Integration of sequencing and epidemiological data for surveillance of SARS-CoV-2 infections in a tertiary-care hospital

Manja Czech-Sioli[1*], Thomas Günther[2], Alexis Robitaille[2], Hannes Roggenkamp[1], Henning Büttner[1], Daniela Indenbirken[2], Martin Christner[1], Mark Lütgehetmann[1], Johannes Knobloch[1], Martin Aepfelbacher[1], Adam Grundhoff[2*], Nicole Fischer[1*]

[1] Institute for Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg Eppendorf, Hamburg, Germany

[2] Virus Genomics Unit, Leibniz Institute for Experimental Virology, Hamburg, Germany

*authors contributed equally

Corresponding author: Nicole Fischer,
Institute for Medical Microbiology, Virology and Hygiene
University Medical Center Hamburg-Eppendorf
Martinistrasse 52
20246 Hamburg, Germany
Email: nfischer@uke.de

Running Title: SARS CoV-2 nosocomial transmission during the 2nd wave.

Abstract

Background: The ongoing COVID-19 pandemic significantly burdens hospitals and other healthcare facilities. Therefore, understanding the entry and transmission of SARS-CoV-2 is critical for effective prevention and preparedness measures. We performed surveillance and analysis of testing and transmission of SARS-CoV-2 infections in a tertiary-care hospital in Germany during the second and third pandemic waves in fall/winter 2020.

Methods: Between calendar weeks 41/2020 and 1/2021 40% of all positive patient and staff samples (284 total) were subjected to full-length viral genome sequencing. Clusters were defined based on similar genotypes indicating common sources of infection. We integrated phylogenetic, spatial, and temporal metadata to detect nosocomial infections and outbreaks, uncover transmission chains, and evaluate containment measures' effectiveness.

Results: Epidemiologic data and contact tracing readily recognize most healthcare-associated patient infections. However, sequencing data reveal that temporally preceding index cases and transmission routes can be missed using epidemiologic methods, resulting in delayed interventions and serially linked outbreaks being counted as independent events. While hospital-associated transmissions were significantly elevated at a moderate rate of community transmission during the second wave, systematic testing and high vaccination rates among staff have led to a substantial decrease in healthcare-associated infections at the end of the second/beginning of the third wave despite high community transmissions.

Conclusions: While epidemiologic analysis is critical for immediate containment of healthcare-associated SARS-CoV-2 outbreaks, integration of genomic surveillance revealed weaknesses in identifying staff contacts. Our study underscores the importance of high testing frequency and genomic surveillance to detect, contain and prevent SARS-CoV-2-associated infections in healthcare settings.

Keywords: SARS-CoV-2 whole genome sequencing; integration of spatial and temporal information; hospital surveillance; molecular epidemiology, containment measurements

Introduction

Despite available testing strategies and vaccination, SARS-CoV-2 remains a major challenge to health systems in most countries, especially now with the omicron variant. [1, 2]. Genomic surveillance of SARS-CoV-2 can provide information on lineages circulating in the human population, possible sources of the virus from other geographic areas and guide the success of containment efforts [3]. In addition, such information on the source of infection and genomic phylogeny is valuable for linkage-of-infection studies in hospitals and other health care facilities as reported from first wave pandemic experience [3-6]. Integrating epidemiologic data, temporal correlations, and genomic information can provide far more accurate information on hospital-acquired infections. This is important to detect and prevent potential threats from healthcare-associated infection chains and thus maximize patient protection.

With the first SARS-CoV-2 case in Hamburg [4], we established genome sequencing to systematically investigate SARS-CoV-2 infections in the University Medical Center Hamburg-Eppendorf (UKE), a tertiary-care hospital in Northern Germany.

We aimed to identify transmission chains and potential vulnerabilities to implement targeted infection control measures by integrating genomic and classical epidemiological data. We launched this surveillance strategy at the start of the second wave. From October 2020 to January 2021, we sequenced 40% of positive cases, staff, and patients. We compare epidemiological and genomic approaches individually with an integrated approach of genomic, temporal, and spatial resolutions to identify healthcare-associated transmissions. In addition, we highlight the benefit of regular staff testing and vaccination on the frequency of healthcare-associated transmissions.

Methods

Study design

We conducted a surveillance study of SARS-CoV-2 infections at the UKE during the 2nd COVID-19 wave in Germany (October 2020 to January 2021). The UKE is one of northern

Germany's largest tertiary care hospitals with approximately 1500 beds with 88.152 in-patients, including 434 COVID-19 patients, in 2020. Patients independent of the reason for their hospital admission were screened for SARS-CoV-2 infection by nasopharyngeal swab and RT-qPCR at the time of admission and if developing symptoms during their hospital stay. Patients were preemptively isolated and admitted to a COVID ward after a PCR-positive result. Regardless of their location and duties, employees were tested if symptomatic, and employees with contact to high-risk patients were preemptively screened for SARS-CoV-2 weekly with a gargle test and RT-qPCR. RT-qPCR for SARS-CoV-2 and viral load were determined in a certified diagnostic laboratory at the UKE [7-9]. All general prevention measures and interventions for SARS-CoV-2 infections at the UKE during the time of the study are listed in Supplementary Table S1.Only samples with Ct values <32 were included in sequencing. The study was approved by the local ethics committee (PV7306).

Categorization of infection sources

Categorization of SARS-CoV-2 infection sources in patients was adapted from the COVID-19 surveillance definitions, European Centre for Disease Prevention and Control (ECDC) [10], see Supplementary Table S2, and based on the time until the first positive SARS-CoV-2 test after hospital admission: Community-associated (CA): 0-2 days; Indeterminate association (IA): 3-7 days; Probable healthcare-associated (pHA): 8-14 days; Definite healthcare-associated (dHA): >14 days.

SARS-CoV-2 amplicon sequencing, bioinformatics, statistics, and visualization

Whole genome amplicon sequencing was performed as previously published [11, 12]. Detailed information on bioinformatics analyses, statistics and visualization is provided in supplementary material.

Cluster assignments

The assignment of phylogenetically related samples to clusters was based on the presence ≥2 identical samples plus samples with ≤3 mutations to the cluster defining sequence.

Contact data

Reported contacts (RC) were determined by interviewing persons for their close contacts 48hrs before testing positive or being symptomatic. Possible contact (PC) on non-COVID wards have been tested positive within a maximum distance of 14 days to each other and resided at the same ward within 30 days prior to being tested positive. For PC between patients and staff on COVID wards patients and staff resided at the same ward within 2-14 days prior to staff being tested positive.

Public data sources

7-day incidence rates of SARS-CoV-2 infections and vaccination rates of the population in Hamburg were obtained from the Hamburg ministry of health (https://www.hamburg.de/corona-zahlen/). Numbers on SARS-CoV-2 PCR tests and positive rates were extracted from data of the Robert-Koch-Institute (https://www.rki.de).The classifications used to define community's transmission rate follow the definition of the CDC and are shown in Supplementary Table S3. ENA accession numbers of all amplicon-based sequences are provided in Supplementary Table S4. Abbreviation used in the manuscript are shown in Table supplementary Table S6.

Results

Cohort description and sampling strategy

With the beginning of the second wave of the SARS-CoV-2 pandemic in Germany in fall 2020, the numbers of COVID-19 patients and SARS-CoV-2 positively tested staff increased accordingly (Figure 1A). To rapidly detect and contain hospital-acquired infections, we

initiated systematic hospital surveillance by sequencing in week 41 of 2020, with the local rate of community transmission being moderate (Figure 1A). All available positive patient and staff specimens were included to identify index cases and transmissions. However, with the increasing incidence in the community after week 45 (high rate of community transmission), we collected random samples to monitor containment of prior outbreaks and detect new clusters of infection and cryptic transmission (Figure 1B). We successfully genotyped in weeks 41 - 45, 37% - 100% (n=118) and weeks 46 - 51, 15% - 80% (n=156) of all cases. In addition, based on epidemiologic studies, 10 cases of suspected clusters were sequenced weeks 52-1.

Clustering of samples by sequence identity

Phylogenetic analysis of all 284 SARS-CoV-2 genomes shows that several samples display identical sequences while multiple samples descending from each other are only rarely observed (Figure 2A). We based our cluster assignment on the presence of at least two samples with identical genomes similar to previous studies [3, 6, 13]. However, different from these studies, which only allowed a distance of ≤ 1 mutation between samples to define phylogenetic clusters, we included samples with ≤ 3 mutations to the cluster defining sequence to detect possible origins of outbreaks or failed containments. We identified 24 clusters with 2 to 43 samples (Figures 2A-B). Some clusters display a high number of samples with identical sequences or low divergence in a short period most likely indicating a common source of infection for each cluster (clusters 1, 3, 4, 5, and 6). Other clusters, e.g., cluster 2, demonstrate higher divergence and spread over a more extended period.

Epidemiological data acquisition

Rapid accumulation of cases with high sequence similarity can indicate a common infection source but does not necessarily prove healthcare-associated transmission. Following the ECDC definitions [10] (supplementary Table S2), patient samples were categorized in community-associated (CA), indeterminate association (IA), probable healthcare-associated

(pHA), and definite HA (dHA) COVID-19 infection sources (Figure 3A). In general, clustered samples show a significantly higher proportion of HA infections (pHA and dHA) than samples outside of clusters (Figure 3B). In particular, clusters 1, 4 - 8, 13, 16, and 18 contain a high proportion of HA infections (Figure 3C). Almost all clusters also contain employees, with no significant difference in the proportions of patients and staff in clustered vs. not clustered samples (Figure 3D).

Integration of spatial and sequencing data

To identify index cases, transmissions, and the role of staff in potential nosocomial infections, we included contact data obtained by interviewing positive tested individuals (reported contacts, RC). Also, we identified possible contacts (PCs) by comparing the physical locations on non-COVID wards where individuals were residing before testing positive. Figure 4A shows the contact network obtained by integrating spatial and genomic information. Contact tracing connects samples located in the same cluster defined by sequence similarity in many cases. However, samples without sequence similarity are also connected by RC and PC, while many samples with sequence similarity or identity are not.

Epidemiologic evaluations are insufficient in defining clusters

To better understand transmissions, we analyzed all clusters identified by sequencing individually. Figure 4B shows a combined representation of all data for cluster 1, the largest cluster including 43 cases, occurring between weeks 41 and 45. 28 of 43 cases share an identical SARS-CoV-2 sequence suggesting a common infection source. Cluster 1, based on contact data, is reported as three separate epidemiologic clusters (Figure 4B). However, integration of genomics reveals identical cluster-defining sequences in all three clusters and identifies additional cases with no previously known contacts, including the index case (#1) in week 41. All cases except #1 and #43 are classified as nosocomial transmissions according to ECDC criteria. However, transmission from the index to patients and staff at weeks 43 - 45 remains unclear because no RCs or PCs were reported. A targeted follow-up revealed that

cases #3, #11, #17, and #37 were mobile workers moving between wards (Figure 4B). After week 45, genomic surveillance assigned six additional samples to cluster 1. Patients #38 and #40 at weeks 47-48 are pHA and linked by PCs. In addition, they show identical sequences (+3 mutations) to #42 (staff), suggesting a common source of infection in the healthcare setting.

Similar to cluster 1, we find five additional clusters (4, 5, 6, 13, and 18) that show explicit properties of HA outbreaks based on the low divergence of sequences and temporal and spatial criteria (Supplementary Figure S1).

Alignment with regional sequence data from GISAID

Although we did not acquire regional SARS-CoV-2 sequence data during weeks 41-45, we obtained data on the local distribution of sequence variants through alignment with GISAID data. A comparison of cluster-defining/prototypic sequences with the GISAID database shows that some clusters contain highly prevalent sequences in Germany and Europe, while others are rarely found (Supplementary Figure S2). For cluster 1, we identified only 29 related sequences of the prototype sequence in GISAID at that time with 12 being sampled in Northern Germany. In contrast, we identified >1000 related sequences of the prototype sequence defining cluster 2.

Similarly to cluster 2, the sequences from clusters 7, and 8 show a high prevalence in Europe during that time (Supplementary Figure S2), making community-associated infections somewhat likely. Furthermore, the high genomic divergence spread over a long period and missing contact data argue against large outbreaks in these clusters (Supplementary Figure S3). Clusters 3, 11, 14, 15, 17, 21, 22, 24 contain only community-associated cases (Supplementary Figure S4) with clusters 15, 17 and 24 containing sequences with a high prevalence in the community.

Comparative analysis of epidemiological and combined data acquisition

Taken together we find clear indications for HA infections in 35 patients applying ECDC criteria suggesting 22 dHA, 13 pHA, 6 IA, and 146 CA infections. 11 of the 35 HA cases (~31%) were not detected as nosocomial infections by RCs and 8 cases (~23%) not detected by PCs (Supplementary Table S5). For 20 of 22 dHA and all pHA cases assignment into clusters by sequencing provided plausible infection routes (clusters 1, 2, 4, 5, 6, 7, 8, 13, 16, 18). Furthermore, we identified mobile workers in cluster 1 and 5 most likely responsible for the transmission from the index to patients and other staff members.

By integrating genomic, temporal and social information, we find that in the majority of cases, transmissions occur from patient to staff on non-COVID wards (clusters 6, 9, 10, 12, 19, 20, and 23) (Supplementary Figures S1/S5) or among staff (clusters 10, 12, 23) (Supplementary Figure S6). For staff working on COVID wards we find clear indications for transmissions from patients to staff in only two cases and a putative additional case (Cluster 3, Figure 5).

Effects of prevention measures on healthcare-associated infections

From the beginning of the second and the third pandemic wave in Germany, FFP2/KN95 masks were mandatory when managing patients regardless of the type of ward, and staff awareness of hygiene measures was high (Supplementary Table S1). However, during the time of our study two critical measures have been introduced to prevent hospital-acquired infections: Systematic testing of staff and the introduction of vaccination for healthcare workers. Starting in January 2021, weekly PCR testing was made mandatory for staff with patient contact with the mean number of PCR tests per week increasing from 2,200 (August - December 2020) to 7,000 (January - August 2021) (supplementary Figure S7A).

A comparison of the incidence rates of SARS-CoV-2 cases in staff and the general population reveals higher numbers in staff for November 2020 - February 2021 (supplementary Figure S8A). This is likely a combination of higher infection rates in healthcare workers [14], and the high number of tests (positive rate in staff 2.3%). However, the positive rate in the general population was estimated at 18%, indicating a higher rate of

undiagnosed cases (supplementary Figures S7A/B). Furthermore, with healthcare workers being preferentially vaccinated starting January 2021 vaccination rates in staff were higher than in the general population (supplementary Figure S7C/D). Strikingly, in the third wave (March – Mai 2021), when infection dynamics were comparable to the second wave, incidence rates among staff dropped below those seen in the general population (positive rates <0.2% and 10-15%, respectively). Statistical analysis shows that the relative incidence in staff was significantly lower in the third wave than in the second wave (Figure 6A). The proportion of HA infections in patients classified by ECDC criteria in the second vs. third wave shows a significant reduction of HA cases (Figure 6B, Supplementary Figure S8B).

Discussion

To our knowledge, our study has the following unique aspects: (1) It is one of the first studies to comparatively address the success of contact tracing, epidemiologic data, and genomic data analysis alone and in an integrated fashion in detecting nosocomial SARS-CoV-2 infections during the pandemic. Our data demonstrate that the combination of genomic data with epidemiological findings can significantly improve the detection and early intervention of healthcare associated transmissions. (2) By a comprehensive analysis of each cluster individually we show that contacts of mobile staff often neglected in studies [15] pose a challenge for tracing, but can be identified by integration of genomic data (cluster 1). However, we also show that sequencing data alone, without incorporating epidemiologic data, are insufficient to identify nosocomial outbreaks. Genetic similarity might also indicate common genotypes in the general population, thereby genetically related cases could have been introduced independently. Alignment with local sequences is essential to distinguish independent entries from transmissions in the hospital, which is particularly important if the rate of community transmission is high.

(3) In contrast to studies from the first phase of the pandemic, we do not find an increased risk of HCW infection on COVID wards; instead, we find HCW infections predominantly on non-COVID wards. Similar to previous studies [3, 5, 13, 15], we show that transmission

among staff and from patients to staff frequently occurred but without further transmission. However, unlike a previous study describing healthcare workers on COVID wards at high risk for nosocomial SARS-CoV-2 infection [5], we find mainly nosocomial transmissions occurring on non-COVID wards. Only cluster 3 suggests patient-to-staff transmission occurring on a COVID ward. Differences, besides different health care systems and containment regimens in the countries, are primarily the periods of study conduction. While Sikkens et al studied transmission events in the first pandemic wave with a low rate of community transmission; our studies focused on the second wave with a moderate - high rate of community transmission and thus, a higher risk of infection by personnel outside of COVID wards.

(4) Our study compares the number of healthcare-associated infections in two distinct episodes with the introduction of two critical interventions, regular weekly screening of healthcare workers with patient contact and preferential vaccination of healthcare workers. In particular, the impact of testing has only been modeled to this point [15]. During the study, several infection control measures were implemented throughout the hospital, including systematic weekly RT-qPCR testing of staff (which was not uniformly regulated in Germany) and the national rollout of vaccination for healthcare workers in January 2021. Although impossible to separate from each other, we show that these measures were highly effective and significantly reduced the incidence of SARS-CoV-2 infection among staff, despite high rates of community transmission. Until March 2021, we detect higher infection rates in staff than in the general population. This is in accordance with former studies [14], but probably also an effect of higher testing frequencies in HCW.

Furthermore, we observe a significant drop of HA infections in patients during the third wave. While we show that 10.5% of all patient cases are healthcare-associated infections during the second wave, similar to studies in the UK [1, 2, 16], this number drops to 3.6%. These effects are most likely the result of several prevention measures; prioritizing vaccinations in healthcare workers, elderly, and risk patients together with stringent testing strategies of hospital personnel. With HA infections in patients dropping already January 2021, we

hypothesize that the increase in testing of staff members, initiated in December 2020 [8], significantly contributed to disrupting transmissions chains in staff and patients alike.

When interpreting our data several limitations have to be considered. We performed the study at one single center. Regulations and measures regarding testing of patients and staff, isolation of patients, and rules for medical staff during break hours were not uniform in every federal state in Germany, and in addition, facilities themselves could introduce additional measures. Thus, the results cannot be generalized in a federal state such as Germany.

In addition to technical limitations in sequencing determined by viral load and RNA integrity, not all samples were available for sequencing (similar to previous studies [3, 13, 15]), and we did not collect local sequences during the study period. However, the alignment with local and regional sequences from GISAID excluded independent entries for most clusters (e.g., clusters 1, 3, 4, 5, 13, 18). Given the limitation that we did not sequence all samples during high community transmission from week 45 and later, we agree to have an incomplete dataset and might miss index cases and contacts.

In summary, this study shows how sequencing data improve to uncover transmission routes of healthcare-associated infections and reveals shortages of classical epidemiological investigations. We further show for the first time that systematic testing regimens, and vaccinations successfully contributed to the reduction of healthcare-associated SARS-CoV-2 infections. Our results served as a rationale for expanding systematic testing for staff regardless of their location. Thus, the results have led to improved retraining and use of personal protective equipment and break policies for staff, particular on non-COVID wards.

NOTES

Acknowledgements

We thank all members of the Institute for Medical Microbiology, Virology and Hygiene of the UKE for their support in collecting the samples, performing RNA extractions and SARS-CoV-2 RT-qPCR. We thank Daniela Indenbirken, Christina Herrde and Kerstin Reumann from the Leibniz Institute for Experimental Virology, Hamburg for excellent technical support in NGS sample preparation.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Carter B, Collins JT, Barlow-Pay F, et al. Nosocomial COVID-19 infection: examining the risk of mortality. The COPE-Nosocomial Study (COVID in Older PEople). J Hosp Infect **2020**; 106(2): 376-84.
2. Bhattacharya A, Collin SM, Stimson J, et al. Healthcare-associated COVID-19 in England: A national data linkage study. J Infect **2021**.
3. Ellingford JM, George R, McDermott JH, et al. Genomic and healthcare dynamics of nosocomial SARS-CoV-2 transmission. Elife **2021**; 10.
4. Sikkema RS, Pas SD, Nieuwenhuijse DF, 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**; 20(11): 1273-80.
5. Sikkens JJ, Buis DTP, Peters EJG, et al. Serologic Surveillance and Phylogenetic Analysis of SARS-CoV-2 Infection Among Hospital Health Care Workers. JAMA Netw Open **2021**; 4(7): e2118554.
6. Snell LB, Fisher CL, Taj U, et al. Combined epidemiological and genomic analysis of nosocomial SARS-CoV-2 infection early in the pandemic and the role of unidentified cases in transmission. Clin Microbiol Infect **2022**; 28(1): 93-100.
7. Norz D, Fischer N, Schultze A, et al. Clinical evaluation of a SARS-CoV-2 RT-PCR assay on a fully automated system for rapid on-demand testing in the hospital setting. J Clin Virol **2020**; 128: 104390.
8. Olearo F, Norz D, Hoffman A, et al. Clinical performance and accuracy of a qPCR-based SARS-CoV-2 mass-screening workflow for healthcare-worker surveillance using pooled self-sampled gargling solutions: A cross-sectional study. J Infect **2021**; 83(5): 589-93.
9. Pfefferle S, Reucher S, Norz D, Lutgehetmann M. Evaluation of a quantitative RT-PCR assay for the detection of the emerging coronavirus SARS-CoV-2 using a high throughput system. Euro Surveill **2020**; 25(9).

10. European Centre for Disease Prevention and Control (ECDC): Surveillance definitions for COVID-19. Available at: https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions.
11. Brehm TT, Pfefferle S, von Possel R, et al. SARS-CoV-2 Reinfection in a Healthcare Worker Despite the Presence of Detectable Neutralizing Antibodies. Viruses **2021**; 13(4).
12. Gunther T, Czech-Sioli M, Indenbirken D, et al. SARS-CoV-2 outbreak investigation in a German meat processing plant. EMBO Mol Med **2020**; 12(12): e13296.
13. Lumley SF, Constantinides B, Sanderson N, et al. Epidemiological data and genome sequencing reveals that nosocomial transmission of SARS-CoV-2 is underestimated and mostly mediated by a small number of highly infectious individuals. J Infect **2021**; 83(4): 473-82.
14. Nguyen LH, Drew DA, Graham MS, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. Lancet Public Health **2020**; 5(9): e475-e83.
15. Lindsey BB, Villabona-Arenas CJ, Campbell F, et al. Characterising within-hospitalSARS-CoV-2 transmission events using epidemiological and viral genomic data across two pandemic waves. Nat Commun **2022**; 13(1): 671.
16. Bhattacharya A, Collin SM, Stimson J, et al. Healthcare-associated COVID-19 in England: A national data linkage study. J Infect **2021**; 83(5): 565-72.

Figure Legends

Figure 1: A: SARS-CoV-2 cases at the hospital between week 41 in 2020 and week 1 in 2021. The patient group includes all inpatients who tested positive for SARS-CoV-2 independent of the reason for their hospital admission (COVID-19 or other). The local incidence per 100,000 is shown in grey. The classification of the rate of community transmission according to the CDC (see Supplementary Table S3) is depicted above. **B**: Percentage of SARS-CoV-2 cases genotyped by sequencing. In weeks 41-45 in 2020, all samples available were chosen for sequencing. In weeks 46-51, samples were randomly chosen for sequencing (Ct<32), and 15-80% each week were successfully genotyped. In weeks 52 (2020) – 1 (2021), specific samples suspected to be related to an infection cluster were sequenced and genotyped.

Figure 2: A: Phylogenetic tree based on divergence of all SARS-CoV-2 positive samples sequenced in the study from the reference sequence NC_045512. Clusters were defined by at least two samples with identical sequences plus additional samples with a maximum distance of 3 mutations. In total, 24 clusters can be identified by sequencing. **B**: Time-dependent visualization of SARS-CoV-2 samples that were clustered by their sequence. The color code shows each sample's distance (number of mutations) from the cluster defining sequence.

Figure 3: Suspected SARS-CoV-2 infection sources in sequenced samples. Categories were based on hospital admission dates and dates of the first positive test according to the European Centre for Disease Prevention and Control [10] (Supplementary Table S2). CA: Community-associated. IA: Indeterminate association, pHA: Probable healthcare-associated, dHA: Definite healthcare-associated. **A**: Time-dependent visualization of SARS-CoV-2 samples in clusters and not classified as a cluster (n.c.) with suspected SARS-CoV-2 infection sources. **B**: Statistical analysis of suspected infection sources in clustered patient samples and cases not classified as a cluster (n.c.). HA: Healthcare-associated infections

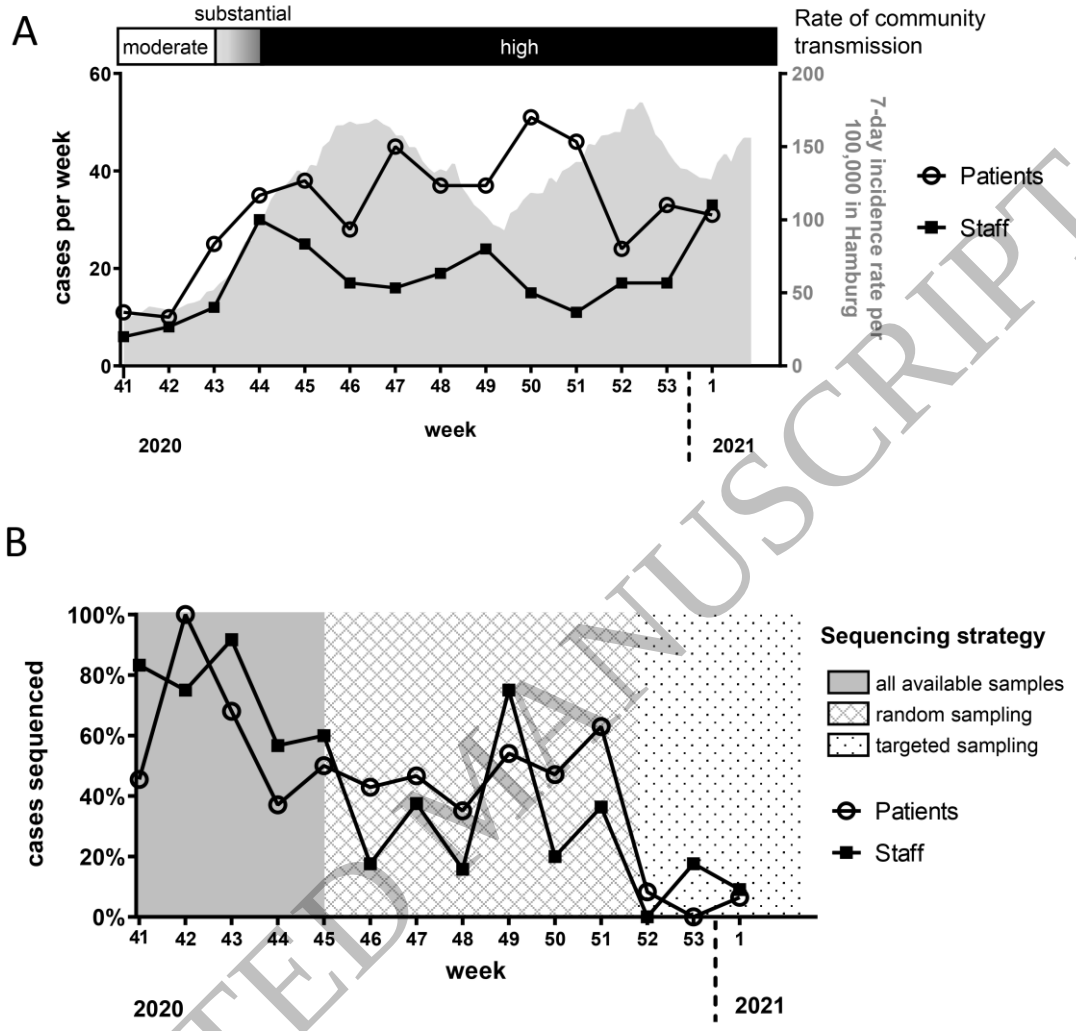
combining pHA and dHA infections. CA: Community-associated infections. Indeterminate associations (IA) were excluded from the analysis. Fisher's exact test calculates a significant difference in the proportions of HA and CA infections between the group of n.c. samples and clustered samples ($p<0.0001$). **C**: Percentages of CA, IA, pHA and dHA infections in sequencing clusters (all clusters and clusters 1 - 24) and n.c. samples. n.a.: not applicable. **D**: Statistical analysis of the proportion of staff members and patients in clustered and n.c. samples. No significant difference was determined between the two groups by Fisher's exact test ($p=0.362$).

Figure 4: A: Contact network on non-COVID-wards. PC nCoV: Possible contacts between patients and staff and among both groups on non-COVID-wards. RC nCoV: Reported contacts on non-COVID wards. Colors and shapes indicate membership of cases in sequencing clusters. n.c.: not classified as a cluster. **B**: Detailed time-dependent visualization of cluster 1, including contact data, suspected infection sources, and sequencing data. For a detailed explanation, see main text. CA: community-associated infections, IA: indeterminate associations, pHA: probable healthcare-associated infections, dHA: definite healthcare-associated infections, n.a.: not applicable.

Figure 5: A: Contact network on COVID-wards. PC CoV: Possible contacts between patients and staff on COVID-wards. **B**: Detailed time-dependent visualization of cluster 3, including contact data, suspected infection sources, and sequencing data. The staff cases #6 and #7 in cluster 3 have PCs to patient cases #2 and 4 and between each other. Transmission from patient to staff is highly likely due to the sequence identity (#7 identical, #6 +1 mutation) and the time distance. Staff case #8 is likely also connected. All patient cases are community-associated and were traced back to a shared household (nursing facility), which explains the low divergence of their sequences.

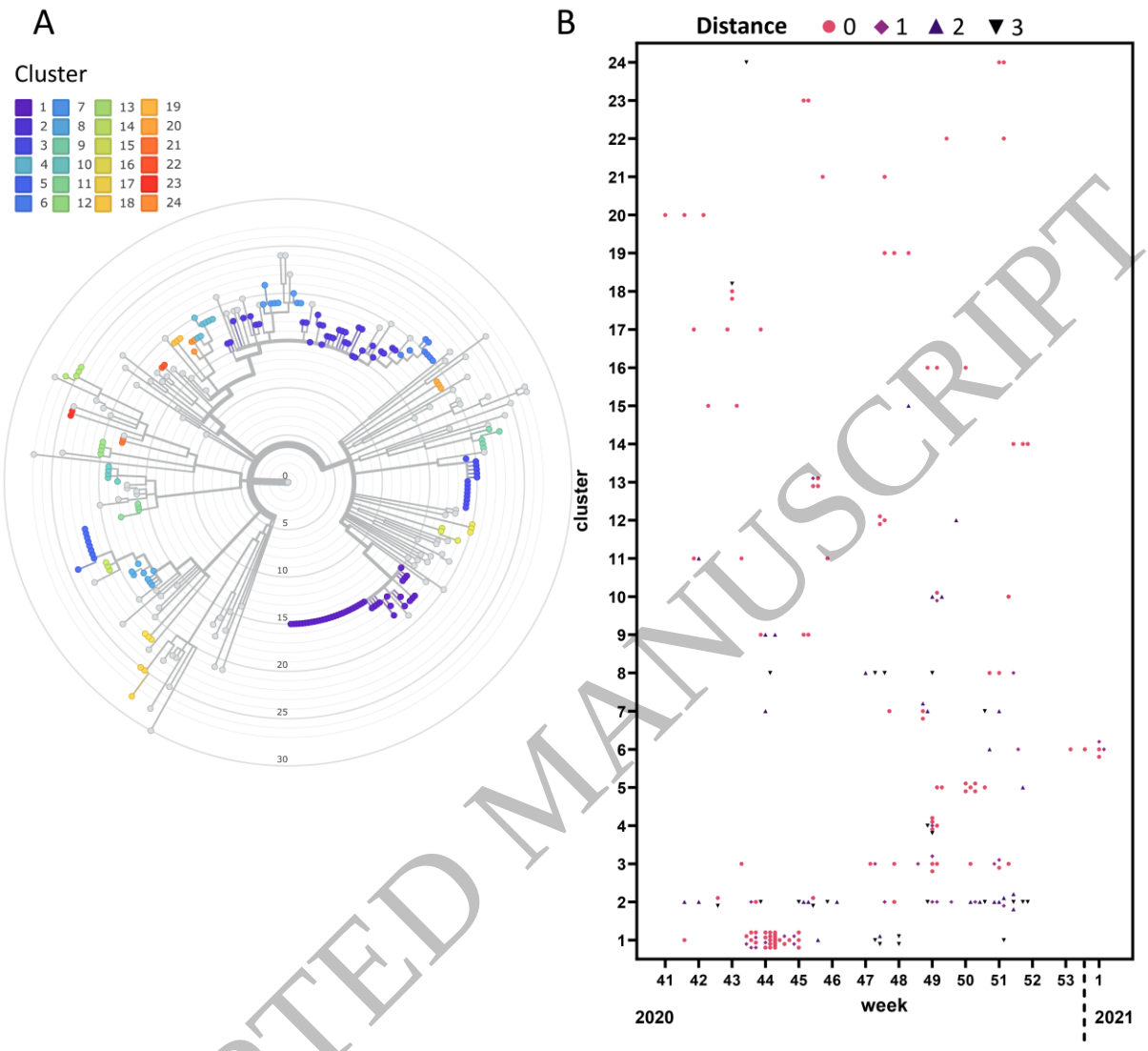
Figure 6: A: 7-day incidence rates of SARS-CoV-2 in staff relative to the incidence in Hamburg. Second and third waves are defined by a 7-day incidence rate above 50 and the low point between the two waves. Statistical comparison by Mann-Whitney-test provides significantly reduced incidences in staff for the third wave compared to the second wave (p=0.0014). **B:** Percentages of healthcare-associated SARS-CoV-2 infections in patients from September 2020 until end of August 2021. HA: Healthcare-associated infections including probable healthcare-associated (pHA) and definite healthcare-associated infections (dHA). Second and third wave definition as in A. Statistical comparison by Mann-Whitney-test provides significantly reduced HA patient proportion for the third wave compared to the second wave (p = 0.0169).

Figure 1



1

Figure 2



1

Figure 3

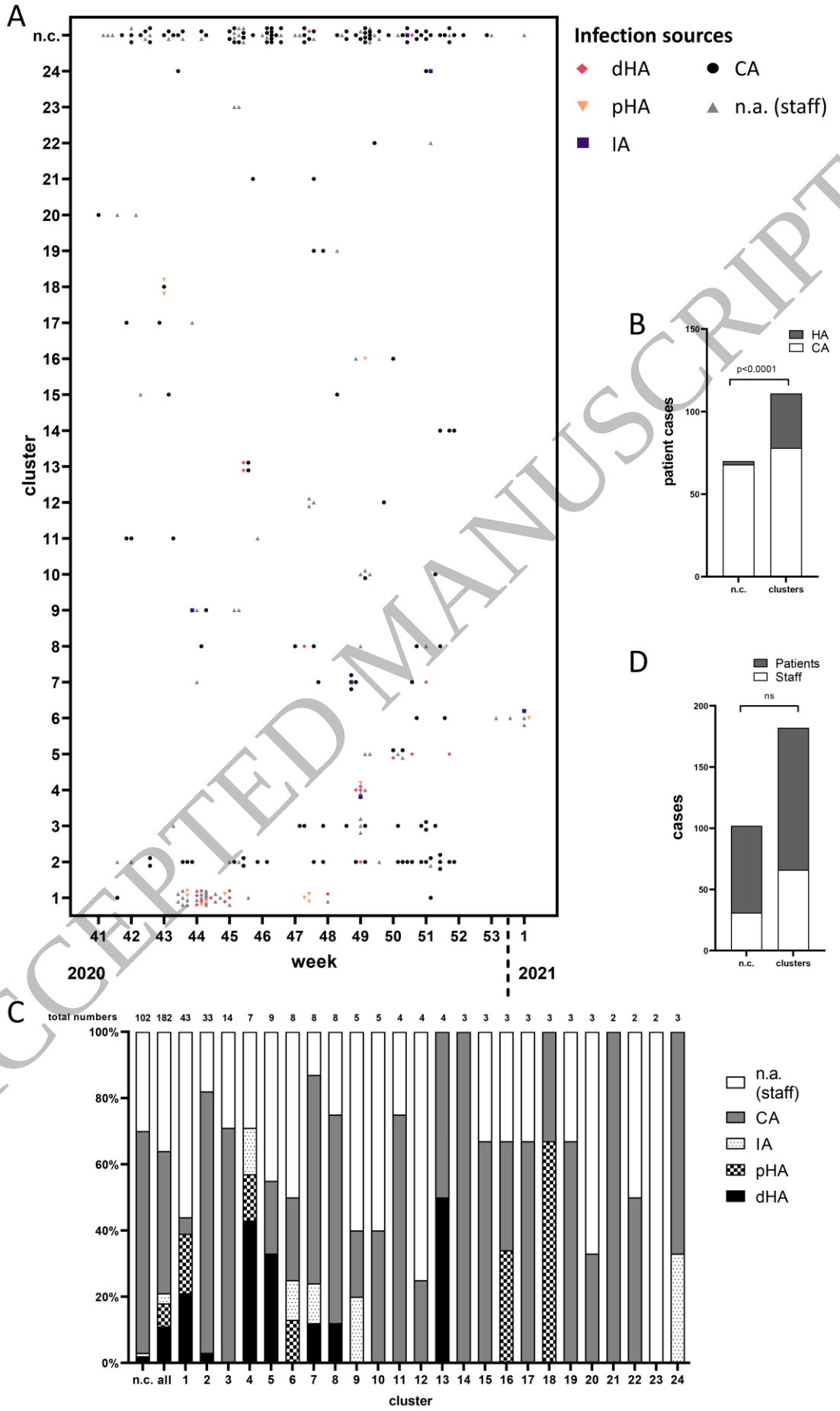
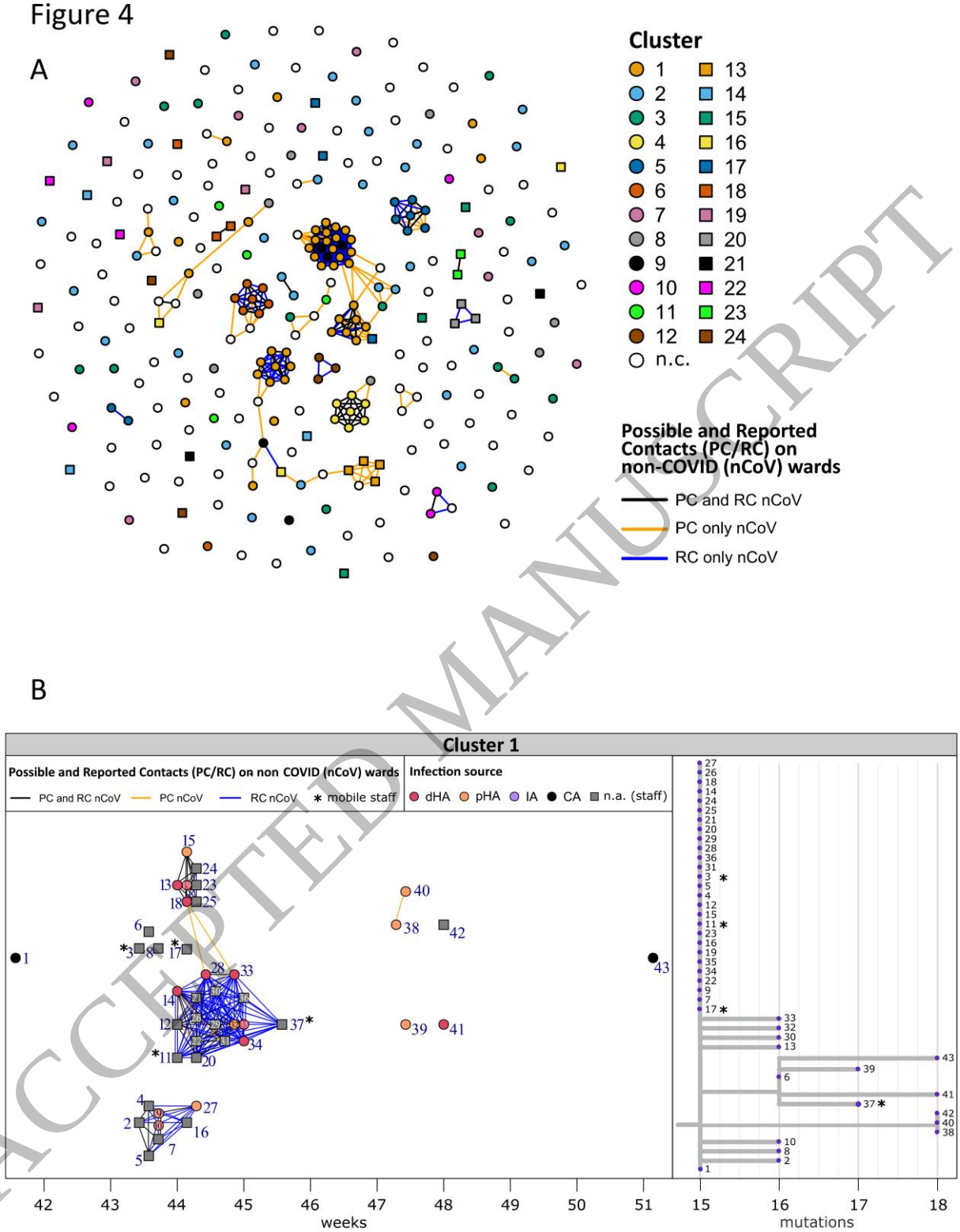
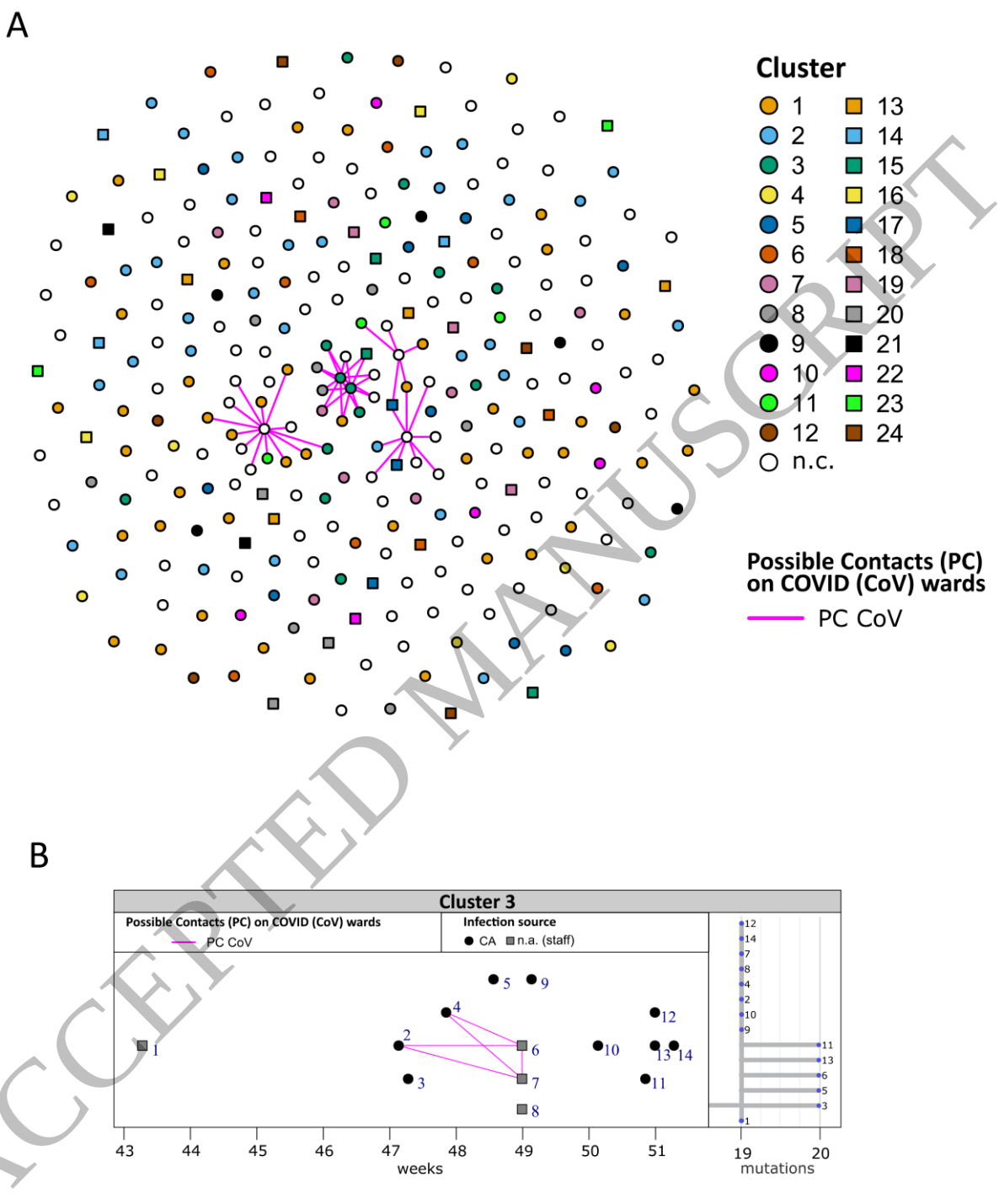


Figure 4



1

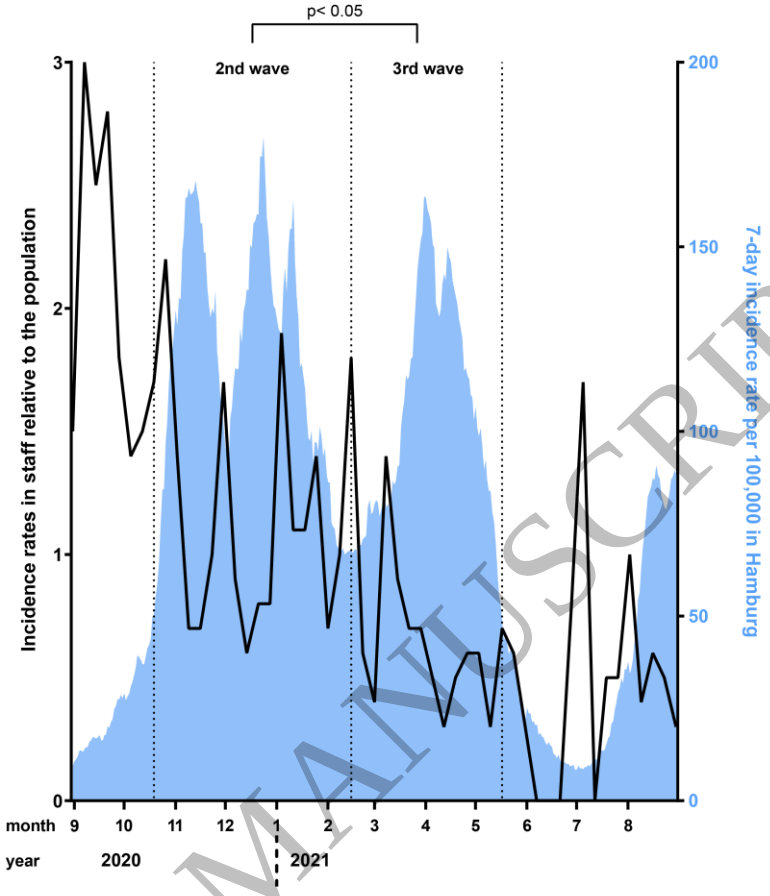
Figure 5



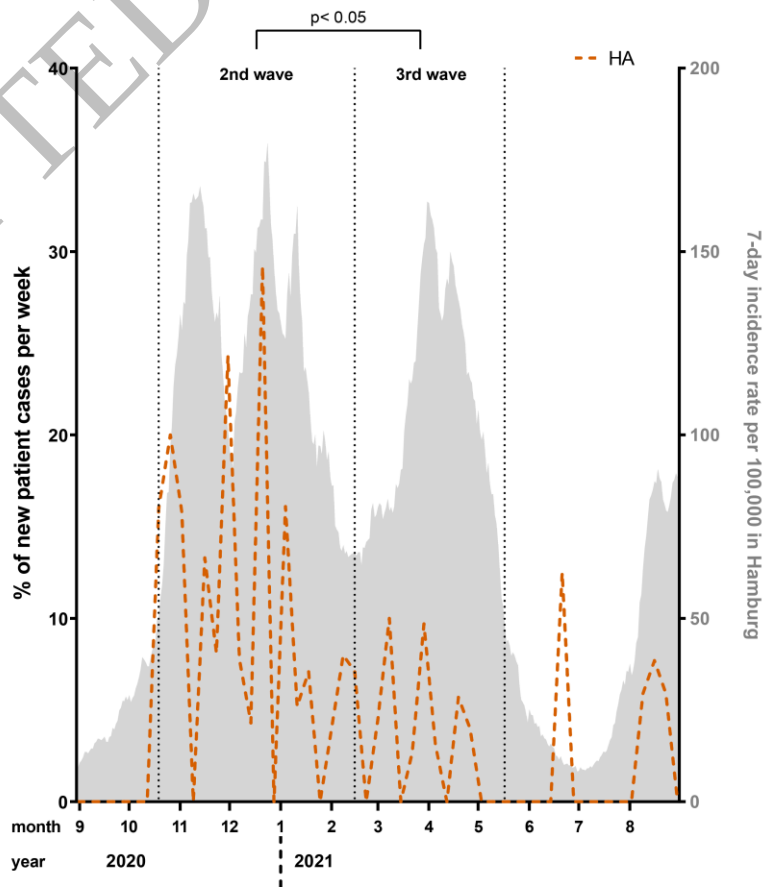
1

Figure 6

A



B



1