The classification, origin, and evolutionary dynamics of severe fever with thrombocytopenia syndrome virus circulating in East Asia

Shaowei Sang [ID][1,2,*], Peng Chen[3], Chuanxi Li[1], Anran Zhang[1], Yiguan Wang[4,*], Qiyong Liu[5,*]

[1]Clinical Epidemiology Unit, Qilu Hospital of Shandong University, No. 107 Wenhua Road, Lixia District, Jinan 250012, People's Republic of China

[2]Clinical Research Center of Shandong University, No. 107 Wenhua Road, Lixia District, Jinan 250012, People's Republic of China

[3]Department of Healthcare-associated Infection Management, School and Hospital of Stomatology, Shandong University, No. 44-1 Wenhua Road, Lixia District, Jinan 250012, People's Republic of China

[4]CAS Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, No. 300 Fenglin Road, Xuhui District, Shanghai 200032, People's Republic of China

[5]National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, WHO Collaborating Centre for Vector Surveillance and Management, No. 55 Changbai Road, Changping District, Beijing 102206, People's Republic of China

*Corresponding author. Shaowei Sang, Clinical Epidemiology Unit, Qilu Hospital of Shandong University, No. 107 Wenhua Road, Lixia District, Jinan, Shandong 250012, China. E-mail: sangshaowei@sdu.edu.cn; Yiguan Wang, CAS Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, No. 300 Fenglin Road, Xuhui District, Shanghai 200032, China. E-mail: wang.yiguan@163.com; Qiyong Liu, National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, WHO Collaborating Centre for Vector Surveillance and Management, No. 55 Changbai Road, Changping District, Changping, 102206, Beijing People's Republic of China. E-mail: liuqiyong@icdc.cn

Abstract

The classification of severe fever with thrombocytopenia syndrome virus (SFTSV) lacked consistency due to limited virus sequences used across previous studies, and the origin and transmission dynamics of the SFTSV remains not fully understood. In this study, we analyzed the diversity and phylodynamics of SFTSV using the most comprehensive and largest dataset publicly available for a better understanding of SFTSV classification and transmission. A total of 1267 L segments, 1289 M segments, and 1438 S segments collected from China, South Korea, and Japan were included in this study. Maximum likelihood trees were reconstructed to classify the lineages. Discrete phylogeographic analysis was conducted to infer the phylodynamics of SFTSV. We found that the L, M, and S segments were highly conserved, with mean pairwise nucleotide distances of 2.80, 3.36, and 3.35% and could be separated into 16, 13, and 15 lineages, respectively. The evolutionary rate for L, M, and the S segment was 0.61×10^{-4} (95% HPD: 0.48–0.73×10^{-4}), 1.31×10^{-4} (95% HPD: 0.77–1.77×10^{-4}) and 1.27×10^{-4} (95% HPD: 0.65–1.85×10^{-4}) subs/site/year. The SFTSV most likely originated from South Korea around the year of 1617.6 (95% HPD: 1513.1–1724.3), 1700.4 (95% HPD: 1493.7–1814.0), and 1790.1 (95% HPD: 1605.4–1887.2) for L, M, and S segments, respectively. Hubei Province in China played a critical role in the geographical expansion of the SFTSV. The effective population size of SFTSV peaked around 2010 to 2013. We also identified several codons under positive selection in the RdRp, Gn–Gc, and NS genes. By leveraging the largest dataset of SFTSV, our analysis could provide new insights into the evolution and dispersal of SFTSV, which may be beneficial for the control and prevention of severe fever with thrombocytopenia syndrome.

Keywords: severe fever with thrombocytopenia syndrome; diversity; origin; phylodynamics; selective pressure

Introduction

Severe fever with thrombocytopenia syndrome (SFTS) is an emerging infectious disease that was first reported in 2009 in rural areas of Hubei Province and Henan Province, China (Yu et al. 2011). SFTS cases were subsequently reported in Japan and South Korea in 2012 (Kim et al. 2013, Takahashi et al. 2014). Although sporadic cases were also found in Vietnam (Tran et al. 2019) and Thailand (Rattanakomol et al. 2022), surveillance data showed that SFTS cases were predominantly from China, South Korea, and Japan. The cumulative number of SFTS cases was 7721 in China by 2018 (Miao et al. 2021), 1697 in South Korea by 2022 (Diseases, K.C.f.P.a.C.o. 2023b), and 835 in Japan by April 2023 (Diseases, J.N.I.o.I. 2023a). In China, SFTS has been becoming prevalent since the first identification, with the number of provinces reporting SFTS cases increasing from 2 in 2009 to 11 in 2012 (Liu et al. 2014) and 25 in 2018 (Miao et al. 2021).

SFTS virus (SFTSV), the causative agent of SFTS, was assigned to the genus *Bandavirus* in the family *Phenuiviridae* of *Bunyavirales* by the International Committee on Taxonomy of Viruses (ICTV). SFTSV is an enveloped RNA virus consisting of three-stranded negative-sense RNA segments: large (L), medium (M), and small (S). The L segment is 6368 nucleotides long, encoding a viral RNA-dependent RNA polymerase (RdRp) (2084 amino acids). The M segment is 3378 nucleotides long, encoding 1073 amino acid precursors of glycoproteins (Gn and Gc). The S segment contains 1744 nucleotides, encoding a nucleoprotein (N) and a nonstructural protein (NS) (Yu et al. 2011).

SFTSV is primarily transmitted to humans through the bite of infected ticks, with *Haemaphysalis longicornis* being the main vector. The virus circulates in a zoonotic cycle involving ticks and various animal hosts, including livestock and wild animals (Li et al. 2022), with the potential risk of infecting humans. However, human-to-human transmission of SFTSV is extremely rare and not conclusively proven in most cases. Although there have been some reports suggesting possible human-to-human transmission in hospital settings or through close contact with infected individuals' blood or bodily fluids (Jiang et al. 2015, Jung et al. 2019, Wu et al. 2022), these instances are exceptional and not considered a significant route of transmission.

Currently, the SFTSV classification remains controversial. Yoshikawa et al. classified SFTSV into C1 to C5 and J1 to J3 by the L segment (140 sequences) or M segment (171 sequences), and C1 to C4 and J1 to J3 by the S segment (211 sequences) according to geographical distribution (Yoshikawa et al. 2015). Using the same nomenclature, Lv identified one more genotype (C6) in the L segment (227 sequences) and one more genotype (J4) in the S segment (436 sequences) (Lv et al. 2017). Fu genotyped SFTSV as A–F by the L segment (126 sequences) or M segment (129 sequences) and A–B, D, and E–F by the S segment (166 sequences), according to genetic distance (Fu et al. 2016). Yun further separated B into B1 to B3 but left some strains unclassified (Yun et al. 2020b). Liu classified SFTSV into five lineages, named A to E, for each segment (Liu et al. 2016a). In addition, the estimated evolutionary rate and the time to the most recent common ancestor (TMRCA) of the SFTSV varied significantly among different studies (Fu et al. 2016, Liu et al. 2016a, Yun et al. 2020a). The genetic diversity and phylodynamics of SFTSV also exhibited discrepancies, which may be attributed to the relatively small sample sizes of the sequences used in previous studies. With the accumulation of SFTSV sequences submitted to GenBank, we were able to systematically and comprehensively study the evolutionary dynamics of SFTS in East Asia.

Methods
Data collection
Complete L, M, and S segments of SFTSV were separately retrieved from GenBank (as of 28 January 2023) with accession numbers, collection dates, and locations. Our dataset consisted exclusively of human-derived sequences. Sequences from other hosts were not included in this study. Sequences from the same strain of the specific segment, unverified sequences and sequences for unidentified sources (e.g. modified strains for vaccine development) were excluded. The full genome of SFTS viruses from the same strain was identified according to the strain name and associated information (such as location, collection date, authors, etc.) in GenBank.

Phylogenetic analysis
Multiple sequence alignment was performed using MAFFT 7 (Katoh and Standley 2013). L, M, and S segments were analyzed separately to detect recombination events using RDP 4 (Martin et al. 2015). A total of seven methods, including RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, and 3Seq, implemented in RDP 4 were used. Recombination detected by at least four of the seven methods with a *P*-value cutoff of .01 corrected with the Bonferroni method was considered recombination. The inferred recombinant sequences were removed, and the procedure was repeated until no more recombination events were detected, leaving 1267 L sequences, 1289 metre sequences and 1438 S sequences (Supplementary Table S1-S3). Then, 1147 strains with complete L, M, and S segments were identified (Supplementary Table S4). The *p*-distances of the L, M, and S nucleotide sequences, including the pairwise distances, overall mean distances, and within- and between-lineage mean distances, were computed using Mega 11 (Tamura et al. 2021).

Maximum likelihood (ML) trees were reconstructed in IQ-TREE 2 (Minh et al. 2020), and 1000 replicates of both the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) and ultra-fast bootstrap approximation (UFboot) (Hoang et al. 2018) were performed to assess tree branch support. The nucleotide substitution models GTR + F + R5 for the L segment, TVM + F + R4 for the M segment, and TVMe + R4 for the S segment were determined and used according to the Bayesian information criterion (BIC) from ModelFinder (Kalyaanamoorthy et al. 2017). The final lineage determination was based on comprehensive phylogenetic analysis, which considered the phylogenetic tree structure, node support values, and genetic distance. The ggtree package in R 4.1.0 was used for visualization and annotation of ML trees (Yu et al. 2018).

Discrete phylogeographic analysis
To infer the phylogeographic history, nucleotide substitution rate and TMRCA of the SFTSV, discrete phylogeographic analysis was independently conducted using the L, M and S segments from the same strain with known dates and locations (country or province in China). To reduce bias from the unbalanced sample size distribution, particularly the overrepresentation of samples from Henan and Hubei provinces, we used ClusterPicker software to downsample these regions. The parameters for ClusterPicker were set at 80% for bootstrap cutoff and 0.5% for within-cluster genetic distance (Ragonnet-Cronin et al. 2013). ClusterPicker identified clusters of genetically similar sequences from Henan and Hubei provinces. From each identified cluster, we randomly sampled one representative strain. This process ensured that the downsampled dataset was more balanced and representative of the overall geographic distribution of SFTSV strains, while maintaining genetic diversity. After downsampling, the dataset for each segment (L, M, and S) consisted of 425 sequences (Supplementary Table S5). Discrete phylogeographic analysis was then performed separately for each segment using BEAST 1.10.4 (Suchard et al. 2018) with the BEAGLE library to improve computational performance. For the analysis, we specified a GTR + I + Γ4 (general time-reversible model with a proportion of invariable sites and four gamma-distributed rate categories) model of nucleotide substitution, a flexible nonparametric skygrid coalescent model, and a relaxed uncorrelated lognormal (UCLN) molecular clock (Drummond et al. 2006). Reconstruction of the SFTSV dispersal

history at the regional level was inferred by defining a continuous-time Markov chain (CTMC) model of discrete geographic traits (i.e. countries or provinces). The Bayesian stochastic search variable selection (BSSVS) procedure was used to computationally infer ancestral migrations efficiently by determining the network of nonzero transition rates from the CTMC symmetric matrix (Lemey et al. 2009). Tracer 1.6 was used to inspect the convergence of the MCMC chains, ensuring that sufficient sampling was achieved (effective sample size > 200). TreeAnnotator 1.10.4 was used to infer the maximum clade credibility (MCC) summary tree with a 10% burn-in. The transmission network inferred from the MCC tree was visualized using StrainHub (de Bernardi Schneider et al. 2020). Historical viral population dynamics for SFTSV were reconstructed using the aforementioned flexible nonparametric skyride coalescent.

Selective pressure

HM745930-HM745932 (strain HB29) reference sequences were used to align the codon positions and identify sites under selective pressure. The underlying selection pressure on the nonrecombinant SFTSV coding sequence (CDS) was examined using Datamonkey 2.0 (Weaver et al. 2018). The codon sites were scanned using four different methods, including mixed effects model of evolution (MEME) (Murrell et al. 2012), fixed effects likelihood (FEL) (Kosakovsky Pond and Frost 2005), single-likelihood ancestor counting (SLAC) (Kosakovsky Pond and Frost 2005), and fast, unconstrained Bayesian approximation (FUBAR) (Murrell et al. 2013). In the cases of MEME, FEL, and SLAC, codons with P-values <.1 suggested regions under positive selection, and a posterior probability >.9 was considered positive for FUBAR. In this study, sites were considered under positive selection, if supported by at least three methods. The positive selection sites in the RdRp and Gn–GC 3D structures (PDB: 6Y6K, RdRp; 7X6U, and Gn–GC) were visualized using PyMOL 2.6.

Ethics statement

The SFTSV sequences were downloaded from GenBank, and the privacy of the patients was not included in the data. Therefore, institutional review was exempted.

Results

Epidemiology of the SFTSV

We obtained a total of 3994 SFTSV sequences from GenBank, including 1267 L segment, 1289 M segment, and 1438 S segment. These sequences were predominantly from China, followed by South Korea and Japan (Fig. 1a). The collection date ranged from 2005 to 2022, with most sequences submitted between 2013 and 2015 in China, in 2013 in Japan, and in 2016 in South Korea (Fig. 1b–d). In China, SFTSV sequences were collected from 10 provinces (Fig. 1a), with 73.2% (2695/3681) and 12.1% (445/3681) originating from Henan Province and Hubei Province, respectively (Fig. 1e–g).

Phylogenetic analysis of the L, M, and S segments

The overall mean pairwise distances of the L, M, and S segments were 2.80, 3.36, and 3.35%, respectively, and the pairwise differences ranged from 0% to 4.90%, 0% to 7.02%, and 0% to 6.37% (Fig. 2), respectively.

The ML tree showed that the L sequences were separated into 16 lineages (A–P), with SH-aLRT and UFboot values on the lineages both equal to 100% (Fig. 2). The within-lineage mean distance was <2%, and the between-lineage mean distance ranged from 3.23% to 4.54% (Fig. 3). Ten lineages were detected in South Korea

from 2012 to 2017, with Lineages G and J cocirculating during this period, and seven lineages cocirculating in 2016. Five lineages were detected in Japan from 2005 to 2018, with Lineage G being the dominant lineage. Twelve lineages were detected in China from 2010 to 2022, among which seven lineages were detected in Hubei and Zhejiang provinces, followed by six lineages in Henan Province. The dominant lineage in Henan Province was Lineage A, followed by Lineages C and O, while in Hubei Province, the dominant lineage was Lineage A, followed by Lineage O (Fig. 4).

The M sequences were separated into 13 lineages, with Lineages D, I and M missing compared to the L segment. Lineages I and M for the L segment were classified as Lineage G for the M segment. Both the SH-aLRT and UFboot values for the lineages were >95% (Fig. 2). The within-lineage mean distance was <2.12%, and the between-lineage mean distance ranged from 3.43% to 6.52% (Fig. 3). Eight lineages were detected in South Korea from 2012 to 2017, four lineages in Japan from 2005 to 2018, and eleven lineages in China from 2010 to 2022. The molecular epidemiology pattern of the M segment was similar to that of the L segment (Fig. 4).

The S sequences were separated into 15 lineages, with Lineages D and I missing but Lineages Q and R emerging compared with the L segment. Lineage I for the L segment was classified as Lineage G for the S segment, and Lineage R for the S segment was classified as Lineage J for the L and M segments. Both the SH-aLRT and UFboot values of the lineages were >95%, except for those of Lineage P (Fig. 2). The within-lineage mean distance was <2%, and the between-lineage mean distance ranged from 2.93% to 5.61% (Fig. 3). Ten lineages were detected in South Korea from 2012 to 2017, five lineages in Japan from 2005 to 2018, and twelve lineages in China from 2010 to 2022. The molecular epidemiology pattern of the S segment was similar to that of the L and M segments (Fig. 4).

In this study, the ML trees of the L, M, and S segments indicated 48 reassortants, accounting for 4.18% (48/1,147). Among these reassortants, L reassortants accounted for 41.77% (20/48), M reassortants accounted for 22.92% (11/48), and S reassortants accounted for 33.33% (16/48) (Supplementary Table S4).

Phylogeographic analysis of the L, M, and S segments

The overall evolutionary rates for the L, M and S segments were 0.61×10^{-4} (95% highest probability density (HPD): 0.48–0.73×10^{-4}), 1.31×10^{-4} (95% HPD: 0.77–1.77×10^{-4}) and 1.27×10^{-4} (95% HPD: 0.65–1.85×10^{-4}) subs/site/year, respectively. The MCC trees of the L, M and S segments consistently indicated that the SFTSV most likely originated from South Korea, with posterior location probabilities of 0.93, 0.93 and 0.72, respectively. Except for the Lineages P and K, the ancestral locations of same lineages inferred from L, M, and S segments were consistent. The TMRCAs of the SFTSV inferred from the L, M, and S segments were 1617.6 (95% HPD: 1513.1–1724.3), 1700.4 (95% HPD: 1493.7–1814.0), and 1790.1 (95% HPD: 1605.4–1887.2), respectively. Except for the four lineages formed by only one sequence (lineages F, L, N, and P), the TMRCAs of the other lineages were mainly around between the late 1700s and early 1900s (Fig. 5a). South Korea contributed to the spread of SFTSV among countries. Hubei Province played a critical role in the geographical expansion of the SFTSV in China, followed by Henan Province and Jiangsu Province. Transitions between Hubei Province and Henan Province were most frequent (Fig. 5b, Supplementary S4, S5). The demographic history was relatively stable, increased around 1990–2000, peaked between approximately 2010 and 2013, and then decreased (Fig. 5c).
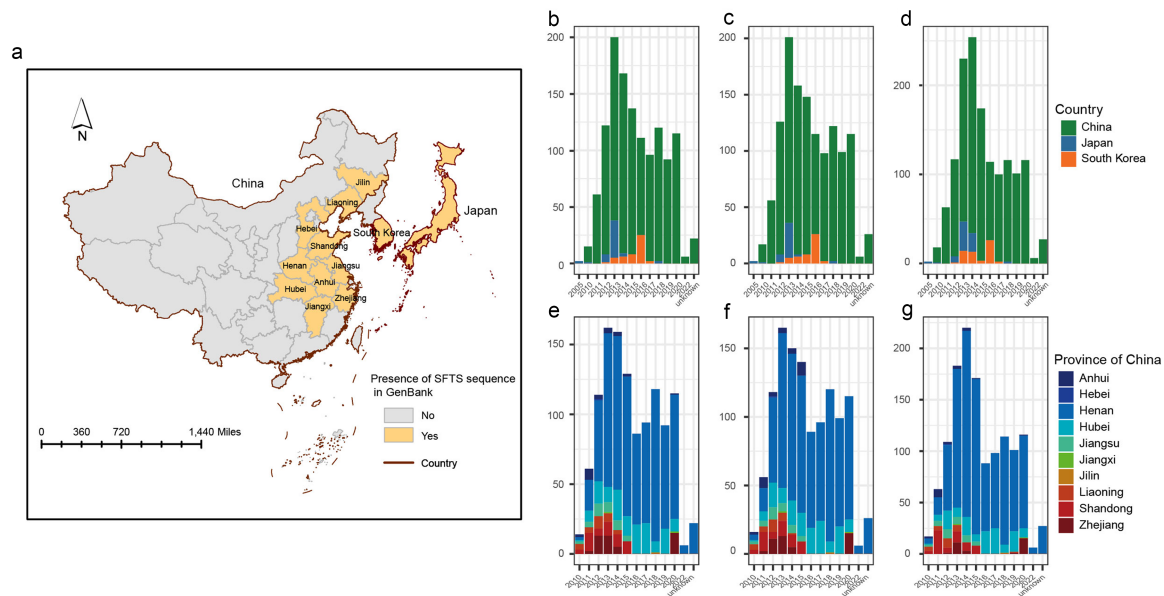
Figure 1. The epidemiology of SFTSV sequences from GenBank. (a) Spatial distribution of SFTSV sequences; (b–d) spatial and temporal distributions of the L, M, and S segments at the country level, respectively; (e–g) spatial and temporal distributions of the L, M, and S segments at the province level, respectively, in China.

Protein structure analysis of adaptive evolutionary sites

Selection pressure analysis revealed multiple codons under positive selection across different genes of the SFTSV genome. In the RdRp gene, codons 2, 1061, 1116, and 1353 were consistently identified under positive selection using four different methods. The Gn–Gc gene showed positive selection at codons 170 and 298, with additional evidence for codons 79 and 984 detected by three methods. In the NS gene, codons 145, 272, and 289 were consistently identified under positive selection, and codon 238 was detected by three methods (Supplementary Table S6, Fig. 6). There is no 3D structure of the NS gene in the PDB database and we are unable to visualize NS gene structure here.

Discussion

In this study, we conducted phylogenetic analysis, phylogeographic analysis, and selective pressure analysis of SFTSV using the most comprehensive and largest dataset available. This dataset encompassed strains from a broad geographic range spanning 18 years from 2005 to 2022. There is no universal approach for classifying viral genetic diversity below the level of a virus species, which is not covered by the ICTV. Published studies on the diversity of SFTSV have used a variety of terms for classification, such as genotype, lineage, and clade. In this study, we established a new nomenclature system for SFTSV based on the genetic distance threshold.

The SFTSV genome sequences showed high conservation, with the L segment and M segment exhibiting the highest and lowest levels of conservation, respectively, and evolutionary rate of M segment was 2.1 times faster of L segment. Previous studies analyzing the diversity of SFTSV have used varying numbers of sequences, partial or whole-genome segments, and different methods for classification, which has led to different genotyping results. Some studies classified genotypes based on the topology of the phylogenetic tree; for example, Yoshikawa et al. named clusters the Chinese lineage and Japanese lineage (Yoshikawa et al.

2015). Fu et al. separated SFTSV into A–F genotypes, with the mean within-lineage distance 0.1–2.6%, and the mean between-lineage distance 3.5–6.2%. Yun et al. further classified the B genotype into B1, B2, and B3 genotypes and left some strains unclassified (Yun et al. 2020b), but did not quantify the mean within- and between-lineage genetic distances. Our study used a large dataset (1267 L sequences, 1289 m sequences, and 1438 S sequences) and classified the lineages based on comprehensive phylogenetic analysis, which considered the phylogenetic tree structure, node support values, and genetic distance. Our classification identified 16, 13, and 15 lineages for L, M, and S segments, respectively, which was significantly more detailed than previous systems We obtained similar topologies using MCC trees, suggesting the robustness of our results. Our classification system used specific genetic distance criteria for lineage definition, with a mean within-lineage distance of <2.1%, and a mean between-lineage distance of 3.0–6.5%, which was more refined than previous studies. Additionally, pairwise distance analysis revealed that the lineages separated well (Fig. 2). Our results of phylogenetic analysis of SFTSV are comparable to those of Rift Valley fever virus (RVFV), a mosquito-borne virus of the family *Bunyaviridae* and genus *Phlebovirus*. SFTSV and RVFV have similar RNA genome structures. The M segment sequences of RVFV were also highly conserved, with pairwise differences in nucleotides ranging from 0% to 5.4%, and were separated into 15 lineages with within-lineage mean pairwise distances ≤0.017 (Grobbelaar et al. 2011). This classification system is useful for discussing epidemiology and transmission, especially when a study indicated that a specific SFTSV clade was statistically associated with SFTS fatality (Dai et al. 2022). However, it should be noted that phylogenetic inference carries statistical uncertainty, and some of the available genome data are noisy, with incomplete genome coverage and errors arising from amplification and sequencing processes (Rambaut et al. 2020).

Reassortment, exclusive to segmented RNA viruses, is an important factor that affects genetic diversity. In our study, the ML trees of the L, M, and S segments exhibited 48 reassortants,
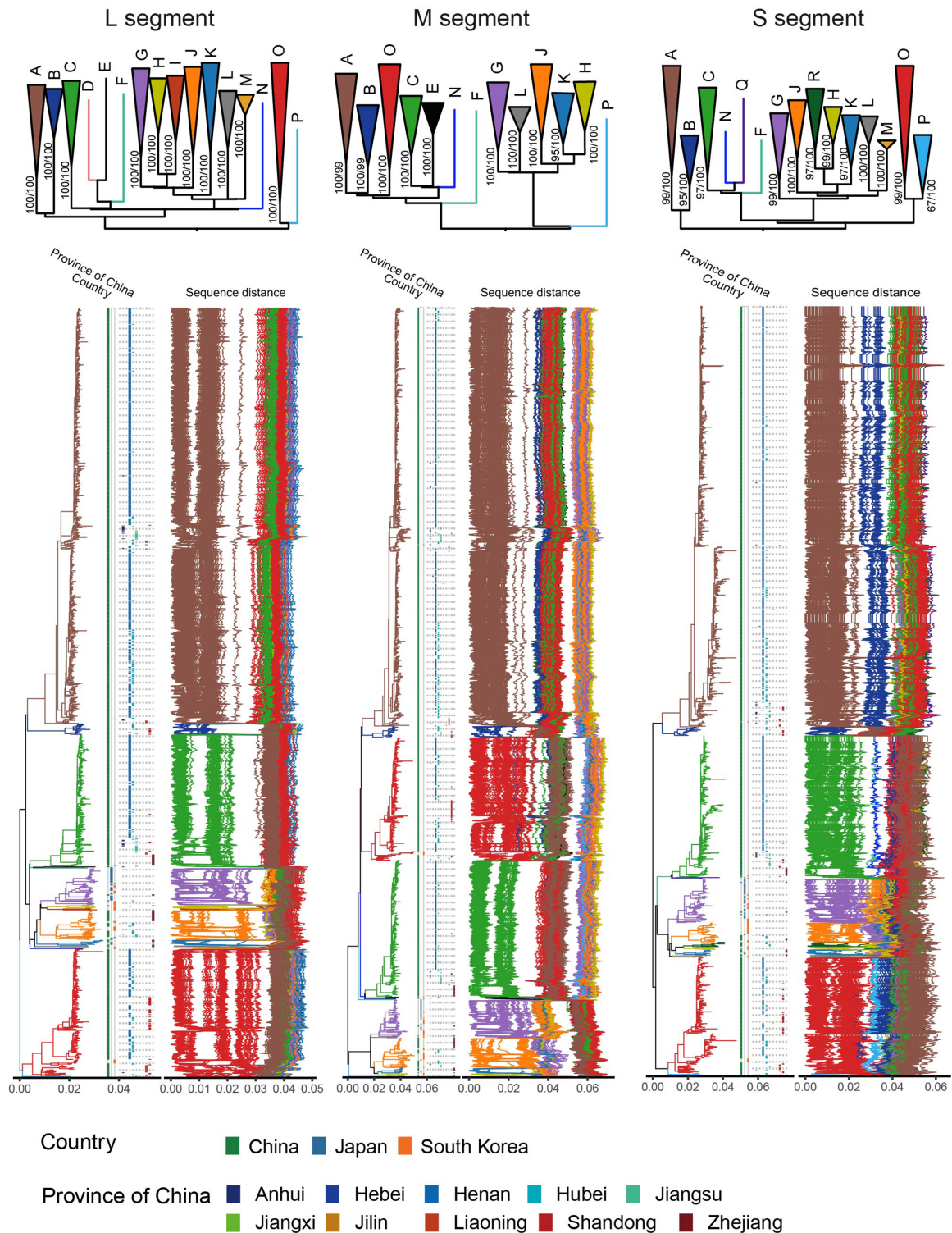
Figure 2. Results of phylogenetic analysis of the L, M, and S segments. The upper panel presents the topology of ML tree, annotated with SH-aLRT and UFboot values for the labeled lineages. The lower panel displays the panoramic view of the ML trees. The outer layers, from left to right, represent the country of origin, the province of China, and the pairwise sequence distance.

accounting for 4.18% of the analyzed strains. The reassortment events were more than previously reported (Shi et al. 2017, Yun et al. 2020a, Dai et al. 2022), likely due to our larger dataset, longer time span covered, multiple countries and provinces, and more refined classification. The L segment had a greater frequency of

reassortment (41.77%). The coexistence of multiple lineages suggests that there are ample opportunities for reassorment to occur, especially in Hubei and Henan provinces in China and in South Korea (Fig. 4). Reassortment is an evolutionary mechanism for host jumps and immune evasion, especially for influenza viruses
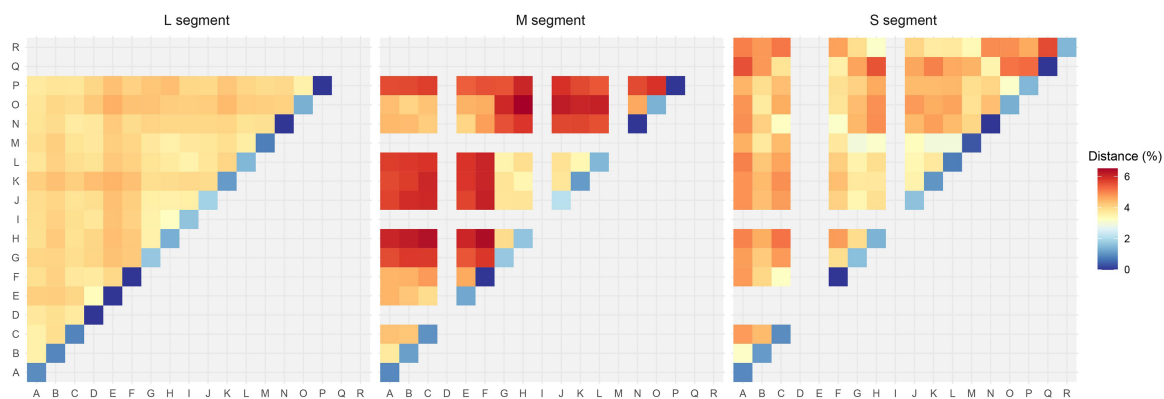
Figure 3. Mean within- and between-lineage distances for L, M, and S segments.
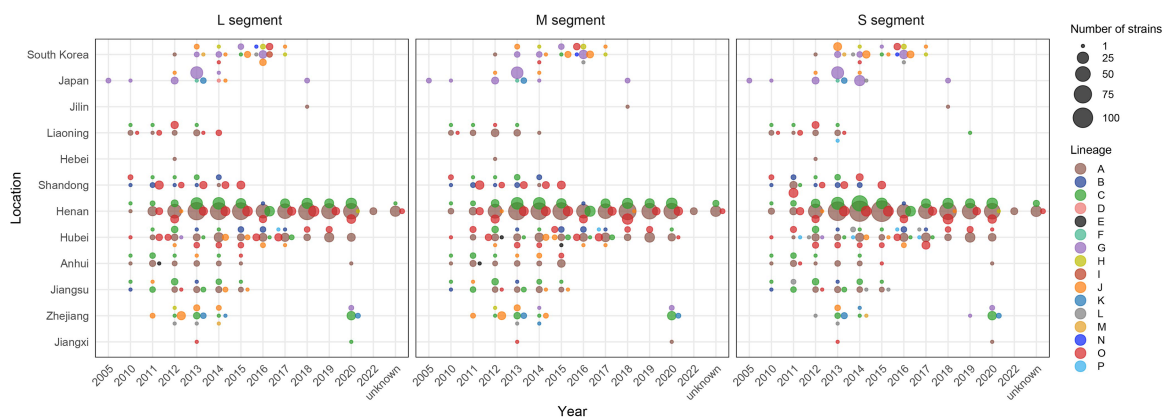

Figure 4. Spatiotemporal distribution of lineages across L, M, and S segments.

(Vijaykrishna et al. 2015). More studies are entailed to address the following questions regarding reassortment not limited to SFTSV: what are the phenotypic consequences of reassortment? Do reassortants have higher fitness or virulence? Do reassortants have a wider host range (Gaudreault et al. 2019)?.

Although China first submitted the sequences in 2010, 2 years before the first submission from South Korea, Japan retrospectively identified the first SFTSV in 2005, the phylogeographic analysis in this study indicated that the SFTSV probably originated from South Korea. In particular, the posterior location probability inferred based on the L and M segments was 0.93, suggesting a highly reliable result. Several studies reported the evolutionary rate and the TMRCA, but these estimates varied remarkably, with evolutionary rates of $(1.07–4.16)\times 10^{-4}$ subs/site/year for the L segment, $(2.08–6.76)\times 10^{-4}$ subs/site/year for the M segment and $(1.09–5.07)\times 10^{-4}$ subs/site/year for the S segment, and a TMRCA of 1719–1868 years for the L segment, 1759–1867 years for the M segment and 1822–1930 years for the S segment (Fu et al. 2016, Liu et al. 2016a, 2016b, Yun et al. 2020a). Our study estimated that the evolutionary rates were 0.61×10^{-4}, 1.31×10^{-4}, and 1.27×10^{-4} subs/site/year for the L, M, and S segments, respectively, which are much lower than those in previous studies. The TMRCA estimates from our analysis were 1617.6 (95% HPD: 1513.1, 1724.3), 1700.4 (95% HPD: 1493.7, 1814.0), and 1790.1 (95% HPD: 1605.4, 1887.2) for the L, M, and S segments, respectively, of which the point estimates were much earlier than that reported in previous studies but the HPD intervals showed some overlap with previous studies. The broad HPD intervals and differences between segments underscore the complexity of estimating virus origins,

especially for segmented viruses with high rates of reassortment. Considering the larger sample size and the broader time range of the sequences included in this study, our results might be more closely related to the nature of the evolutionary process. However, we acknowledge the limitations and complexities inherent in these analyses and suggest that further research is needed to refine these estimates.

Although human-to-human transmission cases have been documented (Jiang et al. 2015, Jung et al. 2019, Wu et al. 2022), the primary infection route to humans is through bites from infected ticks. Therefore, the geographical expansion of the SFTSV may be attributed to the spatial spread of ticks or their vertebrate hosts outside the area of endemicity. *Haemaphysalis longicornis* acts as the main transmission vector for SFTSV. The common hosts of *H. longicornis* are a variety of mammals, including goats, cattle, sheep, yak, donkeys, pigs, deer, cats, rats, mice, hedgehogs, weasels, brushtail possums, and humans (Yu et al. 2011). Human mobilization and/or livestock transport may be one of the transmission mechanisms, as supported by the epidemiological evidence that meat production and milk production are statistically associated with SFTS cases (Jiang et al. 2022). In addition, migratory birds are also regular hosts of *H. longicornis*, which may facilitate the spread of SFTS in the Asia–Pacific region (Yun et al. 2015).

We identified several positively selected sites in the RdRp, Gn-Gc and NS genes but not in the N gene. Of the four sites located on RdRp, Asn2 is on the surface of the endonuclease, Asp1061 is on the fingers, and Ser1116 is on the palm. The endonuclease domain (Walter and Barr 2011), which wraps around the fingers and palm domains, is essential for viral cap-dependent transcription (Vogel
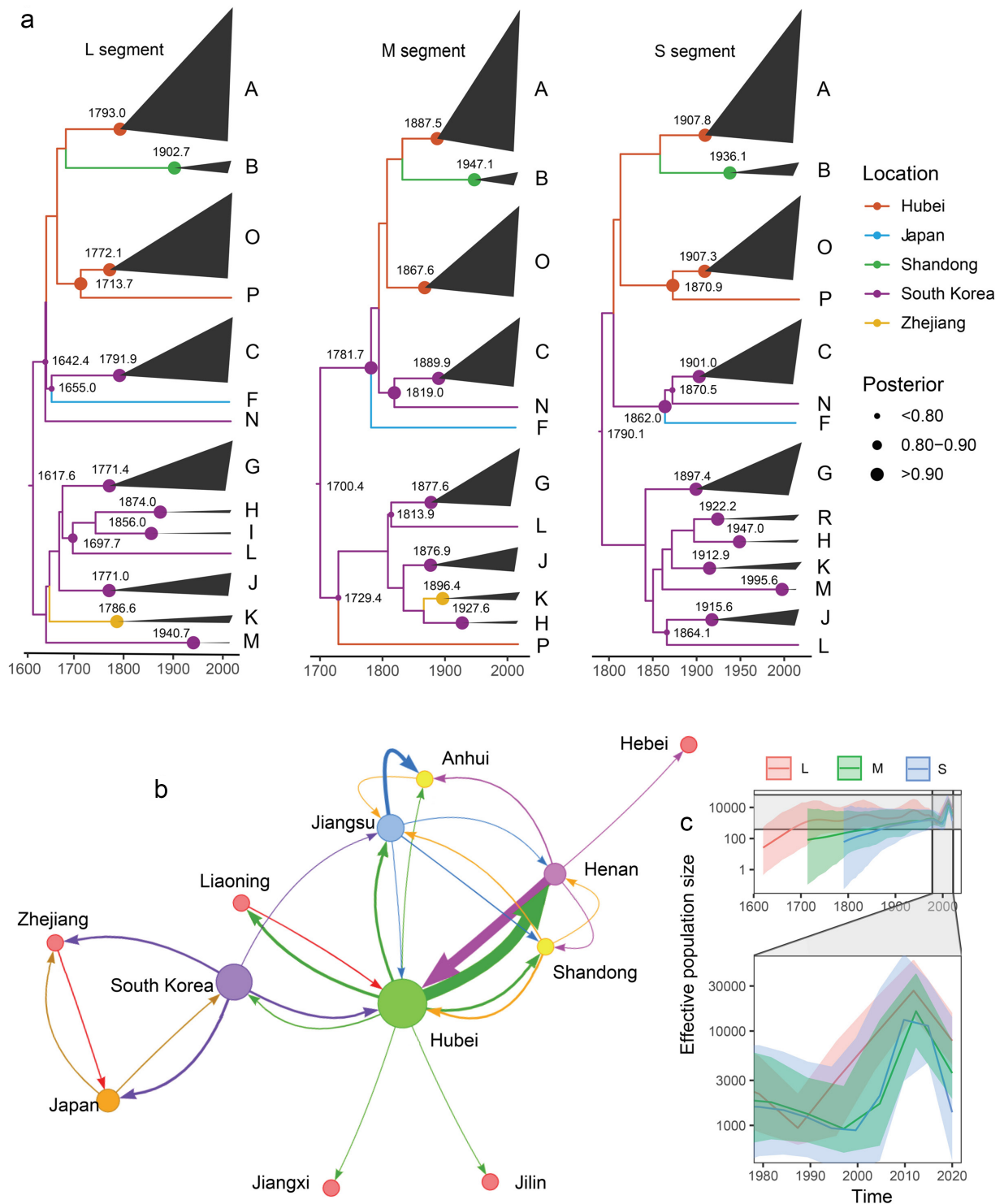
Figure 5. The phylodynamics of SFTSV. (a) The MCC trees of the L, M, and S segments, respectively. The colors correspond to the probable geographic locations. The circle size indicates the posterior of lineages. For the key nodes, the median estimated TMRCAs of lineages are shown; (b) transmission network of the L segment inferred from the MCC tree. The arrows indicate directionality, and their thickness indicates the transition frequency. Node size is scaled by betweenness centrality values, and a higher value reflects the importance of the node as a hub for the traffic of the pathogen. The colors are randomly assigned. (c) The past population dynamics visualized using the Skygrid model. The shaded portion is the 95% Bayesian credibility interval, and the solid line is the posterior median.

et al. 2020). Gn and Gc assemble into a heterodimer on the viral surface, with Gn forming a capsid spike and the Gc portion located below. The ectodomain of Gn consists of domain A, domain B, a β-ribbon, and domain C. The ectodomain of Gc consists of domains I, II, and III (Sun et al. 2023). Of the four sites located on Gn–Gc, two sites (Gln79 and Asp170) are on the surface of Domain A, one site (Ala298) is on Domain B in Gn, and one site (Gln984) is on the surface of Domain III in Gc. Gn and Gc, which are responsible for
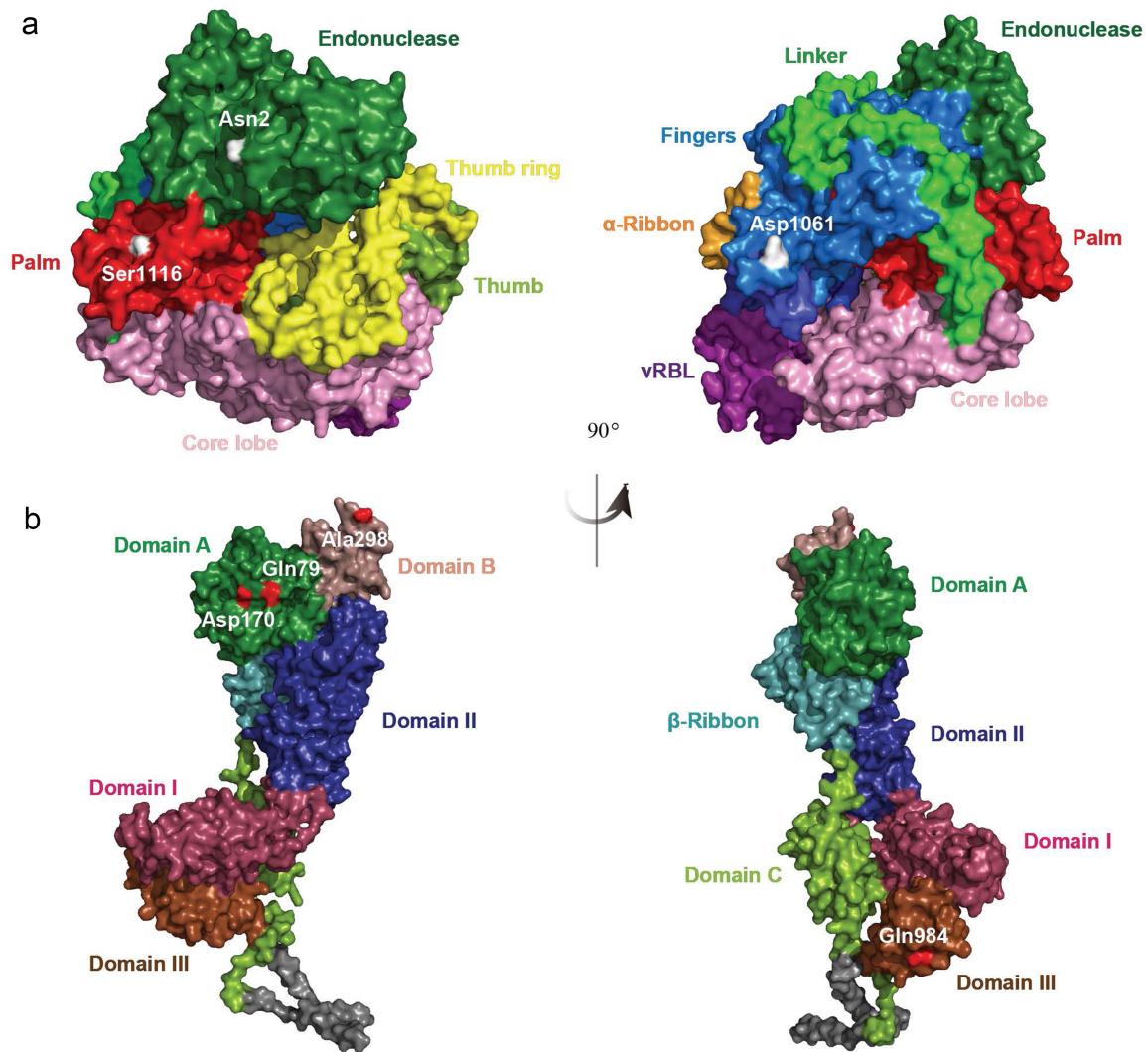
Figure 6. The positive selection sites and domain structures of RdRp (PDB 6Y6K) (a) and Gn–Gc (PDB 7X6U) (b).

cell attachment and membrane fusion, respectively (de Boer et al. 2012, Tani et al. 2016), are two major antigenic components on the viral surface and are the targets of specific neutralizing antibodies (Sun et al. 2023). However, further studies are needed to characterize the impact of these positive selection sites on the efficiency of entry, replication, and production of SFTSV.

Although our study represents the most comprehensive study on the evolutionary dynamics of SFTSV using the largest dataset available to date, the results should be interpreted cautiously given several limitations. First, the sequences were downloaded from GenBank, which means that there is reporting bias, especially in locations with limited resources for SFTSV diagnostic and reporting capacity. Second, researchers usually upload viral segment sequences independently to GenBank, which makes identification of segment sequences from the same strain challenging. In this study, we identified the segment sequences from the same strain by strain name and other information in GenBank, such as location, collection date, or authors, which may introduce mismatch bias. Therefore, the result on reassortants should be interpreted cautiously. Third, some researchers only submitted one or two segments of the same strain to GenBank, which may have led to discrepancies in the number of lineages based on different segments, especially when some lineages had only one or two sequences. Fourth, only human SFTSV was included in this study. SFTSV is known to circulate in a broad range of hosts. This wide host range suggested that SFTSV could adapt to different biological environments. However, the relationship between host immune systems and SFTSV phylogeny was complex and not fully understood. Additionally, the current availability of data from nonhuman hosts was limited. Future research should focus on systematic collection and analysis of SFTSV sequences from diverse hosts across its known geographic range. This broader approach could help elucidate potential host-specific adaptations, refine our understanding of SFTSV evolution, and provide a more comprehensive picture of the virus's evolutionary history and transmission dynamics.

Conclusion

Although SFTS is an emerging infectious disease that was first reported in China in 2009 and SFTSV was retrospectively detected in Japan in 2005, SFTSV likely originated from South Korea between 1617 and 1790. In China, Hubei Province acted as a distribution hub for the spread of the virus. SFTSV is highly conserved, with a low evolutionary rate and a low frequency of reassortment but presents some positive selection sites in the RdRp, Gn–Gc, and

NS genes. This study provides new insight into the diversity and transmission network of SFTSV in East Asia. Regional cooperation should be strengthened to identify and disseminate information regarding the pathways of spread, investigate associated underlying mechanisms, and ultimately improve the coordination of control efforts.

Author contributions

Conceptualization was contributed by S.W.S., Y.G.W., and Q.Y.L., Investigation was done by P.C., C.X.L., and A.R.Z., formal analysis was conducted by S.W.S., P.C., and Y.G.W. writing— original draft by S.W.S., P.C., C.X.L., and A.R.Z; writing – review&editing by S.W.S., Y.G.W. and, Q.Y.L., and funding acquisition by S.W.S.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Conflict of interest: None declared.

Funding

Conflict of interest: None declared.

References

Dai ZN, Peng XF, Li JC *et al.* Effect of genomic variations in severe fever with thrombocytopenia syndrome virus on the disease lethality. *Emerg Microbes Infect* 2022;**11**:1672–82.

de Bernardi Schneider A, Ford CT, Hostager R *et al.* StrainHub: a phylogenetic tool to construct pathogen transmission networks. *Bioinformatics* 2020;**36**:945–47.

de Boer SM, Kortekaas J, Spel L *et al.* Acid-activated structural reorganization of the Rift Valley fever virus Gc fusion protein. *J Virol* 2012;**86**:13642–52.

Diseases, J.N.I.o.I. Investigation and surveillance of SFTS patients in Japan. 2023a.

Diseases, K.C.f.p.a.C.o. Surveillance data on infectious diseases in Korea. 2023b.

Drummond AJ, Ho SY, Phillips MJ *et al.* Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006;**4**:e88.

Fu Y, Li S, Zhang Z *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.

Gaudreault NN, Indran SV, Balaraman V *et al.* Molecular aspects of Rift Valley fever virus and the emergence of reassortants. *Virus Genes* 2019;**55**:1–11.

Grobbelaar AA, Weyer J, Leman PA *et al.* Molecular epidemiology of Rift Valley fever virus. *Emerg Infect Dis* 2011;**17**:2270–76.

Hoang DT, Chernomor O, von Haeseler A *et al.* UFBoot2: Improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;**35**:518–22.

Jiang XL, Zhang S, Jiang M *et al.* A cluster of person-to-person transmission cases caused by SFTS virus in Penglai, China. *Clin Microbiol Infect* 2015;**21**:274–79.

Jiang X, Wang Y, Zhang X *et al.* Factors associated with severe fever with thrombocytopenia syndrome in endemic areas of China. *Front Public Health* 2022;**10**:844220.

Jung IY, Choi W, Kim J *et al.* Nosocomial person-to-person transmission of severe fever with thrombocytopenia syndrome. *Clin Microbiol Infect* 2019;**25**:633e631–633e634.

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–89.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.

Kim KH, Yi J, Kim G *et al.* Severe fever with thrombocytopenia syndrome, South Korea, 2012. *Emerg Infect Dis* 2013;**19**:1892–94.

Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 2005;**22**:1208–22.

Lemey P, Rambaut A, Drummond AJ *et al.* Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009;**5**:e1000520.

Li JC, Zhao J, Li H *et al.* Epidemiology, clinical characteristics, and treatment of severe fever with thrombocytopenia syndrome. *Infect Med (Beijing)* 2022;**1**:40–49.

Liu JW, Zhao L, Luo LM *et al.* Molecular evolution and spatial transmission of severe fever with thrombocytopenia syndrome virus based on complete genome sequences. *PLoS One* 2016a;**11**:e0151677.

Liu L, Chen W, Yang Y *et al.* Molecular evolution of fever, thrombocytopenia and leukocytopenia virus (FTLSV) based on whole-genome sequences. *Infect Genet Evol* 2016b;**39**:55–63.

Liu Q, He B, Huang SY *et al.* Severe fever with thrombocytopenia syndrome, an emerging tick-borne zoonosis. *Lancet Infect Dis* 2014;**14**:763–72.

Lv Q, Zhang H, Tian L *et al.* Novel sub-lineages, recombinants and reassortants of severe fever with thrombocytopenia syndrome virus. *Ticks Tick Borne Dis* 2017;**8**:385–90.

Martin DP, Murrell B, Golden M *et al.* RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* 2015;**1**:vev003.

Miao D, Liu MJ, Wang YX *et al.* Epidemiology and ecology of severe fever with thrombocytopenia syndrome in China, 2010–2018. *Clin Infect Dis* 2021;**73**:e3851–e3858.

Minh BQ, Schmidt HA, Chernomor O *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic Era. *Mol Biol Evol* 2020;**37**:1530–34.

Murrell B, Moola S, Mabona A *et al.* FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 2013;**30**:1196–205.

Murrell B, Wertheim JO, Moola S *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 2012;**8**:e1002764.

Ragonnet-Cronin M, Hodcroft E, Hue S *et al.* Automated analysis of phylogenetic clusters. *BMC Bioinf* 2013;**14**:317.

Rambaut A, Holmes EC, O'Toole A *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;**5**:1403–07.

Rattanakomol P, Khongwichit S, Linsuwanon P *et al.* Severe fever with thrombocytopenia syndrome virus infection, Thailand, 2019-2020. *Emerg Infect Dis* 2022;**28**:2572–74.

Shi J, Hu S, Liu X *et al.* Migration, recombination, and reassortment are involved in the evolution of severe fever with thrombocytopenia syndrome bunyavirus. *Infect Genet Evol* 2017;**47**:109–17.

Suchard MA, Lemey P, Baele G *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;**4**:vey016.

Sun Z, Cheng J, Bai Y *et al.* Architecture of severe fever with thrombocytopenia syndrome virus. *Protein Cell* 2023;**14**:914–18.

Takahashi T, Maeda K, Suzuki T *et al.* The first identification and retrospective study of severe fever with thrombocytopenia syndrome in Japan. *J Infect Dis* 2014;**209**:816–27.

Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol Biol Evol* 2021;**38**:3022–27.

Tani H, Shimojima M, Fukushi S *et al.* Characterization of glycoprotein-mediated entry of severe fever with thrombocytopenia syndrome virus. *J Virol* 2016;**90**:5292–301.

Tran XC, Yun Y, Van An L *et al.* Endemic severe fever with thrombocytopenia syndrome, Vietnam. *Emerg Infect Dis* 2019;**25**:1029–31.

Vijaykrishna D, Mukerji R, Smith GJ. RNA virus reassortment: an evolutionary mechanism for host jumps and immune evasion. *PLoS Pathog* 2015;**11**:e1004902.

Vogel D, Thorkelsson SR, Quemin ERJ *et al.* Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Res* 2020;**48**:5749–65.

Walter CT, Barr JN. Recent advances in the molecular and cellular biology of bunyaviruses. *J Gen Virol* 2011;**92**:2467–84.

Weaver S, Shank SD, Spielman SJ *et al.* Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* 2018;**35**:773–77.

Wu YX, Yang X, Leng Y *et al.* Human-to-human transmission of severe fever with thrombocytopenia syndrome virus through potential ocular exposure to infectious blood. *Int J Infect Dis* 2022;**123**:80–83.

Yoshikawa T, Shimojima M, Fukushi S *et al.* Phylogenetic and geographic relationships of severe fever with thrombocytopenia syndrome virus in China, South Korea, and Japan. *J Infect Dis* 2015;**212**:889–98.

Yu G, Lam TT, Zhu H *et al.* Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol Biol Evol* 2018;**35**:3041–43.

Yu XJ, Liang MF, Zhang SY *et al.* Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med* 2011;**364**:1523–32.

Yun MR, Ryou J, Choi W *et al.* Genetic diversity and evolutionary history of Korean isolates of severe fever with thrombocytopenia syndrome virus from 2013-2016. *Arch Virol* 2020a;**165**:2599–603.

Yun SM, Park SJ, Kim YI *et al.* Genetic and pathogenic diversity of severe fever with thrombocytopenia syndrome virus (SFTSV) in South Korea. *JCI Insight* 2020b;**5**:e129531.

Yun Y, Heo ST, Kim G *et al.* Phylogenetic analysis of severe fever with thrombocytopenia syndrome virus in South Korea and migratory bird routes between China, South Korea, and Japan. *Am J Trop Med Hyg* 2015;**93**:468–74.