# Comparative Genome/Proteome Analysis of Four Serotypes of Dengue Virus Using Bioinformatics Tools

Ayesha Arsha[1], Muhammad Naeem Shahwani,[1] Asma Yousufzai[1], Fazal-ur-Rehman,[2] Muhammad Asif,[1] Nazeer Ahmed[1]*
[1]Department of Biotechnology and Informatics, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta,[2]Department of Microbiology, University of Balochistan, Quetta

**Abstract**

*Dengue viral infections pose threat to almost half of the world's population health. The recent outbreak of dengue fever in Pakistan has invited attention of the scientific community to attempt devising ways and means and initiating programs controlling this menace. This work is also a step forward in the same direction. Utilizing different in-silico approaches like CLC Bio workbench, Protparam and Virus mploc comparative analysis of the full length genome and proteome of dengue virus serotypes was performed. All the four serotypes exhibited high level of similarities both at genome and proteome level. However, variations still existed though insignificant. Concerning prediction of sub-cellular localization of viral capsid, envelop and membrane glycoprotein in the host cell, all three proteins in four serotypes were shown to target endoplasmic reticulum. Endoplasmic reticulum hosting the viral structural proteins emerges as the pivot of the future studies aiming controlling dengue infections.*

**Key words**: Dengue serotypes – in-silico-subcellular localization- Genome-Bioinformatics-Proteome

*Corresponding author's email: (nazirdurrani@yahoo.com)*

## INTRODUCTION

The mysterious viruses are small, infectious, obligate intracellular parasites solely relying on host cells for their existence as they lack the fundamental biochemical machinery (Koonin et al. 2006). Although Viruses are one of the most important biological entities on the earth, still very limited understanding is there of both the number of virus species and the diversity of viral genes and genomes. It is however, beyond doubt that they infect almost all forms of life, including bacteria, archaea and eukaryotes from plants to humans to fungi (Prangishvili et al. 2006).

Among various viral infections, Dengue infection has been observed to cause havoc every now and then putting approximately, 40% of the world population at risk. The estimated figure of dengue infections is 50-100 million people worldwide (WHO 2012). Dengue virus (DENV), which is a mosquito-borne flavivirus causes a wide range of diseases in humans, from a self-limited Dengue fever, to a life-threatening syndrome called Dengue Hemorrhagic Fever (DHF). There are four antigenically different serotypes of the virus: DENV-1, DENV-2, DENV-3 and DENV-4. within the *Flavivirus* genus of the family *Flaviviridae (van der Most et al. 1996)* All four serotypes cause dengue fever (DF), and occasionally the potentially fatal dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS)(Melo et al. 2007). Despite many ongoing vaccine development programs, no prophylactic vaccine is currently available to prevent infection of DENV and the most effective protective measures are those that avoid mosquito bites. Precedents for successful vaccines exist against Flaviviruses (FV) like the one against yellow fever but vaccine design for other FV seems more complicated (Adams and Boots 2006; Seligman and Gould 2004; Thomas et al. 2006). For example, a primary infection with one strain of DV may predispose an individual to Dengue Hemorrhagic Fever, a more severe disease, if infected subsequently with a different DV strain. It is thus critically important to distinguish the common features of these viruses, as well as differences that may be associated with lethality.

This work aims at comparative analysis of the genomes and proteomes of all four serotypes of dengue virus with the purpose of finding the similarities and differences among all four serotypes. Similarities/differences at both DNA and amino acid level along with

predicting localization of viral proteins in the host cell will hopefully help the future work aiming at controlling the complications linked with the Dengue virus and even in vaccine/drug development.

## MATERIALS AND METHODS

### Sequences Retrieval

Complete genome and amino acid sequences of DENV 1-4 were retrieved from the National Center for Biotechnology Information (NCBI) resource (http://www.ncbi.nlm.nih.gov)

### Nucleotide Sequence Analysis

CLC Sequence Viewer creates a software environment enabling users to make a large number of bioinformatics analyses. In this study CLC sequence viewer 6.5.4 version has been used to find out the nucleotide sequence statistics of all four serotypes of dengue virus. http://www.clcbio.com/products/clc-sequence-viewer/ .

### Amino acid sequence Analysis

Protpsaram, a web based bioinformatics tool was used to calculate molecular weight, isoelectric point, sequence length and number of negatively and positively charged amino acids of all dengue serotypes (http://web.expasy.org/protparam/).

### Predicting Subcellular Localization of Viral<Proteins in Host

Virus mPloc which is freely available on web was used to putatively predict localization of viral structural proteins in the host cell. The capsid, envelop and membrane proteins of four dengue serotypes were predicted for their subcellular localization**.**

## RESULTS AND DISCUSSION

### Nucleotide Sequence Statistics of all four serotypes of Dengue virus.

The comparative nucleotide sequence statistics of different serotypes of Dengue represents that all four serotypes possesses RNA genome. Among all serotypes, DENV-1 possesses the largest genome with a length of 10,735bp hence, bearing highest molecular weight of 3,471.58kDa. Furthermore highest GC contents were observed in DENV-4 i.e. 47.1% and lowest in DENV-2 i.e. 45.8%. DENV-1 and DENV-3 contains an equal amount of GC content i.e. 46.7%. Complete details of nucleotide sequence statistics of all serotypes is shown in table 1.

**Table 1 Nucleotide Sequence Statistics of all Serotypes**

| Serotype | Sequence Type | Length | Count | C+G A+U | Frequency |
|---|---|---|---|---|---|
| DENV-1 | RNA | 10,735bp | 3426,2240, 2770, 2299 | 5010 and 5,725 | C+G=0.467 A+U=0.533 |
| DENV-2 | RNA | 10,723,bp | 3553, 2200, 2713, 2257 | 4913 and 5814 | C+G=0.458 A+U=0.542 |
| DENV-3 | RNA | 10,707bp | 3425, 2218, 2783, 2281 | 5,001 and 5,706 | C+G=0.467 A+U=0.533 |
| DENV-4 | RNA | 10,649bp | 3298, 2214, 2804, 2333 | 5,018 and 5631 | C+G=O.471 A+U=0.529 |

### Isoelectric point, molecular weight and nature of amino acids in Dengue serotypes

ProtParam is a web based tool which is widely used to compute different physical and chemical parameters for a query protein (http://www.expasy.ch/tools/protparam.html). The parameters, among others, include the molecular weight, theoretical pI and amino acid composition Isoelectric point (PI) refers to a pH at which a molecule carries no electrical charge. The theoretical Pi of the four Dengue serotypes ranged from 8, 63 to 8, 77. DENV3 was observed with the lowest PI while DENV2 manifested the highest.

The protein molecular weight of the four serotypes revealed that the DENV3 has the minimum while the DENV2 has the maximum molecular weight (377923,2 & 379803 respectively). The number of amino acids for the four serotypes, however, showed the difference of five amino acids between the largest (DENV1 with 3392 aa) and smallest (DENV4 with 3387 aa). The PI, Molecular weight and Number of amino acids for dengue serotypes are shown in table

**Table 2**: **Molecular weight, theoretical pI and Amino Acid sequence length of Dengue serotypes**

| | DENV1 | DENV2 | DENV3 | DENV4 |
|---|---|---|---|---|
| Number of amino acids | 3392 | 3391 | 3390 | 3387 |
| Molecular weight | 378755 | 379803 | 377923,2 | 378371,4 |
| Theoretical pI | 8,71 | 8,77 | 8,63 | 8,76 |

Positively charged and negatively charged amino acid residues were also calculated for dengue virus and comparison was made among the four serotypes. DENV2 contained the highest number of both positively and negatively charged residues followed by DENV4. Figure1 shows the detailed results.

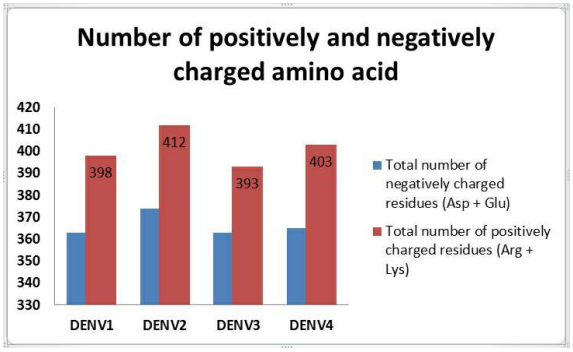


**Fig 1. Comparative analysis of Dengue Serotypes in respect of variation in number of negatively and positively charged amino acids**

### Dengue protein subcellular localization in the host cell

Virus-mPLoc: is a freely available on line tool and is extensively used to predict the subcellular localization of viral proteins within host and virus-infected cells (Shen and Chou 2010). The same tool was used to putatively know the sub cellular localization of capsid, envelop and membrane glycoproteins of all four serotypes of dengue virus. Results are summarized in the following table.

**Table 3: Putative location of Dengue structural proteins in the host**

| Protein | gi number | Amino acid length | Predicted location in host |
|---|---|---|---|
| Capsid protein Dengue virus 1 | gi\|158348409 | 100 | Endoplasmic reticulun |
| Capsid protein Dengue virus 2 | gi\|159024809 | 100 | Endoplasmic reticulun |
| Capsid protein Dengue virus 3 | gi\|164654862 | 100 | Endoplasmic reticulun |
| Capsid protein Dengue virus 4 | gi\|282765486 | 115 | Endoplasmic reticulun |
| Envelope (E) protein [Dengue virus 1 | gi\|158828123 | 495 | Endoplasmic reticulun |
| Envelope (E) protein [Dengue virus 2 | gi\|159024812 | 495 | Endoplasmic reticulun |
| Envelope (E) protein [Dengue virus 3 | gi\|164654853 | 493 | Endoplasmic reticulun |
| Envelope (E) protein [Dengue virus 4 | gi\|73671171 | 495 | Endoplasmic reticulun |
| Membrane glycoprotein Dengue virus 1 | gi\|158828122 | 75 | Endoplasmic reticulun |
| Membrane glycoprotein Dengue virus 2 | gi\|159024811 | 75 | Endoplasmic reticulun |
| Membrane glycoprotein Dengue virus 3 | gi\|158828127 | 75 | Endoplasmic reticulun |
| Membrane glycoprotein Dengue virus 4 | gi\|73671170 | 75 | Endoplasmic reticulun |

All structural proteins of all serotypes of dengue virus are predicted to be localized in the Endoplasmic reticulum of the host. As far as amino acid length of these proteins is concerned, no difference was recorded for each of the three protein types among the serotypes except for capsid protein of DENV4 which was 115 amino acid long as compared to 100 for the rest of three dengue types.

## DISCUSSION

The main focus of this work was to predict with the help of bioinformatics tools the sub-cellular localization of different dengue proteins in the host. Not only the prediction was made for all three structural proteins but also a comparative analysis was made among Dengue serotypes. Surprisingly all structural proteins belonging to all four serotypes were predicted to be localized in the Endoplasmic reticulum of the host cell. In a previous such work these proteins were predicted to be localized in the host plasma membrane(Somvanshi and Seth 2009). The discrepancy probably is due to usage of different versions of the same tool that is Virus ploc and its improved Virus mploc version. As the later was used in this study, we expect more sensitivity. As far as physical and chemical parameters of the amino acid sequences is concerned, DENV2 manifested the highest molecular weight, highest pi and largest sequence length showing the interdependency of these characters on each other. This is in addition to the highest number of both positively and negatively charged amino acids for the same serotype. Among all four serotypes, DENV-4 was observed with relatively high percentage of GC content i.e. 47.1% which indicates that the genome of this serotype is more stable as compared to other serotypes due to the increased percentage of hydrogen bonds between G and C nucleotides. Therefore, this serotype is thermally more stable and high energy is required in order to denature its genome as compare to other serotypes. Contrary to this, the genome of DENV-2 is least thermally stable and cannot tolerate high temperatures and pressures as compared to other three serotypes as its genome contains the less GC content i.e. 45.8%.

## REFERENCES


- Adams B, Boots M (2006) Modelling the relationship between antibody-dependent enhancement and immunological distance with application to dengue. *J Theor Biol* 242: 337-346
- Koonin EV, Senkevich TG, Dolja VV (2006) The ancient Virus World and evolution of cells. Biol Direct 1: 29
- Melo PR, Reis EA, Ciuffo IA, Goes M, Blanton RE, Reis MG (2007) The dynamics of dengue virus serotype 3 introduction and dispersion in the state of Bahia, Brazil. Mem Inst Oswaldo Cruz 102: 905-912
- Prangishvili D, Garrett RA, Koonin EV (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res 117: 52-67
- Seligman SJ, Gould EA (2004) Live flavivirus vaccines: reasons for caution. *Lancet.* 363: 2073-2075
- Shen HB, Chou KC (2010) Virus-mPLoc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites. *J Biomol Struct Dyn* 28: 175-186
- Somvanshi P, ., Seth PK (2009) Comparative Proteome Analysis Of Distinct Variants Of Dengue Virus Using Insilico Approach. *The Internet Journal of Genomics and Proteomics* 4.
- Thomas S, Redfern JB, Lidbury BA, Mahalingam S (2006) Antibody-dependent enhancement and vaccine development. *Expert Rev Vaccines* 5: 409-412
- van der Most RG, Sette A, Oseroff C, Alexander J, Murali-Krishna K, Lau LL, Southwood S, Sidney J, Chesnut RW, Matloubian M, Ahmed R (1996) Analysis of cytotoxic T cell responses to dominant and subdominant epitopes during acute and chronic lymphocytic choriomeningitis virus infection. *J Immunol* 157: 5543-5554