Title: Discovery of severe fever with thrombocytopenia syndrome bunyavirus strains originating from intragenic recombination

Running title: Homologous recombination shapes severe fever with thrombocytopenia syndrome bunyavirus

Authors: Cheng-Qiang He* and Nai-Zheng Ding
College of Life Science, Shandong Normal University, Jinan 250014, China

*Corresponding authors: Cheng-Qiang He, College of Life Science, Shandong Normal University, Shandong Province, China 250014, Tel.: 86-531-8618-0745, E-mail: hchqiang@yahoo.com.cn
This study analyzes available severe fever with thrombocytopenia syndrome virus (SFTSV) genomes and reports that a sub-lineage of lineage I bears a unique M segment recombined from two of three prevailing SFTSV lineages. Through recombination, the sub-lineage has acquired nearly complete G1 associated with protective epitopes from lineage III, suggesting that this recombination has capacity inducing antigenic shift of the virus. Therefore, this study provides some valuable implications for the vaccine design of SFTSV.
In May 2007, three patients with fever, abdominal pain, bloating, nausea, vomiting, gastrointestinal bleeding, and elevated amino-transferases were reported in Henan Province of China (28). Similar cases were also found in Shandong, Jiangsu, Hubei and Anhui provinces (28, 29). Until June 2011, 42 peoples were died from the disease at least (http://www.xyw.gov.cn/zt201106p/). The disease, with an initial case fatality rate of approximately 30%, was termed the severe fever with thrombocytopenia syndrome (SFTS) (7). Recently, it has been reported that SFTS has the potential of person-to-person transmission (1, 2, 8). Its pathogen, designated SFTS virus (SFTSV) (29) or Huaiyangshan virus (HYSV) (5, 30, 31), is a novel member of the phlebovirus genus in the Bunyaviridae family.

SFTSV has a single-stranded negative-sense RNA genome comprising of three segments, S, M and L. Segment M encodes the two viral envelope glycoproteins, G1 and G2 which are involved in immunogenicity and neutralizing or protective epitopes (25). Genetic diversity of SFTSV was reported in recent study (31). In order to control the virus, it is necessary to know the mechanism resulting in its genetic diversity. In this study, we analyze all available SFTSV sequence, and report a prevailing SFTSV lineage with recombinant M gene. This finding might provide important insights into the contribution of recombination in shaping the genetic diversity of SFTSV.

All available SFTSV sequences were collected from the GenBank and aligned with CLUSTALW (24). The alignment file can be available online...
Neighbor-joining (NJ), maximum Parsimony (MP) and maximum-likelihood (ML) phylogenetic trees were respectively constructed employing MEGA5 (23). The Best-Fit substitution model, General Time Reversible (GTR) model was used for ML analysis according to Bayesian Information Criterion within MEGA5 software. Identification methods of recombinants were described as previous reports (11, 13). Briefly, The gene sequence similarity analyses were performed and displayed as graphics with Simplot software (15). The sequence alignment files were sought for potential mosaic viruses using recombination detection programs software package RDP2 (17). At last, incongruent phylogenetic relations of segment M different regions delimited by potential breakpoint(s) were used to determine the recombination event.

According to the available M and L segment sequences, SFTSV can be divided into three lineages (Figure 1). And there were two sub-lineages in lineage I (Figure 1A and C). Interestingly, the larger was found to have significant recombination signals in their M segment after deeply phylogenetic analysis.

At first, the M segment sequences were compared. All members of the potential recombinant group shared high sequence similarity (> 98%) each other (Figure 2A). When representatives of these potential recombinants, lineages I and III (HQ642767/BX-2010/Henan/CHN, HM802203/SD4/China/2010 and JF837594/JS2007-01) were compared, it was found that there were two crossover sites around positions 565 and 1533 if HQ642767/BX-2010/Henan/CHN used as the query (Figure 2B). HQ642767/BX-2010/Henan/CHN and JF837594/JS2007-01 of
lineage III had higher similarity between the two crossover sites, which was contrary to other regions (Figure 2B).

Because the crossover sites depended on the sliding window size, they might be different from the recombination breakpoints. Therefore, RDP2 software package was used to identify the potential breakpoint and determine the recombination event. Six recombination detection programs implemented in RDP2 supported that there was recombination signal with two putative breakpoints (positions 503 and 1410 of the alignment): RDP (16), p = 3.2 × 10^{-3}; Bootscan (17), p = 6.9 × 10^{-3}; Maxchi (22), p = 6.1 × 10^{-4}; Chimaera (20), p = 8.4 × 10^{-4}; Siscan (9), p = 7.1 × 10^{-16}; 3Seq (3), p = 1.3 × 10^{-1}. The result of MaxChi is shown in Figure 3A. The analysis of variable sites of the three viruses also determined the potential breakpoints. In different regions delimited by the two breakpoints, identity of variable sites of HQ642767/BX-2010/Henan/CHN and each putative parent were significantly different (Figures 3B and C) (Fisher’s Exact test, p < 1 × 10^{-6}). The schematic drawing of the mosaic M segment is also shown in Figure 3D.

SFTSV phylogenetic histories of different regions delimited by the breakpoints were re-analyzed to further determine the recombination event in M segment. As predicted, the recombinant group had incongruent phylogenetic histories in different regions of segment M. The recombinant group came forth in lineage I when SFTSV phylogenetic history were reconstructed using the sequence of the region before position 503 (Figure 4A) and after 1410 (Figure 4C). On the contrary, the HR group was clustered into the lineage III between the two putative breakpoints (Figure 4B).
These results provided robust evidence that homologous recombination played a key role in diversity of SFTSV. In recombinant M segment, the region encoding nearly complete protein G1 were descended from lineage III, others from lineage I (Figure 3D). All recombinant strains shared the same recombination event and high sequence identity, and comprised a monophyletic group; suggesting that they should be descended from a common recombinant ancestor. Therefore, this recombinant lineage had unique origin different from other SFTSV strains of lineage I.

For a recombination event, co-infection is necessary. The most suitable hosts for such co-infection events are via vectors, where the virus can persist and undergo vertical transmission, which further increases the probability of recombination. Therefore, the host *Haemaphysalis longicornis* of SFTSV could provide the place of the co-infection and recombination for the virus.

Recombination can influence the reliability of phylogenetic analysis (20). It is true that there is topologic difference between the phylogenetic trees with (Figure 1A) and without (Figure 1B) the recombinant group. In the phylogenetic tree without recombinants, lineages I and III constitute a monophyletic group, which consists with the tree inferred from L segment (Figure 1B and C).

As gene reassortment, recombination processes will allow some viruses to acquire many of the key adaptive mutations in a single step and hence make a major leap in fitness space, which might result in a change of host tropism (14). HR can also cause antigenic shift which is important for the emergence of new viral pathogens (6, 27). Here, HR is found in M segment which encodes the protective epitopes of the
virus in G1 and G2 (25). G1 is also the important virulence determinant (18). Through HR, the lineage I has acquired nearly complete G1 from lineage III. It means that this recombination might cause the change of antigenicity and virulence of lineage I. Thus, this study provides the potential implications for vaccine design of the virus.

HR has been considered as a key genetic diversity mechanism for positive-sense RNA viruses. However, intragenic recombination was considered being rare in negative strand RNA virus (NSRV) including important human pathogens such as influenza, Ebola and hantaviruses (4, 10). These viruses have been thought to generate genetic diversity needing for successful evolution, via mutation and reassortment of those with segmented genomes (19). Nevertheless, NSRV pathogens originating from HR and circulating in the field have been reported in influenza(12, 13), Ebola (26), and hantaviruses (19, 21). Co-infection experiments have also confirmed that HR can occur in the hantaviruses of Bunyaviridae family (19). Here, our study suggest that HR result in the new SFTSV sub-lineage circulating in China; which further determines that HR is an important molecular mechanism of NSRV diversity.

In conclusion, this study provides evidence that HR is a key molecular mechanism resulting in the rapid evolution of SFTSV. Considering that HR could bring unpredicted result in viral pathogen, such as the change of antigenicity and virulence, this study also provide valuable clue for vaccine design of the virus.

Acknowledgement

This work is supported by Projects of Shandong Province Higher Educational Science...
and Technology Program (J09LCD209-12 and J12LE06), and a Shandong Province Young and Middle-Aged Scientists Research Awards Fund (BS2009NY011).

Reference


Figure Legends

Figure 1. The evolutionary history of SFTSV circulating in China inferred from sequences of M and L segments. (A) The phylogenetic history of M segment (deleting recombinants). (B) The evolution history of L segment. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary history was inferred using the Neighbor-Joining (NJ) method. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood (MCL) method and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution. The potential recombinants are marked with "■". The representative of lineage I and III is indicated with “▲” and “●” respectively.

Figure 2. (A) A sequence comparison of the segment M of all recombinant viruses. HQ642767/BX-2010/Henan/CHN is used as the query. The y-axis gives the percentage of identity within a sliding window 500 bp wide centered on the position plotted, with a step size between plots of 40 bp. (B) A sequence comparison of the segment M of HQ642767/BX-2010/Henan/CHN and the two parent lineage representatives, JF837594/JS2007-01 and HM802203/SD4/China/2010. The
recombinant HQ642767/BX-2010/Henan/CHN was used the query. The red vertical lines represent the two crossover sites. The rest is the same as (A).

Figure 3. (A) The analysis of putative breakpoint(s) in M segment employing MaxChi method. The vertical lines represent the putative recombination breakpoints. p-value of recombination analysis was shown near the breakpoints. (B) Alignment of M segment variable sites of HQ642767/BX-2010/Henan/CHN and its parent lineage representatives (JF837594/JS2007-01 and HM802203/SD4/China/2010). The recombinant variable sites identical to JF837594/JS2007-01 and HM802203/SD4/China/2010 are respectively indicated with “▲”and “■”. The putative breakpoints are indicated with “*”. (C) Comparison of the segment M variable sites between HQ642767/BX-2010/Henan/CHN and each of its parent lineage representatives. The y-axis gives the percentage of identity between the recombinant and putative parents. The x-axis represents the nucleotide position of M segment. Fisher’s exact test was employed to test the different significance of these variable sites. And p-value is shown between the two putative breakpoints. (D) A schematic representation of the recombinant M segment. The regions encoding Gn/G1 and Gs/G2 are shown. The recombination region is marked with dark color.

Figure 4. Phylogenetic history of M segment different regions delimited by the putative breakpoints. (A) Phylogenetic history inferred from positions 1 to 503 of segment M. (B) Phylogenetic relationship inferred from positions 504 to 1410. (C)
Phylogenetic tree inferred from positions 1410 to 3379. The MCL of NJ method was employed to reconstruct the phylogenetic history. 2,000 bootstrap replicates were performed to assess the robustness of the clustering. The bootstrap values (more than 70%) were shown under each branch. The phylogenetic history of each region was also reconstructed respectively using MP and ML method. The three methods obtained the same topology. The bootstrap values of MP (Left) and ML (Right) method were listed above the branches including the recombinants. The recombinant strains were marked with “■”.
Figure 2
Figure 3
Figure 4