Universidad de Alcalá

GUÍA DOCENTE

INTRODUCCIÓN AL ANÁLISIS BIOINFORMÁTICO DE DATOS MOLECULARES

Asignatura Transversal

Universidad de Alcalá

Curso Académico 2022/23

1er Cuatrimestre

Aprobada en Junta de Facultad 30/05/2022

GUÍA DOCENTE

Nombre de la asignatura: Introducción al análisis bioinformático de datos moleculares
Código: 100245
Titulaciones en la que se imparte: Grado en Biología, Grado en Biología Sanitaria
Departamento: Ciencias de la Vida Áreas de Conocimiento: Fisiología Vegetal
Carácter: Transversal
Créditos ECTS: 6
Curso y cuatrimestre: 3er y 4º curso / primer cuatrimestre
Profesorado: Eva María del Campo López (Coordinadora)
Horario de Tutoría: Las tutorías presenciales se concertarán previa cita con el docente. Las tutorías no presenciales a través del Aula virtual o cualquier otro medio electrónico no tendrán horario determinado de atención.
Idioma en el que se imparte: Español

1. PRESENTACIÓN

En la última década se ha producido un incremento exponencial de la disponibilidad de aplicaciones informáticas que son útiles para el manejo de la información obtenida en la investigación biológica. En particular, aquellas que analizan secuencias de nucleótidos tanto de genes de interés como de genomas completos. Estas secuencias proceden de diversos tipos de organismos tanto procariotas como eucariotas; estos últimos dotados de hasta tres genomas como es el caso de los organismos vegetales (nuclear, plastidial y mitocondrial). Este conjunto de secuencias contiene una cantidad ingente de información de utilidad en investigación básica, aplicaciones sanitarias y biotecnológicas así como en estudios de biodiversidad. La información disponible ha adquirido tales proporciones que ha provocado el aumento de la demanda de aplicaciones informáticas por parte de la comunidad científica en general que faciliten el análisis de aspectos muy concretos dependiendo del objetivo final de cada estudio particular. Por todo lo expuesto, actualmente es imprescindible un conocimiento básico de las herramientas bioinformáticas de uso más general para poder abordar cualquier trabajo experimental en diferentes campos de la investigación en biología y medicina entre otras.

El objetivo general de esta asignatura es que el estudiante conozca y aprenda a utilizar con eficiencia las herramientas bioinformáticas más comúnmente utilizadas en el análisis de datos moleculares (genómicos, transcriptómicos, proteómicos y metabolómicos). La asignatura será impartida mediante clases magistrales combinadas con actividades y seminarios relacionados con el contenido de las mismas. El estudiante tendrá la oportunidad de resolver cada uno de los problemas planteados a lo largo del desarrollo de la asignatura, haciendo uso de los créditos tanto presenciales como online.

Prerrequisitos y Recomendaciones:

Se recomiendan conocimientos básicos en biología molecular / genética y un nivel de inglés adecuado para la lectura y comprensión de textos científicos.

2. COMPETENCIAS

Competencias genéricas:

1. Adquirir habilidades de aprendizaje que permitan a los estudiantes continuar estudiando de un modo que habrá de ser en gran medida auto-dirigido o autónomo.
2. Adquirir 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.
3. Aplicar los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o interdisciplinarios.
4. Comprender y aplicar adecuadamente de las técnicas necesarias para la resolución de problemas concretos en la investigación en ciencias.
5. Utilizar la lógica, la crítica y autocrítica.
6. Tratar adecuadamente el extenso volumen de información existente en diversas fuentes y recursos.
7. Adquirir habilidad para sintetizar teorías, publicaciones, textos…
8. Trabajar en equipo.
9. Comprender y respetar la diversidad de opiniones y criterios.
10. Comunicarse adecuadamente: expresión oral y escrita.

Competencias específicas:

1. Conocer el desarrollo histórico y los principios básicos de las principales herramientas y recursos bioinformáticos utilizados en investigación.
2. Conocer los principales formatos de datos de secuencias y ser capaz de interpretar la documentación obtenida de bases de datos.
3. Anotar secuencias obtenidas “de novo” y publicarlas en bases de datos internacionales.
4. Realizar análisis bioinformático de secuencias de ácidos nucleicos.
5. Realizar el análisis bioinformático de secuencias de proteínas.
6. Realizar alineamientos globales y locales de secuencias e interpretar la información obtenida.
7. Conocer algoritmos y significado estadístico de los alineamientos de secuencias.

8. Conocer las principales herramientas bioinformáticas utilizadas para realizar reconstrucciones filogenéticas.
9. Conocer las principales herramientas bioinformáticas utilizadas para el análisis de genomas, transcriptomas y proteomas.
10. Conocer y ser capaz de utilizar los principales recursos bioinformáticos para realizar estudios funcionales y estructurales in silico, y analizar resultados de estudios de expresión génica a gran escala.
11. Conocer y ser capaz de utilizar las principales bases de datos moleculares y bibliográficos para documentar y diseñar un abordaje experimental.
12. Elaborar proyectos de investigación que impliquen la obtención de información a partir del análisis bioinformático de secuencias de nucleótidos y proteínas.

3. CONTENIDOS

Bloques de contenido	Horas
Bloque I (Temas 1 a 4)	16
Bloque II (Temas 5 a 8)	16
Bloque III (Temas 9 a 13)	18
Temas *	
Tema 1. Introducción. Nacimiento y desarrollo histórico de la bioinformática. Análisis y manejo de datos genómicos y de patrones de expresión. Bioinformática estructural: plegamiento, modelado y diseño de biomoléculas. Redes biológicas y biología de sistemas. Servicios de flujos de trabajo. Webs de información científica.	
Tema 2. Bases de datos biológicos I. Datos genómicos. Formatos de secuencias. Bases de datos primarias. Bases de datos secundarias. Bases de datos de secuencias de nucleótidos. Bases de datos de secuencias de proteínas. Interconexiones entre bases de datos biológicas.	
Tema 3. Bases de datos biológicos II. Bases de datos especializadas. Bases de datos de anotaciones de genes y genomas. Bases de datos de anotaciones de proteínas. Bases de datos de rutas metabólicas. Bases de datos de organismos modelo. Recuperación de información de bases de datos biológicas.	
Tema 4. Navegadores del genoma. Navegadores generales (Ensembl, UCSC, NCBI Map Viewer, Phytozome, Gramene, otros). Navegadores especie-específicos. Clasificación en función de la tecnología subyacente. Funcionalidades y características. Visualización y recuperación de datos.	
Tema 5. Métodos de obtención de secuencias de ADN. Secuenciación de primera generación (Maxam-Gilbert y Sanger). Automatización de la secuenciación de ADN. Secuenciación de segunda generación (NGS) (pirosecuenciación, Roche 454, Illumina y SOLID). Tecnología de secuenciación de tercera generación (NNGS) (Ion Torrent, semiconductor, DNA nanoball, Heliscope, SMRT, Nanopore, otras).	

Tema 6. Ensamblaje y anotación de genes, genomas y transcriptomas. Tipos de ensamblaje. Introducción al ensamblaje de novo. Estrategia básica del ensamblaje de lecturas cortas. Problemas del ensamblaje de novo de lecturas cortas y soluciones posibles. Evaluación de la calidad y refinado del ensamblaje de novo. Anotación de genes en genomas y transcriptomas. Preparación y envío de secuencias a bases de datos internacionales.
Tema 7. Evaluación de la similitud y alineamientos de secuencias. Bases evolutivas de los alineamientos de secuencias biológicas. Homología, identidad y similaridad de secuencias. Alineamientos globales y locales. Alineamientos de múltiples secuencias. Algoritmos y significado estadístico de los alineamientos de secuencias. Comparación de secuencias: sintenia y evolución molecular. Búsqueda en bases de datos de secuencias con versiones del algoritmo de Smith-Waterman (BLAST y FASTA).
Tema 8. Evolución molecular y análisis filogenético. Selección de marcadores moleculares para análisis filogenéticos. Nucleótidos versus proteínas. Refinamiento de alineamientos múltiples para análisis filogenéticos. Modelos evolutivos. Programas de selección de modelos evolutivos. Métodos de análisis filogenético. Métodos basados en caracteres. Métodos basados en distancias. Método Bayesiano. Herramientas para la construcción de árboles filogenéticos. Estimación de la confiabilidad de árboles filogenéticos.
Tema 9. Introducción a la bioinformática funcional. Análisis en serie de la expresión génica (SAGE). Tecnología de microarrays. Métodos de RNA-seq para análisis de perfiles de expresión génica. Análisis bioinformático e Interpretación biológica de los datos de expresión génica. Aplicaciones y limitaciones de cada una de las tecnologías. Detección de modificaciones epigenéticas.
Tema 10. Proteómica y análisis de proteínas. Predicción de la estructura. Modelado por homología. Validación de la predicción estructuras. Predicción de propiedades fisico-químicas. Predicción de digestibilidad por proteasas. Predicción del perfil hidropático y antigenicidad. Predicción de modificaciones post-traduccionales. Predicción de dominios y motivos. Predicción de la localización subcelular.
Tema 11. Introducción a las redes biológicas. Redes de interacción proteína-proteína (PPIN). Redes de interacción enfermedad-gen. Redes metabólicas. Redes reguladoras de genes. Redes de co-expresión genética.
Tema 12. Metabolómica. Métodos analíticos en metabolómica. Metabolómica no dirigida o global. Metabolómica dirigida. Diseño experimental. Procesamiento y análisis de datos de metabolómica. Estándares de informes de datos de metabolómica. Bases de datos y repositorios de datos de metabolómica. Servidores y software de análisis de datos de metabolómica. Bases de datos de rutas metabólicas. Aplicaciones potenciales de la metabolómica.
Tema 13. Impacto de la bioinformática en diversos ámbitos. Diagnóstico genético. Cáncer y enfermedades infecciosas. Descubrimiento y diseño de fármacos. Medicina regenerativa. Ciencias forenses. Investigación de organismos vegetales como factorías de productos biosanitarios. Mejora de cultivos agrícolas. Detección y vigilancia de patógenos de plantas. Estudios de biodiversidad.
* Los descriptores de los temas pueden variar un poco en función de la respuesta de los estudiantes durante el desarrollo de la asignatura.

4. METODOLOGÍAS DE ENSEÑANZA-APRENDIZAJE.- ACTIVIDADES FORMATIVAS

4.1. Distribución de créditos (especificar en horas)

Número de horas presenciales:	33 h clases teóricas 17 h seminarios
Número de horas del trabajo propio del estudiante:	100 horas
Total horas	150 horas

4.2. Estrategias metodológicas, materiales y recursos didácticos

Actividades no presenciales	**Tareas online a través del Aula virtual: carpeta de tareas** Consistirán en la realización de las actividades-problema planteadas en las clases presenciales y presentación de un informe a través del Aula virtual.
Actividad presencial	**Clases magistrales:** Consistirán en la explicación de las bases teóricas por parte del docente mediante exposición oral con apoyo de un ordenador y un proyector. **Seminarios, que comprenden:** **1- Actividades-problema:** Estas actividades serán planteadas por el docente. Serán resueltas por el estudiante mediante el uso de programas informáticos específicos, instrucciones concretas y la guía del docente. **2- Presentaciones de los estudiantes:** Consistirán en la preparación individual o por grupos de estudiantes de una presentación y su exposición oral basada en resultados obtenidos en al aula de informática y/o de artículos científicos.
Tutorías	Atención personalizada tanto presencial como a distancia sin horario preestablecido.

Estrategias metodológicas:

Clases teóricas:

Consistirán en la exposición oral por parte del docente de los contenidos con el apoyo de un ordenador, un proyector y pizarra. En estas clases se explicarán los aspectos más relevantes en relación con el tema que se trate sin ser excesivamente exhaustivo, evitando la sobrecarga de contenidos. Se seleccionarán contenidos relevantes para la resolución de problemas reales animando al estudiante a profundizar de forma autónoma en aquellos aspectos que más le interesen mediante recursos informáticos y bibliográficos.

Seminarios:

Los seminarios se destinarán a complementar lo aprendido en las clases teóricas con la resolución de actividades-problema y presentaciones que los estudiantes prepararán individualmente o por grupos y que serán expuestas oralmente.

- La metodología docente utilizada en las actividades problema será la conocida como "Aprendizaje basado en problemas". La implementación de esta metodología implicará que un grupo de estudiantes de manera autónoma, aunque guiados por el docente, deberá encontrar la solución a un problema de forma que el conseguir resolverlo correctamente supondrá la necesidad de buscar, comprender, integrar y aplicar los conceptos y métodos explicados en las clases magistrales relacionados con el contenido del problema. Con esta metodología docente, el estudiante es el principal protagonista del proceso de enseñanza-aprendizaje al verse obligado a identificar sus necesidades de aprendizaje y buscar el conocimiento para poder resolver un problema planteado, lo que a su vez origina nuevas necesidades de aprendizaje. Asimismo, esta metodología pretende contribuir a que los estudiantes consigan evaluar sus necesidades de aprendizaje, construir el conocimiento de la materia y trabajar cooperativamente.

Tutorías individualizadas:

En ellas los estudiantes podrán presentar sus dudas al docente ya sea mediante entrevistas presenciales o a través de Internet.

5. EVALUACIÓN: Procedimientos, criterios de evaluación y de calificación[1]

PROCEDIMIENTO

El estudiante dispondrá de dos convocatorias por matrícula: una ordinaria y otra extraordinaria.

Convocatoria ordinaria

Evaluación continua:

Se evaluarán los informes de las actividades-problemas planteados en las clases presenciales y entregados a través del Aula virtual, la elaboración y exposición oral de los seminarios y la realización de una prueba escrita que será optativa.

Evaluación no continua:

Consistirá en un examen final. Los estudiantes que opten por este tipo de evaluación deberán comunicarlo por escrito al Decanato de Biología durante la primera quincena del curso. Esta prueba consistirá en un cuestionario con preguntas muy concretas sobre los contenidos teóricos de la asignatura.

Convocatoria extraordinaria

La convocatoria extraordinaria será exclusivamente para aquellos alumnos que no hayan superado la ordinaria y su nota vendrá dada en su totalidad por un examen final.

Criterios de evaluación:

La asignatura se evaluará en base al grado de adquisición de las competencias generales y específicas indicadas en apartados anteriores por los estudiantes. Se tendrá en consideración:

- Capacidad de delimitación de los problemas planteados.

- Elección de la metodología adecuada cada problema planteado.

- Grado de conocimiento y utilización correcta de los programas adecuados para la resolución de cada problema planteado.

- Desarrollo argumental empleado para la obtención de los resultados y conclusiones.

- Calidad en la redacción y estilo en los informes.

1 *Es importante señalar los procedimientos de evaluación: por ejemplo evaluación continua, final, autoevaluación, co-evaluación. Instrumentos y evidencias: trabajos, actividades. Criterios o indicadores que se van a valorar en relación a las competencias: dominio de conocimientos conceptuales, aplicación, transferencia conocimientos. Para el sistema de calificación hay que recordar la Normativa del Consejo de Gobierno del 16 de Julio de 2009: la calificación de la evaluación continua representará, al menos, el 60%. Se puede elevar este % en la guía.*

- Calidad en la expresión y comunicación oral.

- Calidad en los aspectos formales de los informes y las presentaciones.

Criterios de calificación:

La calificación máxima que se podrá obtener en la evaluación continua con la presentación de los informes y del seminario será de 8,0. Para obtener una calificación superior el estudiante deberá realizar una prueba escrita que será otativa.

En la evaluación no continua, un examen final constituirá el 100% de la calificación.

Las calificaciones posibles serán:

0,0 - 4,9 SUSPENSO
Nivel insuficiente, según la aplicación de los criterios de evaluación.

5,0 - 6,9 APROBADO
Nivel suficiente, según la aplicación de los criterios de evaluación.

7,0- 8,9 NOTABLE
Nivel medio, según la aplicación de los criterios de evaluación.

9,0 - 9,9 SOBRESALIENTE
Nivel alto de conocimientos teóricos y prácticos, capacidad de análisis y reflexión, capacidad de integración y aplicación de los conocimientos, capacidad de búsqueda de información alternativa.

10 MATRÍCULA DE HONOR

Nivel de excelencia en todos los aspectos mencionados anteriormente. Limitada, por legislación universitaria, a un máximo del 5% de los estudiantes matriculados.

Un alumno será calificado como NO PRESENTADO si optando por la modalidad de evaluación no continua, NO se presenta al examen final.

6. BIBLIOGRAFÍA

Bibliografía básica

Baxevanis, A.D., Bader, G.D., Wishart, D.S., 2020. Bioinformatics. John Wiley and Sons, Inc., Chichester, West Sussex, UK; Hoboken, NJ, USA.

Choudhuri, S. and Kotewicz, M., 2014. Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools. Elsevier/AP, Amsterdam.

Lemey, P., Salemi, M. and Vandamme, A-M Eds., 2009.The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. 2nd Ed. Cambridge University Press. Cambridge; UK.

Mukhopadhyay, C.S., Choudhary, R.K., Iquebal, M.A., 2017. Basic Applied Bioinformatics. John Wiley and Sons, Inc., Chichester, West Sussex, UK; Hoboken, NJ, USA.

Pevsner, J., 2015. Bioinformatics and Functional Genomics. John Wiley and Sons, Inc., Chichester, West Sussex, UK; Hoboken, NJ, USA.

Shaik, N.A., Hakeem, K.R., Banaganapalli, B., Elango, R., 2019. Essentials of Bioinformatics, Volume I: Understanding Bioinformatics: Genes to Proteins. Springer. Switzerland.

Enlaces de interés

NCBI: https://www.ncbi.nlm.nih.gov/

EMBOSS programs (EMBL-EBI). https://www.ebi.ac.uk

DDBJ: https://www.ddbj.nig.ac.jp/index-e.html

DOE Joint Genome Institute. https://jgi.doe.gov

KEGG: Kyoto Encyclopedia of Genes and Genomes. https://www.genome.jp

The Genome Reference Consortium. https://www.ncbi.nlm.nih.gov/grc

Mouse Genome Informatics. http://www.informatics.jax.org/

Arabidopsis Information Resource (TAIR). https://www.arabidopsis.org

Q-bank Bacteria database. http://www.q-bank.eu/Bacteria/

Ensembl genome browser. https://www.ensembl.org

UCSC Genome Browser. https://genome.ucsc.edu

PIR: Protein Information Resource. https://proteininformationresource.org

UniProt. https://www.uniprot.org

PROSITE - Expasy. https://prosite.expasy.org

BRENDA: Enzyme Database. https://www.brenda-enzymes.org
CAZy: Carbohydrate-Active enZYmes Database. http://www.cazy.org
PyMOL. https://pymol.org
STRING: Functional protein association networks. http://www.string-db.org

La Universidad de Alcalá garantiza a sus estudiantes que, si por exigencias sanitarias las autoridades competentes impidieran la presencialidad total o parcial de la actividad docente, los planes docentes alcanzarían sus objetivos a través de una metodología de enseñanza-aprendizaje y evaluación en formato online, que retornaría a la modalidad presencial en cuanto cesaran dichos impedimentos.