Detection of a novel parahenipavirus from northern short-tailed shrews (*Blarina brevicauda* [Say, 1823])

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

[1] School of Veterinary Science, College of Life, Environment, and Advanced Science, Osaka Prefecture University, Izumisano, Osaka, Japan

[2] Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, Izumisano, Osaka, Japan

[3] Osaka International Infectious Diseases Research Center, Osaka Metropolitan University, Osaka, Japan

*Corresponding author

Masayuki Horie, DVM, PhD

Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, 1-58 Rinku Orai-kita, Izumisano 598-8531, Osaka, Japan

Email: mhorie@omu.ac.jp

Abstract

Parahenipaviruses (genus *Parahenipavirus*) are non-segmented negative strand RNA viruses belonging to the family *Paramyxoviridae* of the order *Mononegavirales*. Parahenipaviruses have been detected from shrews and rodents, some of which have been reported to be associated with human diseases. Although many parahenipaviruses have been identified thus far, there still exist large phylogenetic gaps in parahenipaviruses, suggesting the existence of yet-to-be-identified parahenipaviruses. In this study, we analyzed public RNA-seq data and identified a novel parahenipavirus from northern short-tailed shrews (*Blarina brevicauda* [Say, 1823]), tentatively named Blarina brevicauda virus (BbV). Based on sequence comparisons between BbV and its most closely related viruses, which meet the International Committee on Taxonomy of Viruses species demarcation criteria, we propose that BbV is a novel species of virus in the genus *Parahenipavirus*. Furthermore, mapping analysis using RNA-seq data derived from multiple tissues of northern short-tailed shrews suggested that BbV is kidney-tropic. This study provides novel insights into the diversity of parahenipaviruses, which would contribute to a deeper understanding of their evolution and control of infectious diseases.

Introduction

The genus *Parahenipavirus* is a newly established genus within the family *Paramyxoviridae* of the order *Mononegavirales*, composed of viruses relatively closely related to the members of the genus *Henipavirus* [1]. The genome of parahenipaviruses contains conserved genes typical of paramyxoviruses, including N, P/V/W/C, M, F, G/H/HN, and L, arranged from the 3' end. Additionally, a unique ORF X, located between the M and F genes, encodes a putative transmembrane protein that is unique to parahenipaviruses within the family *Paramyxoviridae* [2]. The C, V, and W genes overlap with the P gene; the C protein is derived from an alternate reading frame of P, while the V and W proteins are expressed through mRNA editing, involving the insertion of one or two G residues, respectively.

Parahenipaviruses have been mainly detected from shrews and rodents [2–6], some of which are associated with human diseases. In 2022, Langya virus was detected in febrile patients in China, with shrews identified as the natural reservoir [3]. Similarly, Mojiang virus was detected in rats from an abandoned mine where three patients died of severe pneumonia, although a direct causal link remains unclear [4]. While other parahenipaviruses have not been linked to human disease, their natural hosts, such as rodents and shrews, frequently come into contact with humans, highlighting the need to elucidate the diversity of parahenipaviruses for controlling viral infectious diseases.

The diversity of parahenipaviruses is not fully understood. To date, 12 species and several unclassified viruses have been identified [1]. However, significant phylogenetic gaps within the genus *Parahenipavirus* suggest the existence of many unidentified viruses. Further exploration is needed to deepen our understanding of parahenipavirus diversity and evolution as well as to control infectious diseases.

Sequencing data from diverse samples have been deposited in public databases, some of which may contain virus-derived sequences. Efforts to identify viruses using publicly available data have led to the discovery of many novel viruses, demonstrating the usefulness of re-

analyzing archived data for virus discovery [7]. As new public data continue to accumulate over time, they could serve as a valuable resource for identifying additional novel viruses.

In this study, we searched for parahenipaviruses using public RNA-seq data and identified a novel parahenipavirus from the northern short-tailed shrew (*Blarina brevicauda* [Say, 1823]). We determined the nearly complete genome sequence of the virus and tentatively named it Blarina brevicauda virus (BbV). Mapping analyses revealed that viral reads were predominantly detected in kidney-derived data, suggesting that BbV is kidney-tropic. Furthermore, this virus meets the International Committee on Taxonomy of Viruses (ICTV) species demarcation criteria, and we propose BbV as a novel species of virus within the genus *Parahenipavirus*.

Materials and Methods

Identification of Blarina brevicauda virus from public RNA-seq data

RNA-seq data (SRR17216496) was downloaded from NCBI SRA [8] and preprocessed using fastp (v0.23.4) [9] with the options "-l 35 -x -y". The preprocessed reads were then assembled using SPAdes (v4.0.0) [10] with default settings. Contigs of 100 nucleotides or longer were extracted using SeqKit (v2.4.0) [11] and subjected to a BLASTx [12] search against the NCBI clustered nr database [8] on July 29, 2024.

To validate the viral contig, we mapped the RNA-seq data back to the contig using Magic-BLAST (v1.7.2) [13]with the options "-no_unaligned -splice T -word_size 16 -lcase_masking -limit_lookup F". Read depth at each contig position was calculated using SAMtools (v1.19) [14]. Positions with a read depth of 5 or greater were considered reliable, and regions with a read depth below 5 were removed. The mapping pattern was visualized and manually inspected using Geneious Prime (v2023.2.1, https://www.geneious.com). The final contig, deposited in DDBJ under the accession number BR002473, was used for downstream analyses.

Annotation of Blarina brevicauda virus genome

For viral gene annotation, ORFs longer than 300 nucleotides were extracted. The putative amino acid sequences of these ORFs were subjected to BLASTp searches against viral sequences (taxid: 10239) in the NCBI nr database, using the options "-word size 2 -evalue 1E-10." Viral genes were annotated based on the BLAST results.

To identify putative transcription signal sequences, the intergenic regions were extracted and aligned using MAFFT (v7.490) [15] with the E-INS-i algorithm. Based on this alignment and known paramyxovirus signal sequences, the putative transcription signal sequences were manually determined.

The transmembrane domains were predicted using the DeepTMHMM webserver (v1.0.42) [16].

Phylogenetic analyses

Amino acid sequences of the N and L proteins from BbV, other parahenipaviruses, henipaviruses, and three viruses from other genera within the family *Paramyxoviridae* (Supplementary Table 1) were aligned using MAFFT (v7.490) with the E-INS-i algorithm and ambiguously aligned regions were trimmed with trimAl (v1.4.rev22) [17] using the "-strict" option. Phylogenetic trees were inferred by the maximum likelihood method using RAxML-NG (v1.1.0) [18]. The LG+I+G4+F model, selected by ModelTest-NG (v0.2.0) [19], was used for tree reconstruction. The reliability of the phylogenetic tree was assessed by 1,000 bootstrap resamplings using the transfer bootstrap expectation method. The alignments used for the phylogenetic analyses are available in the Supplementary materials.

Calculation of amino acid sequence identities for classification

According to the ICTV species demarcation criteria, BLASTp was performed against viral proteins (taxid: 10239) in the NCBI nr database using the L protein sequence of BbV as the query, revealing Ninorex virus (OQ438286.1) as the top hit. Pairwise alignments of the amino

acid sequences for six major proteins in the *Paramyxoviridae* family (N, P, M, F, G, or L) from BbV and Ninorex virus were conducted using BLASTp. The average sequence identity was calculated based on these pairwise alignments.

Mapping analysis for detection of Blarina brevicauda virus-derived reads

To detect Blarina brevicauda virus-derived reads, we mapped 306 publicly available RNA-seq datasets obtained from animals in the order *Eulipotyphla* (Supplementary Table 2) to the BbV genome sequence using Magic-BLAST (v1.7.2). The number of mapped reads was then counted using SAMtools (v1.19).

Results

Identification of a novel parahenipavirus-like contigs

During our extensive search for viruses in public databases, we detected a viral contig from RNA-seq data derived from the kidney of *Blarina brevicauda* (family *Soricidae*, order *Eulipotyphla*). A BLASTx search using this contig revealed that the top hit was the L protein of Ninorex virus (WJL29506.1) [6], a member of the genus *Parahenipavirus*, with 74.6% amino acid sequence identity (Table 1).

To evaluate the reliability of the contig, we mapped the RNA-seq data from which the contig was derived to the viral contig sequence. Positions with a read depth of 5 or more were considered reliable, and sequences that did not meet this threshold were removed. The final contig length was 16,645 nucleotides. We named this virus Blarina brevicauda virus (BbV).

Annotation of the Blarina brevicauda virus genome

Next, we determined the genome organization of BbV, which contains nine ORFs. Using the amino acid sequences of these ORFs as queries, we performed BLASTp analysis and found sequence similarity to the N, P/V/W/C, M, ORF X, F, G, and L genes of other parahenipaviruses

(Fig. 1a).

To identify the transcription signal sequences, we aligned the intergenic regions and manually determined the sequences, referencing those of other orthoparamyxoviruses (Fig. 1b). The intergenic CTT motif, conserved in most orthoparamyxoviruses, was also present in all BbV intergenic regions. The putative transcription start and end signals were identified as AGGAYHCARG and TWADRRAAAA, respectively. Notably, the mapping data showed a marked decrease in read depth in the intergenic regions (Figs. 1a and S1), suggesting active viral transcription in the samples.

ORF X, located between the M and F genes, is a distinctive gene of parahenipaviruses and is predicted to encode a transmembrane protein. Using DeepTMHMM, we predicted the transmembrane region of BbV's ORF X to be between amino acids 39-60, which is similar to other parahenipaviruses (Fig. 1c).

Paramyxoviruses typically utilize mRNA editing through the insertion of G nucleotide(s) by transcriptional slippage to express P, V, and W proteins. To investigate whether BbV employs the same mechanism, we reanalyzed the BbV mapping data. We detected G nucleotide insertions at position 2,720 on the BbV genome (Fig. 1), which result in the production of three different proteins: the P protein (551 aa) without insertion, the V protein (337 aa) with a single G insertion, and the W protein (281 aa) with two G insertions. Based on the mapping data, the frequencies of expression for these proteins were 40.9%, 44.1%, and 15%, for P, V, and W, respectively (Fig. 1d and Supplementary Table 3).

Evolutionary analyses on Blarina brevicauda virus

To elucidate the evolutionary relationship between BbV and other parahenipaviruses, we performed phylogenetic analyses using the N and L proteins (Fig. 2a and b). In both phylogenetic trees, BbV clustered with Ninorex virus and Wufeng Chodsigoa smithii henipavirus 1, both detected in animals from the family *Soricidae*.

We then calculated the amino acid sequence identities between BbV and Ninorex virus, the closest relative of BbV, for the six major paramyxovirus proteins (N, P, M, G, F, and L). The sequence identity for each protein is shown in Table 1, with an average identity of 54.2%.

Comprehensive detection of BbV reads in public RNA-seq data

To further characterize BbV, we conducted a comprehensive search for BbV-derived reads in public RNA-seq data through mapping analysis. We mapped 306 RNA-seq datasets from animals of the order *Eulipotyphla*, which includes *Blarina brevicauda*, the putative host of BbV, to the viral contig and counted the mapped reads. As a result, BbV-derived reads were detected in four additional RNA-seq datasets (Table 2), all of which belong to the same BioProject as the one where BbV was initially identified. Among the five RNA-seq datasets positive for BbV (including the initial dataset), four were derived from the kidney and one from the skin. The number of viral reads in the kidney-derived datasets is significantly higher than in the skin-derived dataset. Notably, while this BioProject also contains data from the brain, heart, liver, and skin, no BbV-derived reads were detected in these tissues, except for the aforementioned data from the skin (Table 2).

Discussion

Currently, 12 species and several unassigned viruses are reported in the genus *Parahenipavirus* [1], but large phylogenetic gaps remain, and the diversity of parahenipaviruses has yet-to-be elucidated. In this study, we identified the nearly complete genome sequence of a novel paramyxovirus, BbV, from RNA-seq data of *Blarina brevicauda*. Notably, our analysis showed that the average amino acid sequence identity of the six major viral proteins (N, P, M, F, G, and L) between BbV and its closest relative, Ninorex virus, is 54.2%. This meets the ICTV species demarcation criteria for the family *Paramyxoviridae* (average identity below 85%; https://ictv.global/ictv/proposals/2023.018M.Paramyxoviridae_reorg.zip [accessed on Sep 26,

2024]). Based on these findings, we propose that BbV is a novel species of virus within the genus *Parahenipavirus*.

Our data strongly suggest that *B. brevicauda* is the authentic host of BbV. In metaviromic analyses, viral contamination from various sources can sometimes occur, complicating the identification of authentic hosts. However, our mapping analysis identified a significant number of virus-derived reads in the kidney. Further, we showed a marked decrease in read depth in the intergenic regions, suggesting that the virus actively transcribed in the samples. Additionally, closely related parahenipaviruses have also been detected in shrews [2,3,5,6]. These findings indicate that the virus is actively transcribed in the kidney, strongly suggesting that *B. brevicauda* is the authentic host of BbV.

Our data also suggest that BbV is kidney-tropic. A total of 20 RNA-seq datasets from *B. brevicauda* are registered under BioProject PRJNA788430, from which BbV was detected. These datasets are derived from five different tissues (brain, skin, heart, kidney, and liver) apparently from four individuals according to the metadata. Our mapping analysis revealed that viral reads were almost exclusively found in the kidneys of all four individuals. Although viral reads were detected in the skin of one individual, the quantity was significantly lower than in the kidney. Thus, the viral reads in the skin sample might be due to contamination and/or index hopping. Additionally, Hasua virus, Resua virus, Lechcodon virus, and Denwin virus, parahenipaviruses detected from shrews, were suggested to be kidney tropic [20]. Furthermore, many parahenipaviruses were detected from the kidneys [2,5,6,21]. Thus, many of the members of the genus *Parahenipavirus* may be kidney tropic.

In this study, we identified a novel paraphenipavirus from RNA-seq data derived from *B. brevicauda*, contributing to a deeper understanding of the diversity of paraphenipavirus. However, significant phylogenetic gaps remain, and the characteristics of BbV are not yet fully understood. Further epidemiological and virological studies are necessary to better understand this virus.

Acknowledgements

The super-computing resources were provided by 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 KAKENHI grant numbers 21H01199 (MH), 22K19234 (MH), 23K20902 (MH), 24K21922(MH), and 24K18455 (MK) and the 2024 Osaka Metropolitan University (OMU) Strategic Research Promotion Project (Young Researcher) (MK).

References

1. B. Rima, A. Balkema-Buschmann, W. G. Dundon, P. Duprex, A. Easton, R. Fouchier, G. Kurath, R. Lamb, B. Lee, P. Rota, L. Wang, and Ictv Report Consortium, J. Gen. Virol. **100**, 1593 (2019).

2. B. Vanmechelen, S. Meurs, M. Horemans, A. Loosen, T. Joly Maes, L. Laenen, V. Vergote, F. R. Koundouno, N. Magassouba, M. K. Konde, I. S. Condé, M. W. Carroll, and P. Maes, Virus Evol. **8**, veac061 (2022).

3. X.-A. Zhang, H. Li, F.-C. Jiang, F. Zhu, Y.-F. Zhang, J.-J. Chen, C.-W. Tan, D. E. Anderson, H. Fan, L.-Y. Dong, C. Li, P.-H. Zhang, Y. Li, H. Ding, L.-Q. Fang, L.-F. Wang, and W. Liu, N. Engl. J. Med. **387**, 470 (2022).

4. Z. Wu, L. Yang, F. Yang, X. Ren, J. Jiang, J. Dong, L. Sun, Y. Zhu, H. Zhou, and Q. Jin, Emerg. Infect. Dis. **20**, 1064 (2014).

5. S.-H. Lee, K. Kim, J. Kim, J. S. No, K. Park, S. Budhathoki, S. H. Lee, J. Lee, S. H. Cho, S. Cho, G.-Y. Lee, J. Hwang, H.-C. Kim, T. A. Klein, C.-S. Uhm, W.-K. Kim, and J.-W. Song, Viruses **13**, 2020 (2021).

6. M. Horemans, J. Van Bets, T. Joly Maes, P. Maes, and B. Vanmechelen, Virus Evol. **9**, vead065 (2023).

7. J. Kawasaki, S. Kojima, K. Tomonaga, and M. Horie, MBio **12**, e0163821 (2021).

8. E. W. Sayers, E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, and S. T. Sherry, Nucleic Acids Res. **50**, D20 (2022).

9. S. Chen, Y. Zhou, Y. Chen, and J. Gu, Bioinformatics **34**, i884 (2018).

10. A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, and A. Korobeynikov, Curr. Protoc. Bioinformatics **70**, e102 (2020).

11. W. Shen, S. Le, Y. Li, and F. Hu, PLoS One **11**, e0163962 (2016).

12. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden, BMC Bioinformatics **10**, 421 (2009).

13. G. M. Boratyn, J. Thierry-Mieg, D. Thierry-Mieg, B. Busby, and T. L. Madden, BMC Bioinformatics **20**, 405 (2019).

14. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li, Gigascience **10**, (2021).

15. K. Katoh and D. M. Standley, Mol. Biol. Evol. **30**, 772 (2013).

16. J. Hallgren, K. D. Tsirigos, M. D. Pedersen, J. J. Almagro Armenteros, P. Marcatili, H. Nielsen, A. Krogh, and O. Winther, BioRxiv (2022).

17. S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón, Bioinformatics **25**, 1972 (2009).

18. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, and A. Stamatakis, Bioinformatics **35**, 4453 (2019).

19. D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, and T. Flouri, Mol. Biol. Evol. **37**, 291 (2020).

20. V. C. Haring, B. Litz, J. Jacob, M. Brecht, M. Bauswein, J. Sehl-Ewert, M. Heroldova, C. Wylezich, D. Hoffmann, R. G. Ulrich, M. Beer, and F. Pfaff, Microb. Genom. **10**, (2024).

21. Y.-M. Chen, S.-J. Hu, X.-D. Lin, J.-H. Tian, J.-X. Lv, M.-R. Wang, X.-Q. Luo, Y.-Y. Pei, R.-X. Hu, Z.-G. Song, E. C. Holmes, and Y.-Z. Zhang, Cell **186**, 4662 (2023).

Figure legends

Figure 1. Genome organization of Blarina brevicauda virus (BbV).

(A) Schematic representation of the open reading frames (ORFs) and transcription signal sequences of BbV. **(B)** Predicted transcription signal sequences of BbV. **(C)** Prediction of the transmembrane domain of the ORF X protein using DeepTMHMM. **(D)** Schematic diagram illustrating the P, V, and W proteins (left) and the average frequencies of guanine base insertions (right). Detailed data are provided in Supplementary Table 3.

Figure 2. Phylogenetic trees of parahenipaviruses, henipaviruses, and other paramyxoviruses. Maximum likelihood phylogenetic trees were inferred using amino acid sequences of the N **(A)** and L **(B)** proteins. Bootstrap values below 70 are not shown. The scale bar represents the number of amino acid substitutions per site.

Table 1. One-on-one BLASTp comparison between Blarina brevicauda virus and Ninorex virus

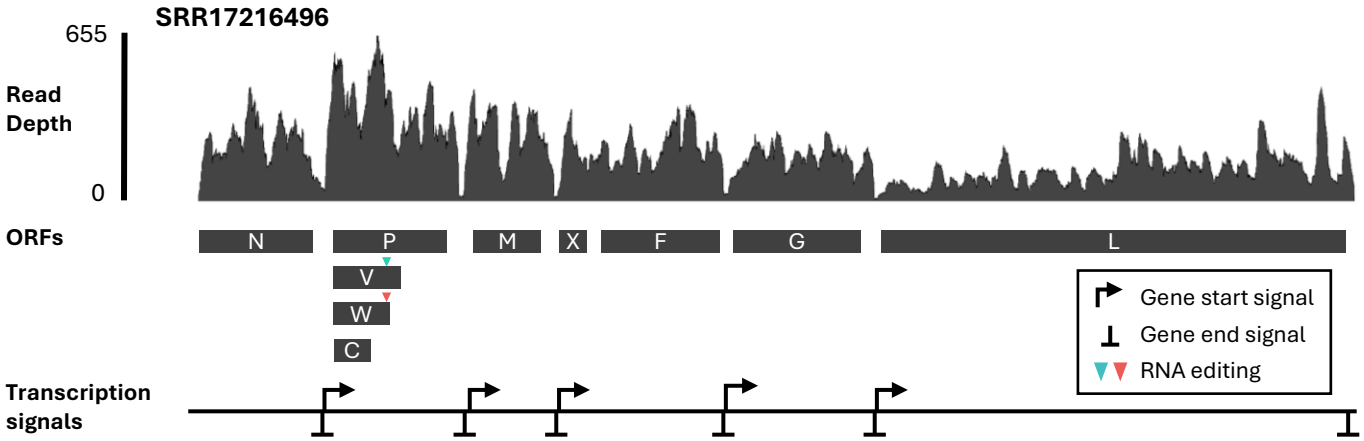
Protein	Identity (%)
N	50.7
P	34.7
M	76.2
F	55.7
G	33.4
L	74.6
Average	54.2

Table 2. Sample information and mapped read numbers in BioProject PRJNA788430

SRA Run accession	Sample name	Mapped read number	Read per million	Tissue
SRR17216479	Shrew217	0	0	Brain
SRR17216480		0	0	Heart
SRR17216481		14,304	124.2	Kidney
SRR17216482		0	0	Liver
SRR17216483		0	0	Skin
SRR17216484	Shrew218	0	0	Brain
SRR17216485		0	0	Heart
SRR17216486		7,081	51.89	Kidney
SRR17216487		0	0	Liver
SRR17216488		0	0	Skin
SRR17216489	Shrew220	0	0	Brain
SRR17216490		0	0	Heart
SRR17216491		18,978	170.74	Kidney
SRR17216492		0	0	Liver
SRR17216493		35	0.26	Skin
SRR17216494	Shrew221	0	0	Brain
SRR17216495		0	0	Heart
SRR17216496		24,085	186.08	Kidney
SRR17216497		0	0	Liver
SRR17216498		0	0	Skin

Figure 1

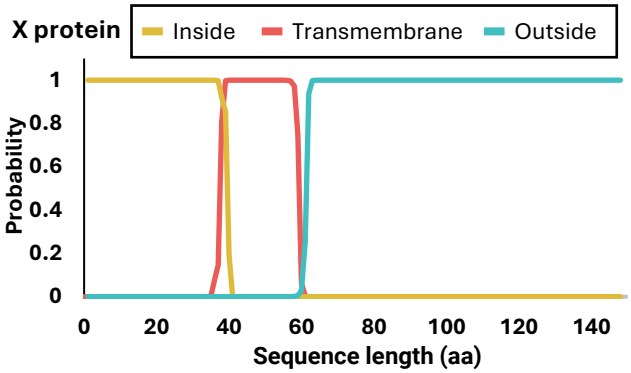
a



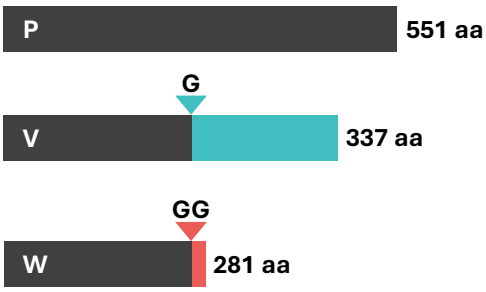
b

	Gene end	Intergenic region	Gene start
N-P	TAATAAAAAA	CTT	AGGATCCAGG
P-M	TTAGAAAAAA	CTT	AGGATTCAAG
M-X/F	TTAAAGAAAA	CTT	AGGATTCAAG
X/F-G	TTATAAAAAA	CTT	AGGATTCAAG
G-L	ATATGAAGAA	CTT	AGGACACAAG
L-trailer	TTAAGAAAAA		

c



d



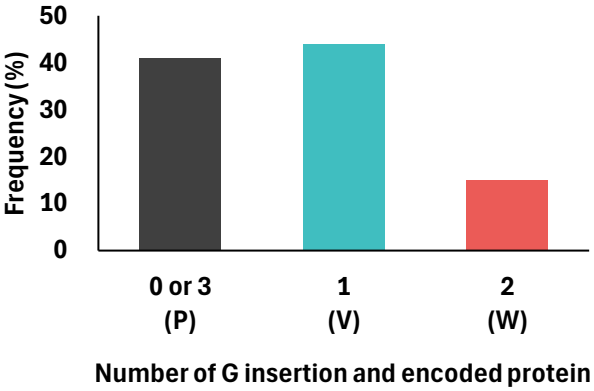


Figure 2

a

N protein

0.20

Shrew
Rodent
Human

98 OM101125.1 Langya virus
86 KF278639.1 Mojiang virus
93 OQ715593.1 Wenzhou shrew henipavirus 1
100 MZ328275.1 Wenzhou Apodemus agrarius henipavirus 1
98 OR713881.1 Hasua virus
OR713879.1 Lechcodon virus
99
100 OM030315.1 Jingmen Crocidura shantungensis henipavirus 2
MZ574409.1 Daeryong virus
97 OK623353.1 Melian virus
93 OK623354.1 Denwin virus
OM030316.1 Wufeng Chodsigoa smithii henipavirus 1
93
OQ438286.1 Ninorex virus
96 BR002473 Blarina brevicauda virus
OR713876.1 Resua virus
99
100 OM030314.1 Jingmen Crocidura shantungensis henipavirus 1
97 MZ574407.1 Gamak virus
OQ970176.1 Crocidura tanakae henipavirus
99 OM030317.1 Wufeng Crocidura attenuata henipavirus 1
ON613535.1 Angavokely henipavirus
97
HQ660129.1 Ghana virus
99 JQ001776.1 Cedar virus
75 AF212302.2 Nipah virus
99 AF017149.3 Hendra virus
AY900001.1 J virus
OK623365.1 Meleucus virus
58 AF014953.1 Canine distemper virus

Parahenipavirus

Henipavirus

b

L protein

0.20

Shrew
Rodent
Human

100 OM101125.1 Langya virus
83 KF278639.1 Mojiang virus
75 OR713881.1 Hasua virus
100 OQ715593.1 Wenzhou shrew henipavirus 1
100 MZ328275.1 Wenzhou Apodemus agrarius henipavirus 1
OR713879.1 Lechcodon virus
100
100 OM030315.1 Jingmen Crocidura shantungensis henipavirus 2
MZ574409.1 Daeryong virus
93
100 OK623353.1 Melian virus
93 OK623354.1 Denwin virus
OM030316.1 Wufeng Chodsigoa smithii henipavirus 1
100
84 OQ236120.1 Chodsigoa hypsibia henipavirus
100 OQ438286.1 Ninorex virus
BR002473 Blarina brevicauda virus
OR713876.1 Resua virus
100
100 OM030314.1 Jingmen Crocidura shantungensis henipavirus 1
100 MZ574407.1 Gamak virus
72 OQ970176.1 Crocidura tanakae henipavirus
99 OM030317.1 Wufeng Crocidura attenuata henipavirus 1
ON613535.1 Angavokely henipavirus
71
HQ660129.1 Ghana virus
99 JQ001776.1 Cedar virus
73 AF212302.2 Nipah virus
100 AF017149.3 Hendra virus
AY900001.1 J virus
OK623365.1 Meleucus virus
100 AF014953.1 Canine distemper virus

Parahenipavirus

Henipavirus