ORIGINAL RESEARCH PAPER
Volume - 10 | Issue - 08 | August - 2021 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

DIFFERENTIAL AMINO ACID USAGE PATTERN OF ENVELOPE GENES OF DENGUE VIRUS

Biological Science

Manisha Ghosh Division of Bioinformatics, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata, India.

Ayan Roy Department of Bioinformatics, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, India.

Surajit Basak* Division of Bioinformatics, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata, India.*Corresponding Author

Shanta Dutta Division of Bacteriology, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata, India.

ABSTRACT

Envelope gene sequences associated with the four serotypes of dengue virus (DENV) have been analyzed to profile characteristic amino acid usage patterns and explore their evolutionary signatures. Present investigation clearly demonstrates an alteration of amino acid usage patterns of the DENV envelope genes and their evolutionary signatures depending on their geographical location of isolation. Comparative analysis associated with the impact of host genome on the envelope genes of the DENV serotypes suggests that host-associated selective constraints vary between Asian and American isolates among the four different serotypes. Molecular docking-based interaction analysis of the envelope proteins with the host (human) receptor DC-SIGN revealed contrasting degrees of binding affinities among the four different serotypes depending on the geographical location of isolation. The present work promises to confer meaningful information that may be exploited suitably to explore the complex evolutionary dynamics and mechanisms of host receptor attachment of DENV envelope proteins.

KEYWORDS

Dengue virus, Envelope genes, Relative amino acid usage, Evolutionary selection, Similarity index, Host selective constraint, Molecular docking

1. INTRODUCTION

Dengue fever is caused by dengue virus which is a mosquito-borne tropical disease [1]. The genome size of dengue virus is approximately 11 kb and the genome are translated as a single, long polypeptide and then cut into ten proteins. Four antigenically different dengue virus (DENV) serotypes (DENV-1 to 4) are known. The dengue viruses are prevalent in tropical and subtropical regions globally. The increasing co-circulation of the four different DENV serotypes, particularly in America and Asia, is not uniform and has implications for patterns in disease severity and hyperendemicity [2].

DENV envelope (E) protein is a major N-glycosylated dimeric membrane fusion protein on the surface of the virion and mediates virus binding and fusion to host cell membrane [3]. Envelope protein has been reported to experience strong selective pressure from the host immune response and targeting it has been suggested to be an effective antiviral strategy against dengue [3,4]. Therefore, we focus on the amino acid usage patterns and explore the evolutionary dynamics of DENV envelope genes. In the present study we studied the contrasting amino acid usage signatures and evolutionary patterns of the DENV envelope genes of the Asian and American origin.

2. MATERIALS AND METHODS

All available envelope gene sequences of the four DENV serotypes were retrieved from ViPR database (https://www. viprbrc.org/ brc/ home.spg?decorator=vipr). CodonW software (http:// www. molbiol. ox.ac.uk/cu) was used to perform correspondence analysis (CoA) on RAAU data of the DENV envelope gene sequences [5,6]. Evolutionary rates of the envelope genes representing Asian and American DENV isolates was calculated using Codeml program of the PAML software package (version 4.7) [7,8]. Similarity index D (A, B) refers to the influence of host genome in shaping viral codon usage patterns [9]. The parameter similarity index, D (A, B) is calculated as:

$$R(A,B)=\frac{\sum_{i=1}^{i=59} a_i X b_i}{\sqrt{\sum_{i=1}^{i=59} a_i^2 X \sum_{i=1}^{i=59} b_i^2}}$$

$$D(A,B)=\frac{1-R(A,B)}{2}$$

where R (A, B) denotes the cosine value of an included angle between A and B spatial vectors. Relative synonymous codon usage (RSCU) value of a particular codon is denoted by 'ai' for a given type of envelope (i.e., Asian and American isolates) protein coding sequence. RSCU value of the same codon for human host is denoted by 'bi'. RSCU values of the envelope genes of Asian and American DENV isolates were calculated using CodonW software. The Codon Usage Database was used to obtain the codon usage values for human genome. D (A, B) refers to the similarity index value that ranges within 0 to 1.0 and greater impact due to host selective constraint is indicated if D (A, B) value close to 1. The DENV envelope protein sequences, representing the four different serotypes, were modelled using the MODELLER 9 software [10] with PDB structure 3J27 (Chain C) as the template based on alignment and identity [11]. Protein crystal structure of the human DC-SIGN (PDB: 1SL4) was retrieved from PDB. Envelope proteins representing the four different DENV serotypes were docked separately with the associated human DC-SIGN receptor employing GRAMM-X server (http:// vakser. compbio.ku.edu/resources/gramm/grammx/).

3. RESULTS AND DISCUSSION

3.1 Amino acid usage pattern and evolutionary selection

We performed Correspondence analysis (CoA) on the relative amino acid usage (RAAU) of the DENV envelope genes separately for each serotype to analyse the amino acid usage patterns. It is a multivariate statistical technique that is frequently used to assess the variation in a dataset and in the present study it was employed to explore the variations among the concerned DENV envelope genes [6,12]. Correspondence analysis resulted in two clusters for each of the four serotypes along the x-axis (Figure 1).

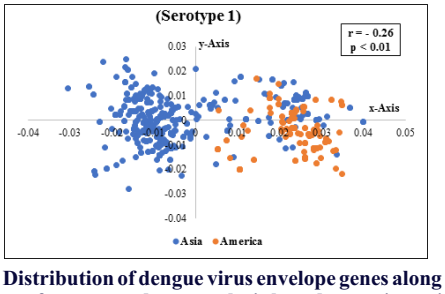

Figure1: Distribution of dengue virus envelope genes along the two major axes of correspondence analysis based on amino acid usage for Serotype 1. Blue points represent envelope genes from Asia and orange colored points represent envelope genes from America. Similar pattern is observed for Serotype 2, Serotype 3, and Serotype 4.

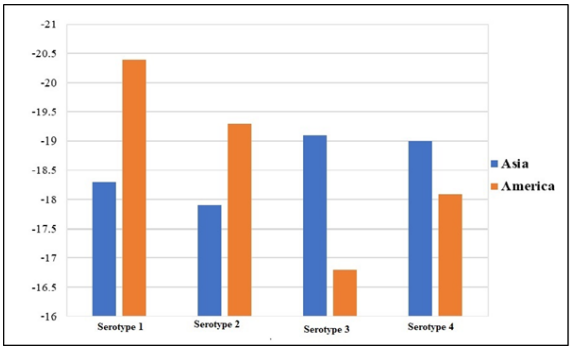

Figure 2: Variation of binding energy between DENV envelope protein and human DC-SIGN receptor among the four different serotypes of dengue virus.

Table 1: Three parameters compared between Asian and American envelope gene sequences of four serotypes of dengue virus.

	Slope between Ka and Ks		Similarity index		Binding energy (kcal/mol)	
	Asia	America	Asia	America	Asia	America
Serotype1	0.0395	0.041	0.119	0.110	-18.3	-20.4
Serotype2	0.0149	0.0154	0.110	0.108	-17.9	-19.3
Serotype3	0.0493	0.0307	0.108	0.123	-19.1	-16.8
Serotype4	0.031	0.0282	0.102	0.103	-19.0	-18.1

For a given serotype, one cluster comprised of the envelope genes of the Asian origin and the other cluster comprised of the envelope genes of the American (North America and South America) origin. This pattern of amino acid usage of envelope genes was consistent for all the four serotypes. This observation indicates that the DENV envelope genes have different amino acid usage signatures depending on their geographical location of isolation.

We have also noted that positions of the genes along the x-axis (Figure 1) significantly correlated with the evolutionary rate of the envelope gene ($P < 0.01$). It indicates towards the governing role of evolutionary selection in shaping the amino acid usage patterns of the envelope genes depending upon their geographical location.

To corroborate our observation, a linear regression line was fitted between Ka and Ks through the origin with the assumption that Ka and Ks should both initially be zero at the moment of lineage divergence. The slopes for the lines for four serotypes of Asia and America have been detailed in Table 1. The observations suggest that the rate of increase of Ka values relative to the increase in Ks values is faster in Asian envelope genes for serotypes 3 and 4 compared to the American counterparts. On the contrary, in case of the serotypes 1 and 2, the rate of increase of Ka values relative to the increase in Ks values was noted to be faster among the envelope genes of American origin compared to those of the Asian origin.

It may be inferred from the above results that the envelope genes representing serotypes 3 and 4 belonging to the American origin have been under stronger evolutionary selection pressure than the DENV envelope genes of the Asian origin. On the other hand, the envelope genes of serotypes 1 and 2 reflecting the Asian origin experience stronger evolutionary selection compared to their American counterparts.

3.2 Influence of host machinery

We, therefore, assess the influence of selection pressure due to human host on the viral codon usage patterns. The similarity index parameter (SiD) has been computed to decipher the influence of human host genome on the envelope genes of DENV serotypes of the Asian and American origins. An in-depth analysis revealed that the SiD values of the envelope genes for the DENV serotype 1 of Asian origin is significantly higher ($P < 0.01$) than the envelope genes for the DENV serotype 1 of the American origin. Similarly, envelope genes for the DENV serotype 2 of Asian origin is significantly higher ($P < 0.01$) than the envelope genes for the DENV serotype 2 of American origin (Table 1). However, the SiD values of the envelope genes of Asian origin for the DENV serotype 3 serotype 4 were significantly lower ($P < 0.01$) than the envelope genes of American origin for the DENV serotype 3 and serotype 4 (Table 1). These observations indicate that the host directed selection pressure was higher on the envelope gene sequences isolated from Asian region for serotypes 1 and 2 whereas, for serotypes 3 and 4 host directed selection pressure was higher on the envelope gene sequences isolated from the American region. The varying degrees of selection pressure on the DENV envelope genes among the Asian and American isolates correlated well with the contrasting evolutionary selection operational on them.

3.3 Molecular interaction between envelope protein and human receptor

Molecular docking of the DENV envelope protein with human DC-SIGN receptor was performed to further evaluate the influence of selection pressure on the binding stability of the envelope proteins with the human receptor DC-SIGN. The binding energy scores of the complex formed between DC-SIGN and envelope proteins separately for each serotype have been detailed in Table 1. An analysis of the binding energy scores of the DENV envelope protein-human receptor DC-SIGN complexes revealed that the binding energy values of the complexes were significantly higher ($P < 0.01$) for serotypes 1 and 2 for the American isolates in comparison to the Asian isolates (Figure 2 and Table 1). On the contrary, the binding energy scores of the complexes for serotypes 3 and 4 were significantly higher ($P < 0.01$) for the Asian isolates in comparison to their American counterparts (Figure 2 and Table 1).

It has been proposed that greater selection pressure due to host on viral genome can be ascertained through higher similarity index value [12, 13, 14]. It was interesting to note that the DENV envelopes genes of the American origin for serotypes 1 and 2 and DENV envelopes genes of the Asian origin for serotypes 3 and 4 were under relaxed host selection pressure compared to their counterparts and exhibited higher binding potential with the human receptor DC-SIGN. In this pretext it is worth mentioning that the event of relaxed selection pressure with the accumulation of beneficial non-synonymous mutations among viral genomes has been reported to significantly contribute to viral proliferation and subsequent progression to infection in host environment [13, 15]. Our observation might point towards the fact that relaxed host selective constraints might provide scope for the accumulation of non-synonymous mutations in the viral populations and considerably alter the fate of attachment of the DENV envelope proteins with the human receptor DC-SIGN. However, further *in vitro* and *in vivo* studies are demanded to draw a definite conclusion.

4. CONCLUSION

The present study highlights the variations in amino acid usage patterns, host selection pressure and interaction with human host receptor DC-SIGN among DENV envelope genes depending on geographical location of isolation of envelope genes. Present findings may provide ample scopes to elucidate the genetic characteristics of DENV with potential host influence. This study also explores the complexities of DENV pathogenesis, targeted towards the development of new generation dengue vaccines.

ACKNOWLEDGEMENT:

Manisha Ghosh is supported by Senior Research Fellowship by Indian Council of Medical Research (ICMR).

Conflict of interest: The authors declare that no conflicts of interest exist.

REFERENCES

1. Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annu Rev Microbiol. 44:649-88. doi: 10.1146/ annurev. mi. 44.100190.003245
2. Sang S, Liu-Helmersson J, Quam MBM, Zhou H, Guo X, Wu H, Liu Q (2019) The evolutionary dynamics of DENV 4 genotype I over a 60-year period. PLoS Negl Trop Dis. 13(7): e0007592. doi: 10.1371/journal.pntd.0007592.
3. Crill WD, Roehrig JT (2001) Monoclonal antibodies that bind to domain III of dengue virus E glycoprotein are the most efficient blockers of virus adsorption to Vero cells. J Virol. 75(16):7769-73. doi: 10.1128/JVI.75.16.7769-7773.2001
4. Fahimi H, Mohammadipour M, Haddad Kashani H, Parvini F, Sadeghizadeh M (2018) Dengue viruses and promising envelope protein domain III-based vaccines. Appl Microbiol Biotechnol. 102(7):2977-2996. doi: 10.1007/s00253-018-8822-y
5. Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press.
6. Peden JF (2000) Analysis of codon usage [Doctoral dissertation]. University of Nottingham.
7. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3(5):418-26. doi: 10.1093/oxfordjournals.molbev.a040410
8. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586-91. doi: 10.1093/molbev/msm088.

9. Nasrullah I, Butt AM, Tahir S, Idrees M, Tong Y (2015) Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. BMC Evol Biol. 15:174. doi: 10.1186/s12862-015-0456-4.
10. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol. 374:461-91. doi: 10.1016/S0076-6879(03)74020-8
11. Zhang X, Ge P, Yu X, Brannan JM, Bi G, Zhang Q, Schein S, Zhou ZH (2013) Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. Nat Struct Mol Biol. 20(1):105-10. doi: 10.1038/nsmb.2463
12. Roy A, Basak S (2021) HIV long-term non-progressors share similar features with simian immunodeficiency virus infection of chimpanzees. J Biomol Struct Dyn. 39: 2447-2454.
13. Roy A, Banerjee R, Basak S (2017) HIV Progression Depends on Codon and Amino Acid Usage Profile of Envelope Protein and Associated Host-Genetic Influence. Front Microbiol. 8:1083. doi: 10.3389/fmicb.2017.01083
14. Khandia R, Singhal S, Kumar U, Ansari A, Tiwari R, Dhama K, Das J, Munjal A, Singh RK (2019) Analysis of Nipah Virus Codon Usage and Adaptation to Hosts. Front Microbiol. 10:886. doi: 10.3389/fmicb.2019.00886
15. Khanlou H, Mechain C, Maunand B (1996) Polymyositis in HTLV-I infected patient. J Intern Med. 239(4):371. doi: 10.1046/j.1365-2796.1996.455793000.