Biologie Moléculaire L2 (BIOL201)

Cours 2:
Structure de la Chromatine
et Structure des génomes

Open Access

Daniel GAUTHERET - Version 2012.1

Le Génome

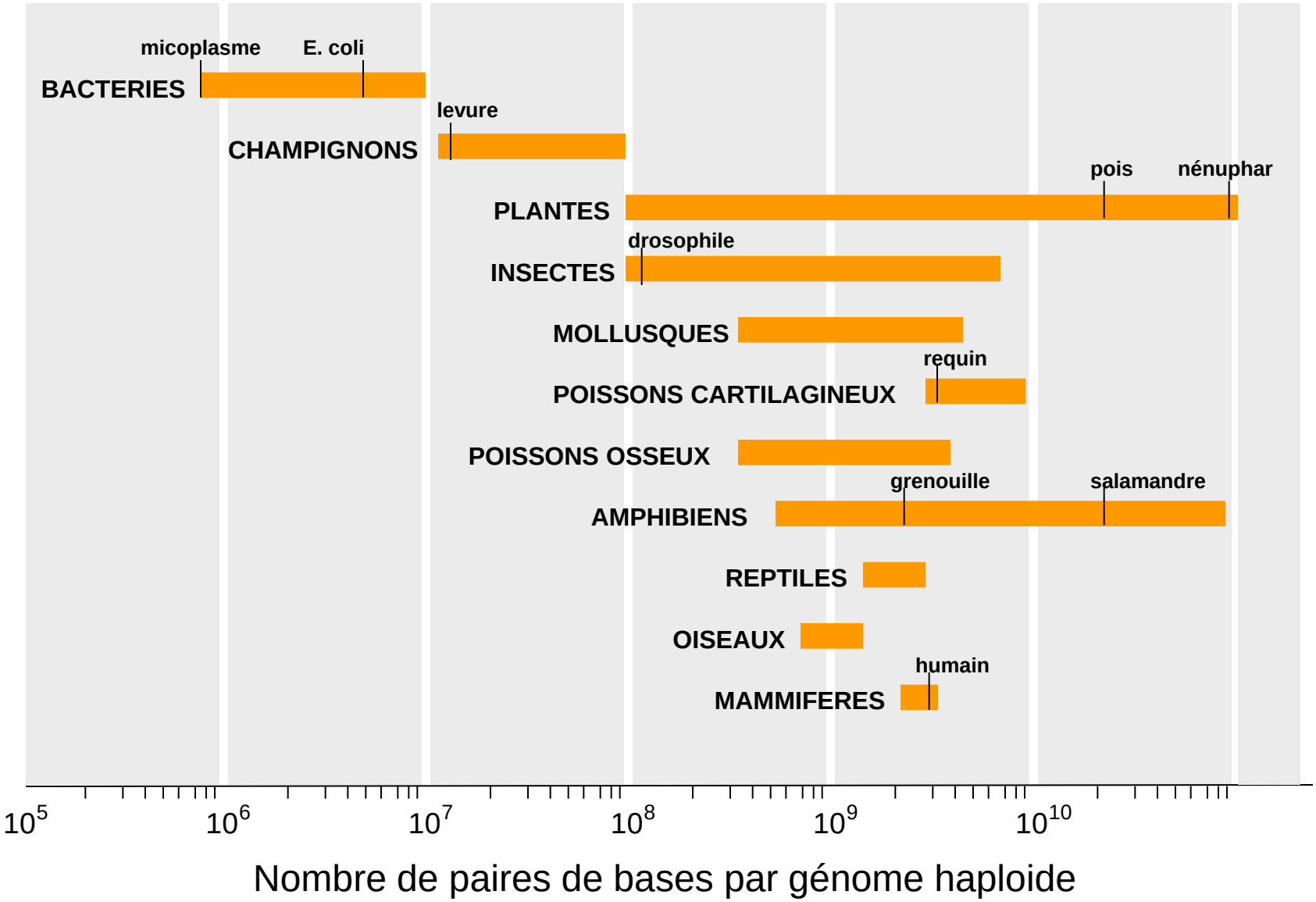
Définition:

Ensemble de l'information héréditaire d'un organisme, présente en totalité dans chaque cellule

Taille des génomes

- Quelques dizaines de milliers de bases pour le génome d'un virus
- Quelques millions de bases pour une bactérie
- 3 milliards de bases pour le génome humain
-16 milliards de bases pour le génome du blé

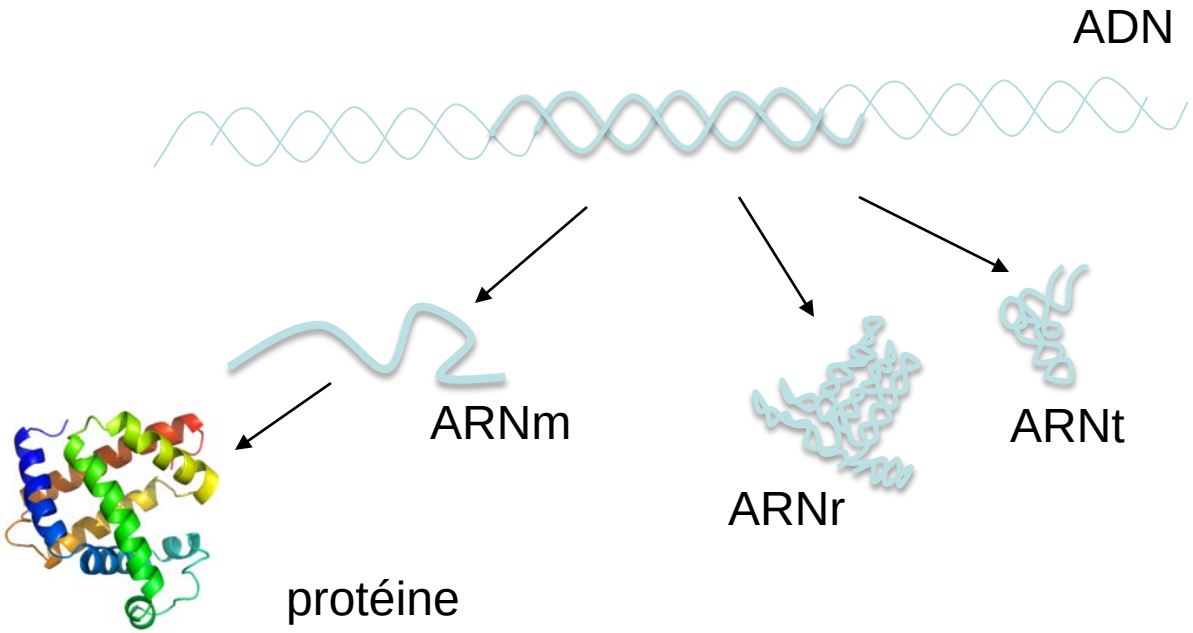
Taille des génomes



Organisme	Bp	Genes	ratio
HIV-1	10 000	10	1000
Haemophilus influenzae	1 830 000	1703	1075
Escherichia coli	4 600 000	4288	1072
Methanococcus jannashchii	1 660 000	1738	955
Amoeba dubia	670 000 000 000	~5000?	134 000 000
Amoeba proteus	270 000 000 000	~5000?	54 000 000
Saccharomyces cerevisiae	13 000 000	5885	2209
Erysiphe cichoracearum (champignon)	1 500 000 000	~10 000?	150 000
Coscinodiscus asteromphalus (diatomée)	25 000 000 000	~5000?	5 000 000
Caenorhabditis elegans	100 000 000	~14 000	7000
Drosophila melanogaster	170 000 000	~12 000	14 000
Arabidopsis thaliana	120 000 000	~10 000	12 000
Lilium formosanum (nénuphar)	36 000 000 000	~15 000?	2 400 000
Zea mays	2 300 000 000	32 000	70 000
Allium cepa (oignon)	18 000 000 000	~20 000?	900 000
Protopterus aethiopicus (dipneuste)	140 000 000 000	~40 000?	3 500 000
Homo sapiens	3 400 000 000	~25 000	100 000

Le gène

- Un gène est une séquence d'acide désoxyribonucléique (ADN) qui spécifie la synthèse d'une chaîne de polypeptide ou d'un ARN fonctionnel (définition Wikipedia)
- Un gène peut donc coder pour un ARN messager ou pour un ARN non-messager (ARNr, ARNt, …)



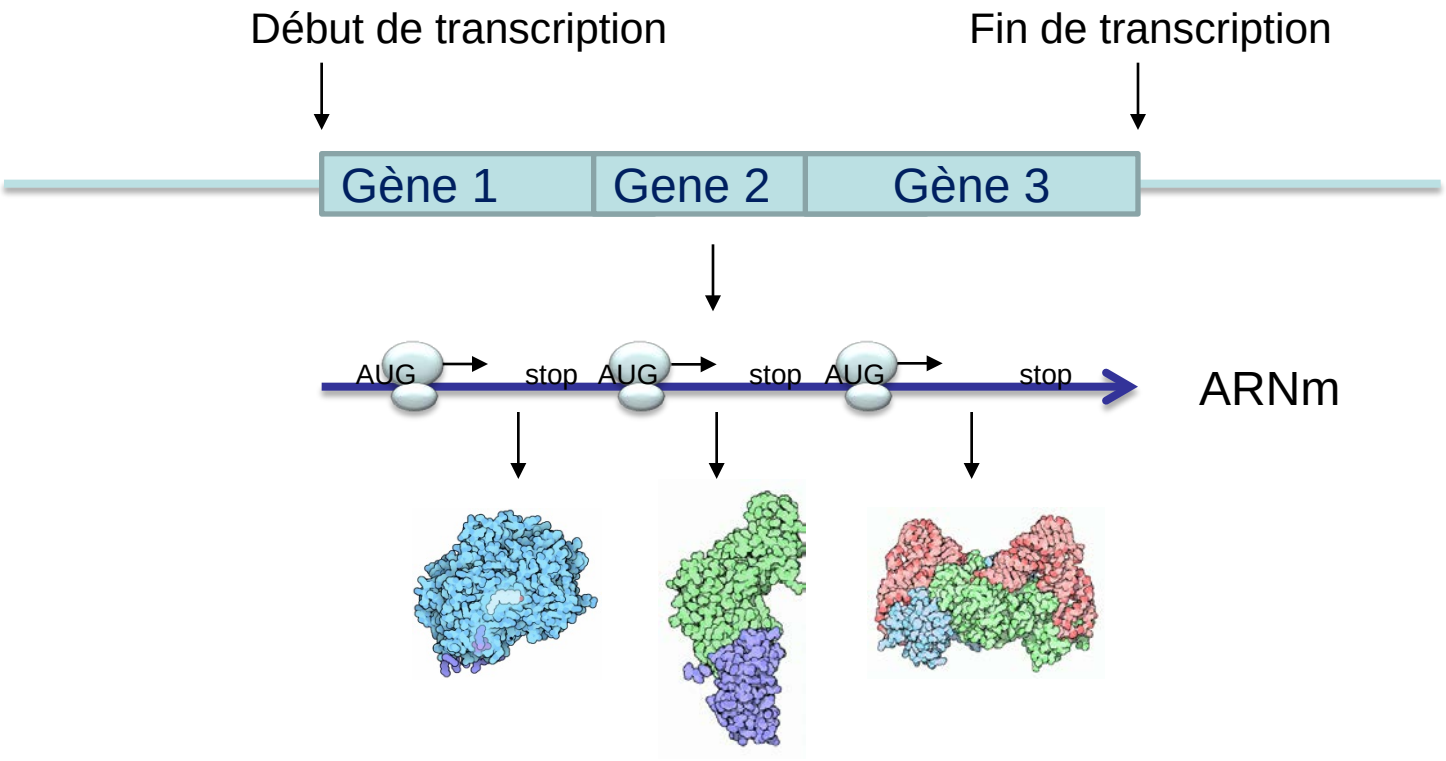
Caractéristiques des génomes <u>procaryotes</u>

- Chromosome circulaire unique
- Une origine de réplication
- Présence possible de petites séquences d'ADN circulaires indépendantes : **les plasmides.**
- Contenu en G+C variable selon les espèces.
    - Ex : 22% chez un parasite *Wigglewortha glossinidia*
    - 67% chez *deinicoccus radiodurans*

Les gènes procaryotes

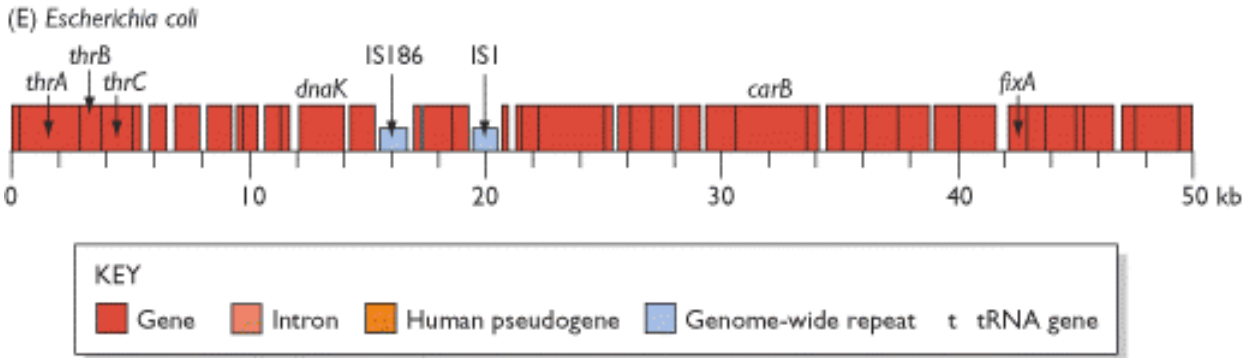
- Fraction codante des génomes élevée.
  - > 90% codant
  - Peu de séquences intergéniques
  - Génome « compact »
- Chez les procaryotes **la séquence des gènes est continue**. **Pas d'intron**
- Gènes organisés en opérons. 600 opérons dans le génome de *Escherichia coli*.

L'opéron


Début de transcription
Fin de transcription
Gène 1
Gene 2
Gène 3
AUG
stop
AUG
stop
AUG
stop
ARNm

Densité des gènes procaryotes

- Longueur gène 950 nt. en moyenne (coli)
- Haute densité en gènes: 95% du génome est transcrit chez E. coli.

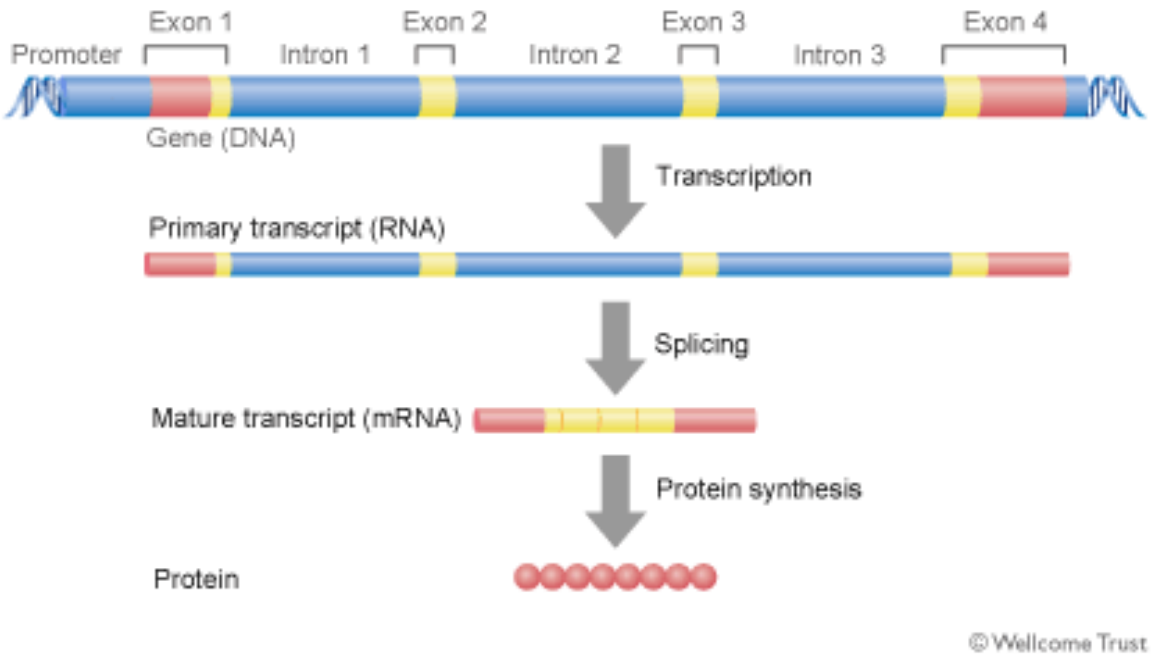


From « Genomes 2 », T.A. Brown

Caractéristiques des génomes <u>eucaryotes</u>

- Dans le noyau
- Taille >> génome procaryote
- Plusieurs chromosomes (homme 23, cheval 32, levure 16, drosophile 4…)
- Plusieurs origines de réplication par chromosome
- Gènes « disloqués » (exons, introns)
- Grandes régions intergéniques de fonction inconnue
- Chromatine!

Le gène de mammifère



transcription

maturation

traduction

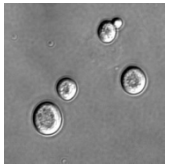
–Gène humain moyen: 27kb, 9 introns, codant:1,3kb , exon moyen: 145 bp, intron moyen:3365 bp.

–Gènes "monstres": dystrophine: 2,4 Mb; Facteur de coagulation VIII: 186 kb, 26 exons; Tinine: codant: 80kb, 178 exons

Densité des gènes eucaryotes

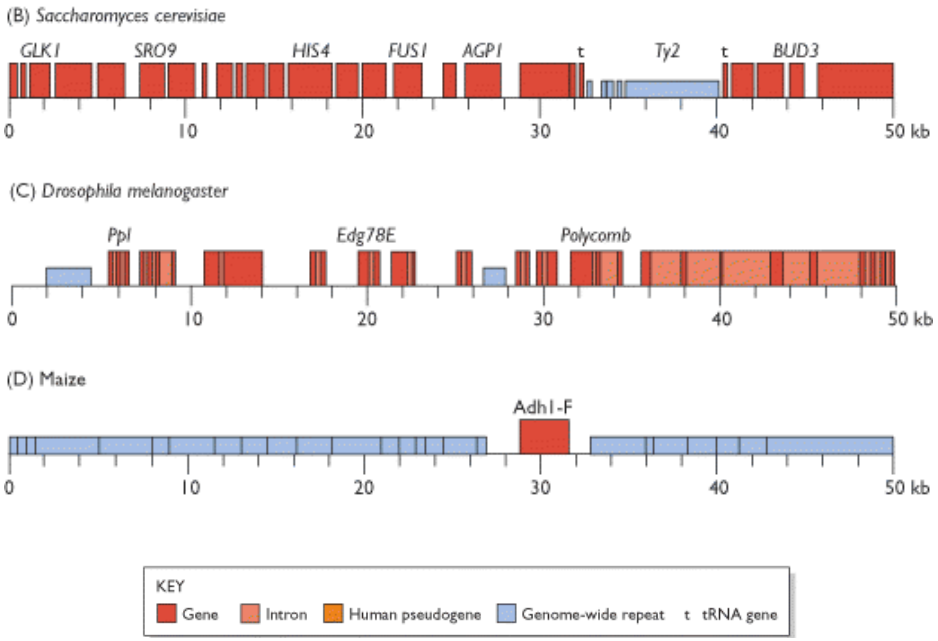
Densité moyenne:

- *S. Cerevisiae*: 1 gène/2kb.



- Drosophile: 1gène/10kb

- Maïs: 1 gène tous les 70kb
- Humain: 1 gène tous les 100kb



From « Genomes 2 », T.A. Brown

Fraction codante et non-codante des génomes

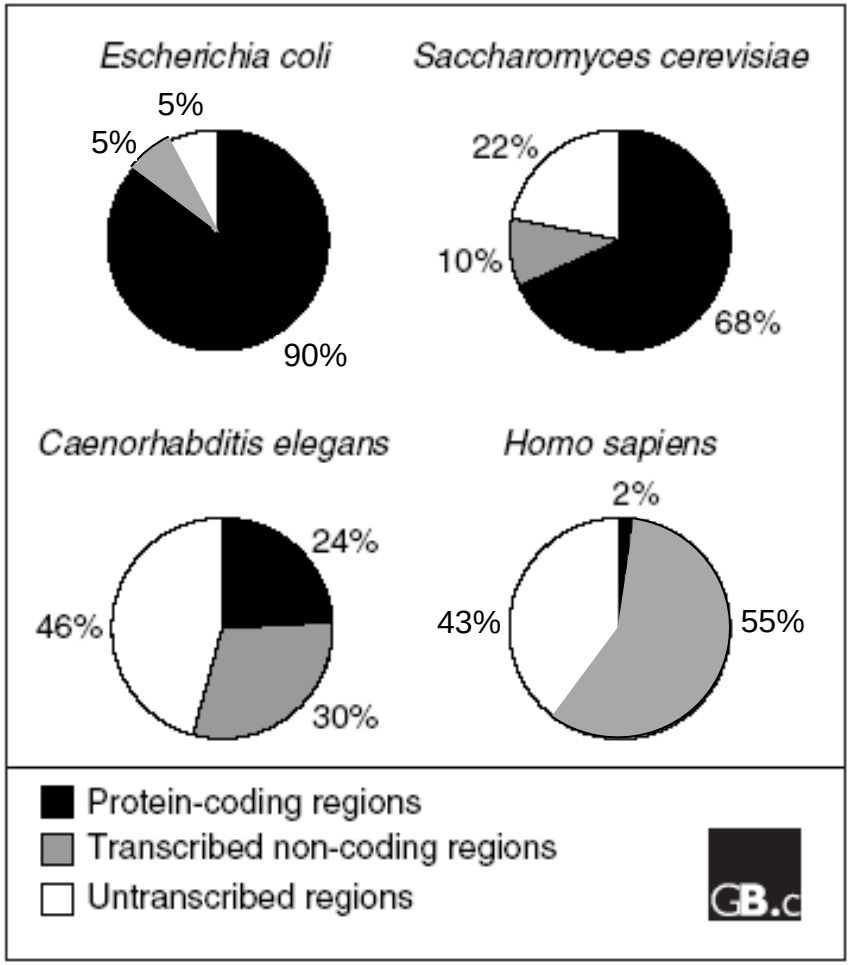
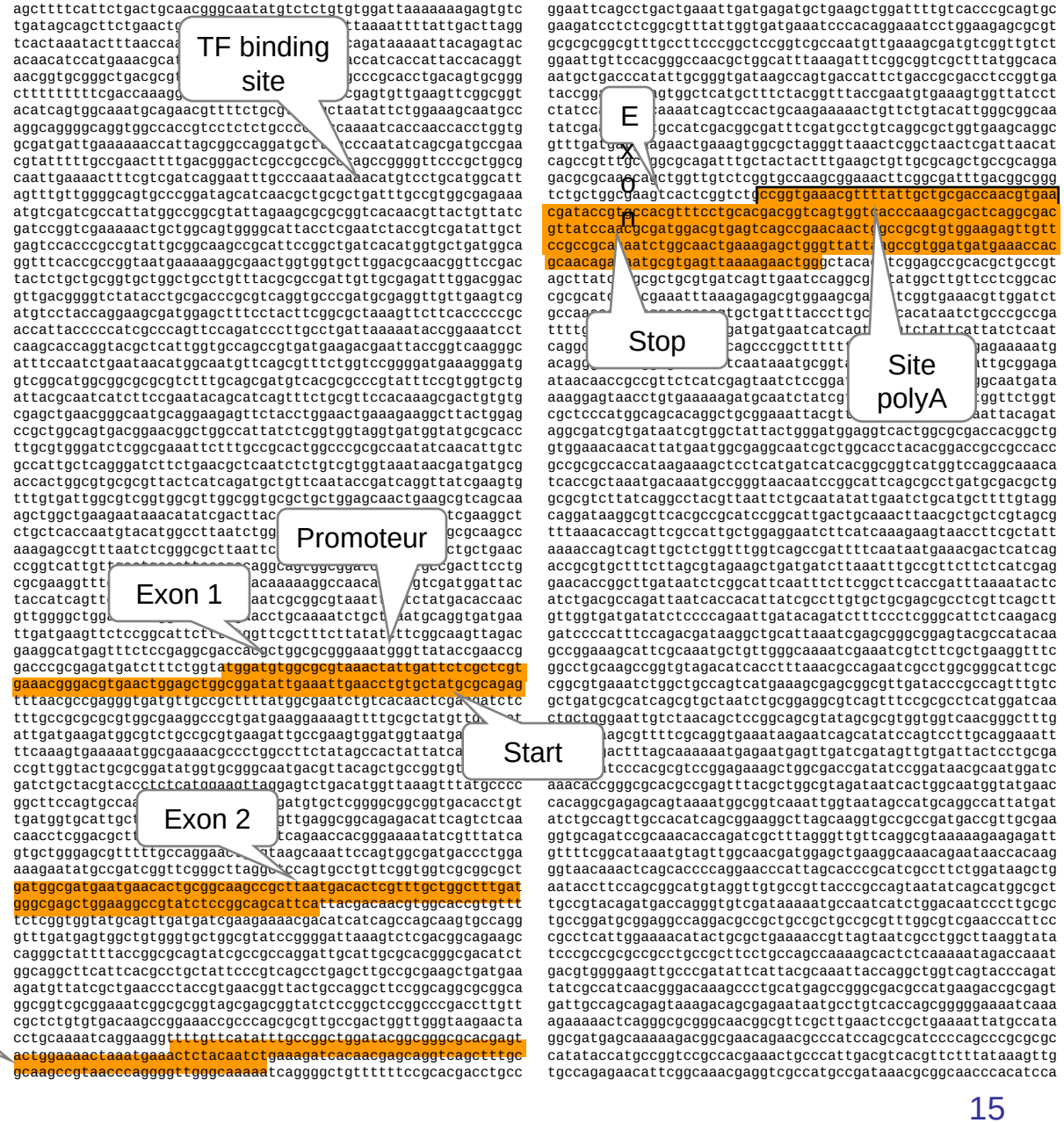


Figure 1
Ratios of the protein-coding, non-coding, and untranscribed sequences in bacterial, yeast, nematode and mammalian genomes. Estimations of the transcribed and protein-coding parts of genomes are based on the sequence length of annotated genes [3,12,13,73]. Estimation of the transcribed portion of the human genome is based on the sequence length occupied by the annotated genes on chromosomes 6, 7, 14, 20, and 22 [5].

Chabalina et al. Genome Biology, 2004

Un gène eucaryote

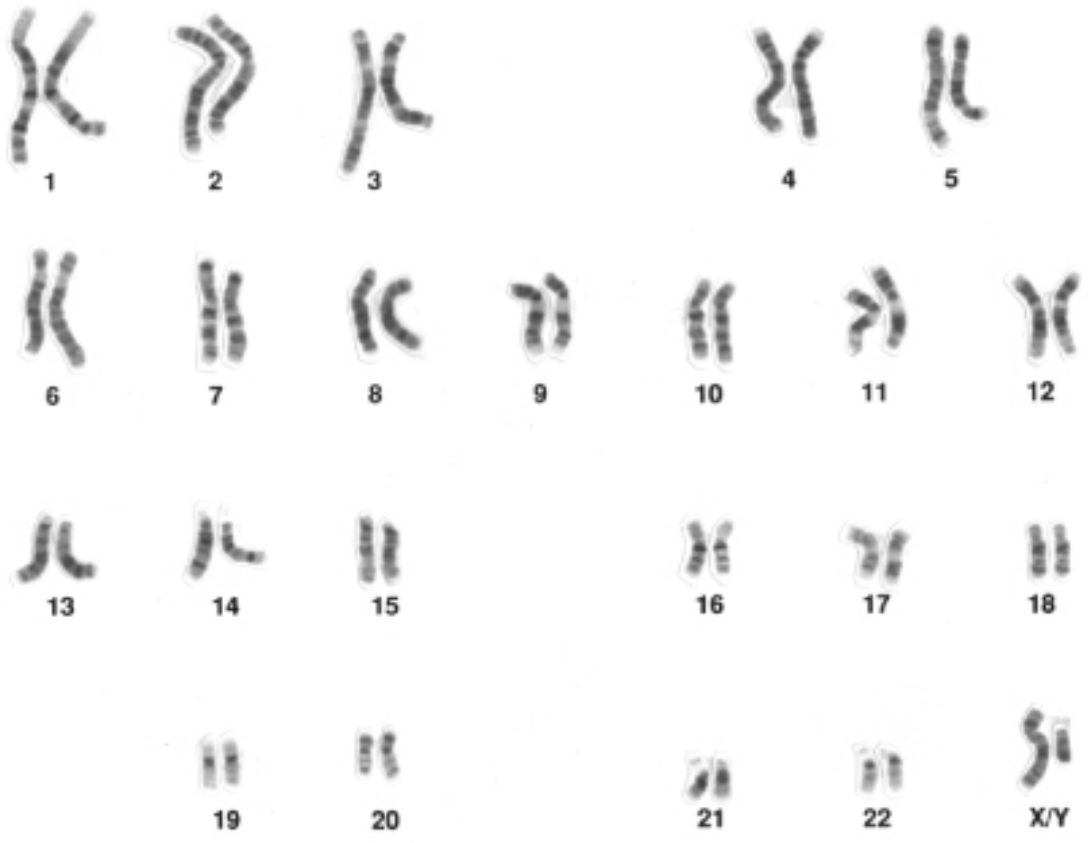


La chromatine: une particularité du génome eucaryote

Ou comment se replie l'ADN à une plus grande échelle que la double hélice

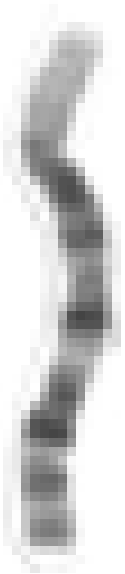
Image: Wikipedia: chromatine

Chromosomes humains



Caryotype

Chromosome 1 humain



Le chromosome contient un double-brin d'ADN de100 millions pb

L'ADN est empaqueté

- Taille du génome humain: 3 Milliard pb soit une taille de 1m si déroulé
- Tient dans une cellule de 6 µm de diamètre.
- Nécessité d'empaquetage de l'ADN

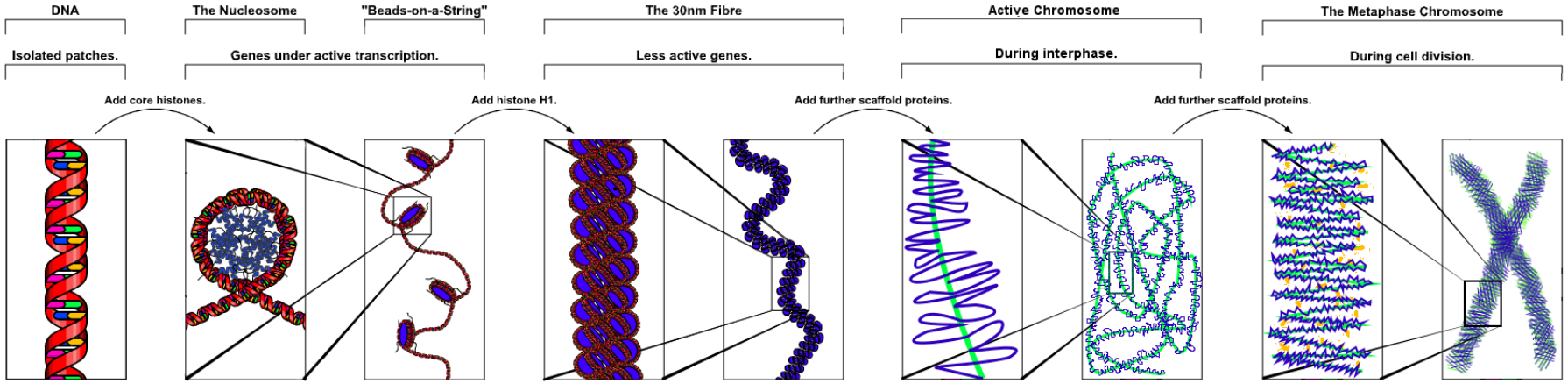


Image: Wikipedia / Richard Wheeler (Zephyris) 2005

D’abord le nucléosome

- 9 protéines de type histone
- Enroulement de l’ADN autour d’un cœur de 8 histones

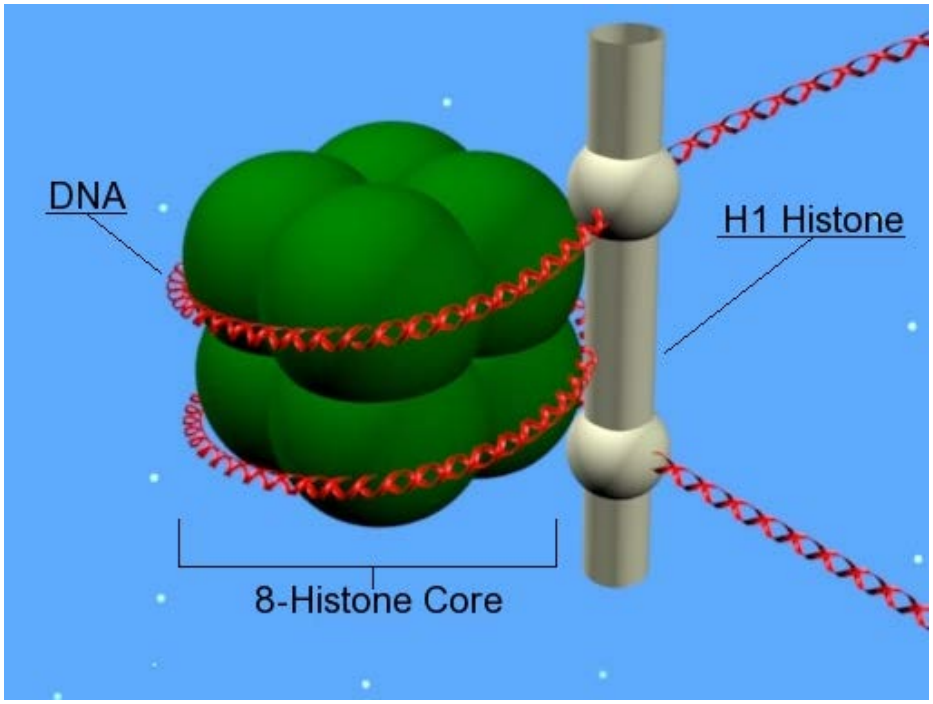
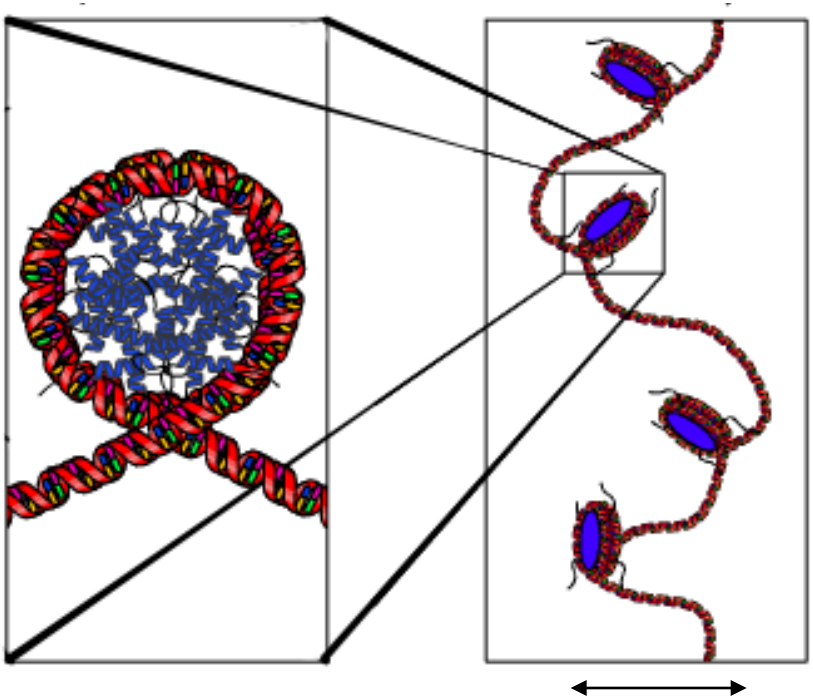


Image: Wikipedia: nucleosome

Filament de 11nm

Sans histone H1. ADN « ouvert ». Expression des gènes possible



11nm

Image: Wikipedia / Richard Wheeler (Zephyris) 2005

Filament de 30 nm

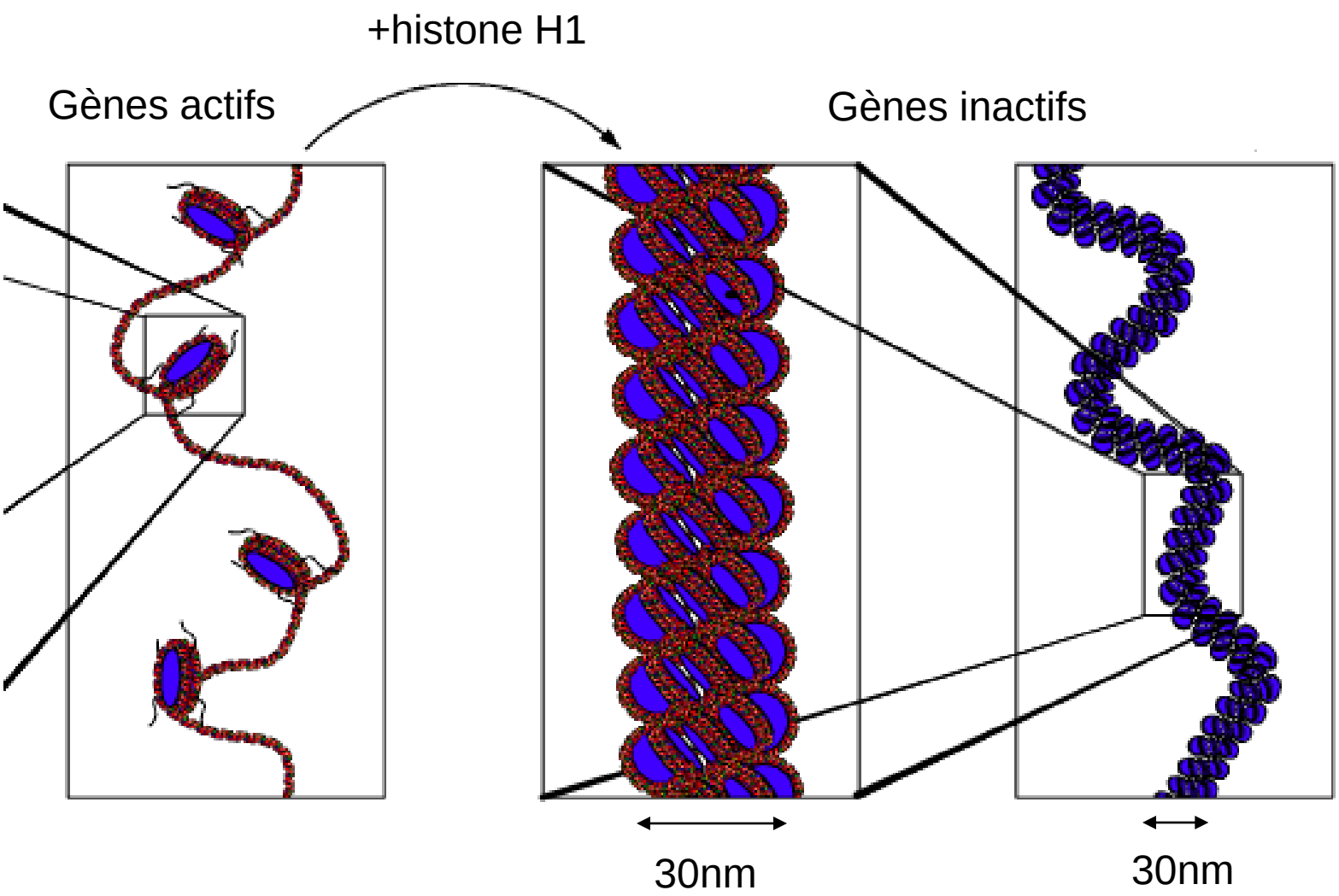
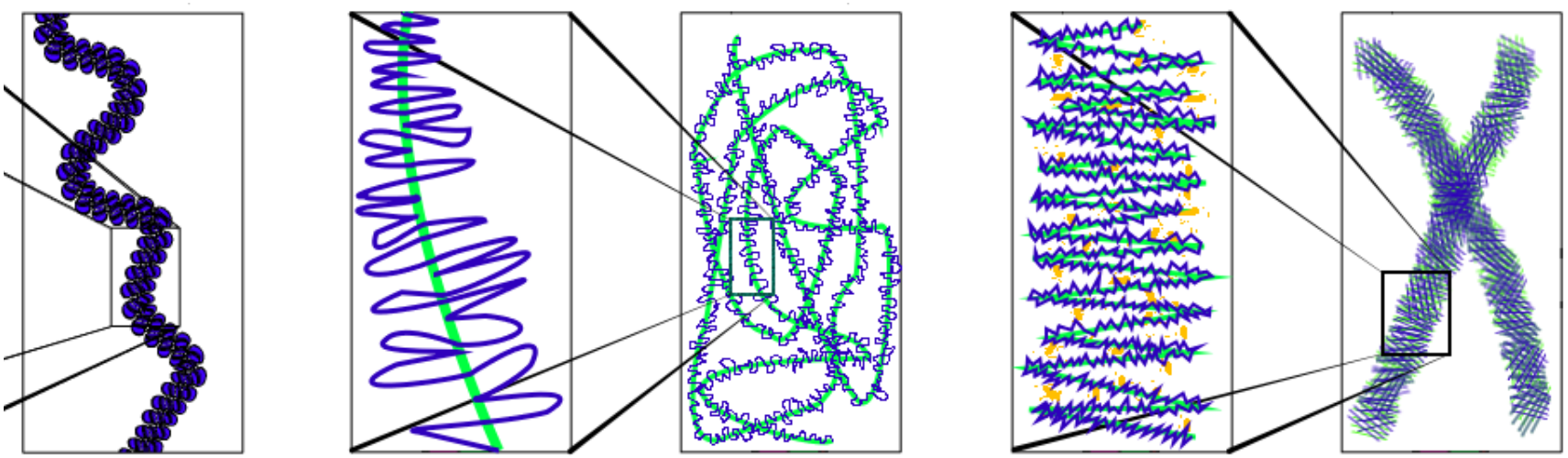


Image: Wikipedia / Richard Wheeler (Zephyris) 2005

Chromosome en interphase et métaphase

Décondensé

Condensé



interphase

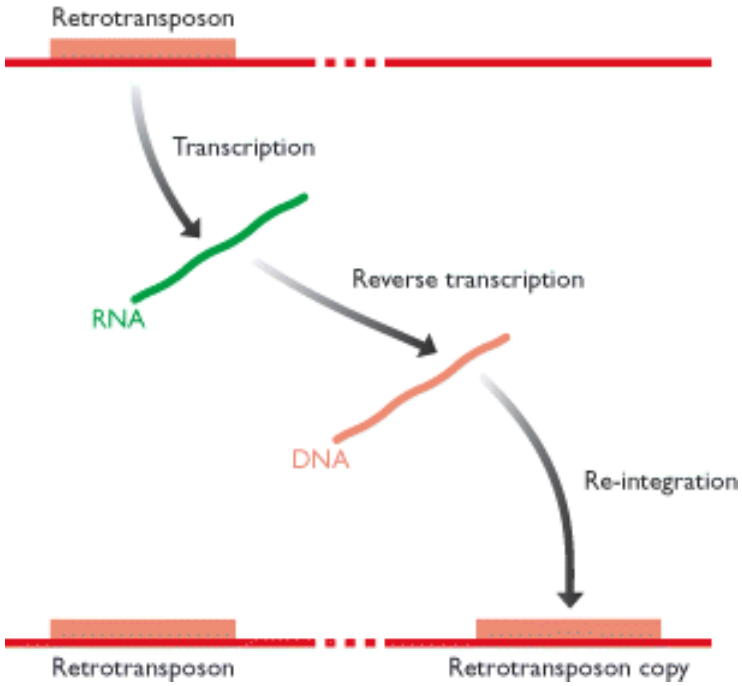
métaphase

Image: Wikipedia / Richard Wheeler (Zephyris) 2005

Junk DNA: l’ADN « poubelle » dans le génome humain

- Un ADN pas si "poubelle" que ça qui joue un grand rôle dans la transformation des gènes et l'apparition de nouveaux gènes.
- Constitué de séquences répétées, principalement de transposons.

Propagation des transposons de type « retroposon »



From « Genomes 2 », T.A. Brown

De nombreux types de transposons

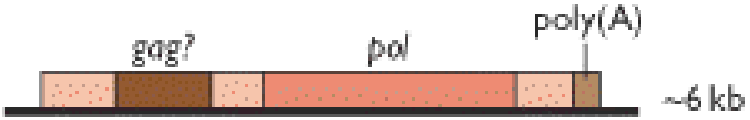
- Les séquences répétées de type transposon représentent plus de 1/3 du génome des vertébrés
- Génome humain: 45% !!

Classes of interspersed repeat in the human genome

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

From « Genomes 2 », T.A. Brown

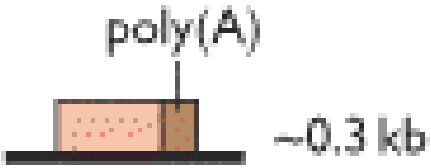
LINES (Long interspersed repeated sequences)



From « Genomes 2 », T.A. Brown

- 21% du génome humain (850.000 copies). Le plus commun de ces éléments se nomme « Line-1 »: 6kb.
- Contiennent un promoteur et 2 gènes
- l'ARN LINE s'assemble avec ses propres protéines et se déplace vers le noyau où l'ARN est reverse-transcrit et s'insère dans le génome au niveau d'une coupure simple brin.
- La machinerie LINE est responsable également de la propagation d’autres éléments

SINEs (Short interspersed repeated sequences)



From « Genomes 2 », T.A. Brown

- Contiennent un promoteur mais pas de gène
- 13% du génome humain. 3 types: Alu, MIR et Ther2/MIR3.
- 1 500 000 SINEs par génome humain
- La plus connue est « ALU », la seule SINE active: 290 bp

Autres séquences répétées

- **Les répétitions simples (Séquences « satellite »)**
  - Répétition parfaite ou non d'un fragment de longueur k.
  - Représentent 3% du génome humain, 0,5% provenant des répétitions de dinucléotides (85% AC ou AT).
- **Segments dupliqués:**
  - blocs de 10–300 kb copiés d'une région à l'autre ou en tandem

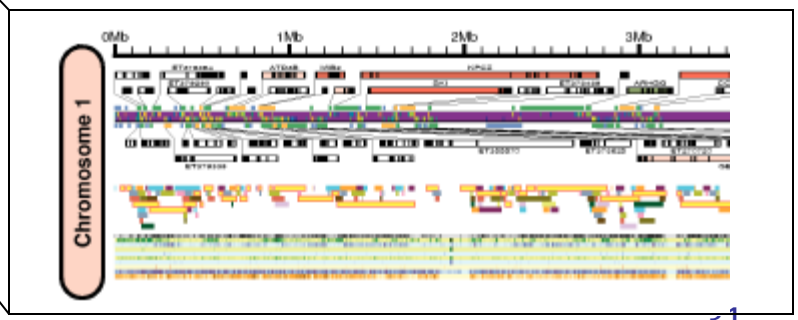
En quoi nos génomes diffèrent-ils?

- Les individus d'espèces différentes ont des génomes différents par la taille, l'ordre et la nature des informations qu'ils contiennent.
- On considère souvent que 2 individus de la même espèce possèdent le « même » génome.
- En fait, le génome de chaque individu est unique. Chez l'homme le génome diffère de 0.5% entre 2 personnes non apparentées

Le génome du Dr. Venter

PloS Biology, 2007

Pour la première fois: deux jeux de chromosomes (maternel et paternel) entièrement séquencés. Différence de 0,5%

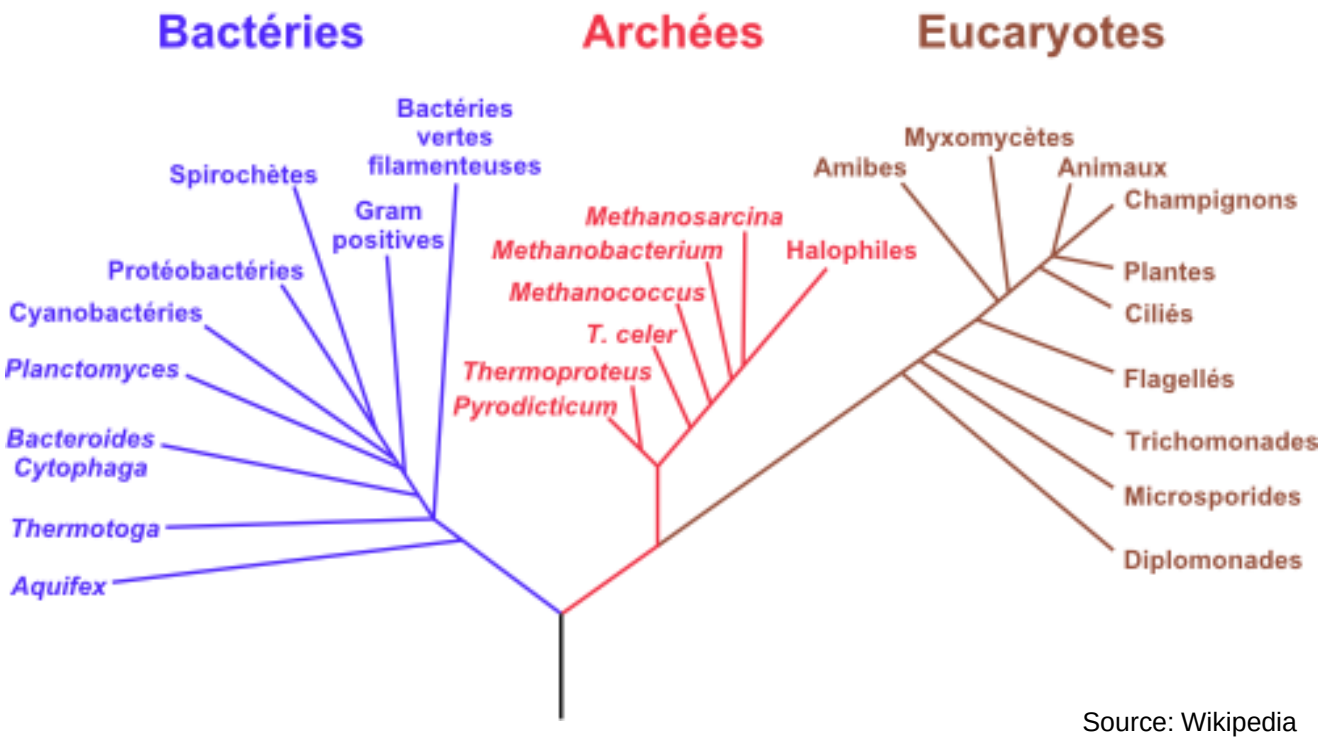


La ressemblance entre génomes

- Homme A / homme B
  - 99,5% identique (0,5% différence)
- Homme/chimpanzé
  - Codant: 98,5% identique
  - Non codant: ~97% identique
  - Quelques duplications/délétions importantes de région de quelques dizaines de kb
- Homme/souris
  - Codant: 90% identique
  - Non codant: la majorité est sans identité apparente, mais on trouve quand même de nombreux segments semblables (« conservés »)
- Homme/poulet
  - Codant: 80% identique
- Homme/poisson
  - Codant: 70% identique

Rappel:
homme et poisson sont très proches!

Arbre phylogénétique de la vie



Source: Wikipedia