Genomic Insights into Zika Virus Emergence and Spread

Nathan D. Grubaugh,[1] Nuno R. Faria,[2] Kristian G. Andersen,[1,*] and Oliver G. Pybus[2,*]
[1]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA
[2]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
*Correspondence: andersen@scripps.edu (K.G.A.), oliver.pybus@zoo.ox.ac.uk (O.G.P.)
https://doi.org/10.1016/j.cell.2018.02.027

The emergence and spread of Zika virus in the Americas continues to challenge our disease surveillance systems. Virus genome sequencing during the epidemic uncovered the timescale of Zika virus transmission and spread. Yet, we are only beginning to explore how genomics can enhance our responses to emerging viruses.

Overview

The emergence of Zika virus in the Americas took the world by surprise. Although the spread of mosquito-borne pathogens to new territories is a recognized threat, no one was prepared for a public health emergency from this little-known flavivirus. Like dengue and chikungunya viruses, Zika is transmitted by *Aedes aegypti* mosquitoes, which thrive in growing tropical cities.

The first confirmed report of Zika virus infection in the Americas occurred in northeastern Brazil in May 2015 (Kindhauser et al., 2016). Its detection was initially met with limited interest, but concerns grew quickly once data indicated a link between Zika virus infection and the development of microcephaly in newborns (e.g., Souza et al., 2016). Once in the Americas, the virus encountered a hemisphere full of immunologically naive human hosts and competent mosquito vectors. Infections with the virus and resultant cases of severe microcephaly subsequently soared in Brazil, and detection in other countries quickly followed (Figure 1). Researchers raced to answer fundamental questions, including (1) when and from where did the epidemic originate, (2) how was the virus spreading, and (3) were severe clinical conditions, such as microcephaly, a new occurrence?

In this commentary, we summarize what we have learned about the spread and evolution of the Zika epidemic. As with the 2013–16 Ebola epidemic in West Africa (Dudas et al., 2017), virus genomic data were used to answer critical epidemiological questions and to complement traditional disease surveillance efforts. The Zika epidemic also posed a distinct set of technical challenges and provided new insights into how an integrated, real-time response system to emerging arboviruses might be achieved.

Delayed Detection and Spread of Zika Virus

Genomic characterization of the Zika epidemic revealed vulnerabilities in current arbovirus disease surveillance systems. Analyses of Zika virus genomes demonstrated that the epidemic in the Americas was likely established in late 2013 in northeastern Brazil, more than a year before the first confirmed case (Figure 1A; Faria et al., 2016, 2017). This surveillance gap, between the arrival and discovery of the virus (Figure 1B) exists, in part, because the symptoms of Zika virus infection are often mild or similar to those of other co-circulating viruses, including dengue and chikungunya (Figure 1C). Using genomics to quantify the true arrival date of Zika virus therefore helps us understand what fraction of microcephaly cases might be attributable to the virus and why microcephaly rates vary among locations (e.g., de Oliveira et al., 2017). The retrospective discovery of this long surveillance gap means that measures to control international spread of the virus were already too late by the time Zika virus was first detected in the Americas (Metsky et al., 2017).

In addition to the timescale of international spread, Zika virus genomic studies also estimated the frequency and sources of virus introductions into specific locations. While the epidemic in the Americas can be traced back to a single "successful" introduction of Zika virus into Brazil, local outbreaks in other countries were likely composed of transmission chains derived from multiple independent introductions (Faria et al., 2017; Metsky et al., 2017). For example, phylogenetic analyses estimated that more than 30 separate introductions of Zika virus to Florida were responsible for the comparatively small outbreak there of ~250 reported cases (Grubaugh et al., 2017). If repeated introductions in a given region are the norm, then responsive local control efforts by themselves may not be able to eliminate infection risk. Rather, the intensity of modern human mobility means that emerging epidemics like Zika require international coordination and an increased focus on surveillance and control efforts in source countries.

A significant challenge for Zika virus diagnostics, and subsequent virus genome sequencing, was the low amounts of virus present in most clinical samples. This problem is caused by the rapid decrease in viremia following the onset of symptoms (Lessler et al., 2016), when most clinical samples were collected. The need for virus genomic data to understand the epidemic, however, led to the development of several sequencing methods that specifically enriched for Zika virus RNA (e.g., Quick et al., 2017). The open sharing of laboratory protocols facilitated the generation of most of the available Zika virus genetic data and resulting insights discussed here.

Where Do We Go from Here?

Despite an estimated 100 million Zika cases in the Americas, less than 300

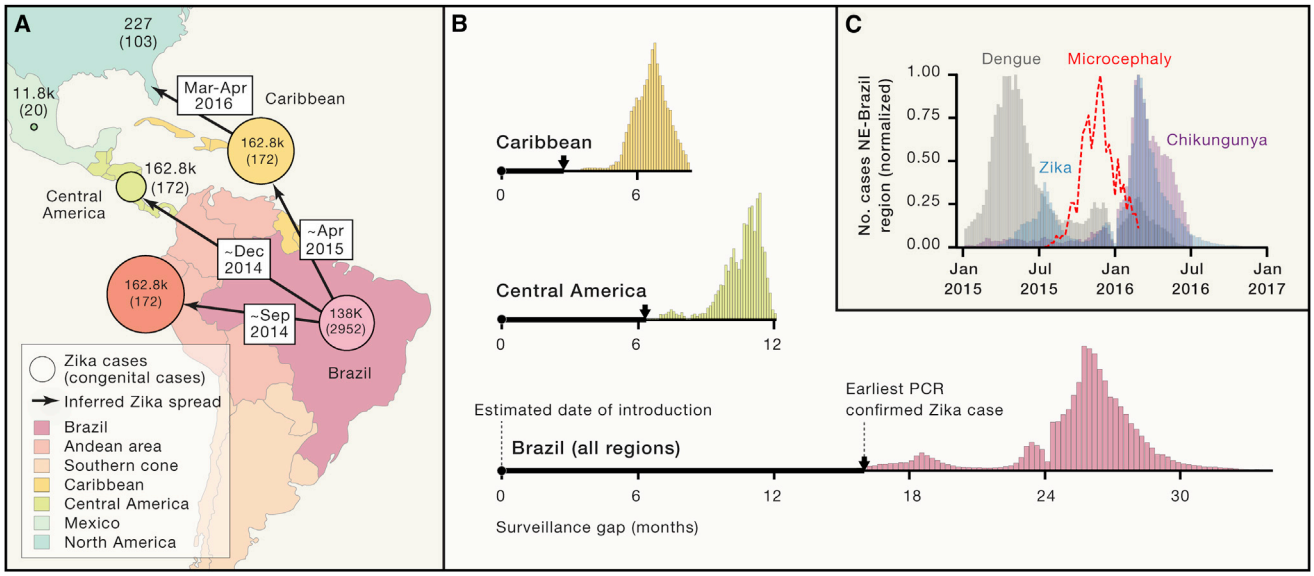

Figure 1. Hidden and Observed Spread of Zika Virus in the Americas

(A) The map illustrates pathways of Zika virus spread from Brazil to other regions in the Americas, as inferred by analyses of virus genomes. Regions are color coded according to Pan American World Health Organization (PAHO) Zika virus surveillance regions. The diameter of each circle is proportional to the cumulative sum of suspected and confirmed Zika virus cases in each region from 2015 to January 2018 (www.paho.org/hq/index.php?option=com_content&view=article&id=12390&Itemid=42090). Numbers for each region correspond to (upper) cumulative reported Zika virus cases from 2015 to January 2018 and (lower, in parentheses) cumulative confirmed congenital syndrome cases associated with Zika virus infection during the same time period. The estimated dates of earliest virus lineage introduction into each region are shown in white boxes (from Faria et al., 2017; Grubaugh et al., 2017; Metsky et al., 2017).

(B) The estimated surveillance gap, which equals the number of months between the earliest virus lineage introduction into a region (mean estimate from virus genome analyses) and the first confirmed PCR case in each region (depicted by vertical arrows). Time series of the number of Zika virus cases in each region are also shown (adapted from PAHO).

(C) Co-circulation of Zika, dengue, and chikungunya viruses in the northeast region of Brazil (NE-Brazil). Each time series was rescaled onto a common axis. Data was obtained from weekly notified cases published by the Secretaria de Vigilância em Saúde, Ministry of Health, Brazil. Rescaled numbers of microcephaly cases (dashed red line) are those from Faria et al. (2016).

complete or partial Zika virus genomes are publicly available (as of January 18, 2018). In comparison, >5% (1,600) of all known Ebola cases were sequenced during the outbreak in West Africa (Dudas et al., 2017). Our ability to genetically track the Zika virus epidemic was limited not only by difficulties in sequencing Zika virus but also by uneven virus sampling through space and time. Reducing surveillance gaps during future arbovirus outbreaks requires a systematic change, moving from passive case reporting and *ad hoc* genomic sampling to continuous and structured virus reconnaissance. Such a strategy should target patients in at-risk areas, together with population-level surveillance in mosquitoes, primates, and other mammals that could establish cycles of sylvatic transmission.

Several lines of evidence suggest that Zika, chikungunya, and dengue viruses can spread through international corridors of human mobility that connect locations with synchronized vector breeding seasons. Recent Zika cases in Angola, for example, raised concerns that a Zika virus lineage from the Americas could have reached Africa. Conversely, the first case of the chikungunya virus ESCA genotype in Brazil most likely originated from Angola (Nunes et al., 2015). Moreover, a Zika outbreak comprising >7,500 infections was reported in Cabo Verde. Yet, due to a dearth of genomic and epidemiological data, we know little about the origins of these outbreaks, the transmission dynamics of the virus, and the genetic diversity of Zika virus strains present in Africa. Much greater investment in virus surveillance in at-risk regions is needed to minimize local surveillance gaps and generate rapid outbreak control responses.

It is still unclear whether specific Zika virus mutations might have contributed to the virus' rapid spread in the Americas or its association with microcephaly. Studies comparing the replication, tropism, and pathogenesis of Zika virus isolates collected worldwide have reported some phenotypic differences, yet most studies agree that all Zika virus lineages can infect human placental cells and are neurotrophic. These results thus suggest that Zika-associated neuropathology is likely not a newly evolved trait, but these studies cannot exclude the possibility that recent Zika virus mutations could have enhanced this trait. For example, genomic studies of past West Nile, chikungunya, and Ebola virus outbreaks have discovered individual mutations that increase viral replication in new hosts and could exacerbate transmission (e.g., Diehl et al., 2016). Two Zika virus mutations that evolved shortly before the epidemic in the Americas (Faria et al., 2016) have been reported to enhance *Aedes aegypti* infectivity (NS1 A188V; Liu et al., 2017) and neurovirulence (prM S17N; Yuan et al., 2017). While these mutations could have elevated the epidemic potential of Zika virus, there is no current epidemiological support of

such a hypothesis. Moreover, mutations that pre-date the epidemic cannot explain the high variation in reported microcephaly rates across the Americas (e.g., de Oliveira et al., 2017). Systematic functional studies of Zika virus mutations from before and during the epidemic, matched with epidemiological data, are required to determine whether recent virus genetic factors have contributed to the epidemic.

A variety of mathematical models have been used to investigate the transmission potential of Zika virus, the risk of neuropathology associated with infection, and the drivers of virus spread. Virus genomic data have the potential to significantly enhance the utility of more traditional epidemiological investigations (Dudas et al., 2017). Phylogenetic birth-death models, for example, provide an accurate way to estimate fluctuations through time in the effective reproductive number of an epidemic (Stadler et al., 2013). The ability of virus genomes to cross-check conclusions drawn from case report data may be particularly valuable in locations where such data are uncertain (e.g., during co-circulation of multiple arboviruses; Figure 1C). Moreover, advances in phylogeographic methods mean that the dynamics of viral spread and persistence across regions can be investigated at high resolution. In one recent example, analyses of Ebola virus genomes revealed that the spread of the epidemic in West Africa involved intense transmission between larger and closer populations (Dudas et al., 2017). Given the scale of the Zika epidemic, however, at least an order of magnitude more virus genome sequences are required to successfully undertake an equally detailed analysis of Zika virus transmission.

Despite the many challenges and questions that remain, the open and collaborative nature of research undertaken by the infectious disease community during recent emerging epidemics gives ample reason for optimism. In contrast to the historical model of data generation by a few genomics centers, the Zika epidemic showed that many smaller laboratories can combine efforts in response to a public health crisis. Open sharing of protocols, epidemiological data, and pathogen genome sequences from laboratories and organizations worldwide played a central role in obtaining insights into the origins, timing, and spread of the Zika epidemic.

ACKNOWLEDGMENTS

We thank the clinicians, epidemiologists, and biologists at the front lines of the epidemic and the many laboratories worldwide that shared protocols and data ahead of print. N.D.G. is supported by NIH training grant 5T32AI007244-33. N.R.F. is funded by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (grant 204311/Z/16/Z). K.G.A. is a Pew Biomedical Scholar and is supported by NIH NCATS CTSA UL1TR001114 and The Ray Thomas Foundation. O.G.P. is supported by the European Union's Seventh Framework Programme (FP7/2007-2013)/European Research Council (614725-PATHPHYLODYN) and by the Oxford Martin School.

DECLARATION OF INTERESTS

O.G.P. has received research funding and consultancy income from Metabiota, San Francisco, CA.

REFERENCES

Diehl, W.E., Lin, A.E., Grubaugh, N.D., Carvalho, L.M., Kim, K., Kyawe, P.P., McCauley, S.M., Donnard, E., Kucukural, A., McDonel, P., et al. (2016). Ebola virus glycoprotein with increased infectivity dominated the 2013-2016 epidemic. Cell *167*, 1088–1098.e6.

Dudas, G., Carvalho, L.M., Bedford, T., Tatem, A.J., Baele, G., Faria, N.R., Park, D.J., Ladner, J.T., Arias, A., Asogun, D., et al. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature *544*, 309–315.

Faria, N.R., Azevedo, R.D.S.D.S., Kraemer, M.U.G., Souza, R., Cunha, M.S., Hill, S.C., Thézé, J., Bonsall, M.B., Bowden, T.A., Rissanen, I., et al. (2016). Zika virus in the Americas: early epidemiological and genetic findings. Science *352*, 345–349.

Faria, N.R., Quick, J., Claro, I.M., Thézé, J., de Jesus, J.G., Giovanetti, M., Kraemer, M.U.G., Hill, S.C., Black, A., da Costa, A.C., et al. (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature *546*, 406–410.

Grubaugh, N.D., Ladner, J.T., Kraemer, M.U.G., Dudas, G., Tan, A.L., Gangavarapu, K., Wiley, M.R., White, S., Thézé, J., Magnani, D.M., et al. (2017). Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature *546*, 401–405.

Kindhauser, M.K., Allen, T., Frank, V., Santhana, R.S., and Dye, C. (2016). Zika: the origin and spread of a mosquito-borne virus. Bull. World Health Organ. *94*, 675–686C.

Lessler, J., Chaisson, L.H., Kucirka, L.M., Bi, Q., Grantz, K., Salje, H., Carcelen, A.C., Ott, C.T., Sheffield, J.S., Ferguson, N.M., et al. (2016). Assessing the global threat from Zika virus. Science *353*, aaf8160.

Liu, Y., Liu, J., Du, S., Shan, C., Nie, K., Zhang, R., Li, X.-F., Zhang, R., Wang, T., Qin, C.-F., et al. (2017). Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. Nature *545*, 482–486.

Metsky, H.C., Matranga, C.B., Wohl, S., Schaffner, S.F., Freije, C.A., Winnicki, S.M., West, K., Qu, J., Baniecki, M.L., Gladden-Young, A., et al. (2017). Zika virus evolution and spread in the Americas. Nature *546*, 411–415.

Nunes, M.R.T., Faria, N.R., de Vasconcelos, J.M., Golding, N., Kraemer, M.U.G., de Oliveira, L.F., Azevedo, Rdo.S., da Silva, D.E., da Silva, E.V., da Silva, S.P., et al. (2015). Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med. *13* https://doi.org/10.1186/s12916-015-0348-x.

de Oliveira, W.K., de França, G.V.A., Carmo, E.H., Duncan, B.B., de Souza Kuchenbecker, R., and Schmidt, M.I. (2017). Infection-related microcephaly after the 2015 and 2016 Zika virus outbreaks in Brazil: a surveillance-based analysis. Lancet *390*, 861–870.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. *12*, 1261–1276.

Souza, W.V., Araújo, T.V., Albuquerque, Mde.F., Braga, M.C., Ximenes, R.A., Miranda-Filho, Dde.B., Bezerra, L.C., Dimech, G.S., Carvalho, P.I., Assunção, R.S., et al. (2016). Microcephaly in Pernambuco State, Brazil: epidemiological characteristics and evaluation of the diagnostic accuracy of cutoff points for reporting suspected cases. Cad. Saude Publica *32*, e00017216.

Stadler, T., Kühnert, D., Bonhoeffer, S., and Drummond, A.J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl. Acad. Sci. USA *110*, 228–233.

Yuan, L., Huang, X.-Y., Liu, Z.-Y., Zhang, F., Zhu, X.-L., Yu, J.-Y., Ji, X., Xu, Y.-P., Li, G., Li, C., et al. (2017). A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. Science *358*, 933–936.