Genetic diversity of circulating Saffold viruses in Pakistan and Afghanistan

Asif Naeem, Takushi Hosomi, Yorihiro Nishimura, Muhammad Masroor Alam, Tomoichiro Oka, Syed Sohail Zahoor Zaidi and Hiroyuki Shimizu

1Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan
2The Meat Inspection Center of Kochi Prefecture, Kochi, Japan
3Department of Virology, National Institute of Health, Islamabad, Pakistan

Human cardioviruses or Saffold viruses (SAFVs) of the family Picornaviridae are newly emerging viruses whose genetic and phenotypic diversity are poorly understood. We report here the full genome sequence of 11 SAFV genotypes from Pakistan and Afghanistan, along with a re-evaluation of their genetic diversity and recombination. We detected 88 SAFV from stool samples of 943 acute flaccid paralysis cases using reverse transcriptase-PCR targeting the 5' untranslated region (UTR). Further characterization based on complete VP1 analysis revealed 71 SAFVs belonging to 11 genotypes, including three previously unidentified genotypes. SAFV showed high genetic diversity and recombination based on phylogenetic, pairwise distance distributions and recombination mapping analyses performed herein. Phylogenies based on non-structural and UTRs were highly incongruent indicating frequent recombination events among SAFVs. We improved the SAFV genotyping classification criteria by determining new VP1 thresholds based on the principles used for the classification of enteroviruses. For genotype assignment, we propose a threshold of 23 and 10% divergence for VP1 nucleotide and amino acid sequences, respectively. Other members of the species Theilovirus, such as Thera virus and Theiler’s murine encephalomyelitis virus, are difficult to classify in the same species as SAFV, because they are genetically distinct from SAFV, with 41–56% aa pairwise distances. The new genetic information obtained in this study will improve our understanding of the evolution and classification of SAFV.

INTRODUCTION

The family Picornaviridae is one of the largest positive-stranded RNA virus families, currently composed of 46 species grouped into 26 genera (Adams et al., 2013; Knowles et al., 2012). The genus Cardiovirus currently consists of two species, Theilovirus and Encephalomyocarditis virus. Cardiovirus has been associated with myocarditis, encephalitis and demyelinating disease in rodents (Brahic et al., 2005; Liang et al., 2008). Encephalomyocarditis virus (EMCV) is widely used as an experimental model for human diseases, such as myocarditis, encephalitis and pancreatitis in rodents. Theiler’s murine encephalomyelitis virus (TMEV), a prototype of the species Theilovirus, is divided into two subgroups: GDVII subgroup strains are highly virulent, whereas Thelier’s original subgroup strains are used as a mouse model for the human demyelinating disease multiple sclerosis (Oleszak et al., 2004; Roos, 2010). A genetically divergent cardiovirus was isolated from a sentinel rat housed with TMEV-seropositive rats in Japan in 1991, and was identified as Thera virus (TRV) (Ohsawa et al., 2003). In 1954, viral isolates were recovered from human clinical specimens in Siberia; these isolates were thought to be linked to a human form of encephalomyelitis called Vilyuisk encephalitis (VE). These isolates of Vilyuisk human encephalomyelitis virus (VHEV) cross-reacted completely with TMEV and weakly with EMCV (Casals, 1963). It remains unclear whether the virus was the human pathogen causing VE, or a possible TMEV contaminant (Pritchard et al., 1992).

Saffold virus (SAFV) was first isolated in WI-38 cells from an archived stool sample of a febrile child (Jones et al., 2007). The initial report was followed by a number of clinical and epidemiological studies signifying the prevalence of SAFV in human populations (Blinkova et al., 2009; Chiu et al., 2008; Drexler et al., 2008; Itagaki et al., 2010, 2011).

The GenBank/EMBL/DDBJ accession numbers determined in this study for SAFV-1 to SAFV-11 are AB747248–AB747258 (complete genomes) and AB747177–AB747247 (VP1), respectively.

Two supplementary figures and two supplementary tables are available with the online version of this paper.
2011; Itagaki et al., 2011; Ren et al., 2009, 2013; Tsukagoshi et al., 2010, 2011; Xu et al., 2009; Zoll et al., 2009). Most of the studies have described the molecular detection of SAFV in children suffering from gastroenteritis (Drexler et al., 2010; Khamrin et al., 2011; Ren et al., 2009, 2013), respiratory tract infection (Itagaki et al., 2010, 2011; Tsukagoshi et al., 2010, 2011) and even cases of acute flaccid paralysis (AFP) (Blinkova et al., 2009). AFP is a complex clinical syndrome, with a wide variety of possible aetiologies and clinical manifestations such as Guillain–Barré syndrome and transverse myelitis (Dietz et al., 1995). Nationwide AFP surveillance is the gold standard for detecting cases of poliomyelitis (World Health Organization, 2012).

The SAFV genome is approximately 8050 nt whereas the 5’ untranslated region (UTR) is 1040 nt containing a type II internal ribosome entry site (IRES) like other members of the species (Drexler et al., 2010). The 3’-UTR is about 120 nt with a poly(A) tail. Phylogenetic analysis based on complete VP1 region suggests the presence of eight distinct SAFV genotypes; SAFV types 1–8 (SAFV-1 to SAFV-8, respectively) (Blinkova et al., 2009; Liang et al., 2008). SAFV-1, SAFV-2 and SAFV-3 are globally distributed, circulating in North and South America (Abed & Boivin, 2008), respiratory tract infection (Itagaki et al., 2010, 2011; Tsukagoshi et al., 2010, 2011) and even cases of acute flaccid paralysis (AFP) (Blinkova et al., 2009). SAFV is a complex clinical syndrome, with a wide variety of possible aetiologies and clinical manifestations such as Guillain–Barré syndrome and transverse myelitis (Dietz et al., 1995). Nationwide AFP surveillance is the gold standard for detecting cases of poliomyelitis (World Health Organization, 2012).

The SAFV genome is approximately 8050 nt whereas the 5’ untranslated region (UTR) is 1040 nt containing a type II internal ribosome entry site (IRES) like other members of the species (Drexler et al., 2010). The 3’-UTR is about 120 nt with a poly(A) tail. Phylogenetic analysis based on complete VP1 region suggests the presence of eight distinct SAFV genotypes; SAFV types 1–8 (SAFV-1 to SAFV-8, respectively) (Blinkova et al., 2009; Liang et al., 2008). SAFV-1, SAFV-2 and SAFV-3 are globally distributed, circulating in North and South America (Abed & Boivin, 2008); Chiu et al., 2008; Drexler et al., 2010; Jones et al., 2007), Europe (Galama et al., 2011; Nielsen et al., 2012; Zoll et al., 2009), China (Ren et al., 2009; Xu et al., 2009), Thailand (Khamrin et al., 2011), Malaysia (Chua et al., 2010), Japan (Itagaki et al., 2010, 2011; Tsukagoshi et al., 2010, 2011) and Australia (Wang et al., 2012), while SAFV-4 to SAFV-8 were previously found only in Afghanistan and Pakistan (Blinkova et al., 2009). Recently, SAFV-6 has also been found in patients with acute respiratory infections in Japan (Tsukagoshi et al., 2011).

SAFV is phylogenetically grouped with TRV, TMEV and VHEV in the species Thielovirus of the genus Cardiovirus (Liang et al., 2008). Traditionally, classification of picornaviruses has been based on both physical properties and serological relatedness. However, sequence-based phylogeny is becoming the gold standard for determining the species (Drexler et al., 2010). Species demarcation criteria for the genus Cardiovirus involves >70% amino acid sequence identity in the complete polyprotein region, >60% amino acid sequence identity in the structural (P1) region, >70% amino acid sequence identity in the 2C and 3CD region, common genome organization and shared natural host (Knowles et al., 2012). On the other hand, the classification of SAFV into a separate species based on the absence of recombination between SAFV and rodent cardioviruses has been proposed (Drexler et al., 2010).

Full genome and near full genome sequences are only available for SAFV-1 (Jones et al., 2007), SAFV-2 (Chiu et al., 2008), SAFV-3 (Zoll et al., 2009), SAFV-5 and SAFV-6 (Blinkova et al., 2009). In this study, we attempted to redefine the precise relationship between SAFV and other members of the species Thielovirus. We determined the complete genome sequence of 11 representative SAFV genotypes from AFP cases in Pakistan and Afghanistan. Genetic information on an expanded number of SAFV isolates enabled us to construct new phylogenies, measure genetic distances, and analyse recombination and geographical distribution. New phylogenies and sequence divergence of SAFV would improve the classification of SAFV in combination with previously published sequence data.

RESULTS

Detection of SAFV from AFP cases by PCR

Using full genome sequences, detailed investigation of the genetic diversity and recombination events in SAFV was performed; we screened SAFV-positive faecal specimens in 943 AFP cases in children aged <15 years through the AFP surveillance in Pakistan and Afghanistan in March–August 2009. Eighty-eight SAFV-positive samples were identified from poliovirus- and enterovirus-negative faecal specimens using reverse transcriptase-PCR (RT-PCR) based on the 5’-UTR. The SAFV-positive samples were further characterized based on the VP1 gene, which revealed 71 SAFV samples belonging to 11 different genotypes. SAFV in all the RT-PCR-positive faecal specimens did not grow in RD or L20B cell lines (data not shown).

Identification of SAFV genotypes based on complete VP1 sequence

The VP1 protein is considered to be the most diverse SAFV protein, and is known for its primary role in determining viral tropism and antigenicity among picornaviruses. We used complete VP1 sequences for new genotype designation based on the same criteria as mentioned previously for enteroviruses (Oberste et al., 1999). 5’-UTR screening followed by VP1-based nested RT-PCR was used for the characterization of the 71 positive-SAFV samples. The remaining 17 samples were negative in the VP1-based nested RT-PCR, presumably due to lower sensitivity of the VP1-based RT-PCR than the 5’-UTR-based one. Phylogenetic analysis of the VP1 region identified three major clusters (SAFV, TRV-TMEV and EMCV). Most of the SAFV strains clustered into eight distinct SAFV genotypes together with the previously known SAFV ones (Blinkova et al., 2009). The remaining eight SAFV strains were clustered into three new genotypes (SAFV-9, SAFV-10 and SAFV-11) (Fig. 1).

Pairwise amino acid and nucleotide distances against previously available SAFV VP1 sequences showed 11 different VP1-based genotypes according to existing genotype criteria (Blinkova et al., 2009). Two SAFV strains (Pak.2457 and Pak.3094) grouped together with a putative genotype 9 SAFV (SAFV-9) from Nigeria (Nig329 strain; GenBank FJ997532, see Table S2 available in the online Supplementary Material). The amino acid and nucleotide distance between SAFV-9 strains in Nigeria and Pakistan...
were 14 and 3%, respectively (Fig. 2). Similarly, we classified the Pak.2325 strain as an independent genotype, SAFV-10, based on the VP1 phylogeny and pairwise amino acid (26–39%) and nucleotide (30–37%) distances from SAFV-1 to SAFV-8. SAFV-10 was also clearly divergent from SAFV-9 both in amino acid (20%) and nucleotide (26%) pairwise distance (Fig. 2). The third new genotype, SAFV-11, had five different strains both from Pakistan and Afghanistan (Fig. 1) with amino acid (17–35%) and nucleotide (25–40%) distance from other SAFVs (Fig. 2).

All SAFV genotypes were mapped to the joint map of Pakistan and Afghanistan for the purpose of geographical distribution. This revealed the sharing of SAFV in both countries, together with the presence of multiple reservoirs, particularly in Pakistan. It also showed that the circulation of SAFV is not geographically restricted to a few areas; instead SAFV is widely distributed (Fig. S1).

**Pairwise distance distributions**

Extended sequence analysis of 11 SAFV genotypes in this study allowed us to re-examine the species and genotype classifications of SAFV. The distribution of pairwise distances between the complete VP1 sequences of 71 SAFV strains from this study and 57 previously known strains were analysed in order to determine whether they could be used for classification of SAFV. For that purpose, the VP1 sequence relationship within the genotype, between SAFV genotypes and rodent cardioviruses, was analysed by comparison of the nucleotide and deduced amino acid sequences of all possible sequence pairs. The relationships were visualized by plotting as a histogram the frequency of pairwise distance scores versus percentage distance (Fig. 2a, b).

For both the nucleotide (Fig. 2a) and amino acid (Fig. 2b) pairwise distance distributions, the scores fell into three major categories. The first peak containing the lowest score with nucleotide distance <23% and amino acid distance <10% reflected relationships between viruses of the homologous type or intra-genotypic distance. The second peak was defined by nucleotide distance scores for pairwise comparisons in the next distribution cluster, ranging from 25–40%, and this was clearly delineated from that of the heterologous genotypes. The last peak corresponded to the heterologous species in the genus *Cardiovirus* (Fig. 2a).

In the amino acid distance distribution (Fig. 2b), the first peak (1–10%) corresponded to the homologous genotypes. The next distribution had three different peaks (15–39%), corresponding to heterologous genotypes with heterologous clusters. The peak with higher scores (34–39%) represented comparisons of viruses from phylogenetically diverse clusters (e.g. SAFV-8 to SAFV-11), whereas the peak with lower scores (15–33%) represented comparisons of viruses from less divergent clusters (e.g. SAFV-3 to SAFV-7). The final peaks represented two species (*Theilovirus* and EMCV) of genus *Cardiovirus*, which appeared to be even more heterogeneous, with scores ranging from 41 to 56% amino acid pairwise distances.

**Complete genome sequence**

Although the full genome or near full genome sequences for SAFV types 1, 2, 3, 5 and 6 have been reported, genetic information on the remaining six SAFV genotypes is limited. To determine the full genome sequences for each of the 11 SAFV genotypes, we selected representative SAFV-positive stool samples that showed a high copy number of SAFV genomic RNA by real-time RT-PCR. The full genomes of all 11 SAFV genotypes were sequenced and named accordingly (Table S2). The availability of the complete genome sequences of all the members of the species *Theilovirus*, such as TMEV, TRV and SAFV, enabled us to perform and present here a comprehensive picture of all the cardioviruses as listed in Table S2.

First, phylogenetic analysis based on the entire capsid region (P1) confirmed that SAFV-1, -2, -3, -5 and -6 strains in this study clustered with previous SAFV strains. In contrast, the remaining six SAFV genotypes, SAFV-4 and SAFV-7 to -11, did not cluster with previous SAFV strains, which confirmed the reliability of the VP1-based genotype identification (Fig. 3a). Based on the phylogenetic analyses of the non-capsid regions, 11 SAFV strains from this study grouped together into a single cluster together with three previous SAFV strains from Pakistan (SAFV-5 and -6) in the P2 region (Fig. 3b). As for the P3 region, clustering of strains in the present study occurred with three previous Pakistani strains, along with SAFV-2 from Finland (Galama et al., 2011), thus suggesting an epidemiological relationship due to recombination (Fig. 3c). The phylogenetic incongruence within the SAFV cluster was quite high, suggesting possible recombination between different SAFV genotypes in the P2 and P3 regions with respect to geographical distribution. Similar patterns of clustering could not be seen in the case of the 5'UTR, 3'UTR and L peptide (Fig. 3d–f).

**Relationship between SAFV and other cardioviruses based on full genome sequence**

In order to re-examine the genetic relationship between rodent and human cardioviruses based on their full genome sequences, we conducted phylogenetic analyses of cardioviruses across different parts of the genome. Based on phylogenetic analyses of nucleotide and amino acid sequences, EMCV species (EMCV and mengoviruses) were found to be highly divergent from TMEV, TRV and SAFV, consistently in different genomic regions (P1, P2, P3, 5'UTR, 3'UTR and L protein) and consisting of a single genetic cluster (Fig. 3). Apart from EMCV, phylogenetic analysis of the 5’UTR, L protein, P1, P2, P3 and 3’UTR showed increased genetic diversity of SAFV in a large cluster, separate from TMEV and TRV (Fig. 3a–f). The distribution of different SAFV
genotypes and other cardioviruses in clusters varied depending on the part of genome that was analysed. Phylogenetic analysis of the P2 and P3 non-structural protein coding regions both for the nucleotide (Fig. 3b, c) and the deduced amino acid sequences (Fig. S2B, C) led to the identification of three clusters, each containing TMEV, TRV and SAFV. The presence of congruent phylogenies of the three clusters suggested the absence of any recombination among the members of these clusters, even in the P2 and P3 regions. In the case of the deduced amino acid sequences corresponding to L protein, SAFV fell into two major clusters, whereas TMEV and TRV fell into separate individual clusters (Fig. S2D), but the distribution of SAFV and other cardioviruses was different in the case of L protein.

**Fig. 1.** Evolutionary relationships of taxa based on complete VP1 sequence. Phylogenetic relationships between cardioviruses and SAFV based on alignments containing nucleotide (a) and amino acid (b) sequences of the complete VP1 region. All 71 SAFV complete VP1 sequences determined in this study are shown in red, while previously available sequences are shown in black. The tree was reconstructed by neighbour joining of 1000 bootstrap pseudo-replicates (values less than 70% are shown) with Kimura two-parameter pairwise distances. Bars represent the number of substitutions per site.

**Fig. 2.** Analysis of complete VP1 sequence relationships based on pairwise distance distributions calculated by the p-distance model: available cardiovirus nucleotide (a) and amino acid (b) sequences. VP1 sequence relationships within a genotype, between SAFV genotypes and other cardioviruses are shown. The relationships are visualized by plotting the frequency of pairwise distance scores versus percentage distance as a histogram. Coloured arrows represent the newly proposed genotype and species thresholds based on complete VP1 region.
phylogeny based on nucleotide sequences. Here, SAFV was distributed into six different clusters, regardless of genotype designation (Fig. S2D).

In order to see whether SAFV is sufficiently divergent from the existing theiloviruses to qualify as a separate species within the genus Cardiovirus, we followed the criteria set by the Picornaviridae study group of the International Committee on Taxonomy of Viruses (ICTV); members of the same species share >70% amino acid identity in the polyprotein, >60% amino acid identity in P1 and >70% amino acid identity in the 2C and 3CD region, share a natural host and share a common genome organization (Knowles et al., 2012). Our analysis based on the 11 SAFV deduced amino acid sequences for the P1 region showed >60% identity to TMEV-DA strain (62–65%) and slightly higher identity to TRV (62–68%). Among SAFV genotypes, the identity range for the P1 region was 77–86%. The 2C and 3CD region alignments for the distance measurements showed all 11 SAFVs to be >70% identical to both TMEV-DA strain (74–75%) and TRV (83–84%), while within SAFV, the identity range in the 2C and 3CD region was 97–98%. Between SAFV and EMCV, the identity range in the P1 region and the 2C and 3CD regions was <70% (45–50 and 46–48%, respectively).

Mapping recombination events among cardioviruses

The phylogenetic relationship among SAFV, and between SAFV and TRV/TMEV, varied depending upon the genome region. In the capsid or P1 region, TRV and TMEV clustered separately from SAFV, while in the non-capsid region, they clustered more closely to SAFV. Therefore, we conducted a recombination analysis among SAFV and TMEV/TRV using several approaches. TRV was the most closely related sequence to the SAFV cluster in the phylogenetic analysis; therefore, we used it as a query sequence against all the representative sequences of TRV, TMEV and 11 SAFV genotypes. Among the SAFV genotypes, the similarity plot clearly showed the highest degree of similarity in the 5’-UTR (90%) and 3’-UTR (80%) (Fig. 4). In contrast, minimum similarity was observed in the structural region or P1 (10–65%). The most divergent region in the capsid-coding region was the VP1 gene (Fig. 4).

For a more specific picture of recombination breakpoints across the genome, bootscanning was performed to map the recombination events among TRV, TMEV and SAFV complete genome sequences. SAFV-11 was used as a query sequence because it was recently found and the percentage similarity was high only among SAFV. Bootscanning showed a close relationship in the P1, P2 and P3 regions of SAFV only, indicating multiple recombination events only among SAFV (Fig. 5). It further showed that SAFV-11 had recombination breakpoints at the junctions of the capsid region (P1) in VP2 and 2A corresponding to SAFV-8, while in the non-capsid region, there were multiple breakpoints located in the 2A, 2B, 2C, 3A, 3B and 3D regions corresponding to SAFV-6, SAFV-4, SAFV-7 and SAFV-5 (Fig. 5). No evidence of recombination between SAFV and other cardioviruses was found by the bootscanning method.

We then applied GARD, a genetic algorithm for recombination detection, on the dataset containing complete genome sequences for SAFV, TMEV and TRV. We found five significant breakpoints with topological incongruence using the Kishino–Hasegawa (KH) test (P<0.01), shown as impulses (Fig. 6). The first potential breakpoint was observed at the junction of VP2 and VP4, the second was at the junction of VP1 and 2A and the remaining points were in the 2C, 3B and 3D regions of the genome (Fig. 6). The resulting six fragments of alignment were used for phylogeny reconstruction by the neighbour-joining method, using the K2P model implemented in MEGA5 software (Tamura et al., 2011).

The phylogenies of the significant (P<0.01) breakpoints showed separation of SAFV, TRV and TMEV into different clusters (Fig. 6). A similar pattern of separation was observed across the genome, regardless of the location of the breakpoint. Phylogenetic trees resulting from the P1 region showed a genotype specific relationship among SAFV only. In the non-structural regions (2A–3D), SAFV recombined independently of genotype, instead recombining based on geographical location. In contrast, TRV and TMEV clustered separately from SAFV throughout the genome (Fig. 6).

DISCUSSION

In this study, we investigated the role of genetic diversity and recombination for reclassification of recently described SAFV. Screening of stool specimens from 943 AFP cases led to the identification of 88 SAFVs belonging to 11 different genotypes. We also report the discovery of three new genotypes of SAFV, with the full genome sequences of 11 genotypes from Pakistan and Afghanistan. The salient features were the high genetic diversity and extensive recombination among different SAFV genotypes.

Diversity of the VP1 gene among SAFV was found to be far greater than in animal cardioviruses. Three more genotypes with high divergence in VP1 were characterized as SAFV-9,
-10 and -11, bringing the number of genotypes from eight to 11 (Fig. 1). All the genotypes of SAFV showed greater diversity than previously thought (Blinkova et al., 2009; Drexler et al., 2010) but similar grouping into a number of well-defined clusters (bootstrap supported), separate from the species Theilovirus.

Full genome or near full genome sequences for only five genotypes of SAFV (SAFV-1 to -3, and SAFV-5 to -6) were available before this study (Drexler et al., 2010). We determined the complete genomes for six new genotypes (Figs 3 and S2). SAFV in Pakistan and Afghanistan defined clusters with their respective types in the P1 region, while exclusive clustering of SAFV observed in the non-capsid part of the genome indicated a geographical relationship between these circulating viruses. Clusters also contained sequences from viruses detected from different regions and years, showing active circulation of different SAFV genotypes. Phylogenetic tree incongruence, particularly in non-structural protein coding regions and UTRs, has also been described before, with the available genotypes ranging from two to five (Blinkova et al., 2009; Drexler et al., 2010; Liang et al., 2008). In this study, the same phenomenon was clearly observed among all 11 genotypes of SAFV.

Multiple recombination events have been reported in the junction region between the structural and non-structural genome regions of SAFV (Blinkova et al., 2009; Drexler et al., 2010). As opposed to previous studies, which had the constraint of limited data, the availability of extensive and diverse sequence data in the present study improved the reliability of recombination analysis. The difference in
terms of phylogenetic relationships is widely used for the
detection of recombination events over the genome, which
sometimes can overlook true breakpoints, particularly with
picornaviruses (Lukashev, 2010). Use of different methods
for analysis of recombination here suggested the presence of
possible breakpoints along with phylogenetic incongruence.

Preliminary evidence of recombination between theiloviruses
and SAFV in the region separating structural and non-
structural genes has been described previously (Blinkova
et al., 2009). However, we were unable to find any evidence
of recombination between theilovirus and SAFV (Figs 5 and
6). Moreover, our results of complete genomes from 11
different SAFV genotypes were consistent with the analysis of
two SAFV genotypes, as described previously (Drexler
et al., 2010). In the model of enterovirus evolution, recombination
takes place strictly between members of the same species
(Lukashev, 2005, 2010; Lukashev et al., 2005).

All three well-separated distributions of pairwise distances
in the VP1 region suggested that SAFV genotypes are easily
defined and segregated for classification purposes (Fig. 2).
The existence of a small number of pairs of variants with
genetic distances at an intermediate position in the
distribution corresponded directly to different geographical
lineages. In view of the extensive VP1 sequence data for
SAFV in our study, for genotype assignment, we refined an
intra-type threshold of 23 and 10% divergence for VP1
nucleotide and amino acid sequences from the previously
proposed VP1 divergence threshold of 25 and 12% for
nucleotide and amino acid sequences, respectively (Blinkova
et al., 2009).

For human enteroviruses, a nucleotide sequence divergence
value of >25% in VP1 (>12% amino acid sequence
difference) is used as an alternative means to classify more
recently discovered types without recourse to extensive

---

Fig. 5. Mapping of recombination breakpoints by bootscanning. For a more specific picture of recombination breakpoints
across the genome, bootscanning was performed to map the recombination events among TRV, TMEV and SAFV complete
genome sequences. SAFV-11 showed close relationships in the L-protein, P1, P2 and P3 regions of SAFV indicating multiple
breakpoints or recombination events at the junctions of structural and non-structural regions of the genome and throughout the
non-capsid. The genome organization of the query sequence (TRV EU542581) is shown above the figure. The type and name of
the strain are indicated.
serological characterization (Oberste et al., 1999). Similarly, thresholds of 12% differentiate inter- from intra-serotype divergences in the VP1 gene of species A (Savolainen et al., 2002) and species B rhinoviruses (McIntyre et al., 2010), respectively, providing the means in principle to detect novel thresholds of 12% differentiate inter- from intra-serotype differentiation between animal and human cardioviruses (Drexler et al., 2010). High genetic diversity, phylogenetic incongruence leading to frequent recombination and distribution patterns of pairwise distances based on the complete VP1 region were the main criteria for discrimination of SAFV genotypes in this study. At the same time, SAFV genotypes shared >70% sequence identity to almost all viral proteins except VP1 from other members of the species Theilovirus. It is rather difficult to treat members of the species Theilovirus (TRV and TMEV) as two genotypes clustering within 11 SAFV genotypes as the same species because they are more heterogeneous, with scores ranging from 41–56% amino acid pairwise distances (Fig. 2b). SAFV and TMEV share a common genome organization, but the L protein of SAFV may have hybrid features of both TMEV and EMCV species (Himeda & Ohara, 2012). Thus, SAFV could be distinguished from the other members of Theilovirus by sequence divergence and intra-typic recombination. Although our results do not fulfill the aforementioned criteria of ICTV, the high degree of diversity in the VP1 region among different members of the species Theilovirus, the frequent intra-typic recombination and the possible differences in the host range for SAFV may keep open questions related to SAFV classification status.

The SAFV genotypes have been difficult to propagate in cell culture, except for SAFV-1, -2 and -3 (Jones et al., 2007; Blomqvist et al., 2012; Zoll et al., 2009). We recently established an infectious clone of SAFV-3, paving the way for future studies to understand the pathogenesis of the virus (Himeda et al., 2011). Full genome sequences determined in this study will help establish the infectious clones of new SAFV genotypes and improve our understanding of the genetic and phenotypic diversity, and classification of SAFV.

METHODS

Biological specimens. All specimens were collected through the AFP surveillance network of Pakistan and Afghanistan from March 2009 to August 2009. The AFP surveillance network is used to investigate any AFP cases in children aged <15 years. The ultimate objective is to detect the circulation of wild polioviruses and/or vaccine-derived polioviruses (VDPV) for global polio eradication. Stool specimens of 943 AFP cases were investigated with 831 samples collected from Pakistan and 112 samples from Afghanistan. In this study, we only used poliovirus- and non-polio enterovirus-negative stool samples, which did not induce apparent cytopathic effects on L20B or RD cell lines.

Detection of SAFV by PCR from AFP cases. RNA extraction from the stool supernatants was performed using a viral RNA extraction kit (Roche) according to the manufacturer’s instructions. All RNA samples were subjected to nested RT-PCR screening based on 5'-UTR as described previously (Drexler et al., 2008) with few modifications. Briefly, 50 μl reactions used a Qiagen OneStep RT-PCR kit, with 600 nmol 1⁻¹ each of first-round primers CF188 and CR990 and 5 μl RNA extract. Amplification involved 30 min at 50 °C; 15 min at 95 °C, and 40 cycles of 30 s at 94 °C, 30 s at 60 °C, and 60 s at 72 °C, with a final elongation step of 10 min at 72 °C. For nested PCR, 50 μl reactions contained 1 μl of first-round PCR product, with 0.25 μl Ex Taq polymerase (TaKara) and 600 nmol 1⁻¹ second-round primers CF204 and CR718. Amplification involved 3 min at 95 °C and 35 cycles of 30 s at 94 °C, 30 s at 60 °C and 60 s at 72 °C, with a final elongation step of 10 min at 72 °C. The same method was used for the VP1 gene-based characterization of positive virus specimens with specific primers (Table S1).

Complete genome sequencing. For each of the genotypes, representative strains with a high viral RNA copy number were tentatively selected for full genome sequencing. RT-PCR was performed using the Titan one tube RT-PCR system (Roche) with specific primers according to the manufacturer’s instructions (Table S1). PCR fragments were purified using the Wizard SV gel and a PCR clean-up system (Promega), which enabled direct sequencing of the PCR products. Based on a sequence comparison (local database search), viruses were confirmed as SAFV. A full genomic sequence for each of the 11 genotypes was obtained by a genome walking method on amplified, partially overlapping fragments of 0.5–0.8 kb in both directions, whereas 5’-UTR was acquired using 5’-UTR primer and 3’-UTR was acquired using KOD-Plus polymerase (Toyobo). These sequences were assembled in Sequencher version 4.2 (Gene Codes) and sequence similarity searches were performed initially using BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and later using an in-house database.

Geographical distribution. In order to demonstrate the geographical distribution of all genotypes across Pakistan and Afghanistan, maps were constructed using HealthMapper 4.3.2, a basic mapping and surveillance program developed by the Communicable Disease Surveillance and Response Department, World Health Organization.

Genetic analysis. The complete nucleotide and protein sequences of cardiovirus genomes were obtained from GenBank (Table S2).
Sequence alignment was performed using CLUSTALW (Higgins et al., 1996). Distance matrices were produced using the Kimura two-parameter method (Kimura, 1980) for nucleotide sequences and the Poisson correction method (Nei & Kumar, 2000) for amino acid sequences, as implemented in MEGA5 software (Tamura et al., 2011). When the rate of nucleotide substitution is nearly the same for all evolutionary lineages and there is no strong transition/transversion (Ts/Tv) bias, the pairwise distance seems to give correct trees more often than other distances, even if sequence divergence is high (Takahashi & Nei, 2000). Matrices were then used for the construction of phylogenetic trees with the neighbour-joining method in MEGA5 software. Bootstrapping was performed using 1000 pseudo-replicates for the calculation of branching confidence. Similarity plots and bootscanning analyses were performed using SimPlot version 3.5.1 (Lole et al., 1999). A sliding window of 600 nt instead of the default value of 200 nt was used in order to improve the signal to noise ratio and the Kimura two-parameter method was used for estimation of distances with respect to nucleotide substitutions. The overall Ts/Tv ratio for the dataset was calculated using MEGA5 software (Tamura et al., 2011). Recombination was further analysed by GARD, a genetic algorithm for recombination detection (Kosakovsky Pond et al., 2006a, b).

ACKNOWLEDGEMENTS

The authors are particularly grateful to the technical staff at NIH, Pakistan, for providing and testing the samples from the AFP case archive for the presence of polio and enteroviruses, as well as Junko Wada for technical assistance. The authors would also like to thank Dr Olen Kew (Centers for Disease Control and Prevention) for valuable discussion and review of the manuscript. Asif Naeem’s PhD studentship was funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan. This study was supported in part by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science, and for research on emerging and re-emerging infectious diseases from the Ministry of Health, Labour and Welfare, Japan.

REFERENCES


Takahashi, K. & Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol Biol Evol 17, 1251–1258.


