

# **SARS-CoV-2 hospital-associated transmission dynamics in São Paulo, Brazil: a retrospective genomic surveillance study**

Darlan S. Candido MSc<sup>1,2\*</sup>, Ingra M. Claro PhD<sup>2,3\*</sup>, Mariana S. Ramundo PhD<sup>2,3\*</sup>, Alessandra Luna-Muschi MD<sup>2,3</sup>, Bernardo Gutierrez MSc<sup>1,4</sup>, Flavia C. S. Sales BSc<sup>2,3</sup>, Jaqueline G. de Jesus PhD<sup>2,3</sup>, Erika R. Manuli MSc<sup>2,3</sup>, Thaís M. Coletti BSc<sup>2,3</sup>, Camila A. M. da Silva BSc<sup>2,3</sup>, Pamela S. Andrade MSc<sup>2,3,5</sup>, Giulia M. Ferreira MSc<sup>2,3,6</sup>, Esmenia C. Rocha MSc<sup>2,3</sup>, Leandro M. de Souza<sup>2,3</sup>, Mariana C. de Pinho<sup>2,3</sup>, Leonardo J. T. Araujo PhD<sup>2,3,7</sup>, Cecilia Salete Alencar PhD<sup>8</sup>, Rinaldo F. Siciliano MD PhD<sup>9</sup>, Rogério Zeigler MD<sup>9</sup>, Cristhieni Rodrigues MD PhD<sup>9</sup>, Maria Emília B. de Souza<sup>10</sup>, Michely F. Vieira<sup>10</sup>, Raquel Keiko L. Ito MD<sup>3,10</sup>, Thais Guimaraes MD PhD<sup>3,11</sup>, Tania M. V. Strabelli MD PhD<sup>9</sup>, Edson Abdala MD PhD<sup>3,10</sup>, Izabel Oliva Marcilio de Souza MD PhD<sup>12</sup>, Elizabeth de Faria MD<sup>13</sup>, Maura Salaroli Oliveira MD PhD<sup>3,14</sup>, Anna S. Levin MD PhD<sup>3,14</sup>, Ester C. Sabino MD PhD<sup>2,3</sup>, Silvia F. Costa MD PhD<sup>2,3†</sup>, Nuno R. Faria, PhD<sup>1,2,15 †</sup>

1 Department of Zoology, University of Oxford, Oxford, OX1 3SZ, United Kingdom.

2 Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

3 Departamento de Moléstias Infecciosas e Parasitárias, Faculdade de Medicina da Universidade de São Paulo, Sao Paulo, 05403-000, Brazil.

4 Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito USFQ

5 Faculdade de Saúde Pública da Universidade de São Paulo, São Paulo, 01246-904, Brazil.

6 Laboratório de Virologia, Instituto de Ciências Biomédicas, Universidade Federal de Uberlândia, Uberlândia, 38400-902, Brazil.

7 Centro de Patologia, Instituto Adolfo Lutz, São Paulo, 01246-000, Brazil.

8 Laboratório de Medicina Laboratorial, Hospital das Clínicas Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

9 Unidade de Controle de Infecção Hospitalar, Instituto do Coração, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

10 Instituto do Cancer do Estado de São Paulo, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

11 Subcomissão de Controle de Infecção Hospitalar, Instituto Central, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

12 Núcleo de Vigilância Epidemiológica, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

13 Centro de Atendimento ao Colaborador, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil

14 Grupo de Controle de Infecção Hospitalar, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, 05403-000, Brazil.

15 MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, SW7 2AZ, United Kingdom.

**Correspondence to:**

Nuno R. Faria

Email: n.faria@imperial.ac.uk (N.R.F.);

Silvia F. Costa

Email: silviacosta@usp.br

\*Authors contributed equally.

†Authors contributed equally.

## **Abstract**

**Background:** Brazil reported its first SARS-CoV-2 case on 26 February 2020 in an international traveler returning to São Paulo, Brazil. By 10 June 2020, 3,898 healthcare workers (HCW) and patients at the Hospital das Clínicas (HC) in São Paulo, the largest hospital complex in Latin America, had tested positive for SARS-CoV-2 RNA. We aimed to provide insight into the transmission of SARS-CoV-2 in healthcare workers and patients, and within and between HC institutes during the early phase of the epidemic in Brazil.

**Methods:** We analyzed epidemiological data from SARS-CoV-2 RT-PCR confirmed cases between 13 March to 10 June 2020. A total of 340 SARS-CoV-2 genomes were generated from healthcare workers and patients from two HC institutes not receiving COVID-19 patients (institutes A and C) and one institute receiving exclusively COVID-19 patients (institute B). Within- and between-institute transmission clusters were identified and within-cluster transmission dynamics was assessed using logistic regression analysis and a suite of phylogenetic genetic analyses.

**Results:** SARS-CoV-2 weekly incidence in healthcare workers was highest in institutes not receiving COVID-19 patients, and decreased by 75%, 54%, 48% for institutes C, A, B, respectively, after universal masking was adopted. We found a total of 86 hospital-acquired patient infections in HC during the study period, 81·4% ( $n=70$ ) in institute C. Of these, 74.3% of these reported after mandatory universal masking. The average cluster size and cluster duration were larger in non-COVID institutes. Sequences from non-COVID institutes were more likely to cluster together than sequences from institute B (odds ratio, OR=4·17 and OR=3·48, for C and A institutes). The proportion of estimated viral importation events from outside the HC complex to the different HC institutes was highest for institute B (83·64%) compared to institute A (67·85%), and C (57·70%). The statistical support for virus migration from non-COVID institutes A and C to institute B was strong (Bayes factor = 113·7 and 84·4, respectively).

**Interpretation:** The hospital-associated SARS-CoV-2 transmission was higher in non-COVID-19 healthcare institutes compared to a COVID-19 healthcare institute during the first epidemic wave in São Paulo, with our data supporting virus infection non-COVID-19 institutes as a source of infection in a COVID-19 institute,—suggesting that risk perception and compliance was lower in non-COVID institutes.

**Evidence before this study (This is a separate panel required by the journal. It is not included in the final word count)**

On 5 October 2020, we searched PUBMED for studies including the following search terms: ("SARS-CoV-2" OR "COVID-19" OR "coronavirus disease 2019") AND ("genom\*" OR "sequenc\*" OR "WGS") AND ("nosocomial transmission" OR "hospital outbreak" OR "hospital-acquired" OR "healthcare-associated" OR "health-care associated"). No search restrictions were applied. Our search retrieved 62 studies, out of which only 25 applied genomic epidemiology to uncover hospital-associated SARS-CoV-2 transmission, 23 original works, and 2 reviews. Most studies covered the early stages of the pandemic (14, 61%), focused in a single hospital (17, 74%), presented evidence for hospital-associated SARS-CoV-2 transmission (20, 87%), and focused on describing individual clusters rather than general transmission patterns (18, 78%, median sample size= 44, range 3–764). There is a general consensus over the importance of universal masking and genomic epidemiology for hospital-associated outbreak control. Studies looking at general transmission patterns show evidence for most cases being linked to super spreading events and highlight the role of healthcare workers in hospital-associated transmission, although patients might be more likely infected by other patients. No studies used genetic data to investigate transmission patterns and dynamics between and within non-COVID and COVID hospitals.

**Added-value of this study**

We aimed to understand the transmission patterns within and between non-COVID-19 and COVID-19 only institutes that are part of the largest hospital complex in Latin America. We show that hospital-associated SARS-CoV-2 transmission was higher in non-COVID institutes, driven by larger clusters and of longer duration, even after mandatory universal masking. We also

uncover between-hospital transmission events and temporal patterns of hospital-associated transmission.

### **Implications of all the available evidence**

While separating COVID-19 from non- COVID-19 patients in different wards/hospitals and mandatory universal masking reduce HCW cases and prevent larger within-hospital outbreaks, adequate HCW risk perception and adherence are extremely important for the effectiveness of these policies. In times of COVID-19 fatigue, hospitals should work closely with HCW to increase awareness and compliance within and outside of the hospital environment.

## **Introduction**

Brazil reported the first confirmed case of SARS-CoV-2 on 26 February 2020 (1), and has since experienced two large continuous COVID-19 waves. As of 13 January 2021, 22,724,232 SARS-CoV-2 cases and 620,641 deaths attributed to SARS-CoV-2 have been reported in Brazil, the highest numbers in Latin America (2). During this period, the State of São Paulo reported 20% of all cases in Brazil (4,298,180), with 1,422,413 reported cases in the city of São Paulo alone.

Studies conducted during the early stages of the COVID-19 pandemic testing for viral RNA or antibodies have reported variable prevalence of SARS-CoV-2 (1.1% to 9.8%) amongst HCW across different countries (3–10). In a university hospital in the United Kingdom, HCW infection rates were higher amongst HCW from COVID-19-dedicated units (22·6%) compared to those working in non-COVID-19 units (8·6%) or working in multiple wards (11·2%) (11). A retrospective analysis of 435 cases amongst inpatients in a hospital in London found that 66 (15%) were definitely or probably acquired at the hospital (12).

Asymptomatic and pre-symptomatic HCW can become inadvertent vehicles of transmission to other HCW and non-COVID patients (13). However, identifying SARS-CoV-2 transmission clusters remains a challenging task in part because of unidentified asymptomatic cases that may be missed during contact tracing, and in part, because consensus virus genome sequences from a transmission cluster are often identical due to the virus' relatively slow evolutionary rate. These limitations can partly be overcome when epidemiological and genomic data are analyzed jointly (14–17). To date, most studies using genomic epidemiology approaches to identify SARS-CoV-2 within-hospital transmission have focused on the identification and description of specific transmission clusters, rather than understanding dynamics of epidemiologically-linked clusters of transmission (12,18–23). Moreover, analyses comparing the dynamics between and within different hospital units in large hospital complexes remain scarce, particularly in Latin America where the SARS-CoV-2 pandemic hit hardest.

Understanding SARS-CoV-2 hospital-associated transmission in the early stages of an epidemic is of great importance to improve healthcare response and preparedness for future outbreaks. Here we investigate the patterns of SARS-CoV-2 early transmission in HC São Paulo, the largest hospital complex in Latin America, where all COVID-19 patients were hospitalized in a dedicated building by combining insights from epidemiological data, genome sequencing, and phylogenetic analysis.

## **Methods**

### **Epidemiological context**

We performed a retrospective study at a large reference teaching tertiary healthcare complex specialized in high complexity cases called Hospital das Clínicas (HC), affiliated with the University of São Paulo, Brazil. The HC complex has approximately 2,200-beds and 22,000 healthcare workers directly involved in patient care offered in nine specialized institutes. From 30 April 2020 to 02 September 2020, one institute was designated as an exclusive COVID-19 hospital (from here on referred to as Institute B). The other institutes were designated for non-COVID-19 patients. HCW were not allowed to move between buildings. Patients with suspected SARS-CoV-2 infection from non-COVID-19 institutes were maintained in individual rooms until RT-PCR confirmation when they would be transferred to the COVID-19 only institute. Universal masking, here defined as mandatory masking to all hospital staff, was adopted at different epidemiological weeks across the institutes: week 15 (institute A), week 17 (institute C), and week 19 (COVID Institute B and other institutes). Detailed information on COVID measures taken across the 3 institutes can be seen in Appendix p 1-3.

### **Study overview: clinical samples and metadata collection**

This study was approved by the national research ethics commission (Comissão Nacional de Ética em Pesquisa) under protocol number CAAE 30127020·0·0000·0068. Patient and HCW SARS-CoV-2 testing were done at the Hospital Central Clinical Laboratory using real-time quantitative polymerase chain reaction (RT-qPCR) on naso-oropharyngeal swabs (**Corman et al., 2020; Waggoner et al., 2020**). All individuals with RT-qPCR positive samples collected between 13 March to 10 June 2020 were included in this study and results from subsequent tests after the first RT-qPCR positive result were excluded. Clinical and epidemiological data collection included age, sex, home address, occupation, unit of work within the hospital, date of onset of symptoms, symptoms, need for hospitalization, and clinical outcome. Geocoding of

residential Zip codes of patients and staff was done by using Google Maps via the geocode function implemented in the R package ggmap. Additional information such as the complete medical history of patients while in the hospital was retrieved only for clustered patients and health workers.

### **Patient classification**

To classify patients according to the time in days between symptom onset and hospitalization, we adapted the Public Health England (PHE) guidelines (24,25). The PHE classification system considers a 14-day period between a SARS-CoV-2 exposure and COVID-19 symptoms, with an average of five days. Patients were classified into one of four groups: Group 1 (community): symptom onset before hospital admission or up to 2 days after hospital admission; Group 2 (Indeterminate hospital-associated): symptom onset between 3-7 days after hospital admission; Group 3 (Suspected hospital-associated): symptom onset between 8-14 days after hospital admission; Group 4 (Hospital-associated): symptom onset >14 days after hospital admission.

### **Genome Sequencing**

Of a total of 3,898 positive individuals, 454 (12%) samples from Institutes A, B, and C were selected for virus genomic sequencing which was performed at the Institute of Tropical Medicine, University of São Paulo, Brazil. Information on the number of samples sequenced per institute can be seen in appendix p 13. Genome amplification was performed using the n-CoV-2019 Artic protocol (<https://artic.network/ncov-2019>) (see supplementary material) and libraries were sequenced using the Oxford Nanopore Technologies portable genome sequencer, MinION. A genome reference-based assembly pipeline was used for consensus sequence generation with a minimum coverage depth of 20x. See appendix p 3-4.

## Analysis

SARS-CoV-2 sequences from Brazil with collection date up to the 20th May 2020 (most recent date in our HCW dataset) (n=1860) were downloaded from GISAID (26–28) and appended to a previously described global dataset of 1,182 viral genomes. The resulting dataset was aligned to the reference NC\_045512·2 using MAFFT v 7·450 (29) and manually edited using AliView. As previously described (30), we further filtered down our dataset by maintaining only sequences with at least 75% consensus sequence coverage. TempEst v·1·5·3 (31) analyses and visual inspection of the alignment in AliView were used to identify and remove sequences with unusual divergence. No recombination signal was found using RDP 4 (32). Three final datasets were generated: Dataset 1 consisted of 2,550 sequences, including 340 sequences from this study; Dataset 2, sequences with >90% coverage (n=2259); and Dataset 3, a reduced version of Dataset 2, including all HCFMUSP sequences with coverage >90% (n=234) (see supplementary material). Pangolin version V3·1·11 (33) was used for lineage assignment.

Maximum likelihood phylogenies was inferred using IQ-TREE v·2·0 (34) under the best substitution model as determined by ModelFinder (35) implemented in the IQ-TREE pipeline. Bayesian time-rooted phylogenies for Dataset 3 were estimated using BEAST v1·10·4 (36) running with BEAGLE (37) and a discrete phylogeographic approach was used to understand the temporal dynamic of hospital-associated SARS-CoV-2 transmission. Hospital-associated clusters were defined according to the content of HC sequences (sequences from this study) and according to the statistical support obtained from the phylogenetic analysis. Compartmentalization analysis was performed using Simmond's Association Index implemented in the Hypothesis Testing Using Phylogenies (HYPHY) (38), and a binomial logistic regression analyses was performed to identify patient characteristics associated with clustering of genomic sequences. Results were reported as the odds ratio (OR) over the baseline variables and p

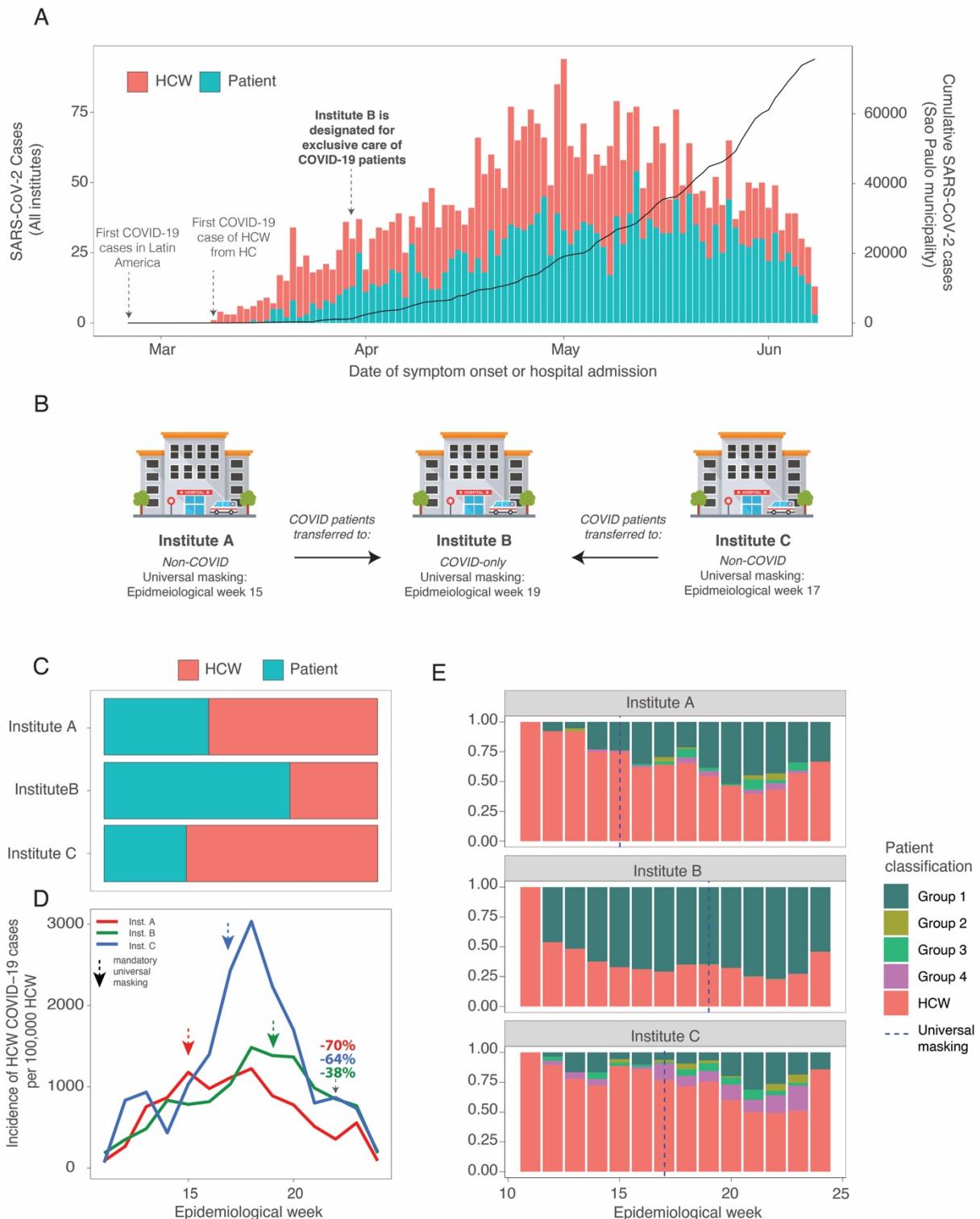
values <0.05 were considered statistically significant. For the household geographical distances analysis, a Mann-Whitney U test was performed in R Studio 1·2·1335. See appendix p 4-9.

## **Results**

### **Epidemiological context**

From 25 February 2020 to 8 June 2020, the municipality of São Paulo, Brazil reported a total of 75,699 COVID-19 cases (Figure 1a). The first SARS-CoV-2 positive cases from HC-FMUSP were reported on epidemiological week 11 (from 9 March 2020), with HCW cases across all nine different institutes. In the following week, the first COVID-19 patient from HC, a community-acquired case, was hospitalized. On 30 March 2020, Institute B was converted into a COVID-19-only institute, while the other institutes were required to transfer COVID-19 patients to Institute B (Figures 1a and 1b).

A total of 12,134 SARS-CoV-2 tests were performed on samples from patients and symptomatic HCW working at the HC-FMUSP hospital complex, of which 3,933 (32%) were positive (appendix p 10). 3,898 SARS-CoV-2 positive individuals were included in this study, 2,008 (51·5%) patients and 1,890 (48·5%) HCW (Figures 1a). A flowchart with the complete study design can be seen in appendix p 10. Given that 91·8% (3,578) of cases from HC were reported by only three institutes - Institute B (2,159 cases, 55·4%), Institute C (716 cases, 18·4%), and Institute A (703 cases, 18%) - we explored the potential differences in the COVID-19 transmission dynamics between non-COVID-19 and COVID-19 institutes (Figure 1B and appendix p 11). A summary of epidemiological and sociodemographic characteristics from all cases in this study can be seen in appendix p 29.



**Figure 1. Epidemiological context of HC SARS-CoV-2 hospital-associated transmission.**

(A) Time series of COVID-19 positive cases across all institutes from Hospital das Clínicas (HC-

FMUSP) and cumulative COVID-19 cases for the municipality of São Paulo. Colors depict whether cases occurred in HCW (red) or patients (blue). The dotted line marks the date of adoption of universal masking. **(B)** Proportion of cases from A, B, and C institutes stratified by HCW/patient. **(C)** Incidence of HCW COVID-19 cases from Institute A (red), Institute B (green), and Institute C (blue) per epidemiological week. The dotted line marks the epidemiological week of adoption of universal masking in each institute. Percentages represent the reduction in the incidence of HCW COVID-19 cases for each institute at week 23, having week 18 as a reference (the week before universal masking was implemented). Percentage colors follow the pattern for line colors and represent the different institutes. **(D)** Proportion of patients according to time between the onset of symptoms and hospitalization. Patients were discretized into four groups: group 1 (community-acquired), group 2 (indeterminate acquisition), group 3 (suspected hospital transmission), group 4 (hospital-acquired) (see methods).

### Epidemiological evidence for hospital-associated transmission

Most COVID-19 cases in non-COVID-19 institutes occurred in HCW, 61·7% in Institute A and 70% in Institute C; in contrast to 32% in Institute B (Figure 1b). Incidence of HCW cases was also higher amongst non-COVID-19 institutes in most epidemiological weeks before the enforcement of universal masking, especially for Institute C, reaching a peak incidence of 3,033/100,000 HCW (Figure 1D). Incidence of HCW cases decreased by 70% in Institute A, 64% in Institute C, and 38% in Institute B, after 7, 5, and 3 weeks of the implementation of universal masking, respectively - note that universal masking was adopted at different epidemiological weeks: 15 (A), 17 (C) and 19 (B) (Figure 1C).

We used the time gap between symptom onset and patient hospitalization as a proxy for SARS-CoV-2 hospital transmission and categorized patients into four groups (see methods). A total of

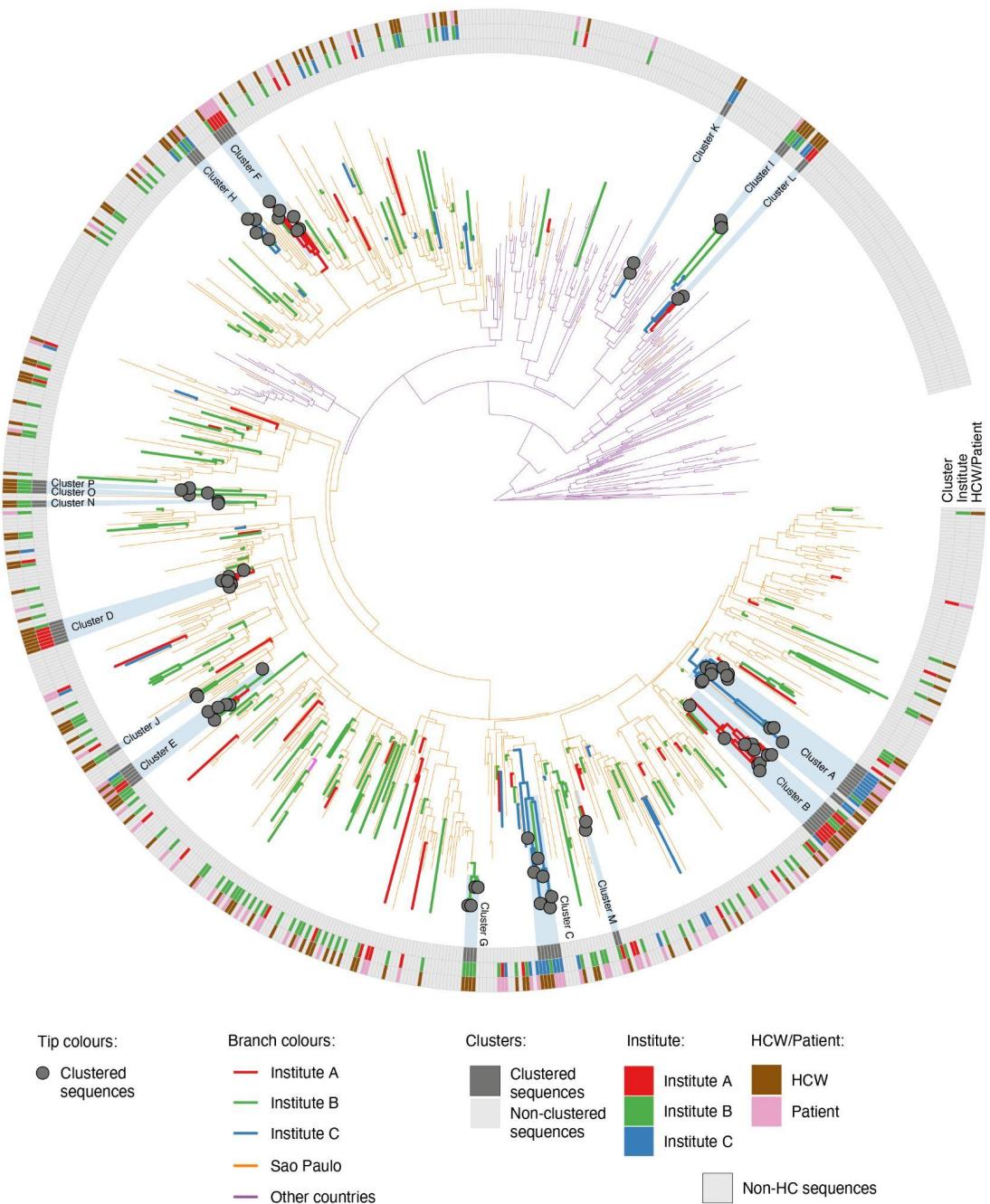
167 patients (8·55 %, total of 1,952 patients) from the three institutes had symptom onset >2 days after hospitalization (groups 2-4) and were possibly linked to hospital-associated transmission (Figure 1d and S3a): 27 (16·2%) belonged to group 2 (indeterminate acquisition), 54 (32·3%) to group 3 (suspected hospital transmission), and 86 (51·5%) to group 4 (hospital-acquired). The majority of group 4 COVID-19 cases occurred at Institute C, 70 (81·4%), while 15 (17·4%) occurred at Institute A, and one at Institute B (1·2%) (Figure 2a, appendix p 32). 52 (74.3%) group 4 cases from Institute C were reported after universal masking was already mandatory (Fig. 1D). The time between hospitalization and symptom onset for group 4 patients ranged from 15 to 175 days (mean=42·5 days, median=27·5 days) (appendix p 12).

### **Hospital-associated SARS-CoV-2 genetic diversity and clustering**

To further support the epidemiological evidence of hospital-associated transmission and characterize its dynamics, we randomly selected 454 SARS-CoV-2 positive samples (hospital workers and patients). From those, we obtained a total of 340 SARS-CoV-2 genomes with coverage >75%, approximately 10% of all reported cases (67 new GISAID submissions, appendix p 13 and 34-45). Most sequences from the three institutes belonged to lineage B·1·1·28, followed by lineages B·1·1·33, B·1 and B·1·1, the main lineages circulating in São Paulo at the time (appendix p 14).

Using datasets B and C (>90% coverage sequences), we were able to identify 16 clusters potentially associated with hospital transmission in the three institutes, comprised by a total of 73 sequences (Figure 2 and appendix p 46). Cluster size ranged from 2-12 sequences from this study, with an average cluster size of 4·6 sequences (median=3·5) (Figures 3a and b, appendix p 47). Within-cluster diversity was on average 1·22 SNPs (median=1, range 0-6 SNPs) (Figure 3D). We have also found maximum within-cluster pairwise diversity to be correlated to cluster

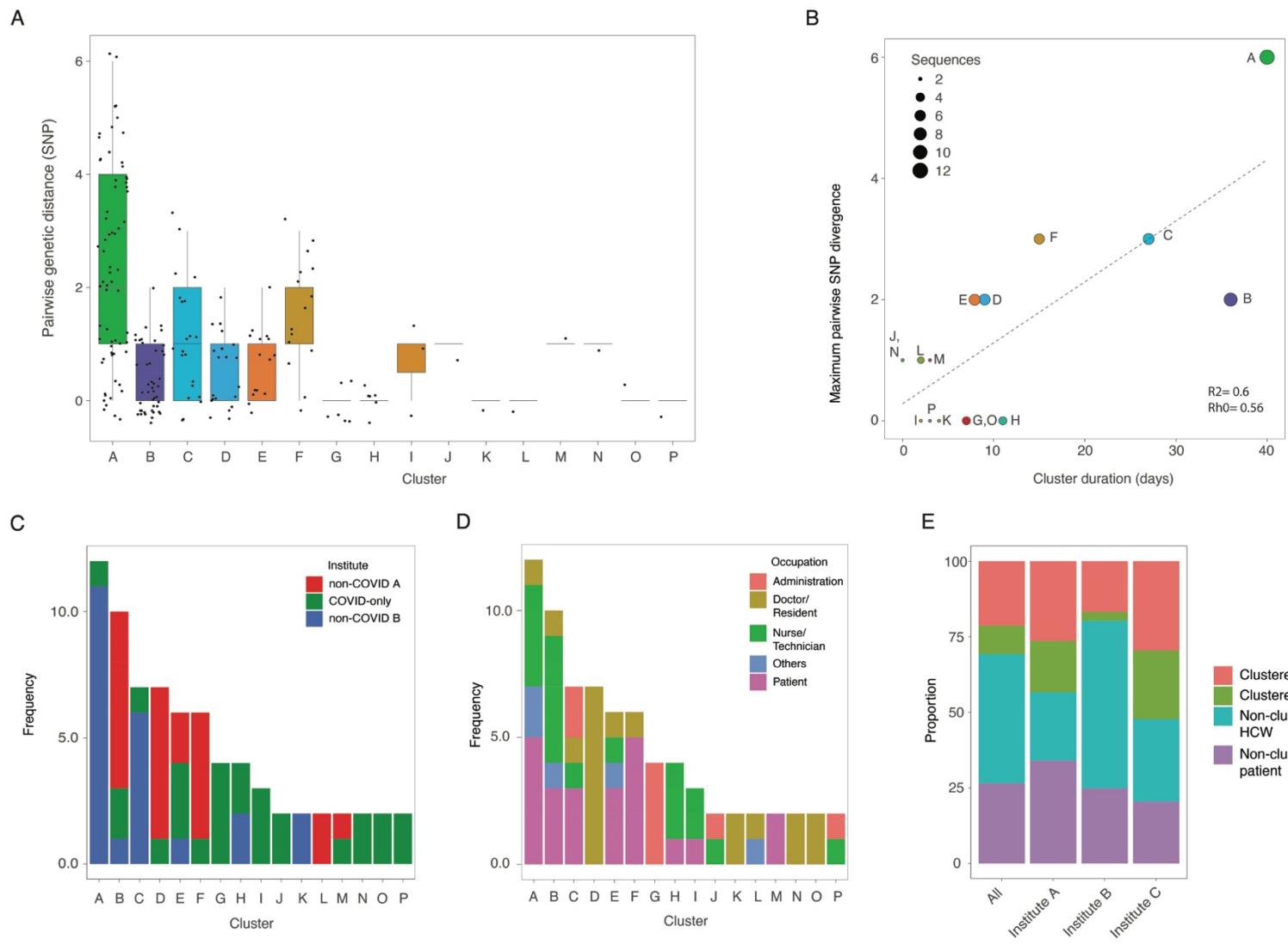
duration (days),  $R^2=0.6$ , and Spearman's rho = 0.56 (Figure 2B). Most clusters, 12 (75%), were defined by one single mutation (appendix p 49).



**Figure 2. Hospital-associated SARS-CoV-2 transmission clusters.** Time-stamped maximum clade credibility phylogeny inferred from dataset 3 (841 sequences), including 234 HC sequences (see methods). Regression of root-to-tip genetic divergence against sampling dates

retrieved an  $R^2$  of 0·57 (appendix p 15). Tips are colored according to hospital-associated transmission clusters and branches are colored according to inferred node location (Institutes A, B and C, São Paulo state, and Other). Heatmaps depict the institute of collection for each HC sample and information on whether a sample belonged to an HCW or a patient. An expanded version of Figure 2 can be found in appendix 2.

Most clusters, 13 (81·25%), were mostly composed ( $\geq 70\%$ ) by sequences from a single institute (Figure 3C). Most clustered sequences belonged to HCW, 50 (68·5%) (figure 3D). Half of the clusters contained only sequences from HCW, while no clusters consisted of sequences from patients only. We also observe that only 19·7% of the sequences from Institute B are clustered; this proportion was 43·4% and 50% for Institute A and Institute C, respectively (Figure 4C). Moreover, despite presenting the largest number of clusters ( $n=6$ ), Institute B clusters had the smallest average size, 2·5 sequences (median=2, range 2-4 sequences), and duration, 2·8 days (median=1·5 days, range 0-7 days), while Institute C had the smallest number of clusters ( $n=3$ ), but the highest average cluster size (median=7, range 2-12 sequences), and the longest average duration, 23·7 days (median=27 days, range 4-40 days). Other cluster characteristics can be seen in appendix p 17 and 47.



**Figure 3. Characteristics of the 16 phylogenetic clusters potentially associated with hospital transmission.** (A) Pairwise genetic distances of sequences in each cluster. (B) Correlation between cluster maximum pairwise genetic divergence and cluster duration. (C) Frequency of sequences in each cluster according to the institute of origin. (D) Frequency of sequences in each cluster according to occupation. (E) Proportion of sequences per institute according to clustering status. Proportion was calculated considering a total of 234 sequences with coverage >90% used for cluster analysis.

#### **Factors linked to Hospital-associated clustering**

To explore the differences between clustered ( $n=X$ ) and nonclustered sequences ( $n=Y$ ) across the three institutes, we used logistic regression models to assess predictors that would best explain such patterns. Model 1 revealed that sequences from non-COVID-19 Institutes C and A were at greater odds for clustering than sequences from Institute B (COVID-only) (OR=4·46; and OR=3·59, respectively) (Table 1). HCW from both Institutes A and C were nine times more likely to cluster than patients from Institute B (OR=9·92; OR=9·21, model 2), while Institute B HCW have a non-significant tendency for higher odds of clustering (OR=2·72,  $p=0·08$ ). Across all institutes, medical residents were the only occupation at greater odds for clustering compared to patients (OR=4·10,  $p=0·0007$ , appendix p 51-53, model 3). OR for different occupations in each institute can be seen in appendix p 51-53.

Table 1. Age and Sex-adjusted Odds ratio and p-values for clustering logistic models 1 and 2

Logistic Model Parameters	Level	aOdds Ratio	p-value
Model 1:			
Variables: Institute + HCW/patient +			
Age + Sex			
Base level: Institute B; Patient	(Intercept)	0·17	<b>0·002</b>
	Institute A	3·48	<b>0·00074</b>
	Institute C	4·17	<b>0·0002</b>
	HCW	1·63	0·21
Model 2:			
Variables: HCW/patient per Institute			
+ Age + Sex			
Base level: Patient.Institute B*	(Intercept)	0·11	<b>0·0034</b>
	HCW.Institute A	7·95	<b>0·0033</b>
	HCW.Institute B*	2·36	<b>0·017</b>
	HCW.Institute C	7·49	<b>0·0043</b>

	Patient.Institute A	4·45	<b>0·027</b>
	Patient.Institute C	7·43	<b>0·0050</b>

HCW: healthcare worker; \* Institute B was COVID-19-only.

To assess the degree of compartmentalization of clustered sequences given different traits, we used the tree-based method Simmonds Association Index (AI). When the analysis was performed on all clustered sequences ( $n=72$ ), a compartmentalization signal was identified for the trait institute ( $AI = 0\cdot45$ , BS=1000). Although all institutes are contributing to the signal, Institute A ( $AI = 0\cdot37$ , BS=1000) and Institute C ( $AI = 0\cdot44$ , BS=999) have higher degrees of population structure when compared to Institute B ( $AI = 0\cdot56$ , BS=998). However, analysis of sequences from each institute separately revealed a compartmentalization signal for traits HCW vs Patients ( $AI = 0\cdot36$ , BS=999) and occupation ( $AI = 0\cdot42$ , BS= 1000) for Institute A sequences only (appendix p 54).

### **Dynamics of hospital-associated SARS-CoV-2 transmission**

We next used genomic data to infer the proportion of imported cases from each Institute, and in turn, infer the extent to which hospital-associated transmission happened within each institute. Time-measured phylogeographic analysis performed on dataset 3 revealed that Institute B had the highest proportion of import-associated cases (from São Paulo, international or other institutes), 83·64% (BCI: 76·64 to 92·70%), followed by Institute A, 67·85% (BCI: 60·38% to 71·70%), and Institute C, 57·70% (BCI: 52·23% to 63·63%) (Figure 4a). These data also imply that Institute A and, especially Institute C, would have had a higher degree of hospital-associated SARS-CoV-2 transmission than Institute B. To validate our results and take potential sampling bias into account, we also estimated the expected percent of imports for

each institute by randomly reshuffling the institute assigned to each sequence. Averages for ten simulations and individual runs are shown in appendix p 18-19, and confirm that the expected average percent of imports for Institute C and Institute A should be higher than the observed while that of Institute should be lower. Similar results are observed when sequences were discretized as from a HCW or patient (appendix p 20-22).

We also identified the location transition rates with strongest statistical support (Bayes Factor > 10) (39). Interestingly, strong statistical support was found for transitions from non-COVID institutes A and C to Institute B (Bayes factor = 113·7 and 84·4, respectively) (Figure 4b). Information on all location transition counts, rates, and Bayes factors can be found in appendix p 55. Temporally, transitions from institutes A and B into institute C peaked 3 weeks after the peak of São Paulo imports into these institutes (Figure 4C). Similar results can be observed for transitions between HCW and patients (Figure 4D).

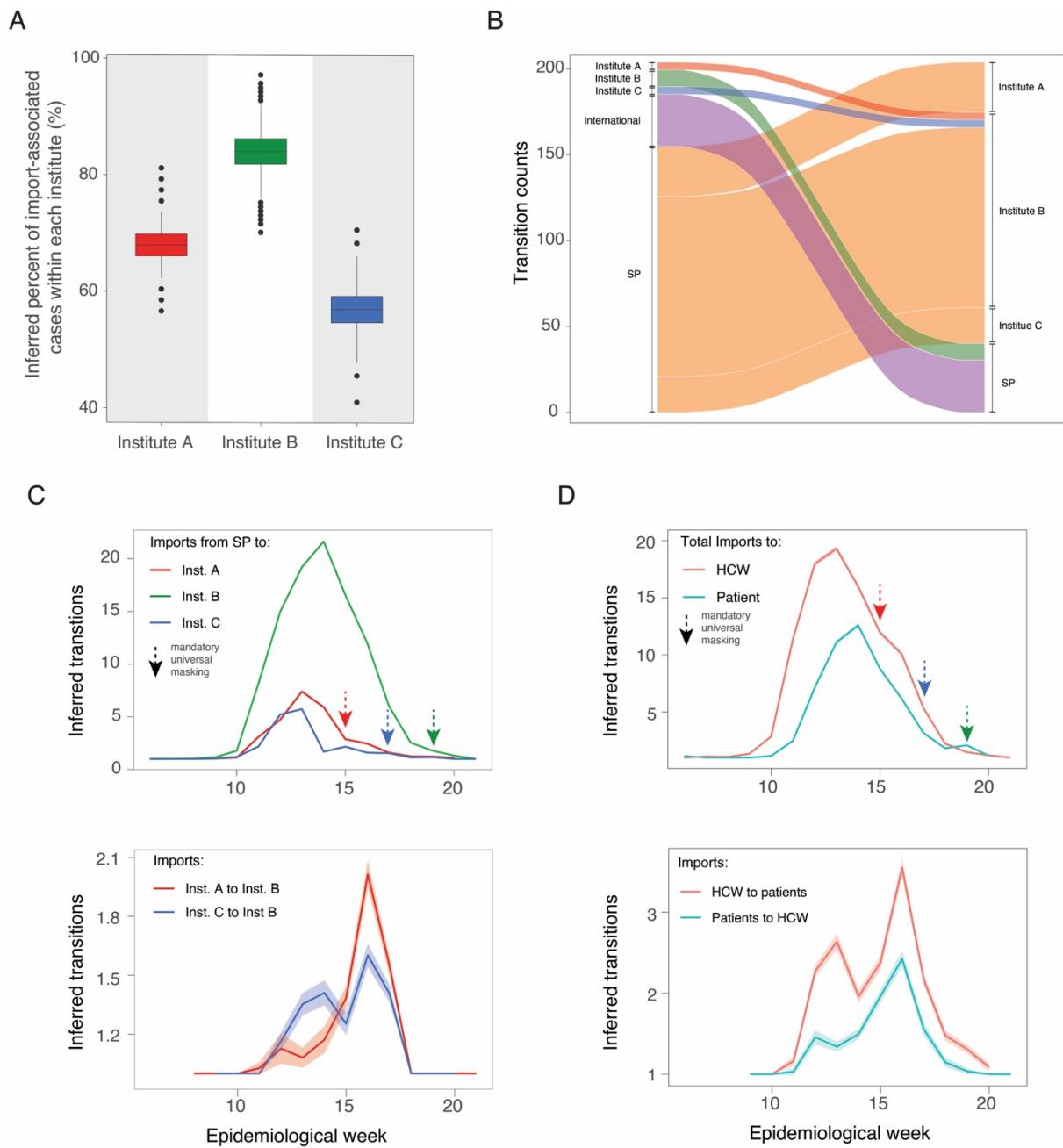


Figure 4. Proportion of SARS-CoV-2 imported cases in HC institutes and inferred transitions into and between institutes. (A) Proportion (%) of inferred total imports to Institutes A, B, and C. (B) Location transition counts (Markov jumps) for transition rates with strong statistical support ( $BF>10$ ). Alluvial plots are proportional to the Markov jumps counts for each specific location transition. Colors identify the location of origin for each transition: Institute A (red), Institute B (green), non-COVID-B (blue), Other (purple), São Paulo (orange). (C) Inferred transitions

(Markov Jumps) from São Paulo to each institute over time (up) and between institutes (down).

(D) Inferred total imported HCW and patient cases (up) and between HCW and Patients (down).

### **Epidemiological links of hospital-associated transmission clusters**

We categorized the epidemiological link for each individual as strong, possible, and unclear (see methods). We find that 31 (42·5%) of the individuals had a strong epidemiological link with at least one other individual from the cluster, 31 (42·5%) had a possible link, and 11 (15%) had an unclear link (appendix p 23 and 47). Full information on epidemiological links for each cluster can be seen in appendix p 24-27. To exclude household interaction between clustered individuals, we used the pairwise distance between sequences in the same cluster and compared it to non-clustered sequences. Clustered sequences tended to come from cases residing in slightly more distant households (median=18·26 km) compared to non-clustered sequences (median= 16·2 km) (p-value=0·07, Mann-Whitney test) - (appendix p 28 and 47-50).

## **Discussion**

While most COVID-19 hospital-associated transmission studies have focused on single hospitals or on mixed hospital data, we provide the first comparison of SARS-CoV-2 transmission across COVID-19 and non-COVID-19 hospitals during the early stages of the pandemic. We provide evidence of higher hospital-associated SARS-CoV-2 transmission in non-COVID-19 hospitals compared to a COVID-19-only hospital using different types of data and analyses. First, we suggest that universal masking was effective in decreasing infections across HCW but not hospital-acquired patient cases in non-COVID-19 institutes. Secondly, we show that positive cases from non-COVID-19 institutes in general and of HCW-only are more likely to be part of a transmission cluster than those from the COVID-19-only hospital. In addition, we estimate that a smaller proportion of the cases in non-COVID-19 institutes were acquired outside of the hospital. Finally, our genetic analyses further identified some level of virus transmission from non-COVID-19 institutes to the COVID-19-only.

Several studies have shown that HCW are at higher risk of COVID-19 infection than the broader community (38,40–43) and have an important role in seeding and amplifying nosocomial SARS-CoV-2 outbreaks to other HCW and patients (44). However, most of this evidence was generated prior to implementation of universal masking, which is effective in reducing the HCW risk of SARS-CoV-2 hospital-acquired infection (45). In turn, our study period includes the early stages of the pandemic and overlaps the progressive implementation of universal masking. We find that incidence amongst HCW was much higher in non-COVID-19 institutes, especially institute C. Considering that universal masking was implemented first at non-COVID institutes (A and C), and assuming that outside work exposure was the same for HCW from the three institutes, these differences could be explained by differences in behavior and risk perception of SARS-CoV-2 among HCW. Adherence to protective measures is correlated to risk perception and HCW tend to associate a higher risk of exposure to contact with infected patients rather

than other HCW (46) (47). However, two studies conducted in São Paulo concluded that HCW who directly provided care to COVID-19 patients were not at higher risk of infection (48,49). During the first COVID-19 wave in the UK, transmission between HCW was the most common form of nosocomial COVID-19 infection (50). Evidence suggests that after the implementation of universal masking, most of the HCW COVID-19 cases were associated with transmission between HCW rather than contact with an infected patient (50) (51). Thus, one of the possible explanations to our findings is that risk perception was higher amongst HCW from institute B, given the awareness of dealing with COVID-19 patients.

Although universal masking was effective in reducing infection amongst HCW, hospital-associated patient cases were still common in institutes A and C. This suggests that patient-to-patient transmission might have also played a role in the nosocomial transmission dynamics. In fact, evidence suggests that patients with hospital-acquired COVID-19 infections are more likely to get infected through contact with other patients in super-spreading events rather than through contact with HCW (52), especially if contacts also had hospital-acquired infections (53). Moreover, the proportion of patient-to-patient transmission almost doubled in the second wave in the UK and became the most common form of COVID-19 nosocomial transmission (50).

Given that Institute B was a COVID-19-only institute, drawing any conclusions on COVID-19 transmission from epidemiological data alone would be challenging, as patients from this institute were already infected at hospitalization and HCW could have been infected outside the hospital. To overcome this limitation, we used SARS-CoV-2 genome sequences from patients and HCW to infer SARS-CoV-2 transmission dynamics. Transmission clusters were larger in the non-COVID-19 institutes, suggesting that transmission in non-COVID-19 institutes involved sustained onward transmission for longer periods. We also showed that individuals from non-COVID-19 institutes were more likely to be part of a transmission cluster, especially if they were

HCW. These findings most likely reflect the within-hospital interactions of HCW, but potentially the social interactions outside work as well. Lunch and smoking breaks are common situations in which unprotected interaction between HCW has been documented (51,54). HCW cases have also been shown to result from unprotected interactions with other HCW in the community (55). Interestingly, our genetic analysis supports these findings when showing that a higher proportion of cases in non-COVID institutes should be associated with hospital transmission. These differences were observed for both HCW and patients.

While evidence for compartmentalization was observed at the institute level, significant links from non-COVID institutes to institute B were inferred from our genetic analysis. Since HCW could not transit between institutes, these inter-institute transmission events could be explained by patients from Institute A and C being transferred to Institute B and/or by HCW potentially interacting with HCW from other institutes outside of the hospital setting. Transportation of patients from non-COVID-19 hospitals to COVID-19 hospitals is critical and specific protocols should be in place to ensure patient and HCW safety (56,57).

Our study has several limitations. Firstly, this is a retrospective study and, as such, it faced limitations regarding access to samples and full metadata patients and sequences. Secondly, 6·5% (n=234, >90% coverage) of all cases were used for cluster analysis to reduce the chances of poor phylogenetic placement. Although this number represents one in every 15 cases, we have likely missed some transmission clusters, especially smaller ones, and intermediate transmission events. In addition, SARS-CoV-2 has a relatively low mutation rate (30) and it is possible that some phylogenetically related sequences might not be an immediate part of the same transmission chain, especially in clusters to which no epidemiological link could be observed. Finally, given that most of our sequencing sampling dates back to before universal

masking was implemented, we were not able to assess the impact of universal masking using genomic data.

By integrating genomic and epidemiological surveillance, hospitals can identify and understand outbreaks and inform targeted infection control interventions. At a time in which new variants are constantly arising and inaccurate risk perception and COVID fatigue become increasingly relevant issues, it is important to emphasize that HCW can become vectors of transmission to other HCW and to non-COVID-19 patients, and interventions towards improving compliance to protective measures should be implemented. Masks should be worn at all times when social distancing is not possible, not only when in contact with COVID-19 patients, including outside and on the way to the hospital. Finally, tighter protective measures (e.g., continuous HCW testing, restriction of visitors, immediate isolation of suspected patients) should also be enforced in non-COVID-19 hospitals, as community introductions can easily become within-hospital outbreaks.

## Funding

Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0) (<http://caddecentre.org/>). N.R.F.: Sir Henry Dale Fellowship: 204311/Z/16/Z and Medical Research Council-Sao Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0); and Bill & Melinda Gates Foundation (INV-034540). DSC is supported by the Clarendon Fund, Oxford University Zoology Department and Merton College. FAPESP supports IMC (2018/17176-8), FCSS (2018/25468-9), JGJ (2019/12000-1), TMC (2019/07544-2), CAMS (2019/21301-5) and ECS (18/14389-0). FFMUSP supports MSR (FFMUSP No. 206.706). The Tropical Medicine Institute from the São Paulo University supports ECR, LMS, MCP, LJTA. CAPES supports PSA (88887.596940/2021-00) and GMF (88887.571465/2020-00).

## **Contributors**

**Conception:** DSC, IMC, MSR, ASL, ECS, NRF, SFC. **Acquisition:** DSC, IMC, MSR, ALM, FCSS, JGJ, ERM, TMC, CAMS, PSA, GMF, ECR, LMS, MCP, LJTA, CSA, RFS, RZ, CR, MEBS, MFV, RKLI, TG, TMS, EA, IOMS, EF, MSO, ASL, ECS, NRF, SFC. **Analysis:** DSC, IMC, MSR, ALM, BG, ECS, NRF, SFC. **Interpretation:** DSC, IMC, MSR, ALM, BG, ASL, ECS, NRF, SFC. **Drafting:** DSC, IMC, MSR, ALM, BG, ECS, NRF, SFC. **Revising:** All authors. Data and materials availability: All data, code, and materials used in the analysis are available in a dedicated GitHub Repository.

## **Declaration of interests**

We declare no competing interests.

## **Acknowledgments**

We thank L. Matkin (University of Oxford) for the incredible support throughout the development of this work. We also thank the contributions of the Núcleo de Vigilância Epidemiológica (NUVE) of the HCFMUSP. We thank the reviewers for their comments and suggestions and GISAID for the essential work of sharing genomic data for the global community and making this work possible. A list with full acknowledgments to all the authors publishing genomic data used in this study can be found in Data S1.

## **Supplementary Material**

Supplementary appendix.

## References

1. Jesus JG de, Sacchi C, Candido D da S, Claro IM, Sales FCS, Manuli ER, et al. Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev Inst Med Trop Sao Paulo.* 2020 May 11;62:e30.
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020 Feb 19;20(5):533–4.
3. Kluytmans-van den Bergh MFQ, Buiting AGM, Pas SD, Bentvelsen RG, van den Bijllaardt W, van Oudheusden AJG, et al. Prevalence and clinical presentation of health care workers with symptoms of coronavirus disease 2019 in 2 dutch hospitals during an early phase of the pandemic. *JAMA Netw Open.* 2020 May 1;3(5):e209673.
4. Mutambudzi M, Niedzwiedz C, Macdonald EB, Leyland A, Mair F, Anderson J, et al. Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants. *Occup Environ Med.* 2020 Dec 9;
5. Barrett ES, Horton DB, Roy J, Gennaro ML, Brooks A, Tischfield J, et al. Prevalence of SARS-CoV-2 infection in previously undiagnosed health care workers in New Jersey, at the onset of the U.S. COVID-19 pandemic. *BMC Infect Dis.* 2020 Nov 16;20(1):853.
6. Jeremias A, Nguyen J, Levine J, Pollack S, Engellenner W, Thakore A, et al. Prevalence of SARS-CoV-2 Infection Among Health Care Workers in a Tertiary Community Hospital. *JAMA Intern Med.* 2020 Dec 1;180(12):1707–9.
7. Vahidy FS, Bernard DW, Boom ML, Drews AL, Christensen P, Finkelstein J, et al. Prevalence of SARS-CoV-2 Infection Among Asymptomatic Health Care Workers in the Greater Houston, Texas, Area. *JAMA Netw Open.* 2020 Jul 1;3(7):e2016451.
8. Mani NS, Budak JZ, Lan KF, Bryson-Cahn C, Zelikoff A, Barker GEC, et al. Prevalence of coronavirus disease 2019 infection and outcomes among symptomatic healthcare workers in seattle, washington. *Clin Infect Dis.* 2020 Dec 17;71(10):2702–7.
9. Lai X, Wang M, Qin C, Tan L, Ran L, Chen D, et al. Coronavirus Disease 2019 (COVID-2019) Infection Among Health Care Workers and Implications for Prevention Measures in a Tertiary Hospital in Wuhan, China. *JAMA Netw Open.* 2020 May 1;3(5):e209666.
10. El-Boghdadly K, Wong DJN, Owen R, Neuman MD, Pocock S, Carlisle JB, et al. Risks to healthcare workers following tracheal intubation of patients with COVID-19: a prospective international multicentre cohort study. *Anaesthesia.* 2020 Nov;75(11):1437–47.
11. Eyre DW, Lumley SF, O'Donnell D, Campbell M, Sims E, Lawson E, et al. Differential occupational risks to healthcare workers from SARS-CoV-2 observed during a prospective observational study. *eLife.* 2020 Aug 21;9.
12. Rickman HM, Rampling T, Shaw K, Martinez-Garcia G, Hail L, Coen P, et al. Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital. *Clin Infect Dis.* 2021 Feb 16;72(4):690–3.

13. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, et al. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility - King County, Washington, March 2020. *MMWR Morb Mortal Wkly Rep.* 2020 Apr 3;69(13):377–81.
14. Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, et al. Phylodynamic applications in 21st century global infectious disease research. *Glob Health Res Policy.* 2017 May 8;2:13.
15. Aggarwal D, Myers R, Hamilton WL, Bharucha T, Tumelty NM, Brown CS, et al. The role of viral genomics in understanding COVID-19 outbreaks in long-term care facilities. *Lancet Microbe.* 2021 Sep 29;
16. Hill V, Ruis C, Bajaj S, Pybus OG, Kraemer MUG. Progress and challenges in virus genomic epidemiology. *Trends Parasitol.* 2021 Dec;37(12):1038–49.
17. Walker A, Houwaart T, Finzer P, Ehlkes L, Tyshaieva A, Damagnez M, et al. Characterization of SARS-CoV-2 infection clusters based on integrated genomic surveillance, outbreak analysis and contact tracing in an urban setting. *Clin Infect Dis.* 2021 Jun 28;
18. Wong RCW, Lee MKP, Siu GKH, Lee LK, Leung JSL, Leung ECM, et al. Healthcare workers acquired COVID-19 disease from patients? An investigation by phylogenomics. *J Hosp Infect.* 2021 Sep;115:59–63.
19. Myhrman S, Olausson J, Ringlander J, Gustavsson L, Jakobsson HE, Sansone M, et al. Unexpected details regarding nosocomial transmission revealed by whole-genome sequencing of severe acute respiratory coronavirus virus 2 (SARS-CoV-2). *Infect Control Hosp Epidemiol.* 2021 Aug 20;1–5.
20. Cheng VC-C, Fung KS-C, Siu GK-H, Wong S-C, Cheng LS-K, Wong M-S, et al. Nosocomial outbreak of COVID-19 by possible airborne transmission leading to a superspreading event. *Clin Infect Dis.* 2021 Apr 14;
21. Løvestad AH, Jørgensen SB, Handal N, Ambur OH, Aamot HV. Investigation of intra-hospital SARS-CoV-2 transmission using nanopore whole-genome sequencing. *J Hosp Infect.* 2021 May;111:107–16.
22. Lucey M, Macori G, Mullane N, Sutton-Fitzpatrick U, Gonzalez G, Coughlan S, et al. Whole-genome Sequencing to Track Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission in Nosocomial Outbreaks. *Clin Infect Dis.* 2021 Jun 1;72(11):e727–35.
23. Sikkema RS, Pas SD, Nieuwenhuijse DF, O'Toole Á, Verweij J, van der Linden A, et al. COVID-19 in health-care workers in three hospitals in the south of the Netherlands: a cross-sectional study. *Lancet Infect Dis.* 2020 Nov;20(11):1273–80.
24. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hossmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis.* 2020 Nov;20(11):1263–72.

25. Hamilton WL, Tonkin-Hill G, Smith ER, Aggarwal D, Houldcroft CJ, Warne B, et al. Genomic epidemiology of COVID-19 in care homes in the east of England. *eLife*. 2021 Mar 2;10.
26. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill*. 2017 Mar 30;22(13):30494.
27. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*. 2017 Jan;1(1):33–46.
28. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. Gisaid's role in pandemic response. *China CDC Wkly*. 2021 Dec 3;3(49):1049–51.
29. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013 Apr;30(4):772–80.
30. Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. 2020 Sep 4;369(6508):1255–60.
31. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016 Jan;2(1):vew007.
32. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*. 2015 May 26;1(1):vev003.
33. Rambaut A, Holmes EC, Hill V, O'Toole A, McCrone J, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *BioRxiv*. 2020 Apr 19;
34. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*. 2020 May 1;37(5):1530–4.
35. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017 Jun;14(6):587–9.
36. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018 Jan;4(1):vey016.
37. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol*. 2012 Jan;61(1):170–3.
38. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005 Mar 1;21(5):676–9.

39. Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, Nourlil J, et al. Phylodynamics and human-mediated dispersal of a zoonotic virus. *PLoS Pathog.* 2010 Oct 28;6(10):e1001166.
40. Fenton L, Gribben C, Caldwell D, Colville S, Bishop J, Reid M, et al. Risk of hospital admission with covid-19 among teachers compared with healthcare workers and other adults of working age in Scotland, March 2020 to July 2021: population based case-control study. *BMJ.* 2021 Sep 1;374:n2060.
41. Shah ASV, Wood R, Gribben C, Caldwell D, Bishop J, Weir A, et al. Risk of hospital admission with coronavirus disease 2019 in healthcare workers and their households: nationwide linkage cohort study. *BMJ.* 2020 Oct 28;371:m3582.
42. Koh D. Occupational risks for COVID-19 infection. *Occup Med (Lond).* 2020 Mar 12;70(1):3–5.
43. Riley S, Ainslie KEC, Eales O, Jeffrey B, Walters CE, Atchison CJ, et al. Community prevalence of SARS-CoV-2 virus in England during May 2020: REACT study. *medRxiv.* 2020 Jul 11;
44. Abbas M, Robalo Nunes T, Cori A, Cordey S, Laubscher F, Baggio S, et al. Explosive nosocomial outbreak of SARS-CoV-2 in a rehabilitation clinic: the limits of genomics for outbreak reconstruction. *J Hosp Infect.* 2021 Aug 27;117:124–34.
45. Seidelman JL, Lewis SS, Advani SD, Akinboyo IC, Epling C, Case M, et al. Universal masking is an effective strategy to flatten the severe acute respiratory coronavirus virus 2 (SARS-CoV-2) healthcare worker epidemiologic curve. *Infect Control Hosp Epidemiol.* 2020 Dec;41(12):1466–7.
46. Fakih MG, Sturm LK, Fakih RR. Overcoming COVID-19: Addressing the perception of risk and transitioning protective behaviors to habits. *Infect Control Hosp Epidemiol.* 2021 Apr;42(4):489–90.
47. Advani SD, Yarrington ME, Smith BA, Anderson DJ, Sexton DJ. Are we forgetting the “universal” in universal masking? Current challenges and future solutions. *Infect Control Hosp Epidemiol.* 2021 Jun;42(6):784–5.
48. Costa SF, Giavina-Bianchi P, Buss L, Mesquita Peres CH, Rafael MM, Dos Santos LGN, et al. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Seroprevalence and Risk Factors Among Oligo/Asymptomatic Healthcare Workers: Estimating the Impact of Community Transmission. *Clin Infect Dis.* 2021 Sep 7;73(5):e1214–8.
49. Oliveira MS de, Lobo RD, Detta FP, Vieira-Junior JM, Castro TL de S, Zambelli DB, et al. SARS-CoV-2 seroprevalence and risk factors among health care workers: Estimating the risk of COVID-19 dedicated units. *Am J Infect Control.* 2021 Sep;49(9):1197–9.
50. Lindsey BB, Villabona-Arenas ChJ, Campbell F, Keeley AJ, Parker MD, Shah DR, et al. Characterising within-hospital SARS-CoV-2 transmission events: a retrospective analysis integrating epidemiological and viral genomic data from a UK tertiary care setting across two pandemic waves. *medRxiv.* 2021 Jul 19;
51. Richterman A, Meyerowitz EA, Cevik M. Hospital-Acquired SARS-CoV-2 Infection: Lessons for Public Health. *JAMA.* 2020 Dec 1;324(21):2155–6.

52. Illingworth CJ, Hamilton WL, Warne B, Routledge M, Popay A, Jackson C, et al. Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections. *eLife*. 2021 Aug 24;10.
53. Mo Y, Eyre DW, Lumley SF, Walker TM, Shaw RH, O'Donnell D, et al. Transmission dynamics of SARS-CoV-2 in the hospital setting. *medRxiv*. 2021 May 1;
54. Schneider S, Piening B, Nouri-Pasovsky PA, Krüger AC, Gastmeier P, Aghdassi SJS. SARS-CoV-2 cases in healthcare workers may not regularly originate from patient care: lessons from a university hospital on the underestimated risk of healthcare worker to healthcare worker transmission. *Antimicrob Resist Infect Control*. 2020 Dec 7;9(1):192.
55. Barry M, Robert AA, Temsah M-H, Abdul Bari S, Akhtar MY, Al Nahdi F, et al. COVID-19 Community Transmission among Healthcare Workers at a Tertiary Care Cardiac Center. *Med Sci (Basel)*. 2021 Jun 30;9(3).
56. Yousuf B, Sujatha KS, Alfoudri H, Mansurov V. Transport of critically ill COVID-19 patients. *Intensive Care Med*. 2020 Aug;46(8):1663–4.
57. Liew MF, Siow WT, Yau YW, See KC. Safe patient transport for COVID-19. *Crit Care*. 2020 Mar 18;24(1):94.

# Supplementary Methods

## Epidemiological Context

The HCFMUSP is composed of nine medical specialty institutes: Central Institute (ICH), Heart Institute (InCor), Cancer Institute (ICESP), Children's Institute (ICr), Radiology Institute (InRad), Psychiatry Institute (IPq), Physical Medicine and Rehabilitation Institute (IMREA), Orthopaedics and Traumatology Institute (IOT), and Long-term auxiliary Hospital (HAS).

Institute B became an institute dedicated to the care of COVID patients. Patients began to be transferred to other institutes and structures were created to receive 300 ICU beds and 300 ward beds, with approximately 6000 HCW. A crisis committee was created and external assistance teams of doctors, nurses and physiotherapists came to help in patient care. All were trained in the correct use of PPE, which included private clothing, N95 masks, gloves and aprons, hats and face shields, and clothing and de-dressing techniques.

Personal protective equipment (PPE) was made available to all HCW. HCW who provided direct patient care to COVID-19 patients wore N95 respirators and scrubs during their entire shifts. When examining or touching patients they added disposable gloves and a gown. During aerosol-generating procedures, they used N95 respirators, a gown, gloves, and eye protection (face shield or goggles). HCW used the same N95 respirator between patients and these were reused by the same HCW for seven shifts or until damaged or soiled. The cleaning staff wore N95 respirators during their entire shift. HCW were trained to don and doff PPE in face-to-face sessions and with videos and posters. All symptomatic HCW were evaluated at a dedicated health service (located in a separate building) and, if indicated, oro-nasopharyngeal swabs were collected. If COVID-19 was confirmed, the HCW received paid leave for 14 days from the onset of symptoms. Clinical and epidemiological data collection (age, sex, home address, occupation, unit of work within the

hospital, date of onset of symptoms, symptoms, need for hospitalization, and clinical outcome) involved the collaboration of several teams that matched sample identification numbers to medical records from patients and health workers on each institute's electronic system.

Supplementary Panel 1 – COVID-19 response measures taken by each institute from the HCFMUSP according to epidemiological week.

Epil week	Institute A	Institute B	Institute C
<b>Week 10 (07/03/2020 - 13/03/2020)</b>	COVID and non-COVID areas; HCW were not allowed to move between areas; For aerosol forming procedures: N95 + glasses + face shield; Triage of patients, HCW and carers (temperature and symptoms)		
<b>Week 11 (14/03/2020 - 20/03/2020)</b>	Restriction of in person events and meetings; No visitors allowed in COVID areas and limited in all other areas;	Entrance areas, elevators, pantries, cafeterias were separated to ensure the safety of health professionals.	
<b>Week 12 21/03/2020 - 27/03/2020)</b>		Outpatient care was suspended	COVID and non-COVID areas
<b>Week 13 (28/03/2020 - 06/04/2020)</b>		Became COVID-only. Guideline use of PPE available for all HCW	
<b>Week 15 (04/04/2020 - 10/04/2020)</b>	Mandatory masking. Surgical masks for all HCW; administrative workers could wear cloth masks.	Focus on identifying symptomatic healthcare workers and immediate leave	HCW servicing patients - surgical masks in inpatient wards and N95 in ICUs; 1 visitor per patient per day
<b>Week 16 (11/04/2020 - 17/04/2020)</b>		HCW not allowed to transit between institutes. Professionals were dedicated exclusively to Institute B	

<b>Week 17</b> <b>(18/04/2020 - 24/04/2020)</b>		Training on the use of PPE for multidisciplinary teams such as nutritionists, hygiene staff, maintenance engineering, psychologists, social workers	Universal masking mandatory (surgical mask)
<b>Week 18</b> <b>(25/04/2020 - 01/05/2020)</b>	N95 for all critical emergency areas, ICUs and surgical centre	Training on the use of PPE for administrative personnel due to an outbreak	5 Covid-19 transmission measures; Daily audits until 10/05
<b>Week 19</b> <b>(02/05/2020 - 08/05/2020)</b>		Mandatory use of surgical masks for all professionals, including administrative personnel.	

### SARS-CoV-2 genome amplification

For SARS-CoV-2 whole-genome sequencing, we used a tilling-amplicon multiplex PCR technique as previously described (3–5). First, the cDNA was synthesized from positive RNA samples using the ProtoScript II First-Strand cDNA synthesis kit (New England Biolabs, UK) and random primers or SuperScriptIV First-Strand Synthesis System (Thermo Fisher Scientific, USA). Subsequently, cDNA was subjected to amplification using the V2 ARTIC scheme (<https://artic.network/ncov-2019>) and Q5 High-Fidelity DNA polymerase (New England Biolabs, UK). After amplification, the AmpureXP purification beads (Beckman Coulter, United Kingdom) were used for product purification and the Qubit dsDNA High Sensitivity Assay on the Qubit 3·0 instrument (Life Technologies, USA) to quantify the amplicons.

### Library preparation and whole-genome sequencing

Sequencing libraries were prepared using a total input of 100ng. The normalized amplicons were submitted to barcode ligation using the EXP-NBD 104 (1–12) and EXP NBD 114 (13–24) Native Barcoding Kits (Oxford Nanopore Technologies, UK). Sequencing libraries were generated using the SQK-LSK109 Kit (Oxford Nanopore Technologies, UK). Finally, 20ng of the library, containing

23 samples and one negative control, were loaded onto an R9·4·1 flow-cell on the MinION device and sequenced using MinKNOW 1·15·1 (Oxford Nanopore Technologies, UK).

### **Bioinformatic analysis**

Guppy software v2·2·7 (Oxford Nanopore Technologies, UK) was used to basecall, demultiplex, and trim the FAST5 files. FASTQ files were mapped to the reference genome of SARS-CoV-2 isolate Wuhan-Hu 1 (GenBank Accession Number MN908947) using minimap2 v·2·28·0 to generate the consensus genomes and SAMtools to convert to a sorted BAM file (Li et al., 2009). Length filtering and the quality test was performed for each barcode using artic guppyplex (<https://artic-network/ncov-2019/ncov2019-bioinformatics-sop.html>). The genome statistics were obtained from SAMtools and Tablet viewer (Milne et al., 2010). To recover consensus sequences, called variants were detected with Nanopolish. Genome regions with a depth of <20-fold were not included in final consensus sequences, and these positions are represented with N characters. Runs with negative control presenting any contamination were discarded.

### **Collation of genomic datasets and sequence quality control**

SARS-CoV-2 sequences from Brazil with collection date up to the 20th May 2020 (oldest collection date in our HCW dataset) (n=1860) were downloaded from GISAID ([6–8](#)) and appended to a previously described global dataset of 1,182 viral genomes ([4](#)). As previously described ([4](#)), we further filtered down our dataset by maintaining only sequences with at least 75% consensus sequence coverage. The resulting dataset was aligned to the reference NC\_045512·2 using MAFFT v 7·450 ([9](#)) and manually edited using AliView. 3' and 5' untranslated regions of each sequence were discarded.

A Maximum likelihood tree was inferred using IQ-TREE v·2·0 ([10](#)) under the best substitution model as determined by ModelFinder ([34](#)) implemented in the IQ-TREE pipeline. TempEst v·1·5·3 ([35](#)) analyses and visual inspection of the alignment in Aliview were used to identify and remove

sequences with unusual divergence for a given date of collection and/or long stretches of polymorphisms. Our final dataset consisted of 2,550 sequences, including 340 sequences from HCFMUSP, 67 novel genomes and 273 previous GISAID submissions from our group (dataset 1). Pangolin version V3·1·11 (36) was used for lineage assignment. For accurate cluster identification, sequences with >90% coverage (n=2259) were maintained for subsequent phylogenetic analysis and initial cluster identification, including 234 sequences from this study. Finally, for Bayesian phylogenetic analysis, dataset B was subsampled to include 200 randomly selected sequences from other countries, all sequences from Sao Paulo state (n=407), and all sequences from this study with coverage >90% (n=234) (dataset 3, n=841 sequences). Only sequences from Sao Paulo state were maintained given that all sequences from this study clustered amongst them.

### **Phylogenetic and Phylogeographic analysis**

Maximum likelihood trees for all datasets were inferred using IQ-TREE v·2·0, with the best substitution model determined by ModelFinder implemented in the IQ-TREE pipeline: GTR+F+R2 (Datasets A and B), GTR+F+R3 (dataset 3). Sequences from this study were scanned for recombination using all methods available on RDP4 (37) and no recombination signal was found. Bayesian time-rooted phylogenies were estimated using Beast v1·10·4 (38) and BEAGLE (39), under an HKY+ $\Gamma$  nucleotide substitution model, a strict molecular clock and an exponential growth coalescent tree prior (40); (27). For population size, a log-normal prior with mu=1 and sigma=0 was used and for growth rate, a la place prior with mean=0 and scale=100 was set. Analyses were run in triplicates for 200 million Markov Chain Monte Carlo (MCMC) steps. Run convergence was assessed using Tracer v·1·7·1 (41). Log and tree files were combined after removal of 10% burn-in from each run using LogCombiner v1·10·4 and summary Maximum Clade Credibility trees were generated from the combined tree files using TreeAnnotator v1·10·4. LogCombiner v1·10·4

was also used to generate a resampled distribution of 1,000 from the combined tree files, which was used for subsequent phylogeographic analysis.

To understand the dynamics of SARS-CoV-2 spread across the institutes we used 4 different discrete trait schemes. For the first trait, location ( $k=5$ ), sequences were assigned one of five locations ( $k=5$ ): (I) Institute A, (II) Institute B, (III) Institute C, (IV) São Paulo (other sequences from São Paulo state) or (V) Other (international sequences). For the discretization scheme two ( $k=3$ ), location2, all institute sequences were assigned one single location (I) HC; while other sequences were assigned as (II) São Paulo and (III) Other To incorporate the transmission dynamics between HCW and patients, discretization scheme 3 ( $k=4$ ) assigned sequences to either (I) HCW, (II) patient, (III) São Paulo, and (IV) Other ( $k=4$ ). Finally, discretization scheme 4 ( $k=8$ ) accounted for both the institute and HCW/Patient trait information by assigning sequences to either (I) Institute A\_HCW (II) Institute A\_patient, (III) Institute B\_HCW, (IV) Institute B\_patient, (V) Institute C\_HCW, (VI) Institute C\_Patient, (VII) São Paulo and (VIII) Other. The number of migration events between each location considered, as well as the total number of imports and exports for each location, were estimated using a Markov Jumps count approach [\(42\)](#) implemented on Beast v1.10.4, under an asymmetric CTMC model for discrete trait reconstruction [\(43\)](#). Migration rates were inferred on a separate run under the same asymmetric model with a Bayesian Stochastic Search Variable Selection procedure (BSSVS) [\(44\)](#), which identifies and limits the number of rates to only one that can actually explain the diffusion process. The proportion (%) of imports for each institute was estimated by considering the maximum number of imports for each trait as equal to the total number of sequences for each trait included in the analysis. By considering imports as any transition from another trait, the remaining proportion, adding up to 100%, would be the proportion of transmission happening within a particular trait (non-imports), and thus representing transmission within each institute or HCW/Patient. To estimate the expected proportion of imports and transitions in a scenario in which clustering was not related to the traits, locations for institute sequences in schemes 1 and 4 were randomly

reordered ten times and ten independent simulations were run for each scheme, under the same evolutionary models described above. The average distribution for imports for each trait was estimated from the results of the ten simulations.

### **Cluster analysis**

Considering the relatively low evolutionary rate of SARS-CoV-2, which accumulates 2-3 mutations a month (27),(45) cluster analysis was performed on sequences with >90% coverage, in order to decrease the chances of incorrect assignment of sequences to transmission clusters. Initial identification of clusters was performed on dataset 2 under an ML phylogeny run on IQ-TREE v·2 (see methods). Clusters were confirmed on dataset 3 under both ML, Bayesian time-rooted trees and Bayesian trait-referenced time-rooted trees. Clusters were defined according to the content of HC sequences (sequences from this study) and according to the statistical support obtained from phylogenetic analysis. HC clusters were considered when >75% of the sequences were from HC, when they were supported by a minimum SH-like approximate likelihood ratio test (SH-aLRT) support of 75, minimum fast bootstrap support of 90, a minimum node posterior support of 0·9, and had at least one defining mutation separating clustered sequences from the nearest sequences in the phylogenetic tree. Statistical support thresholds were defined considering the support thresholds recommended by IQ-TREE and the agreement between the different support statistics across different phylogenetic methods.

The strength of the epidemiological link between clustered sequences was assessed using both hospital-associated metadata and the geocoded residential addresses for patients and HCW. An “epidemiological link” was assigned when individuals worked/were hospitalized in the same ward/floor at the time of symptom onset or when it involved HCWs from the same specialty/division. A “possible epidemiological link” was defined as being hospitalized/working in the same institute at the time of symptom onset, but not necessarily in the same ward/floor or being from the same division/specialty. Finally, an “unclear epidemiological link” was defined as

individuals who worked/were hospitalized at different institutes at the time of symptom onset and no clear epidemiological link could be established.

### **Compartmentalization analysis**

To investigate the association of specific traits and genetic diversity of SARS-CoV-2 in our dataset, we used Simmond's Association Index implemented in the Hypothesis Testing Using Phylogenies (HYPHY) (46). AI calculates an association index for genetic diversity according to different compartments (traits). It assesses population structure by weighting the contribution of each node while also running a bootstrap to provide support for the association index. In our analysis, we considered three separate sets of compartments: Institutes (Institute A, Institute B, Institute C), occupations (doctor, nurse, administration, other, and patient), and Patient/HCW. Doctors and nurses comprise all different levels of training of medical doctors and nursing professionals (including technicians). Analysis was performed in two datasets: (I) all HC sequences >90% coverage (n=234) and (I) clustered sequences only (n=73). Runs were performed under the following conditions: ten relabellings (default), 1000 bootstraps, and a significance threshold of <2% (default).

### **Statistical analysis**

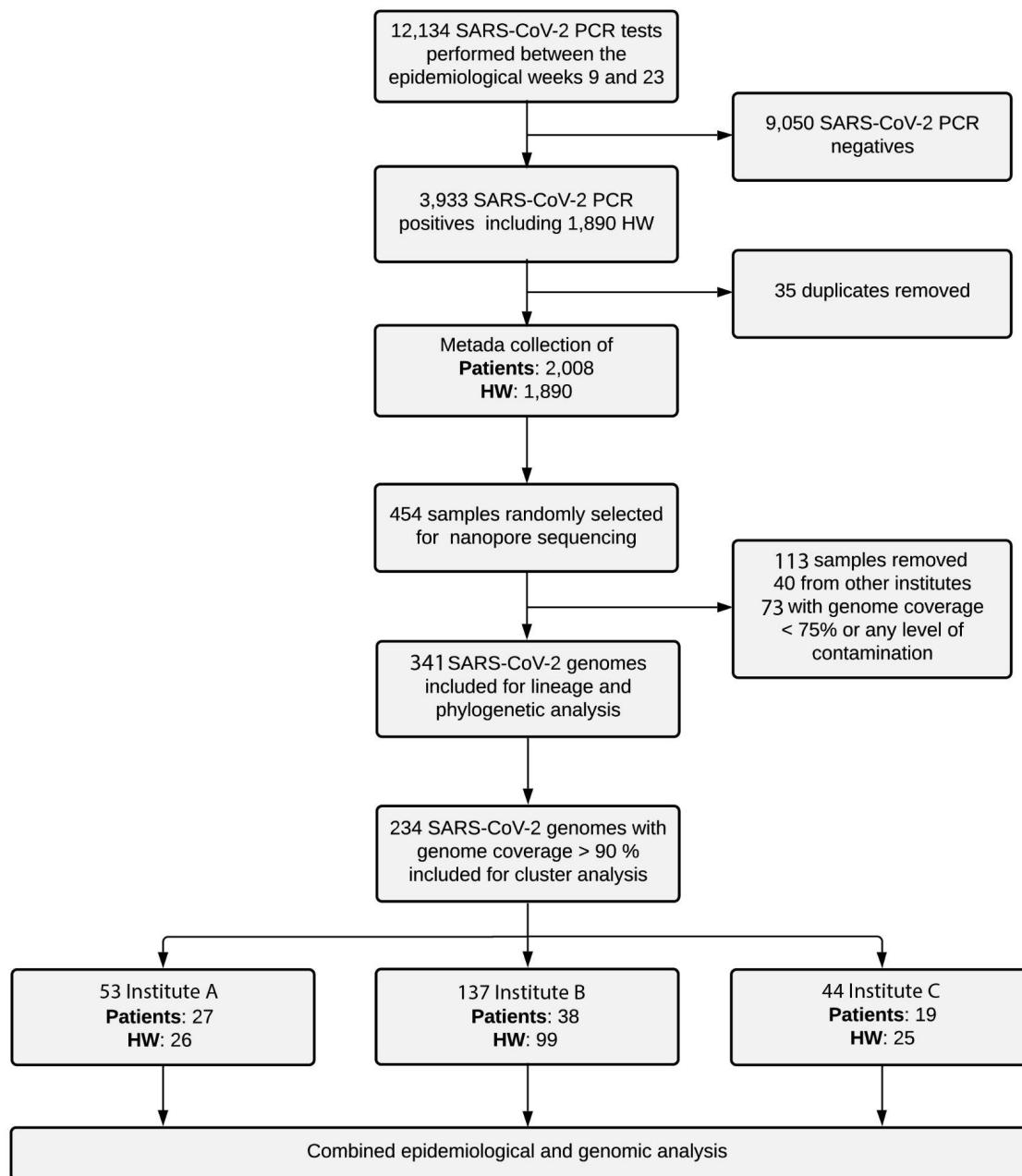
Descriptive analyses are shown as arithmetic means, median, and range. To assess traits that may affect the clustering of genetic sequences, we performed binomial logistic regression analyses, having clustered or non-clustered as the outcome variable, and analyses were controlled for sex and age. For model 1, traits location (Institute A, Institute B, and Institute C) and HCW/patient were used. Baseline variables were Institute B and patient. For model 2, one single trait incorporating both location and HCW/patient was used: HCW.Institute A, patient.Institute A, HCW.Institute B, patient.Institute B, HCW.Institute C, patient.Institute C. Baseline variable was patient.Institute B. For model 3, traits location (Institute A, Institute B, and Institute C) and

occupation (administration, doctor, medical resident, nurse, nurse technician, others, and patient. Baseline variables were Institute B and patient. Finally, model 4 contained one single trait with all possible combinations of location and occupations used in model 3 ( $k=21$ ). The baseline variable was patient. Institute B. Results were reported as the odds ratio (OR) over the baseline variables and p values  $<0.05$  were considered statistically significant. For the household geographical distances analysis, a Mann-Whitney U test was performed in R Studio 1·2·1335.

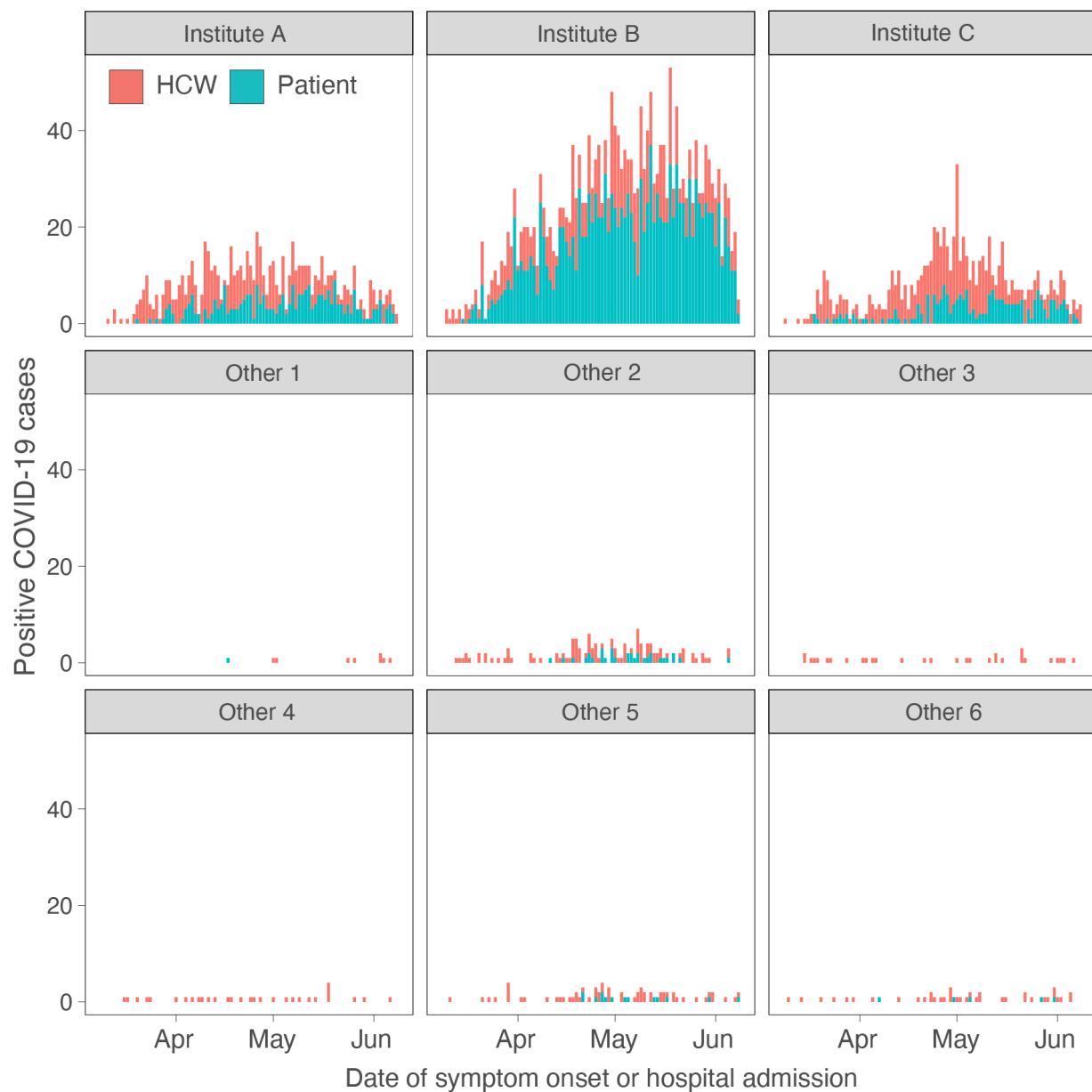
### **Data sharing**

Raw virus reads, and consensus sequences generated in this study can be found at <https://github.com/CADDE-CENTRE/...> XMLs GISAID IDs are available in Table S3.

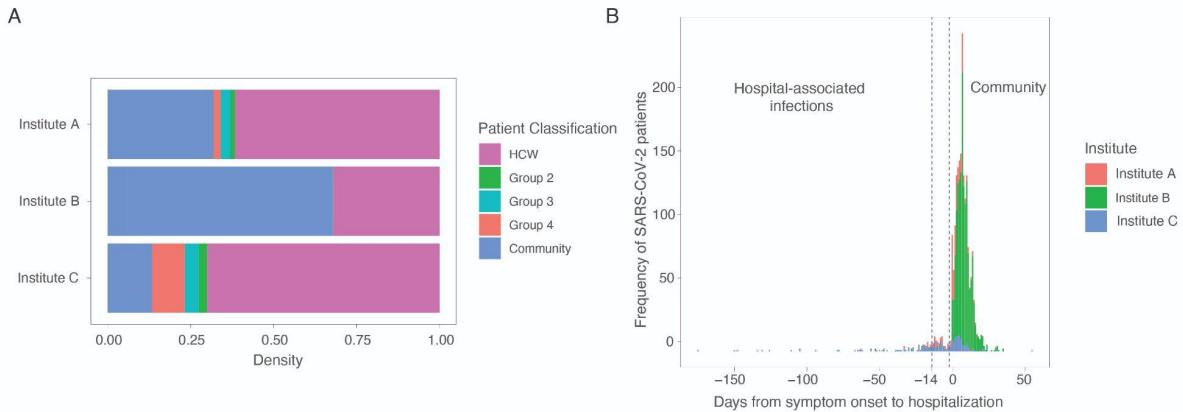
# Supplementary Figures



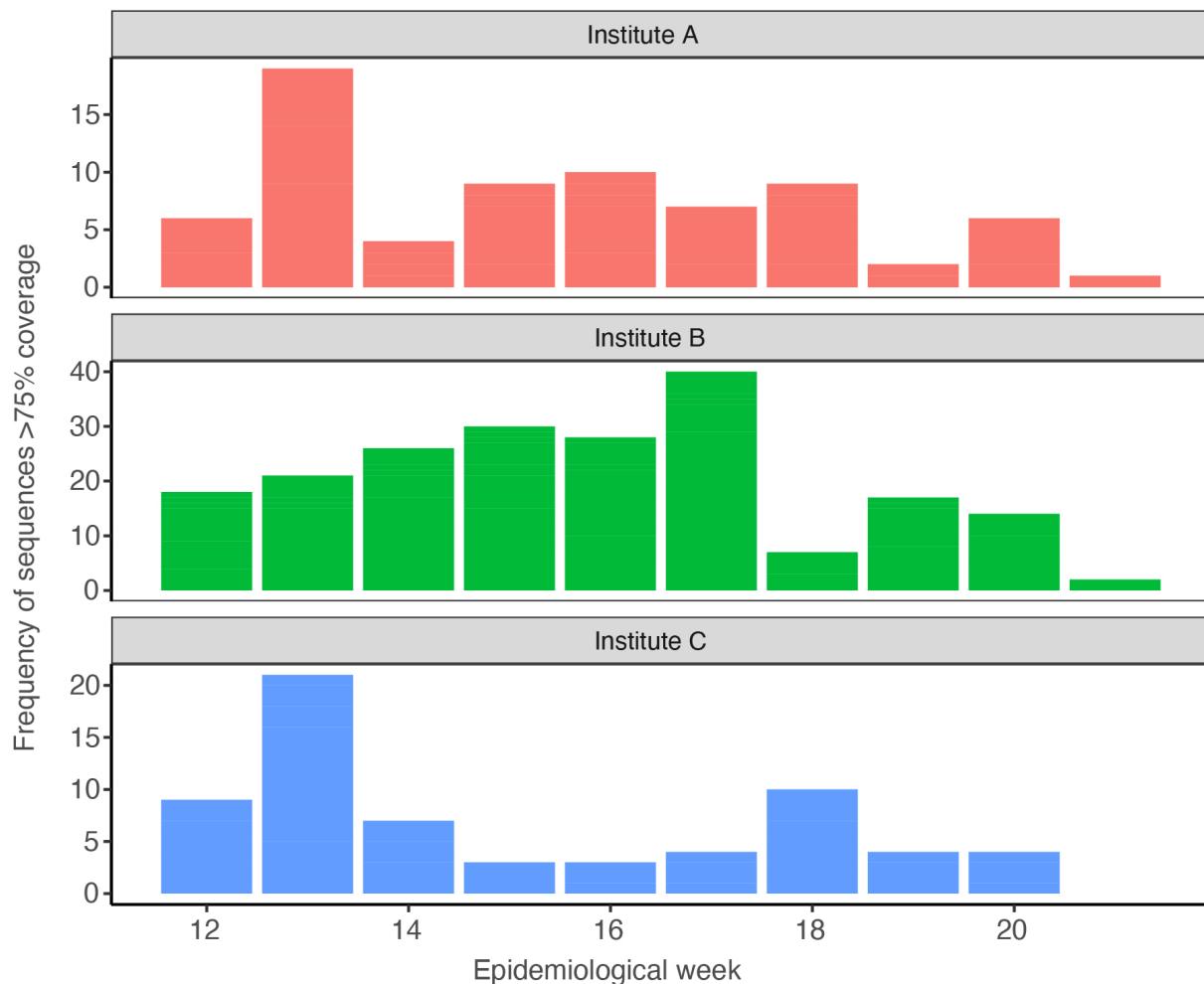
**Figure S1. Fluxogram contains information on the study design.** Out of 3,933 SARS-CoV-2 positive individuals from all HC institutes, minimum metadata was successfully collected for a total of 3,898. Of these, 454 samples from Institute B, Institute A, and Institute C were randomly selected for nanopore genome sequencing using the ARTIC protocol. 340 samples passed our quality control analysis and were submitted to lineage assignment using Pangolin COVID-19 Lineage assigner. 234 samples with coverage >90% were used for cluster analysis. For metadata collection and genome sequencing and quality control, see methods.



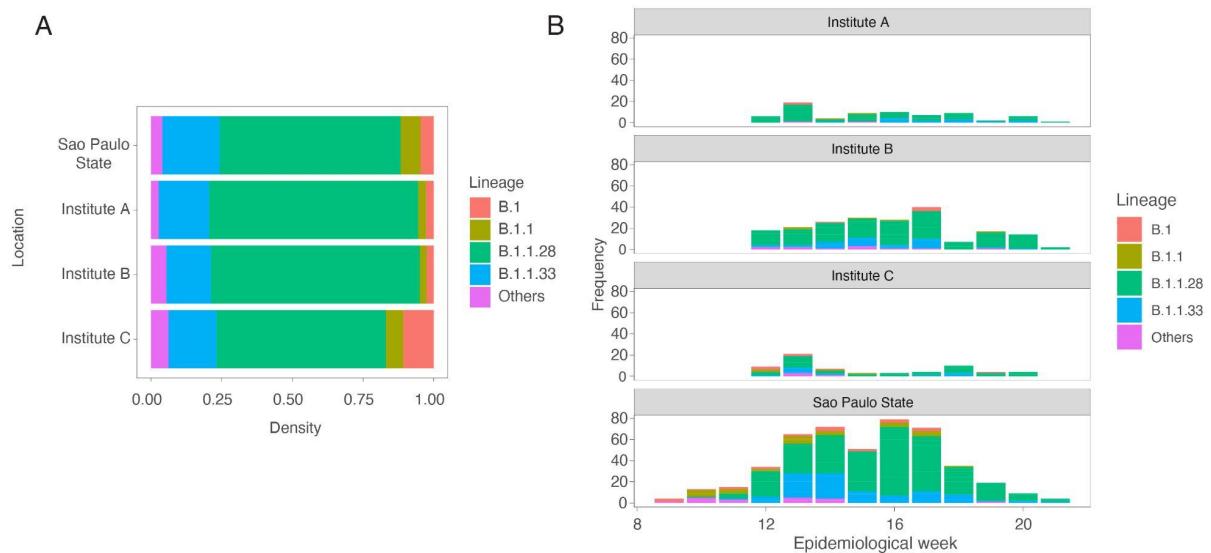
**Figure S2. Time series of 3,898 HC COVID-19-positive cases by the institute of origin.**  
 Colors distinguish patients (blue) from HCW (red). The date of symptom onset was used for health workers and patients who were hospitalized before symptom onset. For community-acquired patients, the date of hospitalization was used (see methods). Institute A, Institute B, and Institute Cconcentrate 91·8% of all HC reported cases.



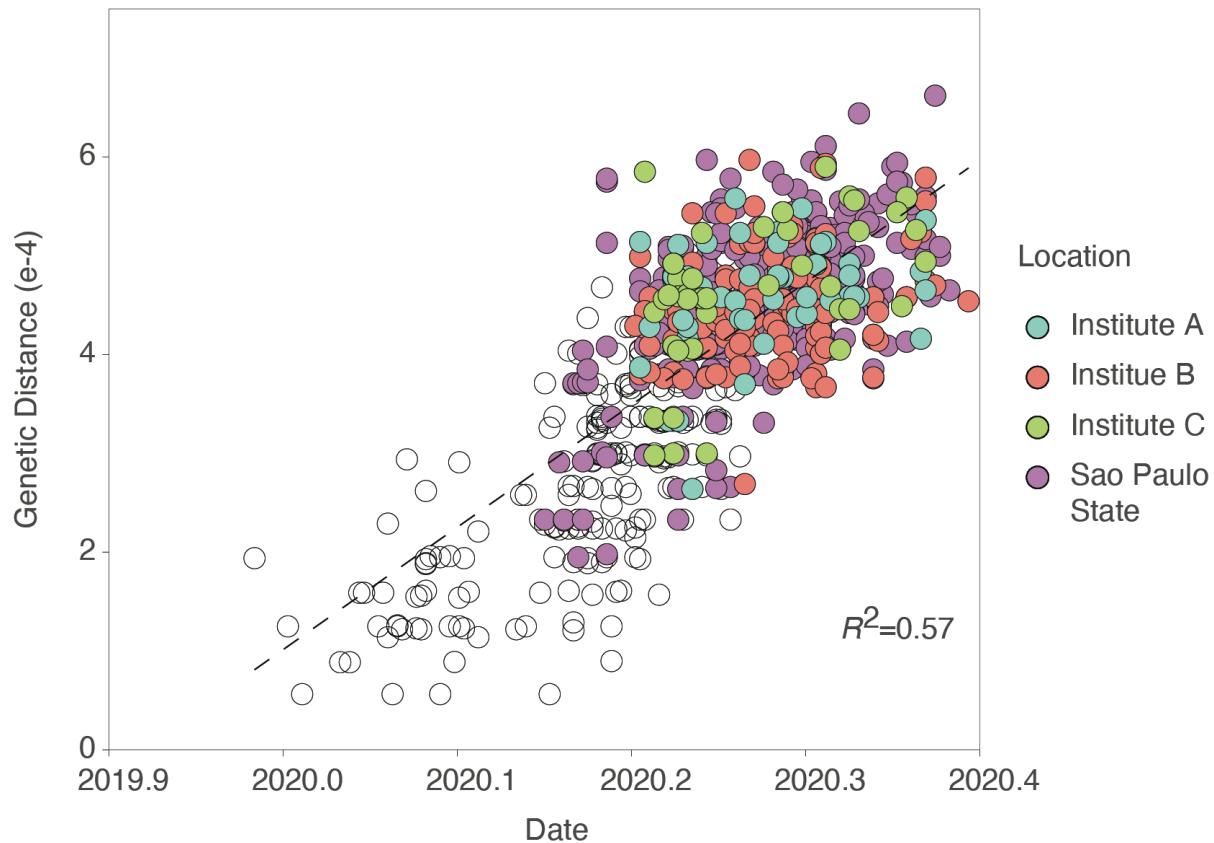
**Fig S3. Classification of Covid-19 positive patients according to the gap between symptom onset and hospitalization.** (A) Proportion of patients belonging to each category per institute. (B) Distribution of COVID-19 positive patients according to the time gap (in days) between hospitalization and symptom onset. Negative values mean that patients were hospitalized prior to symptom onset, while positive values mean that patients were hospitalized after symptom onset. Dotted lines mark patients from groups 3 (symptom onset >2 and <8 days after hospitalization) and 4 (symptom onset  $\geq 8$  and  $<15$  days after hospitalization), to which hospital-associated infection is suspected but inconclusive·



**Fig S4. Time series of COVID-19 sequences with coverage >75% per week per institute of origin.**



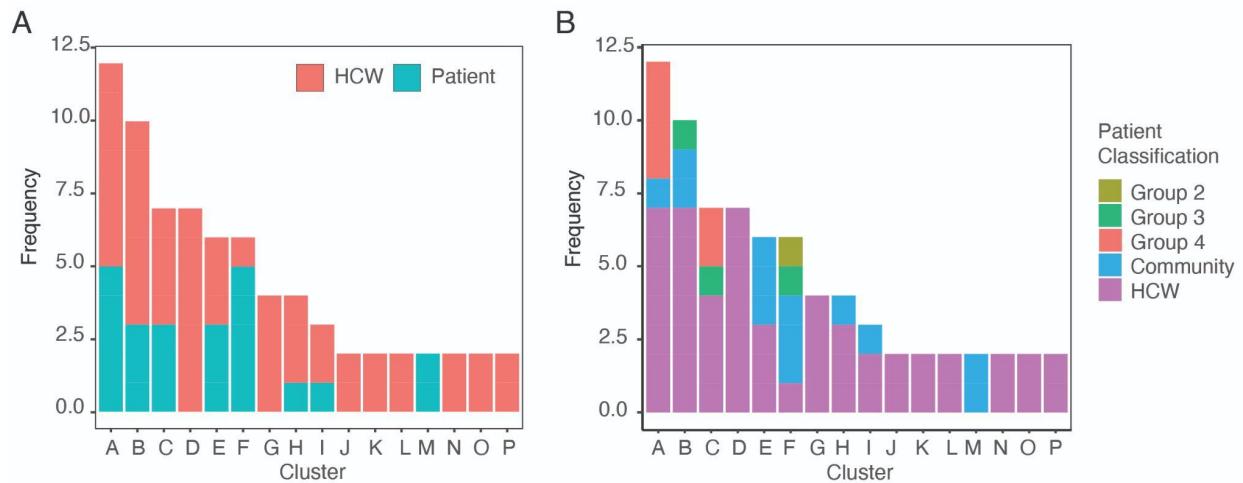
**Fig. S5. Genetic diversity of HC SARS-CoV-2 samples.** (A) Proportion of SARS-CoV-2 lineages per institute (n=340 sequences, >75% coverage) and São Paulo state (n=471 sequences). Lineage assignment was performed using Pangolin COVID-19 Lineage assigner. (B) Time series of lineages per institute/São Paulo per epidemiological week.



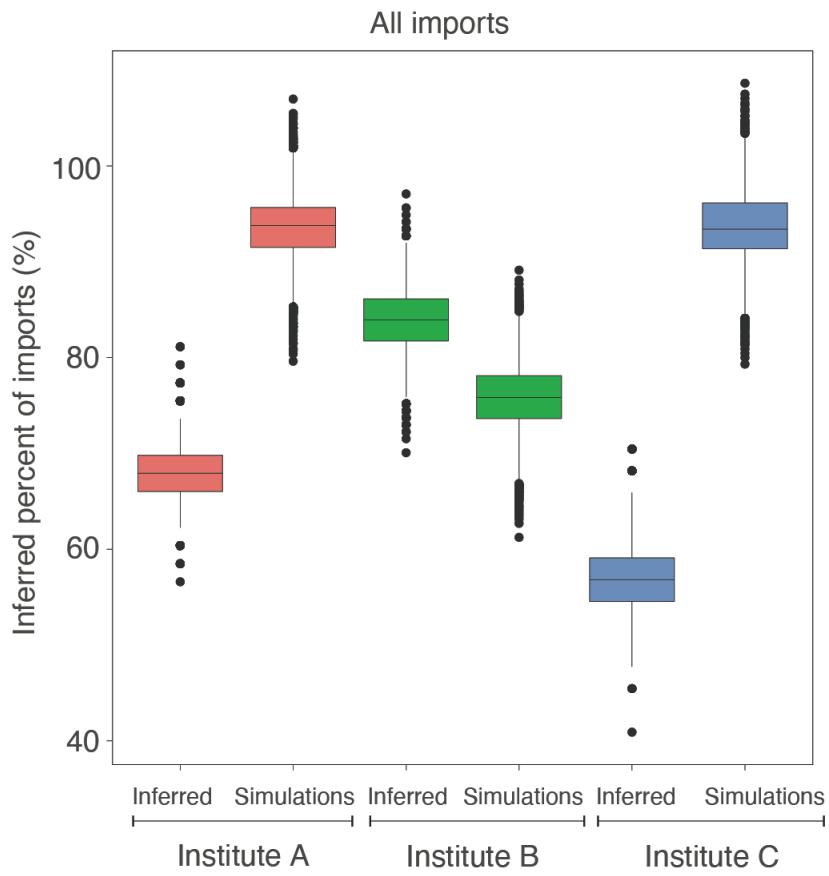
**Fig S6. Regression of root-to-tip genetic diversity and sample collection dates for Dataset C.** Dataset C is composed of 841 sequences, 234 sequences from HC, 407 sequences from the State of São Paulo in Brazil, and 200 international sequences (see methods). Circles are colored according to the location of collection: HC institutes (Institute A, Institute C, and Institute B), São Paulo State (SP, purple), international (white).

**Fig. S7 (separate PDF file)**

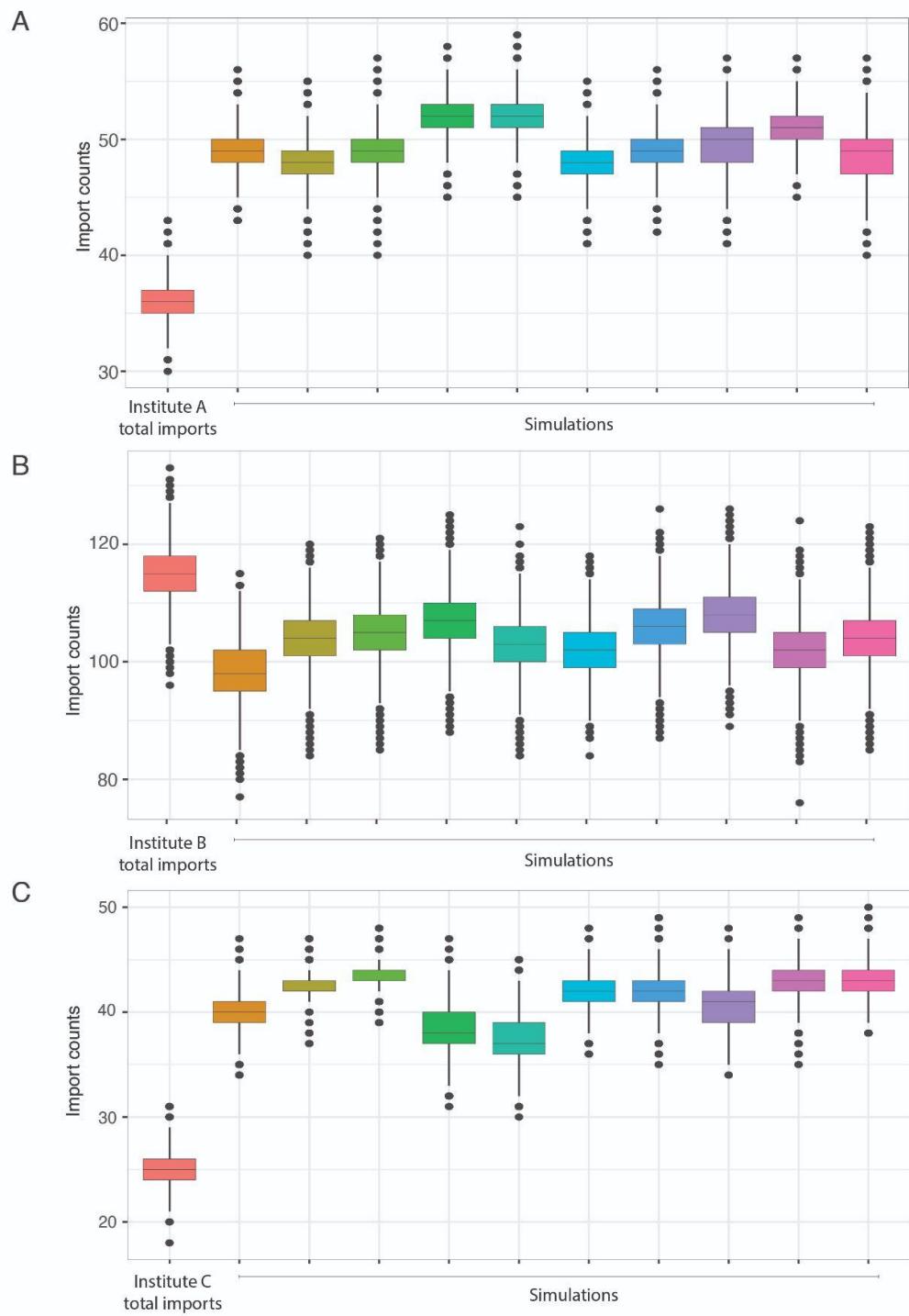
Detailed annotated maximum clade phylogenetic tree. Time-resolved maximum clade credibility phylogeny of 1,182 SARS-CoV-2 sequences, 490 from Brazil (red) and 692 from outside Brazil (blue). The largest Brazilian clusters are highlighted in grey (Clade 1, Clade 2 and Clade 3). States are defined by their 2-letter ISO 3166-1 codes. The MCC tree can be found at our Dryad repository (see Data Availability).



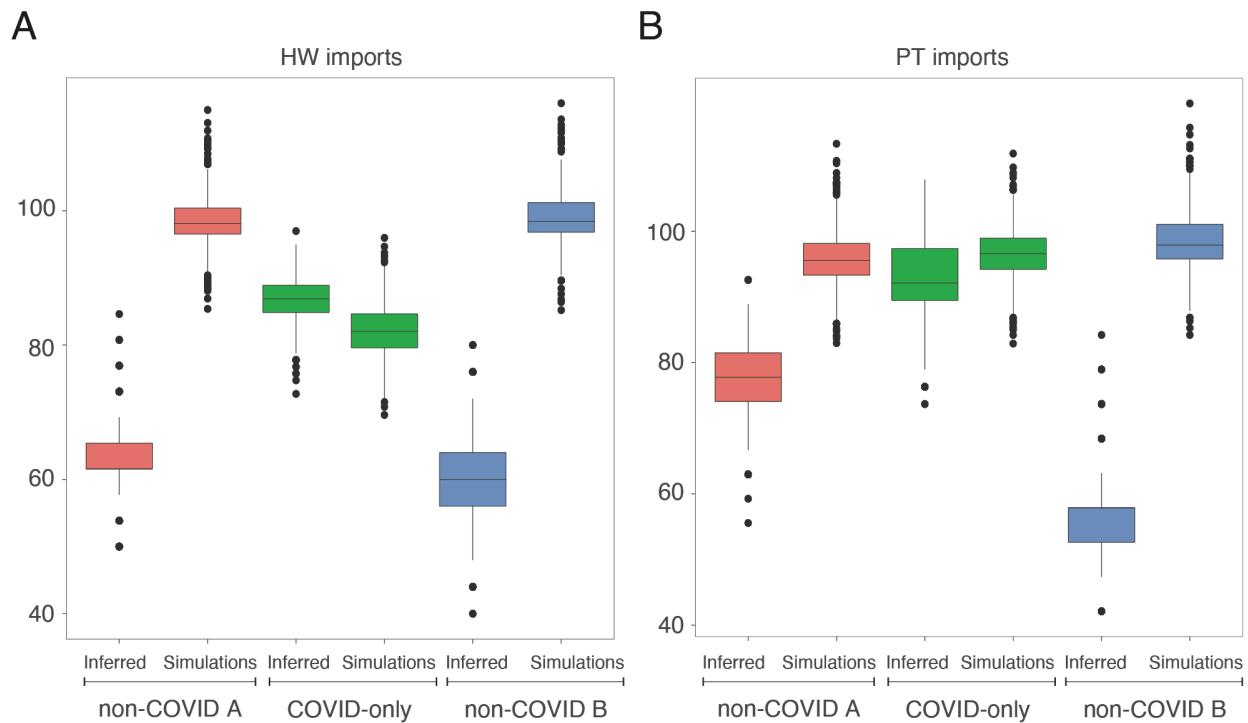
**Fig S8. Characteristics of 16 potential hospital-associated HC transmission clusters.** (A) Frequency of HCW and patients per cluster. (B) Patient classification according to hospitalization per cluster.



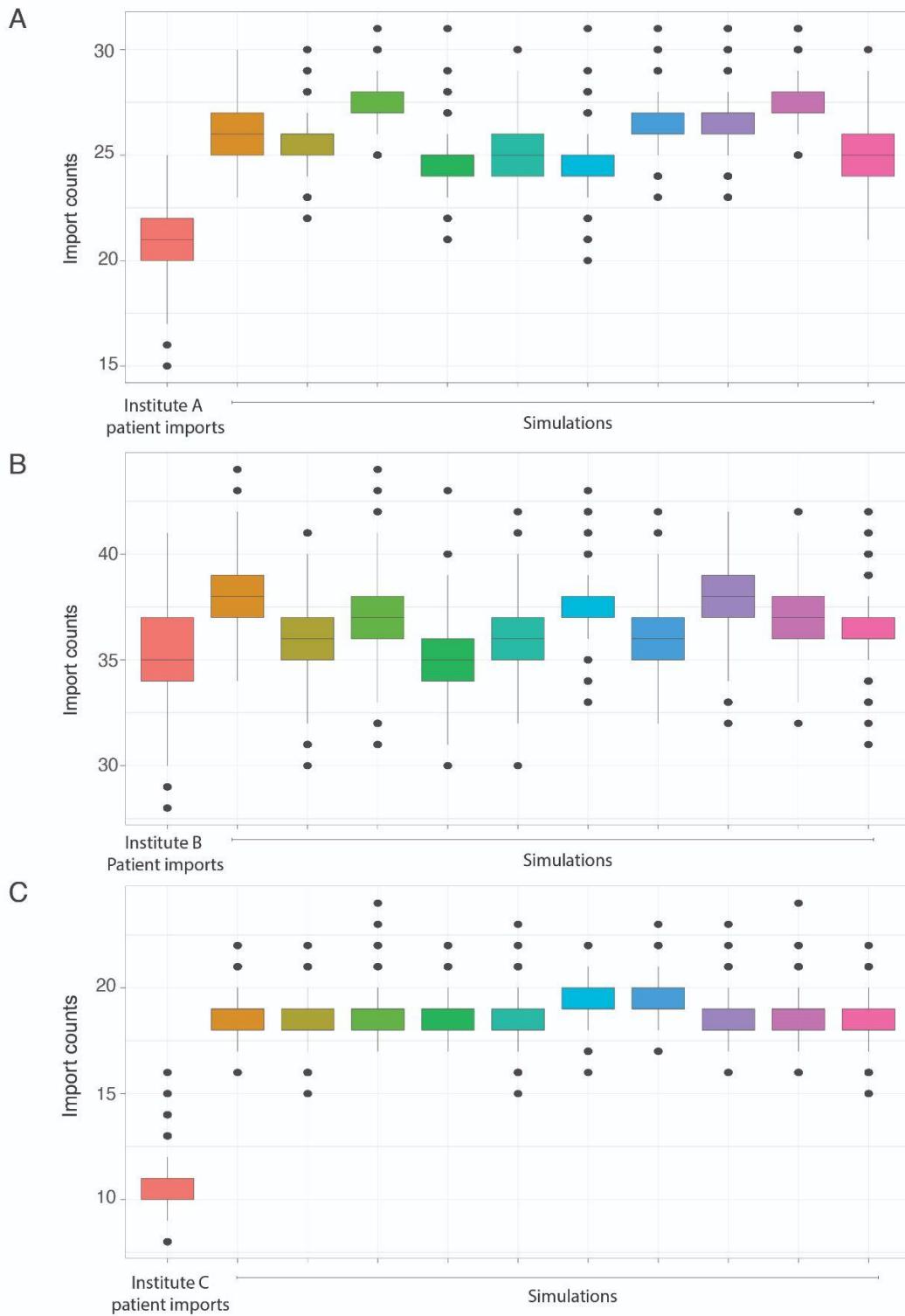
**Fig S9. Inferred proportion of cases of patients and HCWs from HC institutes caused by imports from outside each institute.** (A) Proportion (%) of inferred HCW imports to Institute A, Institute B, and Institute C. (B) Proportion (%) of inferred patient imports to Institute A, Institute B, and Institute C. Colours highlight boxplots from each institute Institute A (red), Institute B(green), Institute C(blue). Imports were inferred from a discrete trait analysis using Markov Jumps counts implemented on Beast v·1·10·4 (see methods for details).



**Fig S10. Distribution of total SARS-CoV-2 import cases per HC institute.** Imports were inferred from dataset C (841 sequences) and using a Markov jumps approach implemented in Beast 1·10·4. An import is considered as any jumps coming from outside the institute. Simulations were performed by reshuffling the institute trait assigned to each HC sequence (see methods). (A) Total Institute A imports. (B) Total Institute B imports. (C) Total Institute C imports.

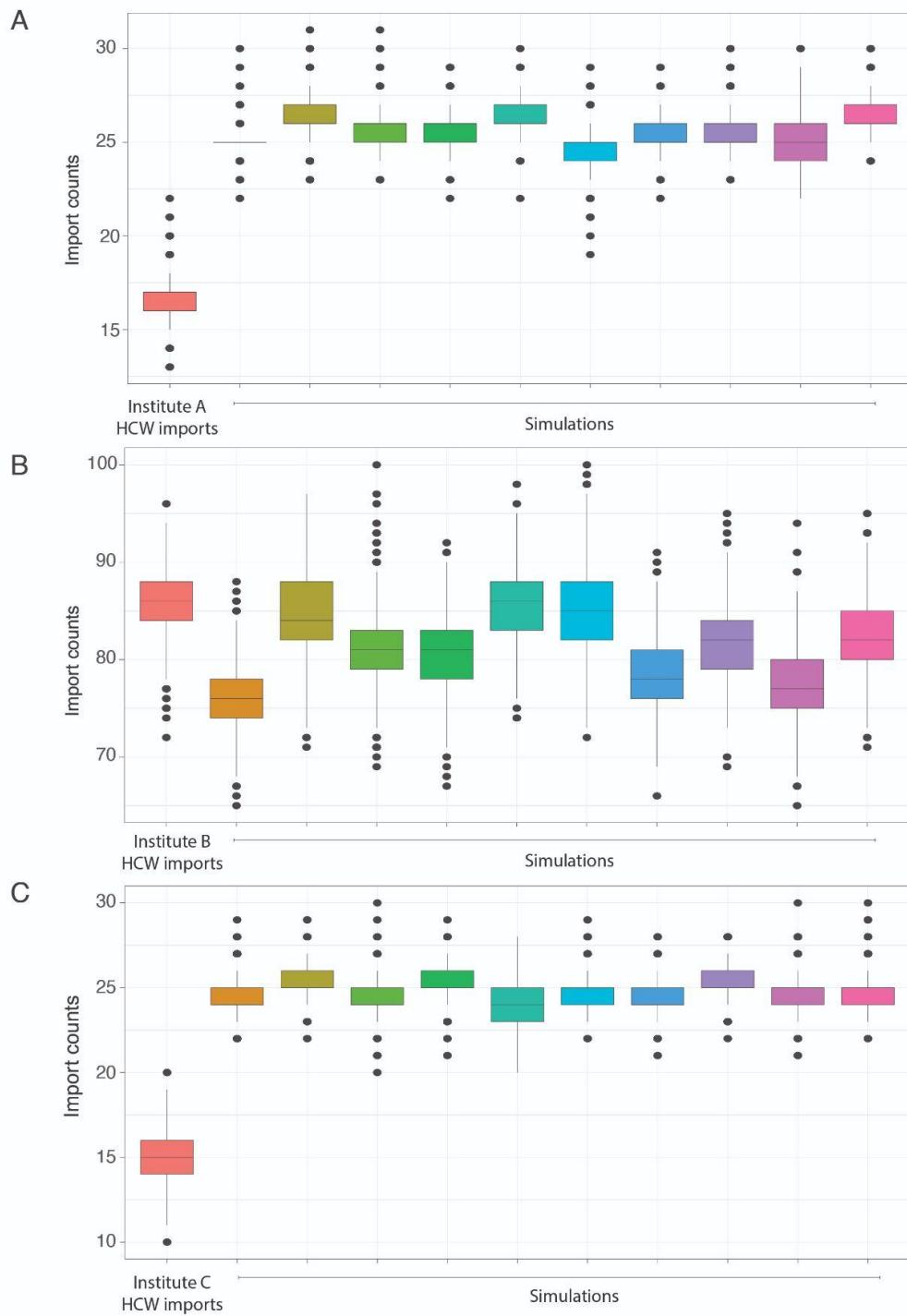


**Fig S11. Inferred proportion of cases of patients and HCWs from HC institutes caused by imports from outside each institute.** (A) Proportion (%) of inferred HCW imports to Institute A, Institute B, and Institute C. (B) Proportion (%) of inferred patient imports to Institute A, Institute B, and Institute C. Colours highlight boxplots from each institute Institute A (red), Institute B(green), Institute C(blue). Imports were inferred from a discrete trait analysis using Markov Jumps counts implemented on Beast v·1·10·4 (see methods for details).

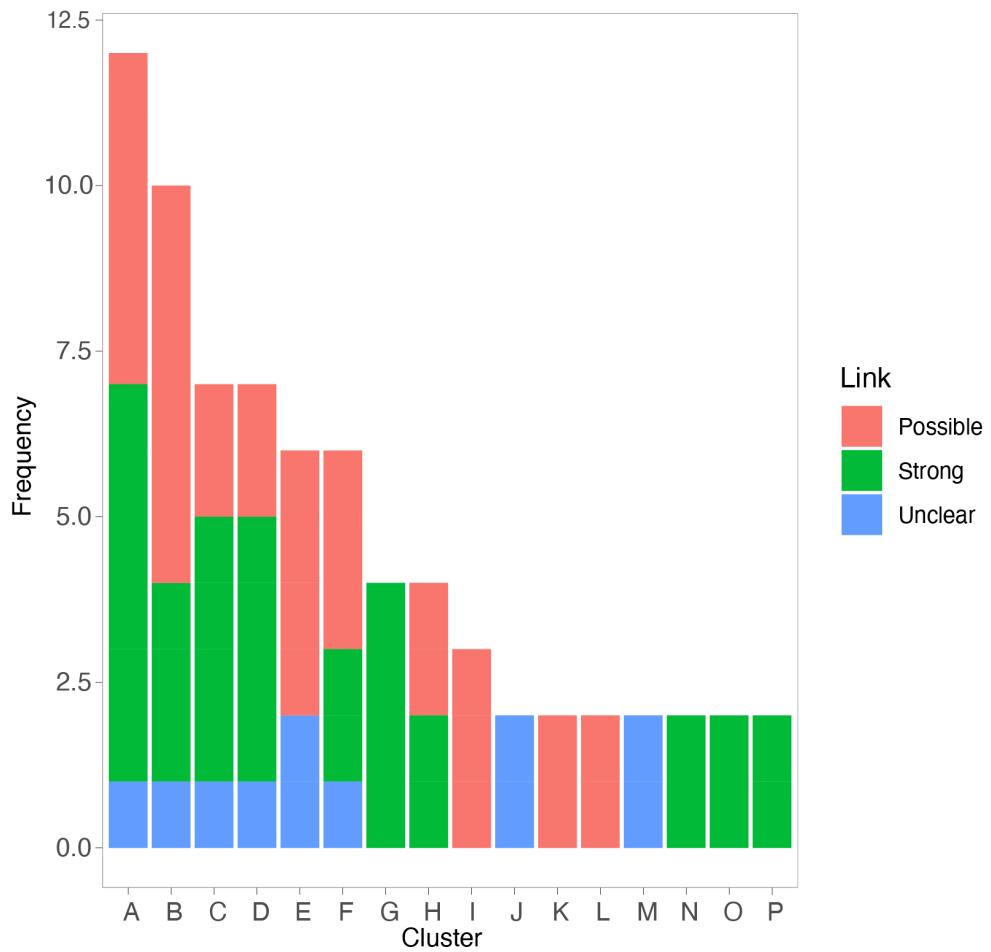


**Fig S12. Distribution of patient SARS-CoV-2 import cases per HC institute.** Imports were inferred from dataset C (841 sequences) and using a Markov jumps approach implemented in Beast 1·10·4. An import is considered as any jumps coming from outside the institute. Simulations were performed by reshuffling the institute trait assigned to each HC sequence (see

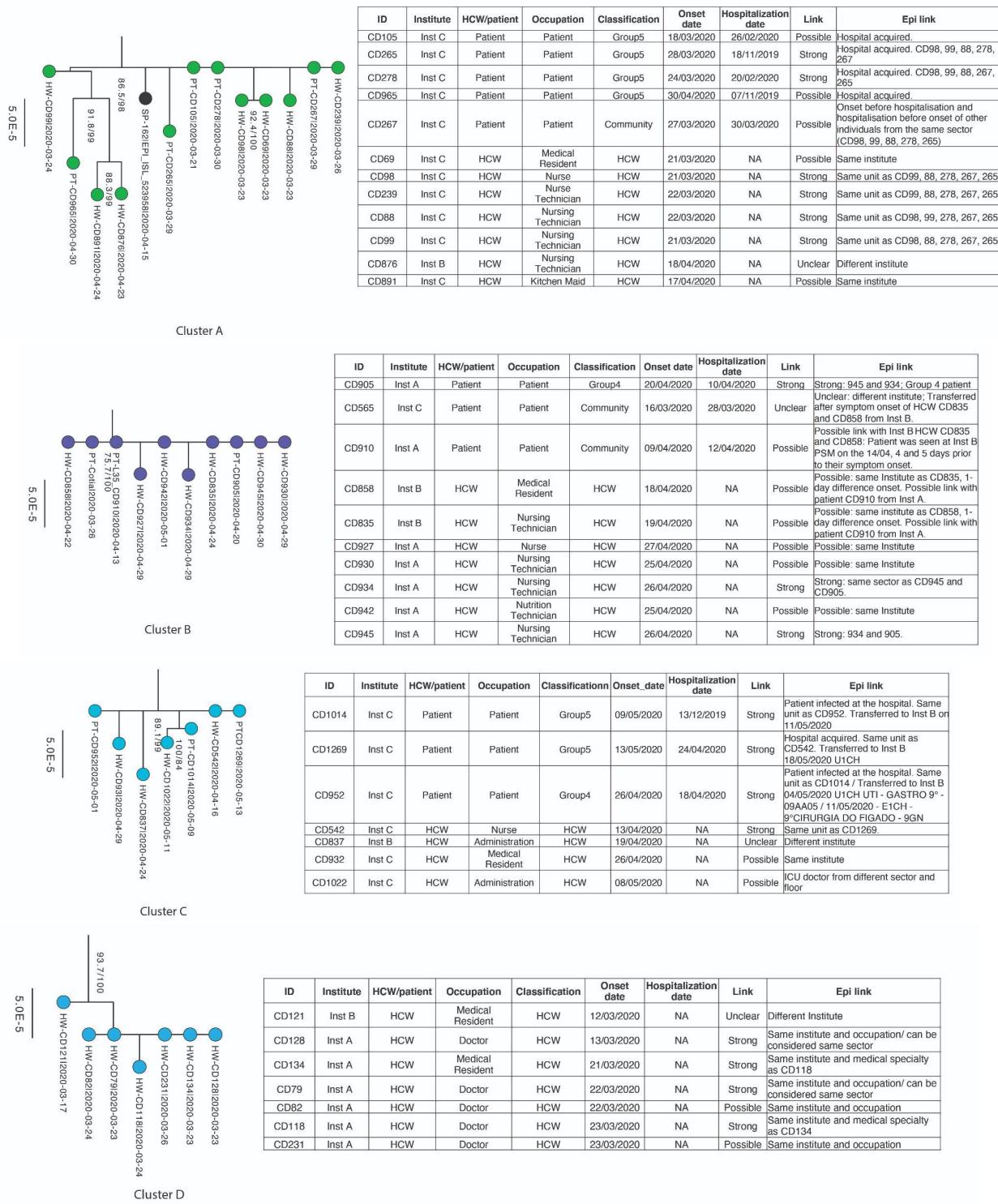
methods). (A) Institute A patient imports. (B) Institute Bpatient imports. (C) Institute Cpatient imports.



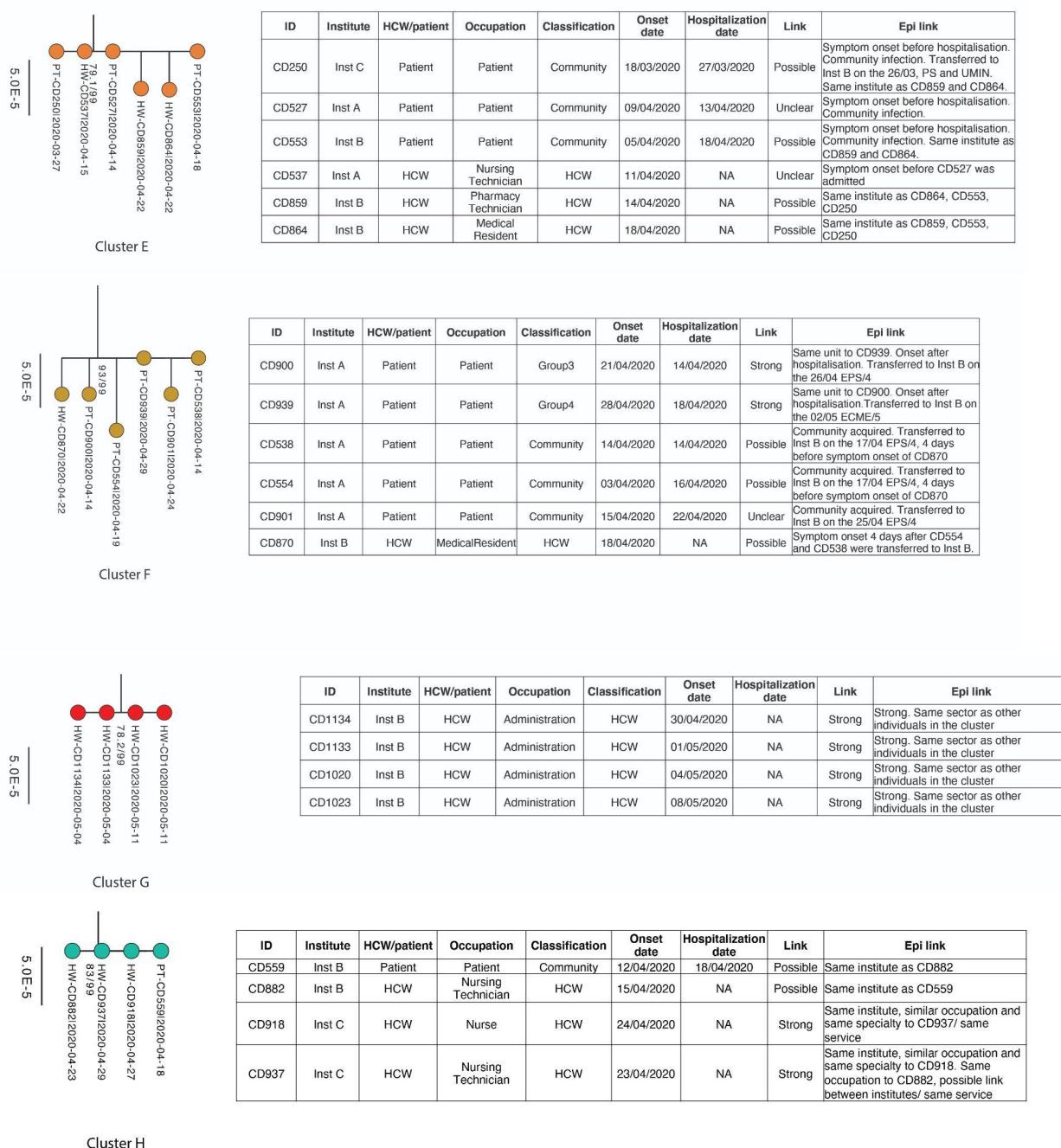
**Fig S13. Distribution of HCW SARS-CoV-2 import cases per HC institute.** Imports were inferred from dataset C (841 sequences) and using a Markov jumps approach implemented in Beast 1·10·4. An import is considered as any jumps coming from outside the institute. Simulations were performed by reshuffling the institute trait assigned to each HC sequence (see methods). (A) Institute A HCW imports. (B) Institute BHCW imports. (C) Institute CHCW imports·



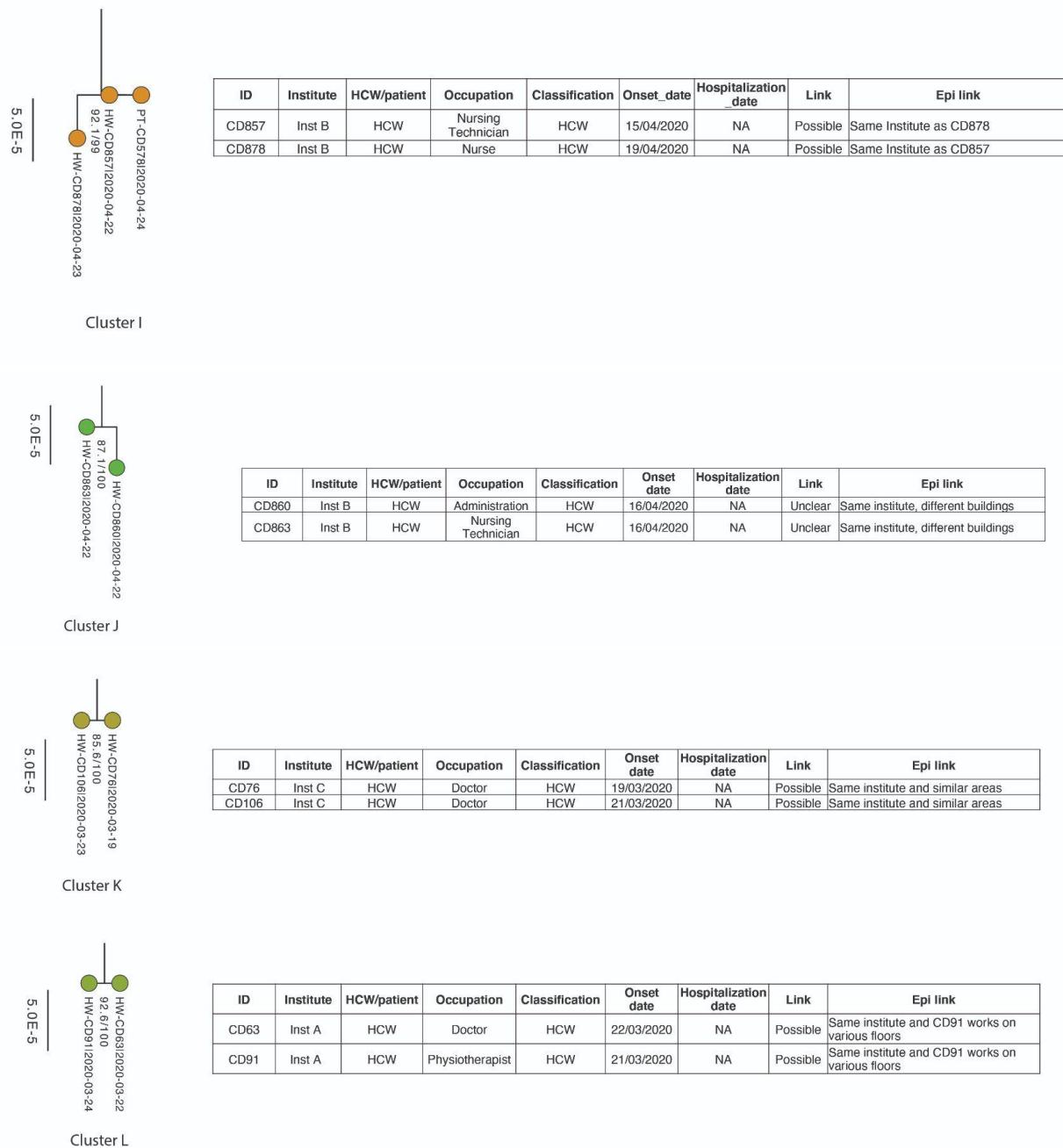
**Fig S14. Frequency of epidemiological link strength by phylogenetic cluster.**



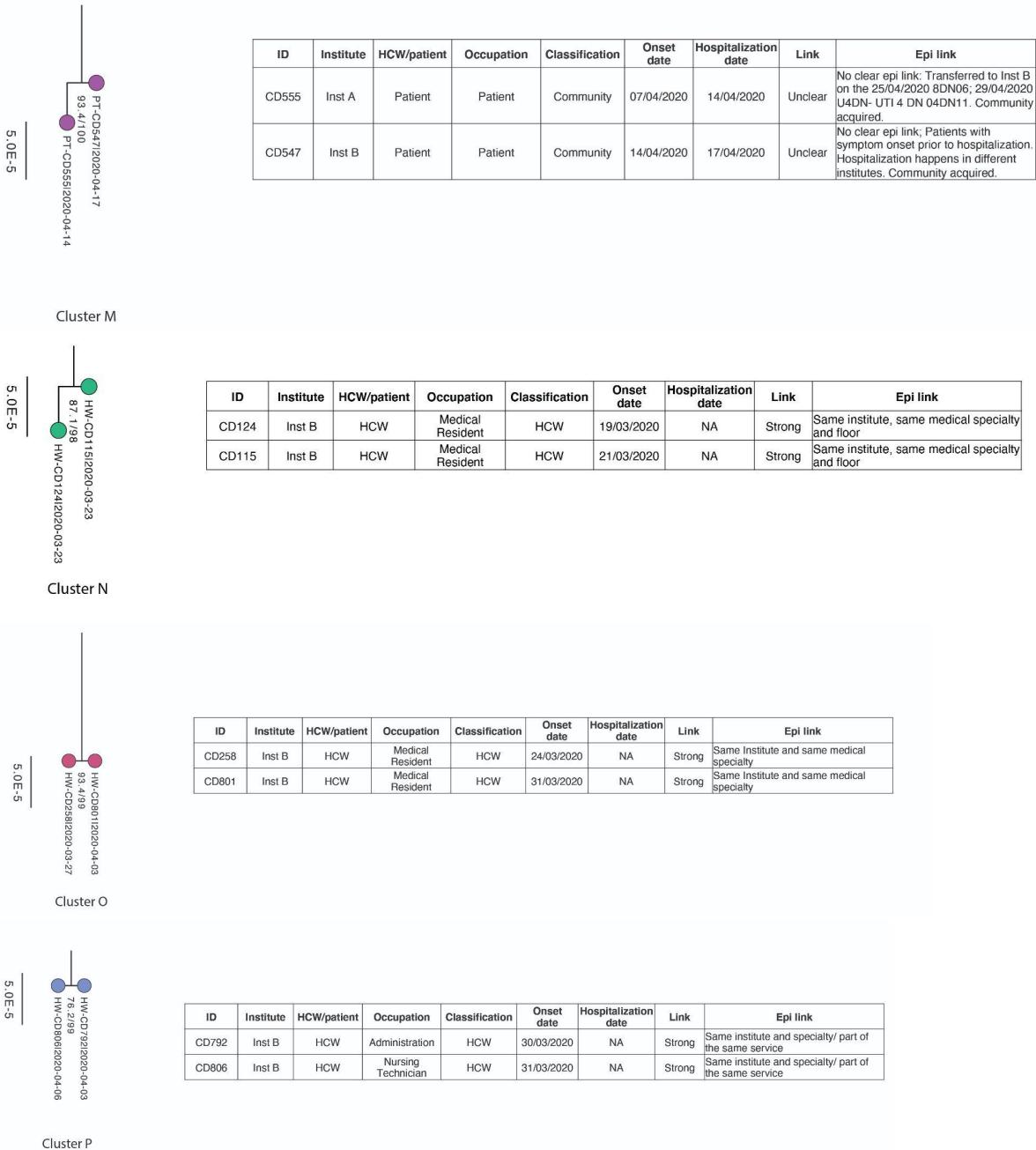
**Fig S15. Maximum likelihood phylogenetic subtrees and epidemiological characteristics of samples in transmission clusters A, B, C, and D. Cluster subtrees were extracted from an ML tree inferred from Dataset C (see methods).**



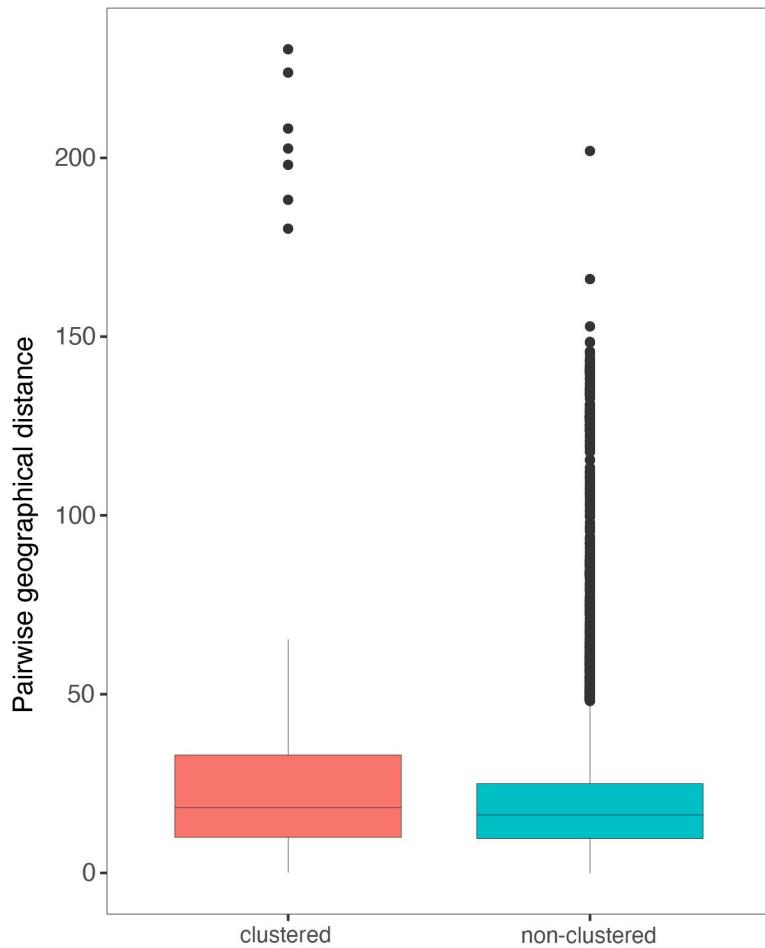
**Fig S16. Maximum likelihood phylogenetic subtrees and epidemiological characteristics of samples in transmission clusters E, F, G, and H.** Cluster subtrees were extracted from an ML tree inferred from Dataset C (see methods).



**Fig S17. Maximum likelihood phylogenetic subtrees and epidemiological characteristics of samples in transmission clusters I, J, K, and L.** Cluster subtrees were extracted from an ML tree inferred from Dataset C (see methods).



**Fig S18. Maximum likelihood phylogenetic subtrees and epidemiological characteristics of samples in transmission clusters M, N, O, and P. Cluster subtrees were extracted from an ML tree inferred from Dataset C (see methods).**



**Fig S19. Comparison between pairwise geographical distances of households of clustered and non-clustered sequences.** For clustered sequences, pairwise household geographical distances were estimated between sequences of the same cluster. For non-clustered sequences, pairwise household geographical distances were estimated between all sequences.

Table S1. Epidemiological and Demographic characteristics of all Hospital das Clínicas Complex (HC) and of Institute B, Institute C, Institute A, and other institutes.

	<b>HC (n = 3898)</b>				<b>Inst B (n = 2159, 55·4%)</b>				<b>Inst C (n = 716, 18·4%)</b>				<b>Inst A (n = 703, 18%)</b>			
	<b>All</b>	Patients	HW	p- valu- e	All	Patients	HCW		All	Patients	HCW		All	Patients	HCW	
		(n = 2008)	(n = 1890)			(n = 1468)	(n = 691)			(n = 215)	(n = 501)			(n = 269)	(n = 434)	
<b>Age</b>	46 (0-101)	60 (0 - 101)	37 (17-84)	<0·001	52 (0-97)	60 (0-101)	38 (17-71)	<0·0001	39 (0-93)	61 (0-93)	35 (20-67)	<0·0001	42 (15-92)	64 (15-92)	37 (19-66)	<0·0001
<b>Sex</b>																
Female	2251 (57·75%)	921 (45·9%)	1330 (70·4%)	<0·001	1112 (51·5%)	665 (45·3%)	447 (64·7%)		456 (63·7%)	92 (42·7%)	364 (72·65%)	<0·0001	474 (67·4%)	140 (52·0%)	334 (76·9%)	<0·0001
Male	1647 (42·25%)	1087 (54·1%)	560 (29·6%)		1047 (48·5%)	803 (54·7%)	244 (35·3%)		258 (36·3%)	123 (57·2%)	137 (27·34%)		229 (32·6%)	129 (48·0%)	100 (23·1%)	
<b>Occupation</b>																
Nursing Technician	557 (14·3%)	-	557 (29·5%)	-	172 (8·0%)	-	172 (24·9%)		156 (21·8%)	-	156 (31·1%)	-	171 (24·3%)	-	171 (39·4%)	-
Doctor	421 (10·8%)	-	421 (22·3%)		188 (8·7%)	-	188 (27·2%)		96 (13·4%)	-	96 (19·2%)		65 (9·2%)	-	65 (15·0%)	
Administrative	295 (7·6%)	-	295 (15·6%)		130 (6·0%)	-	130 (18·8%)		64 (8·9%)	-	64 (12·8%)		56 (8·0%)	-	56 (8·0%)	
Nurse	282 (7·2%)	-	282 (14·9%)		70 (3·2%)	-	70 (10·1%)		95 (13·3%)	-	95 (19·0%)		81 (11·5%)	-	81 (18·7%)	
Physiotherapist	44 (1·3%)	-	44 (2·3%)		16 (0·7%)	-	16 (2·3%)		14 (1·9%)	-	14 (2·8%)		10 (1·4%)	-	10 (2·3%)	
KitchenMaid	33 (0·85%)	-	33 (1·7%)		13 (0·6%)	-	13 (1·8%)		16 (2·2%)	-	16 (2·8%)		4 (0·6%)	-	4 (0·9%)	

Cleaning	22 (0·6%)	-	22 (1·2%)		12 (0·55%)	-	12 (1·7%)		2 (0·3%)	-	2 (0·4%)		0 (0·0%)	-	0 (0·0%)	
RadiologyTechnician	22 (0·6%)	-	22 (1·2%)		5 (0·2%)	-	5 (0·7%)		5 (0·7%)	-	5 (1·0%)		4 (0·6%)	-	4 (0·9%)	
Others	219 (5·6%)	-	219 (10·4%)		76 (3·4%)	-	76 (10·6%)		51 (7·1%)	-	51 (10·2%)		42 (6·0%)	-	42 (9·7%)	
Unknown	15 (0·4%)	-	15 (0·9%)		12 (0·55%)	-	12 (1·7%)		2 (0·3%)	-	2 (0·4%)		1 (0·1%)	-	1 (0·2%)	
<b>Sector</b>																
Inpatient	1099 (28·2%)	933 (46·5%)	166 (8·7%)	<0·0 001	815 (37·7%)	759 (51·7%)	56 (8·1%)		144 (20·1%)	90 (41·9%)	54 (10·8%)	<0·0001	103 (14·6%)	62 (23·0%)	41 (9·4%)	<0·0001
UIC	713 (18·3%)	422 (21·0%)	291 (15·4%)		406 (18·8%)	298 (20·3%)	165 (15·6%)		197 (27·5%)	87 (40·5%)	110 (21·9%)		90 (12·8%)	37 (13·7%)	53 (12·2%)	
Emergency	471 (12·1%)	357 (17·8%)	114 (6·0%)		230 (10·65%)	174 (11·8%)	56 (8·1%)		40 (5·6%)	14 (6·5%)	26 (5·2%)		185 (26·3%)	165 (61·3%)	20 (4·6%)	
Administration	86 (2·2%)	0 (0·0%)	86 (4·5%)		54 (2·5%)	0 (0·0%)	54 (7·8%)		11 (1·5%)	0 (0·0%)	11 (2·2%)		12 (1·7%)	0 (0·0%)	12 (2·8%)	
Outpatient	100 (2·6%)	10 (0·5%)	90 (4·8%)		60 (2·8%)	8 (0·5%)	39 (5·7%)		15 (2·1%)	2 (0·9%)	13 (2·6%)		5 (0·7%)	0 (0·0%)	11 (2·5%)	
Others	589 (15·1%)	59 (2·9%)	530 (28·0%)		180 (8·4%)	34 (2·6%)	165 (23·9%)		156 (21·8%)	18 (8·4%)	138 (27·5%)		153 (21·8%)	5 (1·8%)	148 (34·1%)	
Unknown	840 (21·5%)	227 (11·3%)	613 (32·4%)		414 (19·2%)	214 (14·6%)	200 (28·9%)		153 (21·4%)	4 (1·9%)	149 (29·7%)		149 (21·2%)	0 (0·0%)	149 (34·3%)	
<b>Outcome</b>																
Release	2946 (75·6%)	1061 (52·8%)	1885 (99·7%)	<0·0 001	1504 (69·6%)	814 (55·4%)	690 (99·85%)		596 (83·2%)	98 (45·6%)	498 (99·4%)	<0·0001	543 (77·24%)	109 (40·5%)	434 (100·0 %)	<0·0001
Death	577 (14·8%)	572 (28·5%)	5 (0·3%)		356 (16·5%)	355 (24·2%)	1 (0·15%)		78 (10·9%)	75 (34·9%)	3 (0·6%)		131 (18·6%)	131 (48·7%)	0 (0·0%)	

Transfer	36 (0·9%)	36 (2·0%)	0 (0·0%)		27 (1·2%)	27 (1·7%)	0 (0·0%)		5 (0·2%)	5 (2·2%)	0 (0·0%)		3 (0·4%)	3 (1·1%)	0 (0·0%)	
Unknown	339 (8·7%)	339 (16·9%)	0 (0·0%)		272 (12·6%)	272 (18·5%)	0 (0·0%)		37 (5·2%)	37 (17·2%)	0 (0·0%)		26 (3·7%)	26 (9·6%)	0 (0·0%)	
<b>Municipalit y</b>																
São Paulo City	2568 (65·9%)	1352 (67·3%)	1216 (64·3%)	<0·0 001	1534 (71·0%)	1017 (69·3%)	517 (74·8%)		431 (60·2%)	135 (62·8%)	296 (59·1%)	0·001	404 (57·5%)	157 (58·4%)	247 (56·9%)	0·0008
Others	1055 (27·1%)	656 (32·7%)	399 (21·1%)		605 (28·0%)	451 (30·7%)	154 (22·3%)		175 (24·4%)	80 (37·2%)	95 (19·0%)		211 (30·0%)	112 (41·6%)	99 (22·8%)	
Unknown	275 (7·0%)	0 (0·0%)	275 (14·6%)		20 (0·9%)	0 (0·0%)	20 (2·9%)		110 (15·4%)	0 (0·0%)	110 (21·9%)		88 (12·5%)	0 (0·0%)	88 (20·3%)	

Table S2. Epidemiological and demographic characteristics of groups 2, 3, and 4 patients from three HC complex institutes.

	Patients (n = 167)			p-value
	3 - 7 days (n=27)	8 - 14 days (n=54)	>14 days (n=86)	
	Group 2	Group 3	Group 4	
<b>Age</b>	64 (0 - 92)	63 (25 - 81)	60 (14 - 92)	0·16
<b>Sex</b>				
Female	10 (37·0%)	22 (40·7%)	42 (48·8%)	0·45
Male	17 (63%)	32 (59·3%)	44 (51·2%)	
<b>Institute</b>				
Inst B	0 (0·0%)	3 (5·5%)	1 (1·2%)	-
Inst C	18 (66·7%)	30 (55·5%)	70 (81·4%)	
Inst A	9 (33·3%)	21 (38·9%)	15 (17·4%)	
<b>Sector</b>				
Inpatient unit	13 (48·1%)	29 (53·7%)	41 (47·7%)	-
ICU	11 (40·7%)	18 (33·3%)	38 (44·2%)	
Emergency Room	1 (4·0%)	4 (7·4%)	5 (5·8%)	
Outpatients unit	0 (0·0%)	0 (0·0%)	2 (2·3%)	
Others	2 (7·4%)	1 (1·8%)	0 (0·0%)	
Unknown	0 (0·0%)	2 (3·7%)	0 (0·0%)	
<b>Outcome</b>				
Death	12 (44·5%)	25 (46·3%)	35 (40·7%)	0·93
Discharged	11 (40·7%)	19 (35·2%)	29 (33·7%)	
Transfer	0 (0·0%)	2 (3·7%)	4 (4·6%)	

Unknown	4 (14·8%)	8 (14·8%)	18 (20·9%)	
---------	-----------	-----------	------------	--

Table S3. Sequencing statistics for the Brazilian SARS-COV-2 genomes from this study (n=340).

Isolate	GISAID ID	New GISAID submission	Municipality	State	Collection Date	Mapped Reads	Average depth coverage	Bases covered >10x	Bases covered >25x	Reference covered (%)
CD100	EPI_ISL_476449	No	Sao Paulo	SP	2020-03-24	57472	717·771	27974	27153	89·9
CD1002	EPI_ISL_1172015	No	Sao Paulo	SP	2020-05-06	51838	645·976	26774	25293	81·8
CD1003	EPI_ISL_4463722	Yes	Sao Paulo	SP	2020-05-06	59284	746·913	26275	24424	79·6
CD1008	EPI_ISL_1172016	No	Sao Paulo	SP	2020-05-08	28954	361·532	26156	23357	79·2
CD101	EPI_ISL_476450	No	Sao Paulo	SP	2020-03-23	71837	888·494	29758	28960	95·2
CD1011	EPI_ISL_722006	No	Sao Paulo	SP	2020-05-09	63171	783·391	29222	29203	96·9
CD1014	EPI_ISL_721995	No	Sao Paulo	SP	2020-05-09	52807	657·624	29196	29016	96·2
CD1016	EPI_ISL_4463723	Yes	Sao Paulo	SP	2020-04-23	75669	941·732	29032	29019	96·1
CD1020	EPI_ISL_4463724	Yes	Sao Paulo	SP	2020-05-11	57547	719·699	28007	27286	89·6
CD1022	EPI_ISL_4463725	Yes	Cotia	SP	2020-05-11	131119	1622·29	29449	29209	97·0
CD1023	EPI_ISL_4463726	Yes	Sao Paulo	SP	2020-05-11	103425	1288·57	29268	28537	94·4
CD1025	EPI_ISL_4463727	Yes	Sao Paulo	SP	2020-04-27	86029	1068	26783	24805	85·4
CD105	EPI_ISL_476452	No	Sao Paulo	SP	2020-03-21	20454	252·938	29431	28406	95·9
CD106	EPI_ISL_476453	No	Sao Paulo	SP	2020-03-23	77578	959·98	29158	28215	93·5
CD107	EPI_ISL_476454	No	Sao Paulo	SP	2020-03-23	65714	830·558	26697	25044	83·4
CD109	EPI_ISL_476455	No	Sao Paulo	SP	2020-03-24	41571	523·947	26899	24987	81·8
CD110	EPI_ISL_476456	No	Sao Paulo	SP	2020-03-23	54967	679·223	29177	28384	94·3
CD111	EPI_ISL_476457	No	Sao Paulo	SP	2020-03-21	32555	406·751	27655	25681	87·3
CD1129	EPI_ISL_4463728	Yes	Poa	SP	2020-05-04	212992	2381·59	29232	29028	96·3
CD113	EPI_ISL_476459	No	Sao Paulo	SP	2020-03-20	43796	550·345	27191	25287	83·2
CD1130	EPI_ISL_4463729	Yes	Sao Paulo	SP	2020-04-27	222158	1752·33	23615	22723	75·0
CD1131	EPI_ISL_4297851	Yes	Sao Paulo	SP	2020-04-30	215848	1559·04	27664	25739	87·3
CD1132	EPI_ISL_1171871	No	Sao Paulo	SP	2020-05-04	273611	2804·66	29276	29072	96·3
CD1133	EPI_ISL_4297852	Yes	Sao Paulo	SP	2020-05-04	186352	2297·73	29286	29208	97·2
CD1134	EPI_ISL_4297853	Yes	Embu das Artes	SP	2020-05-04	202136	2529·59	29466	29278	97·2
CD1135	EPI_ISL_4297854	Yes	Sao Paulo	SP	2020-05-05	157917	1920·49	29238	29071	96·3
CD114	EPI_ISL_476460	No	Guarulhos	SP	2020-03-20	41489	521·837	26433	24987	80·7

CD115	EPI_ISL_476461	No	Sao Paulo	SP	2020-03-23	67205	830·874	29834	29184	96074
CD116	EPI_ISL_476462	No	Sao Paulo	SP	2020-03-23	65774	815·213	29615	28988	96·8
CD118	EPI_ISL_476469	No	Sao Paulo	SP	2020-03-24	128482	1626·99	29836	29836	98·6
CD121	EPI_ISL_476471	No	Jandira	SP	2020-03-17	104908	1326·95	29588	29384	96·9
CD122	EPI_ISL_476472	No	Sao Paulo	SP	2020-03-19	178589	2139·45	29147	28704	95·3
CD124	EPI_ISL_476473	No	Sao Paulo	SP	2020-03-23	181356	1898·82	27552	26950	89·1
CD126	EPI_ISL_476475	No	Sao Paulo	SP	2020-03-24	177196	2017·32	28628	28352	92·7
CD1261	EPI_ISL_4463730	Yes	Guarulhos	SP	2020-05-14	18289	226·612	29213	29206	96·9
CD1262	EPI_ISL_4297855	Yes	Sao Paulo	SP	2020-05-14	22872	284·946	26814	25000	83·2
CD1264	EPI_ISL_4297856	Yes	Sao Paulo	SP	2020-05-14	19422	241·971	25801	22681	75·7
CD1265	EPI_ISL_4297857	Yes	Sao Paulo	SP	2020-05-14	12256	152·797	27388	24988	86·2
CD1268	EPI_ISL_4297858	Yes	Sao Paulo	SP	2020-05-14	24900	311·296	27942	26239	87·0
CD1269	EPI_ISL_4297859	Yes	Santo Andre	SP	2020-05-13	19845	244·974	29167	29094	96·4
CD1276	EPI_ISL_4297860	Yes	Sao Paulo	SP	2020-05-12	9924	122·826	28702	27446	92·2
CD1278	EPI_ISL_4297861	Yes	Sao Paulo	SP	2020-05-14	24601	303·813	29458	29457	98·0
CD1279	EPI_ISL_4297862	Yes	Franco da Rocha	SP	2020-05-14	19482	246·765	25865	22679	77·4
CD128	EPI_ISL_476477	No	Sao Paulo	SP	2020-03-23	116345	1473·48	28847	28564	93·5
CD1283	EPI_ISL_4297863	Yes	Sao Paulo	SP	2020-05-15	34797	450·457	25431	22832	76·0
CD1284	EPI_ISL_4297864	Yes	Sao Paulo	SP	2020-05-15	85433	1077·47	28014	26851	88·1
CD1285	EPI_ISL_4297865	Yes	Embu das Artes	SP	2020-05-15	16228	202·795	29270	29003	97·1
CD1287	EPI_ISL_4297866	Yes	Sao Paulo	SP	2020-05-15	110708	1384·22	28880	27227	89·6
CD1288	EPI_ISL_4297867	Yes	Carapicuiba	SP	2020-05-15	70865	886·936	29039	29025	96·2
CD1289	EPI_ISL_4297868	Yes	Sao Paulo	SP	2020-05-15	69833	869·803	26346	23972	80·3
CD1292	EPI_ISL_4463731	Yes	Sao Paulo	SP	2020-05-14	32859	417·543	25827	23644	77·9
CD1293	EPI_ISL_4297869	Yes	Sao Paulo	SP	2020-05-15	104135	1298·02	29458	29269	97·1
CD1297	EPI_ISL_4297870	Yes	Sao Paulo	SP	2020-05-14	101437	1269·58	29274	28864	96·3
CD1298	EPI_ISL_4297871	Yes	Guarulhos	SP	2020-05-15	98654	1234·54	29273	29254	97·1
CD1300	EPI_ISL_4297872	Yes	Diadema	SP	2020-05-15	57045	705·346	29002	28100	94·0
CD1303	EPI_ISL_4297873	Yes	Sao Paulo	SP	2020-05-14	73375	921·437	28821	27582	92·4
CD1305	EPI_ISL_4297874	Yes	Osasco	SP	2020-05-17	76351	952·286	26871	25362	85·4
CD1306	EPI_ISL_4297875	Yes	Sao Paulo	SP	2020-05-20	81809	1004·05	27076	25471	85·4

CD131	EPI_ISL_476479	No	Sao Paulo	SP	2020-03-19	59589	788·547	26108	24493	80·6
CD1316	EPI_ISL_4297876	Yes	Sao Paulo	SP	2020-05-17	47548	595·197	27875	27049	89·0
CD133	EPI_ISL_476480	No	Taboao da Serra	SP	2020-03-23	112350	1434·81	28401	26871	87·1
CD134	EPI_ISL_476481	No	Sao Paulo	SP	2020-03-23	60637	766·611	28958	28353	92·7
CD135	EPI_ISL_476482	No	Sao Paulo	SP	2020-03-24	70126	895·417	28290	27135	90·7
CD139	EPI_ISL_476486	No	Sao Paulo	SP	2020-03-24	114744	1452·87	29408	28978	95·3
CD231	EPI_ISL_476237	No	Francisco Morato	SP	2020-03-26	101131	1272·4	29053	28318	93·4
CD232	EPI_ISL_476238	No	Sao Paulo	SP	2020-03-26	116552	1282·01	28367	26189	87·4
CD234	EPI_ISL_476239	No	Guarulhos	SP	2020-03-26	83874	1050·7	28362	26806	89·8
CD235	EPI_ISL_476240	No	Santo Andre	SP	2020-03-26	103474	1234·57	28585	26618	89·6
CD236	EPI_ISL_476241	No	Carapicuiba	SP	2020-03-24	63095	679·685	26516	23297	78·8
CD237	EPI_ISL_476242	No	Osasco	SP	2020-03-26	102358	1103·6	27206	25196	83·8
CD239	EPI_ISL_476243	No	Mogi das Cruzes	SP	2020-03-23	91282	1159·16	29300	28991	95·3
CD240	EPI_ISL_476244	No	Sao Paulo	SP	2020-03-26	66959	848·48	29675	29675	97·9
CD241	EPI_ISL_476245	No	Embu das Artes	SP	2020-03-26	82197	1044·52	29371	28662	95·7
CD242	EPI_ISL_476246	No	Sao Paulo	SP	2020-03-27	126660	1601·78	29675	29640	97·9
CD243	EPI_ISL_476247	No	Sao Paulo	SP	2020-03-27	95802	1201·52	29046	28048	92·6
CD246	EPI_ISL_476249	No	Sao Paulo	SP	2020-03-27	90130	1049·37	27902	26164	88·1
CD250	EPI_ISL_476250	No	Taboao da Serra	SP	2020-03-25	52724	667·049	29486	29038	96·3
CD252	EPI_ISL_476251	No	Ferraz de Vasconcelos	SP	2020-03-21	72855	784·277	25639	22858	75·8
CD255	EPI_ISL_476252	No	Sao Paulo	SP	2020-03-27	69676	868·835	27424	25565	86·5
CD257	EPI_ISL_476253	No	Sao Paulo	SP	2020-03-27	75996	976·113	29039	27664	93·6
CD258	EPI_ISL_476254	No	Sao Paulo	SP	2020-03-27	71401	904·888	29490	29478	97·1
CD260	EPI_ISL_476256	No	Osasco	SP	2020-03-27	58654	742·317	29674	29670	97·9
CD262	EPI_ISL_476257	No	Sao Paulo	SP	2020-03-24	64287	807·039	27683	25098	85·3
CD263	EPI_ISL_476258	No	Sao Paulo	SP	2020-03-27	54650	662·635	28553	27830	92·5
CD265	EPI_ISL_476259	No	Sao Paulo	SP	2020-03-29	14932	178·592	29667	29363	97·9
CD266	EPI_ISL_476260	No	Sao Paulo	SP	2020-03-27	41575	497·046	29367	28972	97·1
CD267	EPI_ISL_476261	No	Sao Paulo	SP	2020-03-29	36392	434·083	29399	28329	95·3

CD268	EPI_ISL_476262	No	Sao Paulo	SP	2020-03-29	38441	460·513	29668	29201	97·1
CD270	EPI_ISL_476263	No	Sao Paulo	SP	2020-03-30	41739	500·181	29379	28405	94·4
CD271	EPI_ISL_476264	No	Sao Paulo	SP	2020-03-30	32654	400·351	26509	24127	81·2
CD272	EPI_ISL_476265	No	Sao Paulo	SP	2020-03-30	38024	454·495	29673	29659	97·9
CD273	EPI_ISL_476266	No	Sao Paulo	SP	2020-03-30	46695	559·593	29164	28698	96·0
CD274	EPI_ISL_476267	No	Franco da Rocha	SP	2020-03-30	11820	141·811	29178	28727	96·9
CD277	EPI_ISL_476269	No	Sao Paulo	SP	2020-03-31	7798	941·438	28663	26476	88·8
CD278	EPI_ISL_476270	No	Franco da Rocha	SP	2020-03-30	33977	408·999	29118	28443	94·2
CD280	EPI_ISL_476271	No	Guaruja	SP	2020-03-30	7473	903·933	28103	25610	88·1
CD281	EPI_ISL_476272	No	Sao Paulo	SP	2020-03-30	38254	462·456	29001	28463	94·3
CD282	EPI_ISL_476273	No	Sao Paulo	SP	2020-03-30	26792	321·717	28721	27563	93·4
CD283	EPI_ISL_476274	No	Embu das Artes	SP	2020-03-30	48921	586·647	29434	28749	96·1
CD284	EPI_ISL_476275	No	Sao Paulo	SP	2020-03-30	20633	250·424	27038	24094	81·1
CD285	EPI_ISL_476276	No	Embu das Artes	SP	2020-03-30	19201	233·625	25884	22849	78·6
CD286	EPI_ISL_476277	No	Sao Paulo	SP	2020-03-30	19110	228·81	29346	27597	92·9
CD292	EPI_ISL_4297877	Yes	Sao Paulo	SP	2020-03-30	597600	3175·86	29446	29134	96·1
CD293	EPI_ISL_4297878	Yes	Sao Paulo	SP	2020-03-26	519575	4319·42	29838	29689	97·9
CD37	EPI_ISL_4297879	Yes	Sao Paulo	SP	2020-03-16	88136	1104·21	29605	29287	97·4
CD38	EPI_ISL_476488	No	Santana de Parnaiba	SP	2020-03-16	47079	579·429	29515	28991	96·9
CD39	EPI_ISL_4297880	Yes	Sao Paulo	SP	2020-03-16	98513	1230·56	29327	29078	97·4
CD41	EPI_ISL_1084733	No	Sao Paulo	SP	2020-03-16	48675	603·222	29287	29040	97·4
CD43	EPI_ISL_4297881	Yes	Sao Paulo	SP	2020-03-16	24035	303·04	29287	28888	96·5
CD44	EPI_ISL_4297882	Yes	Sao Paulo	SP	2020-03-17	45850	573·558	26872	25549	82·9
CD45	EPI_ISL_4297883	Yes	Sao Paulo	SP	2020-03-17	22577	279·997	29169	27716	94·6
CD46	EPI_ISL_4297884	Yes	Sao Paulo	SP	2020-03-17	42829	533·776	28149	27170	89·3
CD473	EPI_ISL_476372	No	Sao Paulo	SP	2020-04-01	76230	966·8	27332	25152	83·7
CD474	EPI_ISL_476373	No	Sao Paulo	SP	2020-04-01	229448	2825·63	29622	29483	98·7
CD475	EPI_ISL_476374	No	Sao Paulo	SP	2020-04-02	73561	932·902	29204	28712	96·1
CD476	EPI_ISL_476375	No	Osasco	SP	2020-03-20	77453	972·914	29463	29457	97·9
CD477	EPI_ISL_476376	No	Cubatao	SP	2020-04-01	100531	1265·78	29192	28564	95·2

CD479	EPI_ISL_476377	No	Sao Paulo	SP	2020-04-01	137720	1666·51	28386	27507	92·7
CD48	EPI_ISL_4348336	Yes	Sao Paulo	SP	2020-03-17	89974	1140·18	29044	29034	96·4
CD481	EPI_ISL_476378	No	Ribeirao Preto	SP	2020-04-01	73322	929·21	26710	24749	80·4
CD483	EPI_ISL_476379	No	Sao Paulo	SP	2020-04-05	135173	1701·37	29460	29385	97·6
CD484	EPI_ISL_476380	No	Itaquaquecetuba	SP	2020-04-05	126972	1604·03	29467	29457	98·0
CD485	EPI_ISL_476381	No	Diadema	SP	2020-04-05	123389	1536·73	27368	26279	86·3
CD487	EPI_ISL_476383	No	Sao Paulo	SP	2020-04-04	137022	1728·5	29467	29458	98·0
CD488	EPI_ISL_476384	No	Santo Andre	SP	2020-04-03	39061	487·111	29245	27570	95·6
CD49	EPI_ISL_4348337	Yes	Sao Paulo	SP	2020-03-18	50478	634·468	29489	29070	96·5
CD490	EPI_ISL_476386	No	Sao Paulo	SP	2020-04-06	140378	1755·94	29465	29038	97·9
CD491	EPI_ISL_4297885	Yes	Sao Paulo	SP	2020-04-05	28197	342·312	29560	28505	95·2
CD492	EPI_ISL_4297886	Yes	Osasco	SP	2020-04-05	16177	205·257	25995	23778	78·5
CD495	EPI_ISL_4297887	Yes	Sao Paulo	SP	2020-04-03	11925	147·621	27027	23858	81·7
CD496	EPI_ISL_4297888	Yes	Francisco Morato	SP	2020-04-06	24325	299·001	28565	27680	91·7
CD497	EPI_ISL_4297889	Yes	Sao Paulo	SP	2020-04-07	33136	407·832	29432	28084	95·3
CD498	EPI_ISL_4297890	Yes	Sao Paulo	SP	2020-04-07	16591	206·238	26124	23909	80·8
CD499	EPI_ISL_4297891	Yes	Sao Paulo	SP	2020-04-07	30383	373·876	29567	28835	95·4
CD501	EPI_ISL_4297892	Yes	Sao Paulo	SP	2020-04-07	8399	104·272	27820	24799	81·9
CD502	EPI_ISL_4297893	Yes	Sao Paulo	SP	2020-04-07	39566	484·976	28399	26983	91·2
CD503	EPI_ISL_4297894	Yes	Sorocaba	SP	2020-04-08	22631	277·592	29228	28746	95·2
CD504	EPI_ISL_4297895	Yes	Poa	SP	2020-04-07	39593	487·6	29674	29657	98·9
CD505	EPI_ISL_4297896	Yes	Sao Paulo	SP	2020-04-08	24050	297·846	27953	26564	88·2
CD51	EPI_ISL_4468752	Yes	Sao Paulo	SP	2020-03-18	48029	609·5	27241	25667	85·6
CD511	EPI_ISL_4297898	Yes	Sao Paulo	SP	2020-04-10	38607	474·391	28154	26303	88·2
CD515	EPI_ISL_4297899	Yes	Sao Paulo	SP	2020-04-10	10191	122·774	25351	23254	77·2
CD518	EPI_ISL_4297900	Yes	Sao Paulo	SP	2020-04-11	12660	151·008	26360	24008	79·0
CD52	EPI_ISL_4348338	Yes	Sao Paulo	SP	2020-03-15	44635	554·686	29097	28600	96·6
CD520	EPI_ISL_1084739	No	Sao Paulo	SP	2020-04-12	19540	240·413	27693	26564	87·6
CD525	EPI_ISL_672687	No	Sao Paulo	SP	2020-04-11	109980	1380·34	29465	29461	97·9
CD527	EPI_ISL_672688	No	Sao Paulo	SP	2020-04-14	87200	1101·11	29484	29229	97·9
CD528	EPI_ISL_672689	No	Sao Paulo	SP	2020-04-10	77425	896·929	25683	24112	79·5

CD529	EPI_ISL_672690	No	Sao Paulo	SP	2020-04-13	61332	769·405	29048	27807	93·4
CD53	EPI_ISL_4297901	Yes	Sao Paulo	SP	2020-03-18	32518	421·449	26742	25640	84·2
CD530	EPI_ISL_672691	No	Sao Paulo	SP	2020-04-13	94341	1173·91	28787	27663	92·7
CD532	EPI_ISL_672692	No	Sao Paulo	SP	2020-04-14	99021	1152·67	25130	22673	77·1
CD533	EPI_ISL_672693	No	Carapicuiba	SP	2020-04-11	83537	1043·22	29220	28867	96·0
CD534	EPI_ISL_672694	No	Sao Paulo	SP	2020-04-14	113237	1421·98	28798	27919	93·5
CD537	EPI_ISL_672695	No	Carapicuiba	SP	2020-04-15	135664	1703·09	29455	29019	97·0
CD538	EPI_ISL_672696	No	Sao Paulo	SP	2020-04-14	61630	777·989	29263	28951	96·1
CD539	EPI_ISL_672697	No	Sao Paulo	SP	2020-04-15	83953	1053·95	28626	28473	94·4
CD54	EPI_ISL_476428	No	Taboao da Serra	SP	2020-03-18	123850	1543·76	29245	28629	96·1
CD540	EPI_ISL_672698	No	Sao Paulo	SP	2020-04-15	63383	807·511	27724	27514	90·1
CD541	EPI_ISL_672699	No	Carapicuiba	SP	2020-04-15	88326	1117·95	29250	28782	96·2
CD542	EPI_ISL_672722	No	Sao Paulo	SP	2020-04-16	111549	1412·65	29275	28997	96·2
CD543	EPI_ISL_672723	No	Sao Paulo	SP	2020-04-16	96021	1222·33	27501	26754	88·3
CD545	EPI_ISL_672724	No	Sao Paulo	SP	2020-04-17	127010	1563·7	28765	27769	91·7
CD547	EPI_ISL_672725	No	Sao Paulo	SP	2020-04-17	113661	1421·56	29461	29456	97·8
CD548	EPI_ISL_672726	No	Sao Paulo	SP	2020-04-16	147928	1623·33	28358	26902	90·7
CD549	EPI_ISL_672727	No	Fortaleza	CE	2020-04-15	114951	1287·82	27184	24984	83·6
CD550	EPI_ISL_672728	No	Sao Paulo	SP	2020-04-18	134753	1682·33	29612	29257	97·1
CD551	EPI_ISL_672729	No	Santo Andre	SP	2020-04-18	44935	586·03	24853	23013	76·5
CD553	EPI_ISL_672730	No	Sao Paulo	SP	2020-04-18	92380	1175·43	28421	27528	92·1
CD554	EPI_ISL_672731	No	Jandira	SP	2020-04-16	111597	1394·08	29488	29460	97·9
CD555	EPI_ISL_672732	No	Guarulhos	SP	2020-04-14	140320	1727·59	28869	28425	94·4
CD556	EPI_ISL_672733	No	Sao Paulo	SP	2020-04-13	106419	1277·59	27768	25804	88·3
CD557	EPI_ISL_672734	No	Sao Paulo	SP	2020-04-14	172160	2019·94	29461	29027	97·1
CD558	EPI_ISL_672735	No	Ferraz de Vasconcelos	SP	2020-04-20	157736	1984·08	29461	29458	97·9
CD559	EPI_ISL_672736	No	Sao Paulo	SP	2020-04-18	116739	1476·24	29435	29185	96·9
CD56	EPI_ISL_1084735	No	Sao Paulo	SP	2020-03-18	23237	286·746	28809	26612	89·3
CD561	EPI_ISL_672737	No	Sao Paulo	SP	2020-04-20	135627	1441·47	27196	25652	86·7
CD564	EPI_ISL_672738	No	Sao Paulo	SP	2020-04-20	96139	1019·69	25985	24357	79·7
CD565	EPI_ISL_672739	No	Sao Paulo	SP	2020-03-26	66925	848·104	29209	29207	97·0

CD566	EPI_ISL_672740	No	Sao Paulo	SP	2020-04-21	131952	1664·68	29223	28368	93·5
CD567	EPI_ISL_672669	No	Sao Paulo	SP	2020-04-20	134644	1535·12	27745	26550	89·0
CD568	EPI_ISL_672741	No	Sao Paulo	SP	2020-04-19	234778	2795·74	29466	29457	97·9
CD57	EPI_ISL_476430	No	Sao Paulo	SP	2020-03-20	41754	535·563	25413	23064	76·8
CD570	EPI_ISL_672742	No	Sao Paulo	SP	2020-04-22	158065	1932·58	29067	27974	93·6
CD572	EPI_ISL_672743	No	Sao Paulo	SP	2020-04-23	96941	1221·04	27323	25921	86·5
CD578	EPI_ISL_672745	No	Guarulhos	SP	2020-04-20	27160	341·853	28602	27367	91·7
CD61	EPI_ISL_476431	No	Sao Paulo	SP	2020-03-21	73827	925·342	29049	28364	94·3
CD623	EPI_ISL_672700	No	Sao Paulo	SP	2020-04-12	93599	1178·02	29204	28526	95·2
CD63	EPI_ISL_476432	No	Sao Paulo	SP	2020-03-22	86292	1080·67	28981	28508	94·1
CD65	EPI_ISL_476433	No	Sao Paulo	SP	2020-03-18	66129	827·86	28648	27076	91·5
CD66	EPI_ISL_476434	No	Sao Paulo	SP	2020-03-25	108446	1370·37	28172	27686	90·0
CD67	EPI_ISL_476435	No	Sao Paulo	SP	2020-03-19	206616	2584·53	29836	29835	98·7
CD69	EPI_ISL_476436	No	Sao Paulo	SP	2020-03-23	77974	977·901	28569	27104	90·1
CD70	EPI_ISL_476437	No	Sao Paulo	SP	2020-03-23	126712	1589·14	29209	28996	96·0
CD71	EPI_ISL_476438	No	Sao Paulo	SP	2020-03-23	124416	1510·78	28105	27099	88·3
CD72	EPI_ISL_476439	No	Sao Paulo	SP	2020-03-24	186287	2333	29836	29835	98·6
CD73	EPI_ISL_476440	No	Sao Paulo	SP	2020-03-24	78479	1019·37	26043	23749	79·8
CD74	EPI_ISL_476441	No	Sao Paulo	SP	2020-03-25	133179	1665·57	29815	29361	96·9
CD75	EPI_ISL_476442	No	Sao Paulo	SP	2020-03-25	99343	1222·57	26806	24969	81·1
CD76	EPI_ISL_476443	No	Sao Paulo	SP	2020-03-19	116959	1463·61	29833	29385	97·9
CD788	EPI_ISL_722057	No	Niteroi	RJ	2020-04-02	58354	737·503	28131	27542	90·8
CD789	EPI_ISL_722080	No	Curitiba	PR	2020-04-02	79447	997·21	29053	28143	93·5
CD79	EPI_ISL_476445	No	Sao Paulo	SP	2020-03-23	93416	1181·27	28769	27609	90·0
CD790	EPI_ISL_722108	No	Sao Paulo	SP	2020-04-02	42148	529·6	26583	24991	83·3
CD791	EPI_ISL_722004	No	Sao Paulo	SP	2020-04-03	45349	569·229	29453	28808	96·8
CD792	EPI_ISL_722078	No	Taboao da Serra	SP	2020-04-03	40231	503·691	28594	28166	93·5
CD795	EPI_ISL_722079	No	Sao Paulo	SP	2020-04-03	59114	739·384	28831	27964	93·6
CD797	EPI_ISL_722065	No	Sao Paulo	SP	29/03/2020	46821	592·291	28623	27888	91·8
CD798	EPI_ISL_722023	No	Sao Paulo	SP	2020-04-03	53399	669·793	29458	28854	97·8
CD800	EPI_ISL_722119	No	Sao Paulo	SP	2020-04-03	35556	444·949	27822	26197	86·6

CD801	EPI_ISL_722012	No	Sao Paulo	SP	2020-04-03	44273	554·611	29452	28980	97·0
CD802	EPI_ISL_722082	No	Caierias	SP	2020-04-06	35736	448·464	29298	28291	94·0
CD803	EPI_ISL_721987	No	Sao Paulo	SP	2020-04-06	20822	260·668	29457	28282	95·1
CD806	EPI_ISL_722029	No	Cruzeiro	SP	2020-04-06	54086	672·156	29459	29436	97·9
CD807	EPI_ISL_722122	No	Sao Paulo	SP	2020-04-06	30347	378·962	27571	25775	87·7
CD81	EPI_ISL_476446	No	Sao Paulo	SP	2020-03-24	115174	1440·66	29836	29675	98·6
CD812	EPI_ISL_722084	No	Sao Paulo	SP	2020-04-06	30509	383·491	29028	28567	94·4
CD813	EPI_ISL_722063	No	Sao Paulo	SP	2020-04-06	42268	530·851	28099	26720	90·1
CD814	EPI_ISL_722099	No	Sao Paulo	SP	2020-03-16	39044	496·104	26753	24608	81·5
CD815	EPI_ISL_722067	No	Sao Paulo	SP	2020-04-08	34343	431·028	29163	28106	92·5
CD816	EPI_ISL_722068	No	Sao Paulo	SP	2020-04-08	31183	390·924	28599	27360	92·5
CD818	EPI_ISL_722030	No	Sao Paulo	SP	2020-04-08	93195	1166·59	29469	29457	98·0
CD819	EPI_ISL_722031	No	Sao Paulo	SP	2020-04-08	85293	1074·33	29461	29453	98·0
CD82	EPI_ISL_476447	No	Sao Paulo	SP	2020-03-24	118677	1483·26	29809	29182	96·0
CD820	EPI_ISL_722055	No	Sao Paulo	SP	2020-04-08	60449	761·781	27968	27477	90·3
CD822	EPI_ISL_722005	No	Sao Paulo	SP	2020-04-06	70936	893·995	29284	28978	96·9
CD823	EPI_ISL_722000	No	Guarulhos	SP	2020-04-08	114662	1446·86	29085	29059	96·3
CD825	EPI_ISL_722013	No	Mogi das Cruzes	SP	2020-04-09	65666	827·279	29457	29062	97·0
CD827	EPI_ISL_722059	No	Sao Paulo	SP	2020-04-09	93612	1179·65	28447	27272	90·0
CD828	EPI_ISL_722042	No	Sao Paulo	SP	2020-04-09	89017	1122·76	29463	29254	97·9
CD829	EPI_ISL_1172014	No	Sao Paulo	SP	2020-04-09	119250	1501·91	29461	29446	98·0
CD83	EPI_ISL_476448	No	Sao Paulo	SP	2020-03-24	42305	529·617	29455	28602	95·1
CD830	EPI_ISL_722043	No	Sao Paulo	SP	2020-04-06	57666	725·969	29460	29456	97·9
CD832	EPI_ISL_722018	No	Sao Paulo	SP	2020-04-08	84262	1064·81	29459	29260	97·1
CD833	EPI_ISL_722075	No	Mongagua	SP	2020-04-09	71419	907·93	28650	27139	92·7
CD834	EPI_ISL_722090	No	Sao Paulo	SP	2020-04-09	42274	550·586	25131	23123	74·7
CD835	EPI_ISL_722087	No	Sao Paulo	SP	2020-04-24	98892	1248·82	29066	28410	94·5
CD836	EPI_ISL_722032	No	Sao Paulo	SP	2020-04-13	109400	1379·24	29606	29459	97·9
CD837	EPI_ISL_722044	No	Sao Paulo	SP	2020-04-24	92592	1171·59	29461	29184	98·0
CD838	EPI_ISL_722054	No	Embu das Artes	SP	2020-04-14	113913	1422·86	28187	27051	90·1
CD839	EPI_ISL_722033	No	Carapicuiba	SP	2020-04-14	142469	1802·65	29464	29458	97·9

CD840	EPI_ISL_722045	No	Sao Paulo	SP	2020-04-15	104844	1318·98	29570	29461	98·0
CD841	EPI_ISL_721992	No	Sao Paulo	SP	2020-04-15	79095	981·914	29008	28618	95·4
CD842	EPI_ISL_722127	No	Sao Paulo	SP	2020-04-15	86274	1065·42	28390	27236	89·1
CD847	EPI_ISL_722083	No	Sao Paulo	SP	2020-04-16	49956	617·73	28793	28084	93·3
CD848	EPI_ISL_722085	No	Sao Paulo	SP	2020-04-16	45678	566·059	28817	28592	94·4
CD849	EPI_ISL_722034	No	Vargem Grande Paulista	SP	2020-04-16	70820	878·6	29615	29462	97·9
CD851	EPI_ISL_722101	No	Sao Paulo	SP	2020-04-17	42820	531·941	26149	24430	81·5
CD852	EPI_ISL_722014	No	Sao Paulo	SP	2020-04-17	46461	575·073	29445	29242	97·1
CD853	EPI_ISL_722035	No	Sao Paulo	SP	2020-04-17	94658	1172·55	29595	29458	97·9
CD854	EPI_ISL_722114	No	Sao Paulo	SP	2020-04-17	62519	785·366	27799	26314	85·5
CD855	EPI_ISL_722072	No	Itapevi	SP	2020-04-22	65683	814·038	28746	27487	91·8
CD856	EPI_ISL_722009	No	Sao Paulo	SP	2020-04-22	36582	447·696	29459	29215	97·0
CD857	EPI_ISL_722003	No	Sao Paulo	SP	2020-04-22	75045	926·763	29072	29058	96·3
CD858	EPI_ISL_722036	No	Sao Paulo	SP	2020-04-22	62722	771·889	29451	29209	97·0
CD859	EPI_ISL_722025	No	Sao Paulo	SP	2020-04-22	27069	334·514	29459	29457	98·0
CD86	EPI_ISL_476463	No	Sao Paulo	SP	2020-03-19	107501	1307·77	27651	25580	85·7
CD860	EPI_ISL_722015	No	Sao Paulo	SP	2020-04-22	54638	674·279	29021	28588	96·1
CD861	EPI_ISL_722053	No	Santos	SP	2020-04-22	49020	617·106	27943	26378	88·5
CD862	EPI_ISL_722056	No	Itapecerica da Serra	SP	2020-04-22	24229	297·033	27983	26747	89·4
CD863	EPI_ISL_722073	No	Betim	MG	2020-04-22	63248	787·19	28258	27941	91·8
CD864	EPI_ISL_722088	No	Taboao da Serra	SP	2020-04-22	72866	902·116	28811	28388	93·6
CD865	EPI_ISL_722052	No	Sao Paulo	SP	2020-04-22	19277	241·341	27845	26993	88·9
CD867	EPI_ISL_722024	No	Sao Paulo	SP	2020-04-22	56958	698·706	29346	29022	97·0
CD868	EPI_ISL_722112	No	Sao Paulo	SP	2020-04-22	33239	414·859	26864	25696	83·0
CD869	EPI_ISL_722026	No	Sao Paulo	SP	2020-04-22	58877	727·822	29464	29459	97·9
CD87	EPI_ISL_476464	No	Franco da Rocha	SP	2020-03-22	119209	1474·11	29660	29359	96·9
CD870	EPI_ISL_722001	No	Sao Paulo	SP	2020-04-22	171146	2083·47	29098	29074	96·2
CD872	EPI_ISL_4463732	Yes	Sao Paulo	SP	2020-04-22	167208	2078·27	29283	28732	95·3
CD873	EPI_ISL_722046	No	Carapicuiba	SP	2020-04-23	100380	1261·42	29584	29457	98·0

CD874	EPI_ISL_722021	No	Mogi das Cruzes	SP	2020-04-23	132896	1662·19	29461	29063	97·2
CD875	EPI_ISL_722037	No	Sao Paulo	SP	2020-04-23	167929	2082·99	29486	29464	98·0
CD876	EPI_ISL_722047	No	Sao Paulo	SP	2020-04-23	248441	3008·91	29472	29459	97·9
CD877	EPI_ISL_722048	No	Sao Paulo	SP	2020-04-23	108612	1366·6	29485	29471	97·9
CD878	EPI_ISL_721989	No	Sao Paulo	SP	2020-04-23	153805	1892·97	29035	28247	95·3
CD879	EPI_ISL_722027	No	Guarulhos	SP	2020-04-23	127196	1592·73	29458	29457	97·9
CD88	EPI_ISL_476465	No	Cotia	SP	2020-03-23	58372	723·188	28909	28088	93·5
CD880	EPI_ISL_722038	No	Guarulhos	SP	2020-04-23	160425	1979·84	29489	29475	98·0
CD881	EPI_ISL_722129	No	Ribeirao Pires	SP	2020-04-23	205101	2523·22	29625	29463	98·7
CD882	EPI_ISL_722028	No	Sao Paulo	SP	2020-04-23	58656	736·312	29443	29434	97·9
CD883	EPI_ISL_722049	No	Barueri	SP	2020-04-23	121355	1522·82	29462	29456	97·9
CD89	EPI_ISL_476466	No	Osasco	SP	2020-03-24	26016	324·992	28090	25963	88·2
CD891	EPI_ISL_721996	No	Sao Paulo	SP	2020-04-24	46890	585·906	29271	29018	96·2
CD892	EPI_ISL_721997	No	Sao Paulo	SP	2020-04-24	50420	631·056	29271	29018	96·2
CD894	EPI_ISL_722110	No	Sao Paulo	SP	2020-04-24	16008	201·505	27307	24934	83·7
CD895	EPI_ISL_722102	No	Sao Paulo	SP	2020-04-24	42099	530·277	27031	25288	82·0
CD896	EPI_ISL_722091	No	Carapicuiba	SP	2020-04-24	22799	294·833	25057	23352	75·2
CD898	EPI_ISL_722061	No	Osasco	SP	2020-04-24	60293	751·503	29013	26341	91·4
CD899	EPI_ISL_721999	No	Sao Paulo	SP	2020-04-23	26940	338·261	29272	29018	96·2
CD90	EPI_ISL_476489	No	Sao Paulo	SP	2020-03-24	81155	1026·92	29834	29673	97·9
CD900	EPI_ISL_722039	No	Sao Paulo	SP	2020-04-14	53573	669·285	29461	29457	97·9
CD901	EPI_ISL_722008	No	Carapicuiba	SP	2020-04-24	38875	485·618	29457	29204	97·0
CD902	EPI_ISL_722051	No	Sao Paulo	SP	2020-04-24	24677	309·225	29459	29207	97·9
CD905	EPI_ISL_722076	No	Itapecerica da Serra	SP	2020-04-20	38183	480·91	29051	28036	93·5
CD907	EPI_ISL_4297902	Yes	Franco da Rocha	SP	2020-04-24	22528	286·073	25343	23500	76·8
CD91	EPI_ISL_476490	No	Sao Paulo	SP	2020-03-24	75295	953·394	29836	29744	98·6
CD910	EPI_ISL_722019	No	Sao Paulo	SP	2020-04-13	59049	738·823	29278	28999	97·1
CD917	EPI_ISL_722124	No	Sao Paulo	SP	2020-04-27	31168	393·726	28312	26547	88·3
CD918	EPI_ISL_722081	No	Itapevi	SP	2020-04-27	50142	629·492	29183	28061	93·5
CD92	EPI_ISL_476203	No	Sao Paulo	SP	2020-03-24	98925	1246·21	27521	26976	88·3

CD920	EPI_ISL_722109	No	Sao Paulo	SP	2020-04-27	33615	426·912	27836	25533	83·7
CD921	EPI_ISL_722016	No	Carapicuiba	SP	2020-04-25	69007	854·987	29281	29264	97·1
CD923	EPI_ISL_722060	No	Sao Paulo	SP	2020-04-27	101458	1256·98	28156	26849	91·0
CD924	EPI_ISL_722125	No	Jandira	SP	2020-04-28	52450	660·218	27756	26494	89·0
CD926	EPI_ISL_722126	No	Sao Paulo	SP	2020-04-28	87786	1086·25	27998	26905	89·0
CD927	EPI_ISL_722020	No	Sao Paulo	SP	2020-04-28	65734	810·487	29284	29273	97·1
CD928	EPI_ISL_722071	No	Sao Paulo	SP	2020-04-27	82248	1026·99	28853	27425	92·6
CD930	EPI_ISL_722064	No	Sao Paulo	SP	2020-04-29	66769	827·223	28360	27643	91·8
CD931	EPI_ISL_722103	No	Francisco Morato	SP	2020-04-29	52672	662·221	26580	24364	82·0
CD932	EPI_ISL_721993	No	Sao Paulo	SP	2020-04-29	122577	1493·13	28898	28861	95·4
CD934	EPI_ISL_721994	No	Sao Paulo	SP	2020-04-29	98723	1214·15	29068	28857	95·4
CD935	EPI_ISL_722116	No	Sao Paulo	SP	2020-04-29	51982	655·416	27557	25811	86·4
CD937	EPI_ISL_722041	No	Taboao da Serra	SP	2020-04-29	59889	738·907	29451	29243	97·9
CD939	EPI_ISL_721998	No	Sao Paulo	SP	2020-04-29	46828	578·545	29266	28995	96·2
CD94	EPI_ISL_476205	No	Sao Paulo	SP	2020-03-19	107126	1355·36	28832	27759	94·3
CD940	EPI_ISL_722094	No	Sao Paulo	SP	2020-04-29	25961	326·325	25629	23820	76·7
CD941	EPI_ISL_722092	No	Sao Paulo	SP	2020-05-01	30014	383·3	25513	21950	75·3
CD942	EPI_ISL_722086	No	Sao Paulo	SP	2020-04-30	54293	680·262	28846	28414	94·4
CD943	EPI_ISL_722107	No	Sao Paulo	SP	2020-04-29	50676	635·477	26943	25040	83·6
CD945	EPI_ISL_722062	No	Sao Paulo	SP	2020-04-30	78566	974·789	28325	27208	91·1
CD952	EPI_ISL_722040	No	Sao Paulo	SP	2020-05-01	45297	561·732	29460	29457	97·9
CD954	EPI_ISL_722117	No	Sao Paulo	SP	2020-05-01	47146	585·119	28070	26173	86·6
CD957	EPI_ISL_722098	No	Sao Paulo	SP	2020-03-05	53716	691·788	26088	23803	79·9
CD96	EPI_ISL_476207	No	Sao Paulo	SP	2020-03-21	121343	1517·52	27674	27049	88·3
CD965	EPI_ISL_722022	No	Sao Paulo	SP	2020-04-30	70494	872·811	29429	29032	97·1
CD97	EPI_ISL_476208	No	Diadema	SP	2020-03-23	123574	1522·68	27458	26318	86·2
CD971	EPI_ISL_722120	No	Sao Paulo	SP	2020-05-05	94548	1186·08	27780	26518	87·2
CD972	EPI_ISL_722095	No	Aruja	SP	2020-05-05	51812	671·907	24563	23071	76·0
CD973	EPI_ISL_722097	No	Sao Paulo	SP	2020-05-05	50456	634·107	25404	23871	78·7
CD975	EPI_ISL_722111	No	Sao Paulo	SP	2020-05-05	70510	882·501	27093	25595	83·4
CD98	EPI_ISL_476467	No	Sao Paulo	SP	2020-03-23	92861	1154·77	29144	28115	95·2

CD981	EPI_ISL_722093	No	Sao Paulo	SP	2020-05-07	46698	606·117	24757	22924	75·8
CD986	EPI_ISL_722096	No	Sao Paulo	SP	2020-05-08	81399	950·075	25435	22492	76·1
CD99	EPI_ISL_476468	No	Sao Paulo	SP	2020-03-24	18420	227·765	28908	27906	92·5
CD990	EPI_ISL_722077	No	Sao Paulo	SP	2020-05-04	117215	1457·34	29027	28143	92·6
CD991	EPI_ISL_722105	No	Sao Paulo	SP	2020-05-04	75776	950·611	27264	25440	82·8
CD992	EPI_ISL_722121	No	Sao Paulo	SP	2020-05-04	70923	889·795	27524	26012	87·3
CD995	EPI_ISL_722113	No	Franco da Rocha	SP	2020-05-04	50108	630·562	27389	25525	83·5
CD996	EPI_ISL_4463733	Yes	Osasco	SP	2020-05-05	47399	598·43	25368	22970	75·1
CD997	EPI_ISL_1172017	No	Sao Paulo	SP	2020-05-05	76418	943·165	27850	26436	88·8

Table S4. Statistical support for 16 hospital-associated transmission clusters from HC complex.

	Dataset B				Dataset C				
Cluster	alrt	FB	alrt_2	FB_2	alrt	FB	MCC	MCC DTA MJ	MCC DTA BSSVS
A	78·4	100	84·8	100	86·5	98	0·9043	0·9015	0·9002
B	78·2	100	85·9	100	75·7	100	0·9899	0·9922	0·9921
C	90·2	99	88·2	100	89·1	99	1	1	1
D	76·4	100	85·5	100	93·7	100	1	1	1
E	75·9	100	78·6	100	79·1	99	0·9978	1	1
F	91·5	100	90·5	100	93	99	1	1	1
G	91·9	99	92·3	100	78·2	99	0·9906	0·9884	0·9892
H	93·1	100	92·4	100	83	99	0·9996	1	1
I	95	100	94·6	100	92·1	99	1	1	1
J	85·5	100	85·4	100	87·1	100	0·9996	1	1
K	92·4	100	92·2	100	85·6	100	0·9991	0·9988	0·9988
L	84·9	100	90·9	100	92·6	100	0·9994	0·9997	0·9992
M	92·2	100	90·4	99	93·4	100	1	1	1
N	85·7	99	88·1	99	87·1	98	0·9962	0·9969	0·9966
O	92	100	91·8	100	93·4	99	0·9999	1	1
P	77·2	100	76	100	76·2	99	1	1	1

Table S5. Summary of epidemiological and genetic characteristics of 16 hospital-associated transmission clusters from HC complex

Cluster	Size	Institute	Pairwise divergence (SNP)				Collection date			Pairwise geo distance	Epidemiological link			Patient classification			
			Main	Mean	Median	Max	Min	Oldest	Youngest		Median (m)	Strong	Possible	Unclear	Community	>14	HW
A	12	Institute C	2·39	2	6	0	21/03/2020	30/04/2020	40	29331	6	5	1	1	4	7	0
B	10	Institute A	0·40	0	2	0	26/03/2020	01/05/2020	36	20702	3	6	1	2	0	7	1
C	7	Institute C	1·10	1	3	0	16/04/2020	13/05/2020	27	18574	4	2	1	0	2	4	1
D	7	Institute A	0·57	1	2	0	17/03/2020	26/03/2020	9	4574	4	2	1	0	0	7	0
E	5	Mixed	0·67	1	2	0	14/04/2020	22/04/2020	8	10330	0	2	3	2	0	3	0
F	6	Institute A	1·60	2	3	0	14/04/2020	29/04/2020	15	34590	2	3	1	3	0	1	2
G	4	Institute B	0·00	0	0	0	04/05/2020	11/05/2020	7	25264	4	0	0	0	0	4	0
H	4	Mixed	0·00	0	0	0	18/04/2020	29/04/2020	11	36230	2	2	0	1	0	3	0
I	3	Institute B	0·67	1	1	0	22/04/2020	24/04/2020	2	12541	0	3	0	1	0	2	0
J	2	Institute B	1·00	1	1	1	22/04/2020	22/04/2020	0	8410	0	0	2	0	0	2	0
K	2	Institute C	0·00	0	0	0	19/03/2020	23/03/2020	4	14290	0	2	0	0	0	2	0
L	2	Institute A	0·00	0	0	0	22/03/2020	24/03/2020	2	15633	0	2	0	0	0	2	0
M	2	Mixed	1·00	1	1	1	14/04/2020	17/04/2020	3	6899	0	0	2	2	0	0	0
N	2	Institute B	1·00	1	1	1	23/03/2020	23/03/2020	0	2750	2	0	0	0	0	2	0

O	2	Institute B	0·00	0	0	0	27/03/2020	03/04/2020	7	603	2	0	0	0	0	2	0
P	2	Institute B	0·00	0	0	0	03/04/2020	06/04/2020	3	12317	2	0	0	0	0	2	0

Table S6. Defining mutations of 16 hospital-associated transmission clusters from HC complex.

<b>Cluster</b>	<b>Nucleotide</b>	<b>AA</b>
A	C9733T	-
B	G28681T	N:E136D
C	C1912T	-
	C9479T	ORF1a:G3072C
	C14362T	-
D	G27240T	ORF:E13D
E	A1777G	-
F	G19086T	ORF1b:V1467I
	C24096T	S:A845V
G	T9093C	ORF1a:V2943A
H	C3293T	ORF1a:P1010S
I	G17866A	ORF1b:V1467I
	G29422T	-
J	C1884T	ORF1a:A540V
K	G18589T	ORF1b:V1708F
L	C26456T	E:P71L
M	C15738T	-
	C23481T	S:S640F
N	C3874T	-
O	C7869T	ORF1a:S2535L
P	C24023T	-

Table S7. Summary of epidemiological and genetic characteristics of 16 hospital-associated transmission clusters from HC complex per institute.

Institute	Clusters	Cluster size		Duration (days)		Pairwise geographical distance (m)	Epidemiological link (%)		
		n	Mean	Median	Mean		Strong	Possible	Unclear
Institute B	6·00	2·50	2·00	2·83	1·50	10363·40	37·04	40·74	22·22
Institute A	4·00	6·25	6·50	15·50	12·00	18167·63	36·00	40·00	16·00
Institute C	3·00	7·00	7·00	23·67	27·00	18574·48	52·17	43·48	4·35

Table S8. Logistic Regression Models for prediction of outcomes clustered vs non-clustered sequences of 234 HC SARS-CoV-2 positive individuals.

<b>Logistic Model Parameters</b>	<b>Level</b>	<b>aOdds Ratio</b>	<b>p-value</b>
Model 1:			
Variables: Institute + HCW/patient			
+ Age + Sex			
Base level: Institute B; Patient	(Intercept)	0·17	<b>0·002</b>
	Institute A	3·48	<b>0·00074</b>
	Institute C	4·17	<b>0·0002</b>
	HCW	1·63	0·2111
Model 2:			
Variable: HW/PT per Institute + Age +			
Sex			
Base level: Patient.Institute B	(Intercept)	0·11	<b>0·0034</b>
	HCW.Institute A	7·95	<b>0·0033</b>
	HCW.Institute B	2·36	<b>0·01676</b>
	HCW.Institute C	7·49	<b>0·0043</b>
	Patient.Institute A	4·45	<b>0·0266</b>
	Patient.Institute C	7·43	<b>0·00498</b>
Model 3:			
Variables: Institute + Occupation			
+ Age + Sex			
Base level: Institute B; Patient	(Intercept)	0·0757	<b>0·0002</b>

	Institute A	4·4953	<b>0·0002</b>
	Institute C	5·1756	<b>0·0001</b>
		0·0000	0·9892
	Administration	2·8086	0·0936
	Doctor	1·2141	0·7186
	Medical Resident	6·7498	<b>0·0056</b>
	Nurse	3·7721	0·1157
	Nurse Technician	1·3628	0·5404
	Other	1·1490	0·8479
Model 4:			
Variable: Occupations per institute +			
Age + Sex			
Base level: Patient.Institute B	(Intercept)	0·0468	<b>0·0002</b>
	Administration.Institute A	0·0000	0·9931
	Administration.Institute B	5·7929	<b>0·0235</b>
	Administration.Institute C	6·5012	0·2213
	Doctor.Institute A	22·4661	<b>0·0012</b>
	Doctor.Institute B	0·7448	0·8067
	Doctor.Institute C	3·8426	0·1848
	Medical Resident.Institute A	317509584	0·9935
	Medical Resident.Institute B	10·1226	<b>0·0065</b>

	Medical Resident.Institute C	29·6604	<b>0·0185</b>
	Nurse.Institute A	10·4206	0·1323
	Nurse.Institute B	3·1991	0·3898
	Nurse.Institute C	107332973	0·9894
	Nurse Technician.Institute A	7·7753	<b>0·0325</b>
	Nurse Technician.Institute B	2·0146	0·3440
	Nurse Technician.Institute C	7·6647	<b>0·0318</b>
	Other.Institute A	6·4456	0·1067
	Other.Institute B	0·9923	0·9949
	Other.Institute C	10470286	0·9939
	Patient.Institute A	4·6535	<b>0·0241</b>
	Patient.Institute C	9·1490	<b>0·0025</b>

Table S9. Compartmentalization analysis results for 73 clustered sequences from HC complex according to different traits.

Dataset	n	Trait	AI*	Bootstraps
All clustered sequences	73	Institute	0·4517	1000
		HCW/patient	0·7699	993
		Occupation	0·7499	832
		Institute B vs Others	0·5583	998
		Institute A vs Others	0·3675	1000
		Institute C vs Others	0·439	999
Institute A clustered	23	HCW/patient	0·36435	999
		Occupation	0·4203	1000
Institute B clustered	27	HCW/patient	0·9275	433
		Occupation	0·8634	808
Institute C clustered	23	HCW/patient	0·862	540
		Occupation	0·92875	

\*Simmond's Association Index

Table S10. Marjov Jumps counts, BSSVS rates and Bayes factors for all Location trait transitions.

<b>Origin</b>	<b>Destination</b>	<b>Counts</b>	<b>Counts 95% BCI</b>	<b>Rates</b>	<b>Rates 95% BCI</b>	<b>Bayes Factor</b>
Other	SP	30·465	27, 34	2·231	0·8533, 3·7905	<b>29353·794</b>
SP	Institute A	29·107	23, 34	1·333	0·4798, 2·3163	<b>29353·794</b>
SP	Institute B	104·918	94, 113	4·359	1·7017, 7·2442	<b>29353·794</b>
SP	Institute C	20·761	17, 24	1·003	0·3523, 1·7782	<b>29353·794</b>
Institute A	Institute B	4·405	2, 8	0·765	0·02241, 7·1799	<b>113·699</b>
Institute C	Institute B	4·472	1, 7	0·828	0·0824, 1·7819	<b>84·371</b>
Institute B	SP	9·763	1, 19	1·381	0·1162, 2·9797	<b>26·817</b>
Institute B	Institute A	5·410	0, 10	0·856	2·8247E-3, 2·1398	7·112
Other	Institute C	2·150	1, 4	0·618	2·0054E-4, 2·3907	4·249
Institute C	SP	2·163	0, 5	0·844	1·2088E-3, 2·6343	3·248
Institute C	Other	0·722	0, 2	0·637	8·6213E-4, 2·2403	2·946
Institute B	Institute C	2·200	0, 5	0·872	9·9104E-4, 2·8171	1·677
Institute C	Institute A	1·034	0, 7	0·984	3·5649E-5, 3·2025	0·858
Institute B	Other	0·167	0, 1	1·015	9·4685E-4, 3·2935	0·551
Institute A	SP	0·766	0, 3	1·079	5·6372E-4, 3·3461	0·481
SP	Other	0·703	0, 2	1·072	1·8106E-4, 3·3926	0·343
Institute A	Institute C	0·311	0, 1	1·156	1·0516E-4, 3·5092	0·280
Institute A	Other	0·058	0, 1	1·123	4·0284E-4, 3·4349	0·270
Other	Institute A	0·423	0, 1	1·169	6·4777E-6, 3·483	0·172
Other	Institute B	0·821	0, 2	1·159	5·5896E-4, 11·3425	0·156

## Bibliography

1. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hossmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis.* 2020 Nov;20(11):1263–72.
2. Hamilton WL, Tonkin-Hill G, Smith ER, Aggarwal D, Houldcroft CJ, Warne B, et al. Genomic epidemiology of COVID-19 in care homes in the east of England. *eLife.* 2021 Mar 2;10.
3. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc.* 2017 Jun;12(6):1261–76.
4. Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science.* 2020 Sep 4;369(6508):1255–60.
5. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science.* 2021 May 21;372(6544):815–21.
6. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* 2017 Mar 30;22(13):30494.
7. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges.* 2017 Jan;1(1):33–46.
8. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. Gisaid's role in pandemic response. *China CDC Wkly.* 2021 Dec 3;3(49):1049–51.
9. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013 Apr;30(4):772–80.
10. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol.* 2020 May 1;37(5):1530–4.
11. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017 Jun;14(6):587–9.
12. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016 Jan;2(1):vew007.
13. Rambaut A, Holmes EC, Hill V, O'Toole A, McCrone J, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *BioRxiv.* 2020 Apr 19;

14. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015 May 26;1(1):vev003.
15. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 2018 Jan;4(1):vey016.
16. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol.* 2012 Jan;61(1):170–3.
17. Hong SL, Lemey P, Suchard MA, Baele G. Bayesian phylogeographic analysis incorporating predictors and individual travel histories in BEAST. *Curr Protoc.* 2021 Apr;1(4):e98.
18. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst Biol.* 2018 Sep 1;67(5):901–4.
19. O'Brien JD, Minin VN, Suchard MA. Learning to count: robust estimates for labeled distances between molecular sequences. *Mol Biol Evol.* 2009 Apr;26(4):801–14.
20. Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, et al. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol.* 2011 Aug 9;21(15):1251–8.
21. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 2006 May;4(5):e88.
22. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol.* 2021 Jul;19(7):409–24.
23. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005 Mar 1;21(5):676–9.