MOL2NET, International Conference Series on Multidisciplinary Sciences
http://sciforum.net/conference/mol2net-03

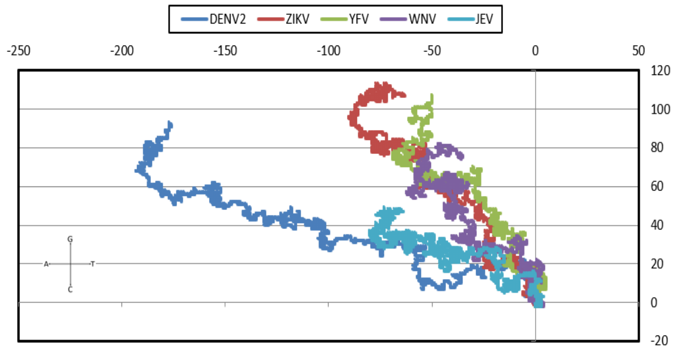
Comparison of Base Distributions in Dengue, Zika and Other Flavivirus Envelope and NS5 Genes

Sumanta Dey[a], Proyasha Roy[a], Ashesh Nandy (Email: anandy43@yahoo.com)[a], Subhash C Basak[b] and Sukhen Das[c]

[a] Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700058, India

[b] University of Minnesota Duluth-Natural Resources Research Institute and Department of Chemistry and Biochemistry, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA

[c] Department of Physics, Jadavpur University, Jadavpur, Kolkata 700032, India



Abstract

Among the nucleotide sequences of flaviviruses, those of the Zika virus and Dengue type 2 virus are believed to share a high degree of similarity. Our study of the nucleotide sequences of the envelope and NS5 genes shows that the sequences of the Dengue type 2 virus are sharply different compared to other human infecting flaviviruses. This is emphatically seen in a 2D graphical representation and distinctly discriminated in terms of relevant RNA descriptors. In this report, we demonstrate this difference through various parameters and consider possible reasons for such variations that seem to have been largely neglected in the literature.

Introduction

The recent epidemic of the Zika virus in South and Central Americas have focused attention on the flavivirus group of viruses which also include all four types of Dengue virus (DENV1, DENV2, DENV3, DENV4), West Nile virus (WNV), Yellow fever virus (YFV) and Japanese encephalitis virus (JEV) besides Zika virus (ZIKV). All these viruses share strong homology at the protein sequence level, and slightly weaker homology at the nucleotide sequence level [1]. Phylogenetic studies place the four DENV serotypes in one clade and, ZIKV is found to be closest to DENV2. It is therefore widely held that ZIKV and DENV2 have the closest relationship among all the aforementioned flaviviruses [2].

Closer examination of the gene and protein sequences of these viruses reveals slightly divergent results. Taking one structural and one non-structural gene, viz., envelope (E) and NS5, as examples for this study, we have found sequence differences between DENV2 and other flaviviruses to be reasonably correct in the case of the protein sequences, although the DENV2 sequence has a fair excess of the lysine amino acids. However, the DENV2 nucleotide sequence of these genes have a widely different base distribution compared to the ZIKV as seen in a 2D graphical representations. In this brief report, we enumerate the various analyses we have done and explore the reasons and ramifications of the differences.

Materials and Methods

The flaviviruses investigated here include 5 sequences selected for our analyses of the envelope (E) genes of DENV2, ZIKV, WNV, YFV and JEV and their corresponding NS5 sequences. DENV2 was particularly chosen because in the literature, it is cited as being phylogenetically closest to ZIKV [2]. The envelope and NS5 genes were selected primarily due to the fact that the envelope protein is responsible for host cell fusion and endocytosis, whereas the NS5 gene encoding the RNA dependent RNA polymerase is important in gene replication [2]; that these are the longer genes in the viral genome also helps in the analysis.

Base distribution and quantification were done using one of the 2D graphical representation methods [3] where an RNA sequence is represented in a 2D grid and for each base, a point is plotted by moving one unit in the negative x-direction if it is an adenine, in the y-direction if it is a guanine, in the x-direction for a thymine and in the negative y-direction if it is a cytosine. Plotting these points successively for each base in the sequence generates a graph that represents the base distribution in the sequence. Using the x, y values to get a weighted center of mass (c.m.), μ_x and μ_y, we can compute a graph radius, g_R, from the origin to the c.m. The g_R becomes representative of the overall spread of the graph and can be considered as the RNA sequence descriptor [4]. We show the base distribution of the flavivirus genes in this 2D representation and compute the sequence descriptors, g_R, for comparison purposes. Comparison is also made between the gene sequences using CLUSTALW [5].

Results and Discussion

The distinction between the nucleotide sequences of DENV2 and the other flaviviruses is perspicuously seen in the 2D graphical representations (Fig. 1 and 2). A large excess of adenine and guanine exist in both the envelope and NS5 genes of DENV2 (Table 1). The Base Distribution Index or Sequence Descriptor, g_R, indicates the magnitude of the spread of the nucleotide sequence graphs. The average g_R values computed for the envelope and NS5 genes of ZIKV, YFV, JEV and WNV are 61.38 ± 11.97 and 181.35 ± 7.26, respectively, both of which are overshot by a large margin by DENV2 - 120.60 for envelope gene and 249.98 for NS5 gene. Such large variations indicate a markedly different codon bias profile of DENV2 on comparing with the other flaviviruses (see Fig. 3 for the graph of the codon bias in envelope genes, data not shown).

Interestingly, a BLAST pairwise analysis of the envelope genes shows only 61% identity between DENV2 and ZIKV, but that can be misleading since a similar analysis between two totally unrelated genes, DENV2 envelope gene (1485 nucleotides) and H5N1 neuraminidase (1410 nucleotides), shows 50% relative identity. This implies that such kind of analysis is not a reliable indicator of identity. Detailed study of the nucleotide sequences shows the 2D graphical representation to be a pointer to the true differences. This aspect seems to have been little explored in the literature.

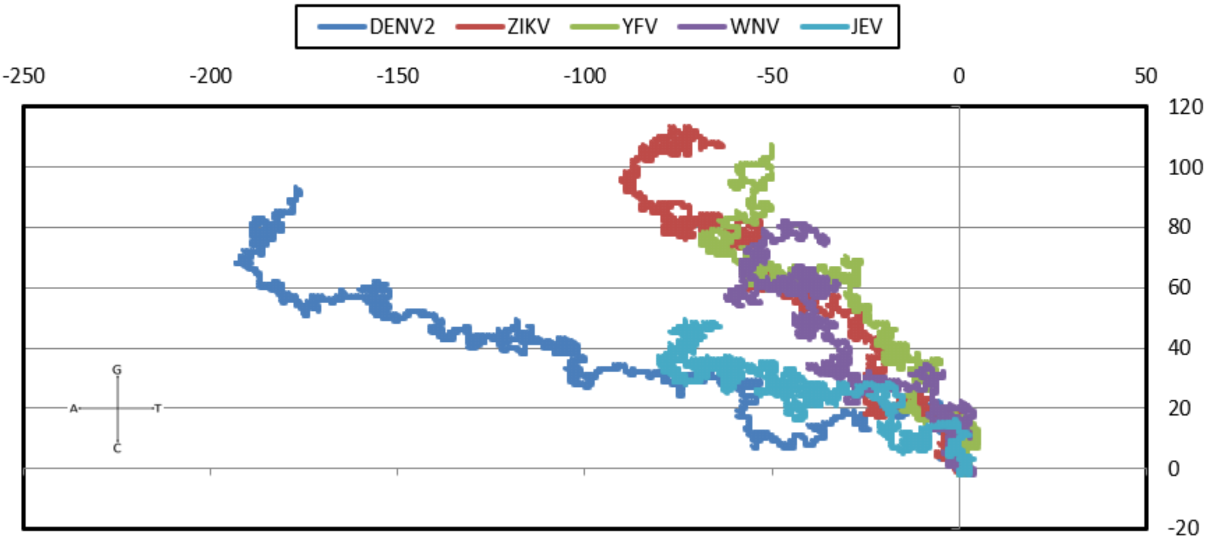


Fig. 1. Flavivirus envelope genes in a 2D graphical representation. Axes AGTC. Deep blue - DENV2, Light blue - JEV, Purple - WNV, Green – YFV, Red – ZIKV.

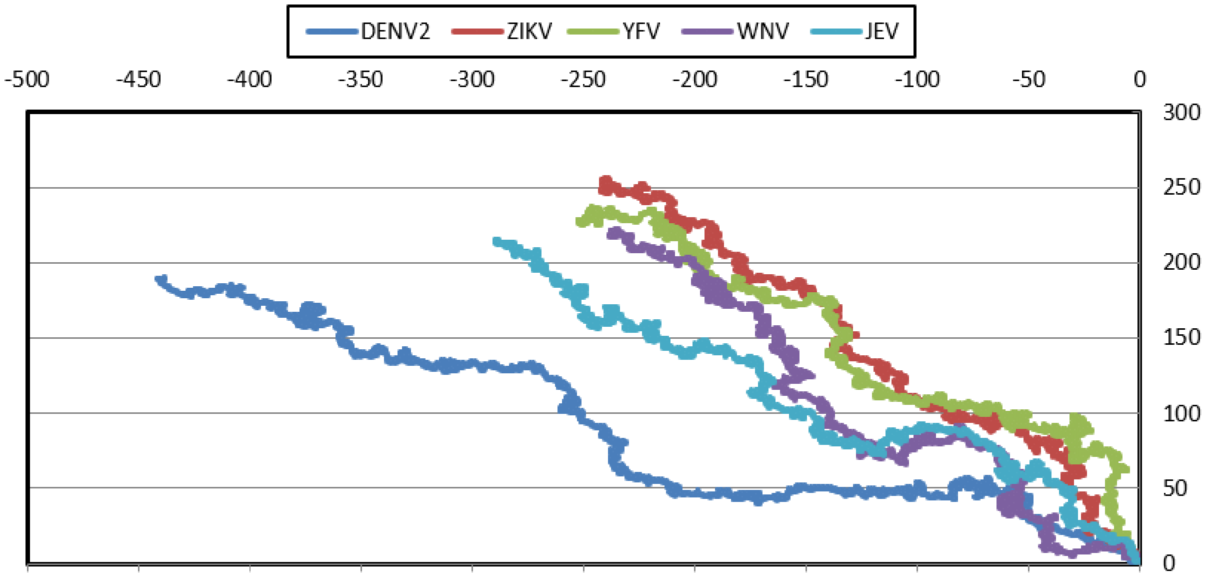


Fig. 2. Flavivirus NS5 genes in a 2D graphical representation. Axes AGTC. Deep blue - DENV2, Light blue - JEV, Purple - WNV, Green – YFV, Red – ZIKV.

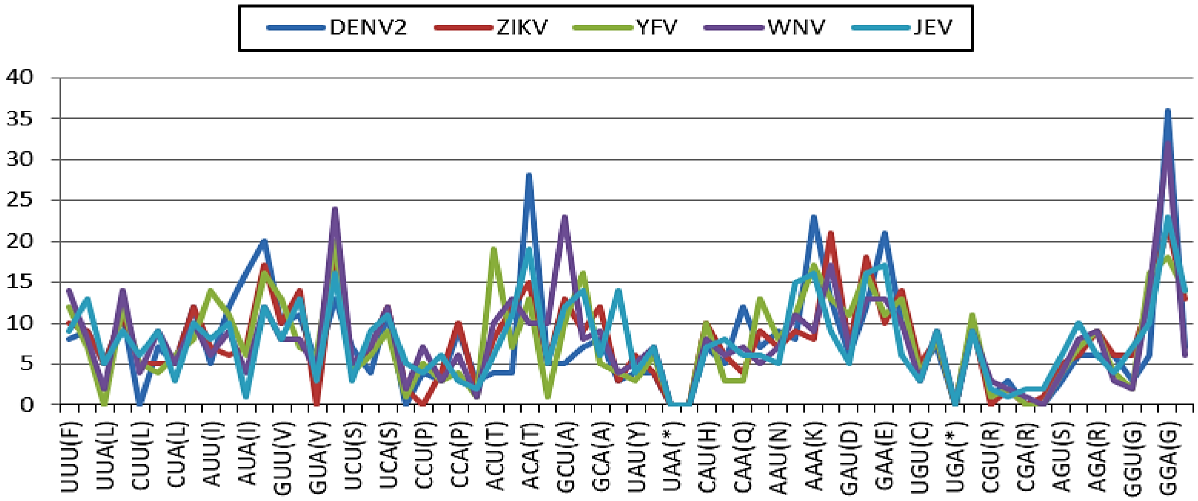


Fig. 3. Flavivirus envelope codon bias

Table 1. Nucleotide Composition and g_R Values

Virus	**Envelope Gene**					**NS5 Gene**				
	A	**C**	**G**	**T**	g_R	**A**	**C**	**G**	**T**	g_R
DENV2	494	291	382	317	120.5958	949	525	715	511	249.9799
ZIKV	400	325	432	337	78.43309	780	566	821	542	182.0870
YFV	405	306	413	355	60.98772	802	563	794	556	181.9147
JEV	319	368	415	326	52.99962	801	594	807	513	189.5519
WNV	390	342	417	354	53.06449	770	595	816	534	171.8647

The protein sequences do not express as much divergence as is shown by the nucleotide sequences. The CLUSTALW pairwise alignment scores of protein and nucleotide sequences of envelope and NS5 are given in Table 2. Although, it is noted that the proteins of DENV2 constitute a slightly higher composition of lysine (K) and threonine (T) arising from the codon usage bias, the availability of synonymous codons largely maintains the close similarity between the protein sequences on translation.

Table 2. CLUSTALW Pairwise Alignment Score for Nucleotide and Protein Sequences

Virus	**Nucleotide Sequences**									
	Envelope Gene					**NS5 Gene**				
	DENV2	ZIKV	YFV	JEV	WNV	DENV2	ZIKV	YFV	JEV	WNV
DENV2	-	49.5	39.9	43.9	51.8	-	56.9	52.2	56.1	56.6
ZIKV	-	-	41.5	49.7	48.8	-	-	52.9	57.6	58.8
YFV	-	-	-	39.7	42.0	-	-	-	53.0	53.2
JEV	-	-	-	-	64.0	-	-	-	-	68.6
	Protein Sequences									
	Envelope Protein					**NS5 Protein**				
	DENV2	ZIKV	YFV	JEV	WNV	DENV2	ZIKV	YFV	JEV	WNV
DENV2	-	52.1	41.8	45.1	43.4	-	67.8	62.1	67.3	68.4
ZIKV	-	-	40.4	54.2	53.2	-	-	63.7	69.0	70.7
YFV	-	-	-	41.6	41.6	-	-	-	63.8	62.3
JEV	-	-	-	-	77.2	-	-	-	-	83.4

In spite of DENV2 being one of the more virulent pathogens and probably the most virulent among the four DENV serotypes, there is little reflection in literature of the acute base distribution difference between DENV2 and the other flaviviruses that has resulted in the codon usage profile of DENV2 being sharply different from the other flavivirus genes. One probable correlation of the disparity in the nucleotide sequences, observed in our analyses, that can be made is between DENV2 codon adaptation and the host tRNA pool [6]. Some models indicate that certain viral codon usage is optimized to match the tRNA pool in the host organism that enables faster translation elongation [7]; there are also speculations that high levels of gene expression result in tRNA starvation which slow down the replication and translation processes and result in modulation of gene expression [8,9]. These factors may combine to create a more efficient replication process for the DENV2 as compared to other flaviviruses.

Conclusions

So far, Zika and Dengue type 2 virus analyses have reported that DENV2 and ZIKV are quite close. We have found that the base distribution patterns of the two viruses are different, and we have shown these explicitly in the case of the envelope and NS5 gene sequences, through the 2D graphical representations. Quantitative estimates show that the graph spreads are about 30 to 50% larger for the DENV2 genes. We speculate that such disparity probably gives DENV2 an edge through its codon usage bias which uses the cellular tRNA pool for optimization of gene expression and regulation, making DENV2 a highly virulent pathogen.

References

1. Hahn YS; Galler R; Hunkapiller T; Dalrymple JM; Strauss JH; Strauss EG. Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. *Virology* 1988 ,162, 167-80.
2. Nandy A; Basak SC. The Epidemic that Shook the World – The Zika Virus Rampage. *Exploratory Research and Hypothesis in Medicine* 2017, 2, 43–56.
3. Nandy A. A New Graphical Representation and Analysis of DNA Sequence Structure. *Current Science* 1994, 66, 309-314.
4. Higgins D.; Thompson J.; Gibson T; Thompson J. D.; Higgins D. G.; Gibson T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, 22, 4673-4680.
5. Raychaudhury C; Nandy A. Indexing Scheme and Similarity Measures for Macromolecular Sequences. *J. Chem. Inf. Comput. Sci.* 1999, *39,* 243-247.
6. Toshimichi Ikemura. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. *Journal of Molecular Biology* 1981, 151, 389-409.
7. Plotkin, J. B.; Kudla, G. Synonymous but not the same: The causes and consequences of codon bias. *Nature Reviews Genetics* 2011, 12, 32–42.
8. Spencer, P. S.; Barral, J. M. Genetic Code Redundancy and Its Influence on the Encoded Polypeptides. *Computational and Structural Biotechnology Journal* 2012, 1,1–8.
9. Diambra LA. Differential bicodon usage in lowly and highly abundant proteins. *PeerJ* 2017, 5(3081).