WP6 Deliverable 6.8,

Report on viral loads, viral genome sequences for prospective cohorts (preliminary report)

Surbhi Malhotra-Kumar, Benoit Visseaux

Universiteit Antwerpen (UANTWERPEN)

French National Institute of Health and Medical Research (INSERM)

Project Classification

Project Acronym:	ORCHESTRA
Project Title:	Connecting European Cohorts to Increase Common and Effective Response to SARS- CoV-2 Pandemic
Coordinator:	UNIVR
Grant Agreement Number:	101016167
Funding Scheme:	Horizon 2020
Start:	1st December 2020
Duration:	36 months
Website:	www.orchestra-cohort.eu
Email:	info@orchestra.eu

Document Classification

WP No:	WP6
Deliverable No:	D6.8
Title:	Report on viral loads, viral genome sequences for prospective cohorts (preliminary report)
Lead Beneficiary:	Inserm / Antwerpen
Other Involved Beneficiaries:	
Nature:	Report
Dissemination Level:	Public
Due Delivery Date:	28.02.2022
Submission Date:	30.06.2022
Justification of delay:	The submission of this deliverable was delayed due to difficulties in obtaining data required from all participating cohorts
Status:	Final
Version:	1.0
Author(s):	Benoit Visseaux, Matilda Berkell, Surbhi Malhotra, Erley Lizarazo Forero

History of Changes

Version	Date	Created/Modified by
1.0		Matilda Berkell, Benoit Visseaux, Surbhi Malhotra-Kumar, Erley Lizarazo Forero
0.2		
0.3		

Executive summary

WP and deliverable context

The present report is part of ORCHESTRA project, a three-year international research project aimed at tackling the coronavirus pandemic. ORCHESTRA provides an innovative approach to learn from the pandemic SARS-CoV-2 crisis, derive recommendations to further management of COVID-19 and be prepared for the possible future pandemic waves. The ORCHESTRA project aims at delivering sound scientific evidence for the prevention and treatment of the infections caused by SARS-CoV-2 assessing epidemiological, clinical, microbiological, and genotypic aspects of population, environment, and socio-economic features. The project builds upon existing, and new large scale population cohorts in Europe (France, Germany, Spain, Italy, Belgium, Romania, Netherlands, Luxemburg, and Slovakia) and non-European countries (India, Perú, Ecuador, Colombia, Venezuela, Argentina, Brazil, Democratic Republic of Congo, and Gabon) including SARS-CoV-2 infected and non-infected individuals of all ages and conditions. The primary aim of ORCHESTRA is the creation of a new pan-European cohort applying homogenous protocols for data collection, data sharing, sampling, and follow-up, which can rapidly advance the knowledge on the control and management of the COVID-19. Within ORCHESTRA project, Work Package 6 (WP6) aims at providing innovative laboratory capabilities combining serology, immunology, viral and human genomics, microbiota, and epigenetic analysis. It aims to describe markers and physiopathology of various COVID-19 outcomes including severe cases, long COVID and vaccine efficiency across various patient populations gathered within ORCHESTRA cohorts.

The objectives of WP6 are distributed in two parts: (1) a retrospective part on frozen samples obtained during 2020 and (2) a prospective part starting in 2021.

Content of the document

The present report describes the cohorts, patient numbers and characteristics, and sampling timepoints subjected to characterization of viral loads and SARS-CoV-2 whole genome sequencing in COVID-19 patients in retrospective and prospective cohorts.

Dissemination level: Public

Core content

Methods

Sample collection and patient cohorts

Cohorts that provided nasopharyngeal swab samples for sequencing and/or genome sequences included:

- University of Verona, Italy (UNIVR) (n=..samples)
- University of Bologna, Italy (UNIBO)
- University Medical Center Groningen, the Netherlands (UMCG)
- Servicio Andaluz de Salud, Spain (SAS)
- FrenchCovid, France

Three centers performed RNA extractions, RT-qPCR, and viral variant sequencing:

- University of Antwerp (UA), Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Belgium (samples sequenced: UNIVR, UNIBO, SAS)
- INSERM, Laboratoire de Virologie, Hôpital Bichat Claude Bernard, France (samples sequenced : FrenchCovid, SAS)
- University Medical Center Groningen, The Department of Medical Microbiology and Infection Control, the Netherlands (samples sequenced: UMCG)

RNA extraction and SARS-CoV-2 RT-qPCR

At UA, RNA was extracted from 350 µL of nasopharyngeal swab storage medium using MagMAX Viral/Pathogen II Nucleic Acid Isolation Kit (ThermoFisher Scientific) on the KingFisher Flex Purification System (ThermoFisher Scientific). Subsequently, reverse-transcriptase Real-Time PCR (RT-PCR) targeting SARS-CoV-2 was performed using QuantStudio 5 Real-Time PCR instrument (ThermoFisher Scientific) and TaqPath COVID-19 CE-IVD RT-PCR kit (ThermoFisher Scientific), targeting three regions (S protein, N protein, and ORF1ab) of the SARS-CoV-2 virus. Data analysis was performed using FastFinder Analysis software (UgenTec, Hasselt, Belgium), where detection of at least two gene targets was considered positive and Ct values above 37 were considered negative, as recommended by the supplier.

At INSERM, extracted RNA was analyzed using the RealStar® SARS-CoV-2 RT-PCR Kit 1.0 (Altona Diagnostics, Hamburg, Germany) targeting two regions (E protein and S protein or non-specified). All Ct values above 40 were considered negative for all these tests as recommended by the manufacturers.

At UMCG, SARS-CoV-2 infection was detected using qRT-PCR assays by targeting: i) the E-gene, ii) the RdRp / N genes employing the AlinityM platform (Abbott Laboratories, Chicago, IL, USA), or, iii) the E / N genes Gene-Xpert (Cepheid). Samples with Ct < 30 were selected for SARS-CoV-2 WGS and organized based on their Ct values for downstream pooling compatibility.

SARS-CoV-2 whole genome sequencing

Whole genome sequencing (WGS) of the SARS-CoV-2 genome was conducted at the University of Antwerp by preparation of multiplexed libraries with the Illumina COVIDSeq kit (Illumina Inc.) using a Zephyr G3 NGS Robot followed by 2 x 74 bp paired-end sequencing on a NextSeq 500/550 instrument (Illumina Inc.). Library preparation for whole genome sequencing was performed at UA using the Illumina COVIDSeq kit (Illumina, Cat. No. 20043675) according to manufacturer's protocol. DNA quantification of the pooled library was performed using the Qubit dsDNA HS Assay kit (ThermoFisher, Cat. No. Q32854). Library denaturation was performed using the NextSeq 550/500 High Output kit v2 with a 1.4 nM Phix Library positive control with 1% spike-in. Sequencing was performed on a NextSeq 500/550 instrument using the NextSeq 500/550 High Output kit v2 with a loading volume of 1300 µl.

At INSERM, whole genome sequencing was conducted using the NEBNext ARTIC SARS-CoV-2 Companion Kit - Oxford Nanopore Technologies (New England Biolabs, Ipswich, Massachusetts, U.S.A.), based on the Artic protocol. Briefly, acid nucleic extraction using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche, Basel, Switzerland) and reverse transcription was performed with LunaScript and random hexamers. Tiling PCR amplification was performed with two pools of primers (ARTIC nCoV-2019 V3 panel). Libraries were then prepared with NEBNext Companion Module for Oxford Nanopore Technologies, Ligation Sequencing (SQK-LSK 109), then sequenced on a GridION analyser using MinION R9.4.1 flow cells (Oxford Nanopore Technologies, Oxford, U.K.).

SARS-CoV-2 WGS was carried at the UMCG using the EasySeq™ RC-PCR SARS-CoV-2 WGS kit (NimaGen BV, Nijmegen, The Netherlands). In short, RNA was extracted from 190 µL of sample using the NucliSense EasyMag (bioMérieux, Lyon, France) and eluted in 50 µL. A cDNA synthesis reaction was prepared with the LunaScript RT supermix (5x) kit (New England Biolabs, Ipswich MA, EUA). 10 µL cDNA input was split into two reactions with separate primer / probe pools used for targeted amplification with the RC-probe. Amplicons were then pooled based on the initial Ct-value according to the manufacturer. Next, 9 pM final Library pools were sequenced on an Illumina MiSeq platform using a Mid Output Kit (2x200-cycles PE) (Illumina, San Diego, CA, USA).

Bioinformatic analysis

At UANTWERP, raw Illumina data quality assessment was performed using FastQC, followed by quality trimming with TrimGalore v. 0.6.7. At UMCG, MiSeq data were processed with MiSeq control software v2.4.0.4 and MiSeq Reporter v2.4. Reference mapping was performed at both centers against the SARS-CoV-2 genome (GenBank: NC_045512.2, or MN908947.3) using the CLC Genomics Workbench v.9.5.3 (Qiagen), and extraction of consensus sequences. At INSERM, SARS-CoV-2 consensus sequences were obtained from long-read sequencing data using the Medaka-based Artic-nCoV workflow v1.1.0 adapted by EPI2ME lab. Clade and lineage assignment was performed using Nextclade (https://clades.nextstrain.org) and Pangolin v.4.0.6 (https://pangolin.cog-uk.io), respectively.

Results

All received samples were subjected to SARS-CoV-2 RT-qPCR and whole genome sequencing analysis, which resulted, to date, in 1570 genome sequences that have been shared within the ORCHESTRA consortium and deposited to GISAID (https://www.gisaid.org/) and the COVID-19 data portal (https://www.covid19dataportal.org/). These genomes were collected from patients experiencing SARS-CoV-2 infections between 29 January 2020 and 8 February 2022 and several different variants of concern, such as 20I/alpha, 21A/I/J/delta, and 21K/L/omicron were observed across the cohorts during the sampling period. Overall, 48.1% of patients were female and the median age was 59 years (IQR: 44-73). Further, 58.7 patients were outpatients, whereas 37.6% of patients experienced hospitalization, 11.4% were admitted to the ICU, and 13.8% died after their COVID-19 infection (**Table 1**).

Overall, 755 of the sequenced genomes were obtained from female patients with COVID-19, and the median age was 59 years (IQR: 44-73, **Table 1**). 61.7% of the sequenced genomes originated from two cohorts in Italy, whereas 22.2%, 10.2%, and 4.30% originated from the Netherlands, Spain, and France, respectively. In total, sequences from 15 different clades were detected where the most abundant were 21J/delta (n = 495), 20A (n = 280), 20E (EU1, n = 260), and 21K/omicron (n = 228, **Table 1**).

An overview of the distribution of variants detected during the sampling period for all cohorts can be seen in **Figure 1**. The majority of the samples collected in the study were clustered around March – April 2020, November 2020 – January 2021, and November 2021 – January 2022. The first intensive sampling period was primarily characterized by genomes belonging to the 20A clade. At the end of 2020, we observed a shift in the most predominant clades, where 20A is still prevalent, but an increased abundance of the 20E (EU1) clade is observed, as well as the emergence of the 20I/alpha clade. From July 2021 through December 2021, the vast majority of the sequenced genomes belonged to a single dominant clade, namely the 21J/delta clade. This was followed by the emergence of the 21K/omicron clade that in January 2022 became dominant.

When comparing combined Ct values between outpatients, hospitalized, ICU-admitted, and deceased patients it was observed that deceased patients carrying 20A, 20B, and 20I (Alpha, V1) subvariants displayed lower viral loads than hospitalized, ICU-admitted, and outpatients indicative of a higher viral load (**Figure 2**). As the prevalence of delta and omicron subvariants increased at the end of 2021 and early 2022, no or few patients in this study were admitted to the ICU or deceased, possible due to the increased percentage of fully vaccinated individuals participating in the study in comparison with the early stages of the pandemic.

Table 1. Patient characteristics for patients with successfully sequenced SARS-CoV-2 genomes submitted to GISAID enrolled in cohorts from Italy, the Netherlands, France, and Spain.

Patient characteristic	N (%)
Total	1570
Female	755 (48.1)
Male	809 (51.9)
Age (median, IQR)	59 (44-73)
Country	
Italy	969 (61.7)
University of Verona	621 (39.6)
University of Bologna	375 (23.9)
Netherlands (University Medical Center Groningen)	348 (22.2)
Spain (Servicio Andaluz de Salud)	160 (10.2)
France (FrenchCovid)	66 (4.2)
Outcome/severity	
Outpatient	921 (58.7)
Hospitalization	591 (37.6)
ICU admission	179 (11.4)
Deceased	216 (13.8)
Clade	
19A	12 (0.8)
19B	7 (0.4)
20A	280 (17.8)
20B	63 (4.0)
20C	23 (1.5)
20D	1 (0.1)
20E (EU1)	260 (16.6)
20H (Beta, V2)	1 (0.1)
20I (Alpha, V1)	125 (8.0)
20J (Gamma, V3)	5 (0.3)
21A (Delta)	2 (0.1)
21I (Delta)	61 (3.9)
21J (Delta)	495 (31.5)
21K (Omicron)	228 (14.5)
21L (Omicron)	7 (0.4)

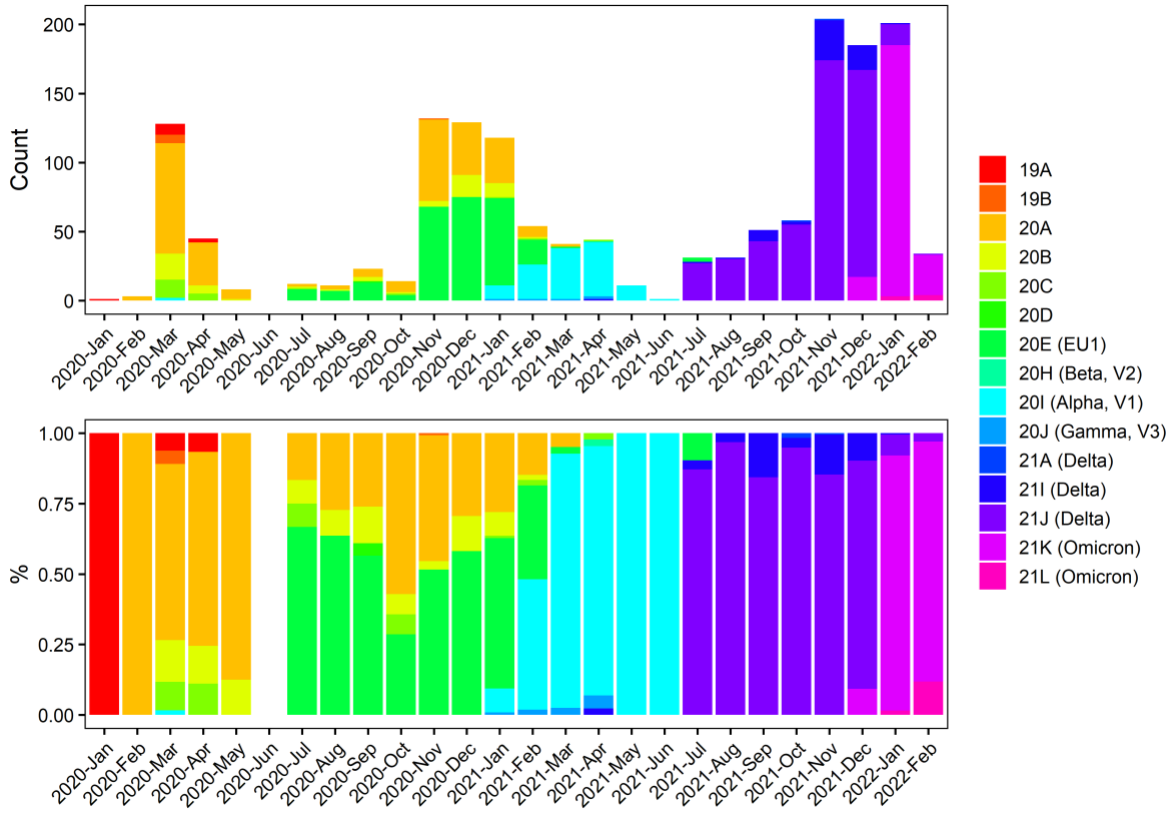


Figure 1. Distribution of SARS-CoV-2 genomes across different clades over the study period of January 2020 to February 2022. See **Supplementary Figure 1** for the distribution of genomes generated across the study period for the different patient cohorts.

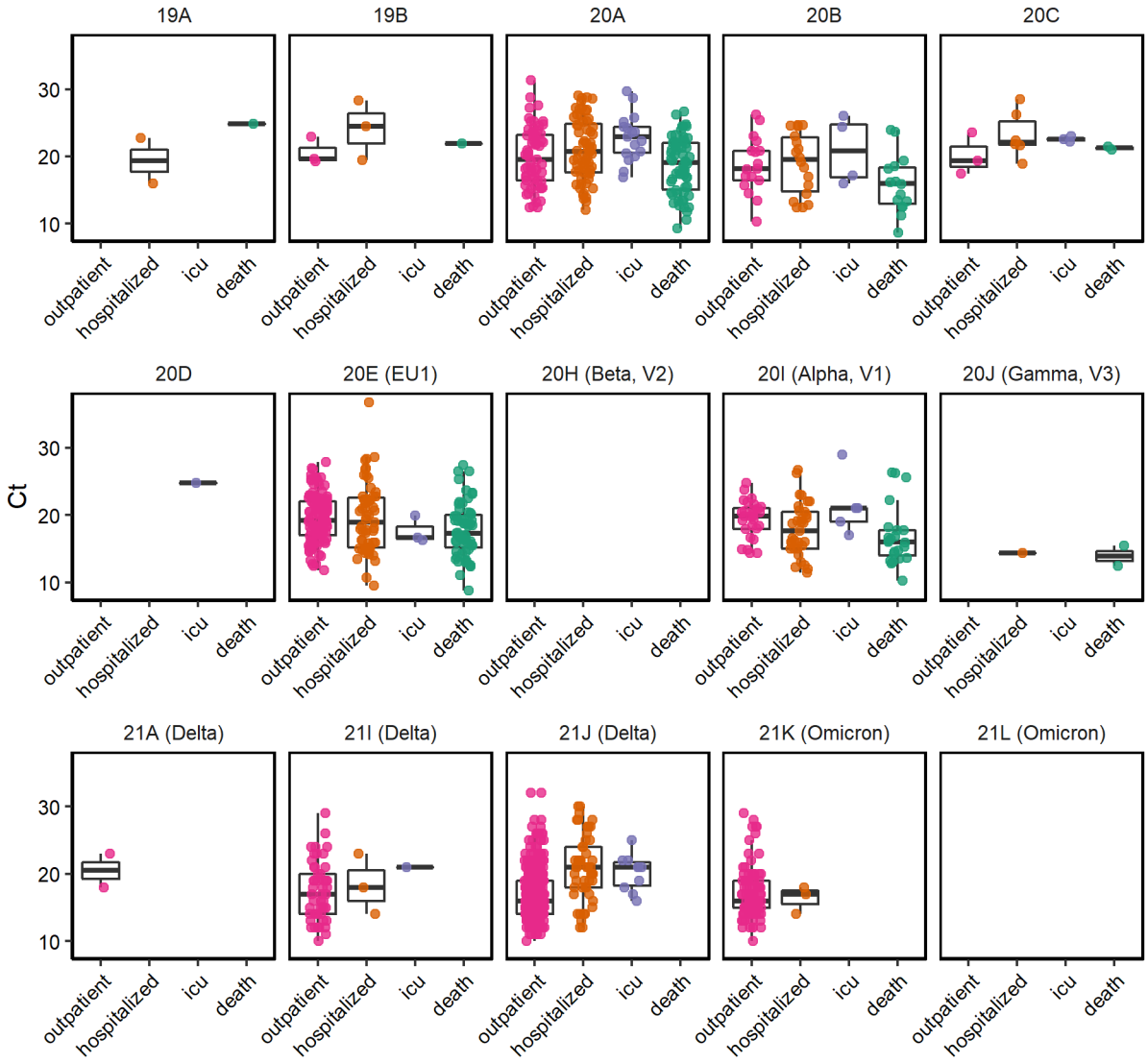
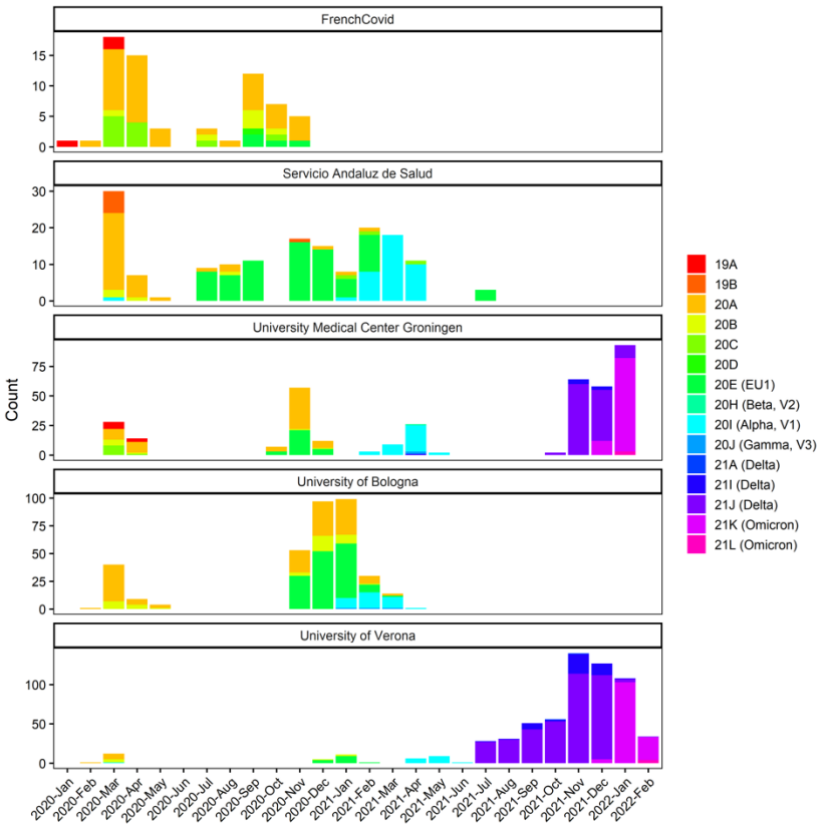
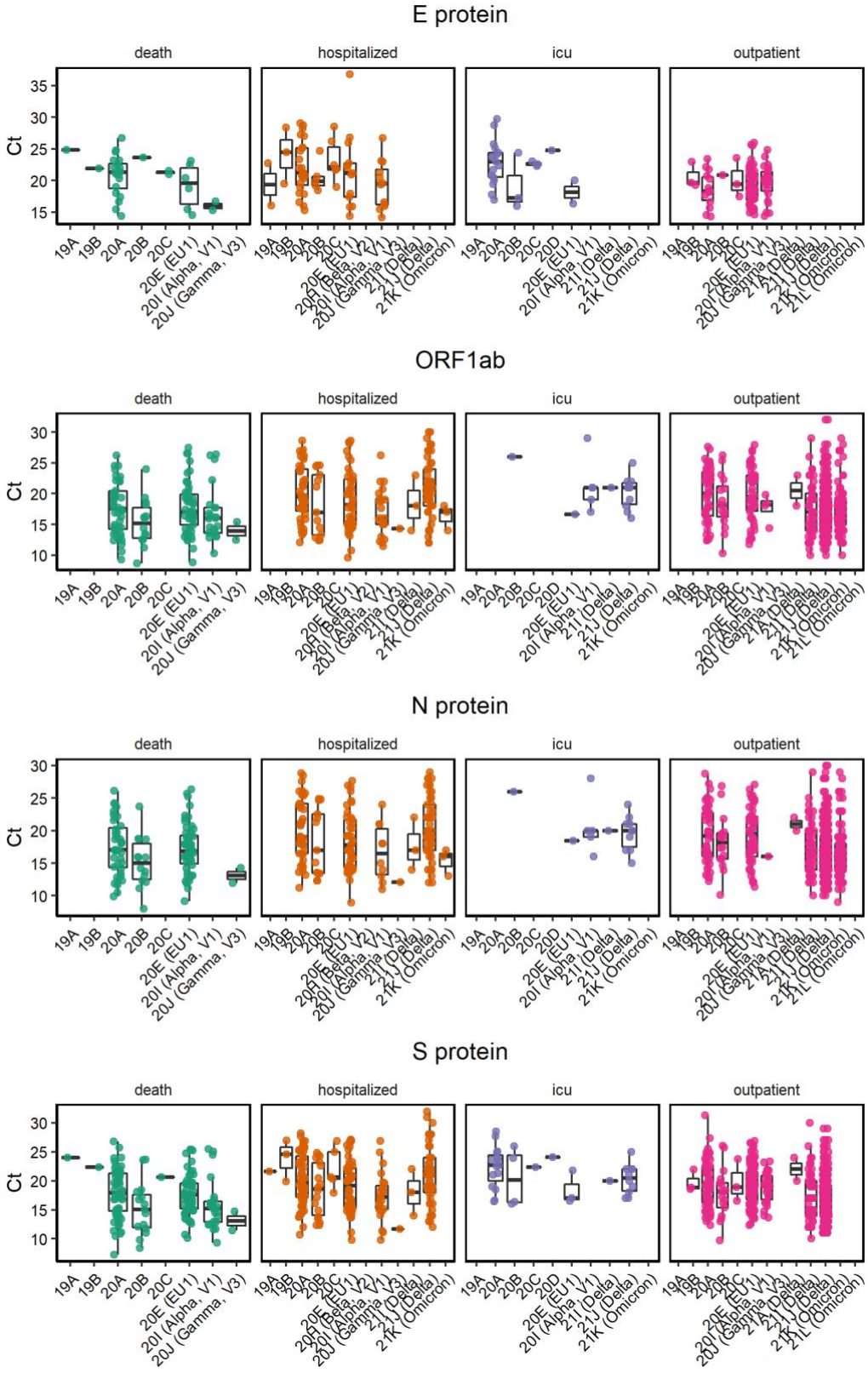


Figure 2. Combined cyclic threshold (Ct) values stratified by Nextstrain clade and patient outcome. The first available gene target based on the following list is utilized: E protein, ORF1ab, N protein, S protein. See **Supplementary Figure 2** for Cts of individual gene targets.

Supplementary Figures



Supplementary Figure 1. Distribution of SARS-CoV-2 genomes across different clades in the different cohorts over the study period of January 2020 to February 2022.



Supplementary Figure 2. Cyclic threshold (Ct) values stratified by Nextstrain clade and patient outcome for all detected gene targets.

Acknowledgments

The WP6 want to acknowledge the cohorts and corresponding teams that provided the samples and associated clinical data (UNIVR; UNIBO; FrenchCOVID cohort; SAS cohort; UMCG).

Appendix

Appended to this deliverable is the metadata submitted together with the SARS-CoV-2 genomes deposited on GISAID.