

RESEARCH ARTICLE

Genetic diversity and evolution of Hantaan virus in China and its neighbors

Naizhe Li¹, Aqian Li¹, Yang Liu¹, Wei Wu¹, Chuan Li¹, Dongyang Yu², Yu Zhu², Jiandong Li¹, Dexin Li¹, Shiwen Wang^{1,3*}, Mifang Liang^{1,3*}

1 Key Laboratory of Medical Virology and Viral Diseases, Ministry of Health of People's Republic of China, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, **2** Department of Microbiology, Anhui Medical University, Hefei, China, **3** China CDC-WIV Joint Research Center for Emerging Diseases and Biosafety, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, P. R. China

* wangsw@ivdc.chinacdc.cn (SW); liangmf@ivdc.chinacdc.cn (ML)



Abstract

Background

Hantaan virus (HTNV; family Hantaviridae, order Bunyavirales) causes hemorrhagic fever with renal syndrome (HFRS), which has raised serious concerns in Eurasia, especially in China, Russia, and South Korea. Previous studies reported genetic diversity and phylogenetic features of HTNV in different parts of China, but the analyses from the holistic perspective are rare.

Methodology and principal findings

To better understand HTNV genetic diversity and gene evolution, we analyzed all available complete sequences derived from the small (S) and medium (M) segments with bioinformatic tools. Eleven phylogenetic groups were defined and showed geographic clustering; 42 significant amino acid variant sites were found, and 19 of them were located in immune epitopes; nine recombinant events and eight reassortments with highly divergent sequences were found and analyzed. We found that sequences from Guizhou showed high genetic divergence, contributing to multiple lineages of the phylogenetic tree and also to the recombination and reassortment events. Bayesian stochastic search variable selection analysis revealed that Heilongjiang, Shaanxi, and Guizhou played important roles in HTNV evolution and migration; the virus may originate from Zhejiang Province in the eastern part of China; and the virus population size expanded from the 1980s to 1990s.

Conclusions/Significance

These findings revealed the original and evolutionary features of HTNV, which will help to illustrate hantavirus epidemic trends, thus aiding in disease control and prevention.

OPEN ACCESS

Citation: Li N, Li A, Liu Y, Wu W, Li C, Yu D, et al. (2020) Genetic diversity and evolution of Hantaan virus in China and its neighbors. *PLoS Negl Trop Dis* 14(8): e0008090. <https://doi.org/10.1371/journal.pntd.0008090>

Editor: Alain Kohl, University of Glasgow, UNITED KINGDOM

Received: January 22, 2020

Accepted: July 8, 2020

Published: August 20, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pntd.0008090>

Copyright: © 2020 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequences are available from the NCBI and ViPR database ([S2 Table](#)).

Funding: This work was supported by National Science and Technology Major Project of China

(2018ZX10711001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Hemorrhagic fever with renal syndrome (HFRS) and hantavirus pulmonary syndrome are endemic zoonotic infectious diseases caused by hantaviruses, which have threatened public health worldwide for decades. However, our knowledge about the emergence and evolution of HTNV still need to be improved. To get more information about HTNV genetic diversity and phylogenetic features with a holistic perspective, we investigated the genetic diversity and spatial distribution of HTNV using all available whole-genome sequences of small (S) and medium (M) segments collected from 18 different regions and using a larger time-scale; 11 phylogenetic groups were defined. The sequences showed geographic clustering, and divergence with recombinant or reassortment events occurred. Using the Bayesian stochastic search variable selection method, geographic origins and migration patterns of HTNV epidemics were deduced. Our data provided important biological information to better understand hantavirus evolution, transmission, and epidemics.

Introduction

Hantaviruses have gained worldwide attention as etiological agents of emerging zoonotic diseases—namely, hemorrhagic fever with renal syndrome (HFRS) in Eurasia and hantavirus cardiopulmonary syndrome in the Americas—with fatality rates ranging from <10% up to 60% [1]. Among all countries, China is the most seriously affected and accounts for over 90% of the total HFRS cases around the world [2, 3]. It has been reported that the HFRS death rate in China was 2.89% during the years 1950–2014 [4]. Although a declining HFRS trend has been observed at a global scale in China, there still exist certain local regions that continue to display increasing HFRS trends [5]. However, the causative agent of the disease remained unknown until the early 1970s, when Lee and colleagues reported on Hantaan virus (HTNV) present in the lungs of its natural reservoir, the striped field mouse (*Apodemus agrarius*) [6].

HTNV, as one of the pathogenic hantaviruses of HFRS, is a member of genus *Orthohantavirus*, family Hantaviridae, in the order Bunyvirales. The virus genome consists of three separate segments of negative-stranded RNA, referred to as small (S), medium (M), and large (L) segments, which encode nucleocapsid protein (NP), two envelope glycoproteins (Gn and Gc), and viral RNA-dependent RNA polymerase, respectively [7]. Recent studies [8–10] reported genetic diversity and phylogenetic features of HTNV in different parts of China, but the analyses from the holistic perspective are rare. Also, the mechanisms underlying the emergence and evolution of HTNV are poorly understood. Understanding the phylogenetic factors contributing to HTNV transmission has important implications for our understanding of its epidemiology and provides insights about surveillance as well as outbreak predictions and preparation.

In this study, we focused on the genetic diversity and evolutionary history of HTNV. Whole-genome sequences were analyzed with bioinformatic tools. The results revealed the phylogenetic relationships among HTNV strains with both Bayesian and maximum-likelihood (ML) methods. To deduce geographic origins and migration patterns of HTNV epidemics, Bayesian evolutionary analysis sampling trees (BEAST) software was used in this study. Furthermore, the recombination and reassortment events were also detected.

Materials and methods

Sequence data set

All the complete S gene and M gene sequences of *Hantaan orthohantavirus* deposited until June 2019 were collected from Virus Pathogen Database and Analysis Resource (ViPR, www.vipr.lanl.gov/).

vibrbrc.org). Only one sequence was retained for the identity sequences with the same strain names. It should be mentioned that sequences from Shandong Province (accession no. KY639536-KY639711) were not defined as HTNV by both ViPR and Genebank but can be organized in *H. orthohantavirus* (see below). These sequences were still analyzed in this study. All the sequences were aligned using MAFFT version 7 [11] with default settings followed by manual refinement. The coding sequences were retained and used for the following analyses [12].

Recombination detection

As recombination seriously affects phylogenetic inference, the whole data set was tested for the presence of recombination signals using RDP4 [13]. RDP, GENECONV, Maximum Chi-squared, Chimaera, Bootscan, Sister Scanning, and 3Seq were used only on events with P values < 0.01 that were confirmed by four or more methods, which were considered recombination events.

Phylogenetic analyses and amino acid substitution analyses

It is reported that recombination sequences will bias the shape of the inferred phylogenetic tree, the branch length, and reconstruction of ancestral sequences [13–15]. To reduce the potential bias, the recombination segments were first removed from the data sets. The phylogenetic relationships of the complete S and M gene sequences were estimated using a Bayesian Markov chain Monte Carlo (MCMC) method as implemented in MrBayes v3.2.2 [16] and an ML phylogenetic inference as implemented in IQ-TREE v1.6.8 [17]. Dobrava virus, Puumala virus (PUUV), Sin Nombre virus, and Thottapalayam virus were designated as the out-group.

For Bayesian MCMC analysis, the suitability of substitution models for our data sets were assessed using jModelTest v2.1.10 [18], which performed a statistical model selection procedure based on the Akaike information criterion. It identified the best-fitting substitution model GTR+ Γ for both data sets of complete S and M gene sequences. The MCMC settings consisted of two simultaneous, independent runs with four chains each, which were run for 20 million generations and sampled every 200 generations with a 25% burn-in.

For the ML analysis, the best-fitting nucleotide substitution model that minimizes the Bayesian information criterion score was selected by ModelFinder [19], implemented in IQ-TREE. An ML tree was constructed using the best-fitting nucleotide substitution model, and statistical robustness of the branching order within the tree was assessed using ultrafast bootstrap support values [20] (1,000 replicates) and the SH-like approximate likelihood ratio test [21] (SH-aLRT, 5,000 replicates). The trees were visualized in FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

The Metadata-Driven Comparative Analysis Tool, supplied by ViPR, was used to analyze the variation of amino acid between the different phylogenetic groups, hosts, and sample collection years. The P value threshold was set to 0.05. If a specific amino acid exists only in one species or one group but not in other species or groups, this amino acid is considered as a “significant amino acid” marker (or synapomorphy).

Natural selection analysis

Natural selection analysis was assessed using the Datamonkey web-based facility [22]. Site-specific selection pressure was assessed using four varying methods, including single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), mixed-effects model evolution (MEME), and fast unconstrained Bayesian approximation (FUBAR) [23–25]. Significance levels were set to $P < 0.05$, and posterior probability (PP) > 0.9 . The SLAC, FEL, and MEME

methods were used to estimate the nonsynonymous to synonymous (dN/dS) ratio, which represents the differential effect of natural selection on these two types of mutations. The lower dN/dS values indicate stronger negative selection against amino acid change. To examine whether episodic diversifying selection had occurred on individual branches, we utilized the adaptive Branch-Site Random Effects Likelihood (aBSREL) [26] method, which is also available on the Datamonkey web server.

Coalescent and evolution analyses

Complete S gene sequences were used to deduce the evolutionary history of HTNV. To avoid potential biases due to sampling heterogeneity, the data set was reduced using CD-HIT [27] by clustering together sequences with a nucleotide sequence identity threshold of 99%. Temporal evolutionary signal in the ML tree was evaluated by TempEst v1.5.1 [28], which plots sample collection dates against root-to-tip genetic distances obtained from the ML phylogeny tree.

The most recent common ancestor (tMRCA), substitution rates, and evolutionary history were estimated using a Bayesian serial coalescent approach implemented in BEAST v1.10.4 [29]. jModelTest v2.1.10 was used to determine the model of nucleotide substitution that best fit the data set, and the data set was subsequently run using GTR+I+ Γ . Both strict and relaxed (uncorrelated exponential and uncorrelated lognormal) molecular clocks were used for the analysis, combined with different tree priors (constant population size, GMRF Bayesian Skyride, and Bayesian skyline were used in this study). Timescale was inferred using an informative substitution rate prior (a lognormal distribution with 95% of the density lying between 1×10^{-4} and 1×10^{-3} substitutions per site per year) previously estimated for the N gene of rodent-borne hantavirus [30]. Models were compared pairwise by estimating the log marginal likelihood via path sampling and stepping stone analysis [31–33]. For each model, an MCMC was run for 100 million generations, with sampling every 10,000 steps. All other priors were left on their defaults. Posterior probabilities were calculated using the program Tracer1.7 after 10% burn-in. The results were accepted only if all the parameters have effective sample sizes over 200.

The selected molecular clock/demographic model (strict clock with Bayesian skyline prior in this study) was then used for the Bayesian phylogeographic inference based on the continuous-time Markov chain process over discrete sampling locations to estimate the diffusion rates among locations, with Bayesian stochastic search variable selection. The MCMC was run for 300 million generations, with sampling every 30,000 generations. Convergence was assessed by estimating the effective sampling dates using Tracer and accepting effective sample size values of 200 or more for all the parameters. A maximum clade credibility (MCC) tree with the phylogeographic reconstruction was selected from the posterior tree distribution after a 10% burn-in using the Tree Annotator v1.10 and was manipulated in FigTree v1.4.4. The migration routes were visualized using SPREAD3 [34]. Migration pathways were considered to be important when they yielded a Bayes factor greater than 15 and when the mean posterior value of the corresponding migration event was greater than 0.50. Bayes factors were interpreted according to the guidelines of Kass and Raftery [35].

Results

HTNV sequences data set

A total of 225 complete S gene sequences and 180 complete M gene sequences from 238 HTNV strains were contained in the data set. The sampling dates of the sequences ranged from 1976 to 2017. Most of the samples were collected during 2011–2015 (Table 1). According to the sample collection area, 18 regions were defined, including Russia, South Korea, and 16 provinces of China (Fig 1). More samples were collected from South Korea and Zhejiang, with 67 and 36 strains,

Table 1. The sampling dates and host distribution in different regions.

Region	Total	Years						Host					
		≤1990	1991–2000	2001–2005	2006–2010	2011–2015	≥2015	UN	Aa	Rn	Hs	Others	UN
Anhui	1							1			1		
Guangdong	1							1					1
Guizhou	23	8		7				8	9	9	5		
Heilongjiang	6				2	3		1	5		1		
Hubei	4	1		1		1		1	2		2		
Hunan	1				1				1				
Jilin	4			1		2		1	2				2
Jiangsu	13					12		1	4	2	5	Mf:2	
Jiangxi	6					5		1	2			Mm:3	1
Liaoning	7					7					7		
Russia	8			1				7	7		1		
Shaanxi	22	2			9	3	8		10	1	11		
Shandong	14					14			14				
Sichuan	2							2				Cr:1	1
South Korea	78	1		11	14	29	12	11	61		11		6
Tianjin	1							1					1
Yunnan	2		1	1								Nc:1	1
Zhejiang	38				28	8		2	31			Mf:5	2
UN	7							7			1	Tt:1	5
Total	238	12	1	22	54	84	20	45	148	12	45	13	20

Abbreviations: Aa, *A. agrarius*; Hs, *H. sapiens*; Mf, *Mirotus fortis*; Mm, *Mus musculus*; Nc, *Nivienter confucianus*; Rn, *R. norvegicus*; Tt, *Tscherskia triton*; UN, unknown

<https://doi.org/10.1371/journal.pntd.0008090.t001>

respectively. Furthermore, Guizhou and Shaanxi also had a higher number of HTNV samples. The samples were collected from nine kinds of hosts, including *A. agrarius* (Aa), *A. peninsulae* (Ap), *Rattus norvegicus* (Rn), *Mus musculus* (Mm), *Nivienter confucianus* (Nc), *Tscherskia triton* (Tt), *Mirotus fortis* (Mf), *Crocidura russula* (Cr), and *Homo sapiens* (Hs). *A. agrarius*, the main reservoir host for HTNV, has the highest number of samples, with a total of 148 (62.2%) strains isolated. The sampling dates and geographic distribution of the sequences are shown in Fig 1.

Recombination events detection

Recombination events were detected among all the 225 complete S gene sequences and 180 complete M gene sequences from 238 HTNV strains collected in this study. The RDP analysis suggested nine recombination events in six HTNV isolates, and all the recombination events occurred in the M gene segment (Table 2). Similar recombination events were detected in three isolates (CGAa31MP7, CGAa31P9, CGHu2). All three isolates were collected from 2004 to 2005 in Guizhou Province [8]. This would imply that these isolates have the same ancestor, which has experienced recombination previously. Although only five methods suggested the recombination event occurred in strain JN131026, the *P* values were all lower than 0.01. Considering the consensus score was higher than 0.6 (0.646), we defined it as a recombinant isolate. *P* values of the event tested in strain A16 were in the high credibility level.

Phylogenetic analyses and amino acid substitution analyses

The recombination isolates were excluded first. Our phylogenetic analysis showed that 11 groups were defined with the S segment, each with a high degree of support (Fig 2). The

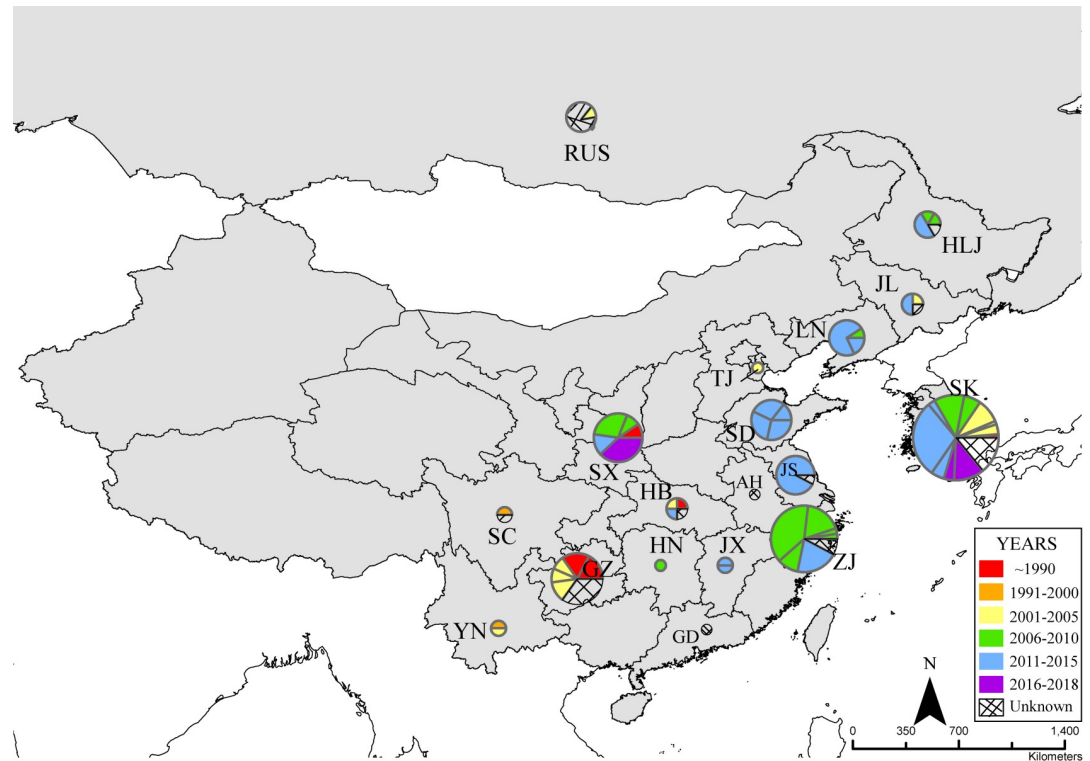


Fig 1. Map showing the sampling dates and geographic distribution of HTNV isolates included in this study. The geographic distribution of the HTNV isolates analyzed in this study. The size of a pie represents the number of isolates in the corresponding region. The pie slices are colored according to the sampling years. AH, Anhui; GD, Guangdong; GZ, Guizhou; HB, Hubei; HLJ, Heilongjiang; HN, Hunan; HTNV, Hantaan virus; JL, Jilin; JS, Jiangsu; JX, Jiangxi; LN, Liaoning; RUS, Russia; SC, Sichuan; SD, Shandong; SK, South Korea; SX, Shaanxi; TJ, Tianjin; YN, Yunnan; ZJ, Zhejiang. *Map source: Natural Earth* (<https://www.naturalearthdata.com/>).

<https://doi.org/10.1371/journal.pntd.0008090.g001>

isolates were mainly clustered together by region. The isolates from South Korea all belonged to group A. All the sequences from Jiangsu Province and one strain from Sichuan (S85_46) Province clustered in group A with sequences from South Korea. Group B was formed with sequences collected from Russia and the northeast of China. The isolates collected in Shaanxi were all clustered in group C, except one strain, 84FLi, which was isolated years ago in 1984. All the three isolates from Hubei constituted group D, and all the isolates from Shandong clustered as a subgroup in group E. The other subgroup of group E contains strains collected from Anhui, Hunan, Shaanxi, Sichuan, Guangdong, and Yunnan. The isolates collected from Zhejiang constitute group G and group H. Group J contains isolates all from Jiangxi. Group K was formed with three isolates from Russia. We noticed that the isolates collected from Guizhou Province were highly divergent. They dispersed in group B, C, F and I, indicating that HTNV has spread in Guizhou for a long period.

There were 10 distinct phylogroups in the M gene tree. The structure of the phylogenetic tree reconstructed by the M gene was identical to the S gene tree, except for seven isolates (CGRn15, CGAa2, CGAa75, CGRn2616, CGRn2618, CGRn45, CGHu3) derived from Guizhou Province. They were classified into group I, but in the S gene tree, they were included in group C. Additionally, isolate N8, collected from Jiangxi, was classified into group G in the M tree but group J in the S tree. This suggested that reassortment events occurred in these strains. The branching order within the lineages differed in S and M trees, which indicates the different

Table 2. Recombination events detected in all HTNV sequences.

Events	Recombinant	Segment	Beginning	Ending	Major Parent	Minor Parent	Method(s) by which Breakpoint Was Detected in RDP4
1	CGAa31MP7	M	1,756	1,916	CGRni1	CGRn15	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
2	CGAa31MP7	M	2,315	3,402	Q32	CGRn2616	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
3	CGAa31P9	M	1,756	1,916	CGRni1	CGRn15	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
4	CGAa31P9	M	2,315	3,402	Q32	CGRn2616	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
5	CGHu2	M	1,756	1,916	CGRni1	CGRn15	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
6	CGHu2	M	2,315	3,402	Q32	CGRn2616	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
7	CGRn93P8	M	2,386	3,402	CGRn93MP8	CGHu2	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq
8	JN131026	M	1,413	2,181	JN131027	158577	MaxChi/Chimaera/Bootscan/SisScan/3Seq
9	A16	M	2,214	3,402	H8205	SN7	RDP/GENECONV/MaxChi/Chimaera/Bootscan/SisScan/3Seq

Abbreviations: HTNV, Hantaan virus; M, medium

<https://doi.org/10.1371/journal.pntd.0008090.t002>

evolutionary history between the M and S segments. It should be clarified that all the branching orders were well supported.

We analyzed the variation of amino acids between the different phylogenetic groups. We noticed that five amino acid sites on NP gene and 37 on Gn/Gc gene were group specific and could be markers to distinguish different groups (Fig 3). No significant variant was found at the highly conserved pentapeptide motif WAASA, which is thought of as the cleavage site on glycoprotein precursor (GPC) polypeptide into the Gn and Gc transmembrane proteins [36]. To determine whether any of these amino acid sites were located in any known immune epitopes, we compared these significant positions against information about experimentally determined immune epitopes curated by the Immune Epitope Database. We found a total of 19 positions that were located in known immune epitopes (Table 3). No significant amino acid marker was found among different hosts or collected times.

Natural selection analysis

The natural selection analysis showed that purifying selection plays a dominant role in HTNV evolution, with very low dN/dS values and an abundance of negatively selected sites (Table 4). The dN/dS values for Gn were the highest, but they still show little evidence of positive selection. In total, 10 positively selected sites in the Gn protein and 11 in the Gc protein were found by MEME. One and 14 positively selected sites in N protein were found by FUBAR and MEME, respectively. We noticed that none of the significant amino acid substitutions sites (Fig 3) were being subjected to positive selection. Evidence of episodic diversifying selection was detected in two branches (CGRn53, HoJo) in the phylogeny of M segment by aBSREL.

Coalescent analyses and evolution of HTNV

To avoid potential biases in the phylogeographic reconstructions due to sampling heterogeneity, we obtained a “nonredundant” subset including 51 clusters. In order to make our data set represent all 15 regions (there was no sequence with exact collected year from Anhui, Sichuan, and Guangdong), isolate E142, Hunan03, and JS10 were added to the data set. Temporal evolutionary signal analyses showed a positive correlation between genetic divergence and sampling time (S1 Fig). It should be mentioned that though the data set we used showed positive correlation by root-to-tip analyses, it failed to pass the date-randomization test (DRT, S2 Fig) [37]. As DRT is stricter than root-to-tip analyses in clock signal evaluation, we think our data set still shows clock signal, which is not so strong. The strict clock with Bayesian skyline prior

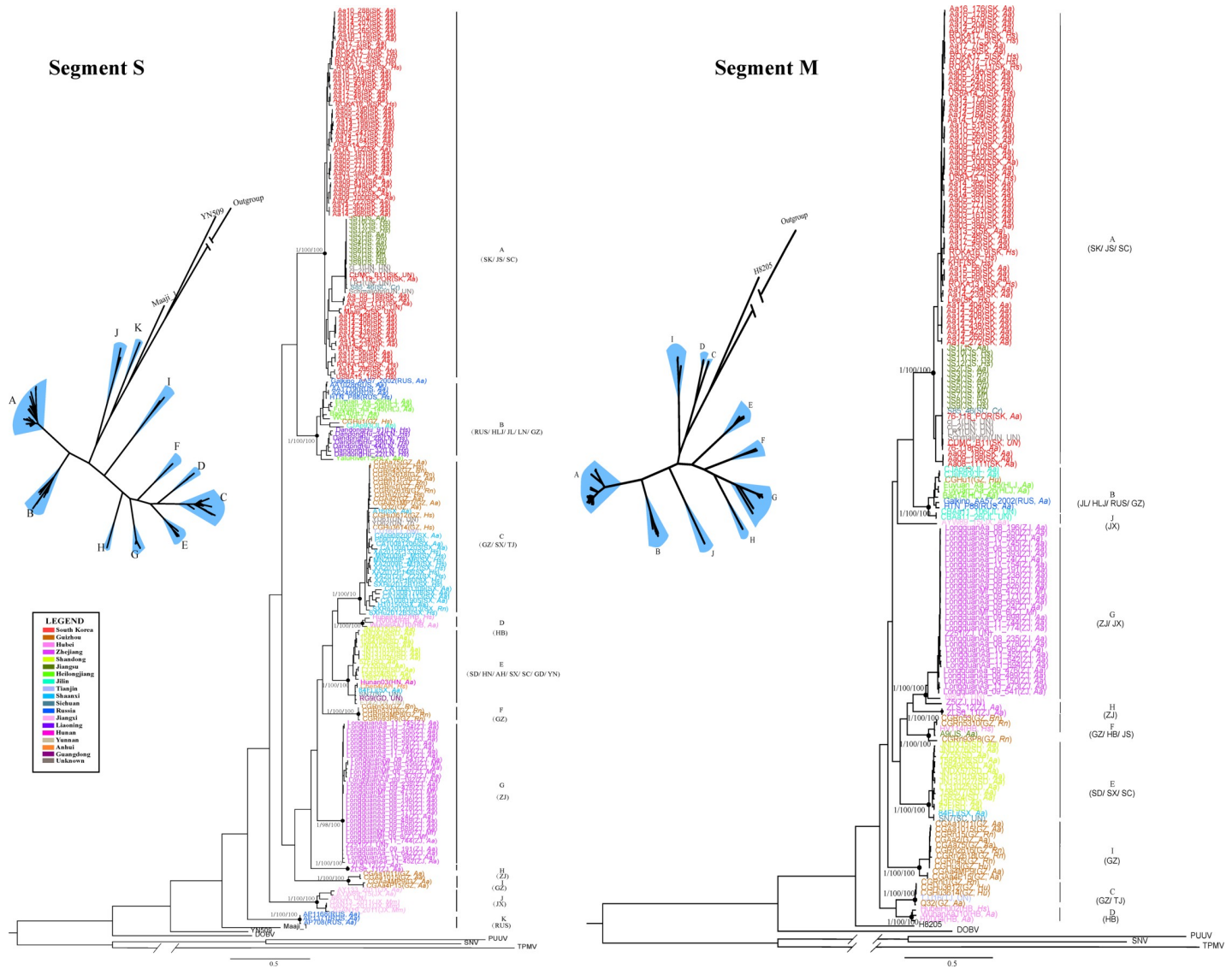


Fig 2. Phylogenetic trees of S and M genes reconstructed by MrBayes. Posterior node probabilities/SH-like approximate likelihood values/ultrafast bootstrap values for major nodes (black circles) are shown. Split network analysis of HTNV was also shown. The region and host features for each sample were labeled after the strains' names. The brackets below the group name indicated the regions associated to each group. AH, Anhui; GD, Guangdong; GZ, Guizhou; HB, Hubei; HLJ, Heilongjiang; HN, Hunan; HTNV, Hantaan virus; JL, Jilin; JS, Jiangsu; JX, Jiangxi; LN, Liaoning; M, medium; RUS, Russia; S, small; SC, Sichuan; SD, Shandong; SK, South Korea; SX, Shaanxi; TJ, Tianjin; YN, Yunnan; ZJ, Zhejiang.

<https://doi.org/10.1371/journal.pntd.0008090.g002>

model yielded a higher log marginal likelihood than the others, indicating the best-fit model for our data set (S1 Table).

The results of our Bayesian phylogenetic analysis showed that HTNV probably first emerged in Zhejiang Province of China (root PP = 0.28), with a most recent common ancestor in 1214 (95% credibility interval 590–1592; Fig 4A). The root PP of Jiangxi, Shaanxi, and Guizhou were also in higher levels. The MCC tree of HTNV obviously showed two separate clusters. One was composed of isolates from the northeast of China and its surroundings, and the other was composed of isolates from the south and middle of China. These indicated that HTNV has been spreading in China for a long time. Reconstruction of the demographic history using the Bayesian skyline plot revealed that the HTNV population size was relatively

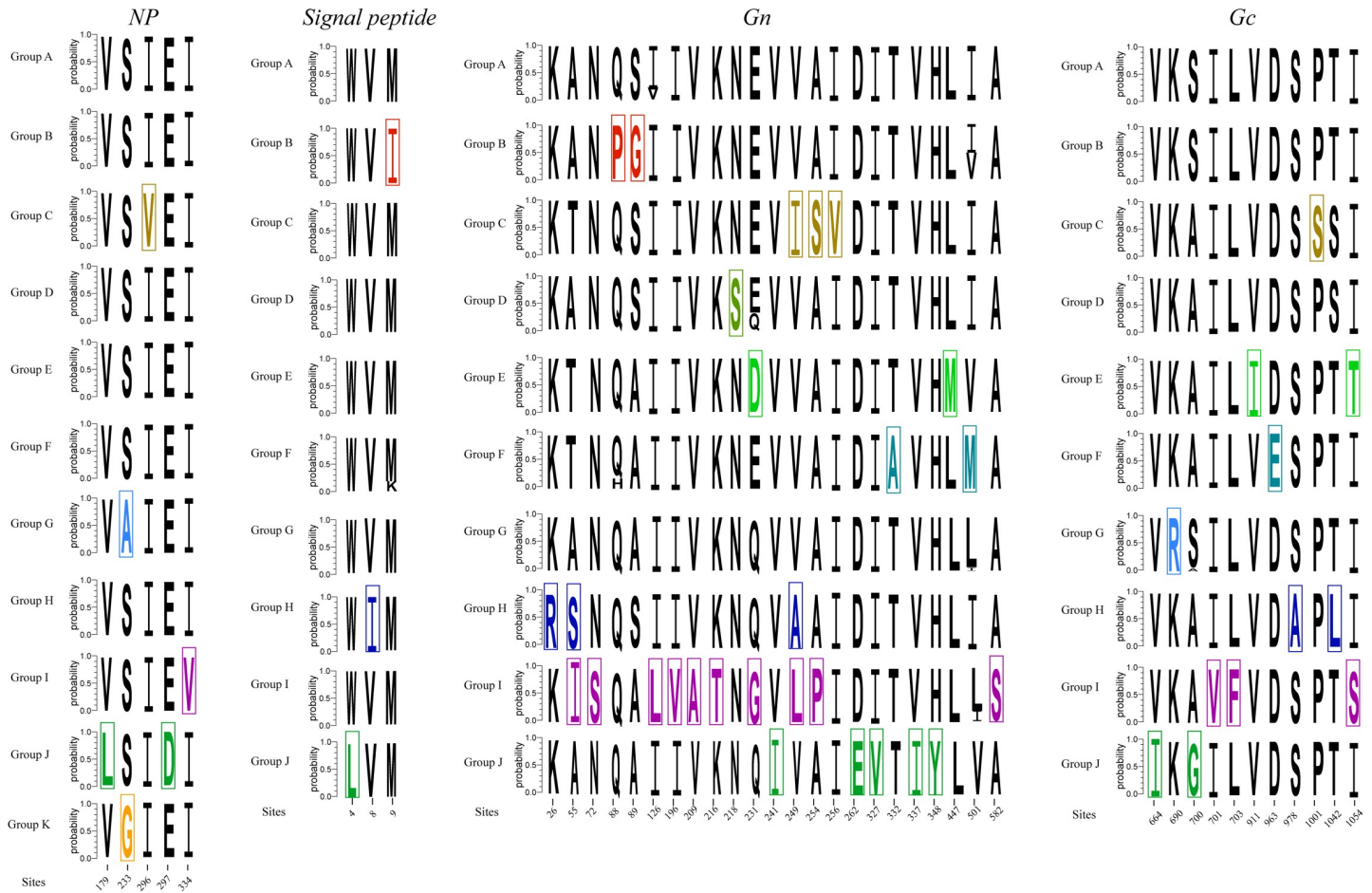


Fig 3. The “significant amino acid marker” of each group. The specific amino acid sites of each group are marked by different colors. NP, nucleocapsid protein.

<https://doi.org/10.1371/journal.pntd.0008090.g003>

constant from 1500 to the 1970s, expanded between the 1980s and 1990s, and then stayed steady from 2000 until now (Fig 3B). The mean substitution rate of 2.0185×10^{-4} substitution per site per year, with a 95% highest posterior density (HPD) that ranged from 9.3435×10^{-5} to 3.2011×10^{-4} , was estimated.

The spatiotemporal linkage of HTNV is shown in Fig 5. Significant location transitions were found mainly in two regions, the northeast and the middle of China (Fig 5A). This indicated that Heilongjiang may be the radiation center of the northeast regions of China and the surrounding areas. Significant migration events were found from Heilongjiang to Liaoning and the Far East of Russia. The migration from Heilongjiang to Khabarovsk was detected at a high credibility level (Bayes factor = 163.0). Shaanxi was assumed as the radiation center of the middle of China. Virus migration from Shaanxi to Tianjin, Yunnan, Hunan, and Guizhou was observed. The number of observed state changes pinpointed Shaanxi as the main source of HTNV epidemics in China, at least in the middle of China, and Heilongjiang as the dominating source of HTNV in northeast of China (Fig 5B).

Discussion

In this study, we used the whole-genome sequences of segment S and M obtained from ViPR to analyze the genetic diversity and evolution of HTNV. Our analyses showed that HTNV can

Table 3. The significant sequence variations' locations on immune epitopes.

Protein	Variation Site	Epitopes					
		Epitope ID	Description	Amino Acid Change	Starting Position	Ending Position	B-/T-Cell Epitopes
NP	179	164302	NGIRKPKHLYVSLPN	V→L	169	183	T
	296	807	AEAAGCSMI	I→V	288	296	-
	334	17746	FSILQDMRNTIMASK	I→V	332	346	T
		149047	FSILQDMRNTI		332	342	T
Signal peptide	8, 9	742143	VMASLVWPV	VM→IM/VI	8	16	T
Gn	209	567849	IVCFVAV	V→A	208	215	T
	216	77254	K216 , G217	K→T			B
	218	77257	N218 , L415	N→S			B
	501	568239	PAITFIIL	I→M,V,L	499	506	T
Gc	664	743275	GVGVPMTDLELDF	V→I	663	677	T
	690	743113	FSLTSSSKYTYRRKL	K→R	677	691	T
		743694	KYTYRRKLTNPLEEA		684	698	T
	700, 701, 703	743887	LTNPLEEAQSIDLHI	S700G/A, I701V, L703F	691	705	T
		742696	AQSIDLHIEIEEQTI		698	712	T
	963	742777	DHINILVTKDIDFDN	D→E	950	964	T
		40611	LVTKDIDFD		955	963	B
	978	744025	NLGENPCKIGLQTSS	S→A	964	978	T
		743579	KIGLQTSSIEGAWGS		971	985	T
		744567	SIEGAWGSGVGFTLT		978	992	T
	1,001	744698	TCLVSLTECPTFLTS	P→S	992	1,006	T
		59502	SLTECPTFL		996	1,004	T
		742922	ECPTFLTSIKACDKA		999	1,013	T
1,042	743186	GKGGHSGSTFRCCHG	T→L/S	1,034	1,048	T	
	744680	STFRCCHGEDCSQIG		1,041	1,055	T	
1,054	743149	GEDCSQIGLHAAAPH	I→S/T	1,048	1,062	T	

Abbreviations: NP, nucleocapsid protein

<https://doi.org/10.1371/journal.pntd.0008090.t003>

be divided into 11 groups. We named the groups based on the S tree for NP gene, which is more stable and has more sequences available. The isolates were geographically clustered. Sequences obtained from the same geographic area clustered together. It is worth noting that sequences from Guizhou were more diverse. We will discuss this later.

A total of 42 significant amid acid variant sites were found among the different groups. More significant variant sites and greater nonsynonymous variations were found for Gn than

Table 4. Summary of selection pressures and codon sites under positive selection.

Gene	SLAC			FEL			FUBAR		MEME		
	dN/dS	N	Sites	dN/dS	N	Sites	N	Sites	dN/dS	N	Sites
NP	0.0339	0	-	0.0341	0	-	1	43	0.0341	14	23, 32, 33, 50, 313, 326, 332, 396, 397, 398, 399, 401, 403, 405
Gn	0.0520	0	-	0.0450	0	-	0	-	0.0450	10	49, 53, 208, 213, 224, 275, 418, 520, 528, 548
Gc	0.0305	0	-	0.0242	0	-	0	-	0.0242	11	675, 724, 725, 780, 766, 860, 896, 925, 992, 1,045, 1,105

“N” refers to the number of positively selected amino acids, and “Sites” refers to amino acids sites that were found to be under positive selection.

Abbreviations: dN/dS, ratio of nonsynonymous to synonymous; FEL, fixed-effects likelihood; FUBAR, fast unconstrained Bayesian approximation; MEME, mixed-effects model evolution; NP, nucleocapsid protein; SLAC, single-likelihood ancestor counting

<https://doi.org/10.1371/journal.pntd.0008090.t004>

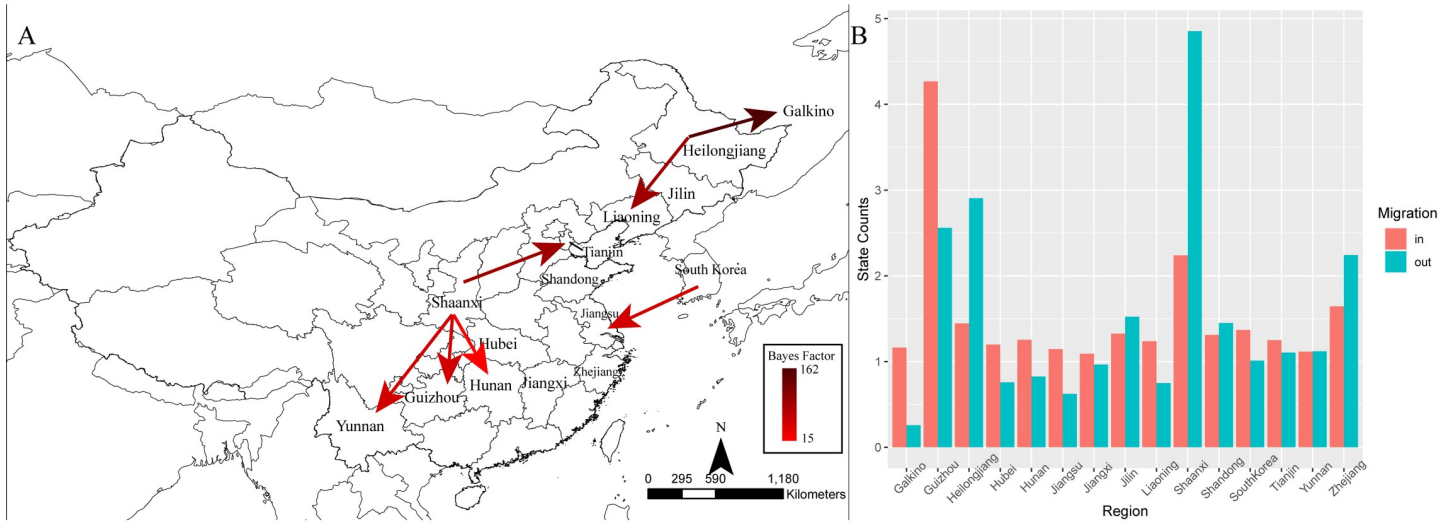


Fig 5. Reconstructed phylogeographic linkage and the total number of location state transitions of HTNV. (A) The phylogeographic linkage constructed reconstruction of the HTNV using BEAST. Color of lines represents the Bayes factor of migration between two regions. The map was created by Spread3 software, and the geographic data were provided by Natural Earth (<https://www.naturalearthdata.com/>). (B) Histogram of the total number of location state transitions inferred from the HTNV data set. BEAST, Bayesian evolutionary analysis sampling trees; HTNV, Hantaan virus.

<https://doi.org/10.1371/journal.pntd.0008090.g005>

437-GQRKVILTKLVIGQ-451 (although the epitope is not recorded in Immune Epitope Database), which showed high antibody binding and replacement at any position of the sequence LTKTLVIGQ (amino acid 443–451), leading to a substantial decrease in reactivity. We found a mutation L447M in group E, which contains isolates collected from Shandong, Shaanxi, and Sichuan. It is also reported [40] that amino acid exchanges of epitope (954-LVTKDIDFD-963) could lead to a loss in reactivity of binding with recombinant human antibodies. We found a mutation, D963E, in group F, which contains isolates from Guizhou, Jiangsu, and Hubei. This might be a clue to the possibility of immune escape in these regions. We found an epitope (ID 742143) located on signal peptide, which showed high binding affinity to HLA-A*0201 molecules and frequencies of epitope-specific cytotoxic T lymphocytes in the peripheral blood mononuclear cells of patients with HFRS and could induce CD8+ T-cell responses to inhibit HTNV replication [41]. Experimentation is needed to determine whether peptides containing these residues can confer escape or can be used to develop serotype-specific reagents for serology-based diagnostics. No significant amino acid marker was found among different hosts and periods, indicating that geographic region is the main factor affecting the genetic diversity of HTNV. The topologies between the M and S segments are different. We inferred that this was due to the different evolutionary history, which was in accordance with others' assumptions [42, 43], but this still needs to be validated in the future.

Our analyses showed that both reassortment and recombination play important roles in HTNV evolution. A number of studies have revealed that genetic reassortment can occur naturally or experimentally between the members of the Bunyavirales [44–49]. Seven of eight reassortment events occurred in M segments found in our study are consistent with Zou and colleagues [8]. We found nine recombination events in our study. All the recombination events occurred in segment M. The reassortment and recombination events found in Gn/Gc envelope glycoproteins were consistent with their functional roles in the viral escape from immunological responses. It should be noted that recombinant A16, isolated in Shaanxi, has both a major parent, H8205, and a minor parent, SN7. Strain H8205 has been classified into Amur virus, which indicates that recombination occurred across different species in

hantavirus. Strain SN7 was isolated in Sichuan, which implies the relationship of the evolution of HTNV between Shaanxi and Sichuan. Evidence of recombination has also been reported for *Tula orthohantavirus* [50] and *Puumala orthohantavirus* (PUUV) [51]; even the recombination in hantavirus is rare [52].

We investigated the evolution and migration of HTNV in China and its surrounding countries using Bayesian phylogeographic inference. Our phylogenetic analyses placed the root of the tree for HTNV in Zhejiang with strong support (Fig 3B). But no migration event was found from Zhejiang to other regions. This may result from the lack of whole-genome data from Zhejiang and its surrounding areas. But the reassortment event of strain N8 indicates the correlation of the isolates in Zhejiang and Jiangxi. Markov jump estimates between different locations pinpointed Heilongjiang as an important source of HTNV epidemics in northeast China and the Far East of Russia. This finding may explain the high level of epidemics in these areas since 1931, when HFRS was first recognized in northeast China [53, 54], due to the migration from Heilongjiang to its surroundings. In a similar pattern, Shaanxi was recognized as the origin of HTNV spread in the middle of China. These can be explained by a scenario in which the virus was first introduced into these areas and then expanded there. But the transmission sources to Heilongjiang and Shaanxi are still unclear. The samples collected from South Korea have been studied before [55–57], so we do not discuss them too much here. But the migration of HTNV from South Korea to Jiangsu indicates the possibility of HTNV import from other countries.

Our analyses showed the high diversity of isolates from Guizhou Province. Phylogenetic analyses showed that isolates collected from Guizhou Province clustered in group B, C, F, and I. Seven of eight reassortment events and seven of nine recombination events occurred in Guizhou. We found the same recombination pattern in strain CGAa31MP7, CGAa31P9, and CGHu2. All of these indicate that different sources of HTNV were transmitted to Guizhou and evolved for a long time in this place [58, 59]. This assumption could be supported by our Markov jump estimate (Fig 5). But our Markov jump estimate as well as the phylogenetic correlation with other regions also indicated the dispersion from Guizhou to other areas, even though no significant migration was found by phylogeographic inference. Furthermore, the PP that Guizhou was the root location was high, so we cannot deny the possibility of Guizhou to be the place of HTNV's common ancestor. With its rich mountainous topography, Guizhou is more than 1,000 meters above sea level, and as much as 92.5% of the province's total area is characterized by mountains. This makes *A. agrarius*, the main host of HTNV, sympatric in rural resident areas more common. The Hengduan Mountains region was hypothesized to have played an important role in the evolutionary history of *Apodemus* since the Pleistocene era [60, 61]. The Yunnan–Guizhou Plateau, which connects and overlaps with the Hengduan Mountains region, is also thought to play an important role in HTNV evolution. But we only found acceptable migration from Shaanxi to Guizhou in our study. This may result from the lack of sequences in Guizhou. To reduce the bias, recombinant and reassortment isolates were removed before the phylogeographic analysis. We also got rid of similar sequences with identity more than 99%. Thus, limited sequences from Guizhou were used for evolutionary and phylogeographic analyses. More gene information should be collected in order to know more about the HTNV evolution and spread in this area.

The most recent common ancestor of HTNV was determined about more than 800 years ago in our study, which is much earlier than when the first patient with HFRS was reported, in the early 1930s [62]. However, HFRS-like disease was described in Chinese writings about 1,000 years ago [63], so we believe HTNV has spread in China for more than 1,000 years. The most common recent ancestor according to Ramsden and colleagues [64] existed 859 years before the present for all rodent hantaviruses estimated and 202 years before the present for

Murinae viruses, with a mean substitution rate for rodent hantavirus of 6.67×10^{-4} substitutions per site per year with a 95% HPD that ranged from 3.86×10^{-4} to 9.8×10^{-4} substitutions per site per year. Partial sequences (275 nt) were used in this study, which can explain the difference in tMRCA estimation. A recent study [65] showed that the estimated rate of nucleotide substitutions for the N gene of all rodent-borne hantavirus was 6.8×10^{-4} substitutions per site per year, which is similar to 6.67×10^{-4} substitutions per site per year. Our results implied that HTNV evolved at a lower speed compared with the other rodent-borne hantaviruses. The estimated rate of nucleotide substitutions of Dobrava virus and PUUV, 3×10^{-4} and 5.5×10^{-4} substitutions per site per year, respectively, supports our hypothesis [30].

The relatively recent origin of HTNV apparently contradicts the virus–host codivergence theory. Previous estimates of evolutionary dynamics in hantaviruses were based on the critical assumption that the congruence between hantavirus and rodent phylogenies reflects codivergence between these two groups because of the divergence of the rodent genera *Mus* and *Rattus*, approximately 10–40 million years ago, which indicates a mean substitution rate in the range of 10^{-8} substitutions per site per year [51, 66, 67]. However, the observation of host–pathogen phylogenetic congruence does not necessarily indicate codivergence. Phylogenetic congruence between a parasite and its host can also arise from delayed cladogenesis, in which the parasite phylogeny tracks that of the host but without temporal association [68]. This could occur if hantaviruses largely evolve host associations by cross-species transmission and related species tend to live in the same area, in which case a pattern of strong host–pathogen phylogenetic congruence could be observed in the absence of codivergence. Our evolutionary rates were estimated directly from primary sequence data sampled at known dates so that they more closely reflect the evolutionary changes undergone by the virus, at least in the short term. And with a mean substitution rate that is closer to other RNA virus, we consider our results to be more reliable than codivergence theory.

It should be noted that the substitution rate and tMRCA of HTNV estimated is not precise and could be improved in the future. As we show in the Results, the data set used for the molecular clock rate and phylogeography analyses did not pass the DRT, which revealed that the molecular clock rate of the data set is potentially unreliable. Therefore, a more accurate substitution rate and tMRCA would be estimated in the future when more sequences with stronger temporal structure are available. Furthermore, the selection pressure analyses showed that purifying selection plays a dominated role in HTNV evolution. It is reported [69] that the presence of strong purifying selection can lead to an underestimation of branch lengths. In our study, we calculated the total branch length of the MCC tree and aBSREL tree using the *compute.brLen()* function [70] supplied by ape package in R (<http://ape-package.ird.fr/>). The results showed that the total branch length for the MCC tree is smaller than the aBSREL tree (8.604 to 14.698), indicating the underestimation of the substitution rate and tMRCA of HTNV.

Our demographic analyses revealed that the HTNV population had expanded from the 1980s to 1990s. Because HTNV is a kind of rodent-borne virus, the population size of HTNV is relevant to rodent populations. Previous studies have revealed that climatic factors can influence HFRS transmission through their effects on the reservoir host (mostly rodents of the family Muridae) and environmental conditions [71–73]. It is reported that global annual average temperatures increased by more than 1.2°F (0.7°C) from 1986 to 2016 relative to temperatures seen from 1901 to 1960 [74]. Global warming may affect rodent winter survival through winter temperatures by a complicated process, and it may also influence the magnitude of HFRS outbreaks through summer climatic conditions (both temperature and rainfall). Additionally, as global climate change accelerates, the amount of annual rainfall increased accurately [75]. Moreover, agriculture improved rapidly in China during the same period, leading to more available food, which increases the reproduction rates of rodents. The steady effective

population from 2000 until now has resulted from the successful strategy for HTNV prevention in China; however, it is worth noting that the uncertainty of demographic analyses increases toward the root of the genealogy, where population history is reconstructed from fewer lineages, and a majority of samples in our data set were collected in recent years. Thus, the population size of the viruses may not be precise, but the increase from the 1980s to 1990s and the constant size after 2000 could be used as a reference for HTNV demographic inference.

However, some caution must be taken when interpreting our results. We have estimated the age of genetic diversity based only on N gene sequences; the evolutionary rate of HTNV may be underestimated. The Gn/Gc envelope glycoproteins are more diverse and considered to evolve at a higher speed. Furthermore, there may be geographic biases due to the imbalance in the collection of sequences from different regions. Because we did the research using whole-genome sequences, partial sequences are not comparable; for instance, partial sequences from different research may be located in different regions of whole genome. The lack of whole-genome sequences in some regions may result in the lack of knowledge about HTNV migration. In addition, lack of sequences of L gene can be obtained now. More gene information should be collected and shared to solve these problems.

In conclusion, our study revealed 11 groups of HTNV and 43 significant amino acid markers for different groups which could be connected with different regions. We hope it could be an indicator of immune effect evaluation of vaccine in different regions. Both recombination and reassortment events can be detected in HTNV. The origin and migration of HTNV were also shown by our analyses. We found that Heilongjiang, Shaanxi, and Guizhou played important roles in HTNV evolution and migration. It is crucial to pay more attention to HFRS prevention and control in these areas. Because rodent populations and activity influence the spread and increase of HTNV, rodent prevention and control is an effective way to reduce the incidence of diseases, which has also been proved by other studies [76, 77]. And a steady effective population from 2000 until now indicates that this is a successful strategy for HTNV prevention. These data provide important insights for better understanding the genetic diversity and spatiotemporal dynamics of HTNV and would be useful for disease prevention and control.

Supporting information

S1 Fig. Root-to-tip regression analyses.

(TIF)

S2 Fig. Comparison of the substitution rates estimated using BEAST from both the original and 20 date-randomized data sets from the DRT. BEAST, Bayesian evolutionary analysis sampling trees; DRT, date-randomization test.

(TIF)

S1 Table. Log marginal likelihoods computed by path sampling and stepping stone sampling.

(XLSX)

S2 Table. Accession numbers of the sequences using in this study.

(XLSX)

Author Contributions

Data curation: Naizhe Li, Yang Liu, Wei Wu, Chuan Li, Dongyang Yu, Yu Zhu.

Formal analysis: Naizhe Li, Aqian Li.

Investigation: Mifang Liang.

Methodology: Naizhe Li, Aqian Li.

Project administration: Jiandong Li, Dexin Li, Shiwen Wang.

Resources: Naizhe Li.

Software: Naizhe Li.

Supervision: Jiandong Li, Dexin Li, Shiwen Wang, Mifang Liang.

Validation: Naizhe Li.

Visualization: Naizhe Li.

Writing – original draft: Naizhe Li.

Writing – review & editing: Naizhe Li, Aqian Li, Yang Liu, Wei Wu, Chuan Li, Dongyang Yu, Yu Zhu, Jiandong Li, Dexin Li, Shiwen Wang, Mifang Liang.

References

1. Papa A. Dobrava-Belgrade virus: Phylogeny, epidemiology, disease. *Antiviral Research*. 95(2):104–17. <https://doi.org/10.1016/j.antiviral.2012.05.011> PMID: 22659378
2. Khaiboullina S, Morzunov S, St. Jeor S. Hantaviruses: Molecular Biology, Evolution and Pathogenesis. *Current Molecular Medicine*. 5(8):773–90. <https://doi.org/10.2174/156652405774962317> PMID: 16375712
3. Jonsson CB, Figueiredo LT, Vapalahti O. A global perspective on hantavirus ecology, epidemiology, and disease. 2010; 23(2):412–41.
4. Jiang H, Du H, Wang LM, Wang PZ, Bai XF. Hemorrhagic Fever with Renal Syndrome: Pathogenesis and Clinical Picture. *Frontiers in cellular and infection microbiology*. 2016; 6:1. Epub 2016/02/13. <https://doi.org/10.3389/fcimb.2016.00001> PMID: 26870699; PubMed Central PMCID: PMC4737898.
5. Lin H, Zhang Z, Lu L, Li X, Liu Q. Meteorological factors are associated with hemorrhagic fever with renal syndrome in Jiaonan County, China, 2006–2011. *International Journal of Biometeorology*. 2014; 58(6):1031–7. <https://doi.org/10.1007/s00484-013-0688-1> PMID: 23793957
6. Lee HW, Lee PW, Johnson KM. Isolation of the Etiologic Agent of Korean Hemorrhagic Fever. *The Journal of Infectious Diseases*. 1978; 137(3):298–308. <https://doi.org/10.1093/infdis/137.3.298> PMID: 24670
7. Plyusnin A, Vapalahti O, Vaheri A. Hantaviruses: genome structure, expression and evolution. *The Journal of general virology*. 1996; 77 (Pt 11):2677–87. Epub 1996/11/01. <https://doi.org/10.1099/0022-1317-77-11-2677> PMID: 8922460.
8. Zou Y, Hu J, Wang ZX, Wang DM, Li MH, Ren GD, et al. Molecular diversity and phylogeny of Hantaan virus in Guizhou, China: Evidence for Guizhou as a radiation center of the present Hantaan virus. *Journal of General Virology*. 2008; 89:1987–97. <https://doi.org/10.1099/vir.0.2008/000497-0> PMID: 18632971
9. Zou Y, Hu J, Wang Z-x, Wang D-m, Yu C, Zhou J-z, et al. Genetic Characterization of Hantaviruses Isolated from Guizhou, China: Evidence for Spillover and Reassortment in Nature. 2008; 1041:1033–41. <https://doi.org/10.1002/jmv>
10. Zuo S-q Li X-j, Wang Z-q Jiang J-f, Fang L-q Zhang W-H, et al. Genetic Diversity and the Spatio-Temporal Analyses of Hantaviruses in Shandong Province, China. *Frontiers in Microbiology*. 2018; 9:1–13. <https://doi.org/10.3389/fmicb.2018.02771> PMID: 30524397
11. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690; PubMed Central PMCID: PMC3603318.
12. Maes P, Klempa B, Clement J, Matthijnsens J, Gajdusek DC, Kruger DH, et al. A proposal for new criteria for the classification of hantaviruses, based on S and M segment protein sequences. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2009; 9(5):813–20. <https://doi.org/10.1016/j.meegid.2009.04.012> PMID: 19393771.

13. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015; 1(1):vev003. <https://doi.org/10.1093/ve/vev003> PMID: [27774277](https://pubmed.ncbi.nlm.nih.gov/27774277/); PubMed Central PMCID: PMC5014473.
14. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics.* 2000; 156(2):879–91. Epub 2000/10/03. PMID: [11014833](https://pubmed.ncbi.nlm.nih.gov/11014833/); PubMed Central PMCID: PMC1461297.
15. Arenas M, Posada D. The effect of recombination on the reconstruction of ancestral sequences. *Genetics.* 2010; 184(4):1133–9. Epub 2010/02/04. <https://doi.org/10.1534/genetics.109.113423> PMID: [20124027](https://pubmed.ncbi.nlm.nih.gov/20124027/); PubMed Central PMCID: PMC2865913.
16. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003; 19(12):1572–4. <https://doi.org/10.1093/bioinformatics/btg180> PMID: [12912839](https://pubmed.ncbi.nlm.nih.gov/12912839/)
17. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution.* 2014; 32(1):268–74. <https://doi.org/10.1093/molbev/msu300> PMID: [25371430](https://pubmed.ncbi.nlm.nih.gov/25371430/)
18. Posada D. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution.* 2008; 25(7):1253–6. <https://doi.org/10.1093/molbev/msn083> PMID: [18397919](https://pubmed.ncbi.nlm.nih.gov/18397919/)
19. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods.* 2017; 14:587. <https://doi.org/10.1038/nmeth.4285> PMID: [28481363](https://pubmed.ncbi.nlm.nih.gov/28481363/)
20. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution.* 2017; 35(2):518–22. <https://doi.org/10.1093/molbev/msx281> PMID: [29077904](https://pubmed.ncbi.nlm.nih.gov/29077904/)
21. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology.* 2010; 59(3):307–21. <https://doi.org/10.1093/sysbio/syq010> PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)
22. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution.* 2018; 35(3):773–7. <https://doi.org/10.1093/molbev/msx335> PMID: [29301006](https://pubmed.ncbi.nlm.nih.gov/29301006/)
23. Kosakovsky Pond SL, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular Biology and Evolution.* 2005; 22(5):1208–22. <https://doi.org/10.1093/molbev/msi105> PMID: [15703242](https://pubmed.ncbi.nlm.nih.gov/15703242/)
24. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection. *Molecular Biology and Evolution.* 2013; 30(5):1196–205. <https://doi.org/10.1093/molbev/mst030> PMID: [23420840](https://pubmed.ncbi.nlm.nih.gov/23420840/)
25. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012; 8(7):e1002764. Epub 2012/07/19. <https://doi.org/10.1371/journal.pgen.1002764> PMID: [22807683](https://pubmed.ncbi.nlm.nih.gov/22807683/); PubMed Central PMCID: PMC3395634.
26. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution.* 2015; 32(5):1342–53. <https://doi.org/10.1093/molbev/msv022> PMID: [25697341](https://pubmed.ncbi.nlm.nih.gov/25697341/)
27. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010; 26(5):680–2. <https://doi.org/10.1093/bioinformatics/btq003> PMID: [20053844](https://pubmed.ncbi.nlm.nih.gov/20053844/)
28. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution.* 2016; 2(1). <https://doi.org/10.1093/ve/vev007> PMID: [27774300](https://pubmed.ncbi.nlm.nih.gov/27774300/)
29. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution.* 2012; 29(8):1969–73. <https://doi.org/10.1093/molbev/mss075> PMID: [22367748](https://pubmed.ncbi.nlm.nih.gov/22367748/)
30. Ramsden C, Melo FL, Figueiredo LM, Holmes EC, Zanutto PMA, Moreli ML, et al. High rates of molecular evolution in hantaviruses. *Molecular Biology and Evolution.* 2008; 25:1488–92. <https://doi.org/10.1093/molbev/msn093> PMID: [18417484](https://pubmed.ncbi.nlm.nih.gov/18417484/)
31. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Molecular Biology and Evolution.* 2012; 29(9):2157–67. <https://doi.org/10.1093/molbev/mss084> PMID: [22403239](https://pubmed.ncbi.nlm.nih.gov/22403239/)

32. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. *Molecular Biology and Evolution*. 2012; 30(2):239–43. <https://doi.org/10.1093/molbev/mss243> PMID: 23090976
33. Baele G, Lemey P. Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. *Bioinformatics*. 2013; 29(16):1970–9. <https://doi.org/10.1093/bioinformatics/btt340> PMID: 23766415
34. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. Spread3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Molecular Biology and Evolution*. 2016; 33(8):2167–9. <https://doi.org/10.1093/molbev/msw082> PMID: 27189542
35. Gao F, Liu X, Du Z, Hou H, Wang X, Wang F, et al. Bayesian phylodynamic analysis reveals the dispersal patterns of tobacco mosaic virus in China. *Virology*. 2019; 528:110–7. <https://doi.org/10.1016/j.virol.2018.12.001> PMID: 30594790.
36. Lober C, Anheier B, Lindow S, Klenk HD, Feldmann H. The Hantaan virus glycoprotein precursor is cleaved at the conserved pentapeptide WAASA. *Virology*. 2001; 289(2):224–9. <https://doi.org/10.1006/viro.2001.1171> PMID: 11689045.
37. Duchene S, Duchene D, Holmes EC, Ho SY. The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data. *Mol Biol Evol*. 2015; 32(7):1895–906. Epub 2015/03/17. <https://doi.org/10.1093/molbev/msv056> PMID: 25771196.
38. Li S, Rissanen I, Zeltina A, Hepojoki J, Raghwanji J, Harlos K, et al. A Molecular-Level Account of the Antigenic Hantaviral Surface. *Cell reports*. 2016; 15(5):959–67. Epub 2016/04/28. <https://doi.org/10.1016/j.celrep.2016.03.082> PMID: 27117403; PubMed Central PMCID: PMC4858563.
39. Liang M, Mahler M, Koch J, Ji Y, Li D, Schmaljohn C, et al. Generation of an HFRS patient-derived neutralizing recombinant antibody to Hantaan virus G1 protein and definition of the neutralizing domain. *Journal of medical virology*. 2003; 69(1):99–107. Epub 2002/11/19. <https://doi.org/10.1002/jmv.10259> PMID: 12436484.
40. Koch J, Liang M, Queitsch I, Kraus AA, Bautz EK. Human recombinant neutralizing antibodies against hantaan virus G2 protein. *Virology*. 2003; 308(1):64–73. Epub 2003/04/23. [https://doi.org/10.1016/s0042-6822\(02\)00094-6](https://doi.org/10.1016/s0042-6822(02)00094-6) PMID: 12706090.
41. Tang K, Cheng L, Zhang C, Zhang Y, Zheng X, Zhang Y, et al. Novel Identified HLA-A*0201-Restricted Hantaan Virus Glycoprotein Cytotoxic T-Cell Epitopes Could Effectively Induce Protective Responses in HLA-A2.1/K(b) Transgenic Mice May Associate with the Severity of Hemorrhagic Fever with Renal Syndrome. *Frontiers in immunology*. 2017; 8:1797. Epub 2018/01/10. <https://doi.org/10.3389/fimmu.2017.01797> PMID: 29312318; PubMed Central PMCID: PMC5732971.
42. Kang HJ, Stanley WT, Esselstyn JA, Gu SH, Yanagihara R. Expanded host diversity and geographic distribution of hantaviruses in sub-Saharan Africa. *J Virol*. 2014; 88(13):7663–7. <https://doi.org/10.1128/JVI.00285-14> PMID: 24741077; PubMed Central PMCID: PMC4054438.
43. Razzauti M, Plyusnina A, Henttonen H, Plyusnin A. Microevolution of Puumala hantavirus during a complete population cycle of its host, the bank vole (*Myodes glareolus*). *PLoS ONE*. 2013; 8(5):e64447. <https://doi.org/10.1371/journal.pone.0064447> PMID: 23717616; PubMed Central PMCID: PMC3661530.
44. Beaty BJ, Sundin DR, Chandler LJ, Bishop DH. Evolution of bunyaviruses by genome reassortment in dually infected mosquitoes (*Aedes triseriatus*). *Science (New York, NY)*. 1985; 230(4725):548–50. Epub 1985/11/01. <https://doi.org/10.1126/science.4048949> PMID: 4048949.
45. Szabo R, Radosa L, Lickova M, Slavikova M, Heroldova M, Stanko M, et al. Phylogenetic analysis of Puumala virus strains from Central Europe highlights the need for a full-genome perspective on hantavirus evolution. *Virus genes*. 2017; 53(6):913–7. Epub 2017/07/01. <https://doi.org/10.1007/s11262-017-1484-5> PMID: 28664467.
46. Ling J, Smura T, Tamarit D, Huitu O, Voutilainen L, Henttonen H, et al. Evolution and postglacial colonization of Seewis hantavirus with *Sorex araneus* in Finland. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2018; 57:88–97. Epub 2017/11/15. <https://doi.org/10.1016/j.meegid.2017.11.010> PMID: 29133028.
47. Lee SH, Kim WK, No JS, Kim JA, Kim JI, Gu SH, et al. Dynamic Circulation and Genetic Exchange of a Shrew-borne Hantavirus, Imjin virus, in the Republic of Korea. *Scientific reports*. 2017; 7:44369. Epub 2017/03/16. <https://doi.org/10.1038/srep44369> PMID: 28295052; PubMed Central PMCID: PMC5353647.
48. Klempa B. Reassortment events in the evolution of hantaviruses. *Virus genes*. 2018; 54(5):638–46. Epub 2018/07/27. <https://doi.org/10.1007/s11262-018-1590-z> PMID: 30047031; PubMed Central PMCID: PMC6153690.

49. Guterres A, de Oliveira RC, Fernandes J, Maia RM, Teixeira BR, Oliveira FCG, et al. Co-circulation of Araraquara and Juquitiba Hantavirus in Brazilian Cerrado. *Microbial ecology*. 2018; 75(3):783–9. Epub 2017/09/01. <https://doi.org/10.1007/s00248-017-1061-4> PMID: 28856421.
50. Nikolic V, Stajkovic N, Stamenkovic G, Cekanac R, Marusic P, Siljic M, et al. Evidence of recombination in Tula virus strains from Serbia. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2014; 21:472–8. <https://doi.org/10.1016/j.meegid.2013.08.020> PMID: 24008094.
51. Sironen T, Vaheri A, Plyusnin A. Molecular evolution of Puumala hantavirus. *J Virol*. 2001; 75(23):11803–10. <https://doi.org/10.1128/JVI.75.23.11803-11810.2001> PMID: 11689661; PubMed Central PMCID: PMC114766.
52. Chare ER, Gould EA, Holmes EC. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *The Journal of general virology*. 2003; 84(Pt 10):2691–703. Epub 2003/09/19. <https://doi.org/10.1099/vir.0.19277-0> PMID: 13679603.
53. Wu W, Guo JQ, Yin ZH, Wang P, Zhou BS. GIS-based spatial, temporal, and space-time analysis of haemorrhagic fever with renal syndrome. *Epidemiol Infect*. 2009; 137(12):1766–75. <https://doi.org/10.1017/S0950268809002659> PMID: 19393118.
54. Zhang WY, Wang LY, Liu YX, Yin WW, Hu WB, Magalhaes RJ, et al. Spatiotemporal transmission dynamics of hemorrhagic fever with renal syndrome in China, 2005–2012. *PLoS Negl Trop Dis*. 2014; 8(11):e3344. <https://doi.org/10.1371/journal.pntd.0003344> PMID: 25412324; PubMed Central PMCID: PMC4239011.
55. Kim JA, Kim WK, No JS, Lee SH, Lee SY, Kim JH, et al. Genetic Diversity and Reassortment of Hantaan Virus Tripartite RNA Genomes in Nature, the Republic of Korea. *PLoS Negl Trop Dis*. 2016; 10(6):e0004650. Epub 2016/06/18. <https://doi.org/10.1371/journal.pntd.0004650> PMID: 27315053; PubMed Central PMCID: PMC4912082.
56. Kim WK, Kim JA, Song DH, Lee D, Kim YC, Lee SY, et al. Phylogeographic analysis of hemorrhagic fever with renal syndrome patients using multiplex PCR-based next generation sequencing. *Scientific reports*. 2016; 6:26017. Epub 2016/05/26. <https://doi.org/10.1038/srep26017> PMID: 27221218; PubMed Central PMCID: PMC4879520.
57. Kim WK, No JS, Lee D, Jung J, Park H, Yi Y, et al. Active Targeted Surveillance to Identify Sites of Emergence of Hantavirus. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2020; 70(3):464–73. Epub 2019/03/21. <https://doi.org/10.1093/cid/ciz234> PMID: 30891596.
58. Wang DM, Wang ZX, Tong YB, Liu M, Cai XH, Hu LJ, et al. [Surveillance on hemorrhagic fever with renal syndrome in Guizhou during 1984–2000]. *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi*. 2003; 24(8):694–6. Epub 2003/10/03. PMID: 14521790.
59. Wang ZX. [Preliminary study on epidemic hemorrhagic fever (EHF) in Guizhou Province]. *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi*. 1989; 10(1):1–5. Epub 1989/02/01. PMID: 2567626.
60. Hille A, Tarkhnishvili D, Meinig H, Hutterer R. Morphometric, biochemical and molecular traits in Caucasian wood mice (*podemus/Sylvaemus*), with remarks on species divergence. *Acta Theriologica*. 2002; 47(4):389–416. <https://doi.org/10.1007/bf03192465>
61. Liu X, Wei F, Li M, Jiang X, Feng Z, Hu J. Molecular phylogeny and taxonomy of wood mice (genus *Apodemus* Kaup, 1829) based on complete mtDNA cytochrome b sequences, with emphasis on Chinese species. *Molecular phylogenetics and evolution*. 2004; 33(1):1–15. Epub 2004/08/25. <https://doi.org/10.1016/j.ympev.2004.05.011> PMID: 15324834.
62. Chen HX, Qiu FX. Epidemiologic surveillance on the hemorrhagic fever with renal syndrome in China. *Chinese medical journal*. 1993; 106(11):857–63. Epub 1993/11/01. PMID: 7908258.
63. Klein SL, Calisher CH. Emergence and Persistence of Hantaviruses. In: Childs JE, Mackenzie JS, Richt JA, editors. *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 217–52.
64. Ramsden C, Holmes EC, Charleston MA. Hantavirus Evolution in Relation to Its Rodent and Insectivore Hosts: No Evidence for Codivergence. 2008. <https://doi.org/10.1093/molbev/msn234> PMID: 18922760
65. Souza WM, Bello G, Amarilla AA, Alfonso HL, Aquino VH, Figueiredo LTM. Phylogeography and evolutionary history of rodent-borne hantaviruses. *INFECTION, GENETICS AND EVOLUTION*. 2014; 21:198–204. <https://doi.org/10.1016/j.meegid.2013.11.015> PMID: 24287104
66. Nemirov K, Henttonen H, Vaheri A, Plyusnin A. Phylogenetic evidence for host switching in the evolution of hantaviruses carried by *Apodemus* mice. *Virus research*. 2002; 90(1–2):207–15. Epub 2002/11/30. [https://doi.org/10.1016/s0168-1702\(02\)00179-x](https://doi.org/10.1016/s0168-1702(02)00179-x) PMID: 12457975.
67. Hughes AL, Robert F. Evolutionary Diversification of Protein-Coding Genes of Hantaviruses. *Molecular Biology & Evolution*. 2000;(10):10.

68. Jackson AP, Charleston MA. A cophylogenetic perspective of RNA-virus evolution. *Mol Biol Evol.* 2004; 21(1):45–57. Epub 2003/09/02. <https://doi.org/10.1093/molbev/msg232> PMID: 12949128.
69. Wertheim JO, Kosakovsky Pond SL. Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol.* 2011; 28(12):3355–65. Epub 2011/06/28. <https://doi.org/10.1093/molbev/msr170> PMID: 21705379; PubMed Central PMCID: PMC3247791.
70. Grafen A. The phylogenetic regression. *Philosophical transactions of the Royal Society of London Series B, Biological sciences.* 1989; 326(1233):119–57. Epub 1989/12/21. <https://doi.org/10.1098/rstb.1989.0106> PMID: 2575770.
71. Zhang WY, Guo WD, Fang LQ, Li CP, Bi P, Glass GE, et al. Climate variability and hemorrhagic fever with renal syndrome transmission in Northeastern China. *Environmental health perspectives.* 2010; 118(7):915–20. Epub 2010/02/10. <https://doi.org/10.1289/ehp.0901504> PMID: 20142167; PubMed Central PMCID: PMC2920909.
72. Clement J, Vercauteren J, Verstraeten WW, Ducoffre G, Barrios JM, Vandamme A-M, et al. Relating increasing hantavirus incidences to the changing climate: the mast connection. *International Journal of Health Geographics.* 2009; 8(1):1. <https://doi.org/10.1186/1476-072X-8-1> PMID: 19149870
73. Tian HY, Yu PB, Luis AD, Bi P, Cazelles B, Laine M, et al. Changes in rodent abundance and weather conditions potentially drive hemorrhagic fever with renal syndrome outbreaks in Xi'an, China, 2005–2012. *PLoS Negl Trop Dis.* 2015; 9(3):e0003530. <https://doi.org/10.1371/journal.pntd.0003530> PMID: 25822936; PubMed Central PMCID: PMC4378853.
74. McCarthy M. Global climate is warming rapidly, US draft report warns. *BMJ.* 2017; 358:j3824. <https://doi.org/10.1136/bmj.j3824> PMID: 28794049.
75. Tian H, Yu P, Cazelles B, Xu L, Tan H, Yang J, et al. Interannual cycles of Hantaan virus outbreaks at the human-animal interface in Central China are controlled by temperature and rainfall. *Proc Natl Acad Sci U S A.* 2017; 114(30):8041–6. <https://doi.org/10.1073/pnas.1701777114> PMID: 28696305; PubMed Central PMCID: PMC5544290.
76. Zheng Z, Wang P, Wang Z, Zhang D, Wang X, Zuo S, et al. The characteristics of current natural foci of hemorrhagic fever with renal syndrome in Shandong Province, China, 2012–2015. *PLoS Negl Trop Dis.* 2019; 13(5):e0007148. Epub 2019/05/21. <https://doi.org/10.1371/journal.pntd.0007148> PMID: 31107874; PubMed Central PMCID: PMC6544330.
77. Tian H, Stenseth NC. The ecological dynamics of hantavirus diseases: From environmental variability to disease prevention largely based on data from China. *PLoS Negl Trop Dis.* 2019; 13(2):e0006901. Epub 2019/02/23. <https://doi.org/10.1371/journal.pntd.0006901> PMID: 30789905; PubMed Central PMCID: PMC6383869.