

Bioinformática y Biología molecular

Bioinformática

13-2-19

Elvira Mayordomo



En los periódicos

- La biotecnología, genética y bioinformática en primera plana
- Todo empezó con el descubrimiento de la estructura del DNA por Watson y Crick en 1953
- En los 90 se inició el proyecto del genoma humano y se clonó a la oveja Dolly
- En el 2000 se anunció la secuenciación completa del genoma humano
- En el 2008 comenzó el proyecto de los 1000 genomas
- Ya hemos llegado al “\$1000 genome”?



Secuenciando DNA

- El objetivo es determinar una secuencia de “nucleótidos” que son las piezas que forman el DNA humano, es decir, la molécula que guarda nuestra información genética
- Desde el punto de vista informático buscamos un string hecho con las letras que representan los nucleótidos
- Conocemos métodos para leer estas secuencias desde los 80, pero con longitudes muy restringidas (hoy unos 1000 nucleótidos)
- Nos interesan moléculas de DNA con cientos de miles



Secuenciando DNA ... ¿Cómo?

- Se generan muchas copias de la molécula de DNA que nos interesa
- Rompemos aleatoriamente esas copias en trozos, idealmente pequeños
- Con alta probabilidad esos trozos se solapan entre sí
- Leemos (“secuenciamos”) los trozos
- Nos quedan muchos (miles de) trozos que son subsecuencias de la que buscamos, con solapamientos
- No tenemos idea de cómo combinarlos, el orden se ha perdido
- Aquí entra la informática ...

→ métodos “next generation”



Detalles con mucha importancia

- Queremos derivar modelos formales de problemas biológicos para encontrarles soluciones algorítmicas
- Pero es imposible olvidarnos del problema biológico original porque la formalización siempre es “burda”
- Todos los datos biológicos son inherentemente inexactos



Métodos computacionales en bioinformática

- Gestión de bases de datos
 - Estadística
 - Algorítmica
-
- En este curso sobre todo algorítmica. La primera parte y algo del resto de
Algorithmic Aspects of Bioinformatics.
Bockenhauer, Bongartz. Springer 2008



Biología molecular ...¿Para qué?

- Necesitamos conocimientos básicos de Biología molecular para poder desarrollar y evaluar modelos abstractos y técnicas para manejarlos
- Trataremos las **proteínas** y los **ácidos nucleicos**

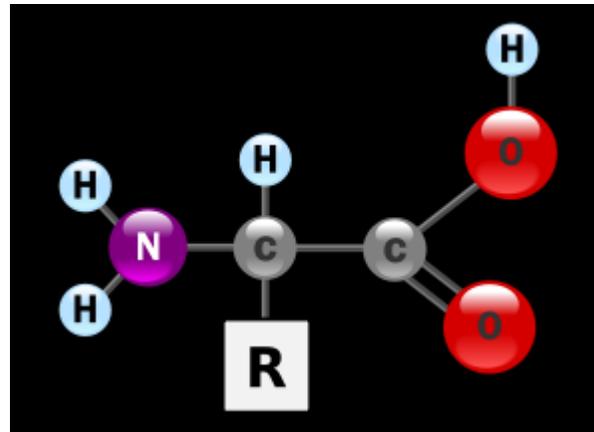


Proteínas

- La clase de moléculas más importantes de los seres vivos
- Funciones: como enzimas (catálisis de procesos metabólicos), en transmisión de señal, mecanismos de defensa, transporte de moléculas, material de construcción

Los aminoácidos forman las proteínas

- Una proteína es una cadena de aminoácidos
- Un aminoácido:



R es una “cadena lateral”

20 aminoácidos

- Según R aparecen 20 aminoácidos distintos que forman parte de las proteínas

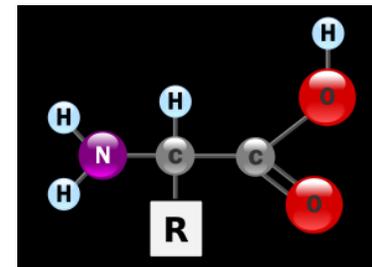
Ala A (H)	Val V (H)	Leu L (H)	Ile I (H)	Phe F (H)
Pro P (H)	Met M (H)	Ser S (P)	Thr T (P)	Cys C (P)
Trp W (H)	Tyr Y (P)	Asn N (P)	Gln Q (P)	Asp D (P)
Glu E (P)	Lys K (P)	Arg R (P)	His H (P)	Gly G (P)

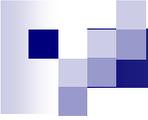
Proteínas = cadenas de aminoácidos

- Los aminoácidos se unen mediante enlaces peptídicos
- Podemos representar una proteína como una cadena leída de (H²N a COOH)

VHLTPEEK ...

- Esto es mucho simplificar, ignorando la estructura espacial ...



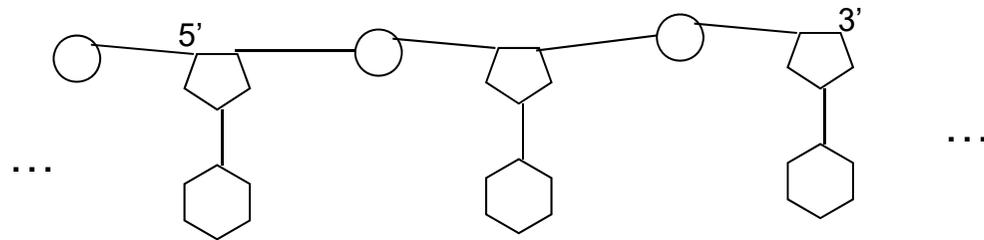
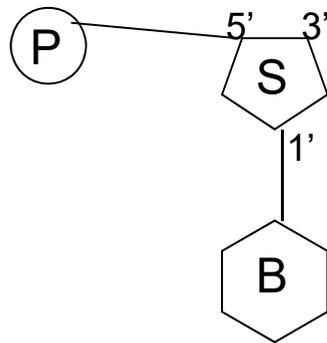


Ácidos nucleicos

- Las moléculas más importantes después de las proteínas
- En todos los seres vivos, son las responsables de codificar y almacenar la información genética
- Permiten la transmisión de información genética de una generación a otra
- Tienen una estrecha conexión con las proteínas: los ácidos nucleicos sirven como mapas para la construcción de las proteínas

Ácido nucleico= cadena de nucleótidos

- Cada ácido nucleico está formado por nucleótidos encadenados





Cadenas ...

- La B (base) caracteriza al nucleótido
- Podemos escribir un ácido nucleico como una secuencia de nucleótidos
- Gran simplificación



DNA y RNA

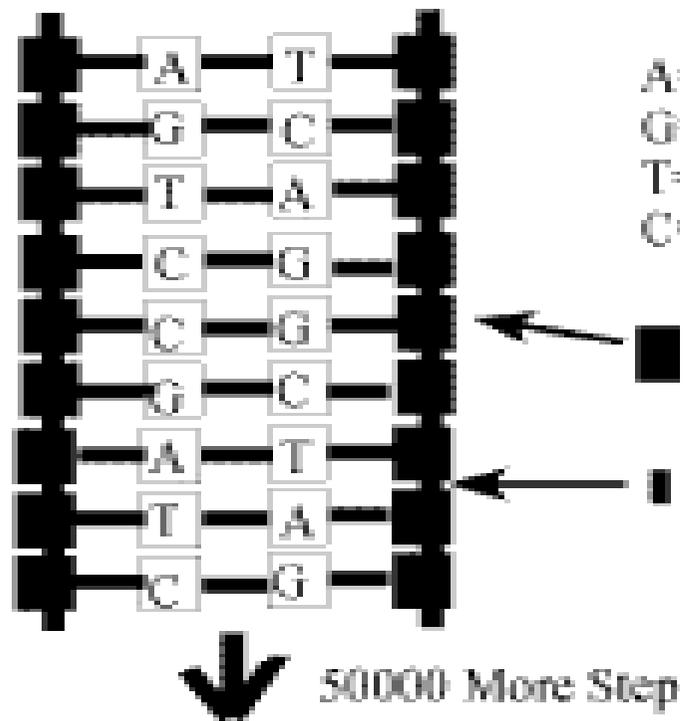
- Los dos tipos de ácidos nucleicos son DNA y RNA
- Se diferencian por los azúcares (S)
- DNA usa cuatro bases A, C, G, T
- RNA usa cuatro bases A, C, G, U



DNA

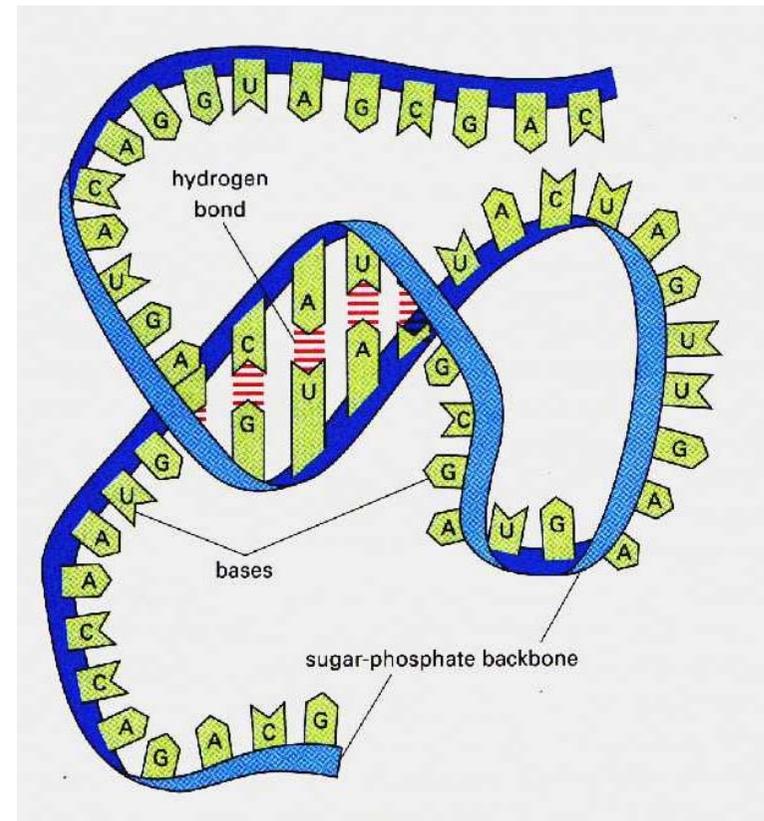
- Está formado por dos cadenas de nucleótidos
 - Complementarias (A-T y C-G)
 - Se leen en direcciones opuestas
 - En forma de hélice

DNA



RNA

- Normalmente una sola cadena
- Trozos de la misma molécula se unen con otros complementarios lo que da formas diversas





Triple estructura

- Primaria: cadena de nucleótidos
- Secundaria: describe los trozos complementarios que están unidos
- Terciaria: cómo está doblada en el espacio

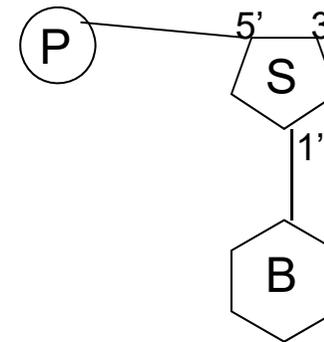
Cómo escribimos DNA

- Usamos la dirección de lectura de la cadena de “arriba” (de 5’ a 3’)

Ejemplo: s=AGACGT es:

s: 5' ...AGACGT ... 3'

\bar{s} : 3' ...TCTGCA ... 5'





Longitudes

- Se mide la longitud en pares de bases (bp)
- Se usan las unidades kbp (1000 bp) y mbp (1000 kbp)

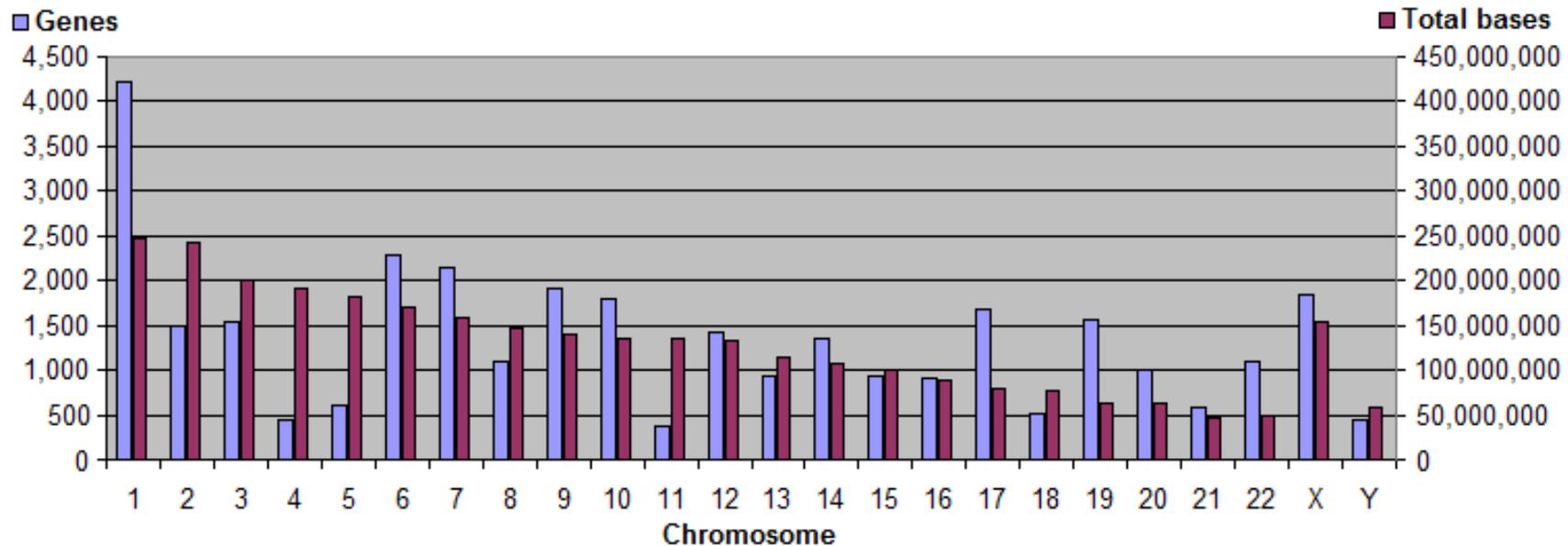
- AGACGT tiene 6 bp
- También se usan para RNA



Información hereditaria

- Una región de DNA que codifica una proteína se llama **gen**
- Una molécula de DNA que tiene varios genes se llama **cromosoma**
- Los cromosomas suelen aparecer en pares: **cromosomas homólogos** (uno materno y uno paterno)
- Las células humanas tienen 46 cromosomas
- Toda la información hereditaria de una célula se llama **genoma**

23 pares de cromosomas humanos

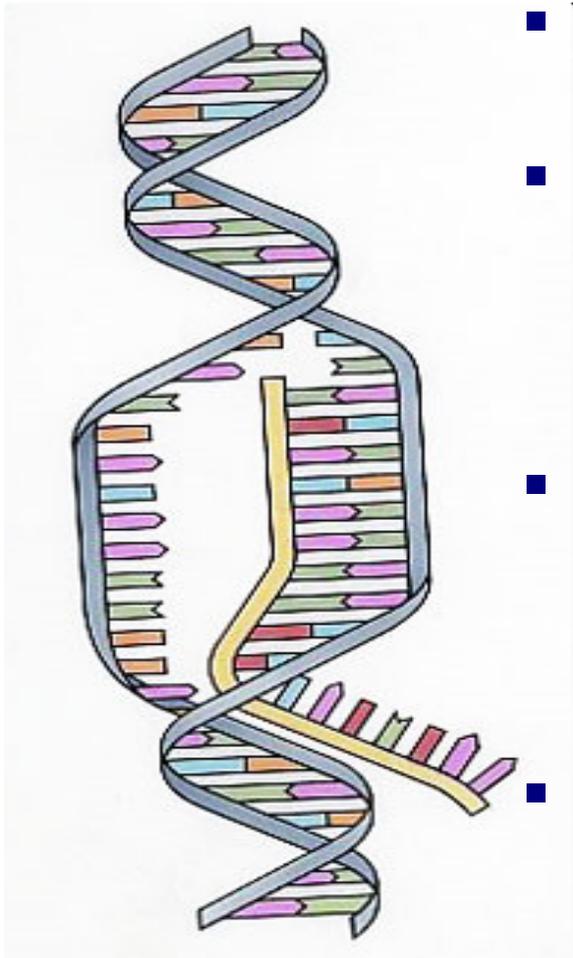




Síntesis de proteínas

- El DNA está en el núcleo y las proteínas se sintetizan fuera (en los ribosomas)
- Hay dos pasos: copia y traducción

Transcripción o copia



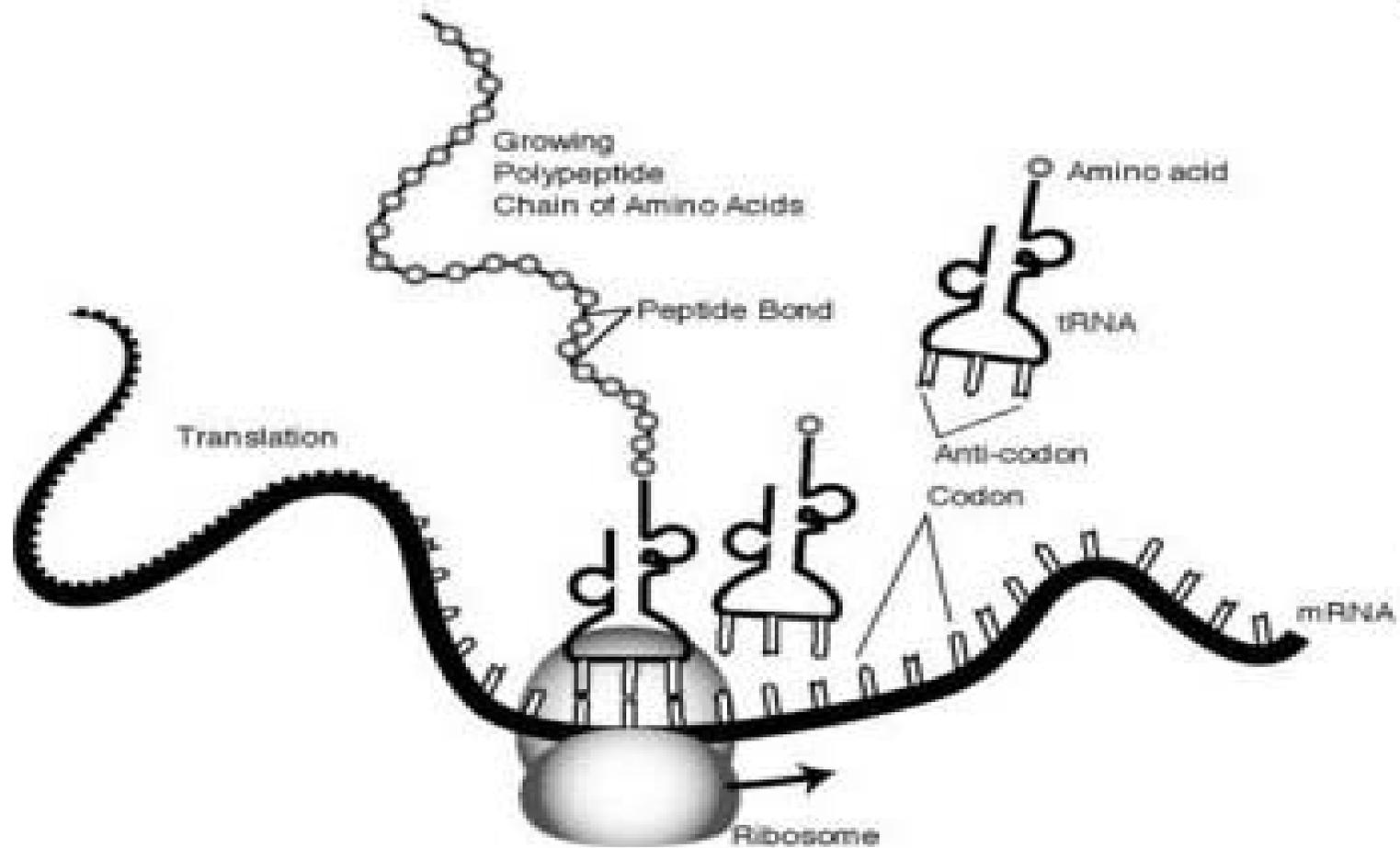
- Se separan las dos copias de DNA
- Se sintetiza una copia complementaria de RNA (amarillo) sustituyendo T por U
- Sólo se copian algunas zonas (los exones) y no los intrones (¿no relevantes?)
- El resultado es el mRNA (RNA mensajero)



Traducción

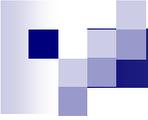
- La información del mRNA se convierte en una secuencia de aminoácidos
- Un codón son 3 bases. Cada codón codifica un aminoácido
- También hay codones que codifican STOP

Traducción



Codón-aminoácido

		Second Position									
		U		C		A		G			
		code	Amino Acid	code	Amino Acid	code	Amino Acid	code	Amino Acid		
First Position	U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U	
		UUC		UCC		UAC		UGC		C	
		UUA	leu	UCA		UAA	STOP	UGA	STOP	A	
		UUG		UCG		UAG	STOP	UGG	trp	G	
	C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U	
		CUC		CCC		CAC		CGC		C	
		CUA		CCA		CAA	gln	CGA		A	
		CUG		CCG		CAG		CGG		G	
	A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U	
		AUC		ACC		AAC		AGC		C	
		AUA		ACA		AAA	lys	AGA		A	
		AUG		ACG		AAG		AGG		G	
	G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U	
		GUC		GCC		GAC		GGC		C	
		GUA		GCA		GAA	glu	GGA		A	
		GUG		GCG		GAG		GGG		G	



Aminoácido-codón

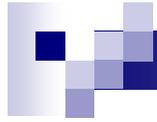
Inverse table

Ala/A	GCU, GCC, GCA, GCG	Leu/L	UUA, UUG, CUU, CUC, CUA, CUG
Arg/R	CGU, CGC, CGA, CGG, AGA, AGG	Lys/K	AAA, AAG
Asn/N	AAU, AAC	Met/M	AUG
Asp/D	GAU, GAC	Phe/F	UUU, UUC
Cys/C	UGU, UGC	Pro/P	CCU, CCC, CCA, CCG
Gln/Q	CAA, CAG	Ser/S	UCU, UCC, UCA, UCG, AGU, AGC
Glu/E	GAA, GAG	Thr/T	ACU, ACC, ACA, ACG
Gly/G	GGU, GGC, GGA, GGG	Trp/W	UGG
His/H	CAU, CAC	Tyr/Y	UAU, UAC
Ile/I	AUU, AUC, AUA	Val/V	GUU, GUC, GUA, GUG
START	AUG	STOP	UAG, UGA, UAA



Algunos datos ...

- En los humanos el DNA cromosómico es de 3.000 millones de bp (pares de bases)
- Contiene relativamente poca información (10-20%)
- Se dice que se ha secuenciado el DNA de un individuo cuando se conocen las zonas que se consideran relevantes (genes y otras)
- También hay DNA mitocondrial ...



Técnicas experimentales

- Leer la sección 2.4



Técnicas experimentales

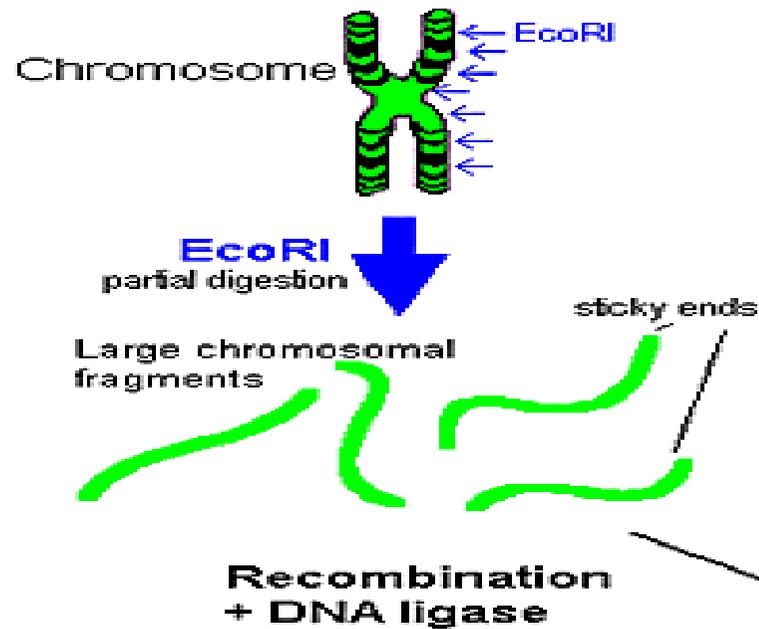
- “Denaturation”: separar las dos hebras del DNA por calor
- Hibridización: asociar bases complementarias o hebras complementarias
- Se puede cortar una doble hebra por un sitio concreto (enzimas de restricción)
- Se pueden reunificar después los trozos (ligasa)



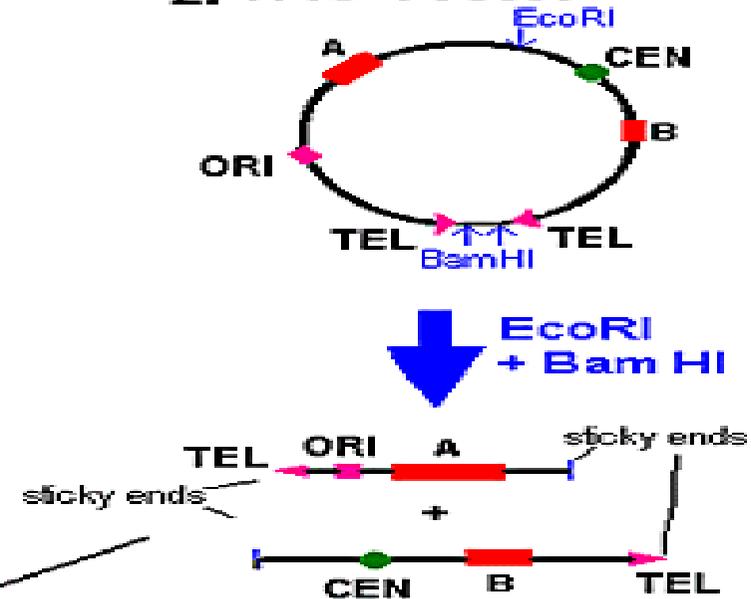
Duplicación de DNA: cloning

- Para los experimentos una sola molécula de DNA no es suficiente, es necesario un gran número de copias idénticas
- Cloning: se inserta el fragmento a copiar en un organismo “host” (anfitrión) , se replica con la reproducción natural y luego se vuelve a extraer
- Los “host” son muy variados, pueden copiar desde 15-50 kbp (bacteria) a varios millones (inserción de cromosomas artificiales)

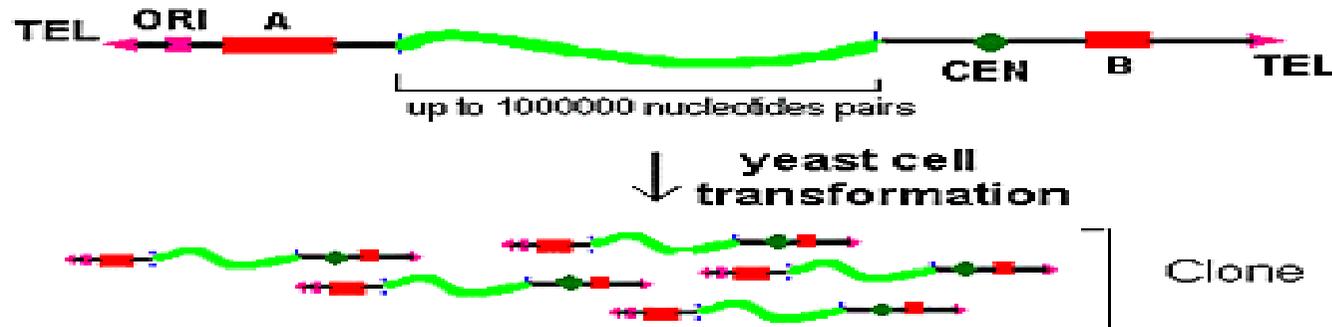
1. Human DNA



2. YAC vector



3. Yeast artificial chromosome with inserted human DNA



Cloning into a Yeast Artificial Chromosome (YAC)



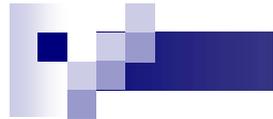
Problemas del clonning

- Contaminación con el DNA del host
- Pérdida de fragmentos completos, cuando la inserción tiene efectos letales en el host
- Dos fragmentos no consecutivos pueden unirse en la clonación (clon quimérico)

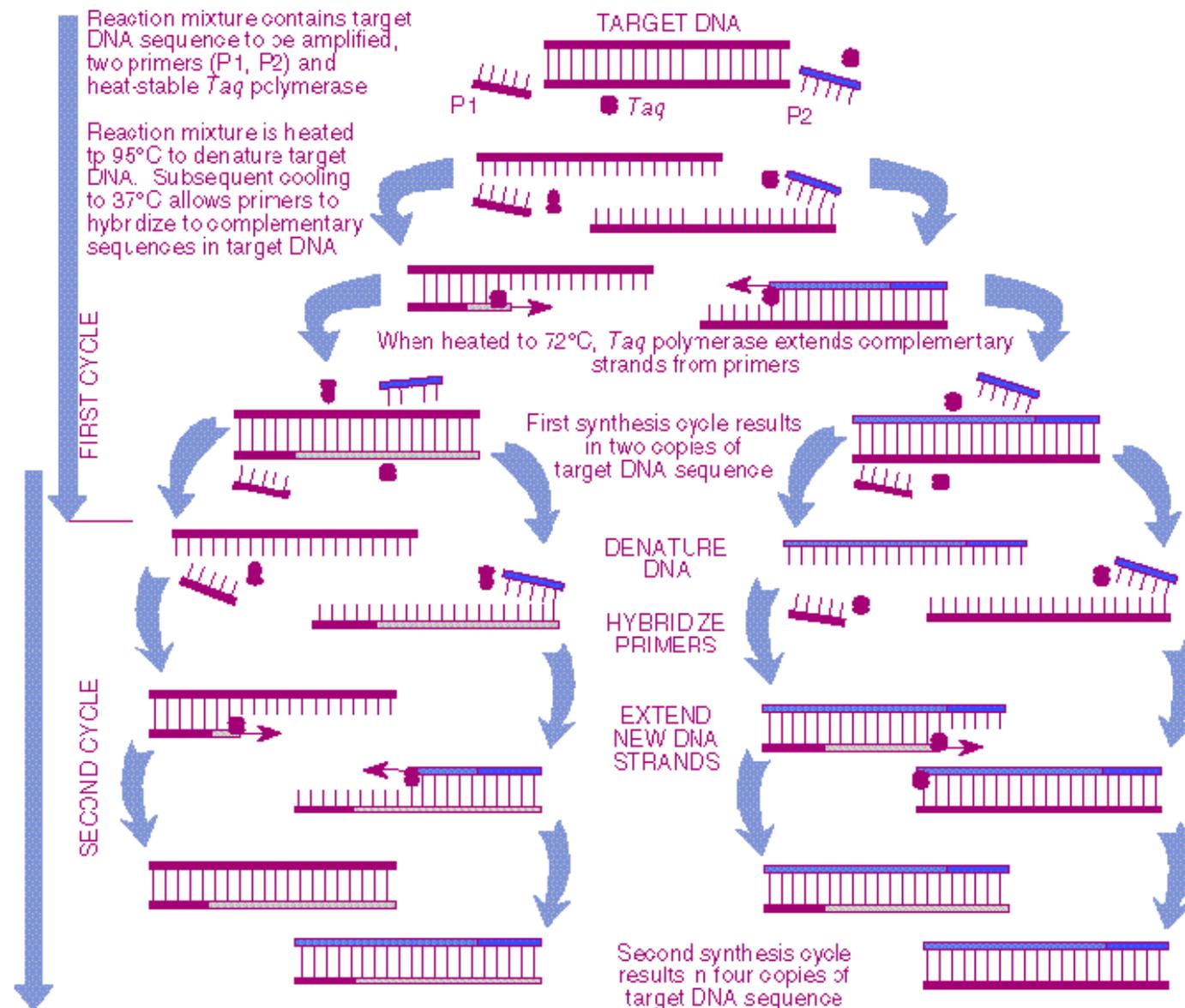


Duplicación de DNA: PCR

- Polymerase chain reaction
- Necesitamos conocer un fragmento inicial y otro final
- En cada paso duplica el número de copias
- Los errores al principio son muy peligrosos ...



DNA Amplification Using Polymerase Chain Reaction



Source: *DNA Science*, see Fig. 13.



Gel electrophoresis

- Se trata de separar los fragmentos por tamaño
- Se meten en gel y se aplica un campo eléctrico, la velocidad es inversamente proporcional al tamaño
- Se separan así por longitudes
- Usando trozos de referencia se puede usar para medir la longitud

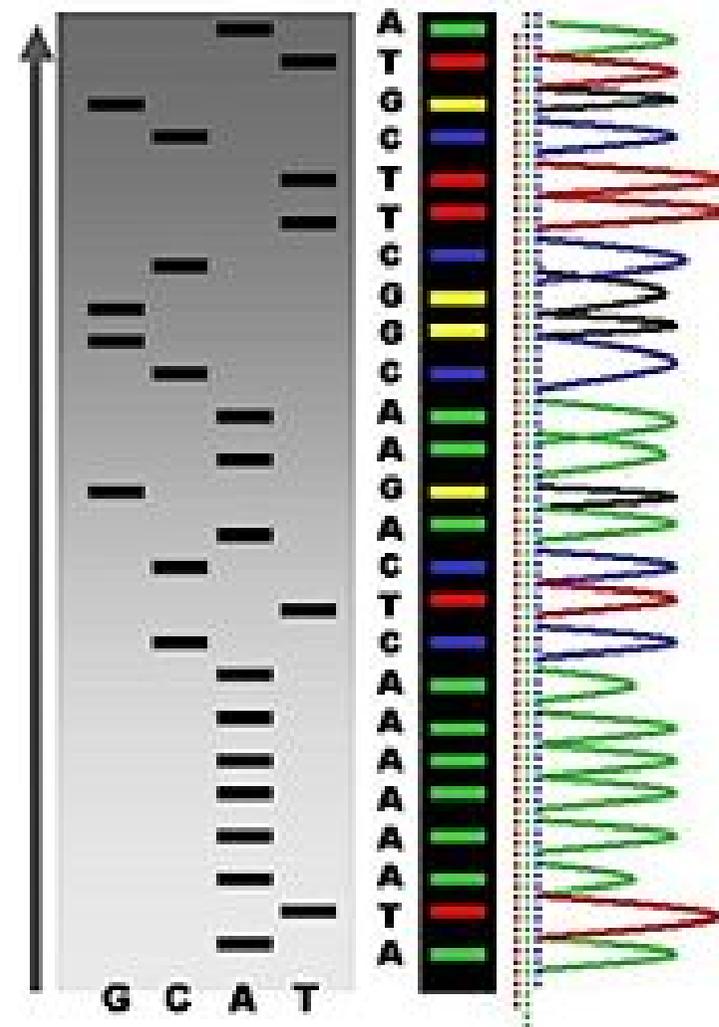


Secuenciar DNA: chain termination method

- El “chain termination method” se basa en el anterior (gel electrophoresis)
- Tenemos un fragmento de DNA desconocido s, hacemos muchas copias
- Paso 1: conseguir que haya 4 tubos de ensayo A, C, G, T cada uno conteniendo los prefijos de s que terminan en A (C,G,T)

chain termination method (2)

- Paso 2: Colocamos los cuatro tubos de ensayo en paralelo y ordenamos por longitud como antes
...





Chain termination method (3)

- Sólo sirve para fragmentos de hasta 1000 bp (más da demasiados errores)
- Puede dar errores de lectura del resultado (llamados errores de secuenciación):
 - Inserción
 - Borrado
 - Sustitución



Experimentos de hibridización

- Para averiguar si un fragmento desconocido s contiene una secuencia t
- Sintetizamos t' la complementaria de t
- Testeamos si s y t' se unen (hibridizan)



DNA chips

- Para hacer varios experimentos de hibridización en paralelo
- Si queremos saber si s contiene t_1, \dots, t_n :
 - Colocamos t'_1, \dots, t'_n en sitios fijos (DNA chip)
 - Hacemos copias etiquetadas de s
 - Dejamos que se unan al chip
 - Lavamos las copias de s sueltas y averiguamos las posiciones de hibridización con las etiquetas



DNA chips: errores

- Falsos positivos
- Falsos negativos

- Se pueden usar los DNA chips para RNA
- ...



Premios nóbel

- Watson, Crick y Wilkins por el descubrimiento de la estructura del DNA, Medicina 1962
- Mullis por el método PCR, Química 1993



El próximo tema ...

- Comparando DNA: Alineamiento