Full-length title:

Emergence in Southern France of a new SARS-CoV-2 variant of probably Cameroonian origin harbouring both substitutions N501Y and E484K in the spike protein

Short title (for the running head):

A new SARS-CoV-2 variant with spike substitutions N501Y and E484K

Author list: Philippe COLSON\textsuperscript{1,2,3}, Jérémy DELEURCE\textsuperscript{1}, Emilie BUREL\textsuperscript{1}, Jordan DAHAN\textsuperscript{4}, Agnès JOUFFRET\textsuperscript{5}, Florence FENOLLAR\textsuperscript{1,2,6}, Nouara YAH\textsuperscript{7}, Jacques FANTINI\textsuperscript{7}, Bernard LA SCOLA\textsuperscript{1,2,3}, Didier RAOULT\textsuperscript{1,2} *

Affiliations: \textsuperscript{1} IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; \textsuperscript{2} Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD), Microbes Evolution Phylogeny and Infections (MEPHI), 27 boulevard Jean Moulin, 13005 Marseille, France; \textsuperscript{3} Assistance Publique-Hôpitaux de Marseille (AP-HM), 264 rue Saint-Pierre, 13005 Marseille, France; \textsuperscript{4} Laboratoire de Biologie Médicale SYNLAB Provence Marseille, 25 rue Rabattu, 13015 Marseille, France; \textsuperscript{5} Laboratoire de Biologie Médicale SYNLAB Provence Forcalquier, rue du Souvenir Français, 04300 Forcalquier, France; \textsuperscript{6} Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD), Vecteurs – Infections Tropicales et Méditerranéennes (VITROME), 27 boulevard Jean Moulin, 13005 Marseille, France; \textsuperscript{7} Aix-Marseille Université, INSERM UMR S 1072, 51 boulevard Pierre Dramard, 13015 Marseille, France.

* Corresponding author: Didier Raoult, IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email:

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Keywords: SARS-CoV-2, variant, emergence, travel, southern France

Word counts: abstract, 198; text, 1,778

Figure: 1; Table: 1 References: 23

Supplementary Material: Figure: 1; Tables: 2; References: 2
SARS-CoV-2 variants have become a major virological, epidemiological and clinical concern, particularly with regard to the risk of escape from vaccine-induced immunity. Here we describe the emergence of a new variant. For twelve SARS-CoV-positive patients living in the same geographical area of southeastern France, qPCR testing that screen for variant-associated mutations showed an atypical combination. The index case returned from a travel in Cameroon. The genomes were obtained by next-generation sequencing with Oxford Nanopore Technologies on GridION instruments within ≈8 h. Their analysis revealed 46 mutations and 37 deletions resulting in 30 amino acid substitutions and 12 deletions. Fourteen amino acid substitutions, including N501Y and E484K, and 9 deletions are located in the spike protein. This genotype pattern led to create a new Pangolin lineage named B.1.640.2, which is a phylogenetic sister group to the old B.1.640 lineage renamed B.1.640.1. Both lineages differ by 25 nucleotide substitutions and 33 deletions. The mutation set and phylogenetic position of the genomes obtained here indicate based on our previous definition a new variant we named “IHU”. These data are another example of the unpredictability of the emergence of SARS-CoV-2 variants, and of their introduction in a given geographical area from abroad.
SARS-CoV-2 has emerged in China in December 2019 and has been declared a pandemic 21 months ago [1]. We have shown since the summer of 2020 that several SARS-CoV-2 variants have emerged in our geographical area and caused distinct epidemics, either successive or superimposed [2,3]. In addition, we described that the origin of these variants was often their introduction from abroad but could also be mink. This was observed by genotyping, as of 09/12/2021, SARS-CoV-2 from almost 40,000 patients using next-generation sequencing (NGS) of complete genomes for more than 22,000 patients and implementing multiple qPCR specific of each variant for a more exhaustive assessment of their spread. Since then and with the emergence of the Alpha variant at the end of 2020, SARS-CoV-2 variants have become a major virological, epidemiological, and clinical concern, particularly with regard to the risk of escape from vaccine-induced immunity [4-7]. Here we describe the emergence in south-eastern France of a new variant of probably Cameroonian origin.

The index case was an adult first diagnosed as infected with SARS-CoV-2 by real-time reverse transcription PCR (qPCR) performed in a private medical biology laboratory on a nasopharyngeal sample collected mid-November 2021 (Table 1). He was vaccinated against SARS-CoV-2 and returned from a travel to Cameroon three days before. He developed mild respiratory symptoms the day before diagnosis. He lives in a small town of southeastern France. Subsequent detection by qPCR of three mutations in the spike gene to screen for variants, as systematically performed in France in case of SARS-CoV-2 positivity, revealed an atypical combination with L452R-negativity, E484K-positivity, and E484Q-negativity (Pentaplex assay, ID Solution, France) that did not correspond to the pattern of the Delta
variant involved in almost all SARS-CoV-2 infections at that time (Table 1). Respiratory samples collected from seven other SARS-CoV-2-positive patients living in the same geographical area exhibited the same combination of mutations screened by qPCR. They were two adults and five children (<15 years of age) (Table 1). The respiratory samples from these eight patients were sent to university hospital institute Méditerranée Infection for SARS-CoV-2 genome sequencing as recommended by French public health authorities. A rapid NGS procedure was launched overnight. It allowed obtaining SARS-CoV-2 genotype identification in ≈8 hours. Briefly, viral RNA was extracted from 200 µL of nasopharyngeal swab fluid using the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer’s instructions. Extracted RNA was reverse-transcribed using SuperScript IV (Thermo Fisher Scientific) and cDNA second strand was synthesized with LunaScript RT SuperMix kit (New England Biolabs) then amplified using a multiplex PCR protocol according to the ARTIC procedure (https://artic.network/) with ARTIC nCoV-2019 V3 panel of primers (IDT, Coralville, IA, USA). Finally, NGS was performed with the ligation sequencing kit and a GridION instrument of Oxford Nanopore Technologies (Oxford, UK) following manufacturer’s instructions. Subsequently, fastq files were processed using the ARTIC field bioinformatics pipeline (https://github.com/artic-network/fieldbioinformatics). NGS reads were basecalled using Guppy (4.0.14) and aligned to the Wuhan-Hu-1 reference genome GenBank accession no. MN908947.3 using minimap2 (v2.17-r941) (https://github.com/lh3/minimap2) [8]. The ARTIC tool align_trim was used to softmask primers from read alignment and to cap sequencing depth at a maximum of 400. The identification of consensus-level variant candidates was performed using the Medaka (0.11.5) workflow developed by ARTIC (https://github.com/artic-network/artic-ncov2019). This strategy allowed assembling the complete genome from NGS reads obtained within 30 min of run for cycle threshold values (Ct) of qPCR comprised between 15 and 27. SARS-CoV-2
Genomes were classified into Nextclade and Pangolin lineages using web applications (https://clades.nextstrain.org/; https://cov-lineages.org/pangolin.html) [10, 11, 13]. They were deposited in the GISAID sequence database (https://www.gisaid.org/) [14] (Table 1). Phylogenies were reconstructed with the nextstrain/ncov tool (https://github.com/nextstrain/ncov) then visualized with Auspice (https://docs.nextstrain.org/projects/auspice/en/stable/).

The analysis of viral genomes revealed the presence of 46 nucleotide substitutions and 37 deletions, resulting in 30 amino acid substitutions and 12 deletions (Figure 1a; Supplementary Tables S1 and S2). Fourteen amino acid substitutions and 9 amino acid deletions are located in the spike protein. Substitutions N501Y and E484K are combined as in the Beta, Gamma, Theta and Omicron variants [5, 15]. Substitution F490S is present as in the Lambda variant, and substitution P681H is present as in the Lambda and Omicron variants. In other structural proteins than the spike, amino acid changes include two substitutions in the nucleocapsid protein and one in the membrane protein. In non-structural proteins, amino acid changes include one substitution in proteins Nsp2, Nsp3, Nsp4, Nsp6, Nsp12 (RNA-dependent RNA polymerase), and Nsp13 (helicase); two substitutions in Nsp14 (3’-5’ exonuclease); four substitutions in Nsp8 (which is part of the replication complex with Nsp7 and Nsp12); and three deletions in Nsp6. Finally, in regulatory proteins, amino acid changes include four substitutions in ORF3a, one in ORF9b and one in ORF8. In addition, codon 27 of ORF8 gene is changed into a stop codon, as in the Alpha variant [16]; some members of the Marseille-4 variant (B.1.160) that predominated in our geographical area between August 2020 and February 2021 also exhibit a stop codon in ORF8 gene but at another position [3].

Nextclade identified a 20A lineage. Pangolin identified a B.1.640 lineage in primary analysis but a B.1 lineage with the -usher (Ultrafast Sample placement on Existing tRee;...
https://genome.ucsc.edu/cgi-bin/hgPhyloPlace) option, which showed the phylogenetic placement of the genomes we obtained as an outgroup of the B.1.640 lineage and their clustering with a genome obtained late October in France (Ile-de-France) (EPI_ISL_5926666). The B.1.640 lineage corresponds to a variant first identified in France in April 2021, in Indonesia in August 2021, and in Republic of the Congo (Brazzaville) in September 2021, and it was involved in a cluster of cases in Brittany, France around mid-October 2021 [17]. As of 09/12/2021, 157 genomes were available from the GISAID database including 92 from France and 36 from the Republic of the Congo. The sets of spike mutations of the B.1.640 lineage and of genomes obtained here are similar, with 11 common nucleotide substitutions and 1 common deletion of 9 codons (Supplementary Figure S1, Tables S1-2). However, spike genes of both lineages differ by 7 mutations. In addition, 25 nucleotide substitutions and 33 nucleotide deletions located elsewhere in the genomes differ between the two genotypes. The pattern of mutations of present genomes hence indicates a new variant, which we named “IHU” (in reference to our institute), based on our previous definition [3]. Phylogeny performed with nextstrain/ncov tool (https://github.com/nextstrain/ncov) also showed that B.1.640 and IHU variants were most closely related between each other but comprised two divergent branches (Figure 1b). Their last common ancestor is estimated to date from January 2021 but no genome is currently available from GISAID that corresponds to it. Accordingly, a new Pangolin clade corresponding to the IHU variant was created on 07/12/2021 that was named B.1.640.2, the old B.1.640 clade being renamed B.1.640.1 (https://github.com/cov-lineages/pango-designation/issues/362). It encompasses present genomes and three other genomes comprising a sister group including the one recovered late October 2021 in France (Ile-de-France) (EPI_ISL_5926666) and two additional genomes obtained from samples collected late November in England (EPI_ISL_7181977) and Wales (EPI_ISL_7402094). As the index case was probably infected with the IHU variant during his
stay in Cameroon, we sought for this variant in GISAID among genomes from this country
but as of 09/12/2021 none of the 556 available genomes belong to the B.1.640.1 or B.1.640.2
lineages.

We analyzed a complete structure of the spike protein of the IHU variant generated by
incorporating its specific mutational profile to the original 20B SARS-CoV-2 (Wuhan-Hu-1
isolate with D614G substitution) [18] and fixing all gaps in the pdb file by incorporating the
missing amino acids with the Robetta protein structure prediction tool
[https://robetta.bakerlab.org/], followed by energy minimization with the Polak-Ribière
algorithm as previously reported (Figure 1c) [19]. In the N-terminal domain (NTD), the 134-
145 amino acid deletion is predicted to significantly affect the neutralizing epitope. Other
changes involve amino acids at positions 96 and 190: in Wuhan-Hu-1 isolate, E96 and R190
induce a turn in NTD secondary structure through electrostatic interactions between each
other. This interaction is conserved between substituted amino acids 96Q and 190S, which
suggests the co-evolution of these changes. In the receptor binding domain (RBD), aside the
well-known substitutions N501Y and E484K, several changes were predicted to significantly
affect the neutralizing epitopes. Particularly, P681H is located in the cleavage site of S1-S2
subunits of the spike and is observed in other variants including the recently emerging
Omicron [15]. Besides, D1139H substitution implies an amino acid involved in the fusion
between the virus and the infected cell. Also, D614G is combined with T859N in the IHU
variant. Interestingly, in the Wuhan-Hu-1 isolate, amino acids D614 and T859 from two
subunits of the trimeric spike are face to face and lock the trimer in a closed conformation.
Substitution D614G allows unlocking the trimer conformation, but this is predicted to be still
easier in case of additional presence of substitution T859N.
Respiratory samples collected until end of November 2021 from four other SARS-
CoV-2 positive patients living in the same city or borough than the index case could be
identified as containing the IHU variant by NGS within 24 hours after their reception (Table 1). All 12 IHU variant-positive samples showed the same combination of spike mutations as screened by real-time qPCR techniques: negativity for 452R and 484Q; positivity for 484K, 501Y [20], and 681H [3]. We also used the TaqPath COVID-19 kit (Thermo Fisher Scientific, Waltham, USA) that provided positive signals for all three genes targeted (ORF1, S, and N). Thus, the IHU variant can be distinguished by screening with qPCR assays from the Delta (L452R-positive) and Omicron (L452R-negative and negative for S gene detection by the TaqPath COVID-19 assay) variants that currently co-circulate in our geographical area.

Finally, scanning electron microscopy using a SUV 5000 microscope (Hitachi High-Techonologies Corporation, Tokyo, Japan) [21] allowed a quick visualization of the virus from a respiratory sample (Figure 1d).

Overall, these observations show once again the unpredictability of the emergence of new SARS-CoV-2 variants and their introduction from abroad, and they exemplify the difficulty to control such introduction and subsequent spread. They also warrant the implementation of genomic surveillance of SARS-CoV-2 that we started from the very beginning of the pandemic in our geographical area as soon as we diagnosed the first SARS-CoV-2 infection [21] and that we expanded during summer 2020 [2,3]. This surveillance has been implemented at the country scale in 2021 through the French Emergen consortium (https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/consortium-emergen). It is too early to speculate on virological, epidemiological or clinical features of this IHU variant based on these 12 cases. For this purpose, respiratory samples from infected patients were inoculated on Vero E6 cells as previously described [22] to be able assessing the sensibility to neutralization by anti-spike antibodies elicited by vaccine immunization, or by prior infection [23].
Acknowledgments

We are thankful to the Emergen French consortium (https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/consortium-emergen). We are also grateful to Laurence Thomas, Claudia Andrieu, Ludivine Brechard, Mamadou Beye, Marielle Bedotto, Elsa Prudent, Sofiane Bakour, Jacques Bou Khalil, and Clio Grimaldier for their technical help.

Author contributions

Conceived and designed the experiments: PC, DR, JF, BLS. Contributed materials/analysis tools: PC, JDe, EB, JDa, AJ, FF, NY, JF. Analyzed the data: PC, DR, BLS, JD, EB, JF, NY. Wrote the paper: PC, JF, DR. All authors approved the last version of the manuscript.

Funding

This work was supported by the French Government under the “Investments for the Future” program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03 and was also supported by Région Provence Alpes Côte d’Azur and European funding FEDER PRIMMI (Fonds Européen de Développement Régional-Plateformes de Recherche et d’Innovation Mutualisées Méditerranée Infection), FEDER PA 0000320 PRIMMI, and by Hitachi High-Technologies Corporation, Tokyo, Japan.

Conflicts of interest

DR has a conflict of interest being a consultant for Hitachi High-Technologies Corporation, Tokyo, Japan from 2018 to 2020. All other authors have no conflicts of interest to declare.
Funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Ethics
This study has been approved by the ethics committee of University Hospital Institute (IHU) Méditerranée Infection (N°2021-029). Access to the patients’ biological and registry data issued from the hospital information system was approved by the data protection committee of Assistance Publique-Hôpitaux de Marseille (APHM) and was recorded in the European General Data Protection Regulation registry under number RGPD/APHM 2019-73.
REFERENCES


resistant epitopes targeted by SARS-CoV-2 antibodies: A global consortium study.


FIGURE LEGENDS

Figure 1. Virological features and scanning electron microscopy image of the SARS-CoV-2 IHU variant

a: Map of the IHU variant genome showing amino acid substitutions and deletions.

b: Phylogeny reconstruction performed using the nextstrain/ncov tool (https://github.com/nextstrain/ncov) then visualized with Auspice (https://docs.nextstrain.org/projects/auspice/en/stable/). The genome of the original Wuhan-Hu-1 SARS-CoV-2 isolate (GenBank accession no. NC_045512.2) was added as outgroup, in addition to SARS-CoV-2 genomes of Pangolin lineages B.1.640.1 and B.1.640.2. X-axis shows time.

c: Representations of the spike of the IHU variant showing the location of all its amino acid substitutions. N-terminal domain (NTD) mutations are in blue; receptor binding domain (RBD) mutations are in red; mutations involved in ACE-2 unmasking are in yellow; mutations at S1-S2 cleavage site are in green; mutations at fusion region are in cyan.

d: Scanning electron microscopy image obtained using a SUV 5000 microscope from a respiratory sample positive for the SARS-CoV-2 IHU variant (Hitachi High-Technologies Corporation, Tokyo, Japan).
Table 1. Main epidemiological and virological features of cases identified with infection with the SARS-CoV-2 IHU variant

<table>
<thead>
<tr>
<th>Case no.</th>
<th>Age</th>
<th>Diagnostic qPCR Ct</th>
<th>Results of qPCR used to screen for the presence of SARS-CoV-2 spike substitutions</th>
<th>Results of the TaqPath COVID-19 qPCR assay (Targets: ORF1, S, and N genes)</th>
<th>Genome GISAID Id.</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Adult</td>
<td>15</td>
<td>L452R-neg.; E484K-pos.; E484Q-neg.; N501Y-n.t.; P681H- n.t.</td>
<td>Pos. for all three genes</td>
<td>EPI_ISL_7552465</td>
</tr>
<tr>
<td>10</td>
<td>Adult</td>
<td>16</td>
<td>L452R-neg.; E484K-pos.; E484Q-neg.; N501Y-n.t.; P681H- n.t.</td>
<td>Pos. for all three genes</td>
<td>EPI_ISL_7552470</td>
</tr>
<tr>
<td>11</td>
<td>Adult</td>
<td>22</td>
<td>L452R-neg.; E484K-pos.; E484Q-neg.; N501Y-n.t.; P681H- n.t.</td>
<td>Pos. for all three genes</td>
<td>EPI_ISL_7552483</td>
</tr>
<tr>
<td>12</td>
<td>Adult</td>
<td>15</td>
<td>L452R-neg.; E484K-pos.; E484Q-neg.; N501Y-n.t.; P681H- n.t.</td>
<td>Pos. for all three genes</td>
<td>EPI_ISL_7601710</td>
</tr>
</tbody>
</table>

* Index case; travelled to Cameroon

Ct, cycle threshold; Id., identifier; neg., negative; N, nucleocapsid; no., number; n.t., not tested; ORF1, open reading frame 1; pos., positive; qPCR, real-time reverse-transcription PCR; S, spike.

All 12 respiratory samples were collected between mid and end of November 2021.
SUPPLEMENTARY MATERIAL

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. Microarray showing the distribution along the SARS-CoV-2 genome and in viral genes of nucleotide changes observed in comparison with the genome of the Wuhan-Hu-1 isolate for the Pangolin B.1.640.1 and B.1.640.2 (IHU variant) lineages.

Genomes were analyzed using the Nextstrain web-tool (https://clades.nextstrain.org/) [1,2]. Representation is adapted from Nextclade sequence analysis web application output (https://clades.nextstrain.org/).
Supplementary Table S1. Comparison of nucleotide mutational patterns of the B.1.640.1 and B.1.640.2 lineages

<table>
<thead>
<tr>
<th>B.1.640.1</th>
<th>B.1.640.2 (IHU variant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C241T</td>
<td>C241T</td>
</tr>
<tr>
<td>G601T</td>
<td>-</td>
</tr>
<tr>
<td>C1191T</td>
<td>C1191T</td>
</tr>
<tr>
<td>A1620G</td>
<td>-</td>
</tr>
<tr>
<td>C2416T</td>
<td>C2416T</td>
</tr>
<tr>
<td>C2455T</td>
<td>C2455T</td>
</tr>
<tr>
<td>C3037T</td>
<td>C3037T</td>
</tr>
<tr>
<td>G6622T</td>
<td>-</td>
</tr>
<tr>
<td>G7328T</td>
<td>-</td>
</tr>
<tr>
<td>C9711T</td>
<td>-</td>
</tr>
<tr>
<td>G9756A</td>
<td>-</td>
</tr>
<tr>
<td>C10029T</td>
<td>T10561C</td>
</tr>
<tr>
<td>-</td>
<td>Deletion 11288-11296</td>
</tr>
<tr>
<td>T11418C</td>
<td>C11514T</td>
</tr>
<tr>
<td>C11956T</td>
<td>C11956T</td>
</tr>
<tr>
<td>C14408T</td>
<td>C14408T</td>
</tr>
<tr>
<td>C16573T</td>
<td>C16869T</td>
</tr>
<tr>
<td>C16689T</td>
<td>C17004T</td>
</tr>
<tr>
<td>-</td>
<td>T17348C</td>
</tr>
<tr>
<td>-</td>
<td>A17916G</td>
</tr>
<tr>
<td>-</td>
<td>C18175T</td>
</tr>
<tr>
<td>-</td>
<td>C18804T</td>
</tr>
<tr>
<td>-</td>
<td>G19348T</td>
</tr>
<tr>
<td>-</td>
<td>T19680C</td>
</tr>
<tr>
<td>C20283T</td>
<td>A21258G</td>
</tr>
<tr>
<td>C21588T</td>
<td>C21588T</td>
</tr>
<tr>
<td>G21848C</td>
<td>G21848C</td>
</tr>
<tr>
<td>Deletion 21968-21994</td>
<td>Deletion 21968-21994</td>
</tr>
<tr>
<td>G22132T</td>
<td>G22132T</td>
</tr>
<tr>
<td>T22191C</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>G22205C</td>
</tr>
<tr>
<td>A22600C</td>
<td>A22600C</td>
</tr>
<tr>
<td>A22743G</td>
<td>A22743G</td>
</tr>
<tr>
<td>T22907A</td>
<td>T22907A</td>
</tr>
<tr>
<td>G23012A</td>
<td>-</td>
</tr>
<tr>
<td>T23030C</td>
<td>T23031C</td>
</tr>
<tr>
<td>T23031G</td>
<td>A23063T</td>
</tr>
<tr>
<td>A23063T</td>
<td>A23063T</td>
</tr>
<tr>
<td>A23403G</td>
<td>A23403G</td>
</tr>
<tr>
<td>C23604A</td>
<td>C23604A</td>
</tr>
<tr>
<td>C24138A</td>
<td>C24138A</td>
</tr>
<tr>
<td>G24368C</td>
<td>-</td>
</tr>
<tr>
<td>G24977C</td>
<td>-</td>
</tr>
<tr>
<td>C25487T</td>
<td>C25487T</td>
</tr>
<tr>
<td>G25563T</td>
<td>G25563T</td>
</tr>
<tr>
<td>A26492T</td>
<td>-</td>
</tr>
<tr>
<td>T26767C</td>
<td>T26767G</td>
</tr>
<tr>
<td>C27513T</td>
<td>C27513T</td>
</tr>
<tr>
<td>C27807T</td>
<td>C27807T</td>
</tr>
<tr>
<td>-</td>
<td>T27833C</td>
</tr>
<tr>
<td>C27972T</td>
<td>C27972T</td>
</tr>
<tr>
<td>-</td>
<td>T28002C</td>
</tr>
<tr>
<td>Deletion 28271</td>
<td>Deletion 28271</td>
</tr>
<tr>
<td>T28297C</td>
<td>T28297C</td>
</tr>
<tr>
<td>G28337T</td>
<td>G28337T</td>
</tr>
<tr>
<td>G28887T</td>
<td>G28887T</td>
</tr>
<tr>
<td>T29377C</td>
<td>T29377C</td>
</tr>
<tr>
<td>G29405C</td>
<td>-</td>
</tr>
<tr>
<td>Deletion 29738-29758</td>
<td>-</td>
</tr>
<tr>
<td>G29779T</td>
<td>G29779T</td>
</tr>
</tbody>
</table>
Genomes were analyzed using the Nextstrain web-tool (https://clades.nextstrain.org/) [1,2].

Spike region is indicated by a grey background.

Supplementary Table S2. Comparison of amino acid mutational patterns of the B.1.640.1 and B.1.640.2 lineages

<table>
<thead>
<tr>
<th>B.1.640.1</th>
<th>B.1.640.2 (IHU variant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M: I82T</td>
<td>M: I82S</td>
</tr>
<tr>
<td>N: D22Y</td>
<td>N: D22Y</td>
</tr>
<tr>
<td>N: T205I</td>
<td>N: T205I</td>
</tr>
<tr>
<td>N: E378Q</td>
<td>-</td>
</tr>
<tr>
<td>ORF1a: P309L</td>
<td>ORF1a: P309L</td>
</tr>
<tr>
<td>ORF1a: E452G</td>
<td>-</td>
</tr>
<tr>
<td>ORF1a: L2119F</td>
<td>-</td>
</tr>
<tr>
<td>ORF1a: A2355S</td>
<td>-</td>
</tr>
<tr>
<td>ORF1a: S3149F</td>
<td>-</td>
</tr>
<tr>
<td>ORF1a: R3164H</td>
<td>-</td>
</tr>
<tr>
<td>ORF1a: T3255I</td>
<td>Deletions ORF1a: S3675-</td>
</tr>
<tr>
<td>ORF1a: T3718A</td>
<td>-</td>
</tr>
<tr>
<td>ORF1b: P314L</td>
<td>ORF1b: P314L</td>
</tr>
<tr>
<td>ORF1b: V1294A</td>
<td>ORF1b: V1294A</td>
</tr>
<tr>
<td>ORF1b: P1570S</td>
<td>ORF1b: P1570S</td>
</tr>
<tr>
<td>ORF1b: V1961F</td>
<td>ORF1b: V1961F</td>
</tr>
<tr>
<td>ORF3a: T32I</td>
<td>ORF3a: T32I</td>
</tr>
<tr>
<td>ORF3a: Q57H</td>
<td>ORF3a: Q57H</td>
</tr>
<tr>
<td>ORF8: Q27*</td>
<td>ORF8: Q27*</td>
</tr>
<tr>
<td>ORF9b: I5T</td>
<td>-</td>
</tr>
<tr>
<td>ORF9b: C37R</td>
<td>-</td>
</tr>
<tr>
<td>ORF9b: P10L</td>
<td>-</td>
</tr>
<tr>
<td>ORF9b: Q18H</td>
<td>-</td>
</tr>
<tr>
<td>S: P9L</td>
<td>S: P9L</td>
</tr>
<tr>
<td>S: E96Q</td>
<td>S: E96Q</td>
</tr>
<tr>
<td>S: R190S</td>
<td>S: R190S</td>
</tr>
<tr>
<td>S: I210T</td>
<td>-</td>
</tr>
<tr>
<td>S: R346S</td>
<td>S: R346S</td>
</tr>
<tr>
<td>S: N394S</td>
<td>S: N394S</td>
</tr>
<tr>
<td>S: Y449N</td>
<td>S: Y449N</td>
</tr>
<tr>
<td>S: F490R</td>
<td>S: F490R</td>
</tr>
<tr>
<td>S: N501Y</td>
<td>S: N501Y</td>
</tr>
<tr>
<td>S: D614G</td>
<td>S: D614G</td>
</tr>
<tr>
<td>S: P681H</td>
<td>S: P681H</td>
</tr>
<tr>
<td>S: T859N</td>
<td>S: T859N</td>
</tr>
<tr>
<td>S: D936H</td>
<td>-</td>
</tr>
<tr>
<td>S: D1139H</td>
<td>-</td>
</tr>
</tbody>
</table>

Genomes were analyzed using the Nextstrain web-tool (https://clades.nextstrain.org/) [1,2].

Spike region is indicated by a grey background.
REFERENCES
