



MRC Centre for
Global Infectious
Disease Analysis

Imperial College
London

Arboviral Genomic Epidemiology: Current Landscape and Implications for Risk Assessment

Nuno R. Faria

Professor in Virus Genomic Epidemiology

Co-Lead Pathogen Genomic Epidemiology Unit

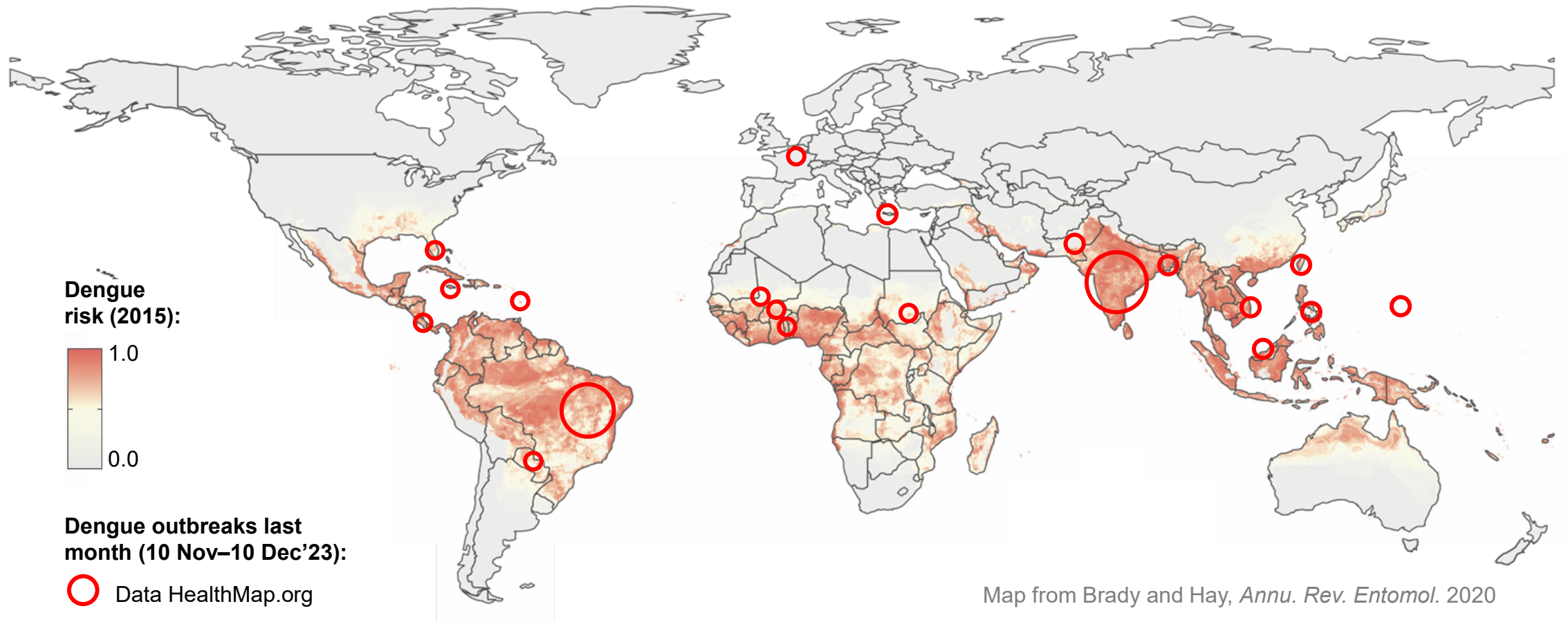
MRC Centre for Global Infectious Disease Analysis

School of Public Health, Imperial College London, UK (SPH-ICL)

Institute of Tropical Medicine, University of São Paulo, Brazil (IMT-USP)

Email: nfaria@ic.ac.uk | LinkedIn: [nmrfaria](https://www.linkedin.com/in/nmrfaria) | Website: www.imperial.ac.uk/people/n.faria

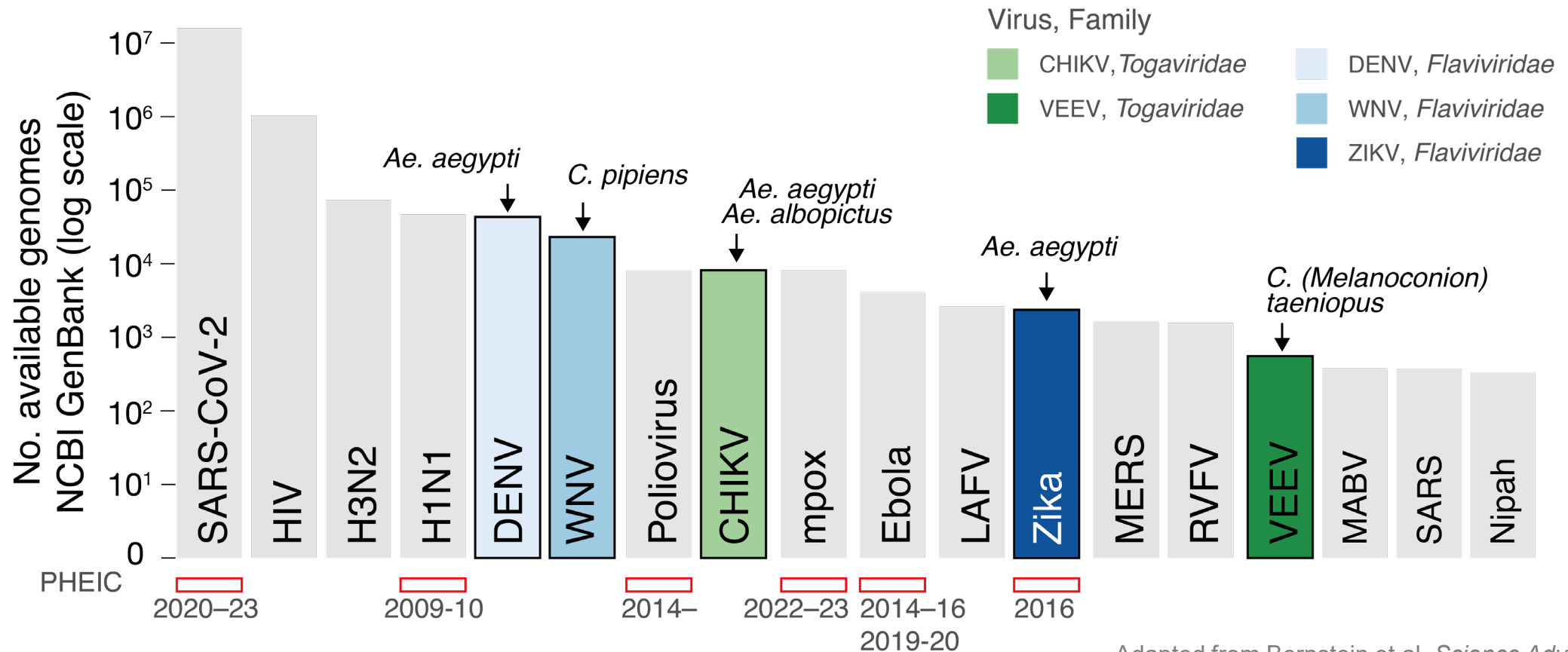
Anthropogenic changes and the global expansion of *Aedes*-borne viruses



Dengue: 3.9 billion people at risk, 390 million infections per year, and **7 genomes per million cases** (2023)

Overview genome sequencing volume for selected pathogenic viral threats

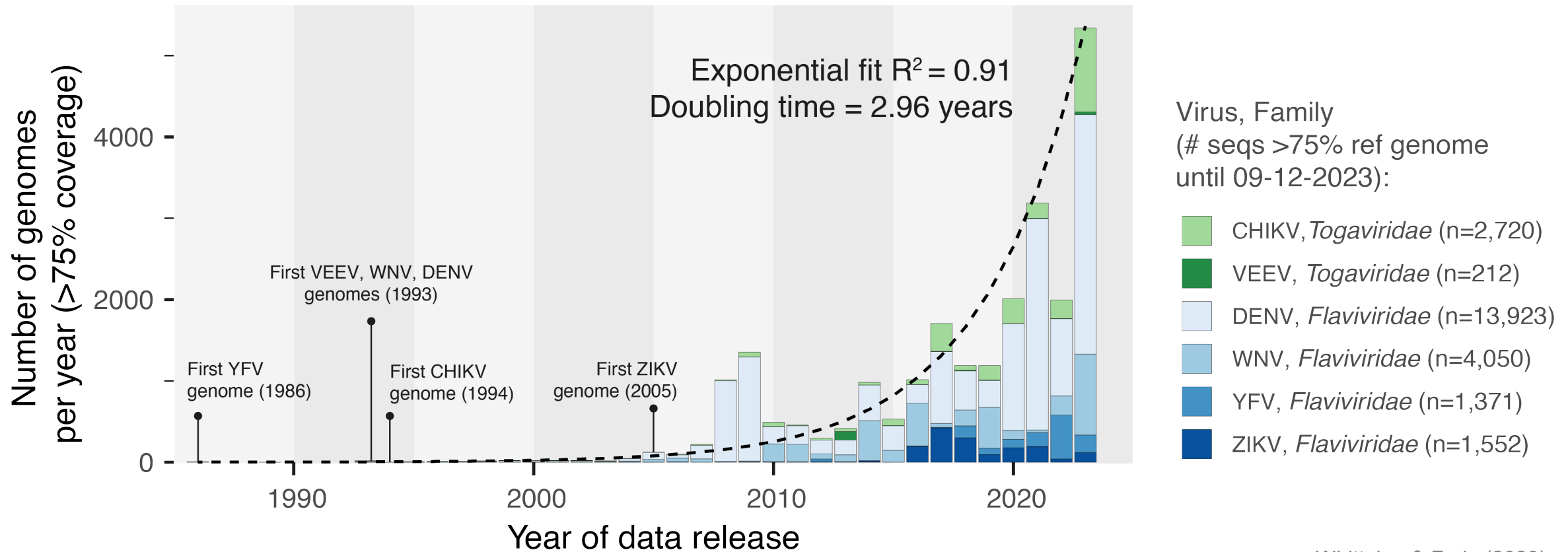
Relatively limited genomic sequencing data available for chikungunya (CHIKV), Zika (ZIKV), and Venezuelan equine encephalitis (VEEV) viruses.



Adapted from Bernstein et al. *Science Advances* 2022

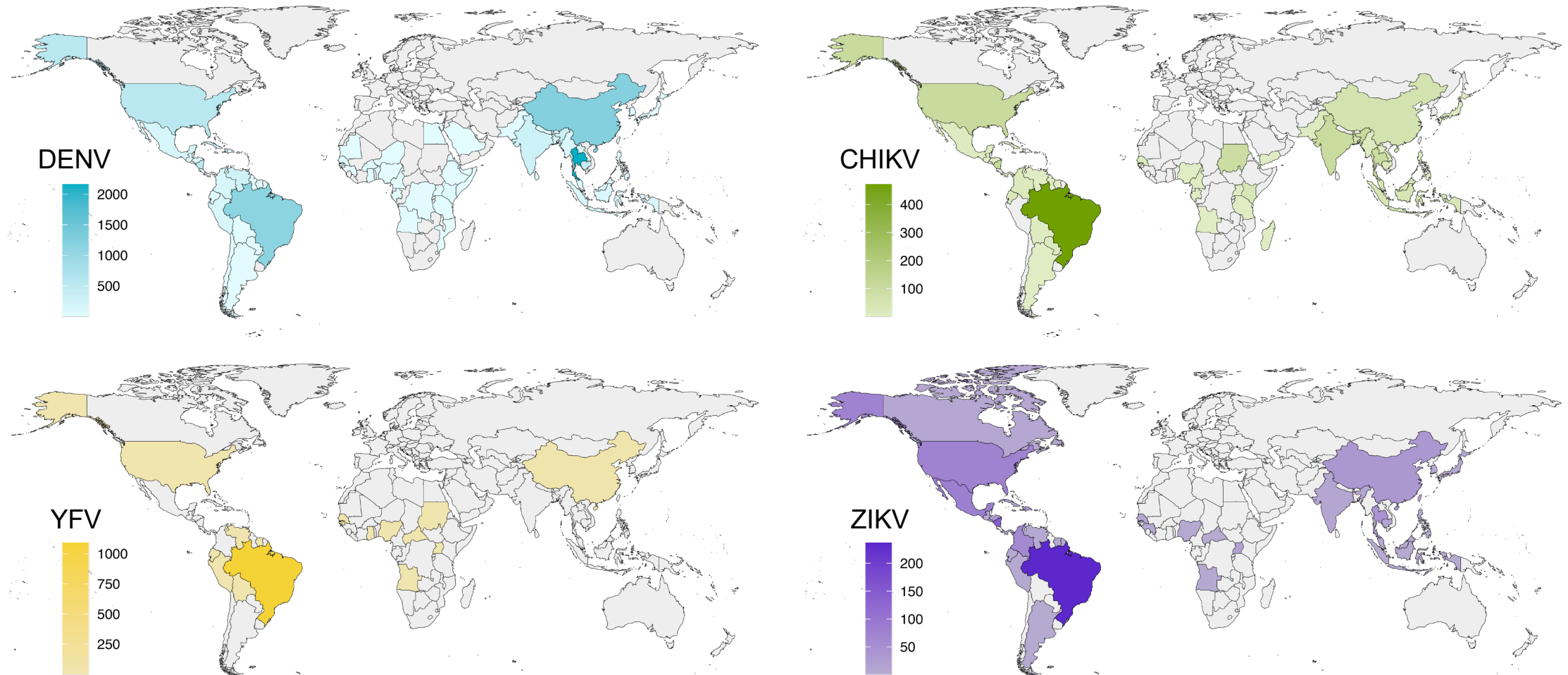
Exponential growth of genomic sequencing data for arboviral threats

Large post-pandemic increase with 5,340 arboviral genome sequences shared in 2023 alone.



Whittaker & Faria (2023)
Data from NCBI GenBank (>=75% genome) (2023)

Mapping sequencing capacity for key *Aedes*-borne viruses – DENV, YFV, ZIKV, and CHIKV

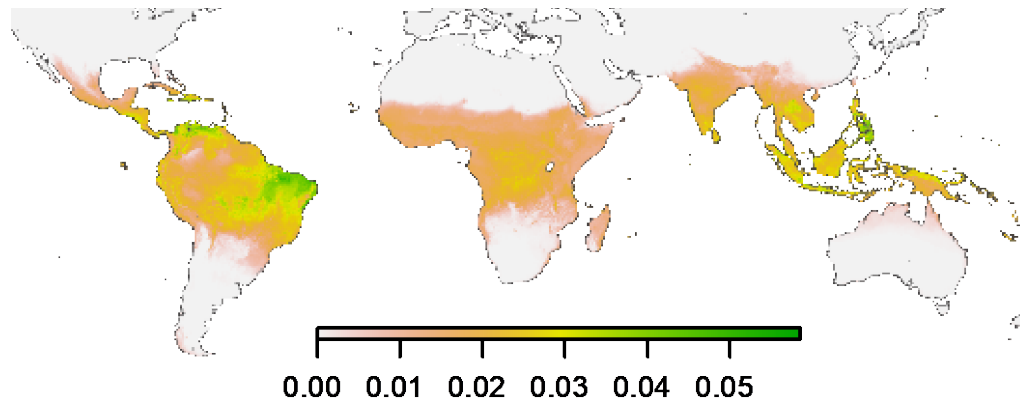


Whittaker & Faria (2023)
Data from NCBI GenBank ($\geq 75\%$ genome) (2023)

How representative are arboviral genome sequencing efforts worldwide?

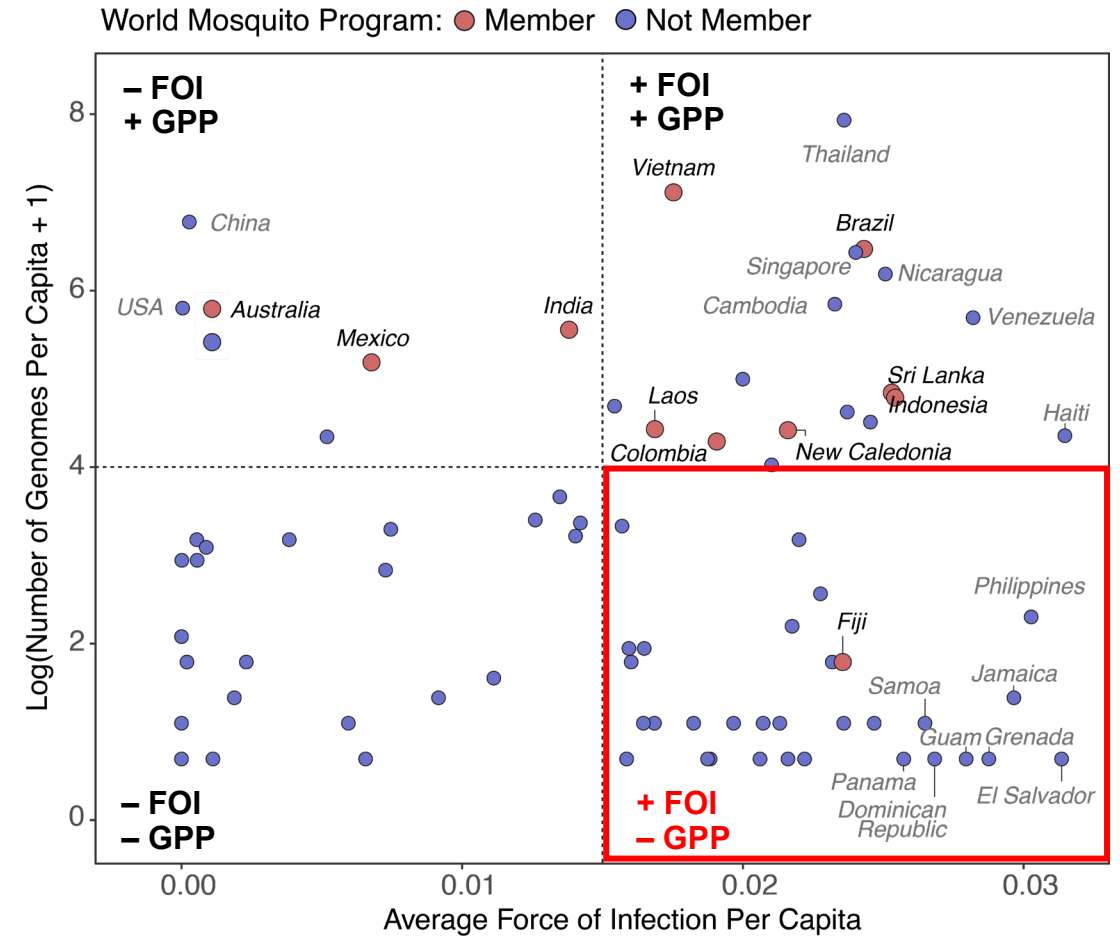
Analysis of DENV sequencing metadata (NCBI)

- Ten countries produce 82% DENV data (Thailand, Vietnam, China, Brazil, Singapore, Nicaragua, Cambodia, USA, Venezuela, India).
- Unbalanced genome data per capita in relation to dengue force of infection: identification of undersampled hotspots of diversity



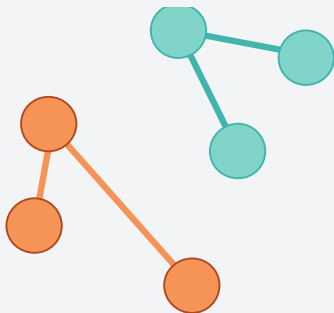
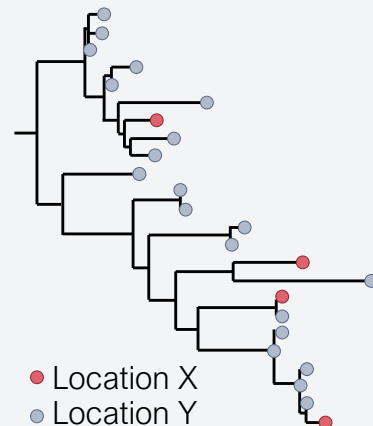
FOI dengue estimates from Cattarino et al. *Science Trans Med* 2020

Current landscape



Whittaker & Faria (2023)
Data from NCBI GenBank ($\geq 75\%$ genome) (2023)

Maximizing the public health benefit of genomic sequencing of arboviral threats

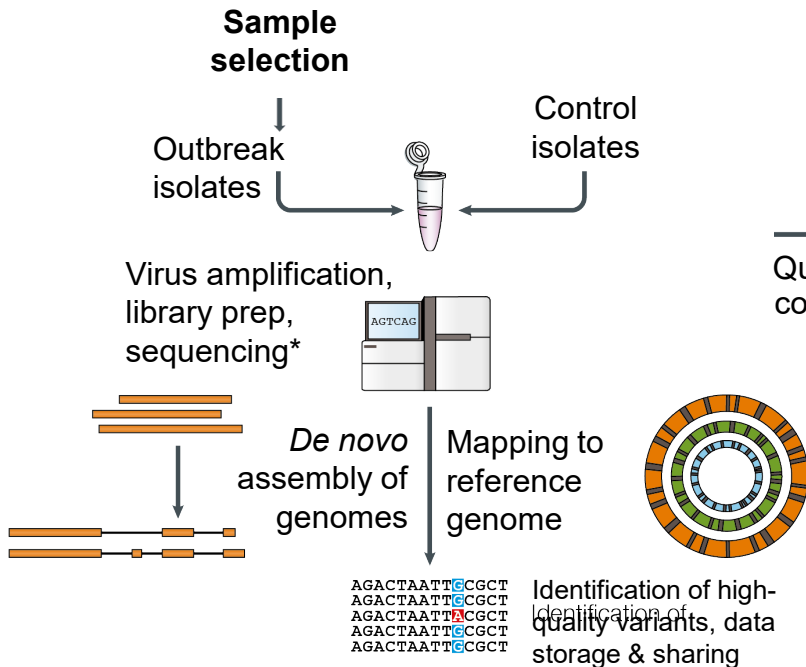
Goal	Pathogen identification, understanding origins of novel outbreaks, spillover/spillback	Providing epidemiological characterisation of lineages in transmission hotspots	Design and monitoring of diagnostics, therapeutics, vaccines, investigation of phenotypic changes	Investigating arbovirus emergence, persistence, evolution and transmission, at different scales and cycles
	Pathogen-agnostic sequencing of human, mosquitoes and non-human primate samples	Non-random, e.g. climate-and mobility-informed sampling	In silico assessment, experimental work (e.g., CHIKV A226V)	Random representative population sequencing
			<pre> AGACTAATTGCGCT AGACTAATTGCGCT AGACTAATTACGCT AGACTAATTGCGCT AGACTAATTGCGCT </pre>	
	Untargeted sequencing	Targeted sequencing		Population sequencing

Adapted from "A guide to implementation for maximum impact on public health", WHO, 2021

Key components of pathogen genomic epidemiology

Genomic sequencing

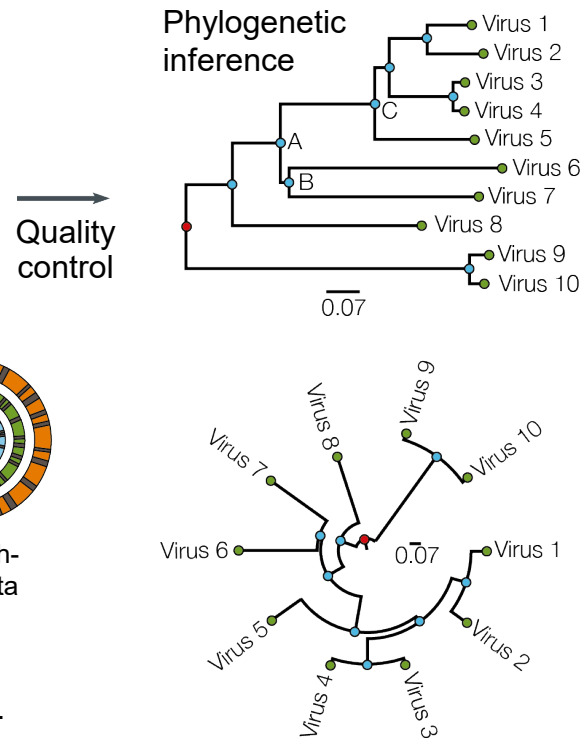
Access to samples and linked metadata



*First, second and third generation sequencers. Typically requires a priori knowledge of the virus (& specific RT-PCR tests).

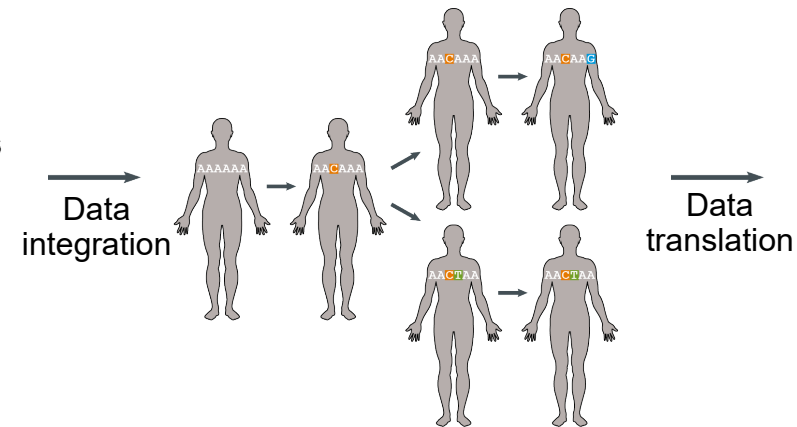
Bioinformatics

Outbreak and background datasets



Genetic Analysis

Phylogenetic inference and ancestral reconstruction



Enzootic and epizootic outbreaks: mosquitos and non-human primate genomes should be included

Translation

Outbreak Responders

Dashboards & visualizations
Diversity, cases and clusters
Transmissibility & severity
Monitor diagnostics & vector control
Climate-, mobility- and genomic-driven forecasts

Public health action

Adapted from "A guide to implementation for maximum impact on public health", WHO, 2021; and Gardy & Loman, *Nature Reviews Microb* 2018

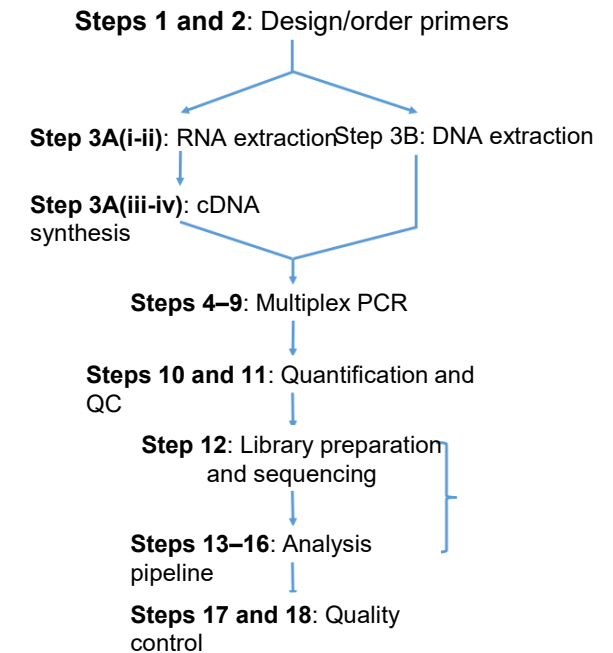
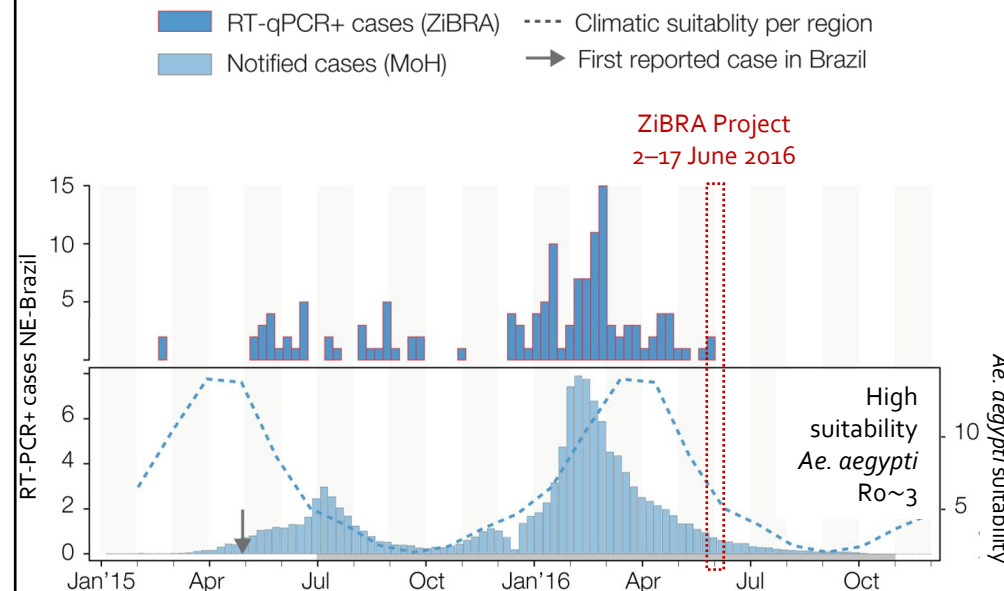
Optimizing the detection of Zika using climate-informed temporal sampling

Targeted sampling and sequencing strategies using *Aedes* suitability and amplicon-based sequencing to investigate introduction and persistence of Zika virus in Brazil.

Monthly climatic suitability for *Aedes aegypti*

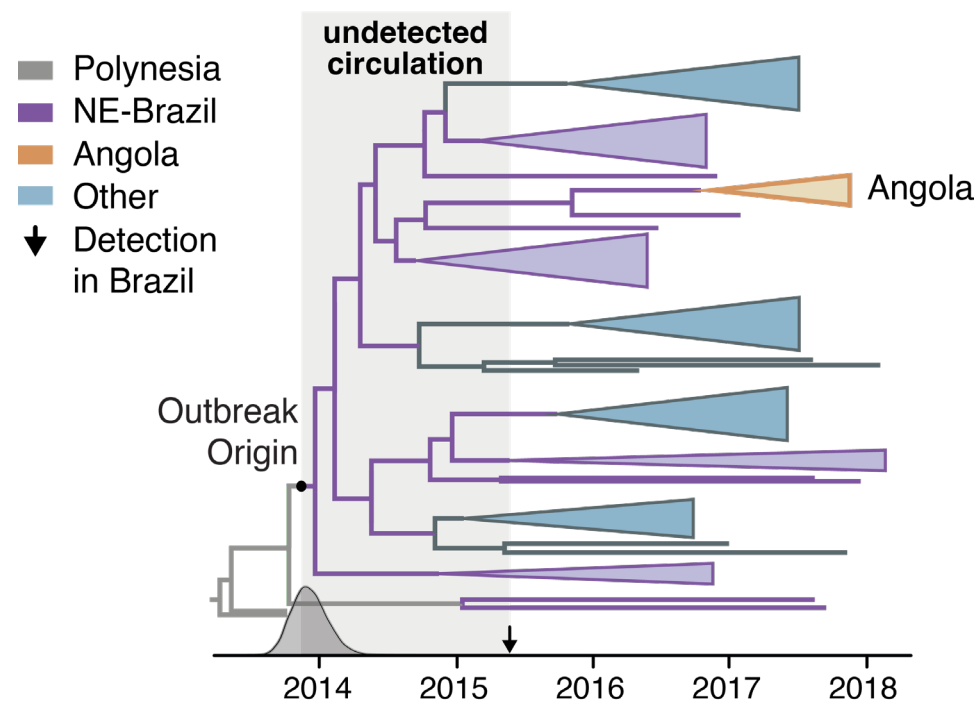
RT-PCR+ Zika cases (ZiBRA mobile lab) and weekly MoH notified cases

Amplicon-based sequencing and protocol optimisation in real-time



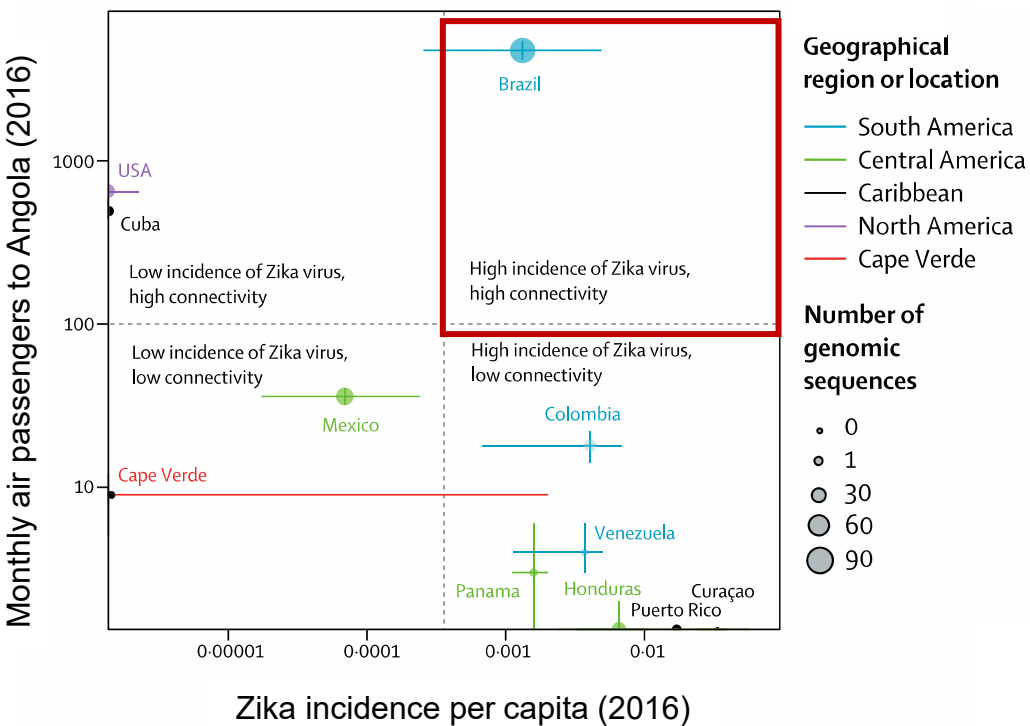
Genomic surveillance across mobility corridors with synchronized suitability of *Aedes aegypti*

Zika silent circulation underlies underreporting in locations at risk for zoonotic spillover



Dating the emergence of arboviral lineages provides a baseline for understanding trends in disease severity

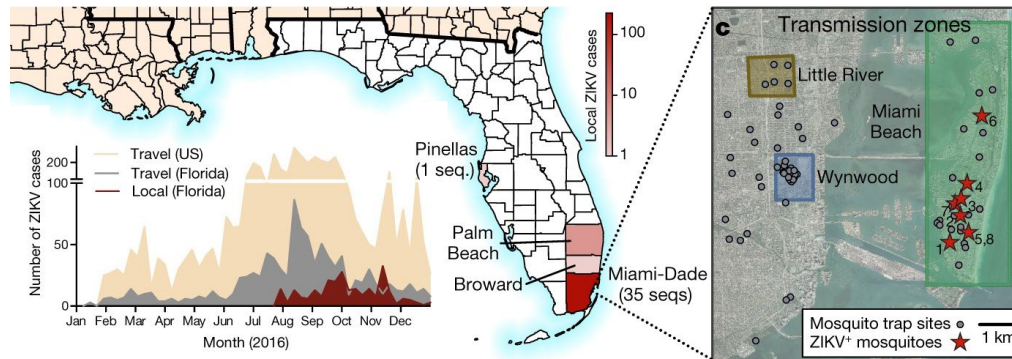
Air travel linked to spread of Zika virus Asian lineage to Angola



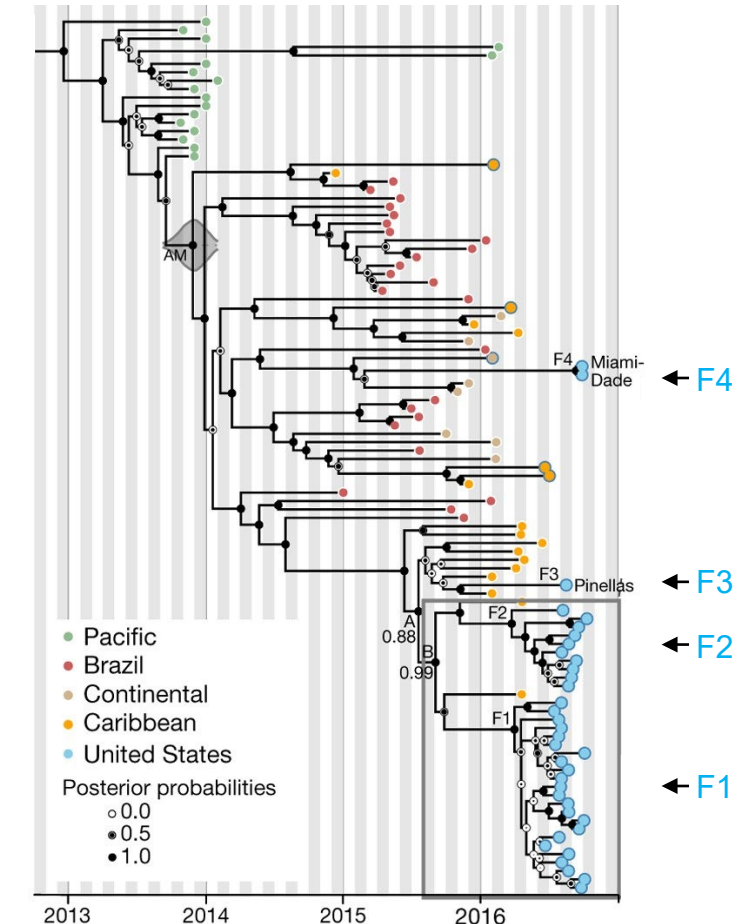
Faria et al. *Science* 2016; Faria et al. *Nature* 2017; Faria et al., *Genome Medicine* 2016; Hill et al., *Lancet Inf Dis* 2019

Genomic data improves risk potential estimates for long-term Zika virus persistence

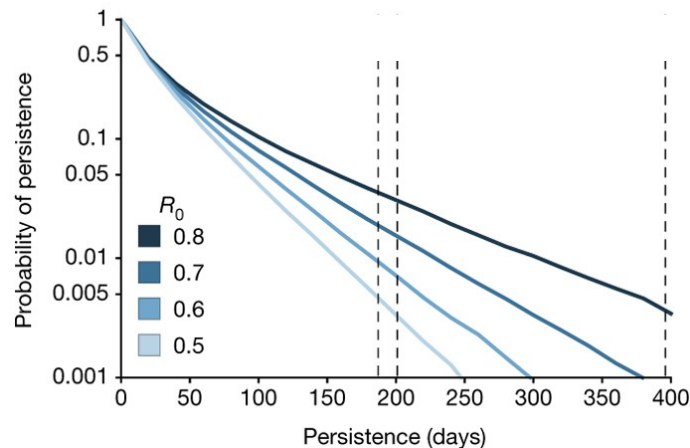
241 locally acquired ZIKV cases in Florida



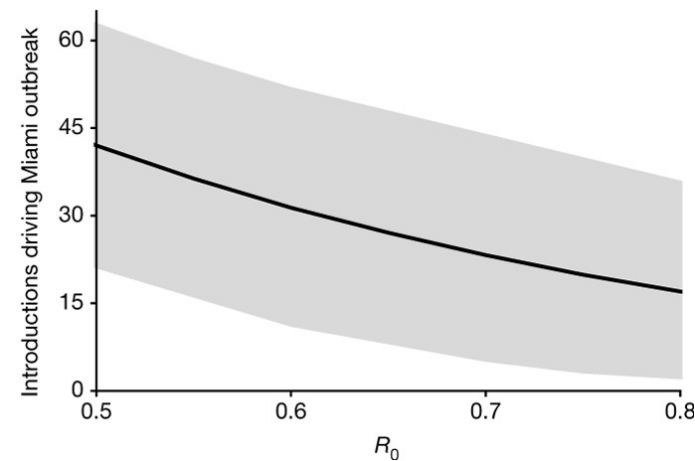
39 genomes => 4 lineages



Prob. lineage >1y ~0.5%



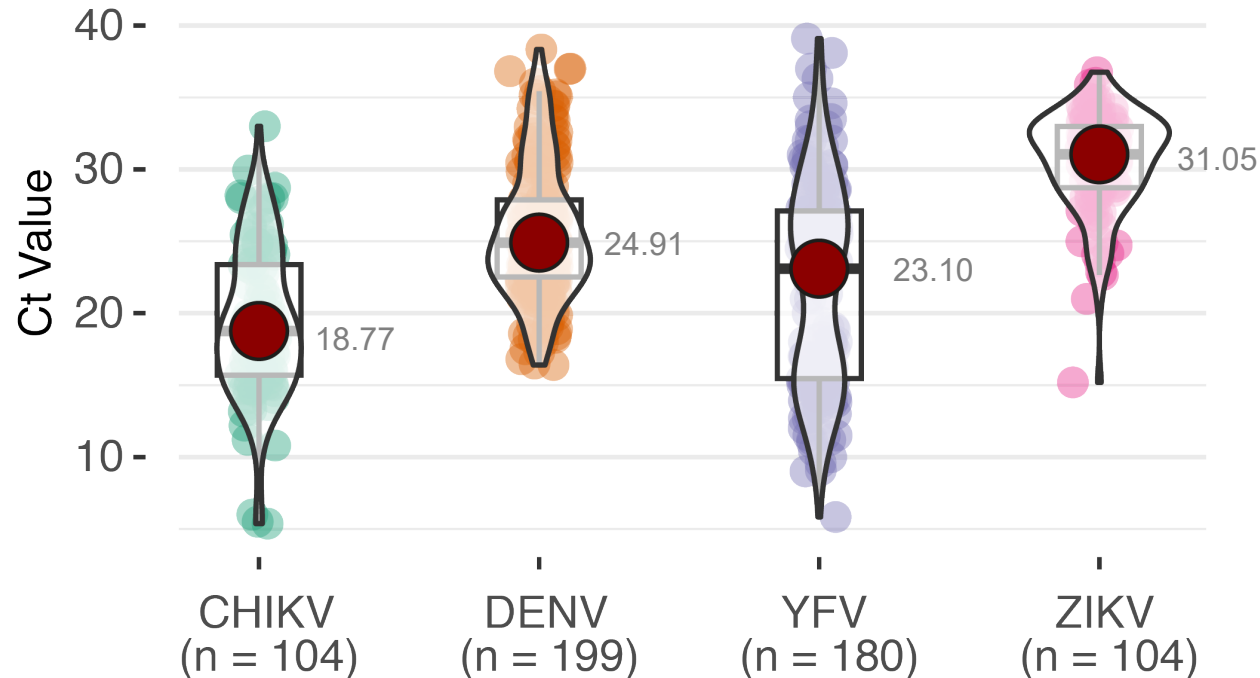
~40 introductions in Miami



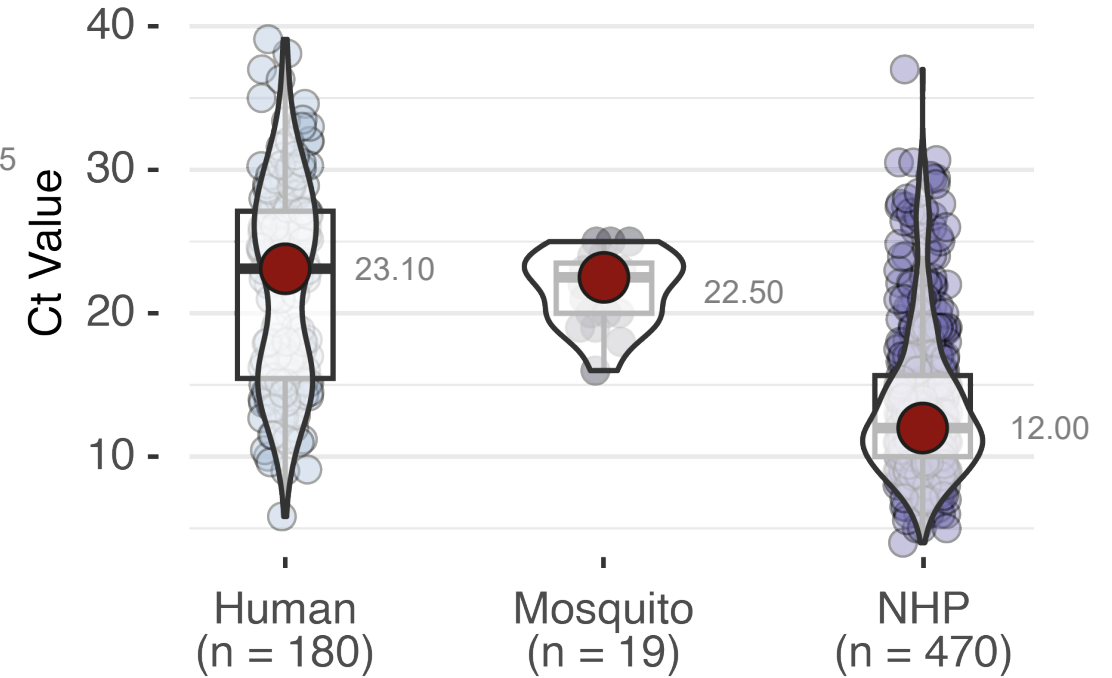
Arboviral transmission dynamics impacts detection and genome sequencing volumes

- ZIKV has lowest viral titres compared to CHIKV, DENV and YFV (human samples).
- YFV high viral titres, particularly for non-human primate samples.

Cycle thresholds (CT) of sequenced genomes (>50% FG) for HUMAN samples



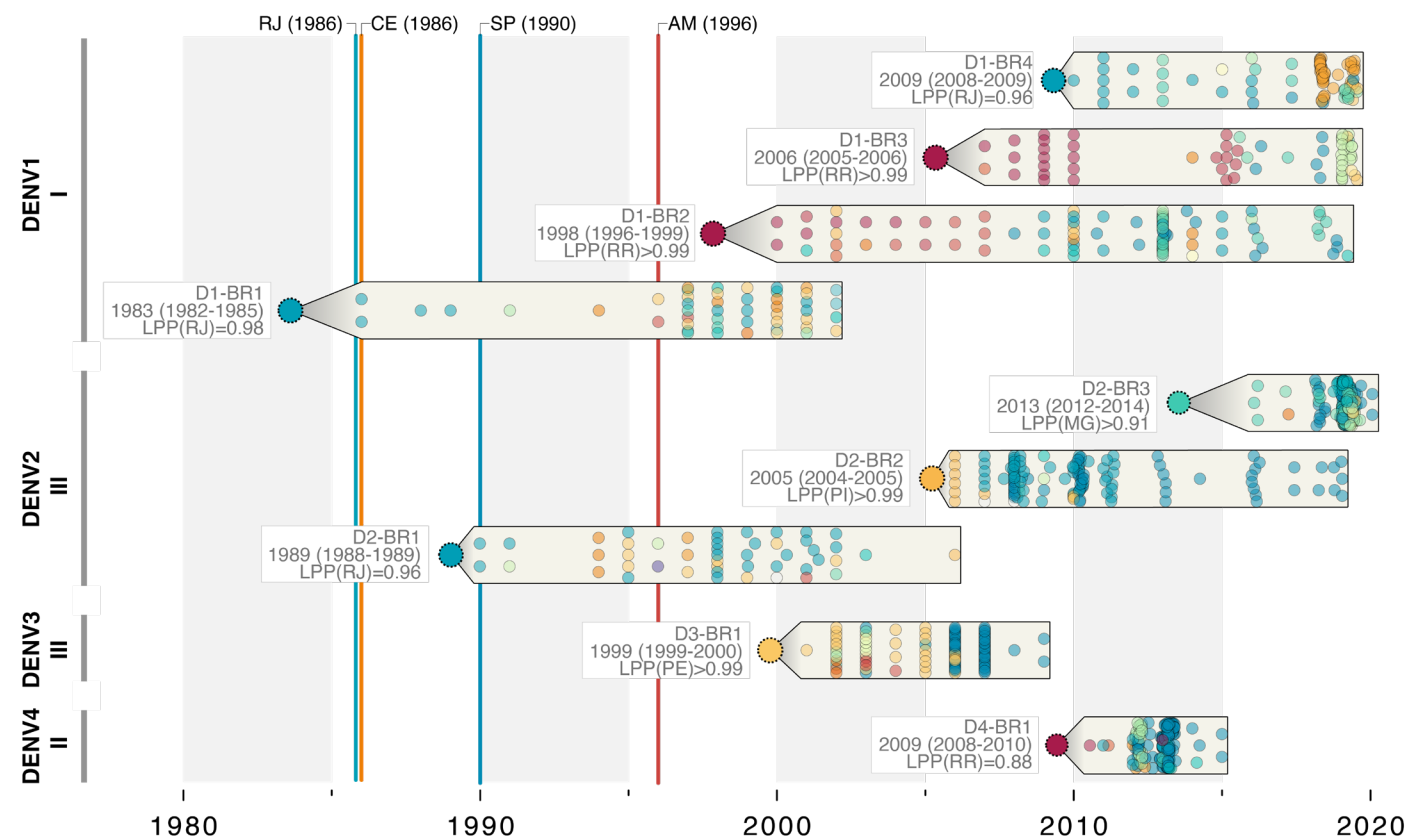
CT values of sequenced YFV genomes (>50% FG) for different HOST species



Data from the CADDE Project (<https://www.caddecentre.org/>) generated at IMT-USP Sao Paulo, Brazil.

Dengue transmission forecasting improves when considering genomic surveillance data

Summary of phylogenetic inference (DENV1–4 from Brazil)



Legend:

DENV1–4: Dengue virus serotypes 1–4

I–III: Dengue genotypes

BR1–BR4: D1–D4 lineages (n>20) in Brazil

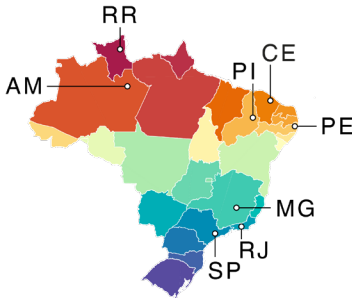
LPP: Location Posterior Probability

Vertical lines: dates and locations of first reported cases in each region

○ Tips (sampling locations)

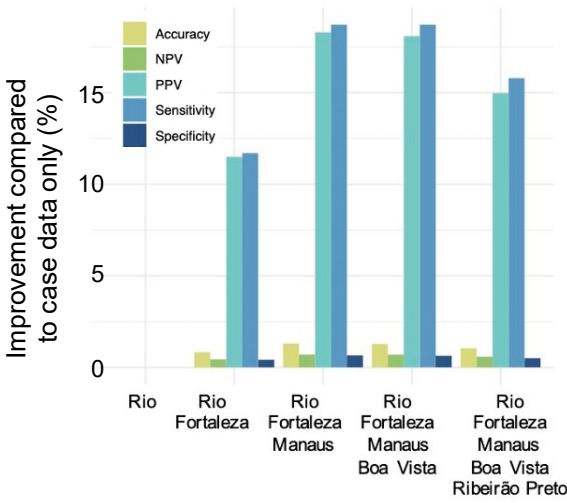
○ Inferred root location of main lineages

Tips coloured according to federal unit of sample collection:



Best performing forecasting models

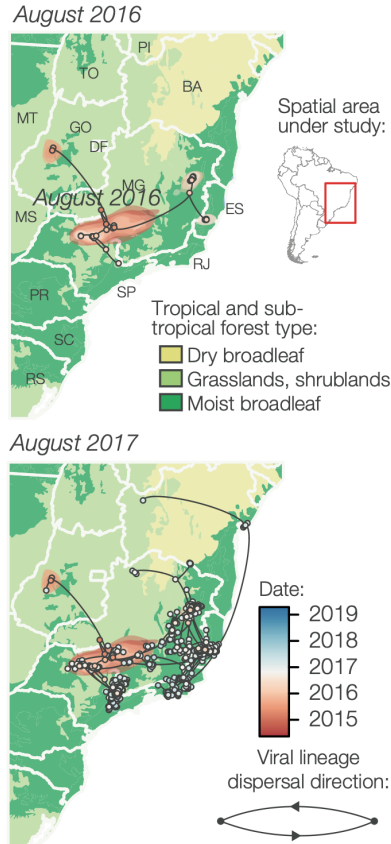
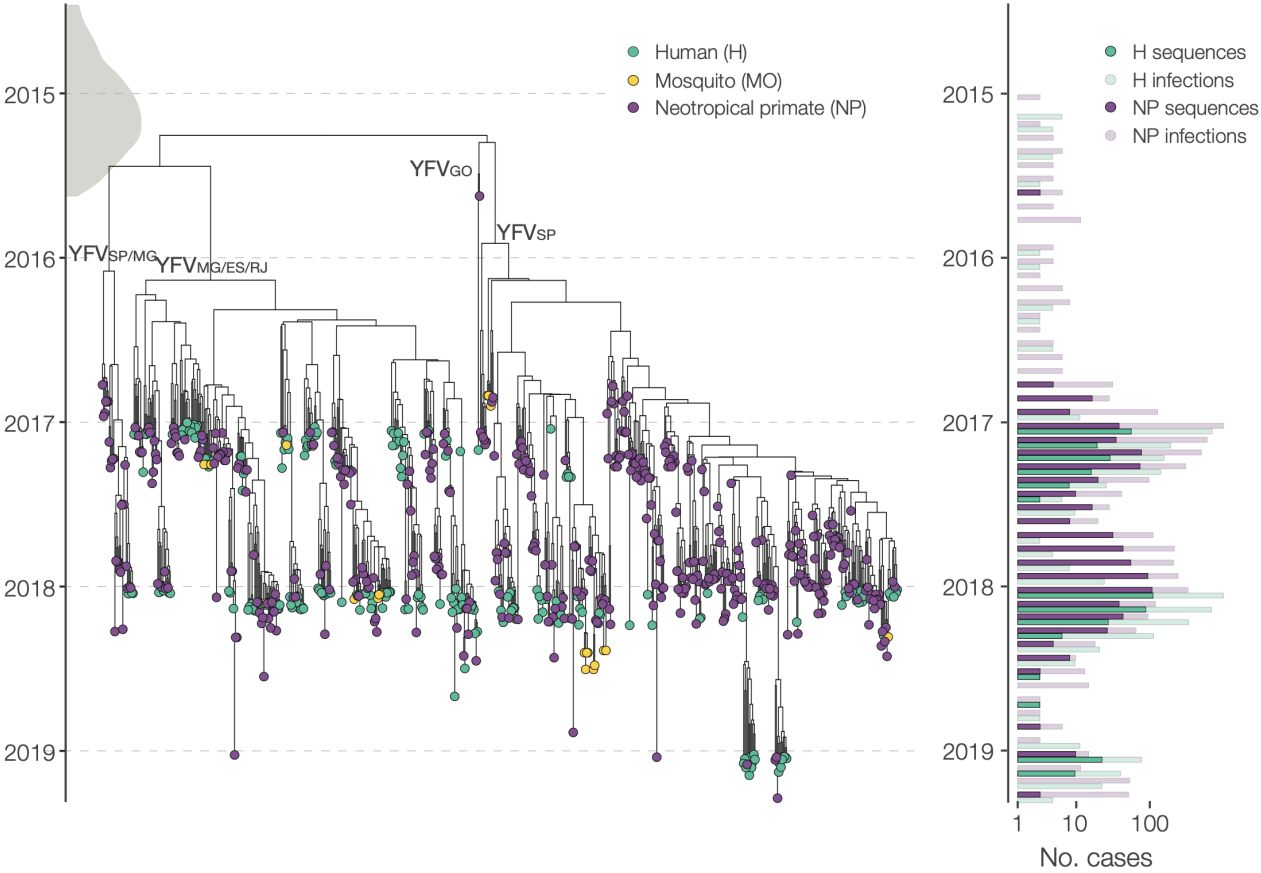
Improved performance in predicted historical expansion of dengue in Brazil when considering dynamics inferred from phylogenetic inference.



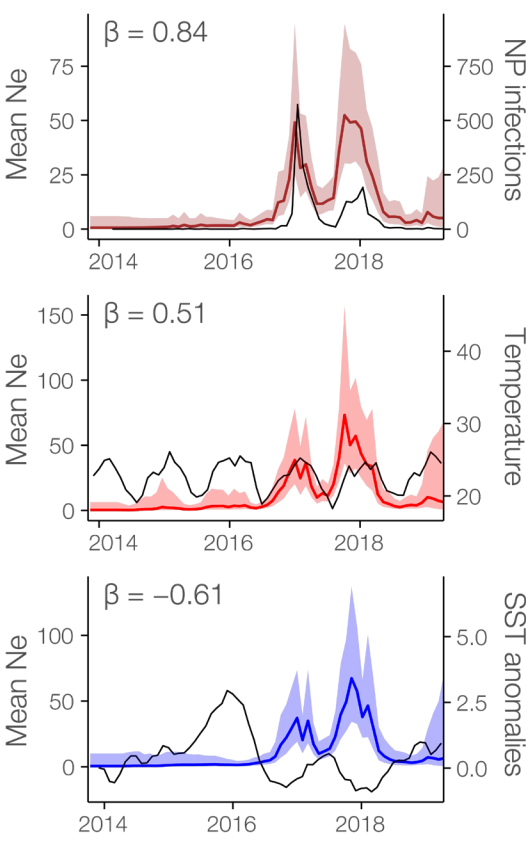
Climate and land-use shape YFV spread

- Collaborative effort across multiple institutions to triple amount of YFV available data: nearly **15%** of confirmed infected humans or NPs.

YFV sylvatic transmission at country-wide scale



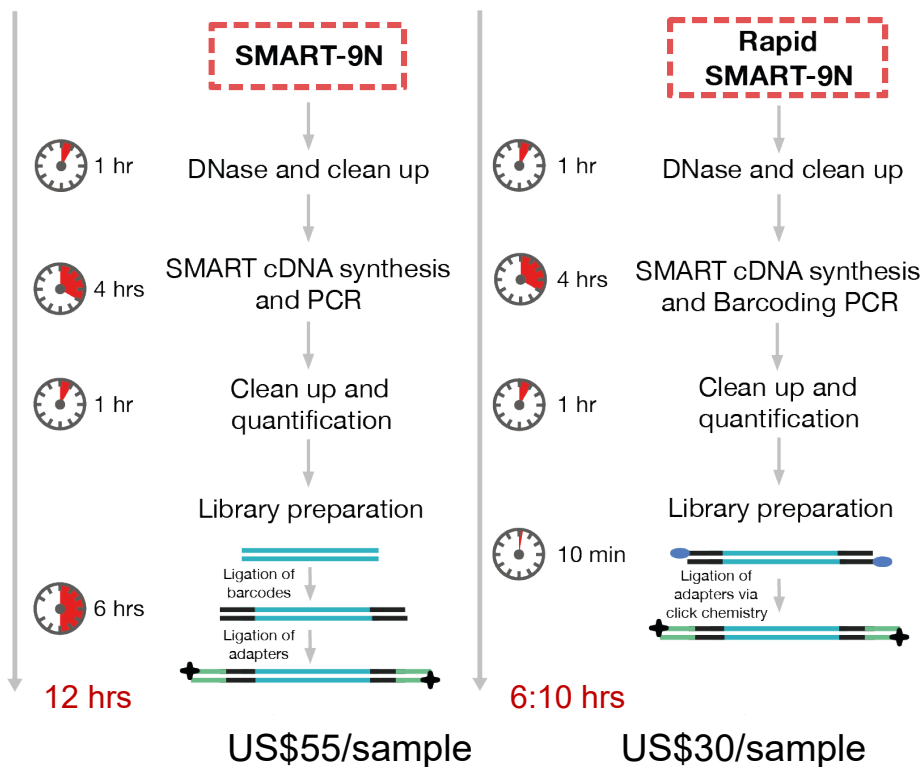
Hypothesis testing



Hill SC et al. BioRxiv under review 2023 (on behalf of the CADDE Project)

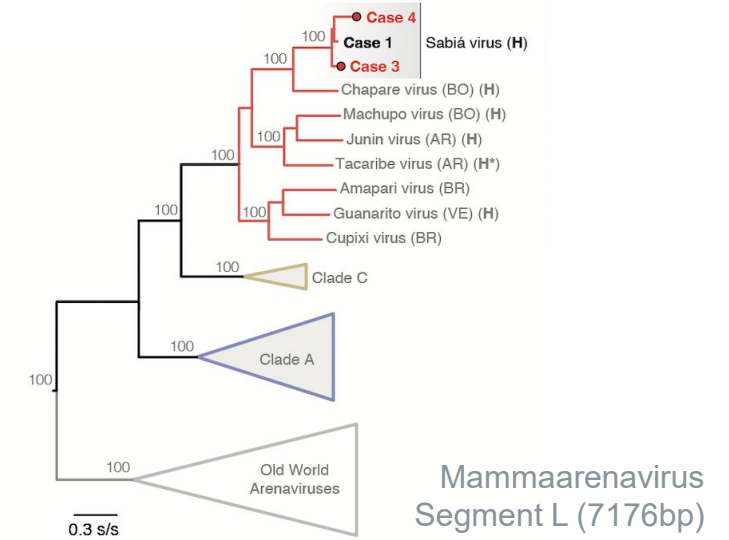
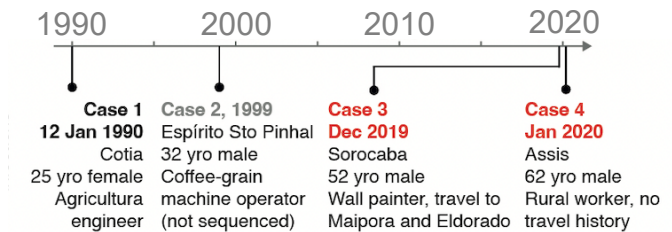
Pathogen-agnostic genomic surveillance for arboviral “Disease X”?

Current low-cost (USD 30-55), portable and sensitive untargeted sequencing protocols



YFV, DENV, ZIKV,
CHIKV, VEEV,
MADV, mpox,
ISV, HAV, **SABV**

Sabiá arenavirus in fatal cases suspected of YFV infections

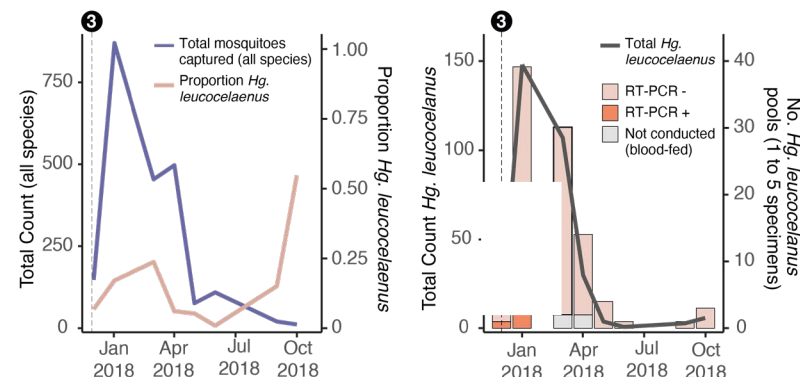
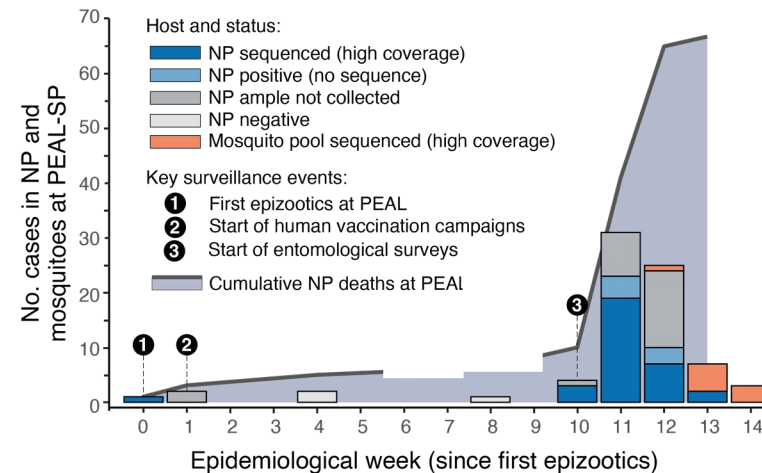
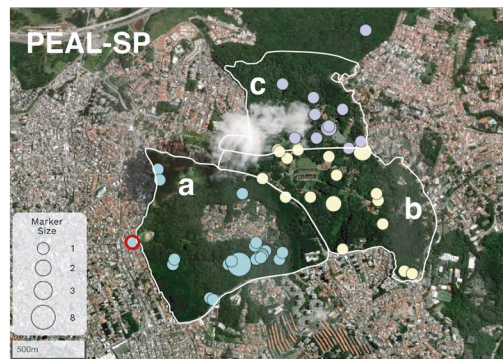
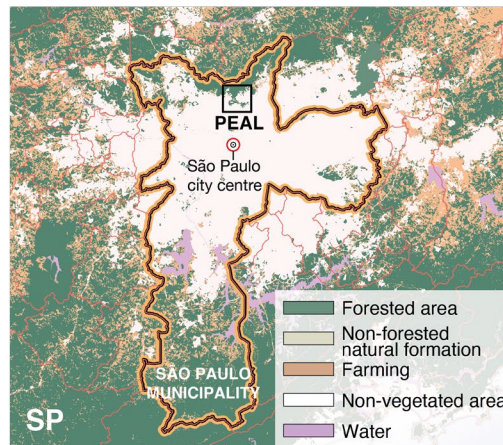


Claro et al. Wellcome Trust Open Research 2023
(Protocols.io: dx.doi.org/10.17504/protocols.io.7w5hpg6)

Claro et al. *in preparation* 2023
Data MRC-FAPESP CADDE Project (2019–23)

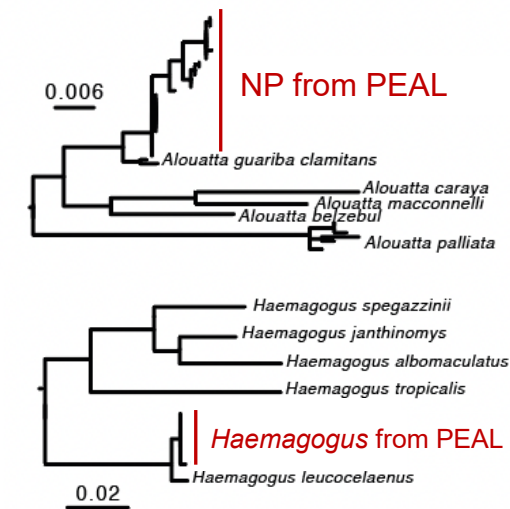
Pathogen-agnostic sequencing protocols identify multiple infections, host and vector species

Epizootic YFV outbreak in São Paulo city



Untargeted genome sequencing

- Simultaneous virus sequencing and identification of species mtDNA and insect-specific viruses (MO)
- *Alouatta guariba clamitans* co-infected with hepatovirus A genotype IA

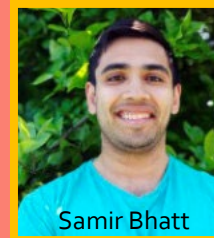
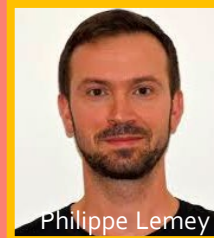
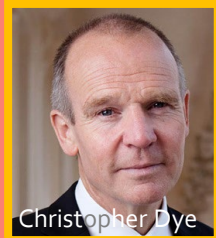
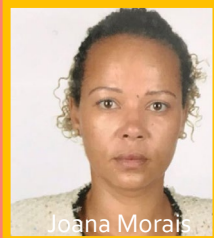
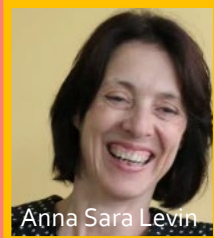
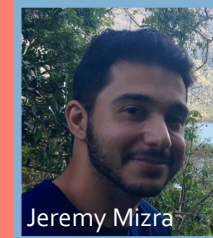
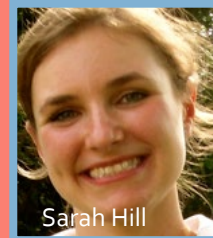
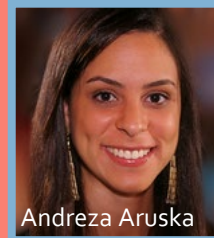
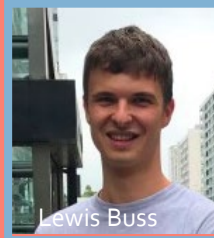
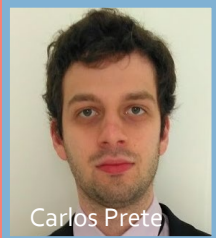
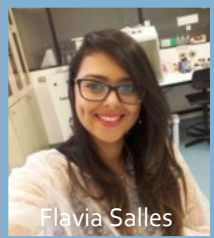
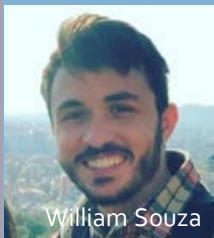
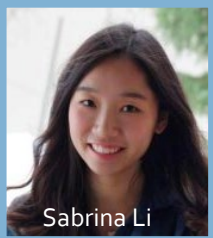
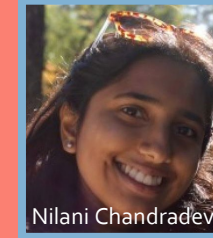
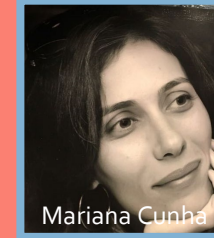
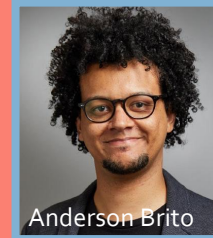
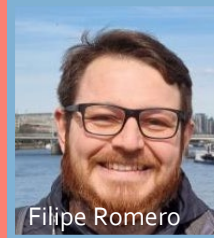
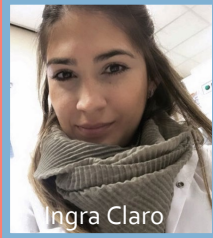


Deus, Claro, Faria, *in preparation* (2023)

Summary and future areas for research

- Optimizing the global detection of arboviral threats: climate, mobility, FOI-informed sampling strategies, focusing on undersampled locations with high transmission.
- Modelling frameworks integrating genomic with other surveillance data (epidemiological, serological, mobility, entomological, climate) to improve arboviral forecasts.
- Investment in low-cost and temperature insensitive diagnostic tools for known, and rare or yet unknown arboviral threats to improve arboviral surveillance *everywhere*.
- Collaborative genomic surveillance as a key component of arboviral risk assessments needs to be underlined by equity and inclusivity principles – and integrated across sectors, emergency cycles, and geographic levels.

Acknowledgments



Funding:
BILL & MELINDA GATES foundation



Institutions:
USP
Universidade de São Paulo

