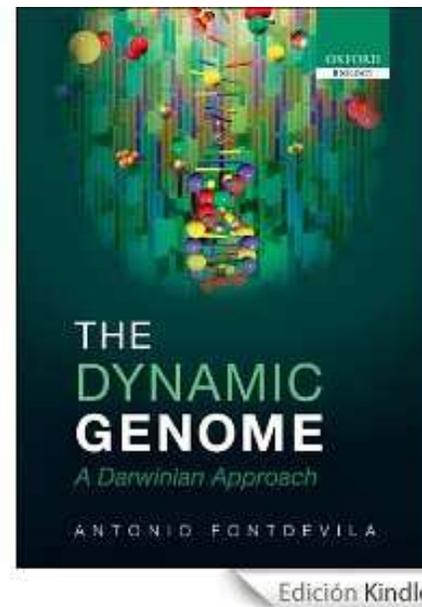
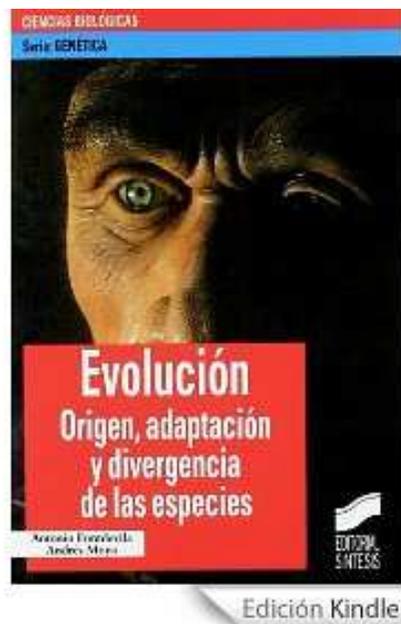


TEMA 3.
ESTRUCTURA Y EVOLUCIÓN DEL GENOMA
EUCARIOTA

- Genes y genomas.
- Estructura del genoma eucariota.
Genómica comparada.
- Evolución del tamaño, del número y de la
- complejidad de los genes.
- Origen y evolución de la complejidad genómica.

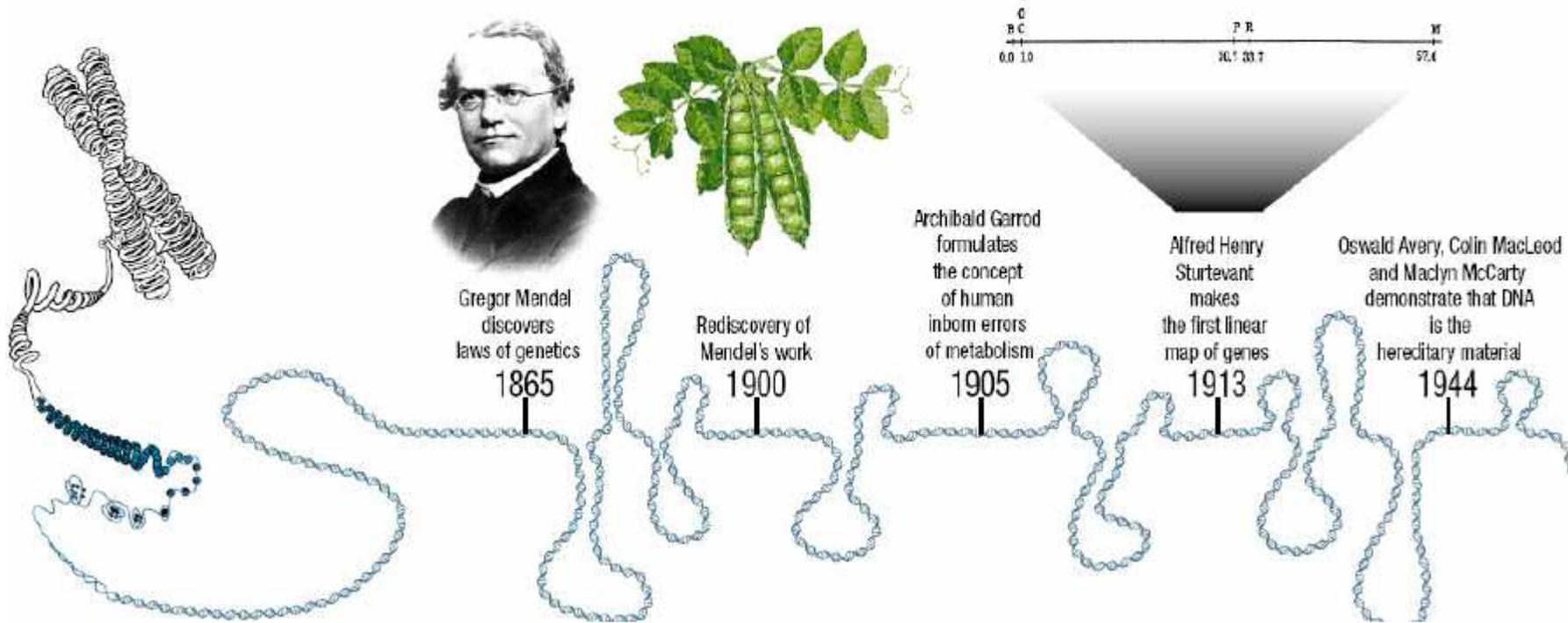
Bibliografía

- 1- Evolución. Origen, adaptación y divergencia de las especies. Antonio Fontdevila y Andrés Moya
- 2- The Dynamic Genome: A Darwinian Approach. Antonio Fontdevila
- 3- Fundamentals of Molecular Evolution
- 4- On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE
- Dan Graur , Yichen Zheng, Nicholas Price, Ricardo B. R. Azevedo , Rebecca A. Zufall, and Eran Elhaik



De los genes a los genomas...

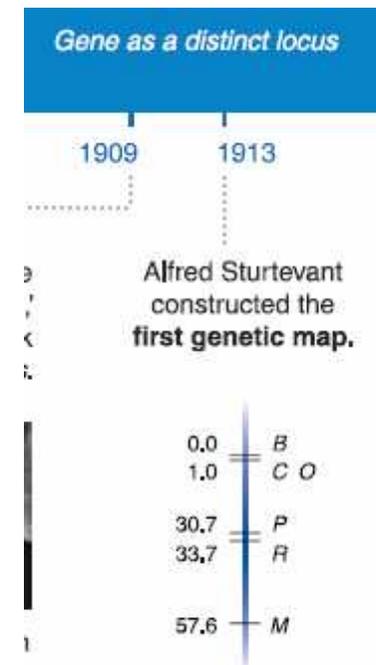
- Genes cómo concepto abstracto (Mendel, 1865)



Genes alineados en los cromosomas (Morgan 1910; Sturtevant 1913)
Genes en el ADN (Avery, 1940)

Gen como una unidad separada de la herencia

- Capacidad de determinar una característica particular de un organismo y la heredabilidad de esta característica
- El primer mapa genético fue creado en 1913
- El mecanismo de herencia mendeliana en 1915, define a un gen como un ente abstracto cuya existencia se refleja en la manera en que se transmiten a los fenotipos entre las generaciones.

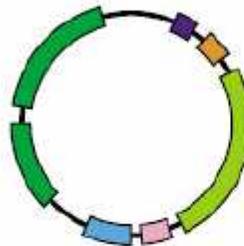


De los genes a los genomas...

- Estructura del ADN, Código Genético ... (1950-60s)



	U	C	A	G	
U	Phe Leu	Ser	Tyr stop stop	Cys Trp	U C A G
C	Leu	Pro	His Gln	Arg	U C A G
A	Ile Met	Thr	Asn Lys	Ser Arg	U C A G
G	Val	Ala	Asp Glu	Gly	U C A G



James Watson and Francis Crick describe the double-helical structure of DNA

1953

Marshall Nirenberg, Har Gobind Khorana and Robert Holley determine the genetic code

1966

Stanley Cohen and Herbert Boyer develop recombinant DNA technology

1972

The Belmont Report on the use of human subjects in research is issued

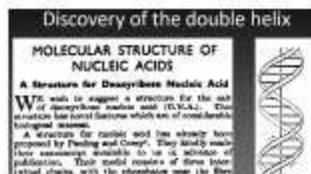
1974

Frederick Sanger, Allan Maxam and Walter Gilbert develop DNA-sequencing methods

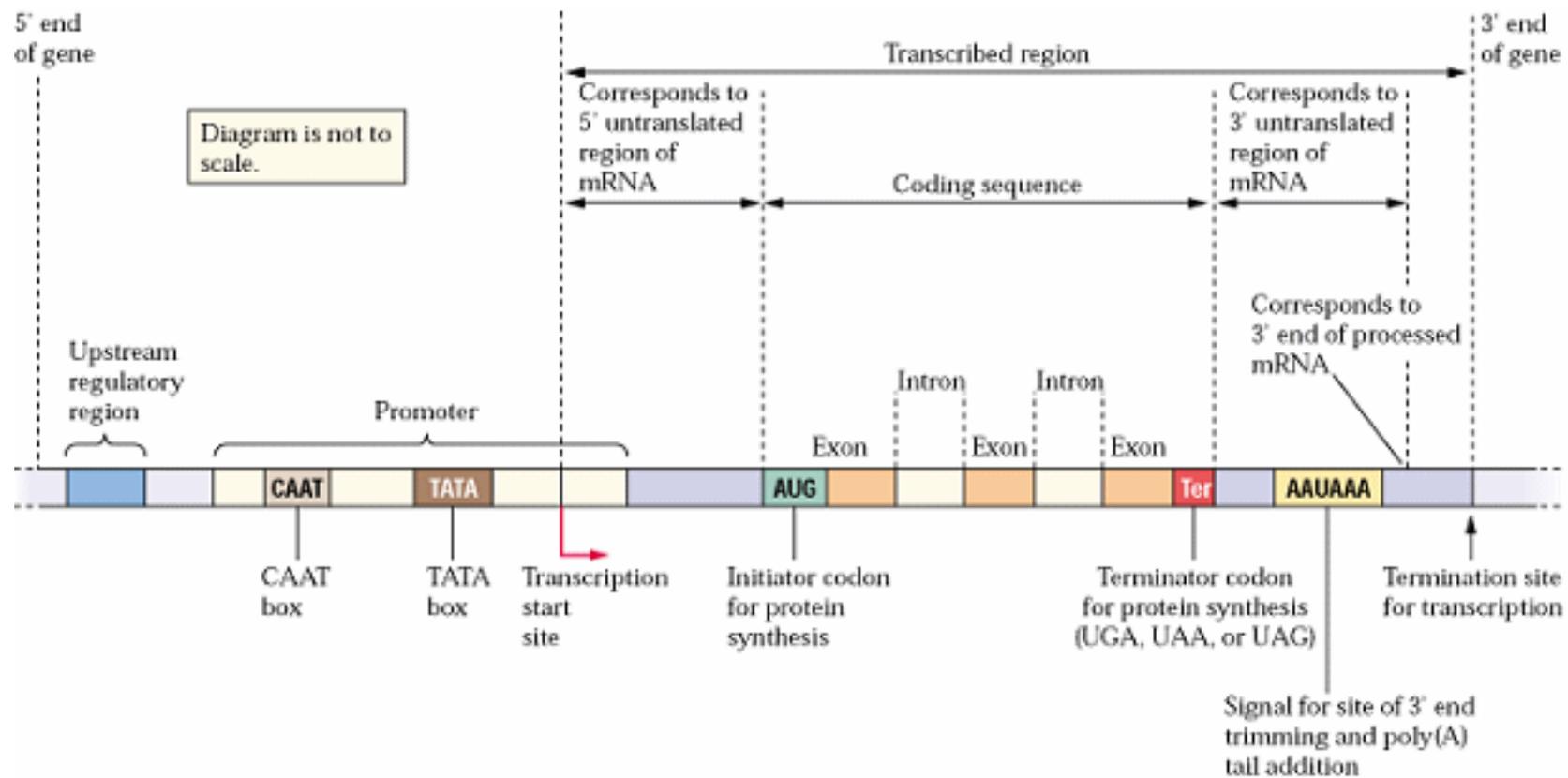
1977

GenBank database established

1982

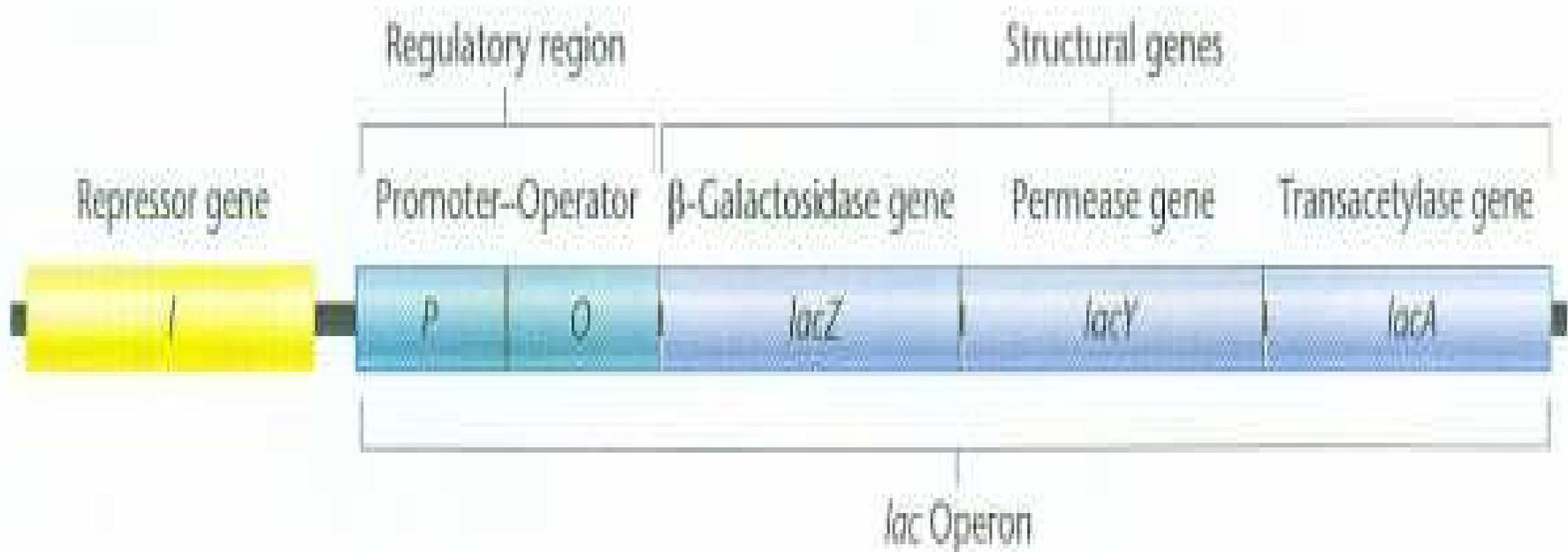


- Genes (secuencia, promotores, intrones... 1970-80s)

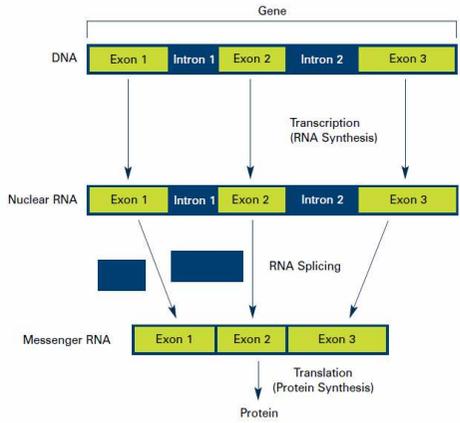


Estructura de un gen tipo

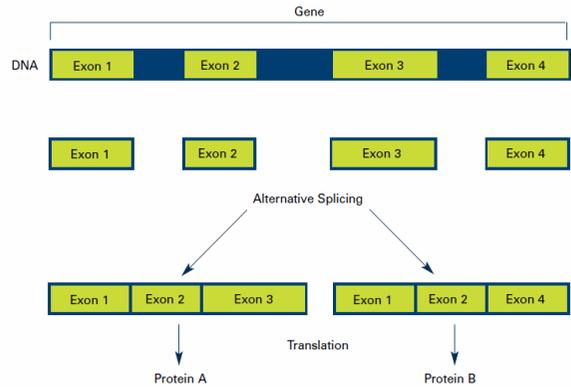
Consiste en una región de ADN formado por secuencias que codifican para una o más proteínas, un "promotor" de secuencia para la unión del ARN polimerasa, y un "operador" secuencia que se unen a genes reguladores



Splicing- alternativo



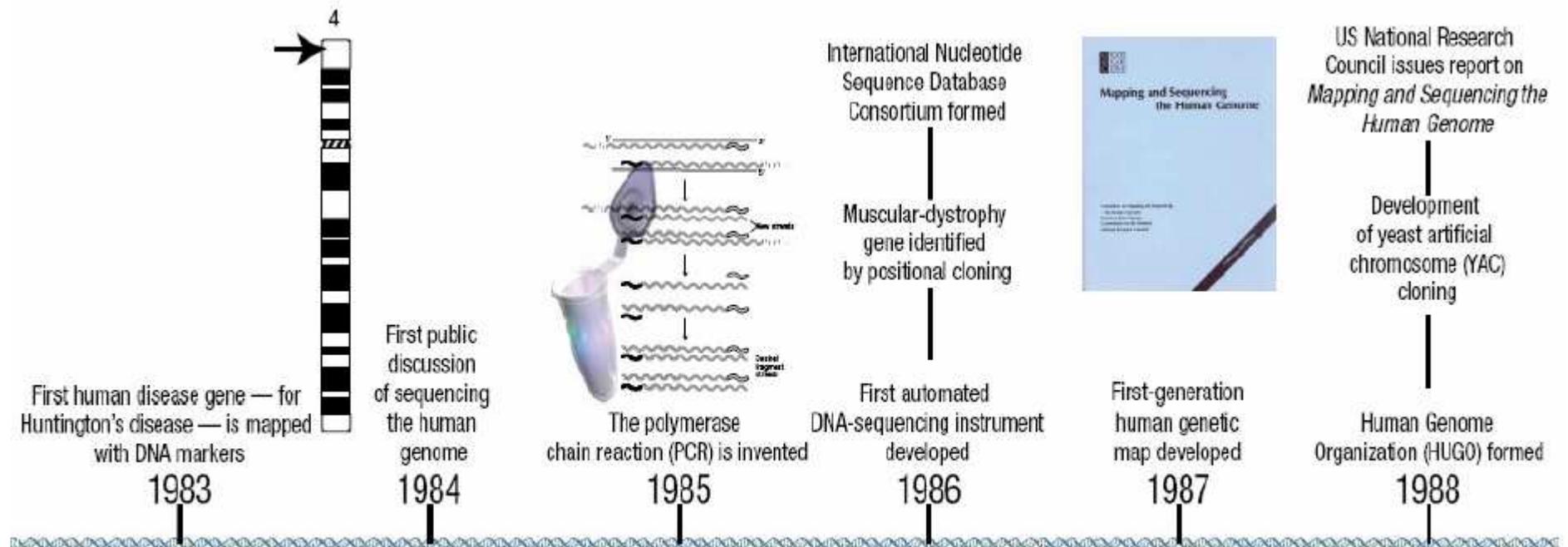
Genes are often interrupted by stretches of DNA (introns, blue) that do not contain instructions for making a protein. The DNA segments that do contain protein-making instructions are known as exons (green).



Arranging exons in different patterns, called alternative splicing, enables cells to make different proteins from a single gene.

De los genes a los genomas...

- PCR (Kary Mullis, 1985)



Secuenciación automática

PCR

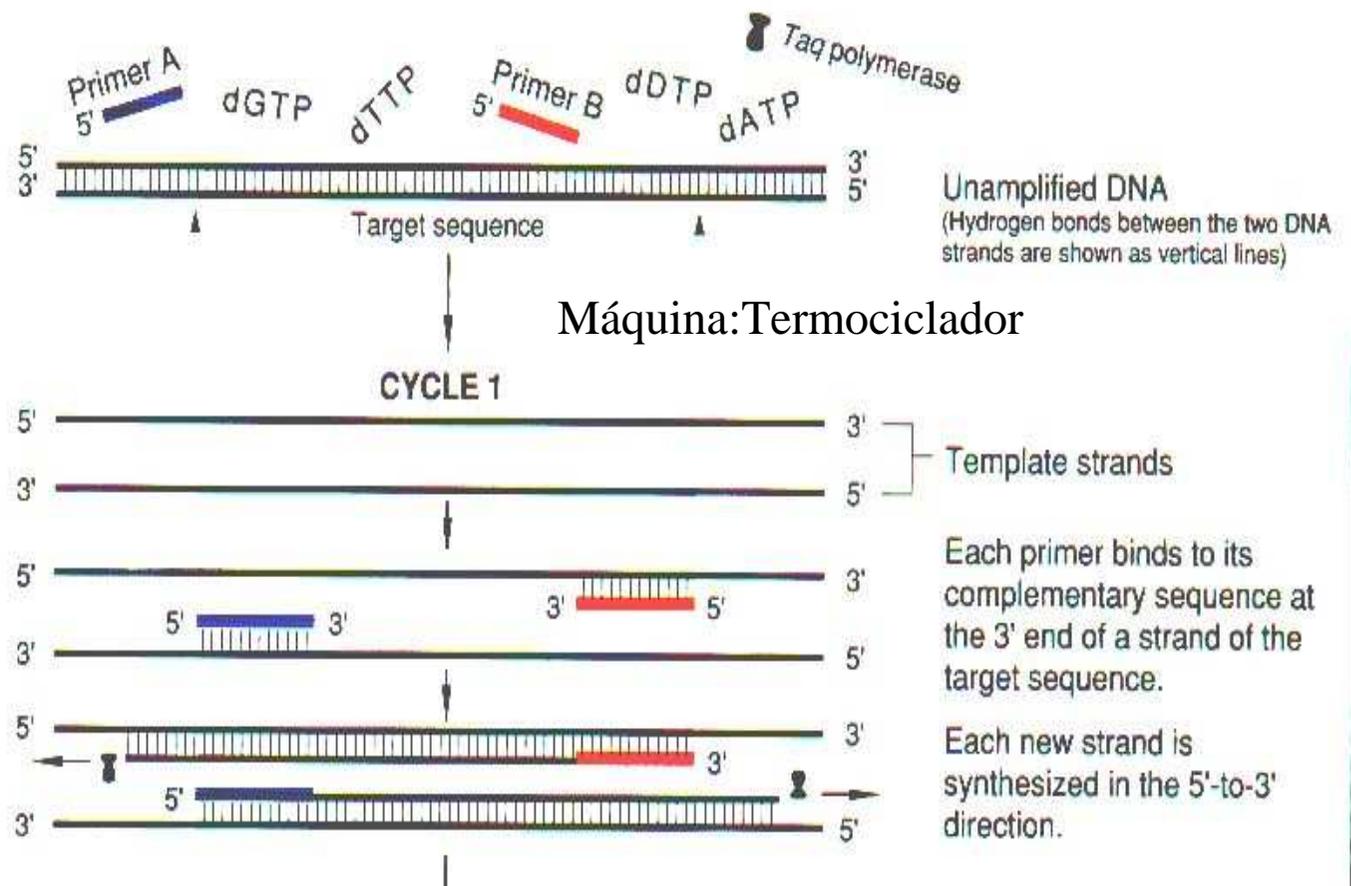
Figure 1. The Polymerase Chain Reaction

Reaction mixture includes DNA sample: two single-stranded primers, each with a 20-base sequence identical to the 5' end of one strand of the target sequence; heat stable *Taq* polymerase; and deoxyribonucleotide triphosphates (dNTPs).

Phase 1 92-95C denat.
Denature unamplified DNA at 95°C to form single-stranded templates.

Phase 2 50-65C hibrid.
Anneal primers to template at about 60°C.

Phase 3 72C polimeriz.
Synthesize new strands at 72°C.



PCR (continuación)

Phases 1 and 2

Denature products of Cycle 1 and anneal primers to template strands.

Phase 3

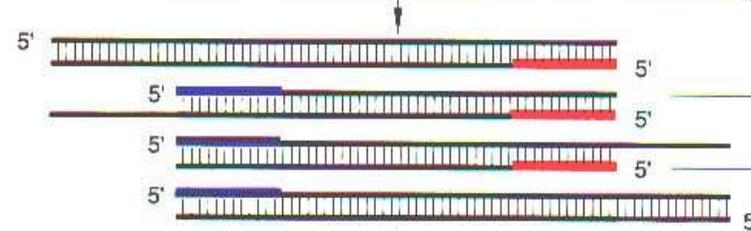
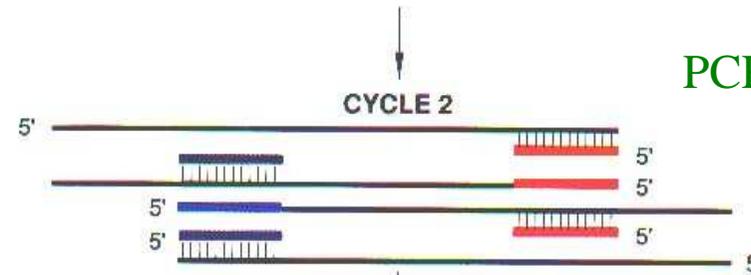
Synthesize new strands.

Phases 1 and 2

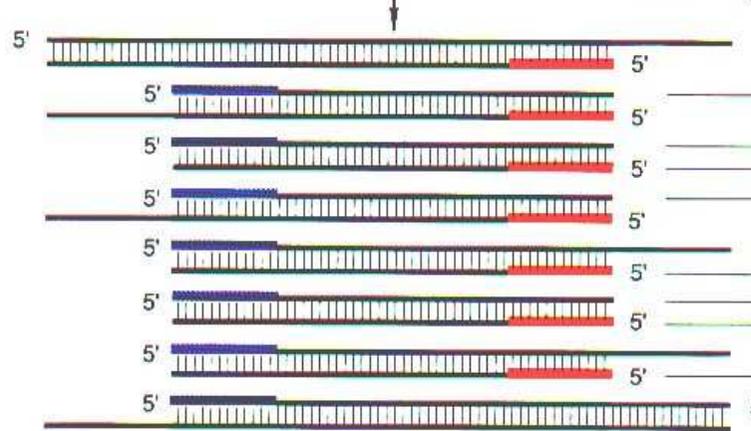
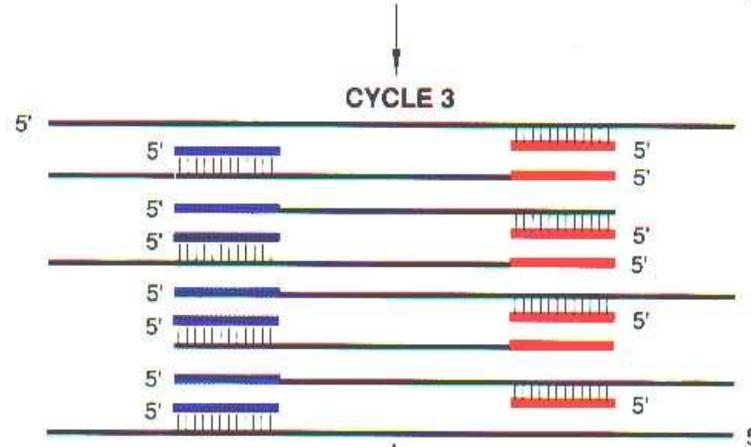
Denature products of Cycle 2 and anneal primers to template strands.

Phase 3

Synthesize new strands.



First two template strands that span the target sequence only



Eight template strands that span the target sequence only

1 ciclo 2^1

2 ciclos 2^2

3 ciclos $2^3 \dots$

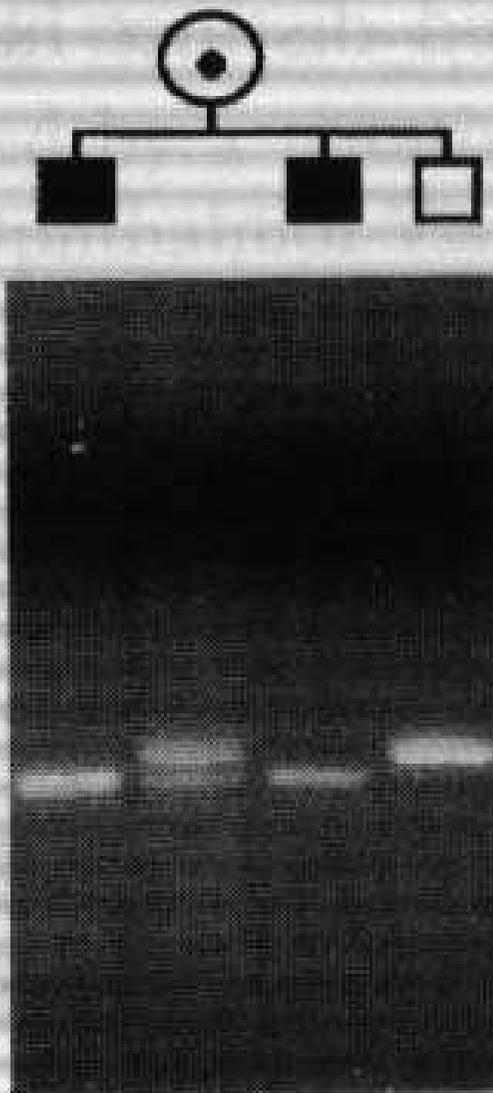
30 ciclos 2^{30}

PCR product of RPGR exon 11

B

bp

400 —
300 —
200 —
100 —



Normal

Mutation

19 bp deletion in family XLRP-236

Fujita et al., Am. J. Hum Genet (1997): 61:571-580

Un gen es una unión de secuencias genómicas capaces de codificar un conjunto coherente de productos funcionales con la capacidad de superponerse.

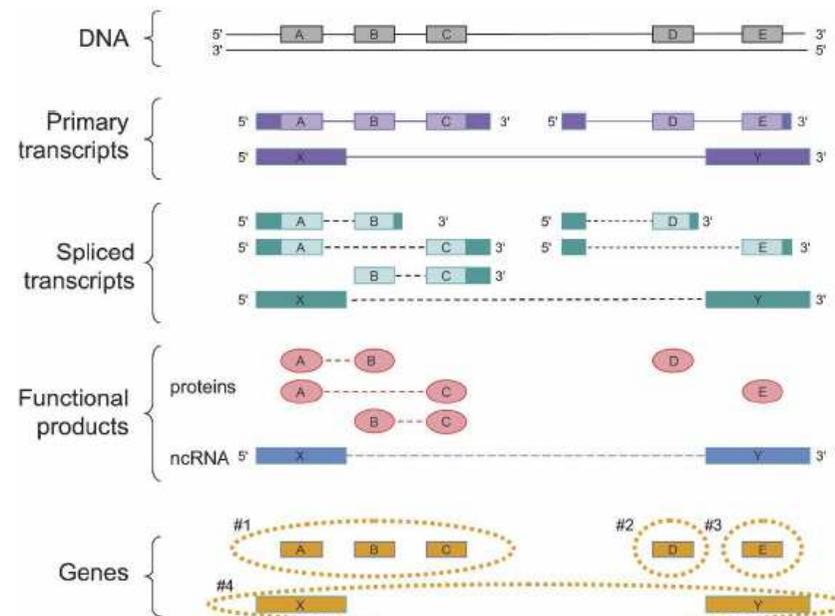
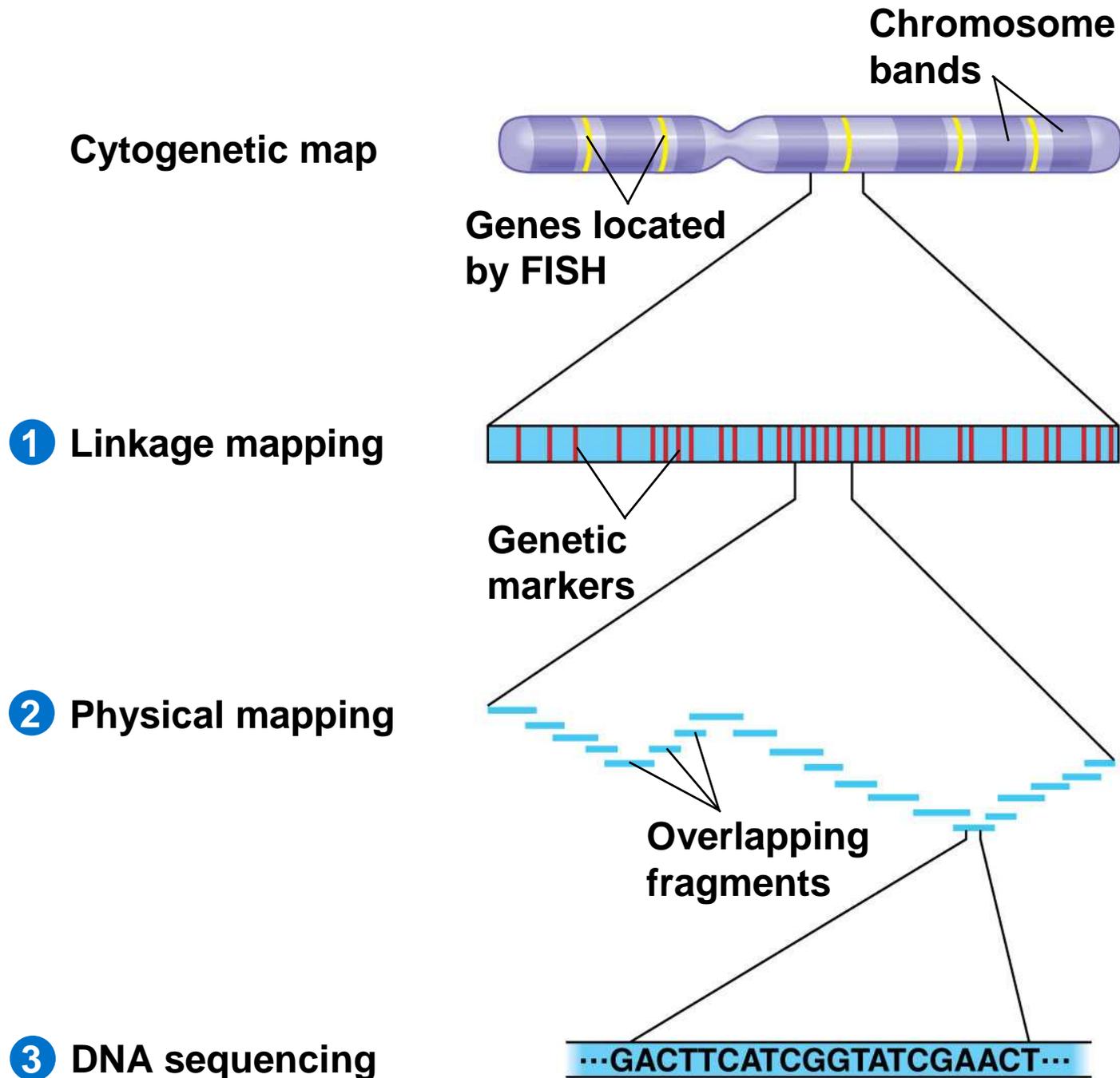
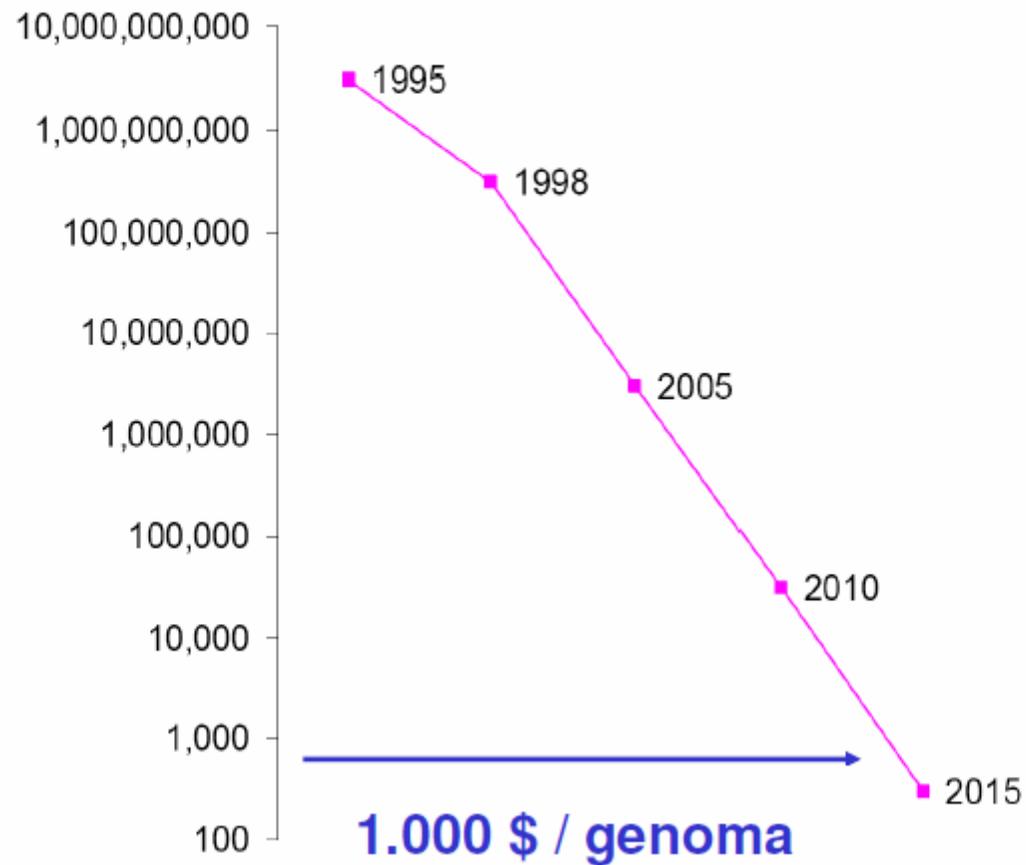


Figure 5. How the proposed definition of the gene can be applied to a sample case. A genomic region produces three primary transcripts. After alternative splicing, products of two of these encode five protein products, while the third encodes for a noncoding RNA (ncRNA) product. The protein products are encoded by three clusters of DNA sequence segments (A, B, and C; D; and E). In the case of the three-segment cluster (A, B, C), each DNA sequence segment is shared by at least two of the products. Two primary transcripts share a 5' untranslated region, but their translated regions D and E do not overlap. There is also one noncoding RNA product, and because its sequence is of RNA, not protein, the fact that it shares its genomic sequences (X and Y) with the protein-coding genomic segments A and E does not make it a co-product of these protein-coding genes. In summary, there are four genes in this region, and they are the sets of sequences shown inside the orange dashed lines: Gene 1 consists of the sequence segments A, B, and C; gene 2 consists of D; gene 3 of E; and gene 4 of X and Y. In the diagram, for clarity, the exonic and protein sequences A–E have been lined up vertically, so the dashed lines for the spliced transcripts and functional products indicate connectivity between the proteins sequences (ovals) and RNA sequences (boxes). (Solid boxes on transcripts) Untranslated sequences, (open boxes) translated sequences.

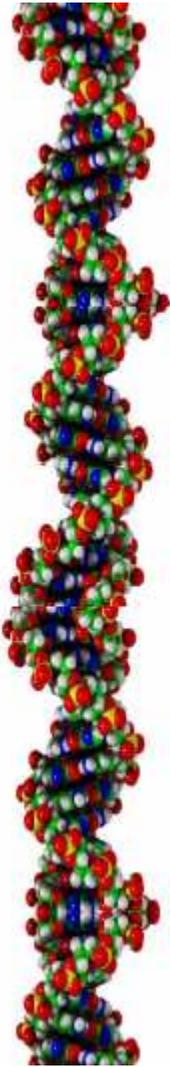


Coste de (re)secuenciar un genoma de mamífero



GENOMAS Y SUS TAMAÑOS

	Especie	Tamaño del genoma (Mb)	Número de genes
	<i>Saccharomyces cerevisiae</i>	12	5770
	<i>Caenorhabditis elegans</i>	97	20598
	<i>Drosophila melanogaster</i>	180	13525
	<i>Arabidopsis thaliana</i>	125	25498
	<i>Fugu rubripes</i>	365	30000
	<i>Mus musculus</i>	3300	30000
	<i>Homo sapiens</i>	3200	30000



Que es el Valor C (*C-value*) ?

Es el concepto que cuantifica el tamaño del genoma nuclear.

“*C-value*” → se refiere al contenido de DNA nuclear en células haploides

El origen del término surgió referente a “contenido” de DNA (C=)

“Contenido de la masa de DNA” en picogramos (10^{12} g) \neq Tamaño en bp

1 pg = 978 Mb (1 pg \approx 1 billón de bp)

Tamaño de genoma (bp) = Masa pg (contenido de DNA) \times $0,978 \times 10^9$

Contenido de DNA (pg) = Número de bp / $0,978 \times 10^9$



Valor C

Cantidad total de DNA por genoma haploide



Polychaos dubium may have the largest genome known for any organism, 670 Gb base pairs of DNA, over 200 times larger than the human genome.

¿Cómo se expresa la cantidad de ADN?

+ **En número de cromosomas**, designado por “n” en organismos haploides y “2n” en diploides donde hay dos copias de cada cromosoma.

+ **En masa de ADN** (pg=picogramos)

+ **Valor C** consiste en expresar la cantidad de ADN total en el genoma de una célula con respecto a la presente en una célula haploide de la misma especie. En células diploides puede ser de 2C o 4C dependiendo del estadio del ciclo celular.

+ **En longitud** : pb= pares de bases en hebra doble. Para moléculas largas se emplea kilobases o mega bases (Mb)

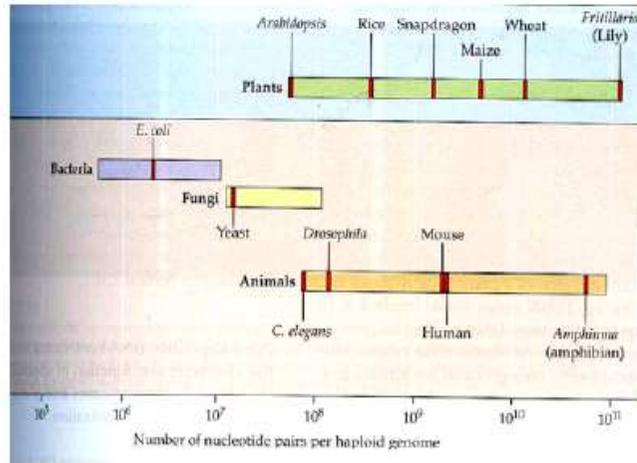
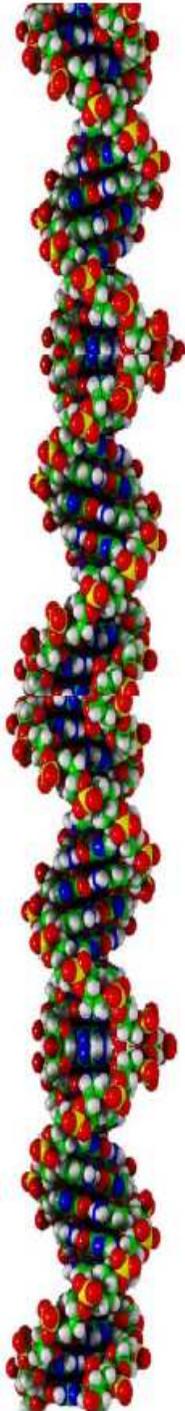
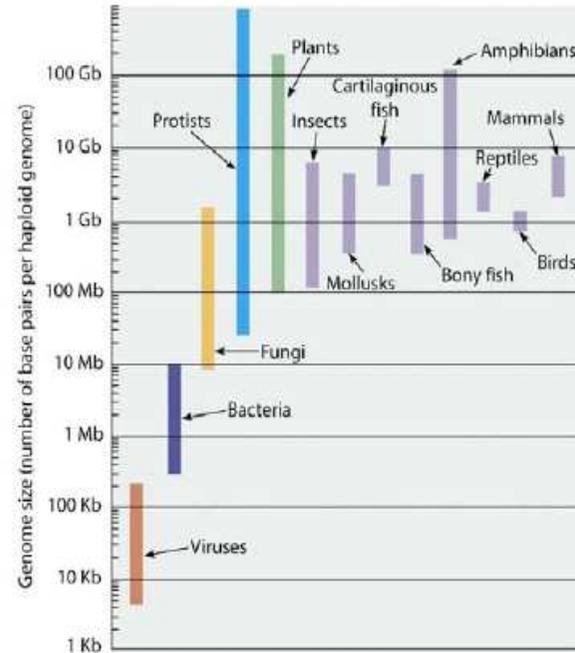


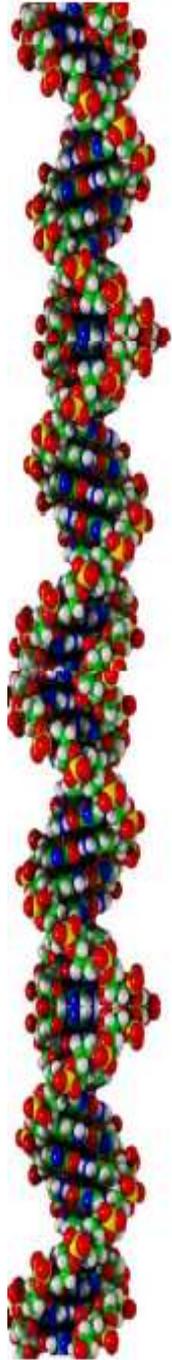
Table 7.2 Sizes of eukaryotic genomes

Species	Genome size (Mb)
Fungi	
<i>Saccharomyces cerevisiae</i>	12.1
<i>Aspergillus nidulans</i>	25.4
Protozoa	
<i>Tetrahymena pyriformis</i>	190
Invertebrates	
<i>Caenorhabditis elegans</i>	97
<i>Drosophila melanogaster</i>	180
<i>Bombyx mori</i> (silkworm)	490
<i>Strongylocentrotus purpuratus</i> (sea urchin)	845
<i>Locusta migratoria</i> (locust)	5000
Vertebrates	
<i>Takifugu rubripes</i> (pufferfish)	400
<i>Homo sapiens</i>	3200
<i>Mus musculus</i> (mouse)	3300
Plants	
<i>Arabidopsis thaliana</i> (vetch)	125
<i>Oryza sativa</i> (rice)	466
<i>Zea mays</i> (maize)	2500
<i>Pisum sativum</i> (pea)	4800
<i>Triticum aestivum</i> (wheat)	16,000
<i>Fritillaria assyiaca</i> (lily)	120,000



Copyright © 2009 Pearson Education, Inc.





Esta falta de correlación entre la complejidad de un organismo y el tamaño de su genoma, es la llamada → **Paradoja del Valor C.**

Arabidopsis thaliana = 125 Mb



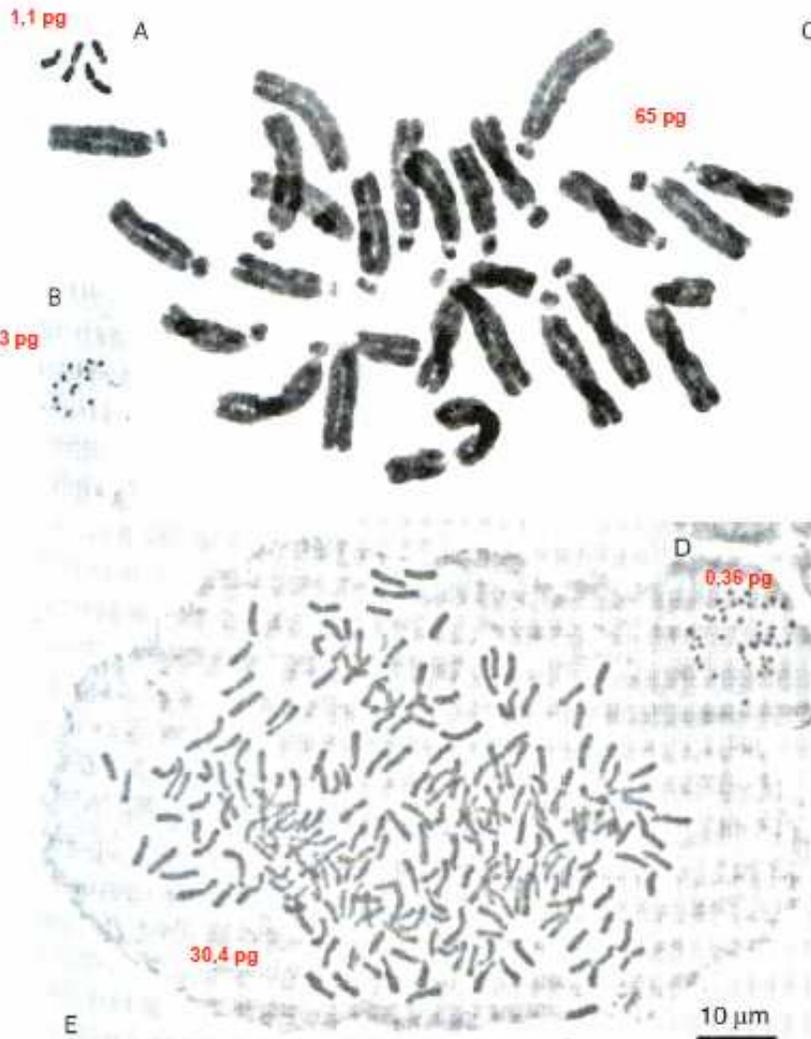
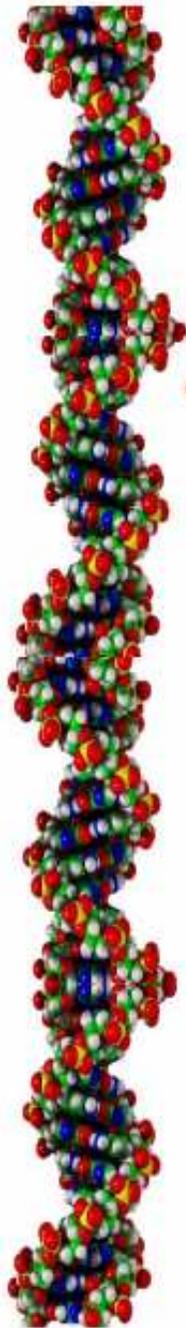
Fritillaria assyriaca = 120000 Mb



Cómo un organismo “simple” tiene más DNA que uno “complejo”..?

Poliploidía en plantas (contenido cromosómico)

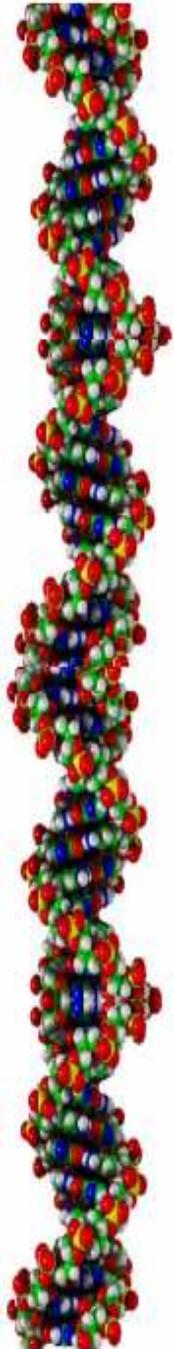




- C
- A. *Brachyscome dichromosomatica* (Asteraceae) ($2x=4$, $1C=1,1$ pg)
 - B. *Myriophyllum spicatum* (Planta acuática de agua dulce) ($2x=14$, $1C=0,3$ pg)
 - C. *Fritillaria sp.* (Liliaceae) ($2x=24$, $1C=65$ pg)
 - D. *Selaginella kraussiana* (Selaginellaceae, planta tropical) ($4x=40$, $1C=0,36$ pg)
 - E. *Equisetum sp.* (arbusto perenne con tallo rizomatoso) ($2x=216$, $1C=30,4$ pg)

Variación del contenido de DNA independiente de los niveles de ploidía.





Arabidopsis thaliana = 125 Mb



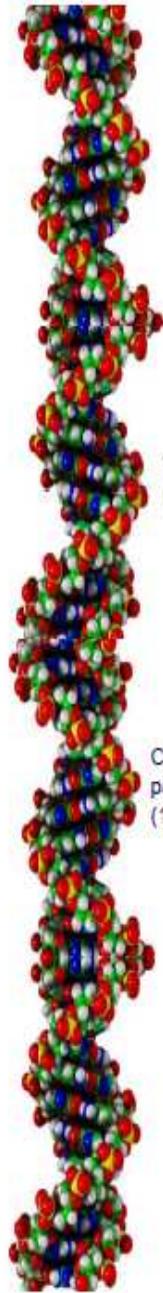
Fritillaria assyriaca = 120000 Mb



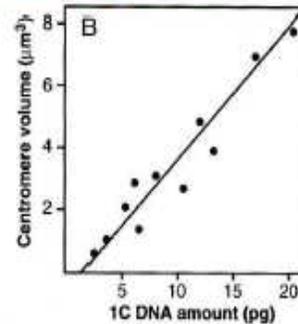
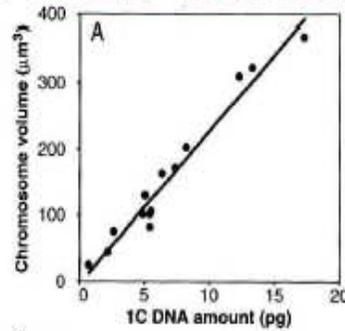
3. Impactos Fenotípicos paralelos a la variación del tamaño del genoma en plantas

→ Tamaño Celular (Cromosomas/Núcleo/célula)

→ **Tamaño cromosómico v nuclear**

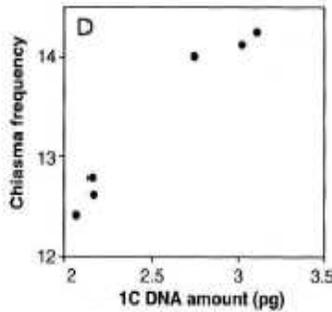
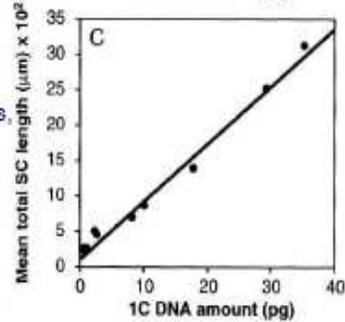


Volumen cromosoma
(14 angiospermas)



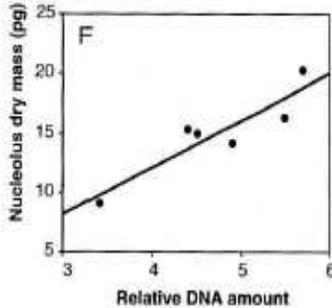
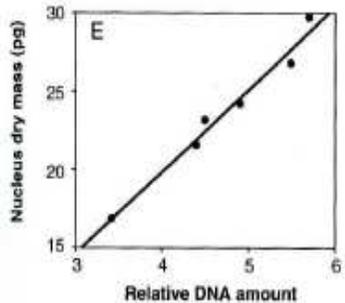
Volumen centrómeros
(11 angiospermas)

Complejos sinaptonémicos,
paquiteno, meiosis
(10 angiospermas)



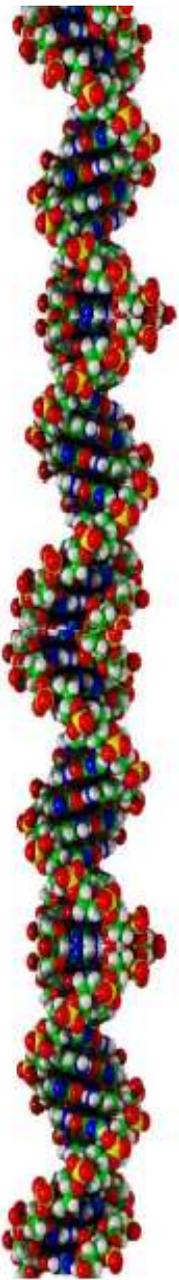
Quiasmas, meiosis
(6 angiospermas)

Masa seca nuclear
(6 angiospermas)

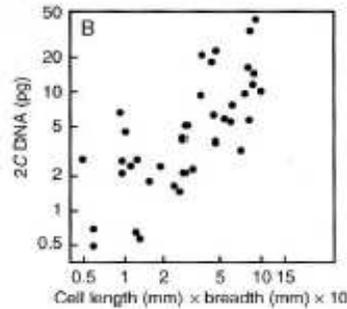


Masa seca nucleolos
(6 angiospermas)



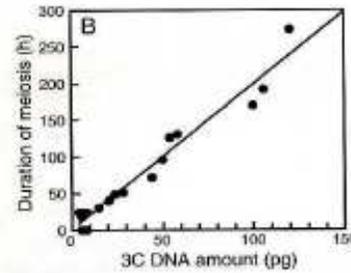
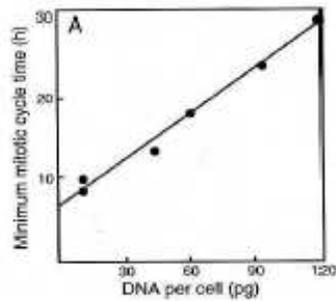


→ **Tamaño Celular**



Dimensión células epidérmicas de hojas (37 angiospermas)

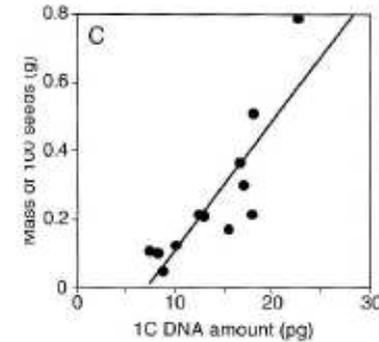
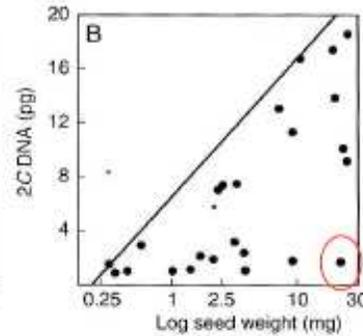
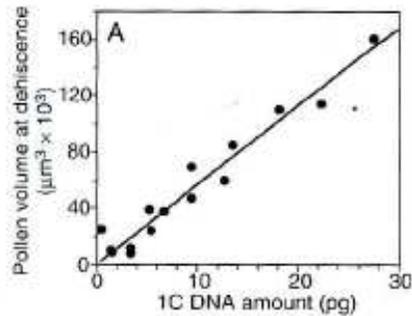
→ **Tasa División Celular (duración)**



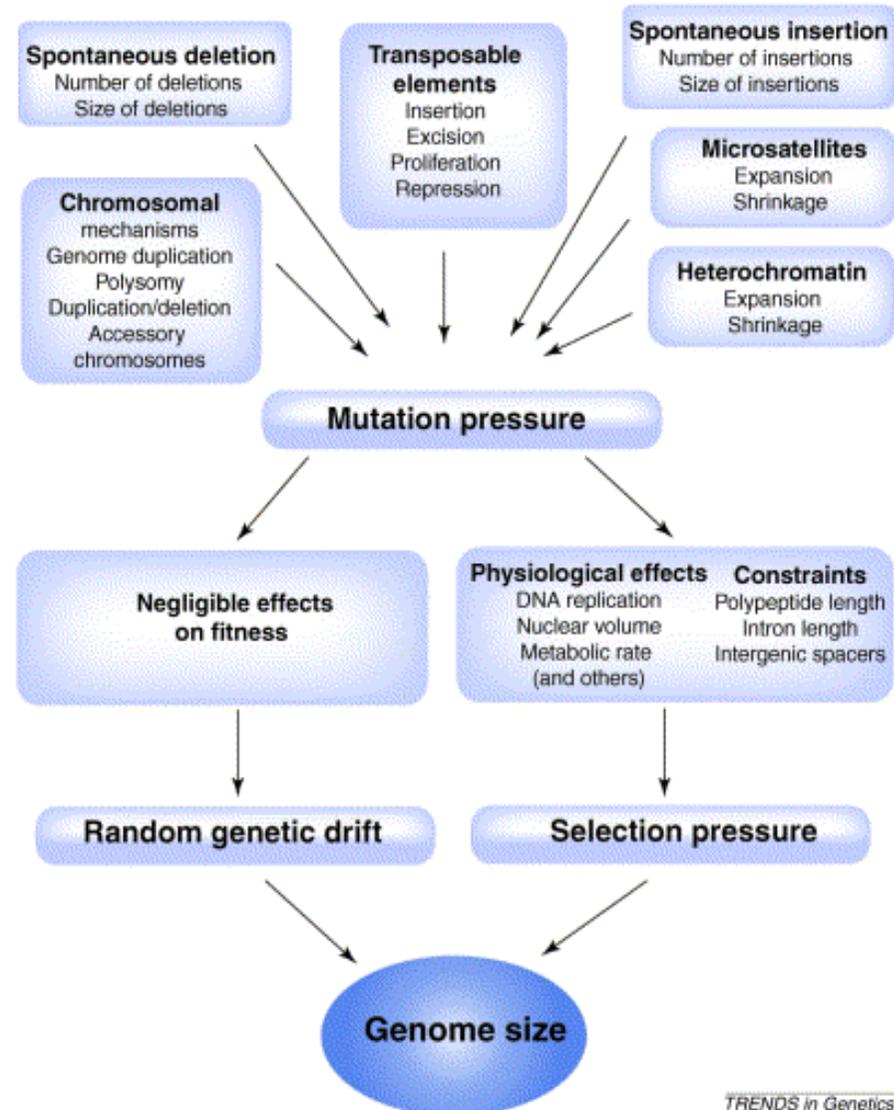
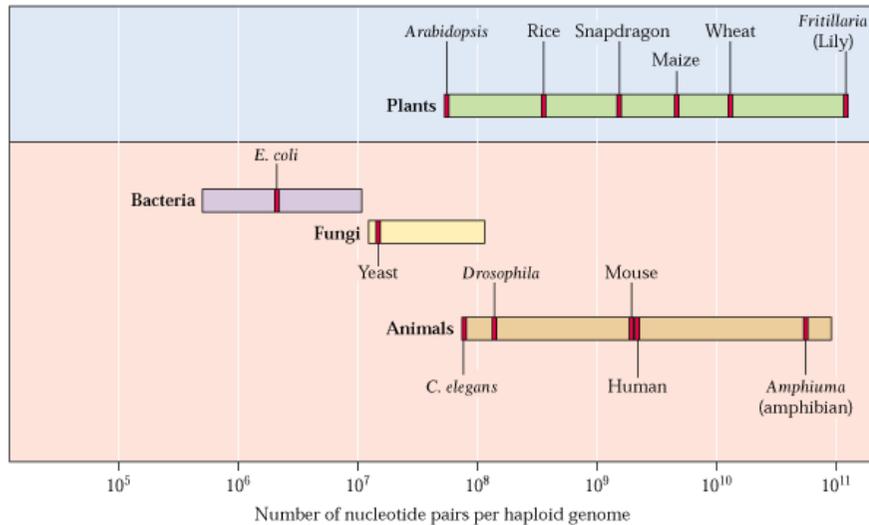
Genomas largos retrasan el proceso de división = células grandes

→ **Polen y semillas**

Genomas pequeños = semillas pequeñas/grandes
Genomas grandes ≠ semillas pequeñas



La paradoja del valor C, o no es una paradoja?



TODAS ESTAS RELACIONES PERMITEN SUPONER QUE

- .- **El tamaño del genoma tiene significado adaptativo**
- .- **El contenido de ADN influencia el fenotipo en dos Formas**
- **A) Por la expresión de su contenido genico**
- **B) Por el efecto fisico de su masa y volumen (nucleotipo)**

Los genomas varían en tamaño, número de genes, y densidad genética

- El genoma de las bacterias y archaeas varían entre 1 a 6 millones de bp. (Mb); mientras los genomas de los eucariotas son muchos más grandes
- El número de genes no está correlacionado con el tamaño del genomas
- Los seres humanos y otros mamíferos tienen una densidad más baja de genes en una longitud dada de ADN

Número de genes y su densidad

- Bacterias de vida libre y arqueas tienen entre 1.500 y 7.500 genes
Hongos unicelulares tienen de aproximadamente 5.000 genes y eucariotas multicelulares hasta al menos 40.000 genes

Table 21.1 Genome Sizes and Estimated Numbers of Genes*

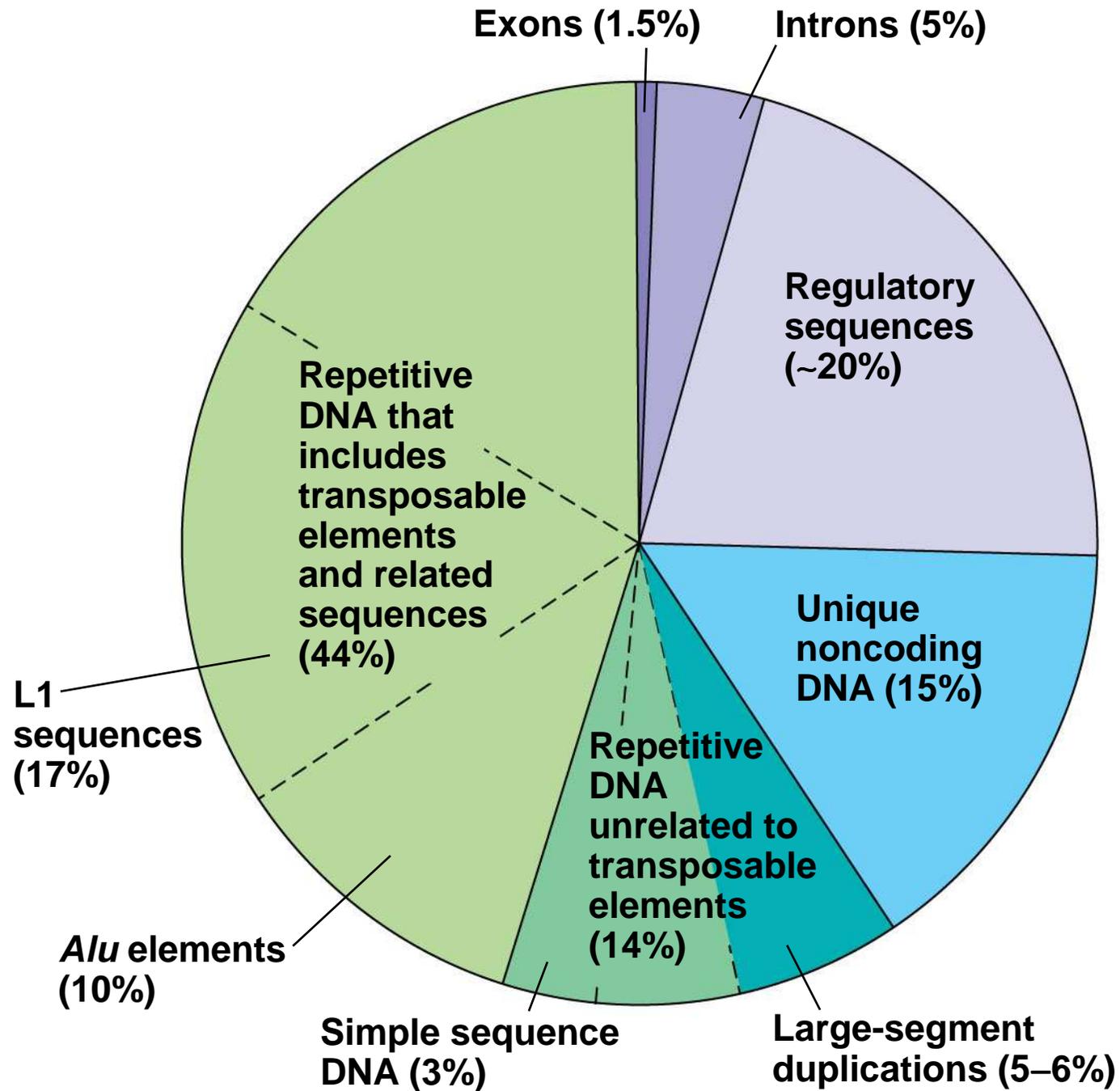
Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Bacteria			
<i>Haemophilus influenzae</i>	1.8	1,700	940
<i>Escherichia coli</i>	4.6	4,400	950
Archaea			
<i>Archaeoglobus fulgidus</i>	2.2	2,500	1,130
<i>Methanosarcina barkeri</i>	4.8	3,600	750
Eukaryotes			
<i>Saccharomyces cerevisiae</i> (yeast, a fungus)	12	6,300	525
<i>Caenorhabditis elegans</i> (nematode)	100	20,100	200
<i>Arabidopsis thaliana</i> (mustard family plant)	120	27,000	225
<i>Drosophila melanogaster</i> (fruit fly)	165	13,700	83
<i>Oryza sativa</i> (rice)	430	42,000	98
<i>Zea mays</i> (corn)	2,300	32,000	14
<i>Mus musculus</i> (house mouse)	2,600	22,000	11
<i>Ailuropoda melanoleuca</i> (giant panda)	2,400	21,000	9
<i>Homo sapiens</i> (human)	3,000	<21,000	7
<i>Fritillaria assyriaca</i> (lily family plant)	124,000	ND	ND

*Some values given here are likely to be revised as genome analysis continues. Mb = million base pairs. ND = not determined.



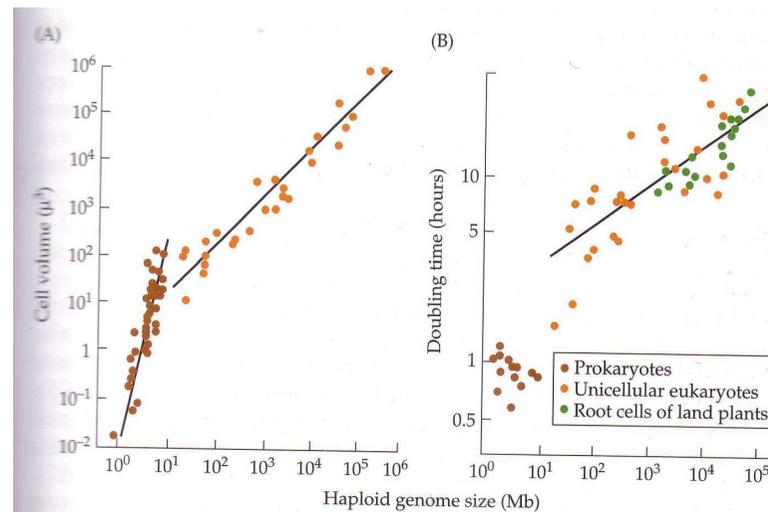
Los eucariotas tienen gran cantidad de ADN no codificante y muchas familias multigénicas

- La mayor parte de los genomas eucariotas no codifica proteínas ni para ARN funcionales
- Hay mucha evidencia que indica que el ADN no codificante (anteriormente llamado "ADN basura") desempeña un papel importante en la célula
- Por ejemplo, los genomas de seres humanos, ratas y ratones muestran alta conservación de unas 500 regiones no codificantes



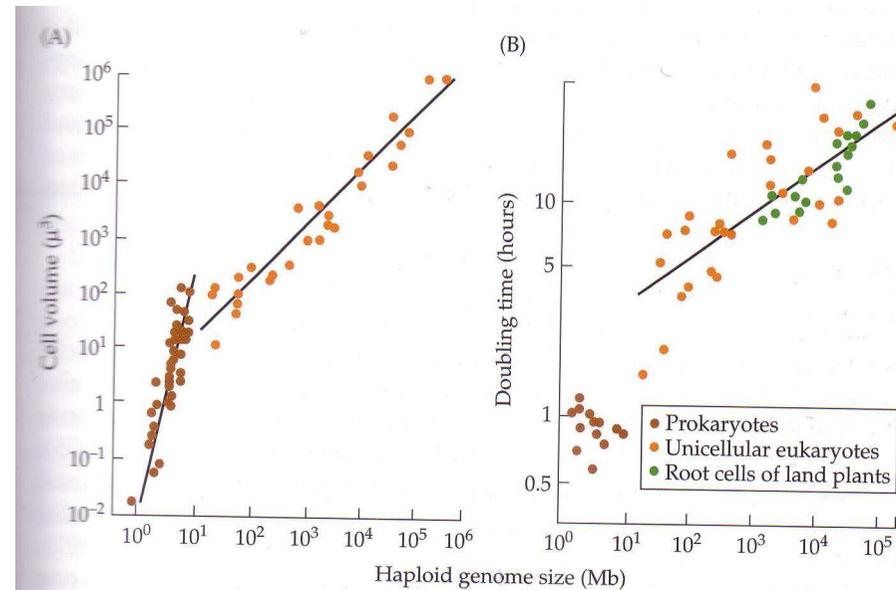
¿Por qué Genomas grandes ?

- **ADN egoísta**
- Los genomas más grandes son el resultado de la proliferación de ADN egoísta
- La proliferación se detiene sólo cuando se está volviendo demasiado perjudicial
- **ADN funcional**
- El contenido del genoma es una consecuencia de la selección natural . Se necesitan genomas más grandes para permitir el tamaño de más núcleo más grande, membrana nuclear más grande , etc

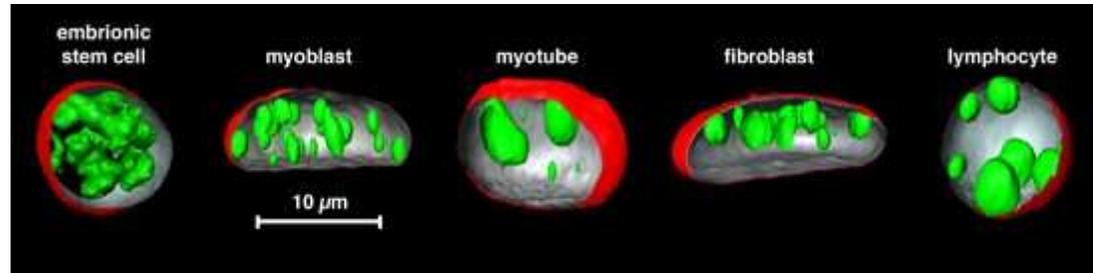


¿ Por qué Genomas pequeños?

- **Costo metabólico:** tal vez las células pierden el exceso de ADN para mejorar la eficiencia energética
 - But DNA is only 2-5% of the dry mass
 - No genome size – replication time correlation in prokaryotes
 - Replication is much faster than transcription (10-20 times in *E. coli*)



Características estructurales del Genoma : organización nuclear



The nucleus must be organized to allow functional transcription and replication

Incredibly dense mesh of chromosomes, cytoskeleton, membranes

Transcription factories / chromosomal territories

“spacer DNA” may affect physical organization in unexpected ways

Inter- and Intra- chromosomal interactions

Entire genome may participate in regulating interactions

