ORCHESTRA

Connecting European Cohorts
to Increase Common
and Effective Response to
SARS- CoV-2 Pandemic

DELIVERABLE

WP6 D6.3

Report on viral loads, viral genome sequences of retrospective cohorts

Dr Benoit Visseaux, Dr Surbhi Malhotra

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

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:	D6.3
Title:	Report on viral loads, viral genome sequences of retrospective cohorts
Lead Beneficiary:	INSERM / UANTWERPEN
Other Involved Beneficiaries:	UNIVR, UNIBO, SAS
Nature:	Report
Dissemination Level:	Public
Due Delivery Date:	31.08.2021
Submission Date:	15.09.2022
Justification of delay:	N/A
Status:	Final
Version:	2.0
Author(s):	Benoit Visseaux, Matilda Berkell, Romain Coppée, Samuel Lebourgeois, Surbhi Malhotra

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

History of Changes

Version	Date	Created/Modified by
1.0	29.10.2021	Benoit Visseaux, Matilda Berkell, Surbhi Malhotra-Kumar
2.0	15.09.2022	Matilda Berkell, Surbhi Malhotra-Kumar

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Table of contents

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

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 largescale 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, 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, the Work Package 6 (WP6) aims at providing innovative laboratory capabilities combining serology, immunology, viral and human genomes, microbiota, and epigenetic analysis. It aims describing markers and physiopathology of various COVID-19 outcomes including severe cases, long COVID and vaccine efficiency across various patients' populations gathered within ORCHESTRA cohorts.

The objectives of the WP6 are distributed in two parts: a retrospective part on frozen samples obtained during 2020 and (2) a prospective part starting in 2021. The goal for the viral genomic part is to allow the description of the SARS-CoV-2 variants (both known variants and novel mutations) according to epidemics waves, stages of SARS-CoV-2 infection (mild vs severe), setting (outpatients vs hospitalized), and outcome (survivors vs non-survivors).

Content of the document

The present report describes the viral genome sequences obtained from the analysis of 1500 frozen NP samples included from retrospective cohorts covering the year 2020.

The objective is to allow the description of the SARS-CoV-2 variants (both known variants and novel mutations) according to epidemics waves, stages of SARS-CoV-2 infection (mild vs severe), setting (outpatients vs hospitalized), and outcome (survivors vs non-survivors).

Dissemination level: Public

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Core content

Methods

Sample collection and patient cohorts

All cohorts participating in ORCHESTRA WP2, WP3, WP4, and WP5 were solicited for availability of biobanked nasopharyngeal swab samples that would be analysed in this study.

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

- University of Verona, Italy (UNIVR)
- University of Bologna, Italy (UNIBO)
- Servicio Andaluz de Salud, Spain (SAS)
- FrenchCovid, France

Two 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)
- INSERM, Laboratoire de Virologie, Hôpital Bichat Claude Bernard, France (samples sequenced: FrenchCovid)

One centre directly provided RT-qPCR and SARS-CoV-2 whole genome sequencing data:

- University Medical Center Groningen, the Netherlands (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.

SARS-CoV-2 whole genome sequencing

Whole genome sequencing (WGS) of the SARS-CoV-2 genome was conducted at UA 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

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

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.).

Bioinformatic analysis

At UA, raw Illumina data quality assessment was performed using FastQC, followed by quality trimming with TrimGalore v. 0.6.7. 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 (https://pangolin.cog-uk.io), respectively, for all generated sequences. All obtained sequences have been provided to the ORCHESTRA consortium and deposited on GISAID (https://www.gisaid.org/).

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Results

All received samples were subjected to SARS-CoV-2 RT-qPCR and whole genome sequencing analysis, which resulted, to date, in 760 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 May 10.

Overall, SARS-CoV-2 genomes from eleven different clades were observed during the collection period across the five cohorts, where clades 20A, 20E (EU1), and 20I (Alpha, V1) were the most common (**Table 1**). 50.7% of the sequenced genomes originated from two cohorts in Italy, whereas 20.5%, 20.1%, and 8.7% originated from the Netherlands, Spain, and France, respectively. An overview of the distribution of variants detected during the sampling period for all cohorts can be seen in **Figure 1**.

When comparing combined Ct values between patients infected with different SARS-CoV-2 variants based on their NextStrain classification, a significant difference was observed (Kruskal-Wallis, $p < 0.001$, **Figure 2**). However, when conducting pairwise comparisons between the different variants, only 20E (EU1) was found to be significantly different from 19A and 20C (Wilcoxon rank sum with Bonferroni post-hoc correction, $p \leq 0.027$).

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	760
Country	
Italy	385 (50.7)
University of Verona	16 (2.1)
University of Bologna	369 (48.6)
Netherlands (University Medical Center Groningen)	156 (20.5)
Spain (Servicio Andaluz de Salud)	153 (20.1)
France (FrenchCovid)	66 (8.7)
Clade	
19A	12 (1.6)
19B	7 (0.9)
20A	285 (37.5)
20B	61 (8.0)
20C	23 (3.0)
20D	1 (0.1)
20E (EU1)	257 (33.8)
20I (Alpha, V1)	107 (14.1)
20H (Beta, V2)	1 (0.1)
20J (Gamma, V3)	5 (0.7)
20I (Delta)	1 (0.1)

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

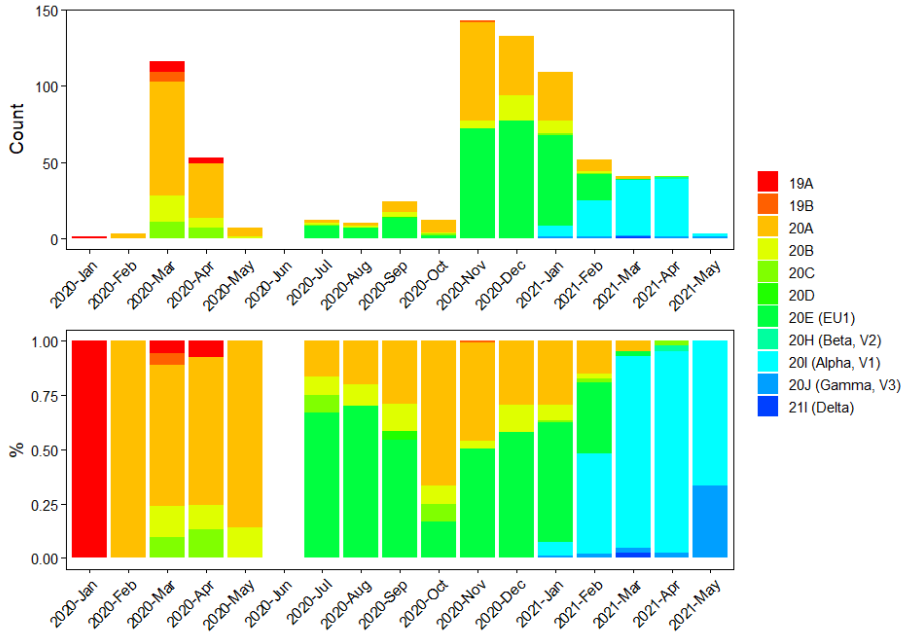


Figure 1. Distribution of SARS-CoV-2 genomes across different clades over the study period of January 2020 to May 2021. See **Supplementary Figure 1** for the distribution of genomes generated across the study period for the different patient cohorts.

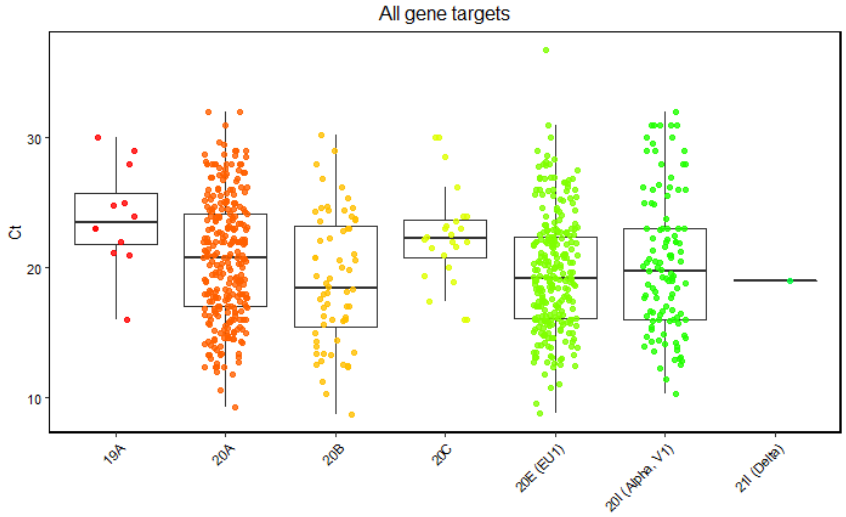
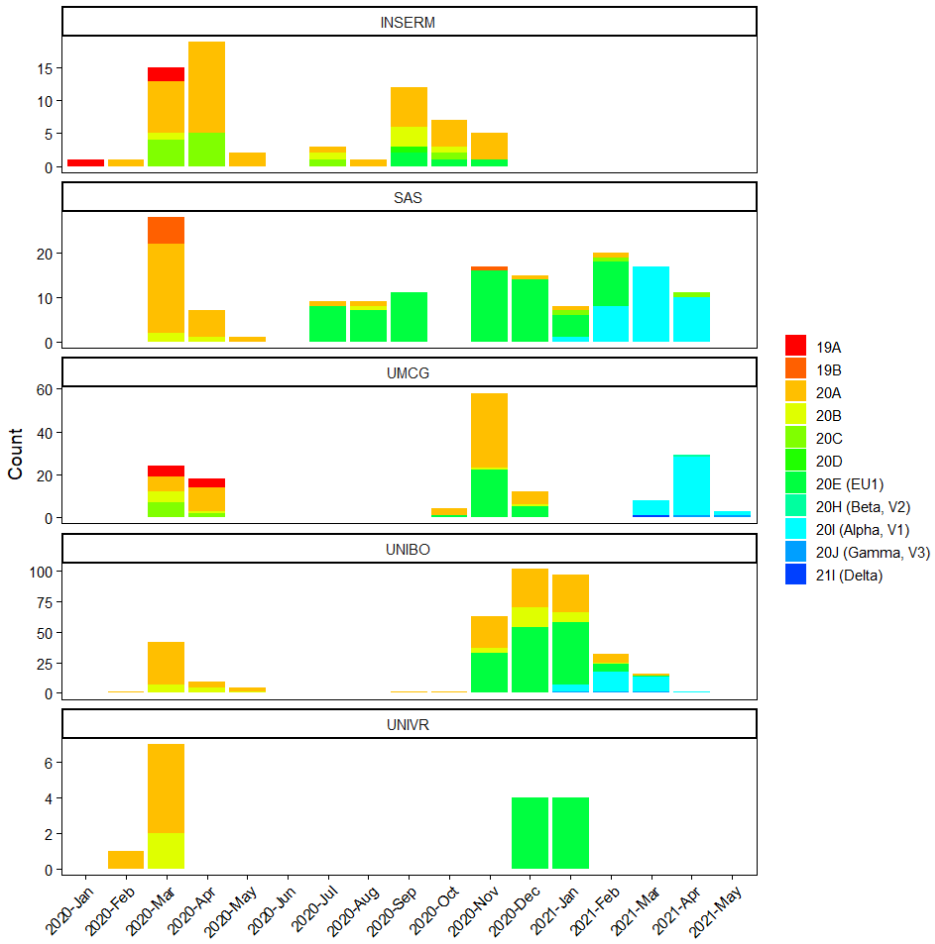


Figure 2. Combined cyclic threshold (Ct) values stratified by Nextstrain clade and patient outcome. The first available gene target based on the following list is utilized: E protein, ORF1ab, N protein, S protein.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

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 May 2021.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Acknowledgments

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

Appendix

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

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167