



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

## GUÍA DOCENTE DE *ANÁLISIS DE SECUENCIAS*

La presente guía docente corresponde a la asignatura Análisis de Secuencias (SEQAN), que se oferta como complementos formativos para el Máster en Bioinformática y Biología Computacional y fue aprobada para el curso lectivo 2017-2018 en Junta de Centro y publicada en su versión definitiva en la página web de la Escuela Politécnica Superior. La guía docente de SEQAN aprobada y publicada antes del periodo de matrícula tiene el carácter de contrato con el estudiante.



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

## ASIGNATURA

### ANÁLISIS DE SECUENCIAS (SEQAN)

#### 1.1. Código

33080

#### 1.2. Materia

Análisis de secuencias

#### 1.3. Tipo

Obligatoria

#### 1.4. Nivel

Máster

#### 1.5. Curso

1º

#### 1.6. Semestre

1º

#### 1.7. Número de créditos

3 ECTS

#### 1.8. Requisitos previos

Para adquirir las competencias de este curso los estudiantes deben tener (requerimientos mínimos):

- Bioquímica a nivel de grado (o cursar complementos formativos en “Bioquímica y Biología Molecular”)
- Competencias básicas en programación (o cursar complementos formativos en “Programación, Linux y Bases de Datos”)
- Estadística a nivel de pregrado



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

- Nivel Inglés B2

## 1.9. Requisitos mínimos de asistencia a las sesiones presenciales

Se plantean dos itinerarios, uno con asistencia obligatoria a clase y otro sin ella. Los estudiantes deberán optar por uno u otro desde el principio del curso y cumplir con los distintos requisitos de evaluación que conlleva cada uno de los modelos, publicados en la presente guía docente (ver apartado 4).

### ITINERARIO CON ASISTENCIA OBLIGATORIA A CLASE

La asistencia es obligatoria al menos en un 80% de las sesiones y realizar/entregar en plazo al menos el 80% de las actividades de las actividades del curso.

Si se dejan de cumplir alguna de estas dos condiciones, el estudiante pasará a ser considerado como “itinerario sin asistencia obligatoria” a efectos de evaluación.

### ITINERARIO SIN ASISTENCIA OBLIGATORIA A CLASE

La asistencia es muy recomendable aunque no obligatoria.

## 1.10. Datos del equipo docente

### Dr. Luis del Peso (Coordinador)

Departamento de Bioquímica

Facultad de Medicina

Laboratorio: 252 (Instituto de Investigaciones Biomédicas)

Teléfono: +34 91 585 44 40

Correo electrónico: [luis.peso@uam.es](mailto:luis.peso@uam.es)

Página web: <https://www.bq.uam.es/miembros/lpeso/index.php>

Horario de tutorías: Petición de cita previa en clase o por correo electrónico.

## 1.11. Objetivos del curso

En este curso se tratan los modelos matemáticos, los algoritmos y los programas informáticos utilizadas para analizar secuencias biológicas. Además de entender las bases teóricas, se estudiará cómo aplicar estos métodos en la práctica. También revisaremos los principales repositorios de secuencias biológicas y como extraer información de ellos.

Las **competencias específicas** que se adquieren en esta asignatura son:

CE1 - Capacidad de aplicar los conocimientos de biología, matemáticas, física y estadística a la bioinformática.

CE2 - Capacidad de aplicar métodos computacionales a la resolución de problemas en el ámbito de la investigación biomédica básica y traslacional.



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

CE3 - Capacidad de analizar, explotar e investigar secuencias biológicas comparativa, evolutiva y funcionalmente.

CE4 - Capacidad de predecir estructuras, funciones, y dinámica de biomoléculas.

CE9 - Capacidad de explotación de la información de las bases de datos biológicas sobre secuencias, estructuras, transcriptomas, genomas, proteomas, etc.

Al final del semestre (objetivos generales), y de cada unidad (objetivos por tema) el estudiante deberá ser capaz de:

<b>OBJETIVOS GENERALES</b>	
G1	Identificar las principales bases de datos de secuencias y estructuras biológicas, así como acceder y extraer información de ellas.
G2	Comprender los principales modelos estadísticos de secuencias biológicas y aplicarlos para obtener hacer predicciones.
G3	Comprender las bases teóricas, tanto conceptuales como algorítmicas, del análisis de secuencias.
G4	Aplicar herramientas bioinformáticas de análisis de secuencias e interpretar los resultados de estos programas.
G5	Interpretar y construir los modelos de secuencias.
G6	Integrar los resultados de análisis de secuencias para resolver problemas biológicos.

<b>OBJETIVOS ESPECIFICOS POR TEMA</b>	
<b>TEMA 1.- Principales base de datos de secuencias</b>	
1.1.	Identificar las principales bases de datos de secuencias y estructuras biológicas
1.2.	Conocer los formatos principales de representación de secuencias
1.3.	Ser capaz de acceder y extraer información de bases de datos de secuencias
<b>TEMA 2.- Modelos estadísticos de secuencias</b>	
2.1.	Comprender los principales modelos estadísticos de secuencias biológicas
2.2.	Aplicar modelos estadísticos para obtener hacer predicciones básicas motivos en secuencias biológicas
<b>TEMA 3.- Alineamiento de pares de secuencias</b>	
3.1.	Comprender las bases teóricas, tanto conceptuales como algorítmicas, del alineamiento de pares secuencias
3.2	Ser capaz de aplicar las herramientas apropiadas para producir alineamientos de pares de secuencias
3.3	Interpretar y evaluar alineamientos de pares de secuencias
<b>TEMA 4.- Alineamiento múltiple y su representación</b>	
4.1.	Comprender las bases teóricas, tanto conceptuales como algorítmicas, del alineamiento múltiple
4.2.	Ser capaz de aplicar las herramientas apropiadas para producir alineamientos múltiples



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

4.3.	Interpretar y construir los principales formatos de representación de secuencias
4.4.	Integrar los resultados de análisis de secuencias para resolver problemas biológicos.
<b>TEMA 5.- Identificación de motivos de secuencia</b>	
5.1.	Comprender las bases teóricas, tanto conceptuales como algorítmicas, de la identificación de motivos de secuencia
5.2.	Ser capaz de aplicar las herramientas apropiadas para identificar motivos de secuencia
5.3.	Integrar los resultados de análisis de motivos para resolver problemas biológicos

## 1.12. Contenidos del programa

### Programa Sintético

- TEMA 1. Principales base de datos de secuencias
- TEMA 2. Modelos estadísticos de secuencias
- TEMA 3. Alineamiento de pares de secuencias
- TEMA 4. Alineamiento múltiple y su representación
- TEMA 5. Identificación de motivos de secuencia

### Programa Detallado

1. Principales base de datos de secuencias
  - 1.1. Crecimiento de las bases de datos
  - 1.2. Tipos generales de bases de datos
  - 1.3. Bases de datos de nucleótidos
  - 1.4. Principales formatos secuencia y características de secuencia (FASTA, GFF)
  - 1.5. Bases de datos de proteínas
  - 1.6. Otras bases de datos
2. Modelos estadísticos de secuencias
  - 2.1. Modelos Multinomiales
  - 2.2. Cadenas de Markov
3. Alineamiento de pares de secuencias
  - 3.1. Introduction
  - 3.2. Matrices de sustitución
  - 3.3. Alineamiento de pares de secuencias
  - 3.4. Estrategias de alineamiento:
    - 3.4.1. Dot-matrix
    - 3.4.2. programación dinámica
    - 3.4.3. BLAST
  - 3.5. Análisis estadístico de las puntuaciones de los alineamientos
  - 3.6. Relevancia biológica



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

#### 4. Alineamiento múltiple y su representación

- 4.1. Estrategias de MSA
- 4.2. Representación de MSA
- 4.3. Contenido de información y Logos
- 4.4. Bases de datos de MSA

#### 5. Identificación de motivos de secuencia

- 5.1. Búsqueda de motivos
- 5.2. Descubrimiento de motivos

## 1.13. Referencias de consulta

### Bibliografía:

#### Libros:

- Introduction to Computational Genomics: A Case Studies Approach. Nello Cristianini and Matthew W. Hahn. Cambridge University Press (2006). (temas/[units](#) 1-4)
- Essential Bioinformatics. Jin Xiong. Cambridge University Press (2006).
- Bioinformatics and functional genomics. Jonathan Pevsner. Wiley-Blackwell. Second edition (2009).
- Bioinformatics --From Genomes to Therapies. Edited by Thomas Lengauer. Wiley-VCH (2007).
- Bioinformatics (Sequence and Genome analysis). David W. Mount. Cold Spring Harbor Laboratory Press (2001)
- Biological sequence analysis. Richard Durbin, Sean Eddy, Anders Krogh and Graeme Mitchison. Cambridge University Press (2007, 10th reprint).

**Material electrónico de trabajo:** los documentos electrónicos de trabajo (material adicional, ejercicios del curso y ejemplos de exámenes, etc.) se publican en la sección de BBM en plataforma Moodle (<http://moodle.uam.es/>)

## 2. Métodos docentes

La metodología utilizada en el desarrollo de la actividad docente incluye los siguientes tipos de actividades:

### \*Clases de teoría:

#### Actividad del profesor

Clases expositivas en las que se realizarán programas y ejercicios ilustrativos. Se utilizará la pizarra, combinada con la explicación de programas en formato electrónico cuya ejecución será proyectada en clase.

#### Actividad del estudiante:



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

*Actividad presencial:* Toma de apuntes, participación en clase respondiendo a las cuestiones planteadas. Resolución de los ejercicios propuestos.

*Actividad no presencial:* lectura del material bibliográfico y de apoyo, estudio de la materia y realización de ejercicios propuestos y de los cuestionarios planteados en la plataforma Moodle.

#### \*Tutorías en aula:

#### Actividad del profesor:

Tutorización a toda la clase o en grupo reducidos con el objetivo de resolver dudas comunes plantadas por los alumnos a nivel individual o en grupo, surgidas a partir de cuestiones/ejercicios señalados en clase para tal fin y orientarlos en la realización de los mismos.

#### Actividad del estudiante:

*Actividad presencial:* Planteamiento de dudas individuales o en grupo y enfoque de posibles soluciones a las tareas planteadas.

*Actividad no presencial:* Estudio de las tareas marcadas y debate de las soluciones planteadas en el seno del grupo.

#### \*Actividades y entregas:

Serán de actividades de diferente naturaleza incluyendo, ejercicios, problemas, discusiones y prácticas con ordenador que se realizarán de forma individual o en grupo.

#### Actividad del profesor:

Explicar y proporcionar instrucciones detalladas de la actividad. Supervisar el trabajo de los estudiantes y proporcionar retroalimentación.

#### Actividad del estudiante:

Las actividades podrán ser planteadas para resolver en el aula o en horas no presenciales. Al finalizar la actividad se entregará el resultado/producto de la misma.

### 3. Tiempo de trabajo del estudiante

		Nº de horas	Porcentaje
Presencial	Exposiciones por el profesor	10 h (45.3%)	22 h (29.3%)
	Resolución de actividades	9 h (40.1%)	
	Realización de pruebas escritas parciales y final	3 h (13.6%)	
No presencial	Preparación sesiones y estudio semanal	14 h (26.4%)	53 h (70.7%)
	Realización de actividades	12 h (22.6%)	
	Preparación del examen (convocatoria ordinaria)	8 h (15.2%)	
	Preparación del examen (convocatoria extraordinaria)	19 h (35.8%)	
<b>Carga total de horas de trabajo: 25 horas x 3 ECTS</b>		<b>75 h</b>	



Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

## 4. Métodos de evaluación y porcentaje en la calificación final

Las pruebas escritas podrán incluir tanto cuestiones teóricas como ejercicios.

### ITINERARIO CON ASISTENCIA OBLIGATORIA A CLASE

La calificación final de la asignatura en la convocatoria ordinaria se obtiene a partir de las notas de las actividades de evaluación continua (Continua) y una prueba final (Final\_ordin) mediante la ecuación:

$$\text{Calificación: } 0.7 * \text{Continua} + 0.3 * \text{Final\_ordin}$$

La evaluación continua implicará al menos tres entregas (que podrán incluir entre otros problemas, ejercicios y test) que se realizarán en el aula o en horas no presenciales.

Para aprobar la asignatura es obligatorio obtener una nota mayor o igual a 5 puntos tanto en la parte de teoría como en la de evaluación continua. En caso contrario, la nota final en actas será

$$\text{Calificación: } \text{Mín}(4.9, 0.7 * \text{Continua} + 0.3 * \text{Final\_ordin})$$

La calificación final de la asignatura en la convocatoria extraordinaria se obtiene a partir de las notas de las actividades de evaluación continua (Continua) y una prueba final (Final\_extra) mediante la ecuación:

$$\text{Calificación: } \text{Max}(0.7 * \text{Continua} + 0.3 * \text{Final\_extra}, \text{Final\_extra})$$

Para aprobar la asignatura en convocatoria extraordinaria es obligatorio obtener una nota mayor o igual a 5 puntos tanto en el examen final (Final\_extra).

### ITINERARIO SIN ASISTENCIA OBLIGATORIA A CLASE

La calificación final de la asignatura en la convocatoria ordinaria y extraordinaria para alumnos que hayan optado por el itinerario sin asistencia obligatoria corresponde únicamente a la prueba final. Esta prueba será más exhaustiva y tendrá más duración que la correspondiente a la trayectoria de evaluación continua, ya que incluirá ejercicios adicionales, relacionados con las actividades intermedias evaluadas en clase.

Para aprobar la asignatura es obligatorio obtener una nota mayor o igual a 5 puntos en el examen final.

**ATENCIÓN:** Cualquier copia descubierta que se haya realizado a lo largo del curso, tanto en cualquiera de las actividades de teoría desarrolladas, ejercicios, y debate como en cualquiera de los apartados de las prácticas, serán penalizadas con rigurosidad. La penalización por copia implica la aplicación de la normativa





Asignatura: Análisis de secuencias  
Código: 33080  
Centro: Escuela Politécnica Superior  
Titulación: Máster en Bioinformática y Biología Computacional  
Nivel: Máster  
Tipo: Obligatoria  
Nº de créditos: 3

interna de la EPS, que supone suspender la convocatoria actual y no poder presentarse a la siguiente.

## 5. Cronograma

Semana	Contenido	Horas presenciales	Horas no presenciales
1	TEMA 1. Principales base de datos de secuencias TEMA 2. Modelos estadísticos de secuencias TEMA 3. Alineamiento de pares de secuencias TEMA 4. Alineamiento múltiple y su representación TEMA 5. Identificación de motivos de secuencia	8	11
2	TEMA 4. Alineamiento múltiple y su representación TEMA 5. Identificación de motivos de secuencia	8	11
3	TEMA 5. Identificación de motivos de secuencia	3	4
	Examen Final convocatoria ordinaria	3	8
	Examen Final convocatoria extraordinaria	3	19