Journal of General Virology (2000), **81**, 307–316. Printed in Great Britain

Analysis of a genomic segment of white spot syndrome virus of shrimp containing ribonucleotide reductase genes and repeat regions

Mariëlle C. W. van Hulten,[1] Meng-Feng Tsai,[2] Christel A. Schipper,[1] Chu-Fang Lo,[2] Guang-Hsiung Kou[2] and Just M. Vlak[1]

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

[2] Department of Zoology, National Taiwan University, Taipei, Taiwan, Republic of China

White spot syndrome is a worldwide disease of penaeid shrimp. The disease agent is a bacilliform, enveloped virus, white spot syndrome virus (WSSV), with a double-stranded DNA genome that probably contains well over 200 kb. Analysis of a 12·3 kb segment of WSSV DNA revealed eight open reading frames (ORFs), including the genes for the large (RR1) and small (RR2) subunits of ribonucleotide reductase. The *rr1* and *rr2* genes were separated by 5760 bp, containing several putative ORFs and two domains with multiple sequence repeats. The first domain contained six direct repeats of 54 bp and is part of a coding region. The second domain had one partial and two complete direct repeats of 253 bp at an intergenic location. This repeat, located immediately upstream of *rr1*, has homologues at several other locations on the WSSV genome. Phylogenetic analysis of RR1 and RR2 indicated that WSSV belongs to the eukaryotic branch of an unrooted parsimonious tree and, further, seems to suggest that WSSV and baculoviruses probably do not share an immediate common ancestor. The present analysis of WSSV favours the view that this virus is either a member of a new genus (Whispovirus) within the *Baculoviridae* or a member of an entirely new virus family.

Introduction

White spot syndrome is a worldwide disease of shrimp. The disease has also been observed in other invertebrate aquatic organisms, such as crab and crayfish, and has a major impact on the economy of the shrimp farming industry in South-east Asia (Flegel, 1997). Infected shrimps show strong signs of lethargy and a reddish coloration of the hepatopancreas (Chou *et al.*, 1995). Moribund penaeid shrimps exhibit patchy discolorations or 'white spots' in the exo-mesoderm under the carapace, hence the name 'white spot syndrome'.

The disease agent is a bacilliform, enveloped virus with a rod-shaped nucleocapsid containing double-stranded DNA. The virus particles have a tail- or flagellum-like extension at one end. Its morphology, nuclear localization and morphogenesis are reminiscent of baculoviruses in insects (Durand *et al.*, 1997). The virus is known as hypodermic and hematopoietic necrosis baculovirus (HHNBV), rod-shaped nuclear virus of *Penaeus japonicus* (RV-PJ), systemic ectodermal mesodermal baculovirus (SEMBV), white spot baculovirus (WSBV) and white spot syndrome virus (WSSV) (Lightner, 1996). A species name, whispovirus, has been proposed (Vlak *et al.*, 1999). Although WSSV has baculovirus characteristics and was an unassigned member of the *Baculoviridae* in the past (Francki *et al.*, 1991), at present it is no longer accepted into this family (Murphy *et al.*, 1995) due to the lack of molecular information.

The WSSV genome consists of a double-stranded DNA estimated to be well over 200 kb in size (Yang *et al.*, 1997). The structure of the DNA (linear or circular) is not yet known. We have begun to analyse the WSSV genome in order to investigate its taxonomic status and to understand the genetic basis of its pathology in such a wide range of invertebrate aquatic organisms including penaeid shrimp. In this paper, we describe the identification and analysis of a 12·3 kb segment of

Author for correspondence: Just Vlak.
Fax +31 317 484820. e-mail just.vlak@medew.viro.wau.nl

The accession numbers of the sequences reported in this paper are AF099142 and AF144620.

the WSSV genome. This segment contained two types of tandem repeat sequence regions and contains several putative open reading frames (ORFs), including two encoding the large (RR1) and small (RR2) subunits of ribonucleotide reductase (RR). Genes encoding this enzyme are often found in large DNA viruses, including African swine fever virus (ASFV) (Boursnell *et al.*, 1991), herpesviruses (Willoughby *et al.*, 1997), poxviruses (Schmitt & Stunnenberg, 1988), iridoviruses (Tidona & Darai, 1997), phycodnaviruses (Li *et al.*, 1997) and baculoviruses (van Strien *et al.*, 1997). The genes for RR1 and RR2 are the first ORFs in WSSV DNA to which putative functions have been assigned. Both genes were used to investigate the ancestral relationships of WSSV to other organisms and viruses including baculoviruses.

Methods

■ **Virus production and purification.** The virus used in this study was isolated from infected *Penaeus monodon* shrimps from Thailand. Infected tissue was homogenized in TN buffer (20 mM Tris–HCl, 400 mM NaCl, pH 7·4). After centrifugation at 1700 ***g*** for 10 min, the supernatant was filtered (0·45 μm filter). The filtrate was injected into healthy *P. monodon* in the lateral area of the fourth abdominal segment to initiate an infection. After 3 or 4 days, haemolymph was withdrawn from moribund shrimps and mixed with modified Alsever solution (Rodriguez *et al.*, 1995) as an anticoagulant. After dilution in TNE (20 mM Tris–HCl, 400 mM NaCl, 5 mM EDTA, pH 7·4), the haemolymph was clarified at 1700 ***g*** for 10 min at 4 °C. The virus particles were removed from the supernatant by centrifugation at 45 000 ***g*** at 4 °C for 1 h and resuspended in TN buffer.

Alternatively, virus was isolated from infected shrimp tissue. The tissue was homogenized in TNE by using a mortar and, after clarification at 1700 ***g*** for 25 min at 4 °C, the supernatant was subjected to discontinuous sucrose density gradient centrifugation. The virus was obtained from the 20/45 % sucrose interface and, after dilution in 3 vols TNE, was sedimented at 80 000 ***g*** for 1 h and resuspended in TN. Virus preparations were stored at −20 °C.

■ **Nucleic acid purification.** Viral DNA was isolated from purified virions by treatment with proteinase K (0·2 mg/ml) and Sarkosyl (1 %) at 45 °C for 2–4 h, followed by phenol–chloroform extraction and dialysis against TE (Sambrook *et al.*, 1989). The purity and concentration of the DNA were determined by agarose gel electrophoresis.

■ **Cloning and mapping of WSSV DNA fragments.** WSSV DNA was digested with restriction enzymes (GIBCO BRL) and the fragments were separated by electrophoresis in 0·7 % agarose gels at 45 V (1·5 V/cm) for 14–20 h. Lambda DNA digested with *Bam*HI/*Eco*RI/*Hin*dIII or with *Hin*dIII was used as a molecular size standard. WSSV DNA fragments were inserted into the plasmid vector pBluescript SK(+) by shotgun cloning of digested WSSV DNA by standard techniques (Sambrook *et al.*, 1989). Fragments were purified from agarose gels by using Glassmax (GIBCO BRL) prior to cloning. Southern hybridization of cloned fragments with digested WSSV DNA was performed by standard techniques (Sambrook *et al.*, 1989).

Restriction enzyme analysis of the cloned fragments with various enzymes, Southern blot hybridization and subcloning were employed to construct a physical map of the 11·3 kb *Bam*HI and 4·8 kb *Xho*I WSSV DNA fragments. Sequence data for the subcloned fragments were also used to establish and confirm the detailed map.

We have adopted the conventional baculovirus nomenclature to indicate gene (lower-case italics) and protein (roman capitals) abbreviations.

■ ***In situ* hybridization.** *In situ* hybridization with cloned fragments was carried out on tissue sections of WSSV-infected shrimp (a gift from P. Voorthuis, Laboratory of Fisheries, Wageningen Agricultural University) by using the ShrimProbe protocol (DiagXotics). Preparations of uninfected shrimp were used as negative controls. Cloned fragments were labelled with digoxigenin according to the DIG System user's guide for filter hybridization (Boehringer Mannheim). The hybridization was visualized by using alkaline phosphatase-conjugated anti-digoxigenin antibody. Bismarck Brown was used to counter-stain the preparations.

■ **DNA sequencing and computer analysis.** Plasmid DNA for sequencing was purified via the QIAprep Miniprep system or JETstar plasmid purification system (Qiagen). Sequencing was done on both ends of the (sub)cloned fragments by using the universal pBluescript forward and reverse nucleotide primers. The complete *Bam*HI fragment was sequenced from both strands by the 'sequence walking' method with custom-synthesized primers. Automatic sequencing was carried out at the Sequencing Facility in the Department of Molecular Biology of Wageningen Agricultural University.

The sequences generated were analysed with UWGCG computer programs (release 9.0). The DNA and the deduced amino acid sequences were compared with the updated GenBank/EMBL, SWISS-PROT and PIR databases by using FASTA and BLAST. To study *rr1* and *rr2* gene similarity, the amino acid sequences of the selected homologues were aligned by the multiple sequence alignment program CLUSTAL W (Thompson *et al.*, 1994). Phylogenetic analysis was performed with the PAUP3.1 program (Swofford, 1993), using CLUSTAL W to produce input files of aligned protein sequences. Bootstrap analysis according to Felsenstein (1993), included in the PAUP package, was used to assess the integrity of the phylogeny produced. The nucleotide sequence of the WSSV *Bam*HI fragment and the *rr2* gene have been deposited in GenBank under accession numbers AF099142 and AF144620, respectively.

The following RR1/RR2 sequences available in GenBank were used in the alignment and phylogenetic analysis: *Homo sapiens* (Hs), X59617/X59618; *Mus musculus* (Mou), K02927/M14223; *Danio rerio* (Dr), U57964/U57965; vaccinia virus (VAC), A28611/M57977; variola virus (VAR), P32984/X69198; *Saccharomyces cerevisiae* (Sc), P21524/M17789; *Schizosaccharomyces pombe* (Sp), P36602/X65115; *Spodoptera littoralis* nucleopolyhedrovirus (SlMNPV), X98924; *Trypanosoma brucei* (Trb), U80910/Y10768; *Spodoptera exigua* nucleopolyhedrovirus (SeMNPV), X97578; *Paramecium bursaria* chlorella virus 1 (CHLV), U42580/P49730; ASFV (isolate Malawi Lil 20/1) (ASFV), P26685/P26713; lymphocystis disease virus 1 (LDV), L63545; Chilo iridescent virus (CIV), AF003534; equine herpesvirus-1 (strain AB4P) (EHV1), P28846/M86664; varicella-zoster virus (strain Dumas) (VZV), P09248/P09247; human herpesvirus-1 (HSV1), 66401/X14112; equine herpesvirus-2 (EHV2), U20824; herpesvirus saimiri (strain 11) (HVS), Q01037; murine herpesvirus-68 (MHV), U97553; *Lymantria dispar* nucleopolyhedrovirus (isolate CI5-6) (LdMNPV), AF081810; *Orgyia pseudotsugata* nucleopolyhedrovirus (OpMNPV), U75930; *Escherichia coli* (Ec), 66408/P00453; *Salmonella typhimurium* (St), P37426/P37427; and *Synechocystis* sp. (Syn), D90913/D90917. The *Cydia pomonella* granulovirus (CpGV) RR1 sequence was provided by D. Winstanley (Horticulture Research International, Wellesbourne, UK).

Results

Analysis of an 11 kb *Bam*HI fragment of WSSV DNA

WSSV DNA was isolated from purified virions and digested with *Bam*HI (Fig. 1). As determined from agarose gels, the sizes of the fragments ranged from about 22 to 3 kb. The size and number of the larger fragments could not be determined accurately due to their poor separation in agarose gels and the possible presence of genetic variation in the WSSV isolate. From the gel, the total size of the WSSV genome was estimated to be over 200 kb, in agreement with estimates published previously (Yang *et al*., 1997). *Bam*HI fragments were inserted into bacterial plasmids by shotgun cloning. *In situ* hybridizations were performed on sections of infected and uninfected shrimp tissue (data not shown), using the plasmids with *Bam*HI fragments as probes, to confirm the virus origin of the inserts. The enlarged nuclei of cells of infected tissue were heavily stained, whereas sections of uninfected shrimp tissue served as a negative control. A commercial probe and specimen were used as a positive control for infection (DiagXotics).

Plasmid inserts of nine hybridization-positive *Bam*HI clones of 18, 14, 11·3, 6·3, 6·0, 5·0, 4·2, 4·0 and 3·0 kb (Fig. 1) were finely mapped and their termini were sequenced. These sequences were subjected to BLAST analysis to identify homologous sequences in databases (GenBank). Most of the terminal sequences tested did not reveal any similarity to known sequences. An 11·3 kb *Bam*HI fragment, however, contained a short stretch of terminal sequences homologous to sequences encoding the small subunit of RR (RR2). This enzyme is responsible for the reduction of ribonucleotides into precursors for DNA replication (deoxyribonucleotides) and consists of two subunits, RR1 and RR2 (Elledge *et al*., 1992). This 11·3 kb *Bam*HI fragment was sequenced and a detailed physical map was derived (Fig. 2).

The WSSV *Bam*HI fragment was 11319 nucleotides long with a G + C content of 42·2 mol%. Seven complete and one partial ORF and two regions with repeated sequences were identified (Fig. 2). One ORF showed a high degree of similarity to *rr1* genes of prokaryotic and eukaryotic organisms and of large DNA viruses, and hence was designated WSSV *rr1*. This 2547 nucleotide ORF, from nucleotide position 5913 (5′ end) to 8459 (3′ end), potentially encodes a protein of 848 amino acids with a theoretical molecular mass of 96 kDa. The initiation codon of the putative *rr1* ORF is located in a favourable context for translation (Kozak, 1989). A TATA box is located 113 bp upstream of the putative RR1 translation start site. A polyadenylation signal is present three nucleotides downstream of the 3′ end of the ORF. At the left-hand end of the *Bam*HI fragment (Fig. 2), there is the 5′ end of an ORF (*rr2*) with high similarity to the small subunit of RR.

Six more non-overlapping ORFs were identified (Fig. 2), ranging in size from 742 (ORF6) to 74 amino acids (ORF3). Fifty amino acids was taken as the minimum size for a putative ORF. An overview of the locations and directions of transcription of these ORFs and the sizes and isoelectric points of the putative proteins is given in Table 1. The putative initiation codons of ORF1, ORF2, ORF4 and ORF6 are in favourable contexts for translation (Kozak, 1989). A TATA box upstream of the ATG was present for ORF1, ORF4 and ORF6, whereas a polyadenylation signal was found only for ORF1 and ORF4 (Fig. 2). ORF4 overlapped direct repeat region (DRR) 1 (Fig. 2). Its putative product is highly acidic (pI = 4·5) and contains repeated domains in the C-terminal half of the protein. ORF5 is small and has no proximal TATA box or polyadenylation signal; the ATG is in an unfavourable context for translation. ORF6 was identified downstream of the putative WSSV *rr1* gene and has two putative ATG start sites in favourable contexts for translation.

Most of the putative proteins encoded by these ORFs had no hydrophobic sequences that might constitute signal sequences or transmembrane domains. The ORF4 protein contained a putative signal sequence, as it has a positively charged domain with a central hydrophobic region at its N terminus. A consensus motif for cleavage between amino acid residues 19 and 20 was present. Only proteins from ORF1 and

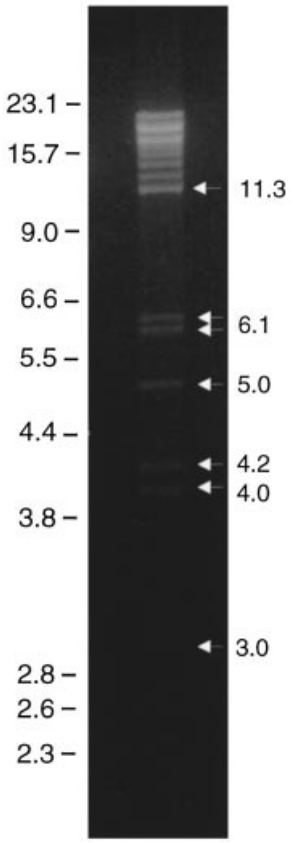

Fig. 1. Electrophoresis in a 0·8% agarose gel of WSSV DNA isolated from purified virus particles and digested with *Bam*HI. The smaller *Bam*HI fragments are indicated by arrows. Lambda DNA digested with *Bam*HI, *Hin*dIII and *Eco*RI and with *Hin*dIII served as size standards. Sizes are shown in kb.

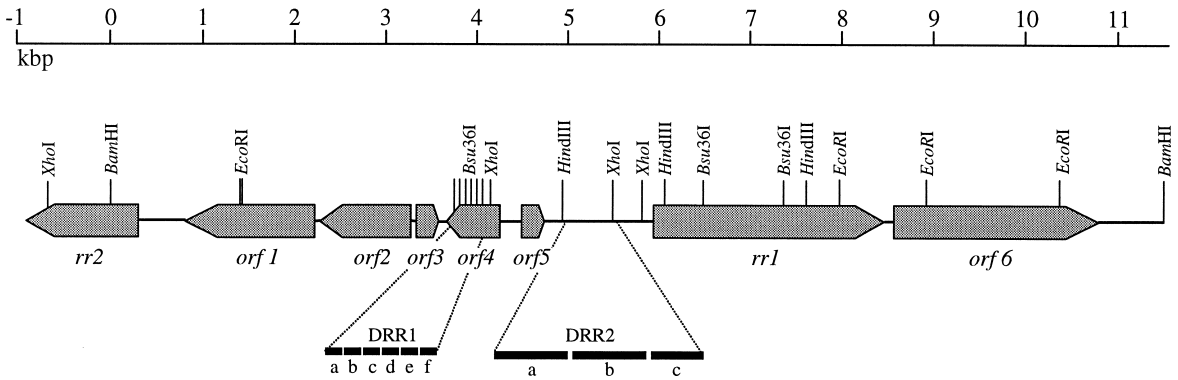

Fig. 2. Physical map of a 12·3 kb segment of WSSV DNA. The top bar shows the location relative to the *Bam*HI site in the RR2 ORF. The locations and directions of transcription of the ORFs are shown below as arrows. The black boxes represent the repeat units in DRR1 and DRR2.

Table 1. Major predicted ORFs and repeat regions on a 12·3 kb WSSV genome fragment

The presence or absence of a TATA box and a polyadenylation signal are shown by + or −, and the Kozak context in which the initiation codon is located is shown as F (favourable) or U (unfavourable). NA, Not applicable; NK, not known.

Name	Location	Amino acids	pI	$M_r \times 10^{-3}$	TATA box	Poly(A) signal	Kozak context
rr2	−943 ← 299	413	4·8	47·6	−	NK	F
ORF1	934 ← 2100	388	8·5	45·2	+	+	F
ORF2	2102 ← 3220	372	6·9	43·2	−	−	F
ORF3	3447 → 3671	74	8·5	8·7	−	−	U
ORF4	3707 ← 4291	194	4·5	22·3	+	+	F
DRR1	3710 − 4033	NA	NA	NA	NA	NA	NA
ORF5	4556 → 4876	106	8·4	12·4	−	−	U
DRR2	4963 − 5649	NA	NA	NA	NA	NA	NA
rr1	5913 → 8459	848	7·8	95·6	+	+	F
ORF6	8540 → 10768	742	5·1	84·4	+	−	F

ORF6 contained stretches of amino acids that might form transmembrane domains. ORF1 had a putative transmembrane region formed by amino acids 4–24. ORF6 contained two putative transmembrane regions at positions 17–37 and 188–204. None of the ORFs contained a canonical nuclear localization signal. Searches against the GenBank, EMBL, PIR and SWISS-PROT databases with the amino acid sequences of ORF1–ORF6, using programs BLAST (Altschul *et al.*, 1997) and FASTA (Pearson & Lipman, 1988), did not reveal significant similarities to any known sequence.

Completion of the putative RR2 ORF

Since only the 5′ end of *rr2* was located on the 11·3 kb *Bam*HI fragment (Fig. 2), a terminal 2·7 kb *Eco*RI–*Xho*I fragment was used to screen an *Xho*I digest of WSSV DNA by Southern hybridization. A 4·8 kb *Xho*I fragment hybridized with this *Eco*RI–*Xho*I fragment and was subsequently cloned and sequenced. This fragment contained a considerable part of the *rr2* gene. The 5′ end of the *rr2* sequence in the *Xho*I fragment was identical to the sequence in the 11·3 kb *Bam*HI fragment, confirming the overlap between these two fragments. To identify the 3′ end of the *rr2* ORF, a 3′ RACE analysis was performed on RNA isolated from infected shrimp tissue (data not shown). The ORF is 1242 nucleotides long and potentially encodes a protein of 413 amino acids with a theoretical molecular mass of 47·6 kDa. Database analysis further confirmed the identity of this ORF as an RR2 homologue. The initiation codon of the putative RR2 protein is in a favourable context for translation (Kozak, 1989).

Repeat regions

Two direct repeat regions (DRR) were detected, DRR1 and DRR2, separated from each other by 929 bp (Fig. 2). Both repeat regions consisted of direct tandemly repeated DNA

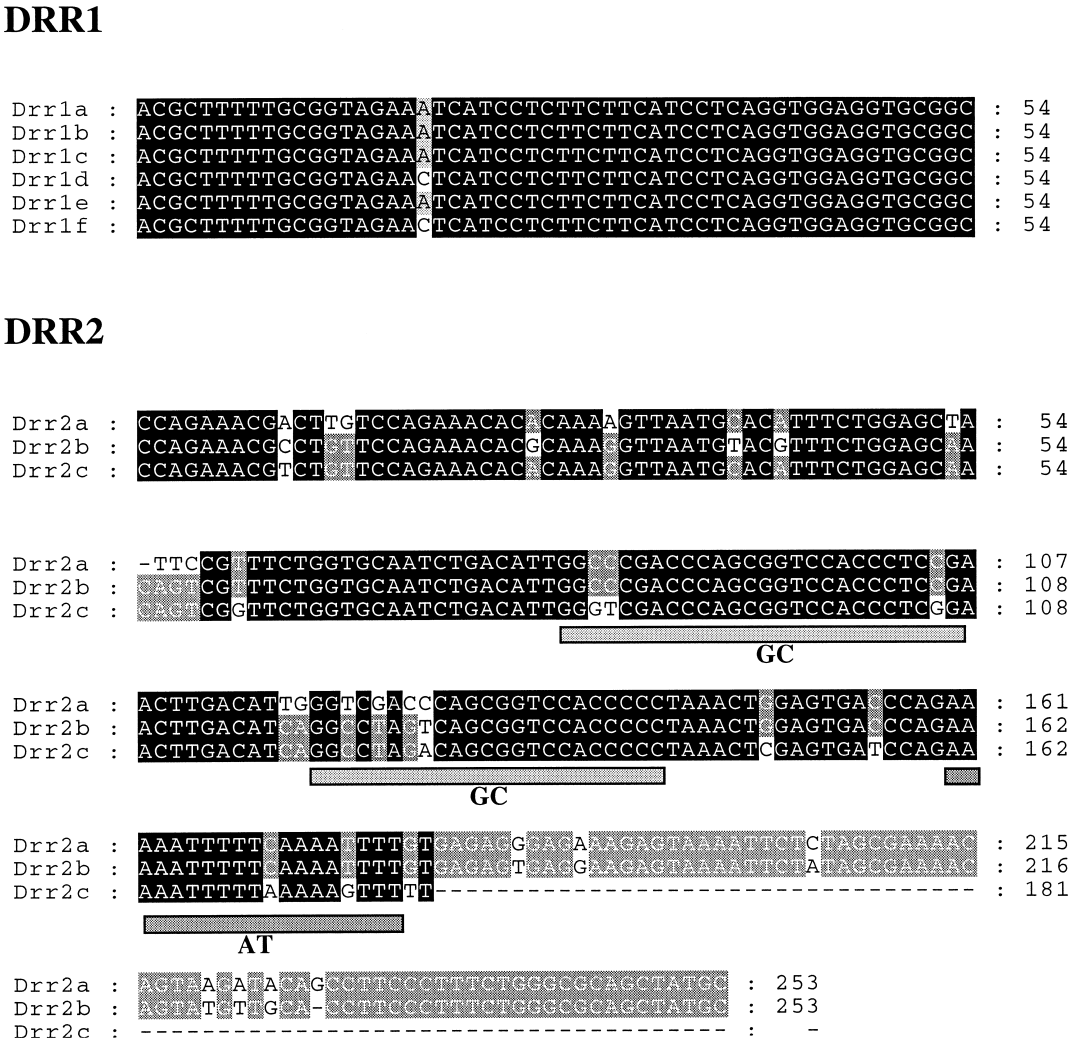

Fig. 3. Nucleotide sequence alignment of DRR1 and DRR2. Shading is used to indicate the occurrence (black 100%, grey > 60%) of identical nucleotides. GC- and AT-rich areas are indicated.

stretches with no inverted repeats or palindromic sequences. DRR1 consisted of six direct repeats of 54 bp with 98% nucleotide sequence conservation (Fig. 3). DRR1 is an intragenic part of ORF4 and results in repeated amino acid stretches in the putative protein.

DRR2 is located 263 bp upstream of the *rr1* gene (Fig. 2) and consists of one partial and two complete direct repeat units of 253 bp (Fig. 3). The conservation of these repeat units was 87%. Comparison of DRR1 and DRR2 with approximately 147 kb of WSSV sequences (M.-F. Tsai & C.-F. Lo, unpublished results) showed that DRR2 is present at multiple, non-overlapping fragments of the WSSV genome, but no sequences homologous to DRR1 were found. In the sequences available, six repeat regions similar to DRR2 were found on non-overlapping fragments. The repeat regions consisted of between two and five complete or partial repeat units. The identity between these repeat units and repeat units DRR2a and DRR2b (Fig. 3) varied between 70 and 90%. When all repeat units were aligned, one conserved AT-rich (93% A + T) region and two conserved GC-rich (both 75% G + C) regions were found (Fig. 3). Comparison of DRR1 and DRR2 with sequences in GenBank did not show any similarity to sequences from other organisms or viruses.

Alignment of WSSV RR proteins

RR is usually composed of two dissimilar subunits, the large subunit (RR1) containing the allosteric regulatory sites and the small subunit (RR2) containing the binuclear iron centre and a tyrosyl free radical (Elledge *et al.*, 1992). This enzyme is important for DNA replication as a provider of dNTP precursors and, hence, is conserved during evolution. The amino acid sequences of WSSV RR1 and RR2 were compared with those from other organisms and viruses. An alignment of

Table 2. Pairwise amino acid identity and similarity (BLOSUM 35) of RR1 and RR2 proteins

WSSV, SeMNPV and OpMNPV RR1 and RR2 amino acid sequences were compared with each other and with the products of other RR genes. Percentage identity/similarity is shown. The RR2 sequences of CpGV and SlMNPV were not available. See Methods for abbreviations. —, Not done.

	RR1			RR2		
Source	WSSV	SeMNPV	OpMNPV	WSSV	SeMNPV	OpMNPV
HS	47/69	48/67	21/37	45/60	43/57	19/36
VAC	45/64	49/67	20/38	44/57	52/69	18/33
SlMNPV	45/65	49/67	21/35	—	—	—
SeMNPV	42/59	—	21/38	34/50	—	19/36
CHLV	40/59	42/61	21/38	39/54	48/68	16/36
ASFV	34/51	35/55	22/37	31/45	34/55	37/13
EC	25/42	24/43	19/34	18/35	21/40	17/36
HVS	24/46	25/45	22/36	19/37	26/50	16/36
LDV	22/36	23/38	19/38	19/36	25/43	14/35
CpGV	19/35	23/37	47/64	—	—	—
OpMNPV	19/34	21/38	—	17/34	19/36	—
LdMNPV (a)	18/34	22/36	83/91	17/35	16/33	79/84
LdMNPV (b)	—	—	—	43/58	48/64	18/36

the complete amino acid sequences was made in CLUSTAL W and then proof-read for inconsistencies (available on request). Most conserved amino acid residues for enzyme activities in the large subunit were also found in WSSV RR1. These include residues involved in substrate reduction and residues located in the cavity of the active site. Residues involved in the radical transfer reaction, as well as two cysteines in the C-terminal end with either a Cys–X_2–Cys or Cys–X_4–Cys motif (Uhlin & Eklund, 1994; Stubbe, 1990), were also conserved. The small subunit (RR2) supplies the reducing capacity of RR and contains an active iron centre and provides electrons via a tyrosyl radical (Schmidt *et al.*, 1998; Liu *et al.*, 1998; Dormeyer *et al.*, 1997). In WSSV RR2, all residues essential for this pathway were conserved.

The alignments showed that the WSSV RR subunits have a high degree of similarity to eukaryotic RR subunits. The pairwise identity and similarity (BLOSUM 35) of the RR1 and RR2 proteins of WSSV, SeMNPV and OpMNPV to each other and to other virus RRs and human and *E. coli* RRs are given in Table 2. The highest relatedness of WSSV RR1 and RR2 was to human RR1 and RR2, whereas the lowest relatedness was observed to OpMNPV and LdMNPV RR1 and RR2. Significant relatedness was found to other baculovirus RRs, such as those from SeMNPV and SlMNPV, and also to Chlorella virus. Furthermore, the WSSV RRs showed a relatively high relatedness to RRs from ASFV. Compared with baculoviruses, WSSV *rr2* has a surprisingly high degree of relatedness to a second *rr2* gene in LdMNPV. This gene, LdMNPV-*rr2b*, is located distally from LdMNPV *rr1* and *rr2a*, which are juxtaposed (Kuzio *et al.*, 1999). WSSV *rr2* has a relatively high degree of relatedness to SeMNPV *rr2*, which also does not co-localize with *rr1* (van Strien *et al.*, 1997; IJkel *et al.*, 1999).

Phylogenetic analysis of WSSV RRs

In order to study the relatedness between RRs of WSSV and eukaryotes, prokaryotes and viruses, phylogenetic trees were constructed from the amino acid sequences of WSSV RR1 and RR2 and those of 26 other RR1 and 24 other RR2 proteins (Fig. 4). Alignments excluding the variable N- and C-terminal sequences were made by using CLUSTAL W and used in the phylogeny study. Maximum parsimony phylogenetic trees were obtained by using PAUP, followed by bootstrap analysis (100 replicates) to determine the 50% majority-rule consensus tree. An heuristic search was performed, where starting trees were obtained by stepwise addition (starting seed 1), and tree bisection–reconnection branch-swapping was performed with the MULPARS function.

The phylogenetic trees for RR1 (Fig. 4*a*) and RR2 (Fig. 4*b*) resembled each other. The bootstrap values of the viruses in the eukaryotic part of the RR2 tree were rather low, making this part of the tree less informative in terms of relatedness. Typically for maximum parsimony, bootstrap values of $\geqslant$ 70% correspond to a probability of $\geqslant$ 95% that the respective clade is a historical lineage. The bootstrap value for vaccinia and variola viruses to form a separate group in the eukaryotic part of the tree was 100% in both trees. The position of the herpesviruses in a separate group was also bootstrap-supported to almost 100% in both trees, which was also the case for the baculoviruses CpGV, LdMNPV and OpMNPV in the RR1

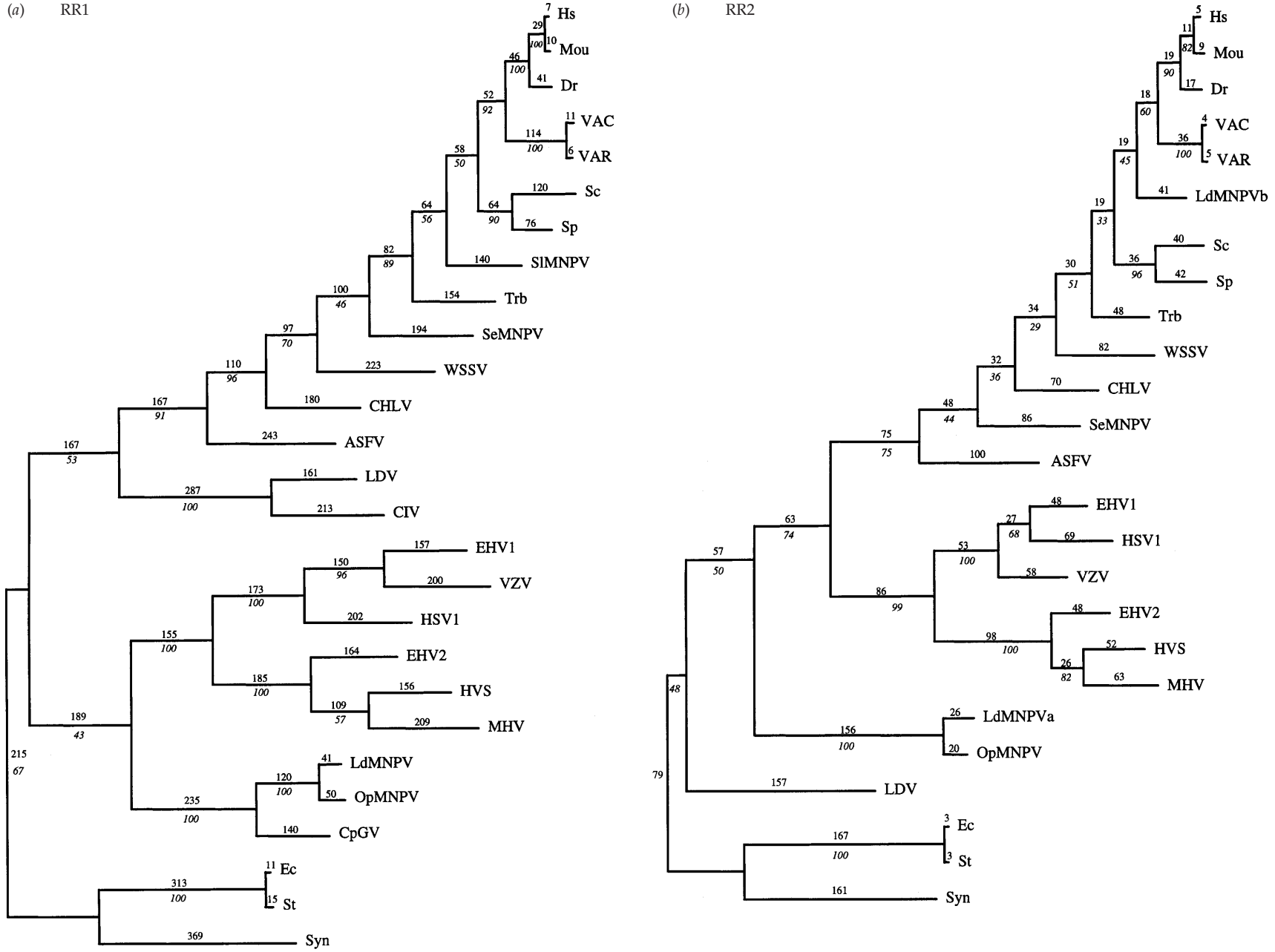

Fig. 4. Bootstrap analysis (100 replicates) of unrooted phylogenetic trees of RR1 (*a*) and RR2 (*b*) proteins constructed with the PAUP heuristic search algorithm. Numbers at the branches indicate branch lengths (normal type) and frequencies of clusters (italics). Outgroups are shown at the bottom. The sequences used were either published or present in GenBank (see Methods for listing). See Methods for abbreviations.

tree, separating these viruses from SeMNPV and SlMNPV. In the RR2 tree (Fig. 4*b*), the same separation was found, except that the extra *rr2b* gene in LdMNPV was found in the eukaryotic branch of the tree. In both the RR1 and the RR2 trees, the phylogenetic status of WSSV was not resolved. In the RR1 tree, WSSV and SeMNPV branched after ASFV and CHLV and before SlMNPV. However, this branching was not clear in the RR2 tree, because of the very low bootstrap values observed for these viruses in the eukaryotic branch of the tree.

Discussion

White spot syndrome is a devastating virus disease in aquatic organisms, in particular in penaeid shrimp. The virus (WSSV) has not been classified by the ICTV (Murphy *et al.*, 1995), but it resembles baculoviruses on the basis of its morphology, morphogenesis and nuclear localization and replication (Durand *et al.*, 1997). Analysis of the viral DNA with restriction enzymes (Yang *et al.*, 1997) gave no further clue as to its taxonomic status. The virus is obtained from many crustacean species, but limited sequence analysis with a two-step PCR procedure seems to suggest that the isolates are very similar if not identical viruses (Lo *et al.*, 1999). The present report is the first to provide functional information on WSSV, by analysing a genomic segment of about 12·3 kb. This segment appeared to contain eight ORFs, including two for the large and small subunits of RR and two direct repeat regions (Fig. 2).

Six ORFs, provisionally named ORF1–ORF6 (Fig. 2), had no homologues in accessible databases (GenBank) and thus appeared to be unique to WSSV. Analysis by terminal sequencing of further *Bam*HI fragments, as well as fragments from *Eco*RI, *Hin*dIII and *Sal*I libraries, showed no significant similarity to sequences in the databases. This information includes the entire sequences of five baculoviruses, AcMNPV (Ayres *et al.*, 1994), *Bombyx mori* (Bm) NPV (Gomi *et al.*, 1999), OpMNPV (Ahrens *et al.*, 1997), LdMNPV (Kuzio *et al.*, 1999) and SeMNPV (IJkel *et al.*, 1999), as well as partial sequence information for other baculoviruses. This is in contrast to baculoviruses, where a large proportion of genes have homologues (Hu *et al.*, 1998). These homologues usually encode proteins that determine the baculovirus character, such as genes involved in DNA replication and transcription (late expression factor genes) or encoding virion structural proteins. On the basis of the limited sequence information obtained, WSSV is distinct from baculoviruses accommodated in the genera *Nucleopolyhedrovirus* and *Granulovirus* (Murphy *et al.*, 1995). The function of the six unique WSSV ORFs with no similarity in databases (GenBank) will be investigated further by molecular and immunological methods.

Comparison of the promoter regions of the eight ORFs on a WSSV segment including the *Bam*HI fragment indicated the presence of common motifs, such as TATA boxes and transcription termination signals. These signals were not present for some ORFs (ORF3 and ORF5), and a favourable Kozak consensus sequence was also absent. These ORFs are possibly not expressed, and transcriptional analysis could indicate whether either of them are functional. Promoter motifs typical of baculovirus transcription, such as the 'early' CAGT and 'late' TAAG (Friesen, 1997; Lu & Miller, 1997), were found scattered in the WSSV sequence but not at appropriate promoter locations for the eight ORFs. This suggests that the transcription strategy is different from that of baculoviruses.

Two ORFs had a high degree of similarity to the large and small subunits of RR and were amenable to further investigation of the taxonomic status of WSSV by phylogenetic analysis. RR is a key enzyme in the DNA replication process, as it catalyses the reduction of ribonucleotides into deoxyribonucleotides (Elledge *et al.*, 1992). Genes involved in nucleotide metabolism are found in pro- and eukaryotic organisms and also in many large DNA viruses. These enzymes, including RR, are therefore excellent tools to study relatedness and phylogeny. Bootstrap analysis of parsimonious phylogenetic trees of RR1 and RR2 genes (Fig. 4) confirmed the location of WSSV in the eukaryotic branch of the RR trees. This positioning of WSSV was supported by the pairwise alignment of the RR proteins, where the highest similarity of the WSSV proteins was found to human and vaccinia virus RRs (Table 2). Several other viruses were found in the eukaryotic branches of the trees, including ASFV, CHLV, SeMNPV and SlMNPV (Fig. 4*a*). In the RR1 tree, ASFV, CHLV and WSSV formed separate clades supported by high bootstrap values. The low bootstrap scores of the SeMNPV and SlMNPV branches indicate that their position of branching remains ambiguous. None of these viruses was located in a separate cluster and hence a common recent ancestor is not likely. The RR2 tree (Fig. 4*b*) is somewhat less informative due to the lower bootstrap values. It is interesting to note that WSSV RR2 was more related to the RR2 of SeMNPV and the RR2b of LdMNPV (Kuzio *et al.*, 1999). The latter two genes are located distant from *rr1* on the viral genome, as is the case for WSSV *rr2*, whereas WSSV had low similarity to *rr2* genes that are juxtaposed to *rr1*, such as OpMNPV *rr2*. These phylogenetic data indicate that WSSV and those baculoviruses that carry *rr* genes do not share an immediate common ancestor.

The region separating the *rr1* and *rr2* genes also contained two domains with multiple repeats. The first repeat domain (DRR1) had six direct repeats of 54 bp, which provides the putative ORF4 product with six repeated domains of 18 amino acids. This protein is highly acidic and its biological function and the significance of these domains remain to be determined. A second domain (DRR2) contained one partial and two complete direct repeats of 253 bp and is located in an intergenic region of the WSSV *Bam*HI fragment. Regions with similar repeats have been identified on at least six non-overlapping WSSV DNA fragments (M.-F. Tsai & C.-F. Lo, unpublished results). These repeats may have functions similar to those of the homologous regions (*hr*s) in baculoviruses. *hr*s also occur at

multiple locations in the baculovirus genome and function as enhancers of transcription and origins of DNA replication (Cochran & Faulkner, 1983; Guarino & Summers, 1986; Kool *et al*., 1995). However, the repeats found in WSSV do not have inverted or palindromic repeats, which appear to be important for their function in baculovirus DNA replication. It remains to be determined whether DRR2-like repeats have a similar function in WSSV replication and/or transcription.

The taxonomic status of WSSV remains to be determined. Originally, this virus was an unassigned member of the *Baculoviridae* (Francki *et al*., 1991). At present, WSSV is no longer accepted in the baculovirus family (Murphy *et al*., 1995). Its large DNA size (> 200 kb) and the lack of significant gene homology with baculoviruses seems to justify its orphan status This taxonomic status is further supported by the phylogenetic information obtained using RR as a tool, where a close ancestral relatedness to baculoviruses could not be demonstrated. The collective information obtained so far, including the sequence information presented in this paper, supports the view that WSSV could be either a representative of a new genus (*Whispovirus*) within the *Baculoviridae* or a representative of an entirely new virus family, Whispoviridae.

We are grateful to Dr E. A. van Strien for assistance in the phylogenetic tree analysis. We thank Dr D. Winstanley (Horticulture Research International, Wellesbourne, UK) for providing us with the CpGV RR1 sequence. This research was supported in part by funds from Intervet International bv, Boxmeer, The Netherlands, and by the National Science Council of Taiwan.

References

Ahrens, C. H., Russell, R. L. Q., Funk, C. J., Evans, J. T., Harwood, S. H. & Rohrmann, G. F. (1997). The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229**, 381–399.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

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

Boursnell, M., Shaw, K., Yáñez, R. J., Viñuela, E. & Dixon, L. (1991). The sequences of the ribonucleotide reductase genes from African swine fever virus show considerable homology with those of the orthopoxvirus, vaccinia virus. *Virology* **184**, 411–416.

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. *Diseases of Aquatic Organisms* **23**, 165–173.

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

Dormeyer, M., Schöneck, R., Dittmar, G. A. & Krauth-Siegel, R. L. (1997). Cloning, sequencing and expression of ribonucleotide reductase R2 from *Trypanosoma brucei*. *FEBS Letters* **414**, 449–453.

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

Elledge, S. J., Zhou, Z. & Allen, J. B. (1992). Ribonucleotide reductase: regulation, regulation, regulation. *Trends in Biochemical Sciences* **17**, 119–123.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5. Department of Genetics, University of Washington, Seattle, WA, USA.

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

Francki, R. I. B., Fauquet, C. M., Knudson, D. L. & Brown, F. (1991). *Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses*. New York: Springer-Verlag.

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.

Gomi, S., Majima, K. & Maeda, S. (1999). Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *Journal of General Virology* **80**, 1323–1337.

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

Hu, Z. H., Arif, B. M., Sun, J. S., Chen, X. W., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1998). Genetic organization of the *Hin*dIII-I region of the single-nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* **55**, 71–82.

IJkel, W. F. J., van Strien, E. A., Heldens, J. G. M., Broer, R., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1999). Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *Journal of General Virology* **80**, 3289–3304.

Kool, M., Ahrens, C. H., Vlak, J. M. & Rohrmann, G. F. (1995). Replication of baculovirus DNA. *Journal of General Virology* **76**, 2103–2118.

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

Kuzio, J., Pearson, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M. & Rohrmann, G. F. (1999). Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* **253**, 17–34.

Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., Rock, D. L. & Van Etten, J. L. (1997). Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology* **237**, 360–377.

Lightner, D. V. (1996). *A Handbook of Pathology and Diagnostic Procedures for Diseases of Penaeid Shrimp*. Special publication. Baton Rouge, LA: World Aquaculture Society.

Liu, A., Sahlin, M., Potsch, S., Sjoberg, B. M. & Graslund, A. (1998). New paramagnetic species formed at the expense of the transient tyrosyl radical in mutant protein R2 F208Y of *Escherichia coli* ribonucleotide reductase. *Biochemical and Biophysical Research Communications* **246**, 740–745.

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. *Diseases of Aquatic Organisms* **35**, 175–185.

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 Press.

Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A. & Summers, M. D. (editors) (1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses.* Vienna & New York: Springer-Verlag.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**, 2444–2448.

Rodriguez, J., Boulo, V., Mialhe, E. & Bachère, E. (1995). Characterisation of shrimp haemocytes and plasma components by monoclonal antibodies. *Journal of Cell Science* **108**, 1043–1050.

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

Schmidt, P. P., Rova, U., Katterle, B., Thelander, L. & Graslund, A. (1998). Kinetic evidence that a radical transfer pathway in protein R2 of mouse ribonucleotide reductase is involved in generation of the tyrosyl free radical. *Journal of Biological Chemistry* **273**, 21463–21472.

Schmitt, J. F. & Stunnenberg, H. G. (1988). Sequence and transcriptional analysis of the vaccinia virus *Hin*dIII I fragment. *Journal of Virology* **62**, 1889–1897.

Stubbe, J. (1990). Ribonucleotide reductases: amazing and confusing. *Journal of Biological Chemistry* **265**, 5329–5332.

Swofford, D. L. (1993). PAUP: phylogenetic analysis using parsimony. version 3.1. Champaign, IL: Illinois Natural History Survey.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Tidona, C. A. & Darai, G. (1997). The complete DNA sequence of lymphocystis disease virus. *Virology* **230**, 207–216.

Uhlin, U. & Eklund, H. (1994). Structure of ribonucleotide reductase protein R1. *Nature* **370**, 533–539.

van Strien, E. A., Faktor, O., Hu, Z. H., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1997). Baculoviruses contain a gene for the large subunit of ribonucleotide reductase. *Journal of General Virology* **78**, 2365–2377.

Vlak, J. M., van Hulten, M. C. W., Lo, C. F. & Kou, G. H. (1999). On the taxonomic position of white spot syndrome virus of penaeid shrimp. In *Abstracts of the 32nd Annual Meeting of the Society for Invertebrate Pathology*, p. 78. Irvine, CA, USA.

Willoughby, K., Bennett, M., Williams, R. A., McCracken, C. & Gaskell, R. M. (1997). Sequences of the ribonucleotide reductase-encoding genes of felid herpesvirus 1 and molecular phylogenetic analysis. *Virus Genes* **15**, 203–218.

Yang, F., Wang, W., Chen, R. Z. & Xu, X. (1997). A simple and efficient method for purification of prawn baculovirus DNA. *Journal of Virological Methods* **67**, 1–4.

Received 20 July 1999; Accepted 20 October 1999