bioRxiv preprint doi: https://doi.org/10.1101/113100; this version posted March 3, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Genetic characterization of the Zika virus epidemic in the US Virgin Islands

**Allison Black[1,2], Barney Potter[1], Gytis Dudas[1], Leora Feldstein[1], Nathan D Grubaugh[3], Kristian G Andersen[3], Brett R Ellis[4], Esther M Ellis[4] & Trevor Bedford[1]**

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [2]Department of Epidemiology, University of Washington, Seattle, WA, USA, [3]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA, USA, [4]United States Virgin Islands Department of Health, Christiansted, USVI

March 1, 2017

**Abstract**

Here we release draft genome sequences of Zika virus (ZIKV) that were sequenced from PCR-positive diagnostic specimens collected by the United States Virgin Islands Department of Health (USVI DoH) as part of their ongoing response to the ZIKV outbreak. We will use these sequences to conduct a genomic epidemiological study of ZIKV transmission in the USVI. We are releasing these genomes in the hope that they are useful for those individuals involved in the public health response to ZIKV and to other groups working to understand Zika virus transmission and evolution.

## Purpose

This work represents a pre-publication sharing of pathogen genomic data of public health significance. We believe that open sharing of sequence data and early analyses is necessary to inform public health response in a timely fashion [1]. We intend for this to serve as a 'marker paper' specifying analyses planned for these data. Please consider the analysis of Zika in the USVI (as described in the 'Genome Announcement' below) as off limits. However, integrating these data as auxiliary data into other sorts of publications is fine. All we ask is for a citation and a heads up. We plan to replace this 'announcement' with a preprint and eventually with a publication. The citation will be continually updated throughout this process.

bioRxiv preprint doi: https://doi.org/10.1101/113100; this version posted March 3, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Genome Announcement

For pathogens causing primarily asymptomatic infections, such as ZIKV, traditional epidemiological methods may yield biased estimates of incidence and outbreak growth due to incomplete case counts. Additionally, contact tracing to infer transmission chains and geographic patterns of spread may also break down when the majority of cases are unrecorded. Genomic epidemiological methods make use of a fundamental principle that the evolution of pathogens occurs on the same timescale as hosts transmit pathogens. Viral replication results in the accumulation of mutations, evidence that an infection has occurred even if asymptomatic or unreported. Thus using sequence data, in combination with temporal and demographic data, we can date when an outbreak started, describe the spatial spread of disease, infer how quickly an outbreak is growing, or determine transmission chains at the individual or the population level.

In collaboration with the US Virgin Islands Department of Health (USVI DoH) we are sequencing Zika virus (ZIKV) full genomes from PCR-positive diagnostic specimens collected during the ongoing DoH response to the ZIKV outbreak. We extracted viral RNA from inactivated serum or urine samples using a QIAamp Viral RNA mini kit (Qiagen). Viral RNA was reverse-transcribed, amplified, and sequenced according to the protocol in Quick et al [2]. Genomes were sequenced on either the MinION or Illumina MiSeq, as indicated, with a subset of samples sequenced on both platforms to ensure the validity of the sequences across platforms.

Our study seeks to describe when ZIKV was introduced to the USVI, how many separate introductions of ZIKV occurred, the degree to which different introductions contributed to local transmission, and describe patterns of ZIKV spread between the different US Virgin Islands. We will also use joint inference of genomic and epidemiological data to estimate the proportion of cases that are asymptomatic, and use inferred case counts to more accurately estimate $R_0$ over the outbreak.

# Preliminary results

To date we have sequenced 11 Zika viruses from the USVI. In Figure 1 we show where they fall in the broader spectrum of viral diversity. It is clear that many of the infections are quite similar, while other infections appear to represent shorter transmission chains following introduction.

bioRxiv preprint doi: https://doi.org/10.1101/113100; this version posted March 3, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

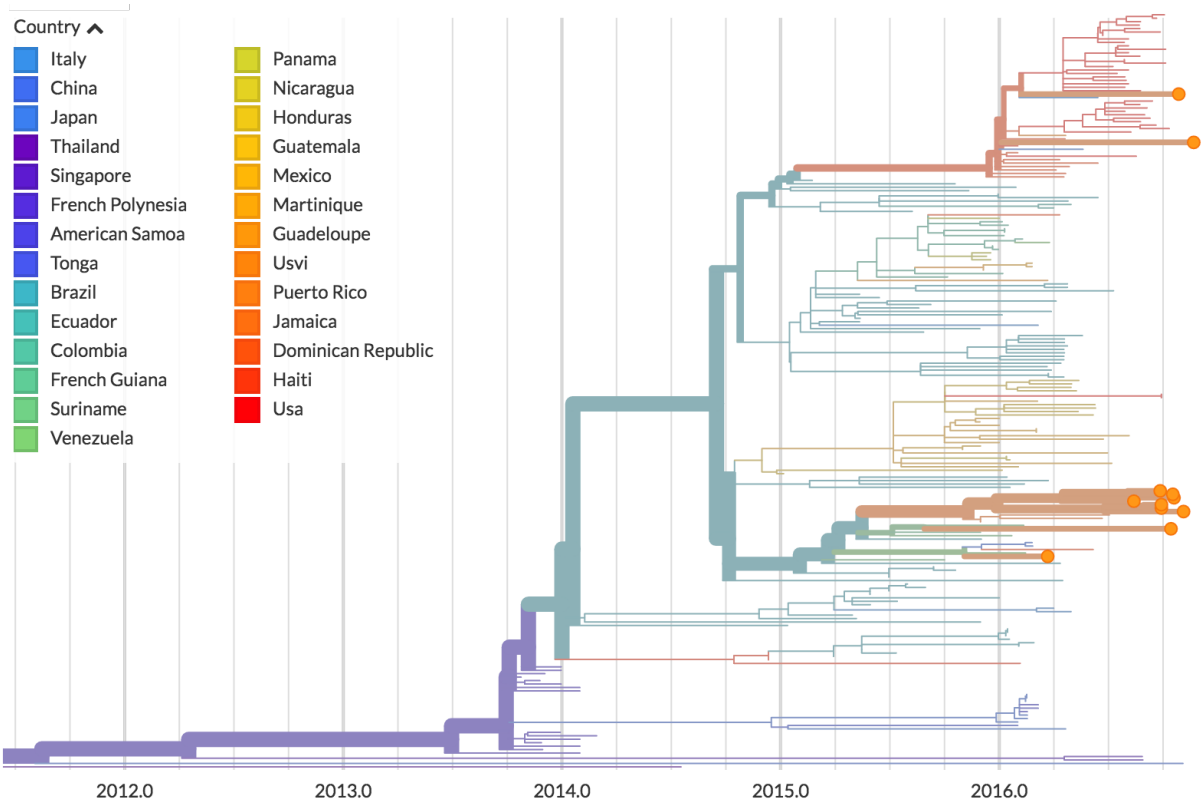


**Figure 1. Preliminary phylogeny showing 11 Zika genomes from the USVI.** This tree is taken from nextstrain.org/zika and shows a time calibrated phylogeny of publicly available Zika genomes.

# Data

All genome sequence data and metadata can be found on GitHub at github.com/blab/zika-usvi.

## References

1. Bedford T. On scientific publishing practices in the face of public health crises. URL http://bedford.io/blog/scientific-publishing-practices/.

2. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, et al. (2017) Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. bioRxiv : 098913.