

**Short  
Communication****Phylogenetic structure of serotype A  
foot-and-mouth disease virus: global diversity  
and the Indian perspective**

Jajati K. Mohapatra, Saravanan Subramaniam, Laxmi K. Pandey,  
Sachin S. Pawar, Ankan De, Biswajit Das, Aniket Sanyal  
and Bramhadev Pattnaik

Correspondence  
Bramhadev Pattnaik  
pattnaikb@gmail.com

Project Directorate on Foot and Mouth Disease, Indian Veterinary Research Institute Campus,  
Mukteswar, Nainital 263 138, Uttarakhand, India

Received 2 November 2010  
Accepted 7 January 2011

Global epidemiological analysis is vital for implementing progressive regional foot-and-mouth disease control programmes. Here, we have generated VP1 region sequences for 55 Indian type A outbreak strains and have included complete VP1 sequences from 46 other countries to obtain a comprehensive global phylogeographical impression. A total of 26 regional genotypes within three continental topotypes, based on a 15% nucleotide divergence cut-off criterion, could be identified. These genotypes correlated with distinct evolutionary lineages in the maximum-likelihood phylogeny. During the last decade, ten genotypes have been in circulation the world over and it was evident that no type A strain has transgressed the continental barriers during this period. A single genotype (genotype 18) within the Asia topotype has been circulating in India with neither any incursion nor any long distance movement of virus out of the country during the last ten years, although close genetic and epidemiological links between viruses from Bhutan and India were revealed.

Conventionally, the VP1 region (1D) sequence has been used for genetic characterization of foot-and-mouth disease virus (FMDV) strains because of its significance in virus attachment and entry, protective immunity and serotype specificity. VP1-based phylogenetic analyses have been used widely to deduce evolutionary dynamics and the epidemiological relationship among the genetic lineages, and in tracing the authentic origin and movement of the outbreak strains (Samuel & Knowles, 2001). Though there has been an exponential growth in the number of FMDV genomic sequences in the public domain in recent years, published epidemiological findings are mostly restricted geographically. Inadequate real-time epidemiological information and nonavailability of sequence data from most of the countries with endemic foot-and-mouth disease (FMD) have stood in the path of understanding the global character of FMDV.

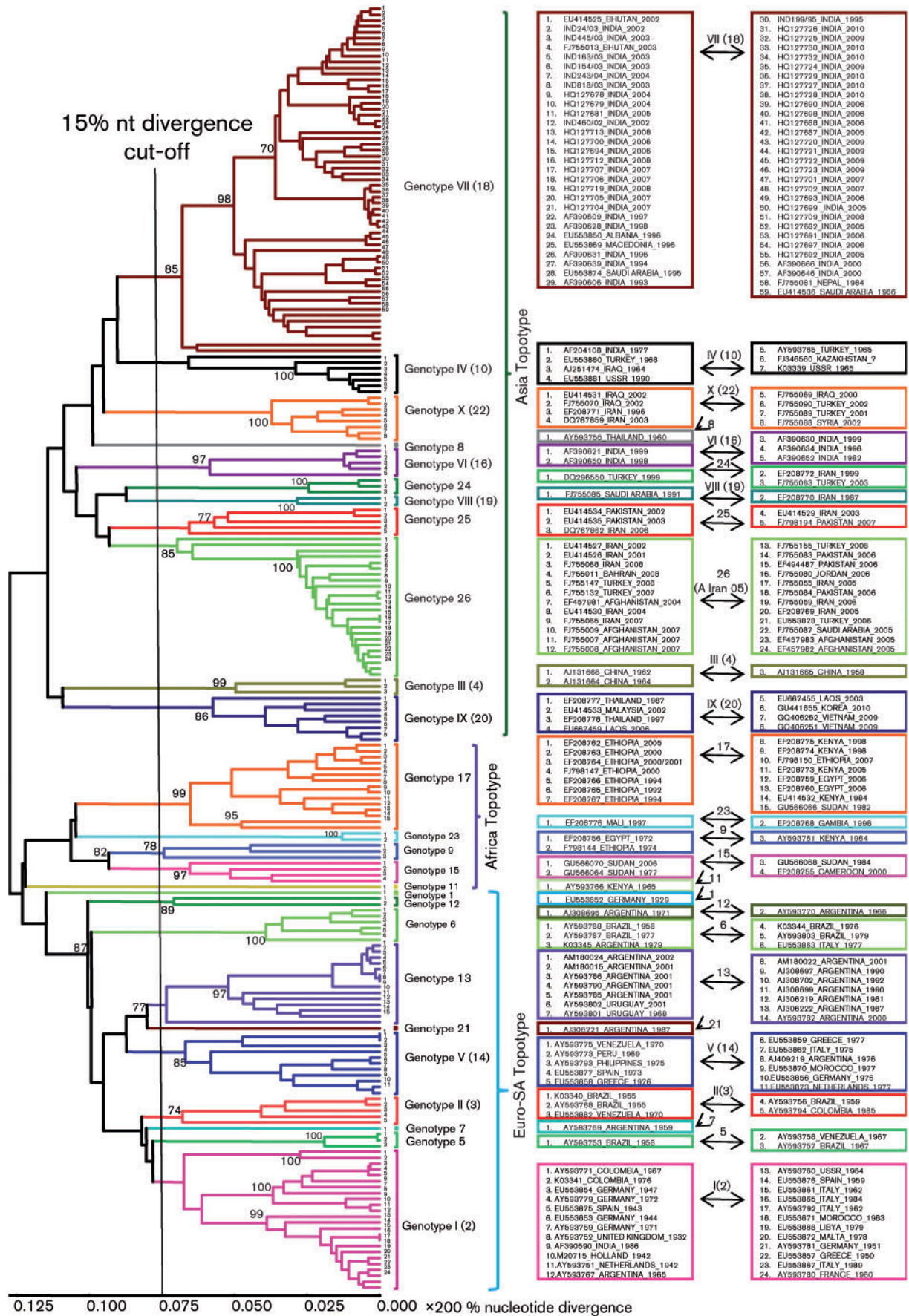
Serotype A is considered to be one of the most diverse serotypes both antigenically and genetically, making control by vaccination very difficult (Kitching, 2005). It has been felt that, despite a significant capacity for molecular characterization being used to generate information on many hundreds of viruses per year in India, there appears to have been no published information on

systematic analysis of the spatial and temporal distribution of topotypes (Rweyemamu *et al.*, 2008). So far only one complete VP1-sequence-based global genotyping study including Indian type A viruses, collected between 1977 and 2001, has been published (Tosh *et al.*, 2002). The type A FMDV population was classified into ten major genotypes in that study, with greater than 15% nucleotide divergence among the genotypes, but that analysis included neither any sequences from Africa nor those of any recent, unique genetic lineages such as 'A Iran 05' from the Middle East (Knowles *et al.*, 2009) or the 'VP3<sup>59</sup>-deletion group' from India (Jangra *et al.*, 2005). To overcome this gap in our knowledge, we attempted an updated and more comprehensive global genotyping to determine the extent of genetic diversity of serotype A and to assess the relationships among the geographically segregated genetic lineages worldwide. Such epidemiological analysis holds the key to implementing sustainable progressive regional FMD control programmes.

For this purpose, complete 1D sequences of 55 Indian type A-outbreak strains collected between 2004 and 2010 as a part of national FMD-surveillance activities and isolated in BHK-21 cells (passage level 3–5) were resolved in this study. The inclusion of 156 GenBank-derived sequences from 46 other countries spread across four continents (21 countries from Asia, nine from Europe, ten from Africa and six from South America) in the phylogenetic reconstruction helped in producing an integrated spatio-temporal

The GenBank accession numbers for the VP1 coding region sequence data reported in this paper are HQ127678–HQ127732.

A supplementary figure is available with the online version of this paper.



**Fig. 1.** UPGMA tree showing a complete VP1-sequence-based global phylogeny and toptype/genotype distribution. Bootstrap support values are indicated only for the major nodes. Inside the boxes, accession numbers or isolate designations followed by the country of origin and year of collection are depicted serially as per their position on the branches in a top-to-bottom direction.

global impression of type A FMDV. Besides this, a detailed study of the molecular epidemiology of type A FMD in India over a period of three decades was performed by including 20 other Indian virus sequences collected between 1977 and 2003 as representatives of different genotypes. This study provides valuable insights into the global distribution of type A FMDV genetic clusters, which can serve as a phylogeographical reference map for keeping track of the evolution and spread of lineages in future.

Genomic RNA extraction from the infected cell-culture supernatant and RT-PCR to amplify the VP1 region were carried out as described previously (Tosh *et al.*, 2002). Nucleotide sequences were generated by using an ABI 3130 Genetic Analyzer (Applied Biosystems) using NK61 and 1C562 primers (for details of these, see Tosh *et al.*, 2002).

The nucleotide sequence alignment was performed using CLUSTAL X version 1.83 (Thompson *et al.*, 1997) and the maximum-likelihood (ML) phylogeny was inferred using PhyML version 3.0 (Guindon & Gascuel, 2003). Selection of the best-fit nucleotide substitution model of evolution was performed using jModelTest version 0.1.1 under the framework of a Bayesian information criterion (BIC) model-selection strategy (Posada, 2008). For this analysis, the HKY85 nucleotide-substitution model, the discrete gamma model with four categories, the nearest-neighbour interchanges algorithm and the approximate likelihood-ratio test for branches (aLRT) (Anisimova & Gascuel, 2006) were selected. The gamma shape parameter was estimated to be 0.423 for the global dataset. Phylogenetic comparisons were also performed using MEGA4 (Tamura *et al.*, 2007) after performing an independent alignment with the CLUSTAL W algorithm (Thompson *et al.*, 1994). The evolutionary history was inferred by using both neighbour-joining (NJ) (Saitou & Nei, 1987) and unweighted pair group mean average (UPGMA) (Sneath & Sokal, 1973) methods. The bootstrap consensus tree inferred from 10 000 replicates (Felsenstein, 1985) was taken to represent the evolutionary history of the taxa analysed. The evolutionary distances used to infer the phylogenetic trees were computed using the Kimura two-parameter method (Kimura, 1980).

It has been suggested by earlier workers, who were engaged in elaborate studies of the epidemiology of picornavirus infections, that approximately 85% identity at the level of VP1 is a realistic cut-off for differentiating between major genotypes (Rico-Hesse *et al.*, 1987; Samuel & Knowles, 2001; Tosh *et al.*, 2002; Vosloo *et al.*, 1992). Such genotype classifications correlated with geographically distinct evolutionary lineages as well. Hence, by using a similar 15% nucleotide-divergence cut-off criterion, we could identify a

total of 26 genotypes as being apparent in the VP1 UPGMA tree (Fig. 1). However, it should be considered that the extent of diversity detected here could be far greater than we currently realize, as surveillance and sampling might not have been foolproof in many parts of the world. We have designated the 26 genotypes by using Arabic numerals 1–26 in the order of their appearance and have also kept the Roman numeric designations (I–X) given in an earlier study for the 10 genotypes (Tosh *et al.*, 2002), to avoid any confusion. The resulting unrooted NJ tree (data not shown) and the rooted ML tree (Fig. 2a and Supplementary Fig. S1, available in JGV Online) clearly show 26 genetically distinct evolutionary lineages/clusters as well, with high bootstrap confidence limits (>70%) and aLRT values (>0.8). Except for genotypes 2 and 14, all other genotypes formed monophyletic lineages in the ML and NJ trees. Though genotypes 2 and 14 formed single clusters in the UPGMA tree, strains were found to be interspersed with genotype 5 and 21, respectively, in the ML and NJ trees. This could be caused by the fact that the ‘15% cut-off’ to delineate genotypes, though logical, is a heuristic choice. Moreover, these genotypes revealed that genotypes with just more than a 15% nucleotide difference (~16–18%) between them are distributed in the same geographical regions. Hence, it is most likely that such clustering is the result of intermediate sequences evolving from the older genotypes, which in turn provide ancestry to the newer genotypes in a stepwise manner (Fig. 2a).

Most of these genotypes (23 of 26) showed a regionally restricted geographical distribution pattern, a few even being confined to a particular country (Fig. 3). All these genotypes could be accommodated within the three broad continental toptypes [Asia, Europe–South America (Euro-SA) and Africa (Knowles & Samuel, 2003)] except for genotypes 2, 14 and 18, which were found to have transgressed their normal continental niches. Genotypes 2 and 14 could be traced to all four continents with endemic FMD, whereas genotype 18 could be found in Asia and Europe. More importantly, all such transcontinental movements of virus occurred before the 21st century and have been attributed to either immigration of people with their livestock to establish colonies, to the importation of livestock and livestock products or to the inadvertent release of old European strains that were extensively used in vaccines in South America during that time (Leforban & Gerbier, 2002; Rweyemamu *et al.*, 2008; Samuel & Knowles, 2001; Valarcher *et al.*, 2008). Overall, the Asia, Euro-SA and Africa continental toptypes comprised 11, 10 and 5 regional genotypes, respectively (Fig. 1). From the UPGMA tree, it is evident that a minimum of a ‘24% nucleotide difference cut-off’ could be a rational criterion to distinguish



**Fig. 2.** ML trees reflecting toptotype/genotype/clade distribution. An Indian type O isolate sequence was used to root the trees and the aLRT values for only the major nodes are shown. (a) ML phylogeny calculated from the global dataset. The numbers against the branches indicate isolate numbers as in Fig. 1. (b) ML phylogeny calculated from the Indian dataset, where isolate designations followed by the state of origin are shown. Accession numbers or references are shown within parentheses. Bars, 0.1 nucleotide substitutions per site.

between the continental toptypes. Genotype 1, comprising an isolate from Germany, happens to be the oldest genotype in this study. In the ML tree, genotype 1 was placed close to the root in accordance with its genealogy. In the Asia toptotype, genotype 8, recorded in Thailand during 1960, was placed close to the root and likewise, for the Africa toptotype, genotype 11 from Kenya appeared to be the ancestral genotype (Fig. 2a). During the last decade, ten genotypes have been in circulation worldwide and it is apparent from the phylogram that no type A strain has jumped the continental barriers during this period.

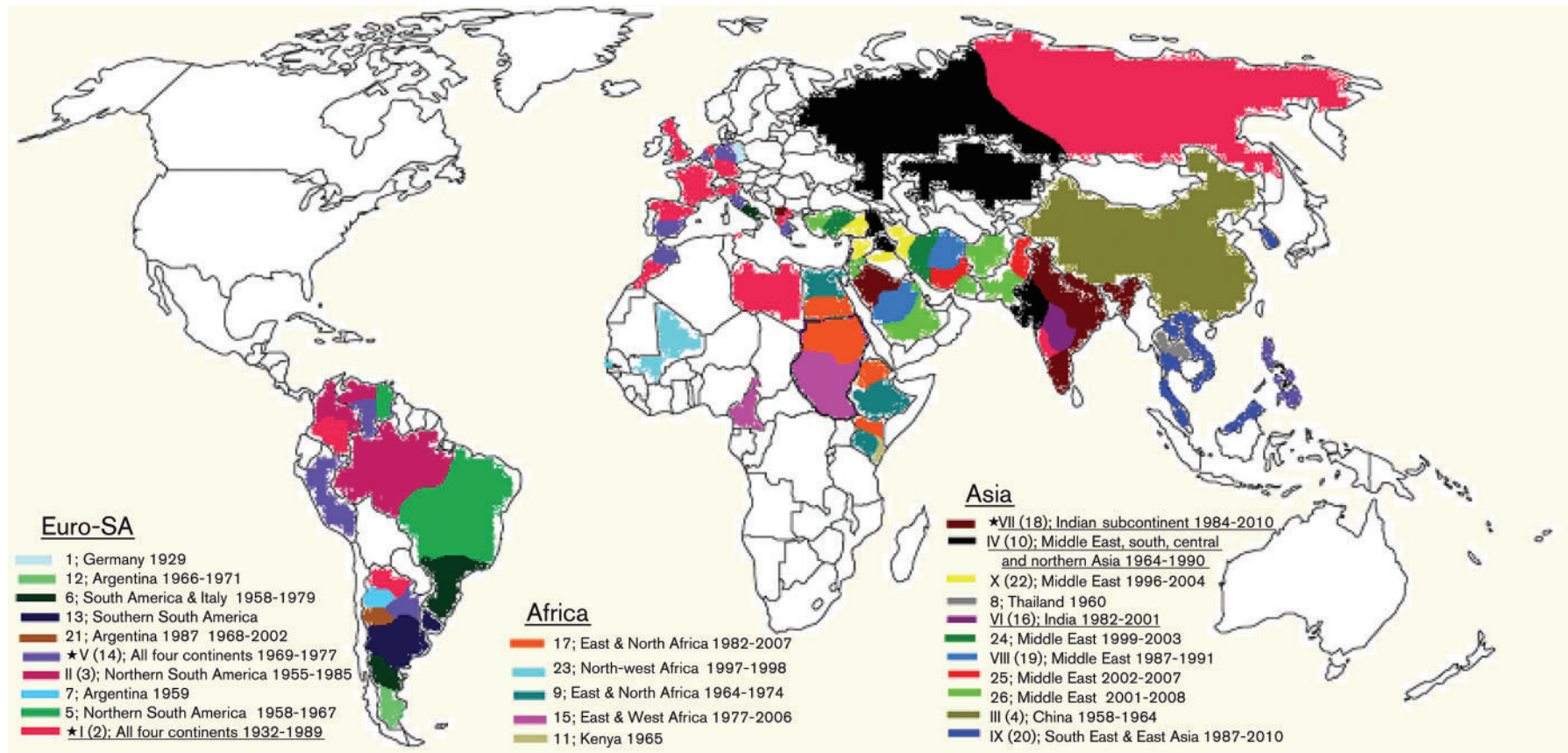
Of the 11 genotypes within the Asia toptotype, six genotypes could be identified only in the Middle East region. Likewise, seven of the ten genotypes within the Euro-SA toptotype could be detected in Argentina only. These two regions might be considered to be hot-spots as far as genotype diversity is concerned. Strict compartmentalization within a country's boundary was less apparent for any of the genotypes indigenous to the Middle East. Hence, this whole region may be considered as an epidemio-geographical unit with respect to the spread of virus strains. Genotype 25, recorded during 2002–2007, from Iran and Pakistan appeared to be the closest neighbour of genotype 26 (A Iran 05 lineage), and two Iranian strains, collected during 2001–2002, clustered as intermediates between genotype 25 and the A Iran 05 lineage. Hence, with the available sequence data, it is tempting to hypothesize that these indigenous historic sequences might have provided the most recent ancestor for the A Iran 05 lineage. A stepwise evolution based on their order of appearance was observed for genotype 20 from South East Asia between 1987 and 2010, indicating rapid strain turnover, probably due to continuing immune selection.

In India, four genotypes [genotype I (2), IV (10), VI (16) and VII (18)] have been documented. Genotypes 2 (Euro-SA toptotype) and 10 (Asia toptotype) were recorded before 1990 and no longer seem to exist in India (Tosh *et al.*, 2002). The epidemiological trend shows an epochal evolution of type A genotypes characterized by a continuous replacement of old genotypes with newer ones, as observed for human enteroviruses (van der Sanden *et al.*, 2010). Population dynamics studies indicate a recent genotype demographic transition from genotype 16 to genotype 18 in 2001. Apparently genotypes 16 and 18, both within the Asia toptotype, evolved independently but have shared the same geo-ecology in the country, as they are placed quite distantly in the ML tree, emerged from two distinct ancestral nodes in the Asia toptotype and have followed different evolutionary trajectories. Each of these

two genotypes appears to share its most recent common ancestor with viruses from two geographically separate regions. Genotypes 18 and 22 (from the Middle East) descended from a common intermediate node while genotypes 16 and 20 (from South-east Asia) showed common ancestral linkage. In the ML tree comprising only Indian isolates, genotype 2 was placed close to the root and followed an independent path of evolution. The other three genotypes (genotypes 10, 16 and 18) have descended and diversified from a common ancestral node (Fig. 2b). Strains from Nepal and Saudi Arabia collected during 1984 and 1986, respectively, clustered in genotype 18 and they appear to be intermediates between genotypes 22 and 18. Based on phylogeographical configuration, it might be suggested that the Indian viruses within genotype 18 are descended from a virus related to that from Nepal and that similar ancestral sequences have also circulated in countries of the Arabian Peninsula.

The 1996 type A outbreak in Albania and Macedonia has been ascribed to the importation of on-the-bone buffalo meat from South Asia. Also, the virus strains revealed close genetic relationships with the then-circulating strains from India and Saudi Arabia within genotype 18 (Leforban & Gerbier, 2002; Tosh *et al.*, 2002). However, in the last decade, it has become evident that a single genotype is circulating in India with neither any incursion of lineages from other countries nor any movement of type A virus out of India, except that some movement has occurred between neighbouring countries of the Indian subcontinent. The A Iran 05 lineage expanded its territory into Pakistan; however, further eastward dissemination into adjoining parts of India has not occurred. Intense surveillance combined with modern molecular techniques should have detected any such incursion into India without fail. Type A viruses from India revealed more than 20% nucleotide divergence from those from South-east Asia and also clustered in separate genotypes. Hence, no epidemiological linkage could be established between contemporaneous Indian and South-east Asian viruses. As far as movement of viruses between India, Bangladesh, Nepal and Sri Lanka is concerned, the molecular phylogenetic analysis is currently handicapped because of lack of sequence data from these countries.

For the VP3<sup>59</sup>-deletion group within genotype 18, the dominant group of virus in recent years in India, the evolutionary sublineage/clade clustering was found to be dictated by geographical isolation. Three different clades could be identified as proof of the fact that though they have descended from a common immediate ancestor they are in the phase of active evolution and diversification.



**Fig. 3.** Global footprints of serotype A FMDV genotypes. Underlined genotypes are those that suggest the grouping of Indian viruses. Genotypes marked with an asterisk denote genotypes which have transgressed their normal continental niches.

Clade 18a, being the oldest clade in this deletion group, was first detected as early as 2002 and circulated up until 2005, being restricted to northern and north-eastern parts of India. Clade 18b has circulated only in north India, while clade 18c was found to be totally restricted to South India. Though the country of origin remains uncertain, phylogenetic relationships suggest that genetically similar viruses (with less than 2% nucleotide difference) belonging to clade 18a of the VP3<sup>59</sup>-deletion group have circulated in both Bhutan (GenBank accession no. EU414525) and the neighbouring state Assam of India (IND 24/2003) during the same time period (Figs 1 and 2). It has been suggested that a nucleotide difference of less than 5% indicates an epidemiological link, and the isolates could be either from the same outbreak or are closely related temporally (Samuel *et al.*, 1999; Vosloo *et al.*, 1992). Although the possibility of airborne spread exists, it is difficult to exclude the possibility of there having been either some trade in live animals or the intermingling of animals from both sides of the border. In any case, such a genetic link underscores the need for rigid border surveillance.

When considering intervention strategies for the control of FMD, it is important to take account of the characteristics of different genetic clusters circulating in various ecological systems along with their routes of movement. The global genotyping and phylogeographical design presented here may serve as a platform in this regard.

## Acknowledgements

This work was supported by the Indian Council of Agricultural Research. The efforts of Dr S. Mahajan, M. V. Sc. scholar in formatting the figures is appreciated.

## References

- Anisimova, M. & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55, 539–552.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696–704.
- Jangra, R. K., Tosh, C., Sanyal, A., Hemadri, D. & Bandyopadhyay, S. K. (2005). Antigenic and genetic analyses of foot-and-mouth disease virus type A isolates for selection of candidate vaccine strain reveals emergence of a variant virus that is responsible for most recent outbreaks in India. *Virus Res* 112, 52–59.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16, 111–120.
- Kitching, R. P. (2005). Global epidemiology and prospects for control of foot-and-mouth disease. *Curr Top Microbiol Immunol* 288, 133–148.
- Knowles, N. J. & Samuel, A. R. (2003). Molecular epidemiology of foot-and-mouth disease virus. *Virus Res* 91, 65–80.
- Knowles, N. J., Nazem Shirazi, M. H., Wadsworth, J., Swabey, K. G., Stirling, J. M., Statham, R. J., Li, Y., Hutchings, G. H., Ferris, N. P. & other authors (2009). Recent spread of a new strain (A-Iran-05) of foot-and-mouth disease virus type A in the Middle East. *Transbound Emerg Dis* 56, 157–169.
- Leforban, Y. & Gerbier, G. (2002). Review of the status of foot and mouth disease and approach to control/eradication in Europe and Central Asia. *Rev Sci Tech* 21, 477–492.
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25, 1253–1256.
- Rico-Hesse, R., Pallansch, M. A., Nottay, B. K. & Kew, O. M. (1987). Geographic distribution of wild poliovirus type 1 genotypes. *Virology* 160, 311–322.
- Rweyemamu, M., Roeder, P., Mackay, D., Sumption, K., Brownlie, J., Leforban, Y., Valarcher, J. F., Knowles, N. J. & Saraiva, V. (2008). Epidemiological patterns of foot-and-mouth disease worldwide. *Transbound Emerg Dis* 55, 57–72.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425.
- Samuel, A. R. & Knowles, N. J. (2001). Foot-and-mouth disease type O viruses exhibit genetically and geographically distinct evolutionary lineages (topotypes). *J Gen Virol* 82, 609–621.
- Samuel, A. R., Knowles, N. J. & Mackay, D. K. J. (1999). Genetic analysis of type O viruses responsible for epidemics of foot-and-mouth disease in North Africa. *Epidemiol Infect* 122, 529–538.
- Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco: Freeman.
- 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.
- 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.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876–4882.
- Tosh, C., Sanyal, A., Hemadri, D. & Venkataramanan, R. (2002). Phylogenetic analysis of serotype A foot-and-mouth disease virus isolated in India between 1977 and 2000. *Arch Virol* 147, 493–513.
- Valarcher, J. F., Leforban, Y., Rweyemamu, M., Roeder, P. L., Gerbier, G., Mackay, D. K. J., Sumption, K. J., Paton, D. J. & Knowles, N. J. (2008). Incursions of foot-and-mouth disease virus into Europe between 1985 and 2006. *Transbound Emerg Dis* 55, 14–34.
- van der Sanden, S., van der Avoort, H., Lemey, P., Uslu, G. & Koopmans, M. (2010). Evolutionary trajectory of the VP1 gene of human enterovirus 71 genogroup B and C viruses. *J Gen Virol* 91, 1949–1958.
- Vosloo, W., Knowles, N. J. & Thomson, G. R. (1992). Genetic relationships between southern African SAT-2 isolates of foot-and-mouth-disease virus. *Epidemiol Infect* 109, 547–558.