

IN SEARCH OF CONSERVED RNA MOTIFS OF DENGUE GENOME OF ALL SEROTYPE: A BIOINFORMATIC APPROACH

Meghna Saha¹, Sayak Ganguli², Sanjay Kumar Paul³, Rayan Das⁴ and Santi Ranjan Dey⁵

Block Epidemiologist Officer, Jhargram, West Bengal¹

Assistant Professor, Post Graduate Department of Biotechnology, St.Xavier's College, Kolkata²

Assistant Professor, Department of Zoology, Rammohan College, Kolkata^{3,5}

Guest Lecturer, Post Graduate Department of Zoology, Asutosh College⁴

Abstract: RNA viruses use small genomes that contain information in both their core sequences and higher-order structures to hijack cellular metabolism and encourage their own replication. By identifying particular sequences that are conserved throughout a collection of related viruses, the majority of functional structures that have been discovered to date. We effectively find numerous hitherto unannotated motifs crucial for viral fitness by flipping the traditional technique, which defines RNA structures first before checking for conservation of these motifs. In addition to identifying possible motifs helpful in the development of antiviral medicines and vaccines, this work demonstrates the ability of RNA structure as a tool for discovering functional elements in viruses. It also paves the way for additional functional element identification in big viral messenger as well as non-coding RNAs. A virus known as dengue virus 1 (DEN-1) was isolated by Walter Schlesinger and Albert B. Sabin. The four closely related viruses that cause dengue diseases are DEN-1, DEN-2, DEN-3, and DEN-4. They are known as serotypes because the antibodies in human blood serum react differently with each of these four viruses. The four dengue viruses are related and share roughly 65% of their genomes despite the fact that there is a great deal of genetic heterogeneity within a single serotype. Despite these variations, all dengue serotype infections result in the same illness and a set of same clinical symptoms. In this research, we looked for a specific or conservative RNA pattern that could be used to neutralize the DENGUE virus, by targeting RNA in future.

Key Words: Dengue, RNA, Serotype, RNA-motif, Dengue-protein

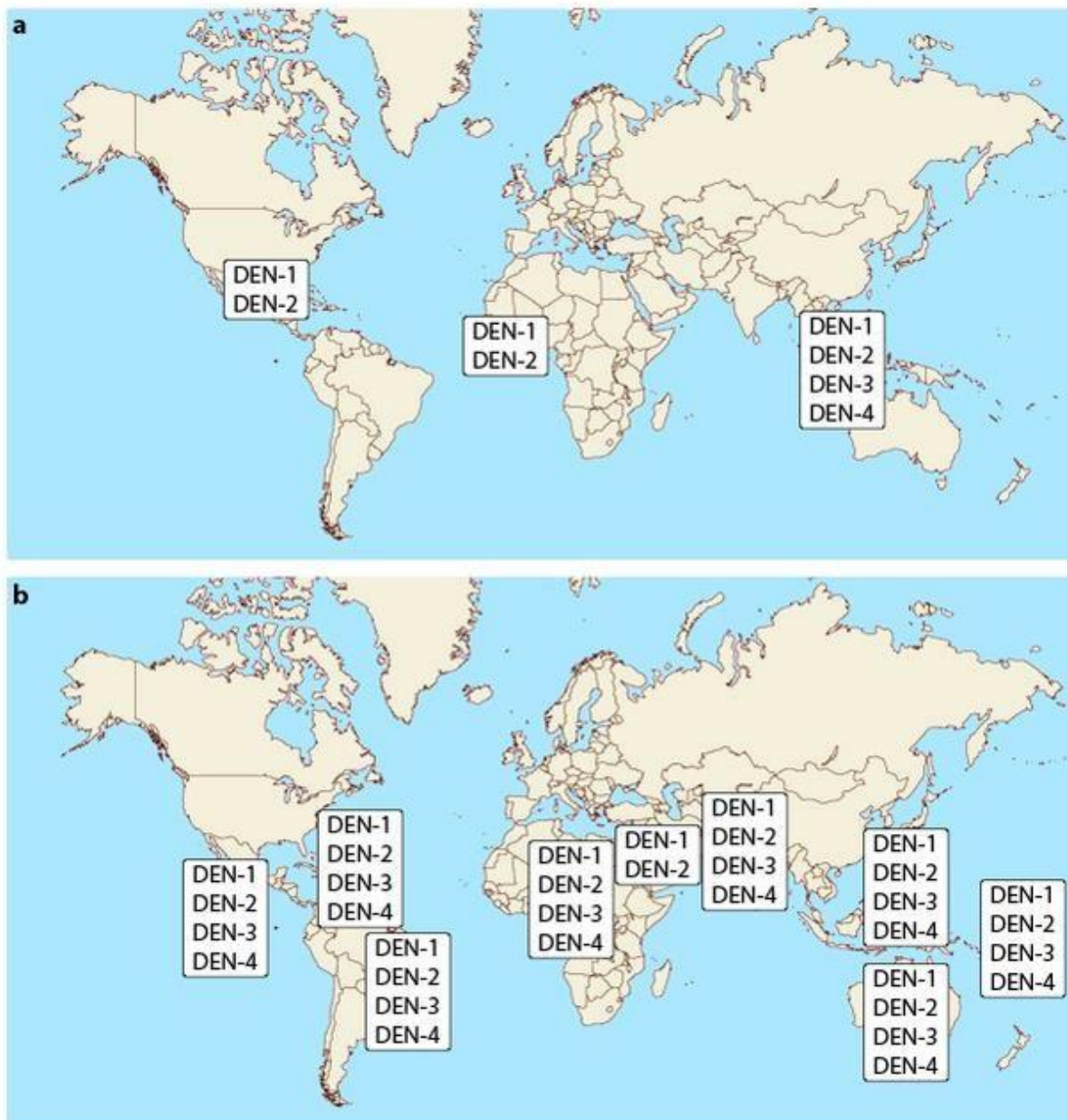
I. INTRODUCTION

A wide range of living things, including bacteria, plants, and animals, can become infected by viruses, which are little agents. The dengue virus is an ultra microscopic entity that can only replicate inside a host organism, like other viruses. The family Flaviviridae's genus Flavivirus contains the dengue viruses. This genus contains a variety of additional viruses that cause human infections and are spread by ticks and mosquitoes in addition to the dengue virus. Yellow fever, West Nile, Japanese encephalitis, and tick-borne encephalitis viruses are all classified as flaviviruses. REN KIMURA and SUSUMU HOTTA discovered the dengue virus in 1943.

These two researchers were looking at blood samples taken from patients in Nagasaki, Japan, during the 1943 dengue epidemic. A year later, the dengue virus was separately isolated by Albert B. Sabin and Walter Schlesinger. The virus that is now known as dengue virus 1 (DEN-1) had been isolated by both teams of researchers. The DEN-1, DEN-2, DEN-3, and DEN-4 viruses are four closely related viruses that cause dengue illnesses. Because each of these four viruses interacts differently with the antibodies in human blood serum, they are referred to as serotypes. Even while there is considerable genetic variation within a single serotype, the four dengue viruses are similar and share about 65% of their genomes. All dengue serotype infections cause the same sickness and similar set of clinical signs, despite these differences.

All four serotypes were discovered in Southeast Asia in the 1970s and both DEN-1 and DEN-2 were discovered in Central America and Africa. The four serotypes were, however, widely dispersed geographically by 2004. Currently, all four dengue serotypes coexist in tropical and subtropical areas of the world (Fig. 1). The four dengue serotypes have similar geographic and ecological niche. Scientists hypothesize that the dengue viruses evolved in nonhuman primates and jumped from these primates to humans in Africa or Southeast Asia between 500 and 1,000 years ago.

Each of panels A and B (Fig. 1) displays a map of the entire world. Green shading and a thin black line delineate the nations. The four dengue virus serotypes are depicted on the map as labeled white boxes. The geographical distribution of dengue serotypes in 1970 is shown in panel a. The geographic distribution of dengue serotypes in 2004 is shown in Panel b. The locations of the DEN-1 and DEN-2 serotypes in Central America and Western Africa are depicted in panel A. Southeast Asia exhibits the DEN-1, DEN-2, DEN-3 and DEN-4 serotypes. All four serotypes are depicted in Central America, the Caribbean, and the Pacific Islands, the Caribbean, South America, Africa, India, Southeast Asia, and Australia. In the Middle East, only DEN-1 and DEN-2 are broadcast. A person obtains immunity to a certain dengue serotype after recovering from an infection with it. After the initial dengue infection, people are shielded from infection by the other three serotypes for two to three months. Unfortunately, it doesn't offer long-term security. Any of the remaining three dengue serotypes can infect a person after that little window of time. Researchers have found that people who have had prior infections are at a lower risk of developing severe dengue illnesses than those who have not.



Aedes mosquitoes in Africa, Southeast Asia, and South Asia transmitted the dengue virus to nonhuman primates in sylvatic cycles until a few hundred years ago, with infrequent emergences into human populations (Anonymous, 2010, Homes and Twiddy, 2003). However, following its breakout from sylvatic cycles, the dengue virus has spread around the world, and the primary lifecycle now only involves transmission between people and *Aedes* mosquitoes (Halstead,

1988). In several vector species, vertical transmission between mosquitoes has also been documented (Haddow *et al*, 2013). The virus has been identified to infect dogs, but additional investigation is required to understand whether dogs or other animals can act as reservoirs or are simply unintentional hosts (Thongyuan and Kittayapong, 2017). Recent research suggests that, depending on the kind of infected cell, host homeostatic systems such as autophagy and the ER stress response, as well as apoptosis, are activated as the virus infects human cells (Ghosh *et al*, 2014). During an infection, the activation of autophagy and ER stress promotes virus proliferation (Datan *et al*, 2016, McLean *et al*, 2011.) Review articles from several research teams attempt to give comprehensive explanations of the dengue life cycle at the cellular level (Zakeri *et al*, 2015, Neufeldt *et al*, 2018). The information needed for RNA virus replication in host cells is encoded in both the genomes' linear sequence and more intricate higher-order structures. A portion of these RNA genome architectures, which have been widely described for viruses with known characteristics, clearly demonstrate sequence conservation. However, it is largely unknown to what extent viral RNA genomes contain functional structural elements that cannot be identified by sequence alone but are nevertheless essential to viral fitness. Here, we develop a structure-first experimental approach and employ it to find 22 motif-like structures throughout the coding sequences of the RNA genomes for the four dengue virus serotypes. At least ten of these patterns influence viral fitness, indicating that RNA structure-mediated control of viral coding sequences occurs to a large but previously unrecognized extent. These viral RNA structures interact with proteins, support a small overall genome architecture, and control viral replication. These motifs are thus restricted at the levels of both RNA structure and protein sequence, making them prospective targets for antiviral agents and live-attenuated vaccines that can overcome resistance. In order to effectively find ubiquitous RNA-mediated control in viral genomes and most likely, other cellular RNAs, conserved RNA structure must first be identified structurally (Boerneke, 2023).

Dengue Virus Genome and Structure (Fig 1 & 2)

One strand of RNA makes up the dengue virus genome. It is known as positive-sense RNA because it may produce proteins by direct translation. Ten genes are encoded by the viral genome (Figure 2). After being translated as a single, lengthy polypeptide, the genome is divided into ten proteins. The dengue virus's RNA genome is depicted on a diagram with labels designating its structural and non-structural parts. The RNA is shown as a horizontal cylinder divided into variously sized, coloured pieces. The 5 prime UTR is a thin black coiled line that extends from the left end of the cylinder and is marked as un-translated RNA. The capsid, designated C and coloured light brown; the membrane, labeled M and coloured orange; and the envelope, labeled E and coloured light brown, blue are the genes encoded by the dengue virus genome, from left to right.

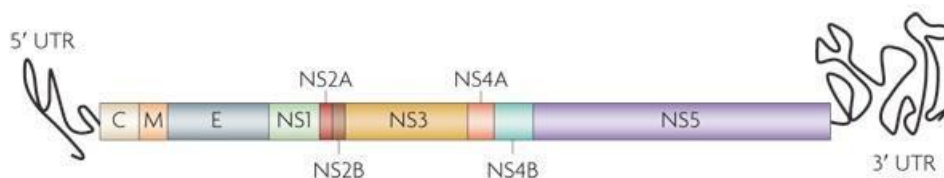


Figure 2: Dengue virus genome:

The dengue virus genome encodes three structural (Capsid [C], Membrane [M], and Envelope [E]) and seven NonStructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) proteins. The non-structural genes NS1 (green), NS2A (red), NS2B (dark brown), NS3 (yellow), NS4A (dark orange), NS4B (teal), and NS5 (purple) are all represented. The longest genes are NS5 and NS2A and NS2B are the shortest. The 3 prime UTR is a section of un-translated RNA located at the right end of the cylinder. The proteins that make up the Capsid (C), Envelope (E) and Membrane (M) are all structural proteins. Nonstructural proteins are NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. These nonstructural proteins are involved in the replication and construction of viruses. The dengue virus is depicted in a schematic with its primary structural elements. A light brown circle encloses an orange hexagon that represents the virus. The brown circle represents the viral envelope, and the hexagon represents the nucleocapsid. The viral genome is represented by a thin crimson substance coiling up into the nucleocapsid. Asymmetrically, seven red lines extend forth from the viral envelope. A green sphere is at the conclusion of each line. E and M proteins are these protrusions. The dengue virus has a roughly spherical shape. Inside the virus is the nucleocapsid, which is made of the viral genome and C proteins. The nucleocapsid is surrounded by a membrane called the viral envelope, a lipid bilayer that is taken from the host. Embedded in the viral envelope are E and M proteins that span through the lipid bilayer. These proteins form a protective outer layer that controls the entry of the virus into human cells. The structure of the dengue virus is roughly spherical, with a diameter of approximately 50 nm (Figure 3). The core of the virus is the nucleocapsid, a structure that is made of the viral genome along with C proteins.

The nucleocapsid is surrounded by a membrane called the viral envelope, a lipid bilayer that is taken from the host. Embedded in the viral envelope are 180 copies of the E and M proteins that span through the lipid bilayer. These proteins form a protective outer layer that controls the entry of the virus into human cells.

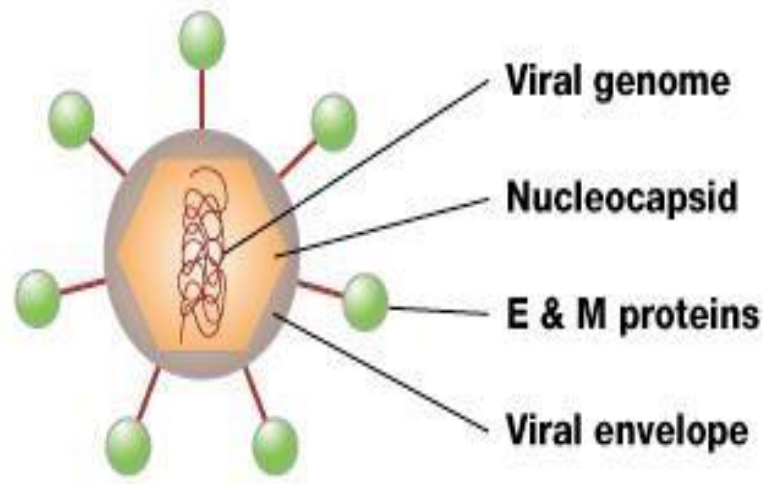


Figure 3: Dengue virus structure

Dengue virus Replication and Infection Cycle (Fig 4)

When the dengue virus binds to a human skin cell (Figure 4), the viral replication process starts. The membrane of the skin cell folds around the virus after this attachment and creates a pouch that seals around the virus particle, forming an endosome. Endosomes are typically used by a cell to take in big molecules and other foreign objects for nutrition. The dengue virus can enter a host cell by taking over this typical cell mechanism. A dengue virus particle infects a eukaryotic host cell, uses the machinery of the host cell to reproduce the viral genome, and then leaves the cell as a fully developed virion. The eukaryotic host cell is shown as a spherical with an atypical shape, a single nucleus, and numerous organelles. A tiny circle that is about 1/200th the size of the eukaryotic cell serves as the virus's representation.

Figure 4 shows replication of the dengue virus the endocytosis process allows the dengue virus to cling to a host cell's surface and enter the cell. The virus releases its contents into the cytoplasm after merging with the endosomal membrane deep inside the cell. The viral genome is released as the virus particle disintegrates. The viral genome is replicated when the viral RNA (vRNA) is translated into a single polypeptide and divided into ten proteins. When newly synthesized RNA and structural proteins protrude from the endoplasmic reticulum (ER), virus assembly takes place on its surface. The trans-Golgi network (TGN) is used to carry the immature virus particles.

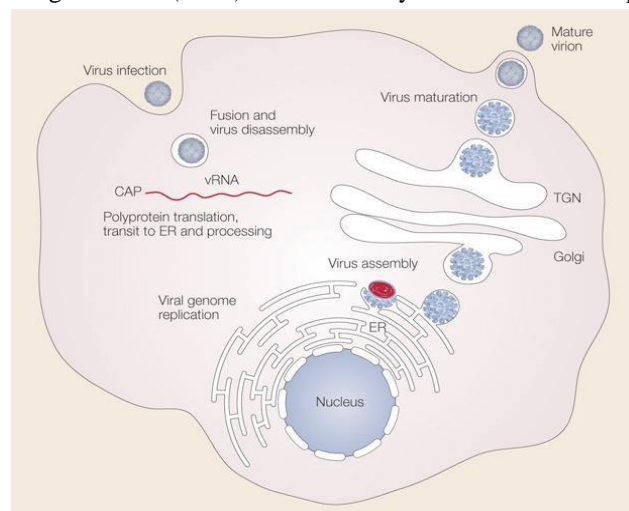


Figure 4

In TGN they mature and convert to their infectious form. The mature viruses are then released from the cell and can go on to infect other cells. Once the virus has entered a host cell, the virus penetrates deeper into the cell while still inside the endosome. Researchers have learned that two conditions are needed for the dengue virus to exit the endosome:

1. The endosome must be deep inside the cell where the environment is acidic.
2. The endosomal membrane must gain a negative charge.

These two conditions allow the virus envelope to fuse with the endosomal membrane, and that process releases the dengue nucleocapsid into the cytoplasm of the cell. In the cytoplasm, the nucleocapsid opens to uncoat the viral genome. This process releases the viral RNA into the cytoplasm. The viral RNA then hijacks the host cell's machinery to replicate itself. The virus uses ribosomes on the host's rough endoplasmic reticulum (ER) to translate the viral RNA and produce the viral polypeptide. This polypeptide is then cut to form the ten-dengue proteins. The newly synthesized viral RNA is enclosed in the C proteins, forming a nucleocapsid. The nucleocapsid enters the rough ER and is enveloped in the ER membrane and surrounded by the M and E proteins. This step adds the viral envelope and protective outer layer. The immature viruses travel through the Golgi apparatus complex, where the viruses mature and convert into their infectious form. The mature dengue viruses are then released from the cell and can go on to infect other cells.

The Hypothesis

The best way to restrict this virus, is to block the vRNA. As the dengue genome is having 65% diversity, we are in search of conserved motifs, that can be used as target(s).

Materials and Methods

Individual coding sequences (CDS) of the corresponding genes encoded by the Dengue virus are to be downloaded from NCBI.

(<https://www.ncbi.nlm.nih.gov/genome/10308>).

All the sequences of the various genes are to be collected and converted to RNA from DNA sequence using fr33 server

Then the best three random motifs in these RNA sequences of each genes are to be identified using MEME suite.

3D structure of these individual identified motifs are to be generated using the RNA composer server. Then the base pair propensities and motifs in the tertiary structures are to be critically analyzed by using DSSR server.

II. RESULTS AND DISCUSSION

The RNA tertiary structure is made up of motifs, and many servers were used to predict the 3D structures. In this instance, the MEME suite's output simply displays any random patterns that were found in the gene after searching. Based on their p-value, each gene in the dengue genome has one of the top three random motifs, the most important RNA motif found in each gene of the dengue virus genome (TABLE I).

The three best motifs identified from each gene were then generated into 3D structures using the RNA composer server. The DSSR server was used to analyse each of these structures. The full results are summarized in (TABLE 2-4), however of the many 3D structure motifs identified, STACKS was shown to be among the most prevalent, followed by SPLAYED APART DI-NUCLEOTIDE and NON-LOOP SINGLE STRANDED SEGMENT patterns.

Since then, it has become a well-studied problem to manipulate and describe the structure of RNA. The backbone torsion is described by six conventional torsion angles (Alpha, Beta, Gamma, Epsilon, and Zeta). The (crank-shaft effect) (Hollbrok et al 1978 and Olson 1982) modifies other torsion angles to compensate for the significant changes in particular torsion angles. As a result, different torsion angle combinations can produce nucleotide morphologies that are identical. Numerous other reports also imply that the majority of RNA motifs serve important roles, either by controlling interactions with other biomolecules or by providing conformational stability.

III. CONCLUSION

The discovery of multiple RNA motifs suggests that RNA 3D structures are complex in form and control biological interactions with other macromolecules. This is supported by the discovery of numerous RNA motifs in the dengue virus genes. These structures need to be investigated in further detail in order to determine their likely functional roles and to analyze them for potential therapeutic targets. The results of the study also support that, though there are progressive change molecular epidemiology of dengue in India (Anoop, M *et al*, 2012), and also point to the role of exotic introductions of viral strains in this process, some conserved motifs are present. The sequences generated would serve as a reference for future studies of the circulating DENV I-IV strains in India.

ACKNOWLEDGEMENT

The authors are thankful to Principal, Rammohan College for her support.

REFERENCES

- [1] Anoop, M, Ashish J. M, Jayakumar, B., Aneesh I, Nai, S., Abraham, R., and. Sreekuma E. Complete genome sequencing and evolutionary analysis of dengue virus serotype 1 isolates from an outbreak in Kerala, South India. *Virus Genes* .2012. 45:1–13. DOI 10.1007/s11262-012-0756-3 Anonymous. "Dengue virus". Pathogen Information (PathInfo). Virginia Bioinformatics Institute, Virginia Tech. Archived from the original on 2010-08-30.
- [2] Boerneke M A, Gokhale N S, Horner S.M. and Weeks, K.M. Structure-first identification of RNA elements that regulate dengue virus genome architecture and replication. 2023. *PNAS*. 120 (15) e2217053120. <https://doi.org/10.1073/pnas.2217053120>
- [3] Datan E, Roy S. G. , Germain G, Zali N, McLean JE, Golshan G, Harbajan S, Lockshin RA, Zakeri Z. "Dengue-induced autophagy, virus replication and protection from cell death require ER stress (PERK) pathway activation". *Cell Death & Disease*. 2016.7 (e2127):e2127. doi:10.1038/cddis.2015.409. PMC 4823927. PMID 26938301.
- [4] Ghosh Roy S, Sadigh B, Datan E, Lockshin RA, Zakeri Z . "Regulation of cell survival and death during Flavivirus infections". *World Journal of Biological Chemistry*. 2014. 5 (2): 93–105. doi:10.4331/wjbc.v5.i2.93 (inactive 31 December 2022). PMC 4050121. PMID 24921001.
- [5] Haddow, A. D. , Guzman H, Popov V. L. , Wood, T. G. , Widen S. G. , Tesh, R. B., Weaver, S. C. "First isolation of Aedes flavivirus in the Western Hemisphere and evidence of vertical transmission in the mosquito Aedes (Stegomyia) albopictus (Diptera: Culicidae)". *Virology*. 2013. 440 (2): 134–9. doi:10.1016/j.virol.2012.12.008. PMID 23582303.
- [6] Halstead, S. B. "Pathogenesis of dengue: challenges to molecular biology". *Science*. 1988. 239 (4839): 476–81. Bibcode:1988Sci...239..476H. doi:10.1126/science.3277268. PMID 3277268.
- [7] Holmes E. C and Twiddy, S. S. "The origin, emergence and evolutionary genetics of dengue virus". *Infection, Genetics and Evolution*. 2003. 3 (1): 19–28. doi:10.1016/s1567-1348(03)00004-2. PMID 12797969.
- [8] McLean, J. E, Wudzinska A, Datan E, Quaglino D, Zakeri Z . "Flavivirus NS4A-induced autophagy protects cells against death and enhances virus replication". *The Journal of Biological Chemistry*. 2011. 286 (25): 22147–59. doi:10.1074/jbc.M110.192500. PMC 3121359. PMID 21511946.
- [9] Neufeldt C. J. , Cortese M, Acosta EG, Bartenschlager R . "Rewiring cellular networks by members of the Flaviviridae family". *Nature Reviews. Microbiology*. 2018. 16 (3):125–142. doi:10.1038/nrmicro.2017.170. PMC 7097628. PMID 29430005.
- [10] Holbrook S R, Sussman J L, Warrant R W and Kim S H. Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J Mol Biol*. 1978; 123:631–60.
- [11] Olson WK. Computational studies of polynucleotide flexibility. *Nucleic Acids Res*. 1982; 10:777–87.
- [12] Thongyuan, S. and Kittayapong, P. "First evidence of dengue infection in domestic dogs living in different ecological settings in Thailand". *PLOS ONE*. 2017. 12 (8): e0180013. Bibcode:2017PLoSO..1280013T. doi:10.1371/journal.pone.0180013. PMC 5576688. PMID 28854207.
- [13] Zakeri Z, Ghosh-Roy S, Datan E, McLean JE, Zandi K, Germain G. "Cell death and virus infection – a short review". *20 Years of Cell Death*. 2015. 1 (1): 292–325. doi:10.13140/RG.2.2.35359.97447.

TABLE I







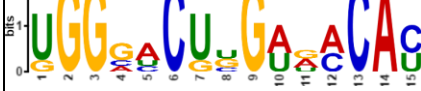

Genome Segment	Motif Logo	Best Motif Sequence
3' UTR		AAACUAUGCUACCUGU
5' UTR		AGGGAGCUAAGCUC
ANCHORED CAPSID PROTEIN C		AUGCUG
CAPSID PROTEIN C		CAGCAGGGAUUUGAAG AGAUGGGGAACAAU
ENVELOP PROTEIN E		UCAACAUAGAAGCAGAA CCU
MEMBRANE GLYCOPROTEIN M		CUGGAAACAUG
MEMBRANE GLYCOPROTEIN PRECURSOR M		UGGGACUGGAGACAC
NON STRUCTURAL PROTEIN NS1		GGAAGUUGAAGACUAUG GCUUUGGAGUAUUCACC ACC

Table 2

	MULTI PL ET	HELIC ES	STE M S	COA X IAL	STA CK SM	ATO 0	HAI R PINAL	INTE N LOO P	NON- SING L E DI- STRA NUCL EO DED TIDE	SPLAY APAR T DI- NUCL EO TIDE	ISOLA WC/W O BBLE APIR	BULG E TURN	POSSIB LE KINK TURN	U TUR N
3'UTR	0	1	2	1	13	0	20	2	7	0	0	0	0	
5'UTR	0	1	1	0	6	2	10	2	4	0	0	0	0	
ANCHORE DCAPSID PROTEIN C	0	0	0	0	5	2	0	0	3	2	0	0	0	
CAPSID PROTEIN C	0	0	0	0	5	1	0	0	3	2	0	0	0	
ENVELOP PROTEIN E	0	1	1	0	11	0	10	4	12	0	0	0	0	
MEMBRAN E GLYCOPRO TEINM	0	1	1	0	9	1	10	4	0	0	0	0	1	
MEMBRAN E GLYCOPRO TEIN PRECURSO RM	0	1	1	0	9	2	10	4	6	0	0	0	1	

Table 3

NON STRUCTURAL L PROTEIN NS1	2	5	5	019	6	3	4	5	16	4	1	2	1
NON STRUCTURAL L PROTEIN NS2A	0	1	0	011	3	0	0	3	5	0	0	0	0
NON STRUCTURAL L PROTEIN NS2B	0	2	3	111	2	2	1	4	9	0	0	0	0
NON STRUCTURAL L PROTEIN NS3	0	1	1	07	2	1	0	4	9	0	0	0	0
NON STRUCTURAL L PROTEIN NS4A	0	0	0	04	0	0	0	3	9	0	0	0	0
NON STRUCTURAL L PROTEIN NS4B	0	0	0	50	2	0	0	3	9	0	0	0	0
PROTEIN 2K	0	0	0	05	0	0	0	3	3	0	0	0	0
PROTEIN PR	0	1	1	08	0	0	0	3	3	0	0	0	0
RNA DEPENDENT RNA POLYMERASE NS5	1	5	7	220	14	4	4	5	12	1	0	0	0

Table 4

