nature medicine

LETTERS

https://doi.org/10.1038/s41591-020-0997-y

Check for updates

Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands

Bas B. Oude Munnink[1], David F. Nieuwenhuijse[1], Mart Stein[2], Áine O'Toole[3], Manon Haverkate[2], Madelief Mollers[2], Sandra K. Kamga[2], Claudia Schapendonk[1], Mark Pronk[1], Pascal Lexmond[1], Anne van der Linden[1], Theo Bestebroer[1], Irina Chestakova[1], Ronald J. Overmars[1], Stefan van Nieuwkoop[1], Richard Molenkamp[1], Annemiek A. van der Eijk[1], Corine GeurtsvanKessel[1], Harry Vennema[2], Adam Meijer[2], Andrew Rambaut[3], Jaap van Dissel[2], Reina S. Sikkema[1], Aura Timen[2,33], Marion Koopmans[1,33] ✉ and The Dutch-Covid-19 response team*

In late December 2019, a cluster of cases of pneumonia of unknown etiology were reported linked to a market in Wuhan, China[1]. The causative agent was identified as the species *Severe acute respiratory syndrome-related coronavirus* and was named SARS-CoV-2 (ref. [2]). By 16 April the virus had spread to 185 different countries, infected over 2,000,000 people and resulted in over 130,000 deaths[3]. In the Netherlands, the first case of SARS-CoV-2 was notified on 27 February. The outbreak started with several different introductory events from Italy, Austria, Germany and France followed by local amplification in, and later also outside, the south of the Netherlands. The combination of near to real-time whole-genome sequence analysis and epidemiology resulted in reliable assessments of the extent of SARS-CoV-2 transmission in the community, facilitating early decision-making to control local transmission of SARS-CoV-2 in the Netherlands. We demonstrate how these data were generated and analyzed, and how SARS-CoV-2 whole-genome sequencing, in combination with epidemiological data, was used to inform public health decision-making in the Netherlands.

Whole-genome sequencing (WGS) is a powerful tool to understand the transmission dynamics of outbreaks and inform outbreak control decisions[4–7]. Evidence of this was seen during the 2014–2016 West African Ebola outbreak when real-time WGS was used to help public health decision-making, a strategy dubbed 'precision public health pathogen genomics'[8,9]. Immediate sharing and analysis of data during outbreaks is now recommended as an integral part of outbreak response[10–12]. Feasibility of real-time WGS requires access to sequence platforms that provide reliable sequences, access to metadata for interpretation, and data analysis at high speed and low cost. Therefore, WGS for outbreak support is an active area of research. Nanopore sequencing has been employed in recent outbreaks of Usutu, Ebola, Zika and yellow fever virus owing to the ease of use and relatively low start-up cost[4–7]. The robustness of this method has recently been validated using Usutu virus[13,14]. In the Netherlands, the first COVID-19 case was confirmed on 27 February and WGS was performed in near to real-time using an amplicon-based sequencing approach.

From 22 January, symptomatic travelers from countries where SARS-CoV-2 was known to circulate were routinely tested. The first case of SARS-CoV-2 infection in the Netherlands was identified on 27 February in a person with recent travel history to Italy and an additional case was identified one day later, also in a person with recent travel history to Italy. The genomes of these first two positive samples were generated and analyzed by 29 February. These two viruses clustered differently in the phylogenetic tree, confirming separate introductions (Fig. 1a).

The advice to test hospitalized patients with serious respiratory infections was issued on 24 February and subsequent attempts to identify possible local transmission chains triggered testing for SARS-CoV-2 on a large scale in hospitals. By 9 March local clusters of epidemiologically related cases of SARS-CoV-2 started to appear in the province of Noord-Brabant. The increase in cases was caused by several co-circulating viruses, and is likely to have been triggered by multiple introductions of the virus following the spring holidays (from 13 to 23 February) with travel to ski resorts in Northern Italy (Fig. 1b). The first intervention was put in place on 9 and 10 March when the prime minister advised people to stop shaking hands and events attended by more than 1,000 visitors were banned in the province of Noord-Brabant. Subsequent analysis identified clusters with local amplification of viruses from patients without any travel history, also outside Noord-Brabant (Fig. 2). This information, combined with the increase in the total number of infections in the Netherlands, led to the decision to implement stricter measures for the whole country to prevent further spread of SARS-CoV-2 on 12 March. All events with more than 100 people attending were canceled, people were requested to work from home as much as possible and people with symptoms such as a fever or cough had to stay at home. On 15 March, this was followed by the closure of schools, catering industries and sport clubs.

In the third phase, sequencing of new cases with emphasis on health-care workers (HCWs) and hospitalized cases was continued.

[1]ErasmusMC, Department of Viroscience, WHO Collaborating Centre for Arbovirus and Viral Hemorrhagic Fever Reference and Research, Rotterdam, the Netherlands. [2]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands. [3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. [33]These authors contributed equally: Aura Timen, Marion Koopmans. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: m.koopmans@erasmusmc.nl

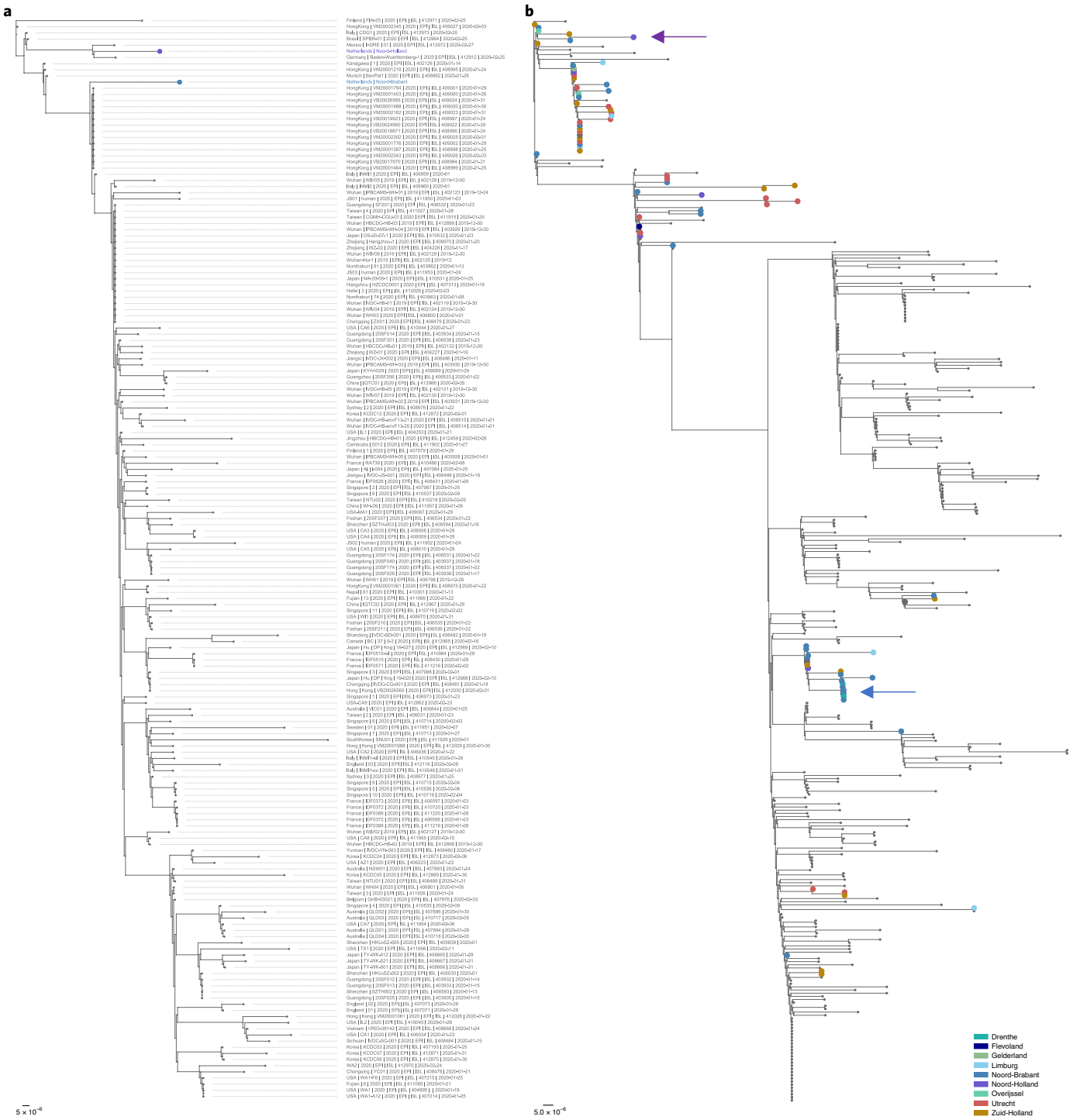


Fig. 1 | Phylogenetic analysis of the first two Dutch SARS-CoV-2 sequences. a,**b**, All sequences that were publicly available on 29 February (**a**) or 9 March (**b**) are included in the analysis. The sequences are colored on the basis of the province of detection. The scale bar represent the amount of nucleotide substitutions per site. Red indicates the Dutch isolates and blue represents SARS-CoV-2 sequences from other countries with recent travel history to Italy.

By 15 March, 189 SARS-CoV-2 viruses from the Netherlands were sequenced, at that moment representing 27.1% of the total number of full genome sequences produced worldwide. The sequences detected in the Netherlands continued to be diverse and revealed the presence of multiple co-circulating sequence types, found in several different clusters in the phylogenetic tree (Fig. 3 and Extended Data Fig. 1). This diversity was also observed in cases with similar travel histories, reflecting that sequence diversity was already present in the originating county, primarily Italy (Fig. 4). In addition to travel-associated cases, an increasing number of local cases was detected through severe acute respiratory infection surveillance; this was not limited to the province Noord-Brabant but SARS-CoV-2 was also increasing in the provinces Zuid-Holland, Noord-Holland and Utrecht, confirming substantial under-ascertainment of the epidemic. The increase in the number of patients with COVID-19 as well as increasing affected geographic areas and occurrence of local clusters provided further support for the increased movement restrictions.

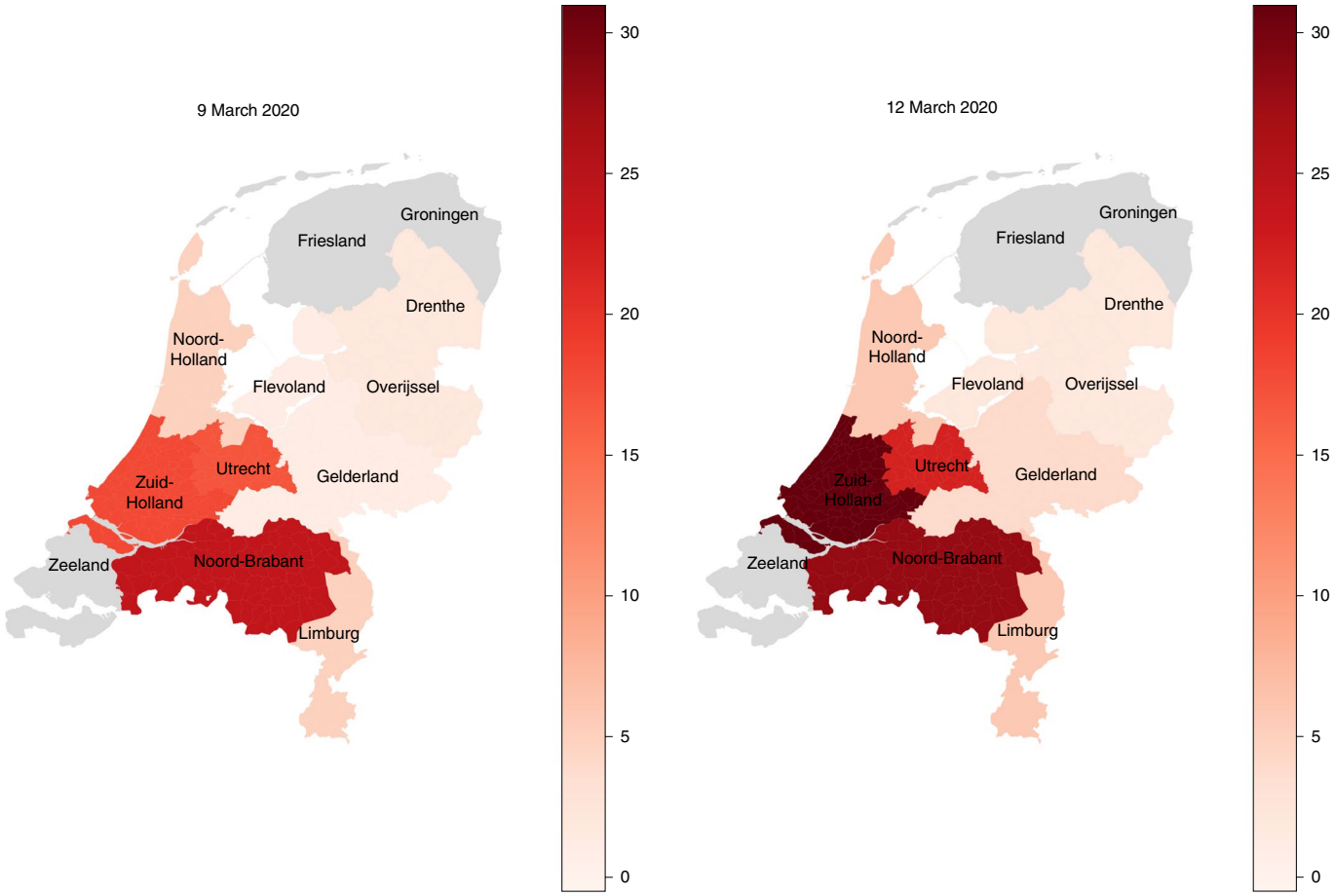


Fig. 2 | Distribution of SARS-CoV-2 sequences from the Netherlands on 9 and 12 March. The shapefile for the map is derived from https://gadm.org. The color scale represents the location and the number of whole-genome sequences generated at the indicated time points.

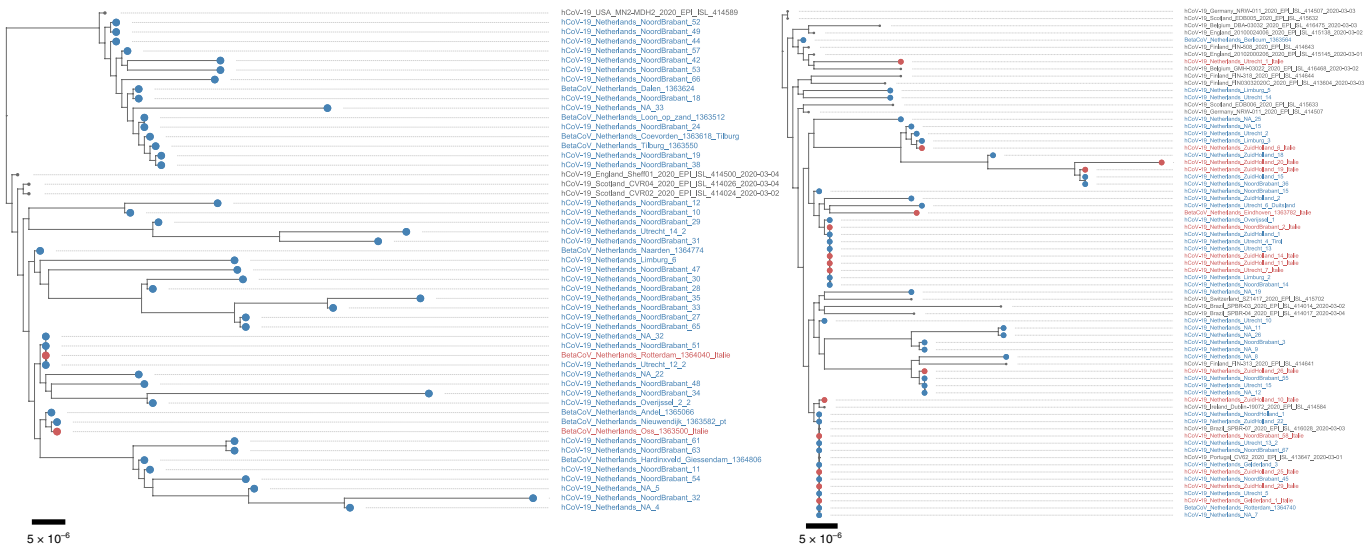


Fig. 3 | Phylogenetic analysis of SARS-CoV-2 emergence in the Netherlands. Zoom-ins of two clusters circulating in the Netherlands. The sequences are colored on the basis of travel history; Dutch patients without travel history are indicated in blue while Dutch patients with travel history to Italy are indicated in red. The scale bars represent the number of substitutions per site.

BEAST analysis revealed that the most recent ancestor of the viruses circulating in the Netherlands dates back to the end of January and the beginning of February (Fig. 4). This is in line with the amplification that occurred in the region (notably Italy and Austria) from which most of the epidemic in the Netherlands was seeded. Most incursions likely occurred during spring break, which is a popular time for winter sports vacations. Retrospective testing showed the presence of the virus in a sample

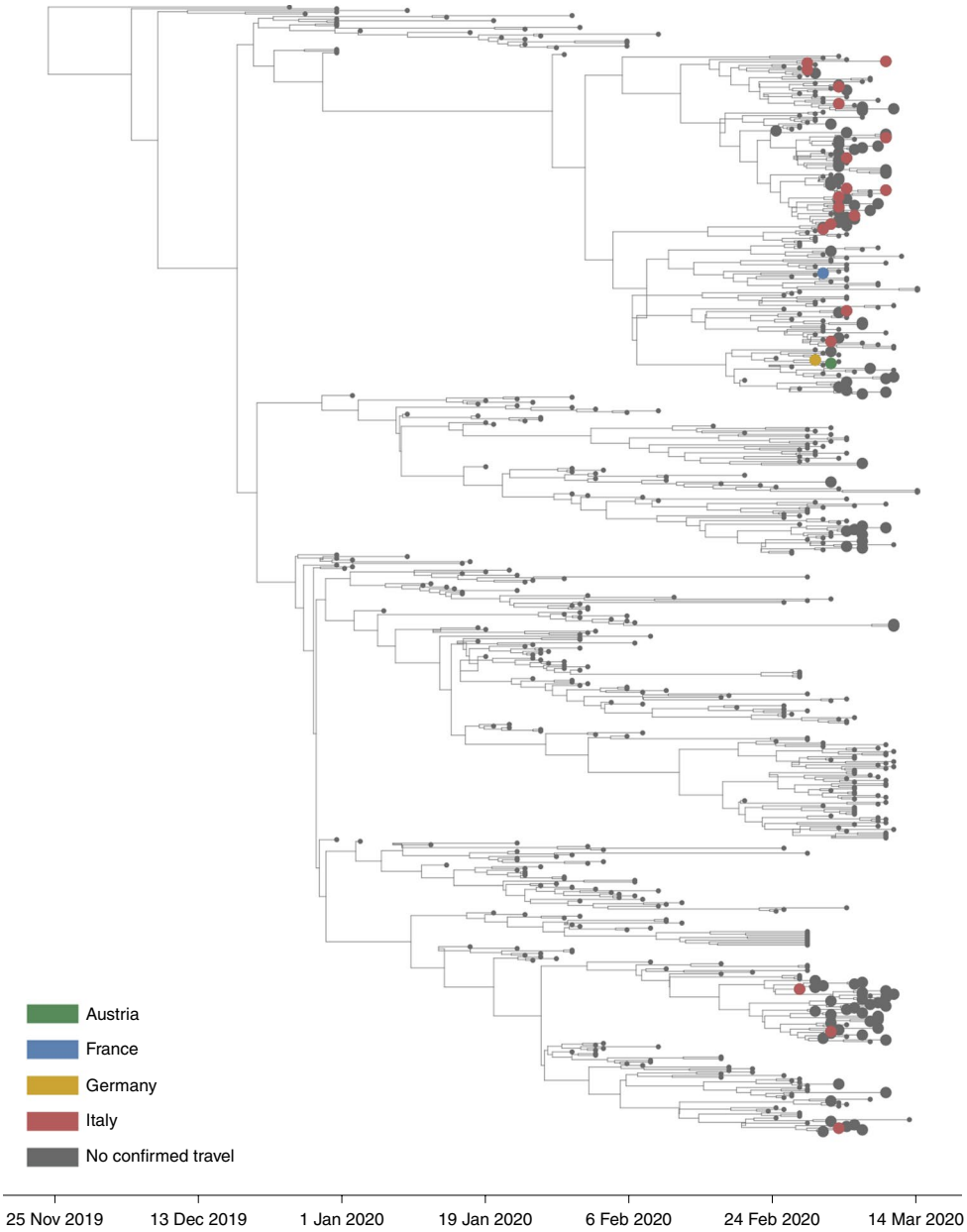


Fig. 4 | BEAST analysis with travel history. Time-resolved visualization of the emergence of SARS-CoV-2 in the Netherlands. Sequences from the Netherlands are depicted with big circles. Green indicates recent travel history to Austria, blue to France, yellow to Germany, and red to Italy.

collected on 24 February in a patient with known travel history to Italy.

In this study, we show that WGS in combination with epidemiological data strengthened the evidence base for public health decision-making in the Netherlands as it enabled a more precise understanding of the transmission patterns in various initial phases of the outbreaks. As such, we were able to understand the genetic diversity of the multiple introduction events in phase 1, the extent of local and regional clusters in phase 2 and the transmission patterns within the HCW groups in phase 3 (among which the absence or occurrence of very limited nosocomial transmission). This information complemented the data obtained from more traditional methods such as contact investigation.

At the time of the study, sequences from the Netherlands made up a substantial part of the total collection of SARS-CoV-2 genomes. Although implementation of WGS in the Dutch disease prevention and control strategy has shown its added value, there were limitations due to the paucity of genomic information available from certain parts of the world, including Italy. The information available from Iran, another major country where the virus was presumably spreading exponentially in the week before the take-off of the epidemic in the Netherlands, was also limited. This sampling bias needs to be considered when drawing conclusions based on genomic data during early stages of an emerging disease outbreak. Without a representative and sizable selection of reference sequences, reliable phylogenetic analysis is difficult. Clustering and conclusions on the origin of viruses may change substantially when virus sequences of other geographical regions are added to the analysis. Moreover, global monitoring of the genetic diversity of the virus is essential to reliably model and predict the spread of the virus. Since early March, the number of publicly available genomes has grown considerably, and the geographic signature in the dataset is becoming increasingly clear. Since its emergence, the global spread of SARS-CoV-2 led to diversification into lineages that reflect ongoing chains of

transmission in specific geographic regions globally, in Europe, and—during the second and third phases—in the Netherlands. The average single nucleotide polymorphism distance between the sequenced viruses in our study was 7.39 and this diversification provided the basis for the use of WGS to investigate possible transmission chains locally (for instance, in health-care settings, where it can be used to inform infection control and prevention when combined with background data on contact histories among others). Moreover, the continued effort will lay the foundation for the enhanced surveillance that will be paramount during the next phase of the pandemic, when confinement measures will gradually be lifted and testing of people with mild symptoms is increased. Given the widespread circulation, the most likely scenario is that SARS-CoV-2 will (sporadically) re-emerge, and discrimination between novel introductions versus prolonged local circulation is important to inform appropriate public health decisions. In addition, owing to genomic mutations, the phenotype and the transmission dynamics of the virus might change over time. Therefore, close monitoring of the behavior of the virus in combination with genetic information is essential as well.

We have used an amplicon-based sequencing approach to monitor the emergence of SARS-CoV-2 in the Netherlands. A critical step in using amplicon-based sequencing is that close, reliable reference sequences need to be available. The primers are designed on the basis of our current knowledge about SARS-CoV-2 diversity and therefore need regular updating. In the future, this may be overcome using metagenomic sequencing. However, at the moment, conventional metagenomic sequencing (Illumina) takes too long for near to real-time sequencing, and nanopore-based metagenomic sequencing is not sensitive enough to allow recovery of whole-genome sequences in a similar fashion and with similar costs compared to amplicon-based nanopore sequencing.

We provide a description of the incursion of SARS-CoV-2 into the Netherlands. The combination of real-time WGS with the data from the National Public Health response team has provided information that helped decide on the next steps in the decision-making. Sharing of metadata is needed within a country but also on a global level. We urge countries to share sequence information to combine our efforts in understanding the spread of SARS-CoV-2. The Global Initiative on Sharing All Influenza Data (GISAID)[15,16] made sharing of sequence information coupled to limited metadata possible in a manner that protects the intellectual property and acknowledges the data providers. However, to fully capitalize on the potential added value of WGS for public health decision-making, systems for combined analysis of data are needed that are in agreement with general data protection rules. We previously developed a model for collaborative exploration of WGS and metadata in a protected sharing environment[17,18]. For truly global collaboration, such systems would need to be further developed and hosted under the auspices of the WHO (World Health Organization).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-020-0997-y.

Received: 21 April 2020; Accepted: 26 June 2020;
Published online: 16 July 2020

References

1. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
2. Gorbalenya, A. E. et al. The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).
3. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
4. Oude Munnink, B. B. et al. Genomic monitoring to understand the emergence and spread of Usutu virus in the Netherlands, 2016–2018. *Sci. Rep.* **10**, 2798 (2020).
5. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
6. Faria, N. R. et al. Zika virus in the Americas: early epidemiological and genetic findings. *Science* **352**, 345–349 (2016).
7. Faria, N. R. et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* **361**, 894–899 (2018).
8. Khoury, M. J. et al. From public health genomics to precision public health: a 20-year journey. *Genet. Med.* **20**, 574–582 (2018).
9. Armstrong, G. L. et al. Pathogen genomics in public health. *N. Engl. J. Med.* **381**, 2569–2580 (2019).
10. Polonsky, J. A. et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos. Trans. R. Soc. B* **374**, 20180276 (2019).
11. Gire, S. K. et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
12. Modjarrad, K. et al. Developing global norms for sharing data and results during public health emergencies. *PLoS Med.* **13**, e1001935 (2016).
13. Oude Munnink, B. B. et al. Towards high quality real-time whole genome sequencing during outbreaks using Usutu virus as example. *Infect. Genet. Evol.* **73**, 49–54 (2019).
14. Oude Munnink, B. B., Nieuwenhuijse, D. F., Sikkema, R. S. & Koopmans, M. Validating whole genome nanopore sequencing, using Usutu virus as an example. *J. Vis. Exp.* https://doi.org/10.3791/60906 (2020).
15. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).
16. Shu, Y. & McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data—from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
17. Amid, C., Pakseresht, N. & Silvester, N. The COMPARE data hubs. *Database* **2019**, baz136 (2019).
18. *Covid-19* (European Bioinformatics Institute; accessed 16 April 2020); https://www.ebi.ac.uk/covid-19

The Dutch-Covid-19 response team

G. J. A. P. M. Oudehuis[4], Janke Schinkel[5], Jan Kluytmans[6,7], Marjolein Kluytmans-van den Bergh[6,7], Wouter van den Bijllaardt[6], Robbert G. Berntvelsen[6], Miranda M. L. van Rijen[6], Peter Schneeberger[8], Suzan Pas[9], Bram M. Diederen[9], Anneke M. C. Bergmans[9], P. A. Verspui van der Eijk[10], Jaco J. Verweij[7], Anton G. N. Buiting[7], Roel Streefkerk[11], A. P. Aldenkamp[12], P. de Man[13], J. G. M. Koelemal[13], D. Ong[13], S. Paltansing[13], N. Veassen[13], Jacqueline Sleven[14], Leendert Bakker[15], Heinrich Brockhoff[15], Ariene Rietveld[16], Fred Slijkerman Megelink[17], James Cohen Stuart[17], Anne de Vries[18], Wil van der Reijden[18], A. Ros[18], Esther Lodder[19], Ellen Verspui-van der Eijk[20], Inge Huijskens[20],

E. M. Kraan[21], M. P. M. van der Linden[21], S. B. Debast[22], N. Al Naiemi[23], A. C. M. Kroes[24], Marjolein Damen[25], Sander Dinant[25], Sybren Lekkerkerk[25], Oscar Pontesilli[25], Pieter Smit[25], Carla van Tienen[25], P. C. R. Godschalk[26], Jorien van Pelt[27], Alewijn Ott[27], Charlie van der Weijden[28], Heiman Wertheim[29], Janette Rahamat-Langendoen[29], Johan Reimerink[30], Rogier Bodewes[30], Erwin Duizer[30], Bas van der Veer[30], Chantal Reusken[30], Suzanne Lutgens[31], Peter Schneeberger[31], Mirjam Hermans[31], P. Wever[31], A. Leenders[31], Henriette ter Waarbeek[32] and Christian Hoebe[32]

[4]Academic Hospital Maastricht, Maastricht, the Netherlands. [5]Amsterdam Medical Centre, Amsterdam, the Netherlands. [6]Amphia Hospital, Breda, the Netherlands. [7]Bernhoven Hospital, Uden, the Netherlands. [8]Bravis Ziekenhuis, Bergen op Zoom & Roosendaal, Roosendaal, the Netherlands. [9]Dienst Gezondheid & Jeugd Zuid-Holland Zuid, Dordrecht, the Netherlands. [10]Elisabeth-Tweesteden Hospital, Tilburg, the Netherlands. [11]RLM, Dordecht, the Netherlands. [12]Foundation PAMM, Eindhoven, the Netherlands. [13]Franciscus Gasthuis & Vlietland, Rotterdam, the Netherlands. [14]MHC Gooi & Vechtstreek, Bussum, the Netherlands. [15]MHC Haaglanden, The Hague, the Netherlands. [16]MHC Hart voor Brabant, Tilburg, the Netherlands. [17]MHC Holland Noorden, Alkmaar, the Netherlands. [18]MHC Kennemerland, Haarlem, the Netherlands. [19]MHC West-Brabant, Breda, the Netherlands. [20]MHC Zuid-Holland Zuid, Dordrecht, the Netherlands. [21]Ijsselland Hospital, Capelle aan den IJssel, the Netherlands. [22]Isala Hospital, Zwolle, the Netherlands. [23]Laboratorium Microbiologie Twente Achterhoek, Hengelo, the Netherlands. [24]Leids Universitair Medisch Centrum, Leiden, the Netherlands. [25]Maasstad Hospital, Rotterdam, the Netherlands. [26]Meander Medical Centre, Amersfoort, the Netherlands. [27]MHC Drente, Meppel, the Netherlands. [28]MHC Flevoland, Lelystad, the Netherlands. [29]Radboud University Medical Center, Nijmegen, the Netherlands. [30]Centre for Infectious Disease Control, Bilthoven, the Netherlands. [31]Foundation Jeroen Bosch Hospital's, Hertogenbosch, the Netherlands. [32]Nursing Home Maastricht, Maastricht, the Netherlands.

Methods

COVID-19 response. This study was carried out in liaison with the national outbreak response team. This team develops guidance on case-finding and containment, based on WHO and European Centre for Disease Prevention and Control recommendations and expert advice, as defined by the crisis and emergency response structure[19,20]. Diagnostics were initially performed on suspected cases with a recent travel history to China, but between 25 and 28 February also suspected cases with travel history to affected municipalities in Northern Italy were tested. Between 1 and 11 March, all suspected cases with travel history to all four provinces in Northern Italy were tested and after 11 March all suspected cases with travel history to Italy were tested. The sequencing effort was embedded in the stepwise response to the outbreak (Extended Data Fig. 2), which evolved from the initial testing of symptomatic travelers including the testing of symptomatic contacts (phase 1), followed by inclusion of routine testing of patients hospitalized with severe respiratory infections (phase 2), to inclusion of HCWs with a low-threshold case definition and testing to define the extent of suspected clusters (phase 3). Depending on the phase and clinical severity, initial contact with patients was established through public health physicians or nurses from the municipal health service (for travel-related cases, contacts of (hospitalized) cases, and patients belonging to risk groups). The different phases in this study were based on observations described in this manuscript. Ethical approval was not required for this study as only anonymous aggregated data were used, and no medical interventions were made on human individuals.

Contact tracing. On 29 January, COVID-19 was classified as a notifiable disease in group A in the Netherlands, with physicians and laboratories having to report any suspected and confirmed case to the Dutch public health services (PHS) by phone. On notification, the PHS initiates source identification and contact tracing, and performs risk assessments. In the early outbreak phase (containment), the PHS traced and informed all high- and low-risk contacts of cases with the aim to stop further transmission. For each case, epidemiological information such as demographic information, symptoms, date of onset of symptoms, travel history, contact information, suspected source, underlying disease and occupation were registered. People were asked to report their travel history for the past 14 days, including potential travel to several countries. Owing to the magnitude of the COVID-19 outbreak, this quickly became impracticable in severely affected regions, and the strategy shifted to registering only data on confirmed cases and informing their high-risk contacts (phase 2) with continued active case-finding in less affected regions. The PHS informed the national public health authority of the Netherlands (RIVM) about all laboratory-confirmed cases. There, a national case registry was kept in which a contact matrix was kept for the first 250 cases.

Sample selection. In the first phase, all samples were selected for sequencing, reflecting travel-associated cases and their contacts. In the second phase, priority was given to patients identified through enhanced case-finding by testing of hospitalized patients with severe acute respiratory infections and continued sequencing of new incursions. In the third phase, the epidemic started to expand exponentially, and sequencing was performed to continue to monitor the evolution of the outbreak. In line with the national testing policy, a substantial proportion of new cases sequenced were HCWs (20%).

SARS-CoV-2 diagnostics. Clinical specimens were collected and phocine distemper virus was added as an internal nucleic acid (NA) extraction control to the supernatant. Clinical specimens included oropharyngeal and nasopharyngeal swabs, bronchoalveolar lavage and sputum. Total NA was extracted from the supernatant using Roche MagNA Pure systems. The NA was screened for the presence of SARS-CoV-2 using real-time single-plex PCRs with reverse transcription for phocine distemper virus, for the SARS-CoV-2 RdRp gene and for the SARS-CoV-2 E gene as described by Corman et al.[21].

SARS-CoV-2 WGS. A SARS-CoV-2-specific multiplex PCR for nanopore sequencing was performed, similar to amplicon-based approaches as previously described[22]. In short, primers for 89 overlapping amplicons spanning the entire genome were designed using primal (http://primal.zibraproject.org/)[22]. The amplicon length was set to 500 base pairs with a 75-base-pair overlap between the different amplicons. The used concentrations and primer sequences are shown in Supplementary Table 1. The libraries were generated using the native barcode kits from Nanopore (EXP-NBD104, EXP-NBD114 and SQK-LSK109) and sequenced on a R9.4 flow cell multiplexing up to 24 samples per sequence run.

Sequence data analysis. The resulting raw sequence data were demultiplexed using qcat (https://github.com/nanoporetech/qcat) or Porechop (https://github.com/rrwick/Porechop). Primers were trimmed using cutadapt[23], after which a reference-based alignment was performed using minimap2[24] to the GISAID sequence EPI_ISL_412973. The run was monitored using RAMPART (https://artic-network.github.io/rampart/) and the analysis process was automated using snakemake[25], which was used to perform near to real-time analysis with new data every 10 min. The consensus genome was extracted and positions with a coverage <30 were replaced with an 'N' with a custom script using biopython and pysam (https://github.com/dnieuw/ENA_SARS_Cov2_nanopore). An overview of the success rate of the sequencing is shown in Supplementary Table 2. Mutations in the genome as compared to the GISAID sequence EPI_ISL_412973 were confirmed by manually checking the alignment. In addition, homopolymeric regions were manually checked and resolved by consulting reference genomes. The average single nucleotide polymorphism difference was determined using snp-dists (https://github.com/tseemann/snp-dists). Human reads were removed by mapping against the human genome (GCF_000001405.26), after which the demultiplexed sequence reads were uploaded to the COVID-19 data portal under the accession numbers ERR4164763–ERR4164952.

Phylogenetic analysis. All available full-length SARS-CoV-2 genomes were retrieved from GISAID on 22 March 2020 (Supplementary Table 3) and aligned with the Dutch SARS-CoV-2 sequences from this study using MUSCLE. Sequences with >10% 'N's were excluded. The alignment was manually checked for discrepancies, after which IQ-TREE[26] was used to perform a maximum-likelihood phylogenetic analysis under the GTR + F + I + G4 model as the best predicted model using the ultrafast bootstrap option with 1,000 replicates. The phylogenetic trees were visualized using custom python and baltic scripts (https://github.com/evogytis/baltic).

BEAST analysis. All available full-length SARS-CoV-2 genomes were retrieved from GISAID[15,16] on 18 March 2020 and downsampled to include only representative sequences from epidemiologically linked cases. Sequences lacking date information were also removed from the dataset. To assess the temporal signal within the data, a maximum-likelihood phylogeny was performed using IQTREE v1.6.8[27] and the root-to-tip divergence was visualized as a function of sample date using TempEst v1.5.1[28] (Extended Data Fig. 3). The correlation coefficient for the root-to-tip analysis was 0.53, which is adequate for subsequent Bayesian analysis as much of this noise is accounted for in the Bayesian model. Bayesian phylogenetic trees were estimated using BEAST v1.10.4[29,30] using an HKY nucleotide substitution model and a strict molecular clock[31]. The analysis was run for 100,000,000 states with an exponential growth prior. Every 10,000 states, trees and parameters were sampled. Log files were inspected in Tracer v1.7.1[32] and Tree annotator v1.10.0 was used to remove the burn-in from the tree files and to infer the maximum clade credibility tree. Reported statistics are shown in Supplementary Table 4. Baltic and custom python scripts (https://github.com/evogytis/baltic) were used to visualize the maximum clade credibility tree.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data produced in this study are available on the COVID-19 data portal under the accession numbers ERR4164763–ERR4164952 and on the GISAID portal under the accession numbers EPI_ISL_413564– EPI_ISL_413591, EPI_ISL_414423– EPI_ISL_414471, EPI_ISL_414529– EPI_ISL_414566 and EPI_ISL_415460–EPI_ISL_415535.

References

19. Kraaij – Dirkzwager, M. et al. Middle East respiratory syndrome coronavirus (MERS-CoV) infections in two returning travellers in the Netherlands, May 2014. *Eurosurveillance* **19**, 20817 (2014).
20. Timen, A. Response to imported case of Marburg hemorrhagic fever, the Netherlands. *Emerg. Infect. Dis.* **15**, 1171–1175 (2009).
21. Corman, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* **25**, 2000045 (2020).
22. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
23. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
24. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
25. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
26. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
27. Quang, B. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
28. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
29. Ayres, D. L. et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173 (2012).

30. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
31. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
32. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

Acknowledgements

We thank the following additional members of the Dutch-Covid-19 response team: staff of the Department of Communicable Disease Control, Public Health Service Amsterdam, MHC Amsterdam, Amsterdam; staff of the Department of Communicable Disease Control, Public Health Service Gelderland-Zuid, MHC Gelderland-Zuid, Tiel; the Laboratorium Medische Microbiologie of JBZ, MHC Hart voor Brabant, Tilburg; staff of the Department of Communicable Disease Control, Public Health Service Utrecht, MHC Regio Utrecht, Utrecht; staff of the Department of Communicable Disease Control, Public Health Service Rotterdam-Rijnmond, MHC Rotterdam-Rijnmond, Rotterdam; staff of the Department of Communicable Disease Control, Public Health Service Zaanstreek-Waterland, MHC Zaanstreek Waterland, Zaandam; staff of the Department of Communicable Disease Control, Public Health Service Brabant Zuidoost, MHC Brabant Zuidoost, Eindhoven. We gratefully acknowledge the originating laboratories, where specimens were first obtained, and the submitting laboratories, where sequence data were generated and submitted to the EpiFlu Database of the GISAID, on which this research is based. All contributors of data may be contacted directly via the GISAID website (http://platform.gisaid.org). We also acknowledge M. de Graaf for her expertise in the interpretation of the early phylogenetic analysis. B.B.O.M., R.S.S., D.F.N., R.M. and M.K. received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 874735 (VEO), 848096 (SHARP JA) and 101003589 (RECoVER) and C.S., B.B.O.M. and M.K. from the European Joint Programme One Health EJP under the grant agreement number 773830 (METASTAVA). Part of this work was funded by the NIAID/NIH contract HHSN272201400008C for T.B., P.L. and M.P.

Author contributions

B.B.O.M., R.S.S., Á.O'T. and M.K. wrote the manuscript, Á.O'T., M.S., M.H. and M.M. set up sample and data collection, B.B.O.M, A.v.d.L., I.C., M.P., P.L., S.v.N., T.B., C.S. and R.J.O generated sequence data, S.K.K., R.M., A.A.v.d.E. and C.G. were involved in sample and data collection, B.B.O.M., R.R.S., D.F.N., A.R., A.M., H.V., A.O., Á.O'T., J.v.D. and M.K. were involved in data analysis and interpretation, B.B.O.M., M.S., M.H., M.M., R.R.S., Á.O'T. and M.K. designed the study. All authors provided critical feedback.

Competing interests

The authors declare no competing interests.

Additional information

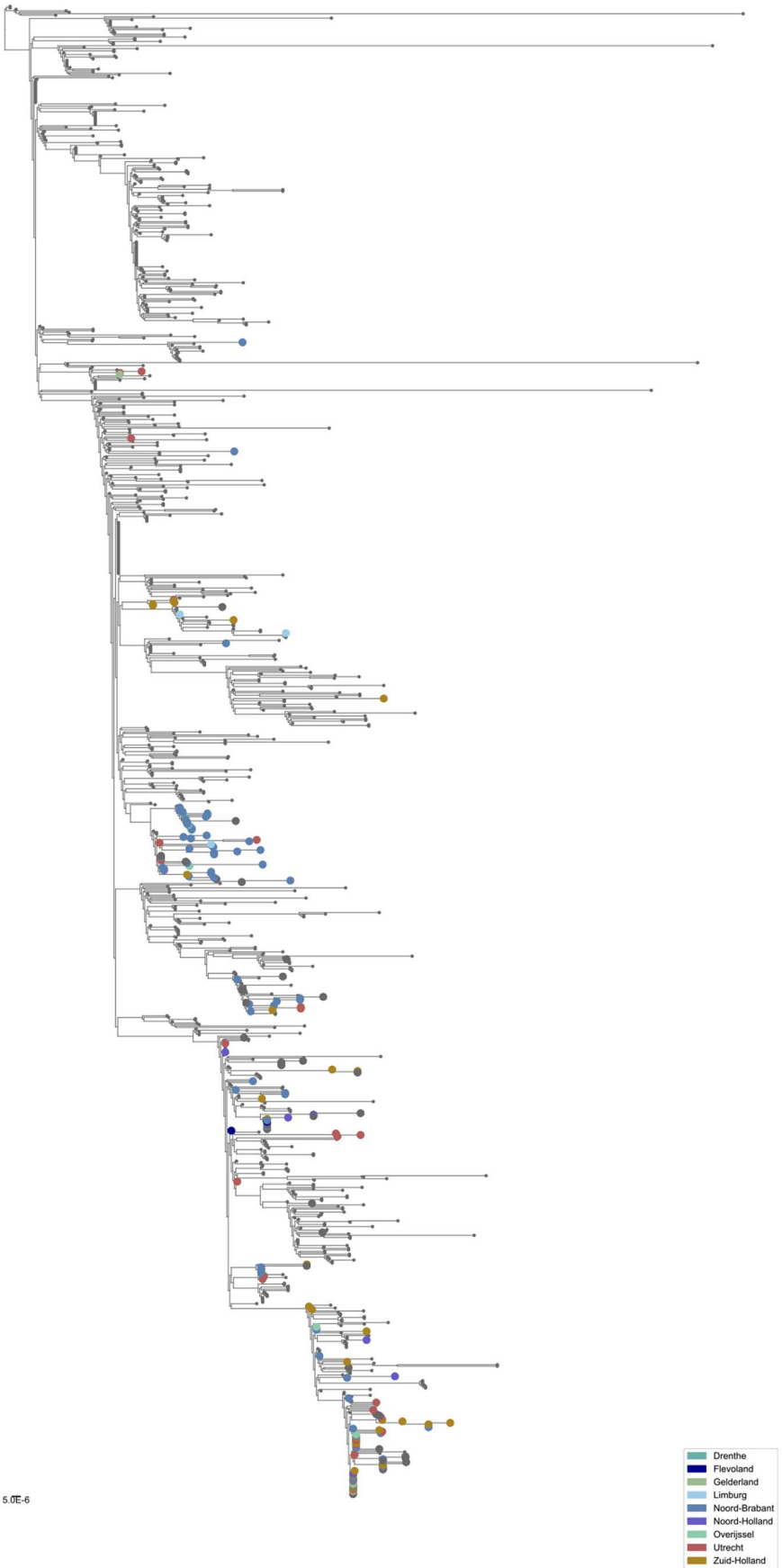
Extended data is available for this paper at https://doi.org/10.1038/s41591-020-0997-y.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-0997-y.

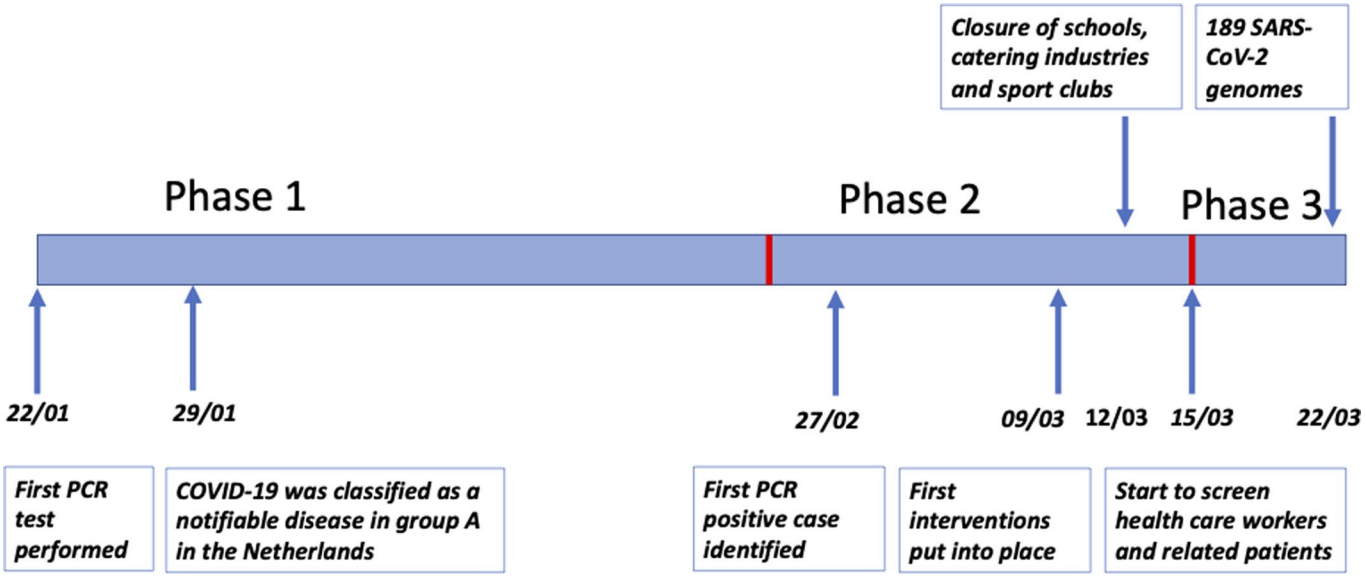
Correspondence and requests for materials should be addressed to M.K.

Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

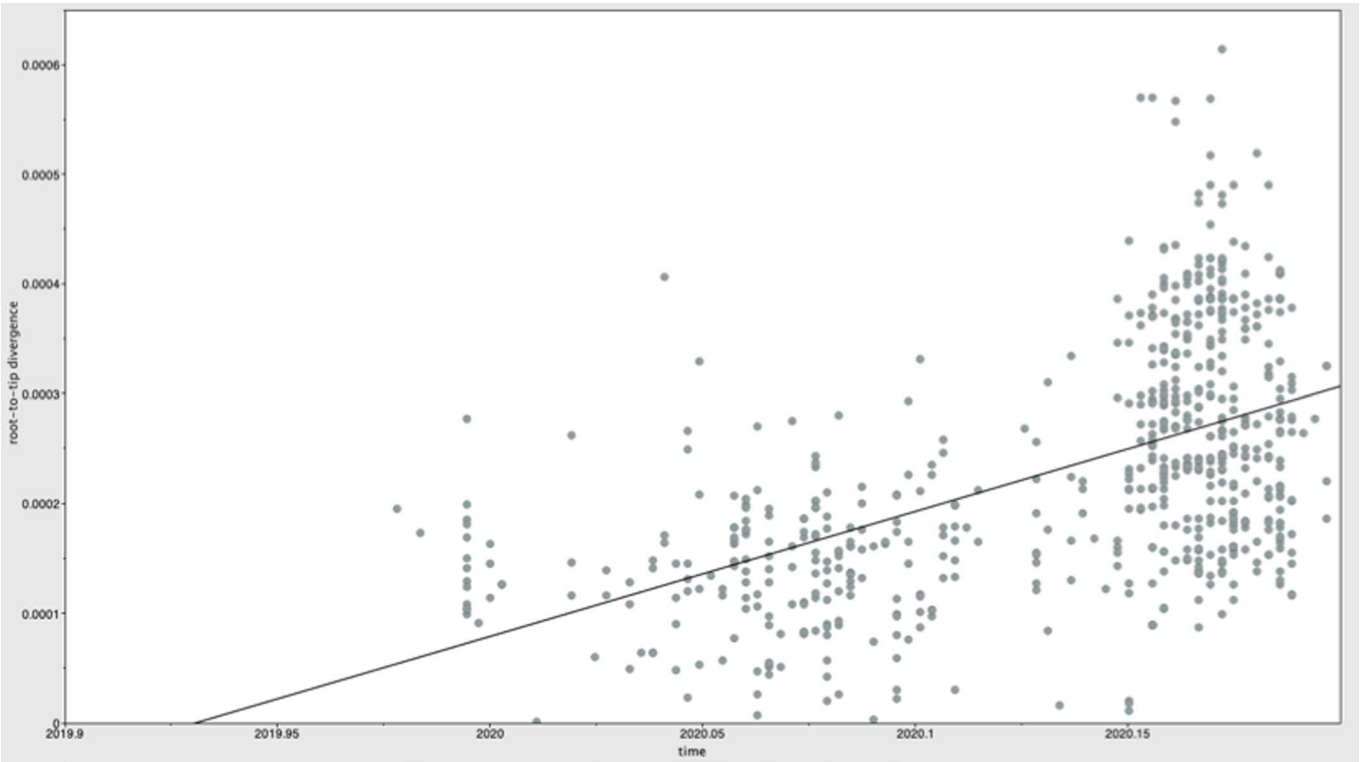
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Full maximum likelihood tree. Sequences are colored based on province of detection. Scale bar represent the number of substitutions per site.



Extended Data Fig. 2 | Timeline of the different phases. Graphical overview of the timeline of the of the different phases in the response to the SARS-CoV-2 outbreak in the Netherlands.



Extended Data Fig. 3 | Root-to-tip analysis. Report of the correlation coefficient for the root-to-tip divergence as a function of sample date.

nature research

Corresponding author(s): Bas Oude Munnink

Last updated by author(s): Apr 22, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software for data collection was used
Data analysis	primal (http://primal.zibraproject.org/) Python3.7 BEAST v1.10.4 MUSCLE v3.8.1551 IQ-TREE multicore version 2.0-rc2 IQTREE v1.6.8 qcat 1.1.0 porechop 0.2.4 RAMPART v1.0.5 Minimap2 2.17-r941 baltic scripts (https://github.com/evogytis/baltic) cutadapt 1.18 snakemake v5.4.5 snp-dists 0.7.0 TempEst v1.5.1 Tree annotator v1.10.0 Tracer v1.7.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data is openly available on GISDAID and on the COVID-19 data portal under the accession numbers ERR4164763-ERR4164952 and on GISAID under the accession numbers EPI_ISL_413564- EPI_ISL_413591, EPI_ISL_414423- EPI_ISL_414471, EPI_ISL_414529- EPI_ISL_414566 and EPI_ISL_415460- EPI_ISL_415535.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size was determined. We had to deal with the information available at the moment as described in the manuscript.
Data exclusions	No data was excluded from the analysis
Replication	The generation of whole genome sequences was not repeated. We have previously validated this sequence method using Usutu virus as described in the manuscript: https://pubmed.ncbi.nlm.nih.gov/31014969/ and https://pubmed.ncbi.nlm.nih.gov/32225162/.
Randomization	There was no randomization performed. In the first phase initial testing of symptomatic travelers according to the WHO and ECDC case definitions including the testing of symptomatic contacts was performed. In the second phase patients hospitalized with severe respiratory infections were also tested. In the third phase health care workers with a low threshold case definition were tested. Positive samples were subsequently sequenced.
Blinding	Blinding was not relevant in this study since we only used positive SARS-CoV-2 material

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging