



# Nearly Complete Genome Sequence of a Novel Phlebovirus-Like Virus Detected in a Human Plasma Sample by High-Throughput Sequencing

Florian Laubscher,<sup>a,b</sup> Samuel Cordey,<sup>a,b</sup> Mary-Anne Hartley,<sup>c</sup> Gael Vieille,<sup>a,b</sup> Noémie Boillat-Blanco,<sup>d,e</sup> Josephine Samaka,<sup>f</sup> Tarsis Mlaganile,<sup>f</sup> Valerie d'Acremont,<sup>c,d</sup> Laurent Kaiser<sup>a,b,g</sup>

<sup>a</sup>Laboratory of Virology, University Hospitals of Geneva, Geneva, Switzerland

<sup>b</sup>University of Geneva Medical School, Geneva, Switzerland

<sup>c</sup>Centre for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

<sup>d</sup>Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland

<sup>e</sup>Infectious Diseases Service, Lausanne University Hospital, Lausanne, Switzerland

<sup>f</sup>Ikara Health Institute, Dar es Salaam, United Republic of Tanzania

<sup>g</sup>Geneva Centre for Emerging Viral Diseases, University Hospitals of Geneva, Geneva, Switzerland

**ABSTRACT** Here, we report a novel phlebovirus-like virus sequence detected in a plasma sample from a febrile adult patient collected in the United Republic of Tanzania in 2014. A nearly complete RNA sequence was generated by high-throughput sequencing on a HiSeq 2500 instrument and further confirmed after repeating the analysis, starting from the initial sample.

Phleboviruses (*Bunyvirales* order, *Phenuiviridae* family) represent single-stranded RNA viruses composed of three segments (S, M, and L). According to the 10th report of the International Committee on Taxonomy of Viruses in 2018 (<https://talk.ictvonline.org/taxonomy/>), the *Phlebovirus* genus is classified into 10 species; its best known member is the mosquito-borne Rift Valley fever virus that infects livestock, wild animals, and humans (1). Several other phleboviruses have been described that infect humans via phlebotomine sandflies or ticks, and although frequently asymptomatic, they may trigger febrile illness or central nervous system infections (2–4).

During a high-throughput sequencing (HTS)-based investigation into the possible viral associations in adult patients with unexplained febrile illness, we detected a nearly complete RNA sequence from a novel phlebovirus-like virus in a plasma specimen collected in March 2014 in Dar es Salaam, United Republic of Tanzania. The specimen originated from an HIV-negative 43-year-old male attending the outpatient department with a 3-day history of fever, headache, and back pain and 2 days of reported pollakiuria. At the time of consultation, the patient was febrile (38.1°C), with a normal blood pressure and pulse rate. Despite extensive microbiological investigations (blood and urine cultures; rapid diagnostic tests for dengue, malaria, and typhoid fever; multiplex PCR targeting tropical causes of fever; and serologies for Epstein-Barr virus, cytomegalovirus, toxoplasmosis, and histoplasmosis), the cause of fever remained unknown.

The plasma sample was prepared for HTS analysis as previously described (i.e., centrifugation, DNase treatment, and RNA extraction by TRIzol) (5). The RNA library was prepared using the TruSeq total RNA preparation protocol (Illumina, San Diego, CA, USA) and run on the HiSeq 2500 platform (Illumina) using the 2 × 100-nucleotide read length protocol. Adapters were trimmed and quality checked using Trimmomatic (v0.33) (parameters included Illuminaclip, TruSeq3-PE-2.fa:1:30:10:4:true; MAXINFO, 40:0.3; and minlen, 36) (6). Human reads were removed by mapping reads against the

**Citation** Laubscher F, Cordey S, Hartley M-A, Vieille G, Boillat-Blanco N, Samaka J, Mlaganile T, d'Acremont V, Kaiser L. 2019. Nearly complete genome sequence of a novel phlebovirus-like virus detected in a human plasma sample by high-throughput sequencing. *Microbiol Resour Announc* 8:e00764-19. <https://doi.org/10.1128/MRA.00764-19>.

**Editor** Jelle Matthijnsens, KU Leuven

**Copyright** © 2019 Laubscher 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/).

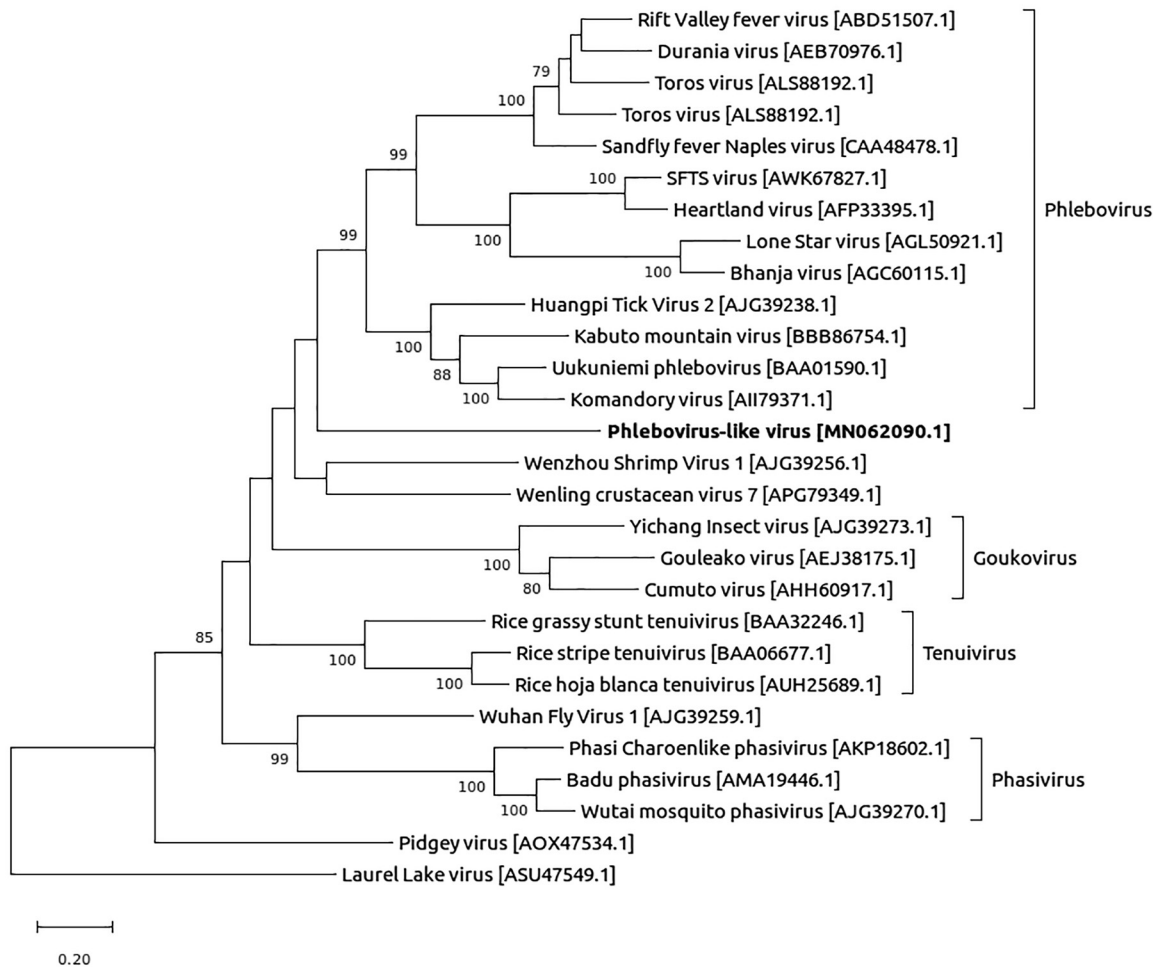
Address correspondence to Samuel Cordey, [samuel.cordey@hcuge.ch](mailto:samuel.cordey@hcuge.ch).

F.L. and S.C. contributed equally to this article.

**Received** 3 July 2019

**Accepted** 6 August 2019

**Published** 29 August 2019



**FIG 1** Phylogenetic tree constructed for the RNA-dependent RNA polymerase using a 500-replicate maximum likelihood method based on the LG model (11), using the Laurel Lake virus (GenBank accession number [ASU47549](#)) as an outgroup. The alignment and tree building were done using SeaView v4.7 (MUSCLE v3.8.31) and the MEGA X software, respectively.

human genome and transcriptome (hg38; Gencode v23) using SNAP (v1.0beta.23; command, paired with default parameters) (7). The *de novo* assembly was made using IDBA-UD (v1.1.3; parameters included `pre_correction`, `-step 20`, `-mink 21`, and `-maxk 100`) (8). Contigs were subjected to a BLAST search (BLASTX, v2.3.0+) (9) against the U-RVDBv12.2 viral database (<https://rvdb-prot.pasteur.fr/>). Then, reads were mapped back to the contigs using SNAP. A total of 4,059, 5,112, and 7,401 reads specific to the S, M, and L segments, respectively, were remapped (median read depths, 337, 129, and 113, respectively). One contig was obtained for each individual segment. The BLAST result showed a maximum of 31% amino acid identity (51% similarity) with other fecal bunyavirus (GenBank accession number [AIB06815](#)) for the S segment, 27% amino acid identity (46% similarity) with Pidghey bunyavirus (GenBank accession number [KX852390](#)) for the M segment, and 31% amino acid identity (50% similarity) with Zaliv Terpenia virus (Uukuniemi phlebovirus) (GenBank accession number [KF767463](#)) for the L segment. Of note, dengue virus-specific sequences were detected but resulted from index hopping. None of the 31 RNA libraries in the same run (including two HTS whole-process controls) were found to be positive for this novel phlebovirus-like sequence. Nevertheless, such discoveries should be interpreted with caution because they could result from a punctual HTS environmental contamination (10). To this end, the analysis was repeated, starting from the initial plasma sample. The novel RNA library was run alone on a HiSeq 4000 instrument (Illumina), which confirmed the presence of this

phlebovirus-like virus sequence (a total of 60,605, 80,846, and 119,580 reads specific to the S, M, and L segments, respectively).

The nearly complete virus genome sequence detected was 1,309, 3,776, and 6,474 nucleotides long for the S, M, and L segments, respectively (G+C contents of 41.0%, 41.6%, and 39.1%, respectively), covering the respective complete coding sequences. A phylogenetic analysis performed on the L segment showed that this viral RNA sequence was related to *Phlebovirus* genus sequences (Fig. 1). Although the detection of this novel phlebovirus-like sequence should be interpreted with caution (i.e., the detection of a sequence does not prove by itself a true human infection), the report of such a sequence is important since phleboviruses are known to infect humans and can trigger febrile illness or central nervous system infections.

**Data availability.** The nucleotide sequences detected from the clinical specimen collected in the United Republic of Tanzania were deposited in GenBank under the accession numbers [MN062090](https://doi.org/10.1093/mra/mn062090), [MN062091](https://doi.org/10.1093/mra/mn062091), and [MN062092](https://doi.org/10.1093/mra/mn062092). The raw sequence data were deposited in the NCBI Sequence Read Archive under BioProject accession number [PRJNA541799](https://doi.org/10.1093/mra/prjna541799).

## ACKNOWLEDGMENTS

We thank Mylène Docquier and Brice Petit from the iGE3 Genomics Platform, University of Geneva, Switzerland.

The study was supported by the Bill and Melinda Gates Foundation (grants OPP1163230 and OPP1163434, Optimization of Sequence-Based Microbial Surveillance) and the Swiss National Science Foundation (grants 310030\_165873 and 320030\_179507).

## REFERENCES

1. Daubney R, Hudson JR, Garnham PC. 1931. Enzootic hepatitis or Rift Valley fever. An undescribed virus disease of sheep, cattle and man from East Africa. *J Pathol* 34:545–579. <https://doi.org/10.1002/path.1700340418>.
2. Ayhan N, Charrel RN. 2017. Of phlebotomines (sandflies) and viruses: a comprehensive perspective on a complex situation. *Curr Opin Insect Sci* 22:117–124. <https://doi.org/10.1016/j.cois.2017.05.019>.
3. Elliott RM, Brennan B. 2014. Emerging phleboviruses. *Curr Opin Virol* 5:50–57. <https://doi.org/10.1016/j.coviro.2014.01.011>.
4. Ergunay K, Ayhan N, Charrel RN. 2017. Novel and emergent sandfly-borne phleboviruses in Asia Minor: a systematic review. *Rev Med Virol* 27:e1898. <https://doi.org/10.1002/rmv.1898>.
5. Petty TJ, Cordey S, Padioleau I, Docquier M, Turin L, Preynat-Seauve O, Zdobnov EM, Kaiser L. 2014. Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. *J Clin Microbiol* 52:3351–3361. <https://doi.org/10.1128/JCM.01389-14>.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
7. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger AL, Luk KC, Enge B, Wadford DA, Messenger SL, Genrich GL, Pellegrino K, Grard G, Leroy E, Schneider BS, Fair JN, Martinez MA, Isa P, Crump JA, DeRisi JL, Sittler T, Hackett J, Jr, Miller S, Chiu CY. 2014. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res* 24:1180–1192. <https://doi.org/10.1101/gr.171934.113>.
8. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>.
9. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
10. Asplund M, Kjartansdottir KR, Mollerup S, Vinner L, Fridholm H, Herrera JAR, Friis-Nielsen J, Hansen TA, Jensen RH, Nielsen IB, Richter SR, Rey-Iglesia A, Matey-Hernandez ML, Alquezar-Planas DE, Olsen PVS, Sicheritz-Ponten T, Willerslev E, Lund O, Brunak S, Mourier T, Nielsen LP, Izarzugaza JMG, Hansen AJ. 2019. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clin Microbiol Infect*, in press. <https://doi.org/10.1016/j.cmi.2019.04.028>.
11. 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>.