

**Computació d'Altes Prestacions i Anàlisi de Big Data**

Codi: 43917

Crèdits: 12

Titulació	Tipus	Curs	Semestre
4313473 Bioinformàtica	OT	0	1

### Professor/a de contacte

Nom: Miquel Àngel Senar Rosell

Correu electrònic: miquelangel.senar@uab.cat

### Idiomes dels grups

Podeu accedir-hi des d'aquest [enllaç](#). Per consultar l'idioma us caldrà introduir el CODI de l'assignatura. Tingueu en compte que la informació és provisional fins a 30 de novembre de 2023.

### Equip docent

José Eduardo Cabrera Díaz

Oscar Lao Grueso

Emanuele Raineri

### Prerequisits

Per a poder fer aquest mòdul és necessari haver aprovat els dos mòduls obligatoris: Programming in Bioinformatics i Core Bioinformatics.

Es recomana tenir el nivell B2 d'anglès o equivalent.

### Objectius

Aquest mòdul pretén proporcionar als estudiants els coneixements i habilitats necessàries (1) per implementar aproximacions d'enginyeria de rendiment a les plataformes d'informàtica modernes i (2) per realitzar anàlisis estadístiques de Big Data.

### Competències

- Comunicar en llengua anglesa de manera clara i efectiva els resultats de les pròpies investigacions.
- Dissenyar i aplicar la metodologia científica en la resolució de problemes.
- Proposar solucions bioinformàtiques a problemes derivats de les recerques òmiques.

- Proposar solucions innovadores i emprenedores en el seu camp d'estudi.
- Tenir coneixements que aportin la base o l'oportunitat de ser originals en el desenvolupament o l'aplicació d'idees, sovint en un context de recerca.
- Utilitzar i gestionar informació bibliogràfica i recursos informàtics en l'àmbit d'estudi.
- Utilitzar sistemes operatius, programes i eines d'ús comú en bioinformàtica, i fer servir plataformes de còmput d'altres prestacions, llenguatges de programació i anàlisis bioinformàtiques.

## Resultats d'aprenentatge

1. Aplicar mètodes estadístics avançats (aprenentatge automàtic, teoria de grafs) per modelar i analitzar problemes bioinformàtics amb dades biològiques massives.
2. Aprendre a entrenar, avaluar i validar models predictius.
3. Aprendre a fer servir les noves plataformes de còmput paral·lel, paradigmes, i el disseny d'aplicacions que requereixen un maneig massiu de còmput i dades.
4. Aprendre noves maneres de modelar, emmagatzemar, recuperar i analitzar tipus de dades abstractes (grafs).
5. Comunicar en llengua anglesa de manera clara i efectiva els resultats de les pròpies investigacions.
6. Conèixer els principis de la paral·lelització de processos.
7. Conèixer els principis de l'emmagatzemament i la gestió de dades massives.
8. Conèixer i aprendre a utilitzar eines de codi font obert per a l'anàlisi paral·lela, distribuïda i escalable mitjançant aprenentatge automàtic.
9. Descriure el funcionament, les característiques i les limitacions de les tècniques, les eines i les metodologies que permeten descriure, analitzar i interpretar l'enorme quantitat de dades produïdes per les tecnologies d'alt rendiment.
10. Descriure i aplicar tècniques de clusterització (clustering) i algorismes de classificació comuns
11. Dissenyar i aplicar la metodologia científica en la resolució de problemes.
12. Generar algorismes de computació paral·lela eficients i aplicacions per a la CID
13. Proporcionar solucions paral·leles a problemes bioinformàtics concrets.
14. Proposar solucions innovadores i emprenedores en el seu camp d'estudi.
15. Tenir coneixements que aportin la base o l'oportunitat de ser originals en el desenvolupament o l'aplicació d'idees, sovint en un context de recerca.
16. Utilitzar i gestionar informació bibliogràfica i recursos informàtics en l'àmbit d'estudi.

## Continguts

### Arquitectura Moderna d'Ordinadors

- Arquitectura de processadors de finalitat general i especialitzada
- Jerarquia de memòria
- Sistemes de clúster
- Infraestructures del núvol i virtualització de sistemes
- Sistema *Middleware* i marcs de programació

### Models de Programació Avançats

- Memòria compartida i programació paral·lela distribuïda
- Shell scripting avançat
- Usant eines del sistema per a anàlisis bioinformàtics
- Principis de l'enginyeria de rendiment (eines i mètodes)
- Computació d'Altes Prestacions amb Python
- Enginyeria de rendiment aplicada a algorismes i eines comunes de bioinformàtica (indexació del genoma, alineament de *reads*, ...).

### Anàlisi de **Big Data**

- Teoria i eines d'estadística avançades en anàlisi de *Big Data* (reducció de dimensionalitat, selecció de variables i Spark)
- Teoria i algorismes d'aprenentatge de màquines. Aplicacions en Bioinformàtica
- Modelat predictiu: mineria de dades, avaluació i validació de models
- Classificació de dades: aprenentatge de Bayes ingenu i arbres de decisió
- Aprenentatge de regles d'associació
- Anàlisi de clusterització: algoritme *k-means*
- Teoria de grafs per *Big Data*

## Metodologia

Seguint una aproximació basada en problemes, l'alumnat aprendrà sobre algorismes, mètodes i plataformes computacionals eficients i els mètodes estadístics que s'aplicaran als problemes de bioinformàtica que tracten amb *Big Data*.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

## Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes teòriques	38	1,52	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
Resolució de problemes a classe i tasques al laboratori biocomputacional	32	1,28	1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 15
Tipus: Autònomes			
Estudi autònom individual	226	9,04	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 16

## Avaluació

El sistema d'avaluació està organitzat en dues activitats principals. Hi haurà, a més, un examen de recuperació. Els detalls de les activitats són:

### Activitats d'avaluació principals

- Portafoli de l'estudiant (70%): treballs fet i presentats per l'alumnat al llarg del curs. Cap de les activitats d'avaluació individuals representarà més del 50% de la nota final.
- Prova teòrica o pràctica individual (30%): per a cadascun dels mòduls principals de l'assignatura s'establirà algun mecanisme d'avaluació individual mitjançant una prova oral o escrita.

### Examen de recuperació

Per poder participar en el procés de recuperació, l'alumnat haurà d'haver participat prèviament en com a mínim l'equivalent a dos terços de la nota final del mòdul en activitats d'avaluació. El professorat informará dels procediments i terminis per al procés de recuperació.

No avaluable

L'alumnat serà qualificat com a "No avaluable" quan el pes de l'avaluació en què ha participat sigui inferior a l'equivalent al 67% de la nota final del mòdul.

Avaluació única

Aquesta assignatura/mòdul no preveu el sistema d'avaluació única.

### Activitats d'avaluació continuada

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Prova teòrica i pràctica individual	30%	4	0,16	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15
Treballs fet i presentats per l'alumnat (portafoli de l'estudiant)	70%	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16

### Bibliografia

El professorat recomanarà la bibliografia actualitzada a cada sessió d'aquest mòdul, i els enllaços estaran disponibles a l'Àrea de l'Estudiant del lloc web oficial del MSc Bioinformatics

### Programari

Linux + SLURM i altres eines de l'entorn Linux

Python i eines del seu ecosistema

R i eines del seu ecosistema