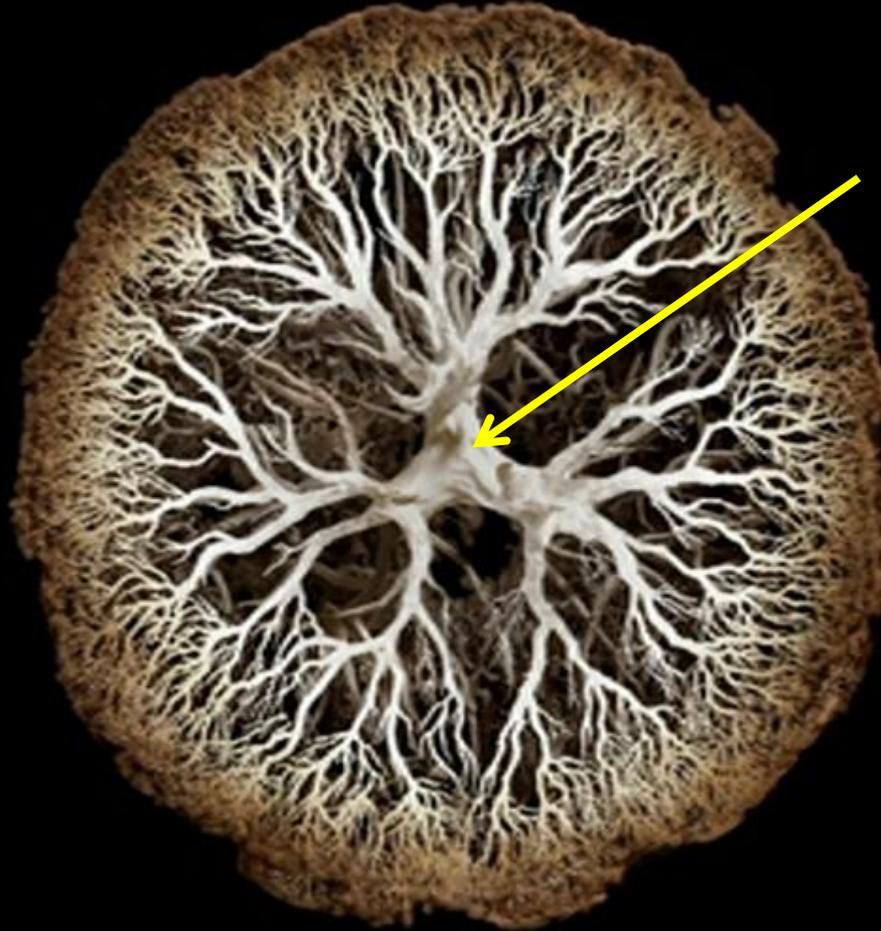


# Last Universal Common Ancestor

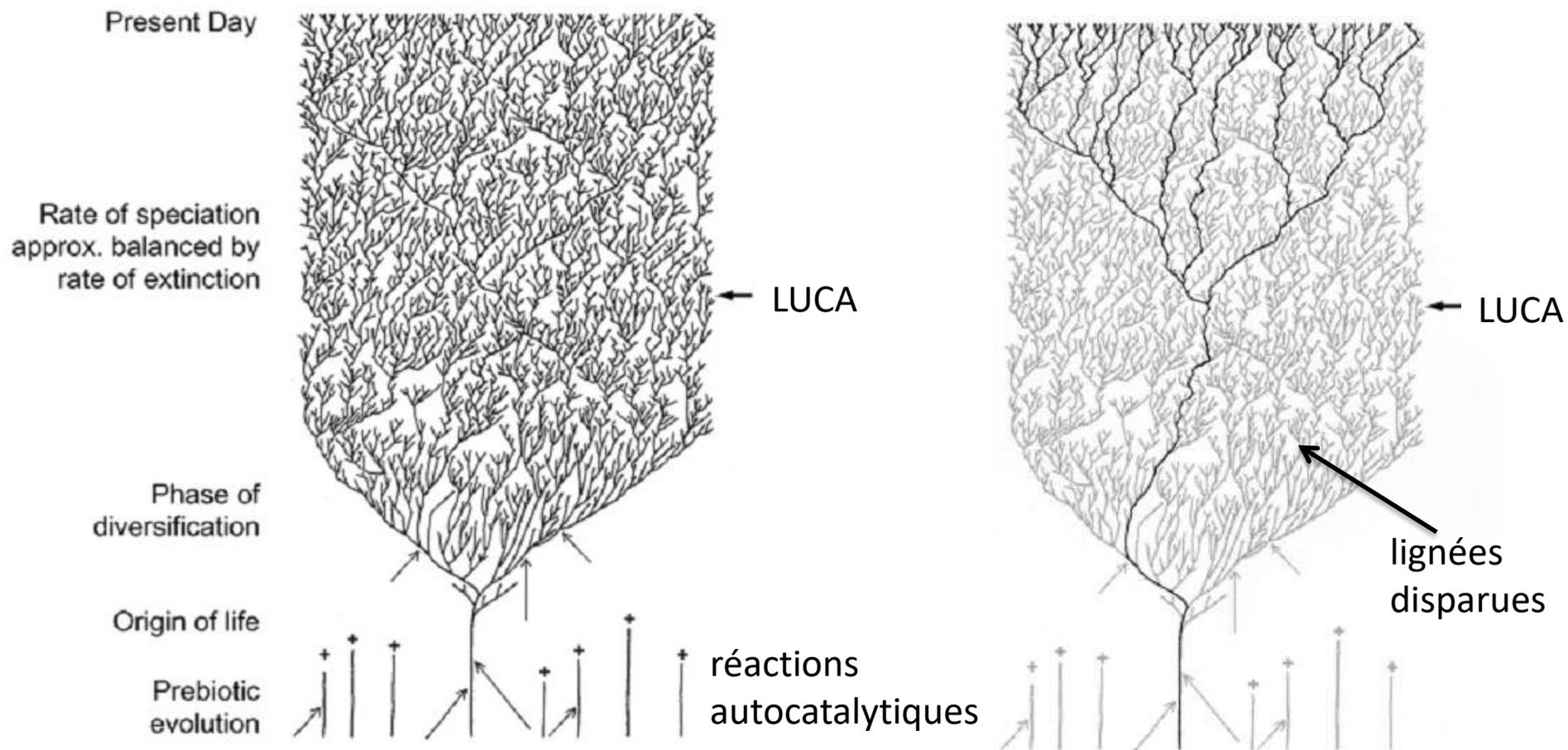


LUCA

Seul au monde ?  
Quel type de membranes?  
Hyperthermophile?  
Production d'énergie?  
Origine?

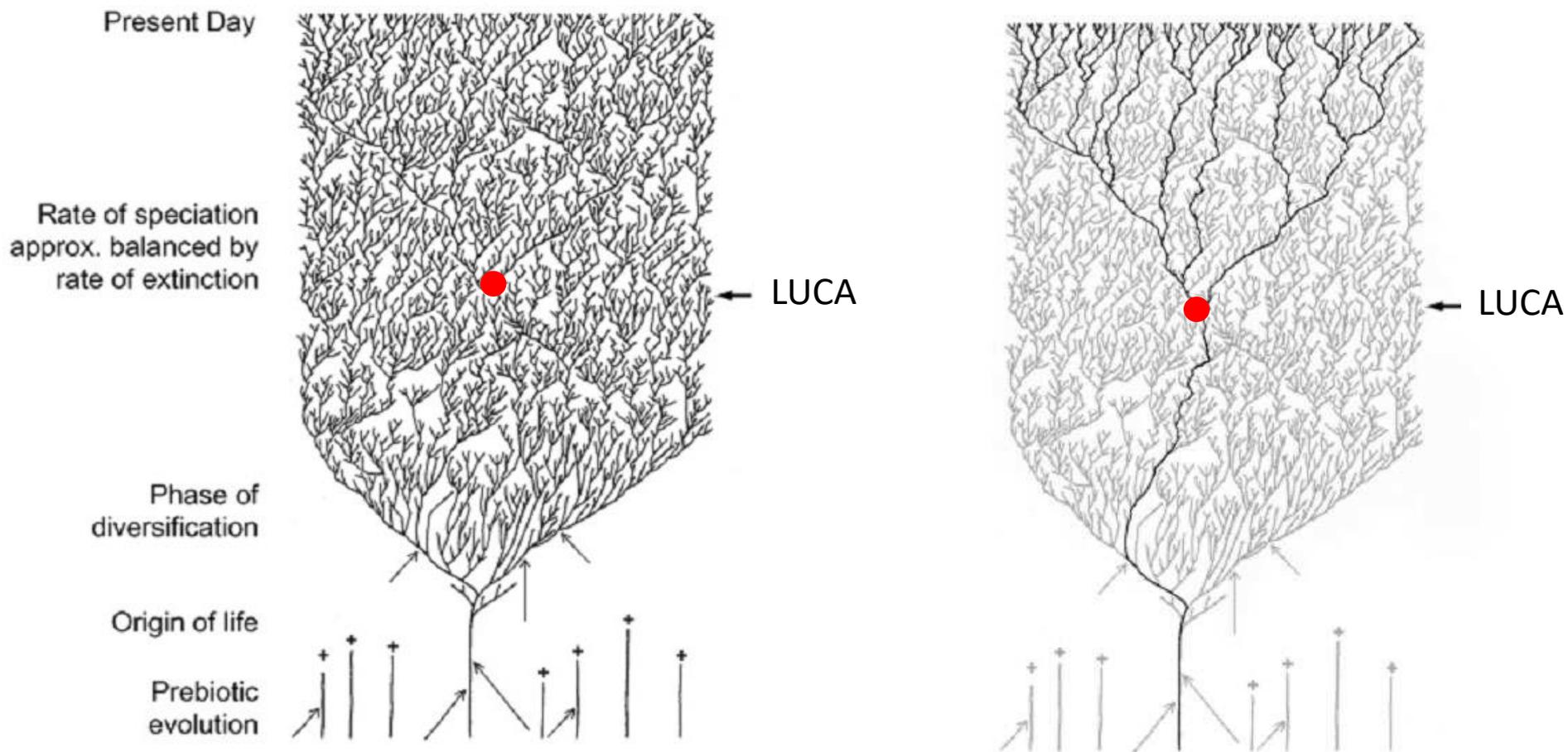
# LUCA n'était pas seul

L'origine et l'évolution de la vie:



LUCA n'était pas la première cellule...  
(<http://web.uconn.edu/gogarten/>)

# LUCA n'était pas seul

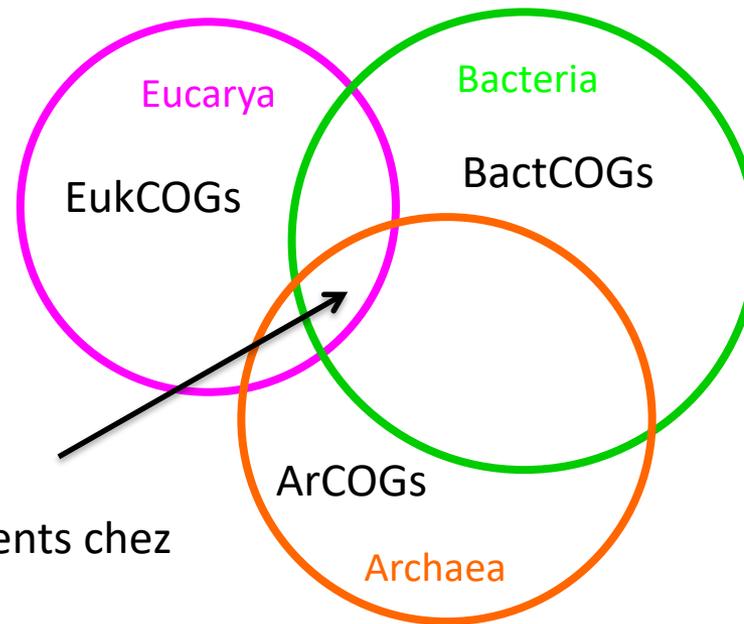


LUCA n'était pas seul... il était membre d'une espèce et il partageaient l'environnement avec d'autres espèces

(<http://web.uconn.edu/gogarten/>)

# Portrait de LUCA - génomique comparative

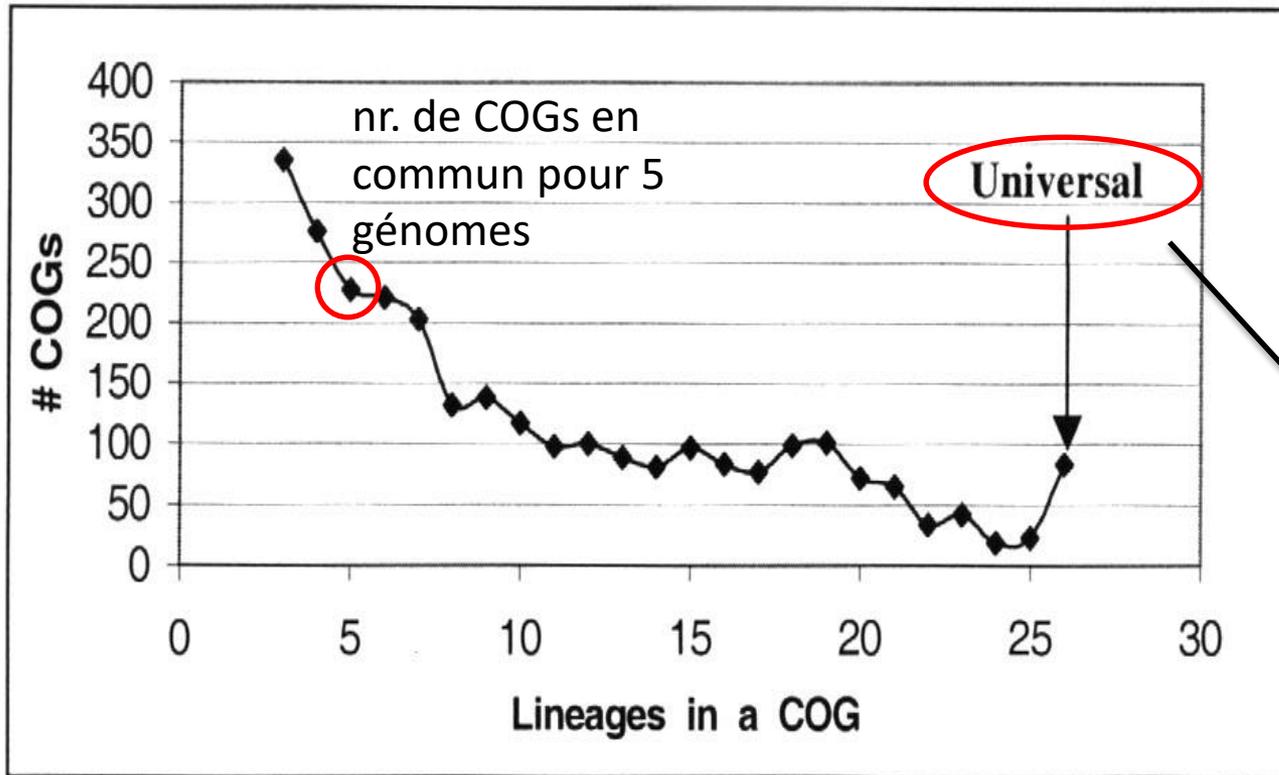
-le contenu en gène du LUCA peut être estimé par la génomique comparative:



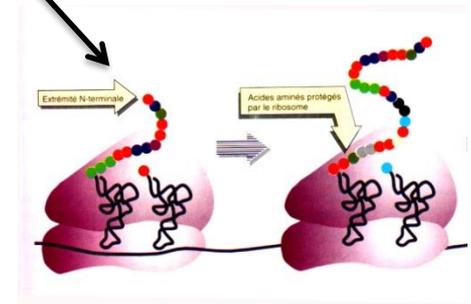
-les gènes universelles  
vraisemblablement présents chez  
LUCA

COG – cluster of orthologous genes

# Portrait de LUCA - génomique comparative



Eugene Koonin



Traduction

**Figure 2.8. Distribution of different phylogenetic lineages in the COG database.** The plot shows the number of protein families (COGs) in a release of the COG database (♦3.4), which included proteins from the given number of phylogenetic lineages out of the total of 26 lineages [827].

# Portrait de LUCA - génomique comparative

Functions	Number of genes
<b><i>Translation and associated functions</i></b>	
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
<b><i>Transcription</i></b>	
RNA-polymerase subunits	2
<b><i>Replication/repair</i></b>	
DNA-polymerase subunit, exonuclease, topoisomerase	3
<b>Total</b>	<b>63</b>

La plupart des protéines universelles font partie du ribosome → ce set n'est pas compatible avec un ribosome fonctionnel

Des nombreuses fonctions importantes (la réplication d'ADN, métabolisme) ne figurent pas sur la liste?!?

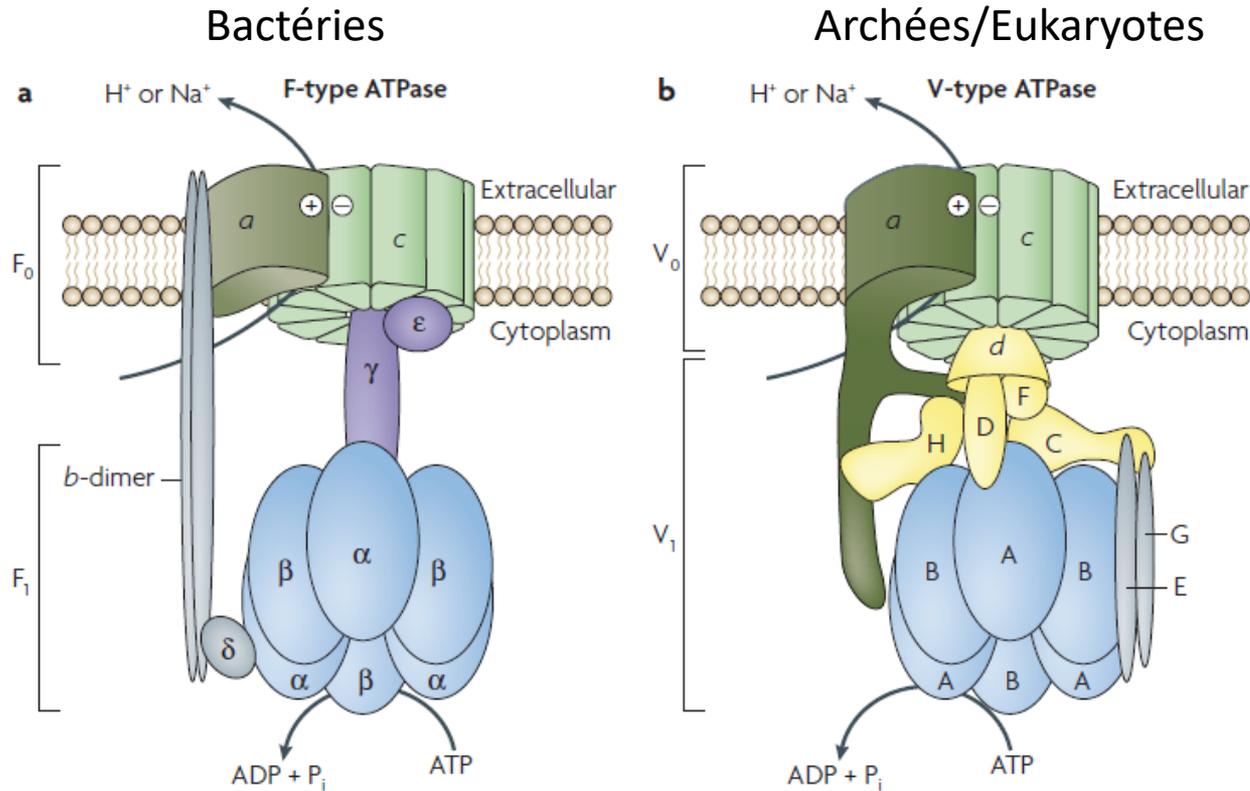
# Portrait de LUCA - génomique comparative

Authors	Data set used in the study	Number of genes/proteins/domains inferred in LUCA	Inference
Kyrpides et al. (1999)	<i>M. jannaschii</i> versus nr-db	324 proteins	LUCA similar in complexity to extant cells
Yang et al. (2005)	174 genomes	49 SCOP superfamilies	
Delaye et al. (2005)	20 genomes	115 Pfam domains	
Ouzounis et al. (2006)	184 genomes	669 ortholog families	
Ranea et al. (2006)	114 genomes	140 CATH protein domains	
Koonin (2003)	~100 genomes	63 ubiquitous genes	Simple LUCA, with a small number of genes
Weiss et al. (2016)	286,514 protein clusters	355 protein families	
Harris et al. (2003)	34 genomes	80 COGs	No strong claim about the complexity of LUCA
Mirkin et al. (2003)	32 genomes	572 proteins/COGs	

Delaye, 2024, Journal of Molecular Evolution

# Portrait de LUCA - production d'énergie

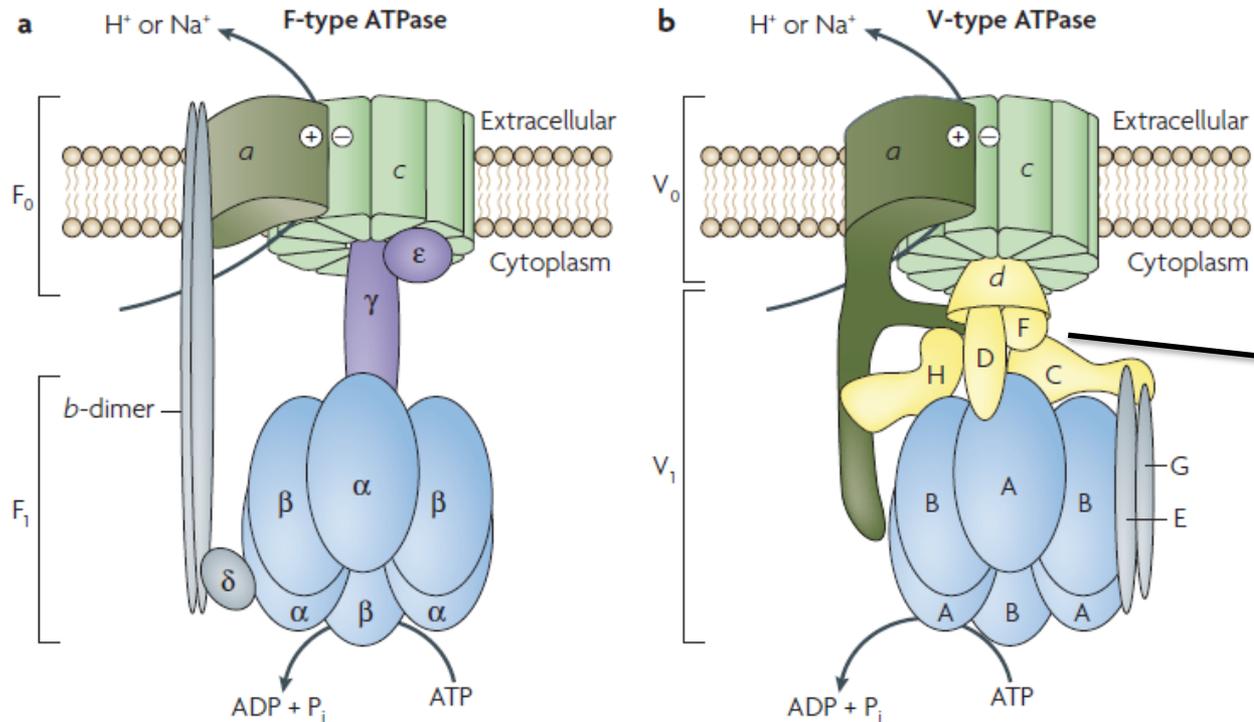
Comment LUCA produisait l'énergie?



-les cellules modernes utilisent les ATP synthétases: couplage du transfert des protons à travers la membrane et la production d'ATP

Mulkiđjanian et al., 2007, Nat. Rev. Microbiol

# Portrait de LUCA - production d'énergie



les protéines formant le lien entre les sous-unités membranaires et les sous-unités ATP synthétase ne sont pas homologues

- lien entre les deux parties est essentiel pour la production d'ATP
- LUCA n'avait pas d'ATP synthétase?

# Portrait de LUCA - production d'énergie

Hypothèse: LUCA utilisait le gradient des protons naturel et ses membranes était perméables aux protons

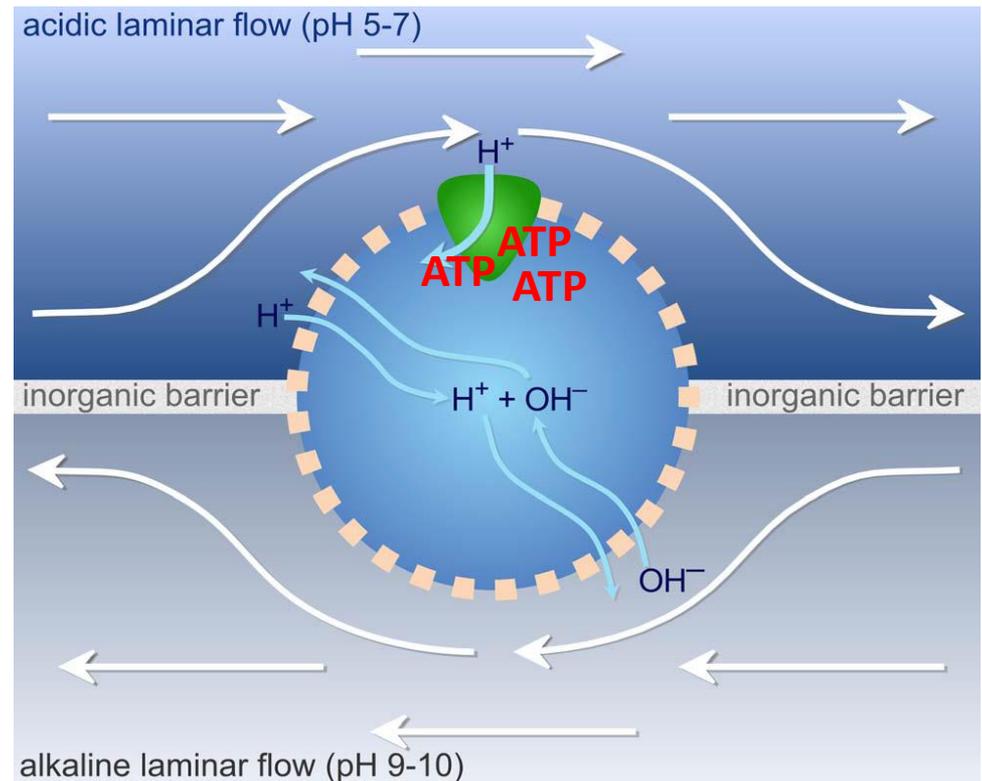
-Pré-requis: Les membranes de LUCA n'étaient pas phospholipidiques (car pas assez perméables aux protons) mais faites des acides gras et d'isoprènes

une **protéine intégrale** positionnée de côté acide convertie le flux des protons en énergie sous forme d'ATP



les sources hydrothermales alcalines, pH 9-11, T= 40-90°C

Lost World



# Portrait de LUCA - production d'énergie

Hypothèse: LUCA utilisait le gradient des protons naturel et ses membranes étaient perméables aux protons

-Pré-requis: Les membranes de LUCA n'étaient pas phospholipidiques (car pas assez perméables aux protons) mais faites des acides gras et d'isoprènes

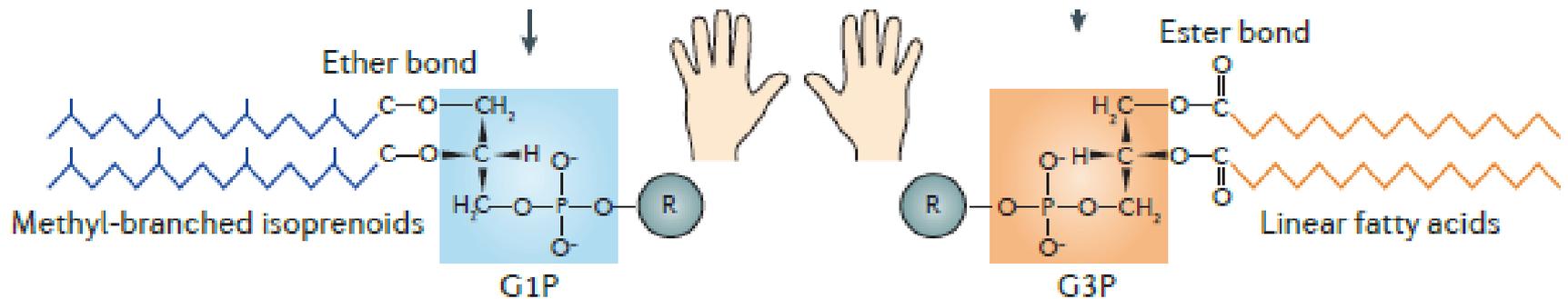


Membranes mixtes ?

Un sujet débattu...

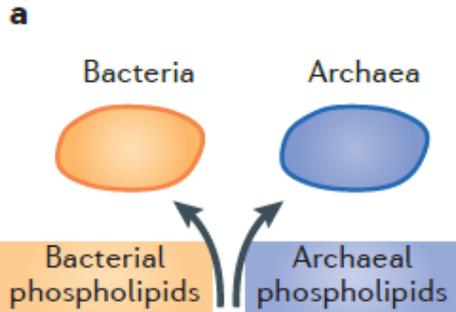
# Portrait de LUCA - nature des membranes

→ Les phospholipides de bactéries/eukaryotes et d'archées sont structuralement similaires mais chimiquement très différents

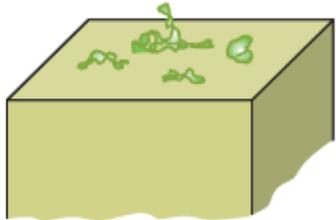


# Portrait de LUCA - nature des membranes

Les hypothèses concernant la nature des membranes de LUCA:



Acellular ancestor;  
no membrane



Surface metabolism on pyrite

→ Koga et al. LUCA n'était pas une cellule

→ les lipides ont émergé indépendamment dans les lignées conduisant aux Archées et Bactéries

→ Critique: LUCA était un organisme déjà assez complexe (plusieurs centaines de gènes) – une telle complexité aurait nécessité des entités individuelles entourés d'une membrane

M  
ci

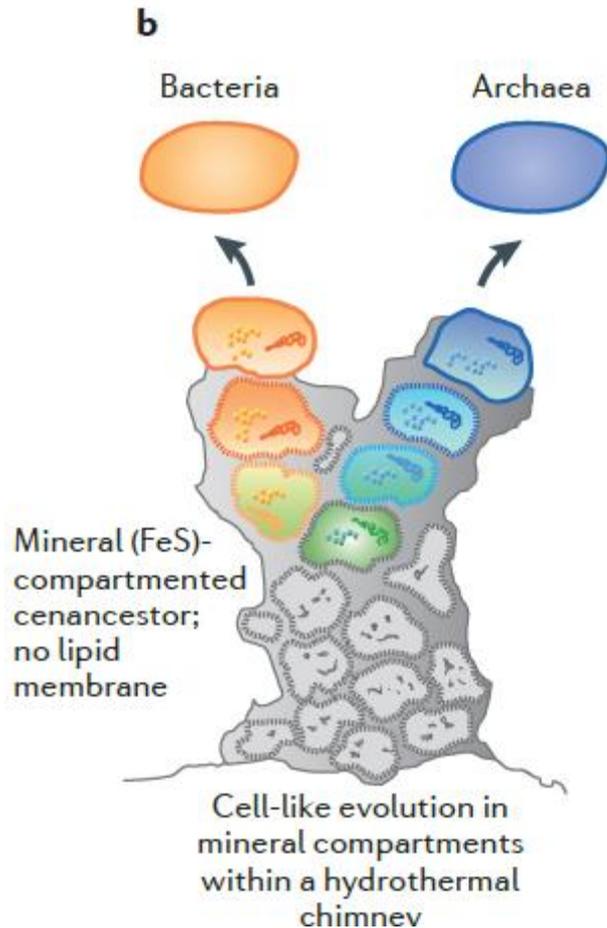
## Lipid membrane-associated proteins

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>• Universally conserved H<sup>+</sup> (or Na<sup>+</sup>) ATPase<sup>32,33,33</sup></li> </ul>  | <ul style="list-style-type: none"> <li>• Synthesis of ATP exploiting a transmembrane ion gradient</li> <li>• Need for a continuous lipid membrane</li> </ul>         |
| <ul style="list-style-type: none"> <li>• Universally conserved components of respiratory chains (cytochrome <i>b</i>, Rieske protein, hydrogenases and dioxygen reductases)<sup>34-37</sup></li> </ul> | <ul style="list-style-type: none"> <li>• Likely presence of a respiratory chain in the membrane that could generate a proton gradient across the membrane</li> </ul> |
| <ul style="list-style-type: none"> <li>• Universally conserved SRP system (SRP domain and its receptor)<sup>38</sup></li> </ul>  | <ul style="list-style-type: none"> <li>• Targeting of proteins to the membrane</li> </ul>  |
| <ul style="list-style-type: none"> <li>• Universally conserved Sec, YidC and Tat pathways<sup>39-41</sup></li> </ul>   | <ul style="list-style-type: none"> <li>• Protein export and insertion into the membrane</li> </ul>   |

CDP-APT, cytidine diphosphate-alcohol phosphatidyltransferase; G1P, sn-glycerol-1-phosphate; G3P, sn-glycerol-3-phosphate; SRP, signal recognition particle; Tat, twin-arginine translocation.

# Portrait de LUCA - nature des membranes

Les hypothèses concernant la nature des membranes de LUCA:



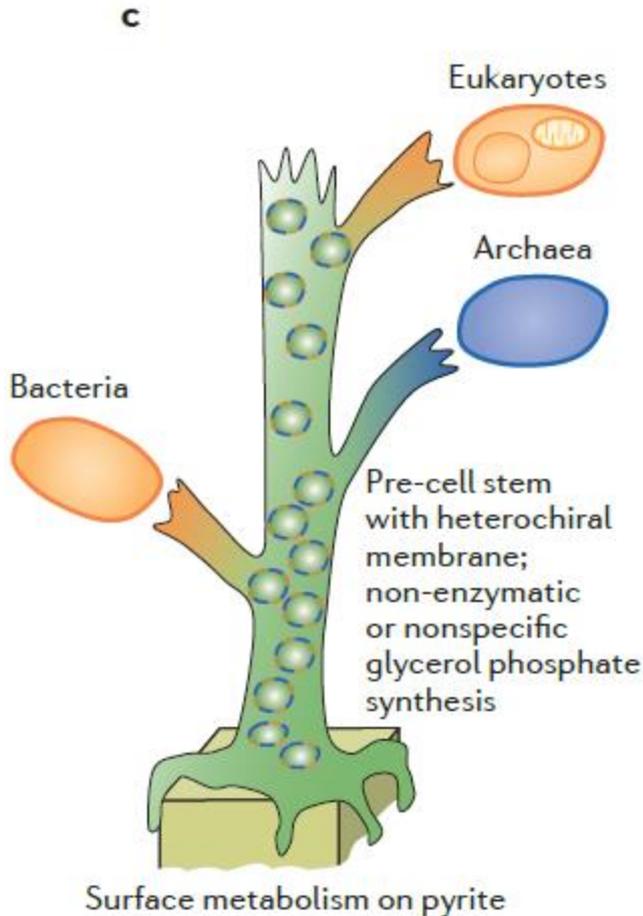
→ Martin et Russell: LUCA a émergé dans une cheminée hydrothermale, il avait des membranes minérales  
→ les lipides spécifiques d'archée et de bactéries sont apparus indépendamment après la séparation des deux lignées

→ Critique: les cheminées hydrothermales ont une durée de vie limitée (entre 100 et 10000 ans), les cellules n'ayant pas de membrane ne peuvent pas se déplacer entre les cheminées

→ LUCA a du apparaître dans une seule cheminée à une vitesse surprenante???

# Portrait de LUCA - nature des membranes

Les hypothèses concernant la nature des membranes de LUCA:



→ Wächtershäuser: les lipides ont émergé très tôt dans l'évolution, les premières membranes ont été synthétisées de manière non-biologique ou alors via les enzymes primitifs sans stéréospécificité

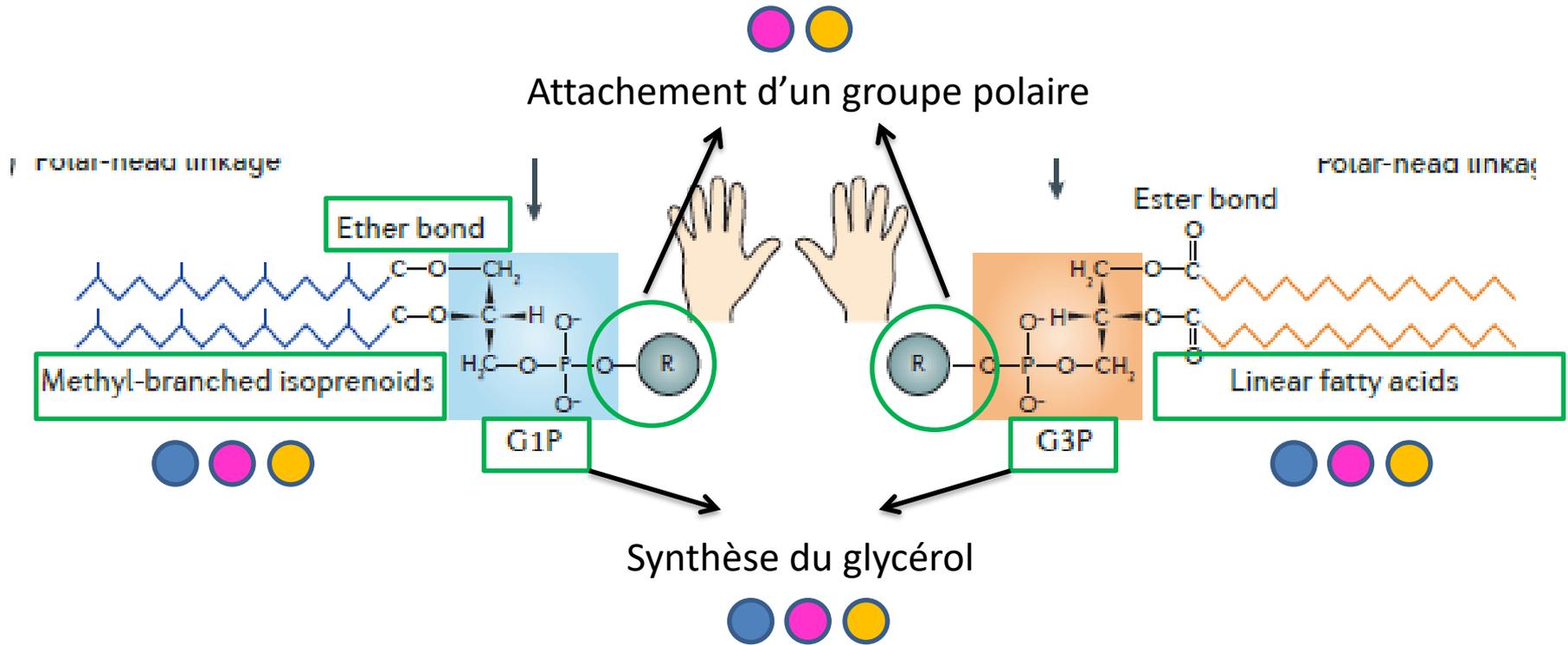
→ les membranes hétérochirales ont été remplacées par les membranes homochirales (plus stables) suite à l'apparition des enzymes stéréospécifiques → apparition d'archées et des bactéries

# Portrait de LUCA - nature des membranes

Table 1 | **Phylogenomic arguments in favour of a cenancestral phospholipid membrane**

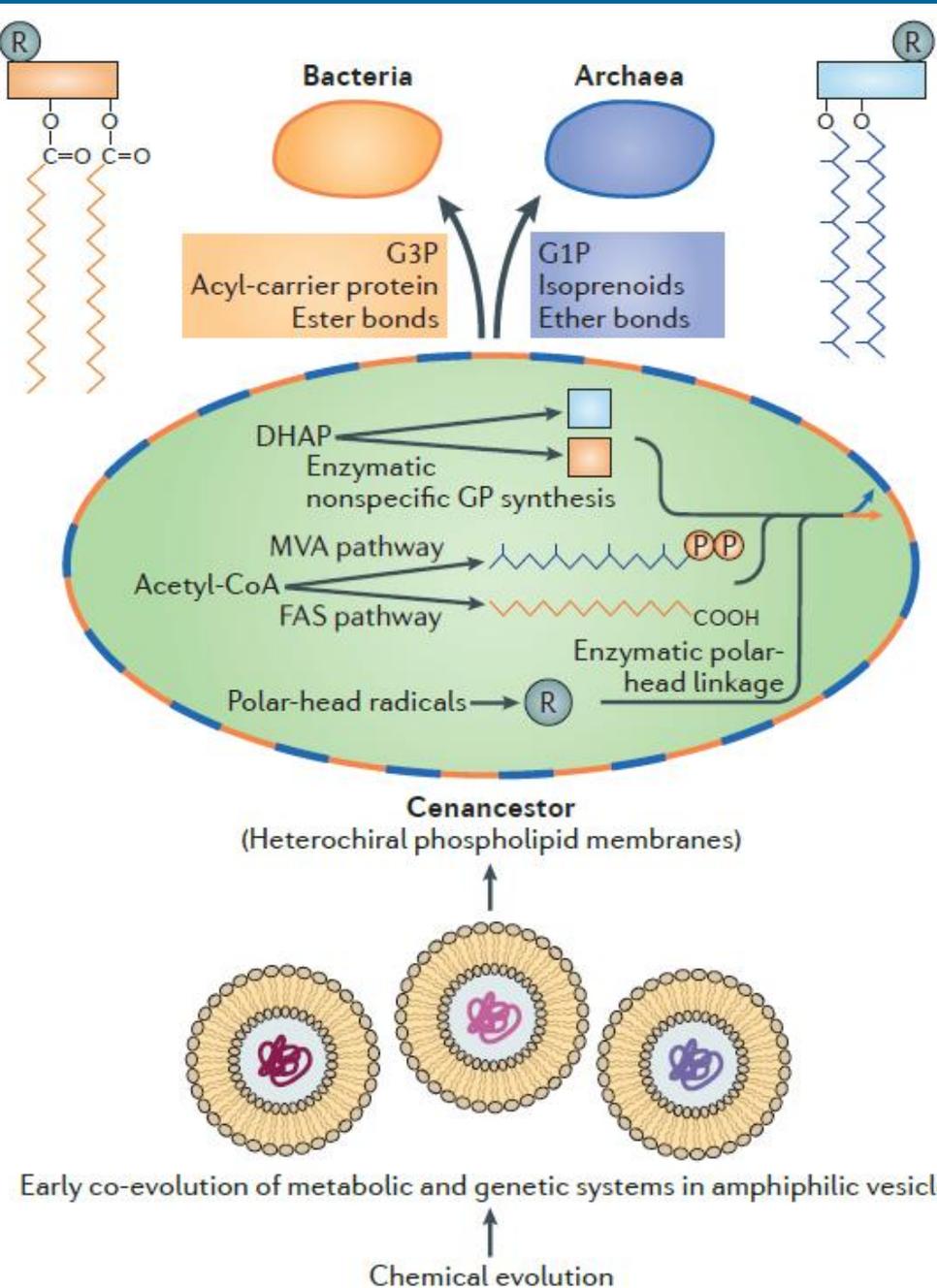
Observations	Implications for the cenancestor
<i>Phospholipid components</i>	
<ul style="list-style-type: none"> <li>• Ancestral presence of at least one member of the two dehydrogenase superfamilies to which the contemporary dehydrogenases synthesizing G1P (archaea) and G3P (bacteria and eukaryotes)<sup>25</sup> belong.</li> </ul>	<ul style="list-style-type: none"> <li>• Possible enzymatic synthesis of G1P and G3P</li> </ul>
<ul style="list-style-type: none"> <li>• Presence of isoprenoids in the three domains of life</li> <li>• Conservation of the mevalonate pathway of isoprenoid synthesis in archaea, bacteria and eukaryotes<sup>26</sup></li> </ul>	<ul style="list-style-type: none"> <li>• Biosynthesis of isoprenoids</li> </ul>
<ul style="list-style-type: none"> <li>• Presence of fatty acids in the three domains of life</li> <li>• Conservation of key enzymes for fatty acid synthesis and degradation in archaea, bacteria and eukaryotes<sup>25,27</sup></li> </ul>	<ul style="list-style-type: none"> <li>• Probable biosynthesis and degradation of fatty acids</li> </ul>
<ul style="list-style-type: none"> <li>• Broad distribution of ether-linked phospholipids and of homologues of the archaeal enzyme superfamily responsible for ether link formation in the three domains of life<sup>30</sup></li> </ul>	<ul style="list-style-type: none"> <li>• Possible hydrocarbon chain attachment to glycerol phosphate (at least) via ether links</li> </ul>
<ul style="list-style-type: none"> <li>• Presence of one representative of the CDP-APT family involved in polar head group attachment in archaea and bacteria<sup>31</sup></li> </ul>	<ul style="list-style-type: none"> <li>• Phospholipid head group attachment in the cenancestor</li> </ul>

# Portrait de LUCA - nature des membranes



● ● ●  
**E**   **A**   **B**

# Membranes



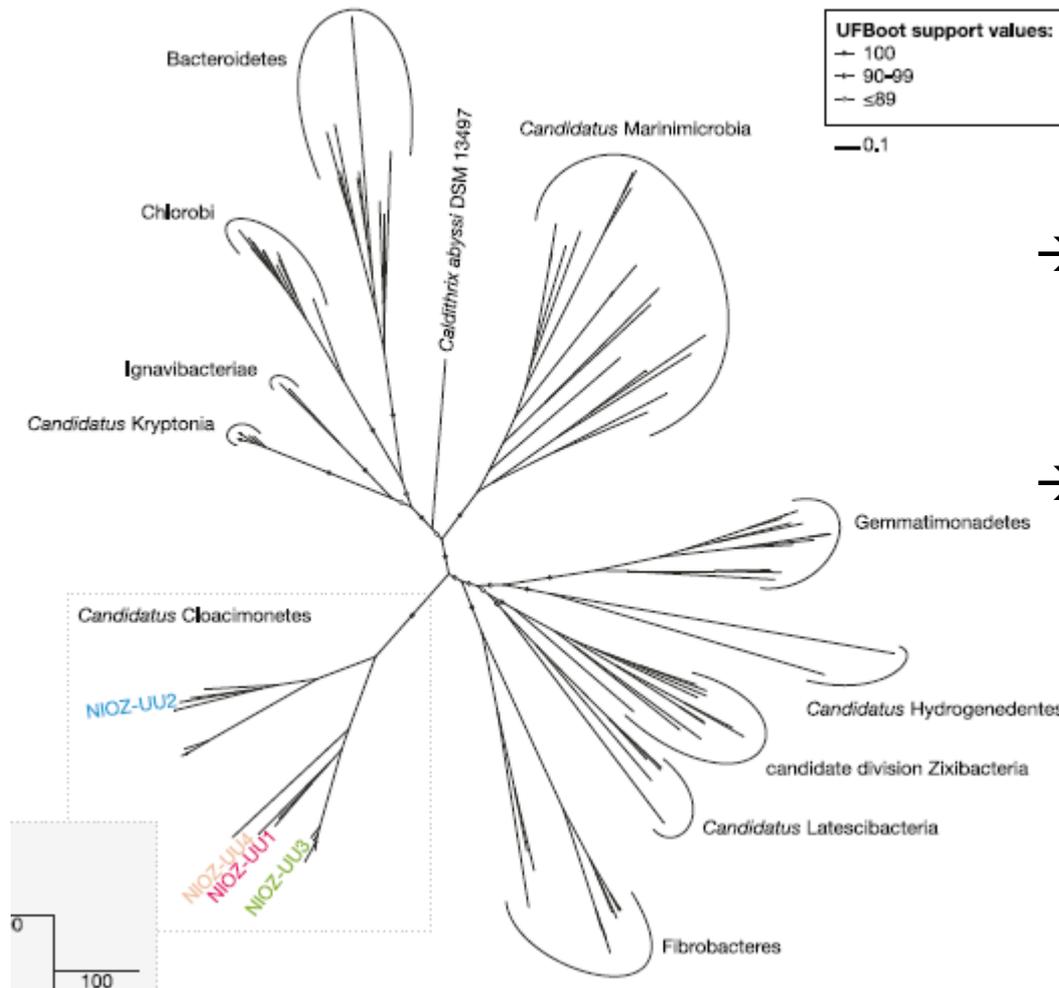
→ Lopez-Garcia et Moreira: LUCA possède une glycerol dehydrogenase (DHAP) non-spécifique → synthèse des lipides hétérochirales

→ LUCA avait le complément des gènes nécessaire pour la synthèse des lipides isoprénoides et d'acides gras (cf. CM sur les Archées)

→ duplication du gène pour la DHAP chez le LUCA et recrutement des paralogues différents dans les deux lignées (archées et bactéries) conduit à une stéréospécificité différente

# Portrait de LUCA - nature des membranes

→ Est-ce que les organismes contenant des membranes mixtes existent ?



→ Reconstruction des MAGs à partir des échantillons prélevés dans la mer noire

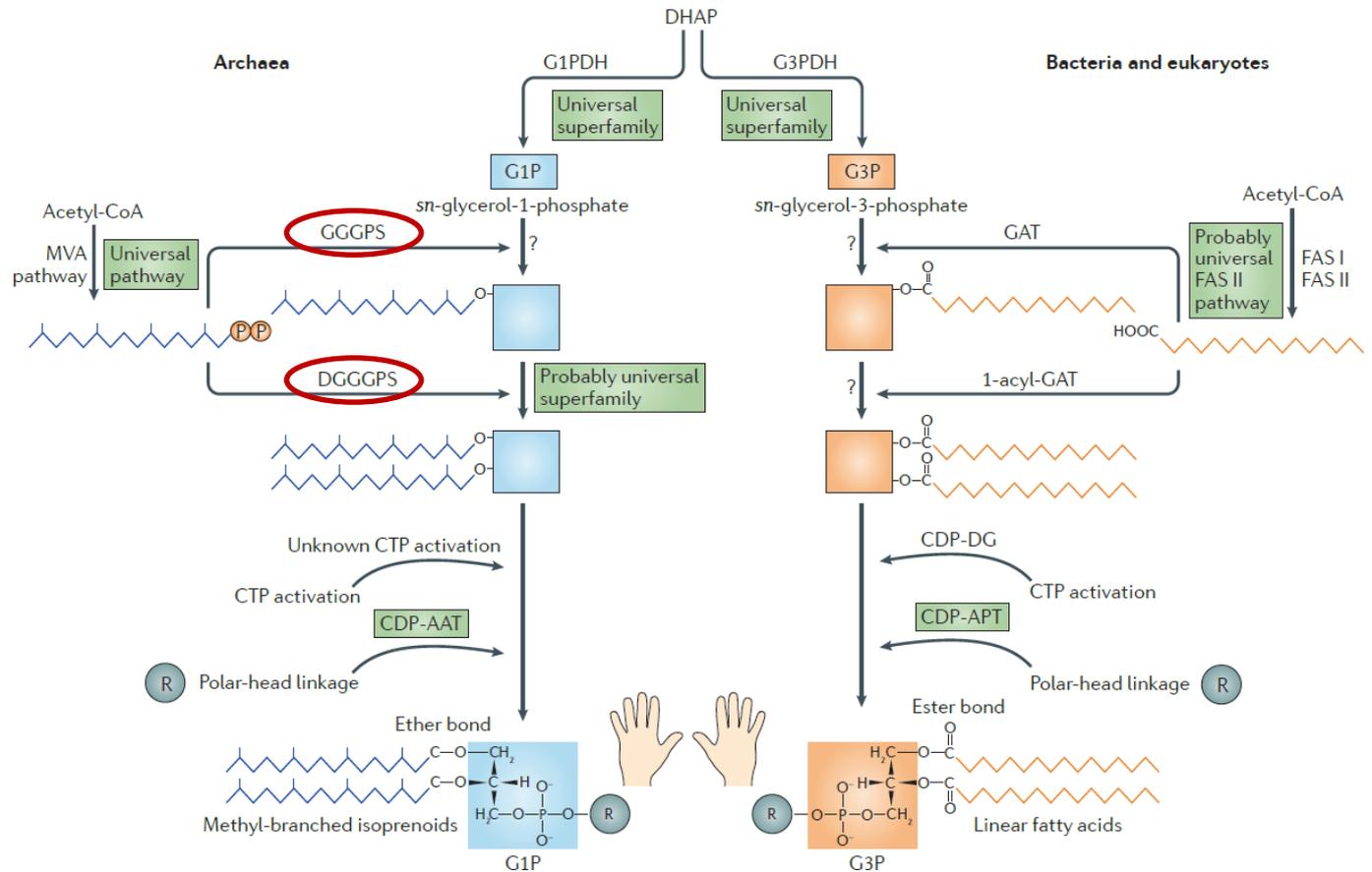
→ 4 MAGs appartenant au clade des Cloacimonetes

Arbre maximum vraisemblance, 43 marqueurs

# Portrait de LUCA - nature des membranes

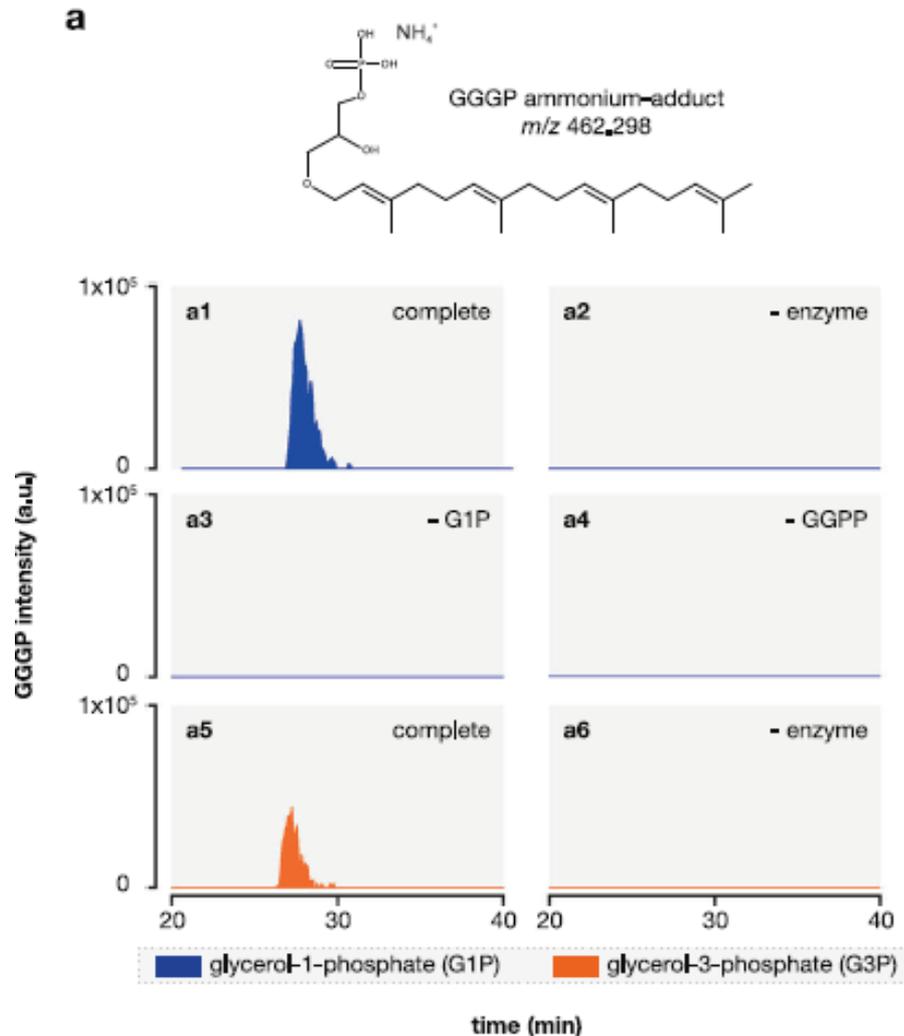
→ Les 4 MAGs codent une voie complète de synthèse de lipides bactériens mais aussi des enzymes impliqués dans la voie de synthèse des lipides isoprénoides !

Rajout de chaînes d'isoprène sur la molécule de glycérol



# Portrait de LUCA - nature des membranes

→ L'expression hétérologue et purification de GGGPS – test d'activité *in vitro*



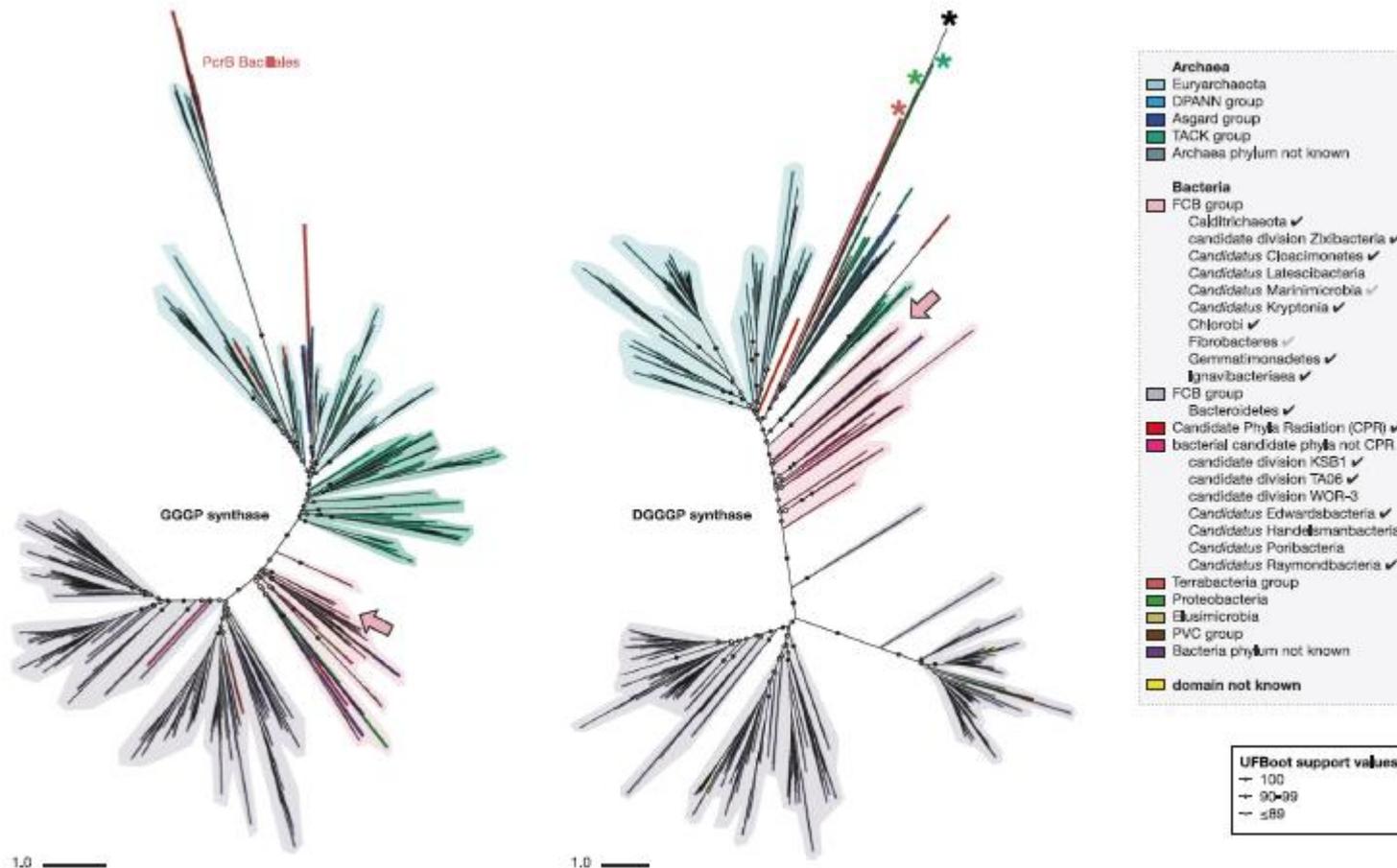
→ Catalyse d'une liaison éther entre chaîne isoprène et G1P ou G3P !



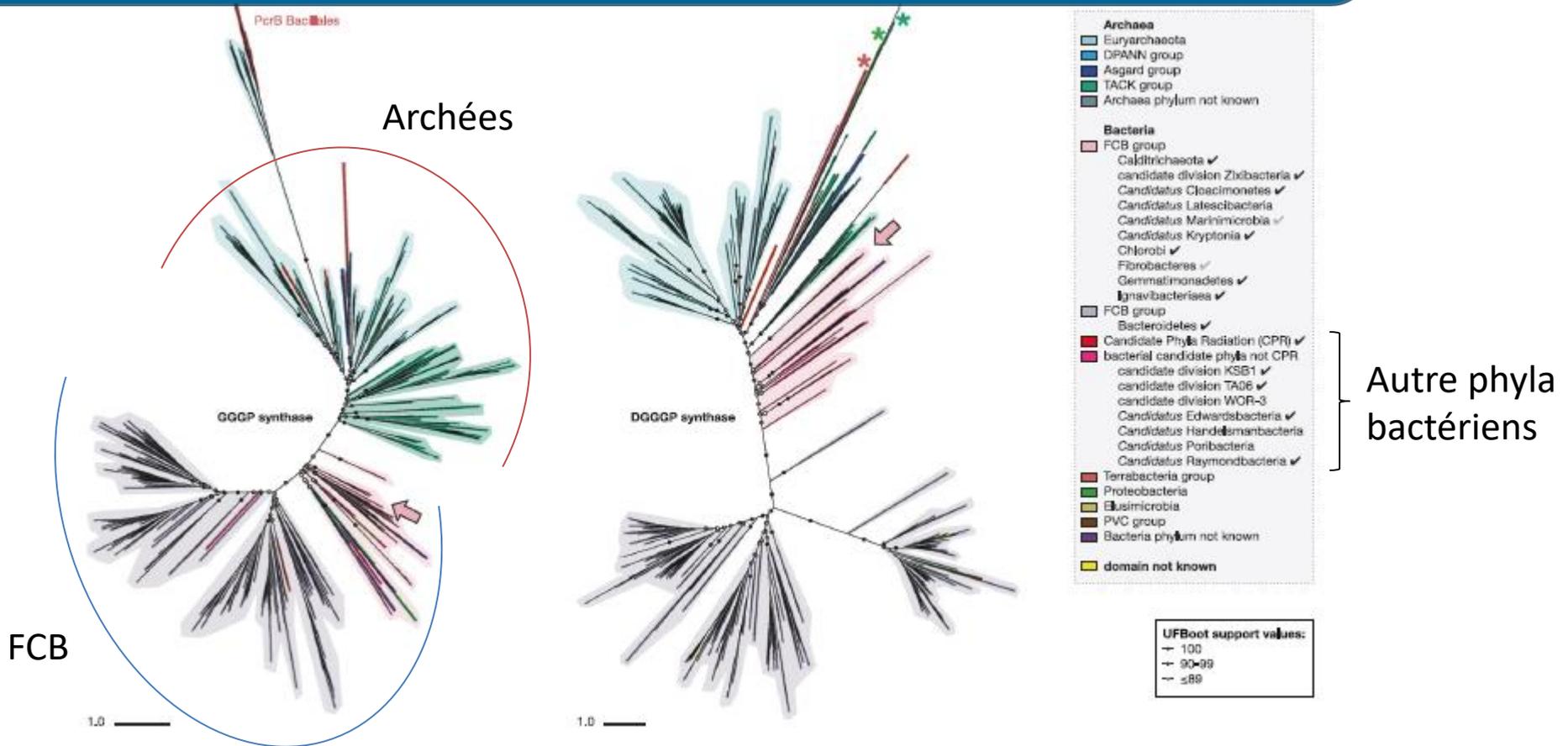
# Portrait de LUCA - nature des membranes

→ Quelle est la distribution de GGGPS et de DGGGPS chez les Bactéries ?

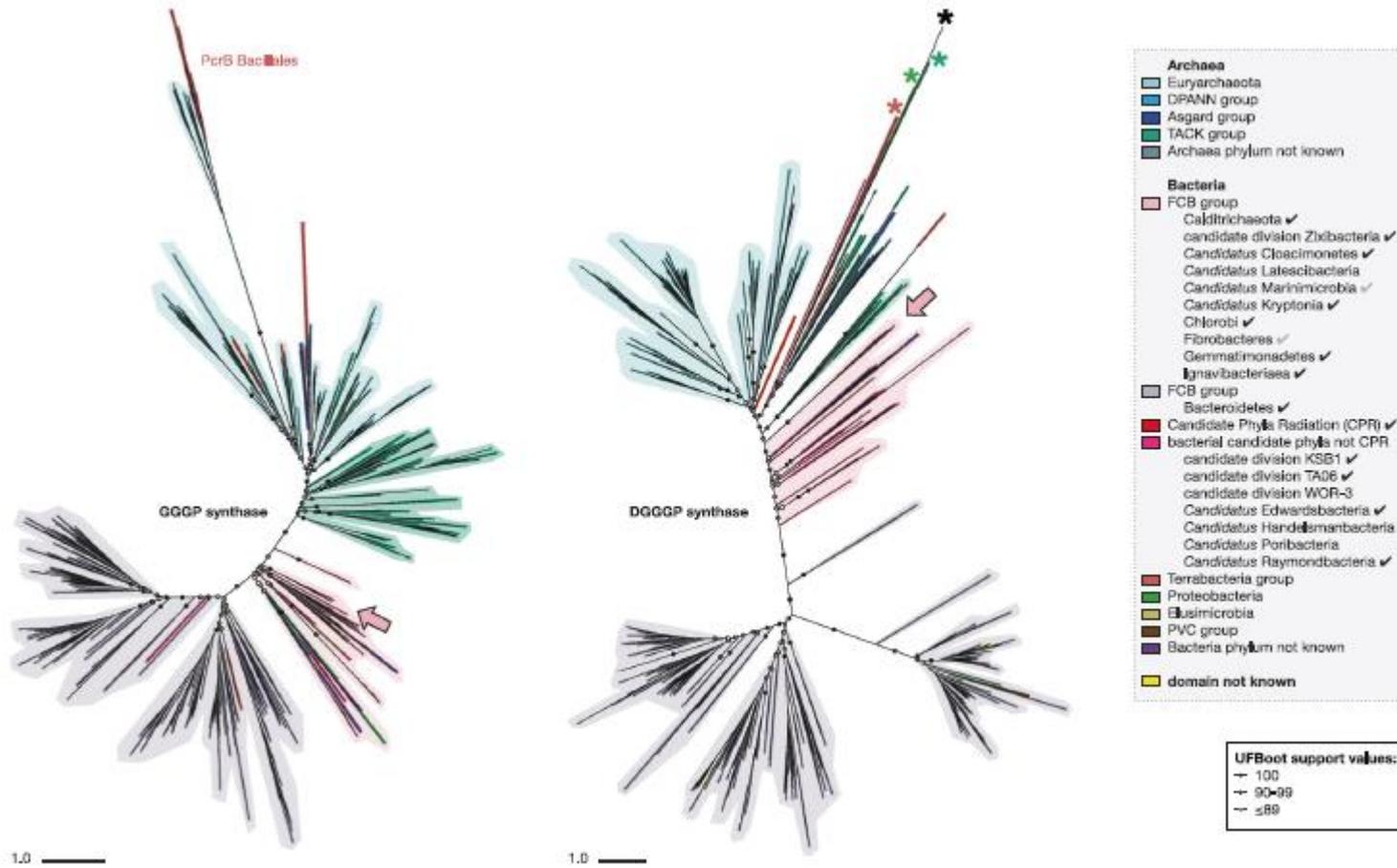
→ Recherche des orthologues par similarité de séquence à partir de ceux identifiés chez les 4 MAGs appartenant au clade des Cloacimonetes



# Portrait de LUCA - nature des membranes



- Large distribution de GGGPS et de DGGGPs chez les FCB Bactéries et (sporadiquement chez d'autres phyla bactérien)
- GGGPS et de DGGGPs sont présents chez l'ancêtre commun de ces bactéries qui ont retenu la capacité de synthétiser les membranes mixtes au cours de leur évolution



→ Les bactéries ont-elles retenu la capacité de former les membranes mixtes depuis LUCA ? Ou ces enzymes ont-elles été transférées à l'ancêtre des bactéries FCB par un ancien HGT depuis les archées ?

→ Dans tous les cas : l'idée d'existence des lipides mixtes est renforcée

→ Les membranes mixtes existent-elles réellement – isolation de ces organismes nécessaire

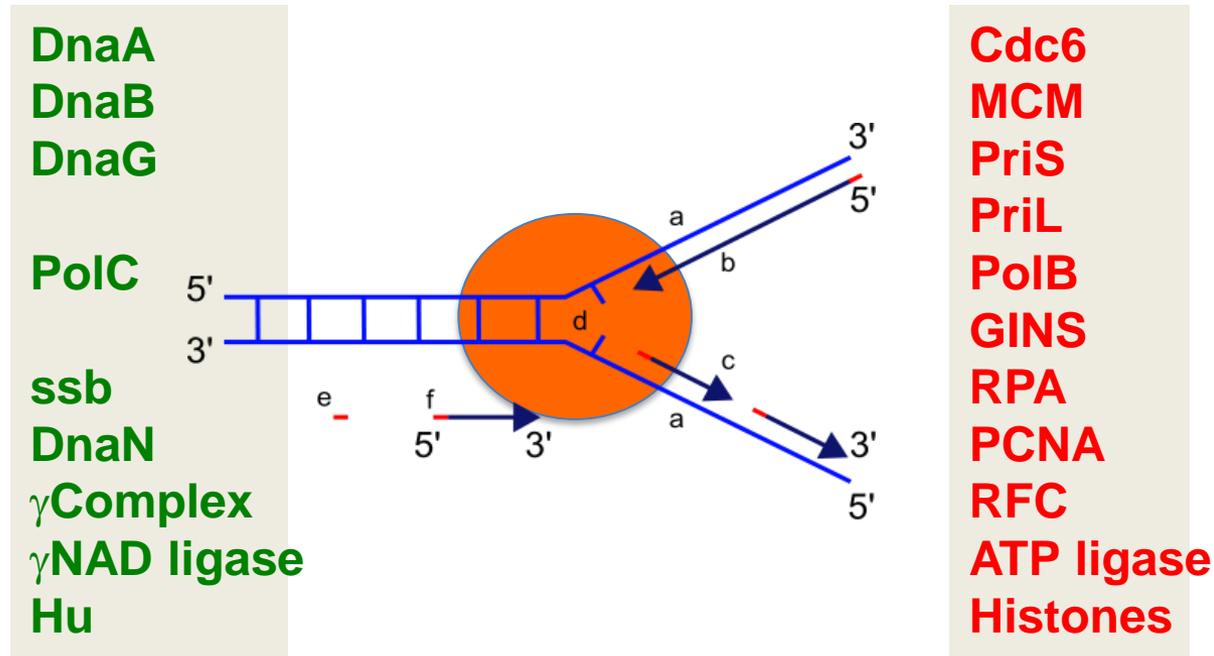
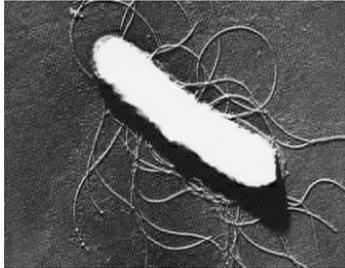
# Portrait de LUCA - ARN ou ADN ?

Functions	Number of genes
<b><i>Translation and associated functions</i></b>	
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
<b><i>Transcription</i></b>	
RNA-polymerase subunits	2
<b><i>Replication/repair</i></b>	
DNA-polymerase subunit, exonuclease, topoisomerase	3
<b>Total</b>	<b>63</b>

Comment LUCA répliquait son matériel génétique?

# Portrait de LUCA - ARN ou ADN ?

Le paradoxe de la réplication d'ADN:

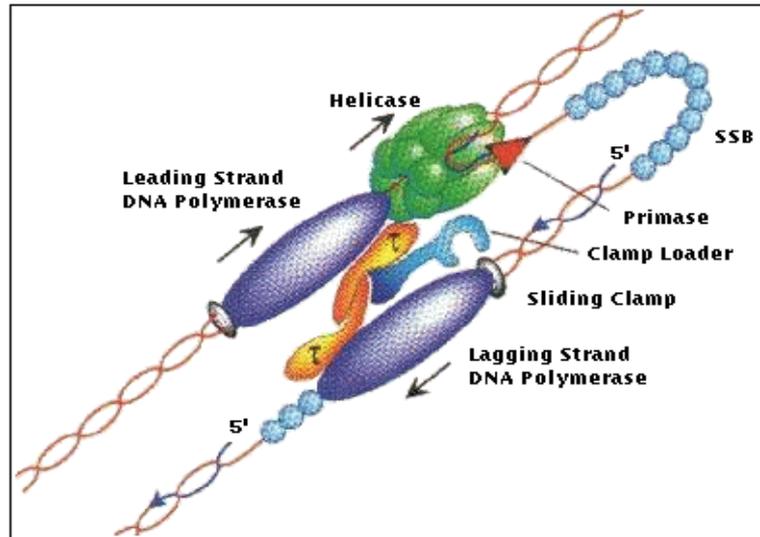


# Portrait de LUCA - ARN ou ADN ?

-les deux systèmes de réplication ne sont pas homologues

*Archaea/Eukarya*

Mcm  
DNA pol (family B)  
Primase

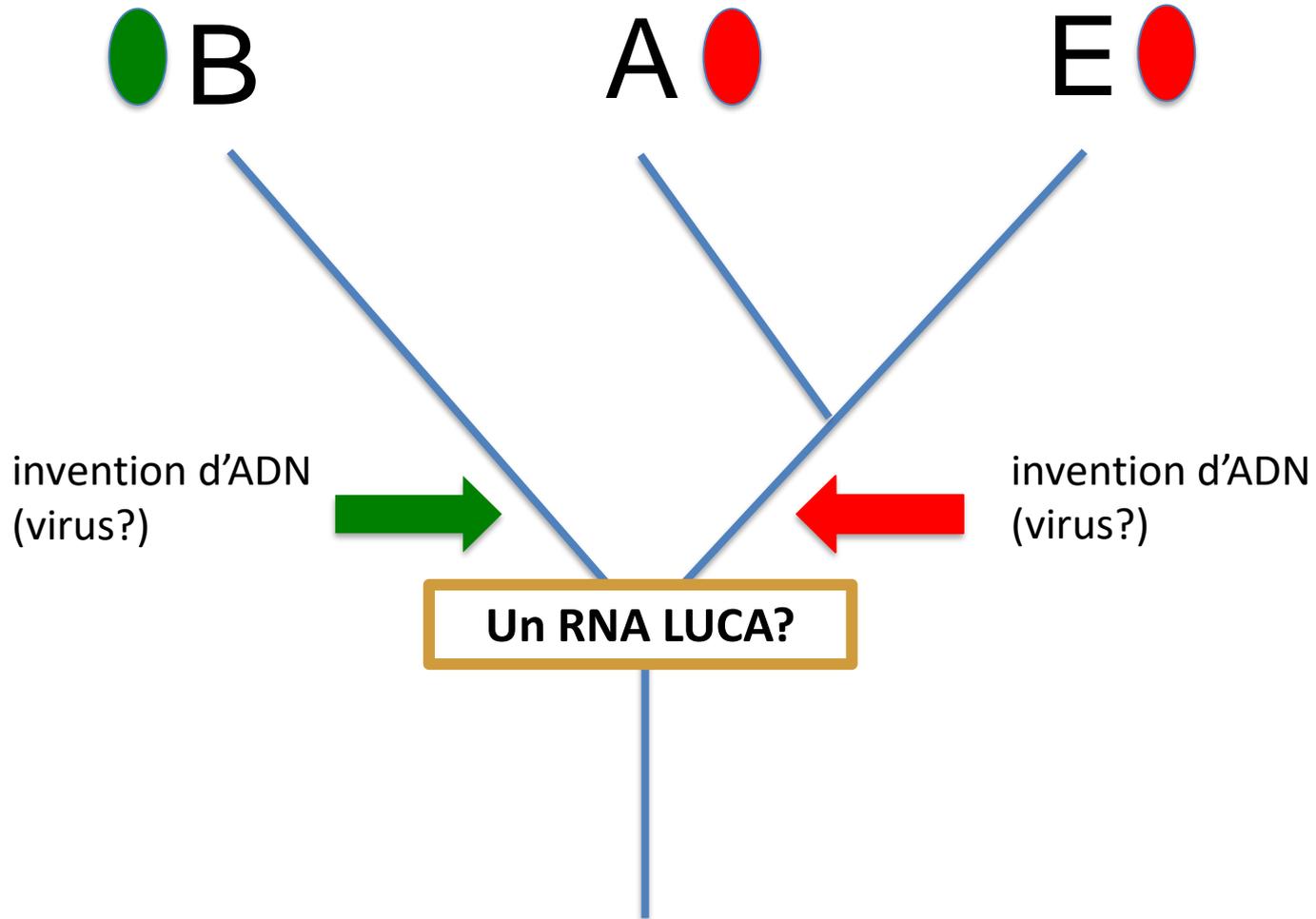


*Bacteria*

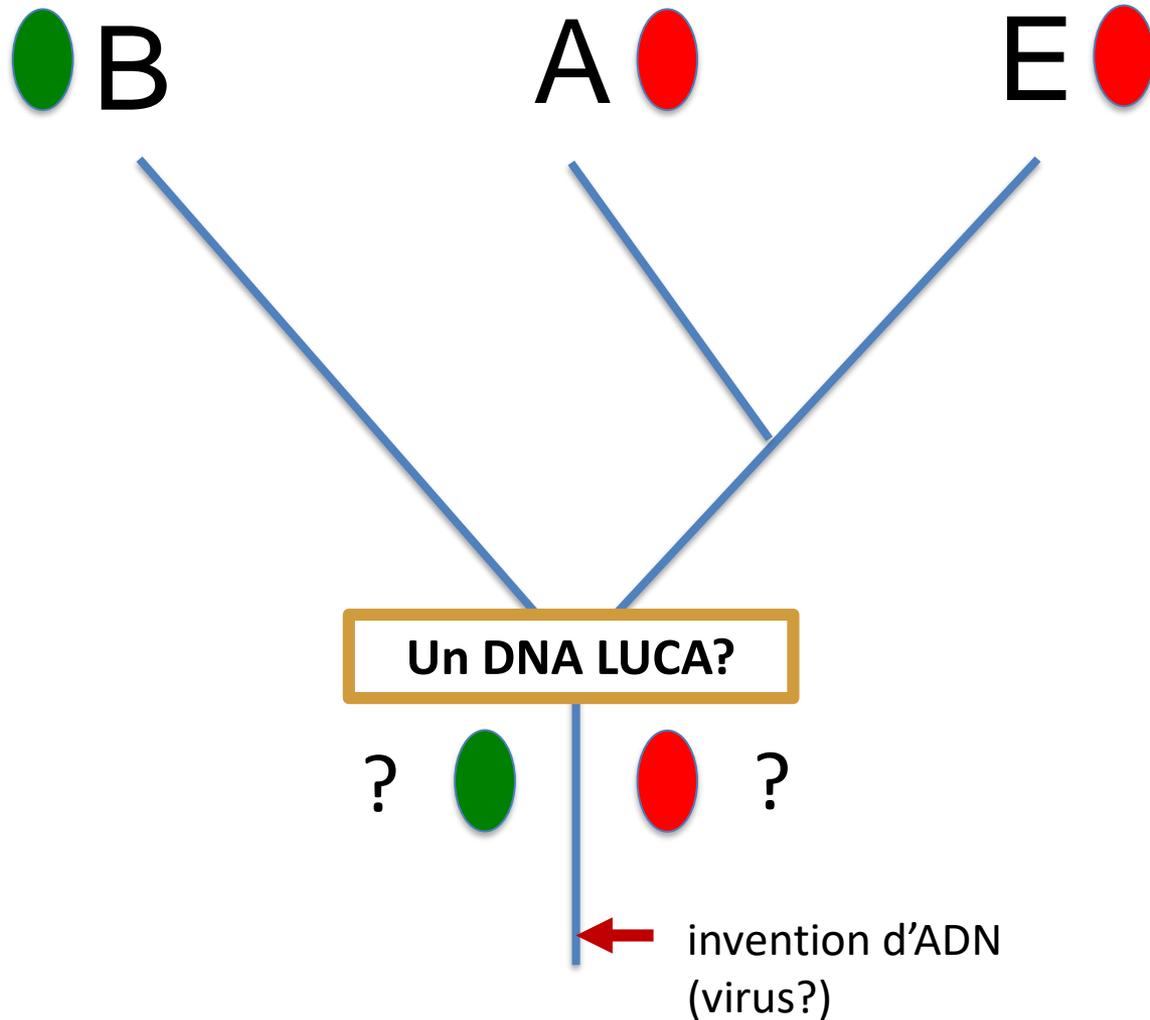
DnaB  
DNA pol III (family C)  
DnaG



# Portrait de LUCA - ARN ou ADN ?



# Portrait de LUCA - ARN ou ADN ?



# Portrait de LUCA - les origines

D'où vient LUCA?

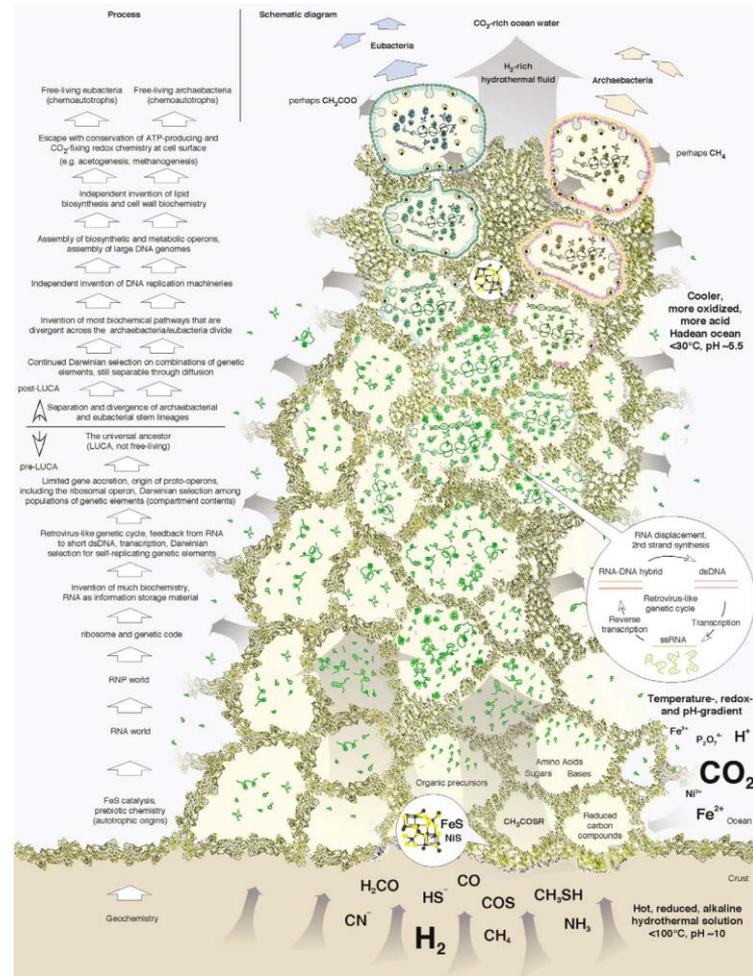


des sources hydrothermales il y a 3.8 Ga



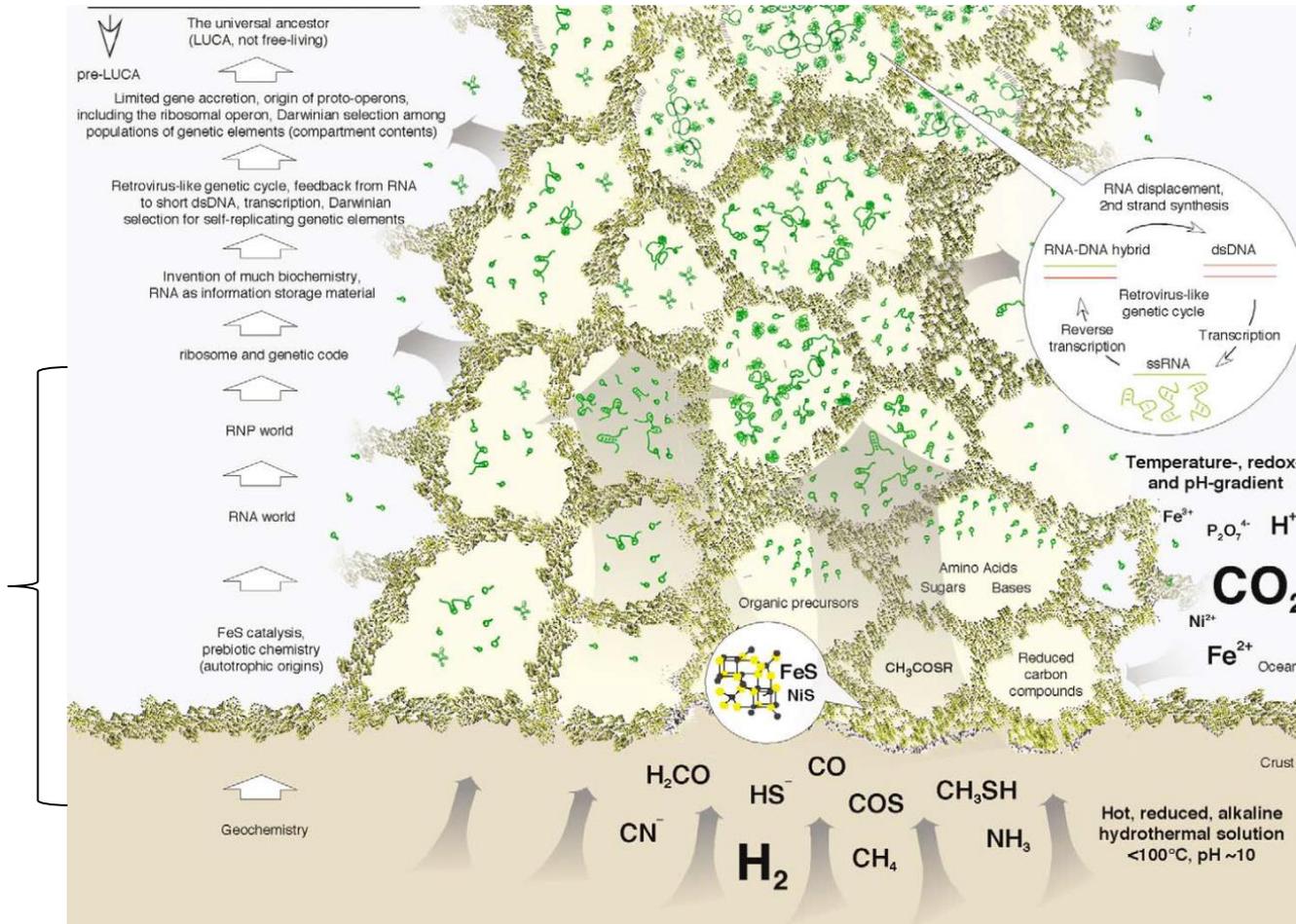
William F. Martin

Eugene Koonin



# Portrait de LUCA - les origines

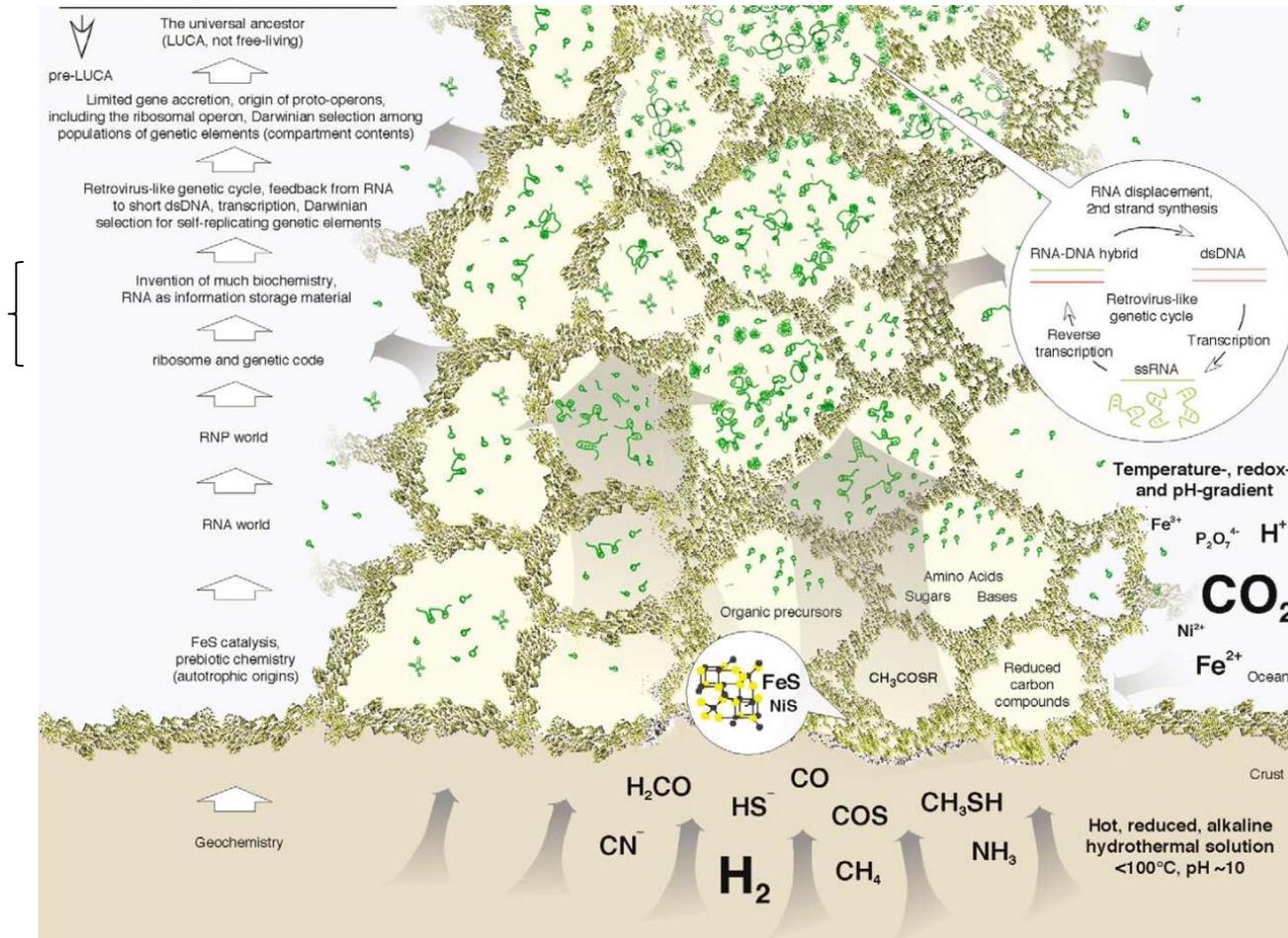
-les molécules organiques simples se combinent pour aboutir à des molécules RNA auto-répliquatives (virus - like)



Martin and Koonin, TRENDS in Genetics Vol.21 2005

# Portrait de LUCA - les origines

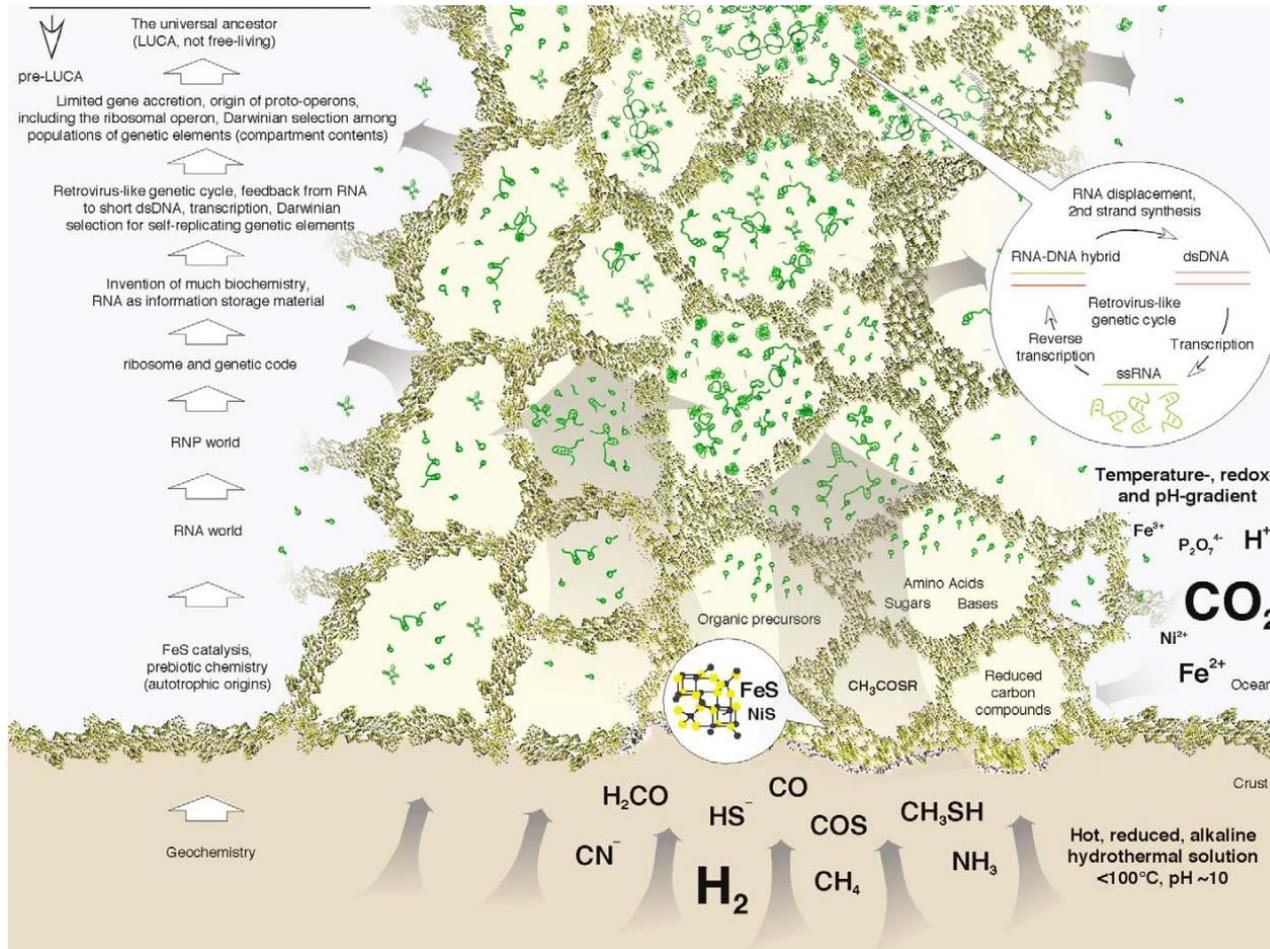
-chaque molécule code pour une ou deux protéines; elles se combinent dans les assemblages de plus en plus complexes – la capacité d'autoréplication est sélectionnée



Martin and Koonin, TRENDS in Genetics Vol.21 2005

# Portrait de LUCA - les origines

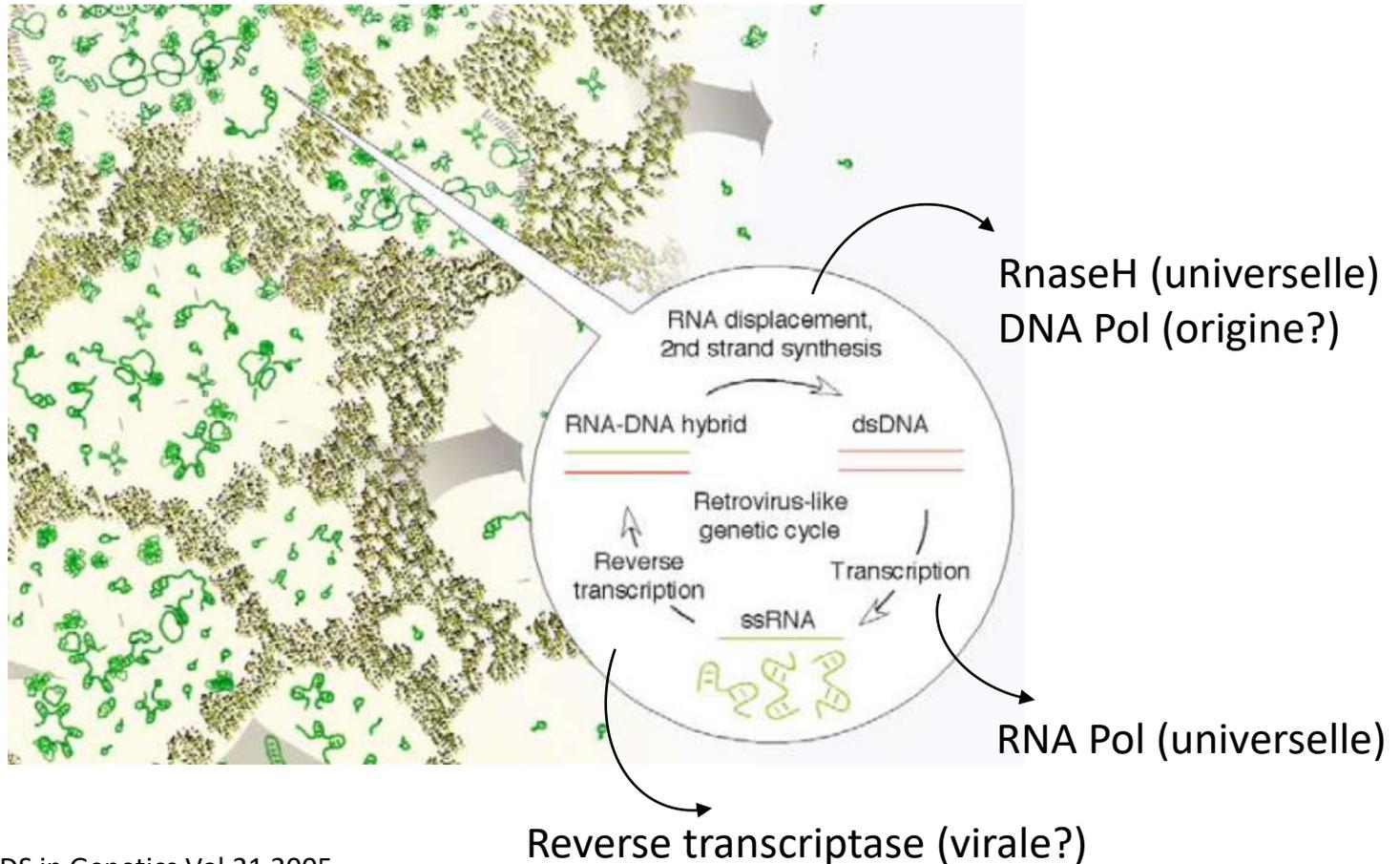
-ces molécules sont confinées dans les cavités de la cheminé (absence des membranes lipidiques? SRP, ATPase?)



Martin and Koonin, TRENDS in Genetics Vol.21 2005

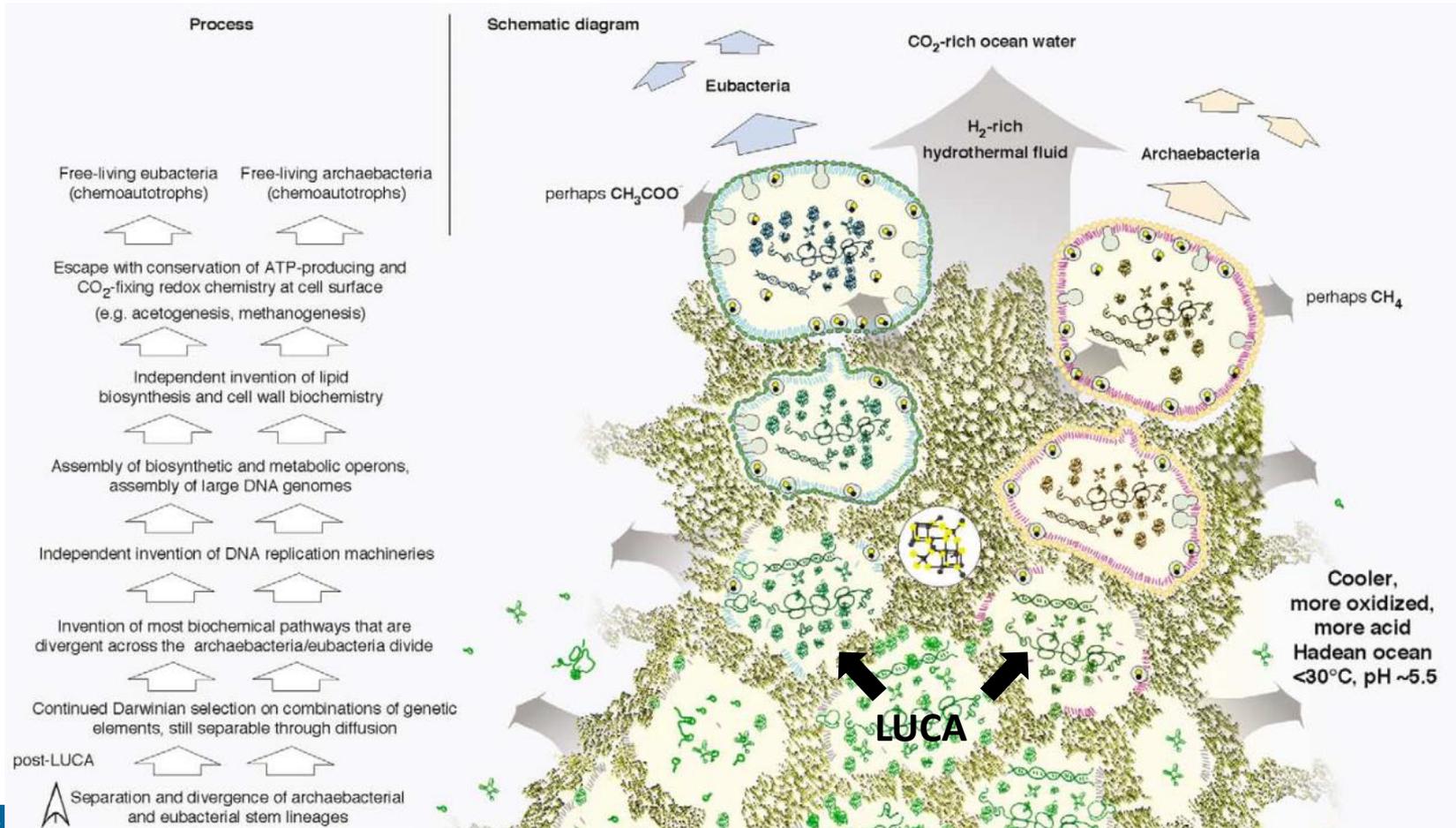
# Portrait de LUCA - les origines

-LUCA réplique son matériel génétique par reverse transcription (RNA → DNA, RNaseH digère l'ARN matrice)

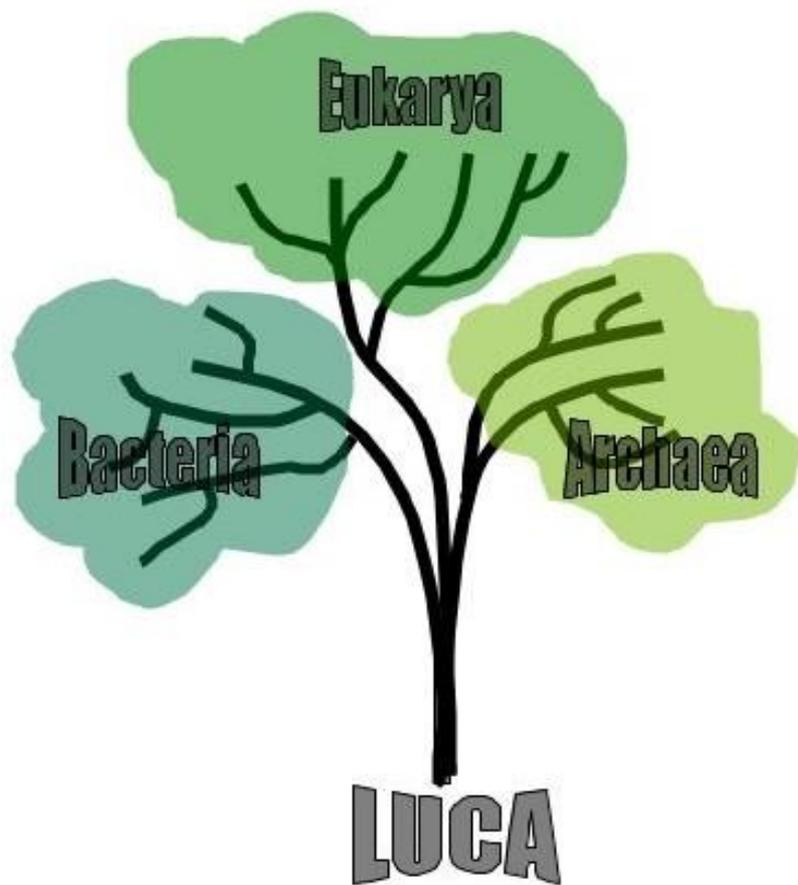


# Portrait de LUCA - les origines

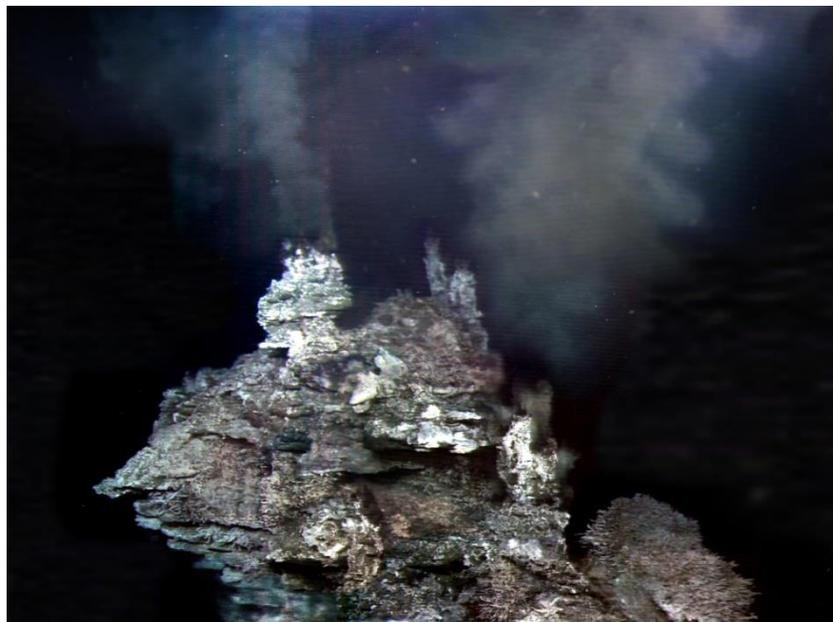
- évolution des systèmes membranaires et de réplication d'ADN dans deux lignées indépendamment → Archées et Bactéries
- les premières Bactéries et Archées sont chimio-litho-autotrophes
- LUCA reste confiné dans la cheminée



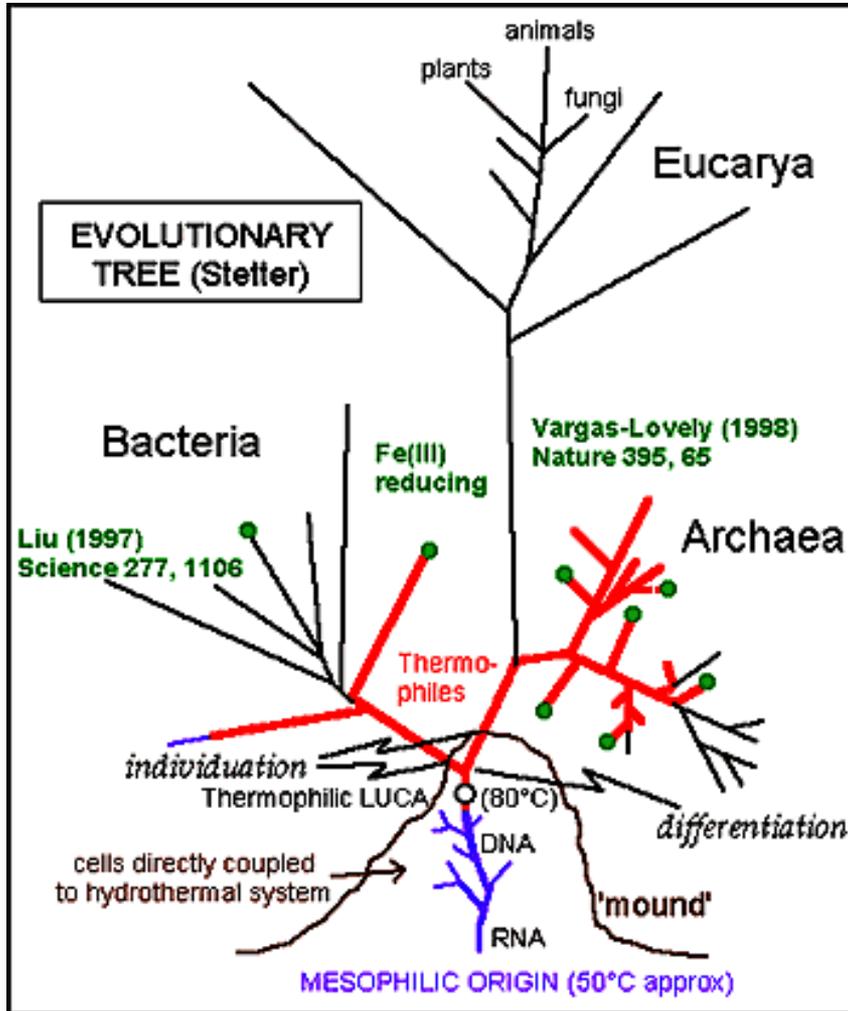
# Portrait de LUCA - un thermophile ?



LUCA était-il un (hyper)thermophile?



# Portrait de LUCA - un thermophile ?

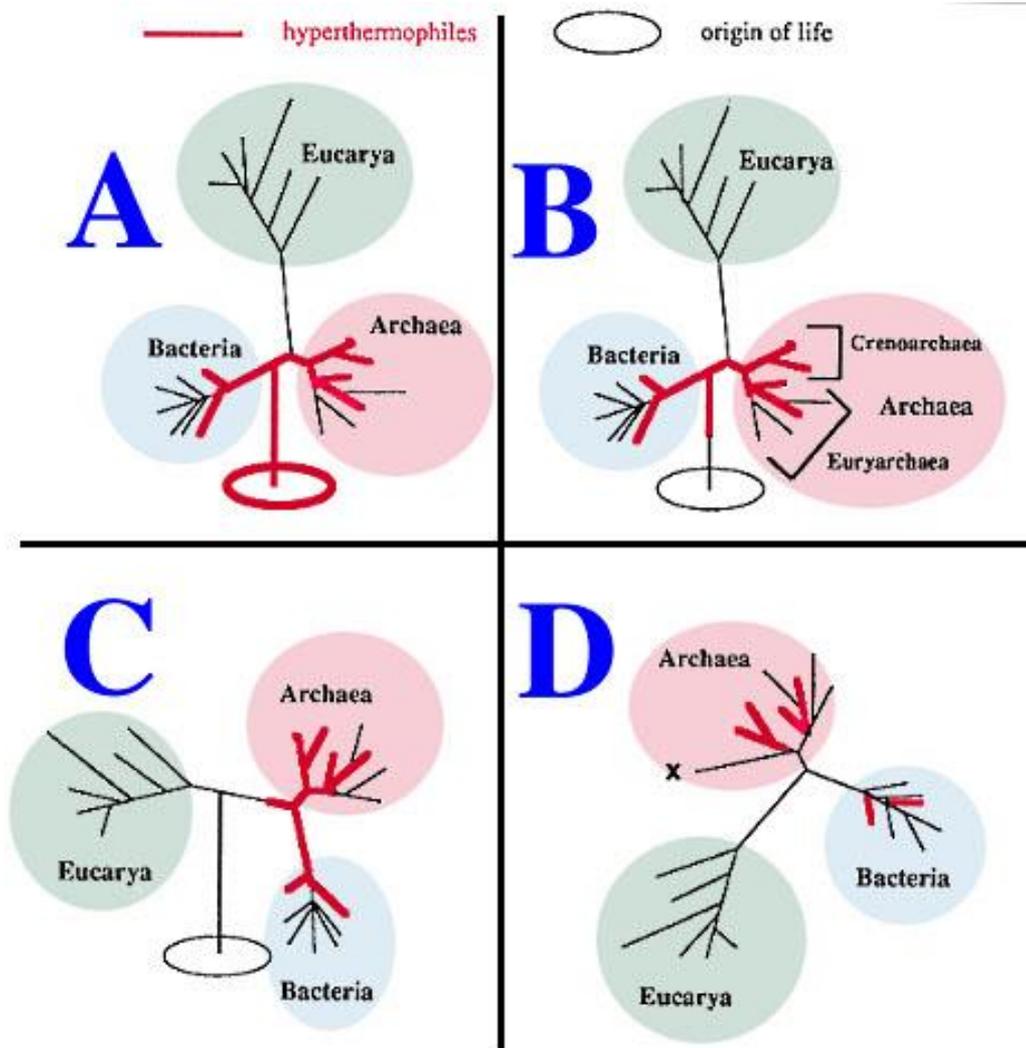


L'arbre universel du vivant (1995) : la racine rouge (C. Stetter)

-Les lignées hyperthermophiles sortent à la base de l'arbre universel

-reflètent la nature hyperthermophile du LUCA?

# Portrait de LUCA - un thermophile ?



Différentes hypothèses sur le positionnement de la racine et la nature de LUCA:

A: LUCA hyperthermophile

B: LUCA mésophile et le dernier ancêtre des Archées et Bactéries est hyperthermophile

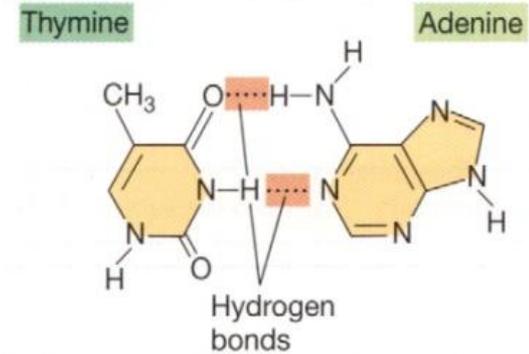
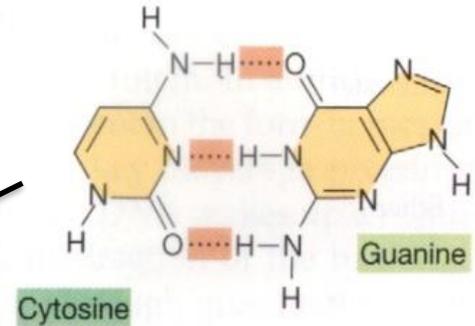
C: LUCA mésophile, la thermophilie est une adaptation plus récente

D: arbre non-enraciné, le dernier ancêtre des bactéries et d'archées est mésophile

# Portrait de LUCA - un thermophile ?

→ le contenu en GC des gènes rRNA est positivement corrélé avec la température optimale de croissance

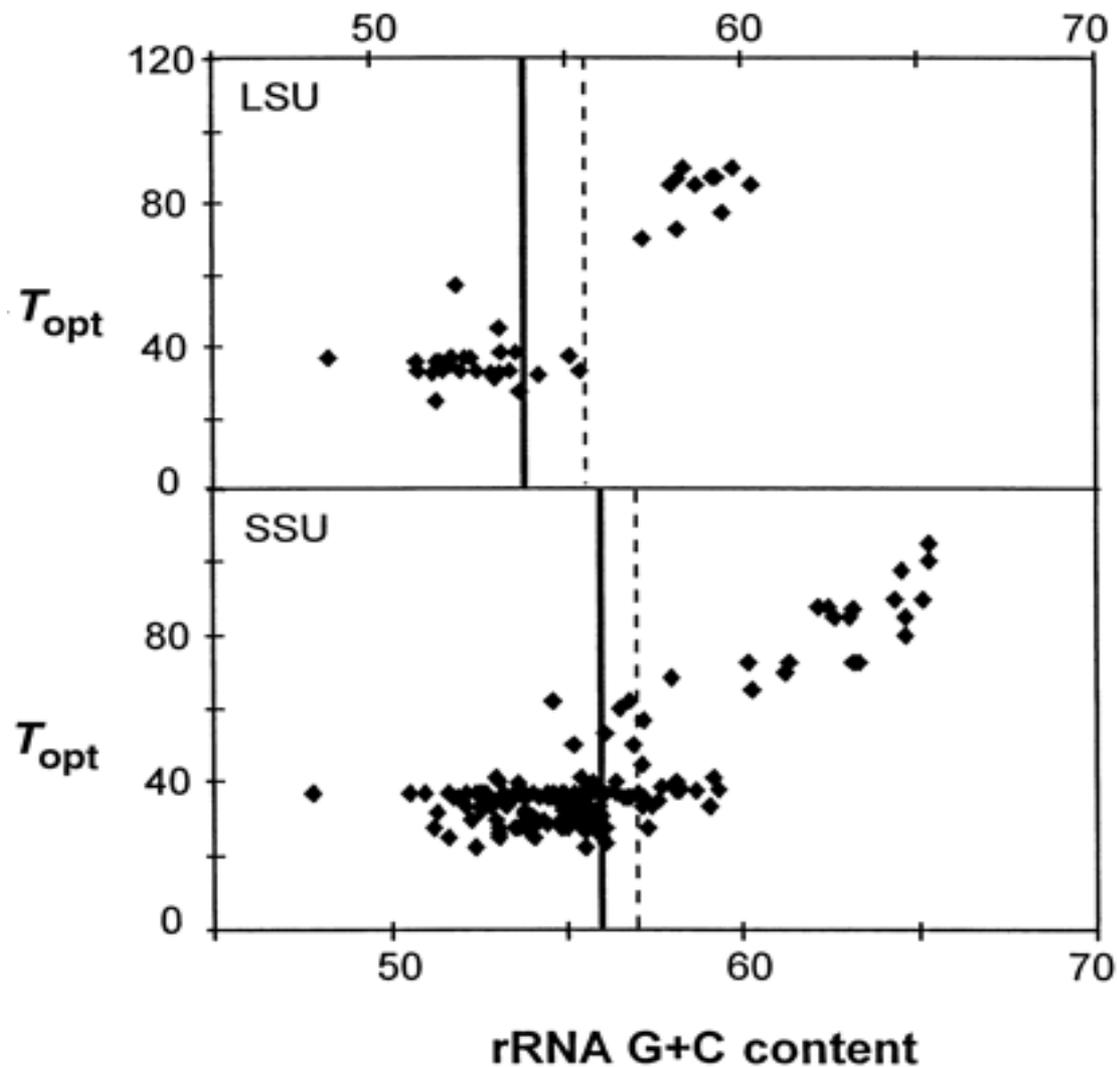
les acides nucléiques à haut % GC sont plus résistantes à la dénaturation



# Portrait de LUCA - un thermophile ?



Manolo Gouy

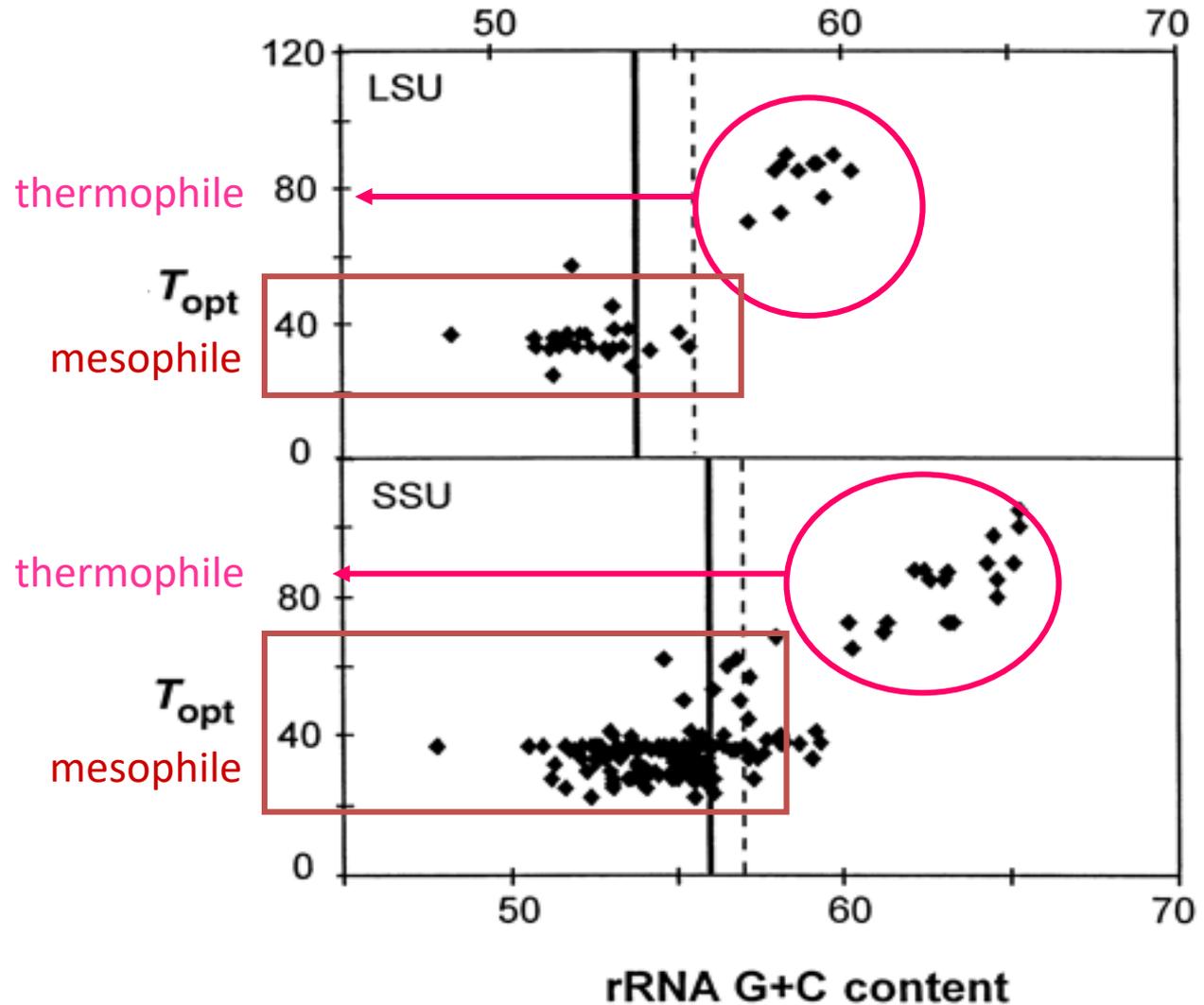


Galtier, Tourasse, Gouy, Science, 1999

# Portrait de LUCA - un thermophile ?



Manolo Gouy



Galtier, Tourasse, Gouy, Science, 1999

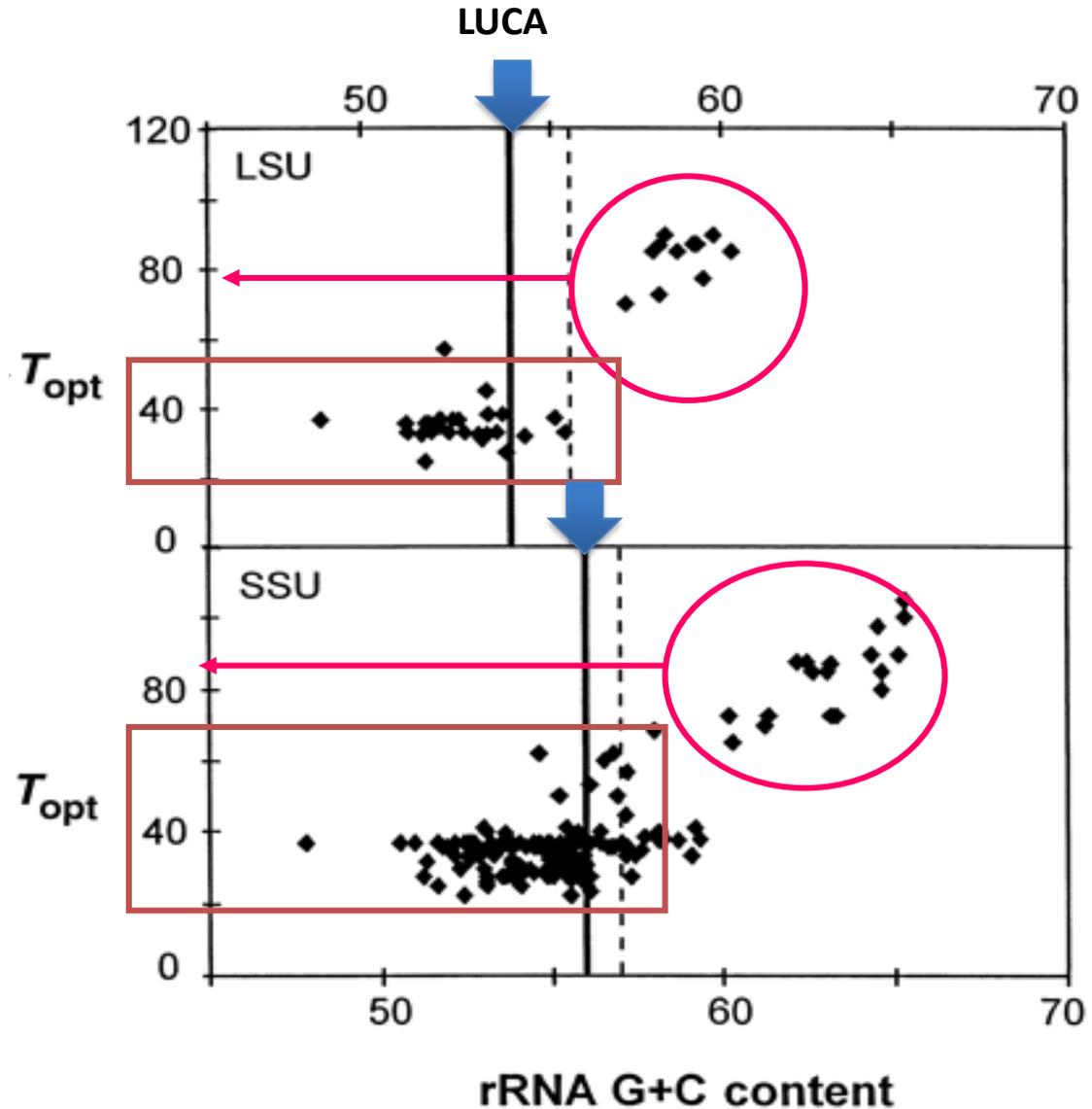
# Portrait de LUCA - un thermophile ?



Manolo Gouy

-reconstruction des séquences ancestrales (LUCA) de rRNA

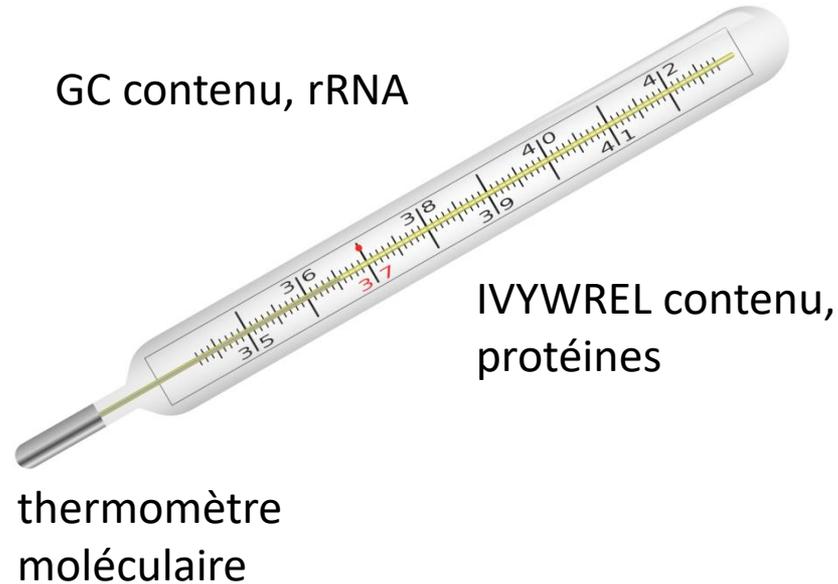
-le contenu en GC de LUCA n'est pas compatible avec une vie à haute température



Galtier, Tourasse, Gouy, Science, 1999

# Portrait de LUCA - un thermophile ?

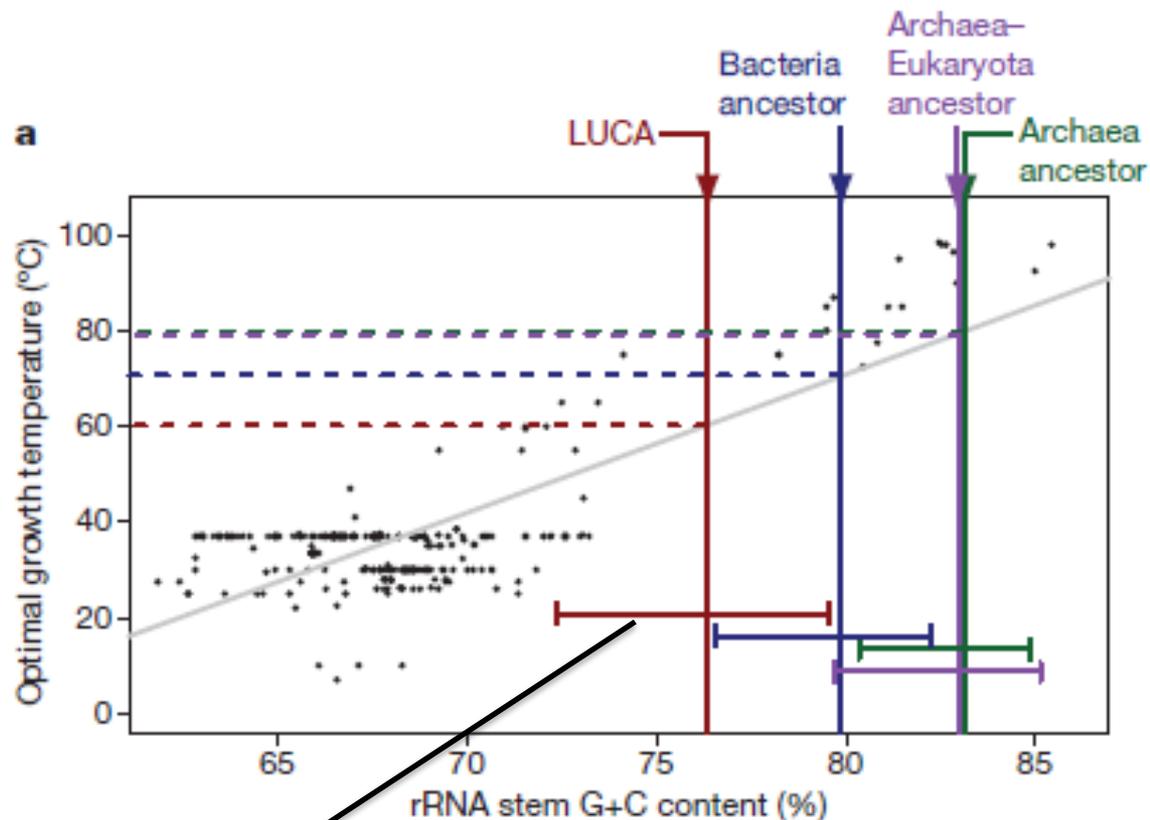
- en 2008 reconstruction des séquences ancestrales de rRNA et des protéines universelles
- corrélation positive entre le contenu en GC et le TOC, les protéines thermorésistantes sont enrichies en IVYWREL



Boussau et al., (2008) Nature 456

# Portrait de LUCA - un thermophile ?

→ Corrélation entre la composition des séquences (rRNA) et la TOC pour les organismes ancestraux



intervalle de confiance 95%

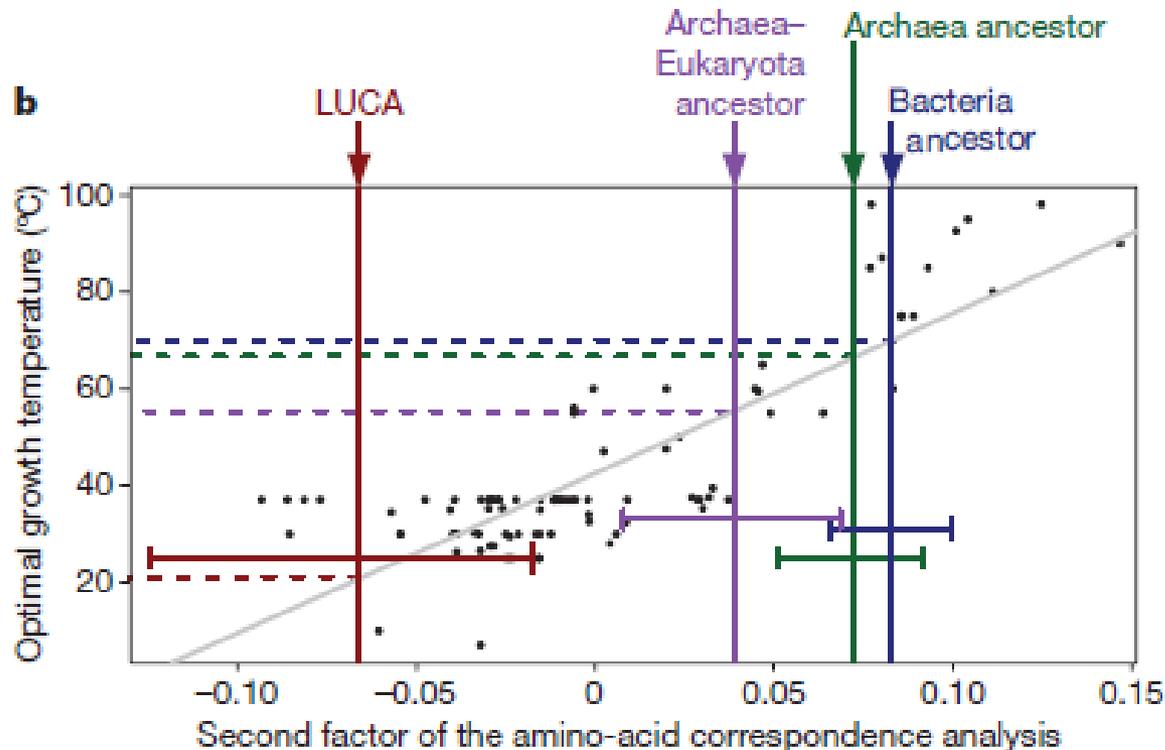
Boussau et al., (2008) Nature 456

# Portrait de LUCA - un thermophile ?

→ Corrélation entre la composition des séquences (protéines) et la TOC pour les organismes ancestraux

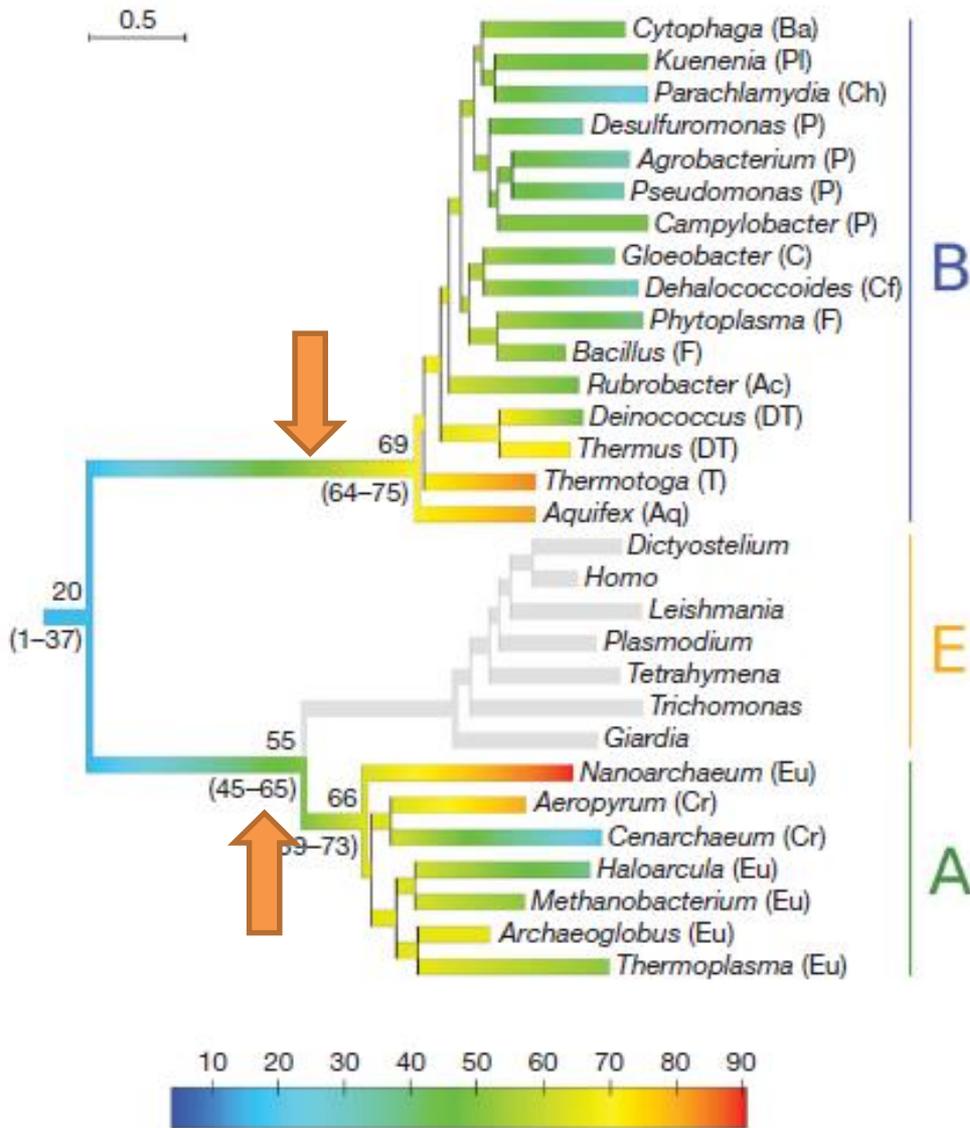
→ LUCA est moins thermophile que ces descendants ( $P < 0.005$ )

→ les ancêtres des bactéries et des archées étaient des organismes thermophiles



Boussau et al., (2008) Nature 456

# Portrait de LUCA – un thermophile ?

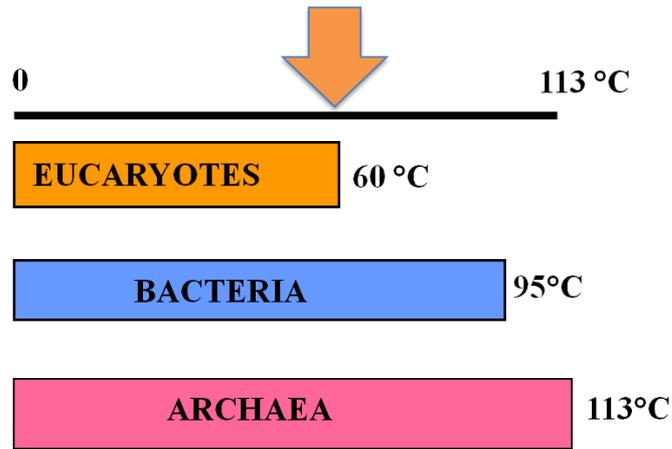


→ LUCA était un organisme mésophile

→ les ancêtres des bactéries et des archées se sont adaptés indépendamment à la thermophilie → à l'origine de leur phénotype procaryote?

# Portrait de LUCA - un thermophile ?

→ pourquoi tous les thermophiles sont procaryotes?

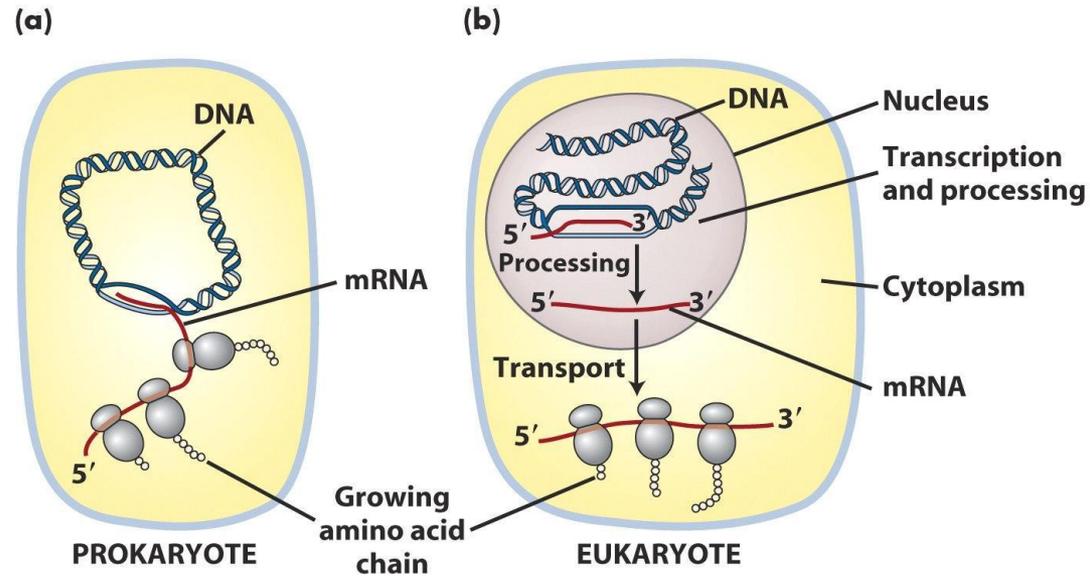


*Oramoeba fumarolia* gen. nov., sp. nov., a new marine heterolobosean amoeboid flagellate growing at 54 °C

Johan F. De Jonckheere<sup>a,b,\*</sup>, Manuela Baumgartner<sup>c</sup>,  
Silvia Eberhardt<sup>c</sup>, Fred R. Opperdoes<sup>a,d</sup>, Karl O. Stetter<sup>c</sup>



Le phénotype procaryote:



- les cellules de petite taille
- couplage de transcription et de traduction
- rapide renouvellement des macromolécules

# Portrait de LUCA - un thermophile ?

→ Temps de demi-vie moyen pour mRNA:

*S. cerevisiae*, 20 min

*Dyctiostelium discoïdum*, 50 min

*E. coli*, 5 min

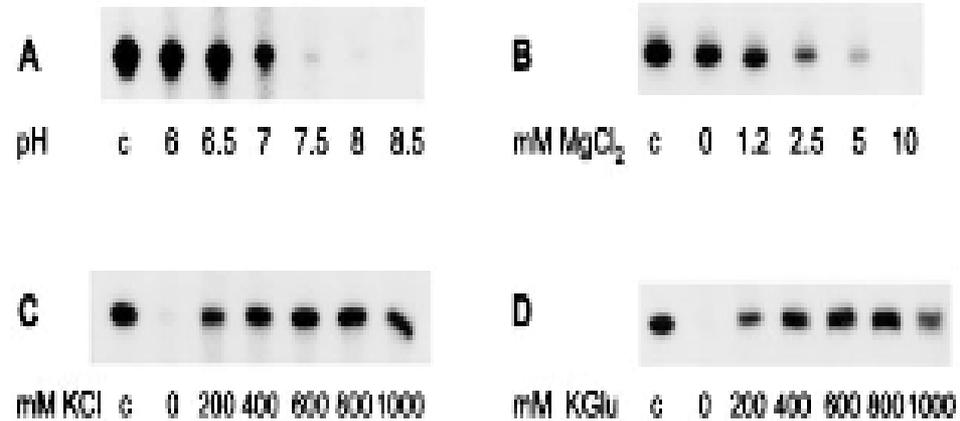
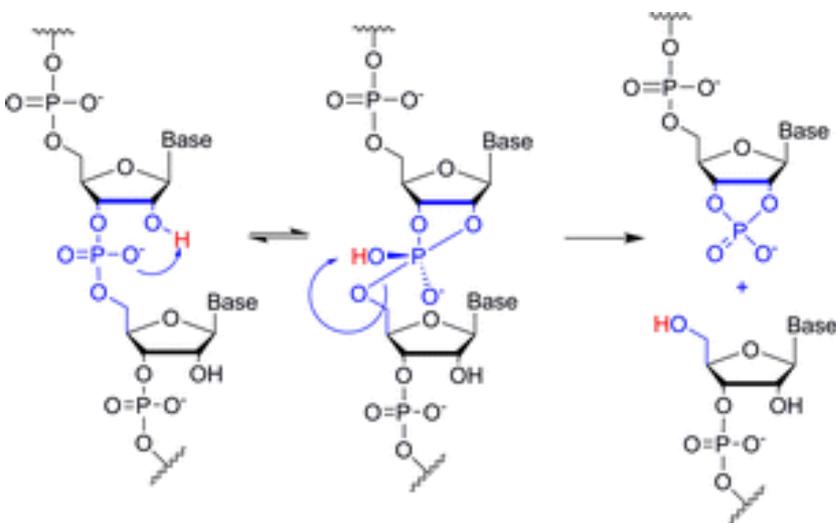
*Sulfolobus sp.*, 4.5 min

Température °C	temps demi-vie (min)*
60	35
70	12
80	4
90	1.3
100	26 sec
110	9 sec

\*en présence de 5mM de MgCl<sub>2</sub>

# Qui était LUCA?

→ mRNA est particulièrement sensible à la thermodégradation (manque de structures secondaires, modifications posttranscriptionnelles)



→ L'hydrolyse de l'ARN est accélérée en présence de magnésium. Le magnésium est indispensable à l'action des ribozymes

# Qui était LUCA?

nature ecology & evolution



Article

<https://doi.org/10.1038/s41559-024-02461-1>

## The nature of the last universal common ancestor and its impact on the early Earth system

Received: 19 January 2024

Accepted: 4 June 2024

Published online: 12 July 2024

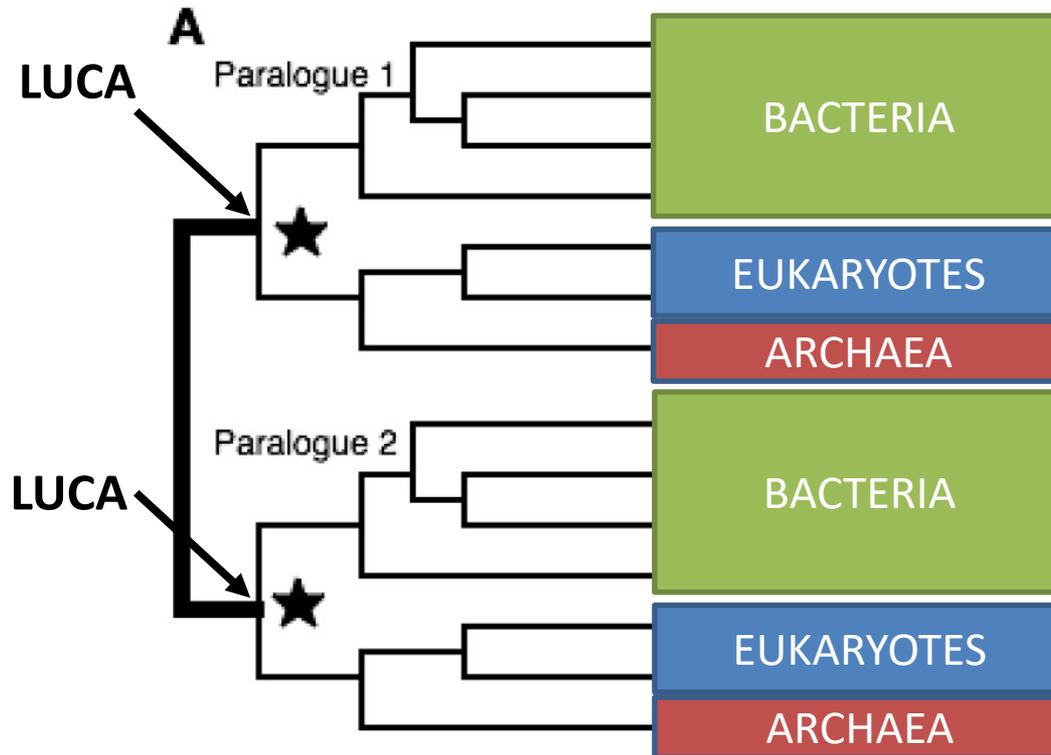
Check for updates

Edmund R. R. Moody<sup>1</sup>, Sandra Álvarez-Carretero<sup>1</sup>,  
Tara A. Mahendrarajah<sup>2</sup>, James W. Clark<sup>3</sup>, Holly C. Betts<sup>1</sup>,  
Nina Dombrowski<sup>2</sup>, Lénárd L. Szánthó<sup>4,5,6</sup>, Richard A. Boyle<sup>7</sup>, Stuart Daines<sup>7</sup>,  
Xi Chen<sup>8</sup>, Nick Lane<sup>9</sup>, Ziheng Yang<sup>9</sup>, Graham A. Shields<sup>9</sup>,  
Gergely J. Szöllősi<sup>5,6,10</sup>, Anja Spang<sup>2,11</sup>, Davide Pisani<sup>1,12</sup>,  
Tom A. Williams<sup>12</sup>, Timothy M. Lenton<sup>7</sup> & Philip C. J. Donoghue<sup>1</sup>

Inférence Bayésienne et l'utilisation des fossiles et **des paralogues universels** pour calibrer l'horloge moléculaire

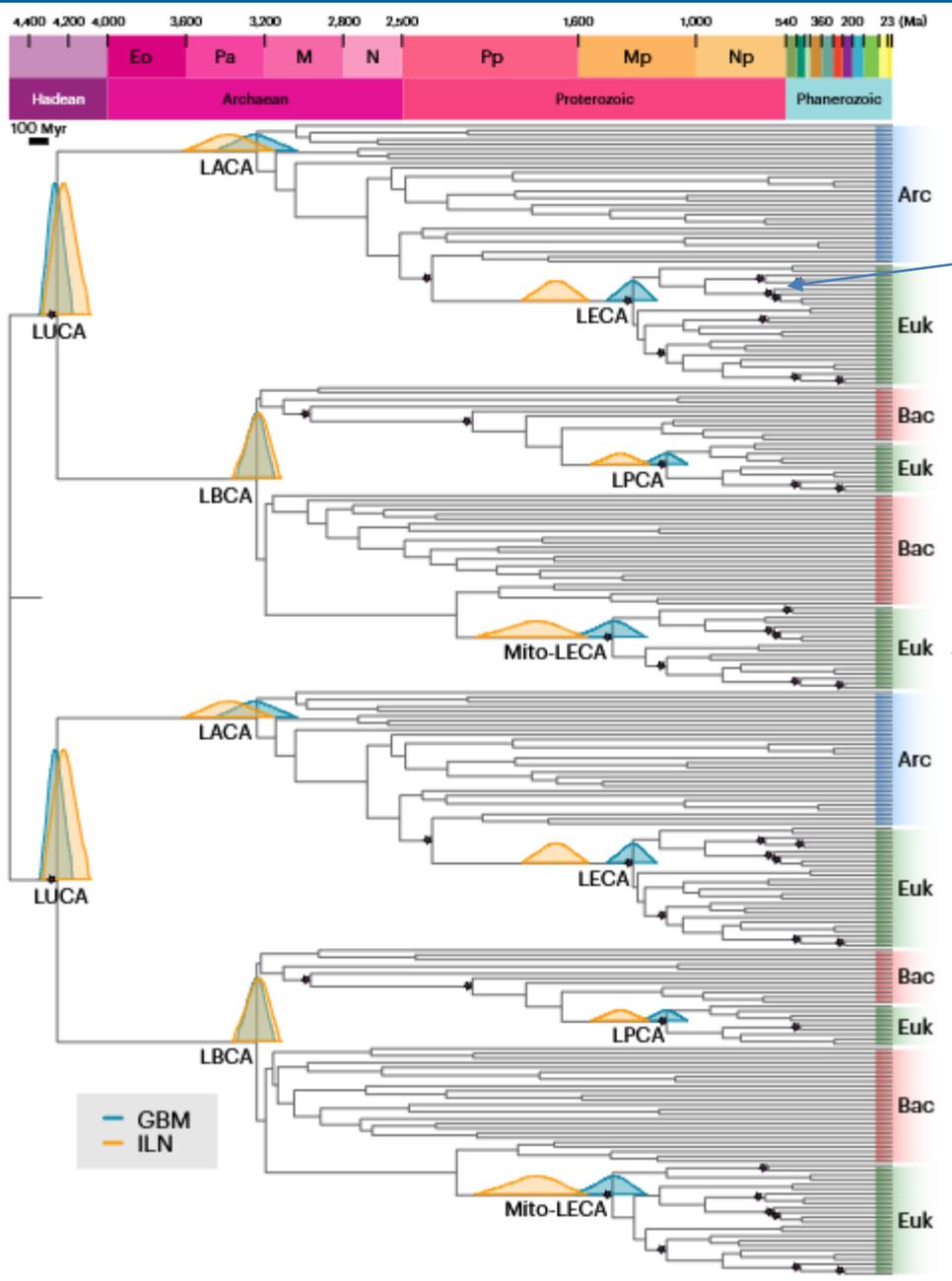
# Qui était LUCA?

Inférence Bayésienne et l'utilisation des fossiles et **des paralogues universels** pour calibrer l'horloge moléculaire



Les paralogues pre-LUCA:

- ✓ SU catalytique et non-catalytique d'ATP synthétase
- ✓ Facteur d'élongation Tu et G
- ✓ SRP (signal recognition particle)
- ✓ Tyr-tRNA synthetase, Trp-tRNA synthetase



Nœuds calibrés  
grâce aux fossiles ★

**LUCA est apparu:  
3.94 Ga to 4.52 Ga**

Plastes

Mitochondrie

— GBM  
— ILN

# Qui était LUCA?

→ Estimation de la probabilité que les familles de gènes trouvés dans la base de données KEGG était présents dans les différents nœuds de l'arbre (réconciliation entre un arbre référence (diapo précédente) et l'histoire évolutive de chaque famille de gènes – méthode probabiliste)



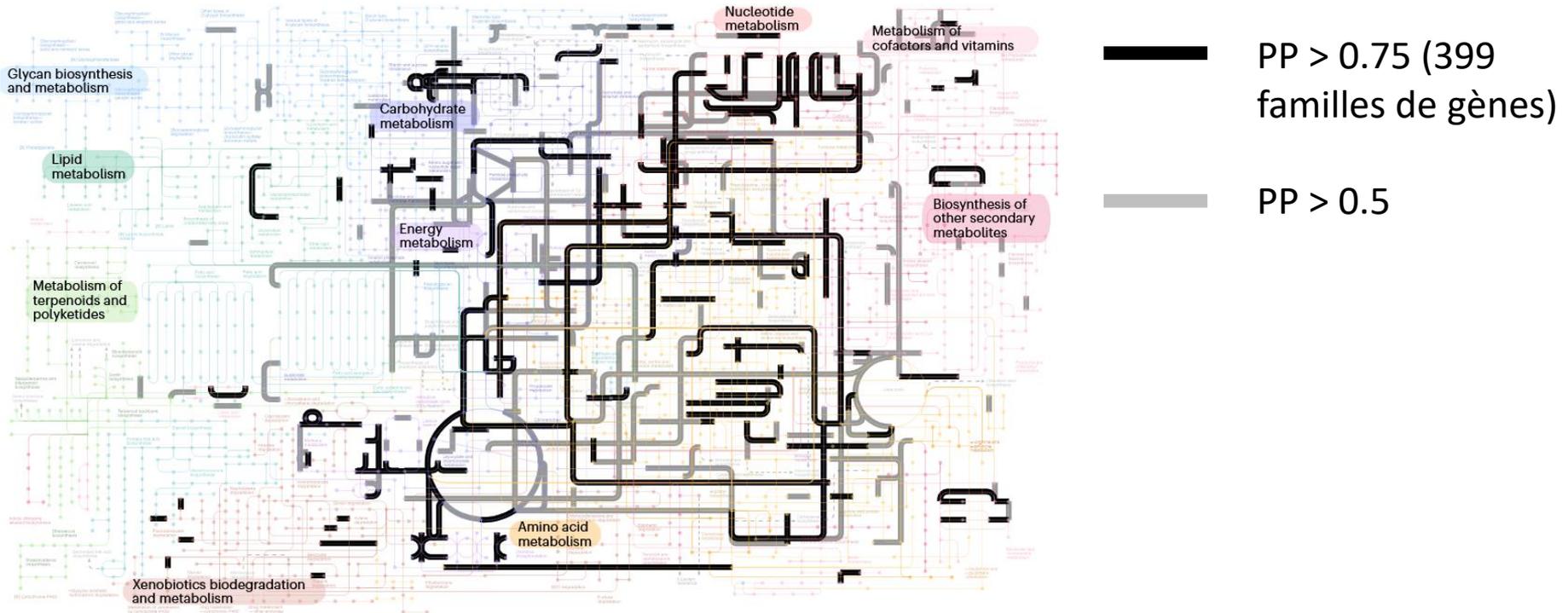
[ [Brite menu](#) | [Copy URL](#) | [Help](#) ]

Change brite type

The screenshot shows the KEGG Orthology (KO) interface. On the left, there is a sidebar with several sections: 'Option' with checkboxes for 'One-click mode', 'Row border', 'shading', and 'Pruning neighbor' (checked); 'Search' with a text input and a 'Go' button; 'ID search' with a text input and a 'Go' button; and 'Join' with a '+' icon. The main content area displays a hierarchical tree of metabolic pathways. The root is '09100 Metabolism', which is expanded to show '09101 Carbohydrate metabolism'. This is further expanded to show a list of specific pathways, each with a right-pointing arrow and a KEGG ID in brackets: '00010 Glycolysis / Gluconeogenesis [PATH:ko00010]', '00020 Citrate cycle (TCA cycle) [PATH:ko00020]', '00030 Pentose phosphate pathway [PATH:ko00030]', '00040 Pentose and glucuronate interconversions [PATH:ko00040]', '00051 Fructose and mannose metabolism [PATH:ko00051]', '00052 Galactose metabolism [PATH:ko00052]', '00053 Ascorbate and aldarate metabolism [PATH:ko00053]', '00500 Starch and sucrose metabolism [PATH:ko00500]', '00520 Amino sugar and nucleotide sugar metabolism [PATH:ko00520]', '00620 Pyruvate metabolism [PATH:ko00620]', '00630 Glyoxylate and dicarboxylate metabolism [PATH:ko00630]', '00640 Propanoate metabolism [PATH:ko00640]', '00650 Butanoate metabolism [PATH:ko00650]', '00660 C5-Branched dibasic acid metabolism [PATH:ko00660]', and '00562 Inositol phosphate metabolism [PATH:ko00562]'. Below this list, other main categories are visible: '09102 Energy metabolism' and '09103 Lipid metabolism'.

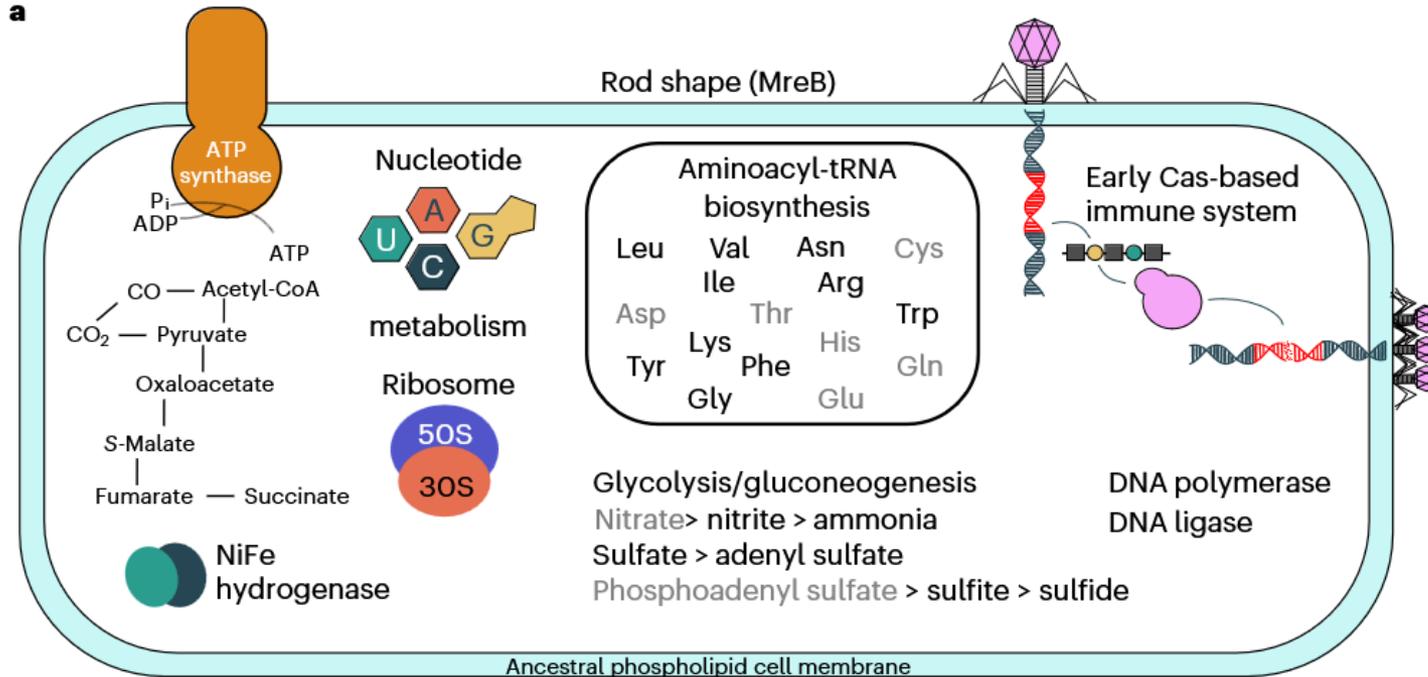
**Cela permet d'inclure bien plus de familles protéique dans l'analyse comparé aux approches de génomique comparative**

# Qui était LUCA?



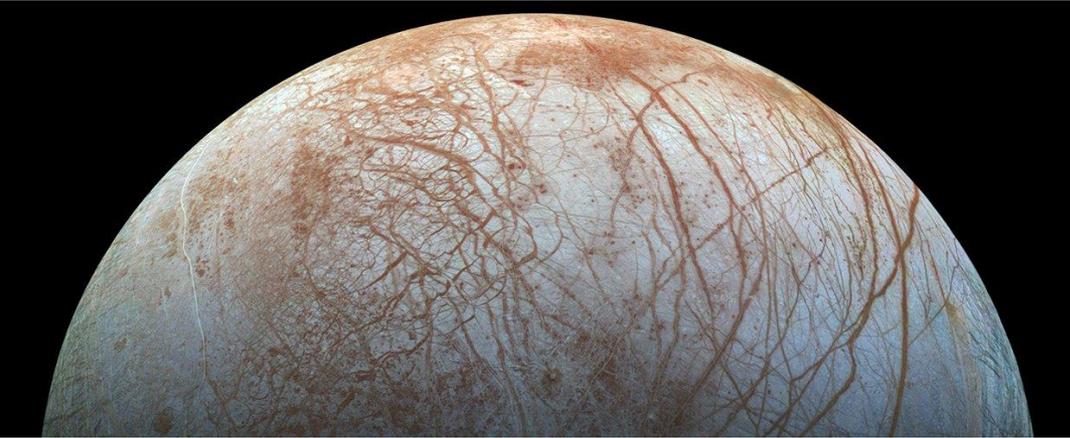
→ Taille estimé du génome 2.75 Mb codant 2,657 protéines !!!!

# Qui était LUCA?



Anaérobie en forme de bâtonnet, organohétérotrophe, pas de méthanogénèse, pas de photosynthèse, système immunitaire (Cas-CRISPR)

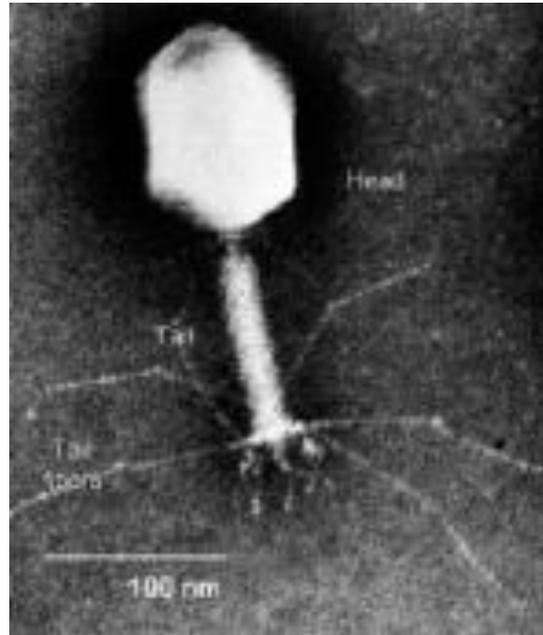
# Qui était LUCA?



“Jupiter’s moon Europa has a subterranean ocean, a rocky seabed, and geothermal heat produced by Jupiter’s gravitational tides. Water, rock and heat were all that were required by LUCA, so could similar life also exist on Europa?”

<https://astrobiology.nasa.gov/news/looking-for-luca-the-last-universal-common-ancestor/>

# Virus: origine et évolution



# Origin and evolution of viruses

- Viruses are by far the most abundant biological entities on Earth and they outnumber all the others put together.
- They infect all types of cellular life including animals, plants, bacteria and fungi.

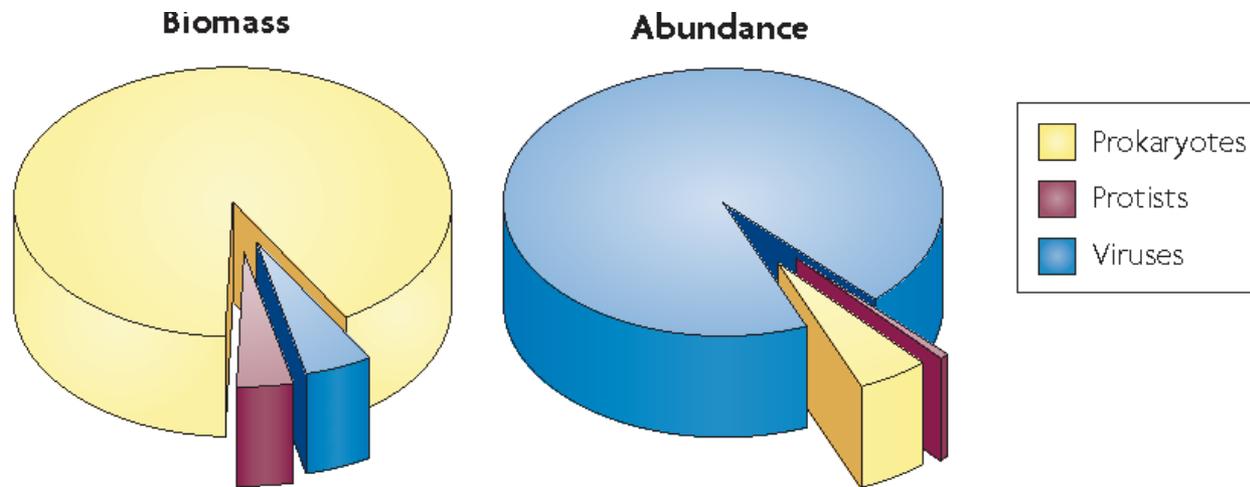


Figure 1 | Relative biomass and abundances of prokaryotes, protists and viruses.

Suttle, CA. (2007) Nat. Rev. Microbiol.

# Origin and evolution of viruses

- $10^{23}$  infections each second ! (in oceans) – viruses are probably responsible for the greatest selection pressure on cellular organisms.
- The entire history of life is the story of virus–host coevolution - the origins and evolution of viruses are an essential component of this process.
- The organization and origins of the enormous virosphere are profound open questions in biology.

What is a virus ?

Hypotheses for the origin of viruses

Are we at the mercy of viruses ?



# Origin and evolution of viruses

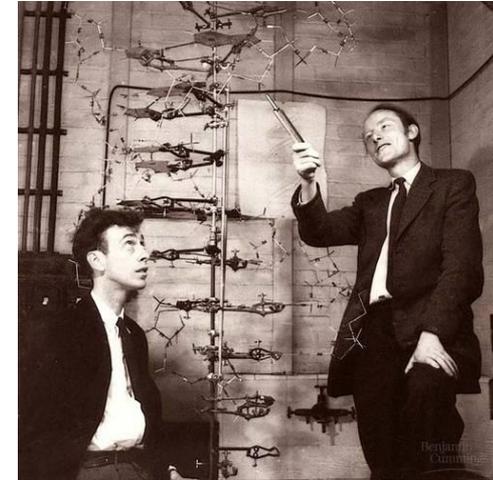
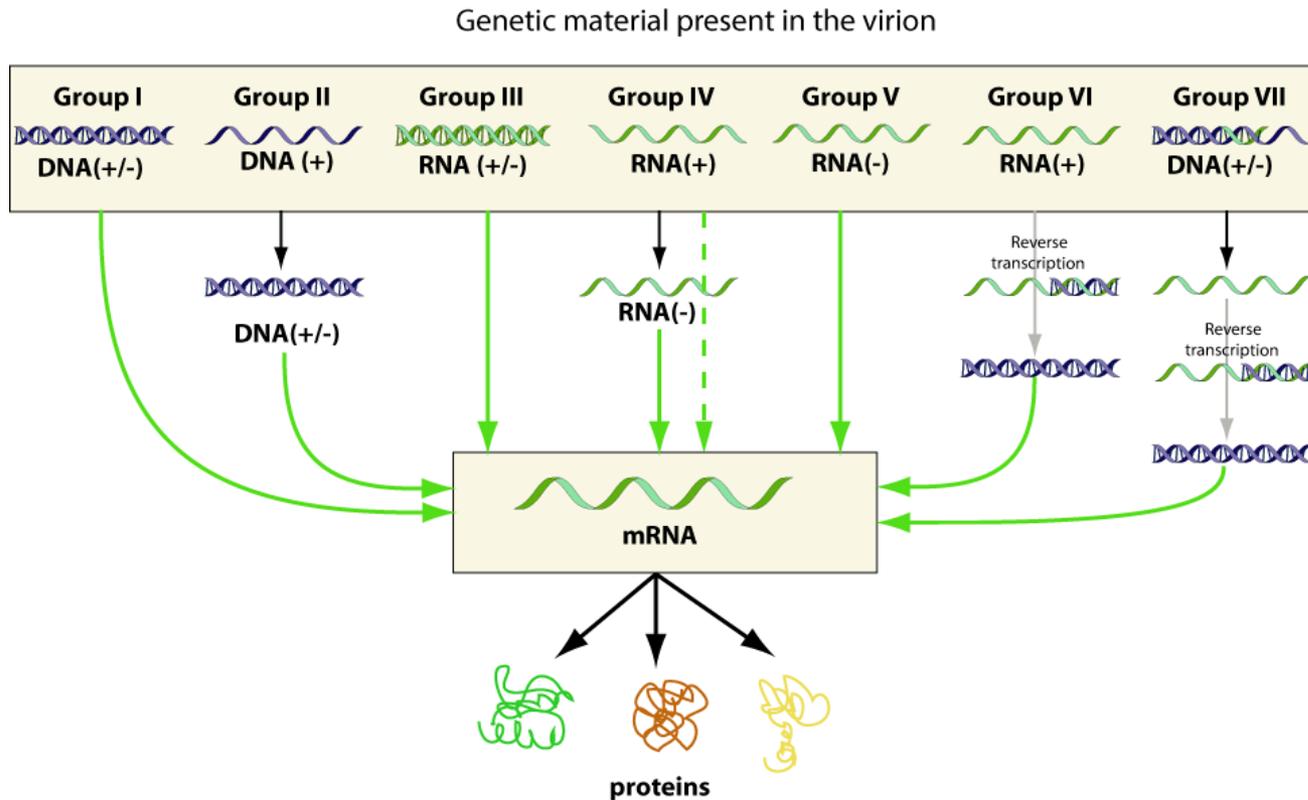
Two confronting hypotheses for the origin of viruses :

→ Viruses are smaller than cells : one step in between non-living (mineral) and living (cells), thus viruses predate the cells → **virus first hypothesis**

→ *contagium vivum fluidum* contains proteins and nucleic acids as cells do, thus viruses evolved from cells by reductive evolution → **virus second hypothesis**

# Origin and evolution of viruses

→ Middle of 20th century: genetic material is used to classify viruses and understand their origin



Discovery of DNA structure in 1953

Baltimore classification

# Origin and evolution of viruses

- Discovery of proviruses: viruses are parts of cellular genomes that escaped and became infections → **escape hypothesis**
- Corrolary : viruses originated independently in the three domains of life
- « true viral genes » do not exist

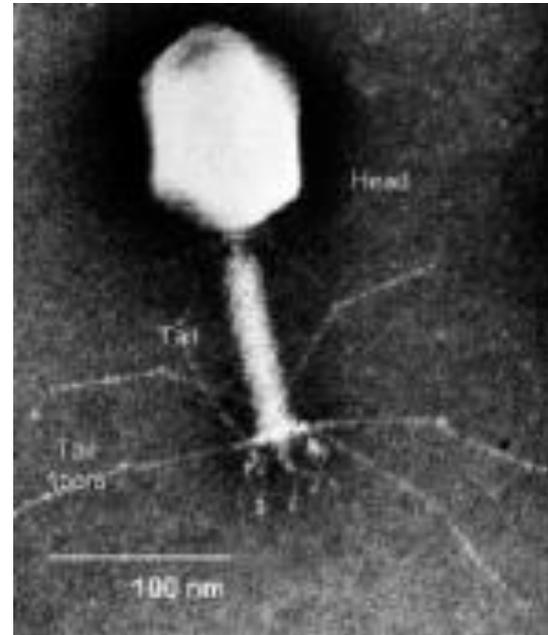
# Virus: origine et évolution



André Michel Lwoff  
(1902-1994)

## Qu'est ce qu'un virus?

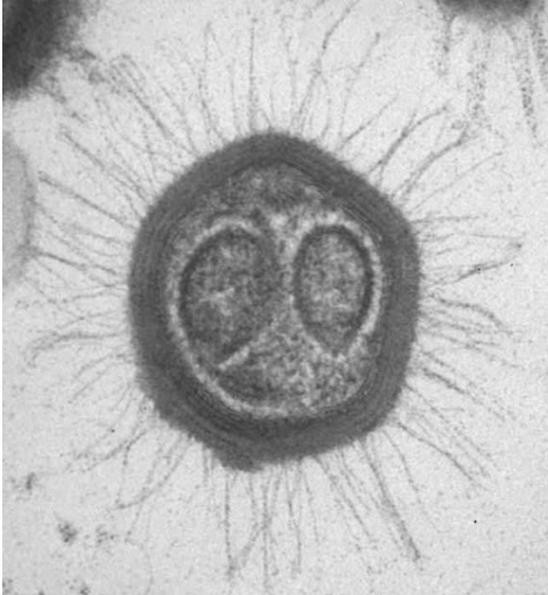
Les virus sont des agents infectieux non-autonomes, ne se reproduisent pas par division binaire, sont constitués de protéines et d'acide nucléique (ADN ou ARN), ne sont pas visibles au microscope optique (passent à travers un filtre 0.22  $\mu\text{m}$ )



Lwoff, A. The concept of virus. *J. Gen. Microbiol.* **17**, 239–253 (1957).

# Virus: origine et évolution

*Acanthamoeba polyphaga* mimivirus (APMV)



→ Un virus géant (0,6  $\mu\text{m}$  de diamètre)!  
Infecte une amibe...



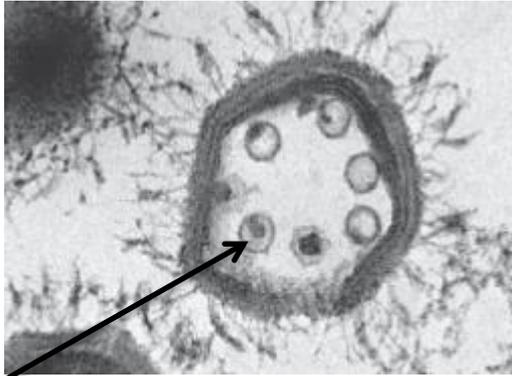
*Entamoeba haemolytica*

1,2 Mpb – environ 900 gènes!  
dont les gènes impliqués dans la traduction,  
mitose, métabolisme des glucides, lipides;  
biogénèse des membranes, motilité...

(LaScola et al., *Science*, 2002; *Science*, 2004)

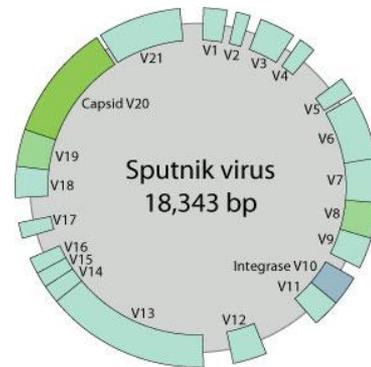
Suzan-Monti et al. *Virus Research* 117 (2006) 145–155

# Virus: origine et évolution



Mimivirus peut tomber malade!

Sputnik, 50 nm, se propage au détriment du Mimivirus



Un virophage!

(LaScola et al., Nature, 2008)

# Virus: origine et évolution

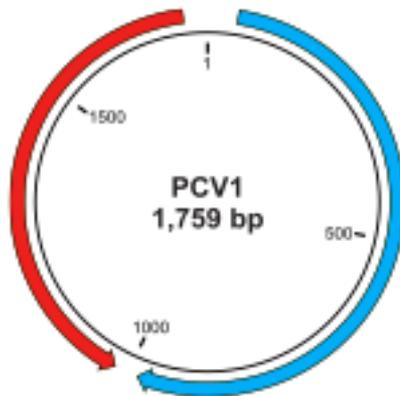
## Pandoravirus

2.5 Megabase (2-3000 gènes)!!!

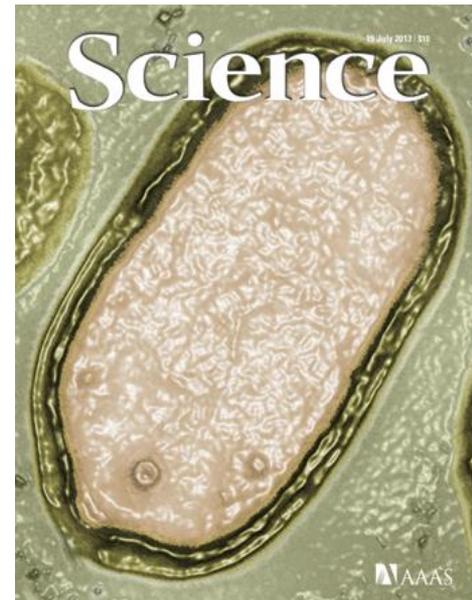


## Qu'est ce qu'un virus?

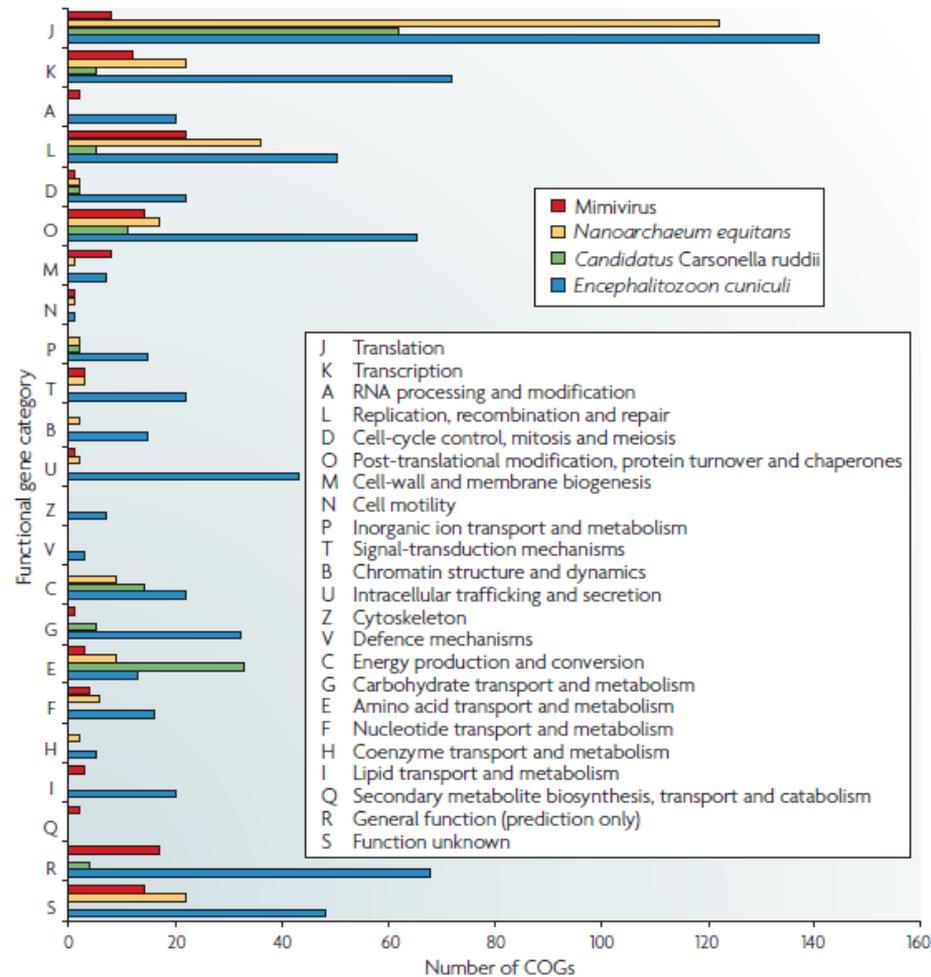
**Circovirus de porc**  
0.0017 Mb (2 gènes)



**Pandoravirus**  
2.5 Megabase (2-3000 gènes)!!!



# Virus: origine et évolution



Quelle est la différence entre Mimivirus et les parasites (intracellulaires) obligatoire avec des génomes très réduits?

« The major difference in *Mimivirus* genome with the genomes of small intracellular bacteria is the absence of genes coding for ribosomal proteins. » - définition négative

# Virus: origine et évolution

ARTICLE

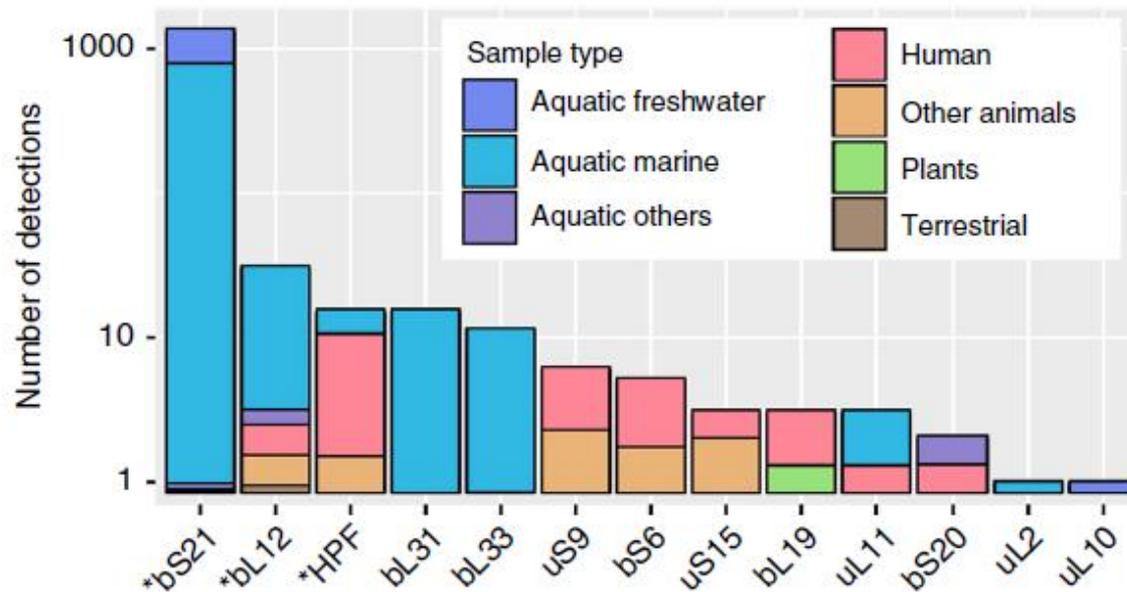
<https://doi.org/10.1038/s41467-019-08672-6>

OPEN

## Numerous cultivated and uncultivated viruses encode ribosomal proteins

Carolina M. Mizuno<sup>1</sup>, Charlotte Guyomar<sup>2</sup>, Simon Roux<sup>3</sup>, Régis Lavigne<sup>4</sup>, Francisco Rodriguez-Valera<sup>5</sup>, Matthew B. Sullivan<sup>6,7</sup>, Reynald Gillet<sup>2</sup>, Patrick Forterre<sup>1</sup> & Mart Krupovic<sup>1</sup>

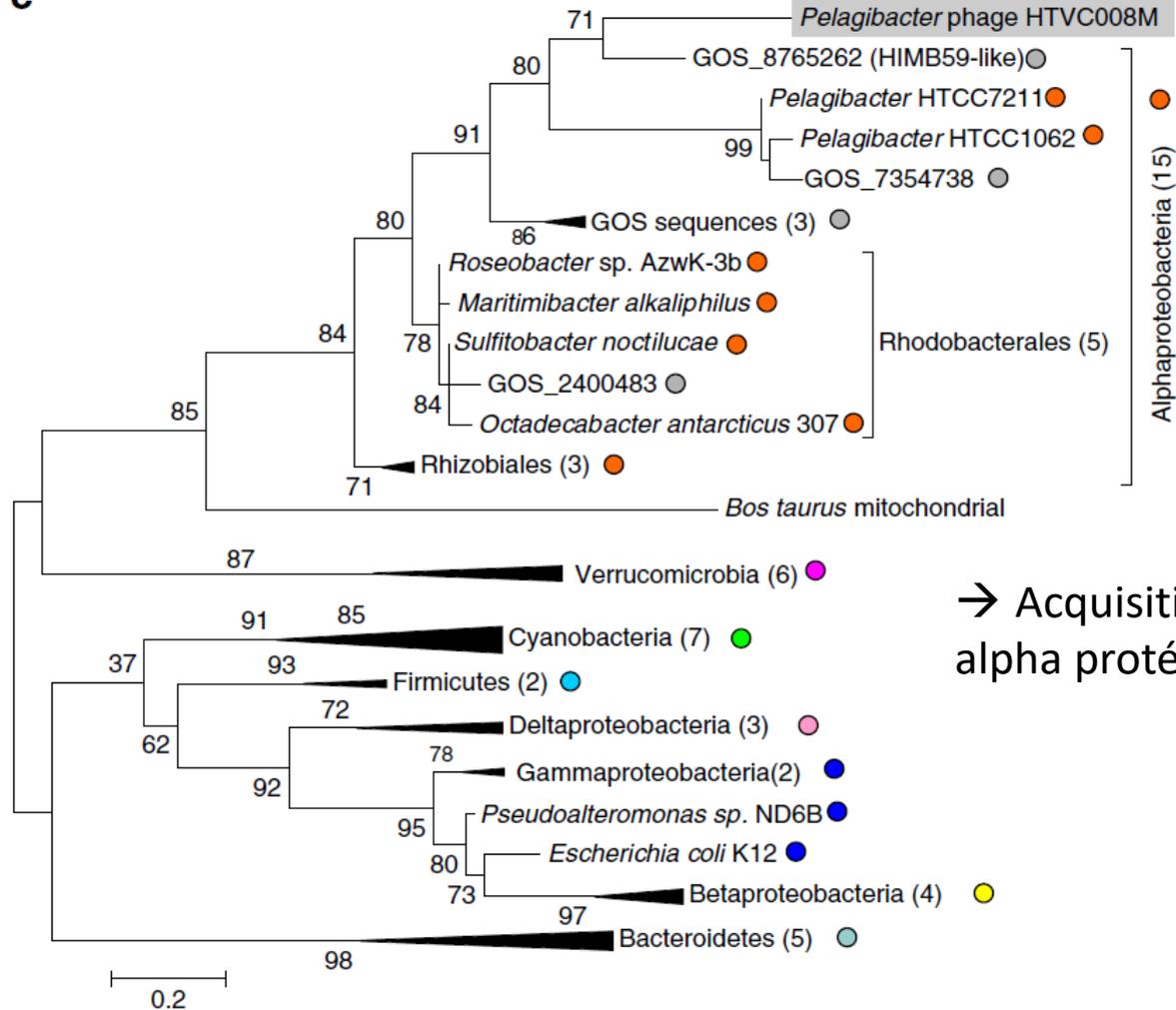
→ 14 different RPs across viral genomes were detected



# Virus: origine et évolution

→ L'arbre phylogénétique pour la protéine ribosomale bS21

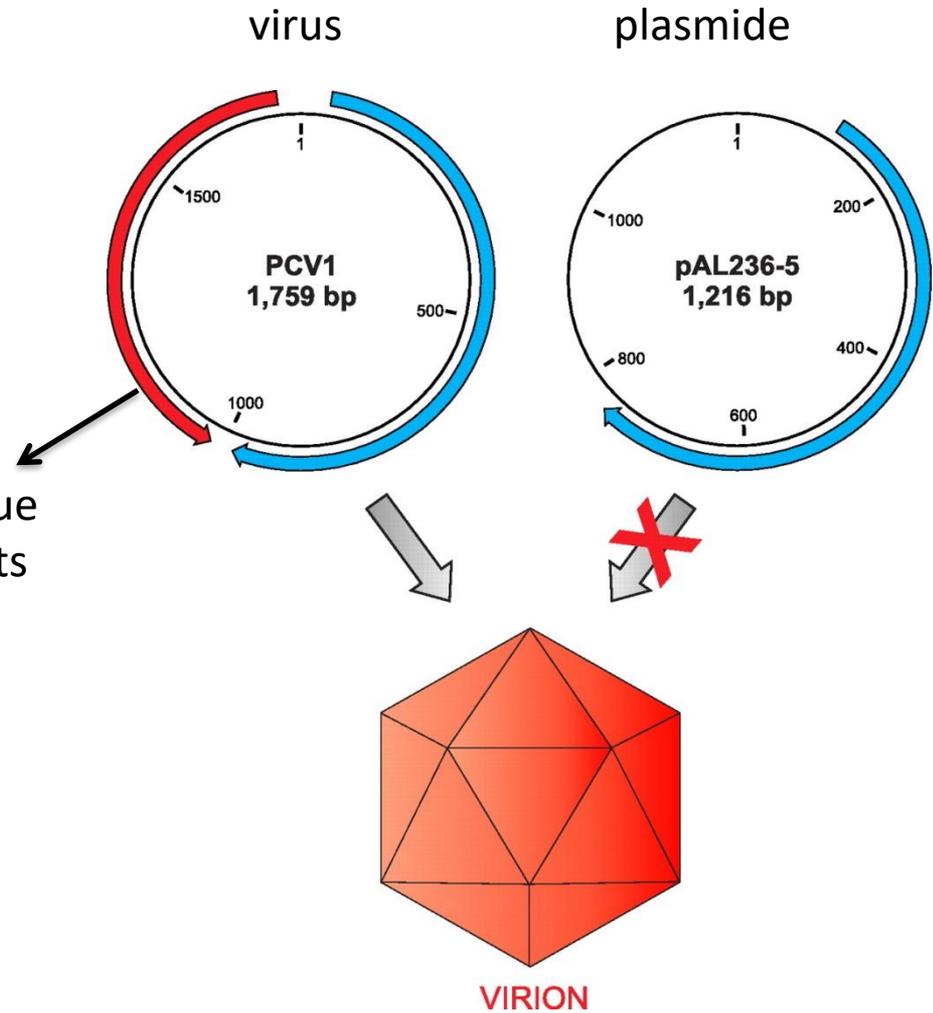
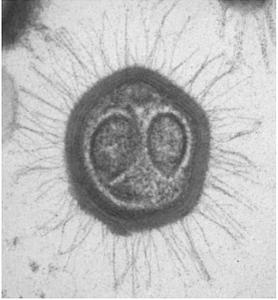
c



← orthologue viral

→ Acquisition par HGT depuis le hôte, une alpha protéobactérie

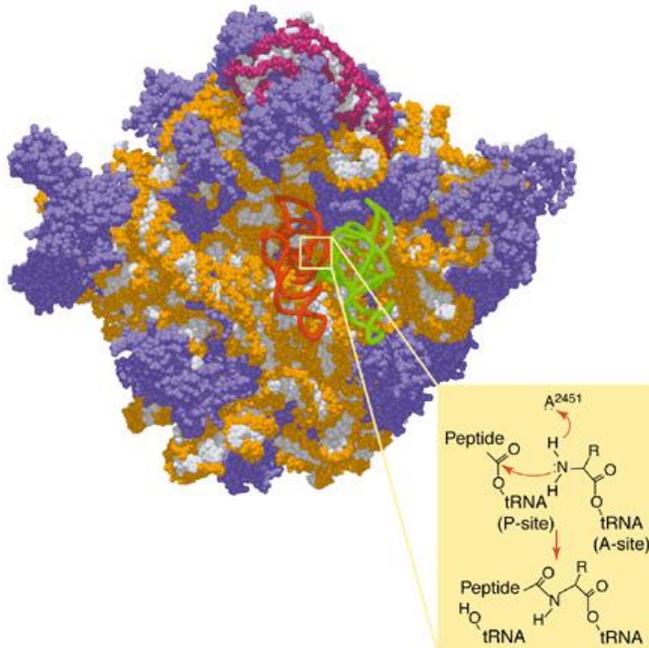
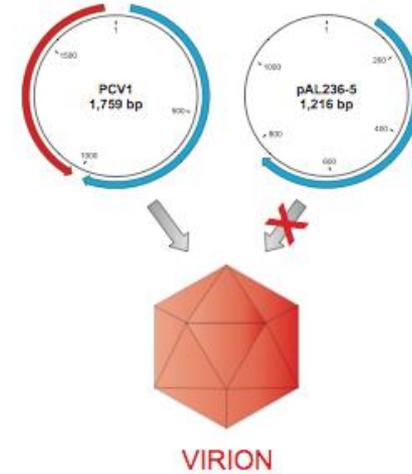
# Virus: origine et évolution



la capacité de former une capside est unique aux virus (distinction avec d'autres éléments génétiques mobiles comme les plasmides)

# Virus: origine et évolution

Par définition un virion doit comporter au moins une protéine codée par le génome viral



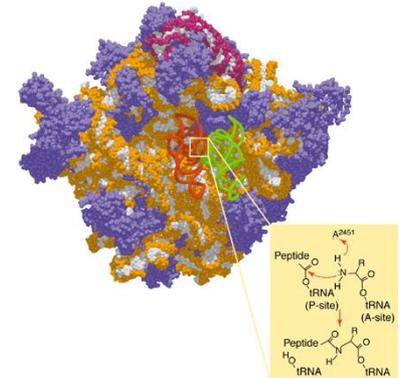
Les virus *modernes* ont émergé après l'apparition des ribosomes ...

# Origin and evolution of viruses

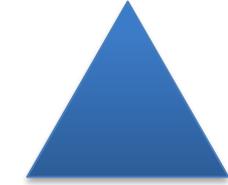
→ late escape hypothesis



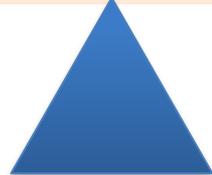
Patrick Forterre



First cell (origin of life)



1st RNA cell



«invention» of ribosome



1st RNA-**Protein** cell

**ORIGIN of VIRUSES**



LUCA

«invention» of DNA



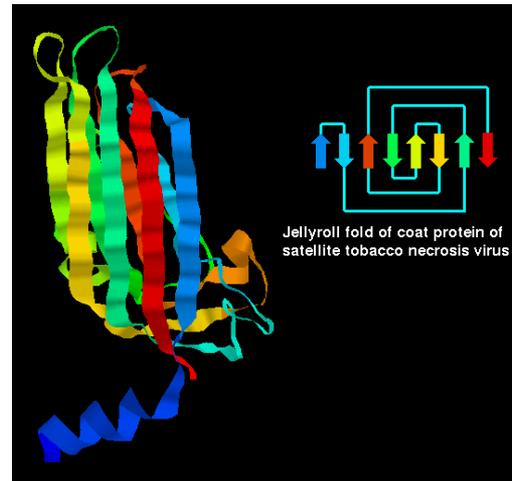
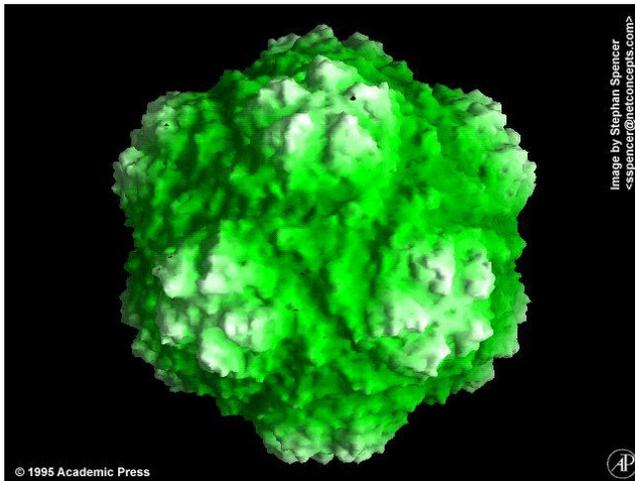
Eukarya      Archaea      Bacteria

# Virus: origine et évolution

« late escape hypothesis »

- les 1<sup>er</sup> virus étaient très simples (capacités fonctionnelles très limitées)

STNV (satellite tobacco necrosis virus), ssRNA

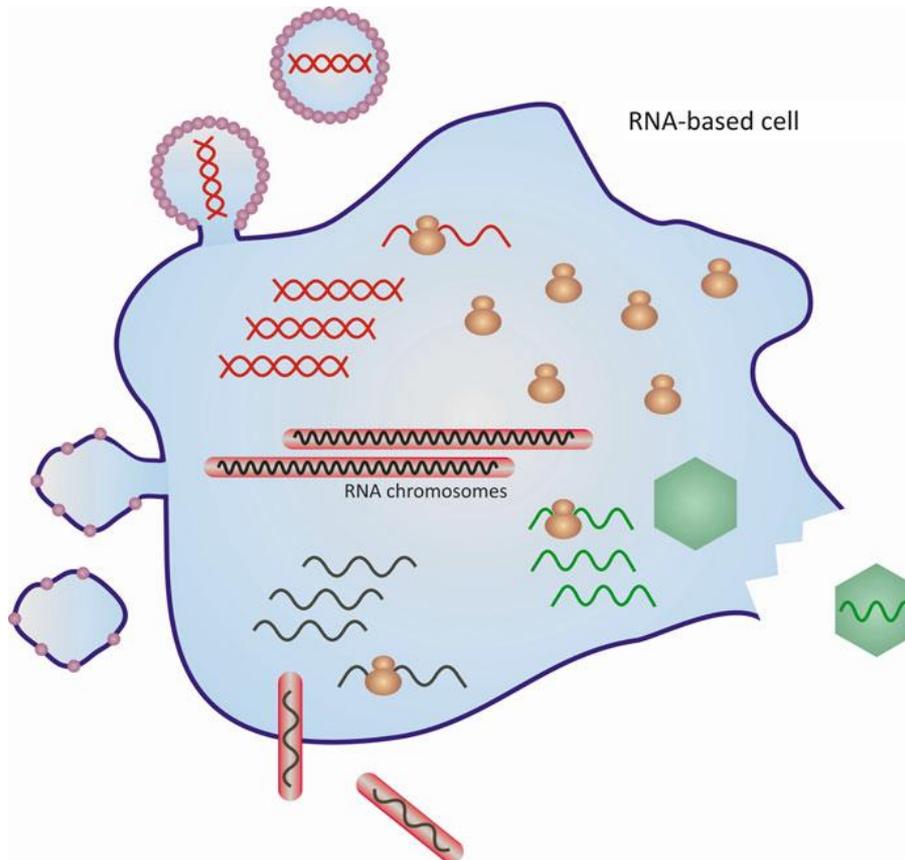


1300 bp, 1 protéine (capside)

-ce virus nécessite la présence d'un autre virus pour répliquer son génome

-le génome des 1<sup>er</sup> virus a été répliqué à l'aide de la machinerie de la cellule hôte

# Origin and evolution of viruses



“The origin of the first viruses boils down to the origin of the capsid proteins, which acquired the ability to package their own genes; subsequent acquisition of additional functions (e.g., for genome replication) would lead to complexification and increased “autonomy” of such viruses.”

# Virus: origine et évolution

Full text provided by [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®



Dennis Bamford



## What does structure tell us about virus evolution?

Dennis H Bamford<sup>1</sup>, Jonathan M Grimes<sup>2</sup> and David I Stuart<sup>2</sup>



= virus « self »

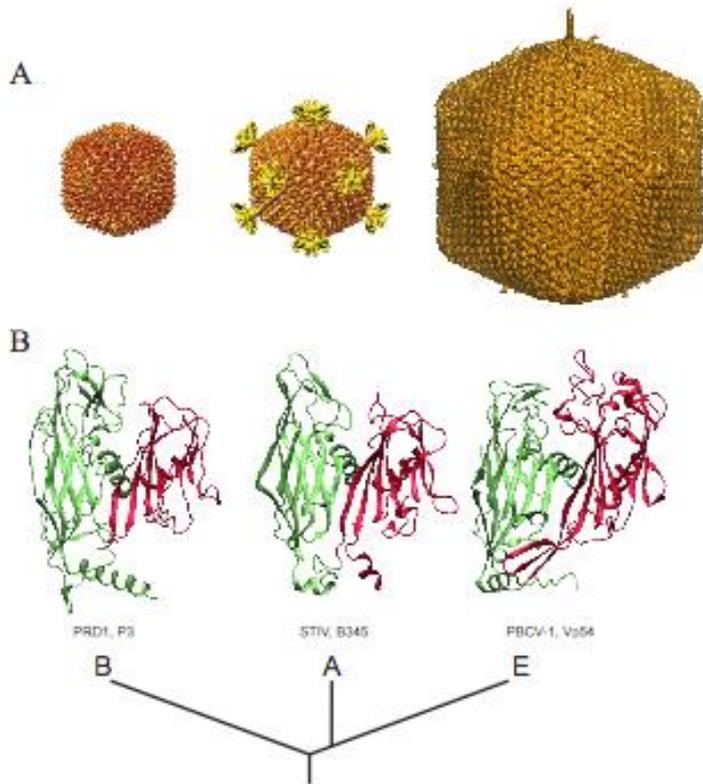
-l'analyse de la topologie des protéines structurales des virions a révélé des similitudes inattendues (*invisible au niveau séquences protéiques*):

→ on peut identifier des lignées virales grâce à l'étude structurale des protéines de capsides

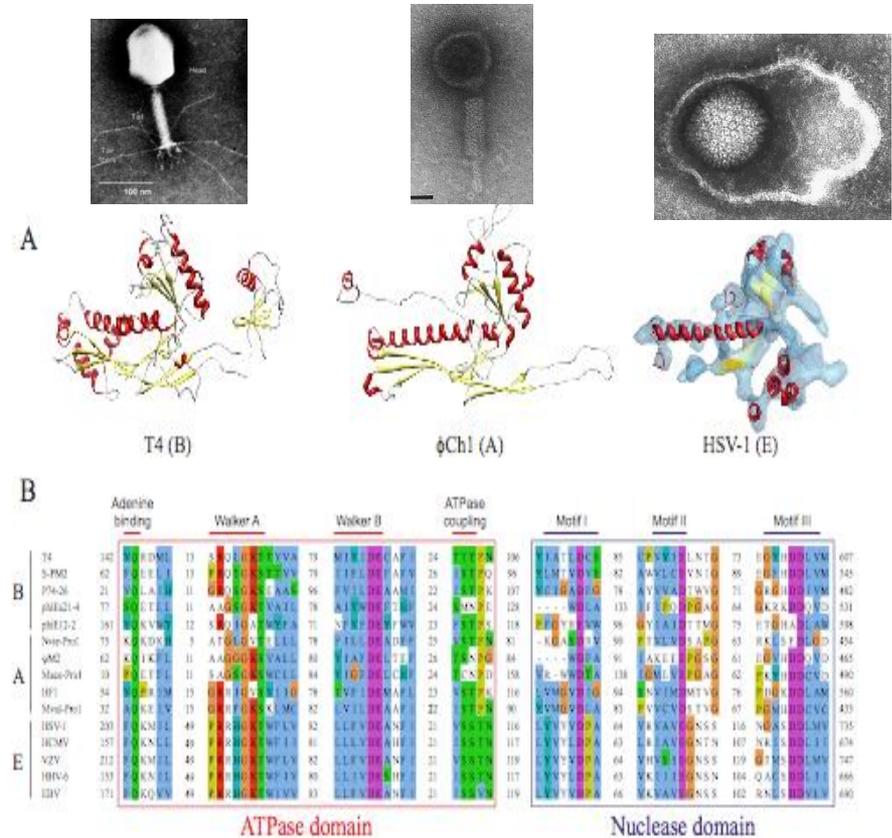
Bamford et al. Current Opinion in Structural Biology 2005, 15:655–663

# Virus: origine et évolution

## La lignée PRD1-like

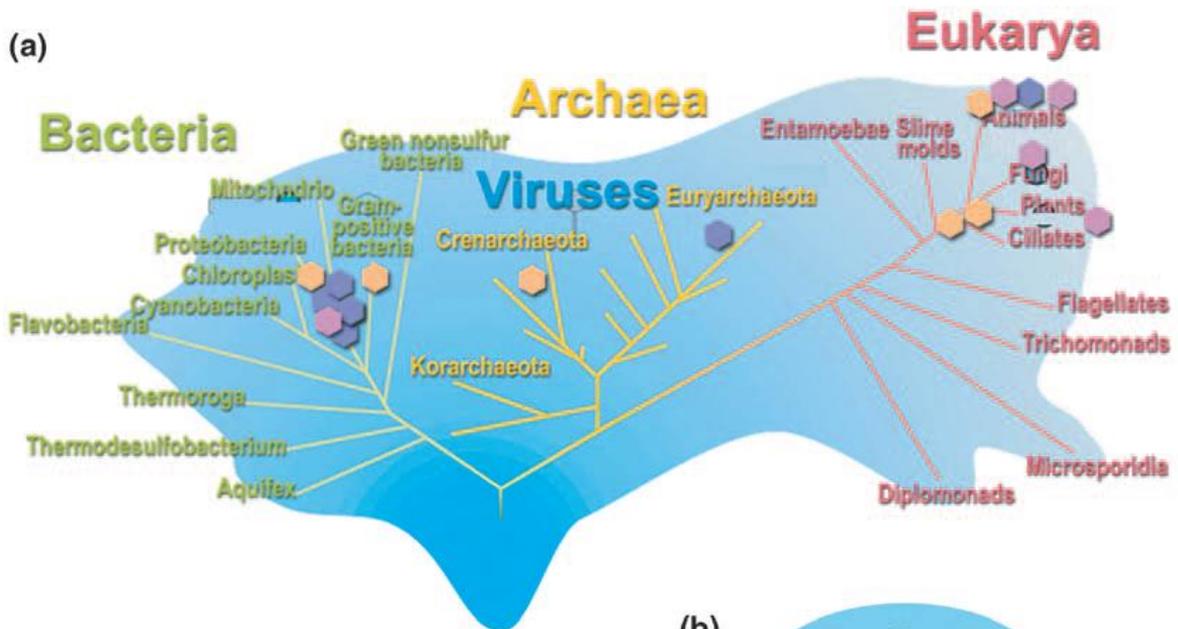


## La lignée HK90-like



# Virus: origine et évolution

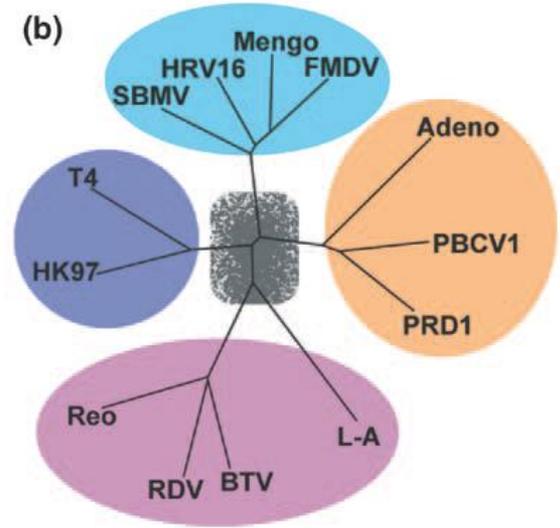
(a)



→ Ces virus ont-ils un ancêtre commun ?

Lineage	Representatives
PRD1-like	PRD1, Bam35, STIV, adeno, PBCV1, PpV01
HK97-like	HK97, T4, P22, φ29, P.f. VLP, HSV-1
BTV-like	φ6, BTV, RDV, Reo, L-A

(b)



Structure-based phylogenetic tree for the viruses

# Origin and evolution of viruses

- Comparative analysis of the major structural proteins across the entire classified virosphere
- Identification of cellular homologs of these viral proteins

## Multiple origins of viral capsid proteins from cellular ancestors

Mart Krupovic<sup>a,1</sup> and Eugene V. Koonin<sup>b,1</sup>

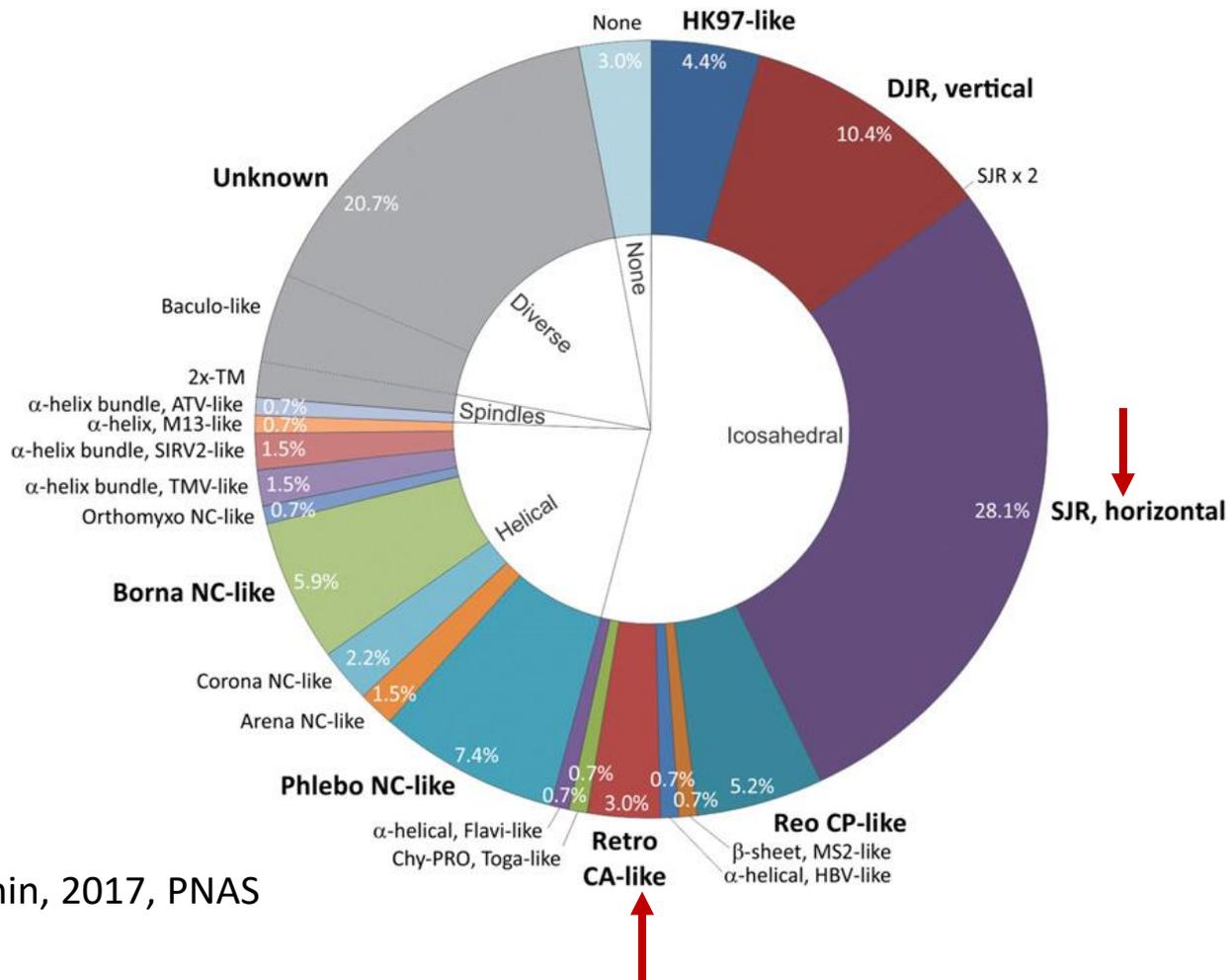
<sup>a</sup>Institut Pasteur, Department of Microbiology, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, 75015 Paris, France; and <sup>b</sup>National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894

Contributed by Eugene V. Koonin, February 3, 2017 (sent for review December 21, 2016; reviewed by C. Martin Lawrence and Kenneth Stedman)

**Viruses are the most abundant biological entities on earth and show** genome replication. Understanding the origin of any virus group is

# Origin and evolution of viruses

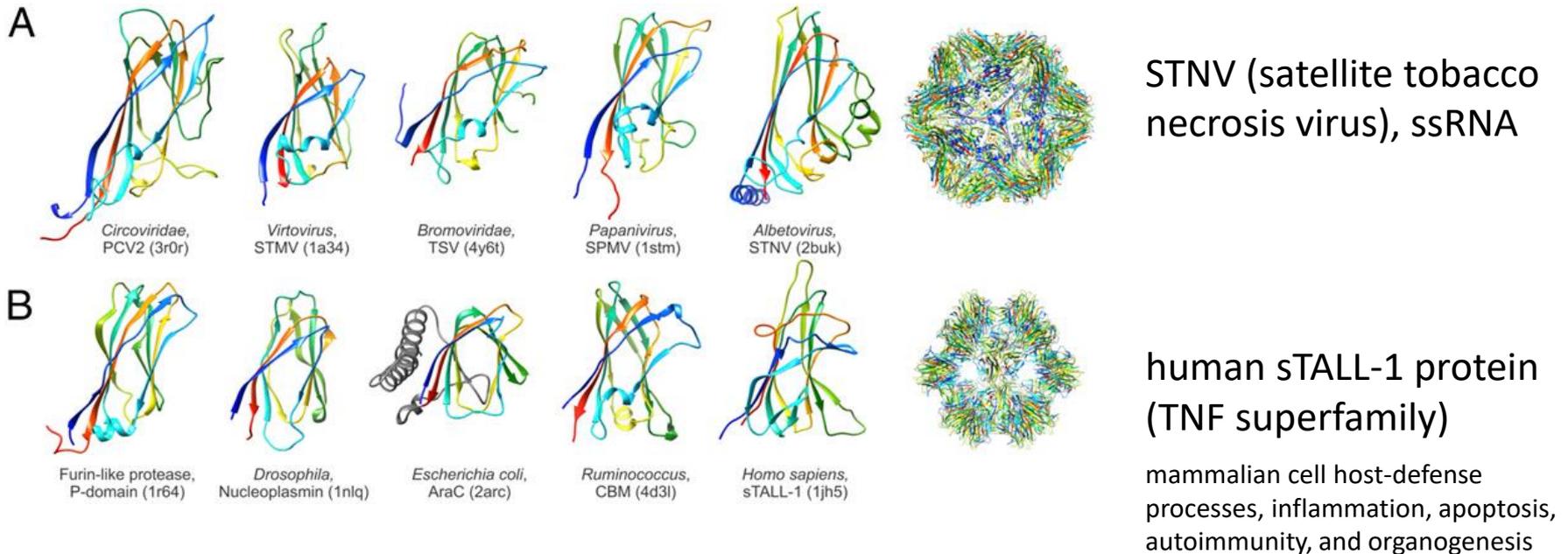
→ The 76.3% of viral taxa for which the fold of the major virion proteins was defined could be divided into 18 architectural classes - “structure-based viral lineages”



# Origin and evolution of viruses

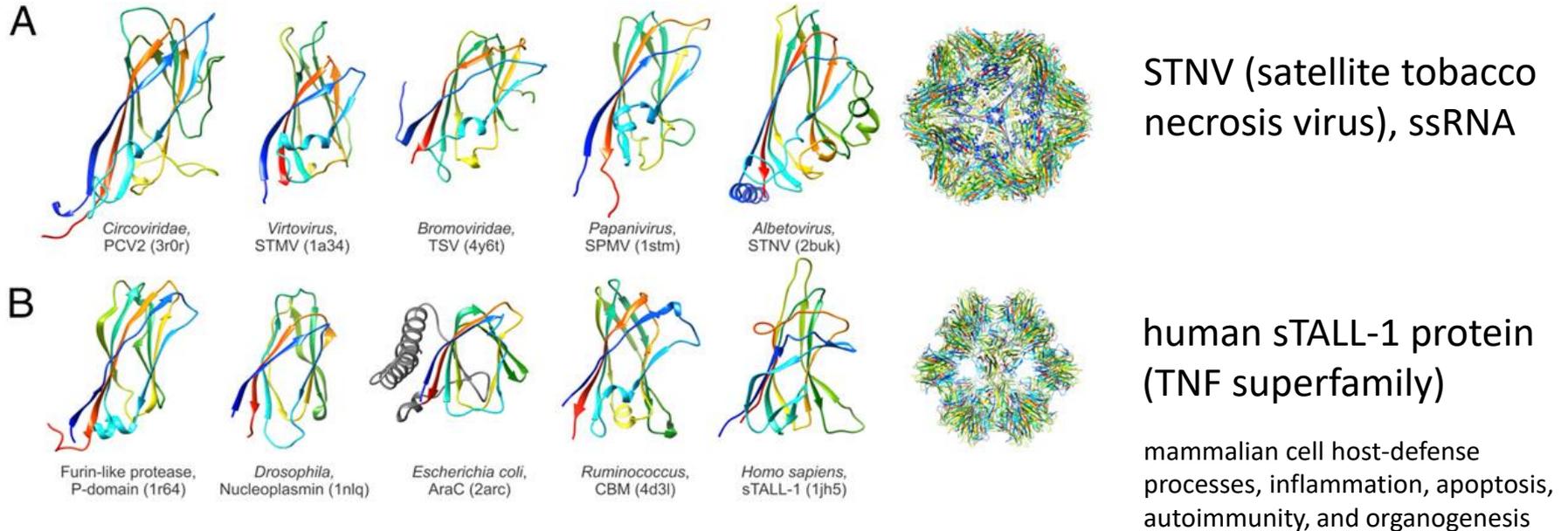
→ Viral and cellular single jellyroll (SJR) proteins

→ SJR is the most prevalent fold among viral CPs (~28% of the CPs)



→ Cellular proteins from the three domains of life with high structural similarity (but no sequence similarity) could be readily identified

# Origin and evolution of viruses



→ Au regard de la similarité structurale entre les protéines SJR virales et cellulaires (et leur capacité de former des particules virus-like) il est probable que les SJR capsides ont évolué grâce au recrutement des orthologues cellulaires.

→ La fonction originale des SRJ dans la reconnaissance des sucres?

→ Le recrutement des protéines SRJ aurait pu être bénéfique immédiatement pour le virus (protection de son matériel génétique et reconnaissance des glycoprotéines à la surface des cellules)

# Origin and evolution of viruses

→ Structure based (Z-score) phylogenetic tree of viral and cellular SJR proteins

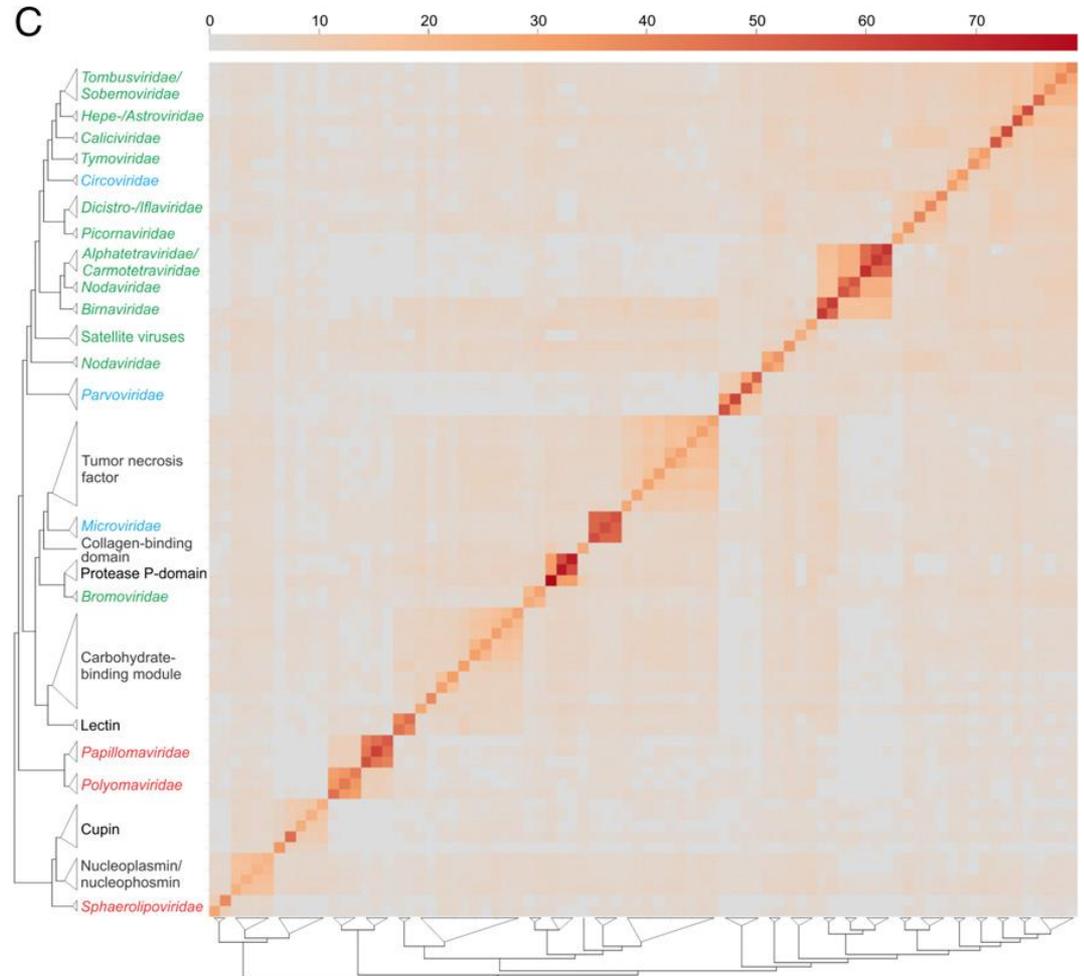
green = RNA viruses

blue = ssDNA viruses

Red = dsDNA viruses

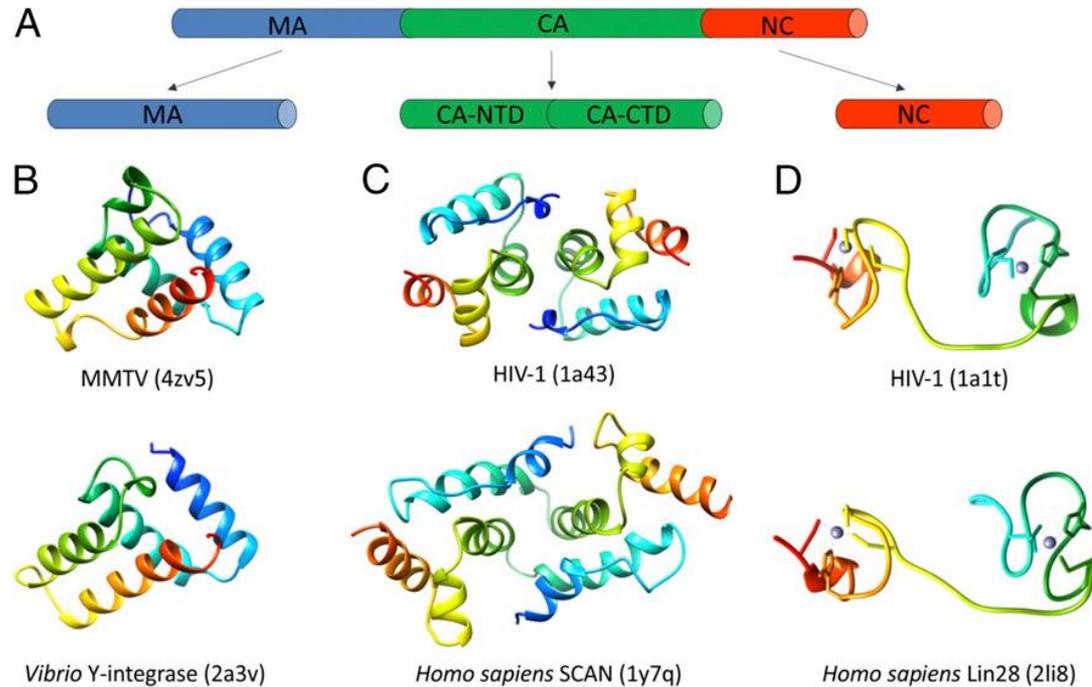
Black = cellular SJR proteins

→ SJR CPs could have evolved from *bona fide* cellular proteins, possibly on several independent occasions



# Origin and evolution of viruses

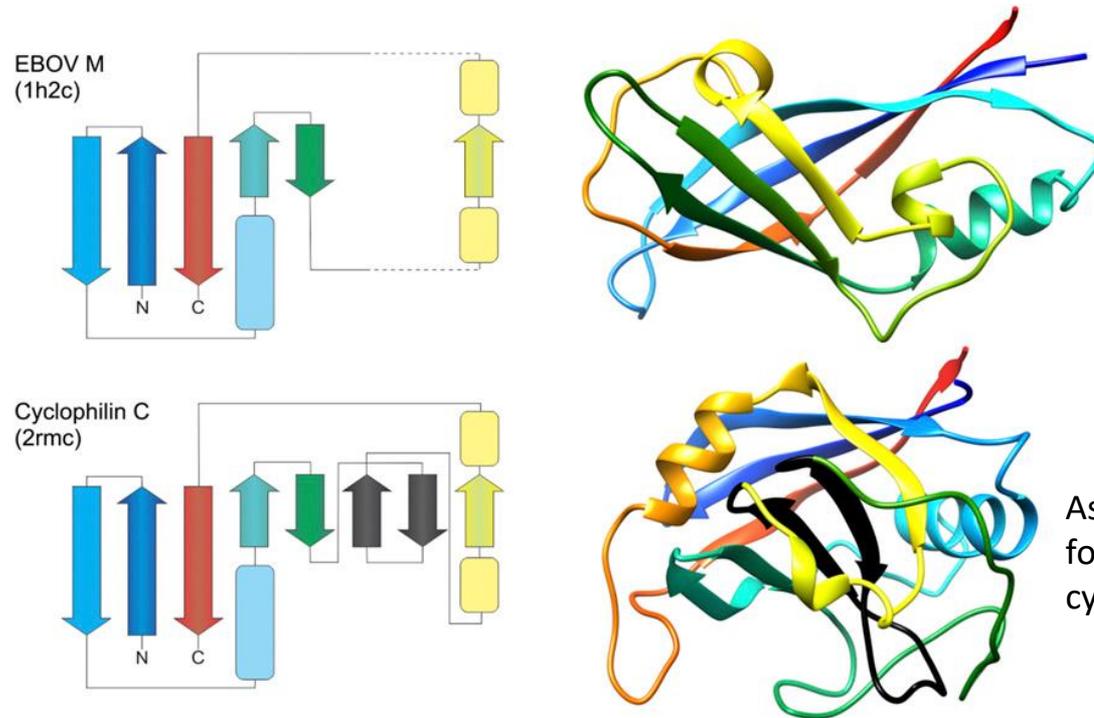
→ the Gag polyprotein of retroviruses (matrix (MA), capsid (CA), and NC proteins)



→ at least two of the three major building blocks of retroviral virions are likely to have evolved from cellular proteins.

# Origin and evolution of viruses

→ Mononegavirales encode matrix proteins that direct virion assembly and budding

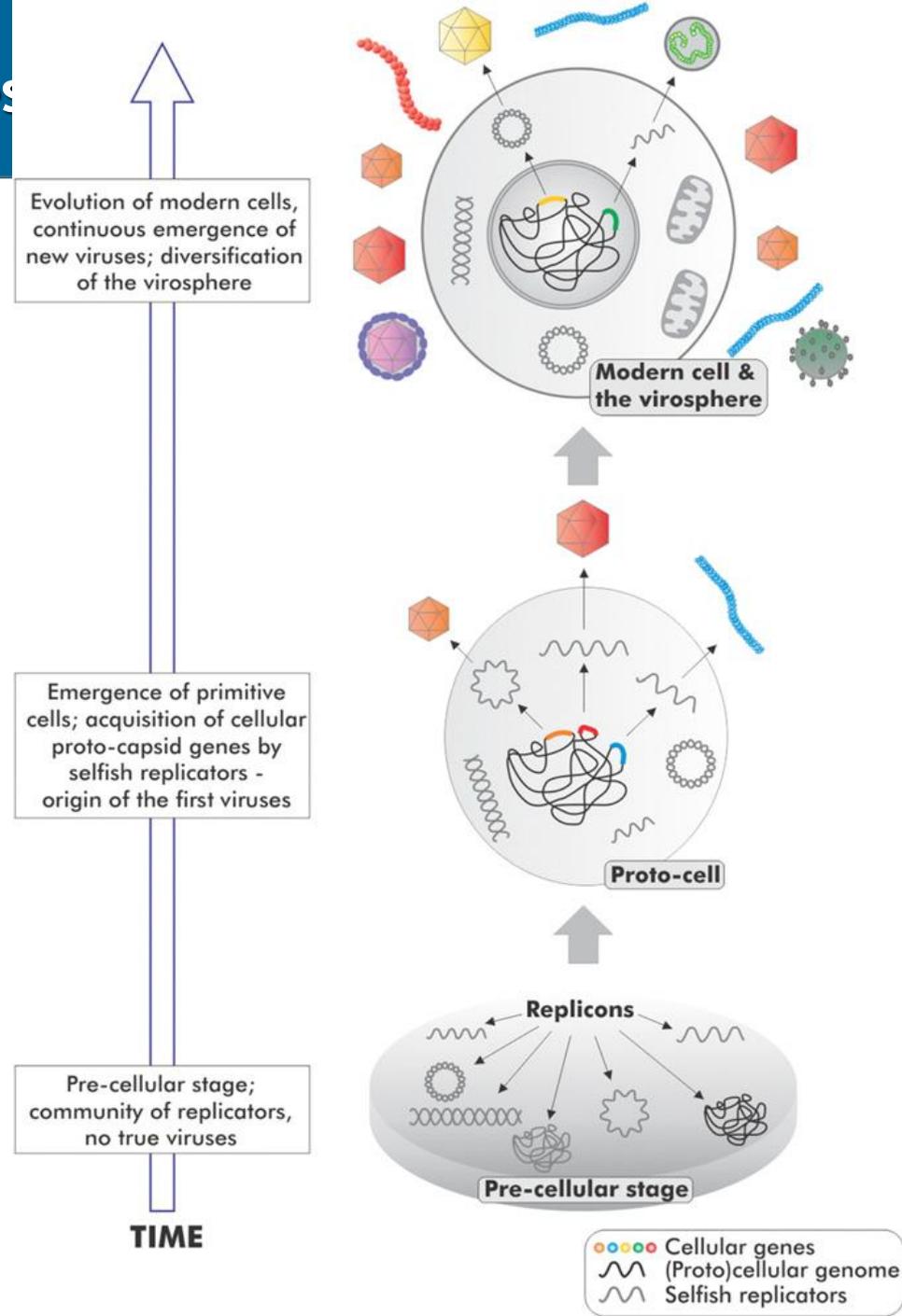


Assist protein folding, receptor for immunosuppressive drug cyclosporin A

→ the matrix protein-encoding gene of mononegaviruses likely evolved from a cyclophilin gene acquired from the host

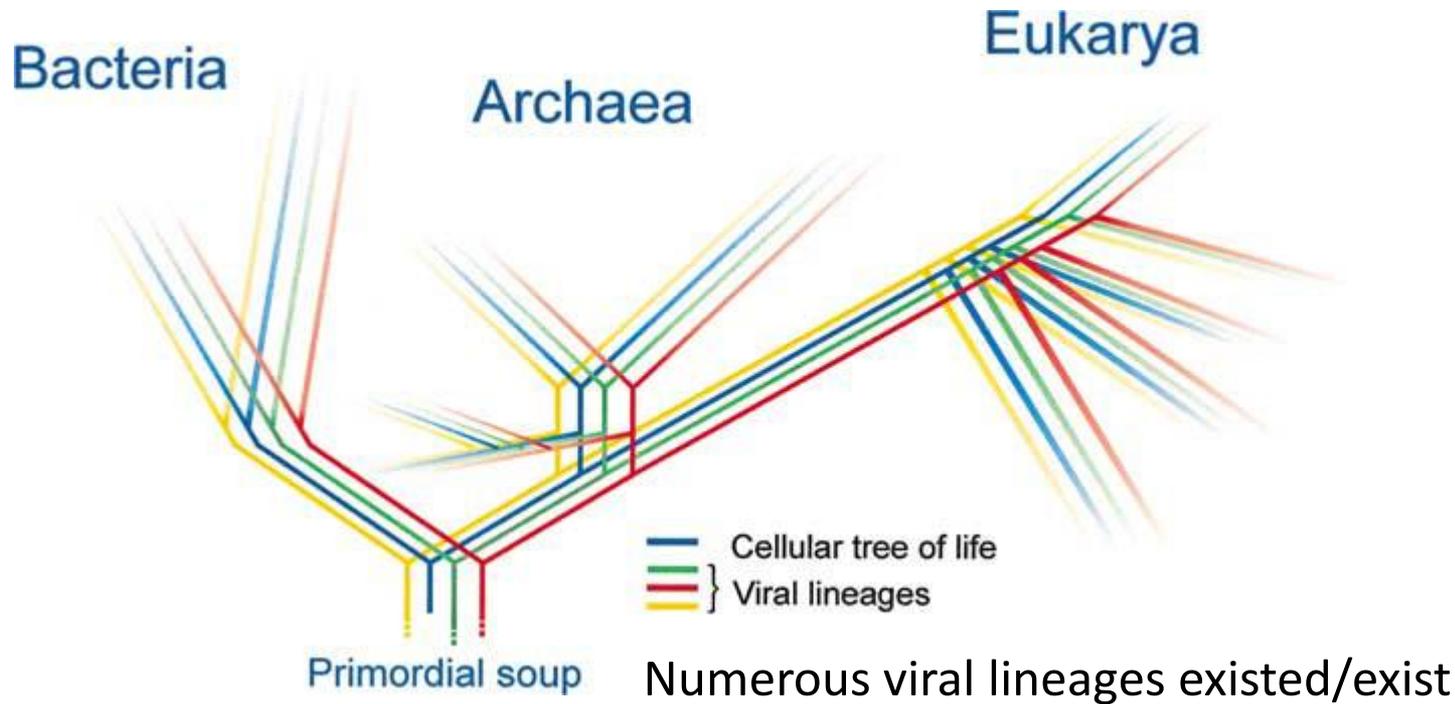
# Origin and evolution of viruses

→ *bona fide* viruses evolved on multiple, independent occasions by recruiting diverse host proteins that became major virion components



# Origin and evolution of viruses

→ Viruses are probably polyphyletic



→ It remains to be determined which part of the cellular tree is covered by which lineage

→ Viruses are a major driver of evolution of cells



ADAPTIVE EVOLUTION

## At the mercy of viruses

**Viruses are responsible for many of the adaptive mutations in the human genome.**

CLAUS O WILKE AND SARA L SAWYER



## Viruses are a dominant driver of protein adaptation in mammals

**David Enard\*, Le Cai, Carina Gwennap, Dmitri A Petrov**

Department of Biology, Stanford University, Stanford, United States

# Virus: origine et évolution

→ l'adaptation de mammifères aux virus a été étudié à travers des protéines impliquées dans l'interaction avec les virus (VIP)

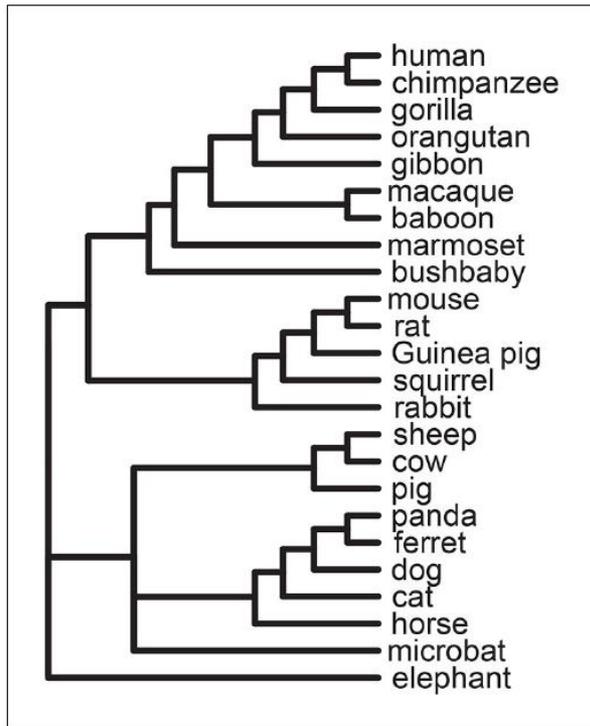


Figure 1. Tree of 24 mammals used in the analysis.

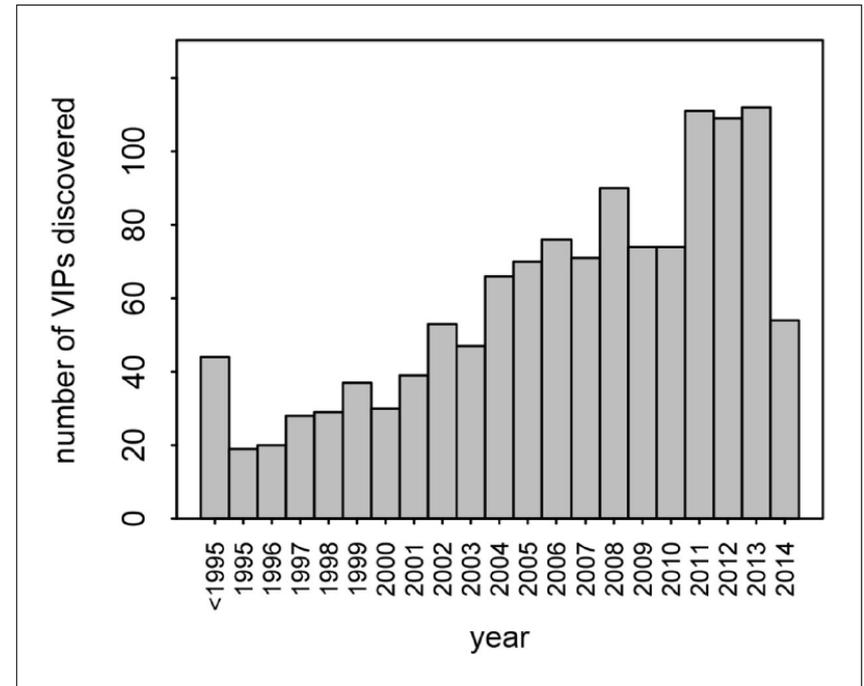


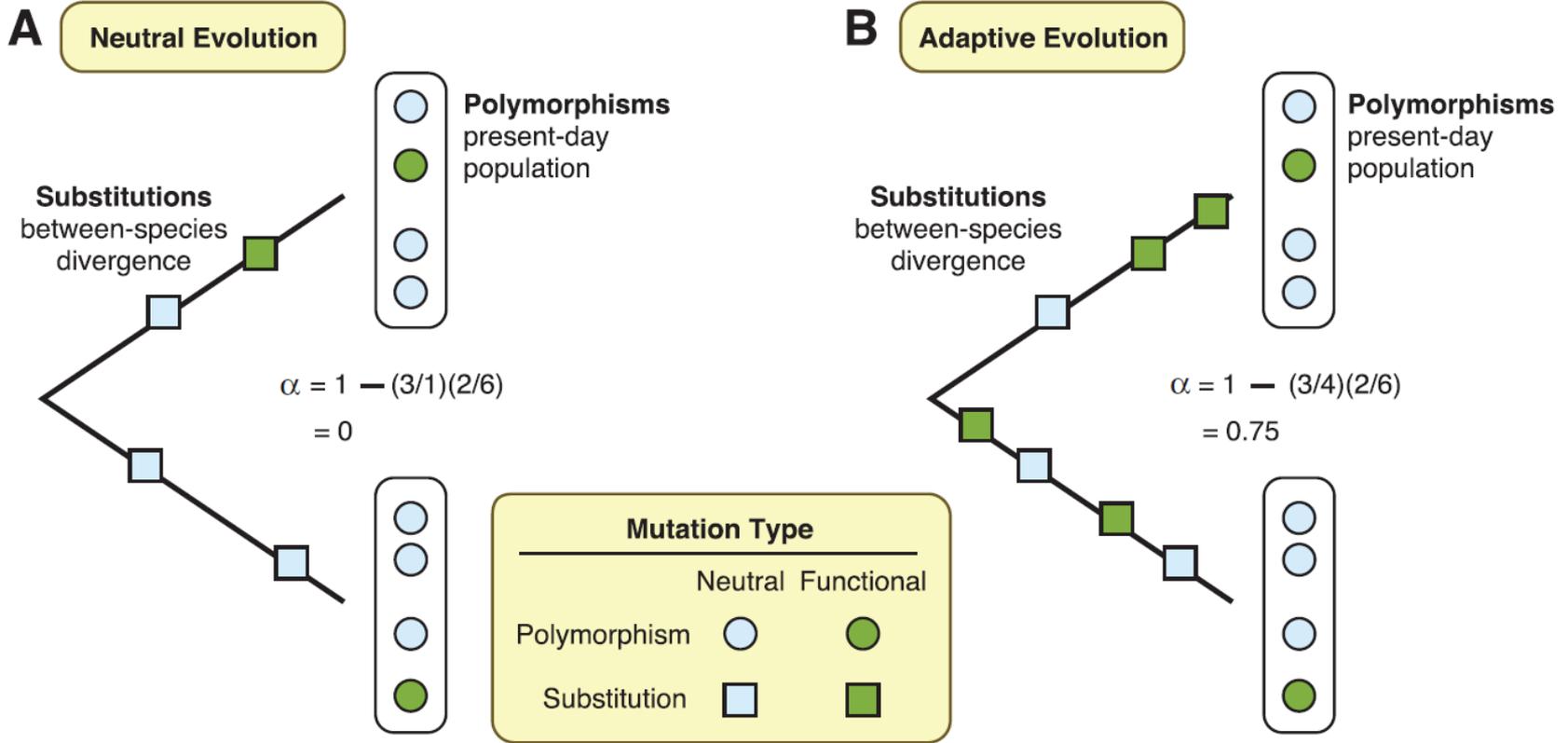
Figure 2. Number of VIPs discovered per year until 2014.

→ ~1300 protéines VIP conservées chez les mammifères

# Virus: origine et évolution

- Les mécanismes à l'origine de changement et diversification des génomes:
  - La théorie neutre: la grande majorité des mutations sont neutres et persistent par chance
  - La théorie d'adaptation: les mutations persistent car elles sont bénéfiques pour l'organisme
  
- Pour mesurer le niveau d'adaptation des protéines on peut calculer les proportions des mutations neutres vs adaptives dans les séquences d'ADN codant pour ces protéines

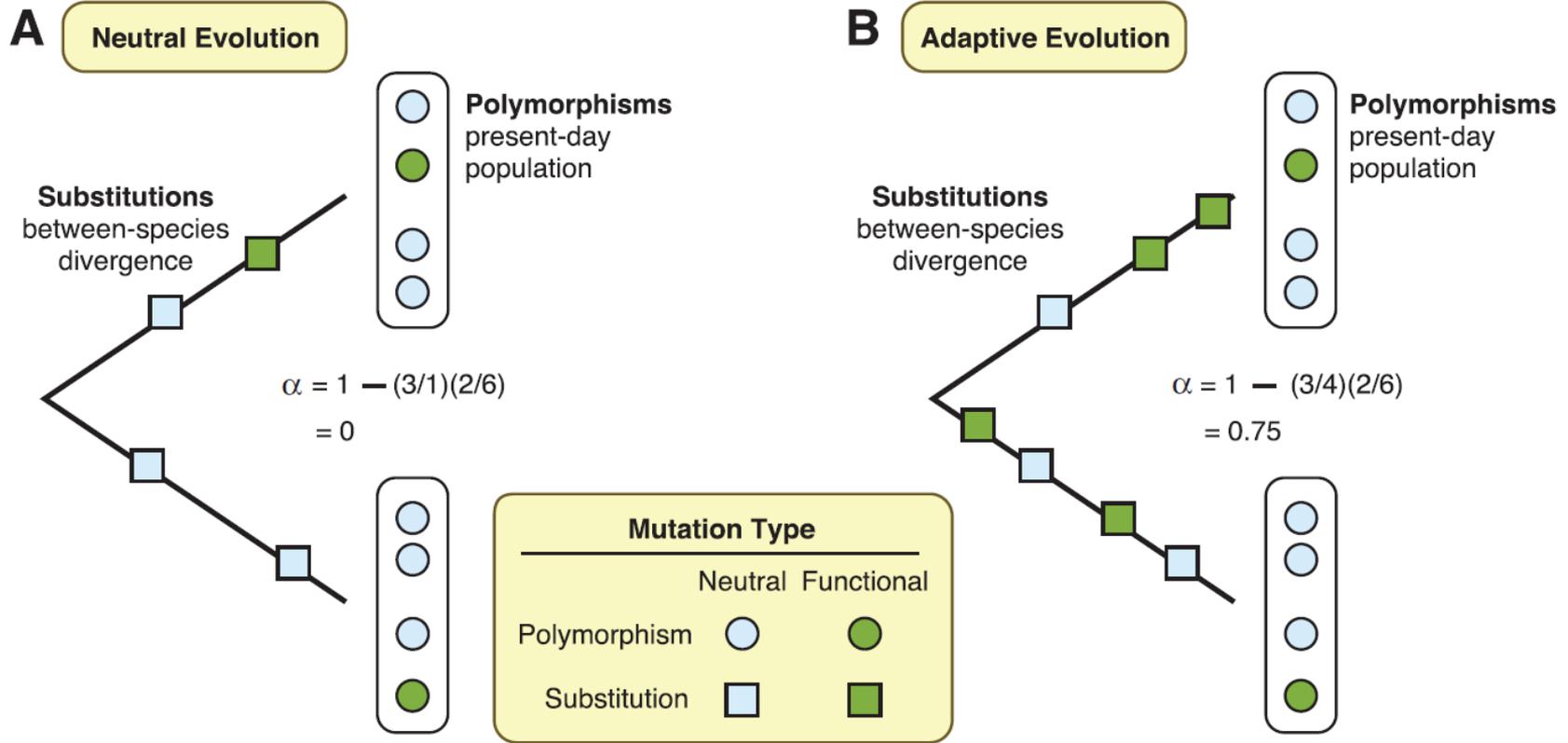
# Virus: origine et évolution



Substitutions = mutations partagé par l'ensemble de la population

Poymorphisme = mutations apparus chez quelques membres de la population

# Virus: origine et évolution

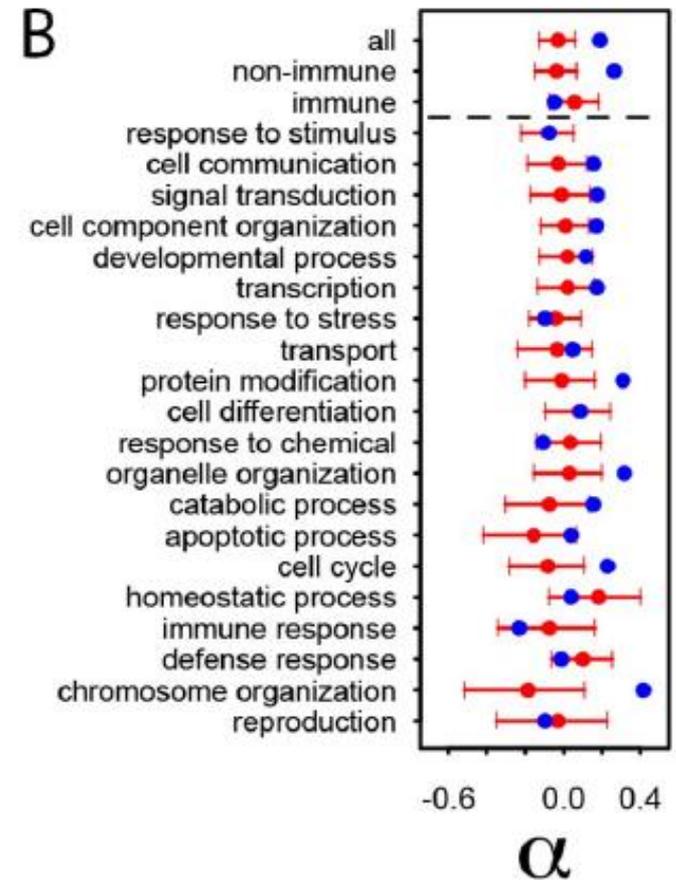
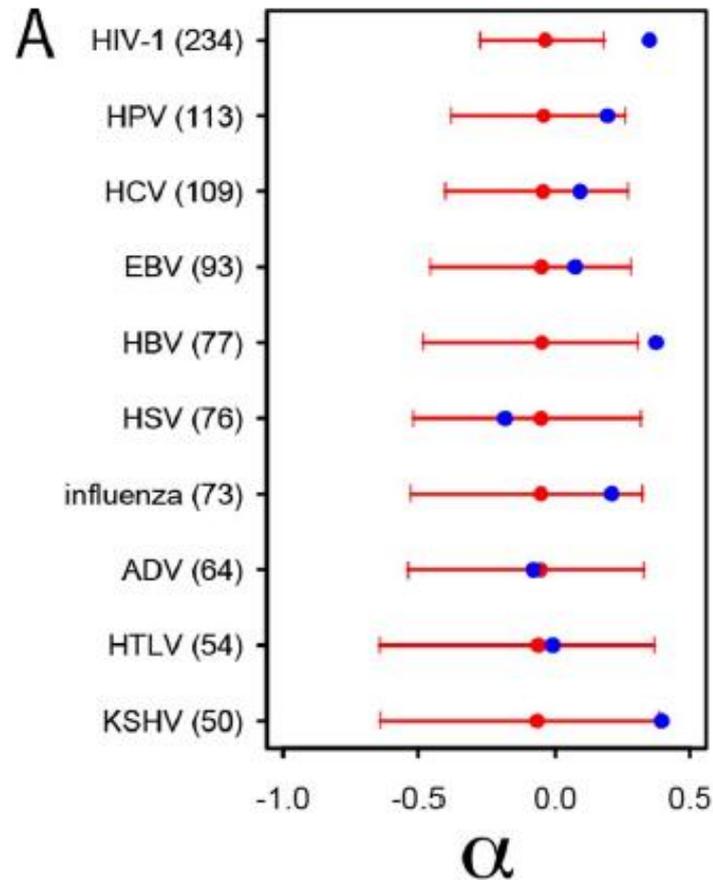


In adaptive evolution the frequency of functional substitutions (mutations which alter the function of the protein) is higher than in neutral evolution

# Origin and evolution of viruses

RED = non VIPs

BLUE = VIPs



# Virus: origine et évolution

- Les virus interagissent avec les protéines conservés (fonctionnellement importantes)
- Ces protéines sont responsable d'une part importante d'adaptation aux virus, malgré la contrainte fonctionnelle importante (perte de fonction délétère)
- Il s'agit des protéines avec des fonctions dans ou en dehors de l'immunité
- Il est estimé (estimation conservative) que les virus ont provoqué environ 30% de changement adaptatif de séquence d'acides aminés au sein de protéines conservées chez les mammifères
- **Les virus sont la force majeure d'évolution adaptative chez les mammifères**

# Virus: origine et évolution

Mammals made by viruses! (If not for a virus, none of us would ever be born.)

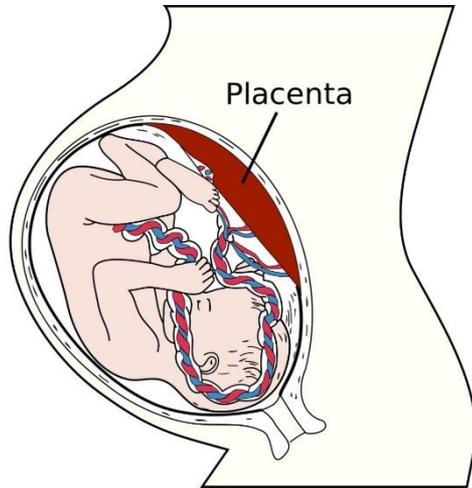


8% du génome human = séquences virales intégrées (rétrovirus)

1,5% du génome human = séquences codantes pour les protéines!

There are more retroviruses in us than there is *us* in us!

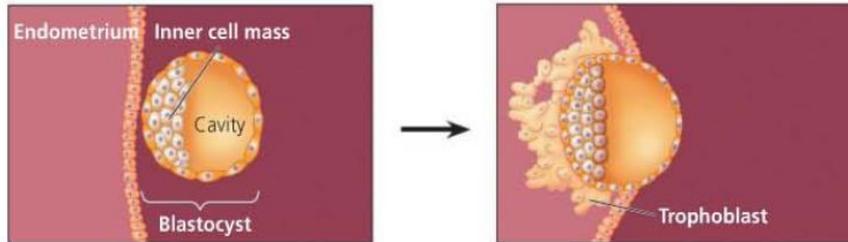
# Virus: origine et évolution



Ornithorhynchus  
(mammifère sans placenta)

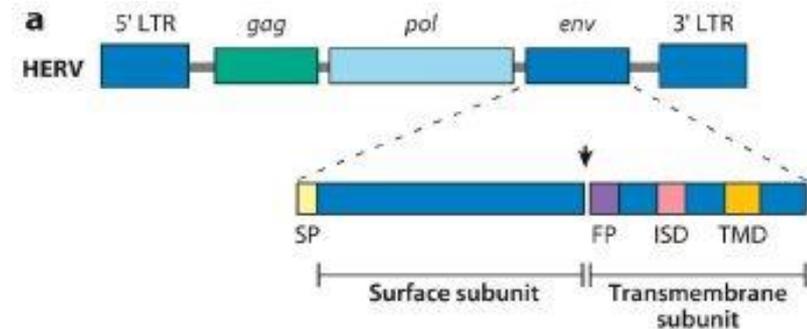
- Placenta est apparue il y a ~150 million d'années, répandue chez la grande majorité des mammifères
- Offre des avantages: protection de l'embryon, thermorégulation, humidité, nutriments

# Virus: origine et évolution

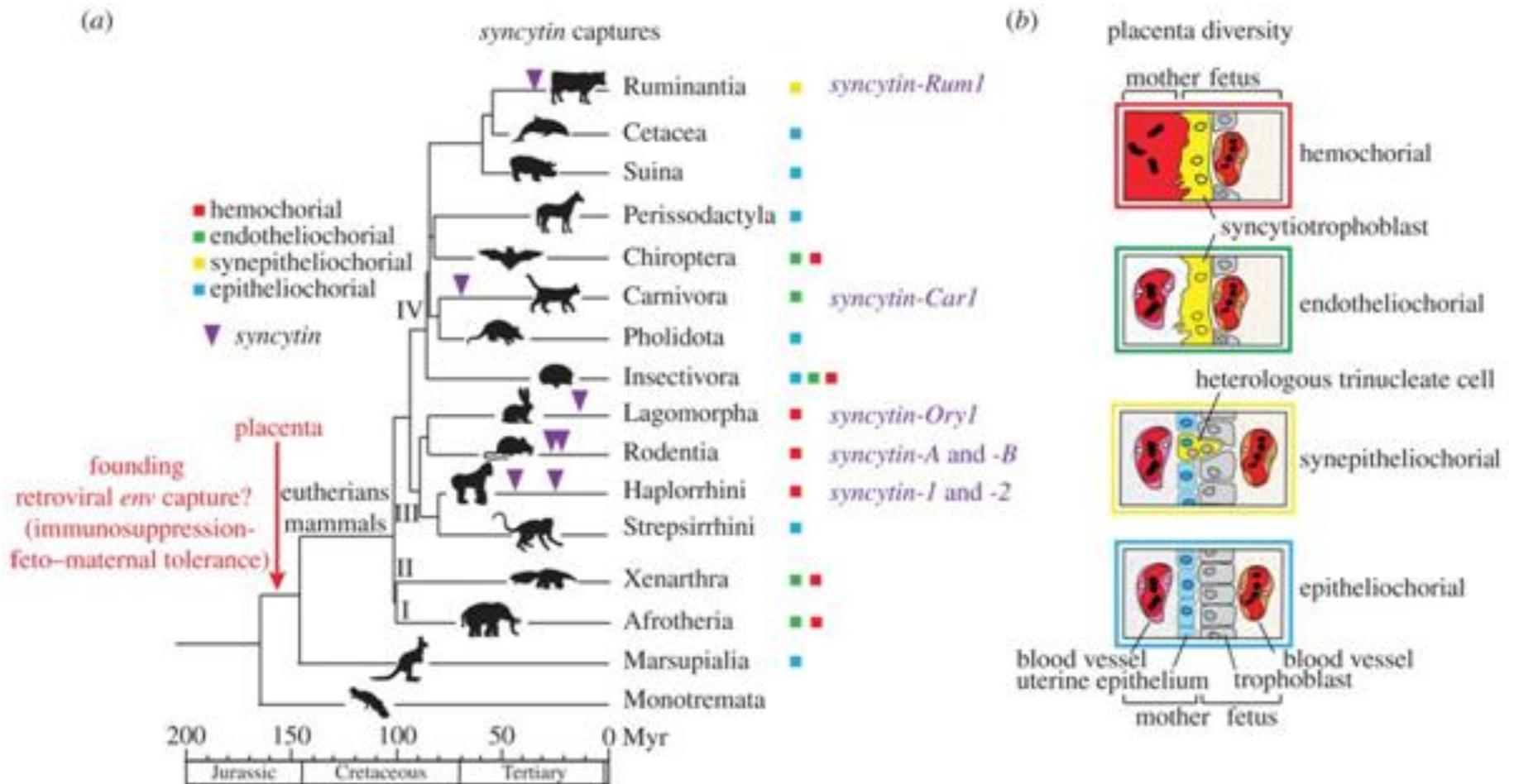


(b) Implantation of blastocyst

- 5 jours après la fertilisation formation de blastocyste
- Le trophoblast envahi la paroi utérine et prolifère
- Les cellules de trophoblaste *fusionnent* pour former placenta → grâce à des protéines nommées syncytines → action *immunosuppressive*
- Syncytines sont d'origine retrovirale!

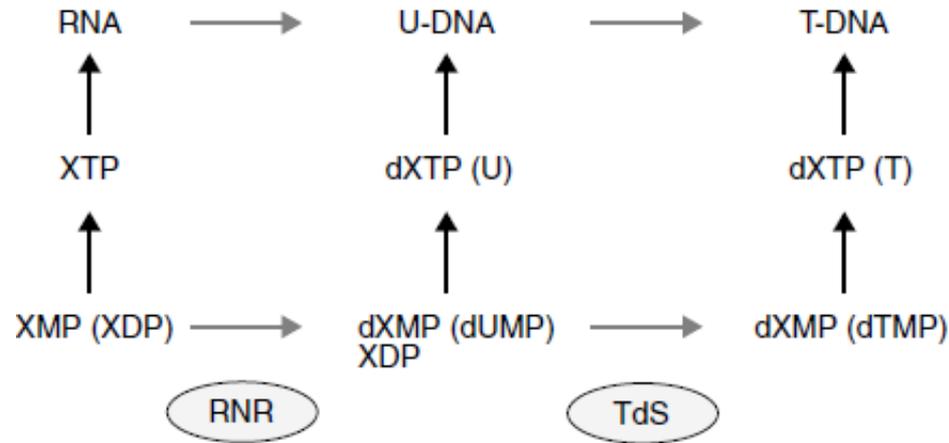


# Virus: origine et évolution



# Origin and evolution of viruses

→ How and when was DNA invented ?



→ RNA to DNA pathway in modern cells :

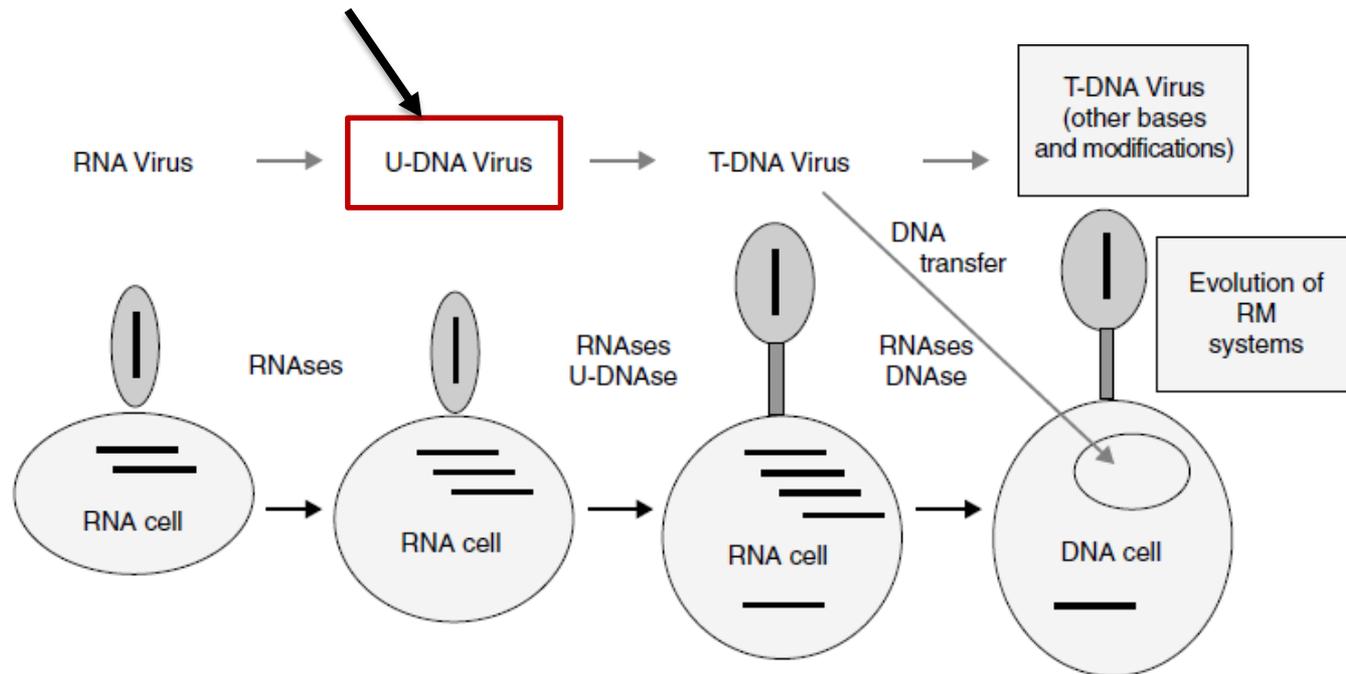
- reduction of ribonucleotides by ribonucleotide reductase
- production of dTMP from dUMP by thymidylate synthases

→ The transition RNA world → DNA world could proceed via these two stages...  
but why ?

# Origin and evolution of viruses

→ DNA was invented as a by-product of the « arms race » between viruses and cells

Selected because resistant to RNases!

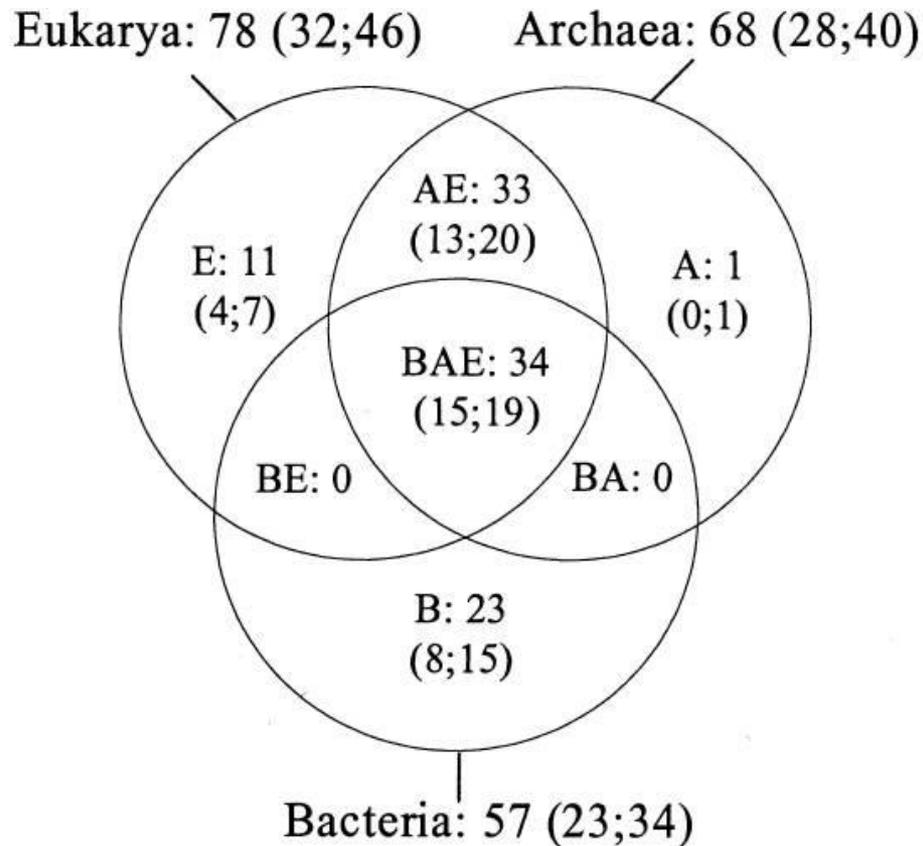


“The evolution of life is a history of virus-host interactions, an incessant “arms race” between viruses and cells”

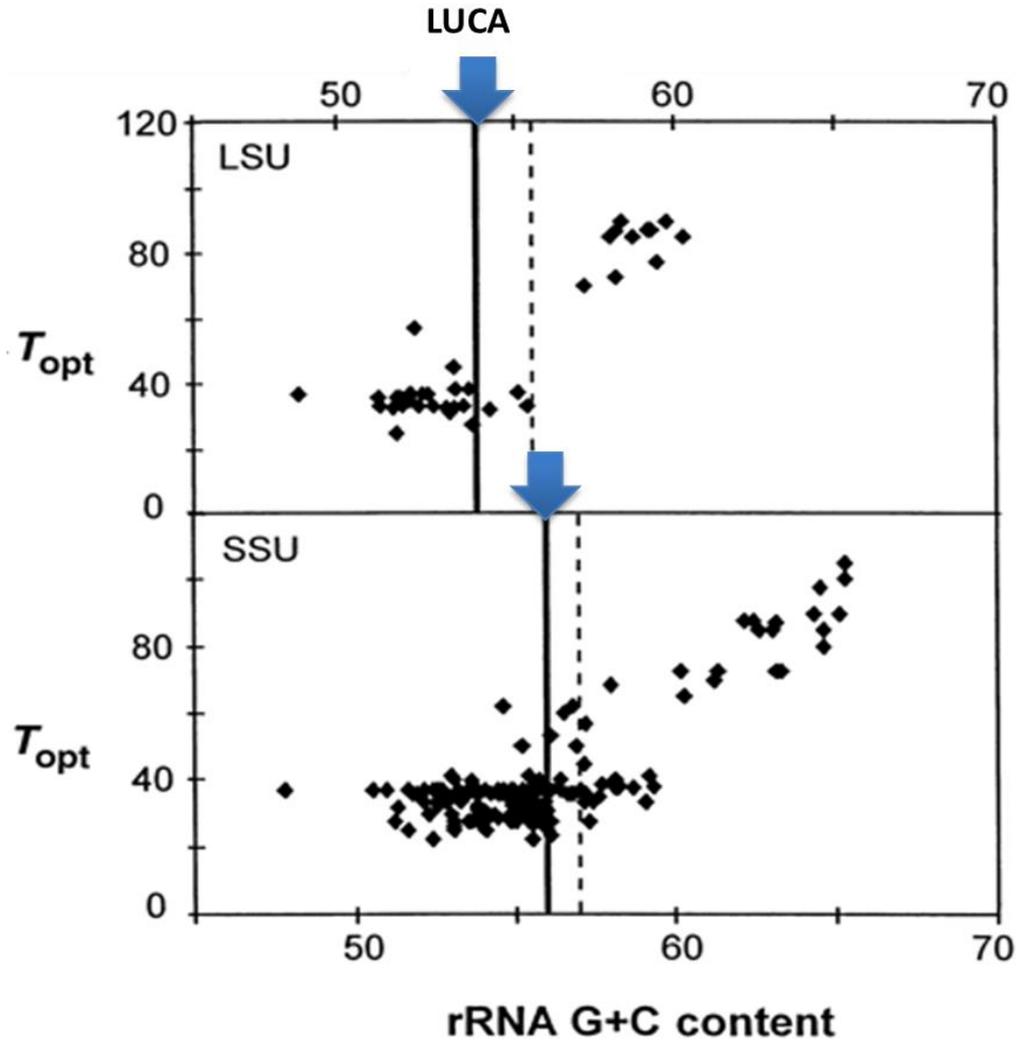
Mizuno et al. (2019) 10:752

# Questions de révision

1. Ce diagramme de Venn montre la distribution des protéines ribosomales au sein de trois domaines du vivant. Proposez 2 scénarii d'évolution du ribosome compatibles avec ces données.



2. La température optimale de croissance ( $T_{opt}$ ) d'un organisme moderne est corrélée positivement avec le % CG de son ARN ribosomal. L'analyse phylogénétique des ARN ribosomaux peut être utilisée pour reconstruire les séquences ancestrales et ainsi déduire la  $T_{opt}$  du LUCA. En s'appuyant sur le graphique ci-dessous expliquez pourquoi ces données suggèrent que LUCA n'était pas un organisme thermophile.



3. De nombreux virus codent les protéines ribosomales. L'un des exemples traités en cours est la protéine bS21 porté par un phage infectant une bactérie du genre *Pelagibacter* sp. Ci-dessous est montré l'arbre phylogénétique (maximum vraisemblance) construit à partir des orthologues de la protéine bS21. Selon cet arbre d'où provient l'orthologue que l'on trouve chez le phage (surligné en gris) ?

