Epidemics 5 (2013) 1–10

Contents lists available at SciVerse ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Evolutionary and molecular analysis of the emergent severe fever with thrombocytopenia syndrome virus

Tommy Tsan-Yuk Lam[a,1], Wei Liu[b,1], Thomas A. Bowden[c,1], Ning Cui[d], Lu Zhuang[b], Kun Liu[b], Yao-Yun Zhang[b], Wu-Chun Cao[b], Oliver G. Pybus[a,*]

[a] Department of Zoology, University of Oxford, Oxford, United Kingdom
[b] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China
[c] Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
[d] The 154 Hospital, People Liberation Army, Xinyang 464000, Henan, PR China

ARTICLE INFO

Article history:
Received 4 July 2012
Accepted 13 September 2012
Available online 26 September 2012

Keywords:
Phlebovirus
Severe fever with thrombocytopenia syndrome
Tick-borne disease
Phylogenetic
Evolution
Nucleoprotein structure

ABSTRACT

In 2009, a novel *Bunyavirus*, called severe fever with thrombocytopenia syndrome virus (SFTSV) was identified in the vicinity of Huaiyangshan, China. Clinical symptoms of this zoonotic virus included severe fever, thrombocytopenia, and leukocytopenia, with a mortality rate of ~10%. By the end of 2011 the disease associated with this pathogen had been reported from eleven Chinese provinces and human-to-human transmission suspected. However, current understanding of the evolution and molecular epidemiology of SFTSV before and after its identification is limited. To address this we undertake phylogenetic, evolutionary and structural analyses of all available SFTSV genetic sequences, including a new SFTSV complete genome isolated from a patient from Henan in 2011. Our discovery of a mosaic L segment sequence, which is descended from two major circulating lineages of SFTSV in China, represents the first evidence that homologous recombination plays a role in SFTSV evolution. Selection analyses indicate that negative selection is predominant in SFTSV genes, yet differences in selective forces among genes are consistent between Phlebovirus species. Further analysis reveals structural conservation between SFTSV and Rift Valley fever virus in the residues of their nucleocapsids that are responsible for oligomerisation and RNA-binding, suggesting the viruses share similar modes of higher-order assembly. We reconstruct the epidemic history of SFTSV using molecular clock and coalescent-based methods, revealing that the extant SFTSV lineages originated 50–150 years ago, and that the viral population experienced a recent growth phase that concurs with and extends the earliest serological reports of SFTSV infection. Taken together, our combined structural and phylogenetic analyses shed light into the evolutionary behaviour of SFTSV in the context of other, better-known, pathogenic *Phleboviruses*.

Introduction

Reports of severe fever with thrombocytopenia syndrome (SFTS) emerged between March and July, 2009, in rural areas of central China. The syndrome was characterized by fever, gastrointestinal symptoms, thrombocytopenia and leukocytopenia in individuals bitten by ticks (Yu et al., 2011). Differential diagnosis initially included human anaplasmosis, leptospirosis and hemorrhagic fever with renal syndrome; however, the causative pathogen could not be ascertained. A few months later, a novel *Phlebovirus* (family *Bunyaviridae*) was reported as the cause of SFTS, through molecular characterization and epidemiologic investigations of the virus isolated from patients' blood (Yu et al., 2011). From June 2009 to September 2010, SFTSV RNA or specific antibodies were detected in the serum of 171 of 241 hospitalized patients who met the case definition for SFTS (in Henan, Hubei, Shandong, Liaoning, Anhui and Jiangsu provinces) (Yu et al., 2011). Two names are used in the literature to refer to the virus: severe fever with thrombocytopenia syndrome virus (Yu et al., 2011; SFTSV; used hereafter) or Huaiyangshan virus (Zhang et al., 2012b). Like many other members of the *Phlebovirus* genus, SFTSV uses ticks as a putative host vector and is likely to be transmitted to human populations from an animal reservoir via blood feeding (Elliott, 2008). Previous studies have detected SFTSV in *Haemaphysalis longicornis* and *Rhipicephalus microplus* ticks collected from a number of domestic animals, including cattle, buffalo, goats, cats and dogs (Zhang et al., 2011, 2012b). Although transmission of SFTSV is currently poorly understood, viral genome sequences obtained from infected ticks on animals are highly similar to those obtained from human isolates, suggesting zoonotic transmission (Zhang et al., 2012b).

* Corresponding author at: Department of Zoology, South Parks Road, Oxford OX1 3PS, United Kingdom. Tel.: +44 01865 271274; fax: +44 01865 310447.
E-mail address: oliver.pybus@zoo.ox.ac.uk (O.G. Pybus).
[1] These authors contributed equally.

1755-4365
http://dx.doi.org/10.1016/j.epidem.2012.09.002

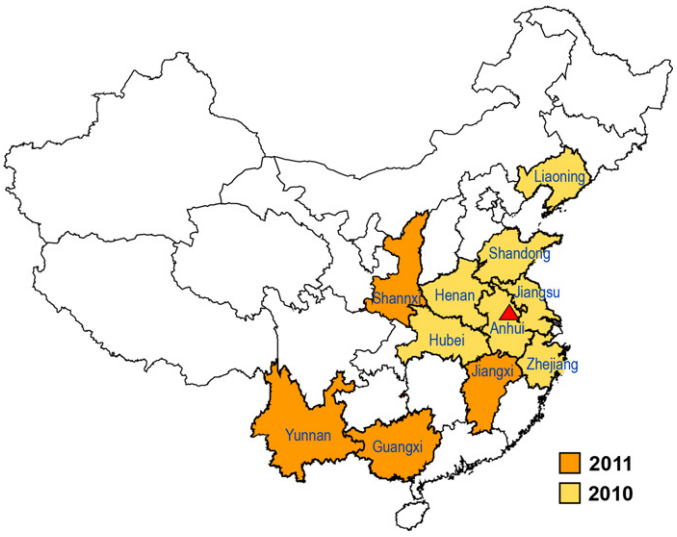

Fig. 1. Map showing the provinces in mainland China that have officially reported SFTSV in 2010–11. The red triangle indicates the first SFTS case (in Dingyuan County, Chuzhou City, Anhui, in September, 2006) that was retrospectively confirmed by laboratory testing (Liu et al., 2012). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

However, a number of human infection clusters have also been identified, suggesting the possibility of human-to-human transmission (Bao et al., 2011; Liu et al., 2012; Xu et al., 2011).

Starting from October 2010, active surveillance for human SFTSV infections has been undertaken by the Chinese Centre for Disease Control and Prevention (CCDC) (Ministry of Health, China, 2010). By the end of 2011, SFTS had been reported in eleven provinces including Henan, Hubei, Anhui, Shandong, Jiangsu, Zhejiang, Liaoning, Yunnan, Guangxi, Jiangxi and Shannxi (Fig. 1). The mortality of SFTSV infections is between 8% and 16%, according to different studies (Table 1).

The genome of SFTSV, like other *Phleboviruses*, comprises three segments: L (large), M (medium), and S (small). Nucleocapsid (N) and non-structural (NS) proteins are encoded in the two non-overlapping ORFs of the S segment, whereas the L and M segments encode the RNA-dependent RNA polymerase (RdRp) and Gn-Gc envelope glycoproteins (GP), respectively (Yu et al., 2011). Whole-genome sequencing and analysis confirmed that SFTSV is a novel *Bunyavirus* and is most closely related to Uukuniemi virus (pairwise nucleotide similarities for the L, M and S segments are 34%, 24% and 29%, respectively; Zhang et al., 2012b). All SFTSV sequences currently available in GenBank are similar regardless of their sampling locations (>90% sequence similarity; Zhang et al., 2012b), suggesting a relatively recent common ancestor.

The broad geographic distribution throughout China, relatively high case fatality rate, and potential for human-to-human transmission render SFTSV a notable threat to public health, both regionally and internationally. Despite this, very little is known about the molecular epidemiology and structure of this zoonotic pathogen. In order to address this paucity of information, we have undertaken comprehensive phylogenetic, evolutionary and structural analyses of SFTSV genetic sequences, and report the complete genome sequence of a new SFTSV isolate, obtained in 2011 from a patient from Henan province. We estimate SFTSV's date of origin, rate of evolution and epidemic history, and test for the presence of viral recombination and reassortment. Through comparative analyses of the genome sequences of SFTSV and related *Phleboviruses* (such as Rift Valley fever virus) we infer structural properties of the SFTSV nucleoprotein and glycoproteins, and explore the effects of natural selection on SFTSV protein evolution.

Table 1
Laboratory-confirmed and fatal cases of SFTSV infection in China.

Study	Laboratory-confirmed cases	Fatal cases	Mortality rate (%)
Yu et al. (2011)	171	21	12.28
Liu et al. (2012)	14	2	14.29
Zhang et al. (2012a)	49	8	16.33
Liu and Liu (2011)	17	2	11.76
Zhang et al. (2011)	33	5	15.15
Yu et al. (2012)	61	5	8.2

Materials and methods

Sample collection, virus isolation and sequencing

Patients who met criteria for probable SFTS cases according to the national guidelines (Ministry of Health, China, 2010) were recruited at an army hospital in the Xinyang administrative district of Henan Province. One patient was confirmed to be infected with SFTSV in June 2010. Briefly, sera samples from this patient were obtained three days following the onset of illness. Serum was inoculated on DH82 cells which were cultured at 37 °C in a 5% carbon dioxide atmosphere with twice-weekly media changes. One isolate, generated by primer-walking strategies as previously described (Zhang et al., 2011) (named WCH-2011/HN/China/isolate97), was used to derive the whole genome sequence. The resulting negative-sense RNA genome was 11,491 nt in length, and has been submitted to GenBank under accession numbers JQ341188, JQ341189 and JQ341190 (L, M and S segments, respectively).

Sequence data collection and alignment

We analysed the nucleotide sequences of all genome segments of all available SFTSV isolates (sampled during 2009–2011) available from GenBank, together with our newly generated genome of WCH-2011/HN/China/isolate97. Each genome segment was aligned using MUSCLE v3.6 (Edgar, 2004) and the ORFs corresponding to the viral RdRp (sited in segment L), GP (segment M), N and NS (segment S) genes were extracted for further analysis. Details of the sequences used in this study are provided in Supplementary Table S1.

Phylogenetic analyses

Phylogenetic trees were reconstructed from the SFTSV alignments using the maximum likelihood (ML) approach implemented in PhyML v3 (Guindon and Gascuel, 2003). We used a General-time-reversible (GTR) nucleotide substitution model and modelled rate heterogeneity among sites using a discrete gamma (Γ) distribution with 4 rate categories. To assess the robustness of the tree topology, we generated 1000 ML bootstrap replicates. Statistical support for phylogenetic clusters was assessed by estimating node posterior probabilities using the Bayesian Markov Chain Monte Carlo (BMCMC) method implemented in BEAST v1.7.1 (Drummond et al., 2012) (further details below). The phylogenetic positions of individual mutations were determined by ancestral sequence reconstruction, using the joint ML method (Pupko et al., 2000) with the GTR+ Γ nucleotide substitution model implemented in HYPHY v1.0b (Pond et al., 2005).

Detecting mosaic sequences

We first screened SFTSV alignments for homologous mosaic structures using the various exploratory methods implemented in Recombination Detection Program v3.29 (Martin et al., 2010), including RDP (Martin and Rybicki, 2000), CHIMAERA (Posada

and Crandall, 2001), GENECONV (Padidam et al., 1999), MAXCHI (Smith, 1992) and 3SEQ (Boni et al., 2007). Potential mosaic structures were defined as those with Bonferroni-corrected *p*-values <0.05 in more than two detection methods mentioned above. Putative mosaic sequences were subsequently analysed using the similarity plot and bootscanning methods implemented in SIMPLOT v3.5 (Lole et al., 1999). The breakpoint positions of putative recombination events were refined using GARD v1.0 (Kosakovsky Pond et al., 2006). For further confirmation, phylogenies of putative mosaic sequences, their potential parents, and outgroup lineages were reconstructed from genome regions either side of the estimated breakpoints. Finally, the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) was used to evaluate the significance of phylogenetic discordance between non-recombinant regions. Confirmed mosaic sequences were excluded from subsequent evolutionary analyses.

Estimating evolutionary timescale and epidemic history

Evolutionary parameters including rates of sequence evolution, tree topologies, epidemic histories, divergence times and substitution model parameters were jointly estimated from each dataset using the BMCMC approach (Drummond et al., 2002) implemented in BEAST v1.7.1 (Drummond et al., 2012). A GTR+ Γ nucleotide substitution model was employed, and a Bayesian skyline plot (BSP) coalescent model was used to estimate effective population size through time (Drummond et al., 2005). Due to the short time period (2009–2011) over which SFTSV isolates had been sampled, we were unable to estimate evolutionary rates directly from the SFTSV sequences themselves. Consequently, we estimated *Phlebovirus*-specific evolutionary rates from the most closely related virus for which sufficient temporal data was available. This was Rift Valley fever virus (RVFV), for which isolates from mammals and mosquitoes have been sampled from 1944 to 2000 and for which complete genomes have been sequenced (Bird et al., 2007). We thus estimated evolutionary rates for each RVFV gene from a heterochronous data set of 33 RVFV whole genomes (available on request) using the dated-tip relaxed molecular clock model implemented in BEAST. Rate variation among branches was modelled using the uncorrelated lognormal distribution (UCLD) model. The resulting evolutionary rate estimates were then encoded as strong prior distributions for the SFTSV evolutionary rate parameters during the subsequent BMCMC analysis of SFTSV. MCMC computation was run for a total 4×10^8 generations, sampled every 1000 generations on average. Multiple MCMC runs were performed to check for consistency and MCMC convergence was evaluated using Tracer v.1.5.1 (Rambaut and Drummond, 2007).

Lineage-specific nucleoprotein gene rate variation

To assess whether evolutionary rates vary among SFTSV lineages, we used the UCLD (Drummond et al., 2006) and 'random local clock' (Drummond and Suchard, 2010) models of rate variation implemented in BEAST. As noted above, the UCLD model accommodates rate variation by assuming the branch-specific rates are lognormally distributed. In contrast, the 'random local clock model' allows rate changes of any magnitude on a phylogeny branch; the new rate is then applied to all branches belonging to the subtree that subtends from that branch. The relative substitution rates of the SFTSV lineages A–E were subsequently calculated from output of these two models.

Selection analyses

We investigated the selection pressures acting on the SFTSV genome, and on those of five other *Phleboviruses*: Rift Valley fever virus (RVFV), Candiru virus (CDUV), Punta Toro phlebovirus (PTTV), sandfly fever Naples virus (SFNV) and sandfly fever Sicilian virus (SFSV). Positively selected sites in the viral genomes were identified using a d_N/d_S approach, specifically, the two-rate fixed effect likelihood (FEL) method (Kosakovsky Pond and Frost, 2005) implemented in HYPHY (Pond et al., 2005). The GTRxMG94 substitution model was used. The M7/M8 codon substitution models implemented in PAML were also used to confirm the positively selected sites identified by the FEL method. As an indicator of the average selective pressure across each gene, an overall d_N/d_S was also estimated for each entire ORF.

Structural analyses of SFTSV nucleoprotein and Gn–Gc glycoprotein

The ancestral nucleotide sequences of SFTSV lineages A, B, C and E were estimated using the ML method described above, then translated into amino acid sequences for structure-based analysis. As the crystal structure of the SFTSV N protein is not yet determined, we used the crystal structure of the related RVFV N protein (PDB accession number 3OV9) as a template to develop structure-guided inferences. The amino acid similarity between the N proteins of RVFV and SFTSV is ~40%. A structure-based sequence alignment of the RVFV N protein (RVFV-N) was created with ClustalW (Larkin et al., 2007) and rendered with ESPRIPT (Gouet et al., 1999) as previously described (Bowden et al., 2009). Residues predicted to be involved in RNA-binding or homo-oligomerisation of SFTSV (calculated with the EBI PISA server; Krissinel and Henrick, 2007) were mapped onto the surface of the hexameric structure of RVFV-N with the program PyMOL (www.pymol.org).

As there are no atomic structures of related *Phlebovirus* Gn or Gc glycoproteins available for a structure-based sequence alignment, the secondary structure of the SFTSV glycoprotein was predicted using the PSI secondary structure prediction server (McGuffin et al., 2000) and mapped onto a sequence alignment of the SFTSV lineages (Figure S2). Five N-linked glycosylation motifs were identified (NXS/TX, where X is any amino acid except proline) and transmembrane regions were predicted using the TMHMM Server (Krogh et al., 2001).

Results

Phylogenetic diversity of SFTSV

The phylogenies estimated from full-length sequences of SFTSV genes demonstrated four major lineages, previously named A, B, C and E (Zhang et al., 2012b; Fig. S1A–D). Phylogenies estimated from partial RdRp and N gene sequences (which include more isolates than the full-length alignments) also revealed a further lineage, named D (no M segment sequences of lineage D are available in GenBank; Fig. S1E, F). These five lineages were supported by high bootstrap support values (>0.7) and posterior clade probabilities (>0.9). All isolates from animals (dogs, cats, goats buffalo and cattle) identified by a previous study (Zhang et al., 2012b) belonged to lineage A. Our new isolate WCH-2011/HN/China/isolate97 was classified in lineage E.

The branching orders of these five lineages are generally consistent among the ML phylogenies estimated from different gene segments (Fig. S1). A small incongruence is observed in the phylogeny of the partial RdRp alignment (Fig. 2A, S1E), in which lineages A, D and E are grouped together, but the ML bootstrap support and posterior clade probabilities for this grouping are low (bootstrap = 13% and probability = 0.10 in Fig. 2A; bootstrap = 50% in Fig. S1E). Therefore we conclude there is no phylogenetic evidence for ancestral reassortment among SFTSV lineages A–E.

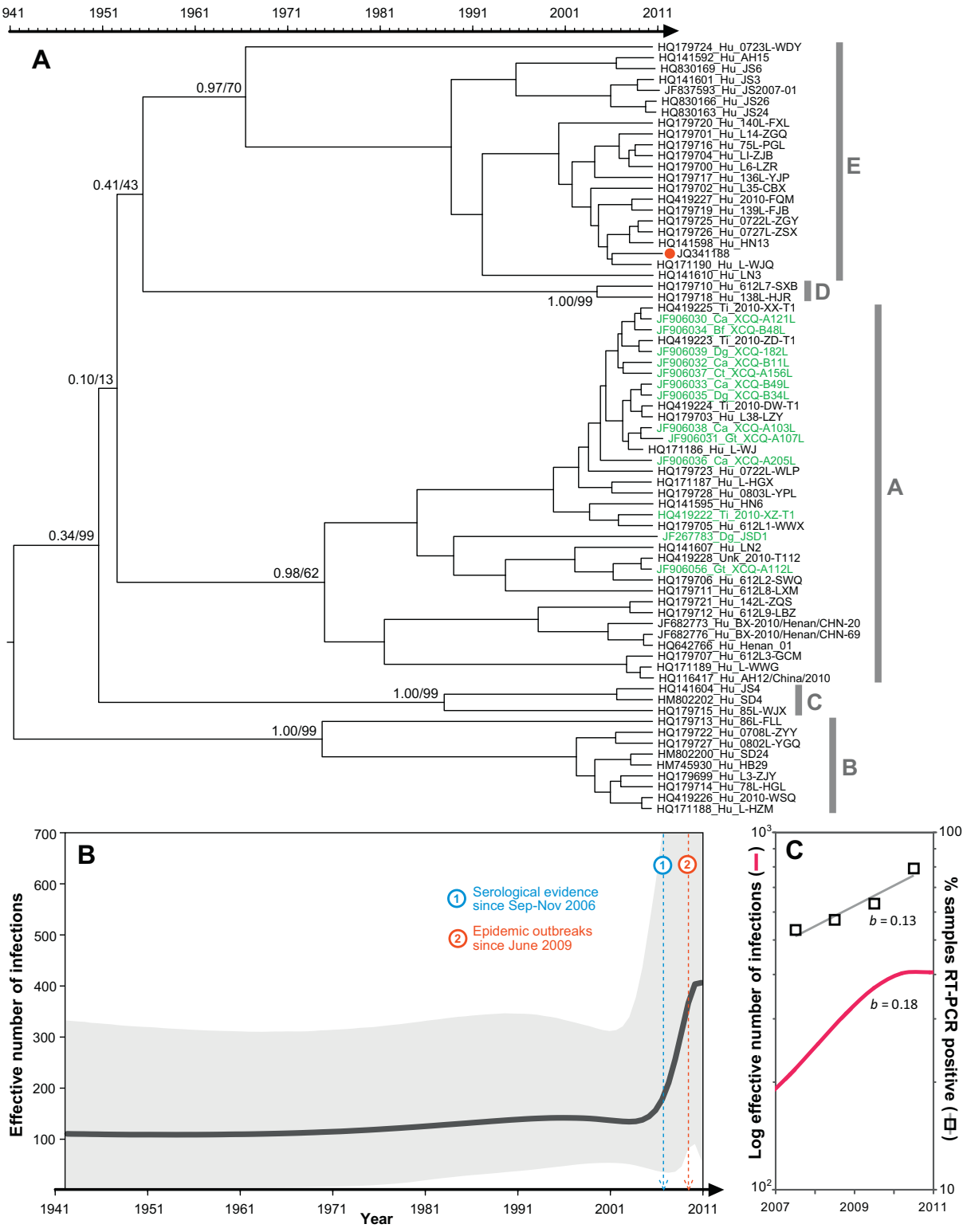

Fig. 2. Bayesian phylogenetic analyses of the RdRp gene of SFTSV. Partial gene sequences of RdRp (2194–3105 nt; $n = 71$) were analysed. This includes 1, 66 and 4 sequences isolated from 2009, 2010 and 2011, respectively. (A) Maximum clade credibility tree summarising the Bayesian phylogenetic analysis. Branch lengths represent time in years. Posterior probabilities and bootstrap support values are indicated next to nodes. Animal isolates are in green. Hu, Gt, Ti, Ct, Dg, Ca, Bf and Unk denote SFTSV isolates from human, goat, tick, cat, dog, cattle, buffalo and unknown species, respectively. The red circle (JQ341188) marks the human isolate 'WCH-2011/HN/China/isolate97' generated in this study. (B) Bayesian skyline plot, estimated from the same alignment. Notable epidemiological events are indicated by dashed lines. The thick grey line represents the estimated effective number of infections and the shaded areas show the 95% HPDs this estimate. (C) Comparison of the skyline plot estimate between 2007 and 2010 with the RT-PCR positive rate (%) among the SFTS cases reported in Henan during the same time. Exponential growth rates are denoted as '*b*'. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

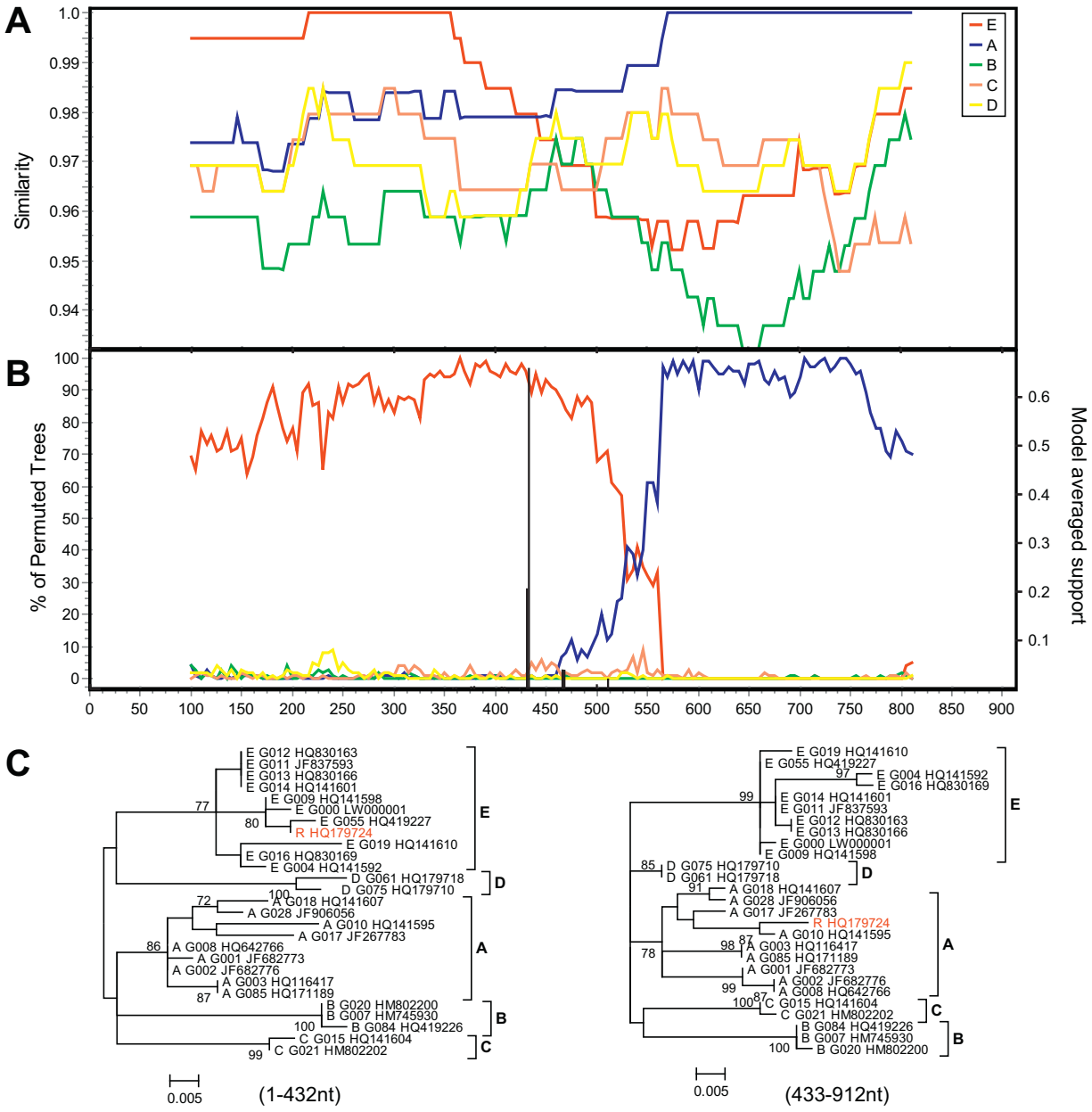

Fig. 3. Putative mosaic structure of the RdRp gene of isolate 0723L-WDY. The partial gene sequence of RdRp (2194–3105 nt; HQ179724) of 0723L-WDY was analysed. (A) Similarity plot, (B) Bootscanning analysis (with GARD estimates of the breakpoint position shown in black) and (C) phylogenetic analysis of the regions on either side of the estimated breakpoint (position 432 nt). Bootstrapping analysis with 1000 pseudo-replicates was performed. Window and step sizes of 150 nt and 5 nt were used in the similarity plot and bootscanning analyses.

Mosaic genome structure in SFTSV

The suite of methods implemented in RDP3 identified putative mosaic structure in the partial RdRp gene sequence of isolate 0723L-WDY (accession number HQ179724). The mosaic structure of this sequence was evident in similarity plot and bootscanning analyses (Fig. 3A and B), both of which indicate that the 3′ and 5′ ends of the viral RNA sequence were most closely related to lineages E and A, respectively. Further analysis using GARD estimated that the mosaic breakpoint is most likely located at nucleotide position 432 of sequence HQ179724. Phylogenetic analyses of the gene regions before and after the breakpoint confirmed the results of the previous methods: isolate 0723L-WDY clustered with lineages E and A, respectively (Fig. 3C). The discordant phylogenies estimated from different regions of 0723L-WDY were significant ($P < 0.05$), according to the Shimodaira–Hasegawa test.

Estimating evolutionary timescale and epidemic history

Phlebovirus-specific evolutionary rates were estimated for the RdRp, GP, N and NS genes, from a heterochronous data set of RVFV isolates using the UCLD relaxed molecular clock model implemented in BEAST. The estimated rates were 2.57 (1.46–3.74) $\times 10^{-4}$ substitutions/site/year for RdRp, 2.99 (1.87–4.07) for GP, 3.67 (1.83–5.73) for N and 3.09 (1.52–4.87) for NS. These estimates were subsequently used to estimate the evolutionary timescale of SFTSV (see Methods). The estimated date of the most recent common ancestor (tMRCA) of SFTSV was consistent among different genes, ranging from year 1864 to 1934 (Fig. 4A), although the credible regions for some of these estimates were large (e.g. 1751–1961 for the NS gene). Collectively, these data suggest that the existing SFTSV lineages A–E originated from a common ancestor that most likely existed 50–150 years ago.

To investigate the epidemic history of SFTSV since its origin, we applied the Bayesian skyline coalescent method to the partial RdRp alignment (71 sequences, 912 nt in length). The resulting plot (Fig. 2B) indicates that the virus maintained a stable effective population size for much of its evolutionary history. However, effective population size starts to rise rapidly after 2005–2006 and the rise continues to 2011 (the date of the most recently sampled isolate).

Lineage-specific nucleoprotein gene rate variation

Visual inspection of the N gene phylogeny suggests that branches of lineage C are shorter than those of other lineages, yet lineage C isolates were not sampled significantly earlier (Fig. S1D, F). This raises the possibility that the evolutionary rate of lineage C is lower than that of other lineages, therefore we analysed SFTSV rate variation in order to explore this hypothesis. Specifically, we estimated the relative evolutionary rates of each SFTSV lineage using the UCLD and 'random local clock' (RLC) models implemented in BEAST (see Methods). The relative rates estimated using both methods were congruent (Fig. 4B). The estimated relative rate of lineage C (UCLD = 0.79; RLC = 0.70) was lower than those of other lineages (which ranged from 0.99 to 1.30); however this difference was not statistically significant, as the credible intervals of the estimates were large and overlapping (Fig. 4B).

Selection analysis

Site-specific selection analyses revealed no sites in the SFTSV genome with significant evidence for positive selection. In addition to the FEL method implemented in HYPHY, this result was confirmed using the M7/M8 models implemented in PAML. For RFVF and SFNV, estimates of the average d_N/d_S ratio of the NS gene were significantly higher than those of other genes. The d_N/d_S ratios for each gene of various different *Phlebovirus* species (Fig. 5) also showed that the GP and NS genes typically have higher d_N/d_S ratios than the RdRp and N genes, which may reflect variation in the strength of selective constraint amongst genome regions.

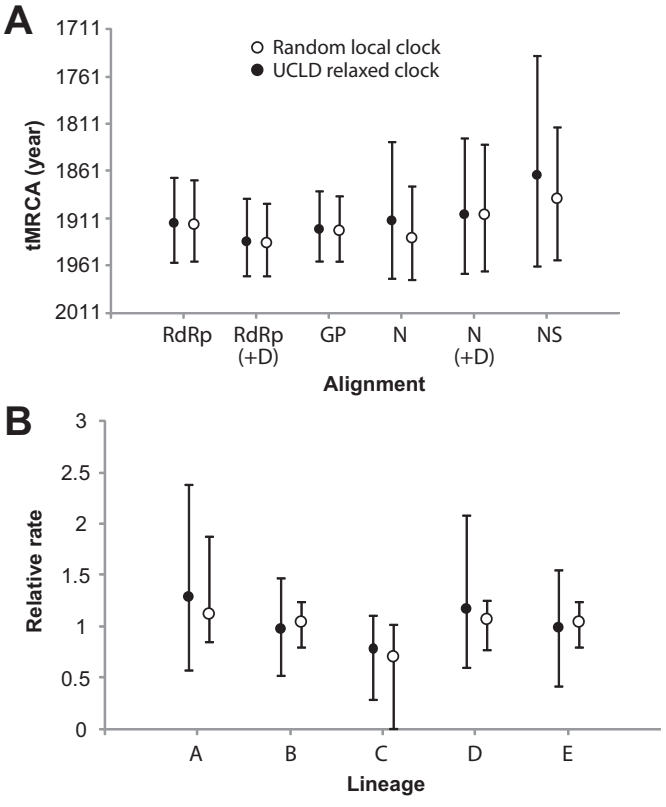

Fig. 4. Bayesian molecular clock analyses. (A) Time of the most recent common ancestor (tMRCA) of different SFTSV gene alignments. The data sets with '+D' in parentheses indicate the inclusion of sequences of lineage D, and hence are partial gene alignments. The 95% HPD of these estimates are indicated by error bars. (B) Relative rates of N gene evolution for different SFTSV lineages. Solid and empty circles represent the estimates from UCLD and RLC molecular clock models respectively.

Structural analyses of SFTSV nucleoprotein and Gn–Gc glycoprotein

In cooperation with the L protein, the *Phlebovirus* nucleoprotein packages genomic RNA to form a winding, irregular structure in the virion (Huiskonen et al., 2009; Raymond et al., 2010). X-ray crystallographic analyses of RVFV-N have revealed a unique structure consisting of thirteen α-helices that assembles into a range of oligomeric forms (Ferron et al., 2011; Raymond et al., 2010) primarily through the interaction of an N-terminal α-helix with the protein surface of an adjacent protomer (Fig. 6A and B; Ferron et al., 2011). Given the limited sequence divergence between the SFTSV and RVFV nucleoproteins (~35%) it is probable that these two proteins share the same fold (Fig. 6C; Chothia and Lesk, 1986).

Our results (Fig. 6) suggest that SFTSV-N also shares similar functionality to RVFV-N. First, we observe a high degree of sequence conservation between RVFV and SFTSV in the interface between adjoining protomers, indicating that the overall mode of oligomerisation is likely to be conserved (Fig. 6A). Of the 75 residues involved in these protein–protein contacts, 32 (43%) are conserved and 43 (57%) are physiochemically similar between RVFV-N and SFTSV-N. Second, we propose that RVFV-N has similar RNA-binding properties to that of SFTSV-N. In their functional analysis, Ferron et al. (2011) identified positively charged residues (R64, K67, and K74) located in the solvent-exposed centre of the RVFV-N hexamer, which are implicated in RNA-binding. These residues are completely conserved between RVFV and SFTSV, indicating that the two viruses likely share a common mode of RNA packaging.

Inferences into the structure of the *Phlebovirus* glycoproteins are far less certain. Although low-resolution electron microscopy structures of icosohedral RVFV have revealed that glycoprotein Gn–Gc complexes assemble into pentameric and hexameric spikes (Huiskonen et al., 2009; Sherman et al., 2009), there is to date a lack of atomic-level information on the structure and function of the putative *Phlebovirus* attachment (Gn) and fusion (Gc) subcomponents. In the absence of such data it is postulated that the Gc glycoproteins of *Phleboviruses* exhibit a class II fusion glycoprotein fold (Vaney and Rey, 2011). This suggestion is based upon a number of properties that are shared with *Alphaviruses* and *Flaviviruses* (Vaney and Rey, 2011) including a predominance of predicted β-stranded secondary structure throughout the Gc. Our analysis suggests that this predominance is also present in the Gc of SFTSV (Fig. S2).

Discussion

The recent emergence of SFTSV and its continuing circulation and geographic spread pose an incessant threat to health in China. Many aspects of the virus' evolutionary biology are poorly understood, such as its date of origin and the possible effects of natural selection, recombination, and random genetic drift on the evolutionary genetics of the virus. Using a combination of approaches we here provide evidence that SFTSV may undergo homologous recombination. Further, our coalescent-based inference suggested an expanding effective population size of the virus over the last decade in the host populations studied here, including the years immediately prior to its discovery and identification.

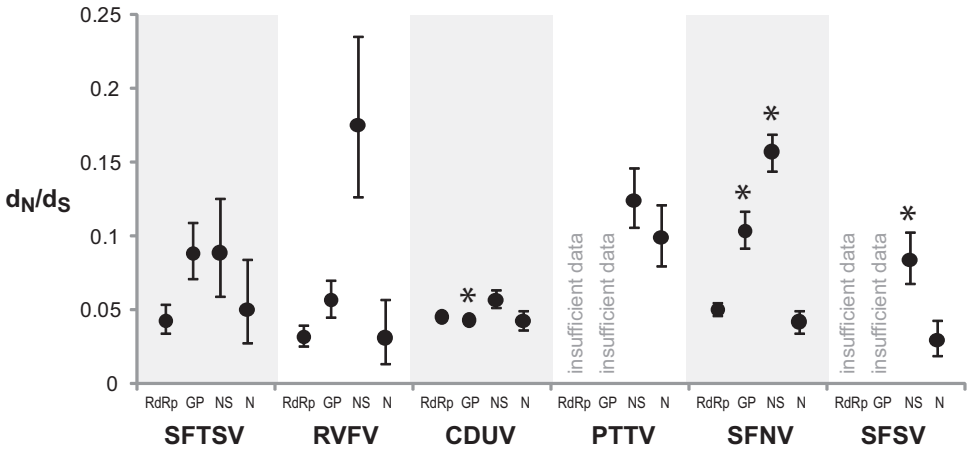

Fig. 5. Estimated d_N/d_S ratios for four genes of six different *Phlebovirus* species. Error bars indicate the 95% confidence intervals of the estimated values. No values are reported for the RdRp and GP genes of PTTV and SFSV because less than three sequences were available. Genes with positively selected sites identified by the FEL approach (at the 0.05 significance level) are marked with asterisks. SFTSV, RVFV, CDUV, PTTV, SFNV and SFSV denote severe fever with thrombocytopenia syndrome virus, Rift Valley fever virus, Candiru virus, Punta Toro phlebovirus, sandfly fever Naples virus and sandfly fever Sicilian virus, respectively.

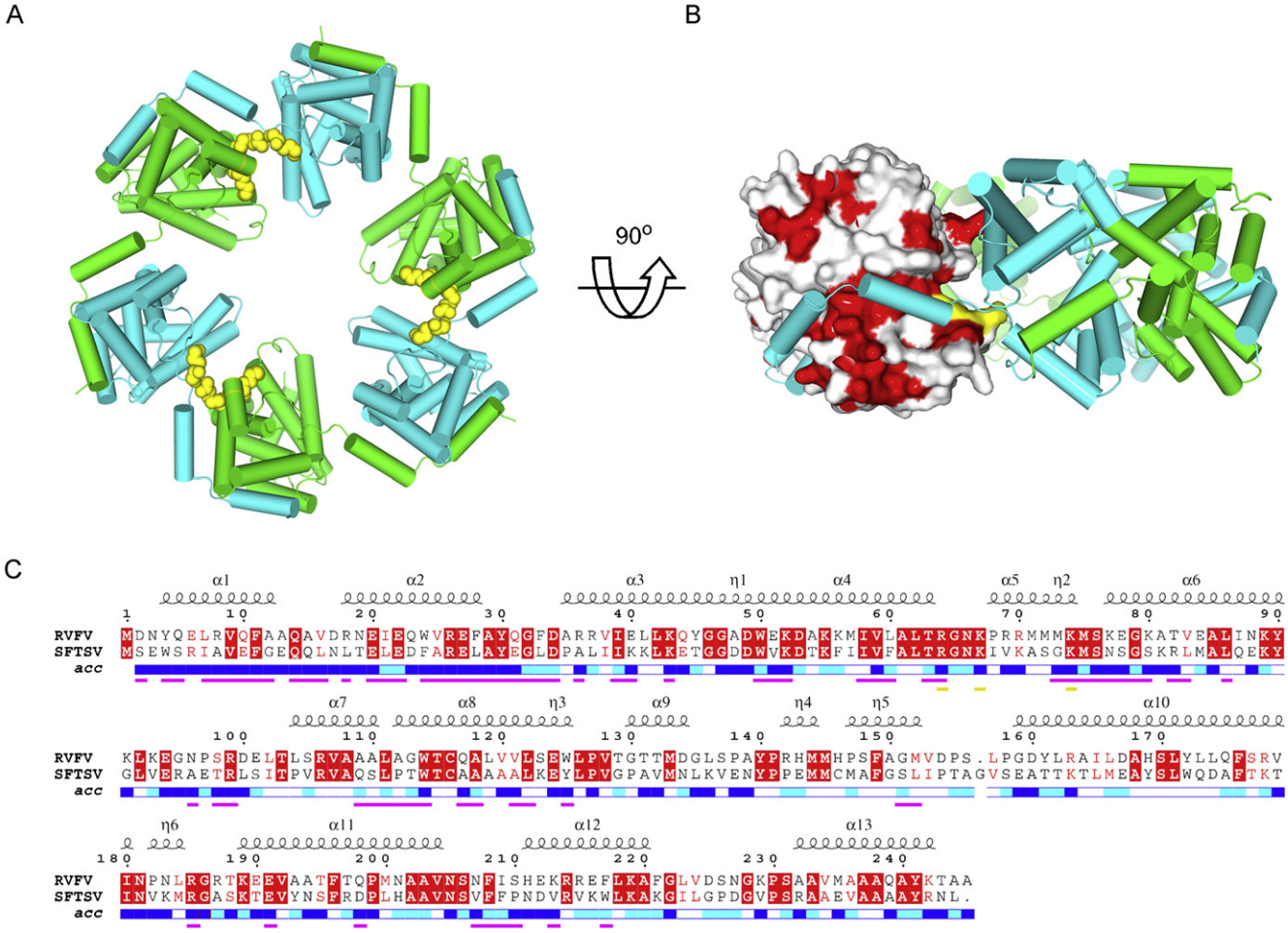

Fig. 6. Structural analyses of the SFTSV nucleocapsid (N) protein. (A) and (B) display the residues conserved between SFTSV and RVFV, mapped onto the surface of the RVFV N protein. (A) Cartoon representation of the RVFV hexamer (PDB accession number 3OV9) with alternating protomers of the hexamer coloured cyan and green. Residues R64, K67, and K74 have been putatively identified as responsible for RNA-binding (Ferron et al., 2011) and are shown as yellow spheres. (B) Close up view of the homotypic interaction between protomers, which reveals a high degree of sequence conservation between RVFV and SFTSV in the protein–protein interface. The surface of one protomer is shown as a surface model and coloured according to residue conservation with SFTSV: red = conserved; white = no sequence conservation. Residues R64, K67, and K74 are coloured yellow. (C) Structure-based sequence alignment of the SFTSV nucleoprotein. Residues highlighted in red (in white text) are fully conserved, residues in red text are partially conserved, and residues in black text are not conserved. Residues that are solvent accessible (as determined by ESPRIPT; Gouet et al., 1999) are highlighted immediately below the sequence; light blue = partially accessible; dark blue = fully accessible. Secondary structure elements, generated from the nucleoprotein structure of RVFV (PDB accession number 3OV9) are shown above the sequence: spiral = α-helix. Pink lines highlight the residues involved in homotypic interactions. Yellow lines highlight positively charged residues in the RVFV nucleoprotein that have been implicated in RNA-binding (Ferron et al., 2011). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

These findings help to provide a better understanding of the evolutionary behaviour of this pathogen.

Recombination is important in shaping the evolutionary genetics of several DNA and RNA viruses (Froissart et al., 2005; Worobey and Holmes, 1999) most notably the human immunodeficiency virus (Kijak and McCutchan, 2005). Homologous recombination is distinct from the process of reassortment, which is widespread among viruses with segmented genomes, including the *Bunyaviruses* and *Orthomyxoviruses*. By exchanging entire genome segments reassortment may confer dramatic changes in viral phenotypes, such as the 'antigenic shift' exhibited by influenza A virus (e.g. Olsen, 2002; Parrish et al., 2008). Reassortants have been suggested to be associated with new outbreaks (Bowen et al., 2001; Gerrard et al., 2004), increased pathogenicity and enhanced transmissibility among vectors and hosts (Beaty and Bishop, 1988; Rodriguez et al., 1998; Shope et al., 1981), and some of these traits have been confirmed experimentally (Bridgen et al., 2001; Perrone et al., 2007; Saluzzo and Smith, 1990; Vialat et al., 2000). Natural reassortment has been reported for various members of the *Phlebovirus* genus, including Candiru virus (Palacios et al., 2011) and RVFV (Sall et al., 1999; Turell et al., 1990). However, in our analysis of all 24 complete SFTSV genomes we found no statistically significant evidence of reassortment between and within different SFTSV lineages.

Although we detected no reassortment in SFTSV, we identified a possible homologous recombination event in the RdRp gene of the 0723L-WDY isolate, which appears to be a mosaic of lineage E and A sequences. Negative-sense single-stranded RNA viruses are thought to recombine comparatively infrequently (Chare et al., 2003), possibly because template-switching during replication is inhibited by the fact that viral RNA is packaged into ribonucleoproteins. Although there have been reports of homologous recombination in *Hantaviruses* (Klempa et al., 2005; Sironen et al., 2001; Zhang et al., 2010; Zuo et al., 2011) and Crimean-Congo hemorrhagic fever virus (Deyde et al., 2006; Lukashev, 2005), only the M and S segments of three *Hantaviruses* (strains KY, HuBJ16 and HuBJ17) (Zhang et al., 2010; Zuo et al., 2011) have phylogenetic support for recombination that is as strong as that presented for SFTSV here. The fitness effects of the identified recombination event are unknown; however, we note that the breakpoint (2626 nt) is located between the functional RdRp domain and the catalytic domain, and the consequences of exchanging these two domains may deserve further investigation.

The inference of natural recombination events from mosaic viral sequences should be undertaken with caution, because laboratory artefacts and contamination may also create similar sequence patterns. According to the literature, the partial RdRp gene sequence of isolate 0723L-WDY originated from a single amplicon amplified by a pair of primers (Zhang et al., 2011). Therefore, it is unlikely that the mosaic was generated by incorrect contig assembly from different specimens. Nonetheless, the possibility that the recombination is artificial cannot be formally excluded unless the virus is re-isolated and sequenced from the original sample, or isolates with the same mosaic structure and breakpoint are found in other samples.

We estimated rates of evolution for the N genes of each major lineage of SFTSV. Although estimated rates tended for be lower for lineage C (Fig. 4B) this difference was not statistically significant, and the d_N/d_S ratio of lineage C was not significantly different to that of the other lineages. Clearly more sequence data is needed to decisively test for the presence or absence of SFTSV rate variation among lineages A–E.

Although the novel emergence of SFTSV and its association with SFTS was first confirmed during the summer of 2009 (Yu et al., 2011), similar symptoms have been reported by hospitals near Huaiyangshan since 2007 (Xu et al., 2011). Liu et al. (2012) report a retrospective study of 13 patient blood samples obtained from two groups of individuals that exhibited SFTS-like symptoms in the spring of 2006. This study confirmed SFTSV infections using indirect immunofluorescence assays and RT-PCR. Similar studies have demonstrated SFTSV sero-positivity in eastern China from 2007 to 2011 (Kang et al., 2012; Xu et al., 2011) and some of these infections may have involved human-to-human transmission (Bao et al., 2011; Liu et al., 2012).

Our reconstruction of the epidemic history of STFSV using coalescent-based methods agrees with and extends these epidemiological reports. The Bayesian skyline plot (Fig. 2B) suggests that the effective number of infections was largely constant for several decades before starting to grow exponentially around the mid-2000s. This transition concurs with the first serological evidence of SFTSV infection in 2006. Furthermore, increases in STFSV detections from 2007 match the apparent increase in viral effective population size. Specifically, between 2007 and 2010 the proportion of samples from suspected SFTS cases in Henan province that tested positive for SFTSV by RT-PCR (Kang et al., 2012) grew exponentially with an estimated exponential growth rate of 0.13 $year^{-1}$ (Fig. 2C). During the same time period, the average exponential growth rate of the Bayesian skyline plot is 0.18 $year^{-1}$ (Fig. 2C). Note, we did not include data from 2011 in this analysis because there are insufficient STFSV sequences from 2011 for comparison.

Although coalescent-based methods for reconstructing epidemic history have proven useful in many instances, and often agree with information from surveillance epidemiology (Biek et al., 2007; Carrington et al., 2005; Lam et al., 2012; Medina et al., 2009), they make a number of strong assumptions and should be interpreted carefully (Frost and Volz, 2010; Pybus and Rambaut, 2009). Here we advise caution in interpreting the Bayesian skyline plot, because its 95% credible limits are large and because isolates from human and non-human samples were combined. More reliable estimates of SFTSV epidemic history should be possible from larger sets of sequence data, which will allow the identification of epidemiologically defined monophyletic groups.

Our molecular clock analysis indicates that the common ancestor of contemporary SFTSV lineages existed 50–150 years ago. We can only speculate on reasons why the virus has not emerged until very recently. First, current knowledge suggests that the susceptible population is older people (Kang et al., 2012). In recent years an increasing proportion of younger adults have moved from rural to urban areas for employment, and consequently more local older individuals have taken up farming work and thus their exposure to high tick densities might have increased. Alternatively, environmental or other changes could have led to a recent expansion of the tick population. Lastly, it is theoretically possible that recent viral mutations have enhanced the virulence and transmissibility of SFTSV in humans.

Our analysis of selective pressures revealed a relatively high d_N/d_S ratio in the NS gene of RVFV, compared to other *Phleboviruses* (Fig. 5). The NS protein is thought to affect virulence, primarily through its antagonism of interferon, helping the virus to evade host antiviral responses (Pepin et al., 2010; Perrone et al., 2007). A higher d_N/d_S ratio in the NS gene may reflect a greater degree of positive selection resulting from the co-evolutionary battle between the virus and host immunity, but could equally well be explained by a reduction in selective constraint on that gene. Other possible explanations for a higher d_N/d_S are recombination, or overlapping reading frames (e.g. PB1-F2 in influenza virus). However, we detected no significant recombination signal in the NS gene and the NS and N ORFs in segment S are not overlapping. We found no codons in SFTSV with significant evidence of positive selection that might reveal clues to its recent emergence. Similar methods applied to other viruses, such as canine parvovirus have provided such evidence (Parrish et al., 2008), although it is common for d_N/d_S

ratios in vector-borne RNA viruses to be low in comparison to non vector-borne diseases (Woelk and Holmes, 2002). Selection analyses may provide greater insights as larger sets of SFTSV genomes become available.

Phlebovirus glycoproteins often exhibit the second highest d_N/d_S ratios (after the NS gene; Fig. 5). Theoretically, this might reflect the high antigenicity of solvent-accessible residues, which could be under selective pressure to evade the host humoral immune response as well as to adapt to their respective cell-surface receptors for attachment and entry (Bowden et al., 2011). However, we found no evidence that individual codons in the SFTSV glycoprotein were under positive selection, although such support was found for the 256th and 806th codons of the CDUV and SFNV glycoproteins, respectively. *Phlebovirus* nucleoproteins (N), on the other hand, are responsible for RNA-binding and -packaging and are thus likely to be more functionally constrained; this is supported by our finding that the estimated d_N/d_S values of *Phleboviruses* N genes are as low as those of the RdRp genes. There is no difference in the N gene amino acid sequences of the ancestors of the major SFTSV lineages, and across the whole SFTSV N gene phylogeny there are few non-synonymous mutations (6 out of 44) on internal branches, again suggesting strong selective constraint on the N gene during SFTSV evolution. Recent elucidation of the first crystal structures of RVFV-N (Raymond et al., 2010) has allowed us to assess whether this predicted constraint can be confirmed in structure. We found that those residues in N proteins that are predicted to be responsible for oligomerisation and RNA-binding are strictly conserved between SFTSV and RVFV, supporting similar modes of oligomerisation and viral genome packaging in these two viruses.

Ethical approval

The research protocol was approved by the human ethics committee of the 154 Hospital in Xinyang, Henan, and all participants provided informed consent.

Conflict of interest statement

No conflicts declared.

Authorship

(i) All the authors have agreed to its submission and are responsible for its contents.
(ii) All the authors have agreed that the corresponding authors may act on their behalf regarding any subsequent processing of the paper.

Acknowledgments

We gratefully thank scientists who collected and sequenced the genomes of SFTSV, and submitted the sequences into GenBank database. TTL is supported by Newton International Fellowship. This study is supported by the China Mega-Project on Infectious Disease Prevention (Nos. 2013ZX10004202-002, 2011ZX10004-001) and National Science Fund for Young Scholars (81222037) to WL. TAB is a Sir Henry Wellcome Postdoctoral Fellow (Grant Number 089026/Z/09/Z). OGP is supported by The Royal Society.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.epidem.2012.09.002.

References

Bao, C.J., Guo, X.L., Qi, X., Hu, J.L., Zhou, M.H., Varma, J.K., Cui, L.B., Yang, H.T., Jiao, Y.J., Klena, J.D., Li, L.X., Tao, W.Y., Li, X., Chen, Y., Zhu, Z., Xu, K., Shen, A.H., Wu, T., Peng, H.Y., Li, Z.F., Shan, J., Shi, Z.Y., Wang, H., 2011. A family cluster of infections by a newly recognized bunyavirus in eastern China, 2007: further evidence of person-to-person transmission. Clinical Infectious Diseases 53, 1208–1214.
Beaty, B.J., Bishop, D.H., 1988. Bunyavirus-vector interactions. Virus Research 10, 289–301.
Biek, R., Henderson, J.C., Waller, L.A., Rupprecht, C.E., Real, L.A., 2007. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. Proceedings of the National Academy of Sciences of the United States of America 104, 7993–7998.
Bird, B.H., Khristova, M.L., Rollin, P.E., Ksiazek, T.G., Nichol, S.T., 2007. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. Journal of Virology 81, 2805–2816.
Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.
Bowden, T.A., Crispin, M., Graham, S.C., Harvey, D.J., Grimes, J.M., Jones, E.Y., Stuart, D.I., 2009. Unusual molecular architecture of the machupo virus attachment glycoprotein. Journal of Virology 83, 8259–8265.
Bowden, T.A., Jones, E.Y., Stuart, D.I., 2011. Cells under siege: viral glycoprotein interactions at the cell surface. Journal of Structural Biology 175, 120–126.
Bowen, M.D., Trappier, S.G., Sanchez, A.J., Meyer, R.F., Goldsmith, C.S., Zaki, S.R., Dunster, L.M., Peters, C.J., Ksiazek, T.G., Nichol, S.T., 2001. A reassortant bunyavirus isolated from acute hemorrhagic fever cases in Kenya and Somalia. Virology 291, 185–190.
Bridgen, A., Weber, F., Fazakerley, J.K., Elliott, R.M., 2001. Bunyamwera bunyavirus nonstructural protein NSs is a nonessential gene product that contributes to viral pathogenesis. Proceedings of the National Academy of Sciences of the United States of America 98, 664–669.
Carrington, C.V., Foster, J.E., Pybus, O.G., Bennett, S.N., Holmes, E.C., 2005. Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. Journal of Virology 79, 14680–14687.
Chare, E.R., Gould, E.A., Holmes, E.C., 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. Journal of General Virology 84, 2691–2703.
Chothia, C., Lesk, A.M., 1986. The relation between the divergence of sequence and structure in proteins. EMBO Journal 5, 823–826.
Deyde, V.M., Khristova, M.L., Rollin, P.E., Ksiazek, T.G., Nichol, S.T., 2006. Crimean-Congo hemorrhagic fever virus genomics and global diversity. Journal of Virology 80, 8834–8842.
Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4, e88.
Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307–1320.
Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular Biology and Evolution 22, 1185–1192.
Drummond, A.J., Suchard, M.A., 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biology 8, 114.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969–1973.
Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.
Elliott, R.M., 2008. Bunyaviruses: General Features, third ed. Elsevier Academic Press, Oxford.
Ferron, F., Li, Z., Danek, E.I., Luo, D., Wong, Y., Coutard, B., Lantez, V., Charrel, R., Canard, B., Walz, T., Lescar, J., 2011. The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. PLoS Pathogens 7, e1002030.
Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., Michalakis, Y., 2005. Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. PLoS Biology 3, e89.
Frost, S.D., Volz, E.M., 2010. Viral phylodynamics and the search for an 'effective number of infections'. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 365, 1879–1890.
Gerrard, S.R., Li, L., Barrett, A.D., Nichol, S.T., 2004. Ngari virus is a Bunyamwera virus reassortant that can be associated with large outbreaks of hemorrhagic fever in Africa. Journal of Virology 78, 8922–8926.
Gouet, P., Courcelle, E., Stuart, D.I., Metoz, F., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305–308.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52, 696–704.
Huiskonen, J.T., Overby, A.K., Weber, F., Grunewald, K., 2009. Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: evidence for GN–GC glycoprotein heterodimers. Journal of Virology 83, 3762–3769.
Kang, K., Tang, X.Y., Xu, B.L., You, A.G., Huang, X.Y., Du, Y.H., Wang, H.F., Zhao, G.H., Chen, H.M., Liu, G.H., Meng, F.J., 2012. Analysis of the epidemic characteristics of fever and thrombocytopenia syndrome in Henan province, 2007–2011. Zhonghua Yu Fang Yi Xue Za Zhi 46, 106–109.
Kijak, G.H., McCutchan, F.E., 2005. HIV diversity, molecular epidemiology, and the role of recombination. Current Infectious Disease Reports 7, 480–488.

Klempa, B., Stanko, M., Labuda, M., Ulrich, R., Meisel, H., Kruger, D.H., 2005. Central European Dobrava Hantavirus isolate from a striped field mouse (*Apodemus agrarius*). Journal of Clinical Microbiology 43, 2756–2763.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution 22, 1208–1222.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22, 3096–3098.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology 372, 774–797.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305, 567–580.

Lam, T.T., Hon, C.C., Lemey, P., Pybus, O.G., Shi, M., Tun, H.M., Li, J., Jiang, J., Holmes, E.C., Leung, F.C., 2012. Phylodynamics of H5N1 avian influenza virus in Indonesia. Molecular Ecology 21, 3062–3077.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Liu, Q., Liu, J., 2011. Fever with thrombocytopenia syndrome virus infection in 17 clinical observation cases. Chinese Journal of Internal Medicine 50, 785–786.

Liu, Y., Li, Q., Hu, W., Wu, J., Wang, Y., Mei, L., Walker, D.H., Ren, J., Yu, X.J., 2012. Person-to-person transmission of severe fever with thrombocytopenia syndrome virus. Vector Borne and Zoonotic Diseases 12, 156–160.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. Journal of Virology 73, 152–160.

Lukashev, A.N., 2005. Evidence for recombination in Crimean-Congo hemorrhagic fever virus. Journal of General Virology 86, 2333–2338.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405.

Medina, R.A., Torres-Perez, F., Galeno, H., Navarrete, M., Vial, P.A., Palma, R.E., Ferres, M., Cook, J.A., Hjelle, B., 2009. Ecology, genetic diversity, and phylogeographic structure of andes virus in humans and rodents in Chile. Journal of Virology 83, 2446–2459.

Ministry of Health, China, 2010. Announcement of the guidelines for preventing and controlling the severe fever with thrombocytopenia syndrome disease. Available at: http://www.moh.gov.cn/publicfiles/business/htmlfiles/wsb/pwsyw/201009/49245.htm

Olsen, C.W., 2002. The emergence of novel swine influenza viruses in North America. Virus Research 85, 199–210.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Palacios, G., Tesh, R., Travassos da Rosa, A., Savji, N., Sze, W., Jain, K., Serge, R., Guzman, H., Guevara, C., Nunes, M.R., Nunes-Neto, J.P., Kochel, T., Hutchison, S., Vasconcelos, P.F., Lipkin, W.I., 2011. Characterization of the Candiru antigenic complex (Bunyaviridae: Phlebovirus), a highly diverse and reassorting group of viruses affecting humans in tropical America. Journal of Virology 85, 3811–3820.

Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.C., Burke, D.S., Calisher, C.H., Laughlin, C.A., Saif, L.J., Daszak, P., 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews 72, 457–470.

Pepin, M., Bouloy, M., Bird, B.H., Kemp, A., Paweska, J., 2010. Rift Valley fever virus(Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. Veterinary Research 41, 61.

Perrone, L.A., Narayanan, K., Worthy, M., Peters, C.J., 2007. The S segment of Punta Toro virus (Bunyaviridae, Phlebovirus) is a major determinant of lethality in the Syrian hamster and codes for a type I interferon antagonist. Journal of Virology 81, 884–892.

Pond, S.L., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences of the United States of America 98, 13757–13762.

Pupko, T., Pe'er, I., Shamir, R., Graur, D., 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Molecular Biology and Evolution 17, 890–896.

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nature Reviews Genetics 10, 540–550.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer

Raymond, D.D., Piper, M.E., Gerrard, S.R., Smith, J.L., 2010. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proceedings of the National Academy of Sciences of the United States of America 107, 11769–11774.

Rodriguez, L.L., Owens, J.H., Peters, C.J., Nichol, S.T., 1998. Genetic reassortment among viruses causing hantavirus pulmonary syndrome. Virology 242, 99–106.

Sall, A.A., Zanotto, P.M., Sene, O.K., Zeller, H.G., Digoutte, J.P., Thiongane, Y., Bouloy, M., 1999. Genetic reassortment of Rift Valley fever virus in nature. Journal of Virology 73, 8196–8200.

Saluzzo, J.F., Smith, J.F., 1990. Use of reassortant viruses to map attenuating and temperature-sensitive mutations of the Rift Valley fever virus MP-12 vaccine. Vaccine 8, 369–375.

Sherman, M.B., Freiberg, A.N., Holbrook, M.R., Watowich, S.J., 2009. Single-particle cryo-electron microscopy of Rift Valley fever virus. Virology 387, 11–15.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16, 1114–1116.

Shope, R.E., Rozhon, E.J., Bishop, D.H., 1981. Role of the middle-sized bunyavirus RNA segment in mouse virulence. Virology 114, 273–276.

Sironen, T., Vaheri, A., Plyusnin, A., 2001. Molecular evolution of Puumala hantavirus. Journal of Virology 75, 11803–11810.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. Journal of Molecular Evolution 34, 126–129.

Turell, M.J., Saluzzo, J.F., Tammariello, R.F., Smith, J.F., 1990. Generation and transmission of Rift Valley fever viral reassortants by the mosquito Culex pipiens. Journal of General Virology 71 (Pt 10), 2307–2312.

Vaney, M.C., Rey, F.A., 2011. Class II enveloped viruses. Cellular Microbiology 13, 1451–1459.

Vialat, P., Billecocq, A., Kohl, A., Bouloy, M., 2000. The S segment of rift valley fever phlebovirus (Bunyaviridae) carries determinants for attenuation and virulence in mice. Journal of Virology 74, 1538–1543.

Woelk, C.H., Holmes, E.C., 2002. Reduced positive selection in vector-borne RNA viruses. Molecular Biology and Evolution 19, 2333–2336.

Worobey, M., Holmes, E.C., 1999. Evolutionary aspects of recombination in RNA viruses. Journal of General Virology 80 (Pt 10), 2535–2543.

Xu, B., Liu, L., Huang, X., Ma, H., Zhang, Y., Du, Y., Wang, P., Tang, X., Wang, H., Kang, K., Zhang, S., Zhao, G., Wu, W., Yang, Y., Chen, H., Mu, F., Chen, W., 2011. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. PLoS Pathogens 7, e1002369.

Yu, B., Zhang, Y., Peng, J.-S., Zhou, G.-Q., Luo, T.-Y., Hu, Q., He, Y.-W., Wang, W., Tian, J.-H., Kong, D.-G., Quan, Y.-X., Dai, Y.-A., Wang, Q.-F., Liu, X.-X., Li, W., Li, W., 2012. Epidemiological study on data involving 61 hospitalized cases with Huaiyangshan hemorrhagic fever in Wuhan, Hubei province. Chinese Journal of Epidemiology 33 (1).

Yu, X.J., Liang, M.F., Zhang, S.Y., Liu, Y., Li, J.D., Sun, Y.L., Zhang, L., Zhang, Q.F., Popov, V.L., Li, C., Qu, J., Li, Q., Zhang, Y.P., Hai, R., Wu, W., Wang, Q., Zhan, F.X., Wang, X.J., Kan, B., Wang, S.W., Wan, K.L., Jing, H.Q., Lu, J.X., Yin, W.W., Zhou, H., Guan, X.H., Liu, J.F., Bi, Z.Q., Liu, G.H., Ren, J., Wang, H., Zhao, Z., Song, J.D., He, J.R., Wan, T., Zhang, J.S., Fu, X.P., Sun, L.N., Dong, X.P., Feng, Z.J., Yang, W.Z., Hong, T., Zhang, Y., Walker, D.H., Wang, Y., Li, D.X., 2011. Fever with thrombocytopenia associated with a novel bunyavirus in China. New England Journal of Medicine 364, 1523–1532.

Zhang, Y., Zhang, H., Dong, X., Yuan, J., Yang, X., Zhou, P., Ge, X., Li, Y., Wang, L.F., Shi, Z., 2010. Hantavirus outbreak associated with laboratory rats in Yunnan, China. Infection, Genetics and Evolution 10, 638–644.

Zhang, Y.Z., He, Y.W., Dai, Y.A., Xiong, Y., Zheng, H., Zhou, D.J., Li, J., Sun, Q., Luo, X.L., Cheng, Y.L., Qin, X.C., Tian, J.H., Chen, X.P., Yu, B., Jin, D., Guo, W.P., Li, W., Wang, W., Peng, J.S., Zhang, G.B., Zhang, S., Chen, X.M., Wang, Y., Li, M.H., Li, Z., Lu, S., Ye, C., de Jong, M.D., Xu, J., 2012a. Hemorrhagic fever caused by a novel Bunyavirus in China: pathogenesis and correlates of fatal outcome. Clinical Infectious Diseases 54, 527–533.

Zhang, Y.Z., Zhou, D.J., Qin, X.C., Tian, J.H., Xiong, Y., Wang, J.B., Chen, X.P., Gao, D.Y., He, Y.W., Jin, D., Sun, Q., Guo, W.P., Wang, W., Yu, B., Li, J., Dai, Y.A., Li, W., Peng, J.S., Zhang, G.B., Zhang, S., Chen, X.M., Wang, Y., Li, M.H., Lu, X., Ye, C., de Jong, M.D., Xu, J., 2012b. The ecology, genetic diversity, and phylogeny of Huaiyangshan virus in China. Journal of Virology 86, 2864–2868.

Zhang, Y.Z., Zhou, D.J., Xiong, Y., Chen, X.P., He, Y.W., Sun, Q., Yu, B., Li, J., Dai, Y.A., Tian, J.H., Qin, X.C., Jin, D., Cui, Z., Luo, X.L., Li, W., Lu, S., Wang, W., Peng, J.S., Guo, W.P., Li, M.H., Li, Z.J., Zhang, S., Chen, C., Wang, Y., de Jong, M.D., Xu, J., 2011. Hemorrhagic fever caused by a novel tick-borne Bunyavirus in Huaiyangshan, China. Zhonghua Liu Xing Bing Xue Za Zhi 32, 209–220.

Zuo, S.Q., Fang, L.Q., Zhan, L., Zhang, P.H., Jiang, J.F., Wang, L.P., Ma, J.Q., Wang, B.C., Wang, R.M., Wu, X.M., Yang, H., Cao, Z.W., Cao, W.C., 2011. Geo-spatial hotspots of hemorrhagic fever with renal syndrome and genetic characterization of Seoul variants in Beijing, China. PLoS Neglected Tropical Diseases 5, e945.