

Técnicas Avanzadas de Biología Molecular y Bioinformática

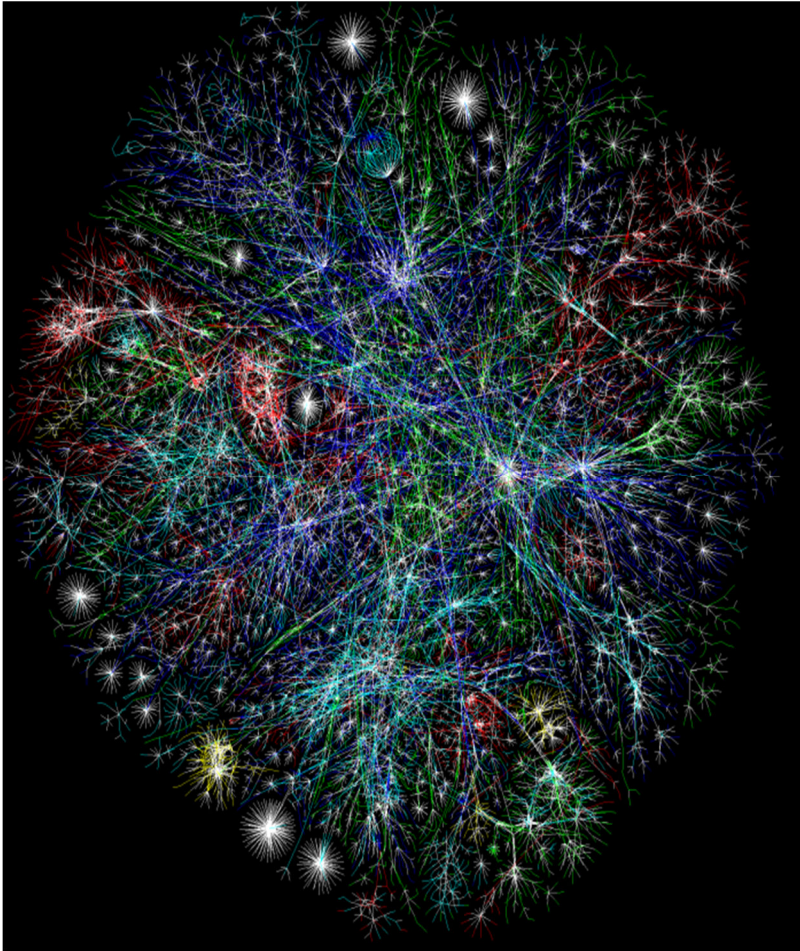
Ascanio Rojas A.

Centro Nacional de Cálculo Científico
Universidad de Los Andes (CeCalCULA), Mérida - Venezuela
email: ascanio@ula.ve

La **Bioinformática** es la aplicación de tecnología de computadores a la gestión y análisis de datos biológicos (1). **Bioinformática** es la rama de las Ciencias de la Computación usada en la investigación, desarrollo o aplicación de herramientas computacionales y aproximaciones para la expansión del uso de datos biológicos, médicos, conductuales o de salud, incluyendo aquellas herramientas que sirvan para adquirir, almacenar, organizar, analizar o visualizar tales datos. A diferencia de la Biología computacional que sería el desarrollo y aplicación de métodos teóricos y de análisis de datos, modelado matemático y técnicas de simulación computacional al estudio de sistemas biológicos, conductuales y sociales.

Una de las principales aplicaciones de la bioinformática es la simulación, la minería de datos (*data mining*), y el análisis de los datos obtenidos en el estudio de moléculas relevantes para la vida, principalmente del ADN/ARN/genoma (Proyecto Genoma Humano) o de las proteínas (cuyo conjunto en un determinado organismo biológico forma su proteoma), así como el diseño y desarrollo de herramientas tales como bases de datos, directorios web, etc.

Informática



Una **red informática** es un conjunto de equipos (computadoras o dispositivos) conectados por medio de cables, señales, ondas o cualquier otro método de transporte de datos, que comparten información (archivos), recursos y servicios.

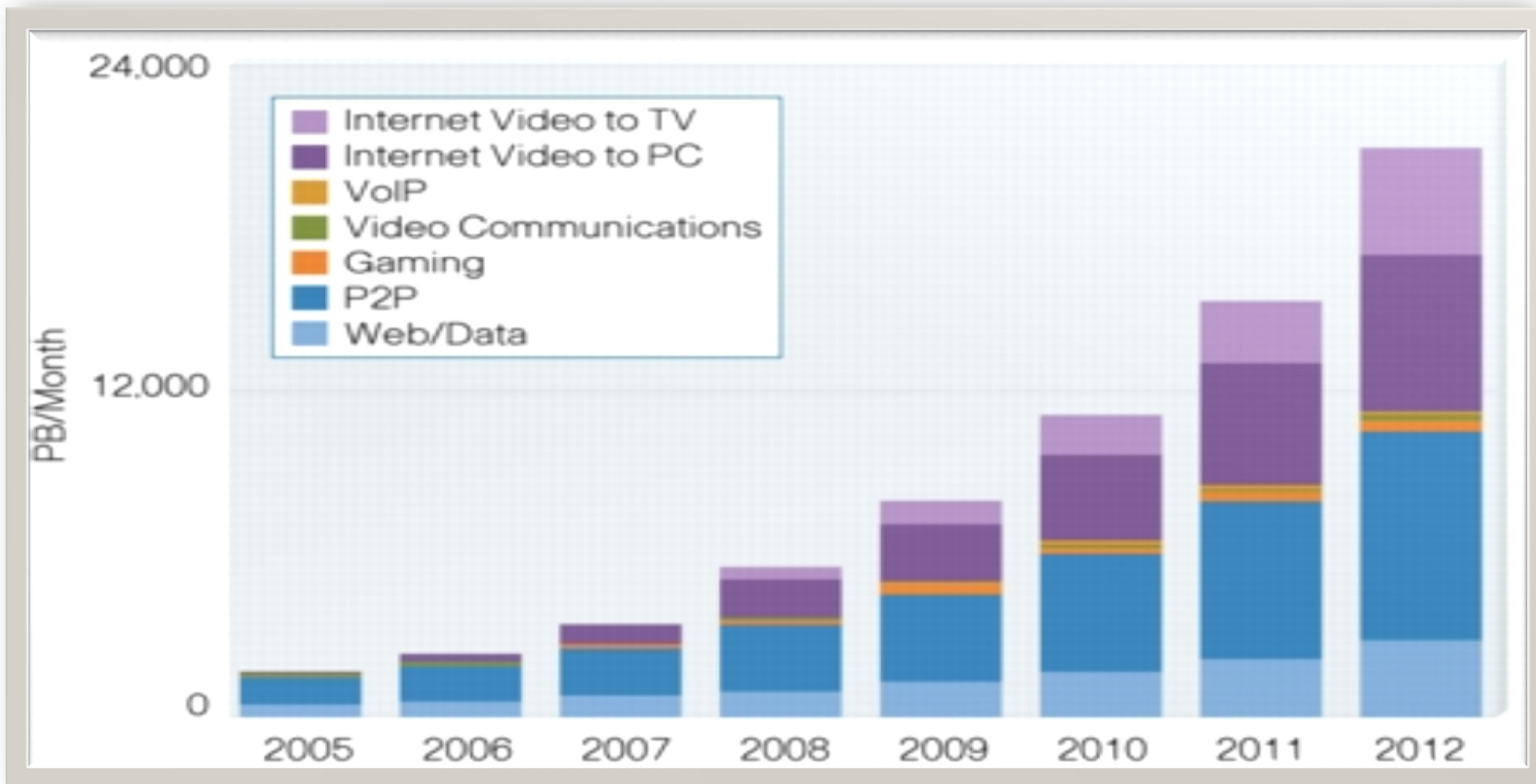
Internet es un método de interconexión descentralizada de redes de computadoras implementado en un conjunto de protocolos denominado TCP/IP y garantiza que redes físicas heterogéneas funcionen como una red lógica única, de alcance mundial. Sus orígenes se remontan a 1969, cuando se estableció la primera conexión de computadoras, conocida como ARPANET, entre tres universidades en California y una en Utah, EE. UU.

World Wide Web (Web) o Red Global Creada en 1990 por el inglés Tim Berners-Lee y el belga Robert Cailliau en el CERN (Ginebra, Suiza)

Asia Pacific / Europe/Middle East/Central Asia/Africa / North America /**Latin American and Caribbean** /RFC1918 IP Addresses / Unknown – White

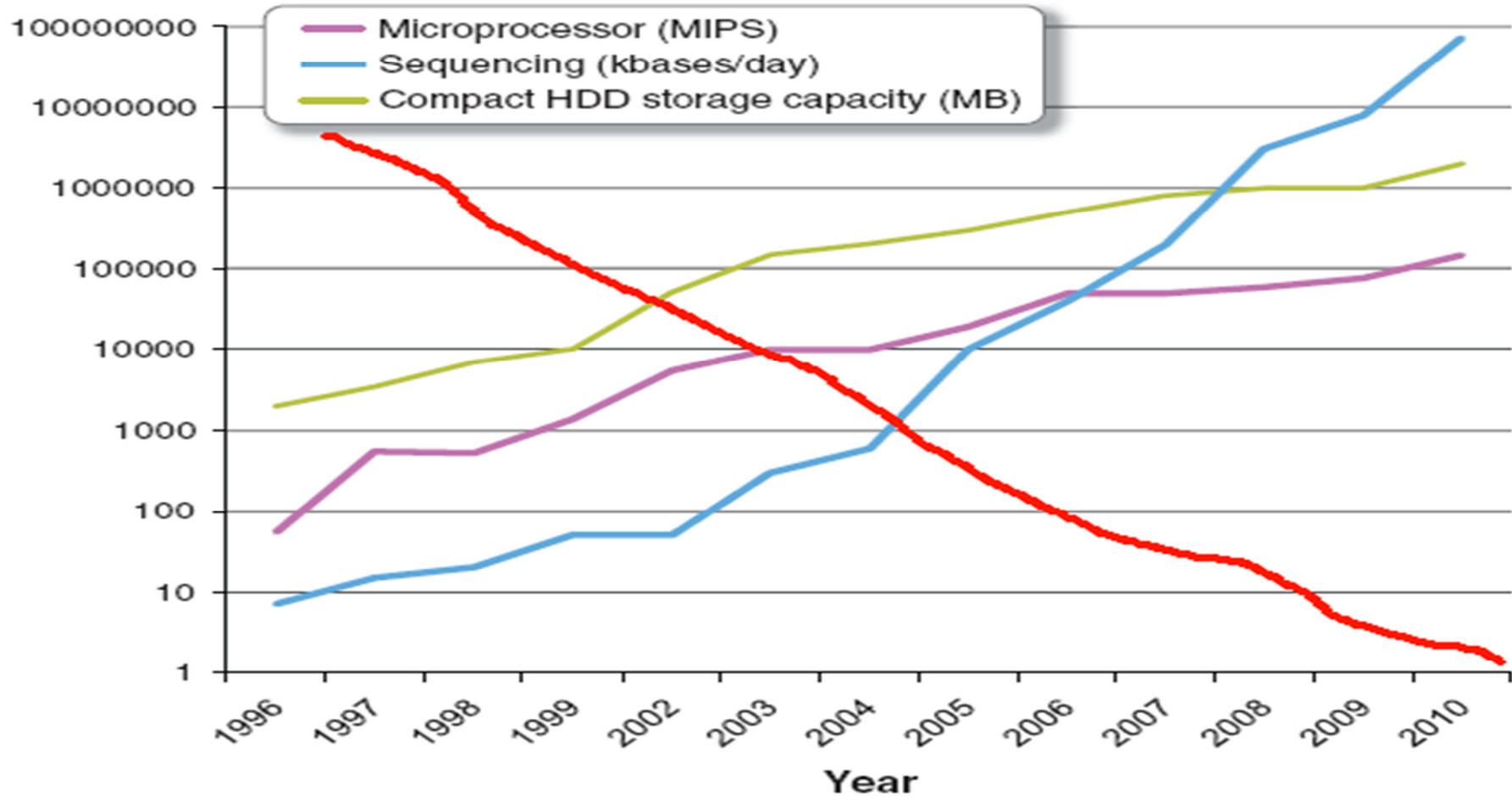
IPv4 4.294.967.296 (2^{32}) direcciones de redes diferentes.

IPv6 340.282.366.920.938.463.463.374.607.431.768.211.456 (2^{128})



Sequencing Progress vs Compute and Storage

Moore's and Kryder's Laws fall far behind



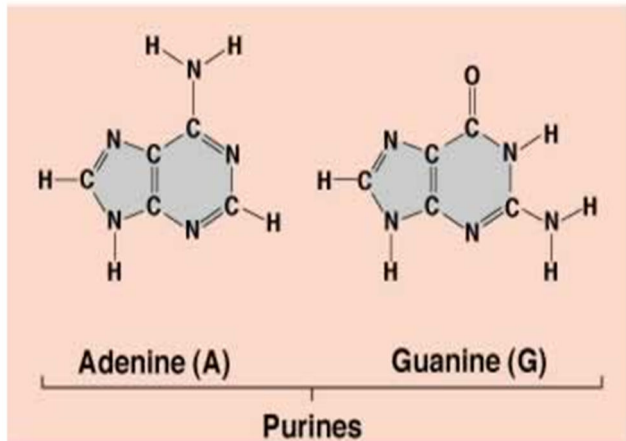
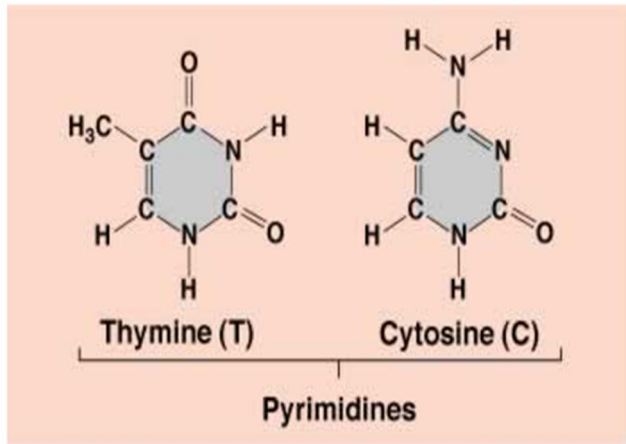
Biología, un poco de historia

- Gregor Mendel describe en 1866 los mecanismos de la herencia
- El bioquímico suizo Friedrich Miescher descubre el ADN en el núcleo de las células, en 1868. El término 'gen' se usa por primera vez en 1909.
- En 1944 se identifica el ADN como el material de la herencia.
- El 28 de febrero de 1953, Francis Crick en el pub Eagle de Cambridge, Inglaterra, y anunció que él y James Watson habían "encontrado el secreto de la vida.", ganan el Nobel por ello.
- Severo Ochoa y Arthur Kornberg dan en 1955 los primeros pasos hacia la interpretación del código genético, (premios Nobel)
- Identificados 23 pares de cromosomas en el ser humano en 1956.
- Científicos de Harvard aíslan el primer gen en 1969. Es de una bacteria.
- En USA se funda en 1976 Genentech, primera ingeniería genética.
- 1977 se decodifica el ADN del phi-X174. 9 genes, 1 cromosoma

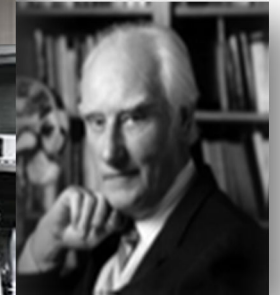
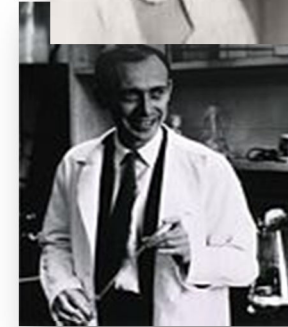
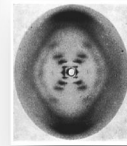
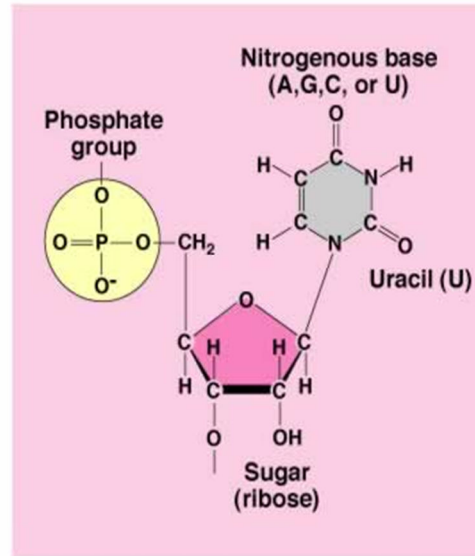


- El gen de la insulina humana se clona en 1978
- Se crean las primeras plantas transgénicas en 1984.
- En 1986 autorizan las pruebas de la vacuna de la hepatitis B, obtenida por ingeniería genética. El Proyecto Genoma Humano inicia en 1990.
- Un tomate genéticamente modificado sale a la venta en California (1994)
- La oveja Dolly, el primer mamífero clonado, nace el 5 de julio de 1996 en Escocia. El único cordero resultante de 277 fusiones de óvulos enucleados con núcleos de células mamarias. Sacrificada el 14 de febrero de 2003.
- Secuenciado el genoma de la levadura de la cerveza en 1996. En la actualidad están secuenciados 812 genomas y existen 3734 proyectos en curso.
- En febrero de 2001 se publica el borrador mapa del genoma humano, en 2003 salen las publicaciones simultáneamente en Science&Nature.
- 2003 inicia el Proyecto ENDODE (the **Encyclopedia Of DNA Elements**) para identificar todos los elementos funcionales en la secuencia del genoma humano. En 2012 anuncian los primeros resultados.

EI ADN



©Addison Wesley Longman, Inc.



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

¹ Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

² Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

³ Von Arx, W. S., *Woods Hole Papers in Phys. Oceanog. Meteor.*, **11** (3) (1950).

⁴ Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey¹. They kindly made their manuscript available to us in advance of

is a residue on each chain every 3.4 Å. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the

Genes

"...and that by means of a known chemical substance it is possible to induce *predictable and hereditary* changes in cells.

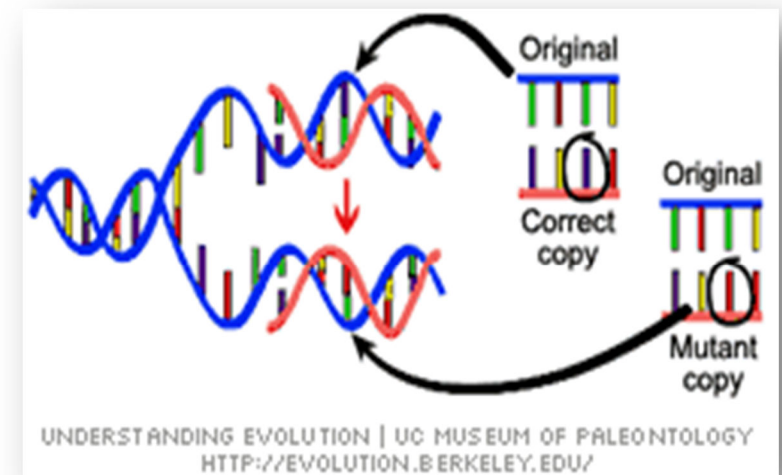
This is something that has long been the dreams of geneticists."

Oswald T. Avery, 1943



Un **gen** es una región localizable de secuencia genómica, lo que corresponde a una unidad de herencia, que está asociada con las regiones reguladoras, regiones transcritas o a cualquier otro tipo de secuencias funcional. *Pearson H (2006).*

"*Genetics: what is a gene?*".



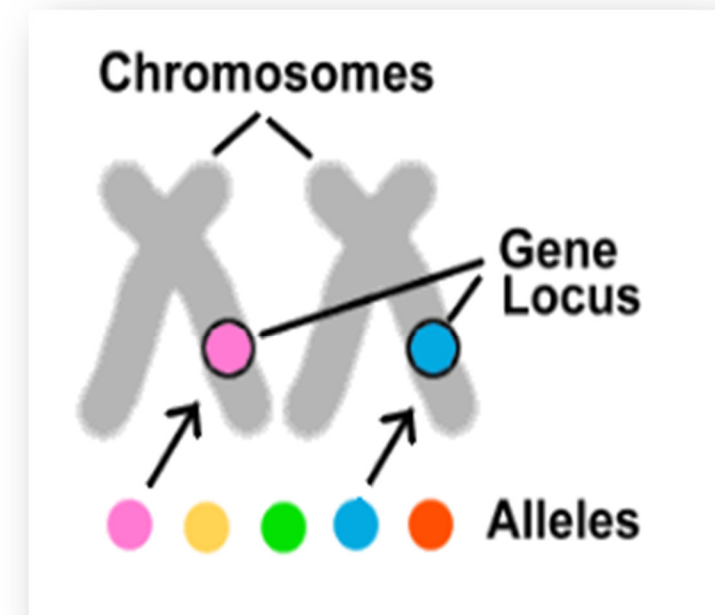
Nature 441 (7092): 398–401.

Genoma



El genoma es el conjunto del material hereditario de un organismo, la secuencia de nucleótidos que especifican las instrucciones genéticas para el desarrollo y funcionamiento del mismo y que son transmitidas de generación en generación.

Cromosoma (del griego *chroma*, color, y *soma*, cuerpo o elemento) es cada uno de los pequeños cuerpos en forma de bastoncillos en que se organiza la cromatina del núcleo celular en la mitosis y la meiosis, cada uno de los cuales se divide longitudinalmente, dando origen a dos cadenas gemelas.

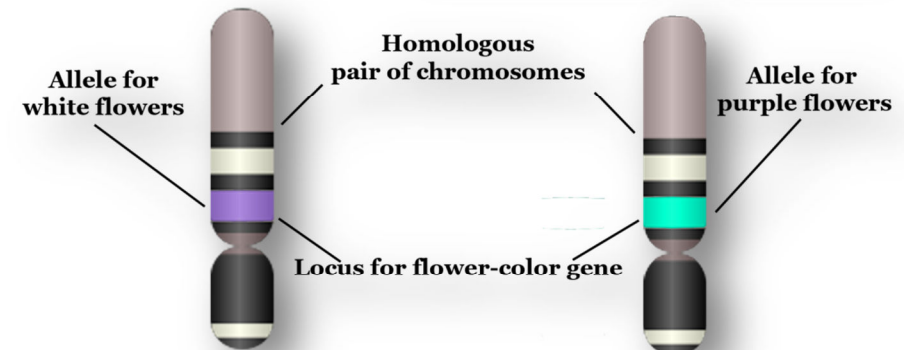
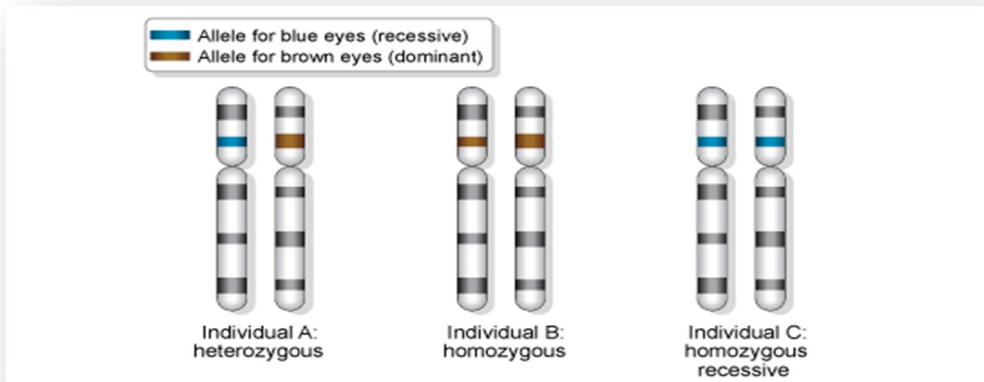
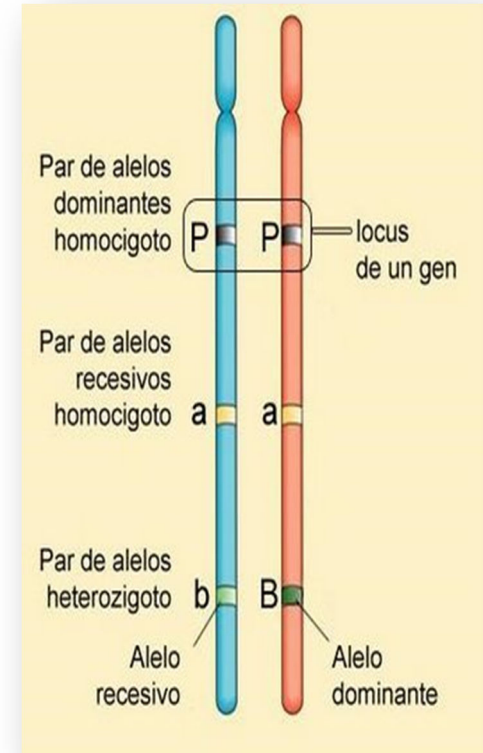


Alelo

Un alelo (del griego: allélon: uno a otro, unos a otras) es la localización espacial de un gen en un cromosoma. Al ser la mayoría de los mamíferos diploides se poseen dos alelos de cada gen. Cada par de alelos se ubica en igual locus o lugar.

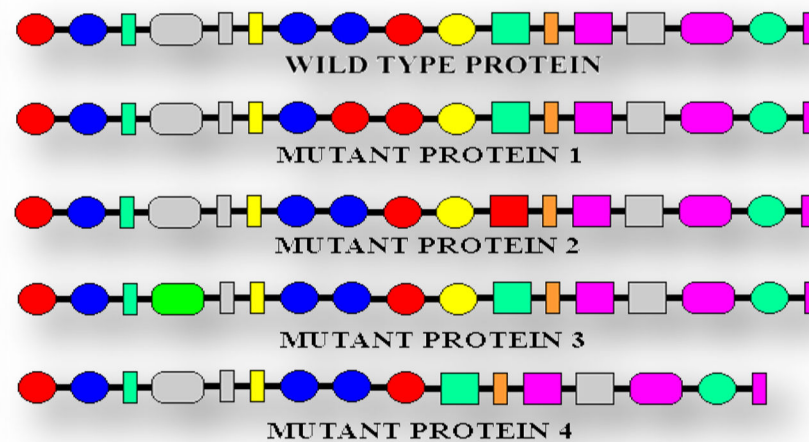
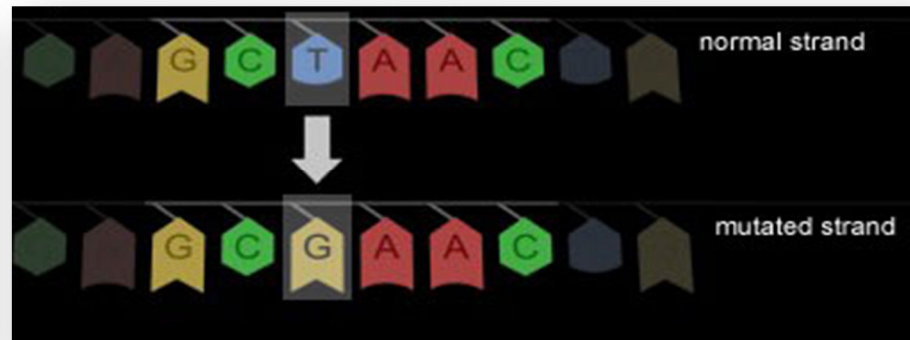
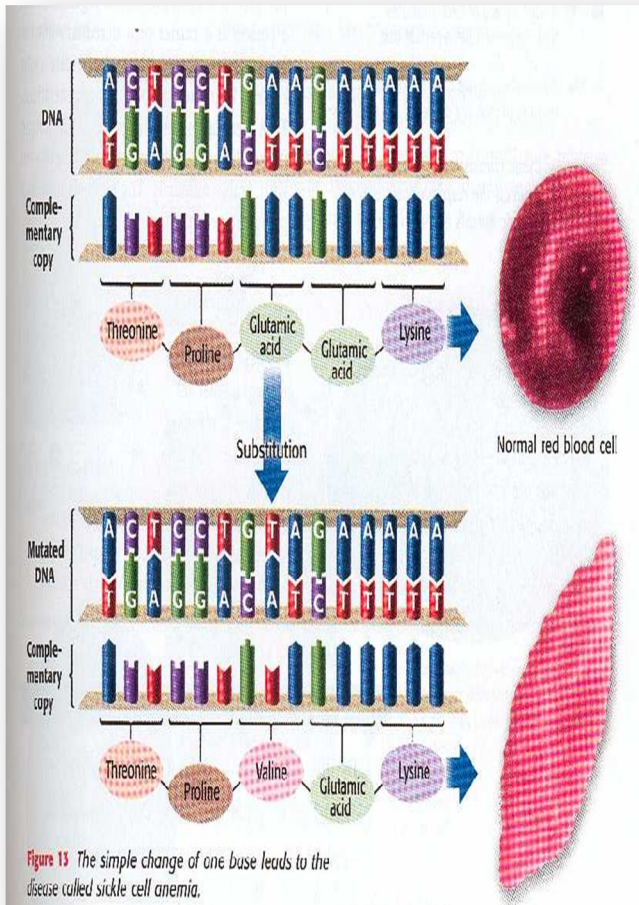
Locus

En biología un locus (del latín locus, lugar; plural loci) es una posición fija sobre un cromosoma, como la posición de un gen o de un marcador.

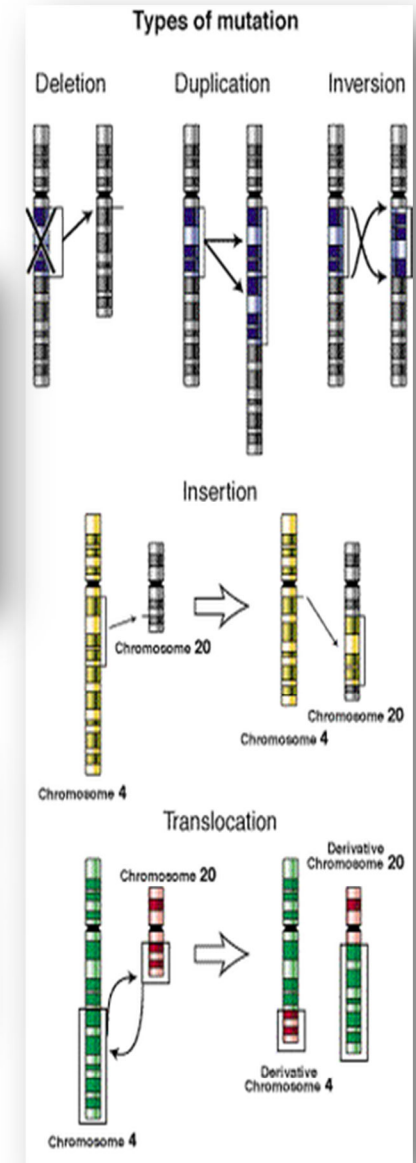


Mutación (del latín *mutare*=cambiar)

La mutación es un cambio permanente y transmisible en material genético (ADN, ARN) de una célula.



5 ALLELES



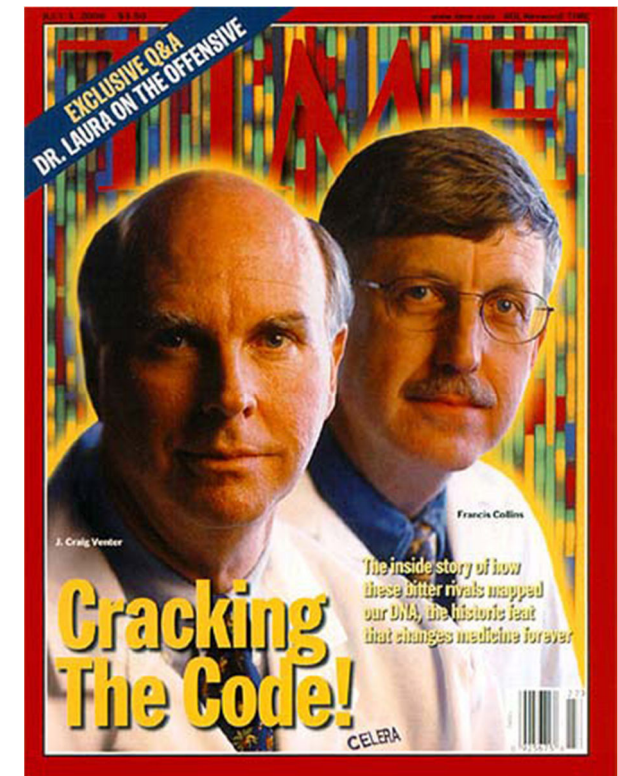
El cuerpo humano tiene **100 trillones** de células.

- En el núcleo hay ~ 2 m de ADN enrollados en una estructura de unos 0,0001 cm, ordenados en **46** cromosomas.
- Todo el ADN de estas células podría estirarse e ir y volver hasta el sol 600 veces (la secuencia llenaría 200 guías telefónicas de 500 páginas)

Genoma La totalidad del ADN de un organismo

<u>Fago λ</u>	5×10^4 pb
<u>Escherichia coli</u>	4×10^6 pb
<u>Levadura</u>	2×10^7 pb
<u>Caenorhabditis elegans</u>	8×10^7 pb
<u>Drosophila melanogaster</u>	2×10^8 pb
<u>Humano</u>	3×10^9 pb
Mitocondrial humano	1.6×10^4 pb

Genómica Conjunto de ciencias y técnicas dedicadas al estudio exhaustivo del funcionamiento, evolución y origen de los genomas.



Los estudios genómicos se caracterizan por su **interdisciplinaridad** debido a que el gran número de datos generados que requiere de conocimientos biológicos, estadísticos e informáticos.

Números del Genoma Humano

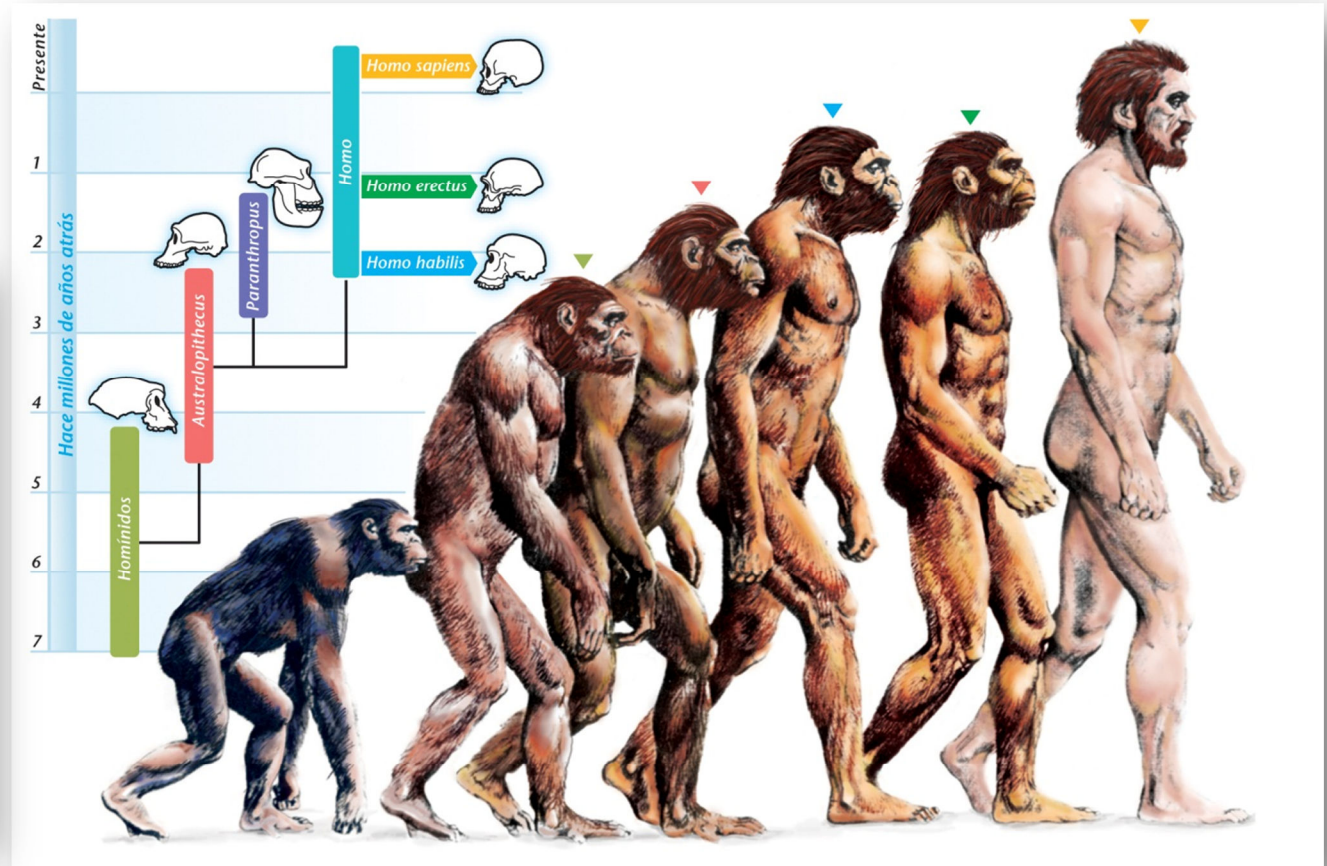
- Nuclear: 3.200 millones pb / Mitocondrial: 16.600 pb
- ~38.000 genes (el doble que la mosca del vinagre, un tercio más que el gusano común y 5.000 genes más que la planta Arabidopsis)
- 99,99% de código es compartido entre humanos (difieren en 1.250 letras)
- De 1.5 a 5 % del genoma codifica proteínas (se estima que existen ~300.000 proteínas). 25 % de genoma no codifica nada o se desconoce su función





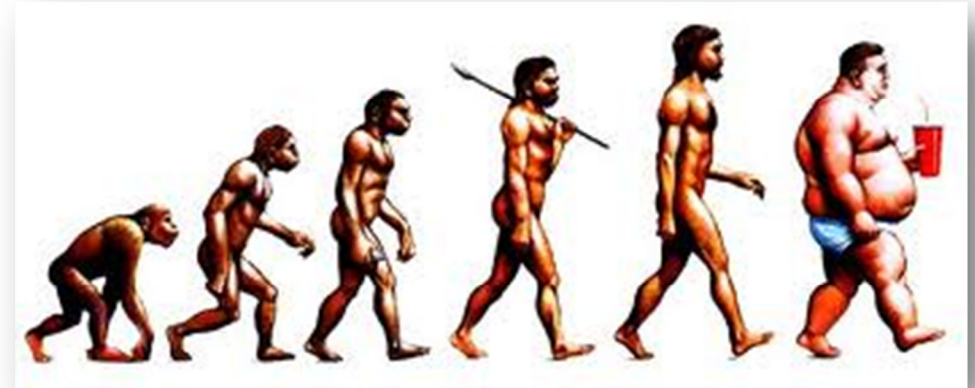
Evolución

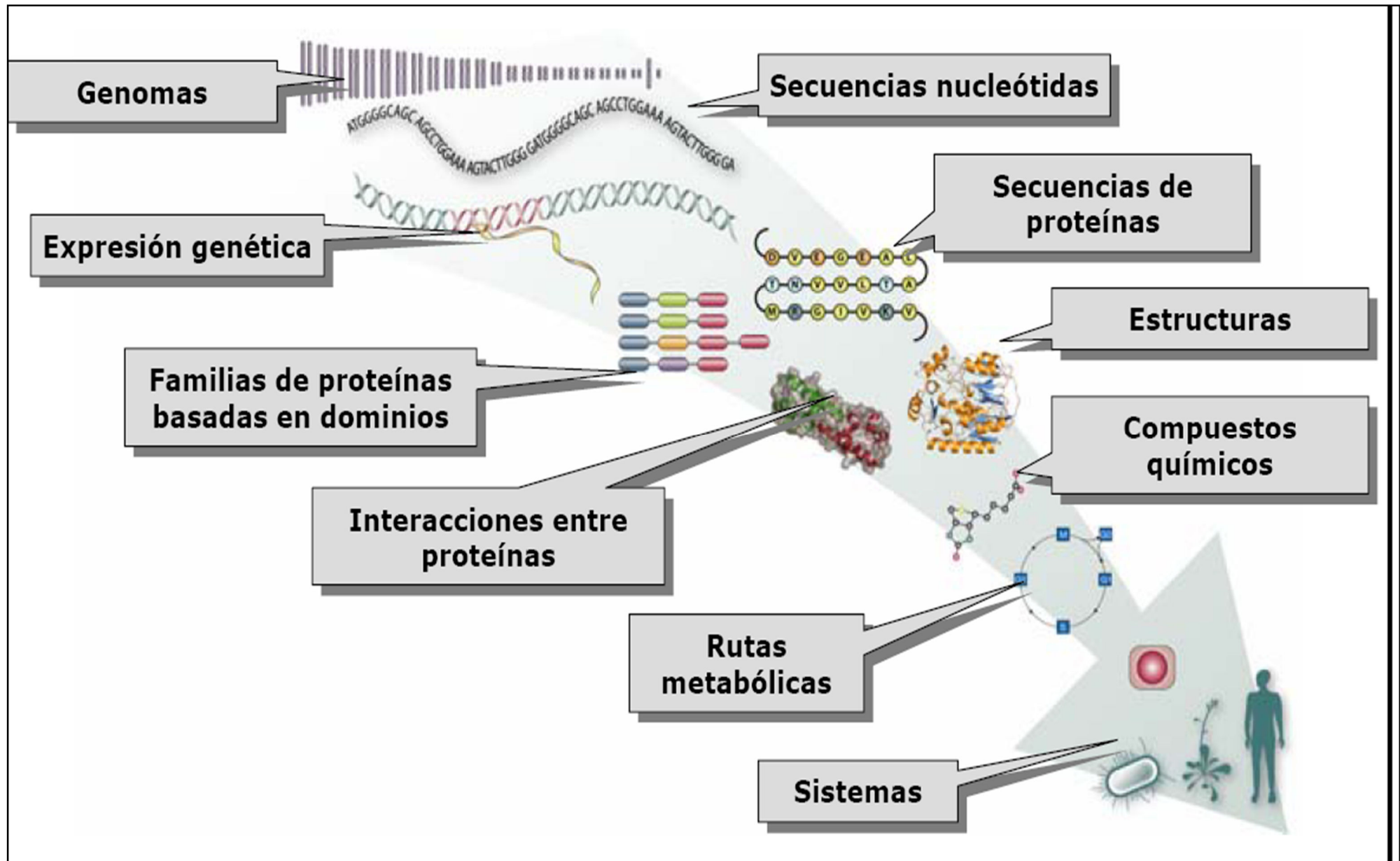
"La evolución es un cambio en la composición genética de las poblaciones. El estudio de los mecanismos evolutivos corresponde a la genética poblacional" (Dobzhansky, 1951).



Selección

La **selección natural** es un mecanismo evolutivo que se define como la reproducción diferencial de los genotipos en el seno de una población biológica.





Rodríguez-Santiago B. et al (2012)

