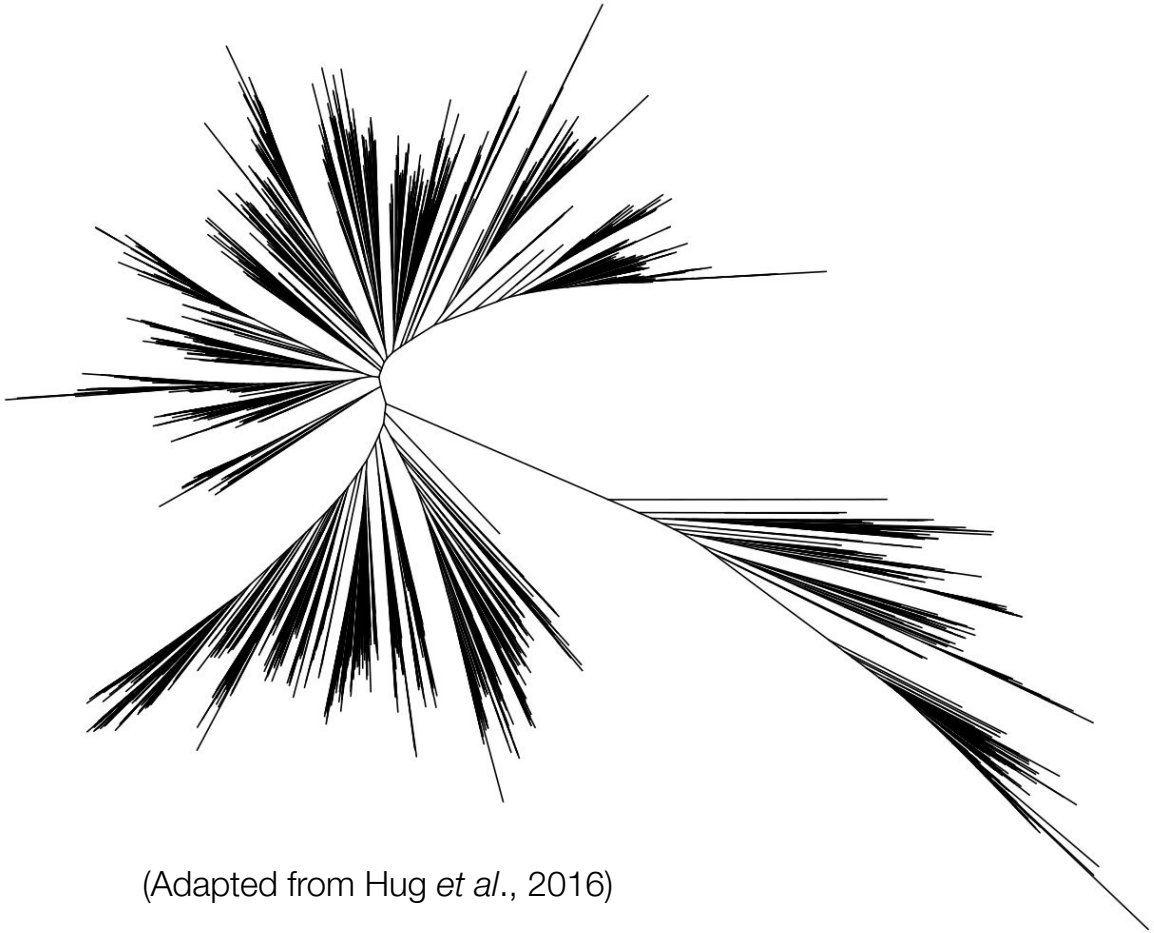
Life Sciences Department Seminar

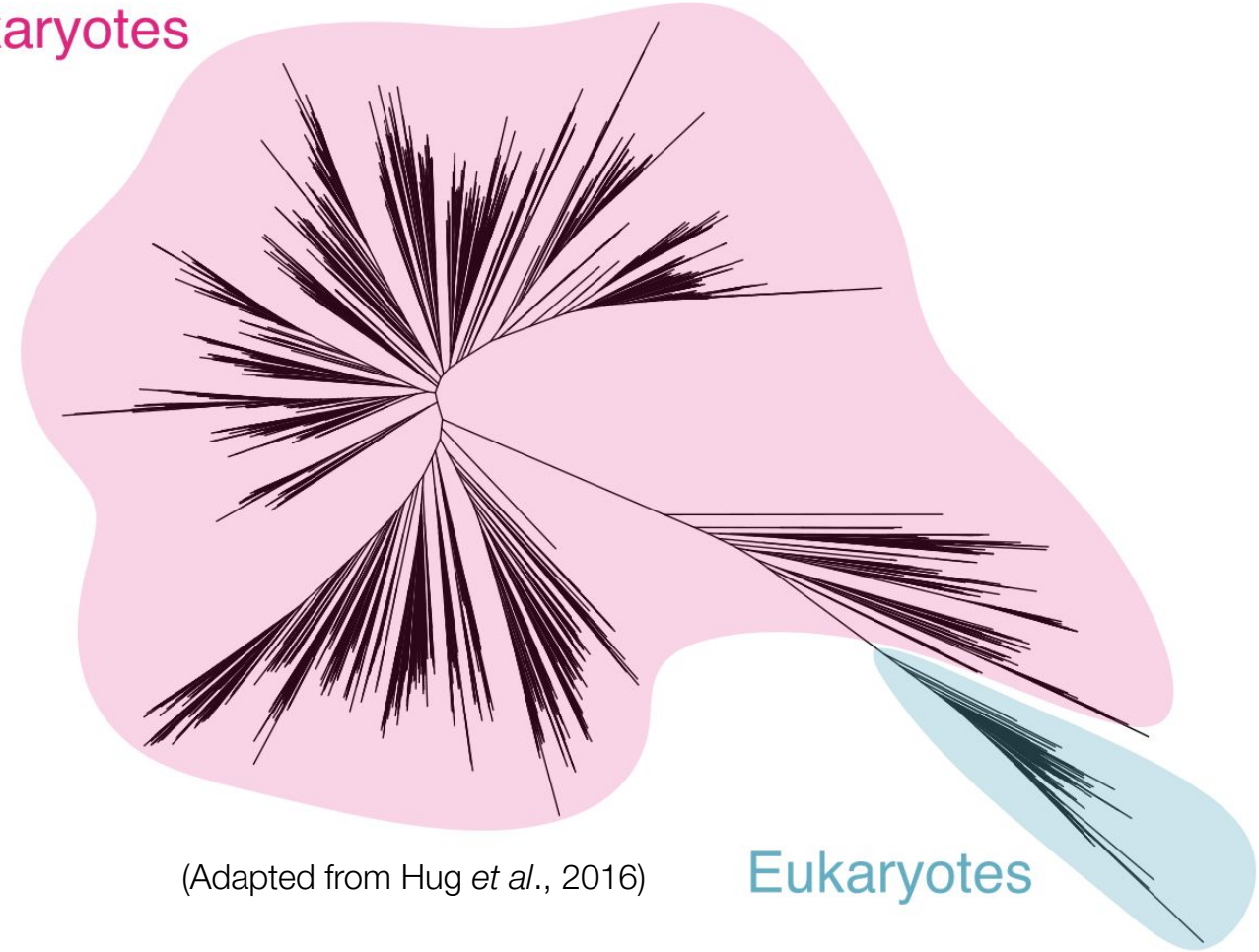
Diverse ancestries reveal complex symbiotic interactions during eukaryogenesis

Saioa Manzano-Morales
Toni Gabaldón
Comparative Genomics - Life Sciences
Barcelona Supercomputing Center



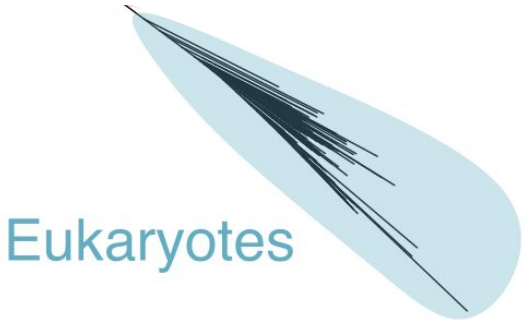
(Adapted from Hug *et al*., 2016)

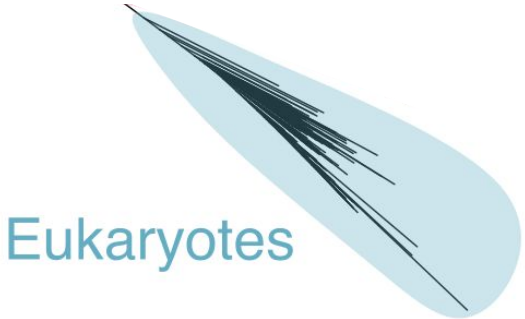
Prokaryotes

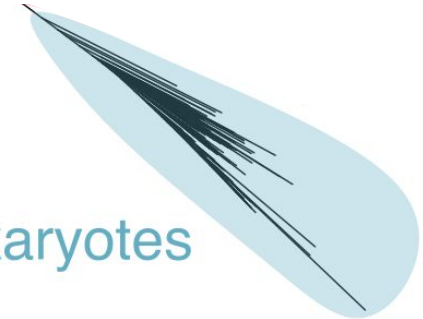


(Adapted from Hug *et al*., 2016)

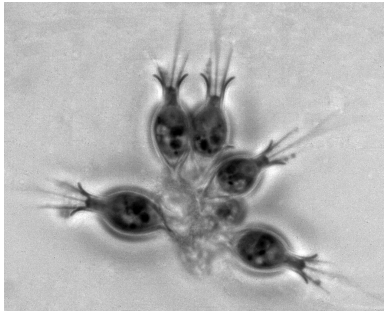
Eukaryotes

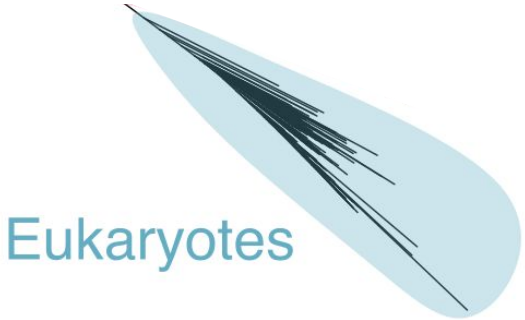

Eukaryotes

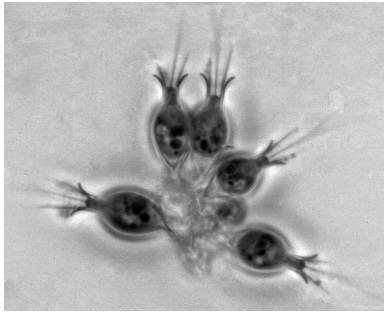

Eukaryotes



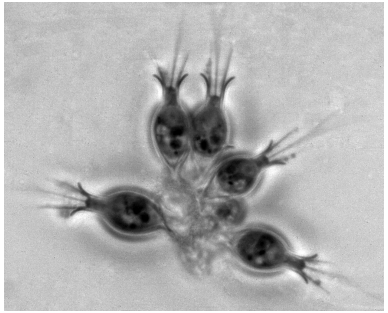
Eukaryotes



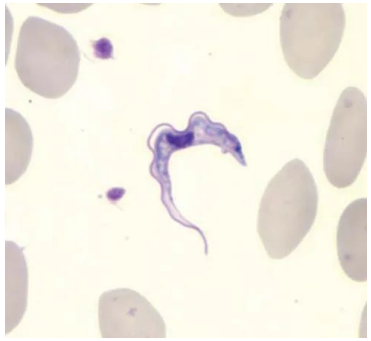

Eukaryotes

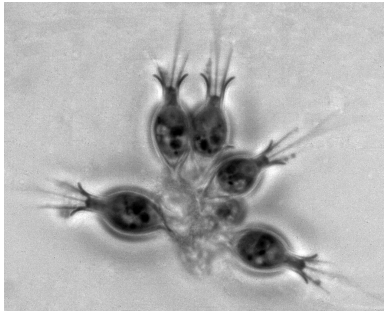


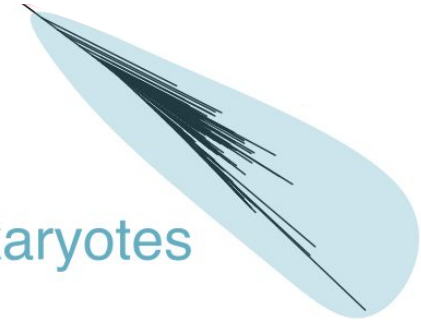
Eukaryotes

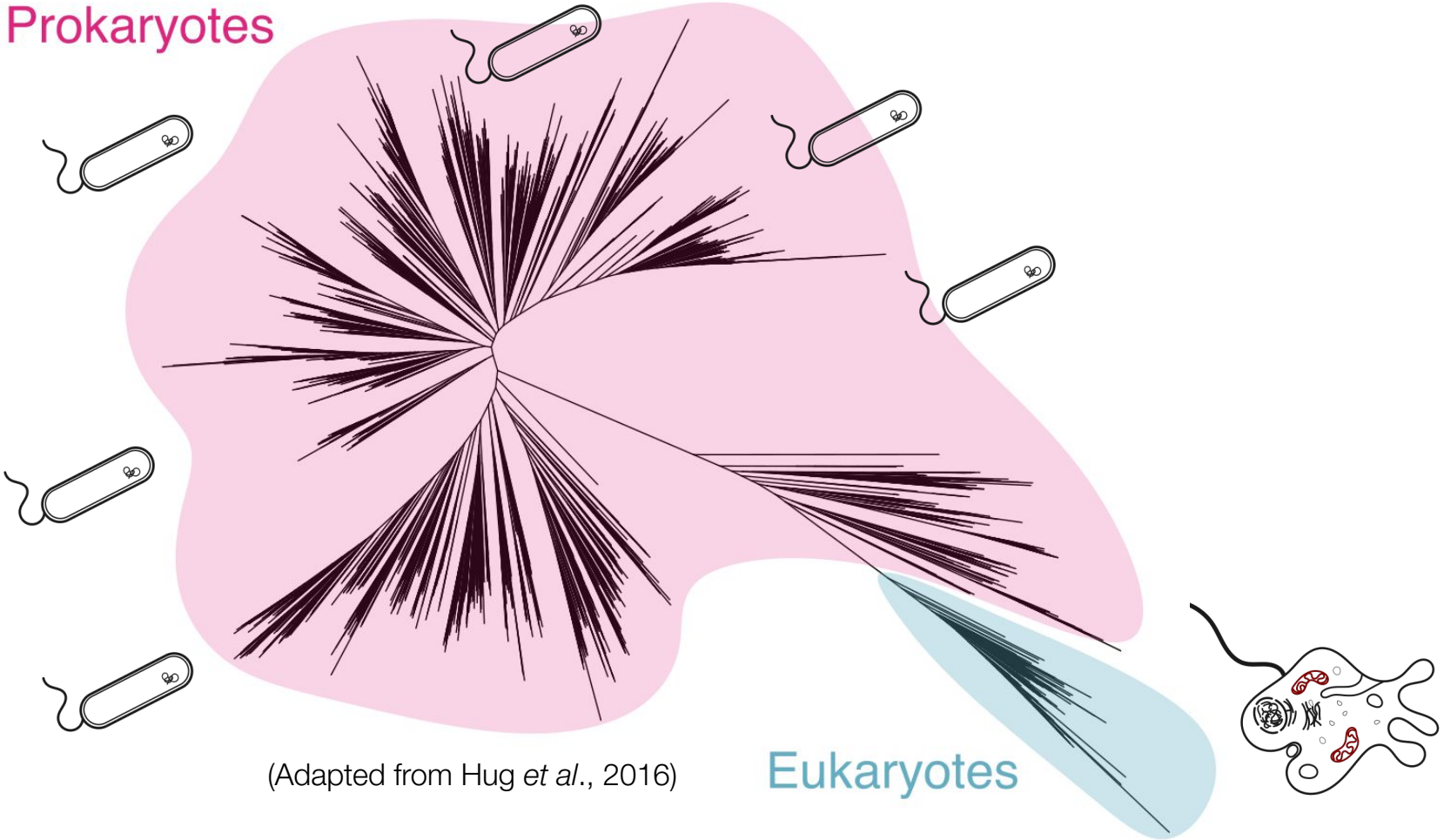


Eukaryotes



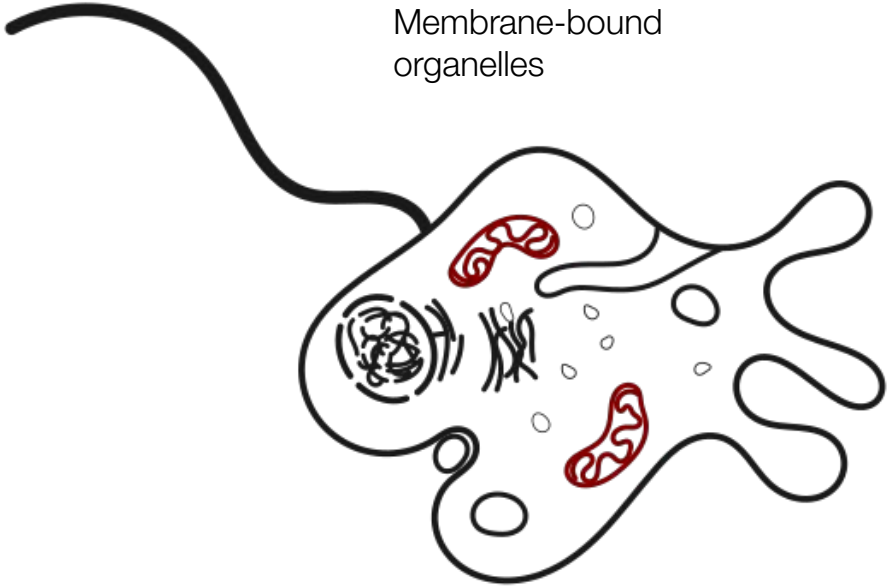



Eukaryotes

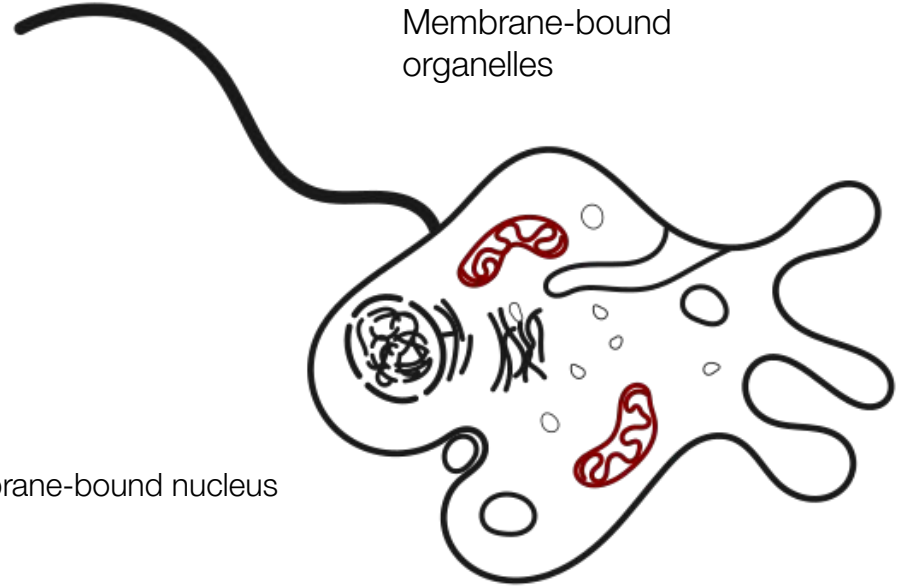


(Adapted from Hug *et al*., 2016)

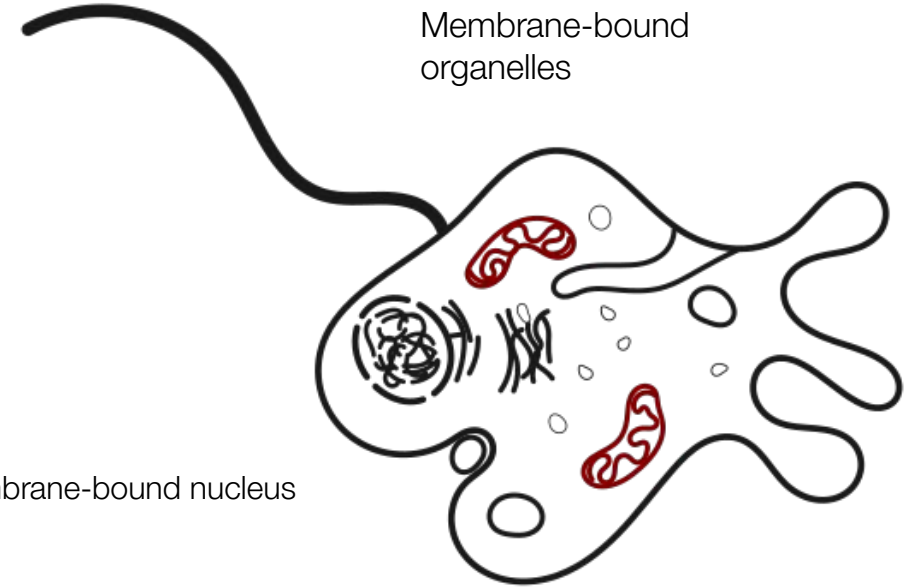
Prokaryotes and eukaryotes: the great divide



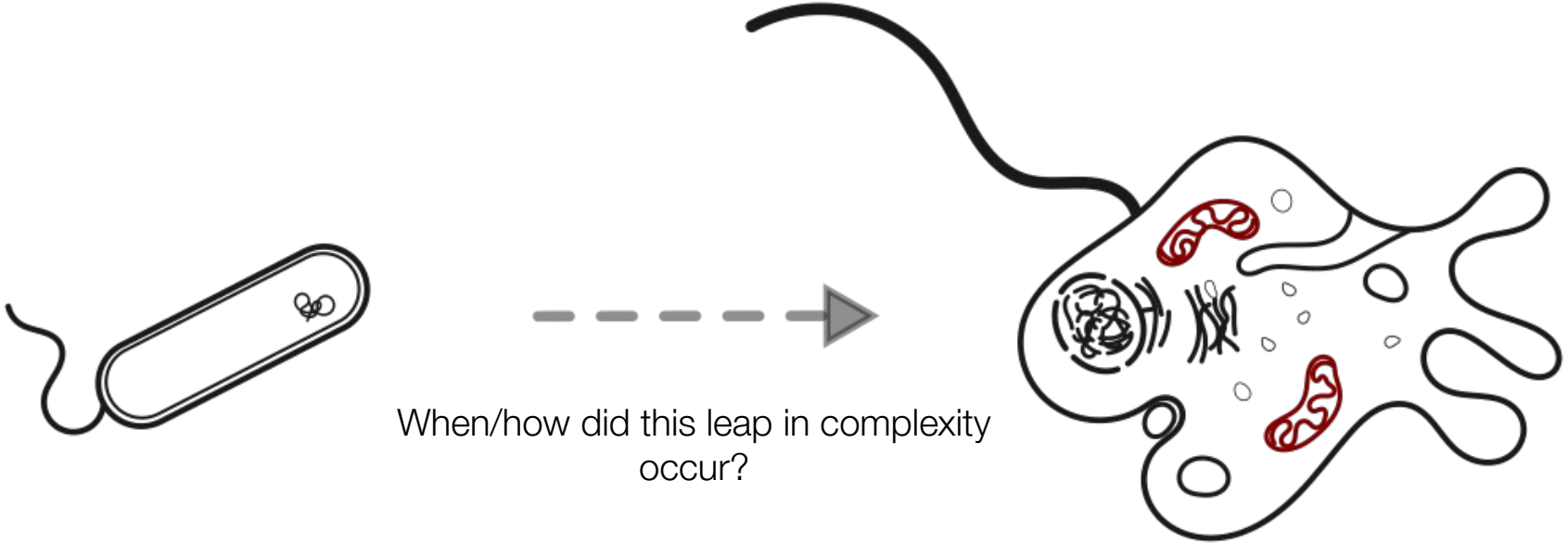
Prokaryotes and eukaryotes: the great divide



Prokaryotes and eukaryotes: the great divide

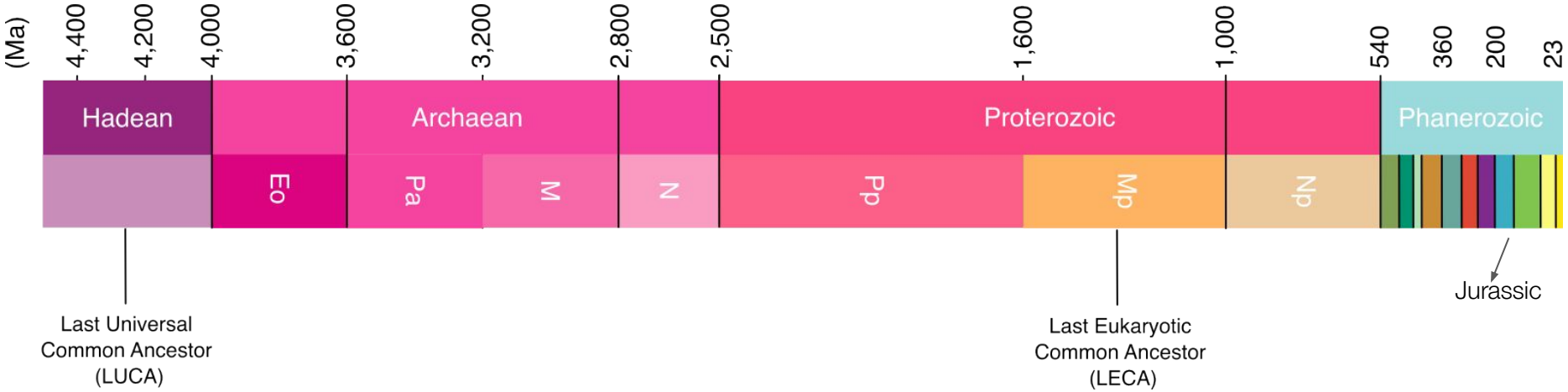


Prokaryotes and eukaryotes: the great divide



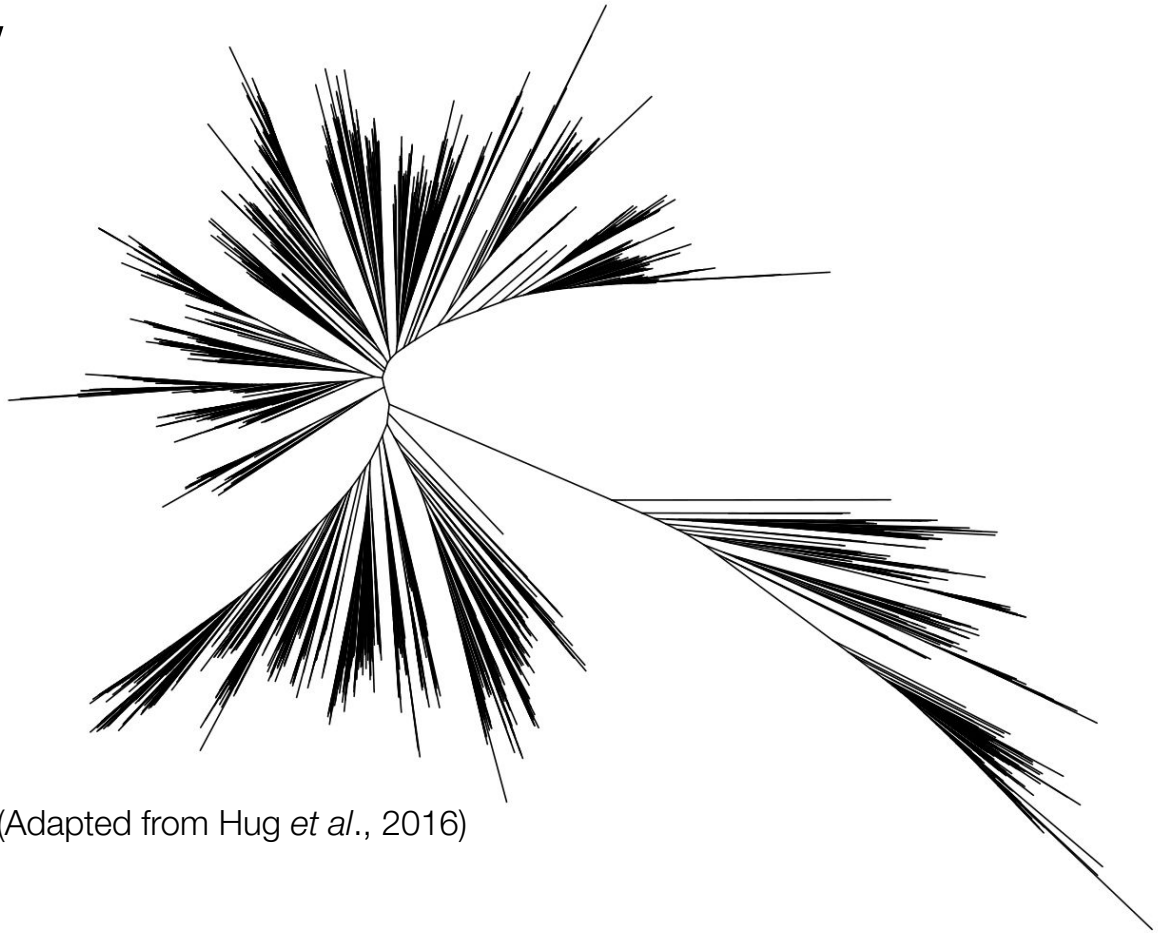
When/how did this leap in complexity occur?

The when



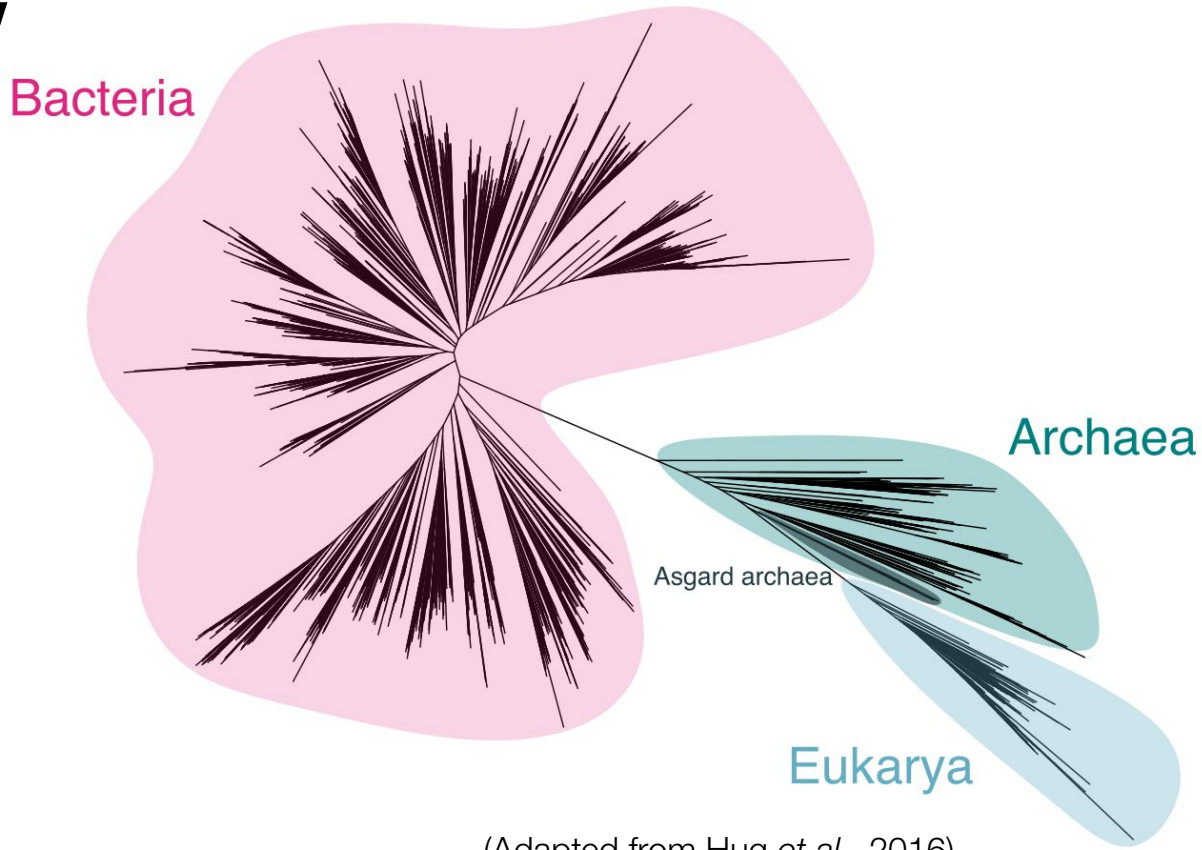
(Adapted from Moody *et al*., 2024)

The how



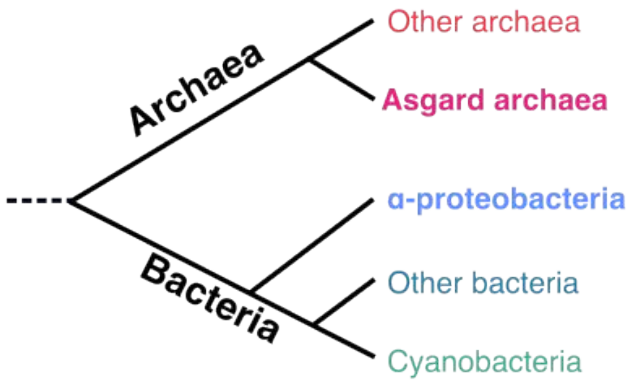
(Adapted from Hug *et al*., 2016)

The how

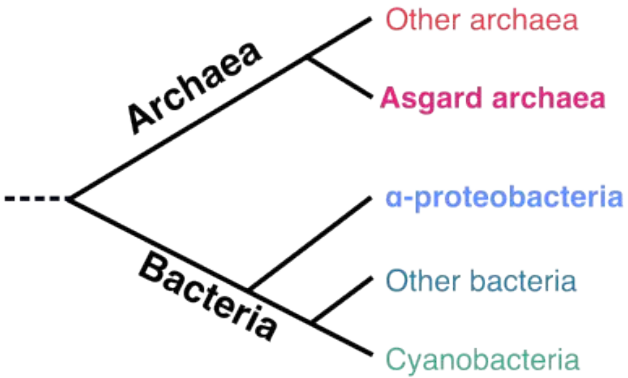


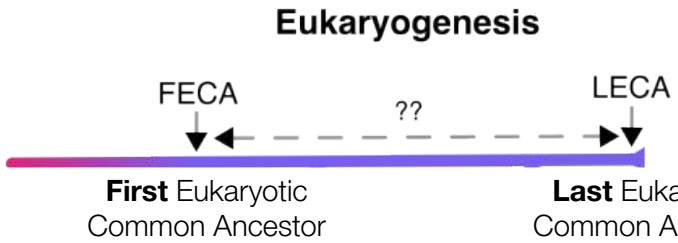
(Adapted from Hug *et al*., 2016)

Eukaryogenesis: a syntrophic affair



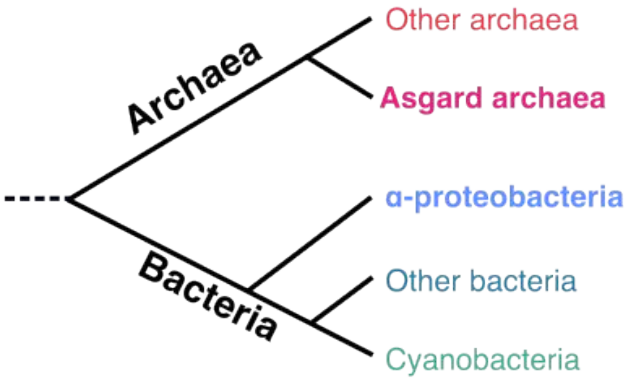
Eukaryogenesis: a syntrophic affair

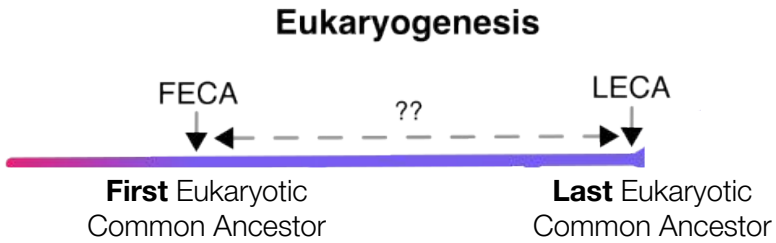




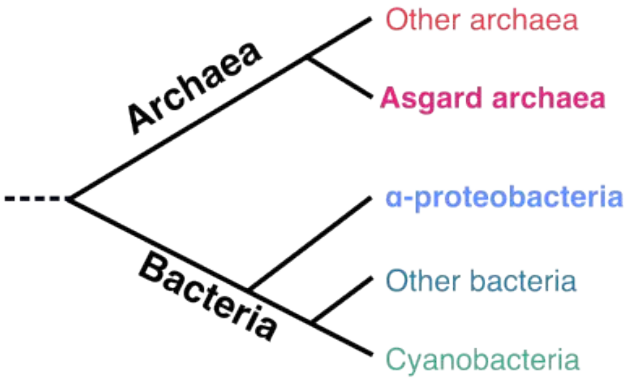
- First non-prokaryotic organism
- Last Common Ancestor of all (extant) eukaryotes

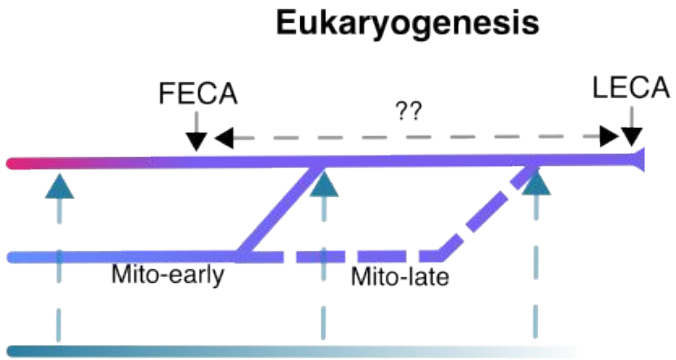
Eukaryogenesis: a syntrophic affair



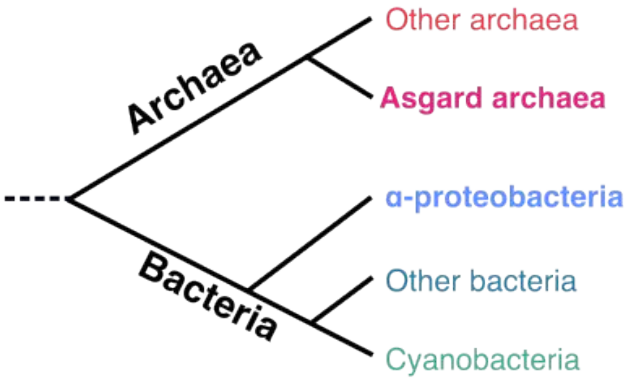


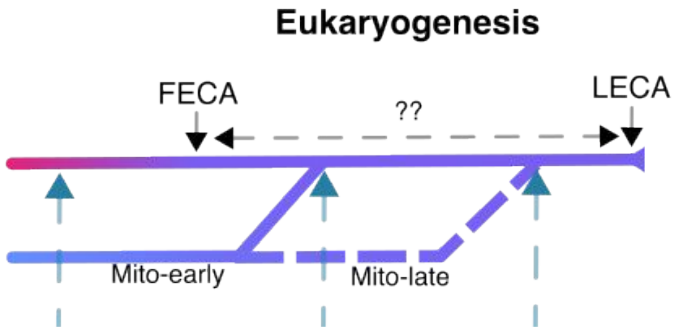
Eukaryogenesis: a syntrophic affair





Eukaryogenesis: a syntrophic affair

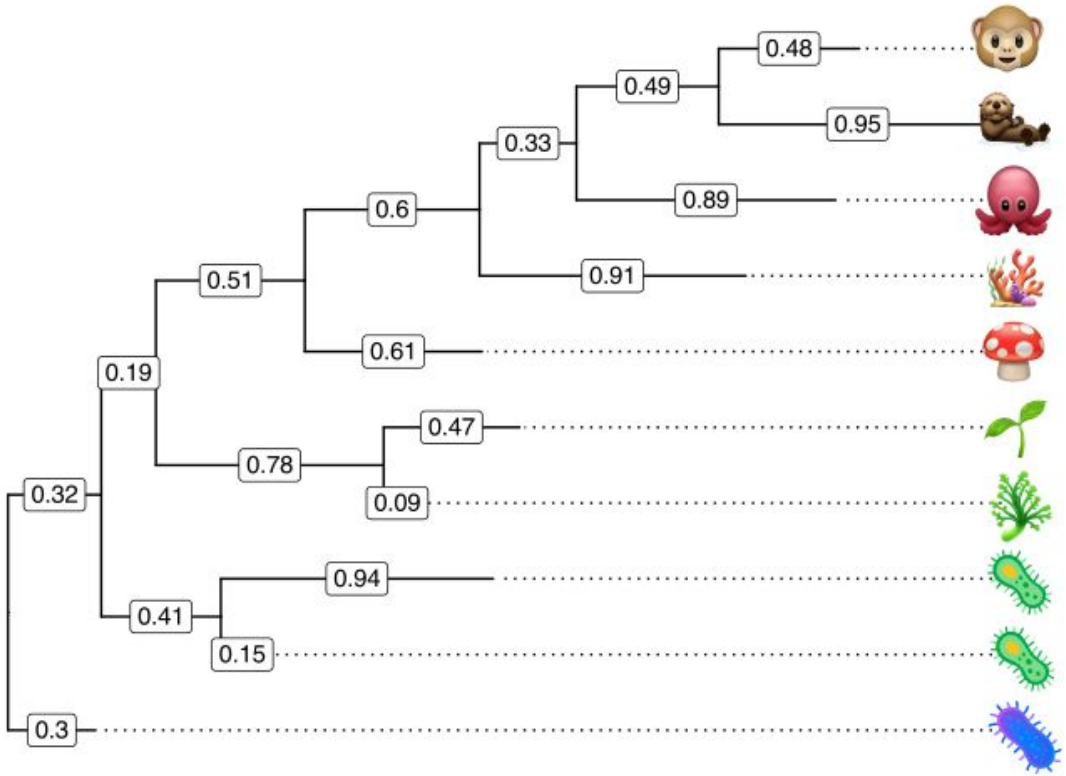




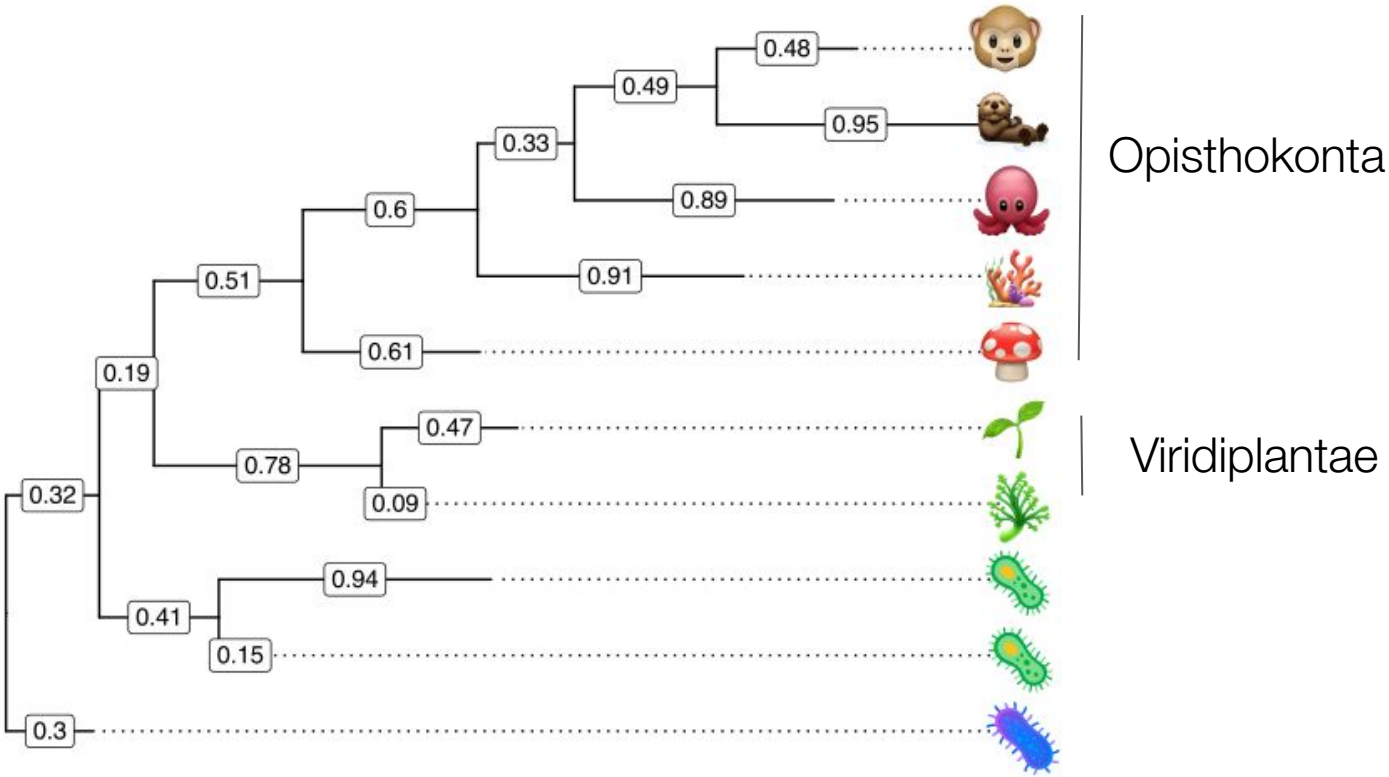
Marina Marcet-Houben

Moisès Bernabeu

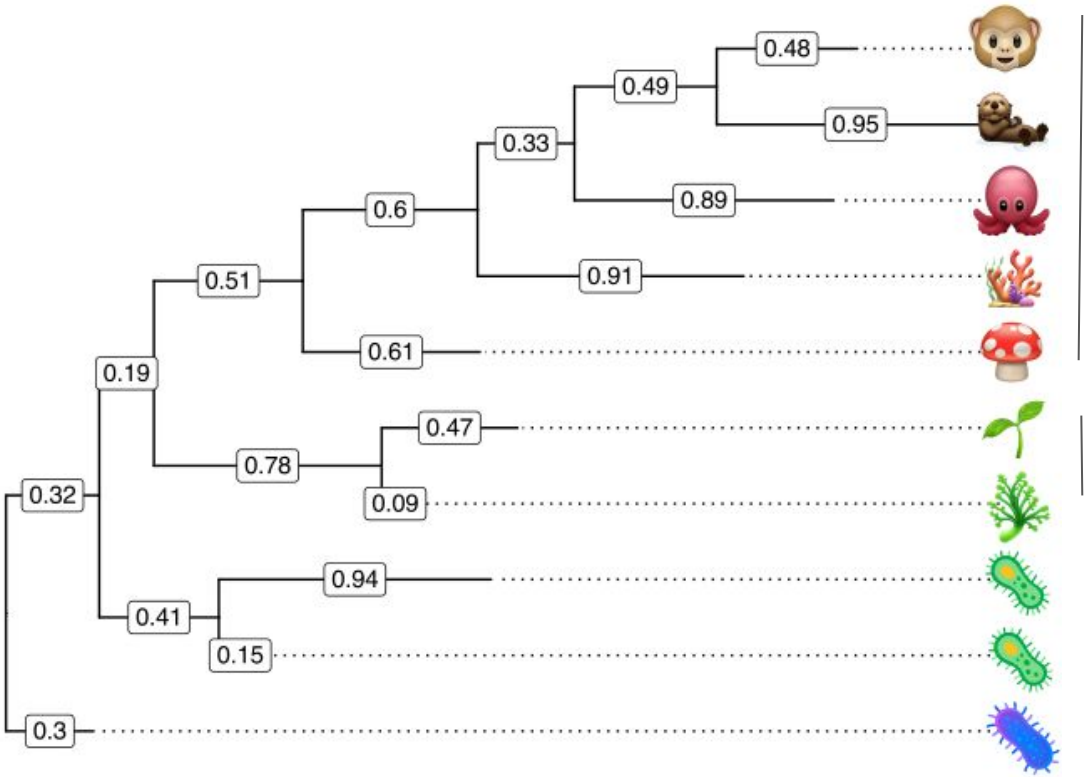
Unveiling the origin of eukaryotic gene families

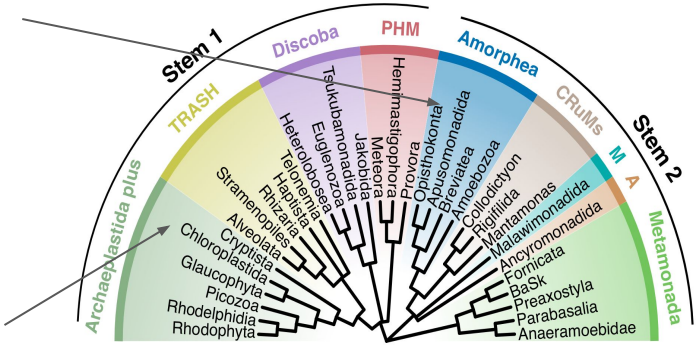


Unveiling the origin of eukaryotic gene families



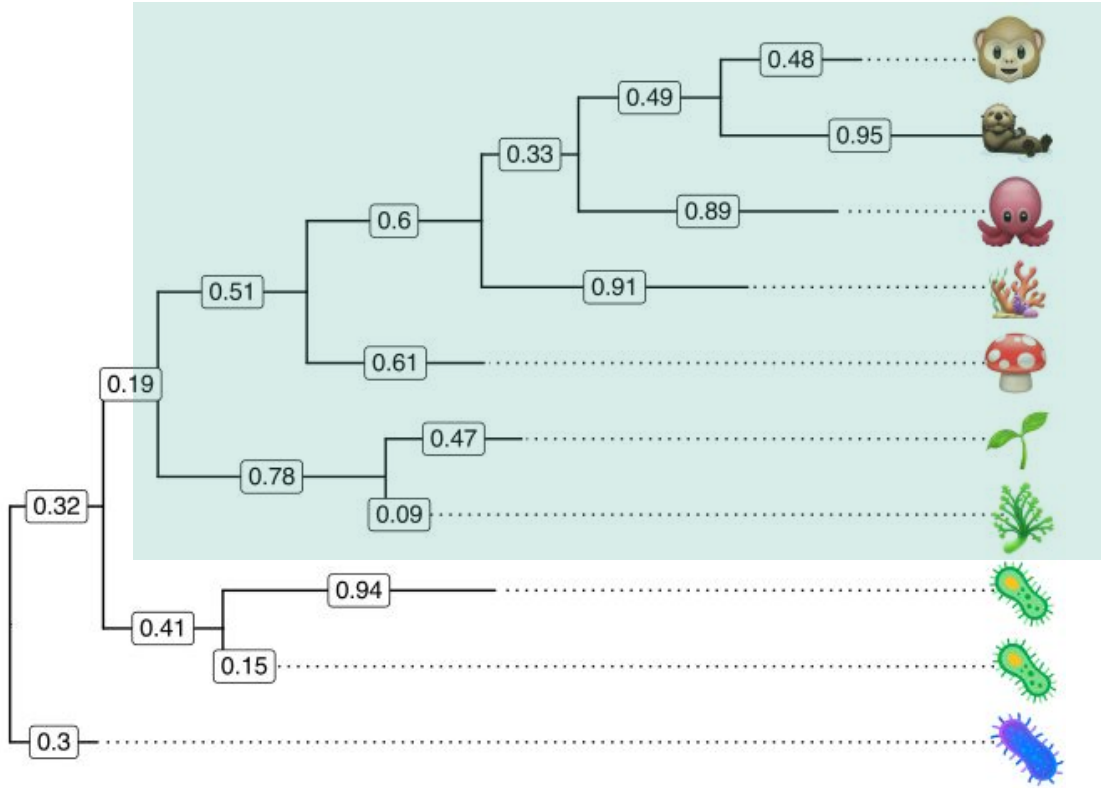
Unveiling the origin of eukaryotic gene families





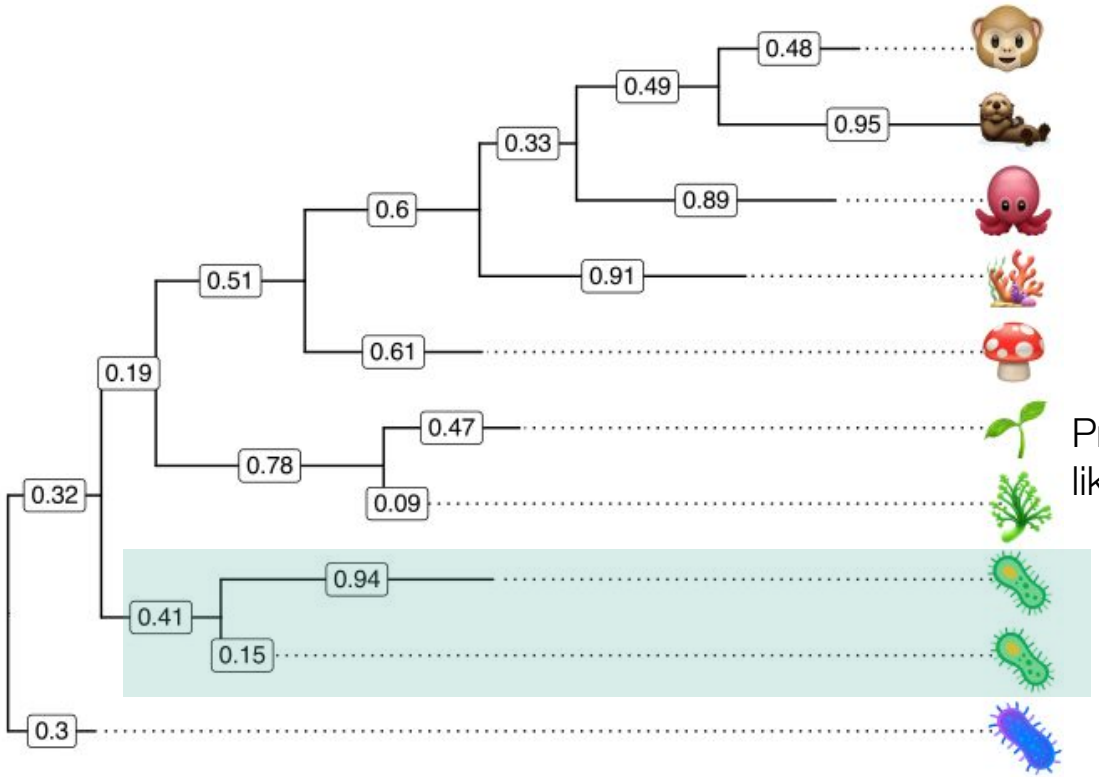
Distribution across two main stems informs of ancient origin

Unveiling the origin of eukaryotic gene families



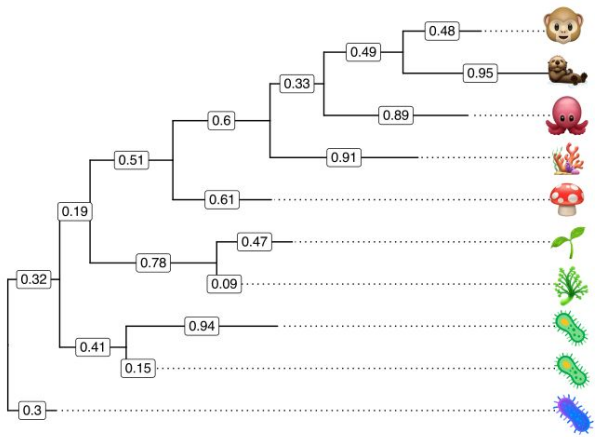
LECA

Unveiling the origin of eukaryotic gene families



Prokaryotic lineage most related is likely donor

Unveiling the origin of eukaryotic gene families

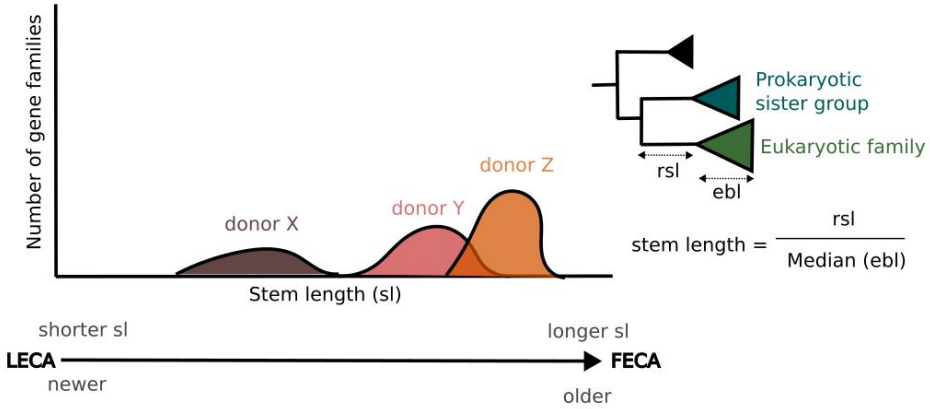


(Pittis & Gabaldón, 2016)

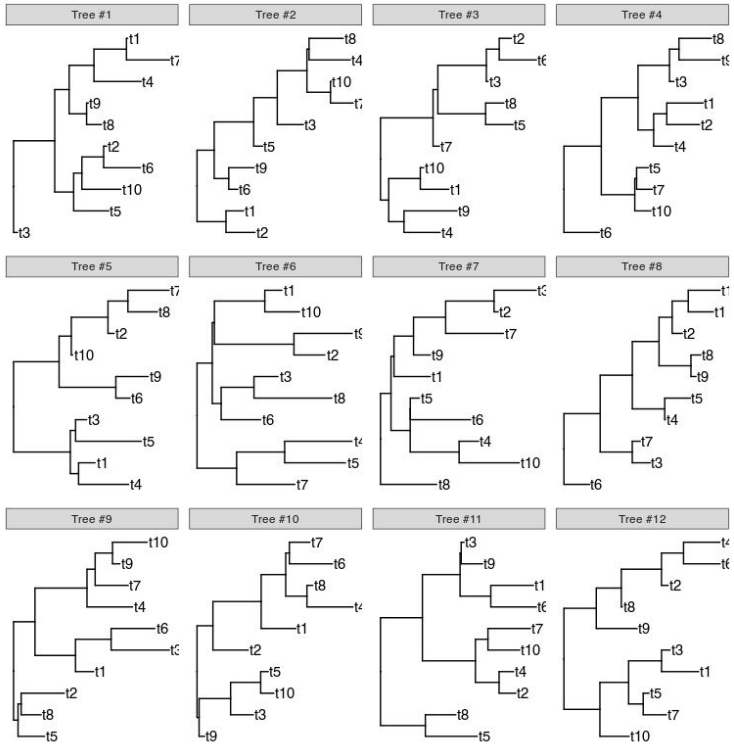
(Bernabeu, Armero & Gabaldón, 2024) (preprint)

Branch lengths in gene trees inform us of evolutionary change → time and rate

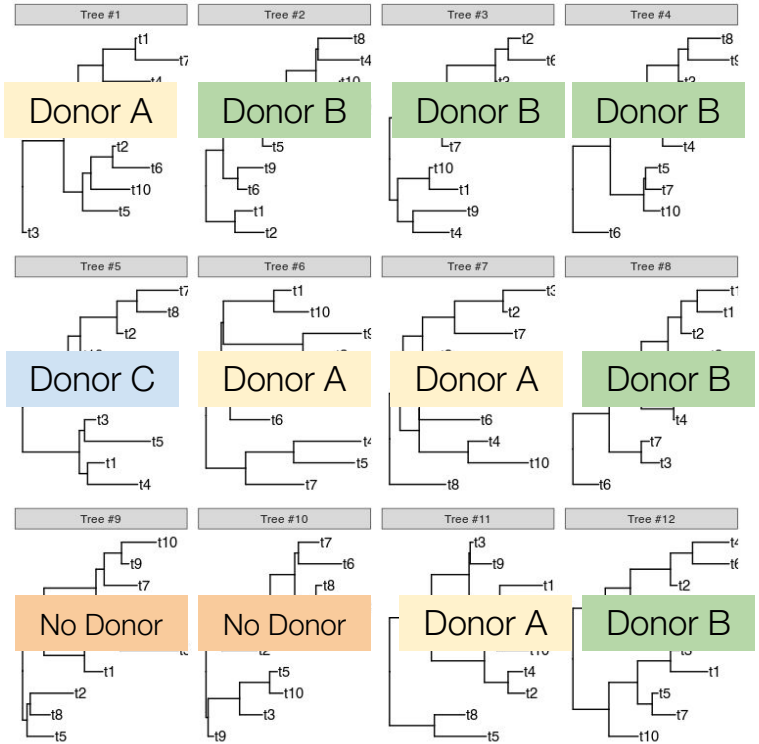
Ratio of branch lengths allows inference of relative timing



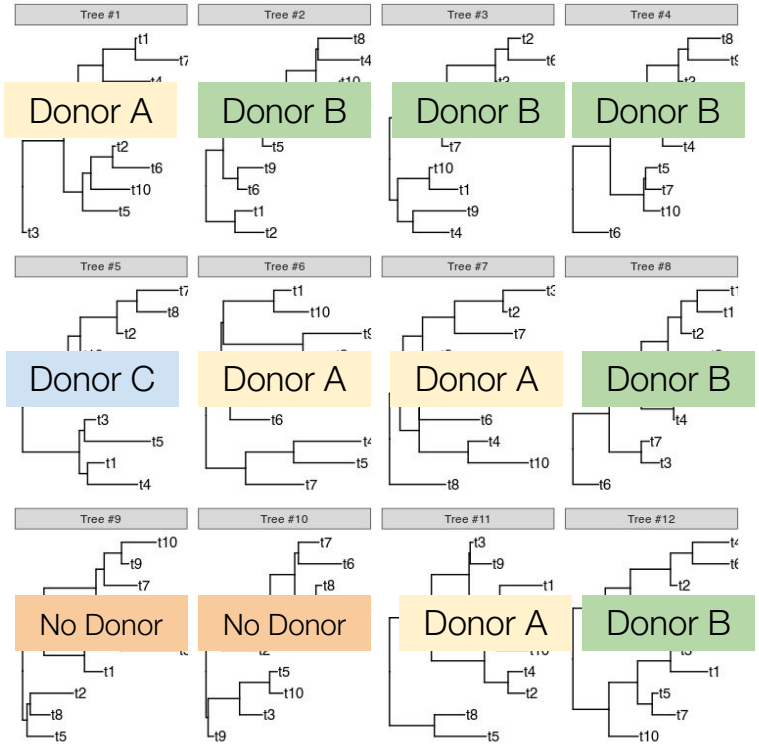
Origin of LECA gene families allows quantification of contributors

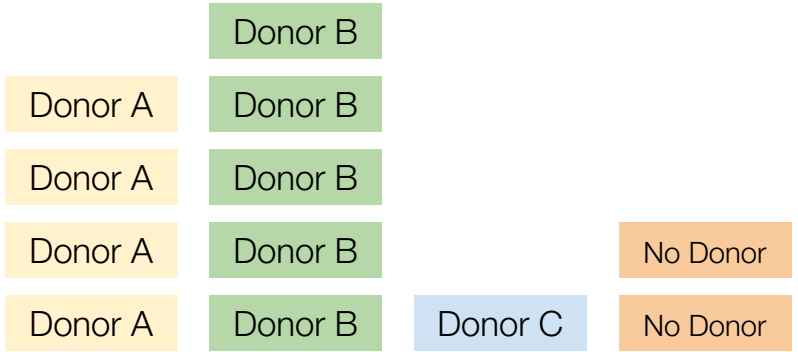


Origin of LECA gene families allows quantification of contributors

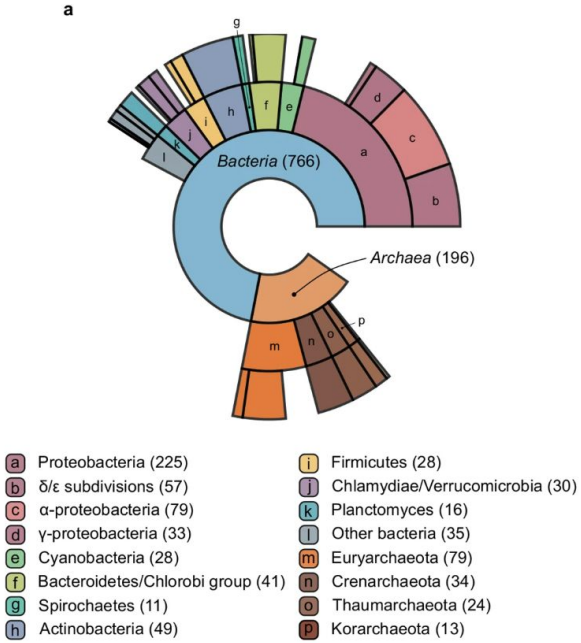


Origin of LECA gene families allows quantification of contributors



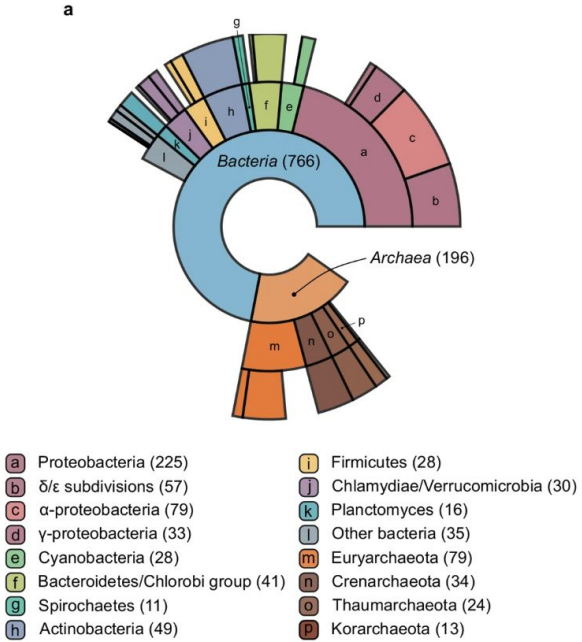


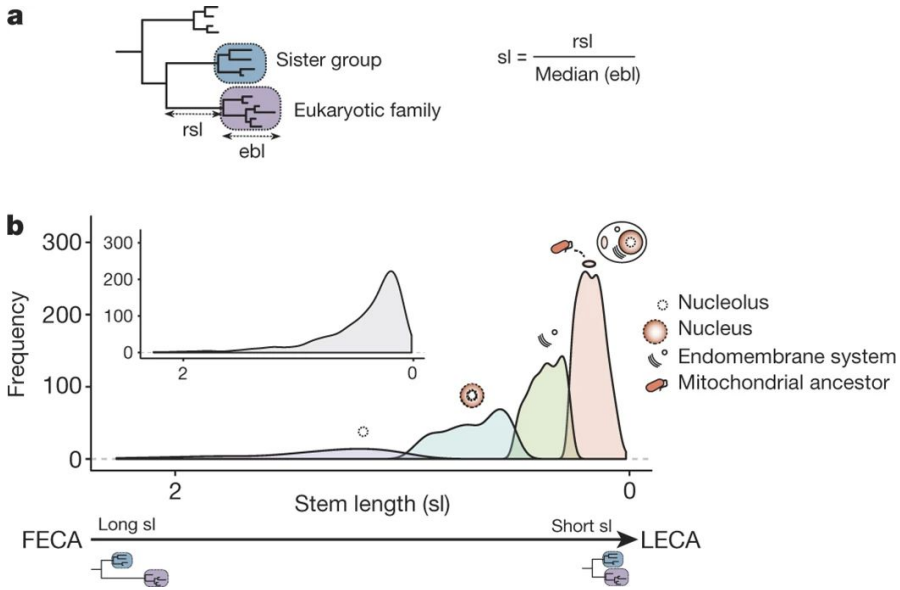
Bacterial contributions outweigh archaeal ones in LECA



(Pittis & Gabaldón, 2016)

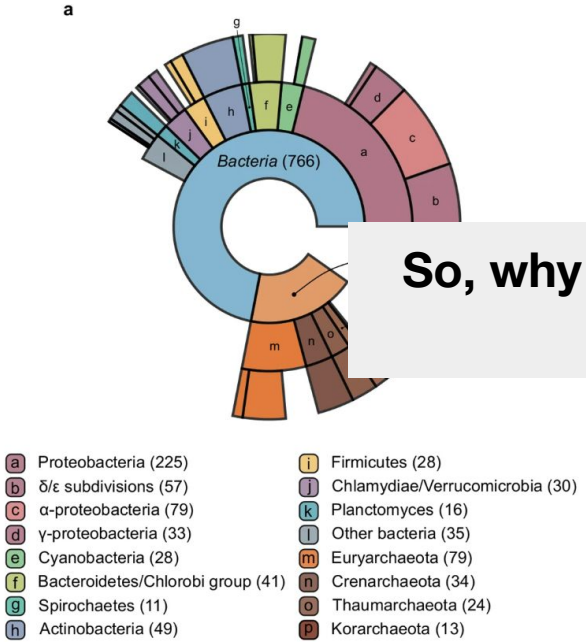
Mitochondrial endosymbiosis was not the kick-off event

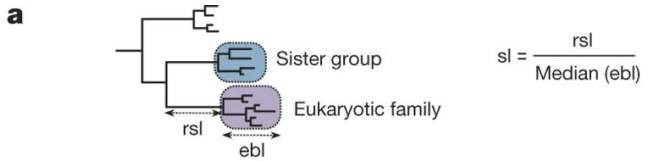




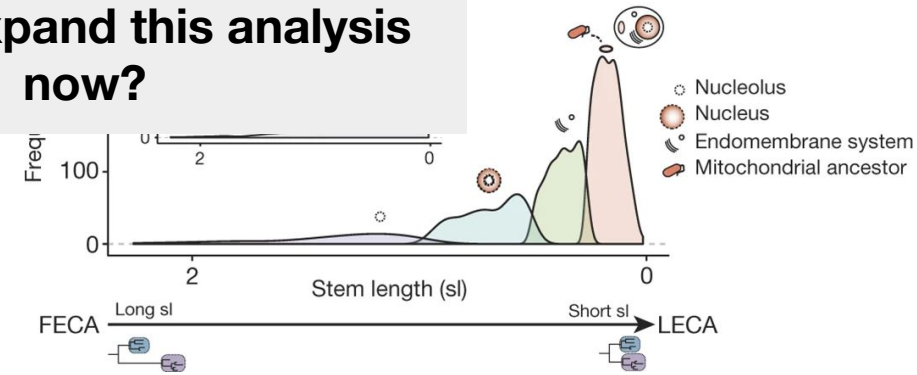
(Pittis & Gabaldón, 2016)

Mitochondrial endosymbiosis was not the kick-off event



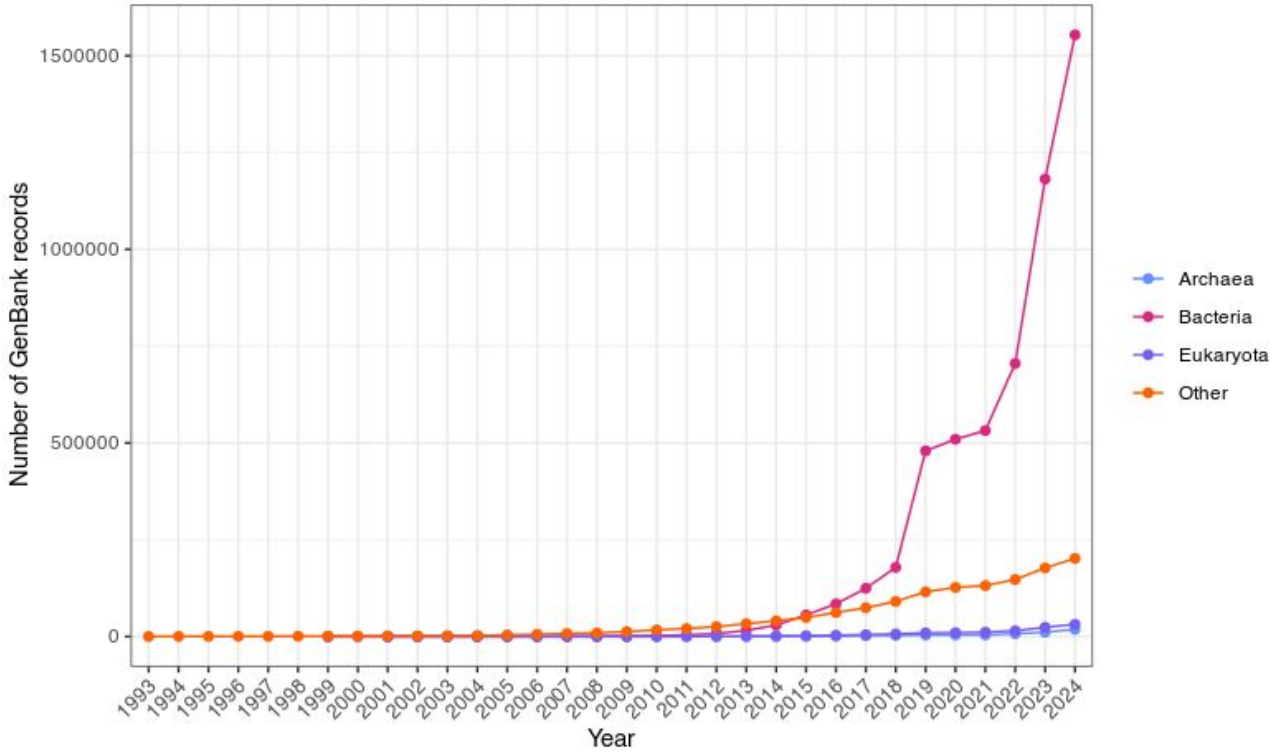


So, why expand this analysis now?

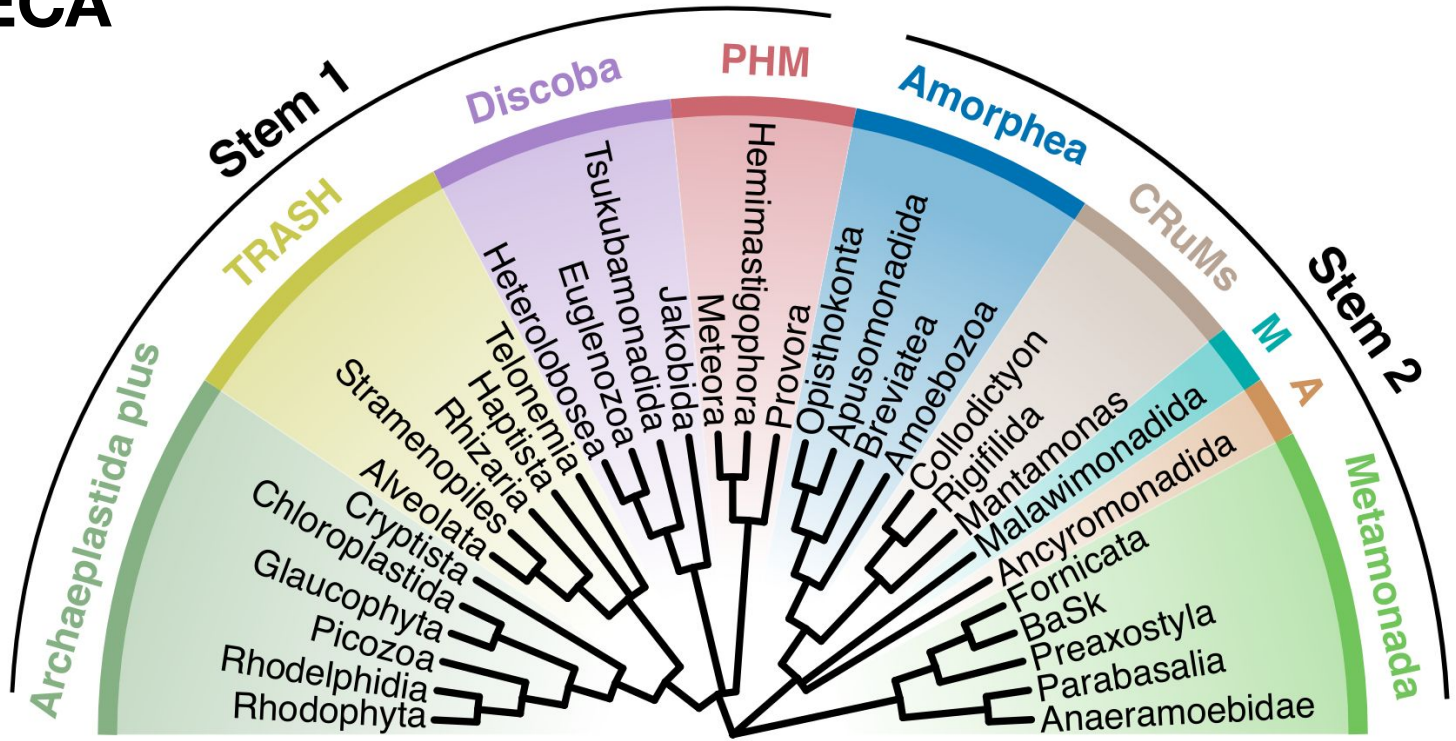


(Pittis & Gabaldón, 2016)

The sequencing boom



New eukaryotic lineages re-shape our understanding of LECA

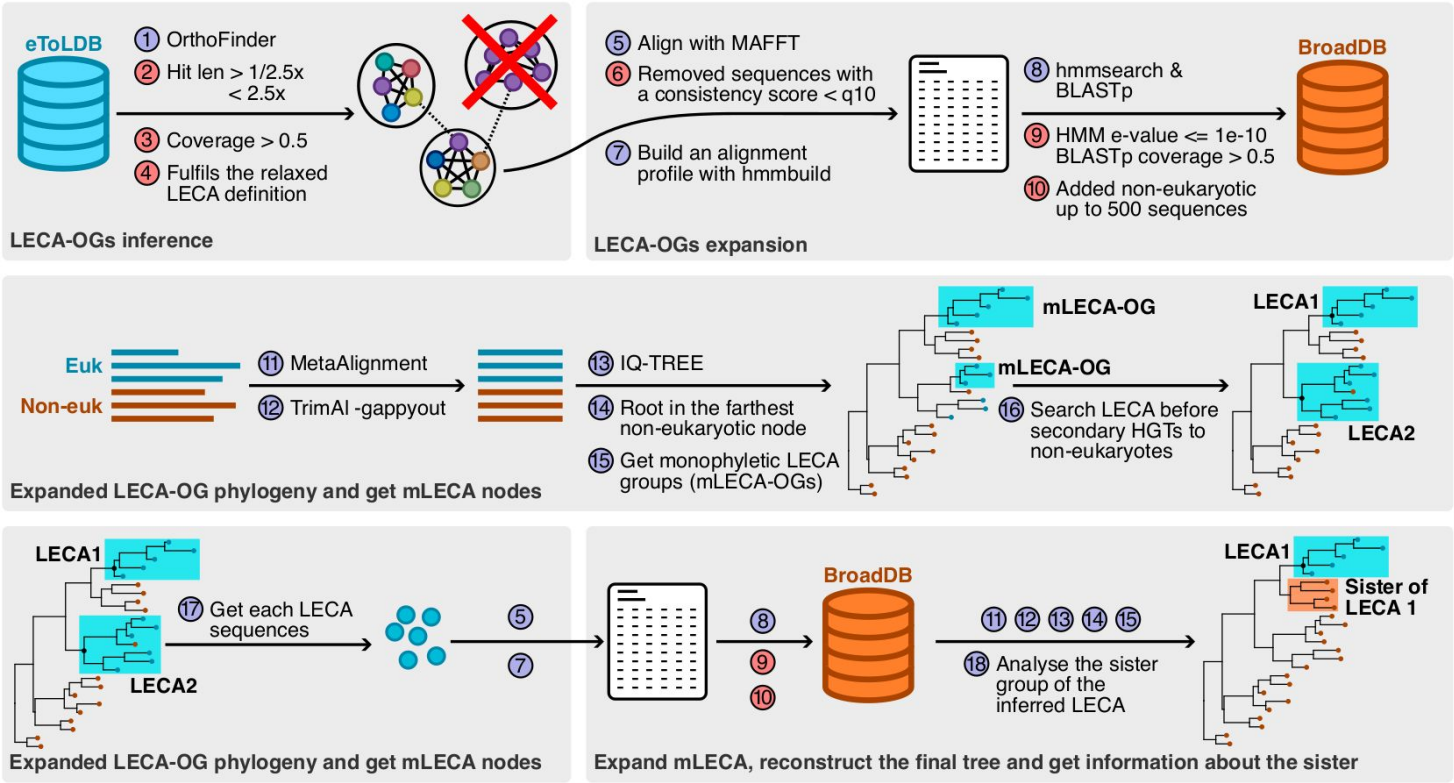


(Bernabeu *et al*., 2024) (preprint)

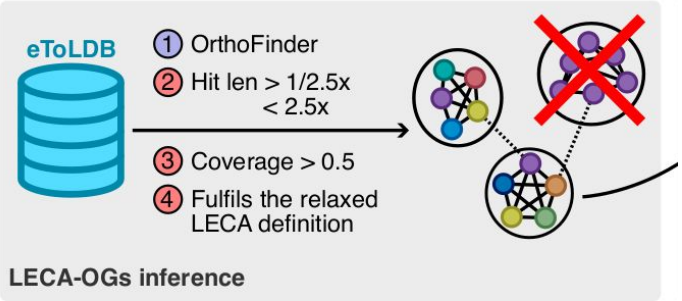
Open questions remain

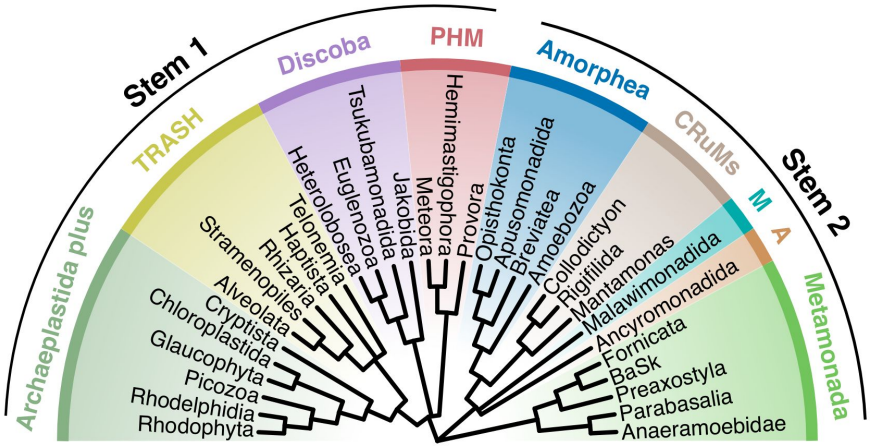
- What did the other donors contribute to the LECA proteome?
- Were there “gene waves” or was it a constant drip?
- Were acquisitions modular?

State-of-the art phylogenetic pipeline for origin inference

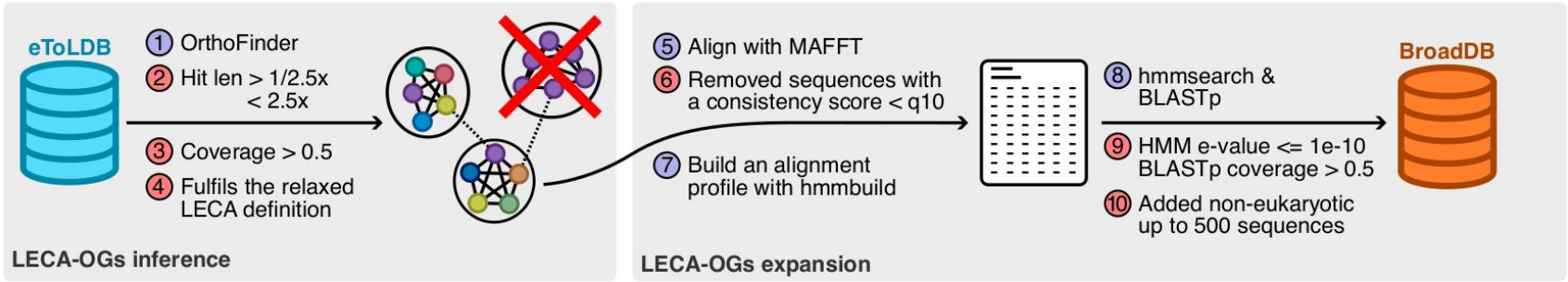


(Bernabeu *et al*., 2024) (preprint)

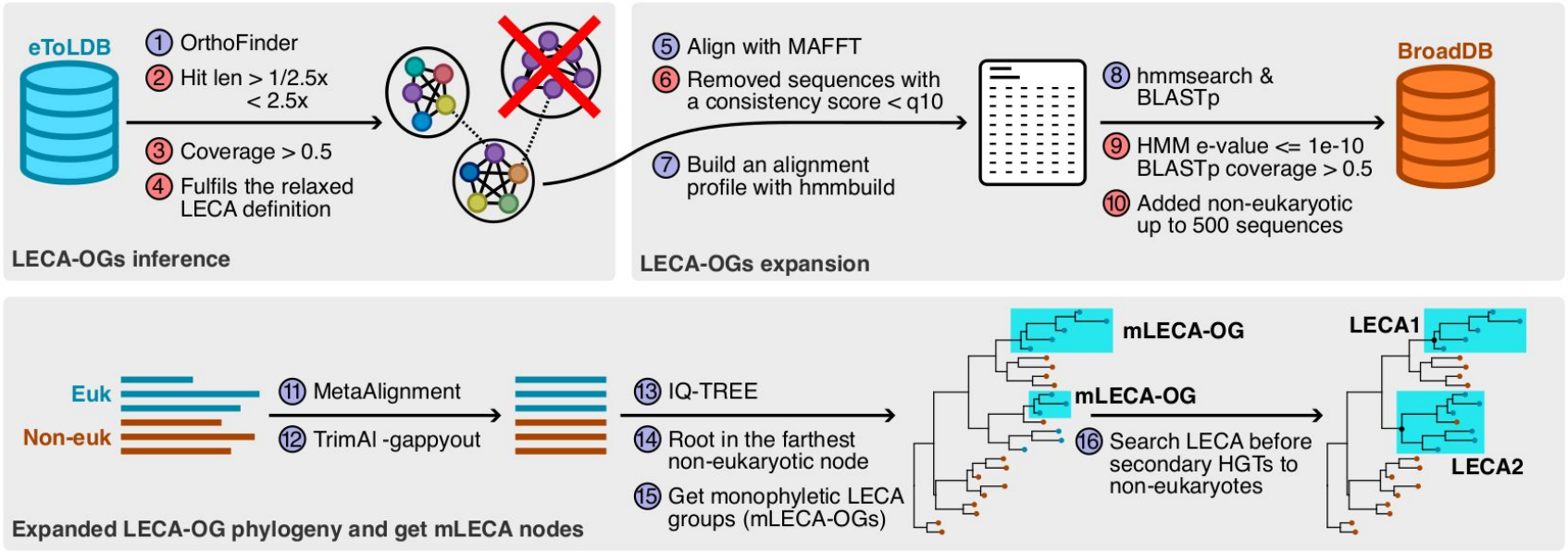




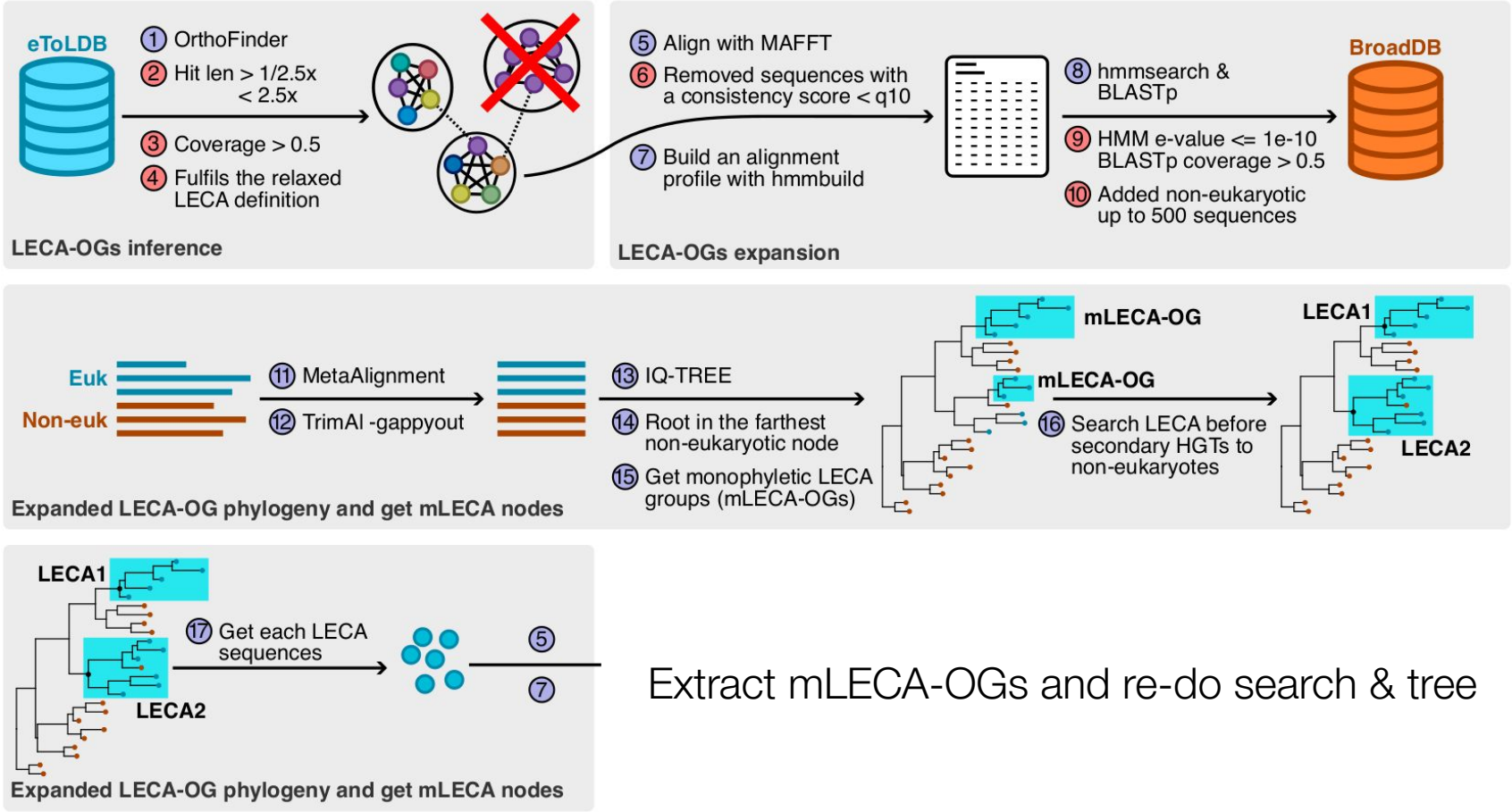
- 3 partially overlapping datasets (100 proteomes/dataset) - eTOLDB A/B/C
- Ortholog inference (OrthoFinder)
- LECA criteria:
  - Present in both stems
  - At least 5 organisms
  - At least 3 supergroups



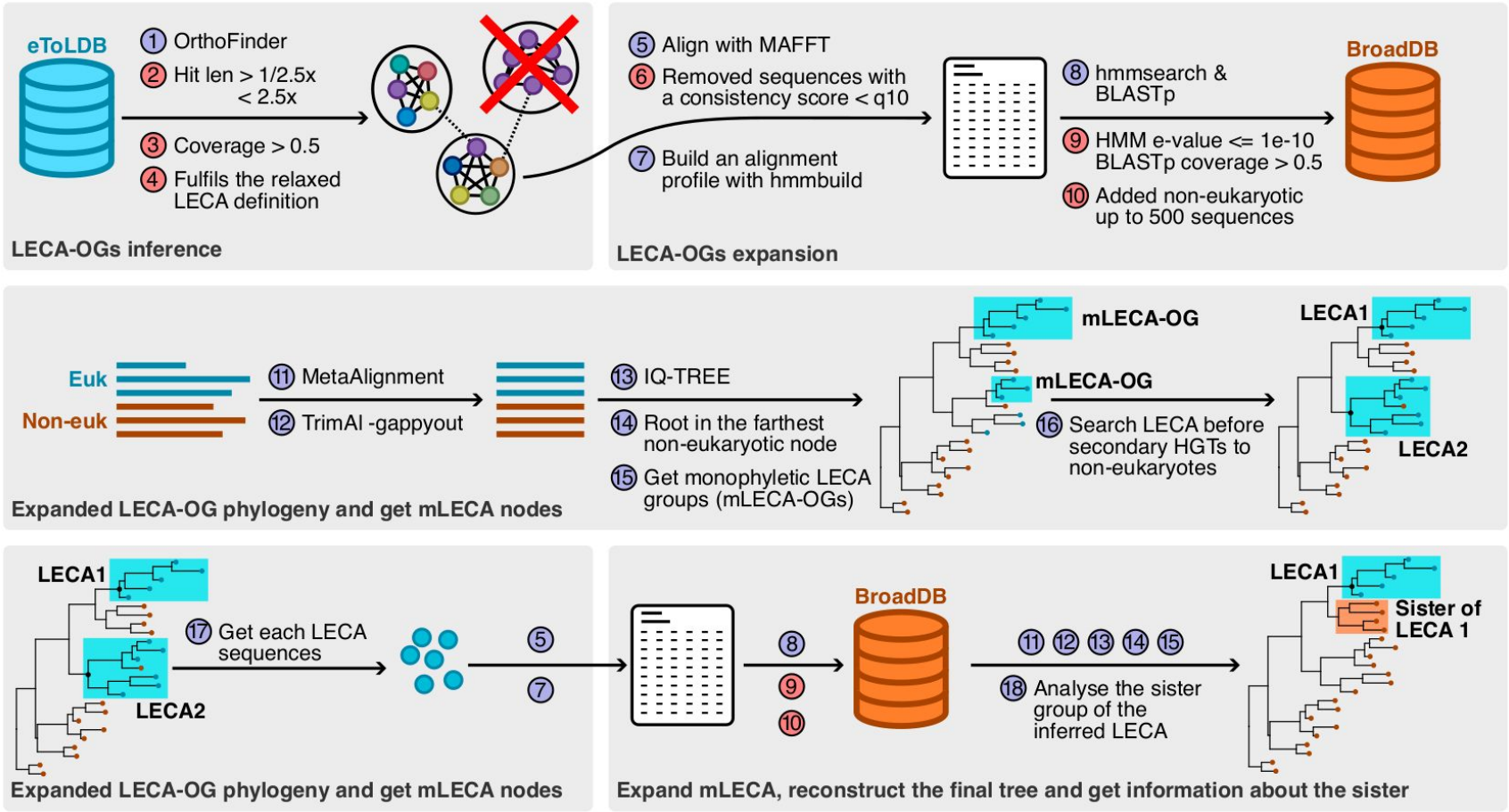
Search HMM of LECA-OG against broad dataset of prokaryotes and viruses

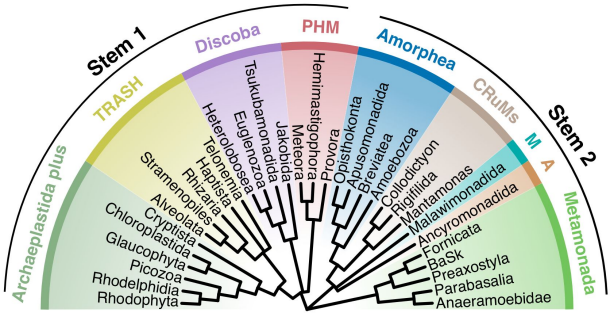


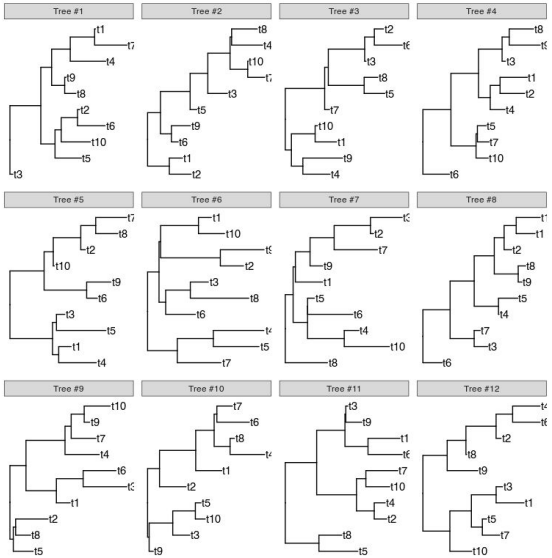
Reconstruct gene tree & find mOGs that fullfill LECA



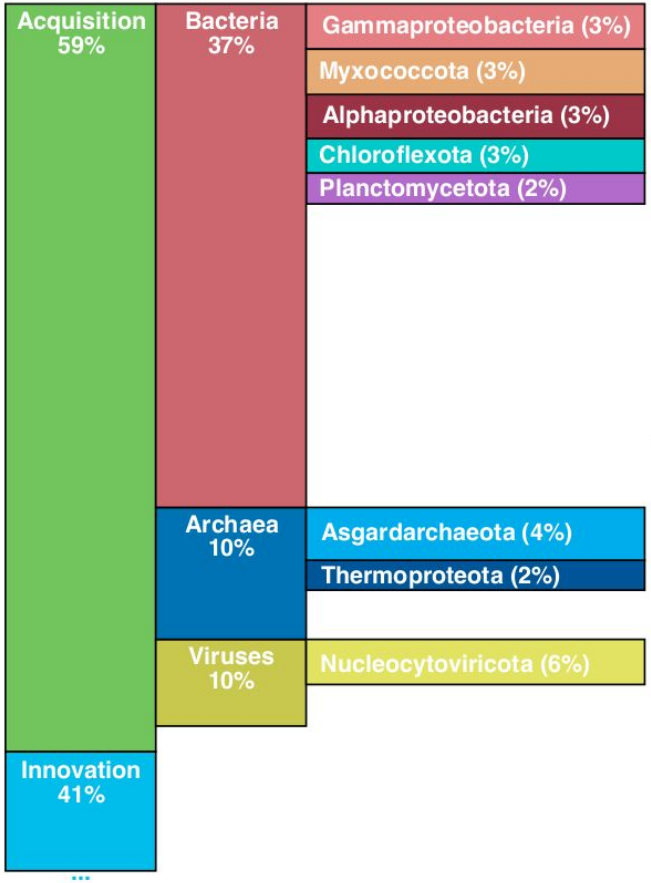
Extract mLECA-OGs and re-do search & tree


eToLDB
1 OrthoFinder
2 Hit len > 1/2.5x < 2.5x
3 Coverage > 0.5
4 Fulfils the relaxed LECA definition
LECA-OGs inference
5 Align with MAFFT
6 Removed sequences with a consistency score < q10
7 Build an alignment profile with hmmbuild
8 hmmsearch & BLASTp
9 HMM e-value <= 1e-10 BLASTp coverage > 0.5
10 Added non-eukaryotic up to 500 sequences
BroadDB
LECA-OGs expansion
Euk
Non-euk
11 MetaAlignment
12 TrimAl -gappyout
Expanded LECA-OG phylogeny and get mLECA nodes
13 IQ-TREE
14 Root in the farthest non-eukaryotic node
15 Get monophyletic LECA groups (mLECA-OGs)
mLECA-OG
mLECA-OG
16 Search LECA before secondary HGTs to non-eukaryotes
LECA1
LECA2
LECA1
LECA2
17 Get each LECA sequences
5
7
Expanded LECA-OG phylogeny and get mLECA nodes
8
9
10
BroadDB
11 12 13 14 15
18 Analyse the sister group of the inferred LECA
LECA1
Sister of LECA 1
Expand mLECA, reconstruct the final tree and get information about the sister

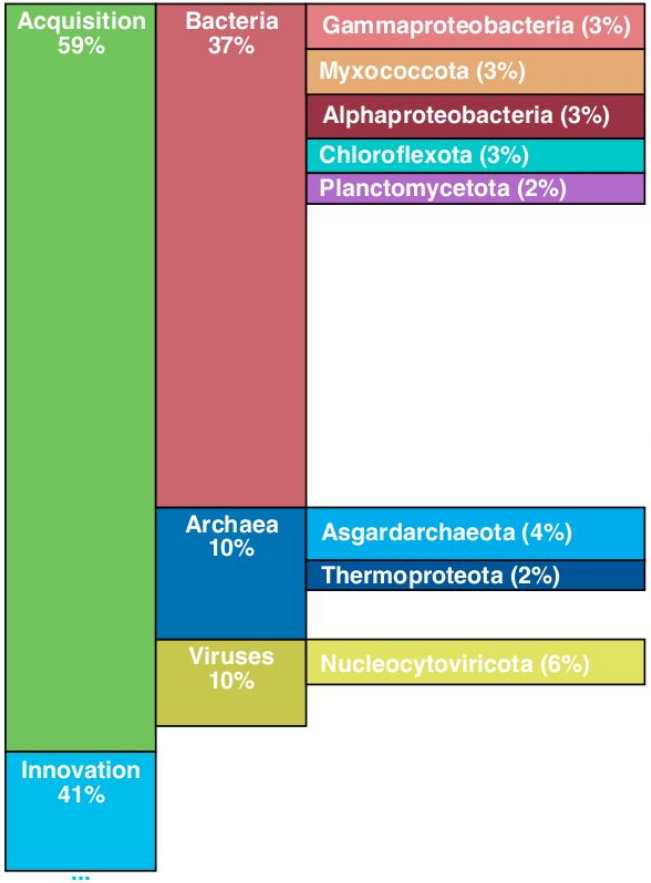

Stem 1
Discoba
PHM
Amorphea
CRuMs
Stem 2
TRASH
Archaeplastida plus
M
A
Metamonada
Tsukubamonadida
Euglenozoa
Heterolobosea
Telonemia
Haptista
Rhizaria
Stramenopiles
Alveolata
Cryptista
Chloroplastida
Glaucophyta
Picozoa
Rhodelphidia
Rhodophyta
Jakobida
Meteora
Hemimastigophora
Provora
Opisthokonta
Apusomonadida
Breviatea
Amoebozoa
Collodictyon
Rigifilida
Mantamonas
Malawimonadida
Ancyromonadida
Fornicata
BaSk
Preaxostyla
Parabasalia
Anaeramoebidae


Tree #1
Tree #2
Tree #3
Tree #4
Tree #5
Tree #6
Tree #7
Tree #8
Tree #9
Tree #10
Tree #11
Tree #12

Profiling the Last Eukaryotic Common Ancestor

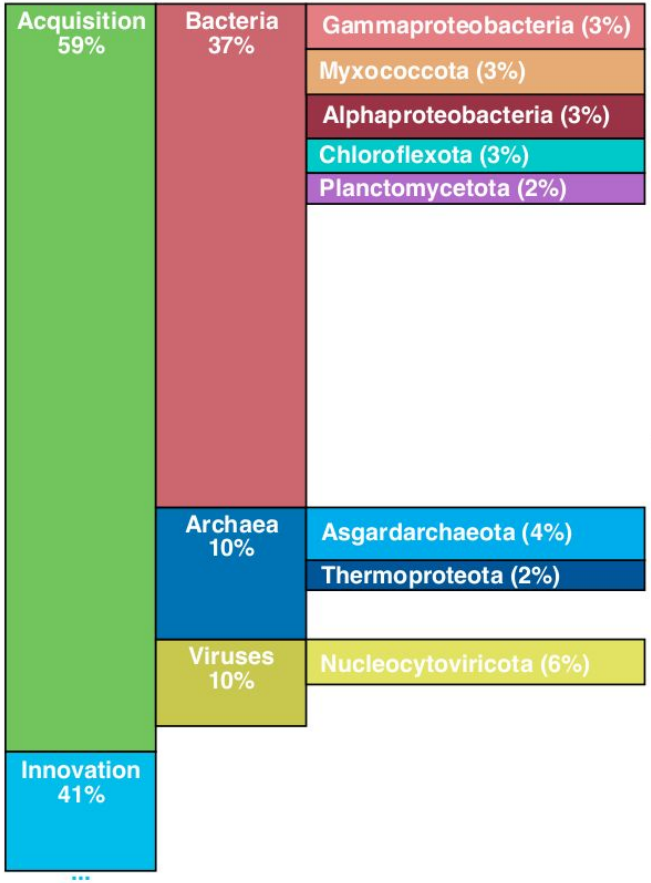


(Bernabeu *et al*., 2024)



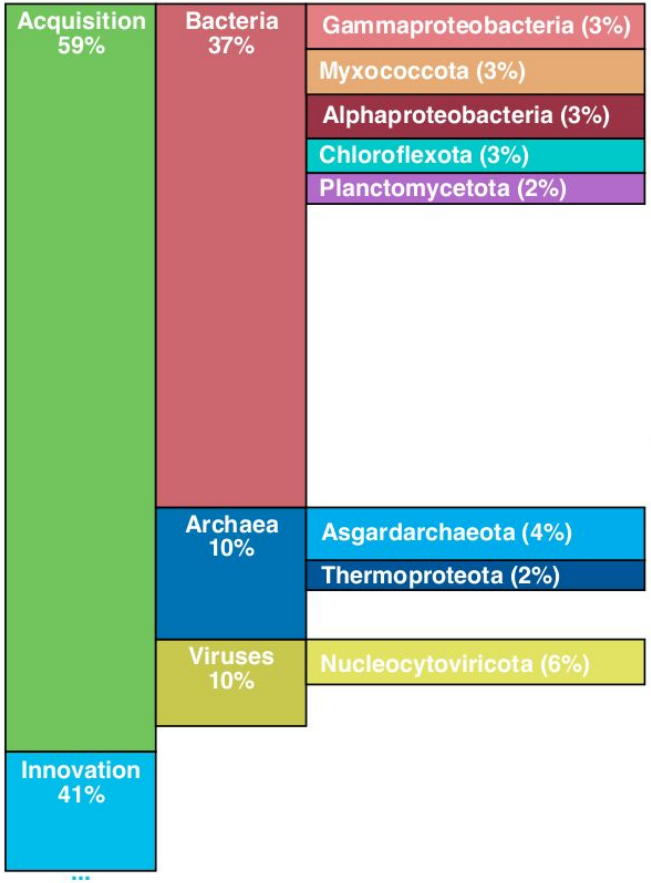
(Bernabeu *et al*., 2024)

- Impact of innovations on LECA proteome



(Bernabeu *et al*., 2024)

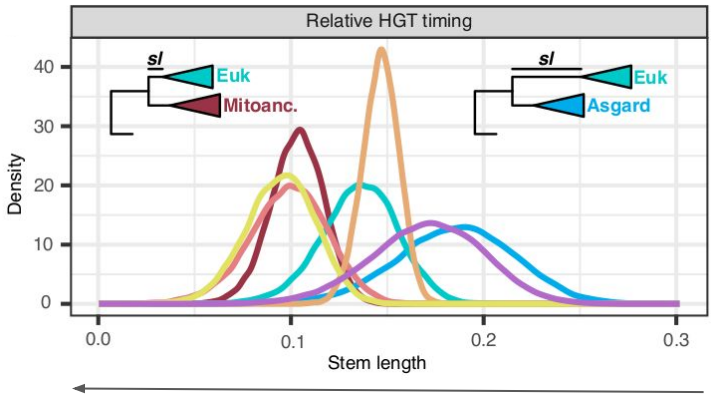
- Impact of innovations on LECA proteome
- Significant contributions from outside Asgard archaea and alpha-proteobacteria



(Bernabeu *et al*., 2024)

- Impact of innovations on LECA proteome
- Significant contributions from outside Asgard archaea and alpha-proteobacteria
- Viral contribution

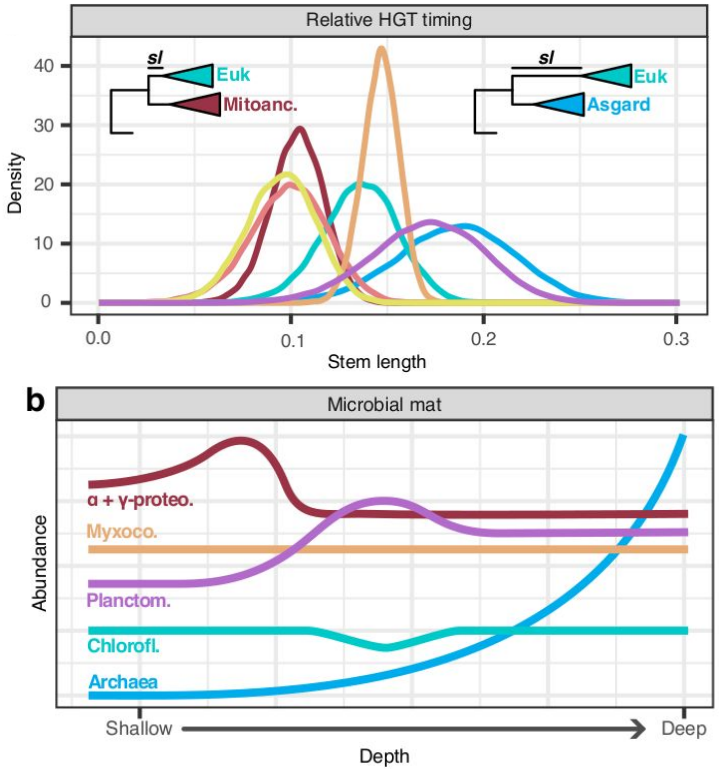
The relative timing of the branch length distributions roughly correlates to the composition of microbial mats



Time

(Bernabeu *et al*., 2024)

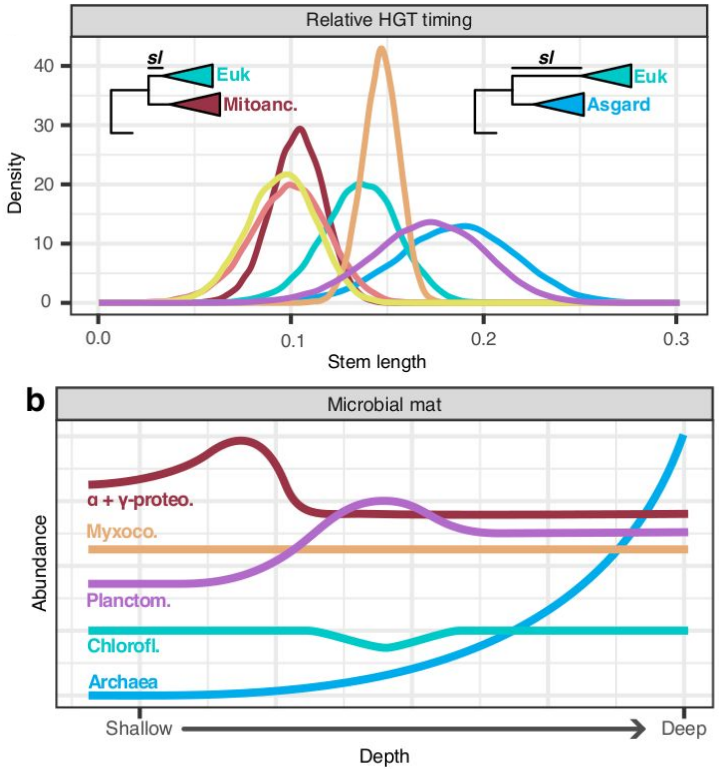
The relative timing of the branch length distributions roughly correlates to the composition of microbial mats



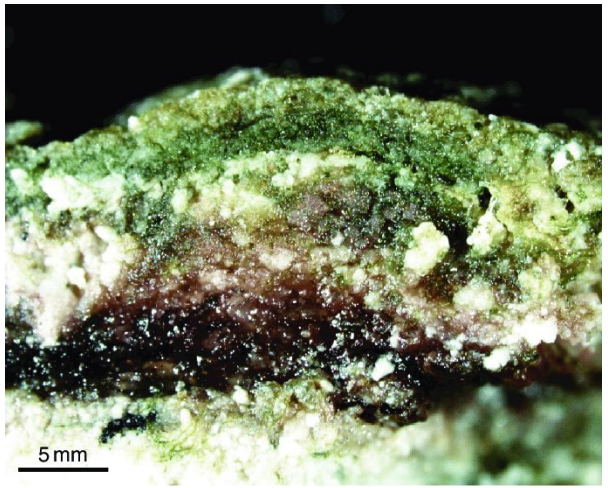
(Bernabeu *et al*., 2024)

(Sanchez-Roman et al., 2009)

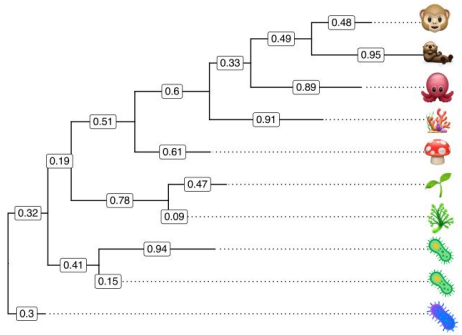
The relative timing of the branch length distributions roughly correlates to the composition of microbial mats



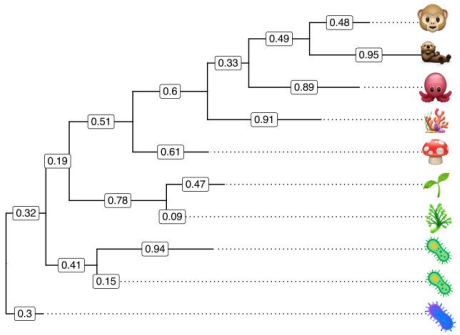
(Bernabeu *et al*., 2024)



(Sanchez-Roman et al., 2009)

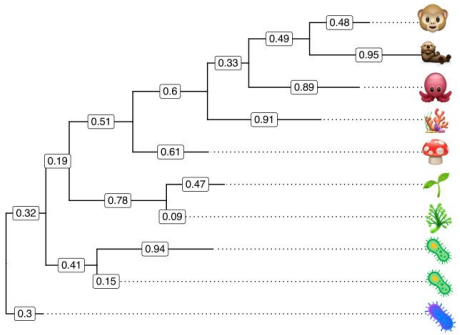

0.48
0.49
0.95
0.33
0.89
0.6
0.91
0.51
0.61
0.19
0.47
0.78
0.09
0.32
0.94
0.41
0.15
0.3

Gene Family



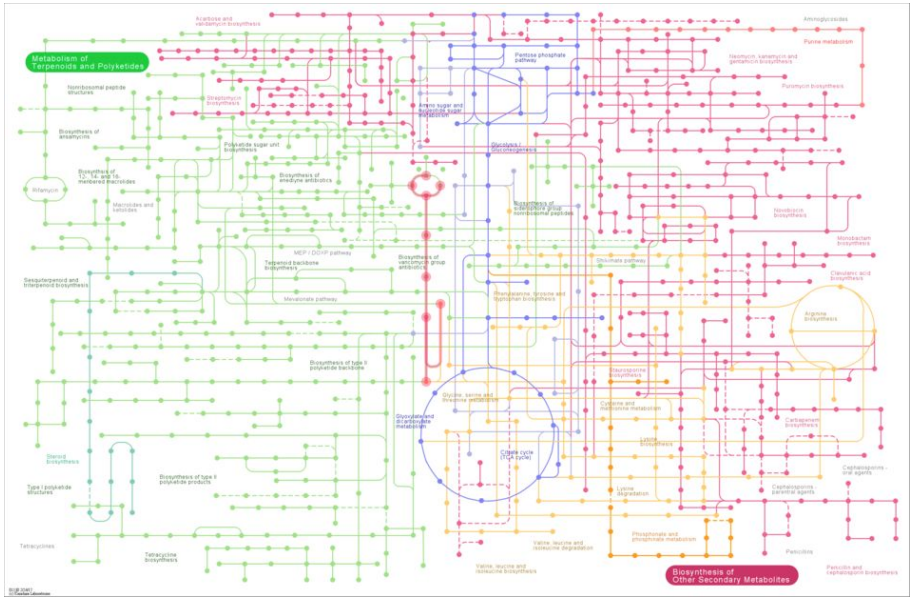
Gene Family

KOFamScan

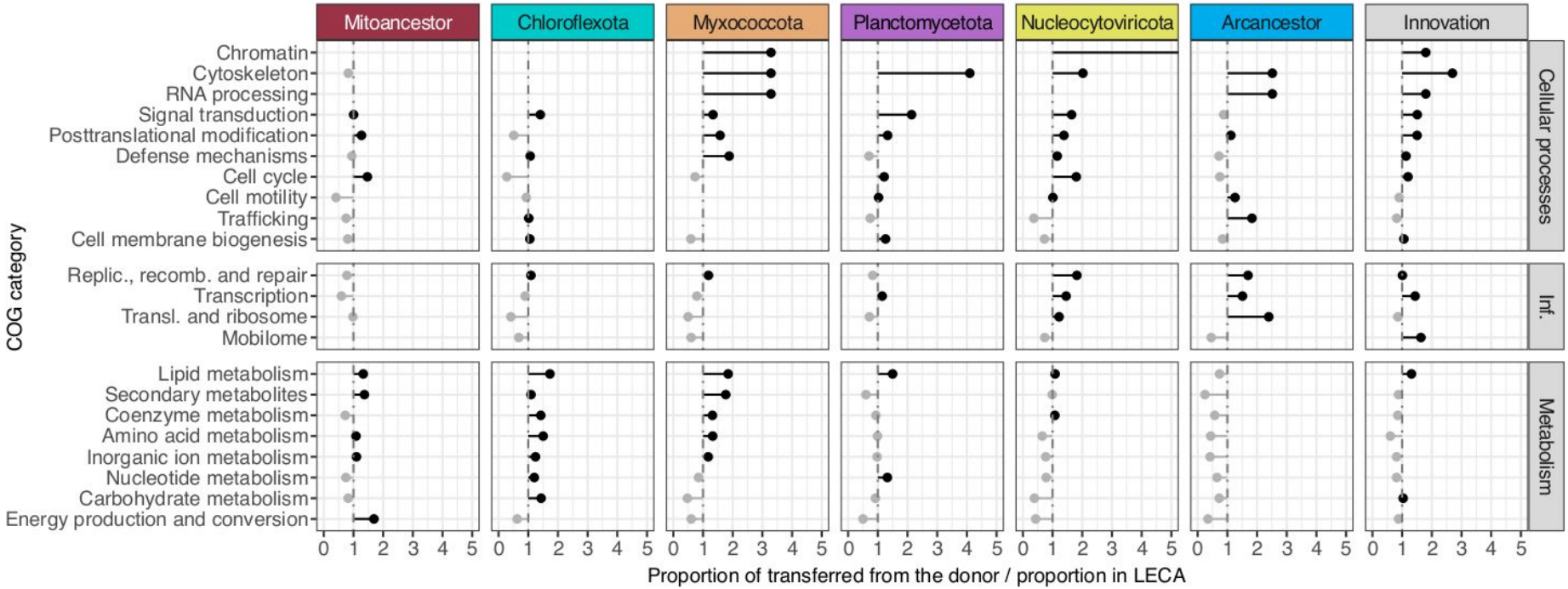


Gene Family

KOFamScan

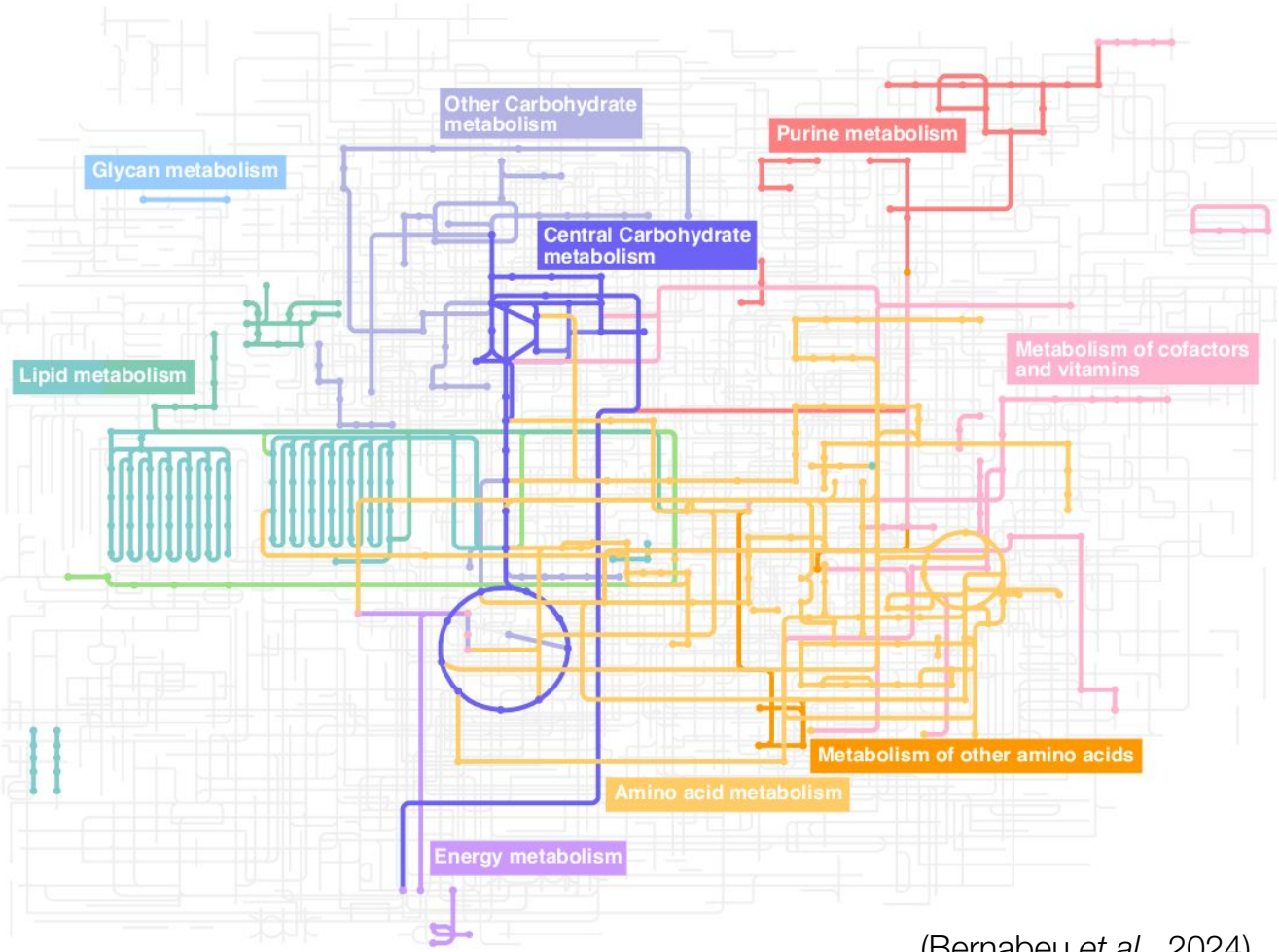


Each donor has left a distinct functional footprint

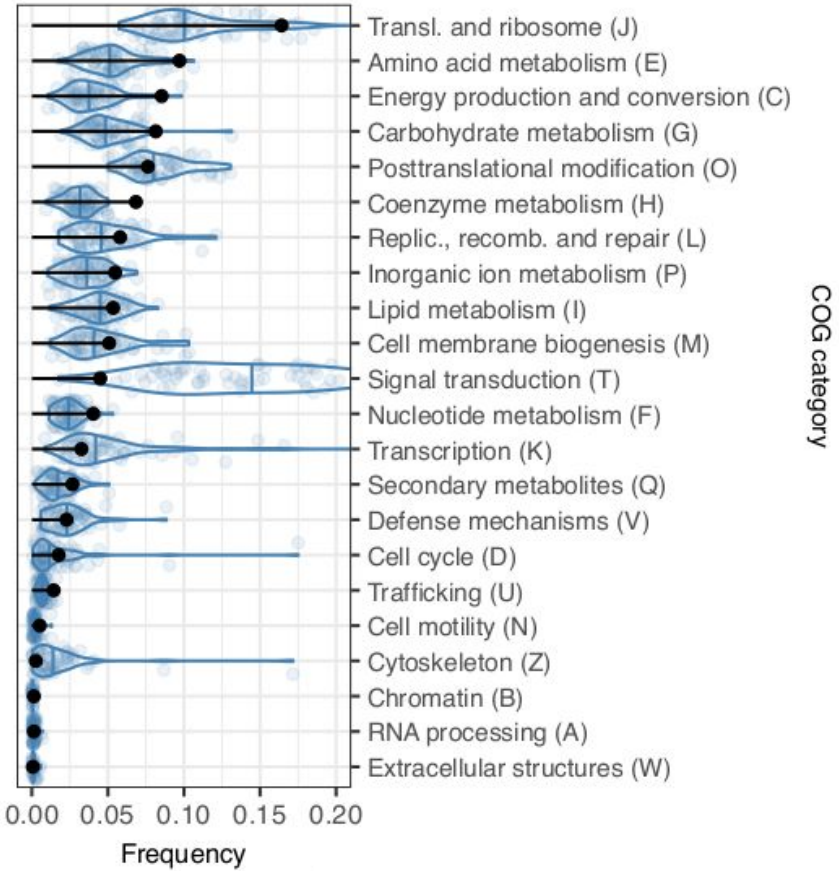


(Bernabeu *et al*., 2024)

LECA was a complex organism, and likely an aerobic heterotroph



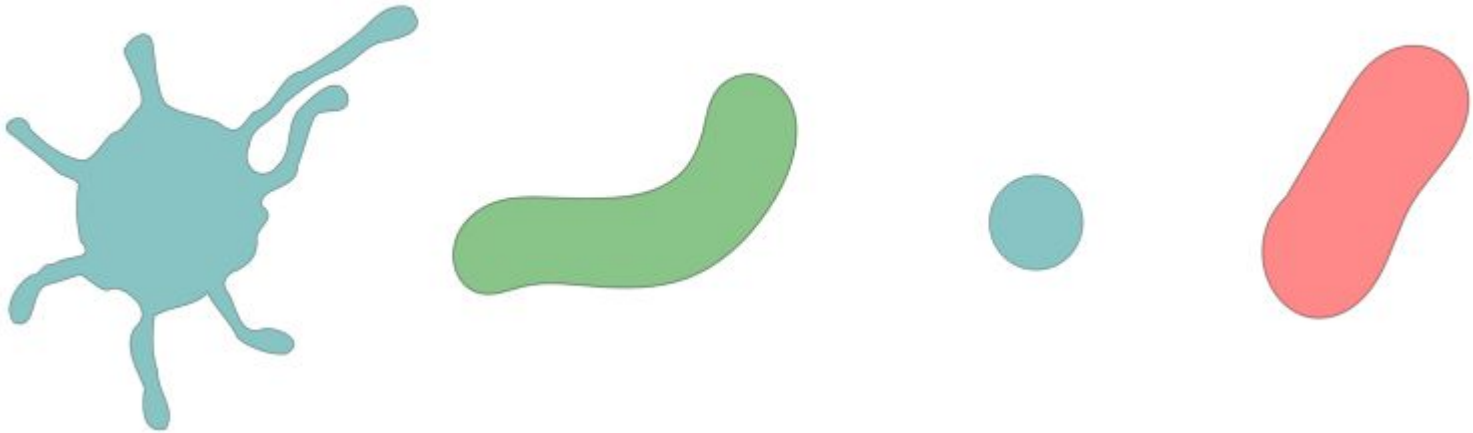
(Bernabeu *et al*., 2024)



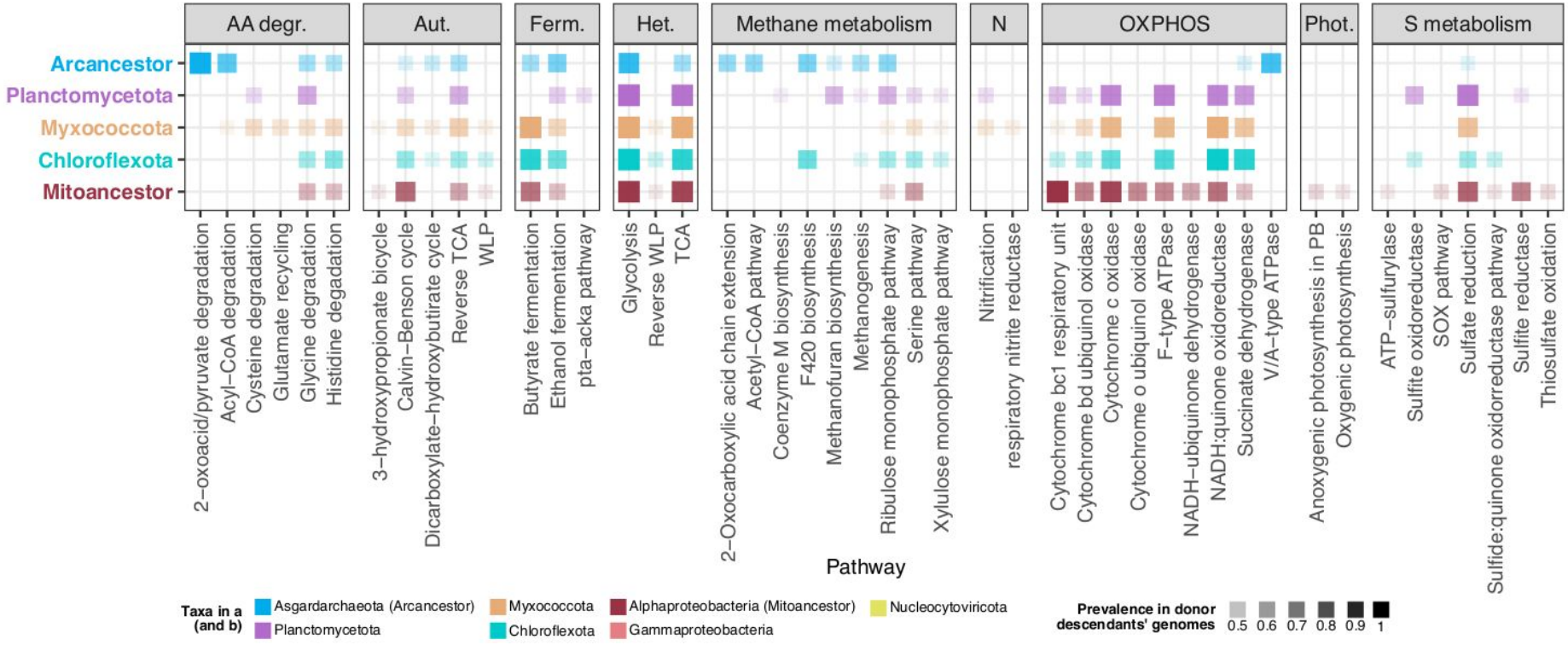
Extant eukaryotes have significantly expanded signaling systems

(Bernabeu *et al*., 2024)

Profiling the donors



The putative metabolism of the donors helps shed light in eukaryogenesis theories



(Bernabeu *et al*., 2024)

TAKE-HOME MESSAGES

1. Diverse gene flow from prokaryotes to the (pre-mitochondrial) proto-eukaryote
2. Significant contribution of viruses: likely as vector
3. Gradual and complex prokaryote-to-eukaryote transition
4. Likely ecological interactions - bacteria-rich environment

TAKE-HOME MESSAGES

1. Diverse gene flow from prokaryotes to the (pre-mitochondrial) proto-eukaryote
2. Significant contribution of viruses: likely as vector
3. Gradual and complex prokaryote-to-eukaryote transition
4. Likely ecological interactions - bacteria-rich environment

TAKE-HOME MESSAGES

1. Diverse gene flow from prokaryotes to the (pre-mitochondrial) proto-eukaryote
2. Significant contribution of viruses: likely as vector
3. Gradual and complex prokaryote-to-eukaryote transition
4. Likely ecological interactions - bacteria-rich environment

TAKE-HOME MESSAGES

1. Diverse gene flow from prokaryotes to the (pre-mitochondrial) proto-eukaryote
2. Significant contribution of viruses: likely as vector
3. Gradual and complex prokaryote-to-eukaryote transition
4. Likely ecological interactions - bacteria-rich environment

Comparative Genomics

Toni Gabaldón

Marina
Marcet-Houben

Moisès
Bernabeu

GA: 724173

GBMF9742

EXCELENCIA SEVERO OCHOA

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

IRB BARCELONA

Life Sciences Department Seminar

Diverse ancestries reveal complex symbiotic interactions during eukaryogenesis

Saioa Manzano-Morales
Toni Gabaldón
Comparative Genomics - Life Sciences
Barcelona Supercomputing Center