

La cellule minimale

- 1/ Définition / Motivations
- 2/ Caractéristiques des cellules actuelles
- 3/ Approche expérimentale
- 4/ Approche bioinformatique

Céline Brochier

Laboratoire de chimie bactérienne

Université de Provence (Aix-Marseille I)

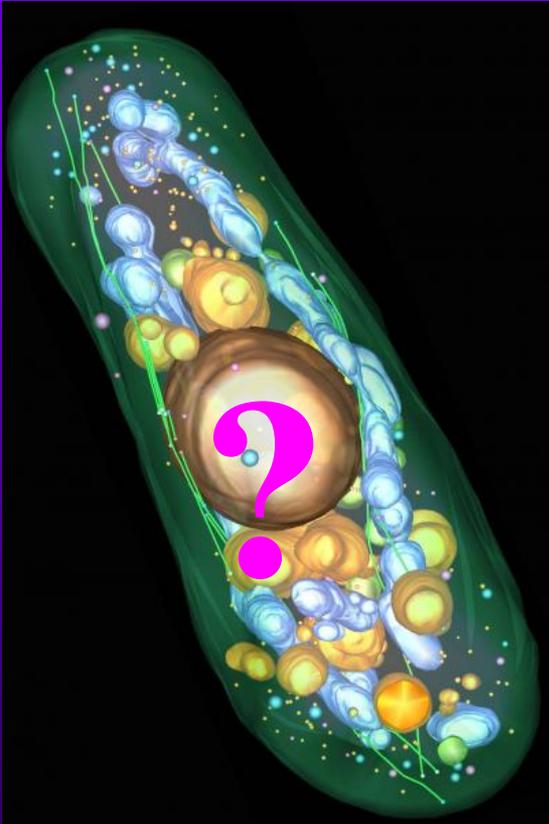
Celine.brochier@ifr88.cnrs-mrs.fr

La cellule minimale : Problématique



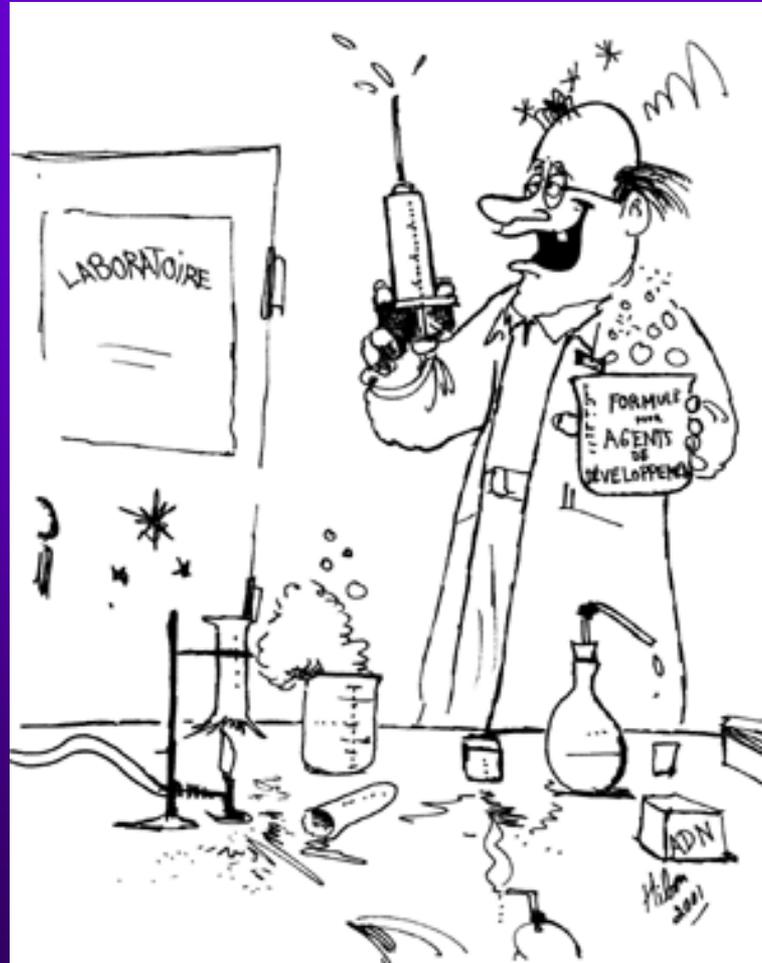
- *Qu'est-ce qui est nécessaire et suffisant pour permettre d'obtenir une cellule*

Motivations? biologiques



Comprendre ce qu'est une cellule
⇒ fonctionnement biologique

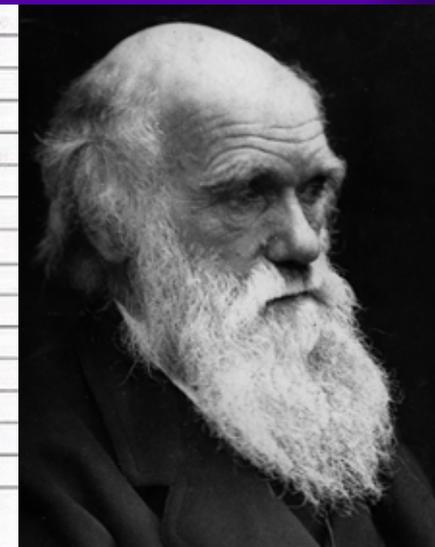
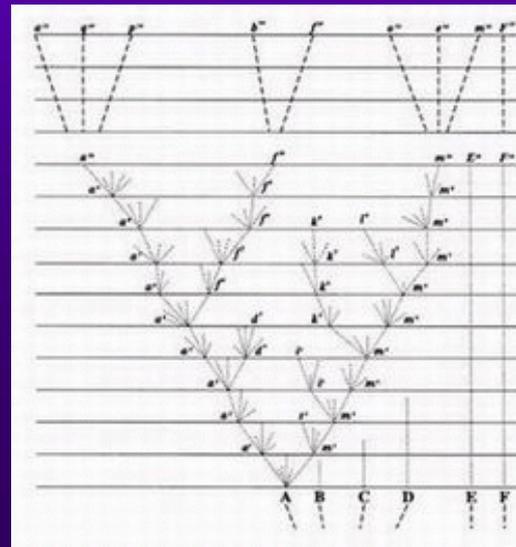
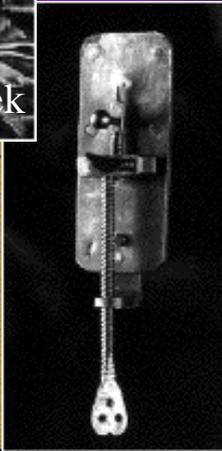
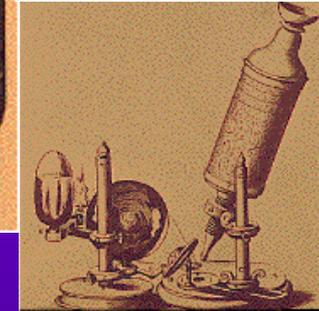
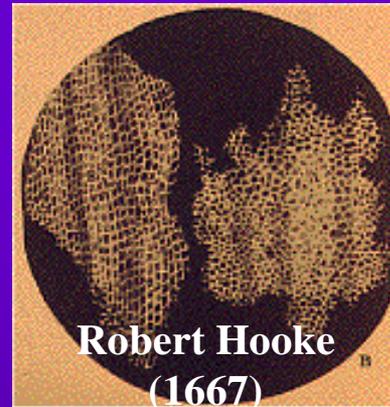
Motivations? biotechnologiques



Motivations? Évolutive

Théorie cellulaire

- Tous les êtres vivants (plantes, animaux, moisissures) sont constitués d'un ensemble d'unités de construction de même type, les cellules (*Matthias Schleiden et Theodor Schwann, 1839*)
- Toute cellule provient d'une autre cellule (*Robert Remak, Rudolf Virchow, 1855*)



Motivations? Exobiologique

- La cellule est l'unité d'organisation structurale des systèmes vivants
- Tous les êtres vivants sont constitués de cellules
- Chaque cellule possède les propriétés caractéristiques du vivant
 - *Un système qui est délimité sur le plan spatial par une membrane semi-perméable de sa propre fabrication, qui est capable de s'auto-entretenir (auto-conservation) et de reproduire un système ayant les même caractéristiques (auto-reproduction) en fabriquant ses constituants à partir d'énergie et/ou de matériaux bruts extérieurs (auto-régulation) (P.L. LUISI)*

La cellule minimale: Un problème simple ?



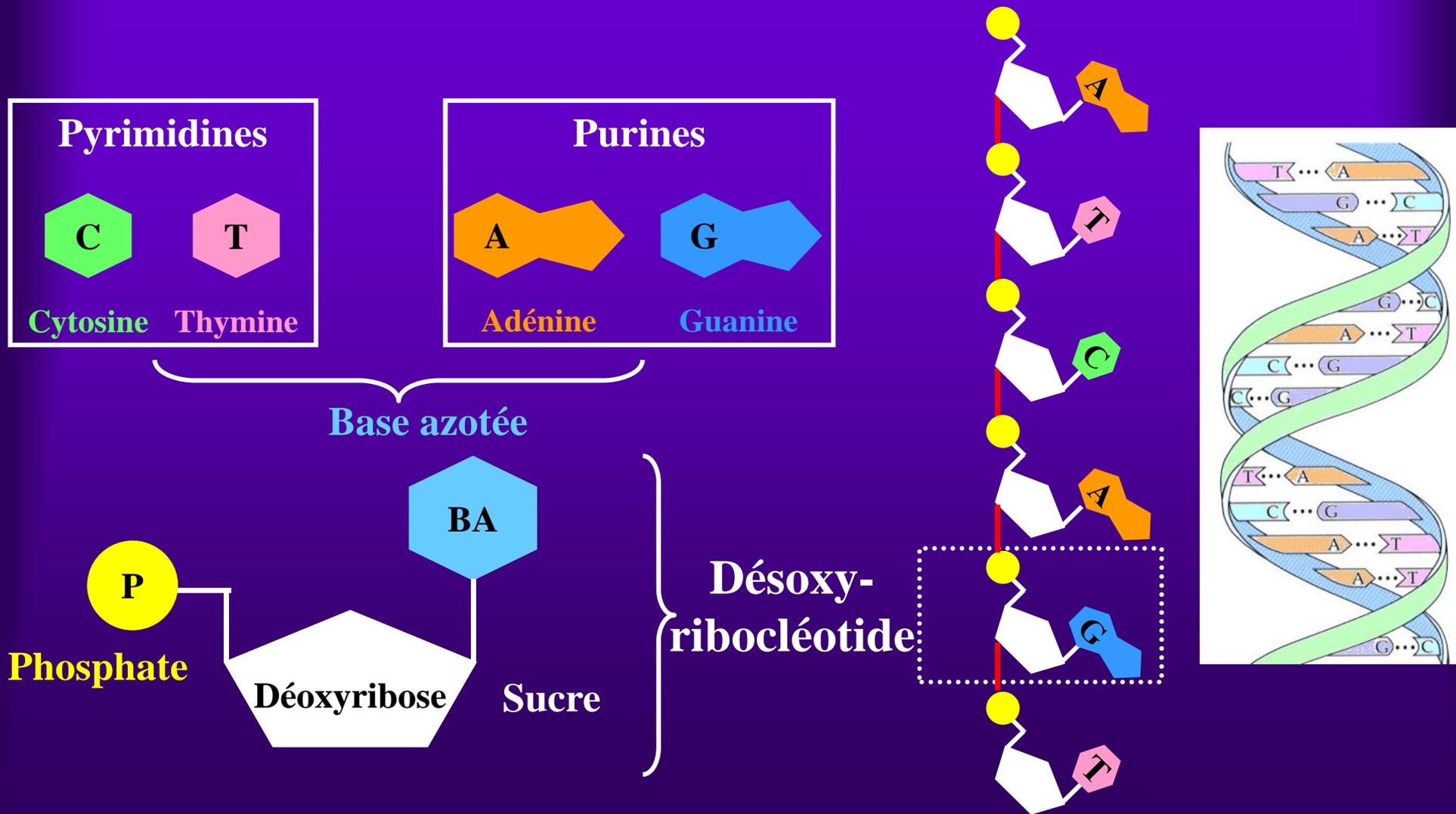
- ⇒ « *Qu'est-ce qui est nécessaire et suffisant pour maintenir une forme de vie cellulaire moderne* » ?
- Doit recouvrir AU MOINS l'ensemble des grandes fonctions communes à toutes les cellules
 - Qu'est-ce qu'une cellule ? et quelles sont ses fonctions / ses caractéristiques?

Biologie XXème siècle

Malgré leurs différences apparentes, toutes les cellules partagent un certain nombre de caractéristiques.

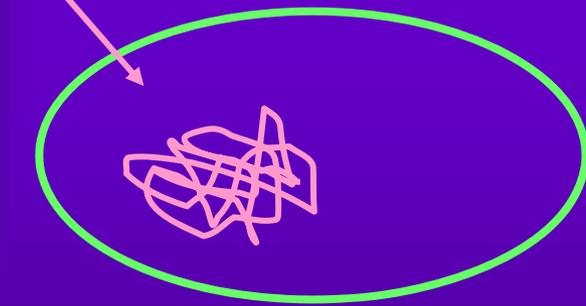


(1) L'ADN support de l'information génétique

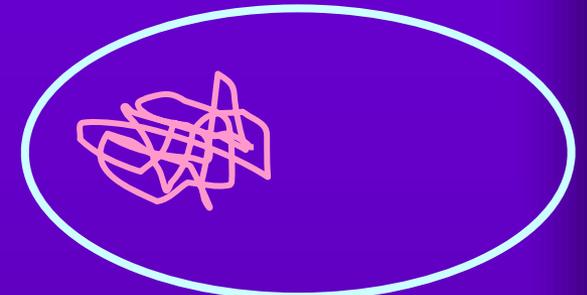


(2) Transmission de l'information génétique

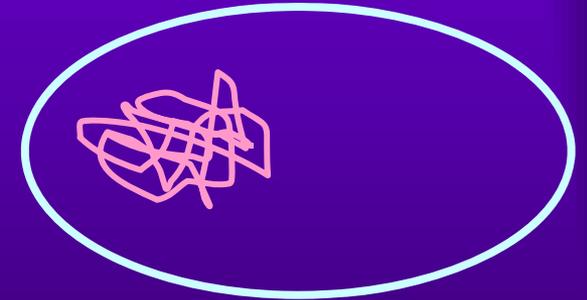
Information génétique
(ADN)



Cellule mère



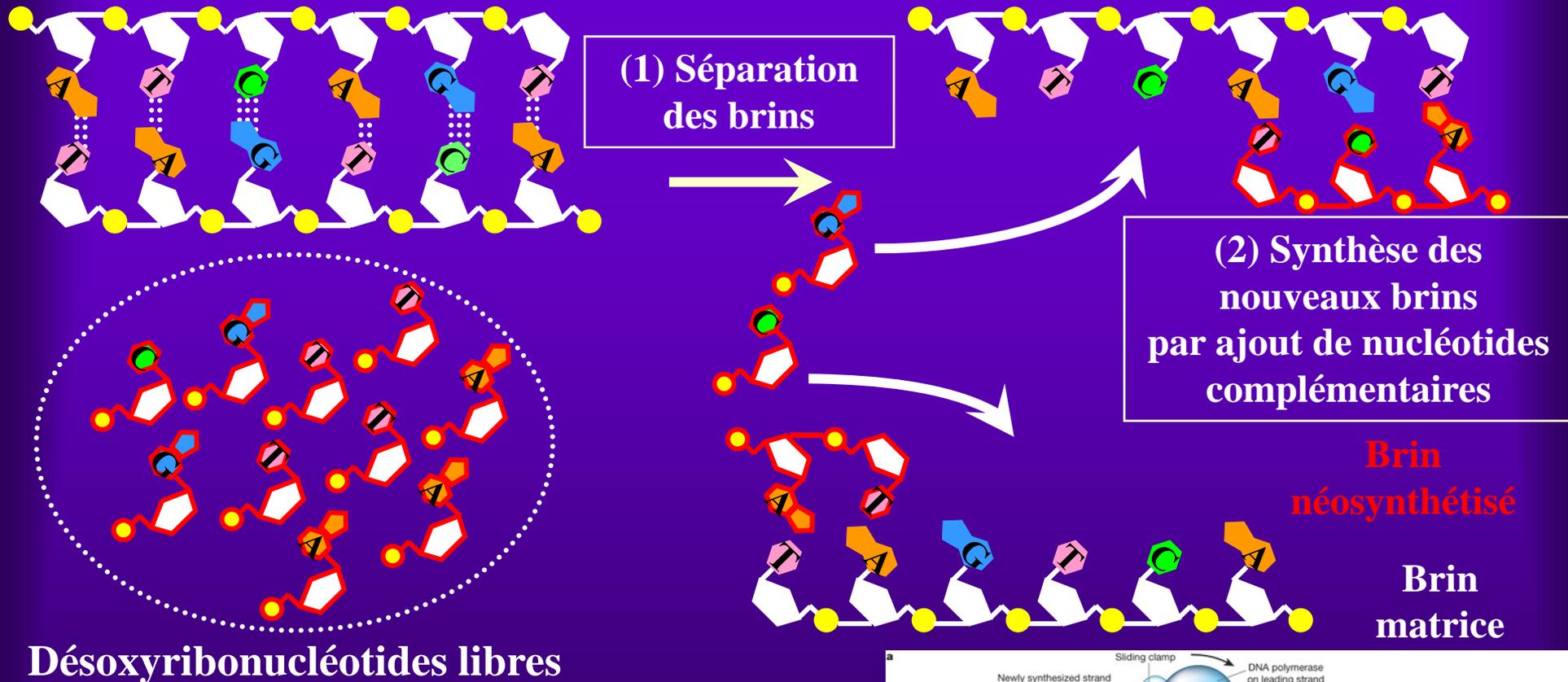
Cellules filles



Lors de la reproduction cellulaire, l'information de la cellule mère est transmise intégralement aux deux cellules filles

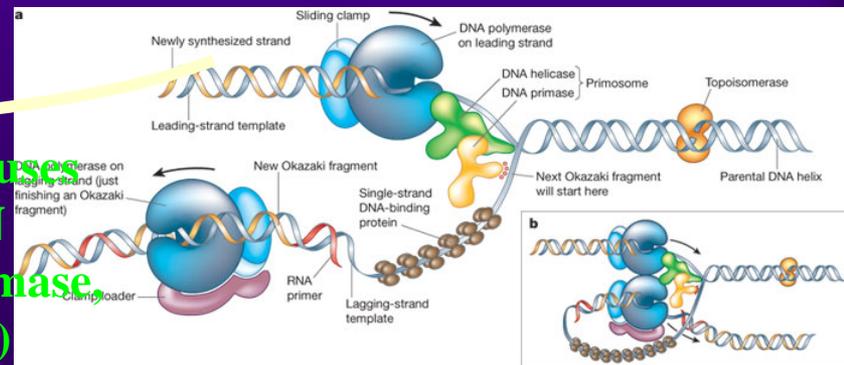
=> Cette information génétique est dupliquée avant la division cellulaire, c'est l'étape de **REPLICATION**

(3) Réplication de l'ADN

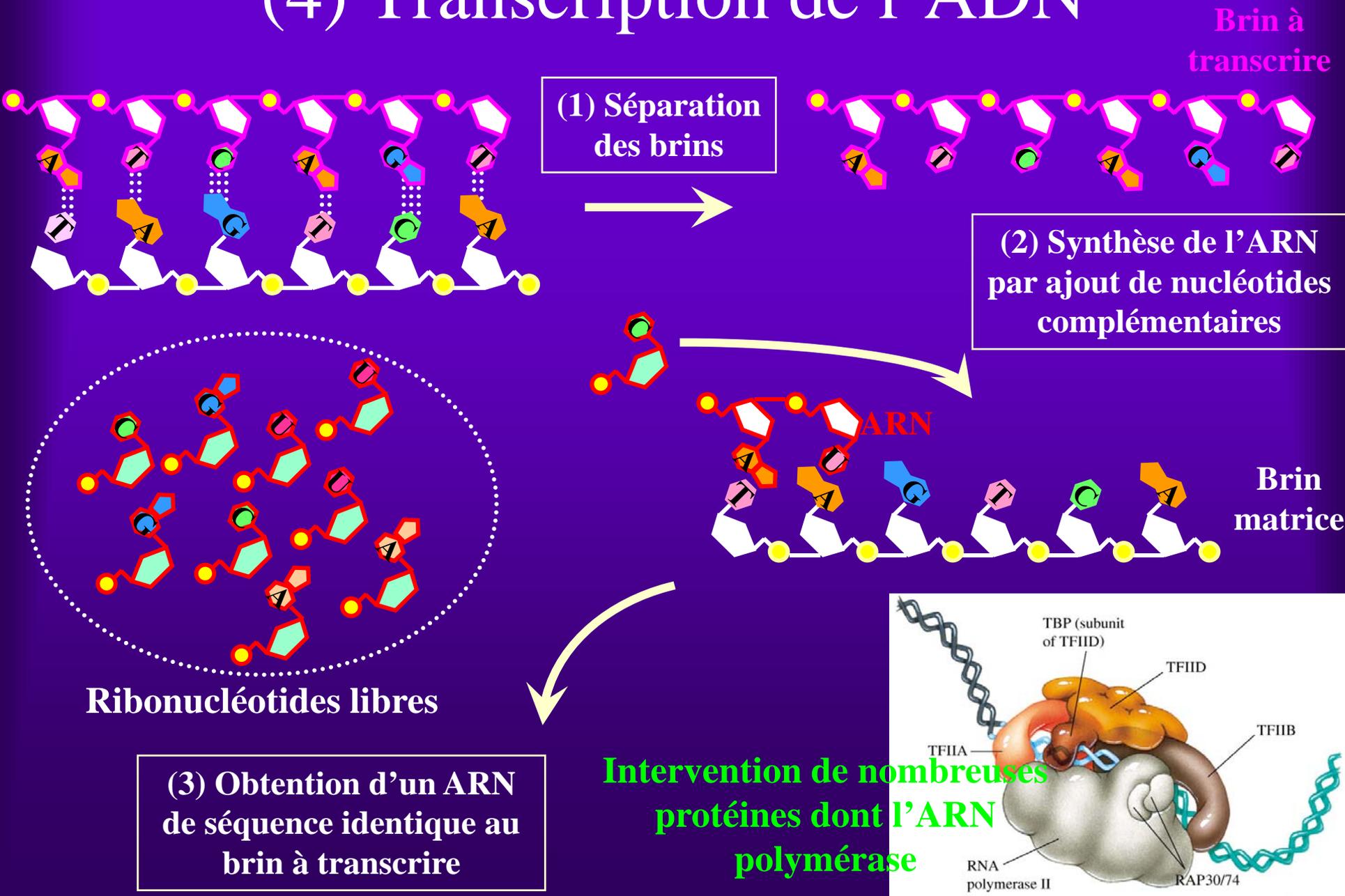


(3) Obtention de deux molécules d'ADN identiques et identiques à l'ADN initial

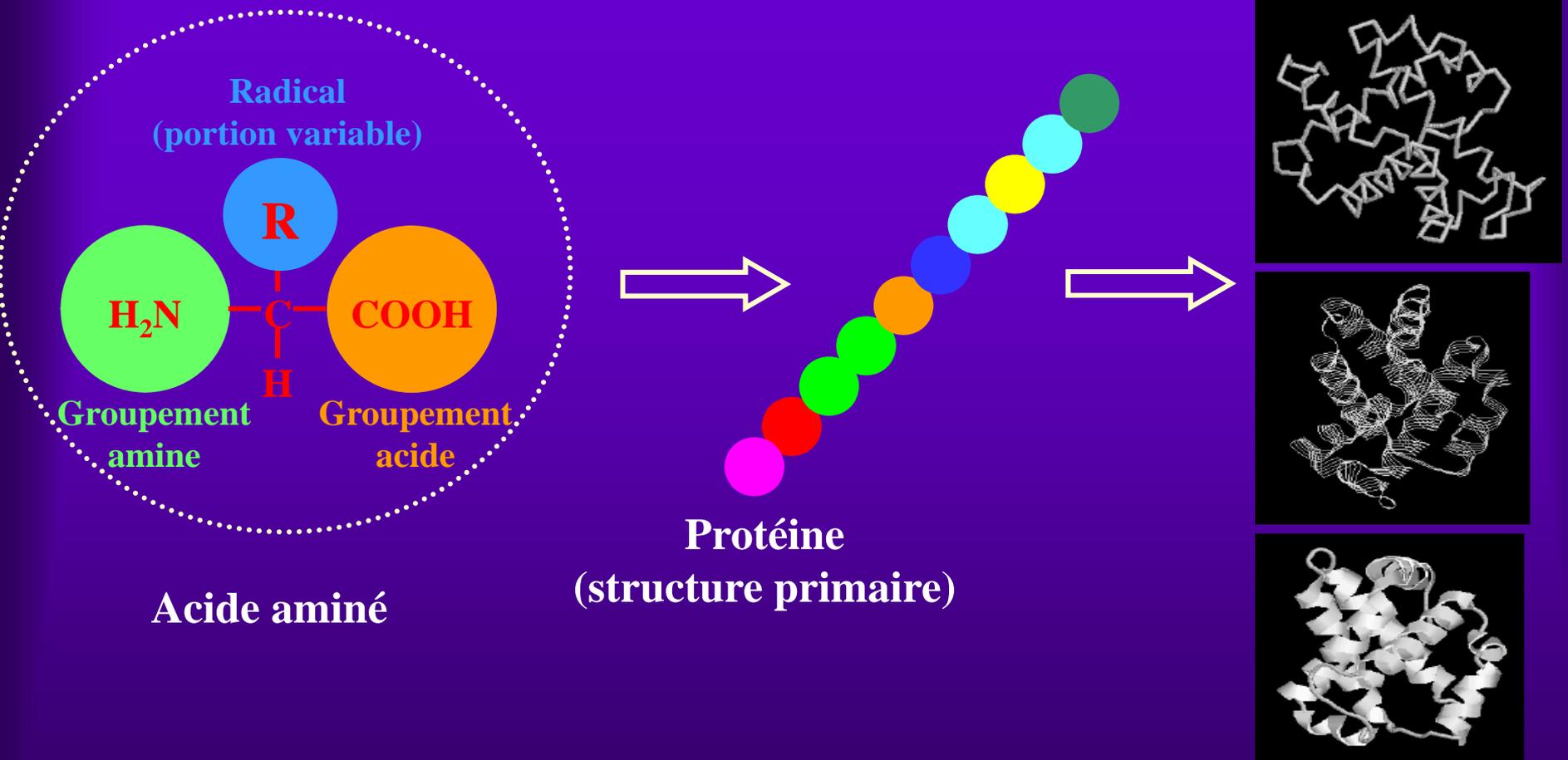
Intervention de nombreuses protéines dont (ADN polymérase, hélicase, primase, ligase, clamp, SSB...)



(4) Transcription de l'ADN



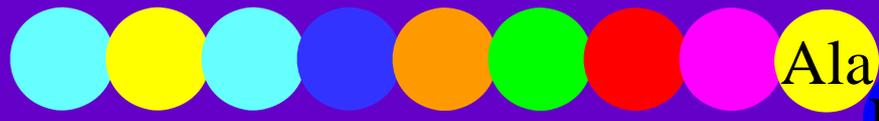
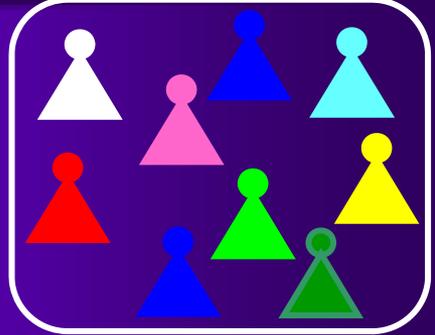
(5) Traduction de l'ARN en protéine



Agents indispensables (catalysent des réactions chimiques, maintient des structures, génération des mouvements, détection des mouvements, réplication, transcription, traduction, régulation de l'expression des gènes...)

(5) Traduction de l'ARN en protéine

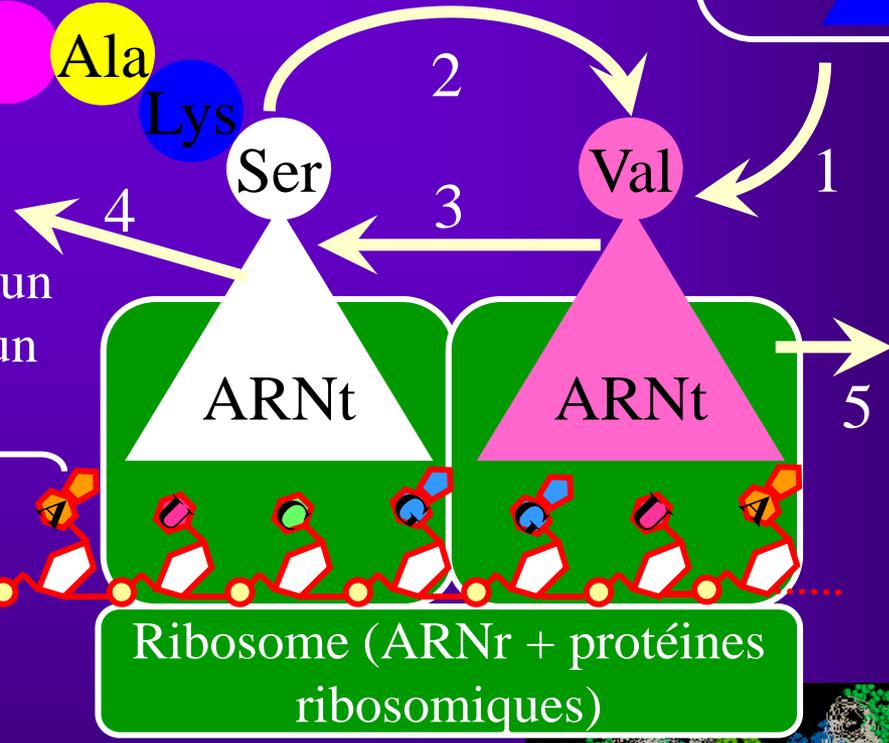
Acides aminés liés à des ARNt



Protéine en cours de synthèse

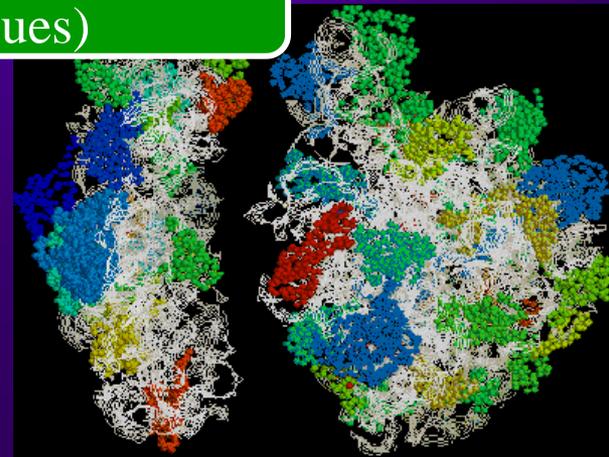
Changement d'alphabet: Passage d'un alphabet à 4 lettres (A, U, G, C) à un alphabet à 20 lettres

ARN



	Gly	Phe	Leu	
Asp	Glu	TCA	GTC	Ser
Ala	GTC	TCA	GTC	Tyr
Val	GAC	TCA	TCA	ochre
Arg	GAT	TCA	TCA	amber
Ser	GAT	TCA	TCA	Cys
Lys	GAC	TCA	TCA	opal
Asn	GAC	TCA	TCA	Trp
Thr	GAC	TCA	TCA	Leu
Met	GAC	TCA	TCA	Pro
Ile	GAC	TCA	TCA	His
Arg	GAC	TCA	TCA	Gln

Nécessite de nombreuses protéines + ARNr + ARNt...



(6) Énergie libre et matériaux bruts

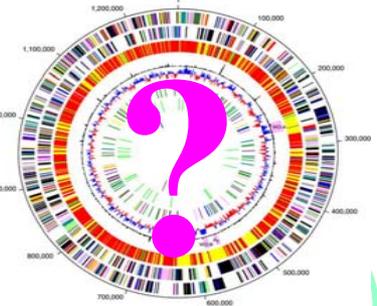
- Toutes les réactions réalisées (synthèse d'ADN, de protéines...) par la cellule nécessitent de l'énergie libre puisée dans l'environnement:
 - Source organique (organotrophe)
 - Source inorganique (lumière = phototrophe / chimique = lithotrophe)
 - Source de carbone
 - Source organique (hétérotrophe)
 - Source inorganique (autotrophe)
 - Toutes les cellules utilisent des petites molécules organiques (sucres, nucléotides, acides aminés, ATP...)
- ⇒ Capacités de fabrication variables selon les organismes

(7) Délimitation spatiale des cellules

- Membrane plasmique = barrière sélective (diffusion ou transport actif)
- Protection, Transport (importation/concentration des nutriments recueillis dans l'environnement et excrétion des déchets), Sensoriel...
- Composition = phospholipides + protéines membranaires
- Fabrication et maintenance de la membrane (métabolisme des lipides), fabrication des protéines membranaires, des pores...

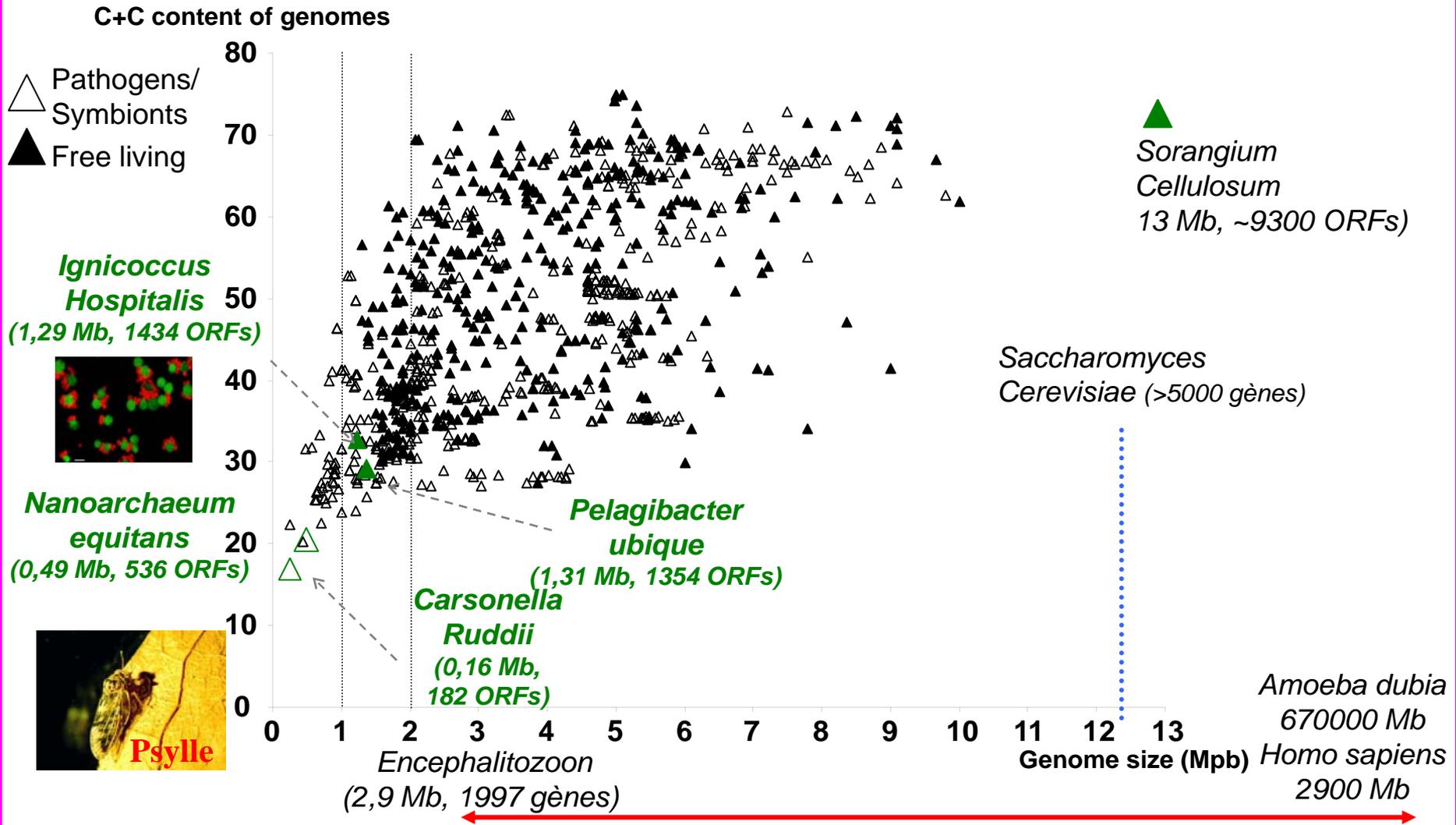


La cellule minimale: Un problème simple ?



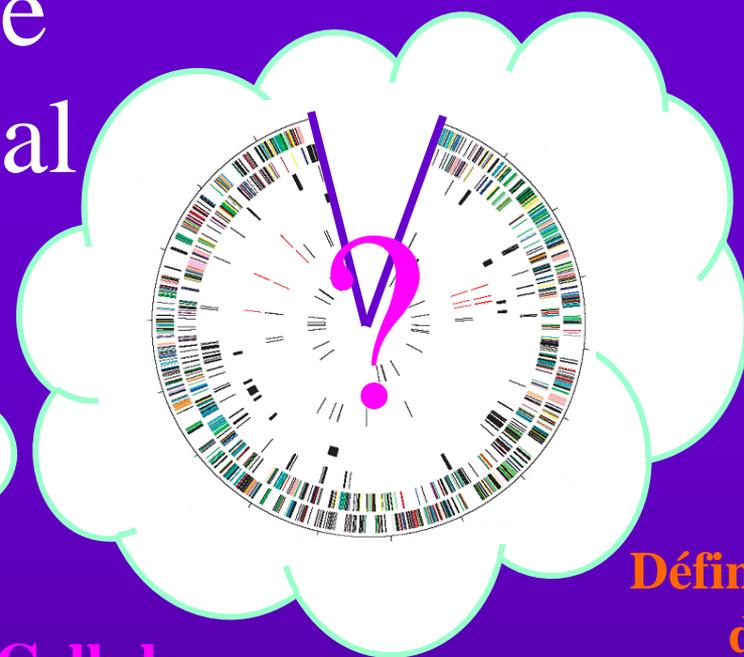
- Caractéristiques communes des cellules
↔ **quelle information génétique?**
- « *Quel est le plus **petit set de gènes** nécessaires et suffisants pour maintenir une forme de vie cellulaire moderne* »
- Doit recouvrir AU MOINS l'ensemble des grandes fonctions communes à toutes les cellules => **Combien de gènes?**

Génomes minimaux actuels



La cellule minimale

↔ Génome minimal

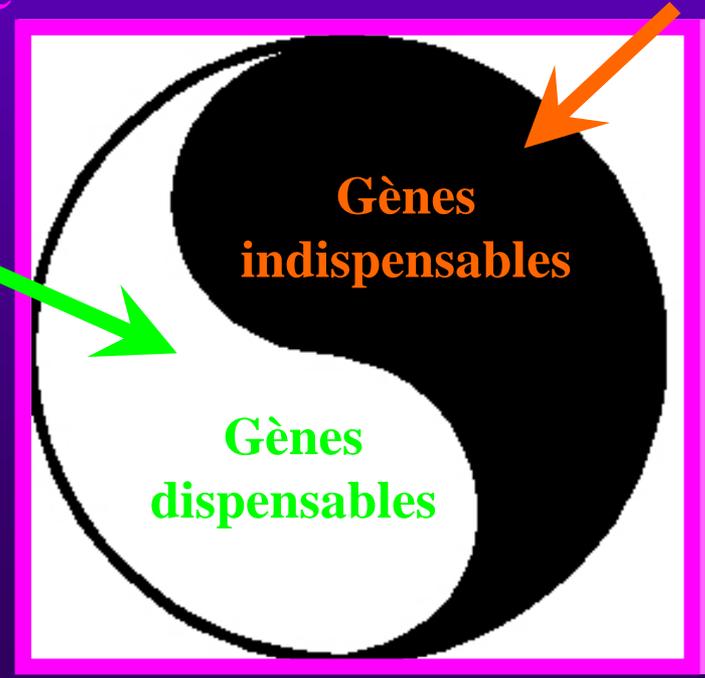


Définissent l'essence de la cellule

Cellule



Autres gènes

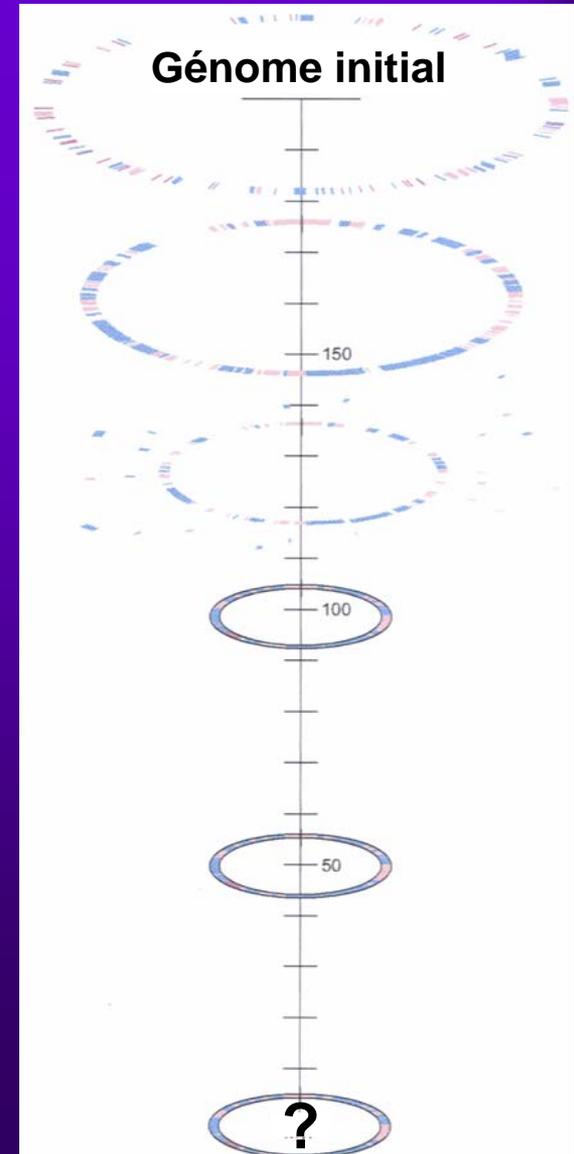
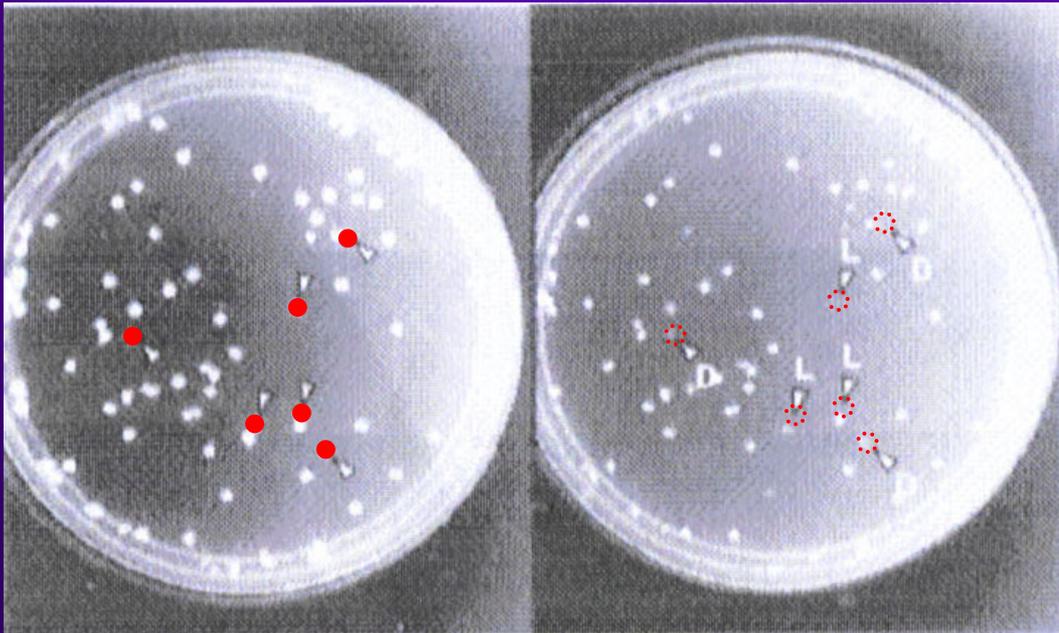


Gènes indispensables

Gènes dispensables

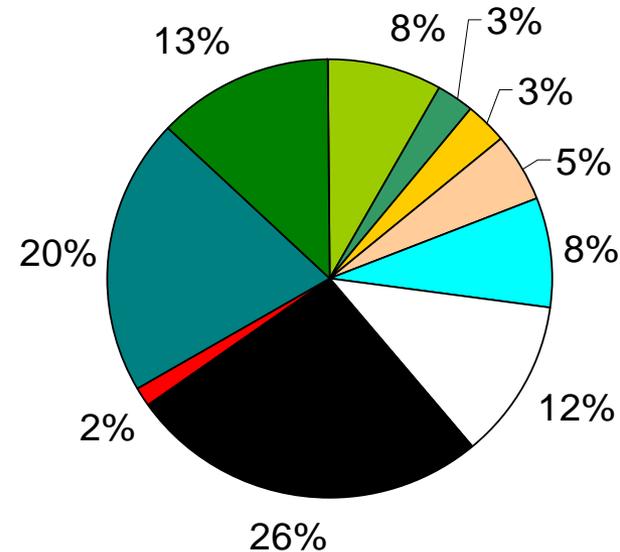
Approche expérimentale (1960<)

- Inactivation des gènes dans un génome un par un (mutagenèse)
=> Conséquences sur le phénotype
 - **Létal (pas de croissance)** \Leftrightarrow gène indispensable
 - **Ralentissement de la croissance**
 - **Croissance normale** \Leftrightarrow gène non indispensable



Modèles

- Bactéries
 - Génomes de petite taille
 - Organismes de petite taille
 - Organismes simples à manipuler
 - Génétique possible



Génome humain

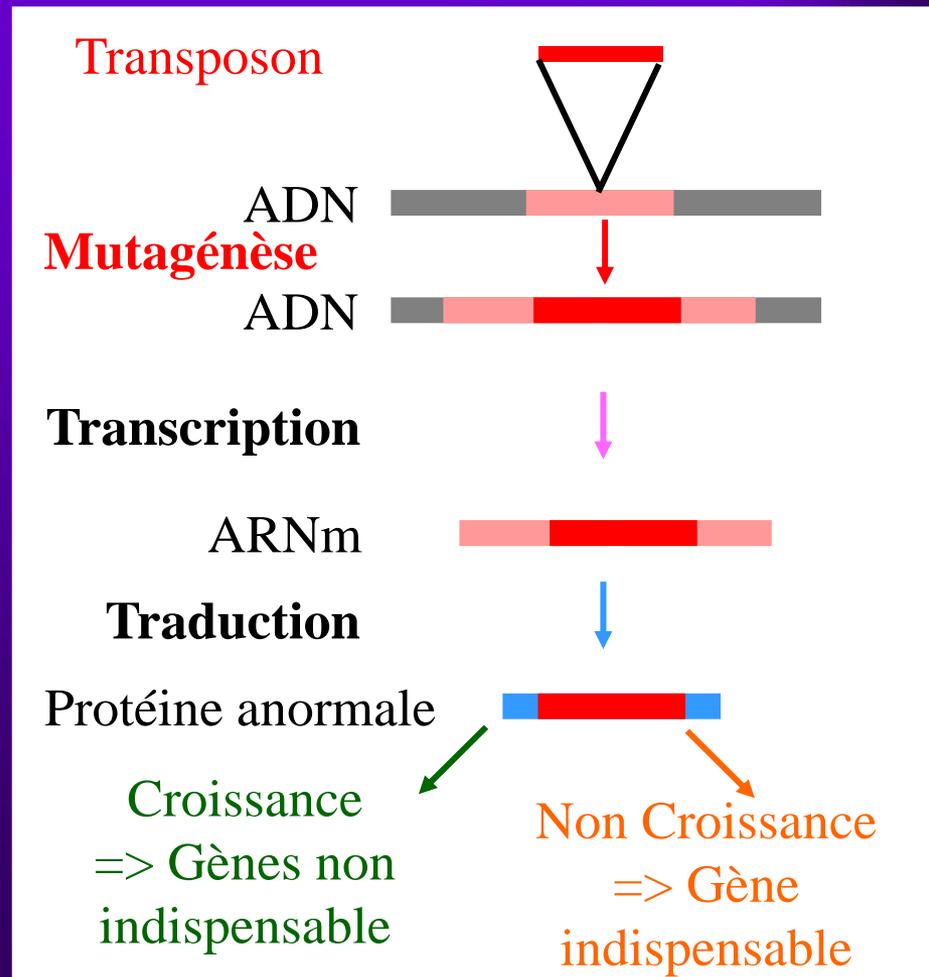
- | | |
|------------------------------|------------------------------------|
| ■ LTR retrotransposons | □ Misc. Séquences uniques |
| ■ DNAs retrotransposons | ■ Introns |
| ■ Séquences répétées simples | ■ Gènes codants pour des protéines |
| ■ Duplications segmentales | ■ LINEs |
| ■ Misc. Hétérochromatine | ■ SINEs |

Mycoplasma genitalium

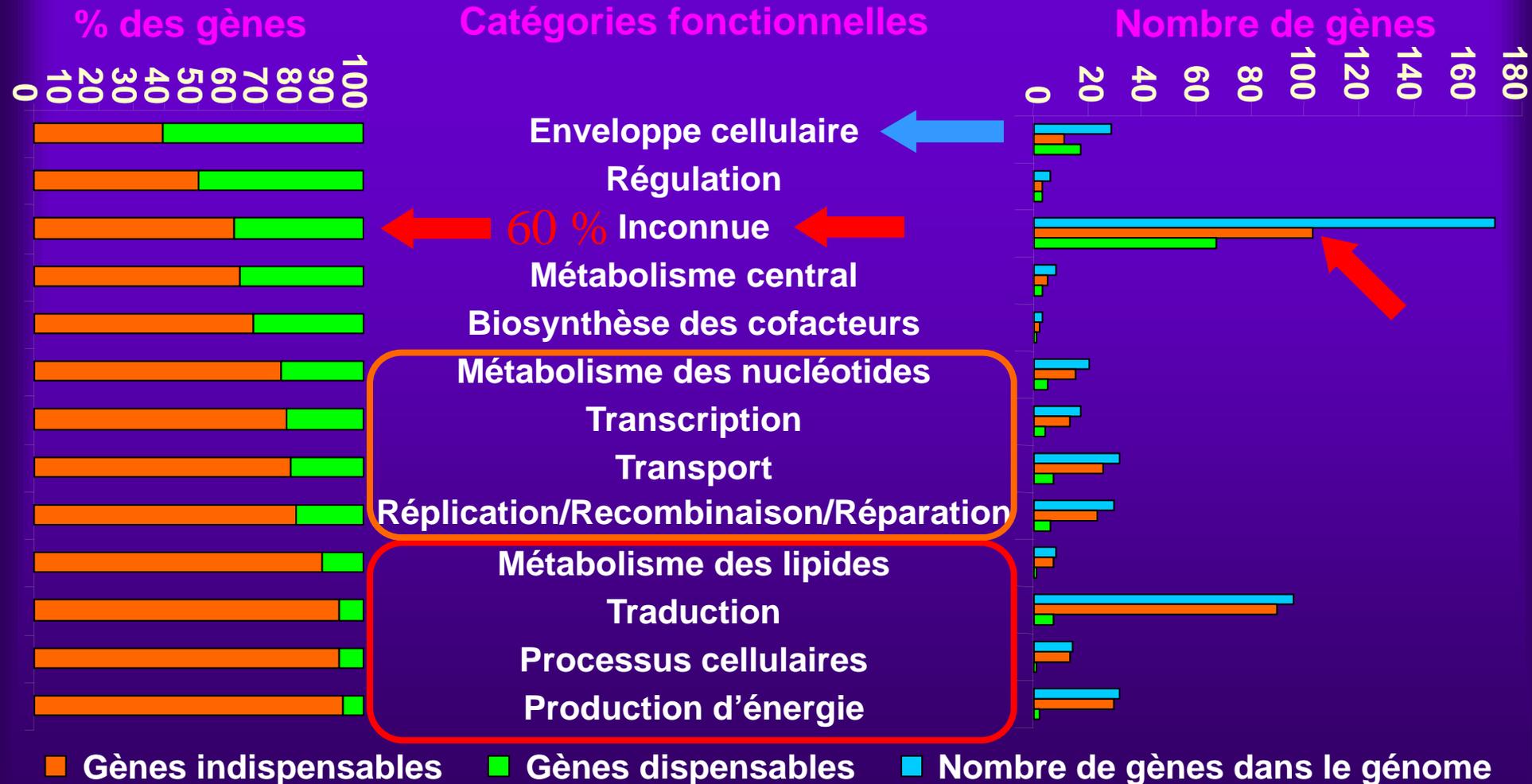


- Bactérie pathogène Gram +
 - Un des plus petits génomes cellulaires (480 gènes pour une taille de 540kb)
 - Génome supposé être composé principalement de gènes indispensables (ex. ~1/3 de ses gènes interviennent dans la fabrication des protéines)
- ⇒ Inactivation des gènes pas mutagenèse massives par des transposons

(Hutchison 1999)



Gènes dispensables / indispensables chez *M. genitalium*



180-215 gènes (1/3-1/2) seraient **NON INDISPENSABLES**

Le set de gènes indispensables est dépendant de l'environnement dans lequel on le définit

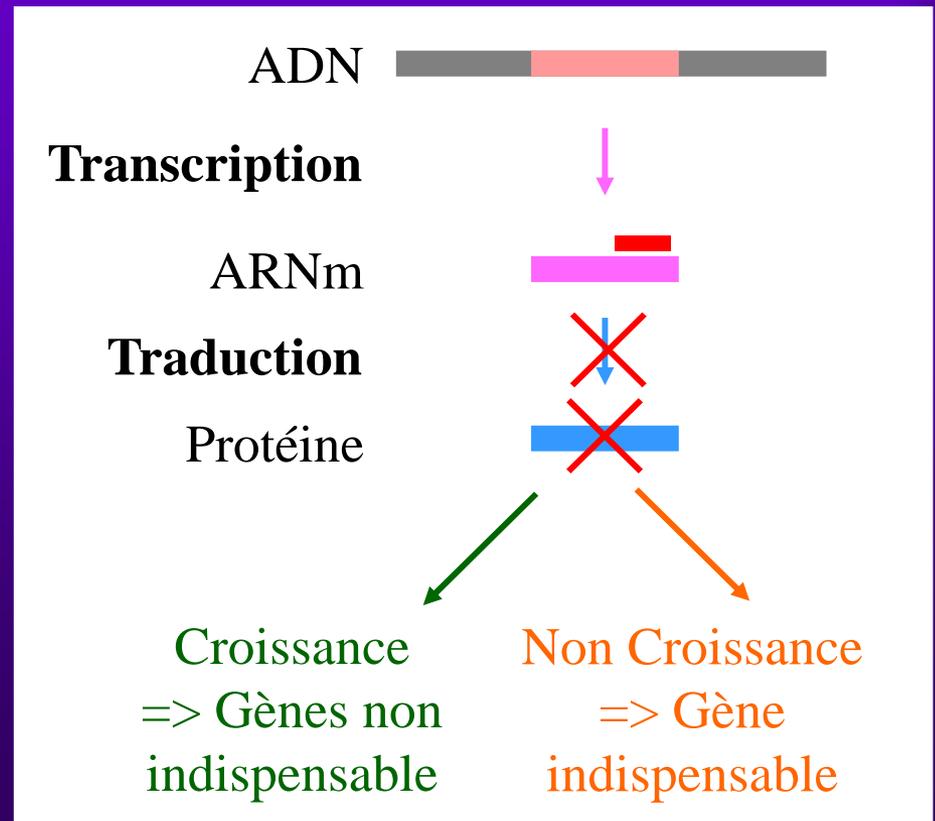
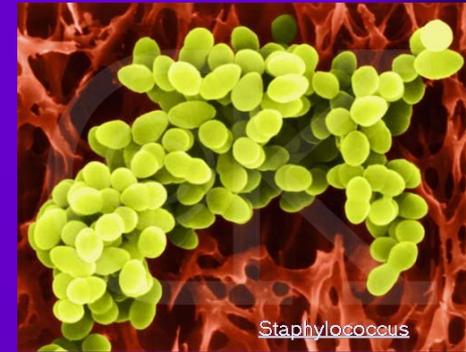
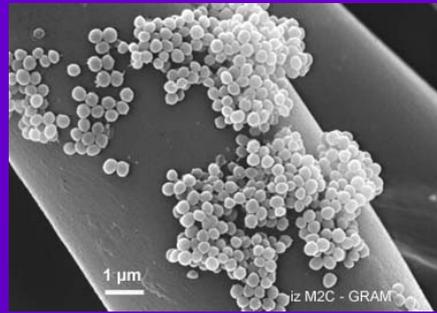
Staphylococcus aureus

- Bactérie pathogène Gram + à bas taux de G+C
- 2595 gènes

⇒ Inactivation des gènes par production d'ARN anti-sens

⇒ 658 induisent un phénotype létal (25%) ⇔ indispensables

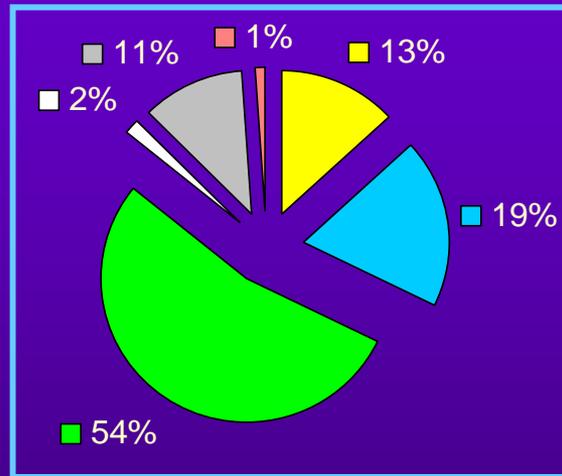
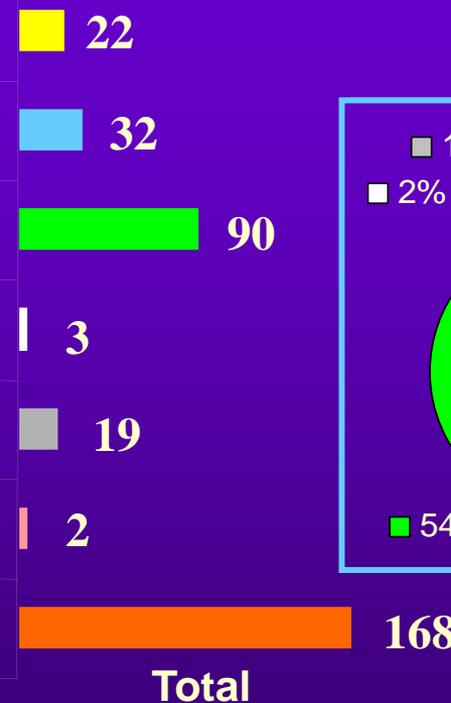
⇒ 168 sont aussi indispensables chez *M. genitalium*



Gènes indispensables chez *S. aureus* et chez *M. genitalium*

Nombres de gènes

0 50 100 150 200



■ **Enveloppe et processus cellulaires**
Paroi, Transport/liaison des protéines, Transduction du signal, Bioénergétique membranaire, Sécrétion des protéines, Division cellulaire, Sporulation

■ **Métabolisme intermédiaire**
Métabolisme des hydrates de carbones, des acides aminés, des nucléotides, des lipides, des coenzymes

■ **Processus informationnels**
Réplication de l'ADN, Modification/Réparation de l'ADN, Ségrégation de l'ADN, Synthèse de l'ARN, Modification de l'ARN, Synthèse des protéines, Modification des protéines

■ **Autres fonctions**

■ **Fonction inconnue**

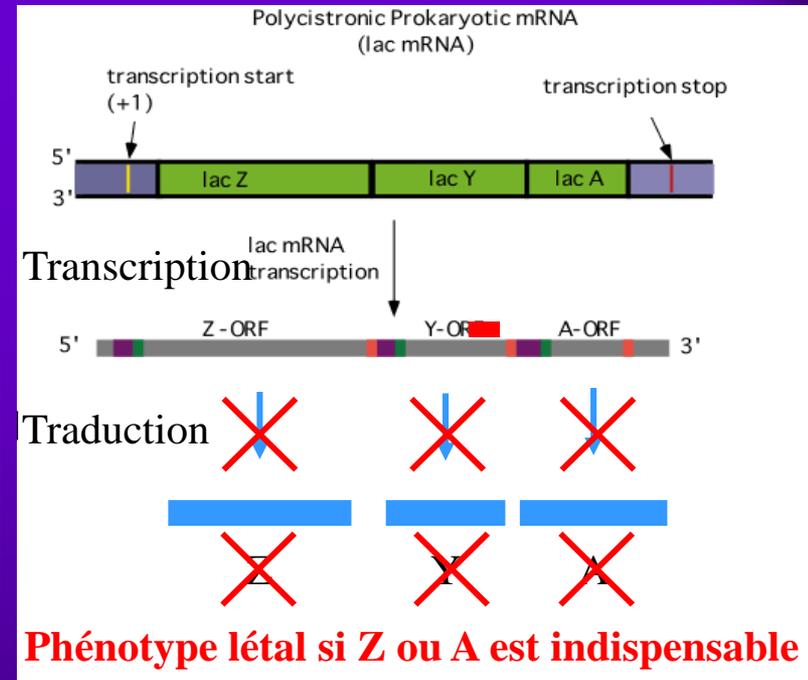
■ **Pas de similarité**

La majorité des gènes indispensables chez *S. aureus* ($658-168=490$) sont absents ou non indispensables chez *M. genitalium*.

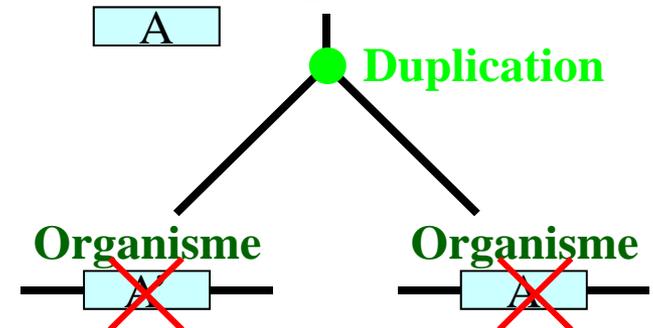
⇒ Il existe des différences majeures entre les organismes : des gènes indispensables chez l'un ne le seront peut-être pas chez l'autre. (Ex. *M.g* ne possède pas paroi contrairement à *S. aureus*)

Limites techniques de ces approches

- L'insertion des transposons:
 - sous-estimation (gènes résistants à l'insertion de transposons)
- L'utilisation de d'ARN anti-sens:
 - surestimation (ARN polycistroniques et d'ARN anti-sens complémentaires de motifs partagés par un ensemble de gènes)
- La présence de paralogues dans les « gros » génomes:
 - sous-estimation (redondance génomique)



Ancêtre de l'organisme considéré



Phénotype viable si A' peut assurer la fonction de A'' ou inversement

Limites conceptuelles de ces approches (1)

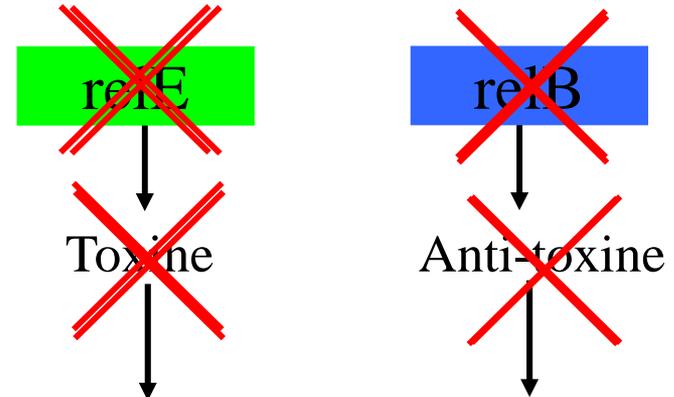
- Dépendance du contexte génomique:
 - Inactivations mutuellement exclusives
 - Inactivation inclusives

Inactivations mutuellement exclusives



Pas de croissance en aérobie et en anaérobie
Croissance en anaérobie
Croissance en aérobie
2 gènes non indispensables séparément
Gènes indispensables collectivement

Inactivations inclusives



Gènes non indispensables collectivement
Pas de croissance
Croissance
2 gènes indispensables collectivement

Limites conceptuelles de ces approches (2)

- Environnement idéal

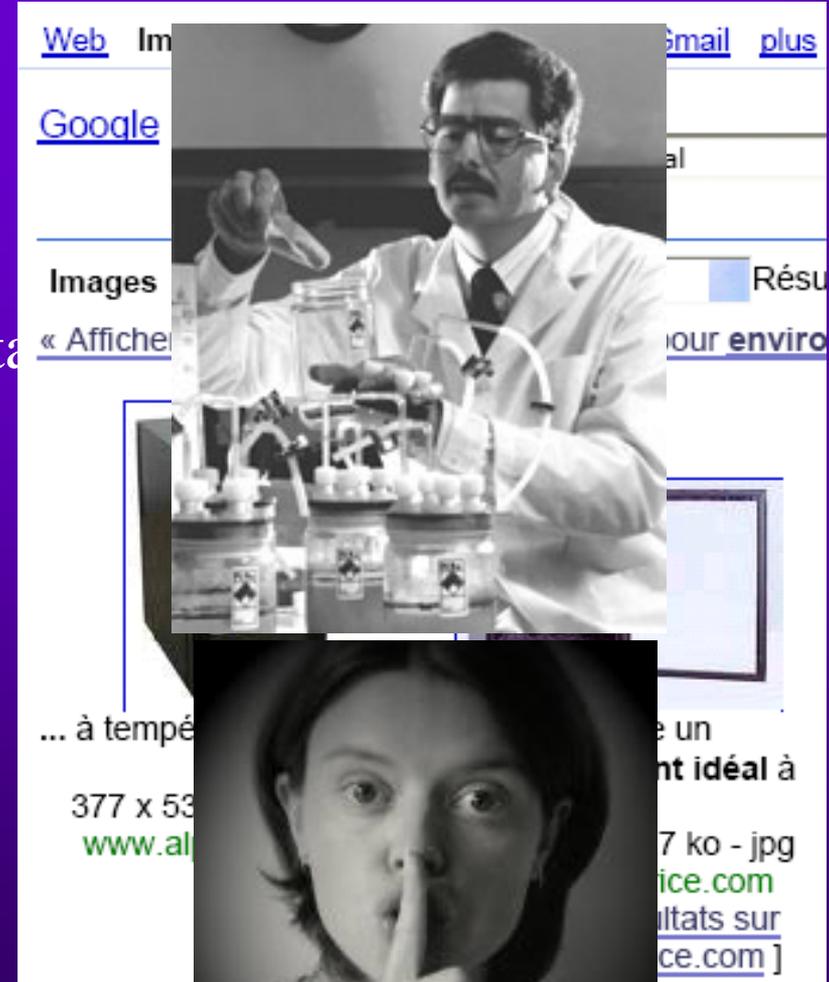
Tous les nutriments essentiels sont fournis

Absence de tout stress environnemental

Le set de gènes essentiels défini est

⇒ spécifique d'un organisme

⇒ associé à des conditions environnementales fixées



Tous les êtres vivants descendent d'un même ancêtre commun LUCA

Eucaryotes

Tous les êtres vivants actuels possèdent dans leur génome tout ce qui est nécessaire à la fabrication d'une cellule
Cette caractéristique a été héritée du dernier ancêtre commun: LUCA

Bactéries

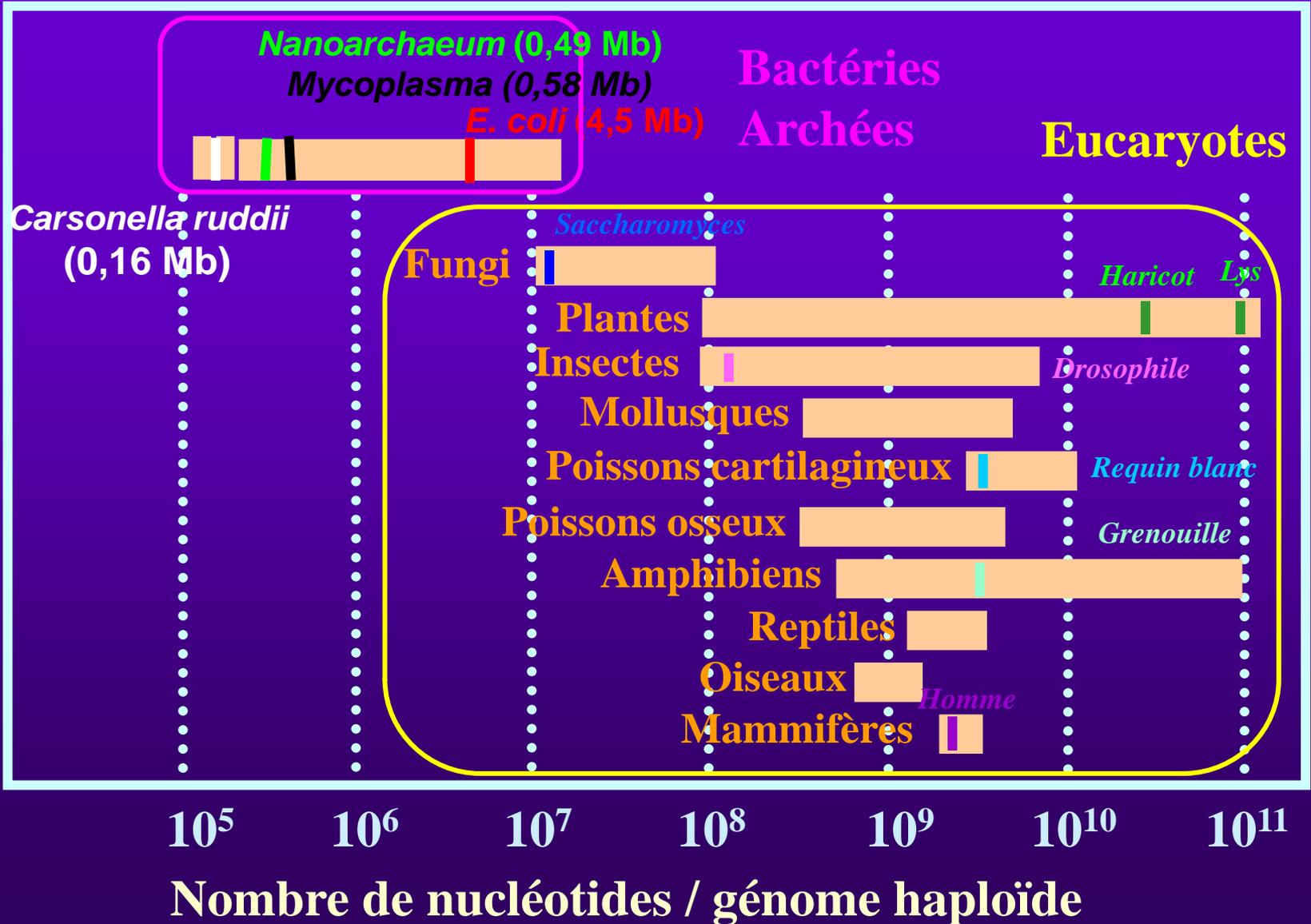
Archées

LUCA

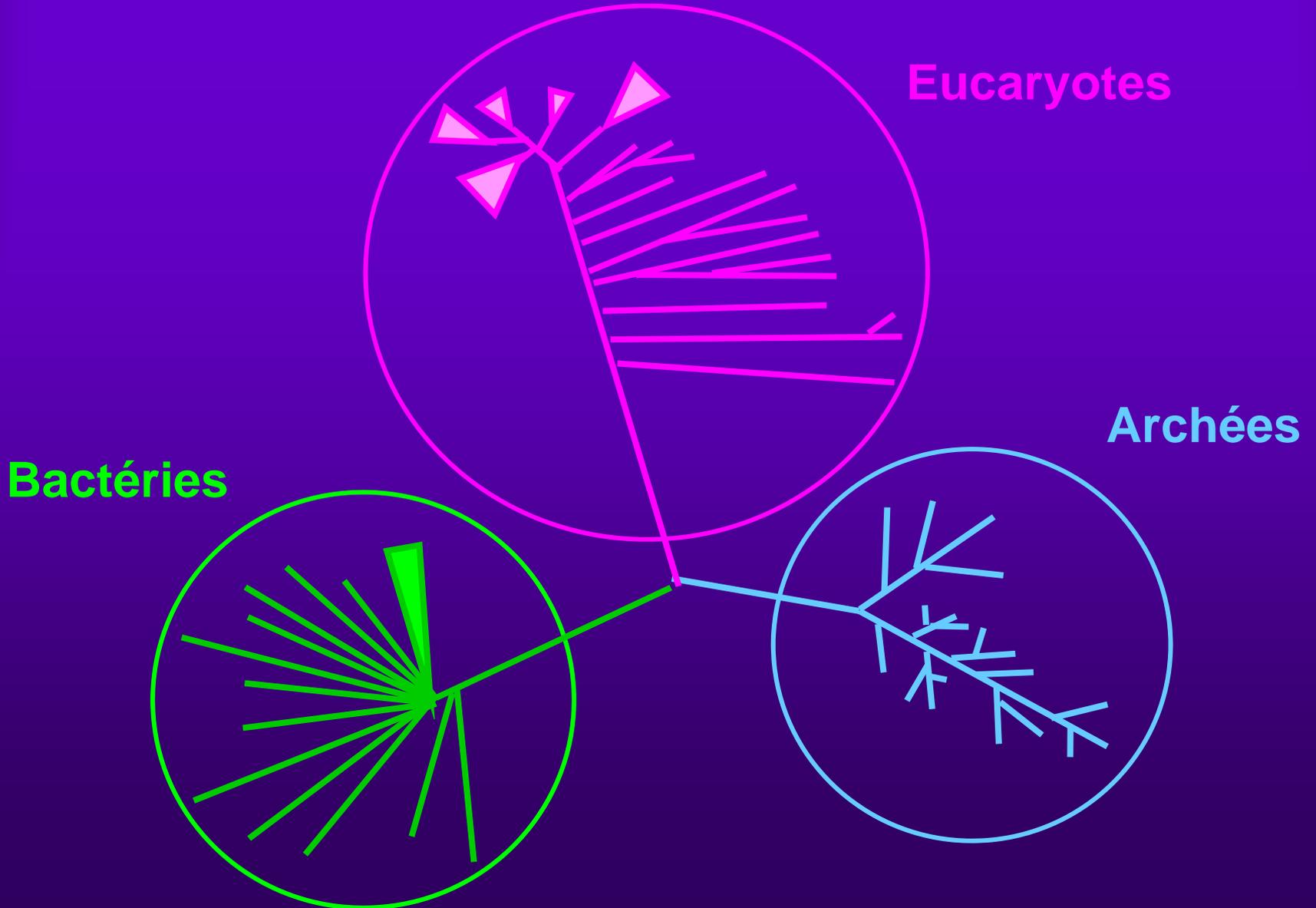
ancêtre cellulaire



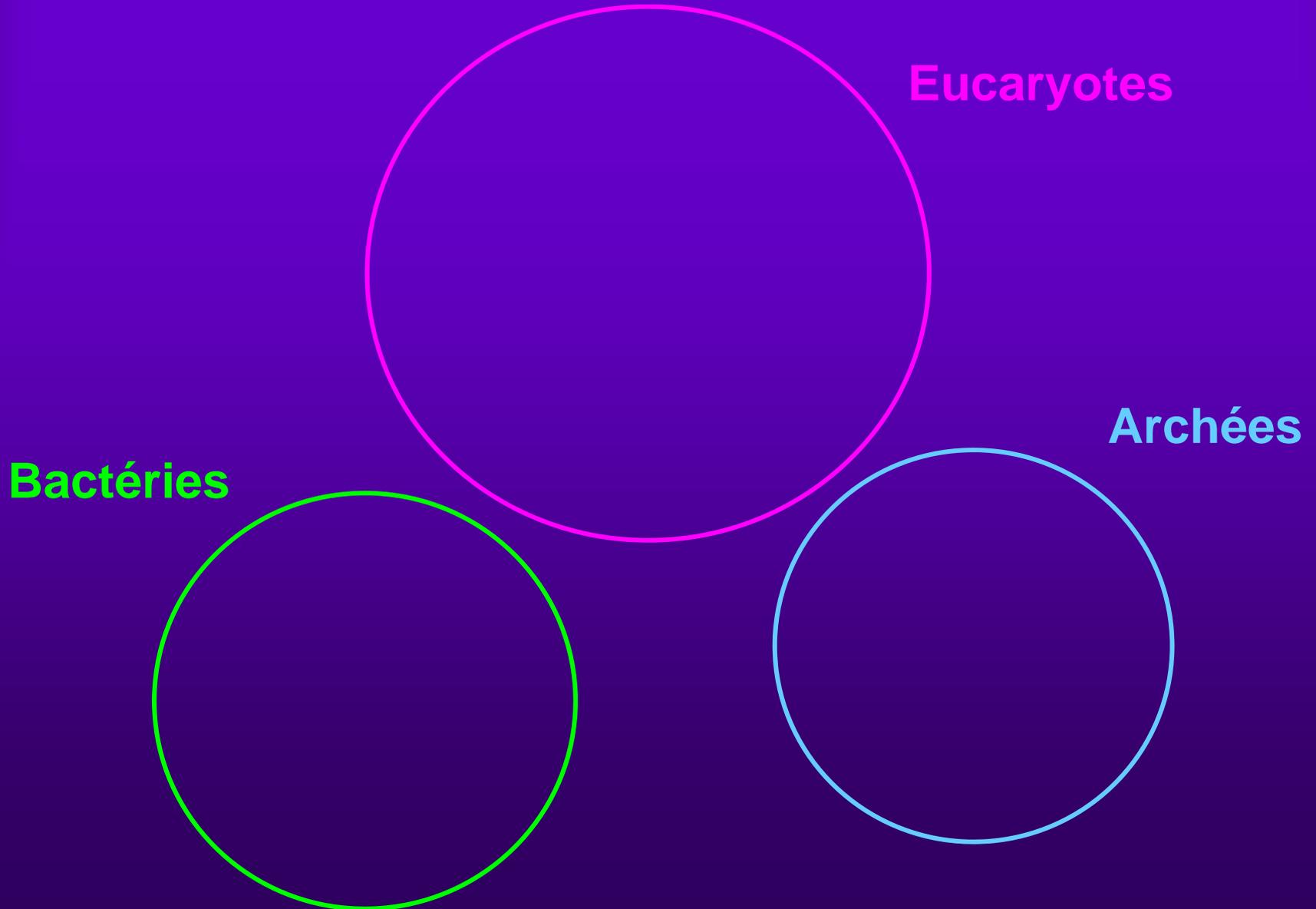
Retour en arrière



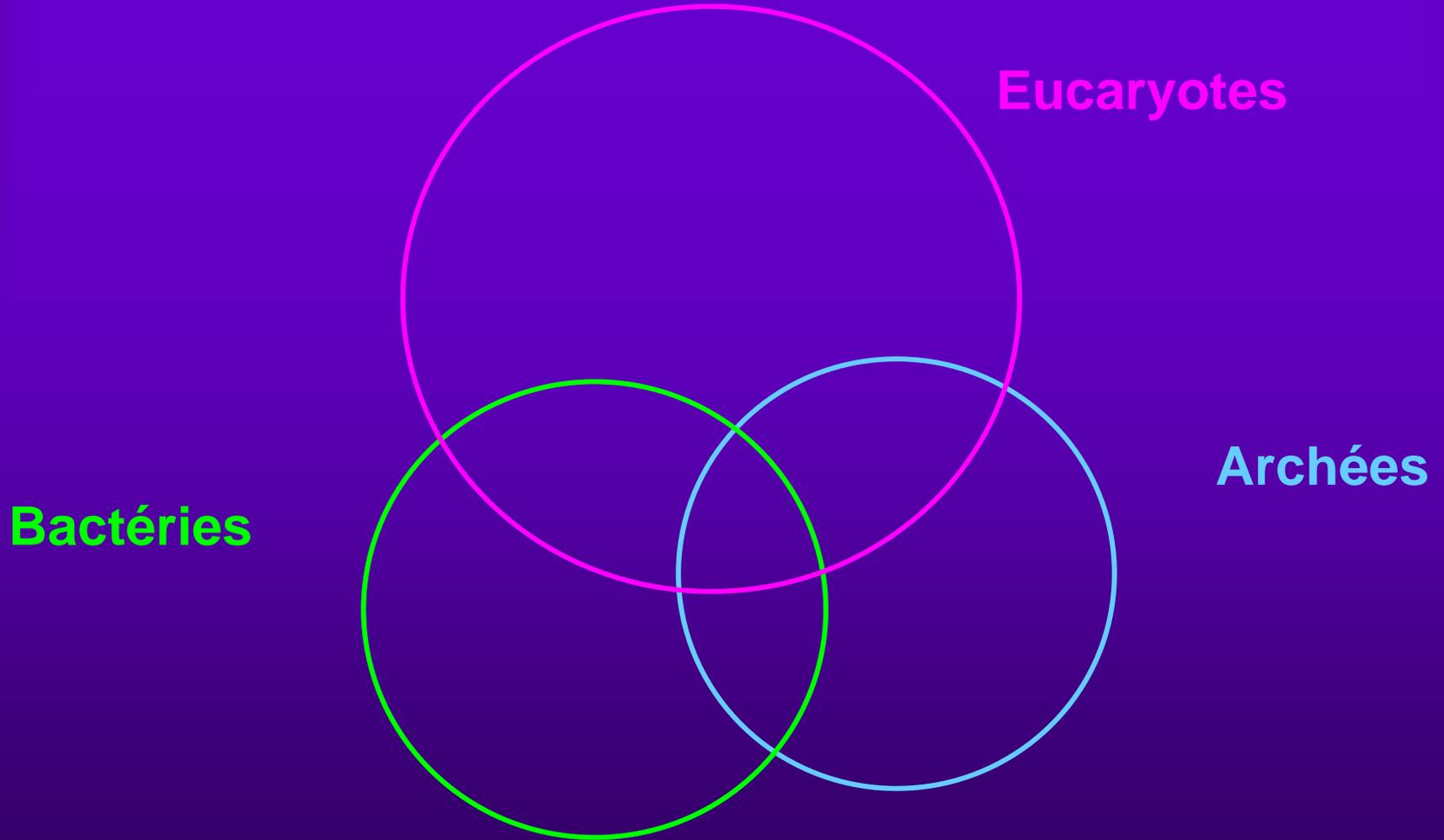
Génomique comparative



Génomique comparative



Génomique comparative



Génomique comparative

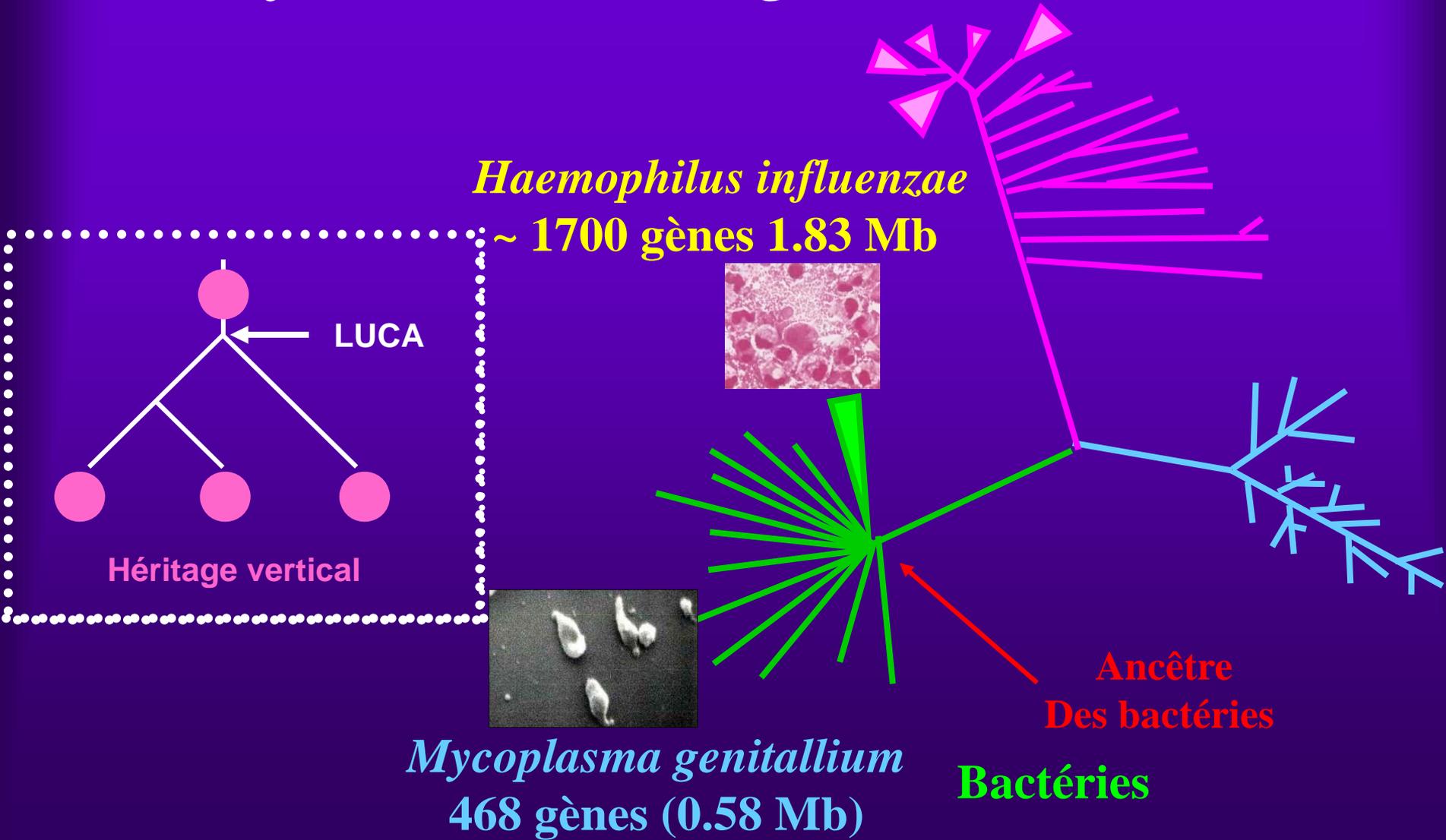
Eucaryotes

Comparer le contenu en gène des génomes
⇒ Identification des gènes hérités de LUCA (i.e. conservés au cours de l'évolution)

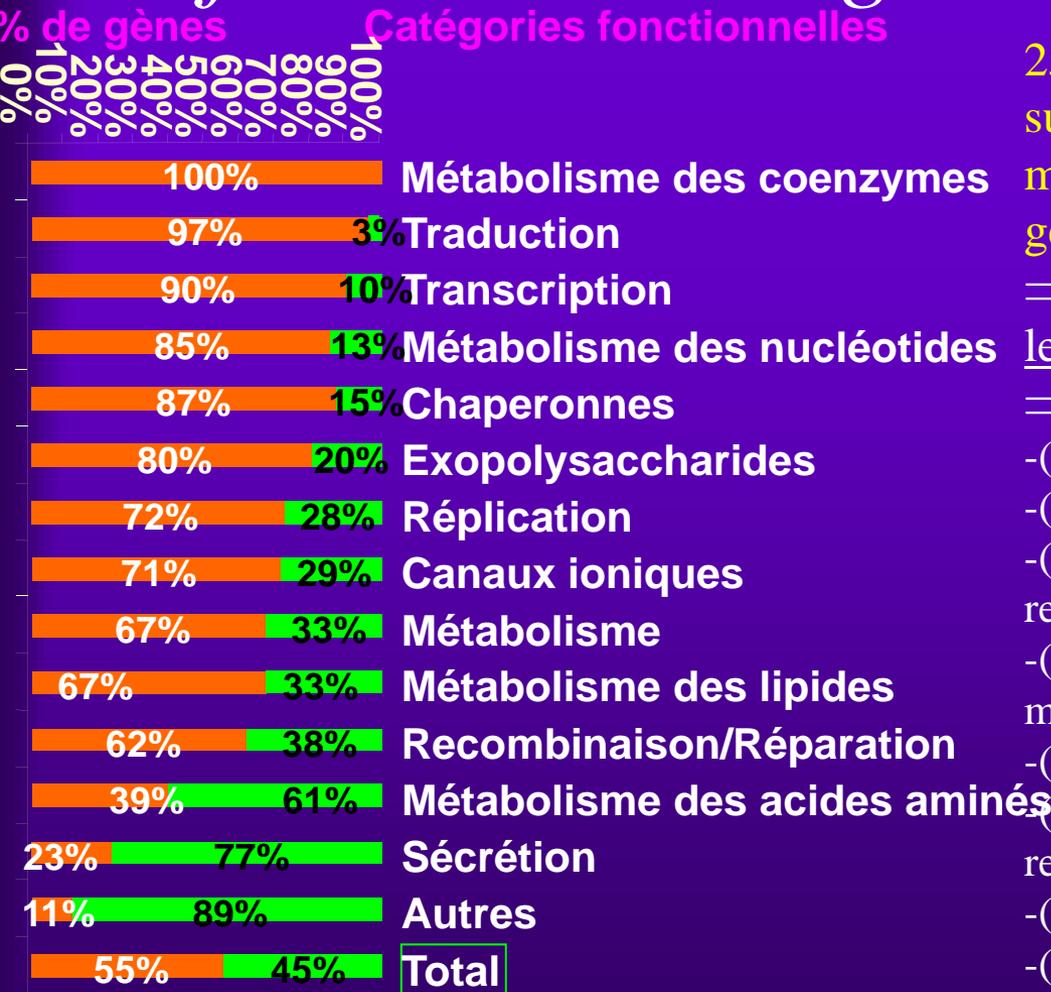
LUCA étant vraisemblablement un organisme cellulaire
⇒ Ce set de gènes doit inclure les gènes définissant « l'essence » de la cellule

Fraction des gènes hérités de LUCA et conservée au cours de l'évolution

Comparaison des génomes de *H. influenzae* et *M. genitalium* (1995)



Comparaison des génomes de *H. influenzae* et *M. genitalium*: 256 gènes



256 gènes supposés être nécessaires et suffisants pour permettre une vie cellulaire moderne et être proche du set minimum de gènes

⇒ Compatible avec les chiffres obtenus par les méthodes expérimentales

⇒ La cellule théorique contient:

- (1) un système presque complet de traduction
 - (2) une machinerie de réplication de l'ADN
 - (3) un système rudimentaire de réparation et de recombinaison de l'ADN
 - (4) un système de transcription presque complet mais sans de facteur de régulation
 - (5) des chaperonnes
 - (6) un métabolisme intermédiaire anaérobie restreint à la glycolyse
 - (7) une voie de biosynthèse des lipides limitée
 - (8) un système d'exportation des protéines
 - (9) un répertoire limité de protéine de transport de
- ⇒ Besoins nutritifs considérables (besoin d'importer tous ses aa, tous ses nucléotides, tous ses acides gras et ses coenzymes complexes)

■ Fraction des gènes présents chez *M. genitalium* et absents chez *H. influenzae*

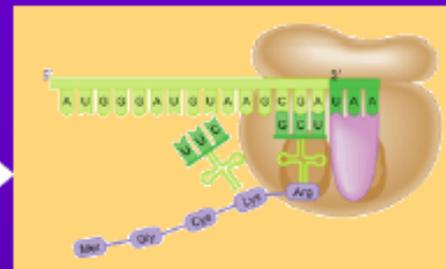
■ Fraction des gènes présents chez *M. genitalium* et conservés chez *H. influenzae*

Comparaison de 100 génomes

Set of universal proteins

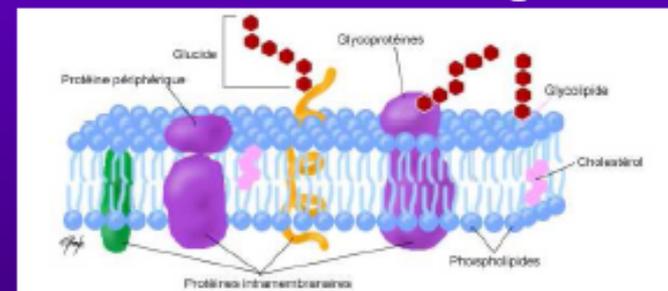
Functions	Number of genes
<i>Translation and associated functions</i>	
Ribosomal proteins	30
Aminoacyl-transfer-RNA synthetases	15
Translation factors	6
Enzymes involved in RNA and protein modification	3
Signal-recognition-particle components involved in secretion	3
Molecular chaperone/protease	1
<i>Transcription</i>	
RNA-polymerase subunits	2
<i>Replication/repair</i>	
DNA-polymerase subunit, exonuclease, topoisomerase	3
Total	63

LUCA had ribosomes & used the universal genetic code



		2nd base in codon			
		U	C	A	G
1st base in codon	U	Phe Leu Leu	Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Tyr
	C	Leu Leu Leu	Pro Pro Pro	His Gln Gln	Arg Arg Arg
	A	Ile Ile Ile Met	Thr Thr Thr	Asn Lys Lys	Ser Arg Arg
G	Val Val Val	Val Val Val	Asp Asp Glx	Gly Gly Gly	
		U	C	A	G

LUCA was a cellular organism



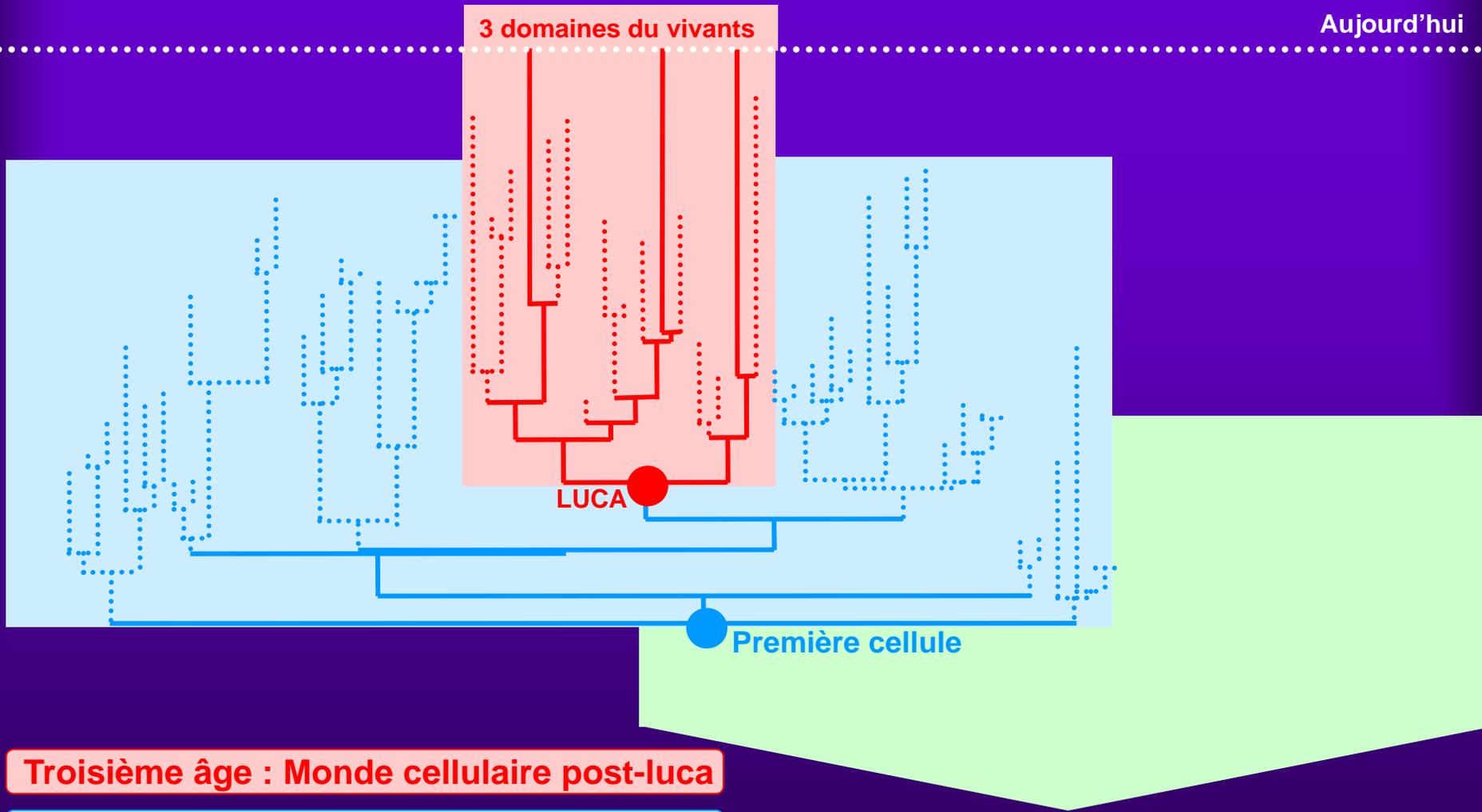
LUCA had a DNA genome ???



The set of universal proteins is not compatible with a viable cell (no metabolism...)

Quelques repères

Aujourd'hui



Troisième âge : Monde cellulaire post-luca

Second âge : Monde cellulaire pré-luca

Premier âge : Monde pré-cellulaire

Origine(s) de la vie

(Forterre et al. MS 2005)

Limites méthodologiques des approches bioinformatiques (1)

GOLD

L'inclusion d'un plus grand nombre de génomes très divergents et/ou de parasites diminue très fortement la taille du set de gènes conservés

=> Tolérance de certaines absences ???

91

Archaeal Ongoing Genomes

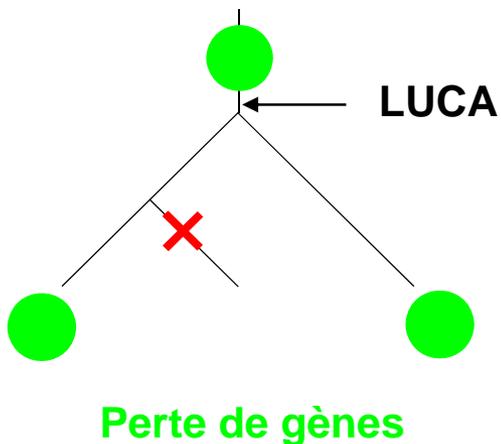
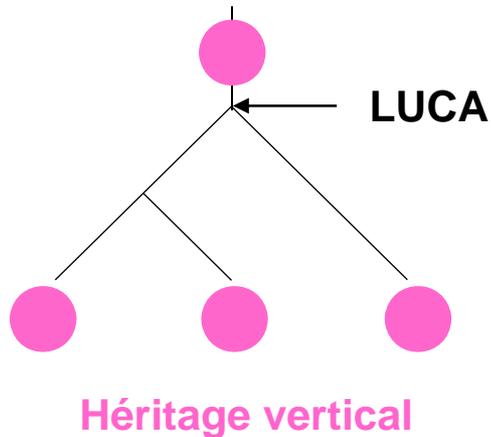
1750

Bacterial Ongoing Genomes

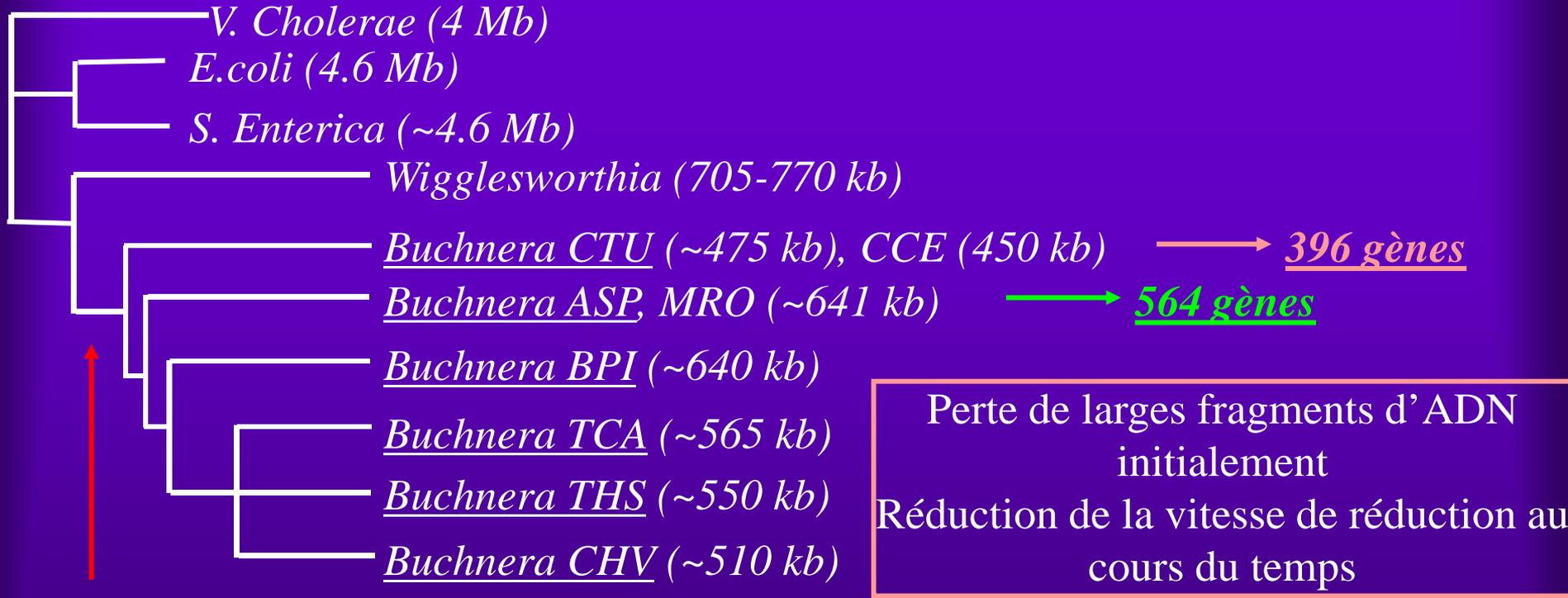
901

Eukaryotic Ongoing Genomes

Limites méthodologiques des approches bioinformatiques (2)



Réduction génomique chez les symbiontes de pucerons (*Gil, 2002*)



Infection unique il y a ~200 Ma (ancêtre 2-2.5 Mb ↔ 1800 à 2500 gènes)

Modifications génomiques très importantes chez le symbionte

-Réduction génomique très importante (taille et nombre de gènes),

0.9% des gènes absents chez *E. coli*

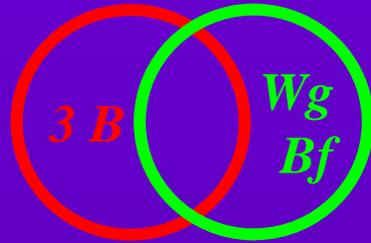
-Augmentation du taux de A+T

-Amplification des gènes impliqués dans la synthèse des acides aminés

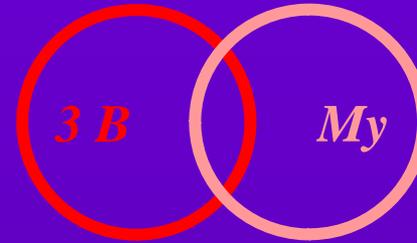
Comparaison des génomes de *Buchnera* (*Klasson 2004*) et de parasites bactériens



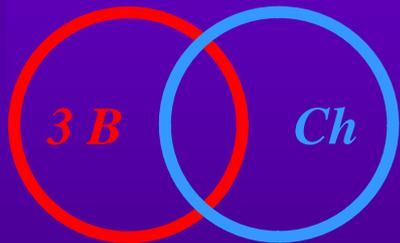
462 gènes
communs



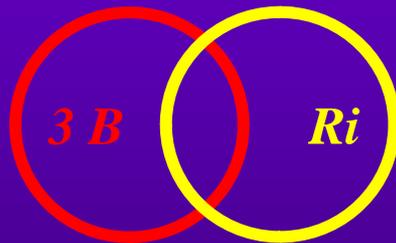
276 gènes
communs



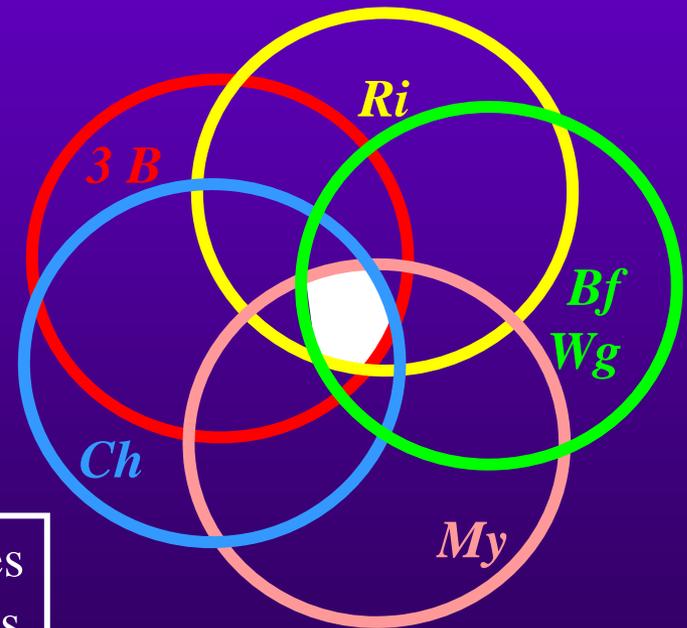
172 gènes
communs



218 gènes
communs



220 gènes
communs



156 gènes
communs

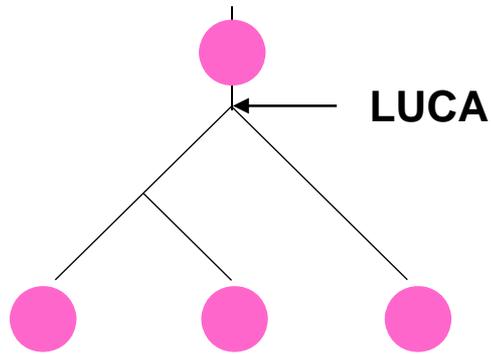
Comparaison des génomes de *Buchnera* (Klasson 2004) et de parasites bactériens

Catégorie	Endosymbiontes et parasites		%
	Endosymbiontes	et parasites	
Traduction	99	86	87%
Transport	8	5	63%
Transcription	13	8	62%
Autres	5	3	60%

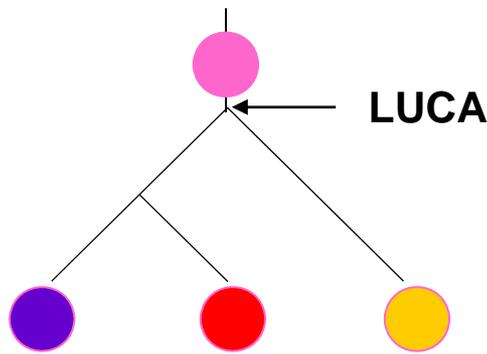
Les phénomènes de réduction génomique reflètent des adaptations à des modes de vie différents

Métabolisme des lipides	6	2	33%
Métabolisme central	3 	1	33%
Régulation	3	1	33%
Métabolisme des cofacteurs	11	3	27%
Métabolisme des acides aminés	10	1 	10%
Structures cellulaires	5	0 	0%
TOTAL	276	156	57%

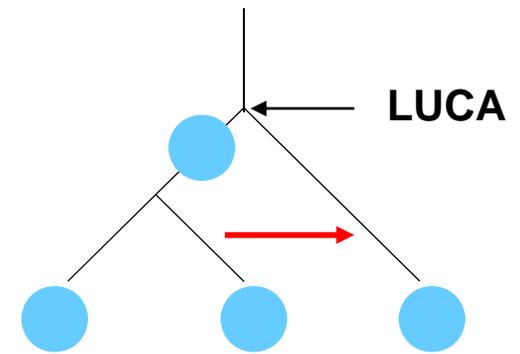
Limites méthodologiques des approches bioinformatiques (2)



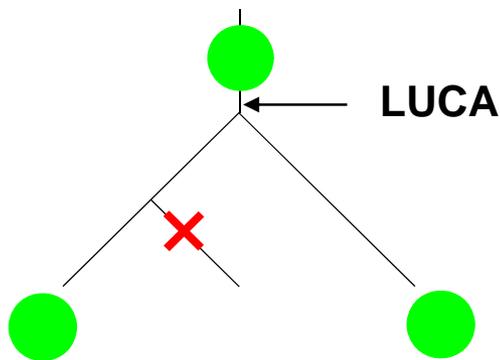
Héritage vertical



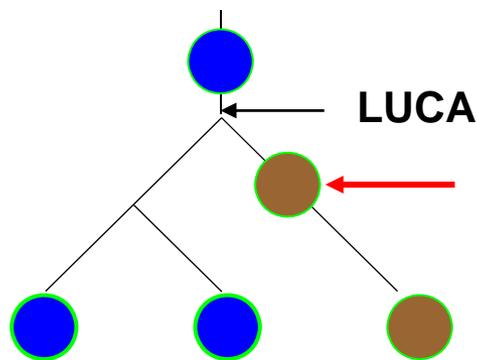
Divergence excessive



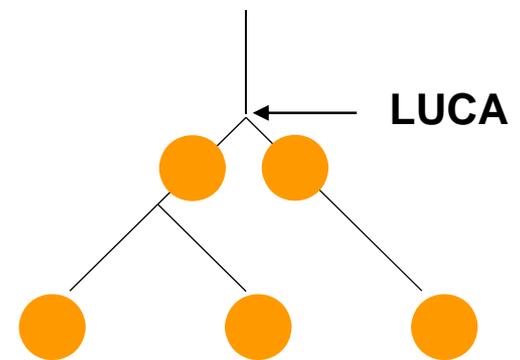
Transfert horizontal de gènes



Perte de gènes



Remplacement non homologue



Convergence fonctionnelle

Limites conceptuelles des approches bioinformatiques

(1) Le génome reconstruit dépend des organismes utilisés pour

=> Développement de nouvelles approches

(2) Certains gènes présents comme non-fonctionnels ont pu être inactivés expérimentalement

(3) Gènes définissant la cellule ?

Recherche de fonctions universelles

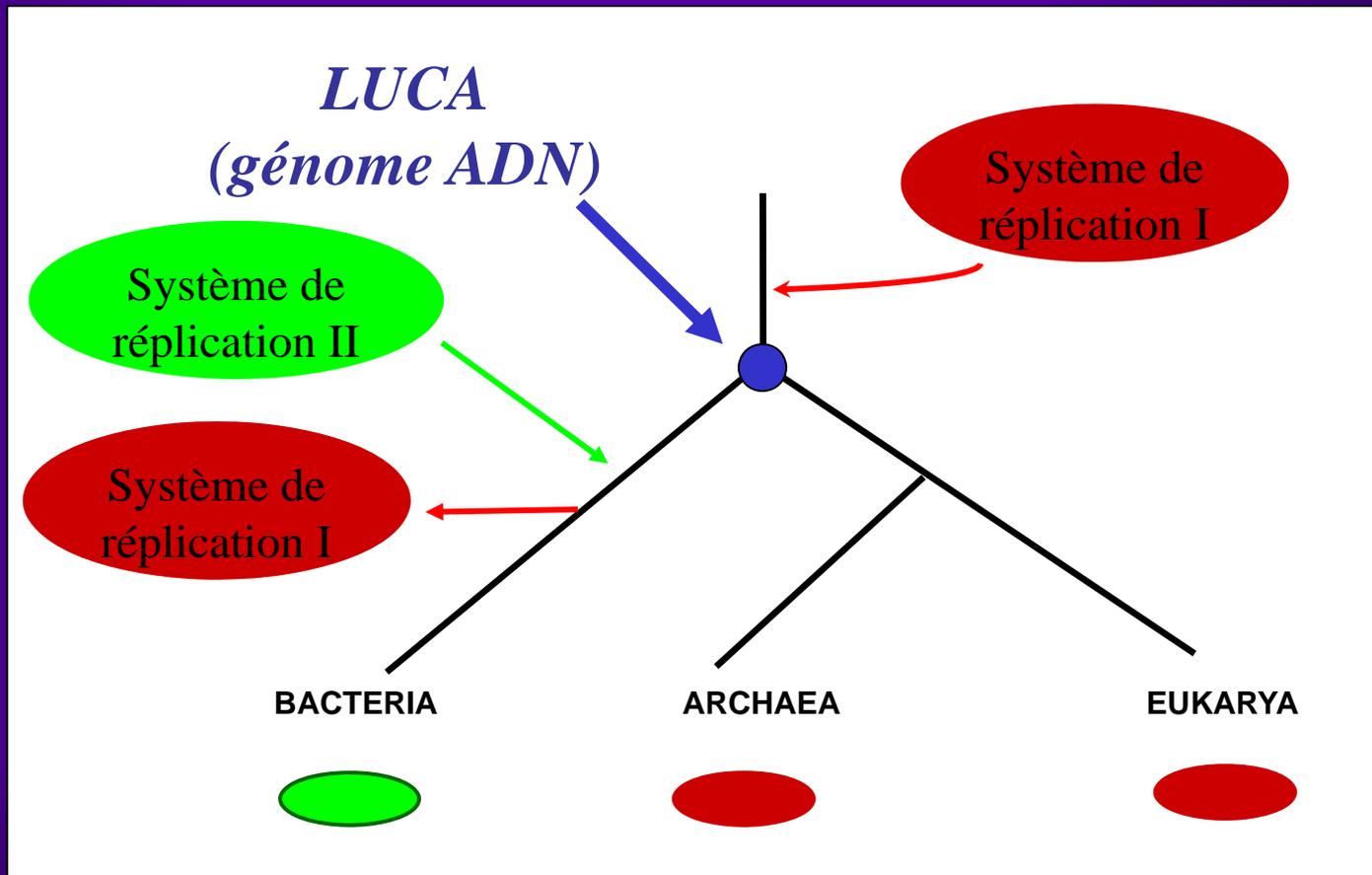
- Approche moins restrictive: l'identification de fonctions semblables conservées chez tous (?) les organismes actuels => présentes chez LUCA
- Utilisation des groupes d'organismes (et non plus organismes individuels) => limite considérablement l'impact des organismes à génomes extrêmement réduits

*~350-1000
gènes*

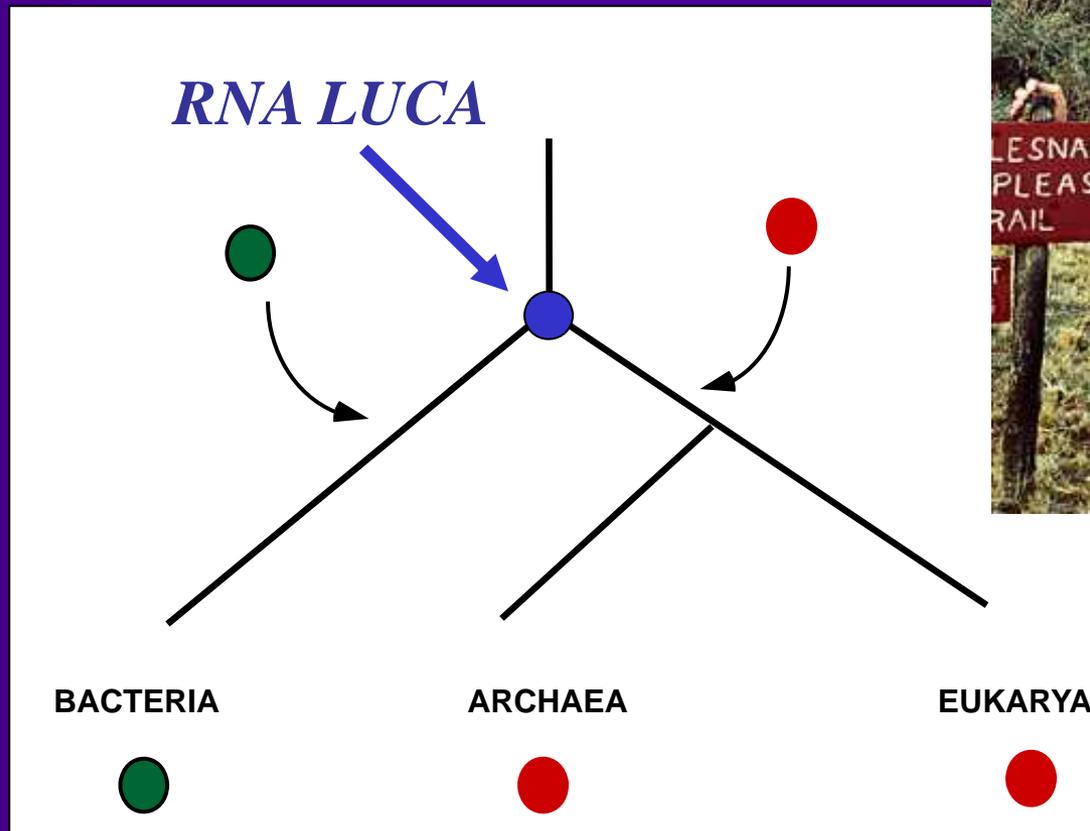
Recherche de fonctions universelles

- ⇒ Inférence de cellules capable de synthétiser *de novo* ses aa, ses nucléotides, des carbohydrates complexes et quelques coenzymes => moindre dépendance vis à vis de l'environnement
- Révèle que des fonctions primordiales sont réalisées par des protéines non orthologues
 - Ex. La majorité des protéines impliquées dans la réplication de l'ADN ne sont pas orthologues entre bactéries et eucaryotes/archées

Remplacement du système de réplication par un système non homologue au cours de l'évolution dans l'une des trois lignées cellulaires



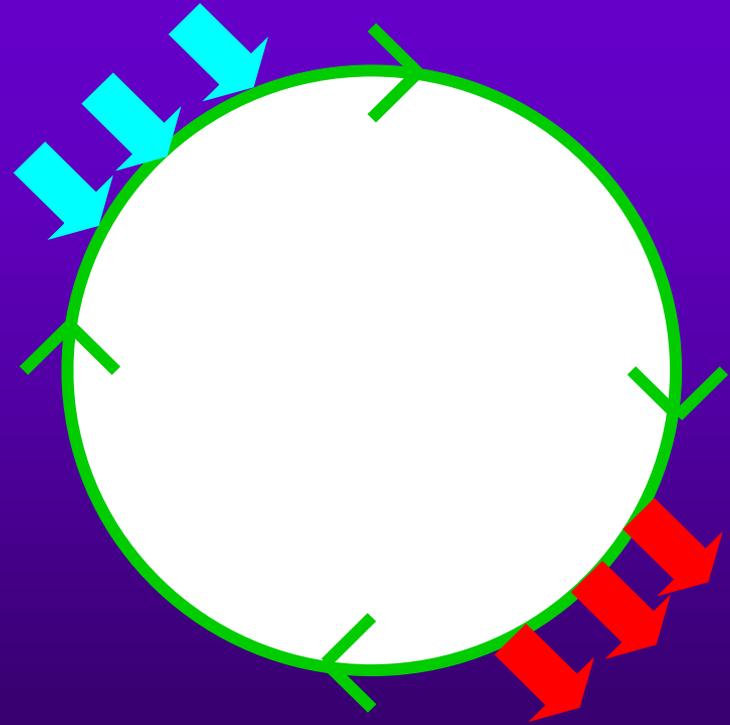
(Mushegian and Koonin, PNAS, 1996)



LUCA était une cellule à ARN et l'ADN a été inventé deux fois indépendamment dans la lignée bactérienne et dans la lignée Archées/Eucaryotes

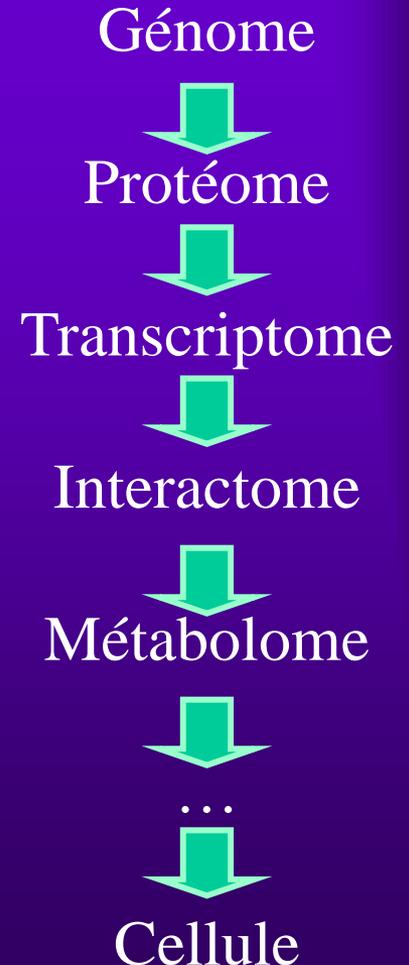
Les génomes sont très plastiques

- Taille des génomes n'est pas infinie
- Leur contenu a beaucoup évolué au cours du temps
- Équilibre entre le gain et la perte de gènes
- Acquisitions d'ADN
- Pertes d'ADN



A retenir (1)

- Construction purement théorique
- Problème très complexe => Solutions multiples
- Expérimentation
 - Solutions ≠ selon les organismes considérés
 - Environnement => influe sur ce qui est indispensable et ce qui ne l'est pas
 - Dépendent du type de cellule considéré
 - Cellule phototrophe / hétérotrophe / parasite / bactérienne / archée / eucaryote...
- Bioinformatique
 - Solutions ≠ selon les génomes/paramètres considérés
 - Problèmes méthodologiques
 - Simplification extrême du problème (régulations, interactions...)



A retenir (2)

Les plus petits génomes actuels > aux génomes minimaux prédits

=> reflète la différence entre les génomes résultants de l'interaction de gènes pendant des millions d'années d'évolution et un génome fabriqué *de novo* par un ingénieur assemblant artificiellement des choses qui ont été inventées indépendamment

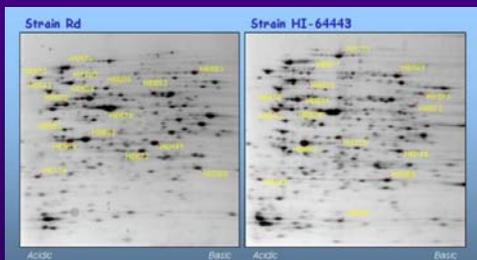
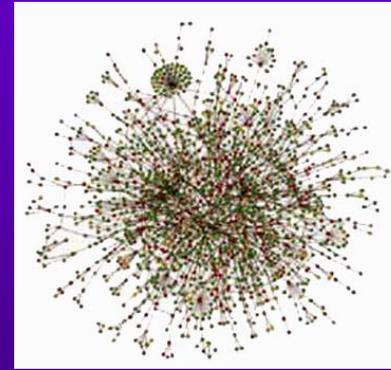
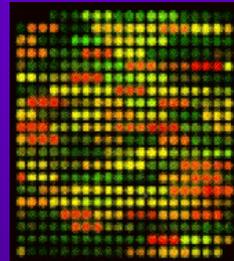
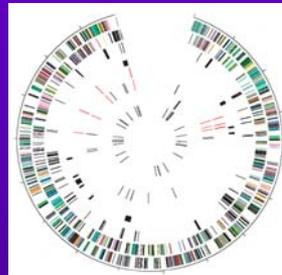
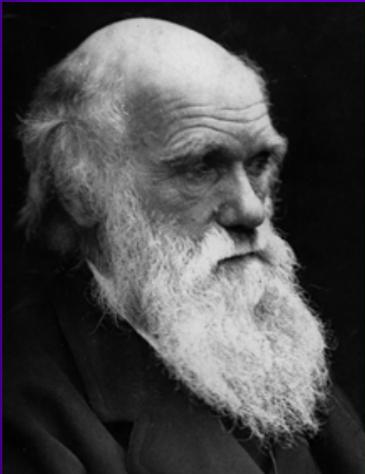


FIGURE 1. Example 2D protein maps for *Haemophilus influenzae*. Cellular proteins for two *H. influenzae* strains were analysed by 2DE. The isolates used were the laboratory Rd strain which has been fully sequenced and a clinical non-typable isolate (HI-64443) widely used in my laboratory. The selected proteins indicated on the protein maps were identified using peptide mass mapping. The identities of these proteins are shown using the ORF notation for the *H. influenzae* Rd genome sequence. High molecular weight proteins migrate at the top of the gel.

