

# CC471 – Biología Computacional

Organizador del Curso:

Prof. Ciro Nuñez Iturri

[ciro.nunez.i@uni.edu.pe](mailto:ciro.nunez.i@uni.edu.pe)



# CC471 – Biología Computacional

- 1) Introducción al curso y sus objetivos
- 2) Organización del curso y contenido
- 3) Conceptos Básicos de Biología Molecular
- 4) Problemas Clásicos de Bioinformática





# 1. Introduccion y objetivos

Pre Requisitos:

- 1)CH061 – Biologia (Conocimientos Básicos de Biología Molecular)
- 2)CC302 – Lenguaje Paralelo Distribuido, Orientado a Objetos  
(Programación, desarrollo y análisis de algoritmos)



# 1. Introduccion y objetivos

Objetivos para el semestre:

Introduccion a la Biologia Computacional:

- 1)Entender y resolver Problemas Fundamentales de Bilología Computacional
- 2)Conocer y aplicar Técnicas y algoritmos para el análisis de datos que se aplican el campo de la Biologia.
- 3)Entender como funcionan los métodos aplicados

Desarrollar la Habilidad para iniciar un trabajo de investigación:

- 1)Practicas con datos reales.





# 1. Introduccion y objetivos - Información del curso

Clases de Teoría : Los Lunes de 17. A 20 hrs. Aula R1/411

Clases de Laboratorio: Los Viernes de 08 a 10 hrs. Sala 1.



# 1. Introduccion y objetivos - Información del curso

1. Examen parcial
2. Examen Final
3. Proyecto de curso
4. 4 Practicas Calificadas
5. 4 Practicas de Laboratorio calificadas.
6. Examen sustitutorio.



## 2. Organización del curso

PROG. ANALITICO	UNIDAD	CONTENIDO	CALENDARIO DE EVALUACIONES
1 (20.08.18 al 24.08.18)	Conceptos introductorios	1. Introducción; 2. Conceptos básicos de Biología Molecular; 3. Problemas clásicos de Bioinformática; 4. Herramientas de recolección y almacenamiento de secuencias de laboratorio;	Laboratorio Dirigido 1
2 (27.08.18 al 01.09.18)	Conceptos introductorios	5. Recursos de Software, introducción a BLAST, CLUSTAL; 6. Cadenas, Grafos y algoritmos	Laboratorio Dirigido 2
3 (03.09.18 al 07.09.18)	Alineamiento de secuencias	1. Introducción al alineamiento de secuencias; 2. comparación de pares de secuencias; 3. Alineamiento de secuencias Global; 4. Alineamiento de secuencias múltiples.	Laboratorio Dirigido 3
5 (10.09.18 al 14.09.18)	Alineamiento de secuencias	5. Cadenas Ocultas de Markov; 6. Metodos exactos, aproximados y heurísticos del alineamiento de secuencias, 7. Problemas derivados del alineamiento de secuencias.	Practica calificada 1; Laboratorio Calificado 1
6 (17.09.18 al 21.09.18)	Clustering	1. El Problema del Clustering; 2. Clustering Jerárquico. 3. Algoritmo Neighbour Joining; y. Algoritmo Average Linkage. 4. Clustering no jerárquico o K-means. 5. EST clustering	Laboratorio Dirigido 4
7 (24.09.18 al 28.09.18)	Arboles Filogenéticos	1. Introducción a la Filogenia.; 2. Algoritmos Comunes; 3. Aplicaciones Biológicas.	Laboratorio Dirigido 5
8 (01.10.18 al 05.10.18)	Arboles Filogenéticos	4. Algoritmos Exactos; 5 Algoritmos Probabilísticos	Practica calificada 2; Laboratorio Calificado 2
9 (08.10.18 al 12.10.18)	Examen Parcial	Examen Parcial	Examen parcial



## 2. Organización del curso

9 (08.10.18 al 12.10.18)	Examen Parcial	Examen Parcial	Examen parcial
10 (15.10.18 al 19.10.18)	Mapeo de Secuencias	1.El Problema de Double Digest y Partial Digest.	Laboratorio Dirigido 6
11 (22.10.18 al 26.10.18)		EXPOCIENCIA	
12 (29.10.18 al 02.11.18)	Mapeo de Secuencias	2; Tecnicas utilizadas en el mapeo de secuencias. 3; Mapeo con non – Unique Probes ; 4. Mapeo con Unique Probes;	Practica calificada 3; Laboratorio Dirigido 7
13 (05.11.18 al 09.11.18)	Mapeo de Secuencias	4; Mapeo con Unique Probes 5. Grafos de Intervalos;	Laboratorio Dirigido 8
14 (12.11.18 al 16.11.18)	Mapeo de Secuencias	6.Mapeo con señales de frecuencias de restricción.	Laboratorio Calificado 3
15 (19.11.18 al 23.11.18)	Introduccion a la estructura de proteinas	1. Fundamentos Biológicos de las Proteinas; 2. Motivacion para la predicción de estructuras de las proteinas	Laboratorio Dirigido 9
16 (26.11.18 al 30.11.18)	Introduccion a la estructura de proteinas	3; Alieneamiento rígido de Proteinas. 4; Técnica de alineamiento por Hashing geométrico	Practica calificada 4
17 (03.12.18 al 07.12.18)	Introduccion a la estructura de proteinas	4. Prediccion de las estructuras de las proteinas	Laboratorio Calificado 4
18 (10.12.18 al 14.12.18)	Examen Final	Examen Final	Examen Final
19 (17.12.18 al 21.12.18)	Examen sustitutorio	Examen sustitutorio	Examen sustitutorio

### BIBLIOGRAFIA

1. Clote, P. and Backofen, R. Computational Molecular Biology, An Introduction. Wiley, 2000.
2. Jones, N. A. and Pevzner, P. A. An Introduction to Bioinformatics Algorithms. The MIT Press. 2004.
3. Mount, D. W. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, 2001





## 2. Organización del curso

### Bibliografía de Consulta

- 1) Clote, P. and Backofen, R. Computational Molecular Biology, An Introduction. Wiley, 2000.
- 2) Jones, N. A. and Pevzner, P. A. An Introduction to Bioinformatics Algorithms. The MIT Press. 2004.
- 3) Mount, D. W. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, 2001
- 4) Materiales del Curso en <http://www.mega.nz>





## PRUEBA DE ENTRADA



# 3. Conceptos Básicos de Biología Molecular

1) Que es la Bio Informática? - El uso de técnicas computacionales, matemáticas y estadísticas para el análisis, interpretación y generación de datos biológicos. Según el National Institute of Health (NIH) es: “ **La investigación desarrollo o aplicación de herramientas computacionales para expandir el uso de datos biológicos, médicos, o de comportamiento, incluyendo las herramientas para adquirir, guardar, organizar, analizar o visualizar dicha información** ”

2) Algunos datos Históricos:

- 1958 - Primera secuencia de Proteína
- 1975 – Primera Secuencia de ADN
- 1986 Desarrollo de PCR (Polymerase Chain Reaction) – Generar copias de fragmentos de ADN
- Inicio de la era genómica.

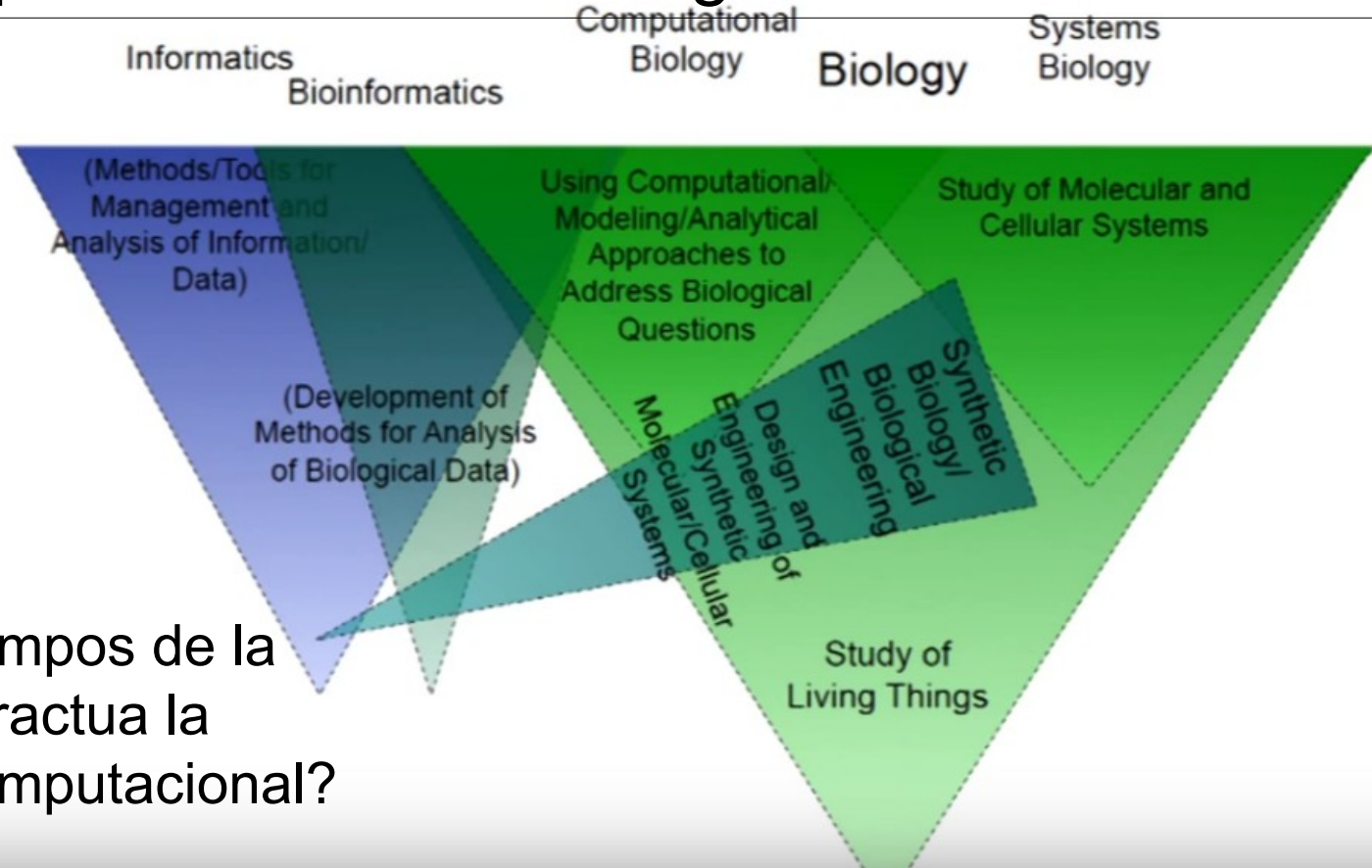


### 3. Conceptos Básicos de Biología Molecular

- 1) Por que usar Biología computacional ?
- 2) Por que ha llegado a ser tan importante en estos ultimos años?



# 3. Conceptos Básicos de Biología Molecular



- Con que campos de la ciencia interactua la Biologia Computacional?



### 3. Conceptos Básicos de Biología Molecular

#### 1) Con que campos de la ciencia interactúa la Biología Computacional?

- **Genómica** - Ciencia que estudia el Genoma (Todos los Genes de un organismo)
- **Transcriptómica** – El estudio de los elementos involucrados en el proceso de transcripción de información del Genoma para la generación de proteínas
- **Proteómica** – Estudio de las Proteínas de los organismos
- **Metabolismo** – Procesos Físicos y químicos
- **Análisis Filogenéticos** – Clasificación de las especies según su información genética.





# 3. Conceptos Básicos de Biología Molecular

## Algunas aplicaciones de la Proteómica y Genómica



# 3. Conceptos Básicos de Biología Molecular

## Algunas aplicaciones de la Proteómica y Genómica

- Entendimiento de los procesos biológicos básicos
- Diagnóstico y tratamiento de enfermedades
- Diseño de medicamentos
- Protección de la vida de plantas y animales
- Entender los principios de la resistencia de las bacterias
- Resolver problemas del medio ambiente
- Desarrollar nuevas fuentes de energía
- Mejorar procesos industriales
- Estudiar los cambios evolutivos en los organismos





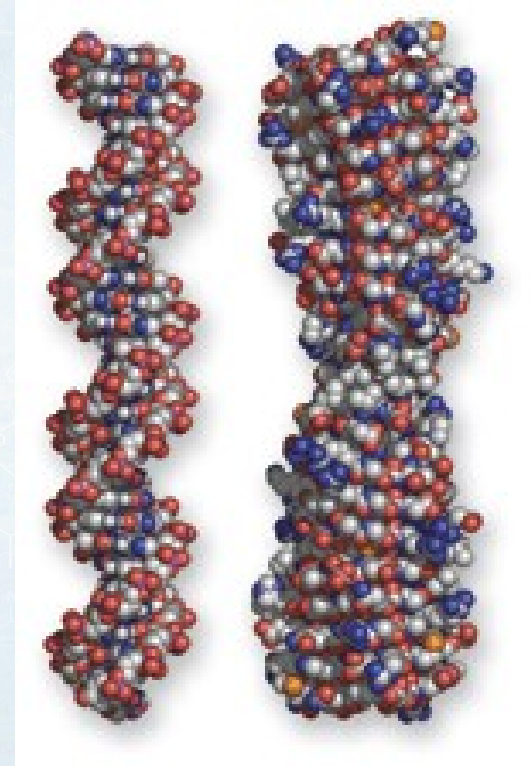
### 3. Conceptos Básicos de Biología Molecular

ADN

mfpA

#### Entender la resistencia de las Bacterias a los Antibióticos

- El antibiótico Fluoroquinolone mata la bacteria de la tuberculosis al adherirse a una proteína vital (complejo DNA-DNA girasa ).
- La bacteria resistente codifica una nueva proteína mfpA parecida al ADN
- mfpA compite con el ADN para adherirse al antibiótico Fluoroquinolone, haciendola resistente a ese antibiotico.



# 3. Conceptos Básicos de Biología Molecular

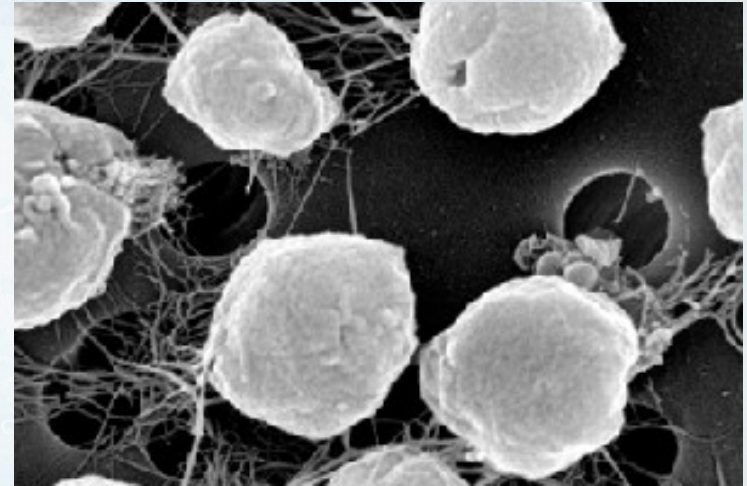
## Methanocaldococcus jannaschii

### Archea que produce metano

Methanocaldococcus jannaschii, produce metano, una importante fuente de energía.

Codifica enzimas que soportan altas temperaturas y presiones.

Puede utilizarse en procesos industriales



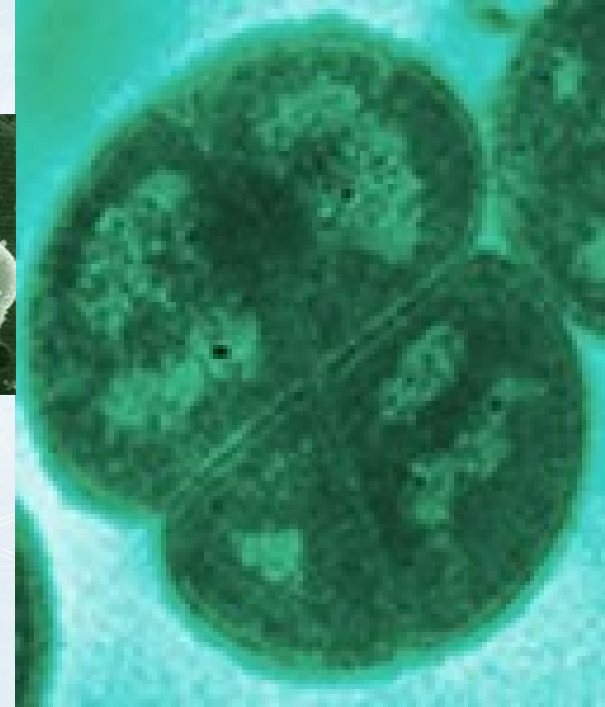
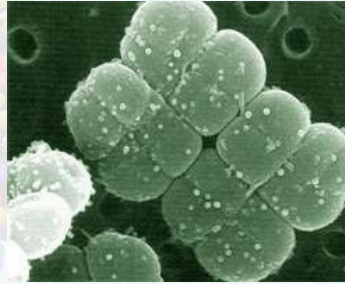


### 3. Conceptos Básicos de Biología Molecular

#### Deinococcus radiodurans

##### **Sobrevive extremadamente altos niveles de radiación**

- Es una bacteria que puede resistir dosis de irradiación hasta de 1500 Kilorads (niveles cientos de veces más altos de lo que soportan otros organismos) Con potencial para limpiar desechos radioactivos.



### 3. Conceptos Básicos de Biología Molecular

#### 1) Que es la Biología Molecular?

- Disciplina Cientifica que tiene como objetivo el estudio de los procesos que se desarrollan en los seres vivos desde el punto de vista molecular.
- Dentro del Proyecto Genoma Humano se definió como “El estudio de la estructura, función y composición de las moléculas biológicamente importantes.”



### 3. Conceptos Básicos de Biología Molecular

- 1) Los Sistemas Biológicos son digitales por naturaleza:  
Las primeras formas de vida basadas en ADN ya estaban guardando, procesando, copiando información digital codificada en 4 letras A,C,G, y T. Esta información genética - digital puede persistir a través de miles de años
- 2) Las ciencias que se benefician del procesamiento de datos, como la Biología computacional, y que demuestran las bondades del procesamiento de grandes cantidades de datos, impulsan el desarrollo y la inversión en nuevas tecnologías para obtener, procesar y analizar cada vez mas data.

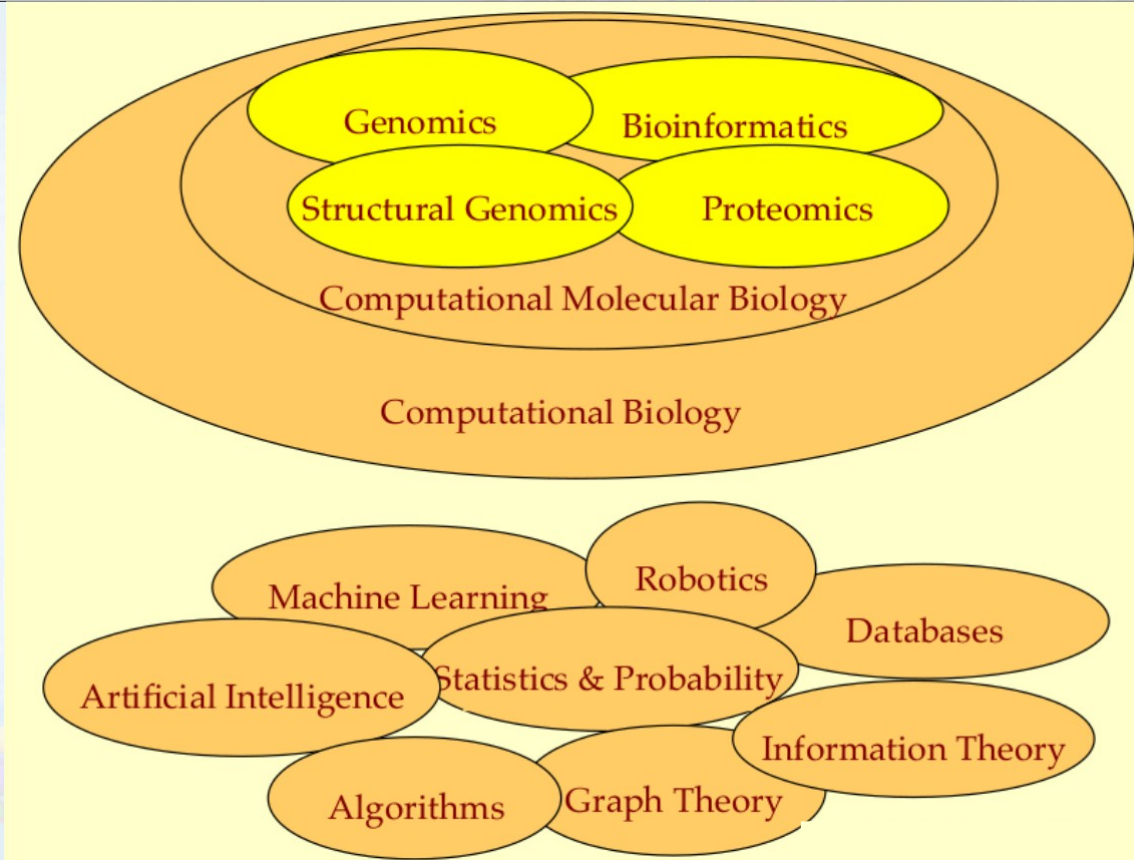


### 3. Conceptos Básicos de Biología Molecular

- 3) Grandes avances en el poder de cómputo en últimos años
- 4) Consideraciones de tiempo de ejecución y requerimientos de memoria
- 5) Algoritmos de Machine learning para reconocer patrones complejos.
- 6) Visualización de información Biológica
- 7) Simulación y modelamiento de información biológica
- 8) Experimentación ética en silicio en vez de en organismos vivos



### 3. Conceptos Básicos de Biología Molecular



# 3. Conceptos Básicos de Biología Molecular

## Objetivos de la biología computacional

**Aprender y Generalizar:** Descubrir estructuras conservadas, modelos de secuencias, estructuras, interacciones, metabolismos y química a partir del estudio de muestras

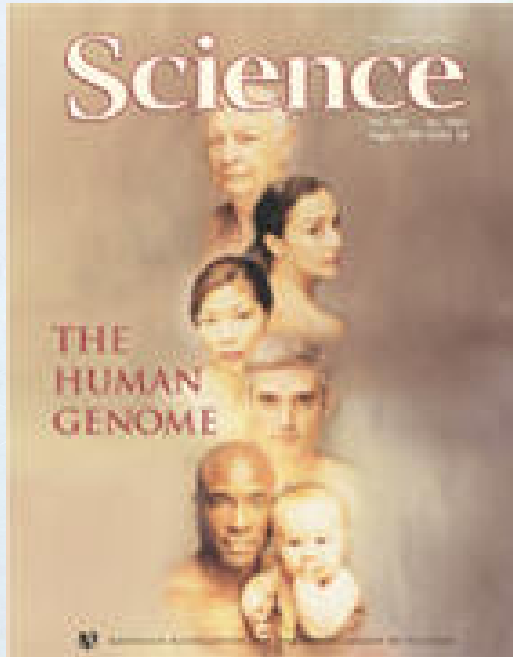
- **Predicción:** Inferir la función o estructura de nuevos genes secuenciados, de genomas, proteínas o proteomas a partir de estas generalizaciones
- **Organizar e integrar:** Desarrollar métodos sistemáticos para estudiar interacciones moleculares, el metabolismo, expresión de genes, señales celulares...
- **Simulación:** Modelar la expresión de genes, la regulación de genes, la configuración de las proteínas, la interacción entre proteínas, metabolismos...
- **Ingeniería:** Construir: Nuevos organismos, funciones, regulación de genes y proteínas
- **Terapia Genética:** Estudiar Genes específicos o mutaciones para el tratamiento de enfermedades.



### 3. Conceptos Básicos de Biología Molecular

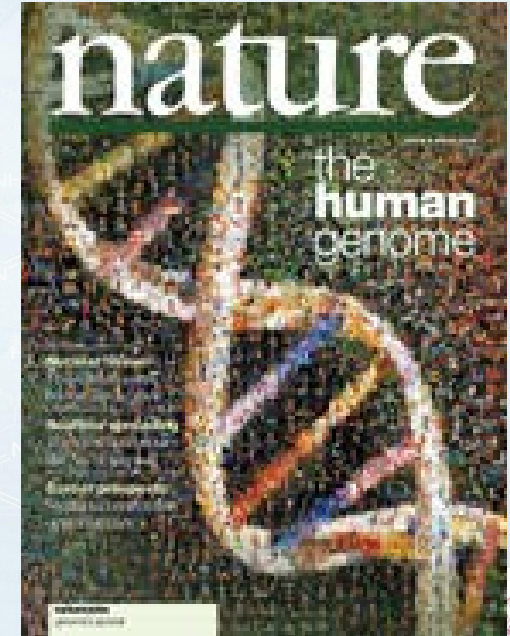
El Genoma Humano es secuenciado - Junio 2003.

Celera



Craig Venter ; Francis Collins

Proyecto Genoma Humano



# 3. Conceptos Básicos de Biología Molecular

El Genoma Humano es secuenciado - Junio 2003.

- 23 pares de Cromosomas
- 3.2 Billones de bases (Base Pairs bp)
- Cerca de 30,000 Genes
- Solo el 2 % del genoma humano “codifica”
- El tamaño promedio de un Gene es 4000 bp
- El Gene que codifica la Distrofina es uno de los mas largos – 2.4 Mbp.
- El 99.9 % del Genoma de 2 personas es identico.



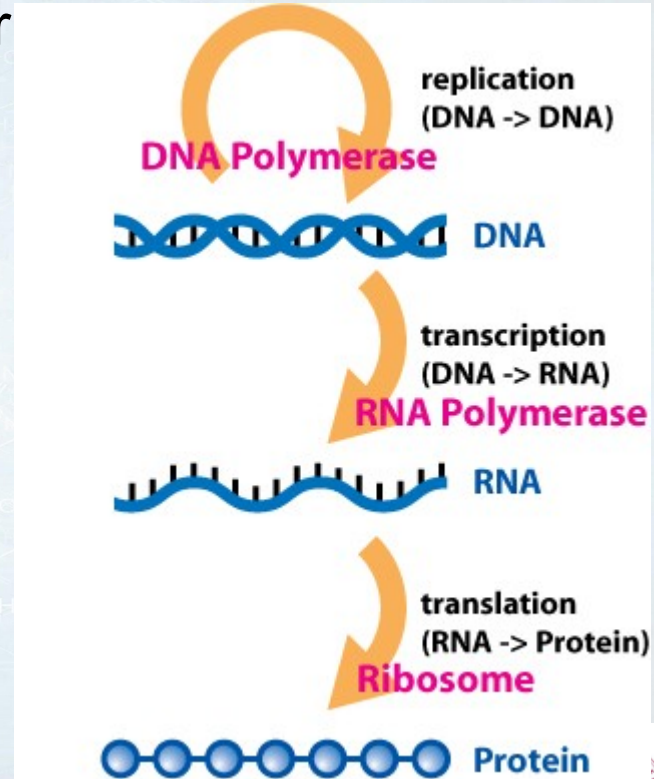


### 3. Conceptos Básicos de Biología Molecular

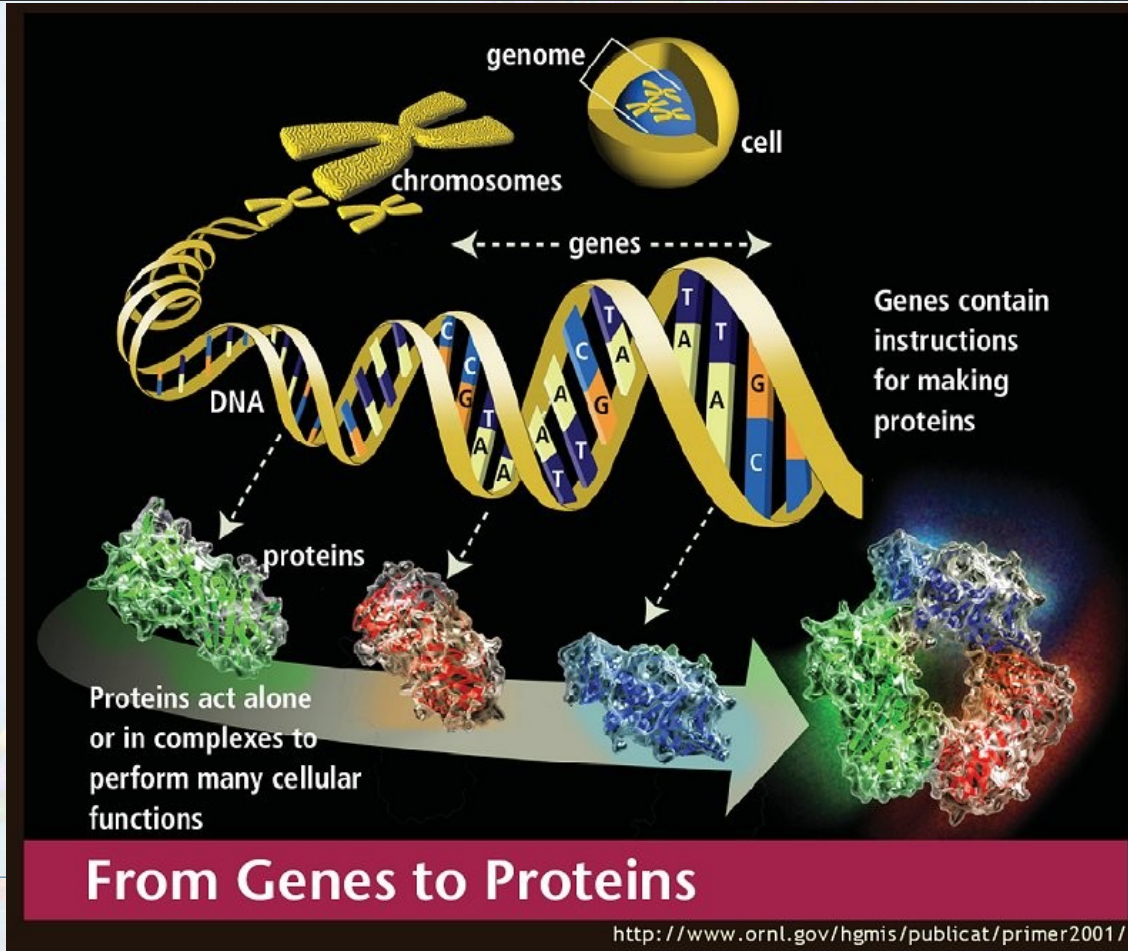
#### El Dogma Central de la Biología Molecular

ADN  $\longrightarrow$  ARN  $\longrightarrow$  Proteína

Muestra cómo la información genética es grabada e interpretada en la célula:  
El Código genético de un organismo se graba en el ADN, el cual se transcribe en el ARN, el cual finalmente se traduce en una proteína.  
Las proteínas llevan a cabo la mayoría de las funciones de la célula, tales como la habilidad de moverse (Motilidad), la regulación y la replicación.  
Algunos Virus y Retrovirus son excepciones a este Dogma



### 3. Conceptos Básicos de Biología Molecular





DNA double helix

Chromosome

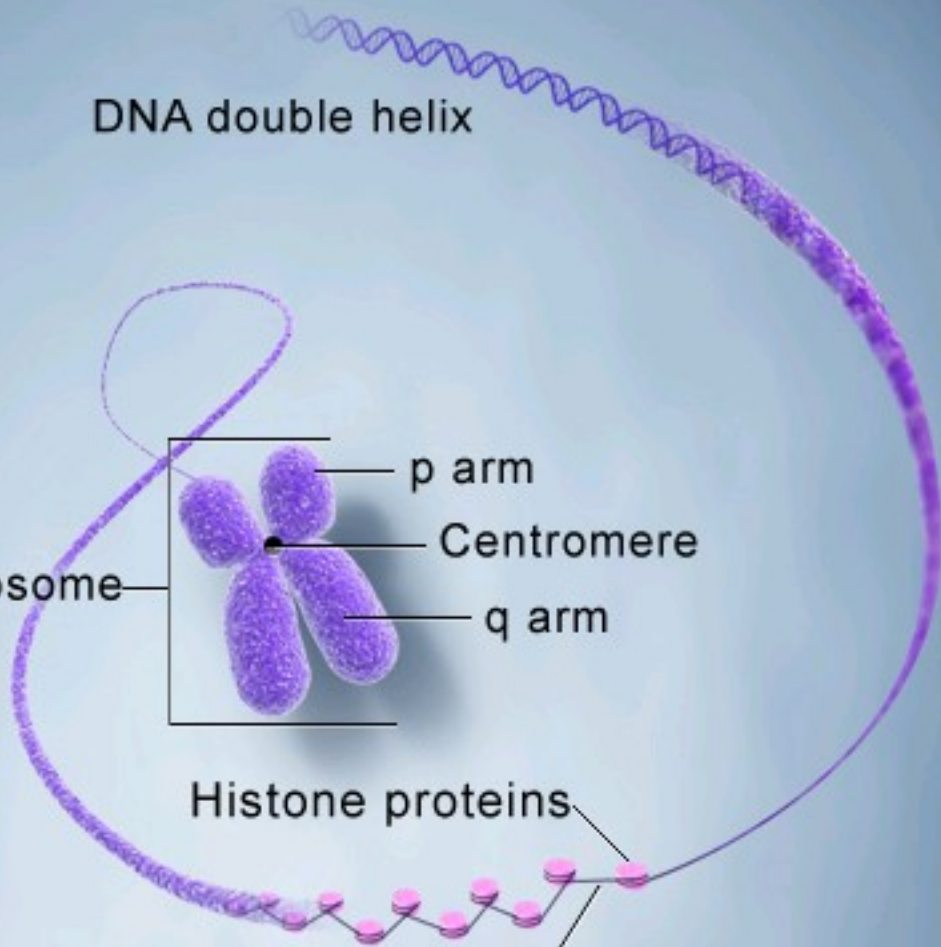
p arm

Centromere

q arm

Histone proteins

DNA



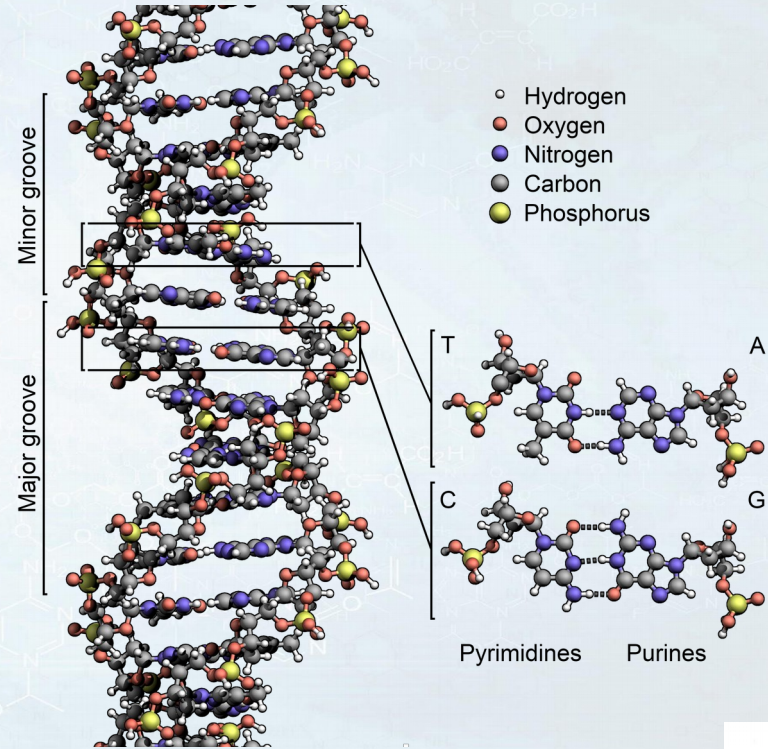
# 3. Conceptos Básicos de Biología Molecular

ADN → ARN → Proteína

ADN – Acido Desoxirribonucleico. Macro molecula que contiene la información genética de un organismo. Contiene regiones llamadas genes que codifican las proteínas a producirse. Existen otras regiones regulatorias que influncian el nivel de expresion de los genes.

Estructura: El ADN está compuesto de 4 Nucleótidos: A(Adenina), C(Citosina), T(Timina) y G(Guanina). A y G son Purinas (Dos anillos) mientras C y T son Pirimidinas (Un anillo).

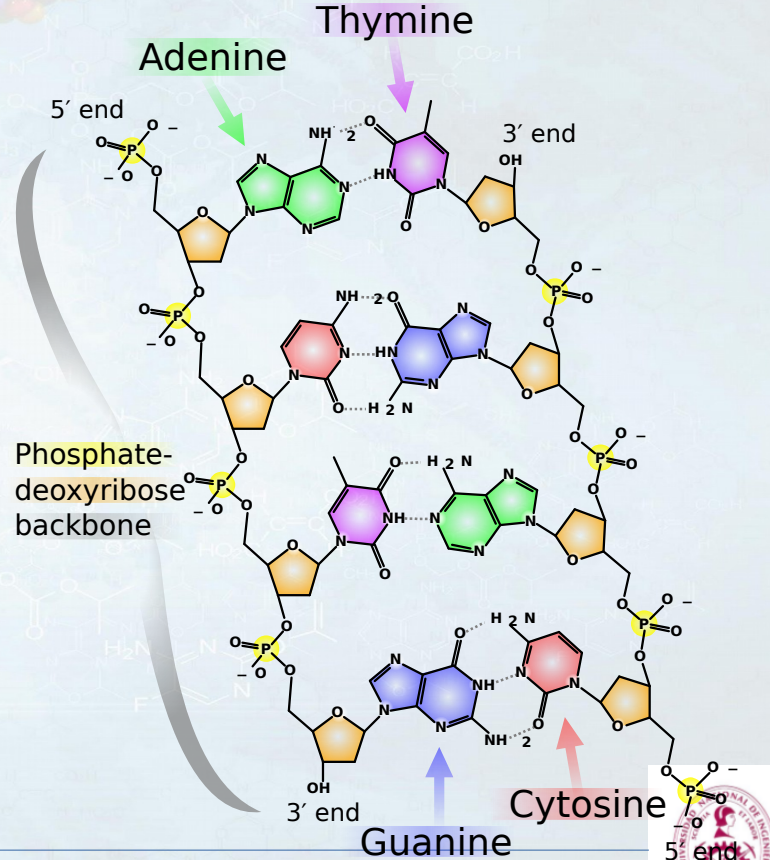
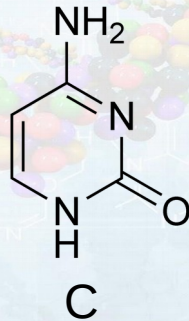
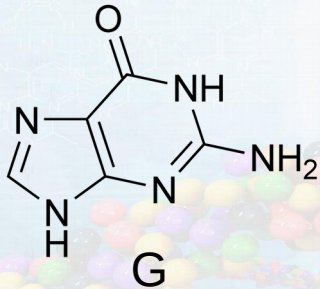
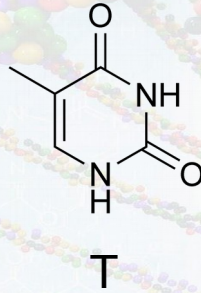
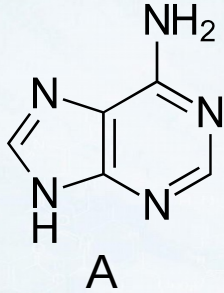
A y T se conectan por 2 enlaces de hidrógeno, mientras que C y G con tres enlaces ,por lo que estos enlaces son más fuertes. -Las bacterias que viven en aguas termales tienen el 80% de G-C





# 3. Conceptos Básicos de Biología Molecular

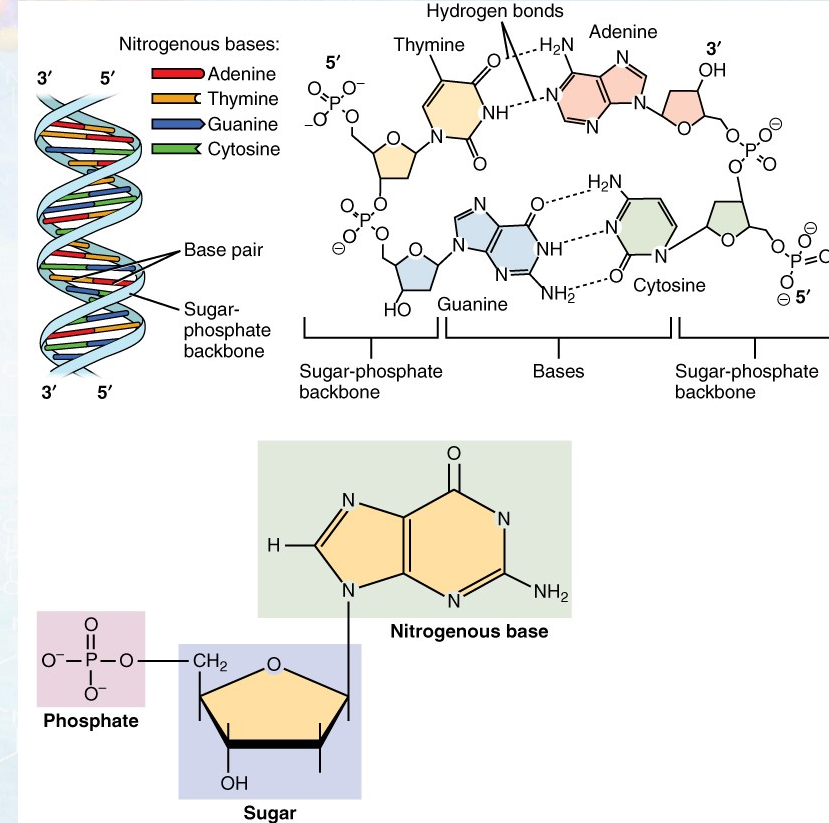
ADN → ARN → Proteína



# 3. Conceptos Básicos de Biología Molecular

ADN → ARN → Proteína

Los dos hilos del ADN son **complementarios**, es decir los enlaces A-T y T-A ; G-C y C-G siempre ocurren en ese orden y entre esos pares de bases. Los Hilos del ADN también tienen **direccionalidad** que se refiere a las posiciones en el anillo de la pentosa, donde se conecta la columna de fosfatos. Generalmente las enzimas ADN y ARN polimerasa se sintetizan en la dirección 5' a 3' y por convención se escriben de 5' a 3'.

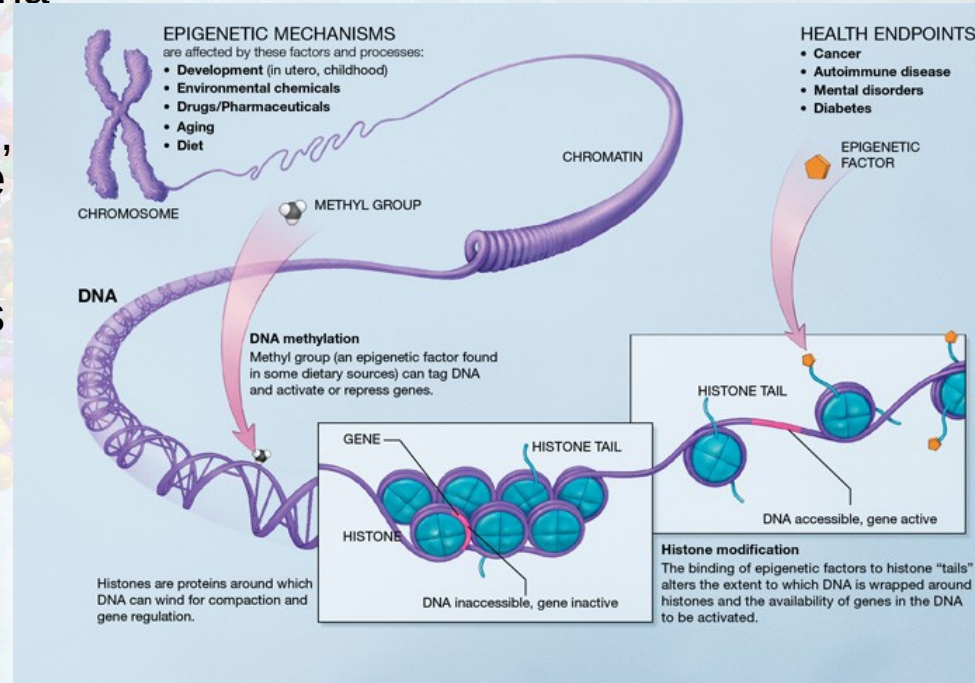




# 3. Conceptos Básicos de Biología Molecular

ADN → ARN → Proteína

En los niveles de Estructura del ADN, aparte de la asociación por pares de bases, (Niveles 1 y 2) Están la torsión de ADN que forma la escalera en forma helicoidal (Nivel 3), y la forma como el ADN se compacta al rededor de pequeñas proteínas llamadas Histones. Estas estructuras de ADN-Histones se compactan aún más en una estructura de cromatina. La expresión de los Genes (o su activación) se puede regular modificando esta cromatina.



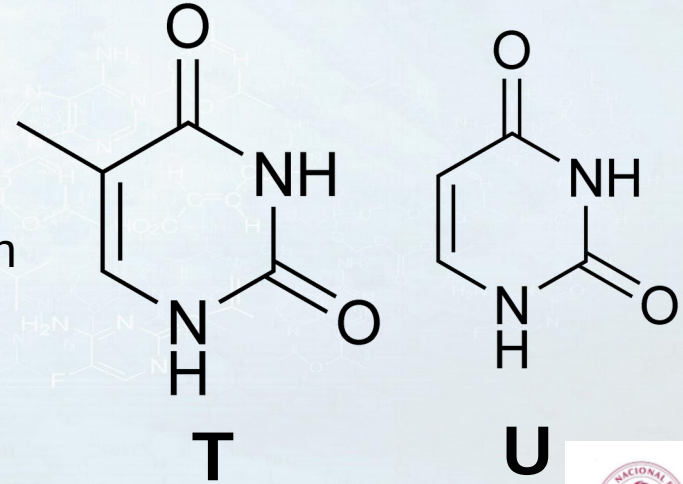
### 3. Conceptos Básicos de Biología Molecular

La Estructura del ADN con sus enlaces débiles de hidrógeno permite que los hilos puedan ser fácilmente separados para la Replicación de ADN, lo que también permite la transcripción (a ARN), la traducción (A proteína), la recombinación de fragmentos y la reparación de ADN, entre otros procesos.

En la replicación la copia de ADN resultante cuenta con un Hilo de origen y un Hilo con bases complementarias nuevas

ADN  ARN  Proteína

La Transcripción, es el proceso por el cual se produce ARN usando ADN como una plantilla. Una enzima -(ARN polimerasa) lee el ADN en la dirección de 3' a 5' y coloca bases complementarias para formar ARN- mensajero (mRNA) el ARN utiliza los mismos nucleótidos que el ADN sólo que reemplaza la Timina por Uracilo.





### 3. Conceptos Básicos de Biología Molecular

ADN → **ARN** → Proteína

El ARN se produce cuando se transcribe el ADN, es estructuralmente similar al ADN, con las siguientes diferencias:

1. El nucleótido Uracilo se utiliza en lugar de Timina
2. ARN contiene Ribosa en lugar de Desoxiribosa. A Desoxiribosa le falta una molécula de oxígeno en la posición 2' encontrada en la Ribosa.
3. El ARN solo tiene un Hilo mientras ADN tiene 2.

Las moléculas de ARN son un paso intermedio para codificar una proteína.

### 3. Conceptos Básicos de Biología Molecular

ADN → **ARN** → Proteína

Existen varios tipos de ARN

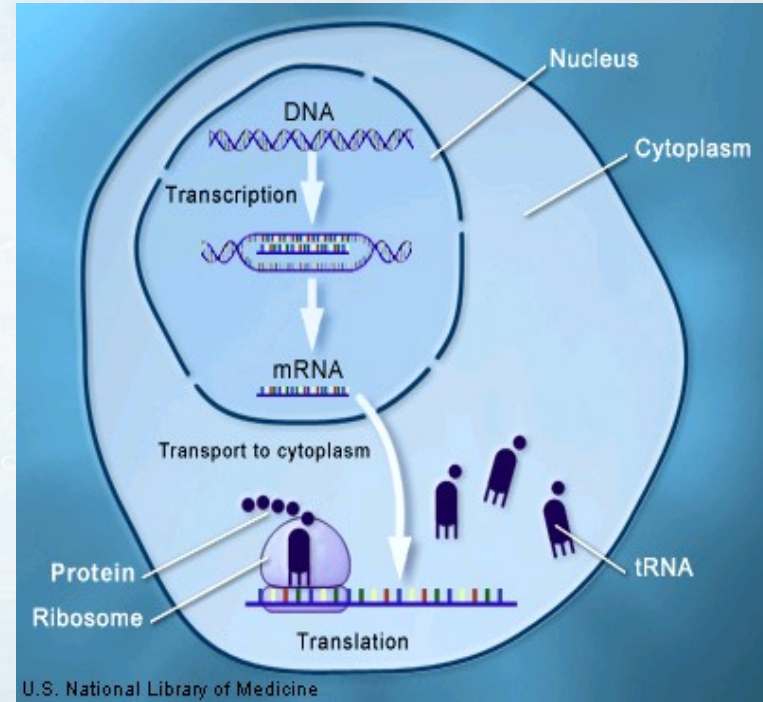
1. ARN – Mensajero (mRNA) Contiene información para hacer proteínas y es traducido en una secuencia de proteína.
2. ARN de Transferencia (tRNA) Especifica la traducción de codón a aminoácido. Contiene tres bases complementarias (anti-Codón) al codón del mRNA, y carga el aminoácido correspondiente a su anti-codón adherido a su extremo 3'
3. ARN – Ribosomal (rRNA) forma el corazón del ribosoma – que es el organelo responsable de la traducción de mRNA a Proteína.
4. snRNA - “Small Nuclear RNA” se encarga de remover introns, es decir segmentos que no codifican fragmentos de proteínas, de las secuencias del mRNA.
5. Existen otras clases de RNA que aún se están estudiando.



### 3. Conceptos Básicos de Biología Molecular

ADN  $\longrightarrow$  ARN  $\longrightarrow$  Proteína

**Traducción.** La estructura primaria de una proteína esta dada por la secuencia de aminoácidos que la conforman. Ya que existen 20 aminoácidos posibles, estos son codificados utilizando los 4 nucleótidos que se utilizan en el mRNA, de hecho se codifican mediante una secuencia de 3 nucleótidos en mRNA denominados codones. Un determinado codon especifica un aminoácido, o constituye un codon de stop, el cual finaliza la traducción de la proteína. Ya que existen 64 secuencias de codones y solo 20 aminoácidos – se dice que el código es degenerado, es decir algunos aminoácidos se especifican por múltiples codones.



# 3. Conceptos Básicos de Biología Molecular

ADN → ARN → Proteína

Codificación de aminoácidos mediante codones en el mRNA

na

SECOND POSITION

FIRST POSITION

THIRD POSITION

	U	C	A	G	
U	phenyl-alanine	serine	tyrosine	cysteine	U
	leucine		stop	stop	C
			stop	tryptophan	A
					G
C	leucine	proline	histidine	arginine	U
			glutamine		C
					A
					G
A	isoleucine	threonine	asparagine	serine	U
	* methionine		lysine	arginine	C
					A
					G
G	valine	alanine	aspartic acid	glycine	U
			glutamic acid		C
					A
					G

\* and start



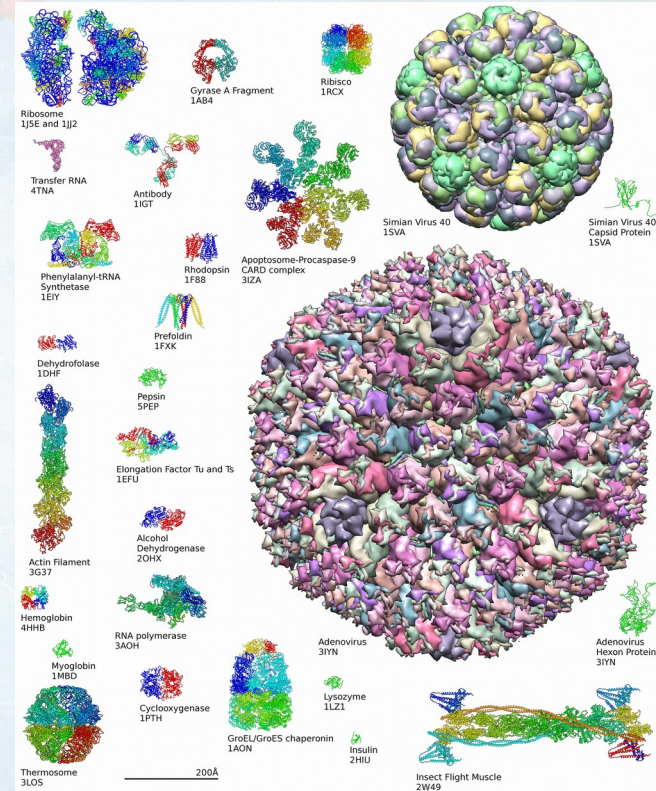


# 3. Conceptos Básicos de Biología Molecular

ADN → ARN → Proteína

Proteína. Es la molécula responsable de la mayoría de las funciones y tareas de la célula. Puede tener diversas funciones: enzimáticas, transporte, sistema inmunológico. Las proteínas están constituidas por cadenas de aminoácidos. Cada Aminoácido tiene propiedades especiales de tamaño, carga, forma, y acidez. Además de la secuencia de aminoácidos, la interacción entre ellos determina la forma tridimensional de la proteína y por consiguiente su función.

**La tarea de determinar la forma de una proteína a partir de su secuencia, es un problema aun no resuelto en Biología Computacional**





# 3. Conceptos Básicos de Biología Molecular

ADN → ARN → Proteína

## Ejemplos de Proteínas:

**La Hemoglobina** Lleva el oxígeno a todas partes del cuerpo

**Canales Iónicos** – Proteínas que controlan el flujo de señales en el cerebro permitiendo el flujo de moléculas en las neuronas

**Enzimas** en la saliva, el estómago y el intestino delgado ayudan a la digestión de los alimentos  
Proteínas en los Musculos Actina y miocina permiten el movimiento muscular

**Anticuerpos** Son proteínas que ayudan a defender el cuerpo contra virus y bacterias.

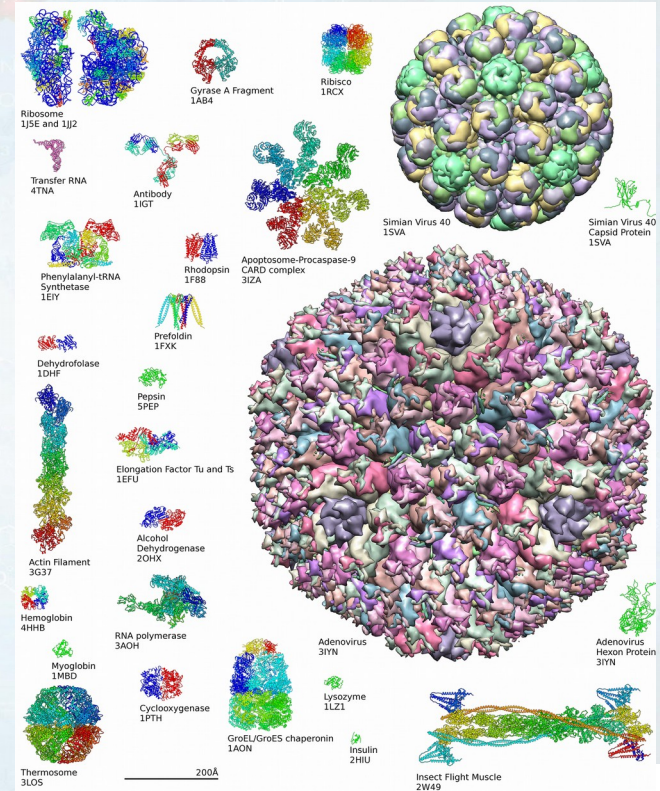




Diagram illustrating the structure and function of a DNA sequence. The sequence is shown with various motifs highlighted in red and green. The sequence is divided into two main regions: Genes and Regulatory motifs.

**Genes:** The top section of the sequence, labeled "Genes", contains the coding region. It is represented by a blue box labeled "Genes" and a red box labeled "Encode proteins". The sequence in this region is highlighted in red.

**Regulatory motifs:** The bottom section of the sequence, labeled "Regulatory motifs", contains the regulatory region. It is represented by a blue box labeled "Regulatory motifs" and a green box labeled "Control gene expression". The sequence in this region is highlighted in green.

Arrows indicate the flow of information from the DNA sequence to the protein and regulatory motifs. A blue arrow points from the "Genes" box to the "Encode proteins" box. A green arrow points from the "Regulatory motifs" box to the "Control gene expression" box. A blue arrow points from the "Genes" box to the "Regulatory motifs" box, indicating the transition from the coding region to the regulatory region.

# 4. Problemas Clásicos de Bioinformática

- 1) Secuenciamiento de Proteínas
- 2) Secuenciamiento de ADN
- 3) Organizar la Información en bases de datos que los investigadores puedan acceder y actualizar
- 4) Programas y algoritmos para el análisis de secuencias
- 5) Alineamiento de secuencias
- 6) Encontrar alineamientos locales en secuencias
- 7) Descubrir alineamientos en múltiples secuencias
- 8) Descubrir relaciones evolutivas utilizando secuencias
- 9) Interpretar la información de manera biológicamente útil
- 10) Predecir la secuencia de proteínas a partir de la traducción de secuencias de ADN.



## 5. Que es lo que sigue?

- Determinar la función de los genes y los elementos que regulan los genes a lo largo del genoma
- Encontrar variaciones en la secuencia de ADN entre las personas y determinar su significado. Estas pequeñas diferencias pueden ayudar a predecir el riesgo de una persona en particular a padecer ciertas enfermedades y su respuesta a ciertos medicamentos.
- Descubrir las estructuras tridimensionales de las proteínas e identificar sus funciones.
- Explore cómo el ADN y las proteínas interactúan entre sí y con el ambiente para crear sistemas vivos complejos.
- Desarrollar y aplicar estrategias basadas en el genoma para la detección temprana, diagnóstico y tratamiento de la enfermedad.
- Secuencia los genomas de otros organismos, para comparar genes similares entre especies.
- Desarrollar nuevas tecnologías para estudiar genes y ADN a gran escala y almacenar datos genómicos de manera eficiente.
- Continuar explorando los problemas éticos, legales y sociales planteados por investigación genómica.

## 5. Que es lo que sigue?

**Proyecto HAPMAP:** (Haplotype Map) El proyecto de mapeo de variaciones Genéticas. Describe las variaciones genéticas, dónde están localizadas en el Genoma y que tan comunes son en diferentes poblaciones a través del mundo

- Descubrimiento de genes relacionados a enfermedades,
- Terapias Genéticas

**Pharmacogenomics:** Farmacología y Genómica.

- Descubrimiento de medicamentos efectivos basados en el perfil genético
- Reducción de reacciones adversas

**ENCODE:** Enciclopedia de elementos funcionales

- El Estudio de la expresión de los genes.



## 6. Recursos en linea

<http://www.ncbi.nlm.nih.gov/> (Nacional Center for Biotechnology Information – NCBI)

<http://www.ebi.ac.uk/> (European Bioinformatics Institute- EBI- EMBL)

<http://www.ddbj.nig.ac.jp/> (DNA Databank of Japan- DDBJ)

<http://www-nbrf.georgetown.edu> (Protein Information Resource- PIR)

<http://www.uniprot.org/> (uniprot) resource of protein sequence and functional information

<http://www.rcsb.org/> (Protein Data Bank- PDB)

<http://ndbserver.rutgers.edu/> (Nucleic Acid Database – NDB)

**NCBI Home**

**Resource List (A-Z)**

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

### Submit

Deposit data or manuscripts  
into NCBI databases



### Download

Transfer NCBI data to your  
computer



### Learn

Find help documents, attend a  
class or watch a tutorial



### Develop

Use NCBI APIs and code  
libraries to build applications



### Analyze

Identify an NCBI tool for your  
data analysis task



### Research

Explore NCBI research and  
collaborative projects



## Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[PubMed Health](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

## NCBI News & Blog

Yellow fever mosquito, 6 other organisms  
in July RefSeq genome annotations

17 Aug 2017

In July, the NCBI Eukaryotic Genome  
Annotation Pipeline released new

August 30 NCBI Minute: NCBI  
Hackathons, a framework for rapid  
prototyping of pipelines for bioinformatics,  
biomedical informatics and genomics



# EMBL-EBI



The home for big data in biology

Our unique Search service helps you explore dozens of biological data resources.

[More about EBI Search >](#)

All

Find a gene, protein or chemical



Example searches: [blast](#) [keratin](#) [bfl1](#)

Find a tool for your data analysis.

[Find a tool >](#)

Share your scientific data with the world.

[Deposit data >](#)

About this image: An [entry from the Tara Oceans expedition](#) in the prototype Image Data Repository, built by scientists at the University of Dundee, EMBL-EBI, the University of Bristol and the University of Cambridge. [Learn more >](#)

## Explore EMBL-EBI and our mission

The European Bioinformatics Institute (EMBL-EBI) shares [data from life science experiments](#), performs [basic research](#) in computational biology and offers an extensive [user training programme](#), supporting researchers in academia and [industry](#). We are part of [EMBL](#), Europe's flagship laboratory for the life sciences.

[More about EMBL-EBI and our impact >](#)

**Services**

We provide freely available

**Research**

We contribute to the

**Training**

We provide advanced

**Industry**

We help disseminate cutting-

**ELIXIR**

We support, as an ELIXIR

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

## UniProtKB

UniProt Knowledgebase

Swiss-Prot (555,100)



Manually annotated  
and reviewed.

TrEMBL (88,032,926)

Automatically  
annotated and not  
reviewed.

## UniRef

Sequence clusters



## UniParc

Sequence archive



## Proteomes



## Supporting data

Literature citations



Taxonomy



Subcellular locations



Cross-ref. databases



Diseases



Keywords



## News



A pseudogene turns into an active DNA methyltransferase dedicated to male fertility

[UniProt release 2017\\_06](#)

Eukaryotic sex: good ideas shared with viruses | Change of cross-references to PATRIC | New file listing deleted and subsequently reinst...

[UniProt release 2017\\_05](#)

A certain taste for light | Extension of controlled vocabulary for PTM to glycosylation sites | Addition of GO annotation to UniRef90 an...

[News archive](#)

## Getting started

 [Text search](#)

Our basic text search allows you to search all the resources available



## UniProt data

 [Download latest release](#)

Get the UniProt data

## Protein spotlight



### Becoming One

July 2017

There are different ways of producing



## Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

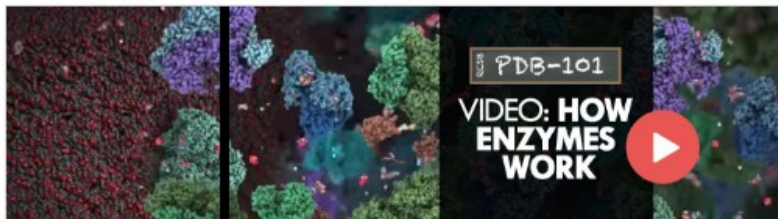
## A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

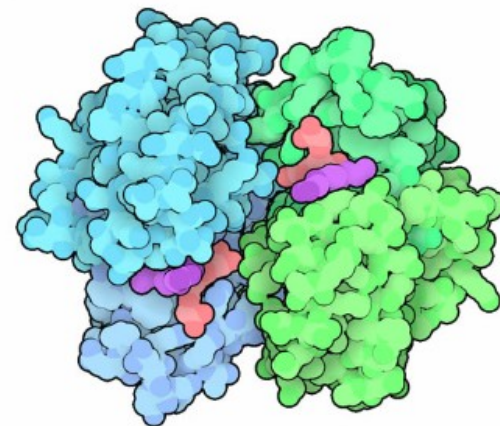
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

### Video: How Enzymes Work



## August Molecule of the Month



Glutathione Transferases

## 6. Recursos en Linea - NCBI

Find bioassays in which a given drug is active  
Find bioassays that test a particular disease or protein target  
Save text searches and set up automated searches with E-mailed results  
Submit data to NCBI  
Submit sequence data to NCBI  
Retrieve all sequences for an organism or taxon  
Find the function of a gene or gene product  
Find expression patterns  
Find genes associated with a phenotype or disease  
Obtain genomic sequence for/near a gene, marker, transcript or protein  
View all SNPs associated with a gene  
Find human variations associated with a phenotype or disease  
Display genomic annotation graphically  
Compare protein homologs between two microbial genomes  
Find sequenced genomes, including those in progress, for a taxonomic group  
Download the complete genome for an organism