

Complete genome sequence of an Indian outbreak strain of chikungunya virus

Eshna Laha,^{1,2} Deepak Jena,¹ Viplov K. Biswas,^{1,3} Sharad Singh,^{1,3} Sunil K. Raghav,¹ Asit K. Pattnaik,^{4,5} Soma Chattopadhyay¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Here, we report the complete genome sequence of an Indian strain of chikungunya virus isolated from an infected patient from Hyderabad, Andhra Pradesh, India, during a massive outbreak in 2005–2006. The genome length spans 11,811 nucleotides and has a poly(A) tail of 29 residues at the 3' end.

KEYWORDS chikungunya virus, CHIKV, whole genome sequence

Chikungunya virus (CHIKV) (family *Togaviridae*, genus *Alphavirus*) is an arbovirus responsible for debilitating arthritis in humans with no commercially available therapeutics or vaccines (1). India witnessed its first major outbreak in 2005–2006 with nearly 1.3 million people in over 13 Indian states affected to date (2).

A Chikungunya virus Indian outbreak strain (CHIKV-IS) was isolated from a CHIKV-infected patient serum during a rapid outbreak in Hyderabad, Andhra Pradesh, in 2005–2006 and stored at -80°C . The virus was adapted in Vero cells in 2008 through multiple serial passages. Post-adaptation, Vero cells were infected with CHIKV-IS at a multiplicity of infection of 0.1, and plaque assay was performed as described earlier (3). A single well-isolated plaque was picked and resuspended in serum-free media and was used to infect fresh Vero cells. Viral RNA was then isolated from the culture supernatant using the QiaAmp Viral RNA Isolation Kit (Qiagen) as per manufacturer instructions. The sequencing library was prepared using the TruSeq Stranded Total RNA Library Prep Kit (Illumina, USA) following the manufacturer's protocol. Purified and enriched products were used to create a paired-end next-generation sequencing library and were sequenced using the NovaSeq 6000 platform (Illumina).

Raw BCL files were demultiplexed using the bcl2fastq tool (v2.20.0.422). All tools used for analysis were run with default parameters unless specified. A total of ~ 12.2 million paired-end reads were generated. The quality of the reads was checked using the FastQC tool (v.0.11.9) (4). The average read length was 36 nt. The adapter sequences and low-quality reads were trimmed and filtered using Cutadapt (v.4.1) for downstream analysis (5). The good quality trimmed reads were mapped to the human genome (GRCh38). All unmapped paired-end reads were extracted using SAMTOOLS (v.1.16.1) and mapped against the CHIKV reference genome (GenBank: [DQ443544.2](#)) using BWA mem (v.0.7.17) (6, 7). Deduplication of aligned reads was done using Picard tools (v.2.18.7). Alignment quality was checked using the SAMTOOLS (v.1.16.1) "--flagstat" option. A total of ~ 2.3 million reads were mapped. We obtained an average coverage depth of $>6,000$ at each nucleotide position, and the GC content of the final genome was 51%. The nucleotide positions 1,052, 4,167, and 5,049 were confirmed by Sanger sequencing and substituted in the consensus genome file. The consensus genome sequences were derived using the BCFTOOLS (v.1.9) "--consensus" option (8).

The CHIKV-IS genome is 11,811 nt long, followed by 29 adenosine residues at the 3'-terminus. The genome spans non-structural (7,422 nt) and structural (3,744 nt) open-reading frames, separated by a 65-nt-long untranslated region (UTR). The 5'- and

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Sunil K. Raghav, sunilraghav@ils.res.in, Asit K. Pattnaik, apattnaik2@unl.edu, or Soma Chattopadhyay, sochat.ils@gmail.com.

Eshna Laha and Deepak Jena contributed equally to this article. The author order was determined based on their contribution to the article

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 16 April 2024

Accepted 25 June 2024

Published 31 July 2024

Copyright © 2024 Laha et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

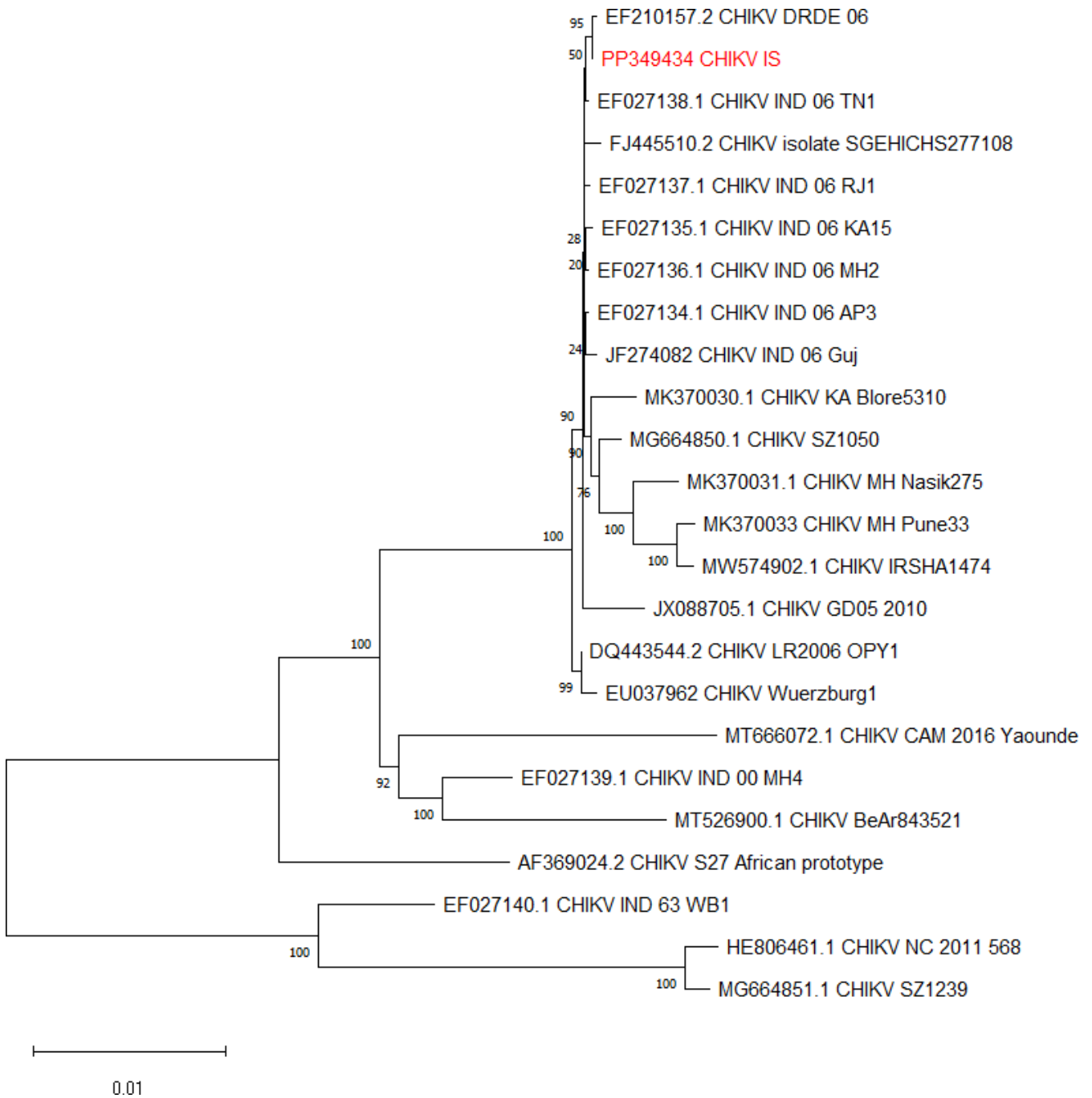


FIG 1 Phylogeny of chikungunya virus strains. Evolutionary history was inferred using the neighbor-joining method with default parameters. The optimal tree is shown and drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum composite likelihood method and are in units of the number of base substitutions per site. MUSCLE tool was used to build the alignment and cleaned using the “pairwise deletion” option. Neighbor-joining method was the statistical method of choice with 1,000 bootstrap replicates. MEGA11 (v.11.0.13) was used to build and analyze the tree.

3'-terminal UTRs are 76 and 498 nt, respectively. The CHIKV-IS lacks the signature A226V mutation in the E1 region, known for vector adaptation, enhanced transmissibility, and higher epidemic potential (9), and also lacks the internal poly(A) within the 3'UTR otherwise present in the prototypic S-27 strain (10). The phylogenetic tree of the CHIKV whole genome sequences from all lineages built using MEGA11 (v.11.0.13) reveals that

CHIKV-IS belongs to the ECSA (East/Central/South African) lineage, sharing nearly 99% sequence identity to several Indian isolates (Fig. 1).

ACKNOWLEDGMENTS

E.L. was supported by the Inspire Fellowship Scheme from the Department of Science and Technology, Government of India. The study was supported in part by funds from the Visiting Advanced Joint Research (VAJRA) program (VJR/2018/000112) of the Department of Science and Technology, Government of India, and DBT Core fund of the Institute of Life Sciences, Bhubaneswar.

E.L.: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, visualization, validation, and writing of the original draft; V.K.B.: data curation, formal analysis, investigation, methodology, resources software, visualization, and validation; D.J.: data curation, formal analysis, resources software, visualization, and validation; S.S.: data curation, formal analysis, investigation, methodology, resources, and writing (reviewing and editing); S.K.R.: resources, software, and writing (reviewing and editing); A.K.P.: conceptualization, investigation, visualization, methodology, fund acquisition, project administration, supervision, and writing (original draft, review, and editing); S.C.: conceptualization, investigation, visualization, methodology, fund acquisition, project administration, supervision, resources, and writing (original draft, review, and editing).

AUTHOR AFFILIATIONS

¹Institute of Life Sciences, Bhubaneswar, India

²Regional Centre for Biotechnology, Faridabad, India

³School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT) University, Bhubaneswar, India

⁴School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, Lincoln, USA

⁵Nebraska Center for Virology, University of Nebraska-Lincoln, Lincoln, USA

AUTHOR ORCIDs

Eshna Laha  <http://orcid.org/0009-0000-6613-8932>

Viplov K. Biswas  <http://orcid.org/0000-0003-3030-0242>

Sharad Singh  <http://orcid.org/0000-0002-2518-9247>

Sunil K. Raghav  <http://orcid.org/0000-0001-9913-4241>

Asit K. Pattnaik  <http://orcid.org/0000-0002-9581-6787>

Soma Chattopadhyay  <http://orcid.org/0000-0001-8378-1014>

FUNDING

Funder	Grant(s)	Author(s)
DST Science and Engineering Research Board (SERB)	VJR/2018/000112	Asit K. Pattnaik Soma Chattopadhyay
DBT-ILS Core Fund		Sunil K. Raghav Soma Chattopadhyay

DATA AVAILABILITY

The complete genome sequence of the chikungunya virus Indian outbreak strain reported here was deposited in GenBank under accession number [PP349434](https://www.ncbi.nlm.nih.gov/nuccore/PP349434). Raw sequence data were deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) under SRA accession number [SRR28383157](https://www.ncbi.nlm.nih.gov/sra/SRR28383157) and BioProject accession number [PRJNA1089339](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1089339).

REFERENCES

1. Santhosh SR, Dash PK, Parida MM, Khan M, Tiwari M, Lakshmana Rao PV. 2008. Comparative full genome analysis revealed E1: A226V shift in 2007 Indian Chikungunya virus isolates. *Virus Res* 135:36–41. <https://doi.org/10.1016/j.virusres.2008.02.004>
2. Sudeep AB, Parashar D. 2008. Chikungunya: an overview. *J Biosci* 33:443–449. <https://doi.org/10.1007/s12038-008-0063-2>
3. De S, Mamidi P, Ghosh S, Keshry SS, Mahish C, Pani SS, Laha E, Ray A, Datey A, Chatterjee S, Singh S, Mukherjee T, Khamaru S, Chattopadhyay S, Subudhi BB, Chattopadhyay S. 2022. Telmisartan restricts chikungunya virus infection *in vitro* and *in vivo* through the AT1/PPAR- γ /MAPKs pathways. *Antimicrob Agents Chemother (Bethesda)* 66:e0148921. <https://doi.org/10.1128/AAC.01489-21>
4. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. 2013. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet* 4:288. <https://doi.org/10.3389/fgene.2013.00288>
5. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 17:10. <https://doi.org/10.14806/ej.17.1.200>
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
8. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
9. Tsatsarkin KA, Vanlandingham DL, McGee CE, Higgs S. 2007. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog* 3:e201. <https://doi.org/10.1371/journal.ppat.0030201>
10. Arankalle VA, Shrivastava S, Cherian S, Gunjikar RS, Walimbe AM, Jadhav SM, Sudeep AB, Mishra AC. 2007. Genetic divergence of Chikungunya viruses in India (1963-2006) with special reference to the 2005-2006 explosive epidemic. *J Gen Virol* 88:1967–1976. <https://doi.org/10.1099/vir.0.82714-0>