



# Genome Sequences of Chikungunya Virus Isolates from Bolivia

 Caio M. B. França,<sup>a</sup>  Roxana Loayza,<sup>b</sup>  Yelin Roca,<sup>b</sup>  Ana Maria Montañó Arias,<sup>b</sup>  Freddy Tinajeros,<sup>c</sup>  Jose R. Loaiza,<sup>d,e,f</sup>  
 Anshule Takyar,<sup>g</sup>  Robert H. Gilman,<sup>h</sup>  Matthew J. Miller<sup>i</sup>

<sup>a</sup>Department of Biology, Southern Nazarene University, Bethany, Oklahoma, USA

<sup>b</sup>U.F. Biología Molecular, Centro Nacional de Enfermedades Tropicales (CENETROP), Santa Cruz, Bolivia

<sup>c</sup>Asociación Benéfica PRISMA, Santa Cruz, Bolivia

<sup>d</sup>Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP), City of Knowledge, Panama City, Republic of Panama

<sup>e</sup>Smithsonian Tropical Research Institute, Balboa, Republic of Panama

<sup>f</sup>Programa Centroamericano de Maestría en Entomología, Universidad de Panamá, Panama City, Republic of Panama

<sup>g</sup>Department of Public Health and Department of Microbiology, University of Oklahoma, Norman, Oklahoma, USA

<sup>h</sup>Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

<sup>i</sup>Sam Noble Museum of Natural History, University of Oklahoma, Norman, Oklahoma, USA

**ABSTRACT** We generated nine coding-complete chikungunya virus genome sequences from blood samples collected during the early 2015 outbreak in Bolivia. Relative to other publicly available chikungunya sequences, the Bolivian samples represent a monophyletic group, suggesting that a single lineage was widely circulating in the country between February and May 2015.

**A**mong the Andean nations of South America, Bolivia has had the highest incidence of chikungunya and postinfection chronic disease (1). In Bolivia, chikungunya virus was first detected in early 2015, with cases of disease peaking between March and May 2015 (Fig. 1). Here, we report nine chikungunya (*Togaviridae: Alphavirus*) genome sequences for isolates from Bolivia.

Febrile patients were screened for chikungunya virus at the Cenetrop national tropical medicine laboratory. We selected nine archived samples (maximum of 1 passage) for sequencing; samples were selected at random (Table 1). All isolates came from blood-extracted RNA (QIAamp viral RNA minikit; Qiagen) with unambiguously positive quantitative PCR (qPCR) tests (Pan American Health Organization [PAHO] diagnostic kits). Seven of the nine samples were from Santa Cruz de la Sierra. We also included one sample from Cochabamba and one sample from Trinidad. We generated cDNA using random hexamers via reverse transcriptase PCR (RT-PCR) (TaqMan reverse transcription reagents; Applied Biosystems). We amplified the chikungunya genome using a multiplex tiled amplicon approach (2). All samples were pooled and sequenced on a single Oxford Nanopore MinION R9.4 flow cell, generating 2,776,384 reads.

Base calling was done in real time using Albacore v2.3.1, which implements quality filtering (QC), using only QC-passed reads in subsequent analyses. We demultiplexed and trimmed adapters and barcodes using qcat v1.1.0 (<https://github.com/nanoporetech/qcat>), which detected barcodes in 2,538,578 reads (>91%) and assigned only 80 out of 2,538,578 (<0.0004%) reads to barcodes BC10 to BC12 (not used in this study, but assignable in the qcat demultiplexing algorithm). This suggests negligible read misassignment during demultiplexing. The average read length for QC-passed reads was 325.4 bp (range, 100 to 3,727 bp).

For our highest read count sample (4866-15), we error corrected, trimmed, and *de novo* assembled reads in Canu v1.9 (3). The resulting assembly was fragmented, so we selected the largest contig to identify the closest whole chikungunya genome on GenBank using BLAST (2 March 2020) (4) to guide reference-based assembly. The BLAST

**Citation** França CMB, Loayza R, Roca Y, Montañó Arias AM, Tinajeros F, Loaiza JR, Takyar A, Gilman RH, Miller MJ. 2020. Genome sequences of chikungunya virus isolates from Bolivia. *Microbiol Resour Announc* 9:e00230-20. <https://doi.org/10.1128/MRA.00230-20>.

**Editor** Simon Roux, DOE Joint Genome Institute

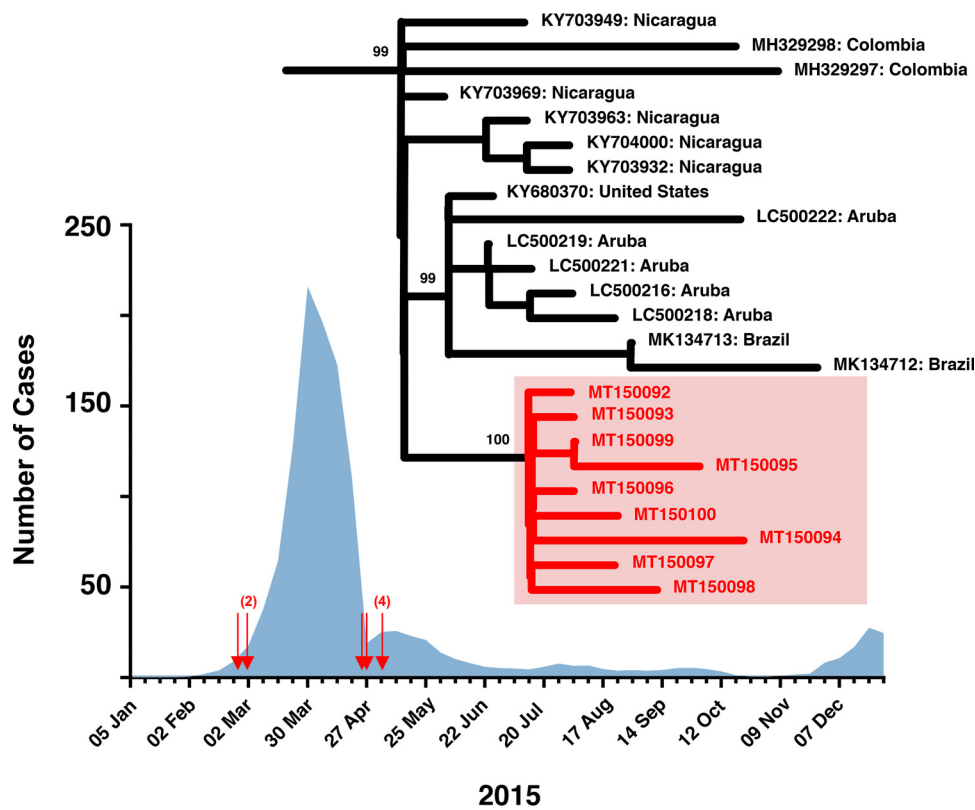
**Copyright** © 2020 França et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Matthew J. Miller, [mjmillier@ou.edu](mailto:mjmillier@ou.edu).

**Received** 8 March 2020

**Accepted** 24 March 2020

**Published** 16 April 2020



**FIG 1** Four-week moving average of chikungunya case count across Bolivia. Red arrows indicate the collection dates for samples sequenced in this study. Maximum likelihood phylogeny was pruned to the clade, including the nine samples. Three nodes have significant approximate bootstrap support (>90).

search returned GenBank accession number [KY703969.1](https://www.ncbi.nlm.nih.gov/nuccore/KY703969.1) as the closest match. We used Canu to correct and trim all nine samples. After trimming, we retained 836,301 reads with an average read length of 345.9 base pairs (range, 184 to 496 base pairs). We mapped these reads to the sequence of [KY703969.1](https://www.ncbi.nlm.nih.gov/nuccore/KY703969.1) using Minimap2 (5) implemented in Geneious v2020.0.5 (6). For each sample, we generated a consensus sequence. All nine consensus sequences along with that from [KY703969.1](https://www.ncbi.nlm.nih.gov/nuccore/KY703969.1) were aligned in Geneious, and we visually corrected homoplasy indel errors, which are common to Oxford Nanopore-derived sequences (7). Default parameters were used for all bioinformatic

**TABLE 1** Results of sequencing efforts for nine chikungunya virus isolates from Bolivia

Sample name	Collection date	City, department	No. of uncorrected reads	No. of corrected and trimmed reads/% mapped	Avg depth of coverage (x)	GenBank accession no.
639-15	24 Feb 2015	Santa Cruz de la Sierra, Santa Cruz	438,120	98,367/100	3,834	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150092">MT150092</a>
4866-15	1 Mar 2015	Santa Cruz de la Sierra, Santa Cruz	495,057	110,170/>99.9	4,055	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150093">MT150093</a>
4990-15	1 Mar 2015	Santa Cruz de la Sierra, Santa Cruz	369,957	69,285/100	2,590	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150094">MT150094</a>
5037-15	26 Apr 2015	Cochabamba, Cochabamba	307,502	61,182/100	2,145	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150095">MT150095</a>
5041-15	4 May 2015	Santa Cruz de la Sierra, Santa Cruz	243,726	34,567/100	1,342	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150096">MT150096</a>
5046-15	4 May 2015	Santa Cruz de la Sierra, Santa Cruz	281,700	53,600/100	2,165	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150097">MT150097</a>
5038-15	4 May 2015	Santa Cruz de la Sierra, Santa Cruz	80,384	28,428/>99.9	922	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150098">MT150098</a>
746-15	4 May 2015	Santa Cruz de la Sierra, Santa Cruz	163,681	12,256/100	525	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150099">MT150099</a>
710-15	27 Apr 2015	Trinidad, Beni	128,959	4,657/>99.9	193	<a href="https://www.ncbi.nlm.nih.gov/nuccore/MT150100">MT150100</a>

tools, unless otherwise specified. The final sequence length for all nine genomes is 11,182 nucleotides (because each sequence was generated from homologous amplicons tiled across the coding region, they have the same start and endpoint), representing 99.6% of the nonstructural and structural coding regions. One sample (710-15) has uncalled bases due to poor coverage in the structural protein-coding region (163 nucleotides; 1.5%); all remaining sequences have no ambiguous bases. The average G+C content for all nine sequences is 50.7% (range, 50.5% to 50.7%).

We downloaded from GenBank the top 400 BLAST hits to sample 4866-15 (2 March 2020) and filtered out sequences without a month and year of sample collection. We aligned remaining sequences with our nine sequences using MAFFT (8), as implemented in Geneious, and trimmed the alignment to the coding region recovered in our sequences. We generated a maximum likelihood phylogeny using IQ-Tree v2.0-rc1 (9). We found that the nine Bolivian sequences are part of the widespread Asian-Caribbean chikungunya genotype and form a unique clade that was part of a larger monophyletic lineage primarily containing sequences from Nicaragua, Aruba, Colombia, and the United States (Fig. 1). The monophyly of our nine samples supports the hypothesis that a single lineage was widely circulating in Bolivia during the early 2015 chikungunya outbreak.

**Data availability.** Genome sequences are available in GenBank under accession numbers [MT150092](#) to [MT150100](#). Sequencing reads are available in the SRA database under BioProject accession number [PRJNA609363](#). The input, output, and complete maximum likelihood phylogenetic tree are available at <https://doi.org/10.6084/m9.figshare.11938047>.

## ACKNOWLEDGMENTS

Funding for this study came from a supplement award from NIH (grant 5D43TW010074) to R.H.G. (principal investigator [PI]), M.J.M., and J.R.L. (coinvestigator) and from start-up funds granted to M.J.M. from the University of Oklahoma.

We thank the Bolivian Ministry of Health and the Santa Cruz Departmental Health Service (SEDES-Santa Cruz) for supporting epidemiological work in Bolivia.

## REFERENCES

- Vargas-Cuentas NI, Roman-Gonzalez A, Yumin T. 2018. Spatial-temporal epidemiology study of Chikungunya disease in Bolivia. *Adv Astronaut Sci Technol* 1:69–80. <https://doi.org/10.1007/s42423-018-0014-4>.
- Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikasaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC, Jr., Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 12:1261–1276. <https://doi.org/10.1038/nprot.2017.066>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
- Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–214. <https://doi.org/10.1089/10665270050081478>.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Rang FJ, Kloosterman WP, de Ridder J. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* 19:90. <https://doi.org/10.1186/s13059-018-1462-9>.
- Katoh K, Kuma K-I, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33: 511–518. <https://doi.org/10.1093/nar/gki198>.
- Nguyen L-T, 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>.