MASTER THESIS OFFER OF CEPLAS
(FOR COMPUTATIONAL MASTER STUDENTS FROM TECHNICAL UNIVERSITY OF MADRID, UPM)

Title of Master Thesis

SARS-CoV-2 whole genome sequencing data analysis for national viral monitoring in public health.

Description of student´s tasks

● Analyse SARS-CoV-2 whole genome sequencing data collected in the national RELECOV project, using analysis pipelines written in Nextflow.

● Post-process the resulting data for its incorporation in the RELECOV's database platform.

● Use data mining to obtain relevant information from the data collected in the project.

● Generate data visualisation plots for its inclusion in the platform.

● Check the meeting of FAIR principles.

Prerequisites: *(languages, informatics skills, bioinformatic skills, other knowledge, etc)*

● Science or health related Grade (BsC)
● Python, R and bash programming languages basic knowledge
● Experience working in Linux terminal

Training Project

This Master Thesis is focused on the development of data analysis and integration protocols needed for the RELECOV project, which our group is co-leading with other units of the National Centre of Microbiology (CNM). The aim of the project is to create a Spanish national network for SARS-CoV-2 sequencing data, from the different autonomous countries, to monitor viral evolution and viral variants effect, which will serve for communication with the public health authorities, in order to ease decision making according to the epidemiological situation.

During this Master Thesis the student will learn how to configure Nextflow pipelines in a High Performance Computing (HPC) environment with a Slurm queue manager. Besides, they will get familiar with the working environment of a Bioinformatics analysis unit of a national public health care centre.

The master student will apply previously acquired knowledge to create post-processing protocols. This will allow the introduction of the data generated in the analysis in the already existing RELECOV project platform, complying with the requirements demanded by the platform.

The student will also improve self-learning skills, in addition to their familiarity with the development of data mining protocols. This will serve for the integration and visualisation of the information generated from the analysis of SARS-CoV-2 sequencing data, in the framework of a national viral surveillance project, always with the guidance of the unit supervisor.

Finally, they will use the knowledge acquired during their master about FAIR data to verify the suitability of the data within the platform with the FAIR principles, creating protocols to fit them in these principles when needed.

Activities that will be performed in the academic internship/ Master Thesis:

● Run nextflow pipelines in a HPC environment with Slurm queue manager to analyse SARS-CoV-2 whole genome sequencing data, to obtain variants, lineage and consensus information

● Post-process sequencing information and integrate it with samples metadata provided using different programming languages and automated processes, for its submission to the RELECOV platform.

● Use programming to automate the data visualisation, to obtain relevant information from the data collected in the project.

● Determine the suitability of the data with the FAIR principles.

● In case that data doesn't suit FAIR principles, create protocols to make it FAIR.

Nº of positions offered:	1
Has the student dealings with underage persons?	NO
Starting date:	01/02/2022
Fecha de fin: **(End date)**	31/06/2022 (Hasta llegar a 280h)
Horas semanales: **(Weekly hours)**	25
Horario jornada laboral: **(Working hours)**	Flexible
Importe Ayuda/Bolsa de estudio:	0 €/mes

(Amount of fellowship / remuneration)	
Tutor académico: (Academic tutor (UPM)) Email:	Joaquín Giner Lamia joaquin.giner@upm.es
Departamento tutor académico: **(Dept. of academic tutor)**	BIOTECNOLOGÍA - BIOLOGÍA VEGETAL
Tutor empresa: **(External tutor)**	Isabel Cuesta De La Plaza
Email tutor empresa: **(Email external tutor)**	isabel.cuesta@isciii.es
Departamento tutor empresa: **(Dept. of external tutor)**	Bioinformatics Unit from the Scientific-Technical Central Units
Ubicación de la estancia de las practicas **(Location of the internship)**	Ctra. de Pozuelo, 28, 28222 Majadahonda, Madrid
ENTIDAD COLABORADORA: **(Collaborating Entity)**	Instituto de Salud Carlos III
A cumplimentar por Oficina Prácticas ETSIAAB: **Créditos a reconocer (Nº ECTS)**:	