

Project approval (Oct 23)

Source of Idea:

1.5 Theories of the Origins of Viruses	17
1.5.1 Viruses Are Remnants of Primeval Genetic Elements	18
1.5.2 Viruses Are the Result of Regressive Microbial Evolution	19
1.5.3 Viruses Are Liberated Autonomous Entities	20
1.5.4 Viruses Are Elements for Long-Term Coevolution	20
1.5.5 Viruses from Vesicles	21

Journal of General Virology (2004), 85, 1971–1980

DOI 10.1099/vir.0.79850-0

Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes

Sandrine Crochu,¹ Shelley Cook,² Houssam Attoui,¹ Remi N. Charrel,¹ Reine De Chesse,¹ Mourad Belhouche,¹ Jean-Jacques Lemasson,¹ Philippe de Micco¹ and Xavier de Lamballerie¹

¹Unité des Virus Emergents, Faculté de Médecine de Marseille, IFR48-IRD UR034, 27 boulevard Jean Moulin, 13005 Marseille, France

²Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Correspondence
Xavier de Lamballerie
Xavier.de-Lamballerie@
medecine.univ-mrs.fr

Flavivirus-related sequences have been discovered in the dsDNA genome of *Aedes albopictus* and *Aedes aegypti* mosquitoes, demonstrating for the first time an integration into a eukaryotic genome of a multigenic sequence from an RNA virus that replicates without a recognized DNA intermediate. In the *Aedes albopictus* C6/36 cell line, an open reading frame (ORF) of 1557 aa with protease/helicase and polyprotein processing domains characteristic of flaviviruses was identified. It is closely related to NS1–NS4A genes of the *Cell Fusing Agent* and *Kamiti River virus* and the corresponding mRNAs were detected. Integrated sequences homologous to the envelope, NS4B and polymerase genes of flaviviruses were identified. Overall, approximately two-thirds of a flavivirus-like genome were characterized. In the *Aedes aegypti* A20 cell line, a 492 aa ORF related to the polymerase of the *Cell Fusing Agent* and *Kamiti River virus* was identified. These flavivirus-related integrated DNA sequences were detected in laboratory-bred and wild *Aedes albopictus* and *Aedes aegypti* mosquitoes, demonstrating that their discovery is not an artefact resulting from the manipulation of mosquito cell lines, since they exist under natural conditions. This finding has major implications regarding evolution, as it represents an entirely different mechanism by which genetic diversity may be generated in eukaryotic cells distinct from accepted processes.

Received 25 November 2003
Accepted 23 February 2004

DOI 10.1099/vir.0.79850-0

The discovery that sequences of flavivirus-related RNA viruses persist in DNA form, integrated into the genome of *Aedes* spp. mosquitoes, supports the theory of *viruses as remnants of primeval genetic elements* (also known as the "**virus-first**" hypothesis). This theory suggests that viruses may have originated from ancient self-replicating entities that existed before the evolution of cellular life, acting as remnants of early genetic systems that predated or coincided with the formation of

modern cells. The integration of RNA virus sequences into the mosquito genome could be seen as an ancient evolutionary relic, showing that viruses have played a long-term role in shaping genomes and evolutionary processes (virus original event).

This finding also ties into the idea that viruses contribute to genetic diversity and evolution, which aligns with the co-evolutionary view, whereby viruses and host genomes have continuously influenced each other over millions of years. The integration of viral genetic material into host genomes is a mechanism that contributes to the diversification of host species and the persistence of viral genetic material across evolutionary time scales.

Task:

- Analyzing **evolutionary relationships between flavivirus-related sequences integrated into Aedes spp. genomes and other known flavivirus sequences**.
- explore how the integrated viral sequences evolved and their relationship to contemporary viruses.

Aim to address:

- How do the **integrated viral sequences** cluster in relation to **known flaviviruses**?
- What is the **evolutionary timeline of the integration** of these sequences into Aedes genomes?
- (From the paper) Can this data shed light on the co-evolution of viruses and mosquitoes?

General Workflow:

1. Data Collection:

- Input: Sequences of Aedes spp. mosquitoes containing flavivirus-related DNA (NCBI or downloaded manually).
- Do: Download these sequences and other relevant flavivirus sequences from databases.

2. Data Formatting:

- Input: Raw sequence data (FASTQ, FASTA).
- Do: Convert downloaded sequences into a standardized FASTA format if needed; will also filter and clean the data (remove low-quality sequences, redundant sequences, etc.).

3. Sequence Alignment:

- Input: Cleaned and formatted FASTA sequences.
- Perform multiple sequence alignments to find homology and prepare the data for phylogenetic analysis.

- Do: align flavivirus-related DNA sequences found in the mosquito genome with complete or partial reference flavivirus
- allow to observe the homology between the mosquito-integrated sequences and the reference flaviviruses—> how much of the viral genome has been integrated? whether the integrated sequences evolving? and how they compare to active viral lineages.

- ***By: MAFFT, MUSCLE. etc

- Query: **Flavivirus-Related Sequences in *Aedes* spp. Mosquitoes**

- integrated sequences within the genomes of *Aedes* spp. mosquitoes (like *Aedes aegypti* and *Aedes albopictus*) that were identified to be closely related to flaviviruses. (**NCBI or, if not readily available, extract the relevant regions from the mosquito genome using BLAST**)

- Reference: **Known Flavivirus Sequences:**

Align the integrated mosquito sequences to a set of reference flavivirus sequences, should be representative of different flavivirus lineages: **Dengue virus (DENV), Yellow Fever virus (YFV), West Nile virus (WNV), Cell Fusing Agent Virus (CFAV) (a mosquito-specific flavivirus), Kamiti River virus (KRV) (another mosquito-specific flavivirus)**

- Specific Genes or Genomic Regions to Align:

Focus on genes that are conserved across flaviviruses and are present in the integrated mosquito sequences, such as (mentioned in the paper):

NS3 (Protease-Helicase), NS5 (RNA-dependent RNA polymerase), Envelope (E) gene

These genes are critical for flavivirus replication and can provide strong phylogenetic signals when comparing across species.

4. Phylogenetic Analysis:

- Input: Aligned sequences.
- Do: Build a phylogenetic tree to visualize relationships between the sequences.
- By: IQ-TREE, FastTree. etc

(could experiment with different models of sequence evolution)

5. Integration event mapping (+1):

- Input: Phylogenetic tree and aligned sequences.
- Do: Compare the **integration points** of viral sequences across mosquito genomes and identify possible co-evolutionary trends.

- **map integration events across different mosquito species:** identifying the points of integration of flavivirus-related sequences across multiple mosquito genomes. could explore how frequently and where these integrations occurred—> are these shared across species or unique to certain populations.

-By:???? ancestral states estimation, R, APE, Geiger, phytools

6. Evolutionary Timeline (+2):

- Input: Phylogenetic data and metadata (e.g., divergence rates).
- Do: Estimate the timing of viral integration into mosquito genomes using molecular clock analysis.
- By:???? (BEAST2) —> time calibration analysis
- **molecular clock analysis** to estimate the evolutionary time when the viral sequences were integrated into the mosquito genomes. (could probably find potential correlations between environmental events and integration events)

7. Output:

- Final phylogenetic tree, co-evolution plots, and timeline analysis.
- generate figures

Dataset:

1. Flavivirus-related Sequences in Aedes spp. Mosquitoes:

- sequences described in the research paper (<https://pubmed.ncbi.nlm.nih.gov/15218182/>) and from the NCBI database for integrated viral sequences.

2. Known Flavivirus Sequences:

- sequences of related flaviviruses like Dengue, West Nile, Yellow Fever, etc.(from NCBI).

3. Reference Mosquito Genome:

- Use publicly available mosquito genome sequences (*Aedes aegypti* and *Aedes albopictus*, mentioned in the paper) for comparison purposes.
- need more from past papers