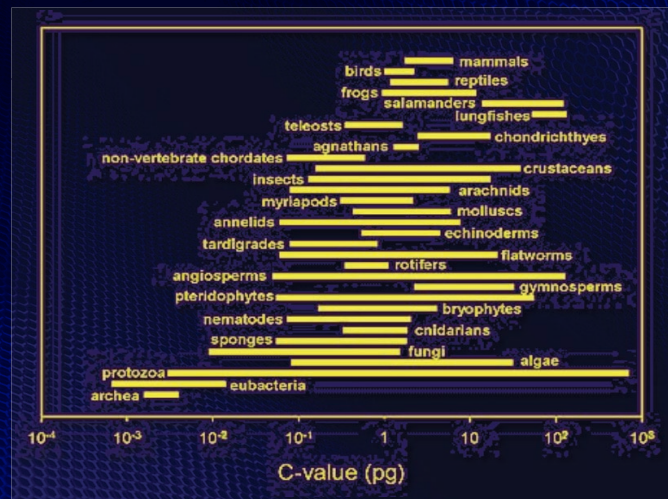


The Problem

“Complexity is hard to define or measure...”

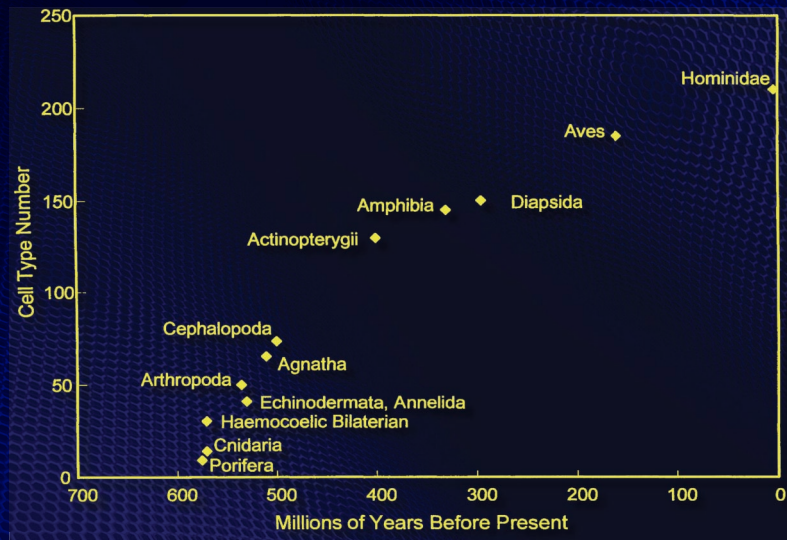
Maynard Smith and Szathmáry, 1995

Genome Size and the Scala Naturae



(from Gregory 2005)

Cell Number and the Scala Naturae

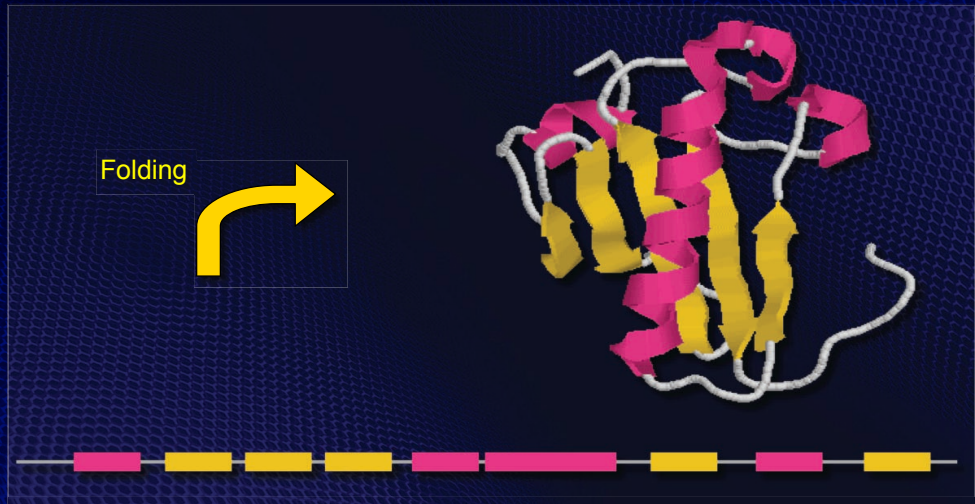


(from Valentine *et al.* 1994)

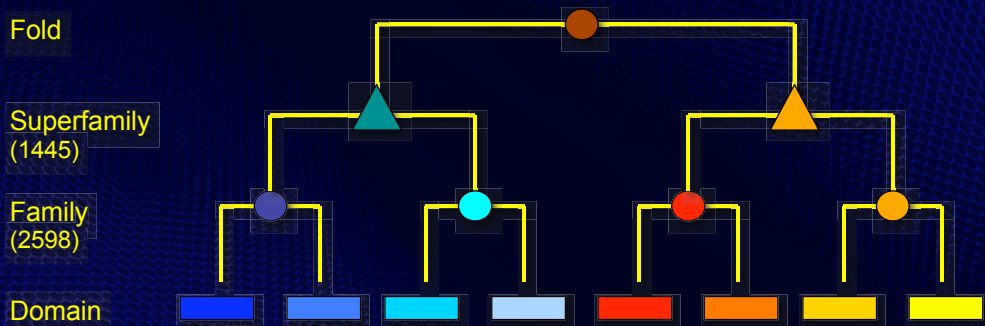
Proteome Size and the Scala Naturae

?

Protein Structural Domains



Protein Domain Classification I



Protein Domain Classification II



Protein 1:  Architecture = A,A,C

Protein 2:  Architecture = D,B

Protein 3:  Architecture = B

Protein 4:  Architecture = C,A,A

Protein 5:  Architecture = D,B,C

Protein Domain Database I

Superfamily ^{1.69}
HMM library and genome assignments server

[PDZ domain-like superfamily assignments to *Homo sapiens* 44_36f \(all transcripts\).](#)

The key to the domain colours is at the bottom of the page.

ENSP00000320758

ENSP00000337459

ENSP00000363475

ENSP00000339862

ENSP00000293813

ENSP00000370298

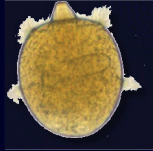
ENSP00000194900

<http://supfam.org>

Protein Domain Database II



238 Bacteria



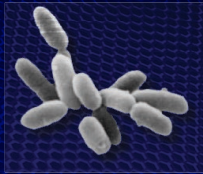
26 Basal Eukaryotes



53 Fungi



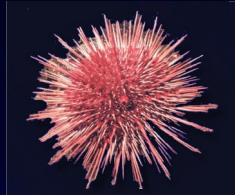
8 Arthropods



27 Archaea



9 Plants

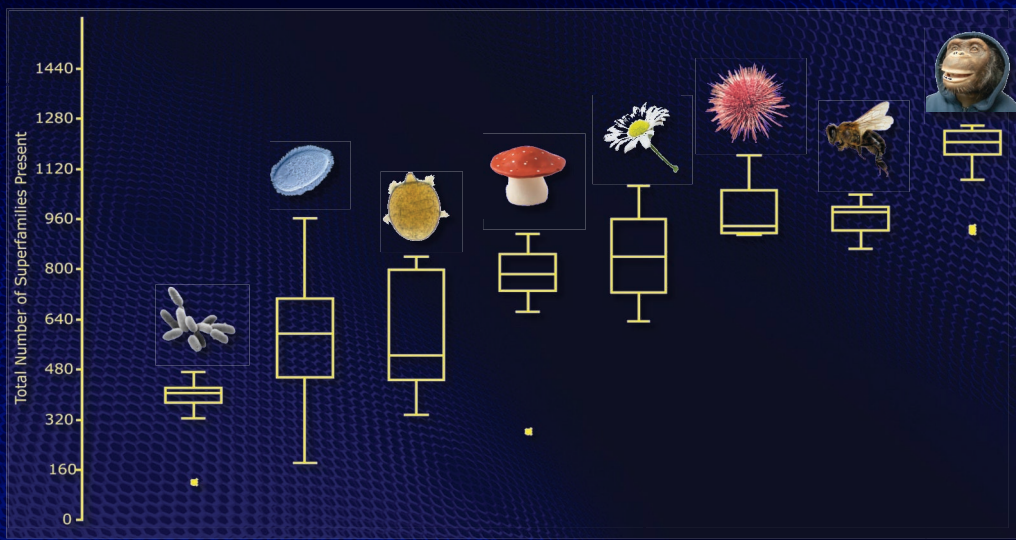


5 Invertebrates (excl. Arthropods)

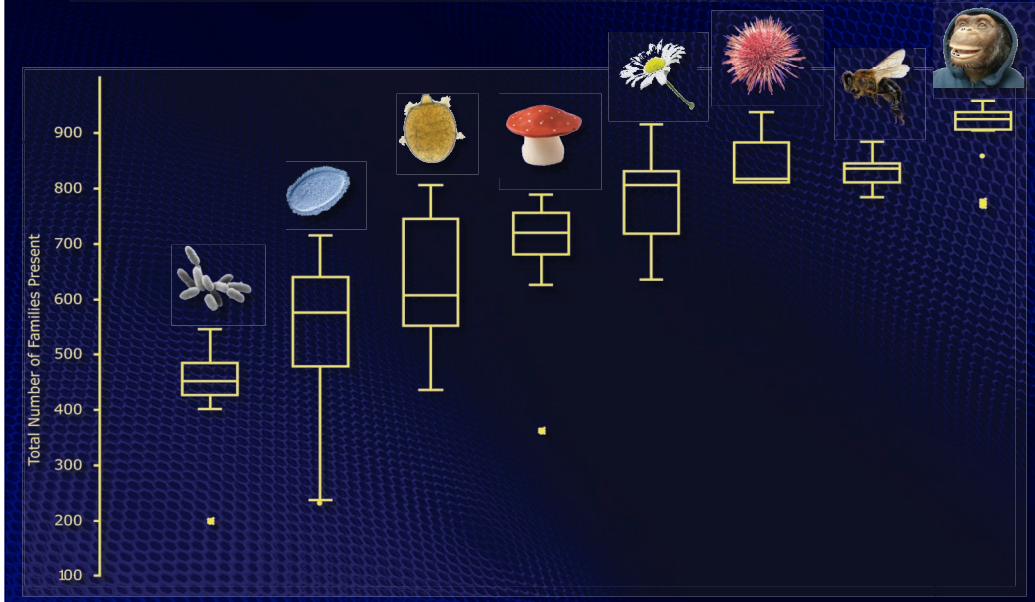


19 Chordates

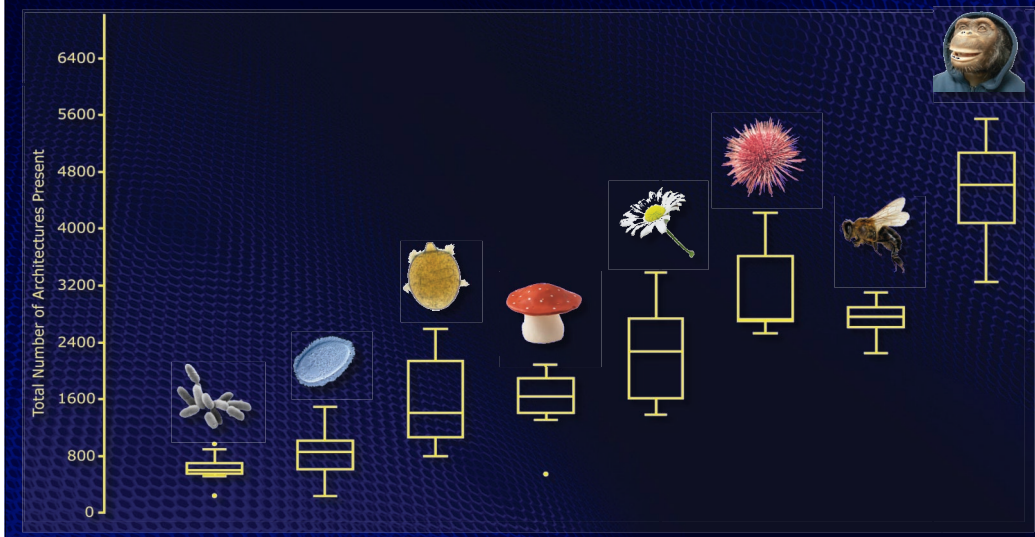
Superfamily Number and the Scala Naturae



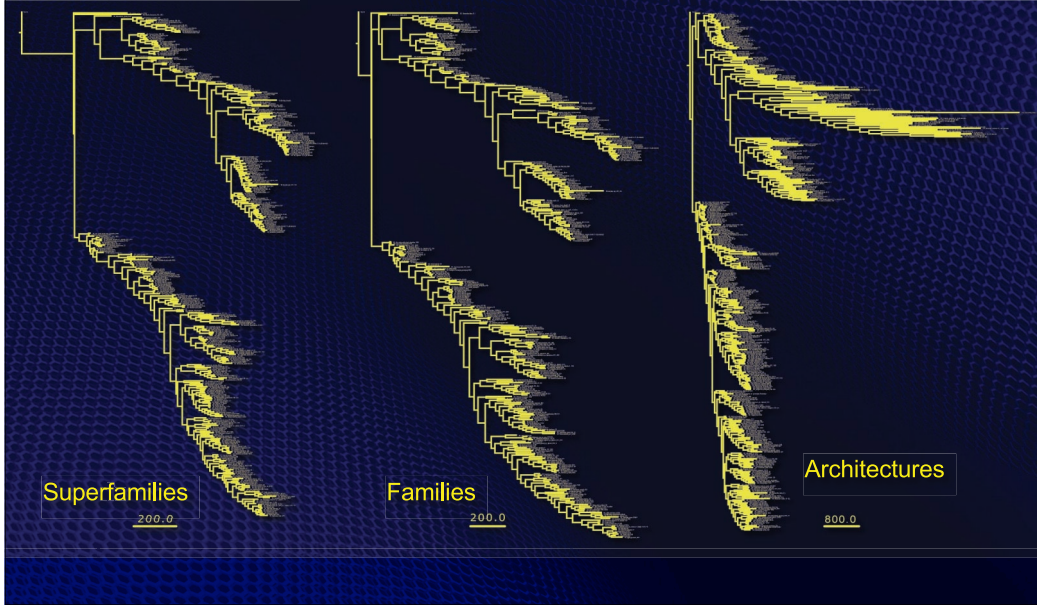
Family Number and the Scala Naturae



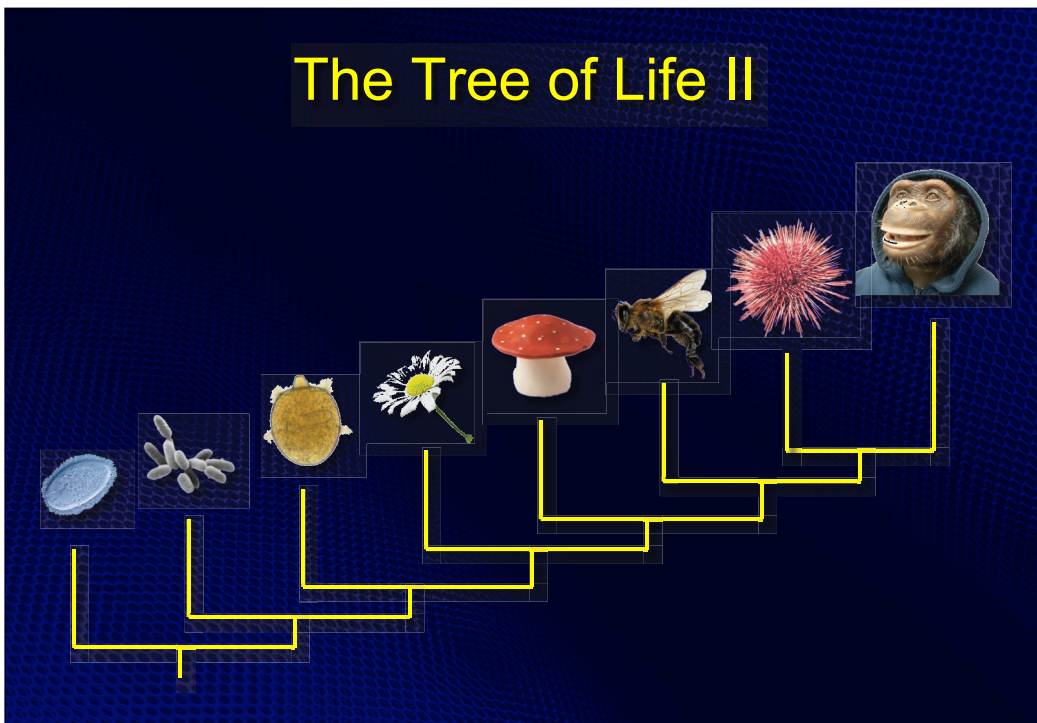
Architecture Number and the Scala Naturae



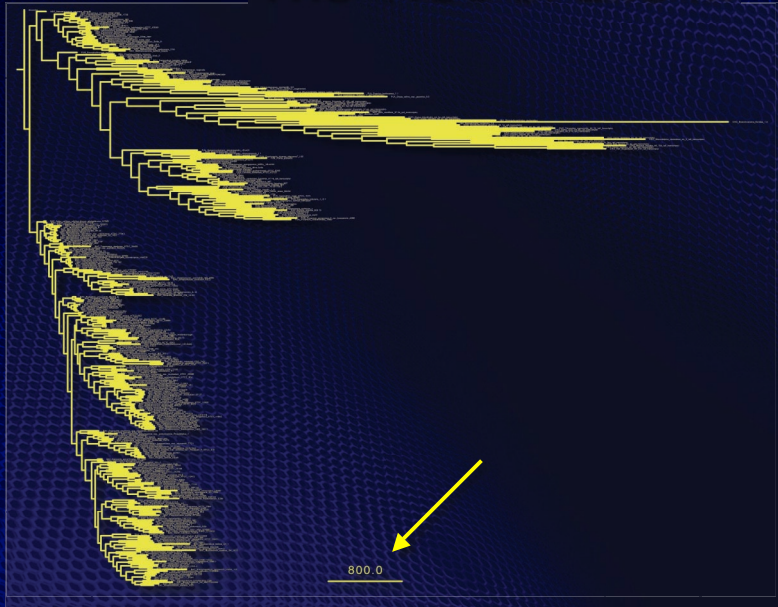
The Tree of Life I



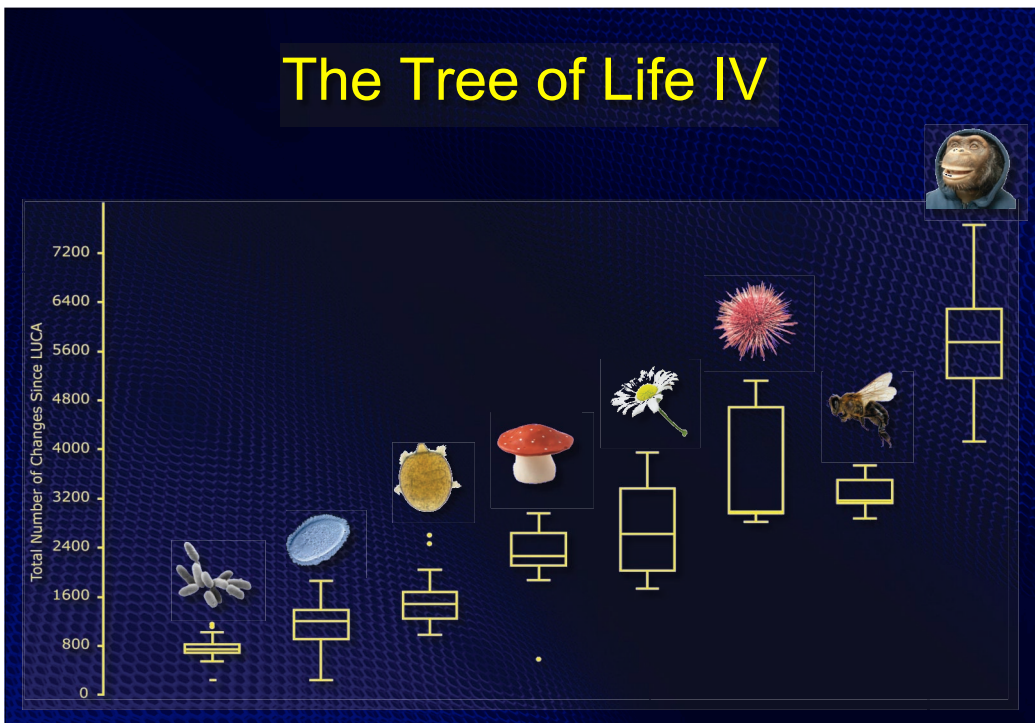
The Tree of Life II



The Tree of Life III



The Tree of Life IV

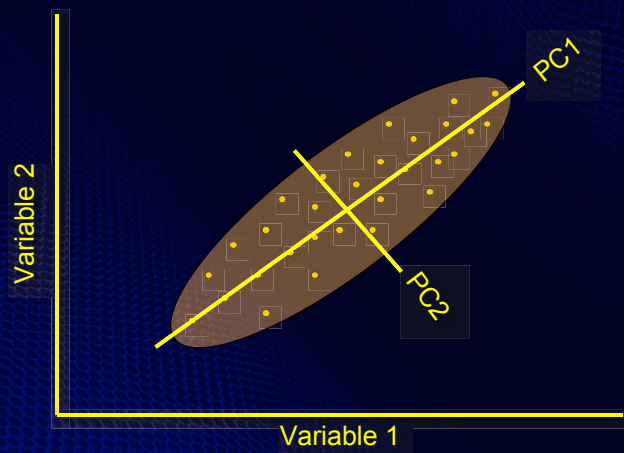


Protein Space I

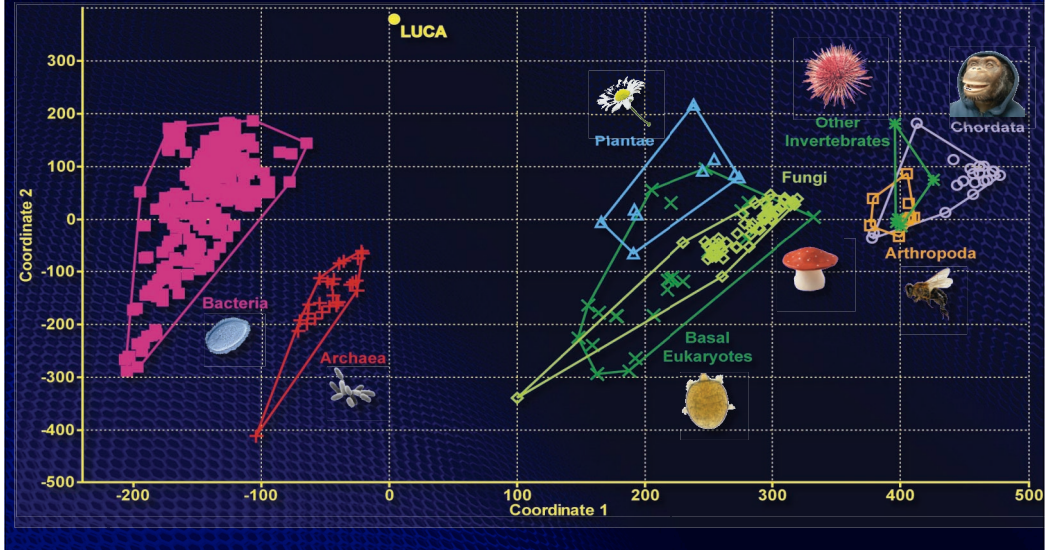


Protein Space II

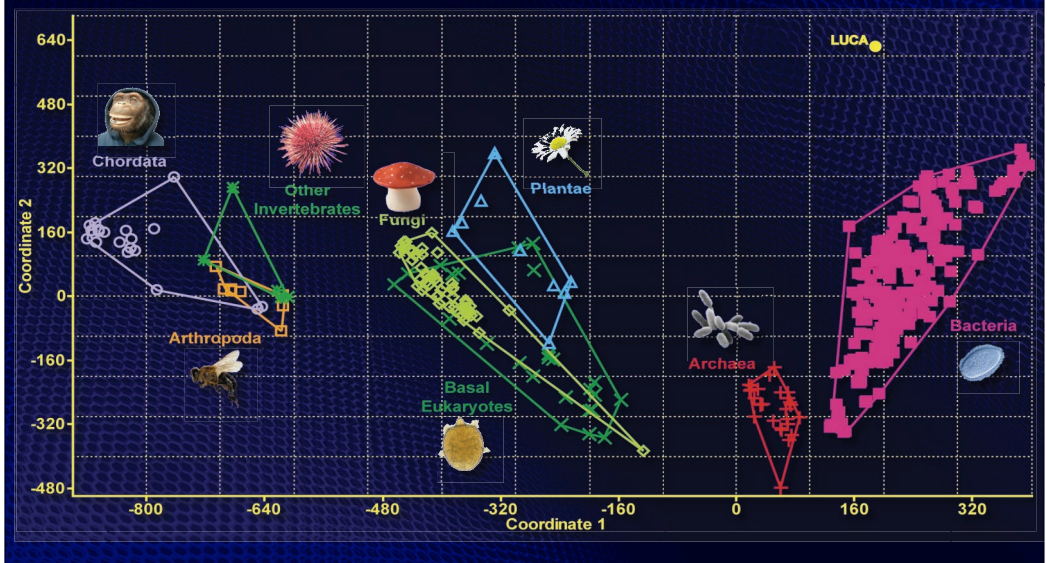
Principal Coordinate Analysis – A Very Brief Introduction



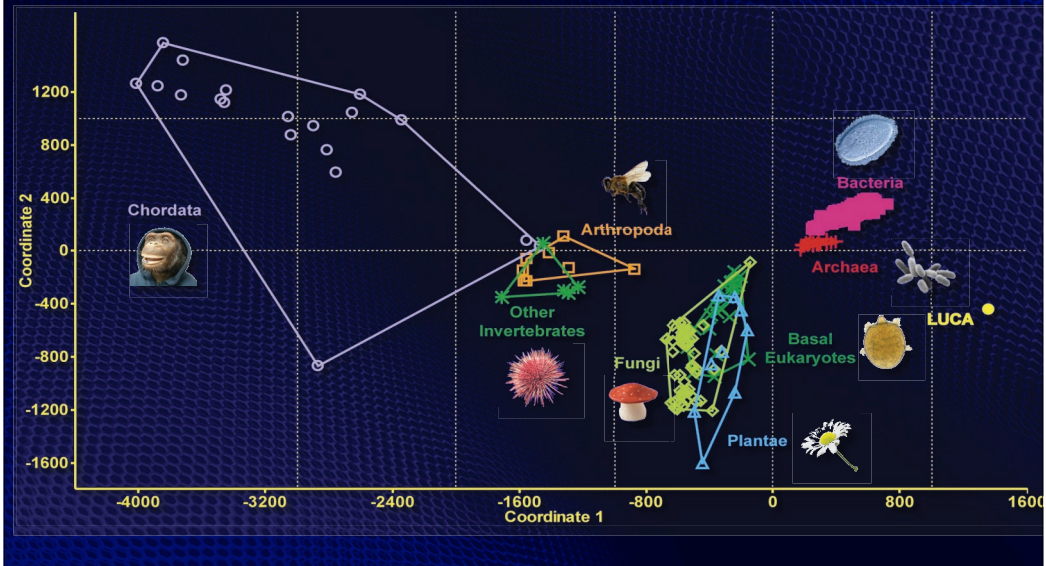
Protein Space III - Superfamilies



Protein Space IV - Families



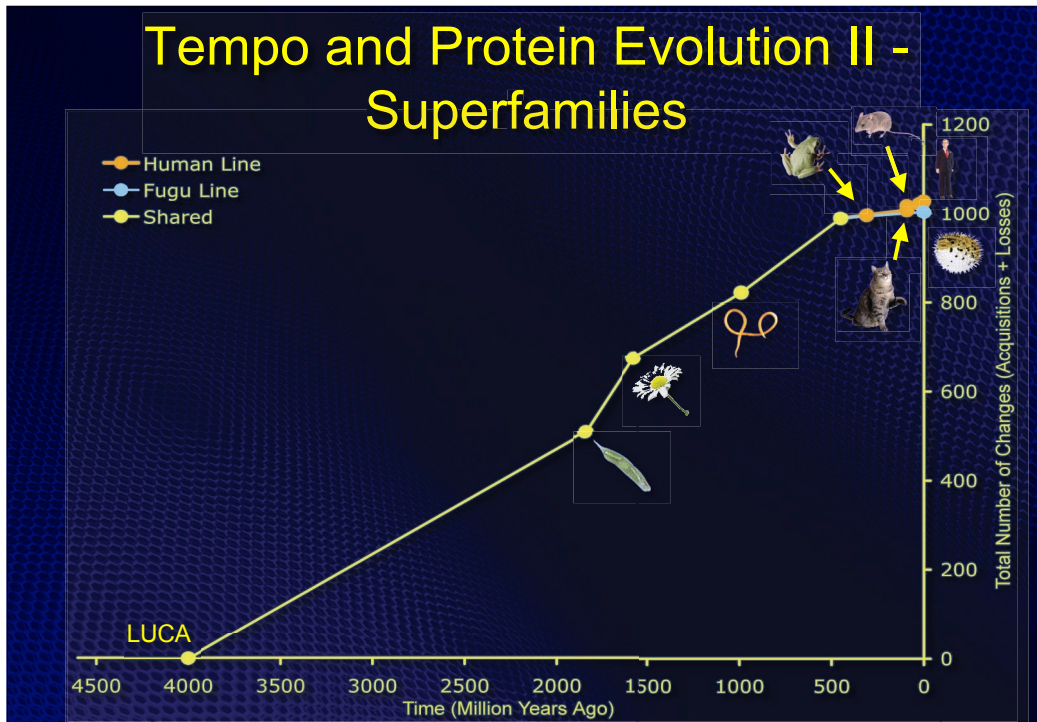
Protein Space V - Architectures



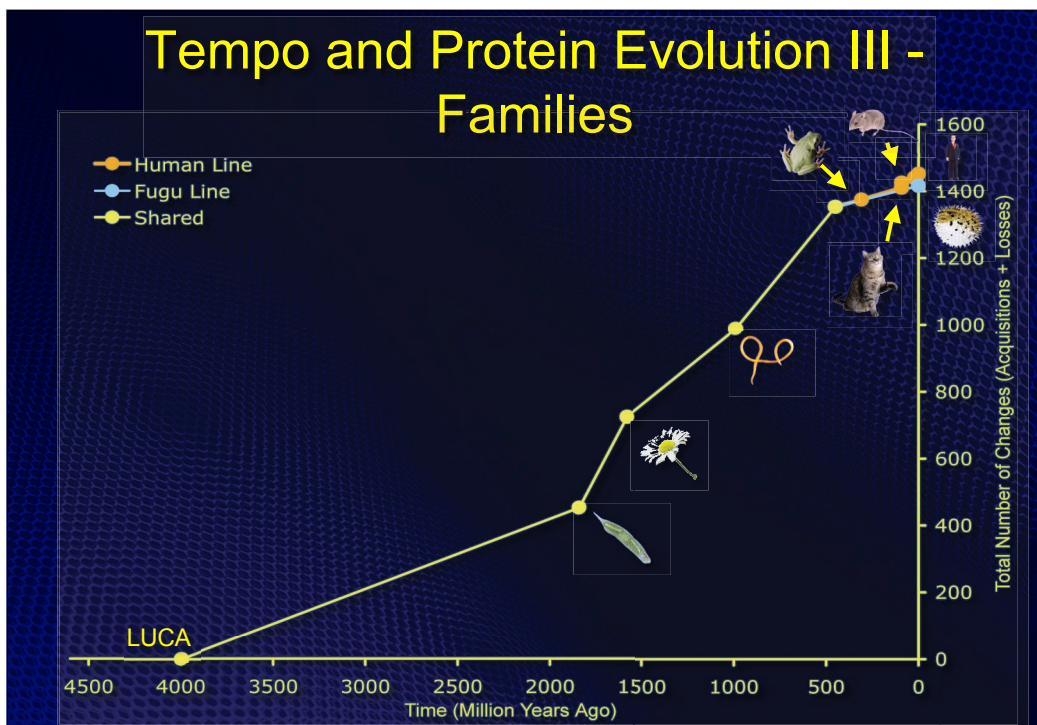
Tempo and Protein Evolution I



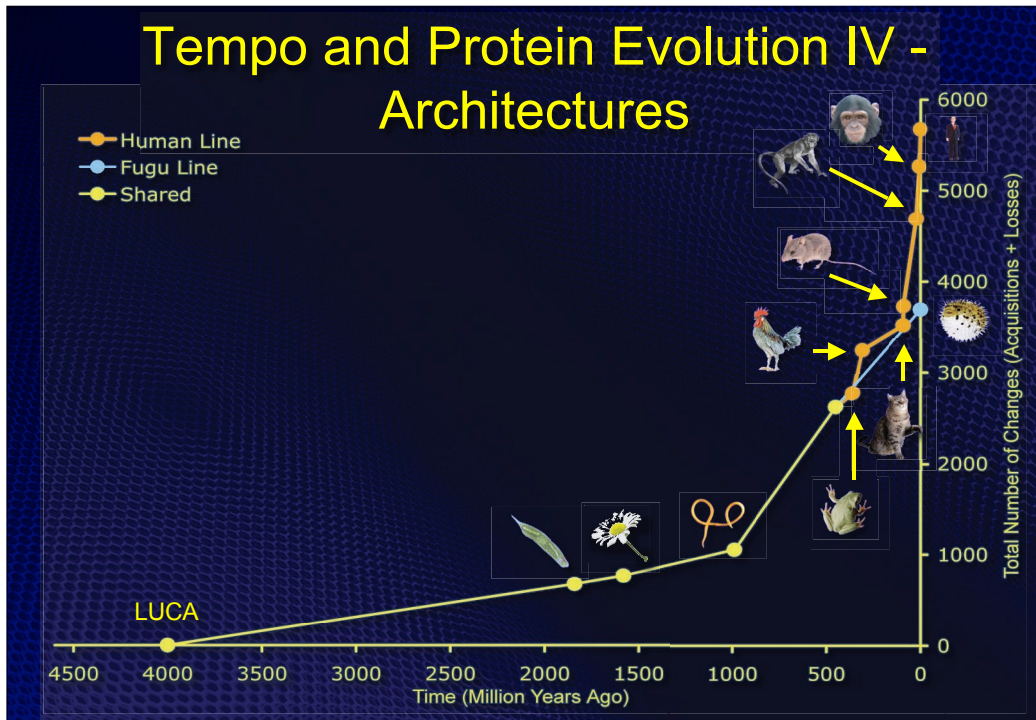
Tempo and Protein Evolution II - Superfamilies



Tempo and Protein Evolution III - Families



Tempo and Protein Evolution IV - Architectures



Conclusions

- Protein domains can help us reconstruct the tree of life and offer a promising new metric for biologic complexity
- Novel domain families and superfamilies have appeared at a roughly constant rate over the history of life and across different groups
- By contrast, novel architectures (new proteins) are acquired at a faster rate in animals, and vertebrates in particular