

1 **Low and high infection dose transmissions of SARS-CoV-2 in the first COVID-19 clusters**
2 **in Northern Germany**

3 Susanne Pfefferle^{1*}, Thomas Günther^{3*}, Robin Kobbe^{2*}, Manja Czech-Sioli¹, Dominic Nörz¹, René
4 Santer², Jun Oh², Stefan Kluge⁴, Lisa Oestereich⁵, Kersten Peldschus⁶, Daniela Indenbirken³, Jiabin
5 Huang¹, Adam Grundhoff³, Martin Aepfelbacher^{1#}, Johannes K. Knobloch^{1#}, Marc Lütgehetmann^{1#},
6 Nicole Fischer^{1#}

7 ¹Institute for Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg-
8 Eppendorf, Hamburg, Germany, ²Department of Pediatrics, University Medical Center Eppendorf,
9 Hamburg, Germany, ³Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Virus
10 Genomics, Hamburg, Germany, ⁴Department of Intensive Care Medicine, University Medical Center
11 Hamburg-Eppendorf, Hamburg, Germany, ⁵Bernhard Nocht Institute, Leibniz Institute for Tropical
12 Medicine, Hamburg, Germany, ⁶Department of Diagnostic and Interventional Radiology and Nuclear
13 Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

14 *These authors contributed equally to this work.

15 #corresponding author: Nicole Fischer,
16 Institute for Medical Microbiology, Virology and Hygiene
17 University Medical Center Hamburg-Eppendorf
18 Martinistrasse 52
19 20246 Hamburg, Germany
20 Phone: +49-40-7410- 55171
21 Fax: +49-40-7410-53250
22 Email: nfischer@uke.de

23 Co-corresponding: Martin Aepfelbacher, m.aepfelbacher@uke.de; Johannes K. Knobloch,
24 j.knobloch@uke.de; Marc Luetgehetmann, mluetgehetmann@uke.de

25 **Keywords:** SARS-CoV-2 infection cluster, viral genomics, metagenomics, viral variants,

26 **2477 words (introduction, results and discussion)**

27 **Running Title:** Intra-host variant transmission of SARS-CoV-2

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

28 **Abstract**

29 **Objectives:** We used viral genomics to deeply analyze the first SARS-CoV-2 infection clusters in the
30 metropolitan region of Hamburg, Germany. Epidemiological analysis and contact tracing together with
31 a thorough investigation of virus variant patterns revealed low and high infection dose transmissions to
32 be involved in transmission events.

33 **Methods:** Infection control measures were applied to follow up contract tracing. Metagenomic RNA-
34 and SARS-CoV-2 amplicon sequencing was performed from 25 clinical samples for sequence analysis
35 and variant calling.

36 **Results:** The index patient acquired SARS-CoV-2 in Italy and after his return to Hamburg transmitted
37 it to 2 out of 132 contacts. Virus genomics and variant pattern clearly confirms the initial local cluster.
38 We identify frequent single nucleotide polymorphisms at positions 241, 3037, 14408, 23403 and 28881
39 previously described in Italian sequences and now considered as one major genotype in Europe. While
40 the index patient showed a single nucleotide polymorphism only one variant was transmitted to the
41 recipients. Different to the initial cluster, we observed in household clusters occurring at the time in
42 Hamburg also intra-host viral species transmission events.

43 **Conclusions:** SARS-CoV-2 variant tracing highlights both, low infection dose transmissions suggestive
44 of fomites as route of infection in the initial cluster and high and low infection dose transmissions in
45 family clusters indicative of fomites and droplets as infection routes. This suggests (1) single viral
46 particle infection can be sufficient to initiate SARS-CoV-2 infection and (2) household/family members
47 are exposed to high virus loads and therefore have a high risk to acquire SARS-CoV-2.

48

49

50

51

52

53

54

55

56 **Introduction:**

57 SARS-CoV-2, first emerged in late 2019 as the alleged cause of a cluster of viral pneumonia cases in
58 Wuhan (1). The virus and its associated disease, COVID-19 have since given rise to a worldwide
59 pandemic. Significant findings addressing epidemiology, containment measures, transmission dynamics
60 and course of the disease have been published worldwide in short intervals. However, questions with
61 regard to the infectivity of the virus, how many viral particles are sufficient for infection, are droplets
62 or fomites the predominant transmission routes, still need to be addressed in more detail.

63 Viral sequence data were immediately publically available and online tools such as GISAID provided
64 graphical interfaces for phylogenetic network analyses (1-6). Viral genomics can further be used to
65 deepen our understanding of infection routes and viral sequence variability, which might be contributing
66 to the clinical course of the disease.

67 Previous reports suggest frequent human-to-human transmission together with a wide range of clinical
68 severity, including patients showing mild or no symptoms with at the same time high concentrations of
69 viral RNA detectable in respiratory samples (7-9).

70 Interestingly, a recent report describes the dynamics of viral shedding and transmissibility of SARS-
71 CoV-2 with more than 44% of transmission occurring prior to symptom onset and viral loads peaking
72 before or at the time of symptom onset (10). This implies that not only health care workers need adequate
73 protection due to more frequent contact with infected patients (with high viral load) but especially
74 household/family members are at high risk due to high exposure and close contact.

75 While the viral load effect on incidence of infection and severity of the disease is currently discussed,
76 however, infecting dose measurements are highly challenging and can only be adequately addressed in
77 animal models.

78 We here use viral genomics and variant calling to follow up viral transmission in the initial COVID-19
79 cluster in Hamburg and other small household clusters occurring at the time in travel returnees in
80 Hamburg. We show here for the first time using viral variant tracing that low infection dose and high
81 infection dose transmissions contribute to the spread of SARS-CoV-2. Overall by analysing 25 clinical
82 samples using viral genomics and variant analysis we find evidence for both, intra-host variant

83 transmission indicative of high viral dose transmissions by droplets and transmission of only one sub-
84 consensus variant indicative of a low dose infection event.

85

86 **Methods**

87 **Patients**

88 The index patient is a physician in his 60s with no underlying diseases. 10d past his initial SARS-CoV-
89 2 positive PCR test he was precautionary admitted to a hospital for 4 days where a chest-CT scan showed
90 characteristic ground-glass opacities (supplementary figure S1).

91 Patient 1, a male laboratory worker in his late 40s, with type II diabetes and high BMI. He developed
92 severe symptoms with dry cough, muscle pain, high fever and acute respiratory distress syndrome
93 (ARDS) with ICU admission and ECMO treatment.

94 Patient 2, a relative of patient 0 living in the same household, in her 60s, no comorbidities known,
95 developed a mild cough 3d after initial testing. She never developed other symptoms of COVID-19
96 disease (supplementary Tables S2-3).

97 Additional 20 oropharyngeal samples were collected between 1st and 8th of March 2020 during SARS-
98 CoV-2 outpatient testing at the University Medical Center Hamburg-Eppendorf, UKE, following the
99 Robert Koch Institute's, RKI, recommendations at that time (individuals with respiratory symptoms
100 returning from countries/regions defined as high risk regions and/or individuals with contacts to
101 positively tested persons).

102 The local ethics committee of the City of Hamburg approved the study (PV7306). All studies were
103 carried out in keeping with local legal and regulatory requirements.

104

105 **Epidemiological investigation**

106 Contact persons of the index patient and patient 1 were identified by the UKE infection prevention and
107 control team together with public health officials. Classification of contacts is summarized in
108 supplementary Table S1. All employees with any contact were tested for SARS-CoV-2 at the day of
109 confirmed infection of the index patient. Persons with positive SARS-CoV-2 PCR were sent to home
110 quarantine until the virus became undetectable and were closely monitored for symptom history.

111 Personnel with high-risk contacts were sent to home quarantine for 14 days and were tested at the end
112 of their quarantine. Employees with low risk contacts remained on duty and voluntary testing was
113 provided on day 7 after the last contact to the index patient.

114

115 **Detection of SARS-CoV-2 RNA via RT-PCR in clinical samples**

116 Oropharyngeal swabs were preserved in modified Amies medium (Copan E-swab, Copan, Italy).
117 Respiratory samples, serum, urine and stool samples were taken without additives, 1:1 mixed with
118 Roche PCR Media kit buffer (Roche, USA). Screening PCR for SARS-CoV-2 RNA was performed as
119 described previously (11, 12).

120

121 **Cell culture and virus isolation**

122 Vero E6 cells (ATCC® CRL-1586) were propagated in DMEM containing 3% FCS, 1% Penicillin/
123 Streptomycin, 1 % L-Glutamine, 1 % Sodium pyruvate, 1 % non-essential amino acids (Gibco/ Thermo
124 Fisher, Waltham, USA) under standard culture conditions. For virus isolation, cells were washed with
125 PBS prior to infection with the native clinical material. After adsorption for 1h at 37°C, fresh DMEM
126 was added. Virus replication was assessed by RT-PCR.

127

128 **Immunofluorescence assays**

129 Assays were performed as described before (13, 14).

130

131 **ELISA for detection of SARS-CoV-2 IgG**

132 A commercially available ELISA for detection of IgG against SARS-CoV-2 (Euroimmun, Lübeck,
133 Germany) was used on the Euroimmunalyzer I (Euroimmun, Lübeck, Germany).

134

135 **SARS-CoV-2 Amplicon Sequencing**

136 SARS-Cov-2 amplicon sequencing was based on the nCoV-2019 sequencing protocol from the Artic
137 network project (artic.network/ncov-2019) with the following modifications. Reverse transcription was
138 performed in the presence of 2.4µM random primer mix (hexamer and anchored-dT primer (dT23VN))

139 After RNase H digestion multiplex PCR was performed using the primer scheme version 3 with 196
140 primers (supplementary Table S4) generating 400bp amplicons. Adaptor ligation for Illumina
141 compatible multiplex sequencing was achieved by using the NEBNext DNA Ultra II Library Prep Kit
142 and the NEBNext Multiplex Oligos (96 Unique Dual Index Primer Pairs; New England Biolabs).
143 Multiplex sequencing was performed on Illumina MiSeq, 500cycle MiSeq v2 kit (Illumina).

144

145 **Unbiased metagenomic RNA sequencing**

146 NGS libraries were prepared from each sample using SMARTer Stranded Total RNA-Seq Kit v2 - Pico
147 Input Mammalian (Takara Bio Europe, Saint-Germain-en-Laye, France). Multiplex-sequencing was
148 performed on an Illumina NextSeq, 300 cycles, PE protocol.

149 The number of reads per sample is summarized in supplementary tables S5-S6. Samples were analyzed
150 using an in-house pathogen detection pipeline (15-18).

151

152 **Bioinformatic analysis of metagenomic - and amplicon sequencing**

153 Illumina paired end amplicon sequencing reads were merged using pear (19), filtered by amplicon length
154 and the presence of correct amplicon primer sequences at both ends and aligned to NC_045512.2 using
155 minimap2 (20) with default settings for short read alignment.

156 Variants were called using freebayes Bayesian haplotype caller v1.3.1 (21) with ploidy and haplotype
157 independent detection parameters to generate frequency-based calls for all variants passing input
158 thresholds (-K -F 0.01). Input thresholds were set to 10 variant supporting reads with a minimum base
159 quality of 30 (-C10 -q30). Only high confidence variants present in > 33% of reads within at least one
160 individual sample were included and annotated using ANNOVAR (22). Variants located in non-coding
161 regions or representing frameshift, stopgain or startloss were deleted. Data were clustered by hclust
162 clustering method ward.D2.

163 To detect subgenomic mRNAs reads containing 10 upstream bases and the 6 core element bases of the
164 transcription-regulating sequence (TRS-L) fused to the 10 bases following the TRSs of the subgenomic
165 mRNA bodies (TRS-B) of the individual subgenomic mRNAs were counted.

166 All metagenomic and amplicon based sequences used in this work (after removal of human sequences)
167 are publicly accessible from the European Nucleotide Archive with the study accession number
168 PRJEB38546.

169

170

171 **Results**

172 In February 2020 a physician (index patient) returned from Trentino, Italy, at this time, not defined as a
173 risk area for contracting SARS-CoV-2 (23). 2 days after his return, he developed symptoms of a cold,
174 stopped his duty and stayed in home quarantine together with a relative thereafter. He was laboratory
175 confirmed as SARS-CoV-2 positive (Figure 1).

176 During these 2 days patient 0 had contact with 131 persons at the workplace. Out of these, 33 were high
177 risk (category I) and 98 were low risk contacts (category II). These contacts and a relative of patient 0
178 were tested negative on February 28th. One high risk workplace contact (patient 1) who had two contacts
179 with Patient 0 developed symptoms 3d later and was subsequently tested positive. Similarly, a relative
180 of patient 0 (patient 2) was tested positive for SARS-CoV-2 at that time. All other high risk contacts,
181 retested at day 7 (n=17) and/or at day 14 (n=28) after the last contact with patient 0 stayed SARS-CoV-
182 2 negative. 40 low risk contacts were tested negative at day 7 after the last contact with patient 0.

183

184 **SARS-CoV-2 infection course in the patients belonging to the index cluster**

185 SARS-CoV-2 viral RNA was monitored by RT-PCR in respiratory specimens from patients 0 - 2 at
186 different time points of infection. High viral loads were observed in all three patients early in the
187 infection, independent of clinical symptoms (Figure 1, Tables S2-S3). In patients 0 and 2 the viral loads
188 declined 1000-fold and 200-fold, respectively, within a 5 day period following the first day of positivity
189 (Figure 1). In the critically ill patient 1, however, viral loads declined only 8-fold and remained high
190 over a prolonged time period. Furthermore, while on ECMO treatment, he also showed moderate viral
191 loads in the serum, suggesting a systemic viral infection (Table S3).

192 Interestingly, from an oropharyngeal swab of patient 2, who stayed asymptomatic, infectious SARS-
193 CoV-2 could be successfully propagated in cell culture (24) suggestive of active viral replication. This

194 is furthermore supported by our data obtained from RNA shotgun sequencing, in which we detect reads
195 spanning the SARS-CoV-2 leader sequence and regions of subgenomic RNAs (sgRNAs) (Table 1).

196 All three patients seroconverted (Figure 1, Table S3), within 10- 14 d after the first positive PCR report.

197

198 **SARS-CoV-2 variant calling allows geographic viral origin analysis and contact tracing**

199 Since we did not obtain high virus coverage by shotgun sequencing over the complete SARS-CoV-2

200 genome in all samples, we performed SARS-CoV-2 amplicon sequencing. This allowed us to thoroughly

201 characterize viral genomes, confirm the relationship of the viral sequences and identify possible

202 sequence variants. We obtained coverage of more than 98% of all SARS-CoV-2 genomes included

203 (n=25) with 23 samples showing >99.5% coverage, the maximum coverage of amplicon sequencing,

204 due to primer positioning at the very 5' and 3' ends. We identified 37 sequence variants, with 26 variants

205 not used for clade or genotype assignment and 9 variants not represented in the GISAID database (Figure

206 2). Overall, mutations within the nucleotide sequence were relatively rare. We find between 2 to 12

207 consensus level variants per sample and only few, 1-2, sub-consensus variants occurring in 30% of the

208 samples. Based on the observed SNPs we can clearly cluster the sequences in distinct variant patterns,

209 pattern I, II (II.1-II.4) and III (Figure 2). The clustering is driven by the presence of few SNPs, frequently

210 found in SARS-CoV-2 sequences. These were described e.g. in the leader sequence, position 241 co-

211 occurring with mutations within the nsp3 (nt 3037), RNAPol (nt 14408) and spike protein (nt 23403).

212 These co-occurring mutations have been recently described as one major SARS-CoV-2 variant

213 occurring in Europe and suggested to be defined as genotype II (6). The earliest representative of this

214 prevalent SNP pattern in GISAID (3, 25) is in an Italian sequence (Italy/CDG1/2020/412973) entered

215 on February 20th. We identified these mutations in one cluster of related infections from persons

216 returning from a weekend of clubbing in Berlin (Figure 2, pattern II.1). In addition to these 4 core

217 variants additional frequent variants are known and here described as sub-clusters of cluster II. Cluster

218 II.2 is defined by an additional frequent mutation of three consecutive bases in the nucleocapsid

219 phosphoprotein, nt 28,881 resulting in two aa changes. This SNP is described in approximately 25% of

220 sequences in GISAID with its first description on February 24th, Netherlands/Berlicum_1363564/2020.

221 Sequences isolated from our index patient, patient 1 and 2 are clustering in pattern II.2. The sequences

222 are highly related defined by an additional SNP at position 160, a variant not described in any sequence
223 currently deposited in any database (3, 25). Thus, the mutation at position 160 in combination with the
224 anamnestic data and the chronological development of test positivity in the three patients, proves that
225 patient 1 and -2 acquired SARS-CoV-2 from patient 0.

226 Interestingly, we identify a unique mutation acquired in patient 1, nt 2393 within orf1ab resulting in an
227 aa substitution Val to Phe at position 710 in the non-structural protein nsp2 (Table 2). The nsp2 protein
228 shows relatively little conservation across coronaviruses and is dispensable for replication *in vitro* (26).
229 However, recent reports suggest that nsp2 can modulate the host cell environment, including
230 transcriptional processes (27).

231 Cluster II.3 separates from cluster II.2 with the presence of one additional SNP at nt27046 within the
232 coding region of the membrane glycoprotein. Pattern II.4 is mainly defined by the SNPs at positions
233 241, 1059, 3037, 14408, 15380, 23403 and 25563. It is less frequent and has been first reported in a
234 sequence from France (France/HF1465/2020) on February 21st. Cluster III (rarely identified) is
235 characterized by two SNPs at positions 1440 and 2891. It was first reported in a sequence from Germany
236 (Germany/NRW/02-1/2020) on February 25th. Finally, pattern I is clearly distinct; this sequence shows
237 12 SNPs with six of them being previously, February 26th, reported in a sequence identified in Norway
238 (Norway/1380/2020), Figure 2B. The SNP at nt 11962 has been rarely reported with its first description
239 in the sequence Wuhan/HBCDC-HB-05/20/20 on January 18th. It was recently suggested as being
240 indicative of genotype I (6).

241

242 **Sub-consensus SARS CoV-2 variant calling reveals intra-host transmissions**

243 We used paired end amplicon sequencing data for variant calling and to define minority sequences
244 variants, so called sub-consensus viral populations. Of the 37 nucleotide positions with SNPs, we find
245 at 14 positions sub-consensus variants: leader region (nts 160 and 199), leader protein (nt 635), nsp2 (nt,
246 2658), nsp3 (nts 5183 and 5284), nsp6 (nts 11083, 11438), nsp10 (nt 13115), RNA polymerase (nt
247 14217), RNA helicase (nt 16726), S (nt 23939), M (nt 27046) and TM (nt 27900). Most of these variants
248 result in non-synonymous mutations with the exception of the variants at nts 5284, 13115 and 14217
249 causing synonymous mutations.

250 Interestingly, we identified three sequence clusters in which we can follow transmission events based
251 on the presence/absence of minority sequences at specific nucleotide positions. Within the first cluster
252 we detected patient 0 carrying a sub-consensus variation at position 160 in the leader region. The two
253 recipients, patient 1 and 2, both carry only one variant at position 160, in both cases G (Var) instead of
254 A (Ref). Considering the index patient showing a variant fraction of 0.5 at position 160 this is highly
255 indicative of low dose viral infection in both transmission events. Similarly, in another family cluster
256 (cluster II.4), patients P21, P22 and P48, we observed only one position with multiple variants, position
257 199. Observed variant fractions are highly suggestive of a low viral dose exposure from P22 or P48 to
258 P21. Interestingly, we observed the transmission of intra-host variants at position 199 from patient P22
259 to P48 (or vice versa) being indicative of high viral load exposure and transmission. Furthermore, we
260 observed the transmission of intra-host variants in a second household cluster, two families returning
261 from a joint vacation. P06, P09 and P10, which are members of one family, are showing intra-host
262 variants at position 11438 indicative of high infection dose exposure being involved in viral
263 transmission. While P09 shows a variant fraction of 0.5 at position 11438, P06 and P10 both have a
264 variant fraction of 0.2. While we cannot make any conclusions on the order of the infection events,
265 however, the presence of identical minority sequence variants in these three patients is highly indicative
266 of high-infection dose transmission between these family members or between the initial sources of the
267 infection and all three family members. Overall, the presence of identical minority sequence is highly
268 suggestive of high viral load exposure occurring in household/family environment.

269

270 **No clinically relevant viral or bacterial co-infections in respiratory samples**

271 We performed unbiased metagenomic RNA sequencing from respiratory specimens of the 25 patients
272 included to detect putative viral and bacterial co-infections (Tables 3, S4-S6).

273 Besides SARS-CoV-2 sequences, no other viral sequences known to be associated with respiratory
274 infections were identified. Furthermore, the identified bacterial sequences in all cases clearly reflected
275 a commensal bacterial flora typical for the oropharyngeal tract. We closely monitored the samples of
276 the index cluster in which we sampled longitudinal respiratory samples of the course of the infection.
277 While we did not observe significant overrepresentation of sequences indicative of bacterial co-

278 infections in the patients 0 or 2, an early throat swap sample from patient 1 exhibited a high proportion
279 of sequences assigned to *Moraxella catharralis* and *Staphylococcus aureus* (24.56% and 19.55%
280 relative abundance, respectively). However, tracheal secretion and BAL fluid from patient 1 at later time
281 points (under ICU treatment) did not find evidence for persistent bacterial co-infection with *Moraxella*
282 *catharrhalis* or any other pathogen known as a cause for respiratory infection (Table 3).

283

284

285 **Discussion**

286 This work describes the first SARS-CoV-2 positive patient in Northern Germany, documents the tracing
287 of all its contacts and describes the subsequent infection cluster. Thorough epidemiological, clinical,
288 virological and metagenomic analyses of the patients in this first cluster together with 20 SARS-CoV-2
289 samples from family and household clusters occurring subsequently in the same geographic area,
290 revealed a number of important insights into transmission, clinical course and molecular epidemiology
291 of SARS-CoV-2/COVID-19.

292 First, based on comparative SNP analysis using 20 SARS-CoV-2 sequences from our study and
293 sequences available in the GISAID database, we identify different virus variant patterns, which allow
294 us to conclude on transmission routes. Transmission of SARS-CoV-2 from the index patient occurred
295 only to two out of 132 contact persons with whom prolonged and unprotected interaction took place.

296 Although we detected mutations between different patient's samples the overall frequency was relatively
297 low. Our observation is different to a recent report in which high frequency of SARS-CoV-2 mutations
298 was described (28). Differences can be explained by the relatively conservative variant calling approach
299 in our analysis with variant frequency >33% being reported compared to 5% variant frequency as cut-
300 off in the recently published report. Similar to previous reports, we did not find any mutation hot-spot
301 genes in our analysis.

302 Within the initial cluster (patient 0, 1 and 2), we identify a single non-synonymous mutation gained in
303 the prevalent virus sequence recovered from patient 1. This mutation has not been described before and
304 results in an aa substitution at position 710 of the non-structural protein nsp2. While we have no
305 indication that this aa substitution has functional consequences for viral replication or host interaction,

306 it is interesting that nsp2 is dispensable *in vitro* and not very well conserved between coronaviruses.
307 Interestingly, mutations within nsp2 have been described (26, 29, 30) suggesting a more stable nsp2
308 protein and thereby contributing to higher pathogenicity (29).
309 Second, to our knowledge we describe the first report of intra-host variant transmissions of SARS-CoV-
310 2 in patients. We report low dose and high dose transmission of SARS-CoV-2 occurring in
311 family/household clusters. In our initial cluster, we observe only low dose viral transmission occurring
312 at the workplace and in the household, respectively. However, in two family clusters in the early
313 epidemic in Hamburg, we observe transmissions of minority sequence variants between household
314 members indicative of high viral load/droplet exposure. While we cannot draw any conclusion on
315 infection dose and disease outcome (e.g. since in the initial cluster the severity of clinical symptoms was
316 independent of viral loads early in infection), our study shows for the first time that in settings of
317 household clusters transmission of minority virus variants occur, which is indicative of high infection
318 dose transmissions by droplets. Our data suggest that (1) low infection dose is sufficient for establishing
319 a SARS-CoV-2 infection and (2) close contacts, household members are at high risk to SARS-CoV-2
320 infection similar to healthcare workers if not taking appropriate protection measurements.

321

322 **Acknowledgement**

323 We are grateful to all staff members at the University Medical Center Hamburg-Eppendorf supporting
324 the diagnosis and management of COVID-19 patients. We thank Svenja Reucher and Kerstin Reumann
325 for excellent technical support and Uwe Ganz for graphical illustration. We gratefully acknowledge the
326 authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database1
327 on which part of this research is based.

328

329 **Author information**

330 SP, RK, TG, AG, MA, JK, ML and NF designed the study. SP, TG, MC, DI, DN, NF performed
331 literature search; SP, RK, TG, RS, MA, JK, ML and NF wrote the manuscript; SP, RK, TG, MC, DN,
332 JO, SK, LO, KP, DI, JK, ML and NF collected the data. TG, AG and JH performed the bioinformatics
333 analysis; SP, TG, MC, DI, ML and NF generated the figures and tables.

334

335 **Declaration of interest**

336 We declare no competing of interest.

337

338

339 **References**

340 1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients
341 with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-33.

342 2. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2
343 genomes. *Proc Natl Acad Sci U S A.* 2020;117(17):9241-3.

344 3. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to
345 reality. *Euro Surveill.* 2017;22(13).

346 4. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of
347 SARS-CoV-2. *National Science Review.* 2020.

348 5. Uddin M, Mustafa F, Rizvi TA, Loney T, Suwaidi HA, Al-Marzouqi AHH, et al. SARS-CoV-
349 2/COVID-19: Viral Genomics, Epidemiology, Vaccines, and Therapeutic Interventions. *Viruses.*
350 2020;12(5).

351 6. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics.* 2020.

352 7. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan,
353 China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med.* 2020.

354 8. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper
355 Respiratory Specimens of Infected Patients. *N Engl J Med.* 2020.

356 9. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus
357 Disease 2019 in China. *N Engl J Med.* 2020.

358 10. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding
359 and transmissibility of COVID-19. *Nat Med.* 2020;26(5):672-5.

- 360 11. Pfefferle S, Reucher S, Norz D, Lutgehetmann M. Evaluation of a quantitative RT-PCR assay
361 for the detection of the emerging coronavirus SARS-CoV-2 using a high throughput system. *Euro*
362 *Surveill.* 2020;25(9).
- 363 12. Wolfel R, Corman V, Guggemos W, Seilmaier M, Zange S, Mueller M, et al. Virological
364 assessment of hospitalized cases of coronavirus disease 2019. *Nature.* 2020;epub ahead of print.
- 365 13. Pfefferle S, Kraehling V, Ditt V, Grywna K, Muhlberger E, Drosten C. Reverse genetic
366 characterization of the natural genomic deletion in SARS-Coronavirus strain Frankfurt-1 open reading
367 frame 7b reveals an attenuating function of the 7b protein in-vitro and in-vivo. *Virology.* 2009;6:131.
- 368 14. Pfefferle S, Schopf J, Kogl M, Friedel CC, Muller MA, Carbajo-Lozoya J, et al. The SARS-
369 coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors.
370 *PLoS Pathog.* 2011;7(10):e1002331.
- 371 15. Alawi M, Burkhardt L, Indenbirken D, Reumann K, Christopheit M, Kroger N, et al. DAMIAN:
372 an open source bioinformatics tool for fast, systematic and cohort based analysis of microorganisms in
373 diagnostic samples. *Sci Rep.* 2019;9(1):16841.
- 374 16. Fischer N, Indenbirken D, Meyer T, Lutgehetmann M, Lellek H, Spohn M, et al. Evaluation of
375 Unbiased Next-Generation Sequencing of RNA (RNA-seq) as a Diagnostic Method in Influenza Virus-
376 Positive Respiratory Samples. *J Clin Microbiol.* 2015;53(7):2238-50.
- 377 17. Fischer N, Rohde H, Indenbirken D, Gunther T, Reumann K, Lutgehetmann M, et al. Rapid
378 metagenomic diagnostics for suspected outbreak of severe pneumonia. *Emerg Infect Dis.*
379 2014;20(6):1072-5.
- 380 18. Gunther T, Haas L, Alawi M, Wohlsein P, Marks J, Grundhoff A, et al. Recovery of the first
381 full-length genome sequence of a parapoxvirus directly from a clinical sample. *Sci Rep.* 2017;7(1):3734.
- 382 19. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd
383 mergeR. *Bioinformatics.* 2014;30(5):614-20.
- 384 20. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.*
385 2018;34(18):3094-100.
- 386 21. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing 2012 [

- 387 22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from
388 high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
- 389 23. Institute RK. Daily situation report 03/04/2020.
- 390 24. Pfefferle S, Huang J, Nörz D, Indenbirken D, Oestereich L, Guenther S, et al. Complete genome
391 sequence of a SARS-CoV-2 strain isolated in Northern Germany *Microbiology Resource*
392 *Announcement.* accepted May 2020.
- 393 25. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to
394 global health. *Glob Chall.* 2017;1(1):33-46.
- 395 26. Graham RL, Sims AC, Brockway SM, Baric RS, Denison MR. The nsp2 replicase proteins of
396 murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral
397 replication. *J Virol.* 2005;79(21):13399-411.
- 398 27. Cornillez-Ty CT, Liao L, Yates JR, 3rd, Kuhn P, Buchmeier MJ. Severe acute respiratory
399 syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in
400 mitochondrial biogenesis and intracellular signaling. *J Virol.* 2009;83(19):10314-8.
- 401 28. Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic diversity of SARS-CoV-2 in
402 Coronavirus Disease 2019 patients. *Clin Infect Dis.* 2020.
- 403 29. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019:
404 The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol.* 2020.
- 405 30. Bal A, Destras G, Geymard A, Valette M, Esmeret V, Frobet E, et al. Molecular characterization
406 of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino-acid deletion in nsp2
407 (Asp268Del). *Clinical Microbiology and Infection.* in press.

408

409 **Figure Legends**

410 **Figure 1:** Overview of transmission events and course of the disease in the first COVID-19 cluster in
411 Northern Germany. SARS-CoV-2 RNA loads (black line, left y-axis) in respiratory samples and SARS-
412 CoV-2 IgG ELISA ratios (green line, right y-axis) from the index patient (centered panel), patient 1
413 (upper panel) and patient 2 (lower panel). Brown triangle indicates initial infection of the index patient
414 in Italy; low and high risk contacts of the index patient are shown as orange and red triangles,

415 respectively. Initial occurrence of symptoms is indicated by an arrow. ICU treatment of patient 1 is
416 labelled with a yellow rectangle. The dashed lines indicate threshold for reporting positive results of the
417 ELISA (= signal to noise ratio, $s/n \geq 1$).

418

419 **Figure 2: (A)** Clustering of viral variants of SARS-CoV-2 sequences recovered from the index patient,
420 patient 1, patient 2 and 19 SARS-CoV-2 sequences from respiratory swabs collected in the same time
421 period in comparison to the reference sequence, NC_045521. Nucleotide positions are indicated at the
422 bottom. Included are only variants with sufficient coverage (>10) and SNP being present in more than
423 33% of all reads in at least one sample. I-III summarizes sequence patterns as defined by SNPs. The
424 frequency of variants is indicated by the heat map ranging from grey (reference), yellow to dark blue
425 (variant). The quality score per individual site is indicated at the top. * indicates members within one
426 family. **(B)** Overview of individual SNPs defining pattern I, II (II.1-II.4) and III with sequence variation
427 in comparison to the reference sequence.

428

429

430 **Supplementary Data:**

431 Supplementary Figure S1: CT scan of the index patient

432 Supplementary Table S1: Category of contact to confirm COVID-19 cases

433 Supplementary Table S2: Clinical parameters obtained in longitudinal samples of P0 and P2.

434 Supplementary Table S3: Results obtained by microbiological diagnostics.

435 Supplementary Table S4: Primers used in multiplex PCR for Amplicon Sequencing

436 Supplementary Table S5: Summary of next generation sequencing data, P0, P1 and P2.

437 Supplementary Table S6: Viral loads and number of SARS-CoV-2 mapped reads and viral genome
438 coverage.

439 Supplementary Table S7: Summary of unbiased metagenomic RNA sequencing data.

440

441

442

443 **Tables:**

444 **Table 1:** Subgenomic RNA detection by shotgun RNA-Seq.

445

Sample	L ^a	S	ORF3a	E	M	ORF6	ORF7a	ORF8	N	Total sg ^b
P01	3	0	0	0	0	0	0	0	6	6
P03	0	0	0	0	0	0	0	0	0	0
P02	0	0	0	0	0	0	0	2	3	5
P06	7	0	4	2	2	0	6	0	14	28
P07	1	2	0	0	0	0	0	0	2	4
P05	0	0	0	0	0	0	0	0	0	0
P09	0	0	0	0	0	0	0	0	14	14
P10	0	0	0	0	0	0	0	0	3	3
P11	0	0	0	0	0	0	0	0	3	3
P04	0	0	0	16	0	0	8	0	3	27
P13	2	0	0	0	0	0	0	0	1	1
P15	0	0	0	0	0	0	0	0	0	0
P16	0	0	0	0	0	0	1	0	2	3
P18	0	0	0	0	0	0	0	0	15	15
P21	7	2	0	0	0	0	0	0	13	15
P22	11	0	0	0	0	1	1	0	13	15
P25	0	0	0	0	0	0	0	0	0	0
P30	0	0	0	0	0	0	0	0	0	0
P33	2	0	10	0	2	0	0	1	59	72
P35	0	0	0	0	5	0	0	0	1	6
P13	3	0	0	0	1	0	0	0	0	1
P39	4	0	0	0	0	0	1	0	20	21
P40	1	2	0	0	0	0	0	0	0	2
P48	8	0	0	0	0	0	1	0	5	6
P35	0	0	0	0	0	0	0	0	0	0
cell culture	340	24	6	10	13	24	18	8	87	190
total ^c	389	30	20	28	23	25	36	11	264	

446 Summary of subgenomic mRNA-detecting reads spanning the leader sequence fused to the genome region
 447 encoding for structural and accessory proteins.

448 ^a leader sequence within the genomic (+) mRNA ^b total number of reads indicating presence of all subgenomic
 449 mRNAs within each sample ^c total number of fragments indicating subgenomic mRNAs across all samples.

450

451

452

453 **Table 2: Nucleotide variants of SARS-CoV-2 sequences in comparison to the reference sequence**
 454 **NC_045512.2.**

Position [nt] (orf)	Index Patient	Patient 1		Patient 2
	d2*	d2*	d9*	d3*
160 (noncoding)	G/A	G/A	G/A	G/A
241 (noncoding)	C/T	C/T	C/T	C/T
2393 (orf1ab, Nsp2)	G	G/T (V710F)	G/T (V710F)	G
3037 (orf1a, Nsp3)	C/T (F924F)	C/T (F924F)	C/T (F924F)	C/T (F924F)
14408 (orf1ab)	C/T (P4715L)	C/T (P4715L)	C/T (P4715L)	C/T (P4715L)
23403 (S)	A/G D614G	A/G D614G	A/G D614G	A/G D614G
28881(orf9, N)	GGG/AAC (RG203/4KR)	GGG/AAC (RG203/4KR)	GGG/AAC (RG203/4KR)	GGG/AAC (RG203/4KR)

455 *time point after SARS-CoV-2 confirmation by PCR.

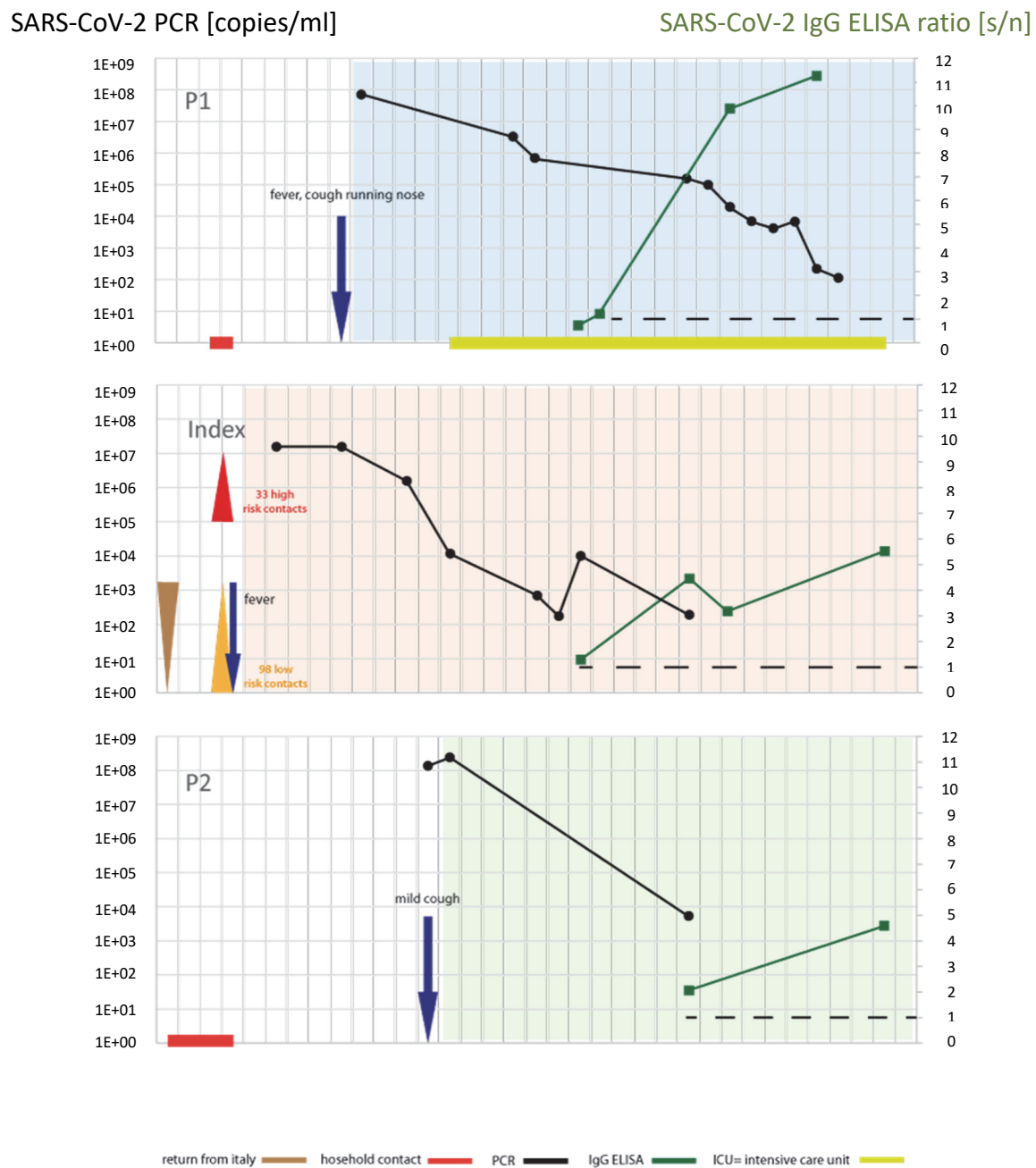
456 **Table 3: Unbiased metagenomic RNA sequencing describing possible co-infections.**

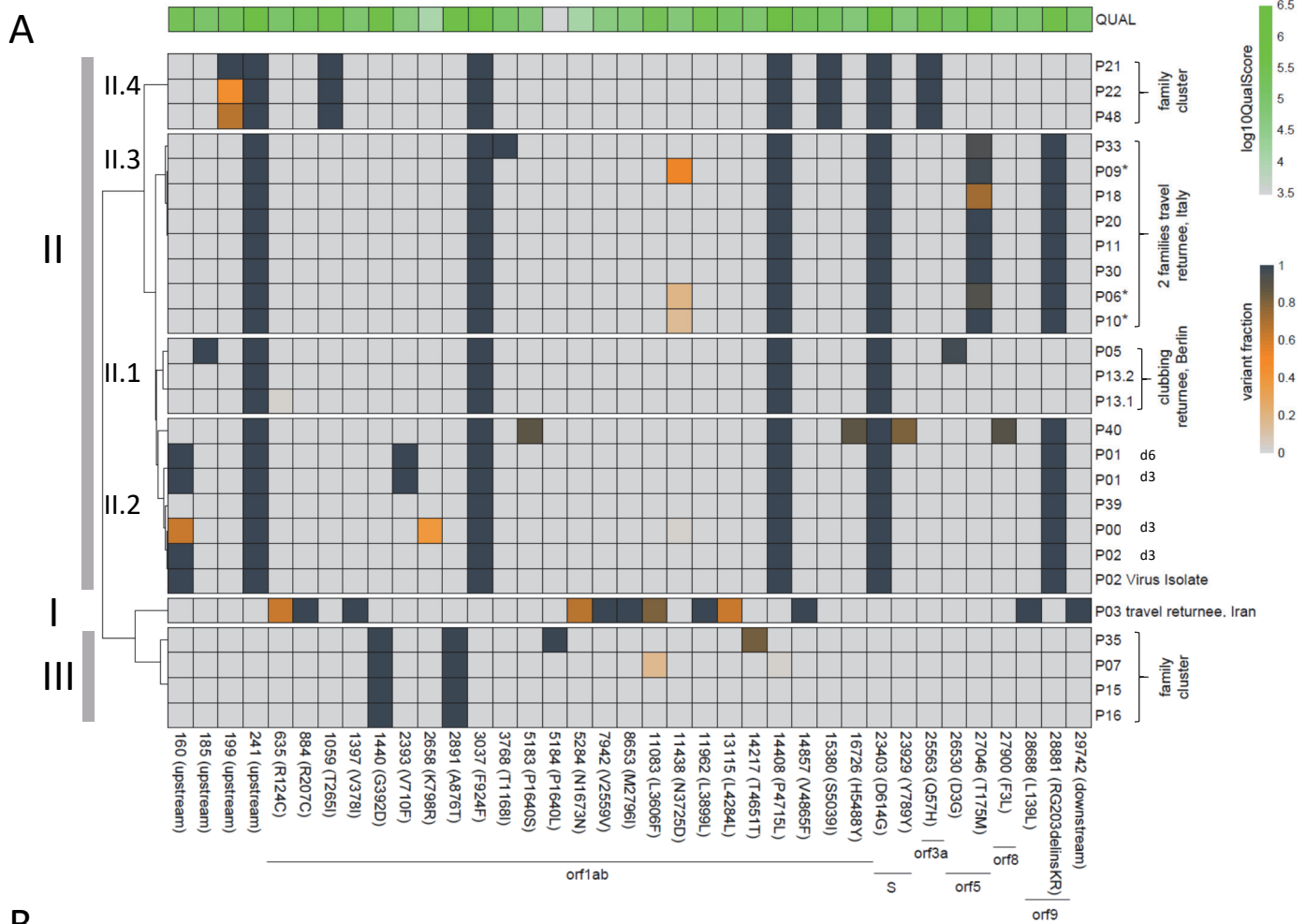
	specimen		SARS-CoV 2		bacteria	
			absolute abundance (# reads)	relative abundance (%)	relative abundance >3 %	
					species name	
Index patient	d1	swab	744	0.016	<i>uncultured bacterium</i>	19.6
					<i>Haemophilus parainfluenzae</i>	13.9
					<i>Fusobacterium pseudoperiodonticum</i>	9.57
					<i>Veillonella parvula</i>	8.51
					<i>Prevotella melinogenica</i>	6.66
					<i>Neisseria sublava</i>	5.7
	d4	swab	–	–	<i>Prevotella melinogenica</i>	29.3
					<i>uncultured bacterium</i>	27.4
					<i>Veillonella sp.</i>	8.5
					<i>Neisseria subflava</i>	5.5
Patient 1	d2	swab	2,782	2.1	<i>Moraxella catharralis</i>	24.56
					<i>Staphylococcus aureus</i>	19.55
					<i>Uncultured bacterium</i>	5.91
					<i>Neisseria subflava</i>	3.6
	d9	BAL	292	8.37	*	*
d10	TS	14	0.021	*	*	
Patient 2	d3	swab	129	0.011	<i>Prevotella melinogenica</i>	18.9
					<i>uncultured bacterium</i>	15.5
					<i>Veillonella atypica</i>	11.03
					<i>Prevotella oris</i>	5.3
					<i>Haemophilus hemaolyticus</i>	4.2
					<i>Prevotella sp.</i>	3.7
					<i>Cambylobacter concisus</i>	3.3

* ECMO related flora, no significant co-infections

457

Figure 1





cluster	nucleotide positions	REF	VAR	1 st description of variant pattern in GISAID	Prevalence based on GISAID entries	Genotype *
I	635; 884 ; 1397 ; 5284; 7942; 8653 ; 11083 ; 11962; 13115; 14857; 28688 ; 29742	C; C; G; C; T; G; G, C, C; G; T, G	T; T; A, T; C; T, T; T; T, T; C; T	02/26/2020 Norway/1380/2020	rare	I
II.1	241 ; 3037 ; 14408 ; 23403	C; C; C; A	T;T;T;G	02/20/20 Italy/CDG1	high	II
II.2	241 ; 3037 ; 14408 ; 23403 ; 28881	C;C;C;A; G	T;T;T;G ;A	02/24/2020 Netherlands/Berlicum_ 1363564/2020	medium	II
II.3	241 ; 3037 ; 14408 ; 23403 ; 27046 ; 28881	C;C;C;A; G; C	T;T;T;G ;T; A	02/24/2020 Netherlands/Berlicum_ 1363564/2020	rare	II
II.4	241 ; 1059 ; 3037 ; 14408 ; 23403 , 25563	C;C;C;C; G;A	T;T;T;T ;G;T	02/21/2020 France/HF1465/2020	low	II
III	1440 ; 2891	G;G	A; A	02/25/2020 Germany/NRW/02- 1/2020	rare	n.a.

Listed are nucleotides identified in all sequences of one cluster; nucleotide positions in bold indicate mutations identified in GISAID. REF: NC_045512.2; Rare: <10%; low 10-25%; medium 25-49%; high >50%; n.a.: not assigned; * Yin; C. Genomics 2020.