

## Genomic characterization and taxonomic position of a rhabdovirus from a hybrid snakehead

Weiwei Zeng · Qing Wang · Yingying Wang ·  
Cun Liu · Hongru Liang · Xiang Fang ·  
Shuqin Wu

Received: 6 January 2014 / Accepted: 18 March 2014 / Published online: 19 April 2014  
© Springer-Verlag Wien 2014

**Abstract** A new rhabdovirus, tentatively designated as hybrid snakehead rhabdovirus C1207 (HSHRV-C1207), was first isolated from a moribund hybrid snakehead (*Channa maculata* × *Channa argus*) in China. We present the complete genome sequence of HSHRV-C1207 and a comprehensive sequence comparison between HSHRV-C1207 and other rhabdoviruses. Sequence alignment and phylogenetic analysis revealed that HSHRV-C1207 shared the highest degree of homology with *Monopterus albus* rhabdovirus and *Siniperca chuatsi* rhabdovirus. All three viruses clustered into a single group that was distinct from the recognized genera in the family *Rhabdoviridae*. Our analysis suggests that HSHRV-C1207, as well as MARV and SCRNV, should be assigned to a new rhabdovirus genus.

Rhabdoviruses are single-stranded and nonsegmented negative-sense RNA viruses that are ubiquitous infectious agents of animal and plant disease. Currently, rhabdoviruses are classified into nine genera: *Cytorhabdovirus*, *Ephemerovirus*, *Lyssavirus*, *Novirhabdovirus*, *Nucleorhabdovirus*, *Perhabdovirus*, *Sigmavirus*, *Tibrovirus*, and *Vesiculovirus*. In addition, some members of this family have not yet been assigned to a genus (<http://www.ictvonline.org/virusTaxonomy.asp>).

Fish rhabdoviruses have been classified as belonging to the genera *Novirhabdovirus*, *Perhabdovirus*, and *Vesiculovirus*. Fish rhabdoviruses have been isolated from a wide range of seawater and freshwater fish, including cyprinidae fish [1, 2], mandarin fish [3], rainbow trout [4], eel [5], and snakehead fish [6]. They pose a significant threat to aquaculture industry around the world. Rhabdoviruses were reportedly isolated from snakehead fish in the early 1980s, and the complete genomic sequences of these viruses have since been determined [6]. There have been no reports of snakehead fish infected with rhabdoviruses in China to date. However, on July 9, 2012, a sudden disease outbreak occurred in a hybrid (*Channa maculata* × *Channa argus*) snakehead fish farm in the city of Foshan, Guangdong Province, China. The observed clinical symptoms and pathological changes were similar to those associated with rhabdovirus infection, e.g., the appearance of petechiae on the tail fin, hepatomegaly, and serious hemorrhage on the surface of the swim bladder. Based on clinical and post-mortem examinations, the disease was preliminary attributed to rhabdovirus infection. In order to confirm the diagnosis in the diseased hybrid snakehead, the Manual of Diagnostic Tests for Aquatic Animals published by the Office International des Epizooties (OIE) recommends the use of primers for infectious hematopoietic necrosis virus (IHNV), spring viremia of carp virus (SVCV), and viral hemorrhagic septicemia virus (VHSV). Two pairs of specific primers designed based on conserved sequences in the glycoprotein (G) gene of snakehead rhabdovirus (SHRV) and *Siniperca chuatsi* rhabdovirus (SCRNV) were used in reverse transcription polymerase chain reaction (RT-PCR) to detect possible rhabdoviruses in the diseased fish. The results showed that no distinct bands were produced by RT-PCR with primers for IHNV, SVCV, VHSV, or SHRV, whereas a specific band was detected when primers for

W. Zeng and Q. Wang contributed equally to this work.

W. Zeng · Q. Wang · Y. Wang · C. Liu · H. Liang · X. Fang ·  
S. Wu (✉)  
Pearl River Fishery Research Institute, Chinese Academy of  
Fishery Sciences, Guangzhou 510380, China  
e-mail: wushuqin001@21cn.com

W. Zeng  
e-mail: zww8810303@163.com

SCRV were used. Sequence analysis showed that this partial gene shared 94.5 % sequence homology with the corresponding region of the SCR genome. The virus was named HSHRV-C1207 and subsequently confirmed by cell culture, electron microscopy, and animal experiments [7]. In the present study, we analyze the complete genome of HSHRV-C1207 and discuss the possible relationship between HSHRV-C1207 and other rhabdoviruses.

Viral RNA was extracted from infected cell cultures using an RNeasy Mini Kit (QIAGEN Co. Ltd., Shanghai, China) according to the manufacturer's instructions. cDNA copies of the HSHRV-C1207 RNA genome were synthesized, cloned, and sequenced according to a previous study [5]. Rapid amplification of cDNA ends (RACE) PCR were performed using a 2nd generation 5'/3' RACE Kit (Roche Diagnostics Ltd., Shanghai, China) following manufacturer's protocol. To extend the 5' region of the cDNA sequence, 5' RACE was performed using a gene-specific primer. An oligo(dT) adapter primer was used for the 3' RACE, after targeting the poly(A) tail region of the viral genomic RNA by using a Poly(A) Polymerase Kit (Takara Biotechnology Co. Ltd., Dalian, China). The subsequent cloning strategy was performed as described in detail previously [4]. The complete genome sequence of HSHRV-C1207 was obtained and deposited in the GenBank nucleotide sequence database under accession number KC519324. Analysis revealed that the whole genome consists of 11,545 nucleotides. It also contains five typical genes encoding nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and RNA-dependent RNA polymerase protein (L). In addition, leader and trailer sequences at the 3' and 5' termini of the HSHRV-C1207 genome display inverse complementarity. Contiguous to each gene are 3 or 4 bases that have been identified as gene junctions. No tandem repeats were found in the genome. Phylogenetic analysis indicated that HSHRV-C1207, together with MARV and SCR, were clustered into a single group with a bootstrap value of 100 %, which was distinct from the other recognized genera.

### The N gene

The sequence of the N gene is believed to begin at the SCR consensus mRNA start sequence AACAG and end at the transcription stop/polyadenylation signal, TATG (A7). This produces an mRNA consisting 1455 nucleotides (nt 73 to 1528) [3]. The deduced sequence of the N protein is composed of 429 amino acids, with a calculated molecular mass of 47.6 kDa, and an isoelectric point (pI) of 5.26. A conserved motif, SPYSA, is present at amino acid residues 289–293. The nucleocapsid protein has two potential tyrosine phosphorylation sites (amino acid

residues 162–170 and 200–208) and eight potential serine/threonine phosphorylation sites (amino acid residues 12–25, 58–66, 156–164, 217–225, 225–233, 296–304, 299–307, and 329–337). Sequence comparison with other rhabdoviruses revealed that the HSHRV-C1207 N protein shares the most homology and similarity with members of the genus *Perhabdovirus*. The highest sequence homology (95.9 %) was found with the corresponding proteins of SCR. It is distinct from the members of the genus *Novirhabdovirus*, with the HSHRV-C1207 N protein exhibiting low amino acid sequence identity (11.8 %) to the corresponding proteins of the novirhabdovirus SHR.

### The P gene

The HSHRV-C1207 P gene is located between nucleotides 1532 and 2502 and is composed of 981 nucleotides. The largest open reading frame (ORF) consists of 888 nucleotides. It encodes a protein of 296 amino acids with a calculated molecular mass of 32.7 kDa and a pI of 4.47. The HSHRV-C1207 P protein contains four potential serine/threonine phosphorylation sites (amino acid residues 56–64, 84–89, 135–143, and 187–195), two potential tyrosine phosphorylation sites (amino acid residues 62–70 and 75–83), and an N-linked glycosylation site at residue 226. Our analysis also revealed the presence of a proline-rich region between amino acid residues 158 and 202. Previous studies on rhabdovirus proteins demonstrated that the P protein has the least conserved primary sequence among the viruses. Using sequence comparison with other rhabdoviruses, we were able to demonstrate that the P protein sequence of HSHRV-C1207 had highest amino acid sequence homology with SCR (94.2 %) and only 12.2 % sequence identity to SHR.

Vesiculoviruses such as vesicular stomatitis Indiana virus (VSIV) and Isfahan virus (ISFV) usually contain an additional ORF within the P gene coding for the putative C protein. This additional ORF has also been found in some novirhabdoviruses such as infectious hematopoietic necrosis virus (IHNV), in some perhabdoviruses, such as eel virus European X (EVEX), and in some unclassified rhabdoviruses such as SCR. We were also able to find a specific ORF at the 3' end of the P gene in HSHRV-C1207 believed to encode a putative C protein of 73 amino acids, with a predicted pI of 9.52. At the present moment, this C protein shows no sequence homology to other known proteins, and its function remains to be elucidated.

### The M gene

The M gene is 944 nucleotides in length and is located between nucleotides 2508 and 3452. It contains an ORF

627 nucleotides in length, located between nucleotides 2535 and 3161, which encodes the M protein. The M protein is composed of 208 amino acids and has a deduced molecular mass of 23.4 kDa and a pI of 8.1. The M proteins of rhabdoviruses form an important component of the viral envelope, with a possible structural role. We determined that the HSHRV-C1207 M protein has seven potential serine/threonine phosphorylation sites, three potential tyrosine phosphorylation sites (amino acid residues 18–26, 36–44, and 82–90), and an N-glycosylation site at residue 99. A PPPY motif that is conserved in all vesiculoviruses is located between residues 19 and 22. This motif is considered to play an important role in interactions with cellular proteins that participate in the last step of virus budding [4]. When the HSHRV-C1207 M protein sequence is compared with that of other rhabdoviruses, it shows the highest amino acid sequence identity to SCRV (95.7 %). This is followed by 23.6 % to EVEX and 10.0 % to the corresponding proteins of SHRV. Unlike SCRV, a novel ORF has not been found within the M gene in HSHRV-C1207.

### The G gene

The G gene consists of 1659 nucleotides extending from nucleotides 3456 to 5114. It encompasses an ORF that is 1527 nucleotides in length and encodes a protein that is 509 amino acids long with a calculated molecular mass of 56.5 kDa and pI of 7.32. The first 15 amino acids of the HSHRV-C1207 G protein most likely include the signal peptide sequence, and amino acids 456–476 are predicted to be a transmembrane region. The HSHRV-C1207 G protein has 15 potential serine/threonine phosphorylation sites, 11 potential tyrosine phosphorylation sites, and a putative N-glycosylation site, which is located at amino acid 387. Comparison of the HSHRV-C1207 G protein sequence with those of other rhabdoviruses demonstrated that its amino acid sequence was most similar to that of SCRV (93.7 %), and only 18.9 % sequence identity to the corresponding proteins of SHRV was observed.

### The L gene

The region between nucleotides 5119 and 10,961 at the 3' end of the genome contains the L gene. An ORF 6237 nucleotides in length is located between nucleotides 5099 and 11,365. The product of the L gene is the L protein, a viral RNA-dependent RNA polymerase consisting of 2079 amino acids, with a calculated molecular mass of 237.7 kDa and a pI of 7.93. Six conserved blocks (I–VI)

have been identified by comparison of SCRV L proteins using multiple sequence alignments [3]. These conserved blocks are located between amino acids 221 and 1715 of the HSHRV-C1207 L protein. Blocks II and III are the most conserved, as they contain the RNA-directed RNA polymerase catalytic domain between amino acids 38 and 10,066 and mRNA-capping domain V between amino acids 1079 and 1301. Compared with L protein sequence of other rhabdoviruses, HSHRV-C1207 shares highest amino acid sequence homology with SCRV (93.3 %), and only 10.0 % homology with the corresponding proteins of SHRV.

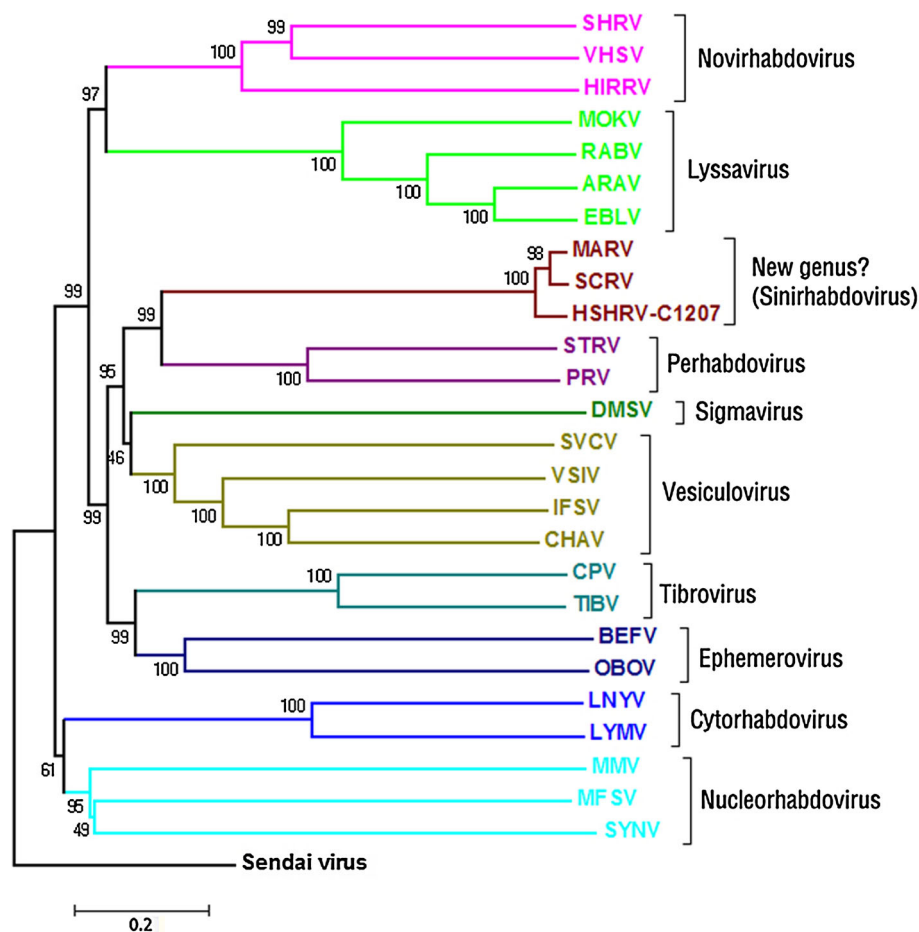
### The 3' leader and 5' trailer regions

The 3' leader region of HSHRV-C1207 comprises the first 72 nucleotides of the genome. This region correlates with the first 71 nucleotides of SCRV and the first 55 nucleotides of SHRV. The 3' leader region of HSHRV-C1207 displays 97.3 and 47.2 % sequence homology to SCRV and SHRV, respectively. The leader sequence begins with nucleotides ACG, which are conserved in all known leader RNA sequences of members of the genera *Ephemerovirus*, *Vesiculovirus* and *Lyssavirus* [8]. The 5' trailer region is composed of 61 nucleotides situated between nucleotides 11,485 and 11,545. It is interesting to note that this region is the same length in SCRV, which has 98.4 % sequence identity. A common feature of rhabdoviruses is the complementarity between the 3' and 5' genome termini, and HSHRV-C1207 displays complementarity of 11 terminal nucleotides of the genome.

### Phylogenetic trees

Since the G protein is located on the surface of virus particles and is believed to be the most divergent among the viral protein products [9, 10], phylogenetic analysis was performed on complete deduced G protein sequences of HSHRV-C1207 and other rhabdoviruses with sequences available in the GenBank database, whereas the fusion protein sequence of Sendai virus was used as an outgroup. Phylogenetic analysis was conducted using MEGA version 5. A neighbour-joining tree was constructed using the maximum composite likelihood model, and the robustness of the tree was tested using 1000 bootstrap replicates.

Analysis based on the predicted amino acid sequences of the G proteins clearly showed 10 distinct phylogenetic clusters with reliable bootstrap support (Fig. 1). Nine of these clusters correspond to the nine recognized genera in the family *Rhabdoviridae*, whereas HSHRV-C1207 was placed in a separate monophyletic lineage together with



**Fig. 1** Phylogenetic relationships between rhabdoviruses based on the complete G protein, using Sendai virus (AAB06281.1) as an outgroup. The tree was generated using the neighbor-joining distance method, and bootstrap values of 1000 tree replicas are shown at each branch node. The existing and newly proposed genera are indicated by different colors. The rhabdoviruses and GenBank sequence accession numbers used in this analysis are as follows: *Monopterus albus* rhabdovirus (MARV; AGZ15720.1), *Siniperca chuatsi* rhabdovirus (SCRV; YP\_802941.1), Hybrid snakehead virus C1207 (HSHRV-C1207; AGI97137.1), sea trout rhabdovirus (STRV; AF434992\_4), perch rhabdovirus (PRV; YP\_007641366.1), spring viremia of carp virus (SVCV; AAW47746.1), vesicular stomatitis Indiana virus (VSIV; AAA48401.1), Isfahan virus (ISFV; YP\_007641385.1), Chandipura virus IB An 9978 (CHAV;

AED98387.1), *Drosophila melanogaster* sigma virus (DMSV; CAA62517.1), Coastal Plains virus (CPV; ADG86362.1), Tibrogargan virus (TIBV; ADG86352.1), bovine ephemeral fever virus (BEFV; AFV92502.1), Obodhiang virus (OBOV; YP\_006200960.1), snakehead virus (SHRV; NP\_050583.1), viral hemorrhagic septicemia virus (VHSV; AGI56027.1), hiram rhabdovirus (HIRRV; NP\_919033.1), Mokola virus (MOKV; YP\_142353.1), rabies virus (RABV; AAB30931.1), Aravan virus (ARAV; YP\_007641395.1), European bat lyssavirus 1 (EBLV; AGQ16846.1), lettuce necrotic yellows virus (LNYV; YP\_425091.1), lettuce yellow mottle virus (LYMV; YP\_002308375.1), maize mosaic virus (MMV; YP\_052854.1), maize fine streak virus (MFSV; YP\_052848.1), *Soenhus* yellow net virus (SYNV; AAA50384.1) (color figure online)

MARV and SCRV, which was supported by bootstrap values of 100 %. This lineage is close to, but distinct from, the perhabdovirus lineage and differs greatly from the other rhabdovirus lineages. These results suggest that HSHRV-C1207, together with MARV and SCRV, may represent a new genus in the family *Rhabdoviridae*, tentatively designated as *Sinirhabdovirus*.

**Acknowledgments** The National Key Technology R&D Program (2012BAD25B02) and The Earmarked Fund for China Agriculture Research System (CARS-46).

## References

- Basic A, Schachner O, Bilic I (2009) Phylogenetic analysis of spring viraemia of carp virus isolates from Austria indicates the existence of at least two subgroups within genogroup Id. *Dis Aquat Org* 85(1):31–40
- Padhi A, Verghese B (2012) Molecular evolutionary and epidemiological dynamics of a highly pathogenic fish rhabdovirus, the spring viremia of carp virus (SVCV). *Vet Microbiol* 156(1–2):54–63
- Tao J-J, Zhou G-Z, Gui J-F (2008) Genomic sequence of mandarin fish rhabdovirus with an unusual small non-transcriptional ORF. *Virus Res* 132(1–2):86–96

4. Morzunov SP, Winton JR, Nichol ST (1995) The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. *Virus Res* 38(2–3):175–192
5. Richard G, van Steven B, Elsa A (2012) Complete genomic sequence and taxonomic position of eel virus European X(EVEX), a rhabdovirus of European eel. *Virus Res* 166:1–12
6. Ahne W, Jorgensen PEV, Olesen NJ (1988) Serological examination of a rhabdovirus isolated from snakehead fish (*Ophicephalus striatus*) in Thailand with ulcerative syndrome. *J Appl Ichthyol* 4:194–196
7. Zeng W-W, Wang Q, Wang Y-Y (2013) Isolation and characterization of a rhabdovirus from snakehead fish (*Ophicephalus striatus*). *J Fish China* 37(9):1416–1424
8. Walker PJ, Wang Y, Cowley JA (1994) Structural and antigenic analysis of the nucleoprotein of bovine ephemeral fever rhabdovirus. *J Gen Virol* 175:1889–1899
9. Delmas O, Holmes EC, Talbi C, Larrous F, Dacheux L, Bouchier C, Bourhy H (2008) Genomic diversity and evolution of the lyssaviruses. *PLoS One* 3:e2057
10. Kuzmin IV, Novella IS, Dietzgen RG, Padhi A, Rupprecht CE (2009) The rhabdoviruses: biodiversity, phylogenetics, and evolution. *Infect Genet Evol* 9:541–553