



Clusteron structure of tick-borne encephalitis virus populations

Sergey Y. Kovalev*, Tatyana A. Mukhacheva

Laboratory of Molecular Genetics, Department of Biology, Ural Federal University, Lenin Avenue 51, Yekaterinburg 620000, Russia

ARTICLE INFO

Article history:

Received 28 September 2012

Received in revised form 29 October 2012

Accepted 30 October 2012

Available online 28 November 2012

Keywords:

Tick-borne encephalitis virus
Clusteron
Evolution
Phylogeography
Classification
Population

ABSTRACT

Tick-borne encephalitis is a natural focal transmissible zoonosis. The causative agent of the disease is a tick-borne encephalitis virus (TBEV) belonging to the genus *Flavivirus* of the family *Flaviviridae* and is widespread in Eurasia. Current TBEV classification based on molecular genetic data comprises three phylogenetically separate subtypes: Far Eastern, European and Siberian (TBEV-Sib). Further differentiation of TBEV isn't developed, making it difficult to investigate the origins, distribution and evolution of the virus. In the present study we determined the nucleotide sequence of the gene E fragment for 282 TBEV-Sib isolates from *Ixodes persulcatus* ticks or their pools from various natural foci in Russia. Analysis of these sequences and sequences obtained from the GenBank database (more than 600), made it possible to cluster TBEV-Sib strains by identical amino acid sequences of a glycoprotein E fragment. In total, 18 groups were identified (from 3 to 285 strains in the group). It was shown that TBEV strains belonging to the same group are phylogenetically related and have a territorial attachment showing either a local or a corridor type distribution. These groups were named as clusterons showed to be the smallest unit of TBEV classification. The grouping of TBEV strains allows characterization of endemic areas both in quantitative and qualitative composition of the clusterons. The approach could be successfully used to record and monitor the TBEV populations.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Tick-borne encephalitis (TBE) is a natural focal transmissible zoonosis. The disease is endemic and widespread in Eurasia from Western Europe to northern Japan. The etiological agent is a tick-borne encephalitis virus (TBEV), which belongs to the genus *Flavivirus* of the family *Flaviviridae*. TBEV causes severe encephalitis in humans and has a significant impact on public health in endemic regions, with an annual incidence of up to 13 thousand people and a case fatality rate ranging from 0.5% to 20% depending on location (Gritsun et al., 2003b; Suss, 2011). TBE epidemiology is closely related to the ecology and biology of ticks. Circulation of TBEV in nature is due to the constant exchange between the ticks and various warm-blooded animals, mainly rodents and birds feeding on the ground. The principal vectors are the ticks *Ixodes persulcatus* and *Ixodes ricinus*.

The TBEV genome presents a single-stranded RNA of positive polarity, approximately 11,000 bases in length, and has a single reading frame encoding a polyprotein (Pletnev et al., 1990). The open reading frame (ORF) is flanked by untranslated regions of the polyprotein. The genome encodes 10 large viral proteins, which are formed as a result of the processing of viral polyprotein. TBEV polyprotein consists of 3414 amino acids (aa). Three viral proteins

are structural, one capsid C protein and two surface M and E proteins which are embedded in the viral membrane of the virion. The other seven are non-structural proteins, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 providing replication of the viral genome into the cell.

The E glycoprotein consists of 496 aa and is a major structural protein of the TBEV membrane responsible for binding to cellular receptors, determination of the tropism and virulence and the formation of virus-neutralizing antibodies (Roehrig, 2003). X-ray analysis revealed the three-dimensional structure of the glycoprotein E (Rey et al., 1995). At neutral pH, glycoprotein E exists as a dimer, each monomer consisting of three domains (I–III) performing different functions, and hydrophobic regions forming the transmembrane “anchor” of the protein (Rey et al., 1995). Glycoprotein E homodimers is located parallel to the lipid membrane on the outer surface of the virus particles (Rey et al., 1995). It was shown that the majority of mutations that alter the properties of a pathogenic virus are grouped within domains of the envelope glycoprotein E (Mandl et al., 2001; Ternovoi et al., 2003; Hayasaka et al., 2004; Romanova et al., 2007). The nucleotide sequence corresponding to the N-terminus of the glycoprotein E, is variable and is therefore often used for genotyping TBEV isolates, for differential genetic diagnosis of TBE and in phylogenetic studies (Hayasaka et al., 1999; Ternovoi et al., 2003; Kovalev et al., 2009, 2010).

Current TBEV classification based on molecular genetic data comprises three subtypes: Far Eastern (TBEV-FE), European

* Corresponding author. Tel.: +7 343 261 68 26; fax: +7 343 350 74 01.
E-mail address: Sergey.Kovalev@usu.ru (S.Y. Kovalev).

(TBEV-Eu) and Siberian (TBEV-Sib) (Ecker et al., 1999; Gould et al., 2001; King et al., 2011). The name of the TBEV subtype is related to its geographical distribution. The amino acid at position 206 of glycoprotein E is successfully used as a subtype marker (Ecker et al., 1999). Classification of TBEV within a subtype is not currently developed. However, attempts to differentiate TBEV-Sib on the basis of structural features of the glycoprotein E and the geographic distribution of strains have been made (Hayasaka et al., 2001; Karan et al., 2006; Golovljova et al., 2008). Thus, one group of researchers distinguished Asian and East European lineages of strains based on the presence of amino acids His, Gln, or Tyr at position 234 of glycoprotein E (Karan et al., 2006). Another group of scientists divided TBEV-Sib into two lineages, Baltic and Siberian, by the presence of Asn or Thr at position 175 and Ala or Thr at position 313 of glycoprotein E. Baltic strains are found in the Baltic countries and the European part of Russia, and Siberian ones are distributed in Western and Eastern Siberia (Golovljova et al., 2008). The above approaches have shown the possibility of using single amino acid substitutions as phylogenetic markers. However, they were not informative enough, due to low resolution.

Despite advances in the study of the genetic diversity of nucleotide sequences, it is still difficult to answer the key questions relating to the formation of TBE foci and the evolution of TBEV. One of the reasons is the absence of a classification of TBEV within subtype. Recently, the possibility of differentiation of TBEV-Sib populations into groups, on the basis of amino acid sequence identity of the glycoprotein E fragment, was shown (Kovalev et al., 2009). This approach proved to be informative and allowed the authors to suggest the hypothesis of anthropogenic dissemination of natural foci of tick-borne encephalitis in Siberia, the Urals and Eastern Europe (Kovalev et al., 2009).

This study shows the development of the approach as a basis for further differentiation of TBEV populations and classification within a TBEV subtype. Also, we have attempted to show the utility of this approach in the registration, recording and monitoring of TBEV populations.

2. Materials and methods

2.1. Virus isolates

The 282 TBEV isolates used in this study were collected during 1966–2011. The isolates were obtained from single ticks *I. persulcatus* or its pools (95.7%), as well as from clinical specimens (4.3%). The isolates were collected from different territories of Russia, Sverdlovsk (193), Perm' (18), Tyumen' (42), Omsk (25), Kurgan (1) Leningrad (1), Chelyabinsk (1) and the Altai (1) regions. Information on their geographical origin, time and source of isolation is given in Supplementary Table S1. The virus isolates obtained before 1986 were passaged in suckling mice, their 10% brain suspensions were lyophilized and stored, without further passages, in the Collection of the Yekaterinburg Research Institute of Viral Infections. Isolates obtained after this period are presented as RNA extracted from ticks or clinical samples that was converted into cDNA.

2.2. RNA extraction

Viral RNA was extracted from 100 µl of tick suspension, blood serum or TE-buffer solution of a lyophilized suspension of suckling mouse brain with lysis solution and then purified with the RNA-sorb extraction Kit (InterLabService Ltd., Moscow, Russia), according to the manufacturer's instructions. The reverse transcription was done using the REVERTA random-primer Kit (InterLabService

Ltd., Moscow, Russia), according to the manufacturer's instructions.

2.3. PCR amplification and sequencing

The fragment of gene E was amplified using nested PCR with internal and external forward and reverse primers as described (Ternovoi et al., 2003), with modifications (Kovalev et al., 2009). Nucleotide sequences of gene E fragment PCR products (506 bp) of TBEV strains were determined using BigDye Terminator v3.1 Cycle Sequencing Kit and ABI PRIZM® 310 Genetic Analyzer, according to the manufacturer's instructions. GenBank Accession Numbers (JX315719–JX316000) of the nucleotide sequences are given in Supplementary Table S1.

2.4. Phylogenetic analysis

Phylogenetic analysis of nucleotide sequences of gene E fragment (without primer sequences (311–762 bp)) and deduced amino acid sequences of glycoprotein E fragment (104–254 aa) were conducted for all analyzed isolates and also 335, 191 and 114 strains of TBEV-Sib, TBEV-Eu and TBEV-FE from GenBank, respectively (Supplementary 1). To avoid confusion, hereinafter, isolate and strain will be referred to as a strain. Molecular phylogenetic analysis was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura, 1980) using MEGA v. 5.05 (Tamura et al., 2011). Construction of phylogenetic networks for each TBEV subtype was performed using Phylogenetic Network Software v. 4.6.1.0 (fluxus-engineering.com) using an algorithm MJ (Median-joining) (Bandelt et al., 1999).

2.5. Geographical distribution of TBEV-Sib clusterons

Information on the territorial distribution of TBEV-Sib clusterons is available as a Supplementary Google Earth map file *Clusterons TBEV-Sib.kml*. The Google Earth program is required to view this file (<http://www.google.com/earth/index.html>). The software provides access to satellite images of high quality Map-based presentation of these data.

2.6. Structural analysis of glycoprotein E

The X-ray crystal structure of glycoprotein E (Protein Data Base [PDB] entry 1SVB) (Rey et al., 1995) was used as the basis for the computer reconstruction of the dimer molecule of TBEV-Sib, strain Zausaev AF527415. Homology modeling was initially performed using ESyPred3D Web Server 1.0 (<http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esyypred/>) (Lambert et al., 2002). The 3D structure of glycoprotein E was visualized using the program RasMol version 2.7.5 (<http://rasmol.org>) (Sayle and Milner-White, 1995).

3. Results

3.1. Characteristics of the clusterons

Phylogenetic analysis showed that all 282 strains analyzed belonged to TBEV-Sib. A comparison of the amino acid sequences of gene E fragment of studied strains and sequences of TBEV-Sib strains from GenBank (total 617) identified 18 groups of strains with the same amino acid sequences. These groups were named as clusterons (Table 1). The number of strains in a clusteron ranged from 3 (0.5%) to 285 (46.2%). Clusteron name consists of two characters, the first is the number of the subtype (1-TBEV-FE, 2-TBEV-Eu and 3-TBE-Sib), the second is a letter attributed to a specific

Table 1
Characteristics of the clusterons of TBEV-Sib.

Clusteron	Prototype strain	GenBank Acc. No.	Number of strains (%)	dS ^a	Clusteron-specific amino acid signature ^b												
					119	120	128	136	151	165	167	175	204	221	228	232	234
3A	Zausaev	AF527415	285 (46.2)	135 ^c	A	A	T	K	V	F	V	T	K	N	K	A	H
3B	Ek125-2007	GU444253	35 (5.8)	6	N	.	.	.	G	Q
3C	Z6	EF566817	16 (2.6)	33	R
3D	Est54	DQ393773	50 (8.1)	68	N
3E	Ek1361-2008	GU444279	5 (0.8)	14	V	.
3F	Ek191-2007	GU444225	39 (6.3)	48	R	.	.
3G	Kokkola-102	DQ451295	9 (1.5)	12	V	N
3H	Aina	AF091006	13 (2.1)	33	V	Q
3I	Ek1669-2006	GU444204	11 (1.9)	3	I
3J	IR99-1m1	AB049348	3 (0.5)	20	Q
3K	1130	EU443275	4 (0.6)	18	.	.	.	R
3L	1699	EU443268	4 (0.6)	34	.	.	I
3M	24	GU143822	3 (0.5)	16	I
3N	IR99-1m4	AB049349	3 (0.5)	19	S
3O	Ek190-1-2007	GU444224	3 (0.5)	1	N	R
3P	Vologda-4-06	FJ214139	8 (1.3)	15	N	Y
3Q	Ek1752-2005	JX315727	5 (0.8)	3	S	.	N	.	.	G	Q	.
3U	886-84	EF469662	4 (0.6)	6	.	S	N
3Unique			117 (19.0)														
Total			617 (100)														

^a Number of synonymous substitutions.

^b Position numbers are given for glycoprotein E.

^c Only strains of the Asian lineage (see below).

Table 2
Characteristics of TBEV clusteron structure.

TBEV subtype	Number of strains	Number of unique strains (%)	Number of strains in clusterons (%)	Number of all aa variable positions (%)	Number of clusteron-specific substitutions	Number of clusterons	Minimum number of strains in the clusteron
TBEV-Sib	617	117 (19.0)	500 (81.0)	70 (46.4)	13	18	3
TBEV-Eu	191	25 (13.1)	166 (86.9)	28 (18.5)	9	10	2*
TBEV-FE	101	24 (23.8)	77 (76.2)	31 (20.5)	10	11	2*
Total	909	166 (18.3)	743 (81.7)	–	–	39	–

* In order to adequately represent the clusteron structure of TBEV-FE and TBEV-Eu the minimum number of strains in the clusteron has been reduced to two due to the low number of sequences.

amino acid signature. Single strains or groups of only two strains with the same amino acid sequence were named as *unique*. The number of these strains was 117 (19.0%). Variability occurred in 70 out of the 151 aa of the glycoprotein E fragment but only thirteen of them were clusteron-specific (forming specific amino acid signatures) (Table 1). Thus, TBEV-Sib could be described as a set of clusterons differing in quantitative and qualitative composition i.e. could be characterized by clusteron structure.

The number of synonymous substitutions (dS) for sequences within one clusteron can differ significantly (from 1 to 135 nucleotides in the studied gene E fragment). However, the dS is not directly related to the number of strains in the clusteron e.g. 3A, 3D and 3B (Table 1) that could indicate the time of divergence between strains.

The number of nucleotide sequences of TBEV-Eu and TBEV-FE strains deposited in GenBank is significantly less than TBEV-Sib; however, a clusteron structure for them was also shown (Supplementary Tables S2 and S3). Characteristics of all three TBEV subtypes clusterons are given in Table 2.

3.2. Phylogenetic analysis

Phylogenetic analysis of nucleotide sequences of all TBEV-Sib strains revealed four phylogenetic lineages named according to their geographical distribution, namely, Asian (prototype strain

Zausaev, AF527415), South Siberian (Aina, AF091006), Eastern European (Baltic) (Est54, DQ393773) and Buryat-Mongolian (886-84, EF469662) (Fig. 2). Asian lineage includes clusterons 3A, 3C, 3E, 3F, 3I, 3K, 3L, 3M and 3N, South Siberian, 3H and 3J, Eastern European, 3B, 3D, 3G, 3O, 3P and 3Q, and finally, the Buryat-Mongolian lineage includes only one clusteron 3U, consisting of strains having transient characteristics with TBEV-FE (Figs. 1 and 2, Supplementary Table S1). As all strains of one group of clusterons generally belong to one phylogenetic lineage and are closely related (Fig. 1), the names of phylogenetic lineages can be attributed to the groups of clusterons as well.

However, there is a rare exception when one clusteron may contain strains originating from different phylogenetic lineages. This phenomenon has been observed exclusively for the Asian group of clusterons. Thus, 15 (5.3%) strains of clusteron 3A, 1 (6.3%) of clusteron 3C, and 1 (2.6%) of clusteron 3F phylogenetically belong to the South Siberian lineage, and 2 (12.5%) strains of clusteron 3C to the Eastern European lineage (Fig. 1). These strains were found only in Western Siberia and the Middle Urals and labeled as 3A², 3C², 3F² and 3C³ respectively (Supplementary Table S1, and Clusterons TBEV-Sib.kml).

We noted that the endemic areas of TBEV-Sib are unique both in qualitative and quantitative clusteron composition (Fig. 2). The territorial distribution of the strains in a clusteron is either by corridor or by local type (Supplementary Clusterons TBEV-Sib.kml). This

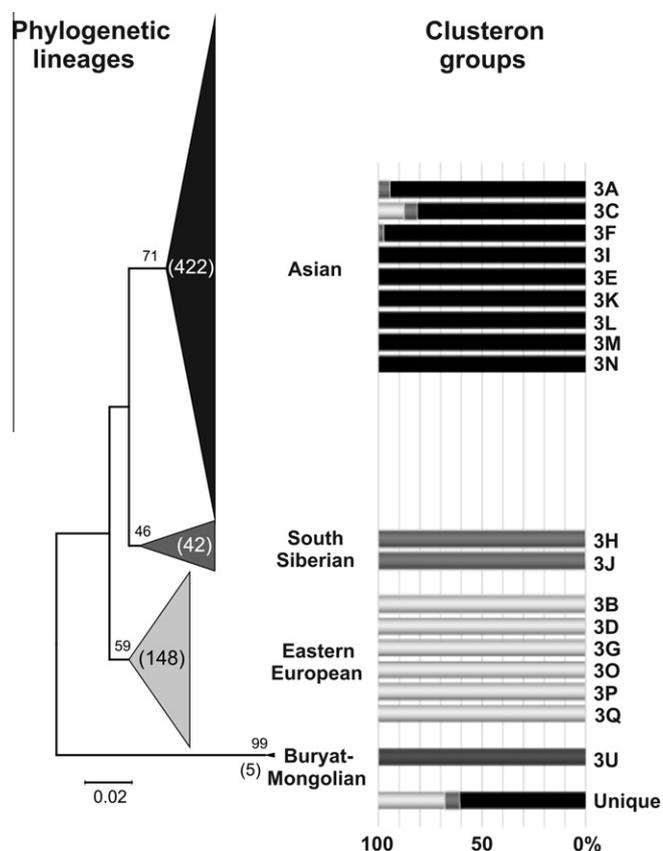


Fig. 1. Correspondence between phylogenetic lineages of TBEV-Sib and groups of clusterons. Phylogenetic lineages are shown on the left and indicated by different colors (the number of strains is shown in brackets); groups of clusterons are shown on the right and colored according to the lineages. For each clusteron as well as for unique strains the percentage of strains belonging to different phylogenetic lineages is indicated.

circumstance allows us to trace the origin and distribution of TBEV strains as has been confirmed by earlier data (Kovalev et al., 2009). Description of the different territories from the standpoint of the clusteron structure of TBEV populations is the purpose of further studies.

Thus, the clusteron is a basic element of the TBEV population structure, consisting of strains having an identical amino acid sequence of a glycoprotein E fragment, as a rule phylogeographically close, and having a certain type of territorial distribution. The clusteron structure of TBEV is determined by the quantitative and qualitative composition of the clusterons, their size and the time and degree of divergence.

For the visualization of the clusteron structure of TBEV subtypes phylogenetic networks were constructed (Fig. 3). The TBEV-Sib clusteron structure is the most complex (Fig. 3A). Differences between TBEV-Sib clusterons can reach up to six amino acid substitutions. At the same time, the clusteron structure of the TBEV-Eu and TBEV-FE is organized much more simply with the differences between clusterons no more than three or four substitutions, respectively (Fig. 3B and C). Phylogeographic features of the clusteron distribution for TBEV-Eu and TBEV-FE were not found (compared to TBEV-Sib).

3.3. Characteristics of amino acid changes in the structure of glycoprotein E fragment

Positions of the clusteron-specific amino acid substitutions for each subtype are usually different from each other (Table 1,

Supplementary Tables S2 and S3). The amino acid signature for each clusteron within the TBEV-Sib is not accidental, because of the large number of phylogeographically related strains. Clusteron-specific amino acid substitutions for TBEV-Sib are located on one lateral surface glycoprotein E (Fig. 4) that indicates their specific functional role. The effect of each of the thirteen amino acid substitutions or their combinations to the structure and properties of glycoprotein E in TBEV-Sib is a separate question that requires further research.

4. Discussion

Any classification is a grouping of objects in order to systematize the material for its ease of perception and the greater efficiency of further studies. Finding reliable discriminatory features, providing comparability and reproducibility of molecular data, allowed the development of a new approach to the differentiation of TBEV populations. Due to the high variability of the TBEV genome, structuring viral populations on the basis of amino acid sequences is the most preferred. Our studies show that TBEV-Sib strains, belonging to the same clusteron, i.e. having the same amino acid sequence of the glycoprotein E fragment, with rare exceptions, have common phylogenetic and geographic origins.

Certainly, full-genome sequences are the most informative for phylogenetic and phylogeographic studies when there are a great number of records in GenBank (hundreds, even thousands), for example, for flaviviruses such as West Nile virus (612 sequences up to date) or Dengue virus (3162 sequences). So far, there are only 59 known genome sequences of TBEV which is not enough to study the structure of virus populations and mechanisms of their origin and distribution. The way out of this situation may be the involvement of parts of genome – individual genes or gene fragment sequences (Kuno et al., 1998; Golovljova et al., 2008; Uzcatgeui et al., 2012). The greatest number of TBEV sequences registered in GenBank is that of gene E – 205 full-length sequences and more than 900 sequences of its fragment of 454 nucleotides (including complete genome and gene E sequences). The use of a TBEV genome fragment allows the direct amplification and sequencing of viral cDNA to identify the virus and to analyze its genetic properties without passaging. This minimizes genetic artefacts caused by adaptation and/or replication by RNA polymerase. It is known that TBEV can change genetic and phenotypic properties relatively quickly, within 4–8 passages, during adaptation to model systems, either cell cultures, or laboratory animals (Mandl et al., 2001; Romanova et al., 2007). That's why the “tick” sequences of their genomes have the greatest interest for studying TBEV evolution. The disadvantage of the use of a relatively small gene E fragment analysis is compensated for by a representative sampling of the strains from all TBEV areas. The reasons for choosing this genome fragment were presented earlier (Kovalev et al., 2009).

Development of the cluster approach has become possible for the following reasons related to the TBEV-Sib. (1) A large number of gene E fragment sequences have been deposited, to date, in GenBank (more than 600). (2) It has been shown that the single amino acid substitutions could serve as phylogenetic markers within this subtype. (3) We have been able to trace the non-random distribution of individual clusterons in the territory. Due to the huge areas of Western Siberia and the Urals and the low human population density, as well as the small number of transport routes linking villages in the past, it was possible to estimate the path and direction of distribution of TBEV strains, to assess the time of divergence more accurately, and to relate these processes to historical events (Kovalev et al., 2009).

The clusteron structure of TBEV can be most clearly presented in the form of phylogenetic networks which give an idea about



Fig. 2. The geographical distribution of groups of TBEV-Sib clusterons. Clusterons are designated by letters those size is proportional to the number of strains in the clusteron.

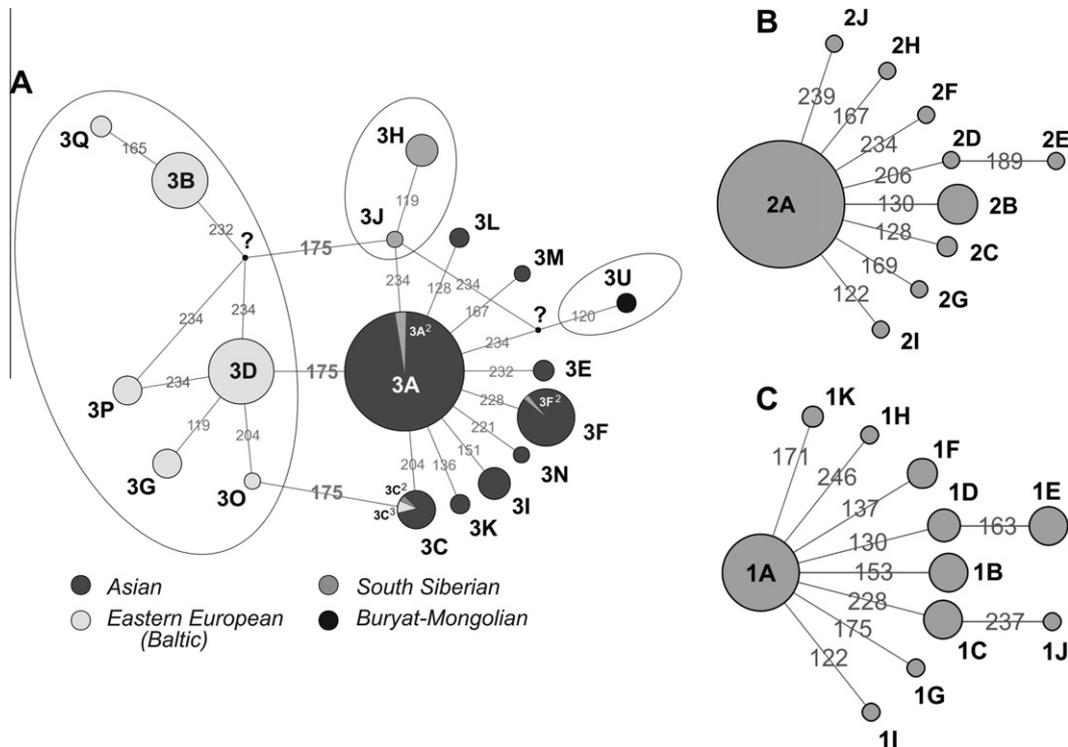


Fig. 3. Phylogenetic clusteron networks: A. TBEV-Sib, B. TBEV-Eu, C. TBEV-FE. Groups of clusterons are marked by ovals; phylogenetic lineages are designated by colors. Area of a circle is proportional to the number of strains in the clusteron. “?” – Clusterons that are not yet discovered. Positions of amino acid substitutions in the glycoprotein E fragment are placed on the lines connecting the clusterons.

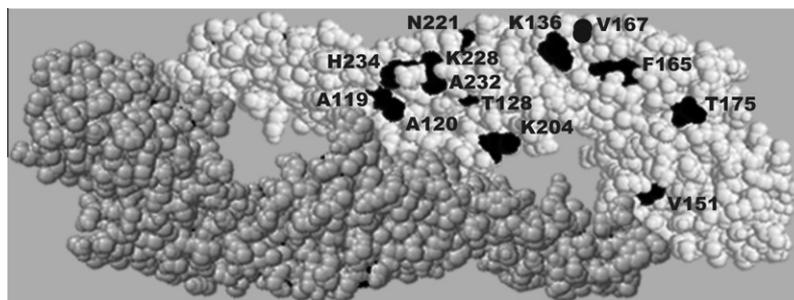


Fig. 4. 3D model of the surface glycoprotein E of TBEV (strain Zausaev AF527415). Two subunits of the dimer are colored in white and gray. The clusteron – specific amino acid positions are colored in black.

evolution of TBEV within the subtype (Fig. 3). As has been shown, the clusterons 1A, 2A and 3A are the most geographically widespread and contain the greatest number of strains. In addition, they are characterized by the largest number of dS (the number of synonymous substitution) that indicates the maximum time of evolution (Table 1, Fig. 2 and 3, Supplementary Table S2 and S3). Thus, the clusterons 1A, 2A and 3A seem to be ancestors of TBEV-FE, TBEV-Eu and TBEV-Sib, respectively.

The TBEV-Sib clusteron structure is much more complex than that of TBEV-Eu and TBEV-FE. The complex structure of TBEV-Sib can be presented in the form of two simple ones, Baltic and Siberian, differing in marker amino acid at position 175 of glycoprotein E (Fig. 3A, Table 1) and in geographical distribution (Golovljova et al., 2008). Such a structure of the Siberian subtype could be explained either by a long evolutionary history or a great variability due to the need of adaptation to different regional transmission cycles. The first scenario is supported by a recent study where the authors concluded that TBEV originated from Siberia with the subsequent formation of two subtypes in the Far East and Europe (Heinze et al., 2012). The fact that the Asian group of TBEV-Sib clusterons (3A, 3C and 3F) includes a small number of strains from other phylogenetic lineages can indicate possible convergence processes in the evolution of TBEV. It's interesting that we couldn't observe the reverse situation i.e. the presence of the strain of Asian lineage in other groups of clusterons (Figs. 1 and 3). Given the fixed direction of convergent changes (from descendant to ancestor), it's possible that the clusterons of the Asian lineage (especially 3A) could be the ancestral group of TBEV-Sib.

The causes and mechanisms of formation of the TBEV clusterons require particular attention. In fact, the key question is about the role of clusteron-specific amino acid signature fixed in the virus populations. Even if the functional importance of the amino acid substitutions is still unknown, the fact of their location on the only lateral surface of glycoprotein E (Fig. 4) indicates the possibility of their involvement in binding to cell surface receptors of the tick (Ecker et al., 1999). Thus, the geographic heterogeneity in the distribution of clusterons can be the result of virus adaptation to the species (Gritsun et al., 2003a), subspecies or local populations of ticks due to co-evolutionary processes.

Applying the common concept of viruses quasispecies (Domingo et al., 1978) to the overall TBEV population, we can conclude that the clusteron structure of the TBEV population can be considered as a spectrum of genetically diverse variants with dominant phenotypes adapted to regional transmission cycles. A wide range of genetic variation (46.4% of the variable aa positions of the studied TBEV-Sib fragment) indicates the high evolutionary potential of TBEV. As the result of such variability, a significant number of unique strains (on average 18.3% of the viral population) is observed (Table 2). However, due to their uniqueness, they are not considered as elements of the clusteron structure. Small clusterons and unique strains seem to be a form of balance between

natural selection and mutation process where these variants of TBEV (“clouds”) are ordered around the large clusterons (“masters”), e.g. 1A, 2A and 3A.

The fact that the strains within one clusteron are phylogeographically related allows to consider it as a smallest unit of TBEV classification. Introduction of a new taxonomic unit is necessary for the study of the origin, distribution and evolution of TBEV which could be carried out at the clusteron level as well as at the level of clusteron groups. In turn, the set of clusterons form a clusteron structure which was found to be the most complex for TBEV-Sib. For TBEV-Eu and TBEV-FE the clusteron structure has been also shown (Fig. 3). However, we couldn't observe a clear difference in geographical distribution between different clusterons. It could be explained by the low number of nucleotide sequences deposited in GenBank for these subtypes compared to TBEV-Sib. It should be noted that the clusteron structure of TBEV is not fixed but it could be improved with the accumulation of new data.

Any endemic TBE territory could be characterized by a unique clusteron structure i.e. the pattern of clusterons with different numbers of strains and the type of distribution in the clusteron (Supplementary, Clusterons TBEV-Sib.kml). Thus, the approach proposed is useful for the monitoring of natural TBEV populations and the creation of local and regional registers of TBEV strains. Ultimately, it would contribute to the establishment of the time and driving forces of TBEV distribution. That will allow determination of the actual time of the emergence of strains in a given territory i.e. to calculate objectively the age of populations and the rate of TBEV evolution. Any statistically significant changes in the clusterons pattern would allow prediction of the alteration of an epidemiological situation and the assessment of trends in the regional characteristics of TBE clinical manifestations, as well as the development of tactics of specific and nonspecific prevention of this infection.

Acknowledgments

We are indebted to S.I. Ibragimov, N.V. Loginovskikh, T.A. Pime-nova, T.E. Snitkovskaya, B.A. Galishev, V.L. Umpelev, N.K. Tokarevich and V.S. Kokorev for the collection of material, and to Dr Keith Chamberlain (Rothamsted Research, UK) for his help in preparing the manuscript. The reported study was partially supported by RFBR, research projects No. 12-04-31263 and No. 13-04-00473.

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.meegid.2012.10.011>. These data include Google maps of the most important areas described in this article.

References

- Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Domingo, E., Sabo, D., Taniguchi, T., Weissmann, C., 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13, 735–744.
- 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.
- Golovljova, I., Katargina, O., Geller, J., Tallo, T., Mittzenkov, V., Vene, S., Nemirov, K., Kutsenko, A., Kilosanidze, G., Vasilenko, V., Plyusnin, A., Lundkvist, A., 2008. Unique signature amino acid substitution in Baltic tick-borne encephalitis virus (TBEV) strains within the Siberian TBEV subtype. *Int. J. Med. Microbiol.* 298, 108–120.
- Gould, E.A., de Lamballerie, X., Zanotto, P.M., Holmes, E.C., 2001. Evolution, epidemiology, and dispersal of flaviviruses revealed by molecular phylogenies. *Adv. Virus Res.* 57, 71–103.
- 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. *Antiviral Res.* 57, 129–146.
- Hayasaka, D., Gritsun, T.S., Yoshii, K., Ueki, T., Goto, A., Mizutani, T., Kariwa, H., Iwasaki, T., Gould, E.A., Takashima, I., 2004. Amino acid changes responsible for attenuation of virus neurovirulence in an infectious cDNA clone of the Oshima strain of tick-borne encephalitis virus. *J. Gen. Virol.* 85, 1007–1018.
- 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.
- Hayasaka, D., Suzuki, Y., Kariwa, H., Ivanov, L., Volkov, V., Demenev, V., Mizutani, T., Gajobori, T., Takashima, I., 1999. Phylogenetic and virulence analysis of tick-borne encephalitis viruses from Japan and far-Eastern Russia. *J. Gen. Virol.* 80 (Pt 12), 3127–3135.
- Heinze, D.M., Gould, E.A., Forrester, N.L., 2012. Revisiting the clinal concept of evolution and dispersal for the tick-borne flaviviruses by using phylogenetic and biogeographic analyses. *J. Virol.* 86, 8663–8671.
- Karan, L.S., Pogodina, V.V., Frolova, T.V., Platonov, A.Y., 2006. Genetic diversity of East European and Asian strains of tick-borne encephalitis virus belonging to Siberian genotype. *Bull. Siberian Med. Suppl.* 1, 24–28 (In Russian).
- Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.
- King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., 2011. *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego, p. 1338.
- Kovalev, S.Y., Chernykh, D.N., Kokorev, V.S., Snitkovskaya, T.E., Romanenko, V.V., 2009. Origin and distribution of tick-borne encephalitis virus strains of the Siberian subtype in the Middle Urals, the north-west of Russia and the Baltic countries. *J. Gen. Virol.* 90, 2884–2892.
- Kovalev, S.Y., Kokorev, V.S., Belyaeva, I.V., 2010. Distribution of Far-Eastern tick-borne encephalitis virus subtype strains in the former Soviet Union. *J. Gen. Virol.* 91, 2941–2946.
- Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N., Cropp, C.B., 1998. Phylogeny of the genus *Flavivirus*. *J. Virol.* 72, 73–83.
- Lambert, C., Leonard, N., De Bolle, X., Depiereux, E., 2002. ESyPred3D: prediction of proteins 3D structures. *Bioinformatics* 18, 1250–1256.
- Mandl, C.W., Kroschewski, H., Allison, S.L., Kofler, R., Holzmann, H., Meixner, T., Heinz, F.X., 2001. Adaptation of tick-borne encephalitis virus to BHK-21 cells results in the formation of multiple heparan sulfate binding sites in the envelope protein and attenuation in vivo. *J. Virol.* 75, 5627–5637.
- Pletnev, A.G., Yamshchikov, V.F., Blinov, V.M., 1990. Nucleotide sequence of the genome and complete amino acid sequence of the polyprotein of tick-borne encephalitis virus. *Virology* 174, 250–263.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291–298.
- Roehrig, J.T., 2003. Antigenic structure of flavivirus proteins. *Adv. Virus Res.* 59, 141–175.
- Romanova, L., Gmyl, A.P., Dzhivanian, T.I., Bakhmutov, D.V., Lukashev, A.N., Gmyl, L.V., Rummyantsev, A.A., Burenkova, L.A., Lashkevich, V.A., Karganova, G.G., 2007. Microevolution of tick-borne encephalitis virus in course of host alternation. *Virology* 362, 75–84.
- Sayle, R.A., Milner-White, E.J., 1995. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374.
- Suss, 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.
- 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.
- Ternovoi, V.A., Kurzhukov, G.P., Sokolov, Y.V., Ivanov, G.Y., Ivanisenko, V.A., Loktev, A.V., Ryder, R.W., Netesov, S.V., Loktev, V.B., 2003. Tick-borne encephalitis with hemorrhagic syndrome, Novosibirsk region, Russia, 1999. *Emerg. Infect. Dis.* 9, 743–746.
- Uzcategui, N.Y., Sironen, T., Golovljova, I., Jaaskelainen, A.E., Valimaa, H., Lundkvist, A., Plyusnin, A., Vaheri, A., Vapalahti, O., 2012. Rate of evolution and molecular epidemiology of tick-borne encephalitis virus in Europe, including two isolations from the same focus 44 years apart. *J. Gen. Virol.* 93, 786–796.