 doi:10.20944/preprints202011.0502.v1

Draft Genome Sequence of Medusavirus stheno, a new member of "Medusaviridae" Isolated from the Tatakai River of Uji, Japan

Koki Yoshida,[a] Ruixuan Zhang,[a] Kimberly G. Garcia,[a] Hisashi Endo,[a] Yasuhiro Gotoh,[b] Tetsuya Hayashi,[b] Masaharu Takemura,[c] Hiroyuki Ogata[a]#

[a]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji 611-0011, Japan

[b]Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

[c]Laboratory of Biology, Department of Liberal Arts, Faculty of Science, Tokyo University of Science, Shinjuku, Tokyo 162-8601, Japan

#Address correspondence to Hiroyuki Ogata, ogata@kuicr.kyoto-u.ac.jp.

Koki Yoshida and Ruixuan Zhang contributed equally to this work. Author order was determined in alphabetical order.

Abstract

"Medusaviridae" is a proposed family of large DNA viruses so far represented by a sole virus isolated from a hot spring. In the present study, we report the isolation and genome sequencing of a new member of this family, medusavirus stheno, discovered from a freshwater sample with *Acanthamoeba castellanii* coculture.

The founding member of "Medusaviridae" of the phylum *Nucleocytoviricota* was previously isolated from a hot spring in Japan (1). The new second member of this family, named medusavirus stheno, was isolated from water/soil samples from the Tatakai River, Uji, Japan.

 doi:10.20944/preprints202011.0502.v1

Samples were filtered with filter paper 43 (Whatman PLC) and 1.2-µm pore size Minisart® Syringe Filter (Sartorius). Ninety µL of solution (18 mL peptone yeast extract-glucose (PYG), 500 µL of amoeba cells) and 9.5 µL of filtered samples were added in a 96-well plate. After 7 days of culture (26°C), 10 µL of supernatant from each well showing delayed proliferation was mixed with 1 mL of PYG, 3 drops of amoeba culture solution in a 24-well plate. After 7 days, the supernatant from wells showing delayed proliferation was serially diluted down to 10^{-11} fold with PYG. Ten µL of each diluted solution was mixed with 90 µL of PYG medium (16 mL PYG and 300 µL amoeba cells) in a 96-well plate. After another 7 days of culture, wells showing a delayed growth with at least 10^{-6} fold dilution were considered to contain viruses without bacteria. Fresh amoeba cells were inoculated with supernatants from these wells in a 75 cm^3 culture flask. After two days, cells were harvested and centrifuged at 538 x g for 5 min twice at 26°C, then the supernatant was centrifugated at 8000 x g for 35 min at 4°C. The resulting pellet was resuspended with 1 mL of phosphate-buffered saline. Centrifugation and resuspension were repeated twice to obtain pellets containing the putative viral particles. Observation of virions in amoeba cells using transmission electron microscopy was performed as described previously (Fig. 1A) (2, 3).

DNA extraction was performed with NucleoSpin Tissue XS kit (Macherey-Nagel GmbH and Co. KG). Sequencing was performed using Nanopore MinION (Oxford Nanopore Technologies, Inc.), HiSeq X Ten (Illumina, Inc.) and MiSeq (Illumina, Inc.). Unicycler (with options SPAdes assembler, Pilon polisher) and Bandage were used for *de novo* hybrid assembly (4, 5), which produced a 362,811 bp contig. Coding sequences (CDSs) were predicted using Prodigal (6) and GeneWise v2.4.1 (7). Putative promoter motifs were identified using MEME version 5.1.1 with E-value < 10^{-50} (8). Homology searches (E-value < 10^{-5}) were performed against NCBI NR using BLASTP (9).

The contig was colinearly aligned with the original medusavirus genome (381 kb, the average nucleotide identity of 79.8%, Fig. 1B) and found to encode 429 CDSs, of which 349 (81%) had their best BLASTP hit in the original medusavirus. Like the original virus, the new medusavirus stheno encoded a complete set of histone domains (H1, H2A, H2B, H3, and H4), but, unlike the original one, H3 and H4 were found encoded in a fused single CDS (ORF_240). Two conserved sequence motifs were identified in

 doi:10.20944/preprints202011.0502.v1

the upstream regions of the CDSs (Table 1). These motifs were also found in the original medusavirus genome.

Data availability. Raw reads and the draft genome sequence of have been deposited in DDBJ (DRA010707) and GenBank (MW018138), respectively.

Acknowledgement

Computational time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University.

Funding Statement. This work was supported by JSPS/KAKENHI (Nos. 18H02279, 16H06279 (PAGS)), The Kyoto University Foundation, and the Collaborative Research Program of the Institute for Chemical Research, Kyoto University (Nos. 2018-32, 2019-34, 2020-31).

References

1. Yoshikawa G, Blanc-Mathieu R, Song C, Kayama Y, Mochizuki T, Murata K, Ogata H, Takemura M. 2019. Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. J Virol 93:1–25.
2. Takemura M. 2016. Morphological and Taxonomic properties of Tokyovirus, the first Marseilleviridae member isolated from Japan. Microbes Environ 31:442–448.
3. Aoki K, Hagiwara R, Akashi M, Sasaki K, Murata K, Ogata H, Takemura M. 2019. Fifteen Marseilleviruses Newly Isolated From Three Water Samples in Japan Reveal Local Diversity of Marseilleviridae. Front Microbiol 10:1152.
4. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:1–22.

5. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: Interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352.

6. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11.

7. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641.

8. Bailey Timothy L. EC. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Second Int Conf Intell Syst Mol Biol 2:28–36.

9. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9.

10. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5.

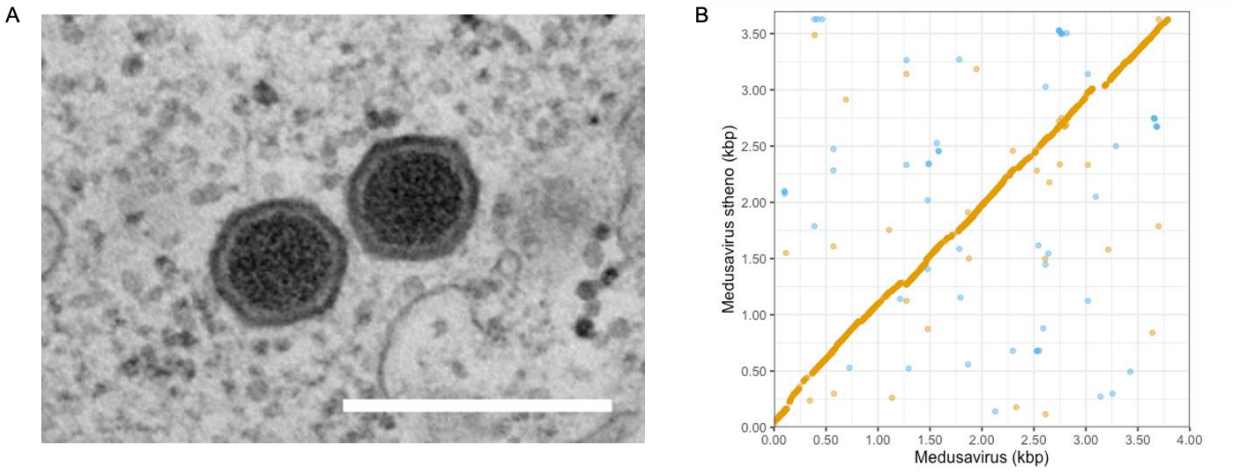
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0502.v1

Figure Legends

Fig 1. Medusavirus stheno virions and geome comparison. (A) Observation of medusaivrus stheno virions by transmission electron microscopy (scale bar 400 nm). (B) Comparison of medusavirus genomes. Dot plot was generated using MUMmer version 3.23 (10).

Table Legends

Table 1. Sequence motifs identified in the upstream regions of medusaviruses.



Medusavirus stheno			Medusavirus		
Consensus sequence	Sites	E-value	Consensus sequence	Sites	E-value
SRCCAYATGAMBTCACATGGC	43	4.1e-162	GCCATRTGAVKTCATRTGGYSRSG	53	8.4e-183
VMMMAMADMAAMAAA	252	5.5e-130	VMAAMAAMARMAAMA	251	3.1e-146