Genomic and epidemiological monitoring of yellow fever virus transmission potential


The yellow fever virus (YFV) epidemic in Brazil is the largest in decades. The recent discovery of YFV in Brazilian Aedes species mosquitoes highlights a need to monitor the risk of reestablishment of urban YFV transmission in the Americas. We use a suite of epidemiological, spatial, and genomic approaches to characterize YFV transmission. We show that the age and sex distribution of human cases is characteristic of sylvatic transmission. Analysis of YFV cases combined with genomes generated locally reveals an early phase of sylvatic YFV transmission and spatial expansion toward previously YFV-free areas, followed by a rise in viral spillover to humans in late 2016. Our results establish a framework for monitoring YFV transmission in real time that will contribute to a global strategy to eliminate future YFV epidemics.

Yellow fever (YF) is responsible for 29,000 to 60,000 deaths annually in South America and Africa (1) and is the most severe mosquito-borne infection in the tropics (2). Despite the existence of an effective YF vaccine since 1937 (3), an estimated >400 million unvaccinated people live in areas at risk of infection (4). Yellow fever virus (YFV) is a member of the *Flaviviridae* family and is classified into four genotypes: East African, West African, South American I, and South American II (5–9). In the Americas, YFV transmission occurs mainly via the sylvatic cycle, in which nonhuman primates (NHPs) are infected by tree-dwelling mosquito vectors such as *Haemagogus* spp. and *Sabethes* spp. (10, 11). YFV transmission can also occur via an urban cycle, in which humans are infected by *Aedes* spp. mosquitoes that feed mostly on humans (12, 13).

Brazil has recently experienced its largest-recorded YF outbreak in decades, with 2043 confirmed cases and 676 deaths since December 2016 (supplementary text and fig. S1) (14). The last YF cases in Brazil attributed to an urban cycle were in Sena Madureira, in the northern state of Acre, in 1942 (15). An intensive eradication campaign eliminated *Aedes aegypti* and YF from Brazil in the 1950s (26), but the vector became reestablished in the 1970s and *Aedes* spp. mosquitoes are now abundant across most of Brazil (17). The consequences of a reignition of urban cycle transmission in Brazil would be serious, as an estimated 35 million people in areas at risk for YFV transmission in Brazil remain unvaccinated (4). New surveillance and analytical approaches are therefore needed to monitor this risk in real time.

**Yellow fever virus outbreak in Brazil, 2016–2017**

Between December 2016 and the end of June 2017, there were 777 polymerase chain reaction (PCR)–confirmed human cases of YF across 10 Brazilian states—mostly in Minas Gerais (MG) (60% of cases), followed by Espírito Santo (32%), Rio de Janeiro (3%), and São Paulo (3%) (18). The fatality ratio of severe YF cases was estimated at 33.6%, comparable to previous outbreaks (19, 20). Despite the exceptional magnitude and rapid expansion of the outbreak, little is known about its genomic epidemiology. Further, it is uncertain how the virus is spreading through space, as well as between humans and NHPs, and analytical insights into the contribution of the urban cycle to ongoing transmission are lacking.

To characterize the 2017 YFV outbreak in Brazil, we first compared time series of confirmed cases in humans (*n* = 683) and NHPs (*n* = 313) reported until October 2017 by public health institutes in MG, the epicenter of the outbreak (Fig. 1, A and B, and fig. S2). The time series are strongly associated (cross-correlation coefficient = 0.97; *P* < 0.001). Both peak in late January 2017, and we estimate that human cases lag behind those in NHPs by 4 days (table S1). NHP cases are geographically more dispersed.

---

1Department of Zoology, University of Oxford, Oxford, UK. 2Computational Epidemiology Lab, Boston Children’s Hospital, Boston, MA, USA. 3Department of Pediatrics, Harvard Medical School, Boston, MA, USA. 4Laboratório de Flavivirus, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil. 5Laboratório de Virologia Molecular, Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. 6Laboratório Central de Saúde Pública, Instituto Octávio Magalhães, FUNED, Belo Horizonte, Minas Gerais, Brazil. 7Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. 8Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. 9Department of Microbiology and Infection, University of Birmingham, Birmingham, UK. 10Department of Biostatistics, University of Oxford, Oxford, UK. 11The Global Health Network, University of Oxford, Oxford, UK. 12Department of Statistics, Harvard University, Cambridge, MA, USA. 13Malaria Atlas Project, Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. 14Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. 15Mathematical Modelling of Infectious Diseases and Center of Bioinformatics, Institut Pasteur, Paris, France. 16GIREN (UMR2300) Génomique Evolutive, Modélisation et Santé, Institut Pasteur, Paris, France. 17KwaZulu-Natal Research, Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa. 18Secretaria de Estado de Saúde de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. 19Núcleo de Doenças de Transmissão Vetorial, Instituto Adolfo Lutz, São Paulo, Brazil. 20Instituto de Medicina Tropical e Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. 21Retrovirology Laboratory, Federal University of São Paulo, São Paulo, Brazil. 22School of Medicine of ABC (FMABC), Clinical Immunology Laboratory, Santo André, São Paulo, Brazil. 23Coordenação de Vigilância Epidemiológica do Estado do Rio de Janeiro, Rio de Janeiro, Brazil. 24Departamento de Vigilância dos Doenças Transmissíveis da Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, DF, Brazil. 25Secretaria de Vigilância em Saúde, Coordenadoria Geral de Vigilância em Saúde Pública, Ministério da Saúde, Brasília-DF, Brazil. 26Organização Pan-Americana da Saúde/Organização Mundial da Saúde (OPAS/OMS), Brasília-DF, Brazil. 27Public Health England, National Infections Service, Porton Down, Salisbury, UK. 28NIHR HPRU in Emerging and Zoonotic Infections, Public Health England, London, UK. 29Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa. 30Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA, USA. 31Department of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA.

†Corresponding author. Email: nuno.faria@zoo.ox.ac.uk (N.R.F.); luiz.alcantara@ioc.fiocruz.br (L.C.J.A.); oliver.pybus@zoo.ox.ac.uk (O.G.P.)
in MG than human cases, which are more concentrated in the Teófilo Otoni and Manhuaçu municipalities (Fig. 1, D and E). Despite this, the numbers of human and NHP cases per municipality are positively correlated (Fig. 1F).

To establish whether human cases are acquired in proximity to potential sources of sylvatic infection, we estimated the distance between the municipality of residence of each human case and the nearest habitat of potential transmission, determined by using the enhanced vegetation index (EVI) (21) (supplementary materials). The average minimum distance between areas with EVI > 0.4 and the residence of confirmed human cases is only 5.3 km. In contrast, a randomly chosen resident of MG lives, on average, ≥51 km away from areas with EVI > 0.4. Similarly, human YFV cases reside, on average, 1.7 km from the nearest NHP case, whereas the mean minimum distance of a randomly chosen MG resident to the nearest NHP case is 39.1 km. This is consistent with YF infection risk being greatest for people who reside or work in forested areas where sylvatic transmission occurs. We find that most human cases (98.5%) were reported in municipalities with estimated YFV vaccination coverage above the 80% threshold recommended by the World Health Organization (WHO). On average, human cases would need to travel 65 km from their place of residence to reach an area where vaccination coverage is <80% (4).

**Fig. 1. Spatial and temporal epidemiology of YFV and CHIKV in Minas Gerais (MG).**

(A) Time series of human (H) YFV cases in MG (676 cases across 61 municipalities)—confirmed by serology, reverse transcription quantitative PCR (RT-qPCR), or virus isolation—during the first YFV epidemic wave (August 2016 to October 2017). (B) Same as in (A) but showing NHP YFV cases (313 cases across 90 municipalities) confirmed by RT-qPCR. (C) Same as in (A) but showing human CHIKV cases (3668 cases across 129 municipalities). (D) Geographic distribution of human YFV cases in MG. (E) Geographic distribution of NHP YFV cases in MG. Figure S3 shows the corresponding geographic distribution of CHIKV cases. (F) Association between the number of human and NHP cases in each municipality of MG (Pearson’s r = 0.62; P < 0.0001; nonparametric Spearman’s rank ρ = 0.32; P < 0.05).

**Risk of YFV urban transmission**

YFV was detected in *Aedes albopictus* mosquitoes caught in MG in January 2017 (22). Further, experiments suggest that *Aedes* spp. mosquitoes from southeast Brazil can transmit Brazilian YFV, although perhaps less effectively than vectors from elsewhere in the country (23, 24). It is therefore important to investigate whether YFV cases in MG occur where and when *Aedes* spp. vectors are active. To do so, we analyzed confirmed chikungunya virus (CHIKV) cases from MG (Fig. 1C).

CHIKV is transmitted by the urban mosquitoes *Ae. aegypti* and *Ae. albopictus* (25). There were 3755 confirmed CHIKV cases in MG during January 2015 to October 2017. The CHIKV epidemic in MG in 2017 began later and lasted longer than the YFV outbreak (Fig. 1C), consistent with the hypothesis that YFV and CHIKV in the region are transmitted by different vector species. However, 29 municipalities with human YFV cases also reported CHIKV cases (Fig. 1D and fig. S3), indicating that YFV is indeed present in municipalities with *Aedes* mosquitoes. The mean

---

**Fig. 2. Age and sex distribution of YFV cases in MG, 2016–2017.**

Red bars show the proportion of observed YFV cases in MG that occur in each age class, in (A) males and (B) females. These empirical distributions are different from those predicted under two models (M1, pale blue bars; M2, orange bars) of urban cycle transmission (see text for details).
YFV vaccination rate in districts with both YFV and CHIKV cases is 72.6% (range = 61 to 78%) (4). Thus, relatively high vaccination rates in the locations in MG where YF spillover to humans occurs, and potentially lower vector competence (23, 24), may ameliorate the risk of establishment of an urban YFV cycle in the state. However, adjacent urban regions (including São Paulo and Rio de Janeiro) have lower vaccination rates (4), receive tens of millions of visitors per year (26), and have recently experienced many human YFV cases (20). Thus, the possibility of sustained urban YFV transmission in southern Brazil and beyond necessitates continual virological and epidemiological monitoring.

We sought to establish a framework to evaluate routes of YFV transmission during an outbreak from the characteristics of infected individuals. Specifically, we assessed whether an outbreak is driven by sylvatic or urban transmission by comparing the age and sex distributions of observed YFV cases with those expected under two models of urban cycle transmission (supplementary materials). In model M1, age and sex variables in vaccination status but are equally exposed to YFV, a scenario that is typical of arboreal transmission (30). Under model M1, predicted incidence is highest among individuals aged 20 to 25 (Fig. 2). In model M2, we assume that the pattern of YFV exposure among age and sex classes follows that observed for CHIKV. The sex ratio of reported CHIKV cases in MG is 0.49 (33% of cases are male) (fig. S4). Under model M2, predicted incidence is highest in females aged >30. The discrepancy between the observed distribution and that predicted under the two urban cycle models indicates that the YF epidemic in MG is dominated by sylvatic transmission. This method shows that age- and sex-structured epidemiological data can be used to qualitatively evaluate the mode of YFV transmission during an outbreak.

Genomic surveillance of the Brazilian YFV outbreak
During a YF outbreak, it is important to undertake virological surveillance to (i) track epidemic origins and transmission hotspots, (ii) characterize genetic diversity to aid molecular diagnostics, (iii) detect viral mutations associated with disease severity, and (iv) exclude the possibility that human cases are caused by vaccine reversion. We generated 62 complete YF genomes from infected humans (n = 33) and NHPs (n = 29) from the most affected Brazilian states, including MG (n = 31), Espírito Santo (n = 8), Rio de Janeiro (n = 2), and Bahia (n = 1) (Fig. 3 and table S3). We also report two genomes from samples collected in 2003 during a previous YFV outbreak in MG from 2002 to 2003 (31). Genomes were generated in Brazil using a combination of methods (tables S5 to S7); half were generated in MG using a MiSeq portable YFV sequencing protocol adapted from (32) (tables S4 and S5). This protocol was made publicly available in May 2017 after the completion of a pilot sequencing experiments using a cultured vaccine strain (supplementary materials). Median genome coverages were similar for samples obtained from NHPs (99%; median cycle threshold value (Ct) = 11) and from human cases (99%; median Ct = 16) (tables S5 to S7).

To put the newly sequenced YFV genomes in a global context, we added our genomes to a pool of 61 publicly available genomes (33, 34). We developed and applied an automated online phylogenetic tool to identify and classify YFV gene sequences (also publicly available, see supplementary materials). Phylogenies estimated by
Fig. 4. Spatial and evolutionary dynamics of YFV outbreak. (A) Frequency of detection of YFV in NHPs in the Americas (50). Circle sizes represent the proportion of published studies (n = 15) that have detected YFV in each primate family and region: SA, South America (except Brazil); CA, Central America; CB, Caribbean; BR1, Brazil (before 2017); BR2, Brazil (this study). (B) Maximum clade credibility phylogeny inferred under a two-state (human and NHP) structured coalescent model. External node symbols denote sample type. Gray bars and labels indicate sample location (RJ, Rio de Janeiro; ES, Espírito Santo; BA, Bahia; others were sampled in MG). Internal nodes whose posterior state probabilities are >0.8 are annotated by circles. Node labels indicate posterior state probabilities for selected nodes. Internal branches are blue for NHPs and red for humans. Figure S8 shows a fully annotated tree. (C) Average number of YFV phylogenetic state transitions (from NHPs to humans) per month. Solid line, median estimate; shaded area, 95% BCI. (D) Expansion of the YFV epidemic wavefront estimated using a continuous phylogeographic approach (35). At each time point the plot shows the maximum spatial distance between phylogeny branches and the inferred location of outbreak origin. Solid line, median estimate; shaded area, 95% BCI. The dashed lines in (B) to (D) indicate when YF was declared a public health emergency in MG (13 January 2017). (E) Reconstructed spatiotemporal diffusion of the YFV outbreak. Phylogeny branches are arranged in space according the locations of phylogeny nodes (circles). Locations of external nodes are known, whereas those of internal nodes are inferred (44). DF, Distrito Federal; GO, Goiás; SP, São Paulo. Shaded regions represent 95% credible regions of internal nodes. Nodes and uncertainty regions are colored according to time.
this tool, along with maximum likelihood and Bayesian methods, consistently place the Brazilian outbreak clade in a single clade within the South America I (SA1) genotype with maximum statistical support (bootstrap = 100%; posterior probability > 0.99) (Fig. 3A and fig. S5).

The outgroup to the outbreak clade is strain BeH655417, a human case sampled in Alto Alegre, Roraima, north Brazil, in 2002. In contrast, isolates sampled during the previous outbreak in MG in 2003 are more distantly related to the outbreak clade within the SA1 genotype (Fig. 3A). Thus, the 2017 outbreak was more likely caused by a YFV strain introduced from an endemic area, possibly northern or center-west Brazil (35), than by the reemergence of a lineage that had persisted in MG. Although low sampling densities mean that this conclusion is provisional, similar scenarios have been suggested for previous Brazilian epizootics (36). The 14-year gap between the current outbreak and the date of the most closely related nonoutbreak strain agrees with the reported periodicity of YF outbreaks in northern Brazil (37), thought to be dictated by vector abundance and the accumulation of susceptible NHP hosts (19, 38).

At least seven humans from MG with PCR-confirmed YFV received a YF vaccine before the onset of symptoms. To test that these occurrences were caused by natural infection, and not by vaccine reactivation, we sequenced the YFV genomes from three of these cases (Fig. 3A and table S3). Our phylogenetic analysis clearly shows that these represent natural infections caused by the ongoing outbreak and are conclusively not derived from the 17DD vaccine strain (which belongs to the West African YFV genotype) (Fig. 3A and fig. S6).

Unifying YFV epidemiology and molecular evolution

Virus genomes are a valuable source of information about epidemic dynamics (39) but are rarely used to investigate specific YFV outbreaks in detail. Here we show how a suite of three analytical approaches, which combine genetic, epidemiological, and spatial data, can provide insights into YFV transmission.

First, we used a Bayesian method (40) to explore potential covariates of fluctuations in the effective population size of the YFV outbreak in 2017. After finding that genetic divergence in the outbreak clade accumulates over the time scale of sampling (Fig. 3B and fig. S6), albeit weakly, we sought to determine which epidemiological time series best describe trends in inferred YFV effective population size. We found that effective population size fluctuations of the YFV outbreak are well explained by the dynamics of both human and NHP YFV cases (inclusion probability: 0.37 for human cases and 0.63 for NHP cases) (table S8). These two YFV time series explain the genetic diversity dynamics of the ongoing outbreak 10^3 times more effectively than CHIKV incidence (inclusion probability < 0.001), which represents transmission by Aedes spp. vectors. One benefit of this approach is that epidemiological data contribute to estimation of the outbreak time scale. By incorporating YFV incidence data into evolutionary inference, we estimate the time of the most recent common ancestor (TMRCA) of the outbreak clade to be late July 2016 (95% Bayesian credible interval (BCI): March to November 2016) (Fig. 3C and fig. S7), consistent with the date of the first PCR-confirmed case of YFV in a NHP in MG (Jul 2016). The uncertainty around the TMRCA estimate is reduced by 30% when epidemiological and genomic data are combined, compared with genetic data alone (Fig. 3C).

Second, to better understand YFV transmission between humans and NHPs, we measured the movement of YFV lineages between the NHP reservoir and humans, using a phylogenetic structured coalescent model (41). Although previous studies have confirmed that YFV is circulating in five neotropical NHP families (Atelidae, Aotidae, Callitrichidae, Pitheciidae, and Cebidae) (Fig. 4A), thus far NHP YFV genomes during the 2017 outbreak have been recovered only from Alouatta spp. (family Cebidae) (39). In this analysis, we used the TMRCA estimate obtained above (Fig. 3C) to inform the phylogenetic time scale (Fig. 4B). All internal nodes in the outbreak phylogeny whose host state is well supported (posterior probability >0.8) are inferred to belong to the NHP population, consistent with an absence of urban transmission and in agreement with the large number of NHP cases reported in southeast Brazil (20). Despite this, we caution that hypotheses of human-to-human transmission linkage should not be tested directly using phylogenetic data alone, owing to the large undersampling of NHP infections. Notably, the structured coalescent approach reveals substantial changes in the frequency of NHP-to-human host transitions through time, rising from zero around November 2016 and peaking in February 2017 (Fig. 4C). This phylogenetic trend matches the time series of confirmed YFV cases in MG (Fig. 1, A and B), demonstrating that viral genomes, when analyzed using appropriate models, can be used to quantitatively track the dynamics of zoonosis during the course of an outbreak (42).

Third, we used a phylogenetic relaxed random walk approach to measure the outbreak’s spatial spread (43) (supplementary materials and methods and table S9). When projected through time and space and time and monitor the likelihood of sylvatic versus urban transmission. We hope that the toolkit introduced here will prove useful in guiding YF control in a resource-efficient manner.

REFERENCES AND NOTES
15. R. A. G. B. Consoli, R. Lourenço-de Oliveira, Principais Mosquitos de Importância Sanitária no Brasil (FioCruz, 1994).