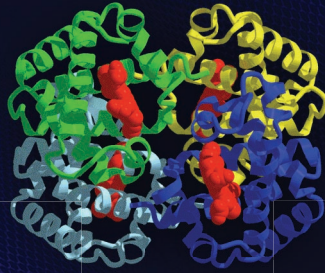


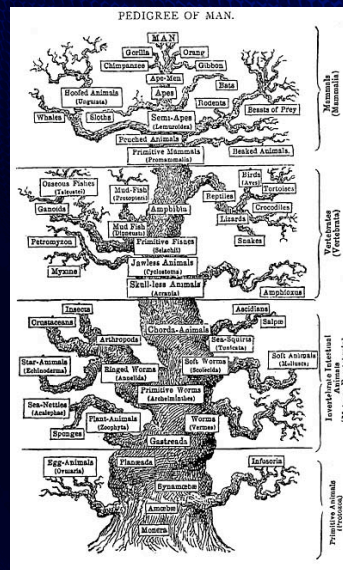
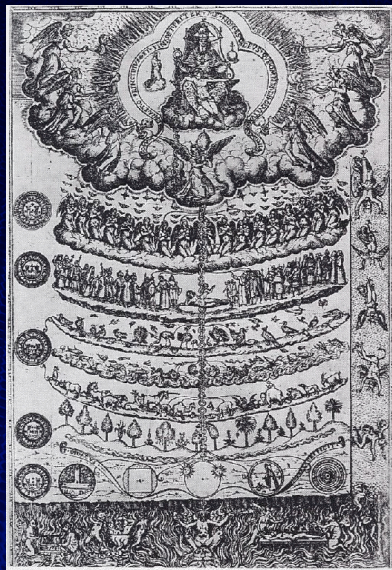
Adventures in protein world reveal the Tree of Life and the evolution of complexity



Graeme T. Lloyd, Philip C. J. Donoghue and Julian Gough



Scala Naturae

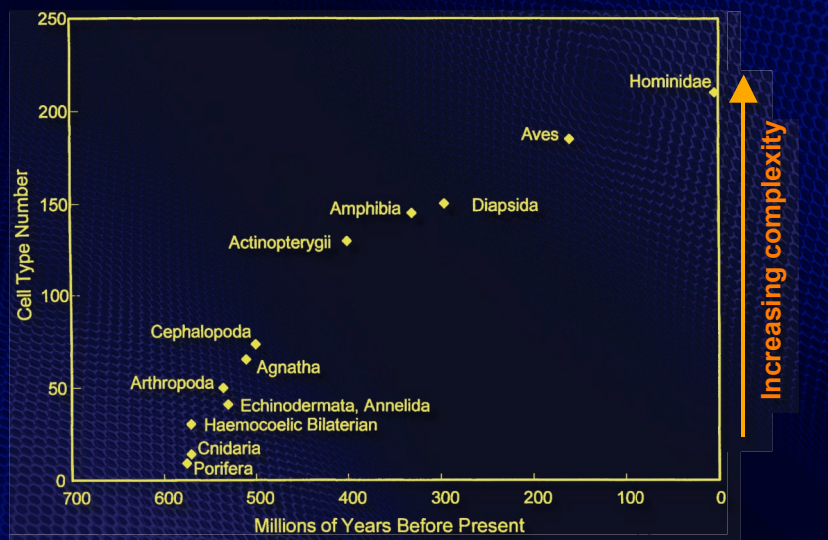


The Problem

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

Maynard Smith and Szathmáry, 1995

Cell Number and Complexity

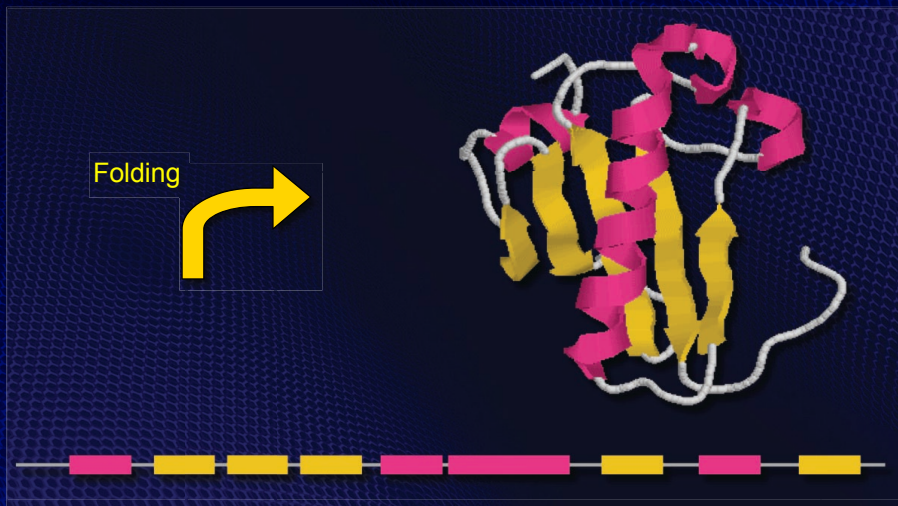


(from Valentine *et al.* 1994)

Proteome Size and the Scala Naturae

?

Protein Structural Domains



SCOP Protein Domain Classification

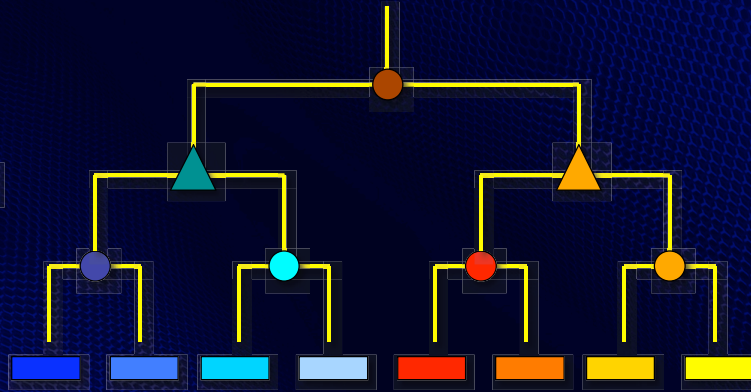
Classes
(7)

Folds
(48)

Superfamily
(1445)

Family
(2598)

Domain
(75930)



Protein Domain Classification



Protein 1: — [A] — [A] — [C] — Architecture = A,A,C

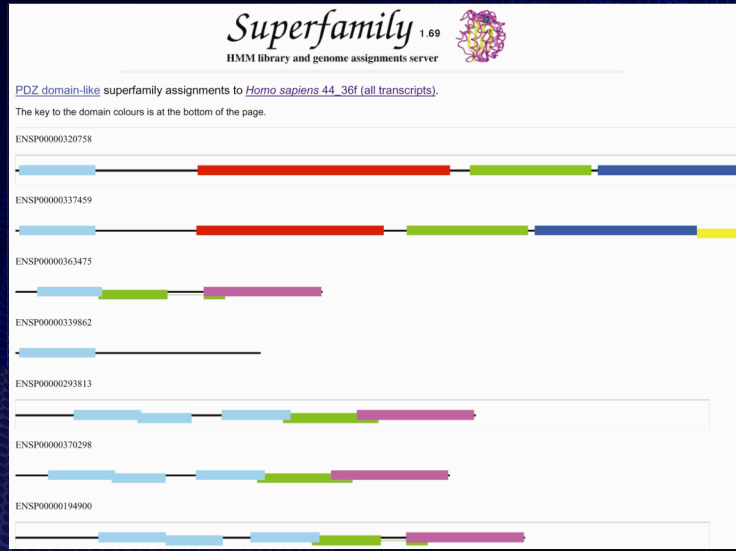
Protein 2: — [D] — — [B] — Architecture = D,?,B

Protein 3: — [B] — Architecture = B

Protein 4: — [C] — [A] — [A] — Architecture = C,A,A

Protein 5: — [D] — [B] — Architecture = D,B

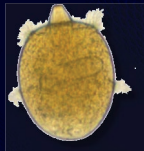
Protein Domain Database I



Protein Domain Database II



238 Bacteria



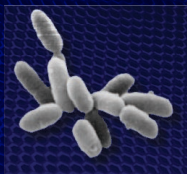
26 Basal Eukaryotes



53 Fungi



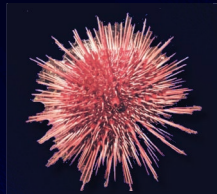
8 Arthropods



27 Archaea



9 Plants

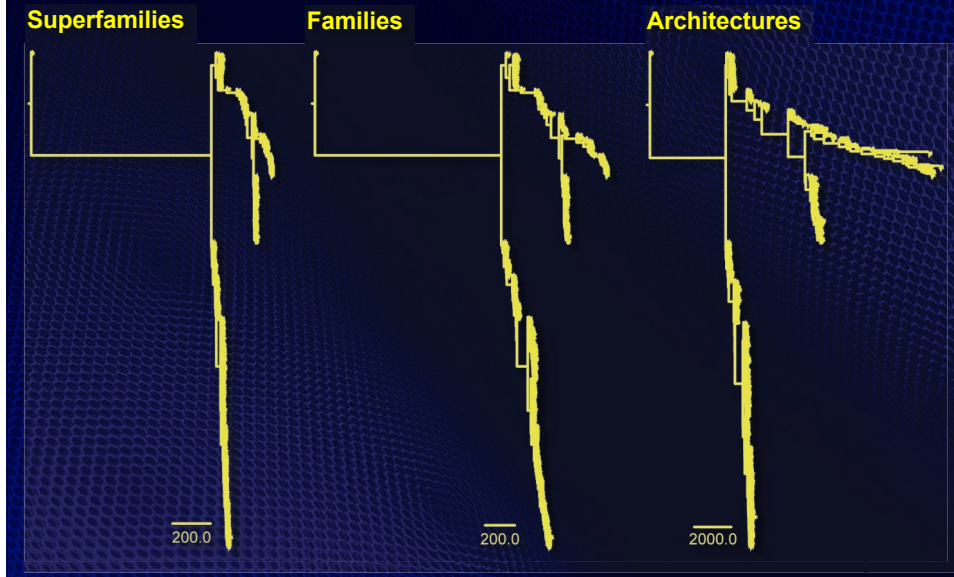


5 Invertebrates (excl. Arthropods)

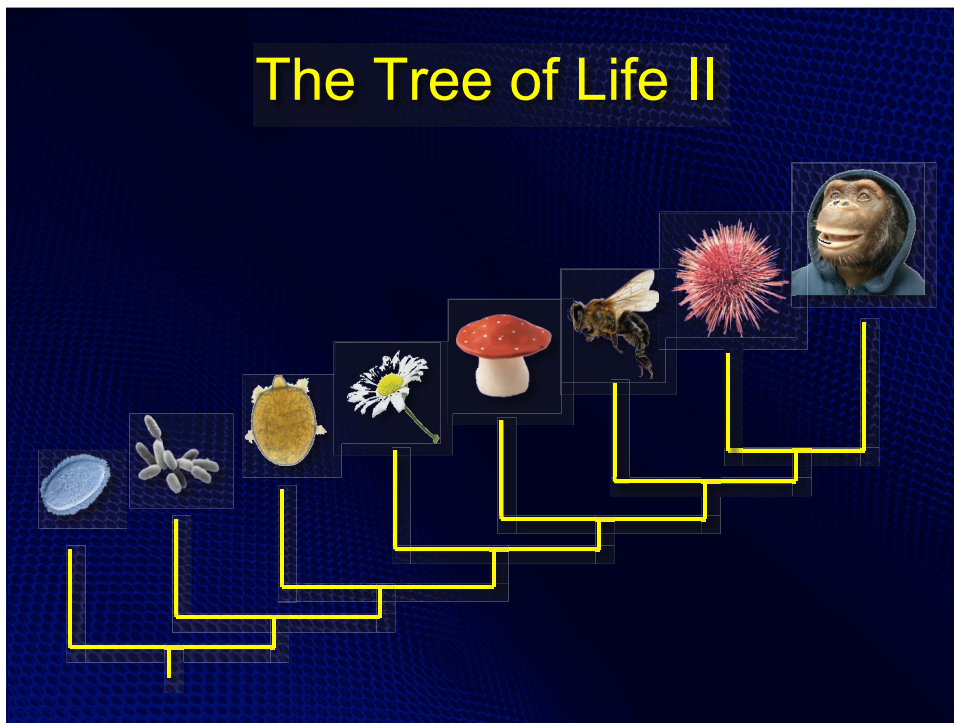


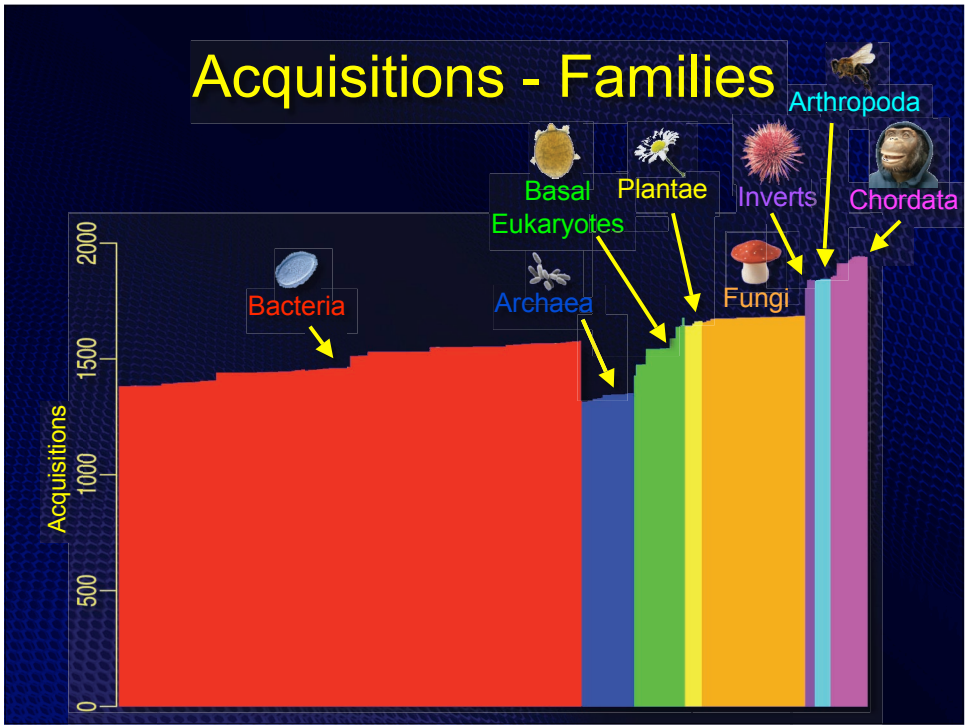
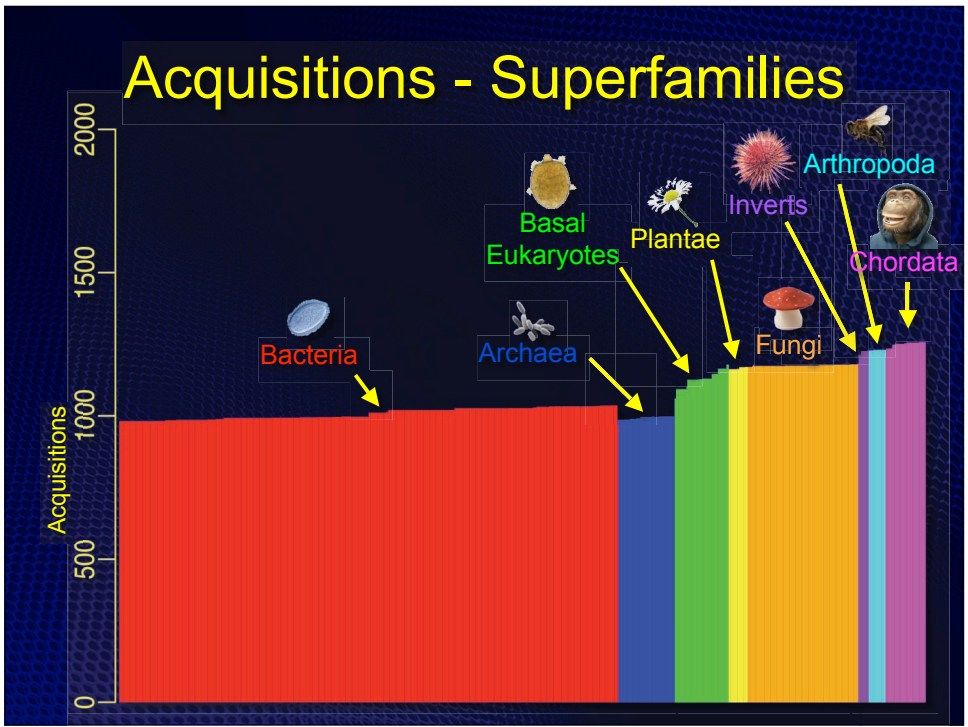
19 Chordates

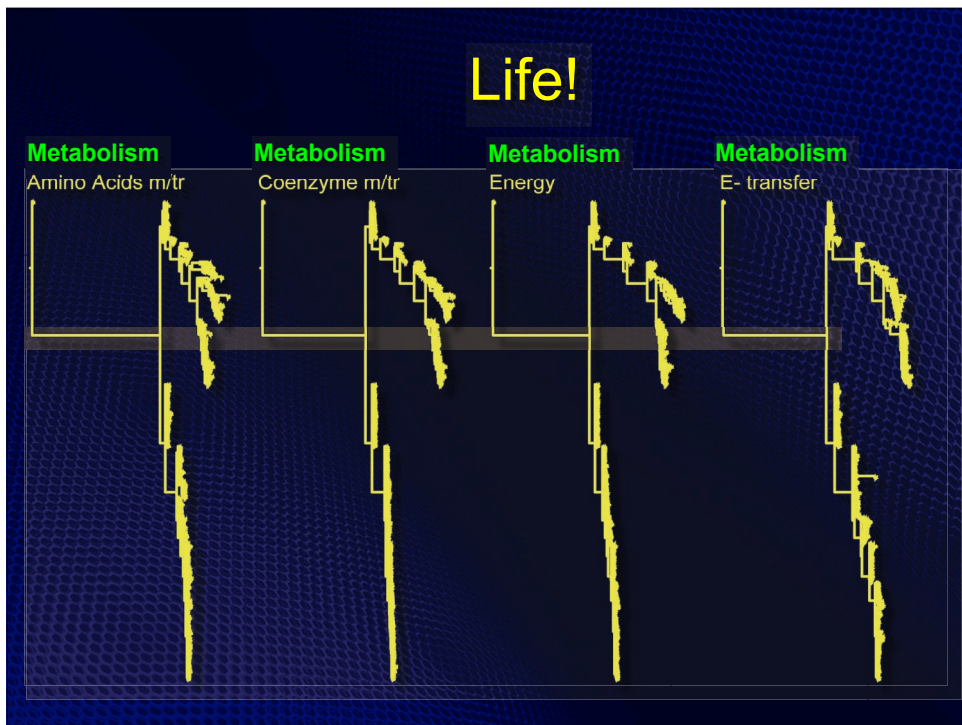
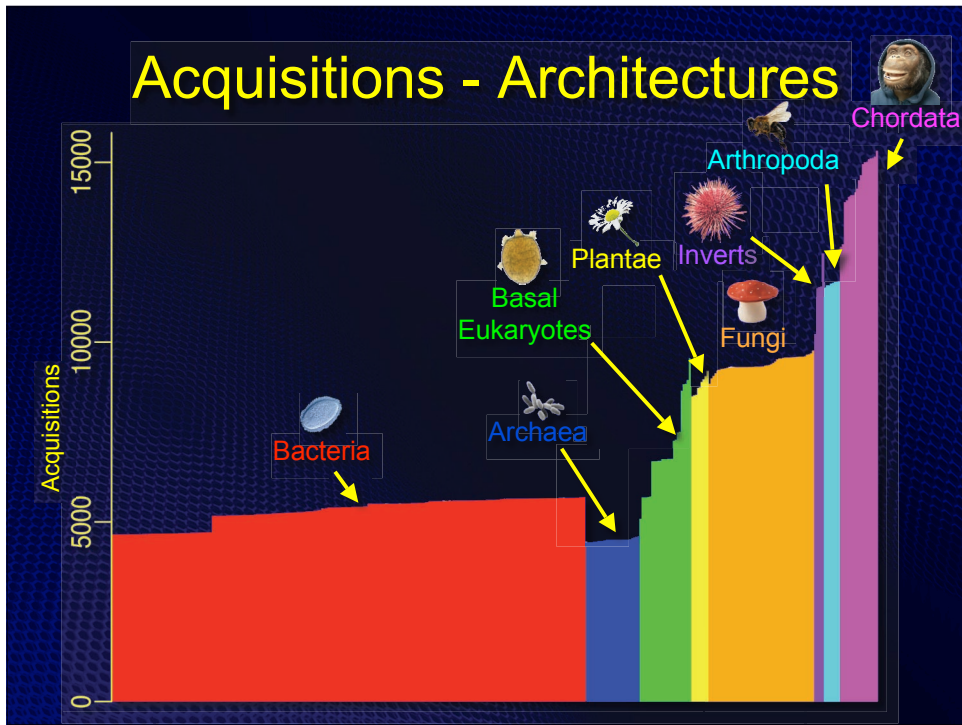
The Tree of Life I



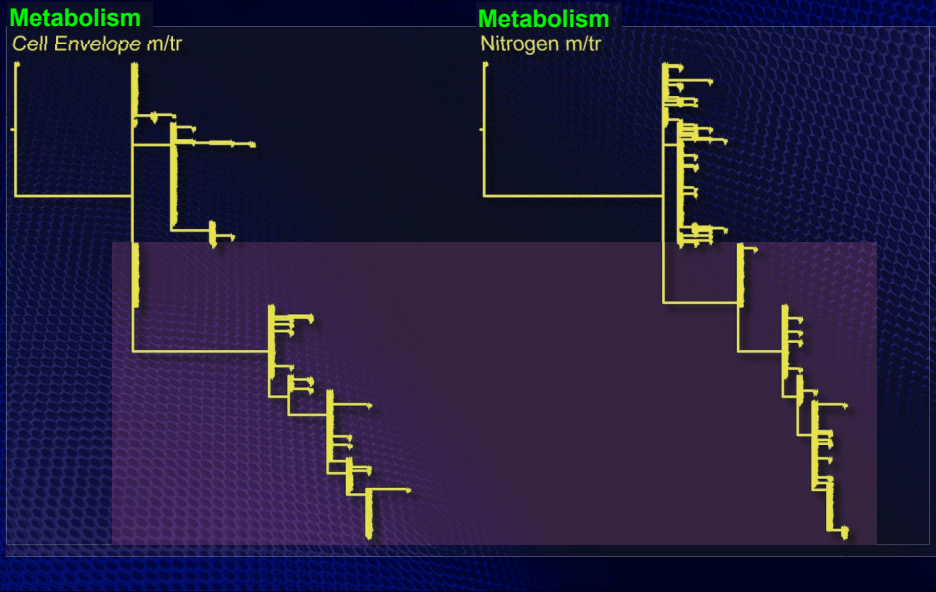
The Tree of Life II



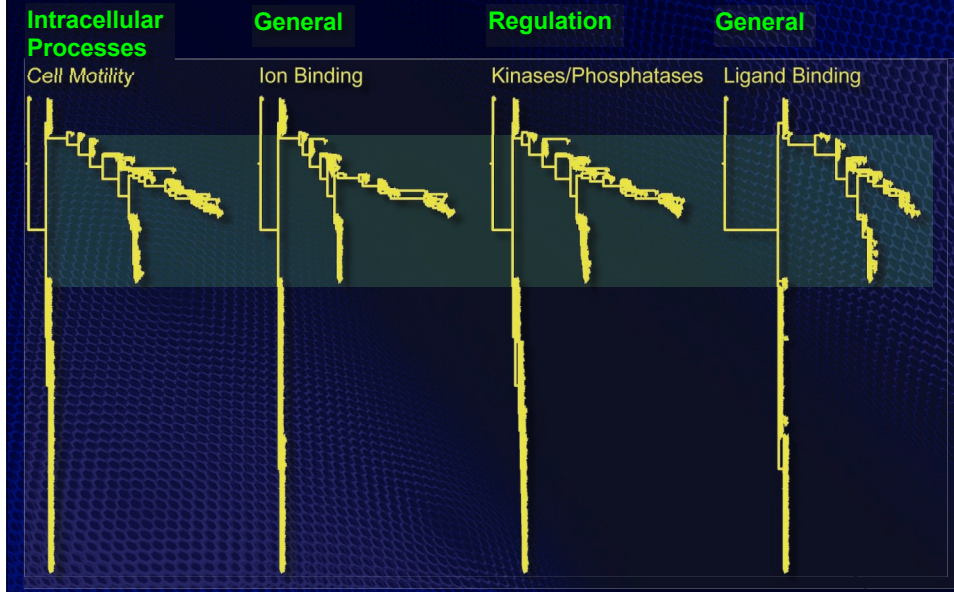




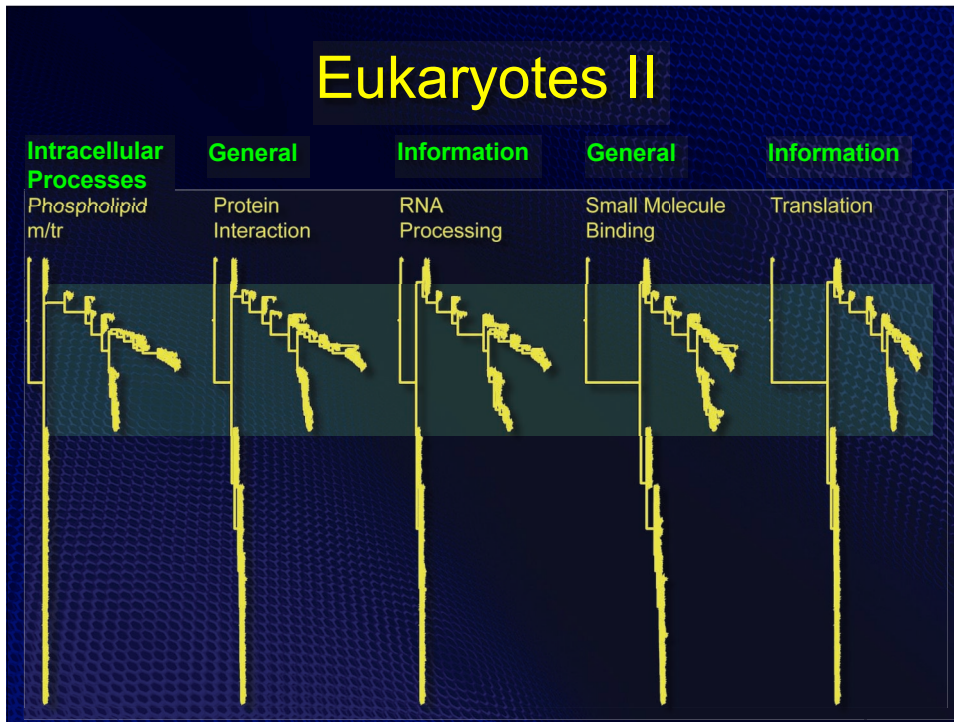
Bacteria



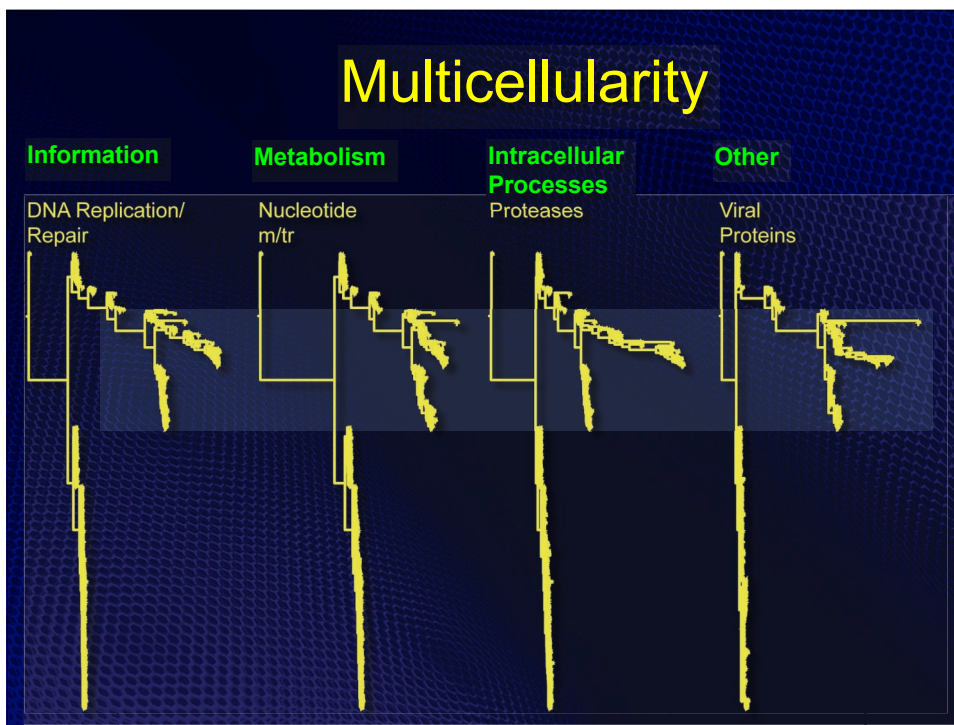
Eukaryotes I



Eukaryotes II



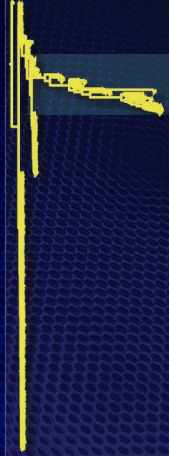
Multicellularity



Metazoa

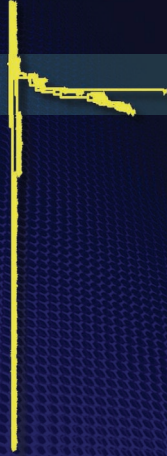
Extracellular Processes

Cell Adhesion



Extracellular Processes

Immune Response



General

Lipid/Membrane Binding



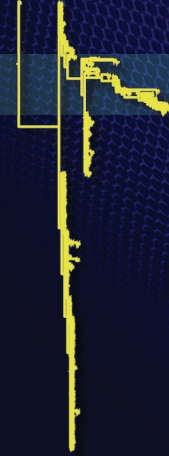
Regulation

Receptor Activity



Extracellular Processes

Toxins/Defence

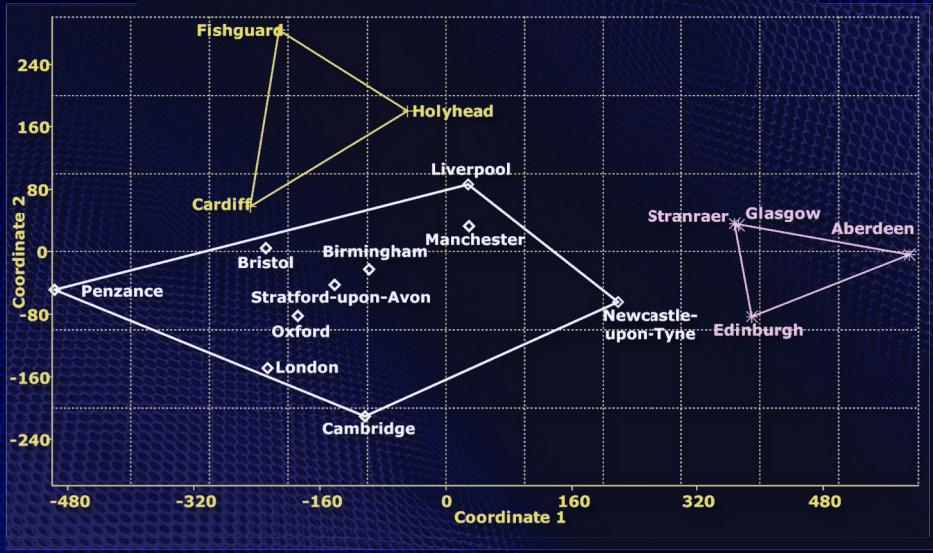


Protein Space

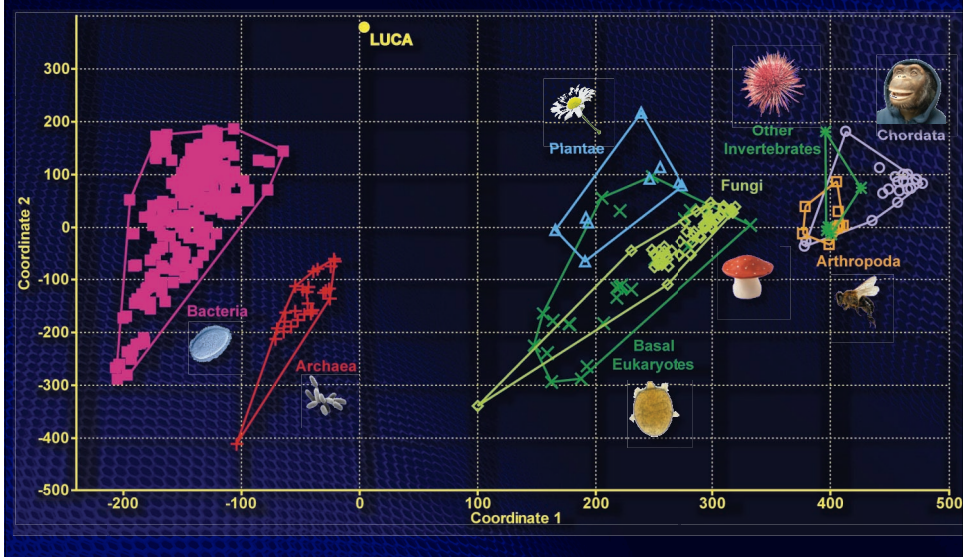


Protein Space

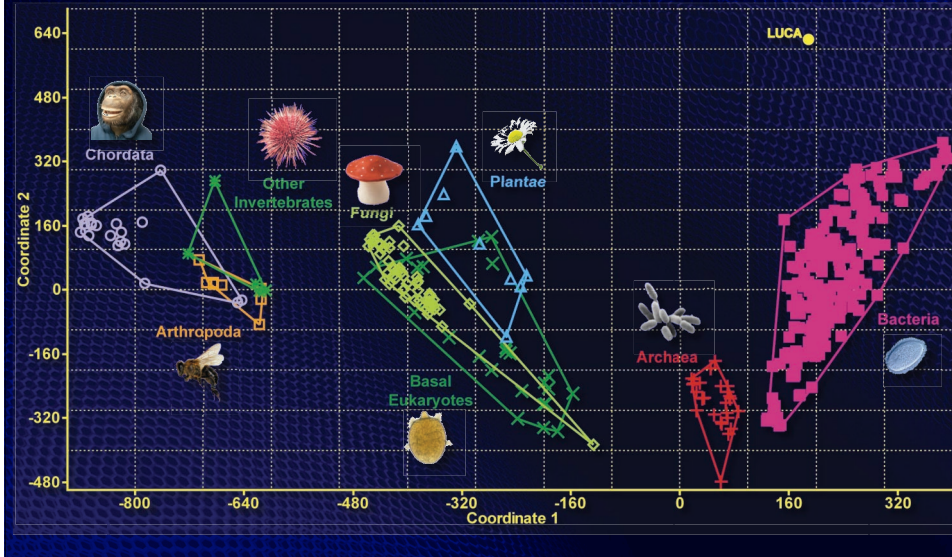
Principal Coordinate Analysis – A Very Brief Introduction



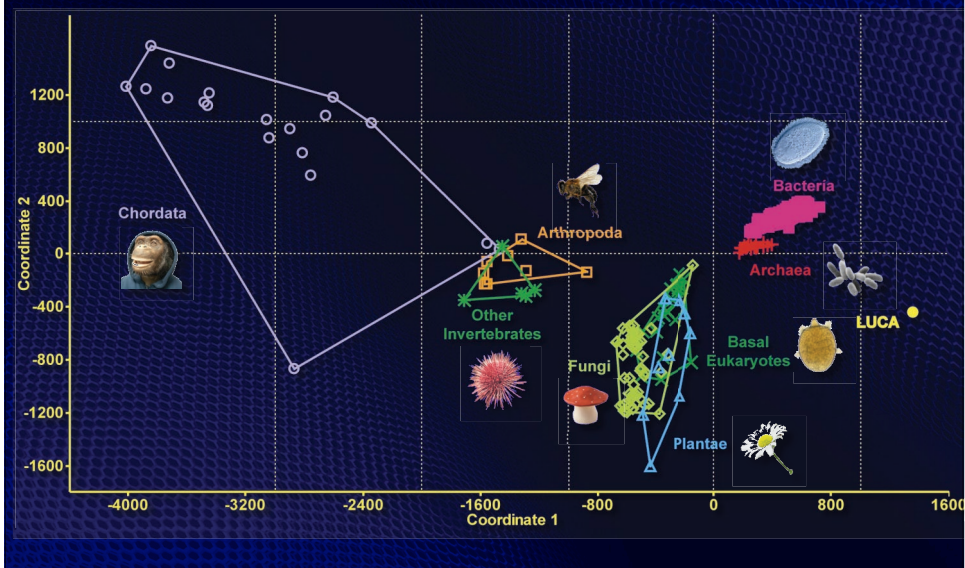
Protein Space - Superfamilies



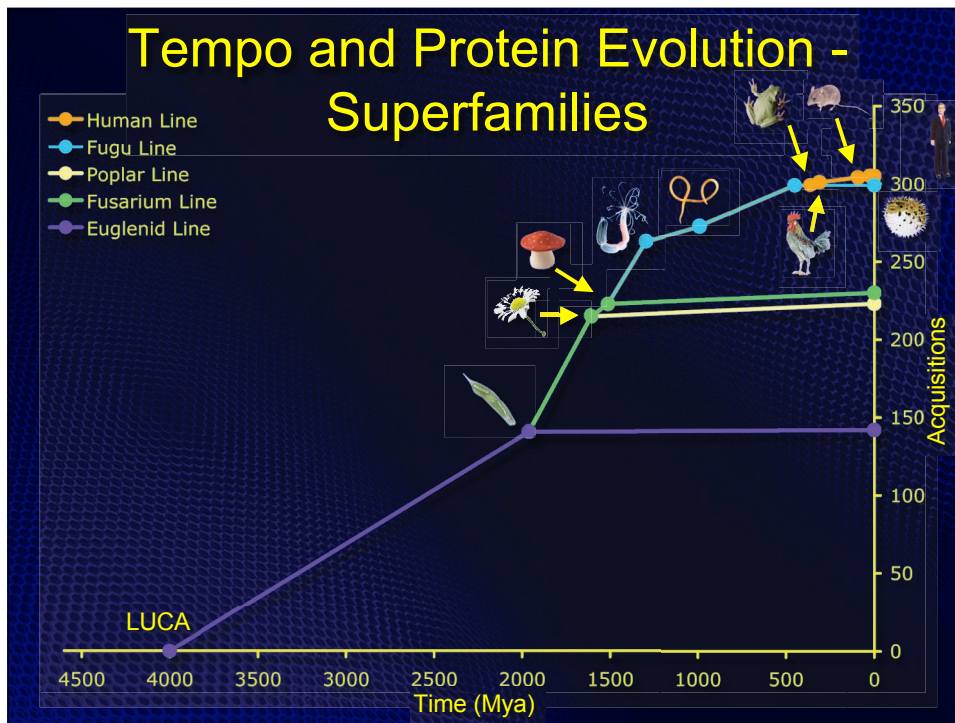
Protein Space - Families



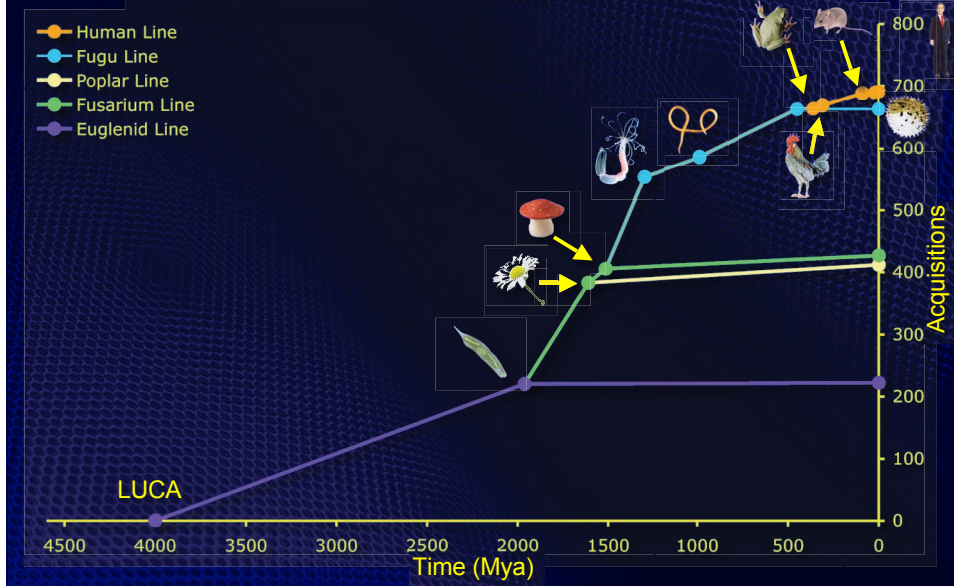
Protein Space - Architectures



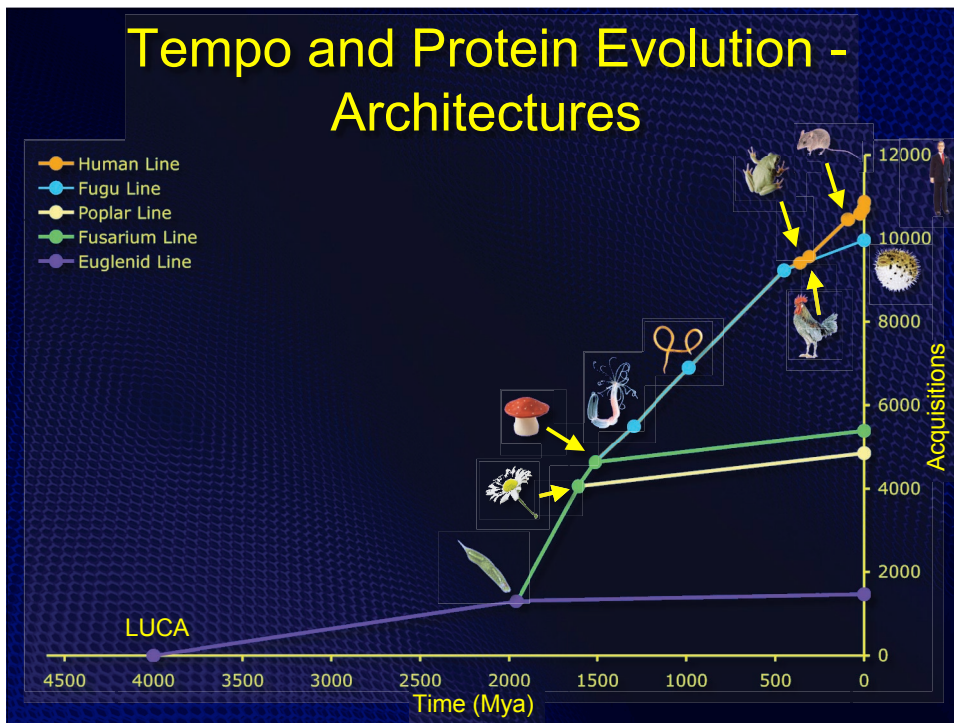
Tempo and Protein Evolution



Tempo and Protein Evolution - Families



Tempo and Protein Evolution - 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