AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Complete Genome Sequence of a New Megavirus Family Member Isolated from an Inland Water Lake for the First Time in India

Anirvan Chatterjee, Farhan Ali, Disha Bange, Kiran Kondabagil

Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India

We report here the isolation and complete genome sequencing of a large double-stranded DNA virus, *Powai Lake megavirus*, for the first time from India. The isolation of a large DNA virus with genome size >1 Mb from India further attests to the prevalence of *Giant* viruses in different environmental niches.

Received 30 March 2016 **Accepted** 10 May 2016 **Published** 16 June 2016

Citation Chatterjee A, Ali F, Bange D, Kondabagil K. 2016. Complete genome sequence of a new megavirus family member isolated from an inland water lake for the first time in India. Genome Announc 4(3):e00402-16. doi:10.1128/genomeA.00402-16.

Address correspondence to Kiran Kondabagil, kirankondabagil@iitb.ac.in.

Nucleocytoplasmic large DNA viruses (NCLDVs), or *Giant* viruses, have genome sizes reaching up to 2.5 Mb (1) and include four families. With their large unique genomes and widespread presence in the aquatic environments (2–6), the NCLDVs are thought to be one of the major vehicles of evolution and are now being explored to revisit evolutionary paradigms, such as origin of eukaryotes (7), evolution of DNA replication system (8), genome packaging systems (9, 10), etc. Sequencing of more NCLDV genomes is necessary to understand their ecological and evolutionary significance.

In this study, we isolated and sequenced the genome of a large double-stranded DNA virus infecting a free-living amoeba, *Acanthamoeba castellanii*. Water samples were collected from Powai Lake, an artificial inland lake in Mumbai, India, and processed for isolation of large viruses against the amoeba host, as per published protocols (11). Transmission electron microscopy revealed icosahedral particles of about 425 nm in diameter comparable to that of some NCLDVs reported previously. Isolated virus was propagated in *Acanthamoeba castellanii*, purified on a sucrose gradient, and the genome was extracted as described earlier (11). Whole-genome shotgun sequencing was performed using Illumina MiSeq 2 × 150-bp paired-end chemistry that yielded 3,481,650 reads. Kraken metagenomics (Illumina BaseSpace webtool), performed with trimmed and quality control (QC)-filtered reads, taxonomically classified 15% of the reads as *Megaviridae*; hence, the isolate was named *Powai Lake megavirus* (PLMV). The G+C content of PLMV (25%) is comparable to that of other *Giant* viruses.

De novo assembly was performed using A5-miseq (12) and evaluated using QUAST (13). The PLMV assembly exhibited a median coverage of 793×, with an N_{50} value of 750,973 bp. All contigs were aligned to the BLAST NR database using MegaBLAST (14, 15), and a consensus FASTA sequence was generated by reordering the 16 contigs using MAUVE (16). The PLMV genome size was found to be 1,208,707 bp, with 996 open reading frames (ORFs), as predicted by GeneMarkS (17). The ORFs were individually annotated using BLASTP (15), and the results were retrieved using custom Python scripts. The annotated genomes were uploaded to NCBI using BankIt Web-based submission tool. Using tRNAscan-SE (18), PLMV was found to encode 5 tRNAs, and the CRISPRFinder Web tool (19–21) detected 3 confirmed clustered regularly interspaced short palindromic repeats (CRISPRs) and 6 CRISPR-like sequences in the PLMV genome. Further, a TransposonPSI (http://transposonpsi.sourceforge.net/) search yielded at least 4 hits in the PLMV genome. The isolation and whole-genome sequencing of a first NCLDV from India presents an opportunity to study their significance in India's rich ecological diversity.

Nucleotide sequence accession number. The complete genome of PLMV has been deposited in the GenBank under the accession no. KU877344.

REFERENCES

1. **Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie JM, Abergel C.** 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science **341:**281–286. http://dx.doi.org/10.1126/science.1239181.
2. **Benamar S, Reteno DG, Bandaly V, Labas N, Raoult D, La Scola B.** 2016. Faustoviruses: comparative genomics of new *Megavirales* family members. Front Microbiol **7:**3. http://dx.doi.org/10.3389/fmicb.2016.00003.
3. **Zhang W, Zhou J, Liu T, Yu Y, Pan Y, Yan S, Wang Y.** 2015. Four novel algal virus genomes discovered from Yellowstone Lake metagenomes. Sci Rep **5:**15131. http://dx.doi.org/10.1038/srep15131.
4. **Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic JM, Ramus C, Bruley C, Labadie K, Shmakova L, Rivkina E, Couté Y, Abergel C, Claverie JM.** 2015. In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*. Proc Natl Acad Sci USA **112:**E5327–E5335. http://dx.doi.org/10.1073/pnas.1510795112.
5. **Boratto PV, Arantes TS, Silva LC, Assis FL, Kroon EG, La Scola B, Abrahão JS.** 2015. Niemeyer virus: a new mimivirus group A isolate harboring a set of duplicated aminoacyl-tRNA synthetase genes. Front Microbiol **6:**1256. http://dx.doi.org/10.3389/fmicb.2015.01256.
6. **Abergel C, Legendre M, Claverie JM.** 2015. The rapidly expanding universe of giant viruses: mimivirus, *Pandoravirus*, *Pithovirus* and *Mollivirus*. FEMS Microbiol Rev **39:**779–796. http://dx.doi.org/10.1093/femsre/fuv037.
7. **Forterre P, Gaïa M.** 2016. Giant viruses and the origin of modern eukaryotes. Curr Opin Microbiol **31:**44–49. http://dx.doi.org/10.1016/j.mib.2016.02.001.
8. **Takemura M, Yokobori S, Ogata H.** 2015. Evolution of eukaryotic DNA

polymerases via interaction between cells and large DNA viruses. J Mol Evol **81:**24–33. http://dx.doi.org/10.1007/s00239-015-9690-z.

9. **Chelikani V, Ranjan T, Zade A, Shukla A, Kondabagil K**. 2014. Genome segregation and packaging machinery in *Acanthamoeba polyphaga* mimivirus is reminiscent of bacterial apparatus. J Virol **88:**6069–6075. http://dx.doi.org/10.1128/JVI.03199-13.
10. **Chelikani V, Ranjan T, Kondabagil K**. 2014. Revisiting the genome packaging in viruses with lessons from the "giants." Virology **466–467:** 15–26. http://dx.doi.org/10.1016/j.virol.2014.06.022.
11. **Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM**. 2004. The 1.2-megabase genome sequence of mimivirus. Science **306:**1344–1350. http://dx.doi.org/10.1126/science.1101485.
12. **Coil D, Jospin G, Darling AE**. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics **31:**587–589. http://dx.doi.org/10.1093/bioinformatics/btu661.
13. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29:**1072–1075. http://dx.doi.org/10.1093/bioinformatics/btt086.
14. **Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA**. 2008. Database indexing for production MegaBLAST searches. Bioinformatics **24:**1757–1764. http://dx.doi.org/10.1093/bioinformatics/btn322.
15. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. 1990. Basic Local Alignment Search Tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
16. **Darling AC, Mau B, Blattner FR, Perna NT**. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394–1403. http://dx.doi.org/10.1101/gr.2289704.
17. **Besemer J, Lomsadze A, Borodovsky M**. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res **29:**2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.
18. **Lowe TM, Eddy SR**. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:** 955–964. http://dx.doi.org/10.1093/nar/25.5.0955.
19. **Grissa I, Vergnaud G, Pourcel C**. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **35:**W52–W57. http://dx.doi.org/10.1093/nar/gkm360.
20. **Grissa I, Vergnaud G, Pourcel C**. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8:**172. http://dx.doi.org/10.1186/1471-2105-8-172.
21. **Grissa I, Vergnaud G, Pourcel C**. 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **36:**W145–W148. http://dx.doi.org/10.1093/nar/gkn228.