JIP06-COVID19-COVRIN-D.2.1.3

Workpackage 2

Responsible Partners:
INSA (36), IZSAM (28)

Contributing partners: All partners

This meeting is part of the European Joint Programme One Health EJP.
This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

GENERAL INFORMATION

European Joint Programme full title	**Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards**
European Joint Programme acronym	**One Health EJP**
Funding	**This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.**
Grant Agreement	**Grant agreement n° 773830**
Start Date	**01/01/2018**
Duration	**60 Months**

DOCUMENT MANAGEMENT

Title OHEJP deliverable	**D-COVRIN.2.1.3 Report on the development of an algorithm for novel virus assessment**
WP and task	**JIP06-WP2-T2.1 Genome analyses /** ST2.1.3 Phylogenetic analyses ("phylogenetic analysis and investigation for potential virulence traits")
Leader	T2.1 Vítor Borges (INSA, 36) / ST2.1.3 Alessio Lorusso (IZSAM, 28)
Other contributors	Adriano Di Pasquale (IZSAM, 28)
Due month of the deliverable	**53**
Actual submission month	**54**
Type ***R: Document, report*** ***DEC: Websites, patent filings, videos, etc.; OTHER***	R **Save date**: 30-June-22
Dissemination level ***PU: Public (default)*** ***CO: confidential, only for members of the consortium (including the Commission Services).***	**PU**
Dissemination ***Author's suggestion to inform the following possible interested parties.***	OHEJP WP 1 ☐ OHEJP WP 2 ☐ OHEJP WP 3 ☐ OHEJP WP 4 ☐ OHEJP WP 5 ☐ OHEJP WP 6 ☐ OHEJP WP 7 ☐ Project Management Team ☒ Communication Team ☒ Scientific Steering Board ☐ National Stakeholders/Program Owners Committee ☐ EFSA ☐ ECDC ☐ EEA ☐ EMA ☐ FAO ☐ WHO ☐ OIE ☐ Other international stakeholder(s): …… Social Media: …… **Other recipient(s):** ……

This meeting is part of the European Joint Programme One Health EJP.
This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

JIP06-COVID19-COVRIN-D.2.1.3

JIP06-WP2-T2.1 Genome analyses
JIP06-WP2- sub-task ST2.1.3 Phylogenetic analyses
D.2.1.3 Report on the development of an algorithm for novel virus assessment

1. Description of the task

The two main overall operational objectives of COVRIN WP2 are: i) to identify drivers for the emergence and spread of SARS-CoV-2; ii) to generate data and build models for risk assessment of SARS-CoV-2.

Within WP2, **Task 2.1 is focused on the analyses of whole genome sequences of SARS-CoV-2 strains circulating in humans and animals.** Task 2.1. involves methodologies exchange among partners [from wet-lab Next-generation sequencing (NGS)-related procedures to Bioinformatics pipelines], as well as **phylogenetic analysis and investigation for potential virulence traits (e.g., mutations known to play a direct role in receptor binding or antibody recognition).** This approach aims not only at **promoting protocol sharing/harmonization**, but also at **mapping evolutionary changes of SARS-CoV-2 viruses isolated within and across human and different animal species**, as a means to identify mutations and recombination events driving adaptation to alternative hosts. In parallel, this task also supports **the selection of strains that may be used for downstream biological characterization in vivo and in vitro.**

This deliverable describes the propaedeutic activities and first outcomes of the **sub-task ST2.1.3 (Phylogenetic analyses).**

2. Description of the deliverable

ST2.1.3 is building bridges between several One Health EJP projects to promote the integration of bioinformatics tools that enable phylogenetic analysis and investigation for virulence traits (e.g., mutations known to play a direct role in receptor binding or antibody recognition) that can aid the identification of drivers for the emergence and spread of new viral threats and the generation of supportive data for risk assessment. This deliverable describes such activities in sections 2.2 - 2.4. However, before getting started with the development of novel bioinformatics tools, IZSAM (P28) also analyzed the performances of five different sequencing protocols to choose the best for downstream activities. These complementary activities (which are *also related to ST2.1.2 – Harmonisation of sampling and methods for NGS (M40 – M50, extended from M48*) are summarized in section 2.1 of the present Deliverable.

2.1. Selection of the best library protocol for SARS-CoV-2 genome sequencing

Nucleic acids purified from nasopharyngeal swabs collected from inpatients and outpatients of the Abruzzi region, central Italy, were tested by a SARS-CoV-2-specific commercially available Real Time RT-PCR. A total of 33 positive samples with cycle threshold (CT) ranging from 16 to 33 was used to compare five different sequencing protocols, including:

- the Sequence-Independent Single-Primer Amplification (SISPA) technique;
- the Arbor SARS-CoV-2 panel, a targeted enrichment approach;
- the ARTIC protocol, a targeted method based on specific multiplex amplifications with specific primers;
- two amplicon-based commercial protocols namely the Swift Amplicon® SARS-CoV-2 Panel kit and the Illumina COVIDSeq Test.

SISPA, Arbor and ARTIC libraries were prepared using Illumina DNA Prep kit and deep sequencing of all samples was performed onto Illumina platforms, MiniSeq or NextSeq500. Bioinformatic analyses were performed as recently described by Di Pasquale and colleagues [1].

The SISPA protocol combined with NGS produced complete sequences only from swabs with CT ≤ 20, instead, Arbor and ARTIC protocols produced complete genome sequences from all samples, reaching optimal levels of vertical (5000x) and horizontal (100%) coverages. The Arbor protocol was more expansive and laborious than ARTIC. The Swift protocol was more cost-effective and less time-consuming than ARTIC, but less efficient in producing reliable data. Finally, the COVIDSeq Test showed the highest performances as complete consensus sequences (100% horizontal coverage) were also obtained from low-input samples (CT > 30) with the best processing time and costs. We conclude that the COVIDSeq Test method is a versatile and scalable method that is immediately applicable for SARS-CoV-2 genomic surveillance and easily adaptable to other pathogens. The protocol was adopted by IZSAM (28) for routine genomic surveillance of SARS-CoV-2.

2.2. Enhancement of *vcf2mst* tool

IZSAM (IT) team developed and published a method for rapid large-scale phylogenetic inferences of SARS-CoV-2 samples, called ***vcf2mst*** [1].

In the COVRIN ST2.1.3, the tool has been enhanced in order to:

- Manage the inclusion/exclusion of genome sub-regions. This is very useful for managing low coverage genomic regions of pathogens and for clustering samples focusing on specific sub-regions of interest (e.g., Spike).
- Accept input from INSA´s algn2pheno (https://github.com/insapathogenomics/algn2pheno) [2] output file format (see section 2.4).
- Accept input from Nextclade (https://clades.nextstrain.org/) output file format.
- Accept input from any CSV file format using a flexible input method.
- Integrate vcf2mst into INSA's ReporTree tool (https://github.com/insapathogenomics/ReporTree) [3].

Noteworthy, the last point is a direct outcome of the previous ones, as the broader variety of possible inputs for vcf2mst made it useful to increase the input flexibility of ReporTree. ReporTree is a flexible pipeline developed in the frame of the One Health EJP project "BeOne: Building Integrative Tools for One Health Surveillance" (https://onehealthejp.eu/jrp-beone/) that facilitates the detection of genetic clusters and their linkage to epidemiological data for multiple pathogens. **This activity reflects the integrative role of COVRIN by building bridges between other ongoing One Health EJP projects and, thus, enhancing our ability to the detect and survey human and animal pathogens.**

All **vcf2mst** enhancements have been packaged and published on the GITHUB repository at the address https://github.com/genpat-it/vcf2mst together with a user manual explaining the different usages. Hence, the tool is freely available. Anyone can download and use it. Currently, this tool was adopted by IZSAM (28) for routine genomic surveillance of SARS-CoV-2.

2.3. Phylogenetic analyses and outbreak investigations

The newly adopted protocol was used by IZSAM (P28) to identify and characterize emerging SARS-CoV-2 variants [4], to depict the early molecular epidemiology of Alpha and Gamma variants in Italy [5], to identify a Delta variant strain from an infected dog [6] (also linked to "*JIP06-WP1-T1.1 - SARS-CoV2 genome detection in livestock, wildlife, pets, and environmental samples (M40 – M63*)") and to describe the early genomic epidemiology of the first documented outbreak of Alpha variant in Italy ([7].

During SARS-CoV-2 diagnostic activities performed in Abruzzo region (central Italy), several strains belonging to the B.1.177.75 lineage tested negative for the N gene, but positive for the ORF1ab and S genes (+/+/- pattern) by the TaqPath COVID-19 CE-IVD RT-PCR Kit manufactured by Thermofisher. By sequencing, a unique mutation, synonymous 28948C > T, was found in the N-negative B.1.177.75 strains. Although we do not have any knowledge upon the nucleotide sequences of the primers and probe adopted by this kit, it is likely that N gene dropout only occurs when 28948C > T is coupled with 28932C > T, this latter present, in turn, in all B.1.177.75 sequences available on public databases. Furthermore, epidemiological analysis was also performed. The majority of the N-negative B.1.177.75 cases belonged to two clusters apparently unrelated to each other and both clusters involved young people. However, the phylogeny for sequences containing the +/+/- (ORF1ab, S, N) pattern strongly supports a genetic connection and one common source for both clusters. However, genetic comparison suggests a connection rather than indicating the independent emergence of the same mutation in two apparently unrelated clusters. This activity highlights once more the importance of sharing genomic data to link apparently unrelated epidemiological clusters.

Moreover, we conducted an analysis of 6515 SARS-CoV-2 sequences sampled in Italy between 29 January 2020 and 1 March 2021 and show how different lineages emerged multiple times independently despite lockdown restrictions. Virus lineage B.1.177 became the dominant variant in November 2020, when cases peaked at 40,000 a day, but since January 2021 this is being replaced by the Alpha variant 'variant of concern-VOC'. In addition, we report a sudden increase in another documented VOC—Gamma—from December 2020 onwards, most likely caused by a single introduction into Italy. We again highlighted how international importations drive the emergence of new lineages and that genome sequencing should remain a top priority for ongoing surveillance in Italy.

In this Deliverable, we also describe the case of a paucisymptomatic dog in Italy infected with SARS-CoV-2 from a household with three confirmed human cases of COVID-19 living in Pesaro (Marche region, Italy). The dog showed high viral RNA titers in the nasal and oropharyngeal swabs. In the nasal swab, SARS-CoV-2 RNA lasted for a least a week. By sequencing, the strain was assigned to the AY.23 lineage one of the sub-lineages of the major SARS-CoV-2 Delta variant of concern (VOC). Although we did not process the swabs of the three human cases, we strongly suspect a human origin of the dog infection. In this regard, AY.23 sequences, although never released thus far in the Marche region, were detected in the neighboring regions. Our findings highlight once more the need for a One Health approach for SARS-CoV-2 surveillance, management, and control, thus preventing viral spillover from animals to humans.

Finally, in this Deliverable, we also describe the analysis of the early spread of Alpha VOC in the town of Guardiagrele (Abruzzo Region, Italy), the first documented case of Alpha variant in Italy. Genomic investigations showed close identity between the sequences from Guardiagrele, forming one distinct clade. This would suggest one or limited unspecified viral introductions from outside to Abruzzo region in early December 2020, which led to the diffusion of Alpha VOC in Guardiagrele and in neighbouring municipalities, with very limited inter-regional mixing.

2.4. Screening of genetic features potentially linked to specific SARS-CoV-2 phenotypes

On behalf of other One Health EJP projects (TELE-VIR, BeONE), INSA (36) is developing bioinformatics tools for:

i) viral detection and genome-based surveillance (https://insaflu.insa.pt) [8], under TELE-VIR project (https://onehealthejp.eu/jrp-tele-vir/);

ii) clustering analysis and rapid association of pathogen genetic clusters with surveillance (clinical /epidemiological) metadata (https://github.com/insapathogenomics/ReporTree) [3], under BeONE project (https://onehealthejp.eu/jrp-beone/), as mentioned in section 2.2; or,

iii) rapidly screening of genetic features potentially linked to specific phenotypes (algn2pheno) (https://github.com/insapathogenomics/algn2pheno) [4], under TELE-VIR project (https://onehealthejp.eu/jrp-tele-vir/);

In this context, as another important goal of COVRIN ST2.1.3 is to employ genomic tools that can facilitate the selection of strains to be used for downstream biological characterization *in vivo* and *in vitro*, IZSAM (P28) applied the "algn2pheno" tool for the selection of viral strains for *in vivo* work in hACE2-transgenic mice (*also related to JIP06-WP2-T2.3 – Development and optimization of animal models (M43 – M60)*). For this, SARS-CoV-2 Spike alignments from IZSAM sequence collection were screened using the algn2pheno tool against two databases: i) "COG-UK Antigenic mutations" (https://sars2.cvr.gla.ac.uk/cog-uk/; hosted by the COG-UK consortium), which currently includes >500 Spike amino acid replacements reported to confer antigenic change; 2) "Pokay" database (https://github.com/nodrogluap/pokay; hosted by University of Calgary, Canada), which is a literature-based compilation of hundreds of SARS-CoV-2 mutations potentially linked to multiple phenotypes. The *algn2pheno* tool allowed the rapid generation of user-friendly reports on repertoire of "relevant" mutations and their potential impact on phenotype, thus facilitating the identification and selection of viruses with potentially distinct phenotypic (e.g. antigenic) profiles for downstream comparative assays.

So far, three sets of experiments have been conducted (for a total of 24 viral strains), mainly including Delta VOC strains and early strains belonging to the B.1 lineage and related offspring. This tool was crucial for variant selection. In addition, the results of the animal work are currently being uploaded and connected to algn2pheno tool so that the observed phenotype is connected back to the selected genotype. Again, this additional projects' interconnection (COVRIN – TELEVIR) promoted by **COVRIN ST2.1.3 shows the advantage of building bridges between ongoing One Health EJP activities to enhance the identification of drivers for the emergence and spread of new pathogen threats and the generation of supportive data for risk assessment.**

3. References

1. Adriano Di Pasquale, Nicolas Radomski, Iolanda Mangone, Paolo Calistri, Alessio Lorusso, Cesare Camma, "SARS-CoV-2 surveillance in Italy through phylogenomic inferences based on Hamming distances derived from functional annotations of SNPs, MNPs and InDels". BMC Genomics 22, 782 (2021). https://doi.org/10.1186/s12864-021-08112-0
2. João D Santos, Carlijn Bogaardt, Joana Isidro, João P Gomes, Daniel Horton, Vítor Borges, "Algn2pheno: a bioinformatics tool for rapid screening of genetic features (nt or aa changes) potentially linked to specific phenotypes" (2022). https://github.com/insapathogenomics/algn2pheno
3. Verónica Mixão, Miguel Pinto, Daniel Sobral, Adriano Di Pasquale, João P Gomes, Vítor Borges. "ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data" (2022); https://github.com/insapathogenomics/ReporTree
4. Amato L, Jurisic L, Puglia I, Di Lollo V, Curini V, Torzi G, Di Girolamo A, Mangone I, Mancinelli A, Decaro N, Calistri P, Di Giallonardo F, Lorusso A, D'Alterio N. Multiple detection and spread of novel strains of the SARS-CoV-2 B.1.177 (B.1.177.75) lineage that test negative by a commercially available nucleocapsid gene real-time RT-PCR. Emerg Microbes Infect. 2021 Dec;10(1):1148-1155. doi: 10.1080/22221751.2021.1933609. PMID: 34019466; PMCID: PMC8205086.
5. Di Giallonardo F, Puglia I, Curini V, Cammà C, Mangone I, Calistri P, Cobbin JCA, Holmes EC, Lorusso A. Emergence and Spread of SARS-CoV-2 Lineages B.1.1.7 and P.1 in Italy. Viruses. 2021 Apr 29;13(5):794. doi: 10.3390/v13050794. PMID: 33946747; PMCID: PMC8146936.
6. Pascucci I, Paniccià M, Giammarioli M, Biagetti M, Duranti A, Campomori P, Smilari V, Ancora M, Scialabba S, Secondini B, Cammà C, Lorusso A. SARS-CoV-2 Delta VOC in a Paucisymptomatic Dog, Italy. Pathogens. 2022 Apr 26;11(5):514. doi: 10.3390/pathogens11050514. PMID: 35631035; PMCID: PMC9143276.
7. Amato L, Candeloro L, Di Girolamo A, Savini L, Puglia I, Marcacci M, Caporale M, Mangone I, Cammà C, Conte A, Torzi G, Mancinelli A, Di Giallonardo F, Lorusso A, Migliorati G, Schael T, D'Alterio N, Calistri P. Epidemiological and genomic findings of the first documented Italian outbreak of SARS-CoV-2 Alpha variant of concern. Epidemics. 2022 Jun;39:100578. doi: 10.1016/j.epidem.2022.100578. Epub 2022 May 13. PMID: 35636310; PMCID: PMC9098518.
8. Borges, V., Pinheiro, M., Pechirra, P. et al. INSaFLU: an automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance. Genome Med 10, 46 (2018). https://doi.org/10.1186/s13073-018-0555-0