medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Evolution and epidemic spread of SARS-CoV-2 in Brazil

Authors: Darlan S. Candido[1,†], Ingra M. Claro[2,†], Jaqueline G. de Jesus[2,†], William M. Souza[3,†], Filipe R. R. Moreira[4,†], Simon Dellicour[5,6†], Thomas A. Mellan[7,†], Louis du Plessis[1], Rafael H. M. Pereira[8], Flavia C. S. Sales[2], Erika R. Manuli[2], Julien Thézé[9], Luiz Almeida[10], Mariane T. Menezes[4], Carolina M. Voloch[4], Marcilio J. Fumagalli[3], Thais M. Coletti[2], Camila A. M. Silva[2], Mariana S. Ramundo[2], Mariene R. Amorim[11], Henrique Hoeltgebaum[12], Swapnil Mishra[6], Mandev S. Gill[6], Luiz M. Carvalho[13], Lewis F. Buss[2], Carlos A. Prete Jr[14], Jordan Ashworth[15], Helder Nakaya[16], Pedro S. Peixoto[17], Oliver J. Brady[18,19], Samuel M. Nicholls[20], Amilcar Tanuri[4], Átila D. Rossi[4], Carlos K.V. Braga[8], Alexandra L. Gerber[10], Ana Paula Guimarães[10], Nelson Gaburo Jr[21], Cecila S. Alencar[22], Alessandro C.S. Ferreira[23], Cristiano X. Lima[24,25], José Eduardo Levi[26], Celso Granato[27], Giula M. Ferreira[28], Ronaldo S. Francisco Jr[8], Fabiana Granja[29], Marcia T. Garcia[30], Maria Luiza Moretti[30], Mauricio W. Perroud Jr[31], Terezinha M. P. P. Castineiras[32], Carolina S. Lazari[22], Sarah C. Hill[1,33], Andreza A. de Souza Santos[34], Camila L. Simeoni[11], Julia Forato[11], Andrei C. Sposito[35], Angelica Z. Schreiber[36], Magnun N. N. Santos[36], Camila Zolini de Sá[37], Renan P. Souza[37], Luciana C. Resende-Moreira[38], Mauro M. Teixeira[39], Josy Hubner[40], Patricia A. F. Leme[41], Rennan G Moreira[42], Maurício Lacerda Nogueira[43], CADDE-Genomic-Network, Neil M Ferguson[6], Silvia F. Costa[2], José Luiz Proenca-Modena[11], Ana Tereza R. Vasconcelos[10], Samir Bhatt[7], Philippe Lemey[6], Chieh-Hsi Wu[44], Andrew Rambaut[45], Nick J. Loman[20], Renato S. Aguiar[37], Oliver G. Pybus[1,], Ester C. Sabino[2,‡], and Nuno Rodrigues Faria[1,2,7,‡,*].

Affiliations:

[1]Department of Zoology, University of Oxford, UK.

[2]Instituto de Medicina Tropical, Departamento de Moléstias Infecciosas e Parasitárias, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

[3]Centro de Pesquisa em Virologia, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, Brazil.

[4]Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

[5]Spatial Epidemiology Lab, Université Libre de Bruxelles, Bruxelas, Belgium.

[6]Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium.

[7]Department of Infectious Disease Epidemiology, Imperial College London, UK.

[8]Institute for Applied Economic Research, Brasília, Brazil.

[9]Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Petrópolis, Brazil.

[10]Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil.

[11]Department of Mathematics, Imperial College London, UK.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[12]Getulio Vargas Foundation, Rio de Janeiro, Brazil.

[13]Department of Electronic Systems Engineering, University of São Paulo, Brazil.

[14]Usher Institute, University of Edinburgh, Edinburgh, UK

[15]Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil.

[16]Departamento de Matemática Aplicada, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil.

[17]Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK.

[18]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK.

[19]Institute for Microbiology and Infection, University of Birmingham, Birmingham, UK.

[20]DB Diagnósticos do Brasil, São Paulo, Brazil.

[21]Divisão de Laboratório Central do Hospital das Clínicas, da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

[22]Instituto Hermes Pardini, Belo Horizonte, Brazil.

[23]Departamento de Cirurgia, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

[24]Simile Instituto de Imunologia Aplicada Ltda, Belo Horizonte, Brazil.

[25]Laboratório DASA, São Paulo, Brazil.

[26]Laboratório Fleury, São Paulo, Brazil.

[27]Laboratório de Virologia, Instituto de Ciências Biomédicas, Universidade Federal de Uberlândia, Uberlândia, Brazil.

[28]Centro de Estudos da Biodiversidade, Universidade Federal de Roraima, Boa Vista, Brazil.

[29]Divisão de Doenças Infecciosas, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brazil.

[30]Hospital Estadual Sumaré, Universidade Estadual de Campinas, Campinas, Brazil.

[31]Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

[32]Department of Pathobiology and Population Sciences, Royal Veterinary College, Hatfield, UK.

[33]Oxford School of Global and Area Studies, University of Oxford, UK.

[34]Departamento de Clínica Médica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brazil.

[35]Departamento de Patologia Clínica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brazil.

[36]Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[37]Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

[38]Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

[39]Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

[40]Centro de Saúde da Comunidade, Universidade Estadual de Campinas, Campinas, Brazil.

[41] Centro de Laboratórios Multiusuários, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

[42]Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, São Paulo, Brazil.

[43]Mathematical Sciences, University of Southampton, Southampton, UK.

[44]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK.

*Correspondence to: sabinoec@usp.br and nuno.faria@zoo.ox.ac.uk

† These authors contributed equally to this work.

Abstract

Brazil currently has one of the fastest growing SARS-CoV-2 epidemics in the world. Due to limited available data, assessments of the impact of non-pharmaceutical interventions (NPIs) on virus transmission and epidemic spread remain challenging. We investigate the impact of NPIs in Brazil using epidemiological, mobility and genomic data. Mobility-driven transmission models for São Paulo and Rio de Janeiro cities show that the reproduction number (R_t) reached below 1 following NPIs but slowly increased to values between 1 to 1.3 (1.0–1.6). Genome sequencing of 427 new genomes and analysis of a geographically representative genomic dataset from 21 of the 27 Brazilian states identified >100 international introductions of SARS-CoV-2 in Brazil. We estimate that three clades introduced from Europe emerged between 22 and 27 February 2020, and were already well-established before the implementation of NPIs and travel bans. During this first phase of the epidemic establishment of SARS-CoV-2 in Brazil, we find that the virus spread mostly locally and within-state borders. Despite sharp decreases in national air travel during this period, we detected a 25% increase in the average distance travelled by air passengers during this time period. This coincided with the spread of SARS-CoV-2 from large urban centers to the rest of the country. In conclusion, our results shed light on the role of large and highly connected populated centres in the rapid ignition and establishment of SARS-CoV-2, and provide evidence that current interventions remain insufficient to keep virus transmission under control in Brazil.

One Sentence Summary

Joint analysis of genomic, mobility and epidemiological novel data provide unique insight into the spread and transmission of the rapidly evolving epidemic of SARS-CoV-2 in Brazil.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Main Text

Severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) is a novel zoonotic betacoronavirus with a 30-kb genome that was first reported in December 2019 in Wuhan, China (*1, 2*). As of 9 June 2020, coronavirus disease (COVID-19) has caused over 7 million cases and 404 thousand deaths globally (*3*). SARS-CoV-2 was declared a public health emergency of international concern on 30 January 2020. The virus can be classified into two phylogenetic lineages, named lineage A and B, that spread from Wuhan before strict travel restrictions were enacted (*4, 5*) and now co-circulate around the world (*6*). The case fatality rate of SARS-CoV-2 infection has been estimated between 1.2 and 1.6% (*7-9*) with substantially higher ratios in those aged above 60 years (*10*). Over 18 to 56% of SARS-CoV-2 transmission is from asymptomatic or pre-symptomatic individuals (*11-13*), complicating epidemiological assessments and public health efforts to curb the pandemic.

While the growth of the SARS-CoV-2 epidemics in China, Italy, and Spain is declining in response to NPIs (*3*), the number of SARS-CoV-2 cases and deaths in Brazil continues to increase (*14*) (**Fig. 1A**). Brazil has now reported 739,503 SARS-CoV-2 cases, the second largest number in the world, and 38,406 deaths (as of 9 June 2020). Over one third of the cases (35%) in Brazil are concentrated in southeast region which includes São Paulo city (**Fig. 1B**), the world's fourth largest conurbation, where the first case in Latin America was reported on 25 February 2020 (*15*). Diagnostic assays for SARS-CoV-2 molecular detection were widely distributed across the national public health laboratory network from early on (*16, 17*). However, several factors including delays in reporting, changes in notification and heterogeneous access to testing across populations, have obfuscated real-time assessments of virus transmissibility based on SARS-CoV-2 case counts (*15*). Given this, death counts caused by severe acute respiratory infections (SARI) reported to the Sistema Único de Saúde (SUS) provide a more accurate measure for SARS-CoV-2 transmission in the country (*18*). Changes in SARS-CoV-2 transmissibility are strongly associated with changes in average mobility (*19, 20*), and can typically be measured by calculating the time-varying reproduction number, R_t, defined as the average number of secondary infections caused by an infected person. $R_t > 1$ indicates a growing epidemic while $R_t < 1$ is needed to achieve a decrease in transmission.

We used a Bayesian semi-mechanistic model (*21, 22*) based on SARI deaths and human mobility data to investigate daily changes in R_t in São Paulo city (12,2 millions inhabitants) and Rio de Janeiro city (6,7 millions inhabitants), the largest urban metropoles in Brazil (**Figs. 1C** and **1D**). Non-pharmaceutical policies in Brazil consisted of school closures implemented between 12 and 23 March 2020 across the countries' 27 federal units/states, and stores closures implemented between 13 and 23 March 2020. In São Paulo city, schools started closing on the 16 March and stores closed four days later. At the start of the epidemics, we found $R_t > 3$ in São Paulo and Rio de Janeiro, and that concurrent with the timing of state mandated non-pharmaceutical interventions (NPIs), R_t values fall close to 1.

Analysis of the reproduction number after NPI implementation highlights several notable mobility-driven features. We find that R_t drops during weekends and holidays due to differences

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

in movement patterns and social isolation. For example, after the implementation of NPIs in São Paulo, the mean *Rt* was 1.0 (95% Bayesian credible interval: 0.7–1.3) during weekends, compared to 1.1 (95% BCI: 0.9–1.5) during weekdays. In addition, there was a period immediately following NPIs, between 21 March and 31 March 2020, where the R_t was consistently less than 1 in São Paulo (**Fig. 1C**). However, after this initial drop, the R_t in São Paulo rises above 1 and shows an increasing trend associated with increased population mobility. For example, this can be seen in the Google transit stations index, that rises from -60% to -52%, and by the decrease in the social isolation index from 54% to 47%. By 4 May 2020, the reproduction number had an estimated value of 1.3 (CI95%: 1.0, 1.6) in both São Paulo and Rio de Janeiro cities (**Table S1**). However, we point out there were instances in the previous 7 days where the 95% credible intervals for R_t had fallen below a value of 1, drawing attention to the fluctuations and uncertainty in the reproduction number for both cities.

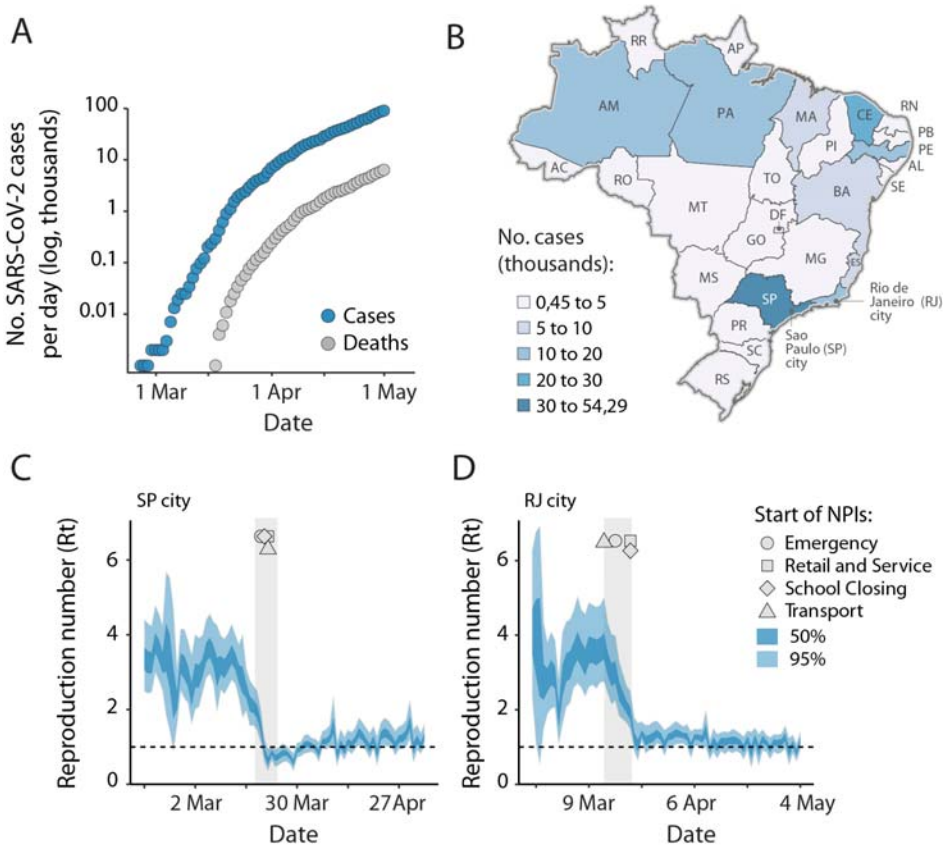

Fig. 1. SARS-CoV-2 epidemiology and epidemic spread in Brazil. (A) Cumulative number of SARS-CoV-2 reported cases (blue) and deaths (grey) in Brazil. **(B)** States are coloured according to the number of cumulative confirmed cases by April 30, 2020. (**C** and **D**) Reproduction number *Rt* over time for the cities of São Paulo (**C**) and Rio de Janeiro (**D**). Reproduction numbers (R_t) were estimated using a Bayesian approach incorporating daily number of deaths and four variables related to mobility data (a social isolation index from Brazilian geolocation company *InLoco*, and Google mobility indices for time spent in transit stations, parks, and the average between groceries and pharmacies, retail and recreational, and workspaces). Dashed horizontal line indicates *Rt*=1. Grey area and geometric symbols represent the time at which NPIs interventions were implemented. Bayesian credible intervals (BICs, 50 and 95%) are shown as

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

shaded areas. The 2-letter ISO 3166-1 codes for the 27 federal units in Brazil (26 federal states and 1 federal district) are as follows: AC=Acre, AL=Alagoas, AM=Amazonas, AP=Amapá, BA=Bahia, CE=Ceará, ES=Espírito Santo, DF=Distrito Federal, GO=Goiás, MA=Maranhão, MG=Minas Gerais, MS=Mato Grosso do Sul, MT=Mato Grosso, PA=Pará, PB=Paraíba, PE=Pernambuco, PI=Piauí, PR=Paraná, RJ=Rio de Janeiro, RN=Rio Grande do Norte, RO=Rondônia, RR=Roraima, RS=Rio Grande do Sul, SC=Santa Catarina, SE=Sergipe, SP=São Paulo, TO=Tocantins.

Early sharing of genomic sequences, including the earliest released SARS-CoV-2 genome, Wuhan-Hu-1, on 10 January (*23*), has enabled unprecedented global preparedness and response (*24, 25*). However, despite the thousands of virus genomes deposited on public access databases, there is a lack of consistent sampling structure that hampers accurate reconstructions of virus movement and transmission using phylogenetic analyses. To investigate how SARS-CoV-2 became established across Brazil, and to quantify the impact of NPIs in virus spatiotemporal spread, we tested a total of 26,732 samples from public and private laboratories using real-time PCR assays and found 7,944 (29%) to be positive for SARS-CoV-2. We then focused our sequencing efforts on generating a spatially representative genomic dataset with curated metadata that maximized the association between the number of sequences and the number of SARS-CoV-2 confirmed cases per state.

We generated 427 SARS-CoV-2 new genomes from Brazil with >75% 200-fold genome coverage that were sampled between the 5 March to 30 April 2020 (**Figs. S1** and **S2** and **Data S1**). The data generated here originated from 84 municipalities across 18 of 27 federal units spanning all regions in Brazil (**Fig. 2A, Fig. S2**). Sequenced samples were obtained from samples collected on average 5 days (median, range: 0 to 29 days) after onset of symptoms and were generated in 3 laboratories using harmonized sequencing and bioinformatic protocols. Despite the use of different diagnostic assays and sequencing primers across sites (**Table S2)**, our data suggests that higher genome coverage samples are characterized by lower real-time PCR cycle threshold (Ct) and shorter times between symptom onset and sample collection (**Fig. S3**). When we include 63 additional available sequences from Brazil deposited in GISAID, a strong spatial representativity was obtained when considering the total number of genomes per state and SARI SARS-CoV-2 cases (n=490 sequences from 21 states, Spearman's correlation, ρ=0.83, P-value < 3.79×10^{-7}, **Fig. 2A**). This correlation ranged from 0.70 to 0.83 (P-values < 0.0001) when considering SARS-CoV-2 deaths or suspected cases, respectively (**Fig. S4**). Moreover, we used an *in-silico* assessment of diagnostic assay specificity for Brazilian strains (n=490) to identify potential mismatches in some assays targeting Brazilian strains. We find that the forward primers of the Chinese CDC and Hong Kong University nucleoprotein-targeting RT-qPCR may be less appropriate for use in Brazil than other diagnostic assays for which few or no mismatches were identified (**Fig. S5, Table S3**).

The time between the date of the first reported case and the date of collection of the first sequence analysed for each state was 4.5 days (**Fig. 2A**). For 8 federal states genomes were obtained from samples collected up to 6 days before the first case notifications. We found that most (n=485/490) Brazilian sequences belong to SARS-CoV-2 lineage B, with only 5 strains belonging to the lineage A (2 from Amazonas, 1 from Rio Grande do Sul, 1 from Minas Gerais and 1 from Rio de Janeiro) (**Data S1 and S2** and **Fig. S6** shows detailed lineage information for each sequence).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

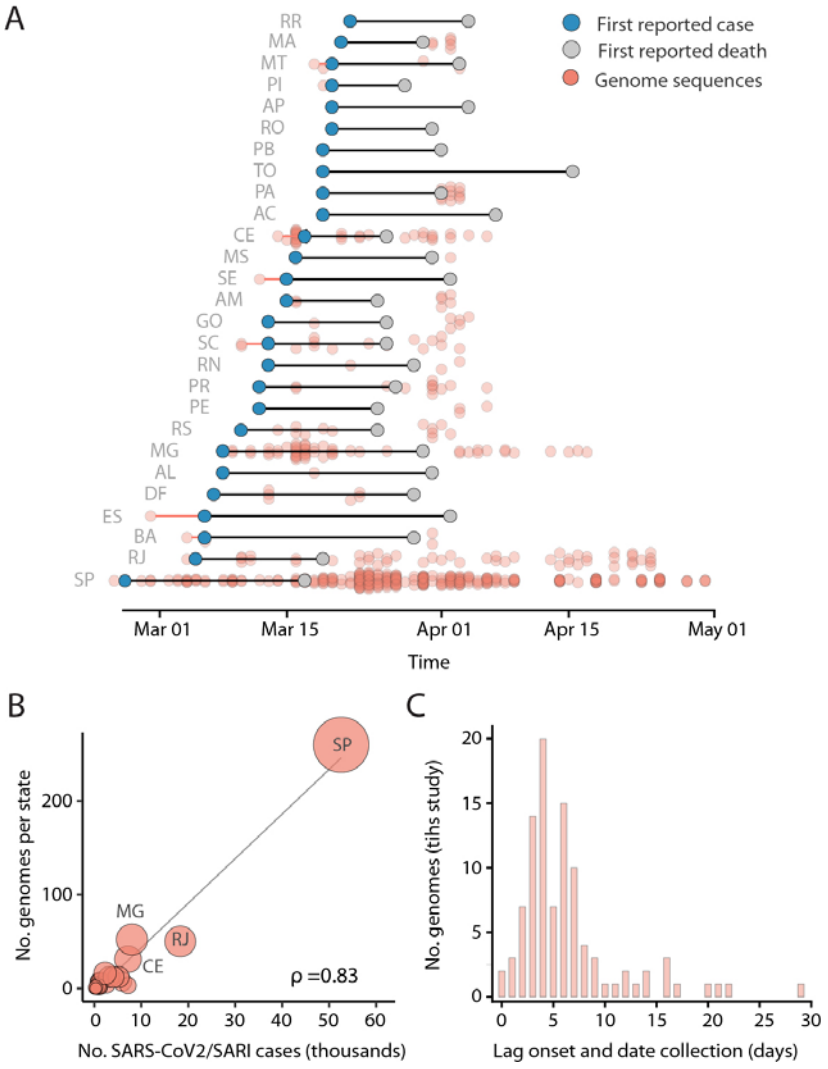

Fig. 2. Spatial representativity in genomic sampling. (A) Dumbbell plot showing temporal intervals between date of collection of sampled genomes, notification of first cases and first deaths in each state. Red lines indicate lag between date collection of first genome sequence and first reported case. The 2-letter ISO 3166-1 codes for the 27 federal units in Brazil (26 federal states and 1 federal district) are shown in the caption of Fig. 1. **(B)** Spearman's rank (ρ) correlation between the number of SARS-CoV-2 confirmed and suspected SARI cases against number of sequences for each of the 21 Brazilian states included in this study (see also Fig. S2). Circle sizes are proportional to the number of sequences for each federal unit. (**C**) Interval between the date of symptom onset and the date of sample collection for the sequences generated in this study.

We used a combination of genomic, spatial, mobility and epidemiological data to investigate changes in transmission patterns in Brazil over time. We first estimated the temporal signal of our global alignment by regressing genetic distances from the tips to the root of a maximum likelihood (ML) tree against sampling dates (**Fig. 3A**, **Fig. S7**). We found sufficient

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

temporal structure in our dataset (r^2=0.51) (**Fig. 3A**). Subsequently, our molecular clock analysis using an exponential parametric tree prior inferred an evolutionary rate of 1.126 x 10^{-3} (95 % BCI: 1.03–1.23 x 10^{-3}) substitutions per site per year (s/s/y), equating to an average of 33 changes per year. This is within the evolutionary rates estimated for other human coronaviruses (*26-29*). We estimate the date of the common ancestor of the SARS-CoV-2 pandemic to around mid-Nov 2019 (median estimate 17 Nov 2019, 95% BCI: 26 Oct 2019–16 Dec 2019), in line with recent findings (*30, 31*). Similar dates of emergence and rates of evolutionary change were obtained using a more flexible non-parametric coalescent model.

Time-measured phylogeographic analyses revealed at least 104 (95% BCI: 101–108) international introductions in Brazil (**Fig. 3A**, **Figs. S8-S9**). This is likely to represent an underestimate of the real number of introductions as we have sequenced, on average, 1 out of 200 confirmed cases (0.5%). Most of these introductions were directed into internationally well-connected states: São Paulo (36% of all imports), Minas Gerais (24%), Ceará (10%) and Rio de Janeiro (8%) (**Fig. S9**). The majority of the Brazilian genomes (75%, *n*=366/490) fall into three monophyletic clades hereafter named as *Clade 1* (*n*=186/490, 38% of all strains from Brazil), *Clade 2* (*n*=161, 33%) and *Clade 3* (*n*=19/490, 4%) (**Fig. 3A, Fig. S10**), which were also identified in a phylogenetic analysis using 13,307 global genomes. We estimate that these clades were introduced from Europe and emerged in Brazil between 22 and 27 February 2020, with upper credible intervals varying from 29 February to 3 March (see also **Fig. S11**). Similar to what has been recently shown in New York City (*32*), this indicates that community-driven transmission became established in Brazil around late February and early March, highlighting that international travel restrictions initiated after this period would have had limited efficacy.

Clade 1 is characterized by an amino acid mutation in the spike protein (G25088T, reference GenBank accession number, NC_045512.2) and circulates predominantly in São Paulo state (n=159, 85.4%, **Figs. S10 and S12**). *Clade 2* is defined by two amino acid mutations in the ORF6 (T27299C) and nucleoprotein (T29148C); this is the most spatially widespread lineage, with sequences from a total of 16 states in Brazil. *Clade 3* (*n*=19, 4%), is concentrated in Ceará state (*n*=16, 84%) and falls in a global cluster with sequences mainly from Europe. In the Amazon region, where the epidemic is expanding rapidly (*22*), we find evidence for multiple national and international introductions, with 37% (*n*=7/19) of all sequences from Pará and the Amazonas states clustering in *Clade 1* and 32% (*n*=6/19) in *Clade 2*. An interactive visualization of the temporal, geographic and mutational patterns in our data can be found in https://microreact.org/project/eyBufSP8UwVNQqBDjxLEuj.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

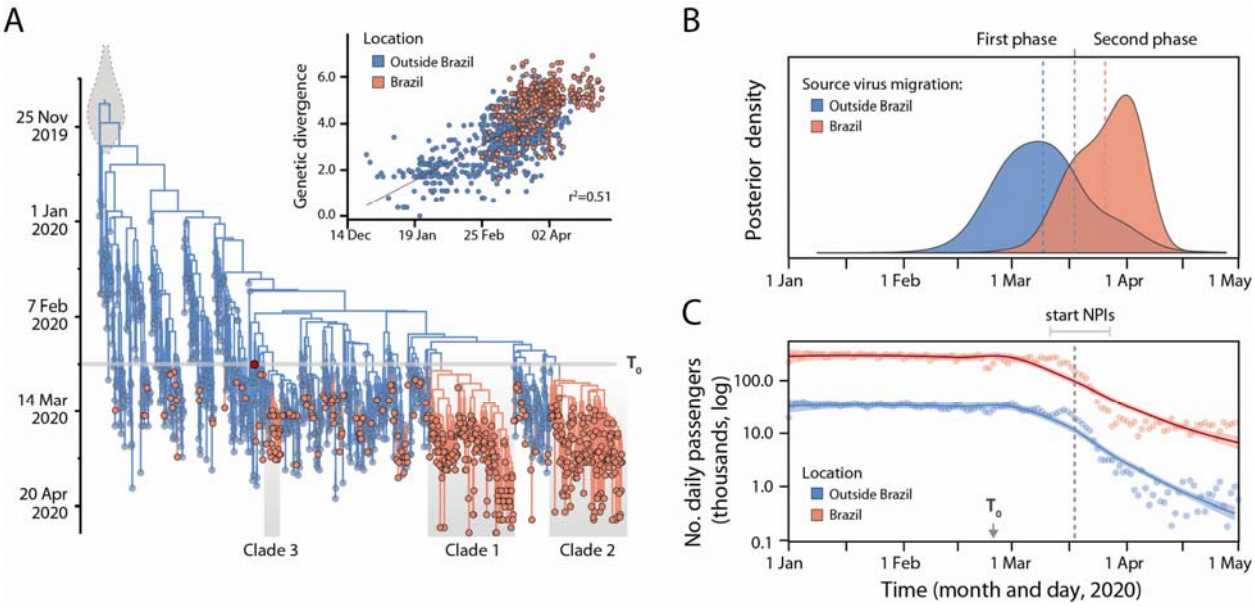

Fig. 3. Evolution and spread of SARS-CoV-2 in Brazil. (**A**) Time-resolved maximum clade credibility phylogeny of 1182 SARS-CoV-2 sequences, 490 from Brazil (salmon) and 692 from outside Brazil (blue). The largest Brazilian clades are highlighted in grey (*Clade 1*, *Clade 2* and *Clade 3*). Detailed dates of emergence for these clades are shown in **Fig. S8**. Inset on the upper left corner shows the root-to-tip regression of genetic divergence against dates of sample collection. (**B**) Dynamics of SARS-CoV-2 import events in Brazil. Dates of international and national (between states) migration events were estimated from sequence data using a phylogeographic approach. The first phase was dominated by virus migrations from outside Brazil while the second phase as marked by virus spread within Brazil. (**C**) Locally estimated scatterplot smoothing (LOESS) of the daily number of international (blue) and national (salmon) air passengers in Brazil in 2020. A detailed annotated phylogeny can be found in **Fig. S8**. T_0 = date of first reported case in Brazil (25 February 2020).

Although metadata on patient's travel history is critical to differentiate international importation and local transmission, such data is only available for the first month of the epidemic in Brazil (*15*). We develop an analysis of genetic data to assess the contribution of international vs. national virus lineage movement events through time (**Fig. 3B**). In the first phase of the epidemic, we find an increasing number of international introductions until the 10 Mar 2020 (**Fig. 2B**). According to line-list epidemiological data on the first 1,468 cases in Brazil (*15*), early cases were predominantly acquired from Italy (26%, *n*=70 of 266 unambiguously identified country of infection) and the USA (28%, *n*=76 of 266) during this period). Our analysis suggests that São Paulo, Minas Gerais, Ceára and Rio de Janeiro states, where the three main clades circulate, received the highest proportion of international imports (**Fig. S9**). After this, the estimated number of international imports decreased concomitantly with the decline in the number of international passengers travelling to Brazil (**Figs. 3B and 3C**, see also **Fig. S13**). In

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

contrast, despite the declines in the number of passengers travelling on national flights (**Fig. 3C**), the number of virus lineage movement events within Brazil continued to increase until around 24 March 2020.

To better understand the contribution of virus spread across spatiotemporal scales in Brazil, we use a continuous phylogeographic model that maps phylogenetic nodes to their inferred origin locations (*33*) (**Fig. 4**). We distinguish branches that remain within a state versus those that cross a state to infer the proportion of within and between state measured virus movement.

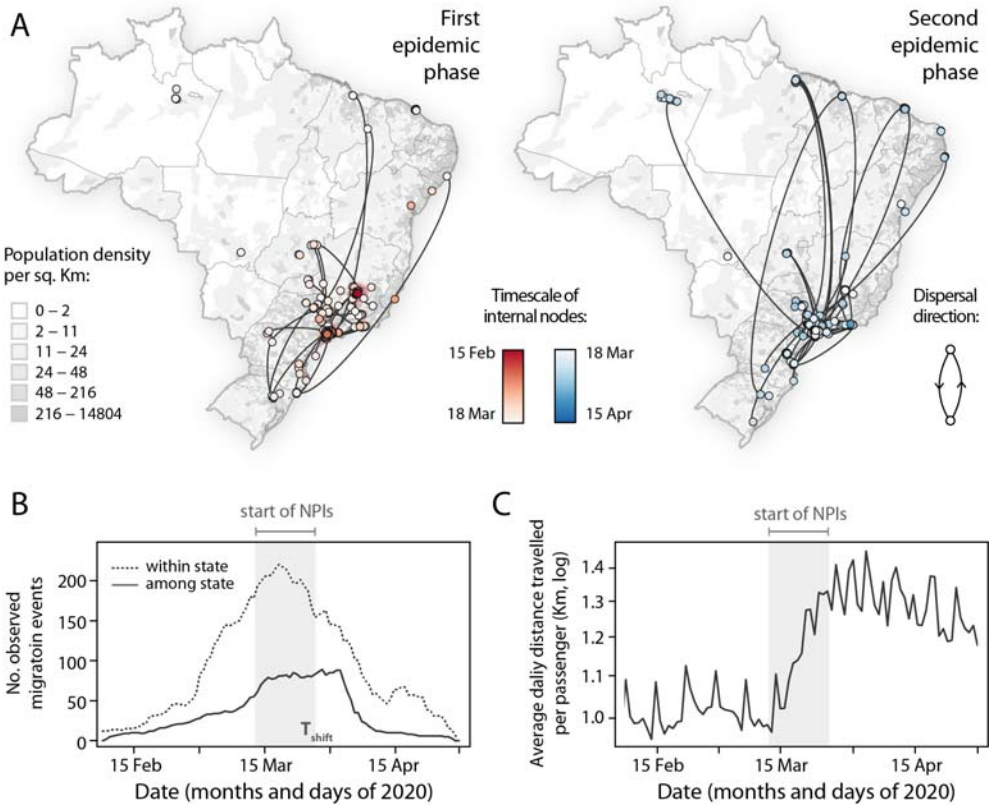

Fig. 4. Spread of SARS-CoV-2 in Brazil. (**A**) Spatiotemporal reconstruction of the spread of Brazilian SARS-CoV-2 clusters containing 3 or more sequences during the first phase (left) and the second epidemic phase (right). Circles represent nodes of the MCC phylogeny and are coloured according to their inferred time of occurrence. Shaded areas represent the 80% high posterior density (HPD) interval and depict the uncertainty of the phylogeographic estimates for each node. Solid curved lines denote the links between sequences and the directionality of movement. Sequences belonging to clusters with less than 3 sequences were also plotted in the map with no lines connecting them. Background population density in 2020 for each municipality was obtained from the Brazilian Institute of Geography (https://www.ibge.gov.br/). **Fig. S14** shows a zoomed version of virus spread in the Southeast region. (**B**) Estimated number of within state and between state virus migrations over time. (**C**) Average distance travelled by an air passenger per day in Brazil calculated using openly available data from the National Civil

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Aviation Agency of Brazil (www.anac.gov.br/en). Light grey boxes indicate starting dates of NPIs across Brazil.

We estimate that during the first epidemic phase, SARS-CoV-2 spread mostly locally and within-state borders. In contrast, the second phase was characterized by long-distance movement events and the ignition of the epidemic outside the southeast region (**Fig. 4A**). Throughout the epidemic, we find that within-state virus movement was on average 2.7-fold more frequent than between-state virus movement. Moreover, our data suggests that within-state virus spread, and to a lesser extent, between-state virus spread, decreased after the implementation of NPIs (**Fig. 4B**). Interestingly, we find that the average route length travelled by passenger increased by 25% during the second phase of the epidemic (**Fig. 4C**), despite a concomitant reduction in the number of passengers flying within Brazil (**Fig. 3C**). These findings emphasize the role of within and between-state mobility as a key driver of both local and inter-regional virus spread, with highly populated and well-connected urban conurbations in southeast region acting as main sources of virus exports within the country (**Fig. S9**).

In conclusion, our study provides the first comprehensive analysis of SARS-CoV-2 spread in Brazil. We found no evidence of community transmission before the first reported case in Brazil, and that large highly connected urban centres drove virus spread across the country. We show that NPIs significantly reduced both virus transmission and spread, but that additional measures are required to maintain transmission at low levels. With recent relaxation of NPIs in Brazil, continued surveillance is required to monitor trends in transmission and the emergence of new lineages in Brazil. Our analysis shows how changes in mobility may impact global and local transmission of SARS-CoV-2, and demonstrates how combining genomic and mobility data can guide traditional surveillance approaches.

References

1. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nature Medicine* **26**, 450-452 (2020).
2. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265-269 (2020).
3. WHO, "Coronavirus disease (COVID-2019) situation reports" (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports, 2020).
4. H. Tian *et al.*, An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 368, 638-642 (2020).
5. M. U. G. Kraemer *et al.*, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 493-497 (2020).
6. A. Rambaut, Holmes, E. C., Hill, V., O'Toole, A., McCrone, J. T., Ruis, C., du Plessis, L., Pybus, O. G., A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* (2020).
7. T. W. Russell *et al.*, Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveillance* 25 (2020).
8. R. Verity *et al.*, Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* 20, 669-677 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

9. J. T. Wu *et al.*, Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine* 26, 506-510 (2020).
10. R. Verity *et al.*, Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* (2020).
11. M. M. Arons *et al.*, Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *The New England Journal of Medicine* 382, 2081-2090 (2020).
12. L. Ferretti *et al.*, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368 (2020).
13. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro surveillance* 25, (2020).
14. Brazilian Ministry of Health. Painel de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil pelo Ministério da Saúde (Available: http://covid.saude.gov.br) (2020).
15. W. M. de Souza et al., Epidemiological and clinical characteristics of the early phase of the COVID-19 epidemic in Brazil. *medRxiv* https://doi.org/10.1101/2020.04.25.20077396 (2020).
16. J. Croda *et al.*, COVID-19 in Brazil: advantages of a socialized unified health system and preparation to contain cases. *Revista da Sociedade Brasileira de Medicina Tropical* 53, e20200167 (2020).
17. L. P. Garcia, E. Duarte, Nonpharmaceutical interventions for tackling the COVID-19 epidemic in Brazil. *Epidemiol Serv Saude* 29, e2020222 (2020).
18. S. B. Oliveira *et al.*, Monitoring social distancing and SARS-CoV-2 transmission in Brazil using cell phone mobility data *medRxiv* doi: https://doi.org/10.1101/2020.04.30.20082172 (2020).
19. S. M. Kissler, "Reductions in commuting mobility predict geographic differences in SARS-CoV-2 prevalence in New York City" (2020).
20. H. J. T. Unwin *et al.*, Report 23: State-level tracking of COVID-19 in the United States (21-05-2020). *https://doi.org/10.25561/79231* (2020).
21. S. Flaxman *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* https://doi.org/10.1038/s41586-020-2405-7, (2020).
22. T. A. Mellan *et al.*, "Report 21 - Estimating COVID-19 cases and reproduction number in Brazil" (2020).
23. Y.-Z. Zhang, Holmes, E. C. "Novel 2019 coronavirus genomes" Virological.org https://virological.org/t/novel-2019-coronavirus-genome/319 (2020).
24. V. M. Corman *et al.*, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveillance* 25, (2020).
25. T. T. N. Thao *et al.*, Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* (2020).
26. M. Cotten *et al.*, Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *Lancet* 382, 1993-2002 (2013).
27. M. Cotten *et al.*, Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* 5 (2014).
28. G. Dudas, Carvalho, L. M., Rambaut A., Bedford, T., MERS-CoV spillover at the camel-human interface. *eLife* https://doi.org/10.7554/eLife.31257.002 (2018).
29. Z. Zhao *et al.*, Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evolutionary Biology* 4, 21 (2004).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.11.20128249; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

30. S. Duchene, Featherstone, L., Hariitopoulou-Sinanidou, M., Rambaut, A., Lemey, P., Baele, G., Temporal signal and the phylodynamic threshold of SARS-CoV-2. *bioRxiv* https://doi.org/10.1101/2020.05.04.077735 (2020).
31. J. Lu *et al.*, Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* (2020).
32. A. S. Gonzalez-Reiche *et al.*, Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* (2020).
33. S. Dellicour *et al.*, A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *bioRxiv* https://doi.org/10.1101/2020.05.05.078758 (2020).

Acknowledgments: A full list acknowledging those involved in the diagnostic and generation of new sequences as part of the CADDE-Genomic-Network can be found in the Supplementary Material. We thank GISAID database for supporting rapid and transparent sharing of genomic data during the COVID-19 pandemic. A full list acknowledging the authors submitting data used in this study can be found in **Data S2**. We thank Lucy Matkin and Josh Quick for logistic support to the CADDE project. **Funding:** This work was supported by the São Paulo Research Foundation (FAPESP) and Medical Research Council CADDE partnership award (MR/S0195/1) (http://caddecentre.org/), FAPESP (2018/14389-0) and the Oxford Martin School. NRF is supported by a Sir Henry Dale Fellowship (204311/Z/16/Z). WMS is supported by FAPESP, Brazil (No. 2017/13981-0). SD is supported by the *Fonds National de la Recherche Scientifique* (FNRS, Belgium). NL, AR and PL are supported by the Wellcome Trust ARTIC network (Collaborators Award 206298/Z/17/Z). PL and AR are supported by the European Research Council (grant agreement no. 725422 – ReservoirDOCS). OJB was funded by a Sir Henry Wellcome Fellowship funded by the Wellcome Trust (206471/Z/17/Z). VHN and CAP were supported by FAPESP (2018/12579-7). JLPM, MRA and CLS are supported by FAPESP (2020/04558-0). MLN is supported by FAPESP (2020/04836-0). LdP is supported by Oxford Martin School; ATRV is supported by CNPq (303170/2017-4) and FAPERJ (26/202.903/20).

Competing interests: Authors declare no competing interests; **Data and materials availability:** All data, code, and materials used in the analysis are available on our Dryad and GitHub repository (upon acceptance).

Supplementary Materials:

Materials and Methods

Figs. S1 to S14

Tables S1 to S3

Captions for Data S1 and Data S2

List of Members of the CADDE-Genomic-Network

Full Reference List