



## Letter to the Editor

## Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del)

A. Bal<sup>1,2,3,4,\*</sup>, G. Destras<sup>1,2,3,\*</sup>, A. Gaymard<sup>1,2,3</sup>, M. Bouscambert-Duchamp<sup>1,2</sup>, M. Valette<sup>1,2</sup>, V. Escuret<sup>1,2,3</sup>, E. Frobert<sup>1,2,3</sup>, G. Billaud<sup>1,2</sup>, S. Trouillet-Assant<sup>3,4</sup>, V. Cheynet<sup>4</sup>, K. Brengel-Pesce<sup>4</sup>, F. Morfin<sup>1,2,3</sup>, B. Lina<sup>1,2,3</sup>, L. Josset<sup>1,2,3,\*</sup>

<sup>1</sup> Centre National de Référence des Virus Respiratoires, Hospices Civils de Lyon, Lyon, France

<sup>2</sup> Laboratoire de Virologie, Hôpital de La Croix-Rousse, Hospices Civils de Lyon, Lyon, France

<sup>3</sup> Université Lyon 1, Faculté de Médecine Lyon Est, CIRI, Inserm U1111 CNRS UMR5308, Virpath, Lyon, France

<sup>4</sup> Laboratoire Commun de Recherche Hospices Civils de Lyon-bioMérieux, Centre Hospitalier Lyon Sud, Pierre-Bénite, France

## ARTICLE INFO

## Article history:

Received 20 March 2020

Accepted 21 March 2020

Available online xxx

Editor: Laurent Kaiser

## To the Editor

In December 2019, a novel coronavirus emerged in China, causing outbreaks of pneumonia [1]. The virus was subsequently identified as a betacoronavirus and named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 is responsible for the coronavirus disease 2019 (COVID-19) pandemic which includes asymptomatic upper and lower respiratory tract infections. Among the first European cases of COVID-19, six were associated with a cluster of transmissions in the French Alps in late January 2020 [2]. The index case of this cluster travelled from Singapore to France and went back to the United Kingdom (UK) where he tested positive for SARS-CoV-2 on February 6th. Here, we aimed to investigate the French cases related to this cluster using meta-genomic next-generation sequencing (mNGS) analysis.

Of the six contact patients who tested positive for SARS-CoV-2, the three samples with the highest viral loads (assessed by RT-PCR targeting the RdRp gene) were selected for mNGS analysis [3]. One nasopharyngeal swab was collected from a patient with an upper respiratory tract infection on February 7th (sample #1, Ct = 31.3). The other two samples were collected from the same asymptomatic

patient on February 8th (sample #2, nasopharyngeal swab, Ct = 31.1) and 9th (sample #3, nasopharyngeal aspirate, Ct = 28.8).

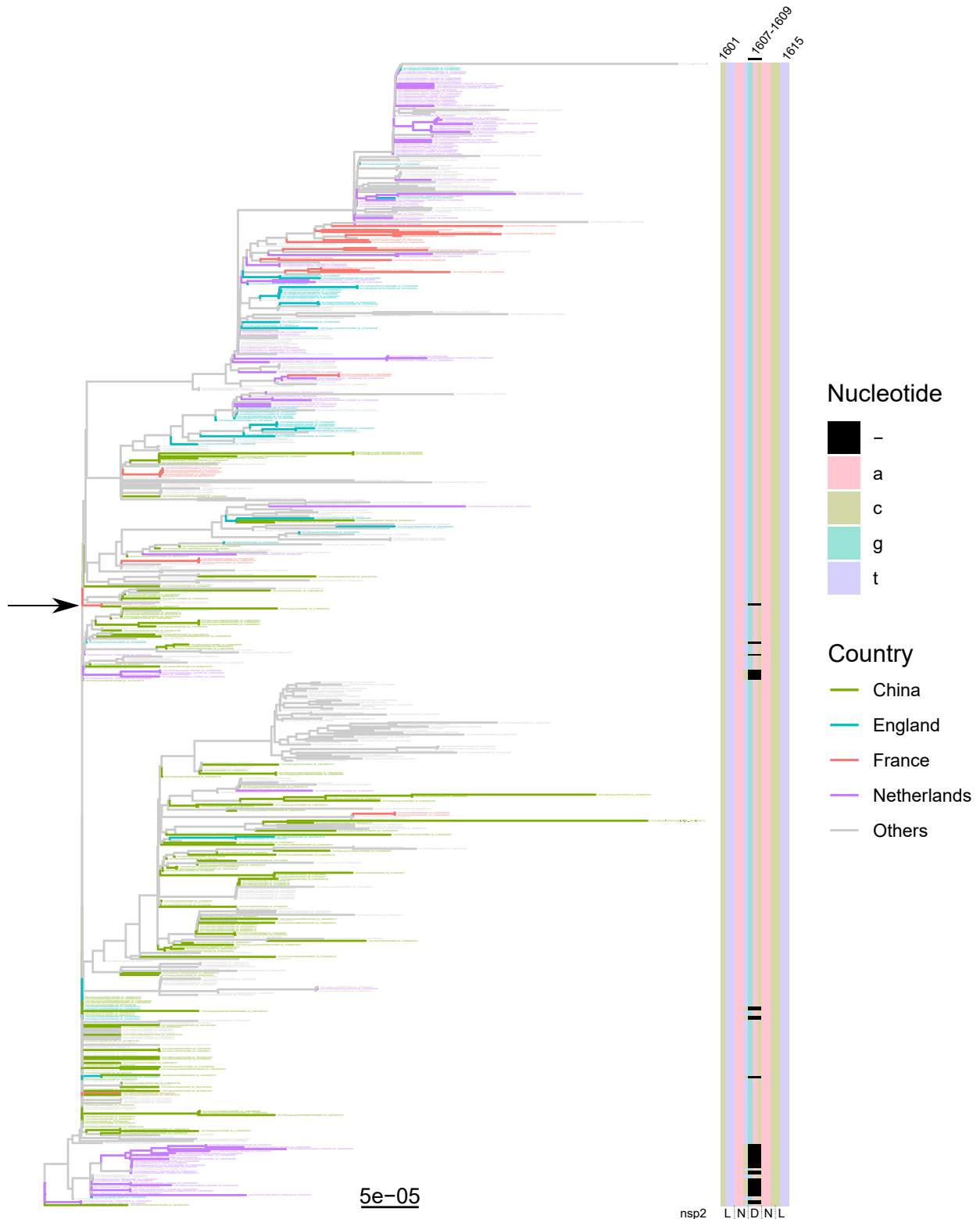
A previously described mNGS protocol was used, but DNase treatment was performed after nucleic acid extraction in order to increase the sensitivity for the detection of RNA viruses [4]. Low-quality and human reads were filtered out, and remaining reads were aligned to the SARS-CoV-2 reference genome (isolate Wuhan-Hu-1, EPI\_ISL\_402125) using the BWA-MEM algorithm. A mean of 19 445 767 reads per sample were generated, of which a mean of 605 243 reads per sample were mapped to the SARS-CoV-2 reference genome. The percentage of genome covered at a minimum depth of coverage of 100x was 38.3% for sample #1, 99.6% for sample #2 and 80.6% for sample #3.

The whole-genome sequence (WGS) generated from sample #2 was deposited on GISAID (Global Initiative on Sharing All Influenza Data) (EPI\_ISL\_410486). The phylogenetic analysis using the 571 WGS of SARS-CoV-2 publicly available (as of March 17th 2020) found that this sequence clustered with a sequence (EPI\_ISL\_408488) collected in Jiangsu, China, on January 19th, suggesting a separate introduction from Asia (Fig. 1). Compared to the reference SARS-CoV-2 sequence, a three-nucleotide deletion in open reading frame 1a (ORF1a) at positions 1607–1609 was identified. This deletion was found in 100% of the reads covering this position with a sequencing depth of 1745x around the deletion. Importantly, this deletion was also identified in 100% of the reads of sample #1 and sample #3 with a depth of 54x and 481x, respectively. Using the CoV-GLUE resource, we found that this mutation leads to a deletion of amino acid 268 in non-structural protein 2 (nsp2) [5]. This deletion in nsp2 (Asp268del) was also characterized in 37/571 (6.1%) of the WGSs available on March 17th (England  $n = 6$ ; The Netherlands  $n = 31$ ). WGS-based phylogenetic analysis found that 15 viruses containing this specific deletion were close to viruses collected in China between December 2019 and early February 2020, while 23 viruses with Asp268del collected in the Netherlands have slightly diverged (Fig. 1).

\* Corresponding author. L. Josset, Hospices Civils de Lyon, Centre National de Référence des Virus Respiratoires, 103 Grande-Rue de la Croix Rousse, 69317, Lyon, France.

E-mail address: [laurence.josset@chu-lyon.fr](mailto:laurence.josset@chu-lyon.fr) (L. Josset).

\* Equal contribution.



**Fig. 1.** Phylogenetic analysis based on whole-genome sequencing (WGS) of SARS-CoV-2. The analysis included 571 WGS of SARS-CoV-2 (>29 000 bp) collected in humans and available on GISAID (Global Initiative on Sharing All Influenza Data) from March 17th, 2020. The following sequences were excluded from the analysis—EPI\_ISL\_406592, EPI\_ISL\_414588, EPI\_ISL\_412900, EPI\_ISL\_408487, EPI\_ISL\_408483 and EPI\_ISL\_406595 because they were outliers, and EPI\_ISL\_413747, EPI\_ISL\_413695—because of incomplete sequences in ORF1ab. The hCoV19/Wuhan/IPBCAMSWH01/2019 strain was used as an outgroup virus. Genetic distances were calculated using the Kimura's two-parameter model (K80) and pairwise deletion. The tree was constructed by the neighbour-joining method using R seqinr and ggtree packages and validated using 1000 bootstrap pseudo-replicates. Sequence from sample #2 (EPI\_ISL\_410486) is indicated by the black arrow. Nucleotide alignment (1601–1615) is depicted as a heatmap on the right panel with the three-nucleotide deletion shown in black. Corresponding amino acid sequence (nsp2: 266–270) for the reference sequence is indicated below the heatmap.

SARS-CoV-2 sequences were not further compared between the two patients due to largely incomplete coverage of the SARS-CoV-2 genome in sample #1. Nonetheless, the longitudinal samples from the asymptomatic patient (sample #2 versus sample #3) were compared using a minimum depth of coverage of 100x in order to make a preliminary assessment of intra-host genetic variability. Three SNVs were noticed between the two samples: C366A (nsp1: S34Y), A20475G (synonymous mutation in nsp15), and T24084A (protein S: L841H), suggesting intra-host evolution of the virus. For all three positions, nucleotides from sample #2 were still detected in sample #3, but as minor variants.

In this short report, we present the first genetic characterization of a COVID-19 cluster in Europe. Despite low viral loads, the mNGS workflow used herein allowed us to characterize the whole-genome sequences of SARS-CoV-2 isolated from an asymptomatic patient in two clinical samples collected 1 day apart. Comparison of these sequences suggests viral evolution with development of quasiespecies. Specific studies using high depth of coverage are needed to explore potential intra-host adaptation. In addition, the present workflow identified a new deletion in nsp2 (Asp268del) which was found in all three samples originating from this cluster. The analysis of 571 WGS identified this deletion in 37 other viruses collected in England (February) and in The Netherlands (March), suggesting the spread of this deletion in Europe. The impact of Asp268del on SARS-CoV-2 transmission and pathogenicity, as well as on PCR performances and antiviral strategies, should be rapidly evaluated in further studies.

### Ethical considerations

Investigations complied with the General Data Protection Regulation (Regulation (EU) 2016/679 and Directive 95/46/EC) and the French data protection law (Law n°78–17 on 06/01/1978 and Décret n°2019-536 on 29/05/2019). Informed consent concerning

the disclosure of information relevant to this publication was obtained from the confirmed cases in France.

### Author contributions

AB and GD have contributed equally to this work.

### Acknowledgements

We would like to thank all the patients, clinicians, laboratory technicians and informatics department who contributed to this investigation. We are also grateful to Véréna Landel and Philip Robinson (DRCI, Hospices Civils de Lyon) for help in manuscript preparation. We thank the authors, the originating and submitting laboratories for their sequence and metadata shared through GISAID on which this research is based. We gratefully acknowledge all the members of CoV-GLUE, Next-strain.org, and virological.org for sharing their analysis in real time.

### References

- [1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33. <https://doi.org/10.1056/NEJMoa2001017>.
- [2] Spiteri G, Fielding J, Diercke M, Campese C, Enouf V, Gaymard A, et al. First cases of coronavirus disease 2019 (COVID-19) in the WHO European region, 24 January to 21 February 2020. *Euro Surveill* 2020;25. <https://doi.org/10.2807/1560-7917.ES.2020.25.9.2000178>.
- [3] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25. <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
- [4] Bal A, Pichon M, Picard C, Casalegno JS, Valette M, Schuffenecker I, et al. Quality control implementation for universal characterization of DNA and RNA viruses in clinical respiratory samples using single metagenomic next-generation sequencing workflow. *BMC Infect Dis* 2018;18:537. <https://doi.org/10.1186/s12879-018-3446-5>.
- [5] CoV-GLUE. Amino acid analysis for the SARS-CoV-2 outbreak. Available from: <http://cov-glue.cvr.gla.ac.uk>.