




A Novel *Polycipiviridae* Virus Identified in *Pteropus lylei* Stools

Sarah Temmam,^a Vibol Hul,^b  Thomas Bigot,^{a,c} Thavry Hoem,^d Christopher Gorman,^b Veasna Duong,^b Philippe Dussart,^b Julien Cappelle,^{d,e,f} Marc Eloit^{a,g}

^aInstitut Pasteur, Biology of Infection Unit, Inserm U1117, Pathogen Discovery Laboratory, Paris, France

^bVirology Unit, Institut Pasteur du Cambodge, Institut Pasteur International Network, Phnom Penh, Cambodia

^cInstitut Pasteur, Bioinformatics and Biostatistics Hub (C3BI), USR 3756 IP CNRS, Paris, France

^dEpidemiology and Public Health Unit, Institut Pasteur du Cambodge, Institut Pasteur International Network, Phnom Penh, Cambodia

^eUMR ASTRE, CIRAD, INRA, Université de Montpellier, Montpellier, France

^fUMR EpiA, VetAgro Sup, INRA, Marcy l'Etoile, France

^gEcole Nationale Vétérinaire d'Alfort, Université Paris-Est, Maisons-Alfort, France

ABSTRACT *Polycipiviridae* is a recently recognized viral family within the order *Picornavirales* with unusual genome organization and phylogenetic placement. Viruses belonging to this family were only reported from arthropod hosts. We describe here the first full genome of a distant polycipivirus-related virus identified in frugivorous bat stools in Cambodia.

Picornavirales consists of nonenveloped viruses characterized by a positive-sense nonsegmented single-stranded RNA (ssRNA) genome and a polyprotein gene expression strategy in which the structural protein module codes for three capsid domains and the nonstructural module codes for the viral helicase and RNA-dependent RNA polymerase (RdRP) (1). Knowledge about picornavirus host range, geographical distribution, and genome organization has exploded due to the democratization of high-throughput sequencing and the identification of novel picorna-like viruses in diverse samples (2). New picornaviruses with a polycistronic genome organization were recently reported in arthropods; *Polycipiviridae* consists of monopartite genomes of 11 kb with four open reading frames (ORFs) in the 5' region (coding for the structural proteins), followed by an intergenic region and a single ORF coding for the replicase complex (3).

Bats are a major mammalian reservoir of viruses (4). Recent metagenomic studies have highlighted the unexpected diversity of viral communities in bats (5, 6). Bat-associated picornaviruses were reported in *Picornaviridae* (e.g., bat kobivirus, hepatovirus, and mischivirus); *Iflaviridae* (bat iflavirus) and *Dicistroviridae* (bat cripavirus), possibly representing a passive carriage through food; and in unassigned groups (e.g., bat-associated posalivirus, fislavirus, felisavirus, and dicibavirus) (7, 8). We report here the characterization of the full-genome sequence of the first bat-associated polycipivirus.

A total of 214 *Pteropus lylei* rectal swabs were collected between May 2015 and July 2016 in Kandal Province, Cambodia. Bats were captured using mistnets; handling and sampling were conducted following the FAO guidelines (9). Swabs were pooled and clarified at 10,000 × *g* for 15 min before ultracentrifugation at 100,000 × *g* for 1 h. Total nucleic acids were extracted from the resuspended pellet with the QIAamp cadior pathogen mini kit (Qiagen, Courtaboeuf, France) according to the manufacturer's recommendations, except that carrier RNA was substituted by linear acrylamide (Life Technologies, Courtaboeuf, France). DNA was digested with the Turbo DNase reagent (Ambion, Life Technologies). Total RNA was further purified with the RNeasy cleanup protocol (Qiagen) and used as the template for next-generation sequencing (NGS)

Citation Temmam S, Hul V, Bigot T, Hoem T, Gorman C, Duong V, Dussart P, Cappelle J, Eloit M. 2019. A novel *Polycipiviridae* virus identified in *Pteropus lylei* stools. *Microbiol Resour Announc* 8:e01662-18. <https://doi.org/10.1128/MRA.01662-18>.

Editor Jelle Matthijssens, KU Leuven

Copyright © 2019 Temmam et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Marc Eloit, marc.eloit@pasteur.fr.

Received 18 December 2018

Accepted 1 March 2019

Published 11 April 2019

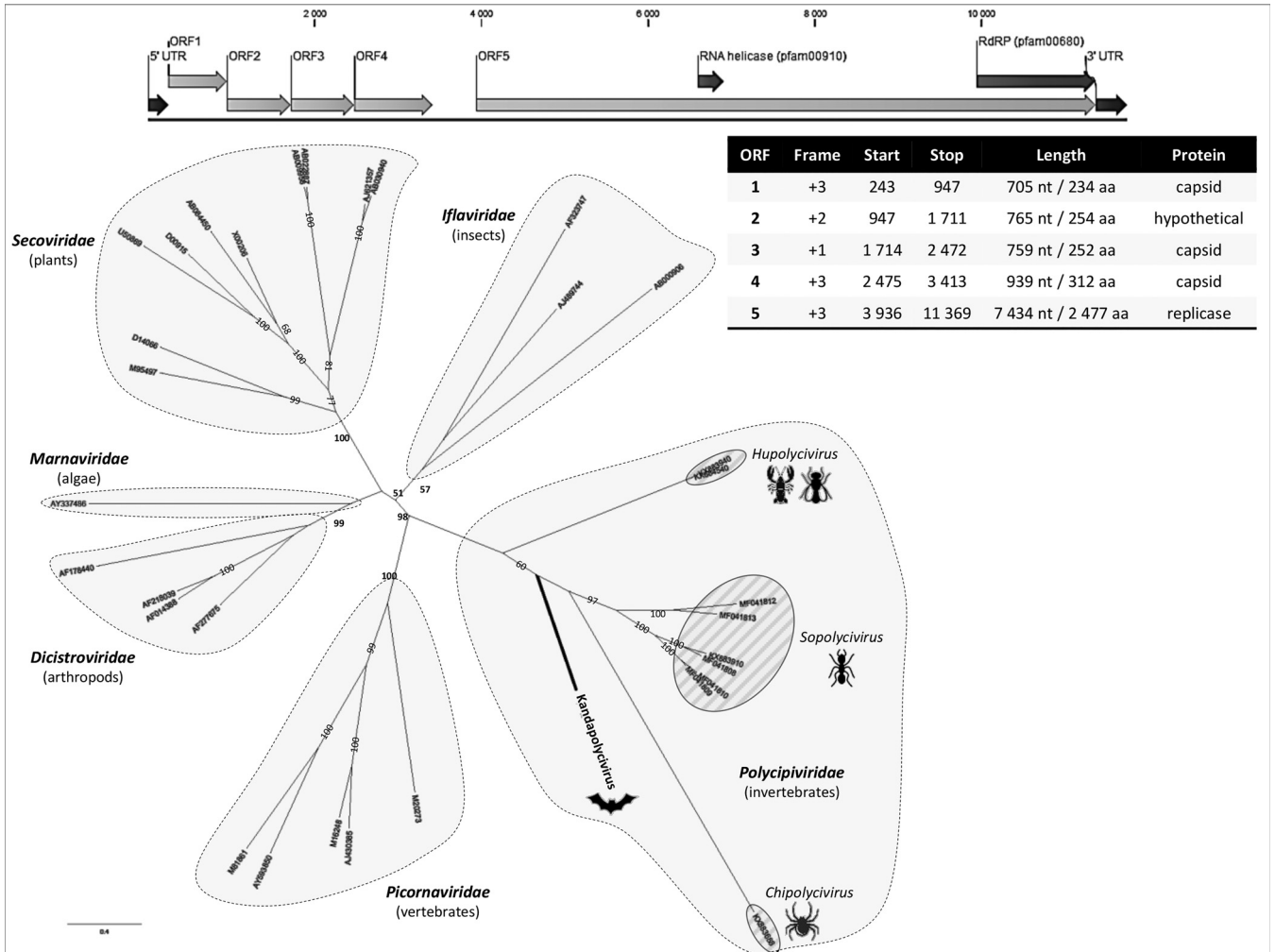


FIG 1 Genome organization of *Kandapolycivirus*, phylogenetic analysis of polycipiviruses, and representative members of the *Picornavirales* order. RNA-dependent RNA polymerase domains were retrieved from Koonin et al. (2) and Olendraite et al. (3) with corresponding accession numbers presented on the tree. Complete amino acid sequences were aligned with MAFFT with the L-INS-I parameter (13). The best amino acid substitution models that fitted the data were determined with ATGC start model selection (14) implemented in PhyML with smart model selection (www.atgc-montpellier.fr/phyml-sm/) using the corrected Akaike information criterion. Phylogenetic trees were constructed using the maximum likelihood (ML) method implemented through the RAxML program under the CIPRES Science Gateway portal (15) according to the selected substitution model. Nodal support was evaluated using the “automatic bootstrap replicates” parameter. Supported nodes (i.e., with bootstrap values above 50) are represented, and bold type indicates nodes defining a family.

library preparation using the SMARTer stranded total RNA-seq kit v2, pico input mammalian (TaKaRa Bio, Saint-Germain-en-Laye, France). Libraries were sequenced in a 2 × 75-bp format on a NextSeq 500 sequencer to produce 45.7 million reads. An in-house bioinformatics pipeline comprised quality check and trimming (AlienTrimmer package [10]), *de novo* assembly (Megahit tool [11]), ORF prediction (https://figshare.com/articles/translateReads_py/7588592), and a sequence search against the protein reference viral database (12; <https://rvdb-prot.pasteur.fr>), followed by the verification that nothing else but viruses were found as better hits when the sequences were subjected to a BLAST search against the whole NCBI/nonredundant (nr) protein database.

A large single contig of 11,745 bp with low amino acid identity to *Polycipiviridae* viruses was obtained. With an average coverage of >3,900× and more than 600,000 reads, this novel virus (tentatively named *Kandapolycivirus*) has a G+C content of 39.25% and the classical genome organization of polycipiviruses, namely, four ORFs in the 5′ part of the genome, among which ORF1 (234 amino acids [aa]), ORF3 (252 aa), and ORF4 (312 aa) code for capsid-like proteins, and a large ORF (ORF5; 2,477 aa) in the

3'-coding region for the replicase module, with RNA helicase and RdRP domains (Fig. 1). ORF2 codes for a protein of 255 aa of unknown function but has several O-glycosylation sites, possibly constituting the fourth capsid ORF that is characteristic of *Polycipiviridae*. Phylogenetic analyses performed on the complete RdRP domain of polycipiviruses and representative *Picornavirales* viruses places *Kandapolycivirus* in the *Polycipiviridae* clade (Fig. 1). Interestingly, *Kandapolycivirus* locates in a distinct putative genus from the *Chipolycivirus* (arachnid-associated viruses), the *Sopolycivirus* (ant-specific viruses), and the *Hupolycivirus* (crustacean- and insect-associated viruses) genera.

Data availability. The genome sequence of *Kandapolycivirus* was deposited in GenBank under accession number [MK161350](https://doi.org/10.1093/mra/mk161350). Raw data corresponding to the *Kandapolycivirus* genome were deposited into the NCBI SRA database under the accession number [PRJNAS16387](https://doi.org/10.1093/mra/prjnas16387).

ACKNOWLEDGMENTS

We thank the agents of the Forestry Administration of Cambodia, Ministry of Agriculture, Forestry and Fisheries for their supervision and help during the capture and sampling of bats.

This work was supported by Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant number ANR-10-LABX-62-IBEID). The collection of biological samples was undertaken in the framework of the ComAcross project with financial support from the European Union (EuropeAid, INNOVATE contract 315-047).

REFERENCES

- Knowles NJ, Hovi T, Hyypiä T, King AMQ, Lindberg AM, Pallansch MA, Palmenberg AC, Simmonds P, Skern T, Stanway G, Yamashita T, Zell R. 2012. *Picornaviridae*, p 855–880. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds), *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, San Diego, CA.
- Koonin EV, Wolf YI, Nagasaki K, Dolja VV. 2008. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat Rev Micro* 6:925–939. <https://doi.org/10.1038/nrmicro2030>.
- Olendraite I, Lukhovitskaya NI, Porter SD, Valles SM, Firth AE. 2017. *Polycipiviridae*: a proposed new family of polycistronic picorna-like RNA viruses. *J Gen Virol* 98:2368–2378. <https://doi.org/10.1099/jgv.0.000902>.
- Hayman DTS. 2016. Bats as viral reservoirs. *Annu Rev Virol* 3:77–99. <https://doi.org/10.1146/annurev-virology-110615-042203>.
- Zheng X-Y, Qiu M, Guan W-J, Li J-M, Chen S-W, Cheng M-J, Huo S-T, Chen Z, Wu Y, Jiang L-N, Chen Q. 2018. Viral metagenomics of six bat species in close contact with humans in southern China. *Arch Virol* 163:73–88. <https://doi.org/10.1007/s00705-017-3570-3>.
- Geldenhuis M, Mortlock M, Weyer J, Bezuidt O, Seaman ECJ, Kearney T, Gleasner C, Erkkila TH, Cui H, Markotter W. 2018. A metagenomic viral discovery approach identifies potential zoonotic and novel mammalian viruses in *Neoromicia* bats within South Africa. *PLoS One* 13:e0194527. <https://doi.org/10.1371/journal.pone.0194527>.
- Yinda CK, Zell R, Deboutte W, Zeller M, Conceição-Neto N, Heylen E, Maes P, Knowles NJ, Ghogomu SM, Van Ranst M, Matthijssens J. 2017. Highly diverse population of *Picornaviridae* and other members of the *Picornavirales*, in Cameroonian fruit bats. *BMC Genomics* 18:249. <https://doi.org/10.1186/s12864-017-3632-7>.
- Oude Munnink BB, Phan MVT, VIZIONS Consortium, Simmonds P, Koopmans MPG, Kellam P, van der Hoek L, Cotten M. 2017. Characterization of Posa and Posa-like virus genomes in fecal samples from humans, pigs, rats, and bats collected from a single location in Vietnam. *Virus Evol* 3:vex022. <https://doi.org/10.1093/ve/vex022>.
- Food and Agriculture Organization of the United Nations. 2011. Investigating the role of bats in emerging zoonoses: balancing ecology, conservation and public health interests. In Newman SH, Field HE, de Jong CE, Epstein JH (eds), *FAO animal production and health manual no. 12*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Crisuolo A, Brisse S. 2014. AlienTrimmer removes adapter oligonucleotides with high sensitivity in short-insert paired-end reads. Commentary on Turner (2014) assessment of insert sizes and adapter content in FASTQ data from NexteraXT libraries. *Front Genet* 5:130. <https://doi.org/10.3389/fgene.2014.00130>.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
- Goodacre N, Aljanahi A, Nandakumar S, Mikailov M, Khan AS. 2018. A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *mSphere* 3:e00069-18. <https://doi.org/10.1128/mSphereDirect.00069-18>.
- Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbx108>.
- Lefort V, Longueville J-E, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol* 34:2422–2424. <https://doi.org/10.1093/molbev/msx149>.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p 1–8. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*. IEEE, Piscataway, NJ.