RESEARCH ARTICLE

Molecular Evolution and Spatial Transmission of Severe Fever with Thrombocytopenia Syndrome Virus Based on Complete Genome Sequences

Jian-Wei Liu[1], Li Zhao[1], Li-Mei Luo[1], Miao-Miao Liu[1], Yue Sun[1], Xiang Su[1], Xue-jie Yu[1,2]*

1 School of Public Health, Shandong University, Jinan, Shandong Province, China, 2 Department of Pathology, University of Texas Medical Branch, Galveston, Texas, United States of America

* xuyu@utmb.edu

Abstract

Severe fever with thrombocytopenia syndrome virus (SFTSV) was a novel tick-borne bunyavirus that caused hemorrhagic fever with a high fatality rate in East Asia. In this study we analyzed the complete genome sequences of 122 SFTSV strains to determine the phylogeny, evolution and reassortment of the virus. We revealed that the evolutionary rate of three genome segments were different, with highest in the S segment and lowest in the L segment. The SFTSV strains were phylogenetically classified into 5 lineages (A, B, C, D and E) with each genome segment. SFTSV strains from China were classified in all 5 lineages, strains from South Korea were classified into 3 lineages (A, D, and E), and all strains from Japan were classified in only linage E. Using the average evolutionary rate of the three genome segments, we found that the extant SFTSV originated 20–87 years ago in the Dabie Mountain area in central China. The viruses were then transmitted to other areas of China, Japan and South Korea. We also found that six SFTSV strains were reassortants. Selection pressure analysis suggested that SFTSV was under purifying selection according to the four genes (RNA-dependent RNA polymerase, glycoprotein, nucleocapsid protein, non-structural protein), and two sites (37, 1033) of glycoproteins were identified as being under strong positive selection. We concluded that SFTSV originated in central China and spread to other places recently and the virus was under purifying selection with high frequency of reassortment.

Introduction

Severe fever with thrombocytopenia syndrome (SFTS) was an emerging tick-borne hemorrhagic fever with high case fatality (12 to 50%), and had been reported in most parts of China, South Korea and Japan [1–5]. The major clinical symptoms and laboratory abnormalities of SFTS were non-specific including fever, gastrointestinal symptoms, myalgia, regional lymphadenopathy, thrombocytopenia, leucopenia, and elevated serum hepatic enzymes [1–3]. Severe

patients would progress to multiple organ failure or death. The causative agent of SFTS was a novel bunyavirus SFTSV, in genus *Phlebovirus* [1, 6]. Heartland virus, Malsoor virus and Hunter Island Group virus were genetically closely related to SFTSV and were isolated from patients in the United States, bats in India and ticks in Australia, respectively [7–9]. SFTSV was detected in *Haemaphysalis longicornis* [1, 10] and was demonstrated to be transmitted by tick transovarially and transstadially [11]. Person to person transmission of SFTS through contact with infected patient's blood or mucus had been reported in several clusters of nosocomial transmission of SFTS in China [12, 13]. Animal hosts of SFTSV included domestic animals and wild mammals, such as goats, cattle, dogs, chickens, rodents and shrews [14–17].

Like all the phlebovirus, SFTSV had a three-segmented RNA genome [1, 5]. The L segment contained 6,368 nucleotides encoding the RNA-dependent RNA polymerase (RdRp), while the M segment contained 3,378 nucleotides coding for the envelope glycoproteins Gn and Gc, which played important roles in receptor binding and entrance into the cells. The S segment with 1,744 nucleotides had an ambisense arrangement, which encoded the nucleocapsid protein (N) in a viral complementary sense orientation, and a non-structural protein (NSs) in a viral sense orientation that interacts with interferon signaling pathways and played an important role in evading innate immunity [18–20]. The termini of the three segments of SFTSV were untranslated regions (UTRs) as other bunyaviruses, which were short and highly conserved. Indeed, the 5' and 3' UTRs of the three segments were extensively complementary, forming approximately a panhandle-like structure [21].

Mutation and recombination/reassortment were the main revolutionary force for viruses, which might increase their virulence and pathogenicity. In segmented-genome viruses, reassortment was an extremely efficient evolutionary force, which made the viruses fit the new environment, even result in a change of host tropism [22]. Reassortment events had been reported in several segmented viruses and they could cause severe diseases, even death. It was reported that reassortment of RNA segment occurred in SFTSV [23–25].

Bayesian analysis method had been utilized widely to infer the origin and transmission of the virus presently, including influenza virus, Peste des Petits Ruminants virus (PPRV), and Ebola virus (EBOV) [26–30]. Phylogenetic analysis demonstrated that SFTSV could be classified into two clades including six genotypes [25, 31]. Bayesian analysis based on the SFTSV single segment suggested that SFTSV likely originated 50–225 years ago [23, 32]. Bayesian phylogeographic analysis of SFTSV sequences in China suggested that the virus likely originated in Huaiyangshan [23]. Only in China, was spatial transmission analyzed [23], but transmission among these countries in East Asia was still unknown. In our study, analysis based on the SFTSV complete genome sequences was performed to determine the phylogeny, evolution, spatial transmission and reassortment of the virus.

Materials and Methods

Data collection

We analyzed only SFTSV that had complete genome sequences for three segments in GenBank (http://www.ncbi.nlm.nih.gov/genbank). The information of each SFTSV strain was collected including GenBank number, host, collecting location and collection date. The collection date of the virus was considered as the date the viral strains was isolated, hence, the range from the collection time to the present was the existence time of the viral strains. In total, 122 SFTSV strains with a time span from 2010 to 2014 were collected from GenBank and analyzed in this study. To analyze the whole genome sequences, the L, M and S segments of each strain were concatenated in the order of L-M-S. The concatenated sequence of L segment, M segment, and S segment of each viral strain was used for phylogenetic analysis.

Phylogenetic Analysis

The L, M, S and concatenated sequences of SFTSV were aligned using the Clustal W in MEGA5.02 software program (http://www.megasoftware.net/). The phylogenetic trees for analyzing molecular evolution were constructed using the maximum likelihood method based on the Kimura 2-parameter model using the MEGA. The confidence of the phylogenetic tree was tested using 1,000 bootstrap replications. The "use all sites" option was chosen for the Gaps/Missing Data Treatment, and the "1st + 2nd + 3rd + Noncoding Sites" was for the codon positions.

Reassortment Analysis

Prior to reassortment detection, the concatenated complete genome sequences were aligned using the multiple alignment program Clustal X. To prevent potential biases during phylogenetic inference, the concatenated sequences were analyzed with Recombination Detection Program version 3.44 (RDP3) that incorporates RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiScan and 3Seq methods [33, 34]. Only reassortment events with p-values≦0.05 which were detected by three or more methods were considered, employing the Bonferroni correction to avoid false positive results.

Selection Analysis

Analysis of selection pressure in individual gene of SFTSV was performed by obtaining mean ratios of non-synonymous (dN) to synonymous (dS) substitutions per site (dN/dS). The overall dN/dS ratios of all genes were calculated by using single likelihood ancestor counting (SLAC) in the HyPhy package [35] on the Datamonkey (http://www.datamonkey.org). The positive selection sites were identified by SLAC, fixed effects likelihood (FEL), internal fixed effects likelihood (IFEL), mixed effects model of evolution (MEME) and the fast unbiased Bayesian approximation (FUBAR) [36]. Only sites detected by at least two methods with statistically significant values were considered as being subject to positive selection under the p-value≦0.05 for SLAC, FEL, IFEL and MEME; posterior probability≧0.95 for FUBAR.

Bayesian Time-Scaled Phylogenetic Analysis

Prior to determine the molecular evolutionary rate of the four data sets, the clock-like behavior of each data set was assessed using a regression of root-to-tip genetic distances inferred from the maximum likelihood (ML) trees against sampling time in the program Path-O-Gen (v1.4, http://tree.bio.ed.ac.uk/software/pathogen/) [37]. To determine the molecular evolutionary rate and estimate the time to the most recent common ancestor (TMRCA) for each of the data sets, Bayesian coalescent phylogenetic analysis was performed using the Bayesian Markov chain Monte Carlo (MCMC) method available in the BEAST package v1.8.0 (http://beast.bio.ed.ac.uk) [27, 32]. The HKY + G nucleotide substitution model and a flexible non-parametric Bayesian skyride coalescent model were used. To allow for rate variation among lineages, an uncorrelated lognormal relaxed molecular clock was used in all cases. We performed the independent runs for 10^5 generations, sampling every 2,000 steps. In addition, to accurately estimate the substitution rate, we repeated the analysis. The nucleotide substitution rate (substitutions/site/year) and the TMRCA (year) values were obtained from Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility (MCC) phylogenetic tree was summarized by using TreeAnnotator (http://beast.bio.ed.ac.uk/treeannotator) and exclusion of the first 10% of the trees as burn-in. The MCC tree with median node heights was visualized in FigTree software v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). To test the strength of

temporal signal in the data, which is essential to the estimation of substitution rates, we repeated the BEAST analysis (using identical parameters) on a data set in which sampling times were randomized on the tips [37, 38]. The 95% highest posterior densities (HPDs) of these randomized sequences were then compared with those of the real data. If these sequences contain clear temporal structure, then the real and randomized data would have different mean estimated substitution rates, and different distributions. Here, the null distribution of mean substitution rates for each data set was generated 5 times per alignment.

Results

Phylogenetic analysis

Only SFTSV strains with complete genome sequences were analyzed in this study. Among the 122 SFTSV strains selected, 108 strains were from China. Chinese SFTSV strains were 61 from Henan, 21 from Jiangsu, 8 from Shandong, 7 from Liaoning, 4 from Hubei, 4 from Anhui, and 3 from Zhejiang. Eight SFTSV strains were from Japan and 6 SFTSV strains were from South Korea. All SFTSV strains were isolated between 2010 and 2014. Phylogenetic analysis classified all SFTSV strains into 5 lineages in each segment, named A, B, C, D and E (Fig 1). Lineage A contained the most viral strains (55 strains) and all strains in lineage A came from China (Henan, Jiangsu, Shandong, Anhui, Liaoning provinces) except for 1 strain (Gangwon/Korea) from South Korea. All SFTSV strains in lineage B were from China, including Henan, Jiangsu, Shandong, Anhui, Liaoning provinces. Lineage C included only 2 strains all from China, with 1 strain from Shandong Province, and the other from Jiangsu Province, while most SFTSV in lineage D were from China (Henan, Hubei, Shandong, Anhui, Liaoning provinces) with 1 strain (KASJH) from South Korea. Interestingly, all the 8 strains from Japan, 3 strains from Zhoushan Archipelago in the East China Sea of Zhejiang Province of China, and 4 strains from Jeju Island of South Korea belonged to lineage E. We also found that different segments of the same strain belonged to different lineage, suggesting segment reassortment, e.g. AHL belonged to lineage E in segment L, but it belonged to lineage B and D in segment M and S, respectively.

Reassortment among SFTSV strains

The concatenated sequences were analyzed by RDP4, suggesting 14 reassortment events in SFTSV complete genomes. Reassortment events existed in M and S segments in eight viral strains: HL/Injected, HL/Egg/G2, HL/Larvae/G2, HL/Nymph/G2, HL/Adult/G2, HN13, 2012YXX1, HN_278. These findings were confirmed by RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiScan and 3Seq and supported by significant p-values of 1.73E-11, 2.35E-8, 3.76E-10, 1.41E-12, 1.72E-9, 5.44E-34, 7.52E-46 respectively. Likewise, the viral strain LN2012-58 was also considered as a reassortant virus in M and S segment because p-values of RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiScan and 3Seq were 2.29E-5, 3.30E-18, 8.45E-6, 1.25E-6, 7.81E-4, 4.22E-14 and 4.07E-5 respectively. Four strains (LN2012-14, LN2012-34, LN2012-41 and LN2012-42) were considered as reassortment events in S segment. These results were found by RDP, GENECONV, Bootscan, SiScan and 3Seq and supported by significant p-values of 1.82E-11, 6.14E-7, 2.18E-11, 1.11E-11, and 1.66E-21, respectively. Besides, strain AHL was a reassortant virus in S segment, with significant p-values of 4.28E-7, 8.40E-8, 9.50E-13, 1.04E-5, 5.10E-4 and 2.41E-12 for RDP, GENECONV, Bootscan, Maxchi, Chimaera and SiScan methods.

Combined with the phylogenetic analysis, secondary screening was performed from the reassortment events analyzed by RDP4. Six strains were identified as reassortment events, including strains LN2012-14, LN2012-34, LN2012-41, LN2012-42, LN2012-58 and AHL. Four SFTSV strains (LN2012-14, LN2012-34, LN2012-41 and LN2012-42) were classified into lineage D in
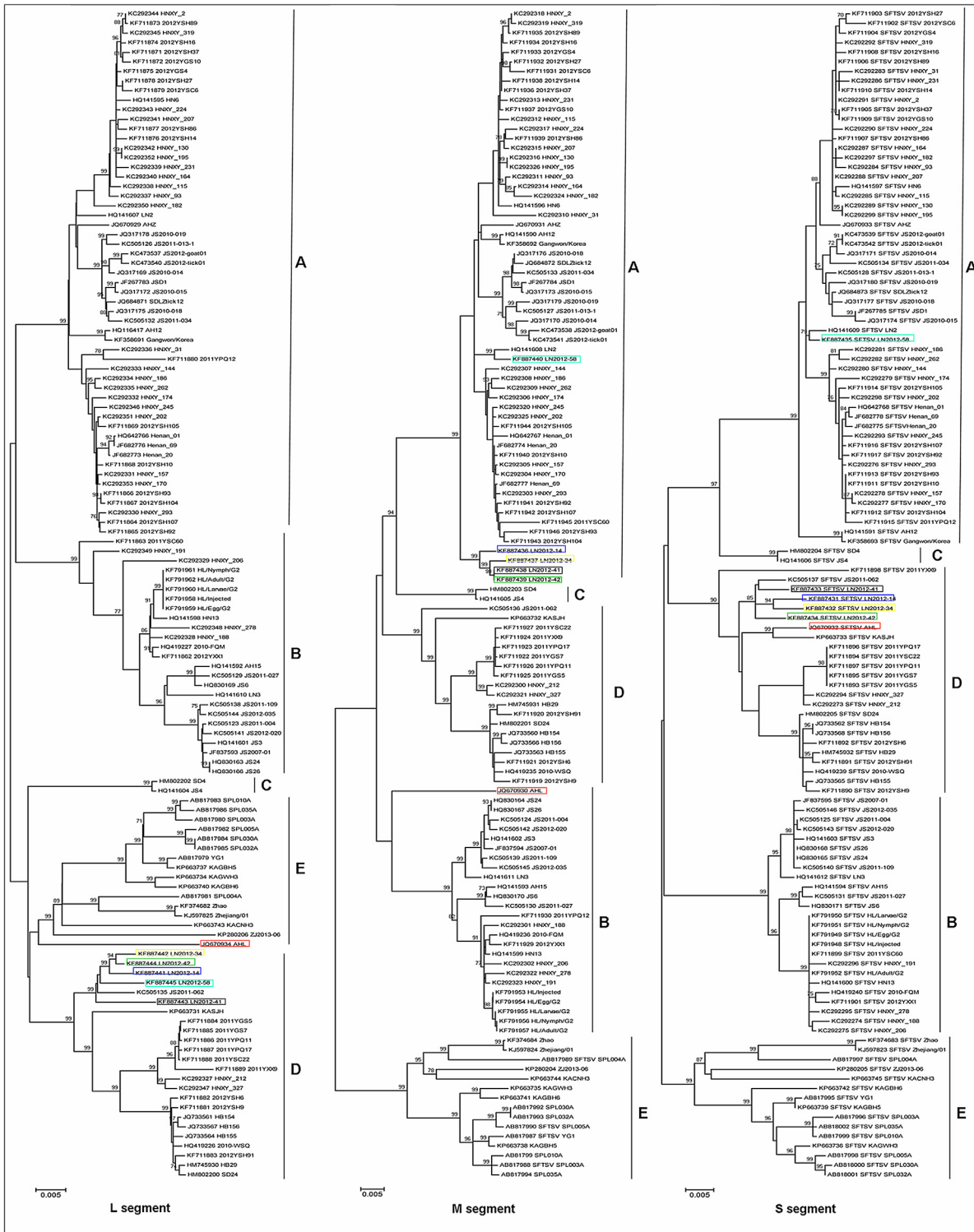
Fig 1. Phylogenetic analysis of the whole segment sequences of L, M, and S segments of 122 SFTSV strains. The maximum likelihood trees were constructed by using MEGA 5.02 software (http://www.megasoftware.net/). SFTSV was classified into 5 lineages labeled as A, B, C, D, and E by each genome segment. GenBank accession number and strain name were labeled on each branch. Bootstrap values ≧70 were labeled at nodes. Scale bar represented nucleotide substitutions per site.

doi:10.1371/journal.pone.0151677.g001

Table 1. Gene-specific global dN/dS ratios estimated using SLAC method.

Genes	dN/dS	95%HPD
RdRp	0.039	(0.035, 0.045)
GP	0.092	(0.082, 0.104)
N	0.026	(0.016, 0.039)
NSs	0.073	(0.056, 0.093)

doi:10.1371/journal.pone.0151677.t001

segment L and S, but they belonged to lineage A in segment M. Strain LN2012-58 belonged to lineage D in segment L, while it belonged to lineage A in segment M and S. Strain AHL belonged to lineage E in segment L, but it belonged to lineage B and D in segment M and S (Fig 1).

Selection Analysis

To assess the selection pressures acting on RdRp, glycoprotein (GP), N and NSs genes, the average dN/dS value measured by using SLAC on online server Datamonkey was shown in Table 1. The dN/dS values of the four genes ranged from 0.026 to 0.092, which indicated that SFTSV was under purifying selection. The highest dN/dS ratio was observed in the GP gene, followed by the NSs, RdRp, N genes.

Two sites (37, 1033) in glycoprotein were identified as being under strong positive selection with one using two different detective methods (SLAC, MEME), the other using four different detective methods (FEL, IFEL, MEME, FUBAR) (Table 2). Other 7 sites were identified as positive selection with only one method FEL or MEME. No positive selection sites were found in RdRp, N and NSs genes.

Estimating Evolutionary Rates

A conservative assessment of the degree of clock-like evolution present in a data set is achieved by fitting a regression of the year-of-sampling against the root-to-tip genetic distance of each sample, measured from the maximum likelihood tree. The regression coefficient (R^2) were 0.2504, 0.2838, 0.2231 and 0.1247 for the data set of concatenated sequence of L-M-S and individual segment sequence of L, M and S, respectively, which strongly supported the presence of molecular clock-like structure.

The average evolutionary rate of each genome segment was different, which was estimated by the Bayesian time-scaled phylogenetic analysis. For the L segment, the mean rate was 4.16E-4 substitutions per site per year (s/s/y) ranging from 8.99E-5 to 9.72E-4 (with 95% highest posterior densities, HPD). For the M segment, the mean rate was 6.76E-4 s/s/y (95%HPD = 3.92E-4 to 1.00E-3 s/s/y). The S segment had a mean rate of 1.09E-3 s/s/y (95%HPD = 5.43E-4 to 1.60E-3 s/s/y). The S segment had a highest evolutionary rate, which was 2.62 times of that of L segment and 1.61 times of that of M segment. To estimate the evolutionary rate of the SFTSV

Table 2. Positive selection sites analysis using SLAC, FEL, IFEL and REL methods.

Methods	SLAC*	FEL*	IFEL*	MEME*	FUBAR#
Positive sites	**1033**	**37**, 323	**37**	**37**, 524, 554, 642, 651, 926, 956, **1033**	**37**

Bold numbers denote amino acids sites under strong positive selection.

*: p-value $\leqq$ 0.05;

#: posterior probability $\geqq$ 0.95.

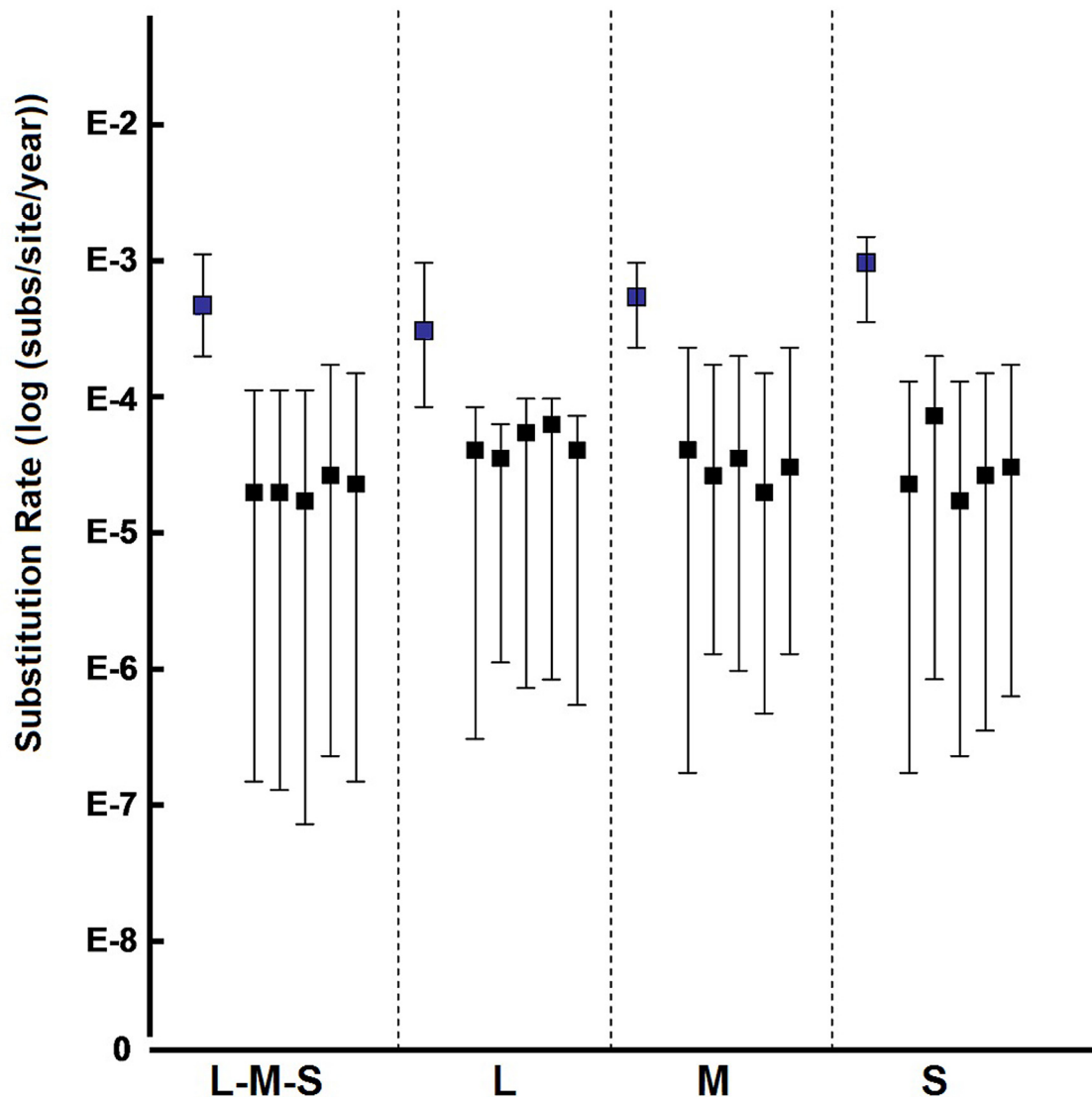doi:10.1371/journal.pone.0151677.t002

Fig 2. Posterior mean and 95% HPDs of the substitution rates estimated from the actual data sets and the 5 tip-date randomizations for the each data set. Substitution rates on the left for each data set were estimated from the actual data sets. Substitution rates on the right for each data set were estimated from the randomized data sets. The mean rates estimated for the data sets were significantly different from those estimated from the randomized data sets.

doi:10.1371/journal.pone.0151677.g002

complete genomes, the complete concatenated genomes were performed as described above. The mean rate of the concatenated sequences was estimated to be 6.73E-4 s/s/y (95% HPD = 2.35E-4 to 1.09E-3 s/s/y), which was similar to M segment. A Bayesian time-scaled MCC tree based on the complete concatenated genomes was constructed.

Randomized analysis suggested that, the mean evolutionary rates of the actual data did not fall in the 95% confidence intervals estimated from the randomized data sets (Fig 2).

Temporal and Spatial Dynamic Analysis

All the viral strains were divided into 5 lineages like the L segment described above (Fig 3). The estimated median TMRCA of SFTSV for all 5 lineages were found to be 1972 (95% HPD = 1927 to 1992). The TMRCA were estimated to be 1987 (95% HPD = 1962 to 2003), 1977 (95%

Fig 3. Time-scaled Bayesian MCC phylogenetic tree based on concatenated SFTSV complete genome sequences. Tree nodes were annotated with posterior probability values (right), estimated median dates of time to most recent common ancestor (TMRCA) and 95% confidence interval of TMRCA (above). Lineages (A, B, C, D and E) were marked with different colors. SFTSV strain names were labeled on each branch. Horizontal axis indicated time in years.

doi:10.1371/journal.pone.0151677.g003

HPD = 1937 to 1996), 1981 (95% HPD = 1947 to 1998), 1983 (95% HPD = 1953 to 2001) for A, B, D, E lineages respectively. Lineages A and C diverged from each other in 1978 (95% HPD = 1940 to 1996). In 1976 (95% HPD = 1930 to 1995), lineage D diverged from lineage A and C. The TMRCA were estimated to be 1974 (95% HPD = 1932 to 1994) for lineage B and E.

Temporal and spatial transmission routes were primarily confirmed using the Bayesian analysis based on the 122 complete concatenated genomes from China, Japan and South Korea. The extant SFTSV originated 42 (95% HPD = 20 to 87) years ago in the Dabie Mountain area in central China encompassing Henan, Hubei, Anhui, and Jiangsu provinces. Around 37 (95% HPD = 18 to 77) years ago, the virus moved to Shandong Province from Henan Province, and the virus was transmitted from Jiangsu Province to Liaoning Province in Northeastern China about 33 (95% HPD = 16 to 67) year ago. SFTSV spread to Zhoushan Archipelago of China, Jeju Island of South Korea, and Japan approximately 31 (95% HPD = 13 to 61) year ago, and there was mutual transmission among these areas.

Discussion

Phylogenetic analysis indicated that all 122 strains in East Asia were divided into five lineages in each segment. In each lineage, there were virus strains from China. SFTSV strains from the mainland of South Korea (Gangwon/Korea and KASJH) were classified into A and D lineages, while four other strains from Jeju Island of South Korea together with SFTSV strains from Zhoushan Archipelago of China and from Japan were classified into lineage E. This was similar to the previous study, which classified SFTSV virus into two clades, Chinese clade and Japanese

clade [31]. Bayesian analysis based on the complete concatenated genomes showed the same result, which further explained the close relationships among the viruses from the islands surrounding the East China Sea including the Zhoushan Archipelago of China and Jeju Island of South Korea, and the Japanese Archipelago. The case fatality rate of SFTS patients in Japan was apparently higher than that in South Korea and Zhejiang Province of China [2, 39]. It is not clear the high case fatality in Japan is because of high virulence of the Japanese SFTSV strains or because of small sample size of SFTS cases from Japan especially when retrospective study were used for diagnosis of SFTS, which might be focused on severe or dead SFTSV patients. Japanese SFTSV strains were classified into the Chinese clade from the mainland of China in the previous study [31], but we did not find this phenomenon because the strains had no complete genome sequences and were not analyzed in our study. Strain AHL belonged to lineage E in segment L, but it belonged to lineage B and D in segment M and S, respectively, suggesting that AHL was probably a reassortant. Our subsequent analysis and previous study have identified AHL as a reassortant [23, 24].

Reassortment is distinct from homologous recombination, which is widespread among viruses with segmented genomes, including the *Orthomyxovirus* and *Bunyavirus* [40, 41]. The reassortment event requires that the viruses must have at least two segmented genomes. When two or more segmented viruses co-infected a single cell simultaneously, the genome segments might be packaged into progeny viruses randomly. Then, the progeny might inherit genomic segments from more than one parent, obtaining increased genetic variability because the offspring might contain novel combinations of genomic segments. This process was an important reason for virus survival and diversification, as illustrated by the antigenic shift commonly observed in influenza virus, another negative strand segmented RNA virus [26]. Natural reassortment has been reported for various members of the *Phlebovirus* genus, including *Candiru* virus and Rift Valley fever virus [34, 42, 43]. Reassortment could increase pathogenicity and enhance transmissibility among vectors and hosts [44]. Six SFTSV strains were identified as reassortants in our analysis, which were isolated from human samples, suggesting that reassortment events occurred frequently in SFTSV hosts.

To assess the evolutionary substitution rate, TMRCA, and divergence of SFTSV lineages and the geographic origin of SFTSV, 122 complete genomes were analyzed by the Bayesian Markov chain Monte Carlo (MCMC) method. From a genetic perspective, substitution rates are critical parameters for understanding virus evolution, given that restrictions in genetic variation within a population of viruses can lead to low adaptability and pathogenicity [45]. Our analysis estimated that the evolutionary rate of three segments was different, with 4.16E-4 s/s/y (95%HPD = 8.99E-5 to 9.72E-4 s/s/y) in L segment, 6.76E-4 s/s/y (95%HPD = 3.92E-4 to 1.00E-3 s/s/y) in M segment and 1.09E-3 s/s/y (95%HPD = 5.43E-4 to 1.60E-3 s/s/y) in S segment, which was similar to that predicted for SFTSV in previous studies and other phleboviruses (E-4 s/s/y) [23, 32, 34]. The mean evolutionary rate of the concatenated sequences was estimated to be 6.73E-4 s/s/y (95%HPD = 2.35E-4 to 1.09E-3 s/s/y). Generally speaking, the M segment should have the highest evolutionary rate in bunyaviruses, as it encoded the two glycoproteins which play a role in interaction with receptors in the cell surface. However, higher substitution rates were observed in the S segment because of the variability seen at the nucleotide level, in the highly variable region of the NSs gene sequence for SFTSV.

Temporal and spatial dynamics of RNA viruses are often reflected by their phylogenetic structure [46]. The inference of divergence events presented facilitated a better understanding of historical divergence and offered further opportunities to study viral demographic history and dispersal events [27]. Maximum clade credibility phylogenetic tree supported the TMRCA for all sampled SFTSV strains were found to be 42 (95% HPD = 20 to 87) years ago, and lineage E isolates were predicted to have diverged at the latest time. SFTSV existed earlier than its first

description in central China and later in Japan and South Korea. The primary reason was that SFTSV was unknown, and lacked the corresponding detective methods and tools for the new virus. Secondly, changes in the working environment increased contact between the vectors and farmers. Lastly, virulence of the virus increased because of mutations and reassortment events. Our analysis indicated that central China was the geographic origin of the most recent common ancestor of SFTSV because of having the highest viral diversity in the area and viral strains distributing in four lineages sans lineage E. Furthermore, another aspect was the TMRCA from the MCC phylogenetic tree. In conclusion, these findings suggest that SFTSV originated in central China, which then spread to eastern China, northeastern China, and from Zhejiang Province in eastern China it spread to Japan and South Korea. Based on the genotype distribution of SFTSV and Bayesian analysis, Fu et al demonstrated that SFTSV strains in Zhoushan Islands were from central China and South Korea, and the virus in Japan were from South Korea [25]. In Fu's and our analysis, SFTSV strains in South Korea were classified into three lineages or genotypes. Within the three lineages, one included SFTSV strains only from Zhoushan Islands, South Korea, and Japan, while the other two included the virus from central and northeastern China, and South Korea. Thus, we inferred that SFTSV strains in the mainland of South Korea were transmitted from Liaoning, Shandong or Jiangsu province in China, while SFTSV strains in Jeju Island of South Korea were from Zhoushan Archipelago of Zhejiang Province of China. SFTSV in Japan were transmitted from Zhoushan Archipelago of Zhejiang Province of China or from Jeju Island of South Korea. However, while these predictions are suggestive of a potential origin for SFTSV, cautions must be exercised in their interpretation because limited SFTSV strains from limited areas were used in the analysis. In our study, only 8 strains from Japan and 6 from South Korea were used, which could affect the lineage classification of the phylogenetic analysis and the judgment of the virus origin and transmission routes. To obtain more accurate data of SFTSV lineage classification, origin and transmission routes, analysis should contain more complete genome sequences from South Korea and Japan. Although ticks were considered as a vector of SFTSV, ticks could not fly and with limited activity areas of the mammals, there must be a host to carry ticks from area to area efficiently. We inferred that the host should be migratory birds, but we had no evidence.

SFTSV glycoproteins exhibited the highest dN/dS ratio among the four genes. Theoretically, this might reflect the high antigenicity, which could be under selective pressure to evade the host humoral immune response as well as to adapt to their respective cell-surface receptors for attachment and entry [47]. Although we found two sites in the SFTSV glycoprotein under positive selection, we did not know if a mutation in the two sites could affect the virulence of the virus. Thus further research should be performed based on reverse genetics to identify the relationship between mutation of the sites and the virulence. The NS protein had the highest dN/dS ratio. We found that there were up to 5.2% diversity among all the amino acids of sampled NS genes. The NS protein is thought to affect virulence, primarily through its antagonism of interferon, helping the virus to evade host antiviral responses [18, 19]. The higher dN/dS ratio in NS gene might be the result from a co-evolutionary battle between the virus and host immunity.

Conclusions

We conclude that SFTSV originated in central China and spread to other places recently and the virus is under purifying selection with high frequency of reassortment.

Acknowledgments

Author Contributions

Conceived and designed the experiments: XJY LZ JWL. Performed the experiments: JWL. Analyzed the data: JWL XJY. Contributed reagents/materials/analysis tools: LML MML YS XS. Wrote the paper: JWL XJY.

References

1. Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL, et al. Fever with thrombocytopenia associated with a novel bunyavirus in China. N Engl J Med. 2011; 364(16):1523–32. Epub 2011/03/18. doi: 10.1056/NEJMoa1010095 PMID: 21410387; PubMed Central PMCID: PMC3113718.

2. Takahashi T, Maeda K, Suzuki T, Ishido A, Shigeoka T, Tominaga T, et al. The first identification and retrospective study of severe fever with thrombocytopenia syndrome in Japan. J Infect Dis. 2014; 209 (6):816–27. Epub 2013/11/16. jit603 [pii] doi: 10.1093/infdis/jit603 PMID: 24231186.

3. Kim KH, Yi J, Kim G, Choi SJ, Jun KI, Kim NH, et al. Severe fever with thrombocytopenia syndrome, South Korea, 2012. Emerg Infect Dis. 2013; 19(11):1892–4. Epub 2013/11/12. doi: 10.3201/eid1911.130792 PMID: 24206586; PubMed Central PMCID: PMC3837670.

4. Zhang X, Liu Y, Zhao L, Li B, Yu H, Wen H, et al. An emerging hemorrhagic fever in China caused by a novel bunyavirus SFTSV. Sci China Life Sci. 2013; 56(8):697–700. Epub 2013/08/07. doi: 10.1007/s11427-013-4518-9 PMID: 23917841.

5. Lei XY, Liu MM, Yu XJ. Severe fever with thrombocytopenia syndrome and its pathogen SFTSV. Microbes Infect. 2015; 17(2):149–54. Epub 2014/12/17. S1286-4579(14)00308-6 [pii] doi: 10.1016/j.micinf.2014.12.002 PMID: 25498868.

6. Wen HL, Zhao L, Zhai S, Chi Y, Cui F, Wang D, et al. Severe fever with thrombocytopenia syndrome, Shandong Province, China, 2011. Emerg Infect Dis. 2014; 20(1):1–5. Epub 2014/01/01. doi: 10.3201/eid2001.120532 PMID: 24378074; PubMed Central PMCID: PMC3884705.

7. McMullan LK, Folk SM, Kelly AJ, MacNeil A, Goldsmith CS, Metcalfe MG, et al. A new phlebovirus associated with severe febrile illness in Missouri. N Engl J Med. 2012; 367(9):834–41. Epub 2012/08/31. doi: 10.1056/NEJMoa1203378 PMID: 22931317.

8. Wang J, Selleck P, Yu M, Ha W, Rootes C, Gales R, et al. Novel phlebovirus with zoonotic potential isolated from ticks, Australia. Emerg Infect Dis. 2014; 20(6):1040–3. Epub 2014/05/27. doi: 10.3201/eid2006.140003 PMID: 24856477; PubMed Central PMCID: PMC4036776.

9. Mourya DT, Yadav PD, Basu A, Shete A, Patil DY, Zawar D, et al. Malsoor virus, a novel bat phlebovirus, is closely related to severe fever with thrombocytopenia syndrome virus and heartland virus. J Virol. 2014; 88(6):3605–9. Epub 2014/01/07. JVI.02617-13 [pii] doi: 10.1128/JVI.02617-13 PMID: 24390329; PubMed Central PMCID: PMC3957954.

10. Park SW, Song BG, Shin EH, Yun SM, Han MG, Park MY, et al. Prevalence of severe fever with thrombocytopenia syndrome virus in Haemaphysalis longicornis ticks in South Korea. Ticks and Tick Borne Dis. 2014; 5(6):975–7. Epub 2014/08/29. S1877-959X(14)00166-6 [pii] doi: 10.1016/j.ttbdis.2014.07.020 PMID: 25164614.

11. Luo LM, Zhao L, Wen HL, Zhang ZT, Liu JW, Fang LZ, et al. Haemaphysalis longicornis ticks as reservoir and vector of severe fever with thrombocytopenia syndrome virus in China. Emerg Infect Dis. 2015; 21(10):1770–6. Epub 2015/09/25. doi: 10.3201/eid2110.150126 PMID: 26402039.

12. Liu Y, Li Q, Hu W, Wu J, Wang Y, Mei L, et al. Person-to-person transmission of severe fever with thrombocytopenia syndrome virus. Vector Borne Zoonotic Dis. 2012; 12(2):156–60. Epub 2011/10/01. doi: 10.1089/vbz.2011.0758 PMID: 21955213.

13. Bao CJ, Guo XL, Qi X, Hu JL, Zhou MH, Varma JK, et al. A family cluster of infections by a newly recognized bunyavirus in eastern China, 2007: further evidence of person-to-person transmission. Clin Infect Dis. 2011; 53(12):1208–14. Epub 2011/10/27. cir732 [pii] doi: 10.1093/cid/cir732 PMID: 22028437.

14. Niu G, Li J, Liang M, Jiang X, Jiang M, Yin H, et al. Severe fever with thrombocytopenia syndrome virus among domesticated animals, China. Emerg Infect Dis. 2013; 19(5):756–63. Epub 2013/05/08. doi: 10.3201/eid1905.120245 PMID: 23648209; PubMed Central PMCID: PMC3647489.

15. Zhao L, Zhai S, Wen H, Cui F, Chi Y, Wang L, et al. Severe fever with thrombocytopenia syndrome virus, Shandong Province, China. Emerg Infect Dis. 2012; 18(6):963–5. Epub 2012/05/23. doi: 10.3201/eid1806.111345 PMID: 22608264; PubMed Central PMCID: PMC3358154.

16. Liu JW, Wen HL, Fang LZ, Zhang ZT, He ST, Xue ZF, et al. Prevalence of SFTSV among Asian house shrews and rodents, China, January-August 2013. Emerg Infect Dis. 2014; 20(12):2126–8. Epub 2014/11/25. doi: 10.3201/eid2012.141013 PMID: 25418111; PubMed Central PMCID: PMC4257798.

17. Ding S, Yin H, Xu X, Liu G, Jiang S, Wang W, et al. A cross-sectional survey of severe fever with thrombocytopenia syndrome virus infection of domestic animals in Laizhou City, Shandong Province, China.

Jpn J Infect Dis. 2014; 67(1):1–4. Epub 2014/01/24. DN/JST.JSTAGE/yoken/67.1 [pii]. PMID: 24451093.

18. Ning YJ, Feng K, Min YQ, Cao WC, Wang M, Deng F, et al. Disruption of type I interferon signaling by the nonstructural protein of severe fever with thrombocytopenia syndrome virus via the hijacking of STAT2 and STAT1 into inclusion bodies. J Virol. 2015; 89(8):4227–36. Epub 2015/01/30. JVI.00154-15 [pii] doi: 10.1128/JVI.00154-15 PMID: 25631085; PubMed Central PMCID: PMC4442386.

19. Qu B, Qi X, Wu X, Liang M, Li C, Cardona CJ, et al. Suppression of the interferon and NF-kappaB responses by severe fever with thrombocytopenia syndrome virus. J Virol. 2012; 86(16):8388–401. Epub 2012/05/25. JVI.00612-12 [pii] doi: 10.1128/JVI.00612-12 PMID: 22623799; PubMed Central PMCID: PMC3421730.

20. Santiago FW, Covaleda LM, Sanchez-Aparicio MT, Silvas JA, Diaz-Vizarreta AC, Patel JR, et al. Hijacking of RIG-I signaling proteins into virus-induced cytoplasmic structures correlates with the inhibition of type I interferon responses. J Virol. 2014; 88(8):4572–85. Epub 2014/01/31. JVI.03021-13 [pii] doi: 10.1128/JVI.03021-13 PMID: 24478431; PubMed Central PMCID: PMC3993744.

21. Matsuno K, Weisend C, Travassos da Rosa AP, Anzick SL, Dahlstrom E, Porcella SF, et al. Characterization of the Bhanja serogroup viruses (Bunyaviridae): a novel species of the genus Phlebovirus and its relationship with other emerging tick-borne phleboviruses. J Virol. 2013; 87(7):3719–28. Epub 2013/01/18. JVI.02845-12 [pii] doi: 10.1128/JVI.02845-12 PMID: 23325688; PubMed Central PMCID: PMC3624231.

22. Kuiken T, Holmes EC, McCauley J, Rimmelzwaan GF, Williams CS, Grenfell BT. Host species barriers to influenza virus infections. Science. 2006; 312(5772):394–7. Epub 2006/04/22. 312/5772/394 [pii] doi: 10.1126/science.1122818 PMID: 16627737.

23. Huang X, Liu L, Du Y, Wu W, Wang H, Su J, et al. The evolutionary history and spatiotemporal dynamics of the fever, thrombocytopenia and leukocytopenia syndrome virus (FTLSV) in China. PLoS Negl Trop Dis. 2014; 8(10):e3237. Epub 2014/10/21. doi: 10.1371/journal.pntd.0003237 PNTD-D-14-00357 [pii]. PMID: 25329580; PubMed Central PMCID: PMC4199521.

24. Ding NZ, Luo ZF, Niu DD, Ji W, Kang XH, Cai SS, et al. Identification of two severe fever with thrombocytopenia syndrome virus strains originating from reassortment. Virus Res. 2013; 178(2):543–6. Epub 2013/09/24. S0168-1702(13)00309-2 [pii] doi: 10.1016/j.virusres.2013.09.017 PMID: 24055465.

25. Fu Y, Li S, Zhang Z, Man S, Li X, Zhang W, et al. Phylogeographic analysis of severe fever with thrombocytopenia syndrome virus from Zhoushan Islands, China: implication for transmission across the ocean. Sci Rep. 2016; 6:19563. Epub 2016/01/26. doi: 10.1038/srep19563 PMID: 26806841; PubMed Central PMCID: PMCPMC4726339.

26. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009; 459(7250):1122–5. Epub 2009/06/12. nature08182 [pii] doi: 10.1038/nature08182 PMID: 19516283.

27. Muniraju M, Munir M, Parthiban AR, Banyard AC, Bao J, Wang Z, et al. Molecular evolution of peste des petits ruminants virus. Emerg Infect Dis. 2014; 20(12):2023–33. Epub 2014/11/25. doi: 10.3201/eid2012.140684 PMID: 25418782; PubMed Central PMCID: PMC4257836.

28. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature. 2015; 524(7563):97–101. Epub 2015/06/18. nature14594 [pii] doi: 10.1038/nature14594 PMID: 26083749.

29. Carroll SA, Towner JS, Sealy TK, McMullan LK, Khristova ML, Burt FJ, et al. Molecular evolution of viruses of the family Filoviridae based on 97 whole-genome sequences. J Virol. 2013; 87(5):2608–16. Epub 2012/12/21. JVI.03118-12 [pii] doi: 10.1128/JVI.03118-12 PMID: 23255795; PubMed Central PMCID: PMC3571414.

30. Tong YG, Shi WF, Liu D, Qian J, Liang L, Bo XC, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature. 2015; 524(7563):93–6. Epub 2015/05/15. nature14490 [pii] doi: 10.1038/nature14490 PMID: 25970247.

31. Yoshikawa T, Shimojima M, Fukushi S, Tani H, Fukuma A, Taniguchi S, et al. Phylogenetic and geographic relationships of severe fever with thrombocytopenia syndrome virus in China, South Korea, and Japan. J Infect Dis. 2015. Epub 2015/03/13. jiv144 [pii] doi: 10.1093/infdis/jiv144 PMID: 25762790.

32. Lam TT, Liu W, Bowden TA, Cui N, Zhuang L, Liu K, et al. Evolutionary and molecular analysis of the emergent severe fever with thrombocytopenia syndrome virus. Epidemics. 2013; 5(1):1–10. Epub 2013/02/27. S1755-4365(12)00046-1 [pii] doi: 10.1016/j.epidem.2012.09.002 PMID: 23438426; PubMed Central PMCID: PMC4330987.

33. Faye O, Freire CC, Iamarino A, de Oliveira JV, Diallo M, Zanotto PM, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl Trop Dis. 2014; 8(1):e2636. Epub 2014/01/15. doi: 10.1371/journal.pntd.0002636 PNTD-D-13-00377 [pii]. PMID: 24421913; PubMed Central PMCID: PMC3888466.

34. Freire CC, Iamarino A, Soumare PO, Faye O, Sall AA, Zanotto PM. Reassortment and distinct evolutionary dynamics of Rift Valley fever virus genomic segments. Sci Rep. 2015; 5:11353. Epub 2015/06/24. srep11353 [pii] doi: 10.1038/srep11353 PMID: 26100494; PubMed Central PMCID: PMC4477411.

35. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21 (5):676–9. Epub 2004/10/29. bti079 [pii] doi: 10.1093/bioinformatics/bti079 PMID: 15509596.

36. Lopes AM, Capucci L, Gavier-Widen D, Le Gall-Recule G, Brocchi E, Barbieri I, et al. Molecular evolution and antigenic variation of European brown hare syndrome virus (EBHSV). Virology. 2014; 468–470:104–12. Epub 2014/08/27. S0042-6822(14)00373-0 [pii] PMID: 25155199.

37. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Molecular biology and evolution. 2010; 27 (9):2038–51. Epub 2010/04/07. doi: 10.1093/molbev/msq088 PMID: 20363828; PubMed Central PMCID: PMCPMC3107591.

38. Ramsden C, Holmes EC, Charleston MA. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Molecular biology and evolution. 2009; 26(1):143–53. Epub 2008/10/17. doi: 10.1093/molbev/msn234 PMID: 18922760.

39. Zhang L, Ye L, Ojcius DM, Lou X, Wang C, feng C, et al. Characterization of severe fever with thrombocytopenia syndrome in rural regions of Zhejiang, China. PLoS One. 2014; 9(10):e111127. Epub 2014/10/31. doi: 10.1371/journal.pone.0111127 PONE-D-14-30025 [pii]. PMID: 25356556; PubMed Central PMCID: PMC4214719.

40. Schrauwen EJ, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA, Herfst S. Reassortment between Avian H5N1 and human influenza viruses is mainly restricted to the matrix and neuraminidase gene segments. PLoS One. 2013; 8(3):e59889. Epub 2013/03/26. doi: 10.1371/journal.pone.0059889 PONE-D-12-35235 [pii]. PMID: 23527283; PubMed Central PMCID: PMC3604002.

41. Gerrard SR, Li L, Barrett AD, Nichol ST. Ngari virus is a Bunyamwera virus reassortant that can be associated with large outbreaks of hemorrhagic fever in Africa. J Virol. 2004; 78(16):8922–6. Epub 2004/07/29. doi: 10.1128/JVI.78.16.8922-8926.2004 78/16/8922 [pii]. PMID: 15280501; PubMed Central PMCID: PMC479050.

42. Palacios G, Tesh R, Travassos da Rosa A, Savji N, Sze W, Jain K, et al. Characterization of the Candiru antigenic complex (Bunyaviridae: Phlebovirus), a highly diverse and reassorting group of viruses affecting humans in tropical America. J Virol. 2011; 85(8):3811–20. Epub 2011/02/04. JVI.02275-10 [pii] doi: 10.1128/JVI.02275-10 PMID: 21289119; PubMed Central PMCID: PMC3126144.

43. Perrone LA, Narayanan K, Worthy M, Peters CJ. The S segment of Punta Toro virus (Bunyaviridae, Phlebovirus) is a major determinant of lethality in the Syrian hamster and codes for a type I interferon antagonist. J Virol. 2007; 81(2):884–92. Epub 2006/10/20. JVI.01074-06 [pii] doi: 10.1128/JVI.01074-06 PMID: 17050607; PubMed Central PMCID: PMC1797479.

44. Rodriguez LL, Owens JH, Peters CJ, Nichol ST. Genetic reassortment among viruses causing hantavirus pulmonary syndrome. Virology. 1998; 242(1):99–106. Epub 1998/03/17. S0042-6822(97)98990-X [pii] doi: 10.1006/viro.1997.8990 PMID: 9501041.

45. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. 2011; 8(2):270–9. Epub 2011/05/20. 15013 [pii]. PMID: 21593585; PubMed Central PMCID: PMC3127101.

46. Biek R, Drummond AJ, Poss M. A virus reveals population structure and recent demographic history of its carnivore host. Science. 2006; 311(5760):538–41. Epub 2006/01/28. 311/5760/538 [pii] doi: 10.1126/science.1121360 PMID: 16439664.

47. Bowden TA, Jones EY, Stuart DI. Cells under siege: viral glycoprotein interactions at the cell surface. J Struct Biol. 2011; 175(2):120–6. Epub 2011/03/29. S1047-8477(11)00081-5 [pii] doi: 10.1016/j.jsb.2011.03.016 PMID: 21440638; PubMed Central PMCID: PMC3137789.