Abstract

Yellow head virus (YHV) is highly pathogenic and was identified as the cause of mass mortalities associated with yellow head disease (YHD) that first appeared in *Penaeus monodon* farmed in Thailand in 1990. By 1992-1993, YHD was widespread throughout the Thai shrimp farming industry, causing losses estimated at ~US$70 million per annum. By the mid 1990s, gross signs consistent with YHD were also being reported in *P. monodon* farmed in many regions of the Indo-Pacific. Due to its high pathogenicity and economic impact, YHV has been listed as a notifiable pathogen by the OIE and the control of YHD remains a significant concern.

At the outset of this study, two genotypic variants of YHV (genotype 1) had been detected in *P. monodon* in Australia (gill-associated virus, GAV, genotype 2) and Vietnam (genotype 3), suggesting that more variants might exist in other regions. The aim of this study was, therefore, to test the hypothesis that genotypic variants existed in *P. monodon* from other locations, and if so, to determine their genetic relationships to the three known genotypes. The study also aimed to improve existing PCR diagnostic protocols to accommodate the detection of all genotypes in the YHV complex.

Fifty-seven field isolates of YH-complex viruses were detected by RT-PCR in tissues of *P. monodon* sampled from nine Indo-Pacific countries. Phylogenetic relationships determined for these isolates using a 671 nucleotide (nt) C-terminal region of the ORF1b gene identified 46 isolates that clustered with the three know genotypes and 11 isolates that clustered in at least three distinct new genotypes. All isolates other than genotype 1 (YHV) were detected in tissues of healthy shrimp. Genotype 4 isolates were detected only in shrimp from India and were slightly less distantly related at the nucleotide level to genotype 5 (85.2% identify) than the other genotypes (80.3%-82.3%). Genotype 6 isolates were only detected in shrimp from Mozambique and were least divergent (3.5%) from genotype 2. One each of three genotype 5 isolates was detected in shrimp from Malaysia, Thailand and the Philippines. The genotype 5 isolate from the Philippines was, however, 6.7% and 7% divergent from the other two isolates, respectively. This level of divergence was greater than found between genotypes 2 and 6 and was similar to that found between isolates of genotype 2 and genotype 3 (~6.7%). This suggests that the Philippine genotype 5 isolate might ultimately be considered as the founding member of a seventh genotype. Genotype 5 isolates were slightly more closely related to genotype 4 (~85.2% identity)

than the other genotypes (83.4%-84.8% identity). Genotype 1 (YHV) isolates were only detected in Thai shrimp affected by YHD. Genotype 2 isolates were detected in Australian shrimp as well as shrimp from Vietnam and Thailand. Genotype 3 had the broadest geographic range, being detected in four countries in Southeast Asia. The finding of single genotypes in Australia (genotype 2), India (genotype 4) and Mozambique (genotype 6) supports the hypothesis that they have evolved independently in geographically-isolated populations of *P. monodon*. The detection of multiple genotypes in Vietnam (genotypes 2 and 3), Malaysia (genotypes 2, 3 and 5) and Thailand (genotypes 1, 2, 3 and 5) suggests that these genotypes have been disseminated by movements of infected *P. monodon* and the trade in live broodstock used for aquaculture.

A ~1.3 kb amplicon at the 5'-terminal region of the ORF3 gene was sequenced for 28 field isolates to examine phylogenetic relationships to assess whether there is evidence of recombination between genotypes. The region, corresponding to N-terminus of gp116 envelope glycoprotein, displayed more sequence variation than the ORF1b amplicon. All isolates of the virulent genotype 1 (YHV) possessed a unique sequence (TILAGIPEKE/D) at the N terminus of gp116 adjacent to the site of endo-proteolysis that cleaves gp116 from the ORF3 polyprotein. In some genotype 1 isolates this unique sequence was followed by a 54 aa deletion that was also not present in other genotypes. The potential role of this unique sequence as a virulence determinant for YHV requires further investigation. Phylogenetic relationships deduced using the ORF3 amplicon sequences were similar to those deduced using the ORF1b amplicon sequence except that genotype 4 was more closely related to genotype 2 than was genotype 3. However, only 18 of the 28 isolates included in the analysis of both ORF1b and ORF3 amplicons clustered in consistent lineages and were assigned as the same genotypes. Inconsistent phylogenies were observed for ten isolates of which six clustered as genotype 3 in ORF1b and as genotype 2 in ORF3, two isolates clustered as genotype 3 in ORF1b and as genotype 5 in ORF3, one isolate clustered as genotype 5 in ORF1b and as genotype 2 in ORF3, and one isolate clustered as genotype 5 in ORF1b and as genotype 3 in ORF3. Discrepancies in genotype assignments were only observed to involve permutations of genotypes 2, 3 and 5 and involved isolates from healthy shrimp originating from Southeast Asia. Sequence analysis of the ~3.2 kb region spanned by the ORF1b and ORF3 amplicons of three putative recombinant viruses VNM-02-H258 (genotype 3/5), IDN-04-H10 (genotype 3/2) and PHL-03-H8 (genotype 5/3) indicated that recombination had occurred at a position just upstream of the ORF1b gene 3'-terminus. These data provide the first evidence of genetic recombination for any

shrimp virus. The high prevalence of recombinants amongst isolates from Southeast Asia has significant implications for diversification, disease emergence and assignment of genotypes for YH-complex viruses.

The region of the genome from the poly[A] tail to the 3'-end of the ORF1b gene (containing all structural protein genes) was sequenced for representative isolates of genotypes 3 and 4. The analysis was conducted to determine whether the evolutionary divergence in the structural protein genes differed significantly from the replicase (ORF1b) gene and to identify conserved motifs likely to be important for protein function and the regulation of RNA transcription and replication. The sequence of the near 3'-terminal genome region of a genotype 5 isolate was also determined to examine whether it possessed an ORF4 gene like genotype 2 or whether it was truncated as in genotypes 1, 3 and 4. Comparisons of the intergenic regions (IGR) upstream of ORF2 and ORF3 identified a conserved sequence 5'-GUCAAUUACACxxAxxUU-3' surrounding the central adenosine residue corresponding to the 5'-terminus of the sub-genomic (sg)mRNAs that is likely to represent the consensus motif used as a transcription regulatory sequence (TRS). A sequence upstream of ORF4 possessed limited homology to the predicted consensus TRS but A>G/U substitutions (genotypes 2, 3, 4 and 5) or a point deletion (genotype 1) occurred at the central critical adenosine residue. It is possible that these mutations explain why a sgmRNA is not transcribed in abundance to allow translation of an ORF4 protein, and why the apparently redundant ORF4 gene has accumulated nucleotide deletions or insertions interrupting its reading in all genotypes except genotype 2. The 3'-terminal genome sequence of genotypes 1, 2, 3 and 4 downstream of the putative ORF4 gene region was extremely highly conserved and was predicted to form a stable hairpin-loop RNA secondary structure with four bulges. Where nucleotide variations occurred in a genotype, other compensatory changes maintained base-pairing and stability of the structure, suggesting that this region is likely to be important for polymerase recognition of the (+) genomic RNA for transcription of (-) genomic RNA.

Conventional and real-time PCR tests for the detection of all genotypes in the YH complex were developed by identifying highly conserved sequences amongst the 57 virus isolates at which primers could be targeted. In the consensus RT-nested PCR, PCR (358 bp) and nested PCR (147 bp) amplicon lengths were kept short to accommodate degraded RNA and pools of two primers were used rather than a single degenerate primer to accommodate all genotypes whist minimizing levels of degeneracy. The consensus real-time PCR used

SYBR-Green chemistry and amplified a 147 bp product using single degenerate primers targeted to the same sites as the nested PCR primer pools. Each PCR method detected the RNA of representatives of all six genotypes. The RT-nested PCR was extremely sensitive, detecting down to a single copy of a GAV synthetic RNA. Phylogenetic analysis using the 95 nt sequence bounded by the nested PCR primers generated genotype associations similar to those generated using the 671 nt sequence, allowing the assignment of genotypes from the amplified products. The consensus RT-nested PCR test has been included in the 5th Edition of the OIE Manual of Diagnostic Tests for Aquatic Animals (2006). The consensus real-time PCR was slightly less sensitive than the RT-nested PCR, detecting down to ~125 copies of the GAV synthetic RNA. However, the test generated products with the expected T_m (77.5°C) with isolates of the six genotypes and showed a linear relationship between input RNA and Ct value up to 10^9 RNA copies. Thus, due to its ability to accurately quantify and compare viral RNA loads in clinical samples, the test could be used to define the infection status of shrimp in relation to threshold levels associated with disease.