

Programació en Bioinformàtica

Codi: 42401
Crèdits: 6

Titulació	Tipus	Curs	Semestre
4313473 Bioinformàtica / Bioinformatics	OB	0	1

La metodologia docent i l'avaluació proposades a la guia poden experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries.

Professor/a de contacte

Nom: Santiago Marco Sola

Correu electrònic: Santiago.Marco@uab.cat

Equip docent

Santiago Marco Sola

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: anglès (eng)

Prerequisits

Per al desenvolupament general del curs, es recomana tenir un nivell B2 d'anglès o similar.

Per al mòdul de programació, és molt recomanable tenir nocions bàsiques de Linux com a usuari (conèixer les eines bàsiques d'edició d'arxius de text i de gestió de carpetes i permissos)

Objectius

Els objectius generals d'aquest mòdul són l'aplicació de les eines i tècniques bàsiques per al desenvolupament en aquesta àrea. Es treballen les capacitats de resoldre reptes i adaptar-se a les tecnologies i paradigmes de la bioinformàtica.

Competències

- Dissenyar i aplicar la metodologia científica en la resolució de problemes.
- Identificar les necessitats bioinformàtiques dels centres de recerca i les empreses del sector de la biotecnologia i la biomedicina.
- Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit.
- 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.
- Treballar individualment i en equip en un context internacional i multidisciplinari.
- 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. Dissenyar i aplicar la metodologia científica en la resolució de problemes.

2. Dissenyar, analitzar i avaluar les prestacions d'infraestructures paral·leles i grans volums de dades.
3. Dominar la creació, la connexió, l'edició i la consulta de bases de dades mitjançant MySQL.
4. Gestionar plataformes paral·leles i bases de dades bioinformàtiques d'acord amb les necessitats existents.
5. Identificar els avantatges i les limitacions de la bioinformàtica i la importància de l'aplicació de noves tecnologies computacionals en investigacions òmiques.
6. Implementar algorismes i tècniques de càlcul estadístic, per a la gestió de grans volums de dades.
7. Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit.
8. 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.
9. Treballar individualment i en equip en un context internacional i multidisciplinari.
10. Utilitzar els llenguatges de programació en bioinformàtica R i Perl.

Continguts

1. Linux (Comandes i shell scripting)

Comandes bàsiques, gestió de usuari, gestió de software i sistema d'arxius

Eines de processament de text i de processament de text

Redireccions, pipes, i filters

Programació de shell scripts amb Bash

2. Llenguatges de programació

Introducció a la programació amb Python dins Bioinformatics

Variables, expressions, tipus de dades, operadors, construccions programàtiques i contextos

Functions, modulis, i subroutines

Programació recursiva

Entrada/Sortida

Depuració de codi

Altres llenguatges de programació: R

3. Estructures de dades i processament de dades

Estructures de dades bàsiques (strings, lists, tuple, sets i dictionaries)

Estructures de dades niades i objectes

Arbres i grafs

Modelatge i representació de dades en bioinformàtica

Formats bàsics en bioinformàtica (FASTQ, SAM, VCF)

Expressions regulars

4. Algorithms in bioinformatics

Introducció a la complexitat algorítmica

Algorismes "Divide and conquer"

Enumeració combinatòria i backtracking

Programació dinàmica

5. Eines i llibreries en bioinformàtica

Eines per a la visualització de dades

Introducció a Biopython

Introducció a NumPy i Pandas

**Llevat que les restriccions imposades per les autoritats sanitàries obliguin a una prioritització o reducció d'aquests continguts.*

Metodologia

La metodologia combinarà treball a l'aula, solució de problemes supervisada a classe, treball no supervisat en el laboratori i treball a realitzar de forma individual a partir de les lectures recomanades i entregues a realitzar. S'utilitzarà una plataforma virtual per a l'entrega dels informes dels treballs.

*La metodologia docent proposada pot experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries.

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			
Solució de problemes a l'aula	14	0,56	1, 3, 6, 7, 8, 10
Treball realitzat a l'aula	20	0,8	1, 2, 4, 6, 7, 8
Treball realitzat en el laboratori	12	0,48	1, 4, 5, 6, 8, 9
Tipus: Supervisades			
Treball realitzat al laboratori a partir de les lectures recomanades	15	0,6	1, 5, 7, 8, 9
Tipus: Autònomes			
Treball realitzat de forma setmanal sobre els entregables i materials proporcionats	83	3,32	1, 2, 6, 7, 9

Avaluació

La metodologia combinarà treball a classe, solució de problemes a classe, treball autònom en el laboratori i fora del laboratori. S'utilitzaran les plataformes virtuals pel seguiment del curs. Cap de les activitats d'avaluació individuals representarà més del 50% de la nota final.

Examen de recuperació

Per poder participar en el procés de recuperació, l'alumne 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. Cal notar que les activitats realitzades dins de classe de forma contínua no poden recuperar-se.

No evaluable

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

*L'avaluació proposada pot experimentar alguna modificació en funció de les restriccions a la presencialitat que imposin les autoritats sanitàries.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Avaluació del treball realitzat per l'alumne durant el mòdul	10%	1	0,04	1, 3, 5, 6, 7, 9, 10
Examens individuals teòrics i pràctics	50%	2	0,08	1, 2, 4, 7, 8
Exàmen final	30%	1	0,04	1, 7, 8
Treball al laboratori, individual i en grup	10%	2	0,08	1, 2, 4, 5, 6, 7, 9

Bibliografia

- Rajkumar Buyya, "High Performance Cluster Computing: Programming and Applications", PH, 1999.
- Bell, Charles; Kindahl, Mats; Thalmann, Lars. "MySQL High Availability". O'Reilly, 2010.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2002).
- GenBank. *Nucl. Acids Res.*, 30(1):17-20. URL <http://nar.oupjournals.org/cgi/content/abstract/30/1/17>.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.*, 28(1):235-242.
- Bessant, C., Shadford, I., Oakley, D. "Building Bioinformatics Solutions with Perl, R and MySQL", Oxford University Press, 2009
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. .The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, 31(1):365-370.
- Christiansen, P., Wall, L., Orwant, J., "Programming Perl". 4th Edition, O'Reilly, 2012
- Mäkinen et al., Genome-Scale Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing. Cambridge Univ. Press, 2015.
- Matloff, N., "The Art of R Programming". No Starch Press Inc., 2011
- Lutz, M., "Learning Python", O'Reilly, 5th edition, 2013
- Siever, E., Figgins, S., "Linux in a nutshell" O'Reilly 2009.
- Sobell, M., "A Practical Guide to Linux. Commands, editors and shell programming". Prentice Hall, 2009.
- Tindall, James., Beginning Perl for Bioinformatics. O'Reilly 2012.

Webs de referència:

- <http://mscbioinformatics.uab.cat>
- <https://cv.uab.cat>

Recursos digitals a la biblioteca UAB:

- <http://www.uab.cat/biblioteques/trobador>
- <http://pagines.uab.cat/bctdigital/>

Programari

Linux (Ubuntu, Bash, linux-tools, etc)
Python 2.7/3.x
Jupyter Notebook / PyCharm
R/RStudio
Matplotlib/Seaborn
Numpy/Pandas