Estructura de genes y genomas

Tipos de secuencias en el genoma humano

TABLE 9-1 Classification of Eukaryotic DNA

Protein-coding genes
- Solitary genes
- Duplicated and diverged genes (functional gene families and nonfunctional pseudogenes)

Tandemly repeated genes encoding rRNA, 5S rRNA, tRNA, and histones

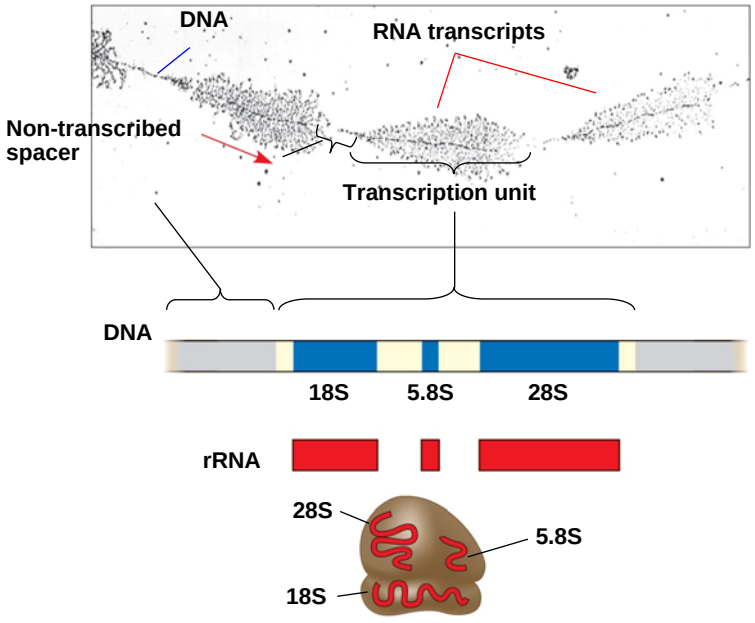
Repetitious DNA
- Simple-sequence DNA
- Moderately repeated DNA (mobile DNA elements)
  - Transposons
  - Viral retrotransposons
  - Long interspersed elements (LINES; nonviral retrotransposons)
  - Short interspersed elements (SINES; nonviral retrotransposons)

Unclassified spacer DNA

Lodish

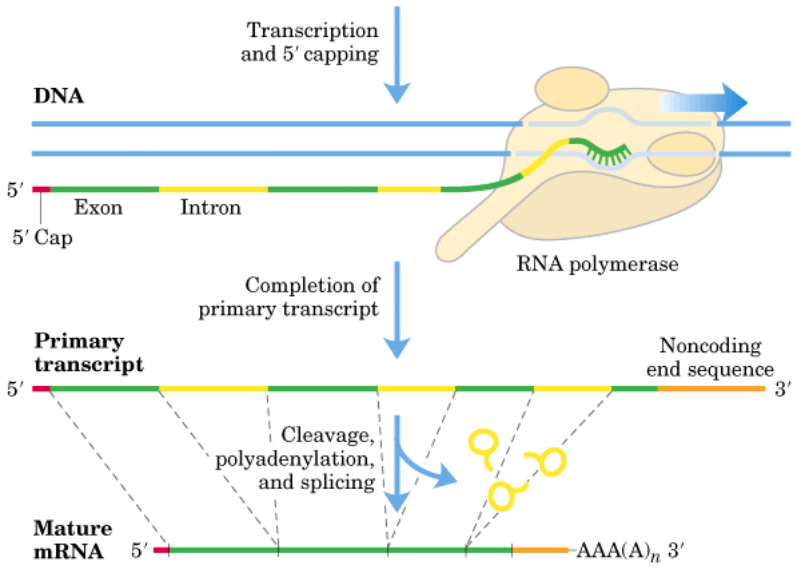
LE 19-17a

Part of the ribosomal RNA gene family

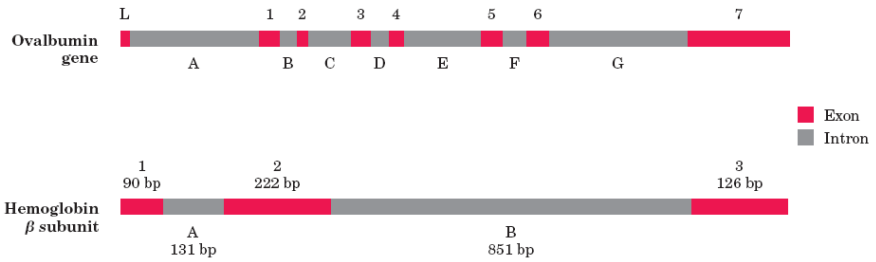


5S rRNA gene: different transcriptional unit (not shown)

transcripción y procesamiento de mRNA

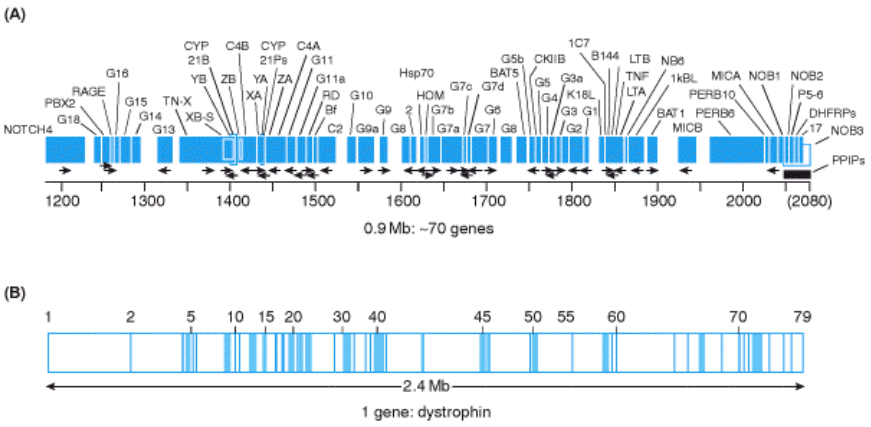


Exones e intrones



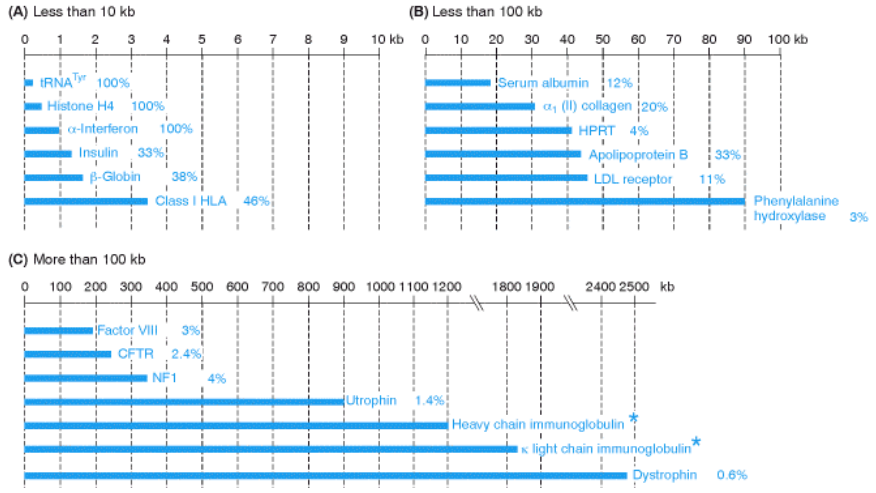
Exones e intrones

gran parte de la secuencia del DNA no es codificante



Exones e intrones

diferentes genes humanos varían mucho en tamaño y proporción de secuencias exónicas



Exones

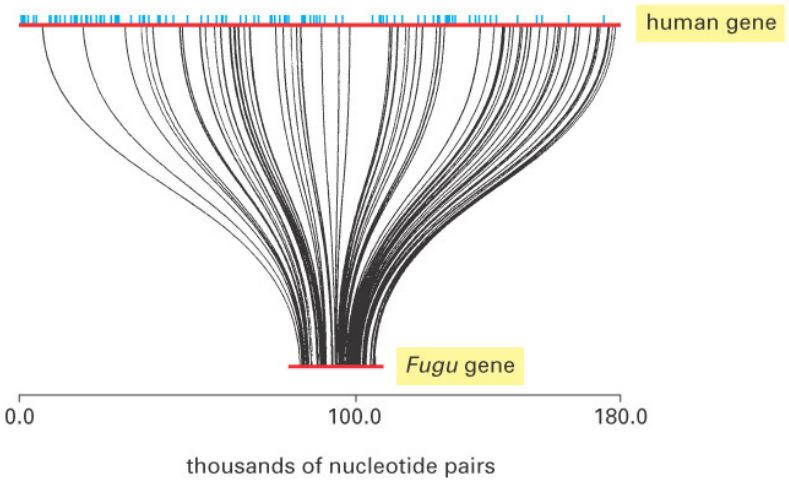


Figure 9-21 Essential Cell Biology, 2/e. (© 2004 Garland Science)

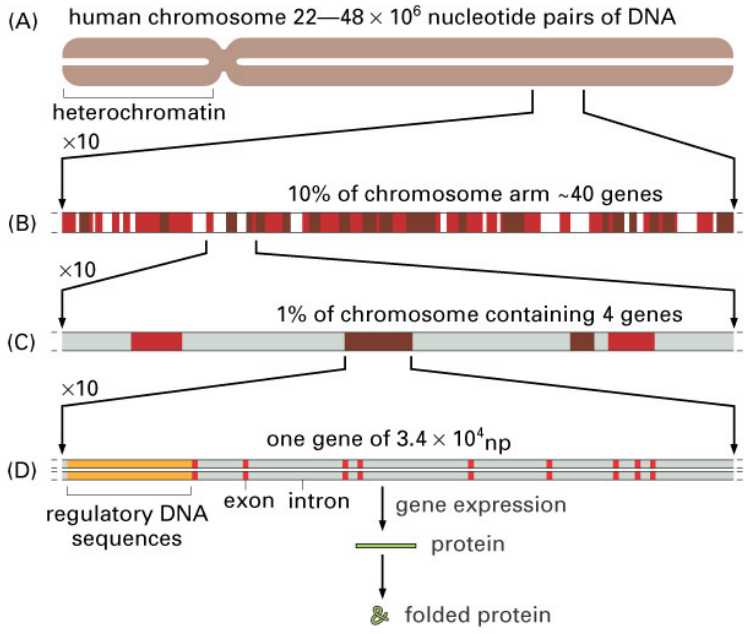
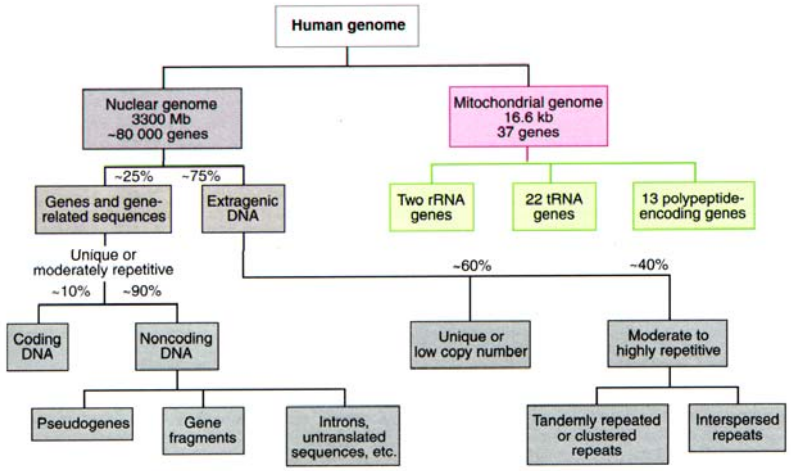


Figure 9-25 Essential Cell Biology, 2/e. (© 2004 Garland Science)

menos del 3% del DNA humano

codifica para proteínas o RNA

gran parte de la secuencia del DNA no se transcribe a RNA



menos del 3% del DNA humano
codifica para proteínas o RNA

ENCODE
Encyclopedia of DNA Elements
nature.com/encode

ENCODE

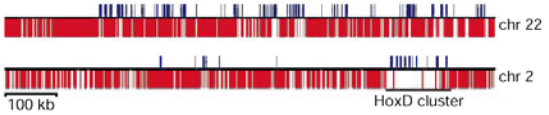
***NATURE* | VOL 489 | 6 SEPTEMBER 2012**

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for **80% of the genome**, in particular **outside of the well-studied protein-coding regions**. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.

Secuencias repetidas dispersas
suman alrededor del 45% del genoma humano

Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	Autonomous	ORF1 ORF2 (pol) AAA	6–8 kb	850,000	21%
SINEs	Non-autonomous	A B AAA	100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous	gag pol (env)	6–11 kb	450,000	8%
	Non-autonomous	(gag)	1.5–3 kb		
DNA transposon fossils	Autonomous	transposase	2–3 kb	300,000	3%
	Non-autonomous	[ ]◀	80–3,000 bp		



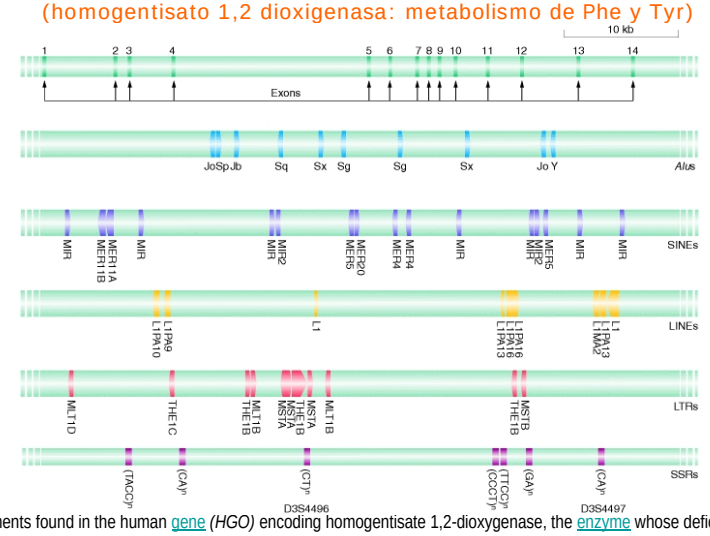
Exons
Repeats

Figure 21 Two regions of about 1 Mb on chromosomes 2 and 22. Red bars, interspersed repeats; blue bars, exons of known genes. Note the deficit of repeats in the HoxD cluster, which contains a collection of genes with complex, interrelated regulation.

IHGSC. *Nature* (2001) **409** 860-921

Ejemplo: Secuencias repetidas dispersas en la secuencia de un gen

(homogentisato 1,2 dioxigenasa: metabolismo de Phe y Tyr)

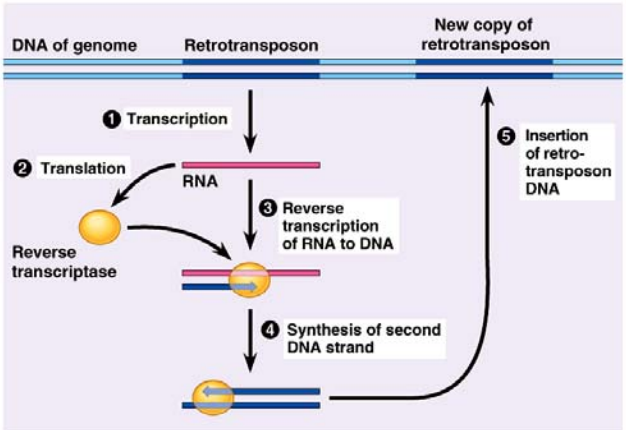


Repetitive elements found in the human gene *(HGO)* encoding homogentisate 1,2-dioxygenase, the enzyme whose deficiency causes alkaptonuria. The first line diagrams the position of the *HGO* exons. The location and direction of *Alus* (blue), SINEs (purple), LINEs (orange), retrotransposon-derived sequences (LTRs, red), and short-sequence repeats (SSRs, maroon) in the *HGO* sequence are indicated by color. (After B. Granadino, D. Beltrán-Valero de Bernabé, J. M. Fernández-Cañón, M. A. Peñalva, and S. Rodríguez de Córdoba, "The Human Homogentisate 1,2-Dioxygenase *(HGO)* Gene," *Genomics* 43, 1997, 115.)

Retrotransposons

- Transposons actually make up over 50% of the corn (maize) genome & > 10% of the human genome.

Most of these transposons are **retrotransposons**, transposable elements that move within a genome by means of **RNA** intermediate, transcript of the retrotransposon **DNA**

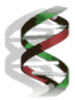


secuencias repetidas dispersas

Summary of Mobile Elements in the Human Genome

Element	% of total genome	copy number
L1 (LINE)	16.9	0.5×10^6
Alu (SINE)	10.6	1.1×10^6
L2 (LINE)	3.2	0.3×10^6
MIR (SINE)	2.5	0.46×10^6
LTR elements	8.3	0.3×10^6
DNA elements	2.8	0.3×10^6
Processed pseudogenes	<1.0	$1\text{-}2 \times 10^4$
Total	~45	$\sim 3 \times 10^6$

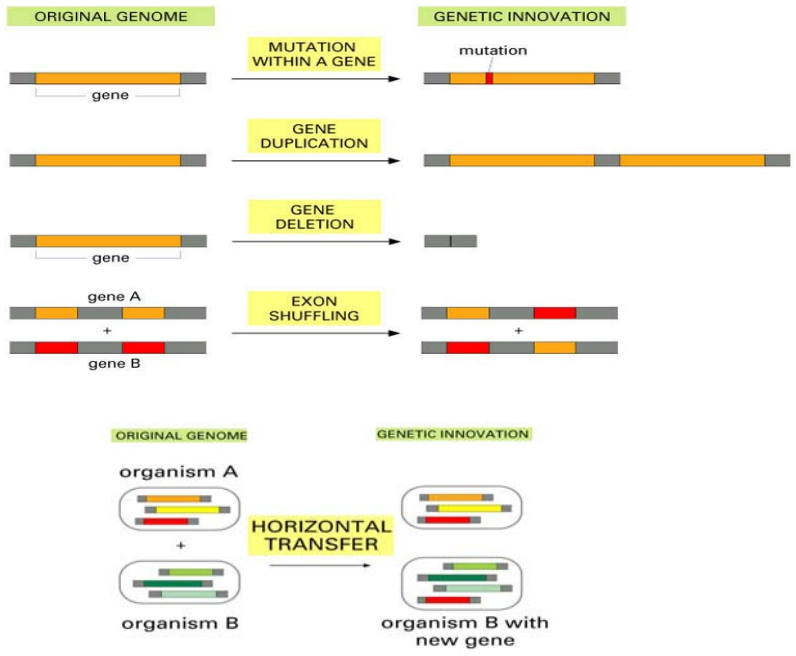
Genome Research 12 (10), 1455-1465 (October 2002) REVIEW: Mammalian Retroelements; P.L. Deininger and M.A. Batzer



Elementos repetidos y evolución de genomas

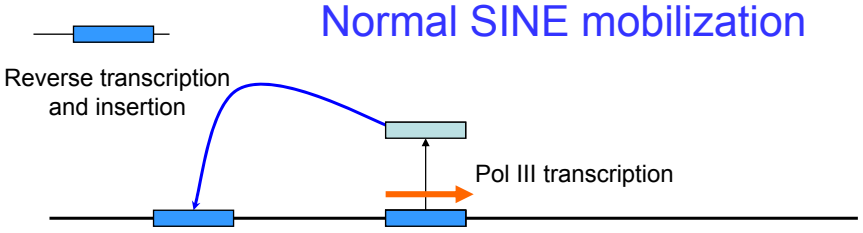
Generating Genetic Variation

- **Five types of change contribute to evolution**
  - Mutation within a gene
  - Gene duplication
  - Gene deletion
  - Exon shuffling
  - Horizontal transfer (rare in eukaryotes)



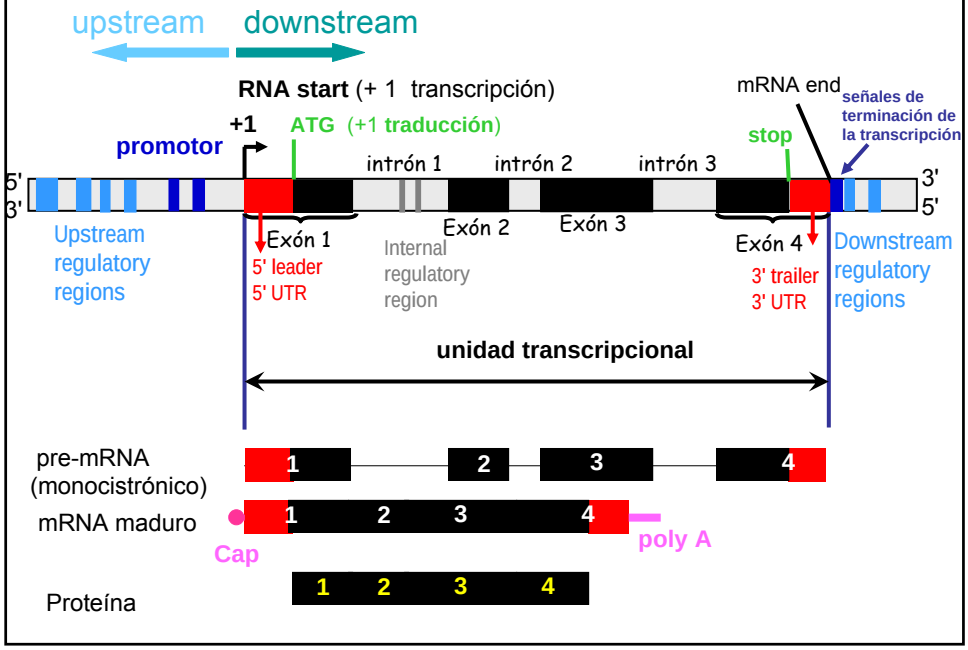
Generating Genetic Variation:

Normal SINE mobilization



1. Usually a single 'master' copy
2. Pol III transcription to an RNA intermediate
3. Target primed reverse transcription (TPRT) – enzymatic machinery provided by LINEs

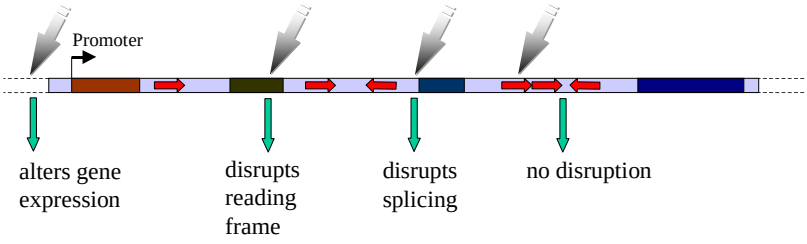
Unidad transcripcional eucariota



Generating Genetic Variation

Mutations within genes

The insertion of mobile elements can disrupt gene structure and function



Lista de enfermedades genéticas como consecuencia de inserción de elementos Alu

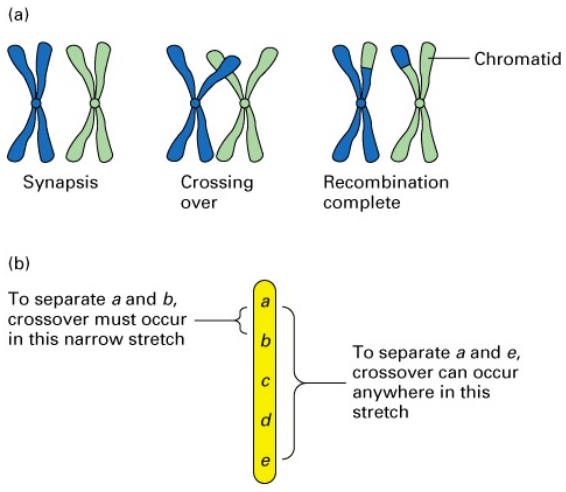
Mobile DNA elements probably had a significant influence on evolution

- Spontaneous **mutations** may result from the **insertion** of a mobile DNA element into or near a transcription unit
- Homologous **recombination** between mobile DNA elements may contribute to gene duplication and other rearrangements, including duplication of introns, recombination of introns to create new genes, and control of gene expression

Functional rearrangements in chromosomal DNA

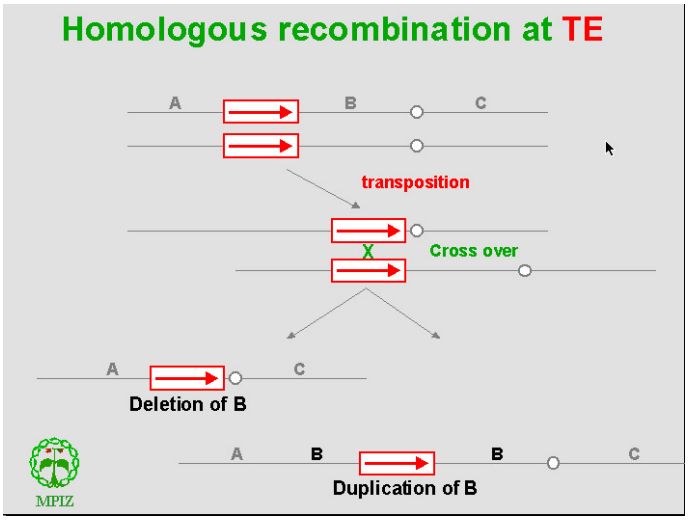
- While the transposition of mobile elements appears to serve no direct immediate function to the organism, several types of rearrangements are beneficial to the organism
- Examples include inversion and deletion of DNA segments and DNA amplification
- These functional rearrangements occur in prokaryotes and eukaryotes

Homologous recombination (meiosis)



¡OJO! en este esquema cada línea representa una molécula de dsDNA

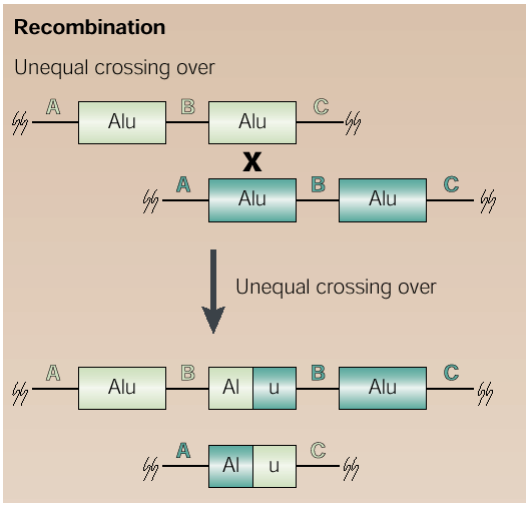
Transposición de un elemento móvil (en uno de los cromosomas de un par) seguida de recombinación homóloga (entre las secuencias de durante la meiosis)
► duplicación del segmento B...



www.mpiz-koeln.mpg.de/ ~rohde/saedler2.html

Recombination: unequal crossing over

Apart from influencing gene expression and function through insertions, the large number of Alu elements, and to a lesser extent L1 elements, may promote homologous recombination events.



Prak & Kazazian. (2000) NRG. 1: 134-144

Duplicación de genes

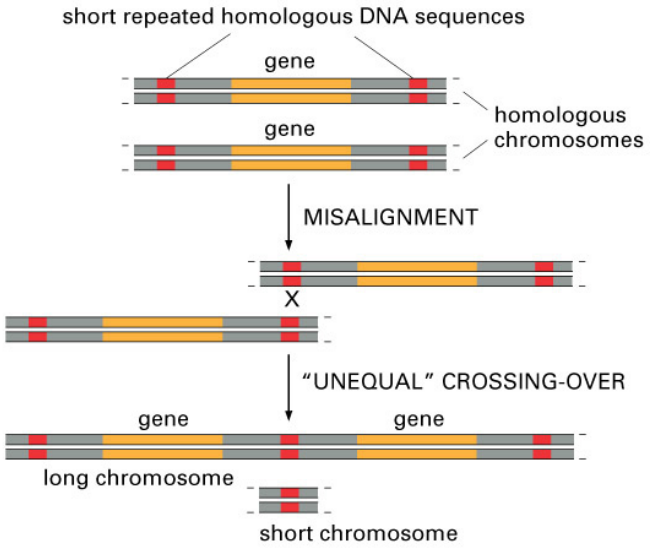
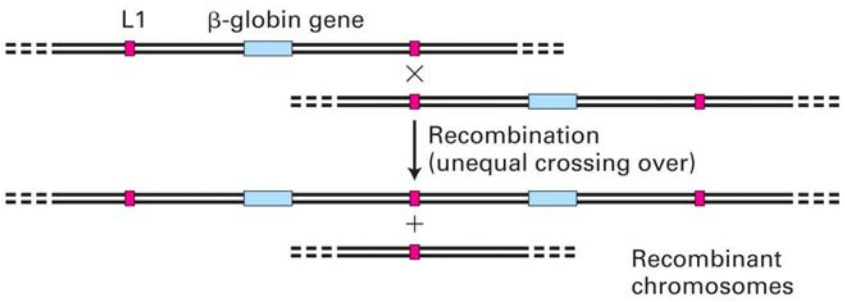


Figure 9-5 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Unequal cossing over…
gene duplication (10.4)



Duplicación de genes ...

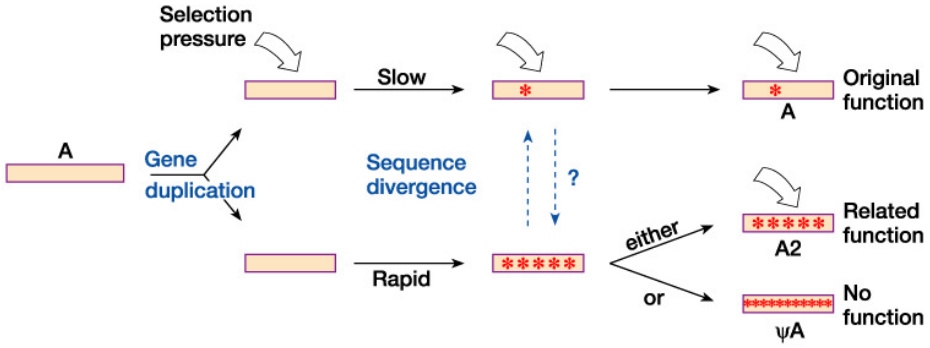


Figure 12-3 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Generating Genetic Variation

- **Gene Duplication: the globin family**
  - A classic example of gene duplication and evolution
  - Globin molecules are involved in carrying oxygen in multicellular organisms
  - Ancestral globin gene (present in primitive animals) was duplicated ~500 Myr.
  - Mutations accumulated in both genes to differentiate them - α and β present in all higher vertebrates
  - Further gene duplications produced alternative forms in mammals and in primates

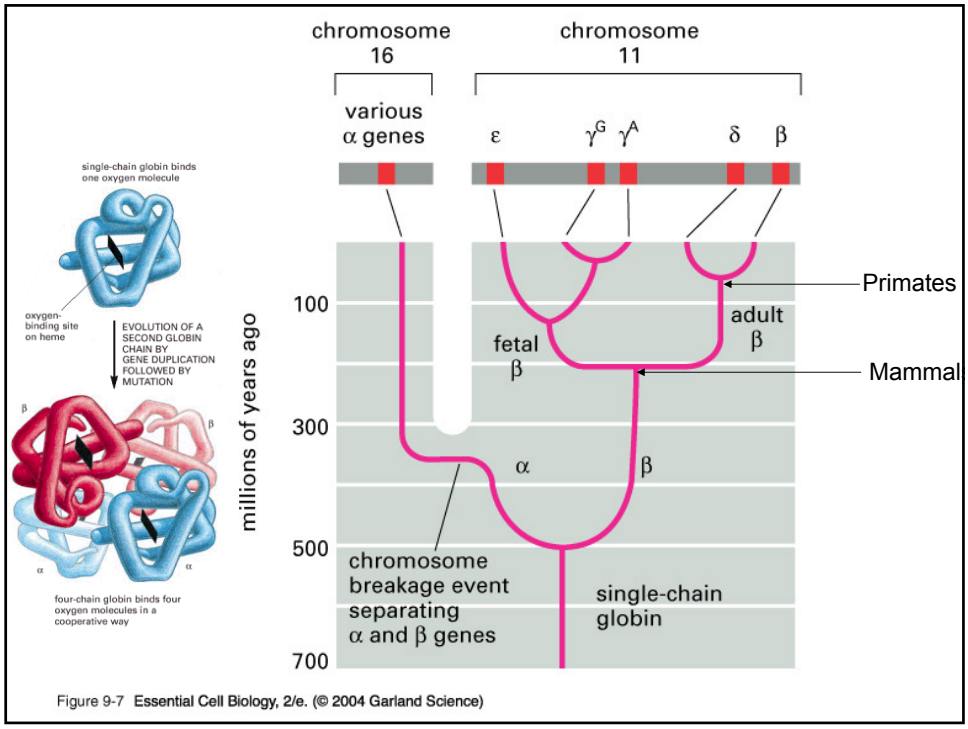
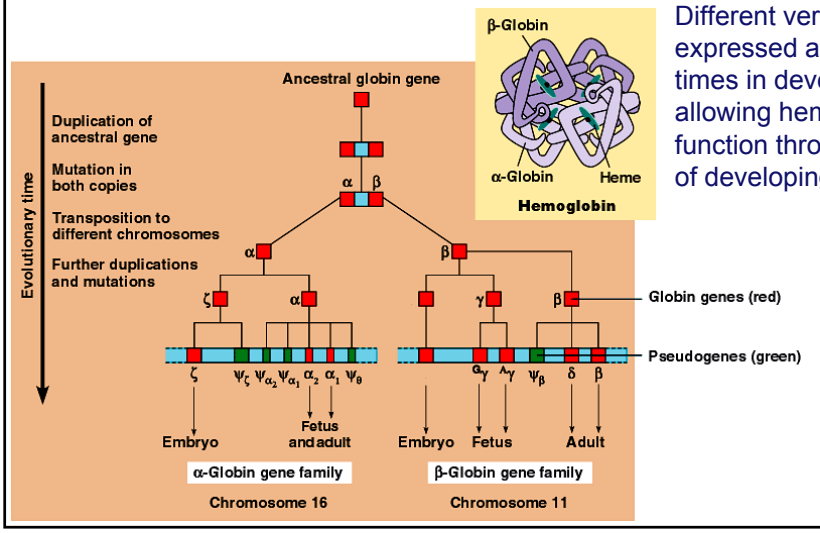


Figure 9-7 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Human globin gene family

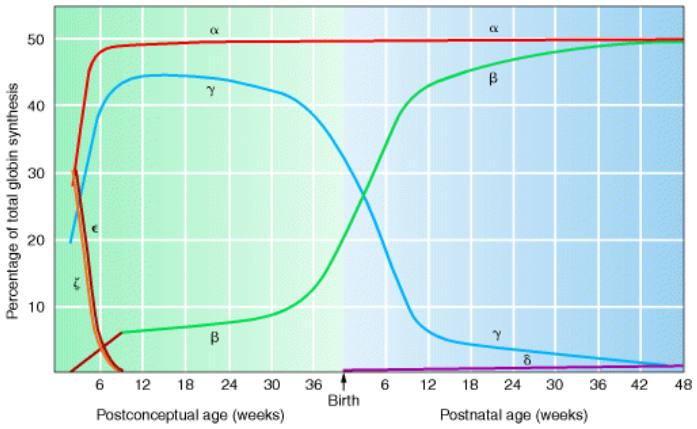
evolved from duplication of common ancestral globin gene



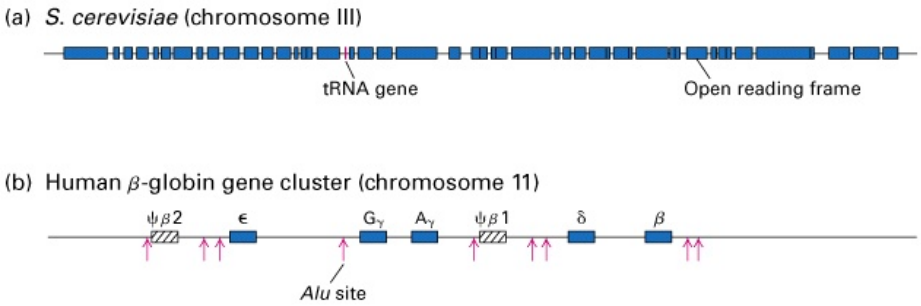
Different versions are expressed at different times in development allowing hemoglobin to function throughout life of developing animal

Expresión de genes de globina en el período fetal y luego del nacimiento

(Griffiths et al 2002)



Genomes of higher eukaryotes contain much nonfunctional DNA, as well as **gene families** derived from a single ancestral gene (e.g. globin)



Amongst eukaryotes, cellular DNA content does not correlate with phylogeny.
Genomes of higher eukaryotes contain much nonfunctional DNA

Duplicación de exones

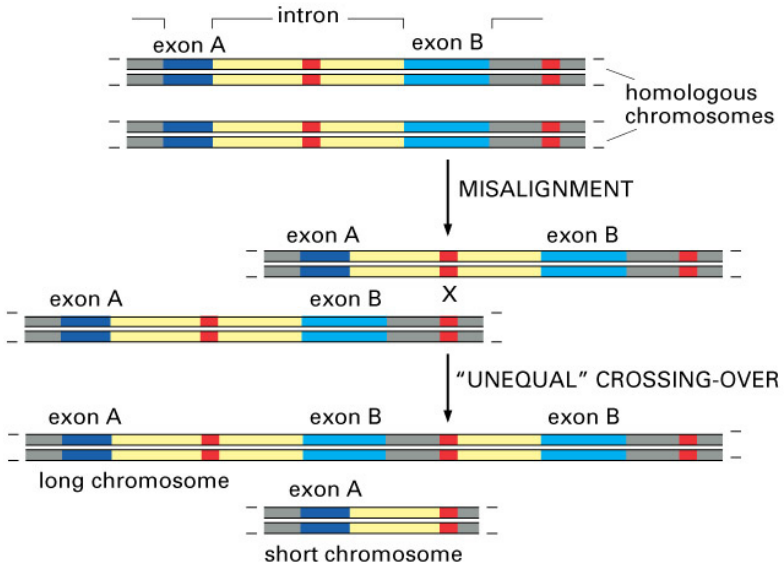


Figure 9-9 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Intercambio de exones por recombinación homóloga entre elementos repetidos dispersos

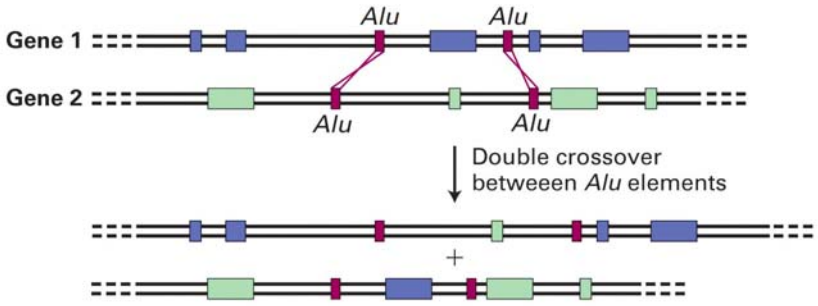


FIGURE 10-17 Exon shuffling via recombination between homologous interspersed repeats. Recombination between interspersed repeats in the introns of separate genes produces transcription units with a new combination of exons. In the example shown here, a double crossover between two sets of *Alu* repeats results in an exchange of exons between the two genes.

Generating Genetic Variation

- Exon Shuffling
  - The exons of genes can sometimes be thought of as individual useful units that can be mixed and matched through exon shuffling to generate new, useful combinations

Generación de variabilidad genética

- Exon Shuffling
  - Los exones pueden considerarse como unidades funcionales que se pueden mezclar mediante el “exon shuffling” para generar nuevas combinaciones y dar origen a proteínas con nuevas funciones

adquisición de exones por transposición

(10.18a)

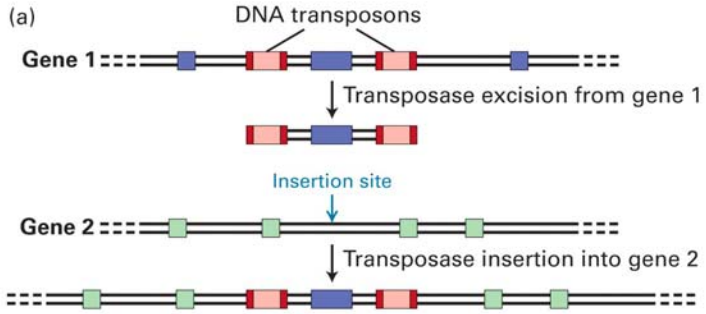
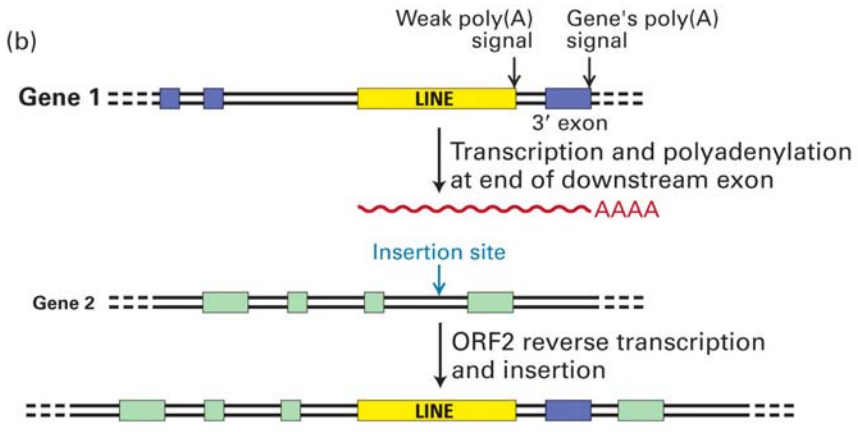
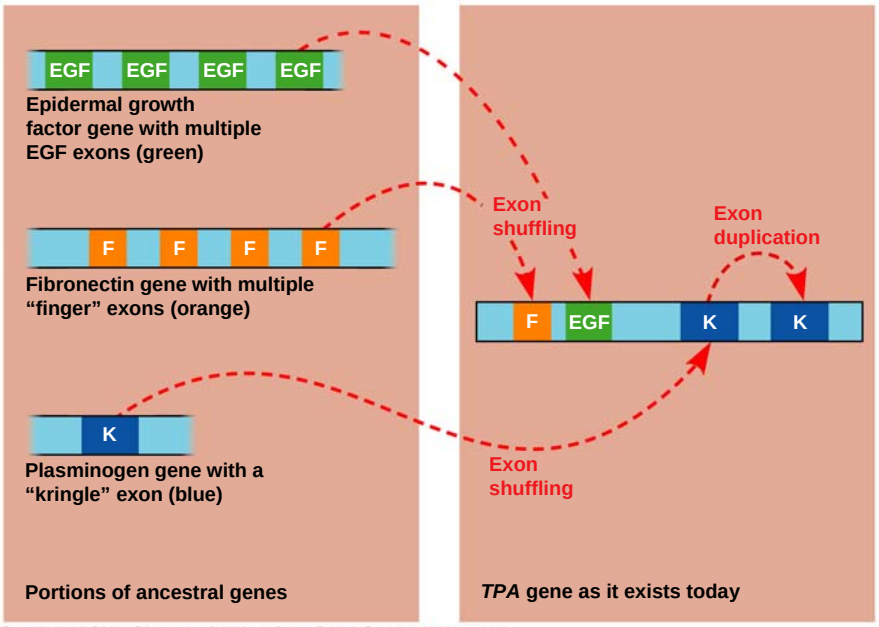


FIGURE 10-18 Exon shuffling by transposition. (a) Transposition of an exon flanked by homologous DNA transposons into an intron on a second gene. As we saw in Figure 10-10, step , transposase can recognize and cleave the DNA at the ends of the transposon inverted repeats. In gene 1, if the transposase cleaves at the left end of the transposon on the left and at the right end of the transposon on the right, it can transpose all the intervening DNA, including the exon from gene 1, to a new site in an intron of gene 2. The net result is an insertion of the exon from gene 1 into gene 2.

Adquisición de un nuevo exón



LE 19-20



Dominios similares en proteínas distintas

Construcción de nuevos genes (sin partir de cero): combinación de exones

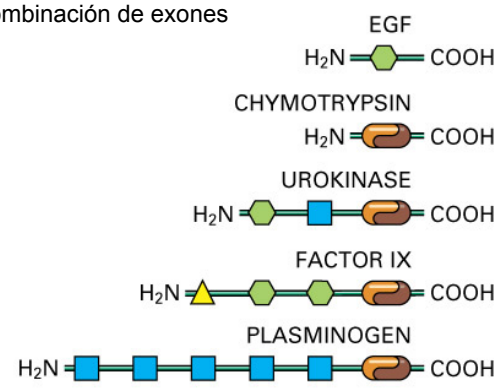
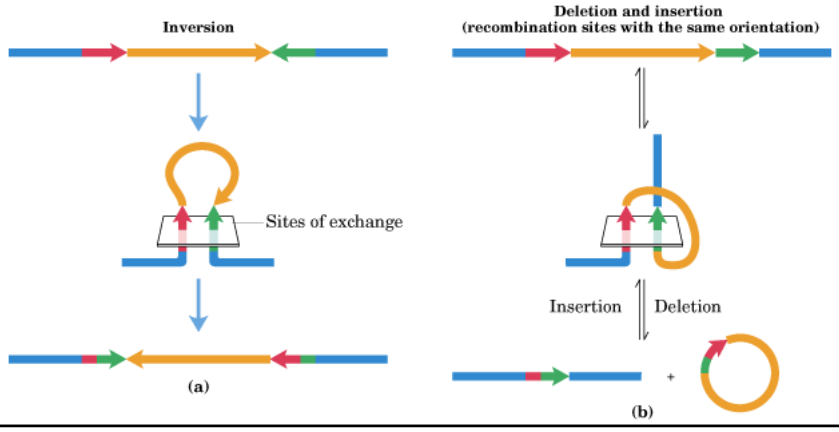


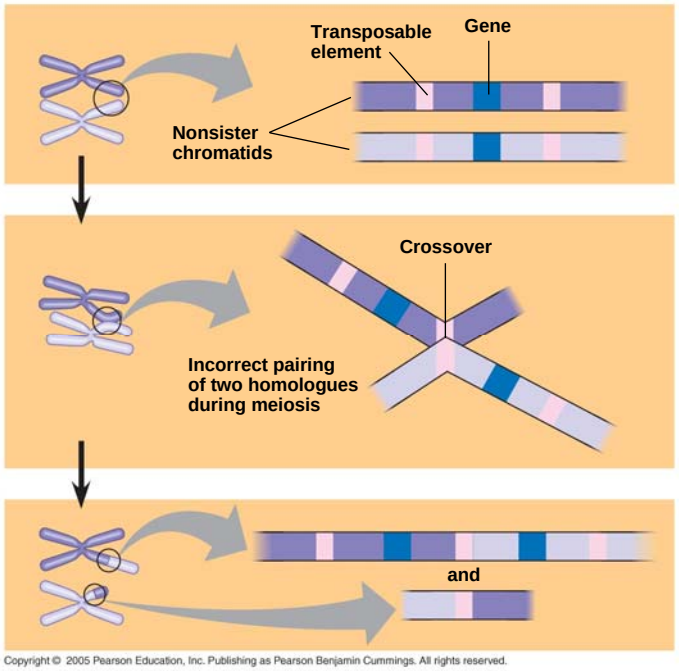
Figure 9-10 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Recombinación intramolecular de elementos repetidos

Repeticiones inversas y directas



LE 19-18

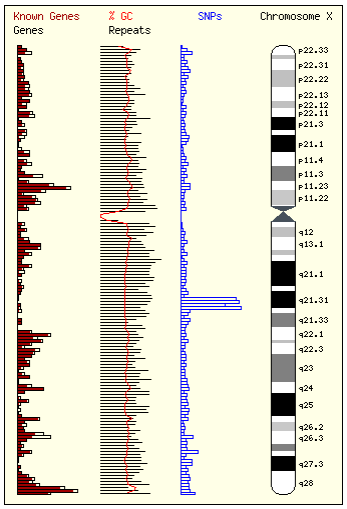


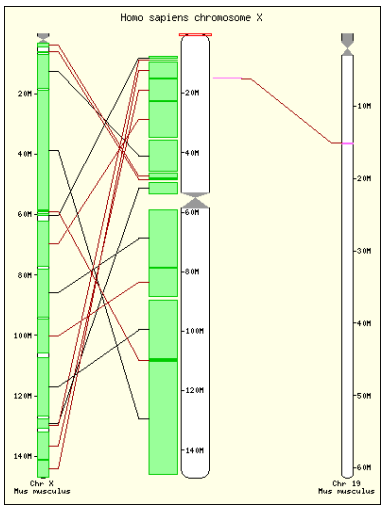
Venkatesh, B., Gilligan, P. and Brenner, S. (2000) *FEBS Lett.* **476**: 3-7.

Comparative vertebrate genomics

	DNA Content (pg)	Genome Size (Mb)	Chromosome No. (n)
Fish			
Fugu *(Fugu rubripes)*	-	390	22
Zebrafish *(Danio rerio)*	1.8	1700	25
Medaka *(Oryzias latipes)*	1.1	1100	24
Amphibians			
Xenopus laevis	3.2	3100	18
Xenopus tropicalis	1.78	1700	10
Bird			
Chicken *(Gallus gallus)*	1.25	1200	39
Mammals			
Human *(Homo sapiens)*	3.5	3000	23
Mouse *(Mus musculus)*	3.5	3000	20
Rat *(Rattus norvegicus)*	3.5	3000	21

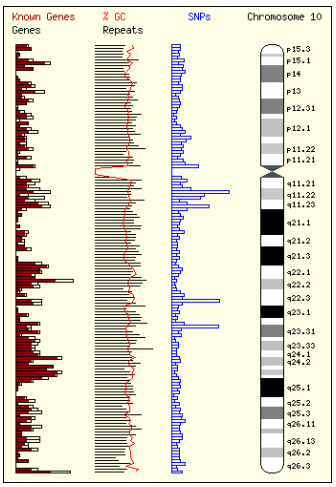
http://www.ensembl.org/Homo_sapiens/

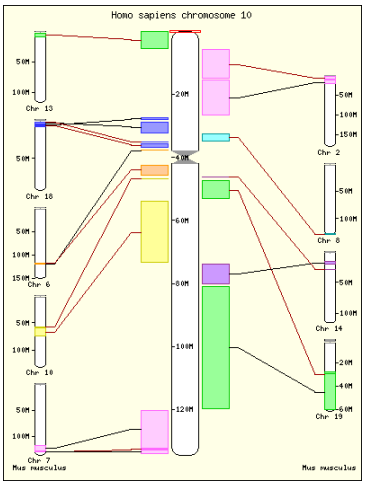




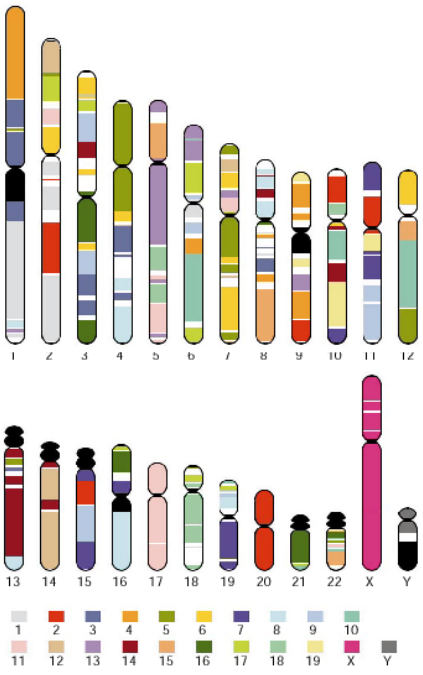
Browse a Chromosome

www.ensembl.org/Homo_sapiens



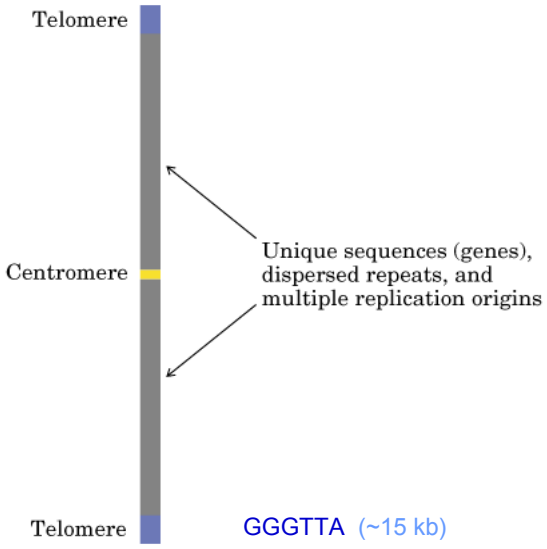


Browse a Chromosome



segmentos conservados entre los genomas del hombre y del ratón (sintenia)

Secuencias **repetidas simples en tandem**:
en regiones teloméricas, centroméricas y otras



GGGTTA (~15 kb)

los *fingerprints* de DNA dependen de las longitudes de los segmentos de **secuencias simples repetidas**

AAGGGTGGGCAGGAAGTGGAGTGTGTGCCTGCTTCCCTTCCCTGTCTTGTCCTGGAAACTCA

λ 33.1 minisatellite

T
C GGGCAGG•AGGGGGAGG

λ 33.5 minisatellite

El número de pares de bases de la secuencia repetida para diferenciar **minisatélites** de **microsatélites** (2 a 6 pb) es arbitrario.

Las secuencias repetidas en tandem (en bloque) son más cortas que las de los elementos repetidos dispersos (provenientes de elementos genéticos móviles) y su número puede ser variable en base a errores de replicación ("tartamudeo" o "deslizamiento" de la DNA pol; "*slippage*": *slipped strand mispairing*) o por eventos de recombinación ("*unequal crossingover*")

Repeated DNA

Minisatellites: VNTRs = Variable **N**umber **T**andem **R**epeats

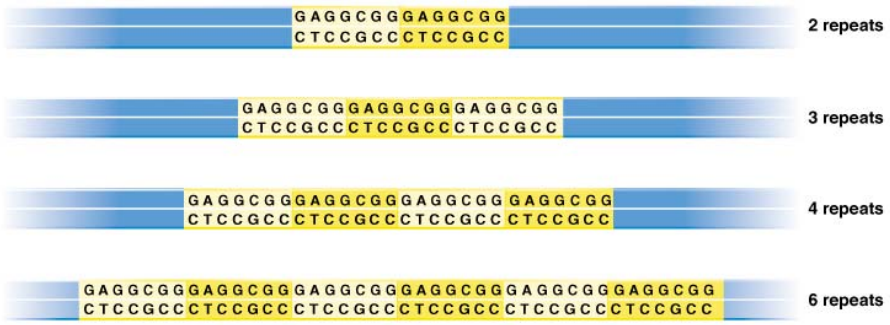
Repeated units of 6 to 100 bp (approx.)

Minisatellites are usually defined as the repetition in tandem of a short (6- to 100-bp) motif spanning 0.5 kb to several kilobases ("Minisatellites: Mutability and Genome Architecture"; G. Vergnaud & F. Denoeud *Genome Res. 2000. 10: 899-907)* .

Microsatellites: STRs = Short **T**andem **R**epeats

SSRs = Simple **S**equence **R**epeats

Repeated units of 2-6 bp

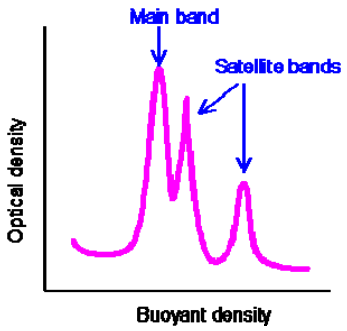


Repeated DNA

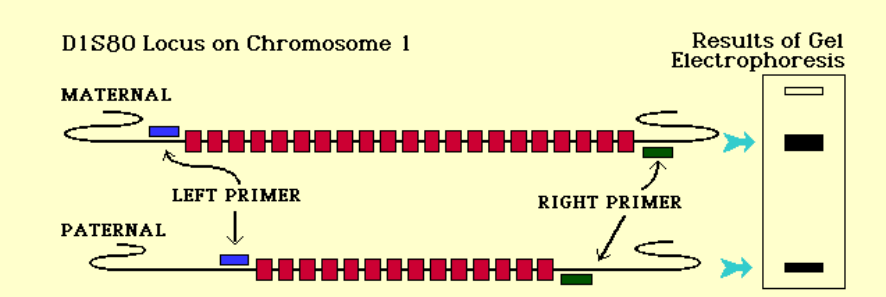
Minisatellites: VNTRs = Variable **N**umber **T**andem **R**epeats

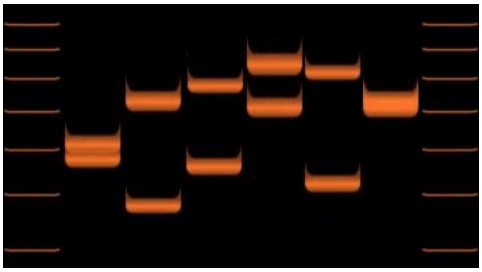
Microsatellites: STRs = Short **T**andem **R**epeats

AATTTTTGTA TTTTTTTTAG AGACGGGGTT TCACCATGTT GGTCAGGCTG ACTATGGAGT
TATTTTAAGG TTAATATATA TAAAGGGTAT GATAGAACAC TTGTCATAGT TTAGAACGAA
CTAACGATAG ATAGATAGAT AGATAGATAG ATAGATAGAT AGATAGATAG ATAGACAGAT
TGATAGTTTT TTTTTATCTC ACTAAATAGT CTATAGTAAA CATTTAATTA CCAATATTTG
GTGCAATTCT GTCAATGAGG ATAAATGTGG AATCGTTATA ATTCTTAAGA ATATATATTC
CCTCTGAGTT TTTGATACCT CAGATTTTAA GGCC



los *fingerprints* de DNA dependen de las longitudes de los segmentos de **secuencias simples repetidas**





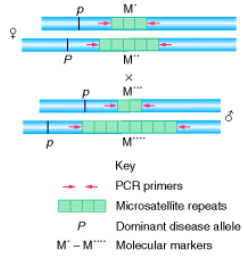
Variations of VNTR (D1S80) allele lengths in 6 individuals.

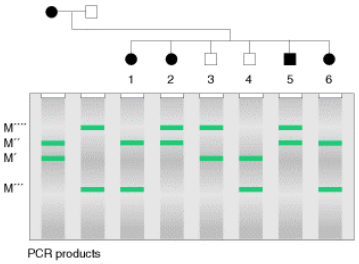
MICROSATÉLITES, SIMPLE SEQUENCE REPEAT (SSR), STR O VNTR

La técnica recibe varias denominaciones:
microsatelite, simple sequence repeat, (SSR),
short tandem repeat (STR) o variable number tandem repeat (VNTR).

Detecta secuencias cortas (microsatélite) de 2-5 bp cuyo motivo se repite muchas veces y está flanqueado por un DNA único.Estas zonas del genoma son hipervariables.

El motivo del microsatelite es usado como **sonda** contra la biblioteca genómica o de cDNA para identificar los clones conteniendo este motivo. Estos clones son posteriormente secuenciados y los primers se diseñan para amplificar los DNA únicos que flanquean el motivo del microsatelite.



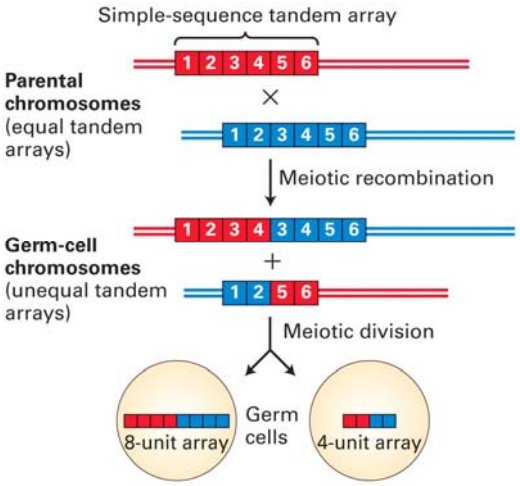


Este método es repetitivo e identifica **locus simples**.
El blanco son regiones hipervariables del genoma.

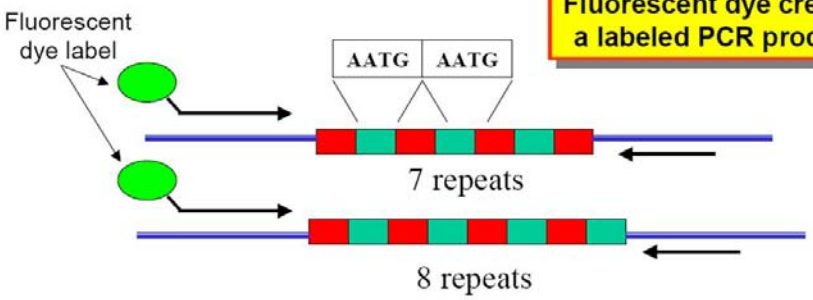
El polimorfismo se detecta como diferencias en el tamaño del producto amplificado. A veces estas diferencias son muy pequeñas (por ej 2bp) por lo que se requiere de tiempo y recursos para efectuarlo.

Esta técnica es apropiada para mapeo, y **fingerprinting.**

Unequal crossing over …..
Differences in lengths of simple-sequence DNA repeats (10.6)



Short Tandem Repeats (STRs)



Fluorescent dye creates a labeled PCR product

the repeat region is variable between samples while the flanking regions where PCR primers bind are constant

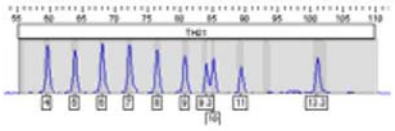
Homozygote = both alleles are the same length

Heterozygote = alleles differ and can be resolved from one another

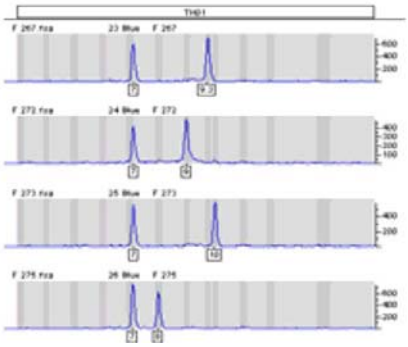
Primer positions define PCR product size

Variation of STRs Among Individuals

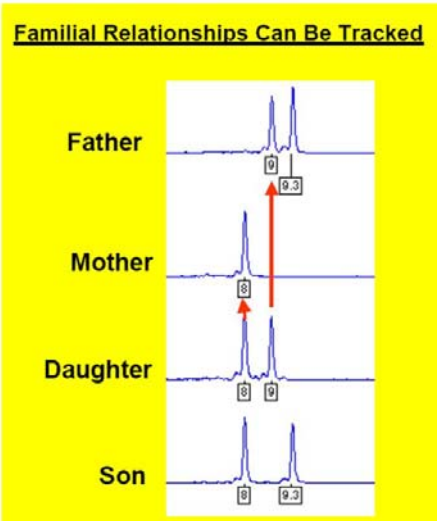
An allelic ladder, which is a mixture of common alleles, is used to convert DNA size to STR repeat number



Unrelated Individuals



Familial Relationships Can Be Tracked



VNTR

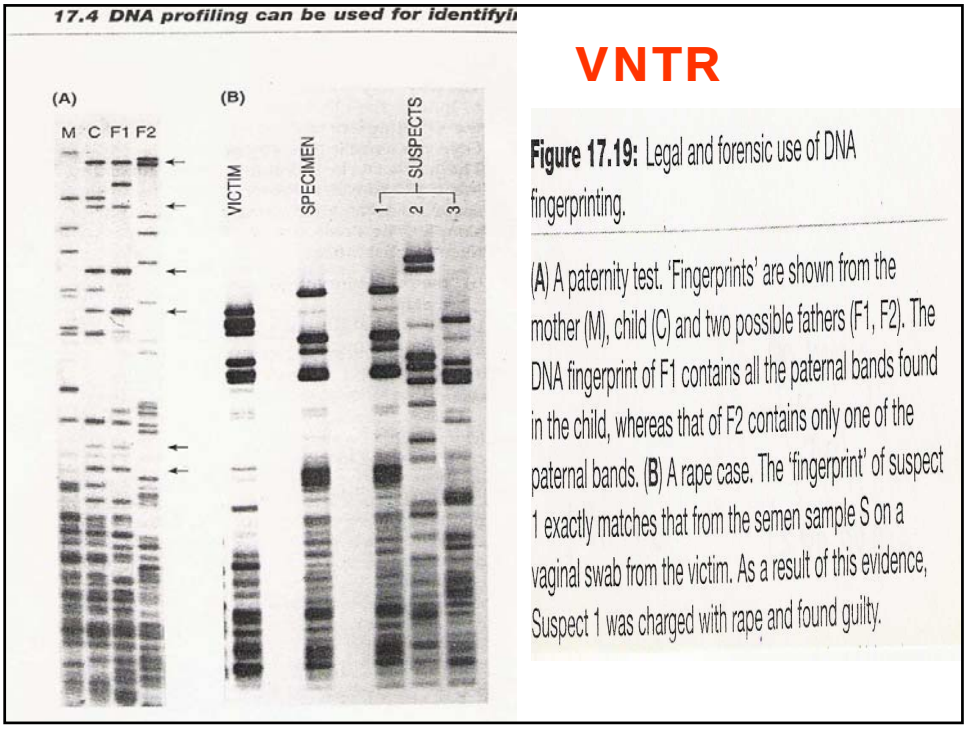
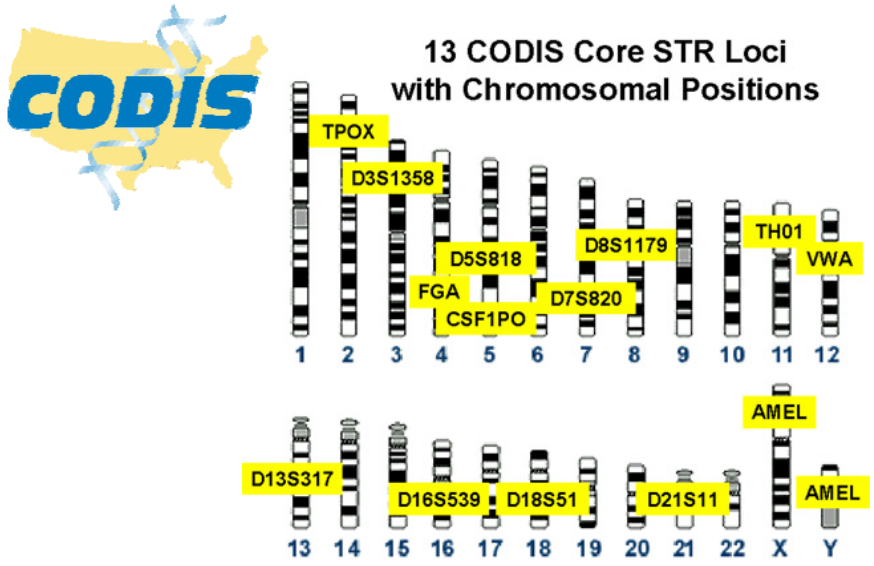


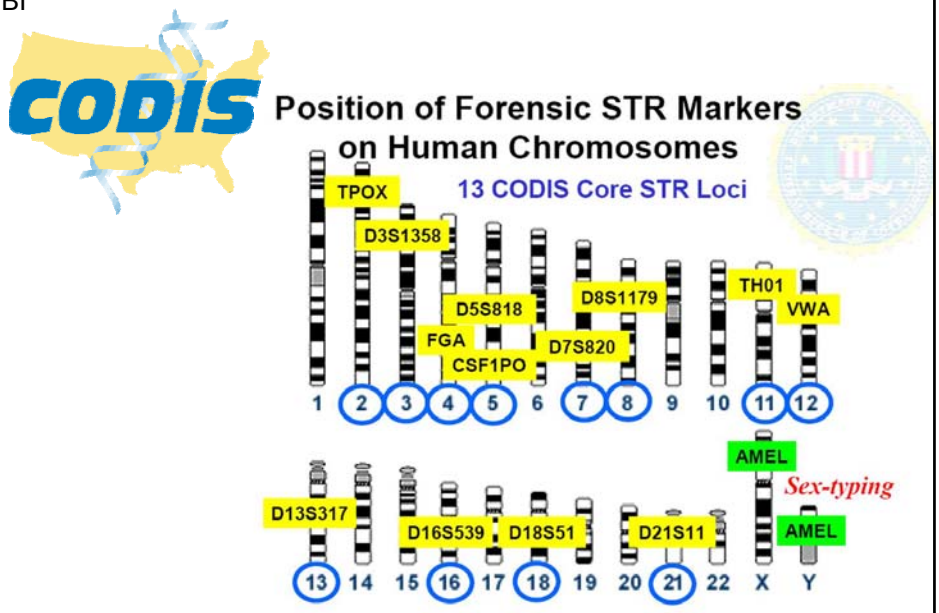
Figure 17.19: Legal and forensic use of DNA fingerprinting.

(**A**) A paternity test. 'Fingerprints' are shown from the mother (M), child (C) and two possible fathers (F1, F2). The DNA fingerprint of F1 contains all the paternal bands found in the child, whereas that of F2 contains only one of the paternal bands. (**B**) A rape case. The 'fingerprint' of suspect 1 exactly matches that from the semen sample S on a vaginal swab from the victim. As a result of this evidence, Suspect 1 was charged with rape and found guilty.

Combined **D**NA **I**ndex **S**ystem (**CODIS**)
FBI



Combined DNA Index System (CODIS)
FBI



fingerprints de DNA: detección de inserciones de **elementos móviles**