VIROLOGICA SINICA, February 2008, 23 (1):31-36
DOI 10.1007/s12250-008-2891-5

CLC number: Q789 Document code: A Article ID: 1674-0769 (2008) 01-0031-06

# Isolation and Complete Genomic Sequence Analysis of a New Sindbis-like Virus*

Jing-jing WANG[1#], Hai-lin ZHANG[2#], Yan-chun CHE[1], Li-chun WANG[1], Shang-hui MA[1], Long-ding LIU[1], Yun LIAO[1] and Qi-han LI[1**]

(*1. Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming 650118, China; 2. Yunnan Institute for Endemic Diseases Control and Prevention, Dali 671000, China*)

**Abstract:** The Sindbis-like virus was first discovered in China in 1986. Its complete genomic sequence consists of more than 11 000 bp encoding more than 3 700 amino acids. It contains a 5' non-transcriptional region (5'-NTR) in a non-structural region, four non-structural proteins (nsP1, nsP2, nsP3, nsP4) regions, capsids in conserved and non-conserved regions and structural E1, E2, E3, 6K regions and a 3' non-transcriptional region (3'-NTR). The Sindbis-IMB was isolated from the blood of a patient suspected to have encephalitis, and was followed by identification and passage. The virus RNA was extracted from virus supernatant in infected cells and the whole genome was divided into 12 fragments; RT-PCR was then performed to amplify the 12 fragments for complete sequencing. The results showed that the whole genomic sequence of Sindbis-IMB consists of 11 717 bp encoding 3 773 amino acids. Homology comparison with other Sindbis-like isolates demonstrated the highest similarity was the YN87448 with a variation of 1% strain isolated in Yunnan Province and the second highest to the SAAR86 strain with a variation of ~1.2%. The nucleotide sequence variations were present in non-structural regions, resulting in amino acids K, E, N, R, H, and L in protein sequences in positions 230, 231, 443,781, 1 582, and 1746 in the new isolation respectively. Furthermore, three additional amino acids--glutamic acid, serine and alanine--were noted in nsp4 terminus as compared to the YN87448 isolate.

**Key words:** Sindbis-like virus; Whole nucleotide sequences; RT-PCR

The Sindbis-like virus is a representative *Alphavirus* in the *Togaviridae* family, which infects the host cells of insects, birds and mammals via mosquitoes and insects; infected humans develop fever, rash, arthritis and on occasion, encephalitis (2, 4). The whole or partial genomic sequences of many Sindbis-like viral strains have been reported worldwide (3). In recent years, epidemics of encephalitis have given the genomic analysis of Sindbis-like viruses an elevated importance and extensive epidemiological significance. The YN87448 strain was first isolated in 1986 from the blood of a patient with fever in Yunnan Province (7). Its genome contains a 5'-NTR in the non-structural region, nsP1, nsP2, nsP3,

Received: 2007-07-23, Accepted: 2007-10-16

* Foundation items: National Nature Science Founds (30670094 and 30560142)

# Equally contributing author

** Corresponding author. Tel: +86-871-8335905,
E-mail: qihanlister@gmail.com

nsP4 regions, capsids in conserved, non-conserved and structural regions, E1, E2, E3, 6K structural regions and a 3'-NTR in non-transcriptional region (4), revealing a high homology to Sindbis-like isolates and in particular, a 99% homology to the SAAR86 strain isolated from Africa. In order to investigate the epidemics of Sindbis-like virus infections in Yunnan and the molecular differences between the YN87448 strain and other Sindbis-like viral isolates, we isolated a Sindbis-IMB strain from the blood of a suspected encephalitis patient in Yunnan Province. The results of virus identification demonstrated a high homology to the YN87448 strain. The sequencing of the whole genome was performed and compared with other isolates.

## MATERIALS AND METHODS

### Virus isolation

24 serum samples were collected from patients with unexplained fever in June, 2006 at Medical Institutions in Eryuan County of Yunnan Province. Serum samples were inoculated into Vero cells (Vero cells were provided by the Viral Immunology Dept. of the Institute of Medical Biology, Chinese Academy of Medical Sciences (CAMS), and samples with obvious cytopathic effect (CPE) observed in 3 continuous passages were inoculated into mice with an age of 1 – 3 days. Each of the samples with obvious CPE observation was inoculated intracranially into 6 to 8 mice for observing fever development in mice. Brain tissues of mice with fever were triturated to 10% suspension for passage amplification in Vero cells. The virus was identified by blood clotting and blood clotting inhibition tests (the Sindbis-like virus blood clotting antigen was provided by Yunnan Institute for Endemic Diseases Control and Prevention).

### Primers design

Based upon the complete sequence of the YN87448 strain, 12 primers were designed (Table 1) and synthesized by Shanghai Sangon Biological Eng- ineering Technology&Service Co.Ltd..

### Extraction of viral RNA, Synthesis and cloning of cDNA

Viral supernatant in infected cells was collected. Total RNA was extracted using the QIAmp Viral RNA reagent kit. cDNA was amplified by Promega Access Quick RT-PCR under the following reaction conditions: 45℃ reverse transcription 40 min; 95℃ predenaturating 2 min; 94℃ denaturating 40 sec; 56℃ annealing 40 sec; 72℃ extension 1 min; 35 cycles; 72℃

Table 1. Primers for RT-PCR Amplification

| Primer No | Forward Primer | Reverse Primer | Target |
|---|---|---|---|
| S1 | attggcggcgtagtacacactattg | tatcggtaactttgcatagc | 1 –1 020 |
| S2 | tgagctgcgaaggctacgta | ctcactcagtgctaagaatt | 901 – 1 940 |
| S3 | gttaagatcataacgcactc | ttctggttagcccttgtgag | 1 821 – 2 850 |
| S4 | ccgtgcaagaagaacatcga | ttcgcactgttgaaagtgat | 2 691 – 3 800 |
| S5 | aagagaatcgaatggatcgcc | aggcacacagttgttgcttg | 3 651 – 4 800 |
| S6 | aagacagggattcagtacta | ctgaacgggacgatataatt | 4 661 – 5 700 |
| S7 | tacggatgtgcctatgtctt | cggcttgtatcacttgtact | 5 561 – 6 720 |
| S8 | gcaagaagtgcctatggata | gccttatttcccctctgatg | 6 611 – 7 590 |
| S9 | aggtagacaacatcacacct | gcacggttcagtatggtgac | 7 501 – 8 720 |
| S10 | tcgacatcctcgaagagaac | agacccttactggttcgtga | 8 551 – 9 660 |
| S11 | ccaaccactgaatggatcat | ggaggtagcttgaatgtctc | 9 561 – 10 700 |
| S12 | caccattcgatcacaaggtc | gaaatgtaaaaacaaaattttg | 10 591 – 11 717 |

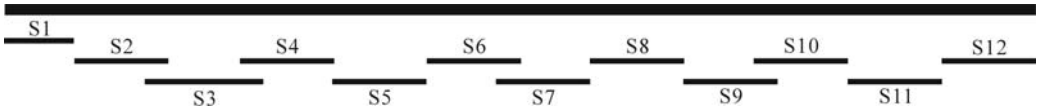


Fig. 1. Cloning approach for Sindbis-IMB complete sequences

extension 10 min. The specific PCR products were cloned into PMD-18T vector (purchased from TaKaRa Corp.) and transformed to *E coli.* DH5α after harvest and purification.

**Sequencing and analysis**

Plasmids were abstracted from the screened clones and the positive clones identified by PCR were sequenced by Sangon Corporation by using RV-M and M13-47 primers in two directions. Sequencing results were analyzed by the OMEGA(Rainbow Techologies, inc.) software package. The cloning approach is shown in Fig. 1. The sequences of other Sindbis-like isolates YN87448, Girdwood, SAAR 86, SINOCK82 and SJ160 to be compared with Sindbis-IMB were obtained from GenBank ( Table 2).

**Construction of phylogenetic trees**

The data of whole genomic sequence from different strains were aligned using MEGA package (version 3.1), and phylogenetic trees were constructed using the neighbor-joining program, from a distance matrix corrected for nucleotide substitutions by the Kimura two-parameter model. Tree stability was assessed via bootstrapping over 500 replicates. The cluster was regarded as strongly supported if the bootstrap value exceeded 95%.

## RESULTS

**Virus Isolation**

One viral isolate that could cause CPE in Vero cells and death of mice was taken from 24 samples and named Sindbis-IMB.

**Whole genomic sequence of Sindbis-IMB**

The whole genomic sequence of Sindbis-IMB was 11 717 bp encoding 3 773 amino acids, of which 7 622 nucleotides were in non-structural regions (excluding 5'-NTR): nsP1, nsP2, nsP3 and nsP4 regions, conserved and non-conserved regions, and capsids in the structural region. The 4059 nucleotides in the structural region (excluding 3'-NTR) were E1, E2, E3 and 6K. The sequencing fragments in two directions were assembled to obtain the whole genomic sequence. After analysis, the ratio of four nucleotides in the whole genomic sequence was determined to be: A: 28.70%; T: 21.22%; G: 25.42%; C: 25.88%.

**Homology comparison between Sindbis-IMB and other isolates**

The homology of the whole genomic sequence of Sindbis-IMB was 98.6% as compared with the other five isolates. Substantial homology was also found to other strains, such as SAAR86 and Girdwood isolates

Table 2. Homology Comparison Between Sindbis-IMB and Other Isolates

| Origins | Isolates | GenBank No. | Homology (%) |
|---|---|---|---|
| China | YN87448 | AF103734 | 99.0 |
| South African | Girdwood | ACU38340 | 98.0 |
| South African | SAAR 86 | ACU38305 | 98.8 |
| Sweden | SINOCK82 | M69205 | 98.0 |
| China | XJ160 | AF103728 | 88.0 |


Fig. 2. Phylogenetic tree of Sindbis-IMB and other isolates.

from South Africa, the XJ160 isolate from Xinjiang in China and the Ockelbo isolate from Sweden (Table 2).

In the constructed phylogenetic tree, the closest genetic relationship was to the YN87448 strain compared with the whole genomic sequence of Sindbis-IMB (Fig. 2).

**Nucleotides and amino acids comparison between Sindbis-IMB and other isolates**

The results of nucleotides and encoding amino acids comparison between Sindbis-IMB and other isolates showed the highest homology to the YN87448 isolate and the second highest to the SAAR86 isolate (Table 3). The main variations were observed in the non-structural region, while the structural region was relatively stable with a mere additional three nucleotides (AAA) at the 5'-terminus in E2 gene.

**Non-structural protein comparison between Sindbis-IMB and other isolates**

The sequence comparison indicated a greatest variation of nucleotide sequences in the non-structural region,

Table 3. Whole Genomic Nucleotides and Amino Acids Comparison Between Sindbis-IMB and Other Isolates

| Regions | Nucleotides | | | | | AminoAcids | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | YN87448 | Girdwood | SAAR86 | SINOCK82 | Xj160 | YN87448 | Girdwood | SAAR86 | SINOCK82 | Xj160 |
| 5′—NTR | 100 | 100 | 100 | 99.0 | 99.5 | | | | | |
| nsp1 | 100 | 99.2 | 98.5 | 97.0 | 96.1 | 100 | 98.5 | 98.5 | 98.7 | 97.3 |
| nsp2 | 99.2 | 97.2 | 97.7 | 98.7 | 96.2 | 99.6 | 98.4 | 98.6 | 98.5 | 96.5 |
| nsp3 | 98.7 | 97.5 | 97.2 | 98.5 | 94.3 | 99.0 | 98.0 | 97.6 | 98.7 | 92.1 |
| Conserved | 99.6 | 99.2 | 99.3 | 99.3 | 97.1 | 99.8 | 99.6 | 99.4 | 99.4 | 97.0 |
| nonconserved | 99.4 | 98.5 | 98.1 | 98.2 | 97.8 | 99.0 | 98.8 | 98.5 | 98.6 | 97.0 |
| nsp4 | 97.2 | 97.8 | 97.5 | 98.0 | 95.0 | 98.0 | 98.4 | 98.8 | 98.7 | 96.0 |
| capisd | 99.5 | 98.2 | 98.0 | 97.8 | 93.4 | 99.7 | 98.7 | 98.5 | 98.2 | 93.0 |
| E3 | 99.3 | 97.9 | 98.1 | 98.0 | 97.6 | 99.5 | 98.4 | 98.2 | 98.6 | 96.0 |
| E2 | 98.8 | 97.3 | 97.0 | 97.9 | 95.3 | 99.2 | 98.0 | 98.3 | 98.7 | 94.1 |
| 6k | 99.0 | 98.1 | 98.5 | 98.2 | 96.3 | 99.5 | 98.8 | 98.9 | 98.7 | 95.0 |
| E1 | 99.4 | 98.3 | 97.5 | 97.3 | 96.0 | 99.7 | 98.9 | 98.6 | 98.7 | 95.0 |
| 3′-NTR | 99.5 | 98.1 | 98.4 | 98.2 | 97.1 | | | | | |

Table 4. Non-structural protein comparison between Sindbis-IMB and other isolates

| | 230 | 231 | 234 | 443 | 781 | 1581 | 1746 | 2523 | 2524 | 2526 |
|---|---|---|---|---|---|---|---|---|---|---|
| IMB | K | E | F | N | R | H | L | Q | S | A |
| YN87448 | R | K | F | K | G | L | P | - | - | - |
| SAAR86 | R | K | L | K | G | L | P | - | - | - |
| Girdwood | R | K | L | K | G | L | P | - | - | - |
| Sinock82 | R | K | L | K | G | L | P | - | - | - |
| XJ160 | R | K | L | K | G | L | P | - | - | - |

amino acids were noted in positions of 230, 231, 443, 781, 1582, 1746, termed K, E, N, R, H, L; and 3 resulting in some variations of amino acids sequences. The sequences of completely different additional amino acids termed Q, S, A were found in the nsP4 terminus (Table 4).

## DISCUSSION

Recent years have seen frequent encephalitis outbreaks. The infectious route and symptoms are similar to Sindbis-like viral infection, while severe cross-reaction occurs in serological reaction, resulting in false diagnosis in clinics. The Sindbis-like virus is a representative Alphavirus in the *Togaviridae* family, which infects insect, bird and mammal etc. host cells via mosquitoes and insects (5). The Sindbis-IMB virus show the characteristics of causing death of mice after the virus was isolated from sera of patients with unknown reason of fever, which is different from the YN87448 isolate with no neurovirulence and no death cause of mice (8). In order to investigate the molecular variations of the isolated Sindbis-IMB virus, whole genomic sequencing of Sindbis-IMB was performed using RT-PCR, demonstrating the highest homology to the YN87448 isolate and second highest to the SARR86 isolate (1).

The whole genomic sequence comparison indicated that though the ratio of four nucleotides were different between Sindbis-IMB and other isolates, such differences would not affect sequence variations of amino acids in the structural region. Compared with the reported sequences of Sindbis-like viruses, the main mutations occurred in positions of 230, 231, 443, 781, 1 582, 1 746, termed K, E, N, R, H, L. Because the YN87448 and Sindbis-IMB strains were isolated from the blood of patients who had fever during different periods of time but in the same place in Yunnan Province, it can be hypothesized that the sequences of these two isolates varied to some extent exhibited mainly in non-structural region after infecting humans via mosquitoes and insects, resulting in changes in virulence. However, no obvious evidence can demonstrate that the Sindbis-IMB originated from the YN-87448 isolate. Further investigations on sequences of more Sindbis-like viruses should be explored. In addition, the 3 additional amino acids termed Q, S, A, which were noted at the nsp4 terminus of Sindbis-IMB. It is reported that the nsp4 has the function of RNA polymerase, which affects synthesis of negative-strand RNA (6), the observed additional three amino acids at 3'-terminus of nsp4 protein needs to be further investigated for their function. Whether such mutations would affect viral neurotoxicity and virus replication has yet to be determined.

All these works have provided experience in further studies of alphavirus, in particular the Sindbis-like viruses. On the other hand, it is suggested that the epidemiological diseases transmitted by mosquitoes and insects in Yunnan Province includes not only the encephalitis, but other Sindbis-like viruses as well, which would help to pay more attention to Sindbis-like viruses except for the encephalitic virus during practical diagnosis in clinics.

## References

1. **Dennis A S, Nancy L D, Sheching L, *et al.*** 1996. Complete nucleotide sequence and full-length cDNA clone of S. A. AR86, a South African Alphavirus Related to Sindbis. **Virology**, 222: 464 – 469.
2. **Dubuisson J, Shlomo L, Nicolas Rug, *et al.*** 1997. Genetic determinants of Sindbis virus neuroinvasiveness. **J Virol** , 71: 2636-3646.

3. **Laine M, Luukkainen R, Toivanen A**. 2004. Sindbis virus and other alphavirus as cause of human arthritic disease. **J Internal Med**, 256: 457-471.
4. **Malherbe H S, Cholmley M, Jackson A.L.** 1963. Sindbis virus infection in man. **South Afr Med J**, 37: 547-552.
5. **Martin P, Richard M K, Oskar R K.** 1998. The alphavirus 3'- nontranslated region:size heterogeneity and ar2rangement of repeated sequence elements. **Virology**, 240 (1): 100-110.
6. **Shailly T, Richard W, Janet L S, *et al.*** 2006. Catalytic Core of Alphavirus Nonstructural Protein Nsp4 Possesses Terminal Adenylyltransferase Activity. **J Virol**, 10: 9962-9969.
7. **Zhou G L, Lian G D, Li L, *et al.*** 1999. Complete nucleotide sequence of the structural region gene of Alphavirus YN87448 strain isolated in China and its relationship to other Sindbis virus. **Virologica Sinica**, 15: 205-210. (in Chinese).
8. **Zhou G L, Lian G D, Li L, *et al.*** 1999. Complete nucleotide sequence of the nonstructural gene of Alphavirus YN87448 strain isolated in China and its relationship to other Sindbis virus. **Chin J Exp Clin Virol,** 13: 314-320. (in Chinese).