Comparison of Genetic Variations in Zika Virus Isolated From Different Geographic Regions

Jooyeon Park, Dept. of Big Data Science, University of Science & Technology, Daejeon, Republic of Korea
Jinhwa Jang, Biomedical Prediction Technology Lab, Korea Institute of Science and Technology Information, Yuseong-gu, Republic of Korea
Insung Ahn, University of Science and Technology, Daejeon, Republic of Korea

ABSTRACT

The Zika virus (ZIKV) belongs to the genus Flavivirus, together with Dengue virus, yellow fever virus, and West Nile virus. The virus, which was first found in Africa in 1947, has spread across the world owing to a lack of effective drugs or vaccines. The complete genome sequence of ZIKV is now available; it includes three structural and seven non-structure genes arranged in the order of capsid, pre-membrane, envelope, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. Two geographically distinct lineages are known, i.e., Asian and African, but ZIKV exhibits differences in clinical progression among regions.

KEYWORDS

Substitution, Transition, Transversion, Variation, Zika Virus

INTRODUCTION

Zika virus (ZIKV), the mosquito-borne Flavivirus, became widely known during a pandemic in Brazil in 2015, in which approximately 1.5 million people were infected (WHO, 2016; Marinho et al., 2016; Jang et al., 2015). Since then, ZIKV transmission has been reported in 64 countries around the world (Stauff et al., 2016). The emergence of newborns with microcephaly was first reported during the outbreak in Brazil, in which 4,700 diseased infants were observed (Calvet et al., 2016). Owing to the severity of the symptoms, the WHO declared a state of emergency against Zika virus infection on February 1, 2016, emphasizing the risk of infection to the global population (WHO, 2016).

ZIKV was first isolated from a rhesus monkey in a study of yellow fever in the Zika forest of Uganda in 1947 (Dick et al., 1952). Over the next 20 years, ZIKV was isolated in East and West Africa in the absence of epidemics, and spread from Africa to Southeast Asia (Moore et al., 1975; Fagbami, 1979). In April 2007, the first pandemic was reported in Yap Island of the Federated States of Micronesia in the Western Pacific Ocean, with about 100 infected patients (Duffy et al., 2009). ZIKV was also responsible for an outbreak in French Polynesia in the South Pacific Ocean in 2013, in which approximately 30,000 people were infected (Cao-Lormeau et al., 2014). Since 2014, ZIKV

DOI: 10.4018/IJHISI.2019070103

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
spread across the Pacific Ocean to South America, Central America, and the Caribbean, and finally reached Brazil, resulting in an explosive outbreak (Fauci & Morens, 2016; Rubin et al., 2016).

The complete genome sequence of ZIKV is now available. The genome is divided into three structural genes, C, prM, and E, and seven nonstructural (NS) genes, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Kuno & Chang, 2007). These genes are arranged in the following order in the genome: 5’-capsid (C)-premembrane (prM)-envelope (E)-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3’ (Zhu et al., 2016). According to our previous study, these 10 genes exhibit differentiation among regions (Jooyeon et al., 2016). A phylogenetic analysis revealed that the genome sequence of ZIKV exhibits regions of conservation and divergence. The capsid, pre-membrane, NS2B, NS3, and NS4A genes tend to be conserved among regions owing to their important roles in the viral life cycle and replication. The envelope, NS1, NS2A, NS4B, and NS5 genes evolve more rapidly to adjust to regional hosts.

ZIKV has two distinct geographical lineages, Asian and African (Enfissi et al., 2016). However, there is not a strong association between geographical lineages and continent-specific viral properties. Even within the Asian lineage, isolates from Southeast Asia and South America have different infection propensities. For example, ZIKV infections in Brazil are characterized by a higher number of babies with microcephaly than infections in Southeast Asia. These observations indicate that sequence variation on each continent must be considered in addition to the differences between Asian and African lineages. In this study, we investigated sequence differences in 10 genes between ZIKV from six continents and examined the relationships between regional features and genetic characteristics.

**MATERIALS AND METHODS**

To investigate the mutation rate for 10 genes in the ZIKV genome, we obtained 123 complete ZIKV genome sequences from NCBI GenBank (www.ncbi.nlm.nih.gov/genbank/). These sequences were isolated from humans on six continents, i.e., Africa, Asia, Europe, Oceania, North America, and South America, from 1968 to 2016. Detailed information for these sequences, including accession numbers, isolated year and isolated continents, is shown in the supplementary document.

**Multiple Sequence Alignments**

Most of the genomes were not annotated. In order to identify the 10 genes, multiple sequence alignments of the 123 genome sequences were generated using Clustal X 2.1 (Larkin et al., 2007) with the following parameters: gap opening penalty, 15; gap extension penalty, 6.66; DNA transition weight, 0.5.

**Determine the Genetic Locus of 10 Genes in ZIKV**

Based on the reference genome, NC012532.1, which was isolated from Uganda, Africa in 2016, the genomic locations of 10 genes within each full-length sequence were determined. Specifically, the capsid (128–440 bp), prM (495–998 bp), envelope (999–2510 bp), NS1 (2511–3566 bp), NS2A (3567–4244 bp), NS2B (4245–4634 bp), NS3 (4635–6499 bp), NS4A (6500–6880 bp), NS4B (6950–7703 bp), and NS5 genes (7704–10412 bp) were determined within the 10,909-bp genome. For the computational analyses, a script written in JAVA was used to separate the complete genome sequences into individual gene loci. The polyprotein genes located between the capsid and prM genes and between the NS4A and NS4B genes were excluded. Figure 1 shows the Gene sequences of ZIKV.

**Figure 1. Gene sequences of the ZIKV**
Enhancing Knowledge Management in a Multi-Center Clinical Trial by a Web-Based Knowledge Medium
www.igi-global.com/chapter/enhancing-knowledge-management-multi-center/25827?camid=4v1a

E-Health and Digital Inclusion
www.igi-global.com/chapter/health-digital-inclusion/73837?camid=4v1a