

63105 - Métodos en bioestadística y bioinformática

Información del Plan Docente

Año académico: 2020/21

Asignatura: 63105 - Métodos en bioestadística y bioinformática

Centro académico: 100 - Facultad de Ciencias

Titulación: 572 - Máster Universitario en Biotecnología Cuantitativa

Créditos: 4.0

Curso: 1

Periodo de impartición: Segundo semestre

Clase de asignatura: Optativa

Materia: ---

1. Información Básica

1.1. Objetivos de la asignatura

To learn how to select appropriate statistical frameworks to model some types of biological data.

To be acquainted with some primary computational tools for the study of genomes, genes and proteins.

To learn general-purpose scripting languages that can be applied to statistical and predictive analyses in molecular biology, genomics and biotechnology.

1.2. Contexto y sentido de la asignatura en la titulación

By using a practical approach, this course introduces some fundamental concepts and algorithms of computational biology and bioinformatics.

1.3. Recomendaciones para cursar la asignatura

Students are expected to be familiar with statistics and molecular biology, and should have basic programming skills.

2. Competencias y resultados de aprendizaje

2.1. Competencias

Relevant probabilistic models for the analysis of biological data.

Statistical inference methods for typical problems.

Construction and validation of predictive and classification models.

Main data formats in bioinformatics.

Key sequence and structure alignment algorithms.

Main approaches for computing phylogenetic trees.

Use of locally installed tools: BLAST+, BWA, HMMER, CLUSTAL-OMEGA, MAMMOTH, MODELTEST, PHYML.

R programming for analysing and plotting biological data.

Handling large data files with Perl one-liners.

2.2. Resultados de aprendizaje

In order to pass, the student should demonstrate the following skills:

1: He/she is able to select the appropriate statistical framework to model some types of biological data.

2: He/she is knowledgeable about fundamental computational tools for the study of genomes, genes and proteins, and their applications in biotechnology and biomedicine.

3: He/she can perform simple programming tasks in the context of biological data

2.3. Importancia de los resultados de aprendizaje

The skills developed while accomplishing these goals are valuable to understand the description of bioinformatics

procedures in the literature and during the course of research projects in this area.

3.Evaluación

3.1.Tipo de pruebas y su valor sobre la nota final y criterios de evaluación para cada prueba

Practical cases (50%): After each session, students are expected to return in time a report which will be marked by the responsible lecturer in order to track individual progress and work.

Class attendance (50%): Students are expected to participate and might optionally be given a final exam with short questions regarding any aspect of the contents of the course.

4.Metodología, actividades de aprendizaje, programa y recursos

4.1.Presentación metodológica general

Lectures will take place in a computer room so that students can carry out their practical work in sessions of 4h. Learning material will be distributed in English in the form of presentations and/or Markdown documents that can be displayed with Rstudio or any Internet browser. Theoretical and practical subjects will be combined in the same sessions.

4.2.Actividades de aprendizaje

Students will review the literature, produce source code to solve practical cases and discuss their solutions with the rest of the class.

4.3.Programa

The course will address the following topics:

Topic 1. Introduction to statistical methods and programming in the R language.

Topic 2. Review of some distributions of probability.

Topic 3. Classical inference and Bayesian methods.

Topic 4. Classification and prediction.

Topic 5. Markov chains and hidden models.

Topic 6. Model validation, false positives and negatives. Receiver Operating Characteristic curves.

Topic 7. Multiple testing and error control. Limitations of Bonferroni, alternatives.

Topic 8. Introduction to processing large data files in the terminal with Perl one-liners.

Topic 9. Bioinformatics file formats: FASTA, FASTQ, PDB/PDBML, Newick.

Topic 10. Dynamic programming algorithms for local and global alignments.

Topic 11. Sequence similarity searches in local databases.

Topic 12. Multiple sequence alignment of nucleotide and protein sequences.

Topic 13. Alignment and superposition of protein structures. Scores.

Topic 14. Phylogenetic trees from distance matrices and evolution models.

Topic 15. Design and analysis of RNAseq experiments

4.4.Planificación de las actividades de aprendizaje y calendario de fechas clave

The course is taught throughout the second semester, from February to June.

The first 20h are taught JL Ojeda Cabrera [jojeda@unizar.es].

The following 20h by B Contreras Moreira [bcontreras@eead.csic.es].

4.5.Bibliografía y recursos recomendados

R programming:

Gentleman,R. 2009. R Programming for Bioinformatics. CRC Press.

Sequence analysis:

Durbin,R., Eddy, S.R., Krogh,A., Mitchison,G. (1999) Biological Sequence Analysis. Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press.

Mäkinen, V., D. Belazzougui, F. Cunial, and A.I. Tomescu. 2015. Genome-Scale Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing. Cambridge University Press.

RStudio:

<https://www.rstudio.com>

Structural Bioinformatics:

http://eead-csic-compbio.github.io/bioinformatica_estructural

Perl:

<http://eead.csic.es/compbio/material/bioinfoPerl>