Cross-species Transmission in the Speciation of the Currently Known 

Murinae-associated Hantaviruses 

Running title: Speciation of Murinae-associated hantaviruses

Xian-Dan Lin1,2*, Wen Wang1*, Wen-Ping Guo1, Xiao-He Zhang2, Jian-Guang Xing3, Sheng-Ze Chen4, Ming-Hui Li1, Yi Chen2, Jianguo Xu1, Alexander Plyusnin1,5, Yong-Zhen Zhang1* 

1 State Key Laboratory for Infectious Disease Prevention and Control, Department of Zoonoses, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Liuzi 5, 102206, Beijing, China; 

2 Wenzhou Center for Disease Control and Prevention, Wenzhou, 325001, Zhejiang province, China; 

3 Wencheng District Center for Disease Control and Prevention, Wencheng, 325005, Wenzhou, Zhejiang province, China; 

4 Yongjia Center for Disease Control and Prevention, Yongjia, 325200, Wenzhou, Zhejiang province, China; 

5 Department of Virology, Infection Biology Research Program, Haartman Institute, University of Helsinki, Finland 

*Contributed to this work equally.

*Correspondence to Dr. Yong-Zhen Zhang, State Key Laboratory for Infectious Disease Prevention and Control, National Institute of Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping Liuzi 5, Beijing, 102206, China.

The total number of words in the main text of the paper is 5528.

The number of words in the Abstract is 202.

The number of figures is 5.

The number of tables is 4.

The number of supplementary tables is 8.
ABSTRACT

To gain more insights into the phylogeny of Dabieshan virus (DBSV) carried by *Niviventer confucianus* and other *Murinae*-associated hantaviruses, genome sequences of novel variants of DBSV were recovered from *Niviventer* rats trapped in the mountainous areas of Wenzhou, China. Genetic analyses show all known genetic variants of DBSV including the ones identified in this study are distinct from other *Murinae*-associated hantaviruses. DBSV variants show geographic clustering and high intra-species diversity. The data suggest that DBSV is a distinct species in the genus *Hantavirus*. Interestingly, DBSV shows the highest sequence identity to Hantaan virus (HTNV) with >7% difference in the sequences of N, GPC, and L proteins; while *N. confucianus* is more closely related to *Rattus norvegicus* (the host of Seoul virus, SEOV) than to *Apodemus agrarius* (the host of HTNV) and Saaremaa virus (SAAV). Further genetic analyses of all known *Murinae*-associated hantaviruses (both established and tentative species) show many of them including DBSV may have originated from host-switching. The estimation of evolutionary rates and divergence time supports the role of cross-species transmission in the evolution of *Murinae*-associated hantaviruses. The detection of positive selection suggests that genetic drift may contribute to the speciation of *Murinae*-associated hantaviruses and adaptation has a role as well.

**Key words:** *Niviventer confucianus*, Hantavirus, Cross-species transmission, Genetic drift, Adaptation, Speciation,
INTRODUCTION

New emerging viral pathogens, e.g., avian and swine influenza viruses (28, 41), severe acute respiratory syndrome-coronavirus (17), and human immunodeficiency virus (16, 51), cause epidemics (or pandemics) in humans by changing or expanding their host range. These pathogens are a considerable threat to human and/or wildlife health, agricultural production, and public security (5, 36). Almost all of the novel viruses have circulated in their reservoir hosts for a long time before emerging in humans or other animals (11, 36, 37). Zoonotic viral pathogens such as hantaviruses and rabies virus show high genetic diversity that depends on natural hosts or geographic origins (7, 22, 43). The role of cross-species transmission in the generation of a new virus species should be studied in greater detail (26), and better understanding of the evolutionary relationship between zoonotic pathogens and their hosts may help in the prevention and control of (re)emerging diseases.

The hantavirus genome consists of three RNA segments: small (S), medium (M), and large (L) segments; they encode respectively the nucleocapsid (N) protein (in some hantaviruses, the nonstructural NSs protein), the glycoprotein precursor (GPC) of the two envelope glycoproteins (Gn and Gc), and the viral RNA-dependent RNA polymerase (RdRP, the L protein) (44). At least 23 established and 30 tentative hantavirus species have been identified worldwide in rodents and insectivores (42). Identification of insectivore-carried hantaviruses has increased especially rapidly during the last five years (22, 24). Each species of these known hantaviruses is specifically associated with one or several closely related rodent or insectivore hosts (42). As the phylogeny of hantaviruses may be congruent with their hosts, hantaviruses are considered to have co-evolved (co-speciated) with their respective rodent or insectivore hosts (20, 21, 23, 25, 38, 39, 43, 44, 52). Recently, Ramsden et al. proposed there was no co-divergence between hantaviruses and their hosts, and the similarities between the phylogenies of hantaviruses and
their hosts are the result of a more recent history of preferential host switching and local adaptation (46). Further studies are needed to determine if this is true.

The association between hantaviruses and their hosts is relatively specific, although host associations can include two or more animal species such as Hantaan virus (HTNV) in mice species *Apodemus agrarius* and *A. peninsulae* (72), and Seoul virus (SEOV) in several rat species (*Rattus norvegicus, R. flavipectus, R. losea, and R. nitidus*) (61, 69, 70). Cross-species transmission (host-switching or host jump) between more distantly related rodent hosts is suggested to occur during the evolution of several hantaviruses, e.g. Topografov virus (TOPV, *Lemmus/Microtus*) (59) and Limestone Canyon virus (LSCV, *Peromyscus/Reithrodontomys* [47]). Among the hantaviruses identified recently in insectivores, evidence for host-switching between different families of insectivores within the order Soricomorpha was reported (4, 23). Recent studies show hantaviruses carried by *Murinae* rodents are related closely to some Soricidae-carried viruses rather than to those associated with Cricetidae rodents (4, 23, 46); and Rockport virus (RKPV) in *Scalopus aquaticus* probably originated from rodents (24). In addition, host jumping of hantaviruses was observed frequently in rodents from the New World (10). Thus, cross-species transmission may be more common than previously shown, and may be the important driving force in the hantavirus evolution and speciation.

Dabieshan (or Da Bie Shan) virus (DBSV) was first isolated from *Niviventer confucianus* (Chinese white-bellied rat) captured in the Dabieshan mountainous area of Anhui Province, China (61). Genetically, the virus is more closely related to HTNV carried by *Apodemus* mice (<15 % amino acid [aa] differences) than to SEOV carried by *Rattus* rats (>25% aa differences) where DBSV was initially recognized as a subtype of HTNV. As the *Niviventer* rat is more closely related to *Rattus* than to *Apodemus*, it was suggested that DBSV jumped to *N. confucianus* from *Apodemus* mice (61). Recently, a new lineage of DBSV was detected in *N. confucianus* captured in Yunnan Province, China (8). Interestingly, DBSV has not been found...
outside these two regions; whereas *N. confucianus* is widespread and abundant in 27 provinces of China (67). To further characterize the genetics of DBSV, we performed a survey in the mountainous areas in Wenzhou, Zhejiang Province where HFRS is endemic (69) and performed (phylo) genetic analyses of the hantaviruses associated with *N. confucianus*. Additionally, to clarify the role of host-switching and subsequent random and directional events in the speciation of *Murinae*-associated hantaviruses, we also analyzed the phylogenetic pattern of cross-species transmission from other *Murinae*-associated hantaviruses.

**MATERIALS AND METHODS**

**Trapping of small animals and screening for hantaviruses**

During the spring and autumn of 2008, rodents were captured in the mountain areas in Wencheng (34°27′–59°27′N, 34°119′–15′120°E) and Yongjia (28°28′–33°28′N; 29°120′–36°120°E) counties in Wenzhou (Fig. 1) where more than 70% of the total area is mountainous. Small animals were trapped using a cage (20cm×20cm×56cm) with a treadle release mechanism using deep-fried dough sticks as bait. The cages were set at five meter intervals according to the protocols described previously (34). Trapped animals were identified by morphological examination according to criteria reported by Chen (9); and further verified using sequence analysis of the cytochrome b (Cyt-b) gene (71). Lung and kidney tissue samples obtained from trapped animals were stored in liquid nitrogen. Hantavirus-specific antigen in lungs was detected using an indirect immunofluorescent assay (IFA) as described previously (68). In addition, hantaviral RNA was detected using RT-PCR as described by Klempa *et al.* (27).

**Preparation of viral RNA and rodent mitochondrial DNA (mtDNA)**
TRIzol reagent (Invitrogen, Carlsbad, CA) was used to extract total RNA from the viral antigen-positive lung tissue samples according to the manufacturer’s instruction. A genomic DNA extraction kit (SBS, Beijing, China) was used to extract mtDNA from rat lung tissue samples according to the manufacturer’s protocol.

Amplification of the viral genome and rodent cytochrome b (Cyt-b) gene and sequencing

To amplify the L, M, and S segment sequences, primer P14 (49) was used to synthesize cDNA using AMV reverse transcriptase (Promega, Beijing). Partial L segment sequences (nt 3,008-3325) were obtained by using nested PCR with two primer pairs (27). The complete M and S segment sequences were amplified as described previously (74). The sequence of the rat Cyt-b gene was recovered using a standard PCR with primers CB1 and CB2 (31).

The obtained viral sequences and rat Cyt-b gene sequences were purified using agarose gel electrophoresis and sequenced using the ABI-PRISM Dye Termination Sequencing kit and an ABI 373-A genetic analyzer.

Phylogenetic analyses

The RDP, GENECONV, bootscan, maximum Chi square, Chimera, SISCAN and 3SEQ recombination detection methods used in RDP3 (33) were employed to detect potential recombinant viral sequences, identify likely parental viral sequences, and localize possible recombination breakpoints. The analyses were performed with default settings for the different test methods; and a Bonferroni corrected P-value cut-off of 0.05. When events were observed with two or more methods and with significant phylogenetic support, the viral sequences were considered recombinant and excluded from this study.

The viral genome and rodent Cyt-b sequences were aligned using Clustal W program (version 1.83). Their nucleotide (nt) and amino acid (aa) identities were calculated using the DNAStar program. The Metropolis-coupled Markov Chain Monte Carlo (MCMC) method in
MrBayes v3.1.2 was employed to estimate the phylogenetic trees. General time reversible (GTR) with a gamma-distribution model of among the site rate heterogeneity and a proportion of invariable sites (GTR+\Gamma+I) was found to be the best model for the open reading frame (ORF) of the S segment and GTR +\Gamma for the ORF of the M segment and the partial L segment was determined using jModelTest version 0.1 (45). Bayesian analysis consisted of 4 million MCMC generations sampled every 100 generations to ensure convergence across two runs of three hot chains and one cold chain. The performance was continued until the average standard deviation of split frequencies was less than 0.01 with a 25% burn-in. Convergence of parameters were assessed by calculating the Effective Sample Size (ESS) using TRACER v1.5 (http://tree.bio.ed.ac.uk/software/tracer) with an acceptable threshold of ESS of over 200. RAxML Blackbox web-server was employed to construct Maximum-Likelihood (ML) trees (54). The Dendroscope program (2.4) was used to visualize the tree files. Bayesian trees were readdressed to construct a tanglegram of rodent host and associated hantavirus using Tree-Map software (2.0b) (21, 59). The Markov model in TreeMap was used to test significance by reconstructing 1,000 hantavirus trees with randomized branches and mapping these random trees onto the fixed host tree. When the level of congruence of the “real” virus tree was no more than that expected between randomly generated trees \(P<0.05\) for both the number of co-divergence events [CEs] and non-co-divergence events [NCEs]), co-divergence was not supported.

Estimating the rates of nucleotide substitution and the TMRCA of Murinae-associated hantaviruses

Bayesian MCMC approach available in the BEAST v1.6.0 software package (14) was used to estimate the rates of nucleotide substitution and divergence time (i.e., the most recent common ancestors [TMRCA]) in Murinae-associated hantaviruses for the S and M segment with uncertainty in all estimates reflected in the 95% high probability density (HPD) intervals.
The ORF sequences of the S and M segment for the year-of-sampling that was available were used to generate the datasets with the recombinant sequences excluded. 48 S and 40 M segment ORF sequences were compiled into the data set 1 and the data set 2, respectively (Table S1).

The DAMBE program was used to determine the level of saturation at each codon position. When saturation was observed at the third position of the S segment and first and third positions of the M segment, these positions were removed. BEAUTi v1.6.0 was used to generate BEAST XML input files with both the strict and uncorrelated lognormal-distribution relaxed molecular clock model. When separate partitions of codon position sites were analyzed, we used GTR+Γ+I for the ORF sequence of the S segment and GTR+Γ for the M segment sequence determined by jModelTest version 0.1: and we used both the constant and the Extended Bayesian Skyline tree before all analyses. Two independent runs were taken for each dataset with sampling every 1,000 generations. Each run was continued until the ESS of all parameters was larger than 200. Tracer v1.5 was employed to summarize, analyze, and visualize the resulting posterior sample. A maximum clade credibility (MCC) tree with a burn-in of 10% of the sampled trees was constructed to summarize the sample of trees produced by each BEAST run using the TreeAnnotator program (v1.6.0). The Bayes Factor (BF) was estimated to determine the best clock and tree prior model with Tracer v1.5. An uncorrected lognormal-distribution relaxed molecular clock model and extended Bayesian Skyline tree prior fit was used for both ORF sequences of the S and M segments according to the BF analysis data (Table S2). The temporal signals in both date set 1 and 2 were also evaluated. Under the best model, BEAST analyses were repeated for the data sets where sampling times were randomized running five times for the randomized data. When the mean rates and 95% HPDs from the real data set had major differences from those from re-sampled data, these samples were considered to contain clear temporal structure.
We report posterior probabilities for the nodes in the MCC tree of >0.7 using FigTree v1.3.1 (http://tree.bio.ed.ac.uk). BEAST analyses were used to estimate rooted phylogenetic trees where a time scale was incorporated according to rates of evolution estimated for each branch of the related viral sequences.

**Determination of Signature Amino Acid Markers**

Based on the relationship of hantaviruses or their hosts, several groups were defined (Tables 2, S3 and S4). Viral N and GPC proteins deduced from S and M segments sequences were aligned using MegAlign in the DNASTar program. If a specific aa exists in one species or one group, but not in other species or group, this aa is considered as a “signature aa” marker (or synapomorphy).

**Analysis of selection pressures**

The program Codeml in the PAML 4.4c software package was employed to detect positive selected sites in the N and GPC proteins (64). In each Murinae-associated hantavirus species, several sequences were selected for detection of positive selection. No sequence was identical to the others (57). Both date set 1 and 2 were tested to determine if they were under positive selection. Three kinds of models (branch-specific, site-specific, and branch-site) were used to detect selective pressure among different branches and at different sites as described by Tang et al. (57). Comparing the models which do not allow for positive selection with the models in which positive selection is allowed, the likelihood ratio test (LRT) was used to find the presence of positively selected sites (3). It was assumed that all branches and sites in the phylogeny had the same $\omega$ ratio in the one ratio model (M0), and that each branch in the phylogeny had an independent $\omega$ ratio in the free-ratio (FR) model. The difference in $\omega$ ratios could be determined by comparing M0 and FR to LRT. The discrete model (M8) was used to estimate $\omega$ for three classes of codons. The variability of selective pressure among sites was estimated by comparing...
M7 and M8. When positive selection (ω > 1) was found, posterior probabilities were estimated for site classes using the Bayes Empirical Bayes (BEB) method (64).

The branch-site model that assumes that the ω ratio varies both among sites and among branches (65, 66) was also used to find positive selected sites; and was used when adaptive evolution occurred at a few time points and only affected a few aa residues. For the branch-site model A (model A), a given virus of interest (species or tentative species) was set as the foreground, and the other viruses as the background. We assumed selective constraint would change across sites both in the foreground and background with a few sites that change only along foreground lineages. There were three ω ratios for the foreground (0 < ω0 < 1, ω1 = 1, ω2 > 1) and two ω ratios for background (0 < ω0 < 1, ω1 = 1) in the branch-site model A. When positive selection (ω2 > 1) was found, posterior probabilities were estimated for site classes using the BEB method. The null model (model A') was the same as model A, but ω2 = 1 was fixed. For the S and M segments sequence, we applied branch-site models to 14 groups on the trees, respectively. Thus 0.0036 was used as the significance level for both S and M segments sequences.

RESULTS

Trapping of Rodents and Screening for hantaviruses

From February to October 2008, a total of 149 small animals belonging to seven species were captured in the mountainous areas of Wenzhou, Zhejiang Province. Of these, 70 (41 N. confucianus, 15 A. agrarius, nine R. flavipectus, two R. losea, and three Suncus murinus) were trapped from Wencheng County and 79 (25 N. confucianus, 25 A. agrarius, two R. losea, one R. nitidus, 25 S. murinus, and one Callosciurus erythraeus) from Yongjia County. All small animals were screened for the presence of hantaviral antigens using IFA; hantaviral antigens
were identified in the lung tissue samples from only three *N. confucianus* from Wencheng (samples Wencheng-Nc-427, Wencheng-Nc-469, and Wencheng-Nc-470) and four *N. confucianus* from Yongjia (samples Yongjia-Nc-15, Yongjia-Nc-38, Yongjia-Nc-58, and Yongjia-Nc-95). Results of the RT-PCR test were in full agreement with the IFA-data.

**Genetic analysis of viral sequences**

To characterize the *N. confucianus*-associated hantavirus found in Wenzhou, the complete hantaviral S and M and also partial L segment sequences were recovered from all positive *N. confucianus* (Table S1). The complete S segment has a total length of 1,725 nt including 36 nt in the 5′ non-coding region (NCR), an ORF encoding the N protein of 429 aa, and the 399 nt-long 3′ NCR. Comparison of these complete S sequences showed they shared 95.3-99.9% nt identity which corresponded to 99.1-100% identity in the deduced aa sequences (Table S5). Further comparison with other known hantaviruses show the novel strains were more closely related to DBSV strains Nc167 and AH09 identified in the Dabieshan mountain areas of Anhui Province (61): 88.7-89.7% nt sequence identity and 98.4-98.8% aa sequence identity, followed by the DBSV strain YN509 identified in Yunnan province (8): 83.4-84%/97.9-98.4%. In agreement with the previous studies (61), the DBSV variants carried by *N. confucianus* were more closely related to HTNV (78.0-79.1%/ 92.1-93.5%) than to SEOV (74.4-75.4%/83-84.6%) and other hantaviruses (34.6-79.9/45.9-93.2%).

The M segment of the novel DBSV variants/strains has a total length of 3,623 nt (3,645 nt in Yongjia-Nc-38) including 46 nt of the 5′ NCR, an ORF for the GPC precursor of 1,133 aa, and 175 nt of the 3′ NCR (197nt in Yongjia-Nc-38). Like the complete S segment sequences, the complete M and partial L segment sequences of these DBSV strains showed similar patterns of relatedness to other DBSV variants (from Dabieshan and Yunnan) and to other known hantaviruses (Table S5).
Phylogenetic relationships of viral sequences

Phylogenetic analysis of the complete coding regions of the S and M segments, and partial L segment sequences was performed using the Bayesian method implemented in the MrBayes v3.1.2 program package. The branching patterns of S-, M-, and L-trees constructed using the classic ML method were similar to those based on the Bayesian method (>70% of bootstrap support values shown in Fig. 2). In the phylogenetic tree based on the S segment ORF, Wencheng and Yongjia variants from Zhejiang identified in this study clustered (Fig. 2A). They formed a well-supported group with the other DBSV strains: Nc167, AH09, and YN509; and showed three geographic lineages (Anhui, Zhejiang, and Yunnan Provinces). Notably, the strains identified in this study were more closely related to strains Nc167 and AH09 isolated in the Dabieshan mountain regions of Anhui Province that shares the border with Zhejiang Province, than to strain YN509 detected in Yunnan Province which is distant from Zhejiang Province (Fig. 1). In agreement with the previous studies (8, 61), DBSV shows a closer evolutionary relationship to HTNV and Amur/Soochong virus (ASV) carried by Apodemus mice than with those carried by Rattus rats, even though Niviventer is more closely related to Rattus species than to Apodemus species (see below, Fig. 3). The topologies of the M- and L-trees were similar to the S-tree (Fig. 2B and 2C). These results support the hypothesis of a cross-species transmission (host-switching) of hantavirus between Niviventer rats and Apodemus mice (61) with a presumed direction of virus transmission not from Apodemus mice to Niviventer rats but in the opposite direction.

Phylogenetic relationships between Murinae-associated viruses and their rodent carriers

Rodent migrations may have led to the current geographical distribution of hantaviruses (43). Phylogenetic analysis of the Cyt-b gene sequences of N. confucianus collected in this study may show the evolutionary relationships between DBSV and its host (N. confucianus). Overall, all available Cyt-b gene sequences of N. confucianus exhibited a high genetic diversity: up to 9.1%.
The *N. confucianus* sequences obtained from Wenzhou showed 9.1% nt divergence from the Yunnan sequences. As shown in Fig. 3, all Cyt-b gene sequences clustered and formed a well-supported *N. confucianus* clade that could be divided into the two well supported lineages (with posterior node probabilities of 0.98 and 1.00). Similar to the results of previous studies (55), *N. confucianus* appears to be more closely related to *Rattus* than to *Apodemus*, and all sequences were grouped into the two major groups corresponding to mice and rats (Fig. 3). The topologies of the trees constructed using classical ML method and the Bayesian method were the same (Fig. 3, only bootstrap values >70% are shown).

At present, the list of *Murinae*-associated hantaviruses defined by the International Committee on Taxonomy of Viruses includes five established species and five tentative species (42). To evaluate the evolutionary relationship between these species and their corresponding rodent hosts, TreeMap 2.0 was used to compare the viral S segment tree and the host Cyt-b gene tree. As shown in Fig. 4, the nodes of the viral phylogeny were similar in topology compared to the associated nodes of the *Murinae* host tree (P<0.05) as measured by CEs (P=0.033±0.004) and the NCEs (P=0.019±0.003) frequency. All known *Murinae*-associated hantaviruses group together with some shrew-born hantaviruses e.g. Tanganya virus (TGNV) from Africa (4, 23, 46). As the location of TGNV is the nearest to the ancestral node separating the shrew-born hantaviruses from other species (Fig. 2) (46), the ancestor of the present known *Murinae*-associated rodent hantaviruses may have originated from the African *Hylomyscus*-like species. If true, at least nine cross-species transmission (host switching/host jump) events could be proposed to occur during evolution of the presently known *Murinae*-associated hantaviruses (Table 1). Six host switching events probably occurred between the different rodent genera: *Hylomyscus* mouse and *Apodemus* mouse, *Hylomyscus* (*-Apodemus*) wood mice and rats, *Bandicota* rat and *Rattus rattus* (*R. tanezumi*), *Bandicota* rat and *R. norvegicus*, *Bandicota* rat and *Niviventer rat*, *Niviventer rat* and *Apodemus* mouse. Another three cross-species
transmission events may have occurred between species within genus *Apodemus* or genus *Rattus*.

**Rates of hantavirus evolution and diversification dates**

The date randomization test was used to determine if the structure and spread of the sequence ages were sufficient to estimate substitution rates and divergence times. The original estimate on data sets 1 and 2 was not recovered in the date-randomized datasets, suggesting that there was sufficient temporal structure in these data. Calculations based on data set 1 using the Bayesian MCMC analysis showed that the mean evolutionary rate in the *Murinae*-associated hantaviruses was $2.0 \times 10^{-4}$ subs/site/year with a 95% HPD from $1.1 \times 10^{-4}$ to $2.9 \times 10^{-4}$ subs/site/year (Table 3). As shown in Fig. 5A and Table 3, the time to TMRCA estimated for all known DBSV variants based on the currently sampled genetic diversity was between 418 and 1,273 years before present (ybp); between 386 and 1,129 ybp for the HTNV variants; and between 78 and 199 ybp for the SEOV variants. Notably, the mean divergence time estimated for DBSV and HTNV was 1,590 ybp, 983 ybp for HTNV and ASV, 581 ybp for Dobrava-Belgrade virus (DOBV) and SAAV, 515 ybp for Gou virus (GOUV) and SEOV.

The evolutionary rate estimated based on data set 2 was similar to that of the S segment (Fig. 5B and Table S6). The evolutionary rates estimated in this study were in agreement with our recent results (29) and also those of Ramsden *et al.* (46); however these rates appeared much higher than the previous estimates assuming a history of co-divergence between hantaviruses and their hosts ($\sim 10^{-6}$-$10^{-7}$ subs/site/year; 20, 39, 52).

**Genetic analysis of viruses originating from cross-species transmission**

For the viruses that may have originated via cross-species transmission, DBSV had $>7\%$ aa sequence difference in the complete N, GPC, and L protein sequences from a sister virus, HTNV (Table 1). $\geq 7\%$ aa sequence difference in all three protein sequences was also observed between
pairs of hantaviruses (Sangassou [SANGV] and SAAV, Thailand virus [THAIV] and SEOV, THAIV and DBSV). However, >7% aa sequence difference was found in GPC and L protein sequences between THAIV and Serang virus (SERV), and only in the GPC protein sequences between HTNV and ASV. Further, the difference was >7% in all three protein sequences in the pairs DOBV/SAAV and SEOV/GOUV.

Distinct hantaviruses possess specific “signature aa” in the N and GPC protein sequences, e.g. HTNV and Puumala virus (PUUV) (52, 73). As shown in Tables 2, S3, and S4, eight of ten species (except SEOV and GOUV) have their own specific “signature aa” in the N protein sequence, and all ten species have their own specific “signature aa” in the GPC protein sequence: SANGV (18/70), DOBV (1/10), SAAV (2/14), THAIV (1/27), SERV (1/18), SEOV (0/3), GOUV (0/4), DBSV (3/24), HTNV (2/8), and ASV (1/10). Remarkably, some “signature aa” are shared within the phylogenetic groups of hantavirus species, e.g. DBSV-HTNV-ASV (16/35), DOBV-SAAV (16/40), and SEOV-GOUV (25/55). Interestingly, some “signature aa” are also shared within the hantaviruses where their hosts are closely related such as the mouse group (SANGV+ ASV+HTNV+ DOBV+SAAV [1/3]) and the rat group (DBSV+SEOV+GOUV+THAIV+SERV [2/2]), suggesting that these viruses may have had similar adaptive selection in mice or rats. In addition, the GPC protein may have faced a higher positive selective pressure than the N protein even if the nt sequence divergences of the S and the M segments were similar (Tables 1, 3, and S6).

Selection pressures in the hantavirus S and M segments

Data sets 1 and 2 were used to analyze the adaptation of Murinae–associated hantaviruses to their respective hosts. The analyses of branch-specific model (FR) did not show that selective pressure varied along the branches for both S and M segments (Tables 4, S7, and S8). Further, the site-specific model (M8) did not identify any sites under positive selection, although the \( \omega \) value was 3.00482 and 2.75162, respectively. For the S segment, no positive selection was
found in all seven species as well when using the branch-site model. However, when closely
related hantaviruses were grouped together, traces of positive selection were found in groups
DOBV-SAAV, DOBV-SAAV-SANGV, GOUV-SEOV, and SERV-THAIV (Tables 4 and S7).
Both LRT and BEB tests did not find these traces significant.

For the M segment, three sites in the DBSV GPC protein may be subjected to positive
selection ($\omega = 4.035147 \ [p = 0.027]$) when using the branch-site model. Presumable positive
selection was observed in ASV, HTNV, and SAAV, but not in SEOV, GOUV, and DOBV. The
results, however, were not statistically significant. Similar to the observations made concerning
the S segment, traces of positive selection were found in the following groups of closely related
viruses: DOBV-SAAV, DOBV-SAAV-SANGV, ASV-HTNV, ASV-DBSV-HTNV, GOUV-SEOV,
SERV-THAIV, and GOUV-SEOV-SERV-THAIV (Tables 4 and S8). Finally, 70 of 139 (50.4%)
positively selected sites in the N protein and GPC proteins appeared to have “signature aa”
markers (Table 4).

DISCUSSION

The data here suggest DBSV identified in China is a distinct species in the Hantavirus
genus. Wang and colleagues first isolated the virus (strain Nc167) from Niviventer rats trapped
in the Dabieshan mountain areas of Anhui Province in 2000 (61). They found the virus was
genetically related to but also distinct from HTNV and showed a 32-fold difference in titers
from HTNV in a two-way cross-neutralization test. The virus is considered as a lineage of
HTNV generated by host-switching from Apodemus mouse to Niviventer rat (61). Recently, the
virus was also found in Niviventer rats collected from Yunnan Province of China. This variant
shared approximately 82% nucleotide sequence identity with strain Nc167 (8). At present, the
virus carried by Niviventer rats is considered a tentative species in the Hantavirus genus and is
designated as DBSV (42). In this study, a new DBSV variant was found in the \textit{Niviventer} rats captured in the mountainous areas of Wenzhou, Zhejiang Province, China. Analyses of the strains identified in this study and those found previously show DBSV is carried by \textit{N. confucianus} in which no other hantaviruses have been found, and exhibits more than 7\% aa difference from any recognized hantaviruses in all three protein sequences. No re-assortants have been found among the known variants of DBSV. Thus, DBSV meets all four criteria for species demarcation in the \textit{Hantavirus} genus proposed by the International Committee on Taxonomy of Viruses (42). Further, the 15.4\% aa difference in the M segment between DBSV and other recognized hantaviruses is also more than the 12\% aa difference proposed by Maes \textit{et al.} as the demarcation criterion of hantavirus species (25). Thus, DBSV should be considered as a distinct species of \textit{Hantavirus}.

Generally, hantaviruses show close association with their respective rodent or insectivore host, and are thought to present a good example of co-divergence of a virus and a host (20, 21, 23, 25, 35, 39, 43, 44, 52). In agreement with the earlier studies (8, 61), our data show DBSV is more closely related to HNTV and ASV than other \textit{Murinae}-associated hantaviruses. For example, they share more signature aa than with any other hantavirus species or group of species, suggesting that they share a common ancestor. Moreover, our field trapping results also suggest that \textit{N. confucianus} has more chances to contact \textit{A. agrarius} than \textit{R. norvegicus}, which was not found. Conversely, \textit{Niviventer} rats have a closer evolutionary relationship with \textit{Rattus} rats than with \textit{Apodemus} mice (Fig. 3). Thus the phylogeny of DBSV, HNTV, and SEOV is not consistent with their respective hosts (Fig. 4). One possible explanation for such a discrepancy would be that the host-switching event occurred between rats and mice. However, the data presented here did not support the direction from \textit{Apodemus} mouse to \textit{Niviventer} rat (61). With the discovery of novel \textit{Murinae}-associated hantaviruses, the direction and time of the host-switching will become clear.
Cross-species transmission was reported previously among the hantaviruses. Cross-species transmission of HTNV from *A. agrarius* to *A. peninsulae*, and even to *Rattus, Niviventer*, and *Mus* species has been reported in China (61, 72, 73). Identification of SEOV in rat species other than *R. norvegicus* is also reported (61, 69, 70). Other studies also report one hantavirus carried by several rodent species (39, 48, 50). Comparing the phylogenies of hantaviruses and their hosts (Fig. 2 and 4, Table 1), we conclude that most Murinae-associated viruses may originate via host switching. Recent work by Ramseden and colleagues suggests there is no co-divergence between hantaviruses and their hosts, and the congruence between the phylogenies of hantaviruses and their hosts is the result of a more recent history of preferential host switching and local adaptation (46). Here, our data show cross-species transmission plays an important role in the speciation of the known Murinae-associated hantaviruses.

Both stochastic events (e.g. genetic drift, bottleneck) and deterministic processes (e.g. selection, adaptation) are occurring in a population. In a large population, even weak selection on a mutant may play an important role in its evolution (1). However, viruses can be particularly susceptible to the effects of genetic drift because inter-host transmission frequently involves population size “bottlenecks” that occur independently of viral fitness (12). Earlier studies found genetic drift effects during both intra-host and inter-host infection of viruses (2, 6, 15). All hantaviruses presumably originate from cross-species transmission except SAAV and GOUV, which show >7% sequence differences from their presumable parental/sister species in at least one of the encoded proteins (Table 1). Further, each species of Murinae-associated hantavirus has its own specific signature aa markers (Table 2). However, our analyses show only a few amino acid sites may have been under weak positive selection in the M segment proteins for ASV, DBSV, DOBV, GOUV, HTNV, and SAAV (Tables 4 and S8). Hantavirus may only infect a small number of individuals when it first jumps into a new rodent population where...
genetic drift effects may play a major role in the fixation of virus mutations in a new rodent host.

Although ASV, GOUV, SAAV, and SERV probably all originate from cross-species transmission, the number and type of aa differences between these viruses and their sister viruses varied greatly (Table 1). The age of the virus in the new host after cross-species transmission may be responsible for the variation, which may be caused by the accumulation of fixation of random mutation in viruses in the new hosts.

Usually, the accumulation of adaptive changes can facilitate the successful colonization in a new host species. HIV (60), influenza virus (13), and SARS corona virus (57) provide good examples. The closer the donor and recipient host species are in phylogenetic space, the fewer changes that are likely to be required for adaptation to the new host (19, 26). Different characteristics of hantaviruses have emerged as adaptations to the distinct genetic environment of their rodent hosts (43). Each hantavirus species or genetic lineages of the same species can possess specific aa “signatures” such as those in the N protein sequences for Puumala virus (52) or the GPC protein sequences for HTNV (73). Similar to previous studies (20, 62), we did not find convincing evidence for positive selection in Murinae-associated hantaviruses. However, each species of Murinae-associated hantavirus has its own specific signature aa markers (Tables 2, S3 and S4). More non-synonymous substitutions have been found in the GPC protein (Table 1), which is known to mediate cell attachment and fusion and to be the major element involved in induction of neutralizing antibodies during hantavirus infection (25). Further, presumably adaptive evolution was detected in the GPC protein especially when the closely related sister viruses were considered a group. Therefore, the adaptation to host species (co-divergence) may have occurred and facilitated the speciation and the further genetic diversity of hantaviruses as well.
Despite advances in understanding the patterns and processes of microevolution in RNA viruses, little is known about the determinants of viral diversification at the macro-evolutionary scale, particularly the processes by which viral lineages diversify into different “species” (26). Several studies show that within host populations both pathogen and host are able to adapt in response to the interactions resulting in co-evolution (18, 56, 58, 63). However, there is also a debate on how the micro-evolutionary changes can influence the patterns of speciation of the interacting species at the macro-evolutionary levels (26). For hantaviruses, there is an ongoing dispute concerning co-evolution/co-adaptation of viruses and their host (43, 46). The data presented here indicate cross-species transmission, the subsequent genetic drift effect, and adaptation in the new host population may substantially contribute to the speciation of the *Murinae*-associated hantaviruses.

ACKNOWLEDGEMENTS

This study was supported by the Chinese Ministry of Science and Technology (Grants 2002DIB40095 and 2003BA712A08-02), and by the State Key Laboratory for Infection Disease Prevention and Control (2011SKLID101).
REFERENCES


580
582
584
586
588
590
592
594
596
598


### TABLE 1. Cross-species transmission events presumably occurring in the evolution of Murinae-associated hantaviruses

<table>
<thead>
<tr>
<th>Species</th>
<th>Genetic distance</th>
<th>Sanger</th>
<th>Neo-species</th>
<th>Difference between pro-species and neo-species</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor Receptor</td>
<td></td>
<td>Pro-species</td>
<td>Neo-species</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td><em>Hylomyscus simus</em></td>
<td><em>Apodemus agrarius</em></td>
<td>0.192</td>
<td>SANGV</td>
<td>SAAV</td>
<td>22.1/11.4</td>
</tr>
<tr>
<td><em>A. agrarius</em></td>
<td><em>A. flavicollis</em></td>
<td>0.182</td>
<td>SAAV</td>
<td>DOBV</td>
<td>13.4/2.3</td>
</tr>
<tr>
<td><em>Bandicota indica</em></td>
<td><em>Rattus norvegicus</em></td>
<td>0.144</td>
<td>THAIV</td>
<td>SERV</td>
<td>16.4/3.0</td>
</tr>
<tr>
<td><em>B. indica</em></td>
<td><em>R. norvegicus</em></td>
<td>0.155</td>
<td>THAIV</td>
<td>SEOV</td>
<td>22.6/13.5</td>
</tr>
<tr>
<td><em>R. norvegicus</em></td>
<td><em>R. rattus</em></td>
<td>0.126</td>
<td>SEOV</td>
<td>GOUV</td>
<td>12.0/1.4</td>
</tr>
<tr>
<td><em>B. indica</em></td>
<td><em>N. confucianus</em></td>
<td>0.186</td>
<td>THAIV</td>
<td>DBSV</td>
<td>24.9/15.6</td>
</tr>
<tr>
<td><em>N. confucianus</em></td>
<td><em>A. agrarius</em></td>
<td>0.202</td>
<td>DBSV</td>
<td>HTNV</td>
<td>21.8/7.9</td>
</tr>
<tr>
<td><em>A. agrarius</em></td>
<td><em>A. peninsulae</em></td>
<td>0.169</td>
<td>HTNV</td>
<td>ASV</td>
<td>17.1/3.5</td>
</tr>
</tbody>
</table>

Note: Pro-species: The presumable parental/sister species of new species.  
Neo-species: The presumable descendant species.
TABLE 2. Comparison of amino acid signatures in the S and M segment amino acid sequence

<table>
<thead>
<tr>
<th>Species</th>
<th>Hosts</th>
<th>No. of signature aa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>SANGV</td>
<td>Hylomyscus simus</td>
<td>18</td>
</tr>
<tr>
<td>DOBV</td>
<td>Apodemus flavicollis</td>
<td>1</td>
</tr>
<tr>
<td>SAAV</td>
<td>A. agrarius</td>
<td>2</td>
</tr>
<tr>
<td>THAIV</td>
<td>Bandicota indica</td>
<td>1</td>
</tr>
<tr>
<td>SERV</td>
<td>R. rattus, R. tanezumi</td>
<td>1</td>
</tr>
<tr>
<td>SEOV</td>
<td>Rattus norvegicus</td>
<td>-</td>
</tr>
<tr>
<td>GOUV</td>
<td>R. rattus</td>
<td>-</td>
</tr>
<tr>
<td>DBSV</td>
<td>N. confucianus</td>
<td>3</td>
</tr>
<tr>
<td>HTNV</td>
<td>A. agrarius</td>
<td>2</td>
</tr>
<tr>
<td>ASV</td>
<td>A. peninsulae</td>
<td>1</td>
</tr>
<tr>
<td>SANGV+DOBV+SAAV</td>
<td>H. simus, Apodemus spp.</td>
<td>14</td>
</tr>
<tr>
<td>THAIV+SEOV+GOUV+DBSV+ASV+HTNV+SERV</td>
<td>B. indica, Rattus spp., Apodemus spp.</td>
<td>8</td>
</tr>
<tr>
<td>DOBV+SAAV</td>
<td>A. flavicollis, A. agrarius</td>
<td>16</td>
</tr>
<tr>
<td>SEOV+GOUV</td>
<td>R. norvegicus, R. rattus</td>
<td>25</td>
</tr>
<tr>
<td>SEOV+GOUV+SERV</td>
<td>R. norvegicus, R. rattus, R. tanezumi</td>
<td>-</td>
</tr>
<tr>
<td>THAIV+SEOV+GOUV+SERV</td>
<td>R. norvegicus, R. rattus</td>
<td>4</td>
</tr>
<tr>
<td>ASV+HTNV</td>
<td>A. agrarius, A. peninsulae</td>
<td>5</td>
</tr>
<tr>
<td>DBSV+ASV+HTNV</td>
<td>N. confucianus, Apodemus spp.</td>
<td>16</td>
</tr>
<tr>
<td>ASV+HTNV+DOBV+SAAV</td>
<td>Apodemus spp.</td>
<td>1</td>
</tr>
<tr>
<td>SANGV+ASV+HTNV+DOBV+SAAV</td>
<td>Mice (H. simus, Apodemus spp.)</td>
<td>1</td>
</tr>
<tr>
<td>DBSV+SEOV+GOUV+THAIV+SERV</td>
<td>Rats (N. confucianus, Rattus spp., B. indica)</td>
<td>2</td>
</tr>
<tr>
<td>THAIV+SERV</td>
<td>B. Indica, R. rattus, R. tanezumi</td>
<td>19</td>
</tr>
<tr>
<td>Summary statistic</td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>TMRCA (DBSV)</td>
<td>802</td>
<td>10.7</td>
</tr>
<tr>
<td>TMRCA (HTNV)</td>
<td>754</td>
<td>9.8</td>
</tr>
<tr>
<td>TMRCA (ASV)</td>
<td>536</td>
<td>6.7</td>
</tr>
<tr>
<td>TMRCA (SEOV)</td>
<td>133</td>
<td>1.4</td>
</tr>
<tr>
<td>TMRCA (GOUV)</td>
<td>206</td>
<td>2.3</td>
</tr>
<tr>
<td>TMRCA (THAIV)</td>
<td>107</td>
<td>1.3</td>
</tr>
<tr>
<td>TMRCA (SERV)</td>
<td>152</td>
<td>1.8</td>
</tr>
<tr>
<td>TMRCA (DOBV)</td>
<td>149</td>
<td>1.8</td>
</tr>
<tr>
<td>TMRCA (SAAV)</td>
<td>318</td>
<td>3.9</td>
</tr>
<tr>
<td>TMRCA (SANGV)</td>
<td>63</td>
<td>0.8</td>
</tr>
<tr>
<td>TMRCA (HTNV-ASV)</td>
<td>983</td>
<td>12.7</td>
</tr>
<tr>
<td>TMRCA (HTNV-ASV-DBSV)</td>
<td>1590</td>
<td>20.4</td>
</tr>
<tr>
<td>TMRCA (DOBV-SAAV)</td>
<td>581</td>
<td>7.2</td>
</tr>
<tr>
<td>TMRCA (DOBV-SAAV-SANGV)</td>
<td>1614</td>
<td>21.6</td>
</tr>
<tr>
<td>TMRCA (SEOV-GOUV)</td>
<td>515</td>
<td>3.6</td>
</tr>
<tr>
<td>TMRCA (THAIV-SERV)</td>
<td>851</td>
<td>10.2</td>
</tr>
<tr>
<td>TMRCA (SEOV-GOUV-THAIV-SERV)</td>
<td>1896</td>
<td>24.8</td>
</tr>
<tr>
<td>Mean rate</td>
<td>2.0E-4</td>
<td>2.7E-6</td>
</tr>
</tbody>
</table>

Note: GM, geometric mean
TABLE 4. Detection of positive selected sites by maximum likelihood (ML) estimation for 48 S segment and 40 M segment sequences of hantavirus.

<table>
<thead>
<tr>
<th>Models</th>
<th>Positive Selected Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S Segment</td>
</tr>
<tr>
<td>Branch Model (M0 vs. FR)</td>
<td>Not allowed</td>
</tr>
<tr>
<td>Site Model (M7 vs. M8)</td>
<td>None</td>
</tr>
<tr>
<td>Branch-site Model A MA' vs. MA</td>
<td>ASV group</td>
</tr>
<tr>
<td></td>
<td>DBSV group</td>
</tr>
<tr>
<td></td>
<td>HTNV group</td>
</tr>
<tr>
<td></td>
<td>SEOV group</td>
</tr>
<tr>
<td></td>
<td>GOUV group</td>
</tr>
<tr>
<td></td>
<td>DOBV group</td>
</tr>
<tr>
<td></td>
<td>SAAV group</td>
</tr>
<tr>
<td></td>
<td>DOBV-SAAV group</td>
</tr>
<tr>
<td></td>
<td>DOBV-SAAV-SANGV group</td>
</tr>
<tr>
<td></td>
<td>HTNV-ASV group</td>
</tr>
<tr>
<td></td>
<td>HTNV-ASV-DBSV group</td>
</tr>
<tr>
<td></td>
<td>SEOV-GOUV group</td>
</tr>
<tr>
<td></td>
<td>THAIV-SERV group</td>
</tr>
<tr>
<td></td>
<td>THAIV-SERV-SEOV-GOUV group</td>
</tr>
</tbody>
</table>

Note: * Positively selected sites are identified with posterior probability p > 95%; **, p > 99%. The positive selection sites are also the “signature” aa markers and are shown in boldface.
Figure Legends

**FIG 1.** A map of China illustrating the location of DBSV variants identified in *Niviventer* rats. The location of trap sites where DBSV was detected in this study is shown as a triangle (▲).

**FIG 2.** Phylogenetic trees based on the entire coding regions of the genome sequences of *Murinae*-associated viruses including DBSV variants obtained in this study. The Bayesian/ML trees were based on the coding sequences of the S (A) and M (B) segments and the partial L (C) segment sequences. Numbers (>0.7/>70%) above or below the branches indicate posterior node probabilities or bootstrap values. A 0.7 posterior node probability over or a 70% bootstrap value was considered a node supported value. All trees were rooted with Thottapalayam virus (TPMV). Scale bar represents number of nucleotide substitutions per site. Rodent hosts abbreviations are: Aa, *Apodemus agrarius*; Ap, *Apodemus peninsulae*; Af, *Apodemus flavicollis*; Nc, *Niviventer confucianus*; Rn, *Rattus norvegicus*; Rr, *Rattus rattus*; Rt, *Rattus tanezumi*; Bi, *Bandicota indica*; and Hs, *Hylomyscus simus*.

**FIG 3.** Phylogenetic relationships between *Niviventer* rats captured in Wenzhou and other rodents found in the GenBank. The Bayesian tree/ML trees were constructed using the cytochrome b gene sequences. The sequences of *Spalax ehrebergi* were used as the out-group. The sequences obtained in this study are shown in bold. Posterior node probabilities/bootstrap values (>0.7/>70%) are shown above or below the branches. The scale bar represents the number of nucleotide substitutions per site.

**FIG 4.** Tanglegram constructed with TreeMap2.0b program illustrating the phylogenies of *Murinae*-associate hantaviruses and their rodent carriers. The host tree on the left was based on cytochrome b gene sequences, and the hantavirus tree on the right was based on the coding sequences of the S segment. The MrBayes v3.1.2 program package was used to construct the phylogenetic trees by the Bayesian method. Numbers (>0.7) above or below branches indicate posterior node
probabilities. The reconciliation analysis showed a significant congruence between
phylogenies of *Murinae*-associated hantaviruses and their hosts (*P*<0.05).

**FIG 5.** The rooted phylogenetic trees with a molecular clock were reconstructed
using the Bayesian MCMC method in BEAST based on the S segment sequences (A)
and the M segment sequences (B). The divergence times, 95% high-probability
density and Bayesian posterior probabilities are given at the nodes leading to each
major hantavirus-specific group.