Vol. 50: 87–93, 2002 | DISEASES OF AQUATIC ORGANISMS
Dis Aquat Org | Published July 8

Complete ORF1b-gene sequence indicates yellow head virus is an invertebrate nidovirus

Nusra Sittidilokratna[1], Richard A. J. Hodgson[2], Jeff A. Cowley[2], Sarawut Jitrapakdee[1], Vichai Boonsaeng[1], Sakol Panyim[1], Peter J. Walker[2,*]

[1]Department of Biochemistry and CENTEX Shrimp, Faculty of Science, Mahidol University, Rama VI Road, Bangkok 10400, Thailand

[2]CSIRO Livestock Industries, 120 Meiers Road, Indooroopilly, 4068 Queensland, Australia

ABSTRACT: We report the sequence of an 8503 nucleotide (nt) region of the genome of yellow head virus (YHV) encompassing the open reading frame (ORF)1b gene. Comparison with the sequence of Australian gill-associated virus (GAV) indicated that the region, comprising ~30% of the YHV genome, commences 268 nt upstream of the putative ORF1a termination codon and continues through ORF1b to a site 30 nt downstream of the ORF2 initiation codon. YHV ORF1a and ORF1b overlap by 37 nt. MFOLD analysis of the overlap and downstream region predicted a 131 nt folding structure ($\Delta G = -47.3$ kcal mol^{-1}) with potential to form an RNA pseudoknot. The structure resides 3 nt downstream of a ribosomal frame-shift 'slippery' sequence (AAAUUUU) and a −1 frame-shift at this site would extend the ORF1 polyprotein by 2616 amino acids (299322 Da). In ORF1b, YHV shares 88.9% amino acid sequence identity with GAV and includes conserved polymerase, metal ion binding, helicase and other domains (Motifs 1 and 3) characteristic of nidoviruses. Compared to GAV, the YHV non-coding region linking the ORF1b and ORF2 genes contains a 263 nt insertion. However, the region contains a conserved core sequence of 46 nucleotides (84.8% identity) that includes a stretch of 20 identical nucleotides surrounding a sub-genomic RNA transcription termination site. The data confirms the taxonomic placement of YHV in the *Nidovirales* and supports biological and topographical evidence that YHV and GAV may be classified as distinct species.

KEY WORDS: Yellow head · Virus · Nidovirus · Polymerase

INTRODUCTION

Yellow head virus (YHV) is a pathogen of farmed shrimp that has been responsible for significant production losses throughout Asia. Yellow head disease was first recognized in *Penaeus monodon* in central Thailand in 1990 (Limsuwan 1991, Flegel et al. 1997). The disease usually occurs in juvenile shrimp. Diseased shrimp swim erratically near the surface of the pond and often display a pale yellow colouration of the cephalothorax region due to yellowing of the underlying hepatopancreas. Histologically, lymphoid organs of moribund shrimp show necrosis and vacuolated cells with hypertrophied nuclei and densely basophilic cytoplasmic inclusions (Chantanachookin et al. 1993). The impact in ponds is rapid, with mortalities often reaching 100% within 3 d of the first sign of disease.

Yellow head virions are enveloped, rod-shaped particles (approx. 40 × 170 nm) with prominent surface projections (11 nm) and an internal striated nucleocapsid (Chantanachookin et al. 1993, Loh et al. 1997, Wang & Chang 2000). The single-stranded RNA genome has been variously reported as having negative or positive polarity (Wongteerasupaya et al. 1995, Nadala et al. 1997, Tang & Lightner 1999). Four structural proteins have been identified in virions (170, 110 to 135, 63 to 67 and 20 to 22 kDa), of which the 135 kDa polypeptide has been shown to be glycosylated (Nadala et al. 1997, Wang & Chang 2000).

*Corresponding author. Email: peter.walker@csiro.au

Morphologically, YHV closely resembles gill-associated virus (GAV) that infects *Penaeus monodon* in Australia and causes a disease with histological characteristics similar to yellow head disease (Spann et al. 1997). On the basis of genome organization and expression strategy, GAV has recently been characterized as the first invertebrate member of the *Nidovirales* —a taxonomic order which also includes the coronaviruses, toroviruses and arteriviruses (Cowley et al. 2000a,b, Enjuanes et al. 2000). As for other nidoviruses, the 5′-end of the (+) single-stranded RNA GAV genome expresses a long polyprotein encoded in 2 different reading frames (ORF1a and ORF1b) that are aligned during translation by a –1 ribosomal frameshift (Cowley et al. 2000b). The ORF1b gene encodes sequence motifs for polymerase, metal ion-binding and helicase domains that are also characteristic of nidoviruses. A limited comparison of short regions of the ORF1b gene has indicated that YHV has significant sequence homology with GAV, suggesting the viruses are closely related and should be regarded as geographic topotypes in the yellow head complex (Cowley et al. 1999, Walker et al. 2001).

In this paper, we describe the complete nucleotide and deduced amino acid sequences of the YHV ORF1b gene and a long non-coding region immediately downstream of ORF1b that together represent approximately 30% of the genome. Analysis of the sequence indicates that YHV, although clearly distinct from GAV, also has characteristics consistent with classification in the *Nidovirales* as a member of a new taxon for which the name *Okavirus* has been proposed (Cowley et al. 2000b, Enjuanes et al. 2000).

MATERIALS AND METHODS

YHV was obtained from moribund *Penaeus monodon* showing signs of yellow head disease that were collected from a farm in Chachoengsao province, Thailand, in July 1998. A gill extract from diseased shrimp was passaged once in *P. monodon* and a stock inoculum was prepared by diluting the first passage gill extract approximately 1/100 in lobster haemolymph medium (Paterson & Stewart 1974). Virions were purified by ultracentrifugation in a continuous Urografin™ gradient (Schering) from clarified haemolymph of 200 juvenile *P. monodon* (average weight 20 g) infected with the stock inoculum as described previously (Wongteerasupaya et al. 1995). The purified virus preparation was examined by negative-contrast transmission electron microscopy (data not shown). Virions displayed typical rod-shaped morphology with visible internal nucleocapsids and prominent surface projections. Some virions displayed unusual hexagonal terminal structures that appeared to derive from a distortion in the envelope. Although some incomplete or degraded virions were observed, the preparation was relatively free of cellular debris. There was no evidence of other viruses that may have been present as adventitious infections. Genomic RNA was extracted from purified particles using Trizol™ reagent (Life Technologies), dissolved in diethyl pyrocarbonate-treated, sterile water and stored at –70°C.

Available YHV sequences (GenBank AF102829, AF148846: Wongteerasupaya et al. 1995) and the GAV ORF1b and ORF2 sequences (GenBank AF227196, AY039647) were used to design RT-PCR primer sites to amplify overlapping fragments encompassing YHV ORF1b. Nucleotide sequences of primer sets YHV1–YHV4 are shown in Table 1. Gene fragments were amplified from YHV RNA by using the SuperScript™ 1-step RT-PCR kit (Life Technologies) according to the manufacturers′ recommended procedure, with minor modification. Briefly, ~20 pg of YHV RNA, a primer set containing 500 nM total primers and 8 U RNasin (Promega) were used in a total reaction volume of 25 µl. For amplification of the 5.0 kb product, the Mg^{2+} concentration was raised to 1.6 mM and 1 U Elongase™ (Life Technologies) was added. RT-PCR was conducted in a Hybaid PCR-Sprint thermal cycler using the following heating cycles: 1 × 50°C/30 min, 94°C/2 min (for cDNA synthesis), 35 × 94°C/30 s, 58°C/30 s, 68°C/5 min (for amplification) and 1 × 70°C/10 min (for end-filling). Amplified products were purified directly (BRESAspin™ PCR Cleanup Kit, GeneWorks) or fractionated in agarose gels and then purified from gel slices (BRESAspin™ Gel Extration Kit).

RT-PCR products were sequenced directly at the Australian Genome Research Facility (Brisbane, Australia) using BIG Dye™ reagent (ABI). Sequences de-

Table 1. PCR primers used to amplify YHV genome fragments

Primer set	Primer code	Sequence (5′ – 3′)	Expected product (kb)
YH1	2s40	TGCTTTGACCGTGTTGACGTCGATGAAGAC	1.6
	2a5	TGACGGTCTTCGTGTAGTTTAGTGTATGCGACTGG	
YH2	2s10	CGACATCACTCCAGACAACATCTG	5.0
	2a12	GTGTGAACACCTTCTTGGCTTCCT	
YH3	2s12	CGCTTCCAATGTATCTGCATGCACC	2.5
	2a15	CATCTTGAATTTGTGTGGGTTGTCA	
YH4	2s37	GCTCCTGATTCCCGGTTATG	1.1
	2a38	GCATATGCCAGAGTCCATCCATTACCGCGT	

termined initially by extension from the PCR primers were used to design new primers for sequence walking in both directions. Overlapping sequence contigs were compiled using SeqEd 1.0.3 (ABI). Sequencing ambiguities and regions significantly different from GAV were confirmed by sequencing re-amplified RT-PCR products.

RESULTS

The sequence of an 8503 nucleotide (nt) region of the YHV genome has been deposited in GenBank (Accession No. AY052786). Predicted translation of the sequence and comparison with GAV (Cowley et al. 2000b) indicated that the region commences 268 nt upstream of the ORF1a termination codon, continues through ORF1b and terminates 30 nt downstream of the ORF2 initiation codon (Fig. 1). ORF1b comprises a single long ORF of 7887 nt. Several minor ORFs in the –2 frame are not preserved in GAV and so are unlikely to be translated. Compared to GAV, the YHV ORF1b sequence contains a 9 nt (3 codons) insertion, a 3 nt (1 codon) deletion and a 3 nt (1 codon) extension immediately preceding the UAA stop codon (Fig. 1). The overall nucleotide sequence identity with GAV ORF1b is 80.5% and the mismatches are generally random. The longest identical stretch of 44 nt occurs in the predicted pseudoknot motif that appears to induce ribosomal frame-shift translation of ORF1b (Cowley et al. 2000b).

MFOLD (Version 2.3) analysis of the 131 nt region of the putative ribosomal frame-shift site using default parameters (37°C, 1 M Na^+) predicted a highly ordered RNA folding structure (ΔG = –47.3 kcal mol^{-1}) with helices of the same length (Fig. 2), as predicted for the equivalent GAV sequence (ΔG = –46.1 kcal mol^{-1}: Cowley et al. 2000b), and potential to form an RNA pseudoknot by base-pairing with a downstream sequence. In the YHV pseudoknot region there are 11 nt variations from the GAV sequence, but most either preserve base paring or are in predicted loop regions. One nucleotide change is predicted to disrupt base-paring, but this does not affect the predicted pseudoknot structure. As for GAV, the putative YHV pseudoknot would provide a mechanism for alignment of the reading frames during translation by a –1 ribosomal frame-shift at the 'slippery' AAAUUUU motif in the ORF1a/ORF1b overlap (Fig. 2). According to the 'simultaneous shift' model (Jacks et al. 1988, Brierley et al. 1989), and as proposed for GAV (Cowley et al. 2000b), the YHV frame-shift site is predicted to occur at the Phe codon (UUC) in this region of ORF1a to generate the read-through amino acid sequence HEANF/SDK (Fig. 2). Indeed, YHV ORF1b commences only 2 codons upstream of the putative slippage site whereas in GAV there is a 23 codon overlap in advance of the site. ORF1a in both GAV and YHV terminates at the same sequence, 9 codons downstream of the putative slippage site. It is also of interest that, in ORF1a, there is relatively poor amino acid sequence identity (4 of 10 changes) between YHV and GAV in the region immediately upstream of the slippage site, suggesting that the 1a/1b junction region has no functional significance in the expressed protein (Fig. 2).

Assuming the frame-shift occurs at the predicted slippage site, YHV ORF1b extends the expressed 1a polyprotein by 2616 amino acids (299 332 Da). Overall, YHV and GAV ORF1b amino acid sequences share 88.9 % identity. We have reported previously that GAV ORF1b encodes conserved polymerase, metal ion-binding (MIB) and helicase domains, and 2 sequence motifs (Motifs 1 and 3) of unknown function, that are characteristic of the ORF1b proteins of nidoviruses (de Vries et al. 1997, Cowley et al. 2000b). Each of these domains and the key universally conserved amino acid residues are also evident in the YHV ORF1b polyprotein (Fig. 3). The 6 combined elements of the YHV polymerase domain (including the unique SDD active site residues of all nidoviruses) differ from GAV in only a single, conservative amino acid change. The elements of the helicase domain also contain only a single, conserva-

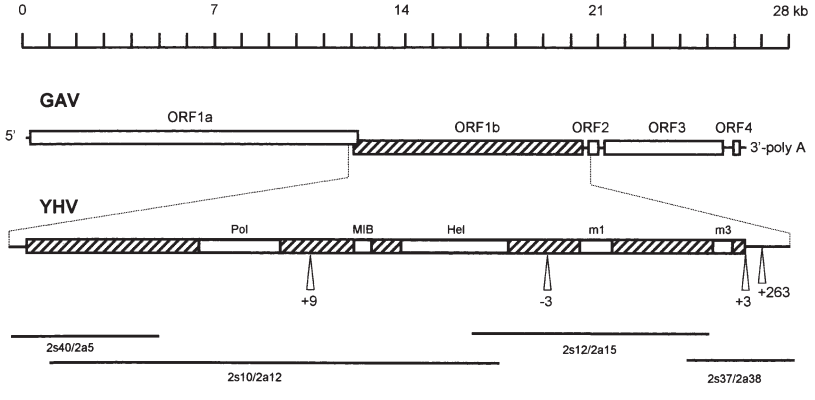

Fig. 1. Schematic illustration of the 8503 nt region of the yellow head virus (YHV) genome described in this paper; gill-associated virus (GAV) genome organization as described in Cowley et al. (2000a). Pol: polymerase domain; MIB: metal ion-binding domain; Hel: helicase domain; m1: conserved Motif 1; m3: conserved Motif 3. The locations of nucleotide insertions (+9, +3, +263) and deletions (–3) in the YHV genome compared to GAV are shown, as are the locations of 4 RT-PCR products amplified using primer sets YHV1-YHV4

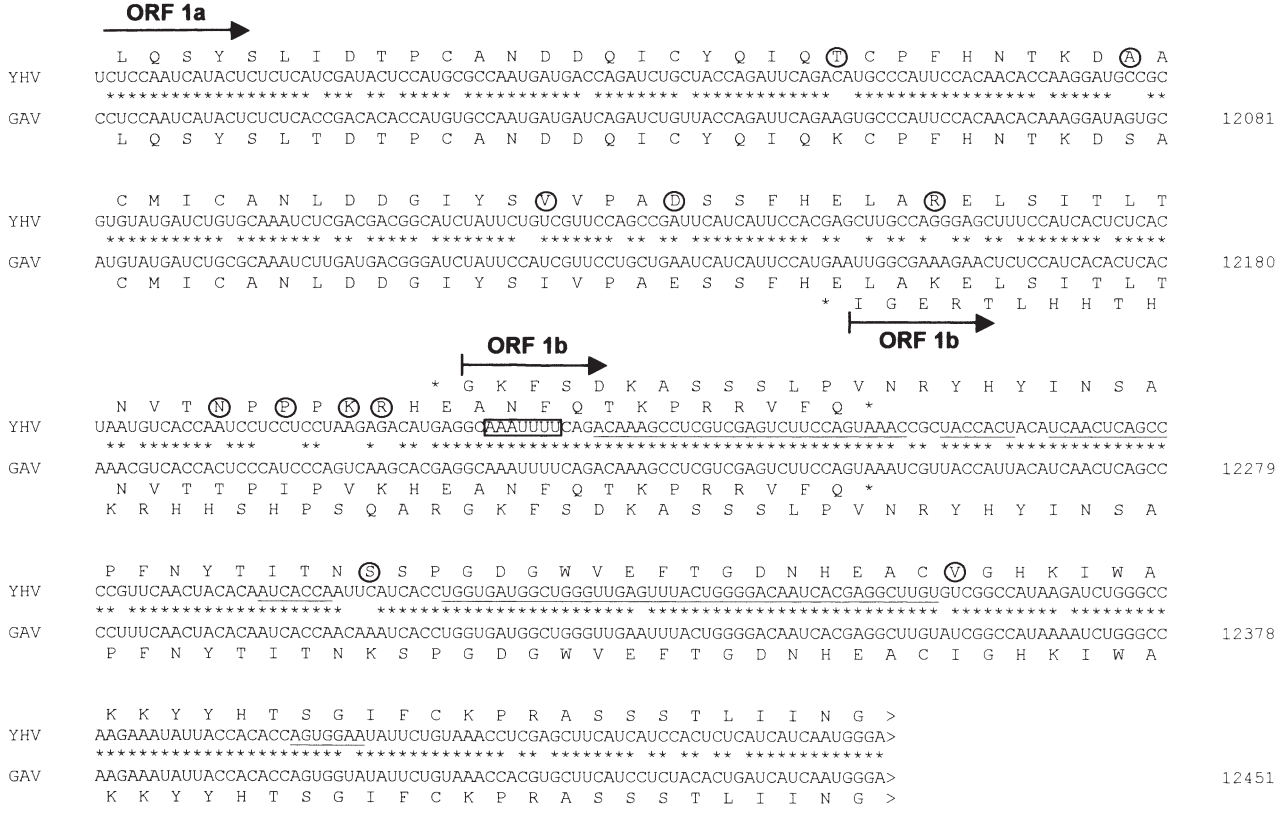

Fig. 2. Comparison of the nucleotide and deduced amino acid sequences of a 469 nt region of the YHV and GAV genomes encompassing the ORF1a/ORF1b overlap. Six elements of the predicted low energy RNA-pseudoknot structure are underlined; the 'slippery' AAAUUUU sequence at the putative ribosomal frame-shift site is boxed; conserved nucleotides are indicated (*); amino acid residues in YHV that differ from GAV are encircled

tive amino acid change and are located within a stretch of 315 amino acids that is the longest region of high sequence identity (99.0%) between the viruses. The Motif 3 sequences of GAV and YHV are absolutely conserved. However, there is substantially more variation in the sequences of the MIB domain and Motif 1. In the MIB domain, all key cysteine residues (10 of 10) and most histidine residues (3 of 5) are conserved, but the overall level of sequence identity is only 86.1%. In Motif 1, amino acid sequence identity is 79.6% and many sequence changes are non-conservative. The high level of sequence identity in the polymerase and helicase domains is characteristic of enzymatic sites that are highly tuned to efficient catalysis. Absolute sequence conservation in Motif 3, which occurs only in nidoviruses (Snijder et al. 1990, den Boon et al. 1991), suggests that this region may also have enzymatic function. In contrast, the MIB domain and Motif 1 are less constrained, as may be expected of binding domains without catalytic activity. Although GAV and YHV are clearly related to other nidoviruses in all significant functional domains and conserved sequence motifs, the overall level of sequence identity in ORF1b is relatively low. An adjusted Clustal X alignment of available ORF1b amino acid sequences of 12 nidoviruses indicated that, excluding all gap positions, GAV and YHV shared 5 to 6% identity with coronaviruses and toroviruses and 3% identity with the torovirus ETV-Berne. Comparatively, coronaviruses shared 10 to 11% identity with ETV-Berne, and arteriviruses shared 7 to 8% identity with ETV-Berne and coronaviruses.

In GAV, the ORF1b gene is followed immediately by a 90 nt non-coding region that includes the identified transcription termination site (5'-ACAACCU…) for a subgenomic RNA encoding the viral nucleoprotein gene, i.e. ORF2 (Cowley et al. 2000a, 2002). In YHV, the corresponding non-coding region between the ORF1b termination codon and the initiation codon for ORF2 is significantly longer, comprising 353 nt (Fig. 4). This variation in genome structure between YHV and GAV was confirmed by sequence analysis on both strands of 2 different YHV isolates (laboratory strain and natural outbreak) and by amplification of RT-PCR products spanning the region in several YHV and GAV isolates using common primer sites (data not shown).

Fig. 3. Clustal W multiple alignment of deduced amino acid sequences of the conserved polymerase (9 non-contiguous domains), metal ion-binding and helicase (7 non-contiguous domains) domains, and Motifs 1 and 3 (3 non-contiguous domains) of YHV, GAV, avian infectious bronchitis virus IBV-M42 (IBV), human coronavirus HCV-229E (HCV) and the equine torovirus ETV-Berne (ETV). The alignment is based on that described by Cowley et al. (2000b). Fully conserved (*), strongly conserved (:), and weakly conserved (.) amino acids are indicated according to amino acid groups assigned by default in the Clustal W programme

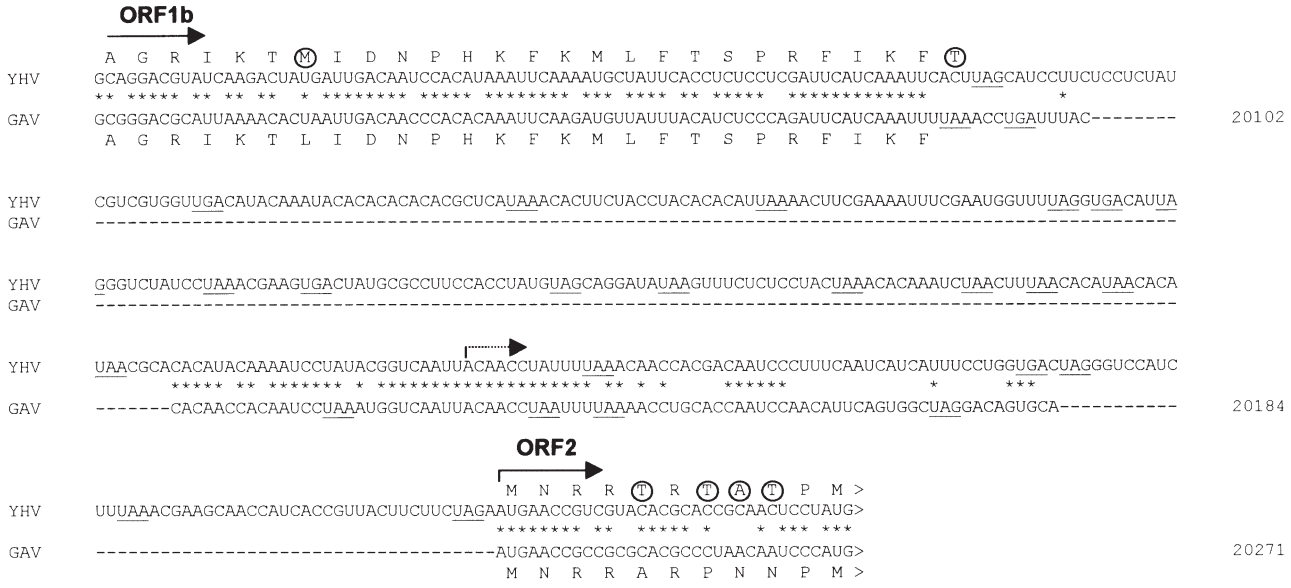

Fig. 4. Comparison of the nucleotide and deduced amino acid sequences of a 470 nt region of the YHV and GAV genomes encompassing the non-coding region between the ORF1b and ORF2 genes. Multiple stop codons in the intergenic region are underlined. The transcription termination site for GAV ORF2 (Cowley et al. 2002) is indicated by the dotted line arrow; conserved nucleotides are indicated (*); amino acid residues in YHV that differ from GAV are encircled

The non-coding region in each virus contains multiple stop codons in each reading frame. In the extended non-coding region of YHV, there is a stretch of 46 nt that shares a high level of identity (84.8%) with the region surrounding the transcription termination site for GAV ORF2. Indeed, at the transcription termination site, there is an unbroken stretch of 20 nt with an identical sequence to GAV, indicating that transcription of YHV ORF2 is also likely to terminate at this site. In the first 11 codons of the ORF2 gene, YHV varies in 4 amino acids from the GAV sequence, suggesting that the level of sequence identity in the nucleoprotein may be far lower than in the ORF1b polyprotein. This has been confirmed in a more complete analysis of the nucleoprotein gene (N. Sittidilokratna et al. unpubl. data).

This paper reports the first extensive sequence data on YHV and includes the complete sequence of the ORF1b polyprotein gene. The data indicates that YHV, like GAV, has genetic features that are consistent with classification as a new taxon (proposed name *Okavirus*) within the order *Nidovirales*. However, the overall level of amino acid sequence identity and significant differences in the length of the ORF1b/ORF2 non-coding region support biological data suggesting that YHV and GAV should be classified as distinct species. Analysis of the structural protein gene region in each virus is currently in progress and should provide further insight into the relationship between these 2 important pathogens of farmed shrimp.

Acknowledgements. The authors gratefully acknowledge the support of the Australian Centre for International Agricultural Research (ACIAR) and the National Center for Genetic Engineering and Biotechnology (BIOTEC) in Thailand. N.S. is a recipient from the Royal Golden Jubilee PhD Program of the Thailand Research Fund and holder of a scholarship from the Faculty of Graduate Studies, Mahidol University.

LITERATURE CITED

Brierley I, Digard P, Inglis SC (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57:537–547

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

Cowley JA, Dimmock CM, Wongteerasupaya C, Boonsaeng V, Panyim S, Walker PJ (1999) Yellow head virus from Thailand and gill-associated virus from Australia are closely related but distinct prawn viruses. Dis Aquat Org 36:153–157

Cowley JA, Dimmock CM, Spann KM, Walker PJ (2000a) Gill-associated virus of *Penaeus monodon* prawns: molecular evidence of the first invertebrate nidovirus. In: Lavi E (ed) The nidoviruses. Kluwer Academic/Plenum Publications, New York

Cowley JA, Dimmock CM, Spann KM, Walker PJ (2000b) Gill-associated virus of *Penaeus monodon* prawns: an invertebrate virus with ORF1a and ORF 1b genes related to arteri- and coronaviruses. J Gen Virol 81:1473–1484

Cowley JA, Dimmock CM, Walker PJ (2002) Gill-associated nidovirus of *Penaeus monodon* prawns transcribes 3′-

coterminal subgenomic mRNAs that do not possess 5′-leader sequences. J Gen Virol 83:927–935
den Boon JA, Snijder EJ, Chirnside ED, de Vries AAF, Horzinek MC, Spaan WJM (1991) Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. J Virol 65:2910–2920
de Vries AAF, Horzinek MC, Rottier PJM, de Groot RJ (1997) The genome organization of the *Nidovirales*: similarities and differences between arteri-, toro-, and coronaviruses. Semin Virol 8:33–47
Enjuanes L, Spann W, Snijder E, Cavanagh D (2000) Nidovirales. In: Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB and 7 others (eds) Virus taxonomy. Academic Press, New York, p 827–834
Flegel TW, Boonyaratpalin S, Withyachumnarnkul B (1997) Progress in research on yellow-head virus and white-spot virus in Thailand. In: Flegel TW, MacRae IH (eds) Diseases in Asian aquaculture. III. Fish Health Section, Asian Fisheries Society, Manila, p 285–295
Jacks T, Madhani HD, Masiarz, FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 55:447–485
Limsuwan C (1991) Handbook for cultivation of black tiger prawns. Tansetakit, Bangkok (in Thai)
Loh PC, Tapay LM, Lu Y, Nadala ECB (1997) Viral pathogens of the penaeid shrimp. Adv Virus Res 48:263–312
Nadala ECB, Tapay LM, Loh PC (1997) Yellow-head virus: a rhabdovirus-like pathogen of penaeid shrimp. Dis Aquat Org 31:141–146
Paterson WD, Stewart JE (1974) *In vitro* phagocytosis by hemocytes of American lobster (*Homarus americanus*). J Fish Res Board Can 31:1051–1056
Snijder EJ, den Boon JA, Bredenbeek PJ, Horzinek MC, Rijnbrand R, Spaan WJM (1990) The carboxy-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related. Nucleic Acids Res 18:4535–4542
Spann KM, Cowley JA, Walker PJ, Lester RJG (1997) Gill-associated virus (GAV), a yellow head-like virus from *Penaeus monodon* cultured in Australia. Dis Aquat Org 31:169–179
Tang KFJ, Lightner DV (1999) A yellow head virus gene probe: nucleotide sequence and application for *in situ* hybridization. Dis Aquat Org 35:165–173
Walker PJ, Cowley JA, Spann KM, Hodgson RAJ, Hall MA, Withyachumnarnkul B (2001) Yellow head complex viruses: transmission cycles and topographical distribution in the Asia-Pacific region. In: Browdy CL, Jory DE (eds) The new wave: Proceedings of the Special Session on Sustainable Shrimp Culture, Aquaculture 2001. The World Aquaculture Society, Baton Rouge, LA, p 227–237
Wang YC, Chang PS (2000) Yellow head virus infection in the giant tiger prawn *Penaeus monodon* cultured in Taiwan. Fish Pathol 35:1–10
Wongteerasupaya C, Vickers J, Sriurairatana S, Nash GL and 6 others (1995) A non-occluded, systemic baculovirus that occurs in cells of ectodermal and mesodermal origin and causes high mortality in the black tiger prawn *Penaeus monodon*. Dis Aquat Org 21:66–77

Editorial responsibility: Carey Cunningham, Aberdeen, Scotland, UK

Submitted: September 6, 2001; Accepted: January 15, 2002
Proofs received from author(s): June 3, 2002