Parallel Adaptations to High Temperatures in the Archean Eon

Samuel Blanquart[a1] Bastien Boussau[b1] Anamaria Necşulea[b] Nicolas Lartillot[a] Manolo Gouy[b]

June 9, 2008

[a] LIRMM, CNRS.
[b] BBE, CNRS, Université de Lyon I.

---

[1] These authors contributed equally to this work.

The Universal Tree of Life

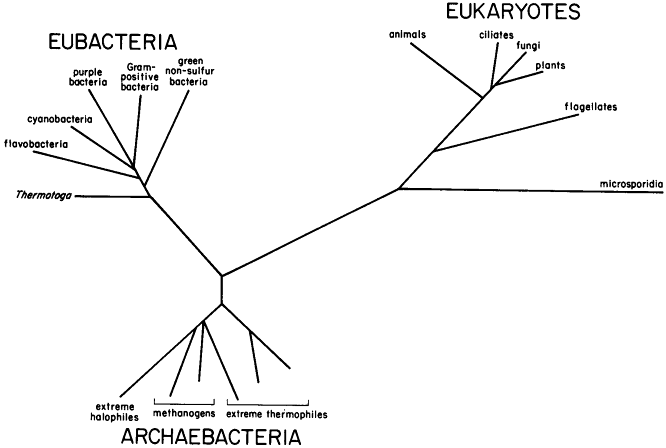


Figure: Universal phylogenetic tree determined from rRNA sequence comparisons [Woese, 1987]. In Procaryotic kingdoms, hyperthermophilic species are the first to diverge [Gribaldo and Brochier-Armanet, 2006], [Gaucher et al., 2003].

Hyperthermophilic Ancestors of Bacteria

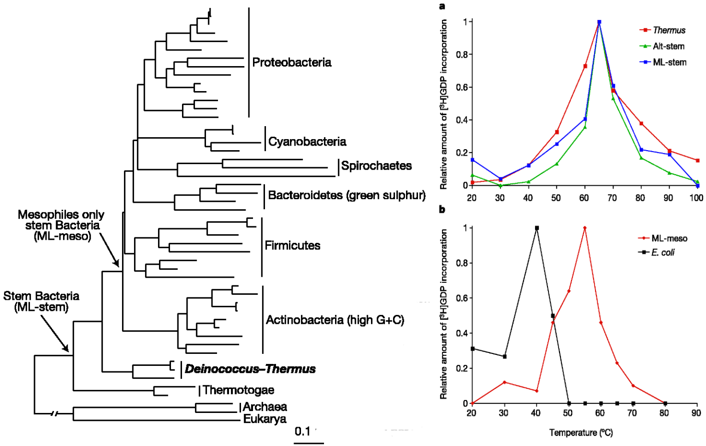


Figure: In silico inference of the EFtu of the bacterial ancestor, in vitro thermostability analysis [Gaucher et al., 2003].

Prokaryotic Ancestors Were Thermophilic: Most Likely also Was LUCA

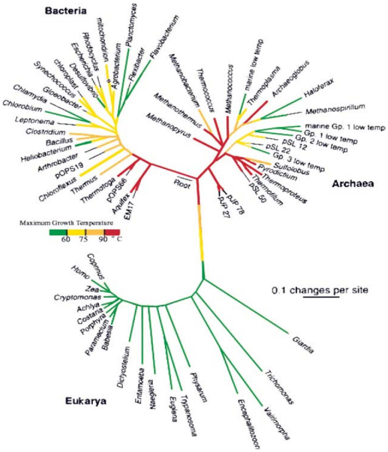


Figure: An intuitive algorithm for inferring the evolution of cellular growth temperatures [Lineweaver and Schwartzman, 2004].

Molecular Thermometers

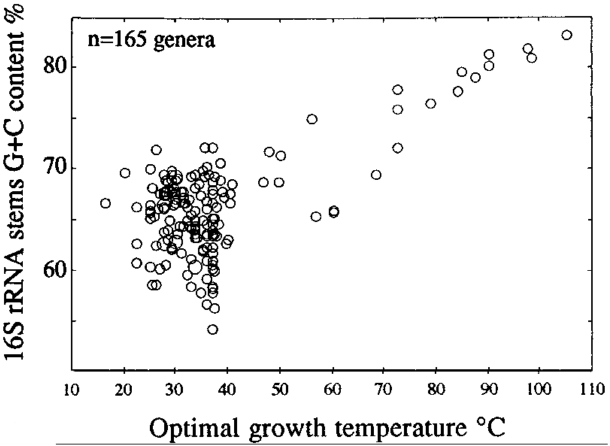


Figure: Procaryotic G+C contents in rRNA stems are correlated to the species optimal growth temperatures (OGT) [Galtier and Lobry, 1997].

Molecular Thermometers

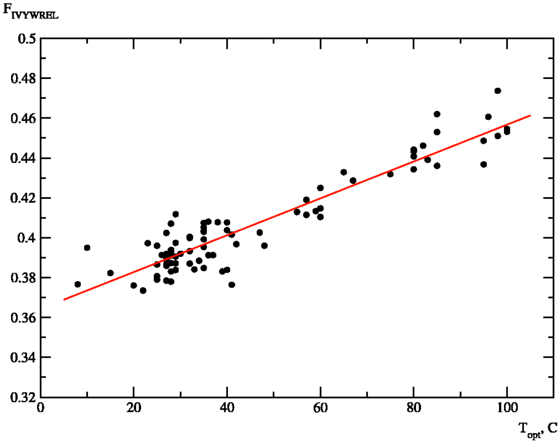


Figure: Procaryotic protein content in amino acids IVYWREL are correlated to the species optimal growth temperatures (OGT) [Zeldovich et al., 2007].

Inferring Ancestral OGTs using Molecular Thermometers

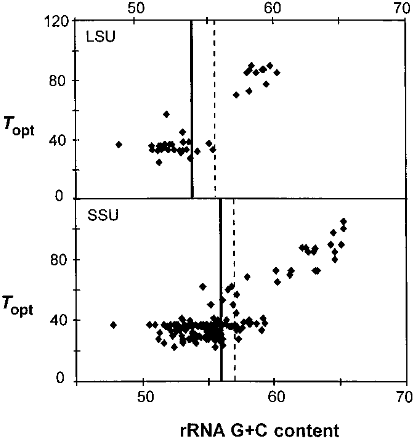


Figure: Inferred G+C contents of LUCA's rRNA SSU and LSU are incompatible with a thermophilic lifestyle [Galtier et al., 1999].

Similar Subsequent Estimates

- Amino acid: LUCA was a hyperthermophilic organism [DiGiulio, 2003], [Brooks et al., 2004].

Similar Subsequent Estimates

- Amino acid: LUCA was a hyperthermophilic organism [DiGiulio, 2003], [Brooks et al., 2004].
- rRNA: LUCA was a non hyperthermophilic organism [Boussau and Gouy, 2006], [Gowri-Shankar and Rattray, 2007].

Similar Subsequent Estimates

- Amino acid: LUCA was a hyperthermophilic organism [DiGiulio, 2003], [Brooks et al., 2004].
- rRNA: LUCA was a non hyperthermophilic organism [Boussau and Gouy, 2006], [Gowri-Shankar and Rattray, 2007].

Do these estimations depend on analysed molecules ?

Similar Subsequent Estimates

- Amino acid: LUCA was a hyperthermophilic organism [DiGiulio, 2003], [Brooks et al., 2004].
- rRNA: LUCA was a non hyperthermophilic organism [Boussau and Gouy, 2006], [Gowri-Shankar and Rattray, 2007].

Do these estimations depend on analysed molecules ?
What probabilistic assumption were made ?

Similar Subsequent Estimates

- Amino acid: LUCA was a hyperthermophilic organism [DiGiulio, 2003], [Brooks et al., 2004].
- rRNA: LUCA was a non hyperthermophilic organism [Boussau and Gouy, 2006], [Gowri-Shankar and Rattray, 2007].

Do these estimations depend on analysed molecules ?
What probabilistic assumption were made ?

- Homogeneous model of sequence evolution [DiGiulio, 2003], [Brooks et al., 2004].

Similar Subsequent Estimates

- Amino acid: LUCA was a hyperthermophilic organism [DiGiulio, 2003], [Brooks et al., 2004].
- rRNA: LUCA was a non hyperthermophilic organism [Boussau and Gouy, 2006], [Gowri-Shankar and Rattray, 2007].

Do these estimations depend on analysed molecules ?
What probabilistic assumption were made ?

- Homogeneous model of sequence evolution [DiGiulio, 2003], [Brooks et al., 2004].
- Non-homogeneous model of sequence evolution [Galtier and Gouy, 1998], [Boussau and Gouy, 2006], [Gowri-Shankar and Rattray, 2007].

In this Work

For the first time in the debate about the early OGT evolution:

In this Work

For the first time in the debate about the early OGT evolution:

- We are able to draw conclusions from non-homogeneous analysis of amino acid sequences.

In this Work

For the first time in the debate about the early OGT evolution:

- We are able to draw conclusions from non-homogeneous analysis of amino acid sequences.
- We infer OGTs for all nodes of the universal cellular tree of life, using both RNA and amino acid sequences.

In this Work

For the first time in the debate about the early OGT evolution:

- We are able to draw conclusions from non-homogeneous analysis of amino acid sequences.
- We infer OGTs for all nodes of the universal cellular tree of life, using both RNA and amino acid sequences.
- Our data:
  - Concatenation of 16S and 23S rRNA sequences for 456 species, 1043 stem positions,
  - Concatenation of 56 proteins for 115 species, 3336 nearly ungaped positions.

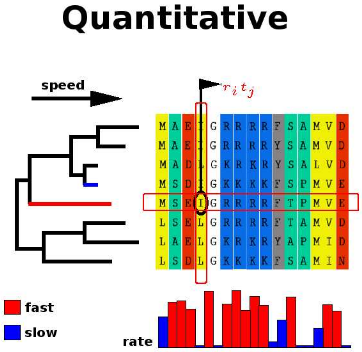
In this Work

For the first time in the debate about the early OGT evolution:

- We are able to draw conclusions from non-homogeneous analysis of amino acid sequences.
- We infer OGTs for all nodes of the universal cellular tree of life, using both RNA and amino acid sequences.

- Our data:
  - Concatenation of 16S and 23S rRNA sequences for 456 species, 1043 stem positions,
  - Concatenation of 56 proteins for 115 species, 3336 nearly ungaped positions.
- Our models:
  - A non-homogeneous ML model defining branchwise G+C frequencies [Boussau and Gouy, 2006],
  - A site- and time heterogeneous Bayesian model of amino acid replacement [Blanquart and Lartillot, 2008].

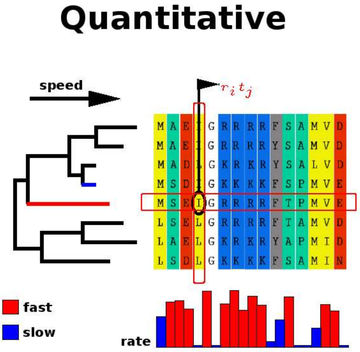
The Amino Acid Replacement Model



```
M A E I G R R R R F S A M V D
M A E I G R R R R Y S A M V D
M A D L G K R K R Y S A L V D
M S D I G K K K K F S P M V E
M S E I G R R R R F T P M V E
L S E L G R R R R F T A M V D
L A E L G K R K R Y A P M I D
L S D L G K K K K F S A M I N
```

- The standard GTR+Γ model is quantitatively (substitution rates) site- (model RAS) and time- (branch length) heterogeneous.

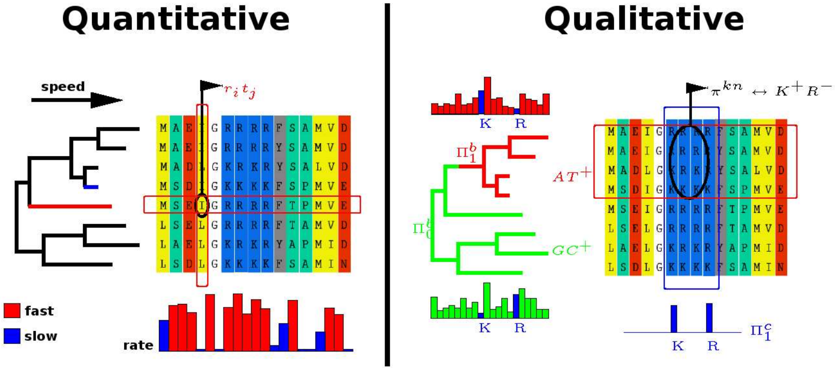
The Amino Acid Replacement Model



M A E I G R R R R F S A M V D
M A E I G R R R R Y S A M V D
M A D L G K R K R Y S A L V D
M S D I G K K K K F S P M V E
M S E I G R R R R F T P M V E
L S E L G R R R R F T A M V D
L A E L G K R K R Y A P M I D
L S D L G K K K K F S A M I N

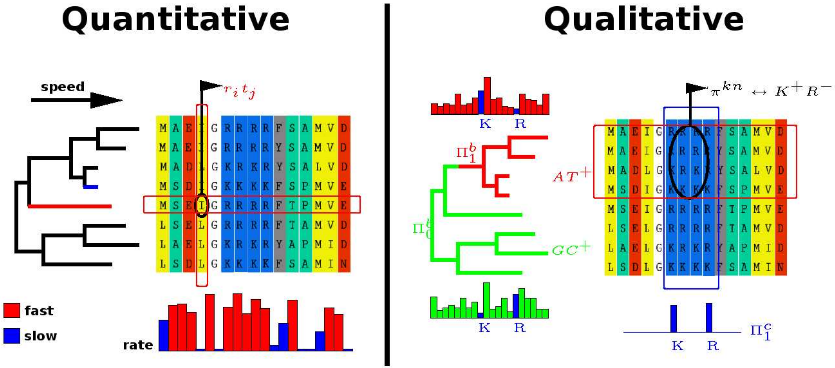
- The standard GTR+Γ model is quantitatively (substitution rates) site- (model RAS) and time- (branch length) heterogeneous.
- It is “single matrix”, and thus qualitatively (replacement probabilities) site- and time- homogeneous.

The Amino Acid Replacement Model



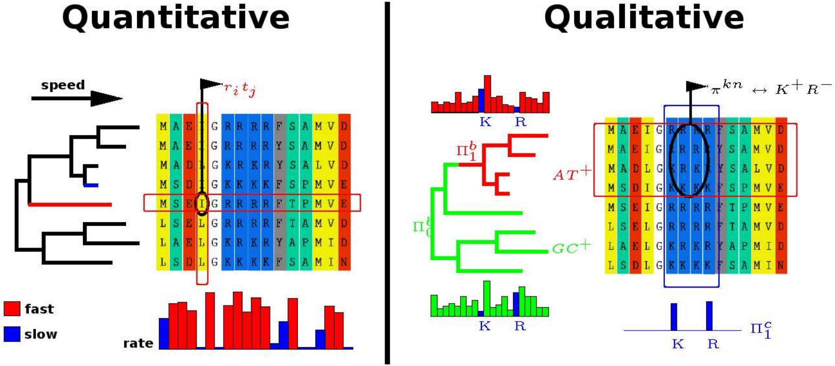
- CAT is qualitatively site- heterogeneous [Lartillot and Philippe, 2004],

The Amino Acid Replacement Model



- CAT is qualitatively site- heterogeneous [Lartillot and Philippe, 2004],
- BP is time- heterogeneous [Blanquart and Lartillot, 2006],

The Amino Acid Replacement Model



- CAT is qualitatively site- heterogeneous [Lartillot and Philippe, 2004],
- BP is time- heterogeneous [Blanquart and Lartillot, 2006],
- CAT+BP is site- and time- heterogeneous [Blanquart and Lartillot, 2008].

Results under the Bacterial Rooting Hypothesis

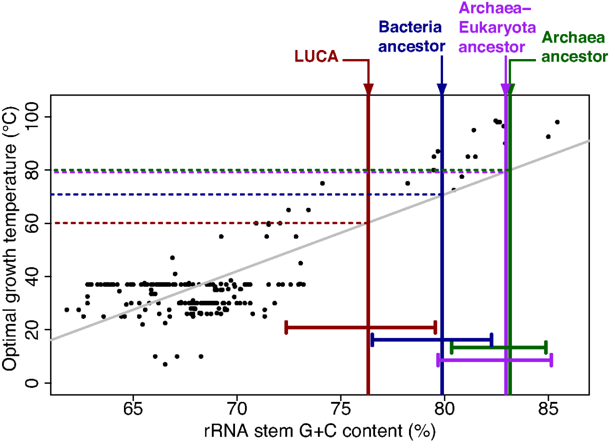


Figure: With rRNA sequences, LUCA is estimated as non thermophilic, in agreement with [Galtier et al., 1999]. Bacterial and Archeal ancestors are inferred as hyperthermophilic, in agreement with [Gribaldo and Brochier-Armanet, 2006] and [Gaucher et al., 2003].

Results under the Bacterial Rooting Hypothesis

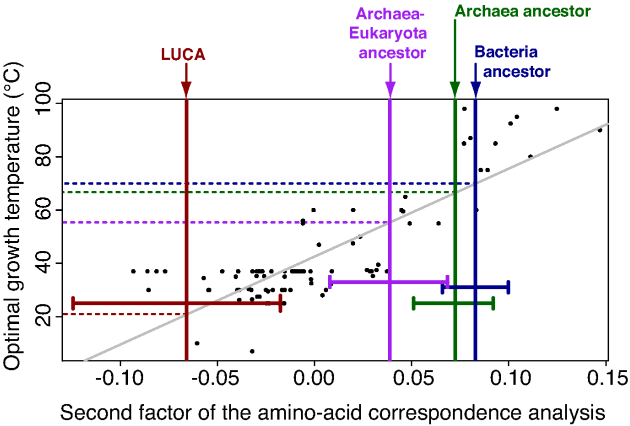


Figure: With protein sequences, LUCA is estimated as non thermophilic, in agreement with [Galtier et al., 1999]. Bacterial and Archeal ancestors are inferred as hyperthermophilic, in agreement with [Gribaldo and Brochier-Armanet, 2006] and [Gaucher et al., 2003].

Results under the Bacterial Rooting Hypothesis

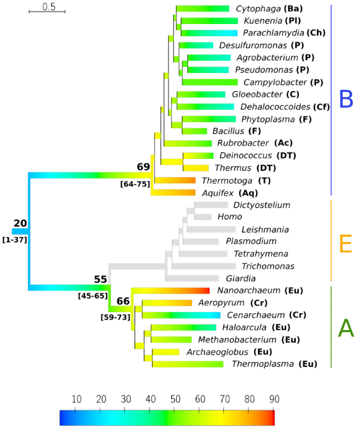


Figure: Convergence to thermophilic way of life from a mesophilic LUCA inferred from amino acid sequences under non homogeneous conditions.

Dependency to the Homogeneity Assumption

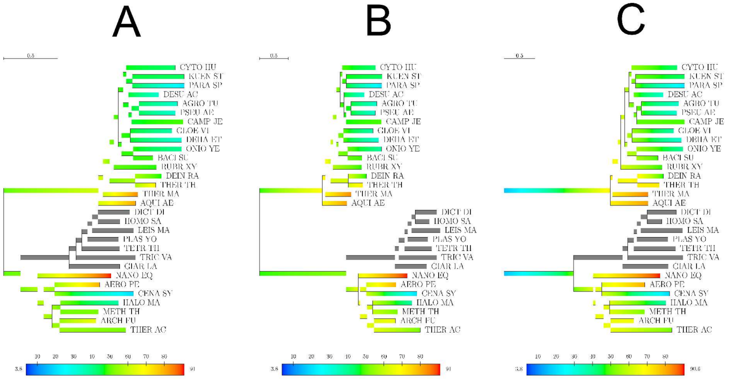


Figure: Dependency of the OGTs to the phylogenetic model. A: GTR, B: CAT [Lartillot and Philippe, 2004], C: CAT+BP [Blanquart and Lartillot, 2008] (A and B, time homogeneous, C, time heterogeneous).

Dependency to the Homogeneity Assumption

	Time Homogeneous			Time Heterogeneous		
	Model	$p_{L>B}$	$p_{L>AE}$	Model	$p_{L>B}$	$p_{L>AE}$
rRNA	GTR	0.24	0.11	GG	0.025	0.000
	Brooks	0.9	1	YR	0.18	0.01
				BP	0.027	0.000
	Model	$p_{L>B}$	$p_{L>AE}$	Model	$p_{L>B}$	$p_{L>AE}$
Protein	GTR	0.022	0.344	CAT+BP	0.000	0.000
	Brooks	0.933	0.983	CAT+YR	0.000	0.000
	CAT	0.008	0.166			

Table: Results significativity. $p_{L>*}$ Pvalue for LUCA growth temperature to be greater than that of its direct descendant, “B” Bacteria ancestor, “AE” Archea Eukaryota ancestor. Models: Brooks [Brooks et al., 2004], GG [Boussau and Gouy, 2006], YR [Yang and Roberts, 1995], CAT [Lartillot and Philippe, 2004], BP [Blanquart and Lartillot, 2006], CAT+BP [Blanquart and Lartillot, 2006].

Dependency to Taxon Sampling

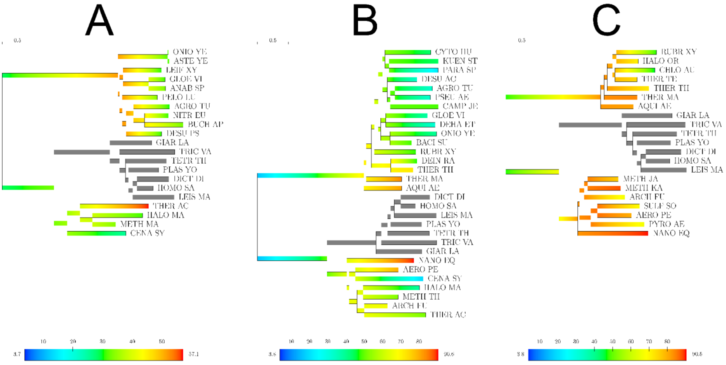
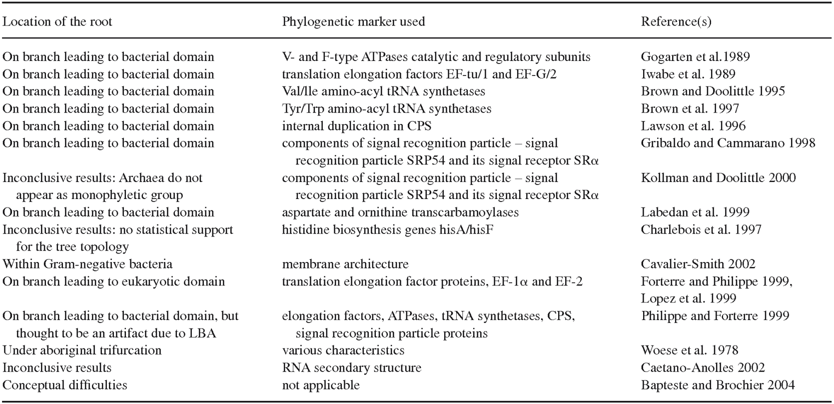


Figure: Amino acid dataset under non homogeneous conditions [Blanquart and Lartillot, 2008], A: mesophilic amino acid dataset, B: complete dataset, C: thermophilic dataset.

The Rooting of the Tree of Life



Location of the root	Phylogenetic marker used	Reference(s)
On branch leading to bacterial domain	V- and F-type ATPases catalytic and regulatory subunits	Gogarten et al.1989
On branch leading to bacterial domain	translation elongation factors EF-tu/1 and EF-G/2	Iwabe et al. 1989
On branch leading to bacterial domain	Val/lle amino-acyl tRNA synthetases	Brown and Doolittle 1995
On branch leading to bacterial domain	Tyr/Trp amino-acyl tRNA synthetases	Brown et al. 1997
On branch leading to bacterial domain	internal duplication in CPS	Lawson et al. 1996
On branch leading to bacterial domain	components of signal recognition particle – signal recognition particle SRP54 and its signal receptor SRα	Gribaldo and Cammarano 1998
Inconclusive results: Archaea do not appear as monophyletic group	components of signal recognition particle – signal recognition particle SRP54 and its signal receptor SRα	Kollman and Doolittle 2000
On branch leading to bacterial domain	aspartate and ornithine transcarbamoylases	Labedan et al. 1999
Inconclusive results: no statistical support for the tree topology	histidine biosynthesis genes hisA/hisF	Charlebois et al. 1997
Within Gram-negative bacteria	membrane architecture	Cavalier-Smith 2002
On branch leading to eukaryotic domain	translation elongation factor proteins, EF-1α and EF-2	Forterre and Philippe 1999, Lopez et al. 1999
On branch leading to bacterial domain, but thought to be an artifact due to LBA	elongation factors, ATPases, tRNA synthetases, CPS, signal recognition particle proteins	Philippe and Forterre 1999
Under aboriginal trifurcation	various characteristics	Woese et al. 1978
Inconclusive results	RNA secondary structure	Caetano-Anolles 2002
Conceptual difficulties	not applicable	Bapteste and Brochier 2004

Figure: Different points of view on the location of the root of the tree of life [Zhaxybayeva et al., 2005].

The Rooting of the Tree of Life

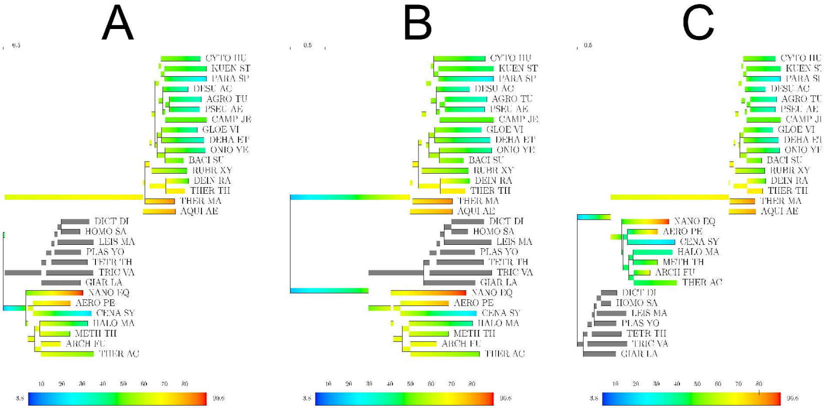


Figure: Non homogeneous amino analysis according to rooting, A: Archea branch, B: Bacteria branch, C: Eukaryota branch.

Setting of the Early Genetic Code

- Our estimation of a mesophilic LUCA and of a subsequent parallel adaptation to thermophily results from a protein content initially depleted in IVYWREL.
- [Fournier and Gogarten, 2007] have also recently proposed that LUCA was depleted in IVYEW, which might be the trace of the early genetic code structure.
- However, our interpretation in terms of adaptation to thermophily has the advantage to explain both rRNA and amino acid patterns.

Archean global temperatures

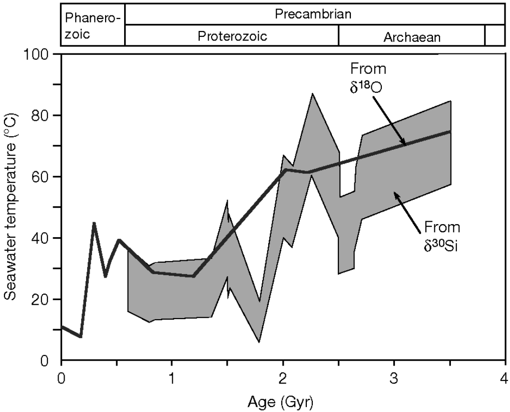


Figure: Global decreasing of ocean temperature over the last 3.5 billion of years [Robert and Chaussidon, 2006].

Bacteria Adaptation to Archean temperatures

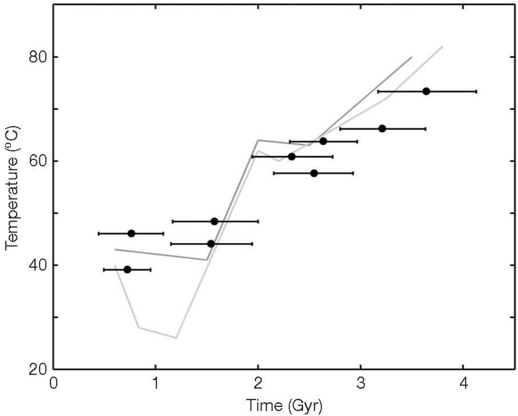


Figure: Melting temperatures of resurected bacterial EFtu over the last 3.5 billion of years [Gaucher et al., 2008]. Some models of Hadean ($< 3.5Gyr$) temperature indicate a possible frozen ocean [Nisbet and Sleep, 2001], [Kasting and Ono, 2006].

The Last Heavy Bombardment

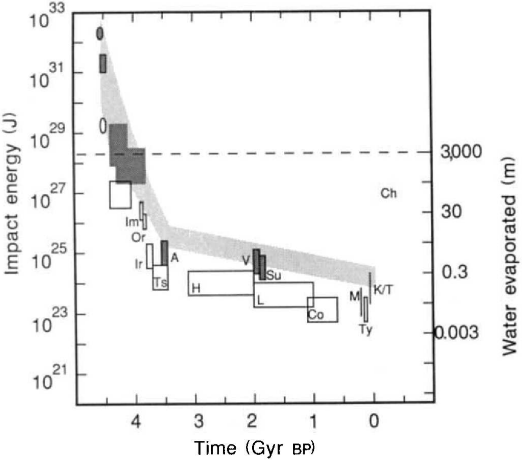


Figure: Energy of meteoritic impact during the earth history [Sleep et al., 1989].

The Forterre Hypothesis

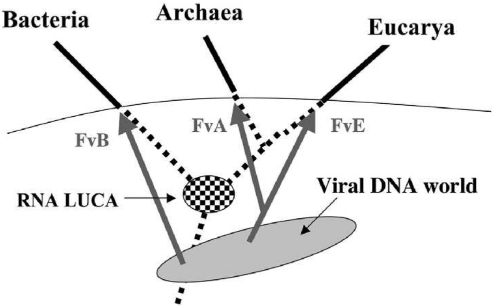


Figure: Horizontal transfer of DNA management ability from several viruses lineages to cellular organisms with RNA genomes [Forterre, 2006]. DNA genomes are more thermostable than RNA ones [Islas et al., 2003].

Acknowledgment

Contributors:

Samuel Blanquart,
Bastien Boussau,
Anamaria Necşulea,
Nicolas Lartillot, and
Manolo Gouy,

Which to thank:

Céline Brochier-Armanet,
Nicolas Galtier,
Marc Chaussidon, and
David Bryant,

For their helpful comments on this work.

Bibliography

- Blanquart, S. and Lartillot, N. (2006).
  A Bayesian Compound Stochastic Process for Modeling Nonstationary and Nonhomogeneous Sequence Evolution.
  Molecular Biology and Evolution, 23(11):2058–2071.

- Blanquart, S. and Lartillot, N. (2008).
  A site- and time-heterogeneous model of amino-acid replacement.
  Mol. Biol. Evol.
  in press.

- Boussau, B. and Gouy, M. (2006).
  Efficient likelihood computations with nonreversible models of evolution.
  Syst. Biol., 55(5):756–768.

- Brooks, D. J., Fresco, J. R., and Singh, M. (2004).
  A novel method for estimating ancestral amino acid composition and its application to proteins of the Last

Dependency to the Prior

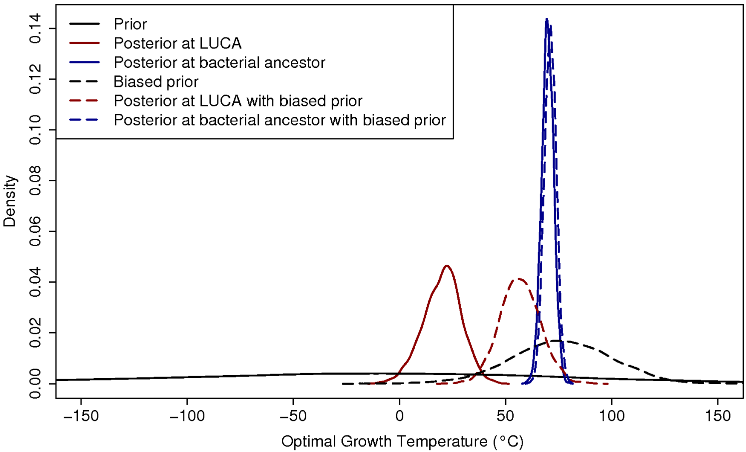


Figure: