



Programa sinóptico de la unidad curricular: **BIOINFORMÁTICA**

Unidad Curricular: <b>Bioinformática</b>					Unidad Responsable: Dpto. de Biología				
Datos Unidad Curricular		Modalidad			Tipo Dedicación		Dedicación Total Unidad Curricular		
Código	Semestre	T	P	L	HSTP	HSTNP	CA	Total Horas por Semana (HS=CA X 3)	Total Horas por Semestre (HS X 16)
181502	8	2	6	0	2	6	4	12	192

Prelaciones: Haber aprobado el 7mo. Semestre o 117 Unidades Crédito.

HSTP: Horas semanales de trabajo que se realiza en el aula o laboratorio y requiere preparación y trabajo adicional

HSTNP: Horas semanales que se realizan en el aula o laboratorio y no requieren de preparación o trabajo adicional

CA: créditos académicos

### Justificación

La bioinformática es la rama de la Biología que usa herramientas computacionales para la creación, almacenamiento y búsqueda en bases de datos biológicas, y para la visualización y el análisis de las secuencias y de la información estructural y funcional de ácidos nucleicos y proteínas, incluyendo el análisis y anotación de genomas así como inferencias de orden evolutivo basadas en su uso.

Muchos de los algoritmos de la bioinformática, en especial los relacionados con las filogenias, fueron inicialmente desarrollados por zoólogos y botánicos analizando caracteres morfológicos. Debido a que en los organismos los caracteres moleculares y los morfológicos con el resultado de un mismo proceso evolutivo, estos algoritmos resultaron igualmente aplicables a partir de datos moleculares. Ello llevó a los zoólogos y botánicos a interesarse en la biología molecular, y a los biólogos moleculares a interesarse en la evolución. El resultado ha sido que la Biología, anteriormente dividida en un bloque molecular y otro orgánico escasamente conexos, se ha reunificado. La bioinformática, cuyos métodos son de interés tanto para biólogos moleculares como orgánicos, ha sido fundamental en esta reunificación que ha traído como consecuencia un progreso revolucionario en áreas del conocimiento biológico tales como la genómica funcional, la estructura biomolecular, el análisis de proteomas, el metabolismo celular, la evolución, la taxonomía, la sistemática, la cuantificación de la biodiversidad, la biología de la conservación, la ingeniería genética, y el diseño de drogas y vacunas.

La unidad curricular Bioinformática permite a los estudiantes de la Licenciatura en Biología integrar y aplicar a la solución de problemas concretos los conocimientos teóricos-prácticos adquiridos en muchas de las asignaturas que han cursado previamente, tales como Biología de Microorganismos, Botánica,



Zoología, Bioquímica, Genética, Biología Celular y Evolución. Igualmente, la Bioinformática capacita a los estudiantes de carreras relacionadas con las ciencias de la vida para la investigación y para el mercado de trabajo, debido a que las herramientas de esta rama del conocimiento son aplicables a una amplísima variedad de problemas biológicos y procesos industriales y comerciales.

### **Requerimientos**

Es deseable que, además de haber cursado la unidad curricular Evolución, el estudiante muestre manejo instrumental del idioma Inglés y que a su vez sea diestro en el uso de computadoras y navegadores.

### **Objetivo General**

Capacitar al estudiante en el campo de la Bioinformática como herramienta necesaria en el desarrollo de su profesión como Biólogo.

### **Objetivos Específicos**

1. Ofrecer un conocimiento general de los términos y conceptos utilizados en la Bioinformática.
2. Enseñar a conocer y usar las principales bases de datos de dominio público de secuencias de genes, genomas, proteínas y conjuntos de datos poblacionales derivados del uso de marcadores moleculares.
3. Enseñar a usar y a hacer aportes a las bases de datos biológicos.
4. Posibilitar la aplicación de métodos para predecir la estructura y función de proteínas y secuencias estructurales de ácidos nucleicos.
5. Mostrar cómo hacer alineamientos por pares y múltiples de proteínas y ácidos nucleicos con base a su homología y similitud.
5. Enseñarle a construir árboles filogenéticos para examinar relaciones evolutivas.
6. Enseñarle el valor práctico del manejo de esta herramienta como identificación, tipaje, diseño de primers, etc.

### **Contenidos**

#### **Unidad I. Introducción y bases de datos biológicas**

**Tema 1.** Conceptos generales de bioinformática. La bioinformática: qué es, objetivos, aplicaciones, limitaciones. Herramientas y lenguajes computacionales empleados en la generación de bases de datos moleculares de carácter biológico.

**Tema 2.** Bases de datos biológicas. Definición y tipos de bases de datos biológicas: nucleótidos, secuencias codificantes y genómicas, proteínas, clínicas y bibliográficas. Acceso a bases de datos biológicos. Precauciones en el uso de las bases de datos biológicas.

#### **Unidad II. Alineamiento de secuencias**

**Tema 3.** Alineamiento de dos secuencias. Base evolutiva. Homología de secuencias y similitud de secuencias. Métodos de análisis basados en la similitud



de secuencias. Métodos de análisis basados en la identidad de secuencias. Matrices de secuencias. Significancia estadística del alineamiento de secuencias.

**Tema 4.** Búsqueda en bases de datos biológicas. Requerimientos. Métodos BLAST, FASTA y Smith-Waterman. Algoritmos exhaustivos y heurísticos. Manejo de programas: BLAST, BLASTN, BLASTX, TBLASTN, TBLASTX.

**Tema 5.** Alineamiento de secuencias múltiples. Función de puntuación (scoring function). Algoritmos exhaustivos y heurísticos. Aspectos prácticos. Manejo de programas como ClustalW y Multialign.

### **Unidad III. Análisis de secuencias**

**Tema 6.** Análisis de secuencias nucleotídicas. Composición de bases, identificación de secuencias codificantes (ORFs), manejo de programas para traducir secuencias de ácidos nucleídos en secuencias aminoacídicas, mapas de restricción. Diseño de oligos para la reacción en cadena de la polimerasa (PCR), manejo de los programas Oligoexplorer y Oligoanalyser.

**Tema 7.** Análisis de secuencias proteicas. Predicción de parámetros fisicoquímicos: masa molecular, punto isoelectrico (pI). Identificación de segmentos transmembranales y secuencias de importación. Mapas peptídicos. Predicción de estructura terciaria. Bases de datos de estructuras terciarias (PDB). Manejo de las herramientas del sitio Expasy (Experte Protein Analysis System).

**Tema 8.** Motivos (motifs) de proteínas y predicción de dominios. Identificación de motivos y dominios en el alineamiento múltiple de secuencias. Bases de datos de motivos y dominios que usan expresiones regulares. Bases de datos de motivos y dominios que usan modelos estadísticos. Bases de datos de familias de proteínas. Descubrimiento de motivos en secuencias no alineadas. Logos de secuencias.

### **Unidad IV. Filogenias moleculares**

**Tema 9.** Conceptos básicos. Terminología de la evolución molecular y de las filogenias moleculares. Filogenias de genes y filogenias de especies. Combinación de datos moleculares y datos morfológicos. Maneras de representar árboles filogenéticos. La dificultad de encontrar el árbol verdadero. Comparación general de los métodos para construir árboles filogenéticos.

**Tema 10.** Selección de genes. Búsqueda y diseño de *primers*. Selección de genes para filogenias profundas y superficiales (deep and shallow). Ventajas y desventajas de los genes mitocondriales. Ventajas y desventajas de los genes nucleares. Genes comúnmente usados para diferentes grupos de organismos.

**Tema 11.** Árboles filogenéticos basados en distancias. Métodos basados en distancias: Algoritmo UPGMA (Unweighted Pair-Group Method with Arithmetic), unión de vecinos. Bases de datos y programas.

**Tema 12.** Árboles filogenéticos basados en caracteres. Máxima parsimonia. Máxima verosimilitud. Enfoque bayesiano. Evaluación de árboles filogenéticos. Bases de datos y programas como MEGA, MUSCLE y BEAST.

**Tema 13.** Teoría coalescente. Evolución neutral: interacción de la deriva genética y las mutaciones. Modelado de la deriva genética y las mutaciones. La



coalescencia bajo diversos escenarios demográficos. Inferencia estadística acerca de las hipótesis y parámetros demográficos. Métodos Monte Carlo. Calibración de relojes moleculares.

#### **Unidad V. Genómica y proteómica.**

**Tema 14.** Mapeo, ensamblaje y comparación de genomas. Mapas de genomas. Secuenciamiento de genomas. Ensamblaje de secuencias de genomas. Anotación de genomas. Genómica comparativa.

**Tema 15.** Genómica funcional. Enfoques basados en secuencias. Enfoques basados en microarreglos. Comparación de microarreglos SAGE y DNA.

**Tema 16.** Proteómica. Tecnología del análisis de la expresión de proteínas. Modificación postraduccional. Sorting de proteínas. Interacciones proteína-proteína.

#### **Trabajos prácticos**

**Semana 1.** Conocimiento, manejo y uso de bases de datos. Términos de búsqueda y los conectores booleanos. Búsquedas por similitud.

**Semana 2.** Organización de los resultados de búsqueda e interpretación de los parámetros de éxito.

**Semana 3.** Alineamientos sencillos e interpretación de los parámetros computacionales de aceptación y rechazo.

**Semana 4.** Los algoritmos de búsqueda y encuentro de las principales bases de datos.

**Semana 5.** Alineamientos múltiples: generación de archivos de entrada, análisis del resultado, edición, validación y criterios de aceptación o rechazo.

**Semana 6.** Análisis de contenido informacional y estructural de secuencias de ADN y diseño de primers.

**Semana 7.** Análisis de contenido informacional y estructural de secuencias de proteínas y búsqueda y validación de motivos.

**Semana 8.** Bases de datos para el análisis profundo de secuencias de proteínas y su función. Redes estructurales y funcionales de proteínas.

**Semana 9.** Métodos generales de generación de árboles filogenéticos y su representación gráfica.

**Semana 10.** Árboles basados en secuencias de organelos y árboles basados en secuencias nucleares. Comparación y críticas.

**Semana 11.** Árboles filogenéticos basados en distancias.

**Semana 12.** Árboles filogenéticos basados en caracteres y programas empleados para su validación. Árboles de árboles.

**Semana 13.** Calibración de relojes moleculares aplicados a la generación de árboles filogenéticos.

**Semana 14.** Análisis computacional de genomas.

**Semana 15.** Análisis computacional de proteomas.

**Semana 16.** Cómo se preparan los archivos de reporte de datos de secuencia para las bases de datos de dominio público.



### **Estrategias Metodológicas**

El Profesor a cargo del curso presentará las bases teóricas de la Bioinformática, su desarrollo histórico y las tendencias más modernas de su uso y aplicabilidad, las cuales serán reforzadas en dos sesiones prácticas por semana de tres horas cada una. Las sesiones prácticas deberán ser ejecutadas por cada estudiante de manera individual en la sala de computación con conexión a Internet, y con el acceso a todos los programas computacionales de los que se hará uso durante el desarrollo del curso. Además de la instrucción recibida, al estudiante le será asignado un problema biológico al cual dará respuesta mediante la aplicación de los conocimientos y habilidades adquiridas durante el desarrollo del curso.

### **Estrategias de Evaluación**

El curso será evaluado por asistencia, participación, evaluaciones cortas, la presentación de resultados de las tareas asignadas, así como por la presentación de un informe tipo monografía de los resultados derivados de la resolución del problema biológico que le fuera asignado al principio del curso.

### **Bibliografía**

- Avise, J.C. 2000. *Phylogeography: the history and formation of species*. Harvard University Press.
- Avise, J.C. 2004. *Molecular markers, natural history and evolution*. Sinauer Associates.
- Claverie, J-M, y C. 2006. *Bioinformatics for dummies*.
- Cracraft, J., y M.J. Donoghue (Eds). 2004. *Assembling the tree of life*. Oxford University Press.
- Felsenstein, J. 2003. *Inferring phylogenies*. Sinauer Associates.
- Freeman, S., y J.C. Herron. 2002. *Análisis evolutivo*. Prentice Hall.
- Gibson, G., y S.V. Muse. 2009. *A primer of genome science*. Sinauer Associates.
- Hall, B.G. 2004. *Phylogenetic trees made easy: a how-to manual*. Sinauer Associates.
- Lecointre, G, y Guyader, H. 2007. *The tree of life: a phylogenetic classification*. Belknap Press of Harvard University Press.
- Lesk, A. 2008. *Introduction to bioinformatics*. Oxford University Press.
- Lynch, M. 2007. *The origins of genome architecture*. Sinauer Associates.
- Nei, M. y S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press.
- Nielsen, R. 2005. *Statistical methods in molecular evolution*. Springer.
- Pagel, M., y A. Pomiankowski. 2007 *Evolutionary genomics and proteomics* Sinauer Associates.
- Polanski, A., y M. Kimmerl. 2007. *Bioinformatics*. Springer.
- Ridley, M. 2004. *Evolution*. Blackwell Publishing.
- Salemi, M. y A-M, Vandamme. 2003. *The phylogenetic handbook: an practical approach to DNA and protein phylogeny*. Cambridge University Press.
- Wakeley, J. 2008. *Coalescent theory: an introduction*. Roberts & Company Publishers.
- Xiong, J. 2006. *Essential bioinformatics*. Cambridge University Press.