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

A R T I C L E   I N F O

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

A B S T R A C T

Tick-borne encephalitis is a natural focal transmissible zoonanthroponosis. 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 zoonanthroponosis. 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 homodimer, 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
strains based on the presence of amino acids His, Gln, or Tyr at position 313 of 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 Eastern part of Russia, and Siberian, or Russian, strains 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 these 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 territories 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 such. Molecular phylogenetic analysis was inferred by using the Maximum Likelihood method 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).

2.2. 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/esypred/) (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-TBEV-Sib), the second is a letter attributed to a specific
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.

### Table 1
Characteristics of the clusterons of TBEV-Sib.

<table>
<thead>
<tr>
<th>Clusteron</th>
<th>Prototype strain</th>
<th>GenBank Acc. No.</th>
<th>Number of strains (%)</th>
<th>dS&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Clusteron-specific amino acid signature&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>3A</td>
<td>Zausaev</td>
<td>AF527415</td>
<td>285 (46.2)</td>
<td>135&lt;sup&gt;c&lt;/sup&gt;</td>
<td>A     A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3B</td>
<td>Elb123-2007</td>
<td>GU444253</td>
<td>35 (5.8)</td>
<td>6</td>
<td>A</td>
</tr>
<tr>
<td>3C</td>
<td>76</td>
<td>EF568817</td>
<td>16 (2.6)</td>
<td>23</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3D</td>
<td>Est54</td>
<td>DQ393773</td>
<td>50 (8.1)</td>
<td>68</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3E</td>
<td>Elb361-2008</td>
<td>GU444279</td>
<td>5 (0.8)</td>
<td>14</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3F</td>
<td>Elb51-2007</td>
<td>GU444225</td>
<td>39 (6.3)</td>
<td>48</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3G</td>
<td>Kokkola-102</td>
<td>DQ451295</td>
<td>9 (1.5)</td>
<td>12 V</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3H</td>
<td>Aina</td>
<td>AF091006</td>
<td>13 (2.1)</td>
<td>33 V</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3I</td>
<td>Elb669-2006</td>
<td>GU444220</td>
<td>11 (1.9)</td>
<td>3</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3J</td>
<td>IR99-1m1</td>
<td>AB049348</td>
<td>3 (0.5)</td>
<td>20</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3K</td>
<td>1130</td>
<td>BI442275</td>
<td>4 (0.6)</td>
<td>18</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3L</td>
<td>1699</td>
<td>BI442268</td>
<td>4 (0.6)</td>
<td>24</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3M</td>
<td>24</td>
<td>GU141822</td>
<td>3 (0.5)</td>
<td>16</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3N</td>
<td>IR99-ln4</td>
<td>AB049349</td>
<td>3 (0.5)</td>
<td>19</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3O</td>
<td>Elb90-1-2007</td>
<td>GU444224</td>
<td>3 (0.5)</td>
<td>1</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3P</td>
<td>Vologda-4-06</td>
<td>FJ214139</td>
<td>8 (1.1)</td>
<td>15</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3Q</td>
<td>Elb752-2005</td>
<td>JX315727</td>
<td>5 (0.8)</td>
<td>3</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3U</td>
<td>886-84</td>
<td>BI469692</td>
<td>4 (0.6)</td>
<td>6</td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td>3Unique</td>
<td></td>
<td>117 (19.0)</td>
<td></td>
<td></td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>617</td>
<td></td>
<td></td>
<td>A       A    T   K   V   F   V   T   R   N   K    A    H</td>
</tr>
</tbody>
</table>

<sup>a</sup> Number of synonymous substitutions.
<sup>b</sup> Position numbers are given for glycoprotein E.
<sup>c</sup> Only strains of the Asian lineage (see below).

### Table 2
Characteristics of TBEV clusteron structure.

<table>
<thead>
<tr>
<th>TBEV subtype</th>
<th>Number of strains</th>
<th>Number of unique strains (%)</th>
<th>Number of strains in clusterons (%)</th>
<th>Number of all aa variable positions (%)</th>
<th>Number of clusteron-specific substitutions</th>
<th>Number of clusterons</th>
<th>Minimum number of strains in the clusteron</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBEV-Sib</td>
<td>617</td>
<td>117 (19.0)</td>
<td>500 (81.0)</td>
<td>70 (46.4)</td>
<td>13</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>TBEV-Eu</td>
<td>191</td>
<td>25 (13.1)</td>
<td>166 (86.9)</td>
<td>28 (18.5)</td>
<td>9</td>
<td>10</td>
<td>2&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>TBEV-FE</td>
<td>101</td>
<td>24 (23.8)</td>
<td>77 (76.2)</td>
<td>31 (20.5)</td>
<td>10</td>
<td>11</td>
<td>2&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>909</strong></td>
<td><strong>166 (18.3)</strong></td>
<td><strong>743 (81.7)</strong></td>
<td><strong>–</strong></td>
<td><strong>–</strong></td>
<td><strong>39</strong></td>
<td><strong>–</strong></td>
</tr>
</tbody>
</table>

<sup>c</sup> 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.

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, 3C, 3E, 3I, 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.

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).
Phylogenetic lineages were constructed (Fig. 3). The TBEV-Sib and TBE-FE are organized much more simply with the differences between TBEV-Sib clusterons no more than three or four substitutions, respectively (Fig. 3B and C). Phylogeographic features of the clusteron distribution for TBEV-Eu and TBE-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; Uzcategui 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 whose 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.
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. Pimenova, T.E. Snihtovskaya, 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


