Identification of a novel ephemerovirus in a water buffalo (*Bubalus bubalis* [Linnaeus, 1758])

Sakiho Imai[1], Mai Kishimoto[1], Masayuki Horie[1,2,3 *]

[1] Laboratory of Veterinary Microbiology, Graduate School of Veterinary Medicine, Osaka Metropolitan University

[2] Osaka International Infectious Diseases Research Center, Osaka Metropolitan University

*Corresponding author

Masayuki Horie, DVM, PhD

Email: mhorie@omu.ac.jp

Abstract

Ephemeroviruses, belonging to the genus *Ephemerovirus* within the family *Rhabdoviridae* of the *Mononegavirales*, are non-segmented, negative-strand RNA viruses that infect artiodactyls and blood-sucking arthropods. Although recent advances in sequencing technology have facilitated the identification of novel ephemeroviruses, thereby expanding our understanding of this viral genus, their diversity remains elusive, as evidenced by phylogenetic gaps between currently known ephemeroviruses. In this study, we analyzed publicly available RNA-seq data and identified a novel ephemerovirus, tentatively named Punjab virus (PBV), in a water buffalo (*Bubalus bubalis* [Linnaeus, 1758]). We obtained two separate PBV contigs from the RNA-seq data; the first contig covers the N, P, and M genes, while the second contig covers the G, α, β, γ, and L genes. Together, these PBV contigs represent 99% of the estimated complete viral genome. Mapping analysis revealed a typical transcriptional gradient pattern commonly observed in mononegaviruses, suggesting that the water buffalo is the authentic host for PBV. Sequence comparisons with its closest relatives indicate that the newly identified virus meets the ICTV species demarcation criteria for sequence divergence. Thus, this study contributes to a deeper understanding of the diversity of ephemeroviruses.

Keywords

Bubalus bubalis, ephemerovirus, water buffalo

Introduction

The diversity of viruses remains largely unknown, with the viruses we know today representing only a small portion of the entire virosphere [10]. Understanding the diversity of viruses is important from various perspectives, such as the control of infectious diseases, the elucidation of viral evolution, and the development of virotherapy. Identification of novel viruses can help with pandemic preparedness, as emerging infectious diseases can be caused by previously unidentified viruses [17]. Additionally, many phylogenetic gaps are present between known viruses [26], which suggests the existence of undiscovered viruses filling the gaps. Furthermore, viruses can be utilized for various treatments and vaccines. Consequently, understanding the diversity of viruses is crucial from a variety of perspectives.

Ephemeroviruses, members of the genus *Ephemerovirus* within the family *Rhabdoviridae* of the order *Mononegavirales*, possess single-stranded negative-strand RNA as their genomes [30]. These genomes encode structural proteins N, P, M, G, and L, as well as several nonstructural proteins that are encoded between the G and L genes. The genes encoding structural proteins are conserved among all the known ephemeroviruses, whereas the genes encoding nonstructural proteins may vary among the viruses.

Ephemeroviruses have been detected in artiodactyls and blood-sucking arthropods [30, 31]. Bovine ephemeral fever virus (BEFV), which is the type virus of the genus *Ephemerovirus*, is an arbovirus proven to cause a disease called BEF in certain ruminants, such as cattle [31]. Animals affected by BEF exhibit symptoms like acute fever, arthralgia, and dropping milk production, which result in economic loss. On the other hand, the pathogenicity of other ephemeroviruses remains unclear, although some ephemeroviruses have been reported in animals showing symptoms [3].

A diverse range of ephemeroviruses have been detected through various methods to date, and recent advancements in metagenomic analyses have expedited these discoveries. However, there still remain large phylogenetic gaps among ephemeroviruses [30], suggesting the presence of undiscovered ephemeroviruses that fill the gaps. Therefore, further exploration of ephemeroviruses is crucial for a more comprehensive understanding of the diversity and potential pathogenicity of ephemeroviruses.

In this study, we investigated the diversity of ephemeroviruses by analyzing publicly available RNA-seq data. We identified novel ephemerovirus sequences, tentatively named Punjab virus (PBV), in RNA-seq data obtained from a water buffalo (*Bubalus bubalis* [Linnaeus, 1758]). Our detailed sequence analysis revealed that the viral genome sequence is divided into two contigs, estimated to cover approximately 99% of the complete genome. Phylogenetic analysis showed that this novel ephemerovirus forms a cluster with Kokolu virus, Puchong virus, and Hayes Yard virus. Moreover, PBV meets the ICTV species demarcation criteria for sequence divergence, suggesting that this virus can be classified as a new species within the genus *Ephemerovirus*.

Material Method

Detection of ephemerovirus-like contigs

RNA-seq data (accession number SRR8476835) were downloaded from the NCBI SRA [24], which were preprocessed by fastp 0.23.2 [6] using the "-l 35 -x -y" options. The preprocessed reads were then assembled by Trinity 2.14.0 [12], SKESA 2.5.1 [27], SPAdes v3.15.5 [23], or metaSPAde v3.15.5 [22] using the default setting. Contigs obtained by Trinity that were 100 nucleotides or more were extracted using SeqKit 2.3.0 [25], and then clustered using CD-HIT v4.8.1 [9] with a threshold of 0.98. The clustered contigs were used for a two-step sequence similarity search as follows.

In the first step, a sequence similarity search was conducted against a custom database containing protein sequences of viruses belonging to the kingdom *Orthornavirae* by MMseqs2 version c48da9d781b81804727b5cccfed7f97cfcc20c9d [28] using the clustered contigs as queries. From the hit contigs, those with *E*-values of less than 10^{-20} and whose top hits (hit sequence with the highest score) were viruses were extracted. Among the extracted contigs, only one representative sequence was used in the subsequent analysis for a group of sequences considered to be isoforms based on the contig IDs.

The second sequence similarity search was performed against the NCBI nr database [24] by BLASTx 2.13.0 [4] with the options "-evalue 1e-20 -max_target_seqs 10 -word_size 2 -lcase_masking" using the extracted contigs as queries. The contigs whose BLAST best hits were viruses were extracted and used in subsequent analyses as virus-like contigs.

Validation of the virus-like contigs by mapping analysis

To validate the accuracy of virus-like contigs, a mapping analysis was performed. The original RNA-seq data (accession number SRR8476835) were mapped to the obtained PBV contigs by HISAT2 2.2.1 [18], and the read depth at each position was calculated using SAMtools 1.16.1 [7]. The positions covered by five or more reads were considered as reliable positions.

Annotation of the virus-like contigs

Open reading frames (ORFs) consisting of more than 256 nucleotides (based on the lengths of ephemeroviral ORFs) were identified in the virus-like contigs using Geneious Prime (https://www.geneious.com). ORFs that spanned transcription signals (see below) were manually corrected. BLASTp searches were conducted against the protein sequences of viruses (taxid:10239) in the NCBI nr database on the BLAST web server using translated sequences of ORFs as queries. The following options were used: Word size: 3; Expect threshold: 10^{-10}.

To identify transcription signals, conserved motifs were searched by MEME 5.5.0 [1] with the options "-mod oops -maxw 10 -nmotifs 3 -dna" using the sequences of intergenic regions. Each identified motif sequence with its flanking 4 nucleotides was extracted, which were aligned by MAFFT v7.490 using the E-INS-i algorithm [15]. Putative transcription signals were determined based on the alignments.

Signal peptide prediction was performed using the SignalP 6.0 web server [29].

Phylogenetic analyses

Phylogenetic trees were inferred using the putative amino acid sequences of N, G, or L proteins of PBV, known members of the genus *Ephemerovirus*, and three other rhabdoviruses (outgroups) (Table S1). The sequences were aligned by MAFFT v7.490 using the E-INS-i algorithm, and the ambiguously aligned regions were trimmed by Trimal v1.4.rev22 with the “-strict” option [5]. Phylogenetic trees were reconstructed by the maximum likelihood method using RAxML Next

Generation 1.1.0 [20]. LG+I+G4, WAG+I+G4+FC, and LG+I+G4+FC models, chosen by ModelTest-NG v0.2.0 [8], were used for the inference of N, G, and L trees, respectively.

Determination of pairwise sequence identities

Amino acid sequences of N, G, or L proteins of PBV and closely related ephemeroviruses (Table S1) were aligned by MAFFT, and then pairwise sequence identities were determined using Sequence Demarcation Tool version 1.2 [21].

Mapping analysis to detect Punjab virus infection

To detect PBV infection, a total of 46,244 publicly available RNA-seq data (accession numbers are available in Supplementary Materials) were downloaded and preprocessed by fastp 0.23.2 with the options "-x -y -l 35". The preprocessed reads were then mapped to the PBV contigs using HISAT2 2.2.1. The numbers of mapped reads were counted using SAMtools. The mapped reads were also manually analyzed to check the accuracy of the mapping.

Results

Identification of a novel ephemerovirus

In our previous study, we performed a large-scale metaviromic analysis and detected many RNA viruses from publicly available RNA-seq data [16]. However, detailed analyses were conducted only for the viral sequences that were close to the full-length genomes. Consequently, many of the detected partial viral sequences have not yet been analyzed well. Therefore, we reanalyzed the BLAST results obtained in the previous study and found that one of the RNA-seq data sets (accession number SRR8476835) obtained from the blood of a water buffalo (*B. bubalis*) [13] contains ephemerovirus-like sequences. To confirm this result, we again performed *de novo* assembly and a two-step sequence similarity search using the resultant contigs. Consistent with the previous result, we detected two ephemerovirus-like contigs whose respective BLAST best hits were Hayes Yard virus N protein (QEA08650.1; 90.3% identity) and Puchong virus L protein (QEA08648.1; 78.4% identity) (Table 1).

To validate the accuracy of obtained ephemerovirus-like contigs, we mapped the original short reads to the contigs and measured the read depths. In this study, we defined positions mapped by five or more reads as “reliable” regions. As a result, we removed some of the extreme terminal sequences of the obtained contigs, resulting in contigs with lengths of 3007 (Contig 1) and 11819 (Contig 2) nucleotides. It is important to note that the mapping pattern showed a typical transcription gradient observed in mononegaviruses, further supporting the assertion that the contigs are derived from an ephemerovirus (Fig. 1).

Characterization of the ephemerovirus-like contigs

To determine the genomic structure of PBV, we extracted ORFs from the contigs and performed BLASTp searches using each of the ORFs as a query. As a result, we identified three and seven ORFs in Contig 1 and Contig 2, showing sequence similarities to N, P, and M and G, Gns, α1, α2, β, γ, L genes of other ephemeroviruses, respectively (Table S2). Because it was initially unclear whether the annotated G gene is full length or not due to its location at the end of contig (Fig. 1b), we characterized the putative G protein *in silico*. The putative G protein sequence was predicted to contain a signal peptide at the N-terminus (Fig. S1a). Furthermore, the putative G protein was alignable with the full-length G proteins of related viruses (Fig. S1b). These results strongly suggest that the annotated G gene is full-length. On the other hand, the stop codon of L gene was not included in the contig (Fig. 1b).

We subsequently performed MEME searches to identify putative transcription signal sequences in the intergenic regions. These searches, in combination with manual curation, identified putative transcription initiation signals 5’-AACAGG-3’ and termination/polyadenylation signals 5’-ATGAAAAAAA-3’ (Fig. 1c).

Phylogenetic analysis

To understand the evolutionary relationships between PBV and other ephemeroviruses, we conducted phylogenetic analyses using the amino acid sequences of the N, G, and L proteins (Figs. 2 and S2). All the trees show that PBV forms a well-supported cluster with Kokolu virus, Puchong virus, and

Hayes Yard virus, and diverged earlier than these three viruses. Additionally, the clade containing PBV and the aforementioned three viruses is closely related to the clade containing BEFV.

Amino acid sequence divergence between Punjab virus and the closely related viruses

To investigate the amino acid divergence between PBV and the closely related viruses, we determined pairwise identities using SDT with the amino acid sequences of N, G, and L proteins from closely related viruses listed in Table S1. The maximum amino acid identities were 91.2% for the N protein (Puchong virus), 64.5% for the G protein (Kokolu virus), and 77.8% for the L protein (Berrimah virus), respectively (Fig. 3).

Mapping analysis to detect infection from other public RNA-seq data

To gain more insight into PBV infection, we searched for PBV-like sequences in public RNA-seq data by mapping analysis. Given that some ephemeroviruses are known to be arboviruses, we mapped short reads from publicly available RNA-seq data of ticks (subclass Acari), mosquitos (family Culicidae), biting midges (family Ceratopogonidae), and bovines (subfamily Bovinae) to PBV contigs, and then counted the number of mapped reads. We detected a small amount of mapped reads from three RNA-seq data sets belonging to the same BioProject from which we originally detected the viral contigs (Table S3). However, we cannot rule out the possibility that these were due to cross-contamination and/or index hopping, and therefore it is unclear whether these samples really contained the virus.

Discussion

To date, 13 species of viruses have been identified in the genus *Ephemerovirus*. However, the divergence of ephemeroviruses remains unclear, as suggested by the presence of phylogenetic gaps [30]. In this study, we identified a novel ephemerovirus in publicly available RNA-seq data obtained from *B. bubalis*. A series of analyses showed that the identified viral sequences possess a typical ephemerovirus genome structure and also exhibit a characteristic transcription pattern of mononegaviruses. Importantly, our analyses demonstrated that the PBV exhibits an amino acid sequence divergence of 8.8%, 35.5%, and 35.2% from the most closely related viral N, G, and L proteins, respectively (Fig. 3). This fulfills the current species demarcation criteria of the genus *Ephemerovirus* in terms of sequence divergence. Although we were unable to obtain a single contig of this virus, and the L protein lacks its C-terminal sequence, our data suggest that the PBV can be classified as a new species within the genus *Ephemerovirus*.

The pathogenicity of PBV remains uncertain. The RNA-seq data (SRR8476835), used for this study was sourced from a blood sample collected from a water buffalo affected by metritis. In the same BioProject (PRJNA514883), there exists additional RNA-seq data (SRR8476836) obtained from another individual also affected by metritis. However, only a few viral reads were detected in this second individual (SRR8476836), creating ambiguity regarding whether the water buffalo was indeed infected by PBV, especially considering cross-contamination and index hopping [19, 32]. Hayes Yard virus, one of the closely related ephemeroviruses, was isolated from a bull (*Bos indicus* [Linnaeus, 1758]) afflicted with a severe ephemeral fever-like illness, but it remains inconclusive whether this virus was the causative agent [3]. Furthermore, while preparing this manuscript, we noted that another study identified a novel ephemerovirus, which can be classified into the same species as PBV, in a febrile cow (Figs. S3) [11]. Further epidemiological studies are essential to improve our understanding of the pathogenicity of ephemeroviruses, including PBV.

The mapping pattern provides strong evidence that *B. bubalis* is a legitimate host for PBV. In viral metagenomic analysis, host identification can sometimes be challenging because samples may contain nucleic acids from viruses of various environmental and dietary origins. Our mapping analysis showed the typical transcription gradient from N to L genes observed in mononegaviruses, implying that PBV was actively transcribing in the samples. Since the RNA was extracted from blood, the likelihood of contamination is minimal. Moreover, the genetically related viruses were also detected from bovines [2, 3, 14]. Considering these points, *B. bubalis* would be an authentic host for PBV. It should be noted that some ephemeroviruses are known to be arboviruses. As PBV was detected in blood samples, it is plausible that the virus could be transmitted by arthropod vectors. Further studies are required to elucidate the transmission route of PBV.

In this study, we only obtained two separate contigs of the PBV genome, but not a single one. Besides using Trinity, we performed *de novo* assembly with several assemblers (SKESA, SPAdes, metaSPAdes), yet we consistently obtained two separate contigs (data not shown). This is likely because the mRNA-seq does not adequately cover the intergenic regions (Fig. 1). Unfortunately, only a few viral reads were detected from RNA-seq data other than the initially detected one (Table S3), making co-assembly unavailable. Further accumulation of data or in-depth molecular epidemiological studies are required to determine the complete genome of PBV.

Together, we identified a novel ephemerovirus from public RNA-seq data, thereby contributing to a deeper understanding of the diversity of ephemeroviruses. However, the virological characteristics of PBV, such as its pathogenicity and infection route, remain unclear. Further identification of infected individuals and the accumulation of sequence information would contribute to the characterization of PBV.

Conflict of interest

The authors declare no conflict of interest in this study.

Acknowledgments

The supercomputing resources were provided by the Human Genome Center, the Institute of Medical Science, the University of Tokyo, and the NIG supercomputer at ROIS National Institute of Genetics. This study was supported by JSPS KAKENHI grant number JP21H01199 (MH) and MEXT KAKENHI grant number JP19H04833 (MH).

References

1. Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. 2015. The MEME Suite. *Nucleic Acids Res.* **43**: W39-49.
2. Balinandi, S., Hayer, J., Cholleti, H., Wille, M., Lutwama, J. J., Malmberg, M. and Mugisha, L. 2022. Identification and molecular characterization of highly divergent RNA viruses in cattle, Uganda. *Virus Res.* **313**: 198739.

3. Blasdell, K. R., Davis, S. S., Voysey, R., Bulach, D. M., Middleton, D., Williams, S., Harmsen, M. B., Weir, R. P., Crameri, S., Walsh, S. J., Peck, G. R., Tesh, R. B., Boyle, D. B., Melville, L. F. and Walker, P. J. 2020. Hayes Yard virus: a novel ephemerovirus isolated from a bull with severe clinical signs of bovine ephemeral fever is most closely related to Puchong virus. *Vet. Res.* **51**: 58.

4. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. **10**: 421.

5. Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**: 1972–1973.

6. Chen, S., Zhou, Y., Chen, Y. and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. **34**: i884–i890.

7. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. and Li, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*. **10**:.

8. Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B. and Flouri, T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **37**: 291–294.

9. Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**: 3150–3152.

10. Geoghegan, J. L. and Holmes, E. C. 2017. Predicting virus emergence amid evolutionary noise. *Open Biol.* **7**:.

11. Golender, N., Klement, E., Ofer, L., Hoffmann, B., Wernike, K., Beer, M. and Pfaff, F. 2023. Hefer valley virus: a novel ephemerovirus detected in the blood of a cow with severe clinical signs in Israel in 2022. *Arch. Virol.* **168**: 234.

12. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat.*

Biotechnol. **29**: 644–652.

13. Gurao, A., Vasisth, R., Singh, R., Dige, M. S., Vohra, V., Mukesh, M., Kumar, S. and Kataria, R. S. 2022. Identification of differential methylome signatures of white pigmented skin patches in Nili Ravi buffalo of India. *Environ. Mol. Mutagen.* **63**: 408–417.
14. Karabatsos, N. 1978. Supplement to International Catalogue of Arboviruses including certain other viruses of vertebrates. *Am. J. Trop. Med. Hyg.* **27**: 372–440.
15. Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**: 772–780.
16. Kawasaki, J., Kojima, S., Tomonaga, K. and Horie, M. 2021. Hidden Viral Sequences in Public Sequencing Data and Warning for Future Emerging Diseases. *MBio*. **12**: e0163821.
17. Kawasaki, J., Tomonaga, K. and Horie, M. 2023. Large-scale investigation of zoonotic viruses in the era of high-throughput sequencing. *Microbiol. Immunol.* **67**: 1–13.
18. Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**: 907–915.
19. Kircher, M., Sawyer, S. and Meyer, M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**: e3.
20. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. **35**: 4453–4455.
21. Muhire, B. M., Varsani, A. and Martin, D. P. 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One*. **9**: e108277.
22. Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**: 824–834.
23. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. and Korobeynikov, A. 2020. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinformatics*. **70**: e102.
24. Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R.,

Trawick, B. W., Pruitt, K. D. and Sherry, S. T. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**: D20–D26.

25. Shen, W., Le, S., Li, Y. and Hu, F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One*. **11**: e0163962.

26. Shi, M., Zhang, Y.-Z. and Holmes, E. C. 2018. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Res.* **243**: 83–90.

27. Souvorov, A., Agarwala, R. and Lipman, D. J. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol.* **19**: 153.

28. Steinegger, M. and Söding, J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**: 1026–1028.

29. Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**: 1023–1025.

30. Walker, P. J., Freitas-Astúa, J., Bejerman, N., Blasdell, K. R., Breyta, R., Dietzgen, R. G., Fooks, A. R., Kondo, H., Kurath, G., Kuzmin, I. V., Ramos-González, P. L., Shi, M., Stone, D. M., Tesh, R. B., Tordo, N., Vasilakis, N., Whitfield, A. E. and Ictv Report Consortium 2022. ICTV Virus Taxonomy Profile: Rhabdoviridae 2022. *J. Gen. Virol.* **103**:.

31. Walker, P. J. and Klement, E. 2015. Epidemiology and control of bovine ephemeral fever. *Vet. Res.* **46**: 124.

32. Wright, E. S. and Vetsigian, K. H. 2016. Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics*. **17**: 876.

Figure legends

Figure 1. Genomic organization of Punjab virus. (a) Genomic organization and transcription profile of Punjab virus. Pink arrow boxes show open reading frames. Short reads from SRR8476835 were mapped to the Punjab virus contigs and visualized. **(b)** Putative transcription signal sequences of Punjab virus.

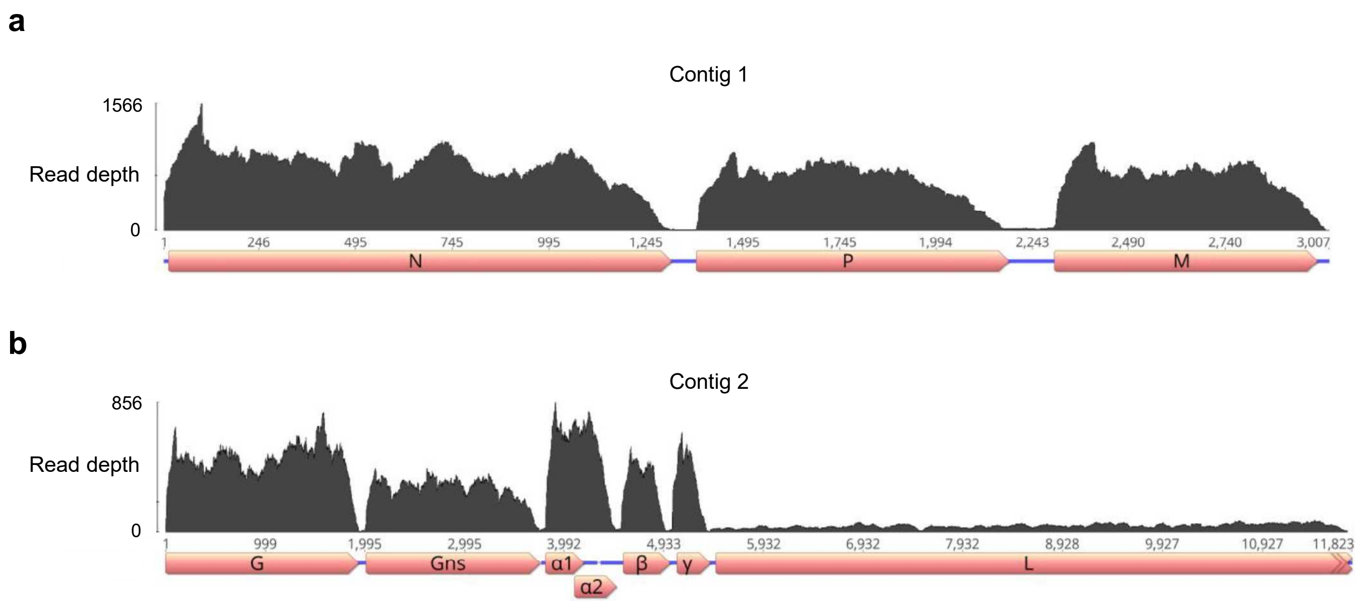
Figure 2. Phylogenetic relationship of Punjab virus and ephemeroviruses. Phylogenetic trees were reconstructed by the maximum likelihood method using amino acid sequences of N **(a)** or L **(b)** protein of Punjab virus, ephemeroviruses, and outgroup rhabdoviruses. Bootstrap values equal to or more than 70 are shown on each branch. The scale bar indicates the number of amino acid substitutions per site.

Figure 3. Sequence divergence of Punjab virus and related ephemeroviruses. Pairwise amino acid sequence identities of N **(a)**, G **(b)**, or L **(c)** proteins between Punjab virus and related ephemeroviruses were determined using Sequence Demarcation Tool [21]. Punjab virus; PUCV, Puchong virus; KoV, Kokolu virus; HYV, Hayes Yard virus; BRMV, Berrimah virus; BEFV, Bovine ephemeral fever virus.

Table 1. The top hits of BLASTx analysis.

	BLAST best hit				
Query	**Accession**	**Virus name**	**Protein**	**Identity (%)**	**Length (aa)**
Contig 1	QEA08650.1	Hayes Yard virus	N protein	90.3	432
Contig 2	QEA08648.1	Puchong virus	L protein	78.4	2098

Figure 1



c

Region	Initiation	Termination
N-P	AACAGG	ATGAAAAAAA
P-M	AACAGG	ATGAAAAAAA
G-Gns	AACAGG	ATGAAAAAAA
Gns-α	AACAGG	ATGAAAAAAA
α-β	AACAGG	ATGAAAAAAA
β-γ	AACAGG	ATGAAAAAAA
γ-L	AACAGG	ATGAAAAAAA

Figure 2

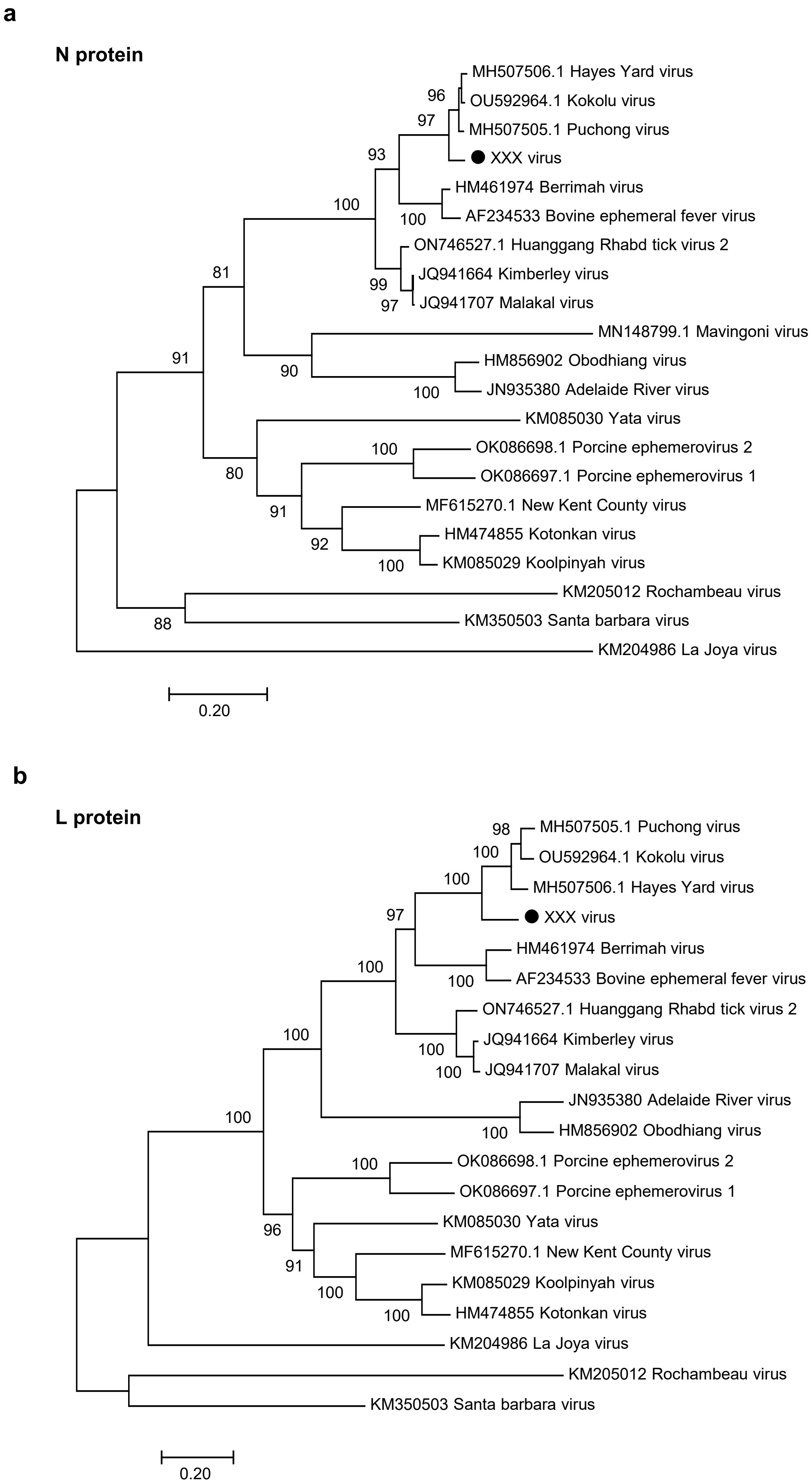
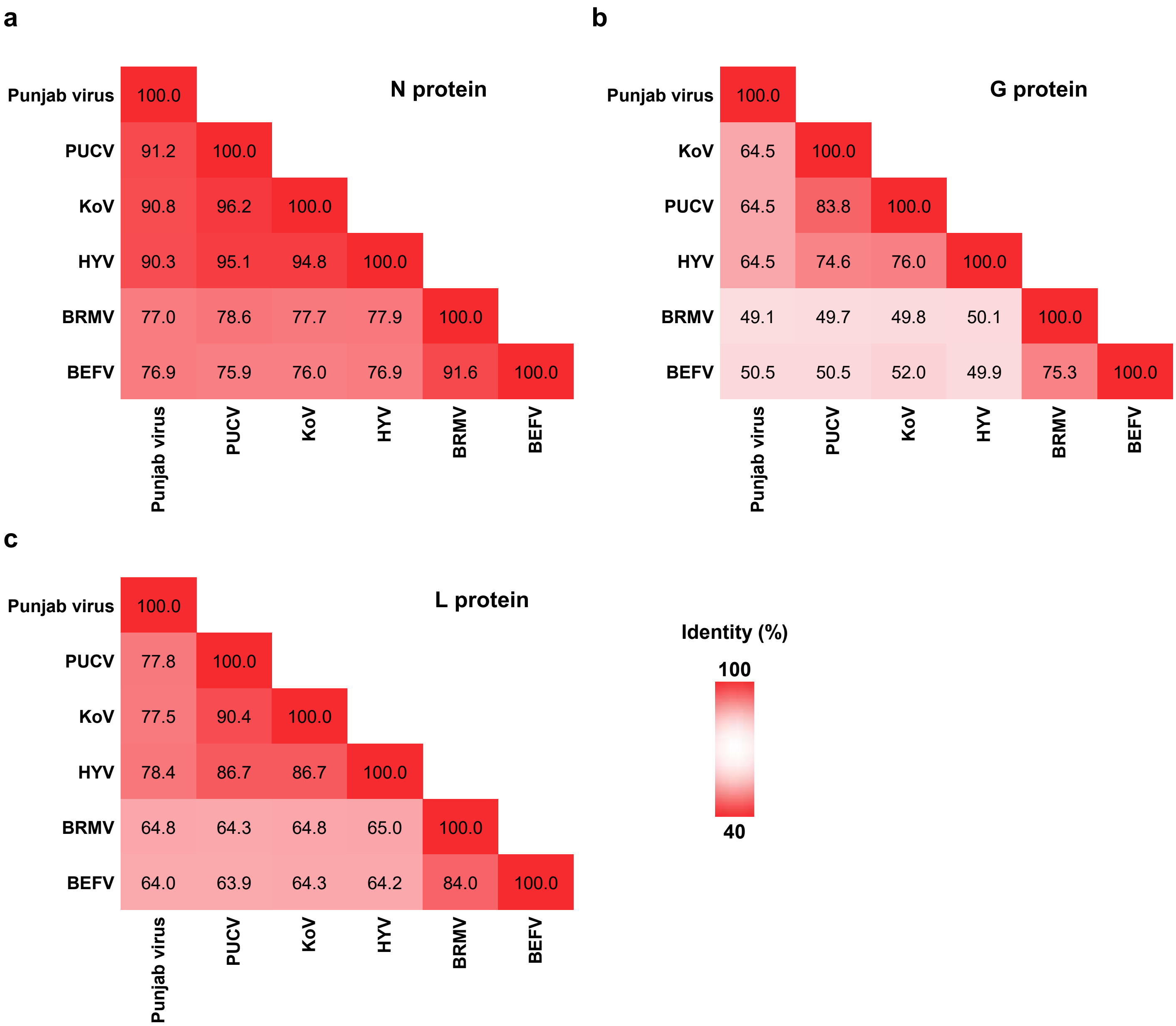

a
N protein
MH507506.1 Hayes Yard virus
OU592964.1 Kokolu virus
MH507505.1 Puchong virus
● XXX virus
HM461974 Berrimah virus
AF234533 Bovine ephemeral fever virus
ON746527.1 Huanggang Rhabd tick virus 2
JQ941664 Kimberley virus
JQ941707 Malakal virus
MN148799.1 Mavingoni virus
HM856902 Obodhiang virus
JN935380 Adelaide River virus
KM085030 Yata virus
OK086698.1 Porcine ephemerovirus 2
OK086697.1 Porcine ephemerovirus 1
MF615270.1 New Kent County virus
HM474855 Kotonkan virus
KM085029 Koolpinyah virus
KM205012 Rochambeau virus
KM350503 Santa barbara virus
KM204986 La Joya virus
96
97
93
100
100
81
99
97
91
90
100
100
80
91
92
100
88
0.20
b
L protein
MH507505.1 Puchong virus
OU592964.1 Kokolu virus
MH507506.1 Hayes Yard virus
● XXX virus
HM461974 Berrimah virus
AF234533 Bovine ephemeral fever virus
ON746527.1 Huanggang Rhabd tick virus 2
JQ941664 Kimberley virus
JQ941707 Malakal virus
JN935380 Adelaide River virus
HM856902 Obodhiang virus
OK086698.1 Porcine ephemerovirus 2
OK086697.1 Porcine ephemerovirus 1
KM085030 Yata virus
MF615270.1 New Kent County virus
KM085029 Koolpinyah virus
HM474855 Kotonkan virus
KM204986 La Joya virus
KM205012 Rochambeau virus
KM350503 Santa barbara virus
98
100
100
97
100
100
100
100
100
100
100
100
96
91
100
100
0.20

Figure 3



N protein

	Punjab virus	PUCV	KoV	HYV	BRMV	BEFV
Punjab virus	100.0					
PUCV	91.2	100.0				
KoV	90.8	96.2	100.0			
HYV	90.3	95.1	94.8	100.0		
BRMV	77.0	78.6	77.7	77.9	100.0	
BEFV	76.9	75.9	76.0	76.9	91.6	100.0

G protein

	Punjab virus	PUCV	KoV	HYV	BRMV	BEFV
Punjab virus	100.0					
KoV	64.5	100.0				
PUCV	64.5	83.8	100.0			
HYV	64.5	74.6	76.0	100.0		
BRMV	49.1	49.7	49.8	50.1	100.0	
BEFV	50.5	50.5	52.0	49.9	75.3	100.0

L protein

	Punjab virus	PUCV	KoV	HYV	BRMV	BEFV
Punjab virus	100.0					
PUCV	77.8	100.0				
KoV	77.5	90.4	100.0			
HYV	78.4	86.7	86.7	100.0		
BRMV	64.8	64.3	64.8	65.0	100.0	
BEFV	64.0	63.9	64.3	64.2	84.0	100.0