DISCLAIMER
This paper was submitted to the *Memorias do Instituto Oswaldo Cruz* on 25 July 2017 and was posted to the Fast Track site on 26 July 2017. The information herein is available for unrestricted use, distribution and reproduction provided that the original work is properly cited as indicated by the Creative Commons Attribution licence (CC BY).

RECOMMENDED CITATION
Delatorre E, Mir D, Bello G. Tracing the origin of the NS1 A188V substitution responsible for recent enhancement of Zika virus Asian genotype infectivity [Submitted]. Mem Inst Oswaldo Cruz E-pub: 26 July 2017. doi: 10.1590/0074-02760170299.

LETTER TO EDITOR

Running title: Origin of NS1 A188V substitution in ZIKV

Title: Tracing the origin of the NS1 A188V substitution responsible for recent enhancement of Zika virus Asian genotype infectivity

Authors: Edson Delatorre[§1], Daiana Mir[§1] and Gonzalo Bello[§*1]

Institutional affiliation:
1. Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil.

§All authors contributed equally

*Corresponding author: Lab. de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz - FIOCRUZ. Av. Brasil 4365, Manguinhos, 21040-360, Rio de Janeiro, RJ, Brazil. Phone: +55 21 3865 8154. E-mail: gbellobr@gmail.com/gbello@ioc.fiocruz.br

ABSTRACT

A recent study showed that infectivity of Zika virus (ZIKV) Asian genotype was enhanced by an alanine-to-valine amino acid substitution at residue 188 of the NS1 protein, but the precise time and location of origin of this mutation were not formally estimated. Here, we applied a Bayesian coalescent-based framework to estimate the age and location of the ancestral viral strain carrying the A188V substitution. Our results showed that the ancestral ZIKV strain carrying the A188V substitution probably arose in Southeastern Asia and circulated in that region for some time (5-10 years) before being disseminated to Southern Pacific islands and the Americas.

Keywords: ZIKV, NS1, infectivity, evolution.

Financial support: FIOCRUZ

Zika virus (ZIKV) is a mosquito-borne pathogen member of the family *Flaviviridae*, genus *Flavivirus*, that was first isolated from a sentinel monkey in Uganda in 1947 (Dick et al. 1952). Until recently, ZIKV was most likely maintained in a sylvatic cycle involving vectors of the genus *Aedes* and non-human African primates (Hayes 2009). Since 2007, however, large epidemics of ZIKV were described in human populations from the Pacific islands and more recently in the Americas (Gatherer & Kohl 2015). Liu *et al.* (2017) reported that the infectivity of ZIKV in *Aedes aegypti* mosquitoes has evolved over time, offering a potential explanation for the recent successful spread of the virus to the Southern Pacific islands and the Americas. The authors showed that a ZIKV isolate representative of the 2015-2016 American epidemics was much more infectious in mosquitoes than a ZIKV strain representative of the 2007–2012 Southeastern Asia epidemics. Further analyses demonstrate that ZIKV infectivity was enhanced by an alanine(A)-to-valine(V) amino acid substitution at residue 188 of the NS1 protein, resulting in increased NS1 antigenaemia in infected hosts that in turn promotes ZIKV infectivity and prevalence in mosquitoes.

The authors suggest that the Asian lineage of ZIKV acquired enhanced infectivity when it spread from the Southeastern Asia to the Southern Pacific around 2013, because residue 188 of the NS1 protein was alanine in ZIKV isolates from the Asian clade collected before 2012, but was mutated to valine in all isolates collected after 2013. This hypothesis, however, was not formally tested using a model-based statistical framework. Here, we performed a Bayesian evolutionary and phylogeographic analysis to reconstruct the spatiotemporal dissemination dynamics of the ZIKV Asian genotype and to properly estimate the age and location of the ancestral viral strain carrying the A188V substitution.

All near-complete ZIKV genome sequences from Asian and Southern Pacific countries with known date of isolation and from American countries sampled during 2014 and 2015 were retrieved from GenBank on March 15th 2017. Only those sequences of imported cases with information about country of infection were included. This resulted in a final data set of 45 ZIKV Asian genotype complete genomes spanning a 50-year period. Complete coding sequences (CDS) were manually aligned and subjected to Maximum Likelihood (ML) phylogenetic reconstruction with PhyML v3.0 (Guindon et al. 2010), under the GTR+Γ_4 nucleotide substitution model selected by jModelTest v1.6 (Posada 2008). Temporal signal of the dataset was verified using Tempest (Rambaut et al. 2016). The spatiotemporal viral diffusion pattern and the ancestral CDS at key internal nodes of the phylogeny were reconstructed using the Markov chain Monte Carlo (MCMC) algorithms implemented in the BEAST v1.8 package (Drummond & Rambaut 2007). The temporal scale was estimated using a relaxed uncorrelated lognormal molecular clock model (Drummond et al. 2006), the GTR+Γ_4 nucleotide substitution model and a flexible Bayesian Skyline coalescent model (Drummond et al. 2005). A discrete state corresponding to the region of isolation (Southeastern Asia, Pacific or America) was assigned for each ZIKV sequence and migration events throughout the phylogeny were reconstructed using both reversible (symmetric) and nonreversible (asymmetric) discrete phylogeographic models (Lemey et al. 2006). MCMC were run sufficiently long (20-100 million MCMC steps) to ensure stationary and convergence of all parameters (Effective Sample Size > 200), through inspection with Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). The Maximum Clade Credibility (MCC) trees were generated with TreeAnnotator v1.8[8] after discarding the 10% burn-in and visualized with FigTree v1.4

(http://tree.bio.ed.ac.uk/software/figtree/). Consensus CDS at key ancestral nodes were computed using the R package SeqinR (http://seqinr.r-forge.r-project.org/).

Analysis of the complete coding regions of 45 ZIKV genome sequences from Southeastern Asia (n = 13, 1966-2016), Southern Pacific (n = 16, 2013-2016) and the Americas (n = 16, 2014-2015) reveals a very strong correlation (R^2 = 0.99) between genetic divergence and sampling time within the ZIKV Asian lineage (Fig. 1A), thus supporting the use of this dataset for molecular clock calibration. The evolutionary rate and the overall time-scale of the Bayesian phylogenetic tree here estimated for ZIKV Asian lineage were fully consistent with those previously reported (Faria et al. 2016) (Table 1). Phylogeographic analyses using both asymmetric (Fig. 1B) and symmetric (data not shown) diffusion models placed the most recent common ancestor of ZIKV Asian genotype epidemic strains in Southeastern Asia (posterior state probability, *PSP* = 1) at around 1999 [95% Bayesian credible interval (BCI): 1995-2003]. From Southeastern Asia, the ZIKV Asian genotype was disseminated to Pacific islands on two independent occasions. The first dissemination originated a large epidemic of ZIKV in Micronesia (Western Pacific) in 2007 (Duffy et al. 2009), but was not further spread to other countries. The second dissemination was dated to 2012 (BCI: 2012-2013) and fuelled recent ZIKV epidemics (2013-2016) in several Southern Pacific islands (Cao-Lormeau et al. 2014, Musso et al. 2014). Our phylogeographic analysis supports a single entry of ZIKV from Southern Pacific (*PSP* = 1) into the Americas at 2013 (BCI: 2012-2013). Ancestral sequence reconstruction at internal nodes of the inferred ZIKV Asian genotype phylogeny place the emergence of the NS1 A188V substitution in Southeastern Asia (*PSP* = 1) at some time between 2003 (BCI: 2000-2006) and 2007 (BCI: 2004-2009) (Fig. 1B).

In summary, we showed that the NS1 A188V substitution associated with enhanced infectivity of ZIKV Asian lineage in *Aedes aegypti* mosquitoes probably arose during viral dissemination among human populations in the Southeastern Asian region, between the early and the middle 2000s. Thus, ZIKV Asian genotype strains carrying the NS1 A188V mutation appear to have spread in the Southeastern Asian region for some time (5-10 years) before being disseminated to Southern Pacific islands and the Americas. The absence of the reversal NS1 V188A mutation at either internal nodes or terminal tips in the ZIKV Asian genotype phylogeny clearly supports some selective advantage for the fixation of the valine amino acid at residue 188 in NS1.

Acknowledgements

E.D. is funded by a fellowship from "Programa Nacional de Pós-Doutorado (CAPES-Brazil)". PNPD/CAPES Postdoctoral Fellowship. D.M. is funded by fellowships from "Agencia Nacional de Investigación e Innovación (ANII-Uruguay)" and "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Brazil)".

Author Contributions

E.D., D.M. and G.B conceived and designed the study, participated in data collection, performed sequence data analyses, and drafted the manuscript. All authors reviewed and approved the final manuscript.

References

Cao-Lormeau, VM, Roche C, Teissier A, Robin E, Berry A, Mallet H, et al. Zika virus, French polynesia, South pacific, 2013. Emerg Infect Dis. 2014; 20: 1085-1086.

Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. Trans R Soc Trop Med Hyg. 1952; 6: 509-520.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006; 4: e88.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005; 22: 1185-1192.

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007; 7: 214.

Duffy MR, Chen T, Hancock T, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 2009; 360: 2536-2543.

Faria NR, Azevedo RSS, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science. 2016; 352: 345-349.

Gatherer D, Kohl A. Zika virus: a previously slow pandemic spreads rapidly through the Americas. J Gen Virol. 2015; 97(2): 269-73.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59: 307-321.

Hayes EB. Zika virus outside Africa. Emerg Infect Dis. 2009; 15: 1347-1350.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2006; 5: e1000520.

Liu Y, Liu J, Du S, Shan C, Nie K, Zhang R, et al. Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. Nature. 2017; 545: 482–486.

Musso D, Nilles EJ, Cao-Lormeau VM. Rapid spread of emerging Zika virus in the Pacific area. Clin Microbiol Infect. 2014; 20: O595-596.

Posada, D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008; 25: 1253-1256.

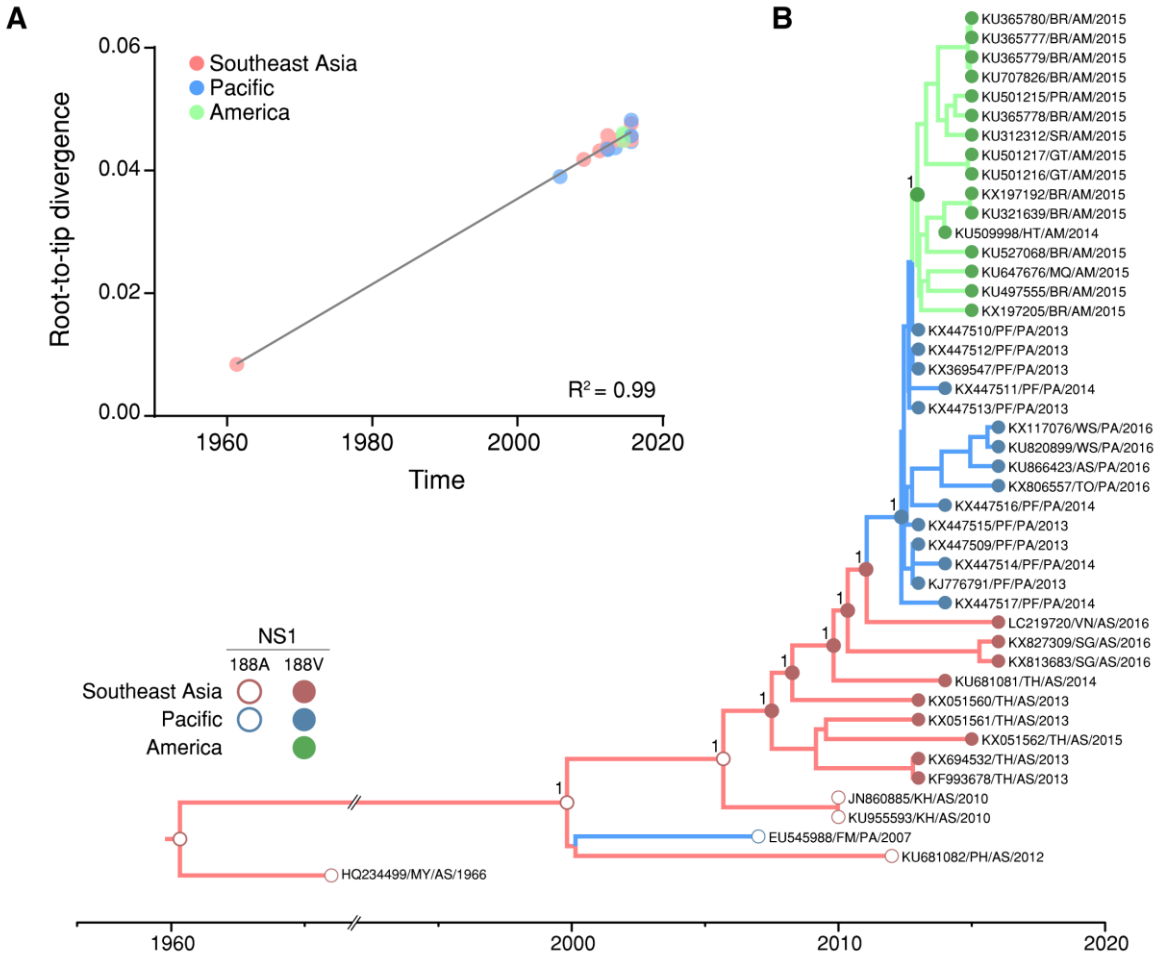
Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016; 2(1): vew007.

Table 1. Posterior estimates of evolutionary parameters of ZIKV Asian genotype.

Study	Phylogeographic model	Evolutionary rate (95% BCI)	TMRCA (95% BCI)		
			Epidemic Asian clade	Southern Pacific clade	American clade
Faria *et al.* 2016	-	1.1×10^{-3} (8.5×10^{-4} - 1.2×10^{-3})	-	-	2014.0 (2013.6-2014.3)
This study	Reversible	9.2×10^{-4} (7.1×10^{-4} - 1.2×10^{-3})	1999.2 (1994.4-2002.8)	2012.4 (2011.9-2012.7)	2013.1 (2012.1–2013.4)
This study	Nonreversible	9.2×10^{-4} (6.9×10^{-4} – 1.2×10^{-3})	1999.2 (1994.6-2002.8)	2012.4 (2011.9-2012.7)	2013.1 (2012.7–2013.4)

FIGURE LEGENDS

Figure 1. Emergence of the A188V substitution at NS1 protein during ZIKV Asian genotype evolution. (A) Correlation between the sampling date of each ZIKV sequence (*n* = 45) and the genetic distance of that sequence from the root of a ML phylogenetic tree. Colors indicate the geographic region of sampling. (B) Bayesian time-scale Maximum Clade Credibility phylogenetic tree estimated from ZIKV Asian genotype genomic sequences (*n* = 45). Branches are colored according to the most probable location state (geographic region) of their descendent nodes as indicated at the legend on the lower left. Reconstructed ancestral key nodes and terminal nodes are highlighted with circles colored according to the location (color) and the amino acid at NS1 188 residue (fill or empty), as indicated at the legend on the lower left. Numbers at key selected nodes represent the posterior probability supports of the clades. All horizontal branch lengths are drawn to a scale of years. Taxon labels include information of GenBank accession number, country of origin, region of origin and year of isolation. Countries represented are: American Samoa (AS), Brazil (BR), Cambodia (KH), Federated States of Micronesia (FM), French Polynesia (PF), Guatemala (GT), Haiti (HT), Malaysia (MY), Martinique (MQ), Philippines (PH), Puerto Rico (PR), Samoa (WS), Singapore (SG), Suriname (SR), Thailand (TH), Tonga (TO), and Vietnam (VN).


A
B
Root-to-tip divergence
0.06
0.04
0.02
0.00
Southeast Asia
Pacific
America
R² = 0.99
1960
1980
2000
2020
Time
NS1
188A 188V
Southeast Asia
Pacific
America
KU365780/BR/AM/2015
KU365777/BR/AM/2015
KU365779/BR/AM/2015
KU707826/BR/AM/2015
KU501215/PR/AM/2015
KU365778/BR/AM/2015
KU312312/SR/AM/2015
KU501217/GT/AM/2015
KU501216/GT/AM/2015
KX197192/BR/AM/2015
KU321639/BR/AM/2015
KU509998/HT/AM/2014
KU527068/BR/AM/2015
KU647676/MQ/AM/2015
KU497555/BR/AM/2015
KX197205/BR/AM/2015
KX447510/PF/PA/2013
KX447512/PF/PA/2013
KX369547/PF/PA/2013
KX447511/PF/PA/2014
KX447513/PF/PA/2013
KX117076/WS/PA/2016
KU820899/WS/PA/2016
KU866423/AS/PA/2016
KX806557/TO/PA/2016
KX447516/PF/PA/2014
KX447515/PF/PA/2013
KX447509/PF/PA/2013
KX447514/PF/PA/2014
KJ776791/PF/PA/2013
KX447517/PF/PA/2014
LC219720/VN/AS/2016
KX827309/SG/AS/2016
KX813683/SG/AS/2016
KU681081/TH/AS/2014
KX051560/TH/AS/2013
KX051561/TH/AS/2013
KX051562/TH/AS/2015
KX694532/TH/AS/2013
KF993678/TH/AS/2013
JN860885/KH/AS/2010
KU955593/KH/AS/2010
EU545988/FM/PA/2007
KU681082/PH/AS/2012
HQ234499/MY/AS/1966
1960
2000
2010
2020