

Outline

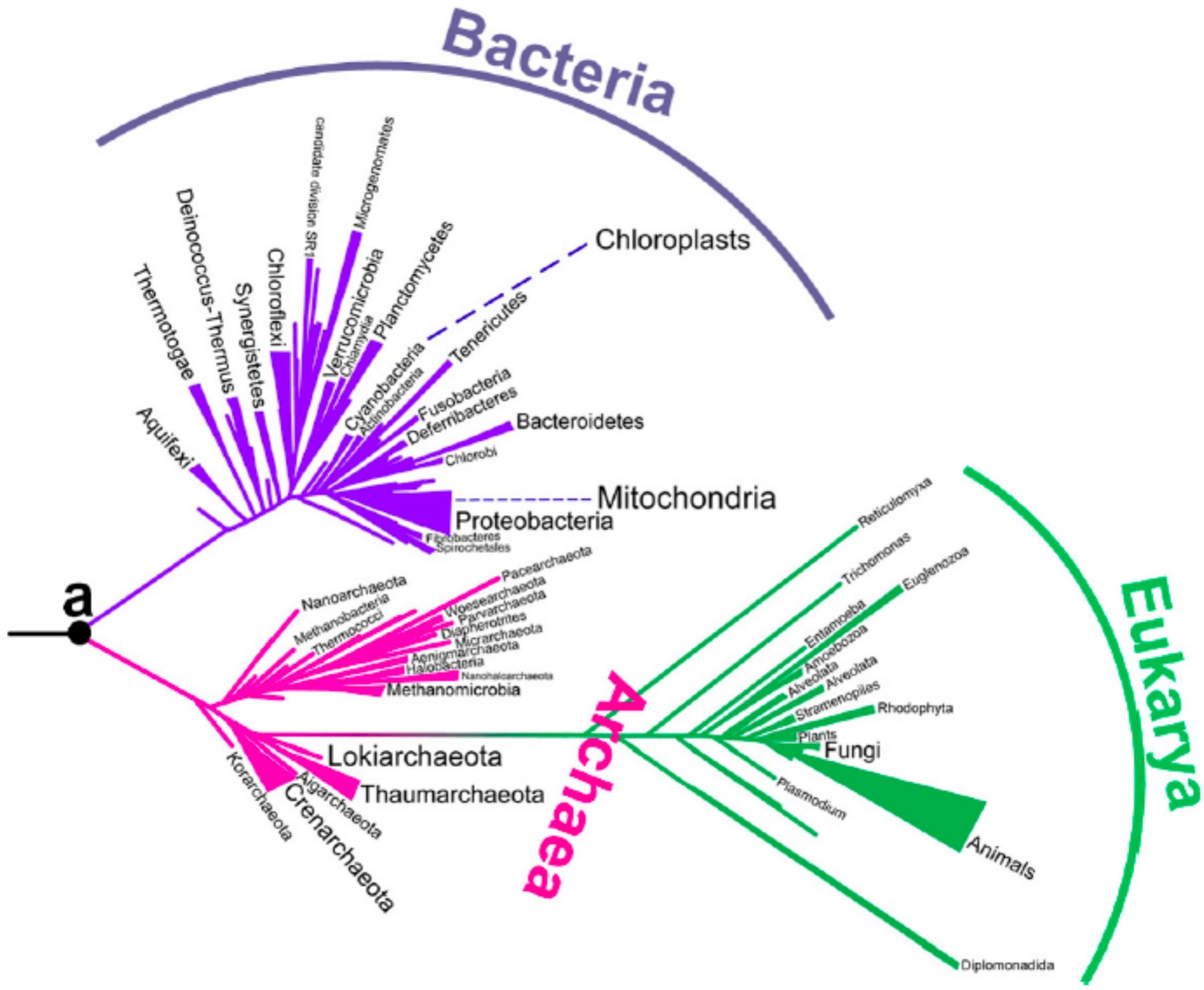
Part I - PPT lecture / Ancestral sequence reconstruction

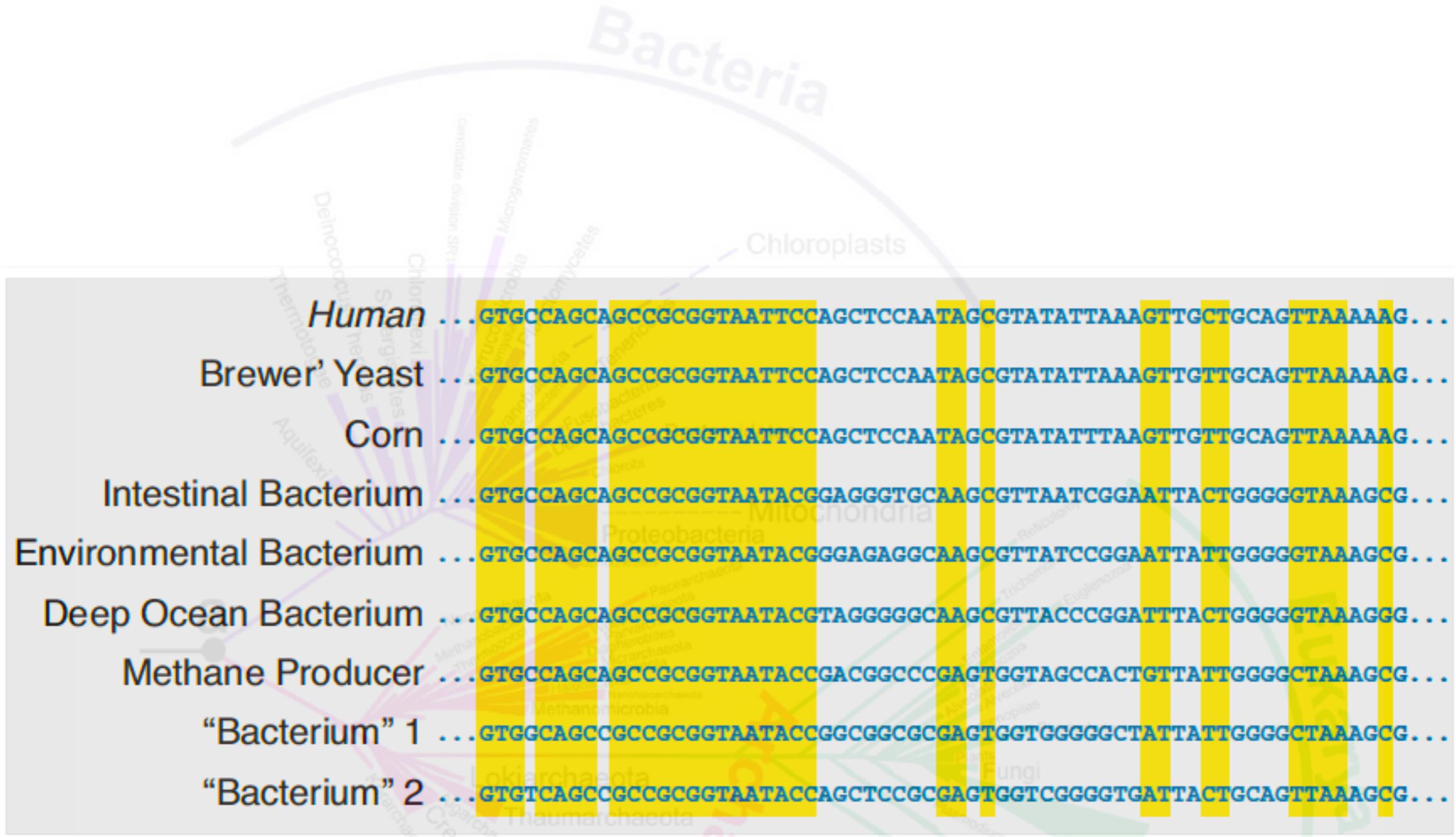
Part II - white board lecture / A very shot introduction to phylogenetic trees: a) maximum parsimony & likelihood, b) predicting ancestral states, c) substitution/evolution-models

Part III - Linux-bash-based hands-on application -
reconstruction of ancestral sequences of modern ADH4
sequences using t-coffee & iqtree

Ancestral Sequence Reconstruction

Gönensin Ozan Bozdag,
1 Subat 2018, @ Hacettepe Biyoloji





An idea

“... two types of information that can be derived from the comparison of different homologous polypeptide chains. First, *it is possible to determine, w/ some probability, the a.a. sequence of their presumed common polypeptide-chain ancestor.*”

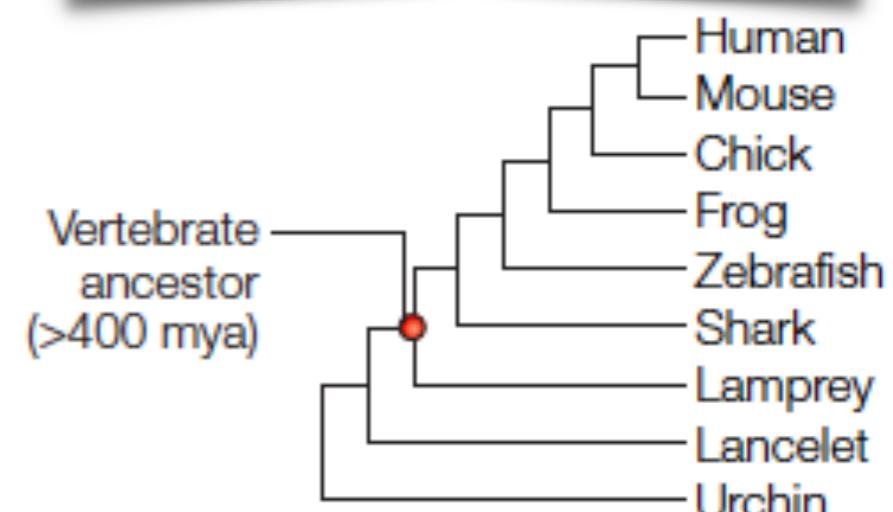
–Pauling & Zuckerkandl 1963

Reconstruct *hypothetical* ancestors

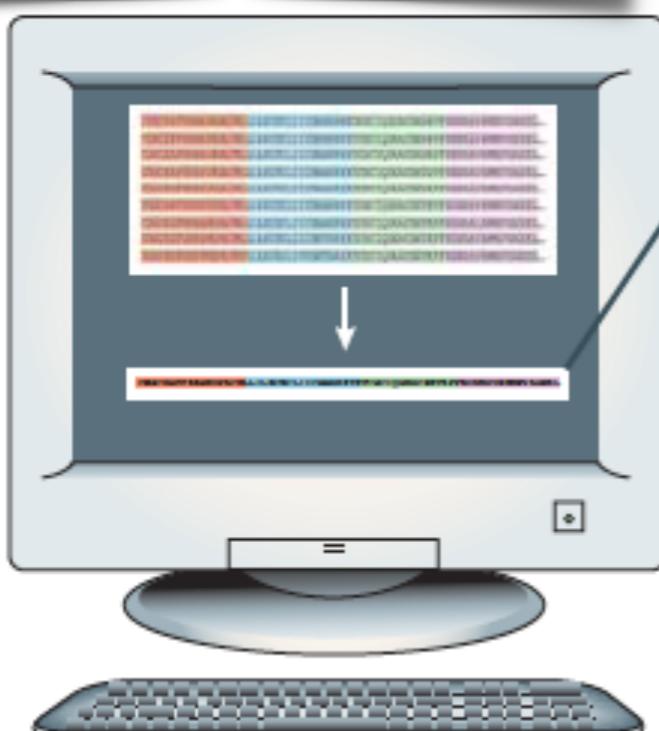
#1 Align sequences

```
CDGCKSFFKRAVEGACMLLILKCKCLIICHHSRRWTKHClQGDACEHSKFKENRAVEMMEVAGEKL...
CDGCKTFFKRAVEGACMLLILKCKCLIICHHSRRYKTCHCIQGRACEHSKFKENRAVEMMEVAGEKL...
CDGCKAFFKKAVEGACMLLILKCKCLIICHHARRYKTCHCVQGRACEKSFKFKENRAVEMMEVAGEKL...
CEGCKAFFKKSVEGACMLLILKCKCLIICHHARRYKTCHCIQGRACEKTDFPKENRAVEMMEVAGEKL...
CEGCKSFFKKSVDGACMLLILKCKCLIICHHARRYKTCHCIQGRACEKTKFRENRAVEMMEVAGEKL...
CEGCKSFFKKSVDPACMLLILKCKCLIICHHARRYKTCHCIQGRACEKTKFKDNRSEVEMMEVAGEKL...
CEGCKSFFKKSMDPACMLLILKCKCLIICHHARRYKTCHCIQGRACEKTKFKDNRRALEMMEVAGEKL...
CEGCKSFLKKSMDPACMLLILKCKCLIICHHTRRYKTCHCIQGRACEKTKFKDNRRAVEMMDVAGEKL...
CDGCKSFLKHTMDPACMFLILKCRCLIVCHHTRYKTCHCIQGRSCEKTKFKDNRRAVEMMEVAGEKL...
```

#2 Infer phylogeny



#3 Reconstruct ancestors

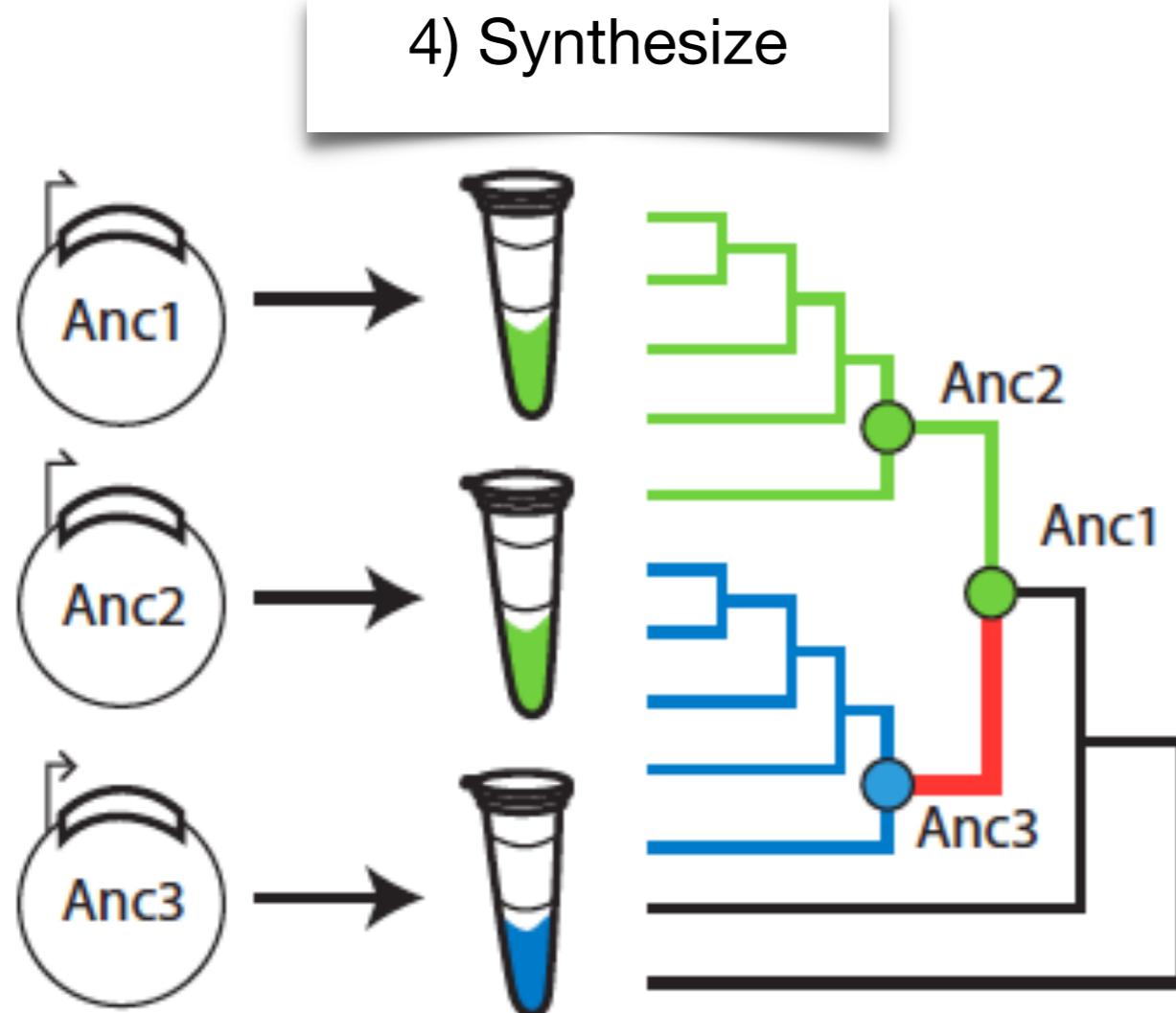


Ancestral sequence

```
CEGCKSFFKKSMDPACMLLILKCKCLIICHHARRYKTCHCIQGRACEKTKFKDNRRAVEMMEVAGEKL...
```

Resurrect *hypothetical* ancestors

AGFRSCHKRRLPGGASWETCSGHKPRELAVFGGRRED Anc1
AGFRSCHKRRLPGGASWETCSGHKPRELAVFGGRRED Anc2
AGFKSCHKRELPGPASFETS SGHKPRELALFGARRED Anc3

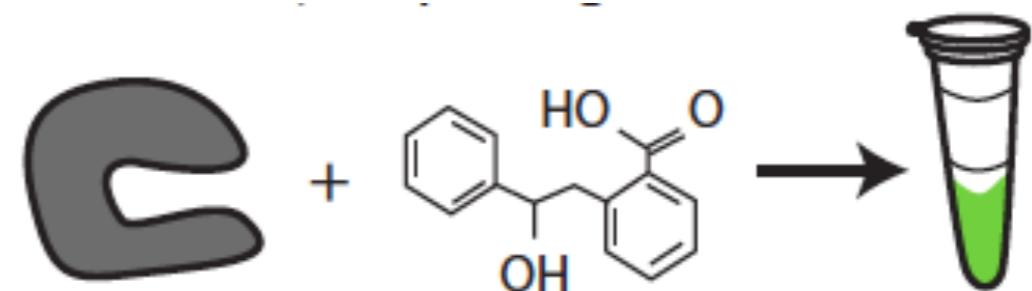


Length, price,
delivery

100-500bp,	\$89, 1-3
501-750bp,	\$149, 1-3
1001-1250,	\$209, 1-3
1251-1500bp,	\$249, 4-7
1501-1750bp,	\$289, 4-7
1751-2000bp,	\$329, 4-7

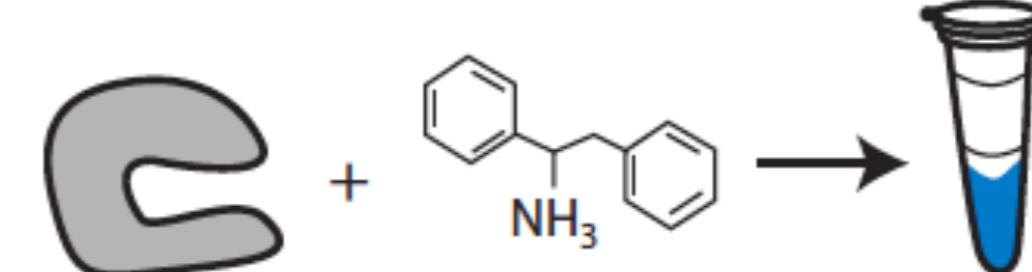
Characterise in a test tube

- Structure



- Function

Human enzyme paralog 2



- Localisation

- Interactions

- Induction

None

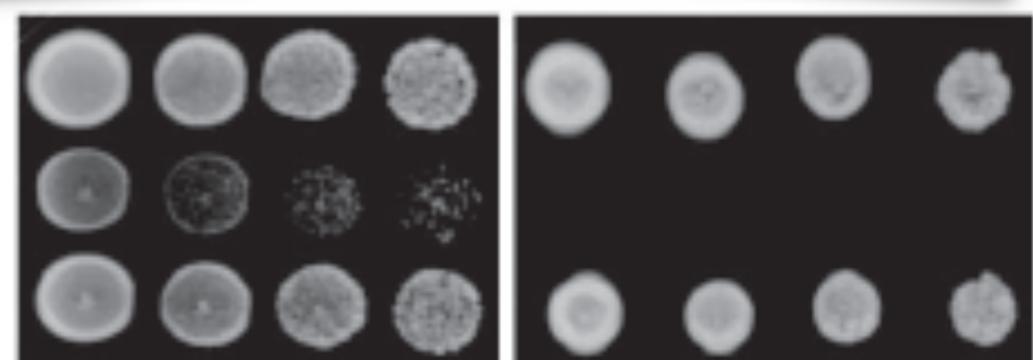
None

Anc.16

WT

16Δ

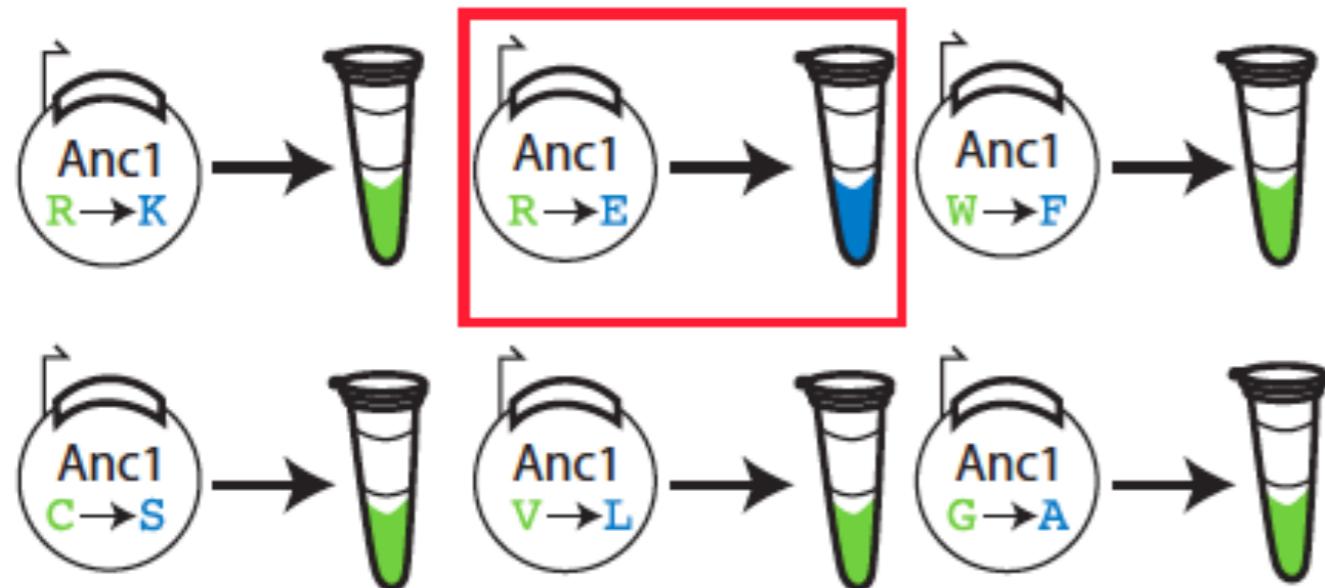
16Δ



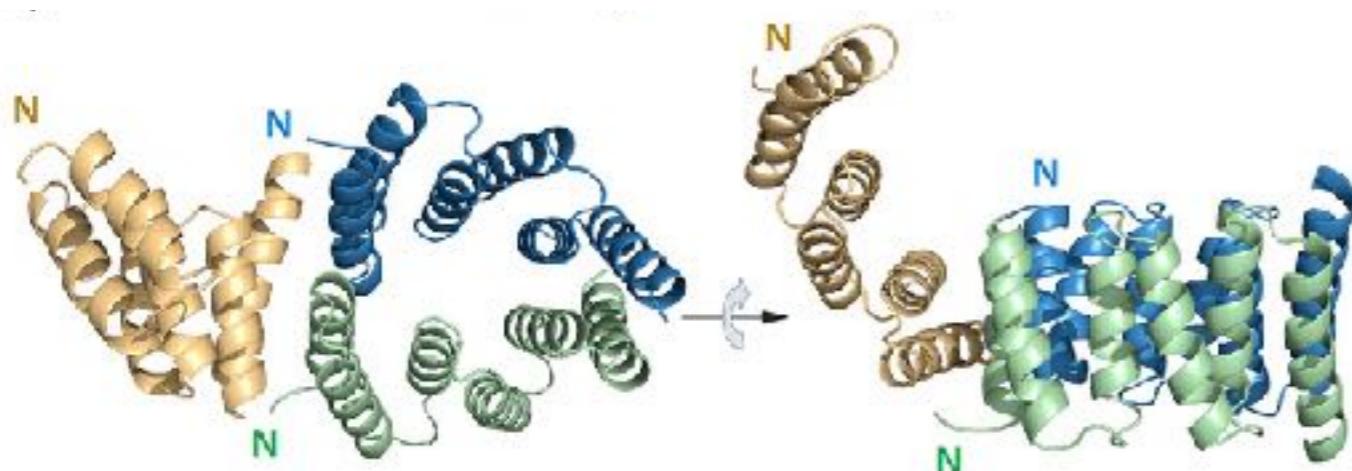
Characterise in cells

Some of the topics studied

AGFRSCHKRR|LPGGASWETCSGHKPRELAVFGGRRED Anc1
AGFKSCHKRE|LPGPASFETSSGHKPRELALFGARRED Anc3



- Protein-protein interactions, epistasis,
- Hormone-receptor co-evolution/divergence
- Evolution of complex molecular machineries
- Ancient enzyme activity
- Evo. of CNVs, order of genes
- What else?

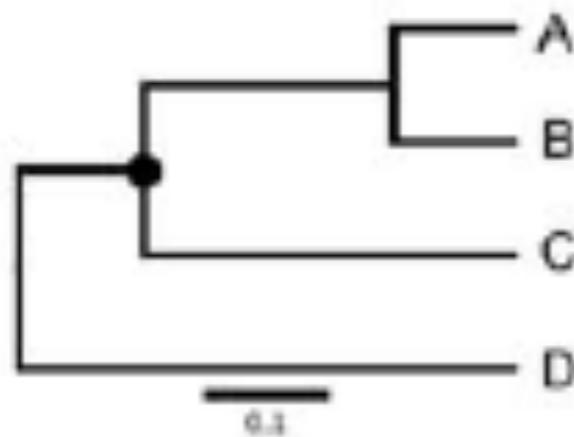


Ancestral genes resurrected

Extant genes	Ancestral gene resurrected	Approximate age of ancestor (years)	Inference method	Refs
Digestive ribonucleases	Ancestral orthologue in LCA of buffalo and ox	5–10 million	Parsimony	38
<i>L1</i> retroposons in mouse	Ancestral parologue [†] in mouse genome	"several million"	Consensus	7
Digestive ribonucleases	Ancestral orthologue in LCA of artiodactyls	~40 million	Parsimony	22
Chymase proteases	Ancestral orthologue in LCA of mammals	~80 million	Parsimony	43
<i>Tc1/mariner</i> transposons	Ancestral parologue in genomes of 8 salmonids	~10 million	Consensus	6
Immune RNases	Ancestral orthologue in LCA of 'higher primates'	31 million	Parsimony, Bayesian distance	27
Pax [§] transcription factors	Ancestral parologue (older than the protostome–deuterostome [¶] ancestor)	600–1,000 million	Bayesian distance	26
Vertebrate rhodopsins	Ancestral orthologue in LCA of archosaurs [#]	240 million	Maximum likelihood	17
Vertebrate short-wave rhodopsins	Ancestral orthologue in LCA of bony vertebrates	>400 million	Maximum likelihood	44
Steroid hormone receptors	Ancestral parologue (older than the protostome–deuterostome ancestor)	600–1,000 million	Maximum likelihood	18
Elongation factor EF-Tu	Ancestral orthologue in LCA of eubacteria	>1 billion	Maximum likelihood	20

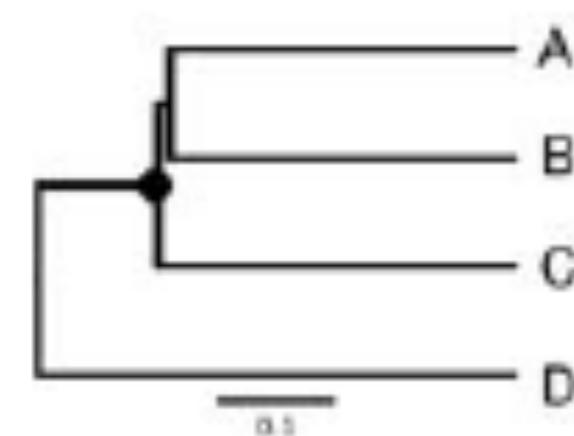
- Time interval: 5 million years - 3.5 billion years

Phylogenetic uncertainty: MP vs ML vs Bayesian?



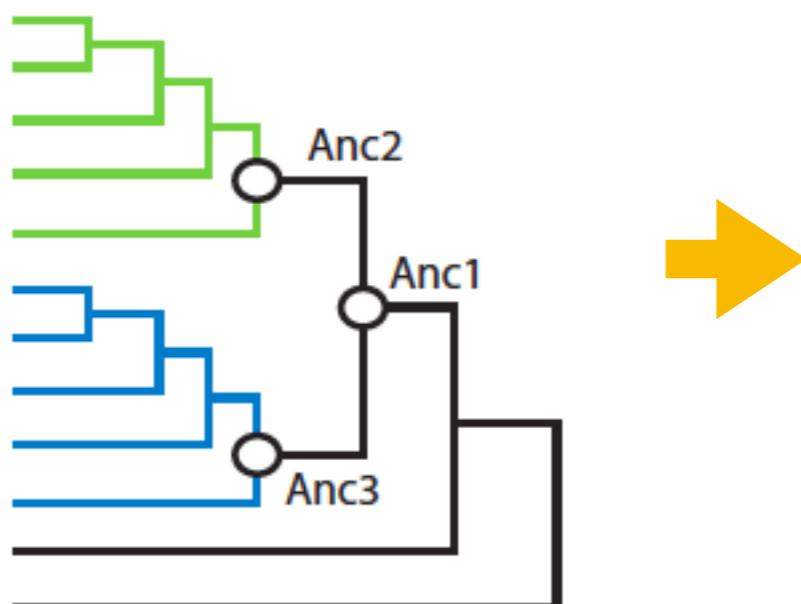
$$P(t|D, \hat{M}, \hat{\theta})$$

- Maximum Parsimony?



$$P(t|D, \hat{M}, \hat{\theta})$$

- Maximum Likelihood?



AGFRSCHKRRRLPGGASWETCSGHKPRELAVFGGRRED Anc1
 AGFRSCHKRRRLPGGASWETCSGHKPRELAVFGGRRED Anc2
 AGFKSCHKRELPGPASFETSSGHKPRELALFGARRED Anc3

?

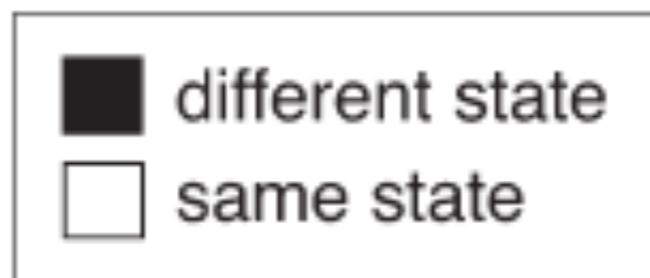
?

?

?

Dealing w/ uncertainty: site-based statistics

A



ultrametric
four taxon



non-ultrametric
four taxon



alcohol
dehydrogenase



steroid-
hormone
receptor

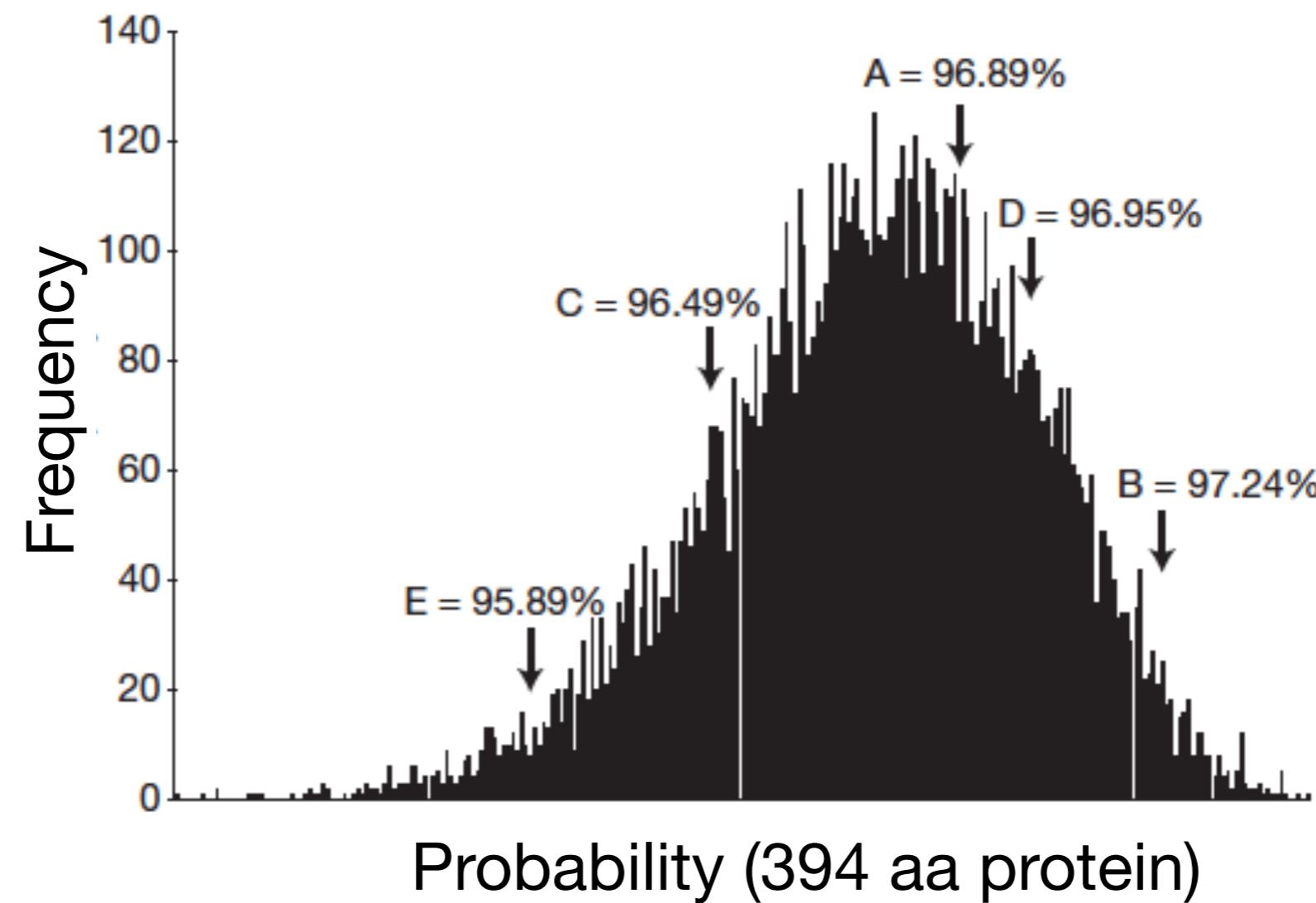


green
fluorescent
protein



Tu family
elongation
factors

The aim is to generate an ancestral 'G' that reflects the ancestral 'P'
- not the exact sequence of the ancestral 'G'



'G': genotype

'P': genotype

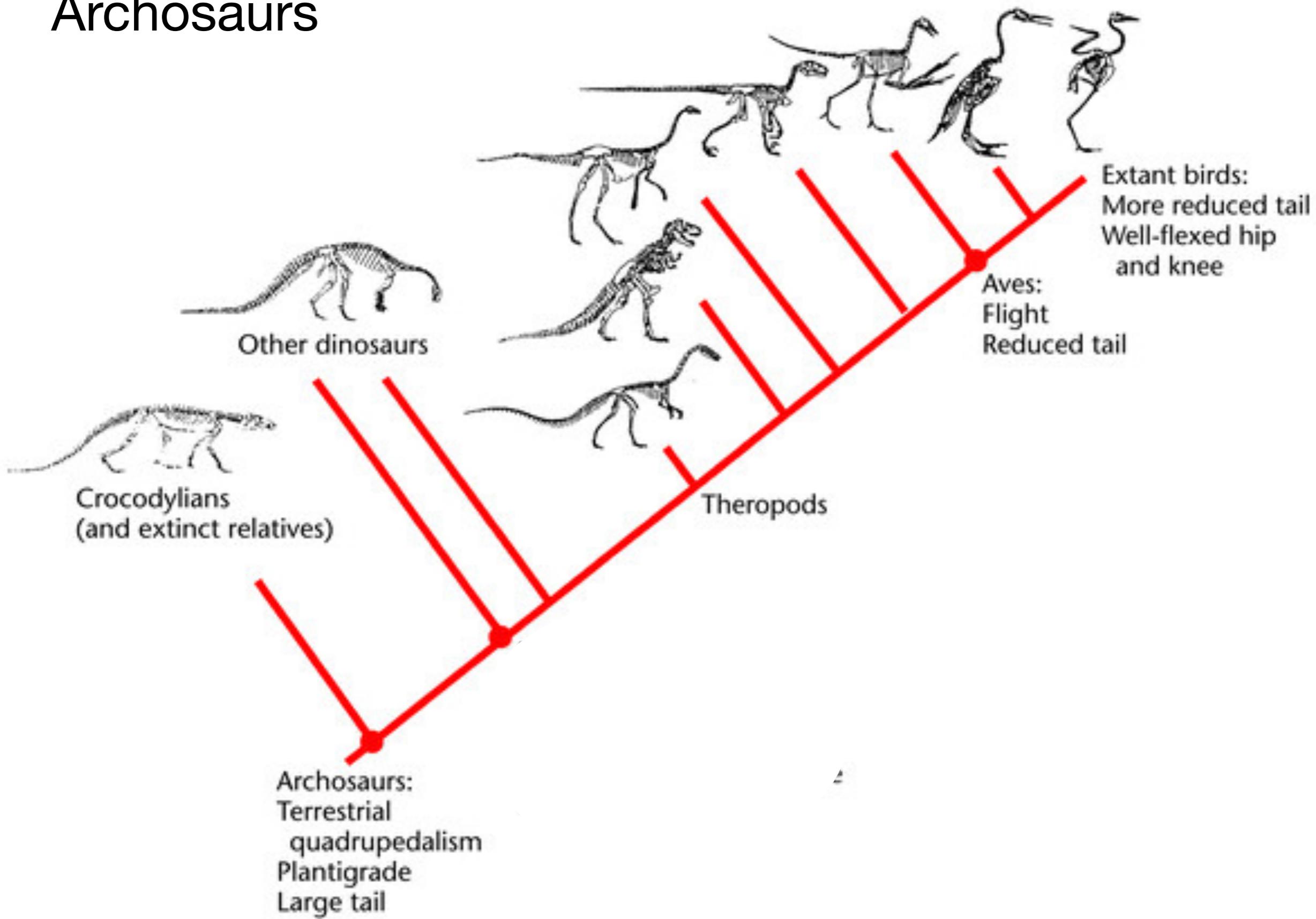
Gaucher et al. 2008

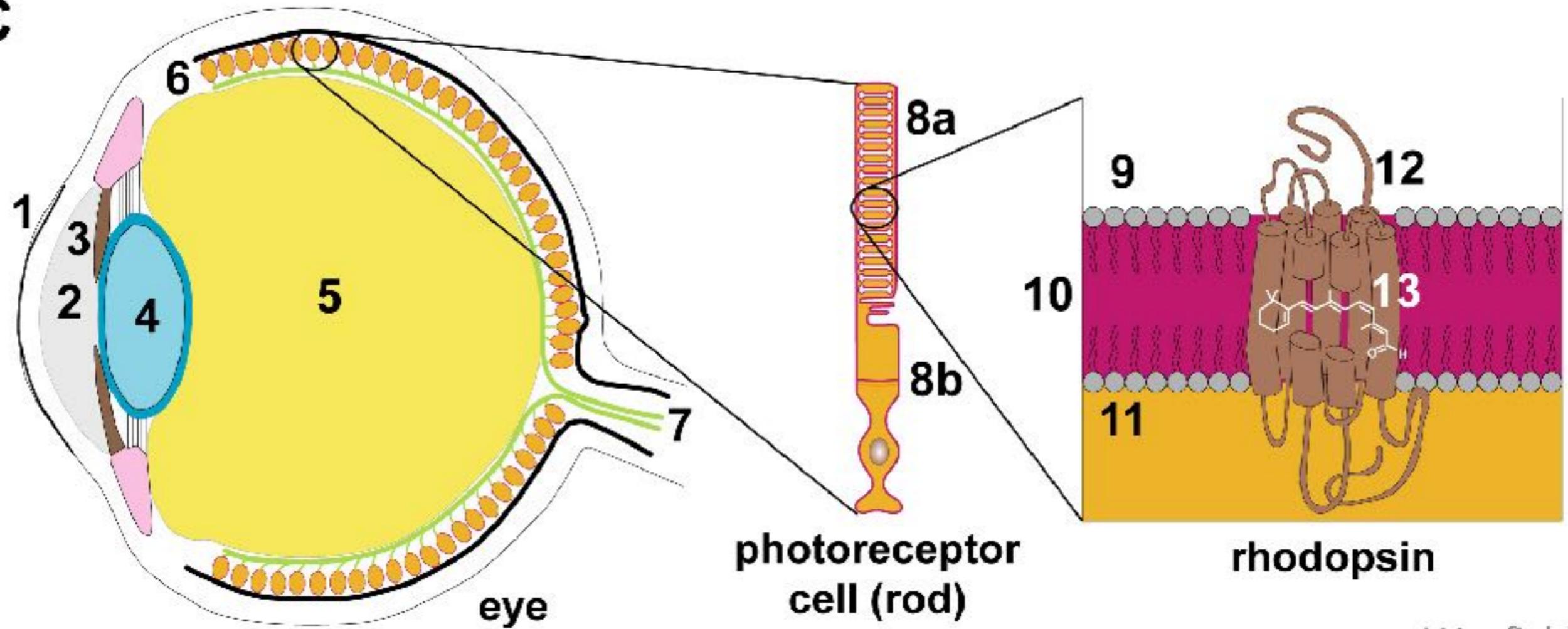
Examples from 3 studies

I - Were dinosaurs active during the night?

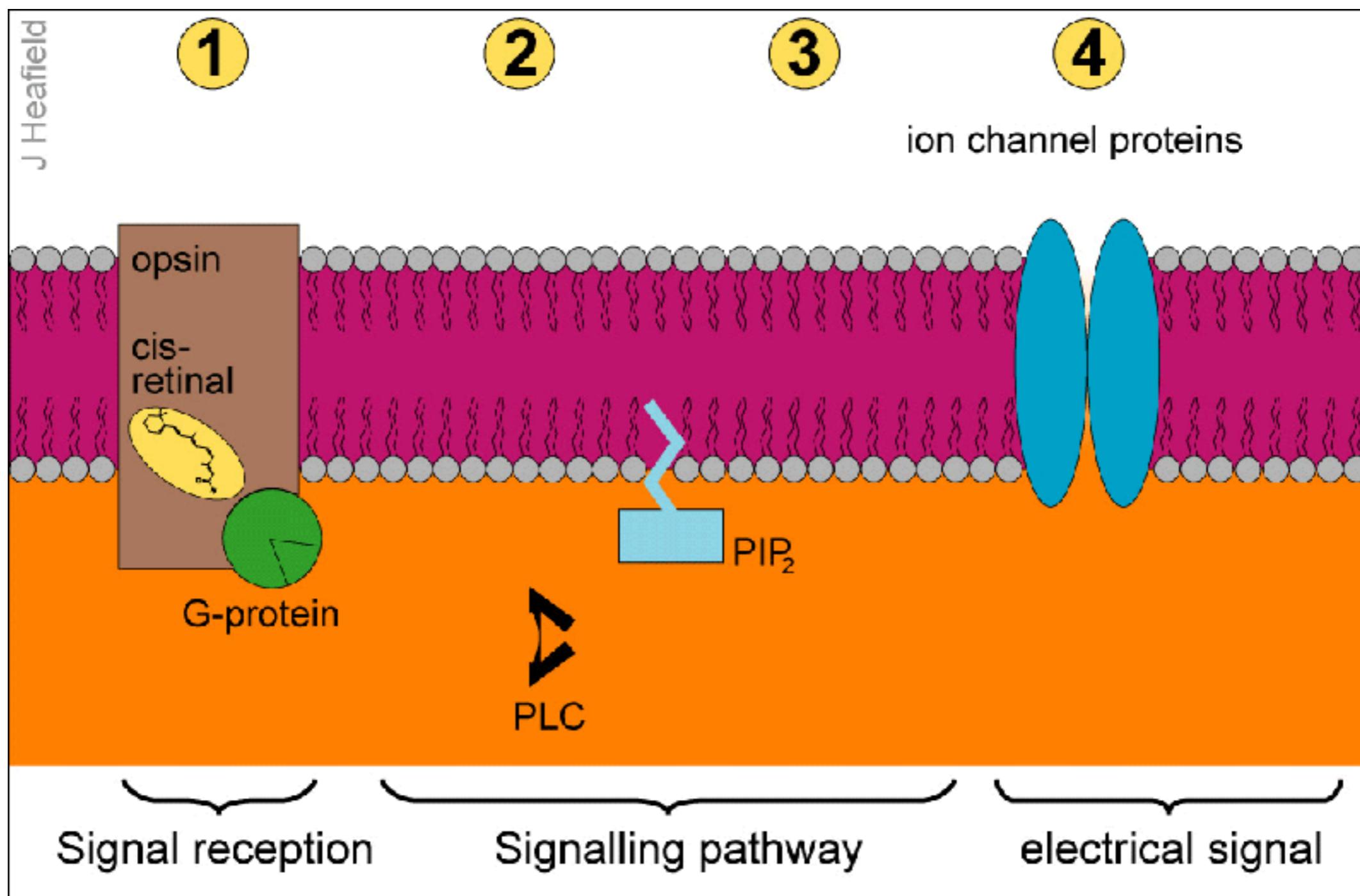


Archosaurs



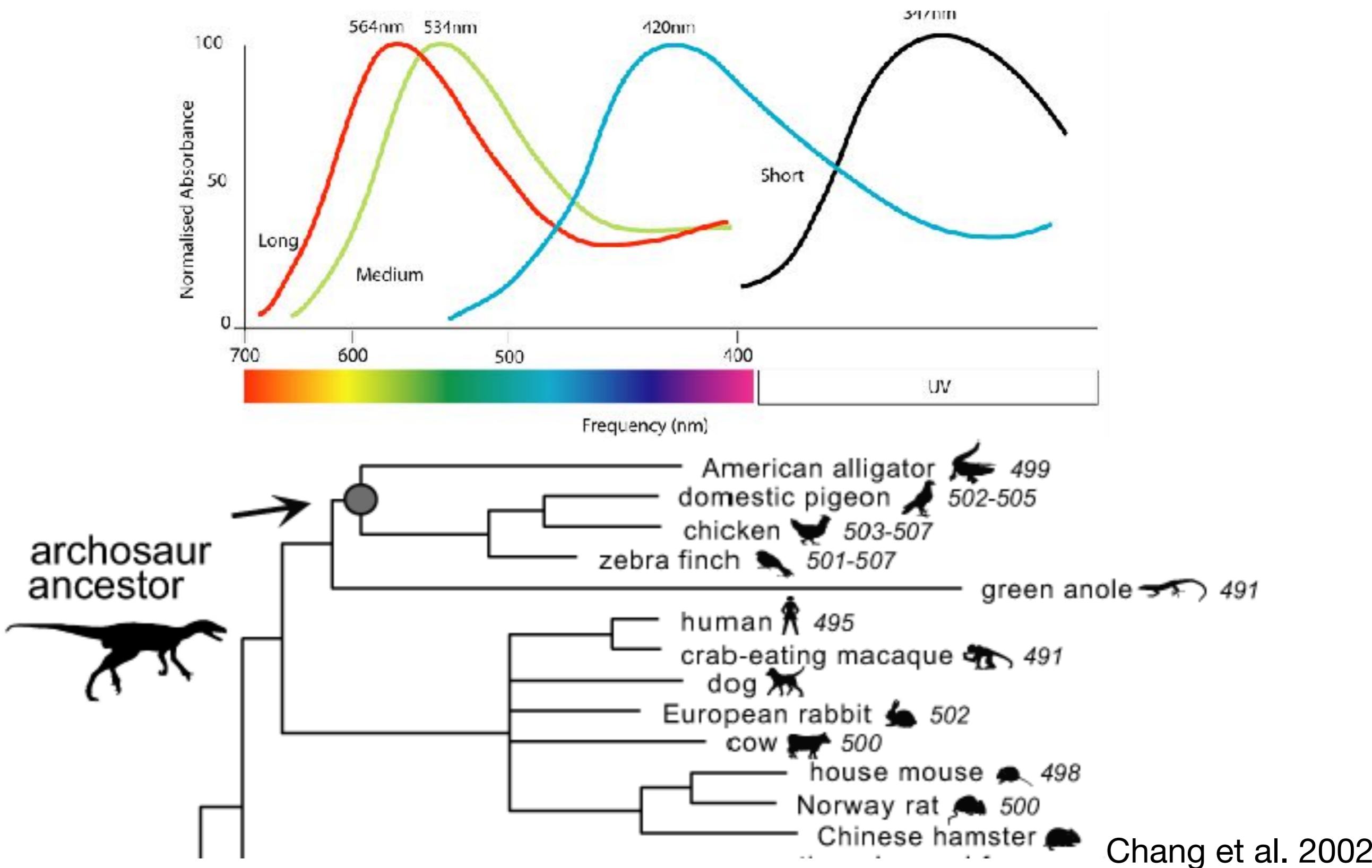
C

Rhodopsin & dim light (=night) vision

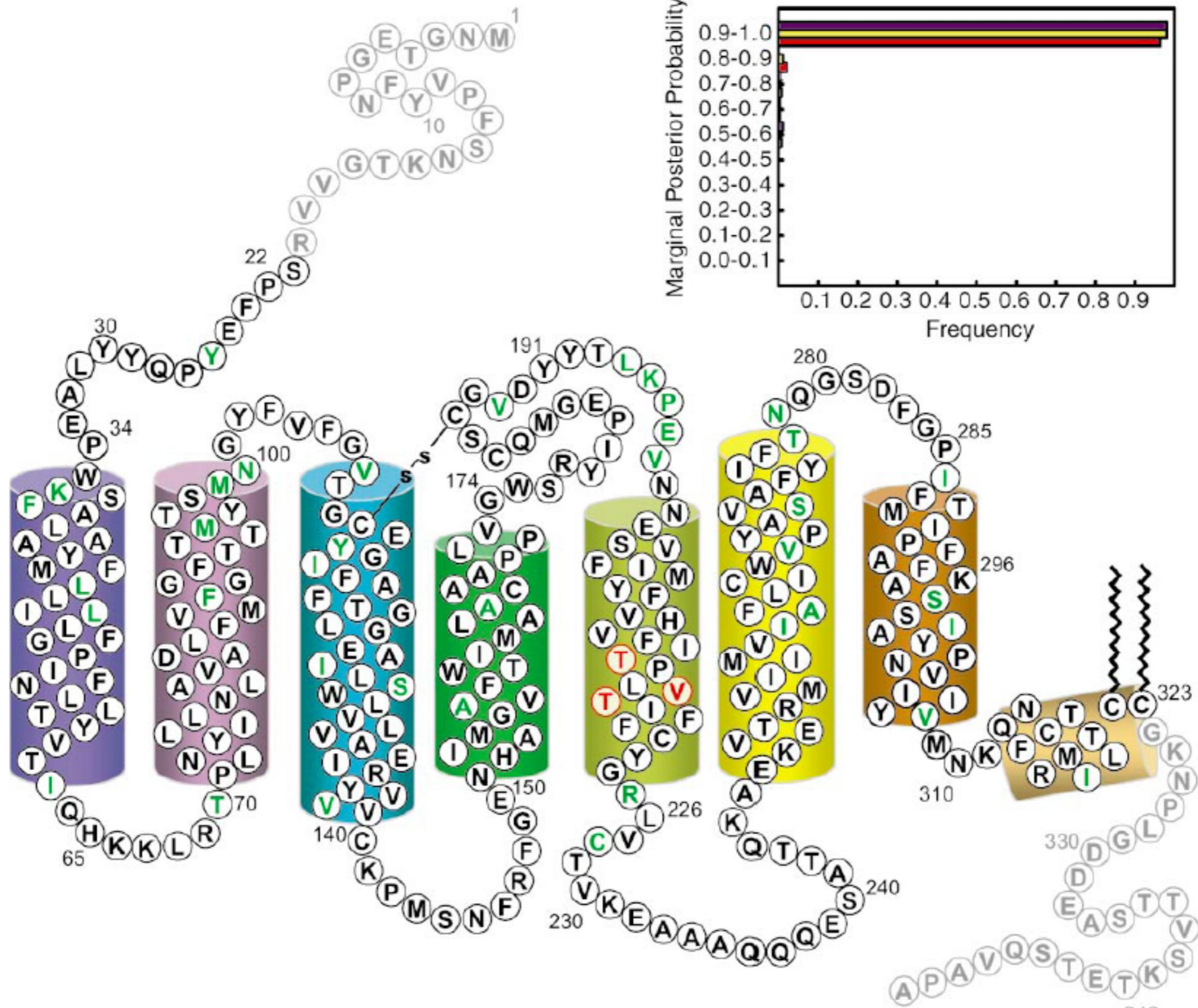


Working hypothesis:

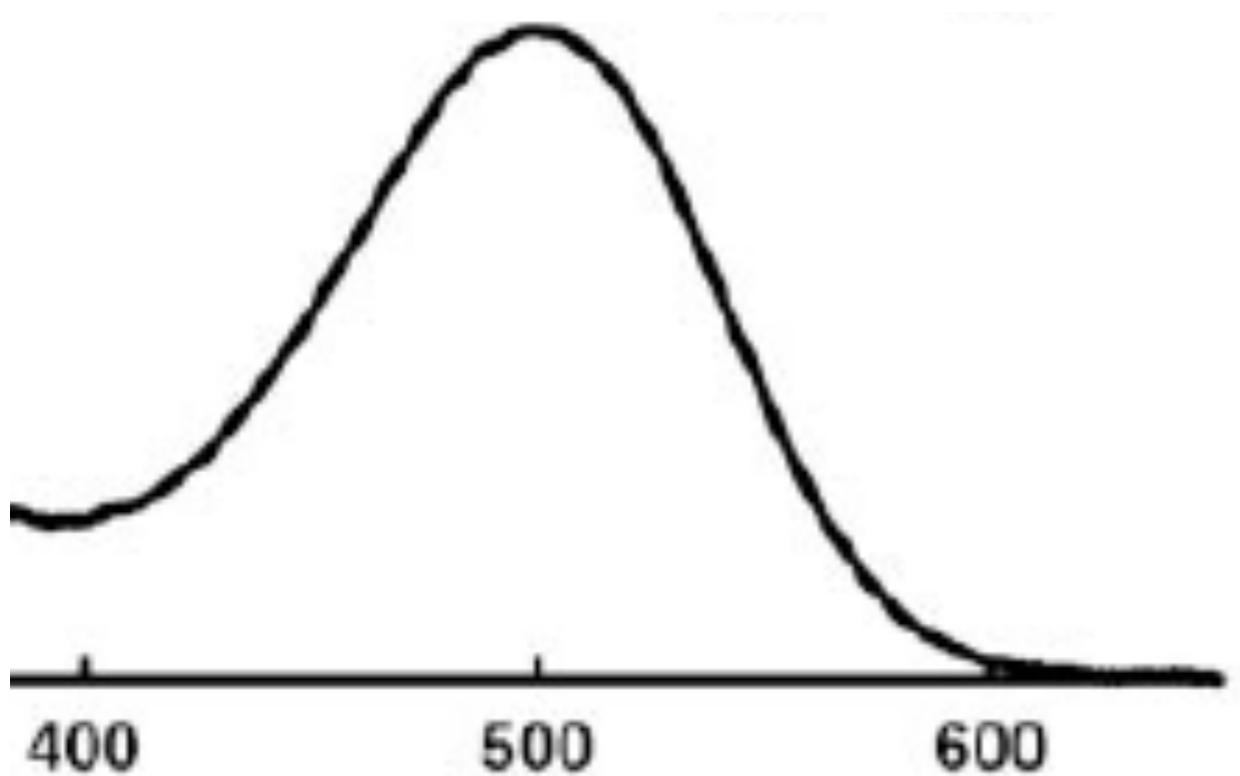
Reconstruct ancestral state(s) & observe light response



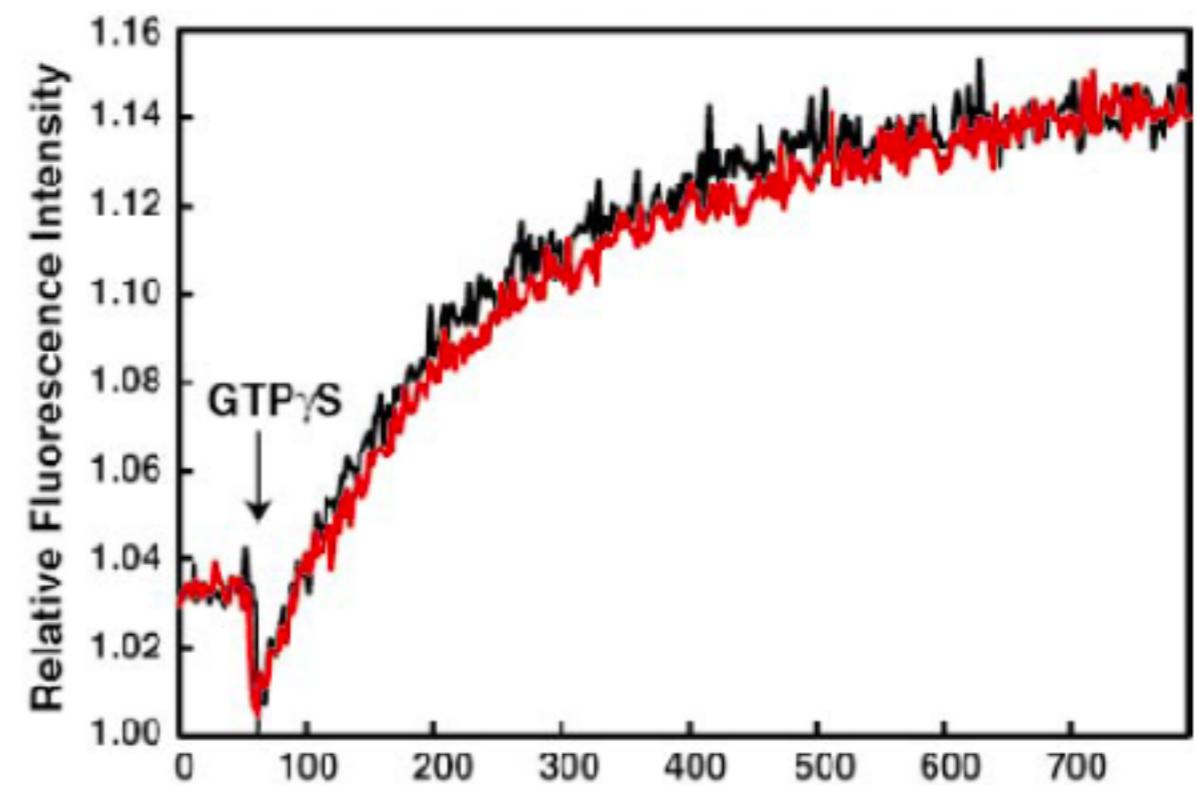
Chang et al. 2002



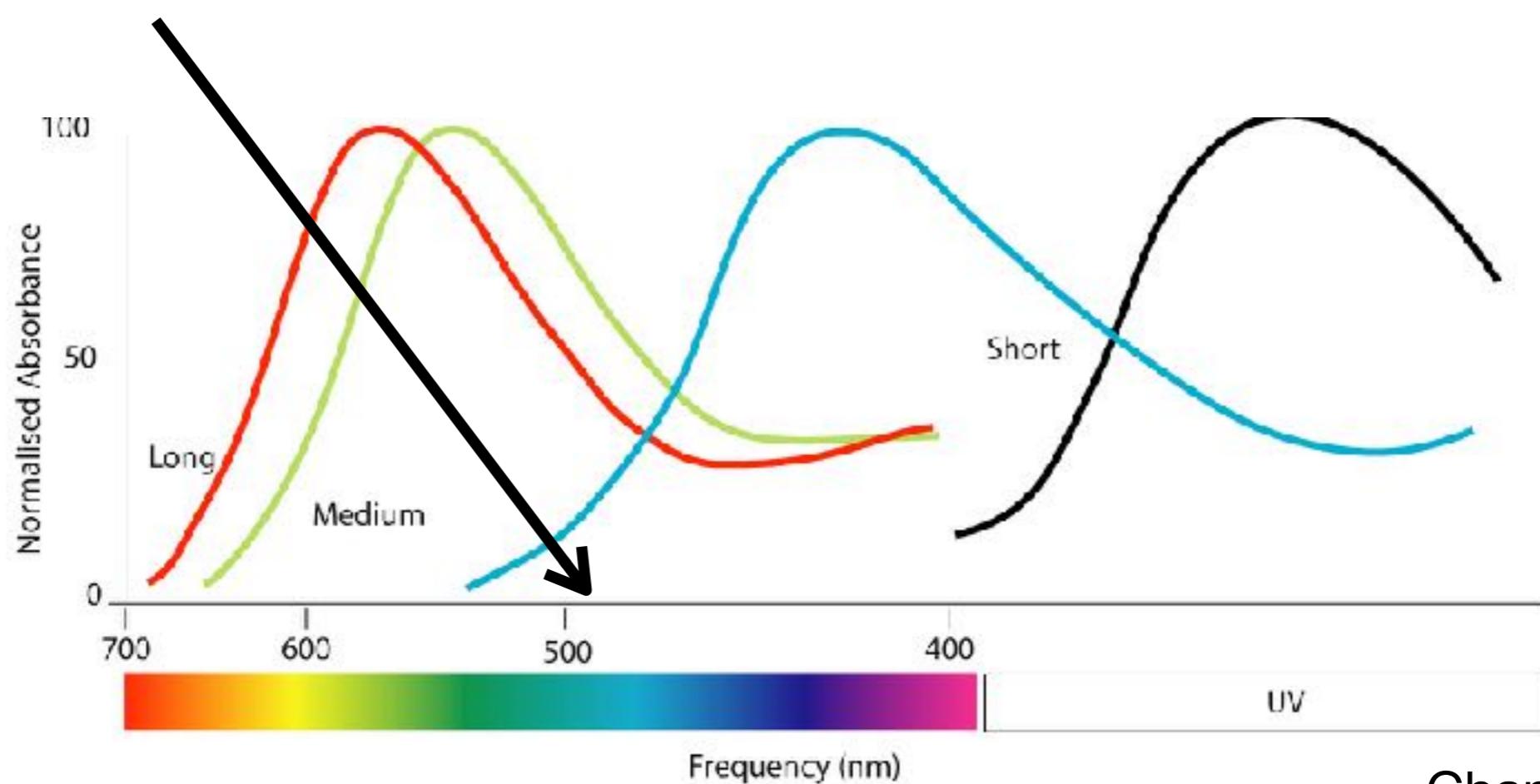
Chang et al. 2002



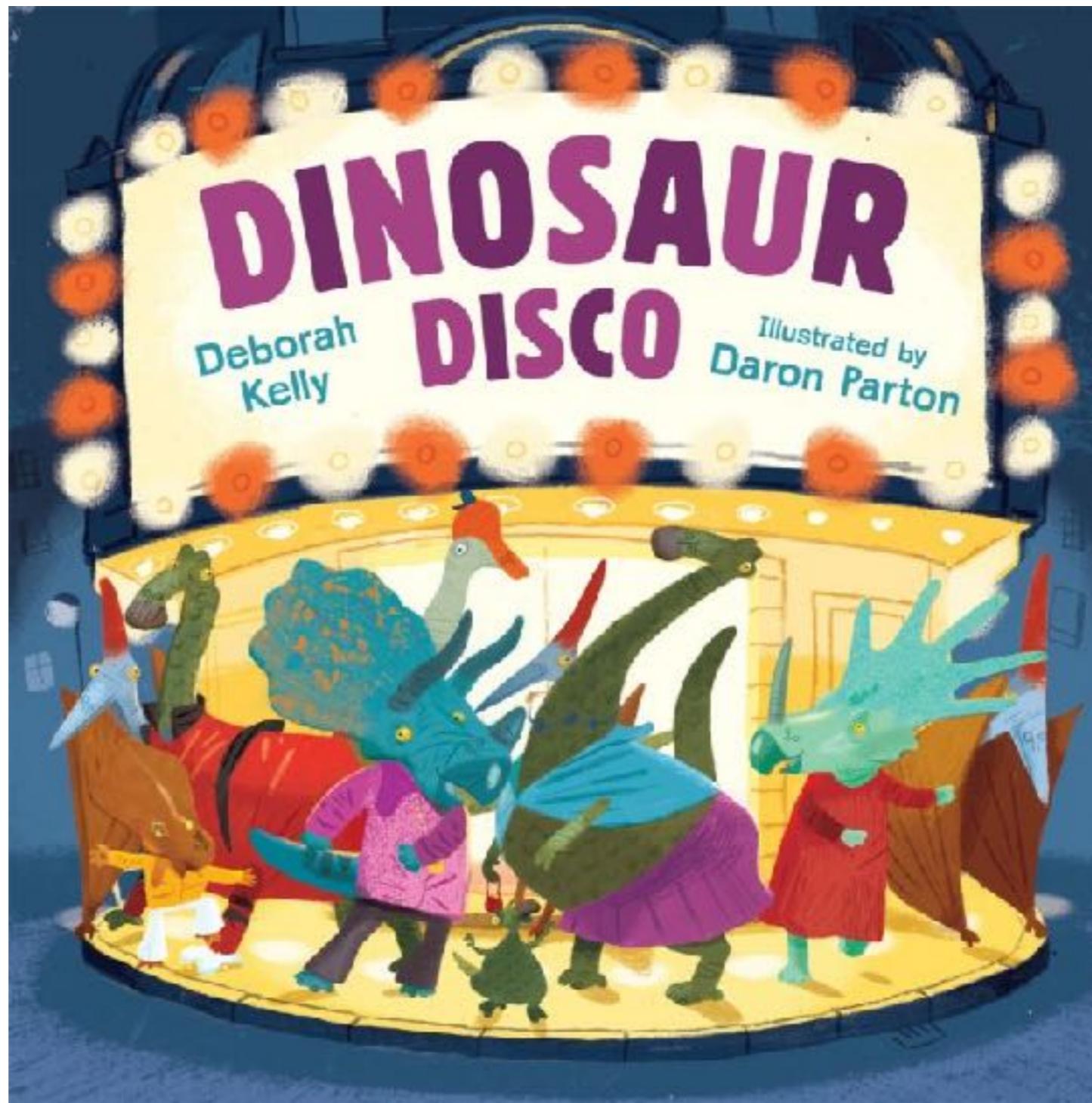
Absorption max (508nm)



Addition of GTP & transducin
activation by ancestral rhodopsin



Conclusion?
Were dinosaurs active during the night?



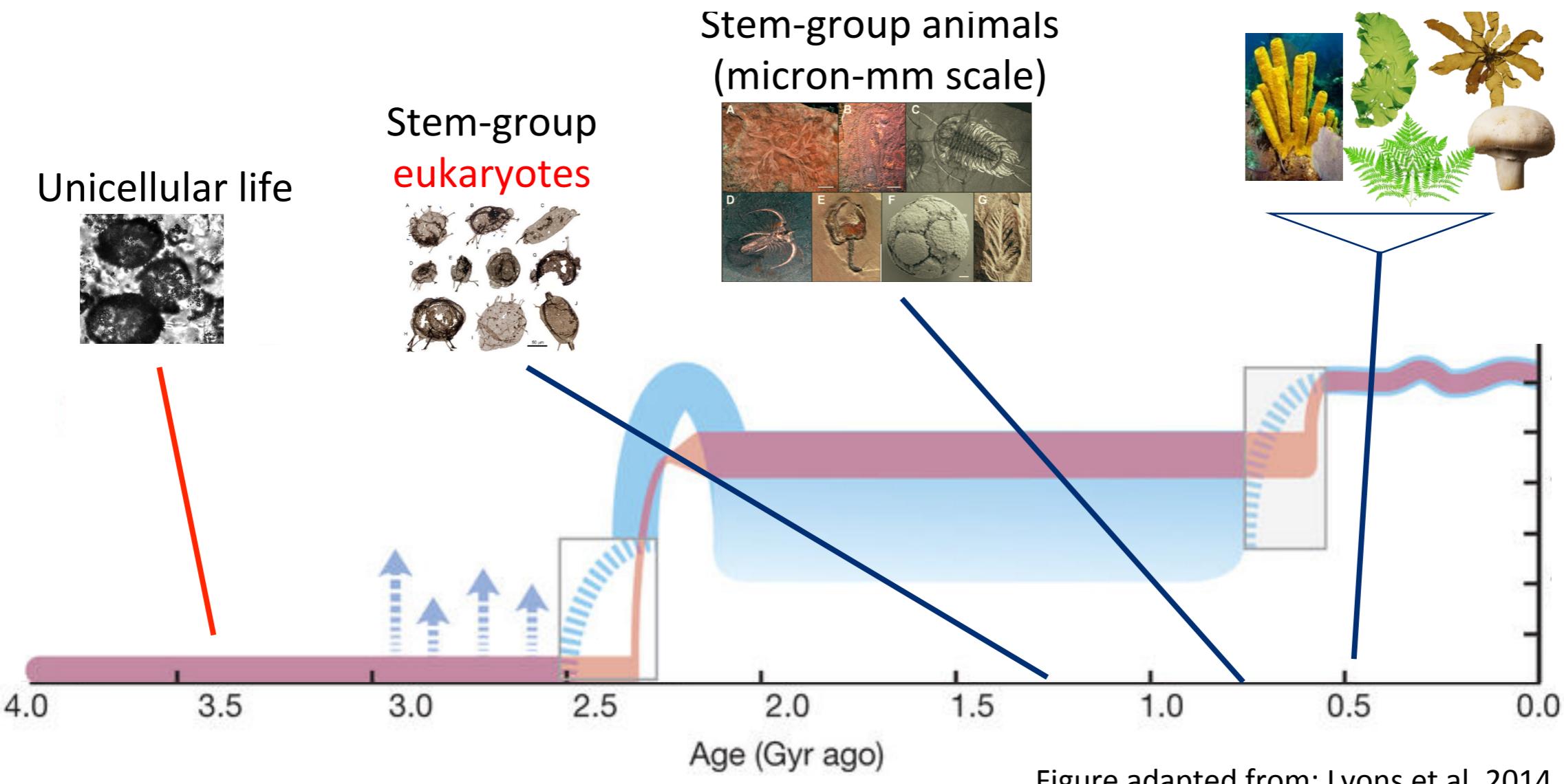
II - Testing adaption of ancient life to cooling ocean temperatures

LETTERS

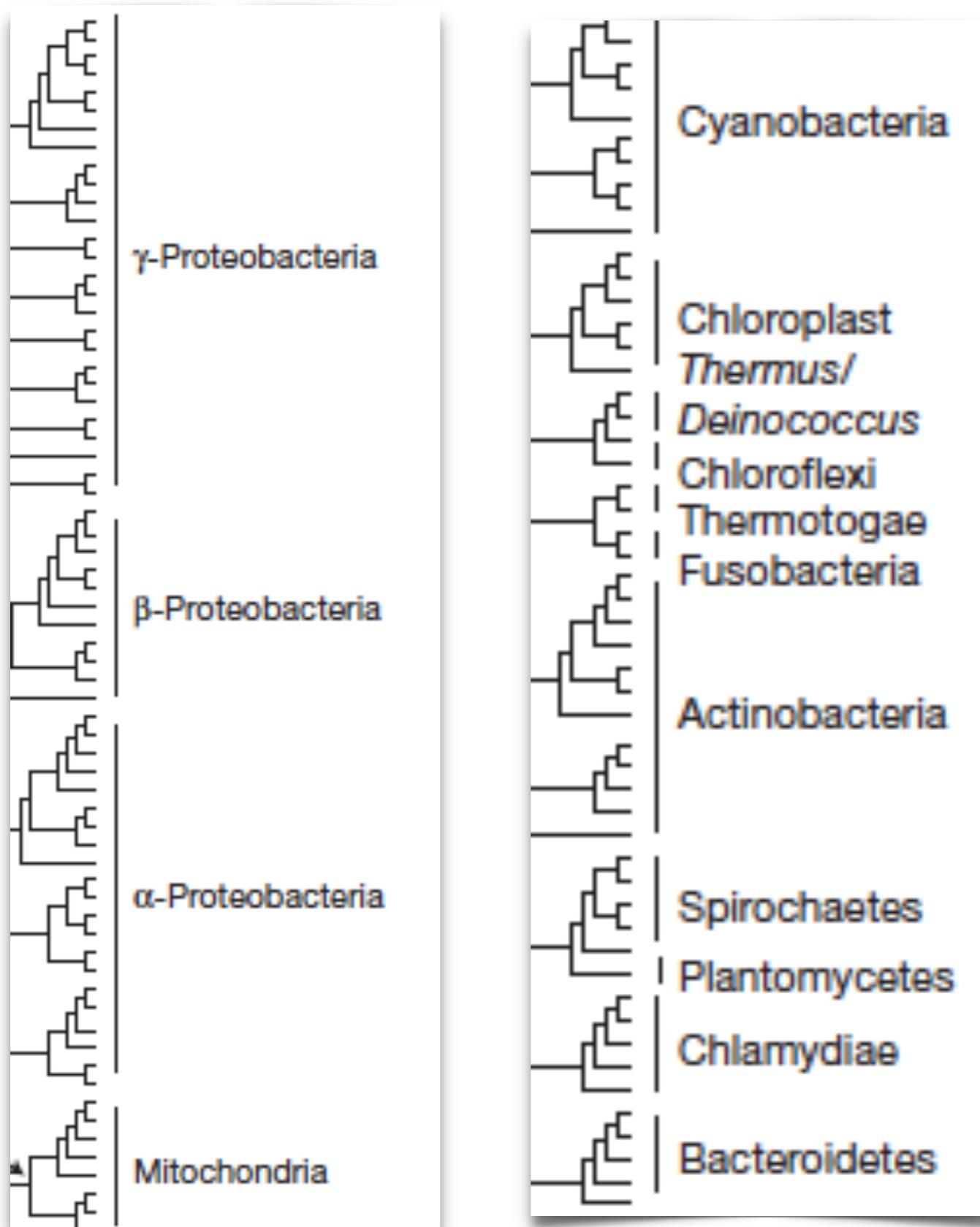
Palaeotemperature trend for Precambrian life inferred from resurrected proteins

Eric A. Gaucher¹, Sridhar Govindarajan² & Omjoy K. Ganesh³

Evolution of life on earth - major events

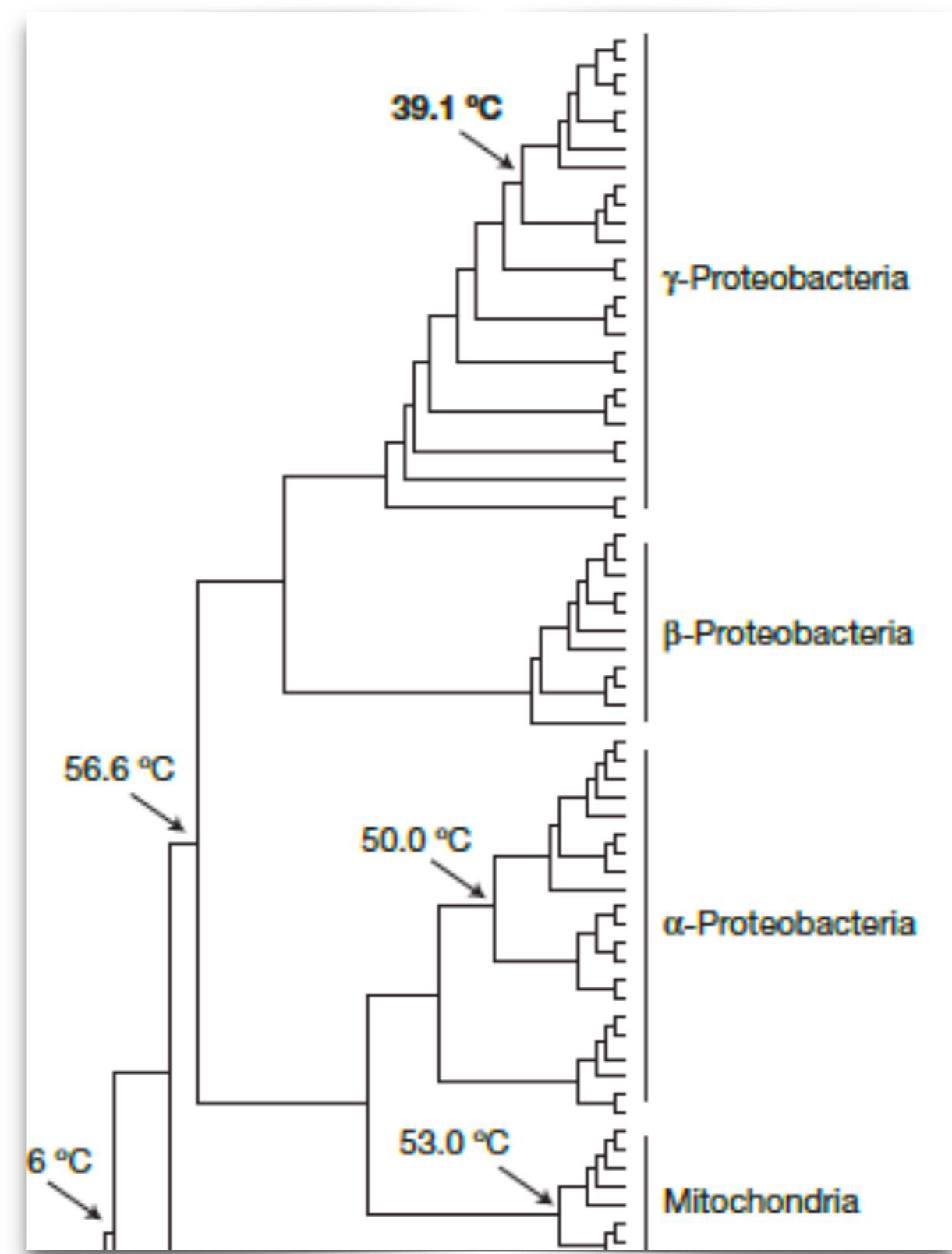
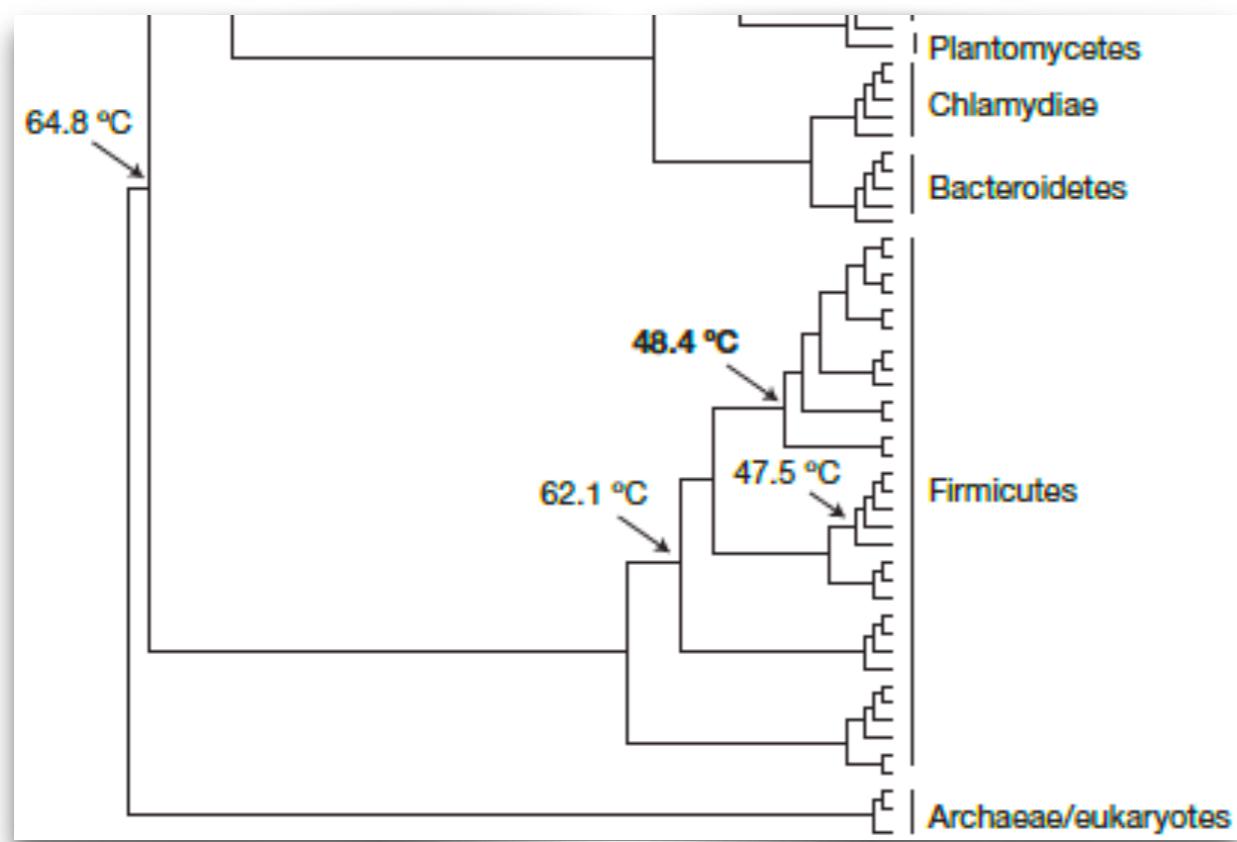


EF-Tu thermal stability ~ temperature of the environment

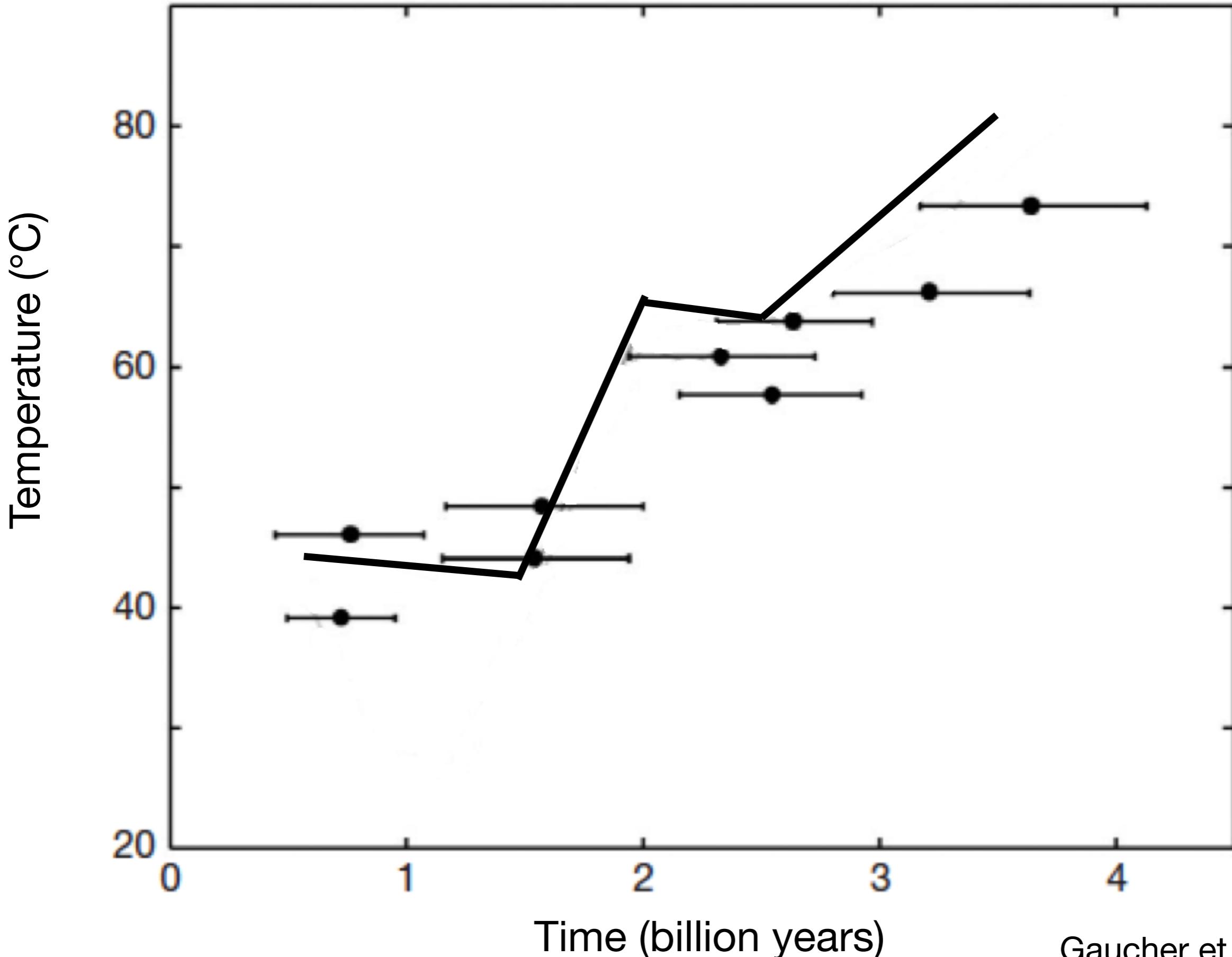


- Hypothesis?

Reconstruction of hypothetical ancestral seq.s & testing their thermal stability

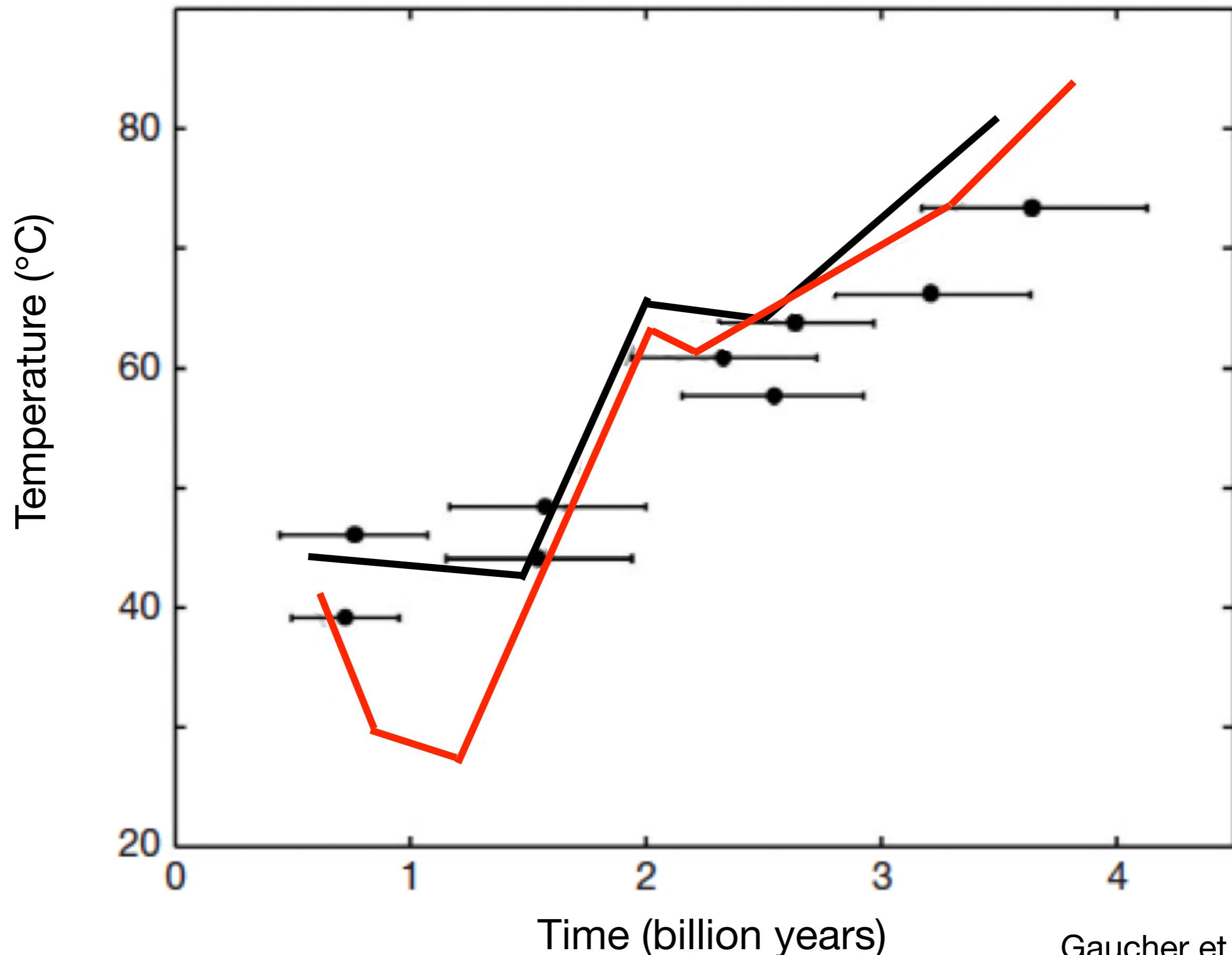


Ancient life adapted to decreasing temperatures



Gaucher et al. 2008

Ancient life adapted to decreasing ocean temperatures



Gaucher et al. 2008

III - Evolution of complex molecular machines

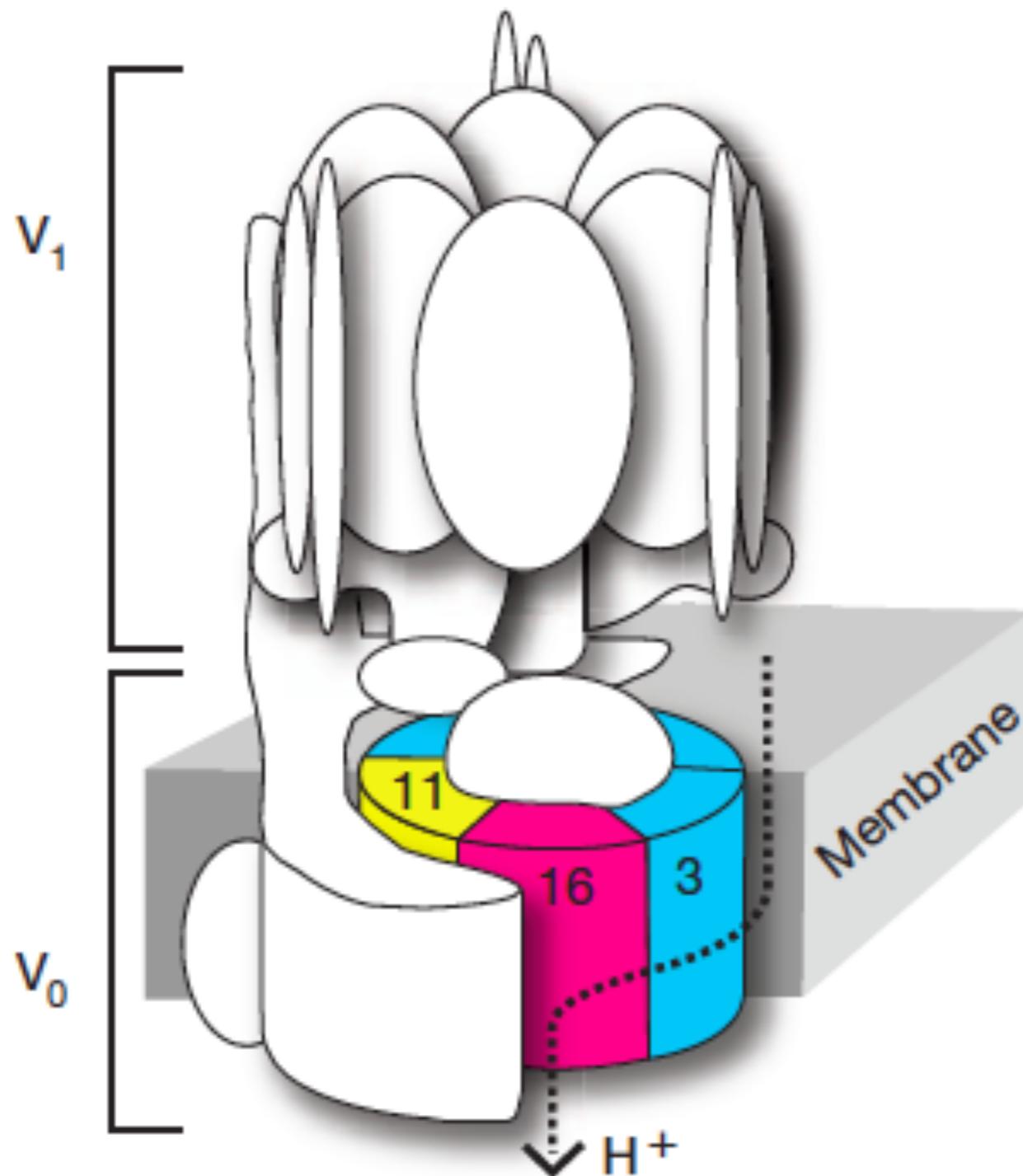
LETTER

doi:10.1038/nature10724

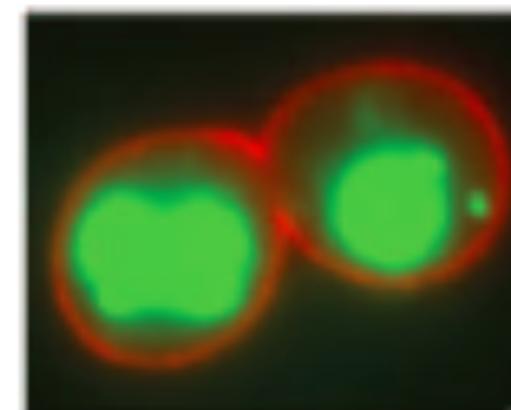
Evolution of increased complexity in a molecular machine

Gregory C. Finnigan^{1*}, Victor Hanson-Smith^{2,3*}, Tom H. Stevens¹ & Joseph W. Thornton^{2,4,5}

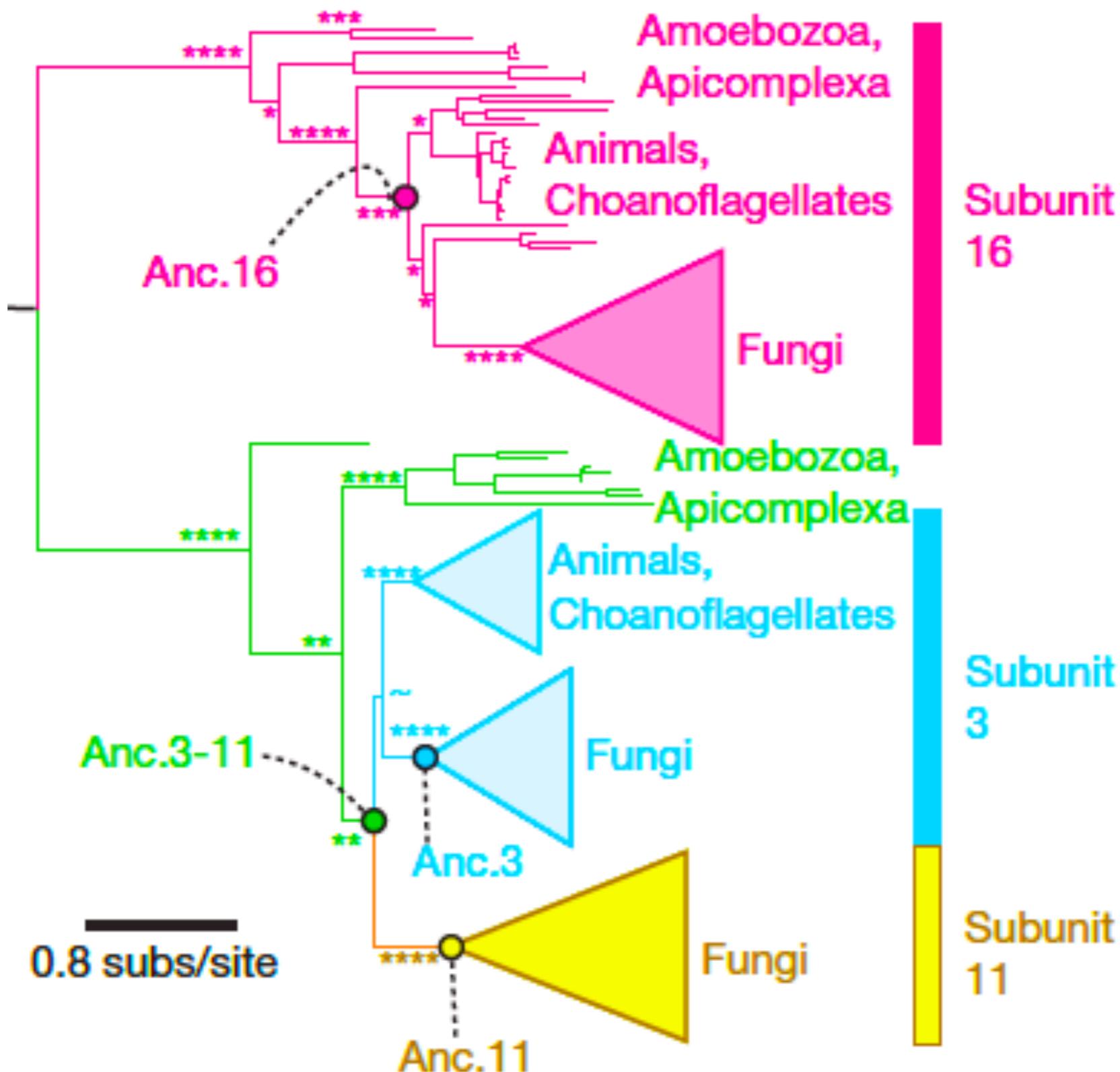
Understanding complex machinery evolution: evolution of eukaryotic V-ATPase subunits



WT

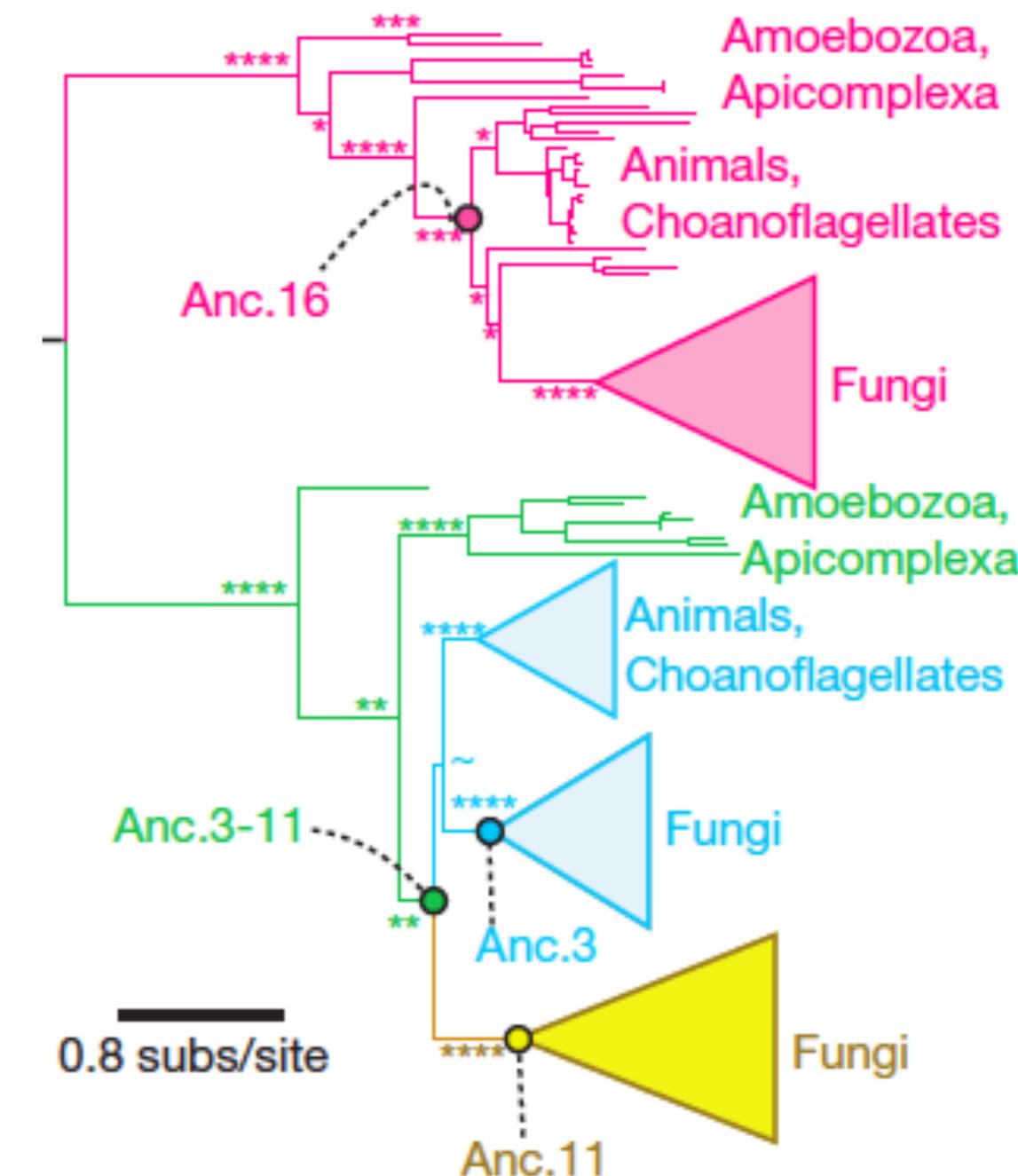
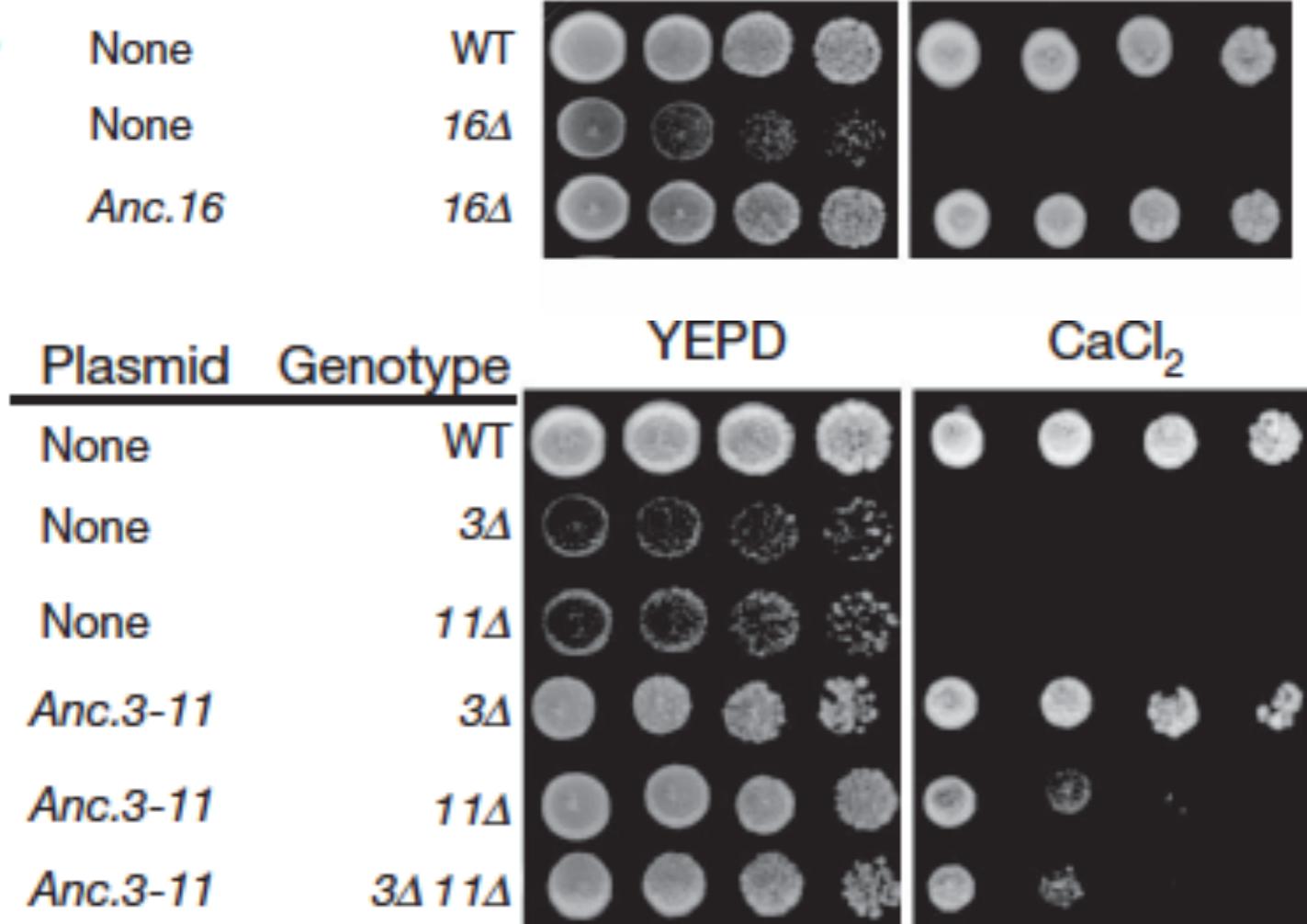


Subunits 11, 3, and 16 - amplification of a common ancestor

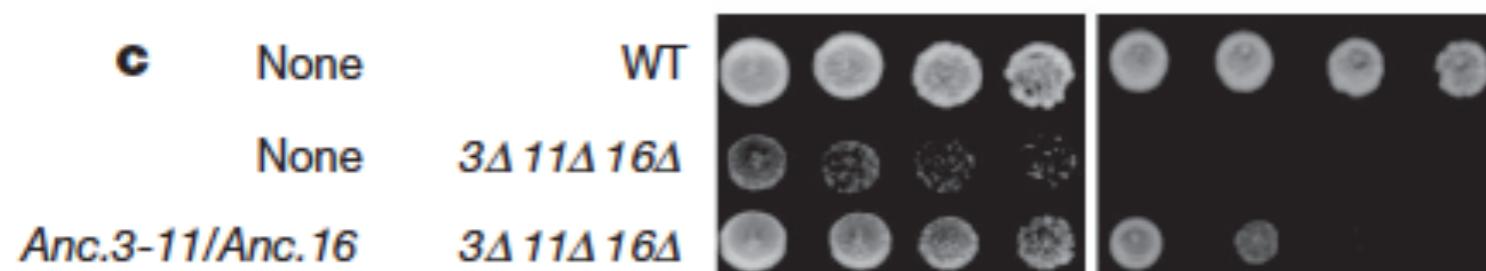


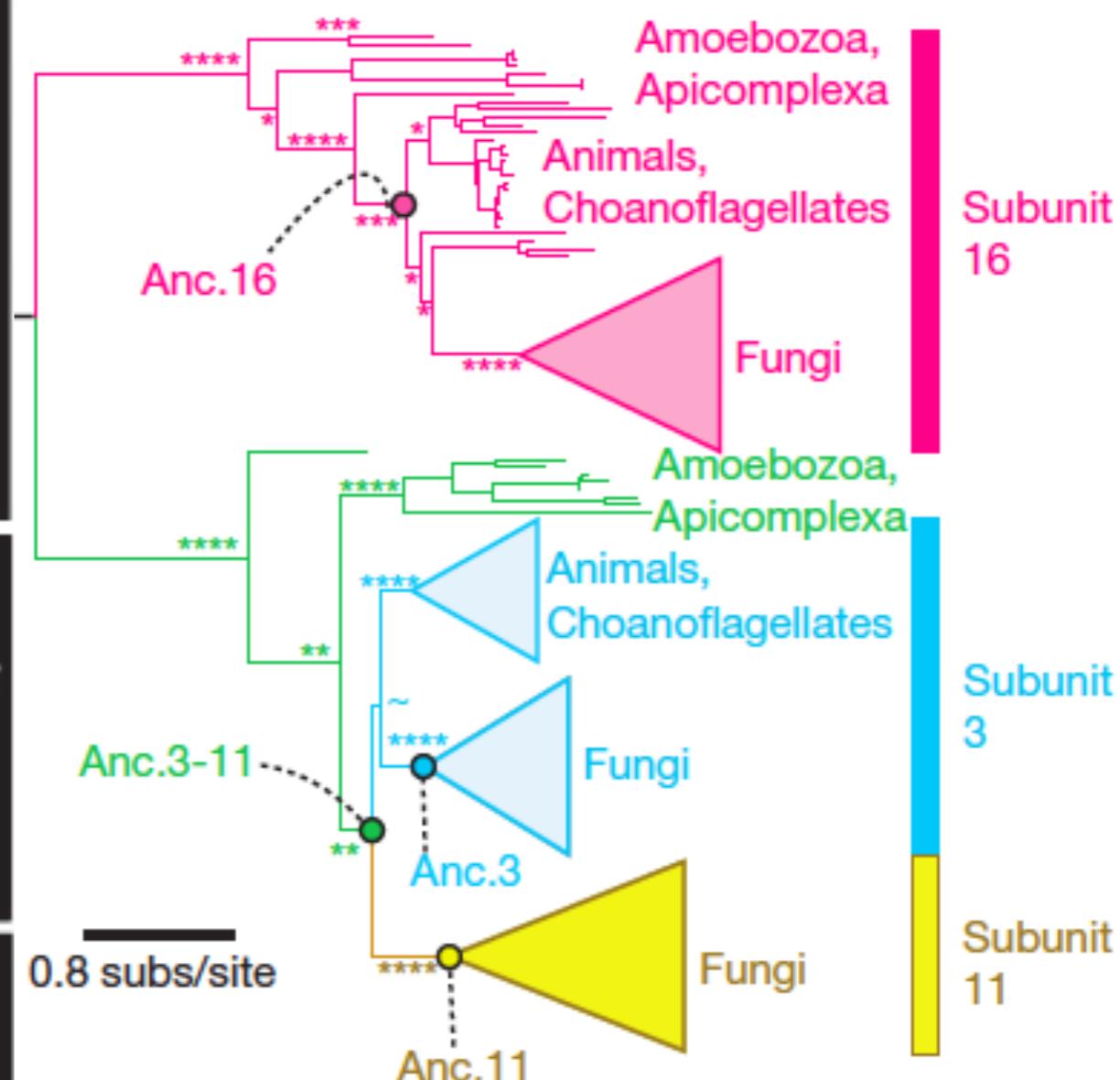
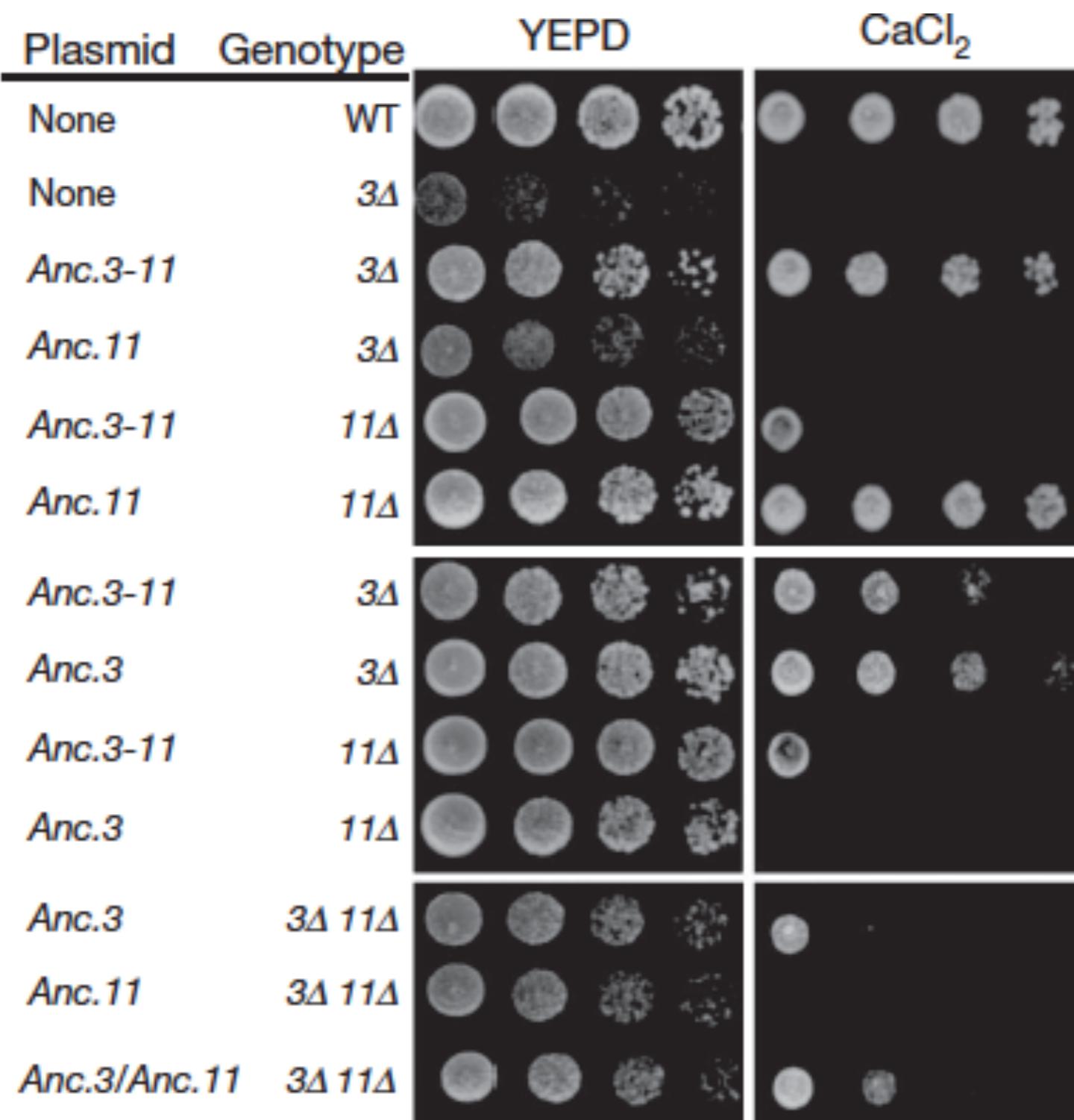
Resurrection of the ancestral sequence(s)/function of the modern duplicates

b

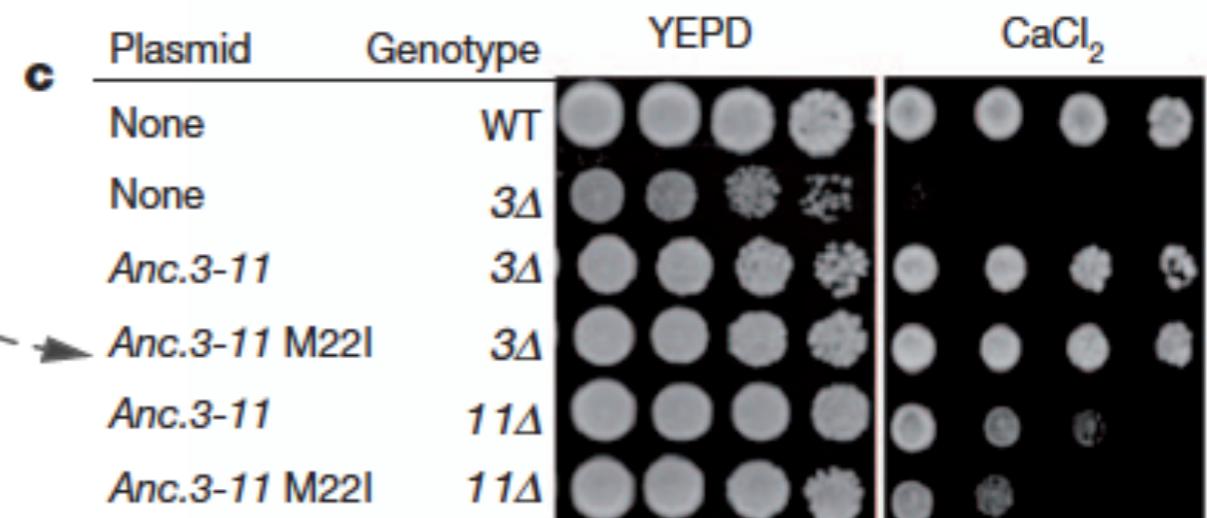
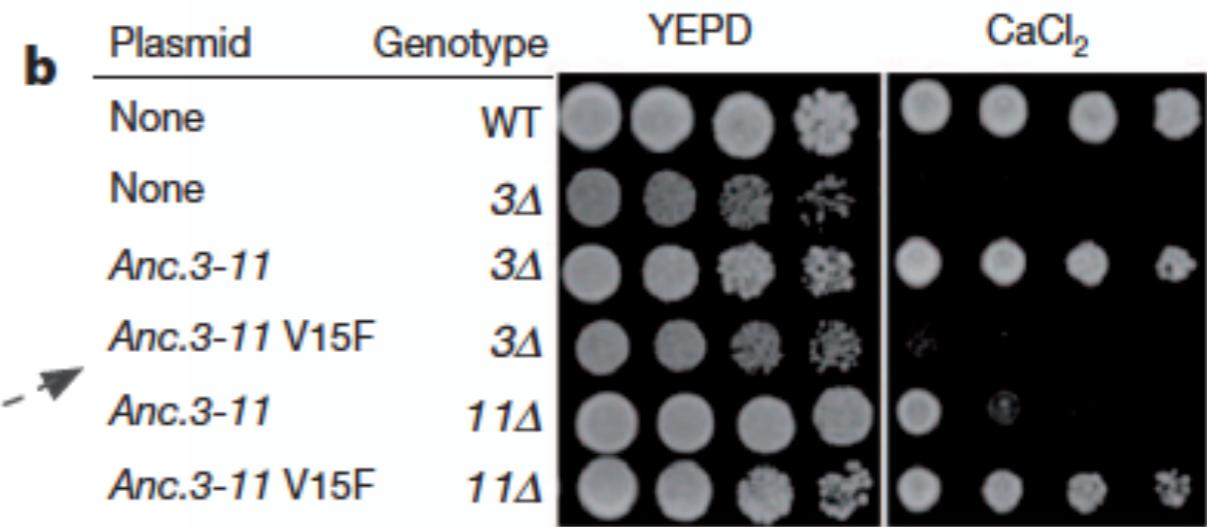
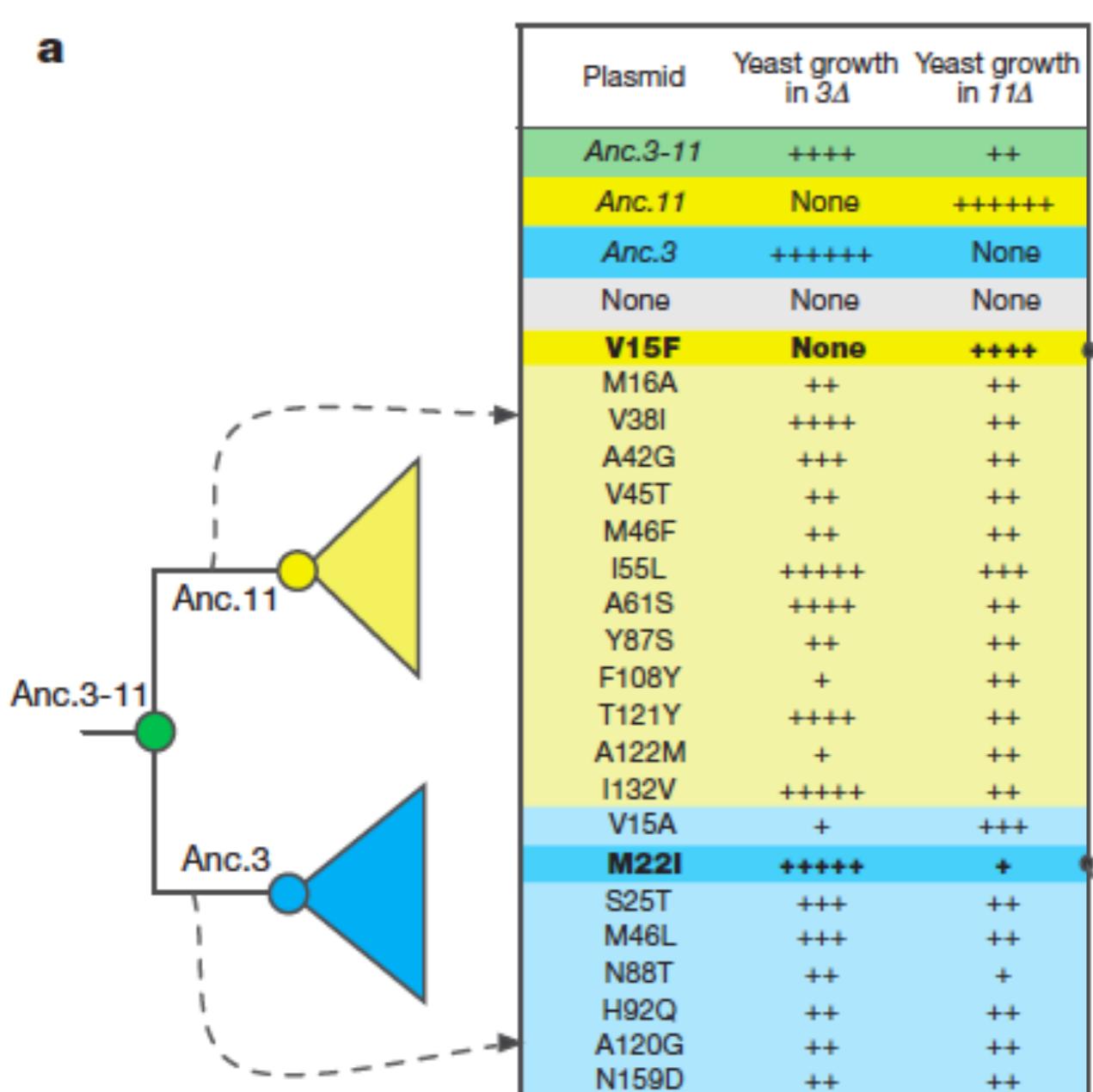


c





Discovering causal mutations for evo. of co-dependence



How did such a complex machinery evolve?

1

We are aware of no other mechanistic analyses of a molecular machine's evolutionary trajectory, so the generality of our observations is unknown. By definition, however, all molecular machines involve

2

In the evolution of any such assembly, additional paralogues could become obligate components because of gene duplication³⁰ and subsequent mutations that cause specific interaction interfaces among them to degenerate.

3

and the complexity of the ring increased. It is possible that specialization of the duplicated subunits allowed increases in fitness, but genome-wide

4

This view of the evolution of molecular machines is related to recent models that explain other biological phenomena—such as the retention of large numbers of duplicate genes and mobile genetic elements within genomes—as the product of degenerative processes acting on modular biological systems²⁷. Although mutations that enhanced the

How accurate?

Statistical tests are helpful but not conclusive

Method	Correctly Inferred Sites	Incorrectly Inferred Sites	Total Sites
PAML_Γ	98.14%	71	3825
FASTML_Γ	98.17%	70	3825
PAML	98.12%	72	3825
PHYLO_Γ	97.88%	81	3825
MP	98.07%	74	3825

A very smart way of testing the accuracy
of reconstructed ancestral sequences:
phenotype profiles of modern and ancient fluorescent
proteins

Reconstructed common ancestors do *not* show outlier phenotypes



Tesekkürler:

Tugce Bilgin

Sibel Kücükyıldırım

Can Elverici

Reyhan Yaka

Kursat Sahin

Mehmet Somel / EKO-EVO Derneği

Hacettepe Biyoloji Bölümü

Part II - white board lecture / topics covered:

- Phylogenetic tree construction methods: maximum parsimony, maximum likelihood
- Estimating ancestral sites using maximum likelihood method
- Why do we need models of evo./substitution while calculating phylogenies?

Part III - Linux-bash-based hands-on application - reconstruction of ancestral sequences of modern ADH4 sequences using t-coffee & iqtree

Hominids adapted to metabolize ethanol long before human-directed fermentation

Matthew A. Carrigan^{a,b,1}, Oleg Uryasev^b, Carole B. Frye^b, Blair L. Eckman^b, Candace R. Myers^c, Thomas D. Hurley^c, and Steven A. Benner^b