


Complete genome sequences of two insect-specific flaviviruses

Jermilia Charles¹ · Chandra S. Tangudu¹ · Andrew E. Firth² · Bradley J. Blitvich¹ 

Received: 23 May 2017 / Accepted: 31 July 2017 / Published online: 16 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract We determined the complete genomic sequences of two previously discovered insect-specific flaviviruses, Marisma mosquito virus (MMV) and Nanay virus (NANV), using a combination of high-throughput sequencing, reverse transcription-polymerase chain reaction, 5' and 3' rapid amplification of cDNA ends and Sanger sequencing. Complete polyprotein amino acid sequence alignments revealed that the closest known relatives of MMV and NANV are Donggang virus (89% identity, 95% similarity) and Nounané virus (53% identity, 70% similarity), respectively. Phylogenetic inference is in agreement with these findings. Potential programmed –1 ribosomal frameshifting sites were bioinformatically identified in the genomes of both viruses.

Viruses in the genus *Flavivirus* (family *Flaviviridae*) can be divided into three distinct groups based on their host ranges and mode of transmission [2]. Dual-host flaviviruses are transmitted horizontally between hematophagous arthropods (i.e. mosquitoes and ticks) and vertebrate hosts. Viruses in the other two groups possess vertebrate-specific or insect-specific host ranges. Insect-specific flaviviruses (ISFs) are further divided into classical ISFs (cISFs) and dual-host affiliated flaviviruses (dISFs). Viruses in the

dISF group phylogenetically affiliate with dual-host flaviviruses despite their apparent insect-restricted phenotypes. Examples of dISFs include Marisma mosquito virus (MMV) and Nanay virus (NANV) in addition to Barkedji virus (BJV), Chaoyang virus (CHAOV), Donggang virus (DONV), Ilomantsi virus (ILOV), Lammi virus (LAMV), Nhumirim virus (NHUV) and Nounané virus (NOUV). MMV was originally isolated from *Ochlerotatus caspius* in Italy in 2001 to 2007 [17]. NANV was originally isolated from *Culex (Melanoconion) ocosa* in Peru in 2009 [6]. All these dISFs have had most, if not all, of their genomes fully sequenced aside from MMV and NANV for which partial envelope and/or NS5 gene sequence data are available. The objective of this study was to fully sequence the genomes of MMV and NANV.

MMV (isolate HU4528/07) and NANV (isolate PRD316/PER/09) were obtained from the World Reference Center for Emerging Viruses and Arboviruses at the University of Texas Medical Branch in Galveston, TX. MMV and NANV had been passaged four and at least five times, respectively in C6/36 (*Aedes albopictus*) cells prior to receipt and each virus underwent an additional passage in C6/36 cells at Iowa State University. Total RNA was extracted using Trizol Reagent (Invitrogen, Carlsbad, CA) and RNA was fragmented using RNase III and assessed for quality using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). Libraries were constructed using the Ion Total RNA-Seq Kit v2 (ThermoFisher, Carlsbad, CA) and barcoded using the Ion Xpress™ RNA-Seq Barcode 1-16 Kit (ThermoFisher). Libraries were assessed for quality and analyzed at the Genomic Technologies Facility at Iowa State University using an Ion Proton Sequencer (ThermoFisher). All ion-Torrent reads were mapped to *Ae. albopictus* and *Ae. aegypti* genomes using Bowtie 2 [13]. Unmapped reads were analyzed using the

Handling Editor: Patricia Aguilar.

✉ Bradley J. Blitvich
blitvich@iastate.edu

¹ Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA

² Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK

sortMeRNA program [12] to remove rRNA-related reads. All remaining reads with Phred values ≥ 33 were subjected to *de novo* SPAdes (ver 3.5.0) assembly [1]. Contigs were mapped to a reference flavivirus genome using LASTZ [9]. Alignment files were manually verified on TABLET [16]. Reverse transcription-polymerase chain reaction and Sanger sequencing were used to close gaps and confirm the accuracy of assembled contigs. The 5' and 3' ends of each genome were identified using 5' and 3' rapid amplification of cDNA ends, respectively. Briefly, a DNA adaptor (5'-rApp/TGGAATTCTCGGGTGCCAAGGT/ddC-3') was ligated to the viral genomic and anti-genomic RNAs using T4 RNA ligase (New England BioLabs, Ipswich, MA). Complementary cDNAs were created using SuperScript III (Invitrogen) and an adapter-specific primer. PCRs were performed using adapter- and gene-specific primers, and amplicons were subjected to Sanger sequencing.

The complete genome of MMV consists of 10,848 nt. (Genbank Accession No. MF139576) and contains a 5' untranslated region (UTR) of 119 nt., a long open reading frame (ORF) of 10,353 nt., and a 3' UTR of 376 nt (Figure 1a). The predicted amino acid sequence of the MMV polyprotein was aligned to other amino acid sequences in the Genbank database revealing that the closest known relatives of MMV are DONV (89% identity, 95% similarity) and ILOV (71% identity, 83% similarity). The complete genome of NANV is slightly smaller (10,804 nt; Genbank Accession No. MF139575) and contains a 5' UTR of 106 nt., a long ORF of 10,299 nt., and a 3' UTR of 399 nt (Figure 1b). Amino acid sequence alignments revealed that the closest known relatives of NANV are NOUV (53% identity, 70% similarity) and NHUV (52% identity, 69% similarity).

The phylogenetic placements of MMV and NANV, relative to other members of the *Flavivirus* genus, were assessed using the Bayesian Markov chain Monte Carlo based method implemented in MrBayes [15]. Complete polyprotein amino acid sequences were aligned using MUSCLE [5] and a phylogenetic tree was constructed

using MrBayes (Figure 2). MMV is most closely related phylogenetically to DONV. Both viruses belong to a distinct clade that also includes CHAOV, LAMV and ILOV. NANV is most closely related phylogenetically to NOUV and these two viruses belong to a distinct clade that also includes BJV and NHUV.

The sequences of MMV and NANV were inspected for potential -1 ribosomal frameshifting (-1 PRF) motifs because several groups of flaviviruses, including dISFs, appear to utilize -1 PRF during translation of their genomic RNA [2, 7, 10, 14]. Such frameshifting occurs at specific sites which normally comprise a slippery heptanucleotide sequence and a 3'-adjacent RNA structure. In eukaryotes, the consensus motif for the slippery heptanucleotide is X₃XXY₂YYZ, where X₃ represents any three identical nucleotides although a number of exceptions occur (such as GGA), YYY represents AAA or UUU, Z is A, C or U, and underscores represent codons in the original reading frame. The 3' RNA structure is normally a stem-loop or pseudoknot and is separated from the shift site by a spacer region of 5–9 nt. PRF has been predicted to occur in the NS2B region of CHAOV, DONV, LAMV, and ILOV [2, 10] and the ability of the identified motifs to stimulate -1 PRF has been verified in reporter constructs [8]. MMV contains a conserved G₃GAU₂UUU shift site sequence followed by a predicted RNA stem-loop structure in the NS2B region, as previously described for CHAOV, DONV and LAMV (Figure 3a). As noted previously, the -1 frame ORF varies considerably in length from six codons in DONV to 107 codons in CHAOV and LAMV. Similarly to DONV, MMV has a six-codon ORF.

The genome of NANV harbors a U₃UUU₂UUU potential shift site that aligns with the U₃UUU₂UUA shift site previously proposed in NOUV (Figure 3b), and tested in reporter constructs [8]. While the NOUV shift site is followed by a compact 13-bp stem-loop, the potential shift site in NANV is followed by a more extended potential stem-loop. Nonetheless, the stem-loop has a stable base

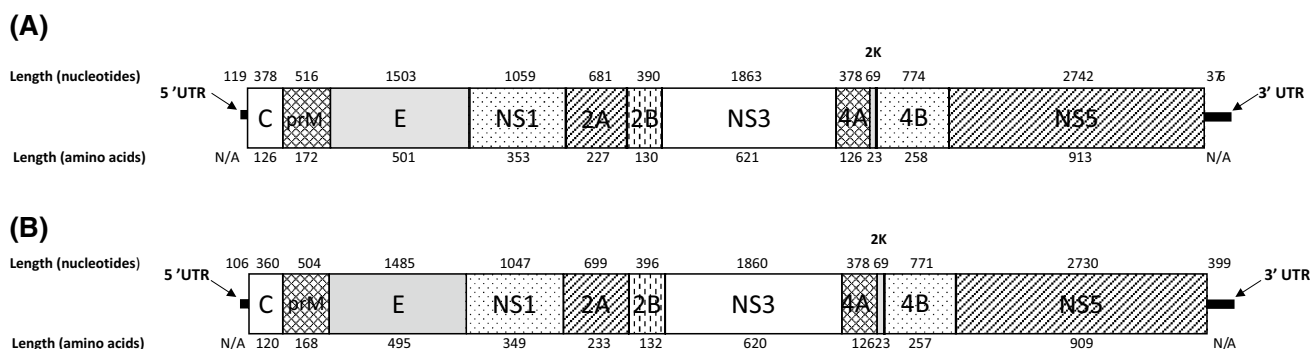


Fig. 1 Schematic representation of each flavivirus genome and polyprotein. Genomic organization of (A) Marisma mosquito virus and (B) Nanay virus. Lengths of the 5' and 3' untranslated regions as well as the structural and nonstructural protein genes are drawn to scale

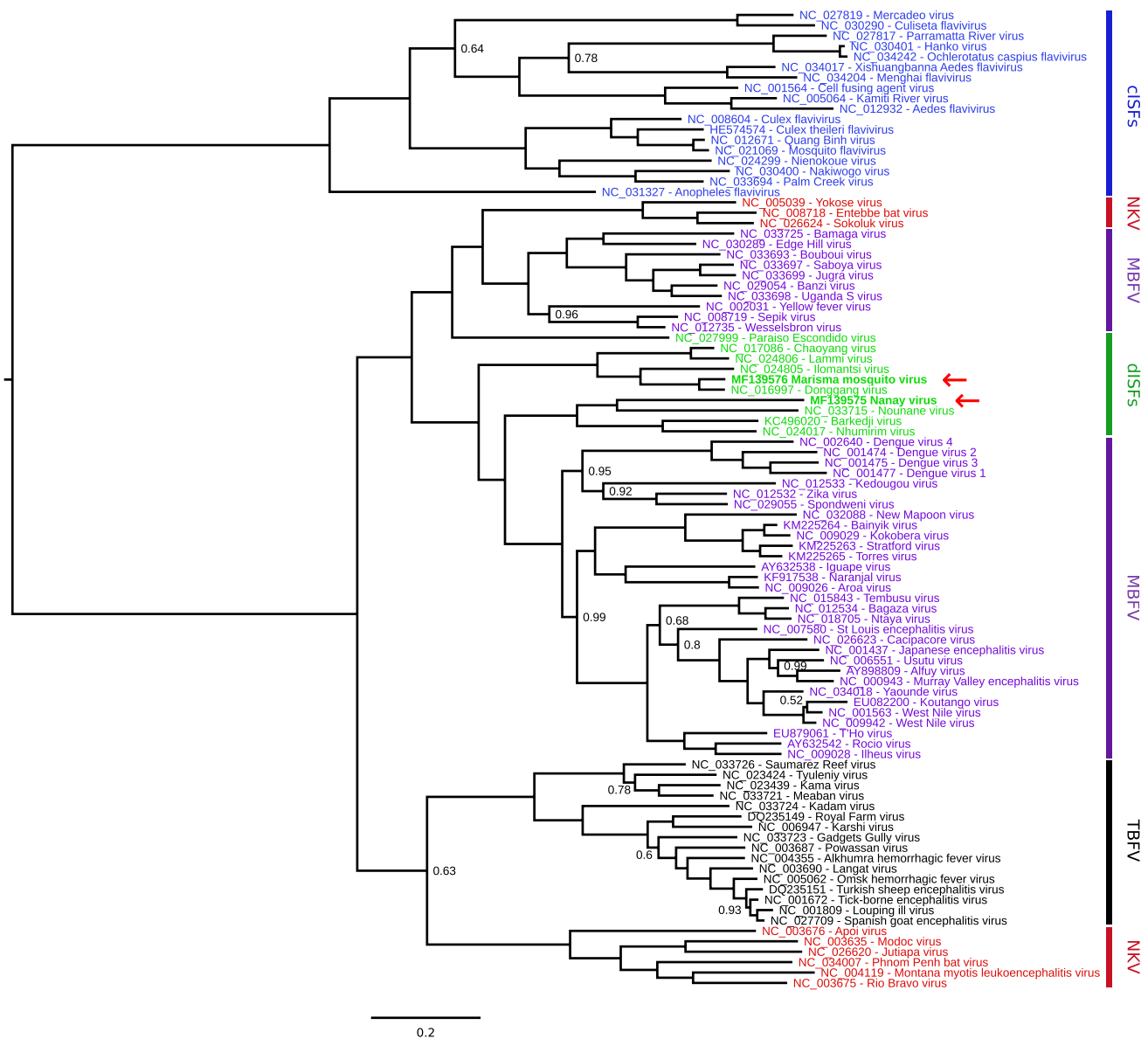


Fig. 2 Phylogenetic tree for genus *Flavivirus*. Complete polyprotein amino acid sequences were aligned using MUSCLE [5]. Regions of ambiguous alignment were excised using Gblocks [3] with default parameters, after which 1604 amino acid positions were retained. A maximum likelihood phylogenetic tree was estimated using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.3 [15] sampling across the default set of fixed

amino acid rate matrices, with one million generations, discarding the first 25% as burn-in. The figure was produced using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). The tree is midpoint-rooted, and nodes are labelled with posterior probability values where different from 1.00. Species names are color-coded as follows: cISFs—blue; dISFs—green; no known vector (NKV) flaviviruses—red; mosquito/vertebrate flaviviruses—purple; tick/vertebrate flaviviruses—black

(seven consecutive Watson-Crick pairs, four being G:C)s and is within the critical 5–9 nt. separation from the shift site. While the simple presence of a shift site and potential RNA structure should be viewed with caution (as they may occur spuriously, and not all RNA structures have the correct geometry to efficiently stimulate frameshifting [4]), the conservation of such features at a similar genomic location between related species lends weight to –1 PRF predictions. Thus, it seems plausible that NOUV

and NANV may represent another group of flaviviruses that utilize –1 PRF.

In summary, we report the complete genome sequences of two previously discovered ISFs, MMV and NANV, and provide bioinformatic evidence that both viruses utilize –1 PRF. It remains to be proven whether dISFs evolved from dual-host flaviviruses or are themselves the precursors but the former theory has been favored [11]. A rapidly growing number of ISFs have been discovered in recent years,

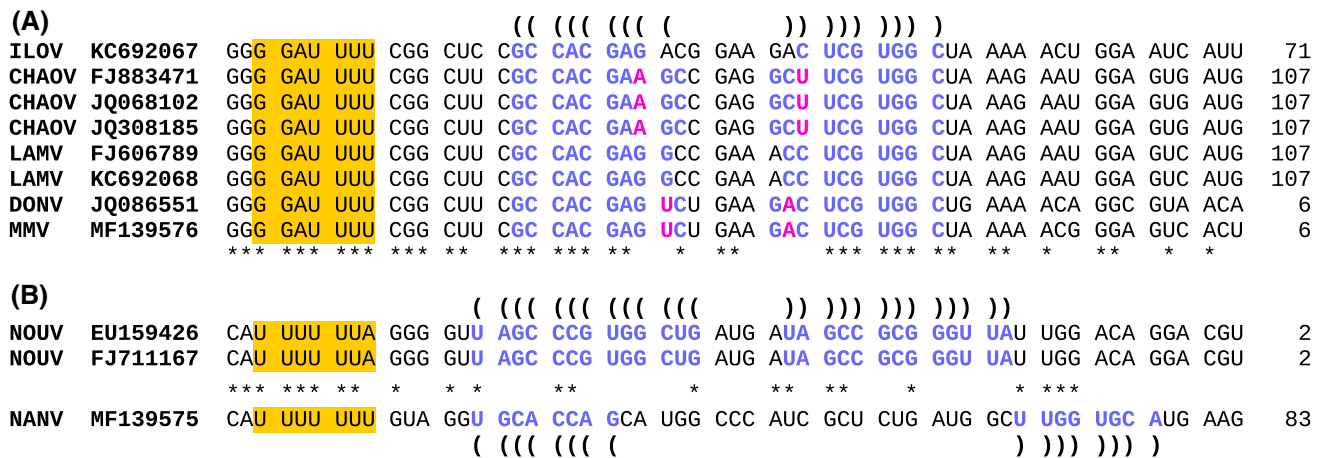


Fig. 3 Predicted sites of ribosomal frameshifting. (a) Marisma mosquito virus and other members of the Chaoyang/Lammi/Donggang/Ilomansi clade have a conserved G_GAU_UUU slippery heptanucleotide (orange highlight) followed by a predicted RNA stem-loop structure (blue letters) in the NS2B region. Compensatory substitutions (i.e. paired substitutions that preserve the predicted base-pair-

ings) are indicated in pink. Parentheses indicate the predicted base-pairings. Numbers in the last column show the length of the -1 frame ORF. (b) A potential shift site (orange highlight) and adjacent RNA stem-loop structure (blue letters), conserved in location but not sequence, in the NS2A region of Nanay virus and Nounane virus

and the availability of complete genome sequence data will allow for more robust comparative genomic studies between dual- and single-host flaviviruses and could, ultimately, provide novel insight into the evolutionary mechanisms that condition their differential host ranges and transmissibilities.

Acknowledgements This study was supported by a grant from the National Institutes of Health (R01AI114720). A.E.F. is supported by a Wellcome Trust grant (106207) and a European Research Council (ERC) European Union’s Horizon 2020 research and innovation programme grant (646891). The authors thank Robert Tesh for providing isolates of MMV and NANV.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The research reported here did not involve the use of human subjects or vertebrate animals.

Ethical statement This study represents original work that has not been submitted to any other journal for publication. The authors have no conflict of interest to declare. No human or animal ethics approval was required for the completion of this study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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
2. Blitvich BJ, Firth AE (2015) Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses* 7:1927–1959
3. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552
4. Chen X, Kang H, Shen LX, Chamorro M, Varmus HE, Tinoco I Jr (1996) A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA. *J Mol Biol* 260:479–483
5. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* 5:113
6. Evangelista J, Cruz C, Guevara C, Astete H, Carey C, Kochel TJ, Morrison AC, Williams M, Halsey ES, Forshey BM (2013) Characterization of a novel flavivirus isolated from *Culex* (Melanoconion) *ocossa* mosquitoes from Iquitos, Peru. *J Gen Virol* 94:1266–1272
7. Firth AE, Atkins JF (2009) A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1’ may derive from ribosomal frameshifting. *Virol J* 6:14
8. Firth AE, Blitvich BJ, Wills NM, Miller CL, Atkins JF (2010) Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. *Virology* 399:153–166
9. Harris R (2007) Improved pairwise alignment of genomic DNA. The Pennsylvania State University, State College

10. Huhtamo E, Cook S, Moureau G, Uzcategui NY, Sironen T, Kuivanen S, Putkuri N, Kurkela S, Harbach RE, Firth AE, Vapalahti O, Gould EA, de Lamballerie X (2014) Novel flaviviruses from mosquitoes: mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. *Virology* 464–465:320–329
11. Junglen S, Korries M, Grasse W, Wieseler J, Kopp A, Hermanns K, Leon-Juarez M, Drosten C, Kummerer BM (2017) Host range restriction of insect-specific flaviviruses occurs at several levels of the viral life cycle. *mSphere* 2(1):e00375–16. doi:[10.1128/mSphere.00375-16](https://doi.org/10.1128/mSphere.00375-16)
12. Kopylova E, Noe L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28:3211–3217
13. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
14. Melian EB, Hinzman E, Nagasaki T, Firth AE, Wills NM, Nouwens AS, Blitvich BJ, Leung J, Funk A, Atkins JF, Hall R, Khromykh AA (2010) NS1' of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *J Virol* 84:1641–1647
15. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542
16. Szlachetko J, Ferri D, Marchionni V, Kambolis A, Safonova OV, Milne CJ, Krocher O, Nachttegaal M, Sa J (2013) Subsecond and in situ chemical speciation of Pt/Al(2)O(3) during oxidation-reduction cycles monitored by high-energy resolution off-resonant X-ray spectroscopy. *J Am Chem Soc* 135:19071–19074
17. Vazquez A, Sanchez-Seco MP, Palacios G, Molero F, Reyes N, Ruiz S, Aranda C, Marques E, Escosa R, Moreno J, Figuerola J, Tenorio A (2012) Novel flaviviruses detected in different species of mosquitoes in Spain. *Vector Borne Zoonotic Dis* 12:223–229