Evolutionary traits of Tick-borne encephalitis virus: Pervasive non-coding RNA structure conservation and molecular epidemiology

Lena S. Kutschera[1] and Michael T. Wolfinger[1,2,*]

[1]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria
[2]Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währingerstraße 29, 1090 Vienna, Austria

Abstract

Tick-borne encephalitis virus (TBEV) is the etiological agent of tick-borne encephalitis, an infectious disease of the central nervous system that is often associated with severe sequelae in humans. While TBEV is typically classified into three subtypes, recent evidence suggests a more varied range of TBEV subtypes and lineages that differ substantially in their 3'UTR architecture. Building on comparative genomics approaches and thermodynamic modelling, we characterize the TBEV 3'UTR structureome diversity and propose a unified picture of pervasive non-coding RNA (ncRNA) structure conservation. Moreover, we provide an updated phylogeny of TBEV, building on more than 220 publicly available complete genomes, and investigate the molecular epidemiology and phylodynamics with Nextstrain, a web-based visualization framework for real-time pathogen evolution.

1 Introduction

Tick-borne encephalitis virus (TBEV) is a zoonotic RNA virus in the genus Flavivirus, family *Flaviviridae*. It is the etiological agent of tick-borne encephalitis (TBE), an infection of the central nervous system that is considered the most common tick-transmitted disease in Eurasia [1], where it occurs in risk or endemic areas that are also referred to as foci [2]. TBEV is transmitted between haematophagous ticks as vectors and vertebrate hosts. Typical reservoir hosts include wild-living animals such as small rodent species. Large vertebrate species, including humans and ungulates like goats, cows, sheep, swine, and deer can become infected but appear not to be competent of transmitting the virus back to ticks [3]. While serological evidence suggest that the majority of human infections is either asymptomatic or subclinical, TBEV is a neurotropic virus that can cause a wide range of life-threatening clinical manifestations comprising febrile, meningeal, meningoencephalitic, poliomyelitic, polyradiculoneuritic and chronic forms (reviewed in [4]), as well as hemorrhagic syndrome [5].

1.1 Flavivirus genome organization

TBEV belongs to the group of tick-borne flaviviruses (TBFVs), which, together with mosquito-borne flaviviruses (MBFVs) and no-known-vector flaviviruses (NKVs) encompass the vertebrate-infecting flaviviruses. Contrary, insect-specific flaviviruses (ISFVs) only replicate in mosquitoes [6]. Flaviviruses are enveloped, single stranded (+)-sense viruses that contain a non-segmented, 5'-capped, non-polyadenylated RNA of approximately 11kB length. The genomic RNA (gRNA) encodes a single open reading frame (ORF) that is flanked by highly structured untranslated regions (UTRs) of variable length [7, 8]. Both UTRs are crucially involved in regulating processes that control different aspects of the virus life cycle, including virus replication, genome cyclisation and packaging, and immune response [9, 10, 11].

*To whom correspondence should be addressed: michael.wolfinger@univie.ac.at

Common architectural traits of flavivirus 3'UTRs comprise autonomous RNA structure formation of distinct domains and the presence of evolutionarily conserved RNA elements with specific functional associations. A hallmark of flavivirus biology is their ability to actively dysregulate the host mRNA turnover machinery by stalling endogenous exoribonucleases [12] at structurally well-defined RNAs in the viral 3'UTR [13]. Homologs of these so-called exoribonuclease-resistant RNAs (xrRNAs) are typically found in one or two copies throughout all ecological groups of the genus Flavivirus [14]. Another element that is characteristic of flavivirus 3'UTRs is the long 3'-terminal stem-loop (3'SL) structure, which is is involved in genome cyclisation and panhandle formation during virus replication [15]. As this element is indispensable for the virus life cycle, absence of a 3'SL homolog in sequence data is indicative of incomplete sequencing or truncation. Several other structured RNAs of known and unknown function are found in flavivirus 3'UTRs, including dumbbell (DB) elements and various stem-loop (SL) structures that are either conserved throughout the entire genus or selectively found in particular virus species [16].

1.2 TBEV diversity correlates with pathogenicity

TBEV has traditionally been classified into three subtypes, European (TBEV-Eur), Siberian (TBEV-Sib) and Far Eastern (TBEV-FE) [17], based on serology, phylogenetic reasoning and geographic location of early samples. Increased sampling over the previous decades resulted in the availability of more isolates that did not cluster with the established subtypes (Figure 1). As such, additional lineages and provisional subtypes have been reported. These include strains that have been isolated in Eastern Siberia and that have accordingly been designated as "Baikalean" isolates, i.e. genotype 4 (TBEV-Bkl-1), represented by the single strain 178-79 [18], and genotype 5 (TBEV-Bkl-2), encompassing strain 886-84 and related East-Siberian isolates [19, 20, 21]. Moreover, isolates found in wild rodents in the Quinghai-Tibet Plateau, China, constitute a Himalayan TBEV subtype (TBEV-Him) [22]. The Siberian subtype, with prototype strain Aina, is considered the most abundant TBEV subtype [23], and is subdivided into five lineages: Zausaev, Vasilchenko, Baltic, Bosnia, and Obskaya [24, 25, 26, 27], with Obsakya also being referenced as TBEV-Ob subtype in the literature [28]. The Far-Eastern subtype encompasses three clades that have been referred to as Cluster I-III [29]. Each cluster has a prototype strain, i.e. Oshima (Cluster I), Sofijn (Cluster II) and Senzhang (Cluster III). The European subtype comprises lineages containing the Neudoerfl and Hypr stains, as well as a recently described Western European lineage that has been found in the Netherlands [30]. The Austrian strain N5-17, isolated from a chamois, shares ancestral roots with the other European clades. Notably, the European TBEV isolates are phylogenetically closer to Louping ill virus (LIV), which is present in the British Isles [31], than to the Eastern TBEV subtypes.

In a recent study, the genomic diversity of established an provisional TBEV subtypes has been assessed and classification of TBEV strains into seven subtypes has been proposed based on a 10% cutoff at nucleotide-level diversity [28]. Despite their antigenic similarity, these viruses differ not only in their phylogeography, but also in virulence and pathogenicity. The Far Eastern subtype is commonly considered the most pathogenic TBEV variant with the highest number of reported cases of severe central nervous system (CNS) disease. Case fatality rates of TBEV-FE are typically in the range of 5%-35% [32], with extreme numbers reaching up to 60% [4]. Contrary, the Siberian subtype typically results in less severe disease with a fatality rate between 1% and 3% [4], but is more often associated with chronic TBE [2]. Albeit case fatality at $<$2% is least for the European subtype [33], disease induced by TBEV-Eur infection is typically biphasic with a viraemic phase associated with fever and myalgia, followed by neurological manifestations of different severity during the second phase, which occurs in 20%-30% of all patients [4].

1.3 Functional RNAs in TBEV non-coding regions

The fundamental biological and biochemical traits that govern TBEV pathogenicity are only beginning to be understood. Besides several subtype-specific differences at the amino acid level [29, 34], the non-coding portions of the TBEV genome are critical determinants of virulence. The 5'UTR is approximately 130nt long and contains evolutionarily conserved RNA struc-

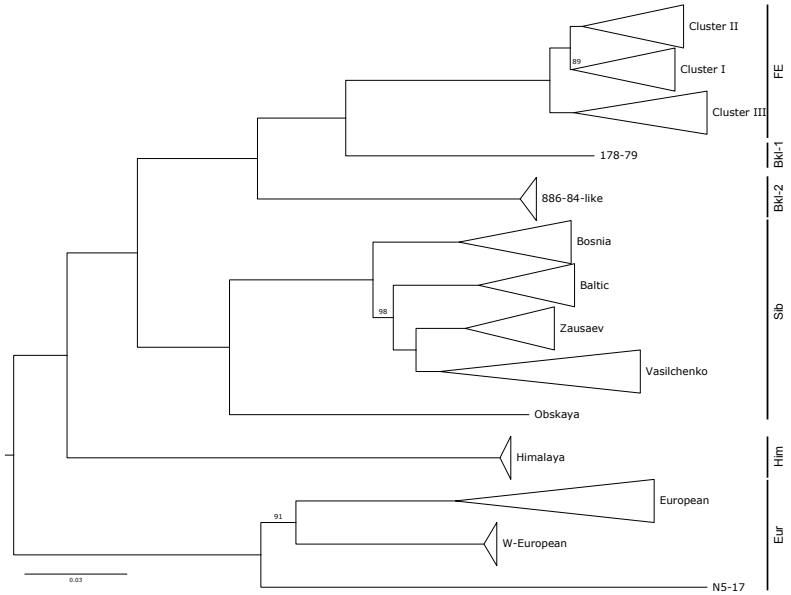

Figure 1: Overview phylogeny of TBEV, depicting the topological arrangement of subtypes TBEV-FE, TBEV-Sib, and TBEV-Eur (all highlighted *sensu strictu* here), as well as novel lineages that do not cluster with these subtypes. Strain 178-79 and the Baikalean lineage share ancestral roots with TBEV-FE. The Himalayan lineage is located ancestral to the clades encompassing TBEV-FE and TBEV-Sib subtypes. The European TBEV strains form a separate clade, with the recently detected Western European and N5-17 lineages being clearly separated from the established TBEV-Eur subtype. The midpoint-rooted maximum likelihood tree is based on 256 complete TBEV genomes. Bootstrap values are shown for branches with a support lower than 100.

tures that are required for replication through cyclisation and panhandle formation. Recently, a cis-acting RNA element in the 5'UTR of TBEV has been identified that mediates neurovirulence by hijacking the host mRNA transport system, thus allowing TBEV gRNA to be transported from the cell body to dendrites of neurons, where it replicates locally [35, 36].

The TBEV 3'UTR is longer than the 5'UTR and contains RNA elements that are involved in mediating immune escape and pathogenicity. It is characterized by a varied architecture of conserved RNA elements that has been associated with differential pathogenicity of particular subtypes [29, 37]. Although a few TBEV-FE isolates exist in public databases that appear to have extremely shortened 3'UTRs of only approximately 50 nt length, typical 3'UTRs range from 350 nt to 760 nt and comprise a 5'-terminal variable region, located immediately downstream of the stop codon, and a 3'-terminal core domain. Earlier comparative studies that have assessed the heterogeneity of these genomic regions at the sequence level have proposed various directed repeat sequences [38, 39]. Considering RNA structure formation [7] allowed for a more fine-grained description and detailed comparative analysis of evolutionarily conserved RNA elements with other tick-borne flaviviruses [14].

An interesting aspect relates to the observation that strains of a particular subtype that have been isolated from ticks or rodents, respectively, are often markedly different from human isolates, particularly in their 3'UTRs. This might be the result of selective pressure on the TBEV 3'UTR during adaptation from tick vector to mammalian host [40]. Likewise, spontaneous deletions of parts of the variable region have been observed after repeated passage [41, 37]. This is complicated by the claim that loss of 3'UTR elements in infected mammalian cells is not quantitative, leading to a scenario that different portions of the viral population have different 3'UTR architectures, each having different replicative fitness when transmitted to ticks [29]. Given that strains resulting in different disease manifestations, ranging from subclinical to encephalitic forms, cluster within the same clades or subcaldes of the TBEV phylogeny [29], the question of functional association and impact on clinical phenotype of different 3'UTR elements is emerging. There are contradictory reports in the literature regarding the influence of deletions in the 3'UTR variable region on the TBEV phenotype. While earlier works proposed no role of the variable

region in virus replication and virulence [41, 42], recent studies suggest increased virulence induced by partial deletions and polyA insertions in mouse models [43, 44].

In this study, we investigate the genetic diversity and phylogeographic spread of TBEV, based on a comprehensive set of more than 220 genomes that encompasses all known and provisional subtypes. By considering also strains that do not cluster with the established subtypes, we provide a unified picture of functional RNA conservation in TBEV 3'UTRs. Moreover, we contextualize genomic and epidemiological data through Nextstrain, providing an updated online resource to study the molecular evolution of TBEV.

2 Methods

2.1 Taxon and metadata collection

Viral genome and annotation data were downloaded from the public National Center for Biotechnology Information (NCBI) Genbank database [45] on 14 January 2021. We obtained 256 TBEV genomes, containing at least full length coding sequences. 221 sequences contained partial or full length 3'UTRs. We extracted metadata for these strains, including location and date of collection, subtype and host species from the Genbank record or from the literature. More than two dozen sequences that have been identified as vaccine strains, highly cell-passaged, or that were the result of cloning, mutants, or duplicates, were filtered out and excluded from downstream analyses. Strain identifiers which were not available through online NCBI resources, but obtained by literature research, were added manually.

Countries were assigned to regions according to the M49 Standard of the United Nations Statistic Division (`https://unstats.un.org/unsd/methodology/m49/`). A categorization of host species was included, which assigns the species of which a particular strain was isolated from to a category according to their role in viral transmission. These categories were vector (mites, ticks), natural host (rodentia, insectivora, aves), non-human dead-end host (large mammalia) or human.

2.2 TBEV subtype and lineage classification

To assign each strain in our dataset to a particular TBEV subtype or lineage, we aligned 221 (near) full-length TBEV genomes with `MAFFT` v7.453 [46] and inferred a maximum likelihood (ML) phylogeny with `iq-tree` v1.6.12 [47] (Figure 1), employing 1000 ultrafast bootstrap replicates [48]. Strains with initially unknown subtype association were assigned to subtypes via distinguishing subtype- or lineage-specific monophyletic clades in the ML tree. Upon classification of strains, we used the multiple nucleotide sequence alignment (MSA) to identify all unique 3'UTR variants within each subtype and lineage, and selected a representative strain for each instance (28 strains, listed in Table S1).

2.3 Characterization of RNA consensus secondary structures

Detection of evolutionary RNA structure conservation builds on the idea of finding homologous RNAs in phylogenetically narrow taxa via RNA family models [49]. Here, we used `infernal` [50] covariance models (CMs) and employed an iterative workflow that has been used recently to characterize evolutionarily conserved RNAs in MBFVs [16]. Starting from a set of TBFV CMs described in [14], we performed an initial screen of TBEV 3'UTRs. Genomic regions that were not covered by the initial CMs were then realigned with `locARNA` v2.0.0RC8 [51] and RNA consensus structures were predicted with `RNAalifold` and `RNALalifold` from the `ViennaRNA` Package v2.4.18 [52].

2.4 Nextstrain Phylogeography

The workflow management system `Snakemake` v5.32.1 [53] was used to build a pipeline for rapid and reproducible deployment of Nextstrain [54] datasets. `augur` v11.0.0 [55] was used

for tracking evolution from nucleotide sequence data, incorporating `MAFFT` v7.453 [46], `iq-tree` v1.6.12 [47], and `TreeTime` v0.8.4 [56], for MSA, ML phylogeny, and timed tree inference, respectively, resulting in a JSON file for Nextstrain visualization. Values for the rate of evolution were chosen according to the molecular clock rate that has been previously described as 5.96e-5 with standard deviation of 6.6e-6 substitutions per site and year [57].

2.5 Data availability

Stockholm MSA files of the conserved RNA structures are available at the `viRNA` GitHub repository (`https://github.com/mtw/viRNA`). The TBEV Nextstrain build is available at `https://nextstrain.org/groups/ViennaRNA/TBEVnext/`. Upon display of the TBEV build in a web browser, all data can be downloaded by scrolling down to the bottom of the page and clicking on the "Download Data" link.

3 Results

To obtain an updated view of TBEV subtype diversity, we compared publicly available genome data from a structural and molecular epidemiological perspective. For the structural analysis, we were particularly interested in the 3'UTR, which constitutes the most divergent part of the TBEV genome. We performed a comparative genomics screen that builds on the concept of predicting consensus structures, i.e. RNA secondary structure that can be formed by all sequences under consideration. We propose the existence of eight distinct, evolutionarily conserved RNA elements in the 3'UTR of different TBEV subtypes. The volatile arrangement of these non-coding RNAs (ncRNAs) at inter- and intra-subtype levels accounts for the diverse TBEV 3'UTR architectures observed in nature.

3.1 TBEV 3'UTRs comprise up to eight distinct RNA elements

Starting from nucleotide multiple sequence alignments of 221 TBEV genomes that covered all known subtypes and lineages, we set out to construct consensus RNA structures of conserved elements. Following up and building on previous work [14], we identified seven distinguishable families of conserved RNA structures that are found as single, double, or triple copies in a variable arrangement in TBEV strains. Together with the poly-A tract that is found in some TBEV-Eur strains, we characterized eight unique ncRNA elements in TBEV 3'UTRs. Inconsistencies with the naming schema of structured TBEV 3'UTR elements in the literature have lead us to introduce here a pragmatic naming system for conserved ncRNA elements.

The structured elements encompass four distinct conserved stem-loop (CSL) elements of unknown function, labeled CSL1 through CSL4, starting from the most 3'-terminal elements. Predicted consensus secondary structures of these elements are shown in Figs. S1 and S2. CSL1, CSL2, and CSL4 are short stem-loop elements that are found in variable copy numbers in different TBEV strains. We have previously shown evolutionary conservation of CSL2 in other tick-borne flaviviruses, where this element was referred to T.SL6 [14]. CSL4, which has a canonical length of approximately 30 nt, has a consensus structure that contains multiple gap characters in the hairpin loop. This is due to the insertion of a poly-A region of variable length in the apical loop of one CSL4 copy in some TBEV-Eur strains. Contrary, CSL3 is markedly longer than the other CSL elements and is interspersed with several interior loops. Comparison of covariance models with `CMCompare` revealed that CSL3 is not structurally homologous to SL-III elements that were found in various mosquito-borne flaviviruses [16].

Besides the CSL elements, TBEV 3'UTRs contain previously described ncRNAs with known functional associations, such as three-way junction elements that form exoribonuclease-resistant RNAs (xrRNAs). These elements, which are ubiquitous in the viral world [58, 59], provide quantitative protection of downstream nucleotides against degradation by endogenous exoribonucleases and have been structurally and mechanistically characterized in different tick-borne flaviviruses [13, 60]. A Y-shaped structure (Y1), together with the the 3'-terminal part of TBEV

3'UTRs, encompassing a small hairpin loop and a long stem-loop element, labeled 3'SL here, comprise the promoter element, which is essential for virus replication.

A structural annotation of TBEV 3'UTRs is shown in Fig. 2, which highlights the differences in architectural organization among and within TBEV subtypes. Selection of strains featured in Fig. 2 was based on subtype diversity, i.e. coverage of all established and prospective subtypes, as well as within-subtype variability of 3'UTR organization. We included at least one representative isolate of each 3'UTR variant that has been present in our data set. Full-length 3'UTRs (where available) are plotted to scale in Figure 2, with colored boxes representing evolutionarily conserved RNA elements. Homologous elements are shown in the same color for each strain.

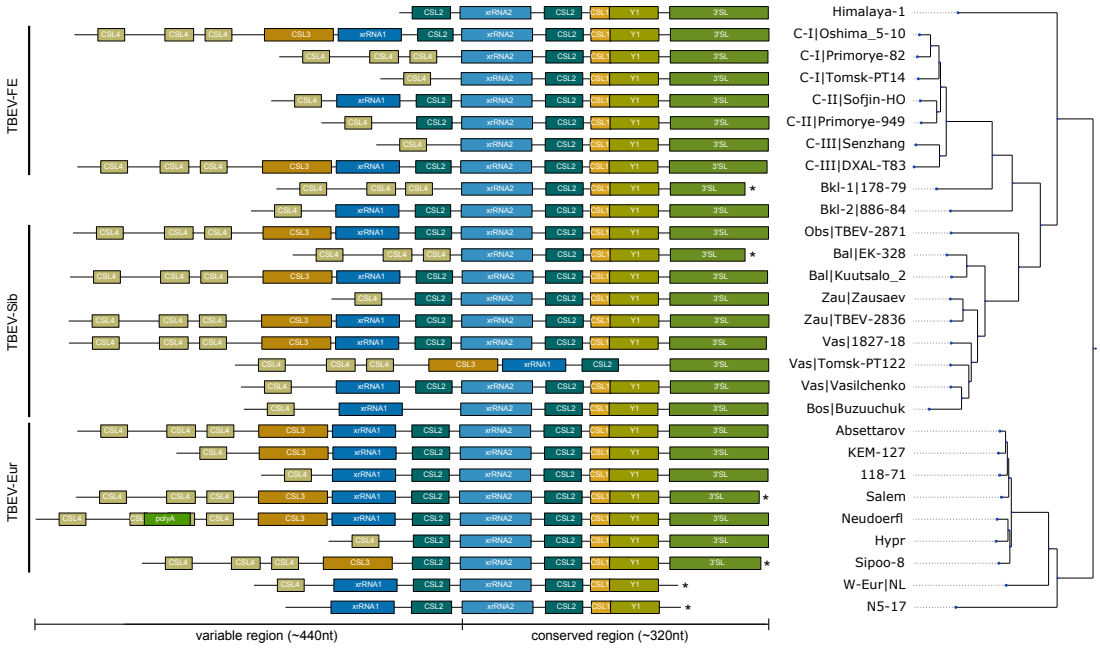

Figure 2: Annotated 3'UTRs of representative TBEV strains. Strains were selected to cover the complete 3'UTR diversity found in publicly available genome data. 3'UTR schemes are plotted to scale, highlighting the variable length and architectural organization of individual strains. Colored boxes represent evolutionarily conserved, structured RNAs. Truncated sequences from incomplete sequencing are marked by an asterisk. The phylogenetic tree on the right has been computed from full genome nucleotide multiple sequence alignments of 28 TBEV strains. Identifiers show strain name, and lineage association where available, i.e. TBEV-FE Clusters I/II/III (C-I/C-II/C-III), the two Baikalean subtypes TBEV-Bkl-1 (Bkl-1), and TBEV-Bkl-2 (Bkl-2), and TBEV-Sib lineages Baltic (Bal), Zausaev (Zau), Vasilchenko (Vas), Bosnia (Bos), Obskaya (Obs). Accession numbers and 3'UTR lengths are listed in Table S1.

Our comparative analysis of TBEV 3'UTRs suggests a common pattern of structural conservation, revealing that both core and variable regions contain specific RNA elements that are characteristic of their location within the 3'UTR. This is particularly marked in the core region, which contains five highly conserved structured RNA elements within approximately 320 nt. These include an exoribonuclease-resistant RNA (xrRNA2), CSL2, CSL1, Y1, and the 3'-terminal long stem-loop structure (3'SL). Consensus secondary structures of core region elements are shown in Fig. S1. The well-conserved architecture of potentially functional RNAs in the core region is observed in all strains considered here, except Tomsk-PT122 (TBEV-Sib).

The variable region ranges between approximately 80 nt (Tomsk-PT-14) and 440 nt (Neudoerfl), and contains up to five different types of RNA elements, i.e. CSL4, poly-A, CSL3, xrRNA1, and CSL2. Consensus secondary structures are shown in Fig. S2. Some of these RNAs occur in multiple copies at particular loci. Most strains contain between one and three copies of CSL4 at the 5'-terminal portion of their 3'UTRs. There are only two exceptions to this observation within our set of strains, i.e. Himalaya-1 and N5-17. The internal poly-A region observed in some European strains is inserted into the hairpin loop of the second CSL4 copy. While a variable number of CSL4 homologs is found in almost all strains, a long variant of the remaining por-

tion of the variable region, containing CSL3, an exoribonuclease-resistant RNA (xrRNA1), and CSL2, is only conserved in some strains. These long 3'UTR variants are found in all subtypes, e.g. Oshima 5-10 (TBEV-FE), 1827-18 (TBEV-Sib), and Absettarov (TBEV-Eur). The structured elements CSL3, xrRNA1 and CSL2 in the variable region appear to be dispensable for virus replication, as other strains either lack a CSL3 element (e.g. Sofjin-HO or Vasilchenko), CSL3 and xrRNA1 (e.g. Hypr or Zausaev), or CSL3, xrRNA1 and CSL2 (e.g. Senzhang or Tomsk PT14).

To illustrate the differences in 3'UTR architectures due to alternative layout of the variable region, we show secondary structure plots of four complete 3'UTRs in Fig. 3. With all strains exhibiting a common core region architecture, we depict representative strains of the European subtype (Neudoerfl, panel A) and the Obskaya lineage of the Siberian subtype (TBEV-2871, panel B), that both exhibit long variants of the variable region. These comprise three CSL4 copies, CSL3, xrRNA1 and one CSL2 element. Insertion of a poly-A stretch results in the formation of a large hairpin loop in the second CSL4 copy of the Neudoerfl strain. The 886-84 strain, representing the Baikalean subtypes, (panel C) has a considerably shorter variable region that contains a single CSL4 copy, xrRNA1, and CSL2, with additional CSL4 structures and CSL3 missing. The Shenzhang strain of the Far Eastern subtype (depicted in panel D) has the shortest 3'UTR in our data set, and contains just one CSL4 element in the variable region.

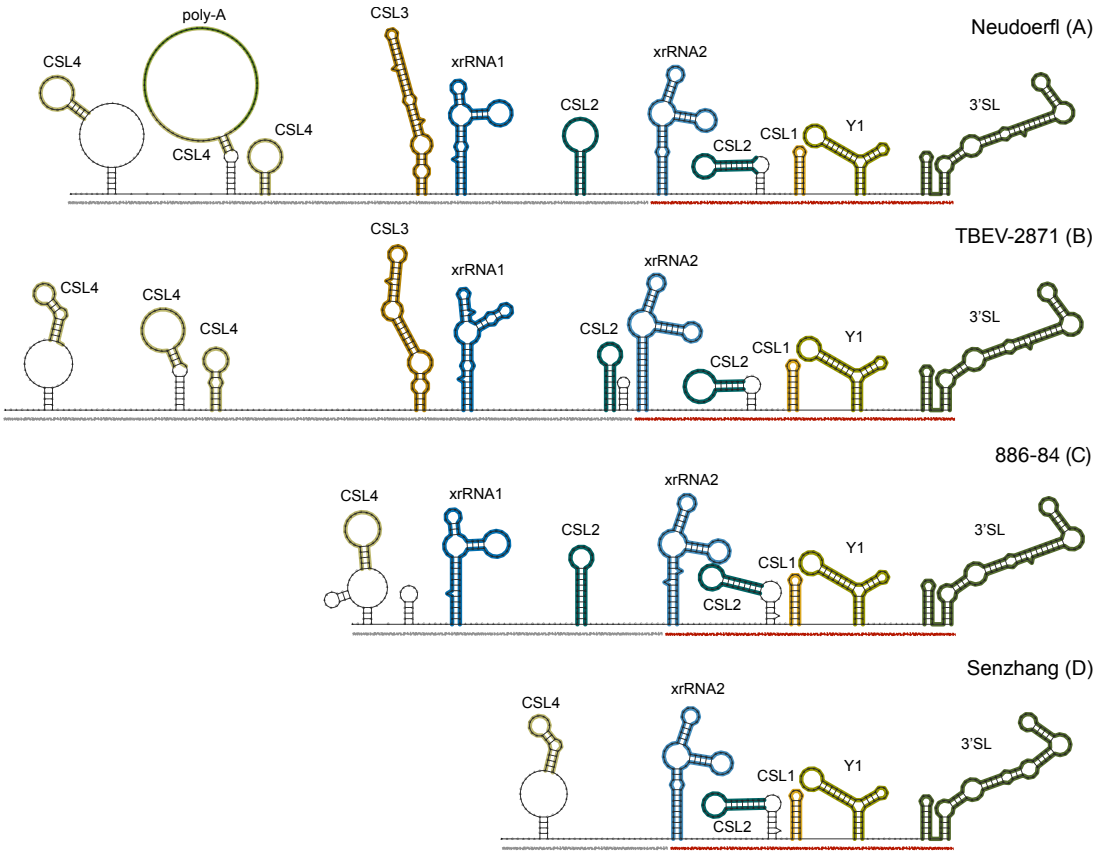

Figure 3: Diverging 3'UTR architecture in four representative TBEV strains, Neudoerfl (A), TBEV-2871 (B), 886-84 (C), and Senzhang (D). Core and variable regions are underlined in red and grey, respectively. Structural homologs of evolutionarily conserved RNA elements are depicted in the same color for all strains (matching the colors used in Fig. 2). Structural components that are not found in all homologs, such as closing stems of some CSL4 and CSL2 elements, or additional short stem loops that are only predicted in particular strains, are not considered conserved.

Some of the recently described strains have a 3'UTR organization that is also found in other lineages. The Western European strain NL and N5-17, while both being truncated at the 3' end, show similar patterns e.g. to strain 886-84. Contrary, other strains exhibit alternative 3'UTR

architectures that are manifested by unique organizational patterns in the variable region that are not observed elsewhere. For example, the 3'UTR of Himalaya-1 does not have a CSL4 element at the 5'-end, but begins with a CSL2 structure. The Buzuuchuk strain, representing the Bosnian lineage of TBEV-Sib, does not have a CSL2 element in the variable region. Sipoo-8, collected in Finland, lacks the exoribonuclease-resistant RNA element in the variable region, although it has three CSL4 elements and CSL3. The most unconventional 3'UTR organization is found in Tomsk-PT122 (Vasilchenko lineage, TBEV-Sib), which is the only example of a strain that exhibits deletions the core region. Tomsk-PT122 is predicted to contain the long version of the variable region described above, while xrRNA2, CSL2, CSL1 and Y elements are are not found in the core region.

3.2 TBEVnext: TBEV molecular epidemiology

To get a comprehensive picture of TBEV phylogeography, considering all publicly available isolates, we sought to to investigate the molecular epidemiology and phylodynamics of TBEV. To this end, we compiled an interactive visualization of the global TBEV spread in Nextstrain, termed TBEVnext, which is publicly available at `https://nextstrain.org/groups/ViennaRNA/TBEVnext`. At the time of writing, TBEvnext v1.0 has been made available with 225 TBEV strains, encompassing all subtypes and lineages.

The overall topology of our TBEV phylogeny (Figure 4a) is in agreement with the literature, featuring Eastern (TBEV-FE, TBEV-Sib), and Western (TBEV-Eur) types as major clades [61]. Strains that are not assigned to one of the larger clades are 178-79 (TBEV-Bkl-1), which is located ancestral to the TBEV-FE clade, and N5-17, which appears ancestral to all other European clades. The East-Siberian/Baikalean/886-84-like (TBEV-Bkl-2) lineage, as well as the Western-European and Himalayan lineages are smaller monophyletic groups that comprise between two and eight isolates. TBEV-Bkl-2 clusters with the Far-Eastern clade, Western European isolates cluster with TBEV-EUR, and the Himalayan isolates share ancestral roots with the clade that comprises all Eastern TBEV isolates. Divergence of the different TBEV subtypes and lineages is exposed in the TBEVnext dataset through an unrooted phylogenetic tree (Figure 4b).

The Nextstrain framework is particularly useful for studying phylogeographic traits. As such, we established a fine-grained geographic location labeling of strains that had relevant location information (i.e. place of isolation) in their metadata. While an exact place of isolation could not be inferred for all strains, sub-national geo-locations are available for many isolates in TBEVnext. This is particularly relevant for studying the spread of different subtypes across geographically extended regions, such as presence of various subtypes in the Russian Federation. An example is Irkutsk Oblast, which shows an accumulation of different subtypes: Western, East-Siberian, Siberian, Far-Eastern and strain 178-79 were collected in this region.

The TBEV Nextstrain dataset features a timed tree, which provides estimates for the time of the most recent common ancestor (TMRCA) associated with each clade (Table 1). The timed tree has been inferred from a ML tree with treetime, assuming a constant evolutionary rate throughout the entire tree. Diverging nucleotide substitution rates have been proposed for different TBEV subtypes [57, 62, 63], hence our numbers should be seen as a rough estimate.

Animation of historic dispersal events across Eurasia within Nextstrain suggests that TBEV originated in Central Russia almost 2700 years ago, and subsequently moved both eastwards, forming the TBEV-Sib and TBEV-FE lineages, and westwards into Europe. While this observation is in agreement with earlier reports [61], the deep splits of the Western types with inferred TMRCAs approximately 1000-1500 years ago raises the question whether TBEV has arrived in Central Europe earlier than previously reported.

One of Nextstrain's strengths is the possibility to augment the dataset with custom traits, in addition to default attributes like date and place of collection. Adopting this feature, we maintain in TBEVnext a mapping of species from which a particular strain has been isolated. It has been previously described that different TBEV subtypes are specialized to particular vector species of Ixodidae. The Eastern subtypes (Far-Eastern, Siberian) are typically transmitted via *I. persulcatus*, whereas the Western subtypes (European) are usually transmitted by *I. ricinus* ticks. While this is broadly supported by our data, metadata analysis of 225 full genome isolates also discloses counter-examples: Nine isolates collected from *I. persulcatus* ticks between 1971

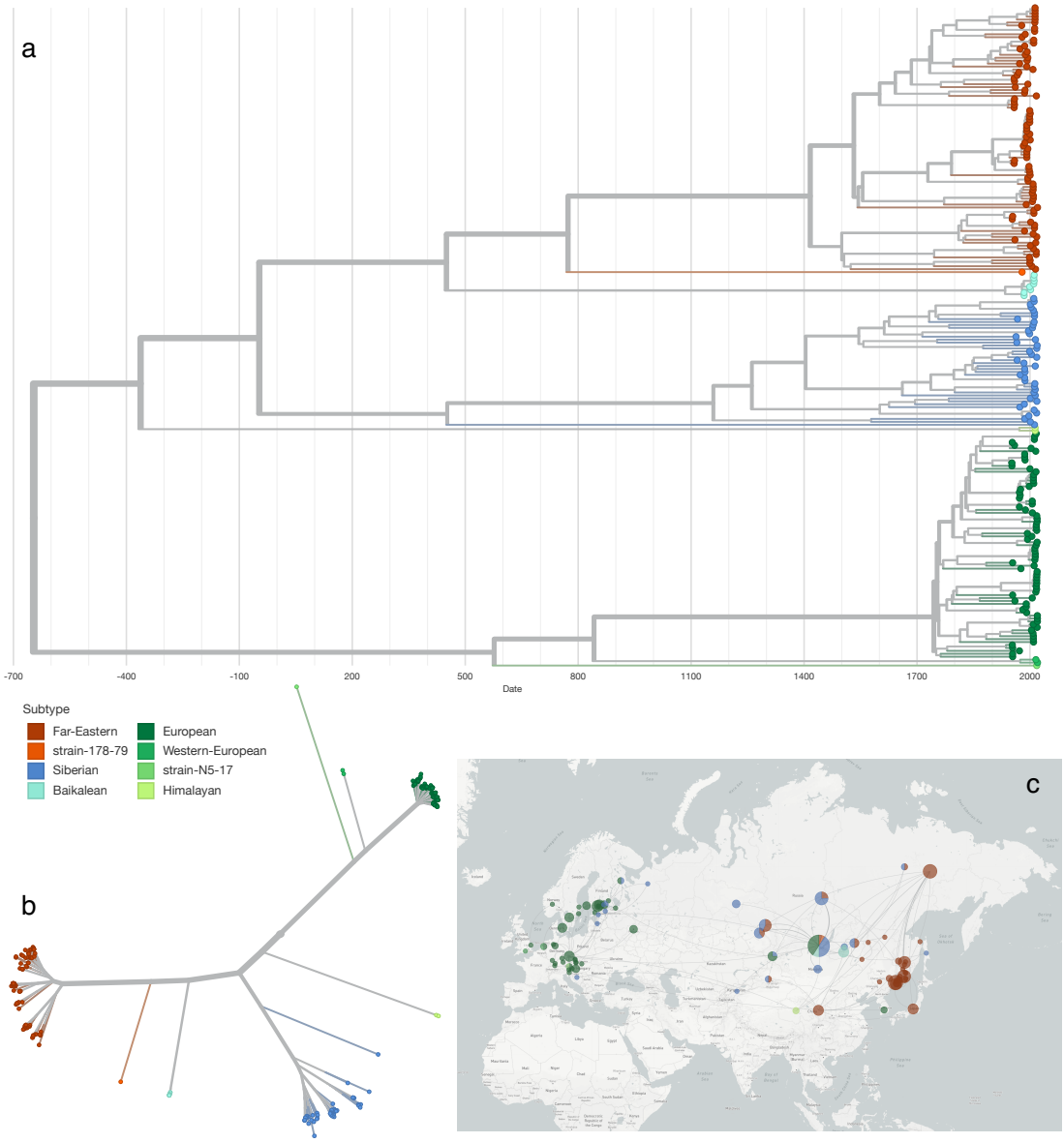

Figure 4: a Visualization of the TBEV phylodynamics in Nextstrain. **b** Unrooted phylogenetic tree, exposing the divergence of TBEV subtypes and novel lineages. **c** Geographic spread of TBEV across Eurasia.

and 2009 in Irkutsk Oblast and the Altai region in Russia carried the European subtype, while two *I. ricinus* ticks carried the Siberian subtype. (Data shown in the interactive visualization of TBEVnext.)

4 Discussion

Increased sample collection and availability of large amounts of genome data in public databases has stipulated research on flaviviruses in recent years. TBEV, being endemic in large parts of Europe and Asia, has been intensively studied, resulting in numerous publications focusing on the functional associations of the variable 3'UTR region in particular strains and viral (neuro)tropism. Some of these studies reported a variety of sequence repeats in the 3'UTR of tick-borne flaviviruses, based on qualitative comparisons at the nucleotide level. Other studies performed computational modelling of RNA secondary structure at the level of individual sequences, yielding a varied picture of stem-loop structures that makes it difficult to compare

Table 1: TMRCA estimates of distinct clades, and associated subtypes of the TBEV phylogeny, extracted from the TBEV Nextstrain dataset. All dates are given as year CE.

Clade	Subtype	TMRCA	Confidence Interval	First Isolation
Siberian	TBEV-Sib	452	(121-703)	1963
Far-Eastern	TBEV-FE	1416	(1287-1512)	1937
Himalayan	TBEV-Him	1971	(1960-1985)	2013
strain 178-79	TBEV-Bkl-1	773	(493-961)	1979
East-Siberian/Baikalean/886-84-like	TBEV-Bkl-2	1933	(1917-1947)	1984
European	TBEV-Eur	1744	(1703-1786)	1951
Western-European	TBEV-Eur	1973	(1962-1987)	2015
strain N5-17	TBEV-Eur	576	(274-809)	2017

structured entities. A unified, reproducible picture of ncRNA structure conservation in different TBEV subtypes has not been available. Here, we set out to obtain an updated view of RNA structure conservation in TBEV 3'UTRs, aiming at capturing a unified picture of the architectural diversity found in different subtypes. To this end, we performed a comparative genomics screen in TBEV 3'UTRs that cover all known subtypes and lineages. The computational approach employed here is based on a thermodynamic model of RNA structure formation, as implemented in the `ViennaRNA` package.

TBEV 3'UTRs comprise two domains, a core domain at the 3'end of the viral genome, and a variable region upstream of the core domain. The 3'-terminal core domain constitutes the most conserved portion of TBEV 3'UTRs, and harbors five different structured ncRNA elements. Two of these, an exoribonuclease-resistant RNA (xrRNA2) and the terminal 3' stem-loop element are also known in other ecological groups of flaviviruses and have been associated with mediating virulence and virus replication [16]. The variable region, on the other hand, contains between two and four distinct structural elements. Observed differences in the length of TBEV genomes that belong to different subtypes is primarily due to heterogeneity of the variable region in TBEV 3'UTRs. Here, another xrRNA copy (xrRNA1) constitutes the only element with known functional association, while simple and extended stem-loop structures of unknown function appear in variable copy numbers. Moreover, an internal poly-A tract is found in several European isolates [64].

TBEV fits well into the picture of selective conservation of RNA structure in the viral 3'UTR, with core and variable regions containing functional RNA elements, i.e. xrRNAs, that are also found in other arboviruses. Unlike other viruses that exhibit strictly lineage-specific patterns of 3'UTR architecture, such as Chikungunya virus [65, 66], TBEV shows a large degree of 3'UTR variability within particular subtypes and lineages. Our data highlight that there is no common 3'UTR architecture associated with certain TBEV subtypes besides the core region. On the other side, we show that the TBEV 3' architectural diversity is realized by combination of a restricted pool of eight structural and one non-structural elements, suggesting that the observed 3'UTR variability is driven by structural, rather than sequence conservation.

Beyond studying evolutionary traits of TBEV at the level of RNA structure conservation, we address the issue also from the perspective of molecular epidemiology. To this end, we provide TBEVnext, a Nextstrain dataset that reveals spatiotemporal and phylodynamic characteristics of the global TBEV spread. Our dataset comprises 225 complete genomes from Eurasia that have been sampled between 1951 and 2019, thereby representing the most comprehensive view of TBEV phylogeographics so far.

In addition to the three major TBEV subtypes TBEV-Eur, TBEV-Sib, and TBEV-Eur, TBEVnext encompasses provisional subtypes like TBEV-Bkl1, TBEV-Bkl-2, and TBEV-Him. The Obskaya lineage of TBEV-Sib, deemed an independent subtype in the recent literature, is considered part of the Siberian clade here, while strain N5-17 is considered part of TBEV-Eur. Analysis of the geographic dispersion of these clades in TBEVnext immediately reveals that they are not specific for particular regions of the world. Lack of an unique association between TBEV

subtypes (clades) and geographic occurrence is particularly evident in Irkutsk Oblast, where different strains of Siberian, Far Eastern, Western, and Baikalean subtypes have been isolated. Likewise, presence of Western strains in the Republic of Korea strains has been reported [67].

TBEVnext features a timed phylogenetic tree that has been inferred from a maximum-likelihood phylogeny under the assumption of a constant evolutionary rate, effectively neglecting the possibility of divergent rates in different parts of the tree. Therefore, our timed tree should be seen as a proxy that is in agreement with proposed TBEV tree topologies and divergence patterns. Importantly, the order of TMRCAs of different clades is in overall agreement with published data that were based on computationally more demanding Bayesian approaches [21].

Availability of global vector/host associations of many strains in TBEVnext yields a comprehensive view on ecological aspects of the global virus spread, and makes unique properties of strains accessible.

In summary, we provide here an updated picture of different evolutionary aspects of tick-borne encephalitis virus. Our data expose patterns of pervasive conservation of individual RNA elements in TBEV 3'UTRs and provide an interactive view of TBEV molecular epidemiology that yields novel insight into the global virus spread.

5 Acknowledgements

We thank Gabor Erdo Mate for fruifutful discussions during the preparation of the manuscript.

References

[1] Daniela Amicizia, Alexander Domnich, Donatella Panatto, Piero Luigi Lai, Maria Luisa Cristina, Ulderico Avio, and Roberto Gasparini. Epidemiology of tick-borne encephalitis (TBE) in Europe and its prevention by available vaccines. *Hum Vacc Immunother*, 9(5):1163–1171, 2013.

[2] Remi N Charrel, H Attoui, AM Butenko, JC Clegg, V Deubel, TV Frolova, EA Gould, TS Gritsun, FX Heinz, M Labuda, et al. Tick-borne virus diseases of human interest in Europe. *Clin Microbiol Infec*, 10(12):1040–1055, 2004.

[3] M Labuda, PA Nuttall, O Kožuch, E Elečková, T Williams, E Žuffová, and A Sabo. Non-viraemic transmission of tick-borne encephalitis virus: a mechanism for arbovirus survival in nature. *Experientia*, 49(9):802–805, 1993.

[4] TS Gritsun, VA Lashkevich, and EA Gould. Tick-borne encephalitis. *Antivir Res*, 57(1-2):129–146, 2003.

[5] Vladimir A Ternovoi, Gennady P Kurzhukov, Yuri V Sokolov, Gennady Y Ivanov, Vladimir A Ivanisenko, Alexander V Loktev, Robert W Ryder, Sergey V Netesov, and Valery B Loktev. Tick-borne encephalitis with hemorrhagic syndrome, Novosibirsk region, Russia, 1999. *Emerg Infect Dis*, 9(6):743, 2003.

[6] Bradley Blitvich and Andrew Firth. Insect-specific flaviviruses: A systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, 7(4):1927–1959, 2015.

[7] Susanne Rauscher, Christoph Flamm, Christian W Mandl, Franz X Heinz, and Peter F Stadler. Secondary structure of the 3'-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities. *RNA*, 3(7):779–791, 1997.

[8] Wy Ng, Ruben Soto-Acosta, Shelton Bradrick, Mariano Garcia-Blanco, and Eng Ooi. The 5' and 3' untranslated regions of the flaviviral genome. *Viruses*, 9(6):137, 2017.

[9] Nicholas J Barrows, Rafael K Campos, Kuo-Chieh Liao, K Reddisiva Prasanth, Ruben Soto-Acosta, Shih-Chia Yeh, Geraldine Schott-Lerner, Julien Pompon, October M Sessions, Shelton S Bradrick, and Mariano A Garcia-Blanco. Biochemistry and Molecular Biology of Flaviviruses. *Chem Rev*, 118(8):4448–4482, 2018.

[10] Sergio M Villordo, Diego E Alvarez, and Andrea V Gamarnik. A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA*, 16(12):2325–2335, 2010.

[11] Luana de Borba, Sergio M Villordo, Nestor G Iglesias, Claudia V Filomatori, Leopoldo G Gebhard, and Andrea V Gamarnik. Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J Virol*, 89(6):3430–3437, 2015.

[12] Gorben P Pijlman, Anneke Funk, Natasha Kondratieva, Jason Leung, Shessy Torres, Lieke Van der Aa, Wen Jun Liu, Ann C Palmenberg, Pei-Yong Shi, Roy A Hall, et al. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host & Microbe*, 4(6):579–591, 2008.

[13] Andrea MacFadden, Zoe O'Donoghue, Patricia AGC Silva, Erich G Chapman, Olsthoorn René C, Mark G Sterken, Gorben P Pijlman, Peter J Bredenbeek, and Jeffrey S Kieft. Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs. *Nat Commun*, 9(1):119, 2018.

[14] Roman Ochsenreiter, Ivo L Hofacker, and Michael T Wolfinger. Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses*, 11(3):298, 2019.

[15] Margo A Brinton and Mausumi Basu. Functions of the 3' and 5' genome RNA regions of members of the genus flavivirus. *Virus Res*, 206:108–119, 2015.

[16] Michael T. Wolfinger, Roman Ochsenreiter, and Ivo L. Hofacker. Functional RNA Structures in the 3'UTR of Mosquito-Borne Flaviviruses. In Dmitrij Frishman and Manja Marz, editors, *Virus Bioinformatics*, pages 65–100. Chapman and Hall/CRC Press, 2021.

[17] Michael Ecker, Steven L Allison, Tamara Meixner, and Franz X Heinz. Sequence analysis and genetic classification of tick-borne encephalitis viruses from Europe and Asia. *J Gen Virol*, 80(1):179–185, 1999.

[18] TV Demina, Dzhioev IuP, IV Kozlova, MM Verkhozina, SE Tkachev, EK Doroshchenko, OV Lisak, AI Paramonov, and VI Zlobin. Genotypes 4 and 5 of the tick-borne encephalitis virus: features of the genome structure and possible scenario for its formation. *Vopr Virusol*, 57(4):13–19, 2012. In Russian.

[19] TV Demina, Yu P Dzhioev, MM Verkhozina, IV Kozlova, SE Tkachev, A Plyusnin, EK Doroshchenko, OV Lisak, and VI Zlobin. Genotyping and characterization of the geographical distribution of tick-borne encephalitis virus variants with a set of molecular probes. *J. Med. Virol.*, 82(6):965–976, 2010.

[20] SY Kovalev and TA Mukhacheva. Reconsidering the classification of tick-borne encephalitis virus within the Siberian subtype gives new insights into its evolutionary history. *Infect Genet Evol*, 55:159–165, 2017.

[21] Renat V Adelshin, Elena A Sidorova, Artem N Bondaryuk, Anna G Trukhina, Dmitry Yu Sherbakov, Richard Allen White III, Evgeny I Andaev, and Sergey V Balakhonov. "886-84-like" tick-borne encephalitis virus strains: Intraspecific status elucidated by comparative genomics. *Ticks Tick-borne Dis*, 10(5):1168–1172, 2019.

[22] Xiaoyi Dai, Guobao Shang, Shan Lu, Jing Yang, and Jianguo Xu. A new subtype of eastern tick-borne encephalitis virus discovered in Qinghai-Tibet Plateau, China. *Emerg Microb Infect*, 7(1):1–9, 2018.

[23] SE Tkachev, IV Babkin, GS Chicherina, IV Kozlova, MM Verkhozina, TV Demina, OV Lisak, EK Doroshchenko, Yu P Dzhioev, OV Suntsova, et al. Genetic diversity and geographical distribution of the siberian subtype of the tick-borne encephalitis virus. *Ticks Tick-borne Dis.*, 11(2):101327, 2020.

[24] Tamara S Gritsun, TV Frolova, VV Pogodina, VA Lashkevich, K Venugopal, and EA Gould. Nucleotide and deduced amino acid sequence of the envelope gene of the Vasilchenko strain of TBE virus; comparison with other flaviviruses. *Virus Res*, 27(2):201–209, 1993.

[25] Irina Golovljova, Sirkka Vene, Katarina Brus Sjölander, Veera Vasilenko, Alexander Plyusnin, and Åke Lundkvist. Characterization of tick-borne encephalitis virus from Estonia. *J Med Virol*, 74(4):580–588, 2004.

[26] Irina Golovljova, Olga Katargina, Julia Geller, Tatjana Tallo, Vassili Mittženkov, Sirkka Vene, Kirill Nemirov, Aleksey Kutsenko, Gelena Kilosanidze, Veera Vasilenko, et al. 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, 2008.

[27] Sergey E Tkachev, Galina S Chicherina, Irina Golovljova, Polina S Belokopytova, Artem Yu Tikunov, Oksana V Zadora, Victor V Glupov, and Nina V Tikunova. New genetic lineage within the Siberian subtype of tick-borne encephalitis virus found in Western Siberia, Russia. *Infect Genet Evol*, 56:36–43, 2017.

[28] Andrei A Deviatkin, Galina G Karganova, Yulia A Vakulenko, and Alexander N Lukashev. TBEV Subtyping in Terms of Genetic Distance. *Viruses*, 12(11):1240, 2020.

[29] Sergei I Belikov, Ilya G Kondratov, Ulyana V Potapova, and Galina N Leonova. The relationship between the structure of the tick-borne encephalitis virus strains and their pathogenic properties. *PLoS One*, 9(4):e94946, 2014.

[30] Margriet Dekker, Gozewijn Dirk Laverman, Ankje De Vries, Johan Reimerink, and Felix Geeraedts. Emergence of tick-borne encephalitis (TBE) in the Netherlands. *Ticks Tick-borne Dis*, 10(1):176–179, 2019.

[31] CL Jeffries, KL Mansfield, LP Phipps, PR Wakeley, R Mearns, A Schock, S Bell, AC Breed, AR Fooks, and N Johnson. Louping ill virus: an endemic tick-borne disease of Great Britain. *J Gen Virol*, 95(Pt 5):1005, 2014.

[32] Uga Dumpis, Derrick Crook, and Jarmo Oksi. Tick-borne encephalitis. *Clin Infect Dis*, 28(4):882–890, 1999.

[33] Reinhard Kaiser. The clinical and epidemiological profile of tick-borne encephalitis in southern Germany 1994–98: a prospective study of 656 patients. *Brain*, 122(11):2067–2078, 1999.

[34] Eliza M Kellman, Danielle K Offerdahl, Wessam Melik, and Marshall E Bloom. Viral determinants of virulence in tick-borne flaviviruses. *Viruses*, 10(6):329, 2018.

[35] Minato Hirano, Kentaro Yoshii, Mizuki Sakai, Rie Hasebe, Osamu Ichii, and Hiroaki Kariwa. Tick-borne flaviviruses alter membrane structure and replicate in dendrites of primary mouse neuronal cultures. *J Gen Virol*, 95(4):849–861, 2014.

[36] Minato Hirano, Memi Muto, Mizuki Sakai, Hirofumi Kondo, Shintaro Kobayashi, Hiroaki Kariwa, and Kentaro Yoshii. Dendritic transport of tick-borne flavivirus RNA by neuronal granules affects development of neurological disease. *Proc Natl Acad Sci USA*, 114(37):9960–9965, 2017.

[37] Vladimir A Ternovoi, Anastasia V Gladysheva, Eugenia P Ponomareva, Tamara P Mikryukova, Elena V Protopopova, Alexander N Shvalov, Svetlana N Konovalova, Eugene V Chausov, and Valery B Loktev. Variability in the 3' untranslated regions of the genomes of the different tick-borne encephalitis virus subtypes. *Virus Genes*, 55(4):448–457, 2019.

[38] Gerhard Wallner, Christian W Mandl, Christian Kunz, and Franz X Heinz. The flavivirus 3'-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology*, 213(1):169–178, 1995.

[39] TS Gritsun, K Venugopal, PM de A Zanotto, MV Mikhailov, AA Sall, EC Holmes, I Polkinghorne, TV Frolova, VV Pogodina, VA Lashkevich, et al. Complete sequence of two tick-borne flaviviruses isolated from Siberia and the UK: analysis and significance of the 5' and 3'-UTRs. *Virus Res*, 49(1):27–39, 1997.

[40] Naveed Asghar, Pontus Lindblom, Wessam Melik, Richard Lindqvist, Mats Haglund, Pia Forsberg, Anna K Överby, Åshild Andreassen, Per-Eric Lindgren, and Magnus Johansson. Tick-borne encephalitis virus sequenced directly from questing and blood-feeding ticks reveals quasispecies variance. *PLoS One*, 9(7):e103264, 2014.

[41] Christian W Mandl, Heidemarie Holzmann, Tamara Meixner, Susanne Rauscher, Peter F Stadler, Steven L Allison, and Franz X Heinz. Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. *J Virol*, 72(3):2132–2140, 1998.

[42] Verena M Hoenninger, Harald Rouha, Klaus K Orlinger, Lisa Miorin, Alessandro Marcello, Regina M Kofler, and Christian W Mandl. Analysis of the effects of alterations in the tick-borne encephalitis virus 3'-noncoding region on translation and RNA replication using reporter replicons. *Virology*, 377(2):419–430, 2008.

[43] Mizuki Sakai, Kentaro Yoshii, Yuji Sunden, Kana Yokozawa, Minato Hirano, and Hiroaki Kariwa. The variable region of the 3'untranslated region is a critical virulence factor in the Far-Eastern subtype of tick-borne encephalitis virus in mouse model. *J Gen Virol*, 95(4):823–835, 2014.

[44] Mizuki Sakai, Memi Muto, Minato Hirano, Hiroaki Kariwa, and Kentaro Yoshii. Virulence of tick-borne encephalitis virus is associated with intact conformational viral rna structures in the variable region of the 3'-utr. *Virus Res*, 203:36–40, 2015.

[45] Karen Clark, Ilene Karsch-Mizrachi, David J Lipman, James Ostell, and Eric W Sayers. Genbank. *Nucleic Acids Res*, 44(D1):D67–D72, 2016.

[46] Kazutaka Katoh and Daron M Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30(4):772–780, 2013.

[47] Lam-Tung Nguyen, Heiko A Schmidt, Arndt von Haeseler, and Bui Quang Minh. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*, 32(1):268–274, 2014.

[48] Diep Thi Hoang, Olga Chernomor, Arndt Von Haeseler, Bui Quang Minh, and Le Sy Vinh. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*, 35(2):518–522, 2018.

[49] Sean R Eddy and Richard Durbin. Rna sequence analysis using covariance models. *Nucleic Acids Res*, 22(11):2079–2088, 1994.

[50] Eric P. Nawrocki and Sean R. Eddy. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935, 2013.

[51] Sebastian Will, Kristin Reiche, Ivo L Hofacker, Peter F Stadler, and Rolf Backofen. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol*, 3(4):e65, 2007.

[52] Ronny Lorenz, Stephan H Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. ViennaRNA Package 2.0. *Algorithm Mol Biol*, 6(1):26, 2011.

[53] Johannes Köster and Sven Rahmann. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, 2012.

[54] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, 2018.

[55] John Huddleston, James Hadfield, Thomas Sibley, Jover Lee, Kairsten Fay, Misja Ilcisin, Elias Harkins, Trevor Bedford, Richard Neher, and Emma Hodcroft. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw*, 6(57):2906, jan 2021.

[56] Pavel Sagulenko, Vadim Puller, and Richard A Neher. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol*, 4(1):vex042, 2018.

[57] Artem N. Bondaryuk, Elena A. Sidorova, Renat V. Adelshin, Evgeny I. Andaev, and Sergey V. Balakhonov. Reporting of New tick-borne encephalitis virus strains isolated in Eastern Siberia (Russia) in 1960-2011 and explaining them in an evolutionary context using Bayesian phylogenetic inference. *Ticks Tick-borne Dis*, 11(5):101496, 2020.

[58] Anna-Lena Steckelberg, Quentin Vicens, and Jeffrey S Kieft. Exoribonuclease-resistant RNAs exist within both coding and noncoding subgenomic RNAs. *MBio*, 9(6):e02461–18, 2018.

[59] Ivar W Dilweg, Alexander P Gultyaev, and René C Olsthoorn. Structural features of an Xrn1-resistant plant virus RNA. *RNA Biol*, 16(6):838–845, 2019.

[60] Hayato Harima, Yasuko Orba, Shiho Torii, Yongjin Qiu, Masahiro Kajihara, Yashiki Eto, Naoya Matsuta, Hang'ombe Bernard M., Yuki Eshita, Kentaro Uemura, Keita Matsuno, Michihito Sasaki, Kentaro Yoshii, Ryo Nakao, William Hall, Ayato Takada, Takashi Abe, Michael T. Wolfinger, Martin Simmunza, and Hirofumi Sawa. An African tick flavivirus forming an independent clade exhibits unique exoribonuclease-resistant RNA structures in the genomic 3'-untranslated region. *Sci Rep*, 2021.

[61] DM Heinze, EA Gould, and NL Forrester. Revisiting the clinal concept of evolution and dispersal for the tick-borne flaviviruses by using phylogenetic and biogeographic analyses. *J Virol*, 86(16):8663–8671, 2012.

[62] EL Subbotina and VB Loktev. Molecular evolution of the tick-borne encephalitis and Powassan viruses. *Mol Biol*, 46(1):75–84, 2012.

[63] Andrei A Deviatkin, Ivan S Kholodilov, Yulia A Vakulenko, Galina G Karganova, and Alexander N Lukashev. Tick-Borne Encephalitis Virus: An Emerging Ancient Zoonosis? *Viruses*, 12(2):247, 2020.

[64] Cristian W Mandl, Christian Kunz, and Franz X Heinz. Presence of Poly(A) in a Flavivirus: Significant Differences between the 3'Noncoding Regions of the Genomic RNAs of Tick-Borne Encephalitis Virus Strains. *J. Virol.*, 65(8):4070–4077, 1991.

[65] Adriano de Bernardi Schneider, Roman Ochsenreiter, Reilly Hostager, Ivo L. Hofacker, Daniel Janies, and Michael T Wolfinger. Updated Phylogeny of Chikungunya Virus Suggests Lineage-Specific RNA Architecture. *Viruses*, 11:798, 2019. PMCID: PMC6784101.

[66] Thomas Spicher, Markus Delitz, Adriano de Bernardi Schneider, and Michael T. Wolfinger. Dynamic molecular epidemiology reveals lineage-associated single-nucleotide variants that alter RNA structure in Chikungunya virus. *Genes*, 12(2):239, 2021.

[67] Seok-Min Yun, Su Yeon Kim, Young Ran Ju, Myung Guk Han, Young Eui Jeong, and Jungsang Ryou. First complete genomic characterization of two tick-borne encephalitis virus isolates obtained from wild rodents in South Korea. *Virus Genes*, 42(3):307–316, 2011.

Evolutionary traits of Tick-borne encephalitis virus: Pervasive non-coding RNA structure conservation and molecular epidemiology
– Supplementary Data –

Lena S. Kutschera[1] and Michael T. Wolfinger[1,2,*]

[1]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria
[2]Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währingerstraße 29, 1090 Vienna, Austria

Table S1: Representative strains used in this study.

Accession	Subtype	Strain	Note	3'UTR length
MG599476.1	TBEV-Him	Himalaya-1	-	384
AB062063.2	TBEV-FE	Oshima 5-10	Cluster I	724
JQ825148.1	TBEV-FE	Primorye-82	Cluster I	509
KJ914682.1	TBEV-FE	Tomsk-PT14	Cluster I	404
AB062064.1	TBEV-FE	Sofjin-HO	Cluster II	518
KU761570.1	TBEV-FE	Primorye-949	Cluster II	466
JQ650523.1	TBEV-FE	Senzhang	Cluster III	408
MN615728.1	TBEV-FE	DXAL-T83	Cluster III	719
EF469661.1	TBEV-Bkl-1	178-79	-	488
KJ633033.1	TBEV-Bkl-2	886-84	-	539
MF774565.1	TBEV-Sib	TBEV-2871	Obskaya	725
DQ486861.1	TBEV-Sib	EK-328	Baltic	471
MG589939.1	TBEV-Sib	Kuutsalo_2	Baltic	727
AF527415.1	TBEV-Sib	Zausaev	Zausaev	455
MH645618.1	TBEV-Sib	TBEV-2836	Zausaev	730
MN114637.1	TBEV-Sib	1827-18	Vasilchenko	727
KM019545.1	TBEV-Sib	Tomsk-PT122	Vasilchenko	556
AF069066.1	TBEV-Sib	Vasilchenko	Vasilchenko	550
KJ626343.1	TBEV-Sib	Buzuuchuk	Bosnia	546
KJ000002.1	TBEV-Eur	Absettarov	-	720
MG210947.1	TBEV-Eur	KEM-127	-	617
KY069120.1	TBEV-Eur	118-71	-	528
FJ572210.1	TBEV-Eur	Salem	-	712
NC_001672.1	TBEV-Eur	Neudoerfl	-	764
U39292.1	TBEV-Eur	Hypr	-	458
MK801805.1	TBEV-Eur	Sipoo-8-Finland-2013	-	645
LC171402.1	TBEV-Eur	NL	W-Eur	442
MG243699.1	TBEV-Eur	N5-17	-	412

*To whom correspondence should be addressed: michael.wolfinger@univie.ac.at

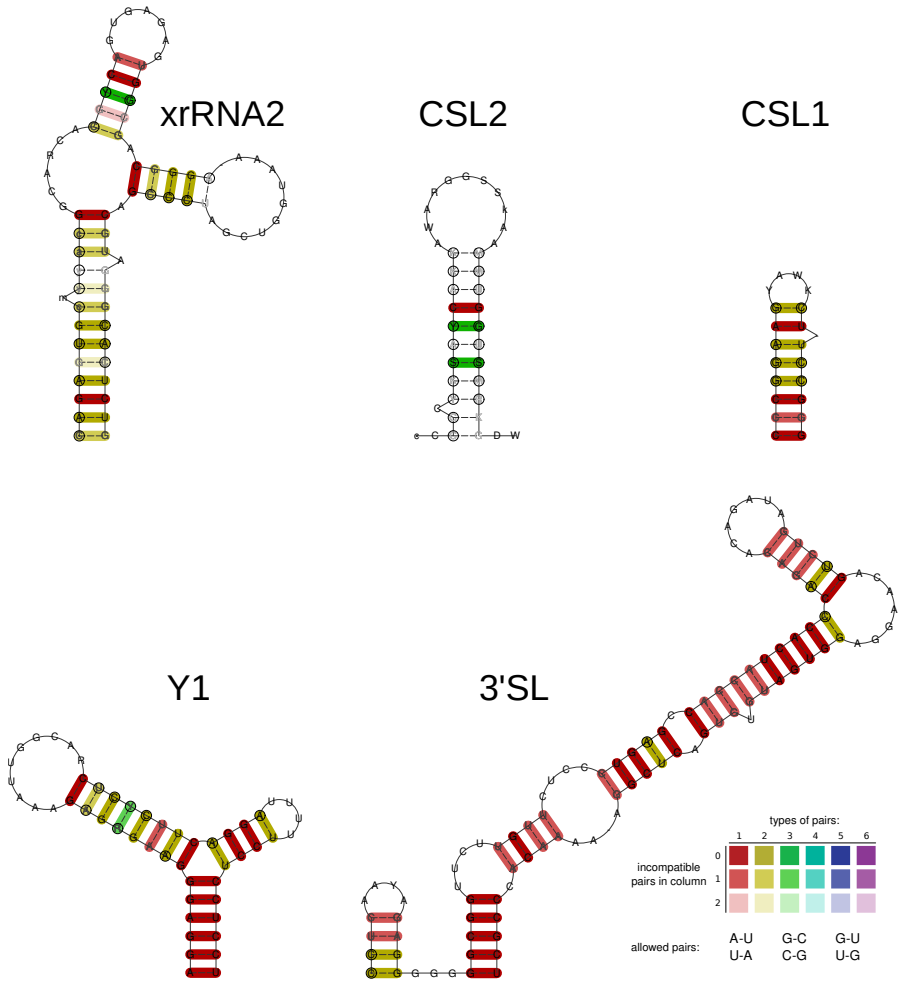

Figure S1: Consensus secondary structures of conserved RNA elements in the core region of TBEV 3'UTRs. Base pair coloring indicates the level of covariation, following the default `RNAalifold` color scheme (bottom right). Nucleotides are labeled in IUPAC notation. Circled nucleotides indicate compensatory mutations in the respective columns of the underlying multiple sequence alignments (MSAs). Structural MSAs for all elements shown here are available at the `viRNA` GitHub repository (`https://github.com/mtw/viRNA`).

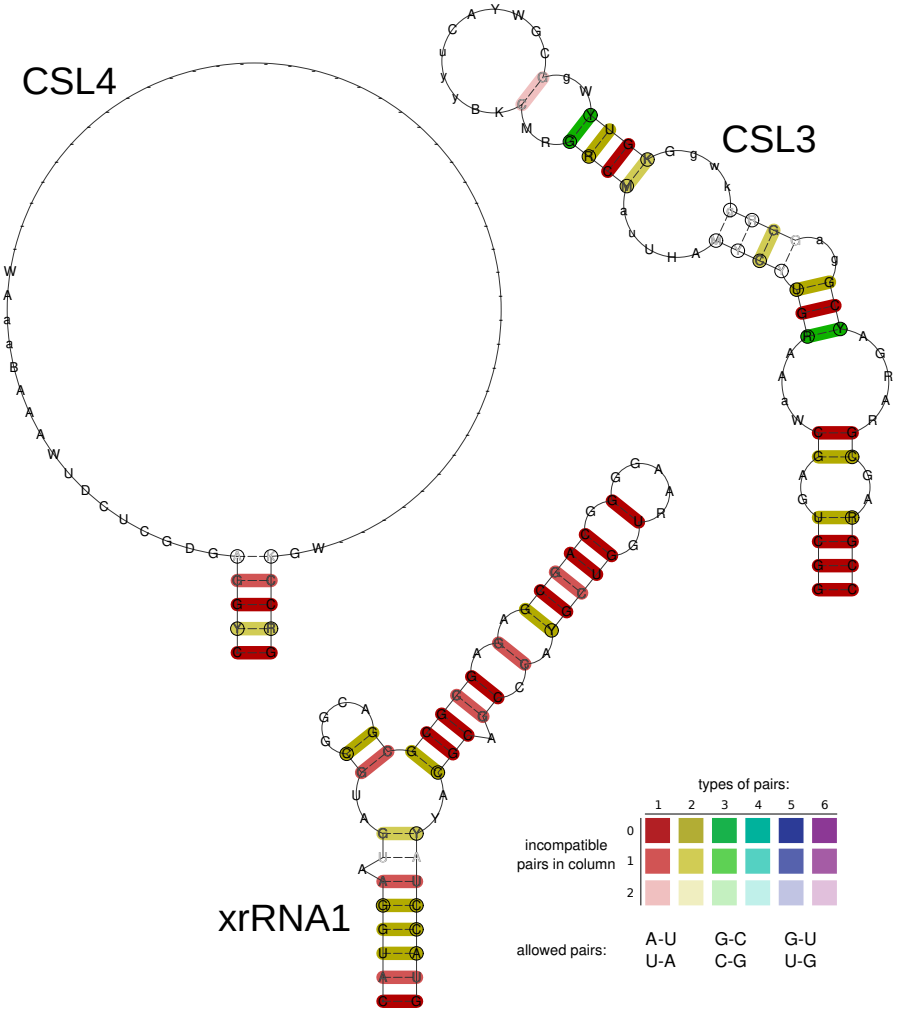

Figure S2: Consensus secondary structures of conserved RNA elements in the variable region of TBEV 3'UTRs. The large amount of gap characters in the hairpin loop of the CSL4 consensus structure is due to an insertion of a poly-A tract in one copy of the CSL4 element, observed in several TBFV-Eur strains. See main text for details and Fig. S1 for a description of base pair coloring.