



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
LICENCIATURA EN CIENCIAS GENÓMICAS

PROGRAMA DE ASIGNATURA

| CLAVE     | NOMBRE DE LA ASIGNATURA        |                |                                      | SEMESTRE |          |
|-----------|--------------------------------|----------------|--------------------------------------|----------|----------|
|           | BIOINFORMÁTICA Y ESTADÍSTICA 2 |                |                                      | CUARTO   |          |
| MODALIDAD | CARÁCTER                       | HORAS SEMESTRE | HORA/SEMANA<br>TEÓRICAS<br>PRÁCTICAS |          | CRÉDITOS |
| CURSO     | OBLIGATORIA                    | 80             | 3                                    | 2        | 8        |
| NIVEL     | AVANZADO                       |                |                                      |          |          |
| TIPO      | TEÓRICO-PRÁCTICA               |                |                                      |          |          |

**OBJETIVO:** Que el alumno comprenda a detalle y profundidad los principales algoritmos utilizados en las ciencias genómicas para el análisis de secuencias, análisis de moléculas y análisis de datos masivos..

**METODOLOGÍA:**

La estructura del análisis de cada algoritmo comprende lo siguiente:

1. *Descripción del problema.*

Plantear la pregunta biológica implicada. Ejemplo.

2. *Análisis del Algoritmo*

Describir el algoritmo que soluciona el problema, mediante el uso de diagramas de flujo, pseudo-código, diagrama de clases u otra técnica de análisis. Si aplica, explicar el proceso de asignación de *scores*.

3. *Fundamento estadístico*

Explicar la estadística que es usada en el algoritmo. Por ejemplo: distribuciones de valor extremo, binomial, hipergeométrica, etc.

4. *Programas computacionales*

Explicar el programa computacional que usa el algoritmo descrito. Esto implica: conocer el programa, entender sus parámetros de entrada, entender los resultados, interpretación biológica de los resultados, ventajas y desventajas del método.

Otros algoritmos y programas relacionados.

Generalmente existen programas relacionados o variantes del programa, y sería muy útil describirlos. Se recomienda abordar los siguientes puntos:

- Criterios de selección del mejor algoritmo de acuerdo al problema.
- Que falta por hacer en el área y sería posible plantear alguna mejora al algoritmo(s).

| Número de Horas: | Contenidos temáticos:  |
|------------------|--|
| 20               | <b>1. Análisis de datos con R</b><br>1.1. Análisis estadísticos de datos usando R<br>1.2 Bioconductor  |
| 15               | <b>2. Análisis de secuencias. Búsqueda de motivos</b><br>2.1 Búsqueda de patrones de proteínas y en secuencias de ácidos nucleicos.  |
| 20               | <b>3. Análisis de estructuras de macromoléculas</b><br>3.1 Propiedades moleculares: accesibilidad al solvente, puentes de hidrógeno, estabilidad, desorden y estructura secundaria. Diseño de primers.<br>3.2 Modelado comparativo, fold recognition y diseño ab initio de |

| <b>Número de Horas:</b> | <b>Contenidos temáticos:</b>  |
|-------------------------|---|
| 15                      | proteínas y ácidos nucleicos.   |
| 10                      | <b>4. Clasificación y Clustering</b><br><b>5. Métodos de high throughput : Microarreglos e inferencia de redes</b><br>5.1 Evaluación de microarreglos.<br>5.2 Normalización de microarreglos.<br>5.3 Análisis de microarreglos.<br>5.4 Inferencia de redes de regulación. |
| 80                      | <i>Total de Horas</i>   |

### **BIBLIOGRAFÍA BÁSICA:**

- R web page : <http://www.r-project.org/>

### **BIBLIOGRAFÍA COMPLEMENTARIA:**

- van Helden J. Regulatory sequence analysis tools. *Nucleic Acids Res.* 2003 Jul 1;31(13):3593-6.
- van Helden J, Rios AF, Collado-Vides J. Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Res.* 2000 Apr 15;28(8):1808-18.
- van Helden J, André B, Collado-Vides J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol.* 1998 Sep 4;281(5):827-42.
- Stormo GD. Consensus patterns in DNA. *Methods Enzymol.* 1990;183:211-21.
- Hertz GZ, Hartzell GW 3rd, Stormo GD. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Comput Appl Biosci.* 1990 Apr;6(2):81-92.
- Schug J, Overton GC. Modeling transcription factor binding sites with Gibbs Sampling and Minimum Description Length encoding. *Proc Int Conf Intell Syst Mol Biol.* 1997;5:268-71.
- Zhou Q, Liu JS. Modeling within-motif dependence for transcription factor binding site predictions. *Bioinformatics.* 2004 Apr 12;20(6):909-16. Epub 2004 Jan 29.
- Bailey TL, Baker ME, Elkan CP. An artificial intelligence approach to motif discovery in protein sequences: application to steroid dehydrogenases. *J Steroid Biochem Mol Biol.* 1997 May;62(1):29-44.
- Ortiz AR, Strauss CE, Olmea O. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* 2002 Nov;11(11):2606-21.
- Contreras-Moreira B, Branger PA, Collado-Vides J. TFmodeller: comparative modelling of protein-DNA complexes. *Bioinformatics.* 2007 Jul 1;23(13):1694-6. Epub 2007 Apr 25

### **SUGERENCIAS DIDÁCTICAS:**

Exposición de los temas por parte del profesor, con la participación activa de los estudiantes.  
Ejercicios en clase y de tarea.

### **SUGERENCIAS DE EVALUACIÓN:**

Ejercicios.  
Participación en clase.  
Proyecto final.  
Presentación del proyecto final.

### **PERFIL PROFESIOGRÁFICO:**

Licenciado(a) en Ciencias Genómicas; Maestro(a) o Doctor(a) en Ciencias de la Computación o Bioinformática.