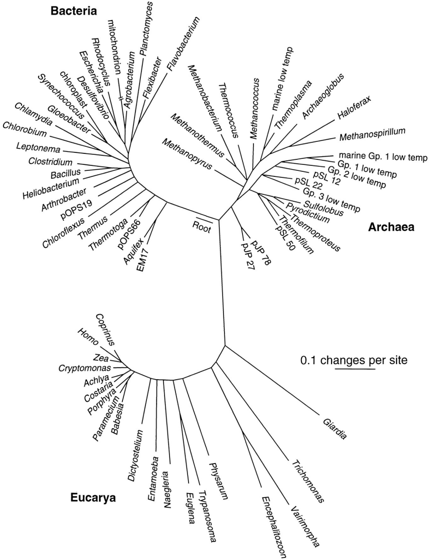
Parallel Adaptations to High Temperatures in the Archean Eon

Bastien Boussau[1*], Samuel Blanquart[2*], Anamaria Necșulea[1],
Nicolas Lartillot[2], Manolo Gouy[1]

[1]LBBE, CNRS, Lyon; [2]LIRMM, CNRS, Montpellier
[*]These authors contributed equally to this work.

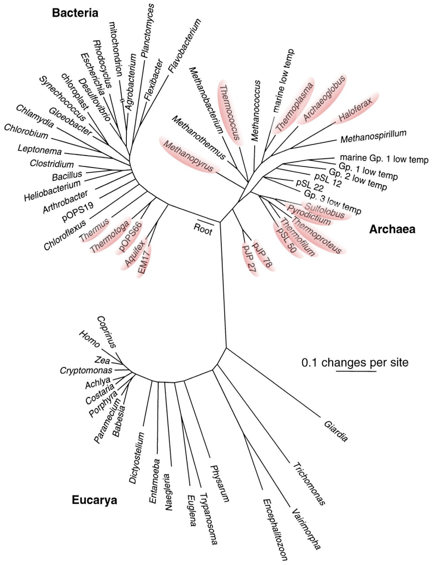
Saturday, June 25th 2008

Thermophiles in the tree of life



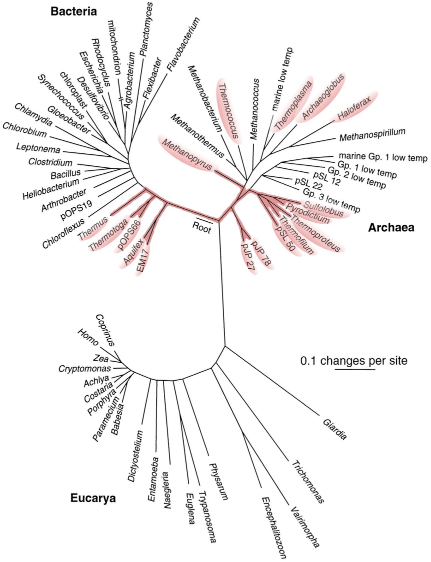
Pace, *Science* 1997.

Thermophiles in the tree of life



Pace, *Science* 1997.

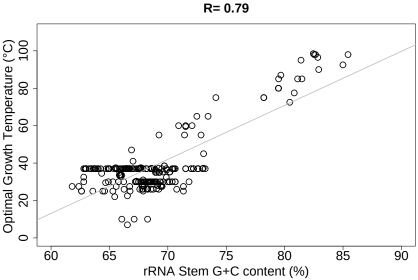
Thermophiles in the tree of life



Pace, *Science* 1997.

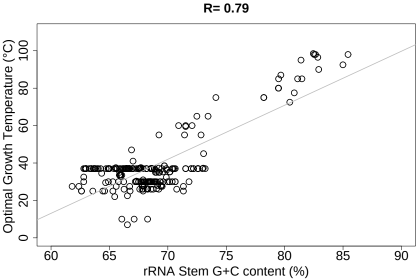
Sequence composition and temperature

Ribosomal RNA

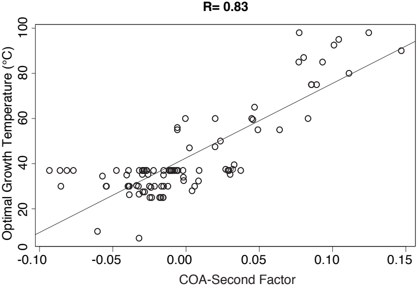


Sequence composition and temperature

Ribosomal RNA

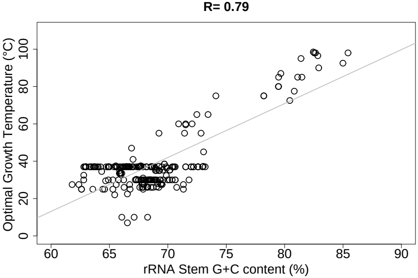


Proteins

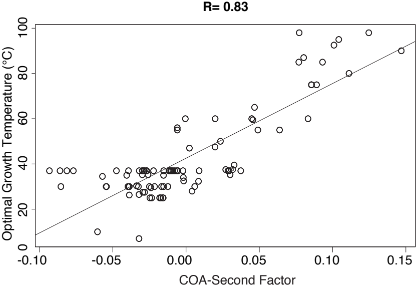


Sequence composition and temperature

Ribosomal RNA



Proteins



rRNA and protein compositions can be used as independent molecular thermometers

Ancestral sequence reconstructions

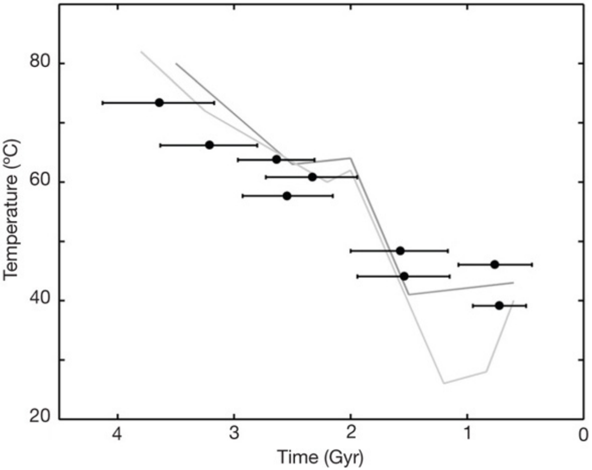
Ribosomal RNA

Galtier *et al.*, 1999:
non-hyperthermophilic LUCA

Proteins

Di Giulio 2003, Brooks *et al.* 2004:
hyperthermophilic LUCA
Gaucher *et al.* 2003, 2008:
hyperthermophilic bacterial ancestor

Gaucher *et al.* 2008



Gaucher *et al.* observe a decrease in optimal growth temperatures from the ancestor of Bacteria to extant organisms.

Ancestral sequence reconstructions

Ribosomal RNA

Galtier *et al.*, 1999:
non-hyperthermophilic LUCA

Proteins

Di Giulio 2003, Brooks *et al.* 2004:
hyperthermophilic LUCA
Gaucher *et al.* 2003, 2008:
hyperthermophilic bacterial ancestor

Ancestral sequence reconstructions

Ribosomal RNA

Galtier *et al.*, 1999:
non-hyperthermophilic LUCA

Proteins

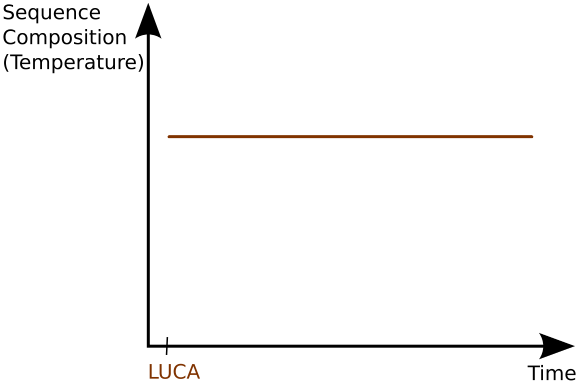
Di Giulio 2003, Brooks *et al.* 2004:
hyperthermophilic LUCA
Gaucher *et al.* 2003, 2008:
hyperthermophilic bacterial ancestor

Differences

Non-homogeneous model of evolution

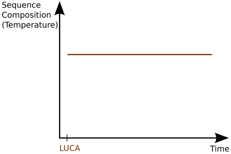
Homogeneous or non-stationary models of evolution

Models of sequence evolution

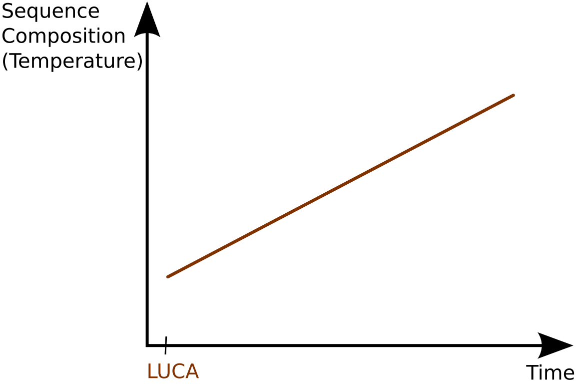


Homogeneous model

Models of sequence evolution

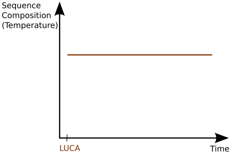


Homogeneous

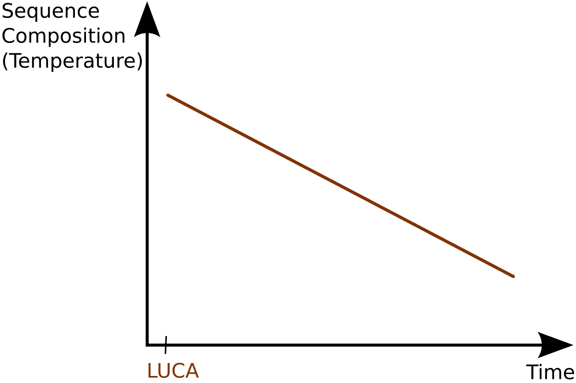


Non-stationary model

Models of sequence evolution

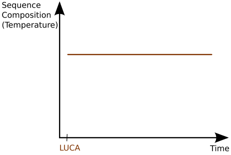


Homogeneous

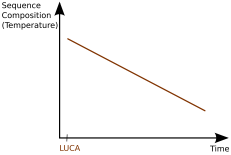


Non-stationary model

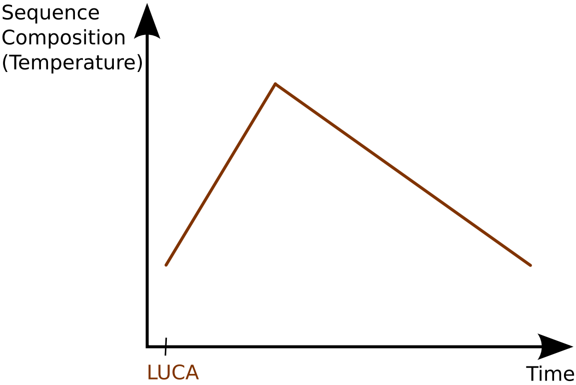
Models of sequence evolution



Homogeneous



Non-stationary

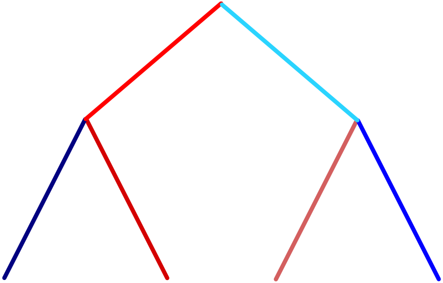


Non-homogeneous model

Models and data used

Ribosomal RNA

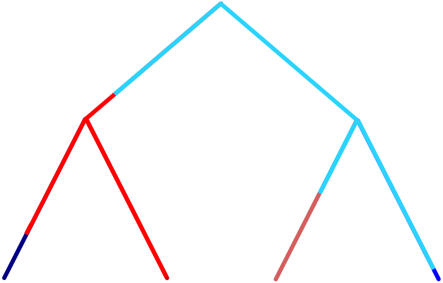
Galtier and Gouy 1998 model



- SSU + LSU genes, stems only
- 456 species
- 1043 sites

Proteins

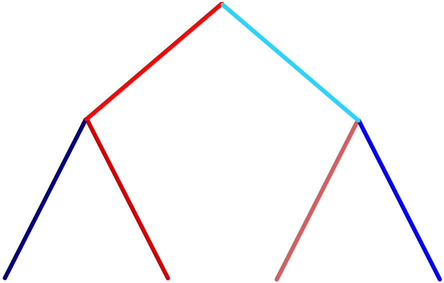
Blanquart and Lartillot 2008 model



- 56 concatenated genes
- 30 species
- 3336 ungapped sites

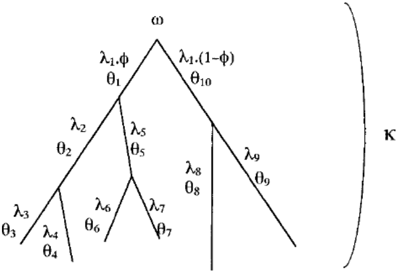
rRNA evolution model: Galtier and Gouy model (1998)

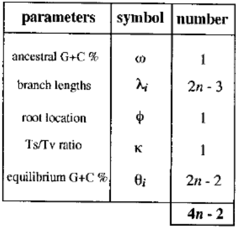
Model



- 1 model per branch
- each model is characterized by an equilibrium G+C content

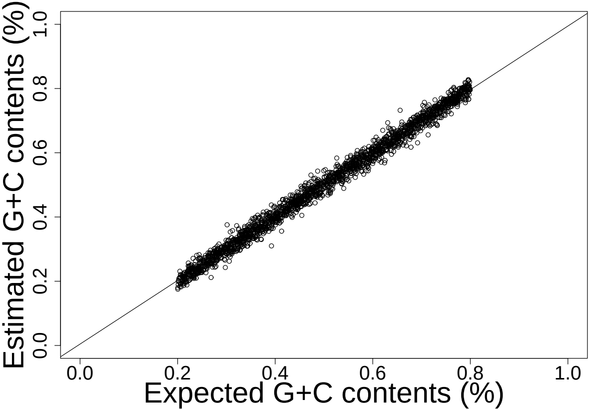
Parameters





parameters	symbol	number
ancestral G+C %	ω	1
branch lengths	λ_i	$2n$ - 3
root location	ϕ	1
Ts/Tv ratio	κ	1
equilibrium G+C %	θ_i	$2n$ - 2
		$4n$ - 2

Ability of the GG98 model to estimate the root G+C



Boussau and Gouy, *Syst. Biol.* 2006.

Protein evolution model: CAT-BP

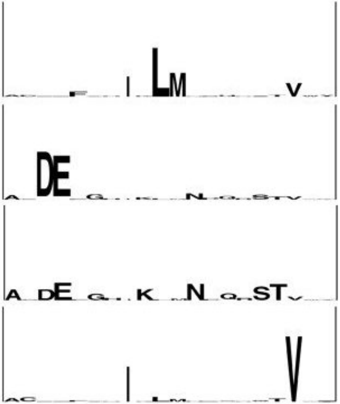
A protein alignment

```
FTALTVVGDGNG
FSALVVIGDKNG
FSALVVVGDKNG
FTALVIVGNRKG
FRAIVVIGNERG
FRAIVVVGNEKG
YKVLVIMGNMDG
FKAFVAIGDYNG
```

Not all amino-acids are allowed at a given position in a protein

Profiles of amino-acids

profile of amino-acids = vector of amino-acid equilibrium frequencies.



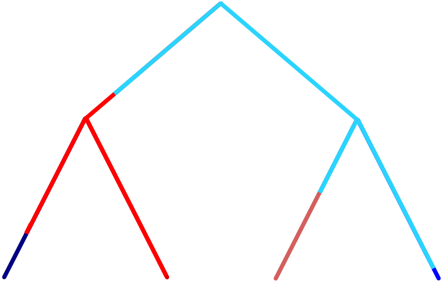
Protein evolution model: CAT-BP

CAT model

FTALTVVGDGNG
FSALVVIGDKNG
FSALVVVGDKNG
FTALVIVGNRKG
FRAIVVIGNERG
FRAIVVVGNEKG
YKVLVIMGNMDG
FKAFVAIGDYNG

- Gamma model: mixture model over rates of evolution
- CAT model: mixture model over profiles of amino-acids

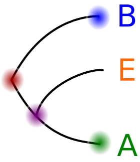
BP Model



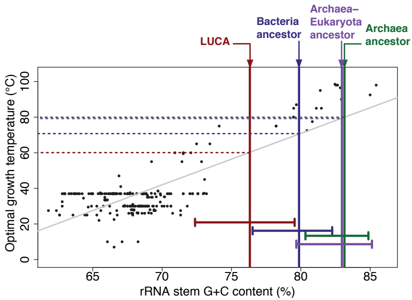
- Breakpoints between profiles are placed along the phylogeny according to a Poisson process.

Blanquart and Lartillot, *Mol. Biol. Evol.*, 2008.

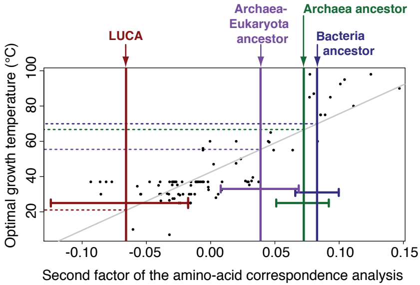
Results



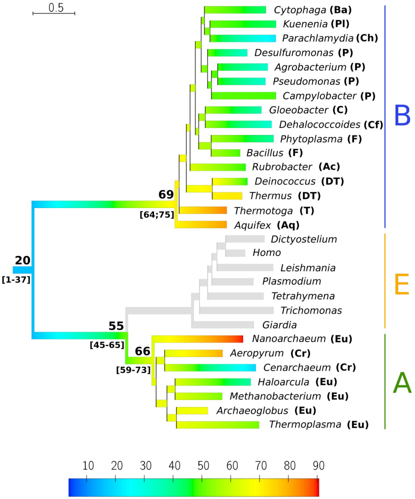
Ribosomal RNA



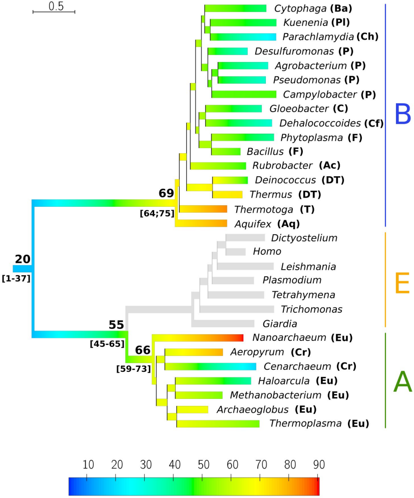
Proteins

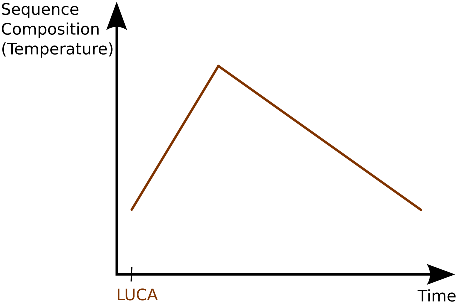


The evolution of Thermophily

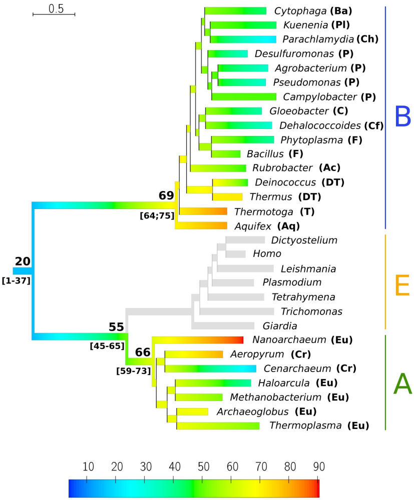


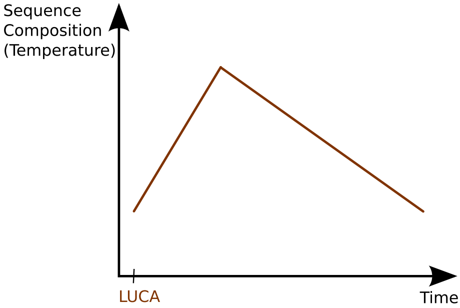
The evolution of Thermophily

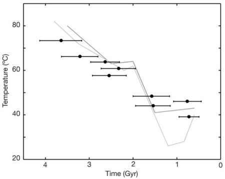




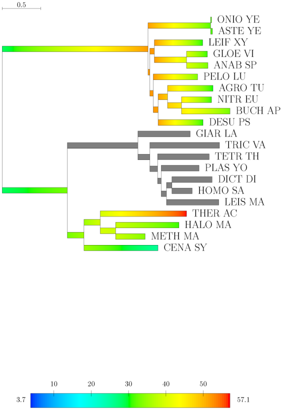
The evolution of Thermophily



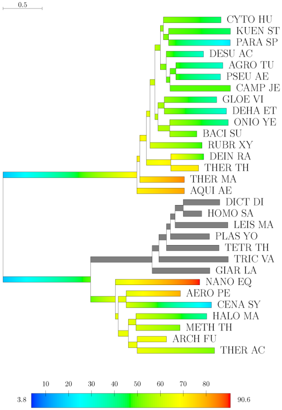




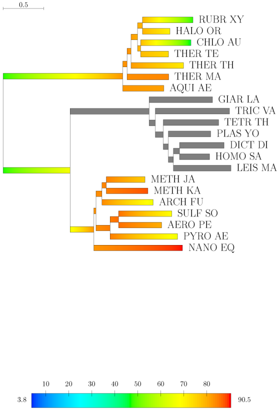
The evidence for parallel adaptations



Mesophilic

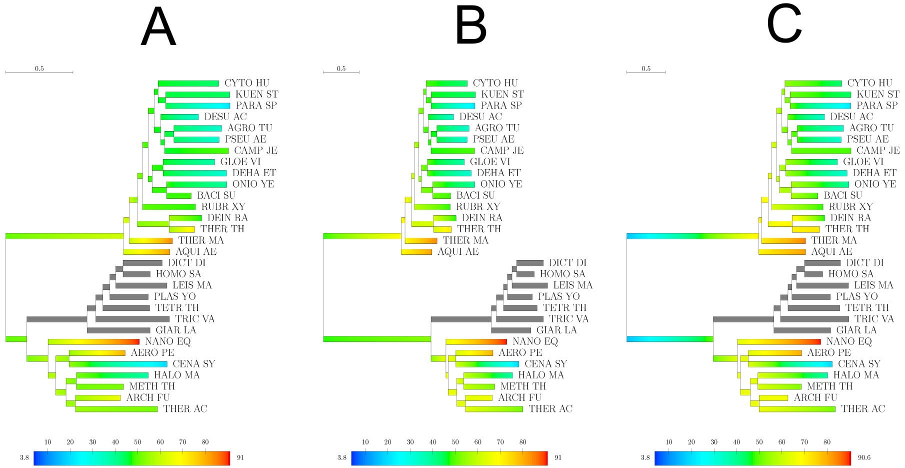


Normal



Thermophilic

The evidence for parallel adaptations: models



GTR CAT CAT-BP

The evidence for parallel adaptations

- obtained in the Maximum Likelihood or Bayesian framework

The evidence for parallel adaptations

- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins

The evidence for parallel adaptations

- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins
- obtained with several non-homogeneous models

The evidence for parallel adaptations

- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins
- obtained with several non-homogeneous models
- robust to changes in the topology

The evidence for parallel adaptations

- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins
- obtained with several non-homogeneous models
- robust to changes in the topology
- robust to changes in taxonomic sampling

The evidence for parallel adaptations

- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins
- obtained with several non-homogeneous models
- robust to changes in the topology
- robust to changes in taxonomic sampling
- robust to changes in prior distributions

The evidence for parallel adaptations

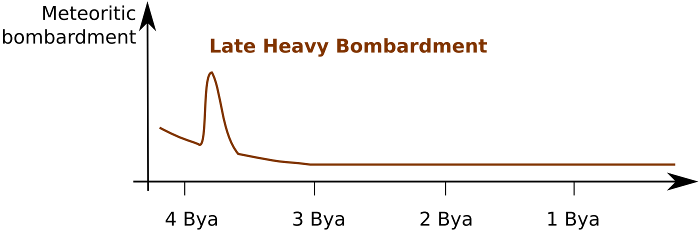
- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins
- obtained with several non-homogeneous models
- robust to changes in the topology
- robust to changes in taxonomic sampling
- robust to changes in prior distributions
- robust to the removal of fast-evolving rRNA sites

The evidence for parallel adaptations

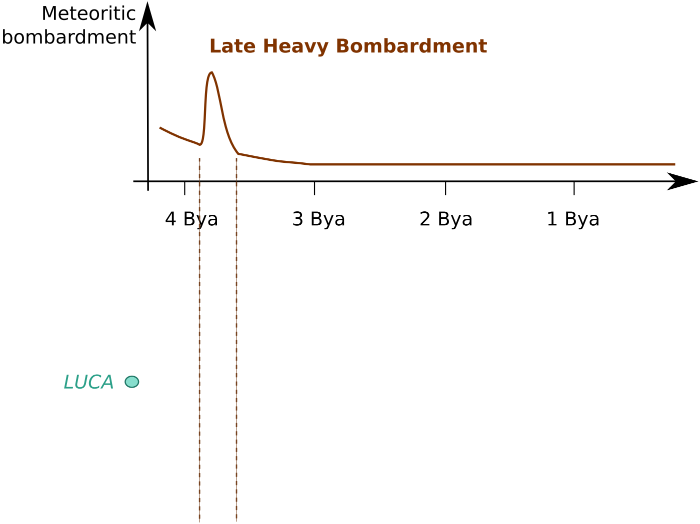
- obtained in the Maximum Likelihood or Bayesian framework
- obtained with rRNA and proteins
- obtained with several non-homogeneous models
- robust to changes in the topology
- robust to changes in taxonomic sampling
- robust to changes in prior distributions
- robust to the removal of fast-evolving rRNA sites
- robust to the removal of rare amino-acids

A pressure for parallel adaptations

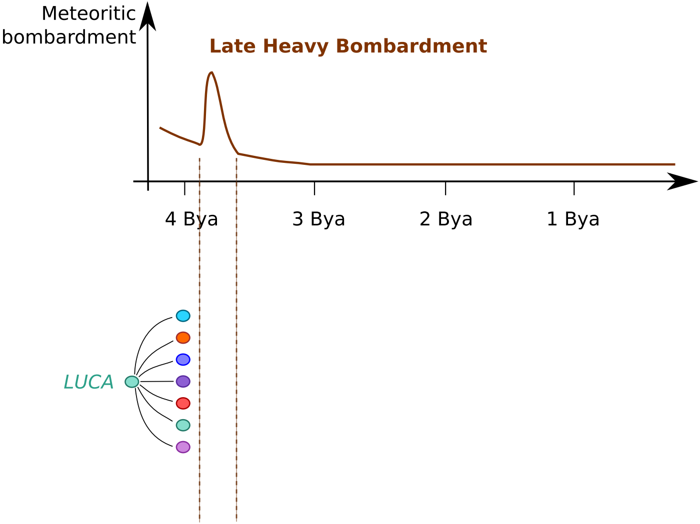
A pressure for parallel adaptations



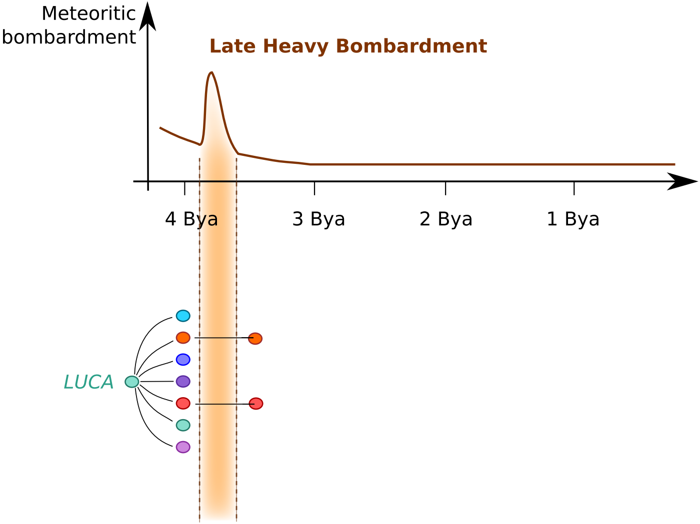
A pressure for parallel adaptations



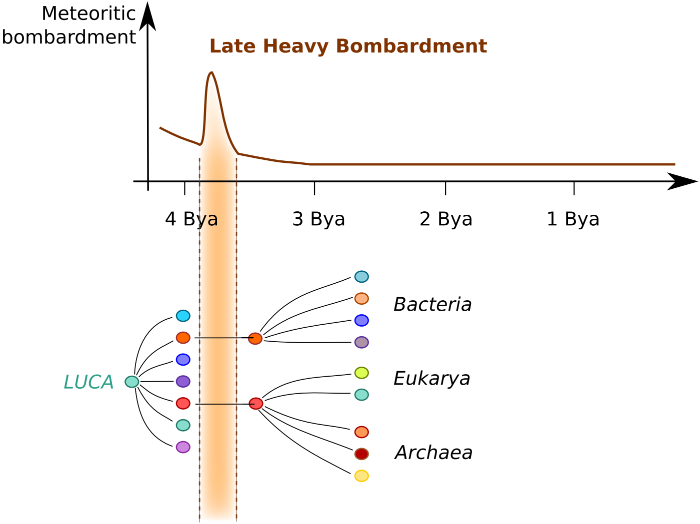
A means for parallel adaptations



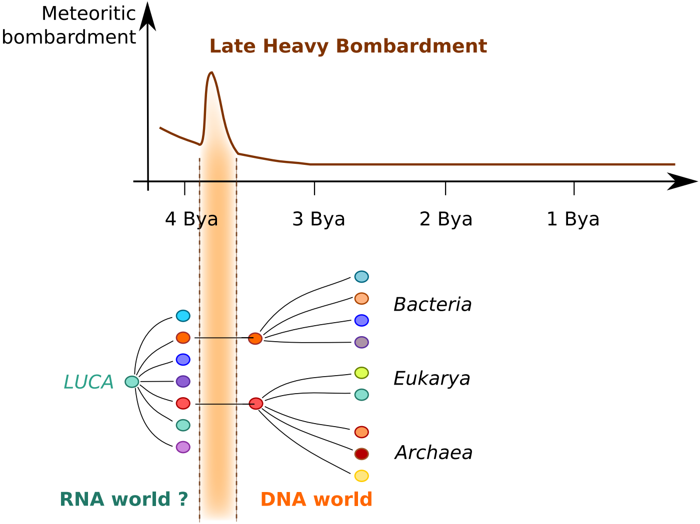
A means for parallel adaptations



A means for parallel adaptations



A means for parallel adaptations



Conclusions

- Only non-homogeneous models can faithfully reconstruct the evolution of sequence composition

Conclusions

- Only non-homogeneous models can faithfully reconstruct the evolution of sequence composition
- Non-homogeneous models find evidence for two phases in the history of thermophily:

Conclusions

- Only non-homogeneous models can faithfully reconstruct the evolution of sequence composition
- Non-homogeneous models find evidence for two phases in the history of thermophily:
    - first an increase from LUCA to its direct descendants

Conclusions

- Only non-homogeneous models can faithfully reconstruct the evolution of sequence composition
- Non-homogeneous models find evidence for two phases in the history of thermophily:
    - first an increase from LUCA to its direct descendants
    - then a decrease from the bacterial ancestor to extant species

Conclusions

- Only non-homogeneous models can faithfully reconstruct the evolution of sequence composition
- Non-homogeneous models find evidence for two phases in the history of thermophily:
  - first an increase from LUCA to its direct descendants
  - then a decrease from the bacterial ancestor to extant species
- Such a scenario is in agreement with several hypotheses concerning the early evolution of the Earth and of life

Acknowledgements

Authors

Bastien Boussau*, Samuel Blanquart*, Anamaria Necşulea,
Nicolas Lartillot, Manolo Gouy
* These authors contributed equally to this work.

Many thanks to...

Céline Brochier-Armanet, David Bryant, Marc Chaussidon, Nicolas Galtier, Antonio Lazcano, Mathilde Paris, and members of the BBE Laboratory.

The evidence for parallel adaptations: roots