Genomic and epidemiological monitoring of yellow fever virus transmission potential

Faria, N. R. [1¶*], Kraemer, M. U. G. [1,2*], Hill, S. C. [1*], Goes de Jesus, J. [3*], de Aguiar, R. S. [4*], Iani, F. C. M. [5,6*], Xavier, J. [3], Quick, J. [7], du Plessis, L. [1], Dellicour, S. [9], Thézé, J. [1], Carvalho, R. D. O. [8], Baele, G. [9], Wu, C.-H. [10], Silveira, P. P. [4], Arruda, M. B. [4], Pereira, M. A. [5], Pereira, G. C. [5], Lourenço, J. [1], Obolski, U. [1], Abade, L. [1,11], Vasylyeva, T. I. [1], Giovanetti, M. [3], Yi, D.[12], Weiss, D.J. [13], Wint, G. R. W. [1], Shearer, F. M. [13], Funk, S. [14], Nikolai, B. [15,16], Fonseca, V.[8,17], Adelino, T. E. R. [5], Oliveira, M. A. A. [5], Silva, M. V. F. [5], Sacchetto, L. [8], Figueiredo, P. O. [8], Rezende, I. M. [8], Mello, E. M. [8], Said, R. F. C. [18], Santos, D. A. [18], Ferraz, M. L. [18], Brito, M. G. [18], Santana, L. F. [18], Menezes, M. T. [19], Brindeiro, R. M. [19], Tanuri, A. [19], dos Santos, F. C. P. [20], Cunha, M. S. [20], Nogueira, J. S. [20], Rocco, I., M. [20], da Costa, A. C. [21], Komninakis, S. C. V. [22], Azevedo, V. [8], Chieppe, A. O. [23], Araujo, E. S. M. [6], Mendonça, M. C. L. [6], dos Santos, C. C. [6], dos Santos, C. D. [6], Mares-Guia, A. M. [6], Nogueira, R. M. R. [6], Sequeira, P. C. [6], Abreu, R. G. [24], Garcia, M. H. O. [24], Abreu, A. L. [25], Okumoto, O. [25], Kroon, E. G. [8], de Albuquerque, C. F. C. [26], Lewandowski, K. [27], Pullan, S. T. [27], Carroll, M. [28], de Oliveira, T. [17,29], Sabino, E. C. [21], Souza, R. P. [20], Suchard, M. A. [30,31], Lemey, P. [9], Trindade, G. S. [8], Drumond, B. P. [8], Filippis, A. M. B. [6], Loman, N. J. [7], Cauchemez, S. [15,16*], Alcantara, L. C. J. [3*¶], Pybus, O. G. [1*¶]

1. Department of Zoology, University of Oxford, OX1 3PS, UK
2. Computational Epidemiology Lab, Boston Children's Hospital, Boston, USA and Harvard Medical School, Boston, USA
3. Laboratório de Flavivírus, Instituto Oswaldo Cruz, FIOCRUZ Bahia, Salvador, Brazil
4. Laboratory of Molecular Virology, Departamento de Genética, Instituto de Biologia, Rio De Janeiro, Brazil
5. Laboratório Central de Saúde Pública, Instituto Octávio Magalhães, FUNED, Belo Horizonte, Minas Gerais, Brazil
6. Laboratório de Flavivírus, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil
7. Institute of Microbiology and Infection, University of Birmingham, B15 2TT, UK
8. Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
9. Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, Herestraat 49, 3000 Leuven, Belgium
10. Department of Statistics, University of Oxford, OX1 3LB, UK
11. The Global Health Network, University of Oxford, UK
12. Department of Statistics, Harvard University, Cambridge, MA, USA
13. Malaria Atlas Project, Big Data Institute, Nuffield Department of Medicine, University of Oxford, Roosevelt Drive, Oxford OX3 7FY, UK
14. London School of Tropical Medicine and Hygiene, London, UK
15. Mathematical Modelling of Infectious Diseases and Center of Bioinformatics, Institut Pasteur, Paris, France
16. CNRS UMR2000: Génomique évolutive, modélisation et santé, Institut Pasteur, Paris, France
17. KwaZulu-Natal Research, Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, South Africa
18. Secretaria de Estado de Saúde de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
19. Laboratory of Medical Virology, Departamento de Genética Instituto de Biologia, Rio de Janeiro, Brazil

20. Núcleo de Doenças de Transmissão Vetorial, Instituto Adolfo Lutz, São Paulo, Brazil
21. Instituto de Medicina Tropical e Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
22. Retrovirology Laboratory, Federal University of São Paulo, Brazil, School of Medicine of ABC (FMABC), Clinical Immunology Laboratory, Santo André, São Paulo, Brazil
23. Coordenação de Vigilância Epidemiológica do Estado do Rio de Janeiro, Brazil
24. Departamento de Vigilância das Doenças Transmissíveis da Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília-DF, Brazil
25. Secretaria de Vigilância em Saúde, Coordenação Geral de Laboratórios de Saúde Pública, Ministério da Saúde, Brasília-DF, Brazil
26. Organização Pan - Americana da Saúde - OPAS/OMS, Brasília-DF, Brazil
27. Public Health England, National Infections Service, Porton Down, UK
28. NIHR HPRU in Emerging and Zoonotic Infections, Public Health England, UK
29. Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa
30. Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA, USA
31. Department of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA

¶Correspondence to:

Nuno Rodrigues Faria
Department of Zoology, University of Oxford, UK
nuno.faria@zoo.ox.ac.uk

Luiz Carlos Junior Alcantara
Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Bahia, Brazil
lalcan@bahia.fiocruz.br

Oliver G. Pybus
Department of Zoology, University of Oxford, UK
oliver.pybus@zoo.ox.ac.uk

Abstract
The yellow fever virus (YFV) epidemic in Brazil is the largest in decades. The recent discovery of YFV in Brazilian *Aedes sp.* mosquitos highlights a need to monitor the risk of re-establishment 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 towards 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 29000–60000 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 classified into four genotypes: East African, West African, South American I, and South American II (*5-9*). YFV transmission occurs mainly via the "sylvatic cycle", in which non-human primates (NHP) are infected by infected tree-dwelling mosquitoes, 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 for decades, with 2043 confirmed cases and 676 deaths since Dec 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 (*16*) but the vector became re-established in the 1970s and *Aedes spp.* mosquitoes are now abundant across most of Brazil (*17*). The consequences of a re-ignition 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 urgently needed to monitor this risk in real-time.

Between Dec 2016 and the end of Jun 2017 there were 777 PCR-confirmed human cases across 10 Brazilian states, mostly in Minas Gerais (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, and between humans and NHPs, and analytical insights into the contribution of the urban cycle to ongoing transmission are lacking.

To characterise the 2017 YFV outbreak in Brazil, we first compare time series of confirmed cases in humans (n=683) and NHP (n=313) reported until October 2017 by public health institutes in Minas Gerais (MG), the epicentre of the outbreak (**Fig. 1A and B, 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 human cases lag those in NHP by 4 days (**Table S1**). NHP cases are geographically more dispersed in MG than human cases, which are more concentrated in Teófilo Otóni and Manhuaçu municipalities (**Fig. 1D and E**). Despite this, the number 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 estimate 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 5.3km. In contrast, a randomly chosen resident of MG lives on average ≥51km away from areas with EVI>0.4. Similarly, human YFV cases reside on average 1.7km from the nearest NHP case, whereas the mean minimum distance of a randomly chosen MG resident to the nearest NHP case is 39.1km. 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 notified 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 65km from their place of residence to reach an area where vaccination coverage is <80% (*4*).

YFV was detected in *Ae. albopictus* mosquitoes caught in MG in Jan 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 Brazil (*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 analysed 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 Jan 2015 to Oct 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 YFV vaccination rate in districts with both YFV and CHIKV cases is 72.6% (range=61-78%) (*4*). Thus, a combination of 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 an urban cycle in MG. For example, an individual's risk of acquiring YFV via the sylvatic cycle depends on their likelihood of travel to forested areas, typically highest among male adults (*27*). In contrast, under a urban cycle, we expect more uniform exposure across age- and sex-classes, similar to that observed for urban cases in Paraguay (*28*) and Nigeria (*29*).

The male-to-female sex ratio of reported YFV cases in MG is 5.7 (i.e., 85% of cases are male) and incidence is highest among males aged 40-49 (**Fig. 2**). We compare this distribution to that expected under two models of urban cycle transmission (**Supplementary Materials**). In model M1, age- and sex- classes vary in vaccination status but are equally exposed to YFV, a scenario that is typical of arboviral transmission (*30*). Under model M1, predicted cases are characterized by a sex ratio ~1 and incidence peaks among individuals aged 20-25 (**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.

During a YF outbreak it is important to undertake virological surveillance to (*i*) track epidemic origins and transmission hotspots, (*ii*) characterise 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 non-human primates (NHP) (*n*=29) from the most affected Brazilian states, including Minas Gerais (*n*=51), Espírito Santo (*n*=8), Rio de Janeiro (*n=2*), and Bahia (*n*=1) (**Fig. 3, Table S3**). We also report two new genomes from samples collected in 2003 during a previous YFV outbreak in MG, in 2002–2003 (*31*). Genomes were generated in Brazil using a combination of methods (**Table S3**); half were generated in Minas Gerais using a MinION portable YFV sequencing protocol adapted from (*32*) (**Tables S4 and S5**). This protocol was made publicly available in May 2017 following pilot sequencing experiments using a cultured vaccine strain (**Supplementary Materials)**. Median genome coverages were similar for samples obtained from NHP (99%; median Ct=11) and from human cases (99%; median Ct=16) (**Tables S5 and S6**).

To put the newly generated YFV genomes in a global context, we added our genomes to 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 this tool, and using maximum likelihood and Bayesian methods, consistently place the Brazilian outbreak strains in a single clade within the South America I (SA1) genotype with maximum statistical support (bootstrap=100%; posterior probability>0.99) (**Fig. 3A**; **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, local isolates sampled during the previous outbreak in MG in 2003 are more distantly related to the outbreak clade within the SA1 genotype (**Fig. 3**). 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 re-emergence of a lineage that had persisted in MG. Although low sampling densities mean 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 non-outbreak 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 7 PCR-confirmed YFV human cases from MG received a YF vaccine ≥3 days before onset of symptoms. To test that these infections were caused by natural infection, and not by vaccine reactivation, we sequenced the YFV genomes of three of these cases (**Fig. 3A, 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**).

Viral genomes are a valuable source of information about epidemic dynamics (e.g. (*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 confirming that genetic divergence in the outbreak clade accumulates over the timescale of sampling (**Fig. 3B, Fig. S6**), we tested which epidemiological time series best describe trends in inferred YFV effective population size. We find 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 better 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 timescale. 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-Jul 2016 (95% Bayesian Credible Interval, BCI: Mar-Nov 2016) (**Fig. 3C, Fig. S7**), consistent with the date of the first PCR-confirmed case of YFV in NHP in MG (Jul 2016). The uncertainty around the TMRCA estimate is reduced by 30% when epidemiological and genomic data are combined, compared to genetic data alone (**Fig. 3C**).

Second, in order to better understand YFV transmission between humans and NHP 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 (Aotidae, Atelidae, Callitrichidae, Pitheciidae, Cebidae; **Fig 4A**), thus far NHP YFV genomes during the 2017 outbreak have been recovered only from *Alouatta* spp. (family Cebidae) (*33*). In this analysis we used the TMRCA estimate obtained above (**Fig. 3C**) to inform the phylogenetic timescale (**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 argue that hypotheses of human-to-human transmission linkage should not be tested directly using phylogenetic data alone, due to the large undersampling of NHP infections. Notably, the structured coalescent approach reveals significant changes in the frequency of NHP-to-human host transitions through time, rising from zero around Nov 2016 and peaking in Feb 2017 (**Fig. 4C**). Remarkably, this phylogenetic trend matches the time series of confirmed YFV cases in MG (**Figs. 1A,B**), demonstrating that viral genomes, when analysed using appropriate models, can be used to quantitatively track the dynamics of zoonosis during the course of an outbreak (*42*).

Third, we measured the outbreak's spatial spread using a phylogenetic relaxed random walk approach (*43*) (**Supplementary Materials**; **Table S9**). When projected through space and time (**Figs. 4D-E**; **Movie S1**), the phylogeny shows a southerly dissemination of virus lineages from their inferred origin in MG towards densely populated areas, including Rio de Janeiro and São Paulo (where YF vaccination was not recommended until Jul 2017 and Jan 2018, respectively). We estimate virus lineages move on average 4.25 km/day (95% BCI: 2.64 to 10.76 km/day) (*44*). This velocity is similar when human YFV terminal branches are removed (5.3 km/day) and therefore most likely reflects YFV lineage movement within the sylvatic cycle and not the movement of asymptomatic infected humans. These rates are higher than expected given the distances typically travelled by NHPs in the region (*45*), and suggest the possibility YFV lineage movement may have been aided by human activity, e.g. transport of infected mosquitoes in vehicles (*46*) or hunting or illegal trade of NHPs in the Atlantic forest (*47, 48*). The epidemic wavefront (maximum distance of phylogeny branches from the inferred epidemic origin) expanded steadily between Aug 2016 and Feb 2017 at an estimated rate of ~3.3 km/day. Therefore by the time YF was declared a public health

emergency in MG (13 Jan 2017; dashed lines in **Figs. 4B-D**), the epidemic had already expanded ~600km (**Fig. 4D**) and caused >100 cases in both humans and NHP (**Fig. 1**). Notably, the first detection in humans in Dec 2016 was concomitant with the outbreak's spatial expansion phase (**Fig. 4D**) and the rise in estimated NHP-to-human zoonoses (**Fig. 4C**); both were likely driven by an increase in the abundance of sylvatic vectors. Thus the outbreak lineage appeared to circulate among NHP in a widening geographic area for several months before human cases were detected.

Epidemiological and genomic surveillance of human and animal populations at risk is crucial for the early detection and rapid containment of YFV transmission. The YFV epidemic in Brazil continues to unfold with an increase in cases since December 2017. Longitudinal studies of NHP are needed to understand how YFV lineages disseminate across South America between outbreaks, and how epizootics are determined by the dynamics of susceptible animals in the reservoir. To achieve the WHO's goal to eliminate yellow fever epidemics by 2026, YF surveillance demands a global, coordinated strategy. Our results and analyses show that rapid genomic surveillance of YFV, when integrated with epidemiological and spatial data, could help anticipate the risk of human YFV exposure through space and time and monitor the likelihood of sylvatic versus urban transmission. We hope that the toolkit introduced here will prove useful in guiding yellow fever control in a resource-efficient manner.

Acknowledgements
We thank FUNED-MG and the Brazilian YFV surveillance network for their essential contributions. N.R.F. thanks J. F. Drexler for sharing data and N. Trovão for discussions. We thank Oxford Nanopore Technologies for technical support. L.C.J.A. thanks QIAGEN for reagents and equipment. This work supported in part by CNPq # 400354/2016-0 and FAPESP # 2016/01735-2. N.R.F. is supported by a Sir Henry Dale Fellowship (204311/Z/16/Z), internal GCRF grant 005073, and John Fell Research Fund Grant 005166. This research received funding from the ERC (grant agreement 614725-PATHPHYLODYN) and from the Oxford Martin School. MUGK acknowledges funding from a Branco Weiss Fellowship, administered by ETH Zurich, a Training Grant from the National Institute of Child Health and Human Development (T32HD040128) and the National Library of Medicine of the National Institutes of Health (R01LM010812, R01LM011965). SD is funded by the Fonds Wetenschappelijk Onderzoek (FWO, Flanders, Belgium). GB acknowledges support from the Interne Fondsen KU Leuven / Internal Funds KU Leuven. ACdC is funded by FAPESP # 2017/00021-9. ACdC and ECS thank Illumina, Zymo Research, Sage Science and Promega for donation of reagents. BBN and SC are supported by the EU's Horizon 2020 Programme through ZIKAlliance (grant 734548), the Investissement d'Avenir program, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (grant ANR-10-LABX-62-IBEID), the Models of Infectious Disease Agent Study of the National Institute of General Medical Sciences, the AXA Research Fund, and the Association Robert Debré. PL and MAS acknowledge funding from the European Research Council (grant agreement 725422-ReservoirDOCS) and from the Wellcome Trust Collaborative Award 206298/Z/17/Z. PL acknowledges support from the Research Foundation, Flanders (Fonds voor Wetenschappelijk Onderzoek, Vlaanderen, G066215N, G0D5117N and G0B9317N).

Author contributions
N.R.F., L.C.J.A., S.C.H., A.M.B.F., M.U.G.K., S.C., and O.G.P. designed the study. S.C.H., J.J.G., R.S.de A., F.C.M.I., J.X., R.D.O.C., J.T., M.G., L.C.J.A. and N.R.F., undertook fieldwork. S.C.H., J.Q., J.J.G., A.C.da C., S.C.V.K., V.F., de O.T. undertook experiments. N.R.F., L. du P., J.T., S.D., G.B., O.G.P., C.-H.W., T.I.V. and P.L. performed genetic analyses. M.U.G.K., S.C., S.F., J.L, U.O., L.A., D.Y. and N.R.F. performed epidemiological and cartographic analyses. B.N., F.M.S. and N.R.F. performed historical YFV review. N.R.F., M.U.G.K., L.C.J.A., S.C. and O.G.P. wrote the manuscript. E.C.S, J.T., L. du P., R.P.S., P.L., de A.C.F.C., R.S.de A., A.M.B.F. edited the manuscript. All other authors were involved in collection, processing, sequencing and bioinformatics of samples and geographic data. All authors read and approved the contents of the manuscript.

Competing interests
N.J.L. and L.C.J.A. received free-of-charge reagents in support of the project from Oxford Nanopore Technologies.

Data and materials availability
Raw data, code, and analysis files are available on GitHub repository (https://github.com/arbospread/YFV-monitoring). See https://github.com/zibraproject/zika-pipeline/tree/master/schemes for MinION sequencing protocols. Genome sequences generated here are available under GenBank accession numbers MH018064-MH018115 and MH484423-MH484434.

Figure 1

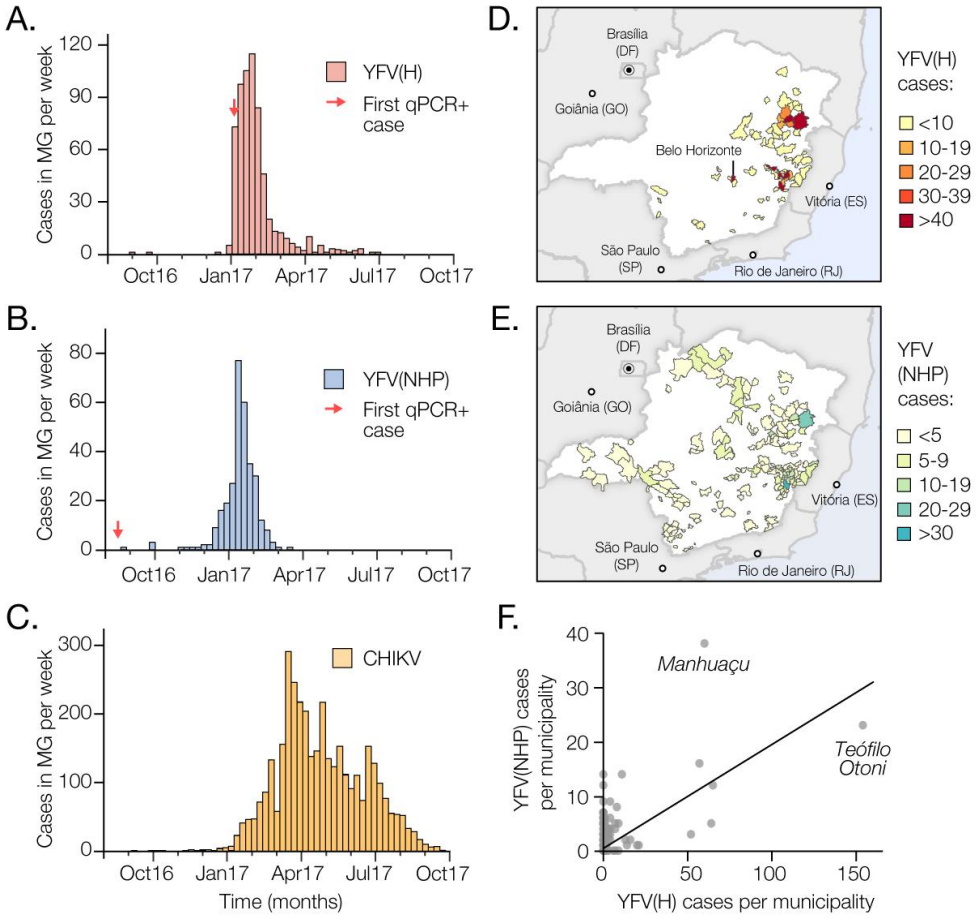

Fig. 1. Spatial and temporal epidemiology of YFV and CHIKV in Minas Gerais, MG. (A) Time series of human YFV cases in MG (676 cases across 61 municipalities) confirmed by serology, RT-qPCR or virus isolation during the first YFV epidemic wave (Aug 2016 to Oct 2017). (**B**) Same as panel A, but showing NHP YFV cases (313 cases across 90 municipalities), confirmed by RT-qPCR. (**C**) Same as panel A, but for 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. **Fig. S2** 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; non-parametric Spearman's rank ρ=0.32; p<0.05).

Figure 2

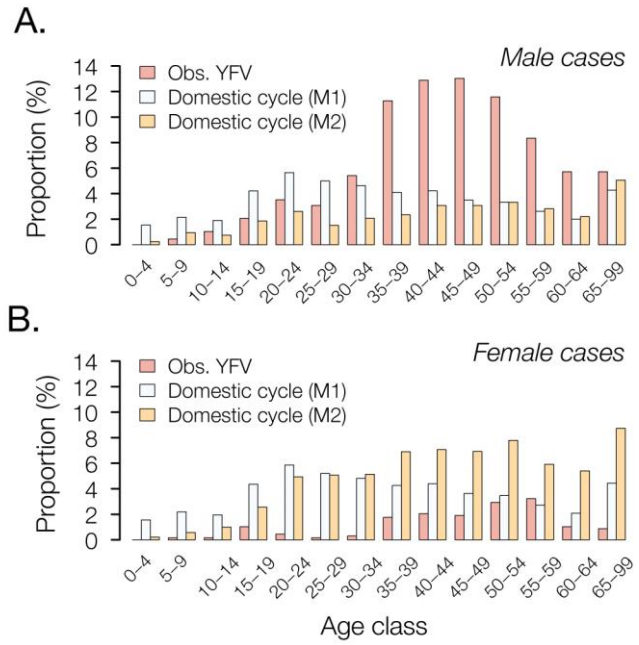

Fig. 2. **Age and sex distribution of YFV cases in Minas Gerais, 2016-2017.** Red bars show the proportion of observed YFV cases in Minas Gerais that occur in each age class, in (**A**) males and (**B**) females. These empirical distributions are different from those predicted under two models of urban cycle transmission (M1 = white bars and M2 = orange bars; see text for details).

Figure 3

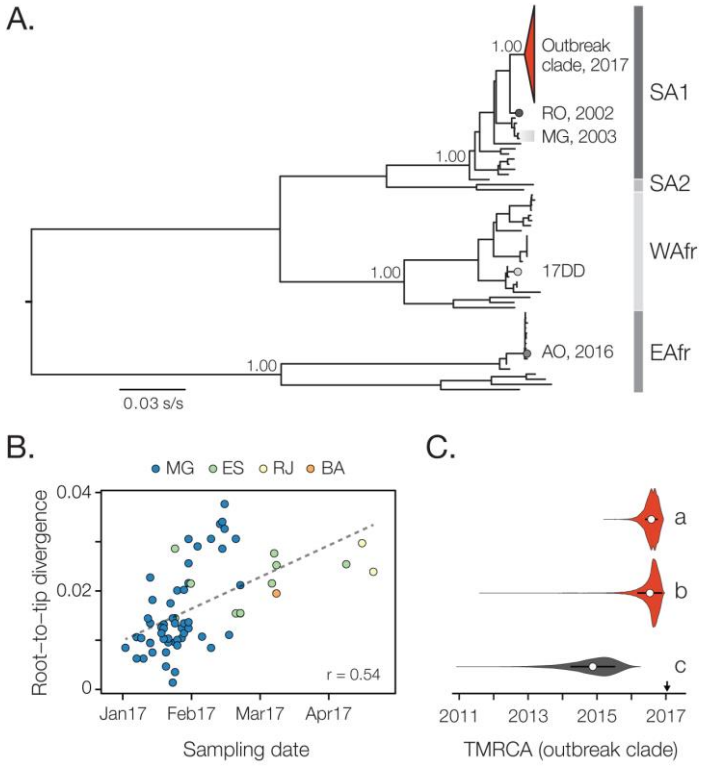

Fig. 3. **Molecular phylogenetics of the Brazilian YFV epidemic**. (**A**) Maximum likelihood phylogeny of complete YFV genomes showing the outbreak clade (red triangle) within the SA1 genotype (see **Figs. 4 and S6** for details). SA2, WAfr and EAfr indicate the South America II, West African, and East Africa genotypes, respectively. For clarity, five YFV strains introduced to Venezuela from Brazil (*49*) are not shown. The scale bar is in units of substitutions per site (s/s). Node labels indicate bootstrap support values. RO 2002 = strain BeH655417 from Roraima. MG 2003 = two strains from the previous YF outbreak in MG in 2003. 17DD = the vaccine strain used in Brazil. AO 2016 = YFV outbreak Angola in 2015-2016 (*13*). (**B**) Root-to-tip regression of sequence sampling date against genetic divergence from the root of the outbreak clade (see **Fig. S6A**). Sequences are coloured by sampling location. (**C**) Violin plots showing estimated posterior distributions (white circle=mean) of the time of the most common ancestor (TMRCA) of the outbreak clade. Estimates were obtained using two different datasets (grey=SA1 genotype, red=outbreak clade) and under different evolutionary models: a=uncorrelated lognormal relaxed clock (UCLN) model with a skygrid tree prior with covariates (specifically, the time series data shown in Figs. 1A-C; see **Fig. S7**); b=UCLN model with a skygrid tree prior without covariates; c=fixed local clock model (see **Supplementary Materials**).

Figure 4

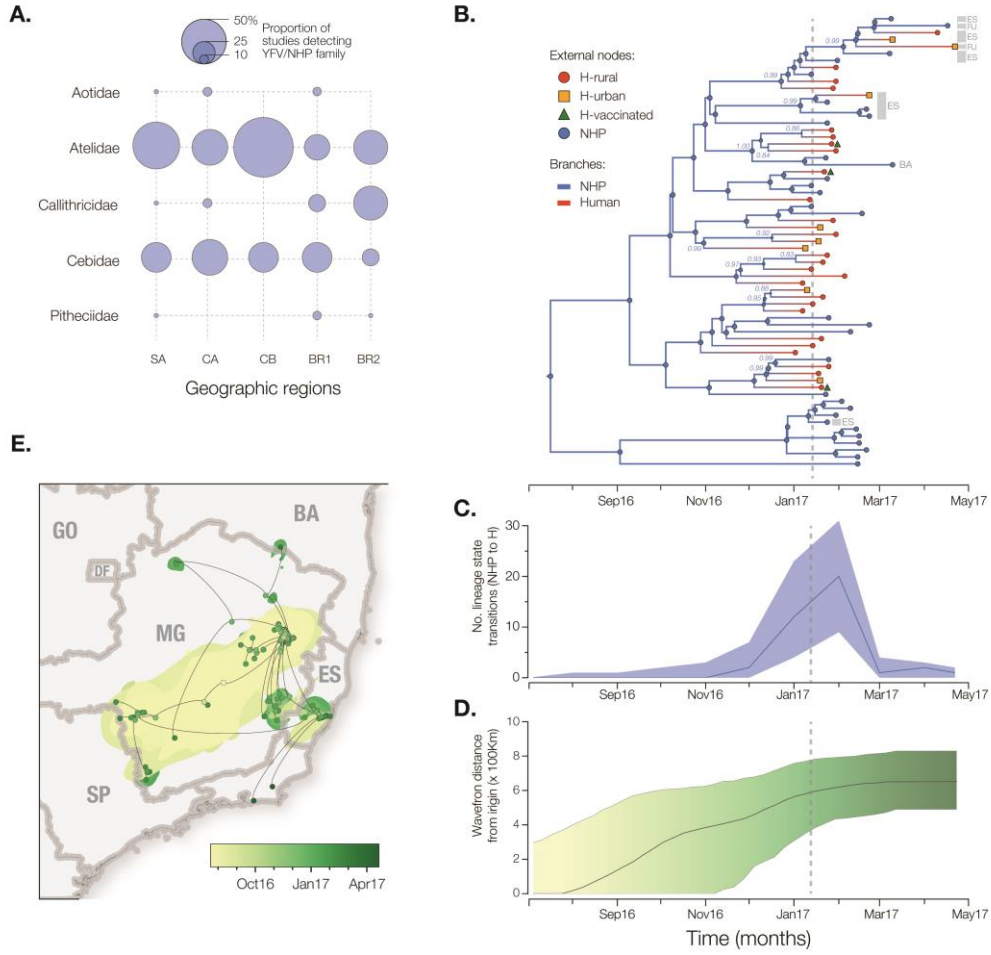

Fig. 4. Spatial and evolutionary dynamics of YFV outbreak. (**A**) Frequency of detection of YFV in non-human primates in the Americas (*50*). Circle sizes represent the proportion of published studies (*n*=15) that have detected YFV in each primate family and region. SA=South America (except Brazil), 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. Grey bars and labels to the 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 coloured blue for NHP, red for human. **Fig. S8** shows a fully annotated tree**.** (**C**). The average number of YFV phylogenetic state transitions (from NHP to human) 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 timepoint 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. **(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, whilst those of internal nodes are inferred (*44*). DF=Distrito Federal, GO=Goiás, SP=São Paulo. Shaded regions show 95% credible regions of internal nodes. Nodes and uncertainty regions are coloured according to time.

Supplementary Materials

Supplementary Text
Materials and Methods
Figs. S1 to S10
Tables S1 to S9
Caption for Movie S1
References (51–109) [Note: References 51-109 are only called out in Supplementary Materials]

Supplementary Materials for

Genomic and epidemiological monitoring of yellow fever virus transmission potential

Faria, N. R. [1¶*], Kraemer, M. U. G. [1,2*], Hill, S. C. [1*], Goes de Jesus, J. [3*], de Aguiar, R. S. [4*], Iani, F. C. M. [5,6*], Xavier, J. [3], Quick, J. [7], du Plessis, L. [1], Dellicour, S. [9], Thézé, J. [1], Carvalho, R. D. O. [8], Baele, G. [9], Wu, C.-H. [10], Silveira, P. P. [4], Arruda, M. B. [4], Pereira, M. A. [5], Pereira, G. C. [5], Lourenço, J. [1], Obolski, U. [1], Abade, L. [1,11], Vasylyeva, T. I. [1], Giovanetti, M. [3], Yi, D.[12], Weiss, D.J. [13], Wint, G. R. W. [1], Shearer, F. M. [13], Funk, S. [14], Nikolai, B. [15,16], Fonseca, V.[8,17], Adelino, T. E. R. [5], Oliveira, M. A. A. [5], Silva, M. V. F. [5], Sacchetto, L. [8], Figueiredo, P. O. [8], Rezende, I. M. [8], Mello, E. M. [8], Said, R. F. C. [18], Santos, D. A. [18], Ferraz, M. L. [18], Brito, M. G. [18], Santana, L. F. [18], Menezes, M. T. [19], Brindeiro, R. M. [19], Tanuri, A. [19], dos Santos, F. C. P. [20], Cunha, M. S. [20], Nogueira, J. S. [20], Rocco, I., M. [20], da Costa, A. C. [21], Komninakis, S. C. V. [22], Azevedo, V. [8], Chieppe, A. O. [23], Araujo, E. S. M. [6], Mendonça, M. C. L. [6], dos Santos, C. C. [6], dos Santos, C. D. [6], Mares-Guia, A. M. [6], Nogueira, R. M. R. [6], Sequeira, P. C. [6], Abreu, R. G. [24], Garcia, M. H. O. [24], Abreu, A. L. [25], Okumoto, O. [25], Kroon, E. G. [8], de Albuquerque, C. F. C. [26], Lewandowski, K. [27], Pullan, S. T. [27], Carroll, M. [28], de Oliveira, T. [17,29], Sabino, E. C. [21], Souza, R. P. [20], Suchard, M. A. [30,31], Lemey, P. [9], Trindade, G. S. [8], Drumond, B. P. [8], Filippis, A. M. B. [6], Loman, N. J. [7], Cauchemez, S. [15,16*], Alcantara, L. C. J. [3*¶], Pybus, O. G. [1*¶]

correspondence to: nuno.faria@zoo.ox.ac.uk (N.R.F.); lalcan@bahia.fiocruz.br (L.C.J.A.); oliver.pybus@zoo.ox.ac.uk (O.G.P.)

This PDF file includes:

Supplementary Text
Materials and Methods
Figs. S1 to S10
Tables S1 to S9
Caption for Movie S1

Other Supplementary Materials for this manuscript includes the following:

Movie S1
Nucleotide alignments, BEAST XML files, epidemiological data, available from the GitHub repository: https://github.com/arbospread/YFV-monitoring

Supplementary Text

Brief overview of YFV in Brazil and YFV urban outbreaks

In the Americas, the YFV jungle/sylvatic cycle involves virus transmission between jungle/sylvatic mosquitos and non-human primates (NHP). In contrast, an urban YFV transmission cycle (sometimes called the domestic cycle) involves virus transmission between humans by *Ae. aegypti* mosquitoes; a secondary role of *Ae. albopictus* in YFV urban epidemics remains controversial (*22-24, 51, 52*).

Several large urban YFV outbreaks have been reported in Brazil since the 17th century, in the states of Pernambuco (Recife, 1685), Bahia (1686-1692), Bahia (Salvador, 1849) and Rio de Janeiro (1849-1908) [reviewed in (*14*)]. After an absence of 20 years, YFV re-emerged in Rio de Janeiro (1928-1929) causing a recorded 738 cases and 478 deaths (*14*). This was the last large urban outbreak in the Americas. The last 3 suspected urban YFV cases were reported in Brazil in 1942, in the state of Acre, Sena Madureira municipality, in northern Brazil (*15*). In 1932, Soper et al. identified sylvatic YFV transmission in Vale do Canaã, Espírito Santo state, in southeast Brazil (*53*). Shortly thereafter, in 1937, Theiler & Smith developed the 17D YFV vaccine (*54*) and in the same year a field vaccine trial started in Brazil. In 1955 Brazil reported the elimination of the Ae. aegypti vector in the country through a national campaign using the insecticide dichloro-diphenyl-trichloroethane (DDT) (*15*). By 1970 Ae. aegypti had been eradicated from most of Latin America (*55*).

After an estimated 30 years, *Ae. aegypti* was re-introduced multiple times to Brazil (*56*) and the vector is now well-established in all municipalities of the country (*17*). Between 1950 to 2016 there were 1366 YFV cases in total across the country, linked to the sylvatic transmission cycle (**Fig. S1**).

Vectors involved in sylvatic YFV transmission in Brazil include mosquitos of the genera *Haemagogus* and *Sabethes*, especially *Hg. janthinomys, Hg. albomaculatus, Hg. leucocelaenus, Sa. Soperdi, Sa. chloropterus, Sa. Cyaneus,* and *Aedes serratus* (*11, 57-60*). Countrywide statistics from Brazil during 2000-2012 indicate that ~86% of sylvatic cases of YFV in humans occur in adult males. They acquire the infection in forested areas and typically have occupations such as agricultural workers, tourists, fishermen, students and truck drivers (*19, 61*). A high proportion of male YFV cases (76%) was observed in a large epidemic of sylvatic YFV in Peru in 1997 (*62*).

Urban yellow fever epidemics have been reported in several African countries and regions, including Nigeria-Lagos (1925-26), Ghana-Accra (1926-27 and 1937) and Gambia-Banjul (1937), Nigeria-Jos (1969), Angola-Luanda (1971), Nigeria-Azare et Bauchi (1979), Nigeria-Sud Ouest (1987), Angola-Luanda (1988), Nigeria-etat de Delta (1995), Nigeria-Kano (2000), Côte d'Ivoire-Abidjan (2001), Guinea-Conakry (2002), Dakar and Touba-Senegal (2002), Bob-Dioulasso-Burkina Faso (2004), and Luanda-Angola (2017) (*62, 63*). A large urban outbreak of YFV in western Nigeria in 1987-1988 exhibited a male:female case ratio of 1.4:1, with 71% of cases being <20 years old (*29*). However direct comparisons of YF epidemiological profiles between outbreaks in South America and Africa are not possible due to several factors including (i) strong heterogeneity in vaccination rates (*4*), (ii) different demographic profiles in each region, (iii) the presence of different YFV genotypes in each region (*36, 64, 65*) and (iv) the presence of other infections, such as hepatitis B, C, or E, that may confound case reporting (*66*).

In the Americas, evidence of urban YF transmission is more limited. Serology indicated a small YFV outbreak in the Department of Santa Cruz, Bolivia, in 1999 (*67*). However, of the 6 confirmed cases (all male), four had travelled outside the city. During 2008-2009, 28 human YFV cases in Paraguay were reported (*28*). Of these, 19 were notified in the departments of San Pedro and Caaguazú and were considered to be sylvatic cases (based on the travel history and occupation of infected patients). The remaining 9 cases occurred in an urban area of the Central Department, Municipio of San Lorenz, and lived <500m from each other: 4 were housewives, 4 were students, and 1 was a dependent. None had travelled during the two weeks prior to symptom onset. The average age of these nine individuals was 25 (range 11-39) and 44% (4/9) were male.

Materials and Methods

Description of epidemiological data from Minas Gerais, Brazil

The 853 municipalities in Minas Gerais are distributed among 28 regional health facilities (Gerências Regionais de Saúde). Local zoonotic surveillance units collected tissue samples from carcasses of NHPs, following the guidelines for epizootic surveillance of YFV (*36*). NHP samples were sent to Fundação Ezequiel Dias (FUNED), Flavivirus Reference Laboratory of the Brazilian Ministry of Health in Oswaldo Cruz Foundation of Rio de Janeiro, and Universidade Federal Minas Gerais for molecular diagnostics. For NHP samples, associated metadata (e.g. primate family, genus or species, date of capture, municipality of sample collection) were obtained directly from local teams or from the National Reportable Disease Information System - SINAN (Sistema de Informação de Agravos de Notificação; http://portalsinan.saude.gov.br/). To detect viral RNA in YFV suspected cases, post-mortem liver NHP samples were tested using YFV RT-qPCR.

For human cases, hospitals or health centers report cases and collect blood/serum samples or tissue samples (fatal cases only). For each tested case, metadata was retrieved from SINAN. Between Jan 2015 and Sep 2017, 2571 samples from patients residing in 212 municipalities of Minas Gerais (MG) with symptoms compatible with YFV infection were tested at (FUNED), located in Belo Horizonte, MG, southeast Brazil (**Fig. S2**), and the Flavivirus Reference Laboratory of the Brazilian Ministry of Health in Oswaldo Cruz Foundation of Rio de Janeiro, in Rio de Janeiro. During the same period, 9555 human samples from patients residing in 362 municipalities of MG were tested for CHIKV infection in the same laboratories (**Fig. S2**) following standard procedures (*68*). Although human and NHP cases are reported via different mechanisms, it is difficult to exclude the possibility that reporting variation among municipalities may play some role in the spatial association of cases. Following Pan American World Health Organization (PAHO) guidelines, YFV human samples were obtained ≤ 6 days after the onset of clinical symptoms, after which they were subjected to RT-qPCR. If samples were obtained >6 days after onset of disease, serological confirmation of YFV through IgM detection was performed. Due to potential cross-reactivity of serological assays, a positive YF serological test performed 6 days after onset of symptoms can indicate one of (i) recent YF infection, (ii) past YF vaccination, or (iii) infection with other circulating flaviviruses, such as Zika virus or dengue virus (*69*). Sex, age, municipality of residence, date of sample collection, and date of onset of symptoms were available for human YFV cases, YFV(H).

From the cases tested in MG between Jan 2015 and October 2017, the following datasets were prepared. Note that the same patient may have been tested with different tests, so the sample size (N) given below equals the total number of individuals with at least one positive test, which may be less than the sum of the number of positive tests:

- Dataset A: YFV(H) cases confirmed either by RT-qPCR (n=159) or by virus isolation (n = 62) or by IgM (n=478) at FUNED (n=683);
- Dataset B: YFV(NHP) cases confirmed by RT-qPCR in liver tissue analysed at the UFMG (N=314);
- Dataset C: CHIKV confirmed by RT-qPCR (n=144) or by IgM (n=3609) at FUNED (N=3755; no virus isolation was performed for CHIKV).
- Dataset D: YFV(H) confirmed either by RT-qPCR (n=159) or by virus isolation (n=62) at FUNED (n=221);
- Dataset E: CHIKV confirmed by RT-qPCR at FUNED (n=144).

The geographic distributions of YFV(H), YFV(NHP) and CHIKV cases are shown in **Fig. 1D**, **Fig. 1E** and **Fig. S3**, respectively. Note that these maps correspond, respectively, to datasets A, B and C described above.

To assess the association between the time series of YFV(NHP) and YFV(H) cases, we computed pairwise cross correlations among datasets A, B, and C, correcting for time lag and assuming that each dataset followed a unimodal distribution across time that covered a single epidemic wave of YFV. The correlations and corresponding P-values are shown in **Table S1**. Note that the time series are compared at the state, not local, level, making the analysis comparison more robust to sampling variation among municipalities.

Model of age-sex distributions under urban and sylvatic transmission cycles

To investigate whether human YF cases result from an urban or sylvatic transmission cycle we examined the age-sex distribution of human YF cases in MG between Dec 2016 (the date of first confirmed human YFV RT-qPCR case) and October 2017 (see **Fig. 1**). We define the "urban cycle" as YFV transmission between individuals mediated by anthropophilic *Ae. aegypti* mosquitoes. This may occur in urban, peri-urban or rural settings (hence the term 'domestic cycle' is sometimes used instead). We developed two models to predict the age-sex distribution of YFV cases expected under an urban cycle. In *model M1*, we assume that exposure to YFV in the urban cycle is independent of sex and age, approximately equivalent to the situation observed during urban YF outbreaks in Paraguay (*28*) and Nigeria (*29*). We reconstructed the resulting age-sex distribution from the underlying population age pyramid in MG (*70*) and from vaccine coverage per birth cohort (*4*). The expected number of individuals of age *a* and sex *s* that are at risk of YFV infection is then:

$$S_1(a,s) = N(a,s).p_U(a,s) \quad (1)$$

where $N(a,s)$ is the number of individuals of age a and sex s in the population and $p_U(a,s)$ is the proportion of unvaccinated individuals in that group. We assume that the proportion of vaccinated individuals is independent of sex in a given birth cohort. The expected proportion of YFV cases that are of age *a* and sex *s* is therefore:

(2)

$$P_1(a,s) = \frac{S_1(a,s)}{\sum_{a',s'} S_1(a',s')}$$

In *model M2*, we assume that, under a scenario of urban cycle transmission, the risk of exposure to YFV for a susceptible individual would be proportional to that seen for CHIKV cases (**Fig. S4**).

Let $C(a,s)$ denote the number of reported CHIKV cases of age *a* and sex *s*. For an individual of age *a* and sex *s*, the relative risk of being reported as a CHIKV case is defined as:

$$RR_{CHIK}(a,s) = \frac{C(a,s)}{N(a,s)} \quad (3)$$

Thus in model M2, the expected proportion of YFV cases that are of age *a* and sex *s* is:

$$P_2(a,s) = \frac{S_1(a,s).RR_{CHIK}(a,s)}{\sum_{a',s'} S_1(a',s').RR_{CHIK}(a',s')} \quad (4)$$

The age-sex distribution of dengue virus cases in Minas Gerais cannot be used as a proxy for that YFV in an urban cycle because, in areas where dengue has been circulating for years, the build-up of dengue immunity in the population will skew the age distribution of cases towards younger age groups. Further, three immunologically-distinct serotypes of dengue virus (1, 2 and 3) co-circulate in the region (*71*) but the serotypes of reported cases are not always known or tested. Additionally, it is not possible to use Zika case counts as a proxy because Zika virus molecular diagnoses were strongly biased towards women of gestational age, in accordance with Ministry of Health guidelines (*72*).

While we can propose simple and generic models to describe the expected age-sex distribution of cases in the YFV urban cycle, it is harder to predict what this distribution might look like in the sylvatic cycle because the propensity of individuals to travel to forested areas will likely depend on complex socio-economic and cultural factors that may exhibit substantial variation among regions, hence it is not feasible to construct a generic model of the age-sex distribution of YFV cases in the sylvatic cycle.

Estimating expected spatial distances to the source of YF infection

Many human YFV cases were reported in cities across the region and the travel history of most cases remains unknown. To assess the likelihood of a sylvatic transmission cycle scenario, in which most infections occur in forested areas, we calculate the average great circle distance between the place of residence of each human case and the nearest location with environmental conditions likely suitable for sylvatic transmission (*73*). We then compare this distance to that expected for typical (non-YFV infected) residents of Minas Gerais, estimated using high-resolution population datasets from 2015 (*74*). We used overall greenness of the environment [Enhanced Vegetation Index (EVI) (*75*)] to identify areas with environmental conditions suitable for sylvatic transmission. The EVI has been previously determined to be the best-fitting predictor of seasonal YFV transmission (*76*). Several thresholds of EVI for each municipality were considered: 0.33 (5%), 0.41 (50%), 0.46 (95%). Great-circle distances were calculated using the "rdist.earth" function in R (*77*). We also

calculated the distance to areas with known occurrences of positive non-human primates, again for both confirmed YF cases in humans and for typical residents of Minas Gerais.

Ethical statements for biological data

The project was supported by the Pan American World Health Organization (PAHO) and the Brazilian Ministry of Health as part of arboviral genomic surveillance efforts. Human samples were previously obtained for routine diagnostic purposes from persons visiting local clinics in Minas Gerais and Rio de Janeiro. Residual anonymized clinical diagnostic samples, with no or minimal risk to patients, were provided for research and surveillance purposes within the terms of Resolution 510/2016 of CONEP (Comissão Nacional de Ética em Pesquisa, Ministério da Saúde; National Ethical Committee for Research, Ministry of Health). We included 121 samples extracted at the Fundação Ezequiel Dias (FUNED), the main central public health laboratory in Minas Gerais (MG) (*sub-study I*). An additional 8 non-human primate (NHP) samples were extracted at the Universidade Federal de Minas Gerais (UFMG) and subsequently sent to FIOCRUZ Bahia for sequencing (**Table S2**). For the NHP samples from UFMG, authorization was obtained by the State Health Department of Minas Gerais and by the Ethics Animal Research Committee of Universidade Federal de Minas Gerais (License CEUA: 347/2017). Human samples were also processed at the Reference Centre for Arbovirus in Rio de Janeiro, the Laboratory of Flavivirus at FIOCRUZ Rio de Janeiro (*sub-study II*). Ethical approval for human samples was obtained from CEP/CAAE: 0026.0.009.000-07, with Institutional Review Board approval numbers 027/2007 and 1.920.256). Samples obtained from the Reference Centre for Arbovirus of São Paulo, Adolfo Lutz Institute (IAL) have been processed in agreement with routine surveillance activities from the Brazilian Ministry of Health and under the CEUA (Comitê de Ética de Uso de Animas em Pesquisa) registration number 02/2011.

Viral RNA isolation and sample processing

Human clinical samples included tissue and serum or plasma. In brief, viral RNA was extracted from 200 µL of clinical sample using QIAmp Viral RNA Minikit (Qiagen) according to the manufacturer's instructions but with several protocol changes. Tissue samples were first homogenised using a TissueLyser LT (Qiagen). A small piece of tissue (~2 mm diameter) was cut using a disposable scalpel and added to a 2 mL Eppendorf tube containing a 5 mm stainless steel bead (Qiagen). 560 µl AVL lysis buffer (Qiagen) was added to each tube and the sample was homogenised for 5 min at 50 Hz on a TissueLyser LT followed by a 10 min incubation at room temperature to lyse virions. Samples were centrifuged at 1,200*g* for 2 min to pellet cellular material, and 500 µL of supernatant was transferred to a new tube containing 500 µL of 100% EtOH. For serum or plasma samples, 200 µL of the sample was added to 560 µL of AVL lysis buffer (Qiagen) and left to incubate for 10 minutes before addition 560 µL of 100% EtOH. RNA extraction was subsequently completed on-site according the manufacturer's protocol for all sample types. To avoid contamination between samples due to the high number of virions, regular glove changes were conducted and parafilm was used to seal the gap between collection tubes and QIAamp Mini columns (Qiagen) during centrifugation. Batches always contained only primate or only human samples and a negative extraction control was processed with every batch. Human samples were linked to a record of clinical information such as date of onset of symptoms, date of sample collection, municipality, state of residence, age, sex, residence type and, when available, vaccine and travel history.

Real-time quantitative PCR (sub-studies I to III)

YFV reverse transcription quantitative real-time PCR (RT-qPCR) was performed on 121 samples using the Superscript III Platinum One-Step qRT-PCR System (Invitrogen) on a StepOnePlus Real-Time PCR machine (Applied Biosystems). The conserved YFV 5' non-coding region was targeted using the primers YFall15F (5' to 3': GCTAATTGAGGTGYATTGGTCTGC), YFall103R (5' to 3': CTGCTAATCGCTCAAMGAACG) and the probe YFall41 (5' to 3': FAM-ATCGAGTTGCTAGGCAATAAACAC-BHQ), based on the previously described Domingo's assay (*78*). Thermocyler conditions consisted of reverse transcription at 45ºC for 15 min, denaturation at 95 °C for 5 min, followed by 40 cycles of denaturation at 95 °C for 10s, and annealing and extension at 60°C for 40s. To check RNA isolation efficiency, we used RNase P as an endogenous positive control. Assays for RNase P used the primers RNaseP-F (5' to 3':AGATTTGGACCTGCGAGCG), RNaseP-R (GAGCGGCTGTCTCCACAAGT), and a probe (FAM-TTCTGACCTGAAGGCTCTGCGCG-BHQ1).

Validation of the sequencing primer scheme for MinION

Two candidate sequencing primer schemes were designed using Primal Scheme (http://primal.zibraproject.org) to amplify 500 bp or 1000 bp overlapping amplicons (*32*) of the complete genome of the YFV South American genotype 1, based on previous reports (*33*), with an overlap length of 75 bp between each neighbouring pair of primers (**Table S4**). The scheme was validated at Public Heath England, UK. cDNA synthesis and multiplex PCR were conducted on RNA extracts from a cultured vaccine strain YFV 17D. PCR products were cleaned using 0.8x Ampure XP (Beckman Coulter) bead cleanups, quantified, and pooled. Libraries for the MinION were constructed using the ligation sequencing kit 1D (SQK-LSK108) and native barcoding kit (EXP-NBD103). The library was sequenced on an R9.4 flow cell (FLO-MIN106). Basecalled reads were aligned to a YFV reference genome using bwa (GenBank accession number: JF912190). Given that the regions overlap, alternate amplicons are amplified in two separate PCR reactions. These are pooled and barcoded together (in previous studies (*32, 79*) these pools were barcoded separately, but this reduces the number of samples per flowcell by half). Mapping the reads to the reference genome showed the scheme provided good coverage across most of the coding-region of the genome. 95% of the genome had a depth of at least 379 reads, and 70% of the genome had a depth of at least 1941 reads. Both the 500bp and 1000bp PCRs with 40 cycles of PCR were tested in May 2017 at Minas Gerais (FUNED) on 7 samples of previously extracted RNA. Following PCR, quantitated dsDNA concentrations were higher for the 500 bp scheme than for the 1000 bp scheme, and therefore this scheme was chosen for all following assays (https://github.com/zibraproject/zika-pipeline/tree/master/schemes).

cDNA synthesis, library preparation and sequencing for MinION (sub-study I)

cDNA was reverse transcribed from viral RNA using the Protoscript II First Strand Sequencing kit (NEB) with random hexamer priming. Multiplex PCR was conducted using Q5 High Fidelity Hot-Start DNA Polymerase (New England Biolabs) and the 500bp sequencing primer scheme (**Table S4**). All samples were subjected to 32-40 cycles of PCR using the thermocycling conditions and reaction conditions described in Quick et al. (*32*). PCR products were purified using a 1x Ampure XP bead cleanup and concentrations were measured using a Qubit dsDNA High Sensitivity kit on a Qubit 3.0 fluorimeter

(ThermoFisher). Library preparation for the ONT MinION was conducted using Ligation Sequencing 1D (SQK-LSK108) and Native Barcoding kit (EXP-NBD103) according to the manufacturer's instructions, but with the changes detailed in (*32*). Amplified DNA and appropriate negative controls were sequenced in barcoded multiplexes of 6–12 samples per MinION run using FLO-MIN106 flow cells. Sequencing was performed without basecalling for 48 hours using MinKNOW. Consensus sequences for each barcoded sample were generated following previously published methods (*32*). Briefly, raw files were basecalled using Albacore, demultiplexed and trimmed using Porechop, and then mapped with *bwa* to a reference genome (GenBank Accession No. JF912190). Nanopolish variant calling was applied to the assembly to detect single nucleotide variants to the reference genome. Consensus sequences were generated; non-overlapped primer binding sites, and sites for which coverage was <20X were replaced with ambiguity code N. Sequencing statistics can be found in **Table S5**.

cDNA synthesis and sequencing using Ion Torrent (sub-study II)

cDNA synthesis was executed with SuperScript III Reverse Transcriptase (ThermoFisher Scientific) and random hexamers. Subsequently, YFV genome amplification was performed using Platinum Taq High Fidelity DNA Polymerase (ThermoFisher Scientific). PCR products were analyzed by electrophoresis gel and purified using PureLink Genomic DNA spin columns (ThermoFisher Scientific). YFV amplicons were quantified using a Qubit dsDNA High Sensitivity kit on the Qubit Fluorometric 2.0 (ThermoFisher Scientific). Sequencing libraries were prepared using 100ng of PCR products with an Ion Plus Fragment Library Kit (ThermoFisher Scientific), according to the manufacturer's instructions. For template amplification, emulsion PCR (emPCR) was performed using the Ion PGM Template OT2 kit and the Ion OneTouch 2 system (ThermoFisher Scientific). Ion Sphere particles (ISPs) were enriched using the Ion OneTouch ES (ThermoFisher Scientific). Enriched ISPs were sequenced using the Ion Torrent Personal Genome Machine Sequencer and the Ion PGM Hi-Q Sequencing kit (ThermoFisher Scientific), with the Ion 316 Chip. Data were collected for up to 8-9h. Reads were extracted, primer trimmed and mapped to a reference using Geneious R9 (9.1.7 version) (*80*). Briefly, primers were trimmed from each read (first 22 nt from 5' end). Reads were extracted based on amplicon size and coverage normalization was performed. Consensus genome sequences were generated by reference mapping to GenBank accession JF912190, and sites for which coverage was <3X were replaced with ambiguity code N. Sequencing statistics can be found in **Table S6**.

cDNA synthesis, library preparation and sequencing using Illumina (sub-study III)

Fourteen specimens were centrifuged at 20,000×g for 20 min and then filtered through a 0.45 µm filter (Merck Millipore, USA). The filtrates were treated with a mixture of nuclease enzymes to reduce background nucleic acids from the host cells and bacteria. RQ1 RNase-Free DNase (Promega Inc), DNase I (Zymo Research), Benzonase (Merck Millipore), RNase A (Zymo Research), RNase ONE (Promega Inc), Turbo DNase (Thermo Fisher) and 10X Turbo DNase buffer were added to the clarified supernatant and incubated at 37°C for 2h. Viral nucleic acids were extracted using a Maxwell 16 automated extractor (Promega Inc). Viral cDNA synthesis from extracted viral RNA/DNA was performed by using 50 pmol of a dodecamer of random primer in a reverse transcription reaction with AMV Reverse Transcriptase (Promega Inc) and RNasin Ribonuclease Inhibitor (Promega Inc). The 2nd strand cDNA synthesis was performed using DNA Polymerase I Large (Klenow) Fragment (Promega Inc), followed by the use of a Nextera XT Sample Preparation Kit (Illumina Inc) to

construct a DNA library with each sample identifiable using dual barcodes. For size selection, we used a Pippin Prep (Sage Science Inc) to select a 400 bp insert (range 200-600 bp). The library was deep-sequenced using the MiSeq Illumina platform with 2 x 300 bp paired ends. Paired-end reads of 2x300 bp generated by MiSeq were demultiplexed using the vendor software from Illumina. Demultiplexed Illumina reads were mapped on the JF912190 reference genome using bwa-mem program (*81*). The genome analysis toolkit (*82*) was used to perform variant calling and generate consensus sequences with a 3x minimum read depth coverage. Sequencing statistics can be found in **Table S7**.

Automated phylogenetic typing tool

We developed an tool that automatically classifies and accurately annotates YFV genome sequences, which is publicly available at https://www.genomedetective.com/app/typingtool/yellowfevervirus.

To build this YFV typing tool, we prepared two reference datasets that include publicly available sequences, one with whole-genomes (*n*=34, length=10,235 bp) and another with envelope gene sequences (*n*=34, length=1,443 bp). The accession numbers for each reference sequence of each genotype are as follows are as follows; for South American genotype 1: JF912190, JF912187, JF912188, JF912189, JF912180, JF912182, JF912185, JF912179, JF912184, JF912183, JF912186; for South American genotype 2: TVP17388, JF912181; for the West African genotype: AF094612, JX898871, JX898872, AY640589, JX898875, JX898874, JX898873, AY572535, AY603338, JX898868, JX898870, JX898876, JX898878, JX898880, X898877, JX898869, YFU54798; and for the East African genotype: AY968064, AY968065, DQ235229, JN620362. To validate the reference datasets, phylogenetic trees were constructed using maximum likelihood (ML) with a general time-reversible model and among-site rate variation modeled using a discretized gamma distribution (GTR + Γ_4), which was inferred as the best-fitting nucleotide substitution model in jModelTest (*83*). Trees were estimated using RAxML v8 (*84*) with 100 bootstrap replicates. All genotype clades are supported by bootstrap values of 100%, with the exception of the West-African genotype in the *env* tree, which is supported by a bootstrap score of 99%.

Classification of query sequences using the YFV subtyping tool involves two steps. The first step identifies the virus species using the basic local alignment search tool (*85*) that searches the RefSeq NCBI Reference sequence database that contains 7952 viruses reference genomes (*86*). The virus species is identified if the alignment score >400, which is the sum of identities minus gaps and mismatches. In addition, the tool also creates a codon alignment and identifies polymorphic sites and genetic diversity in the alignment, and aligns the query sequence to the NC_002031 curated reference sequence (*87*).

The second step involves the reconstruction of a phylogenetic tree with a reference dataset using neighbour-joining (**Fig. S5**). Statistical support for phylogenetic clustering of the query strain with the pre-defined reference genotypes using 1,000 bootstrap replicates. A query sequence is assigned to a particular genotype if clustering is supported by a bootstrap score >70%. The YFV typing tool accepts up to 2,000 sequences per submission and analyses each of sequence independently. At the end of the analysis, a phylogenetic tree is created that displays all query sequences and the reference dataset. A formatted report, estimated phylogenetic tree, and alignments can all be downloaded in multiple formats by the user.

Curation of whole-genome sequence datasets

We screened the GenBank database for published complete YFV genome sequences sampled worldwide using an in-house shell script. Subgenomic gene sequences from vector-borne flaviviruses typically contain insufficient genetic variation for reliable phylodynamic analysis on the time-scale of individual epidemics [e.g. (*43, 88, 89*)]. Therefore we retrieved all publicly available YFV sequences (n=756) from GenBank on 12th June 2018 and retained only complete or near complete YFV genomes (>8000 nt; n=200) (*90*). Sequences were collected from the Caribbean (Trinidad), East Africa (Ethiopia, Uganda and Sudan), Central Africa (Angola), East Asia (China), West Africa (Senegal, Ivory Coast, Ghana, Nigeria and Guinea-Bissau), and South America (Bolivia, Venezuela and Brazil). We then removed i) vaccine associated sequences (except the 17DD vaccine and Asibi strain, accession numbers: DQ100292 and AY640589), ii) sequences from patents, iii) duplicate entries, and iv) unpublished sequences from the Brazilian outbreak. Location and date of collection were retrieved from the original GenBank entry or from original publications. After curation, the "complete dataset" (n=125) contained complete or near complete YFV genomes sampled from 1927 to 2017. Sequences were aligned using MAFFT v.7 (*91*). Maximum likelihood (ML) phylogenetic trees were estimated using RAxML (*84*) under a GTR + Γ_4 nucleotide substitution model, as described previously. ML trees were estimated from i) the "complete dataset" described above (n=125) (**Fig. S6A**), ii) the "SA1 dataset", a subset of the complete dataset containing only South American genotype 1 (SA1) sequences (n=86) (**Fig. S6B**), and the iii) the "outbreak dataset", a subset of the SA1 dataset containing sequences only from the Brazilian outbreak (n=65) (**Fig. S6C**). For each of these three phylogenies, we conducted root-to-tip regressions of sequence sampling date against genetic divergence (*92*) (**Figs. S6A-C**). A preliminary analysis identified 7 anomalous sequences with low genome coverage (<26%), which may represent potential contaminants or mislabelled sequences (*92*) and were therefore excluded from subsequent analyses. All alignments were screened for recombination using the Phi-test available in SplitsTree v.4 (*93*); the null hypothesis of the absence of recombination could not be rejected ($P<0.05$) and lack of recombination was confirmed using the RDP4 package (*94*). The *outbreak dataset* comprises 65 genome isolates, 62 of which were generated by this study, 2 of which were published in (*33*) and 1 published in (*95*) (**Fig. S6C** and **Table S2**).

Bayesian skygrid with epidemiological covariates

Time-scaled phylogenetic trees were reconstructed using the Bayesian phylogenetic inference framework available in BEAST v1.8.4 (*96*). We used a probabilistic approach that combined sequence substitution over an unknown tree, calibrated to a real time scale using a molecular clock model. We used the HKY + Γ_4 nucleotide substitution model and a relaxed molecular clock model, with an underlying lognormal distribution of branch rates (*97*).

For the molecular clock model, we assumed that the outbreak clade exhibited a different clock rate to ancestral paraphyletic lineages, as observed in previous epidemics (*98*) and therefore we used a fixed local clock model (27) on the SA1 dataset; **Fig. S6B**). We also computed a Bayesian skygrid model (*99*) using the outbreak dataset alone (**Fig. S6C**), for which we specified 36 grids (i.e. the approximate number of epidemiological weeks spanned by the duration of the phylogeny). Further, we ran a Bayesian skygrid-based generalized linear model (*40*) with a streamlined prior specification in which effective population size through time is associated with a single covariate, chosen probabilistically from a set of possible covariates, while also accounting for phylogenetic uncertainty. In this analysis we investigated the following set of 3 covariates: i) log-linear YFV(H) case counts (dataset A), ii) log-linear YFV(NHP) case counts (dataset B) and iii) log-linear CHIKV human case

counts (dataset C). Specifically, for each grid point (epidemiological week) we include the log-transformed and standardized number of cases as described in the above section entitled **"**Description of epidemiological data from Minas Gerais, Brazil**"**. The association of each particular covariate with the effective population size dynamics of the outbreak is summarised by a coefficient and an inclusion probability (**Table S8**).

Distributions of the outbreak TMRCA obtained without and with covariates are shown in main text **Fig. 3c** (distributions a and b, respectively). Further, a comparison of the TMRCA estimates with and without YFV case count covariates are shown in **Fig. S7**.

Structured coalescent analyses

Viral lineage transitions among hosts were inferred using a structured coalescent model, as implemented in the MultiTypeTree v6.3.0 package (*41*) for BEAST v2.4.7 (*100*). The analysis was performed on the "*outbreak dataset*" only (see **Fig. S6**). The structured coalescent model also estimates time-scaled phylogenetic trees and state transition histories. It assumes a constant effective population size for each deme (i.e. human vs non-human host states, in this study) and asymmetric transition rates between demes. As in the other analyses above, we used an uncorrelated relaxed molecular clock model with a lognormal distribution prior on the branch rate parameters (*97*) and a HKY+Γ_4 nucleotide substitution model. Default priors were used for the nucleotide substitution model. A lognormal prior was placed on the molecular clock rate parameter, with mean equal to 0.001 substitutions per site per year (in real space) and standard deviation set to 1.

A lognormal prior with mean 0 and standard deviation 4 was used for the effective population sizes of demes and transition rates between demes (=host species states). To ensure that the phylogenetic timescale is well informed we placed a normally distributed prior with mean 0.751 years before the present (and standard deviation 0.18) on the time of the most recent common ancestor (TMRCA) of the tree. When estimating transition rates between host states, two independent runs of 200 million steps were computed, sampling parameters every 20,000 steps. The two chains were combined with LogCombiner, discarding 10% of each chain as burn-in and subsampling only half of the remaining states. Tracer v1.6.0 (http://tree.bio.ed.ac.uk/software/tracer/) was used to check the MCMC analysis for convergence. A maximum clade credibility tree with annotated branches was then generated in TreeAnnotator (**Fig. 4A**; the same tree with detailed taxa information is shown in **Fig. S8**). To recover the number host-switching events through time we counted the number of transitions between demes (host states) across monthly intervals for each tree in the posterior set of structured coalescent trees (migration histories). This count and its 95% HPD interval are shown in **Fig. 4C.**

To test sensitivity to the TMRCA prior used, the analysis was repeated (i) without a prior on the TMRCA and (ii) using a normally distributed prior with mean 0.731 years before the present and standard deviation 0.18, which corresponds to the TMRCA inferred under a standard skygrid model with covariates (see **Fig. 3C**). The inferred posterior distributions of the transition rates between human and NHP host states are shown in **Fig. S9**, where it can be seen that the TMRCA prior does not significantly affect the estimate transition rate dynamics. We also verified that this is the case for the migration histories (data not shown). In addition, the rate of host-transition events from NHP to human - our key result - always clearly deviates from the prior, whereas the reverse rate (from human to NHP) recovers the prior.

Phylogeographic inference in continuous space

Bayesian continuous phylogeographic analyses were performed on the "*outbreak dataset*" only using the skygrid with covariates as the coalescent tree prior (*40*). We first inferred the best fitting continuous diffusion process by performing (log) marginal likelihood estimation using generalized stepping-stone sampling (*101*) on a range of relaxed random walk models, as well as the time-homogeneous Brownian motion process (**Table S9**). Details of the stepping stone sampling approach were as follows: after an initial posterior exploration of 10 million iterations, we collected 1000 samples from each of the 51 power distributions, distributed according to a Beta(0.3,1.0) distribution and sampling at every 1000^{th} iteration. The log marginal likelihood estimates were highly consistent between independent runs in BEAST1.8.4 (*102*).

All the relaxed random walk models strongly outperformed the homogenous Brownian diffusion model. A model with Cauchy distributed diffusion rate variation among branches yielded the highest Bayes Factor (BF) against the homogenous diffusion model, indicating among-branch heterogeneity in branch velocity. The Cauchy model is most strongly preferred among all the relaxed random walk models (**Table S9**).

The Cauchy-distributed phylogeographic model selected above was then used to characterise the outbreak's spatio-temporal epidemic history (*44*). Posterior distributions under the Cauchy models were obtained using Markov chain Monte Carlo (MCMC) sampling as implemented in BEAST v 1.8.4 (*96*). The BEAGLE library v2.1.2 was used to accelerate computation (*103*). MCMC chains were run in triplicate for 250 million generations, sampling every 50,000 steps. MCMC performance was inspected for convergence and for sufficient sampling using Tracer v.1.6.

To summarise virus diffusion over time and space, 1000 post-burn-in phylogenies sampled at regular intervals from the posterior distribution were obtained. The branches of these phylogenies were extracted as vectors, each having start and end spatial coordinates, and start and end dates (i.e. branch duration) in decimal units (*43*). The R package "seraphim" was used to estimate statistics of spatial dissemination, such as dispersal velocity, diffusion coefficients, and evolution of the maximal wavefront distance from epidemic origin (*104, 105*), as well as generating monthly graphical representations of the inferred spatio-temporal spread process (**Movie S1**) using the "spreadGraphic" function (*106*).

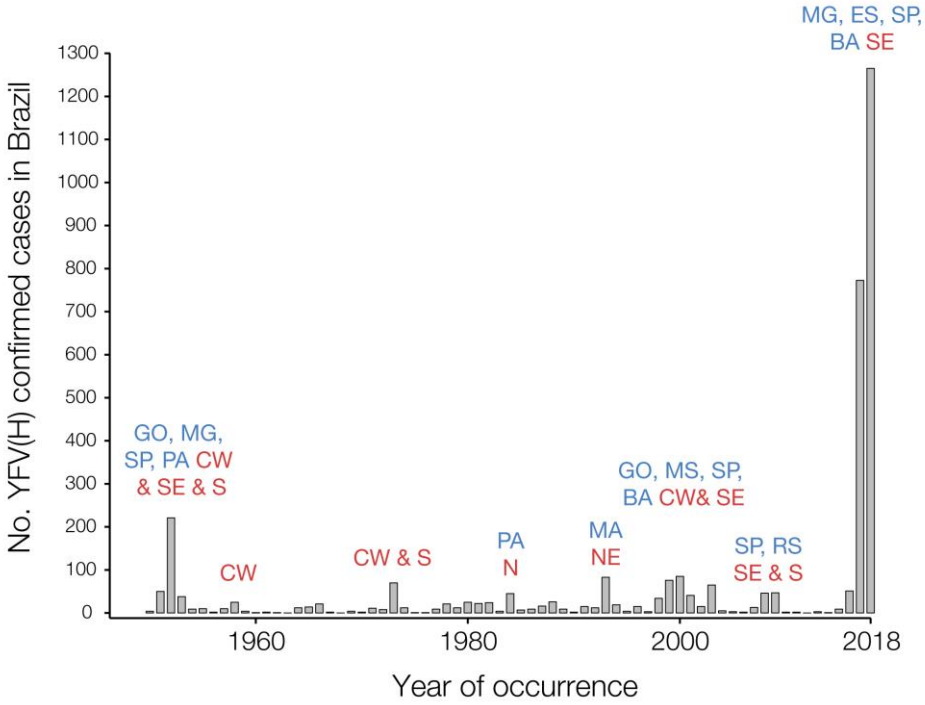

Fig. S1.

Number of YF sylvatic cases in Brazil between 1950 and 2018. Data from 1950 to 1980 are from (*107*); data from 1980 to 2015 are from Brazilian Ministry of Health (MoH) reports (*35*); data under "2017" are from the Brazilian MoH reports from 1st Dec 2016 to 1st July 2017 (*18*); data under "2018" are from 1st July 2017 to 16/05/2018 (*109*). Federal states (in blue): GO=Goiás, MG=Minas Gerais, SP=São Paulo, PA=Pará, MA=Maranhão, BA=Bahia, MS=Mato Grosso do Sul, RS=Rio Grande do Sul. Regions (in red): CW=Centre-West, SE=South-east, S=South, N=North, S=South.

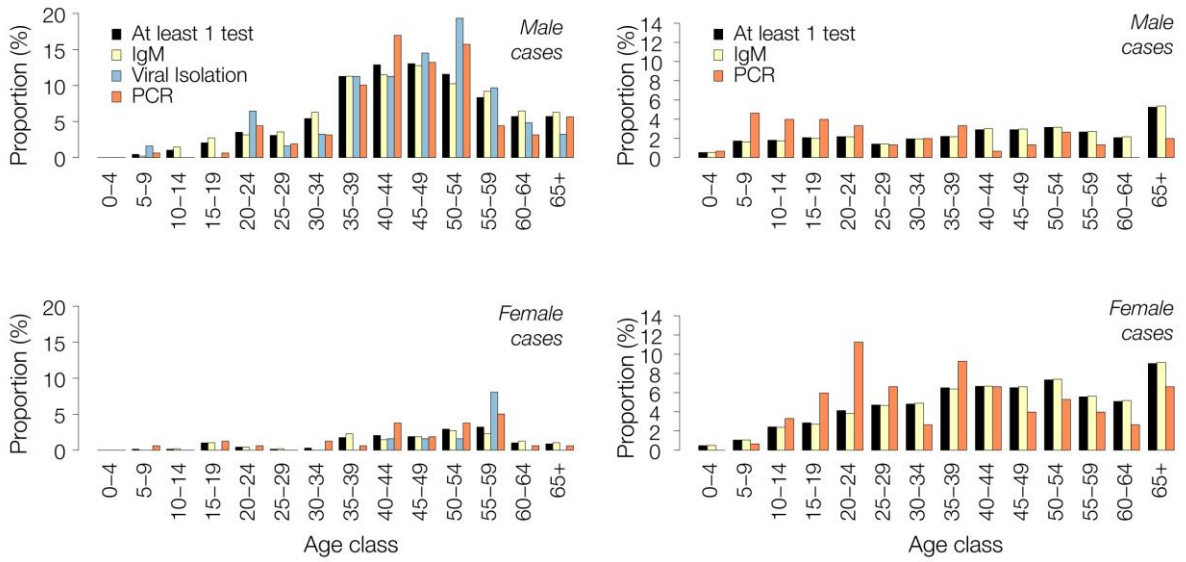

Fig. S2.

Sensitivity of diagnostics for human YFV (left) and CHIKV (right) samples in Minas Gerais. The proportion of samples positive for at least one test (black bars) in each age- and sex-class is shown, together with the proportion of positives for each test separately (IgM=yellow bars, viral isolation=blue bars, RT-qPCR=red bars).

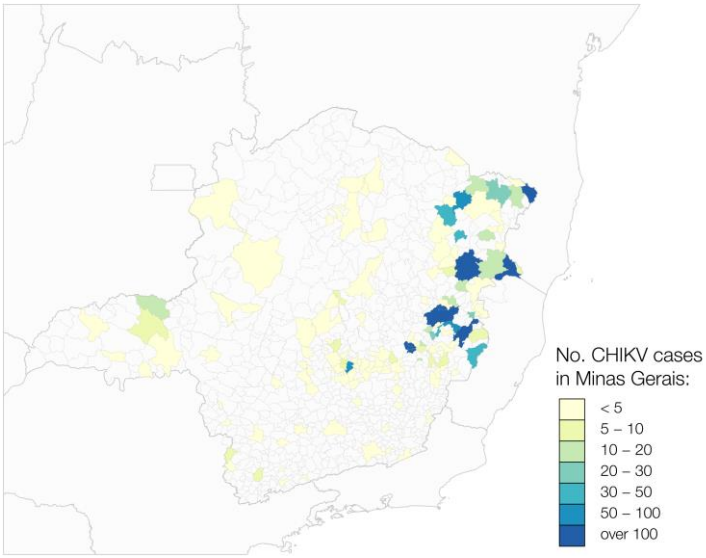

Fig. S3.

Geographic distribution of CHIK cases in Minas Gerais. The figure shows cases confirmed by serology, RT-qPCR, or virus isolation in Minas Gerais from Jan 2015 to October 2017 (corresponding to dataset C).

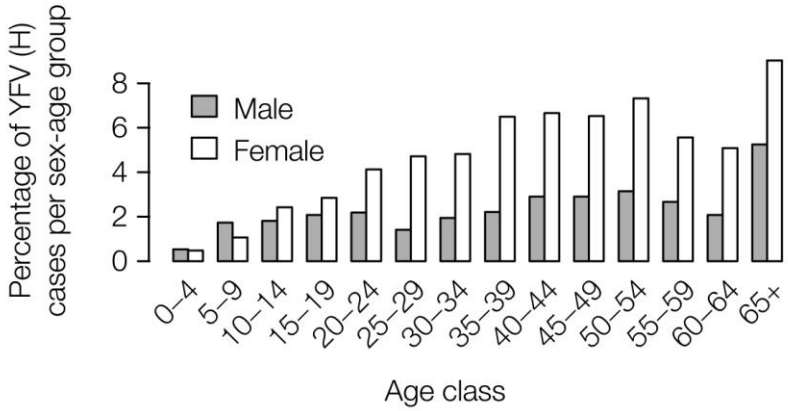

Fig. S4.

Age-sex distribution of reported CHIKV cases in Minas Gerais. The y-axis shows the percentage of CHIV cases belonging to each age- and sex-class that were confirmed by at least one diagnostic test in Minas Gerais, between Jan 2015 and October 2017 (dataset C).

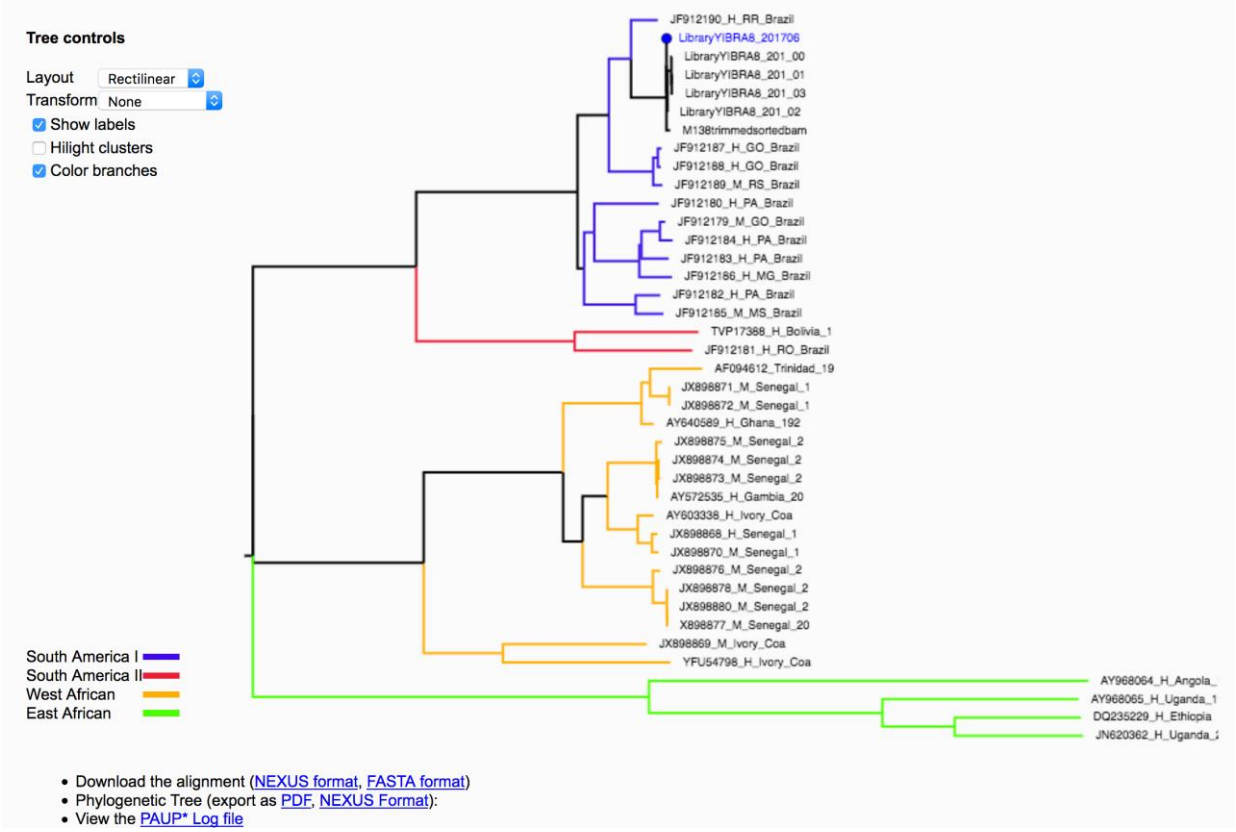

Fig. S5.

Illustration of the output of the online YFV classification tool. The figure shows the ML phylogeny of 6 target sequences analyzed by the tool. The output also provides a link to genome coverage and a more detailed report. The reference dataset is colored according to genotype.

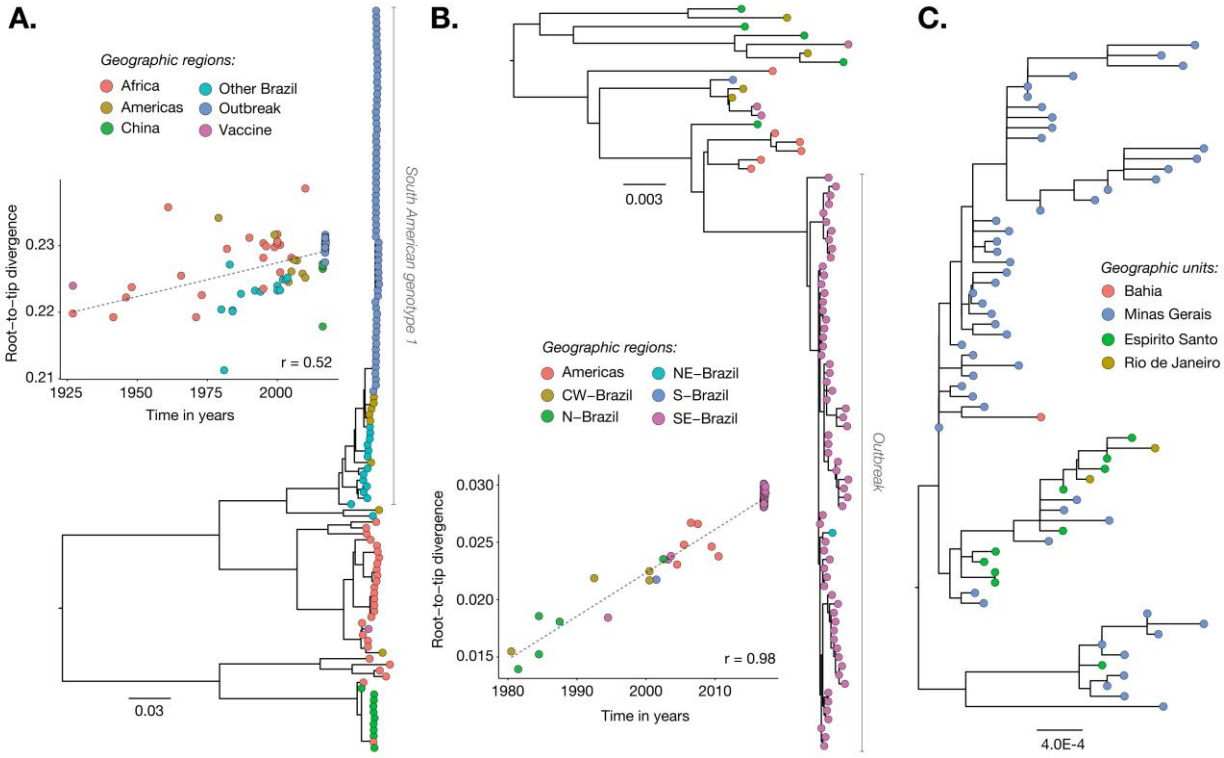

Fig. S6.

Maximum likelihood (ML) phylogenies and temporal signal of YFV near-complete genome sequence datasets. (A) The "complete dataset" (n=125). The figure shows the ML phylogeny for this alignment and its corresponding temporal signal (regression between sampling dates and genetic divergence). Branches are in nucleotide substitutions per site (scale bar). Tips are coloured according to geographic region of sample collection. The position of the SA1 genotype is shown to the right of the tree. (B) The "SA1 dataset" (n=86) that contains South American genotype 1 sequences only. The position of the outbreak clade is shown to the right of the tree. See panel A for further explanation. (C) The "outbreak dataset" (n=65). The corresponding regression analysis for this tree is provided in **Fig. 3**. See panel A for further explanation.

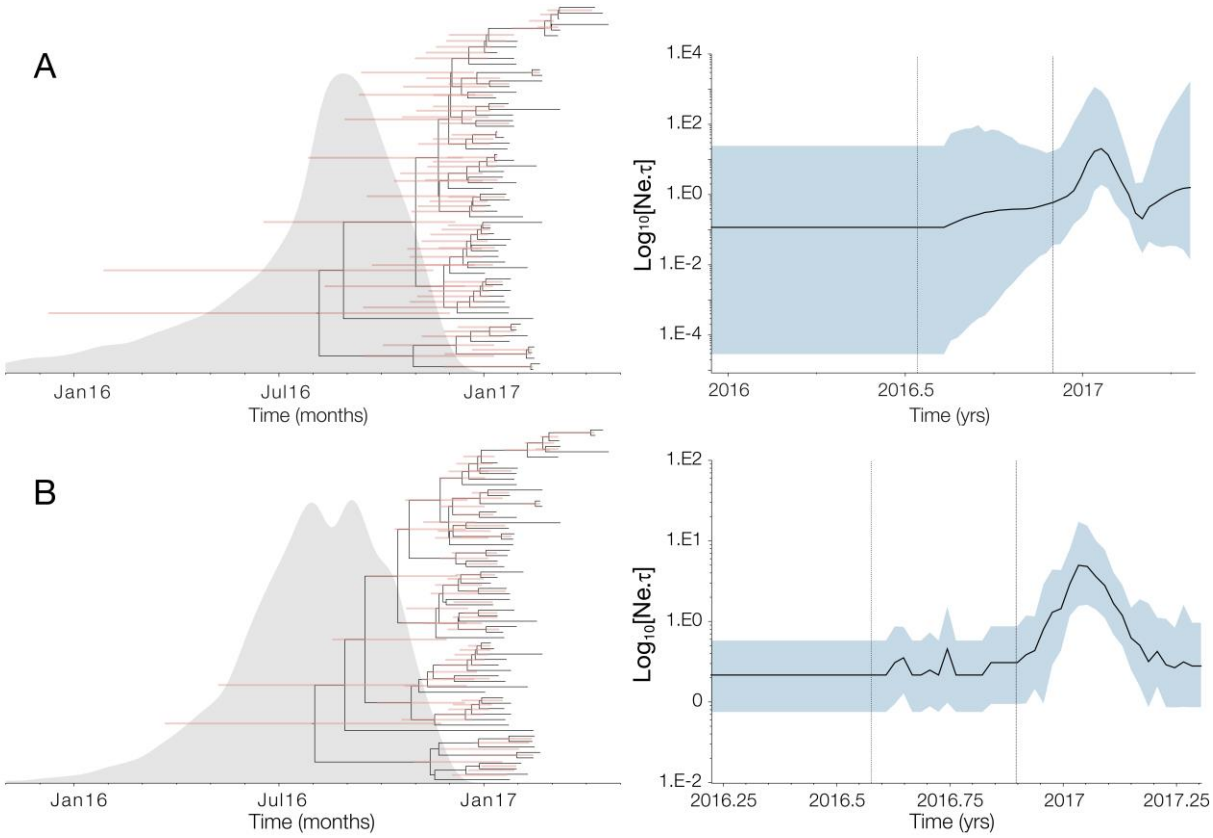

Fig. S7.

Combining virus phylogeny and epidemic time series. The left hand column shows maximum clade credibility trees (MCCs) generated in BEAST, together with their corresponding posterior distributions of the time of most recent common ancestor (TMRCA; grey) of the outbreak clade. Thin horizontal red bars indicate uncertainty in estimates of divergence times of internal nodes in each tree. The right hand column shows the corresponding Bayesian skygrid effective population size estimate. Black dashed line on the left = median estimate, black dashed line on the right = lower 95% highest posterior density (HPD) credible interval TMRCA estimate, and blue shaded area = 95% HPD credible interval. The top row (A) shows the results obtained using the standard skygrid model whilst the bottom row (B) shows the results obtained using skygrid model with covariates (B). Addition of the epidemiological time series data in (B) reduces the statistical uncertainty of the estimated TMRCA parameter by ~30%.

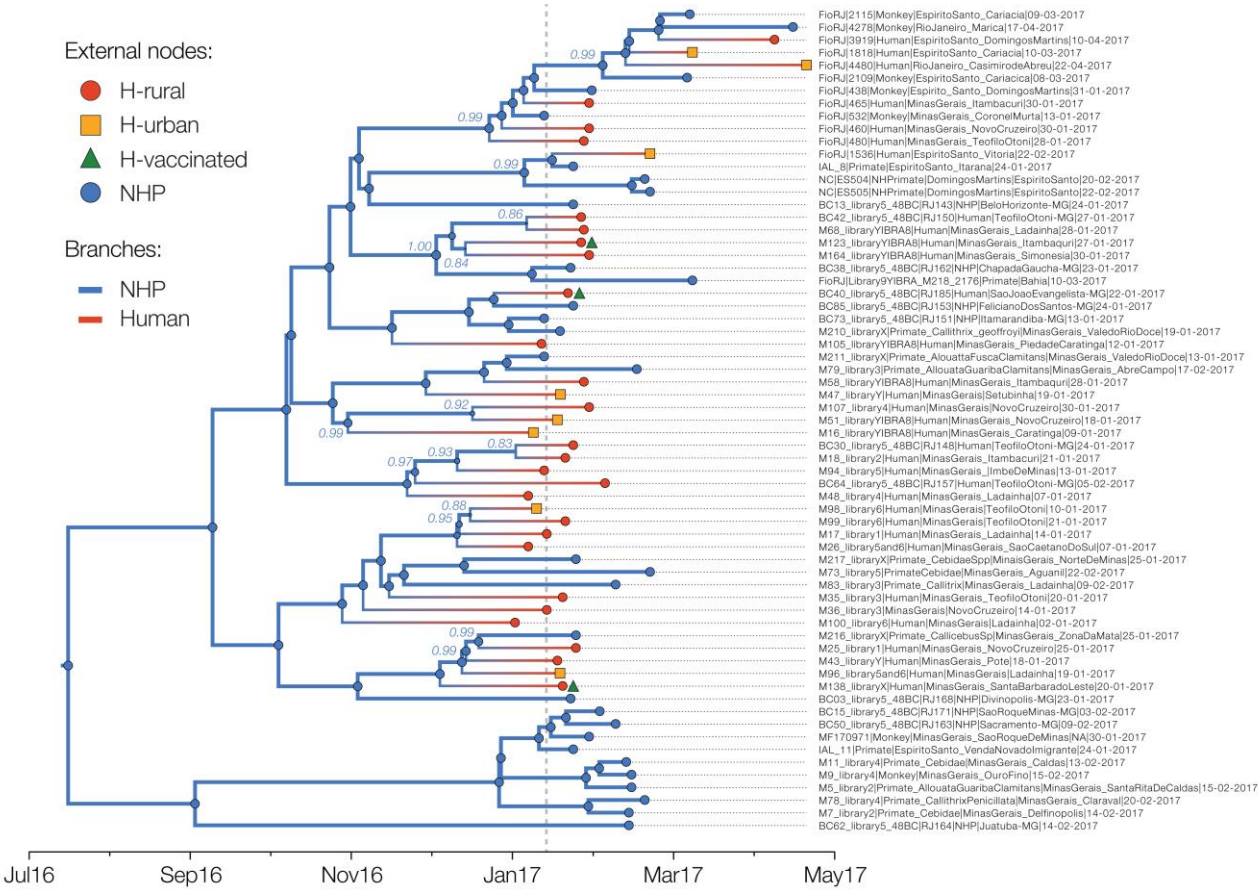

Fig. S8.

Typed maximum clade credibility tree (corresponding to Fig. 4A). The node labels indicate the posterior probability of the most likely host state for each internal node, inferred by the MultiTypeTree package (*41*). This representation does not include information on individual transition events between host states.

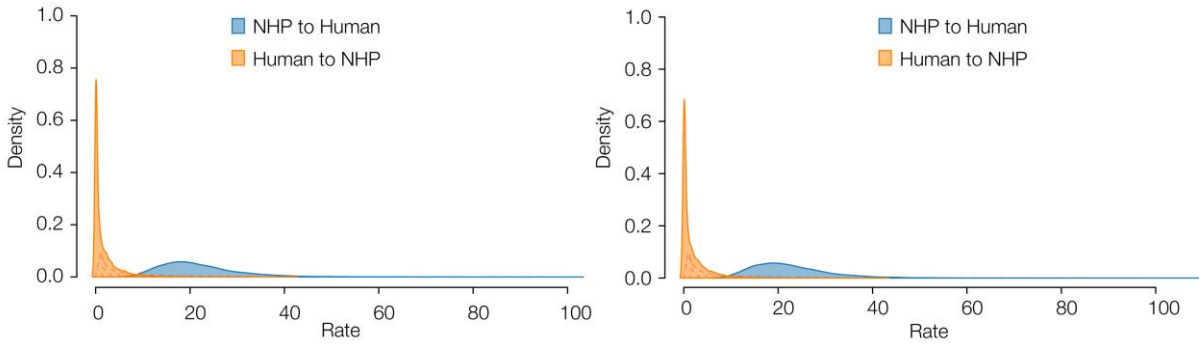

Fig. S9.

Estimated posterior distributions of host state transition rates of the structured coalescent model under different priors. From left to right, (i) using no prior on the TMRCA, and (ii) using a normal distribution with mean 0.73 years before the present and standard deviation of 0.18. The prior distribution used for the migration rates is shaded with dashed red lines.

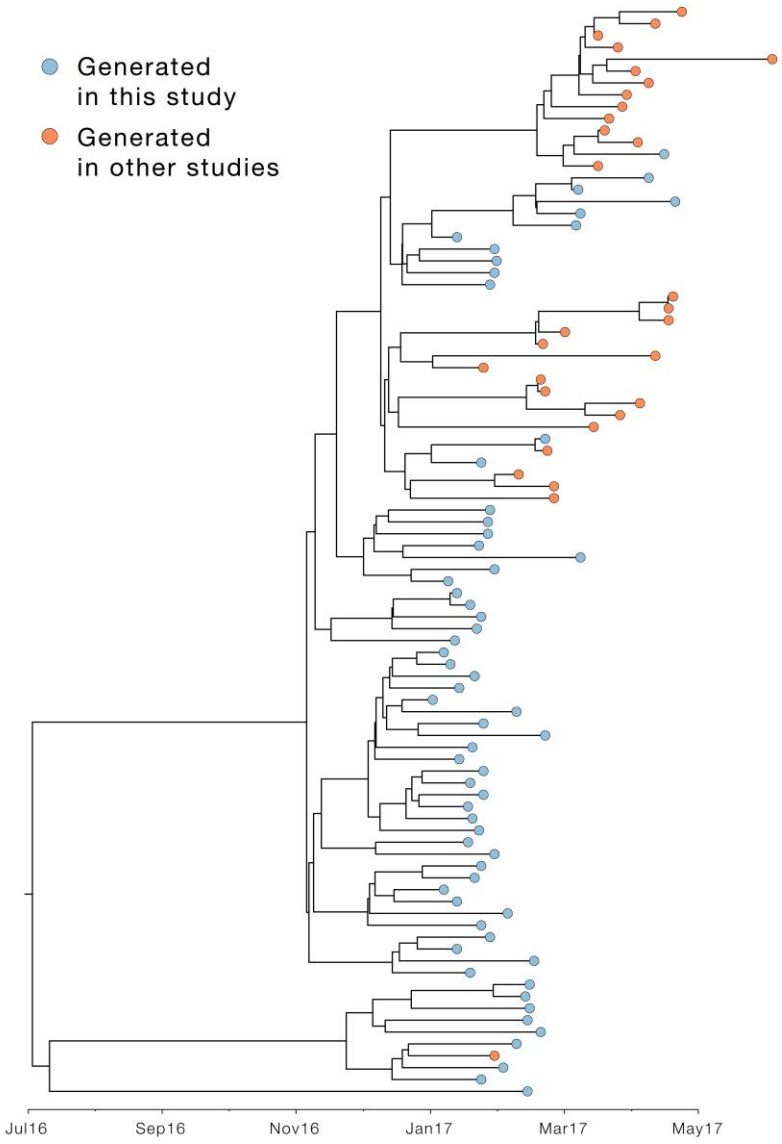

Fig. S10.

Molecular clock phylogeny of the YFV epidemic in Brazil. This tree includes sequences published in studies of the secondary epidemic Blue taxa = generated by this study; orange taxa = generated two separate studies of the secondary epidemic wave of YFV in Brazil (*34*, *109*). Tree was estimated in BEAST v1.8.4 (*96*) using a HKY + Γ_4 nucleotide substitution model and a relaxed molecular clock model, with an underlying lognormal distribution of branch rates (*97*).

Table S1.

Time-series cross correlation analysis. A, B and C correspond to the datasets A, B and C described in the text. Period = the time frame during which the comparison is made. The dataset in brackets is fixed during that period, while the other dataset is shifted temporally to correct for potential lag. *p*-value = the *p*-value of the correlation between the two "auto-correlated" series, accounting for the time-lag. Time lag = the lag between the two series, estimated to the nearest day via linear interpolation. Similar results were obtained when datasets D and B, D and C, and B and E were compared (data not shown).

Comparison	**A vs B**	**A vs C**	**B vs C**
Period	1 Aug 2016 to 1 Oct 2017 (A)	1 Aug 2016 to 1 Oct 2017 (A)	1 Aug 2016 to 1 Oct 2017 (A)
P-value	<0.0001	<0.0001	<0.0001
Cross-correlation	0.971	0.757	0.725
Time lag	B is 4 days ahead of A	C is 62 days behind A	C is 64 days behind B

Table S2.

Laboratories involved in YFV genome sequencing. H=Human, NHP=Non-human primate, MG = Minas Gerais, BA=Bahia, RJ=Rio Janeiro, ES=Espírito Santo.

Institution, State	Sample collection	Sequencing Platform
FUNED, MG	MG	MinION
FIOCRUZ, BA	MG, BA	MinION
FIOCRUZ, RJ	MG, ES, RJ	Ion Torrent™
IAL, SP	ES	MiSeq

Table S3.

Epidemiological data associated with each isolate processed/sequenced in this study. ID=Project identifier; Lab=Laboratory where samples were processed/sequenced; Host=Host species; State: MG=Minas Gerais, BA=Bahia; ES=Espirito Santo; RJ=Rio de Janeiro; Municipality=municipality of residence (unless stated otherwise). Date=Date of sample collection; Ct=RT-qPCR Cycle threshold value. "–"=not available. NHP=non-human primate of unknown species. **Alouatta spp.* suspected. [1]=Patient M26 was a resident of São Caetano do Sul, São Paulo, whose symptoms began on 30th December 2016. Travel history suggests that M26's infection occurred in Itambacuri municipality, neighboring Teófilo Otóni municipality, in Minas Gerais. [2]= Patient M123 was from São Paulo and visited Itambacuri, Minas Gerais. The patient reported receiving a vaccination, but no vaccination date was available. [3]=Patient M138 was vaccinated in Jan 2017 and developed symptoms 3 days later. [4]=Patient RJ185 was vaccinated in Jan 2017, 4 days before sample collection.

ID	Lab	Sample	Host	State	Municipality	Date	Ct	Sex	Age	Residence
M5	FUNED	Liver	*Alouatta*	MG	Sta. Rita de Caldas	15/02/2017	11	-	-	-
M7	FUNED	Tissue	*Cebidae*	MG	Delfinopolis	14/02/2017	6	-	-	-
M9	FUNED	Tissue	*Alouatta*	MG	Ouro Fino	15/02/2017	11	-	-	-
M11	FUNED	Tissue	*Cebidae*	MG	Caldas	13/02/2017	6	-	-	-
M16	FUNED	Liver	Human	MG	Caratinga	09/01/2017	17	M	43	Urban
M17	FUNED	Liver	Human	MG	Ladainha	14/01/2017	12	M	49	Rural
M18	FUNED	Liver	Human	MG	Itambacuri	21/01/2017	15	M	52	Rural
M25	FUNED	Liver	Human	MG	Novo Cruzeiro	25/01/2017	13	M	62	Rural
M26	FUNED	Liver	Human	MG	Itambacuri	07/01/2017	15	M	62	Rural[1]
M35	FUNED	Liver	Human	MG	Teófilo Otóni	20/01/2017	21	M	45	Rural
M36	FUNED	Liver	Human	MG	Novo Cruzeiro	14/01/2017	16	F	55	Rural
M43	FUNED	Liver	Human	MG	Pote	18/01/2017	13	M	43	Rural
M47	FUNED	Liver	Human	MG	Setubinha	19/01/2017	12	M	40	Urban
M48	FUNED	Liver	Human	MG	Ladainha	07/01/2017	17	M	52	Rural
M51	FUNED	Liver	Human	MG	Novo Cruzeiro	18/01/2017	18	M	46	Urban
M58	FUNED	Liver	Human	MG	Itambacuri	28/01/2017	22	M	53	Rural
M68	FUNED	Liver	Human	MG	Ladainha	28/01/2017	20	F	55	Rural
M73	FUNED	Liver	*Cebidae*	MG	Aguanil	22/02/2017	14	-	-	Urban

M78	FUNED	Liver	*Callithrix*	MG	Claraval	20/02/2017	12	-	-	Rural
M79	FUNED	Liver	*Alouatta*	MG	Abre Campo	17/02/2017	11	-	-	Rural
M83	FUNED	Liver	*Callithrix*	MG	Ladainha	09/02/2017	12	M	Adult	-
M94	FUNED	Liver	Human	MG	Imbe de Minas	13/01/2017	14	M	38	Rural
M96	FUNED	Liver	Human	MG	Ladainha	19/01/2017	19	M	56	Urban
M98	FUNED	Liver	Human	MG	Teófilo Otóni	10/01/2017	15	M	72	Urban
M99	FUNED	Liver	Human	MG	Teófilo Otóni	21/01/2017	10	M	58	Rural
M100	FUNED	Liver	Human	MG	Ladainha	02/01/2017	14	M	43	Rural
M105	FUNED	Liver	Human	MG	Pie. Caratinga	12/01/2017	17	M	33	Rural
M107	FUNED	Liver	Human	MG	Novo Cruzeiro	30/01/2017	9	M	35	Rural
M123	FUNED	Serum	Human	MG	Itambacuri	27/01/2017	18	F	47	Rural[2]
M138	FUNED	Serum	Human	MG	St. Barbara Leste	20/01/2017	15	M	50	Rural[3]
M164	FUNED	Serum	Human	MG	Simonesia	30/01/2017	32	F	0	-
M210	UFMG	Liver	*Callithrix*	MG	Sabinópolis	19/01/2017	12	F	-	-
M211	UFMG	Liver	*Alouatta*	MG	José Raydan	13/01/2017	9	F	-	-
M216	UFMG	Liver	*Callicebus*	MG	Luisburgo	25/01/2017	14	F	-	-
M217	UFMG	Liver	*Cebidae*	MG	Chapada Gaúcha	25/01/2017	11	M	-	-
M225	FioCruz	Serum	Human	MG	-	15/03/2003	-	-	-	-
M226	FioCruz	Serum	Human	MG	-	15/08/2003	-	-	-	-
M218	FioCruz	-	NHP	BA	Cordeiros	10/03/2017	-	-	-	-
438	FioCruz	blood	NHP	ES	Domingos Martins	31/01/2017	19	-	-	-
460	FioCruz	Liver	Human	MG	Novo Cruzeiro	30/01/2017	16	M	35	Rural
465	FioCruz	Liver	Human	MG	Itambacuri	30/01/2017	17	M	35	Rural
480	FioCruz	Liver	Human	MG	Teófilo Otóni	28/01/2017	13	F	47	Rural
532	FioCruz	Liver	*Alouatta*	MG	Coronel Murta	13/01/2017	8	F	Adult	-
1536	FioCruz	serum	Human	ES	Vitória	22/02/2017	11	M	31	Urban
1818	FioCruz	serum	Human	ES	Cariacica	10/03/2017	19	M	65	Urban

2109	FioCruz	Liver	*Alouatta*	ES	Cariacica	08/03/2017	7	-	-	-
2115	FioCruz	Liver	*Callithrix*	ES	Cariacica	09/03/2017	11	-	-	-
3919	FioCruz	serum	Human	ES	Domingos Martins	10/04/2017	14	M	45	Rural
4278	FioCruz	Liver	*Alouatta*	RJ	Maricá	17/04/2017	11	M	5	
4480	FioCruz	plasma	Human	RJ	Casimiro de Abreu	22/04/2017	19	M	47	Urban
8	IAL	Liver	NHP *	ES	Itarana	24/01/2017	12	-	-	-
11	IAL	Liver	NHP *	ES	Venda N. Imigrante	24/01/2017	13	-	-	-
RJ143	FioCruz	Liver	*Callithrix*	MG	José Raydan	24/01/2017	10	-	1	-
RJ148	FioCruz	Liver	Human	MG	Teófilo Otóni	24/01/2017	18	M	44	Rural
RJ150	FioCruz	Liver	Human	MG	Teófilo Otóni	27/01/2017	18	M	46	Rural
RJ151	FioCruz	Liver	NHP *	MG	Itamarandiba	13/01/2017	9	-	1	-
RJ153	FioCruz	Liver	*Callithrix*	MG	Felício dos SAntos	24/01/2017	9	-	1	-
RJ157	FioCruz	Liver	Human	MG	Teófilo Otóni	05/02/2017	17	M	35	Rural
RJ162	FioCruz	Liver	*Callithrix*	MG	Chapada Gaúcha	03/02/2017	9	-	-	-
RJ163	FioCruz	Liver	*Callithrix*	MG	Sacramento	09/02/2017	13	-	-	-
RJ164	FioCruz	Liver	NHP *	MG	Juatuba	14/02/2017	11	-	-	-
RJ168	FioCruz	Spleen	*Alouatta*	MG	Bom Jesus do Galho	23/01/2017	18	-	-	-
RJ171	FioCruz	Liver	NHP *	MG	S. Roque de Minas	03/02/2017	9	-	-	-
RJ185	FioCruz	Serum	Human	MG	S. João Evangelista	22/01/2017	14	M	39	Rural[4]

Table S4.

Primer sequences (*n=54*) for the YFV-500bp MinION sequencing scheme.

>YFV_500_1_LEFT	>YFV_500_15_LEFT
GTCTGGTCGTAAAGCTCAGGGA	ATCATCATGGACGAAGCACATTTTT
>YFV_500_1_RIGHT	>YFV_500_15_RIGHT
TGGGGCAGTTGTATTCCATGGA	CCTTCATCCACAAGCACAGGTT
>YFV_500_2_LEFT	>YFV_500_16_LEFT
TCTGAGGACCTTGGGAAAACCT	TCTTGGCCACTGACATAGCTGA
>YFV_500_2_RIGHT	>YFV_500_16_RIGHT
ACCCCCTCAATGAAATCCCTGT	CTCAAAGCACCACTTTCGGTCA
>YFV_500_3_LEFT	>YFV_500_17_LEFT
AACATGACGCAACGAGTTGTCA	ATGACCAGAGGAGAGTCTTCCG
>YFV_500_3_RIGHT	>YFV_500_17_RIGHT
TGTTCCAATTCTCCTGCTTGGC	GGCCAGAACAAACAGCATGACT
>YFV_500_4_LEFT	>YFV_500_18_LEFT
TGCCAAGTTTACCTGTGCCAAA	TGCTCTTGCACTCTGAGGAAGG
>YFV_500_4_RIGHT	>YFV_500_18_RIGHT
CTTTTGTGACTCGCATTGCACC	GCAGCTCCTGGTTTCAAGTCAA
>YFV_500_5_LEFT	>YFV_500_19_LEFT
TGGAATTTGAGCCTCCACATGC	AGGCATGCTGGAAAAGACTAAGG
>YFV_500_5_RIGHT	>YFV_500_19_RIGHT
TCCTTGTGCCACTGGTAAGTCA	TTCTTCTCATAGAGGGCAGGCA
>YFV_500_6_LEFT	>YFV_500_20_LEFT
GCCTCCACCAATGATGATGAAGT	GCACAGAGGAGGGTGTTTCATG
>YFV_500_6_RIGHT	>YFV_500_20_RIGHT
ACACTTGAGCTCTCTCTTGCCA	CACGGTCCACTTCCACAATGTC
>YFV_500_7_LEFT	>YFV_500_21_LEFT
ACAATGTCCATGAGCATGATCCT	AAGATGAAGACTGGACGCAGGG
>YFV_500_7_RIGHT	>YFV_500_21_RIGHT
ACGGACACTCTTTCCTGGACTT	GATGATGGGGACGACTCTCCAA
>YFV_500_8_LEFT	>YFV_500_22_LEFT
CAATACGGCTGGAAGACTTGGG	TGGGGTGGAACATCATCACCTT
>YFV_500_8_RIGHT	>YFV_500_22_RIGHT
GGCAAGCTTCCCTTTTCACCTC	TCTGTTTCCACACTGCGTGTTC
>YFV_500_9_LEFT	>YFV_500_23_LEFT
TCAATTGGGGGTCCAGTTAGCT	CTGTGAATCAAACATCCCGCCT
>YFV_500_9_RIGHT	>YFV_500_23_RIGHT
TCATGAAAGTGCAGTCCAACCG	GGTTCTTTTCTCTGGCCAGGTG
>YFV_500_10_LEFT	>YFV_500_24_LEFT
GTTGGAGGCATGGTGCTTCTAG	TGACACCAGAGCAAAGGATCCA
>YFV_500_10_RIGHT	>YFV_500_24_RIGHT
GGGGTATGGTCTTCTGCATGGA	CGCATAGAATCCACCACCCTCT
>YFV_500_11_LEFT	>YFV_500_25_LEFT
TGACAATGGCTGAGGTGAGACT	TGAATGAGGACCACTGGGCATC
>YFV_500_11_RIGHT	>YFV_500_25_RIGHT
ATCCCATGGCACCTTCTCTTCA	ATCTCCACTCACTGCCATCCTC
>YFV_500_12_LEFT	>YFV_500_26_LEFT
TGGGAAGAGGAAGCTGAGATCAG	TGATACACCACCAGCATGTCCA
>YFV_500_12_RIGHT	>YFV_500_26_RIGHT
TCTCCATCCCATCTACCCTCCA	CTTCCCATGAACAGACCACGTG
>YFV_500_13_LEFT	>YFV_500_27_LEFT
CGAGGGGCCTTTCTCGTTAGAA	CAAGATGAGCTGGTTGGCAGAG
>YFV_500_13_RIGHT	>YFV_500_27_RIGHT
GTCTTGTTTTCCCAGCTCCAGG	CTGCAGATCAGCATCCACAGAG
>YFV_500_14_LEFT	
AACTGAGGTGAAAGAGGAGGGG	
>YFV_500_14_RIGHT	
GGGGTGGCAGTCATCAAGATTG	

Table S5.

Statistics for the sequences generated using the MinION sequencer.

Sample	Host	Ct	Total Reads	Mapped Reads	Bases Covered >10x	Bases Covered >=25x	% Reference Covered
M5	NHP	11	35357	35319	10216	10216	99
M7	NHP	6	29103	29073	10216	10216	99
M9	NHP	11	16838	16781	10216	10216	99
M11	NHP	6	13499	13416	10216	10216	99
M16	Human	17	45277	44706	10216	9946	97
M17	Human	12	12984	12974	10216	10216	99
M18	Human	15	30246	27268	10204	9370	91
M25	Human	13	9477	9463	10216	10214	99
M26	Human	15	116601	111963	10216	10204	99
M35	Human	21	35936	28670	9946	9673	94
M36	Human	16	24004	22049	10216	9358	91
M43	Human	13	4858	4836	10216	9365	91
M47	Human	12	130596	130176	10216	10213	99
M48	Human	17	11516	11119	6460	5767	56
M51	Human	18	37159	36972	10216	10216	99
M58	Human	22	45367	45163	10216	10216	99
M68	Human	20	45925	45642	10216	10209	99
M73	NHP	14	107280	105827	10216	10216	99
M78	NHP	12	24353	24190	10216	10216	99
M79	NHP	11	35816	35449	10216	10216	99
M83	NHP	12	46097	45590	10216	10216	99
M94	Human	14	215524	213371	10216	10216	99
M96	Human	19	22424	20838	8772	5768	56
M98	Human	15	66026	65498	10216	9673	94
M99	Human	10	93181	92822	10216	10216	99
M100	Human	14	148291	147395	10216	10216	99
M105	Human	17	36169	35955	10216	10216	99
M107	Human	9	21532	21417	10216	10216	99
M123	Human	18	31236	30980	10216	10136	98
M138	Human	15	77337	76703	10216	10216	99
M164	Human	32	33076	28899	9075	8346	81
M210	NHP	12	359234	358667	10216	10216	99
M211	NHP	9	342819	342251	10216	10216	99
M216	NHP	14	387989	387305	10216	8916	87
M217	NHP	11	260350	259631	10211	7361	71
M218	NHP	-	70075	69013	10216	10216	99
M225	NHP	-	87647	87506	10216	10216	99
M226	NHP	-	81415	81160	10216	10216	99
RJ143	NHP	10	733	604	8266	3442	54
RJ148	Human	18	17853	6546	9541	9539	93
RJ150	Human	18	26895	25073	9855	9844	96
RJ151	NHP	9	2919	2809	10170	9764	97
RJ153	NHP	9	2857	2758	10172	10166	99
RJ157	Human	17	7531	5025	9258	9256	90
RJ162	NHP	9	6276	6176	10172	10168	99
RJ163	NHP	13	4210	4171	10171	9859	99
RJ164	NHP	11	2563	2549	9851	9403	92
RJ168	NHP	18	3202	2544	9430	8804	89
RJ171	NHP	9	1603	1534	10115	8014	85

RJ185	Human	14	3369	33555	10171	9717	99

Table S6.

Statistics for the sequences generated using the IonTorrent sequencer.

Sample	Host	Ct	Total Reads	Mapped Reads	Bases Covered >1x	Bases Covered >=10x	% Reference Covered
438	NHP	19	223937	217572	9958	9957	97
460	Human	16	44315	41509	9541	8718	81
465	Human	17	207616	202108	9811	9647	94
480	Human	13	47989	44951	9950	9922	96
532	Alouatta	8	283134	277698	10234	9948	97
1536	Human	11	71865	67197	9951	9721	97
1818	Human	19	46210	42614	9273	8824	82
2109	Aloutta	7	324683	315437	10120	10005	97
2115	Callithrix	11	43913	40994	9950	9467	91
3919	Human	14	44821	39888	9735	9139	82
4278	Alouatta	11	44873	41367	9931	9667	97
4480	Human	19	217496	210440	9951	9950	97

Table S7.

Statistics for the sequences generated using the MiSeq sequencer.

Sample	Host	Ct	Total Reads	Mapped Reads	Bases Covered >1x	Bases Covered >=10x	% Reference Covered
8	NHP	12	429568	1507	9774	5307	96
11	NHP	13	574990	3120	9995	7754	98

Table S8.

Coalescent generalized linear model results. Inclusion = probability that the predictor was included in the model. BF = Bayes factor. cEffect = conditional effect size, which represents the estimate of the coefficient conditional on the predictor being included in the model. Both the mean and the 95% highest posterior density credible interval (95% HPD) of the conditional effect size are reported. YFV(H), YFV(NHP) and CHIKV correspond to the time series in epidemiological datasets A to C described above. * HPDs could not be computed for CHIKV as the predictor is too infrequently included in the model.

Predictor	Inclusion	BF	cEffect (95% HPD)
YFV(H)	0.367	1.870	0.959 (0.623, 1.320)
YFV(NHP)	0.633	4.812	1.081 (0.713, 1.466)
CHIKV	0.0003	0.0002	0.761 (NA, NA)*

Table S9.

Generalized stepping-stone (GSS) sampling for each of the continuous diffusion models. Models are ordered according to their Bayes Factor (BF) score, calculated against the Brownian diffusion model (homogeneous diffusion).

Model	**GSS**	**Bayes Factor**
Homogeneous	-16648,01	0.0
Lognormal	-16605,42	42.59
Cauchy	-16580,12	**67.89**

Movie S1.
Reconstructed spatiotemporal diffusion of the YFV outbreak in Brazil.

References

1. T. Garske *et al.*, Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Medicine* **11**, e1001638 (2014).
2. C. I. Paules, Fauci, A. S., Yellow Fever - Once Again on the Radar Screen in the Americas. *The New England Journal of Medicine*, 1397-1399 (2017).
3. M. Theiler, Smith, H. H., The effect of prolonged cultivation in vitro upon the pathogenicity of yellow fever virus. *J. Exp. Med.* **65**, 767-786 (1937).
4. F. M. Shearer *et al.*, Global yellow fever vaccination coverage from 1970 to 2016: an adjusted retrospective analysis. *The Lancet Infectious Diseases* **17**, 1209-1217 (2017).
5. M. R. Nunes *et al.*, Genomic and phylogenetic characterization of Brazilian yellow fever virus strains. *Journal of Virology* **86**, 13263-13271 (2012).
6. J. P. Mutebi, H. Wang, L. Li, J. E. Bryant, A. D. Barrett, Phylogenetic and evolutionary relationships among yellow fever virus isolates in Africa. *Journal of Virology* **75**, 6999-7008 (2001).
7. J. J. von Lindern *et al.*, Genome analysis and phylogenetic relationships between east, central and west African isolates of Yellow fever virus. *The Journal of General Virology* **87**, 895-907 (2006).
8. E. Wang *et al.*, Genetic variation in yellow fever virus: duplication in the 3' noncoding region of strains from Africa. *Virology* **225**, 274-281 (1996).
9. G. J. Chang, B. C. Cropp, R. M. Kinney, D. W. Trent, D. J. Gubler, Nucleotide sequence variation of the envelope protein gene identifies two distinct genotypes of yellow fever virus. *Journal of Virology* **69**, 5773-5780 (1995).
10. M. N. O. Segura, Castro, F. C. C., *Atlas de Culicídeos na Amazônia Brasileira*. (Instituto Evandro Chagas Press, 2007), vol. 167.
11. C. Cardoso Jda *et al.*, Yellow fever virus in Haemagogus leucocelaenus and Aedes serratus mosquitoes, southern Brazil, 2008. *Emerging Infectious Diseases* **16**, 1918-1924 (2010).
12. A. A. Grobbelaar *et al.*, Resurgence of Yellow Fever in Angola, 2015-2016. *Emerging Infectious Diseases* **22**, 1854-1855 (2016).
13. M. U. Kraemer *et al.*, Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015-16: a modelling study. *The Lancet Infectious Diseases* **17**, 330-338 (2017).
14. O. Franco, História da Febre Amarela no Brasil. *Ministério da Saúde, DNERU, Rio de Janeiro*, 208 (1969).
15. R. A. G. B. Consoli, Lourenço-de-Oliveira, R., *Principais mosquitos de importância sanitária no Brasil [online]*. FioCruz (Editora FIOCRUZ, Rio de Janeiro, 1994).
16. J. Vainio, "Yellow fever," *World Health Organization* (1998).
17. M. U. Kraemer *et al.*, The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife* **4**, e08347 (2015).
18. Secretaria de Vigilância em Saúde, Ministério da Saúde, Emergência epidemiológica de febre amarela no Brasil, no período de Dezembro de 2016 a Julho de 2017 (http://portalarquivos2.saude.gov.br/images/pdf/2017/setembro/06/2017_027.pdf) Boletim Epidemiológico no. **48** (2017).
19. P. F. Vasconcelos *et al.*, Epidemic of jungle yellow fever in Brazil, 2000: implications of climatic alterations in disease spread. *Journal of Medical Virology* **65**, 598-604 (2001).

20. PAHO/WHO, "Epidemiological Update Yellow Fever http://www.paho.org/hq/index.php?option=com_topicsandview=readallandcid=2194andItemid=40784andlang=en 16 Feb 2018," *PAHO/WHO Yellow Fever* (2018).
21. F. M. Shearer *et al.*, Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *The Lancet. Global health* **6**, e270-e278 (2018).
22. Instituto Evandro Chagas. http://www.iec.gov.br/portal/ (2018).
23. R. Lourenco-de-Oliveira, M. Vazeille, A. M. de Filippis, A. B. Failloux, Aedes aegypti in Brazil: genetically differentiated populations with high susceptibility to dengue and yellow fever viruses. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **98**, 43-54 (2004).
24. D. Couto-Lima *et al.*, Potential risk of re-emergence of urban transmission of Yellow Fever virus in Brazil facilitated by competent Aedes populations. *Scientific Reports* **7**, 4848 (2017).
25. N. R. Faria, Lourenco, J., Cerqueira, E. M., Lima, M. M., Pybus, O. G., Alcantara, L. C. J., Epidemiology of Chikungunya Virus in Bahia, Brazil, 2014-2015. *PLoS Currents Outbreaks* **1**, (2016).
26. N. R. Faria *et al.*, Genomic and epidemiological characterisation of a dengue virus outbreak among blood donors in Brazil. *Scientific Reports* **7**, 15216 (2017).
27. S. H. Tuboi, Z. G. Costa, P. F. da Costa Vasconcelos, D. Hatch, Clinical and epidemiological characteristics of yellow fever in Brazil: analysis of reported cases 1998-2002. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**, 169-175 (2007).
28. PAHO, "Outbreak of Yellow Fever in Paraguay" (Washington, 2009).
29. A. Nasidi *et al.*, Urban yellow fever epidemic in western Nigeria, 1987. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **83**, 401-406 (1989).
30. S. D. Thiberville *et al.*, Chikungunya fever: a clinical and virological investigation of outpatients on Reunion Island, South-West Indian Ocean. *PLoS Negl Trop Dis* **7**, e2004 (2013).
31. M. Ribeiro, C. M. Antunes, [Yellow fever: study of an outbreak]. *Revista da Sociedade Brasileira de Medicina Tropical* **42**, 523-531 (2009).
32. G. Quick J., N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, k., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., de Jesus, J. G., Giovanetti, M., Hill, S., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G, Andersen, K. G., Loman, N. J., Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols* **12**, 1261-1276 (2017).
33. M. C. Bonaldo *et al.*, Genome analysis of yellow fever virus of the ongoing outbreak in Brazil reveals polymorphisms. *Memorias do Instituto Oswaldo Cruz* **112**, 447-451 (2017).
34. A. Moreira-Soto, Torres, M. C., Lima de Mendonça, M. C., Mares-Guia, M. A., Dos Santos Rodrigues, C. D., Fabri, A. A., Dos Santos, C. C., Machado Araújo, E. S., Fischer, C., Ribeiro Nogueira, R. M., Drosten, C., Sequeira P. C., Drexler, J. .F, Bispo de Filippis A. M., Evidence for multiple sylvatic transmission cycles during the 2016-2017 yellow fever virus outbreak, Brazil. *Clin Microbiol Infect.* S1198-743X(18)30144-7 (2018).
35. Secretaria de Vigilância em Saúde, Ministério da Saúde, Reemergência da Febre Amarela Silvestre no Brasil, 2014/2015: situação epidemiológica e a importância da vacinação preventiva e da vigilância intensificada no período sazonal.

(http://portalarquivos2.saude.gov.br/images/pdf/2015/outubro/19/2015-032---FA-ok.pdf) *Boletim Epidemiológico* **46**, (2015).
36. P. F. Vasconcelos *et al.*, Genetic divergence and dispersal of yellow fever virus, Brazil. *Emerging Infectious Diseases* **10**, 1578-1584 (2004).
37. F. P. Camara, A. L. Gomes, L. M. Carvalho, L. G. Castello, Dynamic behavior of sylvatic yellow fever in Brazil (1954-2008). *Revista da Sociedade Brasileira de Medicina Tropical* **44**, 297-299 (2011).
38. P. F. Vasconcelos *et al.*, Inadequate management of natural ecosystem in the Brazilian Amazon region results in the emergence and reemergence of arboviruses. *Cadernos de Saúde Pública* **17 Suppl**, 155-164 (2001).
39. G. Dudas *et al.*, Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309-315 (2017).
40. M. S. Gill, P. Lemey, S. N. Bennett, R. Biek, M. A. Suchard, Understanding Past Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates. *Systematic Biology*, (2016).
41. T. G. Vaughan, D. Kuhnert, A. Popinga, D. Welch, A. J. Drummond, Efficient Bayesian inference under the structured coalescent. *Bioinformatics* **30**, 2272-2279 (2014).
42. G. Dudas, Carvalho, L. M., Rambaut A., Bedford, T., MERS-CoV spillover at the camel-human interface. *eLife* **31257.001** (2018).
43. O. G. Pybus *et al.*, Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15066-15071 (2012).
44. P. Lemey, A. Rambaut, J. J. Welch, M. A. Suchard, Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* **27**, 1877-1885 (2010).
45. L. Jung, I. Mourthe, C. E. Grelle, K. B. Strier, J. P. Boubli, Effects of Local Habitat Variation on the Behavioral Ecology of Two Sympatric Groups of Brown Howler Monkey (Alouatta clamitans). *PloS One* **10**, e0129789 (2015).
46. E. Flacio, L. Engeler, M. Tonolla, P. Muller, Spread and establishment of Aedes albopictus in southern Switzerland between 2003 and 2014: an analysis of oviposition data and weather conditions. *Parasites & Vectors* **9**, 304 (2016).
47. A. Estrada *et al.*, Impending extinction crisis of the world's primates: Why primates matter. *Sci Adv* **3**, e1600946 (2017).
48. R. A. Nascimento, Montano, R. A. M., An assessment of illegal capuchin monkey trade in Bahia State, Brazil. *Neotropical Biology and Conservation* **8**, 79-87 (2013).
49. A. J. Auguste *et al.*, Enzootic transmission of yellow fever virus, Venezuela. *Emerging Infectious Diseases* **21**, 99-102 (2015).
50. P. R. Stephens *et al.*, Global Mammal Parasite Database version 2.0. *Ecology* **98**, 1476 (2017).
51. B. W. Johnson *et al.*, Vector competence of Brazilian Aedes aegypti and Ae. albopictus for a Brazilian yellow fever virus isolate. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **96**, 611-613 (2002).
52. R. Lourenco de Oliveira, M. Vazeille, A. M. de Filippis, A. B. Failloux, Large genetic differentiation and low variation in vector competence for dengue and yellow fever viruses of Aedes albopictus from Brazil, the United States, and the Cayman Islands. *The American Journal of Tropical Medicine and Hygiene* **69**, 105-114 (2003).
53. F. Soper, Penna, E., Serafim, J., Frobisher, M., Pinheiro, J., Yellow fever whithout Aedes aegypti. Study of a rural epidemic in the Valle do Chanaan, Espirito Santo, Brazil, 1932. *Am J Hyg* **18**, 555-587 (1933).

54. M. Theiler, H. H. Smith, The Use of Yellow Fever Virus Modified by in Vitro Cultivation for Human Immunization. *The Journal of Experimental Medicine* **65**, 787-800 (1937).
55. D. J. Gubler, Dengue, Urbanization and Globalization: the Unholy Trinity of the 21st Century. *Tropical Medicine and Health* **39**, 3-11 (2011).
56. P. Kotsakiozi *et al.*, Tracking the return of Aedes aegypti to Brazil, the major vector of the dengue, chikungunya and Zika viruses. *PLoS Negl Trop Dis* **11**, e0005653 (2017).
57. B. Mondet *et al.*, Isolation of yellow fever virus from nulliparous Haemagogus (Haemagogus) janthinomys in eastern Amazonia. *Vector Borne and Zoonotic Diseases* **2**, 47-50 (2002).
58. P. F. Vasconcelos [Yellow Fever]. *Revista da Sociedade Brasileira de Medicina Tropical* **36**, 275-293 (2003).
59. P. F. Vasconcelos *et al.*, Isolations of yellow fever virus from Haemagogus leucocelaenus in Rio Grande do Sul State, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **97**, 60-62 (2003).
60. P. F. Vasconcelos, Yellow fever in Brazil: thoughts and hypotheses on the emergence in previously free areas. *Revista de Saúde Pública* **44**, 1144-1149 (2010).
61. K. R. Cavalcante, P. L. Tauil, Epidemiological characteristics of yellow fever in Brazil, 2000-2012. *Epidemiol Serv Saúde* **25**, 11-20 (2016).
62. J. Vainio, Cutts, F., "Yellow fever" *WHO/EPI/GEN/98.11* (World Health Organization, Geneva, 1998).
63. World Health Organization - Emergencies preparedness, response [online]. Trends over time (http://www.who.int/csr/disease/yellowfev/trends/en/, 2018).
64. A. A. Sall *et al.*, Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. *Journal of Virology* **84**, 765-772 (2010).
65. J. E. Bryant, E. C. Holmes, A. D. Barrett, Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathogens* **3**, e75 (2007).
66. C. L. B. Gadia *et al.*, Identification of pathogens for differential diagnosis of fever with jaundice in the Central African Republic: a retrospective assessment, 2008-2010. *BMC Infectious Diseases* **17**, 735 (2017).
67. P. Van der Stuyft *et al.*, Urbanisation of yellow fever in Santa Cruz, Bolivia. *Lancet* **353**, 1558-1562 (1999).
68. Secretaria de Estado de Saúde de Minas Gerais, Manejo Clínico Febre Amarela (Belo Horizonte, Minas Gerais, Brazil, 2017).
69. I. B. Rabe *et al.*, Interim Guidance for Interpretation of Zika Virus Antibody Test Results. *MMWR* **65**, 543-546 (2016).
70. C. T. Lloyd, A. Sorichetta, A. J. Tatem, High resolution global gridded data for use in population studies. *Scientific Data* **4**, 170001 (2017).
71. Secretaria de Vigilância em Saúde, Ministério da Saúde, Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 5 de 2018. *Boletim Epidemiológico* **49** (2018).
72. Secretaria de Estado de Saúde de Minas Gerais, "Manual Técnico de Atenção à Saúde e resposta aos casos de infecção pelo vírus Zika em gestantes, fetos e recém-nascidos (v1)" (Belo Horizonte, Minas Gerais, Brazil, 2016).
73. F. M. Shearer, Longbottom, J., Browne, A. J., Pigott, D. M., Brady, O. J., Kraemer, M. U., Marinho, F., Yactayo, S., de Araujo, V. E. M., Nobrega, A. A., Fullman, N., Ray, S. E., Mosser, J. F., Stanaway, J. D., Lim, S. S., Reiner Jr, R. C., Moyes, C. L., Hay, S. I. , Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *The Lancet Global Health* **6**, 270-328 (2018).

74. A. Sorichetta *et al.*, High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data* **2**, 150045 (2015).
75. D. J. Weiss *et al.*, An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J Photogramm Remote Sens* **98**, 106-118 (2014).
76. A. Hamlet *et al.*, The seasonal influence of climate and environment on yellow fever transmission across Africa. *PLoS Negl Trop Dis* **12**, e0006284 (2018).
77. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL: http://R-project.org/.
78. C. Domingo *et al.*, Advanced yellow fever virus genome detection in point-of-care facilities and reference laboratories. *Journal of Clinical Microbiology* **50**, 4054-4060 (2012).
79. N. R. Faria *et al.*, Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546**, 406-410 (2017).
80. M. Kearse *et al.*, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).
81. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
82. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
83. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772 (2012).
84. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
85. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
86. N. A. O'Leary *et al.*, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**, D733-745 (2016).
87. C. M. Rice *et al.*, Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* **229**, 726-733 (1985).
88. N. R. Faria *et al.*, Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352**, 345-349 (2016).
89. M. R. Nunes *et al.*, Phylogeography of dengue virus serotype 4, Brazil, 2010-2011. *Emerging Infectious Diseases* **18**, 1858-1864 (2012).
90. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler, GenBank: update. *Nucleic Acids Research* **32**, D23-26 (2004).
91. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780 (2013).
92. A. Rambaut, Lam, T. T., Fagundes de Carvalho, L., Pybus, O. G., Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2**, (2016).
93. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267 (2006).
94. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**, vev003 (2015).
95. A. Moreira-Soto *et al.*, Evidence for multiple sylvatic transmission cycles during the 2016-2017 yellow fever virus outbreak, Brazil. *Clinical Microbiology and Infection* S1198-743C(18)30144-7 (2018).

96. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214 (2007).
97. A. J. Drummond, S. Y. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, e88 (2006).
98. E. C. Holmes, G. Dudas, A. Rambaut, K. G. Andersen, The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature* **538**, 193-200 (2016).
99. M. S. Gill *et al.*, Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution* **30**, 713-724 (2013).
100. R. Bouckaert *et al.*, BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537 (2014).
101. G. Baele, P. Lemey, M. A. Suchard, Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. *Systematic Biology* **65**, 250-264 (2016).
102. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973 (2012).
103. D. L. Ayres *et al.*, BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology* **61**, 170-173 (2012).
104. S. Dellicour, Rose, R., Pybus, O. G., Explaining the geographic spread of emerging epidemics: a framework comparing viral phylogenies and environmental landscape data. *BMC Bioinformatics* **17**:82 (2016).
105. S. Dellicour, R. Rose, N. R. Faria, P. Lemey, O. G. Pybus, SERAPHIM: studying environmental rasters and phylogenetically informed movements. *Bioinformatics* **32**, 3204-3206 (2016).
106. S. Dellicour *et al.*, Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics. *Molecular Biology and Evolution* **34**, 2563-2571 (2017).
107. Ministério da Saúde, "Manual de Vigilância Epidemiologia de Febre Amarela" (http://bvsms.saude.gov.br/bvs/publicacoes/manual_vigilancia_epid_febre_amarela.pdf) (Brasília, 2004).
108. Secretaria de Vigilância em Saúde, Ministério da Saúde, "Monitoramento do Período Sazonal da Febre Amarela Brasil – 2017/2018" (http://portalarquivos2.saude.gov.br/images/pdf/2018/maio/18/Informe-FA-26.pdf) Informe 26 (Brasília, 2018).
109. M. M. Gomez *et al.*, Genomic and structural features of the yellow fever virus from the 2016-2017 Brazilian outbreak. *The Journal of General Virology* **99**, 536-548 (2018).