Full Paper (Wildlife Science)

Title
Viral-derived DNA invasion and individual variation in an Indonesian population of large flying fox Pteropus vampyrus

Authors
Atsuo Iida¹, Hitoshi Takemae¹², Ronald Tarigan¹, Ryosuke Kobayashi¹, Hirokazu Kato³, Hiroshi Shimoda⁴, Tsutomu Omatsu², Supratikno⁵, Chaerul Basri⁵, Ni Luh Putu Ika Mayasari⁵, Srihadi Agungpriyono⁵, Ken Maeda⁴⁶, Tetsuya Mizutani², Eiichi Hondo¹*

Affiliation
ⁱLaboratory of Animal Morphology, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8602, Japan
²Laboratory of Veterinary Microbiology, Cooperative Department of Veterinary Medicine, Tokyo University of Agriculture and Technology, Sawai, Fuchu, Tokyo 183-8509, Japan
³Biology and Somatology Related Support Section, Nagoya University, Nagoya 464-8602, Japan
⁴Laboratory of Veterinary Microbiology, Joint Faculty of Veterinary Medicine, Yamaguchi University, Yamaguchi 753-8515, Japan
⁵Faculty of Veterinary Medicine Bogor Agricultural University - IPB University
⁶Department of Veterinary Science, National Institute of Infectious Diseases, Tokyo 162-8640, Japan

Running head
Viruses in Indonesian megabat
*Correspondence

Eiichi Hondo; ehondo@nuagr1.agr.nagoya-u.ac.jp; Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8602, Japan
Abstract
Here, we performed next-generation sequencing (NGS) on six large flying foxes (*Pteropus vampyrus*) collected in Indonesia. Seventy-five virus species in the liver tissue of each specimen were listed. Viral homologous sequences in the bat genome were identified from the listed viruses. This finding provides collateral evidence of viral endogenization into the host genome. We found that two of the six specimens bore partial sequences that were homologous to the plant pathogens *Geminiviridae* and *Luteoviridae*. These sequences were absent in the *P. vampyrus* chromosomal sequences. Hence, plant viral homologous sequences were localized to the hepatocytes as extrachromosomal DNA fragments. Therefore, this suggests that the bat is a potential carrier or vector of plant viruses. The present investigation on wild animals offered novel perspectives on viral invasion, variation, and host interaction.

Keywords
endogenization, Indonesia, plant virus, *Pteropus vampyrus*, virus
Introduction

Bats are major natural viral reservoirs and may cause zoonoses [6, 14, 33]. They move over a wide range, harbor numerous live viruses, and directly or indirectly transmit them to other animals and humans [2, 24]. As their biology and ecology are complex, it is difficult to estimate the risk that they could initiate viral epidemics or pandemics [7].

The large flying fox *Pteropus vampyrus* belongs to the family Pteropodidae. This megabat is distributed in Southeast Asia including Indonesia [7]. Fruit bat habitats vary among human societies. Humans may enter into contact with bats by consuming bushmeats. Moreover, sick or injured bats may fall from trees in residential areas [1, 17]. Large flying foxes are natural viral hosts. They can harbor Hendra and Nipah viruses [25, 26, 27]. These viruses should be transmitted from bats to humans either directly or via intermediate hosts such as domestic pigs and horses, long-tailed macaques (*Macaca fascicularis*), crested serpent eagles (*Spilornis cheela*), and wild boars (*Sus scrofa*) [13]. To predict future spillover events from wildlife, it is necessary to characterize the viral species borne by natural hosts and monitor host distribution and behavior. These investigations and analyses are feasible for domestic animals. However, artificial control of wild animals is difficult [15]. Our joint research team consisted of Indonesian and Japanese scientists who studied the risk and control of infectious viral diseases transmitted by flying foxes in Indonesia. Our previous ecological studies investigated the behavioral traits of flying foxes and bat-human interactions in Indonesia [4, 16, 17]. We isolated Pteropine orthoreovirus from an Indonesian population of *P. vampyrus*. This pathogen belongs to the Reoviridae and causes respiratory symptoms in humans [30]. Nevertheless, rapid, exhaustive screening methods are required to estimate the risks of pathogenesis in wild host animals.
The analysis of viral sequences in organisms helps disclose the current viral infection status and transmission risks as well as prior viral invasions as endogenous viral elements. Several virus-derived sequences integrated in the genome are called retroviral elements \[12, 20\]. Certain genes that govern the evolutionary acquisition of novel traits might have been derived from viral endogenizations in ancestral species \[3, 23\]. An earlier report demonstrated that viral sequence endogenization into the host genome could occur in extant species such as bats and may include non-retroviral RNA viruses \[18, 21\]. Thus, investigation of viral homologous sequences in host cells is invaluable for research into infectious diseases, evolutionary modification of the host genome, and virus-host interactions.

In the present study, we estimate conserved viral sequences based on next-generation sequencing (NGS) and \textit{in silico} analysis. We identify the viral homologous sequences in Indonesian \textit{P. vampyrus} and indicate the differences among sequences within each specimen.
Materials and Methods

Animal experiment

All applicable international, national, and/or institutional guidelines for animal care and use were followed. Flying foxes were captured with permission of the Director-General of the Conservation of Natural Resources and Ecosystems of the Indonesian Ministry of Forestry (No. SK.211/KSDAE/SET/KSA.2/7/2016). Permission to collect 100 individuals annually was also specified in the aforementioned document. Permission to trap flying foxes with nets was authorized by the Regional Office of the Environment and Forestry in Garut (No. BA.09/BBKSDA.JABAR-5.1/Res Wil XVIII/2016; September 2016). A wild animal transportation permit was authorized by the Regional Office of the Environment and Forestry in Ciamis (SI.953/BBKSDA-JABAR.5/2016; September 2016). Sample collection was conducted with permission of the Animal Ethics Committee of Bogor Agricultural University, Bogor, West Java, Indonesia.

Sampling

Large flying foxes were collected by net traps around the Leuweung Sancang Conservation Area of West Java, Indonesia in September and October 2016. The collected specimens were transported to Bogor Agricultural University. Liver tissues were excised from the specimens under anesthesia with 2 mg kg\(^{-1}\) xylazine and 10 mg kg\(^{-1}\) ketamine. DNA was extracted from the livers with a DNeasy blood & tissue kit (Qiagen, Valencia, CA, USA). Approximately ~25mg liver pieces were used for the DNA extraction according to the kit protocol. Purified DNA solutions were sent to Novogene Co. Ltd. (Beijing, China) for NGS. The NGS data was deposited to the DNA Data Bank of Japan (DDBJ, ID: DRA011241).
Identification of viral homologous sequences from NGS reads

The NGS reads were homology-searched by BLAST to all viral genome sequences published in NCBI Virus (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/). The e-value threshold was set to < 1E-5. The data are presented in Supplementary Data 1–6. Except for bacterial viruses and phages, the viral species were manually selected and are listed in Supplementary Table 1. Classification, heritable factor, and major host data were added to the listed viral species as annotations. The viruses were sorted by each property. The data are graphically shown in Figures 1A–1C to clarify the viral trends for each specimen.

Phylogenetic analysis

Orthologous sequences of the U6 spliceosomal small RNA and dnajc14 gene were collected from the NCBI nucleotide (https://www.ncbi.nlm.nih.gov/nucleotide/). The query sequences used in the phylogenetic analysis are listed in Supplementary Table 2. Phylogenetic trees were constructed by the neighbor-joining (NJ) method with 1,000 bootstrap replicates in MEGAX (Ver.10.1.8) software (https://www.megasoftware.net/).

Estimation of endogenization into the bat genome

The top hit query sequences in each viral species were used in a BLAST search to the published P. vampyrus genome contigs. The sequences estimated as endogenous elements were mapped into the bat genome and compared to those for other animal genomes based on the published information in NCBI and Ensembl Genome Browser (https://www.ncbi.nlm.nih.gov/; https://asia.ensembl.org/index.html). A CLC Genomics Workbench (CLC Bio, Aarhus, Denmark) was used to reconstruct P. vampyrus genome contigs with the NGS reads derived from the specimens. Contigs > 1,000 bp were collected as predicted genome sequences. The viral
homologous reads unmapped to both the published genomes and the reconstructed contigs were estimated as extrachromosomal sequences.
Results

Identification of viral homologous sequences in large flying foxes

We collected six large flying foxes (*Pteropus vampyrus*) in Garut, West Java, Indonesia. Approximately 1.2–2.0 × 10^8 fragments of 150-bp DNA reads were obtained from each specimen by NGS. The reads were mapped to the viral genome sequences published in the NCBI and the hits were listed as viruses indigenous to the Indonesian large flying fox population (Supplementary Fig. 1, Supplementary Data 1-6). We identified 52–156 (average 74.8) sequences homologous to known eukaryotic viruses from each specimen (Supplementary Table 1).

Sequences homologous to bacterial viruses and phages were excluded because they might have been derived from parasitic hepatic bacteria rather than the bat cells *per se*.

Individual variability of viral homologous sequences in Indonesian bats

The listed viral species were sorted by classification, genetic factors, and major host species. IPB_4 and IPB 41 exhibited similar viral properties. More than half the identified sequences were classified as retroviruses (Figure 1A). RNA viruses were comparatively rare (Figure 1B). The two major viral hosts were mammals and birds (Figure 1C). IPB_4 and 41 comprised numerous plant viruses belonging to the Geminiviridae or Luteoviridae. These data indicate that Indonesian *P. vampyrus* harbored various viral homologous sequences and the number of viral species differed among individuals. Approximately 13% (21/164) of the listed viruses were identified in all bats. These sequences are estimated to be endogenous elements common in the host genome or widespread infectious viruses in the Indonesian population. In contrast, 85% (136/164) of the viruses were found only in single bats. These sequences might be derived from rare infectious viruses that are not endogenized into the host genome. Hence, we classified the listed viruses as follows: (1) DNA viruses identified in all specimens, (2) RNA viruses identified in all specimens,
(3) DNA viruses identified in a single specimen, and (4) RNA viruses identified in a single specimen (Figure 1D). To investigate virus-host interactions, the listed sequences for the selected viral species were mapped onto the published *P. vampyrus* genome (NCBI *Pteropus vampyrus* Annotation Release 100, GCF_000151845.1).

**Type 1: DNA viruses identified in all specimens**

*Bracoviruses* belong to the *Polydnaviridae* and generally infect insects. The viral homologous sequences identified in the bat specimens were mapped onto a U6 spliceosomal small RNA in the *P. vampyrus* genome. Except for the stem loop motif, the U6 sequences are highly conserved among animals (Figure 2A). A phylogenetic tree plotted using the U6 from various species and the conserved *Bracovirus* sequence estimated that the viral sequences diverged from other animals at the common ancestor stage (Figure 2B). We also found homology between a portion of the Pandoravirus dulcis genome and *H2B clustered* which is the *P. vampyrus* histone 4 gene (Supplementary Fig. 2). The sequence identity was lower than that between *Bracovirus* and U6 spliceosomal small RNAs.

**Type 2: RNA viruses identified in all specimens**

Bovine viral diarrhea virus (BVDV) belongs to the *Flaviviridae* and generally infects domestic animals. They were detected as DNA sequences and the viral genomic RNA was probably reverse-transcribed in the bat cells or later endogenized into the host genome. A sequence homologous to BVDV was mapped to the exon of the *dnajc14* cochaperone gene in the *P. vampyrus* genome. The 161-bp sequence is highly conserved between BVDV and vertebrates (Figure 3A). The homologous sequence was a single *dnajc14* exon in numerous vertebrates except cartilaginous fish. However, the 161-bp sequence was conserved in the elephant shark.
Callorhinchus milii. Nevertheless, it was probably not spliced as a single exon (Figure 3B). Thus, the BVDV homologous sequences in dnajc14 were estimated to be ancestral vertebrate traits. In contrast, a phylogenetic tree plotted by using the 161-bp of dnajc14 and the conserved BVDV sequence predicted that the viral sequence had the strongest homology to the bovine sequence (Figure 3C).

Type 3: DNA viruses identified in a single specimen

Tomato yellow leaf curl virus (TYLCV) belongs to the Geminiviridae and generally infects plants. The two 150-bp read sequences completely matched the viral sequences in the IPB_4 bat specimen (Figures 4A, see also Supplementary Fig. 3). The sequences were unmapped both in the published P. vampyrus genome and in the reconstructed contigs (> 1,000 bp) from the NGS reads of the IPB_4 specimen. Thus, the sequences homologous to TYLCV are extrachromosomal DNAs that are not endogenized into the host genome.

Type 4: RNA viruses identified in a single specimen

Beet mild yellowing virus (BMYV) belongs to the Luteoviridae and generally infects plants. A 473-bp DNA contig homologous to the viral genome was reconstructed using six read sequences from the IPB_41 specimen (Figures 4B, see also Supplementary Fig. 4). They were detected as DNA sequences and it was predicted that the viral genomic RNA was reverse-transcribed in the bat cells or later endogenized into the host genome. However, the read sequences were unmapped both in the published P. vampyrus genome and the reconstructed contigs (> 1,000-bp) from the NGS reads of the IPB_41 specimen. Hence, the sequences homologous to BMYV are extrachromosomal DNAs that are not endogenized into the host genome. The contig region overlapped with the open reading frames (ORF) of the viral, coat, and hypothetical proteins.
Discussion

A previous study estimated that the *P. vampyrus* genome size is ~2.32Gb [28]. Thus, the NGS reads obtained in this study (1.2–2.0 × 10^8 150-bp fragments) covered the *P. vampyrus* genome 7.7–10 folds. The DNA reads of the wild bat specimens included various viral homologous sequences. Certain viruses were listed in all specimens and mapped onto the published *P. vampyrus* genome. These sequences were regarded as endogenous viral elements. Their integration in the host genome and the evolutionary virus-host relationships were validated from phylogenetic analyses.

We found that the sequence homologous to Diolcogaster facetosa bracovirus was mapped to the U6 spliceosomal small RNA coding region which is a component of a small nuclear ribonucleoprotein [31]. The sequence was highly conserved among all specimens. The phylogenetic analysis indicated viral sequence invasion into the genome of a common ancestor of extant animals. The sequence then acquired the role of a spliceosomal small RNA immediately after it was integrated into the chromosome. The *Bracovirus* and endogenous U6 sequences have evolved independently. We found homology between a portion of the Pandoravirus dulcis genome and the *histone H2B* gene previously reported. Hence, our methodology was reliable [35]. These sequences are general transitions of viral endogenizations [9]. For BVDV, the 161-bp sequences are homologous to the *dnajc14* exon in vertebrate genomes. Previous reports indicated that the DNAJC14 protein is a potent BVDV replication modulator in host cells [19, 34]. However, the evolutionary relationship between the RNA virus and the cochaperone gene is unknown. Our findings suggest endogenization of the BVDV sequence into the *dnajc14* locus in the common vertebrate ancestor. In contrast, a phylogenetic analysis indicated that the BVDV sequence is highly conserved in bovine *dnajc14*. Therefore, the BVDV genome has been updated
via RNA recombination with the host-derived dnajc14 transcript [5]. The approach of the present study could estimate the evolutionary relationship between the endogenous virus and the host genome.

The sequences that were entirely associated with TYLCV occurred in IPB_4 but were unmapped to both the published *P. vampyrus* genome and the contigs re-constructed from the NGS reads. TYLCV is a DNA virus that causes yellow leaf curl disease in tomato. Its vector is the whitefly *Bemisia tabaci* [8, 10]. A 473-bp contig homologous to BMYV was obtained from IPB_41. The RNA virus BMYV causes cultivated beet plant yield loss [29]. The BMYV homologous sequence was unmapped to the bat genome contigs. Thus, the plant viruses invaded the bat cells and a portion of the RNA genome was reverse-transcribed in BMYV. A replication origin of host-derived reverse-transcriptase might be included in the BMYV genome corresponding to the upstream or 5’-region of the contig. The contigs only partially matched the published BMYV sequences. Therefore, we did not rule out the possibility that the sequences were derived from other viruses related to BMYV and belonging to the *Luteoviridae*. Our data indicate that certain Indonesian large flying foxes harbored plant viruses. Wild flying foxes can become infected with plant viruses because the main food sources of these animals are fruits and flower nectar [11]. Nevertheless, our dataset was too small to confirm that fruit bats are natural plant virus hosts. We only found homologous sequence fragments in the bat tissue and did not validate viral replication, release, or clinical symptoms *in vivo* [32]. Natural bats may nonetheless harbor these viruses among the plants in their habitat.

Here, we identified viral homologous sequences in a wild large flying fox population. The NGS and mapping-based methods classified the sequences as endogenous elements or
extrachromosomal DNAs. We also found specific viral invasions only in particular bat specimens. Individual variations in host tissue viral species profiles are difficult to distinguish using laboratory animal models or only a few wild animal specimens. The present study demonstrated a powerful method for rapidly screening potential reservoir animals in wild populations and detecting potential human and agricultural crop viral pathogens.
Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgements

Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. This study was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS) “Ecological studies on flying foxes and their involvement in rabies-related and other viral infectious diseases”. This program is a collaboration of the Japan Agency for Medical Research and Development (AMED), the Japan International Cooperation Agency (JICA), and the Japan Society for the Promotion of Science KAKENHI (Grant No. 20K21519 to T.M.).
References

1. Akem, E.S. and Pemunta, N.V. 2020. The bat meat chain and perceptions of the risk of contracting Ebola in the Mount Cameroon region. *BMC Publ. Heal.* **20**: 593.

2. Allocati, N., Petrucci, A.G., Di Giovanni, P., Masulli, M., Di Ilio, C. and De Laurenzi, V. 2016. Bat-man disease transmission: zoonotic pathogens from wildlife reservoirs to human populations. *Cell Death Discov.* **2**: 16048.

3. Baba, K., Nakaya, Y., Shojima, T., Muroi, Y., Kizaki, K., Hashizume, K., Imakawa, K. and Miyazawa, T. 2011. Identification of novel endogenous betaretroviruses which are transcribed in the bovine placenta. *J. Virol.* **85**: 1237–1245.

4. Basri, C., Arifin, E.M.Z., Takemae, H., Hengjan, Y., Iida, K., Sudarnika, E., Zahid, A., Soejoedono, R.D., Susetya, H., Sumiarto, B., Kobayashi, R., Agungpriyono, S. and Hondo, E. 2017. Potential risk of viral transmission from flying foxes to domestic animals and humans on the southern coast of West Java, Indonesia. *J. Vet. Med. Sci.* **79**: 1615–1626.

5. Becher, P. and Tautz, N. 2011. RNA recombination in pestiviruses: cellular RNA sequences in viral genomes highlight the role of host factors for viral persistence and lethal disease. *RNA Biol.* **8**: 216–224.

6. Calisher, C.H., Childs, J.E., Field, H.E., Holmes, K.V. and Schountz, T. 2006. Bats: important reservoir hosts of emerging viruses. *Clin. Microbiol. Rev.* **19**: 531–545.

7. Carlson, C.J. 2020. From PREDICT to prevention, one pandemic later. *The Lancet* **1**: E6–7.

8. Czosnek, H. and Laterrot, H. 1997. A worldwide survey of tomato yellow leaf curl viruses. *Arch. Virol.* **142**: 1391–1406.

9. Feschotte, C. and Gilbert, C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13**: 283–296. Published 2012 Mar 16.
10. Ghanim, M. Morin, S., Zeidan, M. and Czosnek, H. 1998. Evidence for transovarial transmission of tomato yellow leaf curl virus by its vector, the whitefly *Bemisia tabaci*. *Virology* **240**: 295–303.

11. Goodwin, R.E. 1979. The bats of Timor: Systematics and ecology. *Bull. AMNH* **163**: 2.

12. Greenwood, A.D., Ishida, Y., O'Brien, S.P., Roca, A.L. and Eiden, M.V. 2017. Transmission, evolution, and endogenization: Lessons learned from recent retroviral invasions. *Microbiol. Mol. Biol. Rev.* **82**: e00044-17.

13. Gumert, M.D., Fuentes, A. and Jones-Engel, L. 2011. Monkeys on the Edge: Ecology and Management of Long-tailed Macaques and their Interface with Humans, Cambridge University Press, Cambridge.

14. Han, H.J., Wen, H.L., Zhou, C.M., Chen, F.F., Luo, L.M., Liu, J.W. and Yu, X.J. 2015. Bats as reservoirs of severe emerging infectious diseases. *Virus Res.* **205**: 1–6.

15. Haydon, D.T., Cleaveland, S., Taylor, L.H. and Laurenson, M.K. 2002. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerg. Infect. Dis.* **8**: 1468–1473.

16. Hengjan, Y., Pramono, D., Takemae, H., Kobayashi, R., Iida, K., Ando, T., Kasmono, S., Basri, C., Fitriana, Y.S., Arifin, E.M.Z., Ohmori, Y., Maeda, K., Agungpriyono, S. and Hondo, E. 2017. Daytime behavior of *Pteropus vampyrus* in a natural habitat: the driver of viral transmission. *J. Vet. Med. Sci.* **79**: 1125–1133.

17. Hengjan, Y., Saputra, V. and Mirsageri, M. 2018. Nighttime behavioral study of flying foxes on the southern coast of West Java, Indonesia. *J. Vet. Med. Sci.* **80**: 1146–1152.

18. Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M. and Tomonaga, K. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**: 84–87.
19. Isken, O., Postel, A., Bruhn, B. Lattwein, E., Becher, P. and Tautz, N. 2019. CRISPR/Cas9-mediated knockout of DNAJC14 verifies this chaperone as a pivotal host factor for RNA replication of pestiviruses. *J. Virol.* **93**: e01714-18.

20. Johnson, W.E. 2019. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* **17**: 355–370.

21. Kondoh, T., Manzoor, R., Nao, N., Maruyama, J., Furuyama, W., Miyamoto, H., Shigeno, A., Kuroda, M., Matsuno, K., Fujikura, D., Kajihara, M., Yoshida, R., Igarashi, M. and Takada, A. 2017. Putative endogenous filovirus VP35-like protein potentially functions as an IFN antagonist but not a polymerase cofactor. *PLoS One* **12**: e0186450.

22. Kunz, T. and Jones, D. 2000. *Pteropus vampyrus*. *Mammal. Species* **642**: 1–6.

23. Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., Kohda, T., Ogura, A., Yokoyama, M., Kaneko-Ishino, T. and Ishino, F. 2006. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**: 101–106.

24. Plowright, R.K., Eby, P., Hudson, P.J., Smith, I.L., Westcott, D., Bryden, W.L., Middleton, D., Reid, P.A., McFarlane, R.A., Martin, G., Tabor, G.M., Skerratt, L.F., Anderson, D.L., Crameri, G., Quammen, D., Jordan, D., Freeman, P., Wang, L.F., Epstein, J.H., Marsh, G.A., Kung, N.Y. and McCallum, H. 2015. Ecological dynamics of emerging bat virus spillover. *Proc. Biol. Sci.* **282**: 20142124.

25. Rahman, S.A., Hassan, S.S., Olival, K.J., Mohamed, M., Chang, L.Y., Hassan, L., Saad, N.M., Shohaimi, S.A., Mamat, Z.C., Naim, M.S., Epstein, J.H., Suri, A.S., Field, H.E. and Daszak, P.; Henipavirus Ecology Research Group. 2010. Characterization of Nipah virus from naturally infected *Pteropus vampyrus* bats, Malaysia. *Emerg. Infect. Dis.* **16**: 1990–1993.
Sendow, I., Field, H.E., Curran, J., Darminto, Morrissy, C., Meehan, G., Buick, T. and Daniels, P. 2006. Henipavirus in *Pteropus vampyrus* bats, Indonesia. *Emerg. Infect. Dis.* 12: 711–712.

Sendow, I., Ratnawati, A., Taylor, T., Adjid, R.M., Saepulloh, M., Barr, J., Wong, F., Daniels, P. and Field, H. 2013. Nipah virus in the fruit bat *Pteropus vampyrus* in Sumatera, Indonesia. *PLoS One* 8: e69544.

Smith, J.D. and Gregory. T.R. 2009. The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained. *Biol. Lett.* 5: 347–351.

Stevens, M. Hallsworth, P.B. and Smith, H.G. 2004. The effects of Beet mild yellowing virus and Beet chlorosis virus on the yield of UK field-grown sugar beet in 1997, 1999 and 2000. *Ann. Appl. Biol.* 144: 113–119.

Takemae, H., Basri, C., Mayasari, N.L.P.I., Tarigan, R., Shimoda, H., Omatsu, T., Supratikno, Pramono, D., Cahyadi, D.D., Kobayashi, R., Iida, K., Mizutani, T., Maeda, K., Agungpriyono, S. and Hondo, E. 2018. Isolation of Pteropine orthoreovirus from *Pteropus vampyrus* in Garut, Indonesia. *Virus Genes* 54: 823–827.

Tycowski, K.T., You, Z.H., Graham, P.J. and Steitz., J.A. 1998. Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell.* 2: 629–638.

Virgin, H.W., Wherry, E.J. and Ahmed, R. 2009. Redefining chronic viral infection. *Cell* 138: 30–50.

Wang, L.F. and Anderson, D.E. 2019. Viruses in bats and potential spillover to animals and humans. *Curr. Opin. Virol.* 34: 79–89.

Yi, Z., Sperzel, L., Nürnberg, C., Bredenbeek, P.J., Lubick, K.J., Best, S.M., Stoyanov, C.T., Law, L.M., Yuan, Z., Rice, C.M. and MacDonald, M.R. 2011. Identification and
characterization of the host protein DNAJC14 as a broadly active flavivirus replication modulator. *PLoS Pathog.* 7: e1001255.

35. Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., Ogata, H. and Takemura M. 2019. Medusavirus, a novel large DNA virus discovered from hot spring water. *J. Virol.* 93: e02130-18.
Figure 1. Distribution of viral sequences in *Pteropus vampyrus* specimens.

A. Graph indicating the distribution of viral species classifications in each specimen. B. Graph indicating the distribution of heritable factors for the viral species in each specimen. C. Graph indicating the distribution of the general viral species hosts of in each specimen. D. Filled columns indicate the viral homologous sequences identified in the specimen. Original data including all candidate virus names are listed in Supplementary Table 1. Five viruses were subjected to analysis to determine whether they were endogenous viral elements (Figures 2–4).
Figure 2. Comparison of Bracovirus and U6 spliceosomal small RNAs.

A. Sequence comparison of Bracovirus and U6 spliceosomal small RNAs from Pteropus vampyrus and other animals. Gray boxes indicate sequences identical to Bracovirus. Lined squares indicate functional motif forming spliceosome complex. B. Phylogenetic tree indicating the relationship between Bracovirus and U6 spliceosomal small RNA sequences in animal genomes. The viral sequence was predicted to be distant from other animal sequences.
Figure 3. Comparison of Bovine viral diarrhea virus (BVDV) and dnajc14.

A. Sequence comparison of BVDV and dnajc14 in Pteropus vampyrus and other animals. Gray boxes indicate sequences identical to BVDV. B. Illustration indicates vertebrate dnajc14 structure. In all species except elephant shark (Callorhinus milii), the 161-bp conserved sequence was found as a single exon. C. Phylogenetic tree indicating the relationship between the conserved BVDV sequences and dnajc14 in vertebrate genomes. Viral sequence was predicted to be the closest to bovine Bos taurus sequence.
Figure 4. Plant virus homologous sequences identified from Pteropus vampyrus.

A. Black line indicates whole Tomato yellow leaf curl virus (TYLCV) genome. Gray boxes under line indicate viral protein coding region. Short lines above viral genome indicate Next generation sequencing (NGS) reads identified in Institut Pertanian Bogor_4 (IPB_4) specimen. B. Black line indicates whole Beet mild yellowing virus (BMYV) genome. Gray boxes under line indicate viral protein coding region. Short lines above viral genome indicate overlap region of the NGS reads identified in IPB_41 specimen.
Supplementary information

Viral-derived DNA invasion and individual variation in an Indonesian population of large flying fox *Pteropus vampyrus*

Atsuo Iida, Hitoshi Takemae, Ronald Tarigan, Ryosuke Kobayashi, Hirokazu Kato, Hiroshi Shimoda, Tsutomu Omatsu, Supratikno, Chaerul Basri, Ni Luh Putu Ika Mayasari, Srihadi Agungpriyono, Ken Maeda, Tetsuya Mizutani, Eiichi Hondo

The supplementary information includes four Additional Figures, two Tables and six Data texts.
Supplementary Fig. 1. Experimental procedure to find viral sequences in wild bats

DNA samples were extracted from the liver tissue of six *Pteropus vampyrus* specimens collected in Indonesia. Next generation sequencing reads were obtained for each specimen. Homology was searched for all viral genome sequences by BLAST to identify the sequence reads homologous to the viral genome. The viral species distribution was estimated from the mapped reads (Figure 1). Reads mapped to the viral sequences were homology-searched by BLAST for *P. vampyrus* genome. Reads mapped to the viruses and genomes were predicted to be endogenous viral elements in the bat genome. Unmapped reads were predicted to be extrachromosomal virus-like sequences in bat liver cells. Typical examples of each trait are described in Figures 2–4.
Supplementary Fig. 2. Comparison of Pandoravirus dulcis and histone H2B

Sequence comparison of Pandoravirus dulcis and histone H2B from Pteropus vampyrus. Gray boxes indicate identical sequences.
| TYLCV genome | Sequence | Bat read |
|--------------|----------|----------|
| 1350         | ATGGAAATGATGTCGTGTTTACATTAGAAATGGCCGCTGGCTGTTCTGT | 1         | ATGGAAATGATGTCGTGTTTACATTAGAAATGGCCGCTGGCTGTTCTGT |
| 1400         | TATCTTTGAAATAGGGGGATTTATCTCCCACCAAAAGCGCCATTCT | 51        | TATCTTTGAAATAGGGGGATTTATCTCCCACCAAAAGCGCCATTCT |
| 1450         | CTCGCTTGAAGGAGCACTGAAGCTTCCCTGTGGCTGAATCCATGATTGT | 101       | CTCGCTTGAAGGAGCACTGAAGCTTCCCTGTGGCTGAATCCATGATTGT |

Supplementary Fig. 3. Tomato yellow leaf curl virus (TYLCV) homologous sequence identified in *Pteropus vampyrus*

Sequence comparison of TYLCV and mapped reads. Two sequences completely matched to plant virus were identified in the bat specimen.
| BMYV genome | 3573 | **CTACACCGTCTAGTATGTTATAGTACGCCTGCTGGTAGGAAC** |
| Bat contig | 1 | **CTCA CACGTAAGATCTATTTGCTGAGAGGAAC** |
| BMYV genome | 3623 | **GATCATAGGGGAAGAAGCGACACGCTGAGAAGCGAGACGCCTCAAGGCT** |
| Bat contig | 51 | **AATCAATGGAAGAAGCGACACGCTCAAGGCGAATTCAGGGA** |
| BMYV genome | 3673 | **CTCAAGCCGTCTCTCTCTCTCGAGAACACACCCGCAACCGAGGCTCCCT** |
| Bat contig | 101 | **ATCAAGCCGTCTCTCTCTCTCGAGAACACACCCGCAACCGAGGCTCCCT** |
| BMYV genome | 3723 | **AGACGAGACGAGAGAGAACCGAAGACGAGGCTGTTCTTACCA** |
| Bat contig | 151 | **AGACGAGACGAGAGAGAACCGAAGACGAGGCTGTTCTTACCA** |
| BMYV genome | 3773 | **AGGGACCATCTTCTCAGCAGAGGTTGTTCCCTCAAAAGACCAATCTCAGGG** |
| Bat contig | 201 | **AGGGACCATCTTCTCAGCAGAGGTTGTTCCCTCAAAAGACCAATCTCAGGG** |
| BMYV genome | 3823 | **GAAGTCTCCAGCGAGCAATCAGTCTCGGCGAGCTCTATCAGACTGCCC** |
| Bat contig | 251 | **GAAGTCTCCAGCGAGCAATCAGTCTCGGCGAGCTCTATCAGACTGCCC** |
| BMYV genome | 3873 | **GCATTCCGTCTAGGATGCTCAAGCCCTACATGATATAAATATCTTCAAT** |
| Bat contig | 301 | **GCATTCCGTCTAGGATGCTCAAGCCCTACATGATATAAATATCTTCAAT** |
| BMYV genome | 3923 | **GGTCAATTGGAGTTTGCCTTCTCGAGGCCTCTTCCAAAATTCGGGTCCA** |
| Bat contig | 351 | **GGTCAATTGGAGTTTGCCTTCTCGAGGCCTCTTCCAAAATTCGGGTCCA** |
| BMYV genome | 3973 | **TCGCTTACGAGCTGGACCGACACTCGTAACCTCAGTCGCCCTTCCTCAACC** |
| Bat contig | 401 | **TCGCTTACGAGCTGGACCGACACTCGTAACCTCAGTCGCCCTTCCTCAACC** |
| BMYV genome | 4023 | **ATTACAAAGTTGGAATCACAAA** |
| Bat contig | 451 | **ATTACAAAGTTGGAATCACAAA** |

**Supplementary Fig. 4.** Beet mild yellowing virus (BMYV) homologous sequence identified in *Pteropus vampyrus*

Sequence comparison of BMYB and contig constructed from the reads from the bat specimen.
| Accession   | Family               | Host     | Type    | Name                                                                 | IPB_1 | IPB_2 | IPB_3 | IPB_4 | IPB_41 | IPB_BG |
|------------|----------------------|----------|---------|----------------------------------------------------------------------|-------|-------|-------|-------|--------|--------|
| NC_001461.1 | Flaviviridae         | Mammal   | RNA     | Bovine viral diarrhea virus 1, complete genome                       |       |       |       |       |        |        |
| NC_001648.1 | Caulimoviridae       | Plant    | DNA     | Cassava vein mosaics virus, complete genome                          |       |       |       |       |        |        |
| NC_001650.2 | Herpesviridae        | Mammal   | DNA     | Equid herpesvirus 2 strain 86/67, complete genome                    |       |       |       |       |        |        |
| NC_001716.2 | Herpesviridae        | Mammal   | DNA     | Human herpesvirus 7, complete genome                                 |       |       |       |       |        |        |
| NC_001747.1 | Luteoviridae         | Plant    | RNA     | Potato leafroll virus, complete genome                               |       |       |       |       |        |        |
| NC_001828.1 | Geminiviridae        | Plant    | DNA     | Tomato mottle Taino virus DNA A, complete sequence                  |       |       |       |       |        |        |
| NC_001932.1 | Geminiviridae        | Plant    | DNA     | Indian cassava mosaic virus DNA A, complete sequence                |       |       |       |       |        |        |
| NC_001938.1 | Geminiviridae        | Plant    | DNA     | Tomato mottle virus DNA A, complete sequence                        |       |       |       |       |        |        |
| NC_002510.1 | Geminiviridae        | Plant    | DNA     | Cotton leaf curl Gezira virus, complete genome                      |       |       |       |       |        |        |
| NC_002555.1 | Geminiviridae        | Plant    | DNA     | Tomato rugose mosaic virus DNA A, complete sequence                |       |       |       |       |        |        |
| NC_002665.1 | Herpesviridae        | Mammal   | DNA     | Bovine herpesvirus 4 long unique region, complete                |       |       |       |       |        |        |
| NC_002766.1 | Luteoviridae         | Plant    | RNA     | Beet chlorosis virus, complete genome                               |       |       |       |       |        |        |
| NC_003378.1 | Geminiviridae        | Plant    | DNA     | Tobacco vein clearing virus, complete genome                       |       |       |       |       |        |        |
| NC_003401.1 | Herpesviridae        | Mammal   | DNA     | Macacine herpesvirus 5, genome                                     |       |       |       |       |        |        |
| NC_003491.1 | Luteoviridae         | Plant    | RNA     | Beet mild yellowing virus, complete genome                          |       |       |       |       |        |        |
| NC_003664.1 | Geminiviridae        | Plant    | DNA     | Tomato chlorotic mottle virus DNA A, complete sequence              |       |       |       |       |        |        |
| NC_003678.1 | Flaviviridae         | Mammal   | RNA     | Pestivirus giraffe-1 H138 complete genome                           |       |       |       |       |        |        |
| NC_003722.1 | Geminiviridae        | Plant    | DNA     | Tobacco curly shoot virus, complete genome                          |       |       |       |       |        |        |
| NC_003743.1 | Luteoviridae         | Plant    | RNA     | Turnip yellows virus, complete genome                               |       |       |       |       |        |        |
| NC_003861.1 | Geminiviridae        | Plant    | DNA     | Sri Lankan cassava mosaic virus-[Colombo] DNA-A,                   |       |       |       |       |        |        |
| NC_003897.1 | Geminiviridae        | Plant    | DNA     | Tomato leaf curl Karnataka virus, complete genome                   |       |       |       |       |        |        |
| NC_004005.1 | Geminiviridae        | Plant    | DNA     | Tomato yellow leaf curl virus, complete genome                      |       |       |       |       |        |        |
| NC_004044.1 | Geminiviridae        | Plant    | DNA     | Tomato yellow leaf curl China virus, complete genome               |       |       |       |       |        |        |
| NC_004147.1 | Geminiviridae        | Plant    | DNA     | Papaya leaf curl virus, complete genome                             |       |       |       |       |        |        |
| NC_004153.1 | Geminiviridae        | Plant    | DNA     | Tomato leaf curl Vietnam virus DNA A, complete genome              |       |       |       |       |        |        |
| NC_004300.1 | Geminiviridae        | Plant    | DNA     | Croton yellow vein mosaic virus, complete genome                    |       |       |       |       |        |        |
| NC_004569.1 | Geminiviridae        | Plant    | DNA     | Tomato yellow leaf curl Malaga virus, complete genome               |       |       |       |       |        |        |
| NC_004583.1 | Geminiviridae        | Plant    | DNA     | Cotton leaf curl Kokhran virus, complete genome                    |       |       |       |       |        |        |
| NC_004657.1 | Geminiviridae        | Plant    | DNA     | Sida golden mosaic Costa Rica virus DNA A, complete                |       |       |       |       |        |        |
| NC_004659.1 | Geminiviridae        | Plant    | DNA     | Sida golden mosaic Honduras virus DNA A, complete                 |       |       |       |       |        |        |
| NC_004661.1 | Geminiviridae        | Plant    | DNA     | Sida yellow vein virus DNA A, complete sequence                    |       |       |       |       |        |        |
| NC_004756.1 | Luteoviridae         | Plant    | RNA     | Beet western yellows virus, complete genome                         |       |       |       |       |        |        |
| NC_005040.1 | Paroviridae          | Insect   | DNA     | Myzus persicae densovirus, complete genome                         |       |       |       |       |        |        |
| NC_005179.1 | Poxviridae           | Mammal   | DNA     | Yaba monkey tumor virus, complete genome                            |       |       |       |       |        |        |
| NC_005319.1 | Geminiviridae        | Plant    | DNA     | Euphorbia leaf curl virus DNA A, complete sequence                 |       |       |       |       |        |        |
| NC_005320.1 | Geminiviridae        | Plant    | DNA     | Tomato leaf curl China virus - [G32], complete sequence           |       |       |       |       |        |        |
| Accession     | Family       | Host     | Type  | Description                                           |
|---------------|--------------|----------|-------|-------------------------------------------------------|
| NC_005347.1   | Geminiviridae| Plant    | DNA   | Pepper yellow vein Mali virus, complete genome        |
| NC_005842.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Iran virus, complete genome           |
| NC_005855.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Sudan virus - [Gezira], complete genome |
| NC_006874.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Madagascar virus-Menabe              |
| NC_008168.1   | Baculoviridae| Insect   | DNA   | Choristoneura fumiferana granulovirus, complete genome |
| NC_008236.1   | Geminiviridae| Plant    | DNA   | Siegesbeckia yellow vein virus-[GD13], complete genome |
| NC_008249.1   | Luteoviridae | Plant    | RNA   | Chickpea chlorotic stunt virus, complete genome       |
| NC_008299.1   | Geminiviridae| Plant    | DNA   | Pedilanthus leaf curl virus isolate Rahim Yar Khan 1   |
| NC_008317.1   | Geminiviridae| Plant    | DNA   | Siegesbeckia yellow vein Guangxi virus, complete genome |
| NC_008492.1   | Geminiviridae| Plant    | DNA   | Corchorus yellow spot virus DNA A, complete sequence  |
| NC_008724.1   | Phycodnaviridae| Chlorella| DNA   | Acanthocystis turfacea Chlorella virus 1, complete    |
| NC_009011.2   | Baculoviridae| Insect   | DNA   | Spodoptera frugiperda MNPV virus, complete genome      |
| NC_009546.1   | Geminiviridae| Plant    | DNA   | Mimosa yellow leaf curl virus DNA-A, complete sequence |
| NC_009607.1   | Geminiviridae| Plant    | DNA   | Tomato severe rugose virus DNA A, complete sequence    |
| NC_010809.1   | Luteoviridae | Plant    | RNA   | Melon aphid-borne yellows virus, complete genome       |
| NC_010837.1   | Geminiviridae| Plant    | DNA   | Blainvillea yellow spot virus DNA-A, complete genome   |
| NC_011096.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Kumasi virus segment A, complete      |
| NC_012206.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Nigeria virus-[Nigeria:2006], complete |
| NC_012492.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Patna virus DNA-A, complete genome    |
| NC_013017.1   | Geminiviridae| Plant    | DNA   | Okra leaf curl virus-[Cameroon], complete genome       |
| NC_013639.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Cameroon virus -                        |
| NC_013668.3   | Alloherpesviridae| Fish   | DNA   | Anguillid herpesvirus 1, complete genome               |
| NC_013668.8   | Alloherpesviridae| Fish   | DNA   | Anguillid herpesvirus 1, complete genome               |
| NC_014066.1   | Geminiviridae| Plant    | DNA   | Okra yellow mosaic Mexico virus DNA A, complete genome  |
| NC_014130.1   | Geminiviridae| Plant    | DNA   | Sida golden mottle virus DNA-A, complete genome         |
| NC_014138.1   | Geminiviridae| Plant    | DNA   | Abutilon Brazil virus DNA A segment, complete sequence |
| NC_014446.1   | Geminiviridae| Plant    | DNA   | Sida golden mosaic Florida virus-Malvastrum DNA-A,      |
| NC_014473.1   | Geminiviridae| Plant    | DNA   | Croton yellow vein virus, complete genome               |
| NC_014542.1   | Geminiviridae| Plant    | DNA   | Tomato leaf curl Oman virus, complete genome           |
| NC_014545.1   | Luteoviridae | Plant    | RNA   | Cotton leafroll dwarf virus, complete genome           |
| NC_014596.1   | Geminiviridae| Plant    | DNA   | Tobacco leaf curl Pusa virus DNA-A, complete genome     |
| NC_014645.1   | Geminiviridae| Plant    | DNA   | Ageratum leaf curl Cameroon virus, complete genome      |
| NC_014707.1   | Geminiviridae| Plant    | DNA   | Papaya leaf crumple virus-Panipat 8                     |
| NC_014745.1   | Geminiviridae| Plant    | DNA   | Okra leaf curl Cameroon virus, complete genome          |
| NC_015043.1   | Geminiviridae| Plant    | DNA   | Sida mosaic Bolivia virus 2 DNA-A, complete genome      |
| NC_015228.1   | Caulimoviridae| Plant   | DNA   | Sweet potato vein clearing virus, complete genome       |
| NC_015323.1   | Adenoviridae | Bird     | DNA   | Fowl adenovirus C, complete genome                      |
| NC_016038.2   | Luteoviridae | Plant    | RNA   | Brassica yellows virus isolate BrYY-ABJ, complete genome |
| NC_016072.1   | Miniviridae  | Ameba    | DNA   | Megavirus chiliensis, complete genome                   |
| Accession | Family       | Host  | Type  | Description                                                                 |
|-----------|--------------|-------|-------|-----------------------------------------------------------------------------|
| NC_016574.1 | Geminiviridae | Plant | DNA   | Abutilon mosaic Brazil virus DNA-A, complete genome                         |
| NC_016578.1 | Geminiviridae | Plant | DNA   | Cleome leaf crumple virus DNA-A, complete genome                            |
| NC_016965.1 | Geminiviridae | Plant | DNA   | Tomato leaf curl Ranchi virus DNA-A, complete genome                        |
| NC_016998.1 | Geminiviridae | Plant | DNA   | Centrosema yellow spot virus DNA-A, complete genome                         |
| NC_016999.1 | Geminiviridae | Plant | DNA   | Macroptilium yellow spot virus DNA A, complete sequence                      |
| NC_017000.1 | Geminiviridae | Plant | DNA   | Macroptilium yellow vein virus DNA-A, complete genome                        |
| NC_017001.1 | Geminiviridae | Plant | DNA   | Macroptilium yellow net virus DNA A, complete genome                        |
| NC_018457.1 | Geminiviridae | Plant | DNA   | Soybean chlorotic spot virus DNA-A, complete sequence                       |
| NC_018464.1 | Peribunyaviridae | Insect | RNA   | Shimbavirus N and NSs genes, segment S, genomic RNA                         |
| NC_018476.1 | Peribunyaviridae | Mammal | RNA   | Simbu virus RdRp gene for RNA-dependent RNA polymerase,                      |
| NC_018571.2 | Luteoviridae | Plant | RNA   | Suakwa aphid-borne yellows virus isolate SABYV-TW19,                       |
| NC_019491.1 | Alloherpesviridae | Fish | DNA   | Cyprinid herpesvirus 1 strain NG-J1, complete genome                        |
| NC_020256.1 | Geminiviridae | Plant | DNA   | Sida mottle Alagoas virus isolate BR:Vsa2:10 segment                        |
| NC_020257.1 | Geminiviridae | Plant | DNA   | Tomato rugose yellow leaf curl virus isolate U2 clone 1                     |
| NC_021205.1 | Geminiviridae | Plant | DNA   | Tomato leaf curl China virus - OX2, complete genome                         |
| NC_021858.1 | Unclassified | Ameba | DNA   | Pandoravirus dulcis, complete genome                                        |
| NC_022005.1 | Geminiviridae | Plant | DNA   | Bean white chlorosis mosaic virus isolate                                   |
| NC_022098.1 | Unclassified | Ameba | DNA   | Pandoravirus salinus, complete genome                                        |
| NC_022229.1 | Geminiviridae | Plant | DNA   | Tomato yellow leaf curl Saudi virus isolate HalI1,                          |
| NC_023034.1 | Geminiviridae | Plant | DNA   | Tomato leaf curl Gandhinagar virus isolate pToGNA15                          |
| NC_023639.1 | Miniviridae   | Ameba | DNA   | Mimivirus terra2 genome                                                     |
| NC_024116.1 | Geminiviridae | Plant | DNA   | Tomato leaf curl Barka virus isolate Tom-55, complete                       |
| NC_024304.1 | Geminiviridae | Plant | DNA   | Solanum mosaic Bolivia virus segment DNA-A, complete                        |
| NC_024450.1 | Herpesviridae | Bird   | DNA   | Falconid herpesvirus 1 strain S-18, complete genome                         |
| NC_024458.1 | Alphaflexiviridae | Plant | RNA   | Pitaya virus X isolate P37, complete genome                                 |
| NC_025680.1 | Geminiviridae | Plant | DNA   | Tomato yellow leaf curl Aphanarquus virus isolate Homra,                     |
| NC_025725.1 | Geminiviridae | Plant | DNA   | Jatropha leaf crumple virus isolate SK1, complete                           |
| NC_026141.2 | Polyomaviridae | Bird   | DNA   | Adelie penguin polyomavirus isolate AdPyV_Crozier_2012,                     |
| NC_026421.1 | Herpesviridae | Mammal | DNA   | Equid herpesvirus 5 strain 2-141/67, complete genome                        |
| NC_026440.1 | Unclassified | Ameba | DNA   | Pandoravirus inopinatum isolate KlaHel, complete genome                     |
| NC_026763.1 | Geminiviridae | Plant | DNA   | Malvastrum yellow vein Cambodia virus isolate O8 segment                     |
| NC_026922.1 | Baculoviridae | Insect | DNA   | Lambdina fiscellaria nucleopolyhedrovirus isolate GR15,                     |
| NC_027128.1 | Iflaviridae   | Insect | DNA   | La Jolla virus putative polyprotein gene, complete cdfs                     |
| NC_027215.1 | Geminiviridae | Plant | DNA   | Tomato yellow leaf curl Mal outbreak virus complete genome                  |
| NC_028045.1 | Circoviridae  | Mammal | DNA   | Tadarida brasilensis circovirus 1, complete genome                          |
| NC_028094.1 | Phycodnaviridae | Algae | DNA   | Chrysochromulina ericina virus isolate CeV-01B, complete                   |
| NC_028124.1 | Geminiviridae | Plant | DNA   | Vinca leaf curl virus isolate RK, complete sequence                         |
| NC_028140.1 | Geminiviridae | Plant | DNA   | Melochia mosaic virus DNA A, complete sequence                              |
| NC_028985.1 | Geminiviridae | Plant | DNA   | Okra leaf curl Oman virus, complete genome, isolate OK-2                   |
| Accession    | Family               | Kingdom  | Type   | Description                                                                 | Genbank ID                  |
|--------------|----------------------|----------|--------|-----------------------------------------------------------------------------|----------------------------|
| NC_029105.1  | Geminiviridae        | Plant    | DNA    | Pavonia yellow mosaic virus isolate BR-Alb51-14 segment                      | NC_029105.1                |
| NC_029906.1  | Luteoviridae         | Plant    | RNA    | Barley virus G isolate Gimje, complete genome                               | NC_029906.1                |
| NC_030656.1  | Poxviridae           | Mammal   | DNA    | Pteropox virus strain Australia, complete genome                            | NC_030656.1                |
| NC_031306.1  | Circoviridae         | unknown  | DNA    | Synedrella leaf curl virus isolate YN3306, complete                         | NC_031306.1                |
| NC_031448.1  | Circoviridae         | unknown  | DNA    | Macroptilium common mosaic virus isolate ALM2_5B segment                     | NC_031448.1                |
| NC_032111.1  | Poxviridae           | Mammal   | DNA    | BeAn 58058 virus, complete genome                                           | NC_032111.1                |
| NC_033774.1  | Tospoviridae         | Plant    | RNA    | Pepper chlorotic spot virus isolate 14YV733 segment L                       | NC_033774.1                |
| NC_034246.1  | Luteoviridae         | Plant    | RNA    | Cowpea polerovirus 1 isolate BE167, complete genome                         | NC_034246.1                |
| NC_034265.1  | Luteoviridae         | Plant    | RNA    | Tobacco virus 2, complete genome                                            | NC_034265.1                |
| NC_034266.1  | Herpesviridae        | Bird     | DNA    | Columbid alphaherpesvirus 1 strain HLJ, complete genome                     | NC_034266.1                |
| NC_035193.1  | Papillomaviridae     | Mammal   | DNA    | Ailuropoda melanoleuca papillomavirus 4 strain gpaam004,                    | NC_035193.1                |
| NC_035473.1  | Geminiviridae        | Plant    | DNA    | Emilia sonchifolia yellow vein Thailand virus isolate                       | NC_035473.1                |
| NC_035481.1  | Geminiviridae        | Plant    | DNA    | Tomato leaf curl purple vein virus isolate BR:793:15                        | NC_035481.1                |
| NC_036582.1  | Poxviridae           | Bird     | DNA    | Flamingopox virus FGPV/09, complete genome                                  | NC_036582.1                |
| NC_037665.1  | unclassified         | Ameba    | DNA    | Pandoravirus maculodenis, complete genome                                   | NC_037665.1                |
| NC_037666.1  | unclassified         | Ameba    | DNA    | Pandoravirus neocaledonia, complete genome                                  | NC_037666.1                |
| NC_037667.1  | unclassified         | Ameba    | DNA    | Pandoravirus quercus, complete genome                                       | NC_037667.1                |
| NC_038438.1  | Geminiviridae        | Plant    | DNA    | Abutilon golden mosaic Yucatan virus isolate Conkai-2007                   | NC_038438.1                |
| NC_038440.1  | Geminiviridae        | Plant    | DNA    | Chenopodium leaf curl virus, complete genome                                | NC_038440.1                |
| NC_038444.1  | Geminiviridae        | Plant    | DNA    | Cotton leaf curl Gezira virus-Cotton A component,                           | NC_038444.1                |
| NC_038449.1  | Geminiviridae        | Plant    | DNA    | Jatropha mosaic India virus-[Lucknow] strain SK-2                           | NC_038449.1                |
| NC_038452.1  | Geminiviridae        | Plant    | DNA    | Malvastrum yellow mosaic Helshire virus isolate Ma179A5                     | NC_038452.1                |
| NC_038458.1  | Geminiviridae        | Plant    | DNA    | Sida golden mosaic Braco virus-[Jamaica:Liguanea:2008]                      | NC_038458.1                |
| NC_038459.1  | Geminiviridae        | Plant    | DNA    | Sida golden mosaic Lara virus strain                                        | NC_038459.1                |
| NC_038460.1  | Geminiviridae        | Plant    | DNA    | Sida golden yellow vein virus-[A11], complete genome                       | NC_038460.1                |
| NC_038461.1  | Geminiviridae        | Plant    | DNA    | Sida yellow leaf curl virus isolate BR:Coi3:07 segment                      | NC_038461.1                |
| NC_038469.1  | Geminiviridae        | Plant    | DNA    | Tomato interveinal chlorosis virus isolate                                 | NC_038469.1                |
| NC_038470.1  | Geminiviridae        | Plant    | DNA    | Tobacco leaf curl New Delhi virus 2, complete genome                        | NC_038470.1                |
| NC_038500.1  | Genomoviridae        | Mammal   | DNA    | UNVERIFIED: Rhinolophus associated gemykbivirus 2                           | NC_038500.1                |
| NC_038509.1  | Luteoviridae         | Plant    | RNA    | Groundnut rosette assistor virus clone N1S4CP coat protein gene, complete cds| NC_038509.1                |
| NC_038798.1  | Geminiviridae        | Plant    | DNA    | Malvastrum yellow mosaic Jamaica virus isolate Ma179A73                    | NC_038798.1                |
| NC_038805.1  | Geminiviridae        | Plant    | DNA    | Rhynchosia rugose golden mosaic                                             | NC_038805.1                |
| NC_038890.1  | Geminiviridae        | Plant    | DNA    | Tobacco leaf rugose virus DNA-A, complete sequence                          | NC_038890.1                |
| NC_038962.1  | Geminiviridae        | Plant    | DNA    | Jatropha leaf curl Gujarat virus, complete genome                           | NC_038962.1                |
| NC_038963.1  | Geminiviridae        | Plant    | DNA    | Tomato latent virus isolate T2+, complete genome                            | NC_038963.1                |
| NC_038967.1  | Geminiviridae        | Plant    | DNA    | VEM begomovirus 1 isolate GtSq11 segment DNA-A, complete                    | NC_038967.1                |
| NC_038968.1  | Geminiviridae        | Plant    | DNA    | VEM begomovirus 2 isolate GtSq5 segment DNA-A, complete                     | NC_038968.1                |
| NC_038969.1  | Geminiviridae        | Plant    | DNA    | VEM begomovirus 3 isolate GtSq10 segment DNA-A, complete                    | NC_038969.1                |
| NC_038970.1  | Geminiviridae        | Plant    | DNA    | VEM begomovirus 4 isolate GtSq8 segment DNA-A, complete                     | NC_038970.1                |
| Accession    | Family         | Host       | Type   | Sample Description                                                                 |
|-------------|----------------|------------|--------|-----------------------------------------------------------------------------------|
| NC_038972.1 | Geminiviridae  | Plant      | DNA    | VEM begomovirus 7 isolate Sp5-4, complete sequence                                 |
| NC_038981.1 | Geminiviridae  | Plant      | DNA    | Pavonia mosaic virus isolate BR-Cor40-14 segment DNA-A,                            |
| NC_038983.1 | Geminiviridae  | Plant      | DNA    | Triumfetta yellow mosaic virus isolate BR-Msj1-10,                                 |
| NC_038988.1 | Geminiviridae  | Plant      | DNA    | Common bean mottle virus isolate CU/Mayabeque 6/2014,                              |
| NC_038992.1 | Geminiviridae  | Plant      | DNA    | Sida golden yellow spot virus isolate BR:Sab889:10,                                |
| NC_040699.1 | Nudiviridae    | Insect     | DNA    | Drosophila innubila nudivirus isolate DINV_CH01M,                                  |
| NC_043122.1 | Geminiviridae  | Plant      | DNA    | Tomato golden leaf distortion virus isolate TO45 coat                               |
| NC_043176.1 | Hantaviridae   | Mammal     | RNA    | Oxbow virus strain Ng1453 glycoprotein gene, complete                              |
| NC_043223.1 | unclassified   | Mammal     | DNA    | Senegalvirus SSV-A contig6 genomic sequence                                         |
| NC_043329.1 | Polydnaviridae | Insect     | DNA    | Diolcogaster facetosa bracovirus segment 29, complete                              |
| NC_043531.1 | Geminiviridae  | Plant      | DNA    | Datura leaf curl virus isolate Sudan-Datura 435-2016,                              |
| Name                                                                 | Accession       | Source | Species            |
|----------------------------------------------------------------------|-----------------|--------|--------------------|
| Diolcogaster facetosa bracovirus                                      | NC_043329.1     | NCBI   |                    |
| Pandoravirus dulcis, complete genome                                  | NC_021858.1     | NCBI   |                    |
| Bovine viral diarrhea virus 1, complete genome                        | NC_001461.1     | NCBI   |                    |
| Tomato yellow leaf curl virus, complete genome                         | NC_004005.1     | NCBI   |                    |
| Beet mild yellowing virus, complete genome                            | NC_003491.1     | NCBI   |                    |
| U6 small nuclear 1 (RNU6-1), small nuclear RNA                        | NR_004394.1     | NCBI   | Homo sapiens       |
| U6 small nuclear RNA (Rnu6), small nuclear RNA                        | NR_003027.2     | NCBI   | Mus musculus       |
| U6 spliceosomal RNA (LOC111730015), ncRNA                             | XR_002776507.1  | NCBI   | Pteropus vampyrus  |
| U6 spliceosomal RNA (LOC112530463), ncRNA                             | XR_003072101.1  | NCBI   | Gallus gallus     |
| U6 spliceosomal RNA (LOC111948540), ncRNA                             | XR_002874553.1  | NCBI   | Oryzias latipes    |
| U6 spliceosomal RNA (LOC116408566), ncRNA                             | XR_004221048.1  | NCBI   | Xenopus tropicalis |
| U6 spliceosomal RNA (LOC113474324), ncRNA                             | XR_003395988.1  | NCBI   | Ciona intestinalis |
| small nuclear RNA U6 at 96A a (snRNA:U6:96Aa), snRNA                 | NR_002081.1     | NCBI   | Drosophila melanogaster |
| novel gene (H2B clustered histone 4-like)                             | ENSPVAG00000017602 | Ensembl | Pteropus vampyrus |
| DNAJC14                                                              | ENSG00000135392  | Ensembl | Homo sapiens       |
| novel gene (dnajc14-like)                                             | ENSPVAG00000017755 | Ensembl | Pteropus vampyrus |
| dnajc14                                                              | ENSDARG000000105398 | Ensembl | Danio rerio       |
| dnajc14                                                              | ENSLACG00000018708 | Ensembl | Latimeria chalumnae |
| novel gene (dnajc14-like)                                             | ENSCMIG0000001493 | Ensembl | Callorhinchus milii |
| dnajc14                                                              | ENSCMIG0000002561 | Ensembl | Callorhinchus milii |
| DNAJC14                                                              | ENSBTAG00000020664 | Ensembl | Bos Taurus        |
| DNAJC14                                                              | ENSSSSCG0000000366 | Ensembl | Sus scrofa        |
| Symbol            | Ensembl ID                     | Database  | Species                  |
|-------------------|--------------------------------|-----------|--------------------------|
| Dnajc14           | ENSMUSG00000025354             | Ensembl   | Mus musculus             |
| dnajc14           | ENSXETG00000021854             | Ensembl   | Xenopus tropicalis       |
| novel gene (dnajc14-like) | ENSACAG00000015368         | Ensembl   | Anolis carolinensis     |
| novel gene (dnajc14-like) | ENSGALG000000037186     | Ensembl   | Gallus gallus           |
| dnajc14           | ENSORLG00000015564             | Ensembl   | Oryzias latipes         |