

SEQUENCAGE DE L'ADN ET ORGANISATION DES GENOMES

Nicolas DUBOIS – PRAG USMB – UFR Sciences et Montagne
nicolas.dubois@univ-smb.fr

QUELQUES DEFINITIONS POUR DEMARRER...

Gène : Fragment d'ADN portant les informations nécessaires à la fabrication d'une ou plusieurs protéines.

Génome : Ensemble de l'information génétique d'un individu. Une copie du génome est présente dans chacun des cellules de l'individu. Le génome se transmet de génération en génération. Par extension, le génome se réfère aussi au support physique de cette information génétique, c'est-à-dire la macromolécule d'ADN.

Génomique : Etude des génomes. Son objectif est de séquencer l'ADN d'un organisme et de localiser sur celui-ci tous les gènes qu'il porte, puis de caractériser leurs fonctions.

LES PROGRES DU SEQUENCAGE

- 1980-1995: séquencer pour répondre à une question donnée: de la biologie à la séquence
 - séquenceurs: tous les laboratoires de biologie moléculaire
 - séquences: des gènes ou des ARNm (< 10 kb)
 - informations biologiques associées aux séquences: riches

phénotype => gène

- >1995: séquençage systématique à grande échelle: de la séquence à la biologie
 - séquenceurs: quelques grands centres de séquençage
 - séquences: grands fragments génomiques, chromosomes, etc ...
 - informations biologiques associées aux séquences: pauvres

gène => phénotype

LES DEFIS AU SEQUENCAGE DE L'ADN : LA TAILLE

SPECIES	CHROMOSOMES	GENES	BASE PAIRS
Human (<i>Homo sapiens</i>)	46 (23 pairs)	28-35,000	~3.1 billion
Mouse (<i>Mus musculus</i>)	40	22.5-30,000	~2.7 billion
Pufferfish (<i>Fugu rubripes</i>)	44	~31,000	~365 million
Malaria Mosquito (<i>Anopheles gambiae</i>)	6	~14,000	~289 million
Sea Squirt (<i>Ciona intestinalis</i>)	28	~16,000	~160 million
Fruit Fly (<i>Drosophila melanogaster</i>)	8	~14,000	~137 million
Roundworm (<i>C. elegans</i>)	12	19,000	~97 million
Bacterium (<i>E. coli</i>)	1*	~5,000	~4.1 million

*Bacterial chromosomes are **chromonemes**, not true chromosomes.

LES DEFIS ... : L'ADN REPETE

⇒ Mesure des cinétiques de réassociation de l'ADN

Après avoir dénaturé l'ADN génomique d'un organisme, on le place dans des conditions ménagées de refroidissement afin de favoriser la renaturation. On mesure la fraction d'ADN non encore renaturé en fonction du temps.

Sur l'axe des Y on porte le pourcentage d'ADN qui reste simple brin par rapport à la concentration totale d'ADN. L'axe des X porte une échelle logarithmique du produit de la concentration initiale de l'ADN (en moles/litre) par le temps écoulé (en secondes).

$$dC/dt = -kC^2$$

C : concentration des séquences simple brin au temps t

k : constante d'association

LES DEFIS ... : L'ADN REPETE

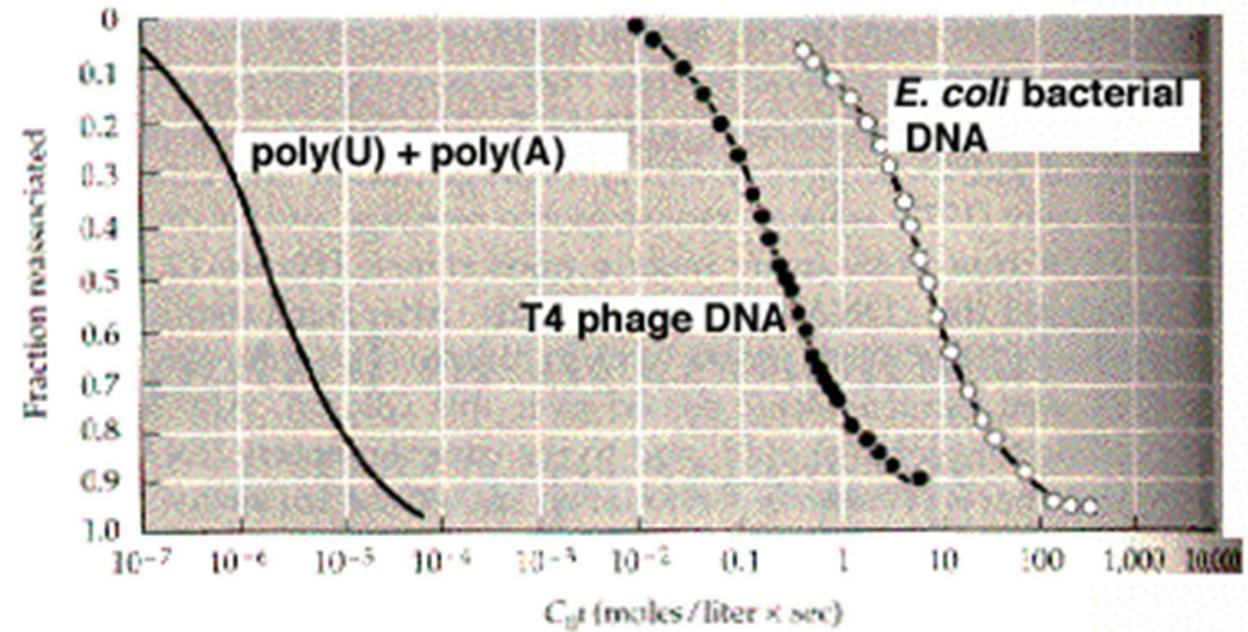
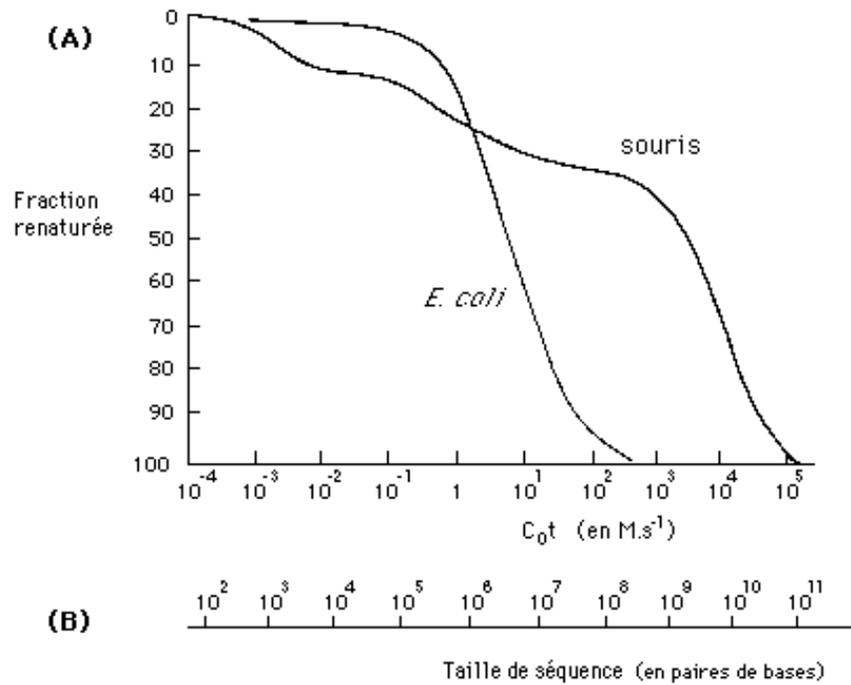
Ces courbes sont appelées **courbes de Cot**.

Une courbe lisse est l'indice d'une renaturation progressive et régulière de l'ADN. C'est ce qui est observé pour des virus, des bactéries et pour des cellules de levure. Par contre lorsque l'on prend de l'ADN d'organismes multicellulaires (*Homo sapiens*), l'aspect est totalement différent. La courbe observée semble le résultat de la mise bout à bout de trois courbes.

Ce qui indique que dans le premier cas il y avait un seul type d'ADN alors que dans le second trois types d'ADN différents doivent coexister : le premier type se renature très rapidement, le second se renature nettement plus lentement, quant au troisième il ne se renature que très lentement. **L'interprétation de cette observation, est que l'ADN qui se renature rapidement n'a aucun mal à retrouver un brin qui lui soit complémentaire parce qu'il contient une (ou des) séquence très fréquente, alors que le troisième type a beaucoup de mal à retrouver son complément parce que cette séquence est " unique ".**

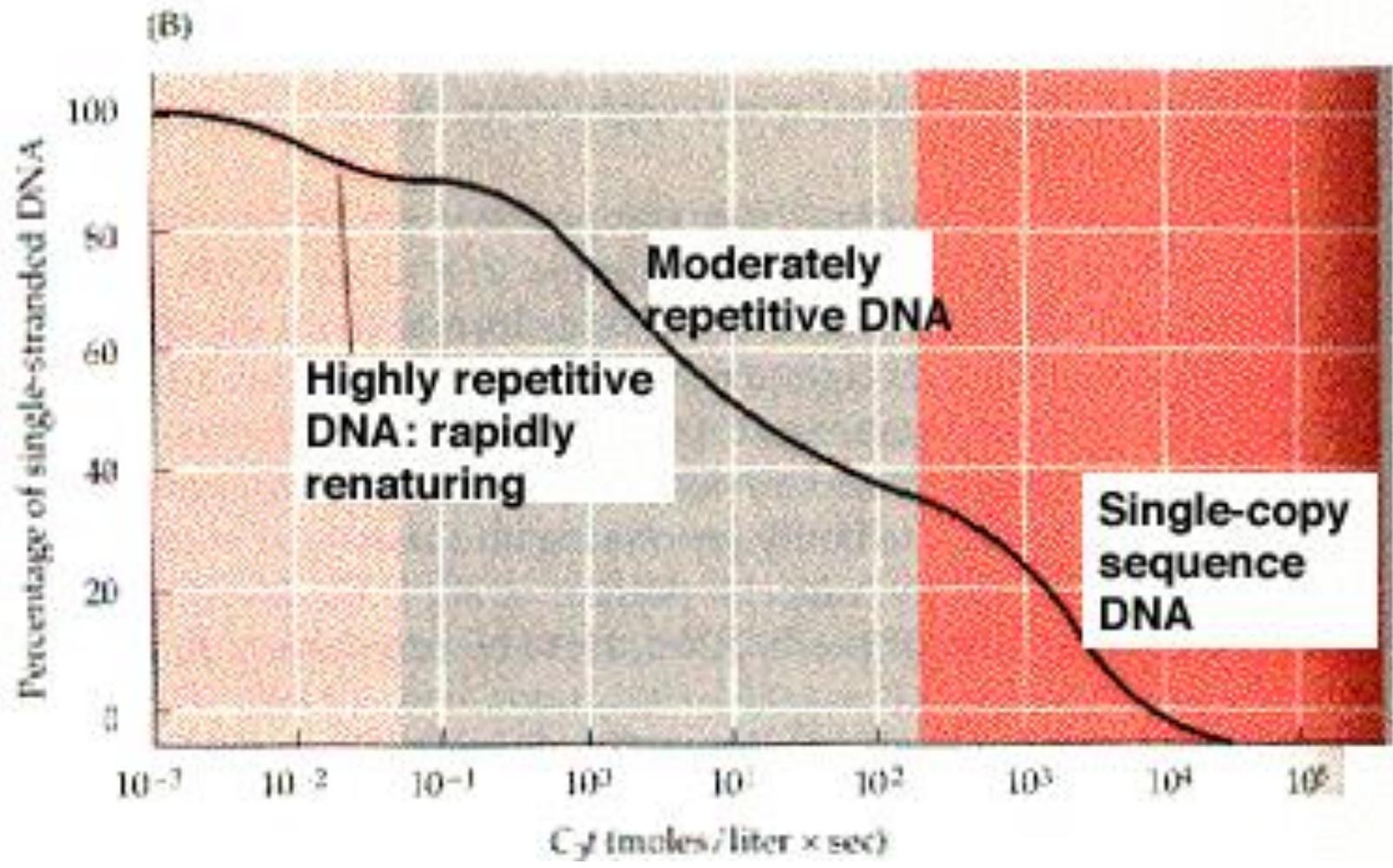
LES DEFIS ... : L'ADN REPETE

⇒ Mesure des cinétiques de réassociation de l'ADN



LES DEFIS ... : L'ADN REPETE

⇒ Mesure des cinétiques de réassociation de l'ADN

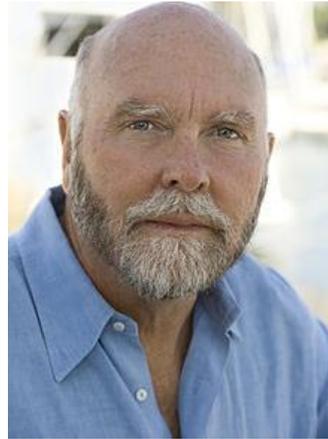


LES PROGRES DU SEQUENCAGE

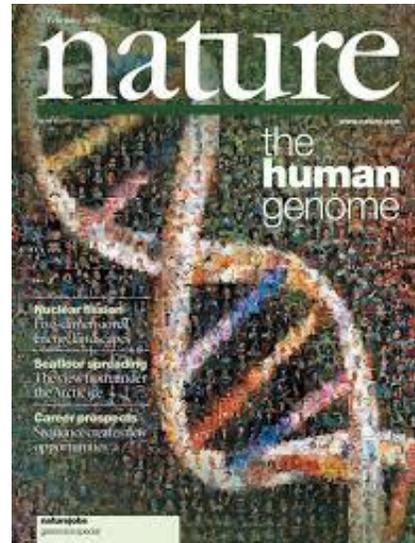
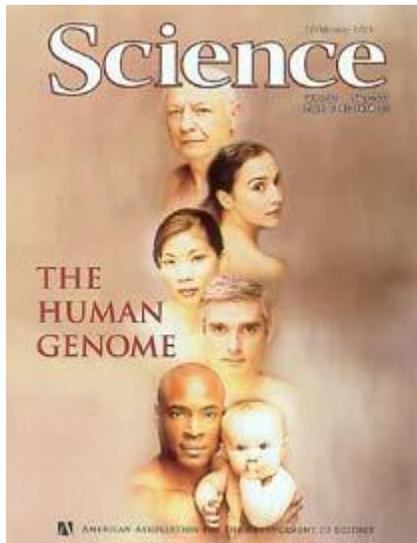
La liste qui suit donne quelques exemples de génomes séquencés parmi les premiers qui le furent.

- **Virus** : 3 778 virus séquencés au 4 mars 2014, dont le VIH (virus d'immunodéficience humaine).
- **Procaryotes** :
 - Archaeobactéries : 319 génomes entièrement séquencés et 447 partiellement au 4 mars 2014.
 - Eubactéries : 12 286 génomes entièrement séquencés et 20 403 partiellement au 4 mars 2014. Exemples :
 - *Escherichia coli*
 - *Agrobacterium tumefaciens*
 - *Haemophilus influenzae* Rd. (premier génome cellulaire séquencé, en 1995)
- **Eucaryotes** : 314 génomes entièrement séquencés et 6660 partiellement au 4 mars 2014. Exemples (les 5 premiers cités sont les 5 premiers publiés, avec la réserve citée pour celui de l'Homme) :
 - *Saccharomyces cerevisiae* (levure, premier eucaryote à être séquencé en 1997. Plusieurs souches séquencées de nos jours)
 - *Caenorhabditis elegans* (ver nématode)
 - *Drosophila melanogaster* (mouche du vinaigre)
 - *Arabidopsis thaliana* (arabette - petite plante de la famille du chou)
 - *Homo sapiens* (nous... Plusieurs individus séquencés dont Watson)
 - *Neurospora crassa* (champignon ascomycète)
 - *Anopheles gambiae* (moustique)
 - *Takifugu rubripes* (fugu : poisson-ballon consommé au japon)
 - *Mus musculus* (souris)
 - *Plasmodium falciparum* (parasite intracellulaire responsable du paludisme)
 - *Oriza sativa* (riz - deux sous-espèces séquencées, *japonica* et *indica*)

LE PROJET GENOME HUMAIN



Consortium public vs **Craig Venter** (Celera Genomics)



Chronologie

- 1977 Méthode de séquençage de l'ADN
- 1985 Naissance du Projet
- 1990 Création d' HUGO. Début du séquençage
- 1996 Carte génétique
- 1998 Carte physique
- 1999 Chromosome 22 terminé
- 2001 Brouillon du génome (~90%)
- 2003 >99% achevé.

L'annotation continue...

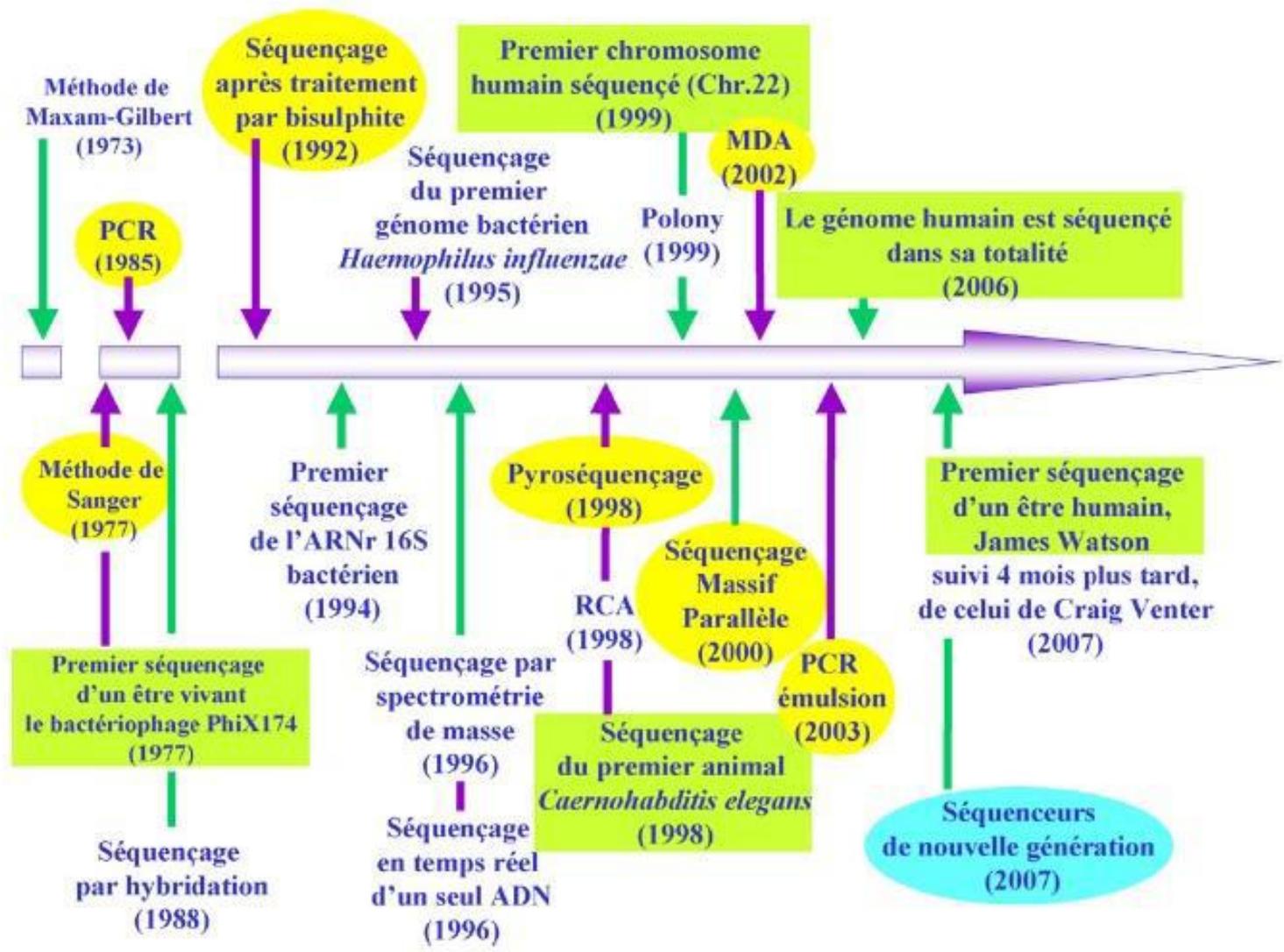
Séquençage achevé en 2004 !

LE PROJET GENOME HUMAIN

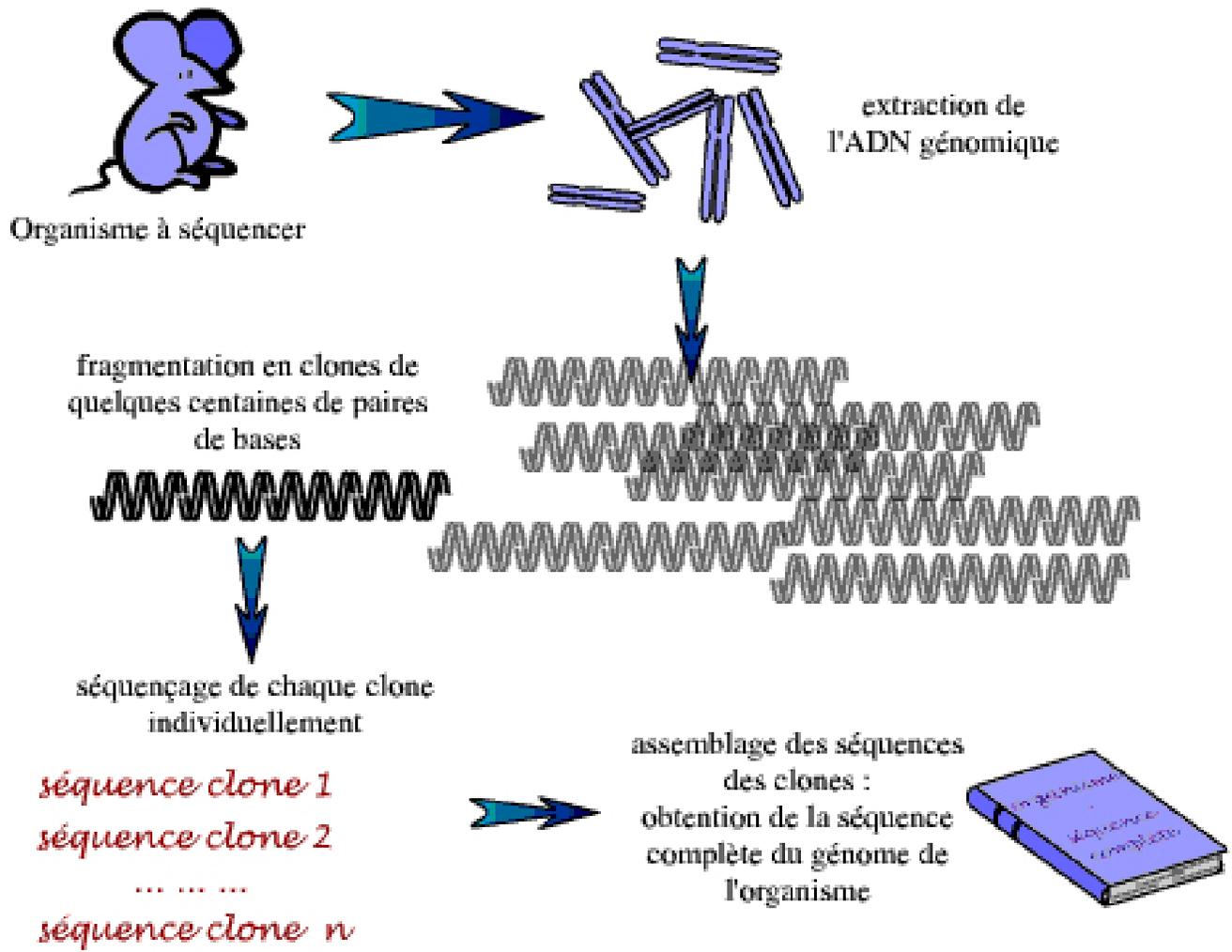
Remarque : On parle de « partie séquençable » du génome humain (2,9 Gb, pour un total de 3,2 Gb). En effet, il n'est pas possible, techniquement, de déterminer la séquence de certaines régions presque exclusivement constituées de séquences répétées, telles que les centromères, les télomères ou les bras courts de certains chromosomes, et cela pour deux raisons : d'une part, parce qu'on éprouve des difficultés à isoler des fragments d'ADN de taille convenable issus de ces régions ; d'autre part, parce qu'on échoue à reconstituer la séquence complète à partir de morceaux de séquences pratiquement identiques. De ce fait, seule la séquence de la partie dite euchromatique du génome peut être effectivement déterminée.

1. SEQUENCAGE DE L'ADN

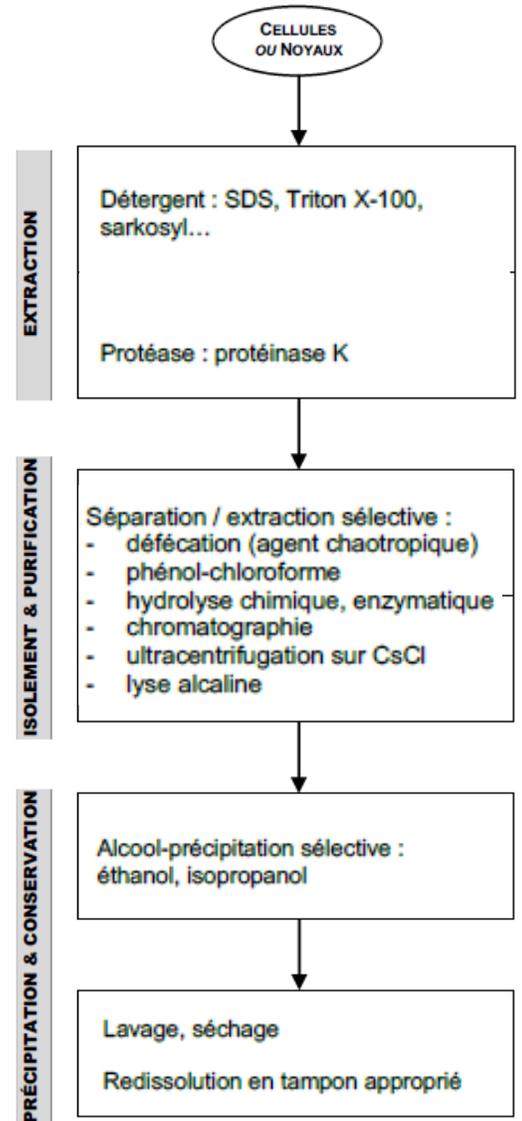
UNE CHRONOLOGIE...



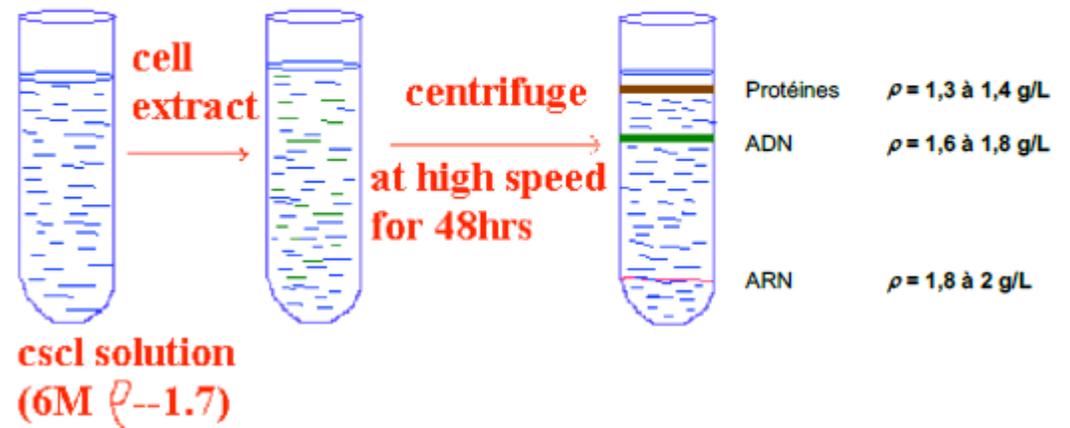
PRINCIPE GENERAL DU SEQUENCAGE D'UN GENOME



EXTRACTION DE L'ADN

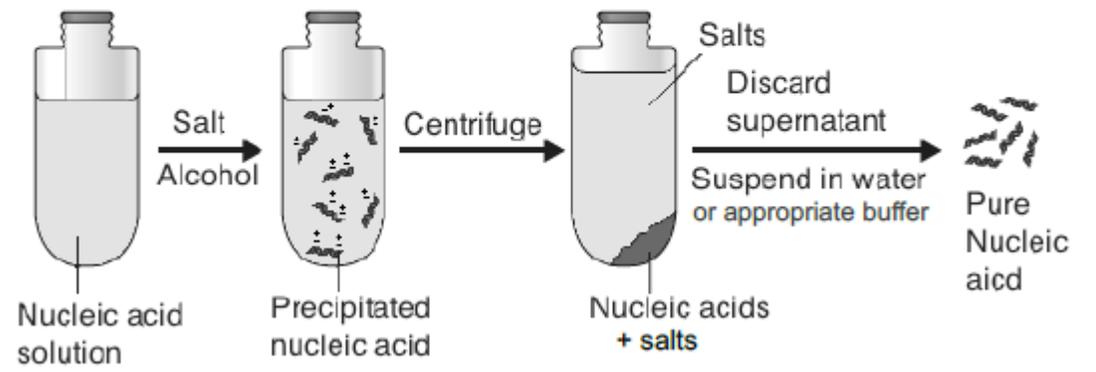


Doc 3 : ULTRACENTRIFUGATION SUR CHLORURE DE CÉSIIUM



Doc 4 : (Co-)PRÉCIPITATION À L'ÉTHANOL EN PRÉSENCE DE CATIONS

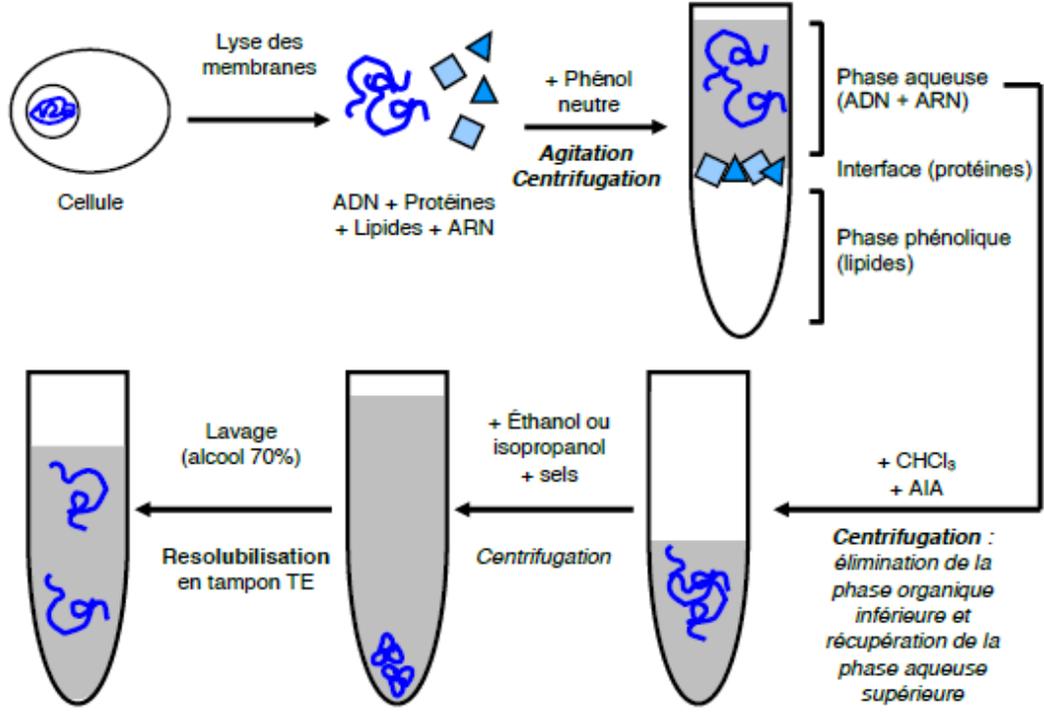
4a. PRINCIPE GÉNÉRAL



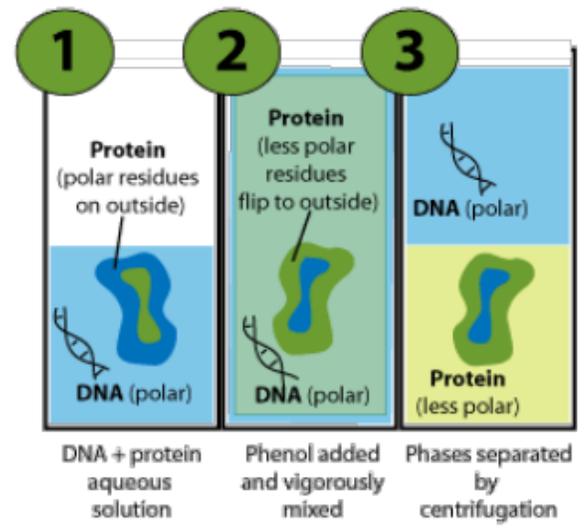
EXTRACTION DE L'ADN

Doc 5 : PURIFICATION DE L'ADN CHROMOSOMIQUE AU PHÉNOL-CHLOROFORME

5a. VUE D'ENSEMBLE



5b. DÉTAIL DE L'ÉTAPE « PHÉNOL »



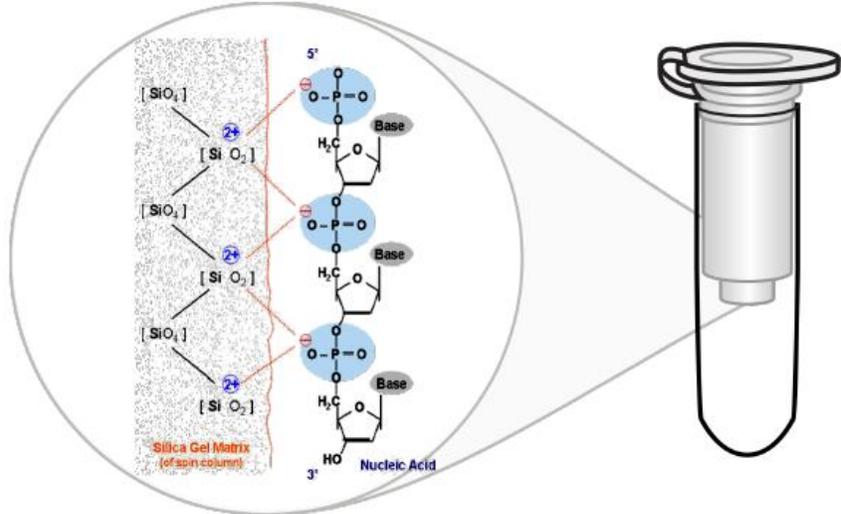
En fait, la majorité des protéines va se concentrer en bas de la phase aqueuse, juste à l'interface avec la phase phénolique inférieure.

Les protéines hydrophobes se retrouvent elles dans la phase inférieure, avec le phénol et les lipides.

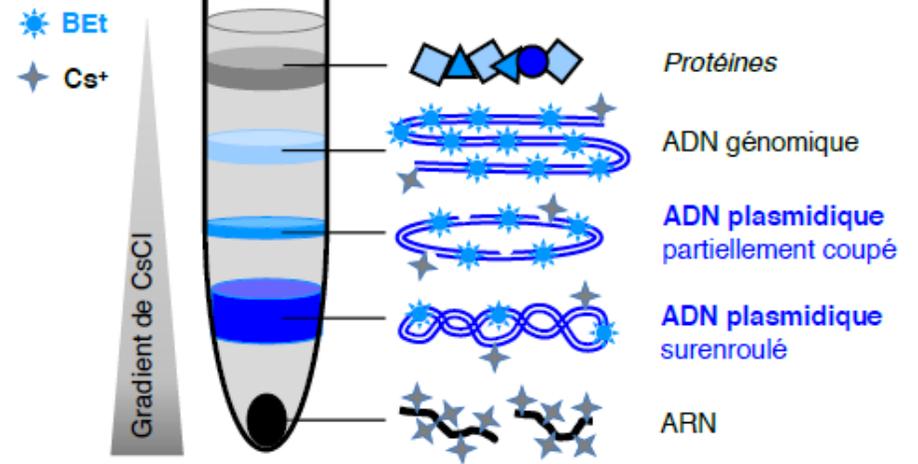
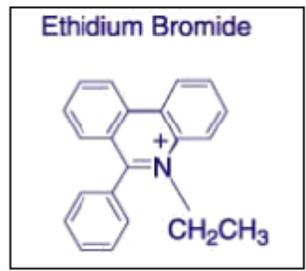
EXTRACTION DE L'ADN

Doc 6 : PURIFICATION DE L'ADN PAR CHROMATOGRAPHIE

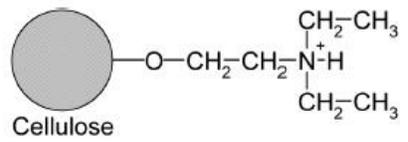
6a. ADSORPTION SUR SILICE (EX DE MINI-COLONNE)



Doc 9 : PURIFICATION DE L'ADN PLASMIDIQUE PAR ULTRACENTRIFUGATION SUR CsCl EN PRÉSENCE DE BET

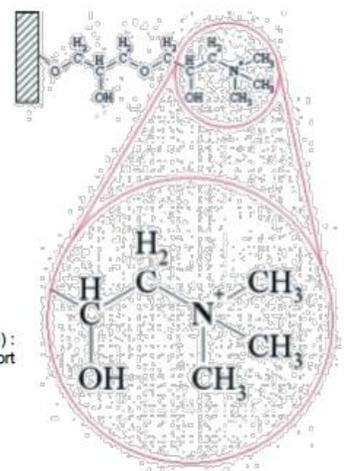


6b. ÉCHANGEUSE D'ANIONS



DEAE-cellulose : échangeur faible

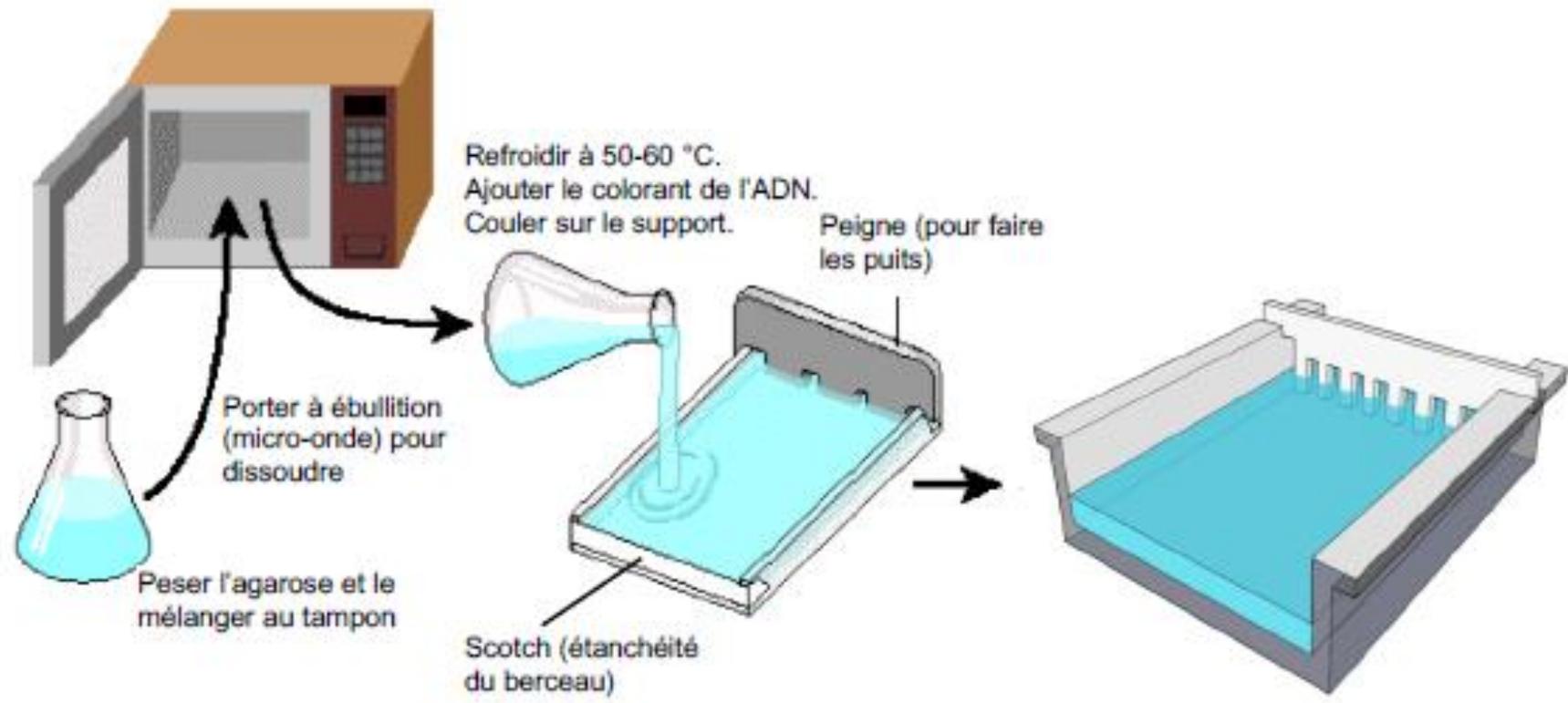
Q-sépharose (ammonium quaternaire) : échangeur fort



PRINCIPE DE L'ELECTROPHORESE

Doc 13 : ÉLECTROPHORÈSE D'ACIDES NUCLÉIQUES SUR GEL D'AGAROSE

13a. PRÉPARATION D'UN GEL D'AGAROSE IMMERGÉ HORIZONTAL



PRINCIPE DE L'ELECTROPHORESE



Download from [Dreamstime.com](https://www.dreamstime.com)
This watermarked comp image is for previewing purposes only.

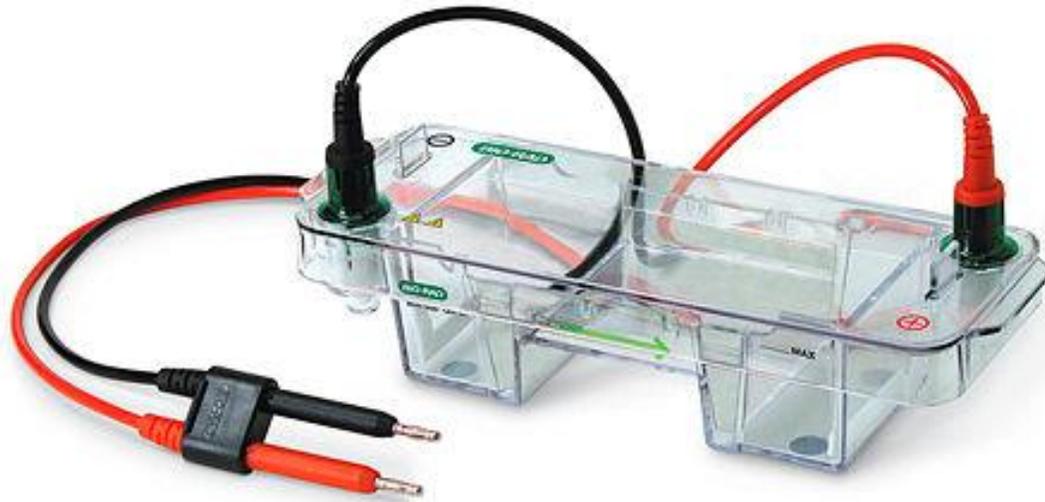
ID 2223462
© Paul Hakin



Download from [Dreamstime.com](https://www.dreamstime.com)
This watermarked comp image is for previewing purposes only.

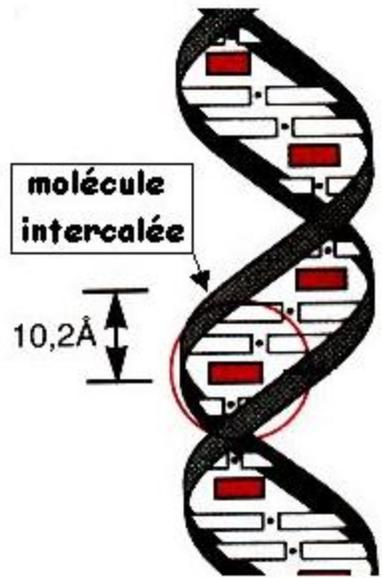
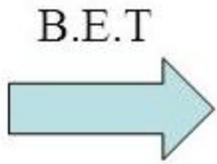
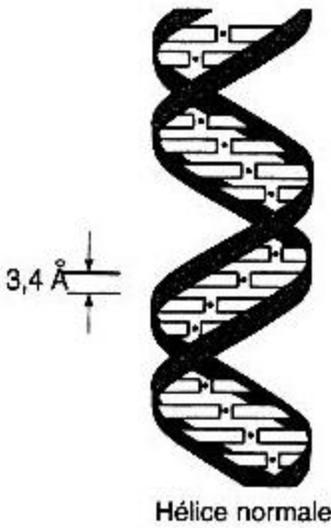
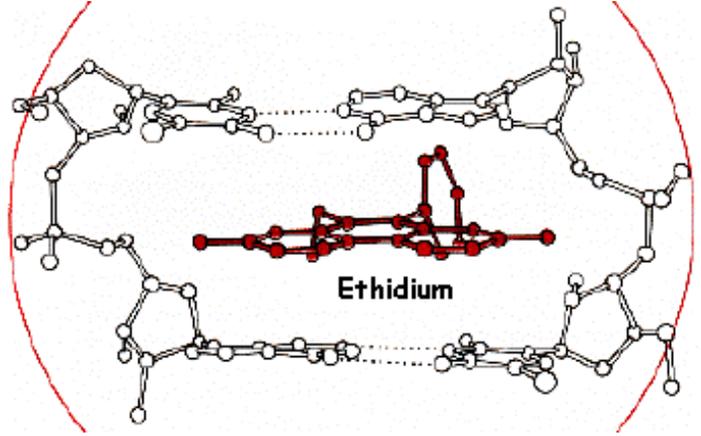
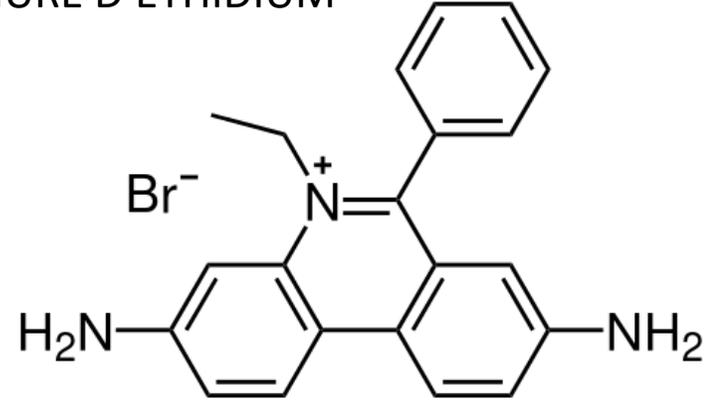
ID 21181684
© Phxchristine | Dreamstime.com

PRINCIPE DE L'ELECTROPHORESE

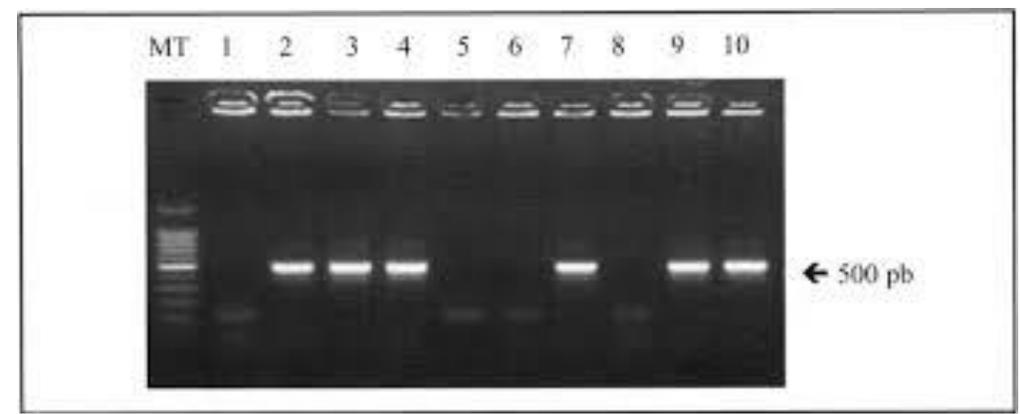
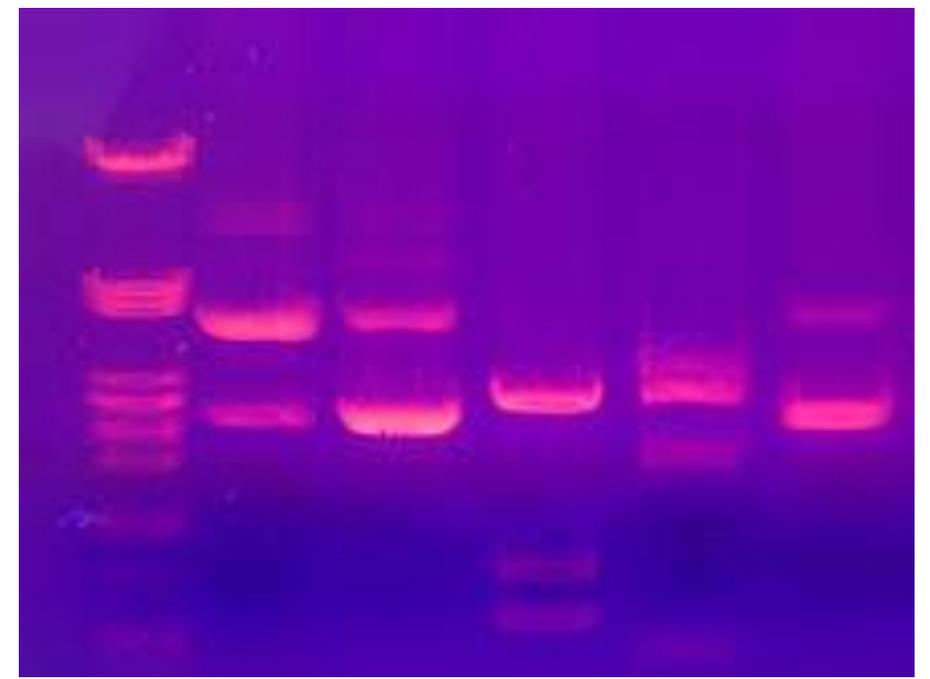
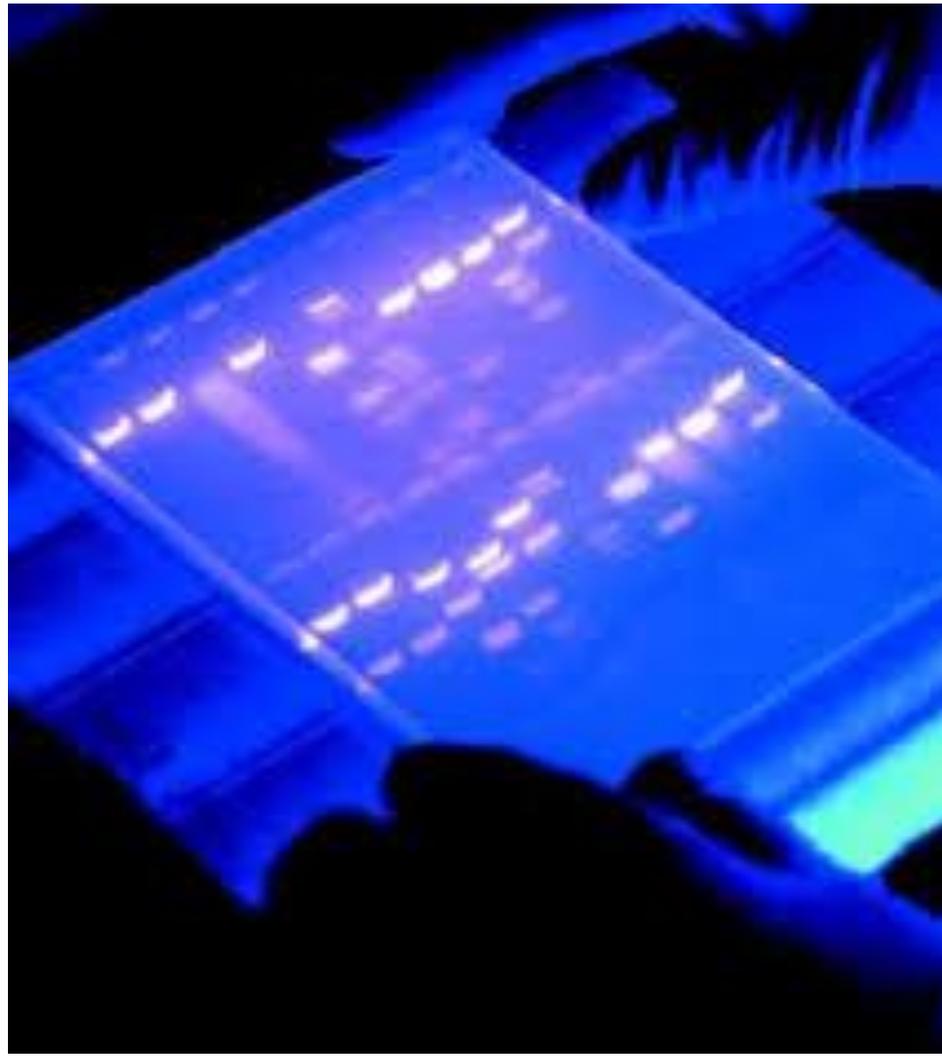


PRINCIPE DE L'ELECTROPHORESE

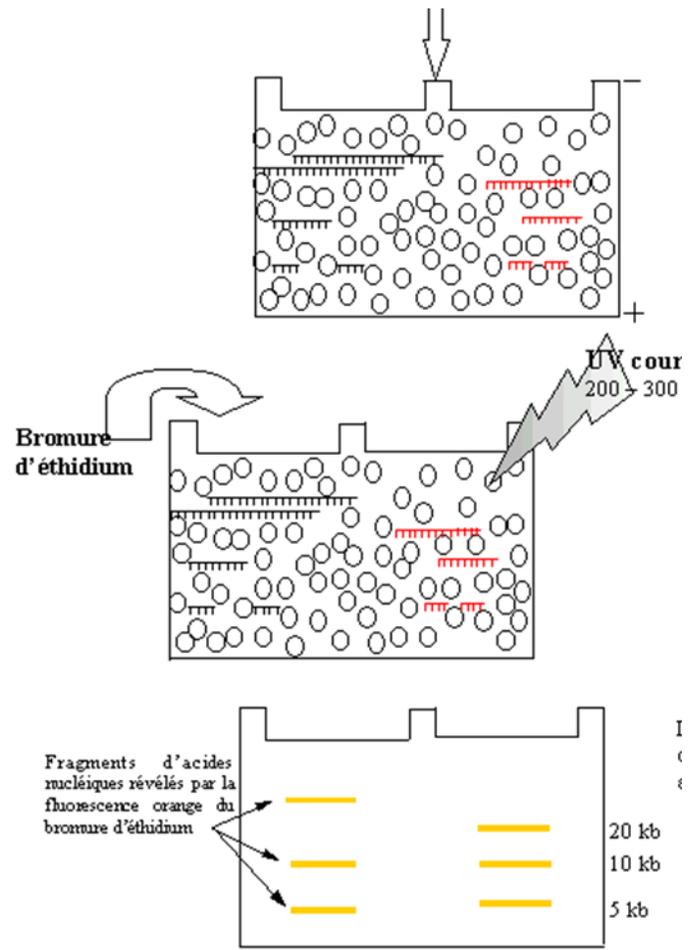
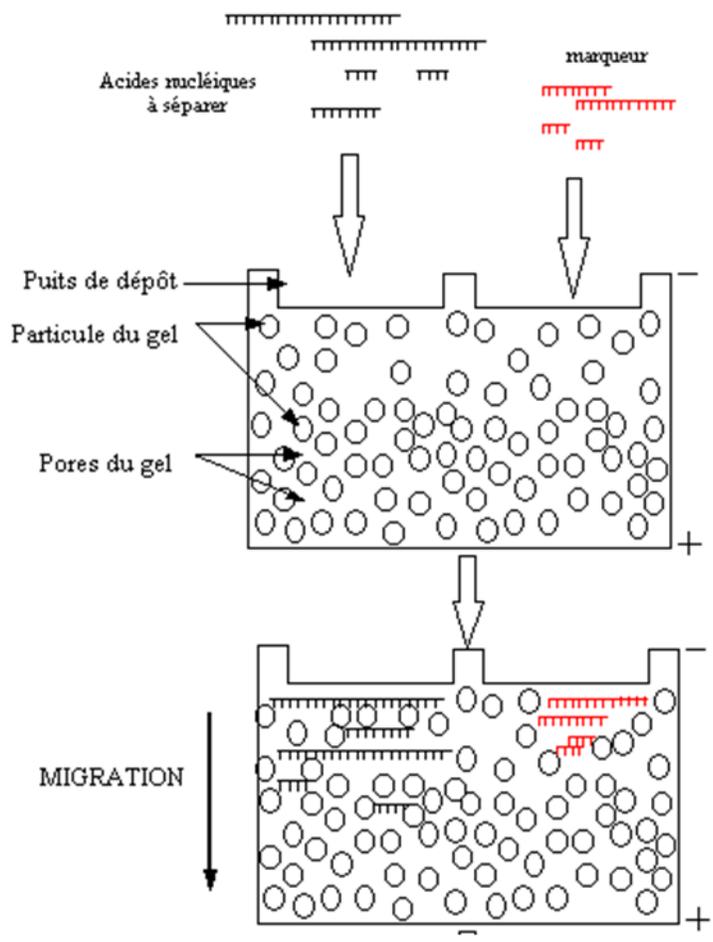
BROMURE D'ETHIDIUM



PRINCIPE DE L'ELECTROPHORESE



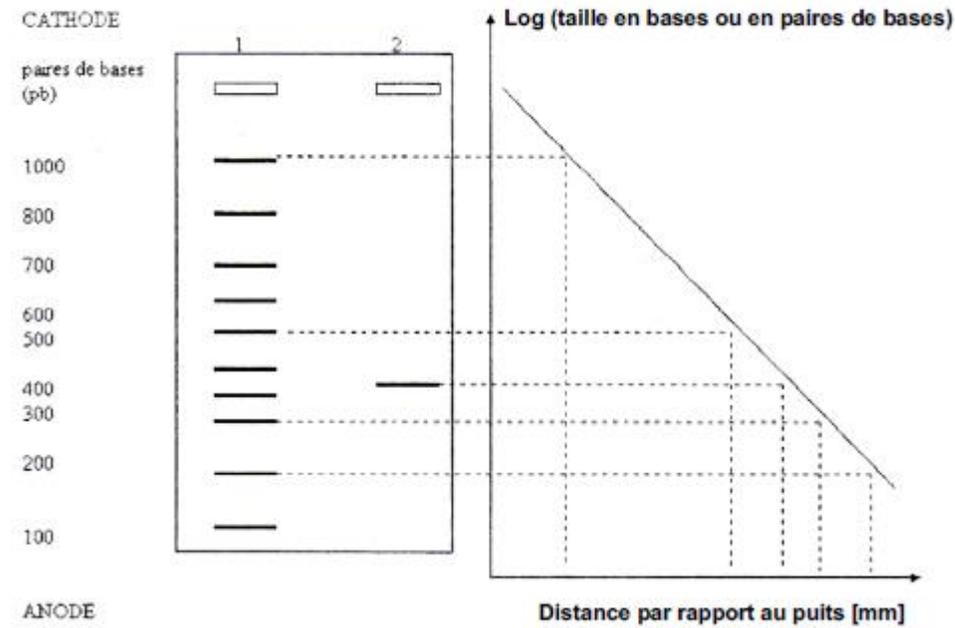
PRINCIPE DE L'ELECTROPHORESE



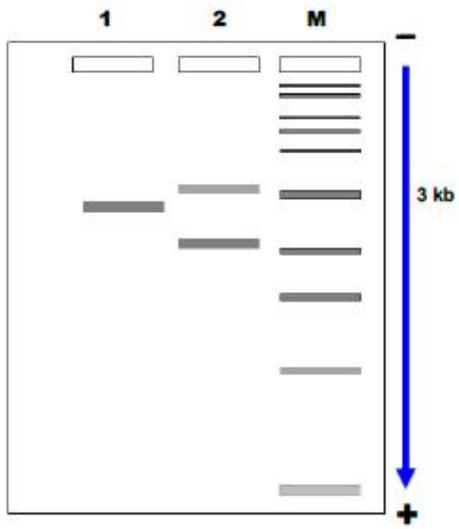
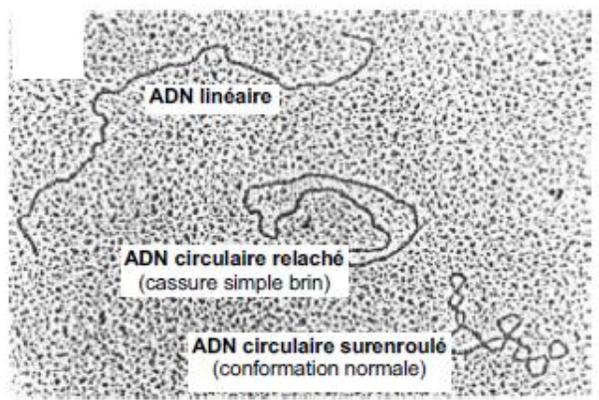
<http://www.inrp.fr/Acces/biotic/biomol/techgen/html/electro.htm>

PRINCIPE DE L'ELECTROPHORESE

13c. ÉTALONNAGE D'UN GEL & DÉTERMINATION D'UNE TAILLE DE FRAGMENT



13d. MIGRATION DES PLASMIDES : LES DIFFÉRENTES FORMES

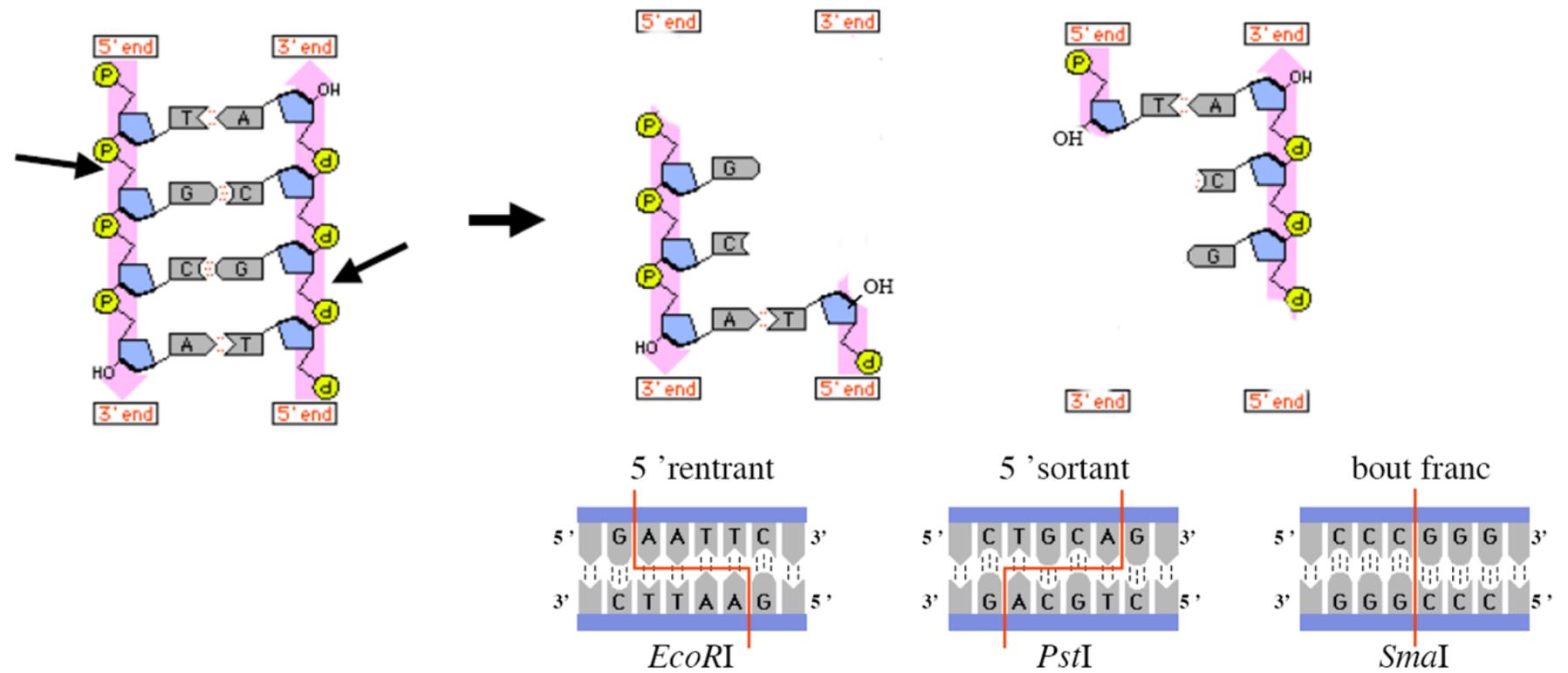


1 : plasmide pUC18 linéarisé (coupure double brin par une enzyme de restriction)
 2 : plasmide pUC18 non linéarisé
 M : marqueur de taille « 1 kb DNA ladder »

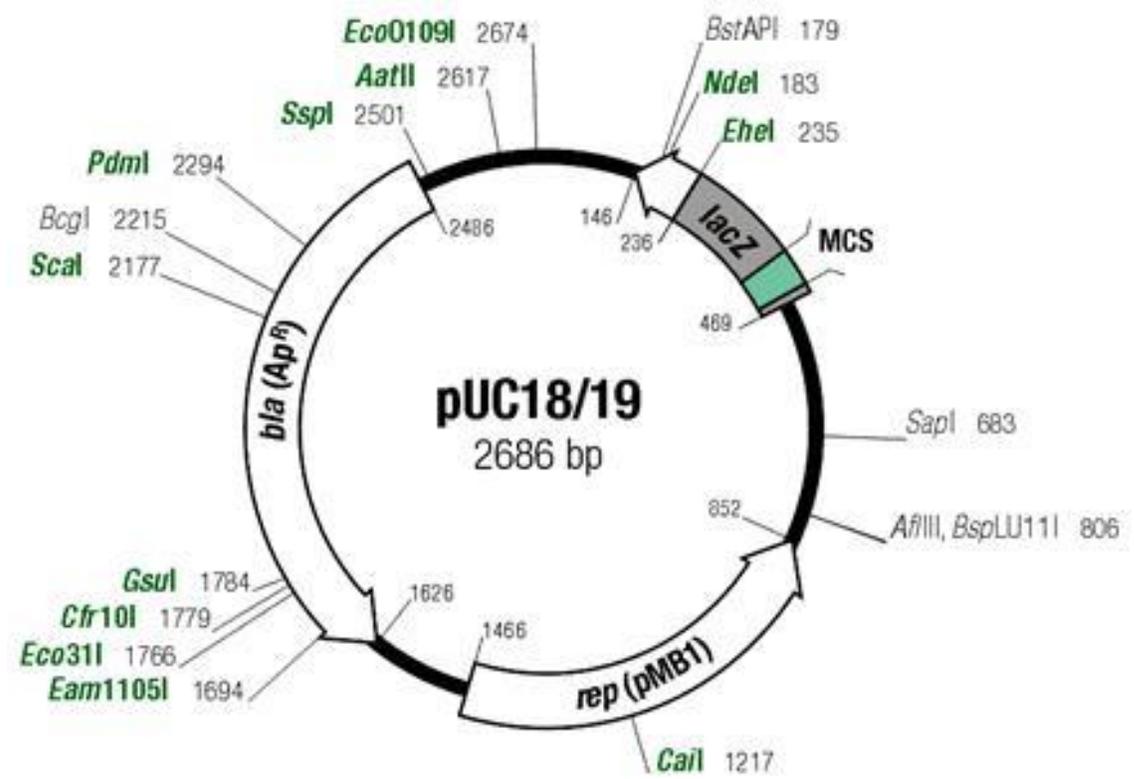
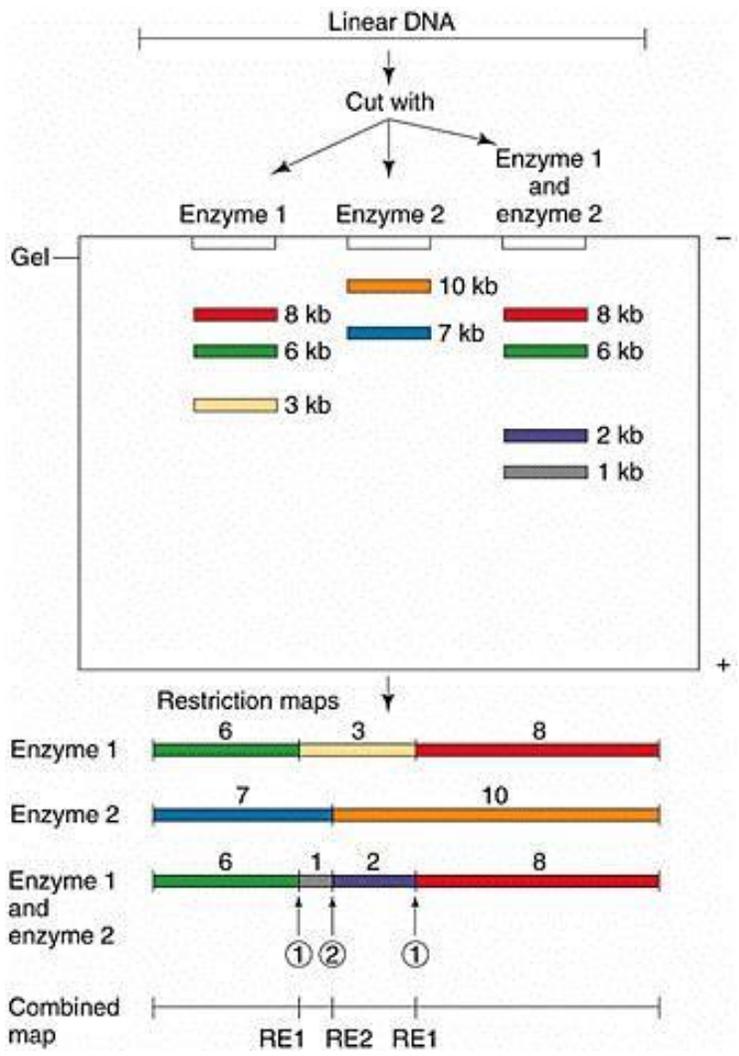
LES ENZYMES DE RESTRICTION

Enzyme de restriction

Enzyme capable de digérer l'ADN (endonucléase) en un endroit présentant une séquence précise, spécifique de l'enzyme. La séquence d'ADN reconnue par l'enzyme de restriction est nommée "site de restriction". Ces sites de restriction peuvent comporter de 2 à une vingtaine de paires de bases. Exemple : l'enzyme EcoR I reconnaît et hydrolyse l'ADN au niveau des séquences GAATTC.



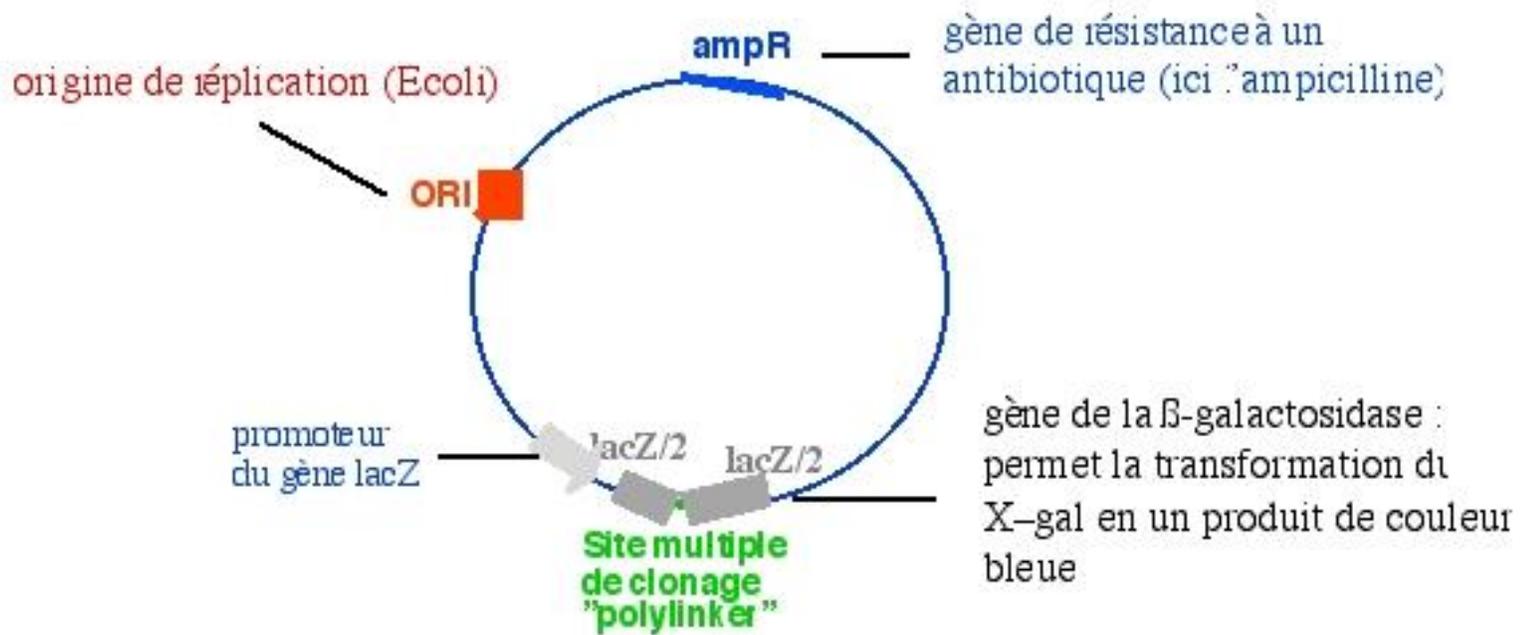
CONSTRUCTION D'UNE CARTE DE RESTRICTION



<http://www.cbs.dtu.dk/staff/dave/roanoke/genetics980211.html>

CLONAGE D'ADN

STRUCTURE D'UN PLASMIDE DESTINÉ AU CLONAGE BACTÉRIEN



GATCGCGTGTGGTGGGATTCCTGCGGATATCCAGGACACGTTGGGCGGCGCTCGAGTCTAGAG

BstI
BamHI

BstXI
RsrI
EcoRI

PstI
EcoRV

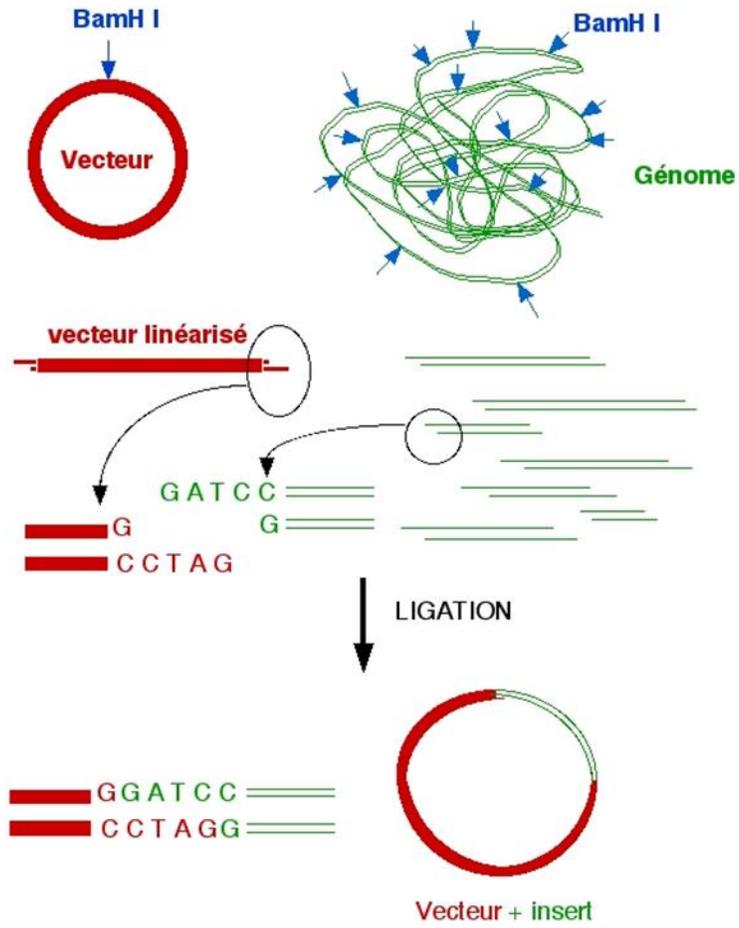
BstXI
NotI
SmaI
EcoS2I
EcoX1
EagI

GdiI
XbaI
PaeR7I
SmaI
PaeR7I
EcoX1
SmaI
EagI
CcoI
PvuI
PvuI
RsaI
Nsp111

<http://www.sesep.uvsq.fr/formation/strucplasm.JPG>

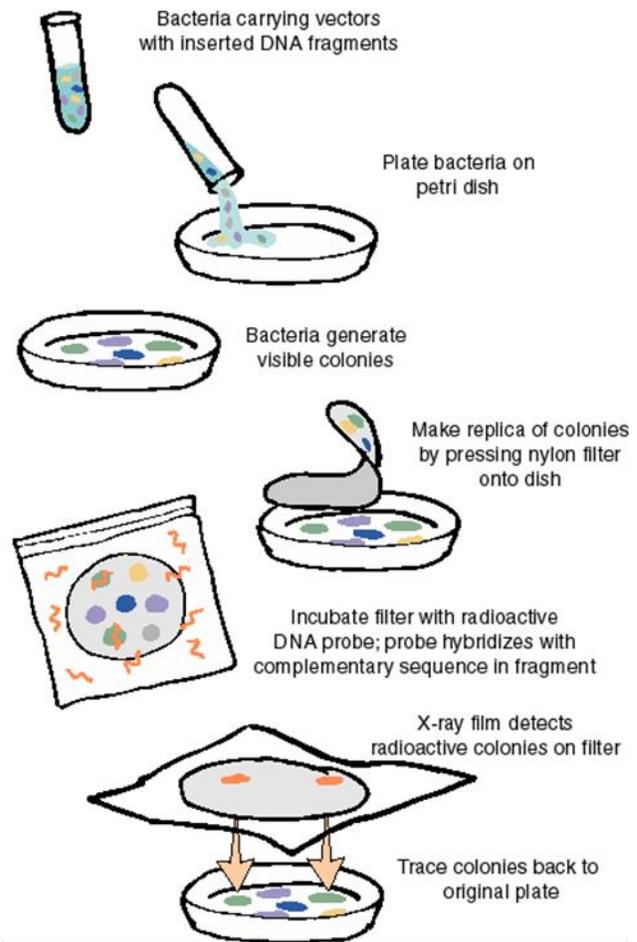
CLONAGE D'ADN

Principe du clonage (cas d'une banque génomique)



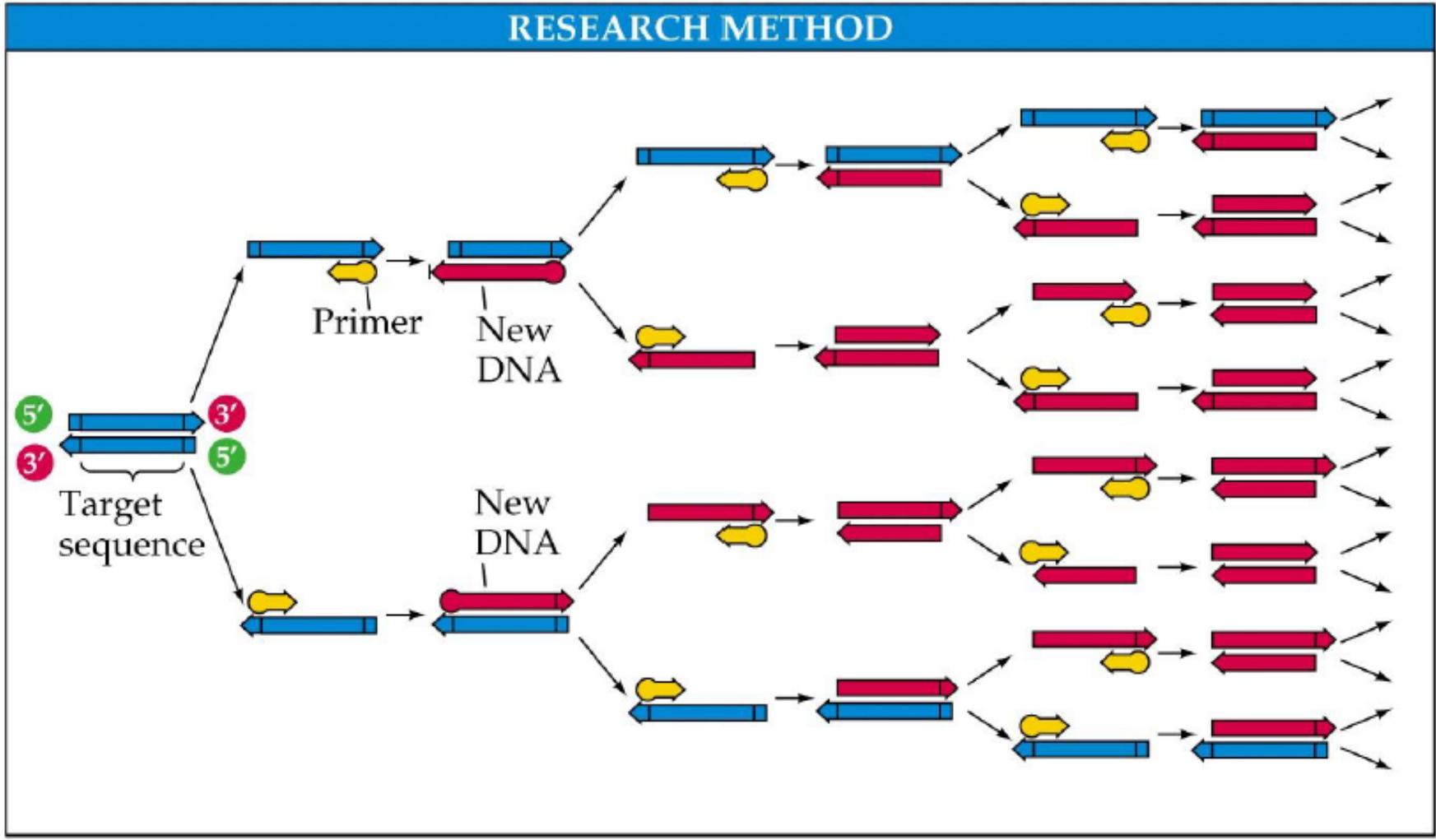
<http://www.sesep.uvsq.fr/formation/princlon1.JPEG>

Screening a DNA library



http://www.fhcrc.org/education/courses/cancer_course/basic/approaches/screening.html

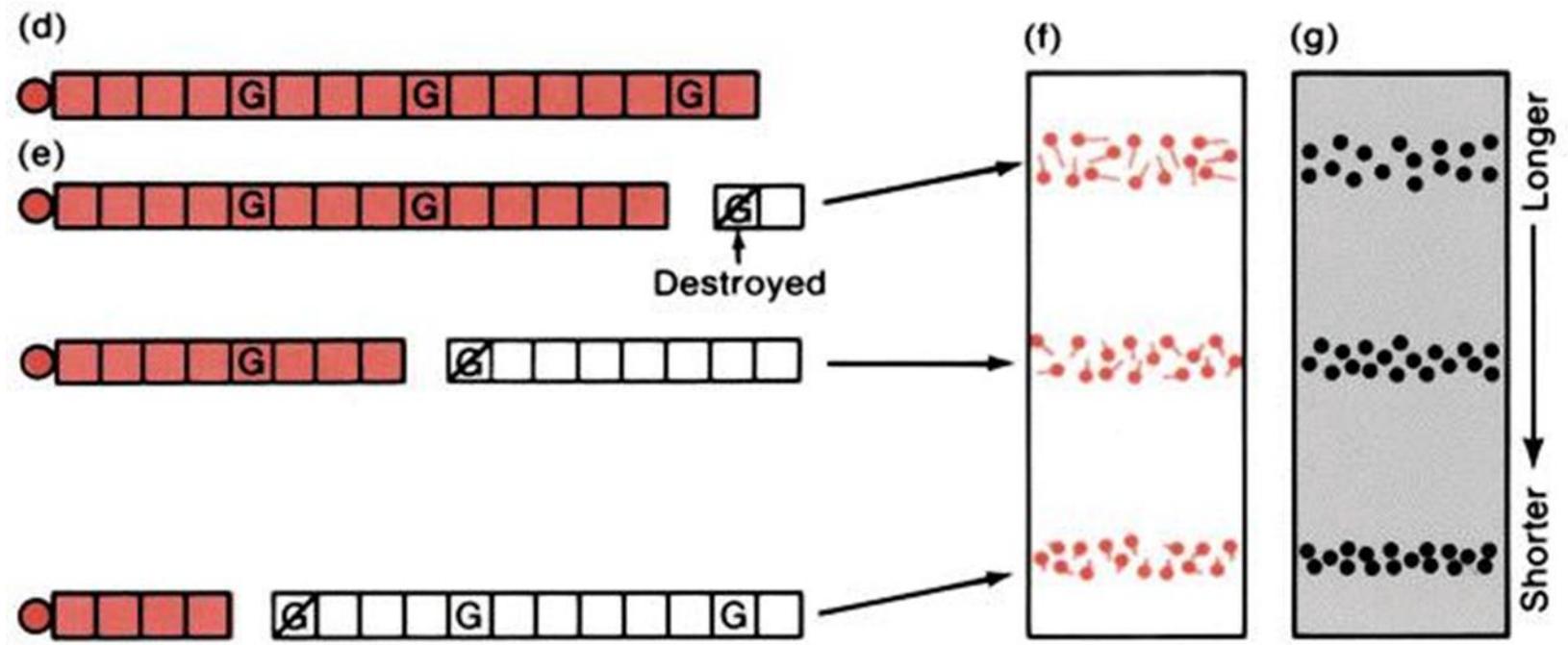
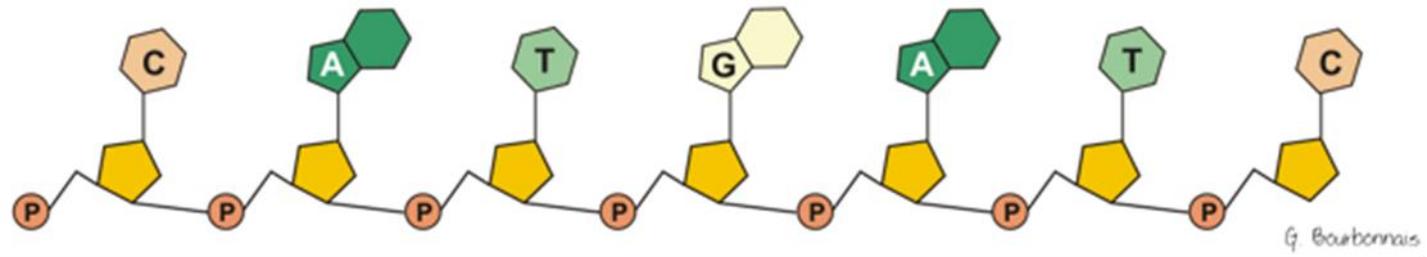
LA PCR



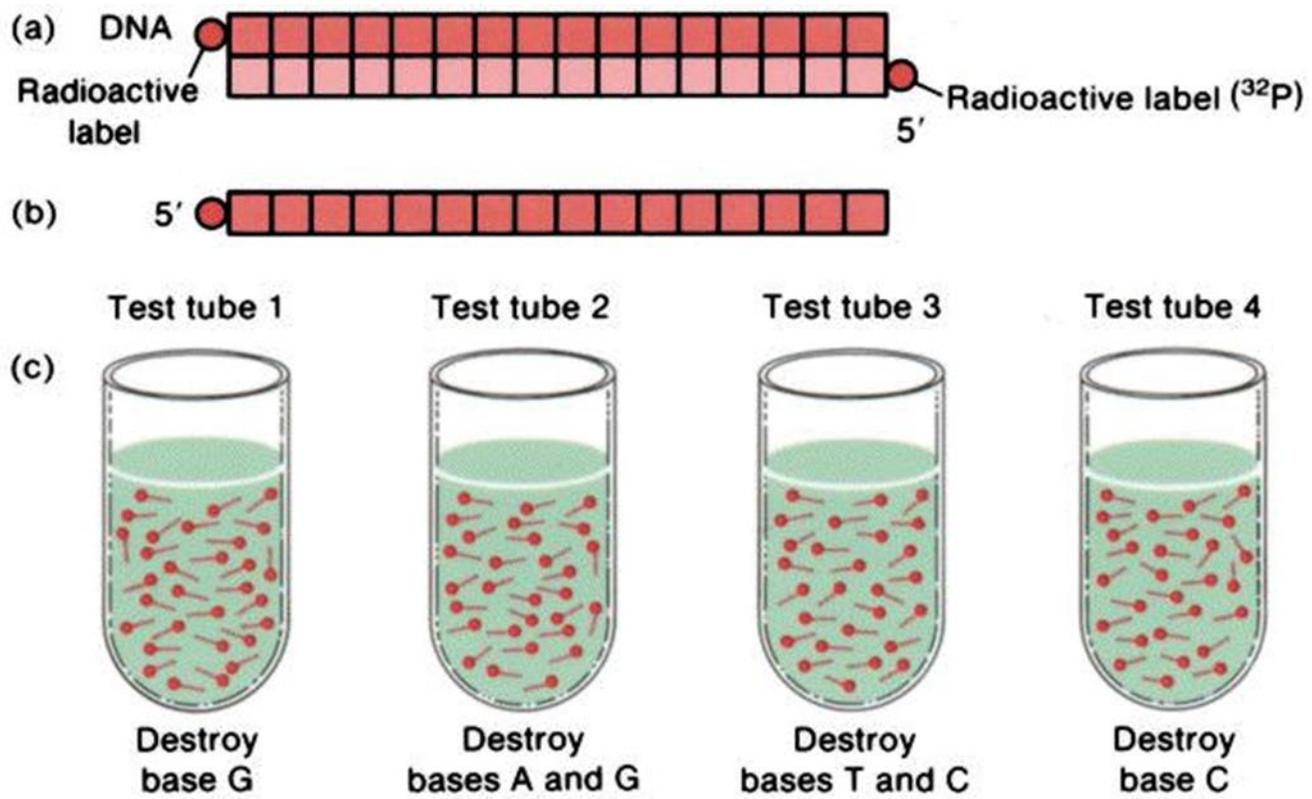
<http://www.rit.edu/~gtfsbi/IntroBiol/images/CH11/figure-11-21.jpg>

© 2001 Sinauer Associates, Inc.

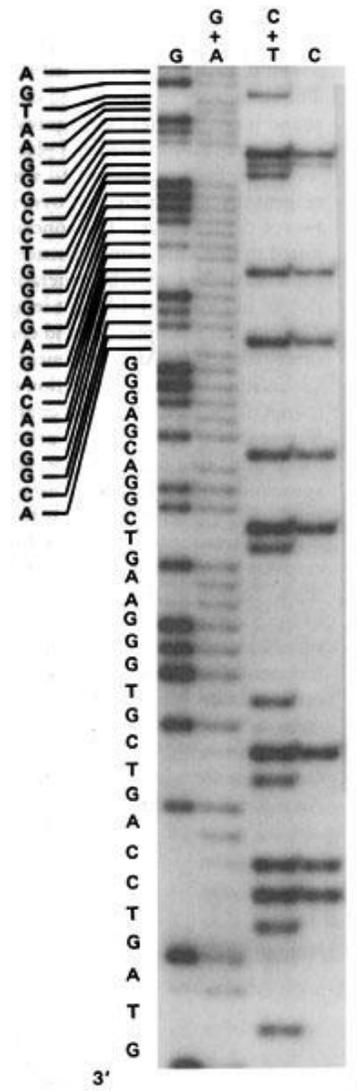
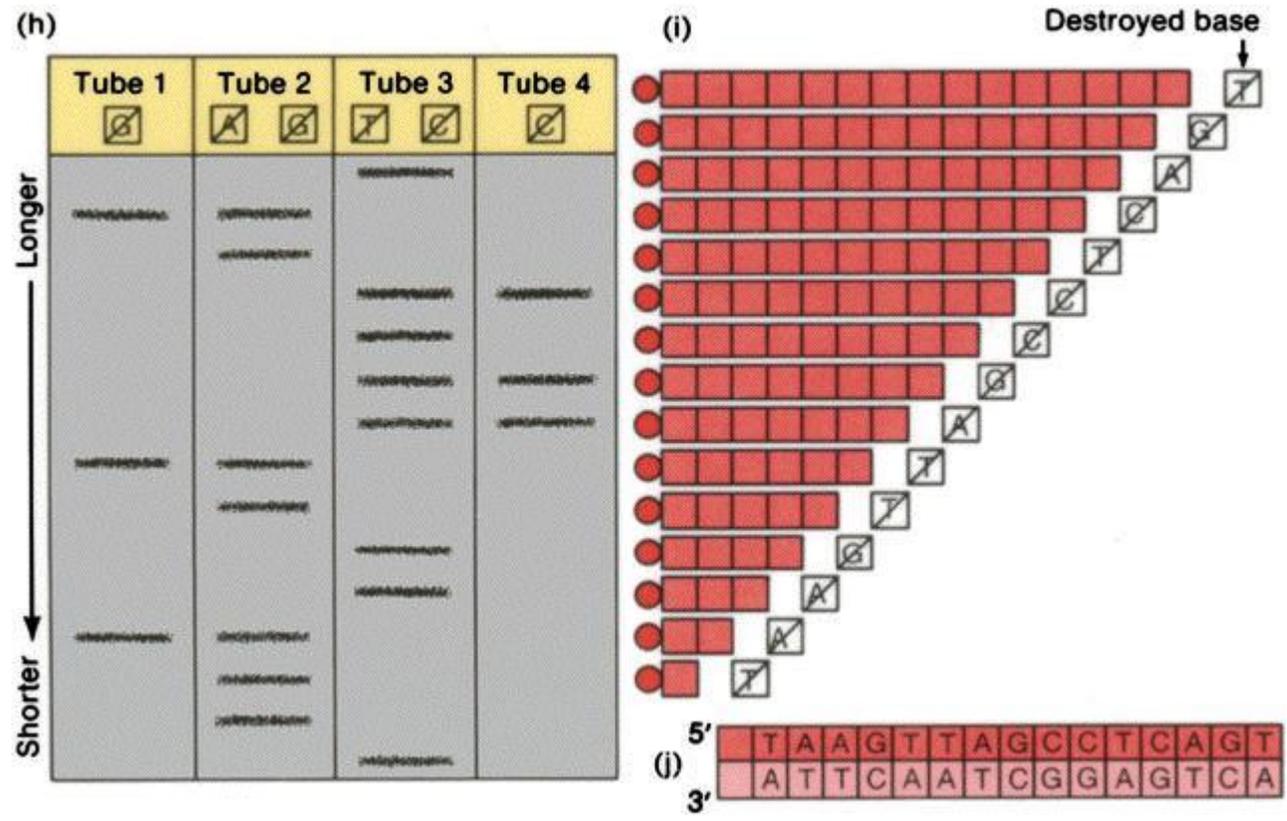
METHODE DE MAXAM ET GILBERT



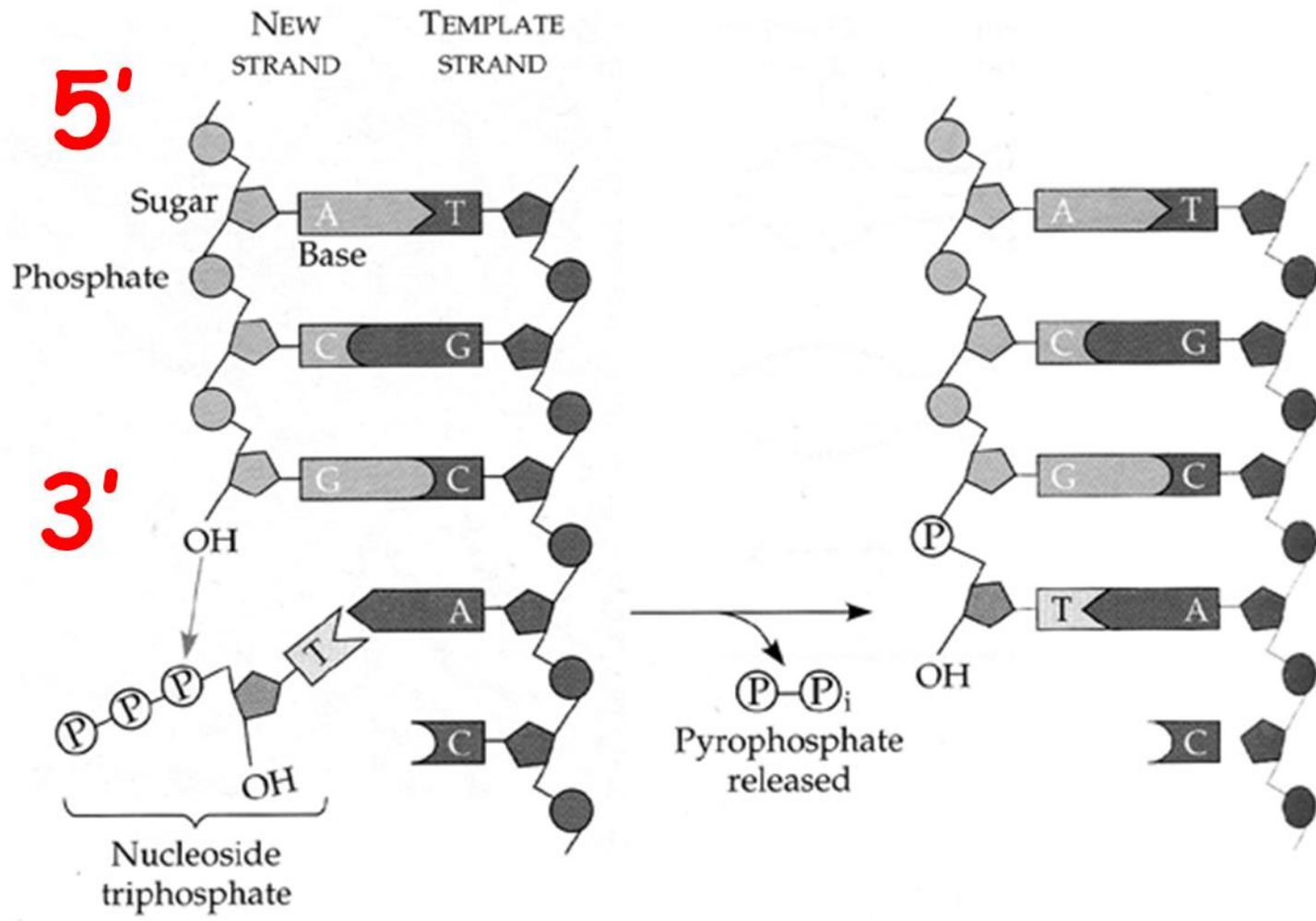
METHODE DE MAXAM ET GILBERT



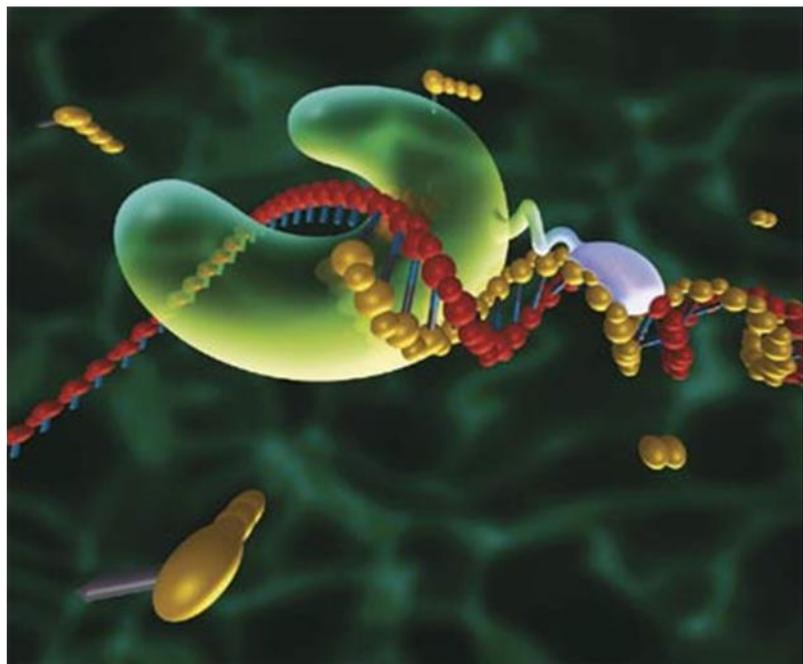
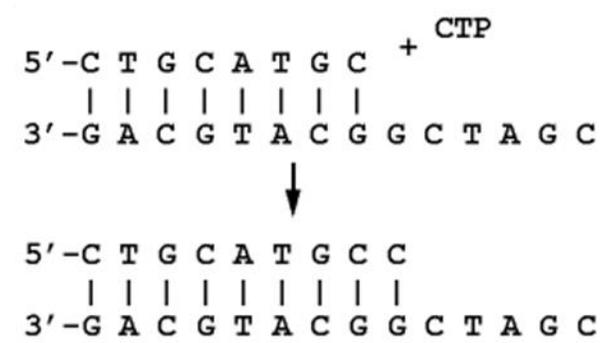
METHODE DE MAXAM ET GILBERT



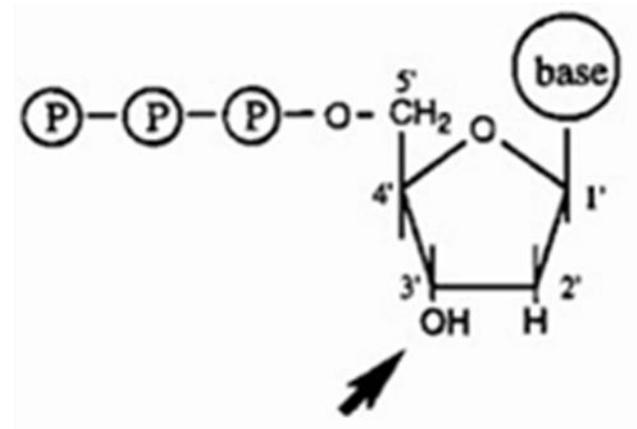
METHODE DE SANGER



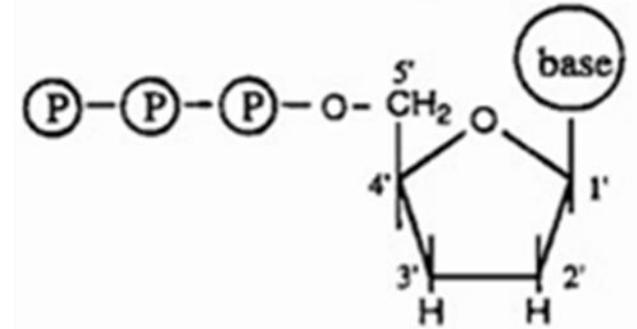
METHODE DE SANGER



http://www.finnzymes.fi/products/pcr/phusion_high_fidelity_dna_polymerase.htm



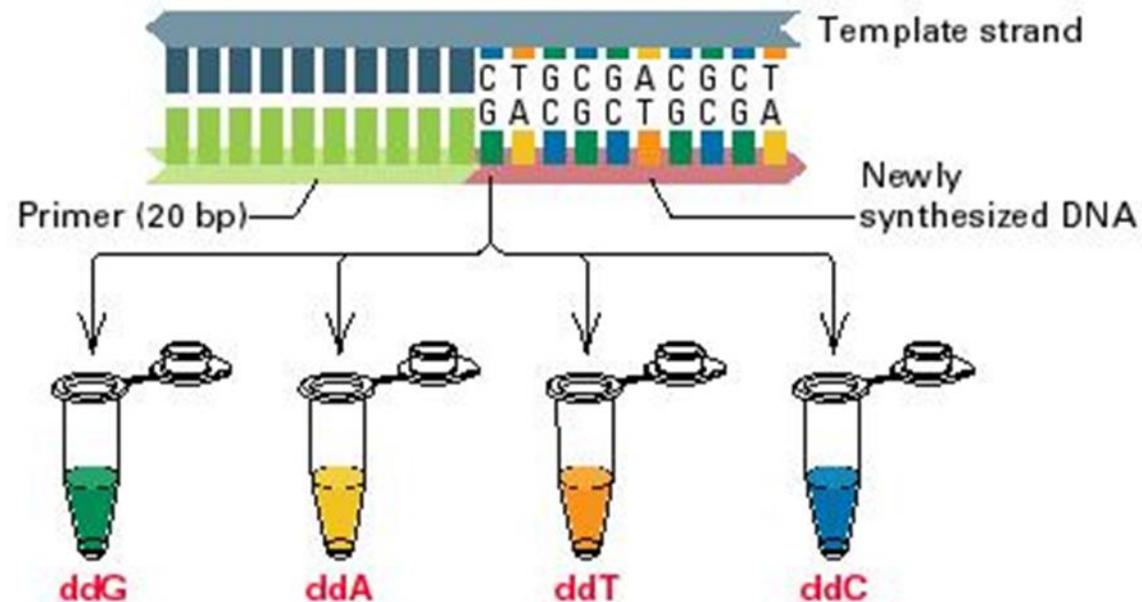
Deoxynucleotide



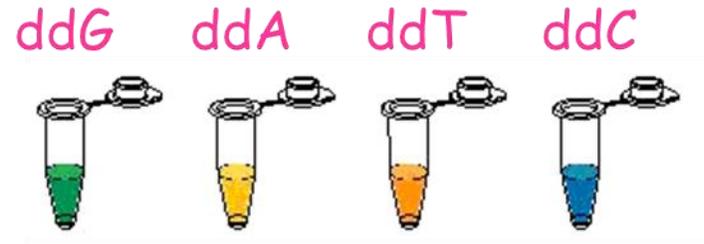
Dideoxynucleotide

METHODE DE SANGER

1. Préparation d'une grande quantité d'ADN (~1µg via miniprep)
Dénaturation et ajout de l'amorce, des 4 dNTP (dont 1
radioactif, S³⁵ ou P³²) et de l'ADN polymérase
Faire 4 aliquots et mettre un peu de ddNTP dans chaque tube
Elongation et terminaison par incorporation de
didéoxynucléotides spécifiques

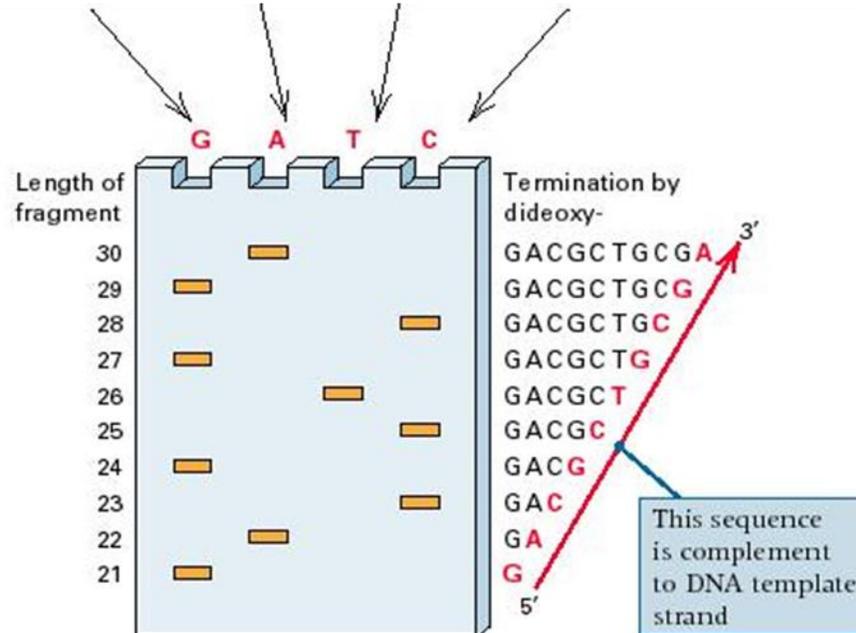


METHODE DE SANGER



5.

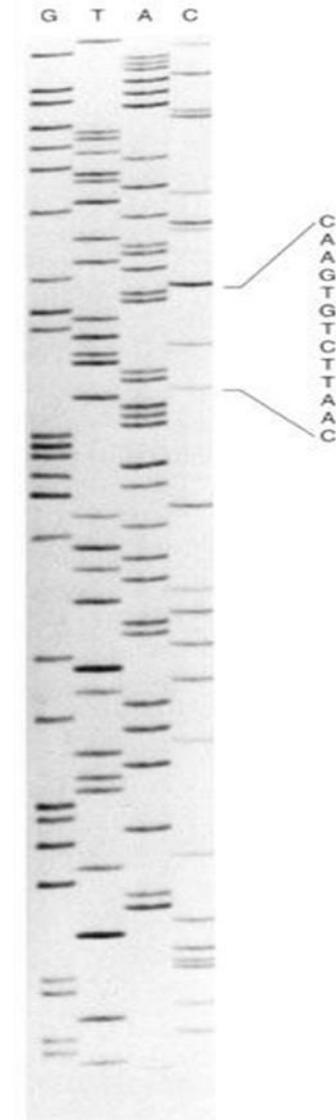
Préparation du gel du polyacrylamide
Dénaturation et dépôt sur gel
Migration par électrophorèse



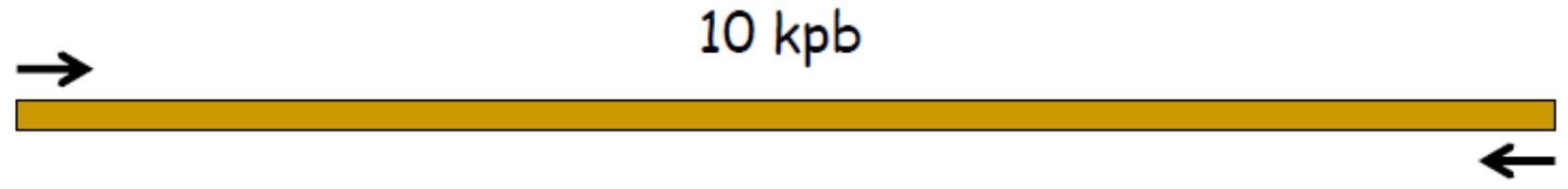
METHODE DE SANGER

8. Transfert du gel sur papier filtre
(Whatman 3MM)
Séchage du gel
Cassette autoradiographique (+ écran
amplificateur)
Développement du film autoradiographique
Lecture sur table lumineuse
Saisie sur ordinateur

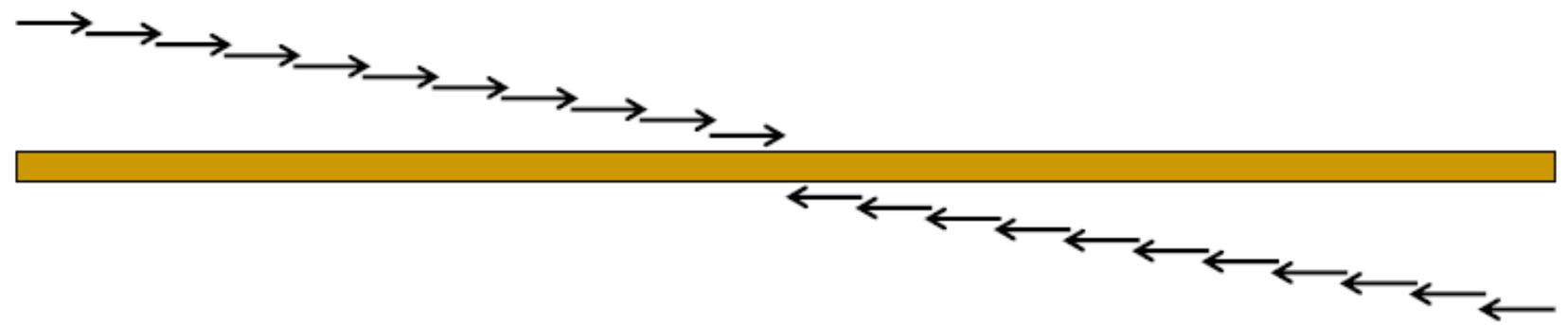
Longueur typique de lecture : 400 nucléotides
Durée totale de l'expérience : environ 3
jours



SEQUENCAGE DE GRANDS FRAGMENTS



Marche sur le chromosome

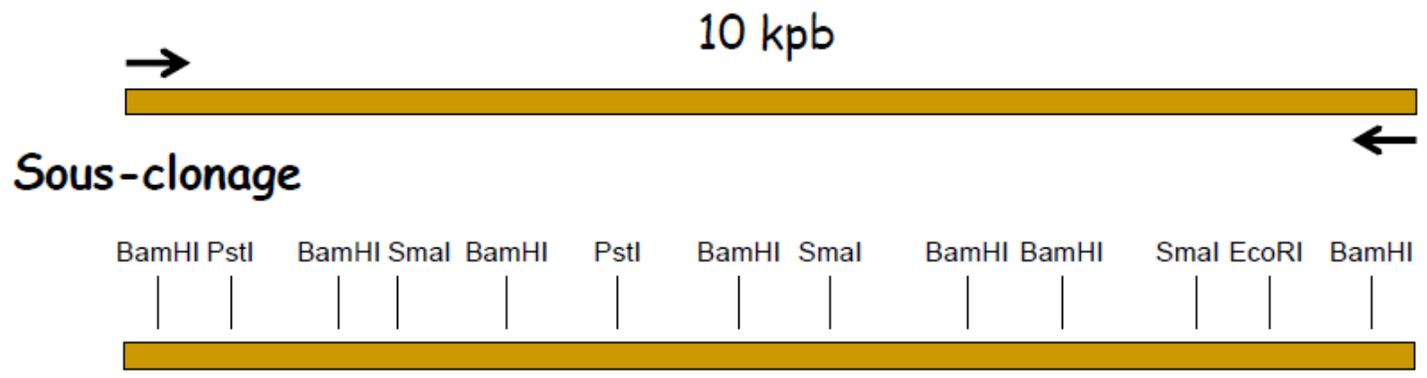


Synthèse de $2 * 10$ amorces

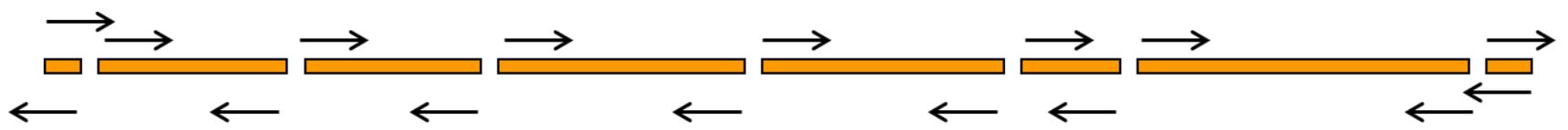
Durée totale du séquençage : $> 11 * 3$ jours

Qualité des séquences → Lecture double brin → ~
3 mois

SEQUENCAGE DE GRANDS FRAGMENTS

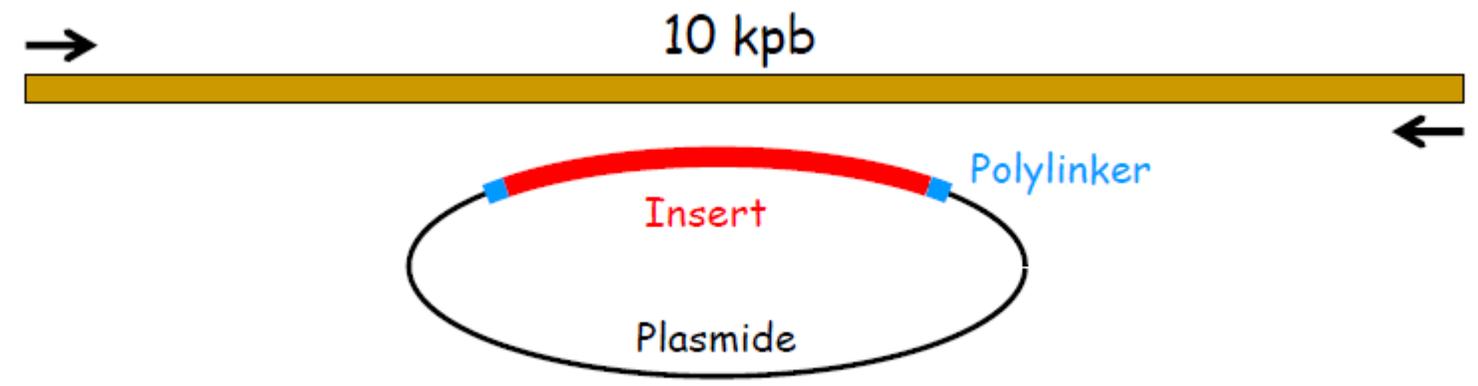


~ 5 enzymes de restriction à 6 nucléotides (HindIII, EcoRI, BamHI, PstI, SmaI)
Carte de restriction pour décider quelles enzymes retenir
Sous-clonage avec BamHI des 8 fragments



Durée totale du séquençage : ~ 5 * 3 jours

SEQUENCAGE DE GRANDS FRAGMENTS



2 enzymes de restriction (5' sortant et 5' rentrant)



Digestion avec une exonucléase



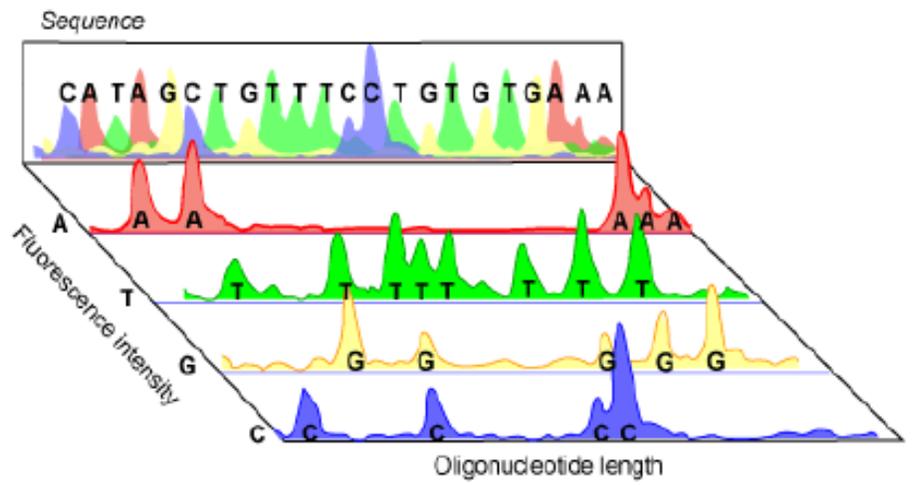
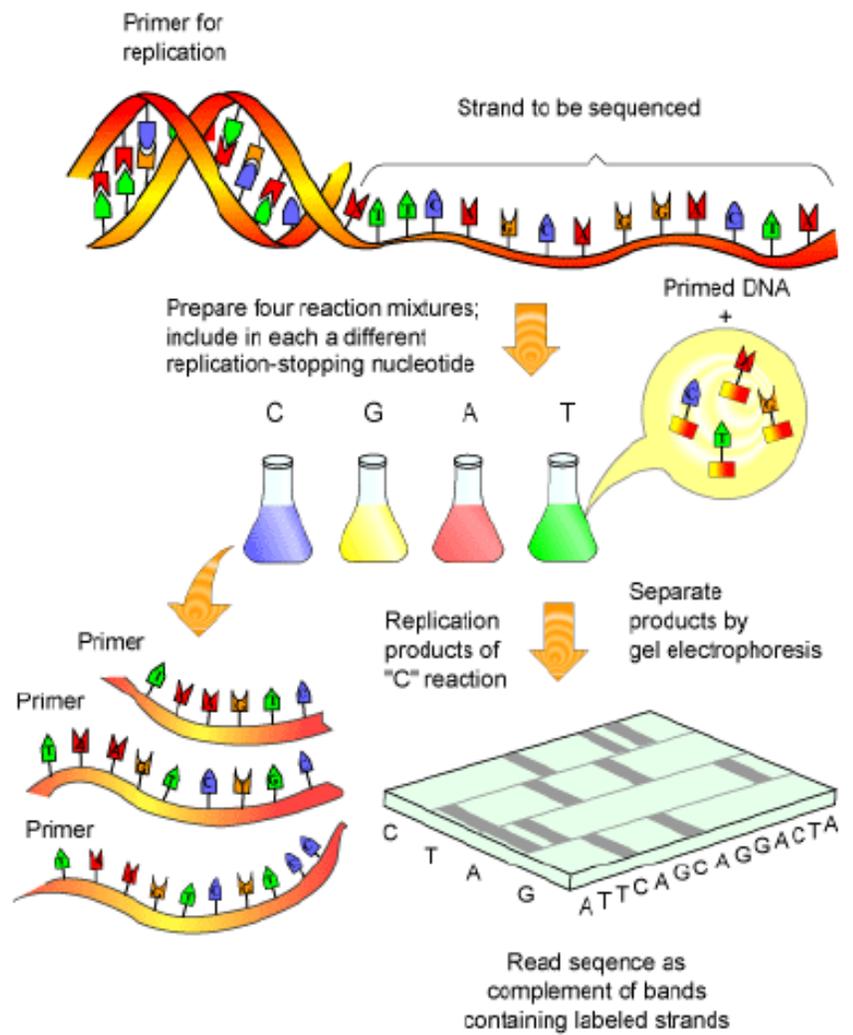
Prélèvements toutes les ~30", ligation, clonage, séquençage → ~10 jours

ETAPES LIMITANTES DU SEQUENCAGE

Temps, coûts, difficulté de robotiser :

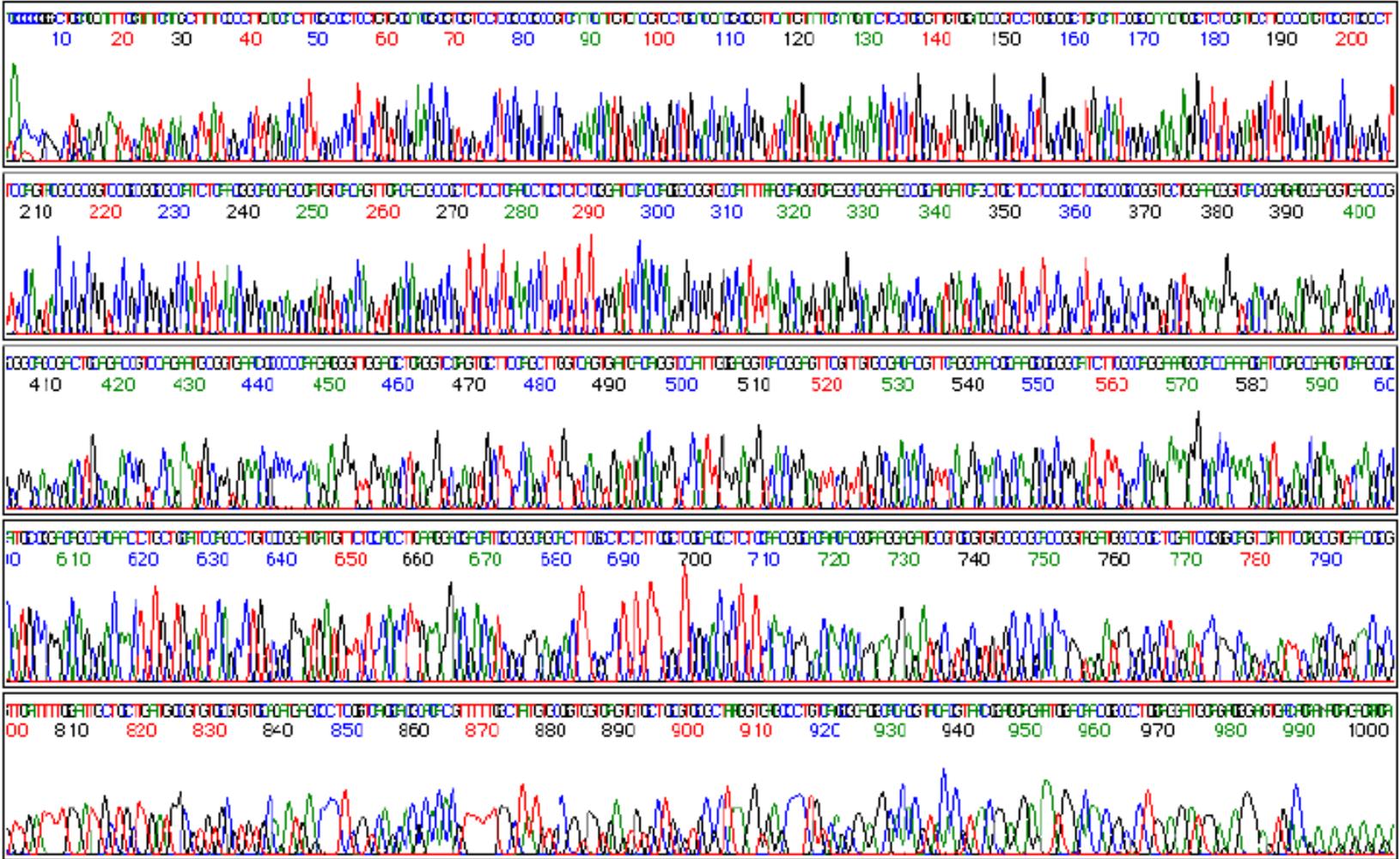
- Obtention de l'ADN (miniprep)
- Autoradiographie : 12h à des jours
- Couler les gels
- Détermination et fabrication des amorces (marche sur le chromosome)
- Sous-clonage

UTILISATION DES FLUOROCHROMES



Cycle sequencing

UTILISATION DES FLUOROCHROMES

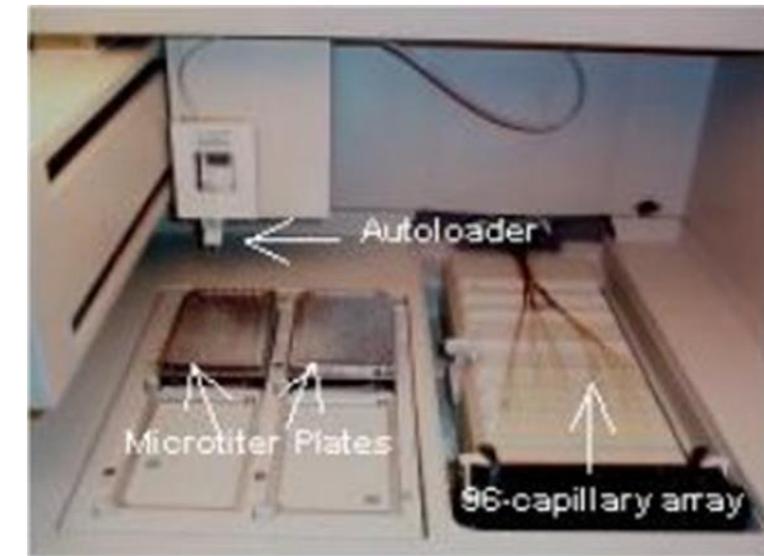


http://www.genotype.de/d/p/d_gel_plasmid_prep.htm

METHODES CAPILLAIRES

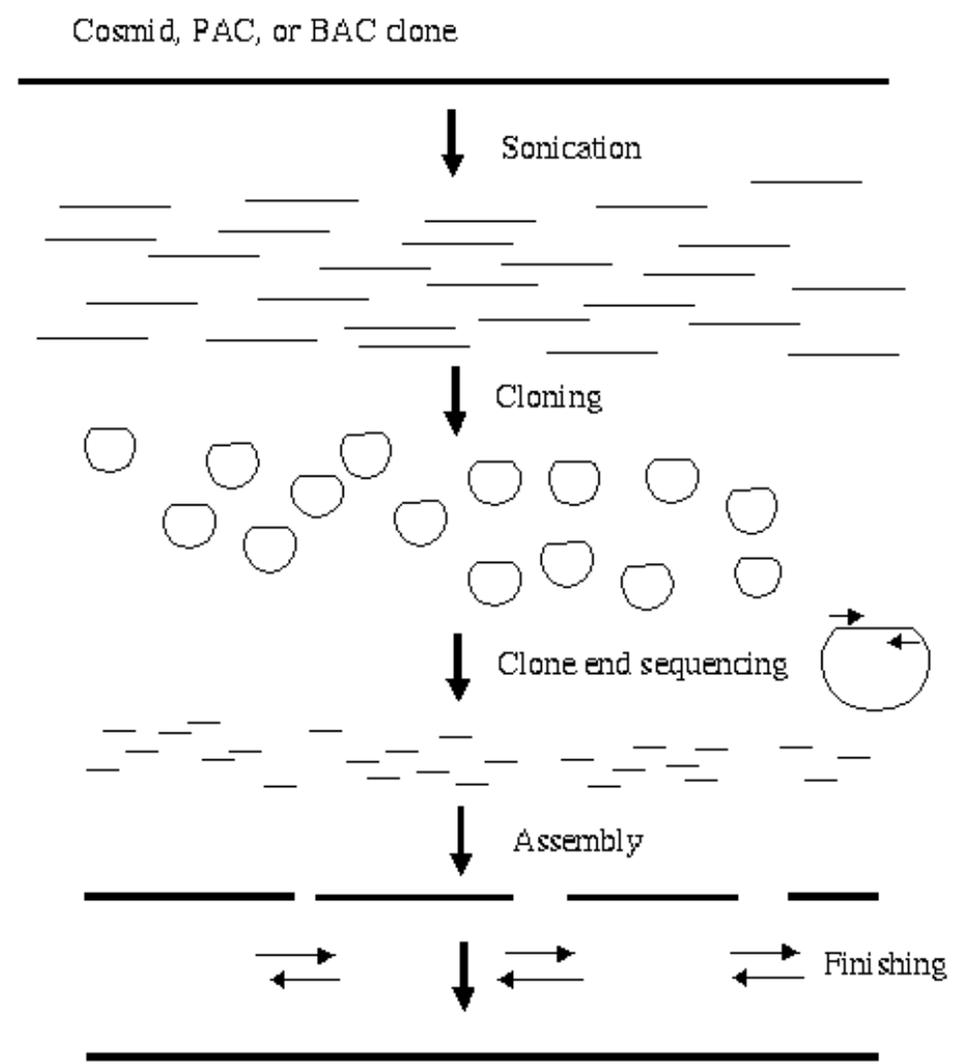


http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp2_2.shtml

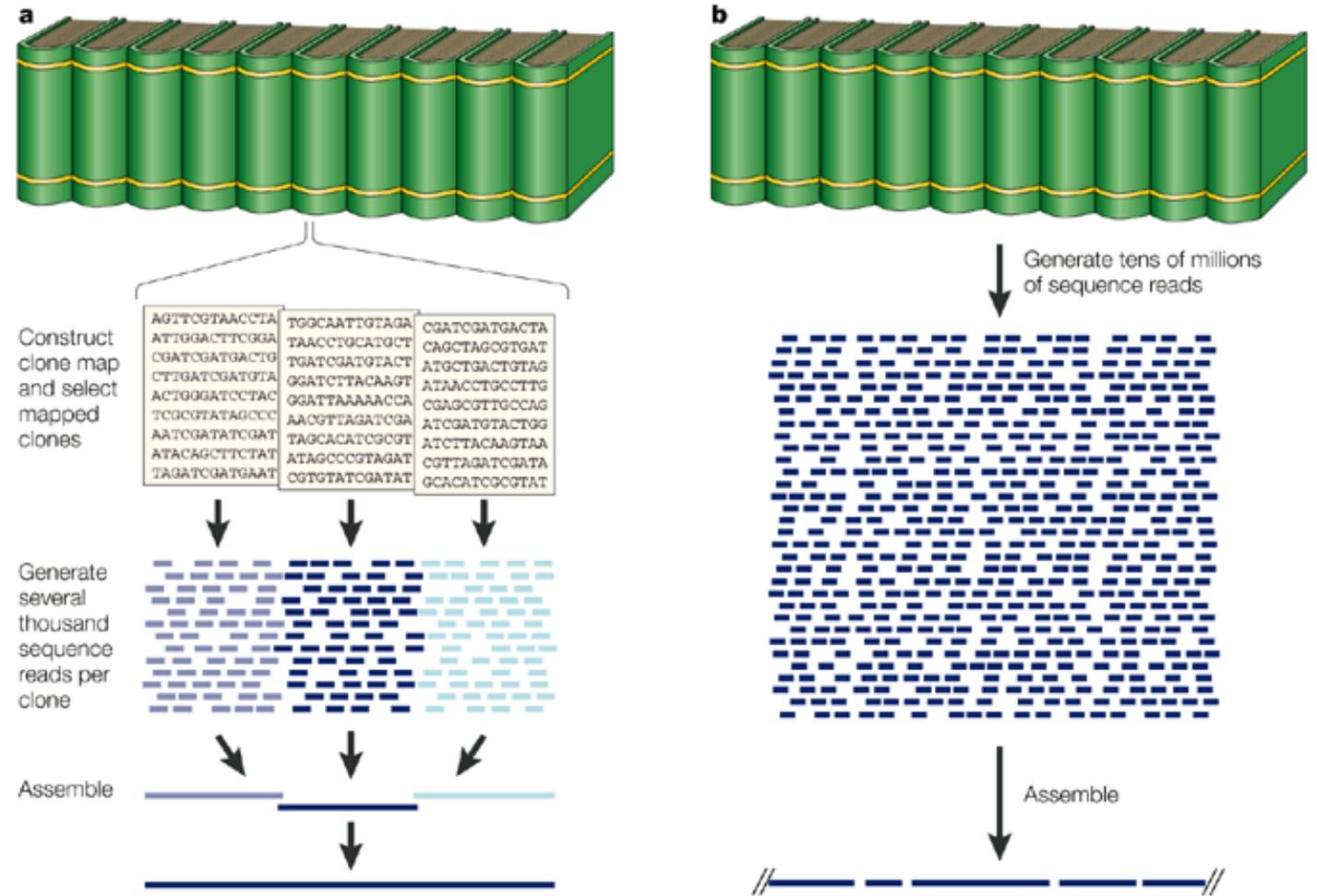


<http://www.aecom.yu.edu/cancer/new/cores/sequencing/abi3700.htm>

SEQUENCAGE PAR SHOTGUN



SEQUENCAGE PAR SHOTGUN

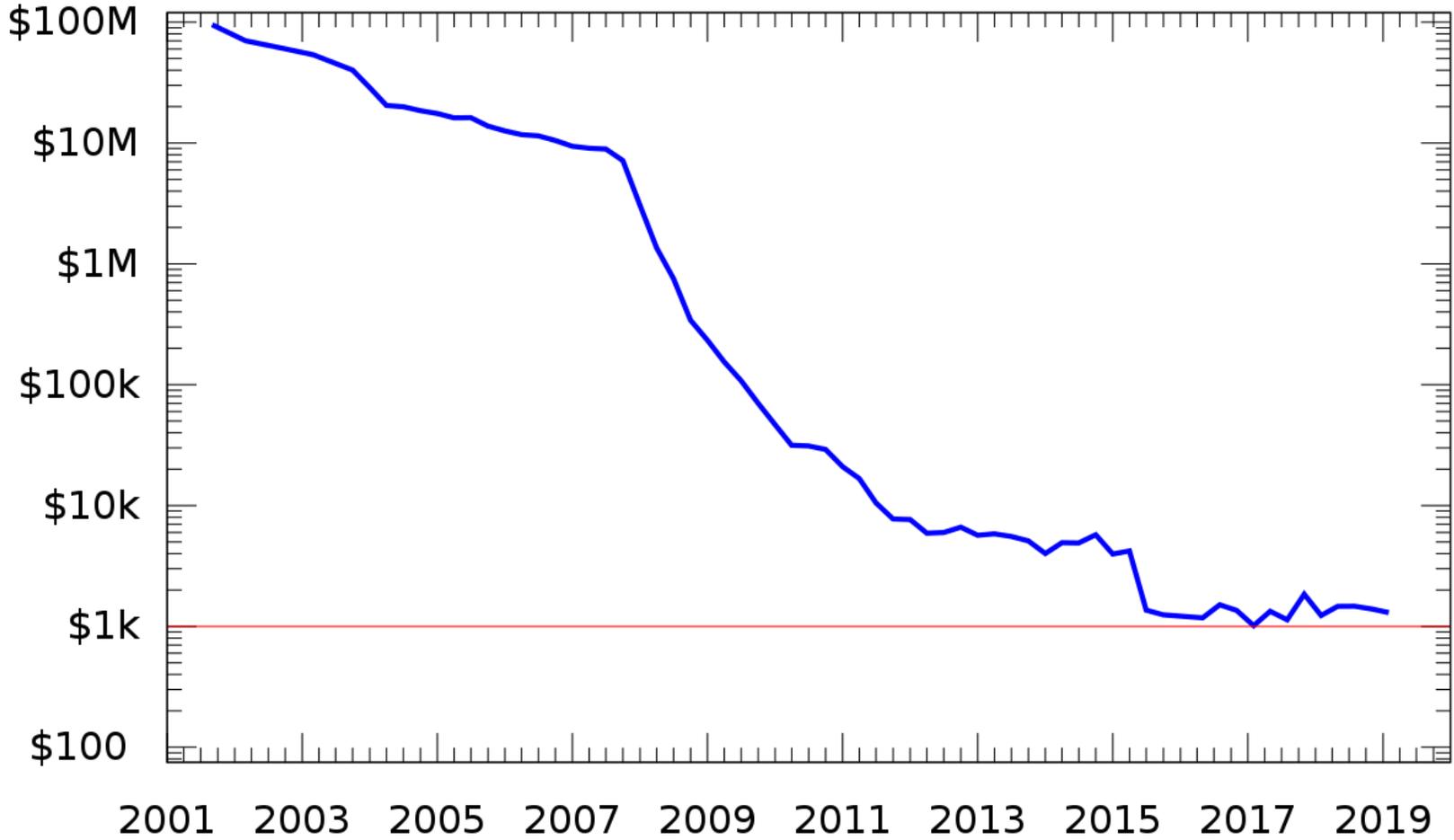


PROGRES DU SEQUENCAGE

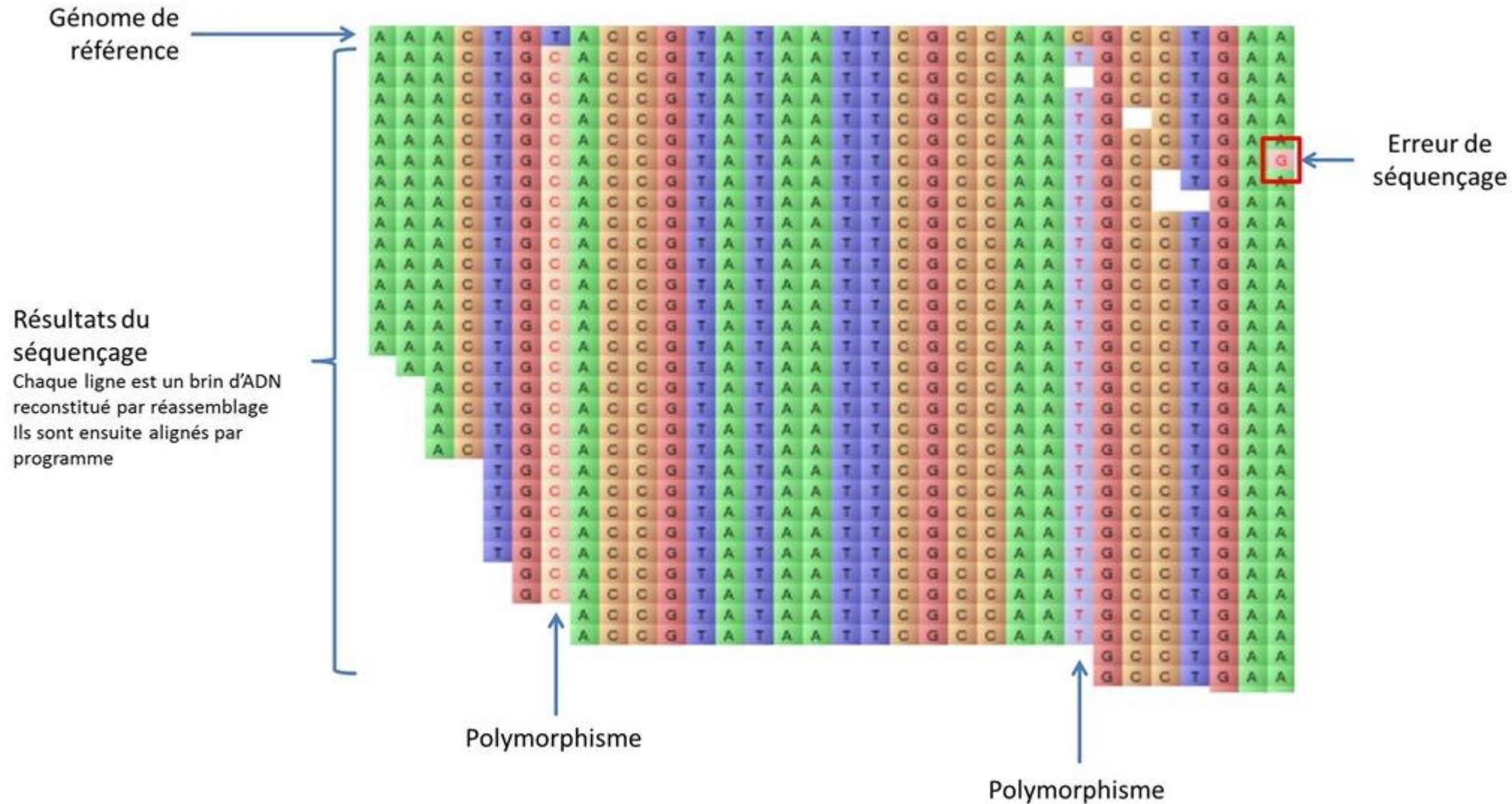
Année	#réactions/pers onne/semaine	Longueur moyenne de lecture	#nucléotides/personne/semaine
1977	4	50	200
1980	20	100	2 000
1990	60	300	18 000
1997	180	500	90 000
1999	500	650	325 000
2000	5000	600	3 000 000

PROGRES DU SEQUENCAGE

Cost to sequence a human genome (USD)



UTILITE DES OUTILS BIOINFORMATIQUES



UTILITE DES OUTILS BIOINFORMATIQUES

Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications

(Archaea/Bacteria/eukaryote/phylogeny)

JAMES R. BROWN* AND W. FORD DOOLITTLE

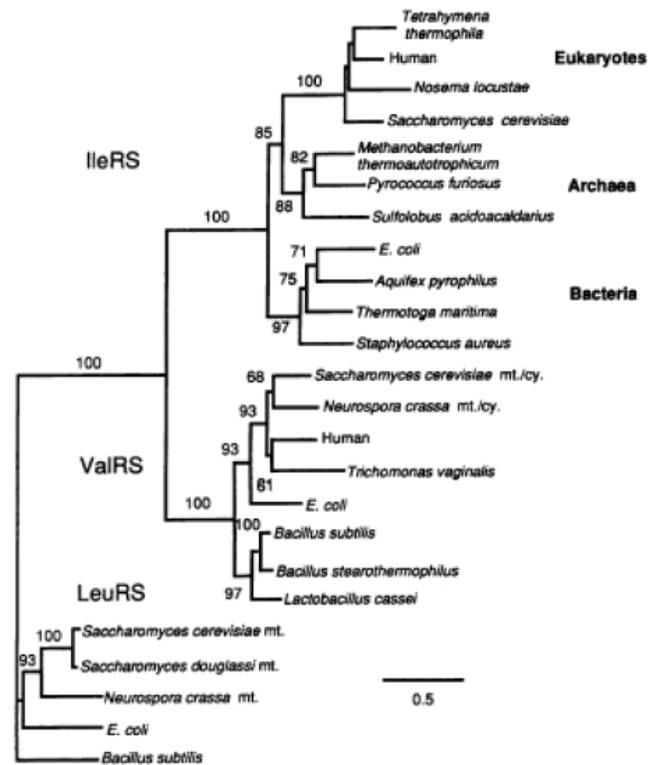


FIG. 3. Neighbor-joining tree of IleRS, ValRS, and LeuRS genes using the program NEIGHBOR (26). The scale represents 0.5 expected number of amino acid replacements per position as determined with the program PROTDIST. Numbers are the frequency of occurrence of nodes that exceeded 50% of 300 bootstrap replicates.

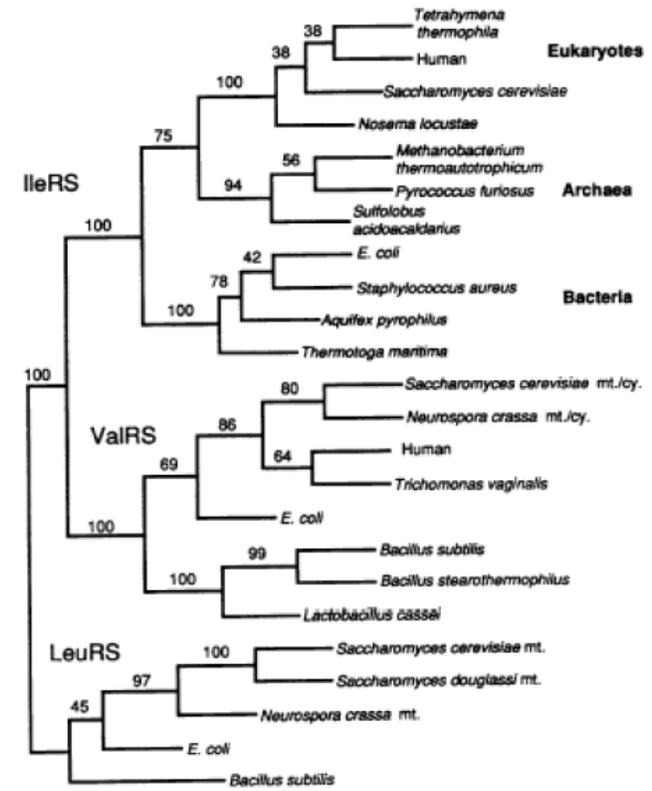


FIG. 2. Consensus maximum parsimony tree of IleRS, ValRS, and LeuRS genes using the program PROTPARS (26). Numbers are the frequency of occurrence of nodes in 300 bootstrap replicates.

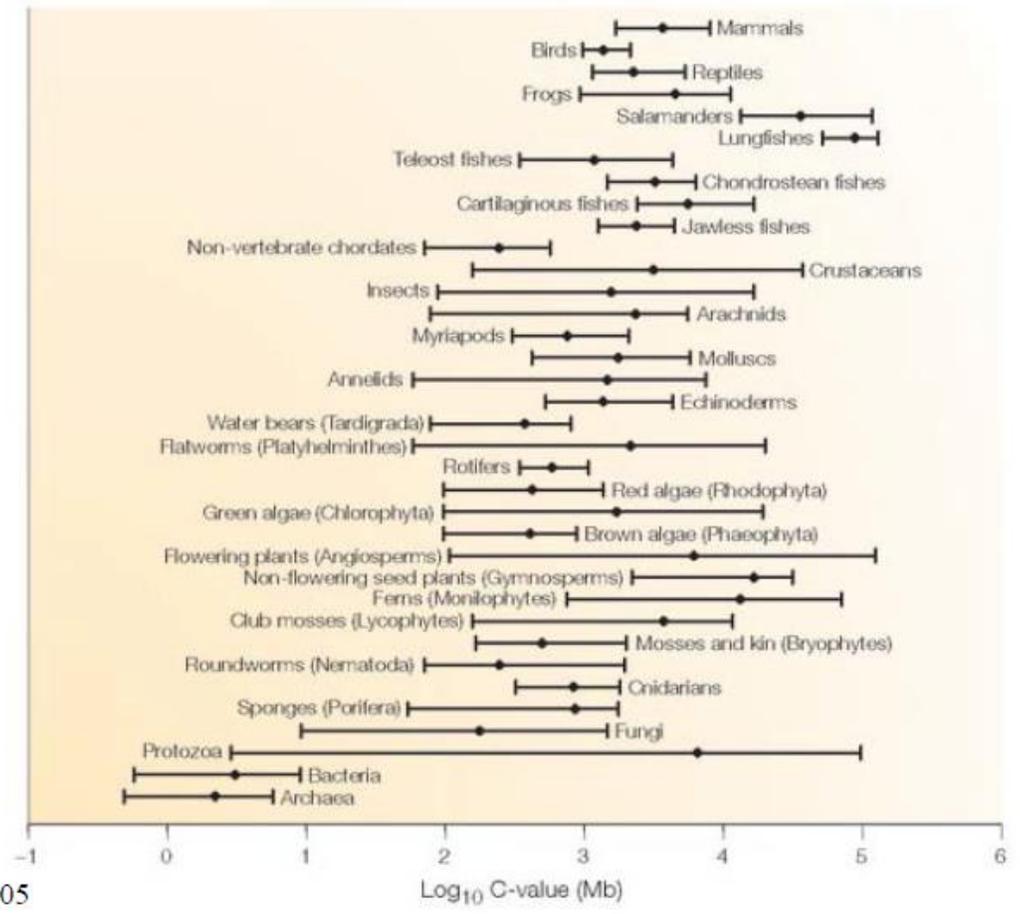
TAILLE DES GENOMES ET « COMPLEXITE »

Taille du génome chez les animaux:
-0,03 pg (30 MB) chez *Meloidogyne graminicola* (Nématode)
-132,83 pg (130GB) chez *Protopterus aethiopicus* (Poisson)

C-value:
contenu en ADN
d'une cellule haploïde

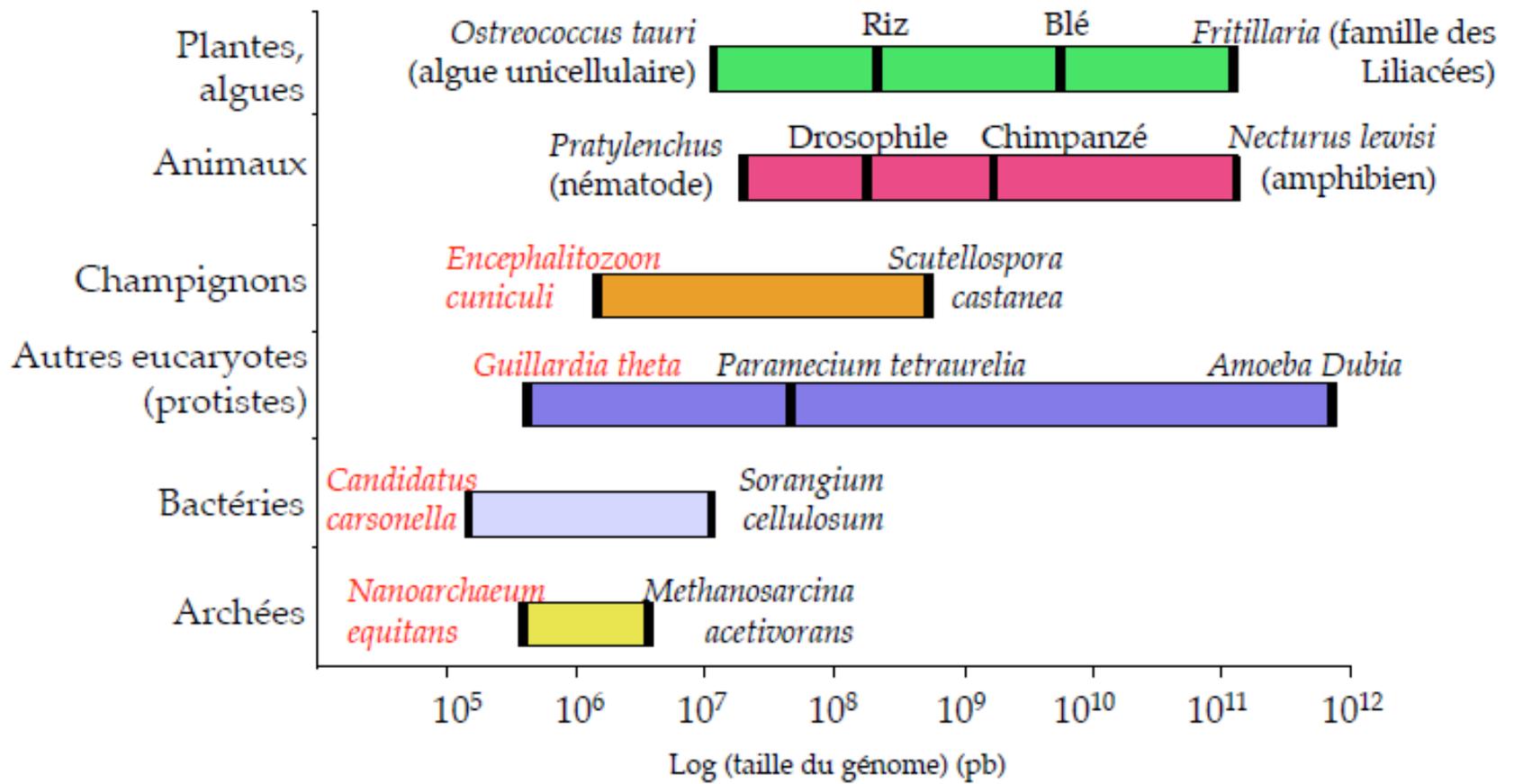
 $1 \text{ pg} = 10^{-12} \text{ g} \Rightarrow 978 \text{ Mb}$

<http://www.genomesize.com/>



Gregory, Nature Reviews Genetics 2005

TAILLE DES GENOMES ET « COMPLEXITE »



Parasites (ou symbiontes) associés à des cellules

Eukaryotic genome size databases (2007)

TAILLE DES GENOMES ET « COMPLEXITE »

La longueur des génomes et le nombre de gènes sont corrélés chez les procaryotes, **pas chez les eucaryotes**.

Le pourcentage de séquences codantes (densité) est très variable en fonction des organismes.

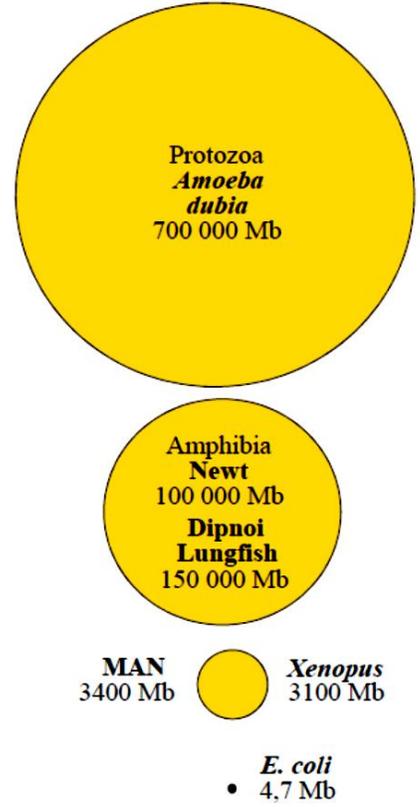
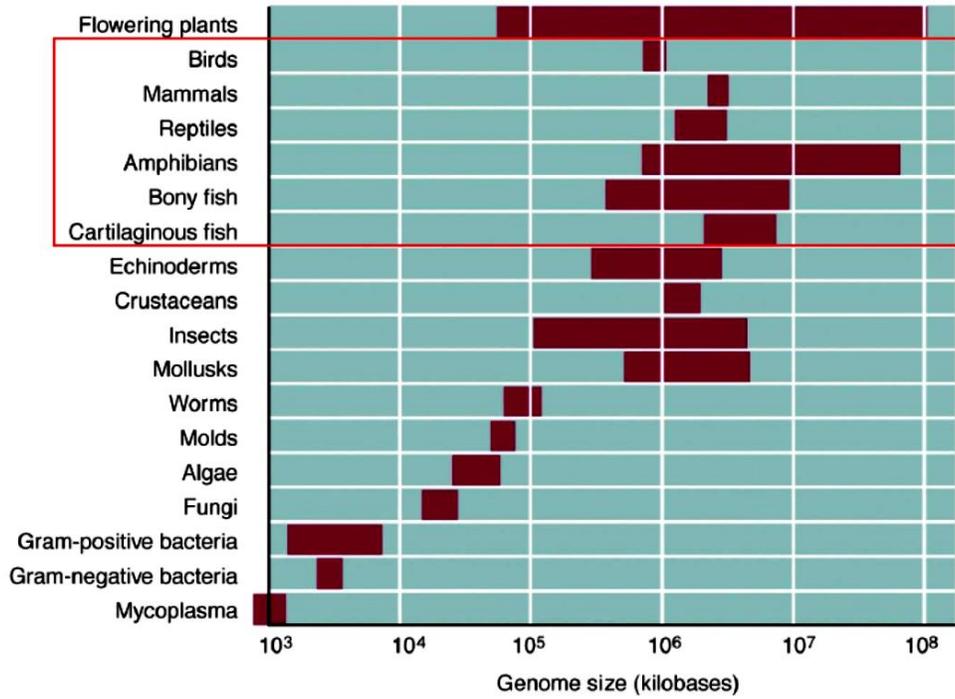
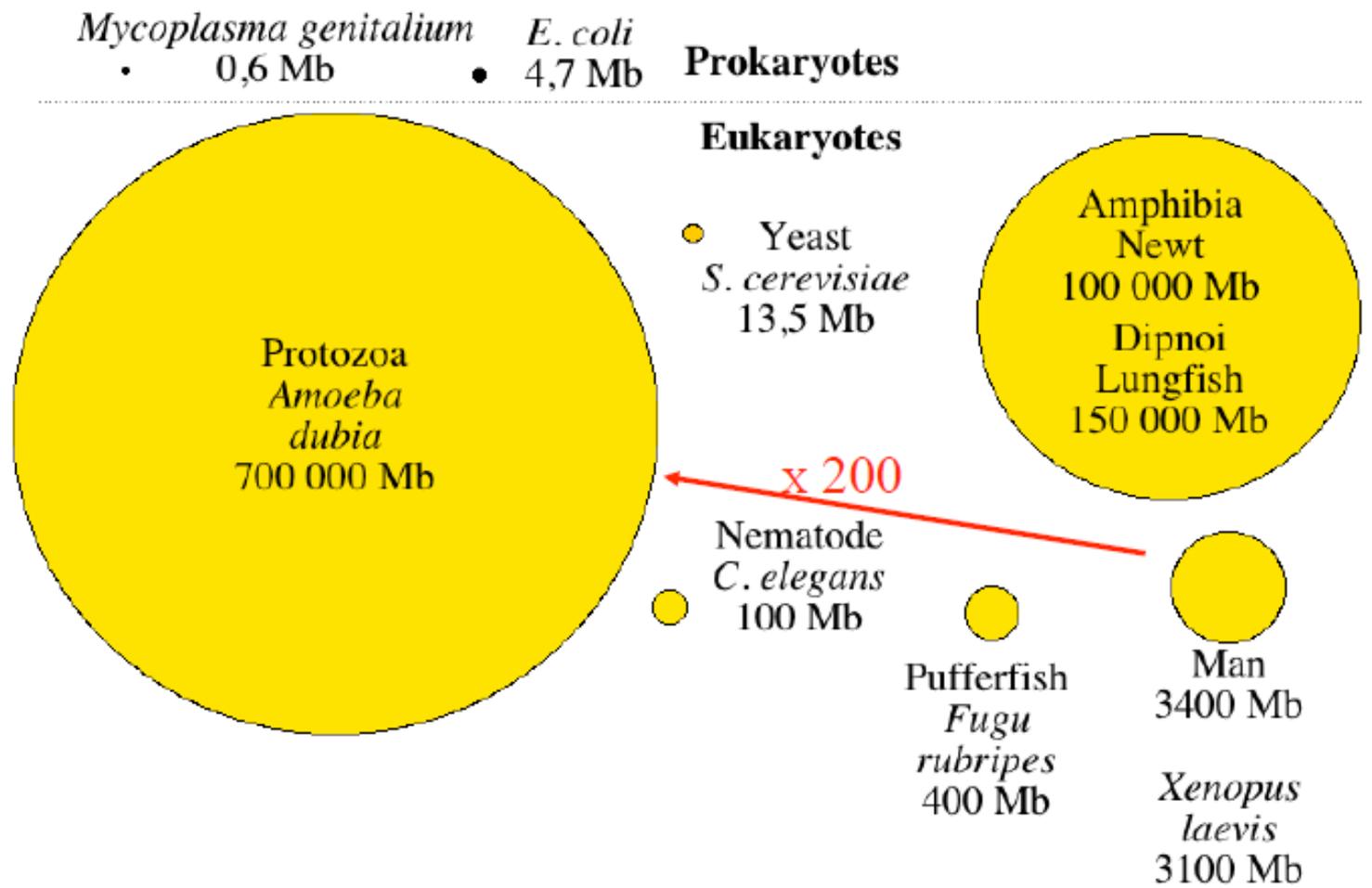


Figure 1.36 The amount of nucleic acid in the genome varies over an enormous range.

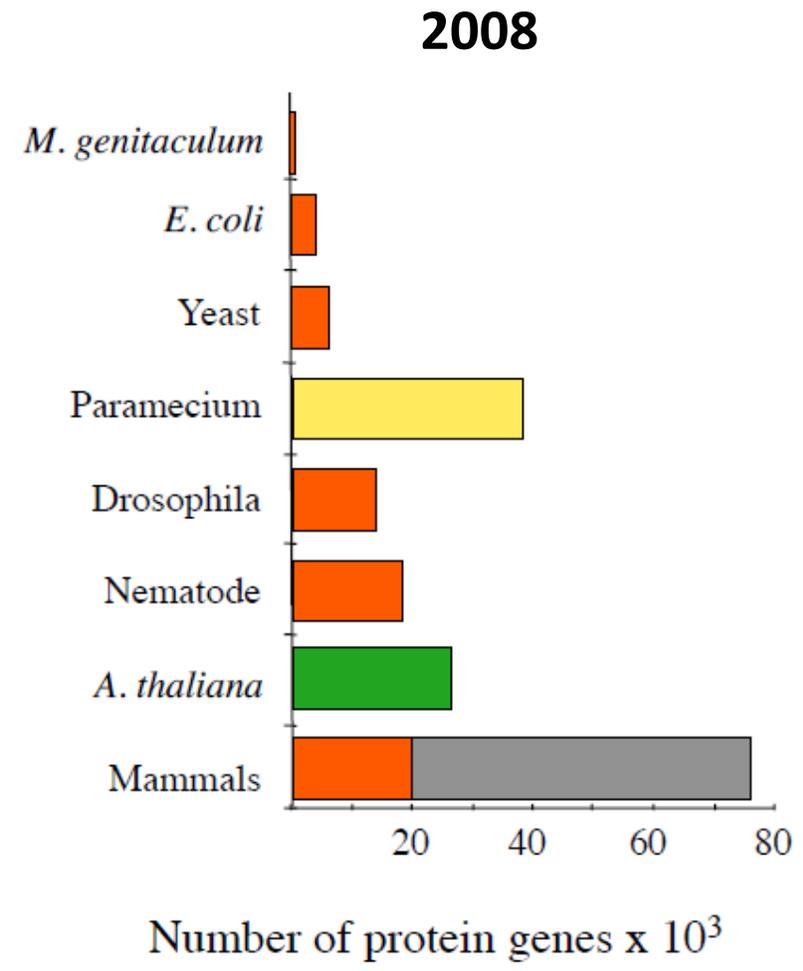
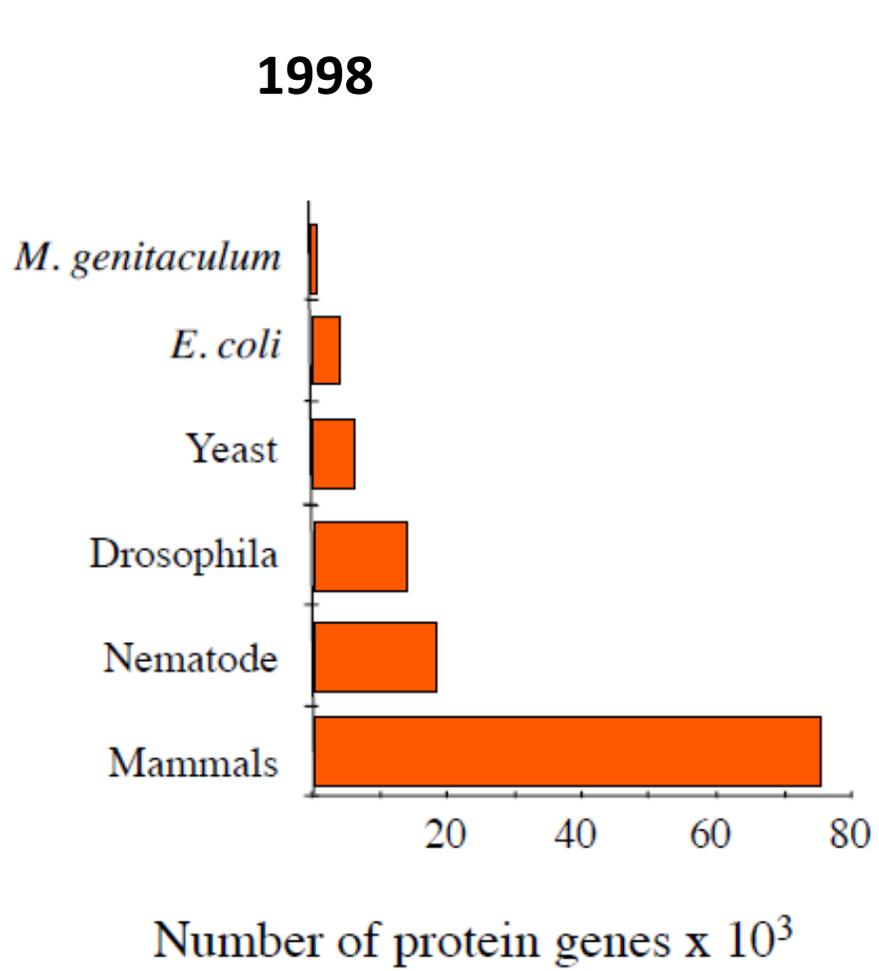
Genome	Gene Number	Base Pairs
Organisms		
Plants	<50,000	<10 ¹¹
Mammals	100,000	~3 × 10 ⁸
Worms	14,000	~10 ⁸
Flies	12,000	1.6 × 10 ⁸
Fungi	6,000	1.3 × 10 ⁷
Bacteria	2-4,000	<10 ⁷
Mycoplasma	500	<10 ⁶
dsDNA Viruses		
Vaccinia	<300	187,000
Papova (SV40)	~6	5,226
Phage T4	~200	165,000
ssDNA Viruses		
Parvovirus	5	5,000
Phage φX174	11	5,387
dsRNA Viruses		
Reovirus	22	23,000
ssRNA Viruses		
Coronavirus	7	20,000
Influenza	12	13,500
TMV	4	6,400
Phage MS2	4	3,569
STNV	1	1,300
Viroids		
PSTV RNA	0	359
Scrapie		
Prion	?	?

TAILLE DES GENOMES ET « COMPLEXITE »

Le paradoxe de la C-value !



ANNOTATION DES GENOMES



Homme vs. *E. coli* : Génome x600, nombre de gènes x4 !!!

ANNOTATION DU GENOME HUMAIN

Technique	Gene estimate	Comments/assumptions
Text book (1990)	100,000	if average size = 30 kb
Genomic sequencing (1994)	71,000	biased toward gene-rich region?
CpG islands	80,000	assumes 66% human genes have CpG islands
EST analysis (1994)	64,000	matching with GenBank; 50% EST redundancy
Chromosome 22 (1999)	45,000	correction for high gene density on chrom. 22
Exofish (2000)	28,000-34,000	Comparison human/fish
EST (2000)	35,000	Number of genes
EST (2000)	120,000	Number of transcripts
First genome draft (2001)	30,000-40,000	Known genes + predictions
Comparison / mouse (2002)	30,000	Known genes + predictions
Finished genome (2004)	20,000-25,000	Known genes + predictions
Finished genome (2007)	20,000	Improved gene annotation

GENES ESSENTIELS

Un gène dont la délétion est létale est dit essentiel.

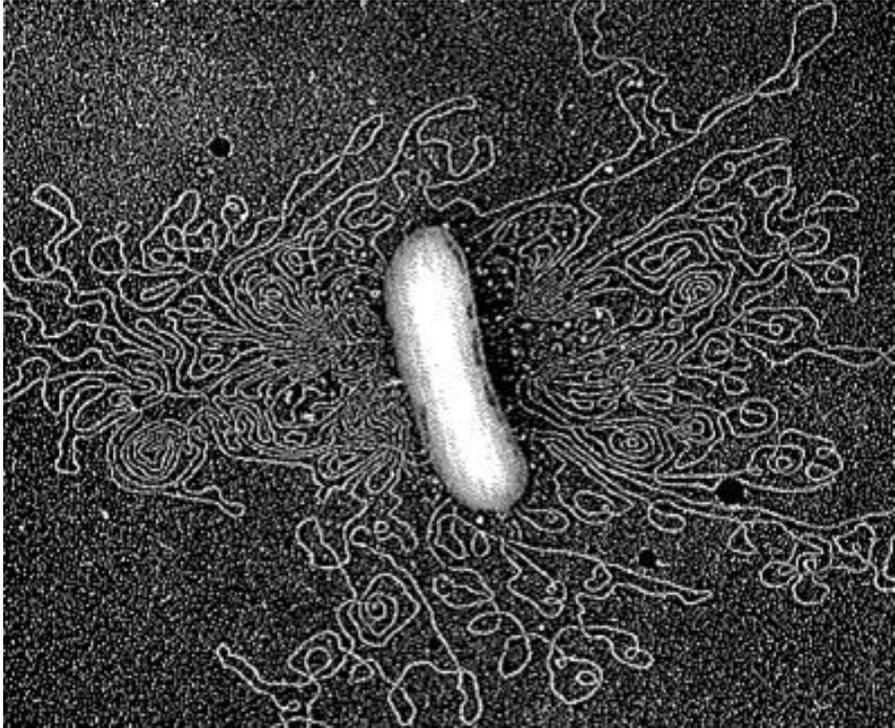
La recherche du sous-ensemble minimal de gènes pour parvenir à faire vivre un organisme vivant est toujours en cours (environ 500 identifiés actuellement).

Figure 3.10 Genome sizes, gene numbers and lethal loci.

Species	Genome (Mb)	Genes	Lethal loci
<i>Mycoplasma genitalium</i>	0.58	470	
<i>Rickettsia prowazekii</i>	1.11	834	
<i>Haemophilus influenzae</i>	1.83	1,743	
<i>Methanococcus jannaschi</i>	1.66	1,738	
<i>B. subtilis</i>	4.2	4,100	
<i>E. coli</i>	4.6	4,288	1,800
<i>S. cerevisiae</i>	13.5	6,034	3,600
<i>D. melanogaster</i>	165	12,000	3,100
<i>C. elegans</i>	97	19,099	
<i>H. sapiens</i>	3,300	100,000	

2. TOPOLOGIE DE L'ADN

UN GENOME BACTERIEN : *E. Coli*



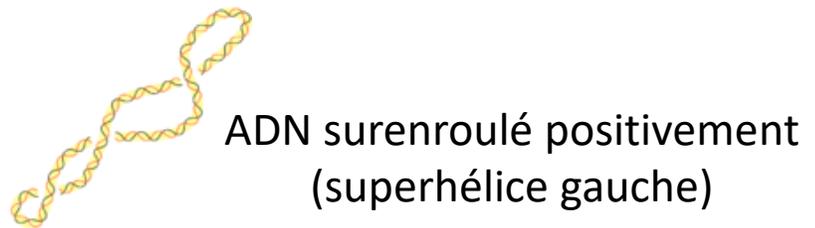
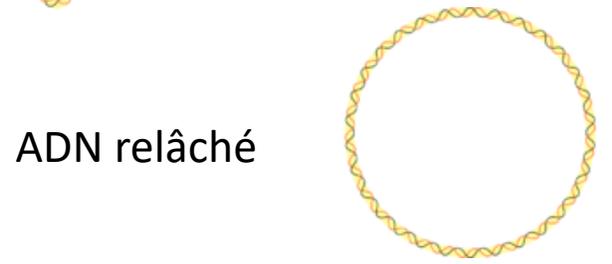
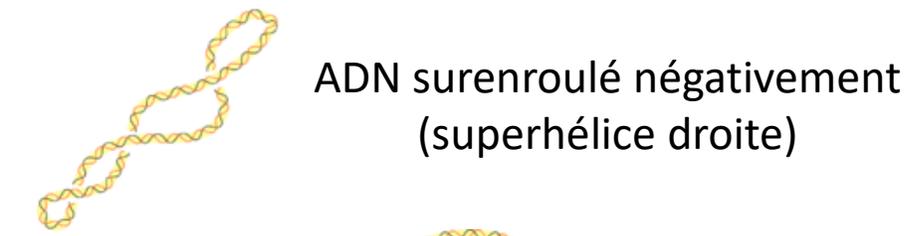
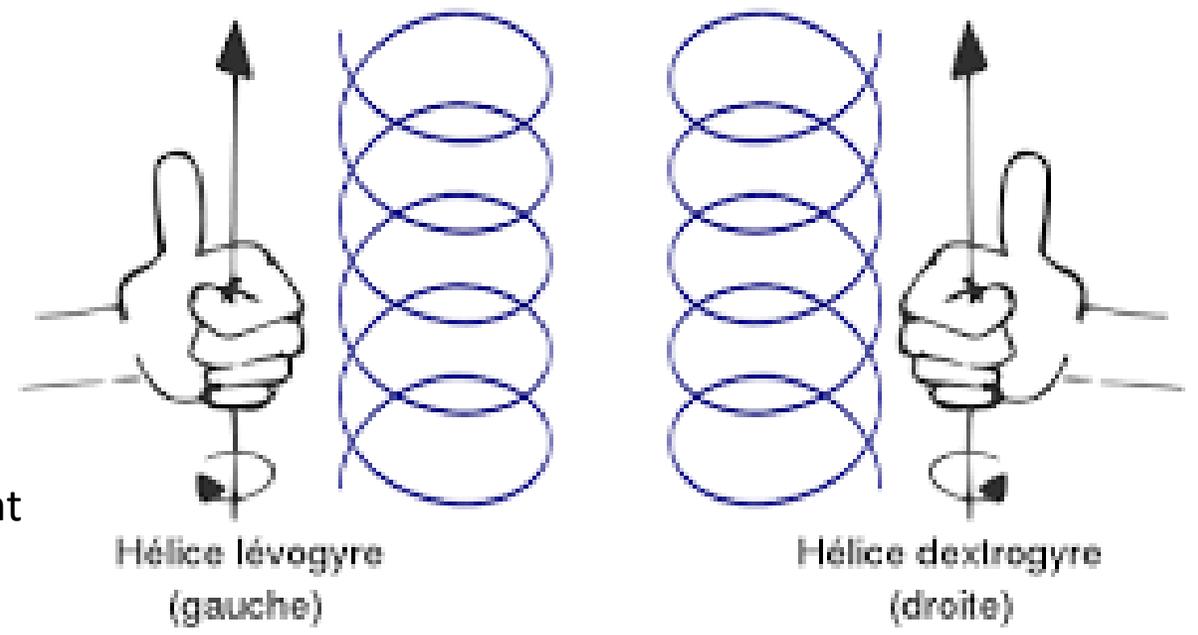
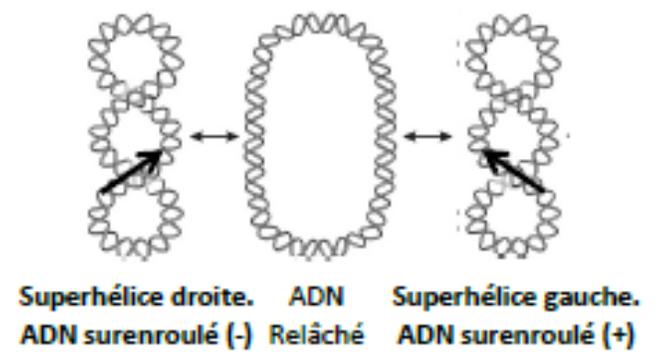
Taille : $4.6 \cdot 10^6$ pb
Double brin d'ADN circulaire

Chromoïde présent en une copie unique
(2 à 4 copies dans bactéries en croissance)

Associé à des protéines HU
(fonction semblable aux histones eucaryotes, mais
structure différente)

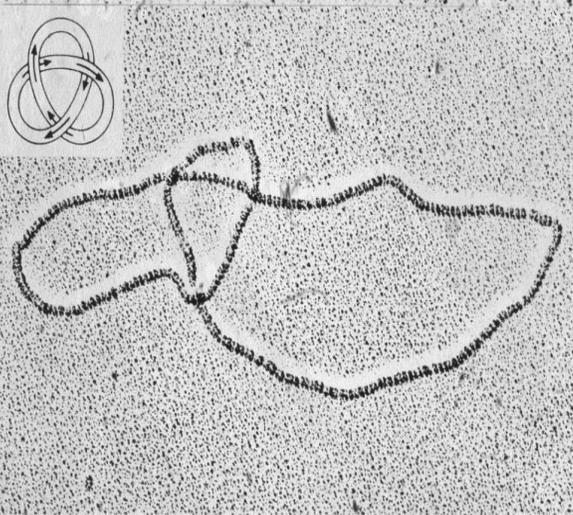
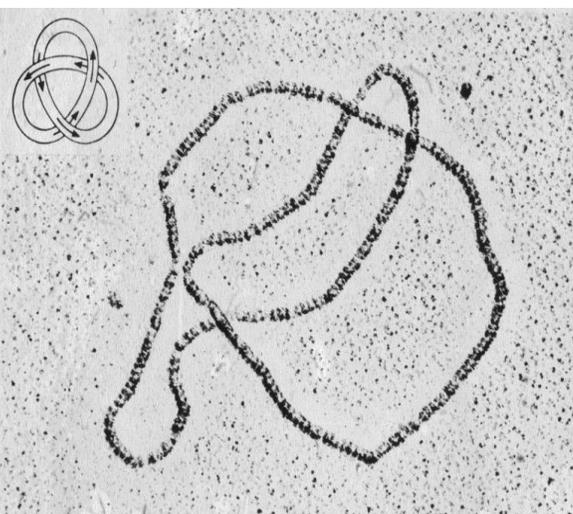
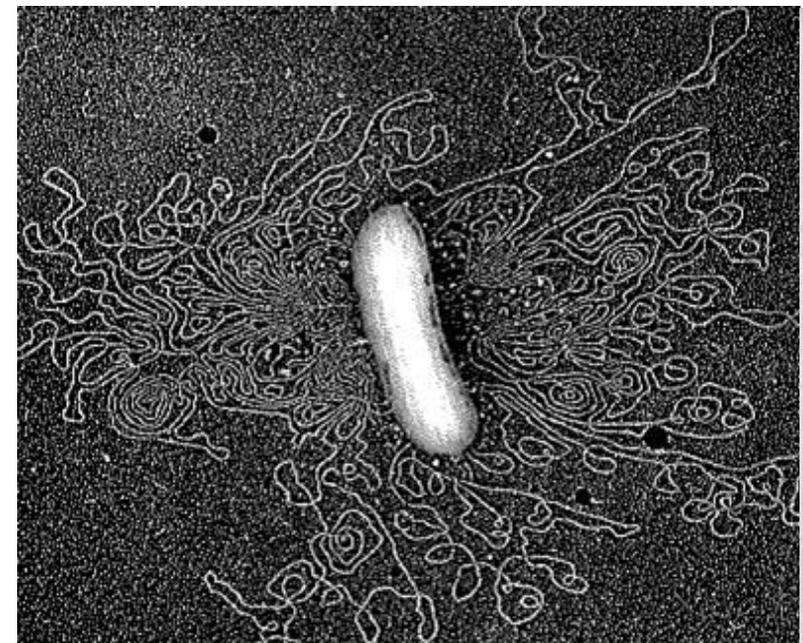
Présence de **surenroulements** de l'ADN
<0 : naturellement
>0 : artificiellement ou temporaire

UN GENOME BACTERIEN : *E. Coli*



Le surenroulement de l'ADN est important pour son conditionnement au sein des cellules. Une molécule d'ADN pouvant être mille fois plus longue que le diamètre d'une cellule, sa condensation pour qu'elle occupe un espace à la mesure de celui d'une cellule est une opération critique. C'est à ce niveau qu'intervient le surenroulement de l'ADN, dans la mesure où il en réduit sensiblement l'encombrement.

UN GENOME BACTERIEN : *E. Coli*



Right- and left-handed dsDNA trefoil knots generated by *E. coli* DNA topoisomerase I acting on form II DNA

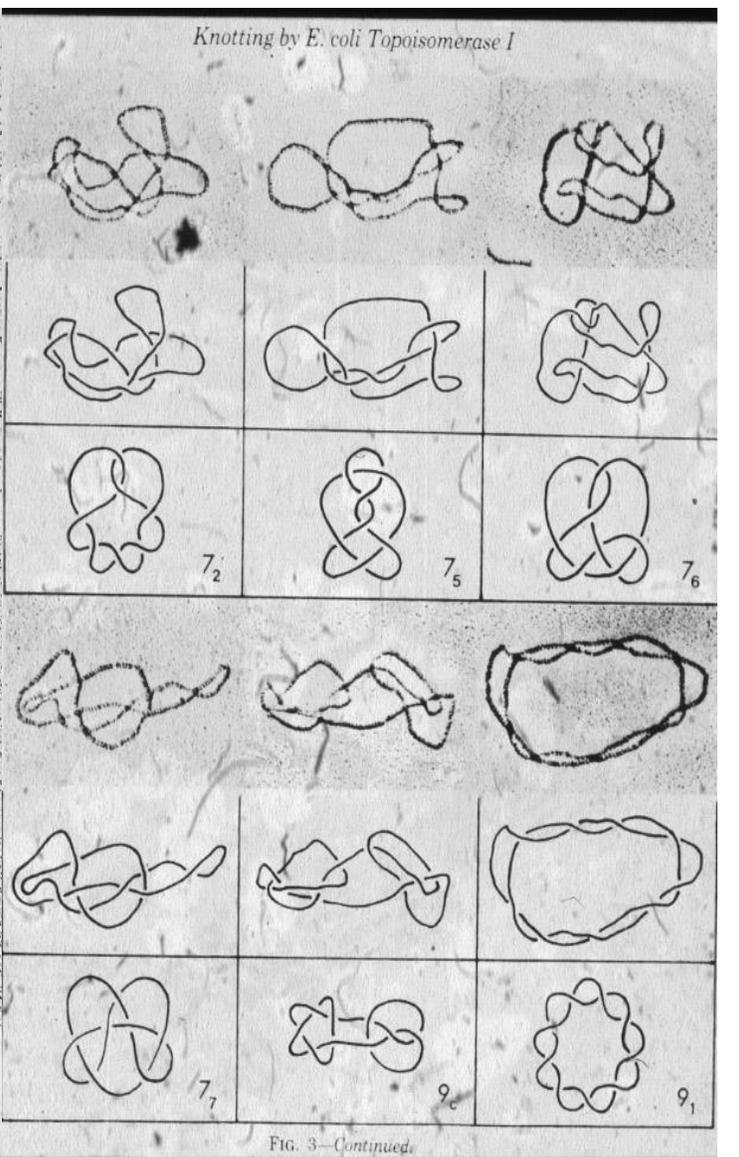


FIG. 3—Continued.

TOPOLOGIE D'UN ADN VIRAL CIRCULAIRE

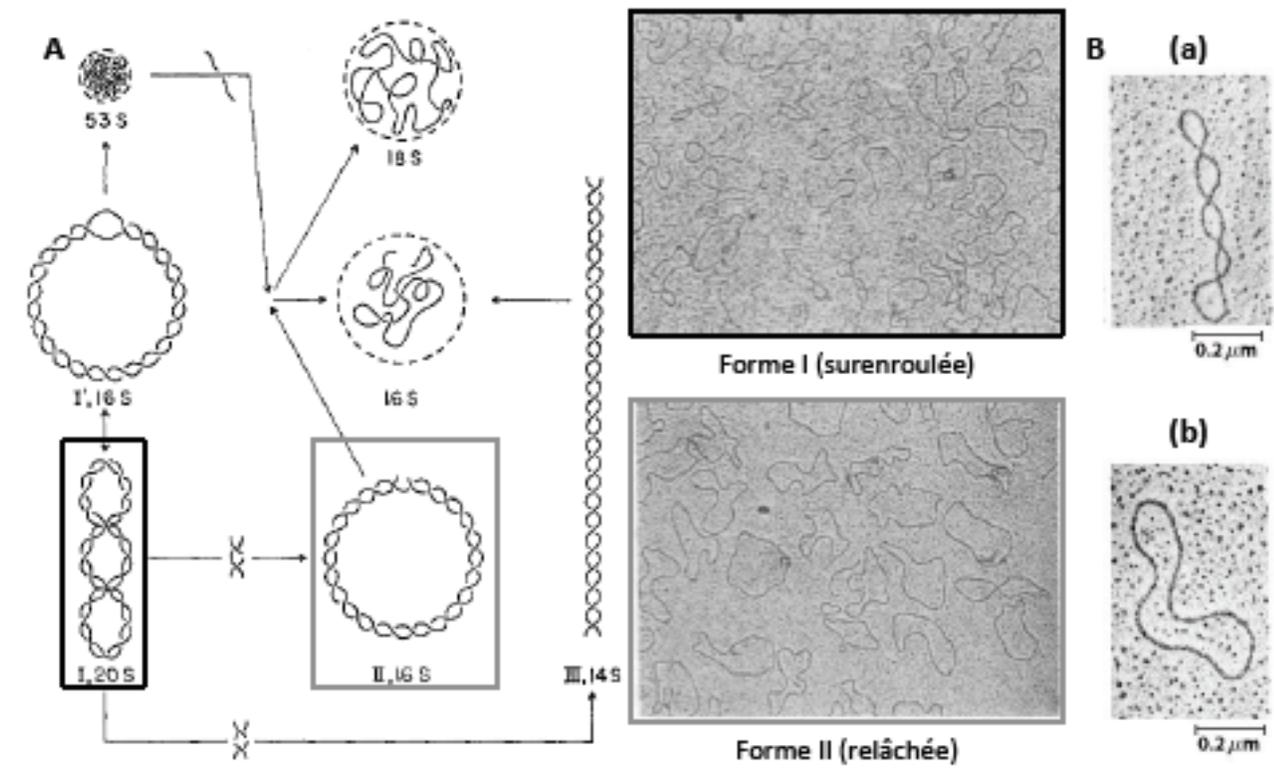


Figure 4 : l'ADN procaryote est surenroulé. A : représentation des différentes formes de l'ADN du polyomavirus (I, II et III) suivant leur coefficient de sédimentation, avec les micrographes électroniques ADN x 21000, des formes I (surenroulée, encadrée en noir) et II (relâchée, encadrée en gris) correspondants. Plus la vitesse de sédimentation est élevée, plus l'ADN est compacté donc surenroulé (d'après Vinograd et al, 1965). B : micrographes électroniques du plasmide pBR322 sous forme relâchée (a) et surenroulée (b) correspondant aux formes I et II de l'ADN du polyomavirus respectivement, observées en figure A.

TOPOLOGIE DES OBJETS CIRCULAIRES

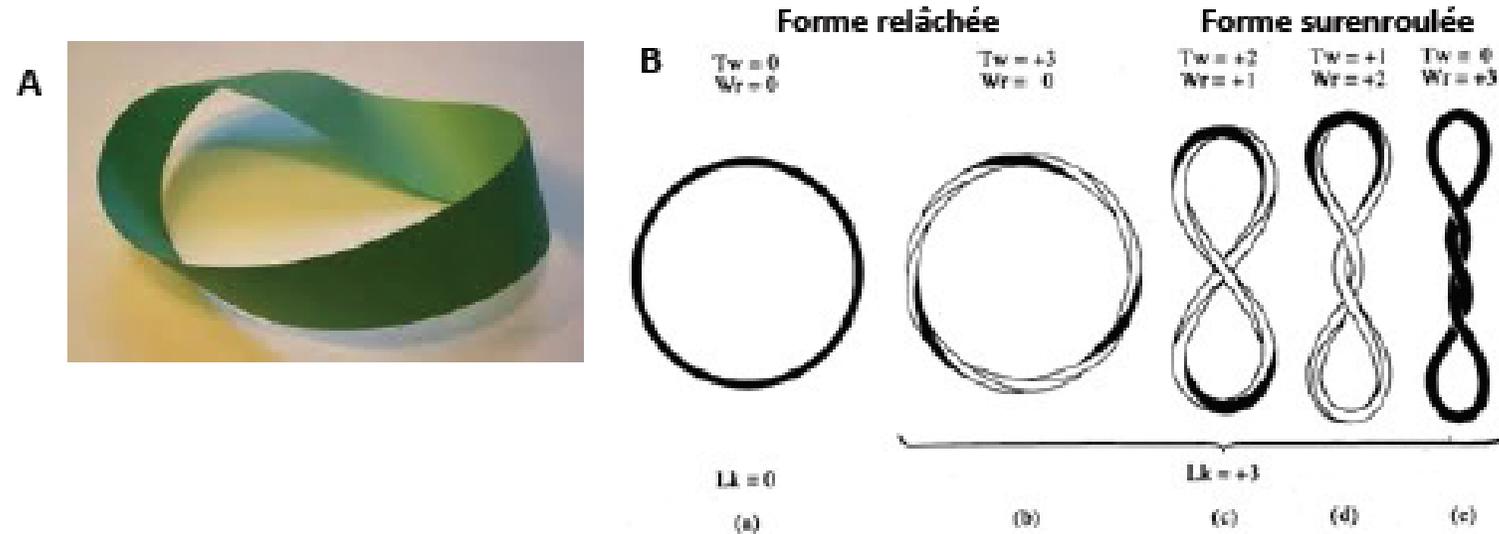


Figure 5 : la molécule d'ADN fermée est un objet topologique. A : photo représentant un ruban de Moëbius. La topologie consiste en l'étude des déformations par transformations continues. Le ruban de Moëbius est un exemple d'objet d'étude de la topologie, comme la molécule d'ADN qui est une structure fermée pouvant se déformer de façon continue sans rupture et réassemblage de brins. B : représentation schématique de la topologie de l'ADN, définie par deux paramètres, le nombre de tours d'hélices T_w et le nombre de supertours W_r , dont la somme définit le nombre total de liens topologiques L_k qui ne varie pas puisque la molécule est fermée. Les figures a et b correspondent à un état relâché puisque $W_r=0$. En b la molécule présente des plis blancs qui figurent trois tours d'hélices, telles deux molécules d'ADN enroulées ($T_w = 3$). En c, d et e, la molécule est surenroulée puisque $W_r \neq 0$. Comme L_k ne varie pas, la déformation de T_w transforme W_r : alors que T_w diminue, W_r augmente. Si T_w devait augmenter, W_r diminuerait. Les plis blancs (b) ont disparu (e), mais la structure n'est pas plane comme en (a), elle est surenroulée : c'est l'effet topologique.

LES ETATS TOPOLOGIQUES DE L'ADN

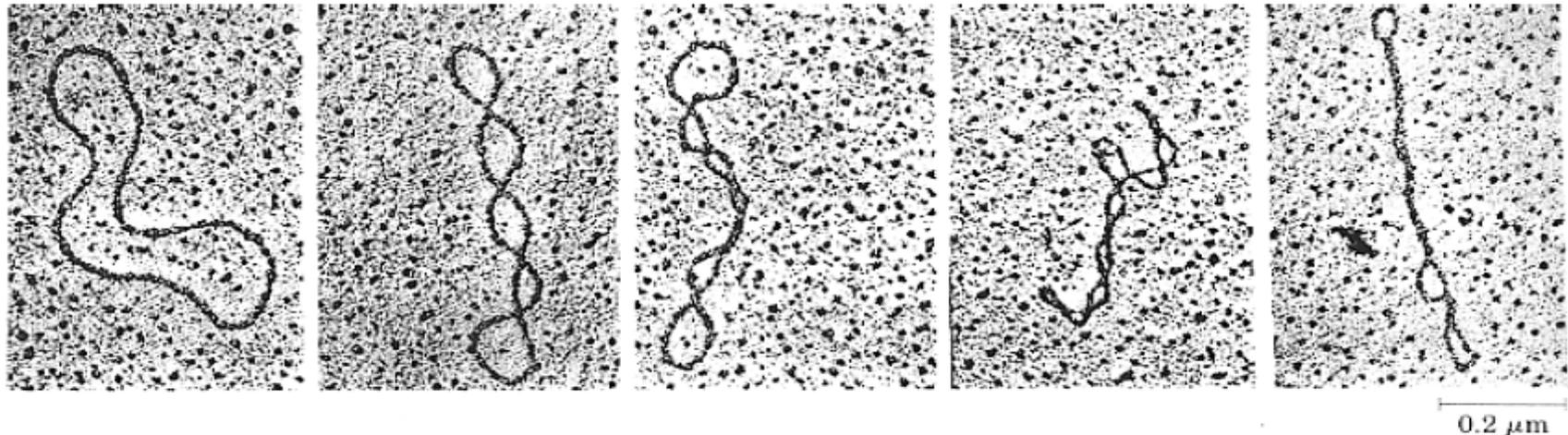
Dans la cellule la même molécule d'ADN double brin peut se présenter sous deux formes (topoisomères) :

état relâché (sans contrainte)

état surenroulé le plus souvent négativement (sens inverse des aiguilles d'une montre)

Forme relâchée

Formes surenroulées



Lehninger et al. Principles of Biochemistry, p.800

*Topoisomères : Deux molécules d'ADN double-brin avec la même séquence
mais avec une conformation différente*

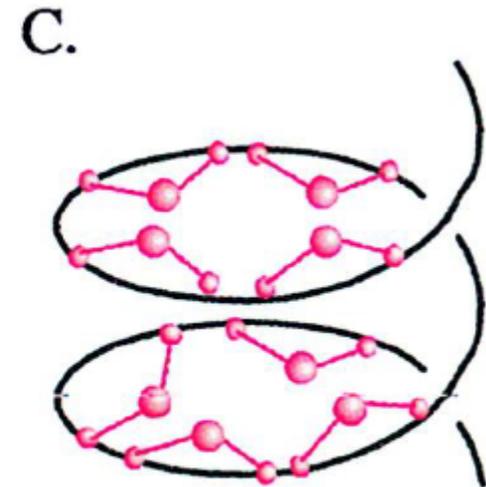
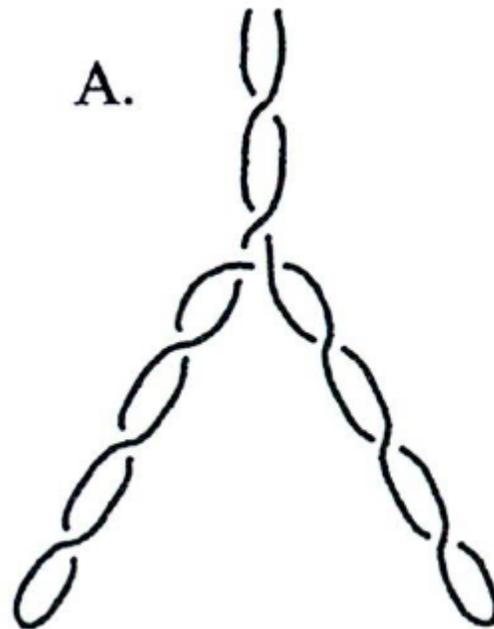
TOURS ET SUPERTOURS

Chaque **topoisomère** est défini par l'**index L** (nombre d'enlacements entre les deux brins).

$L = \text{nb de TOURS (T) entre les deux brins (1 tour tous les 10 nucléotides)} + \text{nb de SUPERTOURS (W) de la double hélice (dans les formes surenroulées)}$.

Deux types de surenroulement:

- **Plectonémique**: en solution, supertours vers la droite
- **Solénoïdal** : dans la cellule, supertours vers la gauche (nucléosome) ou droite



UN GENOME BACTERIEN : *E. Coli*

Tw : « twist number » (nombre de tours)

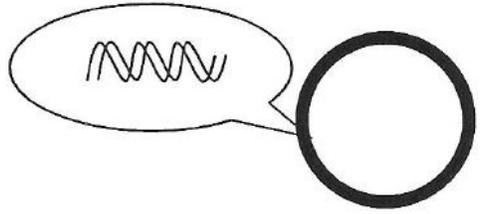
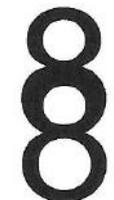
nombre de fois où un brin effectue un tour complet autour de l'autre. Le Tw est proportionnel au nombre de paires de bases appariées.

Wr : « writhe number » (vrillage de l'hélice ou nombre de supertours)

nombre de fois où l'axe de la double hélice effectue un tour complet.

Lk : « linking number » (nombre d'enlacements)

$$L_k = T_w + W_r$$

 <p style="text-align: center;">pas de supertour : $W_r = 0$</p> <p style="text-align: center;">$T_w = N/10,5$</p> <p style="text-align: center;">$L_k = N/10,5 = L_{k0}$</p>	 <p style="text-align: center;">un supertour négatif : $W_r = -1$</p> <p style="text-align: center;">$T_w = N/10,5$</p> <p style="text-align: center;">$L_k = (N/10,5) - 1 = L_{k0} - 1$</p>	 <p style="text-align: center;">deux supertours négatifs : $W_r = -2$</p> <p style="text-align: center;">$T_w = N/10,5$</p> <p style="text-align: center;">$L_k = (N/10,5) - 2 = L_{k0} - 2$</p>
---	---	---

OUVERTURE DE LA DOUBLE HELICE D'ADN

Le surenroulement de l'ADN intervient également lors de la réplication de l'ADN et de sa transcription en ARN.

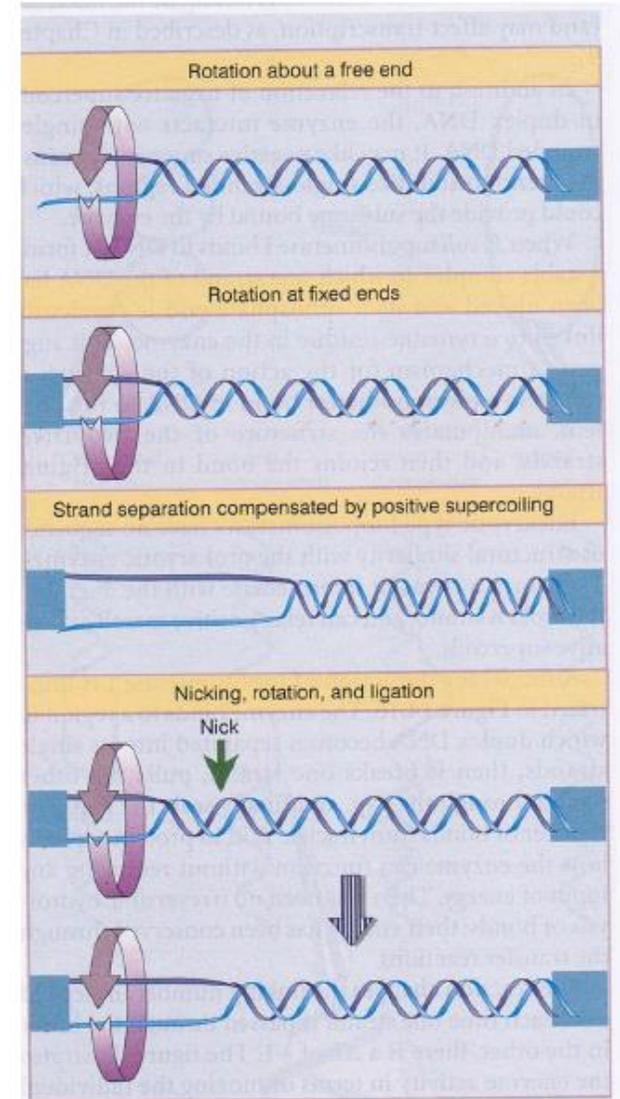
Il résulte du déroulement de la double hélice par l'ADN polymérase et l'ARN polymérase respectivement, déroulement qui introduit des **supertours positifs en amont** et **négatifs en aval**. Des ADN topoisomérases telles que l'ADN gyrase interviennent pour relâcher en partie ce surenroulement.

Toute activité synthétique impliquant l'ADN a besoin de la séparation des deux brins:

réplication
transcription

Comment séparer les brins,

1. Par rotation à partir de terminaisons libres
2. Par rotation à partir de terminaisons fixes
3. Sans rotation: compensation par surenroulement
4. Coupe, rotation, ligation



LES TOPOISOMERASES

-Des enzymes modifient le nb d'enlacements (suppriment ou ajoutent des supertours : les topoisomérases.

Topoisomérase I: coupe sur un brin

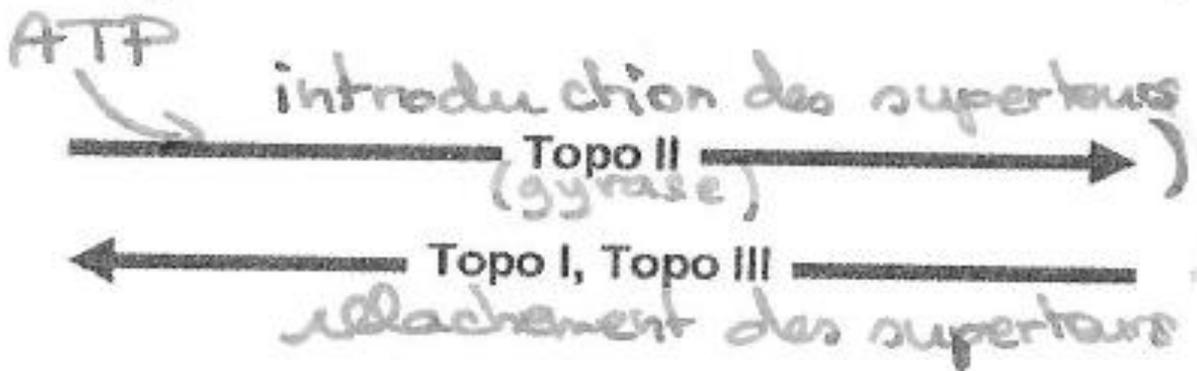
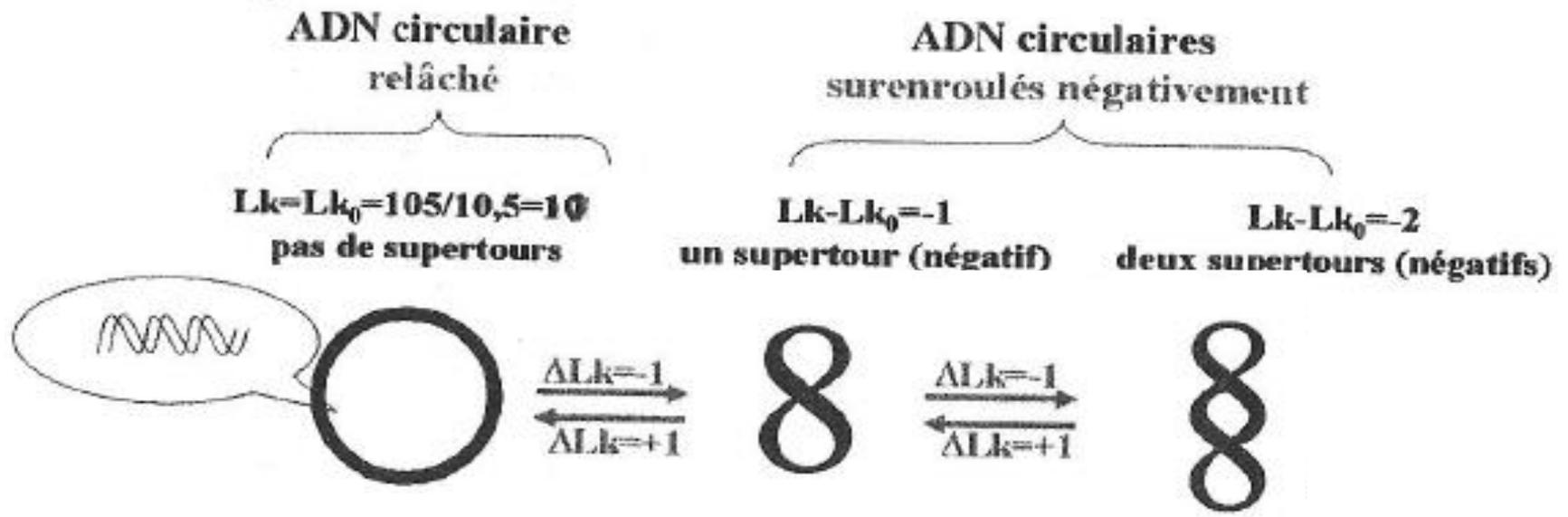
Topoisomérase II: coupe sur les deux brins

puis réassocient les brins après changement du nombre d'enlacements. La plupart agissent en diminuant le nb de tours

- Bactéries:
 - type I: topoisomérase I, III
 - type II: DNA gyrase, topoisomérase IV
- Eucaryotes:
 - type I: topoisomérase I, III
 - type II: topoisomérases II α , II β

Fonctions: réplication, transcription, réparation de l'ADN

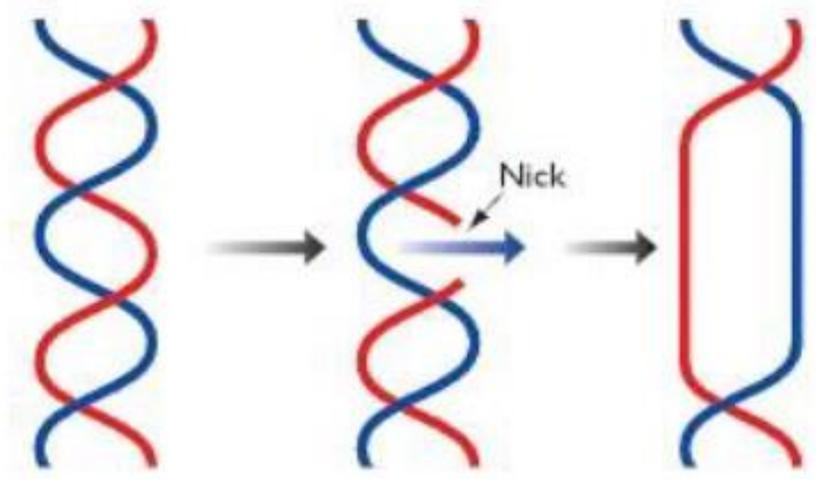
LES TOPOISOMERASES



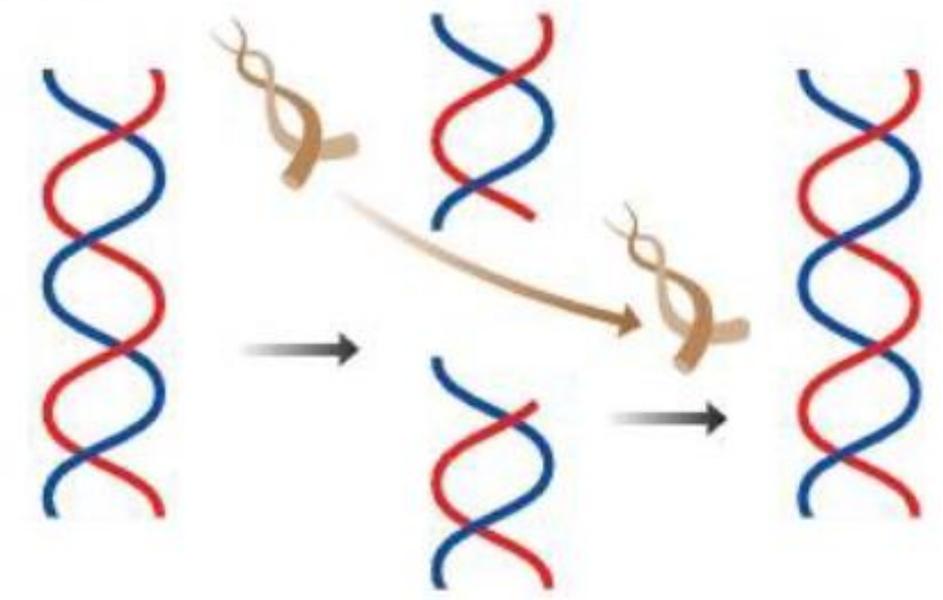
LES TOPOISOMERASES

Topoisomérases I et II

(A) Type I

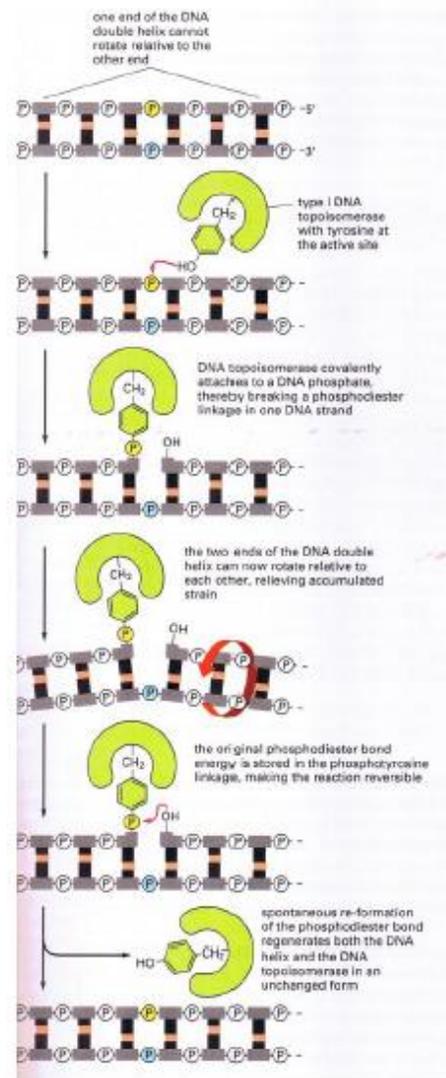


(B) Type II



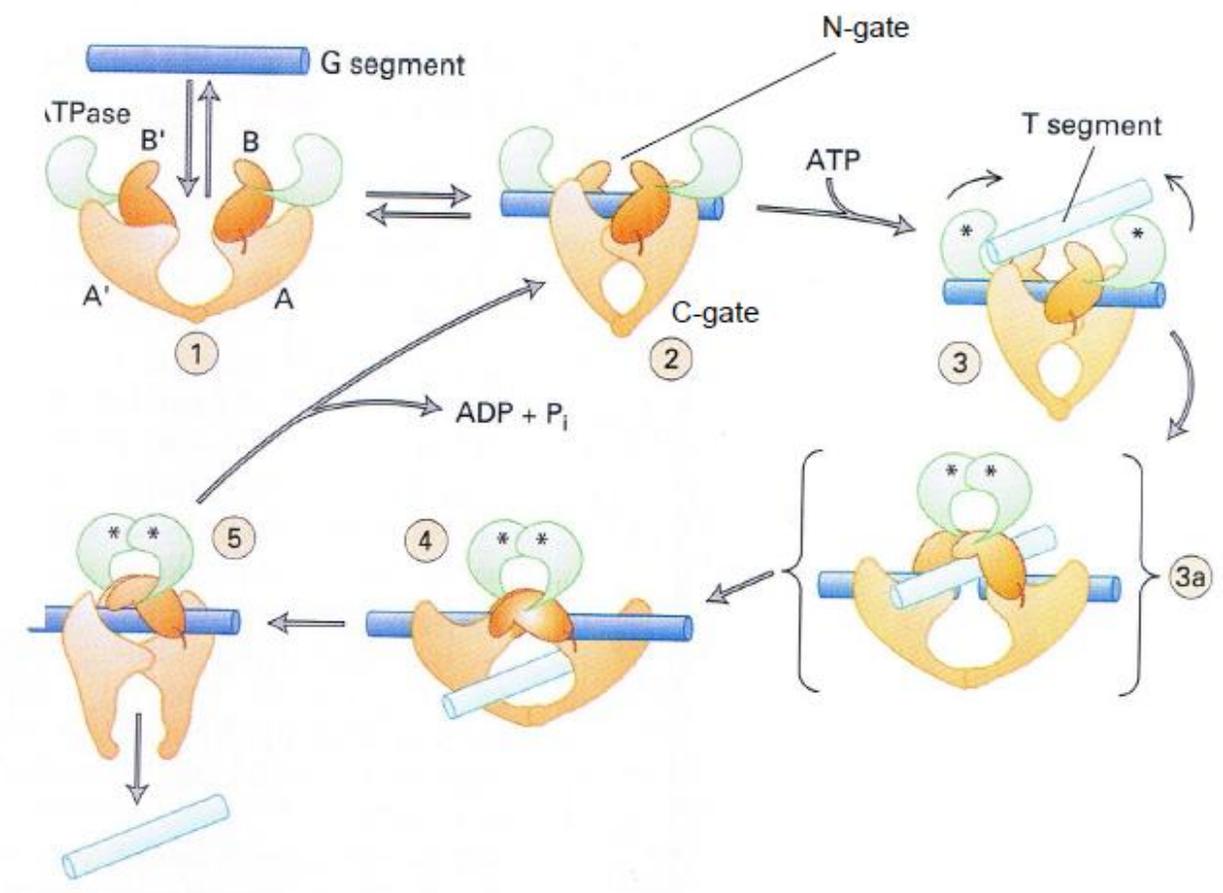
LES TOPOISOMERASES

Topoisomérase I



LES TOPOISOMERASES

Topoisomérase II (gyrase)



*N Patrick Higgins
Nature Structural & Molecular Biology 14, 256 - 258 (2007)*

ENROULEMENT ET SURENROULEMENT

Surenroulement: ADN circulaire (génome procaryotes, mitochondrial eucaryotes, chloroplastes, plasmides)
ADN linéaire (eucaryotes): contraintes créées par des complexes protéiques associés (réplication) impliqué dans la formation du chromosome

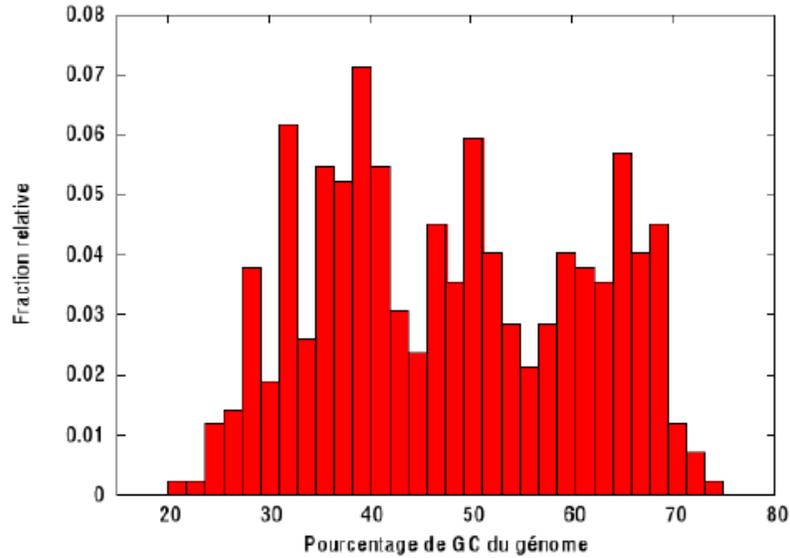
Enroulement: Température: réduit l'enroulement
Force ionique: augmente l'enroulement
Intercalants (bromure d'éthidium): déroulement local double hélice

Topoisomerases

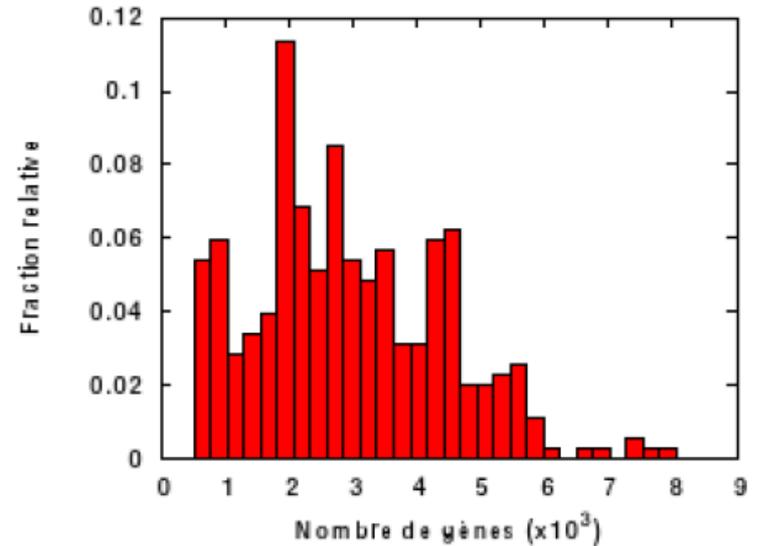
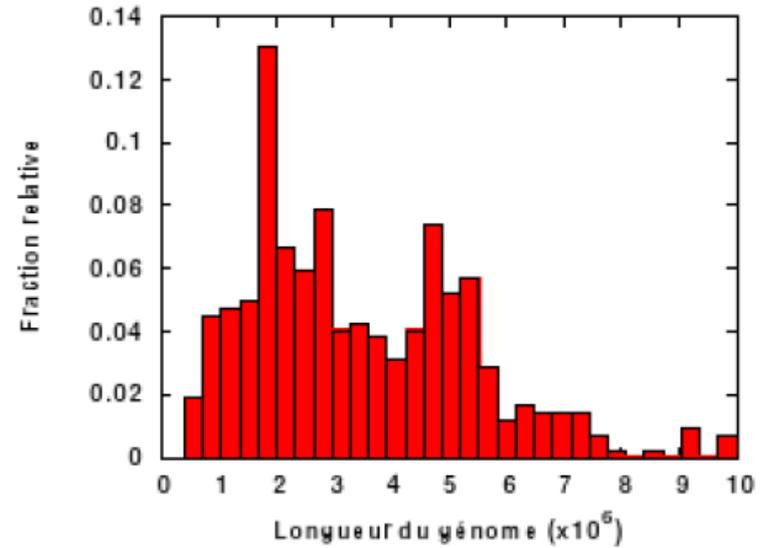


2. ANATOMIE DES GENOMES PROCARYOTES

DIVERSITE DES GENOMES BACTERIENS

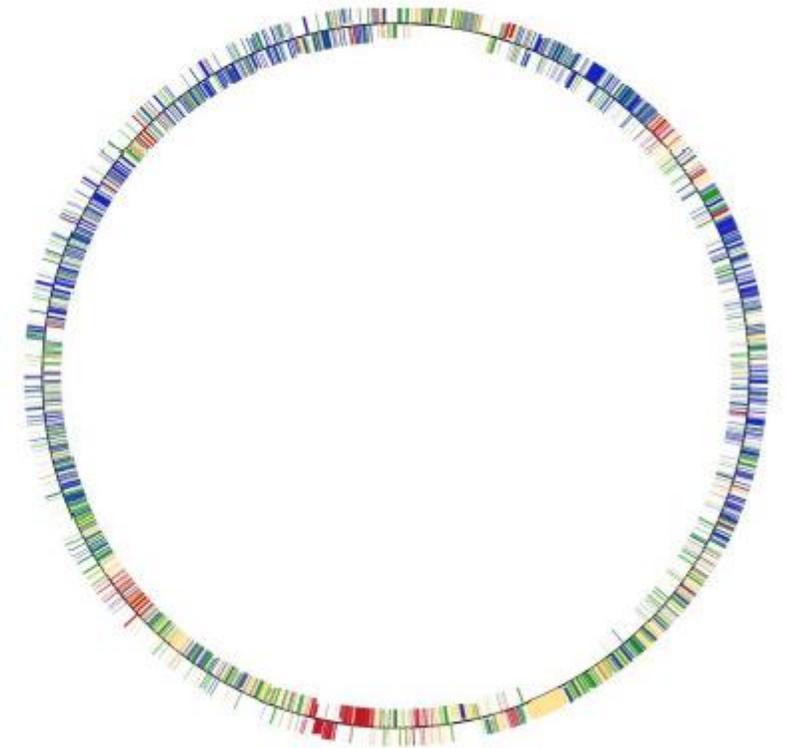


Le nombre de plasmides des génomes bactériens varie de 0 à plus de 20 (dans ce cas 36% du génome)



UN GENOME BACTERIEN : *E. Coli*

Genome Info:	Features:
Refseq: NC 000913	Genes: 4493
GenBank: U00096	Protein coding: 4149
Length: 4,639,675 nt	Structural RNAs: 172
GC Content: 50%	Pseudo genes: 177
% Coding: 85%	Others: 584



UN GENOME BACTERIEN : *L. plantarum*

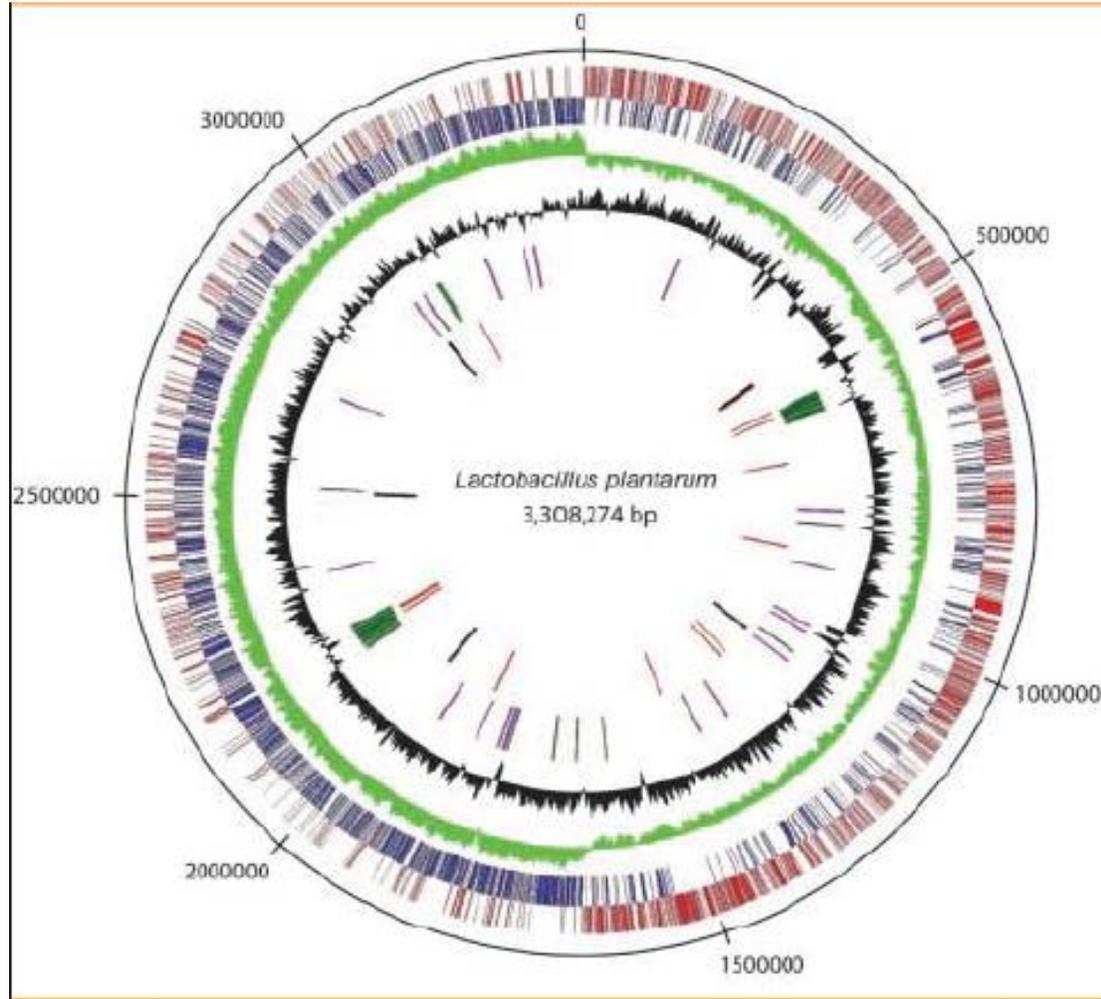
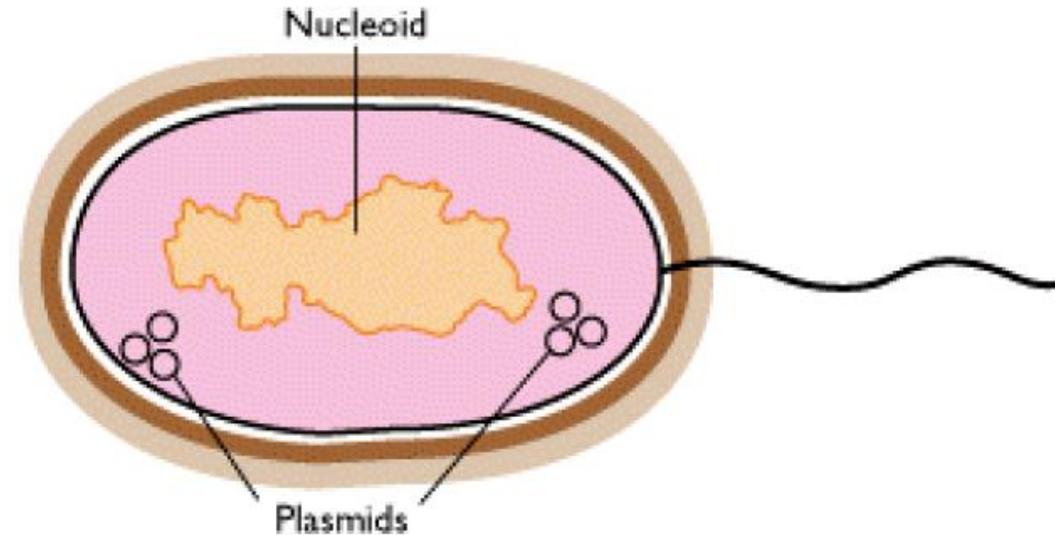


Fig. 1. Genome-atlas view of the *L. plantarum* WCFS1 chromosome, with the predicted origin of replication at the top. The outer to inner circles show (i) positive strand ORFs (red); (ii) negative strand ORFs (blue); (iii) GC-skew (green); (iv) G+C content (black); (v) prophage-related functions (green) and IS-like elements (purple); and (vi) rDNA operons (black) and tRNA encoding genes (red).

Kleerebezem et al. PNAS 2003

STRUCTURE DES L'ADN BACTERIEN



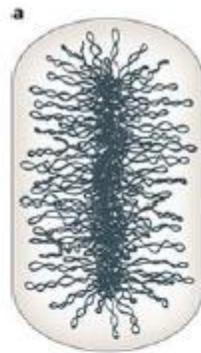
- 1 seul chromosome circulaire
 - . Contient presque exclusivement des gènes
 - . Nombreux gènes organisés en opéron
 - . Pas d'introns dans les gènes
 - . Moins de gènes que chez les eucaryotes
 - . Condensation pour former le nucléoïde par
 - . Supercoiling
 - . association avec des protéines HU, fonctionnellement similaires aux histones mais non homologues en séquence. 1/5 seulement du chromosome est associé aux HU
- 1 ou plusieurs plasmides contenant des gènes de résistance aux antibiotiques

STRUCTURE DE L'ADN BACTERIEN

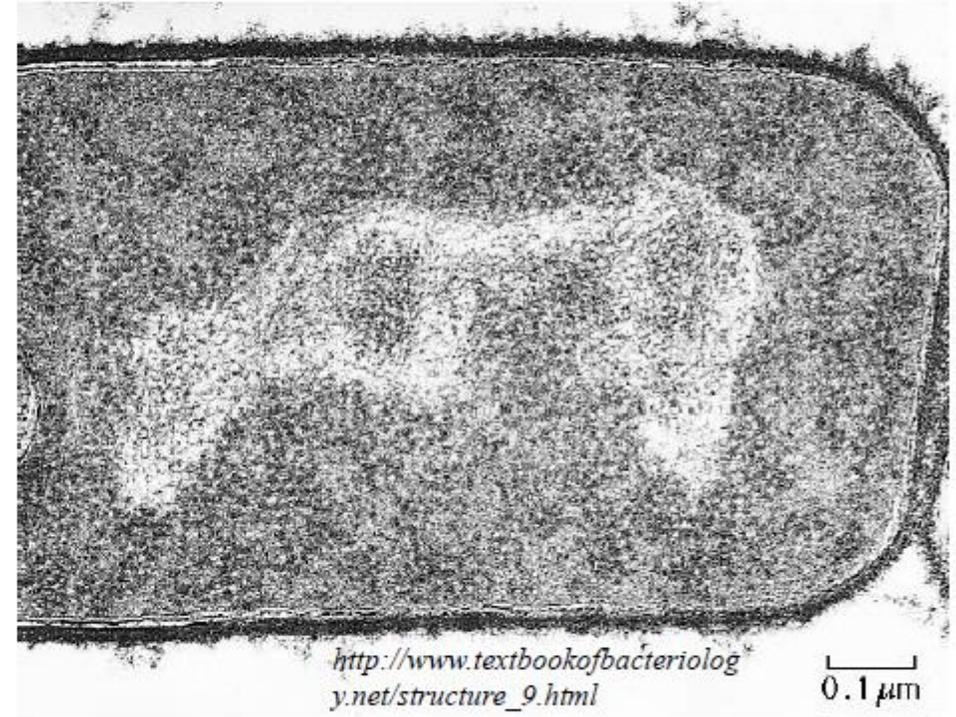
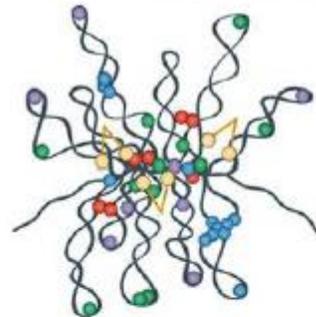
-ADN du « chromosome » bactérien ou nucléoïde (4, 6 Mbp, chez Coli), le plus souvent circulaire attaché à la membrane, associé à des protéines et organisé en domaines (boucles) superenroulés

Gènes (90 % du génome bactérien) souvent organisés en opéron (1 promoteur commun à plusieurs gènes)

-Plasmides : molécule d'ADN circulaire de petite taille (quelques kbp) pouvant se répliquer de façon autonome dans la bactérie à partir d'une origine de réplication unique. Un plasmide peut contenir plusieurs gènes.



b Nucleoid-associated proteins and SMC



Certains plasmides, après modifications, sont utilisés comme vecteur pour transporter des fragments d'ADN lors du clonage moléculaire.

CARTOGRAPHIE DES GENOMES

⇒ Toutes les séquences chez E. coli sont **uniques** ($4,2 \cdot 10^6$ bp)
900 gènes
dont 7 codant pour ARNr et prot ribosomiales (proche oriC).

~4200 protéines

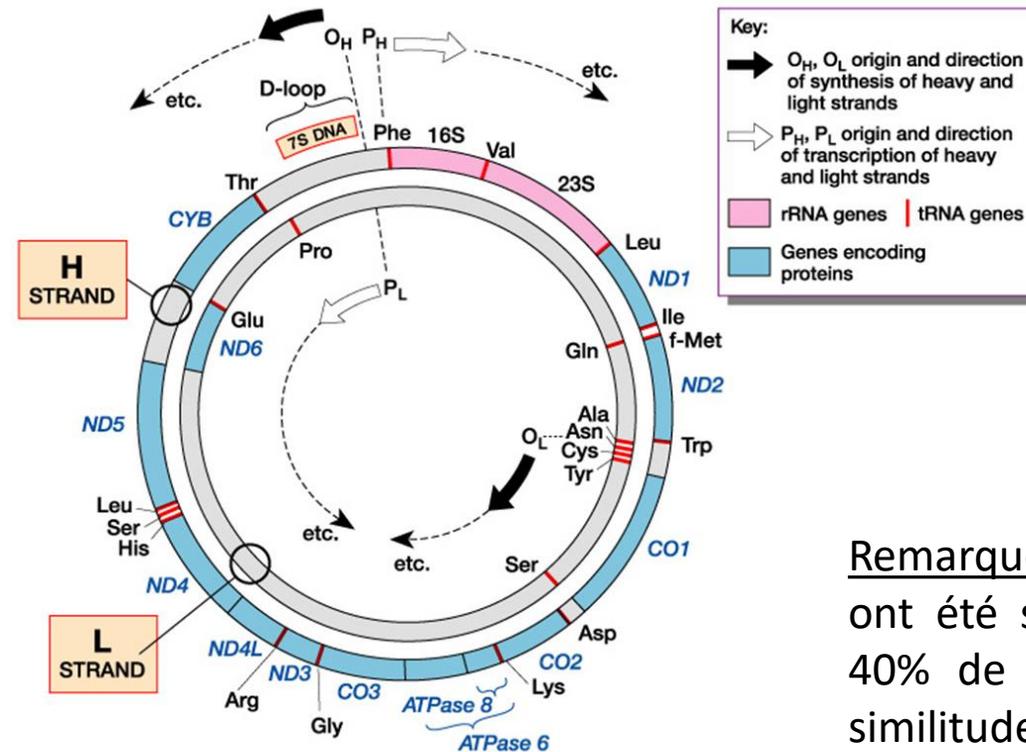
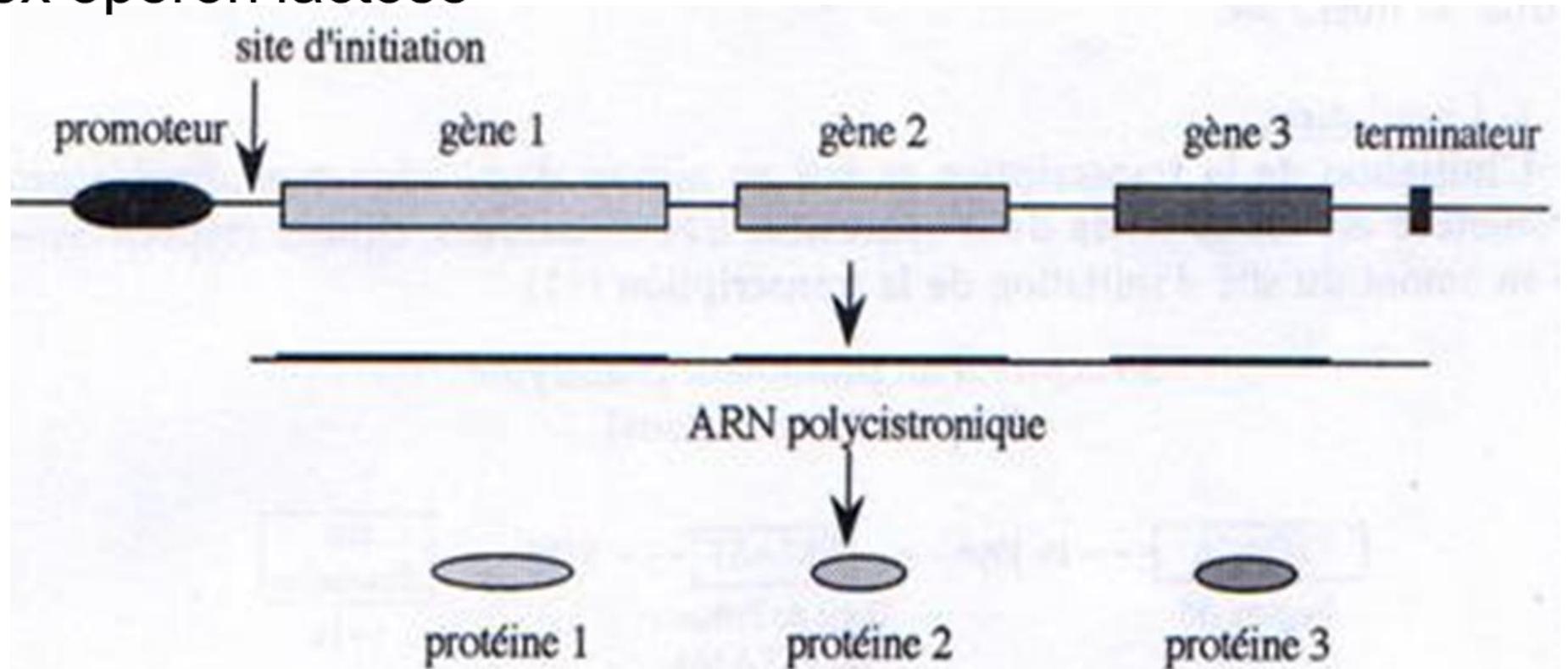


Figure 9-2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

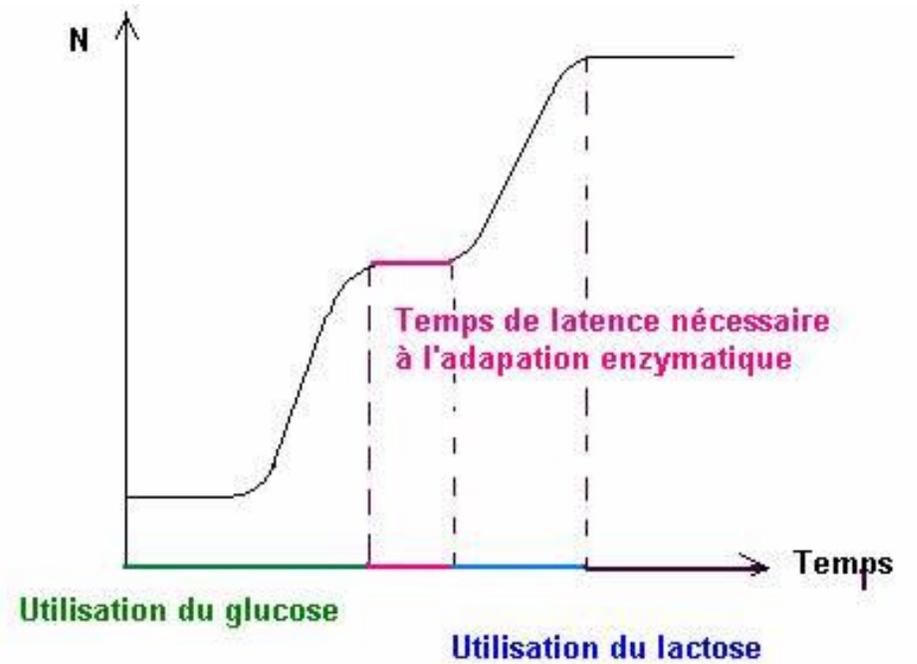
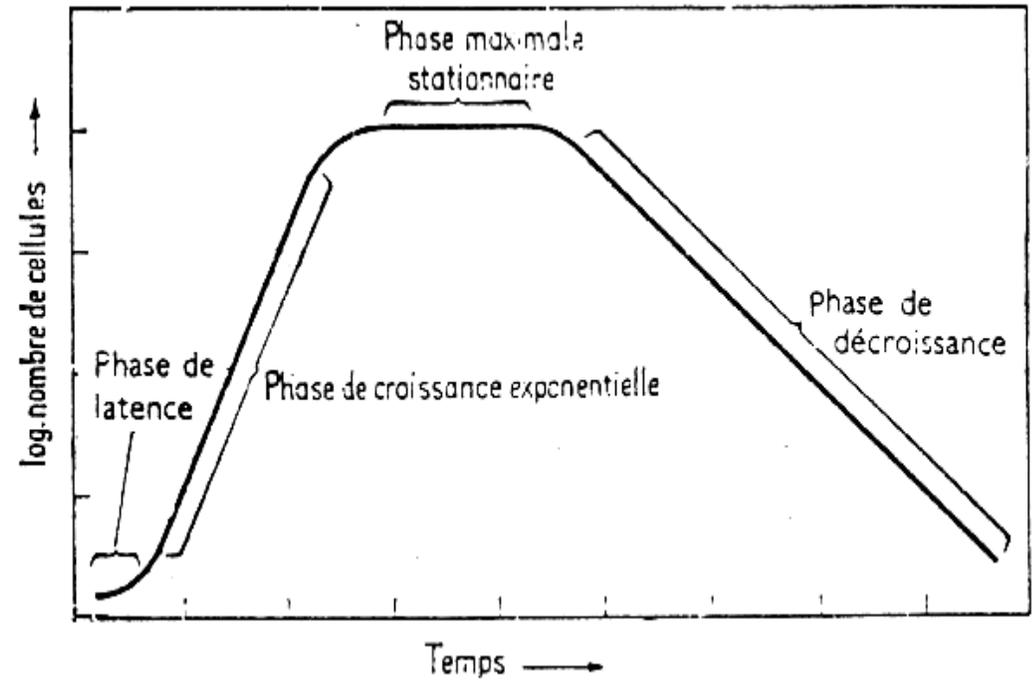
Remarque : Des souches pathogènes d'E. coli ont été séquencées ; elles n'ont qu'environ 40% de leurs gènes en commun (99% de similitudes entre l'Homme et les grands singes...).

ORGANISATION DES GENES EN OPERONS

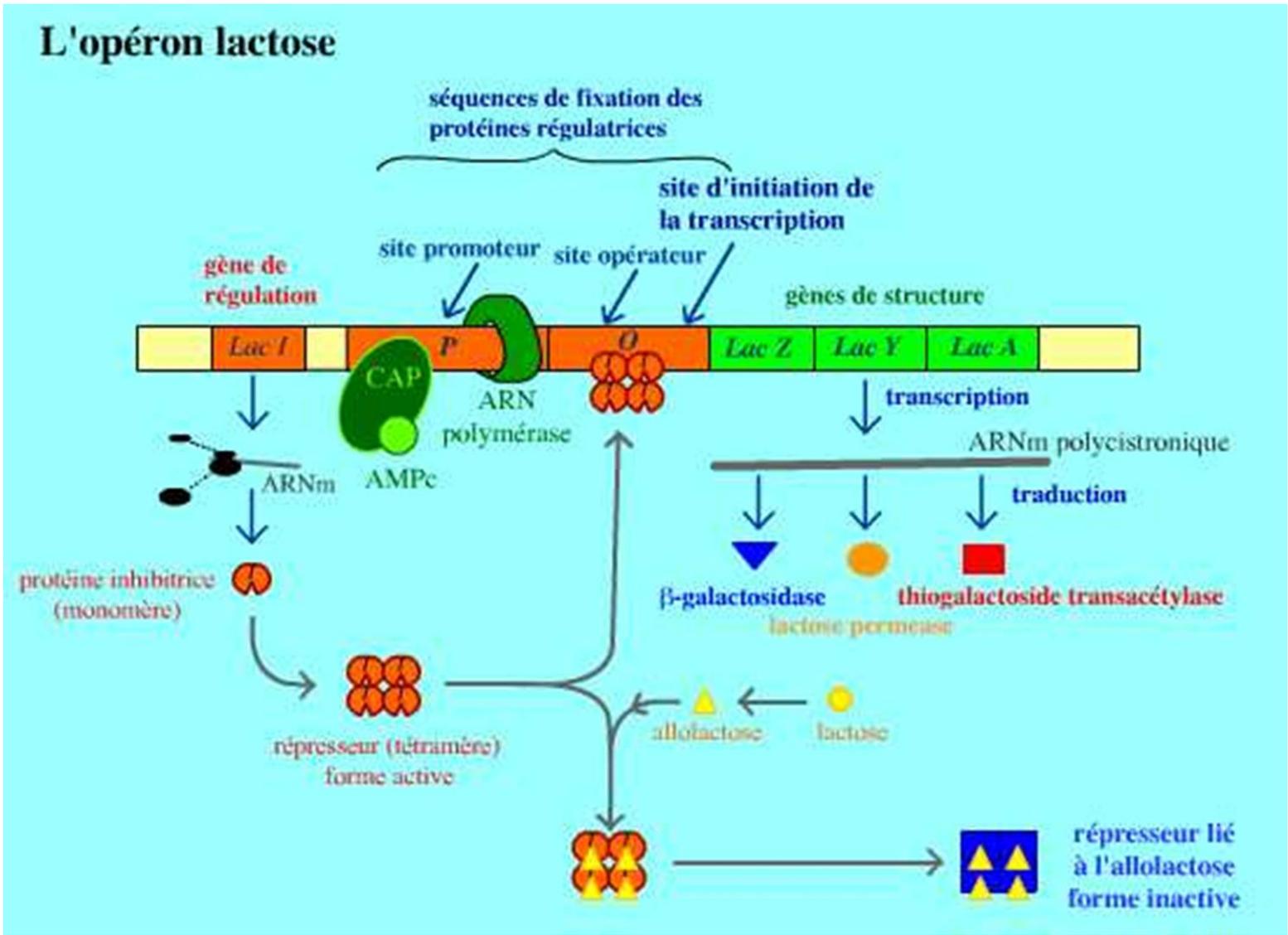
les gènes bactériens peuvent être **polycistroniques** :
ex opéron lactose



PHENOMENE DE DIAUXIE



L'OPERON LACTOSE



LES PLASMIDES

Plasmide :

molécule d'ADN, généralement circulaire, douée de répllication autonome et transmise au cours des générations

- ⇒ certains plasmides peuvent s'intégrer au chromosome, d'autres restent indépendants
- ⇒ portent souvent des gènes non essentiels dans des conditions « normales »
- ⇒ certains ont la possibilité de passer d'une cellule à l'autre (y compris entre espèces)

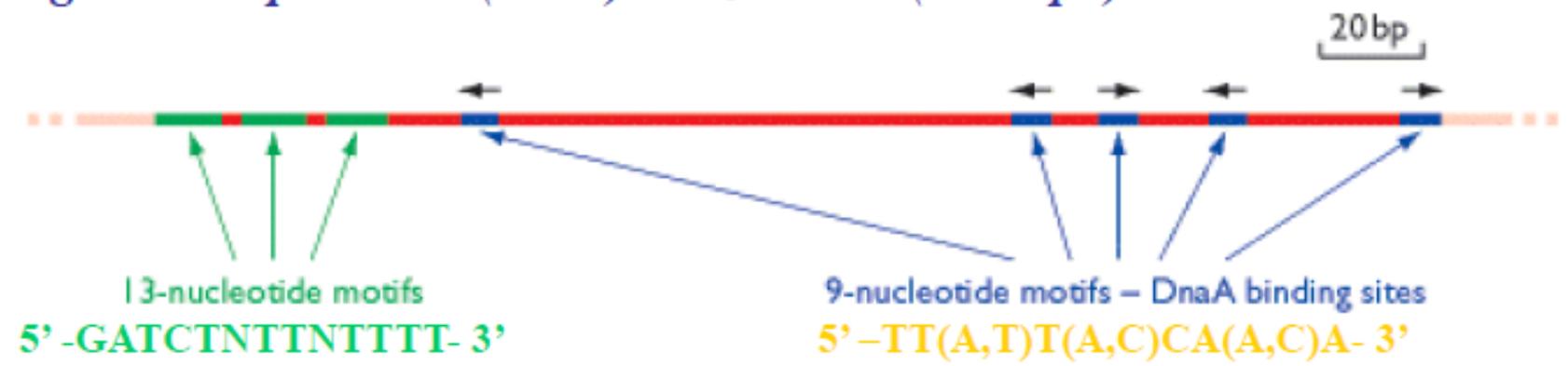
Type of plasmid	Gene functions	Examples
Resistance	Antibiotic resistance	Rbk of <i>Escherichia coli</i> and other bacteria
Fertility	Conjugation and DNA transfer between bacteria	F of <i>E. coli</i>
Killer	Synthesis of toxins that kill other bacteria	Col of <i>E. coli</i> , for colicin production
Degradative	Enzymes for metabolism of unusual molecules	TOL of <i>Pseudomonas putida</i> , for toluene metabolism
Virulence	Pathogenicity	Ti of <i>Agrobacterium tumefaciens</i> , conferring the ability to cause crown gall disease on dicotyledonous plants

LES PLASMIDES

Species	DNA molecules	Size (Mb)	Number of genes
<i>Escherichia coli</i> K-12	One circular molecule	4.639	4397
<i>Vibrio cholerae</i> El Tor N16961	Two circular molecules		
	Main chromosome	2.961	2770
	Megaplasmid	1.073	1115
<i>Deinococcus radiodurans</i> R1	Four circular molecules		
	Chromosome 1	2.649	2633
	Chromosome 2	0.412	369
	Megaplasmid	0.177	145
	Plasmid	0.046	40
<i>Borrelia burgdorferi</i> B31	seven or eight circular molecules, 11 linear molecules		
	Linear chromosome	0.911	853
	Circular plasmid cp9	0.009	12
	Circular plasmid cp26	0.026	29
	Circular plasmid cp32*	0.032	Not known
	Linear plasmid lp17	0.017	25
	Linear plasmid lp25	0.024	32
	Linear plasmid lp28-1	0.027	32
	Linear plasmid lp28-2	0.030	34
	Linear plasmid lp28-3	0.029	41
	Linear plasmid lp28-4	0.027	43
	Linear plasmid lp36	0.037	54
	Linear plasmid lp38	0.039	52
Linear plasmid lp54	0.054	76	
Linear plasmid lp56	0.056	Not known	

L'ORIGINE DE REPLICATION

Une seule origine de réplication (OriC) chez E. coli (~250 pb)



Reconnaissance de origine de réplication par la protéine DnaA => ouverture de la double hélice...

AUTRES GENOMES PROCARYOTES

- Quelques espèces de bactéries possèdent un grand chromosome linéaires, et plusieurs dizaines de petits chromosomes supplémentaires, linéaires ou circulaires.
- Les archées ont, selon les cas, une organisation génomique semblable à celle d'E. coli ou à celle de ces bactéries atypiques. Contrairement aux eubactéries, on trouve parfois des introns dans les gènes d'archées. Il y a également des protéines homologues aux histones.
- De manière générale, les génomes procaryotes évoluent plus rapidement que les génomes eucaryotes car les taux de mutation sont plus élevés, les temps de doublement plus rapides et les transferts horizontaux de matériel génétique sont possibles (ce qui rend la notion d'espèce procaryote très floue...)

AUTRES GENOMES PROCARYOTES

Une situation plus complexe chez les Archaea...

Table 1. A comparison of chromosomal DNA replication in the three domains of life^{a,b}

	Bacteria	Archaea	Eukarya
Chromosome	Linear or circular	Circular	Linear
Replication origin(s)	Single	Single or multiple	Multiple
Origin recognition	DnaA	Cdc6/ORC ^c	ORC
Helicase	DnaB	MCM	MCM
Helicase loader	DnaA and DnaC	Cdc6/ORC ^c	Cdc6 and Cdt1 ^d
Single-stranded DNA-binding protein	SSB	SSB or RPA ^e	RPA
Primase	DnaG	Primase	Pol α /Primase
Sliding clamp	β subunit	PCNA ^f	PCNA
Clamp loader	γ -complex	RFC	RFC
Polymerase	PolC	PolD and/or PolB ^g	PolB

Kelman & Kelman, Trends in Microbiology 2004

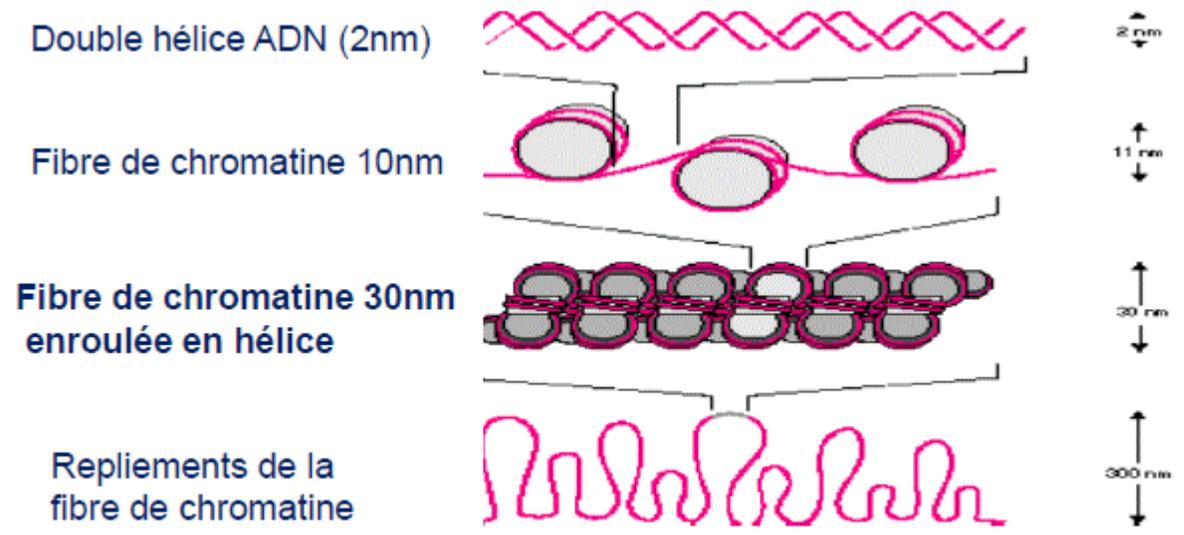
2. ANATOMIE DES GENOMES EUCARYOTES

STRUCTURE DE L'ADN NUCLEAIRE EUCLARYOTE

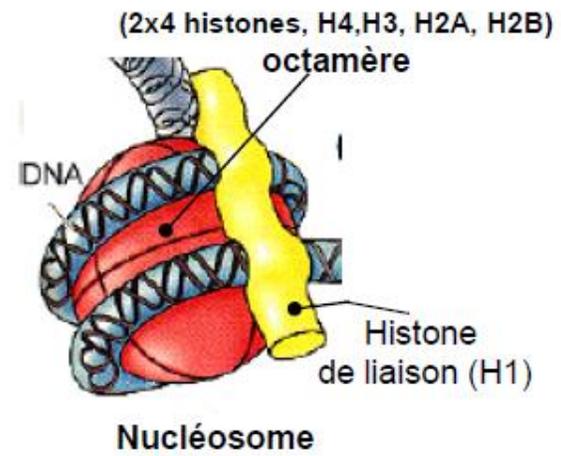
Génome chromosomique

L'ADN génomique, associé à des protéines (histones), est réparti sur plusieurs molécules linéaires de **chromatine**. Chez l'homme 23 molécules par génome haploïde .

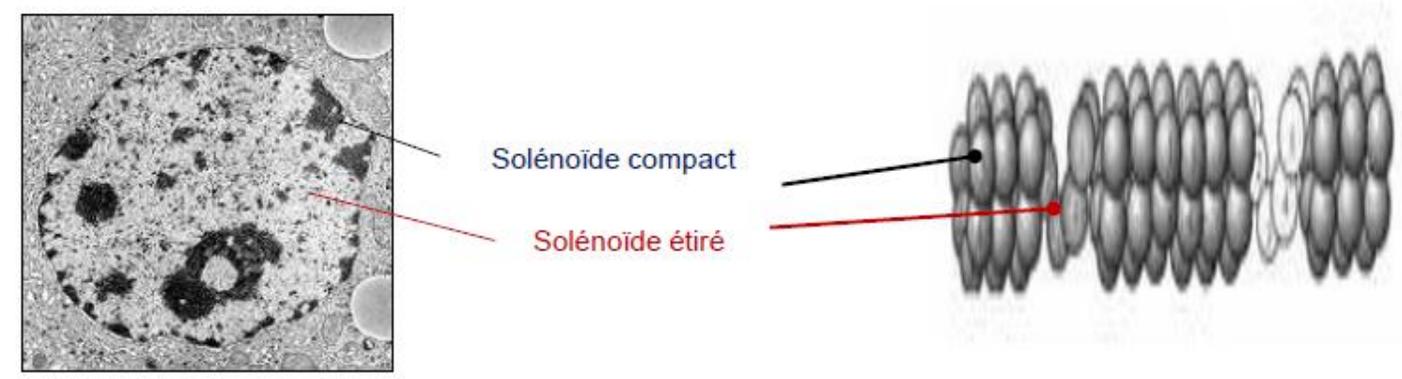
En interphase (G1), chaque molécule de chromatine est une fibre de 30 nm. Elle est constituée de nucléosomes enroulés autour d'un axe et formant une structure « solénoïde » (6 nucléosomes par tour). Elle est repliée en boucles contenant plusieurs tours (environ 50-90 tours correspondant à environ 50 à 100 kbp).



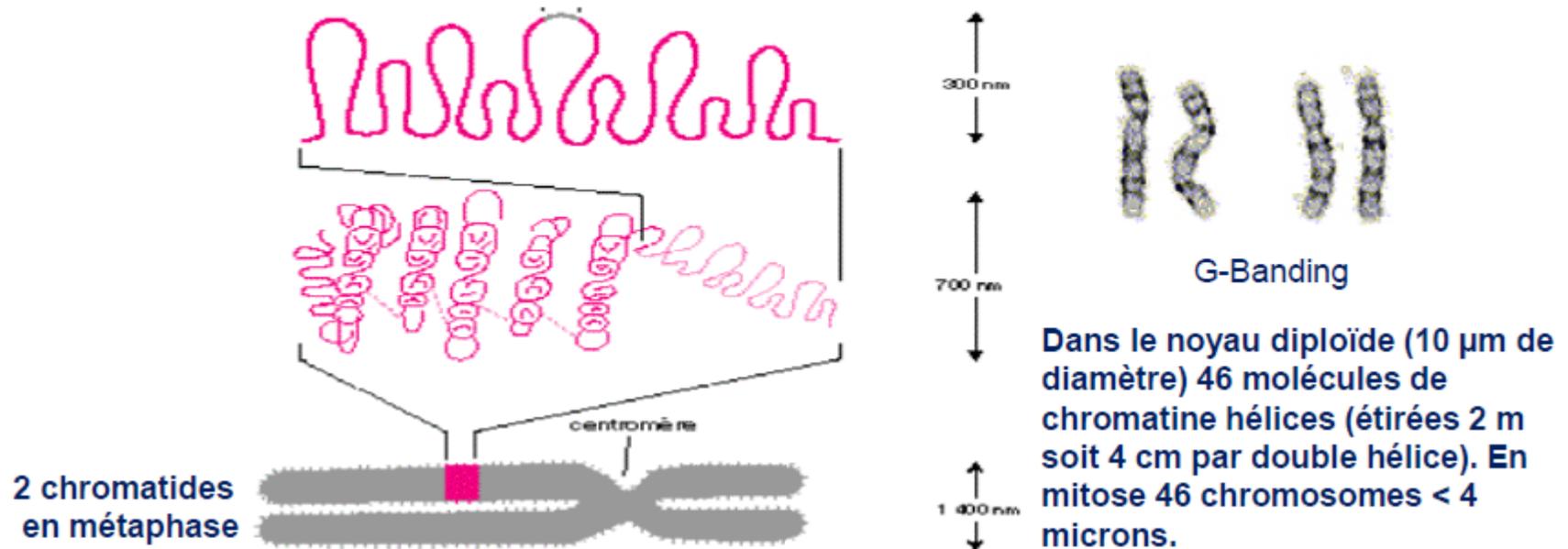
STRUCTURE DE L'ADN NUCLEAIRE EUKARYOTE



En interphase la fibre de chromatine visible en microscopie électronique peut être très compacte (**hétérochromatine**) ou plus relâchée (**euchromatine**). Seuls les gènes contenus dans l'euchromatine peuvent être exprimés.



CONDENSATION DE LA CHROMATINE EN MITOSE



Après la réplication de l'ADN à l'interphase (phase S), la fibre de chromatine est double. Ces deux fibres associées entre elles vont se condenser pour donner un **chromosome qui apparaît en métaphase** sous la forme de deux chromatides associées par le centromère (visible en microscopie optique).

Des techniques de coloration permettent de visualiser des régions correspondant à l'hétérochromatine

CONDENSATION DE LA CHROMATINE EN MITOSE



Chez l'homme, chaque cellule **diploïde** contient **deux génomes** soit **23 paires de fibres de chromatine** après la division cellulaire et avant la réplication de l'ADN. Chaque séquence (chaque gène) sur un locus donné existe donc en deux copies homologues : deux allèles issus des génomes paternel et maternel qui sont le plus souvent identiques (homozygotes) ou légèrement différents (hétérozygotes, allèles polymorphes)

CONDENSATION DE LA CHROMATINE EN MITOSE

Pour être fonctionnelle, chaque molécule de chromatine doit contenir au moins :

- Un **centromère**, séquence ADN satellite constituée d'un motif de 180 nucléotides répété (1 à 10 Mbp) qui permet l'association des chromatides entre elles et leur fixation sur le fuseau mitotique.

- Des **télomères** à chaque extrémité, ADN minisatellite (répétitions de motifs de 6 nucléotides) qui protègent ces extrémités et permettent leur réplication.

- Des **origines de réplication** où se fixe un complexe protéique nécessaire au démarrage de la réplication. Ces origines sont constituées des différents éléments riches en A et T. Quatre pour l'origine de réplication chez la levure (130 nucléotides)

GENOME MITOCHONDRIAL

- ADN circulaire de 17 kbp (10 à 100000 copies/mitochondrie) avec 2 origines de réplication.
- 3 promoteurs (2+1) fonctionnant dans des directions opposées pour la transcription de 2 ARN:
 - PH1
 - PH2
 - PL



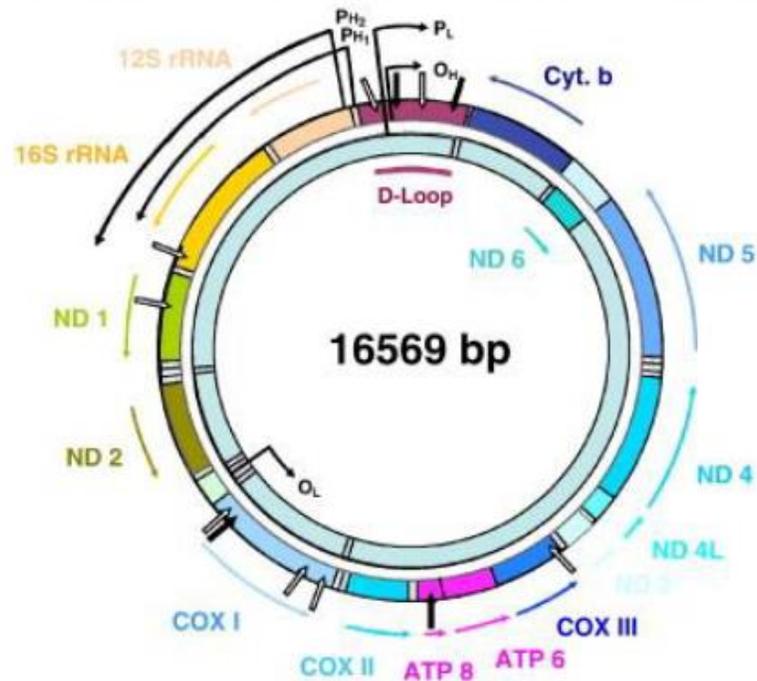
<http://bio3400.nicerweb.com/Locked/media/ch09/DNA-mitochondrion.html>

GENOME MITOCHONDRIAL

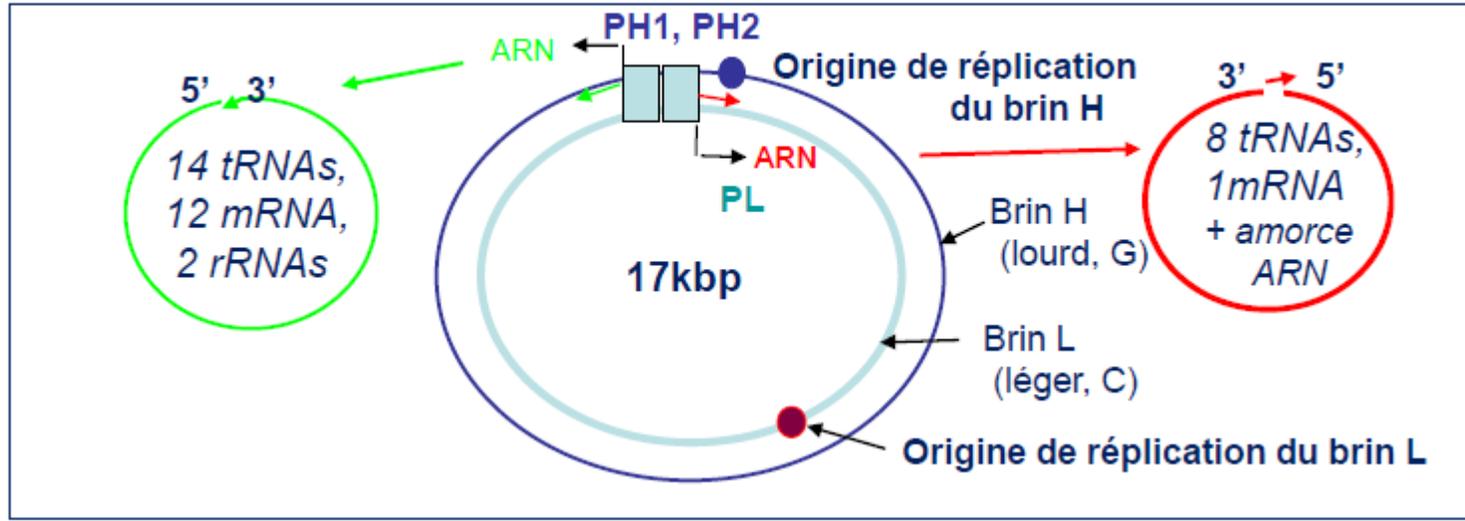
- l'ARN transcrit à partir du **PH1** et **PH2** (coplementaire du brin H) donnera 14 ARNt, 2 ARNr et 12 ARNm codant pour différentes enzymes de la chaine respiratoire.

- l'ARN transcrit à partir du **PL** (complementaire du brin L) donnera 8 ARNt, 1 ARNm codant, et une amorce ARN pour la réplication du brin H

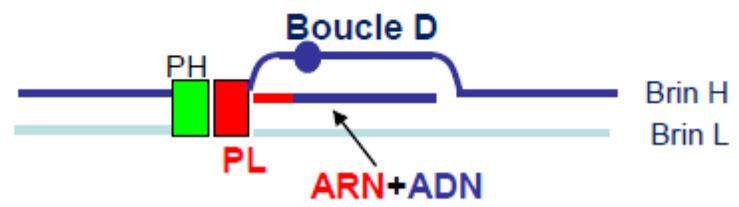
- Au total 37 gènes: 13 protéines (enzymes de la chaîne respiratoire)
24: ARN non codants (ribosomiaux, de transfert)



GENOME MITOCHONDRIAL

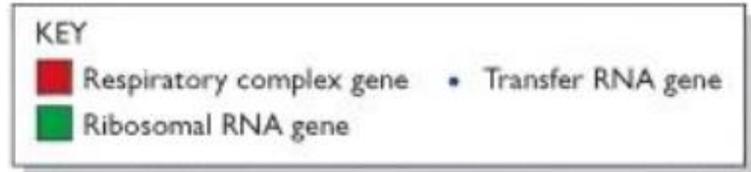
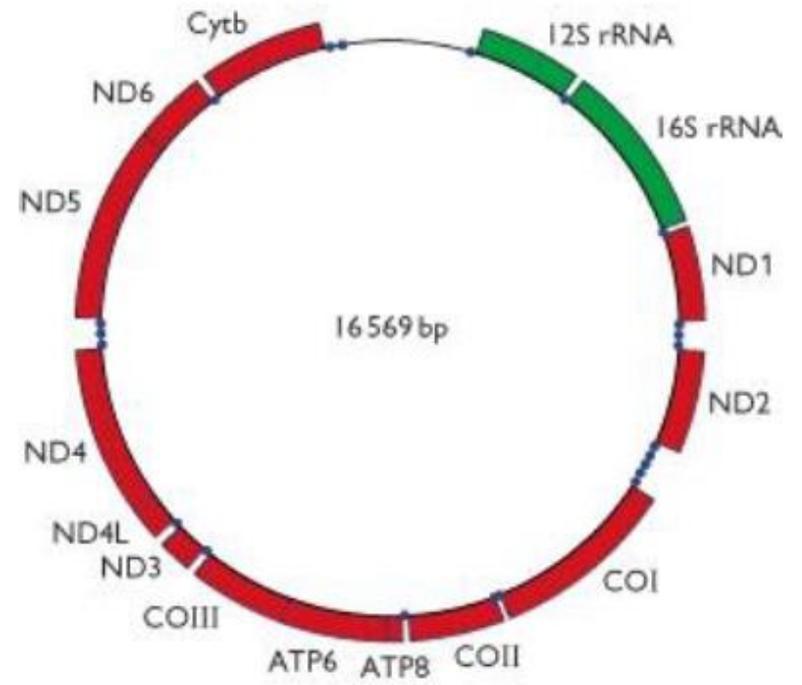
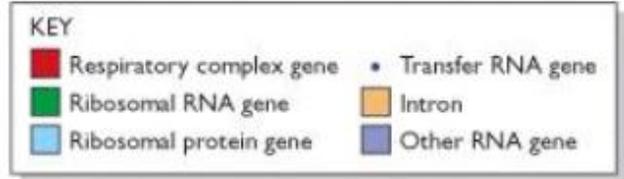
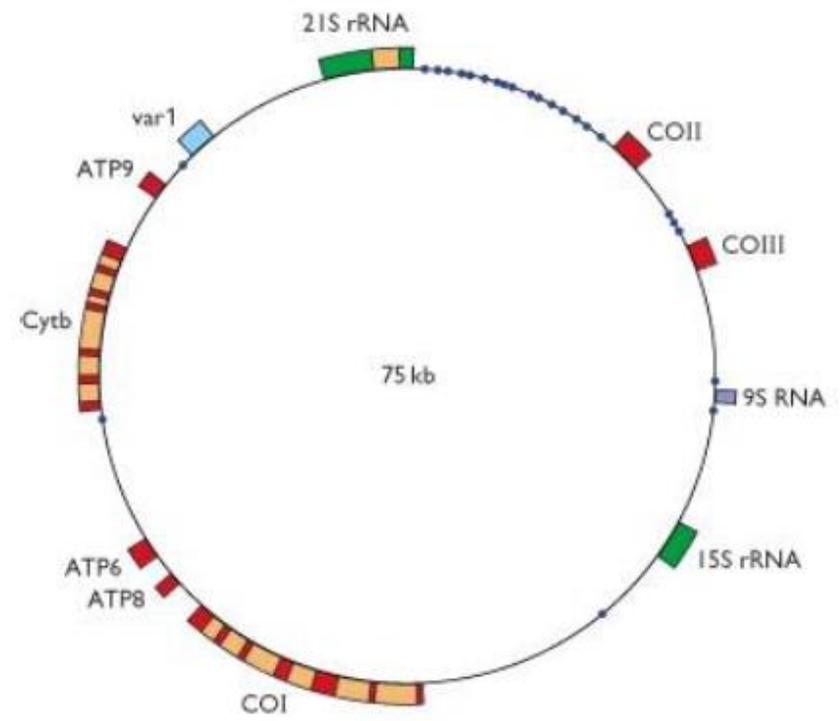


La transcription à partir de PL produit une amorce ARN qui sera allongée pour produire un court fragment d'ADN qui restera associé au brin H formant la boucle D (structure à 3 brins).



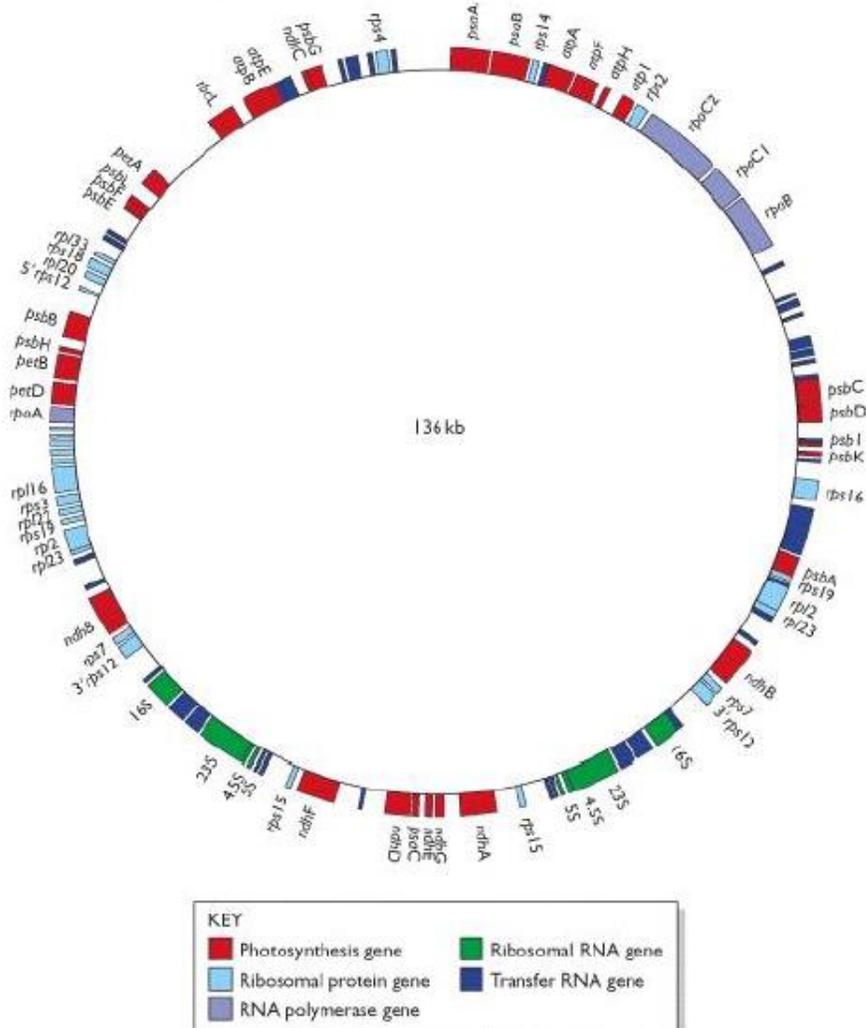
GENOME MITOCHONDRIAL

Exemple: le génome mitochondrial de *Saccharomyces cerevisiae* et de l'homme.



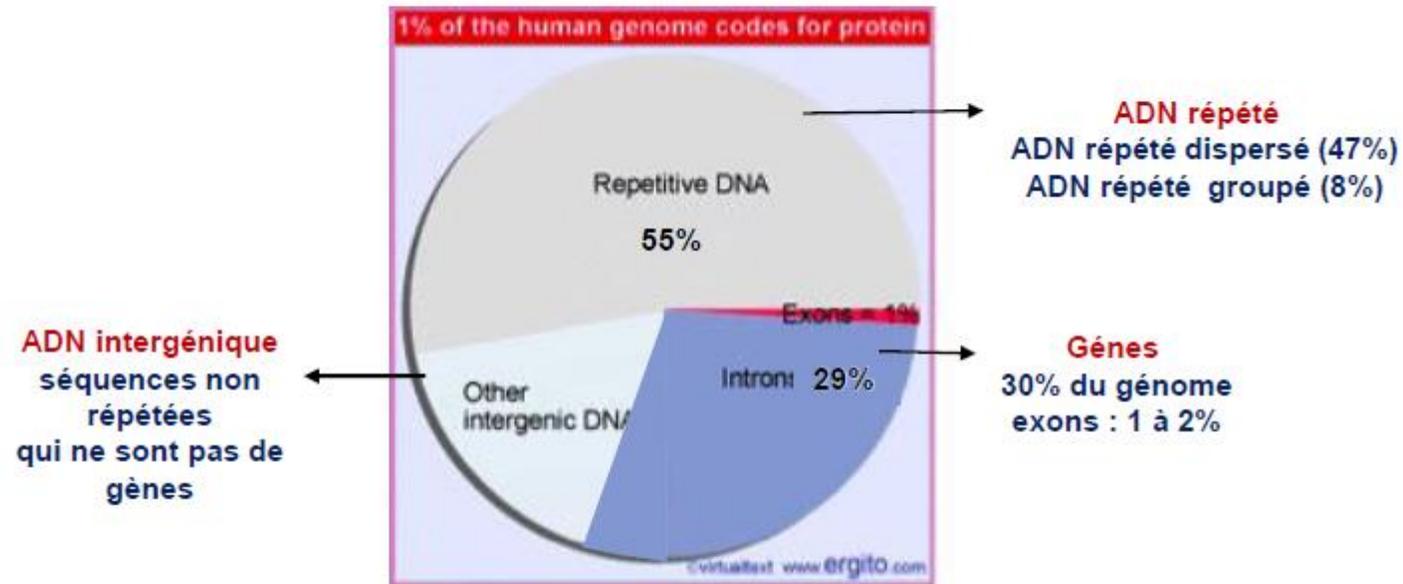
GENOME CHLOROPLASTIQUE

Exemple: le génome chloroplastique du riz.



TYPES DE SEQUENCES (GENOME NUCLEAIRE)

Génome : ensemble de l'ADN (nucléaire, mitochondrial, des chloroplastes) d'une cellule, d'un individu. 3 types de séquences dans le génome nucléaire eucaryote:



Les gènes (25 000) représentent environ 30% du génome humain : Dans ces gènes la partie retrouvée dans les ARN matures (exons) représente 1 à 2% du génome. Seuls les gènes sont transcrits en ARN
En moyenne: un Gène : 40 kbp

ADN REPETE DISPERSÉ

Ces séquences résultent de **transpositions** : déplacement dans le génome de fragments d'ADN dits « éléments transposables » ou **transposons** selon deux mécanismes

Deux types de transposition:

- Transposition **directe**
- **Rétrotransposition**

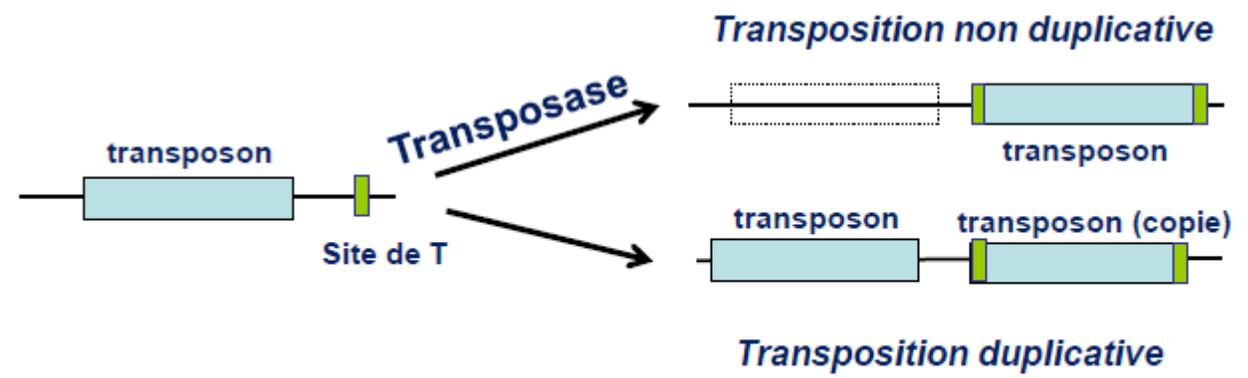


Deux types de transposons

- Type II: **DNA transposons**
- Type I: **rétroéléments**

ADN REPETE DISPERSÉ

A) Transposition directe

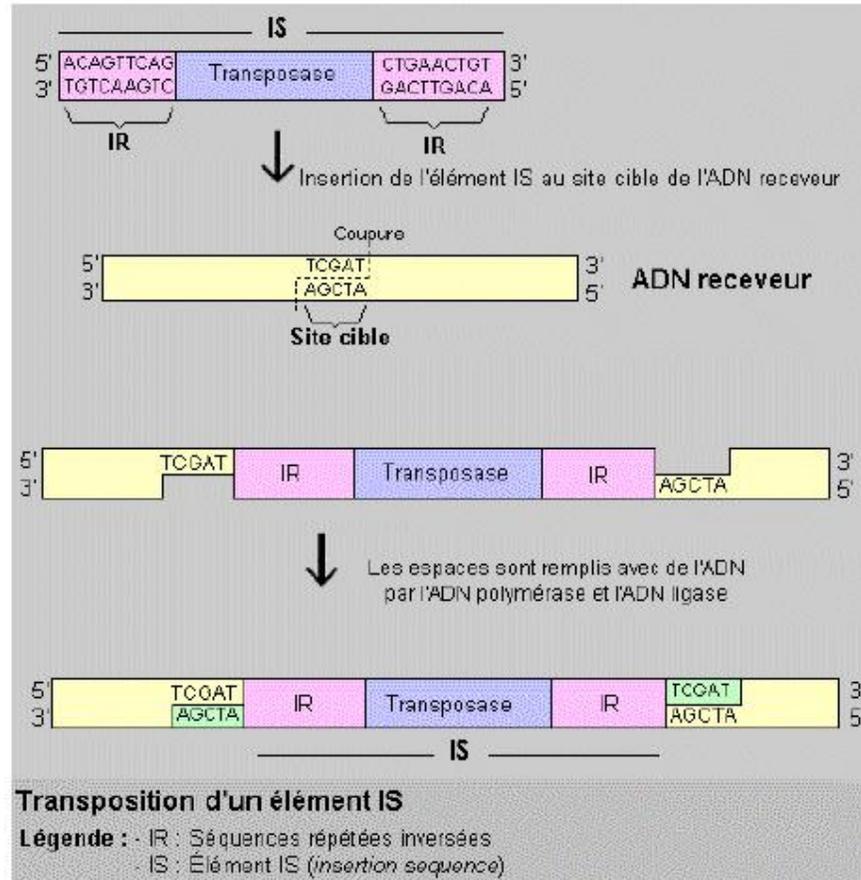


**Type II
DNA transposons**

ADN REPETE DISPERSÉ

A : DNA transposon : Structure d'un élément mobile (DNA transposon)

<http://www.irbv.umontreal.ca/francais/cours/transposon2/mol.htm>



Transposons autonomes
(transposase active)

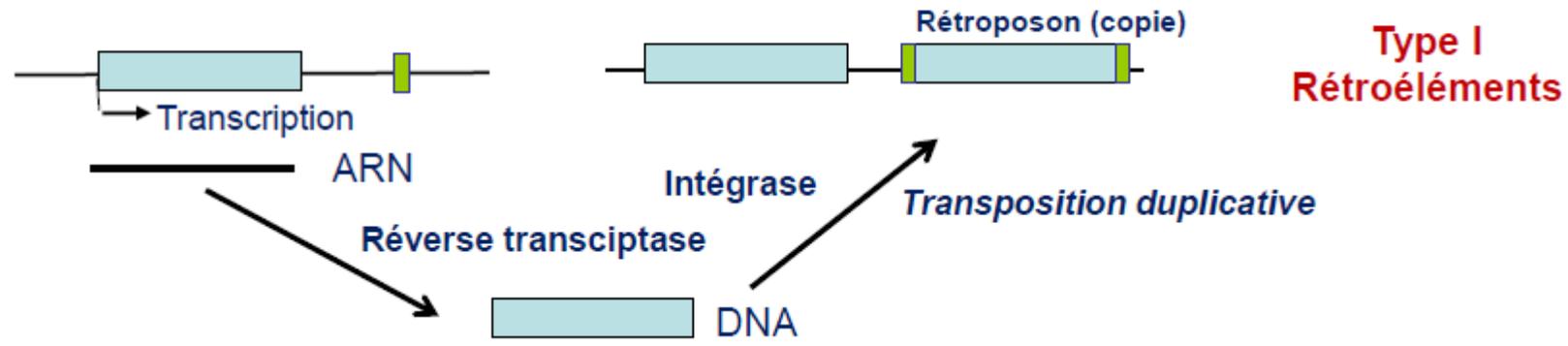
Transposons non autonomes
(pas de transposase active)



Barbara McClintock (1944)

ADN REPETE DISPERSÉ

B) Rétrotransposition (avec un intermédiaire ARN)



ADN REPETE DISPERSÉ

B-Rétroéléments:

Deux types de rétroéléments caractéristiques des génomes eucaryotes:

1- Rétrotransposons

Ont des **LTR** (*long terminal repeat*): séquences répétées directes à chaque extrémité impliquées dans la transcription

Contiennent des séquences codantes pour ARN polymérase réverse transcriptase et d'autres protéines virales (gag)



ADN REPETE DISPERSÉ

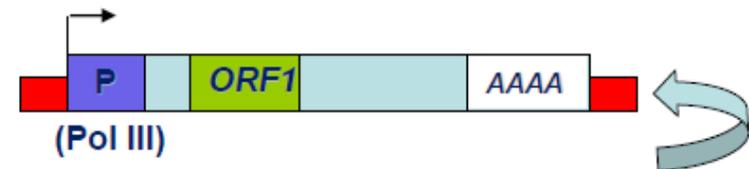
B-Rétroéléments:

Deux types de rétroéléments caractéristiques des génomes eucaryotes:

2- Rétroposons

Pas de LTR. Contiennent un promoteur situé en 5' et une séquence codante pour une transcriptase inverse et/ou une protéine impliquée dans la transposition. Contiennent des séquences codantes pour ARN polymérase et reverse transcriptase (pol). Chez les mammifères deux types de rétroposons :

1-Séquences « LINE » (*long interspersed nuclear elements*). Autonomes, elles codent leur RT (ORF2): 6 à 8 kb; 20% du génome humain

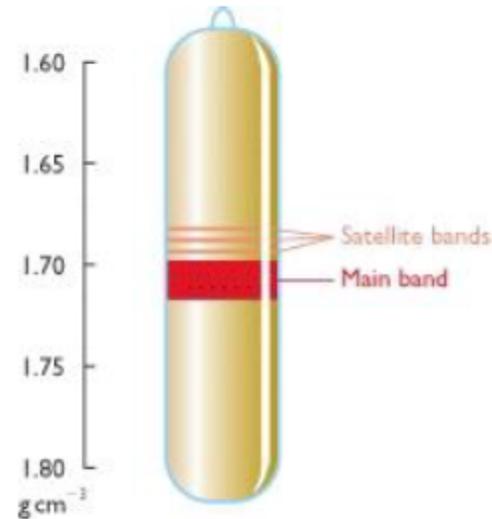


2-Séquences «SINE »(*short interspersed nuclear elements*). Non autonomes. (pas d'ORF2 donc pas de RT) 100 à 300bp; 13% du génome humain (exemple : séquences Alu)

ADN REPETE DISPERSÉ

Séquence satellite : séquence avec des répétitions directes en tandem groupées de motifs composés d'un nombre variable de nucléotides.

Ces répétitions se retrouvent à différents loci du génome. Pour un même motif, le nb de répétitions varie entre les loci

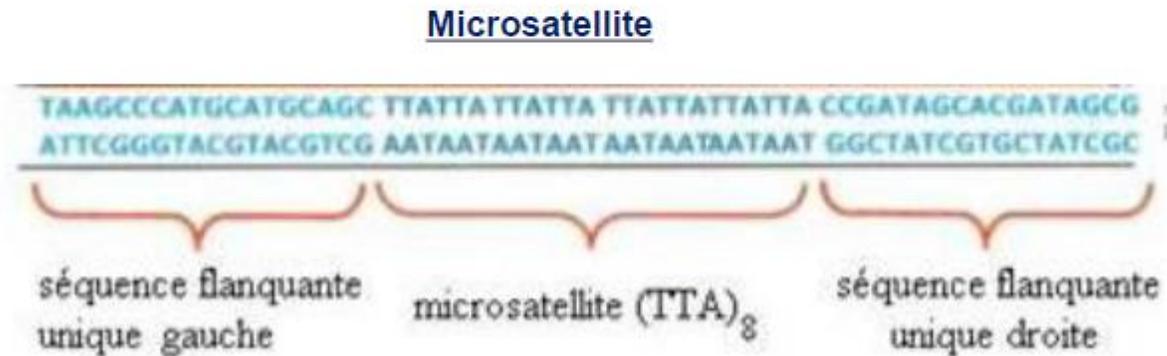


*Genomes. 2nd edition.
Brown TA.
Oxford: Wiley-Liss; 2002.*

	motif	bloc de répétition	exemple
satellite:	2-2000 nt	20 kbp - 10 Mb	centromères
minisatellite :	2-64 nt	0,1-20 kbp	téломères
	<i>(VNTR : variable number of tandem repeats)</i>		
microsatellite:	1-6 nt	10-200 bp	CAG repeats
	<i>(STR : short tandem repeats)</i>		

ADN REPETE DISPERSÉ

Séquence satellite : séquence avec des répétitions directes en tandem groupées de motifs composés d'un nombre variable de nucléotides.
Ces répétitions se retrouvent à différents loci du génome. Pour un même motif, le nb de répétitions varie entre les loci

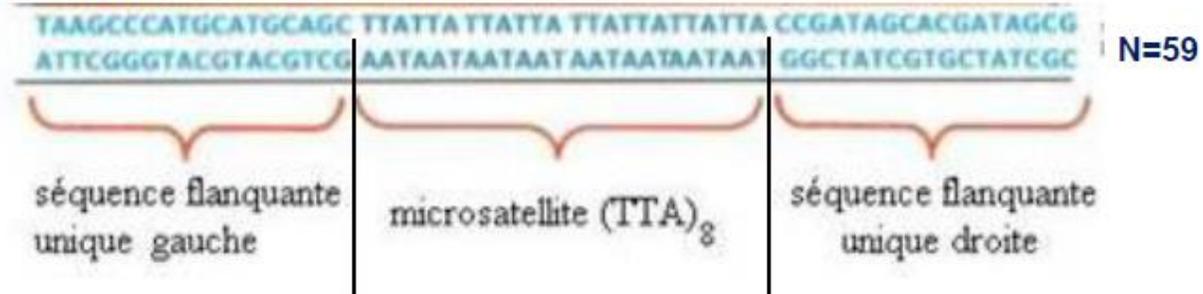


ADN REPETE DISPERSÉ

Polymorphisme des séquences satellites

Les séquences satellites situées sur un même locus sur des chromosomes homologues diffèrent très souvent par le nombre de répétitions du motif sur chaque allèle (**polymorphisme de répétition**)

Microsatellite au locus A : allèle 1



Microsatellite au locus A : allèle 2



Ces deux séquences sont **différentes par le nb de répétitions**. Elles sont polymorphes (polymorphisme de répétition)

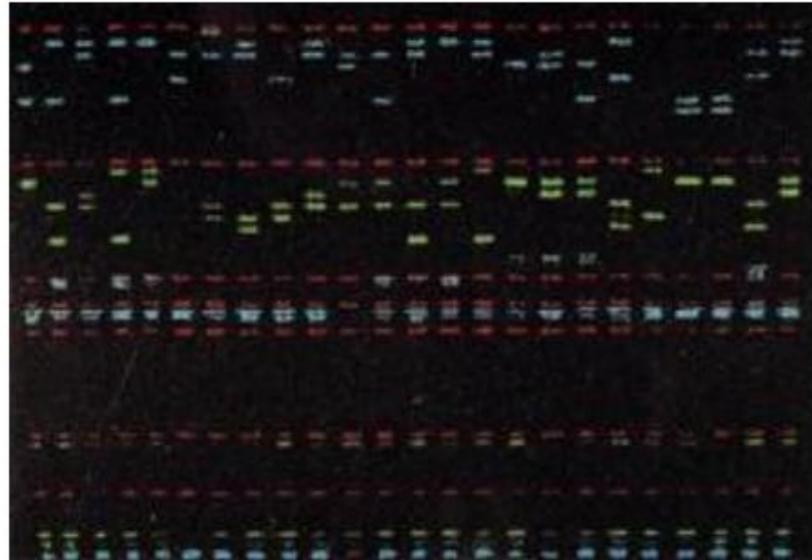
ADN REPETE DISPERSÉ

Polymorphisme des séquences satellites

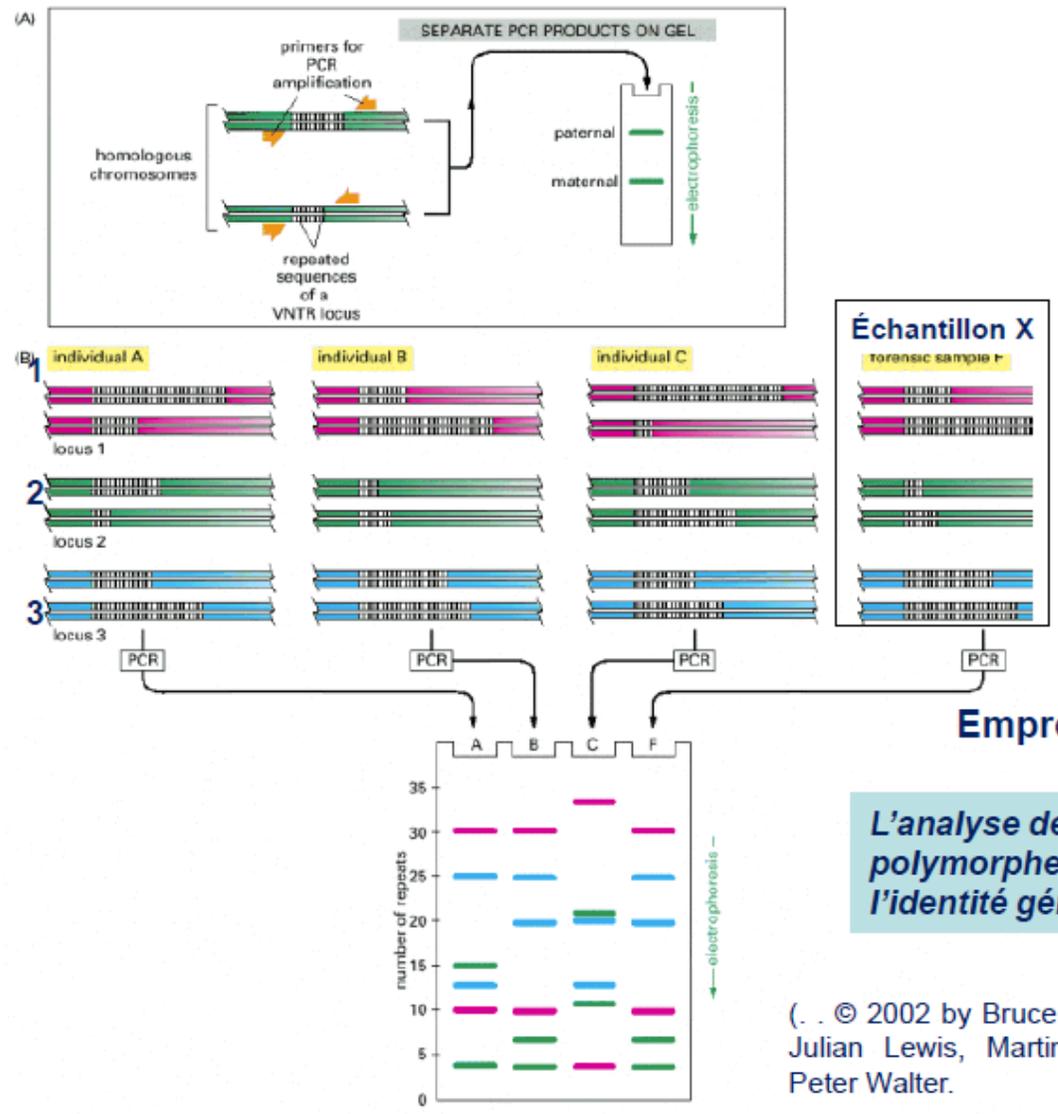
Origine du polymorphisme de répétition : Glissement de la DNA polymérase lors de la réplication, recombinaison asymétrique

Les séquences satellites sont très polymorphes et sont utilisées comme marqueurs pour identifier des individus (profil génétique, test de paternité) et pour cartographier les génomes

allèle: une des deux copies d'une séquence située au même locus sur 2 chromosomes homologues



ADN REPETE DISPERSÉ

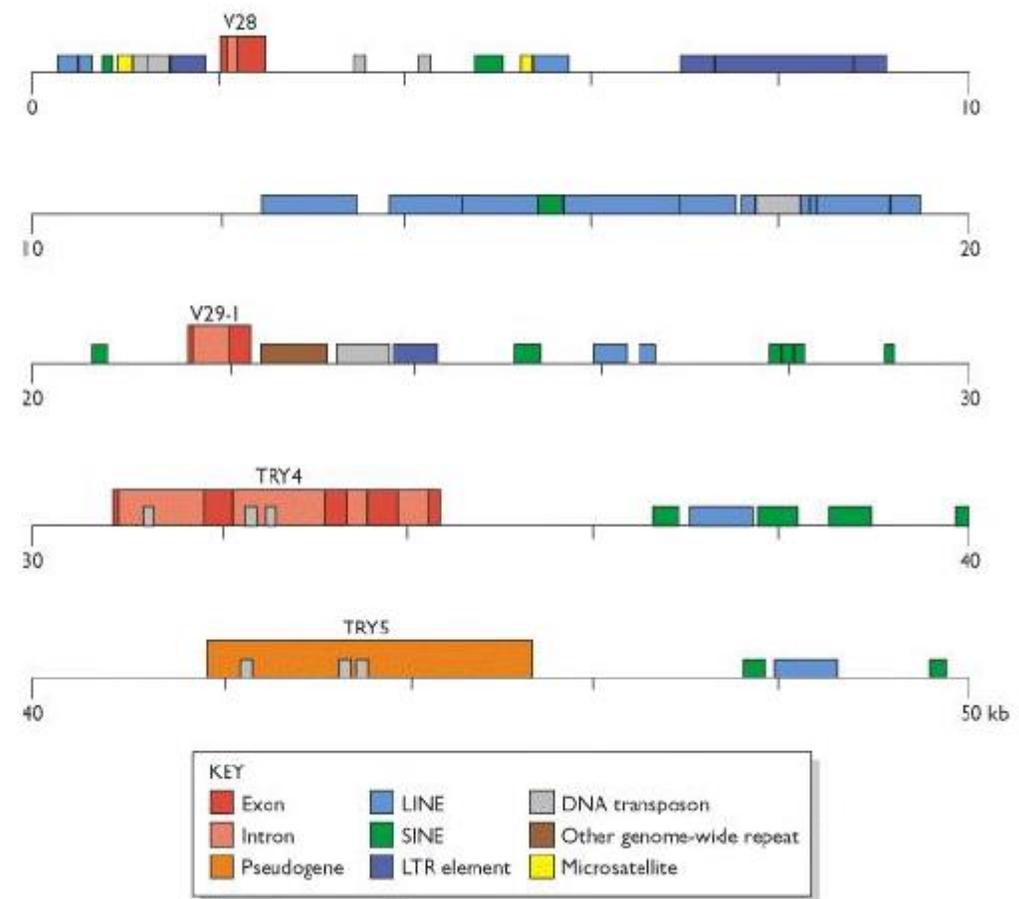


ADN REPETE DISPERSÉ

➤ <u>Gènes</u> (exons + introns)___	30%
➤ <u>ADN répété</u> (séquences répétées)	55%
• Répété dispersé	47%
– DNA transposons (2 à 3 kbp) 300 000	3%
– Rétro-éléments	44%
rétrotransposons	8 %
rétroposons	
LINE (6 à 8 kbp)	20 %
SINE (100 à 300 bp)	13 %
Autres:	3 %
• Répété groupé (ADN satellite)	8%
➤ <u>Séquences intergéniques non répétées</u>	15%

GENES EUCARYOTES

Exemple: fragment de 50 kb du locus du recepteur T-cell, chromosome 7

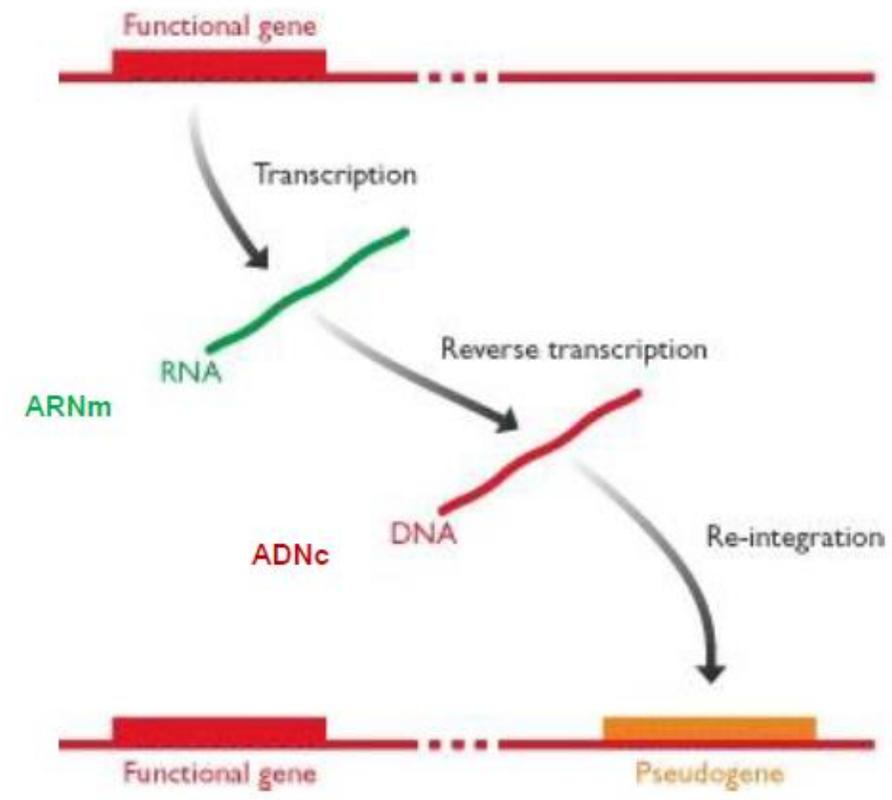


Genomes. 2nd edition. Brown TA. Oxford: Wiley-Liss; 2002

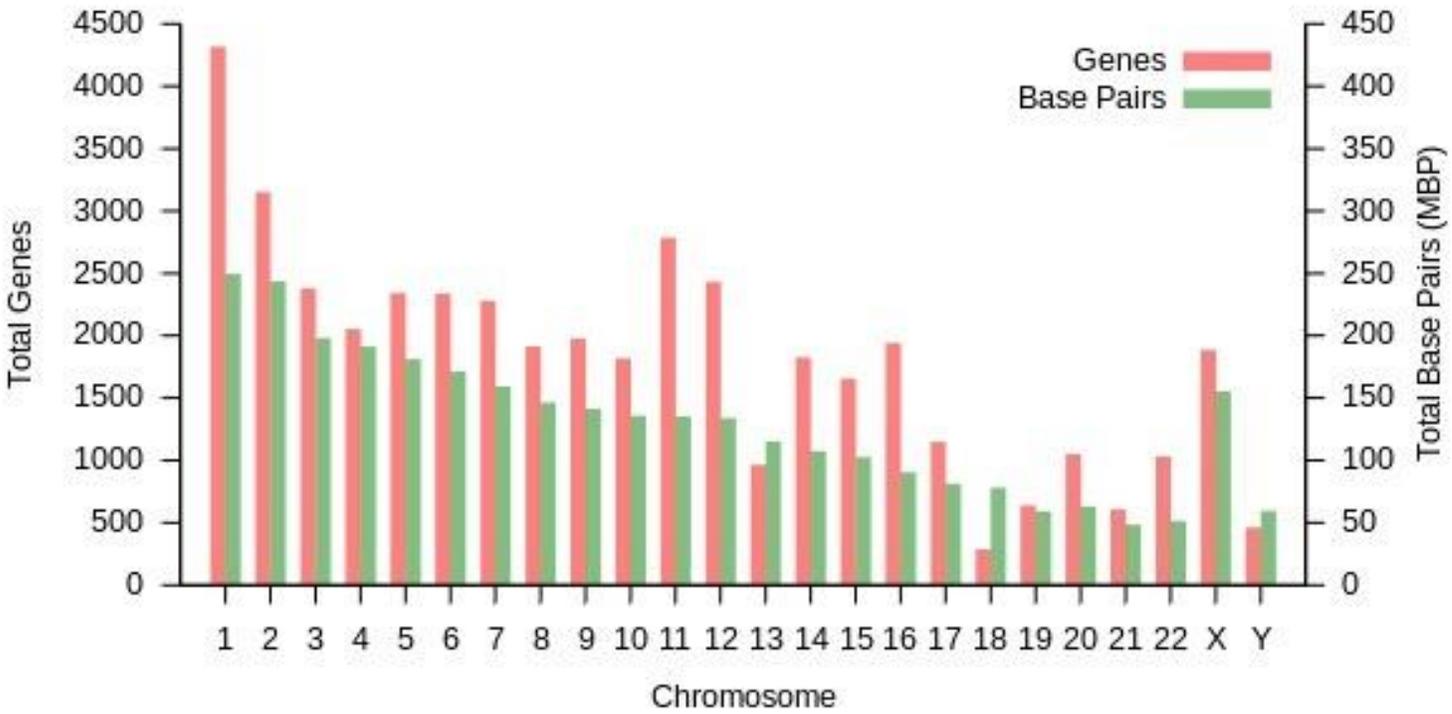
PSEUDOGENES

Pseudogène: copie **non-fonctionnelle** d'un gène

- conventionnel: inactivé à cause d'une mutation
- rétropseudogène (procédé): copie à partir d'un ARNm par transcription reverse et réinsertion



QUELQUES DEFINITIONS POUR DEMARRER...



QUELQUES DEFINITIONS POUR DEMARRER...

	Objets	Données	Outils	Grands projets	Bases de données
Phénotype	Différences au sein d'une espèce	Couleur, structure, taille, ...	Observation clinique, visuelle, épidémiologique	...	PhenomicDB PharmGkb
Génome	L'ADN complet d'une espèce	Séquences ADN et gènes	Séquenceurs	Human Genome Project	GenBank, DDBJ, ENA
Epigénome	Variations d'expression des gènes non liées aux séquences codantes de l'ADN	Biochimie, protéome	Methyl-Seq	Roadmap Epigenomics Project, Epiwin	NCBI Epigenomics Database NGSmethDB (méthylations)
Transcriptome	L'ADN codante des gènes	ESTs, cDNAs (coding DNA)	DNA Microarrays (biopuces)	Human Genome Project	GenBank
Variome	Variations de l'ADN au sein d'une espèce, polymorphismes	SNPs Haplotypes	Séquenceurs	1000 Genome Project	dbSNP (polymorphismes) Patrocles, Jaspar
Protéome	Ensemble des protéines d'une espèce ou d'un type de cellule	Séquence d'acides aminés des protéines + forme 3D	Spectrographie de masse, Y2H (Yeast to Hybrid)	Human Proteome Project	Genbank, CellBase, Human Proteins Reference Database Protein Databank, EMBL-EBI
Bactérome	Gènes des bactéries (ex: flore intestinale)	ADN des bactéries	Séquenceurs + ...	Human Microbiome Project	Microbial Genome Database for Comparative Analysis
Cancérologie	Gènes mutés générateurs de cancers	Gènes mutés générateurs de cancers	Séquenceurs + phénotypie	The Cancer Genome Atlas	The Cancer Genome Atlas, Atlas of Genetics and Cytogenetics in Oncology and Haematology

Source: compilation Olivier Ezratty, 2012