AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

OMICS DATA SETS

A Divergent Strain of *Culex pipiens*-Associated Tunisia Virus in the Malaria Vector *Anopheles epiroticus*

Humberto J. Debat,[a] José M. C. Ribeiro[b]

[a]Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), Córdoba, Argentina

[b]Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases (NIAID-NIH), Rockville, Maryland, USA

ABSTRACT Here, we report the draft genome sequence of a divergent strain of *Culex pipiens*-associated Tunisia virus (CpATV) identified in the malaria vector *Anopheles epiroticus* (CpATV-AnE). CpATV-AnE expands the reference virus sequence, introducing an extended replicase with novel virga-like domains. Our results suggest that the host range of CpATV includes *Anopheles* sp. mosquitoes.

The mosquito *Anopheles epiroticus* is a relevant malaria vector in Southeast Asia (1). Despite its importance in human health, the study of this vector is scarce. Moreover, to our knowledge, no viruses have been associated with this mosquito. The extended use of next-generation sequencing (NGS) is providing a convenient and reliable source for the discovery of novel viruses (2). Here, we assessed an Illumina HiSeq 2000 transcriptome data set (SRA accession number SRX575685) corresponding to total RNA sequencing from wild-collected *A. epiroticus* mosquitoes of Vietnam (3). A total of 9,450,560 paired 100-nucleotide (nt) reads were trimmed of primer sequences using Trimmomatic and *de novo* assembled with Trinity v2.6.6 with standard parameters. The 25,271 transcripts obtained were subjected to BLASTX searches (E value, $<$1e$-$5) against the NCBI RefSeq virus protein database using the NCBI BLAST+ tool. A 1,862-nt contig obtained a significant hit (E value, 0; identity, 97%) with the replicase protein of the recently described *Culex pipiens*-associated Tunisia virus (CpATV; 4). A local BLASTN search using CpATV as a query resulted in the detection of eight more CpATV-like contigs which were extended and polished by iterative mapping of reads and reassembled into a 10,910-nt RNA virus sequence (supported by 1,777 reads; mean coverage, 16.1×), with overall 85.1% similarity and sharing the genomic architecture of CpATV. Strikingly, the previously reported genome of CpATV comprised only 6,816 nt. The ATG the authors postulated as a transcription start site at nucleotide position 331 appears to be an internal methionine which corresponds to a new nucleotide position, 4390, within the revised genome. The *A. epiroticus* strain of CpATV (CpATV-AnE) has four ORFs; ORF1 (62 to 7624-nt coordinates) encodes a 2,520-amino acid (aa) replicase protein (1,441 aa longer than the reported RefSeq, 96.2% identity), as predicted by NCBI ORF finder. Novel functional domains were predicted with the NCBI CD-Search tool (CDD v3.16 database; threshold, 0.001) and the ScanProsite tool covering the replicase, corresponding to an alphavirus-like methyltransferase, a *S*-adenosyl-L-methionine-dependent methyltransferase, a (+) RNA virus RNA helicase, and a tymovirus-like RNA-dependent RNA polymerase (RdRP). The methyltransferase domains additionally detected in CpATV-AnE are consistent with the domain architecture of related replicases of *Virgaviridae*-like viruses. CpATV-AnE has three more ORFs at nucleotide positions 7647 to 8060, 8086 to 10203, and 10286 to 10768, encoding three potential structural proteins of 137, 705, and 60 aa and sharing 94.2%, 86.5%, and 98.8% pairwise identity with reference CpATV proteins, respectively. The

Received 2 August 2018 **Accepted** 6 September 2018 **Published** 25 October 2018

Citation Debat HJ, Ribeiro JMC. 2018. A divergent strain of *Culex pipiens*-associated Tunisia virus in the malaria vector *Anopheles epiroticus*. Microbiol Resour Announc 7:e01026-18. https://doi.org/10.1128/MRA.01026-18.

Editor Irene L. G. Newton, Indiana University Bloomington

Address correspondence to Humberto J. Debat, debat.humberto@inta.gob.ar, or José M. C. Ribeiro, jribeiro@niaid.nih.gov.

CpATV-AnE 81.2-kDa protein encoded in ORF3 has an N-terminal secretory signal-peptide but lacks the tentative merozoite surface protein (MSP7)-like domain reported for CpATV (4). Both ORF2 and ORF4 have the expected *Tobacco mosaic virus*-like viral coat protein domains. Replicase-derived phylogenetic trees cluster both CpATV strains with a group of recently identified invertebrate *Virgaviridae*-like viruses, including Hubei virga-like virus 11 (HvlV11) which shares a mirroring genomic architecture with the revised CpATV genome. It is worth noting that HvlV11 was also initially reported in a truncated form (5) and revised in a subsequent study, which extended the RNA genome more than 4 kb (6). Specific analyses of both CpATV and HvlV11 truncated ends revealed that both comprise highly AT-rich regions (72% and 80%, respectively). This study highlights the importance of curation of RNA virus sequences obtained by NGS, which could be affected by primary sequence context and AT richness during the assembly process.

Data availability. The genome sequence reported here has been deposited at GenBank under the accession number MH634502.

ACKNOWLEDGMENTS

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. J.M.C.R. was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases.

This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

REFERENCES

1. Linton Y-M, Dusfour I, Howard TM, L FR, Manh ND, Dinh TH, Sochanta T, Coosemans M, Harbach RE. 2005. Anopheles (Cellia) epiroticus (Diptera: Culicidae), a new malaria vector species in the Southeast Asian Sundaicus complex. Bull Entomol Res 95:329–339. https://doi.org/10.1079/BER2005364.
2. Koonin EV, Dolja VV. 2018. Metaviromics: a tectonic shift in understanding virus evolution. Virus Res 246:A1–A3. https://doi.org/10.1016/j.virusres.2018.02.001.
3. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arca B, Arensburger P, Artemov G, Assour LA, Basseri H, Berlin A, Birren BW, Blandin SA, Brockman AI, Burkot TR, Burt A, Chan CS, Chauve C, Chiu JC, Christensen M, Costantini C, Davidson VLM, Deligianni E, Dottorini T, Dritsou V, Gabriel SB, Guelbeogo WM, Hall AB, Han MV, Hlaing T, Hughes DST, Jenkins AM, Jiang X, Jungreis I, Kakani EG, Kamali M, Kemppainen P, Kennedy RC, Kirmitzoglou IK, Koekemoer LL, Laban N, Langridge N, Lawniczak MKN, Lirakis M, Lobo NF, Lowy E, MacCallum RM, Mao C, et al. 2015. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science 347:1258522. https://doi.org/10.1126/science.1258522.
4. Bigot D, Atyame CM, Weill M, Justy F, Herniou EA, Gayral P. 2018. Discovery of Culex pipiens associated Tunisia virus: a new ssRNA(+) virus representing a new insect associated virus family. Virus Evol 4:vex040. https://doi.org/10.1093/ve/vex040.
5. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, Qin X-C, Li J, Cao J-P, Eden J-S, Buchmann J, Wang W, Xu J, Holmes EC, Zhang Y-Z. 2016. Redefining the invertebrate RNA virosphere. Nature 540:539–543. https://doi.org/10.1038/nature20167.
6. Debat HJ. 2017. An RNA virome associated to the golden orb-weaver spider *Nephila clavipes*. Front Microbiol 8:2097. https://doi.org/10.3389/fmicb.2017.02097.