Journal Pre-proof

A recursive sub-typing screening surveillance system detects the arising of the ZIKV African lineage in Brazil: Is there risk of a new epidemic?

José I. Kasprzykowski, Kiyoshi F. Fukutani, Helton Fabio, Eduardo R. Fukutani, Larissa C. Costa, Bruno B. Andrade, Artur T.L. Queiroz

Please cite this article as: { doi: https://doi.org/

A recursive sub-typing screening surveillance system detects the arising of the ZIKV African lineage in Brazil: Is there risk of a new epidemic?

José I. Kasprzykowski[1]*, Kiyoshi F. Fukutani[1,2]*, Helton Fabio[1], Eduardo R. Fukutani[1], Larissa C. Costa[1], Bruno B. Andrade[1,2,3,4,5] and Artur T. L. Queiroz[1].

[1] Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil.

[2] Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador, Brazil.

[3] Faculdade de Tecnologia e Ciências (FTC), Salvador, Brazil.

[4] Universidade Salvador (UNIFACS), Laureate Universities, Salvador, Brazil.

[5] Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador, Brazil.

*These authors contributed equally to this article.

Running Head: A surveillance system to detect ZIKV epidemics

Address for correspondence: Bruno B. Andrade, MD, PhD. Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Rua Waldemar Falcão, 121, Candeal, Salvador, Bahia 40296-710, Brazil.

Abstract word count: 51
Main MS text word count: 725
Figures: 01
Tables: 01
Supplementary Figures: 00
Supplementary Tables: 0o

1

Highlights:

-    The Brazilian sequence ZIKV lineage proportion changed across the years

-    The automated surveillance system helps the ZIKV molecular epidemiology

-    A new lineage emerged in Brazil and could be a potential risk of a new epidemic

Abstract

There is currently no system to track the emergence of zika virus (ZIKV) subtypes. We developed a surveillance system able to retrieve sequence submissions and further classify distinct ZIKV genotypes in the world. This approach was able to detect a new occurrence of ZIKV from African lineage in Brazil in 2019.

Keywords: genomic surveillance, virus subtype analysis, database, lineage molecular epidemiology

The zika fever is a self-limiting disease caused by the Zika virus (ZIKV) infection, which is transmitted by *Aedes aegypti* (Weaver et al. 2016). As the first reported epidemics happened in African and Asian continents, ZIKV is classically classified in three lineages: East African, West African, and Asian. Moreover, the Asian lineage is further classified by geographical origin as Micronesian, Cambodian and Malaysian (2012). The strains isolated during the outbreak in Brazil were classified as Asian lineage (Faria et al. 2016). The spread of ZIKV in Brazil started in the Northeast region, further spreading to states of the Central-West and Southeast regions (Campos et al. 2015; Zanluca et al. 2015) and reaching the United States (Bond 2016). Several reports have focused on the description of when and how ZIKV emerged and disseminated in Brazil and South America. Despite some subtyping tools accessible (Fonseca et al. 2019), a system able to perform an automated surveillance of ZIKV subtypes remains unavailable and would be crucial to early detect new potential epidemics. Herein, we developed and implemented an automated system for sequence mining and subtyping, which was first developed to study HIV (Irahe Kasprzykowski et al. 2017), and now adapted to investigate ZIKV, by continuously retrieving data from the NCBI databases. The whole ZIKV-related NCBI nucleotide database was retrieved, comprising all submissions (1748 sequences). All information from the sequences were cross-referenced with other NCBI databases (e.g. Biosample, Pubmed, SRA) and stored in a metadata. Furthermore, all sequences were subtyped using an optimized non-heuristic algorithm based on Smith-Waterman's (Smith and Waterman 1981) against the ZIKV reference genome (Table 1). The alignment of each query sequence with the reference determined the subtype, and it was stored in a MySQL database. All Brazilian submissions [filtered by region of origin (from metadata) and sequence length <200bp] detected in up to December 2019 (248 sequences) were extracted (Figure 1A). The system was developed using JAVA as a main language to retrieve and analyzed sequences, whereas PHP was used to build a frontend and middleware integration. In addition, the MySQL was used as a database management system. The system and data are available at http://saga.bahia.fiocruz.br/saga/.

3

Interestingly, the real-time surveillance detected changes across the 248 Brazilian sequences which were submitted since 2015. In this year, 42.8% of the sequences were characterized as Cambodian and 57.2% as Micronesian. The proportion of sequence submissions with the Cambodia subtype was extremely high (>90%) during 2016, 2017 and 2018. In 2019, there was an inversion on this profile, with submissions from the Micronesian subtype representing 89.2% (Figure 1B). Furthermore, our approach also identified the emergence of 5.4% of all sequences from 2019 which were classified as African lineage (in January and in May). The difference in the distribution of these proportions over time was statistically significant, revealing a substantial change in ZIKV sequences submissions (Figure 1B). These African lineages were isolated from two different regions of Brazil: South, from the Rio Grande do Sul (Almeida et al.) and Southeast, from Rio de Janeiro (Alencar et al.), from different sources: one was isolated from *Aedes albopictus* mosquito species (Alencar et al.) and the second from *Alouatta guariba*, a monkey species (Almeida et al.). The geographic distance between these submissions was >1,500 kilometers and are from two different regions of Brazil. The distinct localities and different sources of isolation are concerning and suggest that this African lineage could be already circulating in Brazil for a significant period of time. Of note, the phylogenetic-based typing tool confirmed the classification.

The real-time surveillance system reported here, which screens all the available ZIKV sequences submitted to online datasets, is novel but has some limitations. (i) This is a database analysis, not an epidemiologic inquiry and under a collection bias. (ii) The algorithm is limited to the sequence information provided and is impractical to assess its quality. Regardless of such limitations, our innovative system was able to detect in real time the emergence of a new ZIKV lineage in two different locations in Brazil. Among the sequences retrieved by our system, two were novel and in preprint publications and reported local data without analyzing the whole sequence universe in the country. By employing a surveillance system able to retrieve sequence

4

submissions and to further classify the distinct ZIKV genotypes in the world we provide a novel tool to aid monitoring and control of ZIKV epidemics.

Ethical Approval: This study used de-identified publicly available sequence data and did not required approval by Institutional Review Board.

Potential conflicts of interest: The authors declare that they have no conflicts of interest.

Contributions: JIK, KFF, BBA and ATQL conceived and designed the study. JIK, HF and ATQL developed the system and database. JIK, KFF, BBA and ATQL performed the data visualization. KFF, DRK and LCC analyzed the data. JIK, KFF, BBA and ATQL wrote the manuscript.

Data availability statement: The datasets generated during and/or analyzed during the current study are available in public databanks and the accession numbers are listed within the manuscript.

Declaration of interests
All the Authors agree and declare that they have not know competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alencar J, de Mello CF, Marcondes CB, Guimarães AÉ, Toma HK, Bastos AQ, et al. Natural infection and vertical transmission of two flaviviruses (Yellow fever and Zika) in mosquitoes in primary forests in the Brazilian state of Rio de Janeiro (Diptera: Culicidae) [Internet]. Available from: http://dx.doi.org/10.1101/688713

Almeida PR de, de Almeida PR, Ehlers LP, Demoliner M, Eisen AKA, Girardi V, et al. Detection of a novel African-lineage-like Zika virus naturally infecting free-living neotropical primates in Southern Brazil [Internet]. Available from: http://dx.doi.org/10.1101/828871

Bond S. CENTERS FOR DISEASE CONTROL AND PREVENTION ISSUES INTERIM GUIDELINES FOR PREGNANT WOMEN AND INFANTS EXPOSED TO ZIKA VIRUS AND PREVENTION OF SEXUALLY TRANSMITTED ZIKA INFECTION [Internet]. Vol. 61, Journal of Midwifery & Women's Health. 2016. p. 392–4. Available from: http://dx.doi.org/10.1111/jmwh.12483_3

Campos GS, Bandeira AC, Sardi SI. Zika Virus Outbreak, Bahia, Brazil [Internet]. Vol. 21, Emerging Infectious Diseases. 2015. p. 1885–6. Available from: http://dx.doi.org/10.3201/eid2110.150847

Faria NR, Azevedo R do S da S, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science [Internet]. 2016 Apr 15;352(6283):345–9. Available from: http://dx.doi.org/10.1126/science.aaf5036

Fonseca V, Libin PJK, Theys K, Faria NR, Nunes MRT, Restovic MI, et al. A computational method for the identification of Dengue, Zika and Chikungunya virus species and genotypes. PLoS Negl Trop Dis [Internet]. 2019 May;13(5):e0007231. Available from: http://dx.doi.org/10.1371/journal.pntd.0007231

Irahe Kasprzykowski J, Ferreira Fukutani K, F'abio H, Maria Prado Barral A, Trancoso Lopo de Queiroz A. HIV-1 Nucleotide Sequence Comprehensive Analysis: A Computational Approach. Curr Bioinform [Internet]. 2017;12(4):303–11. Available from: https://www.ingentaconnect.com/content/ben/cbio/2017/00000012/00000004/art00005

Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol [Internet]. 1981 Mar 25;147(1):195–7. Available from: http://dx.doi.org/10.1016/0022-2836(81)90087-5

Weaver SC, Costa F, Garcia-Blanco MA, Ko AI, Ribeiro GS, Saade G, et al. Zika virus: History, emergence, biology, and prospects for control. Antiviral Res [Internet]. 2016 Jun;130:69–80. Available from: http://dx.doi.org/10.1016/j.antiviral.2016.03.010

Zanluca C, Melo VCA de, Mosimann ALP, Santos GIVD, Santos CNDD, Luz K. First report of autochthonous transmission of Zika virus in Brazil. Mem Inst Oswaldo Cruz [Internet]. 2015 Jun;110(4):569–72. Available from: http://dx.doi.org/10.1590/0074-02760150192

Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage. PLoS Negl Trop Dis [Internet].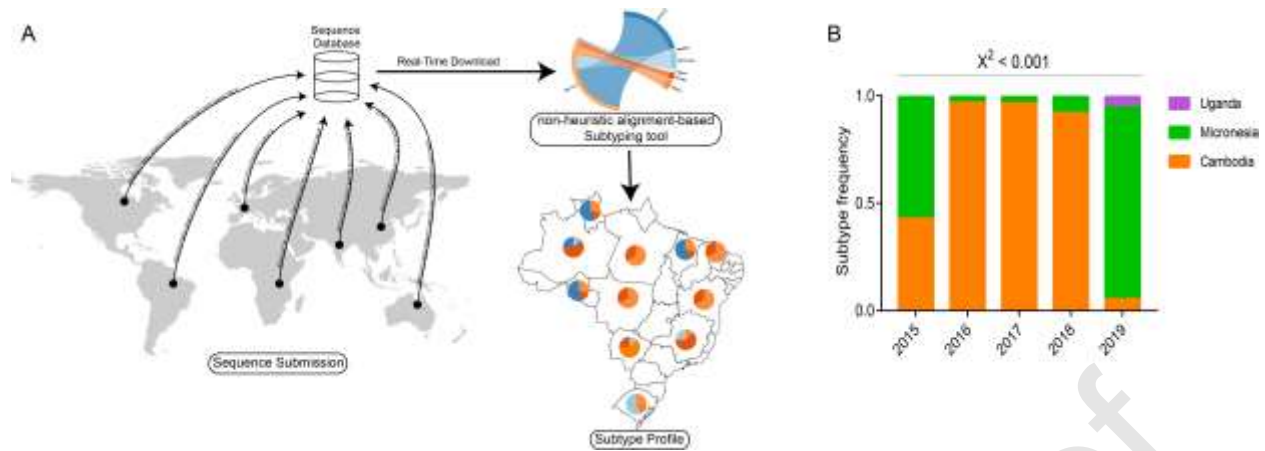 2012 Feb 28 [cited 2019 Dec 30];6(2):e1477. Available from: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0001477

Figure 1. A recursive subtyping-screening surveillance system to track ZIKV sequences.

(A) Illustration of the pipeline developed to perform data acquisition, subtyping processing and

calculation of subtype proportions of ZIKV in Brazil. **(B)** The profile of ZIKV subtypes between

2015 and 2019. Comparison of the ZIKV subtype proportions was performed using the

Pearson's chi-square test. P-value is shown.

Table 1. Reference set used for ZIKV subtyping process: accession number, lineage and strain information.

Accession Number	Lineage	Strain
EU545988	ASIAN	Micronesia
JN860885		Cambodia
HQ234499		Malaysia
HQ234501	AFRICAN	Senegal
HQ234500		Nigeria
HQ234498		Uganda

8