Rocks, clocks and genes from other species

i.e., gene transfers can date the tree of life.

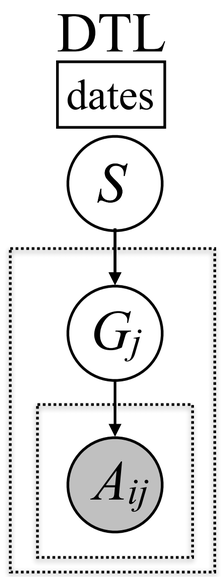
Gergely Szöllősi

MTA-ELTE "Lendület"
Evolutionary Genomics Research Group
Budapest, Hungary

ssolo@elte.hu
@sllsi

Adrian **Davin** Bastien **Boussau** Vincent **Daubin** Eric **Tannier** — LBBE, Lyon

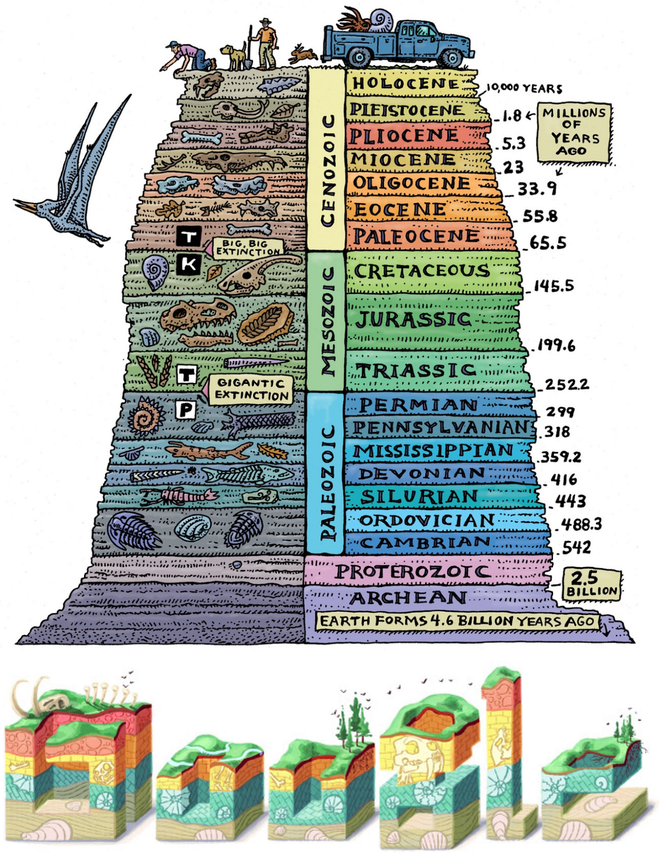
Tom **Williams** — U.Bristol

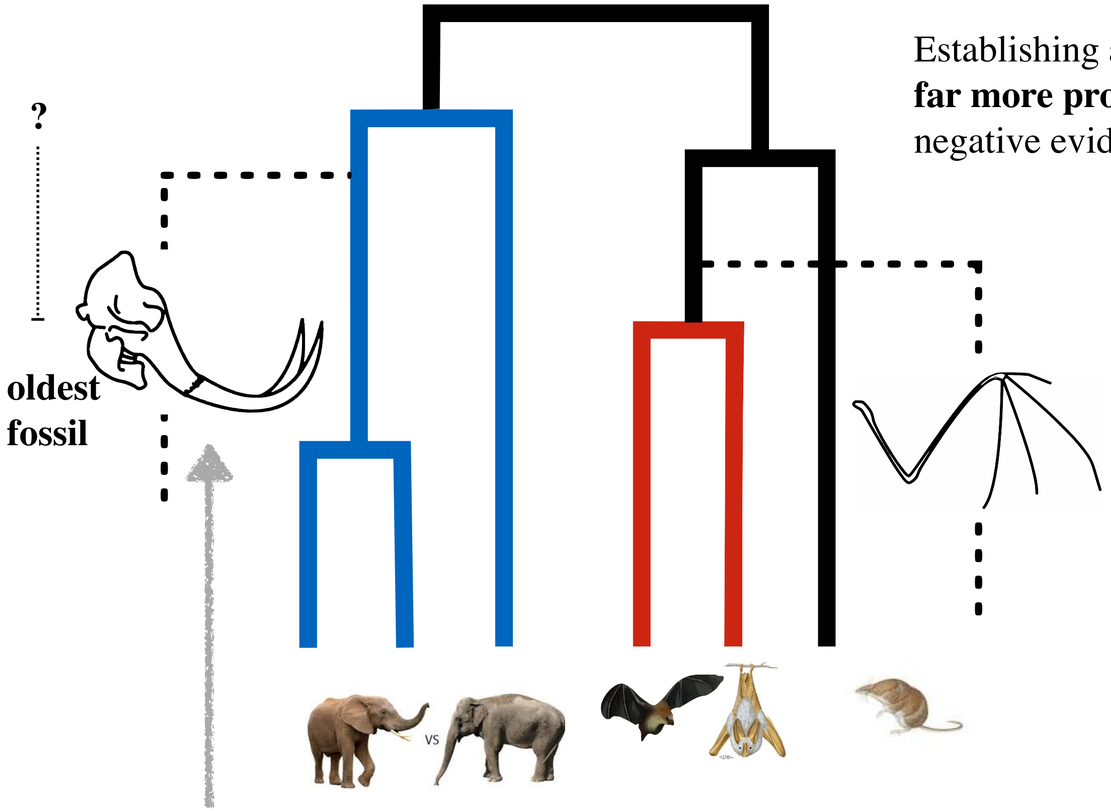


Rocks

ssolo@elte.hu
@sllsi

The geological record is the only source of information concerning absolute time





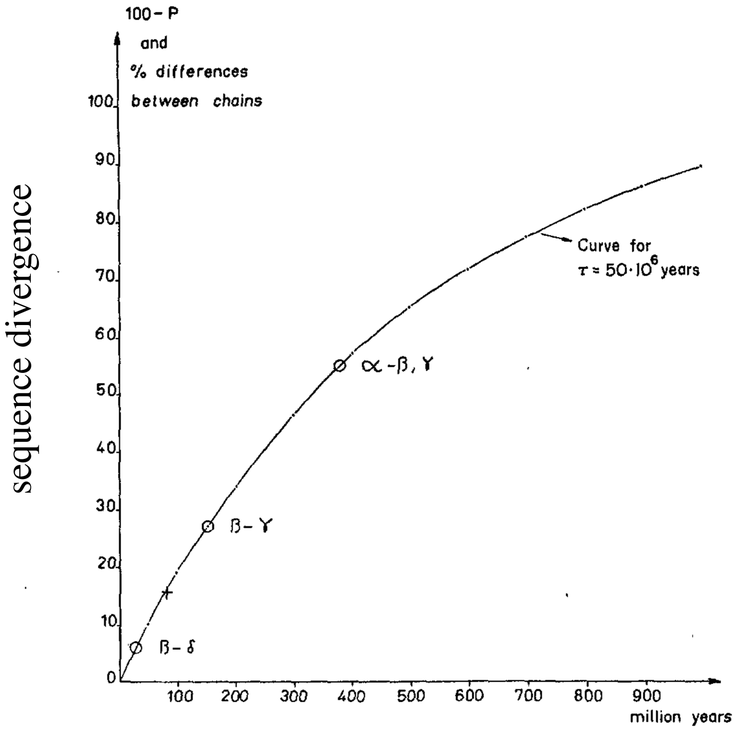
Establishing **a maximum constraint is far more problematic** — it relies on negative evidence — the absence of fossil.

The fossil record is **directly informative on the minimum ages** of clades based on the age of their oldest fossil representative

Clocks

ssolo@elte.hu
@sllsi

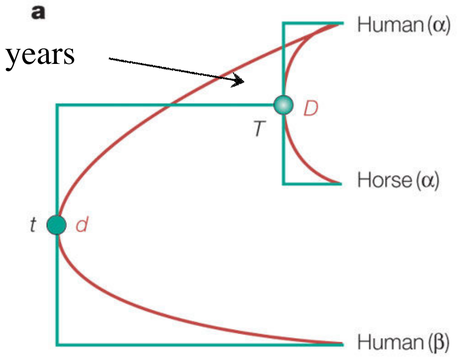
Zukerkandl and Pauling found that the **differences between homologous** amino acid **sequences** from different mammals **are roughly proportional to their time of divergence.**



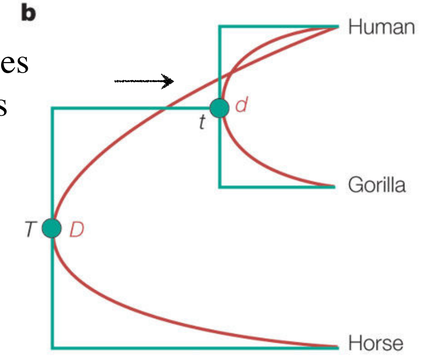
divergence time according to fossils

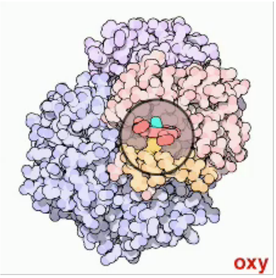
Zukerkandl and Pauling 1965

~130 [100-160] million years
18 substitutions



1 and 2 differences
~11 million years



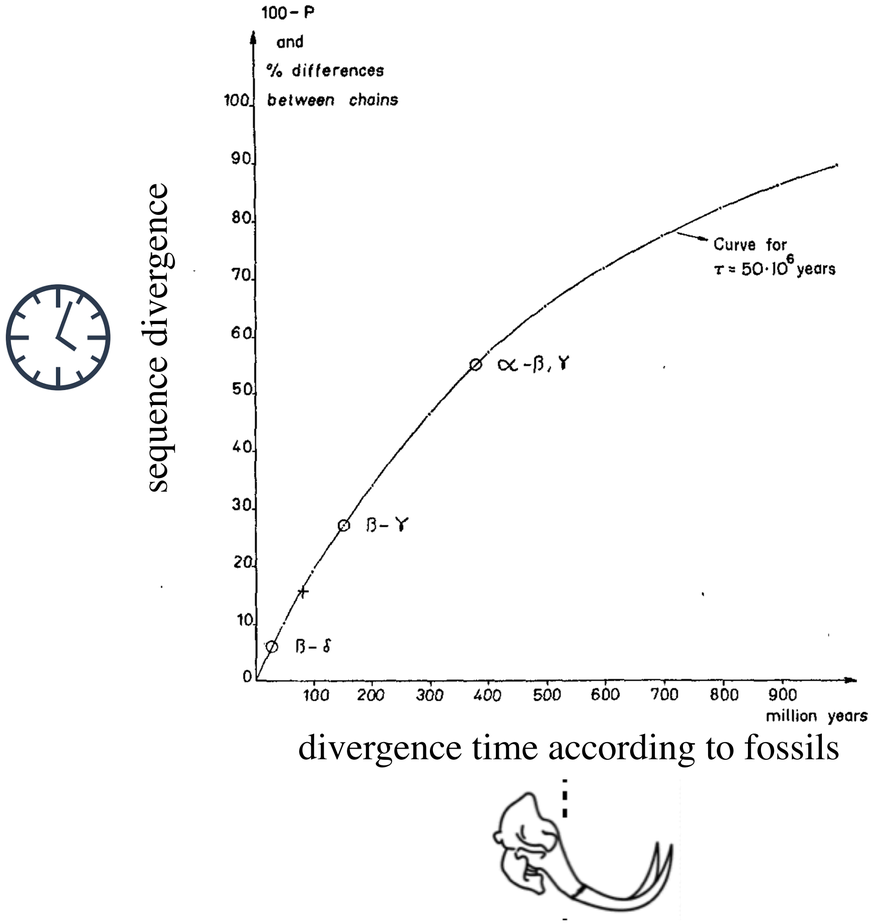


$2\alpha + 2\beta$

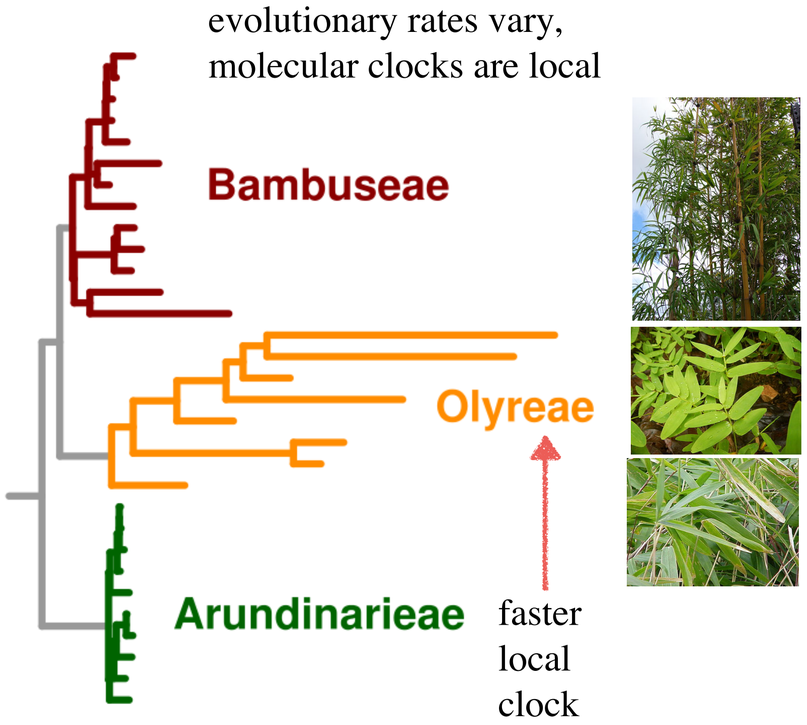
Clocks

ssolo@elte.hu
@sllsi

Zukerkandl and Pauling found that the **differences between homologous** amino acid **sequences** from different mammals **are roughly proportional to their time of divergence.**



Zukerkandl and Pauling 1965



$$\#\ \text{of events} = \text{rate} \times \text{time}$$

Wysocki et al. 2014 & Wikipedia

Rocks & Clocks

Inadequate modelling of the global violation of the molecular clock historically lead to great controversies..

ssolo@elte.hu
@sllsi

Opinion *TRENDS in Genetics* Vol.20 No.2 February 2004

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision

Dan Graur[1] and William Martin[2]

[1]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA
[2]Institut für Botanik III, Heinrich-Heine Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

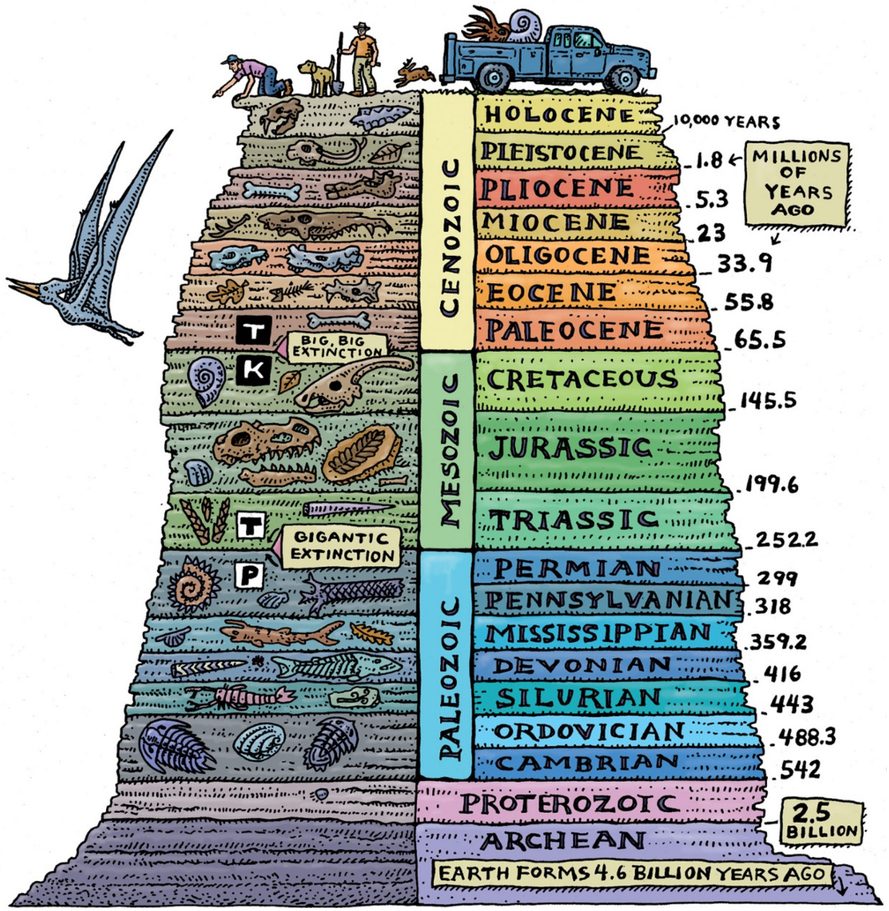
'We demand rigidly defined areas of doubt and uncertainty.' Douglas Adams

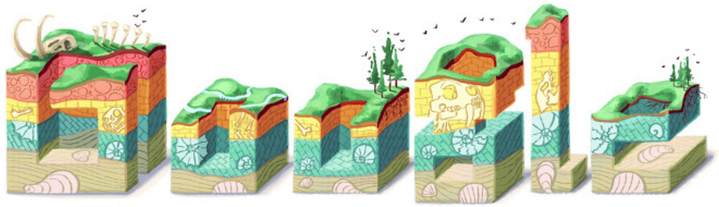
$$\#\ \text{of events} = \text{rate} \times \text{time}$$

Rocks & Clocks

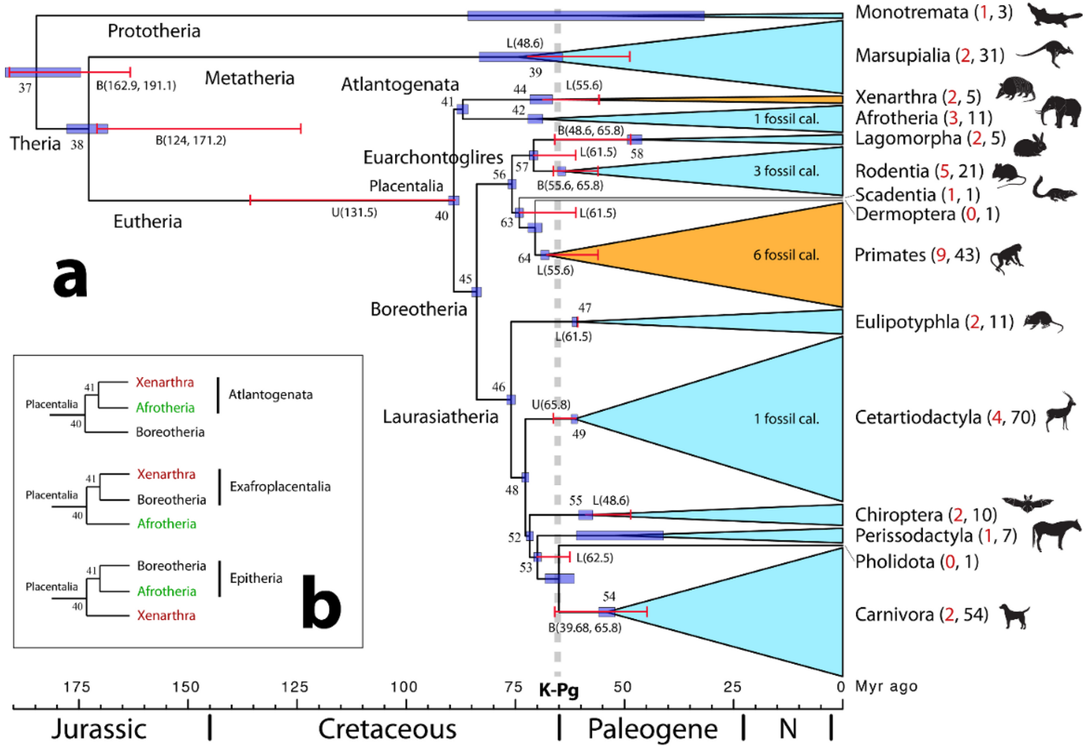
ssolo@elte.hu
@sllsi

Inadequate modelling of the global violation of the molecular clock historically lead to great controversies..





… today, Bayesian RMC methods have resolved most, but not all controversies, using **sequence based local molecular clocks anchored by multiple fossil calibrations.**

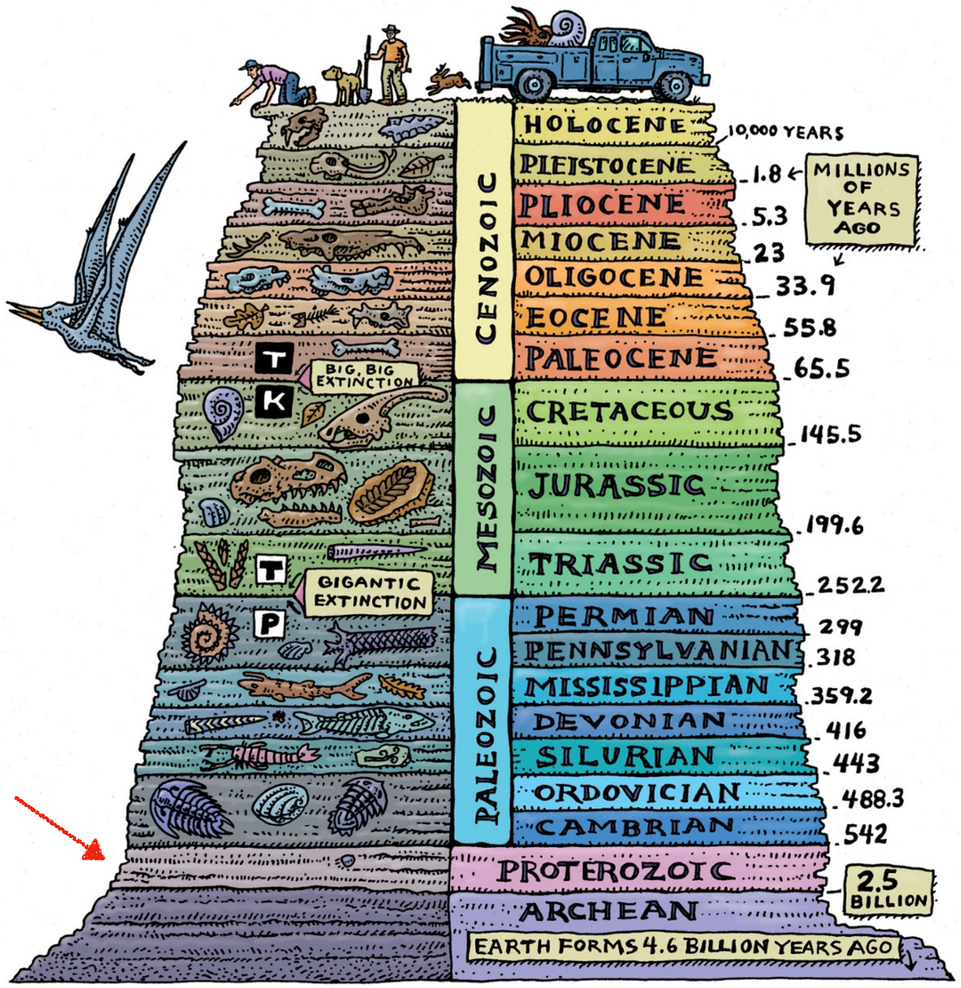


dos Reis et al. 2012

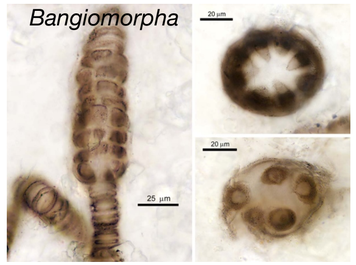
Rocks & Clocks

.. sequence based local molecular clocks anchored by multiple fossil calibrations.

erc GENECLOCKS
ssolo@elte.hu
@sllsi



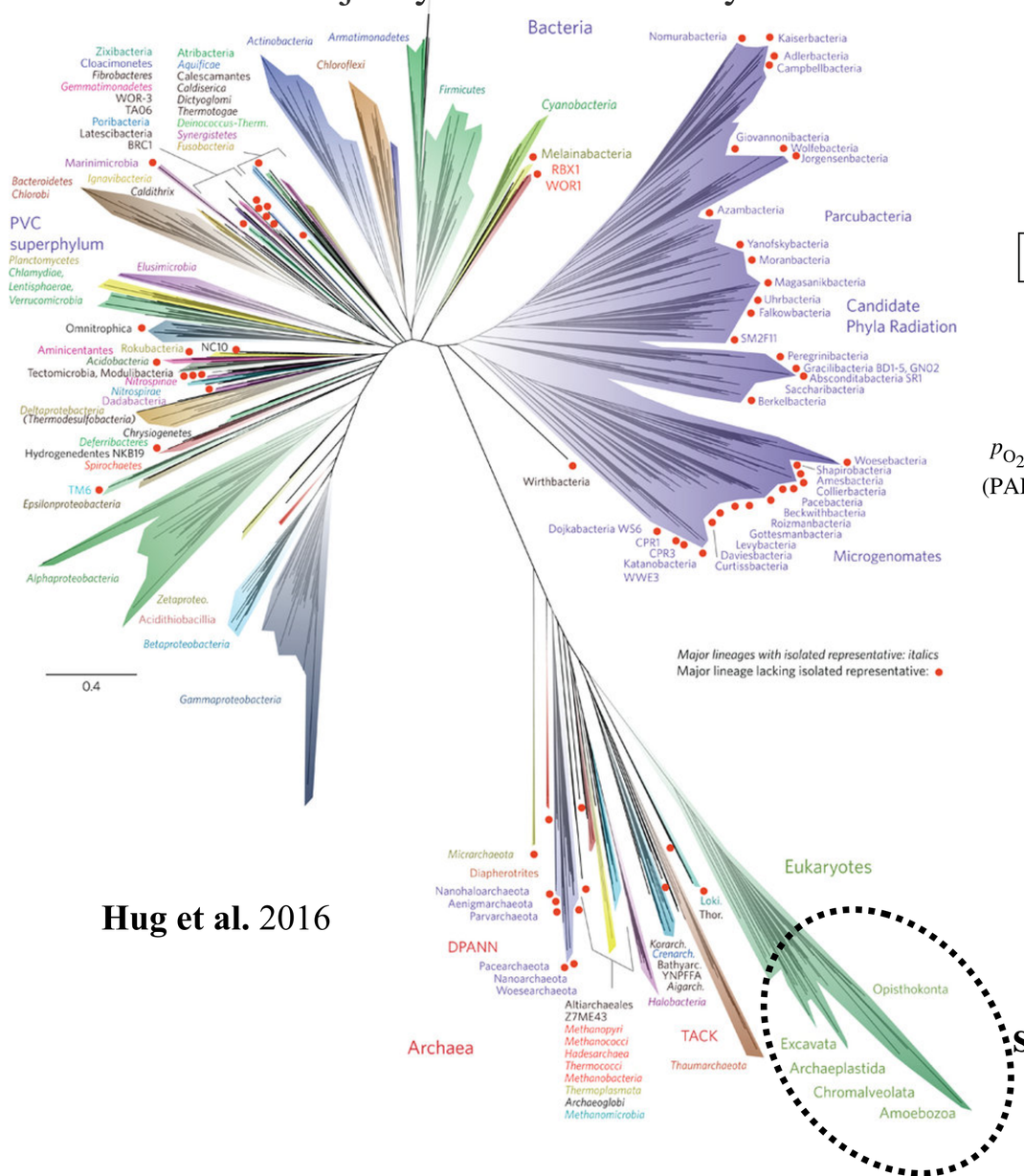
~ 1-1.5 Gya



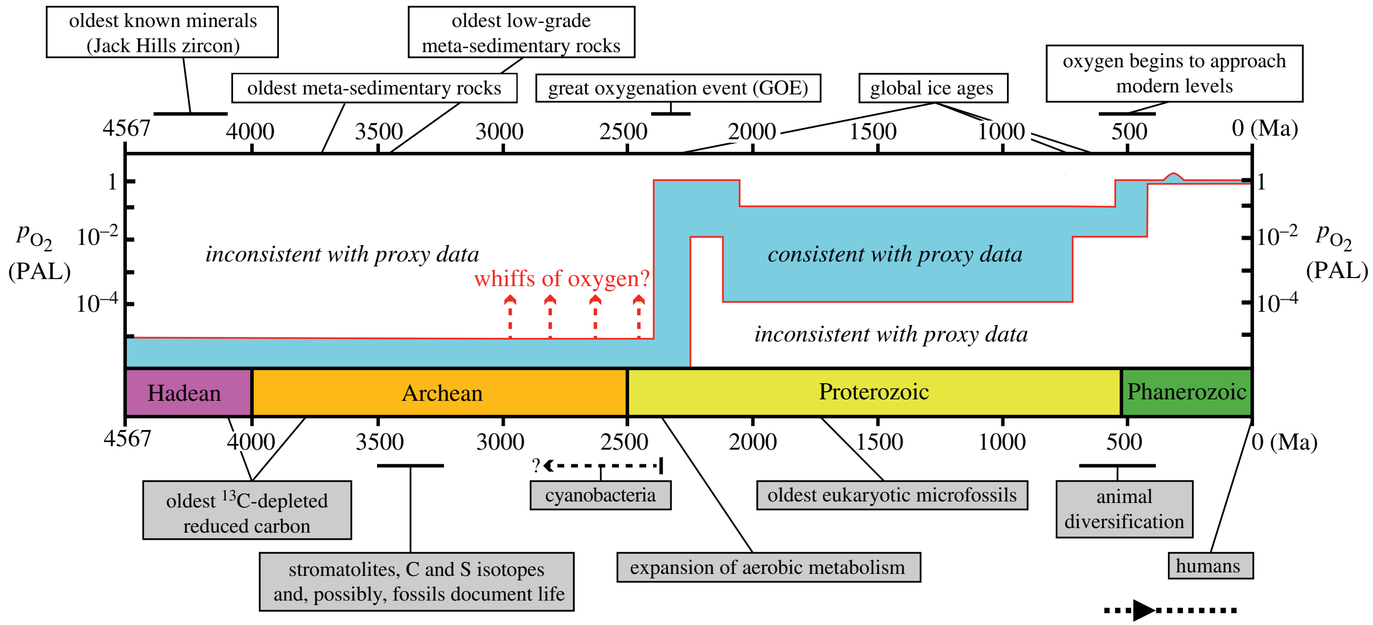
Rocks & Clocks

ssolo@elte.hu
@sllsi

For the majority of life's diversity and most its history **we lack sufficient fossils to anchor local clocks.**



Hug et al. 2016



Knoll et al. 2016

sufficient fossils

sufficient fossils

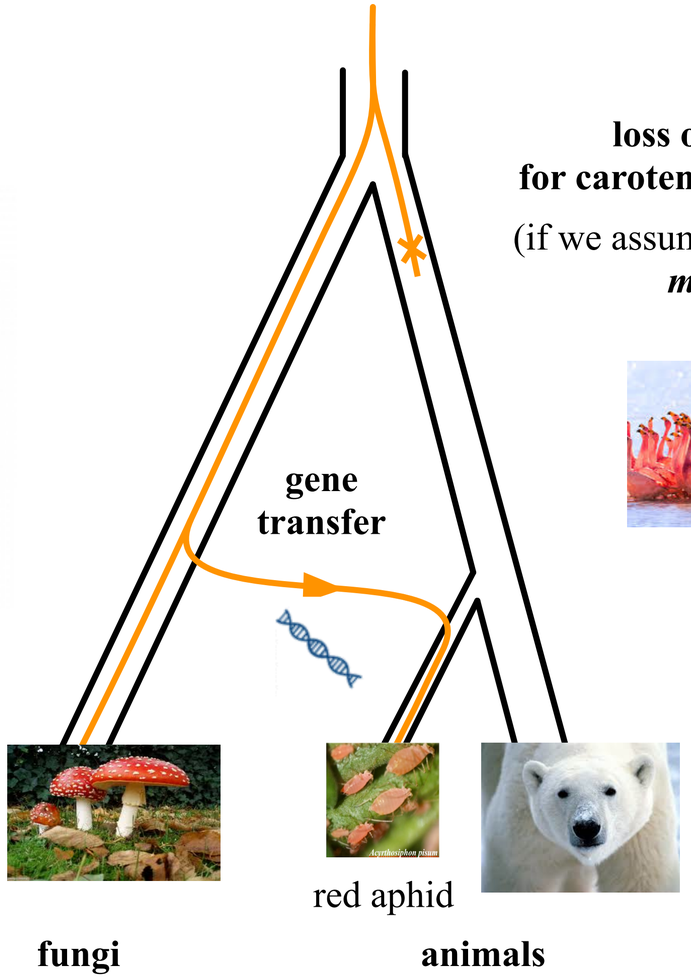
… genes from other species?

ssolo@elte.hu
@sllsi

Horizontal gene transfer is common among unicellular organisms, but examples are know even among animals.

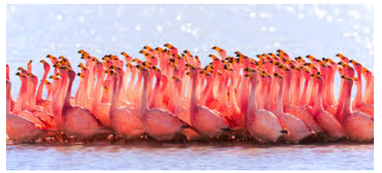
pea aphids

Moran & Jarvik 2010



loss of enzymes
for carotenoid biosynthesis

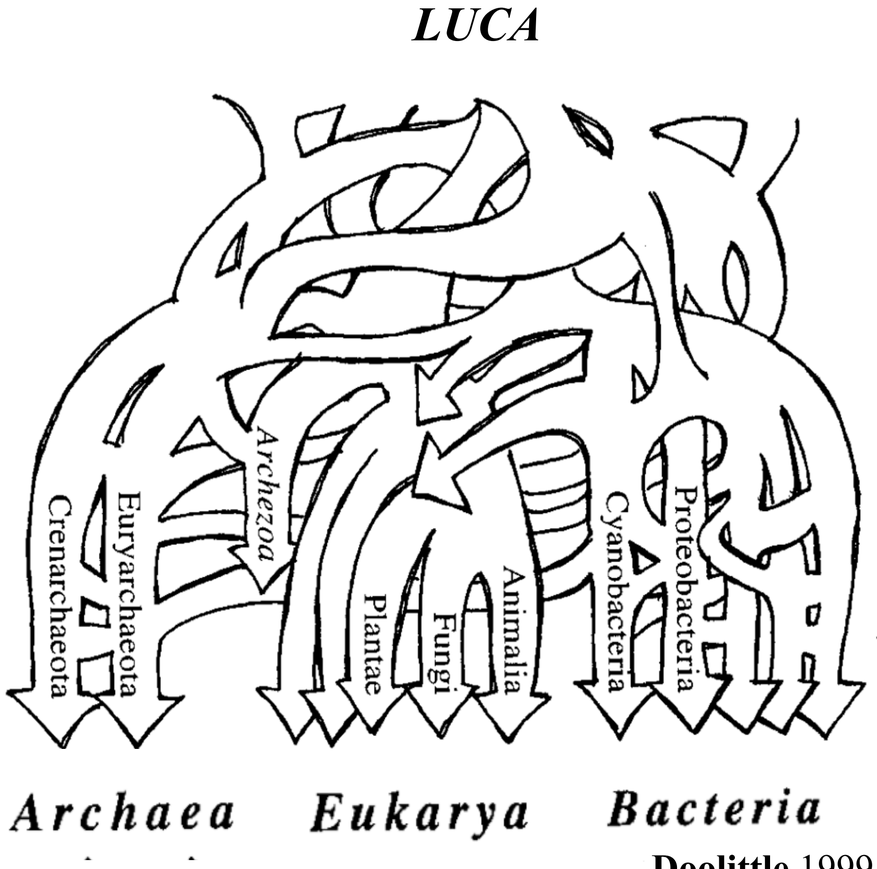
(if we assume only D&L we need ***many losses***)



Horizontal gene transfer as noise

ssolo@elte.hu
@sllsi

Gene transfers result in apparently contradicting gene phylogenies, fungi can seem closely related to aphids. A potentially high rate of transfer esp. early in the evolution of life, suggests that the vertical signal may be drowned in noise.

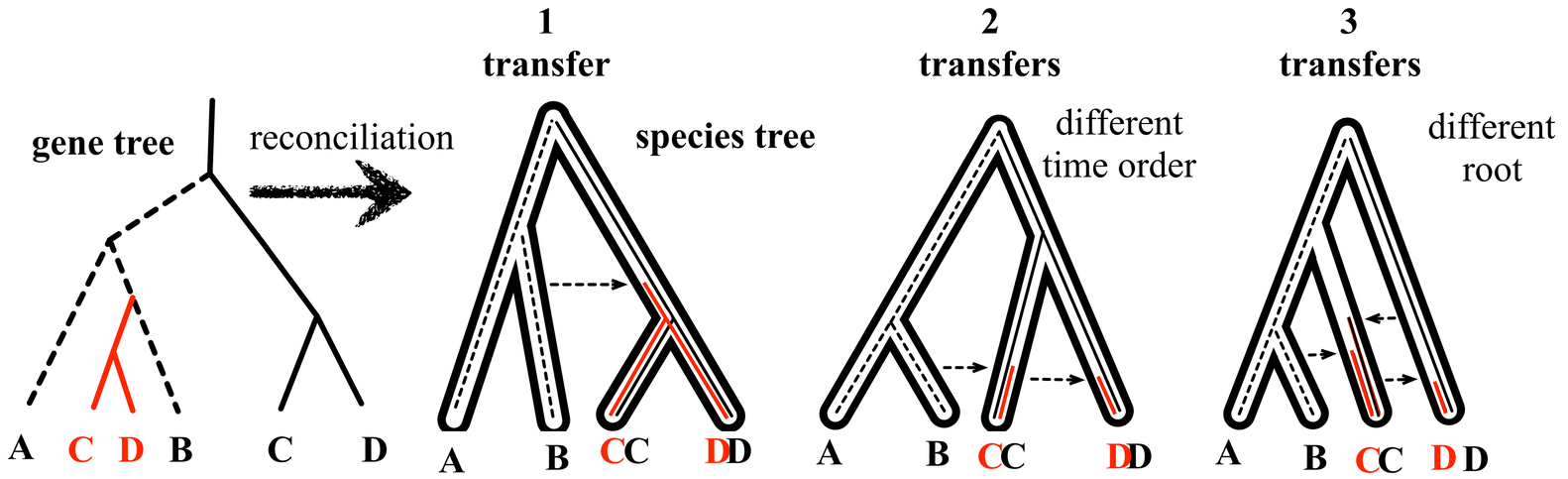


Doolittle 1999

Horizontal gene transfer as information

ssolo@elte.hu
@sllsi

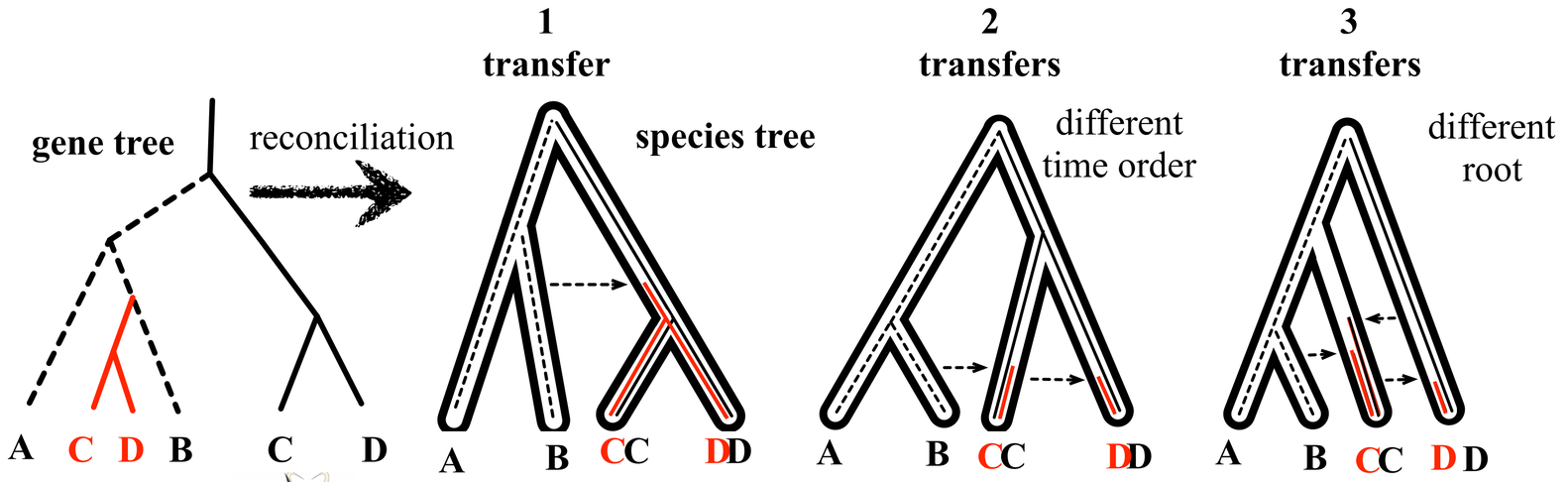
Transfer events, encoded in the topologies of gene trees can be thought of as "*molecular fossils*" that record the order of speciation events.



Horizontal gene transfer as information

ssolo@elte.hu
@sllsi

Transfer events, encoded in the topologies of gene trees can be thought of as "*molecular fossils*" that record the order of speciation events.

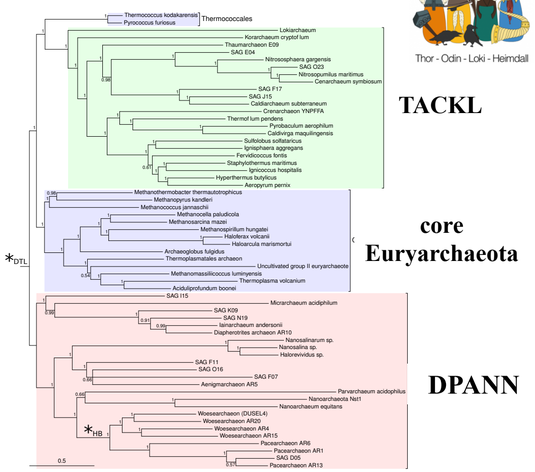


Vincent Daubin

LBBE

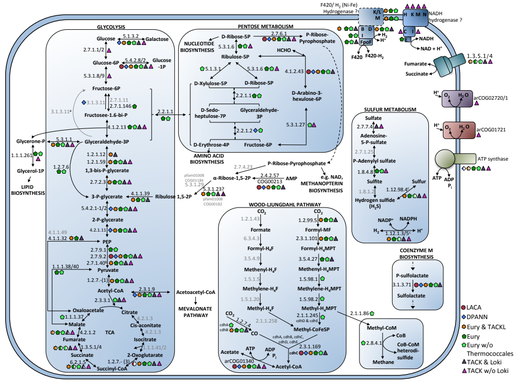
Tom Williams

U.Bristol



Thor - Odin - Loki - Heimdall

LACA's metabolism



An anaerobe that could fix CO_2 to acetyl-CoA and generate acetate and ATP from it.

Rooting archaea wo. an outgroup:

Williams, **Szöllősi**, .. & Embley ***PNAS*** (2017)
Integrative modelling of gene and genome evolution roots the archaeal tree of life

Rooting cyanobacteria wo. an outgroup:

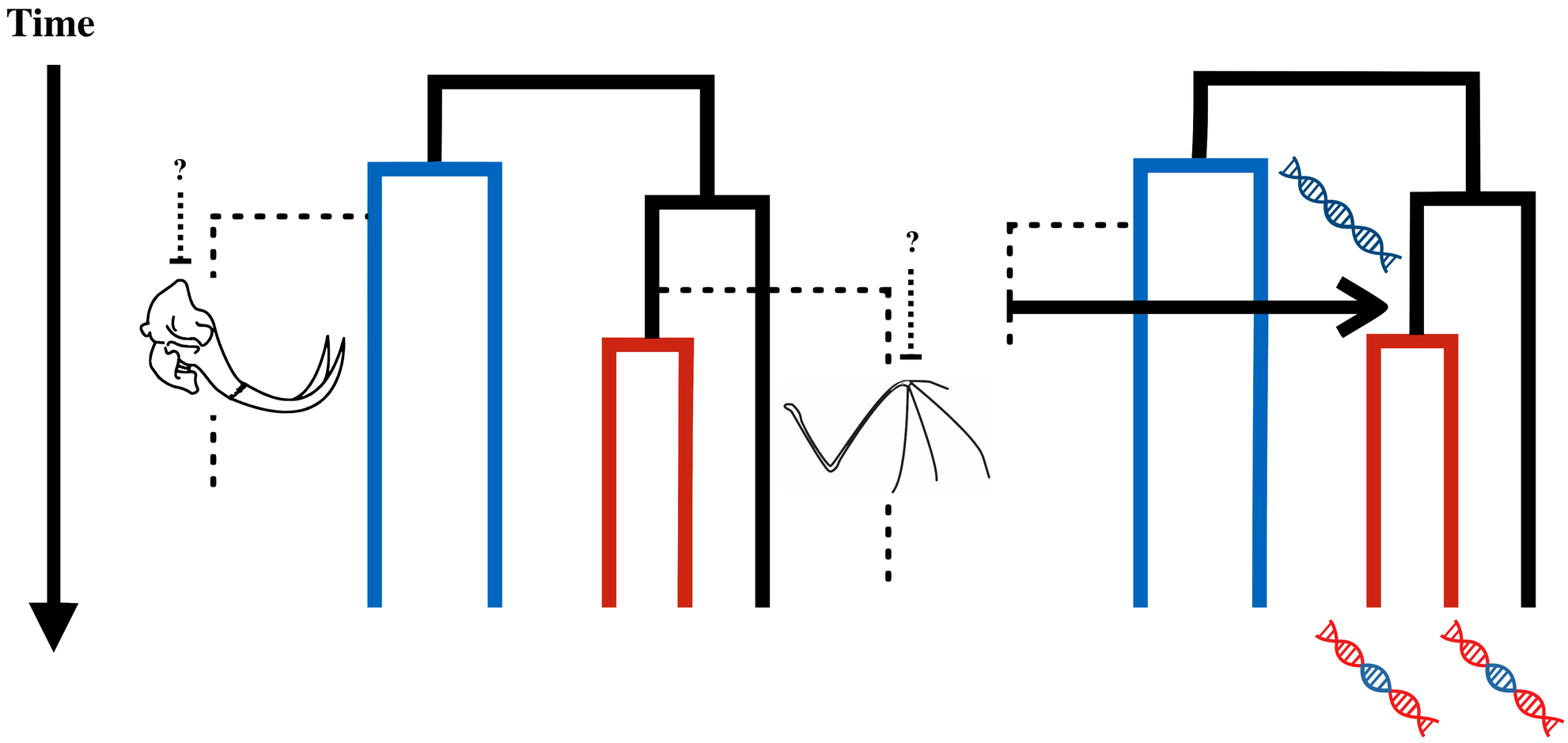
Szöllősi, Boussau, Abby, Tannier & Daubin ***PNAS*** (2012)
Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations

… and genes from other species!

ssolo@elte.hu
@sllsi

Fossils provide **direct evidence on minimum age**, but only **indirect evidence on maximum and relative ages**.

Transfers are not informative on absolute age, but do provide **direct evidence on relative ages**.



http://rdcu.be/KrjA

Joint reconstruction of gene trees and the species tree

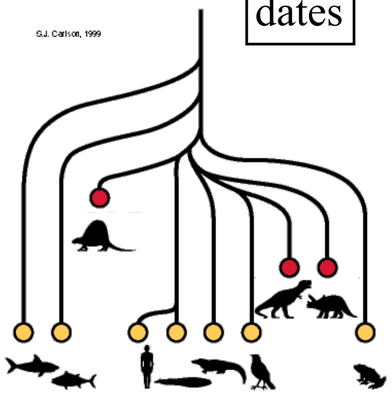
ssolo@elte.hu
@sllsi

Using a hierarchical model where gene trees are generated along the species tree and sequences are generated along gene trees we can *in theory* **jointly infer gene trees and species trees**.

species tree

S

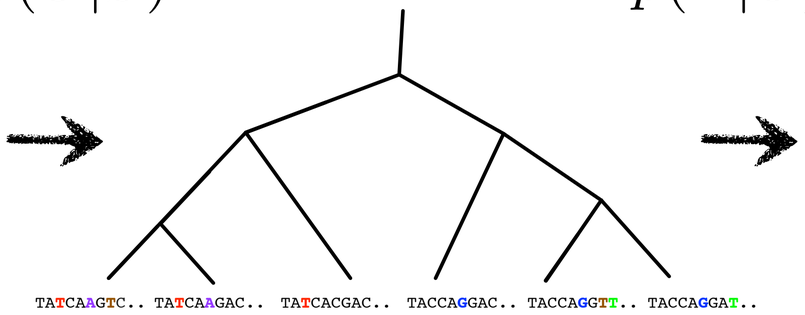
dates



$p(G|S)$

gene trees

G_j



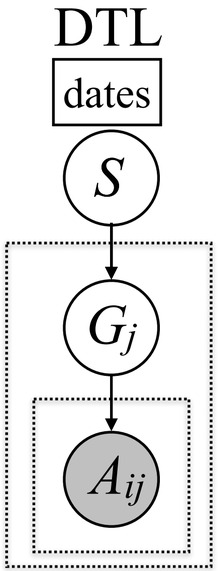
$p(A|G)$

sequences (data)

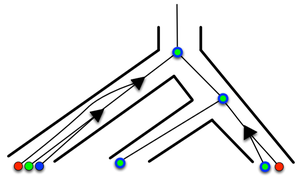
A_{ij}

```
TATCAAGTC..
TATCAAGAC..
TATCACGAC..
TACCAGGAC..
TACCAGGTT..
TACCAGGAT..
```

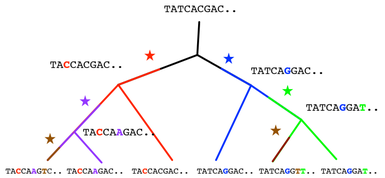
DTL



gene birth and death



substitution events



Species-tree-aware reconstruction of gene trees

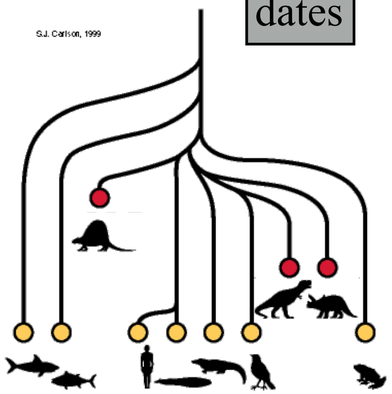
In practice, at present, **fixing the species tree** is much faster computational and still allows us to estimate more accurate gene trees, better ancestral sequences, improved synteny and fewer transfers.

ssolo@elte.hu
@sllsi

species tree
(treated as data)

S

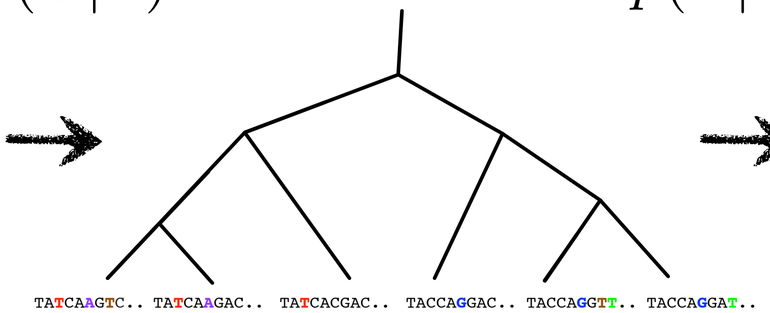
dates



$p(G|S)$

gene trees

G_j



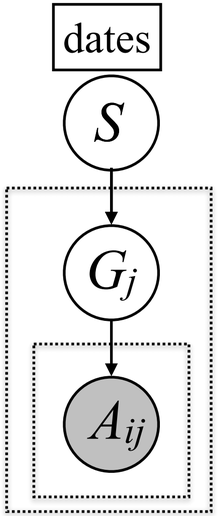
$p(A|G)$

sequences
(data)

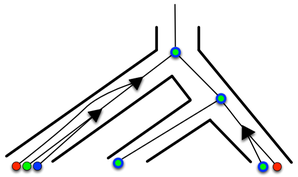
A_{ij}

```
TATCAAGTC..
TATCAAGAC..
TATCACGAC..
TACCAGGAC..
TACCAGGTT..
TACCAGGAT..
```

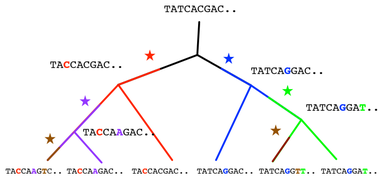
DTL



**gene
birth and death**



**substitution
events**



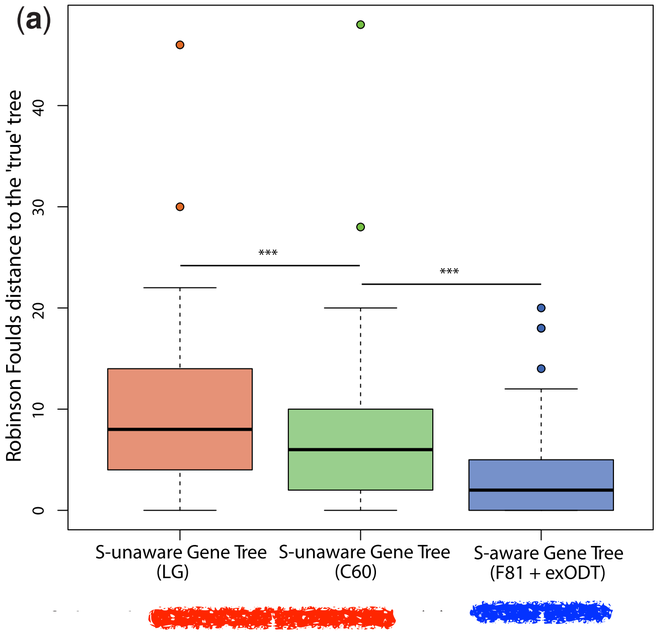
Species-tree-aware reconstruction of gene trees

ssolo@elte.hu
@sllsi

More accurate gene trees, better ancestral sequences, improved synteny and fewer transfers

significantly reduced number of DTL events
(- 24% in Ds, **- 59% in Ts** and - 46% in Ls).

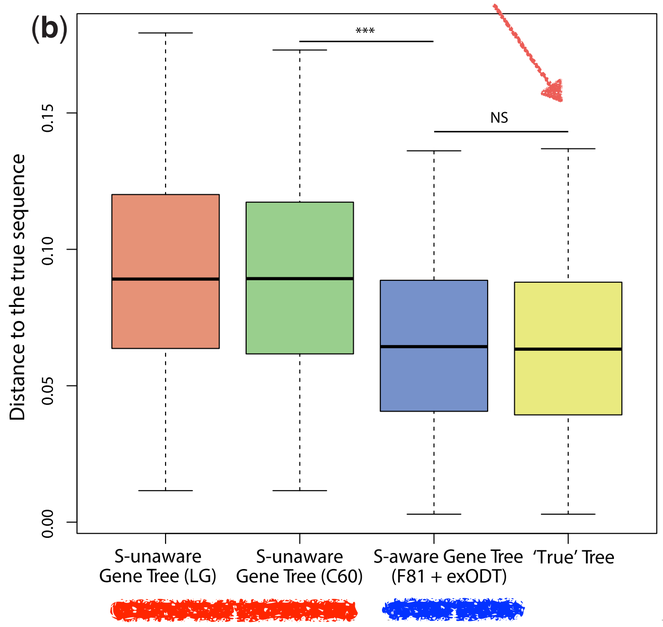
more simulations..



S-aware gene trees

S-unaware gene trees

"real" tree based sequences!



S-aware gene trees

S-unaware gene trees

LeuB

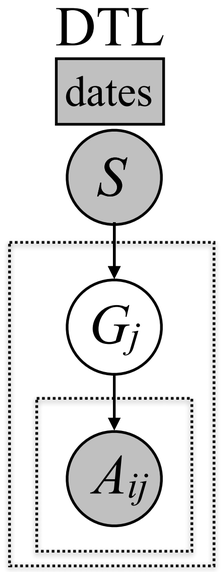
3-isopropylmalate dehydrogenase
E. C. 1.1.1.85

extant Firmicutes

$K_M^{(IPM)}$ (mM)	T_{opt} (°C)	$\Delta G^{\ddagger}_{N-U}$ ($kJmol^{-1}$)
0.2	47	94.9
0.7	53	95.9
1.1	69	100.7

Firmicute ancestor ASR

1.6	**85**	**110.9**
6.8	**78**	**91.4**



implemented in ALE:
http://github.com/ssolo/ALE

Groussin, Hobbs, **Szöllősi**, Gribaldo, Arcus & Gouy ***Mol. Biol. Evol.*** (2015)
Toward More Accurate Ancestral Protein Genotype–Phenotype Reconstructions with the Use of Species Tree-Aware Gene Trees

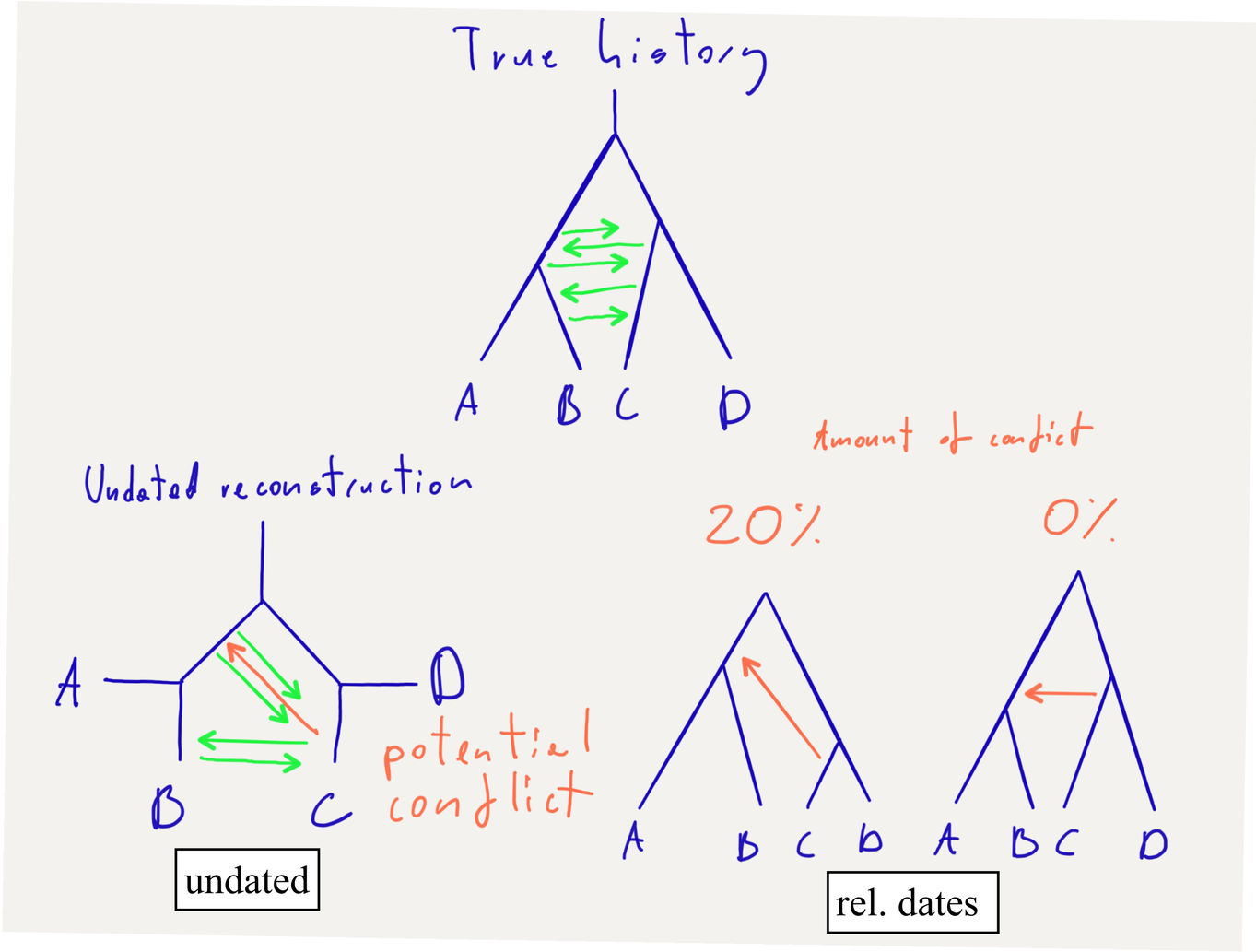
Szöllősi, Tannier, Lartillot & Daubin ***Systematic Biology*** (2013)
Lateral Gene Transfer from the Dead

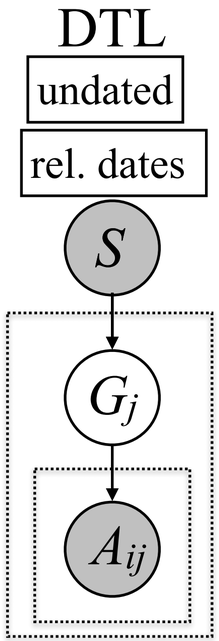
Szöllősi, Rosikiewicz, Boussau, Tannier & Daubin ***Systematic Biology*** (2013)
Efficient exploration of the space of reconciled gene trees

Relative age constrains from transfers

ssolo@elte.hu
@sllsi

Transfers inferred by an “undated” version of the species tree-aware method ALE were input into the MaxTiC (**max**imal **ti**me **c**onsistency) optimisation method to obtain relative age constrains.





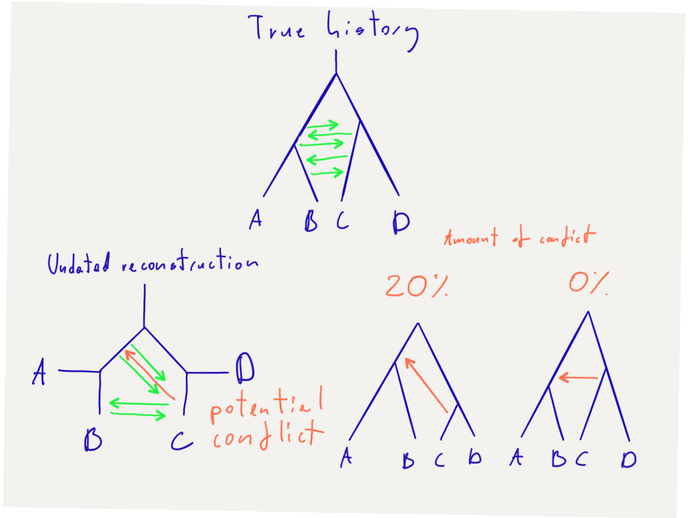
Eric Tannier
Vincent Daubin
LBBE

Chauve, Rafiey, Davin, Scornavacca, Veber, Boussau, Szollosi, Daubin, Tannier **bioRχiv** (2017)
MaxTiC: Fast Ranking Of A Phylogenetic Tree By Maximum Time Consistency with Lateral Gene Transfers

Relative age constrains from transfers

ssolo@elte.hu
@sllsi

Transfers inferred by an “undated” version of the species tree-aware method ALE were input into the MaxTiC (**max**imal **ti**me **c**onsistency) optimisation method to obtain relative age constrains.



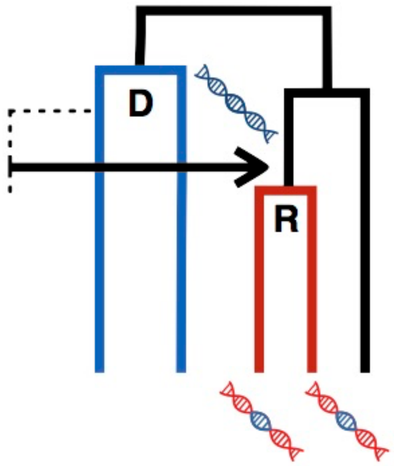
40 Cyano.
60 Arch.
60 Fungi

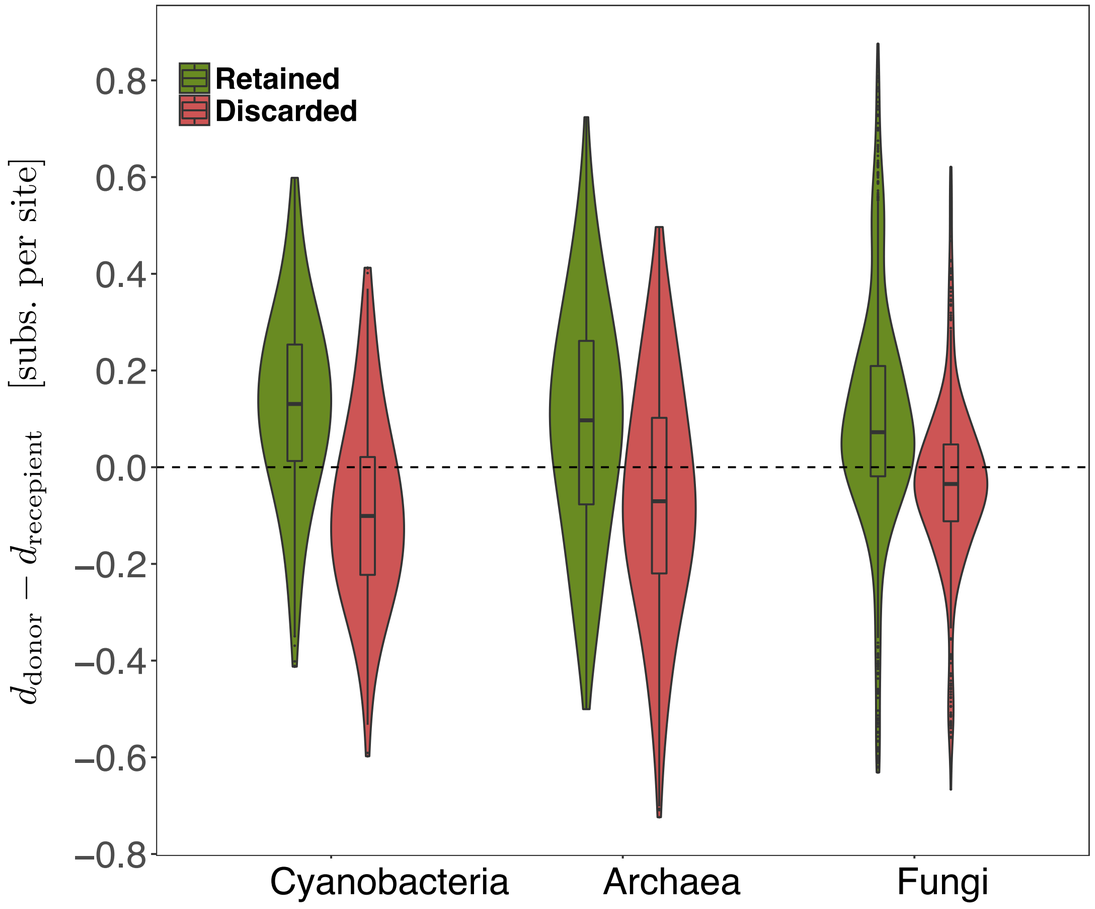
↓

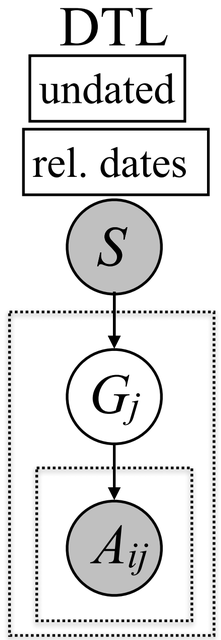
ALE
+
MaxTiC

↓

HGT events
e.g for cyano 3322
informative constraints








Eric Tannier

Vincent Daubin

LBBE

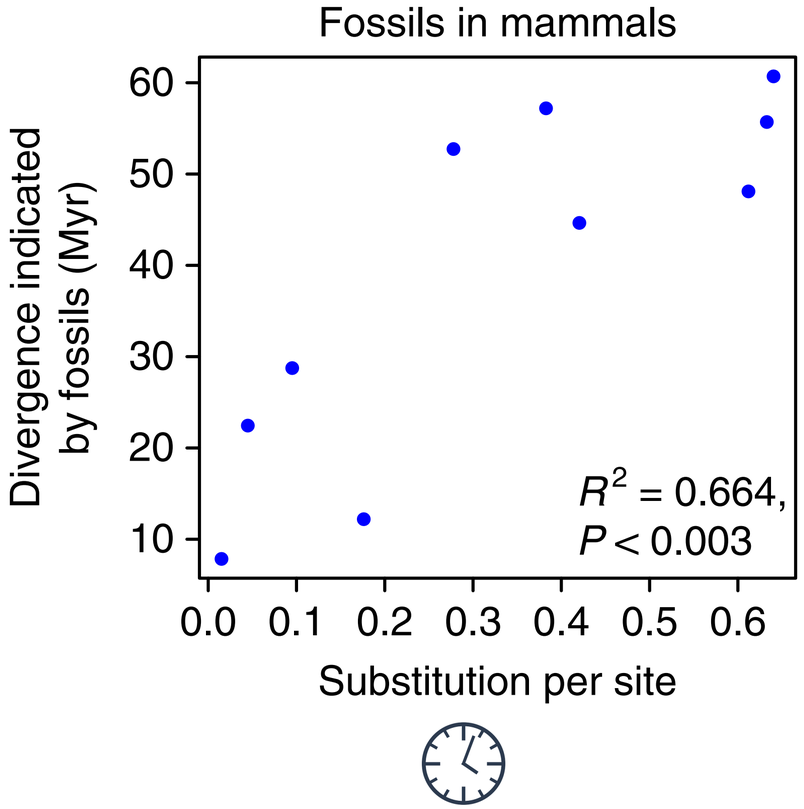
Chauve, Rafiey, Davin, Scornavacca, Veber, Boussau, Szollosi, Daubin, Tannier **bioRχiv** (2017)
MaxTiC: Fast Ranking Of A Phylogenetic Tree By Maximum Time Consistency with Lateral Gene Transfers

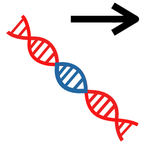
Rocks, clocks and genes from other species

ssolo@elte.hu
@sllsi

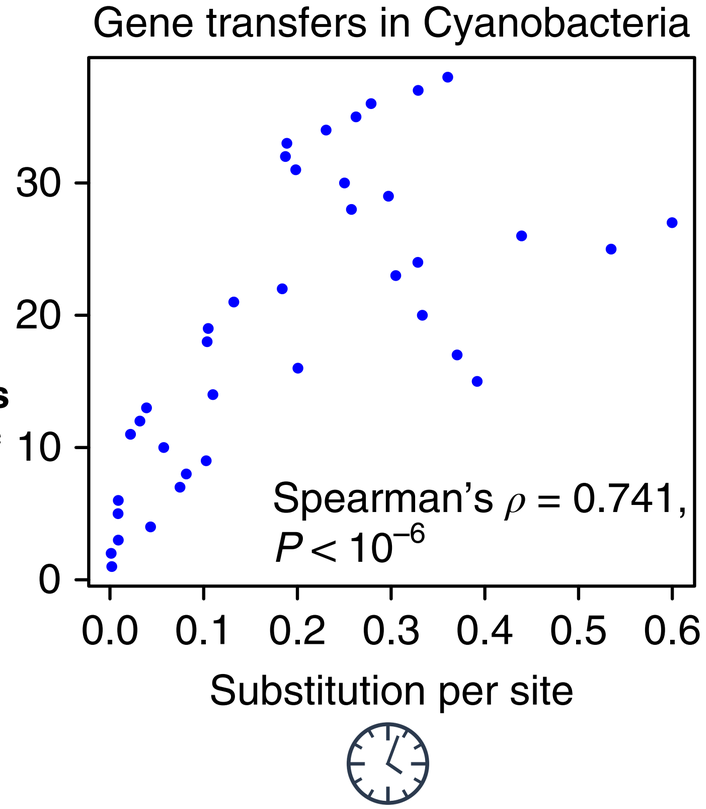
A direct comparison between fossils and transfers is not possible. Following Zuckerkandl and Pauling, we correlated both fossil and transfer based age estimates with sequence divergence:

12 calibrations with maximum age
dos Reis et al. 2012





3322 HGTs
informative constraints



40 genomes

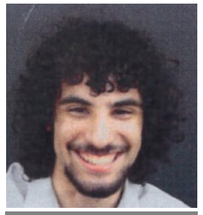
↓

ALE
+
MaxTiC

↓

3322 HGT events
informative constraints

Part of the above correlation may trivially result from the fact that parent nodes are necessarily both older and more distant to extant sequences than their direct descendants ..



Adrian Davin
LBBE

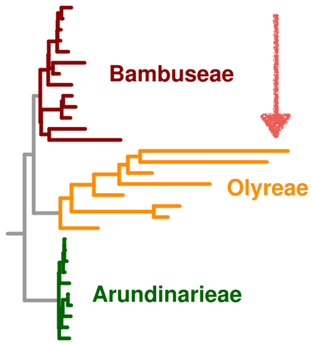
http://rdcu.be/KrjA

Rocks, clocks and genes from other species

ssolo@elte.hu
@sllsi

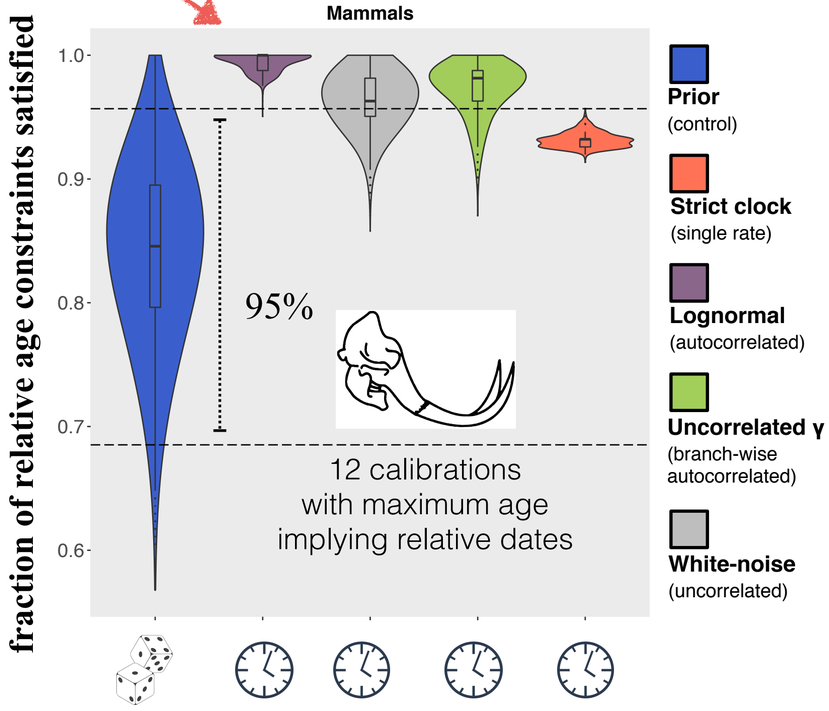
To directly compare relative age constraints, we measured how different **relaxed molecular clock models, without fossil calibrations**, are able to **predict the relative timing of speciations** implied by fossils and by transfers.

faster local clock



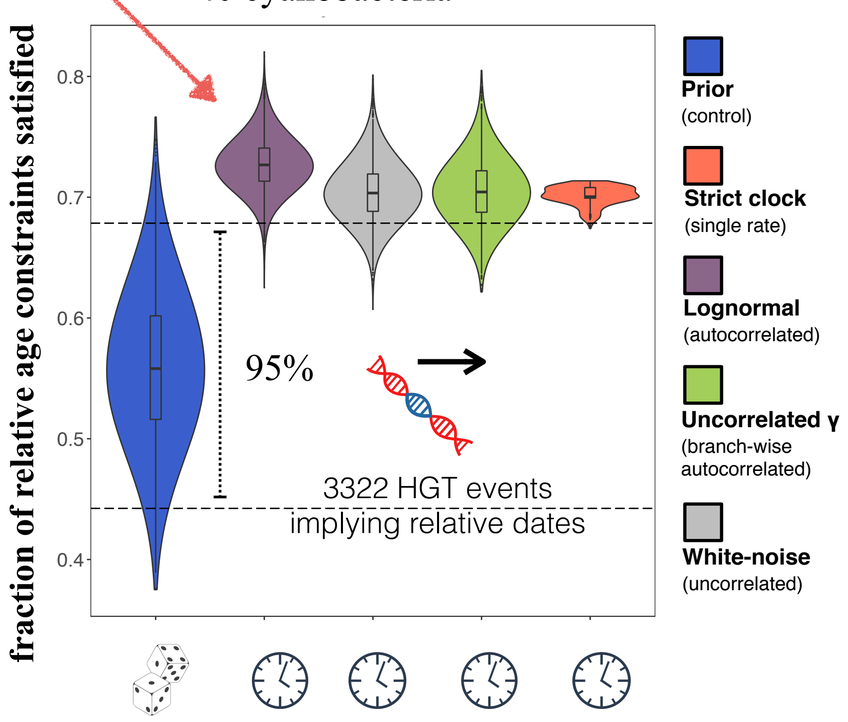
faster local
clock in rodents

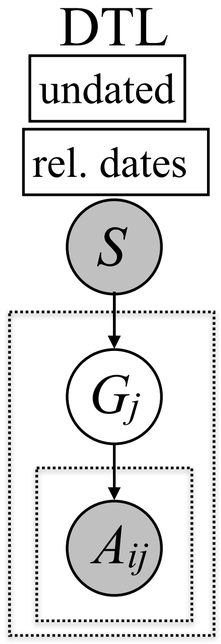
36 mammals



faster local clock
in prochlorococcus

40 cyanobacteria



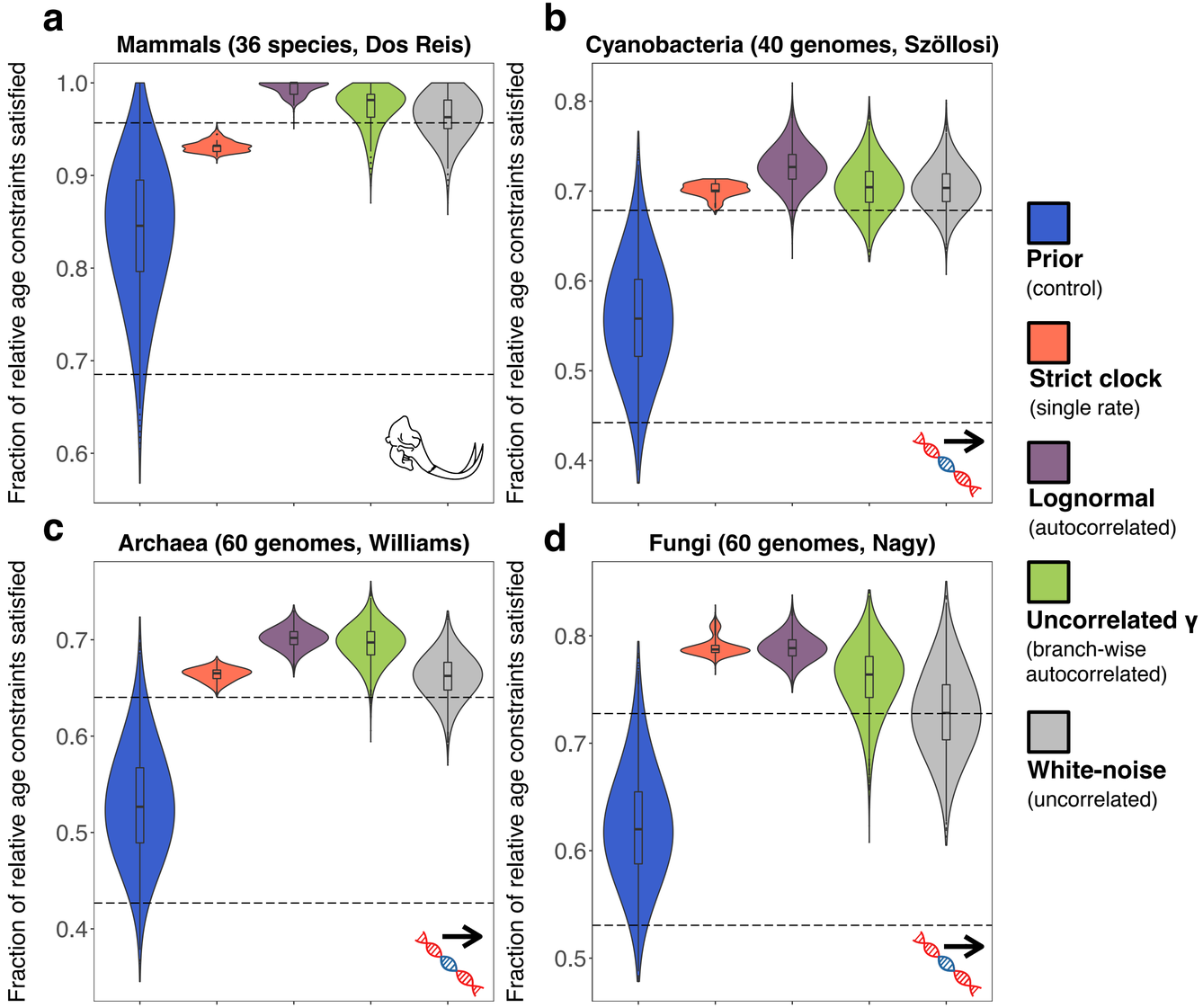


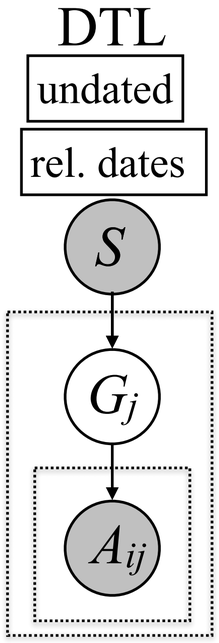
http://rdcu.be/KrjA

Rocks, clocks and genes from other species

ssolo@elte.hu
@sllsi

To directly compare relative ages, we measured how different relaxed molecular clock models, without fossil calibrations, are able to predict the relative timing of speciations implied by fossils and by transfers.





http://rdcu.be/KrjA

nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0525-3

read here:

http://rdcu.be/KrjA

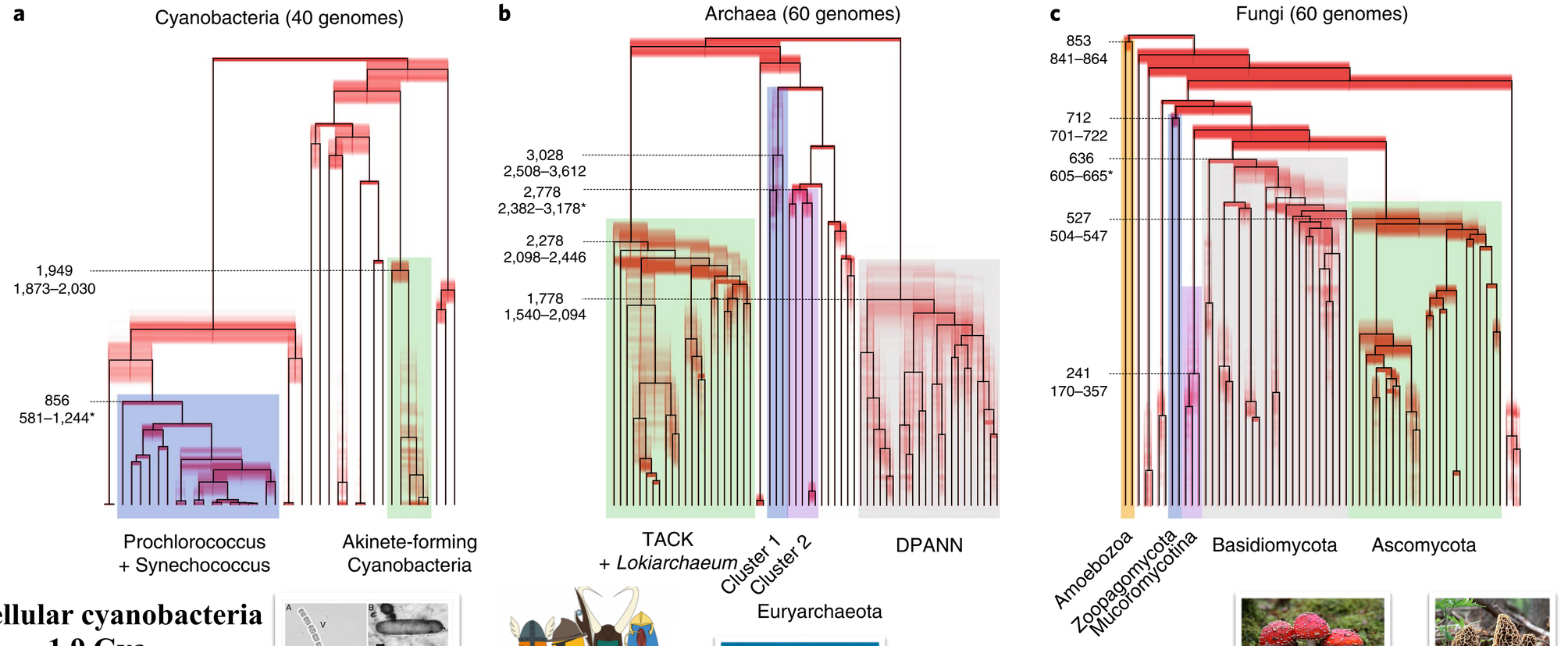
erc GENECLOCKS

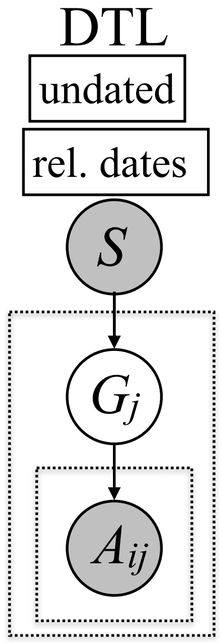
ssolo@elte.hu

@sllsi

Gene transfers can date the tree of life

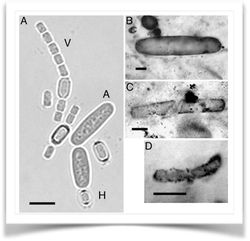
Adrián A. Davín[1], Eric Tannier[1,2], Tom A. Williams[3], Bastien Boussau[1], Vincent Daubin[1*] and Gergely J. Szöllősi[4,5*]





multicellular cyanobacteria
~ 1.9 Gya

Tomitani *et al.* 2006



Thor - Odin - Loki - Heimdall

~ 2.3 Gya

biogenic methane
~ 3.5 Gya

Fungi
0.6-0.8 Gya