



# LA INFORMÁTICA EN EL MÁSTER EN BIOINFORMÁTICA

**Vicente Arnau, Wladimiro Díaz, Ricardo Ferris,  
Miguel Lozano, Sergio Casas**  
Dpto. de Informática. ETSE.

## Primer Cuatrimestre (30 créditos)

Asignaturas	Créditos	Informáticos	Biólogos	Médicos
Bioquímica y biología molecular	9			
Evolución	6			
Genética	9			
Medicina y experimentación clínica	6			
<b>Computadores y sistemas operativos</b>	<b>6</b>			
<b>Programación</b>	<b>9</b>			
<b>Estructuras de datos y bases de datos</b>	<b>9</b>			

## Segundo Cuatrimestre (30 créditos)

Asignaturas	Créditos
<b>Algoritmos en bioinformática</b>	<b>3</b>
Bioinformática estadística	6
Bioinformática estructural	3
Bioinformática evolutiva	6
Nociones básicas de bioinformática y genómica	6
Técnicas ómicas para la obtención masiva de datos	6

## Tercer Cuatrimestre (30 créditos)

Asignaturas	Créditos
Aplicaciones y tendencias en bioinformática	3
Biología de sistemas computacional	6
Estudios in silico en biomedicina	6
<b>Programación y técnicas computacionales avanzadas en bioinformática</b>	<b>3</b>
<b>Trabajo fin de máster (TFM)</b>	<b>12</b>

## Asignaturas orientadas a proyectos

### **Metodología docente:**

Los estudiantes de biología saben poco de informática. Deben adquirir unos conocimientos en programación, bases de datos y sistemas operativos equivalentes a un segundo grado de Ingeniería Informática.

Los datos a utilizar son datos biológicos.

### **Evaluación continua:**

Realización de trabajos sobre datos bioinformáticos.

Examen final (30%-50%)

### **Coordinación:**

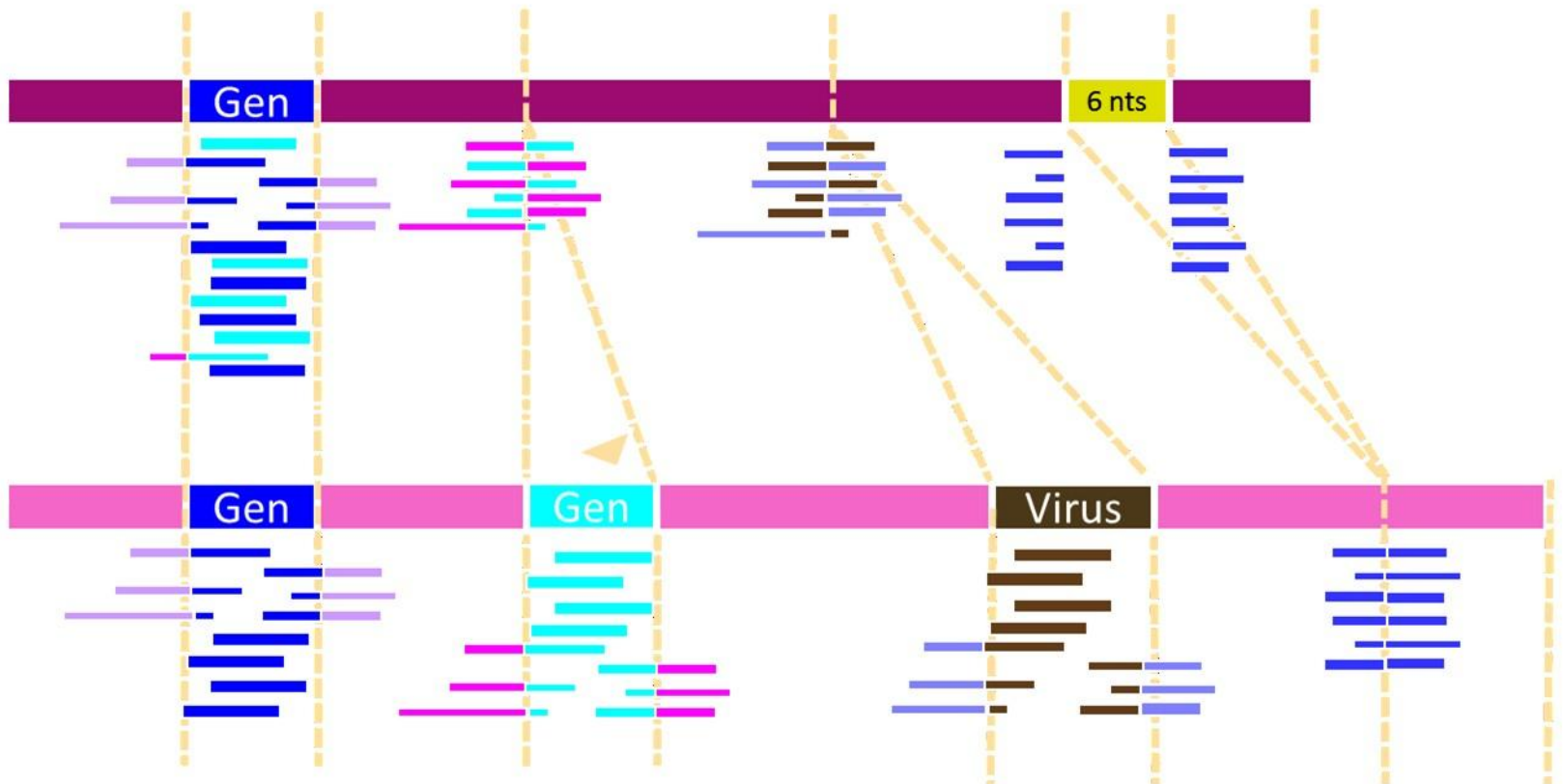
Completa comunicación entre los profesores.



## Computadores y Sistemas Operativos.

Ejemplo de trabajo a desarrollar en el laboratorio.

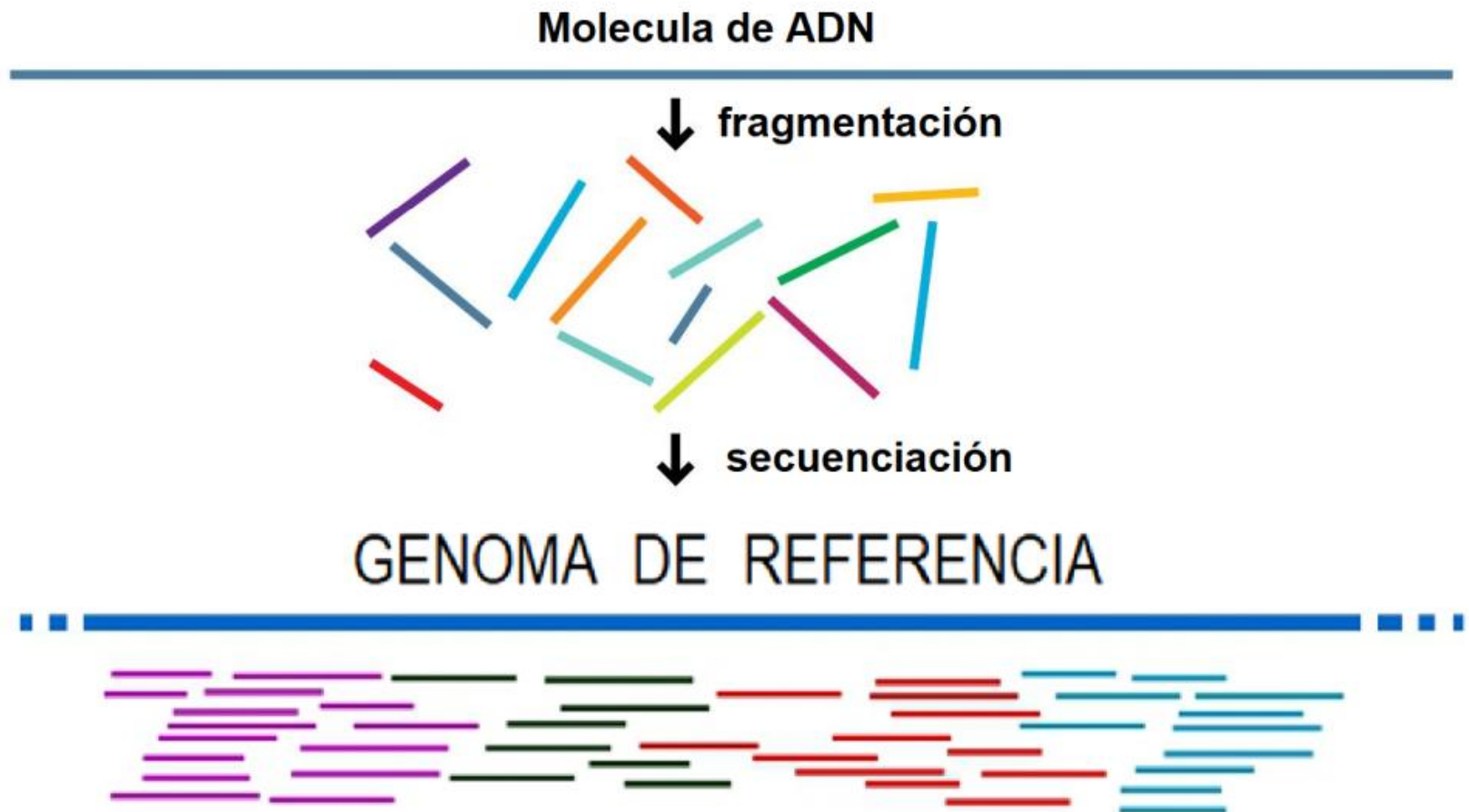
- una duplicación de un gen
- una inserción de virus
- borrado de 6 nt.



## Secuenciación del ADN:

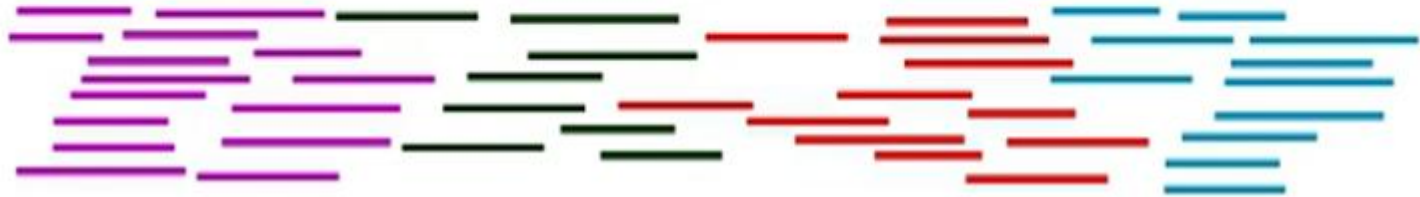
Es el proceso mediante el cual se obtiene la secuencia de adn de una especie.

Se instala el programa **wgsim** y se simula la secuenciación de un genoma. Se ensambla con **BWA**.





# GENOMA DE REFERENCIA



## Ejemplo de fichero SAM/BAM:

```

ref|NC_012920.1|_17073_17268_5:0:0_3:0:0_4 0          ref|NC_012920.1| 14833
    60          100M          *          0          0
ATGATGAAACTTGGGCTCACTCCTTGGCGCCTGCCTGATCCTCAAACACCACAGGCCTATTTCCTAGCCA
EAEEEEEEEEEEEEEEEEEEEE/EAEEEE<<AAAA6EEEEEEEE/EEAAAAEEEEE/AEEEEEEEEEEAE
NM:i:5  MD:Z:12C34T9A23A16C1      AS:i:78  XS:i:0
ref|NC_012920.1|_16181_16280_4:0:0_1:0:0_7 0          ref|NC_012920.1| 4421
    60          60M40S          *          0          0
GCTATCGGGCCATAGCCCGAAAATGTTGGTTATACCGTTCCTACTAATTAATCCCTAACCTGACTAGA
AAAA6EEEEEEEEEEEE6EEEE6EEEE//EE66///6/E//EEE/66666//////////A///A//<<
NM:i:2  MD:Z:15C21C22      AS:i:50  XS:i:0
SA:Z:ref|NC_012920.1|,13998,+,57S43M,60,2;

```



Q1) ¿Que gen se ha duplicado? ¿Donde?

Q2) ¿Donde se ha insertado el genoma de un virus? ¿Qué virus es?

Q3) Obtener la secuencia del virus.

Q4) Obtén la cobertura del genoma de referencia.

Q5) Donde hay mutaciones.

Q6) . . . . .



Comandos básicos para manipulación de ficheros y directorios: **pwd, ls, mkdir, cd, ls, rm, echo, cat, more, mv, touch, alias.**

- Permisos en ficheros. Comandos **chmody chown.**
- Instalación de paquetes: **apt-getinstall.**
- Editores de ficheros: **vi, vim.**
  
- Manejo básico de la shell. Archivo **.bashrcy fstab.**
- Creación de usuarios. Trabajando como **root.**
- Órdenes enlazadas por tuberías.
- Comandos básicos: **head, tail, cut.**
- Lista de órdenes. Uso de variables.
- Estructura condicional “**if-then-else**”.
- Estructura repetitiva **while.**
- La orden **read, comando sort.**
  
- **Grep, wc, sed, awk, uniq, paste, join.**
- Comando **awk: definición, variables, operadores, plantillas, instrucciones de control y funciones. Ejemplos.**
- Comandos para compresión de datos: **tar, gzip, gunzip, bzip2.**



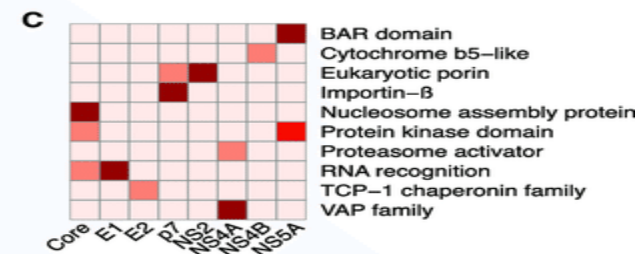
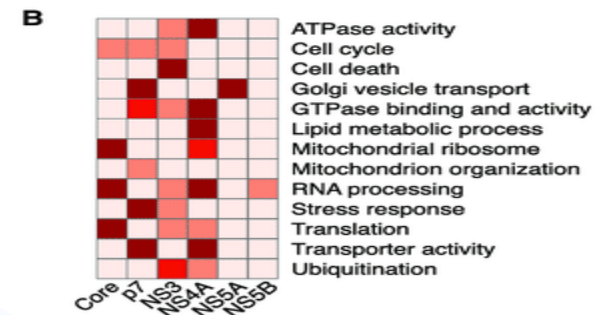
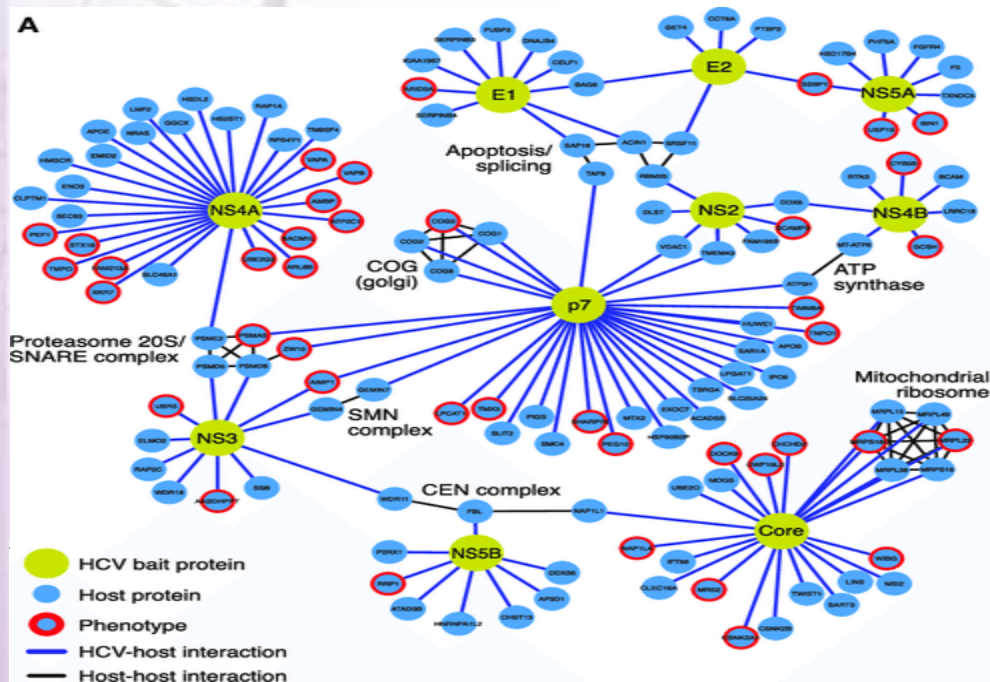
# Programación y estructura de datos:

Manejo de ficheros → sobre ficheros con formato FASTA, FASTQ, SAM, ...

Uso de matrices → matrices de expresión de microarrays

Grafos → grafos de interacciones de proteínas.

Lista ligadas → presencia de k-mers de secuencias de nucleótidos.

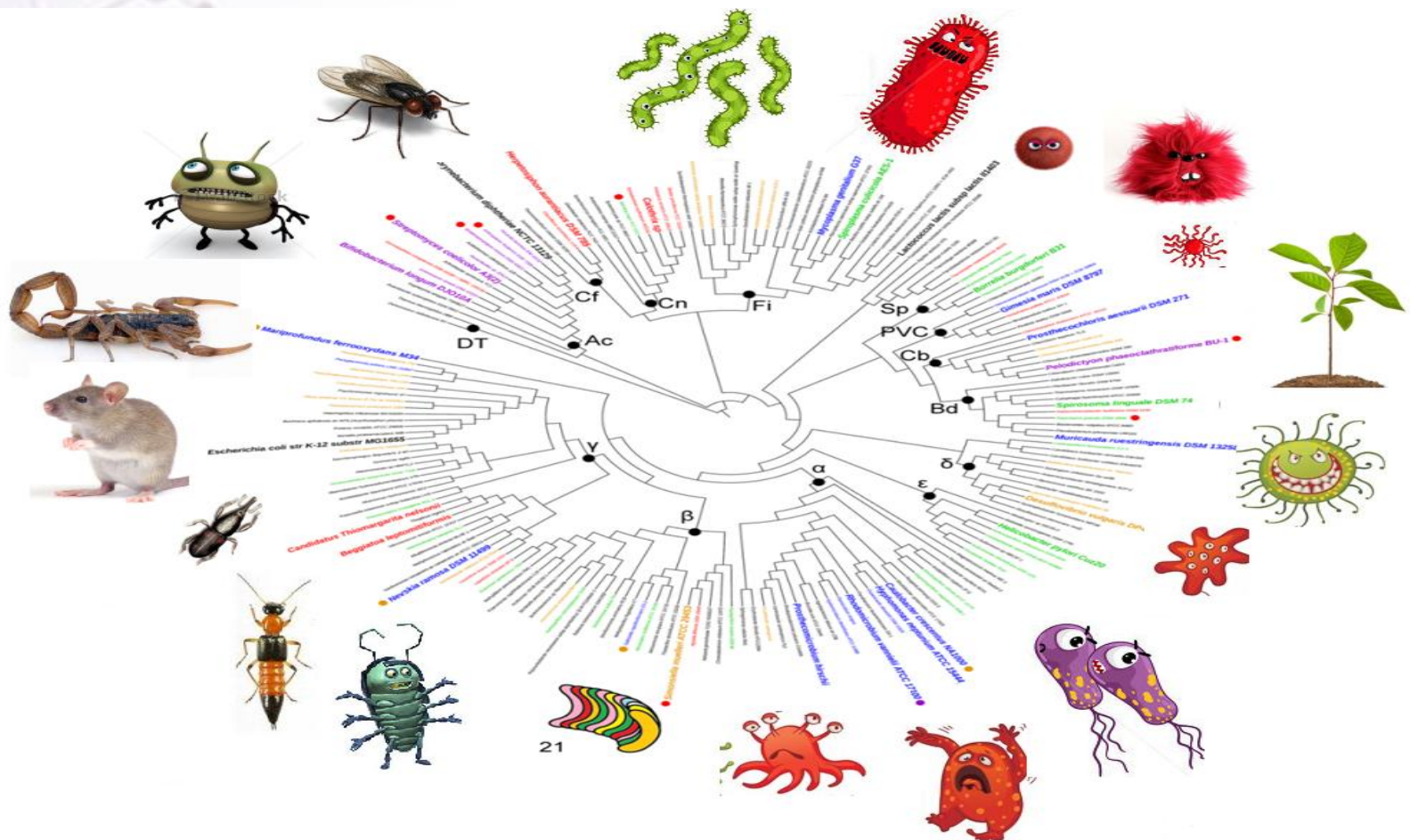


Enrichment (q-value)



## Bases de datos:

La base de datos *Taxonomy* del NCBI es una taxonomía filogenética que incluye a todos los organismos representados en GenBank.



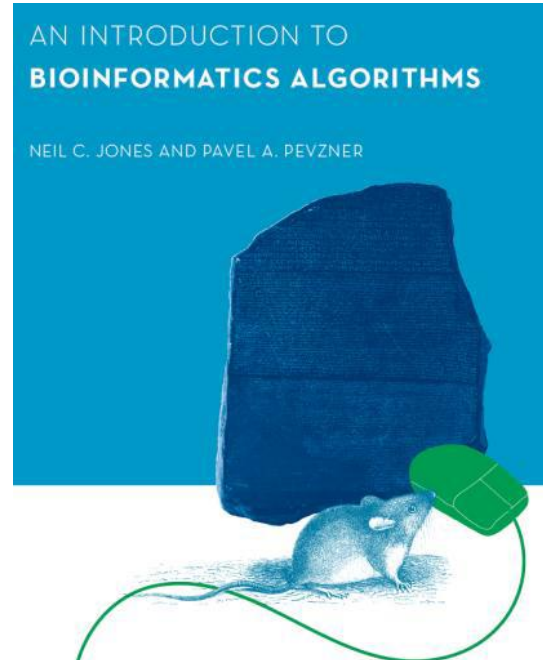


# Algoritmos en Bioinformática:

## Contents in Brief

Subject	4	5	6	7	8	9	10	11	12
Mapping DNA	○								
Sequencing DNA					○				
Comparing Sequences			○	○		○			
Predicting Genes			○						
Finding Signals	○	○						○	○
Identifying Proteins					○				
Repeat Analysis						○			
DNA Arrays					○				
Genome Rearrangements		○							
Molecular Evolution							○		

- Exhaustive Search
- Greedy Algorithms
- Dynamic Programming
- Divide-and-Conquer Algorithms
- Graph Algorithms
- Combinatorial Algorithms
- Clustering and Trees
- Hidden Markov Models
- Randomized Algorithms





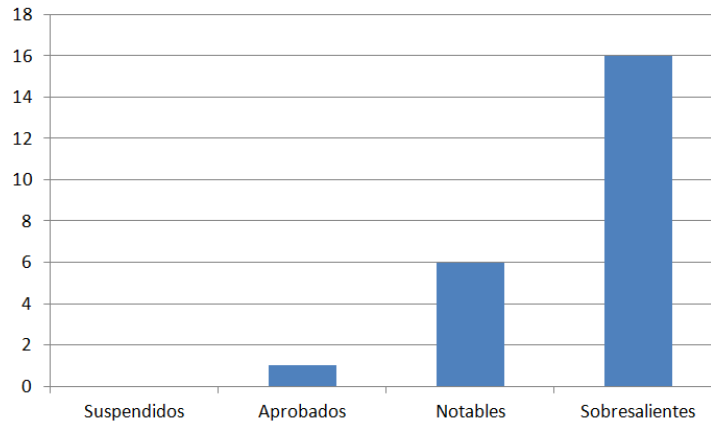
**ubuntu**  
L i n u x

[masterbio2.uv.es](http://masterbio2.uv.es)

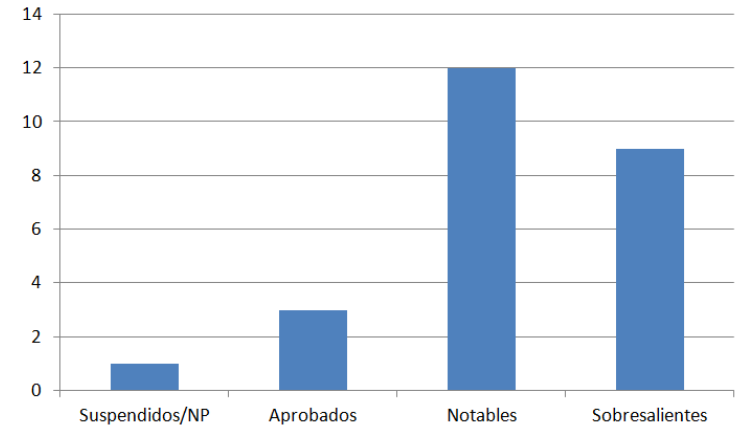


# Ejemplos de resultados académicos del curso 2018-19:

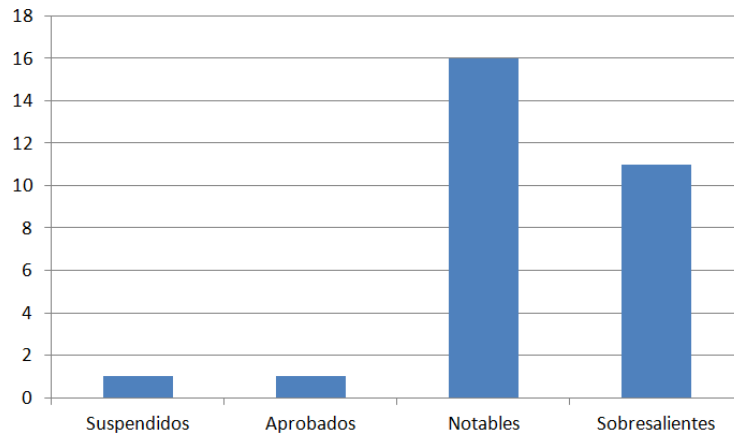
Notas en **Computadores y sistemas Operativos**



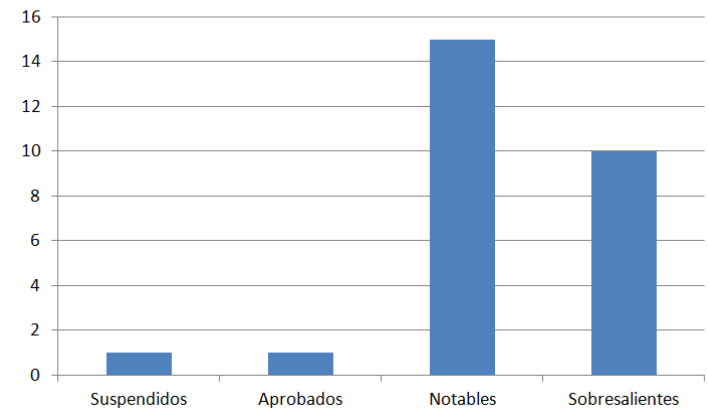
Notas en **Programación**



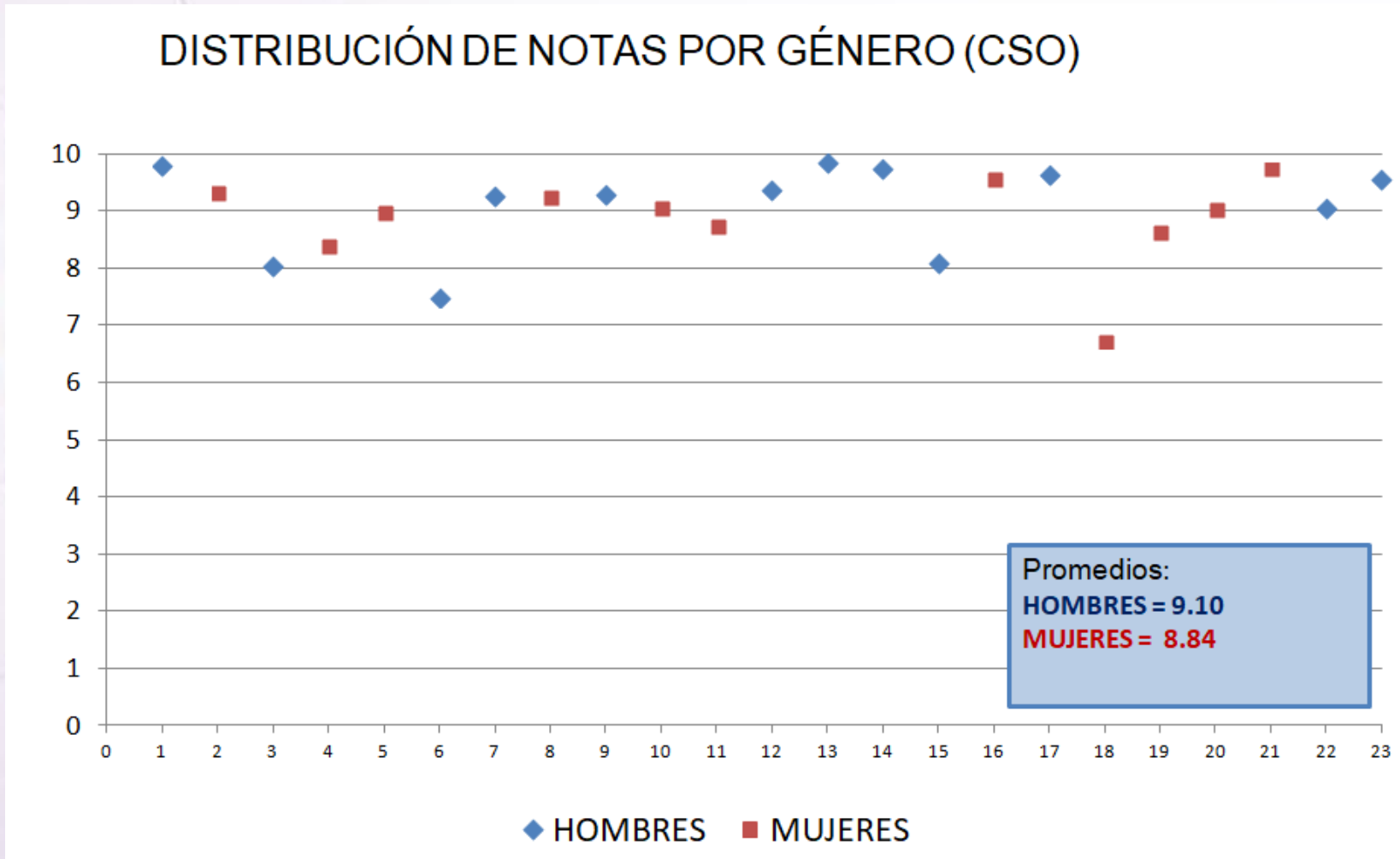
Notas en **Algoritmos en Bioinformática**



Notas en **Bioinformática Estructural**

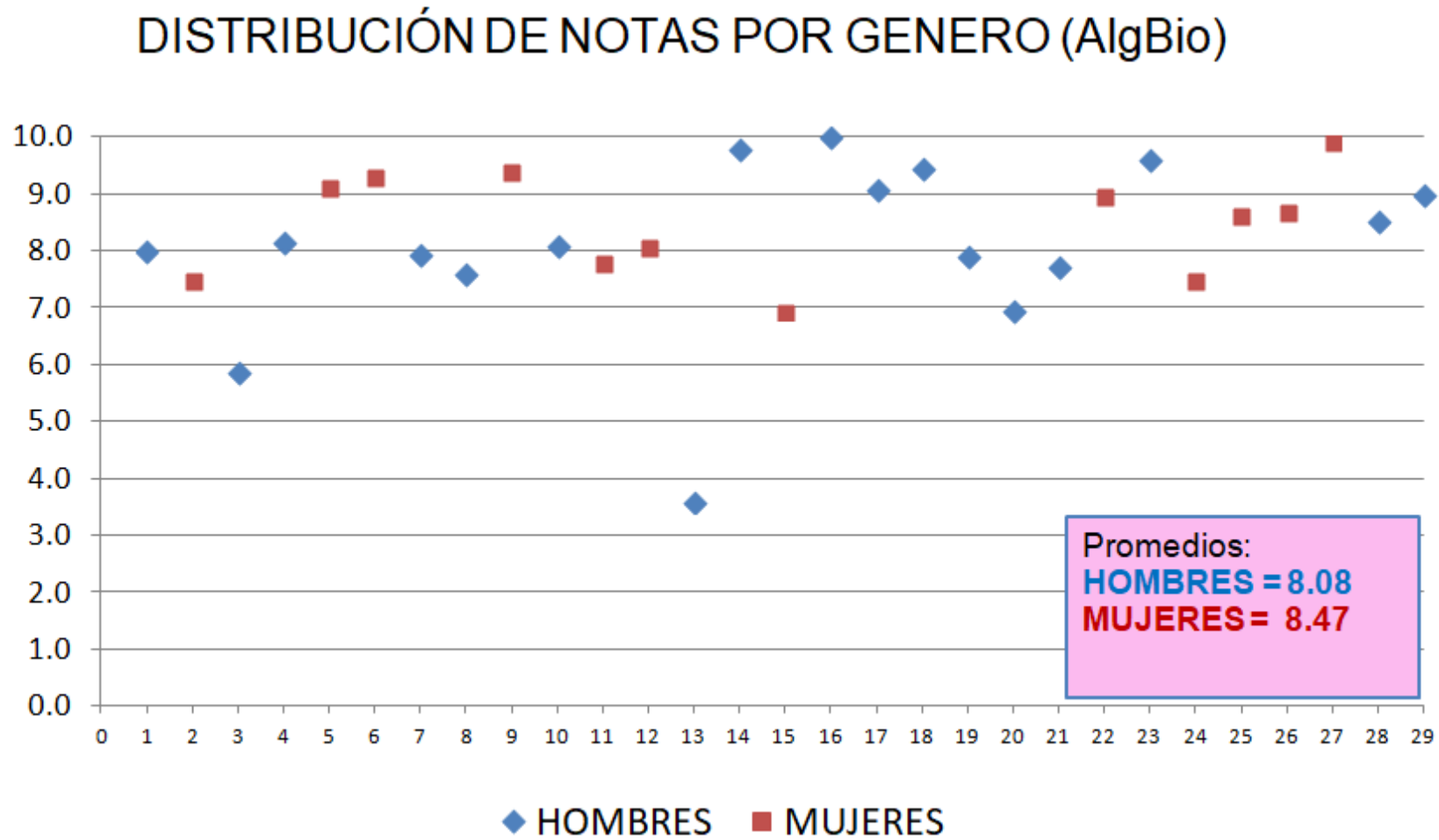


## Resultados académicos por género:





## Resultados académicos por género:





# Máster Oficial en Bioinformática



PRINCIPE FELIPE  
CENTRO DE INVESTIGACION



VNIVERSITAT  
DE VALÈNCIA



Fundación Investigación Clínico de Valencia  
Instituto de Investigación Sanitaria – INCLIVA





# LA INFORMÁTICA EN EL MÁSTER EN BIOINFORMÁTICA

**Vicente Arnau, Wladimiro Díaz, Ricardo Ferris,  
Miguel Lozano, Sergio Casas**  
Dpto. de Informática. ETSE.