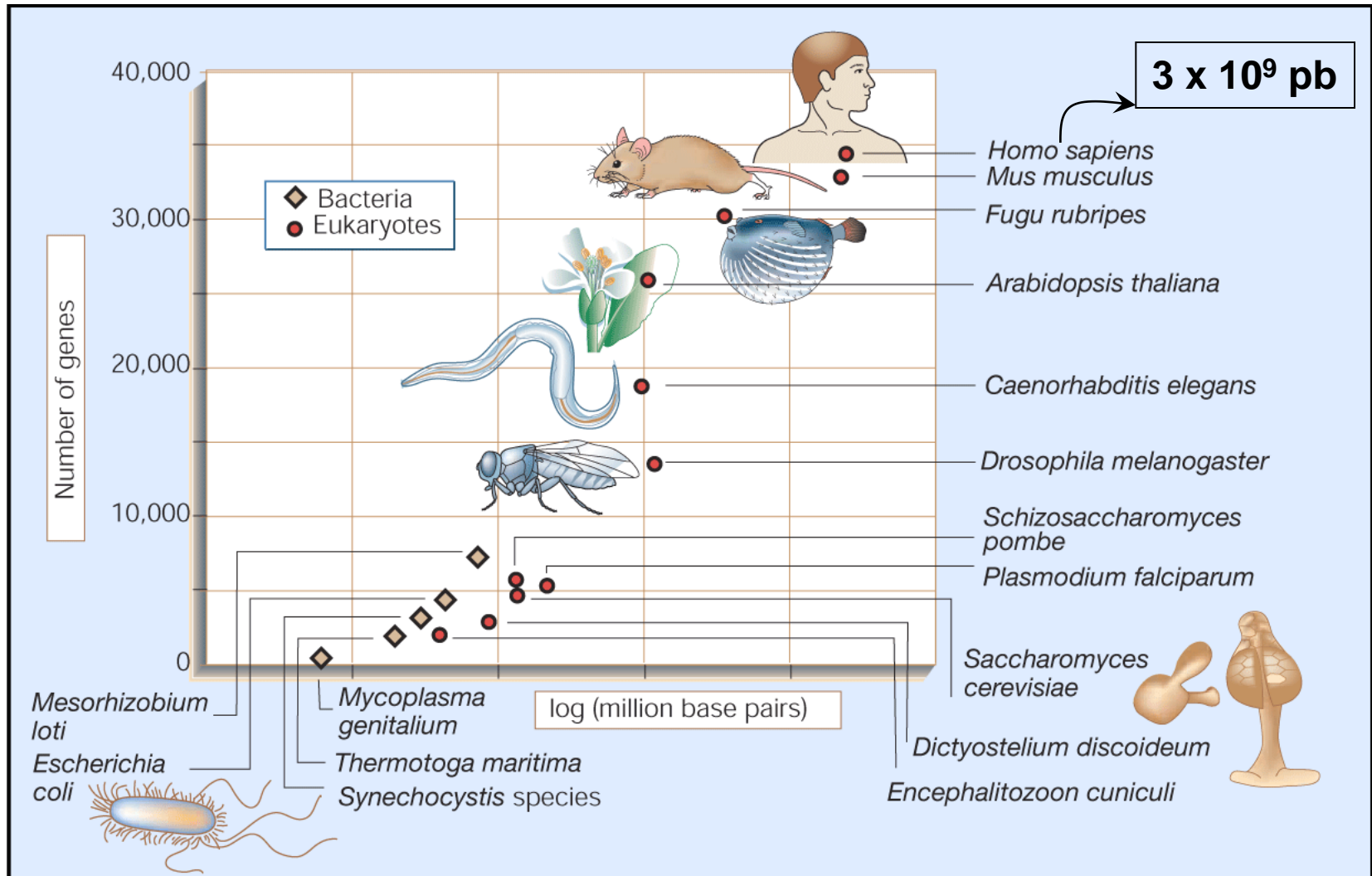


# Thème 1: Génomique

- Génomique:
  - l'ensemble de l'information génétique d'un organisme
  - Taille: nombre de paires de bases de l'ADN constituant tous les chromosomes d'un organisme donné
- Génomique:
  - Science ayant pour objet l'étude des génomes:
    - Séquence (identification des gènes/allèles)
    - Structure
    - Évolution
    - Différences inter-individuelles (haplotypes)

# Génomes



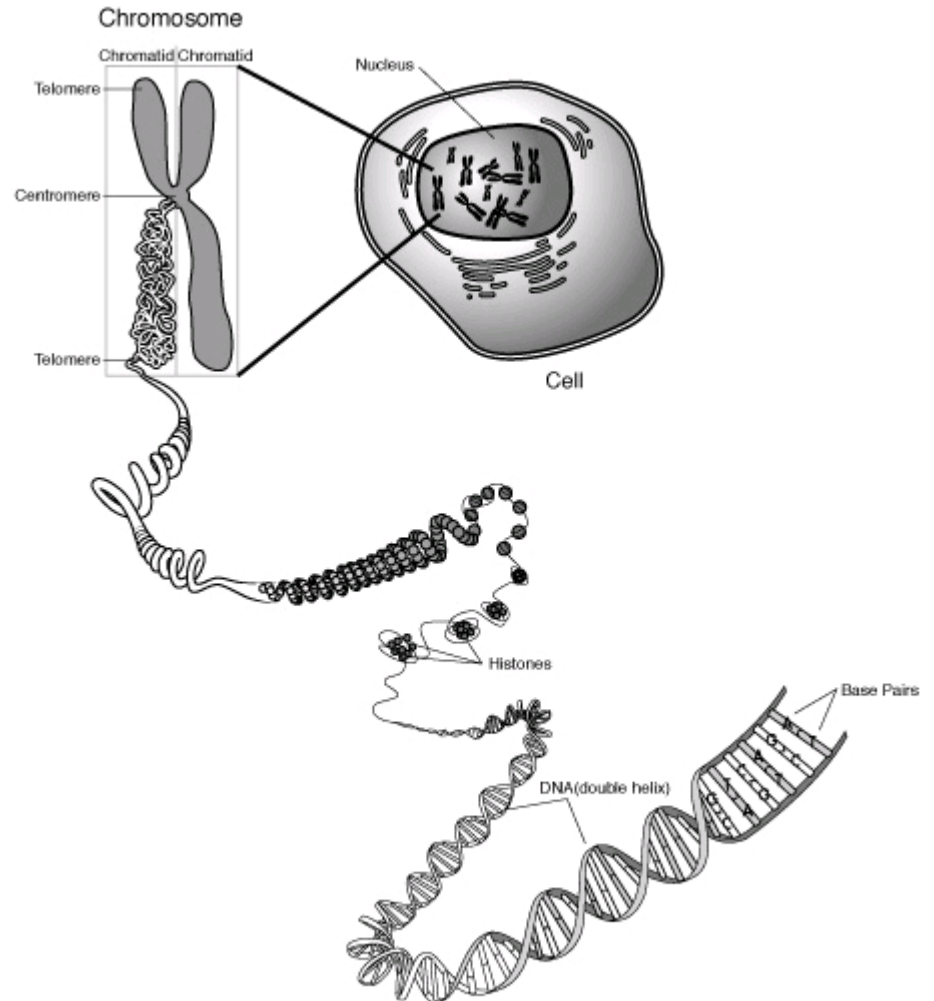
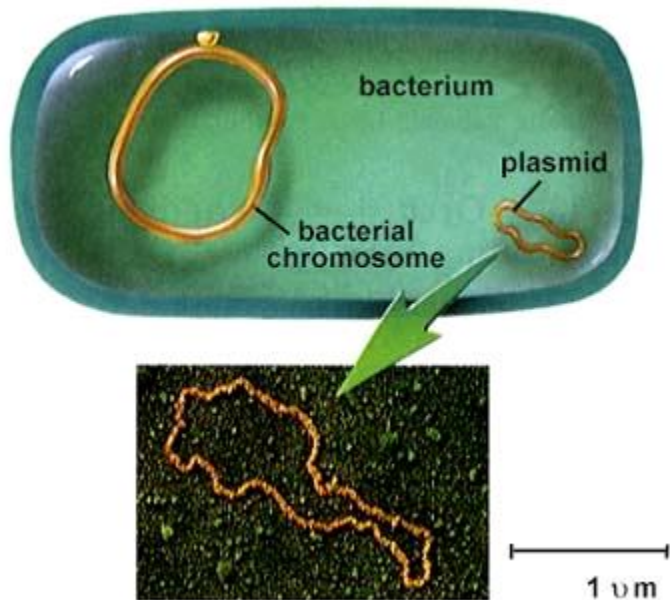
# Génomes bactériens

- **1. Longueur:** de  $10^5$  à  $10^7$  paires de bases. La longueur du génome de *E. coli* = 1.6 mm.
- **2. Génome haploïde:** 1 seule copie de chaque chromosome par bactérie
- **3. Organisation structurale:** le chromosome bactérien est circulaire.
- **4. Informations dans le génome:**
  - Contiennent beaucoup moins de gènes que les génomes eucaryotes
  - Quelques séquences modérément répétées sont présentes – gènes encodant les ARN de transfert et les ARN ribosomiaux.

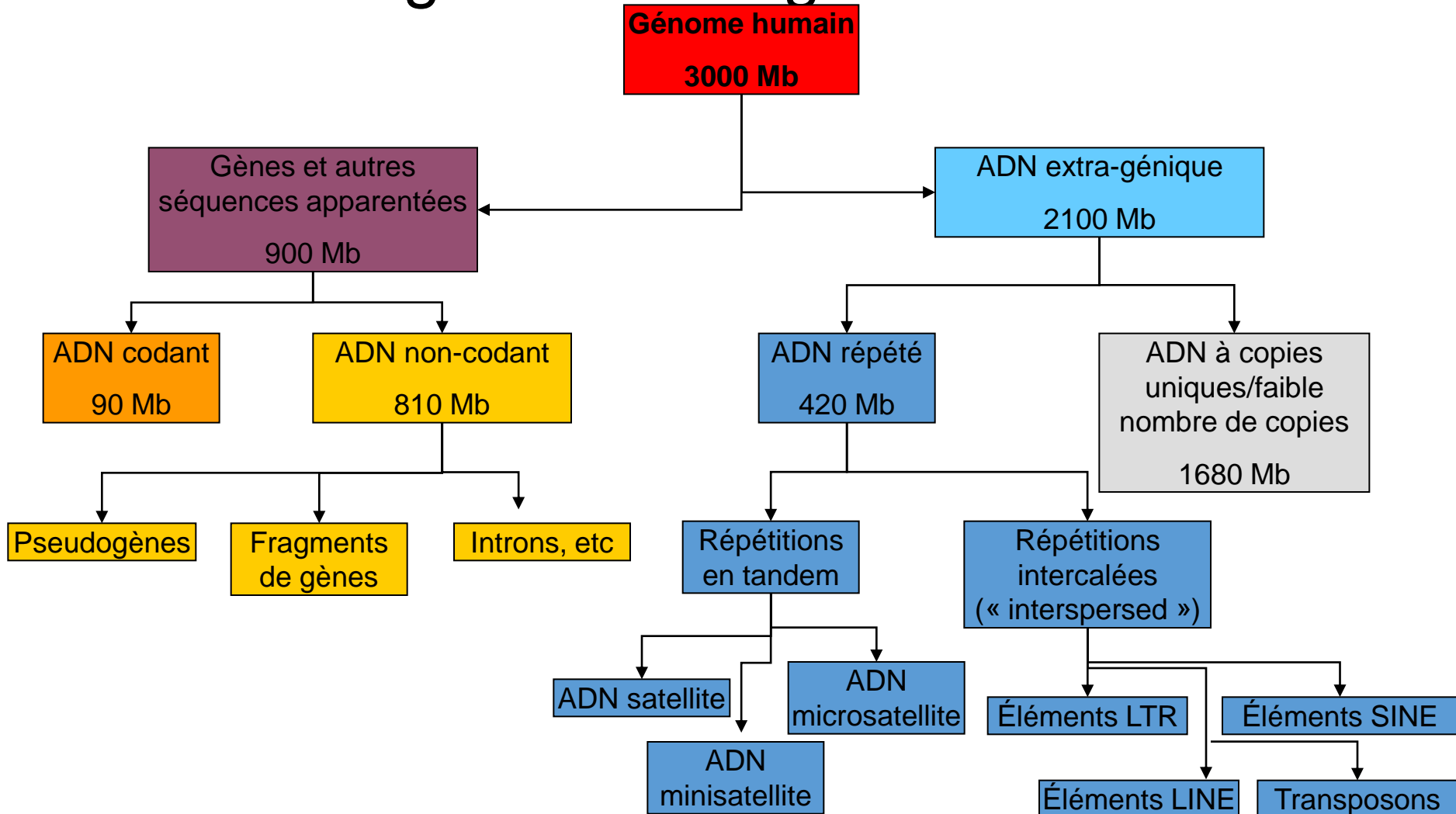
# Génomes eucaryotes

- **1. Longueur:**  $10^7$  à  $10^{11}$  paires de bases par génome haploïde . La longueur du génome humain (par cellule): approx. 1.5 m.
- **2. Génome diploïde**, sauf pour les cellules germinales (spermatozoïdes et ovules).
- **3. Organisation générale:**
  - Plusieurs chromosomes linéaires
  - Le nombre de chromosome varie selon les espèces.
  - Télomères: séquences répétées localisées au deux bouts de chaque chromosome. Essentiel à la réplication de l'ADN.
  - Centromère: situé au milieu de chaque chromosome . Impliqué dans la ségrégation des chromosomes lors de la division cellulaire.
- **4. Information contenue dans le génome:**
  - **a.** beaucoup plus de gènes à copie unique que les bactéries.
  - **b.** séquences modérément répétées (gènes encodant les ARNt et ARNr; gènes encodant les anticorps)
  - **c.** ADN hautement répétés: par exemple l'ADN satellite DNA (très abondant: 20-30% du génome).

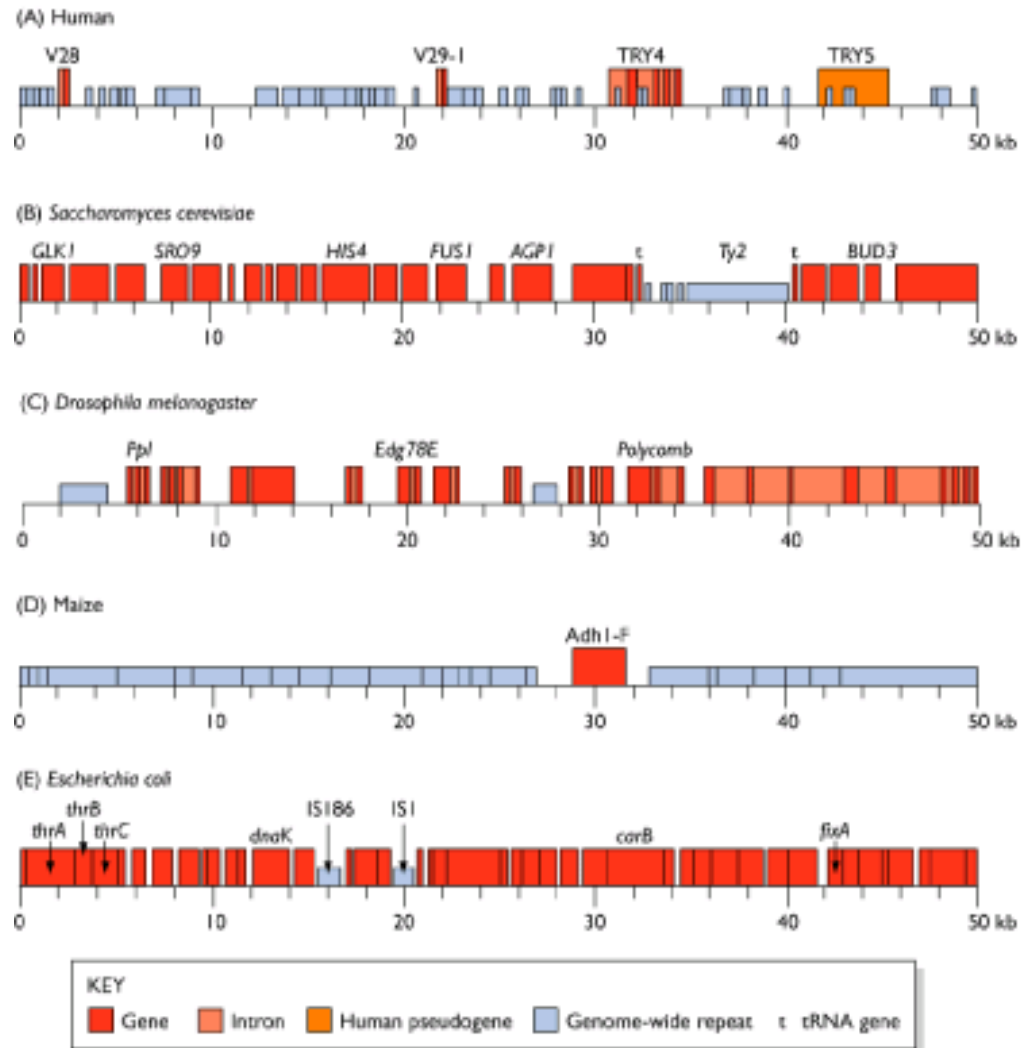
# Chromosomes



# Structure générale du génome humain



# Structure générale des génomes



# Pourquoi étudier les génomes?

- **Donne la listes des «pièces» nécessaires à la formation d'êtres vivants**
- **Identification de gènes responsables de maladies:**
  - Intervention: p.ex. thérapie génique
  - Facilitation du diagnostique/ prédisposition aux maladies
    - P.ex: identification de mutations du gène BRCA-1 qui prédisposent au cancer du sein → établissement d'un test diagnostique → prévention/changement de certaines habitudes de vie.
- **Pharmacogénomique:** relier la réaction individuelle aux médicaments avec la présence d'allèles particulières:
  - Améliorer l'efficacité des médicaments
  - Réduire les effets secondaires indésirables
- **Découverte de traits génétiques utiles en agriculture:**
  - Augmentation de la productivité
  - Résistance des plantes à la sécheresse/parasites



# Outils de la génomique

- **Biologie moléculaire (clonage/PCR):**

- Isolation et amplification d'ADN en vue du séquençage
- Localisation des gènes sur les chromosomes

- **Séquenceurs d'ADN:**

- Séquençage rapide et à grande échelle d'ADN
- Technologie actuelle: séquençage de plusieurs millions de paires de bases par jour

- **Bioinformatique:**

- Programmes permettant l'analyse de données du séquençage:
  - Identification des gènes
  - Comparaison entre séquences

# Difficultés à séquencer et caractériser un génome humain

- **Taille:**

- Humain:  $3 \times 10^9$  pb
- Paires de bases pouvant être séquencées en une expérience: max 400pb/expérience (circa 1990);

- **Complexité:**

- Génome eucaryote est constitué
  - gènes (infime partie du génome)
  - séquence de régulation (e.g. promoteurs)
  - Introns
  - Séquences répétées (majeure partie du génome)
  - ADN « junk »
- Comment trouver les gènes parmi toutes les autres séquences? Les séquences de régulation?

# Difficultés à séquencer et caractériser un génome humain

- **Solutions:**

- Séquencer le génome par petits bouts;
- Développer les outils nécessaires:
  - Trouver des marqueurs pouvant nous aider à nous repérer le long du génome
  - Établir une carte du génome qui nous guidera lors de l'assemblage de la séquence
  - Développer une technologie nous permettant de séquencer l'ADN à grande échelle (automatisation)
  - Mettre en place un système de base de données pour l'entreposage permanent des données de séquences
  - Développer les programmes informatiques permettant d'analyser ces séquences.

# Difficultés à séquencer et caractériser un génome humain

- Pour développer ces outils: commencer par séquencer des génomes plus petits avant de s'attaquer au génome humain:
  - Bactéries (*E. coli*, *H. influenza*, autres)
  - Eucaryotes simples: levures
  - Eucaryotes multicellulaires: *C. elegans*, *D. melanogaster*.

**TABLE 1.3** *Comparison of Gene Content in Some Representative Genomes*

	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>Drosophila</i> <sup>a</sup>	<i>A. thaliana</i> <sup>a</sup>	<i>M. musculus</i>	<i>H. sapiens</i>
Genome size <sup>a</sup> (Mb)	4.6	12.0	120+	115+	2,500+	3,000+
Number of genes <sup>b</sup>	4,300	6,250	13,600	25,500	30,000	50,000
Average gene density (kb)	1.1	1.9	8.8	4.5	80	60
Number of gene families <sup>c</sup>	2,500	4,500	8,000	11,000	10,000	10,000

<sup>a</sup>*Drosophila*, human, and *Arabidopsis* genome sizes indicate sequenced euchromatin, excluding heterochromatin.

<sup>b</sup>Approximate number of genes estimated from original annotation. The human estimate has recently halved to 25,000.

<sup>c</sup>Number of gene families rounded to 500.

# Projet « Génome Humain »

**TABLE 1.1** *Initial Goals of the Human Genome Project (Part 1)*

## From the First 5-Year Plan: 1993–1998

### 1. THE GENETIC MAP

Complete 2 to 5 cM map by 1995

Develop new technology for rapid and efficient genotyping

### 2. THE PHYSICAL MAP

Complete STS map to 100 kb resolution

### 3. DNA SEQUENCING

Develop approaches to sequence highly interesting regions on Mb scale

Develop technology for automated high throughput sequencing

Attain sequencing capacity of 50 Mb per year; sequence 80 Mb by 1998

### 4. GENE IDENTIFICATION

Develop efficient methods for gene identification and placement on maps

### 5. TECHNOLOGY DEVELOPMENT

Substantially expand support for innovative genome technology research

# Projet « Génome Humain »

**TABLE 1.1** *Initial Goals of the Human Genome Project (Part 2)*

## From the First 5-Year Plan: 1993–1998

### 6. MODEL ORGANISMS

- Finish STS map of mouse genome to 300 kb resolution
- Obtain complete sequence of biologically interesting regions of mouse genome
- Finish sequences of *E. coli* and *S. cerevisiae* genomes
- Substantial progress on complete sequencing of *C. elegans* and *D. melanogaster*

### 7. INFORMATICS

- Continue to create, develop, and operate databases and database tools
- Consolidate, distribute, and develop software for genome projects
- Continue to develop tools for comparison and interpretation of genome information

# Projet « Génome Humain »

**TABLE 1.1** *Initial Goals of the Human Genome Project (Part 3)*

## From the First 5-Year Plan: 1993–1998

### 8. ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS (ELSI)

- Continue to identify and define issues and develop policy options
- Develop and disseminate policy regarding genetic testing
- Foster greater acceptance of human genetic variation
- Enhance public and professional education programs on sensitive issues

### 9. OTHER

- Training of interdisciplinary genome researchers
- Technology transfer into and out of genome centers
- Outreach

# Projet « Génome Humain »

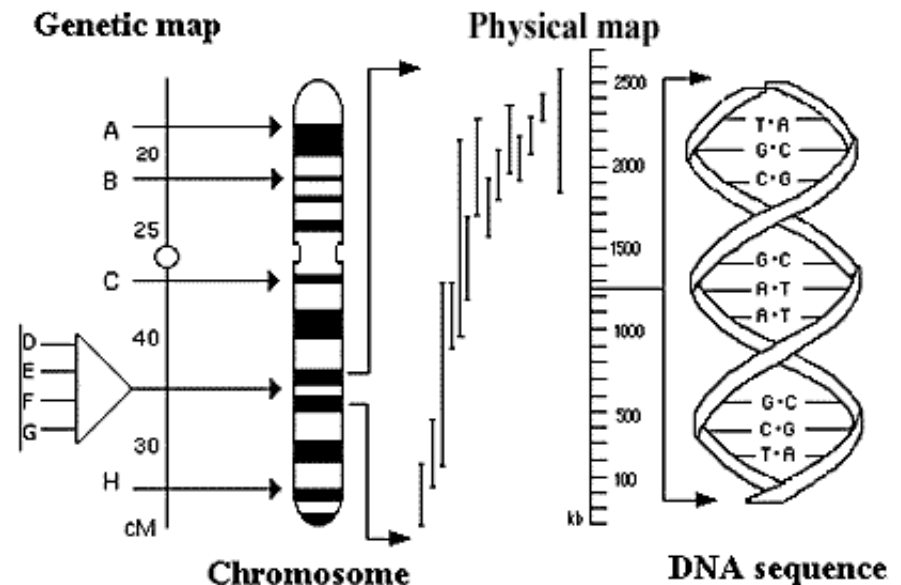
- **Implique les étapes suivantes:**

- 1. Clonage du génome
  - Utilisation de YACs, BACs, cosmides, plasmides
- 2. Cartographie
  - Carte physique, génétique
  - But: donner des points de repère le long de chaque chromosome.
- 3. Séquençage et assemblage des séquences
- 4. Annotation
  - Identification des gènes, séquences régulatrices (promoteurs), séquences répétées, etc.
- 5. Analyse



# Cartes génétiques vs physiques

- Carte génétique:
  - Description de la distance et position relative de gènes ou marqueurs génétiques les uns par rapport aux autres le long du chromosome
  - Exprimée en centimorgans (cM.)
- Carte physique:
  - Taille et position exacte, le long du chromosome, de fragments d'ADN isolés expérimentalement.
  - Exprimée en paires de bases.



**Sequences of base pairs mapping**

# Cartes génétiques vs physiques

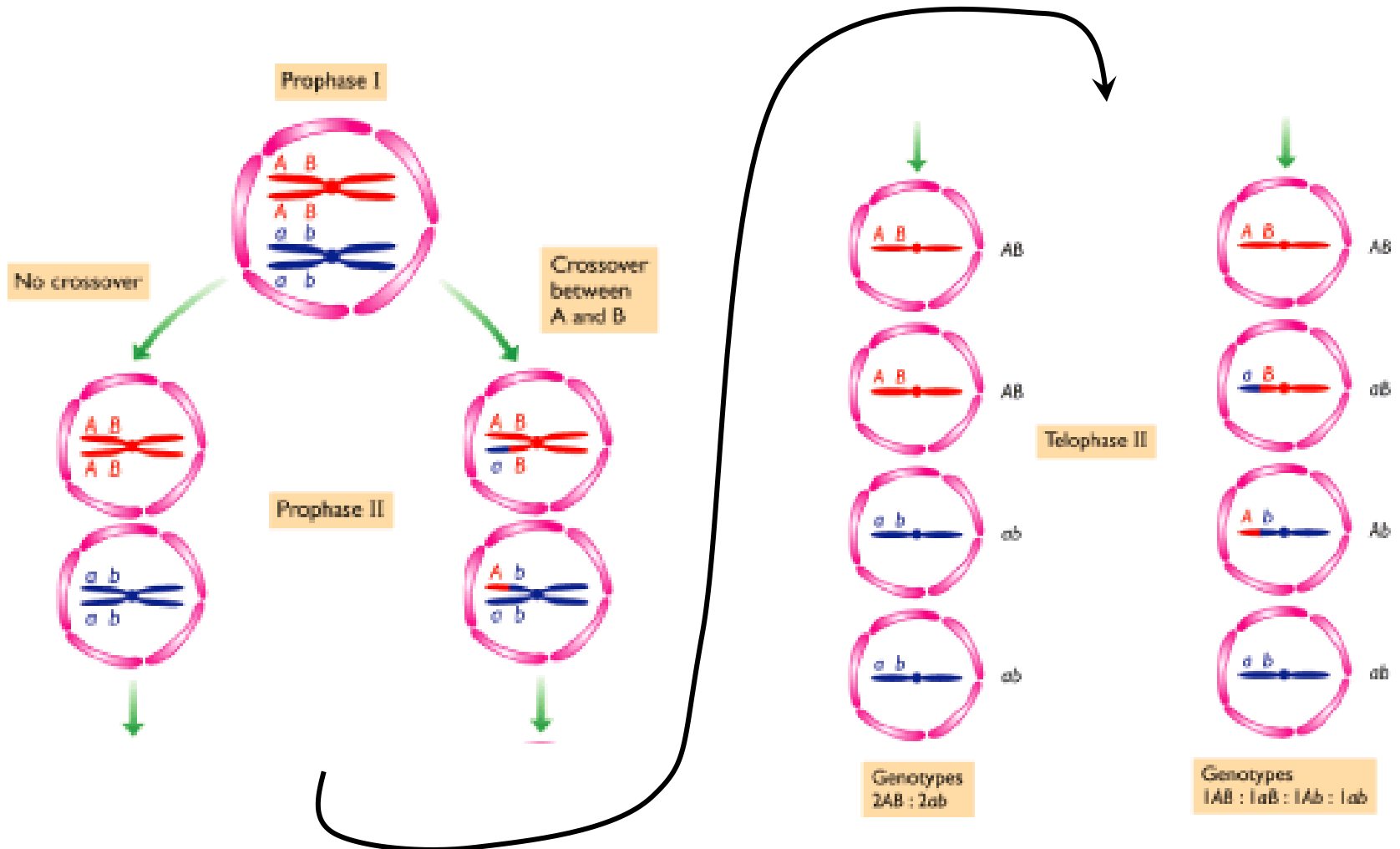
HUMAN GENOME PROJECT GOALS	Resolution
♦ Complete a detailed human genetic map	2 Mb
♦ Complete a physical map	0.1 Mb
♦ Acquire the genome as clones	5 kb
♦ Determine the complete sequence	1 bp
♦ Find all the genes	
With the data generated by the project, investigators will determine the functions of the genes and develop tools for biological and medical applications.	

# Objectif 1: Établir une carte génétique du génome humain

- **Carte génétique:**

- Description de l'ordre/distance relatif de marqueurs génétiques sur le chromosome;
- Résolution: environ 30 Mpb (3 million de paires de bases)
- Marqueurs:
  - Séquence d'ADN
  - Sites de coupure par des enzymes de restriction
  - Phénotype (p.ex.: couleur des yeux)
- Recherche la fréquence avec laquelle ces marqueurs vont **co-ségréger** suite à un croisement génétique;
- Implique l'analyse de la ségrégation des marqueurs au sein de pédigrées (i.e. un couple et ses descendants) ou de croisements contrôlés en laboratoire (e.g. souris, drosophile).

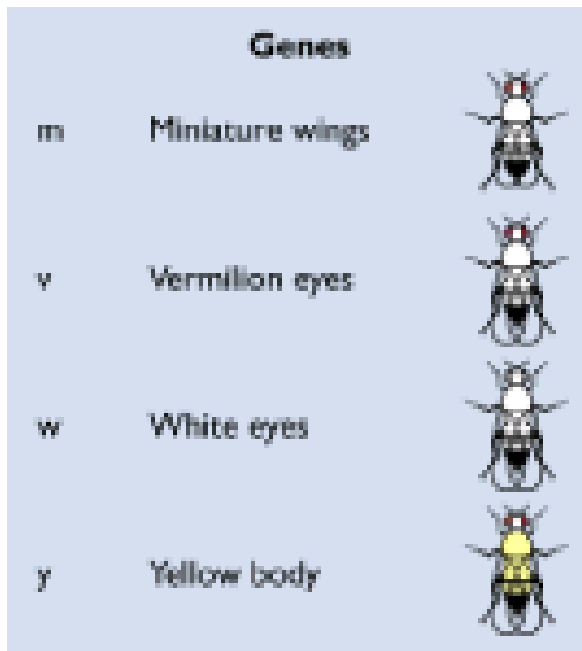
# Recombinaison entre chromosomes



# Objectif 1: Établir une carte génétique du génome humain

- **Unité de distance standard: le centimorgan (cM):**

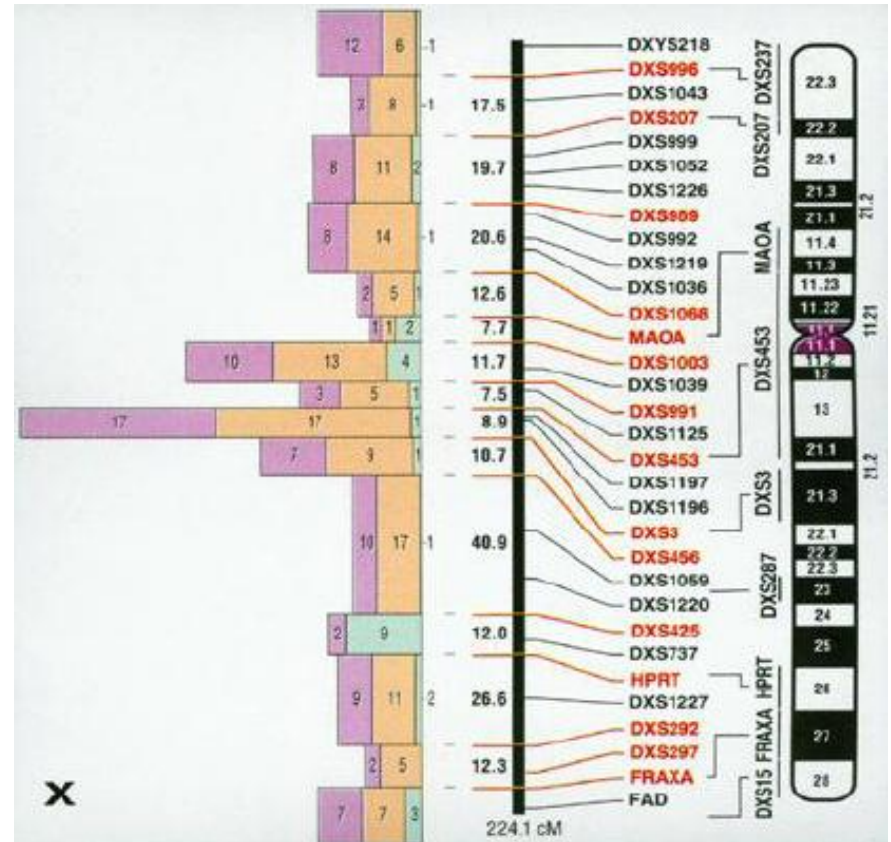
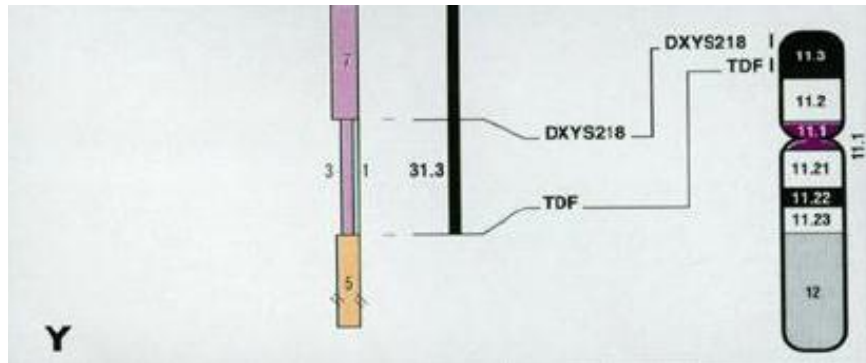
- cM: le pourcentage de la progéniture au sein de laquelle une recombinaison s'est produite entre deux marqueurs.
- Chez l'être humain: 1 cM  $\sim$  1 000 000pb.
- Plus deux marqueurs sont près l'un de l'autre, plus la probabilité qu'une recombinaison ait lieu entre eux est faible.



## Recombination frequencies

- Between m and v = 3.0%
- Between m and y = 33.7%
- Between v and w = 29.4%
- Between w and y = 1.3%

# Carte génétique des chromosomes X et Y humains



La carte génétique complète du génome humain peut être obtenue au lien suivant:  
<http://www.sciencemag.org/cgi/reprint/265/5181/2055.pdf>



# Objectif 2: Établir une carte physique du génome humain

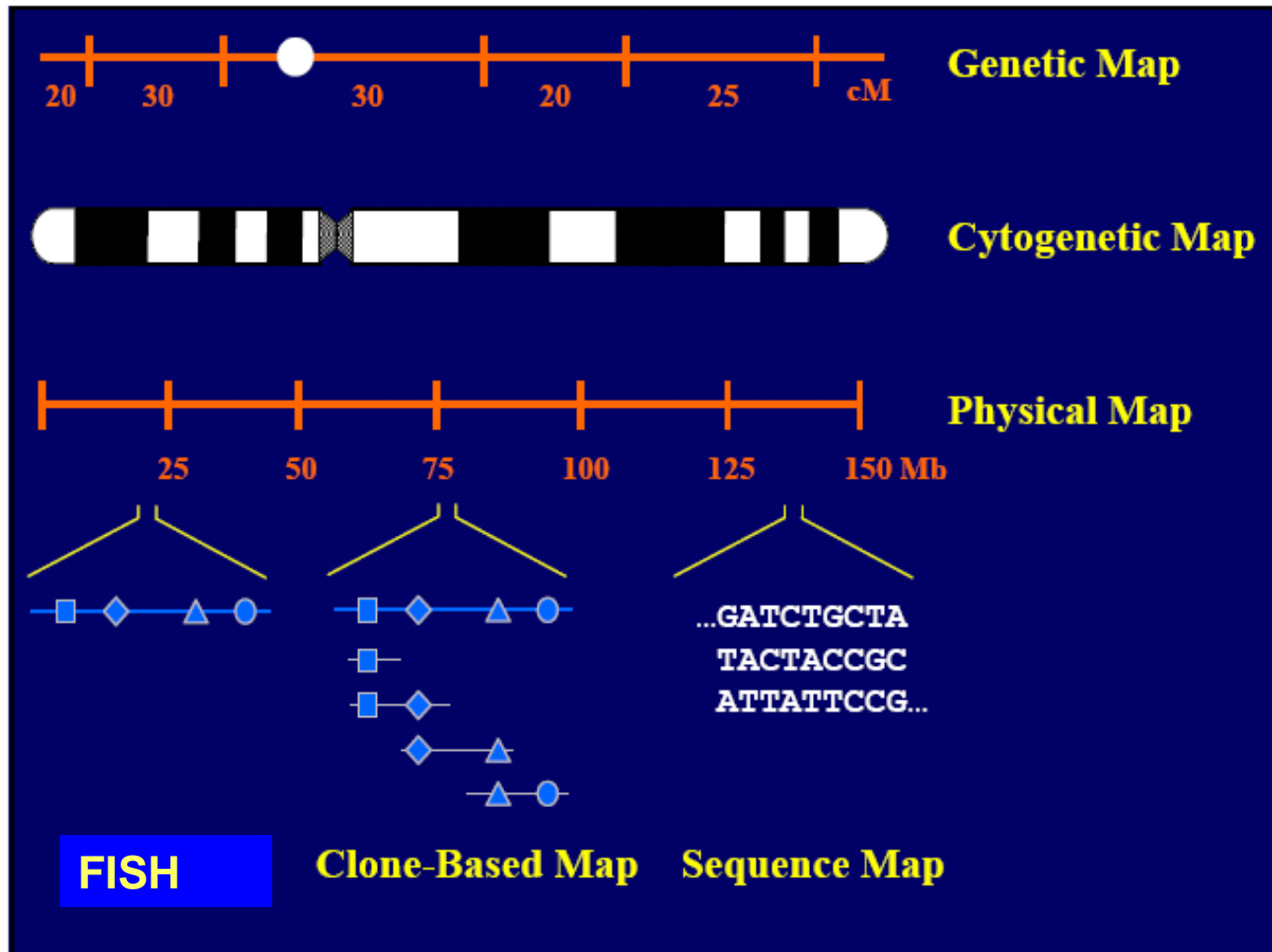
- **Carte physique:**

- Permet la localisation précise et l'isolation de gènes (*clonage positionnel*)
- Étude de l'organisation et de l'évolution des génomes
  - E.g. présence de polymorphismes, régions répétées, réarrangements (duplications, inversions), etc.
- Donne des points de repères pour le séquençage du génome

- **Types de cartes physiques:**

- Basées sur des marqueurs
  - Localisation sur le chromosome (e.g. FISH)
- Basées sur l'assemblage de clones d'ADN génomique
  - E.g. cartographie de restriction, utilisation de STS
- Séquence d'ADN (carte physique ultime)

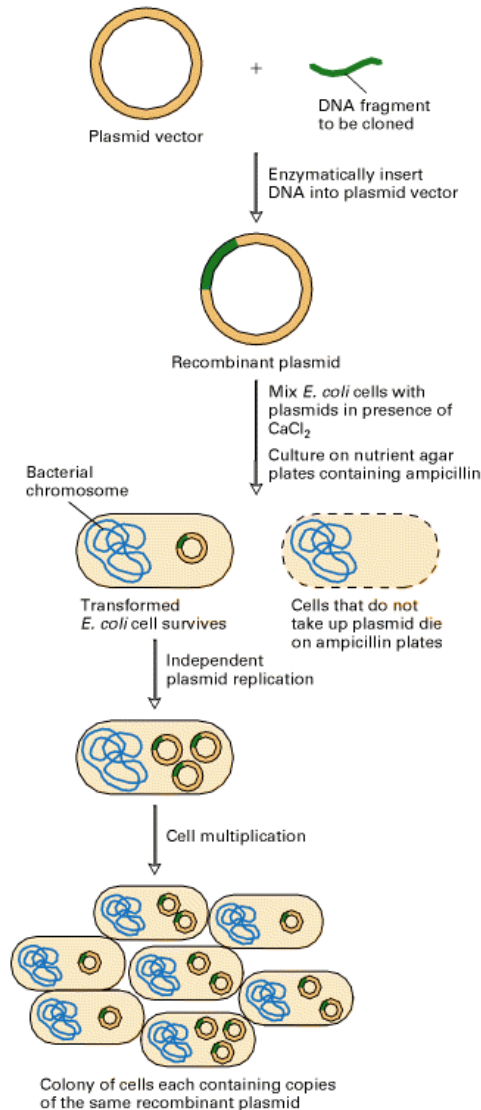
# Objectif 2: Établir une carte physique du génome humain



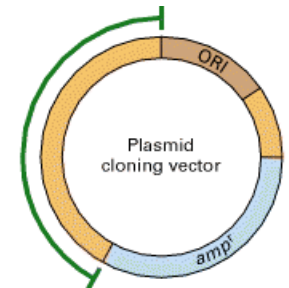


# Vecteurs de clonage

## Plasmides

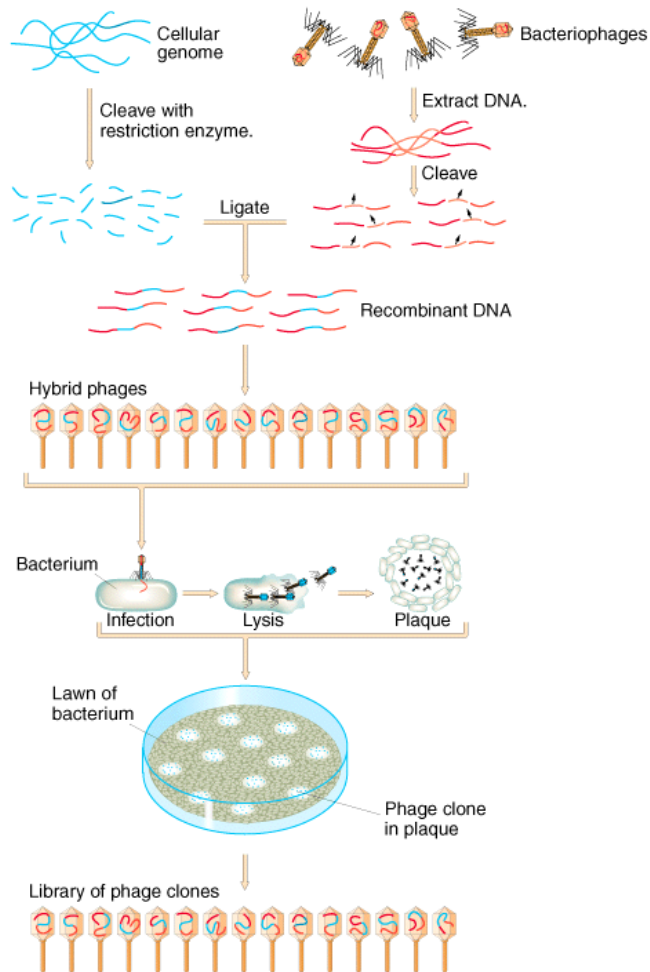


- Plasmide:
  - Petite molécule d'ADN circulaire
- Possède plusieurs caractéristiques importantes:
  - Séquence d'initiation de la réplication (donc: se réplique indépendamment de l'ADN bactérien)
  - Site de clonage (multiple cloning site - MCS)
  - Gène de résistance à un antibiotique (p.ex. ampicilline)
- On commence par couper l'ADN en petits morceaux de 0.5-3 kpb (p.ex. enzymes de restriction)
- On clone ces petits fragments dans des plasmides (donne des plasmides recombinants)
- L'ADN du plasmide se répliquera dans la bactérie hôte, et confèrera à la bactérie la résistance à un antibiotique.
- Avantage: faciles à manipuler/purifier
- Inconvénients: petite taille de l'insert.



# Vecteurs de clonage de génomes

## Bactériophage lambda ( $\lambda$ )

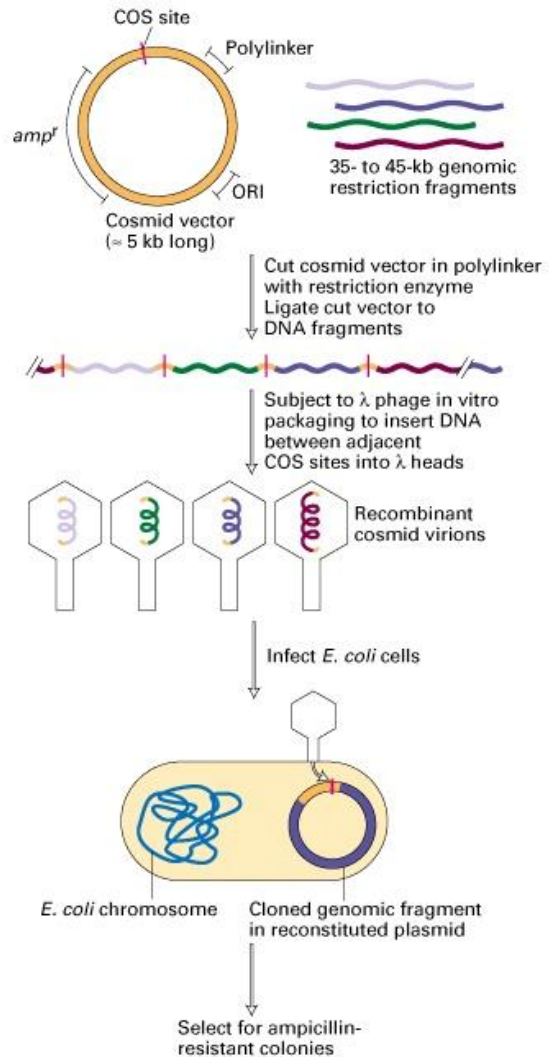


(a)

- Phage  $\lambda$ :
  - Bactériophage qui infecte *E. coli*
  - Site *cos*:
    - petite séquence d'ADN reconstruite chez les bactériophages
    - Essentiel à la réplication de empaquetage
- On commence par casser l'ADN génomique en petits morceaux de 10-15 kpb (p.ex. enzymes de restriction)
- On clone ces petits fragments dans des vecteurs  $\lambda$  (donne des phages recombinants)
- L'ADN du phage sera empaqueté dans une capsid virale
- Avantage: faciles à manipuler/purifier
- Inconvénients: petite taille de l'insert.

# Vecteurs de clonage de génomes

## Cosmides



- Cosmide:
  - Vecteur plasmide contenant ce qu'on appelle un site *cos*
  - Site *cos*:
    - petite séquence d'ADN reconstruite chez les bactériophages
    - Essentiel à la réplication de empaquetage
- On commence par casse l'ADN génomique en petits morceaux de 35-45 kpb (p.ex. sonication)
- On clone ces petits fragments dans des vecteurs cosmides (donne des cosmides recombinants)
- L'ADN situé entre 2 sites *cos* sera empaqueté dans une capsid virale
- Avantage: faciles à manipuler/purifier
- Inconvénients: petite taille de l'insert.

# Vecteurs de clonage de génomes

## Cosmides

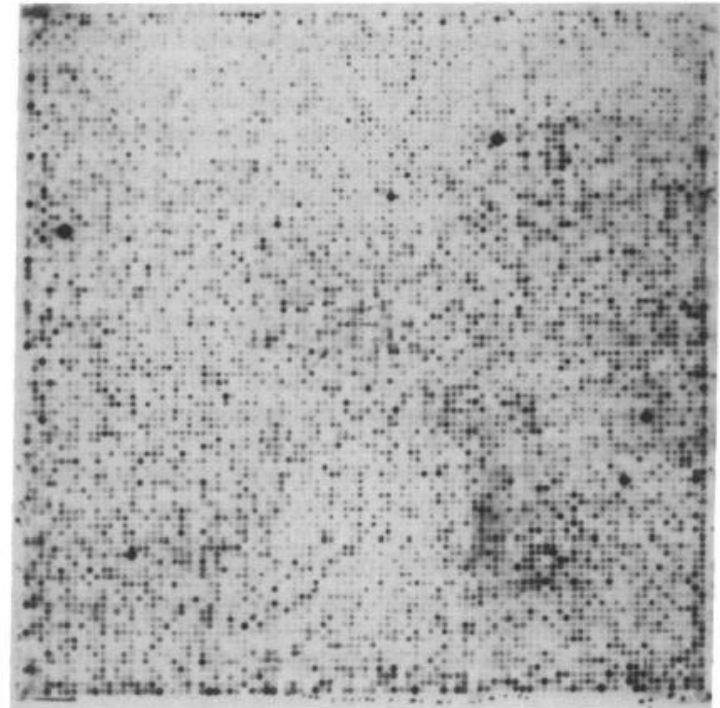
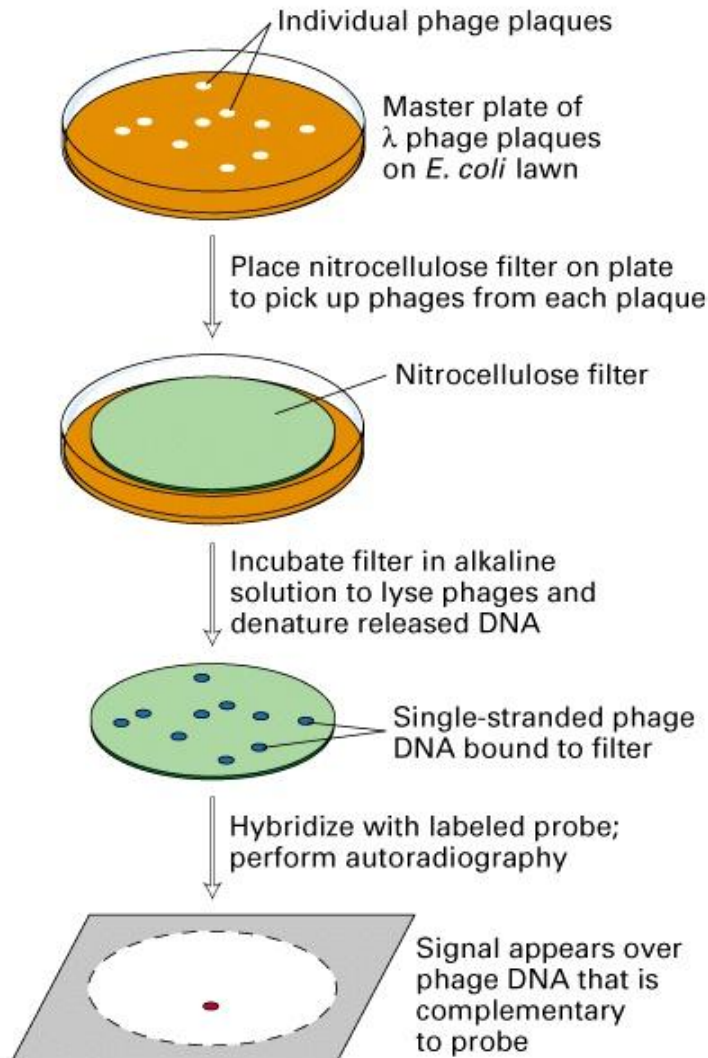
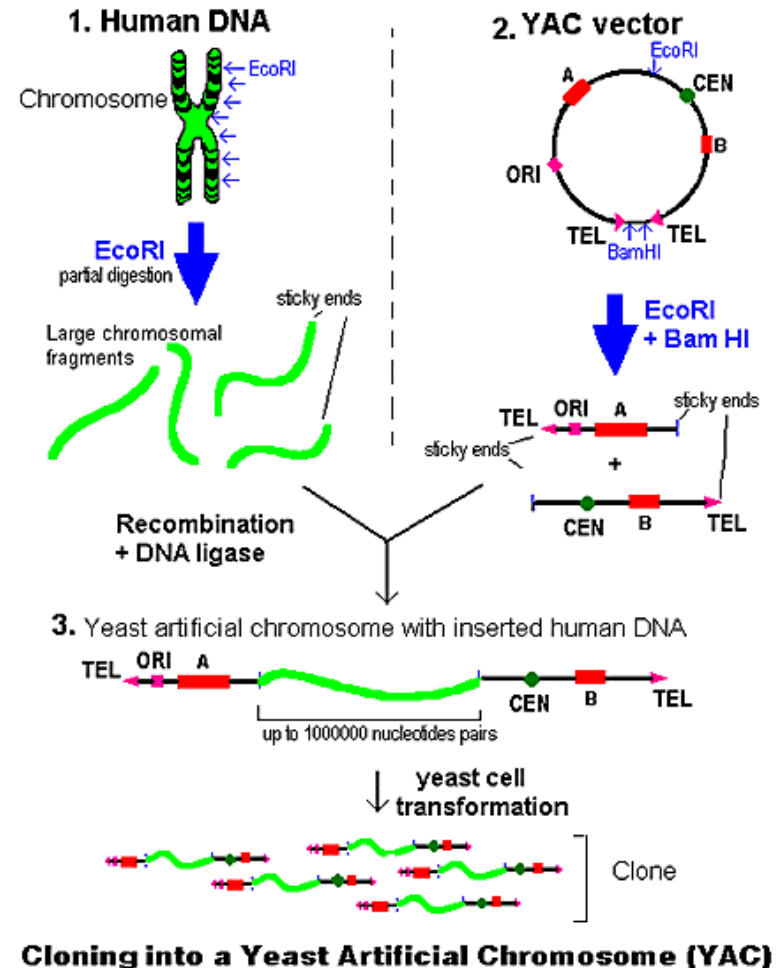


FIG. 2. High-density arrayed display of flow-sorted human X chromosome cosmid library. The cosmid cultures from 96 microtiter plates (9216 wells) were transferred by sterilized metal pins onto a 22 × 22-cm nylon membrane prewetted in medium. Transfer was achieved using a computer-controlled precision X-Y-Z moving device modified in our laboratory for this application. Membranes were incubated on agar plates with 2× YT broth and kanamycin (30  $\mu\text{g}/\text{ml}$ ) overnight at 37°C until colonies were 1–3 mm in diameter. The DNA from the colonies was then *in situ* prepared, denatured, and bound to the membrane. This is the autoradiogram of the membrane after hybridization to a dodecanucleotide probe with the sequence present in the cosmid vector.

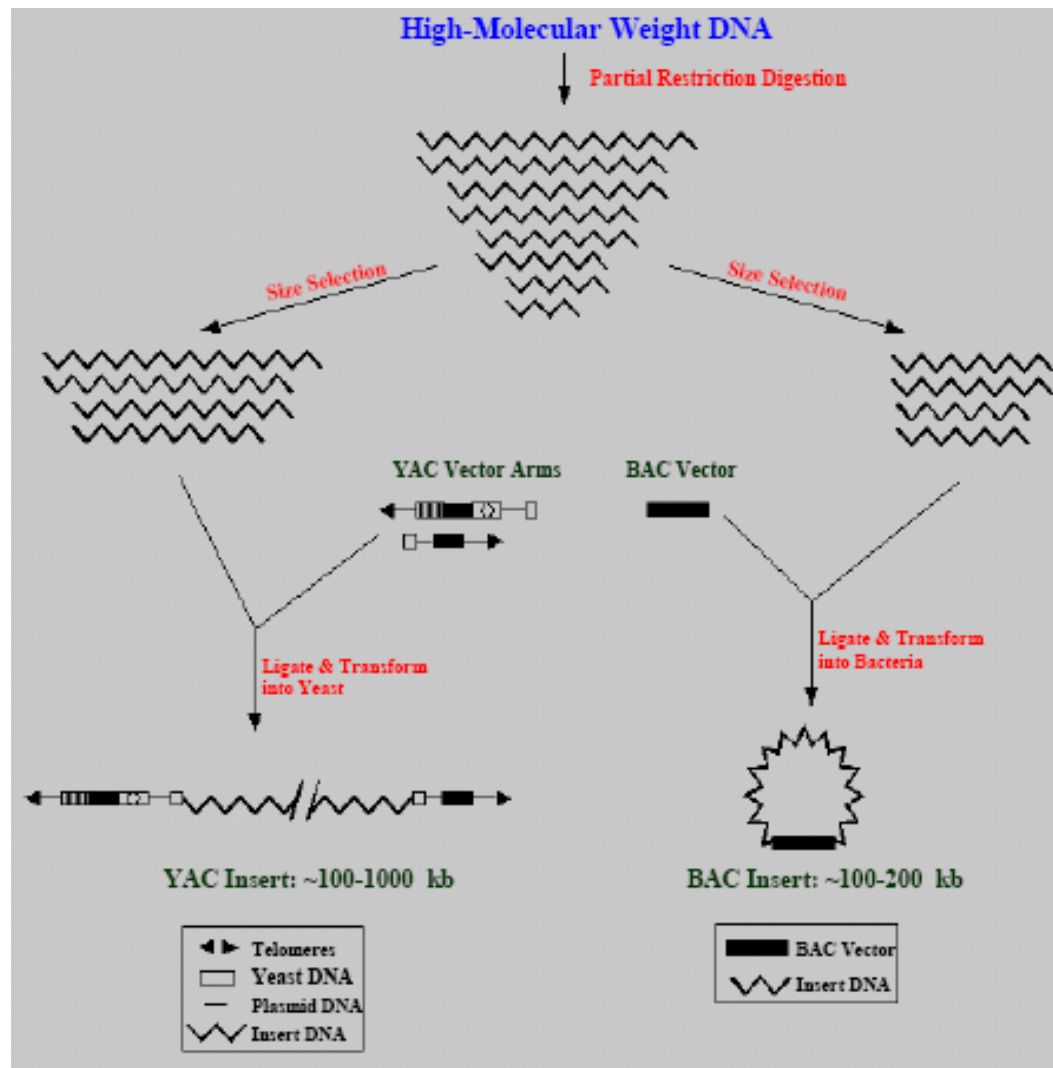
# Vecteurs de clonage de génomes

## Chromosomes artificiels de levure (YACs) et de bactéries (BACs)

- YACs:
  - Possèdent trois composantes retrouvées dans les chromosomes de levures:
    - Origine de réplication (ARS)
    - Télomère (TEL)
    - Centromère (CEN)
- BACs:
  - Similaires aux plasmides
  - Possèdent une origine de réplication de phage (séquence F1) qui leur permet d'accepter de gros fragments d'ADN
- On débute généralement par
  - Isoler un chromosome particulier
  - Couper le chromosome en gros fragments assez gros;
  - Cloner le fragment dans le YAC/BAC.
- Avantages: taille de l'insert:
  - YACS: 100 à 1000 kpb
  - BACs: 100 à 200 kbp
  - BACS: culture dans *E. coli* (plus facile à manipuler)
  - BACS: plus stables que YACs
- Problèmes:
  - Instables
  - Difficile à purifier (p.ex pour séquençage)
  - Croissance lente



# Clonage de longs fragments d'ADN génomique



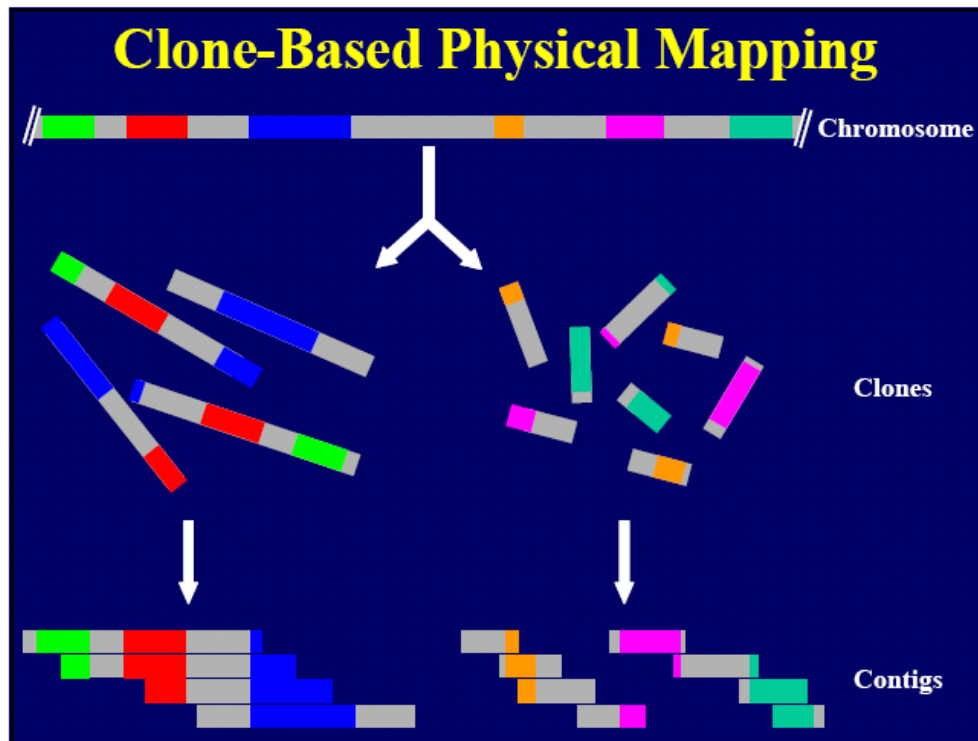


# Quel vecteur choisir?

Vector	Chromosome 19		Human Genome	
	1X	5X	1X	5X
YAC	34	170	2,667	13,335
Cosmid	1,251	6,255	100,002	500,010

- Les chiffres indiquent le nombre de clones de YACs ou de cosmides nécessaires pour cloner l'équivalent de 1 (1X) ou 5 (5X) copies du chromosome 19 ou du génome humain.

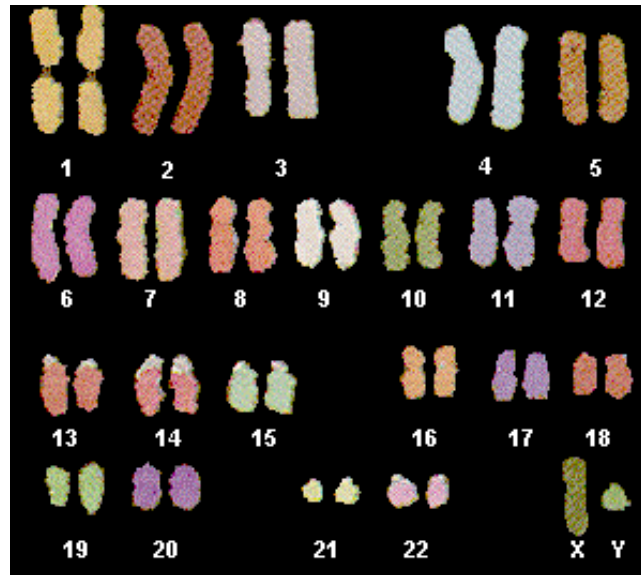
# Assemblage de clones en « contigs »



- Consiste à:
  - Couper le génome en plus petits morceaux;
  - Cloner ces morceaux dans des vecteurs spécialisés
  - Ordonner les fragments clonés
- Ce travail facilitera grandement le séquençage en permettant aux scientifiques:
  - d'avoir accès à une banque de fragments d'ADN génomique clonés
  - de travailler avec des plus petits morceaux d'ADN



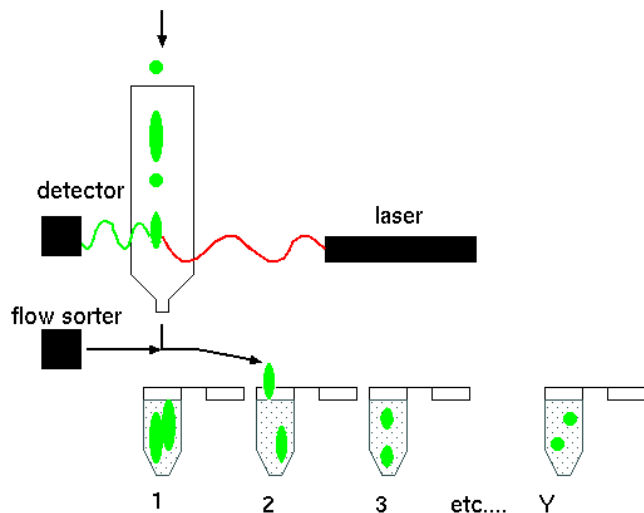
# Isolation de chromosomes



- Il est possible de marquer chaque paire de chromosome avec des sondes fluorescentes différentes;

- Ceci permet de séparer les chromosomes par cytométrie de flux.

- Chaque chromosome peut être ensuite coupé en plus petits morceaux et cloné dans un vecteur approprié.



# Objectif 2: Établir une carte physique du génome humain

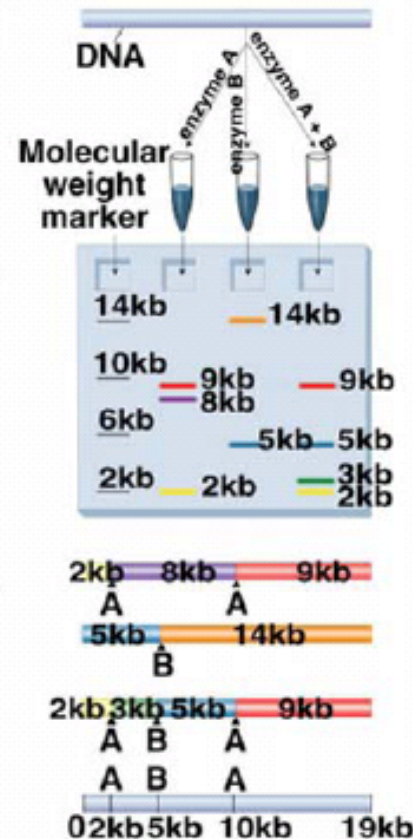
## A. Cartographie avec enzymes de restriction

1. Multiple copies of a segment of DNA are cut with restriction enzymes.

2. The fragments produced by enzyme A only, by enzyme B only, and by enzymes A and B simultaneously are run out side-by-side on a gel, which separates them according to size, smaller fragments running faster.

3. The fragments are arranged so that the smaller ones produced by the simultaneous cut can be grouped to generate the larger ones produced by the individual enzymes.

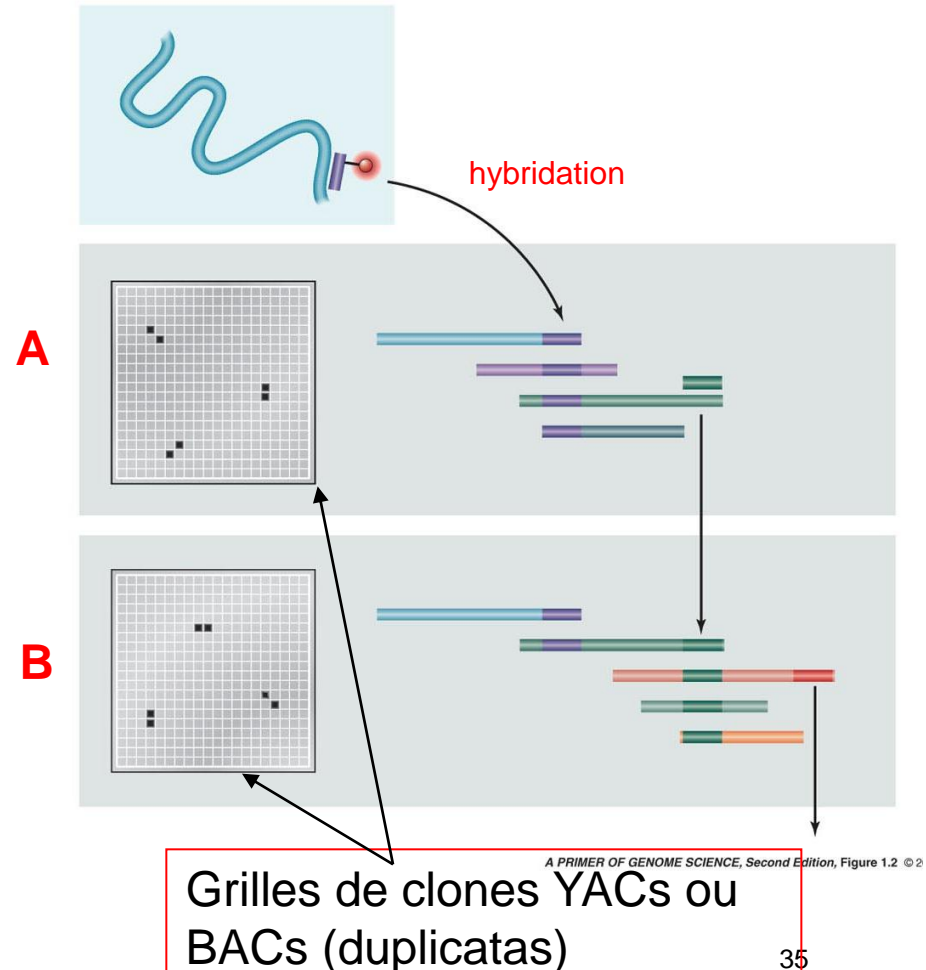
4. A physical map is constructed.



# Objectif 2: Établir une carte physique du génome humain

## B. Cartographie avec « sequence tagged sites »

- Sequence Tagged Sites:
  - Petites séquences d'ADN génomiques
    - Environ 30 000 furent utilisées pour la cartographie du génome humain
  - Chaque STS est unique
  - Dérivés de régions **séquencées** du génome (bouts de clones BACs ou YACs)
  - Permet de cartographier de longs fragments d'ADN génomique (deux fragments d'ADN reconnus par le même STS doivent obligatoirement se chevaucher).



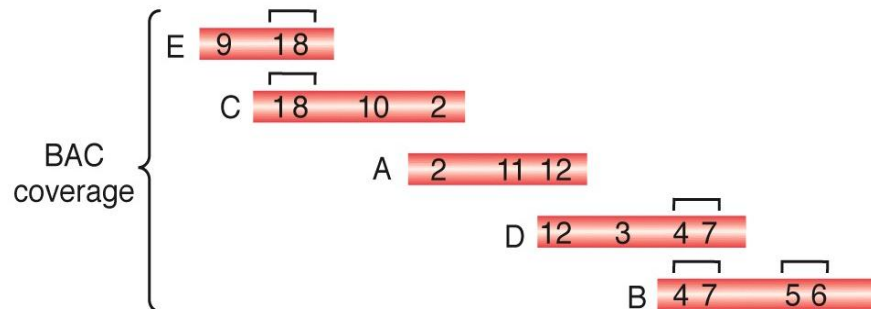
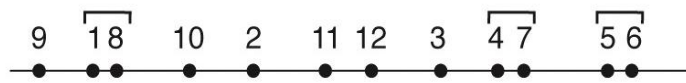
# Objectif 2: Établir une carte physique du génome humain

## B. Cartographie avec « sequence tagged sites »

DATA (STS content of BACs)

		STSs (Sequence-tagged sites)											
		1	2	3	4	5	6	7	8	9	10	11	12
BACs	A	-	+	-	-	-	-	-	-	-	-	+	+
	B	-	-	-	+	+	+	+	-	-	-	-	-
	C	+	+	-	-	-	-	-	+	-	+	-	-
	D	-	-	+	+	-	-	+	-	-	-	-	+
	E	+	-	-	-	-	-	-	+	+	-	-	-

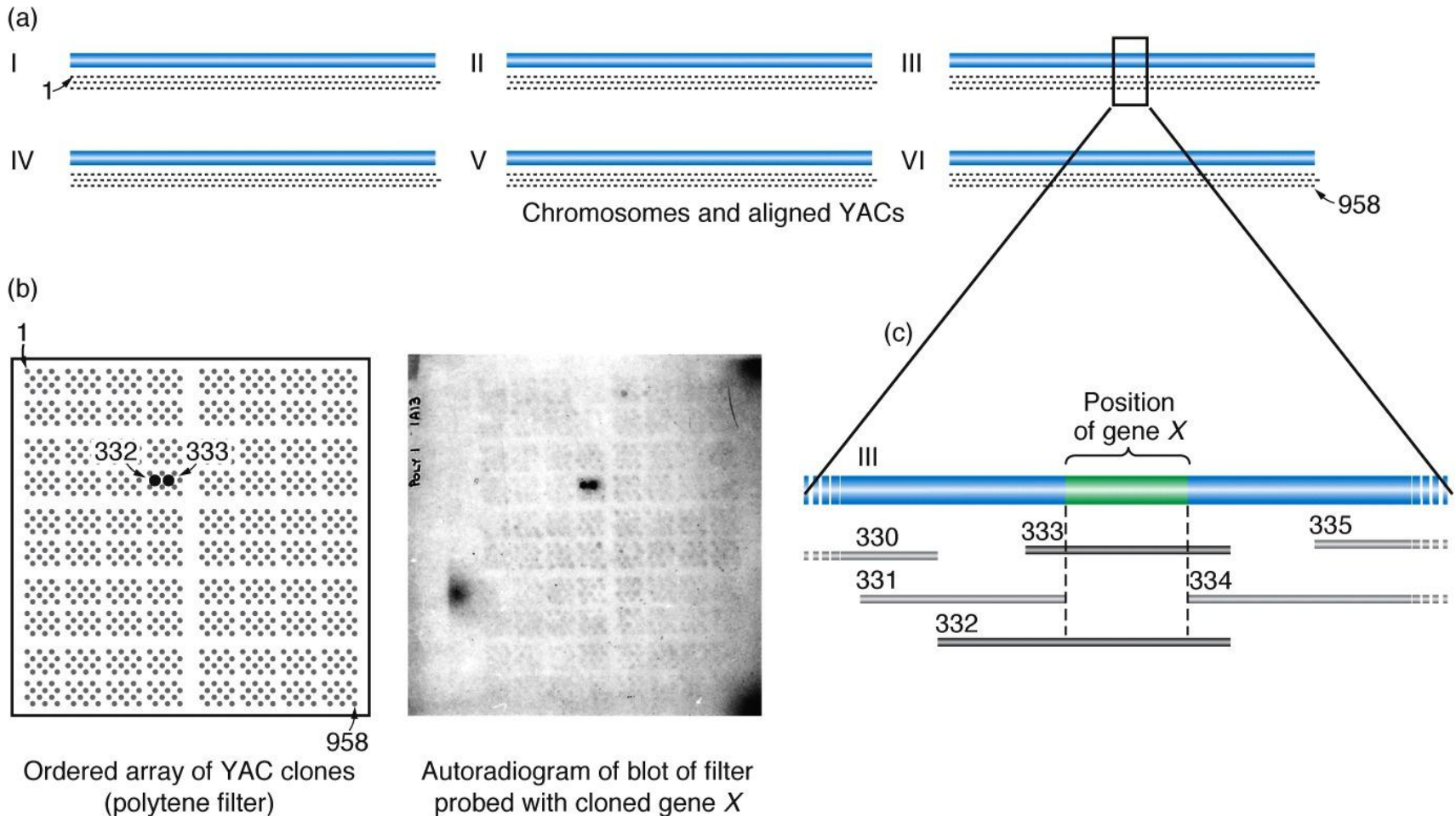
CONTIG  
STS map



Order uncertain

# Objectif 2: Établir une carte physique du génome humain

## B. Cartographie avec « sequence tagged sites »



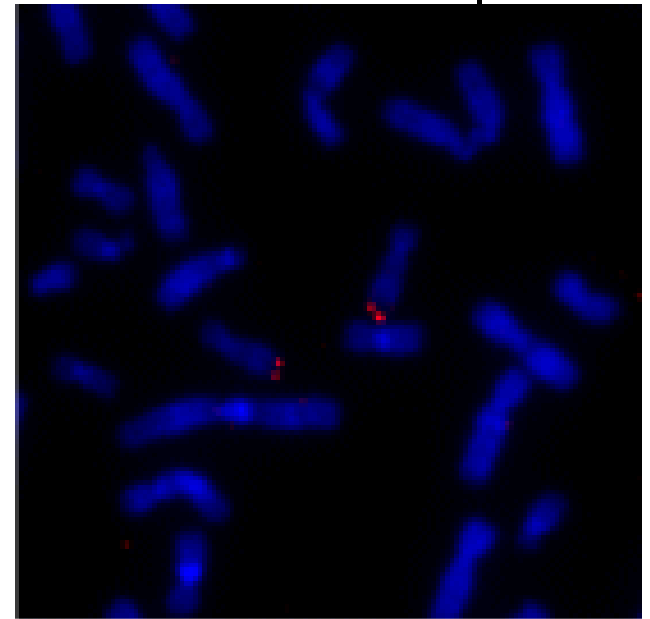
# Fluorescence in situ hybridization (FISH)

- Utilisation de sondes d'ADN (p.ex. STS) marquées avec une molécule fluorescente
- La sonde est incubée avec les chromosomes de l'organisme d'intérêt
- La formation de paires de bases entre la sonde et un chromosome indiquera la position de la séquence complémentaire à la sonde le long du chromosome.

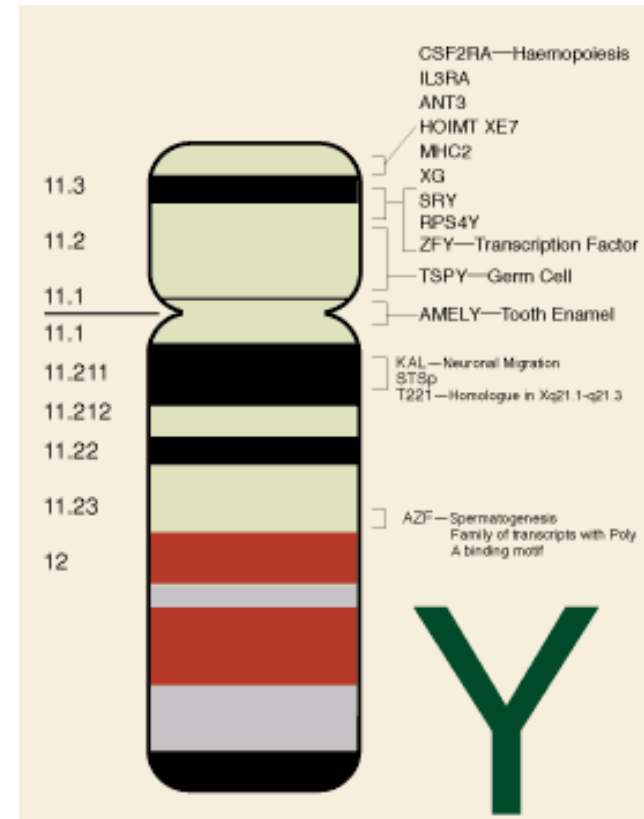
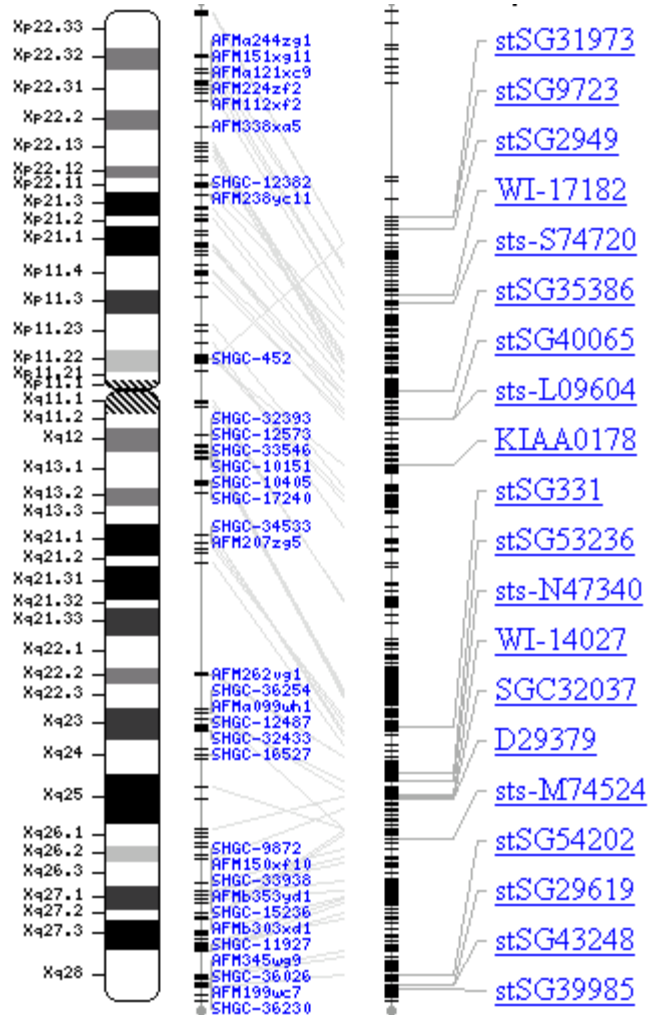


Coloration au Giemsa

FISH out 1.4 kb probe



# Carte physique des chromosomes X et Y humains





# Intégration des cartes génétiques et physiques

