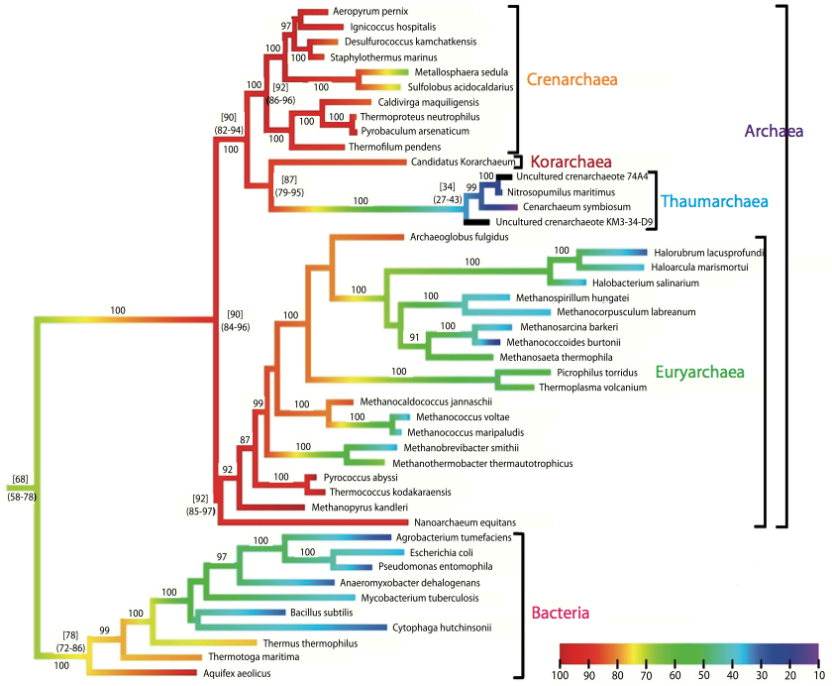
Thermomètres moléculaires: reconstruire les séquences ancestrales pour comprendre l'histoire de l'adaptation à la température le long de l'arbre de la vie

Manolo Gouy

Laboratoire de Biométrie & Biologie Evolutive
CNRS / Université de Lyon



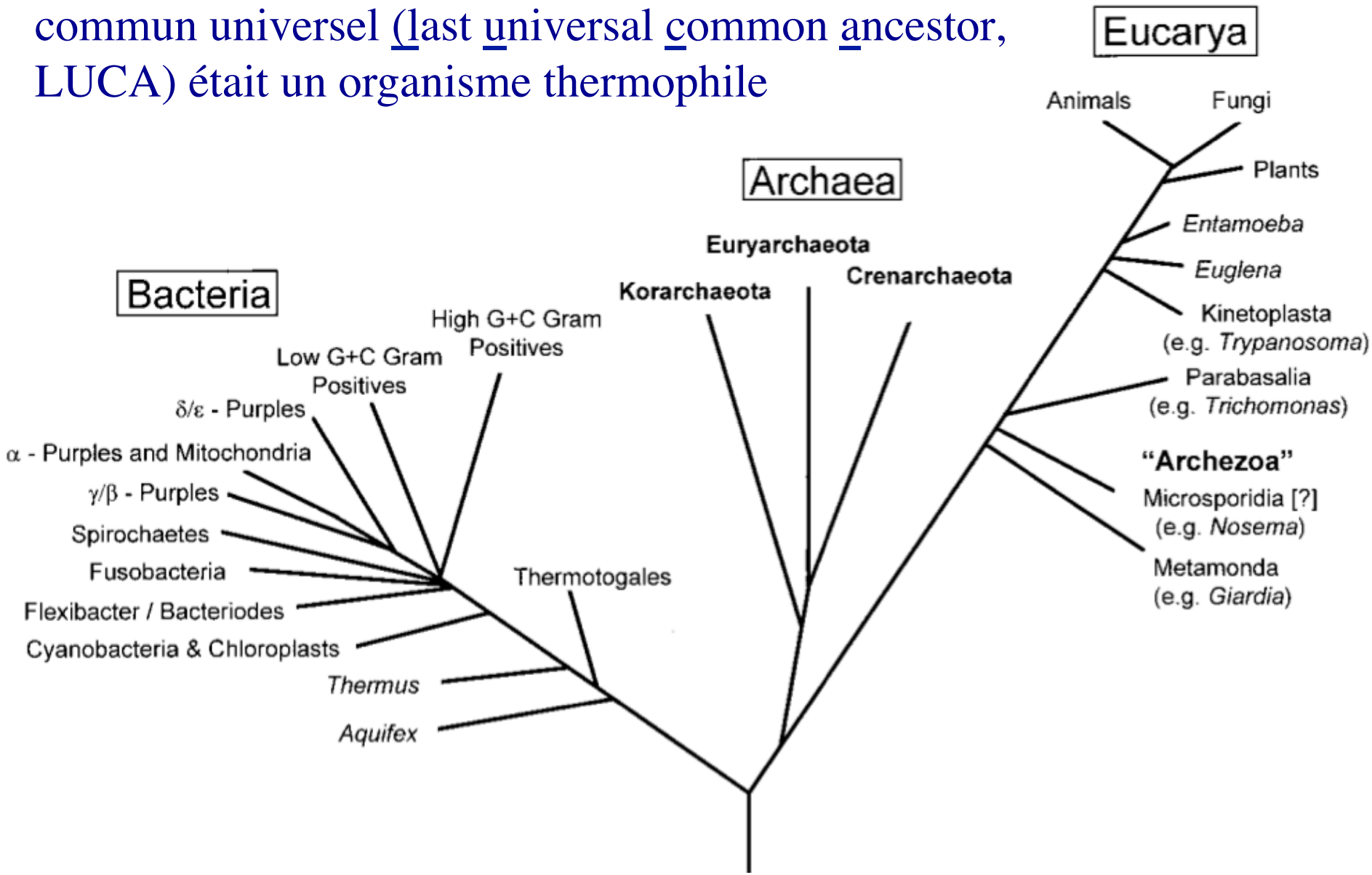
Conférence Société Française d'Exobiologie,
La Baule, 8 octobre 2014

cnrs

UNIVERSITÉ DE LYON

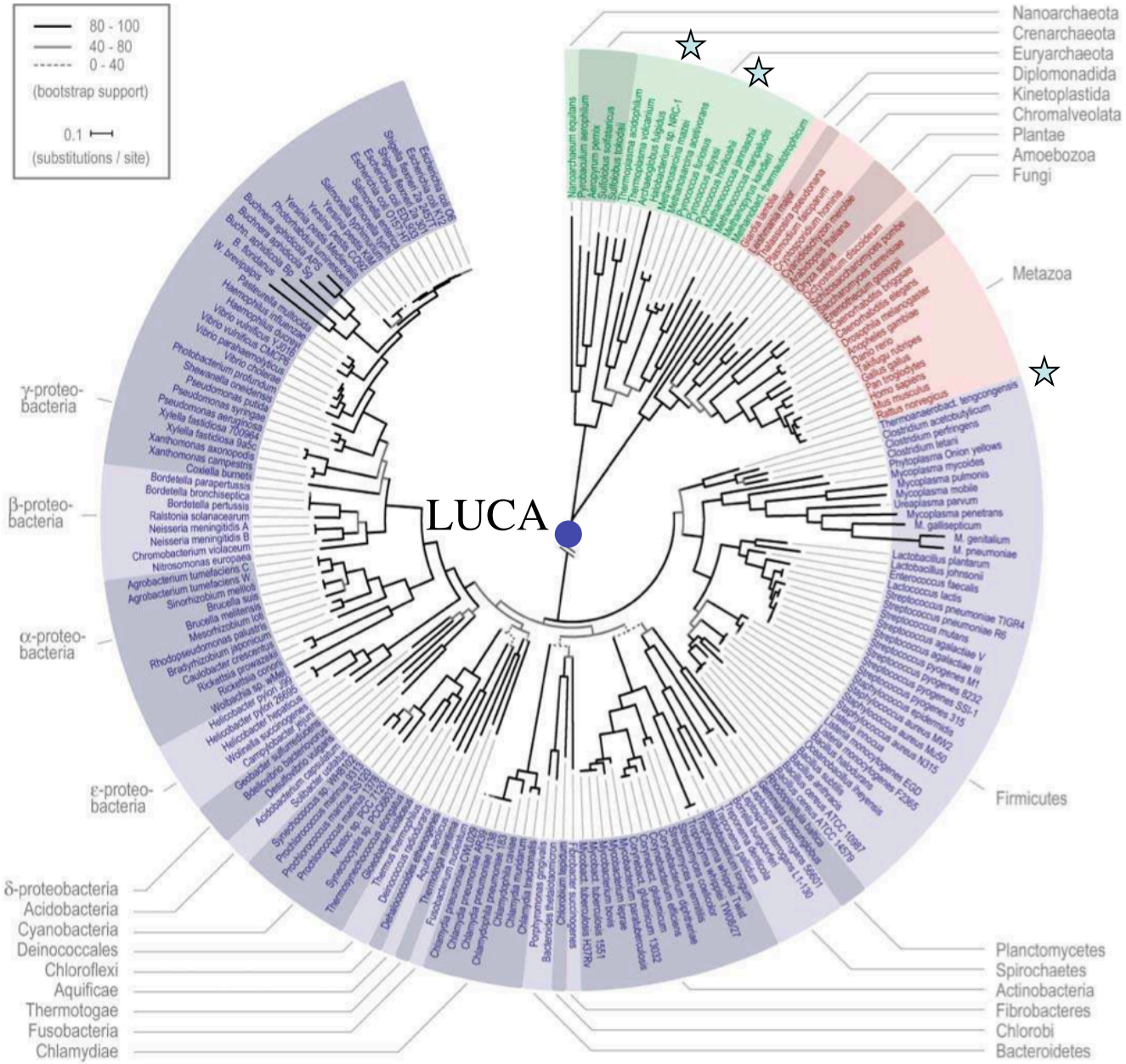
LBBE
BIOMETRIE ET BIOLOGIE EVOLUTIVE

Il a été souvent proposé que le dernier ancêtre commun universel (last universal common ancestor, LUCA) était un organisme thermophile



Vision traditionnelle des trois domaines de la vie

Un autre type d'argument en faveur de la thermophilie de LUCA



Les organismes aux taux d'évolution les plus lents (ayant les plus courtes distances racine-feuille) ☆ *Thermoanaerobacter*, *Pyrococcus*, *Archaeoglobus* sont (hyper)thermophiles

Ciccarelli *et al*., 2006

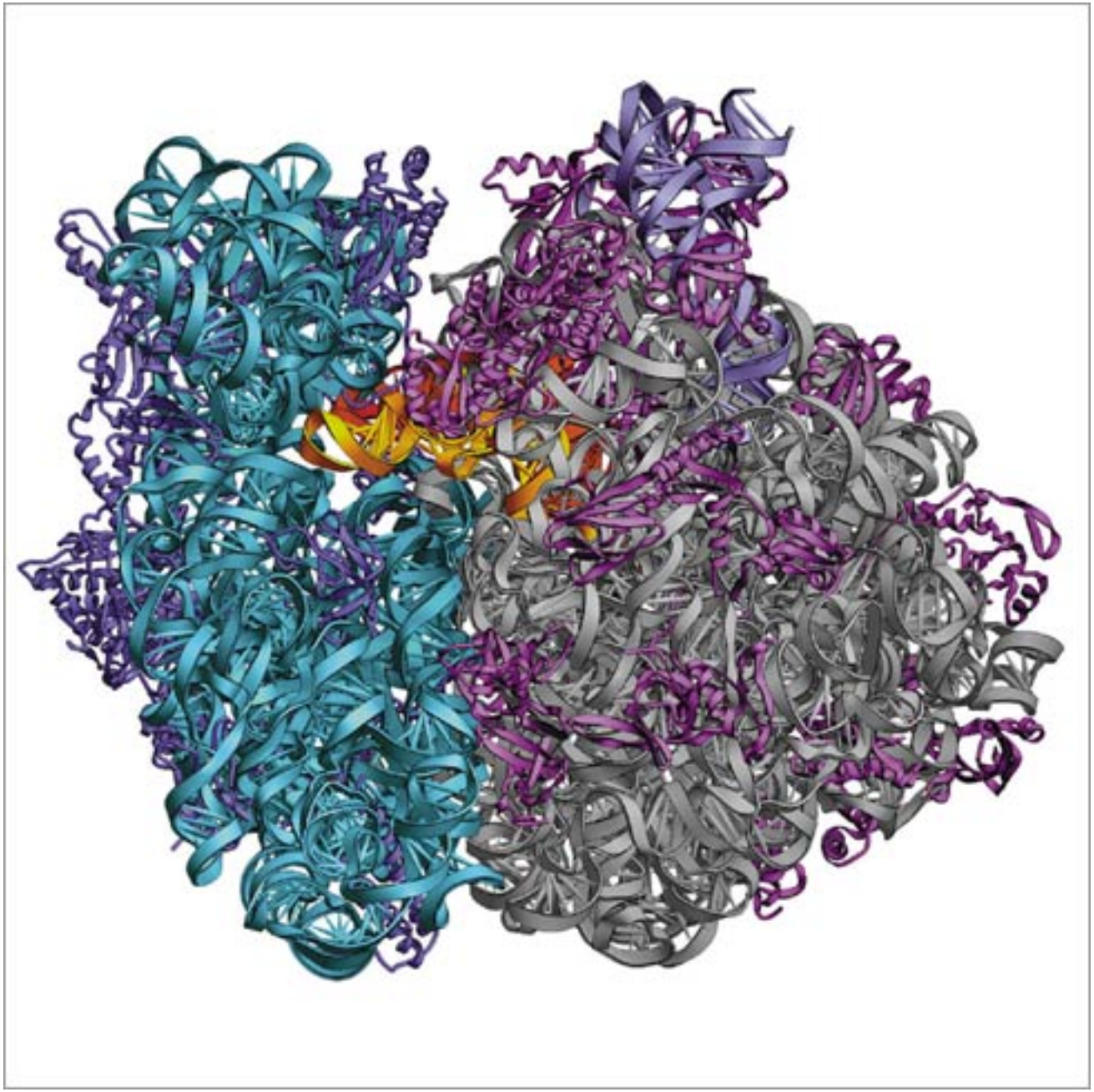
Plan

1. Le lien entre macromolécules biologiques et températures environnementales : les thermomètres moléculaires

2. L'estimation précise des compositions ancestrales nécessite l'emploi de modèles évolutifs non-homogènes

3. Estimation des compositions en bases et en acides aminés des ARN ribosomiques et des protéines de LUCA et des ancêtres des domaines

4. Interprétation des résultats

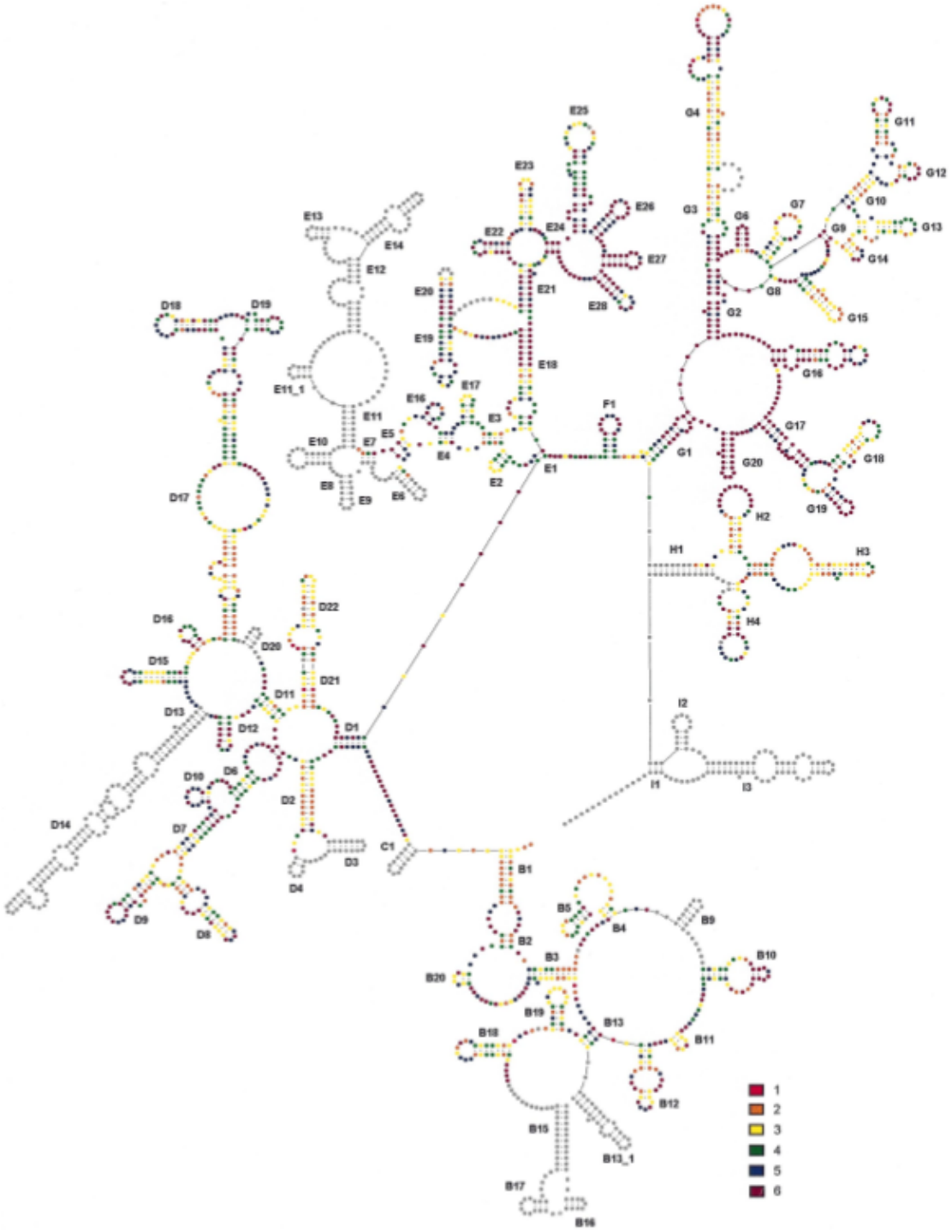


Modèle tri-dimentionnel du ribosome.

La grande sous-unité est à droite.

Les rRNA sont en bleu à gauche et en gris à droite.

Les protéines sont en violet.

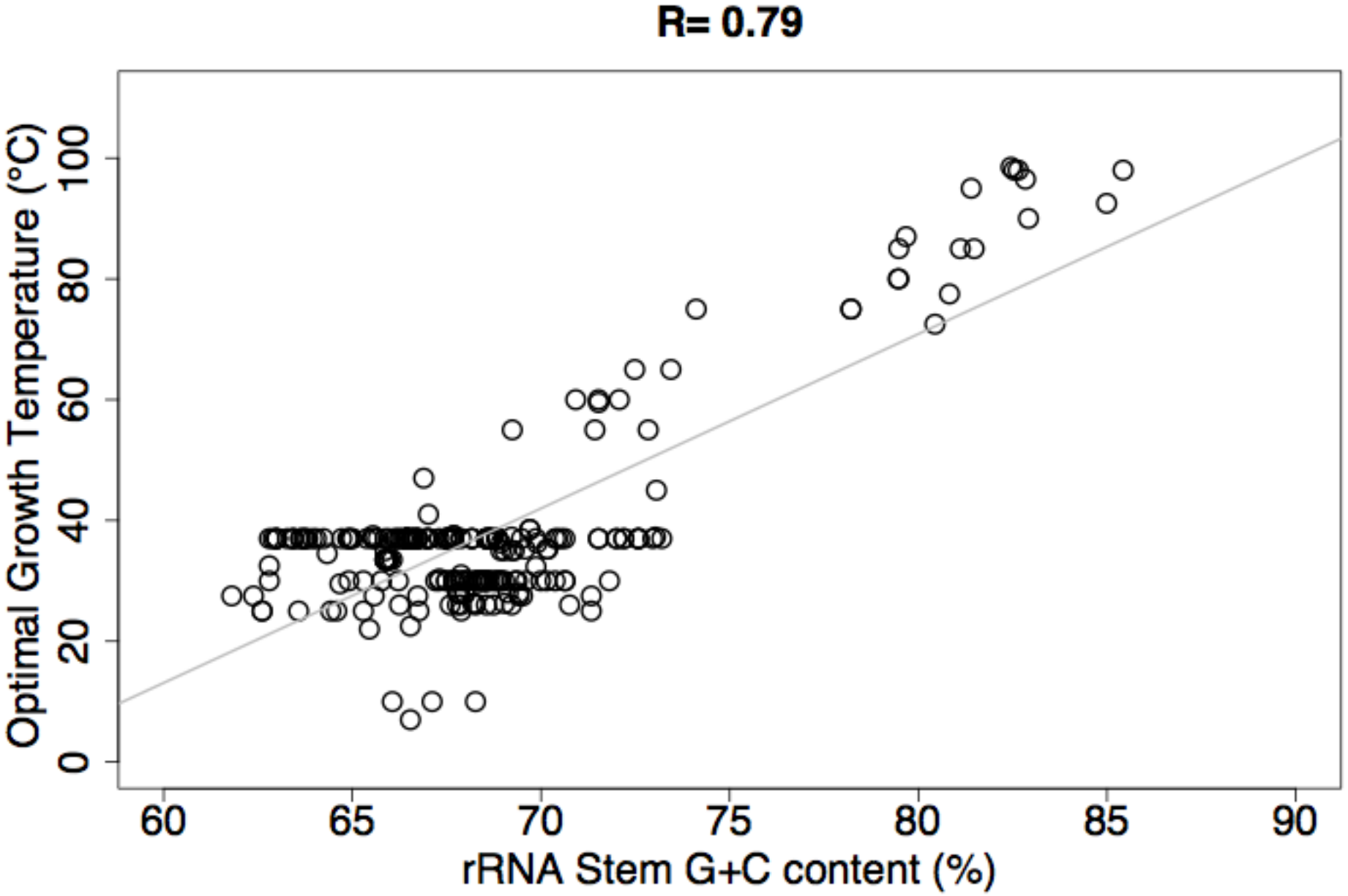


Modèle de la structure secondaire de l'ARN ribosomique de la grande sous-unité.

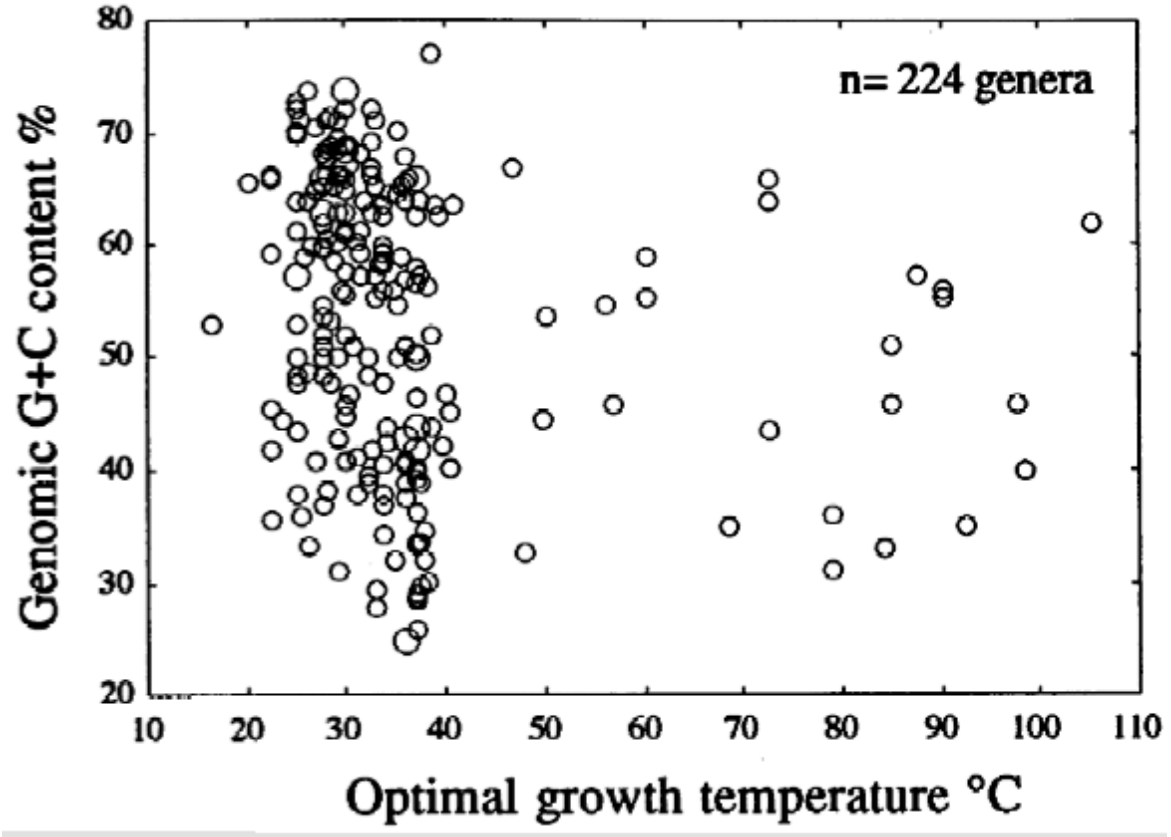
paires G.C: 3 liaisons hydrogène

paires A.U: 2 liaisons hydrogène

Le contenu en G+C des ARN ribosomiques est corrélé à la température optimale de croissance de 275 bactéries et archées



La corrélation GC / T° ne s'applique pas au G+C génomique !



Galtier & Lobry (1997) *J Mol Evol* 44:632

La composition moyenne en acides aminés des protéines est corrélée aux températures optimales de croissance

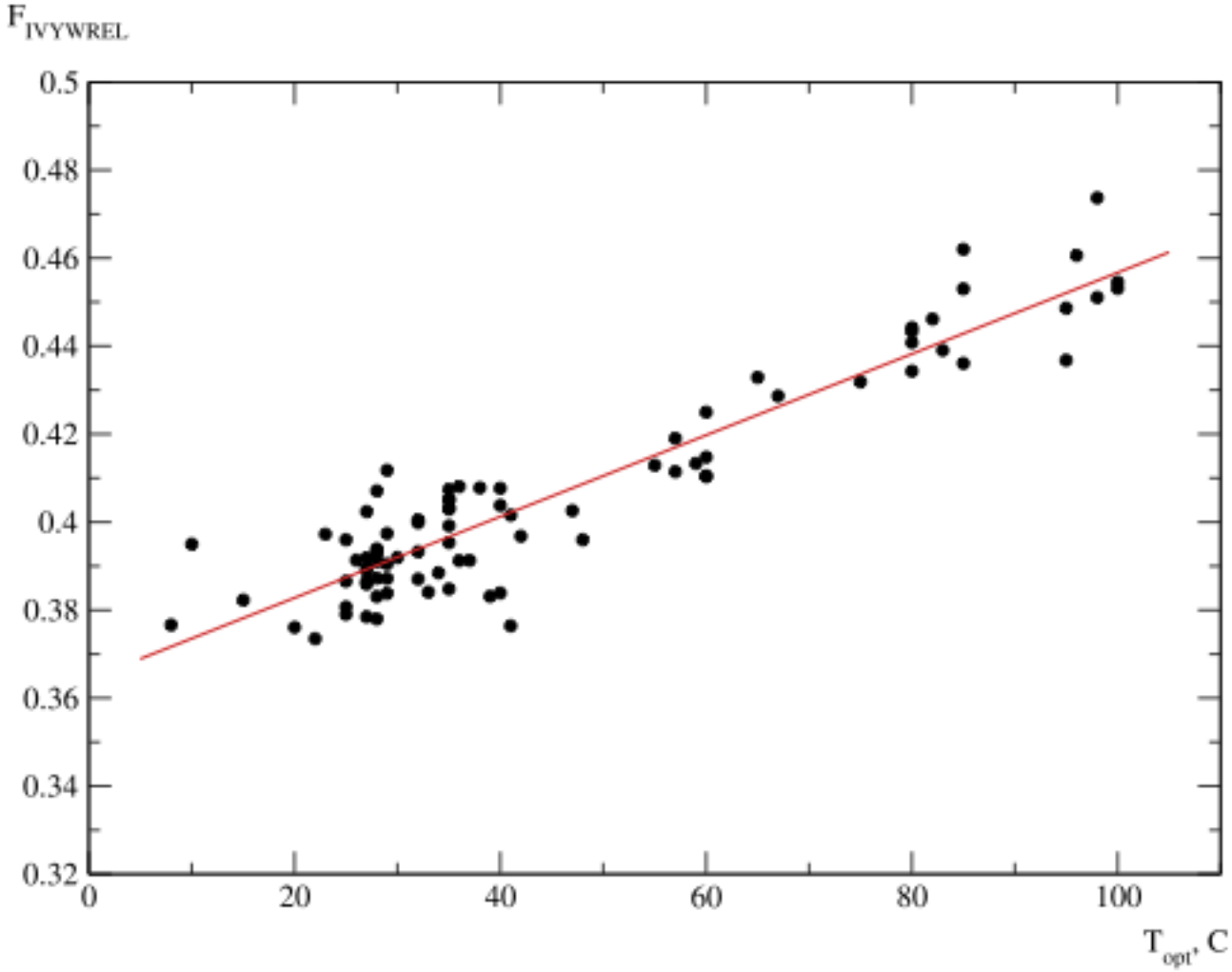
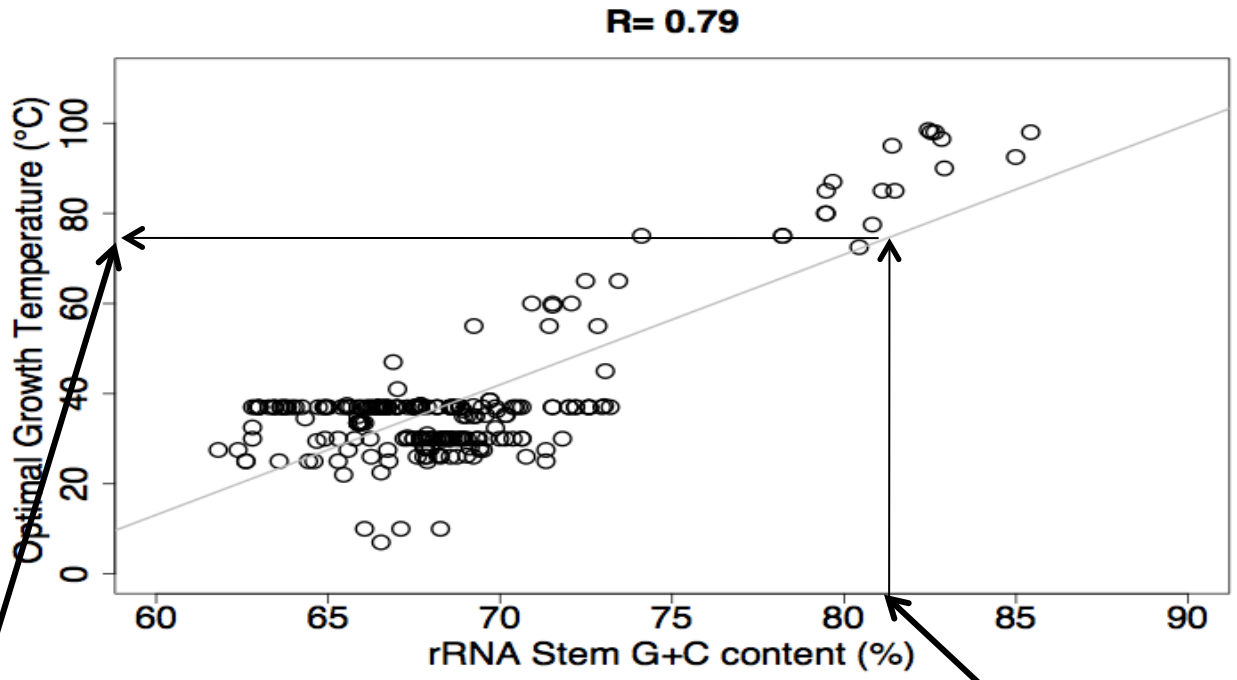


Figure 1. Correlation between the Sum *F* of Fractions of Ile, Val, Tyr, Trp, Arg, Glu, and Leu (IVYWREL) Amino Acids in 86 Proteomes and the OGT of Organisms T_{opt}

Zeldovitch et al. (2007) *PLOS Comp Biol* 3(1):e5

Comment utiliser un « thermomètre moléculaire »



(1) Estimer les compositions ancestrales

(2) en déduire les températures environnementales ancestrales

Plan

1. Le lien entre macromolécules biologiques et températures environnementales : les thermomètres moléculaires

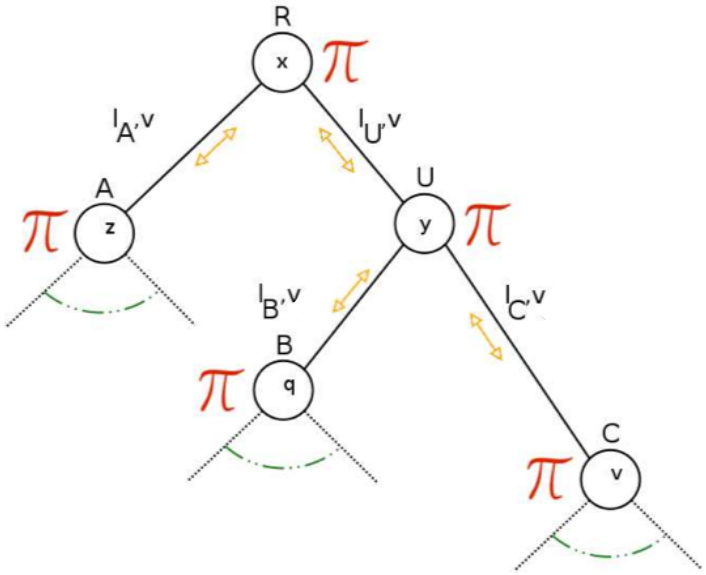
2. L'estimation précise des compositions ancestrales nécessite l'emploi de modèles évolutifs non-homogènes

3. Estimation des compositions en bases et en acides aminés des ARN ribosomiques et des protéines de LUCA et des ancêtres des domaines

4. Interprétation des résultats

Modélisation homogène du processus évolutif

Le processus évolutif est modélisé par les probabilités des substitutions le long des branches, que l'on peut décomposer en une magnitude (l_i) et des taux relatifs (v_i). Classiquement, on suppose que ces probabilités varient entre branches en magnitude (l_i) mais pas en taux relatifs (v) : les taux relatifs de chaque type de substitution (ex: G → T *vs*. G → C) sont invariants tout au long de l'arbre.

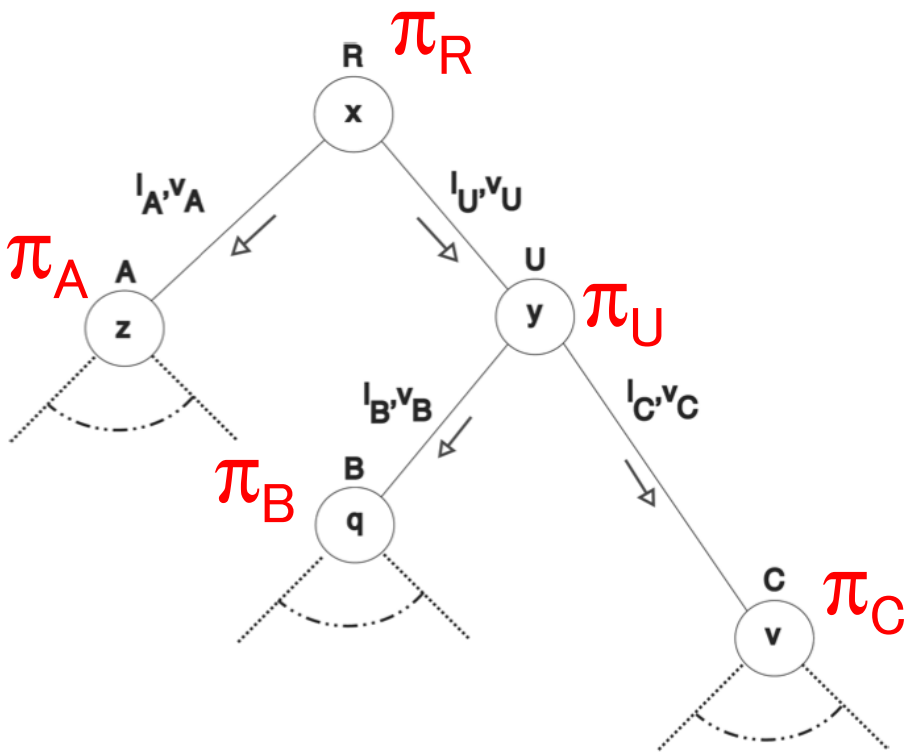


Une conséquence majeure de cette supposition: les séquences ont nécessairement la même composition en bases ou en acides aminés (π) tout le long de l'arbre.

⇒ cette supposition est inadéquate pour utiliser un thermomètre moléculaire

Modélisation non-homogène du processus évolutif

Chaque branche de l'arbre peut avoir sa propre magnitude (l_i) et ses propres taux relatifs (v_i) d'évolution.



- Des compositions différentes aux feuilles de l'arbre peuvent évoluer.
- La reconstruction des compositions ancestrales est plus précise.

Plan

1. Le lien entre macromolécules biologiques et températures environnementales : les thermomètres moléculaires

2. L'estimation précise des compositions ancestrales nécessite l'emploi de modèles évolutifs non-homogènes

3. Estimation des compositions en bases et en acides aminés des ARN ribosomiques et des protéines de LUCA et des ancêtres des domaines

4. Interprétation des résultats

Les données

1. Concaténation des séquences d'ARN ribosomiques de la grande et de la petite sous-unité de 456 organismes (bactéries, archées, eucaryotes).

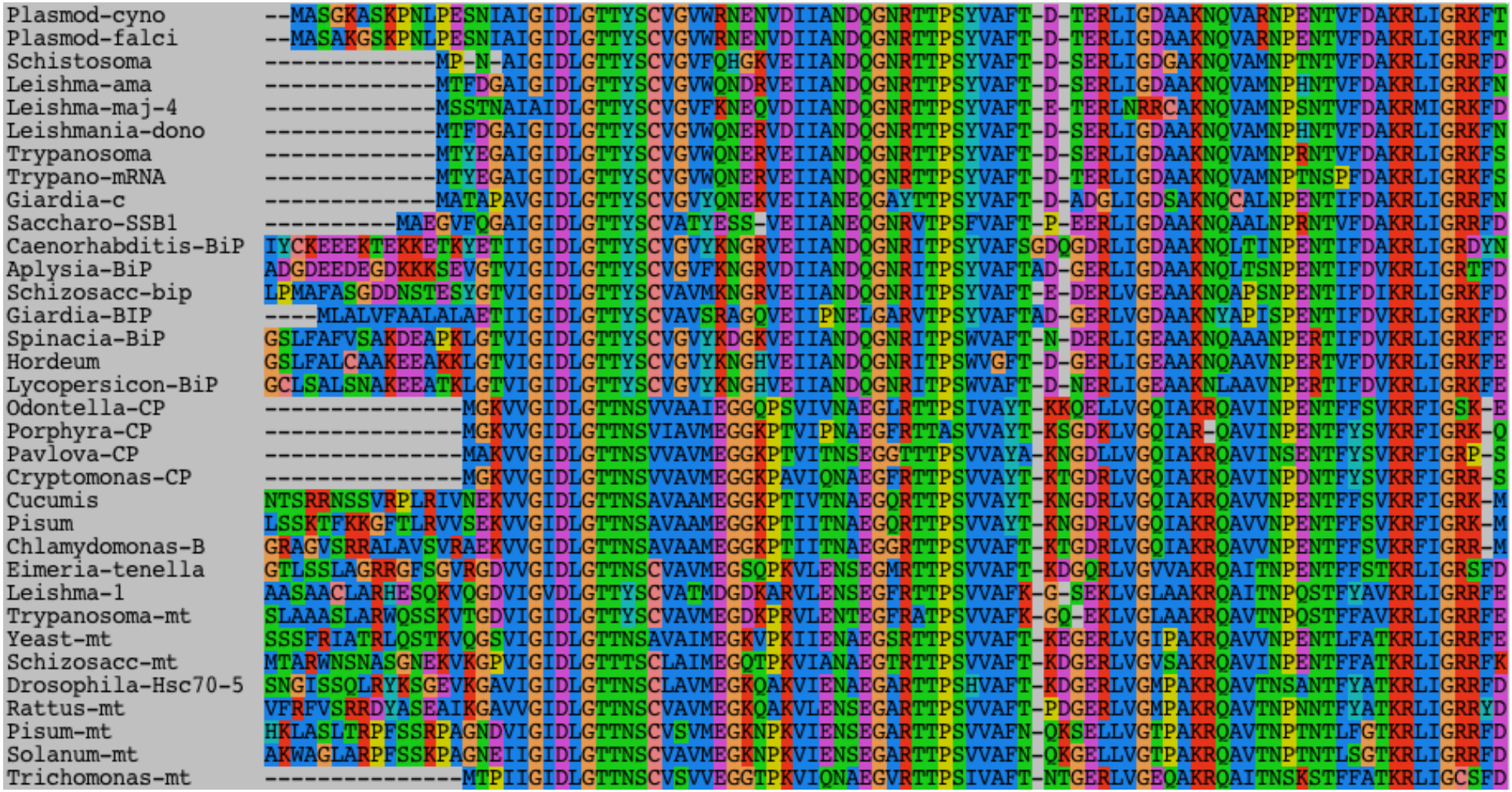
   ▪ Les séquences complètes (2924 sites) ont été utilisées pour estimer la forme de l'arbre phylogénétique universel,

   ▪ seules les régions en double-brin (1043 sites) ont permis d'estimer les G+C ancestraux.

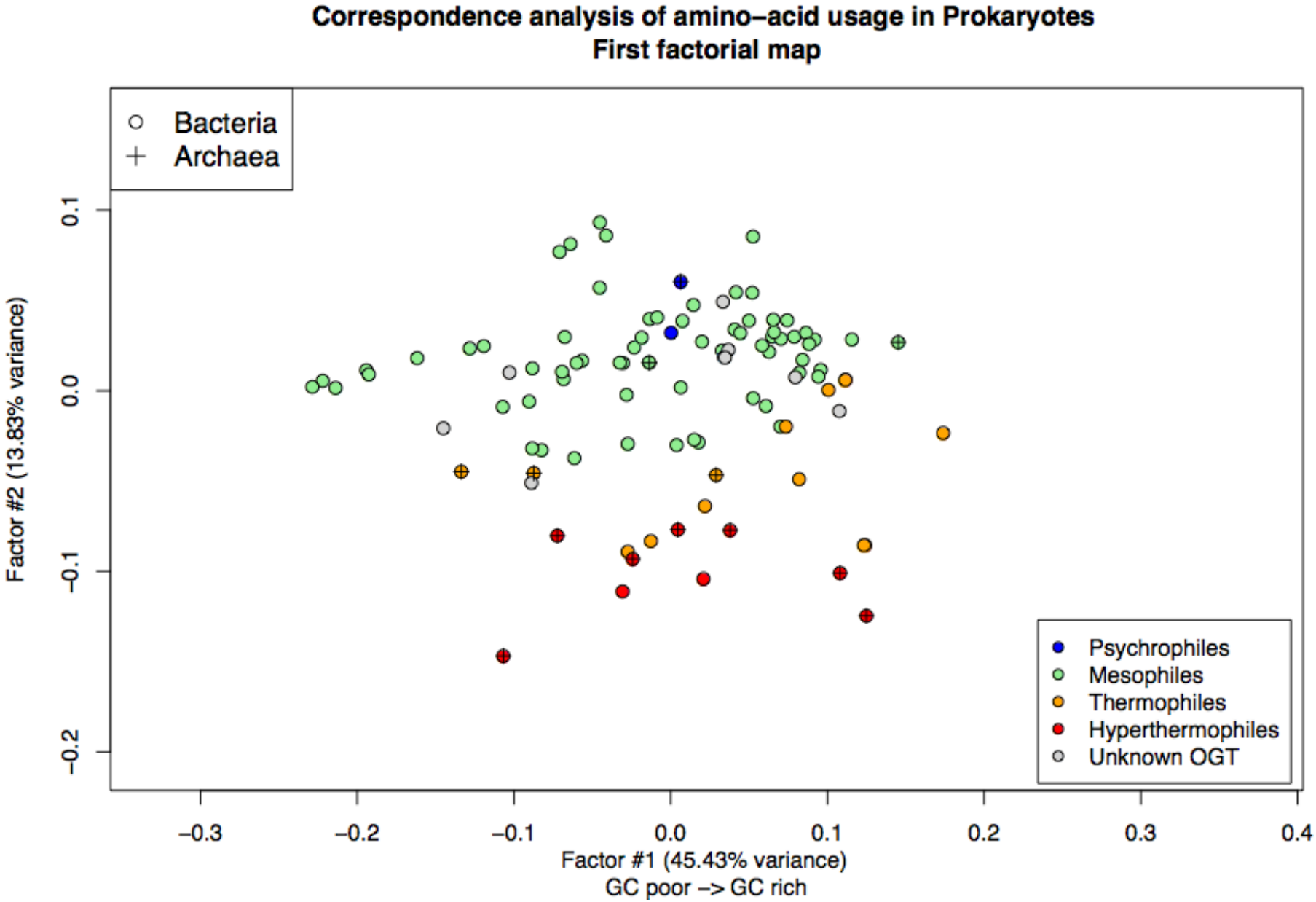
2. Concaténation de 56 séquences protéiques universelles et hautement conservées de 38 organismes (4946 sites).

➔ Estimation des compositions ancestrales en G+C et en acides aminés par des méthodes non-homogènes.

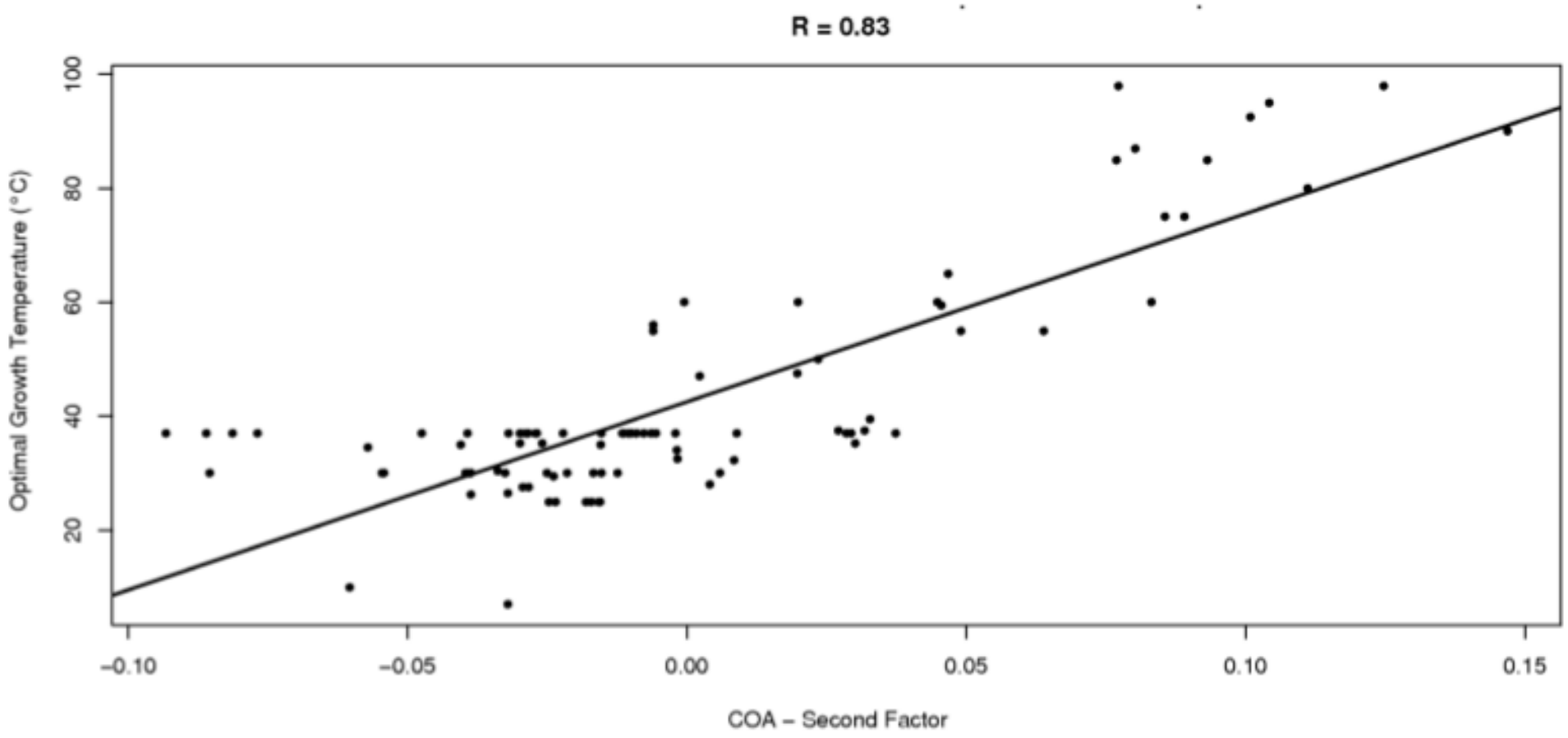
Exemple de données de séquences protéiques



Le thermomètre moléculaire des protéines (1)



Le thermomètre moléculaire des protéines (2)



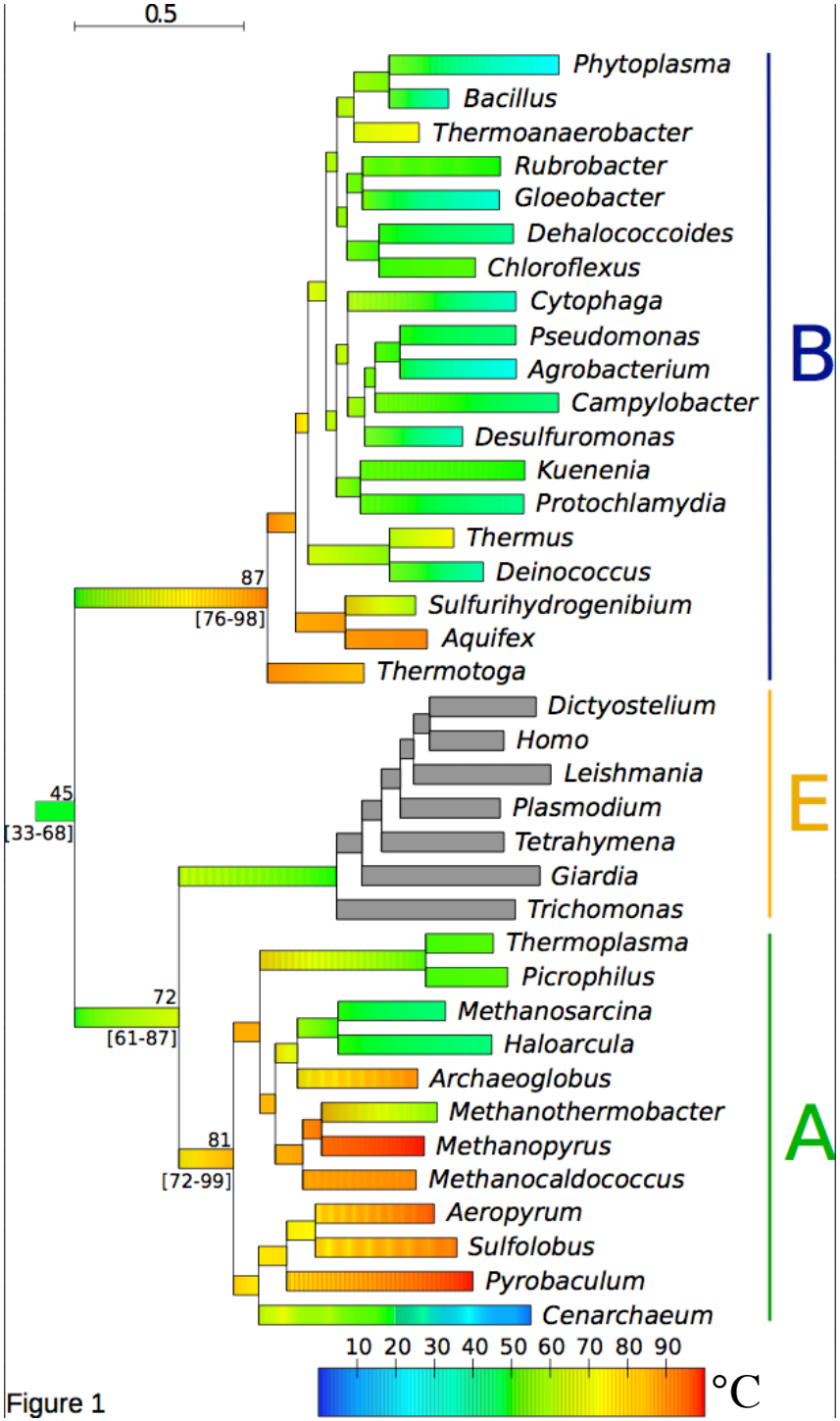


Figure 1

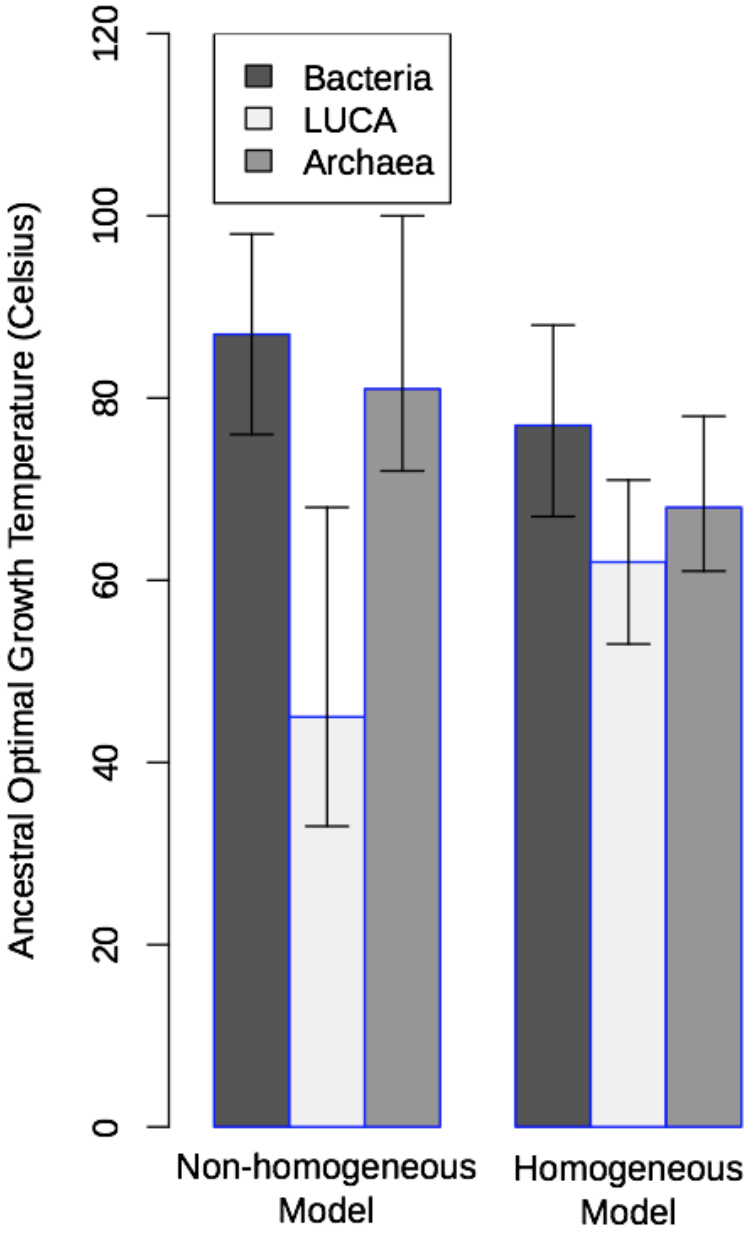
Evolution de la thermophilie le long de l'arbre de la vie prédit par analyse des séquences protéiques

- Les ancêtres des domaines bactériens (B) et archéens (A) sont prédits comme des organismes (hyper)thermophiles.

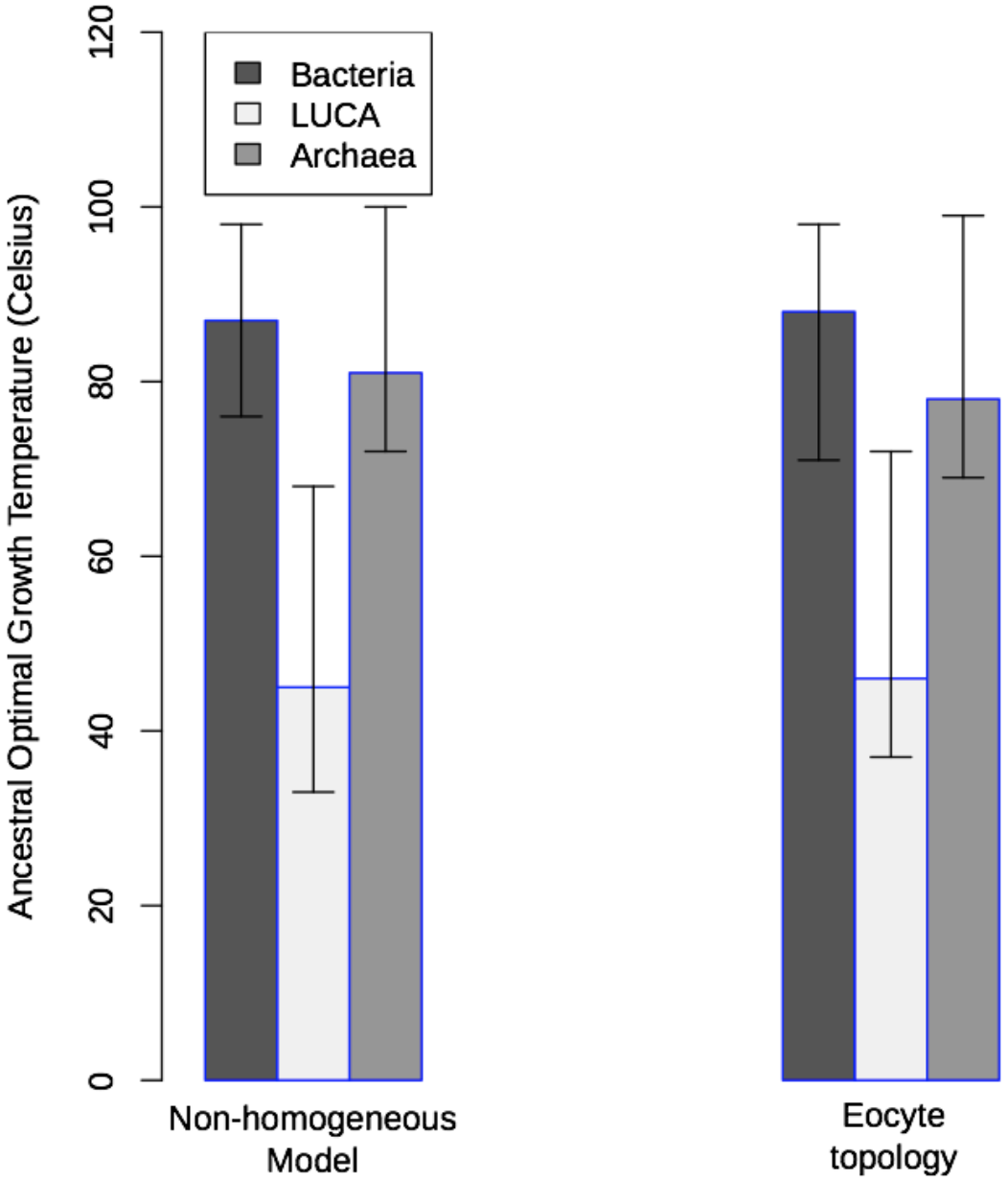
- LUCA est prédit comme un organisme mésophile.

Boussau *et al*. (2008) *Nature* 456:942

Expériences de contrôle : (1) effet du modèle non-homogène



Expériences de contrôle : (2) effet de topologies alternatives



Plan

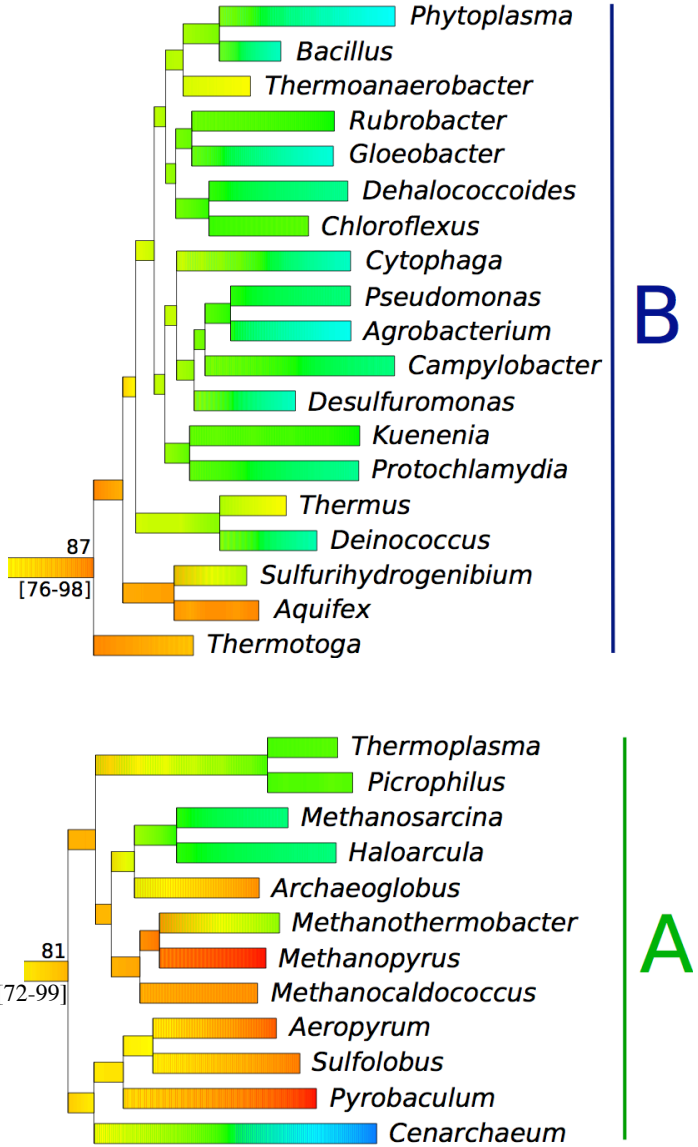
1. Le lien entre macromolécules biologiques et températures environnementales : les thermomètres moléculaires

2. L'estimation précise des compositions ancestrales nécessite l'emploi de modèles évolutifs non-homogènes

3. Estimation des compositions en bases et en acides aminés des ARN ribosomiques et des protéines de LUCA et des ancêtres des domaines

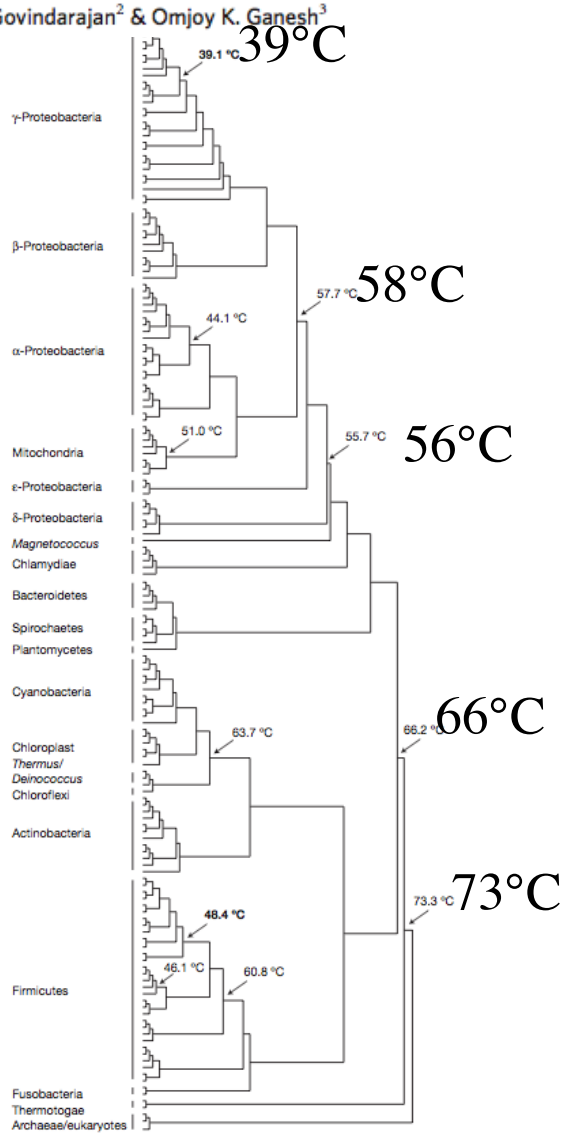
4. Interprétation des résultats

Première inférence:
état ancestral thermophile des bactéries et des archées



Palaeotemperature trend for Precambrian life inferred from resurrected proteins

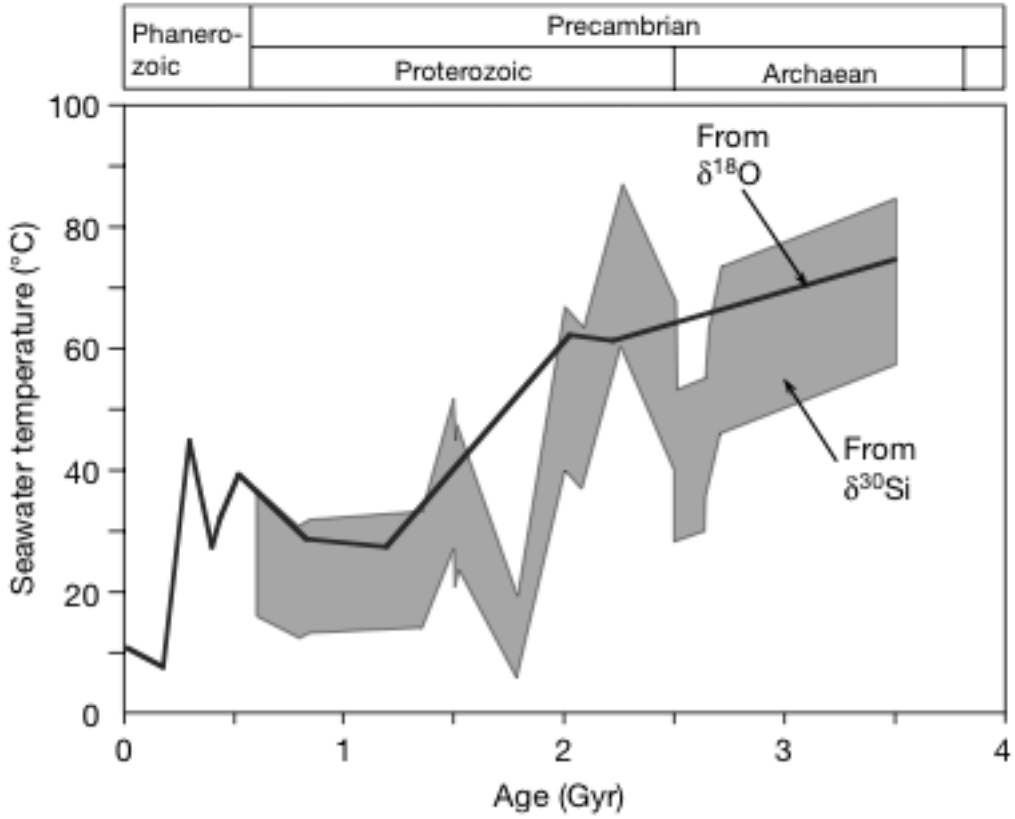
Eric A. Gaucher[1], Sridhar Govindarajan[2] & Omjoy K. Ganesh[3]



A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts

François Robert[1] & Marc Chaussidon[2]

NATURE|Vol 443|26 October 2006



Deuxième inférence:
deux transitions indépendantes vers la thermophilie

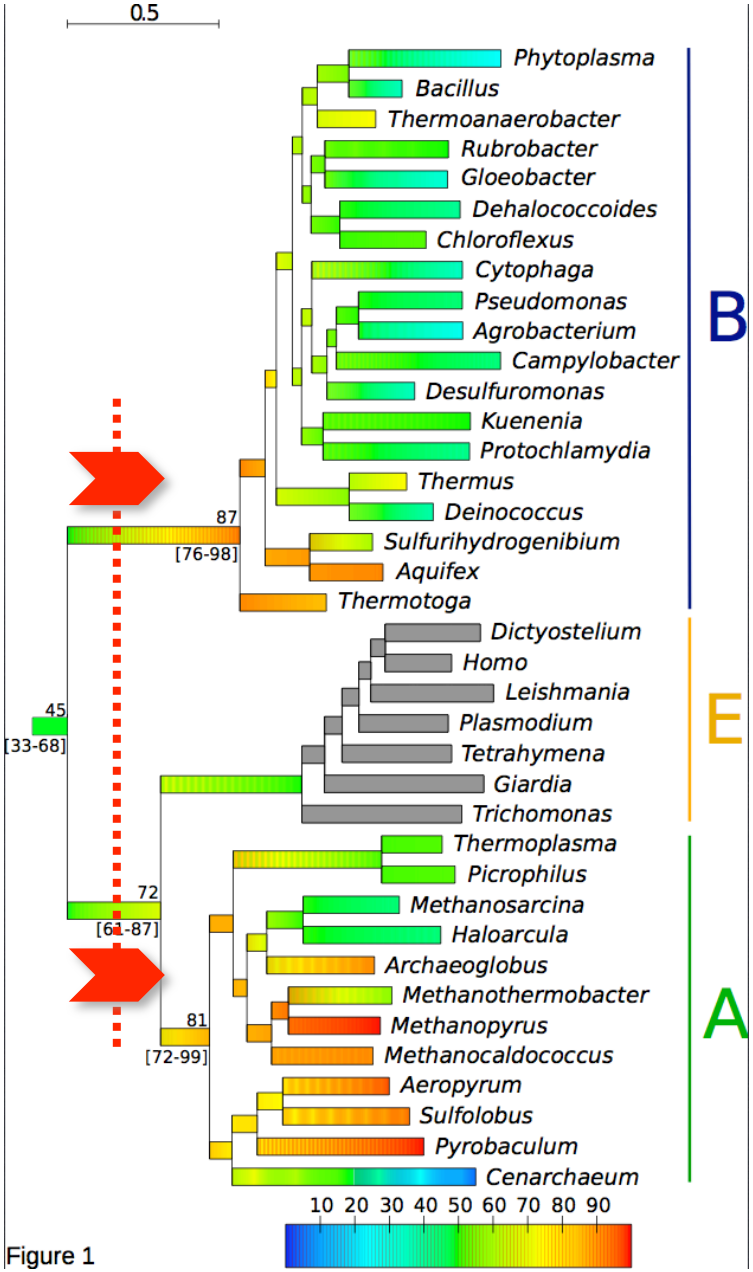


Figure 1

The habitat and nature of early life

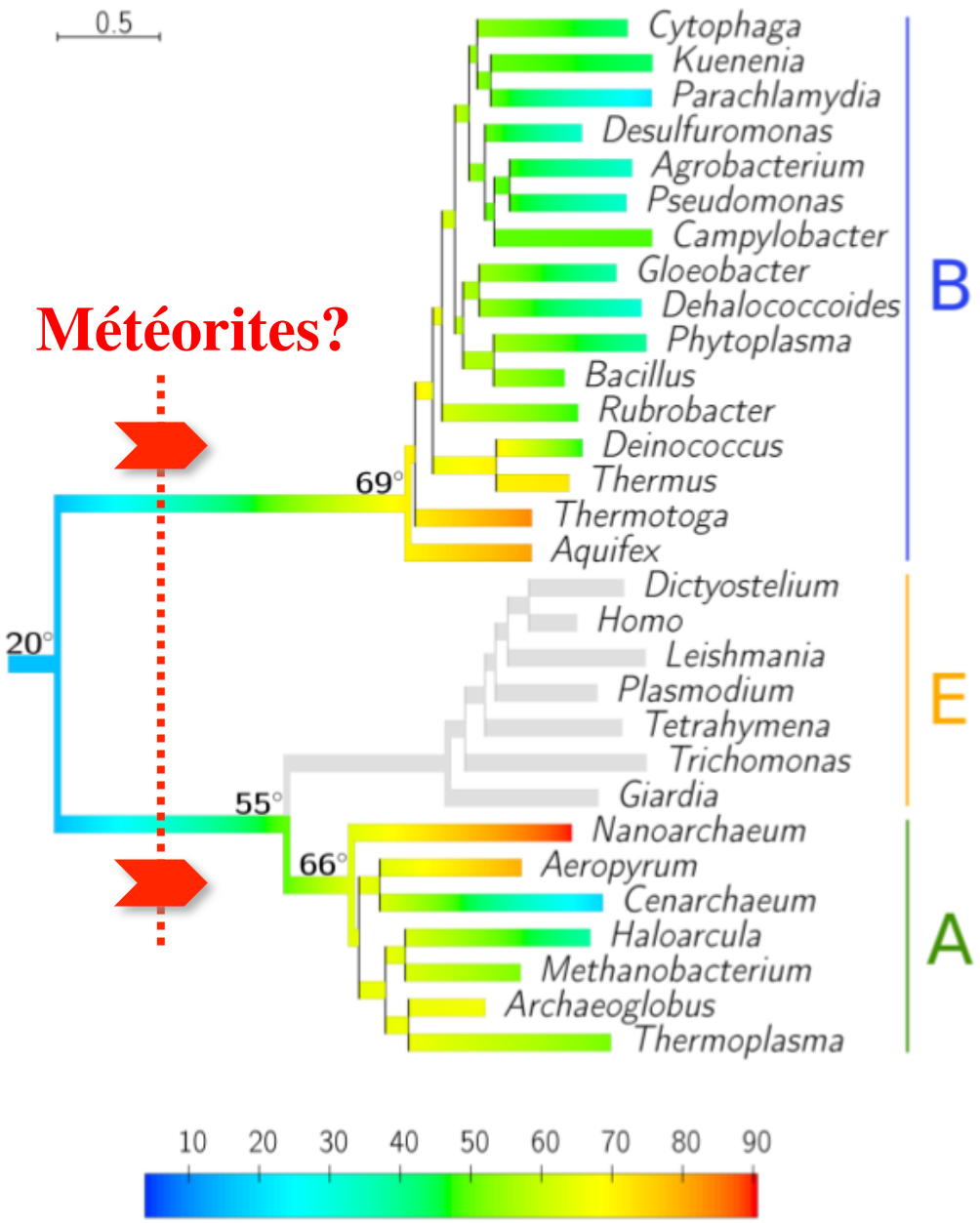
E. G. Nisbet* & N. H. Sleep†

« Earth is over 4,500 million years old.

Massive bombardment of the planet took place for the first 500–700 million years, and the largest impacts would have been capable of sterilizing the planet.

Probably until 4,000 million years ago or later, occasional impacts might have heated the ocean over 100°C. Life on earth dates from before about 3,800 million years ago, and is likely to have gone through one or more hot-ocean 'bottlenecks'.

Only hyperthermophiles (organisms optimally living in water at 80–110°C) would have survived. »


0.5
Météorites?
Cytophaga
Kuenenia
Parachlamydia
Desulfuromonas
Agrobacterium
Pseudomonas
Campylobacter
Gloeobacter
Dehalococcoides
Phytoplasma
Bacillus
Rubrobacter
Deinococcus
Thermus
Thermotoga
Aquifex
Dictyostelium
Homo
Leishmania
Plasmodium
Tetrahymena
Trichomonas
Giardia
Nanoarchaeum
Aeropyrum
Cenarchaeum
Haloarcula
Methanobacterium
Archaeoglobus
Thermoplasma
B
E
A
69°
20°
55°
66°
10 20 30 40 50 60 70 80 90

Remerciements

Bastien Boussau

Nicolas Lartillot

Mathieu Groussin

Laboratoire de Biométrie & Biologie Evolutive

CNRS / Université de Lyon

Anamaria Necsulea

Ecole Polytechnique Fédérale de Lausanne

Samuel Blanquart

Laboratoire d'Informatique Fondamentale de Lille

The molecular signal for the adaptation to cold temperature during early life on Earth

Mathieu Groussin[1], Bastien Boussau[1,2], Sandrine Charles[1], Samuel Blanquart[3] and Manolo Gouy[1]

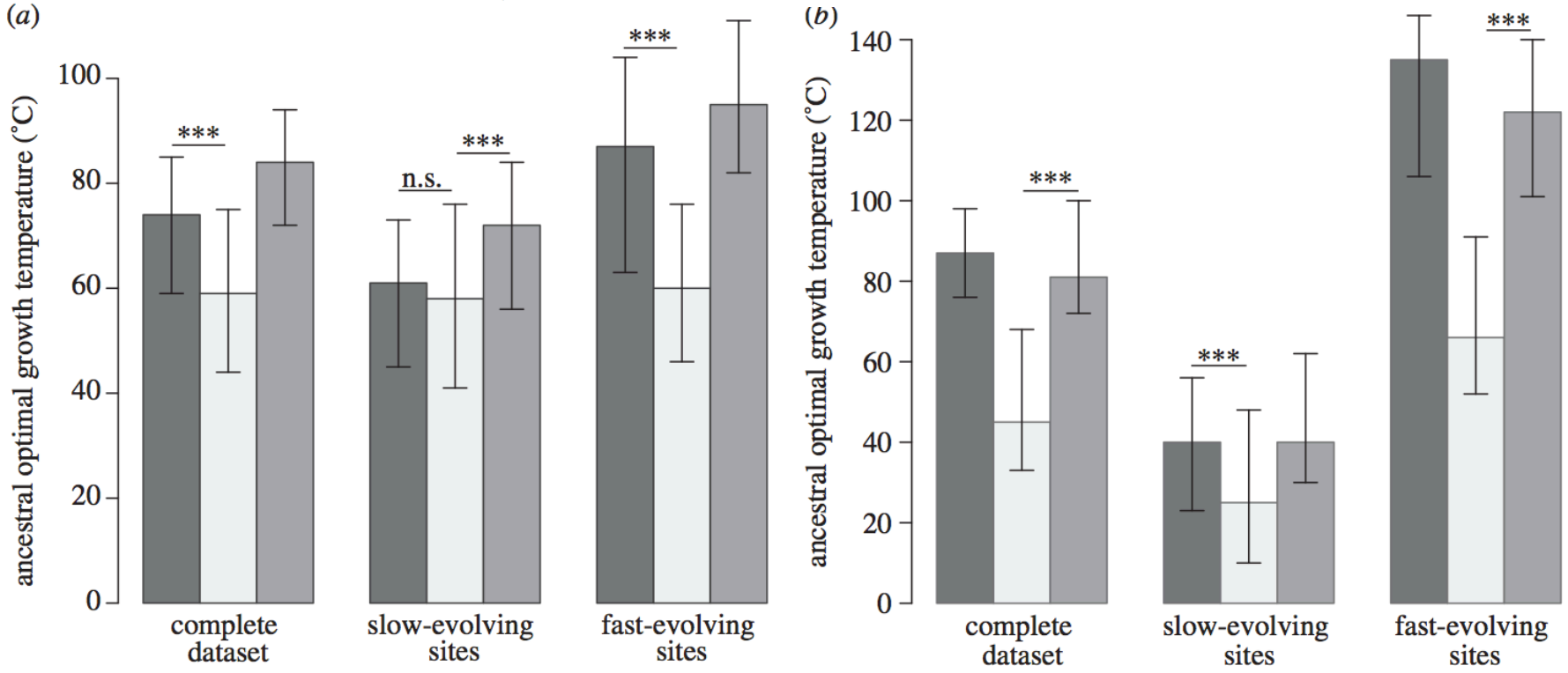
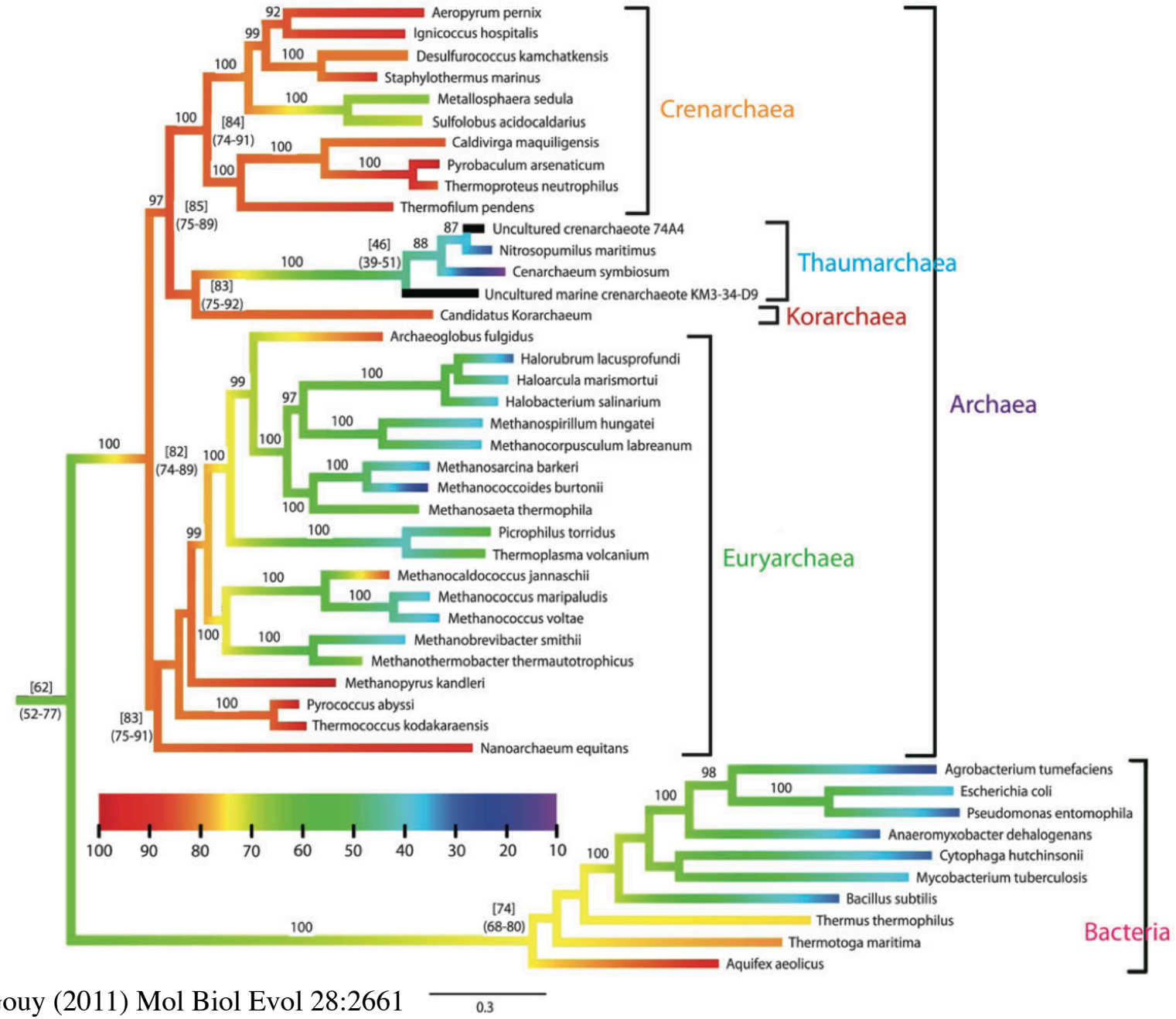


Figure 2. The non-homogeneous models recover the signal for a parallel adaptation to high temperatures within the across-site rate variation. (*a*) rRNA dataset. (*b*) Protein dataset. Ancestral temperatures for domain ancestors and for LUCA were estimated from ancestral compositions inferred with non-homogeneous models, either on all sites of the datasets (complete dataset) or on slow-evolving or fast-evolving sites only. ***p-value < 0.001. n.s. non-significant. Black bars, Bacteria; light grey bars, LUCA; dark grey bars, Archaea.

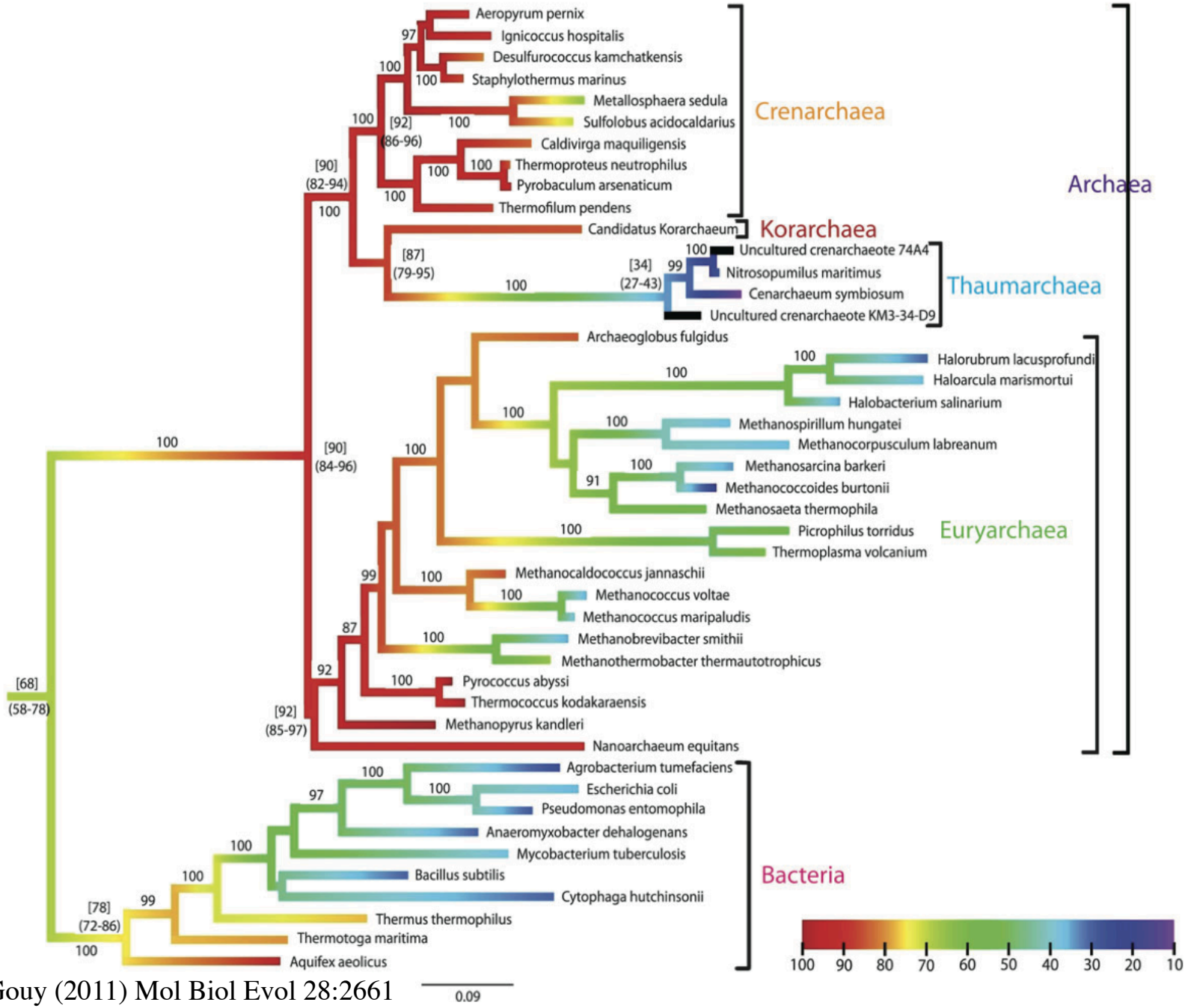
Biol Lett 9: 20130608.

Evolution of OGT from a hyperthermophilic ancestral state over the protein archaeal tree.



Groussin & Gouy (2011) Mol Biol Evol 28:2661

Evolution of OGT from a hyperthermophilic ancestral state over the rRNA archaeal tree.



Groussin & Gouy (2011) Mol Biol Evol 28:2661