Genetic Characterization of Peste des Petits Ruminants Virus from Recently Emerging Wave of Outbreaks in Pakistan

Summary

Peste des petits ruminants (PPR) is endemic in Pakistan and despite exhaustive vaccination, outbreaks are on the rise annually across different parts of the country. Clinical outcome is largely employed to diagnose disease, while detailed investigations on the genomic features of prevailing PPRV usually remain elusive. Here, we present comparative sequence based phylogenomic of field strains from three districts representing different agro-livestock production systems during an emerging wave of outbreaks in 2015, together with complete genome sequencing of one of the selected strains for the first time from Pakistan. The analysis revealed clustering of under-study strains to lineage IV close to isolates from India and China. Investigation of inter- and intra-lineage genetic distances revealed a higher genetic distance between study strain to lineage III viruses than lineages I and II. The strain showed a high percentage of genetic distance from ancestral isolates originating from Nigeria indicating a possible evolution of PPRV. Based on these observations, an integrated cross-protection investigation is warranted in the future, not only to define the protective efficacy of currently applied vaccines, but also to continuously elucidate the genomic and evolutionary nature of circulating viruses in the country to achieve PPR eradication by 2030.

KEYWORDS Peste des petits ruminants, Complete genome, Phylogenomic analysis, Percent identity, Genetic distance, Pakistan

1 INTRODUCTION

Peste des petits ruminants (PPR), caused by peste des petits ruminant virus (PPRV), is a contagious transboundary disease of small domestic and/or wild ruminants (Parida et al., 2015; Baron et al., 2016). The virus belongs to the genus *Morbillivirus* within the family *Paramyxoviridae* and consists of a non-segmented, negative sense, single stranded RNA genome that encodes six structural [nucleocapsid (N), phosphoprotein (P), matrix (M), fusion (F), hemagglutinin (H) and polymerase (L)] and two non-structural (C and V) proteins (Gibbs et al., 1979). The virus has been classified into four distinct lineages (I-IV) on the basis of partial sequence of N (225 bp) and F (322 bp) genes (Parida et al., 2015). All four lineages have been extensively isolated from different African countries whereas, in Asian countries, only lineage IV has been implicated for the epidemics (Parida et al., 2015; Baron et al., 2016).

Pakistan has an agriculture-based economy and livestock plays an integral role in the sustainability of national economy. However, it is challenged by various infectious and non-infectious factors resulting in severe annual losses. Among infectious threats, the occurrence of PPR is the leading infection since first reported in 1991 in Pakistan (Amjad et al., 1996). Despite available disease control measures, a number of outbreaks are being reported across the country with subsequent economic impacts. The economic loss for three PPR outbreaks was estimated to be US$ 12,211 including direct and indirect financial loss; the number of outbreaks throughout a year could result in US$ 240 million annual losses (Abubakar & Munir, 2014). Given the country-wide disease situation and global PPR eradication program by 2030, control of PPR is a benchmark of global food security and poverty alleviation. We have previously provided data on the nature of PPRV in Pakistani small ruminants based on the partial N and F gene sequencing (Munir et al., 2012; Anees et al., 2013). However, in order to offer higher resolution analysis of the virus currently circulating in the country, phylogenomic analysis of the entire genome is imperative. The goal of the current study was to perform a genome based comprehensive phylogenetic and genetic distance analysis of the complete genome of PPRV for the first time in Pakistan, one of the most PPRV-affected countries around the world.

2 MATERIALS AND METHODS

2.1 Study area and sampling

During 2015, an emerging wave of outbreaks were observed in small ruminants originating from different districts of Punjab province of Pakistan. These outbreaks generally infect the entire affected herd, irrespective of species and breed. Additionally, these outbreaks emerged in early winter of 2015 with an unusually high rate of infectivity in herds. Clinical symptoms such as high fever, nasal and lacrimal discharge, erosions in the oral mucosae, severe diarrhoea and abortion in pregnant animals were suggestive of PPR. We collected and processed the clinical samples (blood, mucosal erosion, nasal and lacrimal discharge) from herds of sheep (n = 7) and goat (n = 11) originating from three districts as per convenience in logistics and farmer consent: Lahore (n = 7, two sheep and five goat), Faisalabad (n = 8, three sheep and five goat) and Layyah (n = 3, two sheep and one goat). The selected districts represent traditional and mixed agro-livestock production systems in the country. The number of animals in the herds ranged from 10-150 heads with age ranging from 3 months to 4 years. The breeds of animals included Beetal, Daira Din Panah, Teddy and non-descript for goat, and Kajli, Thalli and Lohi for sheep. Each studied herd had no previous history of vaccination. A majority of animals died within the first five days of clinical disease, with

morbidity and case fatality ranging from 70-90% and 40-100%, respectively. All animals were found to be equally susceptible within infected herds, irrespective of age and breed.

2.2 Genome amplification and sequencing

All 92 samples representing 18 herds were confirmed to be PPRV positive by polymerase chain reaction (Couacy-Hymann et al., 2002). One sample per district was processed for partial sequencing of the N-gene using the primers applied for the detection, while one was processed for complete genome sequencing. Briefly, viral RNA was extracted as per manufacturer's guidelines (QIAamp Viral RNA extraction Mini Kit, Qiagen, Valencia city, CA, USA). The partial N-gene (Couacy-Hymann et al., 2002) and whole genome was amplified by one-step reverse transcriptase polymerase chain reaction (RT-PCR) using virus specific primers (List of primers can be provided upon request). The amplified PCR products were purified (Wizard® SV Gel and PCR Clean-Up System, Promega, Co., Madison, WI, USA) and sequenced in both directions through ABI PRISM Genetic Analyzer 3130x1 version (Applied Biosystems, Foster City, CA, USA).

2.3 Phylogenomic Analysis

The consensus sequence for one complete genome and two partial N sequences were assembled by Geneios® version 8.1.6 and submitted in GenBank under accession numbers (KY967608, MH004284 and MH004285). Partial and complete nucleotide sequences were compared to corresponding GenBank database using BLAST tool at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences were aligned with strains reported earlier from different countries using ClustalW methods in BioEdit® version 5.0.6 (Hall, 1999). Comparative phylogeny using complete genome, partial N (255 bp) gene of under-study isolates and previously reported representative strains of each lineage was estimated by neighbour-joining (1000 replication bootstrap values) method in MEGA® version 6.0 software (Tamura et al., 2013). The percent identity of nucleotide and amino acid of studied isolates was estimated by Pairwise Sequence Comparisons (PASC) analysis with representative strain using Kimura-2 model (1000 bootstrap) in MEGA software. For phylogenomic analysis of the complete genome, Splits Tree program v 4.95 was employed using Neighbor-Net graph method based on pairwise distance estimated by uncorrected *p*-distance and angle split transformation setting (Huson & Bryant, 2006). In order to estimate inter- and intra-lineage genetic distance, PASC analysis was performed using maximum composite likelihood method (d: Transitions + Transversions model) with in MEGA software (Tamura et al., 2013).

3 RESULTS AND DISCUSSION

All clinical samples were found positive, indicating a wide spectrum of susceptibility irrespective of animal breed, age and geographical area. Topology of partial N-gene revealed a wide geographical relationship of circulating PPR strains in the country (Figure 1A) consistent with observations made previously (Balamurugan et al., 2010; Munir et al., 2012; Muniraju et al., 2014). Since partial gene (either N or F gene)-based phylogeny represents an evolutionary pattern of a particular gene only, whole genome sequence analysis is recommended for more precise phylogenetic and evolutionary relationship of the circulating strains in a particular geographical setting (Muniraju et al., 2014), such as Pakistan where full-genomes have not yet been considered. Hence complete genome sequencing was performed using primers and protocols designed to represent local strains.

Genomes of PPRVs have been previously reported to range from 15942 to 15954 nucleotides (nt) in length, due to six-nucleotide indels in noncoding region between M and F gene (Balamurugan et al., 2010). The complete sequence of the under-study isolate was found to be 15948 nt in length, following the characteristic "rule-of-six" feature of paramyxoviruses required for efficient replication. Similar to previously reported genomes, the full genome of this isolate encoded six structural genes in order of 3′-N-P-M-F-H-L-′5 with a 52 nt long leader at 3′ UTR and 73 nt long trailer at 5′ UTR. Total length of coding genes varied in length across the whole genome: N with 1689 nt (55-1744), P with 1655 nt (1748-3402), M with 1484 nt (3406-4888), F with 2410 nt (4892-7302), H with 1957 nt (7306-9262) and L with 6643 nt (9266-1548). All genes were separated by similarly conserved non-coding intergenic trinucleotide (CTT) (Baron et al., 2016).

Based on complete genome sequences, phylogenetic analysis clustered the under-study isolate within lineage IV, close to previously reported isolates from China and India (Figure 1B). A highest nucleotide identity (97.9%) was observed with an Indian strain (KR140086) isolated in 1994 followed by 97.5% with Chinese strain (KX421388) isolated in 2007 and 97.3% with an Indian isolate (KX033350) reported recently in 2016. The reported isolate showed a higher nucleotide divergence (87.5%) with a strain in lineage III (KJ867543) followed by 88.4% with lineage I (EU267273) and 92.1% lineage II (HQ197753) (Table 1). Such a pattern of genetic divergence suggests an ongoing evolution among strains representing different lineages (Muniraju et al., 2014). Individual gene-based comparative analysis showed maximum homology for M-gene than other genes of representative strains in lineage IV (nucleotide: 96.7-98.7%, aa: 94.3-97.7%), whereas, it was found to be varying from 90.5-93.7% for nucleotides and 85.4-86.2% for residues within lineage I-III. The H gene was found to be the most divergent than other genes within lineage IV (nucleotide: 92.8-

95.9%, aa: 96.4-98.1%). Similarly, the nucleotide and residue divergence varied from 85.7 to 91.1% and 79.1%-85.8%, respectively within lineage I-III (Table 1). These findings are consistent with observations made previously where higher divergence and/or more hyper-variability has been reported in H gene than other genes of the PPRV (Yu et al., 1998).

The under-study strain shared high percentage genetic distance with Moroccan and Ethiopian isolates as compared to Chinese and Indian isolates with 12.4% inter-lineage genetic distance between lineage III and II and/ I, 12.1% between lineage IV and III, 10.1% between lineage IV and I, and, 8.2% between lineage IV and II (Figure 2A). The current study presented the first intra-lineage genetic distance analysis of PPRV isolates originating from different countries (Figure 2B). Based on findings of genetic distance, we suggest that the under-study isolate is closer to Chinese and Indian isolates as compared to isolates from Nigeria, considered the ancestral origin of all circulating PPRV strains. This finding also indicates the continuous evolution of PPRV (Muniraju et al., 2014). Comparative residue analysis of the complete N and F gene showed several conserved motifs in our isolate similar to previously reported isolates (Balamurugan et al., 2010). These include three motifs in N-gene: a nuclear export signal motif (^{4}LLKSLALF^{11}), a nuclear localization signal motif (^{70}TGVMISML^{77}) and the RNA binding motif (^{324}FSAGAYPLLWSYAMG^{338}) involved in interaction of N with N monomers of RNA during genomic RNA binding and thought to be required for N-N self-interaction (Yu et al., 1998). Three conserved motifs were also noticed in the F gene: a signal peptide (^{1}MTRVAILAFLFLNAVAC^{19}), cleavage site motif (^{103}GRRTRR^{108}) responsible for virulence and adaptation in the environment and a leucine zipper domain (^{459}LGNAVTRLENAKELLDASDQIL^{380}) involved in maintenance of protein tertiary structure (Lamb & Parks, 2007).

Taken together, assessing the complete genetic nature of field-circulating PPRV strains highlights the level of divergence and genetic differences compared to vaccine or circulating isolates from neighbouring countries. A possible link of intra-lineage genetic diversity of isolates from different countries would be helpful not only in understanding the genetic basis of circulating viruses but also to facilitate the establishment of foundations to exploit such information in designing future vaccine viruses that confer better protection in a given country or situation. Additionally, continuous monitoring of the disease emergence to determine the nature of the virus and to assess the potential of viral evolution, would be a benchmark of success in disease eradication from the globe.

ACKNOWLEDGMENT

None

CONFLICT OF INTEREST

All authors declare no competing of interest

FIGURE LEGENDS

FIGURE 1 Phylogenetic tree of partial N-gene (**A**) and whole genome (**B**) of under-study PPRV sequences compared to representative strains from different geography was constructed using MEGA version 6.0. The study PPRV isolates are marked with black-square (■).

FIGURE 2 Inter and intra-lineage genetic distance for complete genome of PPRV strains are determined using evolutionary network in Splits Tree software. The isolates originated from different countries are highlighted with different colours accordingly.