

Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020

Andreas Walker^{1,2}, Torsten Houwaart^{2,3}, Tobias Wienemann³, Malte Kohns Vasconcelos³, Daniel Strelow³, Tina Senff¹, Lisanna Hülse³, Ortwin Adams¹, Marcel Andree¹, Sandra Hauka¹, Torsten Feldt⁴, Björn-Erik Jensen⁴, Verena Keitel⁴, Detlef Kindgen-Milles⁵, Jörg Timm¹, Klaus Pfeffer³, Alexander T Dilthey³

1. Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

2. These authors contributed equally

3. Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

4. Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

5. Department of Anaesthesiology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Correspondence: Alexander T Dilthey (alexander.dilthey@med.uni-duesseldorf.de)

Citation style for this article:

Walker Andreas, Houwaart Torsten, Wienemann Tobias, Vasconcelos Malte Kohns, Strelow Daniel, Senff Tina, Hülse Lisanna, Adams Ortwin, Andree Marcel, Hauka Sandra, Feldt Torsten, Jensen Björn-Erik, Keitel Verena, Kindgen-Milles Detlef, Timm Jörg, Pfeffer Klaus, Dilthey Alexander T. Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. Euro Surveill. 2020;25(22):pii=2000746. <https://doi.org/10.2807/1560-7917.ES.2020.25.22.2000746>

Article submitted on 25 Apr 2020 / accepted on 04 Jun 2020 / published on 04 Jun 2020

We whole-genome sequenced 55 SARS-CoV-2 isolates from Germany to investigate SARS-CoV-2 outbreaks in 2020 in the Heinsberg district and Düsseldorf. While the genetic structure of the Heinsberg outbreak indicates a clonal origin, reflecting superspreading dynamics from mid-February during the carnival season, distinct viral strains were circulating in Düsseldorf in March, reflecting the city's international links. Limited detection of Heinsberg strains in the Düsseldorf area despite geographical proximity may reflect efficient containment and contact-tracing efforts.

We report on the genetic structure of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in North-Rhine Westphalia, Germany's most populous state (18 million inhabitants). Our analysis includes the 'Heinsberg outbreak' [1], which started in the second half of February 2020 – comprising a superspreading event at a carnival session in Gangelt, a small municipality of ca 12,000 inhabitants on the border between Germany and the Netherlands – and subsequent outbreak dynamics in March, in the state capital Düsseldorf, located 70 km from Gangelt and an international economic and air travel hub of ca 600,000 inhabitants.

Severe acute respiratory syndrome coronavirus 2 genome sequencing

The institute of virology at Düsseldorf University Hospital was one of the first laboratories to offer SARS-CoV-2 diagnostics in North-Rhine Westphalia. A

total of 55 SARS-CoV-2 isolate samples were acquired from diagnostic swabs sent to this institute in February and March 2020. Of these, 10 were directly linked to the Heinsberg outbreak (obtained from medical practices in the Heinsberg district or from patients treated at Düsseldorf University Hospital who were Heinsberg district residents) and 45 originated from the city of Düsseldorf and surrounding districts.

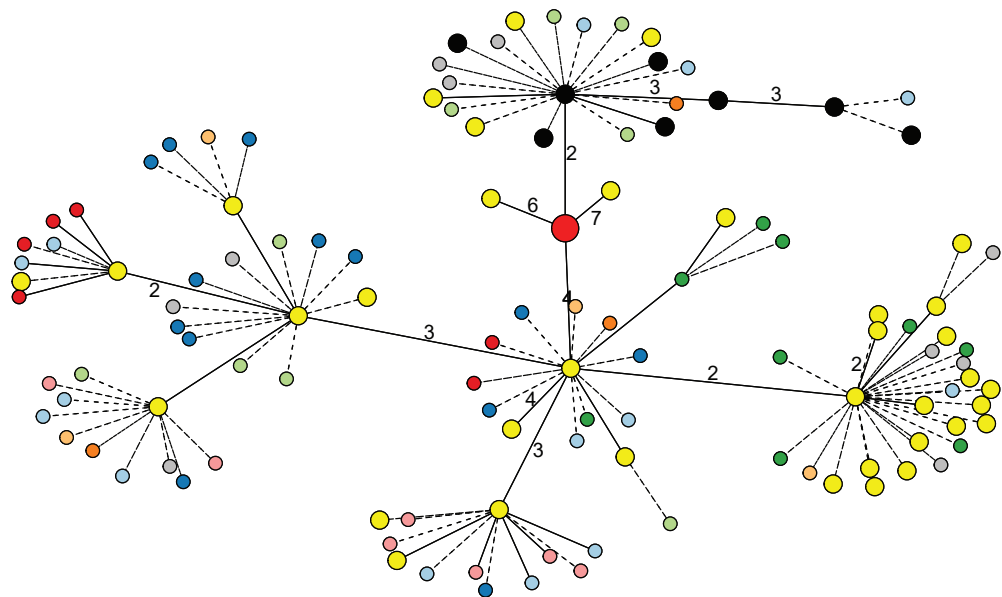
RNA extraction and reverse transcription were carried out as previously described [2]. DNA amplification and sequencing on the Oxford Nanopore platform were carried out according to the Artic protocol [3,4] (Supplementary Text), yielding between 31 and 582 Mb of raw sequencing data per sample (Supplementary Table S1). Bioinformatic analysis was based on the Artic pipelines and additional manual curation was carried out (Supplementary Text), yielding completely resolved genomes with 2–13 polymorphic positions (Supplementary Table S2) relative to the SARS-CoV-2 reference genome [5].

Of note, we observed evidence for ambiguities at polymorphic positions in 11 of 55 samples (Supplementary Table S2); for one such sample (NRW-39; 13 positions called as multi-allelic), PCR was repeated and a separate sequencing run was carried out, confirming the detected ambiguities (Supplementary Text). Further work is necessary to investigate whether ambiguities represent within-patient viral quasispecies.

FIGURE 1

Minimum spanning tree of severe acute respiratory syndrome coronavirus 2 sequences, showing 44 unambiguously^a resolved genomes from the Heinsberg district (n = 8) and the Düsseldorf area (n = 36), Germany, February–March 2020

- Wuhan reference
- Heinsberg district; n = 8
- Düsseldorf area; n = 36
- Netherlands; n = 14
- England; n = 13
- Belgium; n = 9
- US; n = 9
- Iceland; n = 7
- Australia; n = 6
- Scotland; n = 4
- Brazil; n = 3
- Other; n = 10



GISAID: Global Initiative on Sharing All Influenza Data; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; US: United States.

^a Eleven isolate genomes from 55 in our study with ambiguities are not considered ‘unambiguously resolved’ for the purposes of this analysis and are thus omitted.

The original Wuhan SARS-CoV-2 reference genome; and a sample of closely related publicly available SARS-CoV-2 genomes from GISAID (distance to any of the Heinsberg/Düsseldorf genomes of 0 or 1; see Supplementary Text for details) are included in the analysis.

Dashed and solid edges without adjacent numbers indicate distances of 0 and 1, respectively; all other distances are shown explicitly.

In a proof-of-concept experiment, we also successfully sequenced reverse-transcribed viral cDNA from patient material without an intermediate PCR-based amplification step (Supplementary Text), potentially enabling simplified sample preparation and increased read lengths for some samples in the future.

Ethical statement

Our study was Institutional Review Board (IRB)-approved by the ethics committee of the Heinrich Heine University Düsseldorf (#2020–839).

The Heinsberg outbreak

The first cases of SARS-CoV-2 infection in Germany were detected in late January 2020 and could be linked to recent travel to Northern Italy and China [1]. On 24 and 25 February 2020, however, two members of the same household from the Heinsberg district with no known travel history to SARS-CoV-2 risk areas were diagnosed with SARS-CoV-2; by 28 February 2020, the number of confirmed infections in the Heinsberg district had grown to 37; by 22 April 2020, to >1,700 [6]. Contact tracing later showed that many of the early SARS-CoV-2 cases could be linked to a carnival session attended by the two index cases. The carnival event was held on 15 February 2020 in the municipality

of Gangelt, which is part of the Heinsberg district [1]. Epidemiological investigation revealed that the index cases had travelled to the Netherlands, not considered a risk area at the time, 7 days prior to attending the Gangelt carnival session. The ‘Heinsberg outbreak’ represented one of the first large-scale SARS-CoV-2 outbreaks in Germany, seeded by community transmission and amplified by superspreading-type dynamics.

Genomic analysis of 10 SARS-CoV-2 isolates from the Heinsberg outbreak, sampled between 25 and 28 February and including those from the index cases, demonstrated the clonal origin of the outbreak (Figure); all Heinsberg samples shared the same two mutations compared to the SARS-CoV-2 reference genome (Supplementary Table S2). Viral diversity in the Heinsberg samples varied between two and six polymorphic positions relative to the SARS-CoV-2 reference genome, and five distinct viral variants (i.e. haplotypes) could be identified (Supplementary Table S2).

An analysis (Supplementary Text) of other publicly available SARS-CoV-2 sequences did not reveal an obvious origin of the Heinsberg outbreak (Supplementary Table S3); the Heinsberg isolates are not related to early sequences from other German outbreak areas

(Bavaria, Baden-Wuerttemberg), and, despite intense Dutch viral sampling (585 available viral genomes from the Netherlands at the time of analysis), our analysis identified only two closely related isolates from the Netherlands (one collected on 21 March, the other with undefined collection date). The role of the index cases' short vacation in the Netherlands 7 days before the Gangelt carnival session [7], while suggestive in terms of reported SARS-CoV-2 incubation periods thus remains ambiguous [8]. Moreover, large numbers of closely related isolates are circulating in many countries, for example England, Wales, and Iceland (Supplementary Table S3). The small number of polymorphisms shared by all samples in the Heinsberg outbreak ($n=2$), compared with a maximum number of six per-isolate polymorphic positions in the same samples, likely acquired over a period of a few weeks, is compatible with a relatively recent introduction from China.

Düsseldorf outbreak dynamics

The first SARS-CoV-2 cases in Düsseldorf, 70 km from Gangelt, were diagnosed in early March 2020 [9]; as at 21 April 2020, the outbreak had grown to more than 900 confirmed cases [10]. The set of 55 whole-genome-sequenced isolates included 45 samples from Düsseldorf and nearby districts, collected between 3 and 23 March. A minimum spanning tree analysis of 44 unambiguously resolved viral sequences (Figure) showed that there were at least five clusters of viral stains circulating in the Düsseldorf area; the number of polymorphic positions relative to the SARS-CoV-2 reference genome in the Düsseldorf samples varied between 2 and 13 (Supplementary Table S2). Closely related strains (distance 0 or 1) were found in Australia, the United Kingdom, the United States and many other countries (Supplementary Table S3), strongly suggesting multiple independent introduction events. Of note, four 'Düsseldorf area' isolates clustered with the Heinsberg outbreak (Figure 1); of these, two were collected from residents of a district next to Heinsberg, who had been treated at the Düsseldorf University Hospital, and two remained of unclear origin (patient data not available). Thus, there was no evidence for widespread community circulation of Heinsberg-derived SARS-CoV-2 strains in the Düsseldorf area.

Illumina validation

To verify the accuracy of Nanopore-based viral assembly, additional Illumina sequencing was carried out for the first 11 samples, according to date of collection, of our cohort (Supplementary Table S1; Supplementary Text); data analysis was carried out with iVar [11]. For 41 of 45 polymorphic positions identified by either Nanopore or Illumina across the 11 samples, the called alleles agreed; manual inspection of the discordant positions revealed low coverage for two discordant positions and one missed multi-allelic call for each sequencing technology (Supplementary Table S4).

Discussion

Since its emergence in the Chinese city of Wuhan in late 2019, SARS-CoV-2 has infected more than 6 million individuals and led to more than 370,000 deaths worldwide as at 03 June 2020 [12]. As SARS-CoV-2 case numbers and the social and economic consequences of social distancing and lock-down measures continue to rise, many countries are facing difficult trade-offs. Improved methods to characterise the dynamics of viral transmission are urgently needed.

More than 10,000 globally sourced SARS-CoV-2 genomes are publicly available, and powerful data sharing and analysis platforms like the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database [13] and Nextstrain [14] enable the collaborative analysis of viral population structure on a global level. Additional insights into transmission dynamics can be gained from focused investigations of individual outbreaks and by integrating genomic data with classical epidemiology.

Here we have investigated the genetic structure of two SARS-CoV-2 outbreaks, which occurred at two nearby locations in North-Rhine Westphalia using Nanopore sequencing, which has additional applications in many fields such as human genetics [15] and microbial metagenomics [16]. We have demonstrated the clonal origin of the Heinsberg outbreak. This is consistent with available epidemiological data pointing to a carnival session in Gangelt as the epicentre of the outbreak [1]. The lack of association between the Heinsberg samples and other early German outbreak isolates is suggestive of a separate introduction event, possibly via the Netherlands, China, or a third country. By contrast, SARS-CoV-2 isolates circulating in Düsseldorf were highly polyclonal and could be grouped into at least five clusters of viral haplotypes.

Despite the geographical proximity between Heinsberg and Düsseldorf, only four of 36 unambiguously resolved samples from the Düsseldorf area clustered with the Heinsberg outbreak, and two of these were derived from residents of a district neighbouring Heinsberg. Limited detection of Heinsberg strains in the Düsseldorf area may reflect the effectiveness of the contact-tracing efforts conducted by the German public health authorities; of note, 'lockdown'-type restrictions with limits on public gatherings in Germany were only imposed on 23 March 2020 [17], i.e. on the day on which the last sample of our study was collected.

More extensive sampling of SARS-CoV-2 isolates from North-Rhine Westphalia will be required to investigate the effect of various containment measures on transmission chains at a genomic level. Consistent with reports from Iceland [18], New York [19], and data on Nextstrain, our study has demonstrated the simultaneous circulation of distinct viral variants (i.e. haplotypes) in a metropolitan region. In the Heinsberg outbreak, we could identify five distinct variants. As

SARS-CoV-2 genomes continue to diverge as part of ongoing viral evolution, the application of genomic epidemiology [20,21] for the identification and targeted interruption of viral transmission chains will become increasingly feasible.

Data availability

All generated viral genome assemblies have been submitted to GISAID; all generated assemblies and the raw sequencing data are also available on NCBI (BioProject PRJNA627229).

Acknowledgements

This work was supported by the Jürgen Manchot Foundation and by funding from the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; Award number 031L0184B).

We gratefully acknowledge the Authors, the Originating and Submitting Laboratories for their sequence and metadata shared through GISAID. All submitters of data may be contacted directly via GISAID. The Acknowledgments Table for GISAID is part of the Supplement (Supplementary Table S5).

We would also like to thank Nicholas Loman and Josh Quick for advice and discussions.

Conflict of interest

None declared.

Authors' contributions

AW, TH, TW, MKV, JT, KP and ATD conceptualised and designed the study. AW, DS, TS, LH, and OA designed and implemented wet-lab protocols. TH and ATD developed and implemented sequencing data analysis approaches. OA, MA, SH, TF, BJ, VK, and DKM provided clinical samples and epidemiological data and gave input into the study design. All authors have commented on the draft and approved the final version.

References

1. Robert Koch Institute (RKI). Coronavirus Disease 2019 (COVID-19) Daily Situation Report 05/03/2020. Berlin: RKI; Mar 2020. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-03-05-en.pdf?__blob=publicationFile
2. Walker A, Ennker KS, Kaiser R, Lübke N, Timm J. A pan-genotypic Hepatitis C Virus NS5A amplification method for reliable genotyping and resistance testing. *J Clin Virol*. 2019;113:8-13. <https://doi.org/10.1016/j.jcv.2019.01.012> PMID: 30771598
3. Quick, J. ARTIC amplicon sequencing protocol for MinION for nCoV-2019.2020. <https://doi.org/http://dx.doi.org/10.17504/protocols.io.bdp7i5rn>
4. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017;12(6):1261-76. <https://doi.org/10.1038/nprot.2017.066> PMID: 28538739
5. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-9. <https://doi.org/10.1038/s41586-020-2008-3> PMID: 32015508
6. Kreis Heinsberg. Coronavirus im Kreis Heinsberg. [Coronavirus in the district of Heinsberg]. Heinsberg: Kreisverwaltung

Heinsberg; 2020. Available from: <https://www.kreis-heinsberg.de/aktuelles/aktuelles/?pid=5149>

7. Netherlands National Institute for Public Health and the Environment (RIVM). Duitse coronapatiënt niet ziek tijdens verblijf in Limburg. [German corona patient not ill during stay in Limburg]; 26 Feb 2020. Biltoven: RIVM. Available from: <https://www.rivm.nl/nieuws/duitse-coronapatiënt-niet-ziek-tijdens-verblijf-in-limburg>
8. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020;172(9):577-82. <https://doi.org/10.7326/M20-0504> PMID: 32150748
9. Pressedienst Landeshauptstadt Düsseldorf. Zwei Düsseldorfer mit Coronavirus infiziert. [Two Düsseldorfers infected with coronavirus]. Düsseldorf: Pressedienst Landeshauptstadt Düsseldorf; 2020. [Accessed 04 Jun 2020]. Available from: <https://www.duesseldorf.de/medienportal/pressedienst-einzelansicht/pld/zwei-duesseldorfer-mit-coronavirus-infiziert.html>
10. Pressedienst Landeshauptstadt Düsseldorf. Die Coronazahlen vom 21. April. [Coronavirus case numbers 21 April]. Düsseldorf: Pressedienst Landeshauptstadt Düsseldorf; 2020. <https://www.duesseldorf.de/medienportal/pressedienst-einzelansicht/pld/die-coronazahlen-vom-21-april.html>
11. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. 2019;20(1):8. <https://doi.org/10.1186/s13059-018-1618-7> PMID: 30621750
12. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) Situation Report – 135. Geneva: WHO; 03 Jun 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200603-covid-19-sitrep-135.pdf?sfvrsn=39972feb_2
13. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill*. 2017;22(13):30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494> PMID: 28382917
14. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-3. <https://doi.org/10.1093/bioinformatics/bty407> PMID: 29790939
15. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018;36(4):338-45. <https://doi.org/10.1038/nbt.4060> PMID: 29431738
16. Dilthey AT, Jain C, Koren S, Phillippy AM. Strain-level metagenomic assignment and compositional estimation for long reads with MetaMaps. *Nat Commun*. 2019;10(1):3066. <https://doi.org/10.1038/s41467-019-10934-2> PMID: 31296857
17. Robert Koch Institute (RKI). Coronavirus Disease 2019 (COVID-19) Daily Situation Report 23/03/2020. Berlin: RKI; Mar 2020. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-03-23-en.pdf?__blob=publicationFile
18. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med*. 2020;NEJMoa2006100. <https://doi.org/10.1056/NEJMoa2006100> PMID: 32289214
19. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary A, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *medRxiv*. 2020.04.08.20056929 (preprint). <https://doi.org/10.1126/science.abc1917>
20. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. *Nat Microbiol*. 2019;4(1):10-9. <https://doi.org/10.1038/s41564-018-0296-2> PMID: 30546099
21. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet*. 2018;19(1):9-20. <https://doi.org/10.1038/nrg.2017.88> PMID: 29129921

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2020.