

Core Bioinformatics

Codi: 42397

Crèdits: 12

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: Sònia Casillas Viladerrams

Correu electrònic: Sonia.Casillas@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: anglès (eng)

Equip docent

Antoni Barbadilla Prados

Leonardo Pardo Carrasco

Pere Puig Casado

Alfredo Ruíz Panadero

Miquel Àngel Senar Rosell

Jean-Didier Pierre Marechal

Isaac Salazar Ciudad

Oscar Conchillo Solé

Raquel Egea Sánchez

Equip docent extern a la UAB

Cedric Notredame

Emanuel Raineri

Josep Abril

Sebastián Ramos

Prerequisits

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

Objectius

Aquest mòdul se centra en el desenvolupament de diverses eines i recursos bioinformàtics utilitzats habitualment en la recerca de les òmiques. La nostra intenció és que cobreixi diversos aspectes de la bioinformàtica en una sèrie de temes breus, en forma de "tastets". Per tant, no és un mòdul d'accumulació, sinó transversal, que hauria de proporcionar una àmplia gamma d'idees i enfocaments que ofereix la

bioinformàtica, a través d'experts. L'objectiu principal és proporcionar als estudiants els fonaments necessaris per aplicar la bioinformàtica a diferents àrees de la investigació científica. Amb el temps, cada estudiant podrà obtenir tota la profunditat que es proposi sobre qualsevol d'aquests temes, el que finalment representi el seu marc de recerca.

Competències

- Analitzar i interpretar bioinformàticament les dades que es deriven de les tecnologies òmiques.
- Comprendre les bases moleculars i les tècniques experimentals estàndard més comunes en les recerques òmiques (genòmica, transcriptòmica, proteòmica, metabolòmica, interactòmica, etc.).
- 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.
- 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.
- 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. Buscar eines bioinformàtiques específiques i recursos bioinformàtics a la xarxa.
2. Comprendre les bases teòriques, estadístiques i biològiques de què parteixen els programes d'anàlisi bioinformàtica: alineament de seqüències, recerca per similitud i alineament múltiple, predicció d'estructura, anotació de genomes, anàlisi filogenètica i evolutiva.
3. Crear i promoure els algorismes, les tècniques de càlcul i estadística i la teoria per resoldre problemes formals i pràctics derivats de la gestió i l'anàlisi de dades biològiques.
4. Dissenyar i aplicar la metodologia científica en la resolució de problemes.
5. Identificar i aplicar els algorismes en què es basen els programes d'anàlisi bioinformàtica.
6. Identificar i caracteritzar els principals tipus de dades biomoleculars que s'obtenen de les tecnologies òmiques.
7. Proposar solucions innovadores i emprenedores en el seu camp d'estudi.
8. Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit.
9. Sintetitzar i interpretar, de manera lògica i raonada, la informació procedent de bases de dades moleculars i analitzar-la mitjançant eines bioinformàtiques.
10. 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.
11. Utilitzar i gestionar informació bibliogràfica i recursos informàtics en l'àmbit d'estudi.
12. Utilitzar les principals bases de dades moleculars i els principals formats estàndard de dades moleculars, i integrar dades de diferents fonts de dades

Continguts

BLOC 1. ESTADÍSTICA

Inferència estadística

Professor Antonio Barbadilla

- Estadística: pont entre dades i models.
- Tipus de dades
- Població i mostra
- Disseny experimental
- Qualitat de dades

- Exploració de dades
- Distribució de la mostra i llei de grans nombres
- Inferència estadística
- Teorema del límit central
- Estimació de punts
- Estimació de l'interval de confiança
- Hipòtesi
- Elements d'una prova: H_0 , H_1 , prova estadística, valor de p , nivell de significació, errors de tipus I i II, potència
- Prova Z, prova t, prova chi-quadrada, prova de correlació, regressió, anàlisi de variància
- Interpretació de significació estadística
- Proves paramètriques versus no paramètriques
- Selecció de la prova estadística apropiada (arbre de decisió)
- Proves multivariades
- Remostreig

Estadística i processos estocàstics per a l'anàlisi de seqüències.

Professor Pere Puig

a. Fonaments de probabilitat

Conjunts i esdeveniments. Propietats. La probabilitat condicional. Independència. Alfabet i seqüències. Models probabilístics.

b. El model multinomial.

Simulant una seqüència multinomial. Estimació de probabilitats.

c. El paquet seqinr

d. Models de cadena de Markov

Concepte i exemples. Classificació d'estats. Codi r. Simulant una seqüència de cadenes de Markov. Estimació de les probabilitats de transició. La probabilitat d'una seqüència. Usant la cadena de Markov per a la discriminació.

e. Cadenes de Markov d'ordre superior.

Concepte i exemples. Estimació de les probabilitats de transició. Comparació de les cadenes de Markov d'ordre superior.

f. Cadenes ocultes de Markov.

Concepte i exemples. Estimació de paràmetres. Estimació d'estats ocults.

g. Una introducció als models lineals generalitzats.

Conceptes bàsics de GLM. El model logístic. El model de Poisson.

Inferència bayesiana

Professor Emmanuele Raineri

1. Ajust de la corba.

- Estimació de paràmetres de distribucions de probabilitat: binomial, Poisson i gaussiana.
- Exemple: ajustar un conjunt de dades sorollós.
- Validació creuada, sobreajust i regularització.

2. Reducció dimensional.

- Anàlisi de components principals, escalament multidimensional.
- Exemple: distingir els tipus de cèl·lules utilitzant perfils de metilació.

3. Regressió del llaç.
- Selecció de variables en models lineals.
 - Regressió penalitzada: Llaç i Xarxa Elàstica.
 - Exemple: regressió de llaç en R.

BLOC 2. UTILITATS BÀSIQUES

El genoma humà

Professor Alfredo Ruiz

a. Introducció als genomes.

Genomes seqüenciats. Organització i grandària dels genomes eucariotes. Construint un genoma: mètodes NGS per genòmica i transcriptòmica.

b. El genoma humà: ¿on som ara?

Assemblatge actual del genoma humà. El projecte ENCODE: elements funcionals en el genoma humano. Contenido repetitiu del genoma humà.

Bases de dades i formats de seqüència

Professor Oscar Conchillo

a. Formats de seqüència

Nomenclatura. Editors de text. Format FASTA i les seves variants. Format sense format / pla. Format de seqüència de GenBank. Format de seqüència EMBL. GCG, NBRF / PIR, MSA, PHYLIP, NEXUS. Conversió de format.

b. Bases de dades

Concepte. Recerques booleanes. Els comodins i les expressions regulars. Identificadors i números d'accés. Clasificació. Recull de bases de dades NAR. GenBank i altres bases de dades NCBI. EMBL. DDBJ. Meta-bases de dades integrades. Principals bases de dades de nucleòtids, proteïnes, estructura, taxonomia, etc.

Enginyeria de software

Professor Miquel Àngel Senar

- a. Sistema de control de versions amb Git i GitHub.
- b. Estratègies de paral·lelització i HPC.
- c. Computació en el núvol amb Amazon Web Services

Workflows amb Galaxy

Professora Raquel Egea

a. Introducció als gestors de workflows.

Concepte, origen i disseny dels gestors de workflows. Patrons de workflows. Gestors de workflows existents. APIs i servidors web.

b. Galaxy: conceptes bàsics, interfície i usos pràctics.

BLOC 3. BIOINFORMÀTICA ESTRUCTURAL

Estructura de proteïnes

Professors Leonardo Pardo i Óscar Conchillo

a. Introducció

Aminoàcids, proteïnes i enllaços peptídics. Quatre nivells d'estructura de proteïnes. Plegament de proteïnes i estabilitat. Interaccions moleculars. Mètodes experimentals per a la determinació de l'estructura.

b. Motius i dominis

c. Anàlisi

Bases de dades UniProt, PDB, PFAM, CATH i SCOP. Alineament de proteïnes, morphing, superfícies moleculars, potencial electrostàtic molecular.

d. Membrana cel·lular

Proteïnes de la membrana, segments transmembrana.

Modelatge molecular

Professors Leonardo Pardo i Jean-Didier Maréchal

a. Modelatge per homologia

b. Modelatge molecular

Models atòmics. Energia potencial. Mecànica quàntica i molecular. Tècniques d'exploració conformacional.

BLOC 4. GENÒMICA

Alineaments de seqüències

Professor Cedric Notredame

a. Models d'evolució i comparació.

Rellotge molecular. Estructura i evolució de les proteïnes. Matrius de substitució.

b. Programació dinàmica comparacions de seqüències basades

Needleman i l'algoritme de Wunsch. Algorisme de Smith i Waterman. Càlcul de penalitzacions per gaps. Càlcul de l'espai lineal d'algoritmes rodolins.

c. Blast i recerques de base de dades

L'algoritme Blast. Valors-e i estimacions de significació estadística. Estratègies de cerca de base de dades. PSI-Blast i altres enfocaments evolutius.

d. Alineaments múltiples de seqüències: algoritmes i estratègies

Principals aplicacions dels alineaments múltiples de seqüències. Algorismes més comuns. Estratègies d'alineament múltiple de seqüències.

Cerca de gens i de regions reguladores

Professor Josep Abril

a. Predicció de gens

Anotació: concepte, bases de dades, problemes. Cerca de gens: cerca per senyal, cerca per contingut, enfocaments (ab-initio, recerca d'homologia, genòmica comparativa, NGS), avaluació de la precisió del programari.

b. Cerca de regions reguladores.

Motius d'ADN: coincidència exacta, expressions regulars, matrius de pes de posició, arbres de recerca, perfils, algoritmes aleatoris, logotips i pictogrames, programari per a la recerca de motius. Dominis reguladors. H istones. Arquitectura i xarxes CRMs. Xarxes reguladores. Meta-alineament. Conservació, empremta filogenètica i ombrejat filogenètic. NGS.

Genòmica de poblacions
Professor Alfredo Ruiz

- a. Genòmica poblacional sota neutralitat en una població finita.
Introducció. Deriva genètica. Mida efectiva de la població. Probabilitat de fixació de mutacions neutres.
- b. Genòmica de poblacions sota selecció.
Selecció natural. Probabilitat de fixació de mutacions seleccionades. Distribució de fitness de noves mutacions. Tasa d'evolució.
- c. Evolució adaptativa i grandària de la població.

Filogènia i evolució molecular
Professor Sebastián Ramos

- a. Models d'evolució de seqüències.
Seqüència d'ADN. Model de Jukes i Cantor. Models més realistes. Selecció del model.
- b. Filogènia
Concepte. Arbres d'espècies versus arbres de gens. Mètodes de reconstrucció d'arbres: mètodes de distància, parsimònia màxima, probabilitat màxima, inferència bayesiana. Suport. Filogenòmica. Construint arbres amb R.

Biologia de sistemes

Professor Isaac Salazar

- a. Biologia de sistemes de l'edat clàssica i genòmica.
El paradigma de la biologia de sistemes a la llum dels desenvolupaments tecnològics dels últims 100 anys. Colls d'ampolla en la integració de dades.
- b. Modelització matemàtica de circuits moleculars.
Models conceptuals. Des models conceptuals fins a models matemàtics. Formalismes matemàtics. Models basats en dades.
- c. Principis de disseny i organització en circuits moleculars.
Concepte de principi de disseny. Comparacions controlades matemàticament. Anàlisi de viabilitat. Espais de disseny. Biologia sintètica.

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

Metodologia

La metodologia combinarà classes magistrals, resolució de problemes pràctics i casos reals, treball al laboratori de computació, realització de treballs individuals i en equip, lectura d'articles relacionats amb els blocs temàtics i estudi autònom independent. S'utilitzarà la plataforma virtual.

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

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
-------	-------	------	--------------------------

Tipus: Dirigides

Classes teòriques	39	1,56	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Resolució de problemes a classe i tasques al laboratori biocomputacional	39	1,56	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Tipus: Supervisades			
Realització de treballs individuals i en grup	40	1,6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Tipus: Autònomes			
Estudi autònom individual	178	7,12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Avaluació

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

Activitats d'avaluació principals

- Portafoli de l'estudiant (55%): treball fet i presentat per l'alumne al llarg del curs. Cap de les activitats d'avaluació individuals representarà més del 50% de la nota final.
- Prova teòrica i pràctica individual (35%): hi haurà un examen al final d'aquest mòdul. Consistirà en una o dues qüestions de selecció múltiple o de resposta curta per part de cada professor del mòdul.
- Habilitats toves (10%): assistència, puntualitat i participació activa a classe.

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 habilitats toves no poden recuperar-se.

No avaluable

L'alumne 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.

**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
Habilitats suaus	10%	0	0	4, 7, 8, 10
Portafoli de l'estudiant	55%	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Test teòric i pràctic individual	35%	4	0,16	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

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.