

- *Du génome à la cellule*
- *Faire parler les génomes*

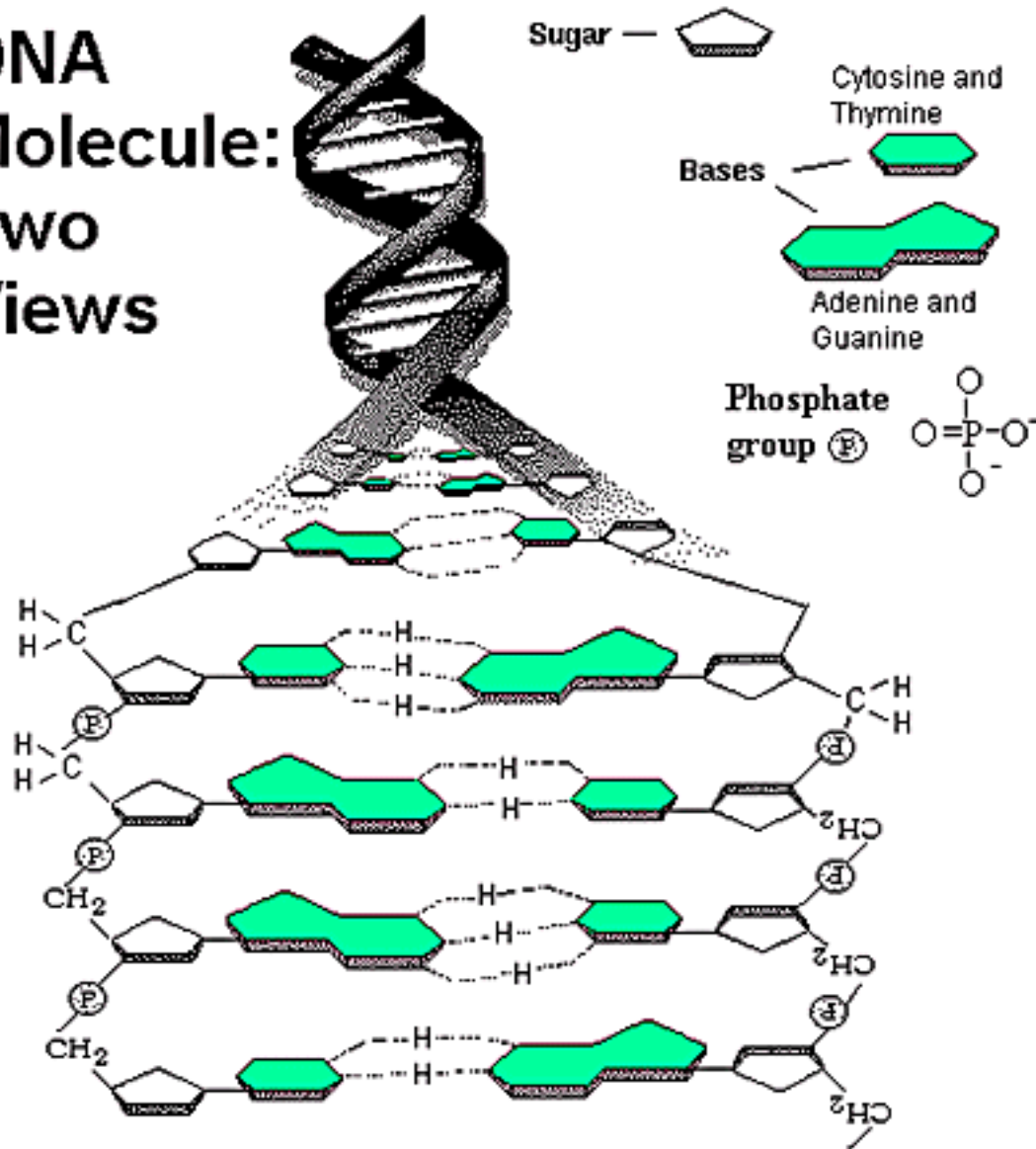
Du génome à la cellule

- *1- L'ADN «la molécule-reine»*
- *2- La bioinformatique conduit à une représentation virtuelle des génomes*
- *3- Les bases de la biologie moléculaire*
- *4- La démarche de la bioinformatique*
- *5- Le temps et le monde vivant*
- *6- Les virus : un monde hétérogène*

1- L'ADN «la molécule-reine»

- i. Structure
- ii. Séquençage
- iii. Etudes bioinformatiques des séquences d'ADN

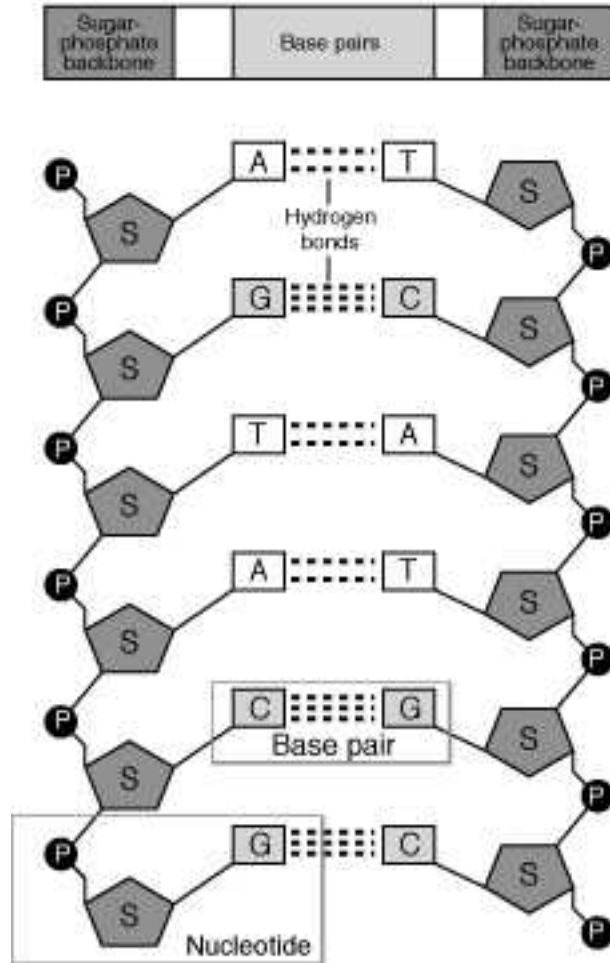
DNA Molecule: Two Views



La structure en double hélice de la molécule d'ADN

James Watson and Francis Crick in 1953

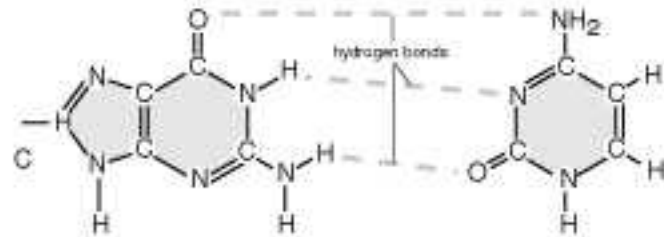
Deoxyribonucleic Acid (DNA)



Nitrogenous Bases

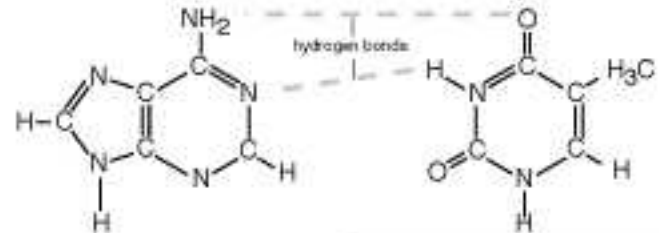
G Guanine

C Cytosine

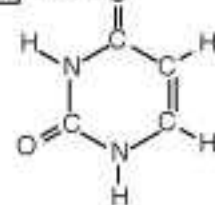


A Adenine

T Thymine



U Uracil



replaces Thymine in RNA

1- L'ADN «la molécule-reine»

- i. Structure
- ii. Séquençage
- iii. Etudes bioinformatiques des séquences d'ADN

Comment séquence-t-on un génome ?

Une réponse rapide : en morceaux !



Lab technician working with sequencing machines
Courtesy of Celera Genomics

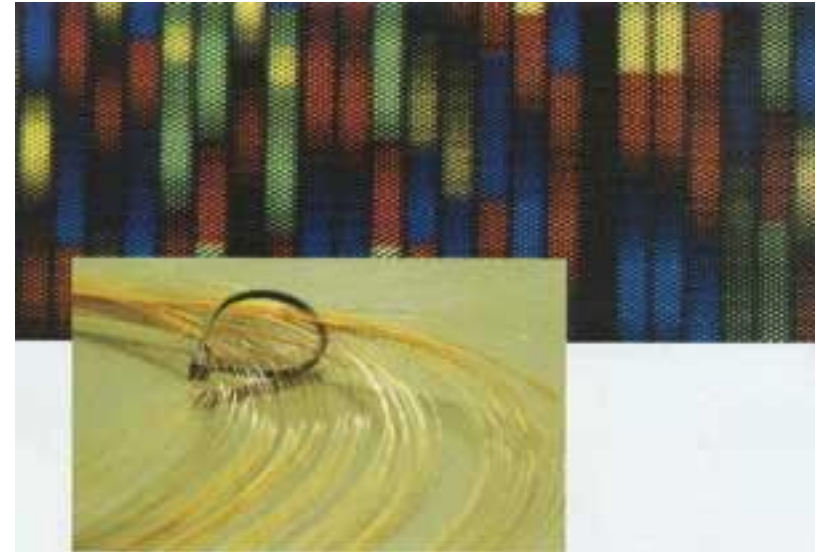


Room filled with sequencing machines
Courtesy of Celera Genomics

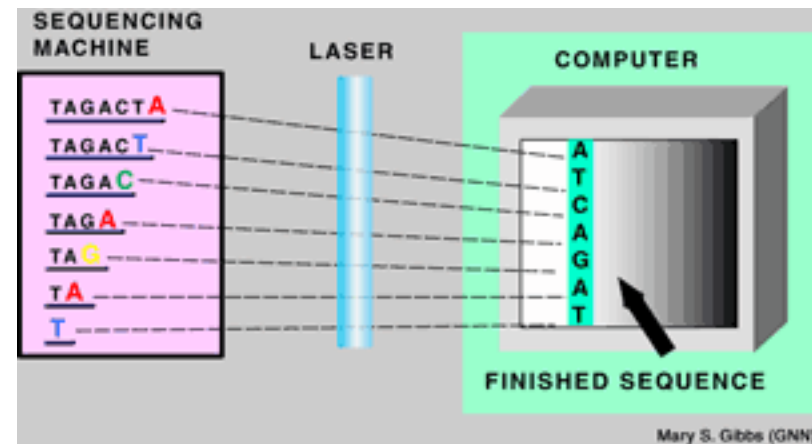
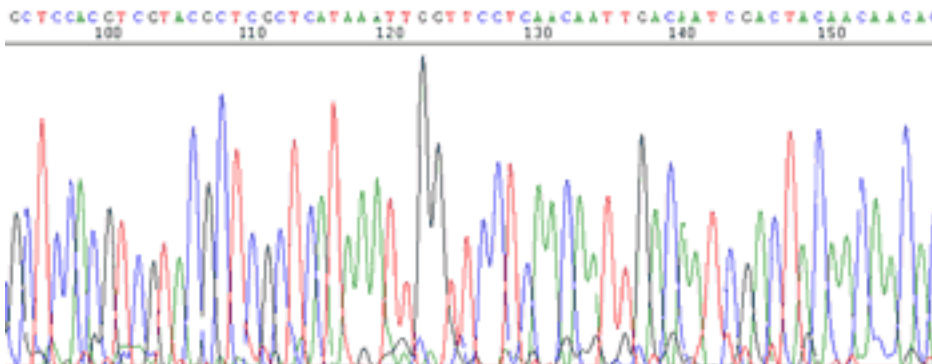
Comment séquence-t-on un génome ?



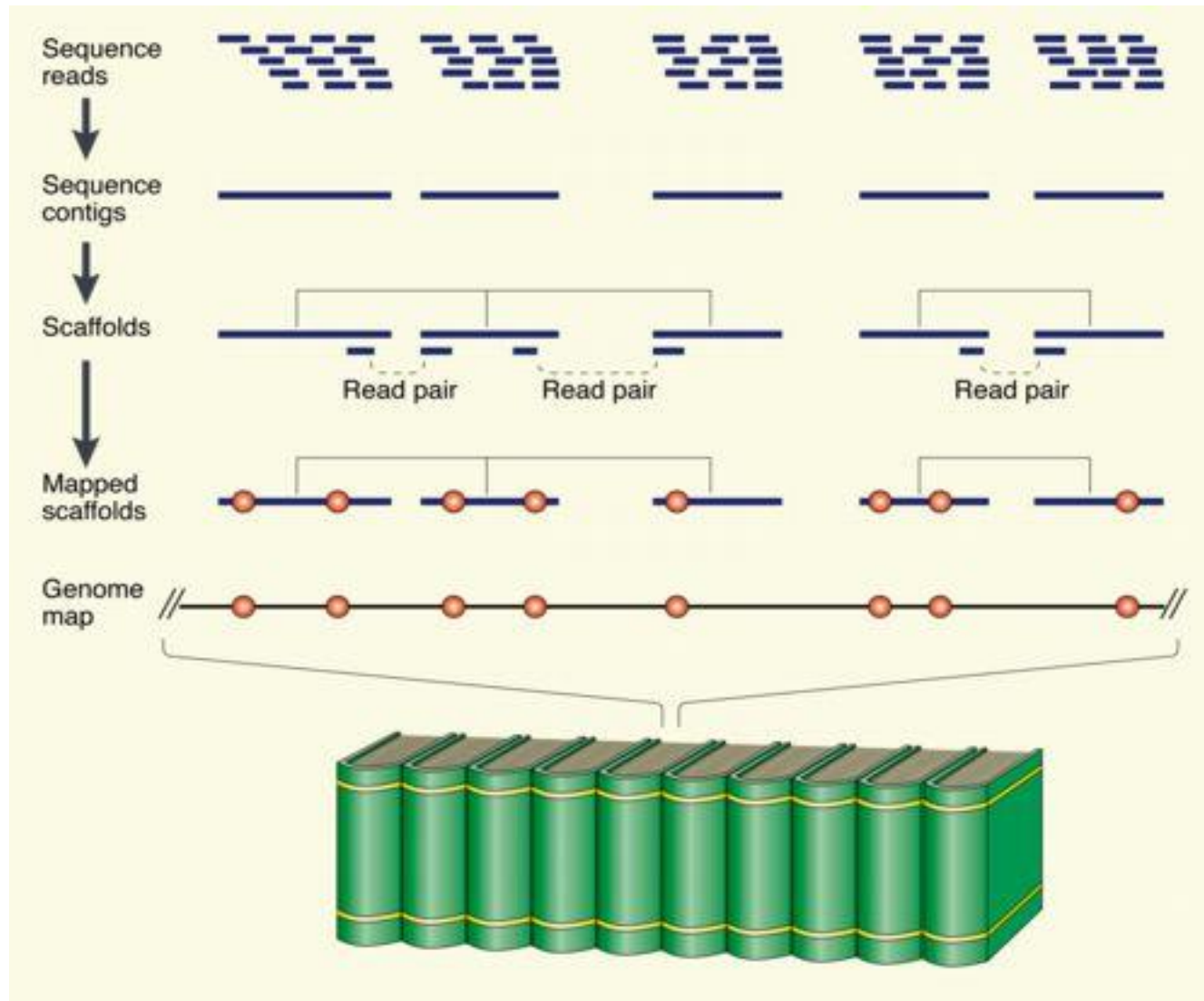
Lab with sequencing machines
Courtesy of Celera genomics



Close up of capillaries from a capillary sequencing machine
Courtesy of Celera Genomics



Comment séquence-t-on un génome ?



1- L'ADN «la molécule-reine»

- i. Structure
- ii. Séquençage
- iii. Etudes bioinformatiques des séquences d'ADN

Longueur des génomes de quelques Eucaryotes

Species	Genome size (Mb)
Fungi	
<i>Saccharomyces cerevisiae</i>	12.1
<i>Aspergillus nidulans</i>	25.4
Protozoa	
<i>Tetrahymena pyriformis</i>	190
Invertebrates	
<i>Caenorhabditis elegans</i>	97
<i>Drosophila melanogaster</i>	180
<i>Bombyx mori</i> (silkworm)	490
<i>Strongylocentrotus purpuratus</i> (sea urchin)	845
<i>Locusta migratoria</i> (locust)	5000
Vertebrates	
<i>Takifugu rubripes</i> (pufferfish)	400
<i>Homo sapiens</i>	3200
<i>Mus musculus</i> (mouse)	3300
Plants	
<i>Arabidopsis thaliana</i> (vetch)	125
<i>Oryza sativa</i> (rice)	430
<i>Zea mays</i> (maize)	2500
<i>Pisum sativum</i> (pea)	4800
<i>Triticum aestivum</i> (wheat)	16 000
<i>Fritillaria assyriaca</i> (fritillary)	120 000

*2- La bioinformatique conduit
à une représentation virtuelle
des génomes*

Fragment d'une séquence d'ADN :

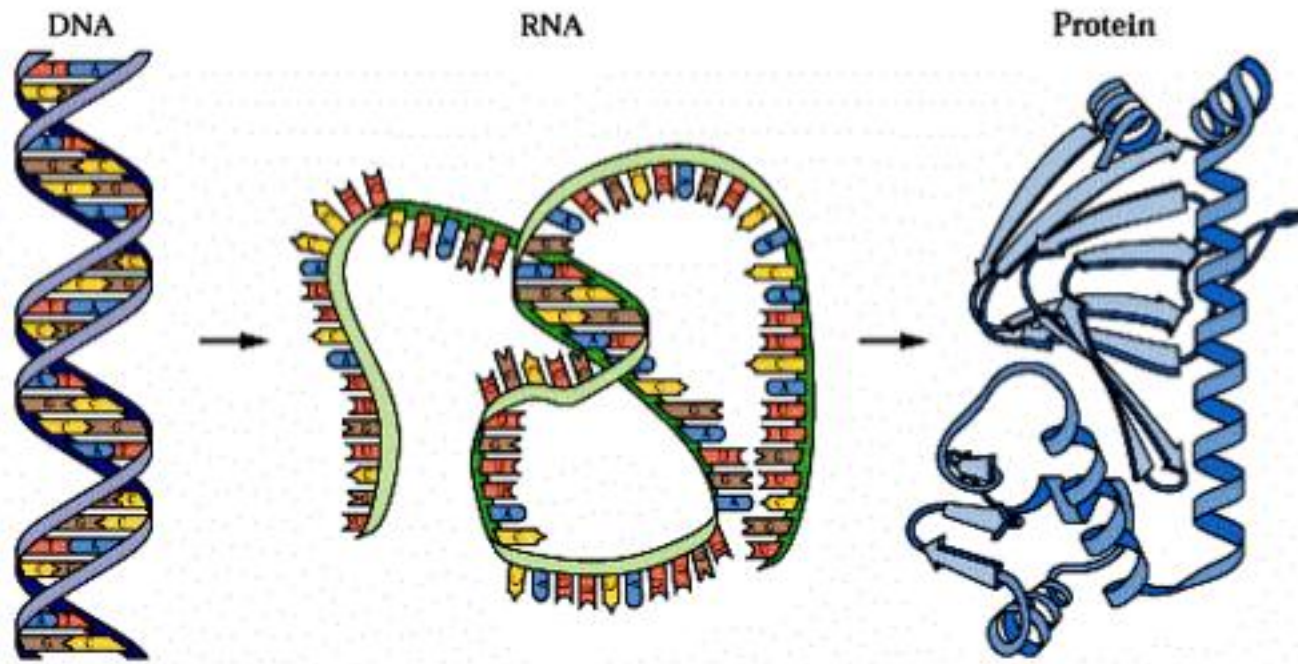
AGTCCGCGGAATACAGGCTCGGT

3- Les bases de la biologie moléculaire

3-1 Le dogme central de la biologie moléculaire

3-1-1 Au niveau des molécules

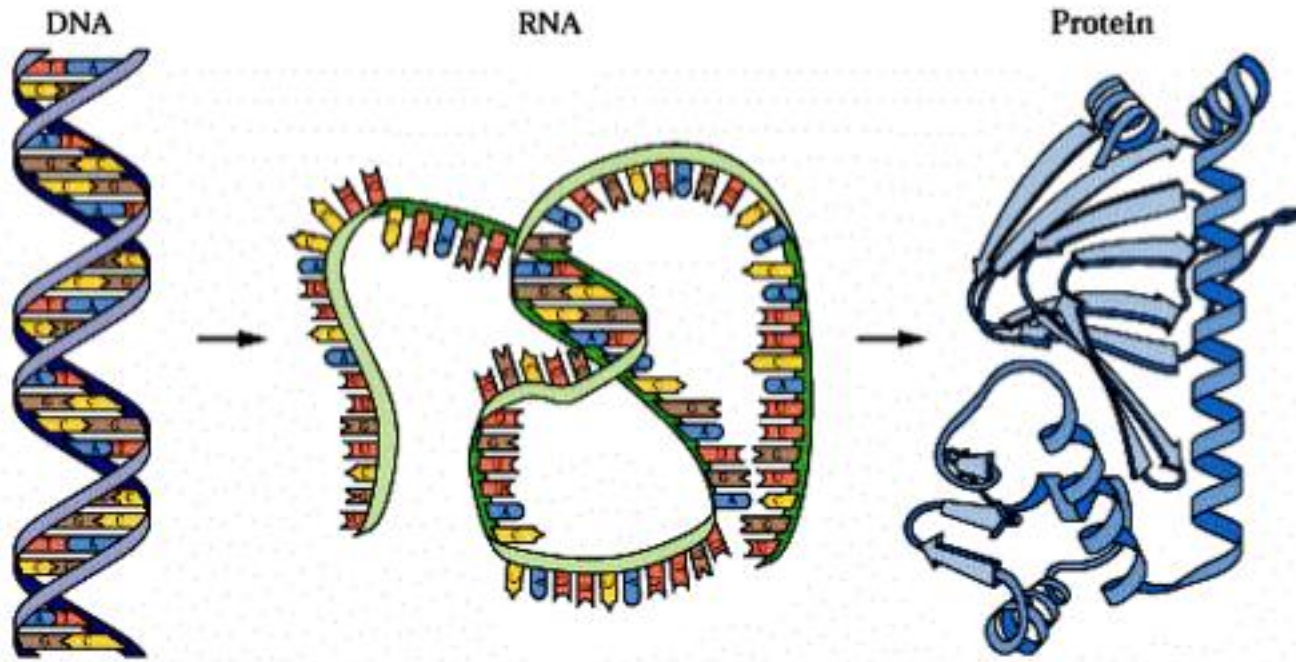
ADN \longrightarrow ARN \longrightarrow Protéine



1960

3-1 Le dogme central de la biologie moléculaire

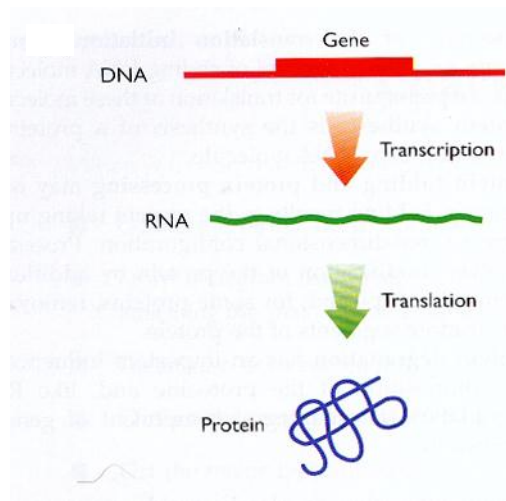
3-1-1 Au niveau des molécules



3-1 Le dogme central de la biologie moléculaire

3-1-2 Au niveau cellulaire

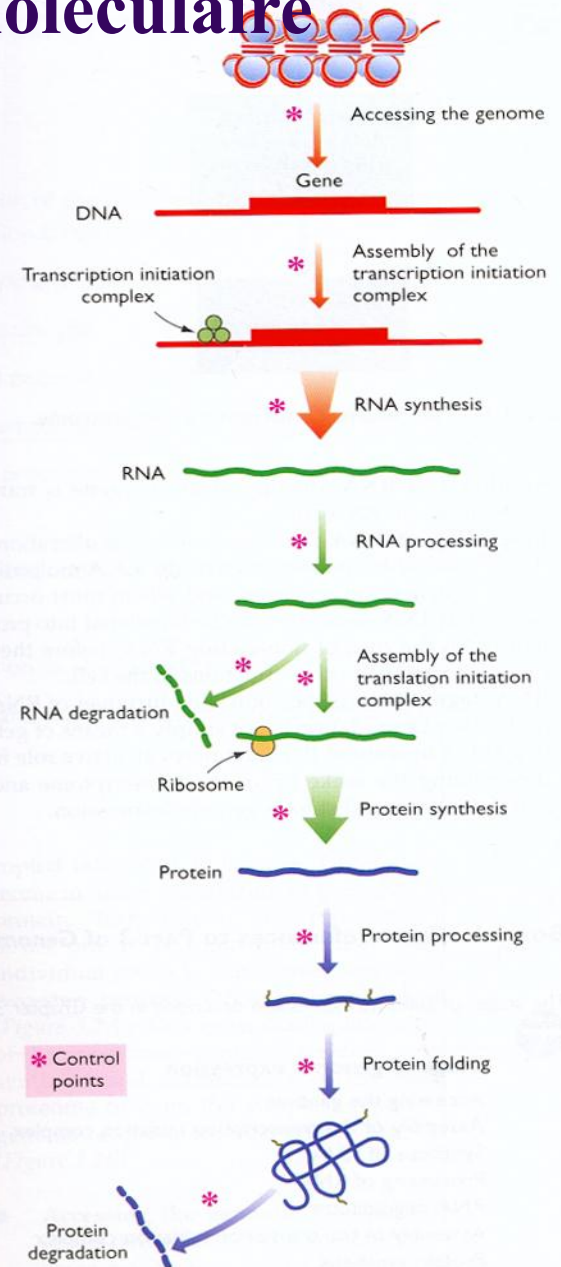
En 1960



En 2000

Transcription

Traduction

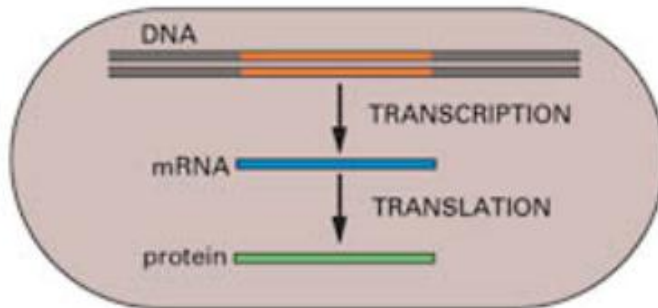


(d'après Brown, 2002, in *genomes*)

3-1 Le dogme central de la biologie moléculaire

3-1-3 Au niveau des organismes procaryotes et eucaryotes

PROCARYOTES



EUCARYOTES

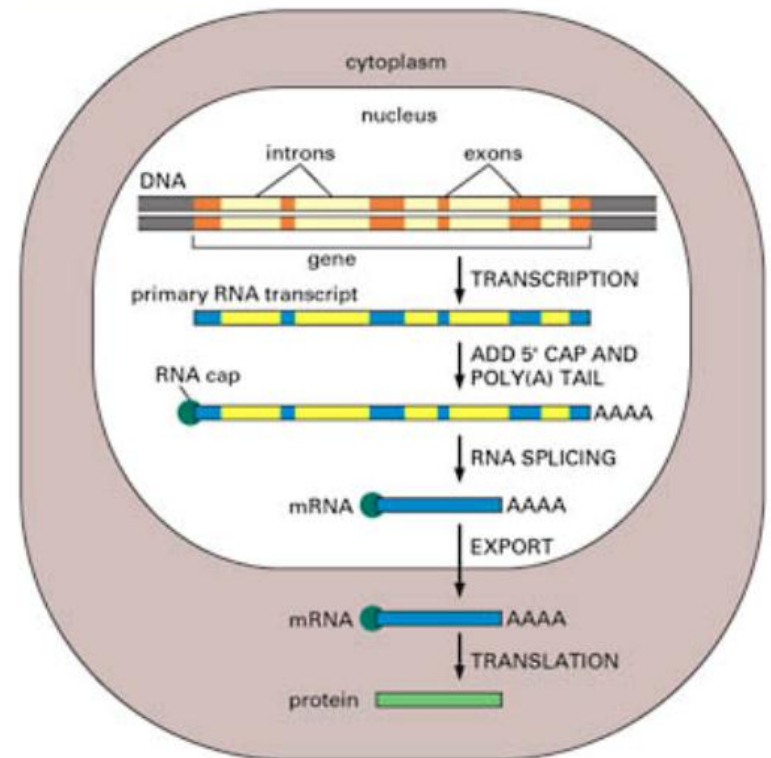


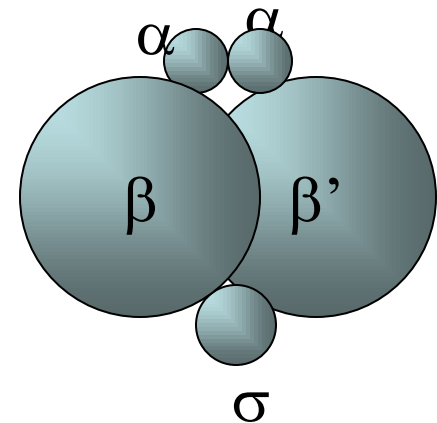
Schéma simplifié

3-2 La machinerie transcriptionnelle

3-2-1 chez les bactéries

Une seule ARN polymérase chez *E.coli* :

- **Découverte en 1958**
- **Contient 5 sous-unités**
- **Environ 7000 molécules par cellule**



3-2 La machinerie transcriptionnelle

3-2-2 chez les eucaryotes

Trois ARN polymérases

ARN polymérase	Synthèse	Constituants
I	- ARN pré-ribosomique	Ensemble de 12 sous-unités plus 2 autres polypeptides
II	- ARN messagers - snoRNA*	Ensemble de 12 sous-unités
III	- Les ARN de transfert - ARN ribosomique 5S - Petits ARN	Ensemble de 12 sous-unités plus 5 autres polypeptides

* : small nucleolar RNA

3-3 La machinerie traductionnelle

3-3-1 Le ribosome est constitué de deux sous-unités

Table 6.1 The compositions of typical prokaryotic and eukaryotic ribosomes. With eukaryotes the lengths of the large rRNAs and the numbers of proteins are not exactly the same in all species; the figures given here refer specifically to mammalian ribosomes

	<i>Prokaryotes</i>	<i>Eukaryotes</i>
Complete ribosome		
Sedimentation coefficient	70S	80S
Molecular mass (daltons)	2 520 000	4 220 000
Number of subunits	2	2
Large subunit		
Sedimentation coefficient	50S	60S
Molecular mass (daltons)	1 590 000	2 820 000
RNA molecules		
number	2	3
sizes	23S = 2904 nucleotides 5S = 120 nucleotides	28S = 4718 nucleotides 5.8S = 160 nucleotides 5S = 120 nucleotides
Number of polypeptides	34	50
Small subunit		
Sedimentation coefficient	30S	40S
Molecular mass (daltons)	930 000	1 400 000
RNA molecules		
number	1	1
sizes	16S = 1541 nucleotides	18S = 1874 nucleotides
Number of polypeptides	21	33

Utilisation en phylogénie moléculaire

Table 6.1 The compositions of typical prokaryotic and eukaryotic ribosomes. With eukaryotes the lengths of the large rRNAs and the numbers of proteins are not exactly the same in all species; the figures given here refer specifically to mammalian ribosomes

	<i>Prokaryotes</i>	<i>Eukaryotes</i>
Complete ribosome		
Sedimentation coefficient	70S	80S
Molecular mass (daltons)	2 520 000	4 220 000
Number of subunits	2	2
Large subunit		
Sedimentation coefficient	50S	60S
Molecular mass (daltons)	1 590 000	2 820 000
RNA molecules		
number	2	3
sizes	23S = 2904 nucleotides 5S = 120 nucleotides	28S = 4718 nucleotides 5.8S = 160 nucleotides 5S = 120 nucleotides
Number of polypeptides	34	50
Small subunit		
Sedimentation coefficient	30S	40S
Molecular mass (daltons)	930 000	1 400 000
RNA molecules		
number	1	1
sizes	16S = 1541 nucleotides	18S = 1874 nucleotides
Number of polypeptides	21	33

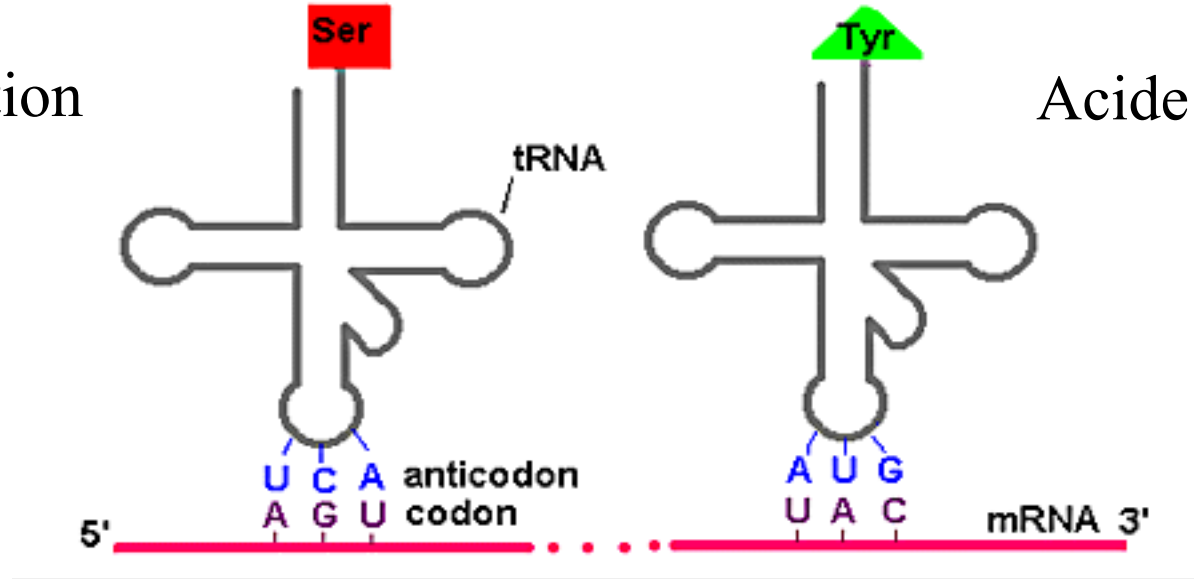
Variations in rRNA sizes in different eukaryotes

<i>Organism</i>	<i>Large rRNA</i>	<i>Small rRNA</i>
Mammals	4718 nucleotides	1874 nucleotides
Amphibia	4110 nucleotides	1825 nucleotides
Yeast	3392 nucleotides	1799 nucleotides

3-3-2 Le code génétique

La traduction

Acide aminé



		2nd base in codon					
		U	C	A	G		
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G	3rd base in codon
	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	

The Genetic Code

(Crick, 1960)

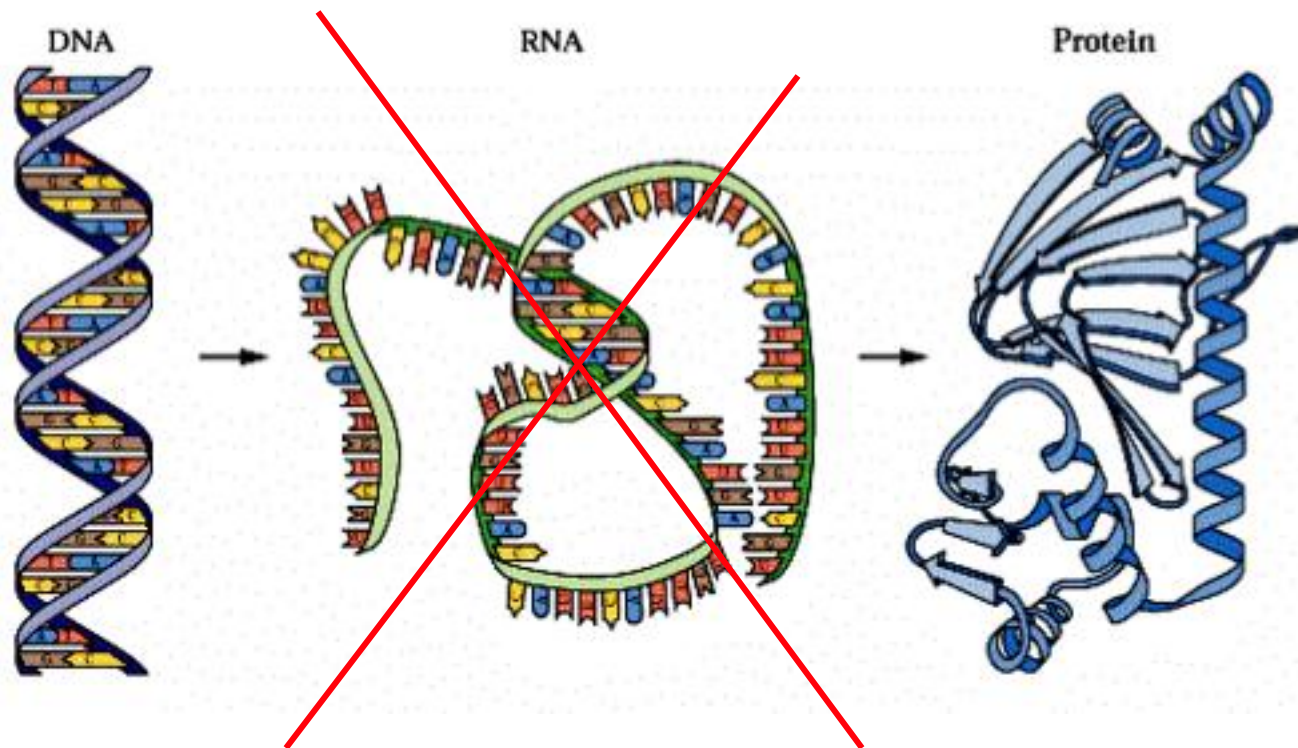
Le code génétique à une lettre est utilisé en bioinformatique

Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

4- La démarche de la bioinformatique

4- La démarche de la bioinformatique

ADN $\xrightarrow{\text{Dédution}}$ Protéine

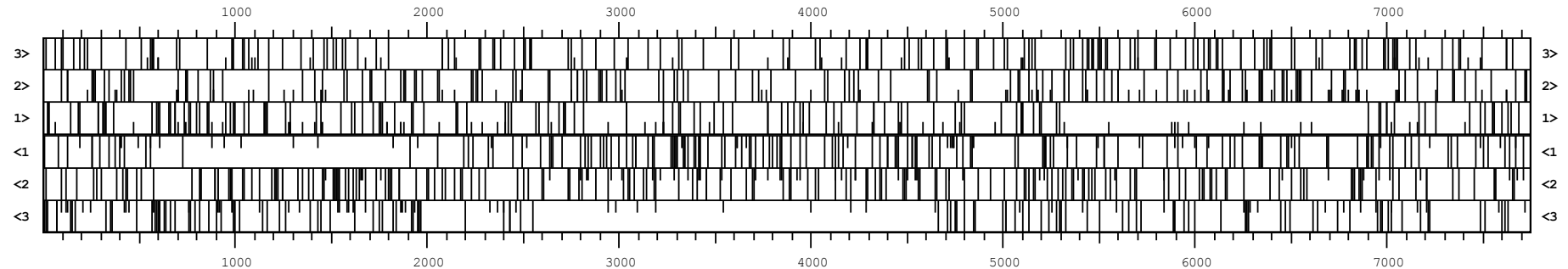


4- *La démarche de la bioinformatique*



Recherche des phases de lecture ouverte
dans une séquence d'ADN procaryote (ou eucaryote sans intron)

logiciel DNA strider (C. Marck)



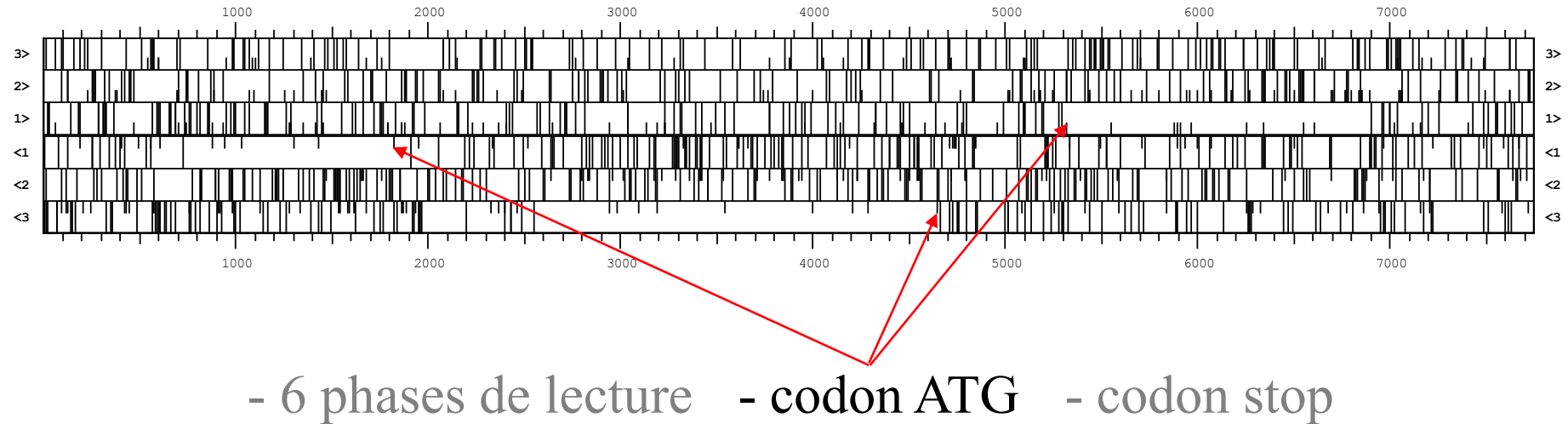
- 6 phases de lecture - codon ATG - codon stop

4- La démarche de la bioinformatique



Recherche des phases de lecture ouverte
dans une séquence d'ADN procaryote (ou eucaryote sans intron)

logiciel DNA strider (C. Marck)

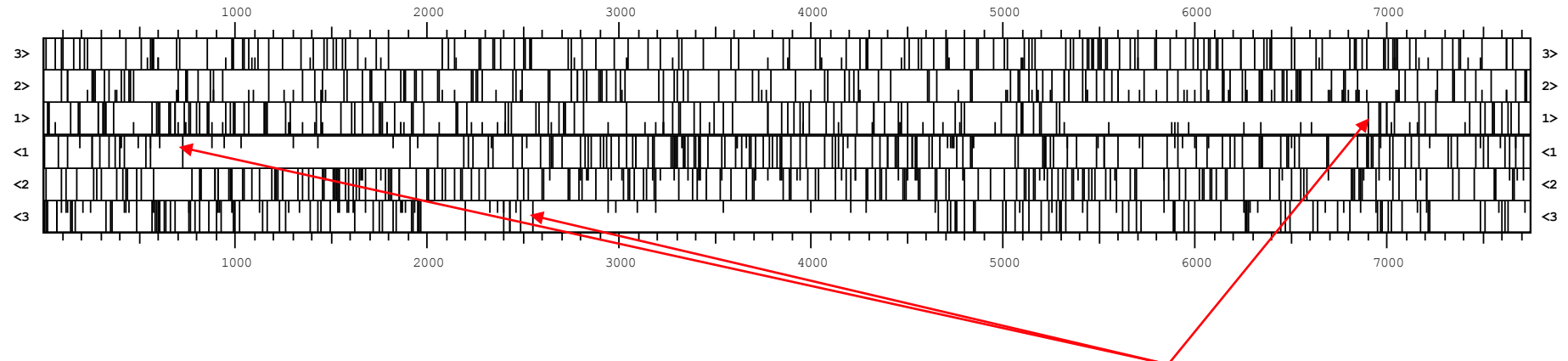


4- La démarche de la bioinformatique



Recherche des phases de lecture ouverte
dans une séquence d'ADN procaryote (ou eucaryote sans intron)

logiciel DNA strider (C. Marck)



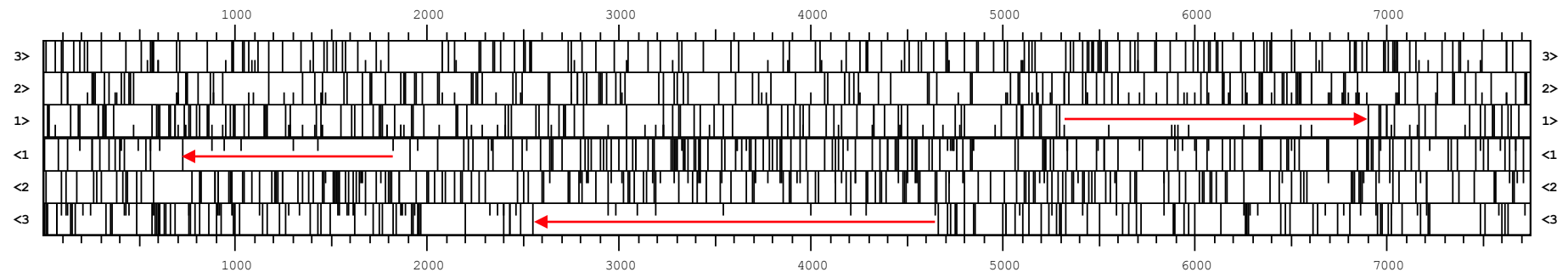
- 6 phases de lecture - codon ATG - codon stop

4- La démarche de la bioinformatique



Recherche des phases de lecture ouverte
dans une séquence d'ADN procaryote (ou eucaryote sans intron)

logiciel DNA strider (C. Marck)



- 6 phases de lecture - codon ATG - codon stop

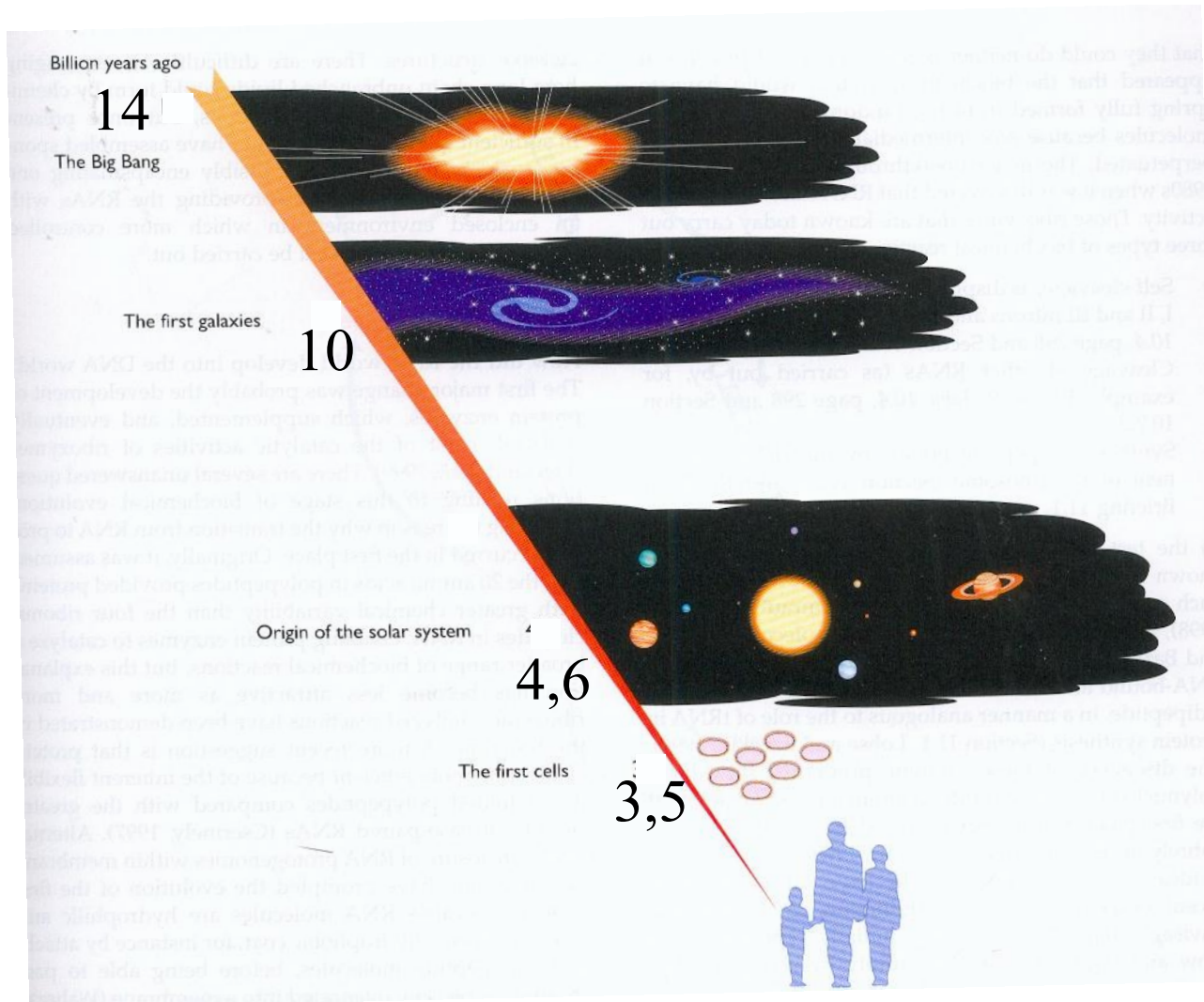
4- La démarche de la bioinformatique



5- Le temps et le monde vivant

5-1 Le temps

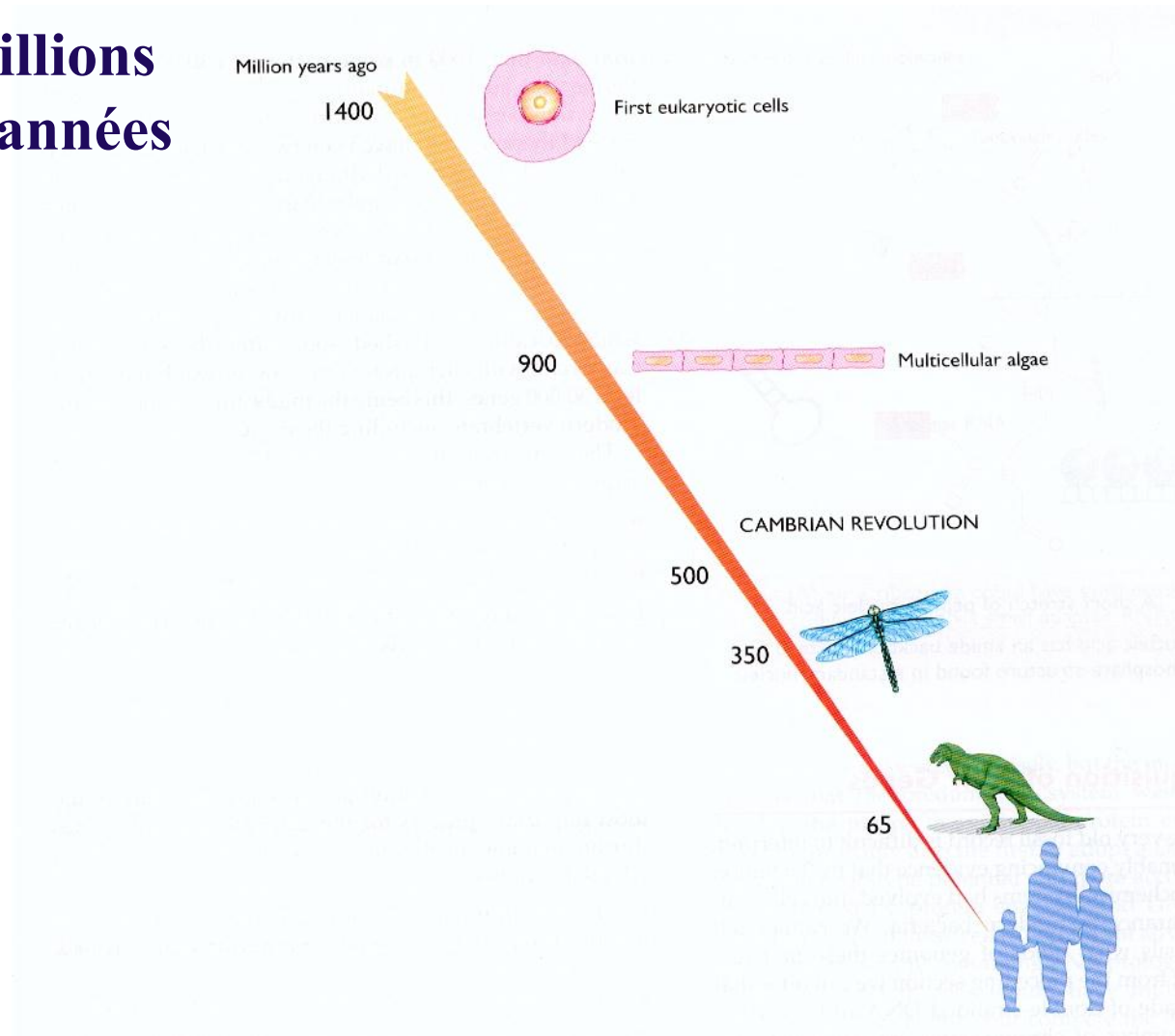
L'échelle du temps en milliards d'années



5-1 Le temps

L'échelle du temps en millions d'années

**Millions
d'années**



5-2 l'ADN est le support de l'information génétique dans les trois domaines du Vivant

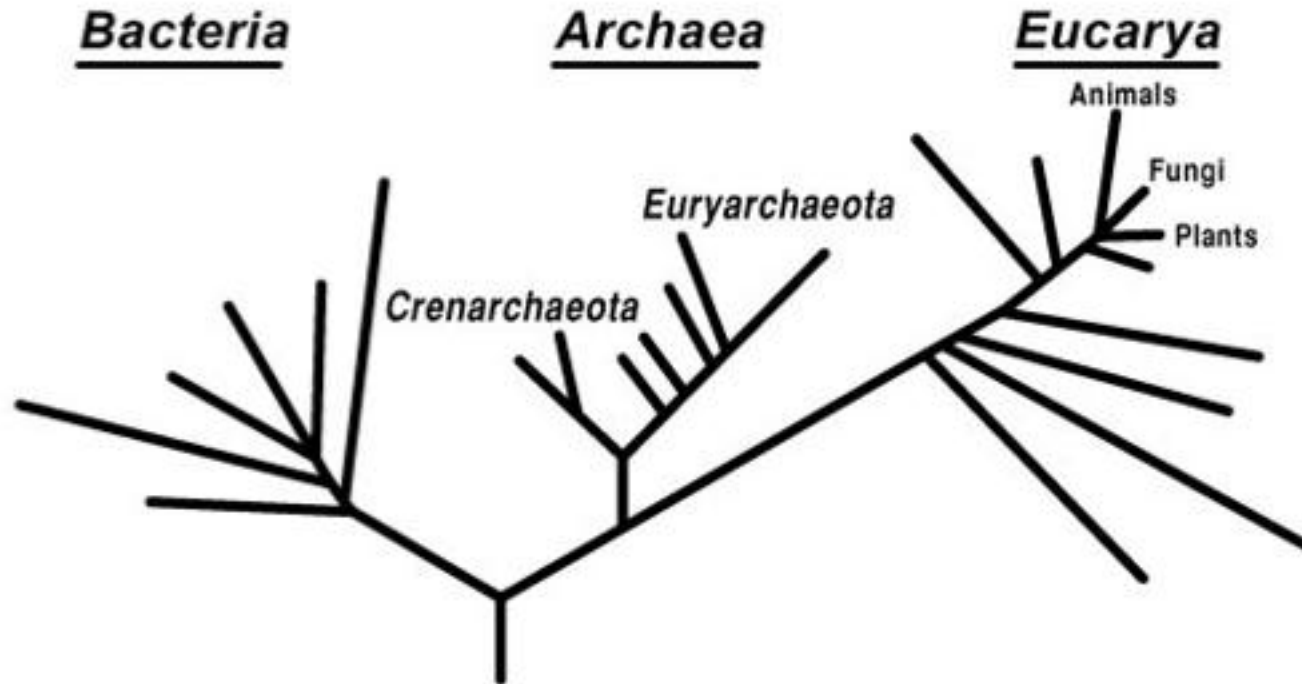
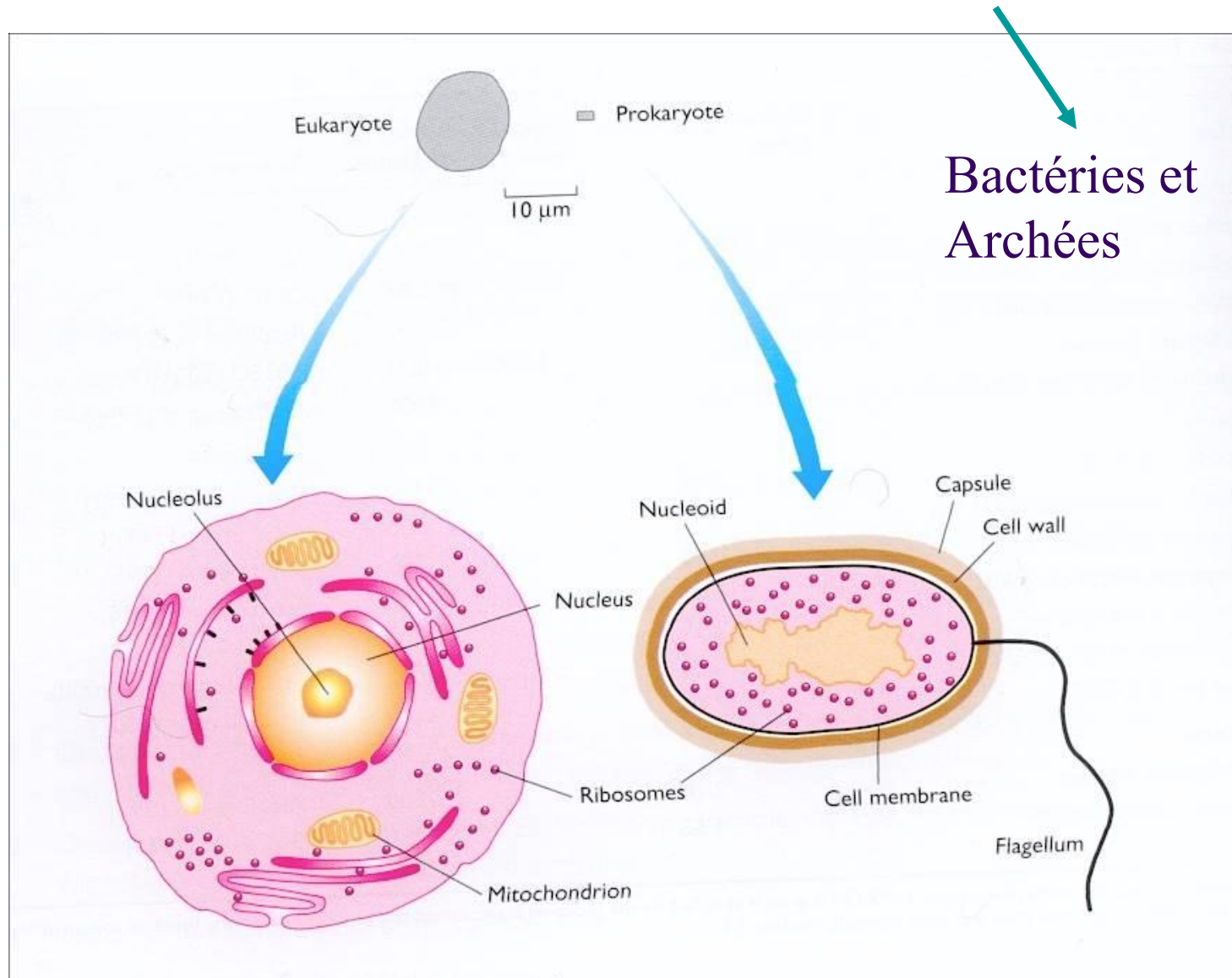


Fig. 1. The universal phylogenetic as determined by ribosomal RNA sequences analyses. [Reproduced with permission from ref. 15 (Copyright 2000, *Proc. Natl. Acad. Sci. USA*).]

5-3 Schéma des cellules eucaryote et procaryote



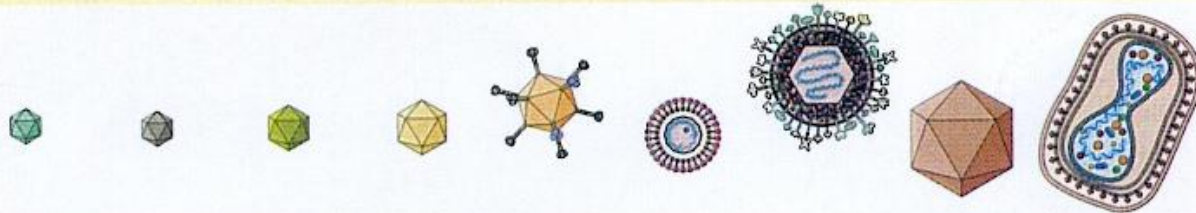
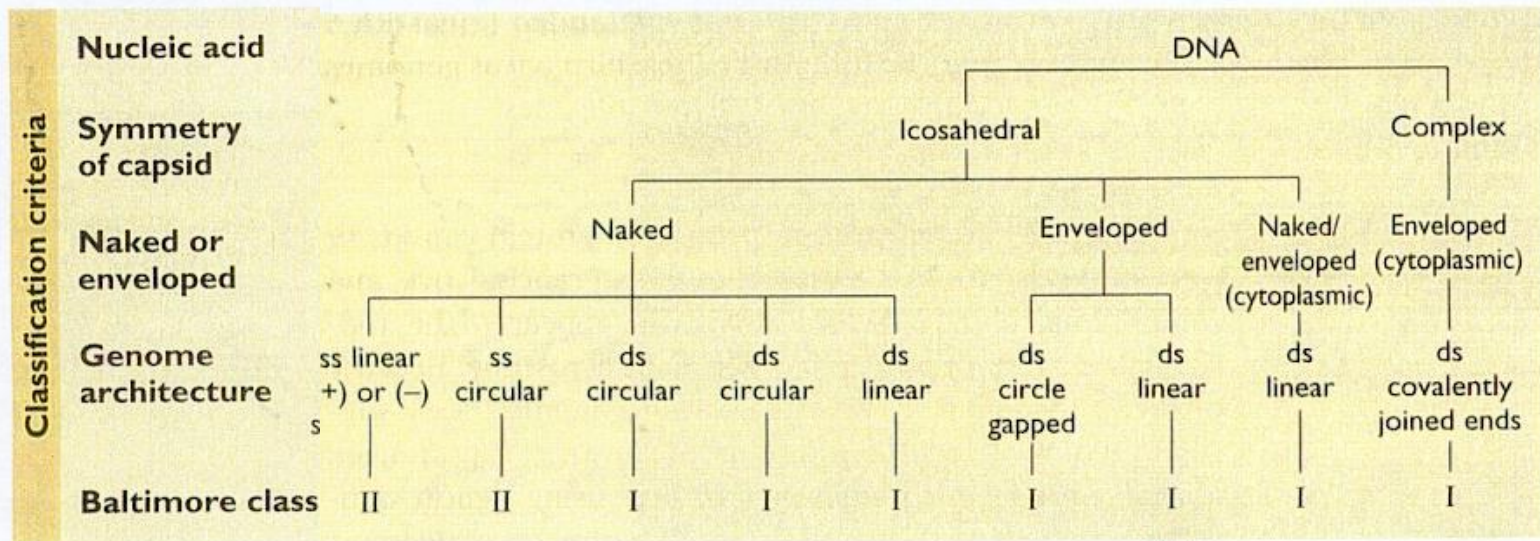
Bactéries et
Archées

5-4 Particularités des Archées

- **Glycoprotéines de la membrane cellulaire différentes des Bactéries et des Eucaryotes**
- **Protéines intervenant dans la réplication, transcription et traduction ressemblent à celles des Eucaryotes**
- **Protéines du métabolisme plus proches des Bactéries que des Eucaryotes**
- **Souvent présentes dans des milieux extrêmes (Méthanophiles, halophiles, thermorésistantes...)**

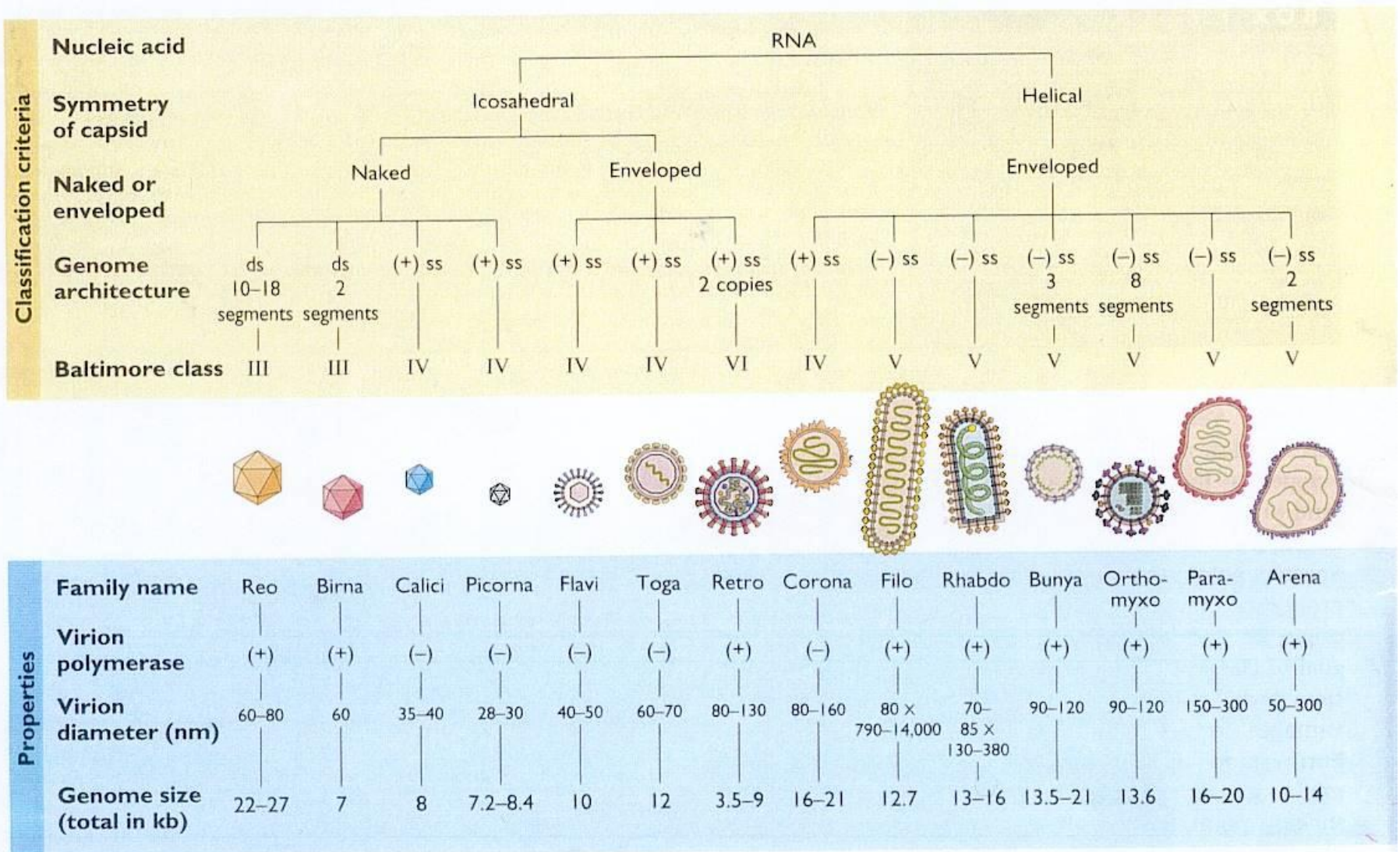
6- Les virus :
un monde hétérogène

6-1 Les virus à ADN



Properties	Family name	Parvo	Circo	Polyoma	Papilloma	Adeno	Hepadna	Herpes	Irido	Pox
	Virion polymerase	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(+)
	Virion diameter (nm)	18-26	12-26	40	55	70-90	42	150-200	125-300	170-200 x 300-450
	Genome size (total in kb)	5	1.8-2.3	5	7-8	36-38	3.2	120-200	150-350	130-280

6-2 Les virus à ARN



D'après Flint *et al*, in Principles of virology

6-3

Les virus sont présents dans les trois domaines du Vivant

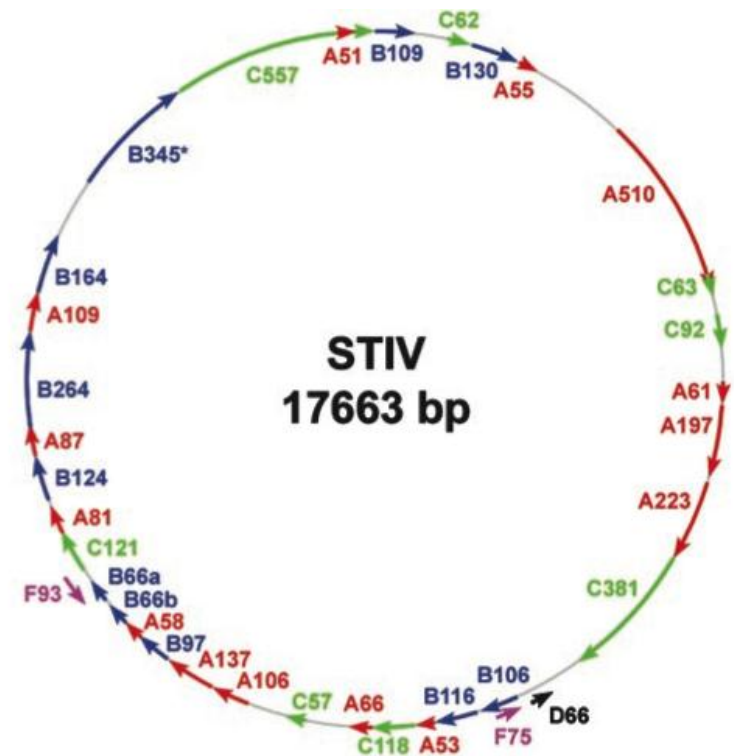
5100 virus connus

Seulement 36 ont été isolés d'Archées

Virus de *Sulfolobus solfataricus*

La capside est similaire à celle de virus de Bactéries et d'Eucaryotes

Ancêtre commun avant la divergence des trois domaines du Vivant ?



G.Rice *et al*, 2004, *PNAS*, 101, p7716

Faire parler les génomes

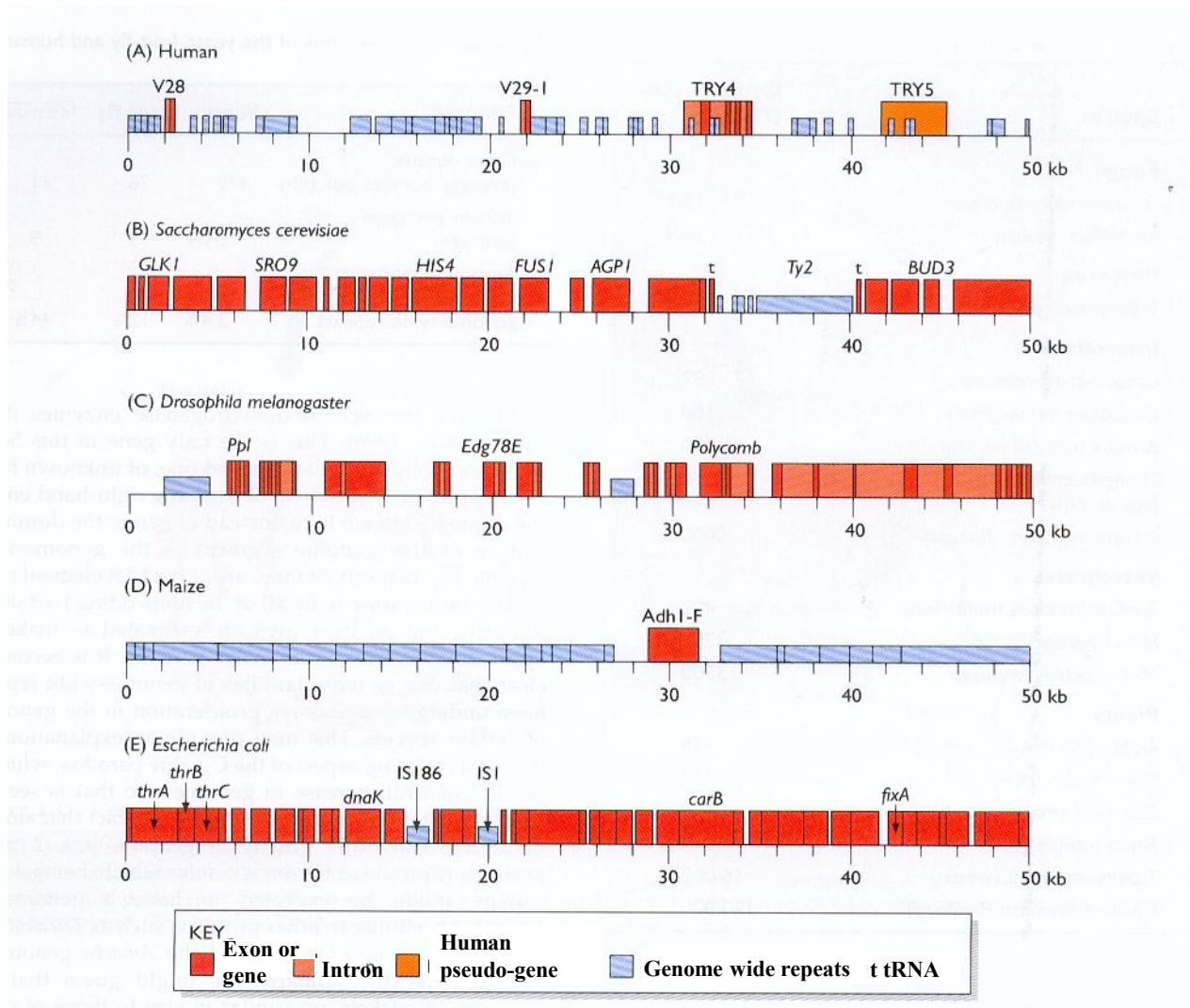
Faire parler les génomes

- *1- Structure des génomes procaryotes et eucaryotes*
- *2- Organisation en opéron des gènes procaryotes*
- *3- Quelles informations sont contenues dans les régions intergéniques ?*
- *4- Les gènes codant des ARN*

Faire parler les génomes

*1- Structure des génomes
procaryotes et eucaryotes*

1-1 Contenu de fragments chromosomiques de quelques espèces



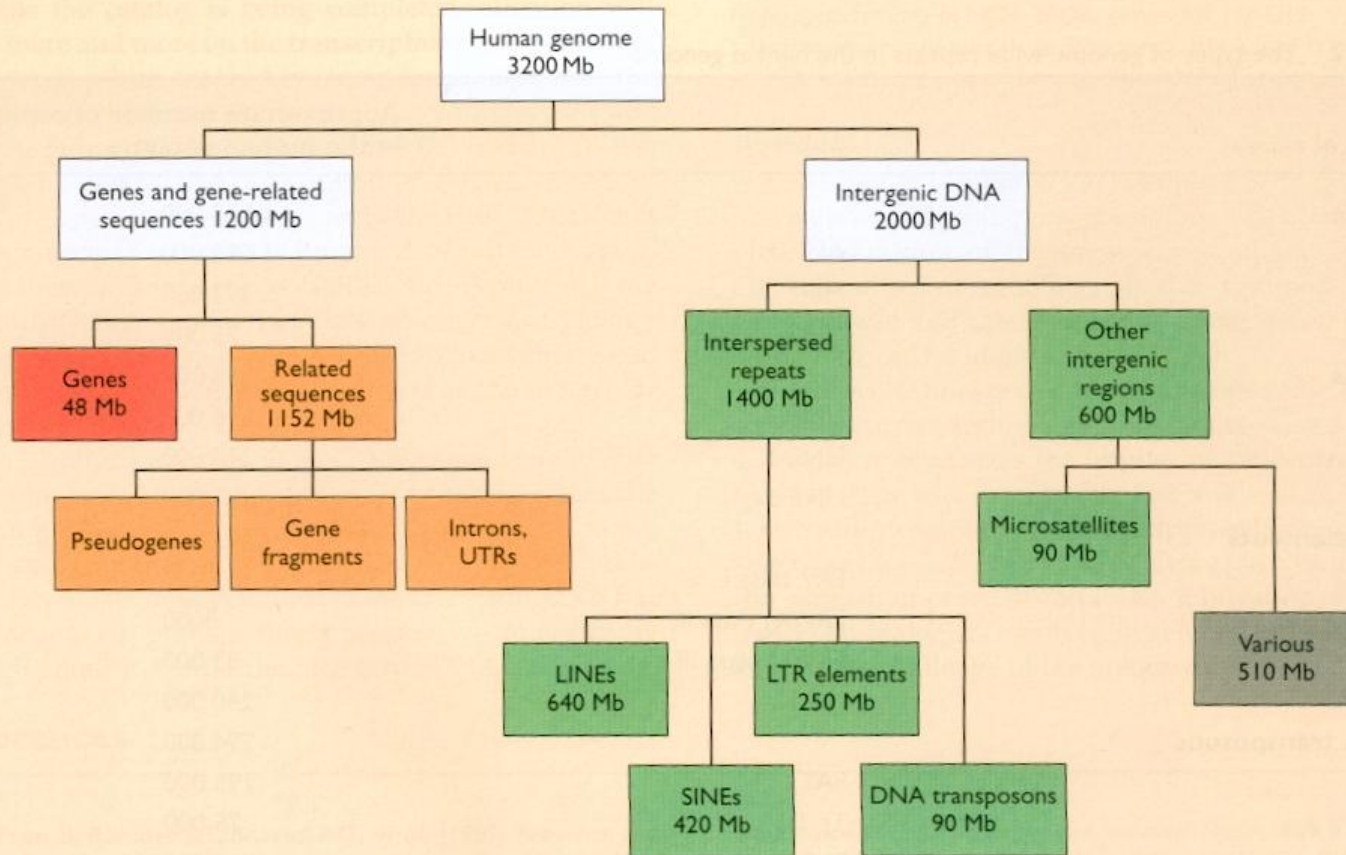
D'après Brown, 2002 in *Genomes*

1-2 Eléments des génomes

Espèce	<i>Helicobacter pylori</i> (Bactérie)	<i>Methanosarcina acetivorans</i> (Archée)	<i>Saccharomyces cerevisiae</i> (levure, Eucaryote)	<i>Homo sapiens</i> (mammifère, Eucaryote)
Chromosome : Nombre Forme Longueur centromère	1 Circulaire Pas de centromère	1 Circulaire Pas de centromère	2 x 16 Linéaire 110 pdb	2 X 23 Linéaire >10 ⁶ pdb
Gènes codant des protéines : % du génome Longueur moyenne d'un gène % de gènes morcelés	91% 945 pdb 0	74% 936 pdb 0	72% 1450 pdb 4%	20%génome transcrit 1 à 4% 20000 pdb environ ≈100%
Gènes codant des ARN : ARN ribosomiques ARN de transfert snARN	0,7% 7 clusters 12 gènes seuls	1,5%	5% 262 copies	0,4% 1300 copies
Séquences répétées : Hautement répétées Moyennement répétées	0% 1% (17 éléments IS)		0% 1,2% (30 copies de Ty)	10% 44%

1-3 Organisation du génome humain

Box 1.4: The organization of the human genome

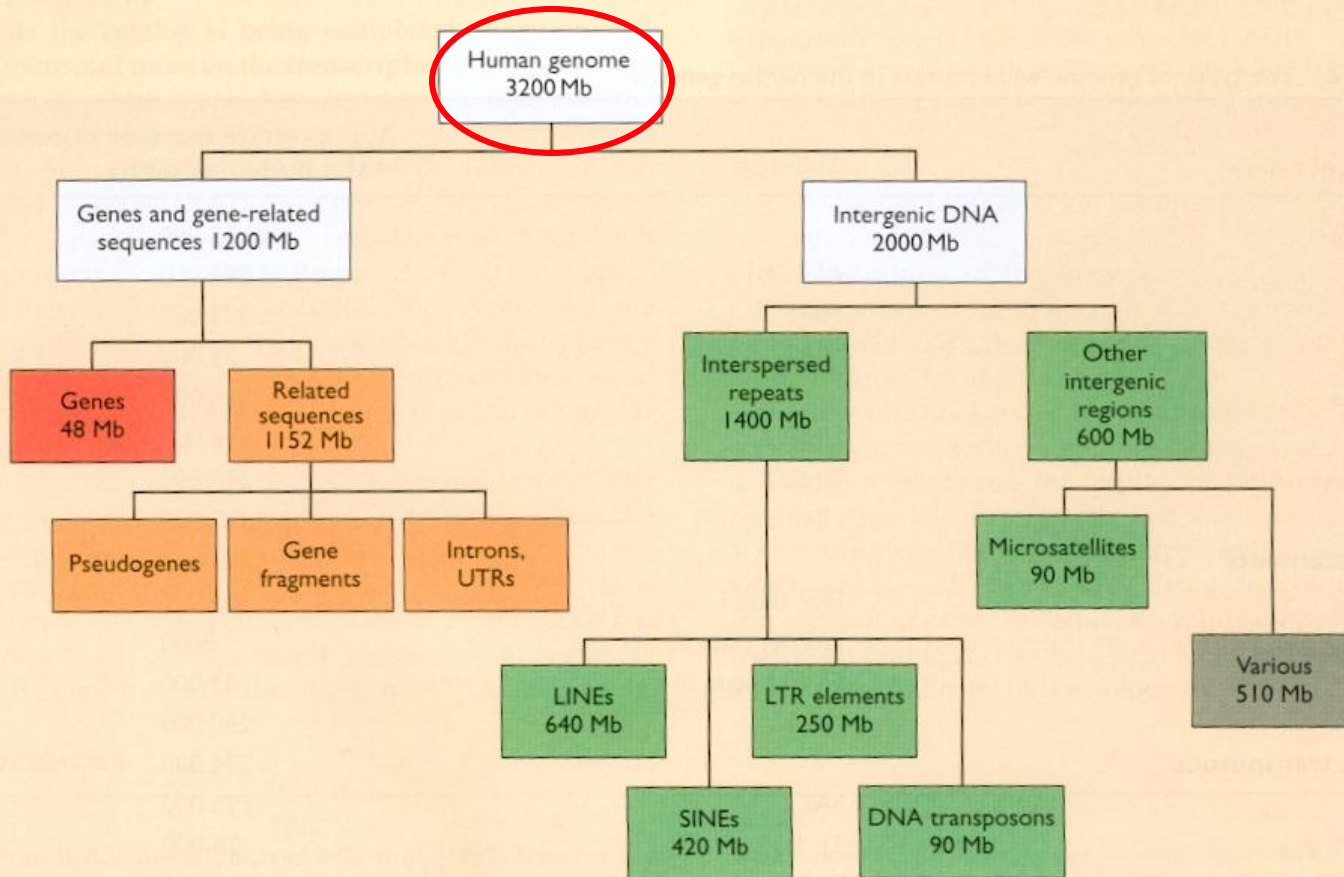


Based on IHGSC (2001) and Venter et al. (2001).

(d'après Brown, 2002, in *genomes*)

1-3 Organisation du génome humain

Box 1.4: The organization of the human genome

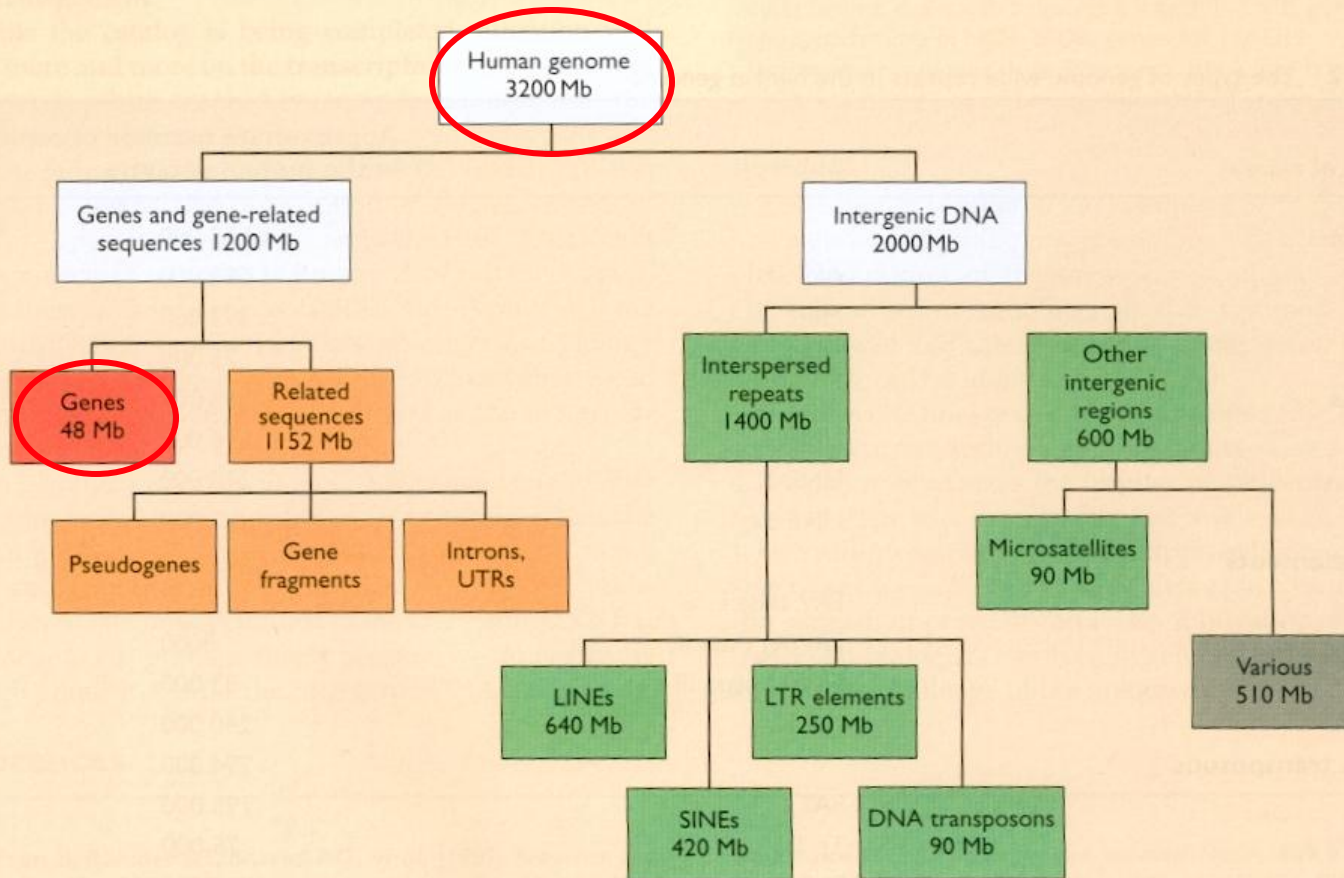


Based on IHGSC (2001) and Venter et al. (2001).

(d'après Brown, 2002, in *genomes*)

1-3 Organisation du génome humain

Box 1.4: The organization of the human genome



Based on IHGSC (2001) and Venter et al. (2001).

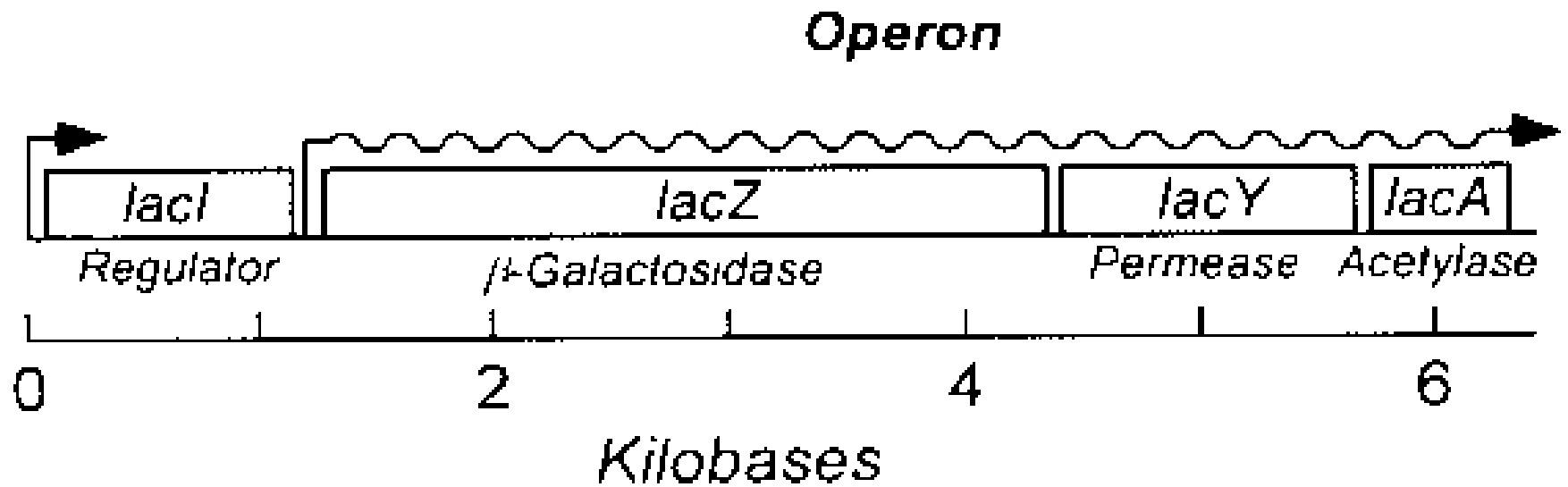
(d'après Brown, 2002, in *genomes*)

Faire parler les génomes

2- Organisation en opéron des gènes procaryotes

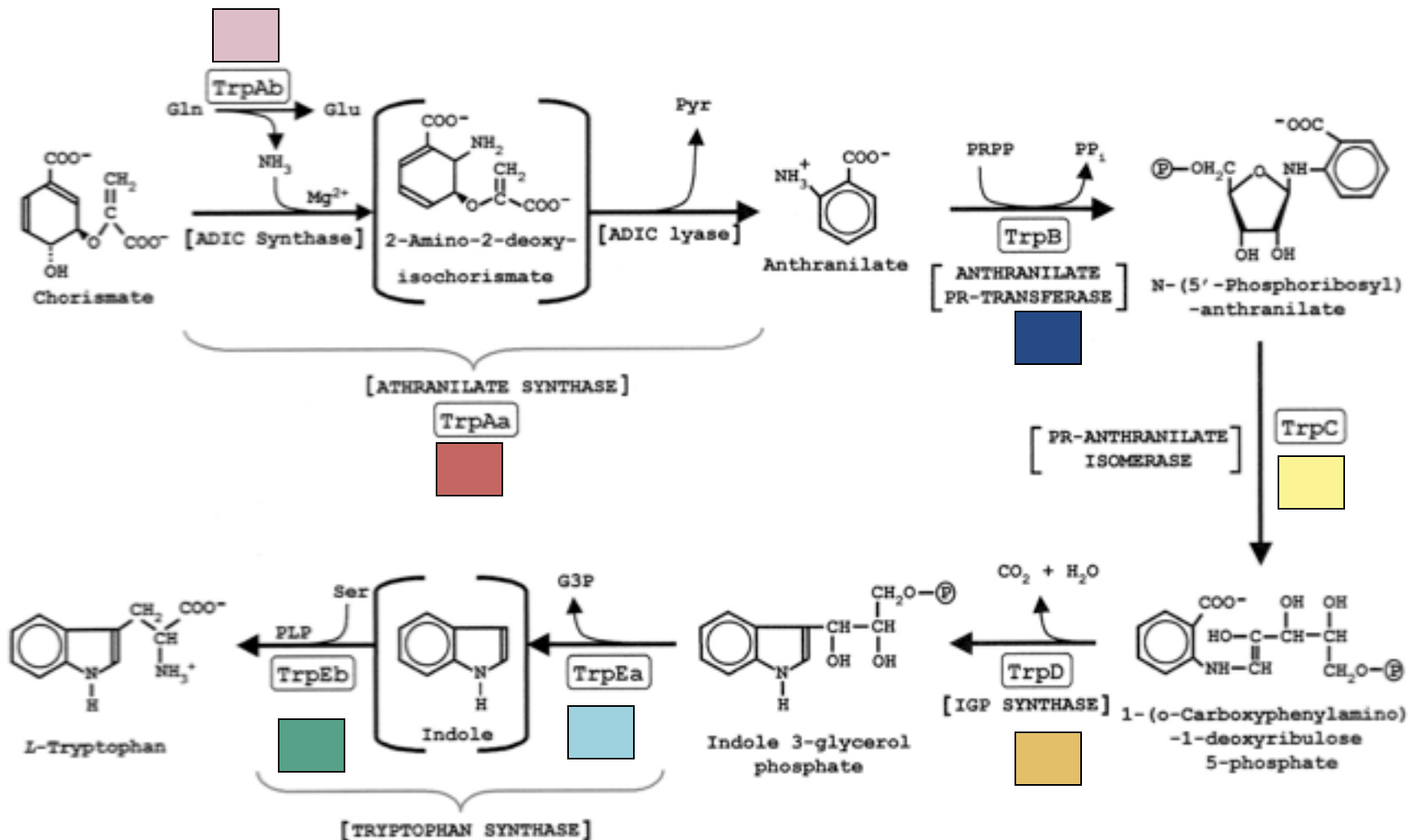
2-1 Les opérons

**Définition : gènes adjacents appartenant
à la même unité de transcription
(*transcription polycistronique*)**

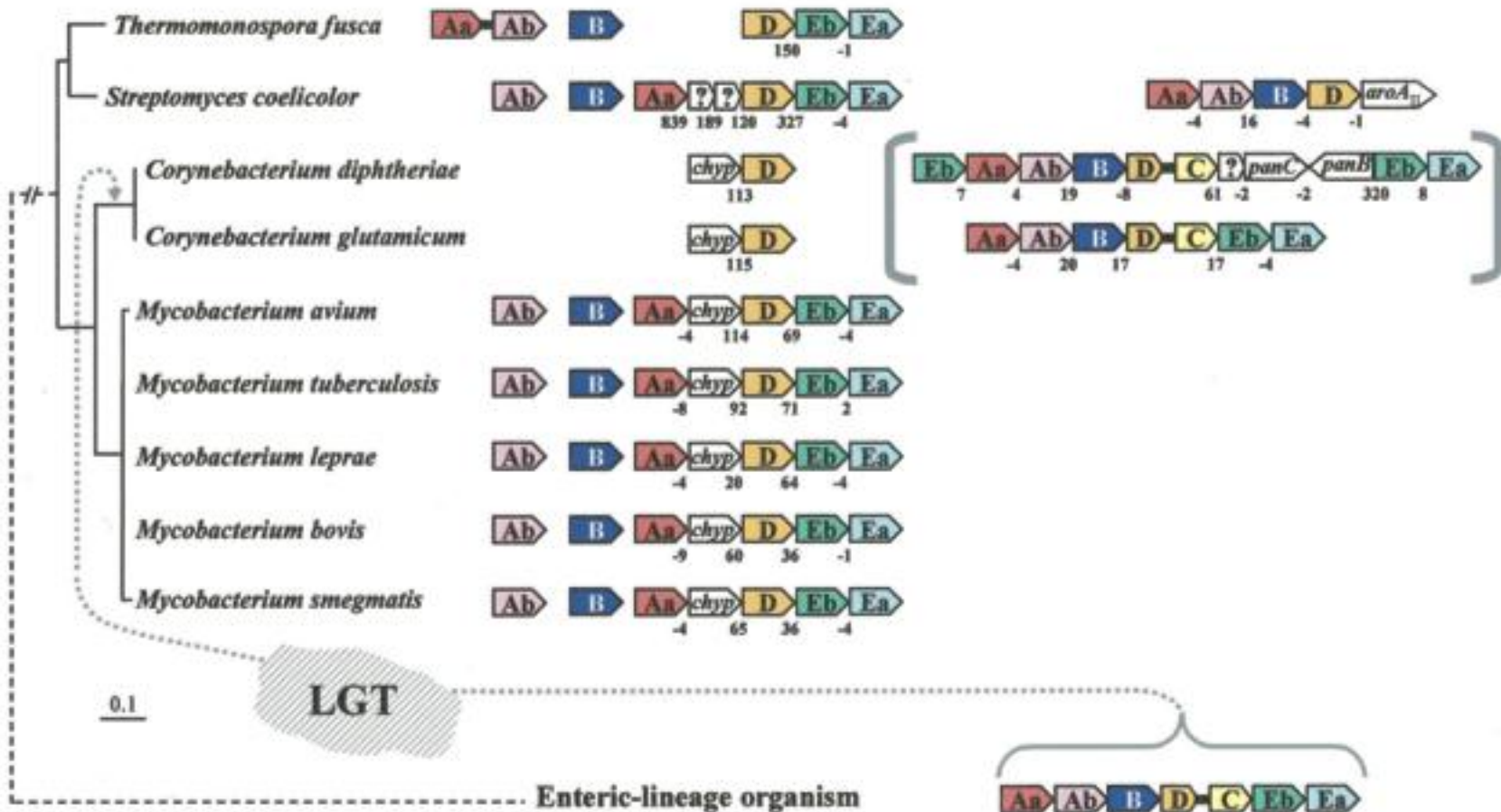


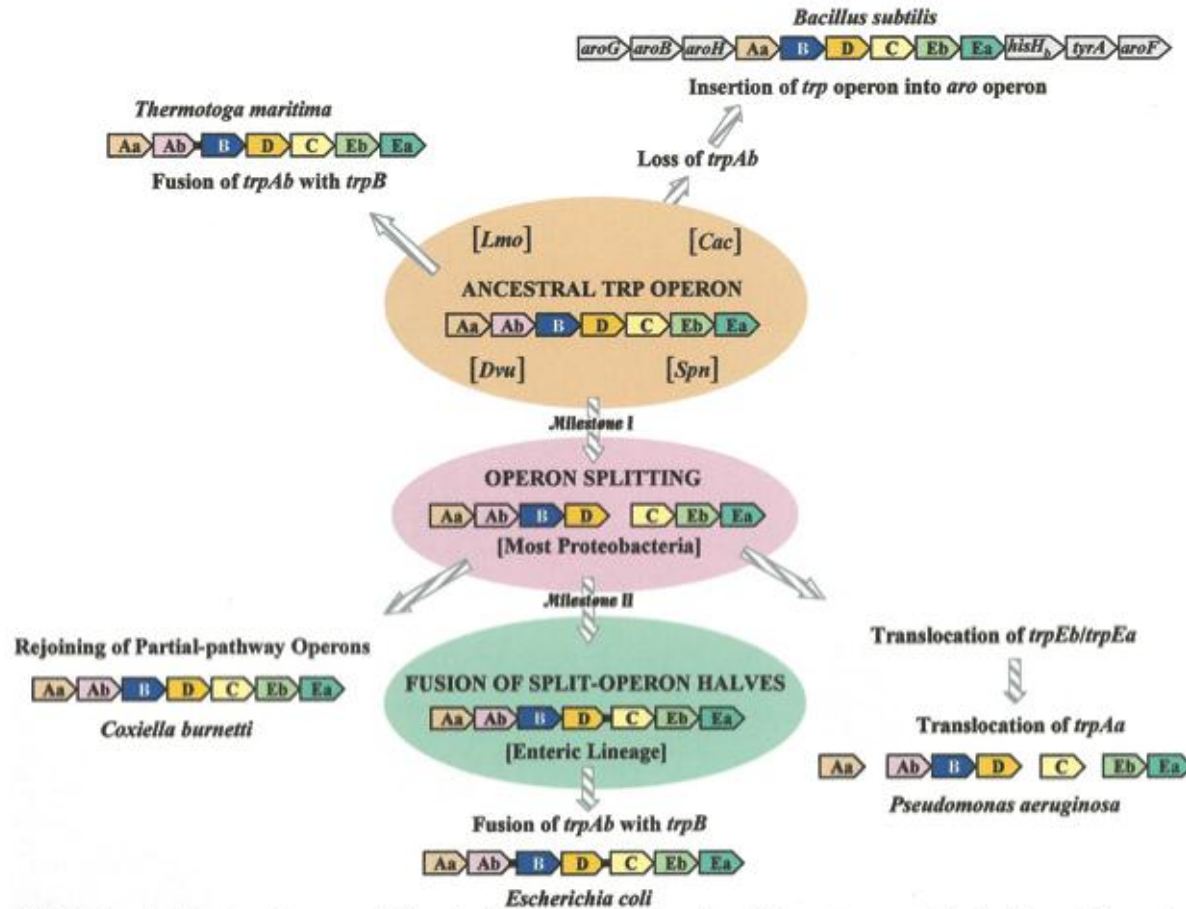
2-2 L'exemple de l'opéron tryptophane

Chaîne de biosynthèse du tryptophane chez *E. coli*



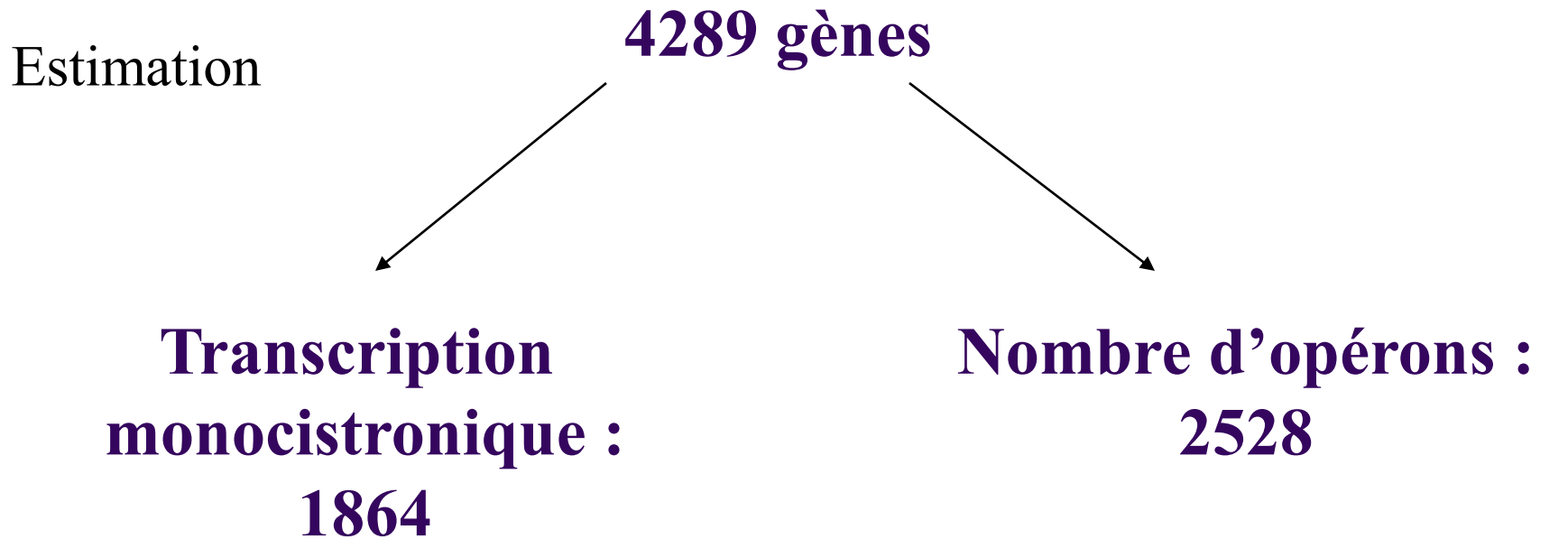
Conservation de l'opéron tryptophane chez différentes bactéries





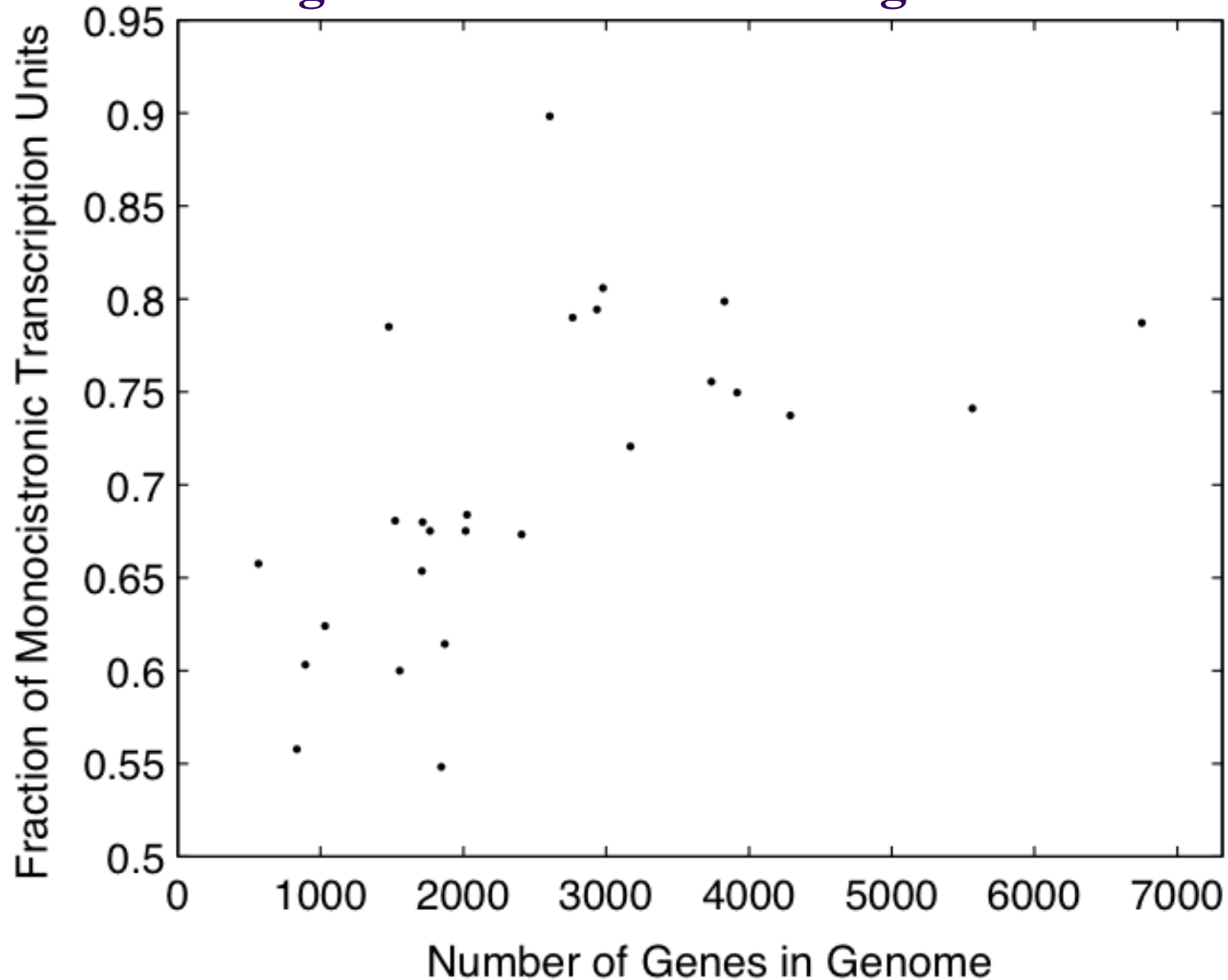
2-3 Les opérons dans leur ensemble

Les opérons de la bactérie *E. coli*

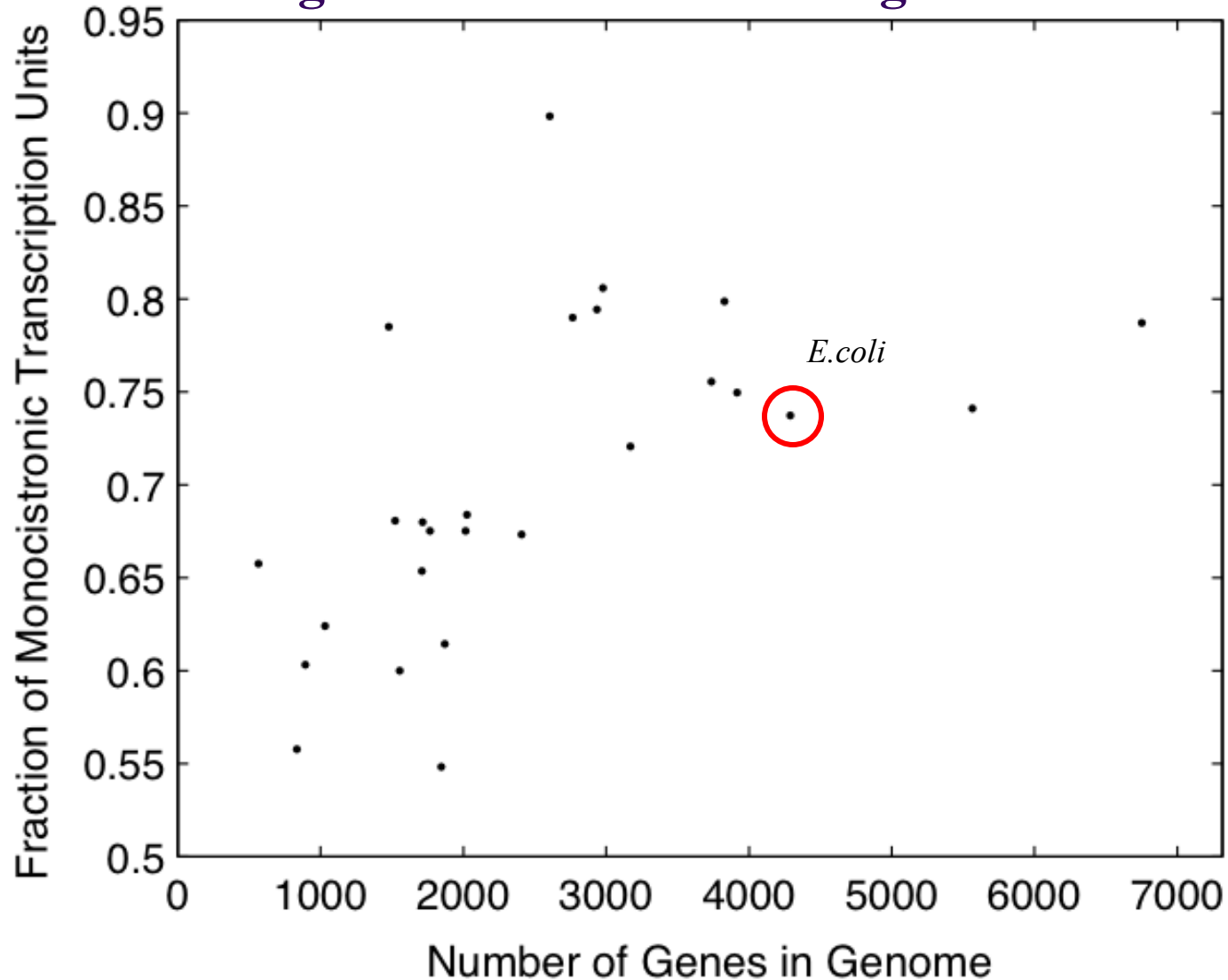


**soit environ 74% des gènes à transcription unique
et en moyenne 1,70 gène par unité de transcription**

**26 génomes de Bactéries et d'Archées :
La fraction de transcription monocistronique
augmente avec la taille du génome**



**26 génomes de Bactéries et d'Archées :
La fraction de transcription monocistronique
augmente avec la taille du génome**



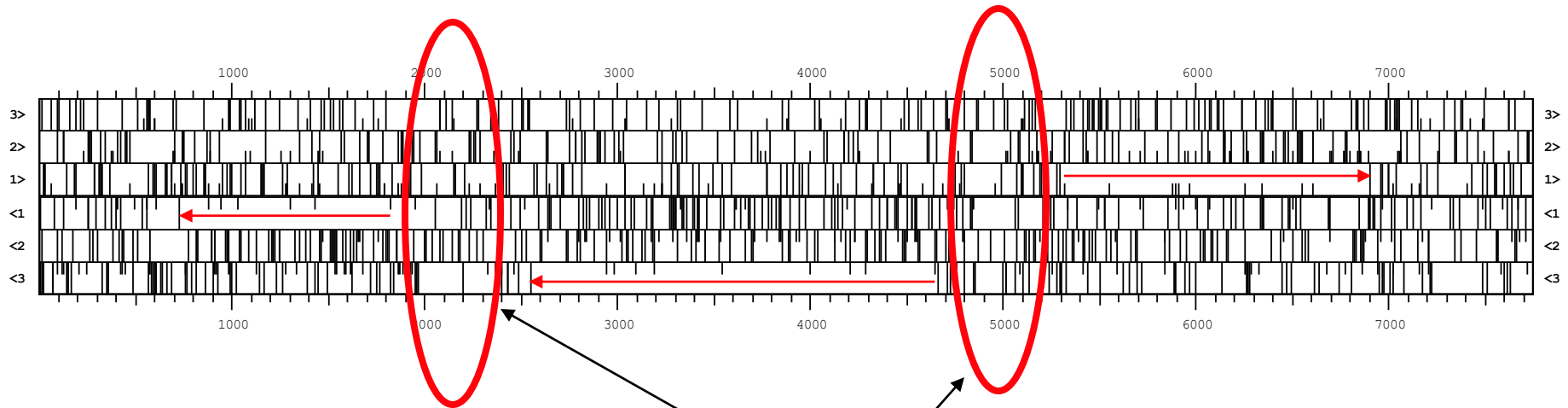
*Plus les génomes sont grands,
moins l'organisation des gènes
en opéron est présente*

Faire parler les génomes

*3- Quelles informations sont
contenues dans les régions
intergéniques ?*

Recherche des phases de lecture ouverte dans une séquence d'ADN procaryote (ou eucaryote sans intron)

logiciel DNA strider (C. Marck)



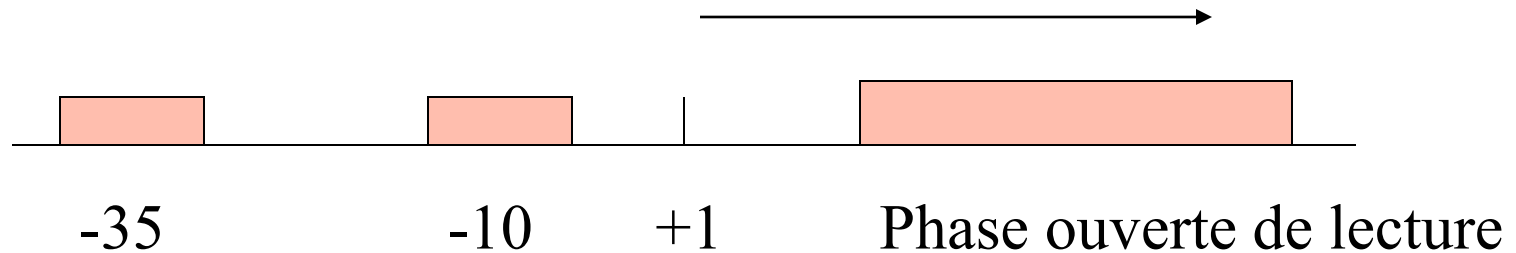
- 6 phases de lecture
- codon ATG
- codon stop

Informations dans ces régions ?
Promoteurs et terminateurs des gènes

3-1 Identification des régions promotrices

*Chez les bactéries :
séquences bien conservées
dans l'ensemble des espèces
donc faciles à identifier*

Promoteur d'un gène d'*Escherichia coli* :



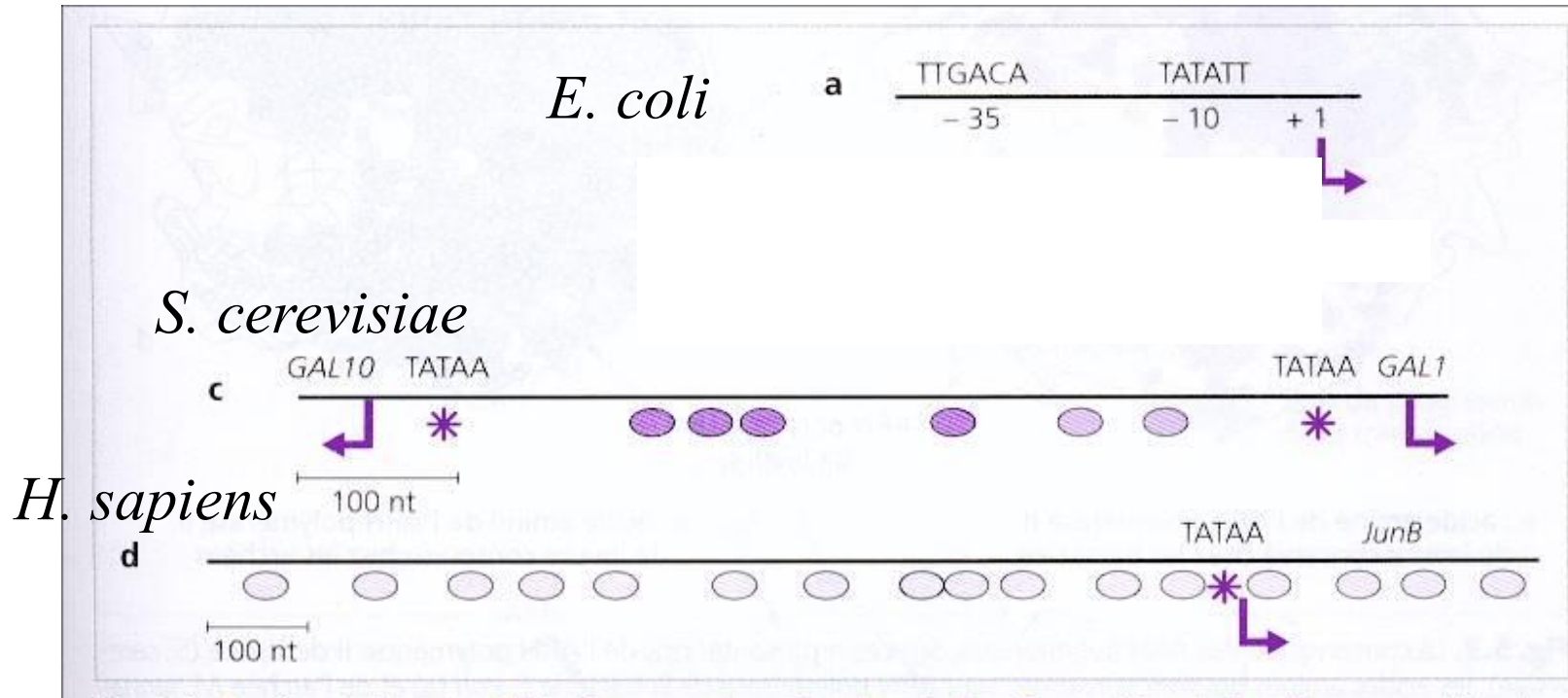
+1 : Démarrage de la transcription

Boîte -35 5'-TTGACA-3'

Boîte -10 (Pribnow box) 5'-TATAAT-3'

*Chez les eucaryotes :
séquences peu conservées
dans l'ensemble des espèces,
donc difficiles à repérer*

Représentation schématique de régions promotrices



From P. Thuriaux, 2004, in *la Levure* (Belin)

3-2 Identification de régions promotrices par génomique comparative

Kellis *et al*, 2003, Nature, vol 423, p241-254

Un exemple d'identification de régions promotrices par génomique comparative

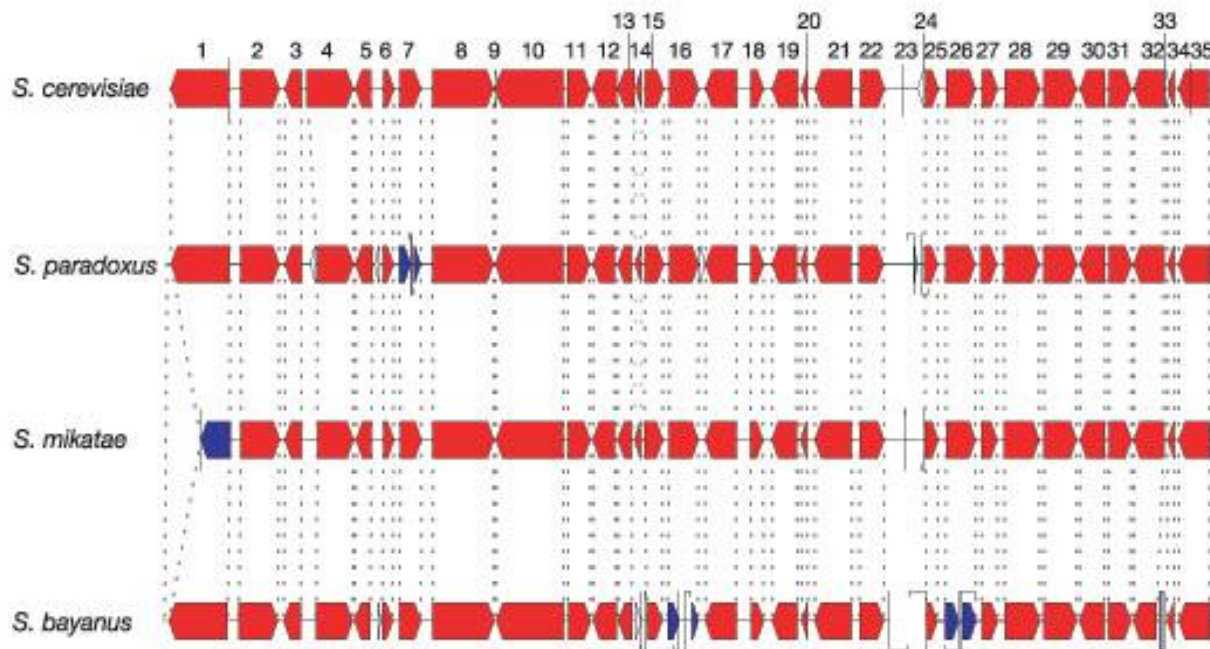
Quatre espèces de levures Saccharomyces

- *S. cerevisiae*
- *S. paradoxus*
- *S. mikatae*
- *S. bayanus*

(<http://genome-www.stanford.edu/Saccharomyces/>).

Un exemple d'identification de régions promotrices par génomique comparative

Comparaison des génomes des quatre levures : Alignements des ORFs



Fragments de 50 kb



ORFs avec homologues



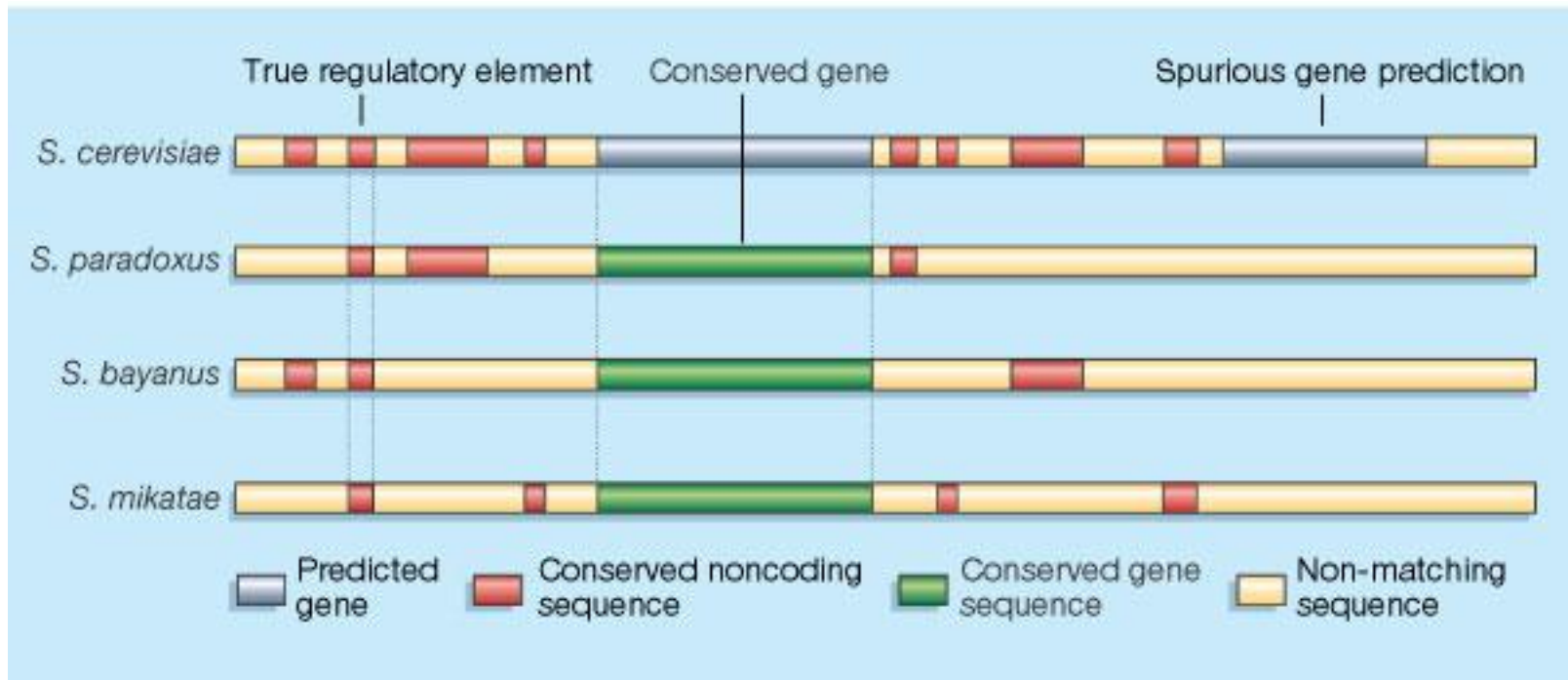
ORFs sans homologue

Un exemple d'identification de régions promotrices par génomique comparative

Conservation des gènes et des intergènes entre *S. cerevisiae* et :

Identité	<i>S. paradoxus</i>	<i>S. mikatae</i>	<i>S. bayanus</i>
<i>Gènes</i>	90%	84%	80%
<i>Régions inter-géniques</i>	80%	70%	62%

Un exemple d'identification de régions promotrices par génomique comparative



*Un exemple d'identification de régions promotrices
par génomique comparative*

*52 régions régulatrices
contenant des motifs distincts
identifiés chez ces levures
Saccharomyces*

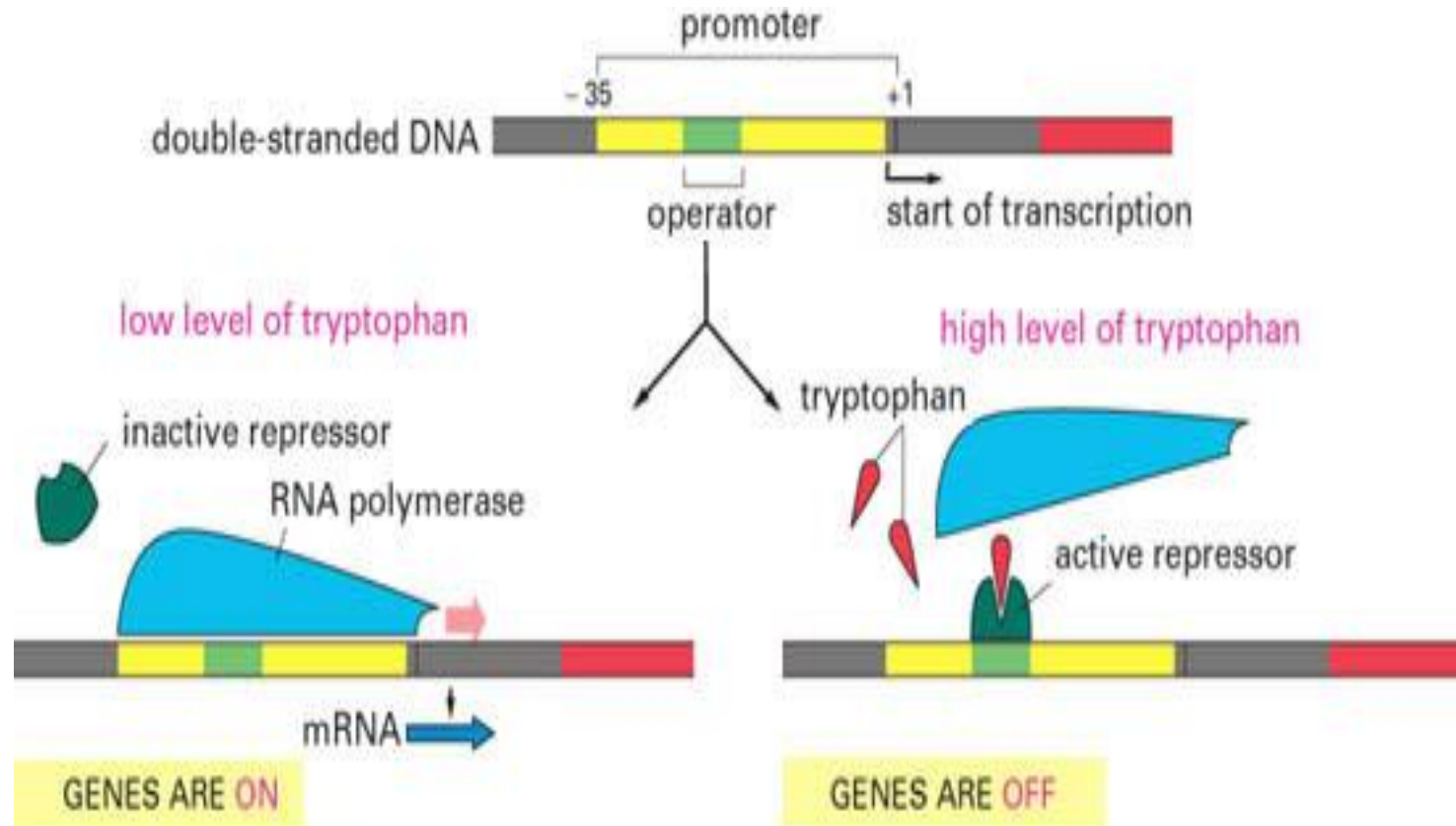
Faire parler les génomes

3-3 Les facteurs de transcription chez les bactéries et les archées

- Perez-Rueda *et al*, 2000, NAR, vol 28, p1838,
- Perez-Rueda *et al*, 2004, Computational Biology and Chemistry, vol 28, p341

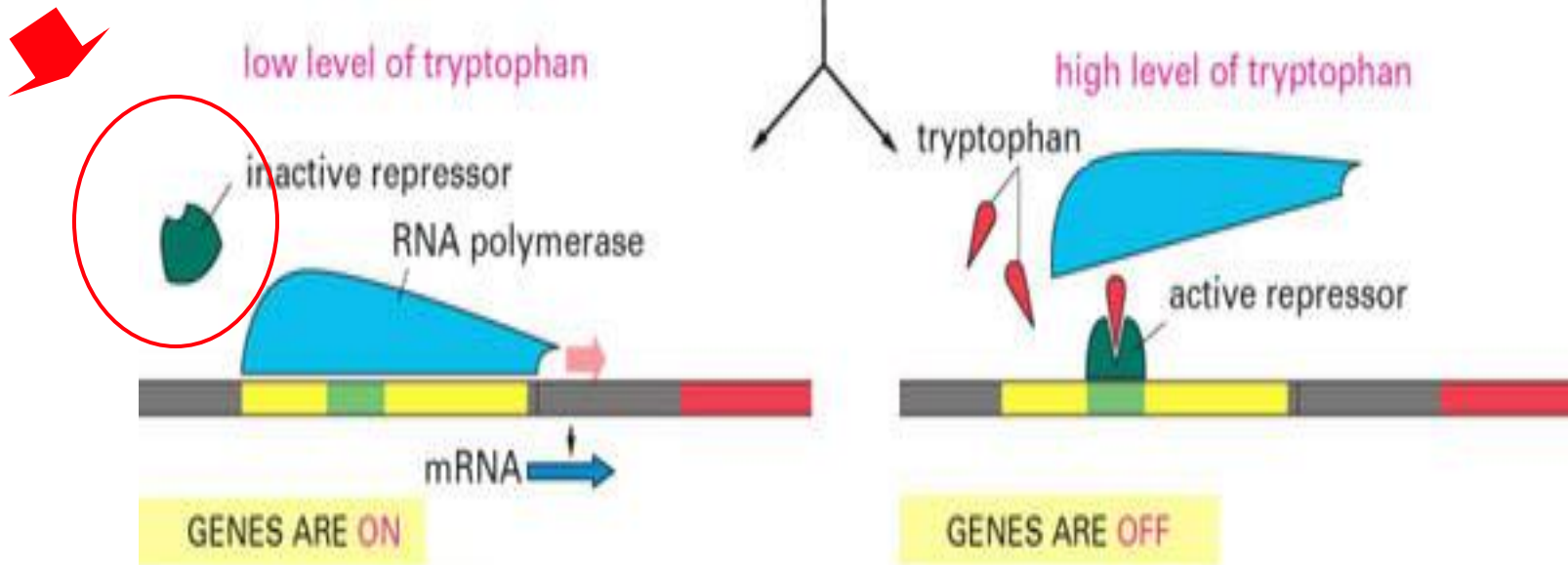
3-3-1 Quel est le rôle des facteurs de transcription ?

Initiation de la transcription



Initiation de la transcription

Facteur
de
transcription



3-3-2 Facteurs de transcription chez *E. coli*

314 protéines régulatrices :

- 35% activateurs**
- 43% répresseurs**
- 22% double fonction**

-Perez-Rueda *et al*, 2000, NAR, vol 28, p1838,

3-3-3 Distribution des facteurs de transcription chez les bactéries et les archées

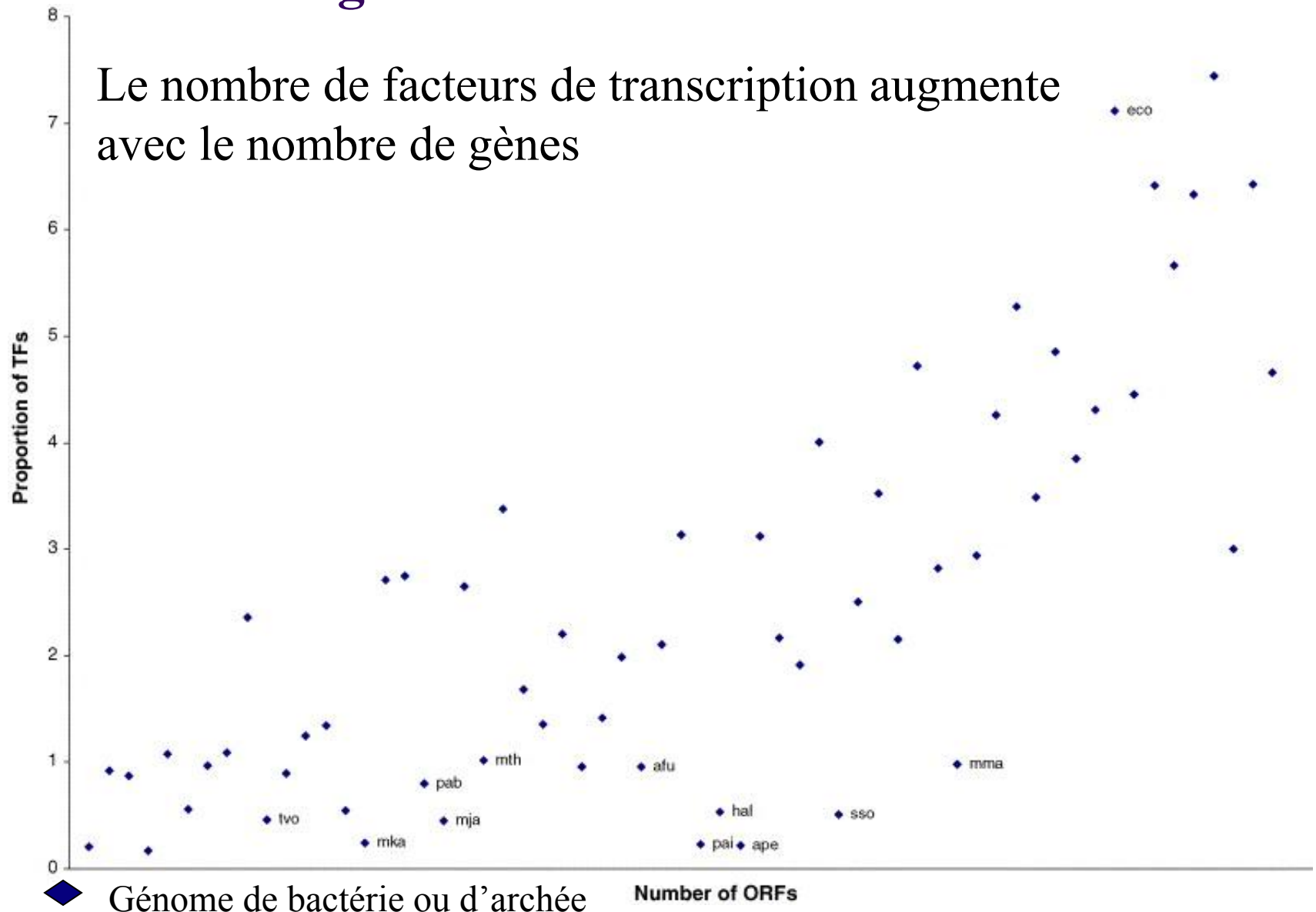
75 familles dans 90 génomes

Etude *in silico*

- Perez-Rueda *et al*, 2004, Computational Biology and Chemistry, vol 28, p341

Distribution des facteurs de transcription dans les génomes de bactéries et d'archées

Le nombre de facteurs de transcription augmente avec le nombre de gènes



Distribution des facteurs de transcription dans les génomes de bactéries et d'archées

FAMILY	ARCHAEA	PROTEOBACTERIA	FIRMICUTE	DIVERSE
AraC/XylS	A-CD--GHIJK-----	ABCDE-GHIJK--NOPQRSTUVWXYZ	-BCDEFGH--KLMNO-	-BCDEF-HIJKL--O--R-
ArsR	ABCDEFGHIJKLMNO-	ABCD--GHIJ--NOP-RSTUVWX	-BCDEFGH---LMNO-	ABCDEFGHIJK---O---S
AsnC	-BCDEFGH-JKLMNOP	ABCD---HI---NOPQRSTUVWXYZ	-BCD-----N--	--C----H-JKL--O---S
BirA	--C-E-GH-JK-MN--	ABC--GHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCDEFGHIJKLMNOP-R-
Fur	A-C-E-G-----OP	ABCD--GHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCD-FGHIJK---O-Q-S
LysR	-B--E--HI-K-----	ABCD--GHIJ-L-NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	A--DEFGHIJK---O---S
MarR	ABCD--G--JK---OP	ABCD--HI---N---R-----	-BCDEFGH---L----	---DE--HIJ-L--O---S
Pho	-----FGH-JK-----	ABCD--G-I---NOP-R-TUVWX	-B-DEFGH---LMNO-	ABCD-F-HIJK---OPQ--
TetR/AcrR	A-CDEFG--JKL----	ABCD--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---LMNOP	AB-DEF-HIJKL--O---S
HypF	-----E--HI-K-M---	-----K--NOP-----	-----F-----O-	A-CD-F-----O----
AslB	-B-D--G--JK-M---	-----G-----NOP--S-UV-X	---D-----O-	-BC--F----K---O----
WrbA	--C-E-G--J-----	-BC--G-I---NOP-R-TUVWX	-BC-----	-----S
Crp	-----	ABCD--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---MNO-	ABCDEFGHIJKLM--O--RS
GalR/LacI	-----	ABCDE-G---NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	-B--E--H-JK---O---S
DnaA	-----	ABCDEFGHI--LMNOPQRSTUVWXYZ	-BCDEFGH--KLMNOP	ABCD-FGHIJKLMNOPQRS
Ebp	-----	ABCDEFGHIJ-LMNOPQRSTUVWXYZ	-BC-EF-H-----O-	ABC-----H--KL---QR-
DeoR	-----	AB-D--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	AB--E--H--K---O---S
LexA	-----	ABCD--G-I---NOPQRSTUVWXYZ	-BCDEFGH---L----	-BCD-F-HIJ-L--O----
OmpR	-----	ABCDEFGHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCDEFGHIJKLM--OP--S
RpiR	-----I-----	AB-D--GHI--NOPQRST---X	-BCDEFGH---LMNO-	-B-----H--K---O----
LuxR/UhpA	-----I-----	ABCD--GHI---NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	---D-FGHIJKL-----S
IclR	-----F-----	AB-D--G-I---NOPQRS-UV-X	-BC-E-----	AB-----H-JK-----S
HipB	-----K-----	AB-D--G-I---NOP-R-T----	-B-DEF-H---L----	-B--E--H--K---O----
Cold	-----F-----	ABCDEFGHI--LMNOPQRSTUVWXYZ	--CDEFGH---LMNO-	ABC-E--HIJK---O-Q-S
MerR	-----E-----	ABCDEF-G-IJK--NOPQRSTUVWXYZ	-BCDEFGH---LMN--	A--DEFGHIJKL--O---S
GntR	--CD-----O-	ABCD--GHI--MNOPQRSTUVWXYZ	-B-DEFGH---LMNO-	-B-DEFGH-JKL--O---S
Ihf	-----OP-----	ABCDEFGHIJKLMNOPQRSTUVWXYZ	-BCDEFGHIJKL-NOP	ABCDEFGHIJKLMNOPQRS
ArgR	-----	-----NOPQ-ST---X	-BCDEFGH---LMNO-	-BC----HIJKL-NO---S
NadR	-----	-----NOPQ-S---X	-----G-----	-----J-----
DniR	-----	A-----HIJK--NOP-R-TUVWX	--CDE--H---L--O-	A-C-----L--O--R-
MetJ	-----	-----NOPQ-ST---X	-----	-----
Crl	-----	-----NOP--T---X	-----	-----
AlpA	-----	-----N-P---T-VWX	-----	-----
FhlD	-----	-----I---NOP-----	-----	-----
FhlC	-----	-----I---NOP-----X	-----	-----
DicA	-----	-----NOP-----	-----	-----
Caif	-----	-----NOP-----	-----	-----
HycA	-----	-----NOP-----	-----	-----
HtgA	-----	-----N-----	-----	-----
DicC	-----	-----N-----	-----	-----
GcvR	-----	--C-----NOP-R-TUVWX	-----	-----
GutM	-----	-----NOP--S-----	-----E-----	-----
HipA	-----	--CD---I---N-----UV--	-----	-----L-----
TrpR	-----	-----NOPQ-ST--WX	-----	-----Q-----
FrvR	-----	-----N-----	--C--F-H-----	-----
ModE	---DE-----	A-CD--G-I---NOPQRS---X	-----	-C-----
NagR	-----	A-----G-----N-P--ST---X	-B--EF-H---M-O-	-B-----

Distribution des facteurs de transcription dans les génomes de bactéries et d'archées

FAMILY	ARCHAEA	PROTEOBACTERIA	FIRMICUTE	DIVERSE
AraC/XylS	A-CD--GHIJK-----	ABCDE-GHIJK--NOPQRSTUVWXYZ	-BCDEFGH--KLMNO-	BCDEF-HIJKL--O--R-
ArsR	ABCDEFGHIJKLMNO-	ABCD--GHIJ--NOP-RSTUVWX	-BCDEFGH---LMNO-	ABCDEFGHIJK---O---S
AsnC	-BCDEFGH-JKLMNOP	ABCD---HI---NOPQRSTUVWXYZ	-BCD-----N--	--C----H-JKL--O---S
BirA	--C-E-GH-JK-MN--	ABC--GHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCDEFGHIJKLMNOP-R-
Fur	A-C-E-G-----OP	ABCD--GHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCD-FGHIJK--O-Q-S
LysR	-B--E--HI-K-----	ABCD--GHIJ-L-NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	A--DEFGHIJK--O---S
MarR	ABCD--G--JK---OP	ABCD--HI---N--R-----	-BCDEFGH---L---	---DE--HIJ-L--O---S
Pho	-----FGH-JK-----	ABCD--G-I---NOP-R-TUVWX	-B-DEFGH---LMNO-	ABCD-F-HIJK---OPQ--
TetR/AcrR	A-CDEFG--JKL----	ABCD--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---LMNOP	AB-DEF-HIJKL--O---S
HypF	-----E--HI-K-M---	-----K--NOP-----	-----FGH---O---	A-CD-F-----O-----
AslB	-B-D--G--JK-M---	-----G-----NOP--S-UV-X	---D-----O---	-BC--F----K---O----
WrbA	--C-E-G--J-----	-BC--G-I---NOP-R-TUVWX	-BC-----O-----	-----H---K---O-----S
Crp	-----	ABCD--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---MNO-	ABCDEFGHIJKL--O--RS
GalR/LacI	-----	ABCDE-G--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	-B--E--H-JK---O---S
DnaA	-----	ABCDEFGHI--LMNOPQRSTUVWXYZ	-BCDEFGH--KLMNOP	ABCD-FGHIJKLMNOPQRS
Ebp	-----	ABCDEFGHIJ-LMNOPQRSTUVWXYZ	-BC-EF-H-----O-	ABC-----H--KL---QR-
DeoR	-----	AB-D--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	AB--E--H--K---O---S
LexA	-----	ABCD--G-I---NOPQRSTUVWXYZ	-BCDEFGH---L---	-BCD-F-HIJ-L--O-----
OmpR	-----	ABCDEFGHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCDEFGHIJKL--OP--S
RpiR	-----I-----	AB-D--GHI--NOPQRST---X	-BCDEFGH---LMNO-	-B-----H--K---O----
LuxR/UhpA	-----I-----	ABCD--GHI--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	---D-FGHIJKL-----S
IclR	-----F-----	AB-D--G-I---NOPQRS-UV-X	-BC-E-----O-----	AB-----H-JK-----S
HipB	-----K-----	AB-D--G-I---NOP-R-T----	-B-DEF-H---L---	-B--E--H--K---O-----
Cold	-----F-----	ABCDEFGHI--LMNOPQRSTUVWXYZ	--CDEFGH---LMNO-	ABC-E--HIJK---O-Q-S
MerR	-----E-----	ABCD--G-IJK--NOPQRSTUVWXYZ	-BCDEFGH---LMN--	A--DEFGHIJKL--O---S
GntR	--CD-----O----	ABCD--GHI--MNOPQRSTUVWXYZ	-B-DEFGH---LMNO-	-B-DEFGH-JKL--O---S
Ihf	-----OP-----	ABCDEFGHIJKLMNOPQRSTUVWXYZ	-BCDEFGHIJKL-NOP	ABCDEFGHIJKLMNOPQRS
ArgR	-----	-----NOPQ-ST---X	-BCDEFGH---LMNO-	-BC----HIJKL-NO---S
NadR	-----	-----NOPQ-S---X	-----G-----	-----J-----
DniR	-----	A-----HIJK--NOP-R-TUVWX	--CDE--H---L--O-	A-C-----L--O--R-
MetJ	-----	-----NOPQ-ST---X	-----	-----
Crl	-----	-----NOP--T---X	-----	-----
AlpA	-----	-----N-P---T-VWX	-----	-----
FhlD	-----	-----I---NOP-----	-----	-----
FhlC	-----	-----I---NOP-----X	-----	-----
DicA	-----	-----NOP-----	-----	-----
Caif	-----	-----NOP-----	-----	-----
HycA	-----	-----NOP-----	-----	-----
HtgA	-----	-----N-----	-----	-----
DicC	-----	-----N-----	-----	-----
GcvR	-----	--C-----NOP-R-TUVWX	-----	-----
GutM	-----	-----NOP--S-----	-----E-----	-----
HipA	-----	--CD---I---N---UV--	-----	-----L-----
TrpR	-----	-----NOPQ-ST--WX	-----	-----Q-----
FrvR	-----	-----N-----	--C--F-H-----	-----
ModE	---DE-----	A-CD--G-I---NOPQRS---X	-----	--C-----
NagR	-----	A-----G-----N-P--ST---X	-B--EF-H---M-O-	-B-----

Distribution des facteurs de transcription dans les génomes de bactéries et d'archées

FAMILY	ARCHAEA	PROTEOBACTERIA	FIRMICUTE	DIVERSE
AraC/XylS	A-CD--GHIJK-----	ABCDE-GHIJK--NOPQRSTUVWXYZ	-BCDEFGH--KLMNO-	-BCDEF-HIJKL--O--R-
ArsR	ABCDEFGHIJKLMNO-	ABCD--GHIJ--NOP-RSTUVWX	-BCDEFGH---LMNO-	ABCDEFGHIJK---O---S
AsnC	-BCDEFGH-JKLMNOP	ABCD---HI---NOPQRSTUVWXYZ	-BCD-----N--	--C----H-JKL--O---S
BirA	--C-E-GH-JK-MN--	ABC--GHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCDEFGHIJKLMNOP-R-
Fur	A-C-E-G-----OP	ABCD--GHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCD-FGHIJK---O-Q-S
LysR	-B--E--HI-K-----	ABCD--GHIJ-L-NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	A--DEFGHIJK---O---S
MarR	ABCD--G--JK---OP	ABCD--HI--N--R-----	-BCDEFGH---L----	---DE--HIJ-L--O---S
Pho	-----FGH-JK-----	ABCD--G-I---NOP-R-TUVWX	-B-DEFGH---LMNO-	ABCD-F-HIJK---OPQ--
TetR/AcrR	A-CDEFG--JKL----	ABCD--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---LMNOP	AB-DEF-HIJKL--O---S
HypF	-----E--HI-K-M---	-----K--NOP-----	-----O-----	A-CD-F-----O-----
AslB	-B-D--G--JK-M---	-----G-----NOP--S-UV-X	---D-----O-----	-BC--F----K---O-----
WrbA	--C-E-G--J-----	-BC--G-I---NOP-R-TUVWX	-BC-----	-----O-----S
Crp	-----	ABCD--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---MNO-	ABCDEFGHIJKL--O--RS
GalR/LacI	-----	ABCDE-G--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	-B--E--H-JK---O---S
DnaA	-----	ABCDEFGHI--LMNOPQRSTUVWXYZ	-BCDEFGH--KLMNOP	ABCD-FGHIJKLMNOPS
Ebp	-----	ABCDEFGHIJ-LMNOPQRSTUVWXYZ	-BC-EF-H-----O-	ABC-----H--KL---QR-
DeoR	-----	AB-D--GHIJ--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	AB--E--H--K---O---S
LexA	-----	ABCD--G-I---NOPQRSTUVWXYZ	-BCDEFGH---L----	-BCD-F-HIJ-L--O-----
OmpR	-----	ABCDEFGHIJK--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	ABCDEFGHIJKL--OP--S
RpiR	-----I-----	AB-D--GHI--NOPQRST---X	-BCDEFGH---LMNO-	-B-----H--K---O-----
LuxR/UhpA	-----I-----	ABCD--GHI--NOPQRSTUVWXYZ	-BCDEFGH---LMNO-	---D-FGHIJKL-----S
IclR	-----F-----	AB-D--G-I---NOPQRS-UV-X	-BC-E-----	AB-----H-JK-----S
HipB	-----K-----	AB-D--G-I---NOP-R-T----	-B-DEF-H--L----	-B--E--H--K---O-----
Cold	-----F-----	ABCDEFGHI--LMNOPQRSTUVWXYZ	--CDEFGH---LMNO-	ABC-E--HIJK---O-Q-S
MerR	-----E-----	ABCD--G-IJK--NOPQRSTUVWXYZ	-BCDEFGH---LMN--	A--DEFGHIJKL--O---S
GntR	--CD-----O-----	ABCD--GHI--MNOPQRSTUVWXYZ	-B-DEFGH---LMNO-	-B-DEFGH-JKL--O---S
Ihf	-----OP-----	ABCDEFGHIJKLMNOPS	-BCDEFGHIJKL-NOP	ABCDEFGHIJKLMNOPS
ArgR	-----	-----NOPQ-ST---X	-BCDEFGH---LMNO-	-BC----HIJKL-NO---S
NadR	-----	-----NOPQ-S---X	-----G-----	-----J-----
DniR	-----	A-----HIJK--NOP-R-TUVWX	--CDE--H--L--O-	A-C-----L--O--R-
MetJ	-----	-----NOPQ-ST---X	-----	-----
Crl	-----	-----NOP--T---X	-----	-----
AlpA	-----	-----N-P---T-VWX	-----	-----
FhlD	-----	-----I---NOP	-----	-----
FhlC	-----	-----I---NOP-----X	-----	-----
DicA	-----	-----NOP	-----	-----
Caif	-----	-----NOP	-----	-----
HycA	-----	-----NOP	-----	-----
HtgA	-----	-----N-----	-----	-----
DicC	-----	-----N-----	-----	-----
GcvR	-----	--C-----NOP-R-TUVWX	-----	-----
GutM	-----	-----NOP--S-----	-----E-----	-----
HipA	-----	--CD---I--N-----UV--	-----	-----L-----
TrpR	-----	-----NOPQ-ST--WX	-----	-----Q-----
FrvR	-----	-----N-----	--C--F-H-----	-----
ModE	---DE-----	A-CD--G-I--NOPQRS---X	-----	--C-----
NagR	-----	A-----G-----N-P-ST---X	-B--EF-H---M-O-	-B-----

Distribution des facteurs de transcription dans les génomes de bactéries et d'archées

*Certains sont universels,
d'autres sont spécifiques
à un groupe d'espèces*

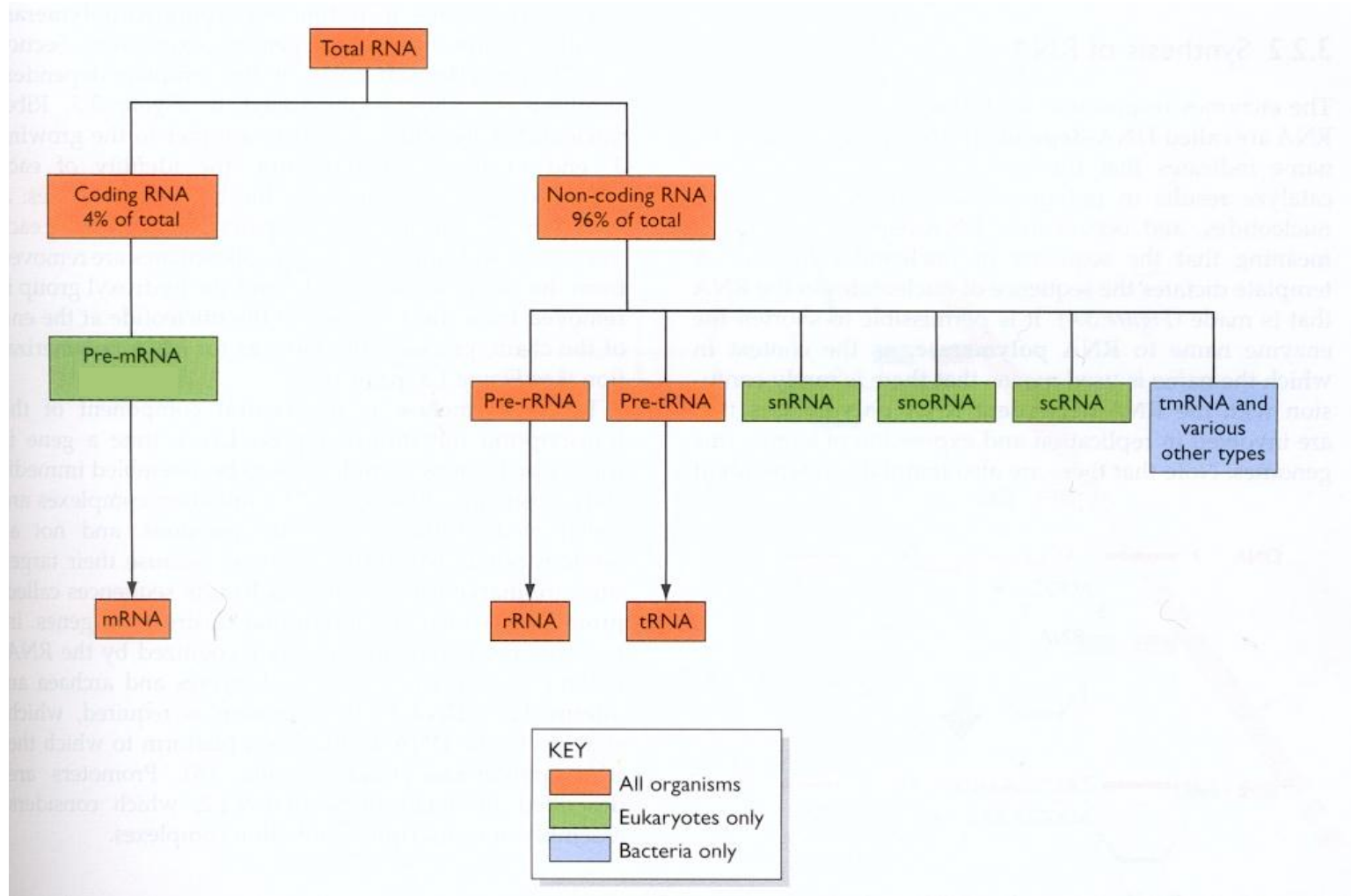
Faire parler les génomes

4- Les gènes codant des ARN



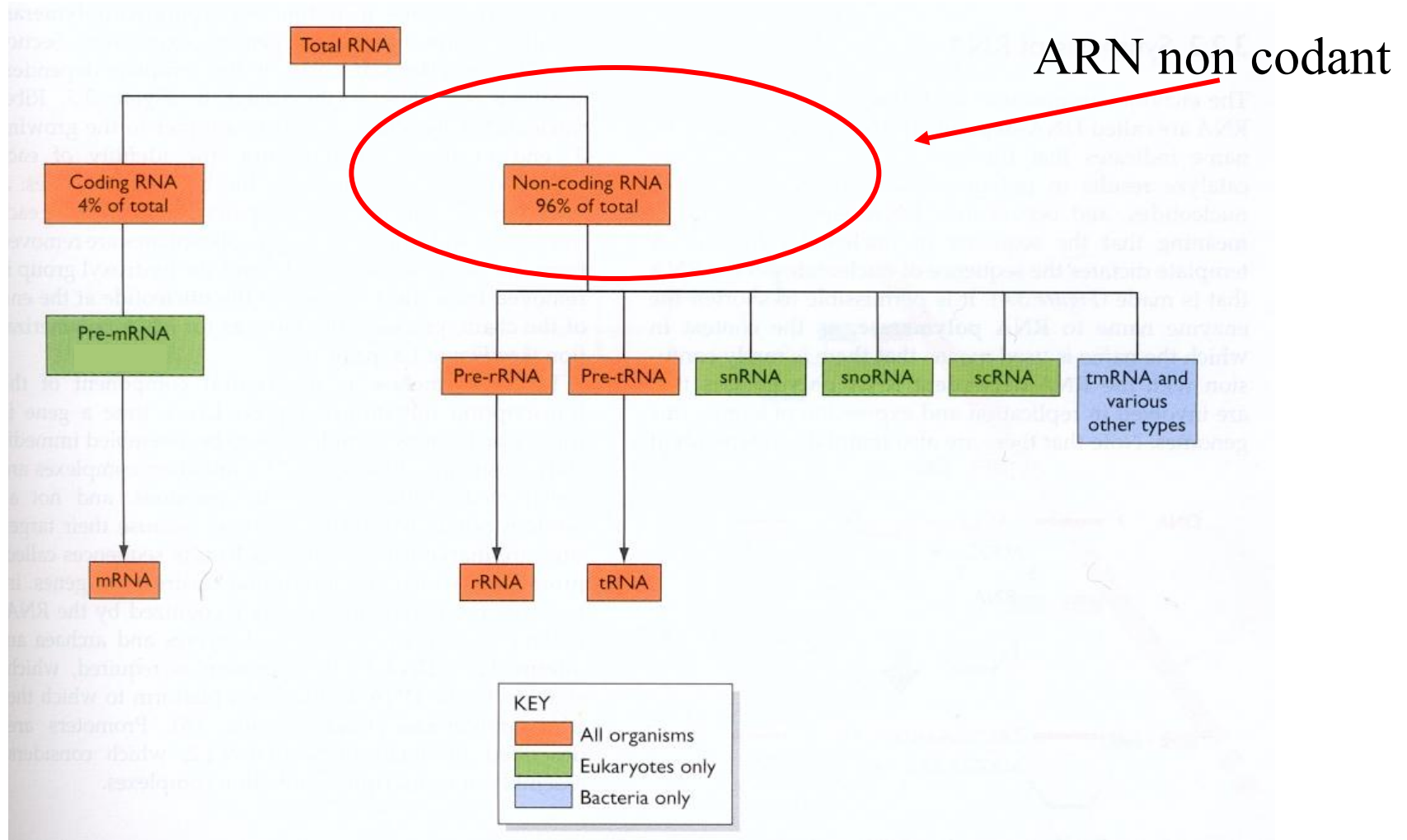
http://www.sanger.ac.uk/cgi-bin/Rfam/genome_dist.pl

4-1 Répartition des différents ARN dans le monde vivant



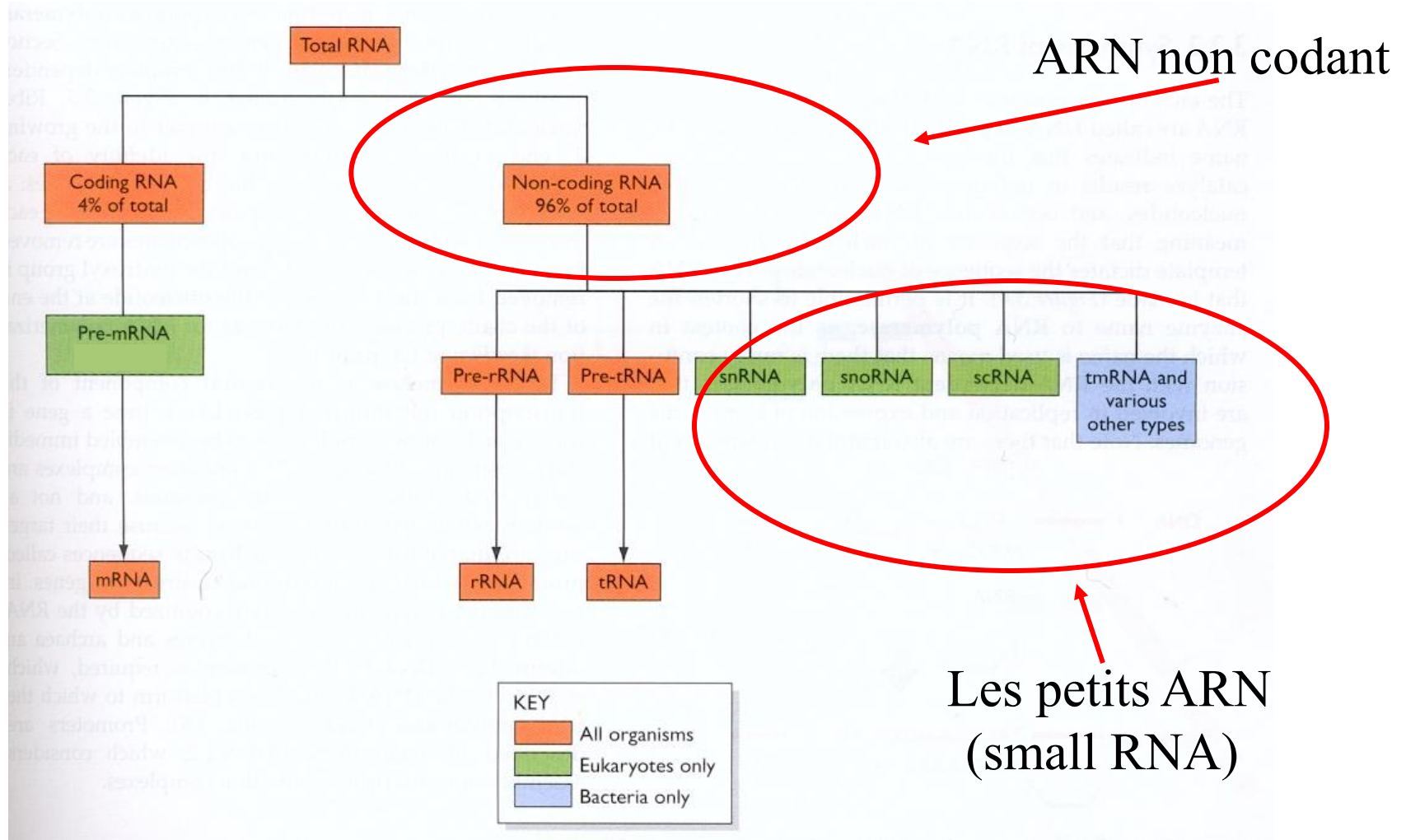
d'après Brown, 2002, in *genomes*)

4-1 Répartition des différents ARN dans le monde vivant



d'après Brown, 2002, in *genomes*)

4-1 Répartition des différents ARN dans le monde vivant



d'après Brown, 2002, in *genomes*)

4-2 Les gènes codant des petits ARN (small RNA genes) sont difficiles à identifier de manière expérimentale et informatique

- **Petite taille**
- **Pas traduit en protéine**
- **Mutants pas repérables (Phénotype ?)**
- **Jusqu'en 2000, seulement 10 gènes sRNA d'*E.coli* connus**
- **Entre 2000 et 2003 : 45 nouveaux gènes**

Comment essayer de les identifier ?

- **Analyse de transcriptome par micro-arrays
d'oligo-nucléotides dans les régions intergéniques (protéines)**
Tjaden B. *et al*, 2002, NAR, vol 30, p 3732
- **Génomique comparative : séquences conservées entre espèces
proches dans des régions apparemment sans gène**
Rivas E. *et al*, 2001, Current Biol, 11, p 1369
- **Biais dans la composition des séquences (%GC riche)**

*4-3 Les gènes codant
des petits ARN
(small RNA genes) chez E. coli*

Hershberg R *et al*, 2003, NAR, vol 31, p1813

4-3-1

Localisation des gènes sRNA chez *E.coli*

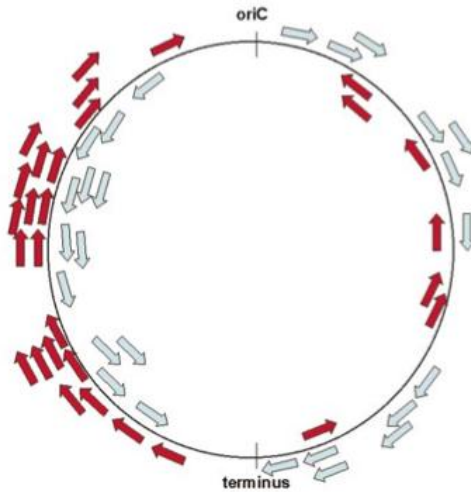


Figure 1. Distribution of the sRNA-encoding genes along the *E.coli* genome. The origin and terminus of replication are indicated. sRNA genes on the leading and lagging strands are colored blue and red, respectively.

Longueur des gènes sRNA

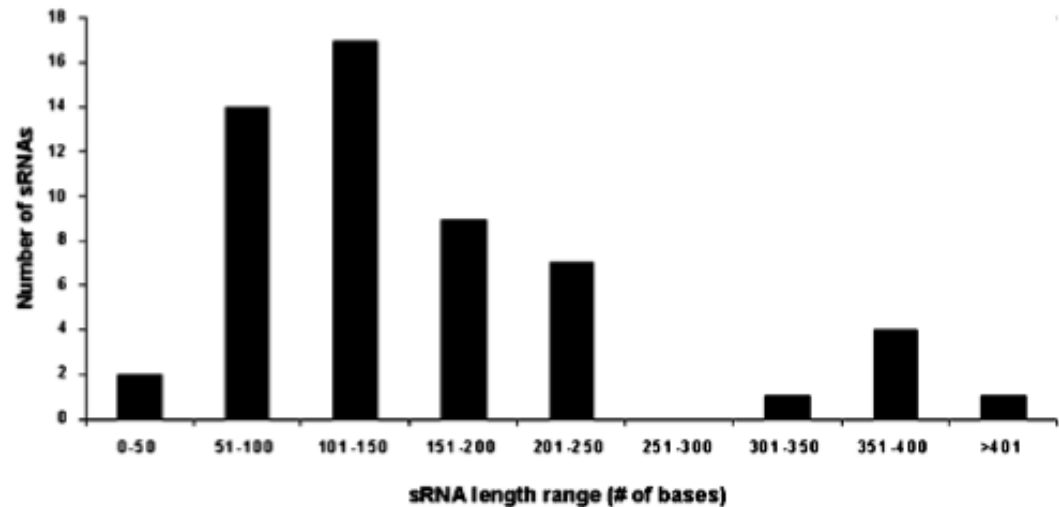


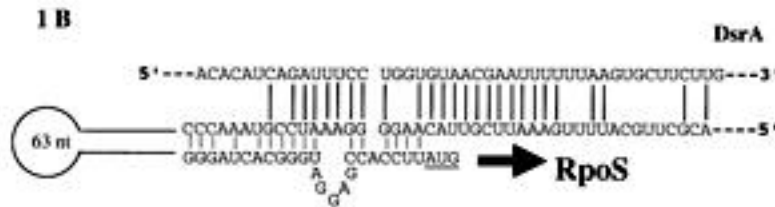
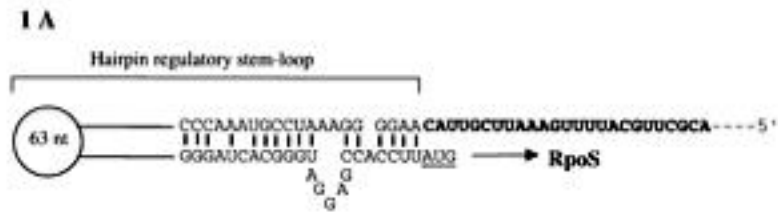
Figure 2. Length distribution of the 55 known sRNA genes.

4-3-2

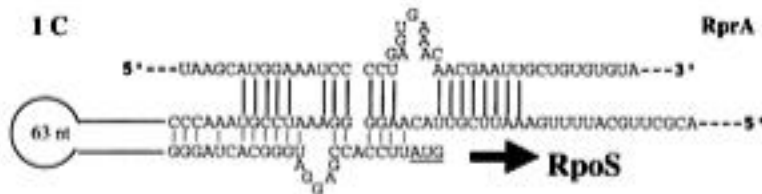
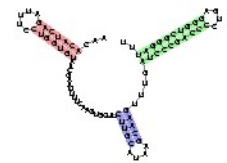
Fonctions déterminées pour 13 des 55 gènes sRNA identifiés

- Appariements avec d'autres ARN
pour réguler leur stabilité ou traduction**
- Liaison avec des protéines
pour réguler leur activité**

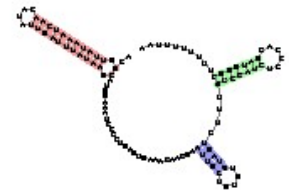
La régulation du gène RpoS (facteur σ) par trois sRNA : Réponse à l'environnement



DsrA



RprA



OxyS



Région 5' du gène RpoS

La régulation du gène RpoS par trois sRNA : Réponse à l'environnement

sRNA and environment 859

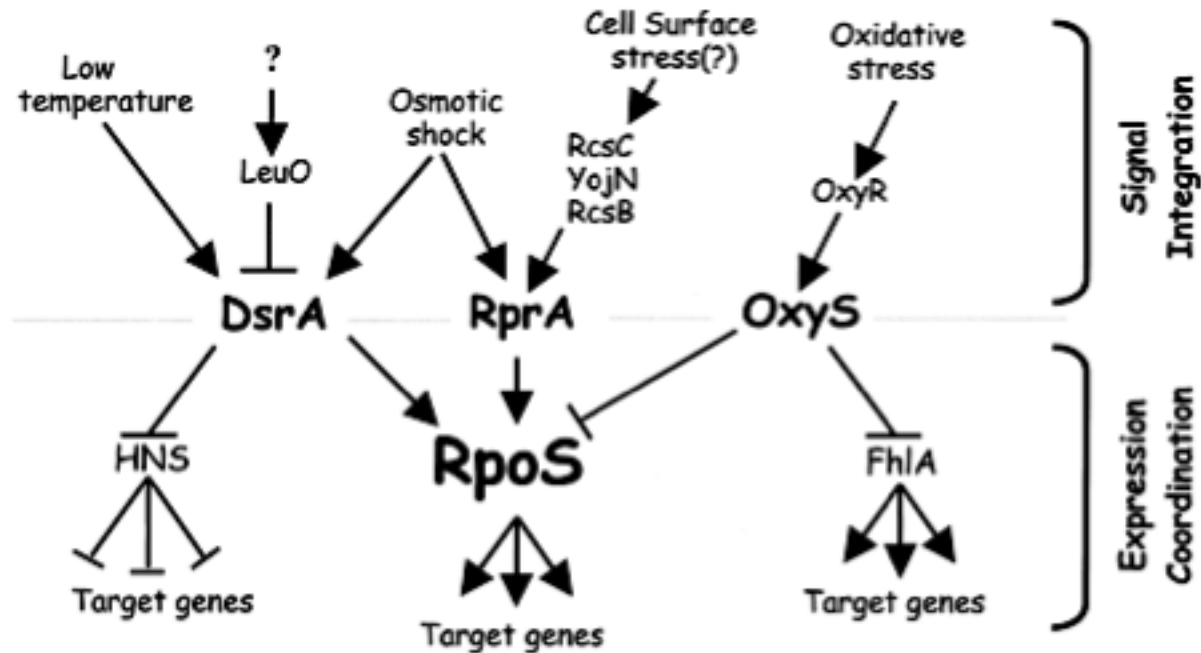


Fig. 2. Environmental signals shown above the line act either directly or through the signalling proteins shown to stimulate levels and activities of the three small RNAs that, in turn, affect RpoS translation. Other known targets of DsrA and OxyS are shown; additional ones may exist, as well as other targets for RprA.

La régulation du gène RpoS par trois sRNA : Réponse à l'environnement

sRNA and environment 859

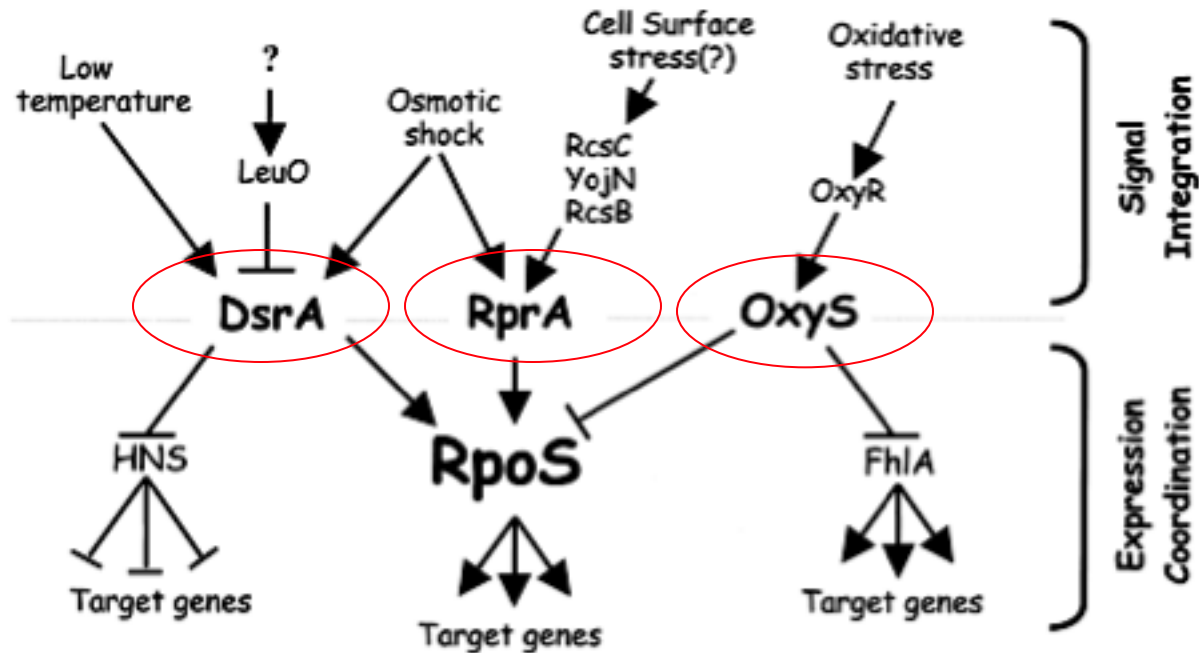


Fig. 2. Environmental signals shown above the line act either directly or through the signalling proteins shown to stimulate levels and activities of the three small RNAs that, in turn, affect RpoS translation. Other known targets of DsrA and OxyS are shown; additional ones may exist, as well as other targets for RprA.

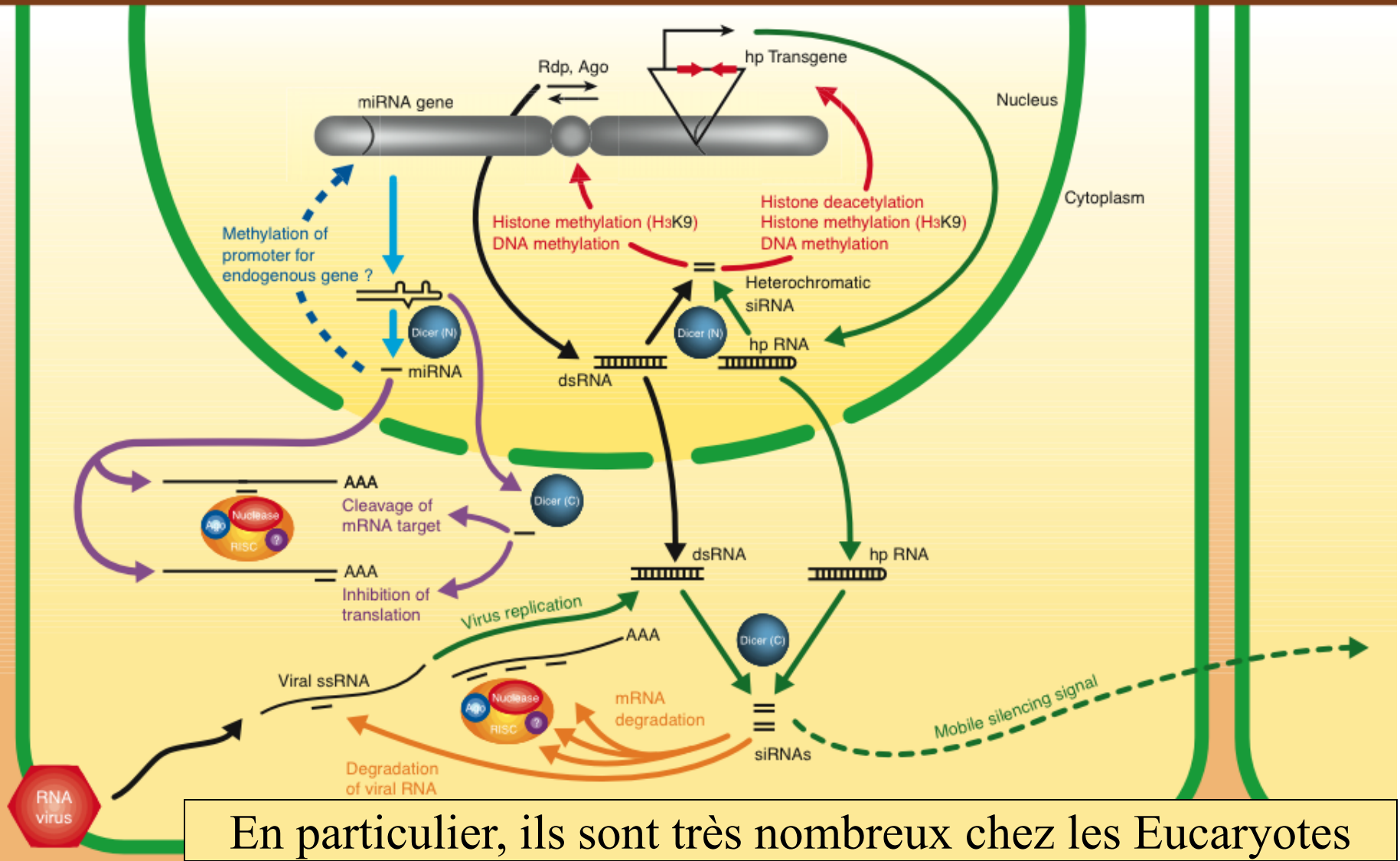
4-4 Conservation des gènes sRNA et des gènes adjacents dans différentes espèces bactériennes

4-5 Les ARN non codant sont universels

ARN non codants (taille en nucléotides)	Fonction cellulaire	Mécanismes
Cible : ADN		
ARN télomérique (400 à 500)	Réplication	Synthèse des télomères
ARN Xist (17000)	Extinction de gènes	Inactivation du chromosome X
Cible : ARN		
ARNase P (370)	Maturation de l'ARN	Clivage de l'extrémité 5' des ARNt
snARN (100 à 300)	Maturation de l'ARN	Epissage des introns de pré-ARNm
snoARN à boîtes C/D et H/ACA (70 à 350)	Maturation de l'ARN	Clivages des préARNr et modifications de nucléotides
gARN (80)	Maturation de l'ARN	Édition de l'ARN
miARN/siARN (21 à 25)	Maturation de l'ARN	Dégradation de l'ARNm
miARN (21 à 25)	Traduction	Inhibition de la traduction
DsrA (90)	Traduction	Stimulation de la traduction
OxyS (100)	Traduction	Inhibition de la traduction
ARN-OUT (69)	Transposition	Régulation
Cible : Protéines		
ARNr, ARNt (70 à 5000)	Traduction	Décodage de l'information génétique
ARNtm (350)	Stabilité	Dégradation des protéines
ARN6S (21 à 25)	Transcription	Régulation de la transcription

The Small RNA World

E. Jean Finnegan and Marjori A. Matzke



En particulier, ils sont très nombreux chez les Eucaryotes

Conclusion

*De l'ADN «la molécule-reine»
au «small RNA world»*