

Recombination may drive the emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis.

Basanta Sahu¹, Prativa Majee¹, Ravi Singh¹, and Debasis Nayak¹

¹Indian Institute of Technology Indore

June 16, 2020

Abstract

Contagious ecthyma is a zoonotic disease caused by the Parapoxvirus virus, Orf (ORFV) affecting sheep, goats, humans and is widely distributed across the world, including India. Here, we have investigated the 2017 ORFV outbreak in goats that occurred in the central Indian state of Madhya Pradesh. The outbreak was characterized by a moderate rate of morbidity (up to 20%) with no mortality. Phylogenetic analysis of four partial genes, such as ORFV011, ORFV020, ORFV059, and ORFV108, revealed the transboundary potential of the virus by showing its relationship with a distinct geographical area. We characterized the complete genome of this strain named Ind/MP and performed comparative genomic analysis. The Ind/MP whole genome consists of 139,807 bp with GC content 63.7%. The genome contains 132 potential open reading frames (ORFs) flanked by inverted terminal repeats (ITRs) of 3,910 bp at both ends having terminal BamHI sites and conserved telomere resolution sequences. Population genetic parameters such as nucleotide diversity (π), selection pressure analysis ($\theta=dN/dS$), etc. suggest that the ORFV resides under purifying selection. The potent molecular markers, such as Simple Sequence Repeats (SSRs) and compound SSRs (cSSRs) were more prevalent within the functional protein with the value 70% and 67%, respectively. A total number of forty recombination events were identified, out of which Ind/MP strain actively participate in twenty-one events suggesting that this strain can recombine for the generation of new variants.

Abstract

Contagious ecthyma is a zoonotic disease caused by the Parapoxvirus virus, Orf (ORFV) affecting sheep, goats, humans and is widely distributed across the world, including India. Here, we have investigated the 2017 ORFV outbreak in goats that occurred in the central Indian state of Madhya Pradesh. The outbreak was characterized by a moderate rate of morbidity (up to 20%) with no mortality. Phylogenetic analysis of four partial genes, such as ORFV011, ORFV020, ORFV059, and ORFV108, revealed the transboundary potential of the virus by showing its relationship with a distinct geographical area. We characterized the complete genome of this strain named Ind/MP and performed comparative genomic analysis. The Ind/MP whole genome consists of 139,807 bp with GC content 63.7%. The genome contains 132 potential open reading frames (ORFs) flanked by inverted terminal repeats (ITRs) of 3,910 bp at both ends having terminal BamHI sites and conserved telomere resolution sequences. Population genetic parameters such as nucleotide diversity (π), selection pressure analysis ($\theta=dN/dS$), etc. suggest that the ORFV resides under purifying selection. The potent molecular markers, such as Simple Sequence Repeats (SSRs) and compound SSRs (cSSRs) were more prevalent within the functional protein with the value 70% and 67%, respectively. A total number of forty recombination events were identified, out of which Ind/MP strain actively participate in twenty-one events suggesting that this strain can recombine for the generation of new variants. **Keywords:** *ORFV*, *Phylogenetics*, *Evolution*, *NGS*, *Recombination*

Introduction

Orf or contagious ecthyma is a neglected zoonotic disease caused by the Orf virus (ORFV), a member of the genus *parapoxvirus*. It primarily affects the sheep and goats, occasionally other ruminants and wild animals as well as animal handlers, veterinarians and para veterinarians (Guo, Rasmussen, Wünschmann, & de La Concha-Bermejillo, 2004; Haig & Mercer, 1998; Inoshima, Ito, & Ishiguro, 2010; Vikøren et al., 2008). The virus is extremely contagious and can exist within the animals' wools and excreta for months to years (Haig & Mercer, 1998). The disease is exhibited by proliferative skin lesions like erythematous papule, macule and scab around the buccal cavity, often leading to weight loss and anorexia. Ulcers are usually large, proliferative, and have 2–4 mm raised crusts and usually get resolved within 1 to 2 months (Gökce, Genc, & Gökce, 2005). Severe symptoms are characterized by the expansion of lesions, pneumonia, and arthritis in goats (De La Concha-Bermejillo, Guo, Zhang, & Waldron, 2003). Early stage diagnosis may be difficult due to asymptomatic nature of pathology (De La Concha-Bermejillo et al., 2003; Moriello & Cooley, 2001). The morbidity rate often reaches up to 100%, resulting in loss of body weight in adults, and feebler kids, thereby negatively affecting the herd economy (Haig & Mercer, 1998). Zoonosis leads to infections in animal handlers by forming painful lesions in hand and can spread to other organs like face and genitals (Andreani et al.; Duchateau, Aerts, & Lambert, 2013; Rajkomar, Hannah, Coulson, Owen, & dermatology, 2016; Turk, Senturk, Dereli, & Yaman, 2014).

The current report investigates the ORFV outbreak associated with the Black Bengal goat breed in India. The Black Bengal is a highly prolific dwarf breed of goat with unique features like excellent mutton quality, early reproductive maturity, high fecundity, and ability to sustain in the hot and humid climate of India (Acharya, 1982). In many tropical countries, sheep and goats act as a potential resource for meat, milk, fiber, and skin. The incidence of diseases and parasitic invasion is one of the major hindrances of goat farming. Among these, ORFV infections cause a significant economic loss. ORFV outbreaks are being reported across many countries over past years (Billinis, Mavrogianni, Spyrou, & Fthenakis, 2012; Gelaye et al., 2016; Kottaridi et al., 2006; Huixia Li et al., 2013; Maganga et al., 2016; Oem et al., 2009). Since 1999, ORFV outbreaks is observed across Indian states, majorly in Meghalaya, Assam (Bora et al., 2012), Tripura (Venkatesan et al., 2018), Odisha (Sahu, Majee, Sahoo, & Nayak, 2019), Uttarakhand (Venkatesan et al., 2011), Uttar Pradesh (Kumar et al., 2014), Kashmir Himalayas (Ahanger et al., 2018) and Tamil Nadu (Nagarajan et al., 2019).

The ORFV has a double-stranded DNA genome (134 to 139 kbp), accommodating ~130 putative genes. Genes in the central region are relatively more conserved and involved in mature virion formation and virus replication. In contrast, genes in the terminal regions are more variable and often attributed to the virulence and immunomodulation, and high frequency of gene recombination (Cottone et al., 1998; Fleming, Lyttle, Sullivan, Mercer, & Robinson, 1995). Despite its global distribution, only fourteen complete genome information is available so far. However, studies on molecular epidemiology of ORFV outbreaks in India, particularly in central India, are limited. The absence of a complete genome sequence of India isolates makes it difficult to comprehend genetic analysis and thus hinder further functional studies. We, therefore, performed molecular detection of ORFV isolates prevalent in central India for the first time performed the complete genome analysis of the circulating strain. Using next-generation sequencing (NGS) platform and comparative genomics approaches, we report recombination events and the phylogenetic and evolutionary status of the Indian ORFV isolate.

2. Materials and Methods

2.1. Goat herds and tissue collection

The study area is located in Dhar district of Madhya Pradesh, a central Indian state (75.30E, 22.59 N). Samples were collected from a private farm during January 2017. Samples (n=10) collected from goats aged between one to eleven months, showing typical Orf skin lesions on their lips and stored at -80 degC for virus isolation and further analysis. Samples were collected from a single herd of unvaccinated goats housed together, and therefore likely to have the same viral strain.

2.2. DNA extraction and virus conformation

Total genomic DNA was isolated from the skin tissue according to the protocol described by Sarker et al.

2017 using a DNeasy blood and tissue purification Kits (QIAGEN, Germany) (Sarker et al., 2017). A total of 25 mg of tissue was aseptically dissected and chopped and transferred into a microcentrifuge tube. After centrifugation for 2 minutes at 800xg the supernatants were subsequently filtered through 5 μ m centrifuge filters (Millipore) and used for viral DNA isolation. The filtrates were RNase A treated by incubating the tube at 56 °C for 10 min. The viral genomic nucleic (DNA) acids were subsequently extracted using Qiagen binding columns (QIAGEN, Germany).

The viral presence was confirmed by PCR utilizing four sets of primers targeting ORFV011, ORFV020, ORFV059, and ORFV108 (Supplement Table S1). The PCR amplified DNA was purified using a MiniElute gel extraction kit (QIAGEN, Germany) and sent for Sanger sequencing (Sanger, Nicklen, & Coulson, 1977). Prior to NGS, we analyzed four partial gene sequences to get a clear picture of intra and inter-strain relationships and constructed a gene-specific phylogenetic tree. This exercise facilitated the understanding of the transboundary potential and helped us to determine the reference genome for mapping, assembly and mutational analysis. For this, we retrieved sequences of ORFV011, ORFV020, ORFV059, and ORFV108 genes from the GenBank (Supplement Table S2). All phylogenetic trees were constructed using a general-time-reversible (GTR) substitution model for the maximum likelihood (ML) phylogeny with 1,000 bootstrap values using MEGA 6.0 (Sarker et al., 2017; Tamura et al., 2013). We also constructed a concatenated phylogenetic tree by taking these genes along with the homologous sequences of complete genomes to determine the reference genome. Virus isolation was attempted by passing clinical samples in primary lamb testis cell and Vero cells (Kottaridi et al., 2006; Mazur & Machado, 1989). However, even after the sixth blind passage, we could not recover the virus. So, we moved ahead with the NGS experiment by isolating viral DNA directly from the clinical samples (Bennett et al., 2017; Burioli, Prearo, & Houssin, 2017; Günther et al., 2017; Sarker et al., 2017).

2.3. Library construction and Illumina NextSeq500 sequencing

The paired-end sequencing library was prepared from the QC-passed viral DNA sample using Illumina TruSeq Nano DNA Library Prep Kit. Approximately 200 ng of QC-passed DNA was fragmented by Covaris M220 to generate a mean fragment distribution of 350 bp. Covaris shearing generates dsDNA fragments with 3' or 5' overhangs. The fragments were then subjected to end-repair. This process converts the overhangs resulting from fragmentation into blunt ends using End Repair Mix. The 3' to 5' exonuclease activity of this mix removes the 3' overhangs, and the 5' to 3' polymerase activity fills in the 5' overhangs followed by adapter ligation to the fragments. This strategy ensures a low rate of chimera (concatenated template) formation. The ligated products were size-selected using AMPure XP beads (Invitrogen, USA). The size-selected products were PCR amplified following temperature cycling profile: 72 degC initial denaturation for 3 min; 95 degC denaturation for 30 s; 12 cycles of 95 degC for 10 s, 55 degC for 30 s, and 72 degC for 30 s; and 72 degC final extension step for 5 min with the index primer as described in the kit protocol. Indexing adapters were ligated to the ends of the DNA fragments, preparing them for hybridization onto a flow cell. The PCR enriched library was analyzed on the 4200 Tape Station system (Agilent Technologies) using high sensitivity D1000 Screen tape as per manufacturer's instructions. Cluster generation and sequencing of the pooled DNA-library was performed as paired-end on NextSeq500 sequencing chemistry according to the manufacturer's instructions.

2.4. Raw sequence data processing, mapping, assembly, and genome annotations

The sequenced raw data were processed to obtain high-quality clean reads using Trimmomatic v0.38 platform to remove adapter sequences, ambiguous reads, and host sequences (Bolger, Lohse, & Usadel, 2014). The high-quality reads were aligned to the reference genome using BWA MEM software (version 0.7.17) (Heng Li & Durbin, 2009). Consensus sequences were extracted using SAMtools (Heng Li et al., 2009). The ORFV genome was annotated using GATU to capture all the potential open reading frames (ORFs) (Tcherepanov, Ehlers, & Upton, 2006). Intergenic regions were further checked for the presence of ORFs using the BWA MEM analysis tool (version 0.7.17). All the ORFs were subsequently extracted into a FASTA file, and similarity searches, including nucleotide (BLASTN) and protein (BLASTP) were performed. Annotation of ORFs as potential genes were established based on shared significant sequence similarity to the known viral

or cellular genes (BLAST E value[?]e-4) or contained a putative conserved domain as predicted by BLASTP. The final ORFV annotations were evaluated with other globally available ORFV isolates to determine the correct methionine start site, assign the proper stop codons, assign the truncation, and validate overlaps.

2.5. Mutations, DNA polymorphism, and evolutionary analysis

Mutations were identified in the newly assembled genome by manually comparing with the reference genome, by utilizing the BioEdit and ExPaSy tools. This led to the identification of unique nucleotide mutations with synonymous or non-synonymous amino acid changes. Next, by taking into account of all fourteen available complete genome sequences, we investigated for identification of polymorphic sites and different evolutionary parameters such as haplotype diversity (Hd), nucleotide diversity (π), selection pressure (θ), neutrality test (D) by using DnaSP (Librado and Rozas, 2009). To estimate selection pressure imposed upon ORFV, non-synonymous (dN) and synonymous (dS) substitution rates and their associated ratios ($\theta=dN/dS$) were estimated by using the bootstrap method with 100 replicates following the Kumar method in DnaSP. To infer the neutrality test, Tajima's D was evaluated based on numbers of polymorphic sites and the average number of nucleotide differences (Librado & Rozas, 2009).

2.6. Detection of Simple Sequence Repeats (SSRs) or Microsatellites

Identification of perfect mono, di, tri, tetra, penta, hexa as well as compound microsatellites were done by IMEx software (Mudunuri & Nagarajaram, 2007). Microsatellites from genomes were extracted using the 'Advance-Mode' of IMEx using the parameters previously used for RNA viruses (Chen et al., 2009) and DNA viruses (Hatcher, Wang, & Lefkowitz, 2015; Wu, Zhou, Zhao, & Tan, 2014). The parameters used were as follows: type of repeat: perfect; repeat size: all; minimum repeat number: 6, 3, 3, 3, 3, 3 for mono, di, tri, tetra, penta and hexanucleotide repeats, respectively. Maximum distance allowed between any two SSRs (dMAX) was 10 nucleotides. Other parameters were used as default. Compound microsatellites (cSSR) were not standardized in order to determine real composition.

2.7. Phylogenetic and recombination analysis

We retrieved fourteen available ORFV complete genome sequences for GenBank database. Additionally, five sequences of *Parapoxvirus* (PPV) and Orthopoxvirus, consisting of two *Pseudocowpox virus* (PCPV), one *Bovine papular stomatitis virus* (BPSV) and two *Monkeypox virus* (MPV), respectively were retrieved from the GenBank. These sequences were aligned by using MAFFT (version 7) (Katoh & Standley, 2013) and then manually edited using the BioEdit (version 7.2) (Hall, Biosciences, & Carlsbad, 2011) tool. MEGA 6 was used to create a phylogenetic tree based on GTR substitution model for the ML phylogeny with 1000 bootstrap (Tamura et al., 2013). To understand the source of genetic variation among all the ORFV complete genomes, we looked for evidence of recombination using the RDP, GENECONV, Bootscan, MaxChi, Chimaera, Siscan, PhylPro, LARD and 3Seq methods contained in the RDP4 program (Martin, Murrell, Golden, Khoosal, & Muhire, 2015). Events that were detected by at least three of the aforesaid methods with significant p-values were considered a plausible recombinant event.

3. Results

3.1. Clinical gross pathological changes

Affected animals showed characteristics of Orf lesions such as papules, pustules, and scabs around their lips. Infected goats showed ulcerated or proliferative lesions in the epidermis around the lips. Other pathological indications include anorexia, weight loss, and appearance of proliferative papillomatous nodules. All infected animals got recovered within 21 to 30 days after the onset of clinical signs. The herd morbidity of the outbreak was recorded by about 20% (n=50) with no mortality. None of the animal handlers of the farm got infected with the virus.

3.2. Virus conformation through PCR and sequencing

To verify the presence of the ORFV, four sets of primers targeting ORFV011, ORFV020, ORFV059, and ORFV108 genes were used for PCR amplification. All ten samples produced the expected DNA size. Since

samples were collected from an unvaccinated herd, we assume that a single strain circulating in the herd. Hence, we sequenced a representative sample and submitted the information to GenBank. Subsequently, gene-specific phylogenetic trees were constructed to infer the genetic relationship and transboundary potential of the circulating strain comparing within the country and with other global isolates (Figure 1A-1D).

Nucleotide BLAST and phylogenetic analysis based on these four genes confirmed that the present isolate has 99%-100% similarity with earlier published Indian isolate Ind/Od/Kh/01/2016 (Sahu et al., 2019). At the global level, the maximum similarity was observed with China/NP/2011, China/FJ-YT2015, China/HuB13/2013, and USA/ORFD/2003 isolates. The concatenated phylogenetic tree analysis indicated that the present isolate Ind/MP/2017 (Ind/MP) is closely related to the China/GO/2012 (Chi/GO) (KP010354) isolate, thus considered for the reference genome for mapping and mutations analysis (Figure 2).

3.3. NGS output and structural analysis of ORFV

On average, ~2.31 GB of high-quality data was generated with 7,934,711 numbers of quality reads by NextSeq 500 NGS platform. These sequences were aligned with Chi/GO strain of ORFV, and subsequently, the reference-based assembly was done by using BWA MEM (version 0.7.17). The total length of the assembled genome exhibited 139,807 bp in length, and the assigned NCBI accession number is MT332357. Like other PPV genomes, the genome possessed a high (63.7%) G+C content. The genome's left-most nucleotide was arbitrarily designated as base 1, and the starting point of the Inverted Terminal Repeats (ITRs), which spanned throughout ORFV001 and ORFV134 having a total length of 3,910 bp. Each ITR is composed of a terminal *Bam* HI site. Telomere resolution motifs are composed of TAAAT, followed by a spacer sequence, ACCCGACC, and six T residues, which form the terminal hairpin loop. (Figure 3). Using NCBI's ORF Finder tool and NCBI's BLAST (Basic Local Alignment Search Tool), we obtained 132 ORFs for a distinct set of genes. Among these, 65 genes orientated positively, and rest 67 have a negative orientation (Figure 4). Genes were numbered according to the method described by Delhon *et al.* (Delhon et al., 2004) and Mercer *et al.* (Andrew A. Mercer et al., 2006) in which the newly recognized genes (12.5 and 107.5) were also observed. The BLASTN and BLASTP results for ORFs varied from 63.85% to 100% and 22.73% to 100% with the existing complete genomes of other ORFV GenBank data (Supplementary File S2).

3.4. Mutations, DNA polymorphism, and evolutionary analysis

In comparison to the reference genome Chi/GO, our analysis showed nearly 488 unique mutations in the current isolate. Among all the predicted genes, 106 coding regions showed substitution mutations, while 26 genes lacked any mutation (Table 1). These observed mutations led to both synonymous and non-synonymous amino acid substitutions. The highest number of synonymous and non-synonymous amino acid substitutions was recorded in RNA helicase NPH-II, RNA-polymerase subunit RPO147, virion core protein P4a precursor and EEV maturation protein, Poly(A)-polymerase catalytic subunit PAPL, NF-kappa pathway inhibitor, DNA-binding protein, Ankyrin/F-box protein, respectively. Besides these, several other predicted proteins contained synonymous as well as non-synonymous mutations that support the notion of maintaining the heterogeneity and virulence of this pathogen. Using DNAsp, we observed 11,545 number of polymorphic sites within the ORFV genome. The nucleotide diversity (π) and haplotype diversity (Hd) were observed to be 0.02815 and 1.000, respectively. Selection pressure analysis ($\theta=dN/dS$), with a value of 0.02911 revealed that ORFV resides under purifying selection. Tajima's D test of neutrality, resulted in significant negative value (-0.14928), suggesting that this virus might be undergoing a period of evolutionary expansion.

3.5. Microsatellite and tandem repeat detection analysis

Our study revealed 1,108 and 94 numbers of SSRs and cSSR, scattered throughout the ORFV genome (Supplementary Table S3 and S4). Dinucleotide repeats were the most abundant (76.5%) in the genome, followed by a trinucleotide (18.14%), and mononucleotide repeats (5.14%). Hexanucleotide repeats were observed at least in number and represented at least 0.18% within the ORFV genome, respectively (Figure 5). There were no SSRs with tetranucleotide and pentanucleotide repeats observed in the ORFV genome. Approximately 89% and 11% of microsatellite motifs were distributed within the coding and non-coding

regions. Among the non-coding region, 4% were present within the UTR while 6% in the intergenic regions, where functional protein and hypothetical protein occupied 70% and 19%, respectively (Supplementary Figure S1). In the case of cSSR, 88% and 12% of microsatellite motifs were distributed within coding or non-coding regions. Within the non-coding region, 5.0% were represented in the UTR while 6% in the intergenic region, where functional and hypothetical (predicted) proteins occupied 67% and 21%, respectively (Supplementary Figure S2).

3.6. Phylogenetic and recombination analysis

To reveal the genetic relationship of the present strain with other PPVs and Orthopoxvirus, a phylogenetic tree based on the complete genomic sequences of 19 poxvirus strains was constructed. It revealed that six ORFV strains originating in goats and seven strains belonging to sheep formed two separate clades except for Ger/D1701 with 61-100% bootstrap support. The present ORFV strain showed a close relationship with Chi/GO and USA/ORFD isolates. Our analysis also showed that all ORFVs were more closely related to PPVs (PCPVs and BPSV) than to Orthopoxvirus (MPVs) (Figure 6). We observed a total of 40 potential recombination events with significant P -values detected across the ORFV genomes (Table 2). A number of recombination events were overlapped within RNA-polymerase subunit (events 1, 17 and 37), Ankyrin/F-box protein (events 4, 5 and 6), hypothetical protein (events 11, 19, 25, 26 and 40) and UTR region (events 3, 9, 10, 14, 23, 31 and 38). Out of 40, potential recombination events, 21 events include Ind/MP as recombinant, minor, and major parental sequences. The recombinant Ind/MP (event 30, 31, 39) were observed within hypothetical proteins, UTR and A-type inclusion protein, using USA/OV-IA82, Chi/SY17, Chi/GZ, Chi/NA1/11 as a minor parent and Chi/NP, Chi/GO, Chi/YX, Chi/GO Ger/D1701 as a major parent. The Ind/MP strain acts as minor and major parental sequences within a total of 12 and 6 events, for the generation of recombinant sequences Chi/YX, Chi/CL17, Chi/GZ, NZ/NZ2, USA/OV-IA82, and Chi/YX, Chi/SJ1, Ger/D1701, Chi/NP, respectively.

Discussion

Although a number of reports suggest the presence of ORFV throughout India, limited information exist on isolates of Central India. We characterized the ORFV isolate from the central Indian state of Madhya Pradesh by utilizing conserved partial genes analysis of ORFV011, ORFV020, ORFV059, and ORFV108. We observed a moderate rate of morbidity (20%) and no mortality during this outbreak, similar to the previous report of ORFV outbreak in North-Eastern state of Assam, India (Bora et al., 2012). The sequence and phylogenetic analysis revealed 99-100% nucleotide similarities with ORFV isolate of Odisha, a state of Eastern India. The global comparison was consistent with the previous study showing its closeness with the Chinese ORFV isolates (Kumar et al., 2014). This indicates the transboundary potential of the present virus isolates to spread to neighboring states or countries. We speculate that trade exchange and transport can be a potential route to establish the epidemiological linkage between the strains isolated from the two distinct geographical areas.

Despite the high number of outbreaks reported from India, no complete genome and associated genomic information are available. Ours is the first ORFV complete genome sequence report from India. NGS was performed from the DNA isolated from the clinical samples, similar to the recently adopted strategy reported in poxviruses, such as *seal parapoxvirus* (SePPV) (Günther et al., 2017), *canid alphaherpesvirus 1* (CHV-1) (Sarker et al., 2017), *kangaroo pox virus* (KPV) (Bennett et al., 2017; Sarker et al., 2017). The Chi/GO strain was assigned reference genome based on constructing a concatenated phylogenetic tree using all the fourteen complete genomes. All these features were relatively the same as the previously published ORFV genome around the world (Chi et al., 2015; Zhong et al., 2019). Like other poxviruses, ORFV genomes contain a large central coding region bound by two identical inverted terminal repeat (ITR) regions (Fraser, Hill, Mercer, & Robinson, 1990; Andrew A Mercer, Fraser, Barns, & Robinson, 1987). The current isolate was observed to have the ITRs of 3910 bp length and span throughout the ORFV001 and ORFV134 with conserved telomere resolution sequences at both ends of its genome, consistent with the previous report (Merchinsky, 1990).

Amino acid substitutions within the immune regulatory genes may have altered clinical manifestation and

refined antiviral response evaluation (Filer et al., 1995; Thomasy & Maggs, 2016). We observed that the highest number of non-synonymous substitutions within immune regulatory genes such as NF-kappa pathway inhibitor, and Ankyrin/F-box protein, etc. These mutations are likely to maintain the heterogeneity and mimicked the virulence of this pathogen. Selection pressure analysis ($\vartheta=dN/dS$) obtained the value 0.02911 suggested that ORFV under purifying selection. This result coincides with the recently reported ϑ value of avipoxvirus and ORFV partial genes, which range between 0.065 and 0.200 (Le Loc'h, Bertagnoli, & Ducatez, 2015; Sahu et al., 2019; Velazquez-Salinas et al., 2018). The synonymous (dN) and non-synonymous mutations (dS) drive the selection pressure to fluctuate the evolution rate.

Due to its polymorphic nature, microsatellites, otherwise known as Short Tandem Repeats (STRs) or Variable Number of Tandem Repeats (VNTRs) are being used for strain demarcation and estimation of evolutionary distance. Such approaches are being in place for a number of viruses such as human cytomegalovirus (hCMV) (Davis et al., 1999; Walker et al., 2001) white spot syndrome virus (WSSV) (Pradeep, Shekar, Karunasagar, & Karunasagar, 2008), Herpes Simplex Virus Type 1 (Deback et al., 2009), Herpes Simplex Virus type 2 (Burrell et al., 2013), Herpesvirus 3 (Avarre et al., 2011), Herpesvirus 6 (Achour et al., 2009), Adenovirus (Houng et al., 2009), Ostreid herpesvirus 1 (Renault et al., 2014), Marek's disease virus 1 (Spatz & Silva, 2007), and Spodoptera littoralis multiple nucleopolyhedrovirus (SpliMNPV) (Atia, Osman, & Elmenofy, 2016). Distribution of classified repeats suggests that dinucleotide GC/CG is more prevalent in most of the ORFV genomes, similar to other DNA viruses like, HPVs (Singh, Alam, Sharfuddin, & Ali, 2014), Caulimoviruses, Geminiviruses (B George, Alam, Kumar, Gnanasekaran, & Chakraborty, 2015; Biju George, Gnanasekaran, Jain, Chakraborty, & Evolution, 2014). Di-nucleotide repeat could form Z-conformation or other alternative secondary DNA to facilitate the recombination activity (Karlin, Campbell, & Mrazek, 1998; Treco & Arnheim, 1986). Here, the presence of higher dinucleotide repeats relative to tri-nucleotide repeats suggests a possible host role in the evolution of dinucleotide repeats within poxvirus genomes. By using a single mononucleotide repeat, Houng et. al, could follow the transmission dynamics of a human adenovirus during an epidemic (Houng et al., 2009). Therefore, these microsatellites could potentially be used as a powerful tool for epidemiological and evolutionary studies for ORFV.

The phylogenetic tree (of complete genome) showed the six goat ORFVs and seven sheep ORFVs formed distinctly separate branches except for Ger/D1701. Interestingly, this result slightly deviated from recently studied phylogenetic analysis based on the complete genome, which forms two host-specific clades (Andreani et al.; Zhong et al., 2019). However, our analysis in comparison to the previous study, showed increase the heterogeneity and unable to maintain the perfection of the host-specific clade. This kind of ambiguity was also observed during study of Phylodynamics analysis of parapoxvirus genus in Mexico (2007–2011), where Ger/D1701 with several other isolates exhibited a separate clade rather than host-specific clade (Velazquez-Salinas et al., 2018). Viruses undergo genetic recombination to form new variants in the population by deleting many of its non-essential genes or by acquiring new host genes (Fleming et al., 1995). Viral genome sequencing elucidates that recombination plays a vital role in understanding the evolution of human and animal pathogens, including *Vaccinia* and *Variola viruses* (Odom, Hendrickson, & Lefkowitz, 2009; Smithson, Purdy, Verster, & Upton, 2014). Remarkably, the comparative genomics approach of ORFV complete genome provides evidence of extensive recombination among the virus. Although several attempts were made to identify recombination events, lack of complete genome sequences made this study exercise unsuccessful (Chi et al., 2015). However, in this study, we identified forty potential recombination events where Ind/MP actively participated in more than 50% events by forming recombinant as well as major and minor parents. Thus, the Ind/MP strain has the potential to evolve via homologous or non-homologous site-specific recombination and can act as a major or minor parent to form new variants.

In conclusion, we report complete genome of circulating ORFV isolate from central India. Additionally, we estimated the transboundary and evolutionary potential of this isolate. Subsequently, by in-depth analysis through comparative genomic approach, we propose future recombination events to drive ORFV evolution and generation of new isolates in the region and globally. We hope that the current genomic information would be greatly useful for further understanding of ORFV biology, epidemiology, and research carried in the front of diagnosis and vaccine development.

SUPPLEMENTARY MATERIAL

Supplementary Figure S1: Distribution of SSRs, Supplementary Table S1: Primers used for virus identification sequencing and phylogenetic analysis, Supplementary Table S2: Sequence used for phylogenetic analysis, Supplementary Table S3: Overview of SSRs in ORFV whole genome sequences, Supplementary Table S4. Overview of cSSRs in ORFV whole genome sequences, Supplementary File 2: Comparative analysis of Ind/MP open reading frames (ORFs).

ACKNOWLEDGEMENTS

BPS is thankful to the University Grant Commission (UGC, Govt. India) for providing Ph.D. fellowship as financial support.

Author Contributions Statement

BPS processed the samples, did the molecular analysis, and written the manuscript. PM constructed the phylogenetic tree and figures. RS has generated the Circos plot. DN designed the study, executed, and written the manuscript.

Live Vertebrate Experiment Statement

No live experiment was conducted in the vertebrate animals for this study. Filed veterinarians collected clinical samples as a routine practice.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

I confirm that we have included the Accession number of the ORFV genome information in the manuscript.

REFERENCES

- Acharya, R. (1982). *Sheep and goat breeds of India* : Food and Agriculture Organization of the United Nations.
- Achour, A., Malet, I., Deback, C., Bonnafous, P., Boutolleau, D., Gautheret-Dejean, A., & Agut, H. (2009). Length variability of telomeric repeat sequences of human herpesvirus 6 DNA. *Journal of Virological Methods*, *159* (1), 127-130. doi:<https://doi.org/10.1016/j.jviromet.2009.03.002>
- Ahanger, S. A., Parveen, R., Nazki, S., Dar, Z., Dar, T., Dar, K. H., . . . Dar, P. J. V. (2018). Detection and phylogenetic analysis of Orf virus in Kashmir Himalayas. *29* (3), 405-410.
- Andreani, J., Fongue, J., Khalil, J. Y. B., David, L., Mougari, S., Le Bideau, M., . . . La Scola, B. Human Infection with Orf Virus and Description of Its Whole Genome, France, 2017.
- Atia, M. A., Osman, G. H., & Elmenofy, W. H. J. S. r. (2016). Genome-wide in silico analysis, characterization and identification of microsatellites in *Spodoptera littoralis* multiple nucleopolyhedrovirus (SpliMNPV). *6* (1), 1-9.
- Avarre, J.-C., Madeira, J.-P., Santika, A., Zainun, Z., Baud, M., Cabon, J., . . . Maskur, M. (2011). Investigation of Cyprinid herpesvirus-3 genetic diversity by a multi-locus variable number of tandem repeats analysis. *Journal of Virological Methods*, *173* (2), 320-327. doi:<https://doi.org/10.1016/j.jviromet.2011.03.002>
- Bennett, M., Tu, S.-L., Upton, C., McArtor, C., Gillett, A., Laird, T., & O'Dea, M. J. V. r. (2017). Complete genomic characterisation of two novel poxviruses (WKPV and EKPV) from western and eastern grey kangaroos. *242* , 106-121.
- Billinis, C., Mavrogianni, V. S., Spyrou, V., & Fthenakis, G. C. J. V. j. (2012). Phylogenetic analysis of strains of Orf virus isolated from two outbreaks of the disease in sheep in Greece. *9* (1), 24.

- Bolger, A. M., Lohse, M., & Usadel, B. J. B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *30* (15), 2114-2120.
- Bora, D. P., Barman, N. N., Das, S. K., Bhanuprakash, V., Yogisharadhya, R., Venkatesan, G., . . . Chakraborty, A. (2012). Identification and phylogenetic analysis of orf viruses isolated from outbreaks in goats of Assam, a northeastern state of India. *Virus Genes*, *45* (1), 98-104.
- Burioli, E., Prearo, M., & Houssin, M. J. V. (2017). Complete genome sequence of Ostreid herpesvirus type 1 μ Var isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. *509* , 239-251.
- Burrel, S., Ait-Arkoub, Z., Voujon, D., Deback, C., Abrao, E. P., Agut, H., & Boutolleau, D. (2013). Molecular characterization of herpes simplex virus type 2 (HSV-2) strains by analysis of microsatellite polymorphism. *Journal of Clinical Microbiology* .
- Chen, M., Tan, Z., Jiang, J., Li, M., Chen, H., Shen, G., & Yu, R. (2009). Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. *FEBS Letters*, *583* (17), 2959-2963. doi:10.1016/j.febslet.2009.08.004
- Chi, X., Zeng, X., Li, W., Hao, W., Li, M., Huang, X., . . . Wang, S. J. F. i. m. (2015). Genome analysis of orf virus isolates from goats in the Fujian Province of southern China. *6* , 1135.
- Cottone, R., Büttner, M., Bauer, B., Henkel, M., Hettich, E., & Rziha, H.-J. (1998). Analysis of genomic rearrangement and subsequent gene deletion of the attenuated Orf virus strain D1701. *Virus Research*, *56* (1), 53-67. doi:https://doi.org/10.1016/S0168-1702(98)00056-2
- Davis, C. L., Field, D., Metzgar, D., Saiz, R., Morin, P. A., Smith, I. L., . . . Wills, C. (1999). Numerous Length Polymorphisms at Short Tandem Repeats in Human Cytomegalovirus. *Journal of Virology*, *73* (8), 6265.
- De La Concha-Bermejillo, A., Guo, J., Zhang, Z., & Waldron, D. J. J. o. v. d. i. (2003). Severe persistent orf in young goats. *15* (5), 423-431.
- Deback, C., Boutolleau, D., Depienne, C., Luyt, C., Bonnafous, P., Gautheret-Dejean, A., . . . Agut, H. J. J. o. c. m. (2009). Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. *47* (3), 533-540.
- Delhon, G., Tulman, E., Afonso, C., Lu, Z., De la Concha-Bermejillo, A., Lehmkuhl, H., . . . Rock, D. L. J. J. o. v. (2004). Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *78* (1), 168-177.
- Duchateau, N. C., Aerts, O., & Lambert, J. (2013). Autoinoculation with Orf virus (ecthyma contagiosum). *International Journal of Dermatology*, *53* (1), e60-e62. doi:10.1111/j.1365-4632.2012.05622.x
- Filer, C., Ramji, J., Allen, G., Brown, T., Fowles, S., Hollis, F., & Mort, E. J. X. (1995). Metabolic and pharmacokinetic studies following oral administration of famciclovir to the rat and dog. *25* (5), 477-490.
- Fleming, S. B., Lyttle, D. J., Sullivan, J. T., Mercer, A. A., & Robinson, A. J. J. J. o. G. V. (1995). Genomic analysis of a transposition-deletion variant of orf virus reveals a 3.3 kbp region of non-essential DNA. *76* (12), 2969-2978.
- Fraser, K. M., Hill, D. F., Mercer, A. A., & Robinson, A. J. J. V. (1990). Sequence analysis of the inverted terminal repetition in the genome of the parapoxvirus, orf virus. *176* (2), 379-389.
- Gelaye, E., Achenbach, J. E., Jenberie, S., Ayelet, G., Belay, A., Yami, M., . . . Lamien, C. E. J. V. j. (2016). Molecular characterization of orf virus from sheep and goats in Ethiopia, 2008–2013. *13* (1), 34.
- George, B., Alam, C. M., Kumar, R. V., Gnanasekaran, P., & Chakraborty, S. J. V. (2015). Potential linkage between compound microsatellites and recombination in geminiviruses: Evidence from comparative

analysis. *482* , 41-50.

George, B., Gnanasekaran, P., Jain, S., Chakraborty, S. J. I., Genetics, & Evolution. (2014). Genome wide survey and analysis of small repetitive sequences in caulimoviruses. *27* , 15-24.

Gökce, H. I., Genc, O., & Gökce, G. (2005). Sero-prevalence of contagious ecthyma in lambs and humans in Kars, Turkey. *Turkish Journal of Veterinary and Animal Sciences*, *29* (1), 95-101.

Günther, T., Haas, L., Alawi, M., Wohlsein, P., Marks, J., Grundhoff, A., . . . Fischer, N. J. S. r. (2017). Recovery of the first full-length genome sequence of a parapoxvirus directly from a clinical sample. *7* (1), 1-8.

Guo, J., Rasmussen, J., Wünschmann, A., & de La Concha-Bermejillo, A. J. V. m. (2004). Genetic characterization of orf viruses isolated from various ruminant species of a zoo. *99* (2), 81-92.

Haig, D. M., & Mercer, A. A. (1998). Ovine diseases. Orf. *Vet Res*, *29* (3-4), 311-326.

Hall, T., Biosciences, I., & Carlsbad, C. J. G. B. B. (2011). BioEdit: an important software for molecular biology. *2* (1), 60-61.

Hatcher, E. L., Wang, C., & Lefkowitz, E. J. J. V. (2015). Genome variability and gene content in chordo-poxviruses: dependence on microsatellites. *7* (4), 2126-2146.

Houng, H. S., Lott, L., Gong, H., Kuschner, R. A., Lynch, J. A., & Metzgar, D. (2009). Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. *J Clin Microbiol*, *47* (7), 2243-2248. doi:10.1128/jcm.01659-08

Inoshima, Y., Ito, M., & Ishiguro, N. (2010). Spatial and temporal genetic homogeneity of orf viruses infecting Japanese serows (*Capricornis crispus*). *J Vet Med Sci*, *72* (6), 701-707. doi:10.1292/jvms.09-0467

Karlin, S., Campbell, A. M., & Mrazek, J. (1998). Comparative DNA analysis across diverse genomes. *Annual review of genetics*, *32* , 185-225. doi:10.1146/annurev.genet.32.1.185

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30* (4), 772-780. doi:10.1093/molbev/mst010

Kottaridi, C., Nomikou, K., Teodori, L., Savini, G., Lelli, R., Markoulatos, P., & Mangana, O. J. V. m. (2006). Phylogenetic correlation of Greek and Italian orf virus isolates based on VIR gene. *116* (4), 310-316.

Kumar, N., Wadhwa, A., Chaubey, K. K., Singh, S. V., Gupta, S., Sharma, S., . . . Mishra, A. J. V. G. (2014). Isolation and phylogenetic analysis of an orf virus from sheep in Makhdoom, India. *48* (2), 312-319.

Le Loc'h, G., Bertagnoli, S., & Ducatez, M. F. (2015). Time scale evolution of avipoxviruses. *Infection, Genetics and Evolution*, *35* , 75-81. doi:https://doi.org/10.1016/j.meegid.2015.07.031

Li, H., & Durbin, R. J. b. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *25* (14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. J. B. (2009). The sequence alignment/map format and SAMtools. *25* (16), 2078-2079.

Li, H., Zhu, X., Zheng, Y., Wang, S., Liu, Z., Dou, Y., . . . Luo, X. J. A. o. v. (2013). Phylogenetic analysis of two Chinese orf virus isolates based on sequences of B2L and VIR genes. *158* (7), 1477-1485.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25* (11), 1451-1452. doi:10.1093/bioinformatics/btp187

Maganga, G. D., Relmy, A., Bakkali-Kassimi, L., Ngoubangoye, B., Tsoumbou, T., Bouchier, C., . . . Berthet, N. J. V. j. (2016). Molecular characterization of Orf virus in goats in Gabon, Central Africa. *13* (1), 79.

- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. J. V. e. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *1* (1).
- Mazur, C., & Machado, R. D. (1989). Detection of contagious pustular dermatitis virus of goats in a severe outbreak. *Veterinary Record*, *125* (16), 419. doi:10.1136/vr.125.16.419
- Mercer, A. A., Fraser, K., Barns, G., & Robinson, A. J. J. V. (1987). The structure and cloning of orf virus DNA. *157* (1), 1-12.
- Mercer, A. A., Ueda, N., Friederichs, S.-M., Hofmann, K., Fraser, K. M., Bateman, T., & Fleming, S. B. (2006). Comparative analysis of genome sequences of three isolates of Orf virus reveals unexpected sequence variation. *Virus Research*, *116* (1), 146-158. doi:https://doi.org/10.1016/j.virusres.2005.09.011
- Merchlinsky, M. J. J. o. v. (1990). Mutational analysis of the resolution sequence of vaccinia virus DNA: essential sequence consists of two separate AT-rich regions highly conserved among poxviruses. *64* (10), 5029-5035.
- Moriello, K. A., & Cooley, J. (2001). Difficult dermatologic diagnosis. Contagious viral pustular dermatitis (orf), goatpox, dermatophilosis, dermatophytosis, bacterial pyoderma, and mange. *J Am Vet Med Assoc*, *218* (1), 19-20. doi:10.2460/javma.2001.218.19
- Mudunuri, S. B., & Nagarajaram, H. A. (2007). IMEx: Imperfect Microsatellite Extractor. *Bioinformatics*, *23* (10), 1181-1187. doi:10.1093/bioinformatics/btm097
- Nagarajan, G., Pourouchottamane, R., Reddy, G. M., Yogisharadhya, R., Sumana, K., Rajapandi, S., . . . Rajendiran, A. J. V. w. (2019). Molecular characterization of Orf virus isolates from Kodai hills, Tamil Nadu, India. *12* (7), 1022.
- Odom, M. R., Hendrickson, R. C., & Lefkowitz, E. J. J. V. r. (2009). Poxvirus protein evolution: family wide assessment of possible horizontal gene transfer events. *144* (1-2), 233-249.
- Oem, J.-K., Roh, I.-S., Lee, K.-H., Lee, K.-K., Kim, H.-R., Jean, Y.-H., & Lee, O.-S. J. V. j. (2009). Phylogenetic analysis and characterization of Korean orf virus from dairy goats: case report. *6* (1), 167.
- Pradeep, B., Shekar, M., Karunasagar, I., & Karunasagar, I. (2008). Characterization of variable genomic regions of Indian white spot syndrome virus. *Virology*, *376* (1), 24-30. doi:https://doi.org/10.1016/j.virol.2008.02.037
- Rajkomar, V., Hannah, M., Coulson, I., Owen, C. J. C., & dermatology, e. (2016). A case of human to human transmission of orf between mother and child. *41* (1), 60-63.
- Renault, T., Tchaleu, G., Faury, N., Moreau, P., Segarra, A., Barbosa-Solomieu, V., & Lapègue, S. (2014). Genotyping of a microsatellite locus to differentiate clinical Ostreid herpesvirus 1 specimens. *Vet Res*, *45* (1), 3. doi:10.1186/1297-9716-45-3
- Sahu, B. P., Majee, P., Sahoo, A., & Nayak, D. J. A. G. (2019). Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha). *12*, 100088.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74* (12), 5463.
- Sarker, S., Roberts, H. K., Tidd, N., Ault, S., Ladmore, G., Peters, A., . . . Raidal, S. R. J. S. r. (2017). Molecular and microscopic characterization of a novel Eastern grey kangaropox virus genome directly from a clinical sample. *7* (1), 1-13.
- Singh, A. K., Alam, C. M., Sharfuddin, C., & Ali, S. (2014). Frequency and distribution of simple and compound microsatellites in forty-eight Human papillomavirus (HPV) genomes. *Infection, Genetics and Evolution*, *24*, 92-98. doi:https://doi.org/10.1016/j.meegid.2014.03.010

Smithson, C., Purdy, A., Verster, A. J., & Upton, C. J. P. o. (2014). Prediction of steps in the evolution of variola virus host range. *9* (3).

Spatz, S. J., & Silva, R. F. (2007). Polymorphisms in the repeat long regions of oncogenic and attenuated pathotypes of Marek’s disease virus 1. *Virus Genes*, *35* (1), 41-53. doi:10.1007/s11262-006-0024-5

Tamura, K., Stecher, G., Peterson, D., FilipSKI, A., Kumar, S. J. M. b., & evolution. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *30* (12), 2725-2729.

Tcherepanov, V., Ehlers, A., & Upton, C. J. B. g. (2006). Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *7* (1), 150.

Thomasy, S. M., & Maggs, D. J. J. V. o. (2016). A review of antiviral drugs and other compounds with activity against feline herpesvirus type 1. *19* , 119-130.

Treco, D., & Arnheim, N. (1986). The evolutionarily conserved repetitive sequence d(TG.AC)n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Molecular and Cellular Biology*, *6* (11), 3934-3947.

Turk, B. G., Senturk, B., Dereli, T., & Yaman, B. (2014). A rare human-to-human transmission of orf. *Int J Dermatol*, *53* (1), e63-65. doi:10.1111/j.1365-4632.2012.05669.x

Velazquez-Salinas, L., Ramirez-Medina, E., Bracht, A. J., Hole, K., Brito, B. P., Gladue, D. P., . . . Evolution. (2018). Phylodynamics of parapoxvirus genus in Mexico (2007–2011). *65* , 12-14.

Venkatesan, G., Balamurugan, V., Bora, D. P., Yogisharadhya, R., Prabhu, M., & Bhanuprakash, V. J. V. I. (2011). Sequence and phylogenetic analyses of an Indian isolate of orf virus from sheep. *47* (3), 323-332.

Venkatesan, G., De, A., Arya, S., Kumar, A., Muthuchelvan, D., Debnath, B. C., . . . Pandey, A. J. V. (2018). Molecular evidence and phylogenetic analysis of orf virus isolates from outbreaks in Tripura state of North-East India. *29* (2), 216-220.

Vikøren, T., Lillehaug, A., Åkerstedt, J., Bretten, T., Haugum, M., & Tryland, M. J. V. m. (2008). A severe outbreak of contagious ecthyma (orf) in a free-ranging musk ox (*Ovibos moschatus*) population in Norway. *127* (1-2), 10-20.

Walker, A., Petheram, S. J., Ballard, L., Murph, J. R., Demmler, G. J., & Bale, J. F. (2001). Characterization of Human Cytomegalovirus Strains by Analysis of Short Tandem Repeat Polymorphisms. *Journal of Clinical Microbiology*, *39* (6), 2219.

Wu, X., Zhou, L., Zhao, X., & Tan, Z. (2014). The analysis of microsatellites and compound microsatellites in 56 complete genomes of Herpesvirales. *Gene*, *551* (1), 103-109. doi:https://doi.org/10.1016/j.gene.2014.08.054

Zhong, J., Guan, J., Zhou, Y., Cui, S., Wang, Z., Zhou, S., . . . Zhai, S. J. V. g. (2019). Genomic characterization of two Orf virus isolates from Jilin province in China. *55* (4), 490-501.

Tables

Gene	Gene product	Number unique synonymous substitutions	Number unique non-synonymous substitutions	Non-synonymous Substitutions
ORFV001	hypothetical protein	0	0	0
ORFV002	NF-kappa-p65 acetylation inhibitor	0	0	0
ORFV005	hypothetical protein	0	1	Lys ⁵⁰ -Gln

ORFV007	dUTPase	0	1	Thr ⁶ -Met
ORFV008	Ankyrin/F-box protein	10	2	Asn ¹⁸⁵ -Asp, Thr ⁴³⁹ -Ala
ORFV009	hypothetical protein	10	2	Val ¹⁶⁴ -Leu, Ala ³⁰⁸ -Thr
ORFV010	EEV maturation protein	10	3	Asn ⁴⁵ -Asp, Val ²⁹³ -Met, Thr ³⁷⁸ -Ala, Ser ²⁹ -Gly
ORFV011	EEV envelope phospholipase	6	1	
ORFV012	hypothetical protein	3	0	
ORFV012.5	hypothetical protein	1	1	Val ¹⁴ -Leu
ORFV013	hypothetical protein	1	0	
ORFV014	RING-H2 motif protein	4	0	
ORFV015	hypothetical protein	4	1	Thr ⁴⁸³ -Ala
ORFV016	hypothetical protein	2	3	Ser ¹⁰⁵ -Gly, Ile ¹⁷⁶ -Met, Tyr ¹⁸⁷ -His
ORFV017	DNA-binding phosphoprotein	0	0	
ORFV018	Poly(A)-polymerase catalytic subunit PAPL	1	3	Ala ³⁶ -Glu, Cys ¹²¹ -Arg, Ile ³² -Val
ORFV019	hypothetical protein	11	0	
ORFV020	DsRNA-binding, interferon resistance	1	0	
ORFV021	RNA-polymerase subunit RPO30	0	0	
ORFV022	Pox virus E6 proteins	4	0	
ORFV023	Membrane protein	4	0	
ORFV024	NF-kappa pathway inhibitor	3	3	Phe ²⁰ -Ser, Ser ³⁴ -Gly, Lys ³⁹ -Glu
ORFV025	DNA-polymerase	10	2	Thr ⁶³⁷ -Ala, Ser ⁸²⁰ -Ala
ORFV026	ERV ALR-like protein (IMV redox protein)	0	0	
ORFV027	Virion core protein	0	1	Asp ⁷⁹ -Glu,

ORFV028	DNA-binding protein	8	3	Thr ¹²¹ -Ala, Ile ³³⁹ -Met, Ser ⁷⁰⁷ -Ala
ORFV029	hypothetical protein	8	2	Thr ²³ -Ile, Gln ⁵⁷ -Arg,
ORFV030	DNA-binding virion protein	7	0	
ORFV031	hypothetical protein	0	0	
ORFV032	DNA-binding phosphoprotein	5	1	Met ⁹⁷ -Ile
ORFV033	IMV membrane protein	3	0	
ORFV034	Telomere-binding protein	5	1	Leu ²⁷⁶ -Met
ORFV035	Virion core protease	6	0	
ORFV036	RNA helicase NPH-II	15	0	
ORFV037	Zn-protease, virion morphogenesis	6	0	
ORFV039	hypothetical protein	0	0	
ORFV038	Late transcription elongation factor	0	0	
ORFV040	Glutaredoxin-like protein	0	0	
ORFV041	hypothetical protein	7	1	Glu ³⁰³ -Ala
ORFV042	RNA-polymerase subunit RPO7	0	0	
ORFV043	hypothetical protein	2	0	
ORFV044	Virion core protein	1	0	
ORFV045	Late transcription factor VLTF-1	2	0	
ORFV046	Myristylated protein	5	3	Thr ¹²³ -Ala, Ala ²¹⁰ -Glu
ORFV047	Myristylated IMV envelope protein	5	0	
ORFV048	hypothetical protein	0	0	
ORFV049	hypothetical protein	2	3	Arg ²²⁰ -Ser, Val ¹⁹⁴ -Met, Ser ¹⁵⁸ -Gly
ORFV050	DNA-binding virion core protein VP8	2	2	Arg ¹⁶⁵ -Lys, Glu ²⁵⁶ -Asp

ORFV051	Membrane protein	2	1	Ser ⁴⁵ -Ala
ORFV052	IMV membrane protein	0	1	Val ⁹³ -Ala
ORFV053	Poly(A)-polymerase small subunit VP39	4	0	
ORFV054	PAPS RNA-polymerase subunit RPO22	2	0	
ORFV055	Late membrane protein	3	0	
ORFV056	RNA-polymerase subunit RPO147	13	0	
ORFV057	Tyrosine phosphatase, virus assembly	2	0	
ORFV058	IMV, viral entry	2	0	
ORFV059	Immunodominant envelope protein	4	1	Thr ¹¹⁰ -Ala
ORFV060	RNA-polymerase associated protein RAP94	2	1	Ser ⁵⁴⁹ -Gly
ORFV061	Late transcription factor VLTF4	2	2	Asp ⁷⁰ -Glu, Asp ¹³⁴ -Glu
ORFV062	DNA topoisomerase type I	3	0	
ORFV063	hypothetical protein	2	0	
ORFV064	mRNA capping enzyme large subunit	11	0	
ORFV065	Virion protein	1	0	
ORFV066	Virion protein	1	0	
ORFV067	Uracil DNA glycosidase	1	1	Thr ⁸ -Pro
ORFV068	NTPase	2	2	
ORFV069	Early transcription factor	3	3	
ORFV070	RNA-polymerase subunit RPO18	0	2	Met ⁵² -Leu, Asp ¹⁶⁷ -Tyr
ORFV071	NPH-PPH downregulator (NTP pyrophosphohydrolase)	3	0	
ORFV072	Transcription termination factor NPH-I	5	0	

ORFV073	hypothetical protein	3	1	Gln ¹²⁰ -Arg
ORFV074	mRNA capping enzyme small subunit	6	0	
ORFV075	Rifampicin resistance protein	0	0	
ORFV076	Late transcription factor VLTF2	0	0	
ORFV077	Late transcription factor VLTF3	0	0	
ORFV078	Thioredoxin-like protein	0	0	
ORFV079	Virion core protein P4b precursor	8	0	
ORFV080	Virion core protein	2	1	Val ¹⁴⁵ -Ala
ORFV081	RNA-polymerase subunit RPO19	0	0	
ORFV082	hypothetical protein	5	0	
ORFV083	Early transcription factor	15	0	
ORFV084	Intermediate transcription factor VITF-3	1	0	
ORFV085	Late virion membrane protein	0	0	
ORFV086	Virion core protein P4a precursor	13	2	Ser ⁶¹¹ -Ala, Thr ²³⁰ -Ala
ORFV087	Virion formation	0	0	
ORFV088	Virion core protein	4	1	Gly ²⁰⁴ -Ser
ORFV089	Virion membrane protein	1	0	
ORFV090	IMV phosphorylated membrane protein	0	0	
ORFV091	IMV membrane protein	1	0	
ORFV092	hypothetical protein	1	1	Thr ⁶² -Ala
ORFV093	Myristylated protein	3	0	
ORFV094	phosphorylated IMV membrane protein	2	0	
ORFV095	DNA helicase	7	0	

ORFV096	Zn-finger protein	1	0	
ORFV098	hypothetical protein			
ORFV097	DNA-polymerase processivity factor	2	1	Ser ²³⁰ -Ala
ORFV099	Holliday junction resolvase	1	0	
ORFV100	Intermediate transcription factor VITF-3	1	0	
ORFV101	RNA-polymerase subunit RPO132	7	0	
ORFV102	A-type inclusion protein/fusion peptide hybrid	3	0	
ORFV103	A-type inclusion protein	0	0	
ORFV104	Viral fusion peptide	3	0	
ORFV105	IMV surface protein	2	0	
ORFV106	RNA-polymerase subunit RPO35	2	0	
ORFV107	Virion morphogenesis	0	0	
ORFV107.5	hypothetical protein	0	0	
ORFV108	DNA packaging protein ATPase	3	0	
ORFV109	EEV glycoprotein	0	0	
ORFV110	EEV glycoprotein	0	0	
ORFV111	hypothetical protein	2	2	Thr ⁸⁹ -Ala, Ala ¹²⁵ -Ser
ORFV112	chemokine binding protein	3	1	Asp ¹⁶⁷ -Glu
ORFV113	hypothetical protein	0	0	
ORFV114	hypothetical protein	7	0	
ORFV115	hypothetical protein	2	0	
ORFV116	hypothetical protein	0	1	Thr ¹³⁵ -Ala
ORFV117	GM-CSF IL-2 inhibition factor	4	0	
ORFV118	hypothetical protein	1	0	
ORFV119	hypothetical protein	1	0	
ORFV120	hypothetical protein	1	1	Phe ¹⁵⁸ -Leu

ORFV121	NF-kappa pathway inhibitor	4	1	Glu ²⁹³ -Asp
ORFV122	hypothetical protein	6	0	
ORFV123	Ankyri/F-box protein	5	3	Gln ¹⁷ -Arg, Val ¹⁵⁴ -Ala, Ile ²²⁸ -Met
ORFV124	hypothetical protein	6	5	Gly ²⁶ -Ser, Ala ⁸² -Val, Thr ³³¹ -Cys,
ORFV125	Apoptosis inhibitor	2	0	
ORFV126	Ankyrin/F-box protein	1	0	
ORFV127	IL-10-like protein	1	1	Ala ¹⁰ -Val
ORFV128	Ankyrin/F-box protein	7	2	Ala ²⁰⁹ -Asp, Leu ⁴²⁷ -Pro
ORFV129	Ankyrin/F-box protein	7	0	
ORFV130	Serine/threonine protein kinase	3	1	Glu ⁴⁰ -Asp
ORFV131	Membrane protein	2	1	Ser ¹⁷⁸ -Gly
ORFV132	VEGF-like protein	3	1	Leu ⁷ -Ile, Asp ⁸⁴ -Glu, Thr ⁹⁹ -Val
ORFV134	hypothetical protein	0	0	

Table 1: Mutational analysis ORF isolates. The number of unique synonymous and unique non-synonymous amino acid substitutions across all isolates sequenced for this study are presented in the table. Individual non-synonymous substitutions are also shown in the column on the right of the table.

Recombination event number	BP* start Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection methods	P-value
1	102311	105339	RNA-polymerase subunit RPO132	Chi/YX	Ger/D1701	Chi/NP	RGBMCST	6.93E-316
2	108034	109113	A-type inclusion protein	Chi/YX	NZ/NZ2	Ind/MP Chi/GO USA/ORFD	RGBMCST	6.54E-36
					Chi/CL17 Chi/SY17	Chi/GO USA/ORFD		

Recombination event number	BP* start in Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection methods	P-value
3	1207	1668	UTR	Chi/CL17	Chi/GZ USA/OV-IA82 Chi/NA1/11 Chi/OV-HN3/12 Ind/MP	NZ/NZ2	RGBMCST	6.53E-68
4	132207	134268	Ankyrin/F-box protein	Chi/CL17	Chi/GZ Chi/GO Chi/NP Chi/YX	Chi/NA1/11 Chi/OV-HN3/12	RGC	2.78E-38
5	130127	131711	Ankyrin/F-box protein	Ger/D1701	Chi/SJ1 Chi/GO Chi/YX Chi/GZ	USA/OV-IA82 Chi/NA1/11 Chi/OV-HN3/12 Chi/SJ1	RGBMCST	5.57E-11
6	130025	130883	Ankyrin/F-box protein	USA/OV-IA82	Chi/CL17 Chi/SY17 Chi/NA1/11 Chi/OV-HN3/12 Unknown (Chi/GZ)	NZ/NZ2	RGBMCST	9.78E-15
7	134408	134735	VEGF-like protein	Chi/GZ	Ind/MP	Chi/CL17	RGBMCST	7.42E-29
8	43796	46928	Virion core protein	Chi/CL17	Chi/GO	Chi/SY17	RGC	6.56E-10
9	6793	9301	UTR	Chi/CL17	Chi/YX	Chi/SY17	RGBMCS	7.68E-15
10	6694	9265	UTR	Chi/GZ	Chi/GO Chi/YX	Chi/SY17	RGBMCST	7.86E-10
11	119339	121370	hypothetical protein	Chi/CL17	Chi/GO Chi/YX	Chi/GZ	RGBMCST	9.86E-11
12	2839	5566	UTR	Ger/D1701	Chi/NA1/11	USA/ORFD	RBMC	9.36E-09

Recombination event number	BP* start Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection methods	P-value
13	113590	114703	chemokine binding protein	Chi/GZ	USA/OV-IA82 NZ/NZ2 Chi/OV-HN3/12 USA/ORFD	Chi/CL17	RGBMCS	9.50E-20
14	2381	2683	UTR	Chi/SJ1	Ind/MP Chi/GO Chi/YX Chi/GZ	Chi/SY17	RBMCS	5.10E-18
15	81359	82959	Thioredoxin-like protein	Chi/CL17	Chi/CL17 Chi/SY17 Chi/YX	Chi/SY17	RGBMCST	5.93E-07
16	44481	44785	RNA-polymerase subunit RPO7	USA/ORFD	Chi/GO Ger/D1701	Chi/YX	RGBMCST	7.95E-04
17	102268	105087	RNA-polymerase subunit RPO132	NZ/NZ2	Ger/D1701	Chi/GO Chi/NP USA/OV-IA82 Chi/NA1/11	RMCT	5.04E-09
18	38787	41928	RNA helicase NPH-II,Zn-protease	Ger/D1701	USA/OV-IA82	Chi/OV-HN3/12 USA/ORFD	RBMCS	5.91E-08
19	13159	14477	hypothetical protein	Chi/GZ	NZ/NZ2 Chi/YX Chi/GO Chi/NP	Ind/MP Chi/GO Chi/NP Chi/SJ1 Chi/SY17	RGBMCST	5.06E-04

Recombination event number	BP* start in Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection methods	P-value
20	15309	16345	DNA-binding phosphoprotein	Chi/GZ	Chi/YX	Chi/CL17	RGBMCS	5.72E-14
21	59685	60433	Immunodominant envelope protein	Chi/D1701	Chi/GO Chi/GZ	Chi/YX	GBMCS	6.45E-13
22	109850	110870	Virion morphogenesis protein, hypothetical protein	Chi/NP	Chi/CL17 Chi/SY17 Chi/GZ	Chi/NP Chi/SJ1	RGBMCST	7.99E-08
23	1672	2991	UTR	Chi/CL17	Chi/SY17 Chi/GZ	Chi/SY17	RGMCS	8.26E-09
24	29748	30996	DNA-binding protein	Chi/GZ	Chi/SJ1	Chi/CL17	RGBMCS	5.95E-04
25	112618	113071	hypothetical protein	Chi/GZ	Ind/MP Chi/GO Chi/YX	Chi/SY17 Chi/CL17	RGMST	5.64E-06
26	119241	120473	hypothetical protein	Chi/GZ	Chi/SJ1	Chi/SY17	RBMCS	7.12E-07
27	26828	27986	DNA-polymerase IMV redox protein	Chi/GZ	Chi/SJ1	Chi/SY17	RGBMCS	5.20E-04
28	81188	82997	Thioredoxin-like protein	Chi/GZ	Chi/SJ1	Chi/SY17	RBMCS	7.99E-03
29	39578	40476	RNA helicase NPH-II	Chi/GZ	USA/ORFD	Chi/SY17	RGBMCS	5.03E-03
30	44373	45334	hypothetical protein	Ind/MP	USA/OV-IA82 Chi/SY17	Chi/NP Chi/GO	RGBMT	8.96E-09
31	5293	6693	UTR	Ind/MP	Chi/GZ USA/OV-IA82	Chi/YX Chi/GO	RGBMCST	5.07E-03

Recombination event number	BP* start Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection methods	P-value
32	126180	126569	Apoptosis inhibitor	Chi/CL17	Chi/SJ1	Chi/GZ	GBT	6.56E-05
33	117522	117772	GM-CSF/IL-2 inhibition factor	Chi/GZ	Chi/NP Chi/YX	Chi/SY17	GBT	3.59E-02
34	97779	101729	Zn-finger protein DNA-polymerase processivity factor Holliday junction resolvase	Chi/CL17 NZ/NZ2	Chi/YX Ind/MP	Chi/SY17 NZ/NZ2 Chi/OV-HN3/12 Chi/NA1/11	RBMCT	9.87E-03
35	65150	67860	mRNA capping enzyme large subunit	USA/OV-IA82	Chi/SJ1 Ger/D1701 Ind/MP	Chi/NA1/11	RBMT	5.94E-05
36	24856	26088	DNA-polymerase	Chi/GZ	Ind/MP	Chi/OV-HN3/12 NZ/NZ2	RBS	2.65E-02
37	105045	105144	RNA-polymerase subunit RPO132	Ger/D1701	Chi/NA1/11	Chi/NP	RGMC	7.09E-04
38	7733	8471	UTR	USA/OV-IA82	Chi/OV-HN3/12 Ger/D1701	Ind/MP Chi/GO USA/ORFD Chi/NA1/11 Chi/OV-HN3/12	RBT	8.35E-04

Recombination event number	BP* start Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant Se- quence(s)	Minor Parental Se- quence(s)	Major Parental Se- quence(s)	Detection methods	P-value
39	105510	106144	A-type inclusion protein	Ind/MP	Chi/NA1/11	Ger/D1701	RBM	7.49E-03
40	12729	14917	RING-H2 motif protein hypothetical protein	USA/OV-IA82	Ind/MP	Chi/NA1/11	RBT	5.64E-03

Table 2: Predicted potential recombination events between ORFV. Details recombination events detected between ORFV isolates were analyzed in this study using RDP4 program. Detection method coding R, G, B, M, C, S, P, L and T represents methods RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD and 3Seq, respectively. The P value for the detection method is shown in bold. To emphasize on the recombination potential of Indian isolate, (Ind/MP) has been bold. Ind/MP strain consists of 21 potential events in which it acts as recombinant, minor and major parental sequences.

Figure Legends

Figure 1: Phylogenetic trees constructed based on the four ORFV genes, namely ORFV011, ORFV020, ORFV059, and ORFV108. The phylogenetic relationship was constructed by GTR model of the maximum-likelihood method using MEGA 6.0 software. Numbers at the branching points indicate the bootstrap support calculated for 1,000 replicates (A) ORFV011, (B) ORFV020, (C) ORFV059 and (D) ORFV108 based on partial nucleotide sequences. Black triangle represents the new isolate mentioned in our study.

Figure 2: Construction of concatenated phylogenetic tree to determine reference genome. Concatenated DNA sequences comprising ORFV011, ORFV020, ORFV059 and ORFV108 genes of ORFV was used to construct a phylogenetic tree with bootstrap value 1000 using GTR model of maximum likelihood method. Black triangle represents the ORFV isolate of present investigation showing its close relationship with the strain KP010354, which is considered as the reference.

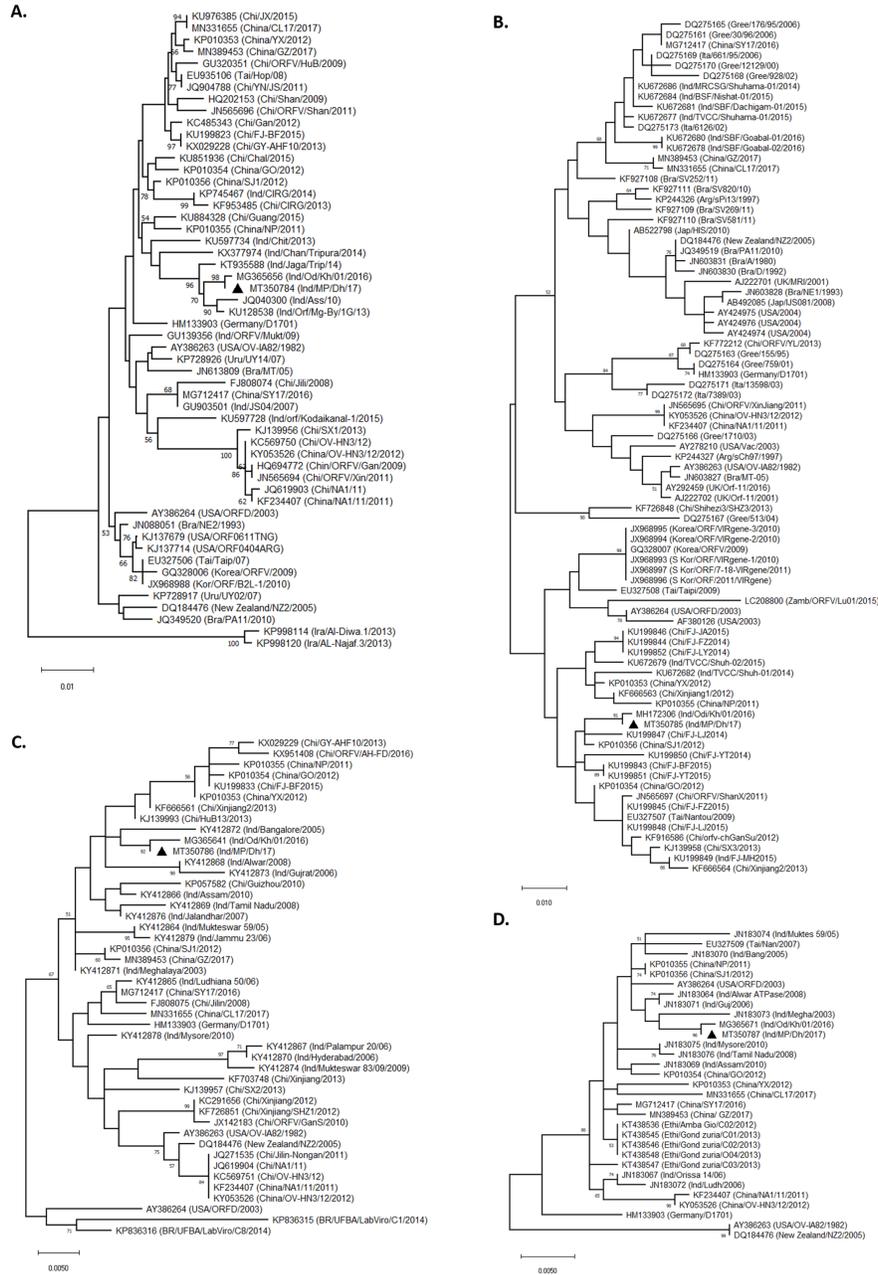
Figure 3: Analysis of ITR-ends of ORFV. (A) The left end (5') sequence alignment of 14 complete ORFV genomes consists of a terminal BamHI site (green box) along with telomere resolution motifs (red box) (ATTTTTT-N(8)-TAAAT). (B) The right end (3') sequence alignment of 14 complete ORFV genomes consists of a terminal BamHI site (green box) along with telomere resolution motifs (red box) (ATTTTTT-N(8)-TAAAT).

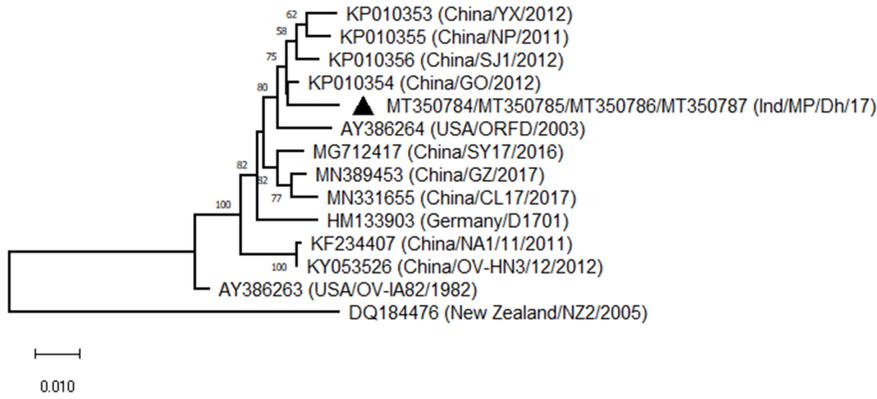
Figure 4: Circos plot illustrating complete genome of the Ind/MP ORFV strain. Schematic presentation of complete genome: from outer to inner track: ORFV genome size with 1Kb division, genes on a negative strand (black), genes on a positive strand (red), SSRs, cSSRs, Synonymous mutation, Non synonymous mutation. These mutations along with SSRs and cSSR were ubiquitously present throughout the genome irrespective of gene orientation.

Figure 5: Classification of SSR distribution in ORFV genome. Distribution of different motifs of SSRs within ORFV genomes. Dinucleotide repeats were found to be the most abundant, followed by

trinucleotide and mononucleotide repeats in the reported genome. Among the di and trinucleotide repeats CG/GC and CGC/GCG having highest number among all classified repeats.

Figure 6: Comparison of PPVs by constructing Phylogenetic tree . Nineteen complete genome sequences including the terminal repetitions, were aligned to construct a phylogenetic tree with bootstrap value 1000 using GTR model of maximum likelihood method. Black triangle represents the ORFV isolates of present investigation showing its relationship with fourteen global strains.





A.

Species/Abbrv	Sequence
1. AY386263_(USA/OV-IA82)	CC - AG AGGCTCCGC GAA - - - AAGTTTTATAAAAAGTTTTG - GAGAGTACCTCTAAGTTTCA
2. AY386264_(USA/ORFD/2003)	CCCCGGAGGTTCCCGGAA - - - AAGTTTTTACAAAAGTTTTGCGAGGAGGCTCGCTGTGCC - - - TCC TA
3. DQ184476_(NZ/NZ2)	TCC - GGAGGTTCCCGGAA - - - AAGTTTT - ATAAAAGTTTT - GAGGAGGT - - - ACCGACC - - - TCC TA
4. HM133903_(Ger/D1701)	CTCCGCCAGCA - TGC GCGGCCCTCGCCGCGGAGCGCCAGCTCCTGACGCACAGCAGCCG - GCCTCCGA
5. KF234407_(Chi/NA11/11)	TCT - AGAGGCTCCCGGAA - - - AAGTTTTATAAAAAGTTTT - GAGAGAGGCC - GACTGCCT - - - - -
6. KP010353_(Chi/YX)	CCCCGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATTA AA
7. KP010354_(Chi/GO)	CCCCGTAGAAAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATTA AA
8. KP010355_(Chi/NP)	TCC - - - - - GA - - - - - AAGCTTTACAAAAGTTTTTC - - - GAGAGGCCGACTACCC - - - - -
9. KP010356_(Chi/SJ1)	CCCCGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCGGCCCAATAAATTA AA
10. KY053526_(Chi/OV-HN3/12)	TCT - AGAGGCTCCCGGAA - - - AAGTTTTATAAAAAGTTTT - GAGAGAGGCC - GACTGCCT - - - - -
11. MG712417_(Chi/SY17)	CCCCGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATTA AA
12. MN331655_(Chi/CL17)	CCCCGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATTA AA
13. MN389453_(Chi/GZ)	CCCCGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATTA AA
14. MT332357_(Ind/MP/Dh/17)	CCCCGTAGAAAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATTA AA

B.

Species/Abbrv	Sequence
1. AY386263_(USA/OV-IA82)	TTGGAGAGGAG - GGCCGA CCTCCCAA - - - - - AGATT - TTGGGSAACGTTT - G - GAGAGGAGGACT - G
2. AY386264_(USA/ORFD/2003)	TTTCGAGAGAGGCGGACCTA CTTCCAACGTCCCGAGAAAAGTT - TTGCAGAAAGTTTGAGAGAGGTCGACT - G
3. DQ184476_(NZ/NZ2)	CTGGAGAGGAG - GGCCGA CCTCCCAA - - - - - AGATT - TTGGGSAACGTTT - G - GAGAGGAGGACT - G
4. HM133903_(Ger/D1701)	TTCCGTATGACAATAGCA CACAGTAGGTCAATATATCTTACTTATCTCC - TCCTGTCTACAGTCTGTCTCTCA
5. KF234407_(Chi/NA11/11)	TTTCGAGAGGAG - GGCCGA CCTCCCAA - - - - - AGATT - TTGGGSAACGTTT - G - GAGAGGAGCACTTG
6. KP010353_(Chi/YX)	CGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAG - - - - A
7. KP010354_(Chi/GO)	CGTAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAG - - - - A
8. KP010355_(Chi/NP)	CCCCGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAG - - - - A
9. KP010356_(Chi/SJ1)	CGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAG - - - - A
10. KY053526_(Chi/OV-HN3/12)	TTTCGAGAGGAG - GGCCGA CCTCCCAA - - - - - AGATT - TTGGGSAACGTTT - G - GAGAGGAGCACTTG
11. MG712417_(Chi/SY17)	CGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAGAAATCA
12. MN331655_(Chi/CL17)	CGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAGAAATCA
13. MN389453_(Chi/GZ)	CGGAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAGAAATCA
14. MT332357_(Ind/MP/Dh/17)	CGTAGAGCCCGGATCCGCTGTTCCGGCGAAGGCAGGACAGACATTTTTTCAGCCCAATAAATAAAGAAATCA

