Virology 530 (2019) 32–38

Contents lists available at ScienceDirect

Virology

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

Adaptive genetic diversifications among tick-borne encephalitis virus subtypes: A genome-wide perspective

Yan Li[a,b,*], Dawei Wang[c], Xiaogang Du[d]

[a] College of Animal Science and Technology, Sichuan Agricultural University, Wenjiang, People's Republic of China
[b] Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Wenjiang, People's Republic of China
[c] State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, People's Republic of China
[d] College of Life Science, Sichuan Agricultural University, Yaan, People's Republic of China

ARTICLE INFO

Keywords:
Tick-borne encephalitis virus
Diversification
Genome-wide analysis
Protein C
Nonstructural protein NS1
Nonstructural protein NS2A

ABSTRACT

Tick-borne encephalitis (TBE) is a severe neurological illness in humans. Tick-borne encephalitis virus (TBEV), the causative agent, can be grouped into Far Eastern, Siberian, and (Western) European subtypes. These subtypes are characterized by diverse vector specificity, host range and clinical manifestations. To provide hints for the decisive genetic factors underlying their epidemic and pathogenic diversities, we performed a genome-wide evolutionary study to evaluate the genetic diversity accompanied with the segregations of TBEV subtypes. The results show that adaptive selection has driven the diversification among these subtypes. Furthermore, the adaptive divergence-related changes have taken place on the proteins C, NS1, and/or NS2A. These results highlight candidate genes for the epidemic and pathogenic diversities, and will be useful in understanding the epidemic and pathogenic features of TBE.

1. Introduction

Tick-borne encephalitis (TBE) is a life-threatening neurological disease in humans. TBE is prevalent over a wide area ranging from Japan and northern China, through far-eastern Russia to Europe (Süss, 2011), and more than 10,000 cases of the disease arise per year (Lindquist and Vapalahti, 2008). Over the past decades, TBE has been characterized by a continued increase in human cases and geographic expansion of the disease-affect area (Csángó et al., 2004; Randolph, 2010; Süss, 2008), and become a growing public health problem (Kunze, 2016).

The etiological agent responsible for TBE is tick-borne encephalitis virus (TBEV). TBEV has a positive-sense single-stranded RNA genome approximately 10.5 kb in length. The genome consists of two untranslated regions at the 5′- and 3′-ends, and one long open reading frame. The open reading frame encodes a polyprotein, which is further co- and post-translationally processed by viral and host cell proteases into three structural proteins (C, prM, and E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Heinz and Allison, 2003).

TBEV is a member of the virus genus *Flavivirus*, of the family *Flaviviridae* (Wengler et al., 1995). This family includes many species which are pathogenic in humans, such as yellow fever, dengue, and Japanese encephalitis (Calisher, 1988; Gritsun et al., 2003a). TBEV isolates are currently divided into Far Eastern (TBEV-Fe), Siberian (TBEV-Sib), and (Western) European (TBEV-Eu) subtypes, which are designated according to their predominant geographic distribution in Eurasian continent (Heinz et al., 2000; Simmonds et al., 2012). Genetic heterogeneity analyses revealed that TBEV-Fe and TBEV-Sib are phylogenetically closer to each other than to TBEV-Eu (Grard et al., 2007; Heinz et al., 2000; Simmonds et al., 2012).

TBE is a vector-borne disease. The principal vectors of TBEV are two species of *Ixodes* ticks, *I. persulcatus* and *I. ricinus*, although the virus has been isolated from 16 species of hard ticks (Pukhovskaya et al., 2018). *I. ricinus* has a role as the main vector of the subtype TBEV-Eu, while *I. persulcatus* transmits primarily the subtypes TBEV-Fe and TBEV-Sib (Ecker et al., 1999; Gaunt et al., 2001; Gritsun et al., 2003c; Hayasaka et al., 2001; Lundkvist et al., 2001; Mavtchoutko et al., 2000; Süss, 2003). Other hard ticks from genera *Dermacentor* and *Haemaphysalis* have served as vectors for local TBE outbreaks (Süss, 2011). In natural environment, the primary hosts of TBEV are small forest mammals. These animals can act as reservoirs for the virus without development of clinical symptoms (Estrada-Peña and de la Fuente, 2014). The subtypes of TBEV demonstrate a geographic difference in preference for

* Corresponding author at: College of Animal Science and Technology, Sichuan Agricultural University, Wenjiang, People's Republic of China.
E-mail address: liyan@sicau.edu.cn (Y. Li).

https://doi.org/10.1016/j.virol.2019.02.006
Received 7 December 2018; Received in revised form 8 February 2019; Accepted 8 February 2019
Available online 12 February 2019
0042-6822/

particular small mammal species that is, the rodents *Apodemus flavicollis*, *A. sylvaticus*, and *Myodes* spp. act as the main hosts and reservoirs for TBEV-Eu in Europe, *A. agrarius* and *A. peninsula* for TBEV-Sib in Siberia, and *Myodes rufocanus* and *Microtus arvalis* for TBEV-Fe in the Far East (Süss, 2011).

Humans are only accidental hosts of TBEV (Bogovic and Strle, 2015). Human infections with TBEV are associated with a range of clinical manifestations from simple fever to severe encephalitis with or without myelitis (Gritsun et al., 2003b). It has been observed that clinical course and outcome are usually associated with the subtype of TBEV infection (Gritsun et al., 2003b; Mansfield et al., 2009; Votiakov et al., 1978). Infection with the TBEV-Fe subtype usually causes a more severe course of the disease, with a case-fatality rate of 20–40% and higher rates of severe neurologic sequelae (Votiakov et al., 1978). In contrast to severe consequences of infection with the TBEV-Fe subtype, the disease caused by the TBEV-Eu subtype often has a biphasic course, producing a case-fatality rate of 0.5–2% and milder, mostly without sequelae (Burke and Monath, 2001). Infection with the TBEV-Sib subtype characteristically induces a chronic progressive form and case-fatality rates of lower than 6–8% (Gritsun et al., 2003b).

Although previous studies have provided comprehensive identifications of the vector specificity, host range and clinical manifestations, the genetic basis for the diversities in these aspects remains, to some extent, deficient (Bakhvalova et al., 2016; Kunze, 2016; Süss, 2011). The spectrum of epidemic and pathogenic features of a pathogen can be a reflection of its adaptations while spreading into new ecological niche. By this logic, detecting potential adaptive genetic changes of TBEV is an important step forward defining and understanding the genetic determinants of epidemic and pathogenic differences. In this study, to dropped helpful hints for the genetic determinants, we assessed the quality and quantity of adaptive genetic diversification that accompanied the phylogenetic segregations of the TBEV subtypes and their epidemic and pathogenic diversities.

2. Materials and methods

2.1. Sequence data collection and alignment

Complete genomes and coding sequences of TBEV field isolates were retrieved from the National Center for Biotechnology Information (NCBI) database. One hundred and sixty-eight sequences were finally used for following evolutionary characterization (Supplementary Table 1). Supplementary Table 1 shows the NCBI GenBank accession numbers, dates of collection, and origins of the included isolates. After sequence alignment was carried out using ClustalW (Thompson et al., 1997), the codon reading frames of the alignment result were checked by eye and the ambiguous codons were manually removed.

2.2. Recombination detection

Recombinant sequence in the retrieved isolates was detected with the RDP v4.56 software package (Martin et al., 2015). The embedded methods BOOTSCAN (Martin et al., 2005), GENECONV (Padidam et al., 1999), Maximum Chi Square (MAXCHI) (Smith, 1992), Recombination Detection Program (RDP) (Martin and Rybicki, 2000), and Sister Scanning (SISCAN) (Gibbs et al., 2000) were adopted for identifying recombination events from the complete coding region. The *P* value cutoff was set to 0.05 and used throughout. Only recombination event detected with three or more methods was taken into consideration.

2.3. Phylogenetic analysis

To infer phylogenetic relationships among the TBEV isolates used in this study, Maximum likelihood (ML) analysis in MEGA5 was performed on the complete coding sequences of the proteins E, as well as those of NS3 and NS5 (Tamura et al., 2011). For each of the datasets, we performed nucleotide substitution model selection with the lowest Bayesian Information Criterion scores (Nei and Kumar, 2000). The Tamura-Nei (TN93) model with gamma-distributed rates for variable sites (five categories) and a proportion of invariant sites fits all three data sets with the lowest score values and was employed in the ML analysis. The reliabilities of the reconstructed trees were evaluated by the bootstrap method with 1000 replications (Felsenstein, 1985).

2.4. McDonald-Kreitman (MK) test on the coding sequences of all individual proteins

The MK test was applied to assess whether adaptive positive diversification has taken place between TBEV-Eu, TBEV-Fe, and TBEV-Sib. In this test, for coding sequences of a protein, if replacement (nonsynonymous) and synonymous changes were due to neutral mutations alone, the ratio of replacement-to-synonymous divergence between phylogenic clusters would be expected to be the same as the ratio of replacement-to-synonymous variations within the clusters. In contrast, if some of the replacements have been influenced by adaptive positive selection, the ratio of replacement-to-synonymous divergence should be greater than that of replacement-to-synonymous variation.

In this study, the MK test was implemented on all 10 individual proteins separately taking into consideration that the viral proteins perform different functions and might have undergone different evolutionary processes (Chambers et al., 1990; Mandl, 2005). Because of the relatively small numbers of divergences and variations in the individual coding sequences, Fisher's exact test of independence was performed and significant excess of replacement divergence between the subtypes can be used to reject the neutral mutation hypothesis in favor of adaptive diversification (McDonald, 2009).

2.5. Divergence scan

The divergences of nucleotide sequences within TBEV-Eu, TBEV-Fe, and TBEV-Sib were assessed by means of the Simplot program (Lole et al., 1999). The reference sequences of TBEV subtypes were randomly chosen as follows: TBEV-Eu, KF151173 and KU885457; TBEV-Fe, FJ402886 and FJ997899; TBEV-Sib, JN003206 and JN003209. SimPlot calculates and graphically plots the percentage identity using a sliding window of 200 nt moving along the complete coding sequences in steps of 20 nt.

2.6. Test for the functional divergence on the coding sequences of putative diversified proteins

The method Gu (2006) in the software program DIVERGE 3.0 was performed to detect the type II functional divergence at the amino acid sites of the putative diversified proteins separately (see below) (Gu, 2006; Gu et al., 2013). Type II functional divergence means that homologous amino acid residues are conserved within individual clusters but their physiochemical properties are different between the clusters (Gu, 2001, 2006). The coefficient of type II functional divergence (θ_{II}) between the clusters is estimated (Gu, 2006), and the Z-score test is implemented for statistical evaluation based on the value of θ_{II} and its standard error (θ_{II} SE) (Gu, 2001; Gu et al., 2013). Under the null hypothesis, the θ_{II} value is equal to zero, which shows that the absence of this type of functional divergence between descendant clusters. A rejection of the null hypothesis would indicate that a considerable change in amino acid physiochemical properties may have taken place - that is, the homologous proteins have potentially undergone functional divergence with the split of the clusters. Furthermore, a posterior probability-based confidence measure is used to predict amino acid sites that could be responsible for the functional divergences with a probability cutoff (Gu, 2006).

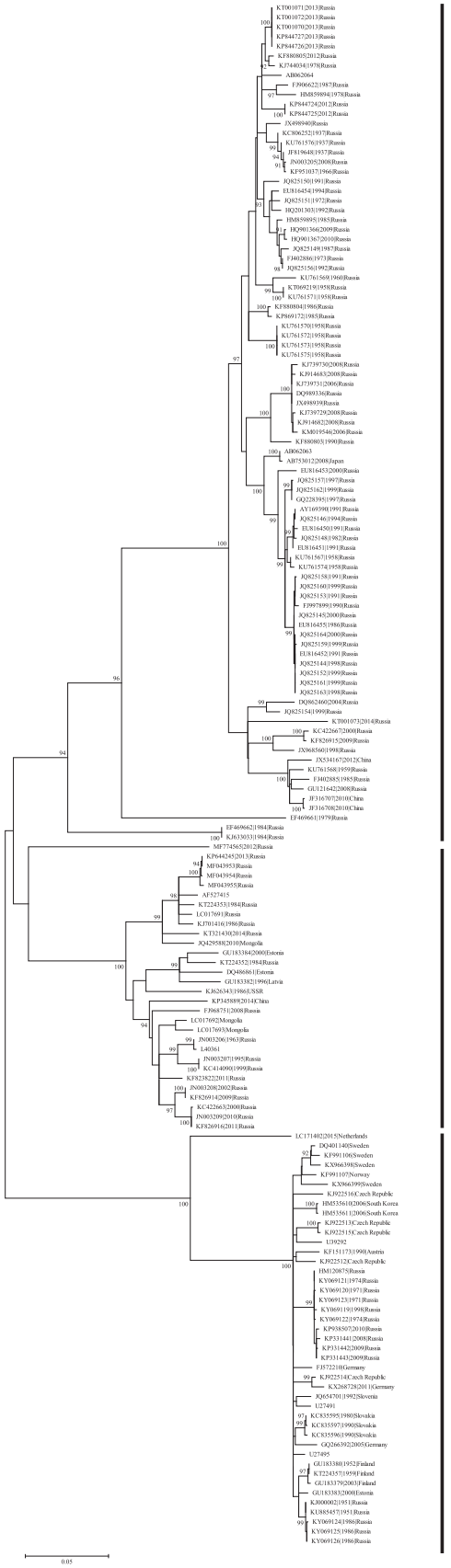

Fig. 1. Maximum-likelihood gene tree for the complete coding sequences of envelope protein E extracted from the 160 complete genome and coding sequences represented in Supplementary Table 1. The numbering of subtypes follows that of previous results (Grard et al., 2007; Heinz et al., 2000; Simmonds et al., 2012). The tips are labeled with GenBank accession numbers, dates of collection, and origins of the isolates. The bootstrap values below 90% are not reported.

3. Results and discussion

3.1. Phylogenetic relationship of TBEV isolates

In order to detect adaptive genetic diversities that accompanied the segregations of the phylogenetic subtypes, it is obviously necessary to state the corrective phylogenetic relationships of TBEV isolates. However, recombination would invalidate the prerequisite, rendering the documents of the evolutionary history very misleading. Therefore, in the first instance, we performed a recombination analysis covering the complete coding sequences of the TBEV isolates with the RDP v4.56 package (Martin et al., 2015). The result showed that nine well-supported recombination events may have taken place in the coding region. In these events, eight putative recombinant isolates (AY182009, GU183381, JQ650522, JQ650523, JQ825147, JQ825155, KJ755186, and KM019545) were identified. These recombinants were removed to eliminate the influence of the recombination, and a total of 160 sequences were enrolled in the following step of the analytical procedures.

We then reconstructed nucleotide trees using the complete coding sequences of E, NS3, and NS5. As shown in Fig. 1 and Supplementary Fig. 1, the clustering structures of the obtained trees are generally similar. Each of the trees comprises three clusters, which were designated subtypes TBEV-Eu, TBEV-Fe, and TBEV-Sib in previous studies (Heinz et al., 2000; Simmonds et al., 2012). The divisions among these subtypes are highly resolved with bootstrap values over 90%. Within each of the subtypes, the isolates were further grouped into a major lineage with further rare disparate isolate(s) (*i.e.,* LC171402 of TBEV-Eu; EF469661, EF469662, and KJ633033 of TBEV-Fe; MF774565 of TBEV-Sib) (Fig. 1). The subsequent analyses focus only on the major lineages of the subtypes because few disparate isolates have heretofore been collected and might have belonged to relict groups substituted later by other TBEV isolates, such as MF774565 of TBEV-Sib (Tkachev et al., 2017).

3.2. Signatures of adaptive diversifications among the TBEV-Eu, TBEV-Sib, and TBEV-Fe subtypes

The MK test has been commonly used method for detecting the occurrence of adaptive positive diversity (McDonald and Kreitman, 1991). This test compares the number of replacement and synonymous changes both between and within phylogenic clusters. Thus, it is obviously that intra-cluster subdivision will generate a downward bias of divergences between the clusters and an upward bias of variations within the clusters. In the case of TBEV, from an evolutionary perspective, TBEV-Fe and TBEV-Sib show clear genetic heterogeneity although they have arisen from a common ancestor and made up a cluster (Fig. 1) (Grard et al., 2007). From epidemic and pathogenic perspectives, they differ from each other in several respects, such as host range and clinical presentations (Gritsun et al., 2003b; Süss, 2011; Votiakov et al., 1978). These features raise further concerns about the implication of the TBEV-Fe-TBEV-Sib subdivision for revealing clear footprints of adaptive genetic diversities between the TBEV subtypes.

In order to avoid or reduce the potential biases from the TBEV-Fe-TBEV-Sib subdivision, we first compared TBEV-Eu with these descendant subtypes separately rather than their aggregation alone. For the comparison between the TBEV-Eu and TBEV-Sib subtypes, the MK test

Table 1
McDonald-Kreitman test on TBEV individual coding sequences between the TBEV-Eu subtype and TBEV-Sib subtype.

TBEV-Eu *vs.* TBEV-Sib	Divergence between subtypes	Variation within subtypes
Protein C		
Replacement	12	25
Synonymous	8	69
Probability[a]		0.007**
PreM		
Replacement	6	36
Synonymous	21	138
Probability		1.000
E		
Replacement	10	67
Synonymous	62	432
Probability		1.000
NS1		
Replacement	14	40
Synonymous	52	284
Probability		0.076
NS2A		
Replacement	10	47
Synonymous	28	223
Probability		0.039*
NS2B		
Replacement	5	19
Synonymous	10	112
Probability		0.132
NS3		
Replacement	15	101
Synonymous	80	527
Probability		1.000
NS4A and 2K		
Replacement	4	23
Synonymous	18	147
Probability		0.745
NS4B		
Replacement	6	37
Synonymous	34	228
Probability		1.000
NS5		
Replacement	18	141
Synonymous	87	732
Probability		0.780

Single asterisk corresponds to *Probability* < 0.05. Double asterisks correspond to *Probability* < 0.01.
[a] The significance according to the Fisher's exact test (two tailed).

Table 2
McDonald-Kreitman test on TBEV individual coding sequences between the TBEV-Eu subtype and TBEV-Fe subtype.

TBEV-Eu *vs.* TBEV-Fe	Divergence between subtypes	Variation within subtypes
Protein C		
Replacement	10	34
Synonymous	11	83
Probability[a]		0.126
PreM		
Replacement	7	51
Synonymous	16	146
Probability		0.803
E		
Replacement	13	78
Synonymous	62	482
Probability		0.482
NS1		
Replacement	15	63
Synonymous	45	295
Probability		0.209
NS2A		
Replacement	18	55
Synonymous	29	216
Probability		0.013*
NS2B		
Replacement	5	21
Synonymous	13	120
Probability		0.179
NS3		
Replacement	18	103
Synonymous	86	534
Probability		0.775
NS4A and 2K		
Replacement	4	38
Synonymous	20	153
Probability		0.793
NS4B		
Replacement	9	42
Synonymous	39	221
Probability		0.672
NS5		
Replacement	30	172
Synonymous	112	807
Probability		0.350

Single asterisk corresponds to *Probability* < 0.05.
[a] The significance according to the Fisher's exact test (two tailed).

was applied to identify the sign of adaptive diversities in all individual proteins. The results showed that there were clear differences in evolutionary pattern among the viral proteins (Table 1). For the coding sequences of protein C and NS2A, the Fisher's exact test of independence indicated that the ratios of replacements *vs*. fixed synonymous divergences were significantly greater than those of replacements *vs.* synonymous variations (Table 1). In addition, the test on that of NS1 exhibited a similar deviation from the neutral mutation hypothesis with a marginal probability (*Probability* = 0.076). However, the results from other coding sequences did not display such pattern of evolution (Table 1). We repeated the MK test with the TBEV-Eu and TBEV-Fe isolates to evaluate whether adaptive diversification has contributed to their evolution. The result of the Fisher's exact test on NS2A showed an excess of replacement divergences between subtypes with a significant probability (Table 2). In contrast, the results from the other protein were nearly opposite from that of NS2A and did not display a signature of adaptive diversity (Table 2).

Following the comparisons between the TBEV-Eu and TBEV-Sib subtypes and between the TBEV-Eu and TBEV-Fe subtypes, we detected the occurrence of adaptive diversity with the TBEV-Fe-TBEV-Sib subdivision. For protein C and NS2A, the statistical results of Fisher's exact test indicated that the ratio of the inter-subtypes replacements to the fixed synonymous divergences was greater than that of the intra-subtypes replacements *vs.* the synonymous variations with significant probabilities (*Probability* < 0.05) (Table 3). Meanwhile, the tests on other proteins revealed that no excess of replacements accumulated with the split of the subtypes.

There are two points worth noting in above results. First, from the view of demography, slightly deleterious nonsynonymous mutations can be fixed during a rapid expansion of a small ancestral population (McDonald and Kreitman, 1991). Therefore, it is also feasible that the accumulation of inter-subtype replacements and the significant results of the MK test can be due to the impact of demographic factors. Under this scenario, a demographic factor should have similar effects on the different genomic regions in the case of no recombination. In this study, however, the incongruous evolutionary patterns of the different coding sequences indicated the effect of demographic factors could be roughly ruled out (Tables 1, 2). Second, under the scenario of heterogeneity in mutation rate across the genome, the rapid accumulation of variations

Table 3
McDonald-Kreitman test on TBEV individual coding sequences between the TBEV-Sib subtype and TBEV-Fe subtype.

TBEV-Sib *vs.* TBEV-Fe	Divergence between subtypes	Variation within subtypes
Protein C		
Replacement	5	30
Synonymous	3	83
Probability[a]	0.044*	
PreM		
Replacement	6	42
Synonymous	9	156
Probability	0.110	
E		
Replacement	7	88
Synonymous	36	524
Probability	0.822	
NS1		
Replacement	7	55
Synonymous	29	336
Probability	0.456	
NS2A		
Replacement	14	53
Synonymous	15	242
Probability	< 0.001**	
NS2B		
Replacement	3	24
Synonymous	8	129
Probability	0.392	
NS3		
Replacement	6	124
Synonymous	47	571
Probability	0.264	
NS4A and 2K		
Replacement	4	33
Synonymous	15	155
Probability	0.753	
NS4B		
Replacement	5	34
Synonymous	21	253
Probability	0.346	
NS5		
Replacement	14	194
Synonymous	56	882
Probability	0.749	

Single asterisk corresponds to *Probability* < 0.05. Double asterisks correspond to *Probability* < 0.01.
[a] The significance according to the Fisher's exact test (two tailed).

on some proteins' coding sequences might bring about the absence of significant statistics on these regions and incongruous evolutionary patterns of viral proteins from the MK test. To put this into perspective, we constructed similarity plots with SimPlot program to assess how related is one coding sequence of isolates within TBEV-Eu, TBEV-Fe, and TBEV-Sib (Lole et al., 1999). As shown in Supplementary Fig. 2, for each subtype, the sequence homology was generally similar throughout genome. In this situation, the incongruence of the evolutionary signatures among the viral coding sequences is difficult to explain under the hypothesis of heterogeneous mutation distribution in the coding regions. Thus, the combined results suggest that adaptive diversifications could have contributed to the accumulation of amino acid changes in protein C, NS2A, and/or NS1 between TBEV-Eu and TBEV-Sib, NS2A between TBEV-Eu and TBEV-Fe, and protein C and NS2A with the TBEV-Fe-TBEV-Sib segregation.

Table 4
The estimates of the type II functional divergence of the candidate proteins.

Protein	Estimate (Z-score)	Divergence-related amino acid sites[b]
TBEV-Eu *vs.* TBEV-Sib		
Protein C	θ_{II}[a] = 0.111 ± 0.050 (2.220*)	61, 81, 98 (at $P_{II} \geq 0.90$) 2, 3, 50, 56, 91, 95, 103 (at $0.90 > P_{II} \geq 0.85$)
NS1	$\theta_{II} = 0.038 \pm 0.021$ (1.810*)	262, 285, 316 (at $P_{II} \geq 0.85$)
NS2A	$\theta_{II} = 0.083 \pm 0.037$ (2.243**)	34, 164, 225 (at $P_{II} \geq 0.95$) 24, 52, 166, 189 (at $0.90 > P_{II} \geq 0.85$)
TBEV-Eu *vs.* TBEV-Fe		
NS2A	$\theta_{II} = 0.111 \pm 0.041$ (2.707**)	30, 126, 175, 182, 227 (at $0.95 > P_{II} \geq 0.90$) 36, 50, 100, 167, 189 (at $0.90 > P_{II} \geq 0.85$)
TBEV-Sib *vs.* TBEV-Fe		
Protein C	$\theta_{II} = 0.036 \pm 0.060$ (0.600)	3, 102 (at $P_{II} \geq 0.85$)
NS2A	$\theta_{II} = 0.114 \pm 0.039$ (2.923**)	34, 35, 36, 50, 124, 127, 164, 167, 187, 224, 227 (at $P_{II} \geq 0.95$) 126, 151 (at $0.90 > P_{II} \geq 0.85$)

Single and double asterisks correspond to $P < 0.05$ and < 0.01 (one-tailed *t*-test), respectively.
[a] Coefficient of type-II functional divergence.
[b] For the TBEV-Eu-TBEV-Sib and the TBEV-Eu-TBEV-Fe comparisons, the numbers of amino acid sites represent their position in the individual coding sequences as aligned to U27495; for the TBEV-Sib-TBEV-Fe comparisons, the numbers of amino acid sites represent their position in the individual coding sequences as aligned to JF819648.

3.3. Amino acid sites experienced functional divergence with the segregations of the TBEV subtypes

It has been widely accepted that the changes of amino acid physiochemical properties at critical sites could contribute to functional diversity associated with different environmental challenges (Li, 1983; Nei, 1987). Hence, using the method Gu-2006 in DIVERGE 3, we identified the impacts of adaptive diversification at individual sites in the candidate proteins which had experienced adaptive diversity with the segregations of the TBEV subtypes. For protein C, NS1, and NS2A of TBEV-Eu and TBEV-Sib, the coefficient of the type II functional divergence (θ_{II}) values were larger than 0 with significance (*Probability* < 0.05) in accordance with the results of the MK test (Table 4). Furthermore, the site-specific scores of the posterior probability ratio showed that quite a number of amino acid sites within these proteins could have undergone functional divergence at $P \geq 0.85$ level (Table 4). For NS2A of TBEV-Eu and TBEV-Fe, the result exhibited a significant deviation from neutrality in favor of the type II functional divergence (*Probability* < 0.01) (Table 4). The site-specific scores indicated that the type II functional divergence could has contributed to adaptive changes at ten sites with a probability of ≥0.85 (Table 4).

Finally, based on the results of the MK test, we attempted to identify the type II functional divergence at individual amino acid sites in protein C and NS2A with the TBEV-Sib-TBEV-Fe subdivision. For NS2A, the θ_{II} value was larger than 0 with significance (*Probability* < 0.01) (Table 4). Furthermore, the site-specific scores showed that the changes at 13 sites could be related to the functional divergence at the *Probability* ≥ 0.85 level (Table 4). In protein C, the posterior probability estimation indicated that two sites might have been influenced by

functional divergence although the θ_{II} value was not significant (Table 4). Taken together, these results support that the type II functional divergence could have contributed to adaptive changes at a number of sites with the phylogenic split of the TBEV subtypes.

On the whole, from a genetic perspective, our analyses suggested that adaptive diversifications have occurred on the viral proteins C, NS2A, and/or NS1 with the divisions of the TBEV subtypes. Unfortunately, because the overall structure of flavivirus NS2A is still lacking while protein C of TBEV shows little sequence identity with those of other flaviviruses, it is presently unknown whether and how the adaptive changes at the identified amino acid sites affect the functions of these proteins. For NS1, it has been known that this protein encompasses three distinct domains: an N-terminal β-roll, an α/β Wing domain, and a C-terminal β-ladder (Akey et al., 2014, 2015; Brown et al., 2016; Wallis et al., 2004). Of note was the our obtained result that all three identified sites lie in C-terminal β-ladder (Edeling et al., 2014). This domain consists an extended β-sheet on one face and a "spaghetti loop" on the opposite face (Akey et al., 2015). The latter has an epitope on the loop face for monoclonal antibody (MAb) binding, which can mediate immune responses against flaviviruses (Chung et al., 2006; Edeling et al., 2014). It is further noteworthy that two of the epitope containing amino acid sites (261 and 315) in West Nile Virus NS1 correspond to the identified sites 262 and 316 in the TBEV homologue (Table 3) (Edeling et al., 2014). Thus, the changes at these sites could bring about a functional benefit to decrease the recognition by anti-NS1 antibodies of hosts in individual geographic areas. Moreover, it has been demonstrated that mosquito-borne flaviviruses utilize NS1 proteins to enhance their acquisition by vectors (Cheng et al., 2016; Liu et al., 2016, 2014). The infection ratios of the mosquitoes that fed on the mice immunized with the dengue virus serotypes 2 NS1 mutant (harbored deletions of antigenic regions that contain the homologues of sites 262 and 316) were lower than those of the mosquitoes that fed on the full-length NS1 immunized mice (Liu et al., 2016). Thus, an alternative explanation for the adaptive diversifications is that these changes may be a genetic trait developed by TBEV to adapt to different vectors while spreading into new ecological niche.

In summary, our results revealed that the TBEV subtypes have experienced adaptive diversifications rendering them good candidates for giving rise to epidemics when disseminated in new geographic locations. The significance of the adaptive diversifications offered helpful hints for the genetic determinants in TBEV vector specificity, host range and clinical manifestations. Of course, further structural and functional studies will be necessary to better understand the highlighted proteins and their adaptive genetic changes. If the inferences are experimentally supported, they would shed more insight into the pathogenesis of TBEV infection and its epidemics over a wide geographic area.

Acknowledgments

We are grateful to Prof. Wen Wang for his helpful comments on the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.virol.2019.02.006.

References

Akey, D.L., Brown, W.C., Dutta, S., Konwerski, J., Jose, J., Jurkiw, T.J., DelProposto, J., Ogata, C.M., Skiniotis, G., Kuhn, R.J., Smith, J.L., 2014. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. Science 343, 881–885.
Akey, D.L., Brown, W.C., Jose, J., Kuhn, R.J., Smith, J.L., 2015. Structure-guided insights on the role of NS1 in flavivirus infection. Bioessays 37, 489–494.
Bakhvalova, V.N., Chicherina, G.S., Potapova, O.F., Panov, V.V., Glupov, V.V., Potapov, M.A., Seligman, S.J., Morozova, O.V., 2016. Tick-Borne Encephalitis Virus Diversity in Ixodid Ticks and Small Mammals in South-Western Siberia, Russia. Vector Borne Zoonotic Dis. 16, 541–549.
Bogovic, P., Strle, F., 2015. Tick-borne encephalitis: a review of epidemiology, clinical characteristics, and management. World J. Clin. Cases 3, 430–441.
Brown, W.C., Akey, D.L., Konwerski, J.R., Tarrasch, J.T., Skiniotis, G., Kuhn, R.J., Smith, J.L., 2016. Extended surface for membrane association in Zika virus NS1 structure. Nat. Struct. Mol. Biol. 23, 865–867.
Burke, D.S., Monath, T.P., 2001. Flaviviruses. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology. Lippincott Williams & Wilkins, London, New York, Tokyo, pp. 1043–1125.
Calisher, C.H., 1988. Antigenic classification and taxonomy of flaviviruses (family Flaviviridae) emphasizing a universal system for the taxonomy of viruses causing tick-borne encephalitis. Acta Virol. 32, 469–478.
Chambers, T.J., Hahn, C.S., Galler, R., Rice, C.M., 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44, 649–688.
Cheng, G., Liu, Y., Wang, P., Xiao, X., 2016. Mosquito Defense Strategies against Viral Infection. Trends Parasitol. 32, 177–186.
Chung, K.M., Nybakken, G.E., Thompson, B.S., Engle, M.J., Marri, A., Fremont, D.H., Diamond, M.S., 2006. Antibodies against West Nile Virus nonstructural protein NS1 prevent lethal infection through Fc gamma receptor-dependent and -independent mechanisms. J. Virol. 80, 1340–1351.
Csángó, P.A., Blakstad, E., Kirtz, G.C., Pedersen, J.E., Czettel, B., 2004. Tick-borne encephalitis in southern Norway. Emerg. Infect. Dis. 10, 533–534.
Ecker, M., Allison, S.L., Meixner, T., Heinz, F.X., 1999. Sequence analysis and genetic classification of tick-borne encephalitis viruses from Europe and Asia. J. Gen. Virol. 80 (Pt 1), 179–185.
Edeling, M.A., Diamond, M.S., Fremont, D.H., 2014. Structural basis of Flavivirus NS1 assembly and antibody recognition. Proc. Natl. Acad. Sci. USA 111, 4285–4290.
Estrada-Peña, A., de la Fuente, J., 2014. The ecology of ticks and epidemiology of tick-borne viral diseases. Antivir. Res. 108, 104–128.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39 (4), 783–791.
Gaunt, M.W., Sall, A.A., de Lamballerie, X., Falconar, A.K., Dzhivanian, T.I., Gould, E.A., 2001. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J. Gen. Virol. 82, 1867–1876.
Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.
Grard, G., Moureau, G., Charrel, R.N., Lemasson, J.J., Gonzalez, J.P., Gallian, P., Gritsun, T.S., Holmes, E.C., Gould, E.A., de Lamballerie, X., 2007. Genetic characterization of tick-borne flaviviruses: new insights into evolution, pathogenetic determinants and taxonomy. Virology 361, 80–92.
Gritsun, T.S., Frolova, T.V., Zhankov, A.I., Armesto, M., Turner, S.L., Frolova, M.P., Pogodina, V.V., Lashkevich, V.A., Gould, E.A., 2003a. Characterization of a siberian virus isolated from a patient with progressive chronic tick-borne encephalitis. J. Virol. 77, 25–36.
Gritsun, T.S., Lashkevich, V.A., Gould, E.A., 2003b. Tick-borne encephalitis. Antivir. Res. 57, 129–146.
Gritsun, T.S., Nuttall, P.A., Gould, E.A., 2003c. Tick-borne flaviviruses. Adv. Virus Res. 61, 317–371.
Gu, X., 2001. Maximum-likelihood approach for gene family evolution under functional divergence. Mol. Biol. Evol. 18, 453–464.
Gu, X., 2006. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. Mol. Biol. Evol. 23, 1937–1945.
Gu, X., Zou, Y., Su, Z., Huang, W., Zhou, Z., Arendsee, Z., Zeng, Y., 2013. An update of DIVERGE software for functional divergence analysis of protein family. Mol. Biol. Evol. 30, 1713–1719.
Hayasaka, D., Ivanov, L., Leonova, G.N., Goto, A., Yoshii, K., Mizutani, T., Kariwa, H., Takashima, I., 2001. Distribution and characterization of tick-borne encephalitis viruses from Siberia and far-eastern Asia. J. Gen. Virol. 82, 1319–1328.
Heinz, F.X., Allison, S.L., 2003. Flavivirus structure and membrane fusion. Adv. Virus Res. 59, 63–97.
Heinz, F.X., Collett, M.S., Purcell, R.H., Gould, E.A., Howard, C.R., Houghton, M., Moormann, R.J.M., Rice, C.M., Thiel, H.J., 2000. Family Flaviviridae. In: Regenmortel, M.H.V.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., Mcgeoch, D.J., Pringle, C.R. (Eds.), Virus Taxonomy 7th International Committee for the Taxonomy of Viruses. Academic Press, San Diego, pp. 859–878.
Kunze, U., 2016. The International Scientific Working Group on Tick-Borne Encephalitis (ISW TBE): review of 17 years of activity and commitment. Ticks Tick-Borne Dis. 7, 399–404.
Li, W.-H., 1983. Evolution of duplicate genes and pseudo genes. In: Nei, M., Koehn, R.K. (Eds.), Evolution of Genes and Proteins. Sinauer Associates, Sunderland, Mass, pp. 14–37.
Lindquist, L., Vapalahti, O., 2008. Tick-borne encephalitis. Lancet 371, 1861–1871.
Liu, J., Liu, Y., Nie, K., Du, S., Qiu, J., Pang, X., Wang, P., Cheng, G., 2016. Flavivirus NS1 protein in infected host sera enhances viral acquisition by mosquitoes. Nat.

Microbiol. 1, 16087.
Liu, Y., Zhang, F.C., Liu, J.Y., Xiao, X.P., Zhang, S.Y., Qin, C.F., Xiang, Y., Wang, P.H., Cheng, G., 2014. Transmission-blocking antibodies against mosquito C-type lectins for dengue prevention. PLoS Pathog. 10.
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. J. Virol. 73, 152–160.
Lundkvist, A., Vene, S., Golovljova, I., Mavtchoutko, V., Forsgren, M., Kalnina, V., Plyusnin, A., 2001. Characterization of tick-borne encephalitis virus from Latvia: evidence for co-circulation of three distinct subtypes. J. Med. Virol. 65, 730–735.
Mandl, C.W., 2005. Steps of the tick-borne encephalitis virus replication cycle that affect neuropathogenesis. Virus Res. 111, 161–174.
Mansfield, K.L., Johnson, N., Phipps, L.P., Stephenson, J.R., Fooks, A.R., Solomon, T., 2009. Tick-borne encephalitis virus – a review of an emerging zoonosis. J. General. Virol. 90, 1781–1794.
Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003.
Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retrovir. 21, 98–102.
Mavtchoutko, V., Vene, S., Haglund, M., Forsgren, M., Duks, A., Kalnina, V., Horling, J., Lundkvist, A., 2000. Characterization of tick-borne enchephalitis virus from Latvia. J. Med. Virol. 60, 216–222.
McDonald, J.H., 2009. Handbook of Biological Statistics, 3rd ed. Sparky House Publishing, Baltimore, Maryland.
McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.
Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford and New York.
Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.
Pukhovskaya, N.M., Morozova, O.V., Vysochina, N.P., Belozerova, N.B., Bakhmetyeva, S.V., Zdanovskaya, N.I., Seligman, S.J., Ivanov, L.I., 2018. Tick-borne encephalitis virus in arthropod vectors in the Far East of Russia. Ticks Tick-Borne Dis. 9, 824–833.
Randolph, S.E., 2010. To what extent has climate change contributed to the recent epidemiology of tick-borne diseases? Vet. Parasitol. 167, 92–94.
Süss, J., 2003. Epidemiology and ecology of TBE relevant to the production of effective vaccines. Vaccine 21, S19–S35.
Süss, J., 2008. Tick-borne encephalitis in Europe and beyond – the epidemiological situation as of 2007. Eurosurveillance 13, 717–727.
Süss, J., 2011. Tick-borne encephalitis 2010: epidemiology, risk areas, and virus strains in Europe and Asia—an overview. Ticks Tick-Borne Dis. 2, 2–15.
Simmonds, P., Becher, P., Collett, M.S., Gould, E.A., Heinz, F.X., Meyers, G., Monath, T., Pletnev, A., Rice, C.M., Stiasny, K., Thiel, H.-J., Weiner, A., Bukh, J., 2012. Family Flaviviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA, pp. 1003–1020.
Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Tkachev, S.E., Chicherina, G.S., Golovljova, I., Belokopytova, P.S., Tikunov, A.Y., Zadora, O.V., Glupov, V.V., Tikunova, N.V., 2017. New genetic lineage within the Siberian subtype of tick-borne encephalitis virus found in Western Siberia, Russia. Infect. Genet. Evol. 56, 36–43.
Votiakov, I., Protas, I.I., Zhdanov, V.M., 1978. Western tick-borne encephalitis, Belarus. Minsk 256.
Wallis, T.P., Huang, C.Y., Nimkar, S.B., Young, P.R., Gorman, J.J., 2004. Determination of the disulfide bond arrangement of dengue virus NS1 protein. J. Biol. Chem. 279, 20729–20741.
Wengler, G., Bradley, D.W., Collett, M.S., Heinz, F.X., Schlesinger, R.W., Strauss, J.H., 1995. Flaviviridae. In: Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., Summers, M.D. (Eds.), Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Vienna & New York, pp. 415–427.