**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 defects 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 infection rate of this virus is mostly agglomerated in the warm regions of America where the mosquito can thrive and infect many individuals (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. The rate of mutation can be of importance due to increasing the chances of new variances in the populations creating higher effect in the next generations. As an example, a strain in Brazil of the Zika virus has been observed to cause birth defects which can affect the new generation coming in (Cugola et.al, 2016).

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 purpose of the study is to understand if human and viral evolution is correlated with the number of mutations in a single population. With this stated, the study hypothesizes that the high mutation rate of Zika virus will increase with bigger human rural 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 hand, 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 nonsynonymous mutations that were according to the protein structure and the 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 utilizing nonsynonymous mutations. 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 must be around -1 to 1, which will discard 0 to be considered a positive correlation between variables.

**Results**

Graphical user interface

Description automatically generated with medium confidence

*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 -1.599e+01 with a slope of 5.980e-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 0.176 resulting in a slight correlation between the two variables.

Results from the statistical tests displayed that a relationship between the population size and the mutation rates of Zika virus was below the line of regression. From the X-Y Scatter plot, a slope of 5.980 e-07 was calculated for the best-fit line. Figure 1, shows some outliers which are mutation rates whose frequency was either too low or too high to be considered in the slope. The correlation tests showed a slight or no correlation between the variables due to the correlation test resulting in 0.176.

**Discussion**

This study suggests that higher mutation rate of the Zika virus will increase due to bigger population sizes. The data presented the rural populations of brazil and the nonsynonymous mutations of a mosquito borne virus. It was inferred that mutational rate had a positive relation with population size , however, the results expressed a very low or no correlation between the variables. This possibility may be that the data does not support the entirety of the hypothesis and that further analysis in data frames should be made. However, it cannot be completely discarded due to inferring that the nonsynonymous mutations of the Zika virus and the rural population of Brazil may be subjected to other environmental factors such as the migration and drift of the different alleles.

On the other hand, further analyses may be made with this type of data. Utilizing a Spearman Correlation test rather than the Pearson Correlation may indicate that their correlation was not linear and rather with valleys and fluctuations through time. It can hypothesize that the mutational rate of Zika virus is dependent of the environmental factors of these rural populations. With this expected, we can infer that rural populations in South America may have genetic factors such as different mutations that can produce a type of antibody against the different strains of Zika virus. This new hypothesis may conclude that mutations are also subjected to genetic diversification and that these mutations are related to the difference in these alleles rather than the populations itself.

To conclude, this study was created to understand how populations can increase or decrease the mutational rate of a mosquito borne virus. These results may be inconclusive, but it gives the evolutionary family a start to understanding why mutations may be dependent of genetic diversity, migration, and strength of mutation. Further analysis, research and tests should be made to create a strong correlation between the variables presented.

**Acknowledgements:**

This study will not be possible without the help of Dr.Mitchell and the course Biology 461: Principle of Evolution. A big thank you and appreciation for his input and help throughout this study should be emphasized and praised.

**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>

Cugola, F., Fernandes, I., Russo, F. *et al.* The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* **534,**267–271 (2016). <https://doi.org/10.1038/nature18296>

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