**Higher mutation rate of zika virus will increase with bigger human population sizes**

**Coralys Maldonado Rivera**

**Evolution 461**

**Introduction**

An arthropod-borne virus known as Zika was first isolated in 1947 from a sentinel monkey in East Africa. The outbreak of Zika in 2015 created a new interest in the mutational and virulence factors. It is believed that Zika can produce microcephaly and other birth defect in the uterus due to being able to persist in the male and female reproductive system (Plourde et.al, 2016). Due to being closely related to viruses of yellow fever and Dengue that can significantly cause illness and high mortality rate, Zika should be considered as another important mosquito-borne virus (Plourde et.al, 2016).

The Zika virus is considered to have a fast mutation rate due to its lack of proofreading DNA during replication. This high mutation rate increases the chance for adaptability to different hosts and results in the creation of new human and zoonotic diseases (Borucki et.al, 2019). In the creation of new non-synonymous mutations, the change of the amino acid sequence can result highly significant. It has also been found that Zika has a high transmission rate in tropical and subtropical regions, where its vector is most likely to thrive. In the last decades the prevalence of Zika in warm tropical regions has increased for example in 2015 Zika cases increased in highly populated countries such as Brazil (Borucki et.al, 2019). The strains created by this virus are thought to be dependent of the region and the amount of population size. Although, mutations are relative to their population size it is also noted that higher populated countries have higher chances of dispersing the virus.

However, it is reported that human population size can influence the rate of mutation where higher size population will increase mutation rate and smaller population will decrease the mutational rate. To further investigate this claim, this paper will aim to compare the mutation rate of the zika virus from a single population size through different years using a Pearson Correlation Test, X-Y Scatter plot and a Linear regression Analysis. The hypothesis states that higher mutation rates of Zika virus will increase with bigger human population sizes.

**Materials and Methods**

*Data and calculations*

To test the hypothesis, data was obtained from the “The World Bank organization” where the population size of rural populations of Brazil contained the number of human population over the years (WorldBank, 2020). On the other the second part of the data was obtained from the paper “Multiscale analysis for patterns of Zika virus genotype emergence, spread and consequence” where a link to GenBank and ViPR databases contained the epidemic-association mutations between time, location, and the number of cases between mutations (Borucki et.al, 2019). The timeline in which each mutation was recorded was between 2006 to 2017. Each of these individual viruses were compare against the entirety of their genome, the non-synonymous mutations, the country, codon, nucleotide length and strain. For this experiment the country, dates and non-synonymous mutations were kept to further analyze.

Human population size measurement was made analyzing the differences between years and the fluctuation of population given in the data. These fluctuations are due to survival rate and environmental factors (Borucki et.al, 2019). The mutational rate was then analyzed by the different non-synonymous mutations throughout the years. These numbers were then divided by the total amount of mutations per population size to obtain the highest amount of population size relative to the number of mutations. This data was then utilized to form statistical tests.

.

*X-Y Scatter Plot and Linear Regression Analysis*

An X-Y scatter plot was generated for the Brazil rural population sample and the mutational rate of Zika virus in RStudio statistical programming (RStudio, 2015). The x-axis was the amount of Brazil human population as the percentage of the relative years of 2006 to 2017, and the y-axis was the mutation rate of Zika virus. Taken from this data, regression line analysis was conducted to observe the relationship between population size and mutation rate.

*Pearson Correlation Test*

A Pearson coefficient test was calculated between the Brazilian population size and the mutation rate of the Zika virus. The test was conducted in RStudio utilizing the code: cor.test(), to produce a 95% confidence interval (RStudio, 2015). This analysis was then utilized to create a correlation plot with the same program. The correlation coefficient result in > 0.08 and was considered to express no correlation between the variables.

**Results**

**Chart, scatter chart

Description automatically generated**

*Figure 1: X-Y Scatter Plot of Mutation rates of Zika virus compared to the Rural population sizes of Brazil. Each dot represents a population of mutations. The y-intercept of the best-fit line was recorded to be 1.883e+01 with a slope of 1.941e-07.*

**Pearson Correlation Test**

The Pearson Correlation Test showed no positive correlation between Mutation rates of Zika virus and the Rural populations sizes of Brazil with a 95% confidence interval. The calculated correlation value was 0.08 resulting in no correlation between the two variables.

Results from the statistical tests displayed no relationship between the population size and the mutation rates of Zika virus. From the X-Y Scatter plot, a slope of 1.941e-07 was calculated for the best-fit line. The correlation tests showed a negative or no correlation between the value giving 0.08.

**Discussion**

This study suggests that higher mutation rate of the Zika virus will increase due to bigger population sizes. This data does not support the original hypothesis that despite that higher mutation rates are a result of bigger populations sizes, the Zika virus mutation rate and the rural populations of Brazil did not result in a correlation. These results may lead us to infer that some statistical errors may have occur or that not sufficient data was found, further, analysis and data recollection should be made for the following study to proceed.

**References**

Borucki, M. K., Collette, N. M., Coffey, L. L., Van Rompay, K. K., Hwang, M. H., Thissen, J. B., Allen, J. E., & Zemla, A. T. (2019). Multiscale analysis for patterns of zika virus genotype emergence, spread, and consequence. *PLOS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0225699

Plourde, A. R., & Bloch, E. M. (2016). A literature review of zika virus. *Emerging Infectious Diseases*, *22*(7), 1185–1192. https://doi.org/10.3201/eid2207.151990

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

*Rural population - Brazil*. (2020). World Bank Organization . (n.d) https://data.worldbank.org/indicator/SP.RUR.TOTL?locations=BR