Infection, Genetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Genetic and antigenic analysis of foot-and-mouth disease virus serotype O responsible for outbreaks in India during 2013

Saravanan Subramaniam, Jajati K. Mohapatra, Biswajit Das, Aniket Sanyal, Bramhadev Pattnaik *

Project Directorate on Foot-and-mouth Disease, Mukteswar-Kumaon, Nainital 263138, Uttarakhand, India

ARTICLE INFO

Article history:
Received 17 May 2014
Received in revised form 6 December 2014
Accepted 6 December 2014
Available online xxxx

Keywords:
FMD virus serotype O
Complete genome
Ind2001
Emergence

ABSTRACT

In recent times, majority of the foot-and-mouth disease (FMD) outbreaks in India are caused by serotype O Ind2001 lineage. The lineage has diverged into four sub-lineages (Ind2001a, b, c and d). We report here the genetic and antigenic analyses of nine Ind2001d isolates that caused outbreaks during April 2013–March 2014 in India. The length of the genomes of outbreak viruses varied between 8153 and 8181 nucleotides without any insertion or deletion in the coding region. Of the nine isolates analyzed antigenically against the currently used Indian vaccine strain INDR2/1975, eight showed good cross serological match (>0.3) indicating optimal antigenic coverage by the vaccine strain. An unprecedented deletion of 22 nucleotides between position 57 and 78 was observed in the 3′ untranslated region of one of the isolates without compromising the virus viability, which imply that partial distortion in SL2 of 3′UTR may not have influence on virus viability at least under *in-vitro* conditions. Recently the Ind2001 lineage has been reported from several countries including Libya and spread of this lineage across a wide geographical area needs to be monitored carefully to avoid any future pandemic.

Foot-and-mouth disease (FMD) virus is the causative agent of the economically most important livestock disease, FMD. The virus is classified within the genus *Aphthovirus* of the family *Picornaviridae*. FMD virus has seven distinct serotypes namely O, A, C, Asia 1 and SAT1-3. The genome of the FMD virus is a positive sense single stranded RNA molecule of ~8.4 kb length that is translated into a single polyprotein, which is proteolytically cleaved to yield mature structural and nonstructural proteins (Belsham, 1993). The coding region, which is flanked on either side by 5′ and 3′ untranslated regions (UTRs), can be divided into four regions viz; L, P1-2A, P2 and P3. The P1 region encodes structural proteins (VP4, VP2, VP3 and VP1), which form the virus capsid. The P2 and P3 regions encode nonstructural proteins (NSP) including viral protease and replicase. The 5′ UTR contains a short fragment called S-fragment (~370 residues), a poly(C) tract of variable length, followed by a large fragment (LF-UTR) of over 700 bases in length. The LF-UTR forms a number of highly conserved secondary structures that include 2–4 pseudoknots (PKs), a cis-acting replication element (cre) (Mason et al., 2002), and Internal Ribosome Entry Site (IRES). Of the five structural domains identified within the LF-UTR (Pilipenko et al., 1989), cre form domain 1 and IRES comprises of domains 2–5. The 3′ UTR, which is approximately 90 bases in length, is composed of two stem–loops followed by a poly (A) tract.

FMD is endemic in India with the prevalence of serotypes O, A and Asia 1 and of which serotype O accounts for about 80% of the total incidences (Subramaniam et al., 2013). The isolates of serotype O collected till date in the country grouped in various genetic lineages of the Middle East-South Asia topotype (ME-SA). The most prominent lineages in the recent years have been Ind2001 and PanAsia (Subramaniam et al., 2013). The Ind2001 lineage has been causing majority of the outbreaks in India since 2008. During April 2013–March 2014 (2013–14), a total of 454 outbreaks due to serotype O FMD virus were recorded in India (PD-FMD Annual Report, 2013-14). Almost 50% of the outbreaks were recorded in the southern peninsular India comprising the states of Karnataka, Tamilnadu, Kerala and Andhra Pradesh. The number of serotype O outbreaks recorded during 2013–14 in India was more than that of recorded in the previous three years [(2012–13 (265 outbreaks), 2011–12 (246 outbreaks) and 2010–11 (150 outbreaks)] (PD-FMD Annual Report, 2013-14). In this study, we are reporting the results of genetic and antigenic analysis of representative serotype O isolates collected from the outbreaks recorded during 2013 in India.

The complete genome sequences were determined from the BHK-21 cell culture adapted isolates ($n = 9$) at passage level 2–3. Total RNA was extracted using RNeasy Mini Kit (QIAGEN). Reverse

* Corresponding author. Tel.: +91 5942 286004; fax: +91 5942 286307.
E-mail addresses: saranvirol@gmail.com (S. Subramaniam), pattnaikb@gmail.com (B. Pattnaik).

http://dx.doi.org/10.1016/j.meegid.2014.12.009
1567-1348/

transcription was performed using oligo d(T)$_{15}$ primer and M-MLV Reverse transcriptase (Promega). The genome was amplified in seven overlapping fragments (SF1F-SF370R, LFIF-DHP2, L463R-NK61, DHP13-DH5, MG33-CTLV10, CTLV2-V4, 3D1081-anchored oligo dT) using Pfu DNA polymerase (Fermentas). The details of the primers used in this study are described earlier (Mathapati, 2012). After the gel purification of PCR products using QIA quick Gel Extraction Kit (QIAGEN), cycle sequencing reactions were performed on a 3130 genetic analyzer (Applied Biosystems). Multiple sequence reads were assembled using the Edit seq module of the Lasergene core suite 10 (DNASTAR, Inc., USA). Percent nucleotide and amino acid identity was calculated using the Megalign module. Bayesian inference method implemented in BEAST v 1.8.0 software package (Drummond and Rambaut, 2007) was used to infer the phylogenetic relationships employing the best fit nucleotide substitution model, TN93 + G as suggested by the MEGA 6.06 software (Tamura et al., 2013). The relaxed uncorrelated lognormal clock and exponential population size model was chosen among the different combinations of molecular clocks and demographic models for this analysis. Model selection was performed by comparing the model marginal log-likelihood through the Akaike's information criterion (Baele et al., 2012). The time to the most recent common ancestor (TMRCA) and evolutionary rate were co-estimated. The log file was analyzed using Tracer v 1.5 and the statistical uncertainties are reflected as 95% highest posterior density value (HPD). Maximum likelihood (ML) phylogenetic analysis was conducted using the MEGA 6.06 software (Tamura et al., 2013). The secondary structures of 3′ UTR were obtained by submitting the relevant sequences to *mfold* program (Zuker, 2003). Two dimensional micro-neutralization test (2D-MNT) was performed using bovine vaccinate serum against the currently used vaccine strain, INDR2/1975 (Subramaniam et al., 2013). The full length genome sequences of the nine FMD virus strains were deposited in GenBank under accession numbers KJ825801–KJ825809.

Excluding the poly (C) tract and poly (A) tail, the length of the genomes of nine FMD virus isolates varied between 8153 and 8181 nucleotides including 51 primer derived nucleotides. Variability in the genome length was due to insertion or deletion of nucleotides in the UTRs. Taking the 'T' succeeding the poly (C) tract as the first base, the consensus length of the LF-UTR was found to be 715 nucleotides. The length of S fragment and LF-UTR between the nine viruses differed by 3 and 2 bases, respectively. No deletions and insertions were observed in the coding region, which contained 6,996 nucleotides encoding a polyprotein of 2332 amino acids. The consensus length of the 3′UTR was 93 bases.

In the ML phylogenetic tree based on VP1 coding region, the isolates were clustered within the lineage Ind2001 of ME-SA topotype, precisely in the sub-lineage Ind2001d (Fig. 1). Complete genome based ML tree (not shown) was also built excluding the Ind2001a sub-lineage (sequence not available), which revealed a comparable clustering pattern to that of VP1 based tree. The Ind2001 lineage is prevalent in the countries of Indian subcontinent including Nepal, Bangladesh, Bhutan and India, but has never been reported from Pakistan (Jamal et al., 2011). Besides the Indian subcontinent, sporadic outbreaks attributable to this lineage have been reported in various countries in the Middle East in 1997, 2001, 2002, 2008, 2009, and more recently in Libya and Saudi Arabia in 2013 (Valdazo-González et al., 2014). The Ind2001 lineage, which re-emerged in the year 2008 in India, continued its supremacy in the field by out-competing the then prevalent O/ME-SA/PanAsia lineage. Since its initial identification in the year 1997, the lineage has diversified into at least four sub-lineages (Ind2001a, b, c and d) (Valdazo-González et al., 2014). The sub-lineage Ind2001b (emerged in the year 2001, circulated actively during 2002) and Ind2001d (emerged in the year 2008 and in circulation since then) were the major causes of serotype O outbreaks in India.

Nucleotide and amino acid divergence of Ind2001d (2013) isolates compared to the prototypic Ind2001 strains and PanAsia in the different genomic regions are given in Table 1. The nucleotide sequence identities among the nine Ind2001d (2013) isolates varied between 97.3% and 99.6%. The Ind2001d (2013) viruses shared nucleotide sequence identities of 93.8–94.3% with the prototypic Ind2001d (2008) isolates and 92.4–92.7% with the Ind2001b isolates. The Indian isolates had 97.0–98.9% nucleotide identities with the 3 isolates of Ind2001d lineage collected from Saudi Arabia, Libya and Bhutan during 2013. With the PanAsian strains, the identities varied between 90.2–90.7% (for PanAsia I) and 91.0–91.3% (for PanAsia II). Altogether, the Ind2001d (2013) isolates had high sequence identities with lineage Ind2001d (2008) followed by Ind2001b.

Phylogenetic relationship among Ind2001 isolates based on complete coding region of 21 isolates (2001–2013) was reconstructed by Bayesian analysis (Fig. 2). The mean nucleotide substitution rate of Ind2001 isolates at complete coding region was estimated to be 4.615×10^{-3} substitutions/site/year (95% HPD, $2.733–7.094 \times 10^{-3}$), which is similar to the rate reported for serotype O/South East Asia/Mya-98 lineage (Valdazo-González et al., 2013). The most recent common ancestor of Ind2001 lineage occurred around 18 years (95% HPD interval of 13–23 years) prior to 2013 (around 1995), 6 years before this lineage was first recognized. The nine Ind2001 isolates of 2013 from India were distributed in two genetic clusters. The isolate, IND142(311)/2013 from India grouped closely with Saudi Arabian and Libyan isolates, and together they shared a common ancestor dates back to 2010. The other eight Indian isolates grouped in one cluster and were derived from a common ancestor dates back to 2009 with an isolate from Bhutan. Interestingly the samples IND142 (311)/2013 and IND142(310)/2013, which were collected from the same place and time were placed in different genetic clusters. As reported earlier (Valdazo-González et al., 2014), the analysis establish a close phylogenetic relationship between the Ind2001 isolates of India, Saudi Arabia and Libya. Considering the limited sequence data, it was difficult to ascertain the origin and transmission of this lineage. It is required to characterize additional Ind2001 viruses from the Indian subcontinent and the Middle East to understand the epidemiology better.

These nine viruses sampled between June 2013 and January 2014 differed at 60 amino acid positions within the polyprotein among themselves. As expected, the maximum number of substitutions (11) was observed in the VP1 protein. Fifty six positions accepted two alternate amino acids and four positions, one each in VP3 (174), 3A (134), 2C (241) and 3D (158) accepted three alternate amino acids. The functional motifs in the LF-UTR, 'RGD' motif in VP1 and the critical sites in the NSPs including the catalytic triad in L and 3C proteins (reviewed in Carrillo et al., 2005) revealed no changes. In the polyprotein of Ind2001d (2013) isolates, 7 amino acids were found to be substituted compared to other Ind2001 isolates (Table 2). Out of the 7 substitutions, 2 each were found in L, 3A and VP1 proteins, and one in VP3 protein. The amino acid changes found in the 3A protein at positions 84 and 132 were non-conservative resulting in charge modification. Position 84 located just downstream to the hydrophobic domain and 132 lies between the two block deletions observed in some serotype O isolates of Asia (Knowles et al., 2001). In the L protein coding region, both the substitutions are located in the hyper-variable region between the two AUG codons. The inter-AUG region has been reported to dictate translation efficiency in FMD virus (Piccone et al., 2010). In the structural proteins, the changes were found at 3 positions including Thr142Pro and Thr158Pro substitutions in the βG-βH loop of VP1. Three codons including L-34, 3A-132

Please cite this article in press as: Subramaniam, S., et al. Genetic and antigenic analysis of foot-and-mouth disease virus serotype O responsible for outbreaks in India during 2013. Infect. Genet. Evol. (2014), http://dx.doi.org/10.1016/j.meegid.2014.12.009