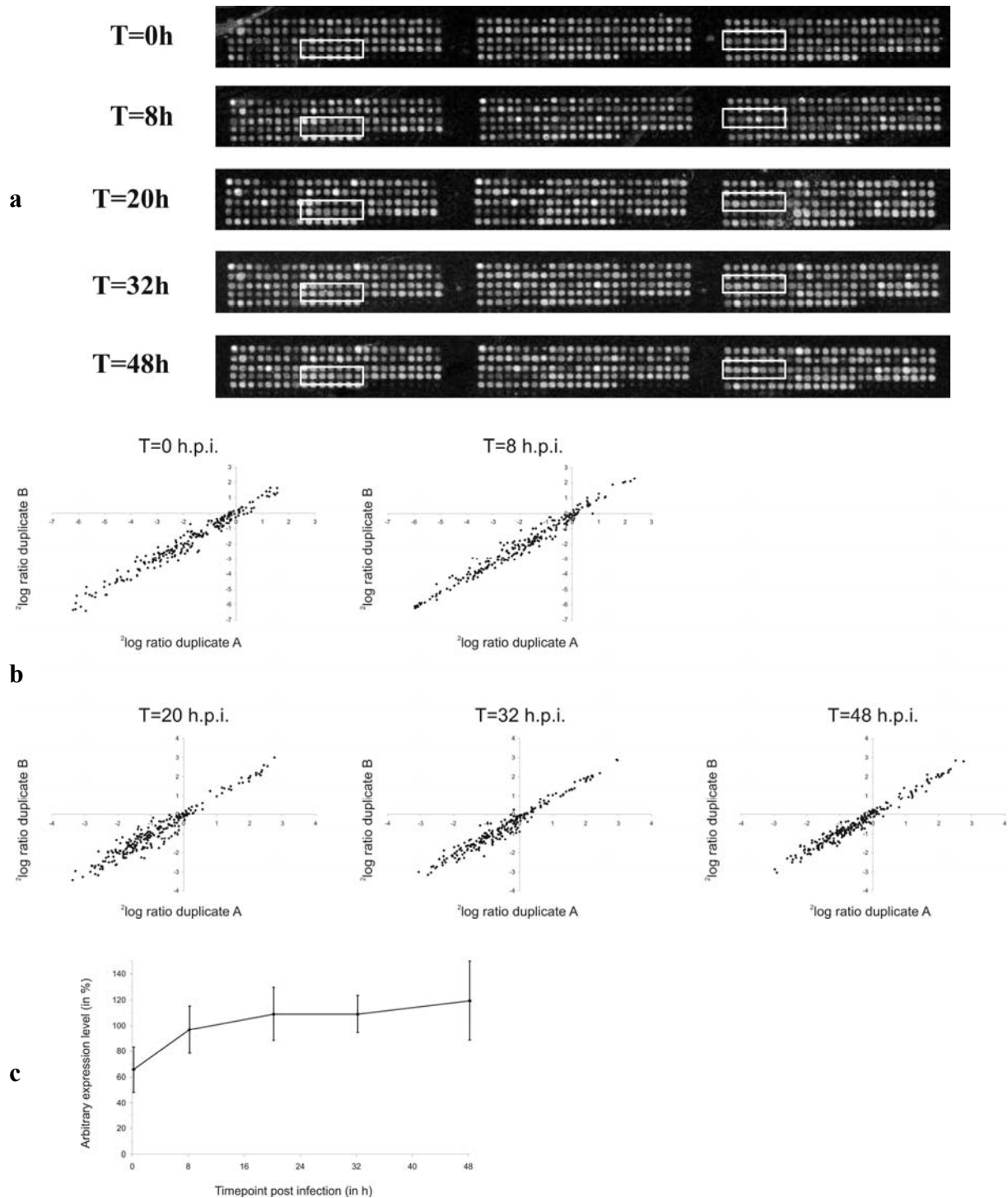
Results

To evaluate WSSV gene expression during infection in shrimp, a viral microarray was designed containing nearly all of the putative WSSV ORFs, based on the published sequences of WSSV-TH, WSSV-CN and TH-96-II (van Hulten *et al.*, 2001a; Yang *et al.*, 2001; Marks *et al.*, 2005a). The WSSV genome contains 9 *hr*s, each consisting of 250 bp repeats with a nucleotide identity between 74% and 91%. For each *hr*, a probe representing the respective *hr* was included on the microarray. In addition to the viral sequences various other probes were included on the microarray as controls. Table 5.1 shows a complete overview of the probes present on the microarray.

Poly(A)$^+$ RNA was isolated from WSSV-TH infected *P. monodon* gill tissue at 0, 8, 20, 32 and 48 h.p.i. The gills are one of the primary target tissues of WSSV infection (Lo *et al.*, 2004). The poly(A)$^+$ RNA was labeled Cy3 and mixed with a Cy5 labeled reference sample to normalize for differences in probes on the microarray slides. In the experiments performed for this time series, a single standard reference sample was used for each hybridization, allowing direct comparison of the various hybridization experiments. Each of the time course Cy3/Cy5 mixtures was hybridized to individual microarrays. Fig. 5.1a shows images of parts of the microarrays that were constructed after the scanning.

Microarray controls

The interpretation of microarray experiments is highly dependent on the quality of the data obtained, as well as on the normalization between the several microarrays used. Therefore, the most important controls are summarized.

Reproducibility: After processing of the data, the Cy3/Cy5 ratio for each gene was obtained in duplicate for the individual time points due to the on-array duplicates. Fig. 5.1b shows that the

a

b

c

Figure 5.1. Pseudocolor microarray images (Cy3 and Cy5 are colored green and red (see original publication), respectively) of one of the on-array duplicate set of probes. An equal mix of red and green result in a yellow pseudocolor, other ratios result in intermediate colors. Probes for WSSV and shrimp genes as well as background and luciferase controls are randomly distributed over the array. For probes representing WSSV genes (examples are boxed) that are red at 0 and 8 hours post WSSV infection, a yellowish color can be observed at 20, 32 and 48 h.p.i., indicating WSSV gene expression. White probes (saturation) indicate a very high Cy3/Cy5 signal **(a)**. Scatterplot of the ^{2}log ratios of the on-array duplicates (A and B) present for each probe on the microarrays. Each point in the graph represents a probe. The trend lines of the scatterings shown in the graphs were almost equal to y=x with a regression coefficient of $R^2 > 0.93$, indicating that the duplicate set of results is very similar **(b)**. Expression level in % of the 16 shrimp genes shown in Table 5.1 (mean expression over the time course is set to 100%; standard deviations are indicated for each time point) **(c)**.

ratios of the duplicates were very similar. The average of the duplicate values was used for further analysis. Additionally, the hybridizations for the five time points were repeated with independently isolated RNA preparations. The data obtained in the duplicate experiment were consistent with the data of the experiment described.

Normalization: To evaluate the normalization procedure (we used the same amount of poly(A)$^+$ RNA for each time point) the transcription profiles of 16 cellular shrimp genes present on the microarray were evaluated (Fig. 5.1c). Although limited information is available for shrimp transcription, we chose to spot probes on our WSSV microarray for shrimp genes that were likely to be expressed constitutively before and during virus infection. Except for time point 0 h.p.i. the difference in expression of the 16 selected shrimp genes is within 20% of the average expression (set at 100%) between the several time points. The expression of some of the individual cellular genes shown in Fig. 5.1c might be influenced by WSSV infection, but the transcriptional profile of the 16 shrimp genes combined confirmed the robustness of the normalization procedure used for analysis of the time course.

Expression WSSV genes
Detectable WSSV genes: Pilot experiments revealed that the probes of *Medicago truncatula* and *Aequorea victoria*, used as background controls, showed a minimal signal compared to the other probes on the WSSV microarray when labeled cDNA of WSSV infected shrimp was hybridized to the microarrays. Except for some false positive signals (see Table 5.1), the Cy3 signal of every WSSV specific probe was below the background threshold at the time points 0 and 8 h.p.i. of the WSSV time course, indicating that we could not detect any WSSV transcripts at these time points. At 20, 32 and 48 h.p.i., we detected 125 of the 158 WSSV ORFs (79%) for which probes were present on the microarray (Table 5.1). The ORFs corresponding to the WSSV probes which did not reach the background threshold at any of the time points (21%) were excluded from further analysis. Four probes representing *hr*s gave a specific signal, indicating transcriptional activity within these *hr*s (Table 5.1). These *hr*s were included in our further analysis.

Transcription profiling: The Cy3/Cy5 ratios of the 125 WSSV genes that we could specifically detect were normalized to the maximal expression of each gene (100%). The constructed transcription profiles are presented in Table 5.2. To examine the relationship between the WSSV genes, these relative expression data were analyzed by hierarchical clustering (Euclidean distance). Fig. 5.2a shows that the WSSV genes clustered into two major groups, boxed yellow and blue. The mean transcription patterns of these clusters are shown in Fig. 5.2b & 5.2c, respectively.

Table 5.1. Complete overview of the PCR amplified DNA fragments spotted on the WSSV microarray, subdivided by WSSV gene probes (according to ORF numbers), shrimp gene probes and negative control probes. Luciferase control probes are not shown.

ORF nr	putative function/ name*	start ORF		end ORF	length ORF	start probe	end probe	length probe	part of gene detected	not detected/ remark
						WSSV-TH ORFs (numbering according to WSSV-TH, acc. no. AF369029; van Hulten *et al.*, 2001a)				
1	in virion (VP28)	1	→	615	615	1	616	616		
2	Protein kinase	710	←	2902	2193	984	1775	792	3'	
						1653	2975	1323	5'	
3	latency related	3118	←	4989	1872	3582	4928	1347		
4		5185	→	8970	3786	7379	8677	1299	3'	
						5176	6495	1320	5'	
5		9056	→	10879	1824	9582	10701	1120		x
6	in virion (vp800)	10834	→	13236	2403	12122	13206	1085	3'	x
						10882	12162	1281	5'	x
7		13311	→	13982	672	13311	13910	600		
8		13979	→	14890	912	14376	14850	475		
9		14923	←	20733	5811	14955	16161	1207	3'	
						16837	18001	1165	5'	x
10		20837	→	21358	522	20837	21349	513		
11		21364	→	22161	798	21365	22160	796		
12		22201	←	22596	396	22229	22662	434		
13		22232	→	22648	417	overlap ORF12; no probe				
14		22685	←	23581	897	22705	23295	591		
15		23591	←	24157	567	23591	24157	567		
16		24265	←	27996	3732	24662	25439	778	3'	
						26882	27959	1078	5'	
17		28024	→	28296	273	28024	28296	273		x
*hr*1										
18		28366	→	28530	165	in *hr*; no probe				
19		28760	→	28960	201	in *hr*; no probe				
20		28957	←	29142	186	in *hr*; no probe				
21		29283	←	29468	186	in *hr*; no probe				
22		29934	←	30149	216	in *hr*; no probe				
23		30426	→	31052	627	30116	31222	1107		
24		31320	→	33485	2166	32365	33427	1063	3'	
						32076	33186	1111	5'	
25		33532	←	35148	1617	33508	34595	1088		
26		35172	→	35402	231	35172	35402	231		x
27	DNA polymerase	35571	→	42626	7056	41523	42614	1092	3'	
						36486	37541	1056	5'	x
28		42667	←	42882	216	42667	42882	216		
29	in virion (vp448)	42935	←	44281	1347	42958	44031	1074		
30	Collagen/ in virion (vp1684)	44350	→	49404	5055	48368	49425	1058	3'	
						47155	48219	1065	5'	x
31	in virion (VP24)	49448	←	50074	627	49443	50074	632		
32		50129	←	50467	339	50135	50467	333		
33		50494	←	51381	888	50494	51024	531		
34	in virion (vp95)	51341	←	51628	288	51429	51628	200		
35		51659	←	51952	294	51677	51952	276		
36		52007	→	55912	3906	54065	55269	1205	3'	
						52228	53259	1032	5'	
37		55999	←	56601	603	55999	56480	482		
38		56598	←	57458	861	56859	57458	600		
39		57509	←	58204	696	57509	58204	696		
40		58285	←	62892	4608	58842	59835	994	3'	x
						61727	62726	1000	5'	x
41		63021	←	65939	2919	63297	64261	965	3'	
						64766	65740	975	5'	
42		65956	→	69795	3840	68624	69621	998	3'	positive at 0 h.p.i.
						66177	67214	1038	5'	x
43		69737	→	72682	2946	71311	72456	1146	3'	
						70137	71111	975	5'	

						WSSV-TH ORFs (numbering according to WSSV-TH, acc. no. AF369029; van Hulten *et al.*, 2001a)				
ORF nr	putative function/ name*	start ORF		end ORF	length ORF	start probe	end probe	length probe	part of gene detected	not detected/ remark
44		72663	→	73253	591	72815	73249	435		
*hr*2										
45		73614	←	73859	246	in *hr*; no probe				
46		73915	←	74106	192	in *hr*; no probe				
47		74151	←	74831	681	in *hr*; no probe				
48		75246	←	75422	177	in *hr*; no probe				
49		75584	→	76210	627	75321	76348	1028		
50		76237	→	76401	165	in *hr*; no probe				
51		76463	→	76714	252	in *hr*; no probe				
52		76776	→	77000	225	76230	77003	774		x
53		77284	←	79815	2532	77972	78957	986	3'	
						78212	79406	1195	5'	x
54	Thymidylate synthase	80046	→	80915	870	79895	81012	1118		
55		81077	→	81751	675	80848	81873	1026		
56		81900	→	83168	1269	81807	83045	1239		
57		83170	→	84000	831	83170	83992	823		
58		84026	→	84919	894	84026	84919	894		
59		85001	←	86197	1197	84862	85854	993		positive at 0 h.p.i.
60		86334	←	87869	1536	86334	86990	657		
61	Protein kinase	87925	←	89667	1743	88143	89540	1398		
62		89955	←	90197	243	89955	90197	243		x
63		90298	→	90744	447	90298	90597	300		x
64		90669	→	91046	378	90794	90983	190		x
65		91003	→	94443	3441	93170	94366	1197	3'	
						91349	92386	1038	5'	x
66		94903	→	96777	1875	95736	96817	1082		
67		97012	←	97242	231	97012	97222	211		
68		97239	←	97394	156	no probe				
69		97587	→	97898	312	97587	97898	312		
70		98032	←	99252	1221	98068	99172	1105		
71	dUTPase	99376	←	100761	1386	99376	100018	643		x
72		100959	→	103865	2907	101441	102490	1050		
73		104007	→	107141	3135	105059	106233	1175	3'	
						104185	105354	1170	5'	
74		107265	→	107570	306	107265	107564	300		
75	in virion (vp357)	107467	→	108789	1323	108178	108924	747		
76		108889	←	109341	453	108889	109341	453		
77		109433	←	110887	1455	109433	110092	660		
78		110964	→	111779	816	110966	111563	598		x
79		111751	→	112419	669	112054	112417	364		
80		112426	→	112812	387	112426	112725	300		
81		112771	←	113784	1014	113185	113784	600		
82		113793	←	117419	3627	114028	115140	1113	3'	x
						116024	117158	1135	5'	x
83		117465	←	117878	414	117465	117878	414		x
84		118025	→	124969	6945	123799	124845	1047	3'	
						118532	119333	802	5'	x
85		125037	←	126416	1380	125070	126154	1085		
*hr*3										
86		126211	←	126876	666	in *hr*; no probe				
87		126782	←	127129	348	in *hr*; no probe				
88		127035	←	127634	600	in *hr*; no probe				
89	latency related	128334	→	132644	4311	131185	132414	1230	3'	
						129522	130720	1199	5'	x
90		132697	←	134976	2280	133013	134063	1051	3'	
						133799	134963	1165	5'	x
91		135031	←	138249	3219	135105	136210	1106	3'	
						137218	138203	986	5'	x

ORF nr	putative function/ name*	start ORF		end ORF	length ORF	start probe	end probe	length probe	part of gene detected	not detected/ remark
	WSSV-TH ORFs (numbering according to WSSV-TH, acc. no. AF369029; van Hulten *et al.*, 2001a)									
92	Ribonucleotide reductase (large subunit)	138330	←	140876	2547	138635	139536	902	3'	
						139437	140552	1116	5'	x
*hr*4										
93		141913	←	142233	321	140982	142276	1295		
94		142498	→	143082	585	142498	142738	241		
95		143118	←	143342	225	143126	143342	217		
96		143569	→	144687	1119	143785	144775	991		x
97		144689	→	146314	1626	145323	146325	1003		x
98	Ribonucleotide reductase (small subunit)	146357	→	147733	1377	146720	147731	1012		
99	Endonuclease	147798	→	148733	936	147798	148733	936		
100		148770	←	151829	3060	149779	151084	1306		
101		152015	→	152788	774	152015	152609	595		
102		152788	→	153624	837	153070	153619	550		positive at 0 h.p.i.
103		153704	→	156274	2571	154992	156256	1265	3'	
						154034	155031	998	5'	x
*hr*5										
104		156538	←	156927	390	in *hr*; no probe				
105		156746	→	156955	210	in *hr*; no probe				
106		157493	→	158107	615	157044	158104	1061		x
107		158204	→	159031	828	158204	159031	828		
108		159076	←	163896	4821	159947	160966	1020	3'	x
						161135	162021	887	5'	x
109	in virion (VP15)	163996	→	164238	243	164053	164238	186		
110		164030	←	164314	285	overlap ORF109; no probe				
111		164346	←	167930	3585	164887	165867	981	3'	
						165932	166990	1059	5'	
112	Class I cytokine receptor/ in virion (vp674)	168000	→	170024	2025	168689	169779	1091		
113		170043	→	172577	2535	171444	172521	1078	3'	
						170471	171059	589	5'	x
114		172701	→	175511	2811	174142	175268	1127	3'	
						173065	174191	1127	5'	x
115		175716	→	175964	249	175716	175964	249		
116		176120	←	177967	1848	176707	177860	1154		
117		178367	←	179251	885	178134	179279	1146		
118	in virion (vp292)	179527	→	180405	879	179249	180217	969		
119		180442	→	181884	1443	180947	181861	915		
120	in virion (vp300)	181937	→	182839	903	181937	182839	903		
121		182911	→	185286	2376	184241	185223	983	3'	
						183051	184337	1287	5'	
*hr*6										
122		185588	→	185818	231	in *hr*; no probe				
123		185843	→	186073	231	in *hr*; no probe				
124		186135	→	186374	240	186135	186374	240		x
125		186534	→	188747	2214	187386	188594	1209	3'	
						187315	188351	1037	5'	
126		188918	→	190420	1503	189444	190505	1062		
127	in virion (vp281)	190500	→	191345	846	190372	191313	942		
128	in virion (vp384)	191349	→	192503	1155	191349	192503	1155		
129		192564	→	193493	930	192488	193513	1026		
130		193553	←	196321	2769	194634	195369	736	3'	
						195256	196248	993	5'	
131		196571	←	197416	846	196571	197416	846		
132		197480	←	198949	1470	197567	198833	1267		
133		198967	←	199479	513	198978	199478	501		x
134		199492	→	203151	3660	201762	202795	1034	3'	
						200650	201747	1098	5'	x

ORF nr	putative function/ name*	start ORF		end ORF	length ORF	start probe	end probe	length probe	part of gene detected	not detected/ remark
135		203364	→	205739	2376	203638	204989	1352	3'	
						203382	204542	1161	5'	
136		205865	→	206029	165	205774	206776	1003		
*hr*7										
137		207118	←	207279	162	in *hr*; no probe				
138		207790	→	207999	210	207790	207939	150		x
139		207992	←	208159	168	no probe				
140		208153	→	210057	1905	208896	209941	1046		x
141		210064	→	210366	303	210064	210366	303		x
142		210519	→	213821	3303	212453	213429	977	3'	
						211483	212401	919	5'	positive at 0 h.p.i.
143		213918	←	218612	4695	214183	215485	1303	3'	
						215902	216995	1094	5'	x
144		218566	←	218859	294	218660	218859	200		x
145		218912	→	219532	621	218912	219529	618		
146		219631	→	220260	630	219631	220260	630		
147		220309	←	221238	930	220313	221238	926		
148		221305	←	221874	570	221305	221874	570		
149	TATA box binding protein/ in virion (vp184)	221977	→	224652	2676	222471	223725	1255	3'	x
						222186	223166	981	5'	x
150		224639	→	225898	1260	224637	225859	1223		x
151	in virion (vp466)	225923	→	227323	1401	226413	227403	991		
152		227329	→	228147	819	227329	228140	812		
153	in virion (VP26)	228221	←	228835	615	228211	228835	625		
154		229074	←	232613	3540	229657	230628	972	3'	x
						230510	231248	739	5'	x
155		232928	→	233281	354	232928	233246	319		x
156		233295	→	233978	684	233295	233492	198		
157		233982	←	234230	249	234007	234204	198		
158		234229	→	235626	1398	235008	236108	1101		x
*hr*8										
159		237222	←	239792	2571	237578	238307	730	3'	
						238145	239420	1276	5'	x
160		239925	→	242285	2361	240988	242199	1212	3'	
						240513	241648	1136	5'	
161		242377	←	243678	1302	242247	243245	999		
162		243701	←	244552	852	243701	244547	847		x
163		244556	←	245341	786	244556	245341	786		x
164		245444	←	245746	303	245444	245746	303		
165		245849	←	250966	5118	246582	247516	935	3'	
						247556	248795	1240	5'	x
166		251400	←	258392	6993	251676	252663	988	3'	x
						252648	253699	1052	5'	x
167		258666	→	276899	18234	275586	276750	1165	3'	
						258772	259853	1082	5'	x
168	in virion (vp68)	277040	←	277246	207	277040	277242	203		
169		277425	→	279614	2190	278753	279541	789	3'	
						277592	278575	984	5'	positive at 0 h.p.i.
170	anti-apoptosis	279667	→	280632	966	279683	280731	1049		
171	Chimeric Thymidine kinase-Thymidylate kinase	280653	→	281849	1197	281031	281782	752		
172		281869	→	282384	516	281869	282372	504		
173		282433	→	282816	384	282433	282811	379		
*hr*9										
174		282829	←	283380	552	in *hr*; no probe				
175		284246	←	284401	156	in *hr*; no probe				
176		284646	←	284843	198	in *hr*; no probe				
177		285406	→	287331	1926	285623	286715	1093		
178		287386	→	288165	780	287386	288164	779		

Title row: **WSSV-TH ORFs (numbering according to WSSV-TH, acc. no. AF369029; van Hulten *et al.*, 2001a)**

WSSV-TH ORFs (numbering according to WSSV-TH, acc. no. AF369029; van Hulten *et al.*, 2001a)									
ORF nr	putative function/ name*	start ORF		end ORF	length ORF	start probe	end probe	length probe	part of gene detected	not detected/ remark
179		288183	←	288866	684	288080	289064	985		
180		289149	←	289343	195	289149	289343	195		x
181		289474	→	289680	207	289474	289680	207		x
182	in virion (VP19)	289998	←	290363	366	289998	290363	366		
183	in virion (vp544)	290501	→	292135	1635	291304	292350	1047		
184		292511	→	292804	294	292511	292792	282		

*hr*s (numbering according to WSSV-TH, acc. no. AF369029)							
hr nr	start *hr*	end *hr*	length *hr*	start probe	end probe	length probe	not detected
*hr*1	28250	30320	2071	29243	30328	1086	x
*hr*2	73550	77150	3601	73526	74259	734	x
*hr*3	126388	128112	1725	126852	128093	1242	x
*hr*4	141139	141827	689	140518	141843	1326	x
*hr*5	156319	157366	1048	156381	157412	1032	
*hr*6	185500	186155	656	185215	186394	1180	
*hr*7	206140	207726	1587	206149	207356	1208	
*hr*8	235672	237156	1485	235726	236816	1091	
*hr*9	283323	285125	1803	283338	284278	941	x

ORFs TH-96-II (numbering according to WSSV-CN, acc. no. AF332093; Yang *et al.*, 2001)								
ORF nr‡	putative function/ name	start ORF		end ORF	length ORF	start probe	end probe	length probe	not detected
CN-479		275207	←	276736	1530	276137	276736	600	
CN-482		277035	→	277574	540	277035	277574	540	
CN-483		277705	→	278079	375	277705	278079	375	
CN-486		278637	→	280976	2340	278637	279236	600	x
CN-489		281128	←	281865	738	281266	281865	600	x
CN-492		282176	→	282586	411	282176	282586	411	
CN-493	in virion (vp35)	282674	←	283360	687	282762	283360	599	
CN-495		283754	→	284014	261	283754	284014	261	
CN-497		284076	←	285773	1698	285174	285773	600	x
CN-500		286077	←	286706	630	286122	286706	585	

Shrimp genes		
ORF	(similar to) acc. no.	length probe
Actin	AF100986	686
Elongation factor 1-alpha	AY117542	301
Cytochrome c oxidase	AW497588	468
Similar to fruit fly 's ubiquitin 52-AA extension protein	AW600779	341
NADH dehydrogenase	AF436051	160
Nucleoside diphosphate kinase	BF024215	427
Ribosomal protein P2	BF024238	352
Calponin-like protein	AW497581	587
Cytochrome B	AF125382	470
Guanine nucleotide-binding protein	BF023988	311
Carbonic anhydrase 1	BF024146	453
ATP-binding subunit of serine protease	BE188550	208
Elongation factor 2	AW618928	307
ATP synthase	AI253861	393
Ribosomal protein S20	BF024253	451
TNF precursor	To be submitted	402

Negative controls		
ORF	(similar to) acc. no.	length probe
GFP (*Aequorea victoria*)	M62653	966
Lyk3 (*Medicago truncatula*)	AY372406	550
Nork-i (*Medicago truncatula*)	AJ418369	500
Enod 12 (*Medicago truncatula*)	X68032	500
HCR4/ Cf-4/9 (*Medicago truncatula*)	AY372416	700

*virion proteins indicated to be present "in virion" have been published by van Hulten *et al.* (2001a), Chen *et al.* (2002a) and Huang *et al.* (2002b)

Table 5.2. (Relative) expression levels of the WSSV ORFs on the microarray, sorted by intensities of the probes.

ORF nr*	putative function/ name†	Time course expression level (in % of maximal expression)					intensity‡	TATA-box§	consensus Poly(A)-signal§	type
		T=0h	T=8h	T=20h	T=32h	T=48h				
1	in virion (VP28)	0	0	44	82	100	+++	0	1	L
27	DNA polymerase	0	0	51	74	100	+++	1	1	L
44		0	0	47	100	76	+++	0	1	L
75	in virion (vp357)	0	0	13	78	100	+++	1	1	L
76		0	0	37	100	90	+++	1	1	L
94		0	0	9	65	100	+++	1	1	L
109	in virion (VP15)	0	0	52	90	100	+++	1	1	L
115		0	0	100	91	76	+++	0	1	E
146		0	0	100	94	92	+++	1	1	E
153	in virion (VP26)	0	0	54	83	100	+++	0	1	L
171	Chimeric tk-tmk kinase	0	0	100	50	42	+++	1	1	E
182	in virion (VP19)	0	0	36	92	100	+++	1	1	L
28		0	0	48	68	100	++	0	1	L
31	in virion (VP24)	0	0	44	82	100	++	0	1	L
55		0	0	100	69	41	++	1	1	E
95		0	0	20	84	100	++	0	0	L
125		0	0	100	97	100	++	1	1	E
135		0	0	39	84	100	++	1	1	L
152		0	0	100	86	88	++	0	1	E
156		0	0	77	84	100	++	0	1	E
168	in virion (vp68)	0	0	40	81	100	++	1	1	L
8		0	0	83	86	100	+	0	1	E
12		0	0	78	87	100	+	1	1	E
23		0	0	100	73	75	+	1	1	E
25		0	0	88	100	83	+	1	1	E
30	Collagen/ in virion (vp1684)	0	0	58	76	100	+	0	1	L
32		0	0	61	89	100	+	1	1	L
33		0	0	45	75	100	+	0	1	L
37		0	0	81	83	100	+	1	1	E
43		0	0	39	86	100	+	0	0	L
49		0	0	78	90	100	+	1	1	E
54	Thymidylate synthase	0	0	12	82	100	+	1	0	L
58		0	0	81	97	100	+	0	1	E
60		0	0	100	90	91	+	0	1	E
65		0	0	46	80	100	+	0	1	L
69		0	0	99	93	100	+	1	1	E
70		0	0	100	89	82	+	1	1	E
81		0	0	100	80	65	+	0	0	E
85		0	0	100	86	84	+	0	1	E
98	Rr (small subunit)	0	0	100	89	83	+	1	1	E
103		0	0	99	100	86	+	0	1	E
107		0	0	91	100	84	+	1	1	E
118	in virion (vp292)	0	0	56	96	100	+	1	0	L
121		0	0	49	77	100	+	1	1	L
126		0	0	100	72	54	+	0	1	E
128	in virion (vp384)	0	0	41	76	100	+	1	1	L
129		0	0	57	79	100	+	0	1	L
131		0	0	87	95	100	+	1	1	E
136		0	0	36	76	100	+	1	1	L
142		0	0	100	94	95	+	0	1	E
143		0	0	45	74	100	+	1	1	L
145		0	0	80	73	100	+	1	0	E
147		0	0	100	92	94	+	1	1	E
157		0	0	52	75	100	+	1	1	L
159		0	0	68	97	100	+	0	1	E
160		0	0	95	100	73	+	0	1	E
161		0	0	80	95	100	+	1	1	E
164		0	0	96	100	84	+	1	1	E
167		0	0	41	97	100	+	0	1	L
170	Anti-apoptosis	0	0	100	60	63	+	0	1	E
173		0	0	100	55	48	+	0	1	E
179		0	0	100	52	51	+	1	1	E
*hr*6		0	0	77	100	90	+			E
2	Protein kinase	0	0	92	77	100	+-	1	1	E
4		0	0	20	74	100	+-	1	0	L
34	in virion (vp95)	0	0	65	76	100	+-	0	1	L
38		0	0	45	74	100	+-	1	1	L
56		0	0	78	83	100	+-	1	1	E
57		0	0	59	92	100	+-	0	0	L

ORF nr*	putative function/ name†	Time course expression level (in % of maximal expression)					intensity‡	TATA-box§	consensus Poly(A)-signal§	type
		T=0h	T=8h	T=20h	T=32h	T=48h				
80		0	0	49	75	100	+-	1	1	L
89	Latency related	0	0	100	59	64	+-	1	1	E
91		0	0	94	96	100	+-	1	1	E
101		0	0	100	64	74	+-	0	1	E
113		0	0	70	76	100	+-	1	1	L
114		0	0	65	77	100	+-	1	1	L
116		0	0	100	88	79	+-	1	1	E
117		0	0	86	79	100	+-	0	1	E
119		0	0	56	91	100	+-	1	1	L
120	in virion (vp300)	0	0	55	88	100	+-	1	0	L
127	in virion (vp281)	0	0	90	100	95	+-	0	1	E
134		0	0	64	91	100	+-	0	0	L
151	in virion (vp466)	0	0	67	70	100	+-	1	1	L
177		0	0	86	87	100	+-	1	1	E
3	Latency related	0	0	47	86	100	+--	1	0	L
7		0	0	51	71	100	+--	0	0	L
9		0	0	83	82	100	+--	1	1	E
10		0	0	60	41	100	+--	0	0	L
11		0	0	100	82	89	+--	0	1	E
14		0	0	47	83	100	+--	0	0	L
15		0	0	74	70	100	+--	0	1	E
16		0	0	90	100	73	+--	1	1	E
24		0	0	80	93	100	+--	1	1	E
29	in virion (vp448)	0	0	84	64	100	+--	1	1	E
35		0	0	50	79	100	+--	0	1	L
36		0	0	43	81	100	+--	0	1	L
39		0	0	58	93	100	+--	0	0	L
41		0	0	63	100	62	+--	0	1	L
53		0	0	76	92	100	+--	1	1	E
61	Protein kinase	0	0	85	94	100	+--	1	0	E
66		0	0	100	73	61	+--	0	1	E
67		0	0	100	89	83	+--	1	0	E
72		0	0	29	100	95	+--	0	0	L
73		0	0	58	80	100	+--	0	1	L
74		0	0	77	100	87	+--	1	0	E
77		0	0	49	93	100	+--	1	0	L
79		0	0	45	77	100	+--	1	1	L
84		0	0	51	72	100	+--	0	1	L
90		0	0	50	78	100	+--	0	1	L
92	Rr (large subunit)	0	0	100	73	66	+--	1	1	E
93		0	0	96	97	100	+--	0	1	E
99	Endonuclease	0	0	100	72	84	+--	0	1	E
100		0	0	47	67	100	+--	0	1	L
111		0	0	100	85	100	+--	0	1	E
112	Class I cytokine receptor/ in virion (vp674)	0	0	86	100	85	+--	0	0	E
130		0	0	48	84	100	+--	1	1	L
132		0	0	81	83	100	+--	0	0	E
148		0	0	63	72	100	+--	0	0	L
165		0	0	100	52	68	+--	0	1	E
169		0	0	83	100	92	+--	1	1	E
172		0	0	100	90	86	+--	1	0	E
178		0	0	83	89	100	+--	1	1	E
183	in virion (vp544)	0	0	42	66	100	+--	0	0	L
184		0	0	65	80	100	+--	0	0	L
*hr*5		0	0	76	100	100	+--			E
*hr*7		0	0	70	100	92	+--			E
*hr*8		0	0	67	73	100	+--			L

*ORF and *hr* numbering in accordance with van Hulten *et al.* (2001a)

†virion proteins indicated to be present "in virion" have been published by van Hulten *et al.* (2001a) and Huang *et al.* (2002b)

‡semi-quantitative levels of gene expression (intensity) were categorized using the following classification (numbers are arbitrary expression units): background threshold-3,000: +--; 3,000-5,000: +-; 5,000-20,000: +; 20,000-40,000: ++; 40,000 or higher: +++

§0 means not present; 1 means present. A TATA-box was considered present in case the sequence TATA(a/t)A appeared within 0 till 300 nt upstream of the translational startcodon; A poly(A)-signal was considered present in case the sequence A(a/t)TAAA appeared within -50 till 300 nt downstream of the translational stopcodon

¶according to Fig. 5.2a
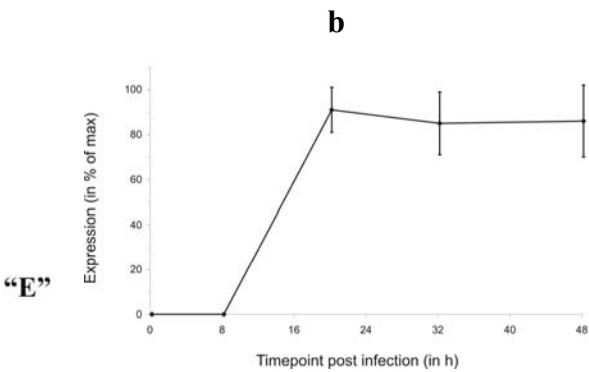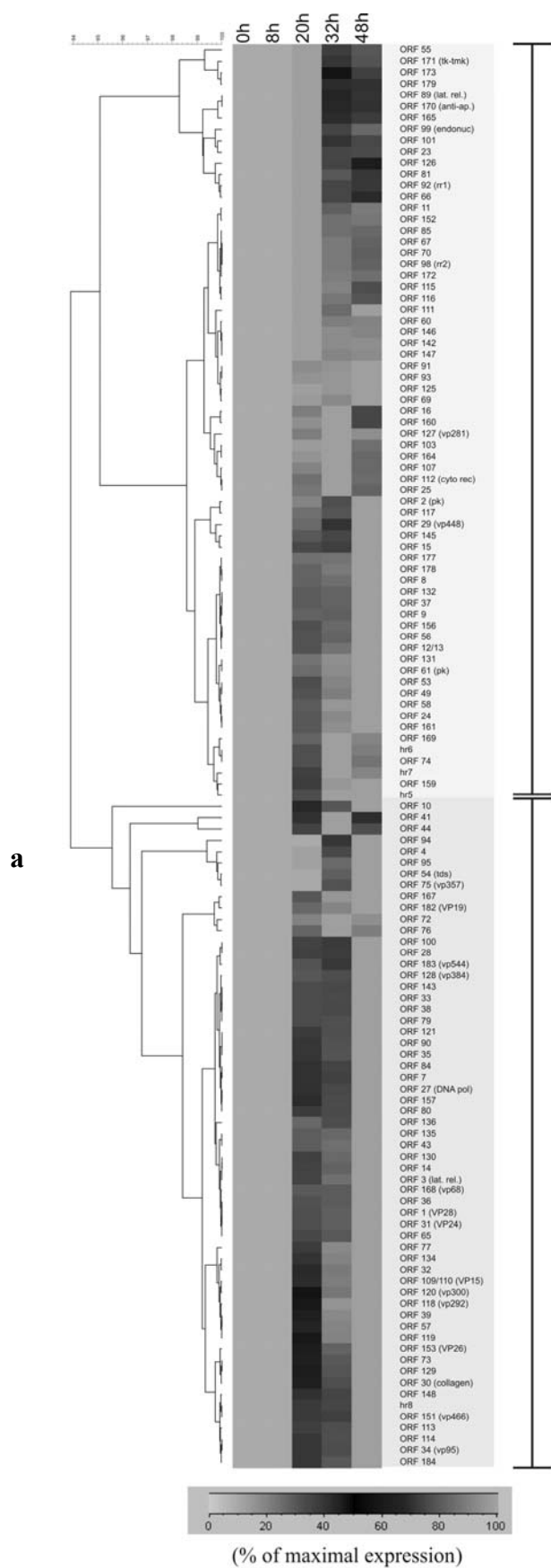
Figure 5.2. Hierarchical clustering (Euclidean distance clustering algorithm) of relative WSSV gene expression profiles in *P. monodon* gill tissue. Columns indicate separate time points, and every row displays the expression profile of a single ORF. The dendogram at the left clusters the ORFs based on the relatedness of their gene expression patterns **(a)**. Mean expression profile (in % of maximal expression; standard deviations are indicated for each time point) of the ORFs of cluster "E" indicated in Fig. 5.2a **(b)**. Mean expression profile (in % of maximal expression; standard deviations are indicated for each time point) of the ORFs of cluster "L" indicated in Fig. 5.2a **(c)**.
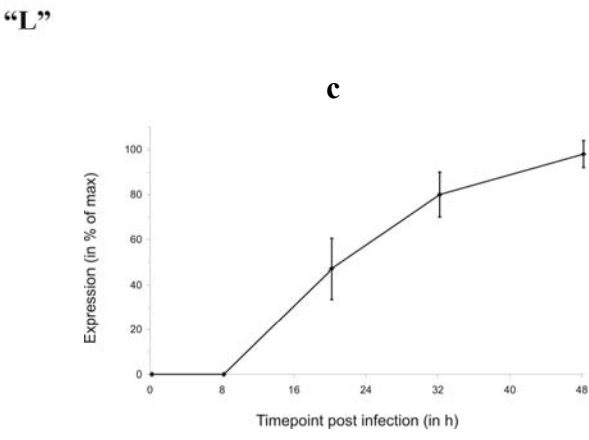
Transcription profiling: The Cy3/Cy5 ratios of the 125 WSSV genes that we could specifically detect were normalized to the maximal expression of each gene (100%). The constructed transcription profiles are presented in Table 5.2. To examine the relationship between the WSSV genes, these relative expression data were analyzed by hierarchical clustering (Euclidean distance). Fig. 5.2a shows that the WSSV genes clustered into two major groups, boxed yellow and blue. The mean transcription patterns of these clusters are shown in Fig. 5.2b & 5.2c, respectively.

After the genes of the "yellow" cluster reached (almost) a maximum amount of mRNA in the shrimp gill tissue at 20 h.p.i., the total amount of transcripts stayed at this level or declined slightly, depending on the gene (Fig. 5.2b). Most WSSV putative early genes, such as the ribonucleotide reductases (*rr1* and *rr2*), the chimeric thymidine kinase-thymidylate kinase (*tk-tmk*) and both protein kinases (*pk*), genes which are considered to be expressed before viral DNA replication, were present within this group (Fig. 5.2a). Therefore, we designated the genes in this cluster "E" ("putative Early-type"). For the "blue" cluster, almost all genes showed a maximal amount of mRNA in the gill tissue at 48 h.p.i. The genes of the "blue" cluster include all major structural virion protein genes (*VP*s), as well as most minor virion protein genes (*vp*s; Fig. 5.2a). Structural virion protein genes are supposed to be expressed after viral replication, and therefore considered late genes. Genes of the blue cluster were designated "L" ("putative Late type").

WSSV gene expression levels

Semi-quantitative levels of gene expression were evaluated by comparing the normalized absolute values of the Cy3 signals at time points of maximum expression (100%) of each particular gene (Fig. 5.3; Table 5.2). These absolute gene expression levels should be interpreted semi-quantitatively, as the DNA fragments spotted on the microarray are different in length and GC content (resulting in different hybridization efficiencies), and various amounts of DNA might have been spotted for different genes and on different microarray slides. Also, not all WSSV transcripts present in the time course samples might be labeled with equal efficiency, especially because it is not known whether all mRNAs of WSSV are polyadenylated. Except for *vp24*, all major structural protein genes show a very high expression level (Table 5.2). Most genes with a relatively high expression level (semi-quantitative) cluster in the putative late type group and encode a consensus poly(A)-signal (Table 5.2). Fig. 5.3 shows that putative early and late type genes, as well as genes of different classes of gene expression level, are distributed randomly over the genome. The presence of a TATA-box in the promoter region of a gene is not positively correlated with its temporal expression class or expression level (Table 5.2).
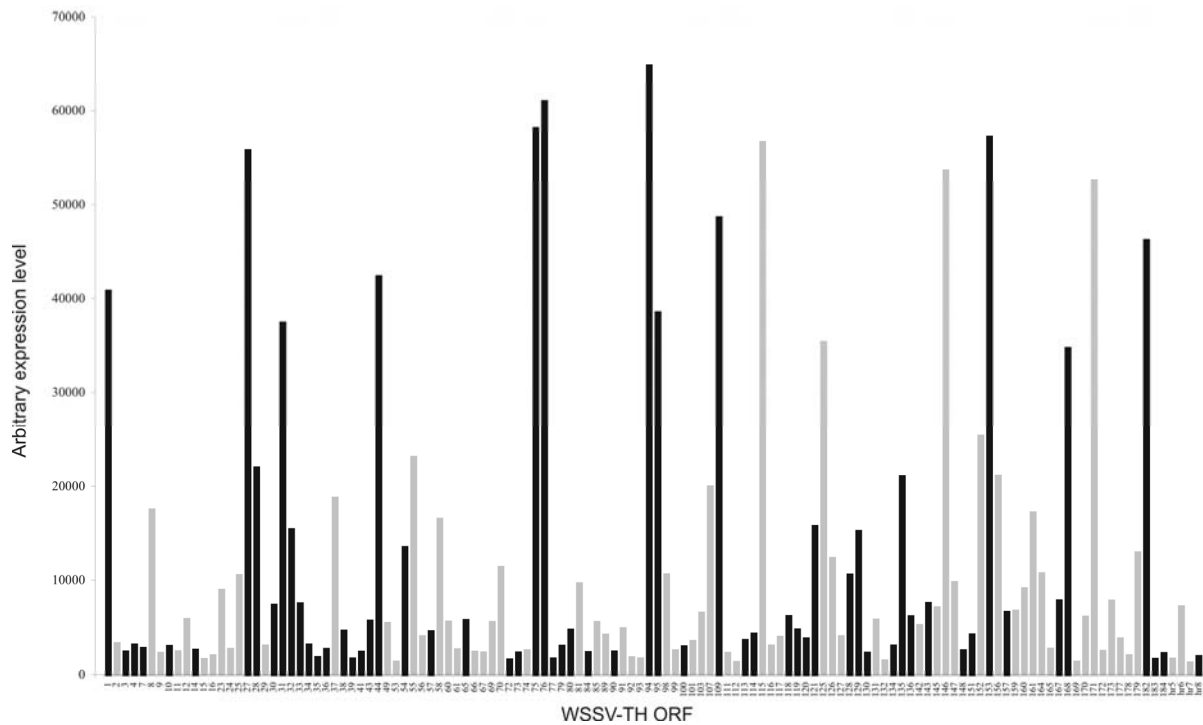
Figure 5.3. Semi-quantitative expression levels (arbitrary units) of WSSV genes. Bars representing genes of cluster "E" are colored gray, bars representing genes of cluster "L" are colored black.

5'/3' Probes

For detection of the 42 largest WSSV genes, two probes per gene were spotted on the microarray, corresponding to the 5' and 3' end, respectively. Of the forty-two 5' end probes, 28 (67%) did not gave a signal above the background threshold during the time course, while for the 3' end probes only 8 (19%) could not be detected. This is probably caused by the RT-reaction, which proceeds from the poly(A)-tail at the 3' end of the transcripts. As most 5' end probes represent parts of the genes which are at least 1.5-2 kb upstream of the poly(A)-tail, probably these 5' ends of the mRNA were not so efficiently reverse transcribed. For the ORFs of which both 5' and 3' ends were detectable, the profile over the time course was very similar, suggesting that both were indeed detecting the same messenger. These data confirm the current annotation of ORFs on the WSSV genome based on the computational analysis by van Hulten *et al.* (2001a) and Yang *et al.* (2001), which are very similar (Marks *et al.*, 2004).

Comparison transcriptome WSSV-TH and TH-96-II

TH-96-II contains a large genomic fragment of ~13.2 kb in a locus known as "variable region ORF23/24", which is absent in WSSV-TH (Marks *et al.*, 2005a). Probes for the 10 genes encoded by the ~13.2 kb fragment of TH-96-II were included on our WSSV microarray (Table 5.1; CN-ORFs). During the complete time course of the previous section, for which WSSV-TH was used, a signal below the background threshold was obtained for these ten probes, excluding the possibility of (a-specific) cross hybridization of WSSV-TH genes to

these probes. Poly(A)$^+$ RNA was isolated from WSSV-TH and from TH-96-II infected *P. monodon* gill tissue at 2 days post WSSV infection, labeled and hybridized to different microarrays, as described in the previous section.

      As the data obtained from the on-array duplicates were very similar (data not shown), the average was used for further analysis. The scatterplot in Fig. 5.4a shows a comparison between the average ratios obtained from infected tissue of WSSV-TH (one microarray slide) and TH-96-II (other microarray slide). The majority of the probes are between the dotted lines, indicating an expression difference of <2 times between WSSV-TH and TH-96-II. Genes showing a two fold expression difference or more are marked.
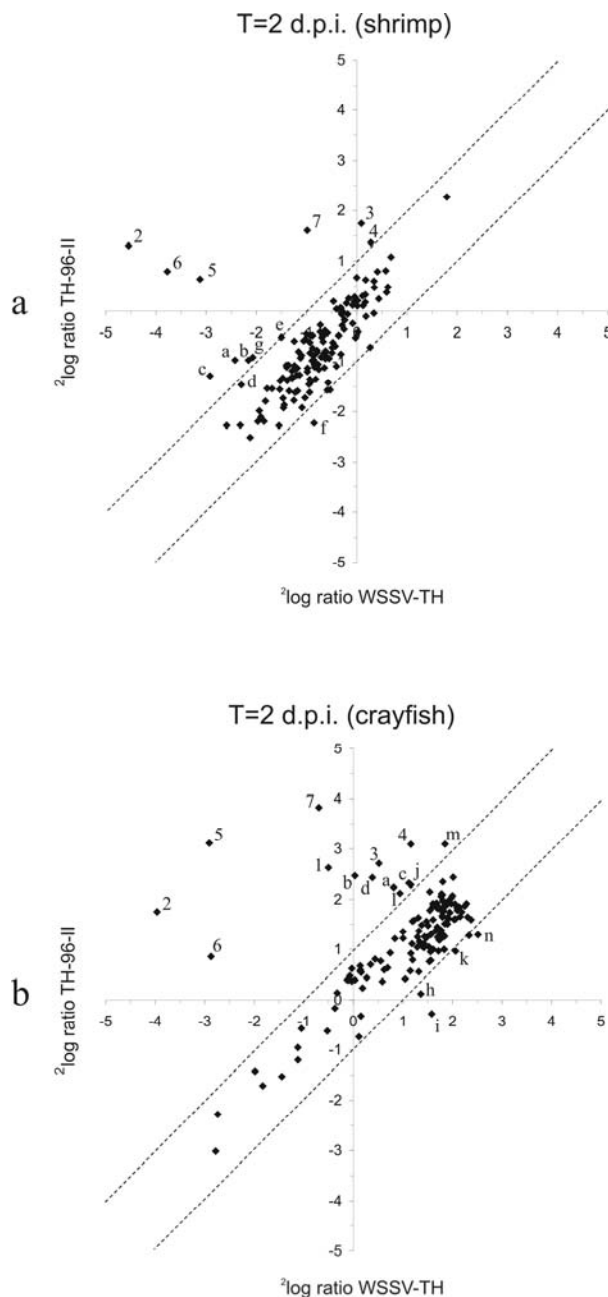
Figure 5.4. Comparison of the transcriptome of WSSV-TH (x-axis) with TH-96-II (y-axis) at 2 d.p.i. in shrimp **(a)** and in crayfish **(b)**. Scatterplot of the expression levels obtained for WSSV-TH and TH-96-II. The dotted line indicate a 2-fold difference in expression between the isolates. The labeled probes (letters or numbers in case of CN-ORFs) are >2 times differentially expressed between WSSV-TH and TH-96-II and represent probes for: 1: CN-ORF479; 2: CN-ORF482; 3: CN-ORF483; 4: CN-ORF492; 5: CN-ORF493; 6: CN-ORF495; 7: CN-ORF500; a: ORF12/13; b: ORF14; c: 3'probe ORF24; d: 5' probe ORF24; e: ORF23; f: ORF77; g: ORF85; h: ORF44; i: ORF118; j: ORF127; k: ORF131; l: ORF145; m: ORF170; n: ORF178.

For the TH-96-II infected *P. monodon* tissue, we could detect 7 of the 10 genes encoded by "variable region ORF23/24" (Table 5.1). CN-ORF486, CN-ORF489 and CN-ORF497 expression (ORF assignment by Yang *et al.*, 2001) could not be detected. CN-ORF482, CN-ORF493 (VP35) and CN-ORF495 were expressed highly, indicated as ++ in Table 5.2. As the CN-ORFs are assigned the background threshold Cy3 value for WSSV-TH, because these genes are absent from this genome, most of the CN-ORFs (marked l to r) appear highly expressed by TH-96-II compared to WSSV-TH in the graphs of Fig. 5.4. ORF12/13 (as these two genes completely overlap, they cannot be distinguished using our microarray), ORF14, ORF23, ORF24 (both 5' and 3' probe) and ORF85 were present around 2 times in excess in the TH-96-II infected *P. monodon* gill tissue, while the only gene that showed a relative expression of >2 times in WSSV-TH infected tissue was ORF77 (Fig. 5.4a).

A similar comparison between WSSV-TH and TH-96-II in the crayfish *A. leptodactylus* (Fig. 5.4b) could detect the same 7 of the 10 genes encoded by "variable region ORF23/24" in TH-96-II infected tissue as detected for *P. monodon*. Also, ORF12/13, ORF14 and ORF24 (5' and 3' probe) were present >2 times in excess in the TH-96-II infected *A. leptodactylus* gill tissue (Fig. 5.4b). Furthermore, ORF127, ORF145 and ORF170 were expressed >2 times in excess in TH-96-II infected tissue, while ORF44, ORF118, ORF131 and ORF178 showed a relative expression of >2 times in WSSV-TH infected tissue (Fig. 5.4b)

Discussion

The WSSV genome contains 184 putative WSSV ORFs based on computational analysis (van Hulten *et al.*, 2001a; Yang *et al.*, 2001). Yang *et al.* (2001) confirmed transcription of around 50 of these ORFs (28%) using RT-PCR on a cDNA cocktail, while Tsai *et al.* (2004) detected 39 minor virion protein genes using a small, one dye (Cy3) based microarray. Using our WSSV microarray analysis (Cy3/Cy5), we could detect transcription of 79% of all putative WSSV ORFs (excluding those encoded within *hr*s; Table 5.1), indicating that most WSSV predicted ORFs are transcriptionally active. However, as we could not detect mRNAs strand specific due to the fact that dsDNA fragments were used as probes on the microarray, it cannot be excluded that some of the signals obtained originate from non-annotated ORFs encoded by the opposite strand. Except for ORF6 (vp800), ORF71 (dUTPase) and ORF149 (TATA-box binding protein), all currently annotated genes were detected. We could also show expression of almost all genes detected previously (Yang *et al.*, 2001; Tsai *et al.*, 2004). Most baculoviruses *hr*s, which have a structure similar to WSSV *hr*s, are non-coding regions (Possee & Rohrmann, 1997). Our data suggest transcriptional activity within WSSV *hr*5, *hr*6, *hr*7 and *hr*8. Further research should elucidate the nature of the transcripts that are detected within these WSSV *hr*s.

Around 20% of the putative WSSV ORFs did not reach the background threshold (Table 5.1). It is possible that these genes are not expressed in our settings (gills), but more

likely it is caused by the detection limits of our microarray experiment. About half of the ORFs not detected encode a consensus poly(A)-signal, indicating their transcripts are probably polyadenylated. Therefore, it is not likely that we could not detect these genes because of our poly(A)-based detection methods. More plausible explanations are (i) a (very) low expression of the gene or (ii) inferior hybridization properties of the probe spotted on the microarray. Other detection methods, such as quantitative RT-PCR reactions, can be used to obtain more information about the expression of these genes.

Our data provide evidence that WSSV gene transcription is regulated in a cascaded fashion. At least two major classes of gene products were distinguished, designated putative early (E) and putative late (L) (Fig. 5.2a). The transcription profiles (Fig. 5.2b & 5.2c) follow the classical expression patterns of these gene types shown for other large invertebrate dsDNA viruses such as baculoviruses in cell culture (Friesen, 1997; Lu & Miller, 1997). In case of baculoviruses, transcription of early genes declines when late gene transcription becomes very high, as the down-regulation of early transcription is directly or indirectly due to late gene expression or viral DNA replication. The fact that the "putative early" cluster of WSSV does not show this decline (Fig. 5.2b) is probably caused by asynchronous infection of the gill tissue. Cells infected in a second or third round of infection in the gills will express early genes at a later stage after injection. The presence of most putative early genes (genes involved in nucleotide metabolism, DNA replication and protein modification) in the "putative early" class and most putative late genes (virion protein genes) in the other class ("putative late") supports the accommodation of transcripts into two classes. The arrangement of WSSV genes into different kinetic classes of gene expression is also supported by the analysis of individual genes using RT-PCR (Tsai *et al.*, 2000a; Tsai *et al.*, 2000b; Liu *et al.*, 2001; Chen *et al.*, 2002b; Marks *et al.*, 2003; Li *et al.*, 2004a; Li *et al.*, 2004b; Tsai *et al.*, 2004). To obtain further support, future microarray studies could include the testing of different inhibitors such as cycloheximide and phosphono-acetic acid to distinguish between gene expression before and after protein synthesis and viral replication, respectively. Such experiments could also shed some light on the classification of the WSSV DNA polymerase and thymidylate synthase as "late" genes by our microarray analysis (Fig. 5.2a), which is inconsistent with previous reports (Chen *et al.*, 2002b; Li *et al.*, 2004b). A synchronized infection would enable a more precise time schedule of gene expression but awaits the availability of a suitable shrimp cell culture system.

In case of large dsDNA viruses like herpesviruses (Ebrahimi *et al.*, 2003) and baculoviruses (Lu & Miller, 1997), late genes are often abundantly transcribed. Our data show similar results for WSSV late genes as most of the highly expressed genes are of the late type. Furthermore, all WSSV major structural protein genes show a very high expression level, except for *vp24* (Table 5.2). The RT-PCRs performed for these genes (Marks *et al.*, 2003) confirm these results, showing high expression levels late in infection and a lower expression level of *vp24*. Also ORF75 and ORF94, two ORFs containing regions of tandem repeats

which are highly polymorphic between several WSSV isolates (Marks *et al.*, 2004; Dieu *et al.*, 2004), show high expression levels (Table 5.2). The protein encoded by ORF75 is most likely located in the virion (vp357; Huang *et al.*, 2002b). No functional data are available for ORF94. Of the three "putative early" genes with a very high expression level, our findings obtained for the chimeric thymidine kinase-thymidylate kinase gene (ORF171) confirmed the results of Tsai *et al.* (2000b) concerning temporal expression class (early) and expression level (very high). No data are available on the function of the other two genes, ORF115 and ORF146.

A comparison in transcriptome was made between the WSSV isolates WSSV-TH and TH-96-II, both in *P. monodon* and in *A. leptodactylus*. The main difference between these isolates is the presence of two large genomic fragments in TH-96-II, ~5.3 kb at a locus known as "variable region ORF14/15" and ~13.2 kb at "variable region ORF23/24", which are both absent in WSSV-TH (Marks *et al.*, 2005a). The genes encoded by the additional fragments are dispensable for infection and replication in these species, as both are permissive host for WSSV-TH (Marks *et al.*, 2004). Our experiments show that most genes encoded by the ~13.2 kb fragment in TH-96-II are transcriptionally active in these two crustacean species, some even to a relatively high level. Therefore it is likely that the gene products, although they are not essential, do have a functional role in both species.

Although the expression level of most ORFs shared by both isolates was similar between WSSV-TH and TH-96-II (Fig. 5.4), some ORFs were differentially expressed (>2 times difference). The expression of ORF12/13, ORF14, ORF23 and ORF24 was higher for TH-96-II compared to WSSV-TH infected gill tissue, both in *P. monodon* and in *A. leptodactylus* (except for ORF23, which is only higher in *P. monodon*). These genes are located at the junction sites of "variable region ORF14/15" and "variable region ORF23/24", but completely present in both isolates. As the TH-96-II 5' upstream regions of ORF14 and ORF24 are absent in WSSV-TH, they could contain important promoter elements involved in expression of these genes. ORF14 is not essential for virus infection and replication in at least *P. monodon*, as a WSSV isolate lacking ORF14 (WSSV-TW, acc. no. AF440570) has been isolated from this species. The significance of the decreased expression of this gene in WSSV-TH infected tissue remains unclear. As the region coding for ORF24 and the complete coding regions of ORF12/13 and ORF23, including the putative promoter regions, are present in all WSSV isolates characterized thus far (Marks *et al.*, 2004), these ORFs most probably have an essential function during virus infection. Recent data suggest a higher virulence of WSSV-TH compared to TH-96-II in *P. monodon* (Marks *et al.*, 2005a). Although the difference in virulence could be explained by a replication advantage of WSSV-TH, having a smaller sized genome, the ORFs differentially expressed between WSSV-TH and TH-96-II in *P. monodon* could also play a role. In this respect, ORF12/13, ORF23, ORF24, but also ORF85, which are all expressed lower in WSSV-TH *P. monodon* infected tissue, and ORF77,

the only gene present >2 times in excess in WSSV-TH infected tissue (Fig. 5.4a), are of interest.

Evaluation of *in vivo* transcription profiles using microarrays provides a first step in understanding WSSV transcription regulation and gene function on a genome-wide scale. Besides providing insights into the basic biology of the virus the microarray can also be used to test the effect of drugs on viral replication and gene expression and consequently produce information that can aid in the development of effective treatments against WSSV.

Materials and Methods

Virus infection
Characteristics of the virus isolates WSSV-TH and TH-96-II were described by van Hulten *et al.* (2001b) and Marks *et al.* (2005a), respectively. *P. monodon* (approximately 35 gram) or *A. leptodactylus* (approximately 35 gram) was injected intramuscularly with purified WSSV, using a relatively high dose to synchronize infection in the gills as much as possible. At various time points after injection, three animals were randomly selected, frozen in liquid nitrogen and stored at -80°C.

Poly(A)$^+$ RNA isolation
Total RNA was isolated from gills as described before (Marks *et al.*, 2003). For each time point post infection gills of three animals were pooled. Approximately 70% of the 184 WSSV genes encode a consensus poly(A)-signal (AAUAAA), or another consensus poly(A)-like signal which could be sufficient for polyadenylation (like AUUAAA), within -50 till 300 bp downstream of their translational stopcodons (Birnstiel *et al.*, 1985; Sheets *et al.*, 1990; van Hulten *et al.*, 2001a; Yang *et al.*, 2001). Therefore, we used poly(A)-based RNA isolation and Cy3/Cy5 labeling methods. Poly(A)$^+$ RNA was purified using the polyATtract mRNA Isolation System III (Promega). The yield of poly(A)$^+$ RNA from total RNA was generally between 0.5% and 1.5% as quantified by measuring A_{260} using a NanoDrop spectrophotometer (NanoDrop Technologies).

Construction WSSV microarrays
WSSV is known to encode several ORFs with high nucleotide identity, accommodated in so called gene families (van Hulten *et al.*, 2001a). To minimize the possibility of WSSV transcripts hybridizing with non-specific probes, we decided to use PCR products (~300-1000 bp in size) instead of oligonucleotides (~50 nt in size) as probes on the WSSV microarrays. The viral array elements include probes for 158 of the 184 putative WSSV-TH ORFs (Table 5.1). Probes for the 22 putative genes encoded within the WSSV *hr*s were not included. Probes for WSSV-TH ORF12 and ORF110 were not spotted as these ORFs are almost completely encoded within the coding regions of ORF13 and ORF109, respectively. For

ORF68 and ORF139, we failed to obtain the respective DNA fragments (for unknown reason). For all WSSV-TH genes larger than 2100 bp (42 genes), in addition to the 3' end probe an extra probe corresponding to the 5' end of the genes was spotted on the microarrays (the 3' end probes were used for quantification purposes, the 5' end probes only as controls). For each WSSV *hr* one probe was included on the microarrays. Probes to detect the ten TH-96-II specific ORFs (CN-ORFs in "Variable region ORF23/24"; Marks *et al.*, 2005a) were also included on the WSSV microarrays. For each WSSV ORF, overlap of the corresponding DNA fragment selected to use as specific probe with (5'/3' UTRs of) neighbouring ORFs was avoided. In addition, probes of the following sources were included on the microarrays (1) 16 cellular shrimp genes to evaluate normalization between the several chips; (2) a set of background controls consisting of 4 genes of the plant *Medicago truncatula* and the jellyfish *Aequorea victoria* GFP gene, each spotted in quintuplicate; (3) the complete coding sequence of the firefly luciferase gene and three partial luciferase clones encompassing the 5', middle and 3' part of the gene, all spotted in quadruplicate (Table 5.1). As the samples were spiked with luciferase mRNA prior to labeling, this allowed correction of the expression ratios between samples for differences in the preparation of labeled cDNA and for differences during the hybridization to the microarrays. The partial luciferase clones were additionally used to monitor the integrity of the labeled sample cDNA. Each of the in total 272 probes was printed in duplicate on each microarray slide to evaluate the consistency of the signals obtained.

Preparation probes

Part of the DNA fragments used as WSSV-TH probes were obtained by performing PCR reactions on DNA clones of the WSSV-TH DNA bank constructed for sequencing of the complete viral genome (van Hulten *et al.*, 2001a), using universal primers. The other part of the WSSV-TH probes and all TH-96-II specific probes (CN-ORFs) were obtained by performing PCR reactions on genomic DNA of WSSV-TH and TH-96-II, respectively, using specific primers designed by Primearray (Raddatz *et al.*, 2001). The cellular shrimp genes, negative controls and luciferase controls were obtained by performing PCR reactions on purified plasmids containing the anticipated fragments, using universal primers. The shrimp genes were originally amplified from a cDNA library of uninfected *P. monodon* available in our laboratory, using specific primers (sequences obtained from GenBank), and cloned into the pGEM-T easy vector (Promega). Sequencing confirmed that these clones contained the anticipated sequences. All PCR reactions were performed using the Expand Long Template PCR System (Roche).

Preparation WSSV microarrays

The PCR products were column purified using the High Pure PCR Product Purification Kit (Roche) and diluted to 0.1 $\mu g \ \mu l^{-1}$ (in a total of 100 μl), as measured by A_{260}. All PCR

products showed a clear band of the appropriate size and concentration by agarose gel electrophoresis. Ten µg of the PCR products was dried to completion and dissolved in 10 µl 5x SSC. Microarrays were prepared by spotting individual DNA fragments on GAPS amino-silane-coated glass slides (Corning) with a PixSys 7500 arrayer (Cartesian Technologies) equipped with Chipmaker 3 quill pins (Telechem). Spotting volumes were 0.5 nl, resulting in a 120 µm spot diameter at a pitch of 160 µm. After drying overnight, the microarrays were rehydrated with steam, snap-dried (95°C-100°C) and UV cross-linked (150 mJ). The slides were soaked twice in 0.2% SDS (2 min), twice in MQ water (2 min) and into boiling MQ water (2 min). After drying (5 min), the slides were rinsed 3 times in 0.2% SDS (1 min), once in MQ water (1 min), submerged in boiling MQ water (2 sec), and air-dried.

Synthesis Cy3/Cy5 labeled cDNA

Purified poly(A)$^+$ RNA preparations were labeled using a standard protocol for cDNA synthesis: 2.5 µg poly(A)$^+$ *P. monodon* RNA (spiked with 1 ng luciferase mRNA (Promega); for *A. leptodactylus* samples 1 µg poly(A)$^+$ RNA was used) and 2.5 µg oligo(dT)$_{21}$ primer were heated to 65°C (3 min) and then placed at 25°C (10 min) to anneal the primer. RNA was reverse transcribed using Superscript II (Invitrogen) according to the protocol of the manufacturer, with the exception that 40% of dTTP was replaced with 5-(3-aminoallyl)-2'dUTP. After precipitation and washing of the cDNA/RNA hybrids, the RNA was hydrolyzed with 0.2 M NaOH at 37°C (10 min). The solution was neutralized with 0.15 M HEPES (pH6.8) and 0.15 M HCl. After precipitation and washing, the cDNA was resuspended in 5 µl 0.1 M carbonate buffer pH 9.3. Dyes were covalently bound to the incorporated amino-groups by adding 5 µl of 5 mM Cy3 or Cy5 reactive dyes (Amersham Pharmacia) in dimethyl-sulfoxide. This mixture was incubated at room temperature in the dark (1 h). Unincorporated dye was removed by performing ethanol precipitation twice, after which the labeled cDNA was dissolved in 5 µl MQ water. Samples under study were labeled Cy3, the reference sample was labeled Cy5. Reference samples used were (i) for the time series: a mixture of poly(A)$^+$ RNA harvested 0, 1 and 2 days after WSSV-TH infection, containing a pool of transcripts representing all of the genes present in the time course and (ii) for the isolate comparison: a mixture of poly(A)$^+$ RNA harvested 0 and 2 days h.p.i. from WSSV-TH as well as from TH-96-II infected shrimp and crayfish gill tissue, providing additionally Cy5 signals for the TH-96-II specific ORFs.

Microarray hybridizations

After prehybridization of the slides for 2 h at 42°C in hybridization buffer (50% formamide, 5x Denhardt's reagent, 5x SSC, 0.2% SDS and 0.1 mg ml^{-1} fish DNA), the slides were washed by dipping in MQ water, followed by dipping in isopropanol. Slides were dried by centrifugation at 160g (1 min). Hybridization occurred in a volume of 65 µl using a covered hybridization frame (Gene Frame 15mm x 15mm (65 µl); ABgene). After heating 65 µl

hybridization buffer containing both Cy3 and Cy5 labeled cDNA sample to 95°C for 1 min, it was loaded into the hybridization chamber. The slides were hybridized during 24 h at 42°C. Following hybridization, the slides were washed in 1x SSC/ 0.1% SDS (5 min), 0.1x SSC/ 0.1% SDS (5 min) and briefly rinsed in 0.1x SSC. Slides were dried by centrifugation at 160x g (1 min).

DNA microarray analysis

Slides were scanned for fluorescence emission with a ScanArray ExpressHT (Perkin Elmer) at 75% laser power and a resolution of 10 μm, using an attenuation of 65% (Cy3) or 60% (Cy5). The resulting Cy3 and Cy5 images were stored as TIFF files and individually processed. For each array element, the integrated optical density was determined within a defined circle, using AIS software (Imaging Research, Canada). Average background values, calculated from the hybridization signals of the *Medicago truncatula* and *Aequorea victoria* probes, were subtracted to correct for non-specific fluorescence. Next, Cy3 or Cy5 signals not reaching 0.5x background value were set to this cut-off (0.5x background value) for the respective dye. Elements for which both the Cy3 and Cy5 signal not reached 0.5x background value were discarded from further analysis.

Normalization of the two samples in each hybridization was done with the mean hybridization signal of the full-length luciferase probes. Finally, the Cy3/Cy5 ratio (R_{ij}) was calculated for each element. The reference sample used on each slide was the same within the time series or within the isolate comparison, allowing direct comparison of the different hybridization experiments. Expression ratios for the on-array duplicates (R_{ij1}, R_{ij2}) were calculated separately and the average of both values was used for further analysis. Improper duplicates ($|{}^2\log(R_{ij1}/R_{ij2})| > 1$) were filtered out. Microsoft Excel was used for organizing data and for statistical analyses. The Cy3 background threshold for WSSV gene expression was set at 1.5 times the expression value initially obtained from the Cy3 background control probes. WSSV genes with a lower Cy3 expression value (below the background threshold) were considered not detected.

Analysis expression data

For WSSV genes, average normalized ratios obtained for each gene were converted into percentages of the maximal expression of each gene over the time series. The WSSV probes with a signal above the background threshold at 0 h.p.i. and showing no increase in ratio later in WSSV infection were discarded from the analysis (5 probes; Table 5.1). The remaining WSSV probes showed a signal below the background threshold at 0 and 8 h.p.i., and the percentages at these time points were set at 0%. Next, cluster and correlation analysis of WSSV transcription profiles was performed using GeneMaths software (Applied Maths). For individual shrimp genes, the ${}^2\log$ expression ratios were normalized by subtracting the average of the ${}^2\log$ ratios for the respective genes over the time course. To enable comparison

with WSSV genes, these normalized ratios were converted to percentages, setting the average expression of the 16 shrimp genes (Table 5.1) over the time course on 100%.

Chapter 6

Transcriptional analysis of the White spot syndrome virus major virion protein genes

Abstract

White spot syndrome virus (WSSV) is a member of a new virus family (*Nimaviridae*) infecting crustaceans. The regulation of transcription of WSSV genes is largely unknown. Transcription of the major WSSV structural virion protein genes, *vp28*, *vp26*, *vp24*, *vp19* and *vp15* was studied, in search for common promoter motifs for coordinate expression. We examined the temporal expression of these genes and both 5' and 3' ends of the mRNA were determined, using infected crayfish gill tissue as a RNA source. RT-PCR showed that all five genes are expressed late in infection compared to the early ribonucleotide reductase large subunit gene. 5' RACE studies revealed a consensus late transcription initiation motif for only two of the five major virion protein genes. This motif was only found in one other upstream region of the putative translational start site of a gene with unknown function (ORF 158). No other conserved sequence motifs could be detected in the sequences surrounding the transcriptional start sites of the five major virion protein genes. All 5' ends were located about 25 nucleotides downstream of an A/T rich sequence, including the consensus TATA-box sequence for *vp15*. The absence of a consensus motif is distinct from gene regulation of other large dsDNA viruses and suggests a unique regulation of WSSV transcription, in line with its unique taxonomic position.

Introduction

White spot syndrome virus is a large, circular, double stranded (ds) DNA virus infecting shrimps and other crustaceans (Wang *et al.*, 1998; Wongteerasupaya *et al.*, 1995). The virion particles, approximately 275 x 120 nm in size, are enveloped and have an ovoid-to-bacilliform shape with a tail-like appendage at one end (Durand *et al.*, 1997; Nadala *et al.*, 1998). Sequencing of three different WSSV isolates revealed that the dsDNA genome is about 300 kb in size (van Hulten *et al.*, 2001a; Yang *et al.*, 2001; GenBank AF440570). Most of the 184 putative ORFs identified on the WSSV genome are unassigned, as they lack homology to known genes in public databases (van Hulten *et al.*, 2001a). Based on its unique morphological and genetic features, WSSV has been accommodated in a new virus family, the *Nimaviridae* (genus *Whispovirus*).

The WSSV particle consists of five major and about 14 minor structural proteins (van Hulten *et al.*, 2002; Huang *et al.*, 2002b). The five major proteins have been named according to their sizes in SDS-PAGE: viral protein (VP) 28, VP26, VP24, VP19 and VP15 (van Hulten *et al.*, 2000a, c, 2002). VP26, VP24 and VP15 are present in nucleocapsid preparations, while VP28 and VP19 are found in envelope fractions of the virions. As an antiserum against VP28 was able to neutralize WSSV infection in the shrimp *Penaeus monodon*, this protein is most likely located on the surface of the virus particle and plays a key role in systemic WSSV infection in shrimp (van Hulten *et al.*, 2001b). A putative function for VP15, a highly basic protein with no hydrophobic regions, is that of a histone-like, DNA-binding protein (van Hulten *et al.*, 2002; Zhang *et al.*, 2001).

Genes of most large dsDNA 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. For proper late gene expression, the motif that contains the transcription initiation site (TIS) often plays a prominent role in the recognition by a virus-coded RNA polymerase (Davison and Moss, 1989; Garcia-Escudero and Viñuela, 2000; Kim *et al.*, 2002; Morris and Miller, 1994; Weir, 2001). Information about pathogenicity, epidemiology, virion structure and the genomic sequence of WSSV is now available, but insight into gene regulation is limited. Transcriptional analysis has been performed on WSSV genes encoding the large and small subunit of ribonucleotide reductase (*rr*), protein kinase (*pk*), the chimeric thymidine kinase-thymidylate kinase (*tk-tmk*) and DNA polymerase (*dnapol*). These genes were identified in *P. monodon* 2 or 4 h post WSSV infection (Liu *et al.*, 2001; Tsai *et al.*, 2000a, b; Chen *et* al., 2002b), indicating that these genes are of the early type. The transcription of *rr1*, *rr2*, *pk* and *dnapol* is initiated 20-28 nucleotides downstream of a TATA box, suggesting a functional role for the TATA box during early transcription. Both *rr* genes and *dnapol* share a consensus WCABT (W=^{a}/t; B=c/g/t) sequence in which transcription is initiated (Chen *et al.*, 2002b). This somewhat degenerated consensus TIS motif could be an early promoter element of WSSV.

The present study focuses on the transcription and regulation of WSSV late genes. As no cell lines or primary crayfish or shrimp cell cultures are available, experiments are performed *in vivo*. Temporal gene expression of the major structural protein genes *vp28*, *vp26*, *vp24*, *vp19* and *vp15,* putative late genes, was studied by transcriptional analysis using RT-PCR. 5' and 3' RACE analyses were used to map the TISs and polyadenylation sites of these major structural protein genes.

Results

All five WSSV major structural virion protein genes are present as single copies in the WSSV genome (van Hulten *et al.*, 2001a). Polyadenylation motifs are present downstream of, or overlapping with, the translational stop codon of the ORFs, suggesting that the transcripts are polyadenylated. The regions upstream of the translational start sites of the major structural protein genes all contain stretches of A/T-rich sequences. Only *vp15* contains a consensus TATA box sequence, 87 nt upstream of the translational start codon. The nucleotides surrounding the methionine start codons of the five major structural protein genes are consistent with the Kozak rule for efficient eukaryotic translation initiation (Kozak, 1989).

RT-PCR

Temporal WSSV gene expression of the major structural virion protein genes was studied in crayfish using RT-PCR. To compare WSSV transcription in crayfish (*Orconectes limosus*) with shrimp (*Penaeus monodon*), both species were infected with WSSV using injection to synchronize infection. The gills, which are well infected by WSSV early in infection, were used for the isolation of RNA. RNA was isolated from crayfish from 2 h post infection till 7 days post infection (d.p.i.), and from shrimp at 1 and 2 d.p.i.. RT-PCR was performed on DNase I treated RNA for the WSSV genes *vp28*, *vp26*, *vp24*, *vp19, vp15* and *rr1* (Fig. 6.1). Actin mRNA of the host was included in this analysis as an internal control for RNA extraction and was detected in every sample (Fig. 6.1). Purified WSSV DNA was taken as a positive control for the PCR and no template as a negative control (Fig. 6.1, C+ and C- respectively). PCR was also performed on RNA that was not reverse-transcribed. These PCRs were all negative, confirming that no detectable amount of viral genomic DNA was present in the RNA samples used. *Rr1* gene transcripts were detected from 6 h p.i. until 7 d.p.i. (Fig. 6.1), which is comparable with data obtained by Tsai *et al.* (2000a). Compared to the *rr1* gene, the *vp* genes are all transcribed late in infection. Transcripts of *vp15* were clearly detected from 16 h p.i., while transcripts of *vp28*, *vp26*, *vp24* and *vp19* were detected from 1 d.p.i. onwards. The RT-PCR suggests that mRNA expression of *vp24* was the lowest (very weak band at 1 d.p.i.) of the major structural protein genes. This is in conformity with the low presence of the VP24 protein in WSSV virions (van Hulten *et al.*, 2002). The RT-PCRs of 1 and 2 d.p.i. showed that there is little if any difference in transcription of the WSSV major structural protein genes between *O. limosus* and *P. monodon*.
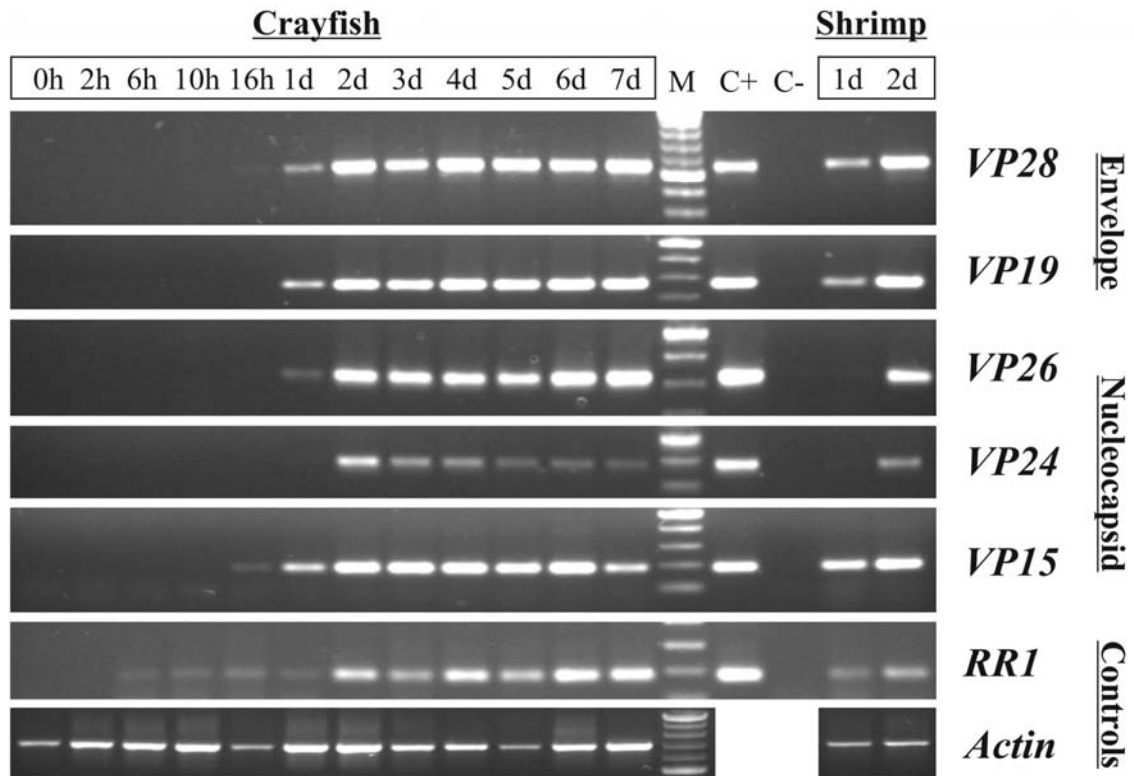
Figure 6.1. RT-PCR time course for the WSSV mRNAs of *vp28*, *vp19*, *vp26*, *vp24*, *vp15* and *rr1*, and the host messenger *actin* in crayfish *O. limosus* and shrimp *P. monodon*. Lane headings show time points hours (h) or days (d) post infection. M represents a 100 bp DNA marker, C+ and C- are respectively the positive and negative control for the PCR.

Transcription initiation sites (TIS)

To determine the TISs of the five major structural protein genes, 5' RACE was performed, using RNA isolated from gills of the crayfish *O. limosus*. A timepoint of 6 days post infection was chosen for the analysis of the mRNA, as abundant mRNAs are present for all analyzed genes (Fig. 6.1). The 5' RACE protocol was slightly modified for *vp28* and *vp24* as long stretches of thymidines are present in the 5' upstream sequences, which could result in aspecific binding of the oligo(dT) anchor primer to the cDNA. Therefore, the cDNA of these genes was 3'-tailed with dTTPs instead of poly(A)-tailing and an oligo(dA) anchor primer was used for the PCR. The 5' RACE PCR products were cloned and for each gene at least three clones were sequenced. The results of the 5' RACE experiments are shown in Fig. 6.2.

The TIS of *vp28* could be located within the nucleotide sequence AAC, 33 nt upstream of the ATG. A single 5' RACE clone was identified with a 5' mRNA terminus 5 nt upstream of the translational start codon and is most probably the result of RNA breakdown, although the possibility of a second TIS cannot be excluded. The TIS of *vp26* is located on GC, 71 nt upstream of the ATG. For *vp24* a TIS was identified on CA, 27 nt upstream of the ATG and for *vp19* the TIS was identified within 64-71 nts from the translational start codon. A minor TIS (one clone) for *vp19* was found 151 nts upstream of the translational start codon (not shown in Fig. 6.2). Transcription from the initiation sites identified for *vp19* is not

Upstream sequences (showing results 5' RACE)

```
VP28: cacgaggttgtcatcaccaaaggtaacctttttttttgtcctcgccgacaaaacgacatcttaataaccaagc
                                                                          5*
VP19: agacaaataggttcgtaaggtgtcctgacaaaaaccgtacacgataaaacaacaggtctttacgttacattga
VP26: taaaaacaaaccctttctgggtatatggtgctacagaagaactaactagctggatccaaccaacacgtaaagg
VP24: tcacaatatatctcagtggtgttttctacgactaaataaaatataacaaactttaatatctgttttttctct
VP15: ataagacaaaaattataaattgggaagaccgatacagtctttcatgacaaaataccccgagaacaaaagattg
                        3*
```

Upstream sequences - continued -   **Downstream sequences**

```
VP28: aacgttcgataaagaaaaaaactcgtcatg - ORF VP28 - taaataaatcgtgcttttttatataga
VP19: cgtacctcttcatcaaacagaaaaaaaatg - ORF VP19 - taaaaatatagaaaacacaacatcagg
VP26: aagaacttccatctaaaacaaagaaaaatg - ORF VP26 - taaagtgggaattgtacaaatataaag
VP24: catgacctttgtacaactttcttcaaaatg - ORF VP24 - aataaacacattttttaaaaatatacc
      5*c                                                               3*
VP15: ttgtctaggaacaaagaaacattaaaaatg - ORF VP15 - taattcttccctgtacaacaactatgt
```

Downstream sequences - continued - **(showing results 3' RACE)**

```
VP28: tagggaatttttaatattacaacaataagaaaataaacaattgaggaaatttataccatatttttattgaccta
                                                         2*
VP19: gacaaaaattataattaaaacacctttttattataaaataaaatttatttgaaacttatgcattgtttgtata
VP26: gttttgtttgaataaaaatacaagtaattttataccatcttttattttctaatcctttgaaatgtatcttgt
                                    2* last t
VP24: tttattttattaaccaattatttacattgtagcttgaaattcatatatatacctttcattgttagggagagat
VP15: tatttaattgattttttttcttctgaataattggaaataataaacatccattgaaacttatgcagtattttt
                                                            2*
```

Figure 6.2. Sequences 100 nucleotides upstream and downstream of the major structural protein genes of WSSV showing the major transcription initiation and polyadenylation sites of each gene, indicated by arrows (TIS by arrow above sequence, polyadenylation by arrow below sequence). The 5' and 3' termini of the different clones sequenced for each gene are underlined (each underlined nucleotide represents a clone). The number with the underlining shows the amount of similar clones. The start codon of each gene is shaded black. The TATA-box for *vp15* is double underlined. The stop codons are shaded dark grey and the poly(A) signals light grey.

regulated by the TATA box, which is located 254 nts upstream of the translational start codon. For *vp15* the TIS was found 24 nts downstream of the TATA-box. Transcription starts 57 nts upstream of the translational start codon of *vp15*.

Analysis of WSSV putative promoter motifs

The TISs of the major structural virion protein genes have been aligned to identify putative consensus motifs in the regions surrounding the TISs of the WSSV major structural protein

genes (Fig. 6.3). Transcription of *vp24* and *vp15* is initiated in exactly the same sequence, TCATGAC. The conserved TCATGAC motif of *vp15* and *vp24* occurs 37 times in the total genome sequence of which three times in a putative promoter region. Except for *vp24* and *vp15*, the sequence is found in the 100 bp sequence upstream of the putative transcriptional start site of ORF158, which encodes a protein with no homology to proteins present in GenBank. This motif, however, is not present in the 5' upstream sequences of the other *vp*s. Transcription of all *vp*s starts 25 nts downstream of an A/T-rich sequence, which includes the TATA-box present in the *vp15* 5' upstream sequence (Fig. 6.3). No other conserved motifs could be detected in the sequences surrounding the transcriptional start sites of the five major structural proteins.

```
VP28-5'UTR : cagctcgcttgccaattgtcctgttacgtactctgtggtttcacgaggttgtcatcacca :  60
VP19-5'UTR : ggcgttgaagaagatatagtgaacagaattgacggaacatgaccaactgcgacaactaac :  60
VP26-5'UTR : gggtagtgggaggtatcgttactgaaattagatcacggcaaggtacacttcatccggttc :  60
VP24-5'UTR : ttctttcgagtatgctgaaacggcagttgctgctcacaatatatctcagtggtgttttct :  60
VP15-5'UTR : cgagacatggctcctgtatctacaaaagagctcgcttaaccacagttgcactcacccggg :  60


VP28-5'UTR : aaggtaacctttttttttgtcctcgccgacaaaacgacatcttaataaccaagcaacgtt : 120
VP19-5'UTR : gctgagtaagtctaaaaagacaaataggttcgtaaggtgtcctgacaaaaccgtacacg  : 120
VP26-5'UTR : gatgtagggttaaattttaaaaacaaaccctttctgggtatatggtgctacagaagaact : 120
VP24-5'UTR : acgactaaataaaatataacaaactttaatatctgtttttttctctcatgacctttgtac : 120
VP15-5'UTR : tacaataagacaaaaattataaattgggaagaccgatacagtctttcatgacaaaatacc : 120
```
                    A/T rich region   **TIS**
```
VP28-5'UTR : cgataaagaaaaaactcgtcatg------------------------------------- : 144
VP19-5'UTR : ataaaacaacaggtctttacgttacattgacgtacctcttcatcaaacagaaaaaaaatg : 180
VP26-5'UTR : aactagctggatccaaccaacacgtaaaggaagaacttccatctaaaacaaagaaaaatg : 180
VP24-5'UTR : aactttcttcaaaatg------------------------------------------- : 136
VP15-5'UTR : ccgagaacaaaagattgttgtctaggaacaaagaaacattaaaaatg------------ : 167
```

Figure 6.3. Alignment of the TISs in the 5' upstream region of the major structural virion protein genes. Shading is used to indicate the occurrence of identical nucleotides (black, 100%; dark gray, 80%; light gray, 60%). The TISs and A/T-rich region are indicated. The translational start codon of each gene is included in the sequence.

Polyadenylation

To determine the polyadenylation sites of the major structural protein genes, a 3' RACE was performed for all five genes on RNA isolated 6 d.p.i. A clear 3' RACE PCR product was obtained for each of the genes. This product was cloned and at least two clones for each gene were sequenced. For each of the genes, the polyadenylation starts 15/16 nucleotides downstream of the poly(A) signal, as shown in Fig. 6.2. Two internal 3' RACE clones (one for *vp28* and one for *vp19*; not shown in Fig. 6.2) were found, probably caused by aspecific binding of the oligo(dT) anchor primer to long stretches of adenosines in the 3' untranslated region.

Discussion

WSSV has a broad host range infecting shrimp, crayfish and other crustaceans. Its dsDNA genome of around 300 kb (van Hulten *et al.*, 2001a; Yang *et al.*, 2001; AF440570) is among

the largest viral genomes, but little is known about gene regulation of this virus. To obtain insight in this process, transcription of five putative late WSSV genes was studied using crayfish as host. The five major structural protein genes were selected as we assumed that their expression is coregulated to secure correct assembly of the virions. The WSSV virus isolate derived from Thailand was used in this study (van Hulten *et al.*, 2001a). The three complete WSSV genome sequences available so far are very homologues (sequence identity of over 99%) and the major differences occur in repeat regions and in a 13 kb region absent in the Thailand isolate (position 31135). Putative promoter sequences analyzed in this study are identical in all three isolates.

The transcription of the WSSV genes was studied using RT-PCR as this technique allows the most sensitive detection of transcripts. Compared to the expression of *rr1* (Tsai *et al.*, 2000a), which is presumed early as it should precede DNA replication to provide dNTPs, the WSSV major structural proteins genes are late (Fig. 6.1). This is in accordance with many structural virion protein genes from other large DNA viruses such as baculoviruses, herpesviruses and African swine fever virus (ASFV). All six WSSV mRNAs studied in the RT-PCR could be detected very late in infection (2 to 7 days p.i., Fig. 6.1), which can be explained by the multiple rounds of WSSV infection which occur in gill tissue. In contrast to the expression of the other major structural protein genes, *vp15* is clearly expressed at 16 h p.i. preceding the other *vp*s, which suggests an earlier role of VP15 in the viral replication and assembly process as DNA binding protein. *Vp15* is the only major structural protein gene with a consensus TATA box (Fig. 6.2). The earlier onset of transcription of *vp15* (Fig. 6.1) could be regulated by this TATA box, as in other WSSV genes which have been shown or are likely to be expressed early in infection (*rr1, rr2, pk* and *dnapol*) (Tsai *et al.*, 2000a; Liu *et al.*, 2001; Chen *et* al., 2002b).

The transcriptional start and stop sites of the *vp*s were determined by 5' and 3' RACE. Splicing events can be excluded for these mRNAs, as has been demonstrated by the cloning and sequencing of cDNAs (Yang *et al*., 2001) and the expression of the VPs in insect cells (van Hulten *et al*., 2002). The polyadenylation sites of the five major structural protein genes, as determined by 3' RACE, were all between 15 or 16 nts downstream of the poly(A) signal (Fig. 6.2). The polyadenylation sites of *vp26* and *vp15* are in agreement with the 3' ends of these genes determined from RNA isolated from WSSV infected *P. japonicus* (Zhang *et al.*, 2001; Zhang *et al.*, 2002b), in these papers referred to as *p22* and *vp6.8* respectively. However, the TIS for *vp26* and *vp15* described by these authors, using sequencing of a single cDNA clone, do not match the TIS found in the present study. Probably part of the 5' end of the mRNAs is missing in these studies (Zhang *et al.*, 2001; Zhang *et al.*, 2002b), because intact 5' ends of mRNA are rarely recovered from cDNA libraries. Polyadenylation of the WSSV genes *rr1*, *rr2*, *pk* and *dnapol* is, respectively, 12, 13, 16 and 17 nts downstream of the polyadenylation signal (Liu *et al.*, 2001; Tsai *et al.*, 2000a; Chen *et al.*, 2002b). All identified WSSV polyadenylation sites are within the range of regular polyadenylation in eukaryotic

mRNAs, which is typically located 15 to 25 nts downstream of the sequence AAUAAA (Fitzgerald and Shenk, 1981).

For all *vp*s, except for *vp19*, only one TIS has been identified. However, the presence of other TISs cannot be excluded as less abundant mRNAs at defined timepoints are difficult to detect in an *in vivo* situation where no absolute synchronized infection can be obtained. For proper late gene expression of large dsDNA viruses, the motif that contains the TIS is often an important viral promoter element. Mutations in the late motif TAAG of baculoviruses (Morris and Miller, 1994), the late TAAAT motif of poxviruses (Davison and Moss, 1989), the TATA initiator of the ASFV (Garcia-Escudero and Viñuela, 2000) and the more diverged initiator YYANWYY (N= any nucleotide; W=^{t}/a) of herpesviruses (Kim *et al.*, 2002; Weir, 2001) often lead to a (dramatic) downregulation of gene expression. The 5' RACE analyses indicate that transcription of the WSSV genes *vp28*, *vp26*, *vp24*, *vp19* and *vp15* did not start at a consensus sequence for all five genes (Fig. 6.3). However, a conserved TIS motif (TCATGAC) was identified for *vp15* and *vp24*. This motif might be involved in WSSV gene regulation, but not for a large group of genes, as it only appears in three WSSV putative promoter regions. Interestingly, all *vp*s contained an A/T rich sequence 25 nts upstream of the TIS (Fig. 6.3) and this sequence could have a function as a promoter element, as has been shown for baculoviruses, poxviruses, and ASFV late genes. For baculoviruses, these A/T tracts serve as a protein binding site (Burma *et al.*, 1994), while for the other viruses, the A/T sequences are thought to play a role in facilitating the unwinding of the double-stranded DNA and, consequently, in the initiation of transcription (Davison and Moss, 1989; Garcia-Escudero and Viñuela, 2000). Further studies are needed to elucidate the role of the *vp15* and *vp24* TCATGAC motif and the A/T tracts in WSSV gene expression. However, the lack of a suitable cell culture system, either a cell line or primary cell cultures, hampers WSSV promoter as well as WSSV transcription studies.

Except for the features mentioned, no other conserved motifs could be detected in the sequences surrounding the transcriptional start sites of the five WSSV major structural virion protein genes. The absence of such a motif is distinct from gene regulation of other large dsDNA viruses and suggests a unique regulation of WSSV transcription, in line with its unique taxonomic position.

Materials and Methods

Virus infection
The virus isolate used in this study originates from infected *Penaeus monodon* shrimp imported from Thailand in 1996 and was obtained as described before (van Hulten *et al.*, 2000a). Crayfish *Orconectes limosus*, equally susceptible to WSSV infection as *P. monodon* and used as a model, was injected intramuscularly with purified WSSV using a 26-gauge needle to initiate infection. At various time points after injection, three crayfish were

randomly selected, frozen in liquid nitrogen and stored at –80ºC. For infection of *P. monodon*, the same infection procedure was followed, using the same virus isolate.

RNA isolation

After removal of the exoskeleton and the underlying epidermis of the frozen crayfish, gill tissue was excised with a scalpel. For isolation of total RNA, 300 mg of frozen gill tissue was homogenized in 4 ml TRIzol Reagent (Invitrogen), using a glass-Teflon homogenizer, and subjected to chloroform extraction, isopropanol precipitation and ethanol washing according to the manufacturer's recommendations. The precipitated RNA was dissolved in 40 µl RNase-free water. The concentration of RNA was quantified by measuring the absorption of the RNA solution at a wavelength of 260 nm. RNA from the gills of shrimp was isolated using the same method.

RT-PCR

For the reverse transcriptase reaction, 20 µg of isolated RNA was treated with DNaseI (Gibco BRL), according to the protocol of the manufacturer. To inactivate the DNaseI, the sample was heated to 65ºC for 10 minutes followed by a phenol/chloroform extraction and ethanol precipitation. The RNA was resuspended in 20 µl RNase free water. Five µg of RNA was used in the reverse transcriptase reaction. The RT reaction was performed using SUPERSCRIPT^{TM} RNase H$^-$ Reverse Transcriptase according to the protocol (Invitrogen). PCR on 1 µl of the obtained cDNA was performed using *Taq* DNA polymerase (Promega) with specific primer sets for the different genes, listed in Table 6.1a. For *vp28*, *vp26*, *vp24*, *vp19* and *vp15* 25 cycles were run, for *actin* and *rr1* 33 cycles were used. The PCR machine used was the Gene Amp PCR System 2400 (Applied Biosystems). Twenty µl of each PCR product was analyzed by agarose gel electrophoresis.

Table 6.1. Primers used for RT-PCR **(a)** and for the 5' and 3' RACE **(b)**.
a. RT-PCR.

Gene	Primer name	Sequence (5'-3')	WSSV sequence coordinates	Size (bp) of PCR product
vp28	VP28-F1	CACAACACTGTGACCAAG	88-105	529
	VP28-R1	TTTACTCGGTCTCAGTGCCAG	616-596	
vp19	VP19-F1	ATGGCCACCACGACTAAC	290363-290346	250
	VP19-R1	CATCGTTATCGTTGGCAGTGTC	290114-290135	
vp26	VP26-F1	ATGGAATTTGGCAACCTAACAAACCTG	228835-228809	304
	VP26-R1	GGGCTGTGACGGTAGAGATGAC	228532-228553	
vp24	VP24-F1	GCTATACTGGCGGGTTTG	50050-50033	384
	VP24-R1	GTCCACTGTTATATCCCTCTTTG	49667-49689	
vp15	VP15-F1	GAAGCTCCAAGACCAAATCCC	164063-164083	176
	VP15-R1	TTAACGCCTTGACTTGC	164238-164222	
rr1	RR1-F1	CATAAAATGGGTTCTAACCAGCAACAATC	140882-140854	187
	RR1-R1	GACGGTCCATGATGTGAGATGC	140696-140717	
Actin	Actin-F1	GAYGAYATGGAGAAGATCTGG		686
	Actin-R1	CCRGGGTACATGGTGGTRCC		

b. 5'/3' RACE.

Gene	Primer name	Sequence (5'-3')	WSSV sequence coordinates
vp28	VP28-RACE-F1	CACAACACTGTGACCAAG	88-105
	VP28-RACE-R1	CGGTCTCAGTGCCAGAGTAGGTGAC	610-586
	VP28-RACE-R2	ATCAATAGAGACGGGGGTG	492-474
vp19	VP19-RACE-F1	ATGGCCACCACGACTAAC	290363-290346
	VP19-RACE-R1	CGGACCCAGCCAGAAGCATC	290060-290079
	VP19-RACE-R2	GTGTCCTTATCAGTGTCAGAATC	290149-290171
vp26	VP26-RACE-F1	GTCATCTCTACCGTCACAGCCC	228553-228532
	VP26-RACE-R1	CCTTGATATCGATCACATTCTTGG	228241-228264
	VP26-RACE-R2	GCTGCACACGTCAATGAGGG	228317-228336
vp24	VP24-RACE-F1	GAGTATACAATAATACAAAGAG	49704-49683
	VP24-RACE-R1	GTAGCATTTAGACCTGGAGACAG	49635-49657
	VP24-RACE-R2	GTCCACTGTTATATCCCTCTTTG	49667-49689
	VP24-RACE-R3	CATAGTTCCTGGTTTGTAATGTGC	49852-49875
vp15	VP15-RACE-F1	ATGGTTGCCCGAAGCTCC	164053-164070
	VP15-RACE-R1	CTTGACTTGCGGGCTCCAGACTTC	164231-164208
	VP15-RACE-R2	GCTCCAGACTTCTTTCCGCTCTTTG	164219-164195
	VP15-RACE-R3	GCTCTTCTTTCCTGCACGCTTCTTC	164172-164148

5' and 3' Rapid Amplification of cDNA Ends (5'/3' RACE)

Both 5' and 3' RACE were carried out using a commercial 5'/3' RACE kit (Roche) following the manufacturer's instructions. Total RNA was isolated from the gills of *O. limosus* 6 d.p.i. as described above. In case of the 3' RACE, first strand cDNA was synthesized using the oligo(dT) anchor primer. The resulting cDNA was amplified using one specific forward primer (see Table 6.1b; RACE-F1 primers) and the anchor primer. For 5' RACE, the primers used for synthesis of the cDNA are described in Table 6.1b (RACE-R1 primers). This cDNA was purified using the High Pure PCR Product Purification Kit (Roche) and a homopolymeric 3' d(A)-tail was added to the cDNA in a mixture with a total volume of 20 μl, using terminal transferase and dATPs included in the kit. Five μl of this mixture was used in a first round PCR performed with an oligo(dT) anchor primer and a nested primer, shown in Table 6.1b (RACE-R2 primers). For *vp28* and *vp24*, the protocol was slightly modified. Instead of poly(A)-tailing, the cDNA was 3'-tailed with dTTPs and the first PCR cycle was performed with an oligo(dA) anchor primer instead of the oligo(dT) anchor primer. For *vp24* and *vp15*, a second round nested PCR was performed using the anchor primer and a 2nd nested primer, shown in Table 6.1b (RACE-R3 primers). The final products of the 5' and 3' RACE were cloned into the pGEM-T easy vector (Promega) and sequenced.

DNA sequencing and computer analysis

Plasmid clones carrying RACE products were sequenced at the Laboratory of Molecular Biology (Wageningen University, the Netherlands), using universal M13 forward and reverse primers. Sequence data were analyzed using the software package DNASTAR4.2 and the data were edited in GeneDoc, version 2.6.000 (Nicholas *et al.*, 1997).

Acknowledgements

Chapter 7

In silico identification of putative promoter motifs of White spot syndrome virus

Abstract

White spot syndrome virus, a member of the virus family *Nimaviridae*, is a large dsDNA virus infecting shrimp and other crustacean species. Although limited information is available on the mode of transcription, previous data suggest that WSSV gene expression occurs in a coordinated and cascaded fashion. To search *in silico* for conserved promoter motifs (i) the abundance of all 4 through 8 nucleotide motifs in the upstream sequences of WSSV genes relative to the complete genome was determined, and (ii) a MEME search was performed in the upstream sequences of either early or late WSSV genes, as assigned by microarray analysis (Marks *et al.*, 2005b). Both methods were validated by alignments of empirically determined 5' ends of various WSSV mRNAs. The collective information shows that the upstream region of early WSSV genes, containing a TATA box and an initiator, is similar to *Drosophila* RNA polymerase II core promoter sequences, suggesting utilization of the cellular transcription machinery for generating early transcripts. The alignment of the 5' ends of known late genes, including all major structural protein genes, identified a degenerate motif (ATNAC) which could be involved in WSSV late transcription. For these genes, only one contained a functional TATA box. However, almost half of the WSSV late genes, as assigned by microarray analysis, did contain a TATA box in their upstream region. This may suggest the presence of two separate classes of late WSSV genes, one exploiting the cellular RNA polymerase II system for mRNA synthesis and the other generating messengers by a new virus-induced transcription mechanism.

Introduction

White spot syndrome virus (WSSV), type species of the virus family *Nimaviridae* (genus *Whispovirus*), is a pathogen of major economic importance in cultured penaeid shrimp (Lightner, 1996; Vlak *et al.*, 2005). Histopathological studies on WSSV infected shrimp have shown that the virus mainly infects tissues of ectodermal and mesodermal origin, such as the stomach, gills, heart, gut, muscle tissue and hematopoietic tissue (Lo *et al.*, 1997; Momoyama *et al.*, 1994; Wongteerasupaya *et al.*, 1995). Infected cells within these tissues are characterized by the appearance of homogeneous hypertrophied nuclei and chromatin margination (Lightner, 1996; Wang *et al.*, 1999b; Wongteerasupaya *et al.*, 1995). WSSV particles have been mainly detected in the nuclei of infected cells, indicating that transcription, replication and virion assembly probably occur in the nucleus (Durand *et al.*, 1997; Wang *et al.*, 1995; Wang *et al.*, 1999b; Wongteerasupaya *et al.*, 1995). It is not clear how the virions are released from the nucleus of an infected cell, but this most likely occurs by budding or by rupture of the nuclear envelope and/or the cell membrane.

The circular ds DNA genome of three WSSV isolates, originating from Taiwan (WSSV-TW), China (WSSV-CN) and Thailand (WSSV-TH), have been completely sequenced (van Hulten *et al.*, 2001a; Yang *et al.*, 2001; Marks *et al.*, 2004). The genome of WSSV-TH has a size of 292,967 bp encompassing 184 open reading frames (ORFs), which are almost equally distributed on both strands (van Hulten *et al.*, 2001a). Thus far, no evidence has been obtained for the occurrence of spliced transcripts. Only 6% of the WSSV ORFs could be assigned a function based on homology with known genes in public databases (van Hulten *et al.*, 2001a). Concerning genes involved in replication and transcription of WSSV, four putative functional proteins have been annotated: a DNA helicase (ORF9), a DNA polymerase (ORF27), a cAMP-responsive element binding protein (ORF66) and a TATA box binding protein (ORF149). Furthermore, several genes involved in nucleotide metabolism, such as both subunits of a ribonucleotide reductase, a chimeric thymidine-thymidylate kinase, a thymidylate synthase, a dUTPase and an endonuclease have been identified on the genome (van Hulten *et al.*, 2001a; Yang *et al.*, 2001). Although present in various other large dsDNA viruses, no RNA polymerase or other genes involved in transcription, e.g. a poly(A)polymerase or mRNA capping enzymes, have (yet) been identified on the WSSV genome. Around 50 major or minor virion protein genes have been identified on the genome (Huang *et al.*, 2002b; Tsai *et al.*, 2004; van Hulten *et al.*, 2000a, c).

Upon infection, expression of the WSSV genes can be divided in at least an early and a late phase (Marks *et al.*, 2005b), while also an immediate-early phase might be present (Liu *et al.*, 2005). The mechanism of the switch between (immediate-) early and late WSSV gene expression, as well as the promoters and regulatory sequences involved, is largely unknown. However, many eukaryotic large ds DNA viruses of >100 kb have a coordinated and cascaded fashion of gene expression (Broyles, 2003; Friesen, 1997; García-Escudero and Viňuela, 2000; Rajčáni *et al.*, 2004). Baculoviruses and herpesviruses (both replicating in the nucleus)

as well as poxviruses and asfarviruses (both replicating in the cytoplasm) express their early genes before viral replication initiates, while late genes are expressed after the onset of viral DNA replication. Both viruses replicating in the nucleus utilize the host RNA polymerase II for early gene transcription (Friesen, 1997; Wagner *et al.*, 1995). However, for late gene transcription, herpesviruses continue to exploit the cellular RNA polymerase II system, while late transcription of baculoviruses occurs by a novel RNA polymerase that is at least partially encoded by the baculovirus genome (Lu and Miller, 1997; Mistretta and Guarino, 2005; Wagner *et al.*, 1995). The viruses replicating in the cytoplasm encode their own RNA polymerase which synthesizes early as well as late mRNAs. This RNA polymerase is encapsidated within the virus particle to enable the initiation of viral gene expression upon arrival in the cytoplasm (Broyles, 2003; Pena *et al.*, 1993).

Despite the differences in gene expression strategies, the above viruses have in common that specific nucleotide motifs involved in transcription initiation, expression kinetics and expression level have been identified in the upstream regions of individual genes. Well known promoter elements used by many viruses are the TATA box and the initiator sequence, which is located at or near the site of transcription initiation (TIS). We hypothesize that conserved promoter motifs play an important role in transcription regulation of WSSV, and they can be identified by *in silico* analysis of upstream regions of WSSV genes. As important promoter motifs are overrepresented in the 5' upstream regions of baculoviruses genes (Ayres *et al.*, 1994), we studied the relative abundance of all 4 through 8 nucleotide motifs in the upstream regions of WSSV genes compared to the complete WSSV genomic sequence. This enumeration strategy was calibrated by testing the eukaryotic large ds DNA viruses mentioned above. To further identify regulatory elements, the nucleotide composition in the upstream regions of WSSV early and late genes, as assigned by microarrays (Marks *et al.*, 2005b), is studied using MEME (Bailey and Elkan, 1994). MEME is an algorithm which searches for conserved motifs in a selected set of sequences, in this case the upstream regions of WSSV ORFs. Experimental validation is obtained by alignments of 5' ends of known WSSV early as well as late transcripts. These alignments include TISs mapped by 5'RACE in previous studies, as well as two newly determined TISs of the major structural protein genes ORF112 and ORF160. Polyadenylation of WSSV early and late genes is studied by alignment of poly(A) sites. Using this approach, we were able to find further support for the presence of coregulated clusters of WSSV genes, as well as to predict putative WSSV promoter elements involved in gene expression of these clusters.

Results

Promoter analysis using the enumeration method

In a search for putative WSSV regulatory promoter elements, we compared the abundance of all 4, 5, 6, 7 or 8 nucleotide motifs in the 100 and 200 nt upstream sequences of all WSSV genes relative to their presence in the complete WSSV genomic sequence. For validation, this enumeration method was applied on the genome sequences of the type species of more extensively studied large ds DNA viruses mentioned in the introduction: *Ac*MNPV (Baculovirus), HHV1 (Herpesvirus), Vaccinia virus (Poxvirus) and ASFV (Asfarviruses).

<u>*AcMNPV, HHV1, Vaccinia virus and ASFV*</u>: Only the analysis of the 4-mers of these viruses is shown, as these will always be included in larger motifs (Table 7.1). Most *Ac*MNPV early genes contain a functional consensus TATA box upstream of the TIS (Friesen, 1997). *Ac*MNPV initiator motifs are composed of the conserved nucleotide sequence

Table 7.1. Frequency of 4- or 5-nucleotide motifs in the 5' upstream regions of the ORFs as compared to the complete genome for the viruses *Ac*MNPV, HHV1, Vaccinia virus, ASFV and WSSV. Only the 15 motifs with the highest relative enrichment are shown for each virus. For *Ac*MNPV, HHV1 and WSSV, sequences that are part of the consensus TATA box (TATA(a/t)A) are underlined, while for Vaccinia virus and ASFV sequences only consisting of A and T residues are *italics*.

4/5-mer (motif)	Occurrence in genome[a] (% of expected occurrence[b])		Occurrence in upstream regions (% of expected occurrence[b])		Relative enrich-ment in up-stream regions	4/5-mer (motif)	Occurrence in genome[a] (% of expected occurrence[b])		Occurrence in upstream regions (% of expected occurrence[b])		Relative enrich-ment in up-stream regions
100 nt upstream						**200 nt upstream**					
***Ac*MNPV**											
taag[c]	393	(29)	90	(114)	4.0	taag[c]	393	(29)	137	(87)	3.0
<u>tata</u>	1314	(66)	172	(149)	2.3	<u>tata</u>	1314	(66)	255	(111)	1.7
<u>ataa</u>	1973	(101)	222	(198)	1.9	<u>ataa</u>	1973	(101)	363	(162)	1.6
<u>atat</u>	1616	(81)	170	(147)	1.8	<u>atat</u>	1616	(81)	268	(116)	1.4
agta	671	(49)	70	(89)	1.8	agta	671	(49)	109	(69)	1.4
aagg	473	(51)	41	(76)	1.5	gata[d]	867	(63)	137	(87)	1.4
gata[d]	867	(63)	74	(94)	1.5	aata	2230	(115)	346	(154)	1.3
cact[e]	612	(64)	52	(95)	1.5	cagt[f]	698	(73)	106	(96)	1.3
aata	2230	(115)	186	(166)	1.4	ctta	393	(28)	59	(37)	1.3
atta	1957	(98)	163	(141)	1.4	tcac[e]	669	(70)	99	(90)	1.3
gtat	949	(68)	79	(98)	1.4	gcta	541	(57)	77	(70)	1.2
cagt[f]	698	(73)	58	(105)	1.4	cccc	190	(42)	27	(52)	1.2
<u>taaa</u>	2716	(140)	222	(198)	1.4	tacc	444	(47)	63	(57)	1.2
aggg	233	(35)	19	(50)	1.4	tagt	737	(53)	104	(64)	1.2
tagt	737	(53)	58	(72)	1.4	cact[e]	612	(64)	86	(78)	1.2
HHV1											
ttag	221	(53)	17	(168)	3.2	<u>ataa</u>	372	(192)	42	(445)	2.3
ctag	182	(20)	13	(60)	2.9	<u>tata</u>	306	(159)	32	(342)	2.2
ctct	658	(76)	38	(180)	2.4	ctag	182	(20)	19	(44)	2.1
tagc	365	(41)	21	(97)	2.4	<u>taaa</u>	454	(234)	47	(498)	2.1
tcta	199	(49)	11	(111)	2.3	catt[g]	339	(83)	35	(176)	2.1
tttt	807	(426)	44	(955)	2.2	taag	254	(60)	26	(127)	2.1
tagg	342	(38)	17	(77)	2.0	ttaa	356	(185)	35	(374)	2.0

4/5-mer (motif)	Occurrence in genome[a] (% of expected occurrence[b])		Occurrence in upstream regions (% of expected occurrence[b])		Relative enrichment in upstream regions	4/5-mer (motif)	Occurrence in genome[a] (% of expected occurrence[b])		Occurrence in upstream regions (% of expected occurrence[b])		Relative enrichment in upstream regions
	100 nt upstream						**200 nt upstream**				
cata	369	(90)	18	(180)	2.0	ctct	658	(76)	64	(152)	2.0
ctta	254	(62)	12	(121)	1.9	ttag	221	(53)	21	(104)	2.0
ccta	342	(39)	16	(75)	1.9	cata	369	(90)	34	(170)	1.9
tctc	892	(103)	41	(194)	1.9	ctta	254	(62)	23	(116)	1.9
ctgt	936	(106)	40	(186)	1.8	tttt	807	(426)	72	(782)	1.8
ctac	492	(56)	21	(99)	1.8	cctt	795	(92)	68	(161)	1.8
ttcc	990	(114)	42	(199)	1.7	aaat	324	(167)	27	(286)	1.7
cact	433	(50)	18	(85)	1.7	ctat	242	(59)	20	(101)	1.7
	Vaccinia virus										
taaa[h]	4421	(94)	472	(140)	1.5	gcac	391	(66)	77	(91)	1.4
aaaa[h]	5439	(115)	564	(167)	1.5	*ataa*	4528	(96)	822	(122)	1.3
ataa	4528	(96)	446	(133)	1.4	*taaa*[h]	4421	(94)	796	(118)	1.3
aaat[h]	4454	(94)	432	(128)	1.4	tact	2010	(85)	353	(105)	1.2
cgcg	318	(107)	30	(141)	1.3	acta	2179	(92)	373	(111)	1.2
gggg	171	(57)	16	(75)	1.3	*aaat*[h]	4454	(94)	755	(112)	1.2
acac	1112	(94)	103	(122)	1.3	*aaaa*[h]	5439	(115)	911	(135)	1.2
aata	5203	(110)	479	(142)	1.3	tgca	832	(70)	138	(82)	1.2
ttaa	3720	(79)	337	(100)	1.3	*aata*	5203	(110)	857	(127)	1.2
tact	2010	(85)	179	(106)	1.3	ctac	1264	(107)	208	(123)	1.2
acta	2179	(92)	190	(113)	1.2	gccg	384	(129)	63	(148)	1.2
tata	4748	(101)	411	(122)	1.2	gcga	500	(84)	82	(97)	1.2
tgaa	1917	(81)	165	(98)	1.2	cacg	586	(98)	96	(113)	1.2
ctaa	1830	(77)	157	(93)	1.2	accc	403	(68)	66	(78)	1.2
gtaa	1921	(81)	164	(97)	1.2	cata	2186	(92)	358	(106)	1.2
	ASFV										
tata[i]	2686	(91)	261	(199)	2.2	*tata*[i]	2686	(91)	405	(154)	1.7
ataa	3146	(107)	278	(214)	2.0	*ataa*	3146	(107)	432	(166)	1.5
aatt	2634	(89)	231	(176)	2.0	*aatt*	2634	(89)	357	(136)	1.5
taat	2697	(91)	227	(173)	1.9	*taaa*	3895	(133)	516	(198)	1.5
taaa	3895	(133)	325	(250)	1.9	*ttaa*	3376	(114)	447	(170)	1.5
aata	3419	(117)	278	(214)	1.8	*aata*	3419	(117)	445	(171)	1.5
ttaa	3376	(114)	274	(209)	1.8	*atat*	3022	(102)	392	(149)	1.5
atat	3022	(102)	238	(182)	1.8	*taat*	2697	(91)	349	(133)	1.5
attt	3890	(131)	306	(231)	1.8	*ttat*	3146	(106)	406	(154)	1.5
ttat	3146	(106)	247	(187)	1.8	*attt*	3890	(131)	495	(187)	1.4
ttta	3895	(131)	296	(224)	1.7	ctaa	1172	(63)	147	(88)	1.4
atta	2697	(91)	203	(155)	1.7	*atta*	2697	(91)	337	(129)	1.4
tatt	3419	(115)	249	(188)	1.6	*tatt*	3419	(115)	426	(161)	1.4
aaaa	6731	(232)	480	(372)	1.6	*ttta*	3895	(131)	482	(182)	1.4
aaat	3890	(133)	272	(209)	1.6	*aaaa*	6731	(232)	828	(321)	1.4
	WSSV (4-mer motif)										
tata	2630	(60)	164	(128)	2.2	tata	2630	(60)	290	(117)	2.0
ataa[j]	3431	(74)	201	(150)	2.0	ataa[j]	3431	(74)	327	(126)	1.7
taaa	3538	(77)	195	(146)	1.9	accc	1333	(88)	127	(148)	1.7
aacc	1774	(79)	92	(142)	1.8	taaa	3538	(77)	333	(128)	1.7
aaaa	6450	(134)	330	(236)	1.8	aaaa	6450	(134)	570	(210)	1.6
accc	1333	(88)	68	(154)	1.8	ccgg	374	(36)	33	(56)	1.6
accg	699	(46)	33	(75)	1.6	cccc	1348	(130)	118	(202)	1.6
tacc	1424	(67)	64	(103)	1.5	cacg	995	(65)	86	(100)	1.5
caac	2792	(125)	123	(190)	1.5	aacc	1774	(79)	150	(119)	1.5
aata	4104	(89)	179	(134)	1.5	accg	699	(46)	56	(65)	1.4
tttt	6450	(160)	281	(240)	1.5	taac[j]	1888	(60)	151	(85)	1.4

4/5-mer (motif)	Occurrence in genome[a] (% of expected occurrence[b])		Occurrence in upstream regions (% of expected occurrence[b])		Relative enrichment in upstream regions	4/5-mer (motif)	Occurrence in genome[a] (% of expected occurrence[b])		Occurrence in upstream regions (% of expected occurrence[b])		Relative enrichment in upstream regions
100 nt upstream						**200 nt upstream**					
taac[j]	1888	(60)	82	(90)	1.5	gtaa	1810	(58)	144	(81)	1.4
cccg	583	(56)	25	(83)	1.5	taag	1402	(45)	109	(61)	1.4
ccgg	374	(36)	16	(53)	1.5	gggt	1333	(91)	103	(124)	1.4
ccgt	940	(64)	40	(94)	1.5	tttt	6450	(160)	491	(216)	1.3
WSSV (5-mer motif)											
ccggg	66	(31)	7	(116)	3.8	<u>tataa</u>	760	(57)	113	(151)	2.6
<u>tataa</u>	760	(57)	73	(195)	3.4	ccccc	292	(137)	41	(342)	2.5
cccgg	66	(31)	6	(99)	3.2	<u>atata</u>	716	(54)	95	(127)	2.4
<u>ataaa</u>	1202	(87)	103	(263)	3.0	<u>ataaa</u>	1202	(87)	147	(188)	2.2
taaaa	1378	(99)	102	(261)	2.6	ccggg	66	(31)	8	(66)	2.1
aaccg	203	(44)	15	(116)	2.6	taaaa	1378	(99)	167	(213)	2.1
gtata	609	(67)	45	(176)	2.6	gtata	609	(67)	73	(143)	2.1
<u>atata</u>	716	(54)	51	(136)	2.5	aaccg	203	(44)	23	(89)	2.0
taccc	331	(75)	22	(178)	2.4	aaaaa	2054	(142)	231	(283)	2.0
acccg	138	(44)	9	(102)	2.3	<u>tatat</u>	716	(56)	79	(110)	2.0
aaaaa	2054	(142)	133	(325)	2.3	acccc	320	(103)	35	(199)	1.9
aacca	644	(96)	41	(216)	2.3	accgg	120	(38)	13	(73)	1.9
ccgta	208	(47)	13	(105)	2.2	cccgg	66	(31)	7	(58)	1.9
gacct	257	(58)	16	(129)	2.2	ctcac[k]	284	(65)	30	(121)	1.9
cgtcg	178	(59)	11	(130)	2.2	taccc	331	(75)	34	(137)	1.8

[a]both strands, excluding *hr*s (present for *Ac*MNPV and WSSV)

[b]expected occurrence is the occurrence of a 4-mer or 5-mer based on random distribution of nucleotides in the complete genome

[c]part of the *Ac*MNPV late initiator sequence (a/g/t)TAAG

[d]part of the *Ac*MNPV upstream activating element with sequence (a/t)GATA(a/t)

[e]part of the *Ac*MNPV downstream activating element with sequence (a/t)CACNG

[f]sequence of the *Ac*MNPV early initiator CAGT

[g]part of the CCATT box

[h]part of the Vaccinia virus late initiator sequence TAAAT and/or the intermediate initiator TAAA(a/t)

[i]sequence of the ASFV late initiator TATA

[j]part of the WSSV putative late TIS motif ATNAC

[k]part of the WSSV putative early initiator (a/c)TCANT

CAGT and (a/g/t)TAAG, for early and late genes, respectively (Blissard and Rohrmann, 1989; Friesen, 1997; Lu and Miller, 1997). Ayres *et al.* (1994) showed that the sequence TAAG occurs less frequently in the whole *Ac*MNPV genome than expected based of the *Ac*MNPV nucleotide composition. The results of the 4-mer motif frequency in the 100 nt upstream of all *Ac*MNPV ORFs analyzed with the enumeration method indeed shows that the TAAG motif frequency is 29% of the expected occurrence in the whole genome (Table 7.1). However, the analysis also shows that this motif has the highest relative enrichment in the upstream regions of the *Ac*MNPV ORFs of all possible 4-mer motifs (4.0 times). Also the baculovirus early promoter motif CAGT is relatively more frequently present in upstream regions (1.4 times), although not as prominent as the TAAG motif. Parts of the TATA box as well as sequences of the well known baculovirus early transcription activating motifs GATA and CACNG (Friesen, 1997) occur relatively often in the upstream regions of the ORFs

(Table 7.1). Compared to 100 nt, the enrichment of the functional motifs in 200 nt upstream of the *Ac*MNPV ORFs is less pronounced (Table 7.1) supporting the experimental observation that in baculoviruses important promoter elements are often located within 100 nt upstream of the translational start codon (Ayres *et al.*, 1994). Analysis of 5-mer motifs of *Ac*MNPV revealed that (a/g/t)TAAG was enriched in the upstream regions of the ORFs, but not CTAAG. Analysis of 6-mer motifs showed a relative enrichment of 3.0 times of the consensus TATA box sequence TATAAA in the 200 nt upstream regions.

For HHV1, the 4-mer nucleotide motifs of known promoter elements were identified by the enumeration method during analysis of the 200 nt upstream sequences, but not when analyzing 100 nt upstream sequences. This supports the view that, in contrast to baculoviruses, most regulatory elements are located more than 100 nt upstream of the HHV1 translational start codons (Rajčáni *et al.*, 2004). Parts of the consensus TATA box, involved in HHV1 early and late transcription (Kim *et al.*, 2002; Rajčáni *et al.*, 2004), occur relatively frequently in the 200 nt upstream of the HHV1 ORFs (Table 7.1). Also the sequence CATT, part of the CCATT boxes which are typically located upstream of the consensus TATA box of HHV1 early genes (Rajčáni *et al.*, 2004), shows a high relative enrichment of 2.1 (Table 7.1).

For both cytoplasmatic viruses Vaccinia virus and ASFV the analysis shows that the late initiator sequences, TAAAT and TATA respectively (Almazán *et al.*, 1992; Broyles, 2003; García-Escudero and Viñuela, 2000), are highly enriched in the 100 nt as well as the 200 nt upstream sequences, although not as prominent as the late TIS of baculoviruses (Table 7.1). Also parts of the sequence TAAA(a/t), essential for Vaccinia virus intermediate gene expression, are enriched (Table 7.1). Furthermore, the analysis shows a considerable enrichment of motifs only consisting of A and T residues. Long stretches of these nucleotides upstream of the transcribed region are typical for Vaccinia virus and ASFV early promoters, as well as for ASFV late promoters (Almazán *et al.*, 1992; Broyles, 2003; García-Escudero and Viňuela, 2000).

WSSV: The same enumeration method was used to analyze the upstream sequences of WSSV ORFs. The analysis of the 4- and 5-mer motifs is shown in Table 7.1. Sequences of the consensus TATA box appear relatively frequently compared to their presence in the complete WSSV genome (Table 7.1). The enrichment of these TATA box sequences is similar to what is observed for *Ac*MNPV and HHV1 (Table 7.1), indicating a functional role for the TATA box in WSSV transcription regulation. Besides the TATA box sequences, the sequence AACC has the highest enrichment in the 100 nt upstream sequences of WSSV ORFs, although not as pronounced as the occurrences of the *Ac*MNPV TAAG motif (Table 7.1). Previous experiments showed that the TISs of the late WSSV envelope protein genes *vp28* and *vp19* start within this exact AACC sequence (Marks *et al.*, 2003) indicating this could be a putative promoter element for late transcription. Furthermore, some motifs consisting of G and C residues, such as the 4-mers CCGG and CCCC and the 5-mers CCGGG and CCGGG

(Table 7.1), and G/C-rich sequences have a relatively high frequency in WSSV upstream regions. Compared to the 100 nt upstream of the ORFs, the results for the analysis of 200 nt are only slightly different and mostly less pronounced (Table 7.1). From the remaining analysis using 6, 7 or 8 nt motifs in the 100 or 200 nt upstream regions (data not shown), it is noteworthy that the enumeration method shows a relative enrichment of the 6-mer consensus TATA box sequence TATAAA of 4.7 times in the 100 nt upstream of the ORFs.

Previously, we showed that the WSSV genes clustered in an early and a late class based on expression profile in shrimp tissue (Marks *et al.*, 2005b). Further analysis within the 100 nt upstream regions of either the WSSV early or late genes using the enumeration method showed that the sequence AACC has the highest relative enrichment of all possible 4-mer motifs for the late genes (2.4 times), while sequences of the TATA box were highly enriched in upstream regions of both gene classes (the sequence TATA showed a relative enrichment of 2.3 times in the 100 nt upstream regions of both gene classes; other data not shown).

MEME

The 100 or 200 nt sequences upstream of all WSSV genes were also studied by MEME (Table 7.2). As multiple classes of coregulated viral genes will be present within these sequences, the MEME settings for this analysis were to identify conserved motifs regardless whether it occurred in the upstream regions of all genes. MEME identified the TATA box as consensus nucleotide motif in these WSSV upstream sequences (Table 7.2). Furthermore this analysis showed that multiple upstream sequences contain stretches of T residues (Table 7.2). Detailed analysis revealed that these are mostly part of the polyadenylation signals (Birnstiel *et al.*, 1985) of the upstream ORFs, and therefore probably not functional as promoter element of WSSV. The outcome of the 100 and 200 nt upstream sequences are very similar, in line with the results of the enumeration method.

Table 7.2. Consensus sequences (4-8 nt) in upstream regions of WSSV genes identified with MEME. Only the best 3 hits of MEME are shown. In case of all WSSV genes, the number of sequences in which the consensus sequence occurred is indicated.

	All WSSV genes (0 or 1 occurrence per individual sequence)	**Early WSSV genes** (1 occurrence per individual sequence)	**Late WSSV genes** (1 occurrence per individual sequence)
100 nt upstream	gtataaaa (TATA box) (61 seqs) tgtttttt (T-rich) (65 seqs) gaggaaga (61 seqs)	gtataaaa (TATA box) tttttttca (T-rich) caacatca	gtataaaa (TATA box) tttgtgta aacc
200 nt upstream	gtataaaa (TATA box) (112 seqs) ttgtttttt (T-rich) (65 seqs) tcctcctc (61 seqs)	agaagagg tatttttt (T-rich) agaaat	gaggaaga tcttttat (T-rich) aaaaatat (A-rich)

For individual analysis of the WSSV early or late kinetic cluster (Marks *et al.*, 2005b) the frequency of a specific motif per individual sequence was set at one, as most WSSV genes belonging to one cluster were considered to be coregulated. Analysis of the 100 and 200 nt upstream regions of either the early or the late class genes identified the consensus TATA box

as putative promoter element (Table 7.2). Previously, we already showed that 37 of the 64 early genes (58%) and 28 of the 58 genes that clustered late (48%) contain a consensus TATA box (Marks *et al.*, 2005b). Specific for the early class, MEME identified the consensus sequences CAACATCA and AGAAT, while for the late class it identified the consensus sequence AACC as well as an A-rich region (Table 7.2). On the other hand, as the early or late kinetic cluster (Marks *et al.*, 2005b) could also consist of subsets of coregulated WSSV genes, an additional MEME analysis was performed in which a motif only had to occur in at least half of the upstream sequences of either the early or the late genes. The outcome was very similar to the results presented in Table 7.2. Interestingly, the TATA box and the AACC motif were identified by the enumeration method as being highly enriched in upstream regions of WSSV ORFs.

Alignments TISs of WSSV genes

To validate both *in silico* methods described above, we compared the outcome with alignments of all known 5' ends of the WSSV early and late class genes. To facilitate comparisons with other viruses, in these alignments the function of the protein encoded by the gene is used to determine its class, either early or late. Early genes often encode enzymes which have functions involved in processes such as nucleotide metabolism, DNA replication, protein modification, viral transcription initiation and host response modulation. Structural virion protein genes often comprise a large part of viral late genes. For nearly all WSSV genes analyzed, this classification matched the results obtained by the microarray study (Marks *et al.*, 2005b).

Early genes: The WSSV genome encodes around 10 genes which, based on their (putative) function, are considered to be early (van Hulten *et al.*, 2001a). For several of these genes the
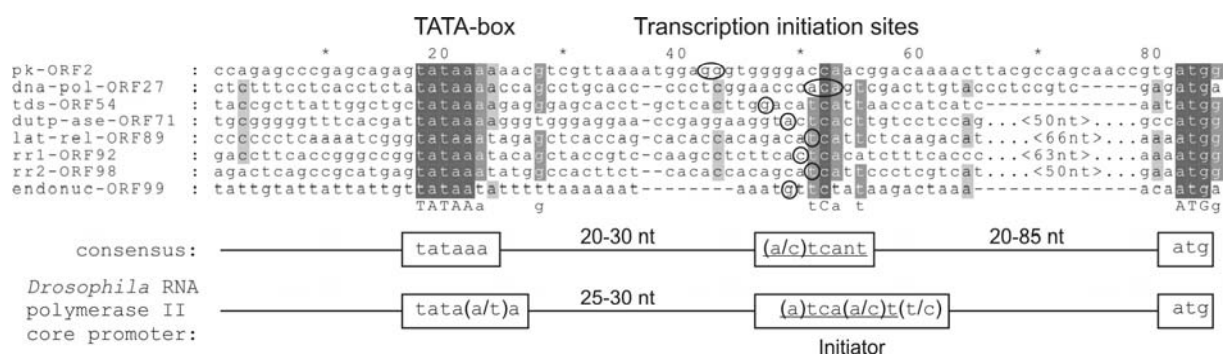
Figure 7.1. Alignment of 5' flanking sequences of WSSV early genes, named after WSSV-TH ORF numbers and the function of their protein product. The transcription initiation site of each gene is encircled. Sequences are aligned by their consensus TATA box, as well as by maximizing the identities around the transcriptional start site. Below, the consensus sequence of the alignment and the *Drosophila* RNA polymerase II core promoter are shown. Similar sequences of the consensus TIS motif and the initiator of the *Drosophila* RNA pol II core promoter are underlined. Abbreviations used and references: *pk*: protein kinase (Liu *et al.*, 2001); *DNA-pol*: DNA polymerase (Chen *et al.*, 2002b); *tds*: Thymidylate Synthase (Li *et al.*, 2004b); *dutp-ase*: dUTPase (Liu and Yang, 2005); *lat-rel*: latency related gene (Hossain *et al.*, 2004); *rr1* and *rr2*: the large and small subunit of ribonucleotide reductase, respectively (Tsai *et al.*, 2000a); *endonuc*: endonuclease (Li *et al.*, 2005a).

5' end of their transcripts has been mapped. RT-PCRs and/or Northern Blots of viral time courses confirmed that these genes were expressed in an early stage during infection (for references see Fig. 7.1). Furthermore ORF89, which is thought to be involved in latency, was empirically shown to be (immediate) early (Hossain *et al.*, 2004; Khadijah *et al.*, 2003). Fig. 7.1 shows an alignment of the experimentally determined transcription initiation sites (TISs) of WSSV early genes. The genes typically contain a consensus TATA box (sequence: TATA(a/t)A; Smale and Kadonaga, 2003). The TIS is located 20 to 30 nucleotides downstream of the consensus TATA box, which is considered to be a functional distance (Arnosti, 2002; Cherbas and Cherbas, 1993; Smale and Kadonaga, 2003). This is between 20 to 85 nucleotides upstream of the translational start codon of the early gene products (Fig. 7.1). When the sequences are aligned by maximizing the identities around the transcriptional start site (Fig. 7.1), a clear consensus transcription initiation motif ((a/c)TCANT) overlapped with the transcriptional start sites. This resembles the RNA polymerase II core promoter motif identified in *Drosophila*, which often consists of a consensus TATA box and/or an initiator with the sequence (A)TCA(+1)(g/t)T(t/c) (Arnosti, 2002; Cherbas and Cherbas, 1993; Hultmark *et al.*, 1986; Smale and Kadonaga, 2003). Similar to WSSV, the initiator of *Drosophila* is typically located 25-30 nt downstream of a TATA box (Arnosti, 2002; Cherbas and Cherbas, 1993; Smale and Kadonaga, 2003). Interestingly, the motif CTCAC, which is part of the identified WSSV consensus sequence (a/c)TCANT and which is the exact sequence of the TISs of the *dutpase* and *rr1* (Fig. 7.1), was also shown to be enriched in upstream regions of WSSV ORFs (Table 7.1).

Late genes: The protein pattern of WSSV particles on an SDS-PAGE gel shows around 8 major WSSV structural virion proteins (Huang *et al.*, 2002b; Tsai *et al.*, 2004; van Hulten *et al.*, 2000a, c). For 6 of these proteins (VP664, VP28, VP26, VP24, VP19 and VP15) the 5' end of the corresponding mRNA has been mapped (Leu *et al.*, 2005; Marks *et al.*, 2003). RT-PCRs and/or Northern Blots of viral time courses confirmed that these genes were expressed in a late stage during infection (Leu *et al.*, 2005; Marks *et al.*, 2003). We completed this analysis by mapping the TISs of the two other major structural protein genes, *vp75* (ORF160) and *vp73* (ORF112). Both *vp75* and *vp73* lack a consensus TATA box (Fig. 7.2a). Using 5'RACE, the TIS of *vp75* was identified within the nucleotide sequence TG, 72 nt upstream of the translational start codon. For *vp73*, the TIS was located at nucleotide residues TC, 220 upstream of the translational start codon (Fig. 7.2a).

When the upstream sequences of all major structural protein genes are aligned by maximizing the identities around the transcriptional start sites (Fig. 7.3), the TISs are present within or very near the nucleotide sequence ATNAC. The transcripts start 20-25 nucleotides downstream of an A/T rich region, which has an average A/T content of 79% compared to 61% of the 200 nt upstream regions of the 8 genes. *Vp15* and *vp19* contain a consensus TATA box, of which only the TATA box of *vp15* is at a functional distance of the TIS (Fig. 7.3;
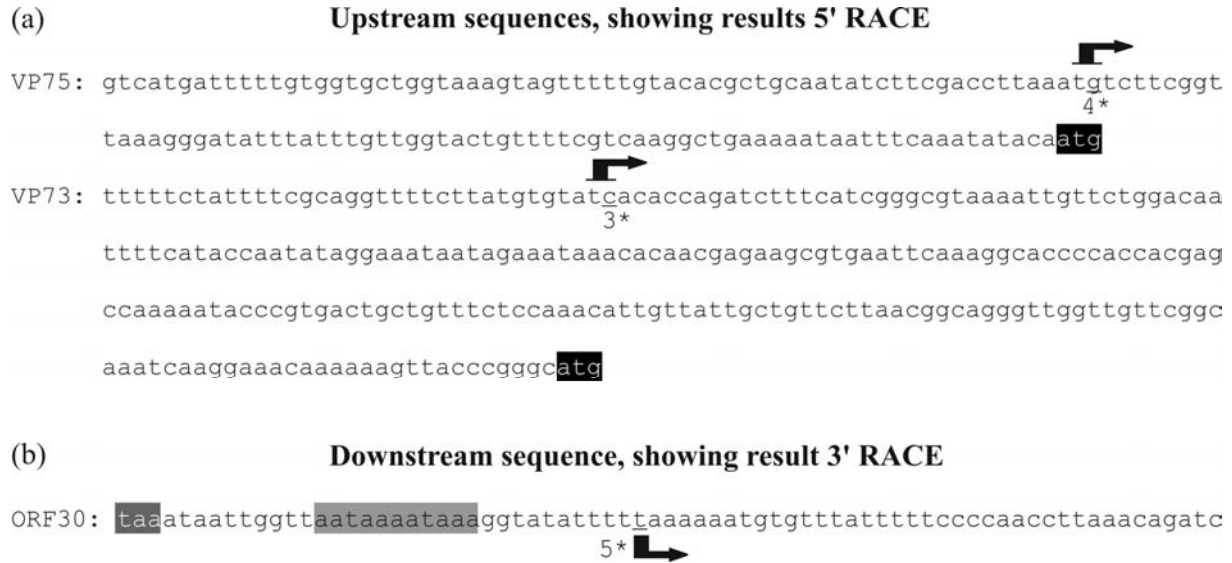
Figure 7.2. Sequences upstream of two major structural protein genes *vp75* (ORF160) en *vp73* (ORF112) of WSSV showing the transcription initiation sites of both genes, indicated by arrows above the sequences. The 5' termini of the different clones sequenced for each gene are underlined. The number beneath the underlining shows the number of similar clones. The start codons of both genes are shaded black **(a)**. Sequence downstream of ORF30 showing the polyadenylation site, indicated by an arrow below the sequence. The 3' terminus of the different clones sequenced is underlined. The number before the arrow represents the number of similar clones sequenced. The stop codon is shaded dark grey and two overlapping poly(A) signals (AATAAA) light grey **(b)**.
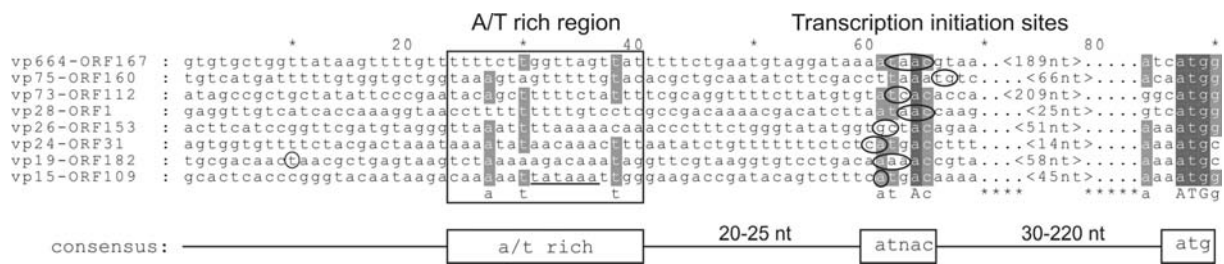
Figure 7.3. Alignment of 5' flanking sequences of WSSV late genes. Name of structural protein genes as well as WSSV-TH ORF numbers are indicated. The transcription initiation site of each gene is encircled. For *vp19* a minor transcription initiation site is also encircled. The TATA box for *vp15* is underlined. The A/T rich region is boxed. Sequences are aligned by maximizing the identities around the transcriptional start site. References: *vp28*, *vp26*, *vp24*, *vp19* and *vp15* (Marks *et al.*, 2003); *vp664* (Leu *et al.*, 2005); *vp75* and *vp73* (this study).

Marks *et al.*, 2003). The length of the TIS to the translational start codon is different for the various genes, ranging from 30 to 220 nt (Fig. 7.3). Interestingly, most of these features were predicted by our *in silico* analysis. The first three nucleotides of the AACC motif identified in the *in silico* analysis (Tables 7.1 and 7.2) are part of the consensus sequence ATNAC, and both contain the AC dinucleotide which is present for almost all genes in Fig. 7.3. Also the sequences ATAA and TAAC, parts of the ATNAC sequence, were identified as putative promoter elements (Table 7.1). The A-rich (and T-rich) sequences identified by MEME are in line with the observation that late genes often contain long stretches of A/T residues upstream of their TIS (Fig. 7.3).

In addition to the 8 major structural proteins, the protein profile of WSSV particles shows a range of about 40 minor virion proteins (Huang *et al.*, 2002b; Tsai *et al.*, 2004). Most of these have not been studied in detail. However, the corresponding messengers are supposed to be late, although 13 of them clustered in the early class during microarray analysis (Marks *et al.*, 2005b). Remarkably, 45% of the minor virion protein genes (18 of the 40) contain a consensus TATA box within 300 nt of the translational start codon. This is in line with the MEME analysis, which also suggested that the TATA box might be involved in late transcription.

Polyadenylation

For various WSSV genes, the site of polyadenylation has been mapped using 3'RACE. We extended this analysis by mapping the polyadenylation site of ORF30, the collagen-like ORF of WSSV (Li *et al.*, 2004a). Polyadenylation of ORF30 starts 32 nt after the translational stop codon, 16 nt after the first poly(A)-signal (sequence AATAAA; Fitzgerald and Shenk, 1981) (Fig. 7.2b).
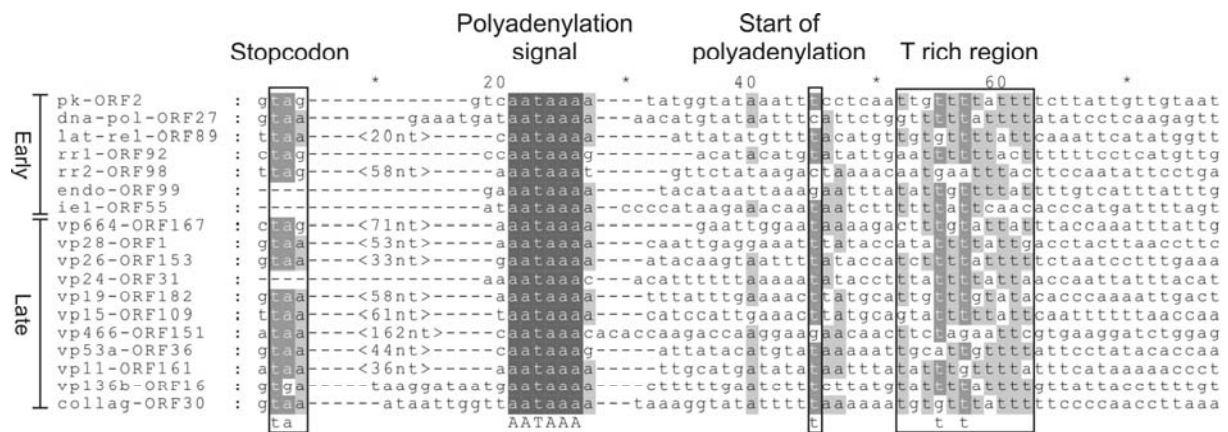
Figure 7.4. Alignment of 3' flanking sequences of WSSV genes. The stop codon, polyadenlylation signal, start of polyadenylation and the T rich region are indicated. Sequences are aligned by stop codon and by polyadenylation signal. Abbreviations used and references: *ie1*: immediate-early 1 (Liu *et al.*, 2005); *vp466* (Huang *et al.*, 2002b); *vp53a*, *vp11*, *vp136b* (Tsai *et al.*, 2004); *collag*: collagen-like ORF (this study). For abbreviations and references of other genes see Figs. 7.1 & 7.3.

Fig. 7.4 shows an alignment of all known polyadenylation sites of WSSV. Polyadenylation typically starts within 11-19 nt after a consensus polyadenylation site. Typically, a T rich region (stretch of about twelve T residues) was identified 8 nt downstream of the poly(A)-site (Fig. 7.4). There seems to be no difference between the polyadenylation sites of early and late genes (Fig. 7.4). A total of 9 WSSV genes were found to be non-polyadenylated (Liu *et al.*, 2005; Tsai *et al.*, 2004). Except for *vp12a* (WSSV-TH ORF34), all these genes lack a consensus poly(A)-signal within -50 to 300 nt of their translational stop codon. Two (*vp31* and *vp13b* encoded by WSSV-TH ORF163 and ORF155, respectively) do however contain the sequence ATTAAA within this region, which in vertebrates is often sufficient for polyadenylation (Sheets *et al.*, 1990), but apparently not in invertebrates or arthropods.

Discussion

In this paper, we used a new enumeration strategy based on a model proposed by Brazma *et al.* (1998) to identify putative WSSV promoter elements. A set of computer scripts was designed, which calculated the difference in nucleotide motif frequencies in the upstream sequences of all genes compared to the complete WSSV genomic sequence. The rationale behind this analysis is that promoter motifs are often thought to be transcription factor binding sites, which are functional upstream of genes. The results obtained with the well studied large ds DNA viruses *Ac*MNPV, HHV1, Vaccinia virus and ASFV (Table 7.1) show that our method is robust in identifying important promoter elements of completely sequenced viral genomes without *a priori* knowledge, as these are often enriched in upstream sequences of viral ORFs. Therefore, this new enumeration method can be useful in the analysis of newly sequenced genomes of large ds DNA viruses. For further analysis of the upstream regions of WSSV genes of the early and late cluster, as assigned by microarray analysis (Marks *et al.*, 2005b), MEME was used. Genes of either cluster might be coregulated by similar mechanisms, utilizing conserved nucleotide motifs. As MEME can identify motifs which have to occur in each individual sequence of a set of submitted sequences, or in a selected number of submitted sequences, it is highly complimentary to the enumeration method. Another advantage of MEME is that it can identify degenerate motifs.

The enumeration method identified various nucleotide motifs (Table 7.1) that were also identified by MEME (Table 7.2) and by the alignments of experimentally determined 5' ends of WSSV mRNAs (Figs. 7.1 and 7.3). These include the consensus TATA box, as well as the nucleotide motif AACC. However, also other nucleotide motifs that were not validated with the other methods, e.g. some motifs rich in C or G residues, were (highly) enriched in WSSV upstream regions and might be involved in WSSV transcription. In accordance with the alignments shown in Figs. 7.1 and 7.3, where most putative promoter elements are located within 100 nt upstream of the ORFs, the nucleotide motifs identified with the enumeration method are most pronounced in the 100 nt upstream of the ORFs (Table 7.1) compared to 200 nt. This suggests that, similar to *Ac*MNPV, most WSSV promoter elements are located within 100 nt upstream of the translational start codons, which is a reflection of the tight package of genes along the WSSV genome. It would be of interest to test the functionality of the sequences (a/c)TCANT and ATNAC, which were identified as the consensus TISs of the WSSV early and late class genes, respectively (Fig. 7.3) and other identified motifs (Table 7.1) in a reporter gene (e.g. luciferase) assay. For testing late promoters in this setup, a co-infection with WSSV should be considered to supply additional viral transcription factors required for late gene expression. In the absence of a suitable WSSV cell system, these reporter gene assays have been performed in the artificial Sf9 insect cell line (Hossain *et al.*, 2004; Liu *et al.*, 2005; Lu *et al.*, 2005) with all its limitations to the interpretation of the results. However, with the recent developments concerning differentiation and growth of

crayfish hematopoietic stem cells *in vitro* (Söderhäll *et al.*, 2005), these experiments might be performed in crayfish cell cultures providing a more convenient and homologous system.

The identification of (putative) promoter elements provides further insight in the transcription mechanisms used by WSSV. The presence of a consensus TATA box for most early genes as well as a conserved transcription initiation motif similar to the *Drosophila* initiator suggest that WSSV uses the host RNA polymerase II transcription machinery for generating early transcripts, as also proposed by Chen *et al.* (2002b) and Liu *et al.* (2005). Previous analysis of WSSV late genes could not reveal any readily apparent dominant nucleotide element used for WSSV late gene expression (Marks *et al.*, 2003). Using the newly available microarray clustering (Marks *et al.*, 2005b), we could now show that around half of the WSSV putative late genes contain a consensus TATA box. This suggests that WSSV might exploit the cellular RNA polymerase II system not only for early but also for (part of) its late mRNA synthesis, similar to some other ds DNA viruses like herpesviruses (Wagner *et al.*, 1995). Only one of the 8 major structural virion protein genes, which are expressed in the late phase of viral infection and most likely are co-regulated to secure correct assembly of the virion, contains a consensus TATA box. Alignment of the 5' ends of the 8 major structural protein genes identified a novel consensus transcription initiation site, ATNAC, downstream of an A/T rich region. The *in silico* analysis further supports the observation that both components might be late promoter elements. This suggests a second pathway for WSSV late gene expression, similar to the late gene expression strategy identified for baculoviruses (Lu and Miller, 1997; Mistretta and Guarino, 2005). However, different from baculoviruses, viral genes required for this pathway, such as a RNA polymerase or late transcription factors, have not been identified on the WSSV genome (van Hulten *et al.*, 2001a; Yang *et al.*, 2001). These genes could however be too much diverged from known homologues to be found based on amino acid homology.

The alignments of the 3' ends of WSSV mRNAs suggest that there is no difference in polyadenylation between early and late mRNAs. The WSSV polyadenylation characteristics of both classes resemble regular polyadenylation in eukaryotic mRNAs, which is typically located 10 to 25 nt downstream of the sequence AATAAA (Fitzgerald and Shenk., 1981; Sheets *et al.*, 1990). Also oligo-T stretches are often present about 30 nt downstream of the poly(A)-signal of eukaryotic genes (Birnstiel *et al.*, 1985). These data indicate that WSSV uses the regular cellular enzymes for polyadenylation of mRNAs. However, other undefined signal pathways of polyadenylation might also be used, as two WSSV genes (dUTPase and TdS) were found to be polyadenylated without a poly(A)-signal present (Li *et al.*, 2004b; Liu and Yang, 2005).

Using a combined approach of *in silico* analysis and experimentally determined data on WSSV transcriptomics, further support was found for the presence of different coregulated classes of WSSV genes. Comparisons with other large ds DNA viruses provided insight in the transcription mechanism of these classes and putative promoter motifs involved. In order to

determine the functionality of these motifs empirically cell culture systems for shrimp will have to be further developed.

Materials and Methods

Virus infection

The virus isolate used in this study, known as WSSV-TH (acc.no. AF369029), originates from infected *Penaeus monodon* shrimp obtained in Thailand in 1996 and was treated as described before (van Hulten *et al.*, 2000c). Crayfish *Orconectes limosus* was injected intramuscularly with purified WSSV using a 26-gauge needle to initiate infection. Three days post infection (d.p.i.), the crayfish were frozen in liquid nitrogen and stored at -80ºC until further use.

5' and 3' Rapid Amplification of cDNA Ends (5'/3' RACE)

Both 5' and 3' RACE were carried out using a commercial 5'/3' RACE kit (Roche) following the manufacturer's instructions. Total RNA was isolated from the frozen gill tissue of three infected crayfish *O. limosus* (harvested 3 d.p.i.) as described before (Marks *et al.*, 2003). In case of the 3' RACE of ORF30, first strand cDNA was synthesized using the oligo(dT) anchor primer. The resulting cDNA was amplified using one specific forward primer (ORF30-RACE-F1: CAGACCCGATTACAGTAGCAG; WSSV-TH location: 48983-49003) and the anchor primer. For the 5' RACE of ORF112 and ORF160, the RACE-R1 primers mentioned below were used for synthesis of the cDNA. This cDNA was purified using the High Pure PCR Product Purification Kit (Roche) and a homopolymeric 3' d(A)-tail was added to the cDNA in a mixture with a total volume of 20 μl, using terminal transferase and dATPs included in the kit. This mixture (5 μl) was used in a PCR, performed with an oligo(dT) anchor primer and a nested RACE-R2 primer (see below). The final products of the 5' and 3' RACE were cloned into the pGEM-T easy vector (Promega) and sequenced. For each 5' and 3' RACE at least 3 clones were sequenced. Primers used for 5' RACE: ORF112-RACE-R1: CGCATATTGTTGTTTGTCGTAG (WSSV-TH location 168230-168209); ORF112-RACE-R2: GACGCGTATCTCAAGTATTCC (WSSV-TH location 168184-168164); ORF160-RACE-R1: CTTGTTGGATTCGGAGCAGTG (WSSV-TH location 240137-240117); ORF160-RACE-R2: GACGGATAATATGGGTGACAAG (WSSV-TH location 240111-240090).

DNA sequencing and computer analysis

Plasmid clones carrying RACE products were sequenced at the company BaseClear (the Netherlands), using universal M13 forward and reverse primers. Sequence data were analyzed using the software package DNASTAR4.2. All sequences data were edited and aligned in GeneDoc, version 2.6.000 (Nicholas *et al.*, 1997).

Promoter analysis using the enumeration method

A set of computer scripts, made in the computer programming language Perl (more information at www.perl.com), was designed to analyze all 4, 5, 6, 7 or 8 nucleotide motif frequencies in the 100 and 200 nt upstream sequences of (putative) ORFs to compare these with a complete viral genome. This method will be referred to as the enumeration method in the remaining part of the article. In case of WSSV, the genome of WSSV-TH as annotated by van Hulten *et al.* (2001a) was used for the analysis. For WSSV and *Autographa californica* Multinucleopolyhedrovirus (*Ac*MNPV), the genes and upstream regions of genes that are partly or completely located in the homologous regions (*hr*s; for WSSV this concerns 24 genes, for *Ac*MNPV 3 genes) and the *hr*s itself were excluded from the analysis to avoid the possibility of finding false motifs due to the high homology of the *hr* sequences. As nucleotide motifs with a regulatory function are expected to be present in multiple upstream regions, only motifs that were present in upstream sequences of at least 5 ORFs were analyzed after running the scripts. Scripts were used for the following procedures: (1) extraction of the upstream region before the selected ORFs, both on the + and − strand; (2) calculation of the X-mer nucleotide motif frequencies in the upstream regions; (3) calculation of the X-mer nucleotide motif occurrences in both strands of the viral genome (without *hr*s); (4) calculation of the relative ratio of the same X-mers between the upstream regions and the viral genome; (5) ranking of the relative ratios for each X-mer from high to low; (6) exclusion of motifs present in less than 5 upstream regions. Accession numbers of genomes analyzed were NC_001623 (*Ac*MNPV); NC_001806 (Human Herpes Virus 1: HHV1); NC_001559 (Vaccinia virus); NC_001659 (African Swine Fever Virus: ASFV) and AF369029 (WSSV).

MEME

The computer program Motif Elucidation using Maximum Expectation maximization (MEME; available at http://meme.sdsc.edu/meme/website/intro.html) was used to search for specific sequence motifs in 100 or 200 nt of the WSSV noncoding sequences upstream of the translational initiation codon. MEME analysis was performed using the sequence of WSSV-TH annotated by van Hulten *et al.* (2001a). A search was performed for the 3 best motifs with a length of 4-8 nt. In case of analyzing the upstream sequences of the WSSV ORFs at large, the occurrence of a specific motif per individual sequence was set at zero to one, but the motif had to occur in at least 60 upstream regions. In case of analyzing upstream sequences of WSSV genes belonging to the early or late kinetic clusters, the frequency of a specific motif per individual sequence was set at one. The 5' noncoding regions were categorized according to class of expression (Marks *et al.*, 2005b): Early (ORF2, ORF8, ORF9, ORF11, ORF12, ORF15, ORF16, ORF23, ORF24, ORF25, ORF29, ORF37, ORF49, ORF53, ORF55, ORF56, ORF58, ORF60, ORF61, ORF66, ORF67, ORF69, ORF70, ORF74, ORF81, ORF85, ORF89, ORF91, ORF92, ORF93, ORF98, ORF99, ORF101, ORF103, ORF107, ORF111, ORF112, ORF115, ORF116, ORF117, ORF125, ORF126, ORF127, ORF131, ORF132, ORF142,

ORF145, ORF146, ORF147, ORF152, ORF156, ORF159, ORF160, ORF161, ORF164, ORF165, ORF169, ORF170, ORF171, ORF172, ORF173, ORF177, ORF178, ORF179) and Late (ORF1, ORF3, ORF4, ORF7, ORF10, ORF14, ORF27, ORF28, ORF30, ORF31, ORF32, ORF33, ORF34, ORF35, ORF36, ORF38, ORF39, ORF41, ORF43, ORF44, ORF54, ORF57, ORF65, ORF72, ORF73, ORF75, ORF76, ORF77, ORF79, ORF80, ORF84, ORF90, ORF94, ORF95, ORF100, ORF109, ORF113, ORF114, ORF118, ORF119, ORF120, ORF121, ORF128, ORF129, ORF130, ORF134, ORF135, ORF136, ORF143, ORF148, ORF151, ORF153, ORF157, ORF167, ORF168, ORF182, ORF183, ORF184).

Chapter 8

General Discussion

White spot syndrome virus (WSSV), a large enveloped double-stranded (ds) DNA virus, is one of the first viruses infecting aquatic invertebrates that has been characterized in detail. Analysis of the gene content and phylogeny showed that WSSV differed profoundly from other large ds DNA viruses and belonged to a new virus family, called *Nimaviridae* (Vlak *et al.*, 2005). WSSV causes a devastating disease in cultivated shrimp around the world, with losses caused by the virus estimated at more than 1 billion euros annually. The research described in this thesis has increased the fundamental insights in the genomic structure and gene regulation (transcriptomics) of WSSV. Furthermore, the data presented has provided a deeper insight in WSSV epidemiology and the molecular basis of WSSV biology. In this chapter the genetic heterogeneity and differences in virulence among WSSV isolates, as observed in chapters 2, 3 and 4, are discussed. Furthermore, the pitfalls of the use of microarrays in viral gene expression profiling (chapter 5) are discussed. The classes of WSSV genes, obtained by the microarray analysis and the promoter motifs identified in chapters 6 and 7, are discussed in the context of transcription regulation of other eukaryotic large ds DNA viruses. Finally, we describe how the results obtained in this thesis may help in the design of novel WSSV intervention strategies.

WSSV virulence and epidemiology

Genetic basis for differences in WSSV virulence
Preliminary studies by Lan *et al.* (2002) and Wang *et al.* (1999a) suggested differences in virulence between various WSSV isolates. To further study differences in virulence and fitness, two WSSV isolates (TH96-II and WSSV-TH) were compared (chapter 4). TH-96-II is a newly characterized archival WSSV isolate from 1996 with the largest genome size (~312 kb) reported at present. WSSV-TH contains the smallest WSSV genome known (~293 kb), mainly due to deletions in the variable region (VR) ORF 14/15 and VR ORF23/24 loci (chapters 2, 3 and 4). Comparison between TH-96-II and WSSV-TH, when administered to shrimp *Penaeus monodon,* showed a higher virulence and competitive fitness for the latter. The molecular basis for this difference has remained unclear. The deletions in VR ORF14/15 and/or VR ORF23/24 in isolate WSSV-TH may contribute to the increased virulence. This could either be caused (i) by the fact that (one of the) TH-96-II genes encoded by VR ORF14/15 or VR ORF23/24, and absent from the WSSV-TH genome, are responsible for attenuation of TH-96-II in *P. monodon* or (ii) by the ~19 kb smaller genome size of WSSV-TH, enabling faster replication to reach a lethal viral dose. Another possibility is that one of

the three proteins encoded by ORF75, ORF94 or ORF125, which are polymorphic between WSSV-TH and TH-96-II (chapter 4), is responsible for the differences in virulence. These hypotheses could be tested by creating WSSV mutants or knock-outs on the proposed loci using site-directed mutagenesis. However, in the absence of the required tools such as a cell culture system and a WSSV BAC clone (a plasmid containing the complete WSSV genome), in which the WSSV genome can easily be manipulated, such studies are not feasible.

The VN isolates identified in chapter 3 may prove useful, as these isolates form a pool of various WSSV genotypes for which both VR ORF14/15 and VR ORF23/24, as well as ORF75, ORF94 and ORF125, are characterized. By determining the virulence and fitness of these VN isolates, the contribution of either of the polymorphic loci might be established. However, the contribution of other genetic polymorphisms to differences in virulence, e.g. one of the WSSV SNPs identified in chapter 2 that may result in amino acid changes in virulence-related proteins, cannot be excluded as long as the genomes of these VN isolates are only partly characterized.

WSSV epidemiology

The comparison of the (complete) genome sequences of widely separated geographical isolates of WSSV showed surprisingly little sequence variation (chapters 2 and 3; Lo *et al.*, 1999; Wang *et al.*, 2000b). This high uniformity in DNA sequences suggests that a single virus isolate (genotype) is responsible for the worldwide outbreaks caused by WSSV. Based on the increasing deletions in VR ORF14/15 and VR ORF23/24 (chapter 3) the results suggest that the spread of WSSV originated from one common ancestor most likely from the coastal areas of Southeast-China. Furthermore, the experiments in chapter 4 seem to suggest that WSSV became more virulent during this epidemic, as isolate WSSV-TH showed an increased virulence compared to the ancestral isolate TH-96-II. Interestingly, a similar scenario was predicted for the shrimp RNA virus Taura Syndrome Virus (TSV; Lotz and Wang, 2004). For various other viruses, including DNA viruses, it has been shown that the virulence increased after serial passaging (Ebert, 1998; Ebert and Bull, 2003). As after the emergence of WSSV in 1991 there probably has been a continuous presence of WSSV-susceptible shrimp in the ever growing shrimp culture, this could be one of the explanations for the increase in virulence during the spread of the virus, not only for WSSV but also for TSV.

However, some comments could be made on the hypotheses mentioned above. The WSSV isolates from Vietnam were collected in 2003/2004, while the three other isolates which were used for the model, WSSV-TW, WSSV-CN and WSSV-TH, were obtained in 1994/1996. To be compatible with the model in chapter 3, this would mean that after the spread of WSSV through Asia from 1991 to 1995, only limited further spread of WSSV took place in Asia. Furthermore, based on the model presented in chapter 3, the ancestral isolate

TH-96-II was expected to be collected at the coast of Southeast-China, but was instead obtained from a shrimp farm in Thailand.

Taken this into account, an alternative explanation could be that WSSV has been introduced in the shrimp farms independently by infected wild shrimp (or another WSSV host) at multiple places, as also suggested by Wongteerasupaya *et al.* (2003). The collection of two genotypic WSSV variants with differences in virulence, as shown in chapter 4, could then be explained by the fact that these WSSV variants were already present in wild shrimp before the emergence of WSSV in the shrimp farming industry in 1991.

Clearly, more information on WSSV epidemiology and the molecular basis for the differences in virulence between WSSV isolates is necessary to favor one model over the other. The genotypes of archival WSSV isolates collected in Taiwan, Japan, the Philippines and further along the coast of Southeast-Asia and analyzed as described in chapter 3 could provide more insight in the early spread of WSSV and to confirm and further detail the model on WSSV spread as shown in chapter 3.

Transcription regulation

Use of DNA microarrays in WSSV gene expression profiling

Microarrays represent a relatively new research tool, which has become common for gene expression profiling. The main advantages of this approach as compared to more conventional techniques such as Northern blot, RT-PCR and Dot blot assays are (i) the large number of genes that can be mapped simultaneously, enabling the identification of various classes of coregulated genes and (ii) the hybridization of a mix of differently labeled samples to the same microarray slide using various dyes for labeling (Murphy, 2002). The latter allows a robust normalization between the individual expression profiles of a batch of samples. However, there are also disadvantages connected to the use of microarrays. These include (i) their limited sensitivity compared to that of other techniques such as RT-PCR (Murphy, 2002) and (ii) the inability to examine expression patterns of alternatively or multiply spliced messengers, or in case of transcriptional read-through (Paulose-Murphy *et al.*, 2001). Though the latter can be a problem for viruses, the occurrence of such messengers has not been observed for WSSV thus far and they have not been detected at the genome sequence level (van Hulten *et al.*, 2001a).

DNA microarrays have been successfully used to study the regulation of viral gene expression of f.e. herpesviruses (Ebrahimi *et al.*, 2003; Paulose-Murphy *et al.*, 2001; Stingley *et al.*, 2000), baculoviruses (Iwanaga *et al.*, 2004) and bacteriophages (Luke *et al.*, 2002). In this thesis a WSSV microarray was generated to study WSSV gene expression, to bypass extensive conventional studies on WSSV gene expression. The expression profiles of the WSSV genes were studied simultaneously using the generated WSSV microarray (chapter 5). Below, the results of chapter 5 are compared to the data presented in various other recent

publications where DNA microarrays were used to study (WSSV) gene expression (Liu *et al.*, 2005; Tsai *et al.*, 2004).

Two complications have to be considered when microarrays are used for profiling viral gene expression over time. A first complication is the enormous variation in the amount of viral mRNAs present in the host cells over the course of infection. At the onset of infection, no viral mRNA is present, while later (some of the) viral mRNAs accumulate to high amounts. This raises the question of what is best to be compared on the DNA microarray slides to analyze a time course for viral gene expression. In various publications describing microarray-based gene expression profiling of WSSV (Liu *et al.*, 2005; Tsai *et al.*, 2004) or other viruses (Paulose-Murphy *et al.*, 2001; Stingley *et al.*, 2000), each of the infected samples (labeled Cy3) is compared to a non-infected sample (reference sample labeled Cy5). The major drawback of this setup is that the signals for the non-infected samples, present on each microarray slide, are very low (generally near the background threshold) and therefore unreliable. This problem was avoided in the array used in this thesis by comparing each of the infected samples to a standard control sample, which contained a pool of transcripts representing every gene present in the time course (see also Luke *et al.*, 2002).

A second complication of viral gene expression profiling using microarrays is the normalization between the individual samples of a time course. Often, normalization is performed using genes which are constitutively expressed ("housekeeping genes") as described for WSSV by Liu *et al.* (2005) and Tsai *et al.* (2004). However, the influence of viral infection on the expression of individual host genes is beforehand unknown. Furthermore, it becomes more evident that genes which were generally considered as housekeeping genes are often differentially expressed (up- or downregulated, host shut-off) under variable conditions (Khimani *et al.*, 2005), e.g. after viral infection. This could be the reason for the fluctuation in the WSSV gene expression profiles obtained by Tsai *et al.* (2004). The use of multiple housekeeping genes instead of one (actin) might also have resulted in more reliable expression profiles. Also normalization based on the complete transcriptomic data obtained from the microarray, the so-called total intensity normalization (Quackenbush, 2002), is not possible due to the increasing amount of viral mRNA during infection over time. Therefore, normalization during the microarray experiments, described in chapter 5, was performed by starting with the same amount of total mRNA (both host and viral mRNAs) for each of the individual samples of the time course. This method is most reliable when the total amount of mRNAs present in the host cells remains similar over time. Although this is unknown for virus infected cells, the expression profiles obtained (see chapter 5: Figs 5.1b, 5.2b & 5.2c) suggest that this normalization method is sufficiently robust for use in WSSV gene expression profiling. Altogether microarrays are a useful tool for WSSV gene expression profiling. However, these studies require an adequate setup, as the interpretation of microarray results is highly dependent on the design of the experiment.

Kinetic classes of WSSV gene expression

Most large ds DNA viruses of eukaryotes, including baculoviruses and herpesviruses, regulate their gene expression in a cascaded fashion (Broyles, 2003; Friesen, 1997; García-Escudero and Viňuela, 2000; Rajčáni *et al.*, 2004). Early genes are defined as genes expressed prior to the onset of viral replication, while the so-called late genes are expressed after viral replication. Both baculo- and herpesvirus families encode immediate-early (IE) genes, which are expressed before protein synthesis (Friesen, 1997; Rajčáni *et al.*, 2004). IE genes often encode important transregulators of viral gene expression and replication. Baculoviruses also encode very late genes, which are expressed when occluded baculoviruses are being formed (Lu and Miller, 1997). Thus for these viruses 3 to 4 different classes of temporarily regulated genes can be distinguished: IE, early, late and very late. For both virus families, the expression profile of a gene in a virus infected cell-line after either blocking protein synthesis (mostly performed with cycloheximide) or replication (with inhibitors such as aphidicolin or phosphono-acetic acid (PAA)) determines its expression class (Broyles, 2003; Friesen, 1997; García-Escudero and Viňuela, 2000; Rajčáni *et al.*, 2004).

The WSSV microarray studies performed to date (chapter 5; Liu *et al.*, 2005; Tsai *et al.*, 2004) have provided some insight into the temporal expression profiles of WSSV genes and indicate the presence of coregulated gene clusters. Viral IE genes are expressed at the earliest stage of infection and are independent of viral protein synthesis (Friesen, 1997). Using the established techniques, including *in vivo* gene expression profiling, it proved very difficult to identify WSSV IE genes, if any (chapter 5). This problem is mainly caused by the asynchrony of infection of the shrimp gill cells *in vivo*, resulting in expression of IE genes at different time points post infection. Moreover, it is difficult to administer *in vivo* protein synthesis inhibitors such as cycloheximide, which is regularly used to bulk up IE genes. Nevertheless an attempt was made by Liu *et al.* (2005), who administered cycloheximide to shrimp before WSSV infection. Using microarray analysis, no less than 60 possible WSSV IE genes were identified, while baculoviruses and herpesviruses possess only around 5 of such genes (Possee and Rohrmann, 1997; Stingley *et al.*, 2000). By further screening using RT-PCR, followed by promoter activity tests with the upstream regions of the possible IE genes in a reporter gene assay, Liu *et al.* (2005) ended up with one candidate IE gene (WSSV-TH ORF55). Two other genes were also assigned IE, although their upstream region showed no promoter activity in the reporter gene assay (ORF242 and ORF418 of WSSV-TW).

Taking the limitation of the use of inhibitors and the asynchronous infection into account, we have tried to differentiate WSSV gene classes later in infection using mRNA samples of the gills of shrimp as a source (chapter 5). Clustering of the individual gene expression profiles showed the presence of two major classes, which were designated "early" and "late" based on the time point at which maximum gene expression was reached. Although the expression profiles of WSSV transcription resemble the corresponding kinetic classes as shown for other viruses (chapter 5), the question remains to what extend this kind of

classification would match a classification based on the use of inhibitors of replication and protein synthesis. However, the presence of most of the predicted early genes (genes involved in nucleotide metabolism, DNA replication and protein modification) in the early class and predicted late genes (virion/structural protein genes) in the other class (late) suggest a similar outcome.

Tsai *et al.* (2004) further focused on the "late" genes and differentiated this class into four subclusters. Two of these subclusters, which together comprise the major part of the genes they analyzed (Cluster C and D in the original paper), can be interpreted as "late" and "very late". Genes of the late subcluster correlated with proteins located in the nucleocapsid, while the proteins encoded by the "very late" genes were often associated with the WSSV envelope (Tsai *et al.*, 2004). These results suggest that similar to other large ds DNA viruses, including baculoviruses and herpesviruses (Friesen, 1997; Rajčáni *et al.*, 2004), (at least) four different kinetic classes of WSSV genes are present.

Model for WSSV gene expression

Based on the data generated by the WSSV microarray (chapter 5) putative WSSV promoter motifs were identified in chapter 7, both *in silico* as well as by the alignment of 5' ends of WSSV mRNAs. The collective data was used to predict some of the mechanisms by which the WSSV genes are expressed. Combined with the results obtained by Liu *et al.* (2005) and Tsai *et al.* (2004), a simple model for WSSV transcription is presented in Fig. 8.1.

The WSSV IE genes are expressed by a yet unknown mechanism immediately after primary infection. One of the IE genes (WSSV-TH ORF55) might be expressed by the cellular RNA polymerase II system, as suggested by Liu *et al.* (2005). Similar to other large ds DNA viruses, at least one of the IE proteins (encoded by WSSV-TH ORF55) may act as a transcription factor for later viral gene expression, as it contains a Cys2/His2-type zinc finger motif. The presence of a consensus TATA-box as well as a conserved transcription initiation motif, (a/c)TCANT, similar to the *Drosophila* initiator (Arnosti, 2002; Cherbas and Cherbas, 1993; Hultmark *et al.*, 1986; Smale and Kadonaga, 2003) upstream of the ORF suggests that WSSV probably uses the RNA polymerase II core promoter for expression of its early genes.

The mechanism used for late gene expression is less clear. Two separate groups of late WSSV genes may exist, one generating messengers by a new virus-induced transcription mechanism and the other exploiting the cellular RNA polymerase II system for mRNA synthesis (chapter 7). As the latter has not been studied in detail, it is not known whether a consensus transcription initiation site is present among these genes. Some of the other motifs identified in chapter 7 (Tables 7.1 & 7.2) may also play a role in early and/or late gene expression, e.g. by binding of transcription factors. Furthermore, Tsai *et al.* (2004) suggested the presence of a "very late" cluster. Detailed promoter analysis of the genes present in the "late" and the "very late" cluster revealed that these two classes do not coincide with the two separate mechanisms of late gene expression shown in Fig. 5.1. About half of the genes present in either the "late"

or the "very late" cluster contains a consensus TATA box sequence (Tsai *et al.*, 2004), and therefore the regulation of the WSSV "very late" genes might be similar to the WSSV "late" genes.
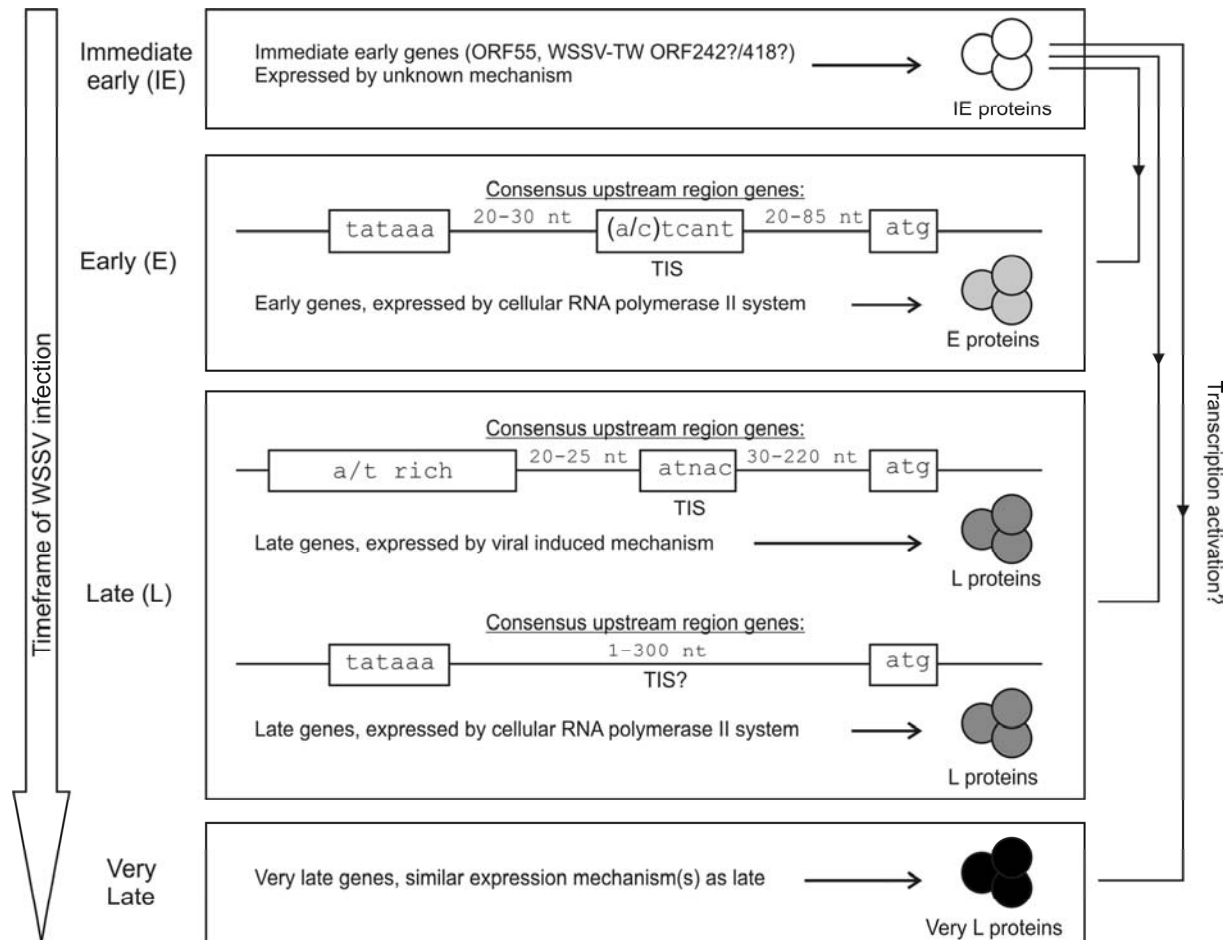
Figure 8.1. Model on WSSV gene expression, as extracted from chapters 5, 6 and 7 and the publications by Liu *et al.* (2005) and Tsai *et al.* (2004). For other motifs that might be involved in early and/or late gene expression: see chapter 7, Tables 1 & 2. TIS: Transcription initiation site; for further explanation: see text.

The confirmation and refinement of this model await the development of a cell line permissive to WSSV. With the recent developments concerning differentiation and growth of crayfish hematopoietic stem cells *in vitro* (Söderhäll *et al.*, 2005), WSSV permissive crayfish cell cultures might be available in the near future. These cells can be used to perform a synchronized WSSV infection, resulting in a more precise time schedule of WSSV gene expression. Furthermore, this will allow the use of drugs as well as the possibility to perform promoter studies, and therefore is an excellent tool to further study WSSV transcription regulation.

Future Applications: WSSV vaccines or intervention strategies

WSSV has become a global problem in aquaculture and its economic impact is enormous (Rosenberry, 2004). Therefore, measures to control this disease are urgently needed. The last

couple of years several observations suggest that shrimp may possess a specific and adaptive immune system, though different from the immune system found in vertebrates (Kurtz, 2005). Recent experiments show a higher survival of WSSV challenged shrimp after administration of WSSV envelope proteins, either VP19 (injected) or VP28 (given orally), compared to control shrimp challenged with WSSV (Witteveldt *et al.*, 2004a, b). These findings may open up ways to vaccinate and protect shrimp against WSSV. Especially VP28 has the potential to be developed into a (commercial) vaccine, as it can be administered orally to shrimp and is therefore convenient to use. During the experiments described in this thesis two other WSSV proteins were identified that are interesting candidates to test in a vaccination pilot study, viz. the proteins encoded by ORF75 and ORF125. Protein sequencing showed that both proteins are present in WSSV virions, and they were named vp357 (Huang *et al.*, 2002b) and vp216 (Zhang *et al.*, 2004), respectively. The microarray analysis performed showed that the corresponding genes are expressed late in infection, as expected for structural protein genes, and -moreover- at (very) high levels (chapter 5, Table 5.2). This is in contrast to most other WSSV minor structural protein genes, which have moderate expression levels. Furthermore, ORF75 and ORF125 contain a tandem repeat region (chapter 2) which, at least in bacteria, is often associated with outer membrane proteins (Denoeud and Vergnaud, 2004). As vp357 also contains a transmembrane domain (van Hulten *et al.*, 2001a) this protein, as well as vp216, may be located in the virion envelope. Being putative envelope proteins, these are the first to come in contact with the host and might therefore be involved in recognition by the host.

Concluding remarks

The work presented in this thesis has advanced our understanding of WSSV being a new virus species. Progress was made on the understanding of how this virus has evolved over the years since its first detection and what impact this may have had on the virulence of the virus. The identification of genetic markers for WSSV strain identification has provided us with tools to study WSSV outbreaks. This identification forms an important part in the fight against this devastating virus, as the origin and mechanisms of new outbreaks of WSSV is still not clear. A major challenge will still be to understand the interaction of the virus with different hosts and the role of the genomic diversity in this interaction. Furthermore, the model presented on WSSV transcription regulation contributed to a better understanding of this newly identified large circular ds DNA virus. Many parallels seem to exist with other DNA viruses, opening new venues for research approaches to further understand the dynamics of this intriguing virus both at the molecular and epidemiological level.

References

Almazan, F., Rodriguez, J. M., Andres, G., Perez, R., Vinuela, E. & Rodriguez, J. F. (1992). Transcriptional analysis of multigene family 110 of African swine fever virus. *J Virol* **66**, 6655-6667.

Anil, T. M., Shankar, K. M. & Mohan, C. V. (2002). Monoclonal antibodies developed for sensitive detection and comparison of White spot syndrome virus isolates in India. *Dis Aquat Org* **51**, 67-75.

Arnosti, D. N. (2002). Design and function of transcriptional switches in *Drosophila*. *Insect Biochem Mol Biol* **32**, 1257-1273.

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

Bailey, T. L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36.

Bigot, Y., Stasiak, K., Rouleux-Bonnin, F. & Federici, B. A. (2000). Characterization of repetitive DNA regions and methylated DNA in ascovirus genomes. *J Gen Virol* **81**, 3073-3082.

Birnstiel, M. L., Busslinger, M. & Strub, K. (1985). Transcription termination and 3' processing: the end is in site! *Cell* **41**, 349-359.

Blissard, G. W. & 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. & Ukkonen, E. (1998). Predicting gene regulatory elements in silico on a genomic scale. *Genome Research* **8**, 1202-1215.

Briggs, M., Funge-Smith, S., Subasinghe, R. & Philips, M. (2004). Introduction and movement of *Penaeus vannamei* and *Penaeus stylirostris* in Asia and the Pacific. *FAO 2004* **10**.

Broyles, S. S. (2003). Vaccinia virus transcription. *J Gen Virol* **84**, 2293-2303.

Burma, S., Mukherjee, B., Jain, A., Habib, S. & Hasnain, S. (1994). An unusual 30-kDa protein binding to the polyhedrin gene promoter of *Autographa californica* nuclear polyhedrosis virus. *Journal of Biological Chemistry* **269**, 2750-2757.

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

Chambers, J., Angulo, A., Amaratunga, D., Guo, H., Jiang, Y., Wan, J. S., Bittner, A., Frueh, K., Jackson, M. R., Peterson, P. A., Erlander, M. G. & Ghazal, P. (1999). DNA microarrays of the complex human cytomegalovirus genome: Profiling kinetic class with drug sensitivity of viral gene expression. *J Virol* **73**, 5757-5766.

Chang, P. S., Lo, C. F., Wang, Y. C. & Kou, G. H. (1996). Identification of white spot syndrome associated baculovirus (WSBV) target organs in the shrimp *Penaeus monodon* by *in situ* hybridization. *Dis Aquat Org* **27**, 131-139.

Chang, Y. S., Peng, S. E., Wang, H. C., Hsu, H. C., Ho, C. H., Wang, C. H., Wang, S. Y., Lo, C. F. & Kou, G. H. (2001). Sequencing and amplified restriction fragment length polymorphism analysis of ribonucleotide reductase large subunit gene of the White spot syndrome virus in blue crab (*Callinectes sapidus*) from American Coastal Waters. *Mar Biotechnol (NY)* **3**, 163-71.

Chantanachookin, C., Boonyaratpalin, S., Kasornchandra, J., Sataporn, D., Aekpanithanpong, U., Supamataya, K., Sriurairatana, S. & Flegel, T. W. (1993). Histology and ultrastructure reveal a new granulosis-like virus in *Penaeus monodon* affected by yellow-head disease. *Dis Aquat Org* **17**, 145-157.

Chen, L. L., Leu, J. H., Huang, C. J., Chou, C. M., Chen, S. M., Wang, C. H., Lo, C. F. & Kou, G. H. (2002a). 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.

Chen, L. L., Lo, C. F., Chiu, Y. L., Chang, C. F. & Kou, G. H. (2000). Natural and experimental infection of White spot syndrome virus (WSSV) in benthic larvae of mud crab *Scylla serrata*. *Dis Aquat Org* **40**, 157-61.

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. & Kou, G. H. (2002b). Transcriptional analysis of the DNA polymerase gene of shrimp White spot syndrome virus. *Virology* **301**, 136-147.

Cherbas, L. & Cherbas, P. (1993). The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* **23**, 81-90.

Chou, H. Y., Huang, C. Y., Lo, C. F. & Kou, G. H. (1998). Studies on transmission of White spot syndrome associated baculovirus (WSBV) in *Penaeus monodon* and *P. japonicus* via waterborne contact and oral ingestion. *Aquaculture* **164**, 263-276.

Chou, H. Y., Huang, C. Y., Wang, C. H., Chiang, H. C. & Lo, C. F. (1995). Pathogenicity of a baculovirus infection causing White spot syndrome in cultured penaeid shrimp in Taiwan. *Dis Aquat Org* **23**, 165-173.

Cochran, M. A. & Faulkner, P. (1983). Localization of homologous DNA sequences interspersed at five regions in the baculovirus AcMNPV genome. *J Virol* **45**, 961-970.

Cowley, J. A., Dimmock, C. M., Wongteerasupaya, C., Boonsaeng, V., Panyim, S. & Walker, P. J. (1999). Yellow head virus from Thailand and gill-associated virus from Australia are closely related but distinct prawn viruses. *Dis Aquat Org* **36**, 153-157.

Croizier, G. & Ribeiro, H. C. T. (1992). Recombination as a possible major cause of genetic heterogeneity in *Anticarsia gemmatalis* nuclear polyhedrosis virus wild populations. *Virus Res* **26**, 183-196.

Davison, A. J. & Moss, B. (1989). Structure of vaccinia virus late promoters. *J Mol Biol* **210**, 771-784.

Delius, H. & Clements, J. B. (1976). A partial denaturation map of herpes simplex virus type 1 DNA: evidence for inversions of the unique DNA regions. *J Gen Virol* **33**, 125-133.

Denoeud, F. & Vergnaud, G. (2004). Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. *BMC Bioinformatics* **5**, 4.

Dhar, A. K., Cowley, J. A., Hasson, K. W. & Walker, P. J. (2004). Genomic organization, biology, and diagnosis of Taura syndrome virus and yellowhead virus of penaeid shrimp. *Adv Virus Res* **63**, 353-421.

Dieu, B. T. M., Marks, H., Siebenga, J. J., Goldbach, R. W., Zuidema, D., Duong, T. P. & Vlak, J. M. (2004). Molecular epidemiology of White spot syndrome virus within Vietnam. *J Gen Virol* **85**, 3607-3618.

Durand, S., Lightner, D. V., Redman, R. M. & Bonami, J. R. (1997). Ultrastructure and morphogenesis of White spot syndrome baculovirus (WSSV). *Dis Aquat Org* **29**, 205-211.

Ebert, D. (1998). Experimental evolution of parasites. *Science* **282**, 1432-1435.

Ebert, D. & Bull, J. J. (2003). Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends Microbiol* **11**, 15-20.

Ebrahimi, B., Dutia, B. M., Roberts, K. L., Garcia-Ramirez, J. J., Dickinson, P., Stewart, J. P., Ghazal, P., Roy, D. J. & Nash, A. A. (2003). Transcriptome profile of murine gammaherpesvirus-68 lytic infection. *J Gen Virol* **84**, 99-109.

Fitzgerald, M. & Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**, 251-260.

Flegel, T. W. (1997). Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. *World J Microbiol Biotechnol* **13**, 433-442.

Friesen, P. D. (1997). Regulation of baculovirus early gene expression. In *The Baculoviruses*, pp. 141-170. Edited by L. K. Miller. New York: Plenum Press.

Gaffney, D. F., McLauchlan, J., Whitton, J. L. & Clements, J. B. (1985). A modular system for the assay of transcription regulatory signals: the sequence TAATGARAT is required for herpes simplex virus immediate early gene activation. *Nucleic Acids Res* **13**, 7847-7863.

Garcia-Escudero, R. & Viñuela, E. (2000). Structure of African swine fever virus late promoters: requirement of a TATA sequence at the initiation region. *J Virol* **74**, 8176-8182.

Garcia-Maruniak, A., Pavan, O. H. O. & Maruniak, J. E. (1996). A variable region of *Anticarsia gemmatalis* nuclear polyhedrosis virus contains tandemly repeated DNA sequences. *Virus Res* **41**, 123-132.

Guarino, L. A. & Summers, M. D. (1986). Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. *J Virol* **57**, 563-571.

Hameed, A. S., Balasubramanian, G., Musthaq, S. S. & Yoganandhan, K. (2003). Experimental infection of twenty species of Indian marine crabs with White spot syndrome virus (WSSV). *Dis Aquat Org* **57**, 157-61.

Hannenhalli, S., Chappey, C., Koonin, E. V. & Pevzner, P. A. (1995). Genome sequence comparison and scenarios for gene rearrangements: a test case. *Genomics* **30**, 299-311.

Hayakawa, T., Rohrmann, G. F. & Hashimoto, Y. (2000). Patterns of genome organization and content in lepidopteran baculoviruses. *Virology* **278**, 1-12.

Hennig, O., Itami, T., Maeda, M., Kondo, M., Natsukari, Y. & Takahashi, Y. (1998). Analyses of hemolymph immunoparameters in kurama shrimp infected with penaeid rod-shaped DNA virus. *Fish Path* **33**, 389-393.

Hossain, M. S., Khadijah, S. & Kwang, J. (2004). Characterization of ORF89--a latency-related gene of White spot syndrome virus. *Virology* **325**, 106-115.

Hsu, H. C., Lo, C. F., Lin, S. C., Liu, K. F., Peng, S. E., Chang, Y. S., Chen, L. L., Liu, W. J. & Kou, G. H. (1999). Studies on effective PCR screening strategies for White spot syndrome virus (WSSV) detection in *Penaeus monodon* brooders. *Dis Aquat Org* **39**, 13-19.

Huang, C., Zhang, L., Zhang, J., Xiao, L., Wu, Q., Chen, D. & Li, J. K. (2001). Purification and characterization of White spot syndrome virus (WSSV) produced in an alternate host: Crayfish, *Cambarus clarkii*. *Virus Res* **76**, 115-125.

Huang, C., Zhang, X., Lin, Q., Xu, X. & Hew, C. L. (2002a). Characterization of a novel envelope protein (VP281) of shrimp White spot syndrome virus by mass spectrometry. *J Gen Virol* **83**, 2385-2392.

Huang, C., Zhang, X., Lin, Q., Xu, X., Hu, Z. H. & Hew, C. L. (2002b). Proteomic analysis of shrimp White spot syndrome viral proteins and characterization of a novel envelope protein VP466. *Mol Cell Proteomics* **1**, 223-231.

Huang, R., Xie, Y., Zhang, J. & Shi, Z. (2005). A novel envelope protein involved in White spot syndrome virus infection. *J Gen Virol* **86**, 1357-1361.

Hultmark, D., Klemenz, R. & Gehring, W. J. (1986). Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. *Cell* **44**, 429-438.

Huynen, M., Snel, B., Lathe, W. & Bork, P. (2000). Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* **10**, 1204-1210.

IJkel, W. F. J. (2001). The genome of *Spodoptera exigua* multicapsid nucleopolyhedrovirus: a study on unique features: PhD Thesis, Department of Virology, Wageningen University (Wageningen).

Iwanaga, M., Takaya, K., Katsuma, S., Ote, M., Tanaka, S., Kamita, S. G., Kang, W., Shimada, T. & Kobayashi, M. (2004). Expression profiling of baculovirus genes in permissive and nonpermissive cell lines. *Biochem Biophys Res Commun* **323**, 599-614.

Jiravanichpaisal, P., Bangyeekhun, E., Söderhäll, K. & Söderhäll, I. (2001). Experimental infection of White spot syndrome virus in freshwater crayfish *Pacifastacus leniusculus*. *Dis Aquat Org* **47**, 151-157.

Jitrapakdee, S., Unajak, S., Sittidilokratna, N., Hodgson, R. A., Cowley, J. A., Walker, P. J., Panyim, S. & Boonsaeng, V. (2003). Identification and analysis of gp116 and gp64 structural glycoproteins of yellow head nidovirus of *Penaeus monodon* shrimp. *J Gen Virol* **84**, 863-873.

Khadijah, S., Neo, S. Y., Hossain, M. S., Miller, L. D., Mathavan, S. & Kwang, J. (2003). Identification of White spot syndrome virus latency-related genes in specific-pathogen-free shrimps by use of a microarray. *J Virol* **77**, 10162-10167.

Khimani, A. H., Mhashilkar, A. M., Mikulskis, A., O'Malley, M., Liao, J., Golenko, E. E., Mayer, P., Chada, S., Killian, J. B. & Lott, S. T. (2005). Housekeeping genes in cancer: normalization of array data. *Biotechniques* **38**, 739-745.

Kim, D. B., Zabierowski, S. & DeLuca, N. A. (2002). The initiator element in a herpes simplex virus type 1 late-gene promoter enhances activation by ICP4, resulting in abundant late-gene expression. *J Virol* **76**, 1548-1558.

Kleckner, N. (1981). Transposable elements in prokaryotes. *Annu Rev Genet* **15**, 341-404.

Knopf, C. W. (1998). Evolution of viral DNA-dependent DNA polymerases. *Virus Genes* **16**, 47-58.

Kool, M., Van den Berg, P. M. M. M., Tramper, J., Goldbach, R. W. & Vlak, J. M. (1993). Location of two putative origins of DNA replication of *Autographa californica* nuclear polyhedrosis virus. *Virology* **192**, 94-101.

Kozak, M. (1989). The scanning model for translation: an update. *Journal of Cell Biology* **108**, 229-241.

Krell, P. J. (1996). Passage effect of virus infection in insect cells. *Cytotechnology* **20**, 125-137.

Kurtz, J. (2005). Specific memory within innate immune systems. *Trends Immunol* **26**, 186-192.

Lan, Y., Lu, W. & Xu, X. (2002). Genomic instability of prawn White spot bacilliform virus (WSBV) and its association to virus virulence. *Virus Res* **90**, 269-274.

Leu, J. H., Tsai, J. M., Wang, H. C., Wang, A. H., Wang, C. H., Kou, G. H. & Lo, C. F. (2005). The unique stacked rings in the nucleocapsid of the White spot syndrome virus virion are formed by the major structural protein VP664, the largest viral structural protein ever found. *J Virol* **79**, 140-149.

Li, L., Lin, S. & Yanga, F. (2005a). Functional identification of the non-specific nuclease from White spot syndrome virus. *Virology* **337**, 399-406.

Li, L., Xie, X. & Yang, F. (2005b). Identification and characterization of a prawn White spot syndrome virus gene that encodes an envelope protein VP31. *Virology*.

Li, Q., Chen, Y. & Yang, F. (2004a). Identification of a collagen-like protein gene from White spot syndrome virus. *Arch Virol* **149**, 215-223.

Li, Q., Pan, D., Zhang, J. H. & Yang, F. (2004b). Identification of the thymidylate synthase within the genome of White spot syndrome virus. *J Gen Virol* **85**, 2035-2044.

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

Lightner, D. V., Redman, R. M., Hasson, K. W. & Pantoja, C. R. (1995). Taura syndrome in *Penaeus vannamei* (Crustacea: Decapoda): gross signs, histopathology and ultrastructure. *Dis Aquat Org* **21**, 53-59.

Lin, S. T., Chang, Y. S., Wang, H. C., Tzeng, H. F., Chang, Z. F., Lin, J. Y., Wang, C. H., Lo, C. F. & Kou, G. H. (2002). Ribonucleotide reductase of shrimp White spot syndrome virus (WSSV): expression and enzymatic activity in a baculovirus/insect cell system and WSSV-infected shrimp. *Virology* **304**, 282-290.

Liu, W. J., Chang, Y. S., Wang, C. H., Kou, G. H. & Lo, C. F. (2005). Microarray and RT-PCR screening for White spot syndrome virus immediate-early genes in cycloheximide-treated shrimp. *Virology* **334**, 327-341.

Liu, W. J., Yu, H. T., Peng, S. E., Chang, Y. S., Pien, H. W., Lin, C. J., Huang, C. J., Tsai, M. F., Huang, C. J., Wang, C. H., Lin, J. Y., Lo, C. F. & 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.

Liu, X. & Yang, F. (2005). Identification and function of a shrimp White spot syndrome virus (WSSV) gene that encodes a dUTPase. *Virus Res* **110**, 21-30.

Lo, C. F., Ho, C. H., Chen, C. H., Liu, K. F., Chiu, Y. L., Yeh, P. Y., Peng, S. E., Hsu, H. C., Liu, H. C., Chang, C. F., Su, M. S., Wang, C. H. & Kou, G. H. (1997). Detection and tissue tropism of White spot syndrome baculovirus (WSBV) in captured brooders of *Penaeus monodon* with a special emphasis on reproductive organs. *Dis Aquat Org* **30**, 53-72.

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. & Kou, G. H. (1996a). White spot syndrome baculovirus (WSBV) detected in cultured and captured shrimp, crabs and other arthropods. *Dis Aquat Org* **27**, 215-225.

Lo, C. F., Hsu, H. C., Tsai, M. F., Ho, C. H., Peng, S. E., Kou, G. H. & Lightner, D. V. (1999). Specific genomic DNA fragment analysis of different geographical clinical samples of shrimp White spot syndrome virus. *Dis Aquat Org* **35**, 175-185.

Lo, C. F., Lei, J. H., Ho, C. H., Chen, C. H., Peng, S. E., Chen, Y. T., Chou, C. M., Yeh, P. Y., Huang, C. J., Chou, H. Y., Wang, C. H. & Kou, G. H. (1996b). Detection of baculovirus associated with White spot syndrome (WSBV) in penaeid shrimps using polymerase chain reaction. *Dis Aquat Org* **25**, 133-141.

Lo, C. F., Wu, J. L., Chang, Y. S., Wang, H. C., Tsai, J. M. & Kou, G. H. (2004). Molecular Characterization and Pathogenicity of White spot syndrome virus. In *Current Trends in the Study of Bacterial and Viral Fish and Shrimp Diseases*, pp. 155-188. Edited by K. Y. Leung. Singapore: World Scientific Publishing Cp. Pte. Ltd.

Lotz, J. M. (1997). Viruses, Biosecurity and Specific Pathogen Free Stocks In Shrimp Aquaculture. *World J Microbiol Biotechnol* **13**, 405-413.

Lotz, J. M. & Wang, S. Y. (2004). The application of biotechnology to understanding the evolution of virulence in emerging shrimp pathogens. In *7th Asian Fisheries Forum (28 Nov - 4 Dec 2004)*. Penang, Malaysia.

Lu, A. & Miller, L. D. (1997). Regulation of baculovirus late and very late gene expression. In *The Baculoviruses*, pp. 193-216. Edited by L. K. Miller. New York: Plenum Press.

Lu, L. & Kwang, J. (2004). Identification of a novel shrimp protein phosphatase and its association with latency-related ORF427 of White spot syndrome virus. *FEBS Lett* **577**, 141-146.

Lu, L., Wang, H., Manopo, I., Yu, L. & Kwang, J. (2005). Baculovirus-mediated promoter assay and transcriptional analysis of White spot syndrome virus orf427 gene. *Virol J* **2**.

Luke, K., Radek, A., Liu, X., Campbell, J., Uzan, M., Haselkorn, R. & Kogan, Y. (2002). Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* **299**, 182-191.

Maeda, M., Itami, T., Mizuki, E., Tanaka, R., Yoshizu, Y., Doi, K., Yasunaga-Aoki, C., Takahashi, Y. & Kawarabata, T. (2000). Red swamp crawfish (*Procambarus clarkii*): an alternative experimental host in the study of White spot syndrome virus. *Acta Virol* **44**, 371-4.

Mari, J., Poulos, B. T., Lightner, D. V. & Bonami, J. R. (2002). Shrimp Taura syndrome virus: genomic characterization and similarity with members of the genus Cricket paralysis-like viruses. *J Gen Virol* **83**, 915-926.

Marks, H., Goldbach, R. W., Vlak, J. M. & van Hulten, M. C. W. (2004). Genetic variation among isolates of White spot syndrome virus. *Arch Virol* **149**, 673-697.

Marks, H., Mennens, M., Vlak, J. M. & van Hulten, M. C. W. (2003). Transcriptional analysis of the White spot syndrome virus major virion protein genes. *J Gen Virol* **84**, 1517-1523.

Marks, H., van Duijse, J. J. A., Zuidema, D., van Hulten, M. C. W. & Vlak, J. M. (2005a). Fitness and virulence of an ancestral White spot syndrome virus isolate from shrimp. *Virus Res* **110**, 9-20.

Marks, H., Vorst, O., van Houwelingen, A. M. M. L., van Hulten, M. C. W. & Vlak, J. M. (2005b). Gene-expression profiling of White spot syndrome virus in vivo. *J Gen Virol* **86**, 2081-2100.

Mayo, M. A. (2002). A summary of taxonomic changes recently approved by ICTV. *Arch Virol* **147**, 1655-1663.

Meignier, B. (1985). Vaccination against Herpes Simplex Virus Infections. In *The Herpesviruses*, pp. 265-296. Edited by B. Roizman & C. Lopez. New York: Plenum Press.

Mistretta, T. A. & Guarino, L. A. (2005). Transcriptional activity of baculovirus very late factor 1. *J Virol* **79**, 1958-1960.

Momoyama, K., Hiraoka, M., Nakano, H., Koube, H., Inouye, K. & Oseka, N. (1994). Mass mortalities of cultured kuruma shrimp *Penaeus Japonicus*, in Japan in 1993: histopathological study. *Fish Path* **29**, 141-148.

Moon, C. H., Do, J. W., Cha, S. J., Yoon, W. J., Kim, S. B., Ko, M. S., Park, M. A., Kim, J. W., Sohn, S. K., Lee, J. H. & Park, J. W. (2003). Highly conserved sequences of three major virion proteins of a Korean isolate of White spot syndrome virus (WSSV). *Dis Aquat Org* **53**, 11-13.

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

Muñoz, D., Murillo, R., Krell, P. J., Vlak, J. M. & Caballero, P. (1999). Four genotypic variants of a *Spodoptera exigua* nucleopolyhedrovirus (Se-SP2) are distinguishable by a hypervariable genomic region. *Virus Res* **59**, 61-74.

Murphy, D. (2002). Gene expression studies using microarrays: principles, problems, and prospects. *Adv Physiol Educ* **26**, 256-270.

Nadala, E. C. B., Jr. & Loh, P. C. (1998). A comparative study of three different isolates of White spot virus. *Dis Aquat Org* **33**, 231-234.

Nadala, E. C. B., Jr., Tapay, L. M. & Loh, P. C. (1998). Characterization of a non-occluded baculovirus-like agent pathogenic to penaeid shrimp. *Dis Aquat Org* **33**, 221-229.

Nicholas, K. B., Nicholas, H. B. J. & Deerfield, D. W. I. (1997). GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* **4**.

Nunan, L. M. & Lightner, D. V. (1997). Development of a non-radioactive gene probe by PCR for detection of White spot syndrome virus (WSSV). *J Virol Methods* **63**, 193-201.

Paulose-Murphy, M., Ha, N. K., Xiang, C. S., Chen, Y. D., Gillim, L., Yarchoan, R., Meltzer, P., Bittner, M., Trent, J. & Zeichner, S. (2001). Transcription program of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus). *J Virol* **75**, 4843-4853.

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

Pena, L., Yanez, R. J., Revilla, Y., Vinuela, E. & Salas, M. L. (1993). African swine fever virus guanylyltransferase. *Virology* **193**, 319-328.

Peng, S. E., Lo, C. F., Lin, S. C., Chen, L. L., Chang, Y. S., Liu, K. F., Su, M. S. & Kou, G. H. (2001). Performance of WSSV-infected and WSSV-negative *Penaeus monodon* postlarvae in culture ponds. *Dis Aquat Org* **46**, 165-172.

Possee, R. D. & Rohrmann, G. F. (1997). Baculovirus Genome Organization and Evolution. In *The Baculoviruses*, pp. 109-140. Edited by L. K. Miller. New York: Plenum Press.

Poulos, B. T., Pantoja, C. R., Bradley, D. D., Aguilar, J. & Lightner, D. V. (2001). Development and application of monoclonal antibodies for the detection of White spot syndrome virus of penaeid shrimp. *Dis Aquat Org* **47**, 13-23.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat Genet* **32**, 496-501.

Raddatz, G., Dehio, M., Meyer, T. F. & Dehio, C. (2001). PrimeArray: genome-scale primer design for DNA-microarray construction. *Bioinformatics* **17**, 98-99.

Rajcani, J., Andrea, V. & Ingeborg, R. (2004). Peculiarities of herpes simplex virus (HSV) transcription: an overview. *Virus Genes* **28**, 293-310.

Rosenberry, B. (1996). World shrimp farming 1996. In *Shrimp News International*. San Diego, California, USA.

Rosenberry, B. (2000). World shrimp farming 2000. In *Shrimp News International*. San Diego, California, USA.

Rosenberry, B. (2002). World shrimp farming 2002. In *Shrimp News International*. San Diego, California, USA.

Rosenberry, B. (2004). World shrimp farming 2004. In *Shrimp News International*. San Diego, California, USA.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular Cloning: A laboratory Manual, 2 edn. New York: Cold Spring Harbor Laboratory.

Sheets, M. D., Ogg, S. C. & Wickens, M. P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* **18**, 5799-5805.

Smale, S. T. & Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annu Rev Biochem* **72**, 449-479.

Söderhäll, I., Kim, Y. A., Jiravanichpaisal, P., Lee, S. Y. & Söderhäll, K. (2005). An ancient role for a prokineticin domain in invertebrate hematopoiesis. *J Immunol* **174**, 6153-6160.

Stingley, S. W., Ramirez, J. J. G., Aguilar, S. A., Simmen, K., Sandri, G. R. M., Ghazal, P. & Wagner, E. K. (2000). Global analysis of herpes simplex virus type 1 transcription using an oligonucleotide-based DNA microarray. *J Virol* **74**, 9916-9927.

Tapay, L. M., Nadala, E. C. B. & Loh, P. C. (1999). A polymerase chain reaction protocol for the detection of various geographical isolates of White spot virus. *J Virol Methods* **82**, 39-43.

Teunissen, O. S. P., Faber, R., Booms, G. H. R., Latscha, T. & Boon, J. H. (1998). Influence of vaccination on vibriosis resistance of the giant black tiger shrimp *Penaeus monodon* (Fabricius). *Aquaculture* **164**, 359-366.

Tidona, C. A. & Darai, G. (2000). Iridovirus homologues of cellular genes--implications for the molecular evolution of large DNA viruses. *Virus Genes* **21**, 77-81.

Tsai, J. M., Wang, H. C., Leu, J. H., Hsiao, H. H., Wang, A. H., Kou, G. H. & Lo, C. F. (2004). Genomic and proteomic analysis of thirty-nine structural proteins of shrimp White spot syndrome virus. *J Virol* **78**, 11360-11370.

Tsai, M. F., Kou, G. H., Liu, H. C., Liu, K. F., Chang, C. F., Peng, S. E., Hsu, H. C., Wang, C. H. & Lo, C. F. (1999). Long-term presence of White spot syndrome virus (WSSV) in a cultivated shrimp population without disease outbreaks. *Dis Aquat Org* **38**, 107-114.

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. & Kou, G. H. (2000a). Transcriptional analysis of the ribonucleotide reductase genes of shrimp White spot syndrome virus. *Virology* **277**, 92-99.

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. & Lo, C. F. (2000b). 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.

Tzeng, H. F., Chang, Z. F., Peng, S. E., Wang, C. H., Lin, J. Y., Kou, G. H. & 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.

Umene, K. (1991). Recombination of the internal direct repeat element DR2 responsible for the fluidity of the a sequence of herpes simplex virus type 1. *J Virol* **65**, 5410-5416.

van de Braak, C. B., Botterblom, M. H., Huisman, E. A., Rombout, J. H. & van der Knaap, W. P. (2002). Preliminary study on haemocyte response to White spot syndrome virus infection in black tiger shrimp *Penaeus monodon*. *Dis Aquat Org* **51**, 149-155.

van Hulten, M. C. W., Goldbach, R. W. & Vlak, J. M. (2000a). Three functionally diverged major structural proteins of White spot syndrome virus evolved by gene duplication. *J Gen Virol* **81**, 2525-2529.

van Hulten, M. C. W., Reijns, M., Vermeesch, A. M. G., Zandbergen, F. & Vlak, J. M. (2002). Identification of VP19 and VP15 of White spot syndrome virus (WSSV) and glycosylation status of the WSSV major structural proteins. *J Gen Virol* **1**, 257-265.

van Hulten, M. C. W., Tsai, M. F., Schipper, C. A., Lo, C. F., Kou, G. H. & Vlak, J. M. (2000b). Analysis of a genomic segment of White spot syndrome virus of shrimp containing ribonucleotide reductase genes and repeat regions. *J Gen Virol* **81**, 307-316.

van Hulten, M. C. W. & 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., Westenberg, M., Goodall, S. D. & Vlak, J. M. (2000c). Identification of two major virion protein genes of White spot syndrome virus of shrimp. *Virology* **266**, 227-236.

van Hulten, M. C. W., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Klein Lankhorst, R. & 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. & Vlak, J. M. (2001b). White spot syndrome virus envelope protein VP28 is involved in the systemic infection of shrimp. *Virology* **285**, 228-233.

Vlak, J. M., Bonami, J. R., Flegel, T. W., Kou, G. H., Lightner, D. V., Lo, C. F., Loh, P. C. & Walker, P. W. (2005). *Nimaviridae*. In *VIIIth Report of the International Committee on Taxonomy of Viruses*, pp. 187-192. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball: Elsevier.

Wagner, E. K., Guzowski, J. F. & Singh, J. (1995). Transcription of the herpes simplex virus genome during productive and latent infection. *Prog Nucleic Acid Res Mol Biol* **51**, 123-165.

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. & Kou, G. H. (1995). Purification and genomic analysis of baculovirus associated with White spot syndrome (WSBV) of *Penaeus monodon*. *Dis Aquat Org* **23**, 239-242.

Wang, C. H., Yang, H. N., Tang, C. Y., Lu, C. H., Kou, G. H. & Lo, C. F. (2000a). Ultrastructure of White spot syndrome virus development in primary lymphoid organ cell cultures. *Dis Aquat Org* **41**, 91-104.

Wang, Q., Nunan, L. M. & Lightner, D. V. (2000b). Identification of genomic variations among geographic isolates of White spot syndrome virus using restriction analysis and Southern blot hybridization. *Dis Aquat Org* **43**, 175-181.

Wang, Q., Poulos, B. T. & Lightner, D. V. (2000c). Protein analysis of geographic isolates of shrimp White spot syndrome virus. *Arch Virol* **145**, 263-274.

Wang, Q., White, B. L., Redman, R. M. & Lightner, D. V. (1999a). Per os challenge of *Litopenaeus vannamei* postlarvae and *Farfantepenaeus duorarum* juveniles with six geographic isolates of White spot syndrome virus. *Aquaculture* **170**, 179-194.

Wang, Y. C., Lo, C. F., Chang, P. S. & Kou, G. H. (1998). Experimental infection of White spot baculovirus in some cultured and wild decapods in Taiwan. *Aquaculture* **164**, 221-231.

Wang, Y. G., Hassan, M. D., Shariff, M., Zamri, S. M. & Chen, X. (1999b). Histopathology and cytopathology of White spot syndrome virus (WSSV) in cultured *Penaeus monodon* from peninsular Malaysia with emphasis on pathogenesis and the mechanism of white spot formation. *Dis Aquat Org* **39**, 1-11.

Wang, Z., Hu, L., Yi, G., Xu, H., Qi, Y. & Yao, L. (2004). ORF390 of White spot syndrome virus genome is identified as a novel anti-apoptosis gene. *Biochem Biophys Res Commun* **325**, 899-907.

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

Witteveldt, J., Cifuentes, C. C., Vlak, J. M. & van Hulten, M. C. W. (2004a). Protection of *Penaeus monodon* against White spot syndrome virus by oral vaccination. *J Virol* **78**, 2057-2061.

Witteveldt, J., van Hulten, M. C. W. & 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., Vermeesch, A. M., Langenhof, M., de Lang, A., Vlak, J. M. & van Hulten, M. C. W. (2005). Nucleocapsid protein VP15 is the basic DNA binding protein of White spot syndrome virus of shrimp. *Arch Virol* **150**, 1121-1133.

Witteveldt, J., Vlak, J. M. & van Hulten, M. C. W. (2004b). Protection of *Penaeus monodon* against White spot syndrome virus using a WSSV subunit vaccine. *Fish Shellfish Immunol* **16**, 571-579.

Wongprasert, K., Khanobdee, K., Glunukarn, S. S., Meeratana, P. & Withyachumnarnkul, B. (2003). Time-course and levels of apoptosis in various tissues of black tiger shrimp *Penaeus monodon* infected with White-spot syndrome virus. *Dis Aquat Org* **55**, 3-10.

Wongteerasupaya, C., Pungchai, P., Withyachumnarnkul, B., Boonsaeng, V., Panyim, S., Flegel, T. W. & Walker, P. J. (2003). High variation in repetitive DNA fragment length for White spot syndrome virus (WSSV) isolates in Thailand. *Dis Aquat Org* **54**, 253-257.

Wongteerasupaya, C., Vickers, J. E., Sriurairatana, S., Nash, G. L., Akarajamorn, A., Boonsaeng, V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B. & 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*. *Dis Aquat Org* **21**, 69-77.

Wu, W., Wang, L. & Zhang, X. (2005). Identification of White spot syndrome virus (WSSV) envelope proteins involved in shrimp infection. *Virology* **332**, 578-583.

Wyban, J. A. & Sweeney, J. N. (1991). Intensive shrimp production technology, pp. 158. Hawaii: High Health Aquaculture Inc.

Yamagishi, J., Isobe, R., Takebuchi, T. & Bando, H. (2003). DNA microarrays of baculovirus genomes: differential expression of viral genes in two susceptible insect cell lines. *Arch Virol* **148**, 587-597.

Yang, F., He, J., Lin, X. H., Li, Q., Pan, D., Zhang, X. B. & Xu, X. (2001). Complete genome sequence of the shrimp White spot bacilliform virus. *J Virol* **75**, 11811-11820.

You, Z., Nadala, E. C., Jr., Yang, J., Van Hulten, M. C. & Loh, P. C. (2002). Production of polyclonal antiserum specific to the 27.5 kDa envelope protein of White spot syndrome virus. *Dis Aquat Org* **51**, 77-80.

Zhang, X., Huang, C., Tang, X., Zhuang, Y. & Hew, C. L. (2004). Identification of structural proteins from shrimp White spot syndrome virus (WSSV) by 2DE-MS. *Proteins* **55**, 229-235.

Zhang, X., Huang, C., Xu, X. & Hew, C. L. (2002a). Identification and localization of a prawn White spot syndrome virus gene that encodes an envelope protein. *J Gen Virol* **83**, 1069-1074.

Zhang, X., Huang, C. H., Xu, X. & Hew, C. L. (2002b). Transcription and identification of an envelope protein gene (p22) from shrimp White spot syndrome virus. *J Gen Virol* **83**, 471-477.

Zhang, X., Xu, X. & Hew, C. L. (2001). The structure and function of a gene encoding a basic peptide from prawn White spot syndrome virus. *Virus Res* **79**, 137-144.

Summary

White spot syndrome virus (WSSV) is a large enveloped DNA virus that infects shrimp and other crustaceans. The virions are approximately 275 x 120 nm in size and have an ovoid to bacilliform shape and a tail-like appendage at one end. Sequencing revealed that the circular, double stranded (ds) DNA genome of WSSV ranges between 293 and 307 kb in size depending on the WSSV isolate. For a sequenced isolate originating from Thailand (WSSV-TH) 184 putative open reading frames (ORFs) were identified on the genome, most of which are unassigned as they lack homology to known genes in public databases. Based on its unique morphological and genetic features, WSSV has been accommodated in the new virus family *Nimaviridae* (genus *Whispovirus*).

WSSV causes serious economic losses in shrimp culture, as 100% cumulative mortalities can be reached within 3-10 days under farming conditions. After its discovery in 1992 in Taiwan WSSV has quickly spread into Southeast-Asia and subsequently to shrimp farming areas all over the world. This thesis aims at obtaining fundamental insights in the genomic structure ("genomics") and transcription regulation ("transcriptomics") of WSSV. This in turn may provide better insight in the molecular basis of WSSV biology and epidemiology, which can be useful in the identification of targets for WSSV intervention strategies.

Alignment of the complete genome sequences of the isolates WSSV-TW, WSSV-CN and WSSV-TH, originating from Taiwan, China and Thailand, respectively, revealed that the sequences were very similar (over 99% sequence identity), suggesting that the isolates are variants of the same virus species (*Whispovirus*) and probably evolved recently from a common ancestor (Chapter 2). Two major polymorphic loci were identified, variable region (VR) ORF14/15 and VR ORF23/24, and both appeared to be genomic regions where large deletions occur. Further polymorphisms included loci with variable numbers of tandem repeats (VNTR loci). Next to VR ORF14/15 and VR ORF23/24, three of these loci, located in the regions coding for ORF75, ORF94 and ORF125, were identified as useful markers in epidemiological and ecological studies. The highly conserved genomic loci, e.g. the gene encoding the major structural virion protein VP26, are useful for reliable monitoring of WSSV infections in PCR based assays.

The observation that the isolate WSSV-TH contains a large deletion in VR ORF23/24 relative to the isolates WSSV-TW and WSSV-CN suggested the evolution and spread of WSSV from a common ancestor, provisionally located near the southeast coast of China. Further support for this model was obtained by the genomic characterization of eight WSSV isolates collected in 2003 and 2004 along the central and south coast of Vietnam (VN) during WSSV outbreaks (Chapter 3). These WSSV-VN isolates contained deletions of intermediate size in VR ORF23/24 relative to WSSV-TW and WSSV-TH. In VR ORF14/15, the WSSV-VN isolates contained deletions of various sizes compared to WSSV-TH. These collective

data suggest that the VN isolates and WSSV-TH have a common lineage, which branched off from WSSV-TW and WSSV-CN early on, and that WSSV possibly entered Vietnam by multiple introductions. Further comparisons among the WSSV-VN isolates revealed that the VNTR loci in ORF75 and ORF125, but not in ORF94, are suitable markers to study local and regional spread of WSSV.

To study the possible effect of the genetic differences on the fitness and virulence of WSSV, two divergent WSSV isolates (TH-96-II and WSSV-TH) were compared (Chapter 4). TH-96-II was a newly characterized archival WSSV isolate from 1996, which has the largest genome size (~312 kb) of all WSSV isolates identified thus far. As TH-96-II does not contain deletions in either VR ORF14/15 or VR ORF23/24, it may be ancestral to all known WSSV isolates. WSSV-TH contains the smallest genome (~293 kb) identified at present, due to large deletions in VR ORF14/15 and VR ORF23/24. Comparison between TH-96-II and WSSV-TH, when administered to shrimp *Penaeus monodon,* showed a higher virulence and competitive fitness for the latter. This may suggest that the virus became more virulent over the years during the epidemic while moving south. This enhanced virulence is possibly caused by the continuous contact with susceptible animals, a behavior also seen with some other emerging viruses. Since the more virulent variant (WSSV-TH) has a smaller genome, it may replicate faster to reach a lethal dose. However, it is also possible that the observed differences in virulence are caused by other genetic polymorphisms between the two isolates.

As WSSV differs profoundly from other large ds DNA viruses and mainly contains unique genes, the mechanism of gene expression and transcription regulation of this new virus was investigated in the second part of this thesis. To study WSSV gene expression on a genome wide scale, a WSSV DNA microarray was constructed containing probes corresponding to nearly all putative WSSV ORFs (Chapter 5). Using a WSSV infection time course we could show expression of at least 79% of the WSSV ORFs included on the microarray in gill tissue of *Penaeus monodon*. Clustering of the transcription profiles of the individual genes showed the presence of two major classes of genes, a putative early and a putative late class, suggesting that the WSSV genes at large are expressed in a coordinated and cascaded fashion. Five genes encoding WSSV major virion proteins (VP28, VP26, VP24, VP19 en VP15), which clustered in the late class, were further confirmed to be late by RT-PCR (Chapter 6). Furthermore, the 5' and 3' ends of the mRNA of these late genes were determined for identification of common promoter motifs.

To search for common conserved WSSV promoter motifs associated with WSSV early or late gene expression, as determined by the microarrays, two *in silico* methods were employed (Chapter 7). The abundance of all 4 through 8 nucleotide motifs in the upstream sequences of WSSV genes relative to the complete genome was determined and the upstream sequences of early or late WSSV genes were analyzed for conserved sequences motifs using MEME. Both methods were complemented by alignments of empirically determined 5' ends of various WSSV mRNAs. The collective information shows that the upstream region of

WSSV early genes, containing a TATA box and an initiator sequence, is reminiscent to *Drosophila* RNA polymerase II promoters, suggesting utilization of the cellular transcription machinery for generating early transcripts. The alignment of the 5' ends of known late genes, including the 5' ends determined in chapter 6, identified a degenerate consensus late transcription initiation motif (ATNAC). Of these genes, only one contained a functional TATA box. However, almost half of the WSSV late genes, as assigned by microarrays, did contain a TATA box in their upstream region. This may suggest the presence of two separate classes of late WSSV genes, one exploiting the cellular RNA polymerase II system for mRNA synthesis and the other generating messengers by a new virus-induced transcription mechanism.

Alignments of the 3' ends of various WSSV mRNAs suggest that there is no difference in polyadenylation between early and late mRNAs. The WSSV polyadenylation characteristics of both classes resemble regular polyadenylation in eukaryotic mRNAs, which is typically located 10 to 25 nt downstream of the sequence AATAAA.

In conclusion, the research performed for this thesis has led to a model on the mechanism of WSSV gene expression, and the promoter motifs involved (Chapter 8). The identification of genetic markers has led to more insight in the quick geographical spread of the virus, and the genetic characterization of WSSV isolates may add to the identification of virulence related factors on the WSSV genome. The fundamental insights obtained in the biology and epidemiology of WSSV in this thesis may help in the identification of WSSV genes which can be targets for WSSV intervention strategies.

Samenvatting

Het Witte-vlekken-syndroom-virus (Engelse naam: "White spot syndrome virus", afkorting: WSSV) is een groot, dubbelstrengs DNA virus dat garnalen en andere kreeftachtigen infecteert. Het WSSV-virion meet, inclusief het lipidemembraan ("envelop") waardoor het omgeven wordt, ongeveer 275 nm bij 120 nm. Het virion heeft een ellips-staafvormig uiterlijk en bezit een staartje aan één van de uiteinden. Na bepaling van de volledige nucleotidenvolgorde bleek het circulaire WSSV genoom, afhankelijk van de oorsprong van het isolaat, 293 tot 307 kilobasenparen (kb) groot te zijn. Voor een isolaat uit Thailand (WSSV-TH) bleek dit genoom 184 potentiële genen, ook wel open leesramen (open reading frames; ORFs) genoemd, te bevatten. Voor de meeste van deze ORFs kon geen mogelijke functie worden gevonden, omdat er geen homologe ORFs in de GenBank aanwezig waren. Vanwege het unieke karakter van zowel het virion als de genetische samenstelling is WSSV in een nieuwe virusfamilie geplaatst met de familienaam "*Nimaviridae*" (nima = draad), binnen het genus "*Whispovirus*" (een samenvoeging van "<u>whi</u>te <u>spo</u>t").

WSSV veroorzaakt wereldwijd grote economische verliezen in de garnalenteelt. In gekweekte garnalen kan WSSV binnen 3 tot 10 dagen een cumulatieve sterfte van 100% veroorzaken. WSSV werd voor het eerst aangetroffen in Taiwan in 1992, waarna het zich snel verspreidde over Zuidoost-Azië. Tegenwoordig wordt WSSV bijna in alle gebieden aangetroffen waar op grote schaal garnalen gekweekt worden. Het doel van dit proefschrift is om fundamenteel inzicht te krijgen in de structuur van het WSSV genoom ("genomics") en de wijze waarop de vertaling in RNA gereguleerd wordt ("transcriptomics"). Hierdoor kan een beter idee verkregen worden over de moleculaire basis van de biologie en de epidemiologie van het virus. Dit kan wellicht leiden tot nieuwe strategieën om dit virus te bestrijden.

Bij vergelijk van de genomen van de isolaten WSSV-TW (afkomstig uit Taiwan), WSSV-CN (afkomstig uit China) en WSSV-TH bleken deze grotendeels identiek te zijn (meer dan 99% nucleotidenovereenkomst). Deze drie isolaten behoren alle tot het genus *Whispovirus* en zijn dus waarschijnlijk recentelijk ontstaan uit een gezamenlijke voorouder. De belangrijkste genetische verschillen tussen de drie isolaten waren twee gebieden in het genoom waar grote deleties voorkomen, variabele regio (VR) ORF14/15 en VR ORF 23/24 genoemd. Verder zijn er verscheidene plaatsen in het genoom die een verschillende hoeveelheid van aaneengesloten, repeterende sequenties bevatten ("VNTR loci"). Drie van deze VNTR loci, namelijk de stukken die coderen voor ORF75, ORF94 en ORF125, bleken, naast VR ORF14/15 en VR ORF23/24, nuttige genetische merkers te zijn voor epidemiologisch en ecologisch onderzoek. De meest geconserveerde delen van het WSSV genoom, bijvoorbeeld het deel dat het belangrijke structurele eiwit VP26 codeert, kunnen gebruikt worden om garnalen betrouwbaar te kunnen onderzoeken op WSSV infectie via PCR testen.

Het feit dat WSSV-TH een grotere deletie bevat in VR ORF23/24 dan WSSV-TW en WSSV-CN leek erop te wijzen dat WSSV zich vanaf de zuidoostelijke Chinese kust vanuit een gemeenschappelijke voorouder verspreid heeft richting Thailand. Ter ondersteuning van dit model werden acht Vietnamese (VN) WSSV isolaten genetisch gekarakteriseerd (Hoofdstuk 3). Deze WSSV-VN isolaten waren verzameld gedurende WSSV uitbraken aan de midden- en zuidkust van Vietnam in 2003 en 2004. De grootte van de genomische deleties in VR ORF23/24 van deze Vietnamese isolaten lag tussen die van WSSV-TW en WSSV-TH in. In VR ORF14/15 hadden alle acht Vietnamese isolaten genomische deleties ten opzichte van WSSV-TH. Deze bevindingen wijzen erop dat de Vietnamese isolaten en WSSV-TH nauw verwant zijn en al vroeg afgetakt zijn van WSSV-TW en WSSV-CN. Mogelijk is WSSV op meerdere plekken Vietnam binnengekomen. Verder vergelijk tussen de Vietnamese isolaten toonde aan dat met name de VNTR loci binnen ORF75 en ORF125 geschikte merkers zijn om plaatselijke en regionale verspreiding van WSSV te bestuderen. De VNTR locus binnen ORF94 bleek hiervoor minder geschikt.

Om het effect van de genetische verschillen op de conditie en virulentie van WSSV te bestuderen werden twee WSSV isolaten, namelijk WSSV-TH en TH-96-II, met elkaar vergeleken (Hoofdstuk 4). TH-96-II is een archivaal isolaat met de grootste WSSV genoomvariant die tot nu toe bekend is (~312 kb). Dit isolaat uit 1996 werd pas onlangs gekarakteriseerd. TH-96-II is mogelijk een voorouder van alle WSSV isolaten die tot nog toe gekarakteriseerd zijn, omdat het geen deleties bevat in VR ORF14/15 en VR ORF23/24. WSSV-TH bevat het kleinste bekende WSSV genoom (~293 kb) dat tot nog toe bekend is. Uit onderzoek bleek dat WSSV-TH niet alleen virulenter was dan TH-96-II, maar TH-96-II ook verdrong gedurende infectie van garnalen (*Penaeus monodon*) met beide isolaten. Dit wijst erop dat WSSV gedurende de verspreiding over de jaren virulenter is geworden. Deze toegenomen virulentie kan mogelijk verklaard worden door het feit dat er steeds weer nieuwe, gezonde groepen garnalen aan het virus worden blootgesteld. Dit scenario wordt ook waargenomen bij andere virussen. Op moleculair niveau wordt het verschil mogelijk verklaard door het feit dat WSSV-TH een kleiner genoom heeft. Een dergelijk klein genoom kan zich sneller vermenigvuldigen, waardoor eerder een dodelijke dosis virus bereikt wordt. Het kan echter ook zijn dat andere genetische verschillen tussen beide WSSV isolaten het verschil in conditie en virulentie veroorzaken.

WSSV is erg verschillend van andere grote, dubbelstrengs DNA virussen en codeert voornamelijk unieke genen. Daarom was het erg interessant de wijze waarop genexpressie van dit nieuwe virus gereguleerd wordt te onderzoeken. Om expressie van alle WSSV genen tegelijkertijd te kunnen vaststellen werd een WSSV DNA microarray gemaakt (Hoofdstuk 5). Gedurende een tijdreeks van WSSV infectie van garnaal kwam tenminste 79% van alle WSSV genen, die met de microarray gemeten konden worden, tot expressie in het kieuwweefsel. Tijdens clustering van de tijdsafhankelijke expressie van de individuele genen bleek een opsplitsing mogelijk in twee grote groepen van genen. Eén groep had al vroeg na

infectie een maximale expressie (vroege groep), terwijl de andere groep later tijdens infectie maximale expressie had (late groep). Dit wijst erop dat de WSSV genen in de loop van de infectie op een georganiseerde wijze tot expressie komen. Met behulp van RT-PCR werd voor vijf belangrijke structurele virion eiwitten (VP28, VP26, VP24, VP19 en VP15) bevestigd dat bijbehorende genen tot de late groep behoorden (Hoofdstuk 6). Ook werden van deze genen de 5' en 3' uiteinden bepaald, met als doel om promoter motieven vinden.

Met behulp van bioinformatica werd gezocht naar geconserveerde WSSV promoter motieven, die samenhangen met de vroege dan wel de late groep genen. Van alle nucleotide motieven met een lengte van 4 tot 8 nucleotiden werd de hoeveelheid, die aanwezig was in de genoomsequenties voorafgaand aan alle genen, vergeleken met de hoeveelheid die aanwezig was in het totale genoom. Verder werden de genoomsequenties voorafgaand aan hetzij de genen van de vroege groep dan wel de genen van de late groep onderzocht op geconserveerde sequentie motieven met behulp van het computer algoritme MEME. Deze analyses werden aangevuld met een vergelijking van experimenteel bepaalde 5' uiteinden van een aantal mRNAs van WSSV. Hieruit bleek dat de genoomsequentie, voorafgaand aan genen uit de vroege groep, sterk lijkt op de promoter sequenties van het RNA polymerase II systeem zoals deze gevonden zijn bij de fruitvlieg (*Drosophila*), omdat beide een TATA box bevatten en ook een geconserveerde sequentie waarin transcriptie start. Dit wijst erop dat WSSV het standaard transcriptie mechanisme, zoals dat al aanwezig is in de gastheercel vòòr WSSV infectie, gebruikt voor de expressie van genen uit de vroege groep. Bij het vergelijken van de 5' uiteinden van een aantal bekende late genen, waaronder die uit hoofdstuk 6, bleek transcriptie van deze genen in een geconserveerd motief (ATNAC) te starten. Van de genen die vergeleken werden was er slechts één waar een functionele TATA box aanwezig was voorafgaand aan dit motief. Echter, van de genen uit de late groep uit de clustering, zoals deze aan de hand van de microarray data gemaakt was, had ongeveer de helft een TATA box in de genoomsequentie voorafgaand aan het gen. Daarom lijkt het erop dat er twee verschillende typen van late genen zijn. Het ene type late genen wordt tot expressie gebracht via het cellulaire RNA polymerase II, terwijl het andere type tot expressie komt via een nieuw, virus-geïnduceerd transcriptie mechanisme.

Aan de hand van het onderzoek dat in dit proefschrift is beschreven is een model opgesteld dat de wijze beschrijft waarop WSSV genexpressie plaatsvindt. In dit model zijn ook de promoter motieven verwerkt die hierbij een rol spelen (Hoofdstuk 8). Het identificeren van genetische merkers heeft geleid tot meer inzicht in het mechanisme van de snelle verspreiding van WSSV tot een hypothese hoe WSSV zich zo snel heeft kunnen verspreiden over de wereld. Het genetisch karakteriseren van WSSV isolaten kan in de toekomst helpen om genen te vinden die betrokken zijn bij de virulentie van WSSV. De in dit proefschrift beschreven fundamentele inzichten in de biologie en epidemiologie van WSSV kunnen mogelijk bijdragen aan het vinden van nieuwe oplossingen voor de bestrijding van WSSV.

Nawoord

Onderzoek doen en het uitvoeren van experimenten is verrassend en spannend. Schijnbaar onbelangrijke "bij-"projectjes blijken opeens zo interessant dat ze uitgroeien tot volwaardige publicaties, terwijl het voornaamste deel van het onderzoek juist meer vragen lijkt op te roepen dan dat het antwoorden geeft. Totaal onverklaarbare uitkomsten en mislukte proeven resulteren uiteindelijk weer in de meest exotische nieuwe hypotheses en proeven. Gelukkig zijn er gedurende die tijd altijd studenten, collega's, vrienden en familie die je op het rechte pad houden, kritische vragen stellen, met nieuwe ideeën komen of aanvullende experimenten doen, dan wel een uurtje met je sporten, een "potje willen vluggeren" of met je naar de kroeg gaan. Ik maak graag van de mogelijkheid gebruik om een ieder die heeft bijgedragen aan dit proefschrift te bedanken. Maar bovenal wil ik alle collega's van "Viro" bedanken voor de open sfeer en gezelligheid op "de Vakgroep". Daardoor kijk ik met erg veel plezier terug op mijn promotieonderzoek!

Toen ik zo'n 4,5 jaar geleden mijn sollicitatie gesprek had zat Mariëlle op de achtergrond stil wat aantekeningen te maken. Pas toen mij een weekje later via de telefoon geen andere keus werd gelaten dan zo snel mogelijk met het project te starten bemerkte ik het enthousiasme en de vastberadenheid waarmee ik ook door je begeleid ben, Mariëlle. Geen passieve houding, er moest aangepakt worden. Ook na je vertrek naar Australië ben je me met raad en daad blijven bijstaan. Je snelle en adequate optreden bij hoofdstuk 5 en bij de discussie heb ik erg op prijs gesteld. Helaas kan je als initiator van het Whispovac project en als co-promoter niet aanwezig zijn vanwege zwangerschap. Voor daarna alvast heel veel plezier en succes in je nieuwe baan.

Just, als promotor had je de belangrijke taak de voortgang en kwaliteit van mijn onderzoek te bewaken. Door me veel vrijheid te geven kreeg ik de mogelijkheid een eigen stempel op het onderzoek te drukken. Daarbij kon ik dankbaar gebruik maken van je wereldwijde connecties met Baculovirologen, Whispovirologen en andere wetenschappers. Verder wil ik je bedanken voor het kritisch corrigeren van al mijn manuscripten. Rob, als tweede promotor was je wat meer aan de zijlijn bij het project betrokken. Toch bleek je altijd zeer goed op de hoogte van wat er gaande was en wist je gedurende meerdere brainstormsessies nieuwe invalshoeken en verbanden te vinden. Daarnaast vond ik het bijzonder prettig dat je altijd tijd vrij kon maken op momenten dat ik behoefte had aan feedback.

Op het lab heb ik nauw samengewerkt met Jeroen en Angela. Jeroen, met name bij de "*in vivo*" experimenten kon ik altijd op je rekenen. Ik heb veel geleerd van je (soms) ongenuanceerde wijsheden. Ik ben erg blij dat je me tijdens mijn promotie terzijde wil staan. Alvast veel succes bij je eigen verdediging. Angela, als "labmanager" heb je ons op een prettige manier goed onder de duim gehouden. Als eerste hulp had je een belangrijke functie wanneer een experiment mislukt was of als ik een verhaal kwijt moest. En als analiste was je

"Mannen" van Schaakvereniging Wageningen, elke partij met of tegen jullie is een avontuur op zich. De donderdagavonden in de Vredehorst, de nachten in Gent en de lange "externe" zaterdagen zijn blijvende hoogtepunten. Ik hoop in de toekomst jullie kleuren weer te mogen verdedigen.

Papa, Mama en Mieke, jullie staan aan de wieg van dit alles. Bedankt voor de onvoorwaardelijke steun en betrokkenheid. Anja en Marthijn, ik bewonder jullie ongebreidelde hoeveelheid energie waarmee alles aangepakt wordt. Bedankt voor de keren dat jullie mij gered hebben uit netelige situaties.

Lieve Carien, het is een voorrecht elke dag te mogen genieten van je vrolijke, ontwapenende karakter, je enthousiasme en je brede belangstelling. Na België en Nederland zijn er nog heel wat landen te bezoeken. Vurig hoop ik dat Canada je eerste bestemming zal zijn!

Curriculum Vitae

Hendrik Marks werd geboren op 9 april 1976 in Heino. In 1994 behaalde hij het gymnasium diploma aan het Gymnasium Celeanum te Zwolle. In datzelfde jaar werd begonnen met de studie Moleculaire Wetenschappen aan Wageningen Universiteit (WU), waar in 2000 het ingenieursdiploma werd behaald. Tijdens de doctoraalfase heeft hij afstudeeronderzoek verricht bij achtereenvolgens de leerstoelgroep Moleculaire Biologie aan de WU, waar het transporteiwit van Cowpea mosaic virus (CPMV) werd onderzocht (Dr. P. Bertens en Dr. J. Wellink), en de afdeling Zoogdiervirologie binnen het onderzoeksinstituut ID-Lelystad, waar werd onderzocht hoe Bovine herpesvirus (BHV) herkenning door het immuunsysteem van de gastheer voorkomt (Dr. F. A. M. Rijsewijk). In 1996 werd de studie één jaar ingewisseld voor het bestuur van de Sportstichting Landbouwuniversiteit Wageningen (huidige SWU Thymos).

Na zijn studie werkte Hendrik een jaar aan het Academisch Ziekenhuis Groningen (AZG, huidige UMCG), waar de rol van MRP eiwitten bij patiënten met de longziekte COPD werd bestudeerd. Vervolgens was hij van april 2001 tot oktober 2005 werkzaam als onderzoeker in opleiding bij het Laboratorium voor Virologie van Wageningen Universiteit, onder begeleiding van dr. ir. M. C. W. van Hulten, prof. dr. R. W. Goldbach en prof. dr. J. M. Vlak. Van het daar uitgevoerde onderzoek, dat gefinancierd werd door Intervet International B.V., staan de resultaten beschreven in dit proefschrift. Vanaf november 2005 is hij werkzaam als Postdoc bij het Department of Zoology, University of British Columbia (Vancouver, Canada).

Account

Marks, H., Vorst, O., van Houwelingen, A. M. M. L., van Hulten, M. C. W. & Vlak, J. M. (2005). Gene expression profiling of White spot syndrome virus *in vivo*. *J Gen Virol* **86**, 2081-2100.

Marks, H., van Duijse, J. J. A., Zuidema, D., van Hulten, M. C. W. & Vlak, J. M. (2005). Fitness and virulence of an ancestral White spot syndrome virus isolate from shrimp. *Virus Res* **110**, 9-20.

Marks, H., Ren, X., Sandbrink, H., Van Hulten, M. C. W. & Vlak, J. M. (2005, submitted). *In silico* identification of putative promoter motifs of White spot syndrome virus.

Dieu, B. T. M., Marks, H., Siebenga, J. J., Goldbach, R. W., Zuidema, D., Duong, T. P. & Vlak, J. M. (2004). Molecular epidemiology of White spot syndrome virus within Vietnam. *J Gen Virol* **85**, 3607-3618.

Marks, H., Goldbach, R. W., Vlak, J. M. & van Hulten, M. C. W. (2004). Genetic variation among isolates of White spot syndrome virus. *Arch Virol* **149**, 673-697.

Marks, H., Ren, X., Witteveldt, J., Sandbrink, H., Vlak, J. M. & Van Hulten, M. C. W. (2004). Transcription regulation and genomics of White spot syndrome virus. In *Diseases in Asian Aquaculture V*. Edited by P. J. Walker.

Zuidema, D., van Hulten, M. C. W., Marks, H., Witteveldt, J. & Vlak, J. M. (2004). Virus-host interactions of White spot syndrome virus. In *Current Trends in the Study of Bacterial and Viral Fish and Shrimp Diseases*, pp. 237-255. Edited by K. Y. Leung. Singapore: World Scientific Publishing Cp. Pte. Ltd.

Marks, H., Mennens, M., Vlak, J. M. & Van Hulten, M. C. W. (2003). Transcriptional analysis of the White spot syndrome virus major virion protein genes. *J Gen Virol* **84**, 1517-1523.

van der Deen, M., Marks, H., Willemse, B. W., Postma, D. S., Müller, M., Smit, E. F., Scheffer, G. L., Scheper, R. J., de Vries, E. G. & Timens, W. (2005, submitted). Diminished expression of MRP1 in bronchial epithelium of COPD patients.

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

Gopinath, K., Bertens, P., Pouwels, J., Marks, H., Van Lent, J., Wellink, J. & Van Kammen, A. (2003). Intracellular distribution of cowpea mosaic virus movement protein as visualised by green fluorescent protein fusions. *Arch Virol* **148**, 2099-2114.

Persbericht

Garnalenvirus wisselt halverwege van overschrijfsysteem

Het wittevlekkenvirus bij garnalen blijkt halverwege de infectie van overschrijfsysteem te wisselen. Dat ontdekte de Wageningse viroloog ir. Hendrik Marks tijdens zijn promotieonderzoek. Marks hoopt op 28 november te promoveren aan de Wageningen Universiteit.

Viroloog Hendrik Marks bestudeerde het wittevlekkenvirus waar reuzengarnalen veel last van hebben. Hij onderzocht het DNA en bekeek hoe en wanneer het virus in actie komt. Uiteindelijk moet de kennis leiden tot bestrijdingsmethoden.

Omdat het garnalenvirus niet lijkt op andere virussen onderwierp Marks het virus aan een grondig onderzoek. De viroloog maakte glazen plaatjes met daarop elk gen van het virus in een apart druppeltje.

Deze zogeheten DNA-chips lieten zien dat een deel van de genen hun werk deed aan het begin van een infectie. Een ander deel kwam juist in de loop van de infectie in actie. Promovendus Marks ontdekte dat het virus in het begin het overschrijfsysteem van de geïnfecteerde garnaal gebruikt. Later stapt het virus over op een heel nieuw soort overschrijfsysteem.

De viroloog vond meer aanknopingspunten om het virus te bestrijden. Zo ontdekte hij in het DNA van alle virusvarianten een zelfde stuk dat voor eiwit VP26 codeert. Elke geïnfecteerde garnaal bevat dit stuk DNA. Een simpele DNA-test kan het stuk aantonen. Het maakt niet uit welke variant van het virus aanwezig is.

Achtergrond

Het garnalenvirus heeft als officiële naam White spot syndrome virus. Het virus ziet eruit als een ellips met staartje. In de ellips zit een dubbele DNA-spiraal. Het virus lijkt niet op andere virussen en is ingedeeld in een aparte familie. In tien dagen kan het virus een hele garnalenkwekerij te gronde richten. Het virus werd voor het eerst ontdekt in 1992 in Taiwan. Inmiddels komt het virus voor in bijna alle streken waar reuzengarnalen gekweekt worden. Het veroorzaakt veel economische schade.

Promovendus Hendrik Marks bevestigde met DNA-onderzoek dat het virus zich waarschijnlijk vanuit Taiwan via Vietnam naar Thailand heeft verspreid. Waarschijnlijk is het virus op meerdere plekken ongeveer gelijktijdig Vietnam binnengekomen.

De viroloog ontdekte verder dat een nieuwe variant van het Thailandvirus sneller ziek maakt dan de oudste Thailandvariant. Toen hij dezelfde garnaal met beide virusvarianten inspoot, won de nieuwe variant het van de oude. Waarschijnlijk komt dat doordat de nieuwe variant wat minder DNA heeft. Daardoor kan de nieuwe variant zich sneller vermenigvuldigen dan de oude en zo drukt de nieuwe variant de oude weg.

Een op het oog gezonde reuzengarnaal. Afbeelding in hoge resolutie beschikbaar bij de onderzoekers.	Elektronenmicroscopische foto van het garnalenvirus. Afbeelding in hoge resolutie beschikbaar bij de onderzoekers.
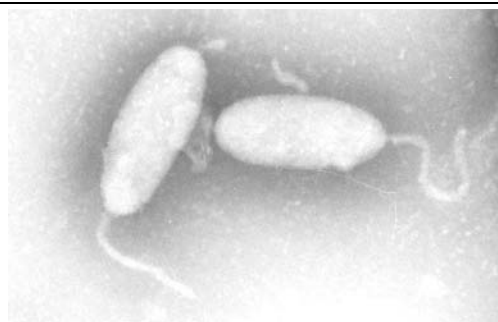
PE&RC PhD Education Statement Form

With the educational activities listed below the PhD candidate has complied with the educational requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 22 credits (= 32 ECTS = 22 weeks of activities)
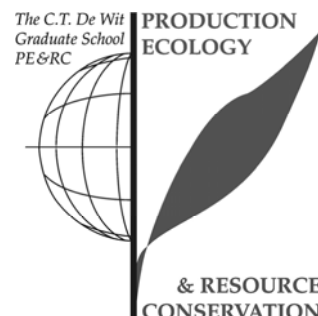
Review of Literature (3 credits)
Transcriptional analysis of White spot syndrome virus (WSSV) genes in
    Penaeus monodon using DNA microarrays (2001)

Writing of Project Proposal (3 credits)
Transcriptional analysis of White spot syndrome virus (WSSV) genes in
    Penaeus monodon using DNA microarrays (2001)

Post-Graduate Courses (4 credits)
Transport processes in health and disease (2000)
Fish vaccination (2002)
Fish immunology (2003)
All two day meetings of the "NWO-CW study group nucleic acids" (2001-2004)

Deficiency, Refresh, Brush-up and General Courses (3 credits)
Veilig werken met radioactieve stoffen en stralingsbronnen (2001)
Techniques for writing and presenting scientific papers (2002)
Career orientation (2004)

PhD Discussion Groups (3 credits)
Weekly work discussion meetings and monthly literature survey meetings with PhD students,
    undergraduate students and staff of the Laboratory of Virology (2001-2005)

PE&RC Annual Meetings, Seminars and Introduction Days (1 credit)
PE&RC annual meeting: "Food insecurity" (2001)
PE&RC annual meeting: "Ethics in science" (2002)
PE&RC annual meeting: "Global climate change and biodiversity" (2003)
PE&RC annual meeting: "Biological disasters" (2004)

International Symposia, Workshops and Conferences (7 credits)
34th Annual Meeting of the Society for Invertebrate Pathology, Noordwijkerhout, The Netherlands (2001)
21st Annual Meeting of the American Society for Virology, Lexington, USA (2002)
23rd Annual Meeting of the American Society for Virology, Montreal, Canada (2004)
24th Annual Meeting of the American Society for Virology, Pennsylvania, USA (2005)
All "Dutch Annual Virology Symposia", Utrecht University, The Netherlands (2002-2005)
All NWO "persdagen", Amsterdam, The Netherlands (2002-2004)
Various one-day workshops in the field of Virology and Microarray (Genomics momentum,
    large dsDNA virus symposium, veterinary viruses, etc.) (2001-2005)

Laboratory Training and Working Visits (1.25 credits)
WSSV promotor studies in cell lines; Laboratory of comparative Immunology/Hematology,
    Uppsala, Sweden (Laboratory of Prof. dr. Söderhäll; 2004)