Journal of General Virology (2005), **86**, 2081–2100 DOI 10.1099/vir.0.80895-0

Gene-expression profiling of *White spot syndrome virus in vivo*

Hendrik Marks,[1] Oscar Vorst,[2] Adèle M. M. L. van Houwelingen,[2] Mariëlle C. W. van Hulten[1]† and Just M. Vlak[1]

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

[2]Plant Research International, Postbus 16, 6700 AA Wageningen, The Netherlands

Correspondence
Just M. Vlak
just.vlak@wur.nl

White spot syndrome virus, type species of the genus *Whispovirus* in the family *Nimaviridae*, is a large, double-stranded DNA (dsDNA) virus that infects crustaceans. The genome of the completely sequenced isolate WSSV-TH encodes 184 putative open reading frames (ORFs), the functions of which are largely unknown. To study the transcription of these ORFs, a DNA microarray was constructed, 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 reached maximal expression at 20 h post-infection (p.i.) (putative early) and 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, the transcriptomes of the WSSV isolates WSSV-TH and TH-96-II, which have differential virulence, were compared at 2 days p.i. The TH-96-II genome encodes 10 ORFs that are not present in WSSV-TH, of which at least seven 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 by using a genome-wide approach provides a better understanding of WSSV transcription regulation and a new tool to study WSSV gene function.

Received 14 January 2005
Accepted 4 April 2005

INTRODUCTION

White spot syndrome virus (WSSV), a member of the virus family *Nimaviridae* (genus *Whispovirus*), is a large, enveloped virus that infects 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 spread quickly (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 double-stranded DNA (dsDNA) genome varies from 293 to 307 kb (van Hulten *et al.*, 2001a; Yang *et al.*, 2001; GenBank accession no. AF440570). The completely sequenced isolate WSSV-TH has a genome size of 292 967 bp. The genome encodes 184 putative ORFs, of which only 6 % could be assigned a putative function based on homology with sequences in public databases. Of the remaining ORFs, five major and approximately 40 minor structural virion-protein genes have been identified (van Hulten *et al.*, 2001a; Huang *et al.*, 2002; Tsai *et al.*, 2004). Additionally, three 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 has focused mainly on WSSV genes that showed homology to known genes, such as the ribonucleotide reductases (Tsai *et al.*, 2000a), the chimeric thymidine kinase–thymidylate kinase (*TK-TMK*; Tsai *et al.*, 2000b), the DNA polymerase

†Present address: CSIRO Livestock Industries, 306 Carmody Road, St Lucia 4067, Brisbane, Australia.

(Chen *et al.*, 2002b), a protein kinase (*PK*) (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 subjected to transcriptional analysis (Marks *et al.*, 2003; Tsai *et al.*, 2004). 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 used successfully to study gene expression of large, dsDNA viruses, such as herpesviruses (Chambers *et al.*, 1999; Stingley *et al.*, 2000; Paulose-Murphy *et al.*, 2001; Ebrahimi *et al.*, 2003), the myovirus bacteriophage T4 (Luke *et al.*, 2002) and the baculoviruses *Autographa californica multiple nucleopolyhedrovirus* and *Bombyx mori nucleopolyhedrovirus* (Yamagishi *et al.*, 2003; Iwanaga *et al.*, 2004). By 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 analysed.

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.*, 2005). 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 with TH-96-II. The ~13·2 kb fragment encompasses 10 ORFs (Marks *et al.*, 2005). By 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.* (2005) also demonstrated a higher virulence in *P. monodon* of WSSV-TH compared with TH-96-II. To correlate this biological feature with differential WSSV gene expression, the complete WSSV transcriptome at 2 days post-infection (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.

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.* (2005), respectively. *P. monodon* (approx. 35 g) or *A. leptodactylus* (approx. 35 g) 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 selected randomly, frozen in liquid nitrogen and stored at −80 °C.

Poly(A)$^+$ RNA isolation. Total RNA was isolated from gills as described previously (Marks *et al.*, 2003). For each time point p.i., 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 that could be sufficient for polyadenylation (e.g. AUUAAA), within −50 to 300 bp downstream of their translational stop codons (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-labelling methods. Poly(A)$^+$ RNA was purified by 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 spectrophotometer (NanoDrop Technologies).

Construction of 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 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 encoded almost completely within the coding regions of ORF13 and ORF109, respectively. For ORF68 and ORF139, we failed to obtain the respective DNA fragments (reason unknown). 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 gene was spotted on the microarrays (the 3′-end probes were used for quantification purposes, the 5′-end probes as controls only). For each WSSV *hr*, one probe was included on the microarrays. Probes to detect the 10 TH-96-II-specific ORFs (CN-ORFs in 'variable region ORF23/24'; Marks *et al.*, 2005) were also included on the WSSV microarrays. For each WSSV ORF, overlap of the corresponding DNA fragment that was selected for use as a specific probe with (5′/3′ untranslated regions of) neighbouring ORFs was avoided. In addition, probes of the following sources were included on the microarrays: (i) 16 cellular shrimp genes, to evaluate normalization between the several chips; (ii) a set of background controls, consisting of four genes of the plant *Medicago truncatula* and the jellyfish *Aequorea victoria* green fluorescent protein (GFP) gene, each spotted in quintuplicate; (iii) the complete coding sequence of the firefly luciferase gene and three partial luciferase clones encompassing the 5′, middle and 3′ parts of the gene, all spotted in quadruplicate (Table 1). As the samples were spiked with luciferase mRNA prior to labelling, this allowed correction of the expression ratios between samples for differences in the preparation of labelled cDNA and for differences during hybridization to the microarrays. The partial luciferase clones were additionally used to monitor the integrity of the labelled sample cDNA. Each of the total of 272 probes was printed in duplicate on each microarray slide, to evaluate the consistency of the signals obtained.

Preparation of probes. Parts of the DNA fragments used as WSSV-TH probes were obtained by performing PCRs 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 parts of the WSSV-TH probes and all TH-96-II-specific probes (CN-ORFs) were obtained by performing PCRs on genomic DNA of WSSV-TH and TH-96-II, respectively, using specific primers designed by using PrimeArray (Raddatz *et al.*, 2001). The cellular shrimp genes, negative controls and luciferase controls were obtained by performing PCRs on purified plasmids containing the anticipated fragments using universal primers. The shrimp genes were originally amplified from a cDNA library of uninfected *P. monodon* that was available in our laboratory by using specific primers (sequences obtained from GenBank) and cloned into the

Table 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.

WSSV-TH ORFs [numbering according to WSSV-TH (GenBank accession no. AF369029; van Hulten *et al.*, 2001a)]							
ORF no.	**Putative function/name***	**Position and direction of ORF**	**ORF length (nt)**	**Position of probe**	**Length of probe (nt)**	**Part of gene detected**	**Not detected/ remark**
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		×
6	In virion (vp800)	10834→13236	2403	12122–13206	1085	3′	×
				10882–12162	1281	5′	×
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′	×
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		×
***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		×
27	DNA polymerase	35571→42626	7056	41523–42614	1092	3′	
				36486–37541	1056	5′	×
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′	×
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		

Table 1. cont.

WSSV-TH ORFs [numbering according to WSSV-TH (GenBank accession no. AF369029; van Hulten *et al.*, 2001a)]							
ORF no.	**Putative function/name***	**Position and direction of ORF**	**ORF length (nt)**	**Position of probe**	**Length of probe (nt)**	**Part of gene detected**	**Not detected/ remark**
38		56598←57458	861	56859–57458	600		
39		57509←58204	696	57509–58204	696		
40		58285←62892	4608	58842–59835	994	3′	×
				61727–62726	1000	5′	×
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′	×
43		69737→72682	2946	71311–72456	1146	3′	
				70137–71111	975	5′	
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		×
53		77284←79815	2532	77972–78957	986	3′	
				78212–79406	1195	5′	×
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		×
63		90298→90744	447	90298–90597	300		×
64		90669→91046	378	90794–90983	190		×
65		91003→94443	3441	93170–94366	1197	3′	
				91349–92386	1038	5′	×
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		×
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		

Table 1. cont.

WSSV-TH ORFs [numbering according to WSSV-TH (GenBank accession no. AF369029; van Hulten *et al.*, 2001a)]							
ORF no.	**Putative function/name***	**Position and direction of ORF**	**ORF length (nt)**	**Position of probe**	**Length of probe (nt)**	**Part of gene detected**	**Not detected/ remark**
78		110964→111779	816	110966–111563	598		×
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′	×
				116024–117158	1135	5′	×
83		117465←117878	414	117465–117878	414		×
84		118025→124969	6945	123799–124845	1047	3′	
				118532–119333	802	5′	×
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′	×
90		132697←134976	2280	133013–134063	1051	3′	
				133799–134963	1165	5′	×
91		135031←138249	3219	135105–136210	1106	3′	
				137218–138203	986	5′	×
92	Ribonucleotide reductase (large subunit)	138330←140876	2547	138635–139536	902	3′	
				139437–140552	1116	5′	×
***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		×
97		144689→146314	1626	145323–146325	1003		×
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′	×
***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		×
107		158204→159031	828	158204–159031	828		
108		159076←163896	4821	159947–160966	1020	3′	×
				161135–162021	887	5′	×
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′	

Table 1. cont.

WSSV-TH ORFs [numbering according to WSSV-TH (GenBank accession no. AF369029; van Hulten *et al.*, 2001a)]							
ORF no.	**Putative function/name***	**Position and direction of ORF**	**ORF length (nt)**	**Position of probe**	**Length of probe (nt)**	**Part of gene detected**	**Not detected/ remark**
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′	×
114		172701→175511	2811	174142–175268	1127	3′	
				173065–174191	1127	5′	×
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		×
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		×
134		199492→203151	3660	201762–202795	1034	3′	
				200650–201747	1098	5′	×
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		×
139		207992←208159	168	No probe			
140		208153→210057	1905	208896–209941	1046		×
141		210064→210366	303	210064–210366	303		×
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′	×
144		218566←218859	294	218660–218859	200		×
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		

Table 1. cont.

WSSV-TH ORFs [numbering according to WSSV-TH (GenBank accession no. AF369029; van Hulten *et al.*, 2001a)]							
ORF no.	**Putative function/name***	**Position and direction of ORF**	**ORF length (nt)**	**Position of probe**	**Length of probe (nt)**	**Part of gene detected**	**Not detected/ remark**
149	TATA box-binding protein/in virion (vp184)	221977→224652	2676	222471–223725	1255	3′	×
				222186–223166	981	5′	×
150		224639→225898	1260	224637–225859	1223		×
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′	×
				230510–231248	739	5′	×
155		232928→233281	354	232928–233246	319		×
156		233295→233978	684	233295–233492	198		
157		233982←234230	249	234007–234204	198		
158		234229→235626	1398	235008–236108	1101		×
***hr*8**							
159		237222←239792	2571	237578–238307	730	3′	
				238145–239420	1276	5′	×
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		×
163		244556←245341	786	244556–245341	786		×
164		245444←245746	303	245444–245746	303		
165		245849←250966	5118	246582–247516	935	3′	
				247556–248795	1240	5′	×
166		251400←258392	6993	251676–252663	988	3′	×
				252648–253699	1052	5′	×
167		258666→276899	18234	275586–276750	1165	3′	
				258772–259853	1082	5′	×
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		
179		288183←288866	684	288080–289064	985		
180		289149←289343	195	289149–289343	195		×
181		289474→289680	207	289474–289680	207		×
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		

Table 1. cont.

TH-96-II ORFs [numbering according to WSSV-CN (GenBank accession no. AF332093; Yang *et al.*, 2001)]						
ORF no.‡	**Putative function/name***	**Position and direction of ORF**	**Length of ORF (nt)**	**Position of probe**	**Length of probe (nt)**	**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	×
CN-489		281128←281865	738	281266–281865	600	×
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	×
CN-500		286077←286706	630	286122–286706	585	

***hr*s [numbering according to WSSV-TH (GenBank accession no. AF369029)]**					
***hr* no.**	**Position of *hr***	**Length of *hr* (nt)**	**Position of probe**	**Length of probe (nt)**	**Not detected**
*hr*1	28250–30320	2071	29243–30328	1086	×
*hr*2	73550–77150	3601	73526–74259	734	×
*hr*3	126388–128112	1725	126852–128093	1242	×
*hr*4	141139–141827	689	140518–141843	1326	×
*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	×

Shrimp genes and negative controls		
ORF	**(Similar to) GenBank accession no.**	**Length of probe (nt)**
Shrimp genes		
Actin	AF100986	686
Elongation factor 1-α	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		
GFP (*A. victoria*)	M62653	966
Lyk3 (*M. truncatula*)	AY372406	550
Nork-i (*M. truncatula*)	AJ418369	500
Enod 12 (*M. truncatula*)	X68032	500
HCR4/Cf-4/9 (*M. 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.* (2002).

pGEM-T Easy vector (Promega). Sequencing confirmed that these clones contained the anticipated sequences. All PCRs were performed by using the Expand Long Template PCR system (Roche).

Preparation of WSSV microarrays. PCR products were column-purified by using a High Pure PCR product purification kit (Roche) and diluted to 0·1 μg μl^{-1} (in a total volume 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. A 10 μg aliquot of each PCR product was dried to completion and dissolved in 10 μl 5× SSC (sodium citrate/sodium chloride) buffer. 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 (95–100 °C) and UV-cross-linked (150 mJ). The slides were soaked twice in 0·2 % SDS (2 min), twice in MilliQ (Millipore)-treated water (MQ water) (2 min) and placed into boiling MQ water (2 min). After drying (5 min), the slides were rinsed three times in 0·2 % SDS (1 min), once in MQ water (1 min), submerged in boiling MQ water (2 s) and air-dried.

Synthesis of Cy3/Cy5-labelled cDNA. Purified poly(A)$^+$ RNA preparations were labelled by using a standard protocol for cDNA synthesis: 2·5 μg *P. monodon* poly(A)$^+$ RNA [spiked with 1 ng luciferase mRNA (Promega); for *A. leptodactylus* samples, 1 μg poly(A)$^+$ RNA was used] and 2·5 μg $(dT)_{21}$ primer were heated to 65 °C (3 min) and then placed at 25 °C (10 min) to anneal the primer. RNA was reverse-transcribed by using Superscript II (Invitrogen) according to the protocol of the manufacturer, with the exception that 40 % of the total dTTP was replaced with 5-(3-aminoallyl)-2′dUTP. After precipitation and washing of the cDNA–RNA hybrids, the RNA was hydrolysed with 0·2 M NaOH at 37 °C (10 min). The solution was neutralized with 0·15 M HEPES (pH 6·8) 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 bound covalently to the incorporated amino groups by adding 5 μl 5 mM Cy3 or Cy5 reactive dyes (Amersham Biosciences) in DMSO. This mixture was incubated at room temperature in the dark (1 h). Unincorporated dye was removed by performing ethanol precipitation twice, after which the labelled cDNA was dissolved in 5 μl MQ water. Samples under study were labelled with Cy3 and the reference sample was labelled with 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 p.i. from WSSV-TH-infected, as well as from TH-96-II-infected, shrimp and crayfish gill tissue, providing additional 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, 5× Denhardt's reagent, 5× SSC, 0·2 % SDS, 0·1 mg fish DNA ml^{-1}), the slides were washed by dipping in MQ water, followed by dipping in 2-propanol. Slides were dried by centrifugation at 160 ***g*** (1 min). Hybridization occurred in a volume of 65 μl, using a covered hybridization frame [Gene Frame 15 mm × 15 mm (65 μl); ABgene]. After heating 65 μl hybridization buffer containing both Cy3- and Cy5-labelled cDNA samples to 95 °C for 1 min, it was loaded into the hybridization chamber. The slides were hybridized for 24 h at 42 °C. Following hybridization, the slides were washed in 1× SSC/0·1 % SDS (5 min), 0·1× SSC/0·1 % SDS (5 min) and rinsed briefly in 0·1× SSC. Slides were dried by centrifugation at 160 ***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 processed individually. For each array element, the integrated OD was determined within a defined circle, using AIS software (Imaging Research). Mean background values, calculated from the hybridization signals of the *M. truncatula* and *A. victoria* probes, were subtracted to correct for non-specific fluorescence. Next, Cy3 or Cy5 signals not reaching 0·5× background value were set to this cut-off (0·5× background value) for the respective dye. Elements for which neither the Cy3 and Cy5 signal reached 0·5× 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 mean 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× the expression value obtained initially from the Cy3 background-control probes. WSSV genes with a lower Cy3 expression value (below the background threshold) were considered to be not detected.

Analysis of expression data. For WSSV genes, mean 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 (five probes; Table 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 by using GeneMaths software (Applied Maths). For individual shrimp genes, the ^{2}log expression ratios were normalized by subtracting the mean 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 mean expression of the 16 shrimp genes (Table 1) over the time course at 100 %.

RESULTS

To evaluate WSSV gene expression during infection in shrimp, a viral microarray was designed that contained 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.*, 2005). The WSSV genome contains nine *hr*s, each consisting of 250 bp repeats with a nucleotide identity of 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 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 Cy3-labelled and

mixed with a Cy5-labelled 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 an individual microarray. Fig. 1(a) shows images of parts of the microarrays that were constructed after the scanning.

Microarray controls

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. 1(b) shows that the ratios of the duplicates were very similar. The mean 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 those 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. 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 the time point at 0 h p.i., the difference in expression of the 16 selected shrimp genes is within 20 % of the mean expression (set at 100 %) between the several time points. The expression of some of the individual cellular genes shown in Fig. 1(c) 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 of WSSV genes

Detectable WSSV genes. Pilot experiments revealed that the probes of *M. truncatula* and *A. victoria*, used as background controls, showed a minimal signal compared with the other probes on the WSSV microarray when labelled cDNA of WSSV-infected shrimp was hybridized to the microarrays. Except for some false-positive signals (see Table 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-infection 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 1). The ORFs corresponding to the WSSV probes that 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 1). These *hr*s were included in our further analysis.

Transcription profiling. The Cy3/Cy5 ratios of the 125 WSSV genes that we could detect specifically were normalized to the maximal expression of each gene (100 %). The constructed transcription profiles are presented in Table 2. To examine the relationship between the WSSV genes, these relative expression data were analysed by hierarchical clustering (Euclidean distance). Fig. 2(a) 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. 2(b and c), respectively.

After the genes of the 'yellow' cluster reached (almost) a maximal 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. 2b). Most WSSV putative-early genes, such as the ribonucleotide reductases (*rr1* and *rr2*), the chimeric *tk-tmk* and both *pk* genes, which are considered to be expressed before viral DNA replication, were present within this group (Fig. 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. 2a). Structural virion-protein genes are supposed to be expressed after virus replication and are 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 maximal expression (100 %) of each particular gene (Fig. 3; Table 2). These absolute gene-expression levels should be interpreted semi-quantitatively, as the DNA fragments spotted on the microarray are different in length and G+C 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 labelled 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 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 2). Fig. 3 shows that putative early- and late-type genes, as

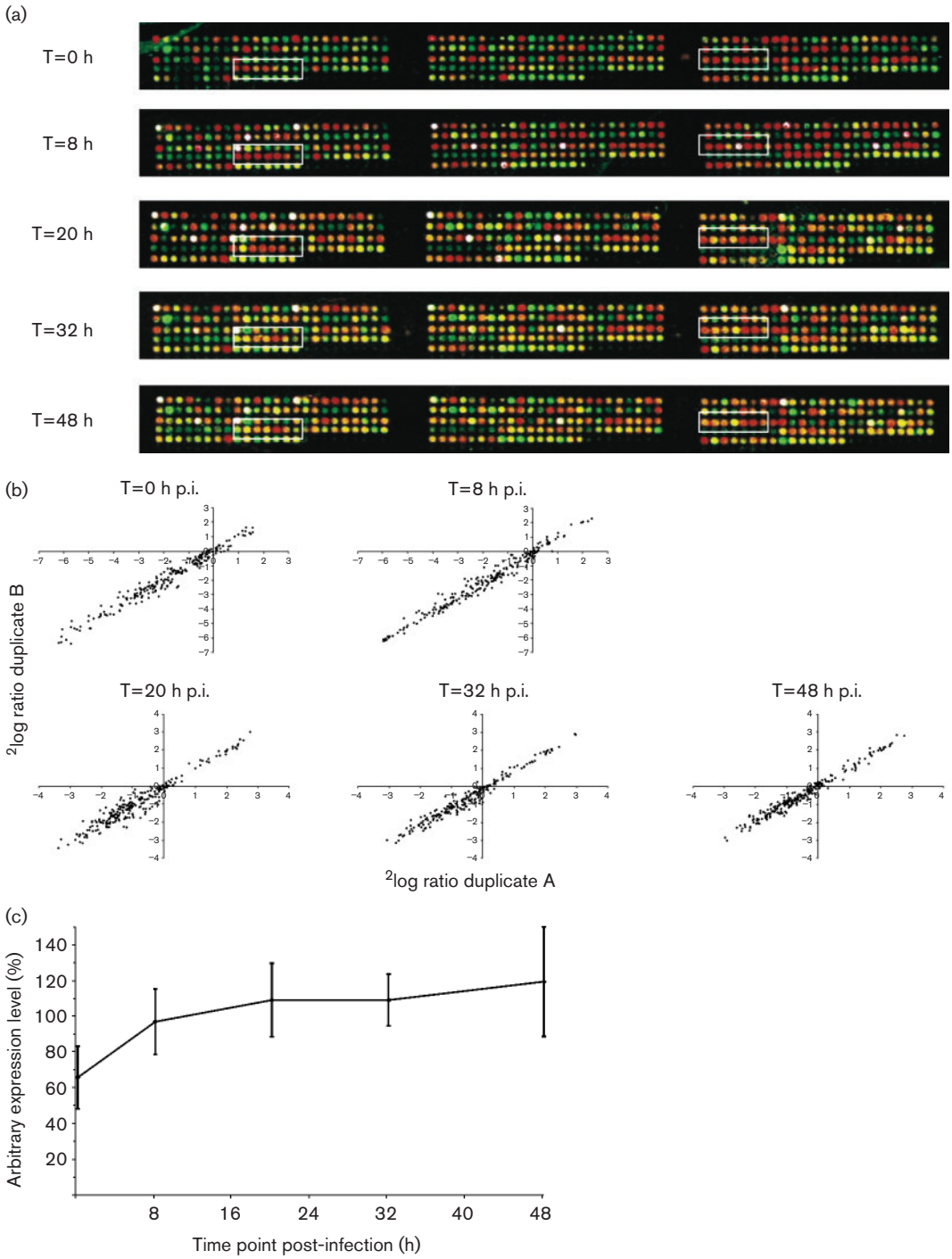

Fig. 1. (a) Pseudocolour microarray images (Cy3 and Cy5 are coloured green and red, respectively) of one of the on-array duplicate sets of probes. An equal mix of red and green results in a yellow pseudocolour; other ratios result in intermediate colours. Probes for WSSV and shrimp genes, as well as background and luciferase controls, are distributed randomly over the array. For probes representing WSSV genes (examples are boxed) that are red at 0 and 8 h post-WSSV infection, a yellowish colour 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. (b) Scatter plot 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 $R^2>0{\cdot}93$, indicating that the duplicate sets of results are very similar. (c) Expression level (%) of the 16 shrimp genes shown in Table 1 (mean expression over the time course is set to 100 %; SD is indicated for each time point).

Table 2. (Relative) expression levels of the WSSV ORFs on the microarray, sorted by intensities of the probes

ORF no.*	Putative function/name†	Time-course expression level (% of maximal expression)					Intensity‡	TATA box§	Consensus poly(A) signal§	Type‖
		T=0 h	T=8 h	T=20 h	T=32 h	T=48 h				
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 thymidine kinase–thymidylate 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	Ribonucleotide reductase (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

Table 2. cont.

ORF no.*	Putative function/name†	Time-course expression level (% of maximal expression)					Intensity‡	TATA box§	Consensus poly(A) signal§	Type‖
		T=0 h	T=8 h	T=20 h	T=32 h	T=48 h				
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
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

Table 2. cont.

ORF no.*	Putative function/name†	Time-course expression level (% of maximal expression)					Intensity‡	TATA box§	Consensus poly(A) signal§	Type‖
		T=0 h	T=8 h	T=20 h	T=32 h	T=48 h				
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	Ribonucleotide reductase (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.* (2002).

‡Semi-quantitative levels of gene expression (intensity) were categorized by using the following classification (numbers are arbitrary expression units): background threshold to 3000, +−−; 3000–5000, +−; 5000–20 000, +; 20 000–40 000, ++; 40 000 or higher, +++.

§0, Not present; 1, present. A TATA box was considered to be present if the sequence TATA(a/t)A appeared 0–300 nt upstream of the translational start codon; a poly(A) signal was considered to be present if the sequence A(a/t)TAAA appeared within −50 to 300 nt downstream of the translational stop codon.

‖According to Fig. 2(a).

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 correlated positively with its temporal expression class or expression level (Table 2).

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′ ends. Of the 42 5′-end probes, 28 (67 %) did not give a signal above the background threshold during the time course, whereas, for the 3′-end probes, only eight (19 %) could not be detected. This is probably caused by the reverse-transcriptase 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 that are at least 1·5–2 kb upstream of the poly(A) tail, these 5′ ends of the mRNA were probably not reverse-transcribed as efficiently. For the ORFs of which both 5′ and 3′ ends

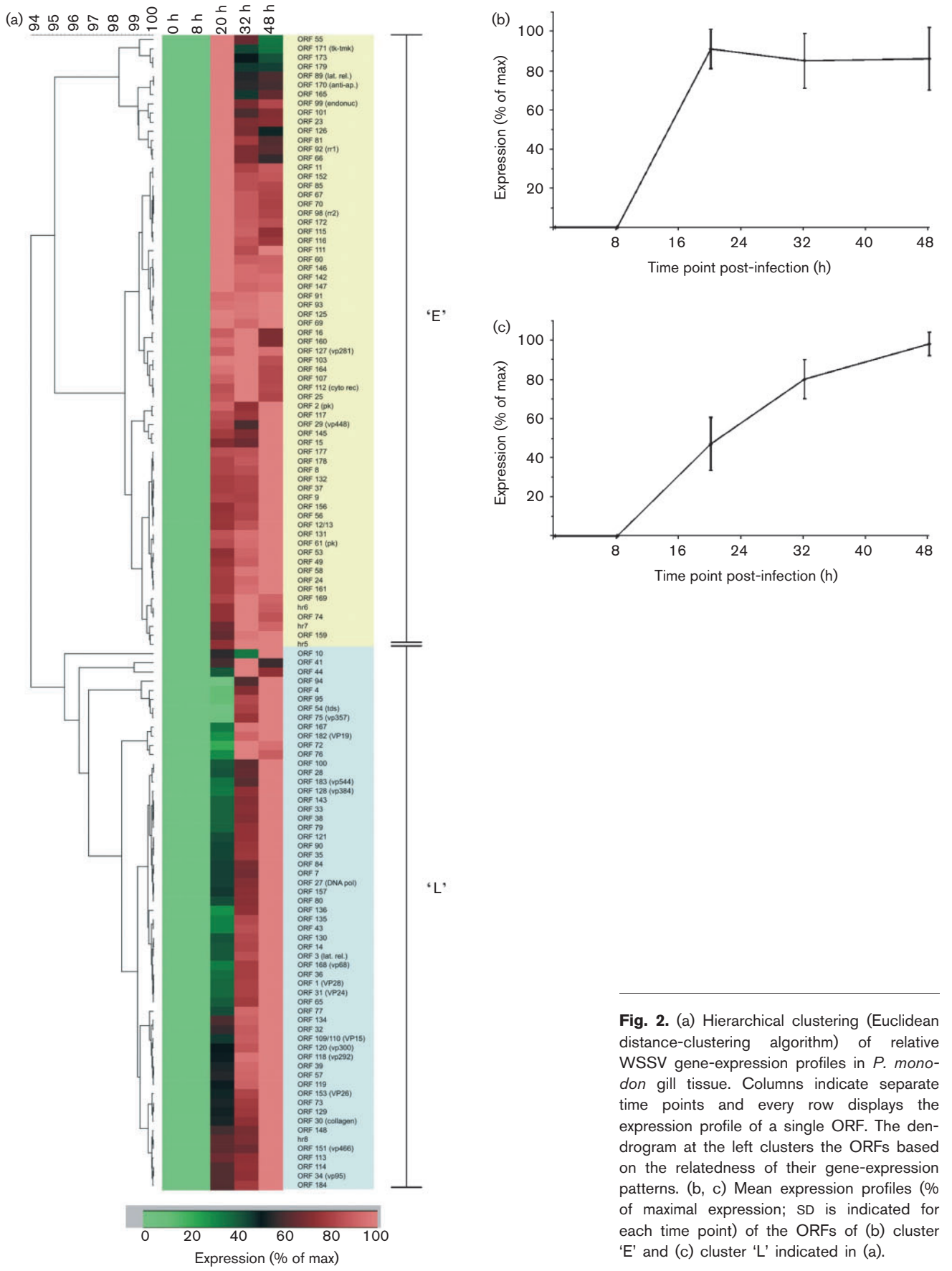

Fig. 2. (a) Hierarchical clustering (Euclidean distance-clustering algorithm) of relative WSSV gene-expression profiles in *P. monodon* gill tissue. Columns indicate separate time points and every row displays the expression profile of a single ORF. The dendrogram at the left clusters the ORFs based on the relatedness of their gene-expression patterns. (b, c) Mean expression profiles (% of maximal expression; SD is indicated for each time point) of the ORFs of (b) cluster 'E' and (c) cluster 'L' indicated in (a).

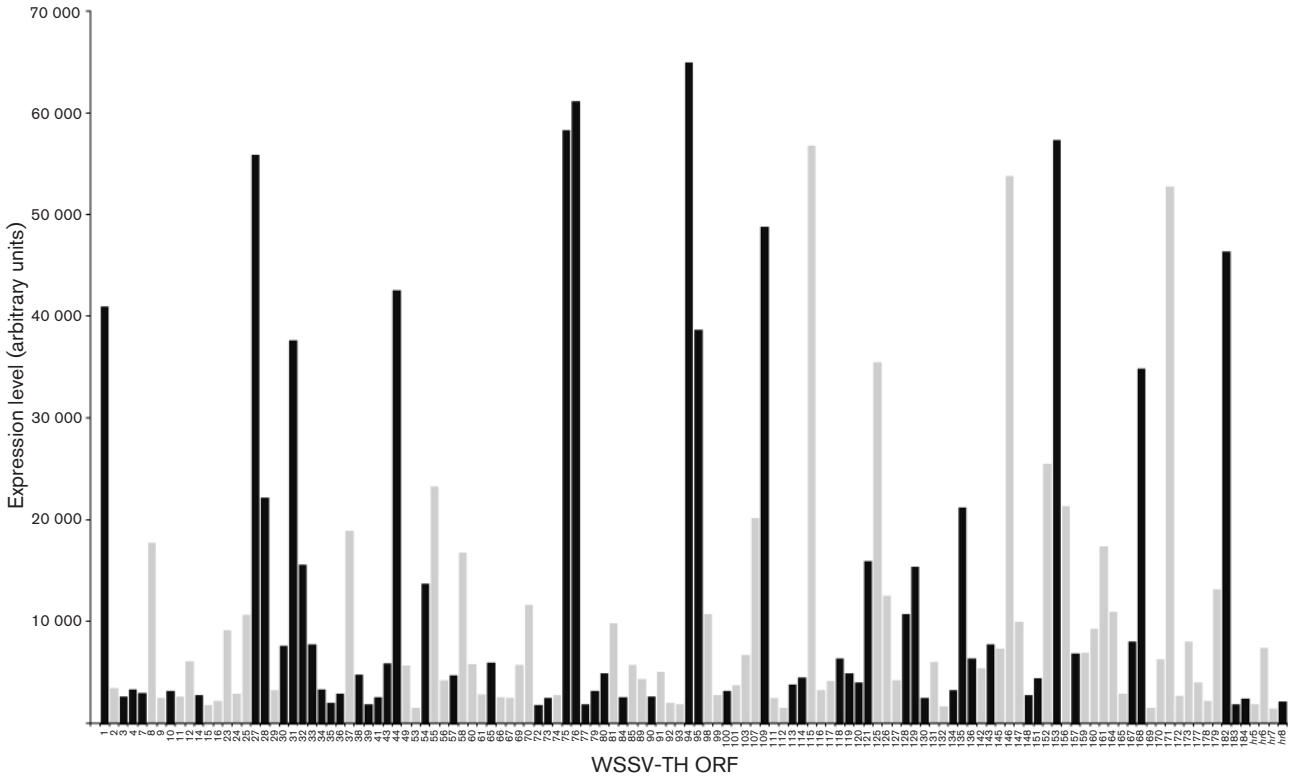

Fig. 3. Semi-quantitative expression levels (arbitrary units) of WSSV genes. Bars representing genes of cluster 'E' are coloured grey; bars representing genes of cluster 'L' are coloured black.

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 analyses by van Hulten *et al.* (2001a) and Yang *et al.* (2001), which are very similar (Marks *et al.*, 2004).

Comparison of the transcriptomes of 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.*, 2005). Probes for the 10 genes encoded by the ~13·2 kb fragment of TH-96-II were included on our WSSV microarray (Table 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 10 probes, excluding the possibility of (non-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, labelled 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 mean was used for further analysis. The scatter plot in Fig. 4 shows a comparison between the mean 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 twofold expression difference or more are marked.

For the TH-96-II-infected *P. monodon* tissue, we could detect seven of the 10 genes encoded by 'variable region ORF23/24' (Table 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 by ++ in Table 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–r) appear to be expressed highly by TH-96-II compared with WSSV-TH in the graphs of Fig. 4. ORF12/13 (as these two genes overlap completely, they cannot be distinguished by using our microarray), ORF14, ORF23, ORF24 (both 5′ and 3′ probes) and ORF85 were present at around $2\times$ excess in the TH-96-II-infected *P. monodon* gill tissue, whilst the only gene that showed a relative expression of $>2\times$ in WSSV-TH-infected tissue was ORF77 (Fig. 4).

A similar comparison between WSSV-TH and TH-96-II in the crayfish *A. leptodactylus* (Fig. 4) could detect the same seven of the 10 genes encoded by 'variable region ORF23/24'

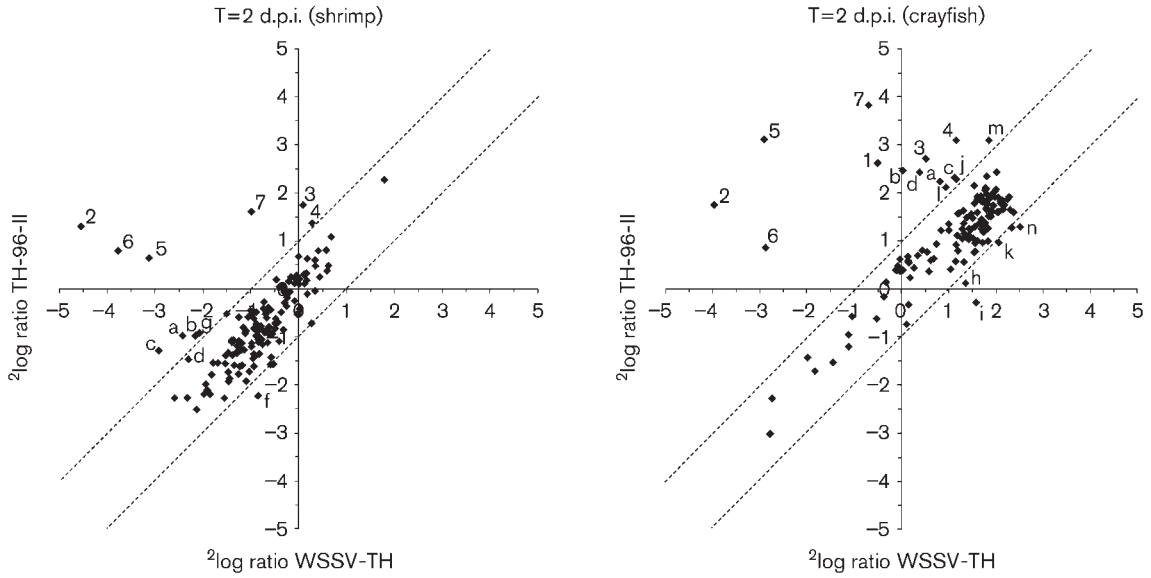

Fig. 4. Comparison of the transcriptomes of WSSV-TH (*x* axis) and TH-96-II (*y* axis) at 2 days post-infection (d.p.i.) in shrimp and crayfish. Scatter plots of the expression levels obtained for WSSV-TH and TH-96-II. The dotted lines indicate a twofold difference in expression between the isolates. The labelled 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.

in TH-96-II-infected tissue as were detected for *P. monodon*. Also, ORF12/13, ORF14 and ORF24 (5′ and 3′ probes) were present $>2\times$ in excess in the TH-96-II-infected *A. leptodactylus* gill tissue (Fig. 4). Furthermore, ORF127, ORF145 and ORF170 were expressed $>2\times$ in excess in TH-96-II-infected tissue, whilst ORF44, ORF118, ORF131 and ORF178 showed a relative expression of $>2\times$ in WSSV-TH-infected tissue (Fig. 4).

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 %) by using RT-PCR on a cDNA cocktail, whilst Tsai *et al.* (2004) detected 39 minor virion-protein genes by using a small, single dye (Cy3)-based microarray. By 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 1), indicating that most WSSV predicted ORFs are transcriptionally active. However, as we could not detect strand-specific mRNAs 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 baculovirus *hr*s, which have a structure similar to that of 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 1). It is possible that these genes are not expressed in gills, but it is more likely that it is caused by the detection limits of our microarray experiment. About half of the ORFs that were not detected encode a consensus poly(A) signal, indicating that 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-PCRs, 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. 2a). The transcription profiles (Fig. 2b and c) 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 the case of baculoviruses, transcription of early genes declines when late-gene transcription becomes very high, as the downregulation of early transcription is due directly or indirectly to late-gene expression or viral DNA replication. The fact that the 'putative-early' cluster of WSSV does not show this decline (Fig. 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, b, 2004; Liu *et al.*, 2001; Chen *et al.*, 2002b; Marks *et al.*, 2003; Li *et al.*, 2004a, b). To obtain further support, future microarray studies could include the testing of different inhibitors, such as cycloheximide and phosphonoacetic acid, to distinguish between gene expression before and after protein synthesis and virus 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. 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 the case of large, dsDNA viruses, such as herpesviruses (Ebrahimi *et al.*, 2003) and baculoviruses (Lu & Miller, 1997), late genes are often transcribed abundantly. 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 2). 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 that are highly polymorphic between several WSSV isolates (Marks *et al.*, 2004; Dieu *et al.*, 2004), show high expression levels (Table 2). The protein encoded by ORF75 is probably located in the virion (vp357; Huang *et al.*, 2002). 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 *tk-tmk* 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 of transcriptomes 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', that are both absent in WSSV-TH (Marks *et al.*, 2005). The genes encoded by the additional fragments are dispensable for infection and replication in these species, as both are permissive hosts 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. 4), some ORFs were expressed differentially ($>2\times$ difference). The expression of ORF12/13, ORF14, ORF23 and ORF24 was higher for TH-96-II- than for WSSV-TH-infected gill tissue, both in *P. monodon* and in *A. leptodactylus* (except for ORF23, which was only higher in *P. monodon*). These genes are located at the junction sites of 'variable region ORF14/15' and 'variable region ORF23/24', but are 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, at least in *P. monodon*, as a WSSV isolate lacking ORF14 (WSSV-TW; GenBank accession 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 encoding 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 probably have an essential function during virus infection. Recent data suggest a higher virulence of WSSV-TH compared with TH-96-II in *P. monodon* (Marks *et al.*, 2005). Although the difference in virulence could be explained by a replication advantage of WSSV-TH, which has a smaller-sized genome, the ORFs expressed differentially between WSSV-TH and TH-96-II in *P. monodon* could also play a role. In this respect, ORF12/13, ORF23 and ORF24, but also ORF85, which are all expressed at a lower level in WSSV-TH-infected *P. monodon* tissue, and ORF77, the only gene present $>2\times$ in excess in WSSV-TH-infected tissue (Fig. 4), are of interest.

Evaluation of *in vivo* transcription profiles by 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 virus replication and gene expression and, consequently, produce information that can aid in the development of effective treatments against WSSV.

ACKNOWLEDGEMENTS

We thank Professor Dr Rob Goldbach, Dr Douwe Zuidema, Dr Marcel Westenberg, Jeroen Witteveldt and Angela Vermeesch for stimulating discussions and their help in the course of these experiments. Roel Staps is acknowledged for his assistance during the hybridizations and scanning of the microarrays. We thank Dr Erik Limpens and Joop Arts for providing us with the *M. truncatula* probes and the shrimp probes, respectively. This work was supported by Intervet International BV, Boxmeer, the Netherlands.

REFERENCES

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

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. & 9 other authors (1999). DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. *J Virol* **73**, 5757–5766.

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., Wang, H.-C., Huang, C.-J. & 9 other authors (2002b). Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus. *Virology* **301**, 136–147.

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.

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.

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.

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.

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

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.

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.

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.

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* (Special Publication of the World Aquaculture Society). Baton Rouge, LA: World Aquaculture Society.

Liu, W.-J., Yu, H.-T., Peng, S.-E. & 10 other authors (2001). Cloning, characterization, and phylogenetic analysis of a shrimp white spot syndrome virus gene that encodes a protein kinase. *Virology* **289**, 362–377.

Lo, C. F., 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.

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

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.

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., 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., 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.

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

Paulose-Murphy, M., Ha, N.-K., Xiang, C. S. & 7 other authors (2001). Transcription program of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus). *J Virol* **75**, 4843–4853.

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.

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

Rosenberry, B. (2002). *World Shrimp Farming 2002* (Shrimp News International no. 15). San Diego, CA: Shrimp News International.

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.

Stingley, S. W., Ramirez, J. J. G., Aguilar, S. A., Simmen, K., Sandri-Goldin, 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.

Tsai, M.-F., Lo, C.-F., van Hulten, M. C. W. & 7 other authors (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. & 7 other authors (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.

Tsai, J.-M., Wang, H.-C., Leu, J.-H., Hsiao, H.-H., Wang, A. H.-J., 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.

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.

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, 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.

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. & Xu, X. (2001). Complete genome sequence of the shrimp white spot bacilliform virus. *J Virol* **75**, 11811–11820.