Genomic features and evolution of the Parapoxvirus during the past two decades

Xiaoting Yao
Northwest Agriculture and Forestry University https://orcid.org/0000-0002-2853-6240
Ming Pang
Northwest Agriculture and Forestry University
Tianxing Wang
Northwest Agriculture and Forestry University
Xi Chen
Northwest Agriculture and Forestry University
Xidian Tang
Northwest Agriculture and Forestry University
Jianjun Chang
Qinghai University
Dekun Chen (chendekun163@163.com)
Northwest Agriculture and Forestry University
Wentao Ma (mawentao@nwafu.edu.cn)
Northwest Agriculture and Forestry University

Research article

Keywords: Parapoxvirus, Poxviridae, nucleotide composition, selection pressure, evolution

Posted Date: July 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-42668/v1

Version of Record: A version of this preprint was published at Pathogens on October 27th, 2020. See the published version at https://doi.org/10.3390/pathogens9110888.

Genomic features and evolution of the Parapoxvirus during the past two decades

Xiaoting Yao[1], Ming Pang[1], Tianxing Wang[1], Xi Chen[1], Xidian Tang[1], Jianjun Chang[2,3], Dekun Chen[1]*, Wentao Ma[1]*

[1] College of Veterinary Medicine, Northwest A&F University, Yangling 712100, Shaanxi Province, China

[2] College of Agriculture and Animal Husbandry, Qinghai University, Xining 810016, Qinghai Province, China

[3] State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, Qinghai Province, China

* Correspondence: mawentao@nwafu.edu.cn

Email addresses: yaoxiaoting@nwafu.edu.cn (Xiaoting Yao), pangming@nwafu.edu.cn (Ming Pang), 2017011072@nwafu.edu.cn (Tianxing Wang), 2017011069@nwafu.edu.cn (Xi Chen), tangxidian@nwafu.edu.cn (Xidian Tang), 383322290@qq.com (Jianjun Chang), chendekun163@163.com (Dekun Chen, Corresponding author), mawentao@nwafu.edu.cn (Wentao Ma, Corresponding author)

Keywords

Parapoxvirus, *Poxviridae*, nucleotide composition, selection pressure, evolution

Abstract

Parapoxvirus (PPV) has been identified in most mammals and poses a great threat to both the livestock production and public health. However, it is still not fully

understood the viral prevalence and evolution of PPV coding sequences. Here, we performed a comparative approach integrating viral genetics, molecular selection pressure and genomic structure to investigate the genomic features and evolution of PPVs. We noticed that although there were significant differences of GC contents between ORFV and other three species of PPVs, all PPVs showed almost identical nucleotide bias, that is GC richness. This reflected a common mechanism which determines GC compositions for virus with similar life cycles. The structural analysis of PPV genomes showed the divergence of different PPV species, which may due to the specific adaptation to their natural hosts. Additionally, we estimated the phylogenetic diversity of segmented genome of PPV. Our results suggested that during the 2010 – 2018 outbreak, the orf virus has been the dominant species under the selective pressure of the optimal gene patterns. Furthermore, we found the mean substitution rates were between 3.56×10^{-5} to 4.21×10^{-4} in different PPV segments, and the PPV VIR gene was evolved at the highest substitution rate. In these protein-coding regions, purifying selection was the major evolutionary pressure, while the GIF and VIR genes suffered the greatest positive selection pressure. These results may provide useful knowledge on the virus genetic evolution from a new perspective which could help create prevention and control strategies.

1. Introduction

Parapoxvirus (PPV) is a chordopoxvirus that causes nonsystemic skin lesion diseases in wild mammals worldwide. PPV belongs to *Poxviridae* and contains orf virus (ORFV), pseudocowpox virus (PCPV), bovine papular stomatitis virus (BPSV) and parapoxvirus of red deer in New Zealand (PVNZ) [1-3]. The PPV genome is approximately 134~139 kb in size, which is terminally connected by hairpin loops. In contrast to other poxvirus genera, PPVs genomes have an extremely high GC content of 65% evenly [4-6]. The genetic structure of PPVs exhibit the general pattern form of a central conserved region, with essential genes in regard to position, spacing and orientation [7]. The terminal variable regions encode genes with vital roles in viral infection process such as the regulation of host immune response to infection and the

determination of host range [6, 7].

PPVs have a global distribution, and the ruminant flocks will be disturbed while virus emerged in many continents [8-10]. This virus is known to infect mammals, such as goats, sheep, cattle, camels, reindeer and other ruminants [11, 12]. Human will also be infected by PPVs, following direct exposing to skin lesions from infected mammals. Regular public health controlling and reporting of human PPVs infections are in place in a lot of regions of the world, containing Europe, South America and South America [11, 13-17]. Until now, there are live PPVs vaccines which may control the disease, but it may not entirely prevent this disease [18].

ORFV is the prototype species of PPVs, which will cause high mortality rates when lesions occur in lips and breast of suckling and grazing animals [6, 19]. Ruminants could be infected with ORFV more than once, although with a shorter period of time to recovery and less obvious pathological changes than in the primary infection [20]. BPSV could infect cattle of all ages, however, clinical signs are usually observed in calves. Like ORFV, repeated infection of ruminants with BPSV is common , indicating that this virus infection may not confer significant immunity [6]. Except for the most important viruses (ORFV and BPSV), PCPV and PVNZ both maintained in mammals. As the zoonotic agents, all of them are worldwide epidemics and threaten the public health. Recently, attenuated vaccines could limit the severity of viral infection but they could not completely prevent the disease [6, 21-25].

Vaccines have been indicated to impose powerful selection pressure on the virus evolution [26, 27]. Viruses may acquire greater virulence through genetic variation and recombination, as a result of the rapid transmission in vaccinated populations. And they usually have a high mutation rate with their low accuracy of reproduction, which may assist the viruses to evade host immune attack and evolve further. The PPV has a large genome, comprising nearly 134~139 kb in size. Consequently, mutation is more likely to occur in PPVs under vaccination and selection pressures [2, 7, 28].

This work reported the global population dynamics of PPVs regarding the evolutionary process and genetic diversity. On the basis of PPVs genomes, the sequence comparison and phylogenetic analysis were performed to investigate the genomic

features of viral strains. From the dataset of PPVs genome sequences, the evolutionary rate was also considered by Bayesian methods. Additionally, we further explored the different genes that have played an important role in PPVs evolution by estimating the genetic diversity and the selection pressure in this study.

2. Materials and Methods

PPVs sequences

PPVs complete genome sequences and nucleotide sequences of the B2L, CBP, F1L, GIF, VIR, VEGF and VLTF-1 genes were downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank/). To explore the emergent time of the earliest PPV isolate, the sequences with an unknown collection date were excluded. The host, collection date and sampling location of each PPV isolates were determined either from GenBank, or were obtained from the related publications. This led to a dataset of 17 complete genome sequences and 776 gene sequences, including 268 B2L genes, 38 CBP genes, 83 F1L genes, 133 GIF genes, 79 VIR genes, 122 VEGF genes and 51 VLTF-1 genes, respectively (Table S1). The resulting sequences were aligned using MUSCLE 3.7 [29].

Gene structure and nucleotide composition

The gene structure of PPV was visualized by comparing their genomic nucleotide sequences using the R program gggenes project. Nucleotide compositional analysis of the PPVs complete genome sequences were measured using BioEdit v7.2.5 [30], including the frequencies of A, T, G, C, AT and GC. The nucleotide occurrence frequencies of the 3rd codon position (T3, G3, C3, and A3) of synonymous codons were measured by the CodonW v1.3.

Compositional change of PPV isolates over time

During 1974 – 2018, 268 B2L genes of PPV strains were downloaded from GenBank and analyzed by maximum-likelihood method in PAML v4.9, with 1000 bootstrap replicates. An online tool – the Interactive Tree Of Life v2 was performed to design the tree [31]. The resulting isolates were composed of 4 species in PPVs population, including ORFV, PCPV, BPSV and PVNZ. Isolates were assigned on the

basis of the PPVs species and the initial collection country. The variation number of PPV was calculated using statistical methods and the plot was drown with the project ggplot2 in R program [32].

Evolution substitution rates

For all PPV genes, the molecular evolution rates were estimated by the MCMC program of BEAST v2.4.8. This program gives a maximum-likelihood estimate of the evolution substitution rate, based on a model that considered a constant evolution rate (molecular clock). We also estimated the confidence intervals for all the parameters.

To obtain the best-fitting models for each viral gene, jModelTest v2.1.7 was performed by Akaike Information Criterion (AIC) [33], which was estimated for each model tested according to the approach of Newton & Raftery [34]. Here, we used general time reversible model, a strict molecular clock and a coalescent model with constant available population size, as performed in the BEAST package. Bayesian Markov chain Monte Carlo (MCMC) analysis was implemented with 10 million steps and was sampled every 10000 steps with 10% burn-in. Tracer v1.6 was used to analyze the results of running BEAST. The effective sample size was greater than 200 for each parameter assessment in the MCMC analysis. Statistical uncertainty in the dataset was estimated in the 95% HPD values.

Genetic diversity and phylogenetic analysis

The genetic diversity (π) of PPV was evaluated as average pairwise difference between viral sequences, using the Tamura-Nei nucleotide substitution models performed in MEGA 5 [35]. BioEdit was implemented to estimated the sequence identity for different gene segments [30], and the heatmap was plotted by R program pheatmap project. For each PPV gene, bayesian phylogenetic tree was inferred by Bayesian Evolutionary Analysis Sampling Trees Software (BEAST) [36]. For all analyses we performed a coalescent constant population model and the general-time reversible model of nucleotide substitution [37]. Based on the correlation between collection date and evolutionary distance for each gene, a strict molecular clock was selected, as evaluated by TempEst [38]. Here, MCMC analysis was run for 10 million

steps, 10% of which was removed as burn-in and parameters sampled every 10000 steps. The effective sample size for all the traces are greater than 200.

Selection pressures in PPVs

Selection pressure was highlighted as the ratio between the average number of non-synonymous nucleotide substitutions (dN) and synonymous (dS) nucleotide substitutions per site (dN/dS), and was implemented in the CODEML program within the PAML software package. Besides, paired models of dN/dS were estimated among nucleotide sequences, containing M0 versus M3, M2a versus M1a and M8 versus M7. We applied the likelihood ratio test to compare the neutral and positive selection models, and chi-square test to indicate the divergence between the compared models. We also calculated the dN and dS values respectively.

3. Results

G and C nucleotide was higher enriched in PPV genomes.

To analyze the basic characteristic of virus genome, the nucleotide contents of PPV whole genomes were calculated. It was shown that the mean compositions (%) of G (32.16 ± 0.32) and C (32.12 ± 0.22) were significantly higher than A (17.95 ± 0.28) and T (17.77 ± 0.25) (Table S2, Fig. 1(A), t-test, $P < 0.01$). The mean percentages of G3 (36.92 ± 1.02) and C3 (37.12 ± 1.33) were also significantly higher than A3 (20 ± 1.02) and U3 (19.94 ± 0.89), highlighting that G- and C- ended codons are more likely occurred in the PPV whole genome sequences (Table S2, Fig. 1(B), t-test, $P < 0.01$). According to the species classification, it was important to mention that similar trend of nucleotide composition was observed among different species, also suggesting higher G/C or G3/C3 contents in PPV genomes. However, there were significant differences in the nucleotide composition parameters trends among these four species (Table S2, Fig. 1C-D, t-test, $P < 0.01$). As shown in Fig. 2, We further compared the nucleotide composition among four different PPV species. Strikingly, there were highly unusual patterns of nucleotide contents, and were significantly different from the nucleotide composition patterns in the various PPV species.

Different genome structures exist in various PPVs.

According to the phylogenetic relationship and genomic structure, PPV genera belongs to *Poxviridae* and includes four species: ORFV, PCPV, BPSV and PVNZ(Fig. 3). The genome of PPV contains large central protein-coding regions encoding functional proteins, such as major envelope protein (B2L), chemokine binding protein (CBP), immunodominant envelope protein (F1L), GM-CSF/IL-2 inhibition factor (GIF), vascular endothelial growth factor (VEGF), viral interferon resistance protein (VIR) and late transcription factor (VLTF). Besides these major structural proteins, different PPVs encode some special accessory proteins, including Ankyrin protein (ANK), IL-10-like protein (vIL-10) and poxviral anaphase promoting complex regulator/ring H2 protein (PACR) (Fig. 3). This suggested that the right end of BPSV genomes were similar with that of other PPV genomes, except for the missing of VEGF existing in the left terminal region of BPSV genomes. It was also of note that PCPV genomes contained one extra VEGF genes in the left end and lacked the vIL-10 gene, while vIL-10 replaced VEGF in the left end of GQ329669.1/PCPV genome. Additionally, PPVs contained more than one ANK genes that may involve in host range.

Population dynamics of PPVs over the past 20 years

In order to explore the change of PPV strains circulating in the world during the past 20 years, 268 B2L gene sequences of PPVs were retrieved from GenBank in this study (Table S1). All B2L isolates were clustered into separate clades on the basis of their phylogenetic tree, suggesting that ORFV was the major species, which circulated mainly in China and India (Fig. S1). In recent years, most of these viruses were from a single species classified as ORFV (Fig. 4A). Consistent with the pattern observed in the phylogenetic tree, the results of case counts also indicated that most of ORFVs were isolated from China and India, and their prevalence increased sharply during 2013 – 2015 (Fig. 4B). Since 2010, ORFV has been the predominant species in circulation throughout the world.

ORFV of genera PPV became predominant species in recent years

To describe the phylogenetic evolution of the prevalent PPVs, the phylogenetic trees were constructed for the seven major structural segments. Phylogenetic trees were partitioned into groups with a 20th percentile cutoff [39]. The PPV strains were not

clustered exactly consistently in each tree of seven segments (Fig. 5). Within each prevalent clade, a large number of ORFVs further formed a main group with few PCPV, BPSV and PVNZ (Fig. 5), demonstrating they may share a common ancestor surviving a bottleneck event. Since 2010, the species diversity of PPVs was reduced with widespread outbreak in many countries (Fig. 4 and Fig. 5), which can be reasonably considered by high sequence similarity of PPVs isolated in the specific time span across broad regions.

In order to confirm this hypothesis, we estimated the temporal dynamic of gene sequence similarities of PPVs by collecting all available PPV sequences during 2000 – 2018. Pairwise identity comparison suggested a decreased diversity in all these seven segments of PPVs, starting mainly in 2010 – 2018 (Fig. 6). At nearly the same time point, the reduced sequence diversity in these major segments suggested that the virus with great fitness was selected, acquired dominance, and led to a widespread outbreak throughout the world. This is the first discovery, as we know, of such genetic bottleneck in the PPV genomes (Fig. 6), despite their prevalence in the world for the last 20 years.

Evolutionary rates and selection analysis of different gene sequences in PPVs

To understand the additional forces affecting PPV evolution, we analyzed the evolutionary rates and the average selective pressure of PPV strains whose isolation date were known. According to the different protein-coding gene sequences, a Bayesian coalescent approach was performed to infer the substitution rates of different segments. Based on the best-fit model, Bayesian calculated the mean evolutionary rates of the seven gene segments, ranging from 3.56×10^{-5} to 4.21×10^{-4} substitutions per site per year (Table 1). The VEGF gene evolved at the slowest rate among the structural segments, with the substitution rate of 3.56×10^{-5} and the highest probability density (HPD) between 4.31×10^{-6} to 9.61×10^{-5}. The VIR gene evolved at the highest estimated mean rate of 4.21×10^{-4} (95 % HPD: $1.28\times10^{-4} \sim 7.32\times10^{-4}$).

The selective pressure for each segment was performed with the phylogenetic analysis by maximum likelihood program (PAML) based on the Model 0 (M0) within codon-based phylogenetic models (CODEML). It supposed a constant substitution rate of all sites and provided average values across all sites. The GIF and VIR genes were

found to be more variable, while the VLTF-1 gene was more conserved. On the basis of Bayesian analysis, likelihood ratio tests were performed to compare the neutral models and the positive selection models. To estimate the selective pressure on each protein-coding segment, we applied a model (M3) for the differences among non-synonymous and synonymous rates (dN/dS) of amino acid residues, and several neutral models (M0, M1 and M7) with proportions of selected codons (M2 and M8) (Table 2). The log likelihood values showed that positive selection models (M2, M3 and M8) were more fitted in each gene regions than neutral models (M0, M1 and M7). The nested comparisons were also performed between the neutral models and the positive models, such as M0 vs M3, M1 vs M2 and M7 vs M8, and determined the better fitness of positive models, indicating that except for the VLTF-1, positive selection pressure occurred at different gene segments within the genome during the PPVs evolution process (chi-square test, $P < 0.05$) (Table 2).

4. Discussion

PPV infections normally occur in animals and even in humans throughout the world [11, 14, 16, 17, 40, 41]. However, comprehensive studies on PPVs are still lack. In the present study, available whole genome sequences and different structural protein-coding sequences were both utilized to investigate the structural and evolutionary genomics of PPVs. In order to identify the common features and differences of PPVs, it is important to analyze genomic patterns of all representative sequenced PPV genomes. To start with, the general nucleotide contents were calculated for the genomes of different PPV species have been shown in Table S2. The results suggested that the PPV coding sequences were GC-rich, especially on the third location of synonymous codons. This finding was consistent with previous researches wherein G and C frequencies were higher than A and T frequencies for alphaherpesviruses, betaherpesviruses, adenoviruses and Rubella virus (RUBV) [42, 43]. It has been known that viruses could take on a wider range of GC compositions than other organisms [44, 45]. Even within the same family, viruses have similar replication or life-cycle strategies, but may show different GC frequencies [46]. For instant, there were

significant differences of GC contents between ORFV and other three species of PPV, while viruses showed almost identical nucleotide bias (Fig. 1, Fig. 2 and Table S2). This supported the insight that there was common mechanism which determine GC compositions for virus with similar life cycles.

These PPVs, although related, could be divided into four subgroups according to the phylogenetic tree (Fig. 3), which contained ORFV, PCPV, BPSV and PVNZ, respectively. To extend the genomic description of these four subgroups of PPVs, we added the characterization of PPV genomes. PPV with large genomes (134~139 kb) encodes nearly complete replisomes [4, 5]. Analysis of the PPV genomes suggested conservation in four subgroups: genes encoding viral structural components (B2L, ANK and F1L), those participated in virion assembly (VEGF), nucleotide metabolism (PACR, CBP, GIF, vIL-10 and VIR), and DNA replication (VLTF). VEGF was known to play an important role in PPV pathogenesis related to vascularization process and epidermal lesion proliferation [6, 47]. Consistent with previous studies, we identified that VEGF was located in the left terminal region of BPSV contrasting the right terminal position of ORFV, PCPV and PVNZ VEGFs [6, 48], suggesting that the divergence of BPSV VEGF may have distinct functions or binding specificities compared with other VEGFs. Notably, PCPV was divergent from other PPVs with deficiency of vIL-10 or left terminal vIL-10. This feature may indicate specific adaptation to their natural hosts [6].

In addition to the genomic characteristics of PPVs, the phylogenetic diversity of segmented genome of PPV may be a crucial contributor to understand the temporal and spatial patterns of viral outbreaks [49, 50]. As shown in this study, the proportion of ORFV strains have undergone a sharply increase in the past 20 years. Our results suggested that PPVs evolved by mutation pressure and natural selection over 20 years of prevalence in animals, and the predominant species, ORFV, was selected recently. Since 2010, ORFV has rapidly increased in many countries, especially in China and India, and is now the predominant PPV species in farmed sheep [51-56]. It indicated that under the selective pressure of the optimal gene pattern by going through the selective bottleneck the ORFV eventually became the dominant species during the 2010

– 2018 outbreak. Combining with the reduced diversity of the 2010 – 2018 PPVs and the close phylogenetic clustering of their gene segments, these results demonstrated that the observed changes of viruses were common in the recent PPVs.

Furthermore, according to the supposition of rate constancy, the differences of viral isolation dates would provide the knowledge about the substitution rate of molecular evolution. The previous study has suggested that the magnitude of evolutionary change may accumulate since the isolation date [57]. A Bayesian coalescent approach suggested the mean substitution rates were between 3.56×10^{-5} to 4.21×10^{-4} in different gene segments, which was comparable with the studies about Newcastle disease virus, hepatitis A and B virus, and infectious bronchitis virus [27, 58, 59]. The highest substitution rate of PPV VIR gene reflected its less important role in viral synthesis and replication process than other gene segments, which may suffer from great natural selection pressure and subject to more substitutions.

When we analyzed the selection profiles of PPV protein-encoding genes, the overall rates of dN/dS were mostly less than 1, except for the GIF and VIR, demonstrating that purifying selection was the major force to drive the evolution of PPVs. However, there were clear differences in this selection profiles of various genes. Among these protein coding genes, GIF and VIR were under greater positive selection than other gene segments. Except for VLTF-1, other structural protein coding genes all had higher dN/dS values, highlighting that positive selection pressure has occurred in most segments of PPVs. Viral fitness could exactly constrain the variability of genes, which was important for virus replication [60]. Therefore, it was reasonable that VLTF-1 mainly suffered from purifying selection in the evolutionary process of PPVs. Additionally, as the angiogenic factor, the VEGF was essential to bind receptor tyrosine kinases and influence embryonic development or tumor neovascularization, note that low mutation rate during the viral evolution in host populations. These results indicated that this selection pressure may improve viral fitness in their infected hosts.

5. Conclusions

In conclusion, this study indicated that PPVs share several common features within different species. However, we noticed that GC contents were significantly different between ORFV and other PPV species, while they showed almost identical nucleotide bias. Besides, the genetic characteristics were also not similar among various PPV species, a phenomenon that may be due to their specific adaptation to natural hosts. We further investigated the evolutionary features of PPVs, and found that ORFV has been the predominant PPV species and rapidly increased in many countries since 2010. PPV isolates have been evolving at a sharply increased rate under the natural selection pressure. Therefore, combining genetic mutation of PPVs, positive selection may reshape the PPV genomes during the evolutionary process. The genetic analysis of PPVs could extend our knowledge of the mechanisms that promote viral evolution.

Author contributions: DKC and WTM conceived and designed experiments; XTY and MP performed all experiments. XTY, TXW and XC collected and analyzed the data. XTY, MP and XDT drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

Conflict of interest: The authors declare that no conflict of interest exists.

Funding information: This work was supported by National Natural Science Foundation of China (31902282), Basic Research Plan of Natural Science of Shaanxi Province (2020JQ-270), Key projects of Science and technology coordination in Shaanxi Province (2020ZDLNY02-06) and Science and Technology Project of Qinghai Agriculture and Pastoral Department (NMSY-2018-07).

REFERENCES

1. Becher P, Konig M, Muller G, Siebert U, Thiel HJ: Characterization of sealpox virus, a separate member of the parapoxviruses. *Arch Virol* 2002, 147(6):1133-1140.
2. Robinson AJ MA: Parapoxvirus of red deer: evidence for its inclusion as a new member in the genus parapoxvirus. *Virology* 1995 Apr 20, 208(2):812-815.
3. D B: The Encyclopedia of Virology plus. *Trends Biochem Sci* 1997 Jan, 22(1):34-35.
4. Wittek R KC, Wyler R: High C + G content in parapoxvirus DNA. *J Gen Virol* 1979 Apr, 43(1):231-234.
5. Büttner M RH: Parapoxviruses: from the lesion to the viral genome. *J Vet Med B Infect Dis Vet Public Health* 2002 Feb, 49(1):7-16.
6. Delhon G, Tulman ER, Afonso CL, Lu Z, de la Concha-Bermejillo A, Lehmkuhl HD, Piccone ME, Kutish GF, Rock DL: Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *J Virol* 2004, 78(1):168-177.
7. Olivero N, Reolon E, Arbiza J, Berois M: Genetic diversity of Orf virus isolated from sheep in Uruguay. *Arch Virol* 2018, 163(5):1285-1291.
8. Inoshima Y MA, Sentsui H: Detection and diagnosis of parapoxvirus by the polymerase chain reaction. *J Virol Methods* 2000 Feb, 84(2):201-208.
9. Hosamani M BV, Scagliarini A, Singh RK: Comparative sequence analysis of major envelope protein gene (B2L) of Indian orf viruses isolated from sheep and goats. *Vet Microbiol* 2006 Sep 10, 116(4):317-324.
10. Chan KW LJ, Lee SH, Liao CJ, Tsai MC, Hsu WL, Wong ML, Shih HC: Identification and phylogenetic analysis of orf virus from goats in Taiwan. *Virus Genes* 2007 Dec, 35(3):705-712.
11. Chakhunashvili G, Carlson BF, Power L, Khmaladze E, Tsaguria D, Gavashelidze M, Zakhashvili K, Imnadze P, Boulton ML: Parapoxvirus Infections in the Country of Georgia: A Case Series. *Am J Trop Med Hyg* 2018, 98(6):1870-1875.
12. Hosamani M SA, Bhanuprakash V, McInnes CJ, Singh RK: Orf: an update on current research and future perspectives. *Expert Rev Anti Infect Ther* 2009 Sep, 7(7):879-893.
13. de Sant'Ana FJ RR, Vulcani VA, Cargnelutti JF, Flores EF: Bovine papular stomatitis affecting dairy cows and milkers in midwestern Brazil. *J Vet Diagn Invest* 2012 Mar, 24(2):442-445.
14. Nougairede A FC, Salez N, Cohen-Bacrie S, Ninove L, Michel F, Aboukais S, Buttner M, Zandotti C, de Lamballerie X, Charrel RN: Sheep-to-human transmission of Orf virus during Eid al-Adha religious practices, France. *Emerg Infect Dis* 2013 Jan, 19(1):102-105.
15. Kitchen M MH, Zobl A, Windisch A, Romani N, Huemer H: ORF virus infection in a hunter in Western Austria, presumably transmitted by game. *Acta Derm Venereol* 2014 Mar, 94(2):212-214.
16. Osadebe LU MK, McCollum AM, Li Y, Emerson GL, Gallardo-Romero NF, Doty JB, Wilkins K, Zhao H, Drew CP, Metcalfe MG, Goldsmith CS, Muehlenbachs A, Googe PB, Dunn J, Duenckel T, Henderson H, Carroll DS, Zaki SR, Denison MR, Reynolds MG, Damon IK: Novel poxvirus infection in 2 patients from the United States. *Clin Infect Dis* 2015 Jan 15, 60(2):195-202.
17. Midilli K EA, Kuşkucu M, Analay H, Erkiliç S, Benzonana N, Yildirim MS, Mülayim K, Acar H, Ergonul O: Nosocomial outbreak of disseminated orf infection in a burn unit, Gaziantep, Turkey, October to December 2012. *Euro Surveill* 2013 Mar 14, 18(11):20425.

18. Tedla M, Berhan N, Molla W, Temesgen W, Alemu S: Molecular identification and investigations of contagious ecthyma (Orf virus) in small ruminants, North west Ethiopia. *BMC Vet Res* 2018, 14(1):13.
19. Mazur C MR: Detection of contagious pustular dermatitis virus of goats in a severe outbreak. *Vet Rec* 1989 Oct 14, 125(16):419-420.
20. McKeever DJ JD, Hutchison G, Reid HW: Studies of the pathogenesis of orf virus infection in sheep. *Journal of comparative pathology* 1988 Oct, 99(3):317-328.
21. de Sant'Ana FJ LF, Rabelo RE, Vulcani VA, Moreira CA Jr, Cargnelutti JF, Flores EF: Coinfection by Vaccinia virus and an Orf virus-like parapoxvirus in an outbreak of vesicular disease in dairy cows in midwestern Brazil. *J Vet Diagn Invest* 2013 Mar, 25(2):267-272.
22. Schmidt C CJ, Brum MC, Traesel CK, Weiblen R, Flores EF: Partial sequence analysis of B2L gene of Brazilian orf viruses from sheep and goats. *Vet Microbiol* 2013 Feb 22, 162(1):245-253.
23. Abrahão JS BI, Mazur C, Lobato ZI, Ferreira PC, Bonjardim CA, Trindade GS, Kroon EG: Looking back: a genetic retrospective study of Brazilian Orf virus isolates. *Vet Rec* 2012 Nov 10, 171(19):476.
24. Cargnelutti JF FM, Teixeira FR, Weiblen R, Flores EF: An outbreak of pseudocowpox in fattening calves in southern Brazil. *J Vet Diagn Invest* 2012 Mar, 24(2):437-441.
25. de Oliveira CH AF, Neto JD, Oliveira CM, Lopes CT, Bomjardim Hdos A, Vinhote WM, Silva AG, Abrahão JS, Kroon EG: Multifocal cutaneous ORF virus infection in goats in the Amazon region, Brazil. *Vector Borne Zoonotic Dis* 2012 Apr, 12(4):336-340.
26. Petrova VN, Russell CA: The evolution of seasonal influenza viruses. *Nat Rev Microbiol* 2018, 16(1):47-60.
27. Zhao Y, Zhang H, Zhao J, Zhong Q, Jin JH, Zhang GZ: Evolution of infectious bronchitis virus in China over the past two decades. *J Gen Virol* 2016, 97(7):1566-1574.
28. Matsumoto H, Setoyama H, Matsuura Y, Ohtani A, Shimizu K, Okada A, Inoshima Y: Sequential detection of pseudocowpox virus and bovine papular stomatitis virus in a same calf in Japan. *J Vet Med Sci* 2019, 81(3):440-443.
29. RC E: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004 Mar 19, 32(5):1792-1797.
30. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999, 41:95-98.
31. Letunic I BP: Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 2011 Jul, 39(Web Server issue):W475-478.
32. Ito K MD: Application of ggplot2 to Pharmacometric Graphics. *CPT Pharmacometrics Syst Pharmacol* 2013 Oct 16, 2:e79.
33. Darriba D, Taboada GL, Doallo R, Posada D: jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012, 9(8):772.
34. Newton MAR, A. E: Approximate bayesian inference with the weighted likelihood bootstrap. *J Roy Statist Soc Ser B* 1994, 56:3–48.
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28(10):2731-2739.

36. Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ: BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2014, 10(4):e1003537.
37. Shapiro B, Rambaut A, Drummond AJ: Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* 2006, 23(1):7-9.
38. Rambaut A, Lam TT, Max Carvalho L, Pybus OG: Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2016, 2(1):vew007.
39. Prosperi MC CM, Fanti I, Saladini F, Pecorari M, Borghi V, Di Giambenedetto S, Bruzzone B, Capetti A, Vivarelli A, Rusconi S, Re MC, Gismondo MR, Sighinolfi L, Gray RR, Salemi M, Zazzi M, De Luca A; ARCA collaborative group: A novel methodology for large-scale phylogeny partition. *Nat Commun* 2011, 2:321.
40. de Sant'Ana FJ RR, Vulcani VA, Cargnelutti JF, Flores EF: Bovine papular stomatitis affecting dairy cows and milkers in midwestern Brazil. *J Vet Diagn Invest* 2012, 24(2):442-445.
41. Kitchen M MH, Zobl A, Windisch A, Romani N, Huemer H: ORF virus infection in a hunter in Western Austria, presumably transmitted by game. *Acta Derm Venereol* 2014, 94(2):212-214.
42. Shackelton LA, Parrish CR, Holmes EC: Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol* 2006, 62(5):551-563.
43. Zhou Y CX, Ushijima H, Frey TK: Analysis of base and codon usage by rubella virus. *Arch Virol* 2012 May, 157(5):889-899.
44. Bronson EC AJ: Nucleotide composition as a driving force in the evolution of retroviruses. *J Mol Evol* 1994 May, 38(5):506-532.
45. Shackelton LA PC, Holmes EC: Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol* 2006 May, 62(5):551-563.
46. Schachtel GA BP, Mocarski ES, Blaisdell BE, Karlin S: Evidence for selective evolution in codon usage in conserved amino acid segments of human alphaherpesvirus proteins. *J Mol Evol* 1991 Dec, 33(6):483-494.
47. Savory LJ SS, Fleming SB, Niven BE, Mercer AA: Viral vascular endothelial growth factor plays a critical role in orf virus infection. *J Virol* 2000 Nov, 74(22):10699-10706.
48. Rziha HJ BB, Adam KH, Röttgen M, Cottone R, Henkel M, Dehio C, Büttner M: Relatedness and heterogeneity at the near-terminal end of the genome of a parapoxvirus bovis 1 strain (B177) compared with parapoxvirus ovis (Orf virus). *J Gen Virol* 2003 May, 84(Pt 5):1111-1116.
49. Perrin A, Larsonneur E, Nicholson AC, Edwards DJ, Gundlach KM, Whitney AM, Gulvik CA, Bell ME, Rendueles O, Cury J *et al*: Evolutionary dynamics and genomic features of the Elizabethkingia anophelis 2015 to 2016 Wisconsin outbreak strain. *Nat Commun* 2017, 8:15483.
50. Pu J, Wang S, Yin Y, Zhang G, Carter RA, Wang J, Xu G, Sun H, Wang M, Wen C *et al*: Evolution of the H9N2 influenza genotype that facilitated the genesis of the novel H7N9 virus. *Proc Natl Acad Sci U S A* 2015, 112(2):548-553.
51. Ahanger SA PR, Nazki S, Dar Z, Dar T, Dar KH, Dar A, Rai N, Dar P: Detection and phylogenetic analysis of Orf virus in Kashmir Himalayas. *Virusdisease* 2018 Sep, 29(3):405-410.

52. Bala JA BK, Abdullah AA, Mohamed R, Haron AW, Jesse FFA, Noordin MM, Mohd-Azmi ML: The re-emerging of orf virus infection: A call for surveillance, vaccination and effective control measures. *Microb Pathog* 2018 Jul, 120:55-63.
53. Bora M BD, Barman NN, Borah B, Bora PL, Talukdar A, Tamuly S: Isolation and molecular characterization of Orf virus from natural outbreaks in goats of Assam. *Virusdisease* 2015 Jun, 26(1-2):82-88.
54. Chan K W LJW, Lee S H, Liao C J, Tsai M C, Hsu W L, Wong M L, Shih H C: Identification and phylogenetic analysis of orf virus from goats in Taiwan. *Virus Genes* 2007, 35 (3):705-712.
55. Flores C GE, Verna A, Peralta A, Madariaga C, Odeón A, Cantón G: Orf virus in human, confirmation in case report from Chile. *Rev Chilena Infectol* 2017 Dec, 34(6):607-609.
56. M Ş: Association of two clusters of Orf virus isolates in outbreaks of infection in goat in the Central Anatolian region of Turkey. *Virusdisease* 2017 Sep, 28(3):345-348.
57. Soylemez O, Kondrashov FA: Estimating the rate of irreversibility in protein evolution. *Genome Biol Evol* 2012, 4(12):1213-1222.
58. Wang H WX, Zheng HH, Cao JY, Zhou WT, Bi SL: Evolution and genetic characterization of hepatitis A virus isolates in China. *Int J Infect Dis* 2015 Apr, 33:156-158.
59. Alvarado Mora MV RC, Gomes-Gouvêa MS, Gutierrez MF, Botelho L, Carrilho FJ, Pinho JR: Molecular characterization of the Hepatitis B virus genotypes in Colombia: a Bayesian inference on the genotype F. *Infect Genet Evol* 2011 Jan, 11(1):103-108.
60. Soderholm J, Ahlen G, Kaul A, Frelin L, Alheim M, Barnfield C, Liljestrom P, Weiland O, Milich DR, Bartenschlager R *et al*: Relation between viral fitness and immune escape within the hepatitis C virus protease. *Gut* 2006, 55(2):266-274.

Fig. 1. Nucleotide content distribution and composition. (A) The mean frequency for A, T, G and C composition in 17 different PPV sequences are shown, suggests higher GC content than AT on the whole. (B) The mean values of the nucleotide content frequency at the 3rd codon position. (C) Based on 4 various species of PPVs, analysis for the overall A, T, G and C composition frequencies. (D) Based on 4 various species of PPVs, analysis for A, T, G and C composition frequencies at the third codon position. Different asterisks represent statistical significance (*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

Fig. 2. Nucleotide composition of PPVs for different species. Frequency of each observed nucleotide composition, reconstructed in BioEdit (v7.2.5), is shown for different species of PPVs.

Fig. 3. Phylogenetic analysis and genome comparison among different species of PPVs. ML phylogenetic tree was reconstructed in phyML (v3.0) and tested with 1000 bootstraps. The predicted genes are represented by block arrows showing the direction of transcription. The genes encoding

for structural proteins and replication-associated proteins of PPVs are shown in different colors.

Fig. 4. Population dynamics of PPVs during the past 20 years. Case counts are presented in the plot. The X-axis represents the year intervals, and the Y-axis represents the number of cases. Different colors represent different species (A) and different collection countries of PPVs (B).

Fig. 5. Phylogenetic analysis of seven structural genes with PPVs during 2000 – 2018. Clades with all of 2000 – 2018 viruses were fully shown. Color of line at right of each leaf node indicates year of isolation (see color bar). Timescale is in years. Vertical lines mark different species of PPVs: gray represents ORFV, purple represents BPSV, black represents PCPV and yellow represents PVNZ.

Fig. 6. Decreased genetic diversity of the seven segments of PPVs during the widespread outbreaks in the world (2000 – 2018). Pairwise comparison of nucleotide sequences of PPVs was plotted as a heatmap. Viruses isolated from 2000 through 2018 were ordered by isolation date from left to right of the x axis and from top to bottom of the y axis. On the axes, the ticks indicated the isolation years. Black dotted line represented the year 2010. Color indicated identity levels from ≤ 30% (blue) to 100% (red).

Fig. S1. The phylogenetic tree of PPV B2L genes. The tree was generated by the maximum likelihood (ML) method using PAML v4.9. The tree was designed by using the online tool "iTOL". Here; different colors represent different collection locations. The color of each leaf node represents different virus species.

Figures

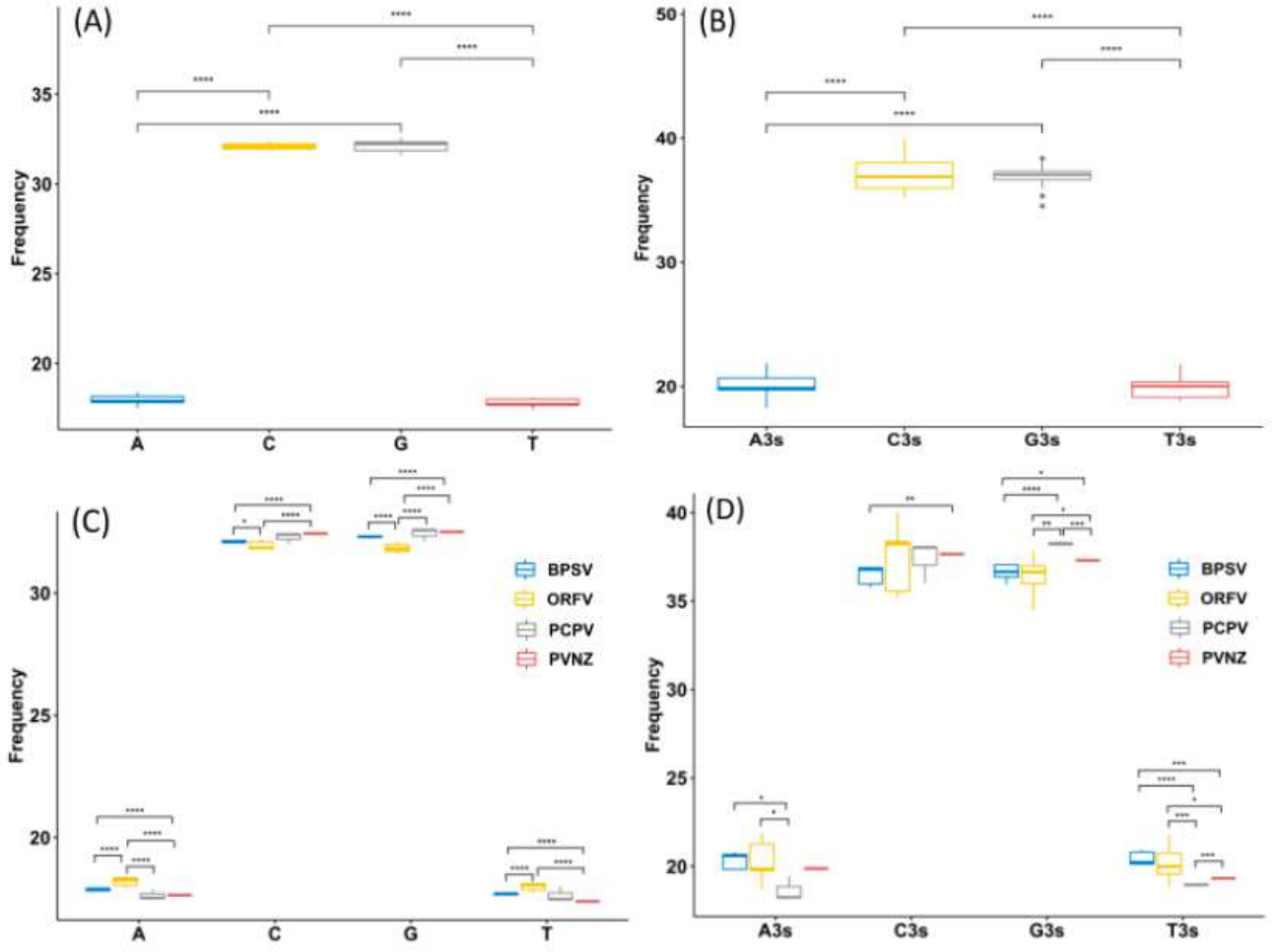

Figure 1

Nucleotide content distribution and composition. (A) The mean frequency for A, T, G and C composition in 17 different PPV sequences are shown, suggests higher GC content than AT on the whole. (B) The mean values of the nucleotide content frequency at the 3rd codon position. (C) Based on 4 various species of PPVs, analysis for the overall A, T, G and C composition frequencies. (D) Based on 4 various species of PPVs, analysis for A, T, G and C composition frequencies at the third codon position. Different asterisks represent statistical significance (*p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001).

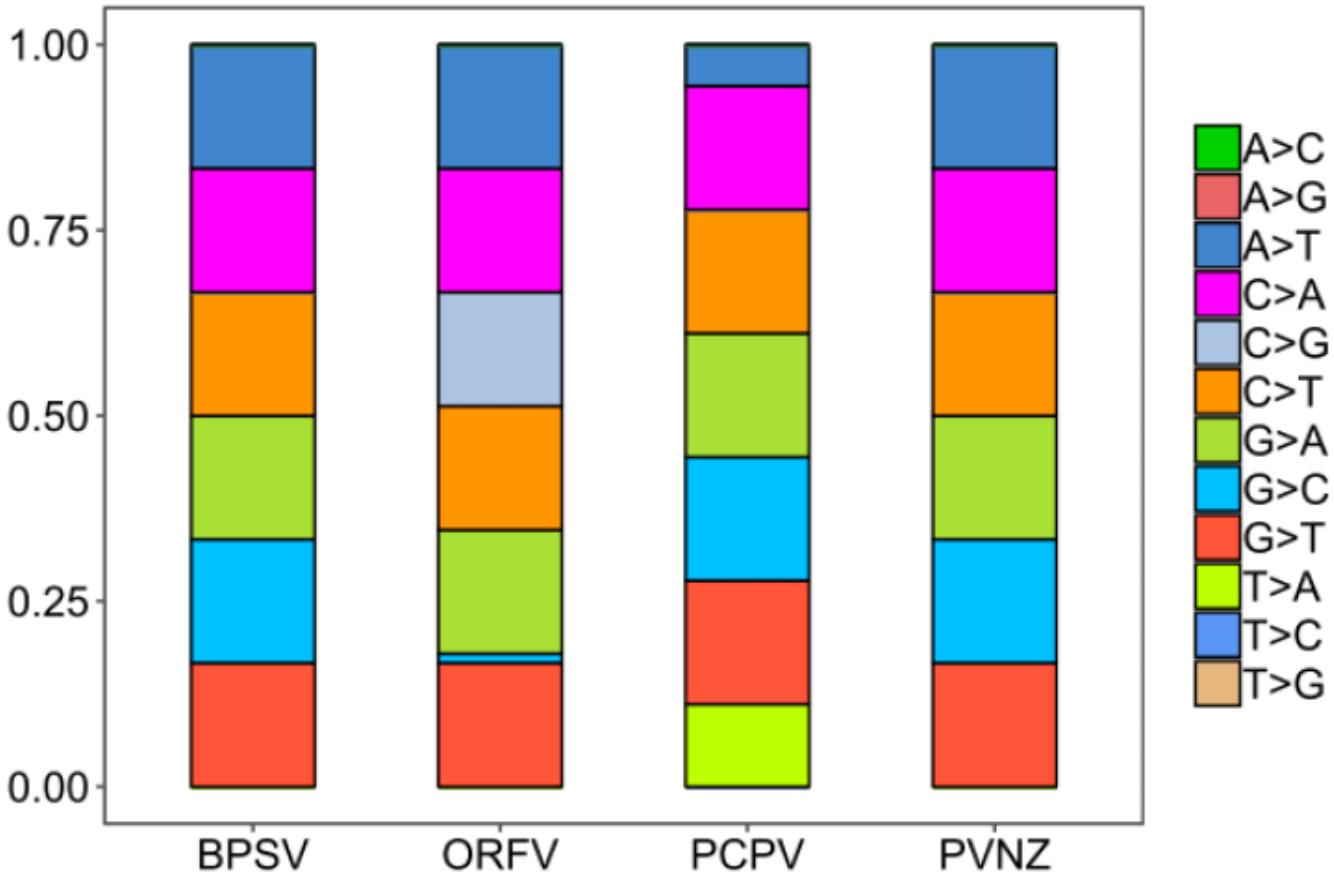

Figure 2

Nucleotide composition of PPVs for different species. Frequency of each observed nucleotide composition, reconstructed in BioEdit (v7.2.5), is shown for different species of PPVs.

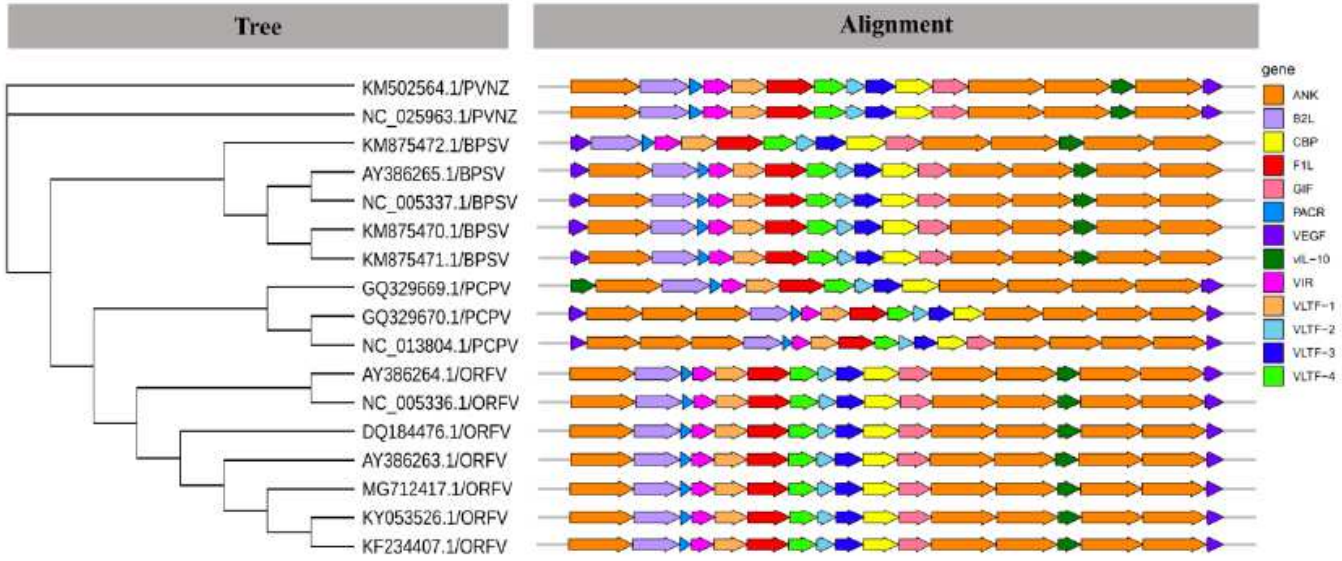

Figure 3

Phylogenetic analysis and genome comparison among different species of PPVs. ML phylogenetic tree was reconstructed in phyML (v3.0) and tested with 1000 bootstraps. The predicted genes are represented by block arrows showing the direction of transcription. The genes encoding for structural proteins and replication-associated proteins of PPVs are shown in different colors.

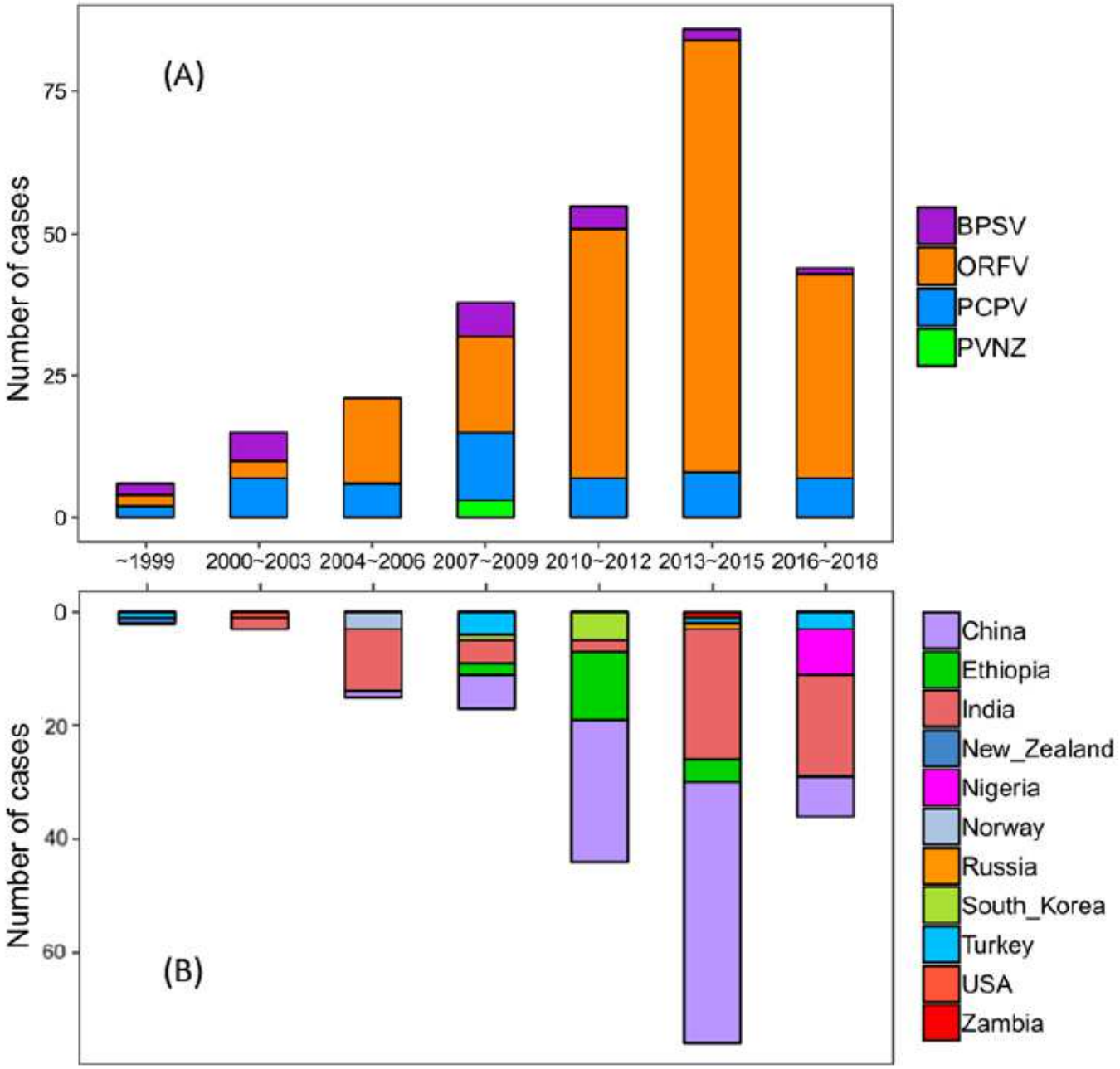

Figure 4

Population dynamics of PPVs during the past 20 years. Case counts are presented in the plot. The X-axis represents the year intervals, and the Y-axis represents the number of cases. Different colors represent different species (A) and different collection countries of PPVs (B).

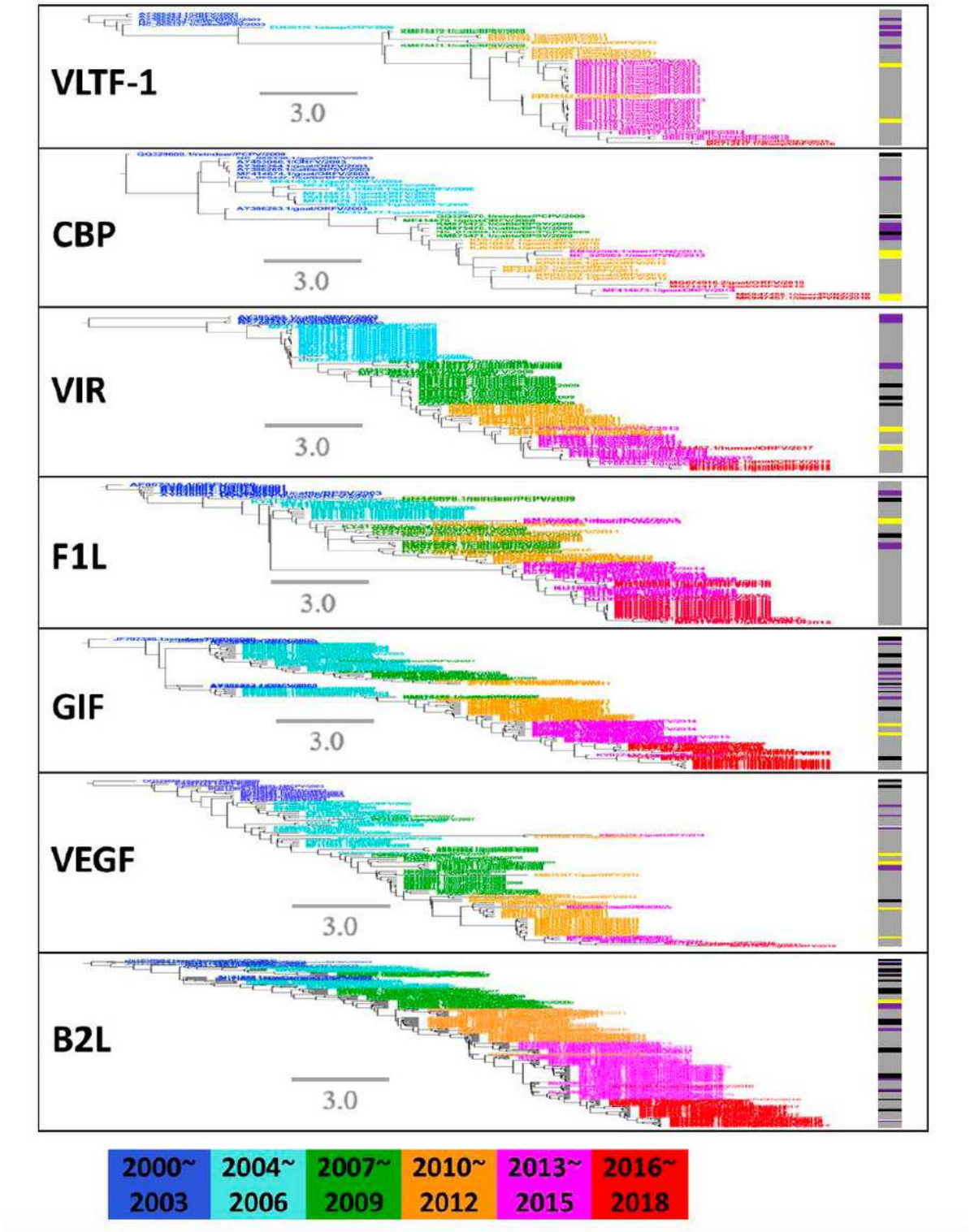

Figure 5

Phylogenetic analysis of seven structural genes with PPVs during 2000 – 2018. Clades with all of 2000 – 2018 viruses were fully shown. Color of line at right of each leaf node indicates year of isolation (see

color bar). Timescale is in years. Vertical lines mark different species of PPVs: gray represents ORFV, purple represents BPSV, black represents PCPV and yellow represents PVNZ.

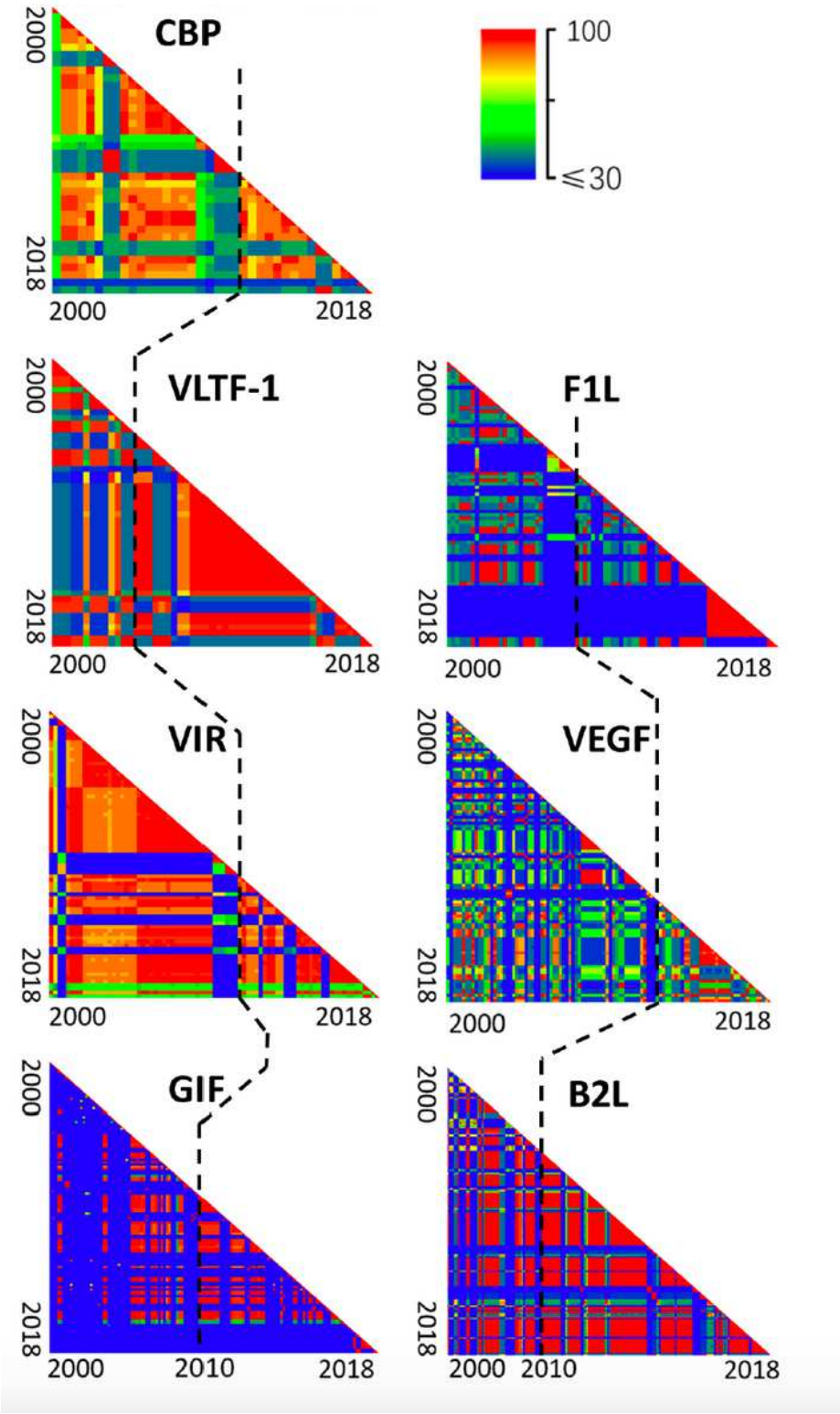

Figure 6

Decreased genetic diversity of the seven segments of PPVs during the widespread outbreaks in the world (2000 – 2018). Pairwise comparison of nucleotide sequences of PPVs was plotted as a heatmap. Viruses isolated from 2000 through 2018 were ordered by isolation date from left to right of the x axis and from

top to bottom of the y axis. On the axes, the ticks indicated the isolation years. Black dotted line represented the year 2010. Color indicated identity levels from ≤ 30% (blue) to 100% (red).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfigure.pdf
- TableS1.xlsx
- supplementarytable2.pdf