International Journal of Social Welfare Promotion and Management
Vol. 3, No. 1(2016), pp. 19-24
http://dx.doi.org/10.21742/ijswpm.2016.3.04

Phylogenetic Analysis of Human-origin Zika Virus Isolated from Different Geographic Regions

Jooyeon Park[1], Jinhwa Jang[2] and Insung Ahn[3]

[1]*Dept. of Big Data Science, University of Science & Technology, 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Korea*
[2]*Biomedical Prediction Technology Laboratory, Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea*
isahn@kisti.re.kr

Abstract

Mosquito-mediated Zika virus, which causes microcephaly in infants of infected pregnant women, has spread to Central America and Asia, and the number of infected individuals has increased worldwide. Zika virus, which belongs to the Flavivirus genus, was first isolated from a monkey in 1947 and then was found to infect humans, subsequently spreading to many countries. Since there are no preventive vaccines or drugs against Zika virus, it is necessary to analyze the phylogenetic characteristics of the virus over time and according to location using genome sequencing analysis. In this study, we conducted phylogenetic analysis of the genome sequence of Zika virus based on 10 genetic loci and analyzed gene functions based on phylogeny according to continent and time. From these studies, we found that the envelope, NS1, and NS5 genes had higher phylogenic accuracy than other genes and that the viral sequences found in these 10 genes had evolved through variations.

Keywords: *Zika virus, Phylogenetic analysis, Gene ontology, Genome, Genetic marker*

1. Introduction

Several reports have described Zika virus infection in many countries, and the risk of that has been increasing worldwide. According to a report by the World Health Organization (WHO) in September 2016, cases of Zika virus infection occurred in Brazil and several countries in Southeast Asia, including Singapore, Philippines, Malaysia, and Vietnam [1][2].

Zika virus was first discovered in a rhesus monkey in the Zika forest in Uganda, Africa in 1947 [3]. Before 2007, Zika virus was regarded as a locally restricted virus occurring only rarely in humans with relatively minor symptoms [4][9]. However, after the first outbreak of Zika virus, which reported about 100 infected patients on Yap Island of the Federated States of Micronesia in the Western Pacific Ocean, approximately 30,000 people were infected in French Polynesia in the South Pacific Ocean in 2013; this was considered the beginning of the Zika virus pandemic [5][6]. Thereafter, in 2014, Zika virus spread to South America, Central America, and the Caribbean across the Pacific Ocean [4][7]. In particular, approximately 0.4–13million people were infected by Zika virus in Brazil in 2015 [8].Moreover, many infants born to infected Brazilian mothers exhibited microcephaly [10].Due to these severe symptoms, the WHO declared a state of emergency against Zika virus infection on February 1, 2016, emphasizing the risk of Zika virus infection to the global population [11].

Article history:
Received (July 26, 2016), Review Result (September 25, 2016), Accepted (October 27, 2016)

Print ISSN: 2205-8435, eISSN: 2207-5321 IJSWPM

The complete genome sequence of Zika virus is now publically available, and the genome sequence of Zika virus is divided into three structural genes (C, prM, and E) and seven nonstructural NS genes (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5), which are found in the following orderin the genome: 5'-capsid(C)-premembrane(prM)-envelope(E)-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'[12][13]. In this study, we aimed to investigate genetic variations through sequence analysis of the 10 genes of Zika virus and evaluate the relationships between genetic features and environmental factors through phylogenetic tree analysis.

2. Genetic and molecular characters of ten genome region in zika virus

The envelope (E) gene, which encodes most surface proteins of Zika virus, plays an important role in viral cycle and receptor binding and fusion in membrane [14]. NS1, a nonstructural protein known as an antigenic marker of Zika virus, plays an important role in viral replication and infection [15]. In addition, NS1 protein binds toextracellular innate and acquired immune response factors, facilitating the effects of the NS1 protein on immune invasion and pathogenesis [16]. Most transmembrane domains are composed of products of NS2A and NS2B genes, which encode small hydrophobic proteins [13]. NS2A is involved in processing of NS1 through cleavage by protease in the cytoplasm [17]. In addition, a protease resulting from binding of NS2B with NS3 cleaves polyprotein together with proteases in the host to facilitate viral replication [18]. Similar to the NS2A and NS2B genes, the NS4A and NS4B genes encode small hydrophobic proteins with multiple transmembrane domains [13]. The NS4 protein interacts with the NS3 and NS5 proteins to mediate membrane localization [17]. Moreover, with NS1 gene, the NS4 in the Asian clade of epidemic Zika virus have recently been reported to have variations in codon usage, suggesting that the NS1 and NS4 genes may have evolved to be compatible with human hosts without major variations in protein sequences [19]. The NS5 gene is the longest and most well-conserved gene among the 10 Zika virus genes [14]. NS5 functions as an RNA polymerase in the cytoplasm and is critical for viral replication [17].Additionally, the NS5 gene of Zika virus degrades the transcription factor, STAT2, which is involved in the innate immune system in human hosts, thereby blocking signal transduction for antiviral gene expression from the interferon to the nucleus [20].

3. Methods

In order to construct phylogenetic trees for the 10 genes of Zika virus, we collected 67 complete genome sequences of Zika from the NCBI GenBank(www.ncbi.nlm.nih.gov/genbank/). Most sequence data were missing detailed locus information for the 10 genes. Thus, the data from these 67 genome sequences were subjected to multiple sequence alignments through Clustal X 2.1 [21], using the following parameters: gap opening, 15; gap extension, 6.66; delay divergent sequences, 30%; and DNA transition weight, 0.5. Thereafter, the genetic locus of each sequence was determined based on locus information of the 10 genes from NC012532.1, the reference genome. For splicing of the complete genome sequences into each gene locus, a script written in JAVA language was used to extract the capsid gene (128–440 bp), prM gene (774–998 bp), envelope gene (999–2510 bp), NS1 gene (2511–3566 bp), NS2A gene (3567–4244 bp), NS2B gene (4245–4634 bp), NS3 gene (4635–6485 bp), NS4A gene (6486–6866 bp), NS4B gene (6936–7689 bp), and NS5 gene (7690–10398 bp) from the total of 10,830 bp. Polyprotein genes that were present between the capsid gene and prM gene and between the NS4A gene and NS4B gene were not considered. Phylogenetic trees of the 67-sequence data for each gene were constructed using MEGA 7 by

applying the maximum likelihood (ML) method [22], the Tamura-Nei model as a substitution model, and the standard genetic code. Gaps and missing data were treated by complete deletion,and the nearest-neighbor-interchange (NNI) method was applied as a heuristic method for ML. Initial phylogenetic trees were determined based on the NJ/BioNJ default values.

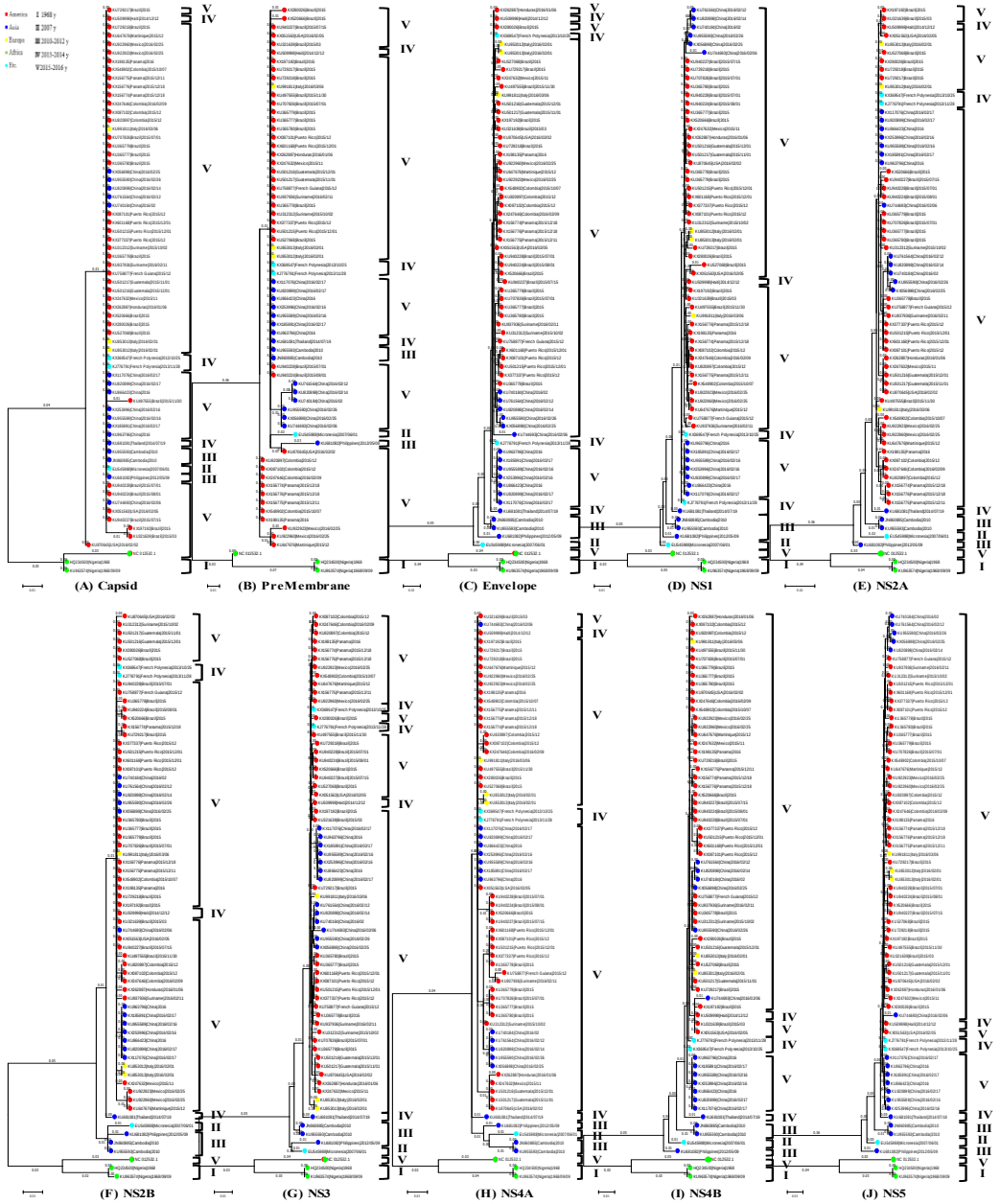


Figure 1. ML tree using 10 genes of zika virus

4. Results and discussion

[Figure 1] show the phylogenetic trees of the 10 structural and nonstructural genes from the 67 Zika viral sequences. Phylogenetic trees were analyzed with classifications following five geographical regions, including America (red), Asia (blue), Europe (yellow), Africa (green),

and other areas (light blue), and five time points, including 1968 (I), 2007 (II), 2010–2012 (III), 2013–2014 (IV), and 2015– 2016 (V).

The phylogenetic trees for the 10 genes commonly showed divergences from the Zika virus sequence isolated from a Nigerian strain isolated in 1968 into outgroups. Thus, Zika virus showed changes in gene sequences over time, which resulted in genetic differentiation from the recent epidemic Zika virus. In contrast, sequences of Zika virus isolated from Micronesia in 2007 and the Pacific Ocean area of Polynesia in 2013 were mixed with other taxa without diverging into outgroups. In other words, the Zika virus that was epidemic in 2007 and 2013 had genetic factors similar to those of recently isolated epidemic Zika viruses found worldwide. In addition, NC012532.1, the reference genome, was isolated from viruses collected in Uganda, Africa, which diverged into outgroups together with the Nigerian branch, resulting in these viruses having positions close to each other. NC012532.1 was collected in 2016, whereas HQ234500 and KU963574 of Nigeria were collected in 1968 and nevertheless diverged into a local outgroup in Africa, suggesting that the sequences of the African clade retained specific features, over time.

In [Figure 1], (A) shows the phylogenetic tree of the capsid gene encoding the protein coat for the Zika virus genome. The branch lengths at the roots of the capsid phylogenetic tree were as short at 0.04 and 0.03, indicating relatively low variation; this could be explained by the observation that most Zika viruses have coat proteins with similar structures. (B)shows the phylogenetic tree of the prM gene, which exhibited highly disperse temporal and geographical distributions because the sequence of the prM gene is conserved among viruses. (C), which shows the tree of the envelope gene, the American and the Asian clade were distinctly diverged, indicating that the distribution of the host human genes differed depending on the region, leading to differences in the sequences of genes encoding components of the viral membrane important for binding to the host. (D)shows the phylogenetic tree of the NS1 gene. This gene was found to show increased divergence depending on the region and time compared to that in other genes, indicating that the NS1 gene had similar sequence distributions depending on region and had evolved to adapt to environments through gradual variations. As shown in (E), the tree of the NS2A gene, unlike those of the other nine genes, showed a scale as high as 0.02. In other words, the NS2A gene had a higher frequency of variations than the other genes, as supported by the mixed positions of the NS2A gene on the phylogenetic tree without forming regional clades. (F) and (G) show the phylogenetic trees of the NS2B and NS3 genes. Owing to the roles of these two proteins as proteases in the host cytoplasm, they were conserved without forming clades by time or region. For the two phylogenetic trees of NS4 genes shown in (H) and (I), the tree of the NS4B gene (I) was more distinctively separated by region and chronologically better ordered than that of the NS4A gene (H). Thus, because NS4A functions in localization to the viral membrane, regional and temporal sequence variations were relatively small, whereas the NS4B gene underwent gradual variations, playing a role in viral evolution. The NS5 gene, shown in (J), exhibited the most distinct temporal distribution on the phylogenetic tree among the 10 genes and formed clades based on the country of origin. These data suggested that NS5, which blocks cellular immune signaling required for viral replication in the host cells, underwent evolution to improve host compatibility.

Notably, the 10 phylogenetic trees showed that the sequences of KU744693 (China, 20160206) were distinct from those from the other Asian areas. In most phylogenetic trees, genes of KU744693 showed protruding nodes or they were localized among sequences found in American regions. Based on these results, we speculated that KU744693 sequences were isolated from patients who were infected by Zika virus in the American regions, or from patients who had characteristics different from those in the other Asian regions.

5. Conclusions

Through sequence analysis of the 10 genetic loci, we investigated variations in the 10 genes and performed phylogenetic analysis of the characteristics of these genes. Based on the genetic loci of the reference genome sequence isolated in Uganda in 2016, 10 genetic loci were extracted from the aligned sequences using a JAVA script, and phylogenetic analysis was then performed for each locus.

We found that the envelope, NS1, and NS5 genes, which were often used as genetic markers of Zika virus, tended to form better regional clades than other genetic loci. Thus, these three genes should be appropriate for representing the environmental distributions of the hosts and explaining viral evolutionary processes. Of the 10 structural and nonstructural genes, the C, prM, NS2B, NS3, and NS4A genes had conserved features and they were not affected by the environment. These genes had mixed distributions on the phylogenetic trees, regardless of the time or region, thereby affecting viral replication in the host. In contrast, the E, NS1, NS2A, NS4B, and NS5 genes exhibited variations that may have been acquired during evolution of host compatibility. These genes showed more distinct distributions according to region and chronology. In addition, most gene sequences became distinct from previous sequences with time, resulting in new sequence features.

Within the 7-month period after March 2016, 14 patients with Zika virus were reported in South Korea [23]. All of these patients were infected after travel to countries with Zika virus epidemics. Ten patients were confirmed to have traveled to Southeast Asian countries, including the Philippines, Vietnam, and Thailand [23]. Moreover, the pandemic area of Zika virus has recently been shown to have moved from Central America to Southeast Asia; because South Korea is geographically close to Southeast Asia, it is important to identify the genetic characteristics of these viruses and perform phylogenetic analysis of Zika viruses isolated from different continents. Such data are expected to reveal significant genetic and phylogenetic factors that can potentially affect the spread of foreign-borne Zika virus in South Korea. In addition, the outcomes will facilitate the selection of major gene targets for the development of vaccines and drugs against Zika virus, which may contribute to preventing the spread of Zika virus worldwide.

Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP (2016M3A9B6915714).

References

[1] WHO Zika situation report, 22 September, **(2016)**.
[2] Singapore Government Ministry of Health, ZIKV report
[3] A.D. Haddow, A.J. Schuh, C.Y. Yasuda, M.R. Kasper, V. Heang, R. Huy and S.C. Weaver, "PLoS Neglected Tropical Disease", Vol. 6, No. 2, **(2012)**.
[4] A.S. Fauci and D.M. Morens, New England Journal of Medicine, Vol. 374, No. 7, **(2016)**.
[5] M.R. Duffy, T.H. Chen, W.T. Hancock, A.M. Powers, J.L. Kool, R.S. Lanciotti and L. Guil-laumot, New England Journal of Medicine, Vol. 360, No. 24, **(2009)**.
[6] A. Roth, A. Mercier, C. Lepers, D. Hoy,S. Duituturaga,E. Benyon and Y. Souares, "Euro Surveillance", Vol.19, No. 41, **(2014)**.
[7] E.J. Rubin, M.F. Greene andL.R. Baden, New England Journal of Medicine, Vol. 374, No. 10, **(2016)**.
[8] WHO Zika situation report, 3 March **(2016)**.

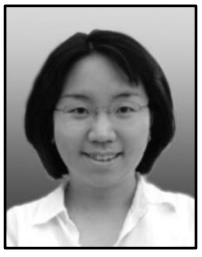
[9] G. Vogel, "Science Magazine American Association for the Advancement of Science", Retrieved/2016.–DOI 10, **(2015)**.
[10] A.S. Oliveira Melo, G. Malinger, R. Ximenes, P.O. Szejnfeld, S. AlvesSampaio and A.M. Bispo de Filippis, "Ultrasound in Obstetrics & Gynecology",Vol. 47,No. 1, **(2016)**.
[11] WHO statements, 1 Feb **(2016)**.
[12] G. Kuno, and G.JJ. Chang, "Archives of virology", Vol. 152, No. 4, **(2007)**.
[13] Z. Zhu, J.F.W. Chan, K.M. Tee, G.K.Y. Choi, S.K.P. Lau, P.C.Y. Woo and K.Y. Yuen, "Emerging microbes & infections", Vol.5, No. 3, **(2016)**.
[14] O. Faye, C.C. Freire, A. Iamarino,O. Faye, J.V.C. de Oliveira,M. Diallo and P.M. Zanotto, "PLoS Neglected Tropical Disease", Vol. 8, No. 1, **(2014).**
[15] H.Song, J. Qi, J. Haywood, Y. Shi and G.F. Gao, "Nature structural & molecular biology", **(2016)**.
[16] W.C. Brown, D.L. Akey, J.R. Konwerski, J.T. Tarrasch, G. Skiniotis, R.J. Kuhn and J.L. Smith, "Nature Structural & Molecular Biology", **(2016)**.
[17] T.J. Chambers, C.S. Hahn, R. Gallerand C.M. Rice, "Annual Reviews in Microbiology",Vol. 44,No. 1, **(1990)**.
[18] J. Lei, G. Hansen, C. Nitsche, C.D. Klein, L. Zhang and R. Hilgenfeld, Science, Vol. 353, 6298, **(2016)**
[19] PK. Russell, "PLOS Neglected Tropical Disease", Vol.10, No. 3, **(2016)**.
[20] A. Grant, S.S. Ponia, S. Tripathi, V. Balasubramaniam, L. Miorin, M. Sourisseau and A. García Sastre, "Cell host &microbe", **(2016)**.
[21] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, "Bioinformatics",Vol. 23,No. 21, **(2007)**.
[22] S. Kumar, G. Stecher andK. Tamura, "Molecular biology and evolution", Vol.33, No. 7,**(2016)**.
[23] MOHW Korea Zika news report, 25 September **(2016)**.

Authors

Jooyeon Park

Student,
University of Science and Technology
KISTI Campus
Dept. of Big Data Science

Jinhwa Jang
Researcher,
Korea Institute of Science and Technology Information
Biomedical Prediction Technology Lab.



InsungAhn
Principal Researcher,
Korea Institute of Science and Technology Information
Biomedical Prediction Technology Lab.
Professor,
University of Science and Technology, KISTI Campus
Dept. of Big Data Science