Journal of General Virology (2010), **91**, 1418–1427 **DOI** 10.1099/vir.0.018887-0

Genomic features and evolutionary constraints in Saffold-like cardioviruses

Jan Felix Drexler,[1,2] Sigrid Baumgarte,[3] Luciano Kleber de Souza Luna,[4] Andreas Stöcker,[1] Patrícia Silva Almeida,[5] Tereza Cristina Medrado Ribeiro,[5] Nadine Petersen,[4] Petra Herzog,[4] Célia Pedroso,[1] Carlos Brites,[1] Hugo da Costa Ribeiro, Jr,[5] Anatoly Gmyl,[6] Christian Drosten[2] and Alexander Lukashev[6]

Correspondence
Christian Drosten
drosten@virology-bonn.de

[1]Research Laboratory for Infectious Diseases, Hospital Professor Edgar Santos, Federal University of Bahia, Salvador, Bahia, Brazil

[2]Institute of Virology, University of Bonn Medical Centre, Bonn, Germany

[3]Laboratory of Virology, Department of Microbiological Consumer Protection, Institute of Hygiene and the Environment, Hamburg, Germany

[4]Clinical Virology Group, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

[5]Department of Paediatrics, Hospital Professor Edgar Santos, Federal University of Bahia, Salvador, Bahia, Brazil

[6]Chumakov Institute of Poliomyelitis and Viral Encephalitides, Moscow, Russia

This study identified the complete genomic sequence of four type 2 and type 3 human Saffold-like cardioviruses (SLCVs) isolated in Germany and Brazil. The secondary structures of the SLCV internal ribosome entry sites (IRESs) were deduced based on RNA base-pairing conservation and co-variation, using an established Theiler's murine encephalomyelitis virus (TMEV) IRES structure as a reference. The SLCV IRES was highly similar to that of TMEV, but motifs critical in TMEV for binding of the polypyrimidine tract-binding protein (PTB) were disrupted. In TMEV, corresponding alterations have been associated with reduced neurovirulence in mice. In the non-structural genome region, there was evidence of multiple intertypic recombination events between different SLCV types. Between viruses of the same type, recombination also occurred in the capsid-encoding genome region. There were apparently no recombination events between mouse TMEV and human SLCV. In another genus of the family *Picornaviridae*, *Enterovirus*, natural recombination occurs strictly within species and can serve as an additional criterion for delimiting species. Accordingly, the results of this study suggest that SLCV and TMEV may represent distinct species within the genus *Cardiovirus*.

Received 2 December 2009
Accepted 2 February 2010

INTRODUCTION

The genus *Cardiovirus* comprises non-enveloped, positive-sense RNA viruses in the family *Picornaviridae*. Until 2007, three rodent cardioviruses, encephalomyocarditis virus (EMCV), Theiler's murine encephalomyelitis virus (TMEV) and rat theilovirus (RTV; Mhg agent), were the only known representatives of this genus. EMCV has been widely used to model human diseases such as myocarditis, encephalitis and pancreatitis in rodents. TMEV serves as a rodent model for the infectious aetiology of chronic and acute neurological demyelinating disease. Sporadic EMCV infections have been described in other mammals, including humans (Kirkland *et al.*, 1989). A TMEV-like cardiovirus has been isolated from a person with Vilyuisk encephalitis, an apparently infectious neurodegenerative disease of indigenous people in Vilyuisk, Siberia (Vladimirtsev *et al.*, 2007). There are, however, serious concerns about whether this virus may have been a contaminant resulting from mouse inoculation; serum antibodies to the virus were found in only a small fraction of Vilyuisk encephalitis patients (Gajdusek *et al.*, 1970; Sarmanova & Chumachenko, 1960).

In addition to these two murine cardioviruses, a novel cardiovirus founding a third, genetically distinct, genetic lineage was recently identified in a cell culture isolate from a child with fever of unknown origin (Jones *et al.*, 2007). This virus was named Saffold virus (SafV). We and others

The GenBank/EMBL/DDBJ accession numbers for the Saffold virus sequences determined in this study are EU681176–EU681179.

have subsequently described Saffold-like cardioviruses (SLCVs) from different continents in children suffering mostly from gastroenteritis (Abed & Boivin, 2008; Blinkova *et al.*, 2009; Chiu *et al.*, 2008; Drexler *et al.*, 2008; Zoll *et al.*, 2009). Serological evidence suggests that most humans acquire SLCV infections by the second year of life (Zoll *et al.*, 2009), similar to human enteroviruses and parechoviruses.

Asymptomatic enterovirus and parechovirus infection is highly prevalent in infants (Lu *et al.*, 2002; Tauriainen *et al.*, 2007; Witso *et al.*, 2006), but both can cause severe disease, such as newborn sepsis and meningitis (Harvala *et al.*, 2009; Pallansch & Roos, 2001). Up to around 25 % of all cases may be contributed by these viruses (Wolthers *et al.*, 2008). In view of the epidemiological similarities between parechoviruses, enteroviruses and SLCV, it is conceivable that SLCV might cause relevant clinical disease in a fraction of infected humans. Moreover, the related TMEV has defined genomic features that correlate with neurovirulence in mice, and it is currently unclear whether a similar pathology might be associated with SLCV in humans. It is thus relevant to analyse the genomic features of SLCV and to understand the mechanisms that define and drive SLCV evolution.

Eight SLCV types have been proposed so far (Blinkova *et al.*, 2009). Unique full-genome sequences have been determined for putative types 1, 2, 3, 5 and 6 (Abed & Boivin, 2008; Blinkova *et al.*, 2009; Chiu *et al.*, 2008; Jones *et al.*, 2007; Zoll *et al.*, 2009). In this study, we determined the complete genome sequences of four SLCV strains from Germany and Brazil, covering types 2 and 3. Integrating these with existing data, we conducted a genetic analysis of SLCV with special emphasis on taxonomic classification, RNA secondary structure and natural recombination.

RESULTS AND DISCUSSION

We determined the complete nucleotide sequences of four SLCVs that had been identified previously (Drexler *et al.*, 2008). These genomes were aligned with all complete or near-complete genomic sequences of rodent theiloviruses and SLCVs deposited in GenBank.

Secondary structure of internal ribosome entry sites (IRESs)

All known picornaviruses contain IRESs in their 5′ non-translated region (NTR). The IRES enables initiation of translation by coordinated binding of canonical initiation factors and IRES-specific *trans*-acting factors (reviewed by Pestova *et al.*, 2001). The 5′NTR of both human and murine cardioviruses contains a type II IRES at genome position nt 567–1040. A model of the TMEV IRES was created previously based on chemical and enzymic probing (Pilipenko *et al.*, 1989, 2001). We used the established TMEV IRES model as a base to generate a secondary structure prediction for SLCV strain D/VI2273/2004 (Fig. 1). Alignment of the SLCV genomes was used to confirm that conserved nucleotide pairs and compensatory mutations among all genomes perfectly supported this structure. The structure prediction corresponded well with the available model of the TMEV IRES. Multiple compensatory mutations preserving paired nucleotides in secondary RNA structure confirmed the validity of the original model. The only minor difference between the IRES of human and murine viruses was in the apical part of the J domain, which was 3 bp shorter in SLCV.

The classical TMEV expresses a 156 aa L* protein from an alternative start codon 13 nt downstream of the canonical

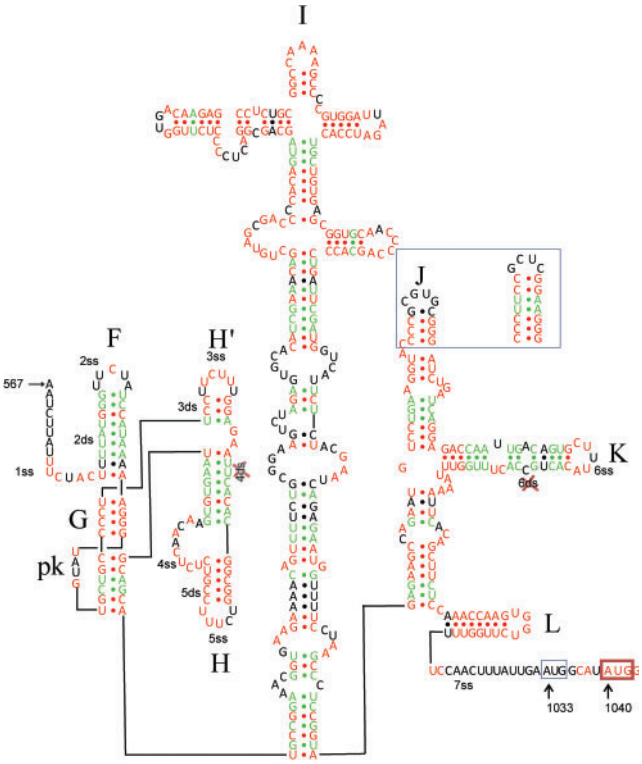

Fig. 1. Model of the secondary structure of the IRES of strain D/VI2273/2004. An authentic AUG initiation codon at position 1040 (numbering according to SafV-1, GenBank accession no. NC_009448) and an additional AUG codon at position 1033 are boxed. Stem–loops are labelled according to Duke *et al.* (1992). The pseudoknot is designated pk. Base pairs conserved on the level of primary and secondary structure, as well as absolutely conserved nucleotides in single-stranded regions of SafVs, are in red. Base pairs supported by compensatory mutations among SafVs are in green. Binding sites of polypyrimidine tract-binding protein (PTB) and its neural homologue (nPTB) are marked according to Pilipenko *et al.* (2001) as ss for single-stranded and ds for double-stranded sites. PTB (nPTB)-binding sites destroyed in SLCVs are indicated with a cross. The boxed J domain indicates differences between the SafV and TMEV IRES.

polyprotein AUG. This protein has been suggested to play a key role in virus persistence. In certain cases, translation of this protein can be initiated from an ACG start codon instead of AUG (Chen *et al.*, 1995; Michiels *et al.*, 1995; van Eyll & Michiels, 2000). None of the analysed SLCVs encoded a full-size L* protein equivalent to that in TMEV. The length of a putative open reading frame originating from an alternative AUG or ACG codon in SLCV was only 24–57 aa. As shown for TMEV, a variant of L* terminating 39 nt beyond the start codon was non-functional (van Eyll & Michiels, 2000). The genomes of five of the 11 SLCVs contained an additional AUG codon 7 nt upstream of the authentic start codon in frame with the putative L* protein. This AUG codon was absent in all TMEVs. However, experimental evidence indicates that this AUG is too close to the IRES and falls outside the 'starting window' in such a way that it cannot initiate translation (Pilipenko *et al.*, 1994).

Polypyrimidine tract-binding protein (PTB)-binding sites

The TMEV IRES contains conserved sequence motifs that are specifically bound by the cellular PTB or its neural homologue (nPTB), regulating the efficiency of IRES-dependent translation initiation (Pilipenko *et al.*, 1995, 2001). Mutations in PTB-binding sites result in attenuation of TMEV neuropathogenicity. It has been shown that translation initiation in the presence of nPTB is more sensitive to distortion of binding sites. A specific double-stranded motif for PTB binding in TMEV was reported to be YRYRY (Y=C/U, R=G/A; Pilipenko *et al.*, 2001). Distortion of this motif completely eliminated PTB binding. For example, a mutation of the 4ds site of TMEV (^{636}CACGC^{640}; nucleotide positions according to TMEV strain GDVII; GenBank accession no. NC_001366) to ^{636}CCCGC^{640} rendered the site non-functional (Pilipenko *et al.*, 2001). In all SLCV genomes, the sequence that structurally corresponds to the TMEV 4ds site is ^{637}UUCAC^{641} (numbering according to SafV-1, GenBank accession no. NC_009448). We can therefore suggest that this sequence cannot bind PTB. Similarly, the putative 6ds site in SLCV has a consensus sequence UGACA, which does not correspond to a proper sequence for a double-stranded PTB-binding motif in TMEV. One virus, D/VI2229/2004, had an additional alteration in the 6ss site. Alterations in any two PTB-binding sites of TMEV result in greatly attenuated neurovirulence (Pilipenko *et al.*, 1995, 2001). In agreement with this, SLCV has not been associated with acute flaccid paralysis and was not found in 400 cerebrospinal fluid specimens from patients with aseptic meningitis, encephalitis and multiple sclerosis (Blinkova *et al.*, 2009; Chiu *et al.*, 2008). Studies on PTB binding to the TMEV IRES were conducted using rabbit PTB and mouse nPTB, and neurovirulence studies were conducted in mice (Pilipenko *et al.*, 2001). So far, there has been a perfect correspondence of results obtained using different host species. However, more experimental data are needed to see how well results from mouse neuroinfection models with TMEV can be transferred to SLCV.

Recombination analysis

For other picornaviruses, including cardioviruses, recombination is a common mechanism of evolution and antigenic variability (Blinkova *et al.*, 2009; Liang *et al.*, 2008). To compare the extent of recombination between SLCV and related picornaviruses, a combined analysis was carried out for theiloviruses, including TMEV and SLCV. Recombination was detected without prior assumptions of phylogenetic ancestry by phylogenetic incompatibility analysis (Simmonds & Welch, 2006). Likely recombination events were detected mostly between structural and non-structural genome regions (Fig. 2a). Analyses carried out separately on datasets containing only TMEV or only SLCV gave results very similar to each other and to the combined matrix (Fig. 2b, c), suggesting similar propensities for recombination in TMEV and SLCV.

To obtain a more specific positioning of recombination breakpoints in different genome regions, bootscan analysis was conducted. Multiple recombination events were detected in the junction region between the structural and non-structural genome portion, as exemplified in Fig. 2(d). Several recombination events detected by phylogenetic compatibility analysis mapped to the non-structural protein region and involved viruses infecting the same host species. These could be confirmed by bootscan analysis (e.g. Fig. 2e). Such a recombination pattern is very similar to that observed in human enteroviruses (Lukashev *et al.*, 2005; Simmonds & Welch, 2006). Interestingly, some recombination events were found in the capsid-encoding genome region. The clearest of these cases involved viruses within the same SLCV type. The German SLCV type 2 isolate D/VI2229/2004 was recombinant relative to Brazilian and American genotype 2 isolates (Fig. 2f). The crossing point in this case mapped approximately to position 1800 of the genome, i.e. the VP2 capsid protein gene. In comparative bootscan analyses among murine theiloviruses, similar intratypic recombination events in the capsid protein gene were frequently detected (data not shown).

We created Bayesian phylogenetic trees for different genome regions to confirm the detected recombination events (Fig. 3). Trees for the VP4–VP2 and VP1 genome regions were created using the 3′ and 5′ terminal 1000 nt of the P1 genome region, respectively. In the VP4–VP2 genome region, strain D/VI2229/2004 (a type 2 virus by VP1-based classification) was an outgroup against other type 2 SLCV with high posterior probability. In VP1, this virus grouped, albeit less reliably, with Brazilian isolate BR/118/2006 (also a type 2 virus by VP1-based classification), confirming the results of the bootscan analysis (Fig. 2f). In this case, recombination in the P1 genome region was confined to the same type.

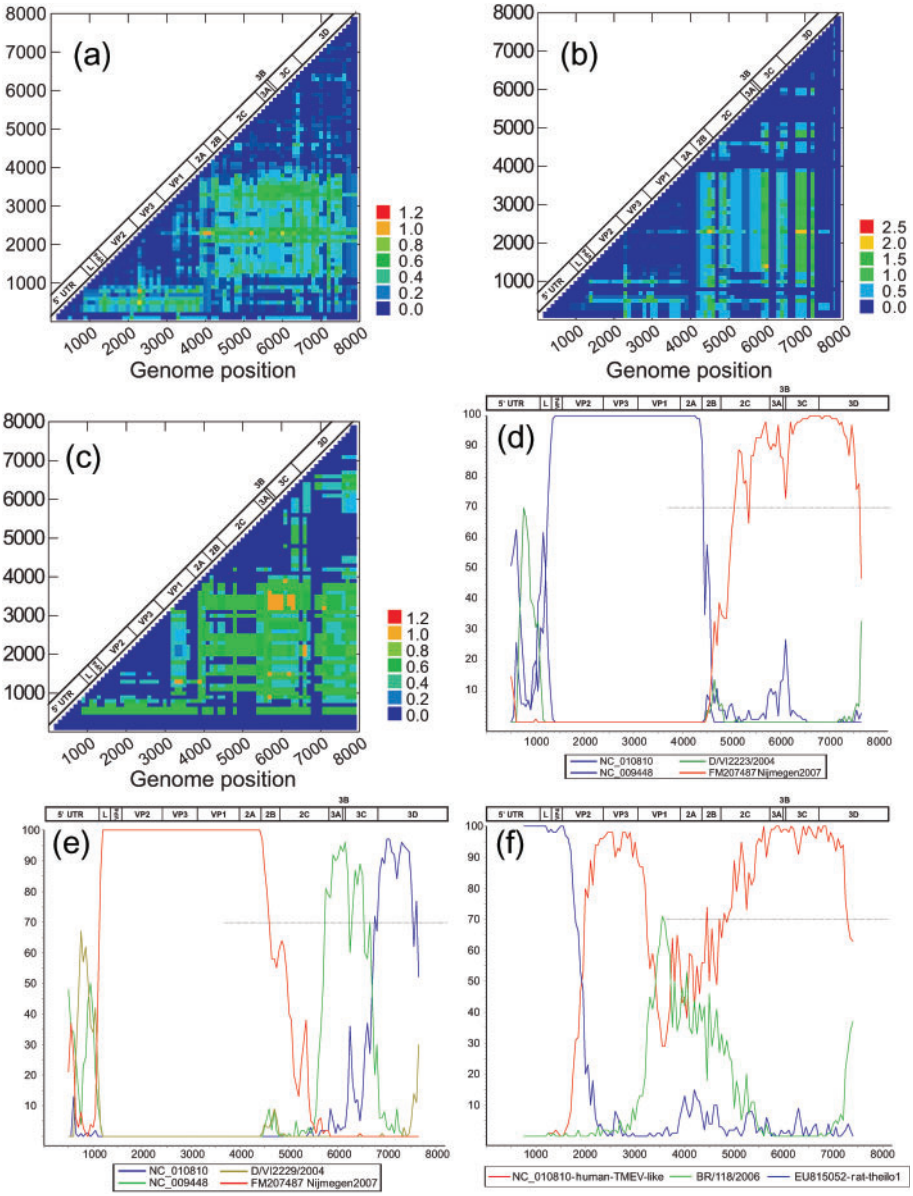

Fig. 2. Recombination analysis of cardioviruses. (a–c) Phylogenetic compatibility matrix of all TMEV-like viruses (a), SafVs (b) and mouse TMEVs (c). Colours indicate the number of phylogeny violations per clade required to order the trees generated from each combination of fragments (*x*- and *y*-axis). Window, 250 nt; step, 100 nt; 70 % bootstrap cut-off. (d–f) Bootscan graphs of SLCVs. (d) Typical recombination event on the border between structural and non-structural genome regions. Query strain, D/VI2229/2004; window, 1000 nt. (e) Typical recombination event within the non-structural genome region. Query strain, D/VI2223/2004; window, 1000 nt. (f) Intratypic recombination in the structural genome region of strain D/VI2229/2004; window, 1500 nt. The *x*-axis indicates the genome position of the sliding window centre, and the *y*-axis shows the percentage of bootstrap pseudoreplicates that support grouping of the query sequence with other cardioviruses. The dotted line indicates a 70 % bootstrap cut-off.

There were also multiple phylogenetic conflicts between the VP4–VP2 and VP1 genome regions involving other types of SafVs. These discrepancies did not affect the integrity of groups but rather the relationships of different types to each other. For example, the SafV-3 group clustered reliably with the SafV-2 group in VP1, but was an outgroup among SLCVs in VP2–VP4. It is quite likely that such recombination events occurred only occasionally and facilitated the emergence of different virus types, as contemporary SafV types are phylogenetically distinct (Fig. 3) (Blinkova *et al.*, 2009). Phylogenetic tree incongruence between the VP4–VP2 and VP1 genome regions

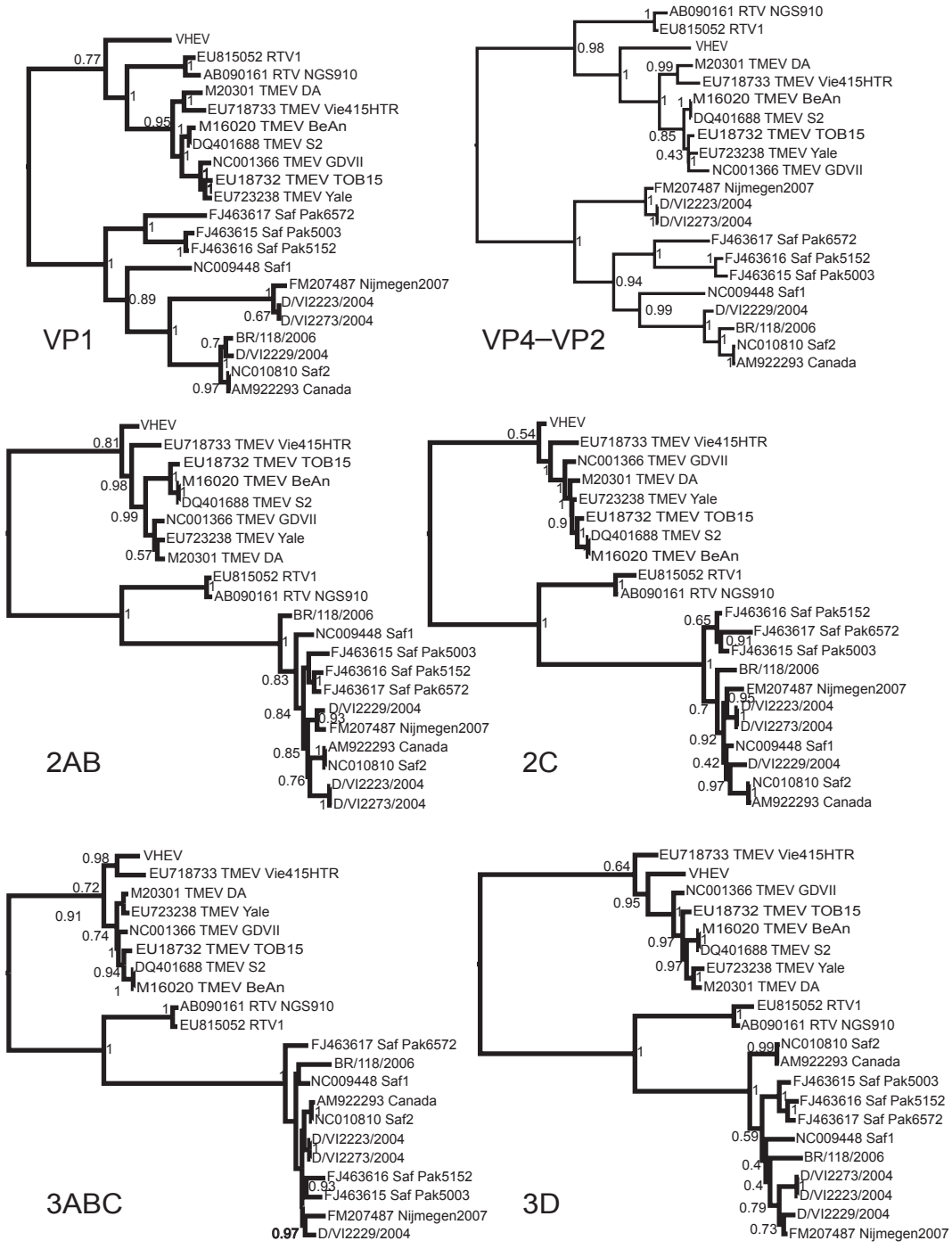

Fig. 3. Phylogenetic relationship of murine, rat and human theiloviruses in six genome regions. Trees were built basing on nucleotide sequences using the Bayesian algorithm implemented in MrBayes v.3.1.2 software (Ronquist & Huelsenbeck, 2003). Numbers at tree nodes are posterior probabilities. For readability, several insignificant posterior values were removed from the tree. EMCV was included in the alignments to provide correct tree rooting, but is omitted from the figure.

was also common among murine theiloviruses, suggesting recombination events in the capsid-encoding genome region. These findings confirmed low but detectable phylogenetic conflict within the P1 genome region detected by compatibility matrices (Fig. 2a–c).

Bayesian phylogenetic trees were also created for different parts of the non-structural region of the genome (Fig. 3). Regions 2AB, 2C, 3ABC and 3D were selected to have comparable length (~800–1400 nt). There were multiple significant phylogenetic conflicts among these genome

regions in both human and murine cardioviruses. Only groups of almost identical viruses (Saf2 NC_010810 and AM922293; D/VI2223/2004 and D/VI2273/2004) were supported throughout the non-structural part of the genome. In one case, a group of three cardioviruses identified in Pakistan was supported perfectly in the 2AB and 3D genome regions, less reliably in 2C and was distorted in 3ABC, suggesting multiple recombination events or mosaic recombination. Therefore, it can be concluded that recombination in cardioviruses is highly prevalent in the non-structural part of the genome. Importantly, recombination in the non-structural protein genome region of cardioviruses occurred only within groups of murine and human viruses, and never between human and mouse viruses.

Substitution rate analysis

Along with recombination, a high substitution rate is a key feature of RNA virus evolution. Frequent recombination and pseudo-reversions may undermine the validity of substitution rate analysis. Therefore, we selected two pairs of SLCVs (NC_010810 and AM922293; D/VI2223/2004 and D/VI2273/2004) that were highly similar over the whole genome and that did not show any signs of recombination in phylogenetic analysis (Fig. 3). These sequence pairs could therefore be used to estimate ratios of synonymous and non-synonymous substitutions in different genome portions (Table 1). When pairs of closely related sequences were analysed, both synonymous and non-synonymous variation in structural and non-structural genome regions were very similar.

Among all SLCVs, non-synonymous sequence variation per site was increased about tenfold in P1 compared with the P2+P3 genome regions (Table 1, Fig. 4a). The higher level of amino acid sequence variation in the P1 genome region can be attributed to high variability among serotypes. The synonymous variation was higher in P1 than in the P2+P3 genome regions, in contrast to human enterovirus B (HEV-B), which exhibited similar (and higher) rates of synonymous variation and nucleotide sequence diversity over the genome (Fig. 4b). Similar observations were found for murine TMEVs (data not shown). The reason for lower synonymous sequence variation in different genome regions of SLCV is not clear. Several explanations may be plausible for this phenomenon. Firstly, cardioviruses might have emerged recently and not had time to diverge. This is unlikely considering the substantial and constantly growing number of known distinct genotypes and the global distribution of both human and murine theiloviruses. Secondly, RNA secondary structure in coding regions could restrict variation in cardioviruses. Many RNA viruses from different families have been suggested, by means of mathematic modelling, to contain a genome-scale ordered RNA structure (GORS) (Simmonds *et al.*, 2004), which was later confirmed by experimental evidence (Davis *et al.*, 2008). It was suggested that such structures could protect the virus from RNases or shield it from recognition by intracellular defence systems. The highest level of GORS was correlated with persistent infection. The murine cardiovirus genome was predicted to contain a moderate level of GORS, whilst enteroviruses were suggested to have a low level. We tested the level of secondary structure conservation in TMEV, SLCV and HEV-B to see how well SLCV corresponded to TMEV. The folding energy of TMEV RNA differed from the mean folding energy of its scrambled replicates by -1.5 SD (Z-score) and the Z-score of enterovirus RNA was close to 0, as reported previously (Simmonds *et al.*, 2004). SLCV strain D/VI2223/2004 produced a Z-score of 0.4, which corresponds to the absence of a conserved secondary structure. Therefore, lower RNA sequence divergence in the non-structural genome region of cardioviruses cannot be explained by the presence of GORS and requires further study.

Conclusions

Taking all the data from this study, we suggest a hypothesis that might at least in part explain the limited sequence

Table 1. Nucleotide sequence diversity per site calculated with DNAsp (Librado & Rozas, 2009)

Pi, Pairwise diversity, Syn, synonymous substitutions; NS, non-synonymous substitutions.

Virus group	P1 genome region	P2+P3 genome region
All human SafVs	Pi (Syn)=0.648	Pi (Syn)=0.458
	Pi (NS)=0.115	Pi (NS)=0.013
	NS/Syn=0.177	NS/Syn=0.028
D/VI2223/2004 and D/VI2273/2004	Pi (Syn)=0.037	Pi (Syn)=0.027
	Pi (NS)=0.0010*	Pi (NS)=0.0019
	NS/Syn=0.02*	NS/Syn=0.07
NC_010810 (USA) and Can112051-06	Pi (Syn)=0.040	Pi (Syn)=0.041
	Pi (NS)=0.0026	Pi (NS)=0.0044
	NS/Syn=0.065	NS/Syn=0.107

*This value is unreliable because it results from a single substitution.

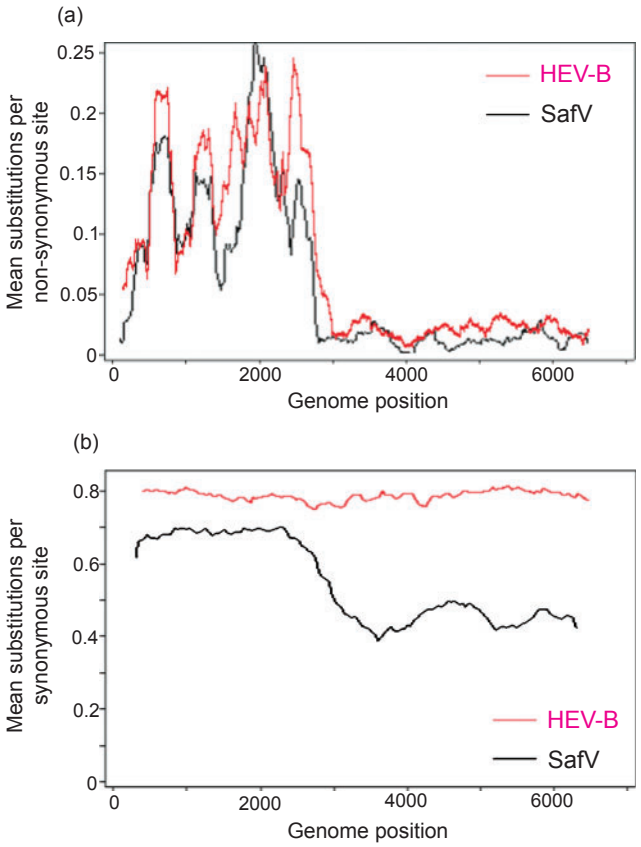

Fig. 4. (a) Nucleotide sequence diversity per non-synonymous site in SafVs and human enterovirus B (HEV-B) over the genome (VP4–3D regions) computed with DNAsp (Librado & Rozas, 2009). Window, 200 nt; step, 10 nt positions. (b) Nucleotide sequence diversity per synonymous site in SafVs and HEV-B over the genome (VP4–3D regions) computed with DNAsp. Window, 200 nt; step, 10 nt positions. All available SLCV genomes were used. To prevent bias from unequal sampling, 11 HEV-B genomes were chosen randomly for analysis to match cardiovirus selection in number and abundance of serotypes.

variation within cardiovirus groups. Recombination plays a key role in the microevolution of picornaviruses and similar viruses with a high polymerase error rate. Each new picornavirus genome contains on average one new mutation, most of which are deleterious (Drake, 1993; Ward & Flanegan, 1992). Recombination may thus be thought of as a mechanism of recreation of functional genomes from several impaired ones. On a larger scale, frequent recombination facilitates the independent evolution of picornavirus genome fragments. New enterovirus recombinant forms emerge very frequently (Leitch *et al.*, 2009), and recombination has been viewed mainly as a mechanism that contributes to the diversity of picornaviruses. We suggest that, on a global scale, constant recombination also acts as a stabilizing force that preserves the global species integrity. In line with the concept of independent evolution of different genome regions, in the structural genome region, recombination conserves (sero)-type sequence by shuffling structural genome regions of the same type, and in the non-structural genome region, frequent recombination averages the global (worldwide) consensus sequence of a species.

In addition to this report, earlier studies on the genus *Enterovirus* have shown that recombination is strictly intraspecific (Lukashev *et al.*, 2003; Oberste *et al.*, 2004; Simmonds & Welch, 2006). The presence or absence of natural recombination may therefore be used as an additional criterion for delimiting *Enterovirus* species (Lukashev, 2005). Our study justifies an extension of this principle to the genus *Cardiovirus*, suggesting that the three cardiovirus groups (TMEV, rat TMEV and SLCV) represent distinct species.

METHODS

Genome sequencing. The SLCVs sequenced here were identified previously in stool samples from 844 patients sampled between 2004 and 2006 with clinical symptoms of diarrhoea (Drexler *et al.*, 2008). Based on the published genome of SafV (GenBank accession no. EF165067) and murine cardioviruses, primers spanning the complete genome were designed. cDNA was produced using a Superscript III kit (Invitrogen), followed by amplification of ~200 bp fragments located at the 5′ and 3′ genome termini and in the central genomic 2A region. Following nucleotide sequencing of these PCR products, specific primers were designed for each virus and the whole genome was amplified from clinical material in two overlapping ~4 kb amplicons using an Expand High Fidelity^{PLUS} kit (Roche). These PCR products were sequenced directly on both strands by primer walking. The 3′ terminus was amplified using a GeneRacer kit (Invitrogen).

Genetic analyses. All complete or semi-complete TMEV, RTV and SLCV sequences available in GenBank (Table 2) were aligned using CLUSTAL W (Thompson *et al.*, 1994). One EMCV sequence was included in alignments intended for phylogenetic analysis. Alignments of the coding genome regions based on translation were joined to 5′NTR sequences aligned separately. 3′NTR sequences were omitted from the analysis because they were lacking in four out of seven of the complete SLCV sequences available in GenBank. Sequence handling was performed with BioEdit v.7.0.5.2 software (Hall, 1999). Nucleotide sequence diversity was calculated using DNAsp v.5 (Librado & Rozas, 2009). Substitution rates per site were calculated with MEGA v.4.0 (Tamura *et al.*, 2007). All positions containing gaps and missing data were eliminated from the dataset for phylogenetic analysis. Phylogenies were calculated using a Bayesian likelihood-based algorithm implemented in MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). A general time reversible (GTR) substitution model with six possible nucleotide state transitions was chosen, assuming a gamma-shaped distribution of rate variation across sites with the expectation of invariable sites. Each analysis was run over 10^6 generations and trees were sampled every 100 generations, resulting in 10^4 final trees. Trees were annotated with TreeAnnotator v.1.4.8 from the BEAST package (Drummond & Rambaut, 2007) with a 2000-tree burn-in and visualized with FigTree v.1.1.1 (http://tree.bio.ed.ac.uk/software/figtree/). Similarity plot and bootscanning analyses were performed using SimPlot v.3.5.1 (Lole *et al.*, 1999). Phylogenetic compatibility matrices were created using Simmonics v.1.6 (Simmonds & Smith, 1999). Three sequences that were very highly similar to each other and did not show signs of recombination in preliminary tests (AM922293, EU376394 and D/VI2273/2004) were omitted from this analysis. Conservation of

Table 2. Cardiovirus sequences used in the analysis

Virus	Strain	GenBank accession no.	Reference	Note
EMCV		NC_001479	Duke *et al.* (1992)	Used as an outgroup
TMEV	GDVII	NC_001366	Law & Brown (1990)	
TMEV	Yale	EU723238	Liang *et al.* (2008)	
TMEV	Vie415HTR	EU718733	Liang *et al.* (2008)	
TMEV	TOB15	EU718732	Liang *et al.* (2008)	
TMEV	S2	DQ401688	Myoung *et al.* (2007)	
TMEV	DA	M20301	Ohara *et al.* (1988)	
TMEV	BeAn	M16020	Pevear *et al.* (1987)	
TMEV	VHEV	EU723237, M80888, M94868	Liang *et al.* (2008); Pritchard *et al.* (1992)	
RTV	RTV1	EU815052	Drake *et al.* (2008)	
RTV	NGS910	AB090161	–	
SafV-1		NC_009448	Jones *et al.* (2007)	
SafV-2		NC_010810	Chiu *et al.* (2008)	Identical to EU376394
SafV-2	Can112051-06	AM922293	Abed & Boivin (2008)	No 5′NTR
SafV-6	Pak6572	FJ463617	Blinkova *et al.* (2009)	Up to 102 3′ terminal aa lacking
SafV-5	Pak5152	FJ463616	Blinkova *et al.* (2009)	Up to 102 3′ terminal aa lacking
SafV-5	Pak5003	FJ463615	Blinkova *et al.* (2009)	Up to 102 3′ terminal aa lacking
SafV-3	Nijmegen2007	FM207487	Zoll *et al.* (2009)	
SafV-2	D/VI2229/2004	EU681176	This study	
SafV-3	D/VI2223/2004	EU681179	This study	
SafV-3	D/VI2273/2004	EU681178	This study	
SafV-2	BR/118/2006	EU681177	This study	

RNA secondary structure was carried out as described by Simmonds *et al.* (2004). Briefly, RNA sequence was shuffled using the CDLR method implemented in Simmonics. Folding energies of the original and scrambled sequences were calculated using the Zipfold online service (Zuker, 2003). The *Z*-score was taken as a measure of structure conservation. Differences in original sequence folding energies compared with mean folding energies of scrambled sequences were expressed as SD.

Analysis of 5′NTR secondary structure was performed manually based on published reports of enzymic probing (Pilipenko *et al.*, 2001) and base conservation among SLCVs.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (grant no. DR772/2-1) and the Russian Foundation for Basic Research. Additional support was obtained from European Union projects RiViGene, EVA and EMPERIE.

REFERENCES

Abed, Y. & Boivin, G. (2008). New Saffold cardioviruses in 3 children, Canada. *Emerg Infect Dis* **14**, 834–836.

Blinkova, O., Kapoor, A., Victoria, J., Jones, M., Wolfe, N., Naeem, A., Shaukat, S., Sharif, S., Alam, M. M. & other authors (2009). Cardioviruses are genetically diverse and cause common enteric infections in South Asian children. *J Virol* **83**, 4631–4641.

Chen, H. H., Kong, W. P., Zhang, L., Ward, P. L. & Roos, R. P. (1995). A picornaviral protein synthesized out of frame with the polyprotein plays a key role in a virus-induced immune-mediated demyelinating disease. *Nat Med* **1**, 927–931.

Chiu, C. Y., Greninger, A. L., Kanada, K., Kwok, T., Fischer, K. F., Runckel, C., Louie, J. K., Glaser, C. A., Yagi, S. & other authors (2008). Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. *Proc Natl Acad Sci U S A* **105**, 14124–14129.

Davis, M., Sagan, S. M., Pezacki, J. P., Evans, D. J. & Simmonds, P. (2008). Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. *J Virol* **82**, 11824–11836.

Drake, J. W. (1993). Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci U S A* **90**, 4171–4175.

Drake, M. T., Riley, L. K. & Livingston, R. S. (2008). Differential susceptibility of SD and CD rats to a novel rat theilovirus. *Comp Med* **58**, 458–464.

Drexler, J. F., Luna, L. K., Stocker, A., Almeida, P. S., Ribeiro, T. C., Petersen, N., Herzog, P., Pedroso, C., Huppertz, H. I. & other authors (2008). Circulation of 3 lineages of a novel Saffold cardiovirus in humans. *Emerg Infect Dis* **14**, 1398–1405.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Duke, G. M., Hoffman, M. A. & Palmenberg, A. C. (1992). Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J Virol* **66**, 1602–1609.

Gajdusek, D. C., Brown, P., Petrov, P. A. & Brown, P. (1970). V. Vilyuisk encephalitis in the Yakut Republic (U.S.S.R.). *Am J Trop Med Hyg* **19**, 146–150.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Harvala, H., Robertson, I., Chieochansin, T., McWilliam Leitch, E. C., Templeton, K. & Simmonds, P. (2009). Specific association of human parechovirus type 3 with sepsis and fever in young infants, as

identified by direct typing of cerebrospinal fluid samples. *J Infect Dis* **199**, 1753–1760.

Jones, M. S., Lukashov, V. V., Ganac, R. D. & Schnurr, D. P. (2007). Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. *J Clin Microbiol* **45**, 2144–2150.

Kirkland, P. D., Gleeson, A. B., Hawkes, R. A., Naim, H. M. & Boughton, C. R. (1989). Human infection with encephalomyocarditis virus in New South Wales. *Med J Aust* **151**, 176–178.

Law, K. M. & Brown, T. D. (1990). The complete nucleotide sequence of the GDVII strain of Theiler's murine encephalomyelitis virus (TMEV). *Nucleic Acids Res* **18**, 6707–6708.

Leitch, E. C., Harvala, H., Robertson, I., Ubillos, I., Templeton, K. & Simmonds, P. (2009). Direct identification of human enterovirus serotypes in cerebrospinal fluid by amplification and sequencing of the VP1 region. *J Clin Virol* **44**, 119–124.

Liang, Z., Kumar, A. S., Jones, M. S., Knowles, N. J. & Lipton, H. L. (2008). Phylogenetic analysis of the species Theilovirus: emerging murine and human pathogens. *J Virol* **82**, 11545–11554.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.

Lu, C. Y., Lee, C. Y., Kao, C. L., Shao, W. Y., Lee, P. I., Twu, S. J., Yeh, C. C., Lin, S. C., Shih, W. Y. & other authors (2002). Incidence and case-fatality rates resulting from the 1998 enterovirus 71 outbreak in Taiwan. *J Med Virol* **67**, 217–223.

Lukashev, A. N. (2005). Role of recombination in evolution of enteroviruses. *Rev Med Virol* **15**, 157–167.

Lukashev, A. N., Lashkevich, V. A., Ivanova, O. E., Koroleva, G. A., Hinkkanen, A. E. & Ilonen, J. (2003). Recombination in circulating enteroviruses. *J Virol* **77**, 10423–10431.

Lukashev, A. N., Lashkevich, V. A., Ivanova, O. E., Koroleva, G. A., Hinkkanen, A. E. & Ilonen, J. (2005). Recombination in circulating enterovirus B: independent evolution of structural and non-structural genome regions. *J Gen Virol* **86**, 3281–3290.

Michiels, T., Jarousse, N. & Brahic, M. (1995). Analysis of the leader and capsid coding regions of persistent and neurovirulent strains of Theiler's virus. *Virology* **214**, 550–558.

Myoung, J., Hou, W., Kang, B., Lyman, M. A., Kang, J. A. & Kim, B. S. (2007). The immunodominant $CD8^+$ T cell epitope region of Theiler's virus in resistant C57BL/6 mice is critical for anti-viral immune responses, viral persistence, and binding to the host cells. *Virology* **360**, 159–171.

Oberste, M. S., Maher, K. & Pallansch, M. A. (2004). Evidence for frequent recombination within species *Human enterovirus B* based on complete genomic sequences of all thirty-seven serotypes. *J Virol* **78**, 855–867.

Ohara, Y., Stein, S., Fu, J. L., Stillman, L., Klaman, L. & Roos, R. P. (1988). Molecular cloning and sequence determination of DA strain of Theiler's murine encephalomyelitis viruses. *Virology* **164**, 245–255.

Pallansch, M. A. & Roos, R. P. (2001). Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses. In *Fields Virology*, 4th edn, pp. 724–743. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott-Raven.

Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I. & Hellen, C. U. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci U S A* **98**, 7029–7036.

Pevear, D. C., Calenoff, M., Rozhon, E. & Lipton, H. L. (1987). Analysis of the complete nucleotide sequence of the picornavirus Theiler's murine encephalomyelitis virus indicates that it is closely related to cardioviruses. *J Virol* **61**, 1507–1516.

Pilipenko, E. V., Blinov, V. M., Romanova, L. I., Sinyakov, A. N., Maslova, S. V. & Agol, V. I. (1989). Conserved structural domains in the 5′-untranslated region of picornaviral genomes: an analysis of the segment controlling translation and neurovirulence. *Virology* **168**, 201–209.

Pilipenko, E. V., Gmyl, A. P., Maslova, S. V., Belov, G. A., Sinyakov, A. N., Huang, M., Brown, T. D. & Agol, V. I. (1994). Starting window, a distinct element in the cap-independent internal initiation of translation on picornaviral RNA. *J Mol Biol* **241**, 398–414.

Pilipenko, E. V., Gmyl, A. P., Maslova, S. V., Khitrina, E. V. & Agol, V. I. (1995). Attenuation of Theiler's murine encephalomyelitis virus by modifications of the oligopyrimidine/AUG tandem, a host-dependent translational cis element. *J Virol* **69**, 864–870.

Pilipenko, E. V., Viktorova, E. G., Guest, S. T., Agol, V. I. & Roos, R. P. (2001). Cell-specific proteins regulate viral RNA translation and virus-induced disease. *EMBO J* **20**, 6899–6908.

Pritchard, A. E., Calenoff, M. A., Simpson, S., Jensen, K. & Lipton, H. L. (1992). A single base deletion in the 5′ noncoding region of Theiler's virus attenuates neurovirulence. *J Virol* **66**, 1951–1958.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Sarmanova, E. S. & Chumachenko, G. G. (1960). Etiologic studies of Viliuisk encephalomyelitis. 3. Serologic studies. *Curr Probl Med Virol* **1**, 216–219.

Simmonds, P. & Smith, D. B. (1999). Structural constraints on RNA virus evolution. *J Virol* **73**, 5787–5794.

Simmonds, P. & Welch, J. (2006). Frequency and dynamics of recombination within different species of human enteroviruses. *J Virol* **80**, 483–493.

Simmonds, P., Tuplin, A. & Evans, D. J. (2004). Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. *RNA* **10**, 1337–1351.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Tauriainen, S., Martiskainen, M., Oikarinen, S., Lonnrot, M., Viskari, H., Ilonen, J., Simell, O., Knip, M. & Hyoty, H. (2007). Human parechovirus 1 infections in young children – no association with type 1 diabetes. *J Med Virol* **79**, 457–462.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

van Eyll, O. & Michiels, T. (2000). Influence of the Theiler's virus L* protein on macrophage infection, viral persistence, and neurovirulence. *J Virol* **74**, 9071–9077.

Vladimirtsev, V. A., Nikitina, R. S., Renwick, N., Ivanova, A. A., Danilova, A. P., Platonov, F. A., Krivoshapkin, V. G., McLean, C. A., Masters, C. L. & other authors (2007). Family clustering of Viliuisk encephalomyelitis in traditional and new geographic regions. *Emerg Infect Dis* **13**, 1321–1326.

Ward, C. D. & Flanegan, J. B. (1992). Determination of the poliovirus RNA polymerase error frequency at eight sites in the viral genome. *J Virol* **66**, 3784–3793.

Witso, E., Palacios, G., Cinek, O., Stene, L. C., Grinde, B., Janowitz, D., Lipkin, W. I. & Ronningen, K. S. (2006). High prevalence of human enterovirus A infections in natural circulation of human enteroviruses. *J Clin Microbiol* **44**, 4095–4100.

Wolthers, K. C., Benschop, K. S., Schinkel, J., Molenkamp, R., Bergevoet, R. M., Spijkerman, I. J., Kraakman, H. C. & Pajkrt, D. (2008). Human parechoviruses as an important viral cause of sepsislike illness and meningitis in young children. *Clin Infect Dis* **47**, 358–363.

Zoll, J., Erkens Hulshof, S., Lanke, K., Verduyn Lunel, F., Melchers, W. J., Schoondermark-van de Ven, E., Roivainen, M., Galama, J. M. & van Kuppeveld, F. J. (2009). Saffold virus, a human Theiler's-like cardiovirus, is ubiquitous and causes infection early in life. *PLoS Pathog* **5**, e1000416.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406–3415.