SUPPLEMENTARY INFORMATION

A single unidirectional piRNA cluster similar to the *flamenco locus* is the major source of EVE-derived transcription and small RNAs in *Aedes aegypti* mosquitoes

Eric Roberto Guimarães Rocha Aguiar[1,2], João Paulo Pereira de Almeida[1], Lucio Rezende Queiroz[1], Liliane Santana Oliveira[3], Roenick Proveti Olmo[1,4], Isaque João da Silva de Faria[1], Jean-Luc Imler[4], Arthur Gruber[2], Benjamin J Matthews[5], João Trindade Marques[1,4]*

[1]Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 30270-901, Brazil

[2]Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, BA, CEP 40101-909, Brazil

[3]Department of Parasitology, Instituto de Ciências Biomédicas, USP, São Paulo, SP, 05508-000, Brazil

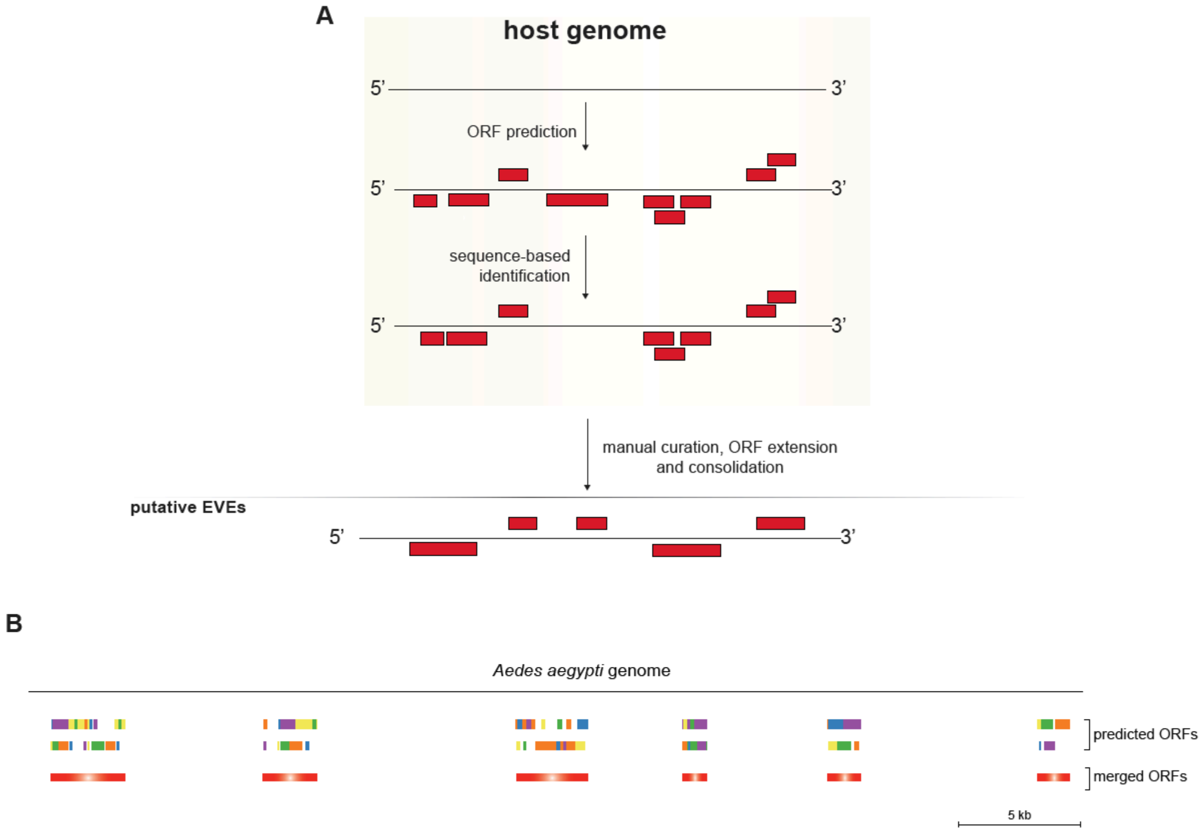
[4]Université de Strasbourg, CNRS UPR9022, Inserm U1257, 67084 Strasbourg, France.

[5]Department of Zoology, University of British Columbia, V6T 1Z4, Vancouver, Canada.

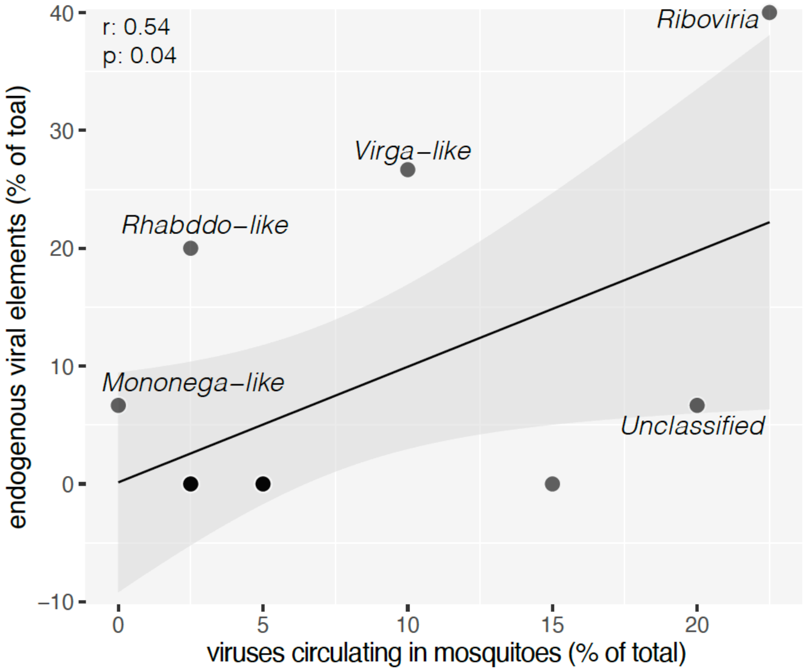
* Correspondence: jtm@ufmg.br

Running title: A *flamenco-like locus* in *A. aegypti* mosquitoes

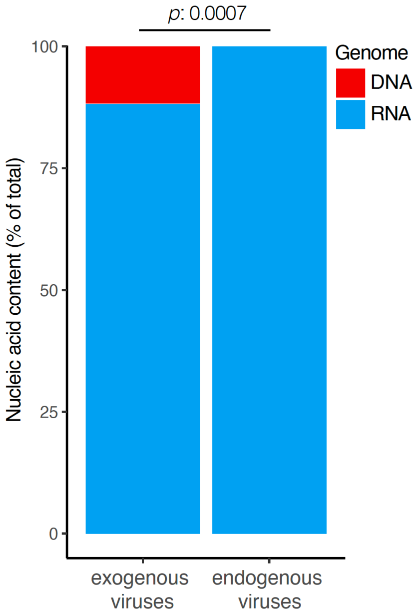
Keywords: Endogenous Viral Elements; EVE; *A. aegypti*; *flamenco locus*; RNA interference; piRNAs



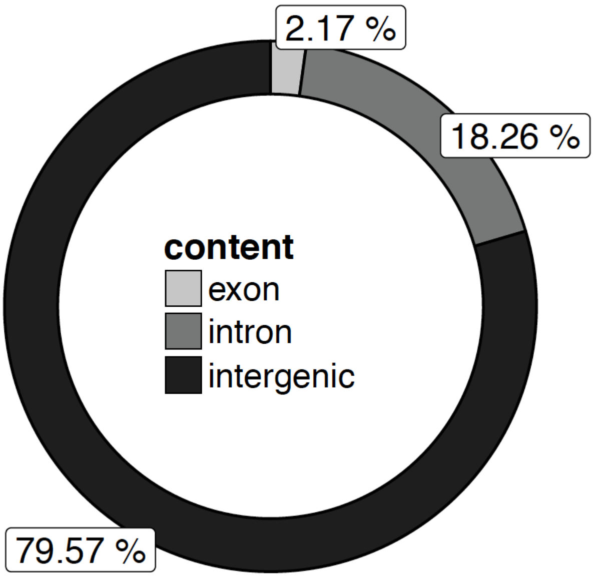
Supplemental Figure S1. Integrative strategy to identify EVEs in *A. aegypti* mosquitoes. (A) Approach applied to identify EVEs based on the reference genome that was used for de novo predictions of ORFs. Potential viral ORFs were identified based on sequence similarity to virus sequences in the GenBank database. EVEs were subjected to ORF extension when necessary (see panel B). Non-redundant EVEs were manually curated and annotated. (B) ORF extension consisted of merging adjacent EVEs (within 150 nt of each other) that were (1) related to the same exogenous virus and (2) annotated in the same genomic strand.



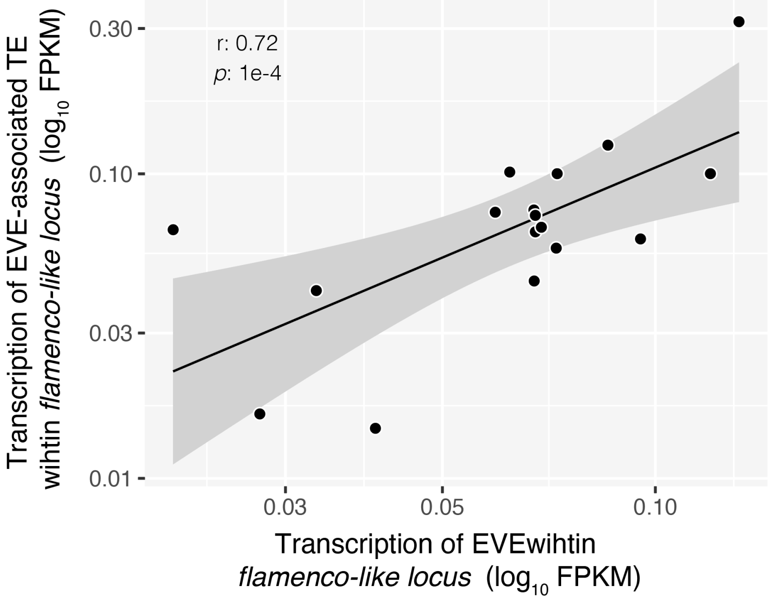
Supplemental Figure S2. Correlation between the abundance of EVEs and unclassified viruses. Scatter plot shows the correlation between EVEs and viruses that remain unclassified by ICTV. Correlation was carried out using *Pearson* correlation test. r (*Pearson* correlation) and *p* value are indicated.



Supplemental Figure S3. Sequences derived from DNA viruses are underrepresented in EVEs. Enrichment was performed using Fisher’s exact test.



Supplemental Figure S4. Percentage of exon, intron and intergenic regions in the *Aedes aegypti* genome. Analysis was performed considering genomic regions of *A. aegypti* genome version AaegL3.



Supplemental Figure S5. Transcription of adjacent EVEs and TEs within EVE cluster 38 are correlated. r (*Pearson* correlation) and *p*-value are indicated.