Sequence analysis and genetic classification of tick-borne encephalitis viruses from Europe and Asia

Michael Ecker, Steven L. Allison, Tamara Meixner and Franz X. Heinz
Institute of Virology, University of Vienna, Kinderspitalgasse 15, A-1095 Vienna, Austria

The epidemiology of tick-borne encephalitis virus was investigated by comparative sequence analysis of virus strains isolated in endemic areas of Europe and Asia. Phylogenetic relationships were determined from the nucleotide and amino acid sequences of the major envelope (E) protein of 16 newly sequenced strains and nine previously published sequences. Three genetic lineages could be clearly distinguished, corresponding to a European, a Far Eastern and a Siberian subtype. Amino acids characteristic for each of the subtypes (‘signature’ amino acids) were identified and their location in the atomic structure of protein E was determined. The degree of variation between strains within subtypes was low and exhibited a maximum of only 2.2% at the amino acid level. A maximum difference of 5.6% was found between the three subtypes, which is in the range of variation reported for other flaviviruses.

Introduction

Tick-borne encephalitis (TBE) virus is a human-pathogenic member of the genus Flavivirus within the family Flaviviridae. It is endemic in many European countries, the Asian parts of Russia, northern China and northern Japan (Monath & Heinz, 1996; Takashima et al., 1997; Hou et al., 1997), and has a significant impact on public health in certain geographical regions.

Based on serological analyses of isolates from Europe and the Far East, TBE virus has been subdivided into two closely related subtypes, a European subtype, primarily transmitted by Ixodes ricinus, and a Far Eastern subtype, with Ixodes persulcatus as its main vector (Clarke, 1964). This distinction, as well as the close relationship of the two subtypes, was later confirmed by sequence analysis of a European and a Far Eastern prototype strain (Mandl et al., 1988, 1989; Pletnev et al., 1986, 1990). Consistent with their close antigenic relationship, immunization studies in mice have shown a high degree of cross-protection between European and Far Eastern subtype strains (Holzmann et al., 1992; Vorobyova et al., 1996). The disease caused by these viruses (also referred to as central European encephalitis and Russian spring–summer encephalitis) has been reported to be more severe in Asia than in Europe, with case fatality rates of up to 30% and 1–2%, respectively. This may be due to virulence differences between the subtypes, but other factors such as differences in hospitalization rates and recording of mild cases have not yet been ruled out. Vaccination with a formalin-inactivated whole virus vaccine is an effective means of protection from disease (Kunz, 1996).

A genetically close relative of the European TBE virus subtype is louping ill virus, which is also transmitted by I. ricinus, but has a different ecological cycle. It causes disease primarily in sheep and red grouse, and is prevalent in the upland sheep grazing areas of Scotland, northern England, Wales, southwest England and Ireland (McGuire et al., 1998; Gould et al., 1997).

Studies on the molecular epidemiology of TBE virus have focused on the major envelope glycoprotein (E). This protein is responsible for essential functions during virus entry, including receptor binding and membrane fusion (Rice et al., 1996), and thus induces neutralizing antibodies and protective immunity (Heinz & Roehrig, 1980). Its three-dimensional structure has been determined by X-ray crystallography (Rey et al., 1995).

Flaviviruses, like RNA viruses in general, exhibit a high mutation frequency and thus have the potential for rapid change and evolution (Domingo et al., 1997). The available data on European TBE virus isolates, however, indicate that this virus is remarkably stable under natural ecological...
条件（Heinz & Kunz, 1981, 1982; Guirakhoo et al., 1987）和没有主要抗原变异性。

因此，E蛋白基因序列的9个菌株（表1）在欧洲和亚洲被确定。为了获得更多信息，我们确定了E蛋白基因序列的16个菌株（表1）在欧洲，俄罗斯的欧洲部分和亚洲的东部地区（包括中央西伯利亚和远东地区）以及分析了E蛋白基因序列与不同菌株之间的relationshp together with the strains for which the E protein gene sequence is already available. The sites of isolation of the different strains included in this study (Table 1) and the geographical distribution of Ixodes vectors are depicted in Fig. 1.

Methods

- Stocks of the viruses to be sequenced were prepared as 20% (w/v) suspensions of infected suckling mouse brain in medium 199 with Hanks’ salts, pH 7.6, containing 15 mM HEPES, 15 mM HEPPS and 1% neomycin. Viral RNA was prepared directly from virus stocks by proteinase K digestion, phenol–chloroform extraction and subsequent ethanol precipitation as described previously (Mandl et al., 1988). cDNA synthesis, PCR and sequence analysis were performed exactly as described by Mandl et al. (1997, 1998). The nucleic acid sequence of the 1488 nucleotide long E gene of each of the 16 newly sequenced strains was deposited in the GenBank database.

To determine the genetic relatedness among TBE virus isolates, phylogenetic trees were constructed from the aligned nucleic acid sequences (multiple sequence alignment program of the Microgenie 4.0 package, Beckmann) using the distance (DNADIST/FITCH) and parsimony (DNAPARS) programs from version 3.52c of PHYLIP (Felsenstein, 1993). The reliability of the trees was tested by bootstrap resampling analysis using the SEQBOOT program (Felsenstein, 1993).

Results and Discussion

The two different methods for generating phylogenetic trees yielded trees with similar topology, three of which revealed three distinct genetic lineages (Fig. 2) that can be designated as subtypes. Subtype 1 (European subtype) includes the European prototype strain Neudoerfl and other strains from Austria, Switzerland, France, Germany, Hungary, Czech Republic, Slovenia, Croatia, Finland, Belarus and the European part of Russia. These results are in agreement with earlier reports in which protein-based methods assigned strains Absettarov (Chumakov et al., 1984) and N256 (Tscherkhanovskaya et al., 1993) to the European subtype. Subtype 2 (Far Eastern subtype) consists of the Far Eastern prototype strain Sofin, other strains from the Far East (East Russia, China, Japan and Korea) and is not subject to major antigenic variation.

Subtype 3 (Far Eastern subtype) consists of the Far Eastern prototype strain Sofin, other strains from the Far East (East Russia, China, Japan and Korea) and is not subject to major antigenic variation. Therefore, the E protein gene sequences of nine strains (Table 1) and the geographical distribution of the different strains included in this study (Table 1) and the geographical distribution of Ixodes vectors are depicted in Fig. 1.

Methods

- Stocks of the viruses to be sequenced were prepared as 20% (w/v) suspensions of infected suckling mouse brain in medium 199 with Hanks’ salts, pH 7.6, containing 15 mM HEPES, 15 mM HEPPS and 1% neomycin. Viral RNA was prepared directly from virus stocks by proteinase K digestion, phenol–chloroform extraction and subsequent ethanol precipitation as described previously (Mandl et al., 1988). cDNA synthesis, PCR and sequence analysis were performed exactly as described by Mandl et al. (1997, 1998). The nucleic acid sequence of the 1488 nucleotide long E gene of each of the 16 newly sequenced strains was deposited in the GenBank database.

To determine the genetic relatedness among TBE virus isolates, phylogenetic trees were constructed from the aligned nucleic acid sequences (multiple sequence alignment program of the Microgenie 4.0 package, Beckmann) using the distance (DNADIST/FITCH) and parsimony (DNAPARS) programs from version 3.52c of PHYLIP (Felsenstein, 1993). The reliability of the trees was tested by bootstrap resampling analysis using the SEQBOOT program (Felsenstein, 1993).

Results and Discussion

The two different methods for generating phylogenetic trees yielded trees with similar topology, three of which revealed three distinct genetic lineages (Fig. 2) that can be designated as subtypes. Subtype 1 (European subtype) includes the European prototype strain Neudoerfl and other strains from Austria, Switzerland, France, Germany, Hungary, Czech Republic, Slovenia, Croatia, Finland, Belarus and the European part of Russia. These results are in agreement with earlier reports in which protein-based methods assigned strains Absettarov (Chumakov et al., 1984) and N256 (Tscherkhanovskaya et al., 1993) to the European subtype. Subtype 2 (Far Eastern subtype) consists of the Far Eastern prototype strain Sofin, other strains from the Far East (East Russia, China, Japan and Korea) and is not subject to major antigenic variation. Therefore, the E protein gene sequences of nine strains (Table 1) and the geographical distribution of the different strains included in this study (Table 1) and the geographical distribution of Ixodes vectors are depicted in Fig. 1.

Methods

- Stocks of the viruses to be sequenced were prepared as 20% (w/v) suspensions of infected suckling mouse brain in medium 199 with Hanks’ salts, pH 7.6, containing 15 mM HEPES, 15 mM HEPPS and 1% neomycin. Viral RNA was prepared directly from virus stocks by proteinase K digestion, phenol–chloroform extraction and subsequent ethanol precipitation as described previously (Mandl et al., 1988). cDNA synthesis, PCR and sequence analysis were performed exactly as described by Mandl et al. (1997, 1998). The nucleic acid sequence of the 1488 nucleotide long E gene of each of the 16 newly sequenced strains was deposited in the GenBank database.

To determine the genetic relatedness among TBE virus isolates, phylogenetic trees were constructed from the aligned nucleic acid sequences (multiple sequence alignment program of the Microgenie 4.0 package, Beckmann) using the distance (DNADIST/FITCH) and parsimony (DNAPARS) programs from version 3.52c of PHYLIP (Felsenstein, 1993). The reliability of the trees was tested by bootstrap resampling analysis using the SEQBOOT program (Felsenstein, 1993).

Results and Discussion

The two different methods for generating phylogenetic trees yielded trees with similar topology, three of which revealed three distinct genetic lineages (Fig. 2) that can be designated as subtypes. Subtype 1 (European subtype) includes the European prototype strain Neudoerfl and other strains from Austria, Switzerland, France, Germany, Hungary, Czech Republic, Slovenia, Croatia, Finland, Belarus and the European part of Russia. These results are in agreement with earlier reports in which protein-based methods assigned strains Absettarov (Chumakov et al., 1984) and N256 (Tscherkhanovskaya et al., 1993) to the European subtype. Subtype 2 (Far Eastern subtype) consists of the Far Eastern prototype strain Sofin, other strains from the Far East (East Russia, China, Japan and Korea) and is not subject to major antigenic variation. Therefore, the E protein gene sequences of nine strains (Table 1) and the geographical distribution of the different strains included in this study (Table 1) and the geographical distribution of Ixodes vectors are depicted in Fig. 1.

Methods

- Stocks of the viruses to be sequenced were prepared as 20% (w/v) suspensions of infected suckling mouse brain in medium 199 with Hanks’ salts, pH 7.6, containing 15 mM HEPES, 15 mM HEPPS and 1% neomycin. Viral RNA was prepared directly from virus stocks by proteinase K digestion, phenol–chloroform extraction and subsequent ethanol precipitation as described previously (Mandl et al., 1988). cDNA synthesis, PCR and sequence analysis were performed exactly as described by Mandl et al. (1997, 1998). The nucleic acid sequence of the 1488 nucleotide long E gene of each of the 16 newly sequenced strains was deposited in the GenBank database.

To determine the genetic relatedness among TBE virus isolates, phylogenetic trees were constructed from the aligned nucleic acid sequences (multiple sequence alignment program of the Microgenie 4.0 package, Beckmann) using the distance (DNADIST/FITCH) and parsimony (DNAPARS) programs from version 3.52c of PHYLIP (Felsenstein, 1993). The reliability of the trees was tested by bootstrap resampling analysis using the SEQBOOT program (Felsenstein, 1993).

Results and Discussion

The two different methods for generating phylogenetic trees yielded trees with similar topology, three of which revealed three distinct genetic lineages (Fig. 2) that can be designated as subtypes. Subtype 1 (European subtype) includes the European prototype strain Neudoerfl and other strains from Austria, Switzerland, France, Germany, Hungary, Czech Republic, Slovenia, Croatia, Finland, Belarus and the European part of Russia. These results are in agreement with earlier reports in which protein-based methods assigned strains Absettarov (Chumakov et al., 1984) and N256 (Tscherkhanovskaya et al., 1993) to the European subtype. Subtype 2 (Far Eastern subtype) consists of the Far Eastern prototype strain Sofin, other strains from the Far East (East Russia, China, Japan and Korea) and is not subject to major antigenic variation.
Genetic classification of TBE viruses

Fig. 1. Geographical distribution of the TBE virus isolates investigated in this study. The distribution of the major tick vectors is indicated by dotted lines (Ixodes ricinus) and broken lines (Ixodes persulcatus) (adapted from Monath & Heinz, 1996).

Fig. 2. Phylogenetic tree illustrating the genetic relationship of TBE virus isolates (generated by the DNADIST/FITCH program). The tree was rooted using Omsk haemorrhagic fever (OHF) virus as the outgroup virus. The scale bar corresponds to expected number of substitutions per nucleotide site.

Japan), as well as strains from Latvia, Ukraine and the European part of Russia. The new subtype 3, which is slightly more closely related to subtype 2 than to subtype 1, includes the central Siberian strains Aina and Vasilenko (Gritsun et al., 1993, 1997). Partitioning of these subtypes (Fig. 2) was supported by bootstrap resampling analysis (100 and 500 replicate data sets) with a value greater than 86% (> 80% is highly significant; Zharkikh et al., 1992) and by generating phylogenetic trees from translated amino acid sequences (data not shown). Partial sequence data of numerous other isolates from the European and Asian part of Russia have also revealed three distinct genetic lineages of TBE virus (V. Zlobin, personal communication).

The degree of variation within the subtypes is very low. The European subtype viruses display a maximum of 2-2.2% amino acid divergence, which is consistent with the previously described antigenic homogeneity of TBE viruses within Europe (Heinz & Kunz, 1981, 1982; Guirakhoo et al., 1987). This also holds true for strain Pan, which appears distinct at the nucleotide level (Fig. 2), but it is nearly identical to the other strains at the amino acid level. It is striking that isolates from Belarus (N256), St Petersburg (Absettarov) and A52 from Kumlinge Island (located between Sweden and Finland) differ from the Austrian prototype strain Neudoerfl by only a single
Fig. 3. For legend see facing page.
amino acid in the E protein. It is also remarkable that strains Absettarov (St Petersburg) and Ljub.I (Slovenia) differ only by two amino acids although they were isolated 42 years apart.

Similar to the European subtype viruses, the Far Eastern subtype viruses also differ by no more than 2-2% at the amino acid level, although they are geographically even more widespread. Surprisingly, some isolates from Europe (RK1424, Crimea) and the European part of Russia (T-blood) cluster with the Far Eastern strains. This relationship has also been revealed by comparative sequence analysis of the NS5 protein gene and the 3' non-coding region (Wallner et al., 1995) and may be linked to the association with the vector I. persulcatus in the case of strains RK1424 and T-blood (Fig. 1). Strain Crimea, however, is a major exception. It is closely related to the Far Eastern strains but was isolated in Ukraine from I. ricinus, which otherwise was always associated with European subtype viruses. Further investigations will be necessary to clarify this issue.

The third distinct genetic lineage formed by the two central Siberian strains Vasilchenko and Aina differs from European subtype isolates by 3-6-5-6% and from Far Eastern subtype isolates by 3-8-5-6% at the amino acid level, which is similar to the differences between European and Far Eastern subtype strains (3-8-5-6%). It is therefore justified to classify these strains as members of a third subtype. This is in agreement with previously published reports in which antibody-binding experiments (Rubin & Chumakov, 1980; Pomelova et al., 1992) and sequence analysis of the 3' non-coding region (Wallner et al., 1995) demonstrated antigenic and genetic differences between strain Aina and both of the previously established subtypes.

In general, the degree of amino acid variability was low (0-5-6%) compared to that at the nucleotide level (1-16-9%), suggesting that selection pressure during evolution has favoured conservation of the E protein sequence. A similar pattern has also been observed in the analysis of yellow fever virus strains (Wang et al., 1997). Alignment of all available E gene sequences from TBE virus isolates revealed that 1027 of the nucleotide positions in the E gene (69%) are absolutely conserved, whereas 461 (31%) contain at least one nucleotide difference. All amino acid differences in comparison to the European prototype strain Neudoerfl are depicted in Fig. 3. As expected, the differences did not affect functionally important sequence elements in the E protein. Also, a pentapeptide at amino acid positions 320-324, which has been described as unique to all tick-borne flaviviruses (Gao et al., 1993), was absolutely conserved among all newly sequenced TBE virus strains. On the other hand, four strains (Vasilchenko, Aina, Hypr, Ljub.I) exhibit an amino acid change within the type-specific hypervariable region at amino acids 232-234 that has been proposed as a genetic marker for distinguishing between individual flaviviruses (Shiu et al., 1992).

Amino acid variations were found at 60 different positions, 33 of which were present in more than one strain. Some of these repeated amino acid differences match the positions of the 'serological' amino acids described previously (Gritsun et al., 1995). Several of these are characteristic of the European subtype, the Far Eastern subtype and the Siberian subtype, and thus represent 'signatures' for each of the subtypes. These are depicted in Fig. 3 and Fig. 4 (E, F and S, respectively). The only position where each of the subtypes contained a characteristically unique amino acid was position 206 (Fig. 3). This amino acid is positioned immediately before the hexapeptide EHLPTA (amino acids 207-211), which is part of a loop (Fig. 4) that is longer in tick-borne flaviviruses than in mosquito-borne flaviviruses (Shiu et al., 1991). On the basis of amino acid 206 and the other 'signature' amino acids, TBE virus subtypes are easily distinguishable from each other, and a tentative assignment of new TBE viruses to one of these subtypes should be possible by determination of only small parts of the E gene.

Fig. 4 displays the location of all of the variable amino acid
positions of the TBE virus subtypes listed in Fig. 3 in the three-dimensional structure of protein E of strain Neudoerfl (Rey et al., 1995). Most of these positions are exposed on the upper or lateral surface, although some face the membrane, and only three positions (39, 115 and 363) are completely buried inside the protein. Inspection of the location of the ‘signature’ amino acids in the structure does not reveal direct interaction between these residues. Only one amino acid change (residue 260) was at a position that previously appeared to be absolutely conserved among flaviviruses (Zanotto et al., 1996).

These results show a clear separation of TBE viruses into three genetic lineages, designated subtypes. Based on the ‘signature’ amino acids we found markers characteristic for each of the subtypes. Within the subtypes TBE virus exhibits a high degree of stability with a divergence of only 2-2% at the amino acid level. Even the maximum difference between members of different subtypes (5-6%) is still within the range of variation found with mosquito-borne flaviviruses, which has been reported to be 4-2% for Japanese encephalitis (Ni et al., 1995; Nam et al., 1996), 5-2% for yellow fever (Lepiniec et al., 1994), 3% for dengue type 1 (Chu et al., 1989), 10% for dengue type 2 (Lewis et al., 1993), 5% for dengue type 3 (Lanciotti et al., 1997) and 4% for dengue type 4 viruses (Lanciotti et al., 1994). Therefore it is justifiable to classify TBE virus as a single virus species. The significance of sequence variation within and between subtypes with respect to biological properties, such as virulence and vector competence, remains to be investigated.

The authors are grateful to Dr M. Vorobyova for providing us with virus isolates and Dr V. Zlobin for sharing unpublished data. We thank Walter Holzer and Angela Dohnal for their excellent technical assistance.

References


Mandi, C. W., Heinz, F. X., Stöckl, E. & Kunz, C. (1989). Genome sequence of tick-borne encephalitis virus (Western subtype) and


Received 28 July 1998; Accepted 27 August 1998