

Core Bioinformatics

Código: 42397
Créditos ECTS: 12

Titulación	Tipo	Curso	Semestre
4313473 Bioinformática / Bioinformatics	OB	0	1

Contacto

Nombre: Sònia Casillas Viladerrams

Correo electrónico: Sonia.Casillas@uab.cat

Equipo docente

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

Uso de idiomas

Lengua vehicular mayoritaria: inglés (eng)

Equipo docente externo a la UAB

Cedric Notredame

Emanuel Raineri

Josep Abril

Sebastián Ramos

Prerequisitos

Se recomienda un nivel B2 de inglés o equivalente.

Objetivos y contextualización

Este módulo se centra en el desarrollo de diversas herramientas y recursos bioinformáticos comúnmente utilizados en la investigación de las ómicas. Nuestra intención es que cubra varios aspectos de la bioinformática en una serie de temas breves, en forma de "catas". Por lo tanto, no es un módulo acumulativo, sino uno transversal, que debe proporcionar una amplia gama de ideas y enfoques que ofrece la bioinformática, de la mano de expertos. El objetivo principal es proporcionar a los estudiantes la base necesaria para aplicar la bioinformática a diferentes áreas de la investigación científica. Con el tiempo, cada

estudiante podrá obtener toda la profundidad que se proponga sobre cualquiera de estos temas, el que finalmente represente su marco de investigación.

Competencias

- Analizar e interpretar bioinformáticamente los datos que se derivan de las tecnologías ómicas.
- Comprender las bases moleculares y las técnicas experimentales estándares más comunes en las investigaciones ómicas (genómica, transcriptómica, proteómica, metabolómica, interactómica, etc.).
- Diseñar y aplicar la metodología científica en la resolución de problemas.
- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- Proponer soluciones bioinformáticas a problemas derivados de las investigaciones ómicas.
- Proponer soluciones innovadoras y emprendedoras en su campo de estudio.
- Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
- Utilizar sistemas operativos, programas y herramientas de uso común en bioinformática, así como, manejar plataformas de cómputo de altas prestaciones, lenguajes de programación y análisis bioinformáticos.
- Utilizar y gestionar información bibliográfica y recursos informáticos en el ámbito de estudio.

Resultados de aprendizaje

1. Buscar herramientas bioinformáticas específicas y recursos bioinformáticos en la red.
2. Comprender las bases teóricas, estadísticas y biológicas, en las que se basan los programas de análisis bioinformático: alineamiento de secuencias, búsqueda por similitud y alineamiento múltiple, predicción de estructura, anotación de genomas, análisis filogenético y evolutivo.
3. Crear y promover los algoritmos, las técnicas de cálculo y estadística y la teoría para resolver problemas formales y prácticos derivados de la gestión y el análisis de datos biológicos.
4. Diseñar y aplicar la metodología científica en la resolución de problemas.
5. Identificar y aplicar los algoritmos en los que se basan los programas de análisis bioinformático.
6. Identificar y caracterizar los principales tipos de datos biomoleculares que se obtienen de las tecnologías ómicas.
7. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
8. Proponer soluciones innovadoras y emprendedoras en su campo de estudio.
9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
10. Sintetizar e interpretar, de forma lógica y razonada, la información procedente de bases de datos moleculares y analizarla mediante herramientas bioinformáticas.
11. Usar las principales bases de datos moleculares, los principales formatos estándares de datos moleculares e integrar datos de distintas fuentes de datos
12. Utilizar y gestionar información bibliográfica y recursos informáticos en el ámbito de estudio.

Contenido

BLOQUE 1. ESTADÍSTICA

Inferencia estadística

Profesor Antonio Barbadilla

- Estadística: puente entre datos y modelos.
- Tipos de datos
- Población y muestra
- Diseño experimental
- Calidad de datos
- Exploración de datos

- Distribución de la muestra y ley de grandes números
- Inferencia estadística
- Teorema del límite central
- Estimación de puntos
- Estimación del intervalo de confianza
- Hipótesis
- Elementos de una prueba: H_0 , H_1 , prueba estadística, valor de p , nivel de significación, errores de tipo I y II, potencia
- Prueba Z, prueba t, prueba chi-cuadrada, prueba de correlación, regresión, análisis de varianza
- Interpretación de significación estadística
- Pruebas paramétricas versus no paramétricas
- Selección de la prueba estadística apropiada (árbol de decisión)
- Pruebas multivariadas
- Remuestreo

Estadística y procesos estocásticos para el análisis de secuencias.

Profesor Pere Puig

a. Fundamentos de probabilidad

Conjuntos y eventos. Propiedades La probabilidad condicional. Independencia. Alfabeto y secuencias. Modelos probabilísticos.

b. El modelo multinomial.

Simulando una secuencia multinomial. Estimación de probabilidades.

c. El paquete seqinr

d. Modelos de cadena de Markov

Concepto y ejemplos. Clasificación de estados. Código r Simulando una secuencia de cadenas de Markov. Estimación de las probabilidades de transición. La probabilidad de una secuencia. Usando la cadena de Markov para la discriminación.

e. Cadenas de Markov de orden superior.

Concepto y ejemplos. Estimación de las probabilidades de transición. Comparación de las cadenas de Markov de orden superior.

f. Cadenas ocultas de Markov.

Concepto y ejemplos. Estimación de parámetros. Estimación de estados ocultos.

g. Una introducción a los modelos lineales generalizados.

Conceptos básicos de GLM. El modelo logístico. El modelo de Poisson.

Inferencia bayesiana

Profesor Emmanuele Raineri

1. Ajuste de la curva.

- Estimación de parámetros de distribuciones de probabilidad: binomial, poisson y gaussiana.
- Ejemplo: ajustar un conjunto de datos ruidoso.
- Validación cruzada, sobreajuste y regularización.

2. Reducción dimensional.

- Análisis de componentes principales, escalamiento multidimensional.
- Ejemplo: distinguir los tipos de células utilizando perfiles de metilación.

3. Regresión del lazo.

- Selección de variables en modelos lineales.
- Regresión penalizada: Lazo y Red Elástica.
- Ejemplo: regresión de lazo en R.

BLOQUE 2. UTILIDADES BÁSICAS

El genoma humano

Profesor Alfredo Ruiz

a. Introducción a los genomas.

Genomas secuenciados. Organización y tamaño de los genomas eucariotas. Construyendo un genoma: métodos NGS para genómica y transcriptómica.

b. El genoma humano: ¿dónde estamos ahora?

Ensamblaje actual del genoma humano. El proyecto ENCODE: elementos funcionales en el genoma humano. Contenido repetitivo del genoma humano.

Bases de datos y formatos de secuencia

Profesor Oscar Conchillo

a. Formatos de secuencia

Nomenclatura. Editores de texto. Formato FASTA y sus variantes. Formato sin formato / plano. Formato de secuencia de genbank. Formato de secuencia EMBL. GCG, NBRF / PIR, MSA, PHYLIP, NEXUS. Conversión de formato.

b. Bases de datos

Concepto. Búsquedas booleanas. Los comodines y las expresiones regulares. Identificadores y números de acceso. Clasificación. Recopilación de bases de datos NAR. GenBank y otras bases de datos NCBI. EMBL. DDBJ. Meta-bases de datos integradas. Principales bases de datos de nucleótidos, proteínas, estructura, taxonomía, etc.

Ingeniería de software

Profesor Miquel Àngel Senar

a. Sistema de control de versiones con Git y GitHub.

b. Estrategias de paralelización y HPC.

c. Computación en la nube con Amazon Web Services

Workflows con Galaxy

Profesora Raquel Egea

a. Introducción a los gestores de workflows.

Concepto, origen y diseño de los gestores de workflows. Patrones de workflows. Gestores de workflows existentes. APIs y servidores web.

b. Galaxy: conceptos básicos, interfaz y usos prácticos.

BLOQUE 3. BIOINFORMÁTICA ESTRUCTURAL

Estructura proteica

Profesores Leonardo Pardo y Óscar Conchillo

a. Introducción

Aminoácidos, proteínas y enlaces peptídicos. Cuatro niveles de estructura proteica. Plegamiento y estabilidad de proteínas. Interacciones moleculares. Métodos experimentales para la determinación de la estructura.

b. Motivos y dominios.

c. Análisis

Bases de datos UNIPROT, PDB, PFAM, CATH y SCOP. Alineamiento de proteínas, morphing, superficies moleculares, potencial electrostático molecular.

d. Membrana celular

Proteínas de membrana, segmentos transmembrana.

Modelado molecular

Profesores Leonardo Pardo y Jean-Didier Maréchal

a. Modelado por homología

b. Modelado molecular

Modelos atómicos. Energía potencial. Mecánica cuántica y molecular. Técnicas de exploración conformacional.

BLOQUE 4. GENÓMICA

Alineación de secuencia

Profesor Cedric Notredame

a. Modelos de evolución y comparación.

Reloj molecular. Estructura y evolución de las proteínas. Matrices de sustitución.

b. Programación dinámica comparaciones de secuencias basadas

Needlman y el algoritmo de Wunsch. Algoritmo de Smith y Waterman. Cálculo de penalizaciones por gaps. Cálculo del espacio lineal de algoritmos pareados.

c. Blast y búsquedas de base de datos

El algoritmo Blast. Valores-e y estimaciones de significación estadística. Estrategias de búsqueda de base de datos. PSI-Blast y otros enfoques evolutivos.

d. Alineamientos múltiples de secuencias: algoritmos y estrategias

Principales aplicaciones de los alineamientos múltiples de secuencias. Algoritmos más comunes. Estrategias de alineamiento múltiple de secuencias.

Búsquedas de genes y de regiones reguladoras

Profesor Josep Abril

a. Predicción de genes

Anotación: concepto, bases de datos, problemas. Búsqueda de genes: búsqueda por señal, búsqueda por contenido, enfoques (ab-initio, búsqueda de homología, genómica comparativa, NGS), evaluación de la precisión del software.

b. Búsqueda de regiones reguladoras.

Motivos de ADN: coincidencia exacta, expresiones regulares, matrices de peso de posición, árboles de búsqueda, perfiles, algoritmos aleatorios, logotipos y pictogramas, software para la búsqueda de motivos. Dominios reguladores. Histonas Arquitectura y redes CRMs. Redes reguladoras. Meta-alineamiento. Conservación, huella filogenética y sombreado filogenético. NGS.

Genómica de poblaciones

Profesor Alfredo Ruiz

a. Genómica poblacional bajo neutralidad en una población finita.

Introducción. Deriva genética. Tamaño efectivo de la población. Probabilidad de fijación de mutaciones neutras.

b. Genómica de poblaciones bajo selección.

Selección natural. Probabilidad de fijación de mutaciones seleccionadas. Distribución de fitness de nuevas mutaciones. Tasa de evolución.

c. Evolución adaptativa y tamaño de la población.

Filogenia y evolución molecular

Profesor Sebastián Ramos

a. Modelos de evolución de secuencias.

Secuencia de ADN Modelo de Jukes y Cantor. Modelos más realistas. Selección del modelo.

b. Filogenia

Concepto. Árboles de especies versus árboles de genes. Métodos de reconstrucción de árboles: métodos de distancia, parsimonia máxima, probabilidad máxima, inferencia bayesiana. Apoyo. Filogenómica.

Construyendo árboles con R.

Biología de sistemas

Profesor Isaac Salazar

a. Biología de sistemas de la edad clásica y genómica.

El paradigma de la biología de sistemas a la luz de los desarrollos tecnológicos de los últimos 100 años.

Cuellos de botella en la integración de datos.

b. Modelización matemática de circuitos moleculares.

Modelos conceptuales. Desde modelos conceptuales hasta modelos matemáticos. Formalismos matemáticos. Modelos basados en datos.

c. Principios de diseño y organización en circuitos moleculares.

Concepto de principio de diseño. Comparaciones controladas matemáticamente. Análisis de viabilidad.

Espacios de diseño. Biología sintética.

Metodología

La metodología combinará clases magistrales, resolución de problemas prácticos y casos reales, trabajo en el laboratorio de computación, trabajo individual y en equipo, lectura de artículos relacionados con los bloques temáticos y autoestudio independiente. Se utilizará la plataforma virtual.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases teóricas	39	1,56	1, 2, 3, 4, 5, 6, 8, 9, 10, 7, 12, 11
Resolución de problemas en clase y tareas en el laboratorio biocomputacional	39	1,56	1, 2, 3, 4, 5, 6, 8, 9, 10, 7, 12, 11
Tipo: Supervisadas			
Realización de trabajos individuales y en grupo	40	1,6	1, 2, 3, 4, 5, 6, 8, 9, 10, 7, 12, 11

Tipo: Autónomas

Evaluación

El sistema de evaluación está organizado en tres actividades principales. Habrá, además, un examen de recuperación. Los detalles de las actividades son:

Actividades de evaluación principales

- Portafolio del estudiante (55%): trabajo hecho y presentado por el alumno a lo largo del curso. Ninguna de las actividades de evaluación individuales representará más del 50% de la nota final.
- Prueba teórica y práctica individual (35%): habrá un examen al final de este módulo. Consistirá en una o dos cuestiones de selección múltiple o de respuesta corta por parte de cada profesor del módulo.
- Habilidades blandas (10%): asistencia, puntualidad y participación activa en clase.

Examen de recuperación

Para poder participar en el proceso de recuperación, el alumno deberá previamente haber participado en como mínimo el equivalente a dos tercios de la nota final del módulo en actividades de evaluación. El profesorado informará de los procedimientos y plazos para el proceso de recuperación. Nótese que las habilidades blandas no pueden recuperarse.

No evaluable

El alumno será calificado como "No evaluable" cuando el peso de la evaluación en la que ha participado sea inferior al equivalente al 67% de la nota final del módulo.

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Habilidades blandas	10%	0	0	4, 8, 9, 7
Portafolio del estudiante	55%	0	0	1, 2, 3, 4, 5, 6, 8, 9, 10, 7, 12, 11
Test teórico y práctico individual	35%	4	0,16	1, 2, 3, 4, 5, 6, 8, 9, 10, 7, 12, 11

Bibliografía

El profesor recomendará la bibliografía actualizada en cada sesión de este módulo, y los enlaces se pondrán a disposición en el Área del Estudiante del sitio web oficial de MSc Bioinformatics (http://mscbioinformatics.uab.cat/base/base3.asp?sito=bioinformaticsintranet&anar=module_2&item=&subitem=)