Sequence of Genome Segments 1, 2, and 3 of the Grass Carp Reovirus (Genus Aquareovirus, Family Reoviridae)

Qin Fang,*† Houssam Attoui,‡ Jean François Philippe Biagini,‡ Zuoyan Zhu,* Philippe de Micco,‡§ and Xavier de Lamballerie§

*State Key Laboratory of Freshwater Ecology and Biotechnology, †Hydrobiology Institute, Wuhan Institute of Virology, CAS, Wuchang, 430071, Wuhan, Hubé, China; ‡Unité des Virus Emergents, Laboratoire de Virologie Moléculaire, EFS Alpes-Méditerranée; and §Laboratoire de Virologie Moléculaire, Tropicale et Transfusionnelle, Faculté de Médecine de Marseille, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13005 Marseille cedex 5, France

Received June 20, 2000

The genome segments 1, 2, and 3 of the grass carp reovirus (GCRV), a tentative species assigned to genus Aquareovirus, family Reoviridae, were sequenced. The respective segments 1, 2, and 3 were 3949, 3877, and 3702 nucleotides long. Conserved motifs 5' (GUUAUUU) and 3' (UUCAUC) were found at the ends of each segment. Each segment contains a single ORF and the negative strand does not permit identification of consistent ORFs. Sequence analysis revealed that VP2 is the viral polymerase, while VP1 might represent the viral guanylyl/methyl transferase (involved in the capping process of RNA transcripts) and VP3 the NTPase/helicase (involved in the transcription and capping of viral RNAs). The highest amino acid identities (26–41%) were found with orthoreovirus proteins. Further genomic characterization should provide insight about the genetic relationships between GCRV, aquareoviruses, and orthoreoviruses. It should also permit to precise the taxonomic status of these different viruses.

Key Words: aquareovirus; GCRV; GCHV; grass carp; double-stranded RNA; polysegmented fish reoviruses.

MATERIALS AND METHODS

Cell Culture and Virus Propagation

Ctenopharyngodon idellus kidney (CIK) cells were grown as monolayers at 28°C in Eagle's minimum essential medium supplemented with 10% fetal bovine serum (FBS). The virus propagation was realized as previously described (6). Briefly, confluent monolayers of CIK were infected with a virus stock at 5 PFU/ml. The cells were overlaid with maintenance medium containing 2% FBS and incubated for 3 days at 28°C. Culture supernatant (500 ml) was recovered and clarified by centrifugation at 8000g for 30 min. This was followed by pelleting the virus at 40,000g for 3.5 h.

Sequence Determination of Segments 1, 2, and 3

Determination of partial sequences. Genomic RNA was extracted from virus using the proteinase K phenol-chloroform protocol as
The dsRNA segments were run for 1 h (7 V/cm) in a 15-cm-long 1% agarose gel-containing 0.5 μg/ml ethidium bromide in TAE buffer. The bands were visualized by UV transillumination and excised using a scalpel blade. The dsRNA was purified from agarose using the RNaid kit (Bio 101) as described by the manufacturer. Each segment (2 μg) was heat denatured at 70°C for 15 min in 90% DMSO. The reverse transcription was performed at 42°C for 1 h in a final volume of 20 μl containing 50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.2 mM each dNTP, 0.2 μg of DMSO-denatured dsRNA (the final DMSO concentration in the reaction mixture was 4.5%), 40 U of RNase inhibitor (Gibco BRL), 0.1 μM hexanucleotide mixture, and 200 U of Superscript II reverse transcriptase (Gibco BRL). The resulting cDNA was treated with the Klenow fragment of DNA Polymerase I as described elsewhere (8). Blunt end cloned into the EcoRV site of pZERO2.0 vector (Invitrogen) and recombinant vectors were transfected into competent TOP10 E. coli by electroporation. Colony screening permitted the identification of the longest cDNA insert from each segment which was sequenced using M13 universal primers, the D-Rhodamine DNA sequencing kit and an ABI Prism 377 sequence analyzer (Perkin–Elmer).

Determination of 5' and 3' end sequences. Virus dsRNA was extracted using a guanidinium isothiocyanate procedure, using the RNA Now kit (Biogentex) according to manufacturer's instructions. RNA segments 1, 2, and 3 were separated and extracted as described above. The single primer amplification technique was realized as described before (9). Briefly, a 3'-amino blocked oligodeoxyribonucleotide (primer A: 5'-po4-aggtctcgtagaccgtgcacc-nh2-3') was ligated to both 3'-OH termini of the dsRNA, using 10 U of T4 DNA ligase (Boehringer Mannheim). The tailed dsRNA was recovered using the RNaid kit and denatured by heating at 99°C for 1 min in presence of 15% dimethyl sulfoxide. cDNA copies of the genomic RNA were

**FIG. 1.** Strategies for the cloning of segments 1 (A), 2 (B), and 3 (C) of GCRV.
synthesized using a complementary primer (primer B: 5’-gggtgcatc-3’) and 200 U of MMuLV Superscript reverse transcriptase (Gibco BRL).

Primers designated Pr5’ and Pr3’ were designed from the partial sequences obtained as described above. The remainder of each segment sequence was PCR amplified using primers B and Pr5’ or Pr3’ as necessary, according to the scheme shown in Fig. 1. PCR reactions were carried out using 2.5 U of Taq polymerase (Gibco BRL) and 0.5 μM of each primer. Thermal cycling parameters were as follows: one cycle of denaturation (90°C, 10 min) followed by 40 cycles of denaturation (94°C, 50 s), annealing (55°C, 50 s) and extension (72°C, 2 min). The cycling program ended by an extension step at 72°C for 10 min.

The amplicons were analyzed by agarose gel electrophoresis, then ligated into the PGEM-T cloning vector. The recombinant vector was transfected into competent XL-blue E. coli and the insert sequence determined as described above.

Sequence Analysis

The longest open reading frame (ORF) of each segment was determined. The putative protein sequences were compared to the sequences in the databases using the NCBI gapped BLAST 2.0 program. Sequence specific primers for the PCR amplification of the viral cDNA were designed with the help of the Oligo software (National Biosciences Inc.).

RESULTS AND DISCUSSION

Segments 1, 2, and 3 of GCRV genome were characterized and their sequences deposited in the GenBank database under Accession Nos. AF260511, AF260512, and AF260513, respectively. The cloning of each segment was realized in three successive steps. In the first step, the cDNA resulting from reverse transcription of dsRNA using random hexanucleotides was cloned into the pZERO2.0 vector. The screening of the recombinant clones permitted to identify an insert of 2300 bp from segment 1, another insert of 800 bp from segment 2, and a third one of 2500 bp from segment 3. The sequence of each of these inserts was determined and used to design primers for the amplification of the 5’ (second step) and 3’ (third step) ends of each segment. The sequences, locations and orientations of these primers are given in Table 1.

This three-step analysis permitted the full-length sequence determination of segments 1, 2, and 3 of GCRV. The length of the segments, those of the corresponding encoded putative proteins and the 5’ and 3’ non-coding regions are given in Table 2. Each segment was found to contain a single ORF and the analysis of the negative strand sequences did not permit to identify any consistent ORF.

As previously observed with other members of the family Reoviridae (2), segments 1, 2, and 3 of GCRV were found to have conserved terminal sequences. All positive strands of each dsRNA segment had the motif “5’-GUUAUUU-3’" in common at the 5’ end, and the motif “5’-UUCAUC-3’" in common at the 3’ end. Moreover, the first and last nucleotides of each segment are inverted complements. In previous studies of Reoviridae genomes, comparable conserved motifs have been reported (2). They are supposed to act as sorting signals, bringing a single copy of each genome segment into the nascent viral capsid (10, 11). These GCRV specific motifs, together with a number of segment-specific inverted terminal repeats that were found in
the 5’ and 3’ regions of each segment, could interact by homologous base pairing thus holding the RNA transcripts in a circular form. Similar situations were previously reported in the case of other viruses with segmented genomes (10–14).

One interesting feature in the nucleic acid sequences was that the initiator AUG in the three segment was located at position 13 from the 5’ end. Segments 1 and 2 AUGs obey to the Kozak’s rule [(−3)A/GNN-AUGG(+1)] of strong initiation codons, with an A at position −3. In the case of segment 3, position −3 corresponds to a C, and therefore partially conforms to Kozak’s rule.

The theoretical proteins deduced from segments 1, 2, and 3 of GCRV were compared to those of other Reoviridae members deposited in the databases. Significant amino acid identities (26–41%) were found with structural orthoreovirus proteins in the order given in Table 3, with the highest identity values being with orthoreovirus polymerase sequences. This finding is in agreement with the physical resemblance in electron microscopy between orthoreoviruses and aquareoviruses. This finding can also permit to speculate on the physical position of these proteins in the viral structure with regard to the analysis of the orthoreovirus structure. Hence, VP1 of GCRV might represent the “turret” protein at the five-fold axis. It is also noticeable that this protein in orthoreoviruses carries the enzymatic function of guanylyl transferase and methyl transferase. It is involved in the capping process of the nascent RNA transcripts in the primary transcription cycle during infection. Further investigations are needed to confirm this function in the GCRV VP1.

The VP2 of GCRV contains the signature motifs for the Reoviridae RDRP: The motif SG is found at position 688 and the motif GDD at position 739. These features together with the high degree of protein identity (41%) with the RDRP of orthoreovirus, strongly support that VP2 of GCRV is the viral RDRP. This protein is most probably located in the virus core.

Based on comparison to the VP3 of orthoreoviruses, GCRV VP3 might represent the NTPase, helicase protein which is involved in the transcription and capping of viral RNA.

The comparison of GCRV sequences to homologous segments of aquareoviruses was hampered by the lack of genome sequence data in that genus. The only comparison that was possible is that of the G + C content of GCRV sequence with that of sequences currently available from aquareoviruses, namely segment 10 of Strip bass reovirus (Aquareovirus genogroup A, Accession No. U83396) and segment 10 of Coho Salmon reovirus (Aquareovirus genogroup B, Accession No. U90430). Similar values ranging between 52 and 55% were calculated from all of these sequences. The comparison of the G + C content of GCRV segments 1, 2, and 3 to that of homologous orthoreovirus genes was also realized. The average value of the G + C content of orthoreovirus genes 1, 2, and 3 was found to be 46%.

In conclusion, the significant similarities between GCRV and orthoreovirus proteins are most unusual concerning viruses belonging to separate genera. This situation has never been reported before in the family Reoviridae. In order to identify or rule out further homologies to orthoreoviruses and aquareoviruses, further genomic characterization is required. The remainder of the genome of GCRV (segments 4–11) and the genome of a representative member of each Aquareovirus genogroup should be sequenced. These further characterizations should provide insight about the genetic relationships between GCRV, aquareoviruses, and orthoreoviruses. It should also permit to precise the taxonomic status of these different viruses.

REFERENCES


765


