

First complete genomic characterization of two tick-borne encephalitis virus isolates obtained from wild rodents in South Korea

Seok-Min Yun · Su Yeon Kim · Young Ran Ju ·
Myung Guk Han · Young Eui Jeong ·
Jungsang Ryou

Received: 3 October 2010 / Accepted: 14 January 2011 / Published online: 1 February 2011
© Springer Science+Business Media, LLC 2011

Abstract We determined for the first time the complete genome sequences of two Korean strains of the tick-borne encephalitis virus (TBEV), designated KrM 93 and KrM 213, isolated from the lung tissues of wild rodents in 2006. The genomes are 11,097 nucleotides (nt) in length and consist of a 132 nt 5'-noncoding region (NCR), a 10,245 nt open reading frame (ORF) containing 10 viral protein-coding regions (3,415 amino acids), and a 720 nt 3'-NCR. Compared with the 31 fully sequenced TBEV strains currently available, KrM 93 and KrM 213 show genomic nucleotide (and deduced amino acid) sequence divergences ranging from 1.8 (0.7) to 19.2 (26.6)% and 1.9 (0.8) to 19.3 (26.7)%, respectively. Phylogenetic and recombination analyses based on the complete genome sequence were performed to identify genetic variations and relationships between the TBEV strains. These showed that the Korean TBEV strains clustered with the Western subtype rather than with Far-Eastern or Siberian subtypes, and phylogenetic trees derived from capsid (C), envelope (E), non-structural (NS) 4B and NS5 regions represented the same branching pattern shown by the complete genome-based tree. Although no recombination events were identified in these two Korean strains, 11 putative recombination events

were identified within the NS5 regions or in the 3'-NCRs of TBEV strains in general. The results provide insight into the genetics of TBEV strains to understand the molecular epidemiology, genetic diversity, and evolution of TBEV.

Keywords Complete genome sequences · Tick-borne encephalitis virus Korean strains · Phylogenetic analysis · Recombination analysis · Western subtype

Introduction

Tick-borne encephalitis (TBE) is the most important flavivirus infection of the human central nervous system (CNS) in Europe, Russia, and Far-Eastern Asia, including Japan and China. More than 10,000 cases of the disease are reported in endemic regions annually [1, 2]. As the causative agent of TBE, Tick-borne encephalitis virus (TBEV), belonging to the family *Flaviviridae* and genus *Flavivirus*, is an enveloped virus containing a single positive RNA genome ~11-kb long and consisting of a single open reading frame (ORF), flanked by 5'- and 3'-noncoding regions (NCRs). The ORF is cotranslationally and posttranslationally cleaved by viral and cellular proteases to form the capsid (C), premembrane/membrane (prM/M) and envelope (E) proteins, and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) [3].

Until now, TBEV has been subdivided into three subtypes: Far-Eastern, Western or European, and Siberian [2, 4]. The main tick vector of the Western subtype is *Ixodes ricinus*, and *Ixodes persulcatus* is the main vector for the other subtypes. The three main subtypes are all closely related both antigenically and genetically, but the Far-Eastern subtype causes more severe encephalitis symptoms and produces more acute infections than do the Western and

Electronic supplementary material The online version of this article (doi:10.1007/s11262-011-0575-y) contains supplementary material, which is available to authorized users.

S.-M. Yun · S. Y. Kim · Y. R. Ju · M. G. Han ·
Y. E. Jeong · J. Ryou (✉)
Division of Arboviruses, National Institute of Health, Korea
Centers for Disease Control and Prevention, Osong Health
Technology Administration Complex (OHTAC),
187, Osongsaengmyeong2(i)-ro, Gangoe-myeon,
Cheongwon-gun, Chungcheongbuk-do,
South Korea
e-mail: zenith@nih.go.kr

Siberian subtypes, respectively [5, 6]. Generally, the Western subtype of TBEV is endemic in Central Europe and the European region of Russia. On the other hand, the Far-Eastern subtype of TBEV is prevalent in Russia and Far-Eastern Asia (China and Japan). The Siberian subtype of TBEV also exists in eastern Russia and in western Siberia. However, exceptional forms of TBEV distribution have been identified, such as the cocirculation of three subtypes in the Baltic countries and the existence of the Siberian TBEV subtype in Finland, which do not fit into the previously known distribution pattern of TBEV subtypes [7–9].

In South Korea, although TBE infections have not been reported among humans, we have identified molecular evidence of TBEV infections in infesting ticks of the wild animals or collected ticks, and have isolated TBEV from lung tissues of the wild rodent, *Apodemus agrarius* [10, 11]. In addition, using phylogenetic analysis based on the complete E gene we confirmed that the Korean strains belong to the Western subtype, and have the Western subtype signature amino acid in the E protein region [12]. The five known Korean strains have a high degree of sequence identity in the E gene. Here, we chose two Korean strains from different geographical origins, KrM 93 and KrM 213, to determine the complete genomic sequence.

In this study, we report the first complete genomic sequence identification of two TBEV strains that were isolated from wild rodents in South Korea in 2006. We aimed to sequence the complete genome of these strains to analyze their genetic characteristics to better understand the molecular evolution of the Korean strains compared with other TBEV strains.

Materials and methods

Viruses and cell lines

The two strains KrM 93 and KrM 213 were isolated from the lung tissues of *Apodemus agrarius* captured in two different provinces, Gurye-gun in the Jeollanam-do and Hapcheon-gun in the Gyeongsangnam-do, respectively. The viruses were propagated in 1-day-old suckling mice by intracerebral inoculation and the brains from moribund mice were homogenized into 10% (w/v) suspensions in phosphate-buffered saline (PBS; pH 7.0) containing 10% fetal bovine serum (FBS; GIBCO, Grand Island, NY), with penicillin (100 IU/ml) and streptomycin (100 µg/ml; P/S, GIBCO). Subsequently, 10% brain suspensions from the inoculated suckling mice were passaged three times on confluent monolayers of BHK-21 cells (ATCC No. CCL-10; American Type Culture Collection, Manassas, VA) in minimum essential medium (MEM; GIBCO) supplemented

with 8% FBS and 1% P/S in 5% CO₂ at 37°C. Culture fluids were collected after 4 days and stored at –80°C as the working virus stock for full-genome sequencing.

Primer design, RNA extraction, and cDNA synthesis

To amplify the complete genome sequence of the two Korean strains, 10 pairs of primers were designed based on the complete sequence of the Neudoerfl TBEV strain (GenBank Accession no. U27495). The primer nucleotide sequences and positions are listed in Table S1. Viral RNA was extracted with 1-ml TRIzol reagent (Invitrogen, Carlsbad, CA) from 200-µl viral stock, according to the manufacturer's protocol. Single-strand cDNA was synthesized using viral RNA and random hexamers through SuperScript™ III First-strand synthesis system for reverse transcription polymerase chain reaction (RT-PCR; Invitrogen) according to the manufacturer's protocol.

PCR amplification and TA cloning for complete TBEV genome sequencing

For the full-length genomic sequencing, eight overlapping parts containing the whole viral genome (Fig. S1A) were amplified by PCR from cDNA as mentioned above, using the *Ex Taq*™ DNA polymerase (Takara, Shiga, Japan) and gene-specific primer sets are listed in Table 1. PCR was performed using a GeneAmp® PCR system 9700 (Applied Biosystems, Foster City, CA) at 94°C for 5 min, followed by 30 cycles of amplification consisting of 94°C for 30 s, annealing at 52–58°C according to the primer sets and 72°C for 1 min 30 s, with a final extension at 72°C for 5 min.

To identify the sequences at the ends of the 5'- and 3'-noncoding regions (NCRs), the 5'-end cap structure was decapped using tobacco acid pyrophosphatase (Epicentre Tech, Madison, WI) and ligated with T4 RNA ligase (New England BioLabs Inc., Beverly, MA) for end-to-end ligation as described previously [13, 14] (Fig. S1B). Subsequently, circularized RNA was used as a template for cDNA synthesis using the SuperScript™ III First-strand synthesis system for RT-PCR (Invitrogen) and the TBE-18R primer. A first round of PCR amplification of the junction region containing the 5'- and 3'-NCRs was applied using the TBE-17F and TBE-18R primer set, and then the TBE-19F and TBE-20R primer set was used for nested PCR amplification.

The amplification products of RT-PCR or RT-nested PCR were purified using QIAEX II Gel extraction Kits (Qiagen, Valencia, CA) according to the manufacturer's protocol and sequenced after cloning into pGEM® T-easy vectors (Promega, Madison, WI), using ABI Prism BigDye Terminator Cycle Sequencing kits and ABI 3730xl

Table 1 Unique amino acid differences in the complete genome between the two Korean strains and four established vaccine strains

Amino acid position	TBEV Korean strain		Vaccine strain			
	KrM 93	KrM 213	Neudoerfl	K23	205	Sofjin-HO
C-43	M	M	M	M	M	L
C-54	A	A	A	A	A	V
C-64	N	N	N	N	N	K
C-111	L	L	L	L	I	L
M-41	L	L	L	L	I	M
M-61	V	V	V	V	I	V
E-51	S	S	N	K	N	N
E-83	A	A	A	T	A	A
E-122	G	E	E	E	E	E
E-136	K	K	K	R	K	K
E-153	A	A	A	A	A	V
E-167	V	V	I	V	V	V
E-201	E	K	E	E	E	E
E-228	K	K	K	K	R	K
E-317	A	A	A	A	T	I
E-460	L	L	L	L	M	V
E-463	A	A	A	A	A	V
E-479	S	S	S	S	G	S
NS1-4	C	C	C	C	C	G
NS1-103	P	P	P	S	P	P
NS1-123	M	M	M	I	M	M
NS1-175	P	P	P	P	P	S
NS1-346	V	V	V	V	V	I
NS2A-40	I	I	I	L	I	I
NS2A-41	V	V	V	I	V	V
NS2A-108	T	T	T	I	T	T
NS2A-127	D	D	E	D	A	A
NS2A-169	V	V	V	V	A	V
NS2A-183	C	C	C	C	Y	C
NS2A-201	I	I	V	I	I	I
NS2B-100	L	L	I	L	L	L
NS3-2	D	D	D	D	D	G
NS3-10	G	G	G	N	G	G
NS3-62	S	S	S	S	S	Y
NS3-174	K	K	K	R	K	K
NS3-184	V	V	V	V	S	T
NS3-218	I	I	I	I	T	I
NS3-372	A	A	A	A	V	A
NS3-380	T	T	T	T	T	A
NS4B-15	T	T	T	A	T	T
NS4B-16	V	V	A	V	V	V
NS4B-21	Q	Q	R	Q	H	H
NS4B-24	P	P	P	P	L	V
NS4B-55	I	I	I	I	M	I
NS4B-72	A	A	A	A	V	A
NS4B-88	F	F	F	F	F	L

Table 1 continued

Amino acid position	TBEV Korean strain		Vaccine strain			
	KrM 93	KrM 213	Neudoerfl	K23	205	Sofjin-HO
NS4B-90	V	V	V	V	I	V
NS4B-178	I	I	T	I	I	I
NS4B-198	S	S	S	S	S	A
NS4B-213	A	A	A	A	V	A
NS5-15	R	R	R	R	K	R
NS5-51	M	M	V	M	M	M
NS5-114	K	K	K	K	R	K
NS5-130	M	M	M	M	M	V
NS5-253	R	R	K	R	R	R
NS5-389	K	K	K	R	K	K
NS5-434	H	H	R	H	Q	Q
NS5-503	S	S	S	S	F	S
NS5-522	K	K	K	K	E	G
NS5-563	K	K	K	K	R	K
NS5-666	C	S	C	C	C	C
NS5-676	G	G	G	G	G	S
NS5-677	K	K	K	K	R	G
NS5-749	V	V	V	V	I	V
NS5-781	N	N	N	S	C	C
NS5-786	V	V	A	V	V	V
NS5-830	K	K	R	K	K	K
NS5-832	M	M	M	M	V	A
NS5-855	K	K	R	K	K	K
NS5-873	K	K	K	K	R	K

Bold letters represent unique amino acid substitutions

sequencer (Applied Biosystems) at Macrogen Inc. (Seoul, South Korea). At least three to five clones of plasmids containing precise inserts were sequenced in both directions using universal primers, T7 and SP6, and sequencing primers of each part are listed in Table S1.

Sequencing results were assembled using the SeqMan program implemented in DNASTAR software (version 5.0.6; DNASTAR Inc., Madison, WI) to determine the final assembly of complete genomic sequences. The full-length genomic sequences of KrM 93 and KrM 213 determined in this study have been deposited in GenBank under accession numbers, HM535611 and HM535610, respectively.

Multiple sequence alignment, sequence, and phylogenetic analyses

The nucleotide sequences of 31 fully sequenced TBEV strains, complete polyprotein sequences from two strains (MDJ-01 and Senzhang), and complete E gene sequences from 47 TBEV strains were used in sequence alignment. The strains used for phylogenetic analyses are listed in

Table S2. These sequences were collected from the GenBank database and their geographical origin, source of isolation, year of isolation, classification of subtype, and GenBank accession numbers are listed in Table S2. Multiple sequence alignment were performed using Clustal W [15] implemented in the MegAlign or BioEdit (version 7.0) programs. Sequence divergences and similarities of aligned nucleotide or deduced amino acid sequence were also assessed using MegAlign.

For detecting recombination events within complete genomic TBEV sequences, we used various algorithms, such as RDP [16], GENECONV [17], CHIMAERA [18], BOOTSCAN [19], MAXCHI [20], SISCAN [21], and 3SEQ [22] as implemented in RDP3 [23]. In general, we used the default settings for each algorithm in RDP3 and only potential recombination events detected by all algorithms were regarded as significant ($P < 0.05$).

Finally, phylogenetic analyses were performed based on the following regions of TBEV: complete genome (nucleotide and deduced amino acid sequence), 5'- and 3'-NCRs, three structural protein regions, and seven nonstructural protein regions. Phylogenetic trees were estimated by the maximum likelihood method using the general time-reversible (GTR) nucleotide substitution model or the LG amino acid substitution model. Branch support was calculated by the approximate likelihood ratio test (aLRT) using PHYML implemented in the program Seaview, version 4 [24, 25]. All trees were drawn using MEGA software version 4 [26].

Results

Genomic organization

The genomes of KrM 93 and KrM 213 were both 11,097-nt long. Their genomic organization consisted of a 132-nt 5'-NCR, a 10,245-nt ORF containing 10 viral protein-coding regions and a 720-nt 3'-NCR. Sequence positions, lengths, and protease cleavage sites are listed in Table S3. The base composition of the genome sequence of KrM 93 was A = 25.02%, G = 31.89%, T = 20.83%, and C = 22.27%. For KrM 213: A = 25.04%, G = 31.85%, T = 20.83%, and C = 22.29%.

Comparative analyses of full-length nucleotide and deduced amino acid sequences

The nucleotide (and deduced amino acid) sequence divergence of KrM 93 and KrM 213 ranged from 1.8 (0.7) to 19.2 (26.6)% and 1.9 (0.8) to 19.3 (26.7)%, respectively. A comparison of sequence identities between the two Korean strains and the other fully sequenced TBEV strains showed

that KrM 93 and KrM 213 had high identity with the strains belonging to the Western subtype with 97.3 (98.8) to 98.2 (99.3)% for KrM 93 and 97.2 (98.7) to 98.1 (99.2)% for KrM 213. These identity values were higher than with the Far-Eastern subtype with 83.8 (76.6) to 84.3 (94.1)% for KrM 93 and 83.8 (76.5) to 84.2 (94.0)% for KrM 213, or the Siberian subtypes with 84.5 (92.2) to 85.2 (94.3)% for KrM 93 and 84.5 (92.1) to 85.1 (94.2)% for KrM 213. In particular, KrM 93 showed the highest sequence identity with German strain Salem and the lowest identities with Russian strains Kavalerovo and Chinese strain MDJ-01. Likewise, the KrM 213 strain also showed the highest sequence identity with German strain Salem and Czech strain 263 and lowest identities with Russian strain Kavalerovo and Chinese strain MDJ-01.

The deduced amino acid sequences of KrM 93 and KrM 213 revealed 276 amino acid substitutions in the entire polyprotein region, compared with those of four TBEV vaccine strains: Neudoerfl, K23, Sofjin-HO, and 205. Among these, unique amino acid substitutions were observed in the E gene at position 122 (E → G) in KrM 93, and in the E gene at position 201 (E → K) and the NS5 gene at position 666 (C → S) for KrM 213 (Table 1). These amino acid changes represented nonconservative substitutions, as described previously [27].

Comparative analyses of 5'- and 3'-NCRs

All the 5'-NCRs of the two Korean strains were 132-nt long and had the same length as five strains belonging to the Western subtype (AS33, 263, Hypr, Toro-2003 and Neudoerfl) and three strains belonging to the Siberian subtype (Vasilchenko, Zausaev and Kolarovo-2008). When the 5'-NCR nucleotide sequences of the two Korean strains were aligned with 27 other TBEV strains, 26 TBEV strains including the two Korean isolates begin with an AGA TTT TCT TGC sequence at positions 1–12. However, this sequence was not conserved in the Salem, K23, or EK-328 strains. Among the Western subtype strains, the German strains Salem and K23, had shorter 5'-NCRs, with 21 and 23 nucleotide deletions at the start points, respectively, as shown in Fig. 1.

According to previous reports, TBEV strains have various sequence lengths in the 3'-NCR that were not associated with parameters such as isolation source, geographical origin or year of isolation [28, 29]. In this study, all the 3'-NCRs of the two Korean strains were 720-nt long, as with German strain AS33 isolated in 2005. The Neudoerfl and 263 strains had the longest 3'-NCRs with an additional long poly (A) sequence as shown in Figure S2. As shown in Figs. 1 and S2, all TBEV strains analyzed had highly conserved 5' and 3' complementary sequences (CS) in the 5'- and 3'-NCRs.

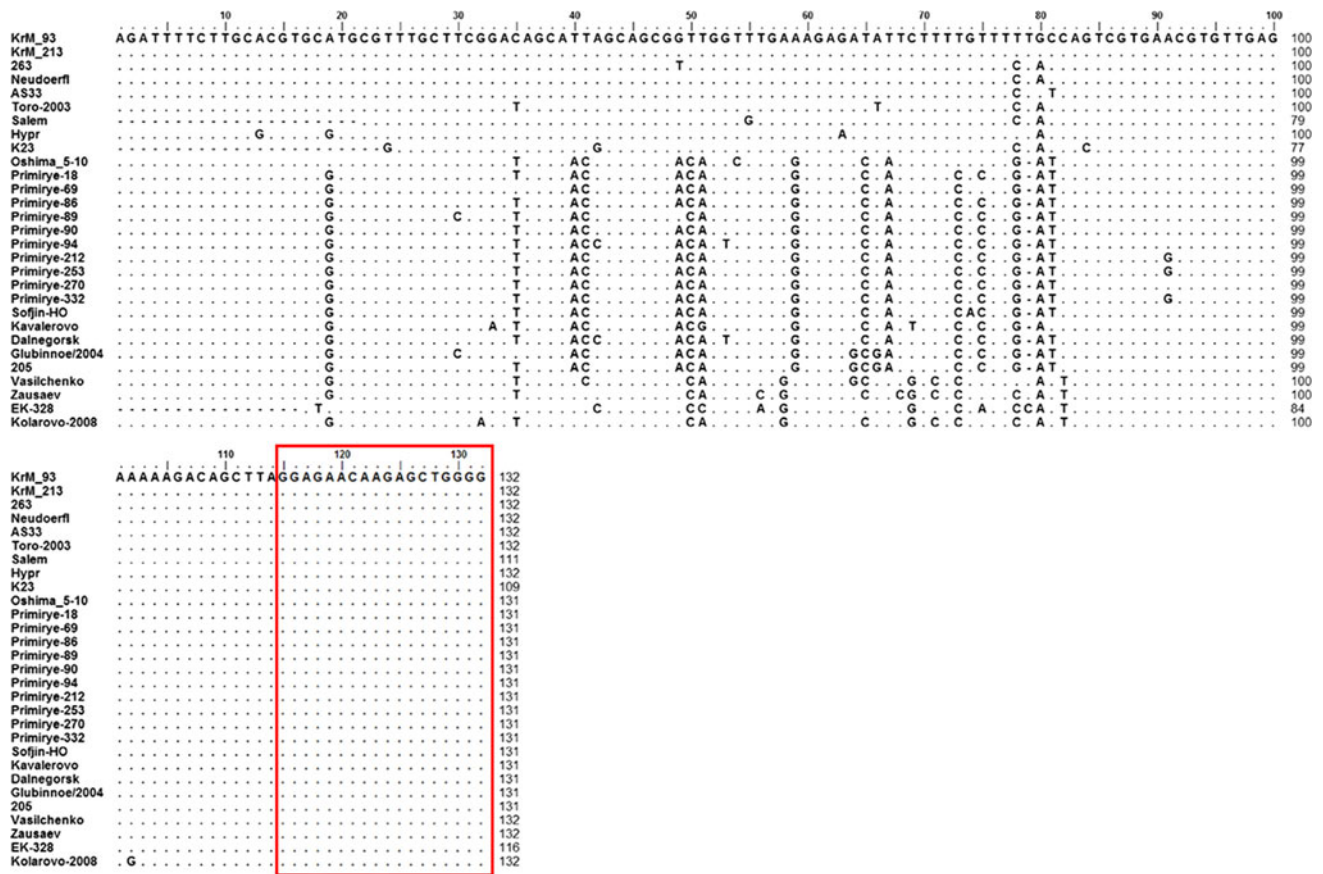


Fig. 1 Nucleotide sequence alignment of the 5'-noncoding region (NCR) of 29 fully sequenced TBEV strains, including the two Korean isolates. Compared with the sequence of Korean strain KrM 93,

identical and deleted regions are indicated by *dots* and *hyphens*, respectively. The *red open boxes* indicate conserved sequences (CS) designated 5' CSA as described previously [30]

Phylogenetic and recombination detection analyses

To further analyze the genetic relationship and evolution of TBEV strains, we performed phylogenetic analyses based on the various genomic regions of the fully sequenced TBEV strains, including two strains with complete poly-protein sequences (Chinese strains Senzhang and MDJ-01). Phylogenetic trees derived from nucleotide and deduced amino acid sequences of the complete genome indicated that the two Korean strains analyzed in this study belong to the Western TBEV subtype. A complete genome-based phylogenetic tree was divided into three distinct clusters and the two Korean strains were closely related to the German strain Salem and Czech strain 263 with high branch support, as shown in Fig. 2a and b. To compare this with the results of complete genome-based phylogenetic analysis, we also performed phylogenetic analyses based on structural (C, prM, M, and E genes), NS protein regions (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 genes) and on the 5'- and 3'-NCRs. As shown in Figure S3, the phylogenetic analyses of the C, E, NS1, NS4A, NS4B, and NS5 regions indicated that the branch pattern of

phylogenetic trees were similar to the complete genome-based phylogenetic tree. There were three distinct clusters: Far-Eastern, Siberian, and Western subtypes. The two Korean strains clustered with other TBEV strains belonging to the Western subtype. All phylogenetic trees analyzed in this study also showed that the Korean strains belonged to the same subgroup as the Western subtype, with high branch support.

To identify any recombination events in the 31 fully sequenced TBEV strains, including the two Korean isolates, we determined potential recombinant sequences, confirmed potential major/minor parental sequences, and potential recombination breakpoint sites using various methods implemented in the RDP3 program. Although putative recombination events were not observed in the two Korean strains analyzed here, we identified 11 putative recombination events, mostly in the 3'-NCRs (Table 2; Fig. 3). Nine putative recombinant strains (Russian strains Primorye-270, 178-79, Primorye-253, Primorye-332, Primorye-212, Primorye-90, Primorye-86, Primorye-69, and Primorye-18) were identified with the same major and minor parental sequences: Russian strain Primorye-94 and

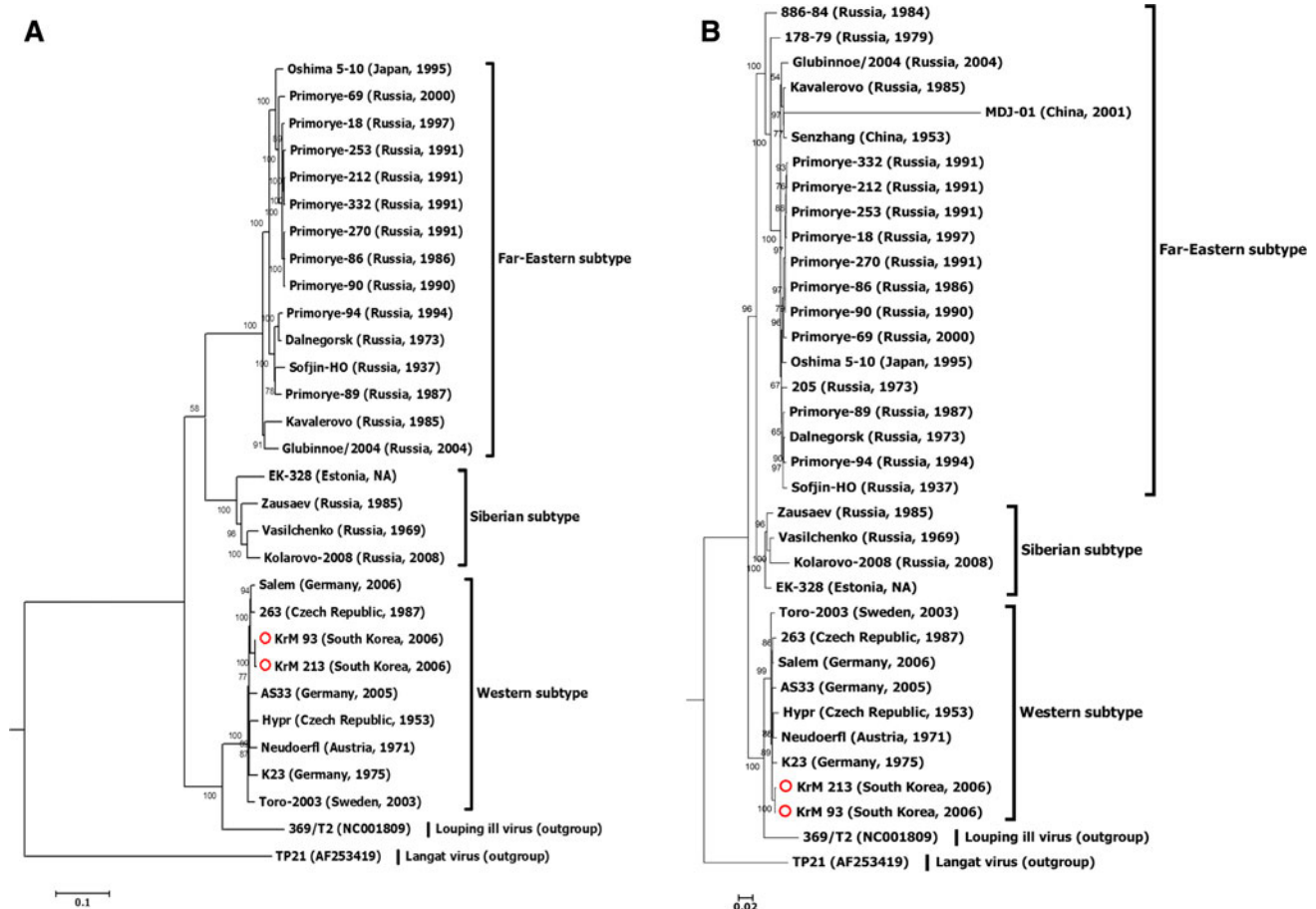


Fig. 2 Phylogenetic trees constructed using the maximum likelihood (ML) method between the two Korean strains and the other TBEV strains based on the **a** complete genome and **b** complete amino acid region. The ML trees were rooted with the sequences of the Louping ill virus (369/T2 strain, GenBank accession no. NC001809) and Langat virus (TP21 strain, GenBank accession no. AF253419) belonging to the tick-borne flavivirus (TBFV) group. Numbers at

each branch of the tree show approximate likelihood ratio test (aLRT) values. The *scale bar* indicates the nucleotide/amino acid substitutions per site. Each strain is identified by strain name followed by geographical origin and the year of isolation, except for the two outgroups. The Korean strains analyzed in this study are marked with *red open circles*

German strain Salem, respectively. The 10th putative recombination event was observed independently in Russian strain Dalnegorsk, with its major parent Primorye-94 and minor parent Primorye-89; its breakpoint site was found between positions 10,284 (NS5 region) to 10,816 (3'-NCR). Finally, the 11th putative recombinant, Russian strain Primorye-89, appeared to have major and minor parental sequences: Russian strain Sofjin-HO and Austrian strain Neudoerfl, respectively.

Discussion

We suggested previously that TBE might exist in South Korea. First, there was molecular evidence of TBEV infection in mammalian hosts (spleen or lung tissues of wild rodents and boars) and in potential vector ticks (*Haemaphysalis longicornis*, *H. flava*, *H. japonica*, and

Ixodes nipponensis) collected from South Korea [10, 11], although these ticks had not been previously reported as TBEV vectors. Second, we reported on the isolation and identification of five TBEV strains from lung tissue homogenates of wild rodents captured in South Korea using in vitro and in vivo experiments [10]. Contrary to our expectations, the sequence comparisons and phylogenetic analyses based on the complete E gene when compared with other TBEV strains indicated that all the five Korean strains belong to the Western subtype [12].

In this study, we determined the complete genome sequence of two Korean strains from different geographical locations and compared them with other fully sequenced TBEV strains. This is the first report of complete genome sequences of TBEV strains isolated from South Korea. The genomes of KrM 93 and KrM 213 are both 11,097-nt long and consist of 5'- and 3'-NCRs and ORFs containing 10 protein-coding regions (C, prM/M, E, NS1, NS2A, NS2B,

Table 2 Mean *P*-values and breakpoint positions of 11 putative recombinant events using seven different recombination detection methods

Recombinant strain	Recombinant detection method							Breakpoint position ^a
	RDP	GENOCOV	BOOTSCAN	MAXCHI	CHIMAERA	SISSCAN	3SEQ	
Primorye-270	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10449–10548
178-79	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10448–10710
Primorye-253	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10437–10584
Primorye-332	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10449–10597
Primorye-212	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10449–10772
Primorye-90	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10449–10596
Primorye-86	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10449–10596
Primorye-69	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10450–10669
Primorye-18	8.032×10^{-31}	1.070×10^{-30}	5.894×10^{-30}	6.390×10^{-9}	4.409×10^{-9}	8.706×10^{-14}	4.447×10^{-11}	10449–10696
Dalnegersk	2.911×10^{-19}	1.820×10^{-20}	1.661×10^{-20}	1.328×10^{-9}	8.271×10^{-3}	4.701×10^{-12}	1.097×10^{-17}	10284–10816
Primorye-89	1.822×10^{-7}	2.353×10^{-6}	5.384×10^{-9}	7.738×10^{-4}	4.249×10^{-3}	1.086×10^{-2}	9.894×10^{-5}	10444–10581

^a Nucleotide positions are indicated based on their position number in alignment

NS3, NS4A, NS4B, and NS5). They are similar to other viruses belonging to the *Flavivirus* genus. The 5'- and 3'-ends of the two Korean strains have paired CS regions with GGA GAA CAA GAG CTG GGG at the 5'-end and CGG TTC TTG TTC TCC at the 3'-end, as in other TBEV strains. Several studies reported CS regions at the 5'- and 3'-ends in the genomes of mosquito-borne, tick-borne, and unknown vector flaviviruses [30–32]. According to a recent study on tick-borne flaviviruses, hybridization between the 5'- and 3'-CS regions is necessary for RNA synthesis and viral replication [33].

Compared with the 31 other fully sequenced TBEV strains, the seven strains belonging to the Western subtype exhibited over 97 and 98% identities to KrM 93 and KrM 213 in nucleotide and deduced amino acid levels, respectively. Sequence similarity estimations for nucleotide positions using Simplot software also showed that KrM 93 and KrM 213 showed higher identity with the Western subtype strains than with the other subtypes at each genomic region (data not shown). These results suggest that the two Korean strains are closely related to the Western subtype strains, as with our previous results from analyses of the complete E genes [12]. Given that unique amino acid substitutions in flavivirus genomes influence viral attenuation [34], we also investigated the amino acid substitutions between the two Korean strains and four currently used TBEV vaccine strains. The KrM 93 and KrM 213 strains had three unique nonconservative amino acid substitutions in the E and NS5 protein regions. In particular, two of the three amino acid substitutions E122 (E → G) in KrM 93 and E201 (E → K) in KrM 213 identified in domain II of the E protein might affect virulence by increasing the positive charge of the E protein. For instance, it was reported that a unique amino acid substitution in the TBEV E gene at position 122 (E → G) produced significant attenuation of neuroinvasiveness in adult mice [35]. In addition, one nonconservative amino acid substitution in the NS5 region might also attenuate virulence by influencing functional RNA-dependent RNA polymerase (RdRp) activity, although a substitution at position 666 (C → S) in the NS5 region has not been confirmed to affect flaviviral virulence.

Several phylogenetic analyses of TBEV have focused on the E protein gene sequences to determine the genetic relationship and evolution among the TBEV strains [8, 9, 28, 36]. These sequences are associated with important biological functions such as inducing neutralizing antibodies and protective immune responses [34]. However, to investigate more detailed and improved phylogeny of TBEV, a complete genome-based phylogenetic analysis is indispensable. Here, we performed a larger phylogenetic analysis than a previous study [37], using 31 fully sequenced TBEV strains and two complete polyprotein

Fig. 3 BOOTSCAN evidences (pairwise distance model, with a window size 200, step size 20, and 1,000 Bootstrap replicates) for the identification of 11 putative recombinant TBEV strains: **a** Primorye-270, **b** 178-79, **c** Primorye-253, **d** Primorye-332, **e** Primorye-212, **f** Primorye-90, **g** Primorye-86, **h** Primorye-69, **i** Primorye-18, **j** Dalnegorsk, and **k** Primorye-89

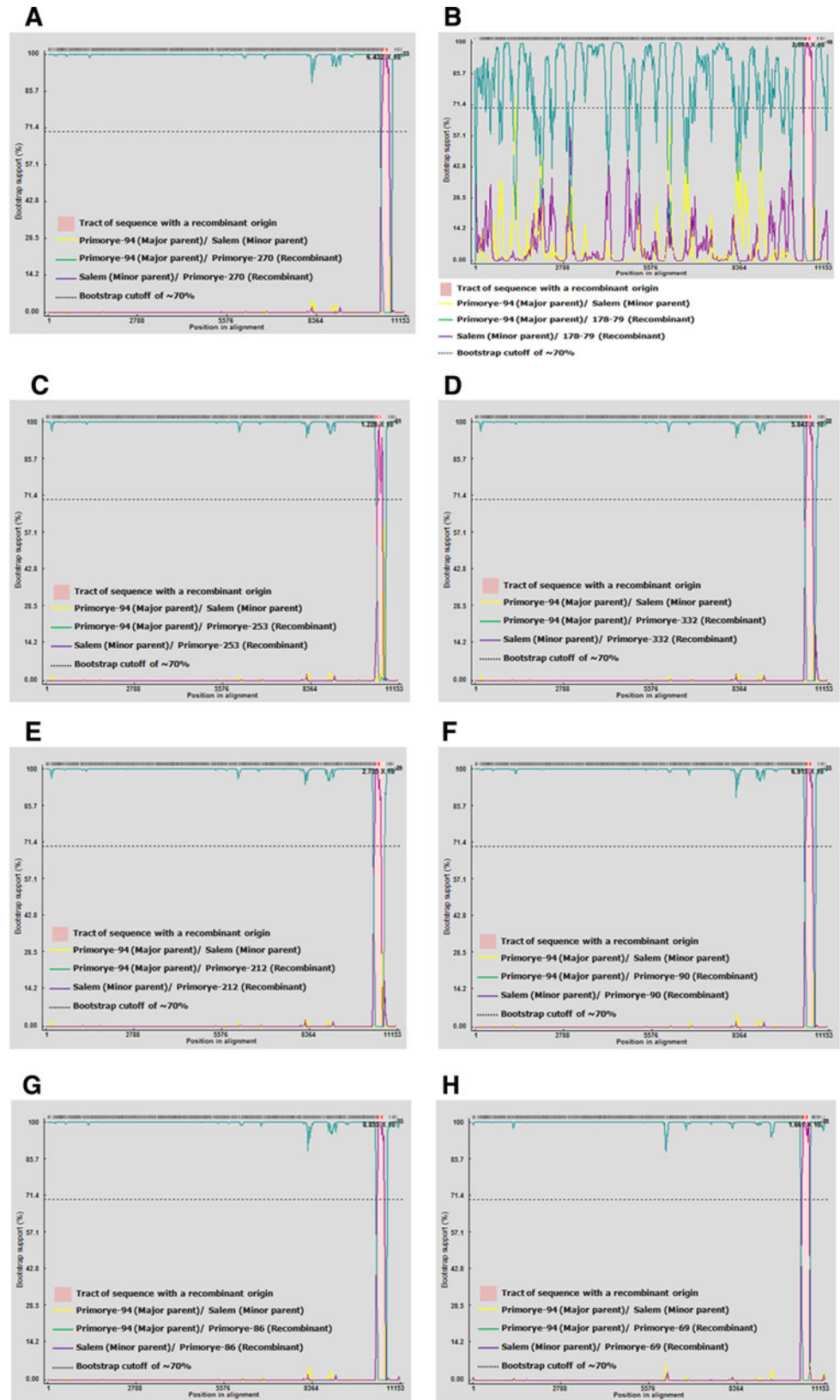
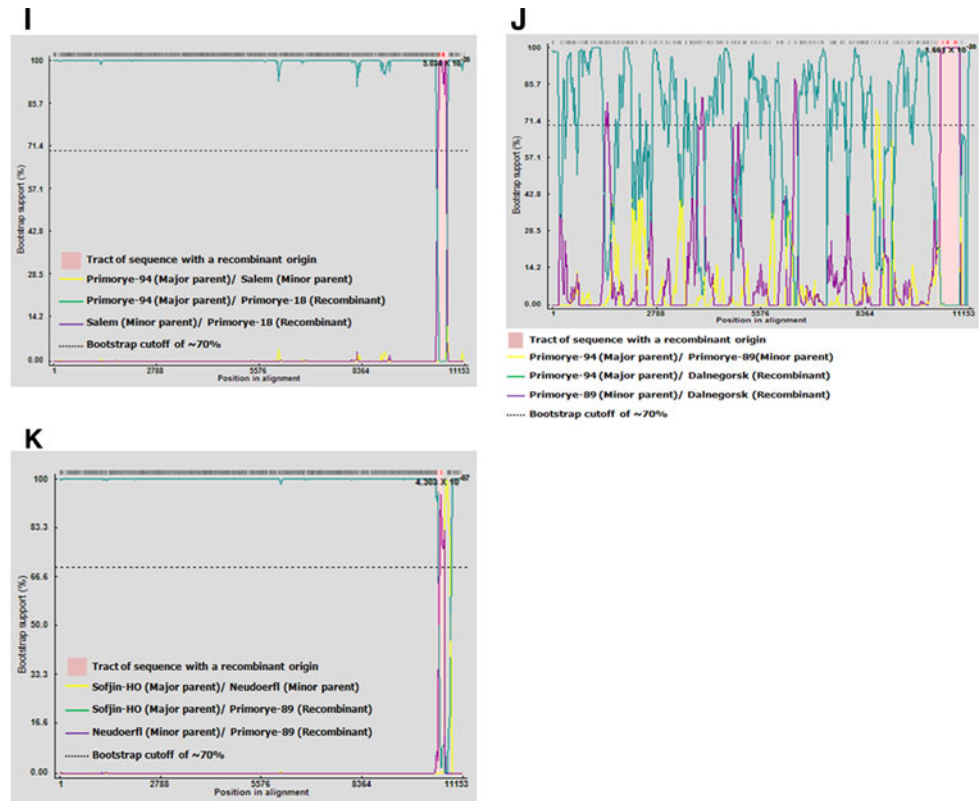


Fig. 3 continued



sequences. The complete genome-based phylogenetic tree showed that viruses were obviously divided three distinct subtypes and revealed that the two TBEV Korean strains are more closely related to the Salem and Czech 263 strains of the Western subtype than to other TBEV strains. Phylogenetic trees based on each genomic region, including the 5'- and 3'-NCRs, showed that phylogenetic trees derived from C, E, NS1, NS4A, NS4B, and NS5 were similar to that based on the complete genome sequence. In particular, the trees obtained from the C, E, NS4B, and NS5 regions represented the same branch pattern as shown by the complete genome-based tree. This showed that the Far-Eastern subtype is phylogenetically more closely related to the Siberian subtypes than to the Western subtype. These results suggest that the C, E, NS4B, and NS5 gene sequences could be useful genetic markers for a robust phylogenetic analysis of TBEV strains.

Finally, we analyzed the available 31 complete TBEV genome sequences, including the two Korean strains to identify any possible recombination events among the TBEV strains. These analyses may explain why the Korean strains are more closely related to the Western subtype, despite the geographical location. Previous studies reported recombination in flaviviruses, including Dengue virus (DENV), West Nile virus (WNV), Japanese encephalitis virus (JEV), and Saint Louis encephalitis virus (SLEV) [38–41]. However, these results were only based on the E gene sequence,

except for DENV type 1 and WNV, which were based on the analysis of complete genome sequences [39, 42]. No evidence for recombination events was detected in tick-borne flaviviruses [40]. In this study, 11 putative recombination events within the NS5 region or the 3'-NCR of TBEV strains were identified by various recombination detection methods, while none was detected in the two Korean strains we studied. The 10 strains with putative recombination events (Primorye-270, 178-79, Primorye-253, Primorye-332, Primorye-212, Primorye-90, Primorye-86, Primorye-69, Primorye-18, and Primorye-89 strains) exhibited intersubtypic recombination between Far-Eastern subtype strains as the major parent and Western subtype strains as the minor parent. One recombinant strain (Dalnegorsk) showed intrasubtypic recombination. Furthermore, all the recombinants showing intersubtypic recombination involved TBEV isolates from humans (Primorye-94 and Sofjin-HO strains), monkeys (Salem strain), or from a tick-borne vector (Neudoerfl strain).

This is the first report in which all TBEV strains for which sequences are available were analyzed for detecting putative recombination events. Inter- or intrasubtypic recombinations were possible in the NS5 or 3'-NCR. In addition the intersubtypic recombinant strains showed that recombination events had occurred between two TBEV strains isolated from different isolation sources. This suggests that coinfection with two different TBEV subtypes

that normally infect different hosts can cause intersubtypic recombination.

In summary, we determined here for the first time the complete genome sequence of these two Korean TBEV strains. We investigated their molecular characteristics by comparative genetic, phylogenetic, and recombination analyses based on the full-length nucleotide and deduced amino acid sequences, and compared them with the available fully sequenced TBEV strains. These results provide insight into the genetics of TBEV strains, which is necessary for understanding the molecular epidemiology, genetic diversity, and evolution of TBEV around the world. A further direction of this study will be to construct full-length infectious cDNA clones of Korean strains using reverse genetics. This will help future research on TBEV replication, virulence and pathogenesis, and facilitate the generation of recombinant TBEV vaccines.

Acknowledgments This research was funded by an intramural grant of the Korea National Institute of Health (Grant number 2010-N53006-00).

References

- U. Dumpis, D. Crook, J. Oksi, *Clin. Infect. Dis.* **28**, 882–890 (1999)
- M. Ecker, S.L. Allison, T. Meixner, F.X. Heinz, *J. Gen. Virol.* **80**(Pt 1), 179–185 (1999)
- T.J. Chambers, C.S. Hahn, R. Galler, C.M. Rice, *Annu. Rev. Microbiol.* **44**, 649–688 (1990)
- L. Lindquist, O. Vapalahti, *Lancet* **371**, 1861–1871 (2008)
- V.N. Bakhvalova, V.A. Rar, S.E. Tkachev, V.A. Matveev, L.E. Matveev, A.S. Karavanov, A.K. Dobrotvorsky, O.V. Morozova, *Virus Res.* **70**, 1–12 (2000)
- T.S. Gritsun, P.A. Nuttall, E.A. Gould, *Adv Virus Res.* **61**, 317–371 (2003)
- I. Golovljova, S. Vene, K.B. Sjolander, V. Vasilenko, A. Plyusnin, A. Lundkvist, *J. Med. Virol.* **74**, 580–588 (2004)
- A.E. Jaaskelainen, T. Tikkaoski, N.Y. Uzcategui, A.N. Alekseev, A. Vaehri, O. Vapalahti, *Emerg. Infect. Dis.* **12**, 1568–1571 (2006)
- K. Lundkvist, S. Vene, I. Golovljova, V. Mavtchoutko, M. Forsgren, V. Kalnina, A. Plyusnin, *J. Med. Virol.* **65**, 730–735 (2001)
- S.Y. Kim, S.M. Yun, M.G. Han, I.Y. Lee, N.Y. Lee, Y.E. Jeong, B.C. Lee, Y.R. Ju, *Vector Borne Zoonotic Dis.* **8**, 7–13 (2008)
- S.Y. Kim, Y.E. Jeong, S.M. Yun, I.Y. Lee, M.G. Han, Y.R. Ju, *Med. Vet. Entomol.* **23**, 15–20 (2009)
- S.M. Yun, S.Y. Kim, M.G. Han, Y.E. Jeong, T.S. Yong, C.H. Lee, Y.R. Ju, *Vector Borne Zoonotic Dis.* **9**, 287–293 (2009)
- J.H. Nam, S.L. Chae, S.Y. Won, E.J. Kim, K.S. Yoon, B.I. Kim, Y.S. Jeong, H.W. Cho, *Am. J. Trop. Med. Hyg.* **65**, 388–392 (2001)
- S.I. Yun, S.Y. Kim, C.M. Rice, Y.M. Lee, *J. Virol.* **77**, 6450–6465 (2003)
- J.D. Thompson, D.G. Higgins, T.J. Gibson, *Nucleic Acids Res.* **22**, 4673–4680 (1994)
- D. Martin, E. Rybicki, *Bioinformatics* **16**, 562–563 (2000)
- M. Padidam, S. Sawyer, C.M. Fauquet, *Virology* **265**, 218–225 (1999)
- D. Posada, K.A. Crandall, *Proc. Natl Acad. Sci. USA* **98**, 13757–13762 (2001)
- D.P. Martin, C. Williamson, D. Posada, *Bioinformatics* **21**, 260–262 (2005)
- J.M. Smith, *J. Mol. Evol.* **34**, 126–129 (1992)
- M.J. Gibbs, J.S. Armstrong, A.J. Gibbs, *Bioinformatics* **16**, 573–582 (2000)
- M.F. Boni, D. Posada, M.W. Feldman, *Genetics* **176**, 1035–1047 (2007)
- L. Heath, E. van der Walt, A. Varsani, D.P. Martin, *J. Virol.* **80**, 11827–11832 (2006)
- M. Gouy, S. Guindon, O. Gascuel, *Mol. Biol. Evol.* **27**, 221–224 (2010)
- S. Guindon, O. Gascuel, *Syst. Biol.* **52**, 696–704 (2003)
- K. Tamura, J. Dudley, M. Nei, S. Kumar, *Mol. Biol. Evol.* **24**, 1596–1599 (2007)
- S. Vрати, R.K. Giri, A. Razdan, P. Malik, *Am. J. Trop. Med. Hyg.* **61**, 677–680 (1999)
- D. Hayasaka, L. Ivanov, G.N. Leonova, A. Goto, K. Yoshii, T. Mizutani, H. Kariwa, I. Takashima, *J. Gen. Virol.* **82**(Pt 6), 1319–1328 (2001)
- G. Wallner, C.W. Mandl, C. Kunz, F.X. Heinz, *Virology* **213**, 169–178 (1995)
- A.A. Khromykh, H. Meka, K.J. Guyatt, E.G. Westaway, *J. Virol.* **75**, 6719–6728 (2001)
- P. Leyssen, N. Charlier, P. Lemey, F. Billoir, A.M. Vandamme, E. De Clercq, X. de Lamballerie, J. Neyts, *Virology* **293**, 125–140 (2002)
- C.W. Mandl, H. Holzmann, C. Kunz, F.X. Heinz, *Virology* **194**, 173–184 (1993)
- R.M. Kofler, V.M. Hoenninger, C. Thurner, C.W. Mandl, *J. Virol.* **80**, 4099–4113 (2006)
- P.C. McMinn, *J. Gen. Virol.* **78**(Pt 11), 2711–2722 (1997)
- C.W. Mandl, H. Kroschewski, S.L. Allison, R. Kofler, H. Holzmann, T. Meixner, F.X. Heinz, *J. Virol.* **75**, 5627–5637 (2001)
- W. Melik, A.S. Nilsson, M. Johansson, *Arch. Virol.* **152**, 1027–1034 (2007)
- V.A. Ternovoi, E.V. Protopopova, E.V. Chausov, D.V. Novikov, G.N. Leonova, S.V. Netesov, V.B. Loktev, *Emerg. Infect. Dis.* **13**, 1574–1578 (2007)
- E.C. Holmes, M. Worobey, A. Rambaut, *Mol. Biol. Evol.* **16**, 405–409 (1999)
- B.E. Pickett, E.J. Lefkowitz, *Virol. J.* **6**, 165 (2009)
- S.S. Twiddy, E.C. Holmes, *J. Gen. Virol.* **84**, 429–440 (2003)
- N.Y. Uzcategui, D. Camacho, G. Comach, R. Cuello de Uzcategui, E.C. Holmes, E.A. Gould, *J. Gen. Virol.* **82**, 2945–2953 (2001)
- H.J. Tolou, P. Couissinier-Paris, J.P. Durand, V. Mercier, J.J. de Pina, P. de Micco, F. Billoir, R.N. Charrel, X. de Lamballerie, *J. Gen. Virol.* **82**, 1283–1290 (2001)