

## **Essentials of bioinformatics for functional genomics: basis and training**

### **Course description:**

With the rise of the Next-Generation Sequencing technologies, it is more accessible every day to sequence whole genomes at an affordable price. Research groups can get extensive genomic data from their interest species without asking big consortium and external sponsors.

The access to powerful new computer digital platforms, genomic databases and genome analysis free software are essential tools for future ground-breaking work in the Data Science analysis from the large biological biobanks that our centre holds.

We intend to bring more knowledge and capacity to our human resources, with new skills in genome analysis, so that the research centre can keep up with the digitalization and informatization of our research results.

Participants will explore the use of available resources and tools to manage, share, analyze and interpret metagenomics data for proper identification and annotation of the gene structures for further analysis.

Both eukaryotes and prokaryotes have genes hidden across their genomes differently, but the procedure to identify them is similar, although the effectiveness differs significantly.

In this practical course, the students will learn the particularities of gene identification in bacteria as well as in higher organisms. The system will include methodologies for a good annotation on the identified genes, description of the annotation terms and databases for retrieving high-quality annotations about processes, metabolic pathways, and protein domains to understand the gene function better.

The workshop is divided into two sections, a theoretical section and a “Hands-on” practical part with exercises.

The second section of the workshop involves participants learning via pre-recorded lectures, live presentations and instructor trainer Q&A sessions. Practical experience “Hands-on” will be developed through group activities and computational exercises.

## **Objectives**

- To create an awareness of a wide range of bioinformatics tools and sufficient experience to use those tools in basic investigations with a relatively high degree of user independence.
- To show the limitations and ambiguities found in commonly used Bioinformatics resources.
- To become aware of specific tools for NGS data analysis (transcriptomes, variant analysis).
- To create simple scripts that automate bioinformatics tasks, thus enlarging the scope of action of the tools applied to large datasets.
- To obtain results from the ongoing project

**Instructor** - António Munoz Mérida, Bioinformatics Expert, CIBIO-InBIO, Universidade do Porto, Campus de Vairão, Rua Padre Armando Quintas. 4485-661 Vairão, Portugal

**Duration** - 30 hrs (through one week)

**Date:** 24 a 30 agosto 2021

## **HANDS-ON demonstration applications (ONLINE WORKSHOP)**

The course will be online and no Basic understanding of molecular biology or particular computing expertise will be assumed.

Informatic requirements; Due to the demonstration requirement for High Performance Computing (HPC) for HANDS-ON Inscriptions are limited to 14 participants . Access to the workshop online platform and to the HPC virtual machines will be sent to participants e-mails a day before the course starts.

**Funding:** This workshop was supported by the project META-MINE "Mining the microbiomes from marine wood-digesting bivalves for novel lignocellulose depolymerizing enzymes" no âmbito da ERA-NET MARINEBIOTECH (MBT), Direção Regional da Ciência e Transição Digital-M2.2.a/F/03/2017

# Programme

24 a 30 agosto 2021

Horário 9:30 às 12:30 e das 13:30 16:30 hora dos Açores

## Metagenomics data Analysis I

- Introduction to Linux basic commands
- QC & host decontamination, assembly
- NGS Data Quality Control, Assembly and Pre-processing
- Read Mapping, Variant Prediction and Visualization of results

## Metagenomics data Analysis II

- Introduction to metagenomics data analysis. Sequence databases
- Basis on the gene identification in prokaryotes and eukaryotes
- Methods to predict gene structures in genomes. Exercises
- Genome Databases
- Hands-on- Exercises for gene finding (genome annotation)
- Available software for genome Analysis
- Exercises for Quality Control
- Exercises De-novo Assembly
- Gene Finding
- Hands-on- Exercises for gene Functional Annotation (gene annotation)
- Basis on the gene annotation, functional terms and databases
- Blast and interProScan
- Gene Annotation and Enrichment Analysis
- Functional interpretation
- Hands-on Metagenomics Taxonomic classification and Scripting
- Taxonomic Analysis, Differential Abundance
- Scripting and Automation of Bioinformatic tasks

- Pipelines and Workflows. Final wrap-up