International Journal of Bio-Science and Bio-Technology
Vol.7, No.6 (2015), pp.25-32
http://dx.doi.org/10.14257/ijbsbt.2015.7.6.03

A Comparative Study of Genetic Variations between Dengue Virus Subtypes 1 and 2

Jin-Hwa Jang[1, 2], Youngmahn Han[1] and Insung Ahn[1, *]

[1]*Biomedical Prediction Technology Laboratory, Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon 305-806, Republic of Korea*
[2]*Laboratory of Computational Biology and Bioinformatics, Institute of Public Health and Environment, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea*
[*]*Corresponding Author: isahn@kisti.re.kr*

Abstract

Because of global climate change, mosquitos which are vectors for the dengue virus (DENV) are increasing in population and the incidence of dengue fever and dengue hemorrhagic fever is increasing every year in countries not only in the subtropics but also in other regions worldwide. There are 4 subtypes of DENV, each differing in infectivity and virulence. This study compared and analyzed genetic variation in the genomic sequences of DENV subtypes 1 and 2. The results showed that while non-synonymous codon variation was high in DENV-1, synonymous codon variation was also high in DENV-2. Additionally, there were differences in the genetic variation in each subtype according to the year or region, and incidence levels increased during periods when sequence variation was high. The results of this study can be used to compare the subtype-specific genetic variation characteristics of DENV to predict the incidence rates and sequence variations of DENV according to the region or year.

Keywords: *Dengue virus, bioinformatics, whole genome, phylogeny, codon variation*

1. Introduction

Dengue fever is an infectious disease that has been rapidly increasing in incidence in recent years. According to a study by the World Health Organization (WHO), the incidence of dengue fever has increased by more than 30-fold in the past 50 years, and more than 1 hundred million patients are diagnosed each year in more than 100 countries [1]. DENV, which causes dengue fever and dengue hemorrhagic fever, is an ssRNA virus that is a member of the *Flaviviridae* family with an 11-kbp genome that includes 3 structural proteins genes (capsid, membrane, and envelope genes) and 7 nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) [2]. DENV is classified into 4 serotypes (subtypes 1–4) and among these, infection by DENV-2 and DENV-4 is most common, while infection by DENV-1 causes more severe symptoms. Particularly in Asia, DENV-2 and DENV-3 cause moresevere disease, including secondary dengue fever and the serotypes differs in virulence and antigenicity [3]. Recently, the infection patterns of DENV suggest that rather than a single infection by a single virus, many infections are caused by more than 2 viruses or by viruses with genetic mutations, in which case the symptoms are more severe [4]. In addition to that , there are differences in the sequence evolution rate; while DENV-3 has a faster evolution rate than DENV-1 and DENV-2, DENV-4 evolution is slower. Among the DENV-2 genotypes, the American/Asian genotype has a much faster substitution rate than the Asian genotype [5]. Dengue fever is endemic to tropical regions in Asia, the South Pacific, Africa, and the American

ISSN: 2233-7849 IJBSBT

continent, while dengue hemorrhagic fever is mostly found in South Asia, South East Asia, the Pacific region, and Latin America. However, with an increase in the number of South East Asian tourists and because of global climate change, mosquitoes, the vector for DENV, are increasing in number and consequently increasing dengue fever infection [1]. As such, analysis of the genetic sequence of DENV in each region can provide a basis for vaccine development.

We confirmed in a previous study that DENV has a unique genomic pattern for each subtype.In addition to that, while most formed groups according to the region within each subtype, in some cases, a similar pattern was observed and extended beyond regional boundaries. These results show that DENV is transmitted among humans through various routes and that changes in genomic patterns between continents on the codon level are mostly in the form of transitions, demonstrating that the intensity of the mutations are not as strong as originally anticipated [6]. Therefore, in this study, we determined the genomic sequence of DENV subtypes 1 and 2 by year and region, analyzed patterns of sequence mutations, and confirmed the phylogenetic sequential differences according to geographic location and time. By using the SimFluVar program [7], which calculates and visualizes the genetic substitution patterns of 2 different groups, we analyzed the codon-based sequence mutations.

2. Materials and Methods

2.1 Collection of Nucleotide Sequence

In the present study, 88 whole genome sequences (UTR5_C_M_E_NS1_NS2A_NS2B_NS3_NS4A_2K_NS4B_NS5_UTR3) included in the Dengue Virus subtype 1 and 2 of The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genbank/) in the United States were collected and used. These sequences included dengue virus subtype 1 and 2 sequences found in the 5 continents of North America, South America, Oceania, Africa, and Asia. The data used spanned over a period of 26 years, from 1985 to 2010 in dengue virus subtype 1 and 27 years, from 1983 to 2009 in subtype 2.

2.2Phylogenetic Analysis

For phylogenetic analysis of whole genome sequences of the dengue virus, multiple sequence alignment was performed on the collated sequences using the ClustalW program (gap open penalty = 40, gap extension penalty = 0.8) [8]. Using the aligned sequences that included the gaps generated from multiple sequence alignment, phylogenetic correlation was performed using MEGA (version 6.06) program [9] to create the maximum-likelihood tree (circular cladogram) based on Tamura-Nei substitution model [10]. Also, we calculated overall mean distance, within group mean distance, between group mean distance, net between group mean distance, within subpopulation mean diversity, entire population mean diversity, inter-populationdiversity, and coefficient of differentiation using MEGA (version 6.06) program.

2.3 Comparison of Codon Substitution Pattern

To ascertain the codon substitution pattern of the dengue virus, the continents on which each subtype was found were grouped individually, and the substitution rate of the third base of each codon was calculated. Analysis was performed using the SimFluVar (http://lcbb.snu.ac.kr/simfluvar) program, which enabled sequence analysis between different genome groups. The SimFluVar is an analytical tool for calculating the codon substitution patterns of influenza virus. Designed to compare

a large number of nucleotide sequences, SimFluVar provides precise patterns of codon variations between two viral groups, especially for the influenza virus. SimFluVar also provides the useful functions, such as editing and visualization of the result matrix [6].

3. Results

Phylogenetic analysis of DENV subtypes 1 and 2 showed that individual subtypes formed clades (Figure 1A). This demonstrates that even within the genomic sequence of DENV, there are differences according to subtype. Based on the pattern of sequence mutations per region for subtype 1 from 1985 to 2010, sequences in viruses extracted from cases in the USA from 1986 to 1996 showed very similar characteristics. However, sequences from 2010 were similar to those in viruses extracted from cases in Vietnam in 1997. Viruses from cases in Mexico, in North America, showed similar characteristics to DENV from cases in Nicaragua, in Central America. The genomic sequences of DENV from cases in Vietnam from 1997 to 2006 formed a group with those in Brazil in 2008 and Colombia in 1998. In subtype 2, viruses from cases in the USA from 1986 to 2004 were nearlysimilar to viruses from cases in Vietnam in the 1990s. In contrast, viruses extracted from cases in the 2000s in Vietnam were more similar to those in Colombia and Nicaragua. Additionally, DENV sequences from countries in Asia such as China, Taiwan, or Vietnam were differentiated from viruses from other continents and were included in a separate group.

Figure 1B shows the results of producing a transition matrix of the codon mutational pattern for DENV subtypes 1 and 2, by using SimFluVar, displayed as a heat map view. The results revealed numerous codon mutations between subtypes 1 and 2, and the mutations occurred not only in synonymous codons close to the diagonal line, but also in non-synonymous codons far from the diagonal line. By using MEGA 6, the overall mean distance between the 2 groups was calculated to be 0.51, the within group mean distance was 0.07, the between group mean distance was 0.98, and the net between group mean distance was 0.91. In addition to that, when diversity was calculated, the within subpopulation mean diversity was 0.07, the entire population mean diversity was 0.51, the inter-population diversity was 0.44, and the coefficient of differentiation was 0.86.

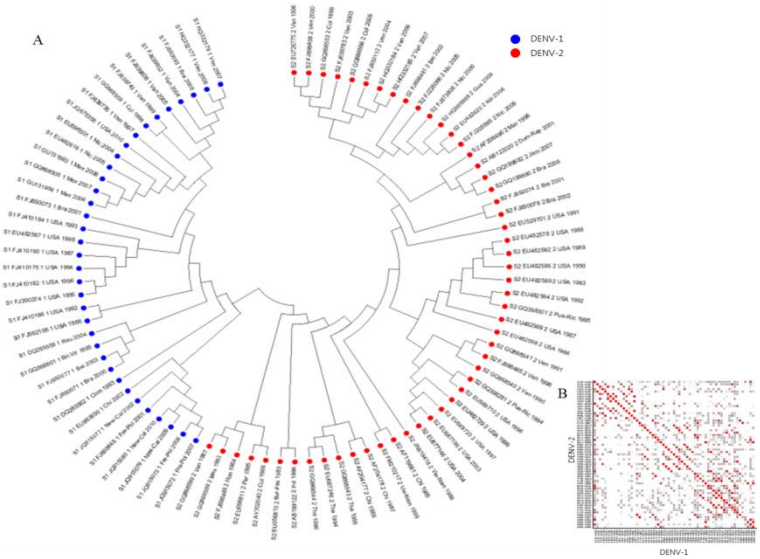

Figure 1. A; Molecular Phylogenetic Analysis of Dengue Virus Subtype 1 and 2 by Maximum Likelihood Method B; Codon Variation Pattern between Dengue Virus Subtype 1 and 2

A; The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood (-72272.7682) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 88 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 10453 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [9]. Blue and red color for subtype 1 and 2. B; Comparisons of codon variation pattern between dengue virus subtype 1 and 2 using SimFluVar [7]

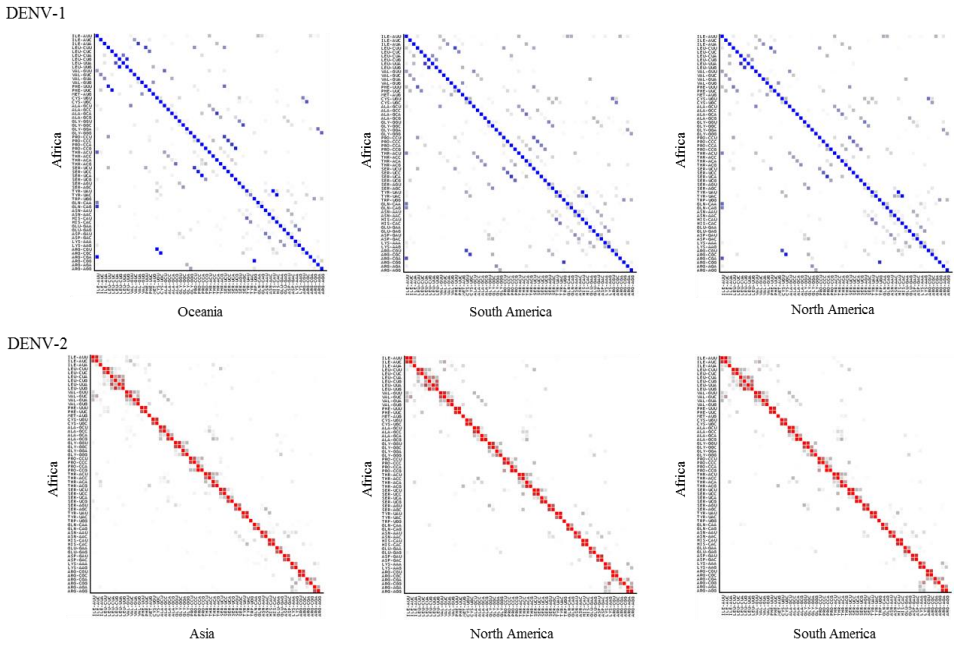

Figure 2. Heat Map View of Transition Matrix According Continents from Dengue Virus Subtype 1 (Blue) and 2 (Red)

Transition matrix consist mainly of two tabs: a tab for viewing the output in Table, format and a tab for viewing the output of graphs in the heat map format. The control box on the rightside enables users to change shapes and colors in the heatmap. The heat map scale function allows users to enlargeor reduce the heat map, and the outline color functionassigns colors to the outline of the small rectangle blocks inthe Heat map view. High, middle, and low colors determinethe manner in which colors are displayed within the Heat mapmatrix. The color scale sets up the scale of color distribution.The closer this value is to zero, the clearer the distinction isbetween the blocks [7].

Figure 2 shows the sequence mutation differences between DENV subtypes 1 and 2 according to region as determined using SimFluVar [7]. For DENV-1, the transition matrix results were similar in Oceania, South America, and North America with respect to Africa, and a similar topology was observed on the heat map view. Thus, the locations of codon mutations were similar between the 3 groups, with the only difference in the mutation rate, leading to differences in only the color intensity of each point. In contrast, in DENV-2, the diagonal line, although straight, formed squares in certain areas. This shows that although the nucleotide sequence is changing, synonymous codon mutations frequently occur, in which replacement with the same amino acid occurs. There were no

mutations in the protein-forming sequences. In DENV-1, non-synonymous codon mutations were more frequent than synonymous codon mutations, indicating that the nucleotide sequence is changing, leading to changes in the amino acid sequence.

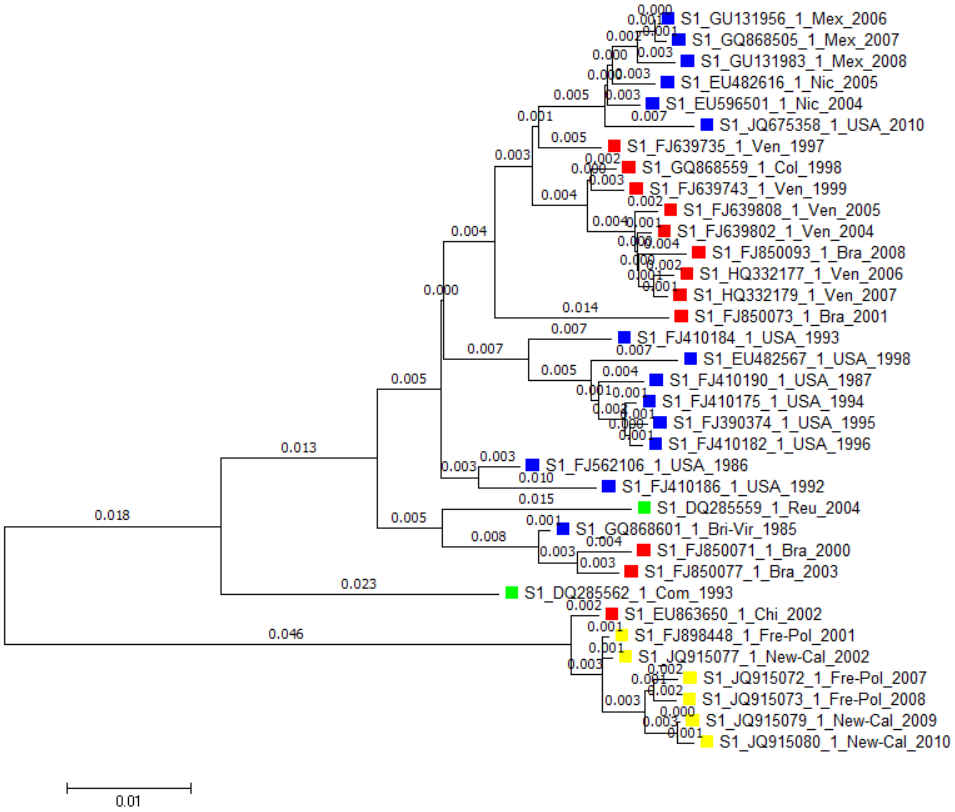

Figure 3. Molecular Phylogenetic Analysis of Dengue Virus Subtype 1 by Maximum Likelihood Method

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood (-32493.9018) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (above the branches). The analysis involved 35 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 10587 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Green, red, blue, and yellow color represents the African, South American, North American, and Oceanian Continent, respectively.

Figure 3 shows the results of the correlation analysis among whole genome sequences of the 4 strains of dengue virus. The 4 subtypes of dengue virus each displayed their own phylogenetic genomic patterns, and within each subtype, most were located in close lineage to the Americas (South and North America). They showed different patterns compared to other phenotypes, based on the outbreaks on the continents of Asia, Africa, and Oceania.In particular, strains from South America and North America form a large group among other groups, indicating that the DENV genomic sequence has similar characteristics within the American continent. In contrast, DENV found in Chile in 2002

showed characteristics that werenearlysimilar to the DENV found on Oceania, a different continent. This is caused by the transcontinental migration of mosquitos, which are the hosts or vectors of DENV.

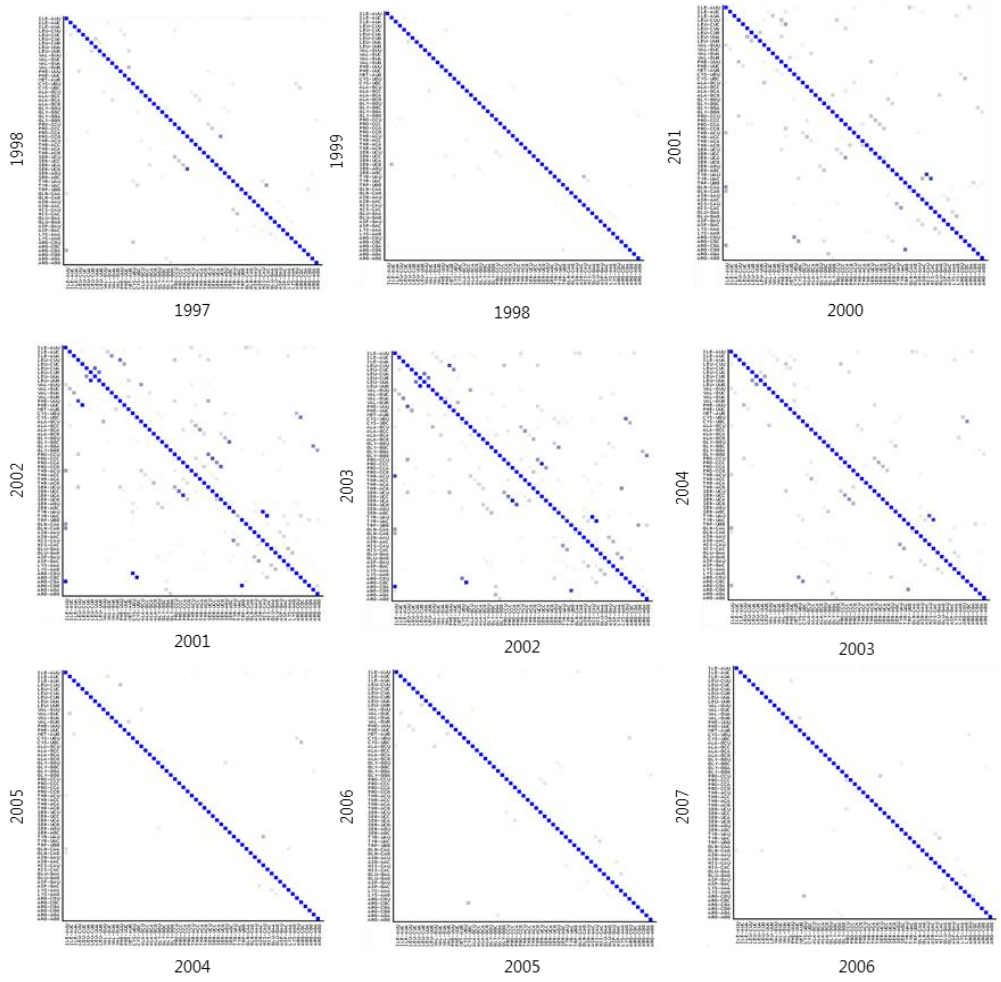

Figure 4. Heat Map View of Transition Matrix According Years from Dengue Virus Subtype 1 from South America

Transition matrix consist mainly of two tabs: a tab for viewing the output in Table format and a tab for viewing the output of graphs in the heat map format. The control box on the right side enables users to change shapes and colors in the heat map. The heat map scale function allows users to enlarge or reduce the heat map, and the outline color function assigns colors to the outline of the small rectangle blocks in the Heat map view. High, middle, and low colors determine the manner in which colors are displayed within the Heat map matrix. The color scale sets up the scale of color distribution. The closer this value is to zero, the clearer the distinction is between the blocks [7].

Figure 4 shows the trend in genome sequence mutations for DENV subtype 1 extracted in South America from 1997 to 2007 that was determined using SimFluVar [7]. The graph shows that there were no large sequence mutations from 1997 to 1999. However, transition matrix results from 2000 and 2001 showed higher sequence mutation rates compared to the previous year, with the highest rates observed from 2001 to 2003. Although these values were not as high as in previous years, more sequence mutations were observed in 2003–2004 compared to 1997. From 2004 to 2007, there were few sequence mutations according to the transition matrix results. In conclusion, during the

11 years from 1997 to 2007, DENV found in South America showed no sequence mutations in the early years, showed high levels of non-synonymous codon mutation rates in the early-to-mid 2000s, and then showed no mutations in the late 2000s.

4. Discussion and Conclusions

In our previous study, according to the phylogeny of the genome sequences of type 1 dengue virus used in the present study, the dengue virus reported in South America in 2002 (EU863650) was observed to show a pattern similar to that of the virus found in Oceania in 2001 and 2002 [6]. This virus, which was first isolated from a patient with dengue fever in Chile, was found to have a high sequence homology with Pacific DENV-1 genotype IV viruses, which was also the case in the present study [11]. Such high similarity of genetic sequences among different continents may be possible due to the free movement of humans, virus hosts, mosquitoes, and the media between the continents. Therefore, the worldwide dispersion pattern of dengue virus would show that transmission can occur to any location, regardless of the geographical area. In particular, DENV-2 and DENV-3 have caused severe secondary dengue infections in Asia [12]. Because virulence and antigenicity can vary depending on the viral serotype, measures should be taken to prevent worldwide transmission of highly virulent virus strains. Accordingly, sequence analysis of dengue viruses found in each region and prediction of their sequence mutations can provide a basis for studies for the development of vaccines and other treatment modalities for potential future virus outbreaks. Each subtype of the dengue virus had genomic patterns that were unique to each subtype, and within each subtype, they were grouped based on the continents they were isolated from within each subtype. However, there were a few cases that showed similar patterns, overcoming the geographic barriers posed by the continents. These results show that transmission of the dengue virus between humans can occur via various routes. However, changes in the genomic patterns at the level of the codons between the continents were majorly due to transition, and thus, the severity of the mutation was low. The results of the genome sequence analysis confirmed that dengue viruses were transmitted between different continents, not just between different countries. Further, measures to inhibit domestic introduction of tropical infectious diseases, such as the dengue virus, are warranted in the future.

In this study, we analyzed the genomic sequence differences between DENV subtypes 1 and 2 and analyzed which codon mutations occurred for each subtype per region and year. There was a clear difference in the genomic sequences of the 2 groups; moreover, the phylogenetic tree revealed the formation of 2 clades. Furthermore, this difference in sequence affected the virulence of the virus, and infection by the DENV-1 subtype showed more severe symptoms compared to infections by the other subtypes [3]. The overall mean distance between DENV-1 and DENV-2 was 0.51, which means that only 50% of their sequences were similar. In contrast, the within group mean distance was 0.07, showing a 7% difference in sequence within subtype.In addition to that, the sequence mutation differences for each subtype per region and year showed that while the level of non-synonymous codon mutations was high in DENV-1, levels of synonymous codon mutations were higher in DENV-2. Thus, while in DENV-1 genomic sequence mutations were accompanied by amino acid sequence change, in DENV-2, although nucleotide sequence mutations were present, there were few amino acid mutations. In DENV-2, no mutations were observed in the protein sequence or structure, but sequence mutations were observed in nucleotide sequence analysis. As such, meaningful results were obtained by using computational and informational analyses in addition to biological analysis. These differences in sequence mutations are likely related to the fact that DENV-1 has low infectivity compared to the other subtypes, but shows higher virulence on infection. Moreover, an increase in virulence because of the high sequence mutation

rate has been reported in a study examining the influenza virus [13].In addition to that, upon analysis of the sequence mutations of the South American DENV-1 virus by year, we confirmed a high mutation rate in some years. In particular, from 2001 to 2003, which was a period of high sequence mutations, the total number of dengue fever patients increased in America [14]. Based on these results, convergent research into subtype-specific sequence mutations in DENV and their relationship with increases in virulence and numbers of affected patients should be examined in the future by using a combination of biological analysis and informational technology.

Acknowledgements

This research was supported by Biomedical Prediction Technology Laboratory at the Korea Institute of Science and Technology Information (K-15-L03-C02-S03)

References

[1] "World Health Organization", **(2015)**, Denguecontrol.http://www.who.int/denguecontrol/en/.
[2] T. J.Chambers, S. H. Chang, R. Galler and C. M. Rice, "Flavivirusgenome organization, expression, and replication", Annu. Rev. Microbiol, vol. 44, no. 649, **(1990)**.
[3] S. S.Twiddy, E. C. Holmes and A. Rambaut, "Inferring the Rate and Time-Scale of Dengue Virus Evolution", Mol. Biol. Evol, vol. 20, no. 122, **(2003)**.
[4] M. Worobey, A. Rambaut and E. C. Holmes, "Widespread intra-serotype recombination in natural populations of dengue virus", Proc Natl Acad Sci., vol. 96, no. 7352, **(1999)**.
[5] L. B. Figueiredo, T. Sakamoto, L. F. L. Coelho, E. S. de Oliveira Rocha, M. M. G. Cota, G. P. Ferreira, J. G. de Oliveira and E. G. Kroon, "Dengue Virus 2 American-Asian Genotype Identified during the 2006/2007 Outbreak in Piauı", Brazil Reveals a Caribbean Route of Introduction and Dissemination of Dengue Virus in Brazil.PLosOne, vol. 9, no. 8, **(2014)**, e1045162014.
[6] I. Ahn, J. H. Jang, Y. Han and T. Q. Tung. "Phylogenetic Analysis of Dengue viruses using Bioinformatics Techniques", Advanced Science and Technology Letters, vol. 105, no. 12, **(2015)**.
[7] I. Ahn, J. H. Jang, H. Y. Kim, J. H. Lee and H. S. Son, "A Visualization Tool for Calculating the GeneticSubstitution Patterns Between Two Different Groups", Evol. Bioinform, vol. 11, no. 179, **(2015)**.
[8] M. A. Larkin, G. Blackshields, N. P. Brown, R. P. A. Chenna, McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, W. Clustal and X. Clustal, version 2.0. Bioinformatics, vol. 23, no. 2947, **(2007)**.
[9] K. Tamura, G. Stecher, D. Peterson, A. Filipski and S. Kumar, "MEGA6: Molecular Evolutionary Genetics Analysis", version 6.0. Mol. Biol. Evol., vol. 30, no. 2725, **(2013)**.
[10] K. Tamura and M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees", Mol. Biol. Evol., vol. 10, no. 512, **(1993)**.
[11] C. Cáceres, V. Yung, P. Araya, J. Tognarelli, E. Villagra and J. Fernández, "Complete nucleotide sequence analysis of a Dengue-1 virus isolated on Easter Island", Chile Archives of Virology, vol. 153, no. 1967, **(2008)**.
[12] S. S. Twiddy, E. C. Holmes and A. Rambaut, "Inferring the Rate and Time-Scale of Dengue Virus Evolution", Mol. Biol. Evol., vol. 20, no. 122, **(2003)**.
[13] K. S. Li, Y. Guan, J. Wang, G. J. D. Smith, K. M. Xu, L. Duan and J. S. M. Peiris, "Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia", Nature, vol. 430, no. 209, **(2004)**.
[14] M. G. Teixeira, M. C. Costa, F. Barreto and M. L. Barreto, "Dengue: twenty-five years since reemergence in Brazil", Cad. Saude. Publica., vol. 25, no. S7, **(2009)**.

