Genome Sequence of Ruloma Virus, a Novel Paramyxovirus Clustering Basally to Members of the Genus *Jeilongvirus*

Bert Vanmechelen,ᵃ Sien Meurs,ᵃ Zafeiro Zisi,ᵃ Joëlle Goüy de Bellocq,ᵇ Magda Bletsa,ᵃ Philippe Lemey,ᵃ Piet Maesᵃ

ᵃKU Leuven–University of Leuven, Department of Microbiology, Immunology and Transplantation, Laboratory of Clinical and Epidemiological Virology, Rega Institute for Medical Research, Leuven, Belgium
ᵇThe Czech Academy of Sciences, Institute of Vertebrate Biology, Brno, Czech Republic

ABSTRACT   We report here the complete genome sequence of ruloma virus, a novel paramyxovirus detected in a Machangu's brush-furred rat from Tanzania. Ruloma virus has the longest orthoparamyxovirus genome reported to date and forms a sister clade to all currently known members of the genus *Jeilongvirus*.

The recently established genus *Jeilongvirus* within the subfamily *Orthoparamyxovirinae* contains seven recognized species, all identified in rodents and bats (1). However, several more putative species have been reported, also in other mammals, indicating that they might be more prevalent and have a wider host range than previously assumed (2–4). Jeilongviruses are set apart from other paramyxoviruses by the presence of one or two additional genes (compared to the traditional N-P/V/C-M-F-G-L genome organization), encoding transmembrane proteins, between the F and G genes. Additionally, members of the rodent subclade of the genus *Jeilongvirus* are characterized by the long but variable length of their G genes and their overall large genome size (5).

We performed next-generation sequencing on cDNA from kidney and spleen tissue of a Machangu's brush-furred rat (*Lophuromys machangui*) caught by snap-trapping in Tanzania (9°02′26.7″S, 33°34′23.5″E) in 2013 (see also reference 6). Prior to Illumina sequencing, the tissue was subjected to total RNA extraction using the RNeasy minikit (Qiagen), with additional freeze-thaw steps before and after tissue homogenization and with on-column DNase treatment (see reference 6 for a more detailed description). The RNA Quantifluor system (Promega) and a Bioanalyzer 2100 device, using an Agilent RNA 6000 Nano chip (Agilent Technologies), were used for quantitation and quality assessment. Next, the Ribo-Zero rRNA removal kit (Illumina) was used to deplete rRNA, and sequencing libraries were prepared with the NEXTflex rapid Illumina directional RNA-Seq library prep kit (PerkinElmer). Paired-end sequencing (2 × 150 bp) was performed on an Illumina NextSeq 500 instrument (Illumina) at Viroscan3D (Lyon, France). Following adapter trimming using Trimmomatic v0.36, CLC Genomics Workbench v10.0.1 was used for *de novo* assembly. Paramyxoviral contigs were identified using a tBLASTx v2.5.0+ search against a custom database containing all recognized and putative paramyxovirus species for which complete genomes are available. Based on these contigs (average coverage depth, 8×), primers were designed to determine the complete genome sequence by reverse transcriptase PCR (RT-PCR) and subsequent Sanger sequencing, and 5′ and 3′ rapid amplification of cDNA ends (RACE) using the 5′/3′ RACE kit (Roche) was used to complete the genome ends. Genome annotation was performed using the NCBI ORFfinder (www.ncbi.nlm.nih.gov/orffinder) and TMHMM Server v2.0 (www.cbs.dtu.dk/services/TMHMM) for transmembrane protein prediction. All tools were run with default parameters unless otherwise specified.

Based on its location (Rungwe District, Tanzania) and host of origin (*Lophuromys machangui*), this virus was given the name ruloma virus. Its genome consists of 20,454 nucleotides, in compliance with the "rule of six" for paramyxovirus genomes, and
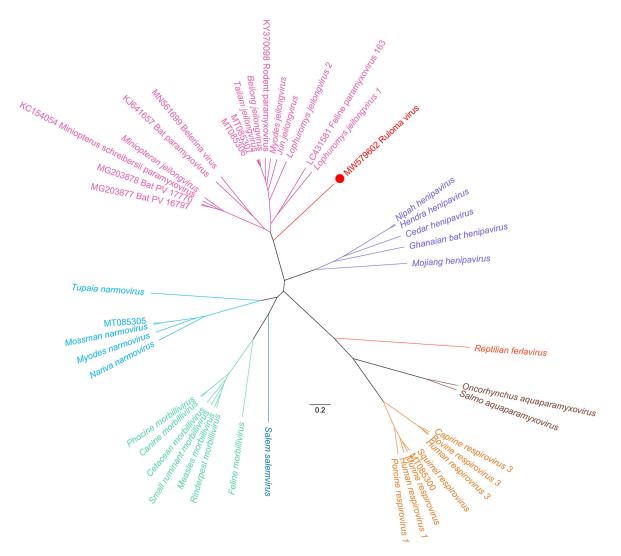
FIG 1 Phylogenetic tree based on the deduced amino acid sequence of the six major open reading frames (N-P-M-F-G-L) of all (putative) member species of the subfamily *Orthoparamyxovirinae* for which (near-) complete genome sequences are available. Ruloma virus is marked with a red circle. Proteins were aligned separately using MAFFT v7.310 (10). The resulting alignments were concatenated and trimmed using TrimAl v1.4.rev15 (gappyout setting) (11). Following model selection using IQ-TREE, BEAST v1.10.4 was used to infer phylogenetic trees, employing a Yule process for tree priors and LG+G4+I as an amino acid substitution model (12–15). TreeAnnotator v1.10.4 was used to summarize a maximum clade credibility tree, which was visualized using FigTree v1.4.3. Sequences are colored according to the genus in which they cluster. For recognized species, the accession number as provided in the ICTV master species list was used (16). For putative species, the accession number is shown in the tree. All nodes had a posterior support of 1, with the exception of the node between MT085307 and Tailam virus, which had a support value of 0.74.

represents the largest mammalian paramyxovirus genome reported thus far (7). The genome has an average GC content of 33.24% and shares little similarity with known paramyxoviruses, as evaluated by pairwise sequence comparison (8) with Mount Mabu Lophuromys virus 1 and 2, showing the highest degree of nucleotide identity (both 41 to 42%). The genome structure, 3'-N-P/V/C-M-F-TM-G-L-5', follows the typical paramyxovirus genome layout but includes an additional gene encoding a transmembrane protein between the F and G genes (9). Similar genome organizations are found only in members of the genus *Jeilongvirus*, which sometimes can even encode a second additional gene (SH) between this TM gene and the F gene. In accordance with its genome structure, phylogenetic analysis based on the protein sequence of all currently available (putative) orthoparamyxovirus species clusters ruloma virus as a sister lineage to all known members of the genus *Jeilongvirus* (Fig. 1). Interestingly, with a length of 784 amino acids, the G protein of ruloma virus is remarkably long, a trait thus far observed only in members of the "rodent clade" of the genus *Jeilongvirus* (5).

REFERENCES

1. Balkema-Buschmann A, Dundon WG, Duprex WP, Easton AJ, Fouchier RAM, Kurath G, Lamb RA, Lee B, Rima BK, Rota PA, Wáng L. 2018. ICTV proposal 2018.011M.A.v1. https://talk.ictvonline.org/ictv/proposals/2018.011M.A.v1 .Paramyxoviridae.zip. Accessed 11 February 2021.

2. Sakaguchi S, Nakagawa S, Mitsuhashi S, Ogawa M, Sugiyama K, Tamukai K, Koide R, Katayama Y, Nakano T, Makino S, Imanishi T, Miyazawa T, Mizutani T. 2020. Molecular characterization of feline paramyxovirus in Japanese cat populations. Arch Virol 165:413–418. https://doi.org/10 .1007/s00705-019-04480-x.

3. Vanmechelen B, Vergote V, Merino M, Verbeken E, Maes P. 2020. Common occurrence of Belerina virus, a novel paramyxovirus found in Belgian hedgehogs. Sci Rep 10:19341. https://doi.org/10.1038/s41598-020-76419-1.

4. Wu Z, Han Y, Liu B, Li H, Zhy G, Latinne A, Dong J, Sun L, Du J, Zhou S, Chen M, Kritiyakan A, Jittapalapong S, Chaisiri K, Buchy P, Duong V, Yang J, Jiang J, Xu X, Zhou H, Yang F, Morand S, Daszak P, Jin Q. 2021. Decoding the RNA viromes of rodent lungs provides new visions into the origin and evolution pattern of rodent-borne diseases in mainland Southeast Asia. Res Square https://doi.org/10.21203/rs.3.rs-17323/v1.

5. Vanmechelen B, Bletsa M, Laenen L, Lopes AR, Vergote V, Beller L, Deboutte W, Korva M, Avsic Zupanc T, Gouy de Bellocq J, Gryseels S, Leirs H, Lemey P, Vrancken B, Maes P. 2018. Discovery and genome characterization of three new Jeilongviruses, a lineage of paramyxoviruses characterized by their unique membrane proteins. BMC Genomics 19:617. https://doi.org/10.1186/s12864-018-4995-0.

6. Bletsa M, Vrancken B, Gryseels S, Boonen I, Fikatas A, Li Y, Laudisoit A, Lequime S, Bryja J, Makundi R, Meheretu Y, Akaibe BD, Mbalitini SG, Van de Perre F, Van Houtte N, Těšíková J, Wollants E, Van Ranst M, Drexler JF, Verheyen E, Leirs H, de Bellocq JG, Lemey P. 2020. Unravelling the evolutionary relationships of hepaciviruses within and across rodent hosts. bioRxiv https://doi.org/10.1101/2020.10.09.332932:2020.10.09.332932.

7. Calain P, Roux L. 1993. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. J Virol 67:4822–4830. https://doi.org/10.1128/JVI.67.8.4822-4830.1993.

8. Bao Y, Chetvernin V, Tatusova T. 2014. Improvements to pairwise sequence comparison (PASC): a genome-based Web tool for virus classification. Arch Virol 159:3293–3304. https://doi.org/10.1007/s00705-014-2197-x.

9. Rima B, Balkema-Buschmann A, Dundon WG, Duprex P, Easton A, Fouchier R, Kurath G, Lamb R, Lee B, Rota P, Wang L, ICTV Report Consortium. 2019. ICTV virus taxonomy profile: Paramyxoviridae. J Gen Virol 100:1593–1594. https://doi.org/10.1099/jgv.0.001328.

10. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.

11. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

12. Heled J. 2012. Sequence diversity under the multispecies coalescent with Yule process and constant population size. Theor Popul Biol 81:97–101. https://doi.org/10.1016/j.tpb.2011.12.007.

13. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol 25:1307–1320. https://doi.org/10.1093/molbev/msn067.

14. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.

15. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016. https://doi.org/10.1093/ve/vey016.

16. ICTV. 2019. ICTV master species list 2019.v1. https://talk.ictvonline.org/files/master-species-lists/m/msl/9601/download. Accessed 11 February 2021.