

Markus Glaß, Danny Misiak, Claudia Misiak, Simon Müller, Alexander Rausch, Katharina Angermann, Mariann Hoyer, Ramona Zabel, Astrid Kehlen, Beate Möbius, Jessica Weickert, Stefan Hüttelmaier and Matthias Karrasch\*

# Fast forward evolution in real time: the rapid spread of SARS-CoV-2 variant of concern lineage B.1.1.7 in Saxony-Anhalt over a period of 5 months

<https://doi.org/10.1515/labmed-2021-0092>

Received July 28, 2021; accepted October 4, 2021;

published online November 3, 2021

## Abstract

**Objectives:** Random mutations and recombinations are the main sources for the genetic diversity in SARS-CoV-2, with mutations in the SARS-CoV-2 spike (S) receptor binding motif (RBM) representing a high potential for the emergence of new putative variants under investigation (VUI) or variants of concern (VOC). It is of importance, to measure the different circulating SARS-CoV-2 lineages in order to establish a regional SARS-CoV-2 surveillance program. We established whole genome sequencing (WGS) of circulating SARS-CoV-2 lineages in order to establish a regional SARS-CoV-2 surveillance program.

**Methods:** We established a surveillance program for circulating SARS-CoV-2 lineages by performing whole genome sequencing (WGS) in SARS-CoV-2 isolates. Specimens were collected over a period of 5 months from three different sites. Specimens were collected from both patients suffering from COVID-19 and from outpatients

without any clinical signs or symptoms; both in a tertiary university hospital, and two private laboratories within an urban area of eastern part Germany.

**Results:** Viral WGS from the 364 respiratory specimens with positive SARS-CoV-2 RT-PCR comprised 16 different SARS-CoV-2 lineages. The majority of the obtained sequences (252/364=69%) was assigned to the variant of concern (VOC) Alpha (B.1.1.7). This variant first appeared in February in our samples and quickly became the dominant virus variant. All SNP PCR results could be verified using WGS. Other VOCs found in our cohort were Beta (B.1.351, n=2) and Delta (B.1.617.2, n=1).

**Conclusions:** Lineage analysis revealed 16 different virus variants among 364 respiratory samples analyzed by WGS. The number of distinct lineages dramatically decreased over time in leaving only few variants, in particular, the VOC Alpha or B.1.1.7. By closer inspection of point mutations, we found several distinct mutations of the viral spike protein that were reported to increase affinity or enable immune escape. Within a study period of only 5 months, SARS-CoV-2 lineage B.1.1.7 became the dominant lineage in our study population. This study emphasizes the benefit of SARS-CoV-2 testing by WGS. The increasing use of WGS to sequence the entire SARS-CoV-2 genome will reveal additional VIUs and VOCs with the potential to evade the immune system and, thus, will be a promising tool for data mining of relevant information for epidemiological studies. SARS-CoV-2 lineage monitoring using WGS is an important surveillance tool for early detection of upcoming new lineages of concern.

**Keywords:** B.1.1.7; B.1.351; B.1.617.2; Germany; SARS-CoV-2; Saxony-Anhalt; variant-of-concern; whole genome sequencing (WGS).

Stefan Hüttelmaier and Matthias Karrasch contributed equally to this work.

\*Corresponding author: Matthias Karrasch, Department of Laboratory Medicine, Unit III, Molecular Diagnostic Section, Halle University Hospital, Halle (Saale), Germany, E-mail: Matthias.karrasch@uk-halle.de

Markus Glaß, Danny Misiak, Claudia Misiak, Simon Müller, Alexander Rausch and Stefan Hüttelmaier, Institute of Molecular Medicine, Section for Molecular Cell Biology, Charles Tanford Protein Centre, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

Katharina Angermann, A&A Praxis für Laboratoriumsmedizin und Med. Mikrobiologie, Halle (Saale), Germany

Mariann Hoyer and Ramona Zabel, amedes MVZ für Laboratoriumsdiagnostik und Mikrobiologie Halle/Leipzig GmbH, Halle (Saale), Germany

Astrid Kehlen, Beate Möbius and Jessica Weickert, Department of Laboratory Medicine, Unit III, Molecular Diagnostic Section, Halle University Hospital, Halle (Saale), Germany

## Introduction

Like in other RNA viruses, random mutations and recombinations are the main sources for the genetic diversity in SARS-CoV-2 [1]. As this might lead to new potentially

dangerous mutations, whole genome sequencing (WGS) of SARS-CoV-2 has been identified as an important technology for surveillance [2]. Viral entry into human cells is determined through the spike protein ORF S [3]. Mutations in the SARS-CoV-2 spike receptor binding motif might lead to potential variants of concern (VOC) [4]. E.g., spike protein point mutation D614G led to a massive increase in infectivity of SARS-CoV-2, and viruses carrying this mutation became dominant in early 2020 [5, 6]. Another mutation of the spike protein that severely increased the infectivity is the N501Y substitution, prevalent in variants of concern like B.1.1.7 and B.1.351 [7]. In this study, we established a surveillance program for circulating SARS-CoV-2 lineages, in particular for the VOCs Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1) and Delta (B.1.617.2), from the area of Halle (Saale), located in the German federal state of Saxony-Anhalt. We performed whole genome sequencing of virus genomes obtained by tiled amplicon sequencing of samples collected between January and May 2021. While we observed a highly diverse lineage distribution at the beginning of the collection period, VOC B.1.1.7 first appeared in February, and then rapidly became the dominant form of the virus found in our samples.

## Materials and methods

### Sample collection

SARS-CoV-2 isolates were collected from COVID-19 patients as well as routine nasopharyngeal swabs from patients without any clinical signs and symptoms in a tertiary university hospital within an urban area of eastern part of Germany. Real-time qPCR (RT-PCR) was performed from patients' respiratory samples to analyze the viral genome copy number following our internal standard operation procedures. In case of SARS-CoV-2 positivity, a single nucleotide polymorphism (SNP) discrimination assay was established to discriminate N501Y and del69/70 variants. For SARS-CoV-2 whole genome sequencing, only SARS-CoV-2 positive samples with a Ct-value <25 were used.

### NGS library preparation, sequencing and data processing

Sequence libraries were constructed from total RNA extracts using the Illumina COVIDSeq Test kit (Illumina, San Diego, CA, US). Shortly, first-strand cDNA, serving as template for two distinct PCR reactions is synthesized using sets of primers generating amplicons that cover the whole SARS-CoV-2-genome. The generated amplicons were tagged with adapter sequences followed by another PCR, which adds primers including index-sequences. A pooled library of all samples was used

for sequencing. Samples were sequenced with an Illumina NextSeq 1000 (Illumina, San Diego, CA, US) in either single-end 100 nt or paired-end 50 nt mode.

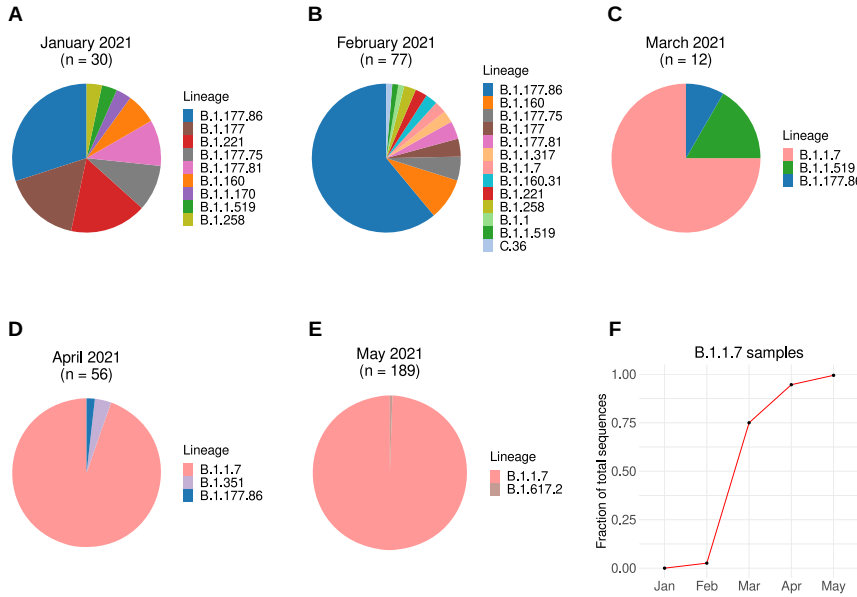
Low quality read ends as well as remaining parts of sequencing adapters were clipped off using Cutadapt v2.8 [8]. The processed sequencing reads were aligned to the SARS-CoV-2 reference genome (RefSeq NC\_045512.2) using the BWA aligner v0.7.17 [9]. Duplicated reads were removed using Picard tools v2.25 (broadinstitute.github.io/picard/). Virus consensus genome sequences were extracted from the mapped and deduplicated reads using bcftools v1.10.2 [10]. Lineage assignments according to the nomenclature proposed in Rambaut et al. [11] were obtained using pangolin v2.4.2 (github.com/cov-lineages/pangolin). Global Pairwise sequence alignments were obtained using the R-package Biostrings v2.61.1 [12].

Publicly available SARS-CoV-2 genome sequences from the whole federal republic of Germany collected between January and May 2021 were obtained from GISAID (gisaid.org; [13]). Only complete genome sequences not flagged with low coverage were used. Lineage assignments were obtained as described above for our own samples.

## Results

### Variant distribution

In total, 364 respiratory specimens from different patients with positive SARS-CoV-2 RT-PCR were collected between January and May 2021 and subsequently sequenced. Patients were either hospitalized because of respiratory symptoms due to COVID-19, or were outpatients who received SARS-CoV-2 swabs for surveillance reasons. The sequenced samples comprised 16 different SARS-CoV-2 lineages. During January and February 2021, we observed a quite diverse mixture of SARS-CoV-2 lineages among our samples (Figure 1A and B). In February 2021 we observed for the first time the variant of concern (VOC) B.1.1.7, also referred to as Alpha. Strikingly, the fraction of this variant increased tremendously and B.1.1.7 quickly became the dominant lineage in our samples (Figure 1C–F). This enforcement of B.1.1.7 could also be observed by considering SARS-CoV-2 genome sequences from the entire country of Germany which was collected in the same time period, and submitted to GISAID (gisaid.org). However, it appeared that B.1.1.7 approached earlier in some other regions of Germany, since this variant started to become dominant in the whole republic already in February 2021 (Supplemental Material, Figure S1). All samples classified as B.1.1.7 collected in February and March 2021 (n=11) were positively tested for B.1.1.7 via SNP PCR prior to sequencing. Besides B.1.1.7 we found three other samples, classified as VOCs, namely B.1.351 (Beta, n=2) in April and, most recently, B.1.617.2 (Delta, n=1) in May 2021.



**Figure 1:** Detection of different SARS-CoV-2 lineages by WGS. (A–E) Fractions of distinct lineages found between January and March 2021. (F) Fraction of samples classified as lineage B.1.1.7.

### Mutation analysis

Mutations were determined via pairwise sequence alignment (PSA) between the SARS-CoV-2 reference genome (RefSeq NC\_045512.2) and the respective consensus sequences. The median number of detected single nucleotide variations (SNVs) per consensus sequence increased over time from 20 in January to 39 in May 2021. The increase coincides with the rise of B.1.1.7, however, when considering only genome sequences from this variant, numbers were also slightly increasing, indicating the ongoing evolution of the virus (Figure 2A).

By far, the most common SNVs were C to U transitions (42% of all SNVs, Figure 2B). This is in agreement with previous reports and might indicate the activity of RNA editing enzymes like APOBEC or ADAR proteins from the human hosts [14, 15].

We could find several mutations in the spike protein (S) that have been described to either increase the affinity

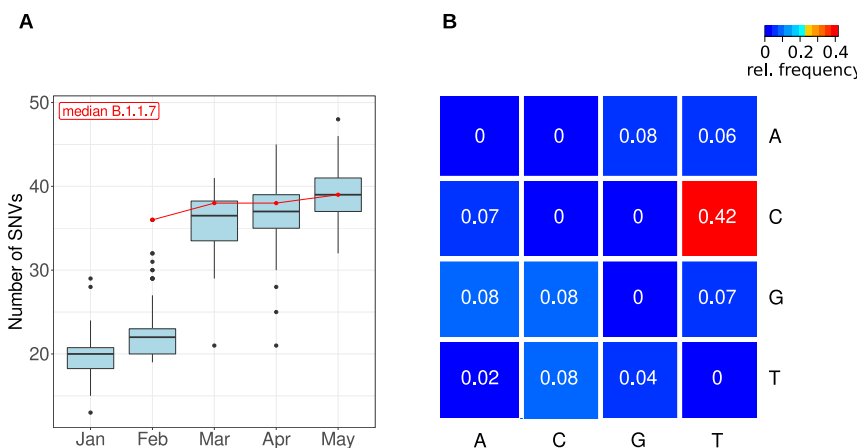
to the human ACE2 receptor or are under suspicion to confer immune evasive potential to the virus.

In all 364 samples the SNV causing the S:D614G substitution was found. This mutation was reported to increase the infectivity and viruses harboring this mutation rapidly became the most prevalent form [6].

The S:N501Y mutation was exclusively found in all but one samples classified as B.1.1.7 (n=251) and in both samples classified as B.1.351. This mutation has also been reported to increase infectivity [7] and is characteristic for the VOCs B.1.1.7 and B.1.351.

Another mutation of regard that we found exclusively in the 11 samples classified as B.1.160 or B.1.160.31, respectively is the S:S477N mutation. This mutation was associated with resistance to neutralisation by multiple monoclonal antibodies [16].

Another point mutation located in the receptor binding domain (RBD) of the spike protein reported to enhance binding affinity to the human ACE2 receptor is the N439K



**Figure 2:** SNV distribution. (A) Distribution of SNV numbers per SARS-CoV-2 genome obtained from samples collected at the indicated months. (B) Relative frequency of distinct types of point mutations in all sequenced samples. Row names indicate reference bases, column names indicate mutated bases.

**Table 1:** Known mutations of the spike protein reported to enhance infectivity or to enable immune escape found in the sequenced SARS-CoV-2 genomes.

Mutation	No. samples	Lineage(s)	PubMed ID (PMID)
D614G	364	All	32697968
N501Y	253	B.1.1.7, B.1.351	33413740
S477N	11	B.1.160, B.1.160.31	33535027
N439K	3	B.1.258	33621484
L452R	2	C.36, B.1.617.2	33758861, 33758899
Q677H	1	C.36	33594385
E484K	2	B.1.351	33846703
E484Q	2	B.1.1.7, B.1.177.86	33952596

mutation [17]. Among other SARS-CoV-2 variants, this mutation was found in the lineage B.1.258 and we recovered this mutation exclusively in the three samples classified as B.1.258.

The S:452R point mutation was found in the two samples classified as C.36 and B.1.617.2, respectively. This mutation was reported to strengthen the binding affinity and to enable the escape from neutralising antibodies [18].

In the sample classified as C.36 we also found another previously reported SNV in the spike protein, Q677H, which, so far, has not been shown to alter virulence, however, according to its location in the RBD, it has been speculated to influence viral properties [19].

Two different reported substitutions of the glutamic acid 484 of the spike protein were found: E484K and E484Q. While the first was found exclusively in the two B.1.351 samples, the latter was found in one sample classified as B.1.1.7 and another classified as B.1.177.86. Both mutations have been described to reduce antibody neutralisation [20, 21]. Table 1 summarizes the findings on mutations of the spike protein.

## Discussion

As reported in previous studies, there is a high need for SARS-CoV-2 lineage detection for the implication of further COVID-19 infection control measures. This was shown in particular during the initial SARS-CoV-2 outbreak report from Hamburg [22], but also in various other outbreak publications, e.g. from the biggest German meat processing plant, located in North Rhine Westphalia [23], from the district of Heinsberg in North Rhine-Westphalia [24], as well as from seroprevalence studies [25].

Applying WGS allowed us to directly infer the virus lineages instead of testing for only a few variants of interest via single nucleotide polymorphisms (SNP)-PCR [26]. SARS-CoV-2 WGS was able to determine the whole set of circulating SARS-CoV-2 lineages from the local region of interest. Furthermore, WGS allowed us to find and investigate all mutations in the virus consensus sequences, and not only those used for defining lineages. This facilitated a deeper insight into the evolution of SARS-CoV-2 as well as the distribution of certain mutations. SARS-CoV-2 WGS from the reported urban area in eastern part of Germany showed a highly variable lineage distribution that changed from a rather diverse population to a population highly dominated by the VOC Alpha (B.1.1.7) in only a few months. This development resembled the development of virus lineages in the entire country of Germany, for the same time period as reflected by publicly available data from GISAID. However, the spread of B.1.1.7 in the considered area showed a time-lag compared to the data from Germany, suggesting that this variant reached other regions of the country earlier. This is in line with a previous report, in which SARS-CoV-2 spread was seen first in western and southern regions of Germany, where early mobility restrictions were implemented [27]. Mobility restrictions following lockdown measures did not happen homogeneously in Germany, a greater mobility reduction occurred in western and southern states as compared to the eastern states of Germany (e.g. Saxony-Anhalt). The VOC Gamma (P.1) was not detected in our study, however, Beta (B.1.351) and Delta (B.1.617.2) were detected, in two and one patient samples, respectively. Thus, also our study underlines the need for regional surveillance programs, in order to be able to react quickly on changing virus lineage distribution, since differences in virus genotype can impact viral load, support continuous shedding of the virus, promote spread of SARS-CoV-2 and might impact the effectiveness of vaccination strategies against SARS-CoV-2.

**Acknowledgments:** The authors would like to thank Anke Cachay, Monique Dreher, Anastassia Teuber, Gabriele Unger and Nancy Wurg (Department of Laboratory Medicine, Unit III, Molecular Diagnostic Section, Halle University Hospital) for specimen handling, SARS-CoV-2 RNA isolation/inactivation and SNP-PCR, as well as Michael Böttcher and Ghanem El Kassem (Charles Tanford Protein Centre, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany) for cooperation with the Illumina Mini-Seq. We further acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's hCov19 Database (see Supplemental Material, Table S1).

**Research funding:** None declared.

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

**Informed consent:** Informed consent was obtained from all individuals included in this study.

**Ethical approval:** The local Institutional Review Board deemed the study exempt from review.

## References

- Daniloski Z, Jordan TX, Wessels HH, Hoagland DA, Kasela S, Legut M, et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* 2021;184:92–105.e16.
- Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol* 2020;6:veaa075.
- Pillay TS. Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. *J Clin Pathol* 2020;73:366–9.
- Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2021; 592:116–21.
- Isabel S, Grana-Miraglia L, Gutierrez JM, Bundalovic-Torma C, Groves HE, Isabel MR, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci Rep* 2020;10:14031.
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182: 812–27.e19.
- Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* 2021;26. <https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106>.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBNet J* 2011;17:10–2.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience* 2021; 10. <https://doi.org/10.1093/gigascience/giab008>.
- Rambaut A, Holmes EC, O’Toole A, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5:1403–7.
- Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. R package version 2.60.1 2021.
- Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob Chall* 2017;1:33–46.
- Azgari C, Kilinc Z, Turhan B, Circi D, Adebali O. The mutation profile of SARS-CoV-2 is primarily shaped by the host antiviral defense. *Viruses* 2021;13. <https://doi.org/10.3390/v13030394>.
- Simmonds P. Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. *mSphere* 2020;5. <https://doi.org/10.1128/mSphere.00408-20>.
- Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 2021;29:477–88.e4.
- Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 2021;184:1171–87.e20.
- Tchesnokova V, Kulakesara H, Larson L, Bowers V, Rechkina E, Kisiela D, et al. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. *J Clin Microbiol* 2021;59:e0092121.
- Hodcroft EB, Domman DB, Snyder DJ, Oguntuyo K, Van Diest M, Densmore KH, et al. Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. *medRxiv* 2021.
- Jangra S, Ye C, Rathnasinghe R, Stadlbauer D, Personalized Virology Initiative study group, Krammer F, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *Lancet Microbe* 2021;2:e283–4.
- Verghese M, Jiang B, Iwai N, Mar M, Sahoo MK, Yamamoto F, et al. Identification of a SARS-CoV-2 variant with L452R and E484Q neutralization resistance mutations. *J Clin Microbiol* 2021;59: e0074121.
- Pfefferle S, Günther T, Kobbe R, Czech-Sioli M, Nörz D, Santer R, et al. SARS Coronavirus-2 variant tracing within the first coronavirus disease 19 clusters in northern Germany. *Clin Microbiol Infect* 2021;27:130.e5–130.e8.
- Günther T, Czech-Sioli M, Indenbirken D, Robitaille A, Tenhaken P, Exner M, et al. SARS-CoV-2 outbreak investigation in a German meat processing plant. *EMBO Mol Med* 2020;12: e13296.
- Laufs J, Elsbernd H, Hölzl F, Winter KD. Coronavirus (SARS-CoV-2) – Ausbruch in Nordrhein-Westfalen (Kreis Heinsberg) – Erstdiagnose einer COVID-19-Infektion bei einem Ehepaar [Coronavirus (SARS-CoV-2) – Outbreak in North Rhine-Westphalia (District of Heinsberg) – First diagnosis of a COVID-19 manifestation in a couple]. *Dtsch Med Wochenschr* 2020;145:665–9.
- Tönshoff B, Müller B, Elling R, Renk H, Meissner P, Hengel H, et al. Prevalence of SARS-CoV-2 infection in children and their parents in Southwest Germany. *JAMA Pediatr* 2021;175:586–93.
- Harper H, Burridge A, Winfield M, Finn A, Davidson A, Matthews D, et al. Detecting SARS-CoV-2 variants with SNP genotyping. *PLoS One* 2021;16:e0243185.
- Schlosser F, Maier BF, Jack O, Hinrichs D, Zachariae A, Brockmann D. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc Natl Acad Sci U S A* 2020;117: 32883–90.

**Supplementary Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/labmed-2021-0092>).