Biochemical and Biophysical Research Communications xxx (2017) 1–7

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Host influence in the genomic composition of flaviviruses: A multivariate approach

Diego Simón [a], Alvaro Fajardo [b], Martín Sóñora [b], Adriana Delfraro [c], Héctor Musto [a, *]

[a] Laboratorio de Organización y Evolución del Genoma, Unidad de Genómica Evolutiva, Facultad de Ciencias (FC), Universidad de la República (UDELAR), Iguá 4225, Montevideo 11400, Uruguay
[b] Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, FC, UDELAR, Uruguay
[c] Sección Virología, FC, UDELAR, Uruguay

ARTICLE INFO

Article history:
Received 16 May 2017
Received in revised form
9 June 2017
Accepted 15 June 2017
Available online xxx

Keywords:
Flavivirus
Base composition
Dinucleotides
Codon usage
Amino acids

ABSTRACT

Flaviviruses present substantial differences in their host range and transmissibility. We studied the evolution of base composition, dinucleotide biases, codon usage and amino acid frequencies in the genus *Flavivirus* within a phylogenetic framework by principal components analysis. There is a mutual interplay between the evolutionary history of flaviviruses and their respective vectors and/or hosts. Hosts associated to distinct phylogenetic groups may be driving flaviviruses at different pace and through various sequence landscapes, as can be seen for viruses associated with *Aedes* or *Culex* spp., although phylogenetic inertia cannot be ruled out. In some cases, viruses face even opposite forces. For instance, in tick-borne flaviviruses, while vertebrate hosts exert pressure to deplete their CpG, tick vectors drive them to exhibit GC-rich codons. Within a vertebrate environment, natural selection appears to be acting on the viral genome to overcome the immune system. On the other side, within an arthropod environment, mutational biases seem to be the dominant forces.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

The genus *Flavivirus* belongs to the family *Flaviviridae,* together with *Hepacivirus*, *Pegivirus* and *Pestivirus*. According to the International Committee of Virus Taxonomy, the genus comprises 53 species with wide global distribution, as well as an increasing number of unclassified or tentative species [1]. They are positive-sense single-stranded RNA viruses of about 11 kb, with a 5′ type I cap structure and lacking a poly(A) tail at the 3′ end. Their genome is translated in a single polyprotein which is cleaved in three structural proteins (C, prM and E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [2].

Despite the similarity in their genomic organization, there are substantial differences in the host range and transmissibility among them. Most known species are arboviruses, which are transmitted horizontally between hematophagous arthropods and susceptible vertebrate hosts, and are classified in mosquito-borne flaviviruses (MBFV) and tick-borne flaviviruses (TBFV). However, some species only replicate in bats or rodents with not-known vector associated to them (NKV). Furthermore, several species only infect mosquitoes, which are referred to as insect-specific flaviviruses (ISFV) [3–9].

The taxonomic relationship among flaviviruses has been extensively investigated through different approaches, originally based on antigenic cross-reactivity in neutralization, complement fixation and hemagglutination inhibition assays [10,11]. Lately, phylogenetic reconstructions based on nucleotide and amino acid sequences allowed a deeper understanding of the diversity of the genus. Several methodological approaches were followed to analyze different genes and complete coding regions [3–8,12–14]. As a result of these efforts it was found that the general pattern of the inferred phylogenetic relationships correlates with the main epidemiological aspects as host range, vectors and related diseases. Nevertheless, the comparison of different analyses evidences phylogenetic incongruences that difficult a proper definition of the taxonomic relationships.

Another approach to understand both the evolution and the phylogenetic relationships, is to analyze compositional properties of each virus, such as base composition, dinucleotide biases, codon

* Corresponding author.
E-mail addresses: dsimon@fcien.edu.uy (D. Simón), afajardo@cin.edu.uy (A. Fajardo), msonora@cin.edu.uy (M. Sóñora), adelfraro@gmail.com (A. Delfraro), hmusto@gmail.com (H. Musto).

http://dx.doi.org/10.1016/j.bbrc.2017.06.088
0006-291X/© 2017 Elsevier Inc. All rights reserved.

Please cite this article in press as: D. Simón, et al., Host influence in the genomic composition of flaviviruses: A multivariate approach, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.06.088

usage and amino acid frequencies. These features can be defined as molecular signatures. Taking advantage of the great number of sequences available, in the present communication we update the analyses of these genomic compositional properties in the genus *Flavivirus* within a phylogenetic framework.

2. Materials y methods

2.1. Dataset construction

Coding sequences (CDS) available from all viral types belonging to the genus *Flavivirus* were retrieved from the ViPR database (Virus Pathogen Database and Analysis Resource) of the National Institute of Allergy and Infectious Diseases, available at http://www.viprbrc.org [15]. Tamana bat virus, an unclassified *Flavivirus*, was excluded from these analyses for being highly divergent [16]. Information about hosts (arthropod and/or vertebrate), status as arbovirus and human pathogenicity were obtained from Arbovirus Catalog, Virus-Host DB and/or ViPR. For details, see Supplementary Table 1.

2.2. Phylogenetic analyses

Polyprotein sequences of each species were aligned with MUSCLE [17]. Regions of ambiguous alignment were excised with GBlocks v.0.91b [18,19]. The optimal amino acid substitution model was inferred with ProtTest 3 [20]. With the selected evolutionary model (WAG + F + G + I), maximum likelihood phylogenetic trees were constructed through PhyML v.3.0 [21]. A bootstrap test with 500 replicates was used to evaluate the robustness of each node. FigTree v1.4.2 (available at http://tree.bio.ed.ac.uk/software/figtree) was used to edit the final trees.

2.3. Hosts genomes

Aedes, *Culex* and *Ixodes*, as the main arthropod hosts, and *Gallus*, *Homo* and *Mus*, representing the three major groups of vertebrate hosts in the sample (fowl, primates and rodents, respectively), were analyzed in a bigger extent. The CDS were obtained from Ensembl repositories [22]: *H. sapiens* and *M. musculus*, from CCDS project; *G. gallus*, from Ensembl genome browser 88; *A. aegypti*, *C. pipiens* and *I. scapularis*, from Ensembl Metazoa.

2.4. Compositional analysis

For each viral genome, different compositional properties were obtained with R package seqinr [23]. Only the CDS of the whole polyproteins were considered, discarding some other putative proteins. The same properties were obtained also for the set of hosts.

2.5. Principal component analysis

Principal component analysis (PCA) is a statistical procedure to reduce the multidimensionality of the data. PCA allows the observation of patterns, either forming clusters or gradients. Interpreting the weights given by the original numerical variables is also informative. In order to infer associations between the axes and some compositional features, Pearson correlations were tested. PCA were performed in R with function pr.comp [24]. The proportion of variance explained by the ten principle components are displayed as Supplementary Fig. 2. Graphical representations were constructed with scatter3dplot package [25].

3. Results and discussion

3.1. Frequency of bases

Fig. 1 shows the inferred phylogenetic relationships between the different flaviviruses analyzed in this study. The sequences cluster in four monophyletic groups: MBFV (blue), NKV (green), TBFV (red) and ISFV (purple) as previously described [7,17,25].

The complete compositional analyses are presented as Supplementary Table 2. Overviewing some descriptive statistics of these analyses, the four main groups display heterogeneity in their base composition. In relation to the other groups, ISFV are A/G-poor and U-rich, MBFV are G/U-rich and A-poor, while NKV are G/C-poor and A/U-rich. Finally, TBFV are A/U-poor and G-rich. ISFV are characterized by their relatively high pyrimidines content, opposite to the rest of the groups which tend to be enriched in purines. TBFV are clearly the richest in GC%, which stands for all codon positions while the contrary is true for NKV. These patterns agree with previous studies [26,27] and also coincide with the observed values in their hosts (see Table S3). For instance, *Ixodes*, the predominant vector within TBFV group, has a mean coding GC (cGC) of 58% (and a GC3 of 72%). *Culex* spp. also exhibit high GC% (cGC of 55%; GC3 of 69%), while *Aedes* spp. presents a mean cGC slightly lower than vertebrate hosts analyzed (i.e.; *Gallus*, *Homo* and *Mus*).

The differences in base composition between mosquitoes are also extensive to those flaviviruses that replicates in these insects. The mean cGC values for all flaviviruses which have either *Aedes* or *Culex* spp. as their main hosts are 49.3% ± 2.3 and 50.5% ± 1.3, respectively. ISFV includes two distinct clades: ISFV-*Aedes* (49.1% ± 2.0) and ISFV-*Culex* (51.1% ± 2.1); both subgroups are mainly associated with each genus. Although these intervals overlap, the trend is consistent and suggests an effect of the GC% of the mosquito host in the viral base composition.

Other groups of viruses that are believed to be insect-specific have been described [3]. Two distinct clades, phylogenetically related to MBFV, seem to circulate strictly in mosquitoes since none have been isolated from vertebrates or cell lines derived from them [28,29]. The species within MBFV that shows an ISFV-like phenotype could be separated in two clades; ISFV-2: CHAOV, DONV, ILOV, and LAMV; and ISFV-3: BARKV, NHUV, and NOUV. The topology observed suggests that the emergence of these clades occurred in two independent events, as was previously suggested [3,30]. Moreover, these groups are also primarily associated with aedine (ISFV-2) or culicine (ISFV-3) mosquitoes (see Supplementary Table 1). ISFV-2 (48.9% ± 0.6) have lower mean cGC values than ISFV-3 (50.9% ± 1.1).

3.2. Dinucleotides

Concerning dinucleotide observed/expected values (see Supplementary Table 2), our results confirm previous reports [31], which can be summarized as follows: UpA is consistently underrepresented in all flaviviruses; TBFV have the lowest mean values (0.46 ± 0.03), while the other groups present similar values among them. The three groups associated with a vertebrate host (i.e.; MBFV, NKV and TBFV) also present biases in CpG, being NKV the most biased (0.34 ± 0.05); the mean values of ISFV-2 (0.66 ± 0.03) and ISFV-3 (0.65 ± 0.05) were intermediate between MBFV (0.52 ± 0.09) and ISFV (0.86 ± 0.09). CpA and UpG are the dinucleotides with more over-represented values.

The biases in CpG and UpA, which are almost universal for vertebrate RNA viruses, have been associated with an antiviral effect [27,31–34], although the mechanisms remain unclarified.

Please cite this article in press as: D. Simón, et al., Host influence in the genomic composition of flaviviruses: A multivariate approach, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.06.088

Download English Version:

https://daneshyari.com/en/article/5504704

Download Persian Version:

https://daneshyari.com/article/5504704

Daneshyari.com