G + C content differs in conserved and variable amino acid residues of flaviviruses and other evolutionary groups

Raphaëlle Klitting, Ernest Andrew Gould, Xavier de Lamballerie

PII: S1567-1348(16)30400-2
DOI: doi: 10.1016/j.meegid.2016.09.017
Reference: MEEGID 2931

To appear in:

Received date: 8 July 2016
Revised date: 1 September 2016
Accepted date: 19 September 2016

Please cite this article as: Klitting, Raphaëlle, Gould, Ernest Andrew, de Lamballerie, Xavier, G + C content differs in conserved and variable amino acid residues of flaviviruses and other evolutionary groups, (2016), doi: 10.1016/j.meegid.2016.09.017

G+C content differs in conserved and variable amino acid residues of flaviviruses and other evolutionary groups

Raphaëlle Klitting[A], Ernest Andrew Gould[A] and Xavier de Lamballerie[A,B]

A. UMR "Emergence des Pathologies Virales" (EPV: Aix-Marseille university - IRD 190 - Inserm 1207 - EHESP), 27 bd Jean Moulin, 13385, Marseille, France

B. Institut hospitalo-universitaire Méditerranée-Infection, 27 bd Jean Moulin, 13385, Marseille, France

Author's e-mail : raphaelle.klitting@posteo.de, eag@ceh.ac.uk, xavier.de-lamballerie@univ-amu.fr

Corresponding author : R. Klitting **E-mail:** raphaelle.klitting@posteo.de

Abstract

Flaviviruses are small RNA viruses that exhibit genetic and ecological diversity and a wide range of G+C content (GC%[1]). We discovered that, amongst flaviviruses, the GC% of nucleotides encoding conserved amino acid (AA) residues was consistently higher than that of nucleotides encoding variable AAs. This intriguing phenomenon was also identified for a wide range of other viruses, and some non-viral evolutionary groups. Here, we analyse the possible mechanisms underlying this imbalanced nucleotide content (in particular the role of the specific G content and the AA composition in flaviviral genomes) and discuss its evolutionary implications. Our findings suggest that one of the most simple characteristics of the genetic code (*i.e.*, the G or G+C content of codons) is linked with the evolutionary behaviour of the corresponding encoded AAs.

Keywords: Flavivirus, Evolution, G+C content

---

[1] *Abbreviations :* AA, Amino Acid ; APOIV , Apoi virus; CA-Cons, Complete Alignment Conserved; CA-Var , Complete Alignment Variable; CDS , complete coding sequences; CXFV , Culex-borne flavivirus; DENV, dengue virus; GC%, G+C content ; ISFV , insect specific flaviviruses; JEV, Japanese encephalitis virus; MBFV , mosquito-borne flaviviruses ; MODV , Modoc virus; MMLV , Montana Myotis Leukoencephalitis virus; NKV , viruses with no known vector; ORF, open reading frame; PW , pairwise; RBV , Rio Bravo virus; TBEV, tick-borne encephalitis virus; TBFV , tick-borne flaviviruses; WNV, West Nile virus; YFV , yellow fever virus