Coding Complete Genome for the Mogiana Tick virus, a Jingmenvirus isolated from ticks in Brazil

Erika C Villa[a], Sandra R Maruyama[b], Isabel KF de Miranda-Santos[b], Gustavo Palacios[a] and Jason T Ladner[#a]

Corresponding author: jason.t.ladner.ctr@mail.mil

a Center for Genome Sciences, United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland, 21702, USA
b Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, SP, Brasil

Abstract
Mogiana tick virus (MGTV) is a segmented Jingmenvirus isolated in 2011 from cattle ticks in Brazil. Here, we present a coding complete genome for MGTV isolate MGTV/V4/11, including all four segments. MGTV is evolutionarily related to the Jingmen tick virus isolates SY84 and RC27.

The Jingmenviruses are a group of segmented, and likely multicomponent, viruses that are evolutionarily related to the unsegmented viruses of the genus *Flavivirus (1)*. Although the Jingmenvirus group has only recently been recognized, with the first virus descriptions published in 2014 (2, 5), known viruses in this group are diverse, globally distributed and capable of infecting a wide range of animal hosts (1, 2, 4–6). Here, we report the coding complete genome (i.e. only missing portions of terminal untranslated regions) (7) of MGTV/V4/11 isolated from cattle ticks (*Rhipicephalus microplus*) collected from Holstein bulls in Ribeirão Preto, state of São Paulo, Brazil (2). MGTV was one of the earliest Jingmenviruses to be reported, and at the time of publication, the segmented nature of the genome was not understood. Therefore, only the two genome segments with detectable sequence homology to flaviviruses were originally reported (2). We revisited the dataset of (Maruyama et al. 2014) and assembled coding complete sequences for all four genome segments.

We downloaded the raw Illumina sequence reads from the NCBI Short Read Archive (GenBank: SRA055953/SRR525284), which includes 11,898,134 paired-end reads, 2x150bp. We assembled the reads, *de novo*, using SPAdes v3.7.1 (8) and identified MGTV genome segments through sequence similarity (BLASTN) to the published genome of Jingmen tick virus (JMTV) isolate SY84 (GenBank: KJ001579-KJ001582). We then realigned the reads to the MGTV contigs using Bowtie2 v2.0.6 (9), recalled the consensus using Samtools v0.1.18 (10) along with custom scripts (https://github.com/jtladner/Scripts/blob/master/reference-based_assembly/consensus_fasta.py) and masked any positions with <3x coverage in support of the consensus. We used Cutadapt v1.9.dev1 (11) to remove Illumina adaptors, Prinseq-lite v0.20.3 (3) to trim and filter low-quality reads/bases and Picard (http://broadinstitute.github.io/picard) to remove duplicates. We only used bases with Phred quality score ≥20 in consensus calling.

The *de novo* assembly resulted in 534,445 contigs, four of which exhibited significant similarity to JMTV isolate SY84. Segments 1 and 3 were previously published as nonstructural protein 5 (Genbank: JX390986.1) and nonstructural protein 3 (Genbank: JX390985.1), respectively. We extended segment 1 by 573 nt on the 5' end and 543 nt on the 3' end, and we extended segment 3 by 241 nt on the 5' end and 60 nt on the 3' end. We found no mismatches between our assemblies and those of (Maruyama et al. 2014). Segments 2 and 4 were previously unpublished. MGTV/V4/11 contains putative open reading frames congruent with JMTV NSP1 (Seg 1), NSP2 (Seg 3), VP1 (Seg 2) and VP2-3 (Seg 4) (5), and genetic divergences between MGTV and JMTV SY84 and RC27 ranged from 9.7-12% at the nucleotide-level and from 3.2-7.6% at the amino acid-level. Given these similarities, MGTV may belong to the same species as JMTV.

Accession number(s). Genbank records JX390985 (segment 3) and JX390986 (segment 1) have been updated to coding complete segments. New annotations have been deposited into Genbank as accession numbers KY523073 (segment 2) and KY523074 (segment 4).

Virus	Segment*	Length	Dataset	Note	Accession
Mogiana Tick virus (MGTV)	1	2963	SRA055953	Updated, now with complete coding region	JX390986.2
	2	2629		Previously unpublished	KY523073
	3	2705		Updated, now with complete coding region	JX390985.2
	4	2728		Previously unpublished	KY523074

*Segment numbers for MGTV are according to Qin et al. 2014.

Acknowledgements
Work at the U.S. Army Medical Research Institute of Infectious Diseases was funded by the Defense Threat Reduction Agency, project 1881290. Opinions, interpretations, conclusions and recommendations are those of the authors and do not necessarily reflect the official policy or position of the U.S. Army.

References

1. **Ladner JT**, **Wiley MR**, **Beitzel B**, **Auguste AJ**, **Dupuis AP 2nd**, **Lindquist ME**, **Sibley SD**, **Kota KP**, **Fetterer D**, **Eastwood G**, **Kimmel D**, **Prieto K**, **Guzman H**, **Aliota MT**, **Reyes D**, **Brueggemann EE**, **St John L**, **Hyeroba D**, **Lauck M**, **Friedrich TC**, **O'Connor DH**, **Gestole MC**, **Cazares LH**, **Popov VL**, **Castro-Llanos F**, **Kochel TJ**, **Kenny T**, **White B**, **Ward MD**, **Loaiza JR**, **Goldberg TL**, **Weaver SC**, **Kramer LD**, **Tesh RB**, **Palacios G**. 2016. A Multicomponent Animal Virus Isolated from Mosquitoes. Cell Host Microbe **20**:357–367.

2. **Maruyama SR**, **Castro-Jorge LA**, **Ribeiro JMC**, **Gardinassi LG**, **Garcia GR**, **Brandão LG**, **Rodrigues AR**, **Okada MI**, **Abrão EP**, **Ferreira BR**, **da Fonseca BAL**, **de Miranda-Santos IKF**. 2014. Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in Rhipicephalus microplus ticks from Brazil. Mem Inst Oswaldo Cruz **109**:38–50.

3. **Schmieder R**, **Edwards R**. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**:863–864.

4. **Shi M**, **Lin X-D**, **Vasilakis N**, **Tian J-H**, **Li C-X**, **Chen L-J**, **Eastwood G**, **Diao X-N**, **Chen M-H**, **Chen X**, **Qin X-C**, **Widen SG**, **Wood TG**, **Tesh RB**, **Xu J**, **Holmes EC**, **Zhang Y-Z**. 2015. Divergent Viruses Discovered in Arthropods and Vertebrates Revise the Evolutionary History of the Flaviviridae and Related Viruses. J Virol **90**:659–669.

5. **Qin X-C**, **Shi M**, **Tian J-H**, **Lin X-D**, **Gao D-Y**, **He J-R**, **Wang J-B**, **Li C-X**, **Kang Y-J**, **Yu B**, **Zhou D-J**, **Xu J**, **Plyusnin A**, **Holmes EC**, **Zhang Y-Z**. 2014. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. Proc Natl Acad Sci U S A **111**:6744–6749.

6. **Webster CL**, **Waldron FM**, **Robertson S**, **Crowson D**, **Ferrari G**, **Quintana JF**, **Brouqui J-M**, **Bayne EH**, **Longdon B**, **Buck AH**, **Lazzaro BP**, **Akorli J**, **Haddrill PR**, **Obbard DJ**. 2015. The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. PLoS Biol **13**:e1002210.

7. **Ladner JT**, **Beitzel B**, **Chain PSG**, **Davenport MG**, **Donaldson EF**, **Frieman M**, **Kugelman JR**, **Kuhn JH**, **O'Rear J**, **Sabeti PC**, **Wentworth DE**, **Wiley MR**, **Yu G-Y**, **Threat Characterization Consortium**, **Sozhamannan S**, **Bradburne C**, **Palacios G**. 2014. Standards for sequencing viral genomes in the era of high-throughput sequencing. MBio **5**:e01360–14.

8. **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin AV**, **Sirotkin AV**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA**. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477.

9. **Langmead B**, **Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357–359.

10. **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.

11. **Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**:10–12.