2ND EDITION

COURSE ON

COMPUTATIONAL BIOSCIENCES USING HPC SYSTEMS

6, 7, 8 FEBRUARY 2024
@ NOVA SCHOOL OF SCIENCE AND TECHNOLOGY
DEPARTMENT BUILDING | ROOM 204

The Research Unit on Applied Molecular Biosciences (UCIBIO), in collaboration with the Associated Laboratory for Green Chemistry (LAQV) and the Laboratory of Instrumentation and Experimental Particle Physics (LIP), under the framework of the EuroCC Project, will instruct the 2nd Edition of the Course on Computational Biosciences using HPC systems - a hands-on course tailored for researchers that are interested in starting High Performance Computing (HPC) projects.

The full course covers a total of six parts, divided into two independent modules. Each module requires a registration and they're limited to 15 places each.
There is also an **optional**, bonus module, providing an introduction to HPC. The module is, however, required for those who don't have any knowledge on the matter.

MODULE 0
Introduction to HPC

MODULE 1
Omics

MODULE 2
Molecular Modeling

	FEB 5	FEB 6	FEB 7	FEB 8
OPTIONAL MODULE	10AM - 12PM **Introduction to HPC** online hands-on tutorial that teaches users how to use the National Distributed Computing Infrastructure's HPC cluster.	09AM - 12:30PM High-throughput Sequencing Data	09AM - 12:30PM Phylogenomics	09AM - 12:30PM Molecular Dynamics
		02PM - 05:30PM Transcriptome Assembly	02AM - 05:30PM Structure-based Virtual Screening	02AM - 05:30PM Hybrid Quantum Mechanics/Molecular Mechanics

REGISTER HERE

20€ PER MODULE
FREE LUNCH ON THE 7TH FOR THOSE WHO REGISTER FOR BOTH MODULES

UCIBIO Applied Molecular Biosciences Unit · i4HB Institute for Health and Bioeconomy · INCD Infraestrutura Nacional de Computação Distribuída · LAQV requimte · LIP · EURO² PORTUGAL · PRR Plano de Recuperação e Resiliência

ADITIONAL INFO ABOUT THE MODULES

MODULE 1

High-throughput Sequencing Data

Ana Rita Grosso | 3h | FastQC, Bowie, STAR and Kallisto (Docker images)

1. Presentation
   - Introduction to high-throughput sequencing data
   - Data repositories
   - Methods for alignment to genome and transcriptome
2. Hands-On
   - Download high-throughput sequencing data
   - Assess quality with FastQC
   - Alignment to transcriptome with Kallisto
   - Alignment to genome with Bowtie/STAR
3. Conclusions
   - What could be done downstream?

The development of high-throughput sequencing data has provided complete collections of genomes and transcriptomes, unveiling the complexity of cell biology and the underlying dysregulation of diseases. In this Module, researchers will learn how to obtain and analyze big data using High-Performance Computing. Such knowledge will provide them with the tools to complement their research projects with multi-omics profiles.

MODULE 1

Transcriptome Assembly

Pedro M. Costa | 3h | Trinity and Blast/Pfam (Docker images)

1. Presentation
   - Introduction to transcriptome assembly and annotation
2. Hands-On
   - Transcriptome assembly with Trinity
   - Open reading frame prediction with Trinity
   - Transcriptome annotation with Blast
3. Conclusions
   - What could be done downstream?

In this module, we will explore "hands on" the main steps needed to assemble transcriptomes using reference mapping as standard in model organisms from murines to the zebrafish; and, most importantly, assembly when reference mapping is not available. Essential steps, requirements, validation and pitfalls will be addressed.

MODULE 1

Phylogenomics

Patrícia H. Brito | 3h | Iqtree, RaxML

1. Presentation
   - Phylogenetic analysis using maximum-likelihood.
   - Phylogenomics and challenges of whole genome inferences
2. Hands-On
   - Phylogenomic analysis of concatenated datasets
   - Model selection and model partitioning
   - Strategies for high-performance computing in phylogenomics

Dataset: Multi-gene alignment of a group of tetrapodes.

Phylogenomics is the application of phylogenetic inference methods to large genomic datasets to infer the evolutionary history of species and their genes. These analyses provide the backbone of evolutionary studies aiming to study the speciation process, population demographics or gene family evolution. This module will provide the basics of phylogenetic inference using maximum-likelihood and Bayesian methods with a hands-on tutorial that takes advantage of HPC resources.

MODULE 2

Structure-based Virtual Screening

Arménio Barbosa, Sérgio Sousa | 3h | Autodock, Vina, Datawarrior

1. Brief introduction to structure-based drug discovery
2. Small molecule library curation
3. Protein preparation
4. Molecular docking
5. Virtual screening
6. Conclusion: Comparison of software performance in screening

Molecular System: Main Protease of SARS-CoV-2

Protein-ligand docking is applied to screen virtual databases of millions of compounds predicting ligand's binding properties to specific biomolecular targets of interest. In this module, researchers will learn how to prepare large virtual screening runs using protein-ligand docking tools taking advantage of HPC resources to screening thousands or millions of molecules.

ADITIONAL INFO ABOUT THE MODULES

MODULE 5

Molecular Dynamics

Arménio Barbosa, Sérgio Sousa | 3h | GROMACS, AMBER

1. Brief introduction to molecular dynamics simulations
2. GROMACS
   - System preparation
   - MD run scripts: GPU and mpi
   - Performance analysis
3. AMBER
   - System preparation
   - MD run scripts: GPU and mpi
   - Performance analysis
4. Conclusion: Comparison of software performance in GPU and mpi in the considered system

Molecular System: Main Protease of SARS-CoV-2

Molecular Dynamics simulations enable the study of the dynamic properties of biological systems through time with atomistic detail. In this module, researchers will learn how to take advantage of HPC and GPU resources to simulate the interaction between a potential drug-like molecule and an enzyme of pharmacological interest at the nanosecond to microsecond time scale.

MODULE 2

Hybrid Quantum Mechanics/Molecular Mechanics

Arménio Barbosa, Sérgio Sousa | 3h | ORCA

1. Brief Introduction to QM/MM Methods
2. Preparation of a QM/MM model
3. Geometry Optimization
4. Relaxed Scan
5. Transition State Characterization
6. Energy Profile Determination
7. Conclusion: QM method selection, parallelization approach, and size dimension

Molecular System: Main Protease of SARS-CoV-2

QM/MM methods allow the study of chemical reactions in large systems with great accuracy. This hybrid approach allows the study of the catalytic site with quantum mechanics, while the rest of the system is treated with classical molecular mechanics. This module will introduce the basics of QM/MM simulations in the study of biomolecules using HPC resources.