The Journal of Infectious Diseases

MAJOR ARTICLE

Population-based SARS-cov-2 whole genome sequencing and contact tracing during the COVID-19 pandemic in Switzerland

Nanina Anderegg [1 2 *], Tiana Schwab [1 *], Loïc Borcard [5], Catrina Mugglin [1 3 4], Bettina Keune-Dübi [3], Alban Ramette [5 **] and Lukas Fenner [1 ** §]

[1] Institute of Social and Preventive Medicine, University of Bern, Switzerland; 2 Centre for Infectious Disease Epidemiology and Research, School of Public Health, University of Cape Town, South Africa; 3 Cantonal Physician's Office, Gesundheitsamt, canton of Solothurn, Solothurn, Switzerland; 4 Department of Infectious Diseases, Inselspital, Bern University Hospital, University of Bern, Switzerland; 5 Institute for Infectious Diseases, University of Bern, Bern, Switzerland

Background: Testing and contact tracing (CT) can interrupt transmission chains SARS-CoV-2. Whole genome sequencing (WGS) can potentially strengthen these investigations and provide insights on transmission.

Methods: We included all laboratory-confirmed COVID-19 cases diagnosed between June 4 to July 26, 2021, in a Swiss canton. We defined CT clusters based on epidemiological links reported in the CT data and genomic clusters as sequences with no single nucleotide polymorphism (SNP) differences between any two pairs of sequences being compared. We assessed the agreement between CT clusters and genomic clusters.

---

* Contributed equally
** Contributed equally
§ Correspondence: Lukas Fenner, University of Bern, Bern, Switzerland, Phone: +41 31 631 35 21, E-Mail: lukas.fenner@ispm.unibe.ch
Alternative correspondence: Nanina Anderegg, University of Bern, Bern, Switzerland, nanina.anderegg@ispm.unibe.ch

DOI: 10.1093/infdis/jiad074

Results: Of 359 COVID-19 cases, 213 were sequenced. Overall, agreement between CT and genomic clusters was low (Kappa coefficient=0.13). Out of 24 CT clusters with at least two sequenced samples, 9 (37.5%) were also linked based on genomic sequencing but in four of these, WGS found additional cases in other CT clusters. Household was most often reported source of infection (101, 28.1%) and home addresses coincided well with CT clusters: In 44 out of 54 CT clusters containing at least two cases (81.5%), all cases of the cluster had the same home address. However, only a quarter of household transmission was confirmed by WGS (6 out of 26 genomic clusters, 23.1%). A sensitivity analysis using ≤1 SNP differences to define genomic clusters resulted in similar results.

Conclusions: WGS data supplemented epidemiological CT data, supported the detection of potential additional clusters missed by CT, and identified misclassified transmissions and sources of infection. Household transmission was overestimated by CT.

Key words: COVID-19, contact tracing, whole genome sequencing, transmission, molecular cluster, public health, household transmission

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly contagious respiratory virus, the causative agent of the coronavirus disease 2019 (COVID-19) pandemic, and a global public health threat. As of July 2021, there have been over 560 million confirmed SARS-CoV-2 infections globally and 6.4 million deaths have been attributed to COVID-19 (1). In Switzerland, almost 4 million COVID-19 cases have been confirmed and around 14,000 people have died from COVID-19 (2). In the absence of effective and safe vaccines until the end of 2020, the pandemic could only be fought by non-pharmaceutical interventions (e.g., ban of events and gatherings, wearing masks, social distancing), extensive testing of symptomatic and asymptomatic persons, and consistent contact tracing (CT) to isolate and quarantine cases and contacts, respectively, and to identify and interrupt transmission chains.

The advancements of high-throughput sequencing technologies and their widespread accessibility have enabled the monitoring of SARS-CoV-2 mutations and variants (3,4). By comparing SARS-CoV-2 whole genome sequences, mutations can be identified and used to provide high-resolution insights that are not captured by CT-based epidemiological investigations (5), such as transmission dynamics and viral evolution, possibly even in real-time (6–8), and can be further used to evaluate the effectiveness of implemented public health interventions (9). Genomic data and phylogenetic analysis have been used to investigate the impact of superspreading events (10), monitoring the introduction and spread of new cases and variants by returning travellers (11–13) and transmission chains in high-incidence settings (14). However, the implementation of routine genomic surveillance is logistically challenging, requiring a comprehensive sequencing infrastructure and coherent integration with the CT

system (8), and it is unknown how well routine CT using epidemiological links between cases performs compared to genomic analyses. In this study we compared population-based CT and viral WGS data from COVID-19 patients in a Swiss canton during a 7-week period just before the start of the fourth epidemic wave and the nationwide spread of the delta variant.

METHODS

Study design, patient data and samples

All laboratory-confirmed COVID-19 cases (Antigen- or PCR-positive) diagnosed between the 4th June to 26th July 2021 in the canton of Solothurn (~280,000 inhabitants) were included for CT and for analysis in this study. CT measures, including data collection and isolation/quarantine requirements, were based on the Epidemics Act law and in line with the recommendations by the Federal Office of Public Health (FOPH): Positively tested persons were routinely contacted by the CT team (Cantonal Physician's Office) for structured, in-depth telephone interviews shortly after diagnosis and during follow-up until the end of isolation on the 10th day. Collected data included sociodemographic information, place of residence, workplace or school, symptoms, potential source of infection (self-reported), activities, visited places, and people in close contact prior to the positive test result. The latter were themselves contacted by the CT team to impose a quarantine of 10 days and – if they did test positive at some later stage – for the in-depth telephone interviews and identification of further close contacts. CT was supported by a partially automated workflow to ensure that accurate and complete data on all positively tested persons and links to their contacts were entered in the electronic cantonal CT database.

Per order of the Cantonal Physician's office WGS was done for all persons tested SARS-CoV-2 PCR-positive. All persons tested positive by rapid antigen tests were recommended to provide an additional sample for PCR-testing and genome sequencing. WGS was performed at laboratories participating in the National Genomics Surveillance Project (5), according to published protocols described in detail in (5,15). Sequences were stored at the international Global Initiative on Sharing Avian Influenza Data (GISAID; https://www.gisaid.org/) (16).

Definitions

Cases were defined as persons tested SARS-CoV-2 PCR-positive by rapid-antigen or PCR test. *Contacts* of a case were defined as people in close contact with the case (less than 1.5 meters for more than 15 minutes) without adequate protection, such as wearing a face mask, in the last 48 hours prior to symptom onset or diagnosis if asymptomatic, as recommended by the FOPH (2). Contacts that tested SARS-CoV-2 positive during their quarantine time were defined as *secondary cases*. Cases not reported as a contact of another case at the time of the positive test were defined as *initial cases*.

We differentiated between *CT clusters* (and *networks*) and *genomic clusters:* CT clusters and networks were based on established links between cases and contacts through the CT data. Networks included cases and all contacts, regardless of whether a contact became a secondary case or not, while CT clusters only included initial and secondary cases. Genomic clusters were defined as SARS-CoV-2 genome sequences with no single nucleotide polymorphism (SNP) differences between any two pairs of sequences being compared. To evaluate clustering concordance between CT and WGS, we also assessed the effects of defining genomic clusters based on one to five SNP differences between any two pairs of sequences.

Analyses

SARS-CoV-2 genome sequences were analysed using Pangolin (version 3.1.2) to infer the lineage of each sequence. Multiple sequence alignments were generated with MAFFT and subjected to maximum-likelihood phylogenetic reconstruction using IQ-TREE 2 (5,17). A distance matrix was generated using the aligned reads based on the number of nucleotide differences and the resulting distance matrix was used to generate a minimum spanning tree (MST) (18). The displayed MST graph was pre-processed with igraph (version 1.2.10) and generated using ggnetworks (v0.5.10) (https://github.com/briatte/ggnetwork).

We used descriptive statistics to report baseline data of cases and contacts. We investigated the trends in the proportion of cases with available sequences, the contribution of different SARS-CoV-2 variants, and sizes of CT networks and clusters over time. We estimated the serial interval for COVID-19 using CT clusters, by extracting all pairs of initial and secondary cases and used the mean time between their positive tests as a proxy for the serial interval. Within genomic clusters, all cases were linked and there was no obvious initial case since we did not know the specific transmission chain within a genomic cluster. To estimate the serial interval of COVID-19 using genomic clusters, we therefore defined its potential sources of infection as all other cases in the same cluster who had test dates prior or equal to the case's test date. With this, we estimated a range for the serial interval, based on the average minimum and maximum time differences between test dates of a cases and their potential sources of infection.

We examined the agreement between CT and genomic data based on the total number of cases in the CT cluster, the number of sequences available and sequence identities within the CT clusters. We visually compared CT and genomic clusters and examined to what extent additional information collected during CT (meta-data including place of residence, workplace, school, and self-reported source of infection) corresponded with CT and genomic clusters. In a sensitivity analysis, we checked the change in agreement between CT and genomic data when using a more relaxed definition of a genomic cluster, allowing for one SNP difference between a pair of sequences. In addition, we quantified the agreement between CT and genomic clusters (inter-rater-reliability) for each week of the study period using the Cohen's kappa statistic (19) with a threshold of 0 to five SNP differences between two sequences defining a genomic cluster. All analyses were done using the software packages Stata (version 16.1) and R (version 4.1.3).

Data availability

The genome assembly sequences were deposited to the European Nucleotide Archive (ENA) and project and sample accession numbers are listed in **Supplementary Table 1**.

Ethics statement

The Ethics Committee of the Northwestern and Central Switzerland approved this study (EKNZ; reference no. 2021-02240).

RESULTS

Over the course of the study period (June 6th to July 29th, 2021), 359 SARS-CoV-2 cases and 460 contacts were recorded in the canton of Solothurn (**Table 1**). Most cases, 289 (80.5%) were initial cases, while 70 cases (19.5%) were secondary cases. Of all cases, 213 (59.3%) had WGS data available, with a lower proportion of sequenced cases at the beginning and end of the study period (**Table 1**, **Figure 1A**). Gender and age distributions were similar across cases and contacts and amongst sequenced and non-sequenced cases (**Table 1**). Most cases contacted reported their household as the most likely source of infection (28.1%), followed by travel (15.6%) and work (11.7%) (**Table 1**). The study period captured the spread of the delta variant with the delta variant dominating from end of July onwards (**Figure 1B)**.

CT clusters

Overall, the data from CT identified 289 different CT networks and clusters. Both CT network and CT cluster sizes did not change substantially over time, although there was a slight decrease in the weekly average network size towards the end of the study (**Figure 1C and 1D**). The mean sizes for CT networks and CT clusters were 2.83 (range 1 to 20) and 1.24 (range 1 to 5) respectively. Based on CT linkages, 112/289 (38.8%) CT networks consisted of an initial case with no reported contact (network size of 1), while 58/289 (20%) CT networks consisted of a single case and contact (network size 2, **Supplementary Figure 1**). Based on 70 links between initial and secondary cases from CT clusters, we estimated an average serial interval of 4.4 days (95% CI: 3.7-5.0). If we restricted the 70 CT links to those confirmed by genomic sequencing, the estimate of the serial interval was similar with an average of 4.8 days (95% CI: 2.9-6.6).

Genomic clusters

Alpha and delta variants clearly separated into clades on the minimum spanning tree (**Figure 2**). In total, 160 genomic clusters were obtained using no SNP differences as threshold to define clusters. Of these, 133 (83.2%) were singleton genomic clusters and 27 (16.9%) consisted of >1 cases (**Supplementary Table 2**). Based on 53 links within genomic clusters, we estimated an average minimum serial interval of 2.5 days (95% CI 1.7–3.2 days) and an average maximum

serial interval of 5.1 days (95% CI 3.9–6.3 days). Most genomic clusters occurred within a short time frame, however the largest genomic cluster (eight cases) spanned 19 days.

Comparing CT and genomic clusters

Of 289 CT clusters, 106 (36.7%) had no cases with available SARS-CoV-2 sequences available, 25 (8.7%) were partially sequenced, containing at least one case with an available SARS-CoV-2 sequence, and 158 (54.7%) were fully sequenced. However, many of the fully sequenced clusters were single-case CT clusters (**Table 2**, **Supplementary Figure 2**). Out of 24 CT clusters which were >1 in size and contained at least two sequenced samples, nine (37.5%) were also linked based on genomic sequencing ("correct" CT linkage), whereas 15 CT clusters contained different sequences ("incorrect" CT linkage). For four of the nine correct CT clusters the same sequence was found outside of the CT cluster ("incomplete" CT linkage), for the remaining five correct CT clusters the sequence was not found outside the CT cluster ("complete" CT linkage, **Table 2**, **Figure 3 and Figure 4**, **Supplementary Figure 2**). Of 159 CT clusters with only one sequence available, 107 (67%) were complete, the sequence not being identified elsewhere (**Figure 3**).

When relaxing our definition of a genomic cluster, allowing for up to one SNP difference between any two pairs of sequences, the number of genomic clusters decreased from 160 to 135. The more relaxed threshold led to slightly fewer "incorrect" CT linkages (11 instead of 15 CT clusters) and conversely slightly more "correct" CT linkages (13 instead of 9). However, it also led to a much larger proportion (>50%) of CT clusters being "incomplete" (**Supplementary Figure 3**). This trade-off was reflected in Cohen's kappa statistic, which showed that throughout the agreement between the CT and genomic clusters was low for all SNP thresholds (average of 0.13 to 0.10) with high uncertainty (**Supplementary Figure 4**).

Comparing CT and genomic clusters based on reported household, workplace, and other potential sources of infection

When comparing clusters with the additional meta-data collected during contact tracing, place of residence (both complete home address and ZIP code) showed poor agreement for genomic clustering but better agreement for CT clustering and persons with the same home address were more commonly found in CT clusters than genomic clusters (**Figure 5**, **Supplementary Figure 5, Supplementary Figure 6**): Of 54 CT clusters with at least two cases, only 5 (9.3%) did not contain cases with the same home address. The remaining 49 (90.7%) CT clusters contained at least two cases with the same home address, with 44 (81.5%) being single-home-address clusters. However, looking at the genomic data of the 44 single-home-address CT clusters, we found that 10 (66.7%) out of 15 CT clusters with >1 sequence available were misevaluated according to genomic data (the sequences within single-home-address CT cluster were different) and only 5 (33.3%) were confirmed by genomic data. Looking at genomic clusters containing at least two cases, most did not contain any cases with the same home address (15/26, 57.7%). The remaining 11 (42.3%) contained at least two cases with the same home address, but only 6

(23.1%) were single-home-address clusters and persons with the same home addresses were often present in multiple different genomic clusters (**Figure 5**). Using the ≤1 SNP threshold to define genomic clusters, 8 (25%) of 32 were single-home-address clusters. In the four largest CT clusters, 76.5% (13/17) of all cases and 92.3% (12/13) of secondary cases reported their household as the most likely source of infection. However, the most likely source of infection in genomic clusters was more varied: In only one of the three largest genomic clusters, most cases reported the same location, namely school, as the most likely source of infection (**Supplementary Figure 7**). For the remaining two, an average of 4.5 different exposures were reported.

Workplace and school showed poor agreement with both CT and genomic clustering (**Supplementary Figure 5, Supplementary Figure 8**). Out of 31 CT clusters containing at least two cases with a reported work address, most (26 cases, 83.9%) were distinct workplace clusters, in which all cases reported a different work address, and only 5 (16.1%) contained cases with the same workplace. For genomic clusters these numbers were similar, with 14/16 (87.5%) distinct-workplace clusters and only 2 (12.5%) clusters containing cases with the same work address. We observed similar results for schools (**Supplementary Figure 8**).

DISCUSSION

Comparing CT and sequencing data from COVID-19 patients shows an overall low agreement and that WGS can provide additional links between cases which CT did not necessarily capture. The potential source of infection is difficult to identify based on CT data and household transmissions are less supported in genomic clusters. Including data from cases in the entire Canton of Solothurn, spanning a 7-week period before the start of the fourth epidemic wave, provided a comprehensive picture of the epidemiologic situation in the general population of the canton of Solothurn, and extended our analysis beyond an outbreak investigation in a limited setting. By comparing CT data to population-based sequencing data, we expanded our analysis beyond self-reported contacts to more objective WGS data.

Routine CT relies on complete and correct recall of contact events by cases, the correctness of information given by the interviewed person, and the willingness of the population to cooperate with the CT team during the interviews (20–22). In our study specifically, additional CT information, such as workplace or school address, was not disclosed for many cases. The limitations of routine CT might be particularly true in the context of highly transmissible infectious disease such as SARS-CoV-2 and during a pandemic with strict measures with substantial economic and social consequences (23). We were also limited to cases within the canton of Solothurn and might have missed cases in clusters if the transmission chain included people living outside of Solothurn. This is likely in Switzerland, where it is common for people to work in a canton different to the one they live in or visit friends and relatives living in another canton.

The limitations of WGS sequencing in this study include that the analysis of sequencing information was done retrospectively, and sequencing of cases was incomplete. Some clinical specimens were no longer available, or sequencing was technically not possible due to too low viral loads and consequently many CT clusters were not fully sequenced. This resulted in a relatively small number of cases, highlighting the need for large databases to overcome the technical and logistical challenges of WGS to target a higher proportion of sequenced results and achieving larger overlapping datasets.

In addition to the limitations of CT and WGS, the introduction of a new delta SARS-CoV-2 variant with changed characteristics (24) such as increased transmissibility may have also impacted the agreement between CT and genomic data. However, the average size of case-contact networks and CT clusters did not change substantially during the study period, and we did not find a higher agreement between CT and genomic clusters in the beginning of our study, when the alpha variant dominated. Since we did not see substantial changes, we used the time between positive tests as a proxy to estimate the serial interval of COVID-19 transmissions across the study period, and while the time between positive tests based on CT data may not have been true transmission chains, the estimated serial interval based on genomic clusters was similar to the one based on CT clusters. These estimates also matched the serial interval of COVID-19 determined in current literature of 4-8 days (25).

In our data almost a third of CT clusters were incomplete, with identical sequences found across multiple CT clusters. These potentially missed links by CT would need to be further investigated to be confirmed. WGS data of COVID-19 cases in a hospital outbreak in Portugal (6) showed that transmission events were mostly driven through health care workers rather than transmissions between patients within the same room. Similar studies used WGS data to confirm or dispute presumed transmissions and investigate sources of infections in hospital outbreaks and even the local population (7,8,26,27). Multiple efforts have showed the added value of real-time or near-real-time genomic sequencing of SARS-CoV-2 in investigating community outbreaks, identifying transmission chains and clarifying probable sources of infection, albeit requiring a comprehensive sequencing infrastructure and coherent integration with the CT system (10,13,14,28,29).

There was poor agreement when matching additionally collected information such as home addresses or self-reported sources of infection with observed genomic clusters and large CT clusters were more likely to correspond with patient self-reported sources of infection than large genomic clusters. These will be biased towards households and workplaces, where people spend more time and are more likely to know of a COVID-19 case compared to public places such as restaurants or hospitals. However, households and workplaces do not necessarily reflect exposures with the highest transmission risk. A meta-analysis of SARS-CoV-2 household transmissions estimated the household secondary attack rate as 19% (30). In our data genomic sequencing showed that cases from the same households were often distributed among several genomic clusters, indicating that infections may have occurred elsewhere. This is similar to what

has been seen in investigations of tuberculosis transmissions, an infection similarly transmitted through aerosols and where household transmissions have a smaller contribution to the overall population transmission even in high TB endemic settings (31–33).

Overall, the overestimation of household transmissions by CT data alone requires further investigation as this could influence future CT procedures, which focus more on identifying multiple sources of infection rather than quarantine of household members. Despite a low to moderate agreement between CT and genomic clusters in our data, at a population level, CT is an effective public health measure (21,34,35), especially when combined with isolation and physical distancing measures (36). Effective CT does entail a high logistic burden, particularly as case numbers increase (21). Here surveillance-based sequencing can be invaluable in better understanding and explaining transmissions clusters and beyond that, offers additional potential in identifying outbreaks more rapidly and can guide efficient viral containment strategies and prospective preventative measures (6–8,28,29) when implemented in real-time.

ACKNOWLEDGEMENTS

We would like to thank all persons who contributed data to this study. We are indebted to the contact tracing team of the canton of Solothurn, Switzerland, for their diligent daily work. The authors also thank the diagnostic laboratories in Switzerland for analysing the genomes.

Authors' contributions

Conception and design: CM, LF. Contact tracing data collection: CM, BKD, LF. Genomic data collection: BKD, CM, LB, AR, LF. Statistical analysis: NA, LB, AR, TS. Genomic analysis: LB, AR. Wrote the first draft of the paper NA, TS, AR, LF. NA, TS, LB, BKD, AR, and LF revised it based on comments from all authors. All authors reviewed and approved the final version of the manuscript.

Sources of support

This work was supported by the Institute of Social and Preventive Medicine (ISPM) and the Institute of Infectious Diseases (IFIK) of the University of Bern. TS and LF are supported by the National Institute of Allergy and Infectious Diseases (NIAID) through grant no. 5U01-AI069924-05. NA was supported by funding from the Swiss National Science Foundation (grant no. P500PM_203010).

Competing interests

All authors declare that they have no conflicts of interest.

REFERENCES

1. World Health Organization. WHO Coronavirus (COVID-19) Dashboard [Internet]. [cited 2022 Apr 21]. Available from: https://covid19.who.int
2. Federal Office of Public Health (FOPH). Coronavirus: Situation in Switzerland [Internet]. [cited 2022 Apr 21]. Available from: https://www.bag.admin.ch/bag/en/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/situation-schweiz-und-international.html
3. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo 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 1;20(11):1263–71.
4. The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe. 2020 Jul 1;1(3):e99–100.
5. Goncalves Cabecinhas AR, Roloff T, Stange M, Bertelli C, Huber M, Ramette A, et al. SARS-CoV-2 N501Y Introductions and Transmissions in Switzerland from Beginning of October 2020 to February 2021—Implementation of Swiss-Wide Diagnostic Screening and Whole Genome Sequencing. Microorganisms. 2021 Apr;9(4):677.
6. Borges V, Isidro J, Macedo F, Neves J, Silva L, Paiva M, et al. Nosocomial Outbreak of SARS-CoV-2 in a "Non-COVID-19" Hospital Ward: Virus Genome Sequencing as a Key Tool to Understand Cryptic Transmission. Viruses. 2021 Apr 1;13(4):604.
7. Francis RV, Billam H, Clarke M, Yates C, Tsoleridis T, Berry L, et al. The Impact of Real-Time Whole-Genome Sequencing in Controlling Healthcare-Associated SARS-CoV-2 Outbreaks. J Infect Dis. 2022 Jan 5;225(1):10–8.
8. Walker A, Houwaart T, Finzer P, Ehlkes L, Tyshaieva A, Damagnez M, et al. Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting. Clin Infect Dis. 2022 Mar 23;74(6):1039–46.
9. Geoghegan JL, Ren X, Storey M, Hadfield J, Jelley L, Jefferies S, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. Nat Commun. 2020 Dec 11;11(1):6351.
10. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2017 Mar 30;22(13):30494.
11. Moser AI, Campos-Madueno EI, Sendi P, Perreten V, Keller PM, Ramette A, et al. Repatriation of a patient with COVID-19 contributed to the importation of an emerging carbapenemase producer. J Glob Antimicrob Resist. 2021 Dec;27:267–72.
12. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinforma Oxf Engl. 2019 Feb 1;35(3):526–8.
13. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960 Apr 1;20(1):37–46.
14. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. J Clin Med. 2020 Apr;9(4):967.
15. World Health Organization. Digital tools for COVID-19 contact tracing: annex: contact tracing in the context of COVID-19, 2 June 2020 [Internet]. Geneva: World Health Organization; 2020. Available from: https://apps.who.int/iris/handle/10665/332265

16. Madewell ZJ, Yang Y, Longini IM Jr, Halloran ME, Dean NE. Factors Associated With Household Transmission of SARS-CoV-2: An Updated Systematic Review and Meta-analysis. JAMA Netw Open. 2021 Aug 27;4(8):e2122240.
17. Martinez L, Shen Y, Mupere E, Kizza A, Hill PC, Whalen CC. Transmission of Mycobacterium Tuberculosis in Households and the Community: A Systematic Review and Meta-Analysis. Am J Epidemiol. 2017 Jun 15;185(12):1327–39.
18. Lalor MK, Anderson LF, Hamblion EL, Burkitt A, Davidson JA, Maguire H, et al. Recent household transmission of tuberculosis in England, 2010–2012: retrospective national cohort study combining epidemiological and molecular strain typing data. BMC Med. 2017 Dec;15(1):105.
19. Fetzer T, Graeber T. Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment. Proc Natl Acad Sci. 2021 Aug 17;118(33):e2100814118.
20. Davis EL, Lucas TCD, Borlase A, Pollington TM, Abbott S, Ayabina D, et al. Contact tracing is an imperfect tool for controlling COVID-19 transmission and relies on population adherence. Nat Commun. 2021 Sep 13;12(1):5412.
21. Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health. 2020 Oct 1;74(10):861–6.
22. Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis. 2020 Oct 1;20(10):1151–60.
23. Douglas J, Geoghegan JL, Hadfield J, Bouckaert R, Storey M, Ren X, et al. Real-Time Genomics for Tracking Severe Acute Respiratory Syndrome Coronavirus 2 Border Incursions after Virus Elimination, New Zealand. Emerg Infect Dis. 2021 Sep;27(9):2361–8.
24. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray KA, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat Med. 2020 Sep;26(9):1398–404.

FIGURE AND TABLES

Table 1: Characteristics of cases and contacts included in this analysis.

	Cases			Contacts
n (%)	Overall	Sequenced	Not sequenced	Overall
Total	359 (100)	213 (100)	146 (100)	460 (100)
Initial_cases	289 (80.5)	176 (82.6)	113 (77.4)	-
Secondary_cases	70 (19.5)	37 (17.4)	33 (22.6)	-
Sex				
Female	164 (45.7)	95 (44.6)	69 (47.3)	224 (48.7)
Male	189 (52.6)	115 (54.0)	74 (50.7)	226 (49.1)
Other	6 (1.7)	3 (1.4)	3 (2.1)	10 (2.2)
Age	31 (21-46)	31 (21-44)	33 (21-48)	29 (15-49)
Exposure				
Household	101 (28.1)	61 (28.6)	40 (27.4)	-
Travel	56 (15.6)	40 (18.8)	16 (11.0)	-

Work	70 (19.5)	27 (12.7)	15 (10.3)	-
Store/Public Transport	35 (9.7)	14 (6.6)	21 (14.4)	-
Restaurant/Bar/Club	22 (6.1)	17 (8.0)	5 (3.4)	-
School/Nursery	18 (5.0)	10 (4.7)	8 (5.5)	-
Healthcare	12 (3.3)	7 (3.3)	5 (3.4)	-
Event	9 (2.5)	5 (2.3)	4 (2.7)	-
Other/Unknown	64 (17.8)	32 (15.0)	32 (21.9)	-
Vaccinated	55 (15.3)	41 (19.2)	14 (9.6)	-

Table 2*:* Comparison of CT clusters to genomic clusters. CT clusters for which at least one case in the cluster had a sequence available, by original CT cluster size (the number of total cases in the CT cluster), the number of cases within the cluster who had a sequence available, and the number of different sequences within the clusters (the number of different genomic clusters that these cases belong to). Numbers in bold correspond to a match (all cases of a CT cluster were also linked by genomic sequencing), and numbers in italic to a "mismatch" (the cases of a unique CT cluster belonged to different genomic clusters).

CT cluster size	No. of cases per cluster with sequence available	Number of different sequences contained in single CT clusters			
		1	2	3	5
1	1	140			
2	1	19			
2	2	**6**	*9*		
3	2	**1**	*3*		
3	3	**1**	0	*1*	
4	2	**1**	0		
4	3	0	0	*1*	
5	5	0	0	0	*1*

DOI: 10.1093/infdis/jiad074

Figure 1: The SARS-CoV-2 pandemic in the canton of Solothurn, Switzerland, between June 4 and July 26, 2022. Panel A: Number of COVID-19 cases reported to the public health authorities with the proportion of samples with available sequence. **Panel B**: Variant composition among samples with available sequence. **Panel C**: Average size (95% confidence interval) of case-contact networks (cases with their secondary cases or contacts). **Panel D:** Average size (95% CI) of contact tracing (CT) clusters (primary and secondary cases). Number of networks/clusters used to estimate the weekly mean are indicated in red and dashed lines correspond to the overall mean for the whole study period.

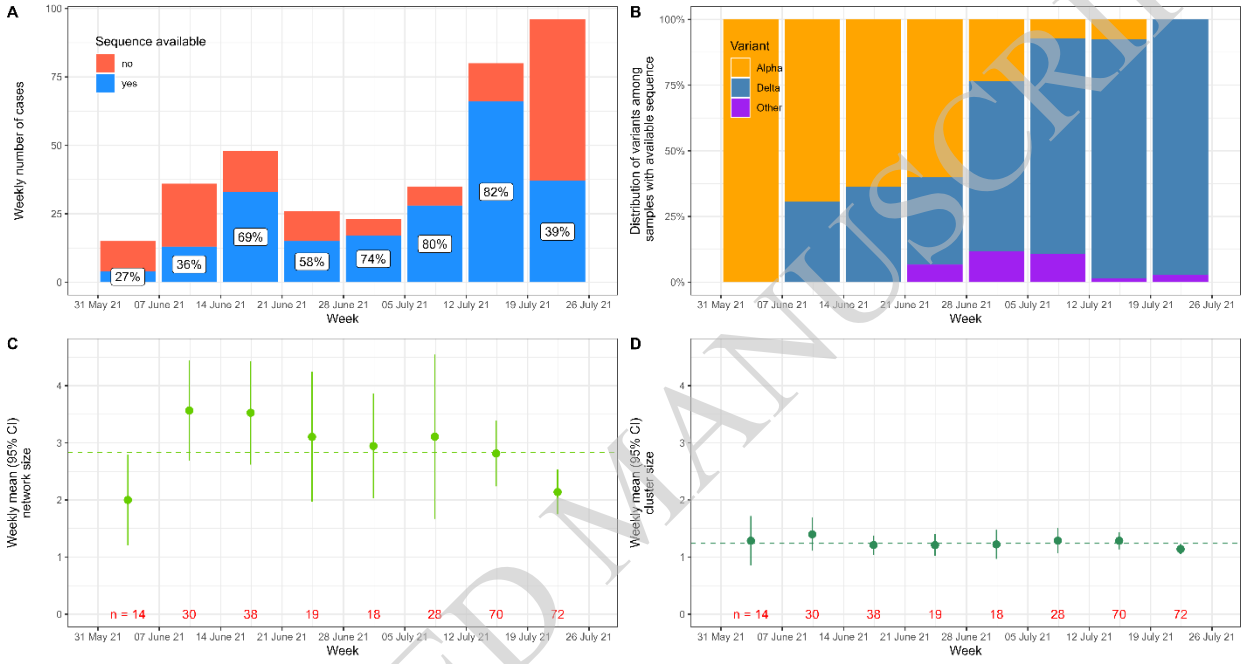

Figure 2: Minimum spanning tree of genomic differences among SARS-CoV-2 sequences. Pangolin lineages and vaccination status are indicated by different colours and shape, respectively.

Figure 3: Agreement between contact tracing (CT) linkage and genomic sequencing. Dots on the left represent CT clusters with at least one sequence available, stars indicate singleton clusters. Dots on the right represent the different genomic clusters i.e. unique sequences. A line connects a CT cluster to a genomic cluster if the sequence was found within the CT cluster.

For CT clusters with >1 sequence available within a cluster, linkage could be either *correct and complete* (all sequences were the same and the sequence was not found outside of the cluster - green), or *correct and incomplete* (all sequences were the same but the sequence was found outside of the cluster - blue), or *incorrect* (the sequences were different - red).

If only one sequence was available in a cluster, the cluster could either be *complete* (unique sequence which was not found elsewhere - orange) or *incomplete* (other cases with the same sequence existed but the links were missed).

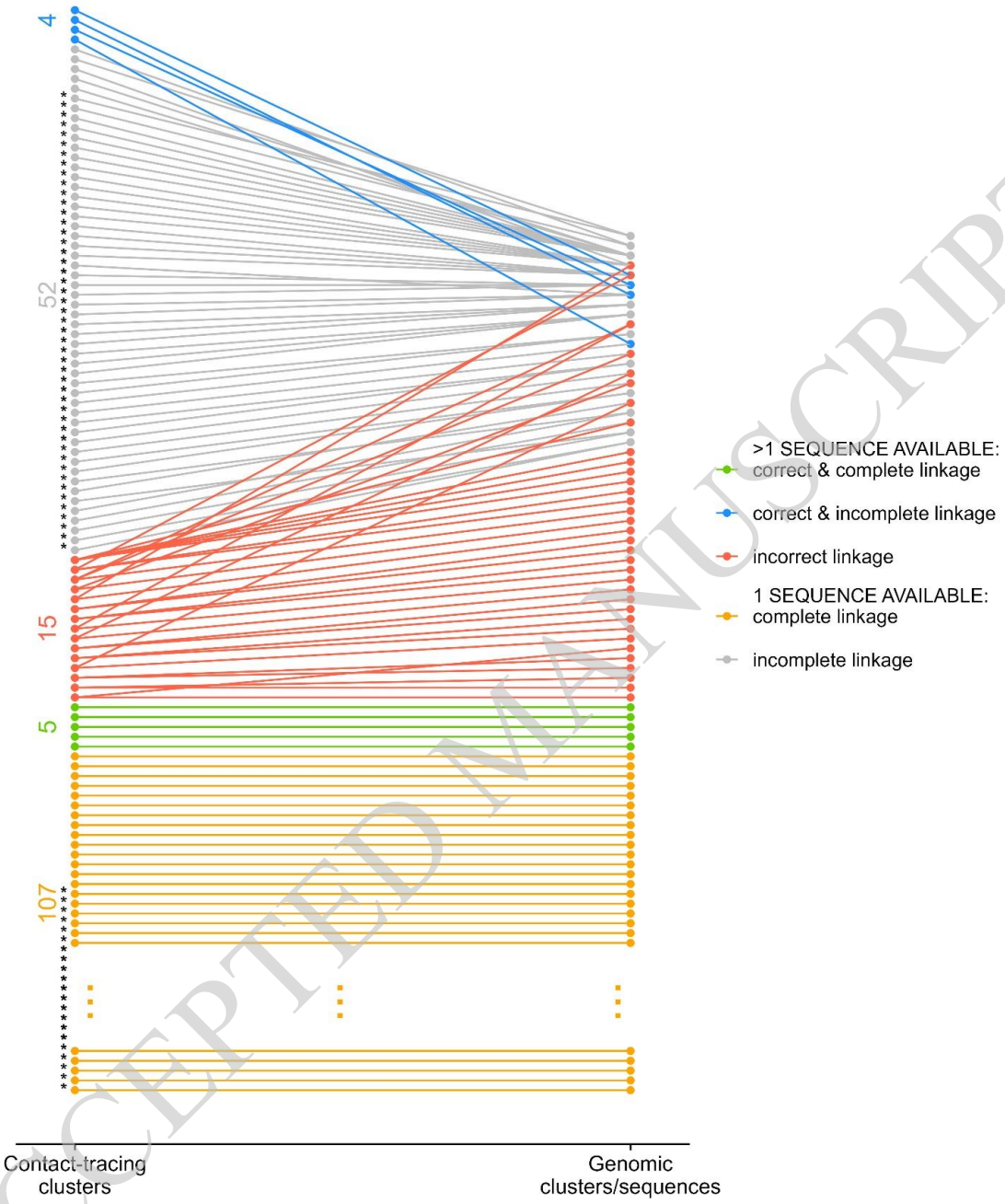

Figure 4: Comparison of CT and genomic clusters. Dots represent individual cases; lines represent links between initial and secondary cases reported to the CT team. The resulting CT clusters (dots connected by lines) are shown by size and date of positive test. The colours (and letters above cases) correspond to cases matched by sequencing (same genomic cluster). White dots are cases belonging to a single-case genomic cluster (a unique sequence not found for any other case). Black dots are cases for whom no sequencing was available.

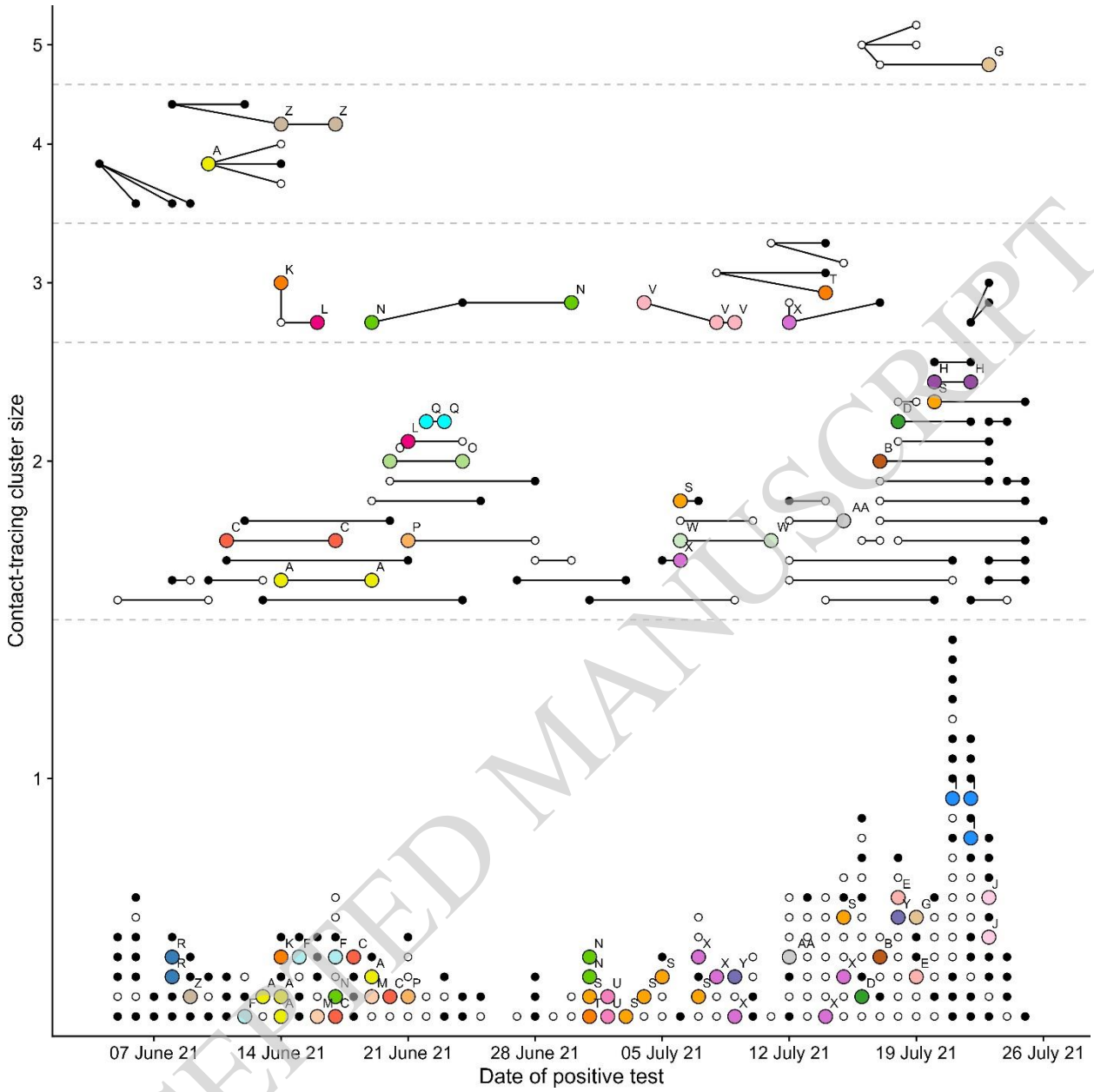

Figure 5: Comparison of CT (A) and genomic clusters (B) for persons in the same household (defined by the same home address). (A) shows CT clusters with cases (dots) connected by lines if a link was reported to the CT team, (B) shows genomic clusters with cases (dots) connected by lines if they shared the same genomic sequence (0 SNP difference). The colours (and letters above the cases) show which cases have reported the same home address (street name and number). For better visibility, only home-addresses that were reported for three or more cases are shown. Additionally, in 55 cases the same home-address was reported by exactly two cases (Supplementary Figure 6).