DOMAR
PHD PROGRAM IN MARINE SCIENCE, TECHNOLOGY AND MANAGEMENT

# **Course title:** Omic Technologies: Variant Calling Using Sequencing Data

**Modality: CFT- Transversal Training Course**

**Orientation:**

- ☐ Sustainable use of Marine Resources
- ☐ Integral Management of the Sea

**Dates:**

Thursday, **September 28, 2023**
Friday, **September 29, 2023**

**Timetable:**

10:00-13:00
14:30-16:30

**Duration**:

**Lectures** totalling **10 hours**
**Online work** totalling **15 hours**.

**Location:**

DOMAR Videoconference Room, Torre-CACTI building, Campus As Lagoas-Marcosende, Universidade de Vigo (Spain)

**Language:** English

**Academic coordinators**:

| Name | Institution | e-mail |
|---|---|---|
| David Posada | Universidade de Vigo (Spain) | dposada@uvigo.es |

**Lecturers**:

| Name | Institution | e-mail |
|---|---|---|
| Fernando Cruz | CNAG-CRG (Barcelona, Spain) | fernando.cruz@cnag.crg. |

**General description:**

This course offers an overview of current massive parallel sequencing data paying special attention to those produced by the Illumina short-read sequencing technology from genomic DNA and its main applications to Genomics Research. Additionally, provides a basic training in bioinformatics techniques for handling, managing and analysing sequencing data. We will follow the main workflows of the Broad Institute for calling Single Nucleotide Variants (SNVs) or Indels with GATK. Additionally, we will do several exercises for handling, filtering and extracting information from VCFs and GVCFs. Finally, these data will allow the estimation of classical population genetics statistics, keeping in mind that they could also be computed at genomic scale.

**Contents:**

1. Classification of Genomic Variants.
2. Massive Parallel Sequencing Data.
3. Standard File Formats for sequencing and genomic variants data.

4. Bioinformatics Tools (FastQC, BWA, Picard, GATK, VCFTools…)
5. Mapping-based approach for Variant Calling (GATK Best Practices)
6. Progress in the field.

**Teaching methodologies:**
The course will focus on the basic tools and approaches for variant calling using massive parallel sequencing data, paying special attention to the Single Nucleotide Variants (SNVs). It will be organized in two classroom sessions of 5 hour each that combine theory and hands-on exercises. Students should bring their own computer with the required programs, a VM, conda environments or containers installed in advance. They will be complemented with 15 hours of online exercises that will require the students to work individually. The classes will include short practical exercises to facilitate the assimilation of theoretical concepts and methodological approaches.

**Evaluation system:**
The assistance and participation of the following items:

- Lectures: 20%
- Classroom exercises: 30%
- Online exercises: 50%

Contributions to the final qualification are shown as percentages.

**Brief CV of the lecturers:**
Fernando Cruz obtained his PhD at the Department of Genetics (University of Vigo). After carrying out several works in Sympatric Speciation, Population and Molecular Genetics, he obtained my PhD in Biology at the University of Vigo in 2005 with the Thesis entitled “Distribución genómica de microsatélites en moluscos” a study about the mutational and evolutionary dynamics of Tandem Repeats in Molluscs. During his postdoctoral research in European Universities (TCD, Uppsala and Lausanne), he applied bioinformatic approaches to study adaptive evolution, selective pressure in domestic species and mammalian genes. Back in Spain he joined the EBD-CSIC focusing on Population and Conservation Genomics. Since 2014, is member of the CNAG Genome Assembly and Annotation Team (Barcelona), where he is using large sequencing datasets to obtain high-quality reference genomes of non-model organisms with commercial and conservation interest. Additionally, he often estimates genome-wide genetic diversity and find ways to answer several questions about the assembled genomes. Fernando has contributed to 23 peer-reviewed publications on the fields of Bioinformatics, Genomics, Evolutionary Biology and Genome Assembly. Please visit https://publons.com/researcher/2711489/fernando-cruz/

**Relevant References:**
Four scientific works related with SNPs and Population Genomics:

1. Mediterranean Mussel Pangenome (2022). Genome Biology, DOI: 10.1186/s13059-020-02180-3
2. Lynx genome (2016). Genome Biology, DOI: 10.1186/s13059-016-1090-1
3. SNP Panel For The Iberian Lynx (2017). BMC Genomics, DOI: 10.1186/s12864-017-3946-5
4. Genomics of Dog Domestication (2008). MBE, DOI: 10.1093/molbev/msn177

**Teaching experience includes:**
Lecturer of this Doctorate Course “Variant Calling Using NGS Data” (Campus Do Mar, University of Vigo) in 2015, 2017 and 2019. Since course 2021-2022 he is teaching Algorithms for Sequence Analysis in Bioinformatics, a subject of the Bachelor’s Degree in Bioinformatics (ESCI-UPF).