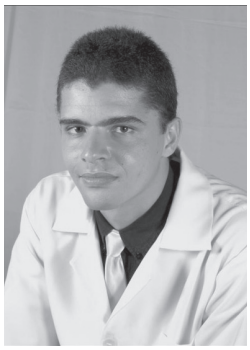
RECIIS
Electronic Journal of Communication Information & Innovation in Health
[www.reciis.cict.fiocruz.br]
ISSN 1981-6286

SUPPLEMENT – BIOINFORMATICS AND HEALTH

Original Articles

Investigation of polymorphisms in the genome of dengue virus

DOI: 10.3395/reciis.v1i2.Sup.102en



Roberto Fabian Santos de Araújo
Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas, Universidade Estadual de Santa Cruz, Bahia, Brazil
menatw@gmail.com

Nicolas Carels
Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas, Universidade Estadual de Santa Cruz, Bahia, Brazil / Centro de Desenvolvimento Tecnológico em Saúde, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil
nicolas.carels@gmail.com

Paulo Roberto Santana de Melo
Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas, Universidade Estadual de Santa Cruz, Bahia, Brazil / Laboratório de Patologia e Biologia Molecular, CPqGM, Bahia, Brazil
pmelo11@yahoo.com.br

Diego Frias
Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas, Universidade Estadual de Santa Cruz, Bahia, Brazil
diego.cepedi@gmail.com

Abstract
In the present study, we investigated the nucleotide polymorphisms among the four genotypes of the Dengue virus - DENV-1, DENV-2, DENV-3 and DENV-4 - by qualitative and quantitative characterization. We studied the Nucleotide Substitution Density (NSD) along each genome identifying the regions with higher degree of mutation and/or conservation. Then, we calculated the Average Nucleotide Substitution Density (ANSD) for each serotype. We observed that the ANSD of DENV-2 is larger than for DENV-1 by 44.21%, DENV-3 by 85%, and DENV-4 by 163.31%. In contrast to DENV-2 and DENV-4, DENV-1 and DENV-3 showed a similar mutational behavior. The NS5 domain from DENV-2 that corresponds to the RNA-dependent RNA polymerase also has a higher mutation rate compared to that of the other DENVs. This suggests that the polymorphism and the virulence can be correlated in DENV-2, which could contribute to the understanding of the disease evolution.

Keywords
Dengue, NS5, polymerase, virulence, polymorphism

Introduction

The Dengue virus (DENV) is the most important flavivirus that causes human disease in Brazil (Figueiredo, 1998). Because Dengue is potentially fatal and has a fast geographical dissemination, it attracted the attention of the Brazilian public authorities in the 90s. The simultaneous presence of two or more Dengue serotypes in the same area - hyperendemicity - increases the risk for complications, such as hemorrhagic fever (FHD) and shock syndrome (SCD), caused by an exaggerated immunological response of the host. Therefore, it is essential to understand how viral polymorphisms can influence the pathogenicity (GUBLER, 1997; 1998).

Dengue is the arbovirus with the largest diffusion in the world. It is found in tropical and subtropical areas where approximately 3 million people are upon risk of infection. Dengue has been identified in more than 100 countries and 2.5 billion people are living in endemic areas (GUZMAN; KOURI, 2002). 50-100 million cases with hundreds of thousands severe cases of the disease (FHD/SCD) and thousands of deaths (about 25.000/year) are reported each year depending on the viral epidemic activity. Dengue has 4 distinct serological types: DENV-1, DENV-2, DENV-3 and DENV-4. They differ in antigenicity and do not induce cross protection, but they have the same epidemiology and cause similar diseases in human. The viral cycle of all serotypes involves human hosts and the mosquito *Aedes aegypti* as the vector (GUBLER, 2002; GONCALVEZ et al., 2002).

The origins of Dengue viruses circulating in Brazil were determined based on the phylogeny of nucleotide sequences. DENV-1 was found to derive from the Caribbean islands while isolates from DENV-2 originated more specifically from Jamaica. Both viruses were probably introduced through the Caribbean (MIAGOSTOVICH et al., 1998; WALNUT et al., 1991). DENV-3 was isolated for the first time in Brazil in 1999 from a patient who was returning from travel in Nicaragua (FIGUEIREDO, 2000).

The majority of FHD/SCD cases were associated with secondary and heterotypical infections, especially during the DENV-2 epidemic with DENV-1 background (FIGUEIREDO, 2000). The successive epidemics of Dengue in Brazil, caused by DENV-1 and DENV-2, culminated with FHD/SCD occurrences. With the continuous circulation of the two serotypes 1 and 2 together, and with the additional high risk of the introduction of other types, Dengue became an important health problem in Brazil (FIGUEIREDO, 2000). Although it has not been possible to establish a clear correlation between one serotype or a particular genotype of DENV and the severity of an epidemic, some DENV-2 and DENV-3 genotypes were associated with DHF. In general, Asian genotypes seem to be more virulent than those initially found in the Americas and in the South Pacific (MESSER et al., 2003; RICO-HESSE et al., 1997; WATTS et al., 1999). In the case of DENV-2, phylogenetic analyses showed that only the American native genotypes were associated with FD while the Asian genotypes were correlated with FHD (CLYDE et al., 2006).

The viral genome of approximately 11 kb encodes a poly-protein that is cleaved during and after translation in three structural proteins (capside C, membrane protein M, and envelope glycoprotein E) and seven non-structural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5) (LINDENBACH; RICE, 2003). Among the non-structural viral proteins, the most extensively characterized are NS3, its cofactor NS2B, and the NS5 protein. The NS3 contains some catalytic domains including a serine protease that requires the NS2B as cofactor. NS3 also shows nucleoside-triphosphatase activity and helicase functions required for the viral RNA synthesis (ARIAS et al., 1993; FALGOUT et al., 1991; II et al., 1999). The protein NS5 acts as a viral RNA-dependent RNA polymerase, besides its role as a metyltransferase (NOMAGUCHI et al., 2003; TAN et al., 1996; EGLOFF et al., 2002). The viral RNA-dependent RNA-polymerase has a well-known low fidelity (DOMINGO; HOLLAND, 1997).

The RNA viruses show a significant genetic variability due to the high intrinsic rate of mutation associated to its RNA-dependent RNA polymerase (DRAKE; HOLLAND, 1999), to its fast replication rate and to its huge population size (HOLMES; TWIDDY, 2003). In this study, we analyzed the degree of genomic polymorphism for the different Dengue virus serotypes. The largest genomic polymorphism was found in DENV-2.

Materials and Methods

A total of 3,278 dengue viral sequences were extracted from the public database EMBL (European Molecular Biology Laboratory), comprising 953 of DENV-1, 1,077 of DENV-2, 1,167 of DENV-3 and 531 of DENV-4. The SRS LION system (Sequence Retrieval System, Release 7.1.3.2) was used to retrieve the sequences at EBI (European Bioinformatics Institute, http://srs.ebi.ac.uk/).

The local processing of files and sequences was performed with Perl scripts. Sequences of each serotype were treated separately. We first eliminated the sequences recognized as part of larger fragments, retaining only the largest non-redundant sequences. Then, the sequences of each serotype were globally aligned one by one with the ClustalW program (http://www.ebi.ac.uk/Tools/clustalw/index.html), using a complete genome sequence of the corresponding serotype as a reference. The reference sequences for the 4 DENVs were selected in the NCBI database (National Center for Biotechnology Information - http://www.ncbi.nlm.nih.gov/). The consistency of each alignment was verified by visual inspection using the program Seaview (http://pbil.univ-lyon1.fr/software/seaview.html).

We calculated the degree of polymorphism of each serotype by counting in each column of the multiple global alignment (~10,700) the number of nucleotides that differ from the corresponding nucleotide in the reference. In other words, if mi is the number of aligned sequences in the i-th column and nri is the amount of nucleotides that match the nucleotide of the reference sequence in that column, then the degree of polymorphism in the i-th position is given by: pi = (mi - nri)/mi. The graph of pi vs i presents a dense succession of peaks that makes the

interpretation difficult (data not shown). Given this, we introduced the function Nucleotide Substitution Density $NSD_j = \frac{1}{L}\sum_{i=1+(j-1)L}^{jL} p_i,\ j = 1,2,\ldots\frac{N}{L}$. NSD is defined for a uniform genome partition in segments of length L. In this study, we selected a length interval of 100 bases that allows smooth curves of NSD along the genome (Figure 1). The graph of NSD for the four serotypes was constructed with the program Scilab 4.1.1 (http://www.scilab.org) using the function smooth that carried out a cubical smoothing by spline. The number of aligned sequences varied with the genome position and the serotype. The number of sequences in the region of the envelope was approximately 5 times larger, in average, than in the remaining intervals. The number of sequences outside the envelope region was never below 100. In order to ensure comparison consistency for each serotype, we calculated the Average Nucleotide Substitution Density ANSD as $ANSD = \frac{L}{N}\sum NSD_j$.

Results and Discussion

After eliminating the redundancy among the 3,278 Dengue sequences retrieved from EMBL, the sequence samples were 757 for DENV-1, 848 for DENV-2, 866 for DENV-3 and 398 for DENV-4. Hence, the size of these samples is large enough to be representative and ensure statistical significance in nucleotide polymorphism analyses.

The curves of the degree of nucleotide polymorphism along the genome for the 4 DENV serotypes are shown in Figure 1. We observed that serotypes 1 and 3 present similar polymorphism profiles, but that they differ substantially from the profiles of serotypes 2 and 4. However, the latter two serotypes differ much more from one another. The Average Density of Nucleotide Substitution of DENV-2 was significantly larger than that of the other serotypes, i.e. 44.21% larger than DENV-1, 85% larger than DENV-3 and 163.31% larger than DENV-4. The region encoding for the envelope protein E presented the highest amount of aligned sequences (5 times more) (data not shown). In this study, we further investigated the NS5 region that encodes the RNA polymerase (ACKERMANN; PADMANABHAN, 2001; NOMAGUCHI et al., 2003; TAN et al., 1996; Egloff et al., 2002). In relation to nucleotide substitutions in this interval, we observed two different patterns: the first one is shared by DENV-1 and DENV-3 and the second is common to DENV-2 and DENV-4, although DENV-2 has a significantly larger level of nucleotide substitutions than the other serotypes. One may speculate that polymerase alteration could promote increase in the mutation rate and explain the largest degree of polymorphism observed in DENV-2. Such hypothesis is currently being investigated by molecular modeling at the protein level. Moreover, since the largest variability within a serotype is associated with the largest degree of mutation, it is possible to think that it could be related to increased virulence and adaptability. The virulence, together with the host susceptibility and environment factors, account for the viral pathogenicity. Amongst these, virulence appears to be a factor obviously dependent upon the genome sequence. Given the existence of four serotypes of Dengue with the same genomic structure, it is possible to presume that the difference in pathogenicity has a functional cause that finds its origin in the sequence. Nucleotide polymorphism analyses help us to identify genomic regions that could be related to Dengue virus virulence.

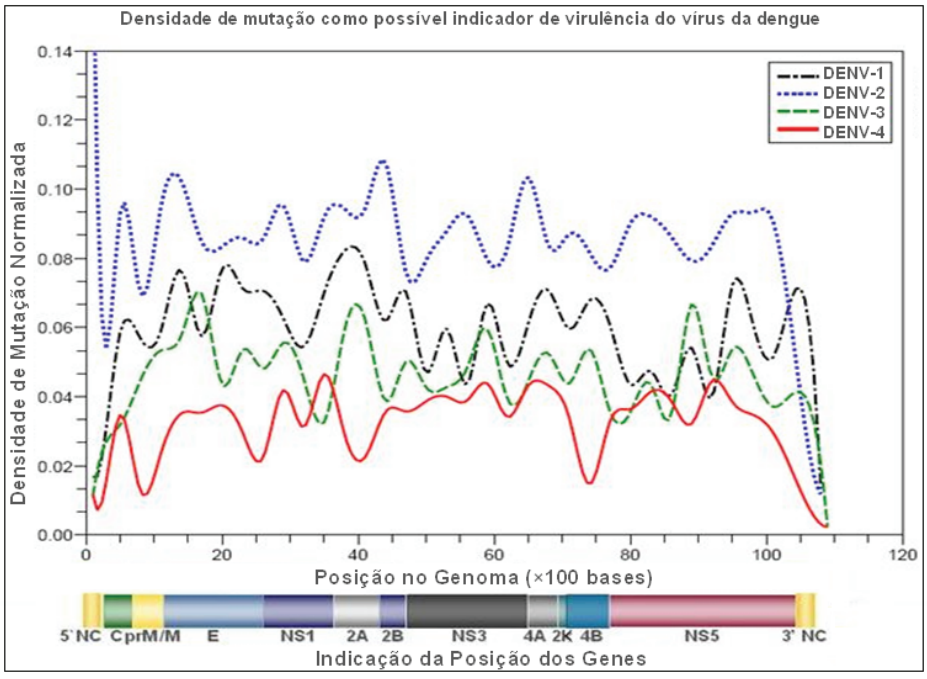


Figure 1 - Density of Nucleotide Substitution along the genome sequence for each serotype of Dengue. The genome was divided in intervals of 100 bases to compute the nucleotide substitution density. The curves were smoothed with spline (program Scilab 4.1.1).

Acknowledgments

This work has been supported by the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) providing fellowships to E. Cardim, W. Reis and N. Carels. We thank Eduardo Costa for the management of NBCGIB computers.

Bibliographic references

ACKERMANN, M.; PADMANABHAN, R. De novo synthesis of RNA by the dengue virus RNA-dependent RNA polymerase exhibits temperature dependence at the initiation but not elongation phase. **J. Biol. Chem.**, v.276, p.39926-39937, 2001.

ARIAS, C.F.; PREUGSCHAT, F.; STRAUSS, J.H. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. **Virology**, v.193, p.888-899, 1993.

CLYDE, K.; KYLE, J.L.; HARRIS, E. Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. **Journal of Virology**, v.80, p.11418-11431, 2006.

DOMINGO, E.; HOLLAND, J.J. RNA virus mutations and fitness for survival. **Annual Review Microbiology**, v.51, p.151-178, 1997.

DRAKE, J.W.; HOLLAND, J.J. Mutation rates among RNA viruses. USA. **Proceedings of National Academy of Science**, v.96, p.13910-13913, 1999.

EGLOFF, M.P. et al. An RNA cap (nucleoside-2 -O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. **EMBO Journal**, v.21, p.2757–2768, 2002.

Falgout, B. et al. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. **Journal of Virology**, v.65, p.2467–2475, 1991.

FIGUEIREDO, L.T. History, present and future of Dengue fever in Brazil. In: Travassos da Rosa, A.P.A.; Vasconcelos, P.F.C.; Travassos da Rosa, J.F.S. (Eds.). **An Overview of Arbovirology on Brazil and Neighboring Countries**. Belem: Instituto Evandro Chagas, 1998. p.154-163.

FIGUEIREDO, L.T. The Brazilian flaviviruses. **Microbes and Infection**, v.2, p.1643-1649, 2002.

GONÇALVEZ, A.P. et al. Diversity and evolution of the envelope gene of dengue virus type 1. **Virology**, v.303, p.110–119, 2002.

GUBLER, D.J. Dengue and dengue hemorrhagic fever: its history and resurgence as a global public health problem. In: **Gubler**, D.J.; **Kuno**, G. (Eds.) **Dengue and Dengue Hemorrhagic fever**. London, UK: CAB International, 1997. p.1-22.

GUBLER, D.J. Dengue and dengue hemorrhagic fever. **Clinical Microbiology Reviews**, v.11, p.480-496, 1998.

GUBLER, D.J. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. **Trends Microbiol.**, v.10, p.100-103, 2002.

GUZMAN, M.G.; KOURI, G. Dengue: an update. **Lancet**, v.2, p.33-42, 2002.

HOLMES, E.C.; Twiddy, S.S. The origin, emergence and evolutionary genetics of dengue virus. **Inf. Genet. Evol.**, v.3, p.19-28, 2003.

LI, H. et al. The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. **Journal of Virology**, v.73, p.3108–3116, 1999.

LINDENBACH, B.D.; RICE, C.M. Molecular biology of flaviviruses. **Advances in Virus Research**, v.59, p.23-61, 2003.

MESSER, W.B. et al. Emergence and global spread of a dengue serotype 3, subtype III virus. **Emerging Infectious Diseases**, v.9, p.800–809, 2003.

MIAGOSTOVICH, M.P. et al. Molecular epidemiology of Den-2 virus in Brazil, **Mem. Inst. Oswaldo Cruz**, v.93, p.625–630, 1998.

NOGUEIRA, R.M.R.; ZAGNER, S.M.O.; MARTINS, I.S.M. Dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS) caused by serotype 2 in Brazil. **Mem. Inst. Oswaldo Cruz**, v.86, p.269-274, 1991.

NOMAGUCHI, M. et al. De novo synthesis of negative-strand RNA by dengue virus RNA dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. **Journal of Virology**, v.77, p.8831–8842, 2003.

RICO-HESSE, R. et al. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. **Virology**, v.230, p.244–251, 1997.

TAN, B.H. et al. Recombinant dengue type 1 virus NS5 protein expressed in Escherichia coli exhibits RNA-dependent RNA polymerase activity. **Virology**, v.216, p.317–325, 1996.

WATTS, D.M. et al. Failure of secondary infection with American genotype dengue 2 to cause dengue hemorrhagic fever. **Lancet**, v.354, p.1431–1434, 1999.

About the authors

Roberto Fabian Santos de Araújo

Graduated in Biomedicine from the State University of Santa Cruz – UESC (Bahia). He worked with protein biochemistry and enzymatic activity. He also obtained a scholarship of scientific initiation (IC) from the Foundation for the Support to Research from the State of Bahia – FAPESB. He developed his work on dengue first at the Laboratory of Bioinformatics from UESC and then at the Nucleus of Computational Biology and Management of Biotechnological Information – NBCGIB, UESC.

Nicolas Carels

Graduated in Agronomical Sciences from the Faculty of Agronomical Sciences of Gembloux (Belgium). He is PhD in Phytopathology from that faculty and also PhD in Genetics from de University Pierre and Marie Curie (Paris, France). In Europe, he carried out research on the genome organization and gene distribution in plants and mammals at the Jacques Monod Institute (Paris, France), the Zoological Station 'Anton Dohrn' (Naples, Italy) and the Center for Astrobioly from the National Institute of Space Technology (Madrid, Spain), successively. In Brazil (State University of Santa Cruz – Bahia), Nicolas worked with Diego Frías (co-author) on the automatic gene search. This research was applied to the gene localization in the genome of the fungal agent of witches' broom disease of cacao.