

Voyage Au Cœur Du Génome

Jean-Philippe.Vert@ensmp.fr

ENSMP - CG - Bioinformatique

17 mars 2005

Une brève histoire

1866 : Lois de l'hérédité (Mendel)

1909 : Morgan et les drosophilistes

1944 : L'ADN est le support de l'hérédité (Avery)

1953 : Structure de l'ADN (Crick et Watson)

1966 : Code génétique (Nirenberg)

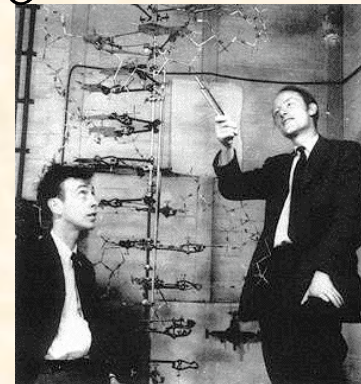
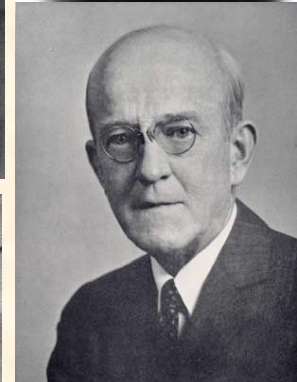
1960-70 : Génie génétique

1977 : Méthode de séquençage (Sanger)

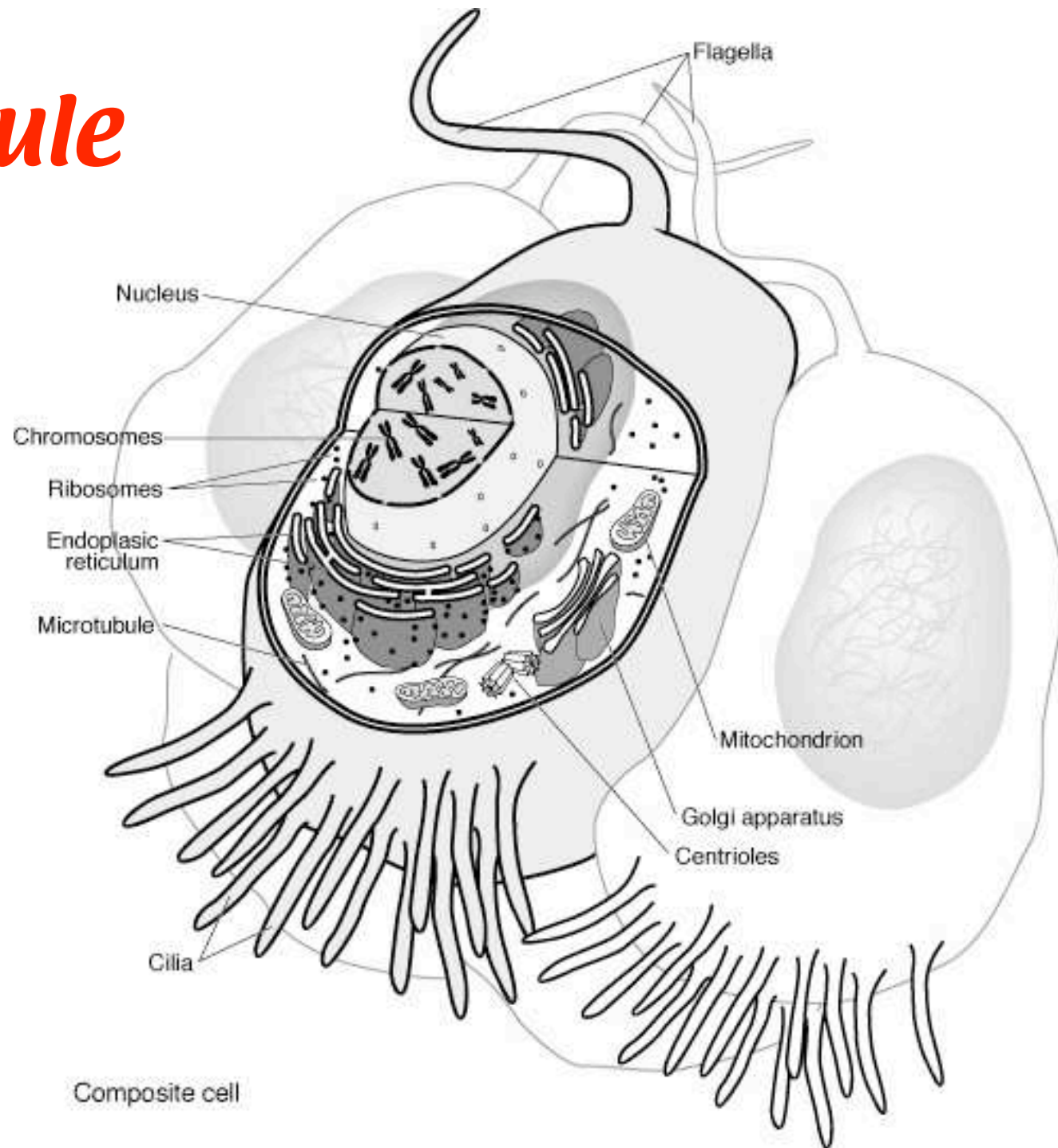
1982 : Création de Genbank

1990 : Lancement du Projet Génome Humain

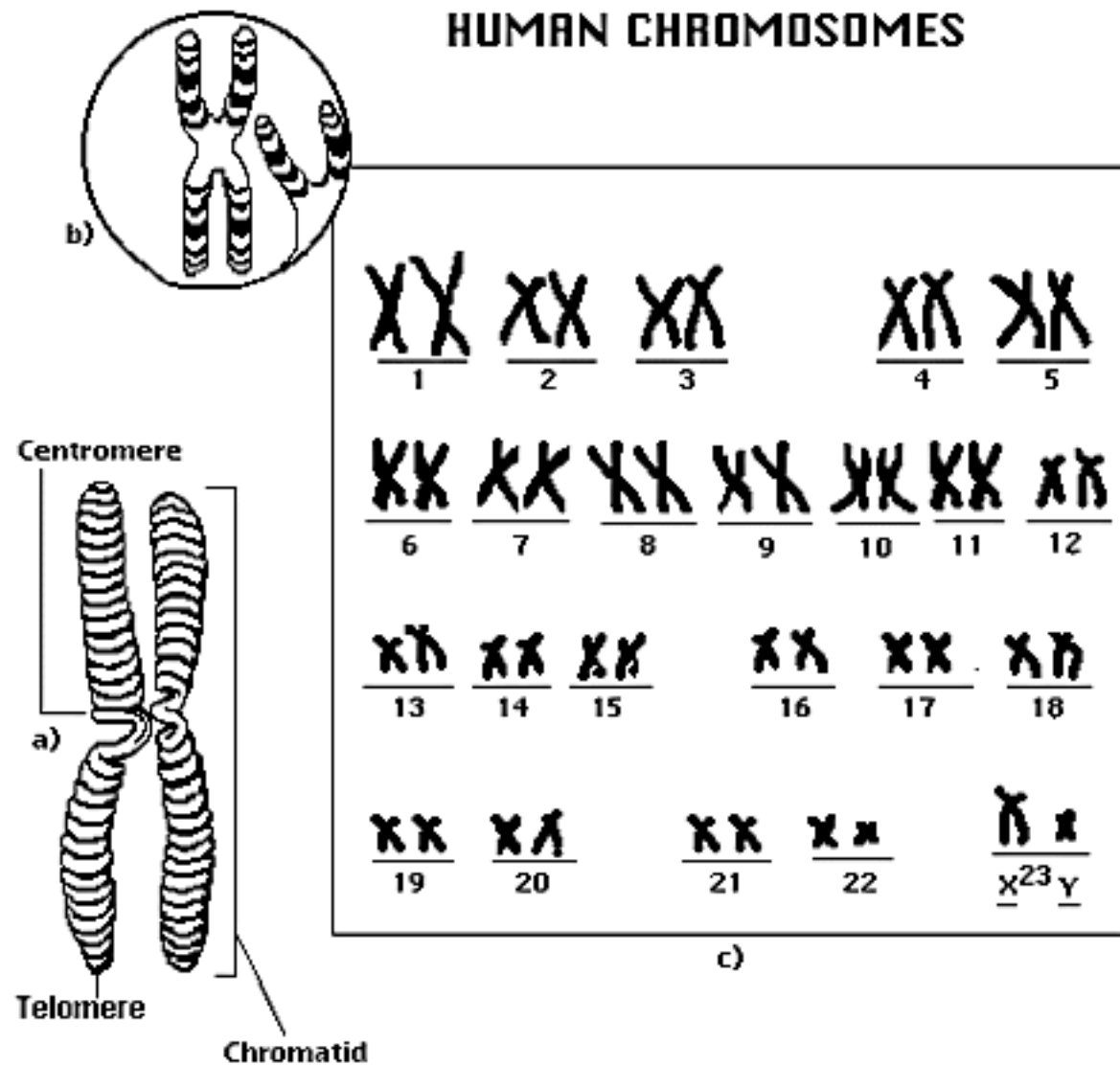
2003 : Complétion du Projet Génome Humain



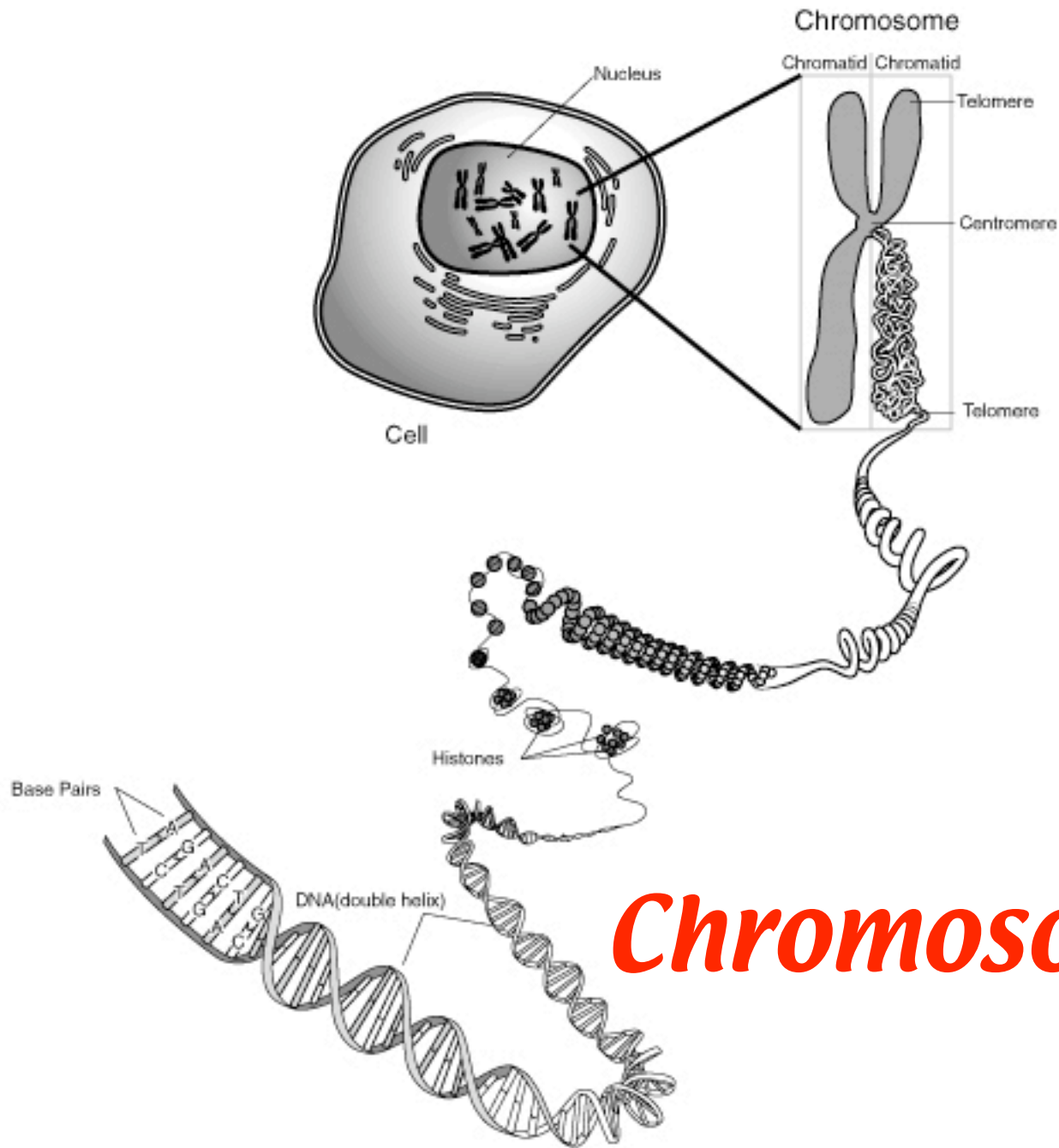
La cellule



HUMAN CHROMOSOMES

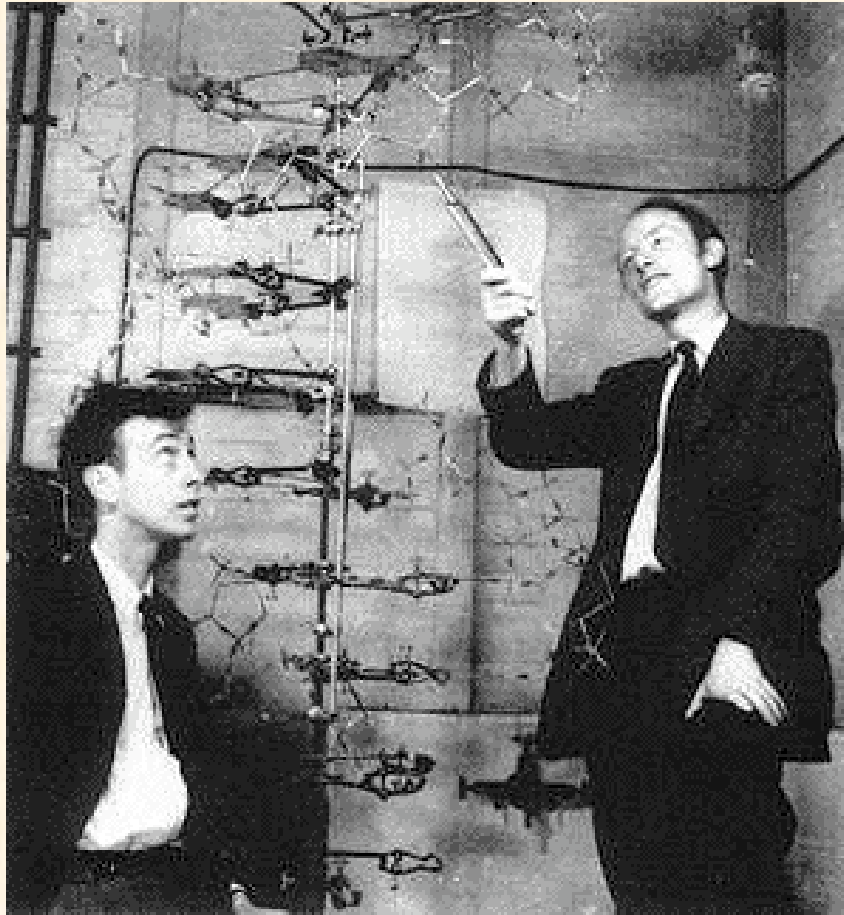


Chromosomes



Chromosomes et ADN

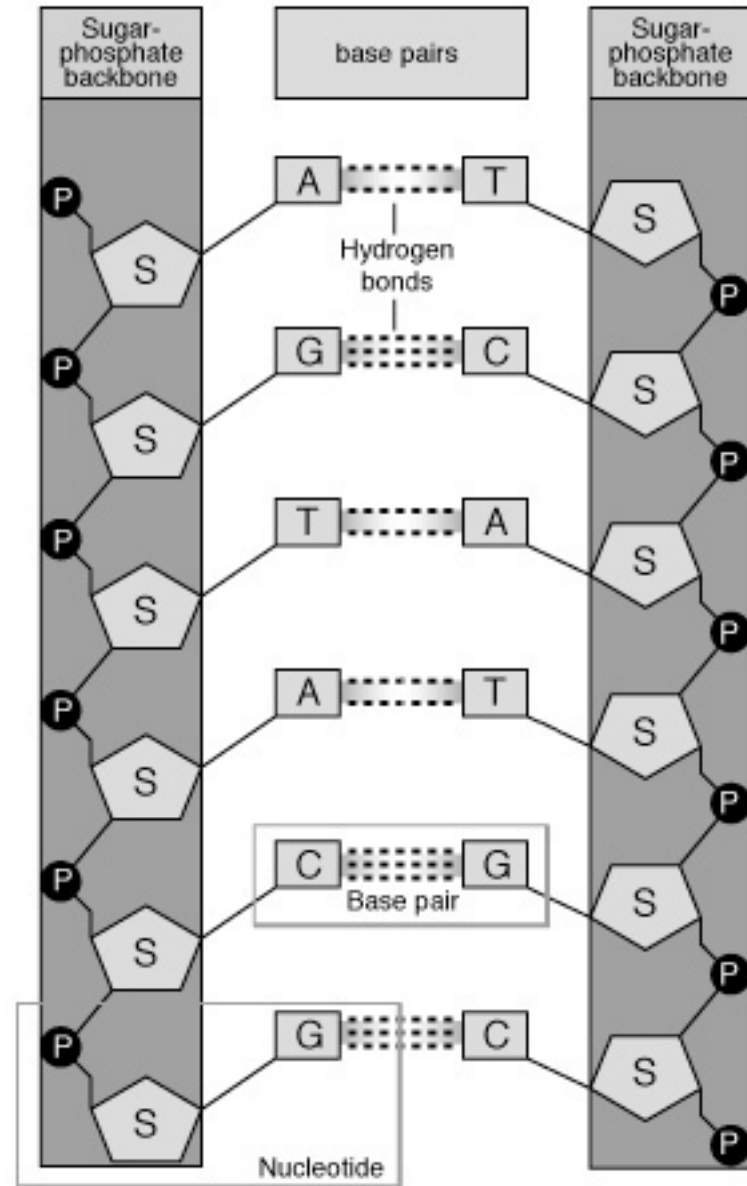
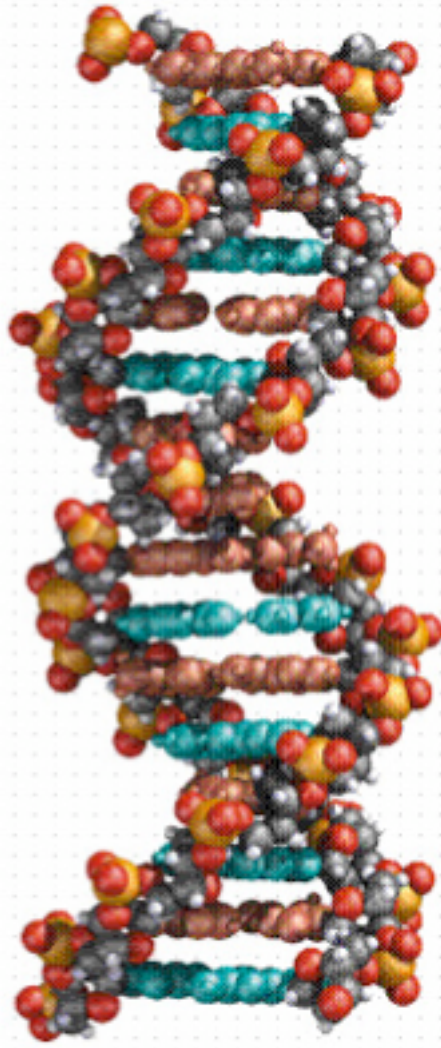
Structure de l'ADN



« We wish to suggest a structure for the salt of desoxyribose nucleic acid (D.N.A.). This structure have novel features which are of considerable biological interest. »

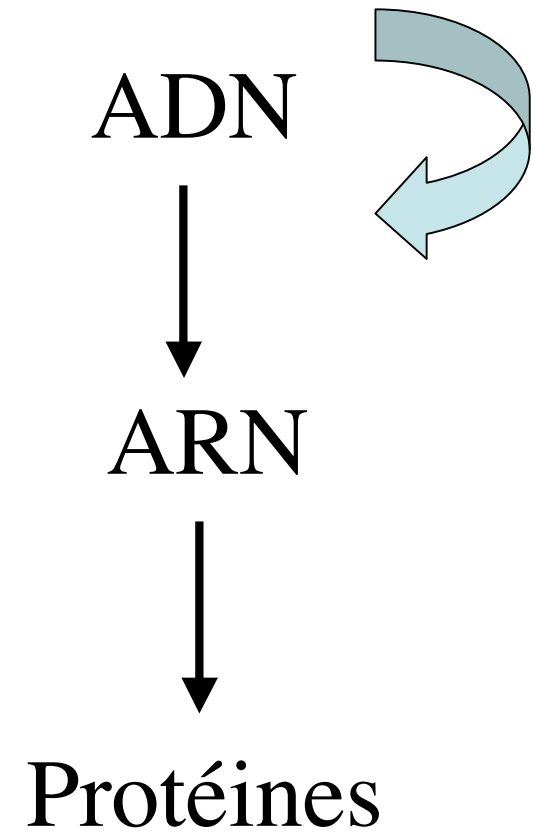
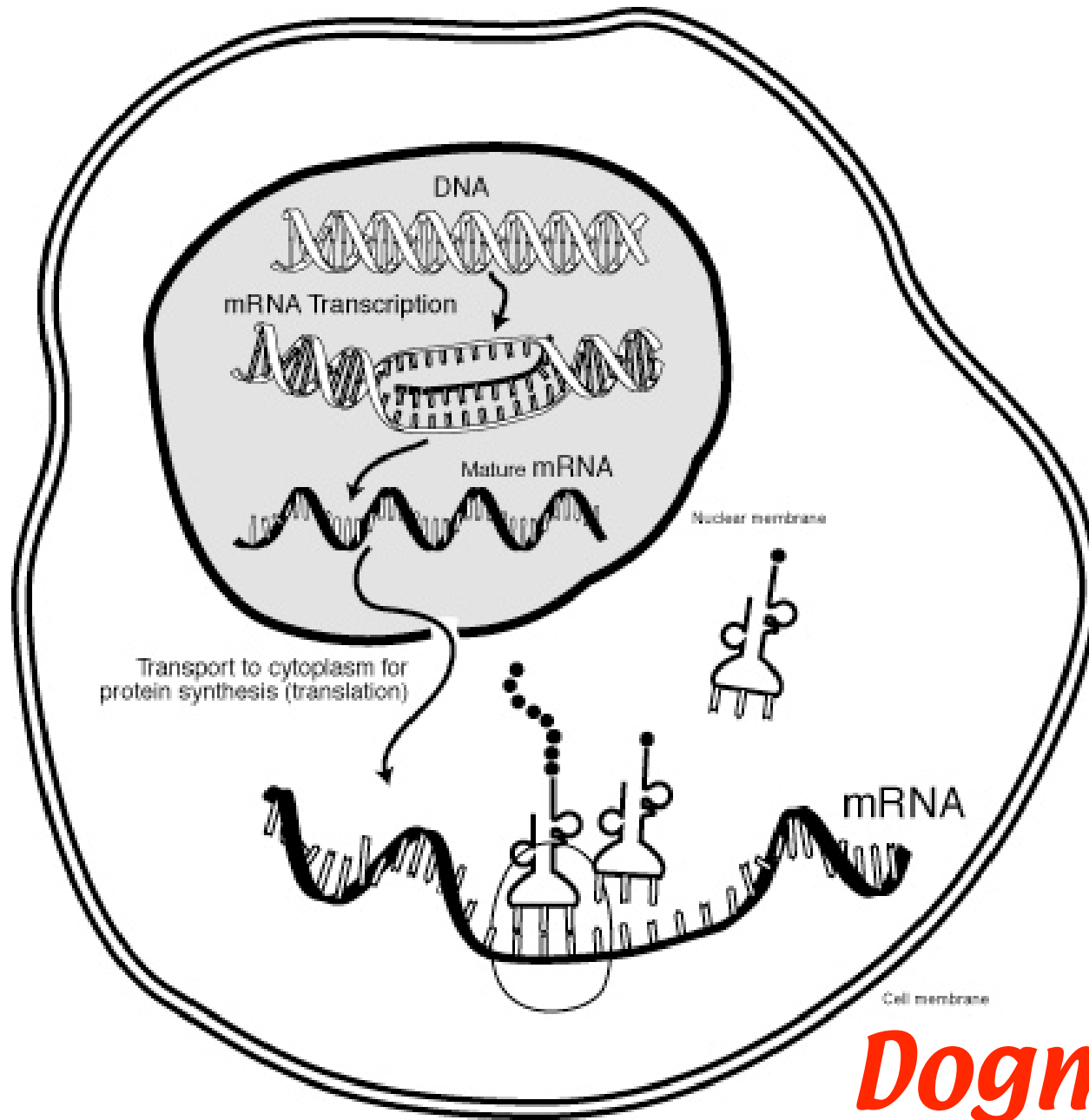
J.D. Watson & F.H.C. Crick (1953) Molecular structure of Nucleic Acids. *Nature*, **171**:737-738.

La double hélice d'ADN



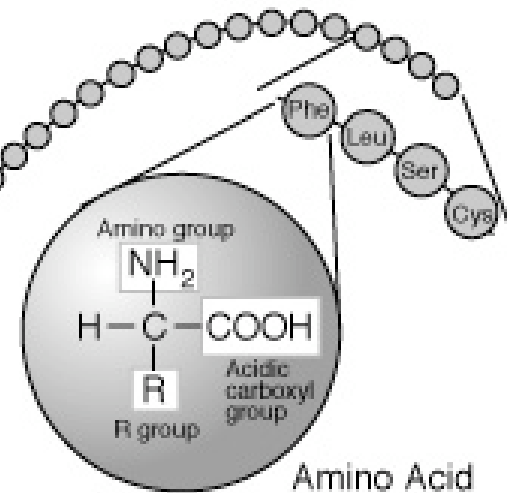
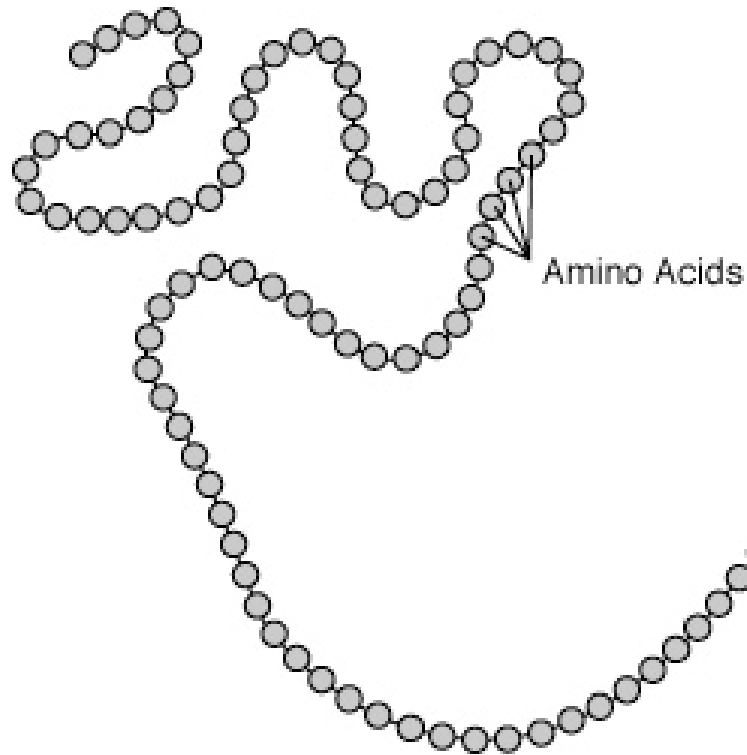
Génomes

Organisme	Chromosomes	Taille du génome
Bactéries	1	400,000 à 10,000,000
Levure	12	14,000,000
Mouche	4	300,000,000
Homme	46	6,000,000,000



Dogme central

Protéines



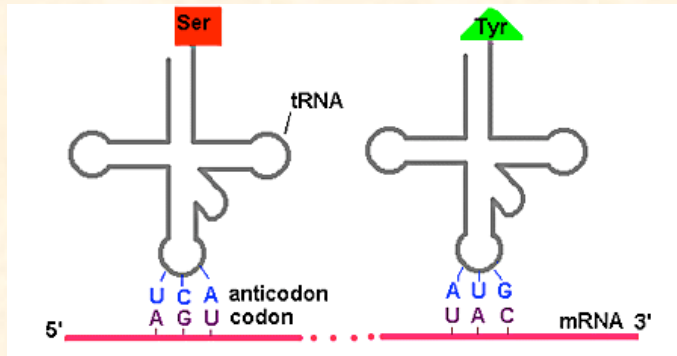
ADN = 4 lettres (ATCG)



ARN = 4 lettres (AUCG)



Protéine = 20 lettres (Acides Aminés)



2nd base in codon

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

1st base in codon

3rd base in codon

The Genetic Code

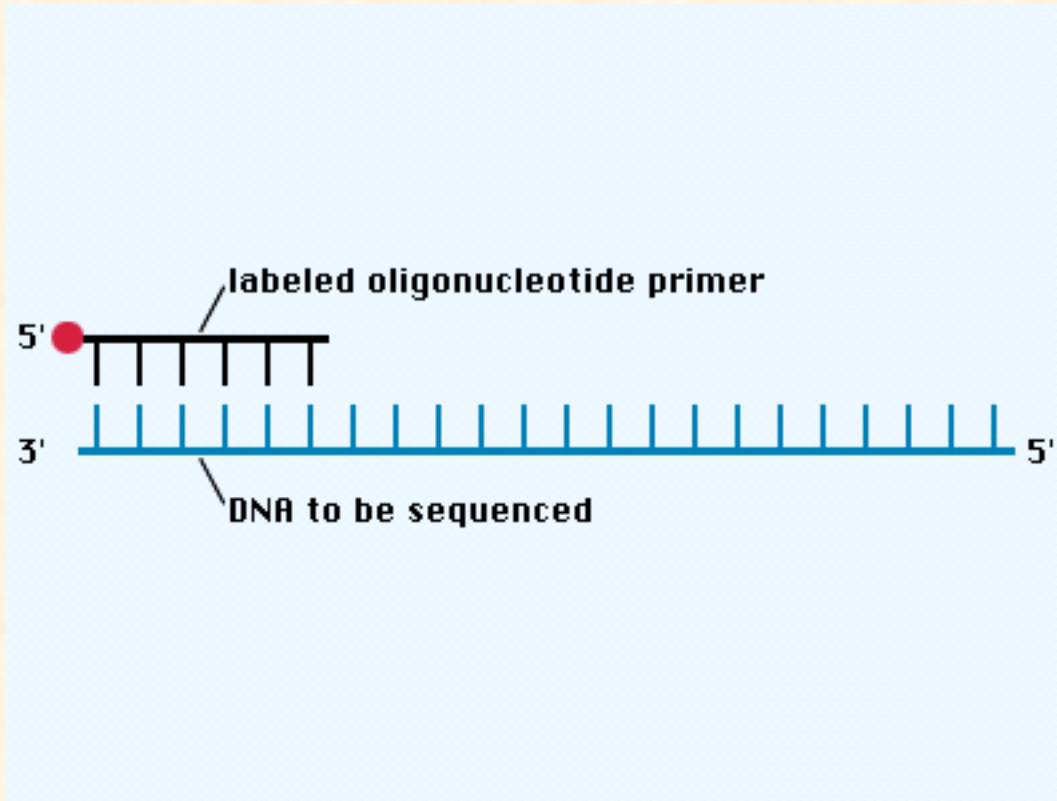
1 acide aminé

=

3 nucléotides

Code génétique

Séquençage de génome

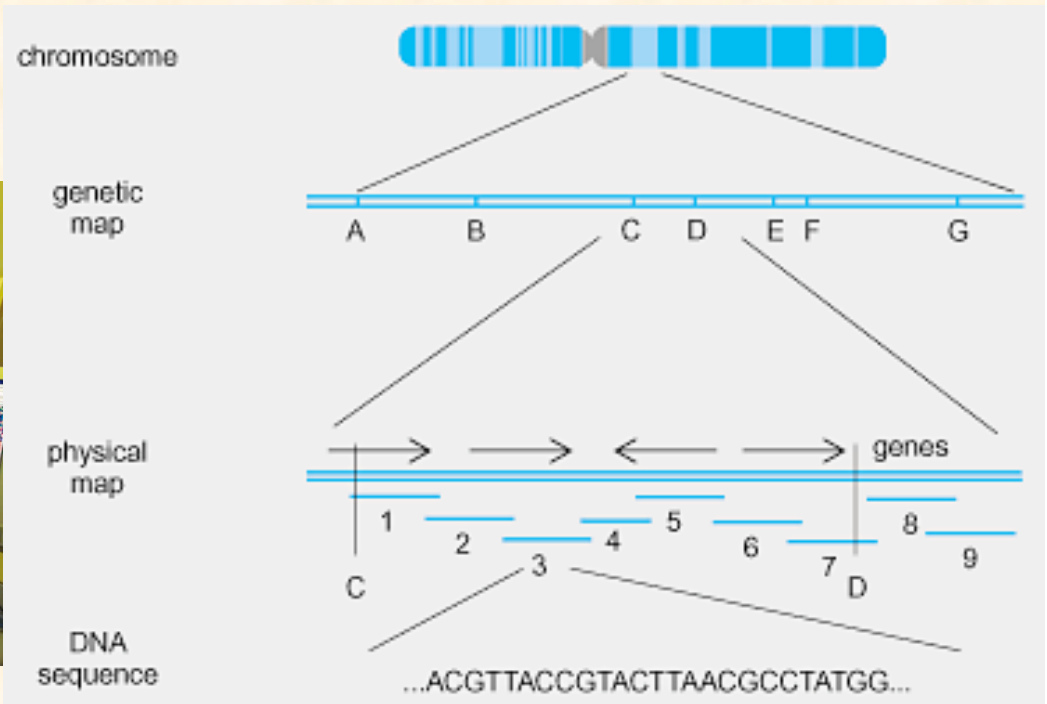


Frederick Sanger, Nobel 1958/1980

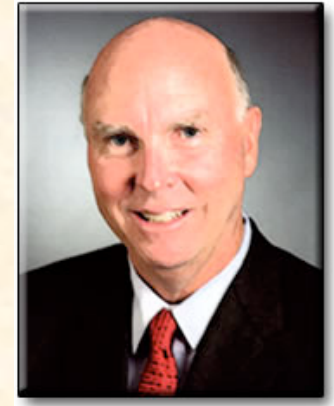


Projet Génome Humain (1990-2003)

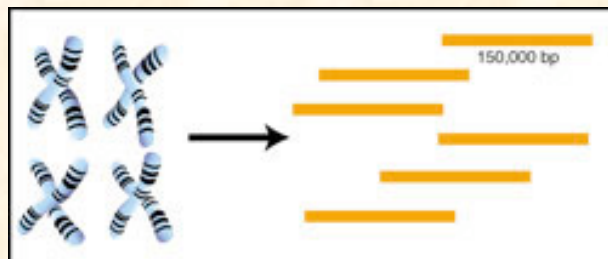
- But : séquencer les 3,000,000,000 de bases du génome humain
- Consortium de 20 laboratoires, 6 pays
- Coût : environ \$3,000,000,000



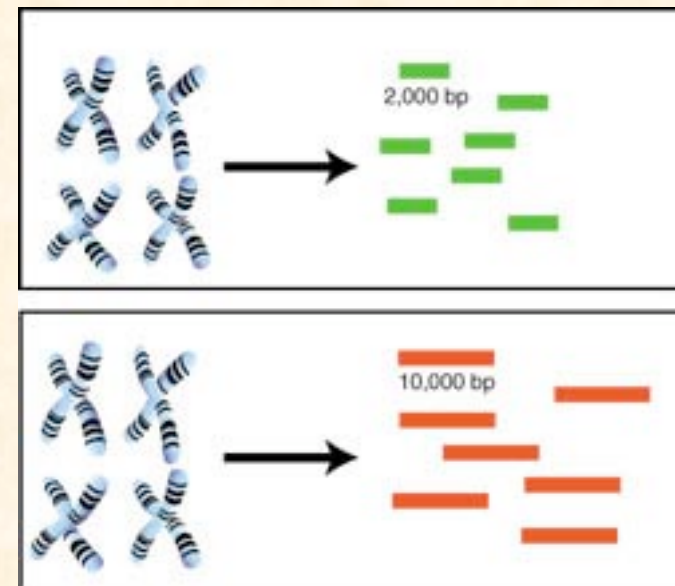
La course Public - Celera (1998-2003)



BAC to BAC
(HGC)



Whole genome
shotgun (Celera)



Février 2001 : Ebauches du génome humain



Avril 2003: Version finale!

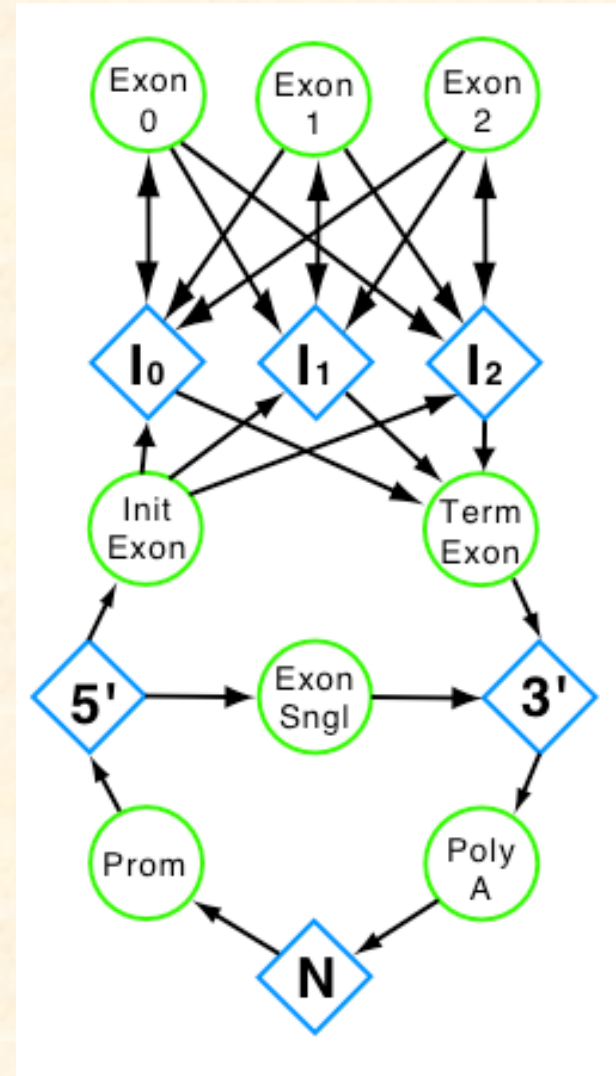
50 Years of DNA: *From Double Helix to Health*
A Celebration of the Genome

Annotation des génomes

Détection de **gènes codants** :

- Comparaison avec ARN séquencés
- Prédiction in silico (modèles probabilistes)
- Génomique comparative

Détection d'autres signaux (régulation. gènes non codants) difficile



Première découvertes...

- Peu de gènes (20~25,000)
- Les régions codantes ne représente que 1,2% du génome
- On ignore le rôle de 97% de l'ADN
= **ADN poubelle** (*junk DNA*)

La théorie de «l'ADN égoïste»

- 40% du génome est composé d'**éléments transposables** qui se répliquent
- 1,000,000 de copies de Alu (300bp), mais peu actifs
- Leur seule fonction est de **survivre** dans le génome



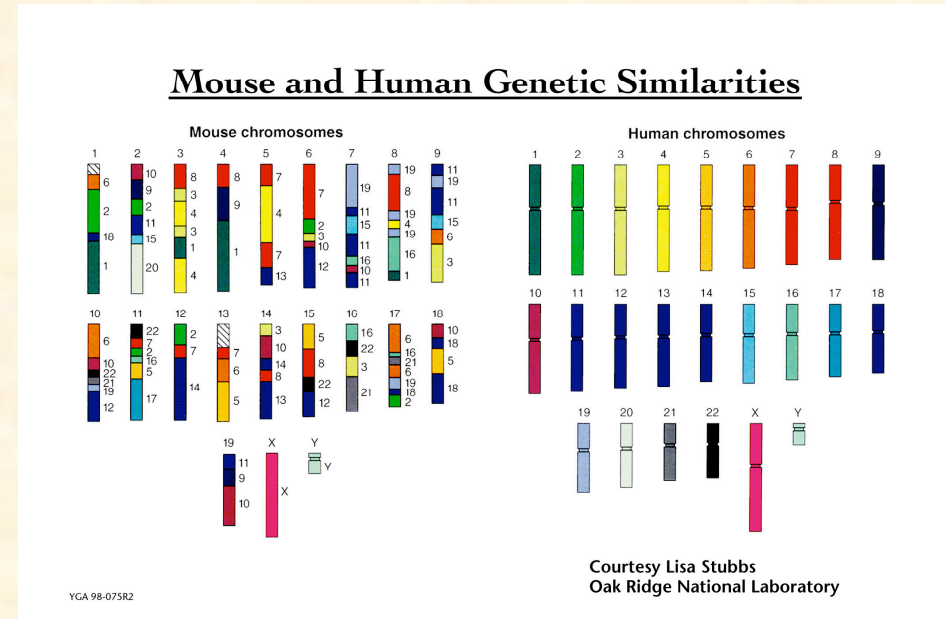


Empruntes phylogénétiques

99% de gènes homologues entre l'homme et la souris

5% du génome sous pression sélective, dont 1,2% codant

Que sont les 3,8% d'ADN poubelle sous pression de sélection?



	42	43
Souris	TAAGCAGTGGCAGGGC--CAG-GCTGAGCTTATCAGTCTCCAGCCAGCCCTGCCACACACATATATAGCCA	
Lapin	GGAGCAGTGACTAGGC--CCA-GCTGGGCITATCAGCCTCACAGCCAGCCCTGCCTGGAGACATAAATAGGCCA	
Homme	TGAGCAGCAACAGGGC--CAGGGCTGGGCTTATCAGCCTCCAGCCAGACCCTGGCTGCAGACATAAATAGGCC	
Poulet	GAGTGATTTCTTGGGCTGCGGCGCTG-GCITATCTGGTGC GGA ACT--GCCCTGG-TG----CATAAATAGCGGC	

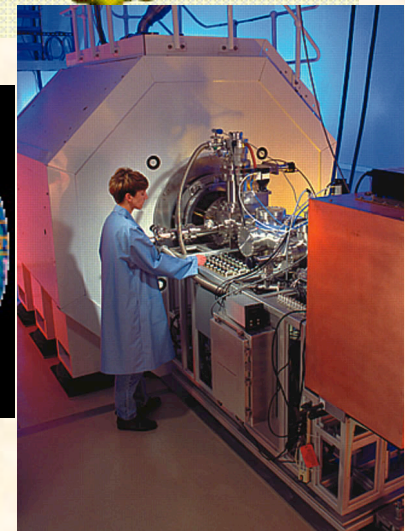
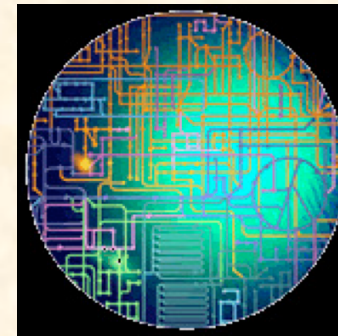
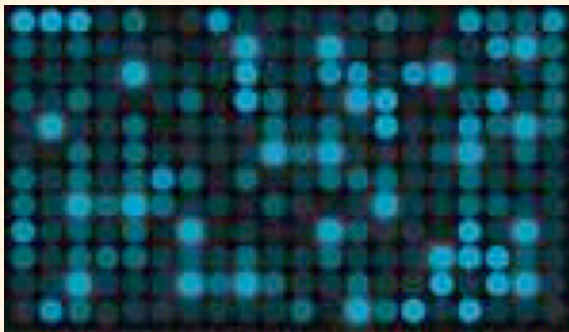
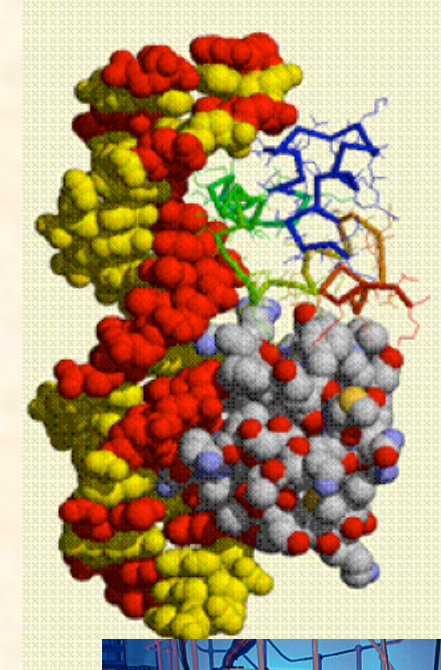
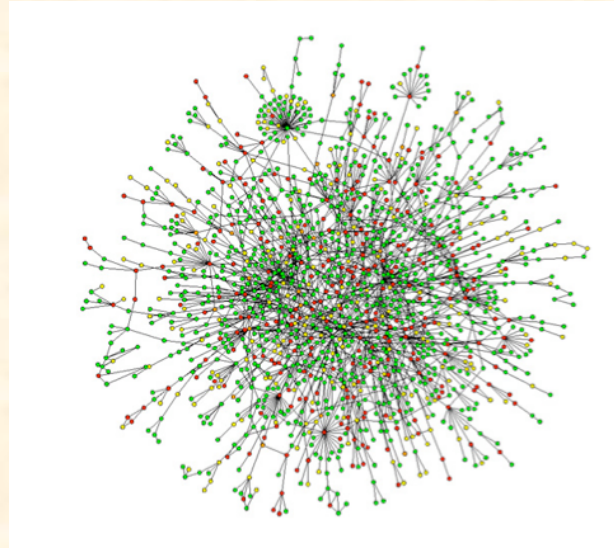
Régions ultraconservées...

481 éléments de plus de 200bp identiques à 100% entre l'homme et la souris

77% en dehors des régions protéiques

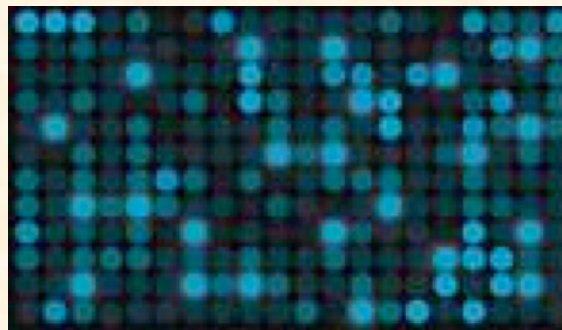
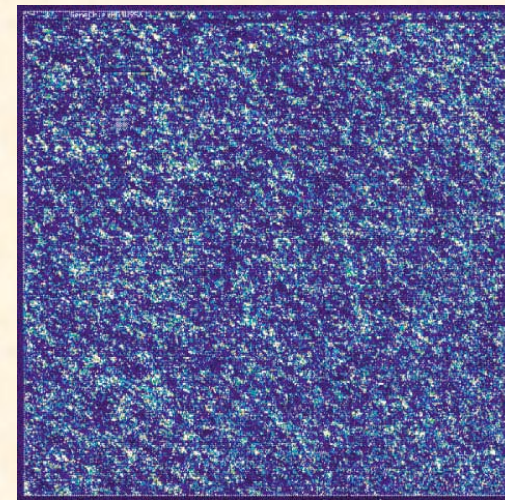
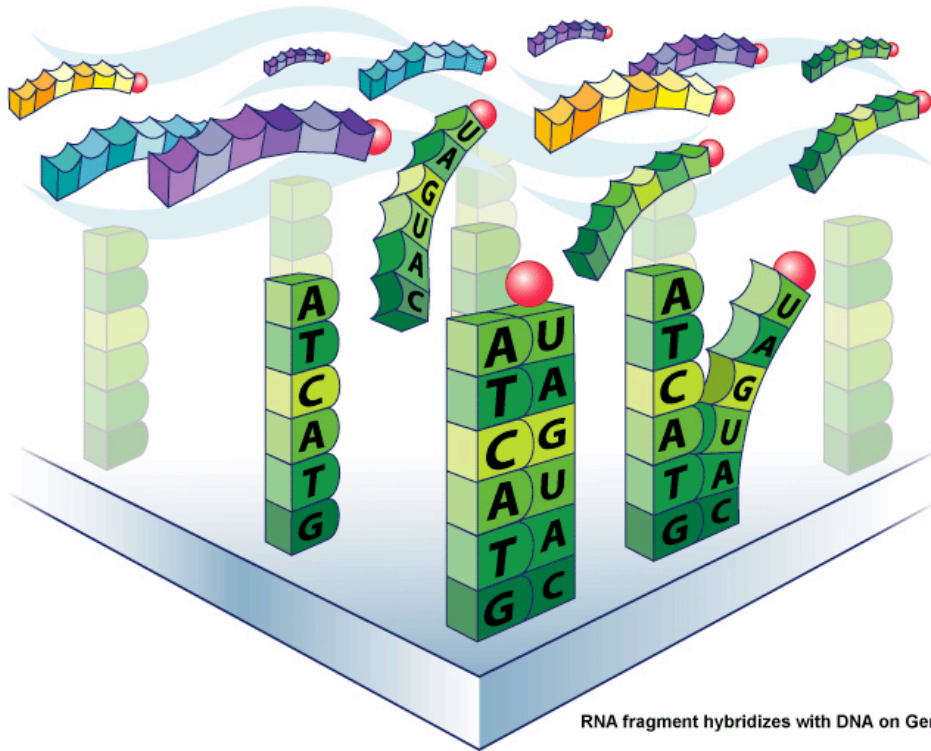
"These ultra-conserved elements are long, they evolved rather rapidly, and they are now evolutionarily frozen. We don't know of a biomolecular mechanism that would explain them," Haussler said.

Nouvelles technologies



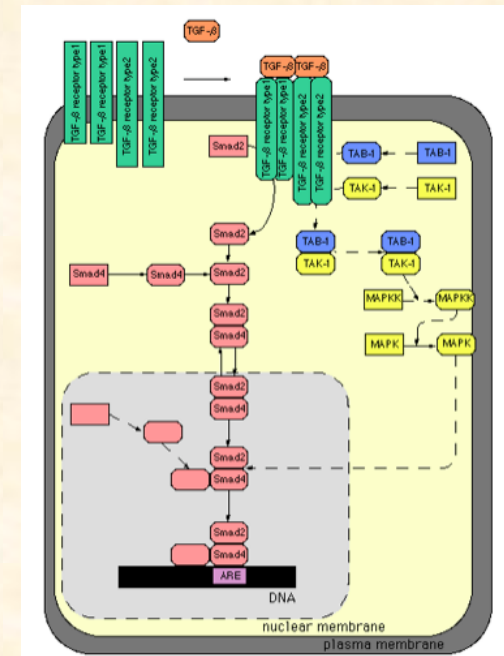
Exemple : puces à ADN

RNA fragments with fluorescent tags from sample to be tested



Nouveaux challenges

- Comprendre les **fonctions** des gènes
- Comprendre les **interactions** entre molécules
- Comprendre le rôle des régions **non protéiques**
- Etudier les **variations** entre individus



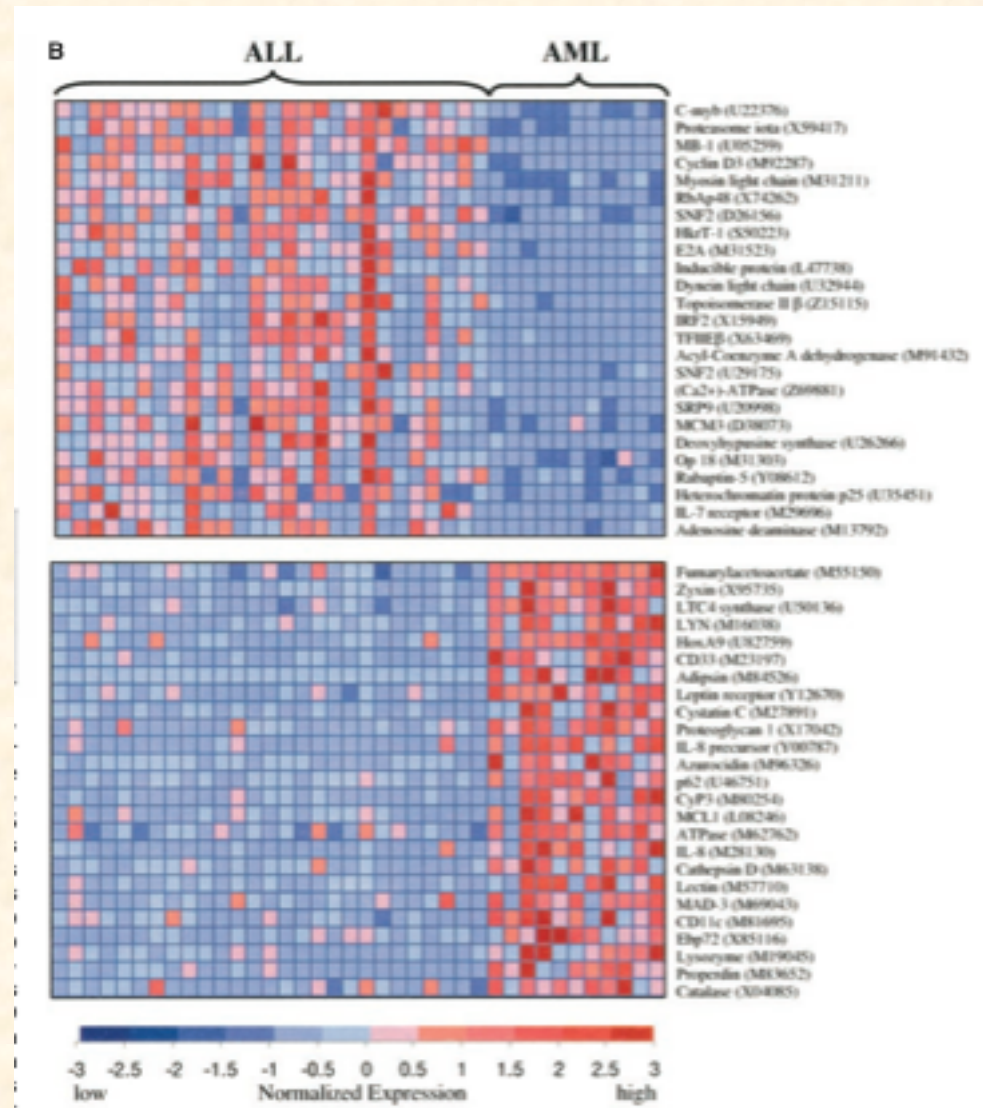
Retombées en biologie

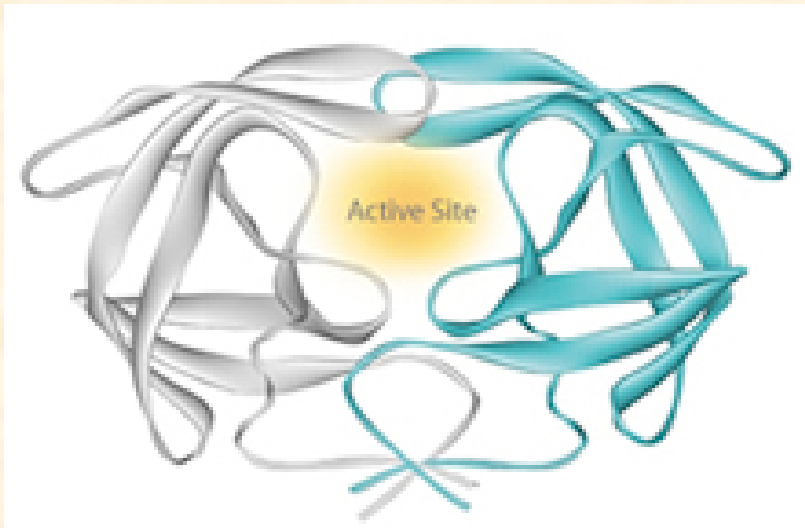
- Structure et fonctions des gènes
- Structure et fonction de l'ADN poubelle
- Interactions
- Phénomènes épigénétiques
- Evolution des espèces
- Variations entre individus

Retombées en médecine

- Identifier les gènes impliqués dans les maladies génétiques
- Gènes de prédisposition aux maladies
- Médecine individualisée (pharmaco-génomique)
- Diagnostic précoce
- Nouvelles cibles thérapeutiques
- Protéines recombinantes, siRNA...
- Thérapie génique
- Vaccin génétique etc...

Ex: classification de tumeurs





1989 : détermination de la structure de la protéase HIV

1996 : mise sur le marché de 3 inhibiteurs

Ex: rational drug design

Retombées en agronomie

- OGM
- Compréhension et optimisation des processus de production

Retombées en énergie/environnement

- Biofuels
- Amélioration des rendements
- Bio-détection de polluants
- Remédiation environnementale
- Séquestration de carbone

Retombées en droit/philosophie/société

- Qu'est-ce que « moi » (ADN, clones, ...)
- Brevetabilité du vivant
- Bioéthique

Retombées économique

- Explosion (bulle?) de l'industrie santé / biotech
- Besoin thérapeutiques croissants (vieillesse de la population, dégradation de l'hygiène de vie)
- Contraintes environnementales croissantes
- CA biotech : \$35Mds en 2002, \$300Mds en 2010...

Pour aller plus loin...

