A ‘double tale’ of accretion in the structure of biological networks

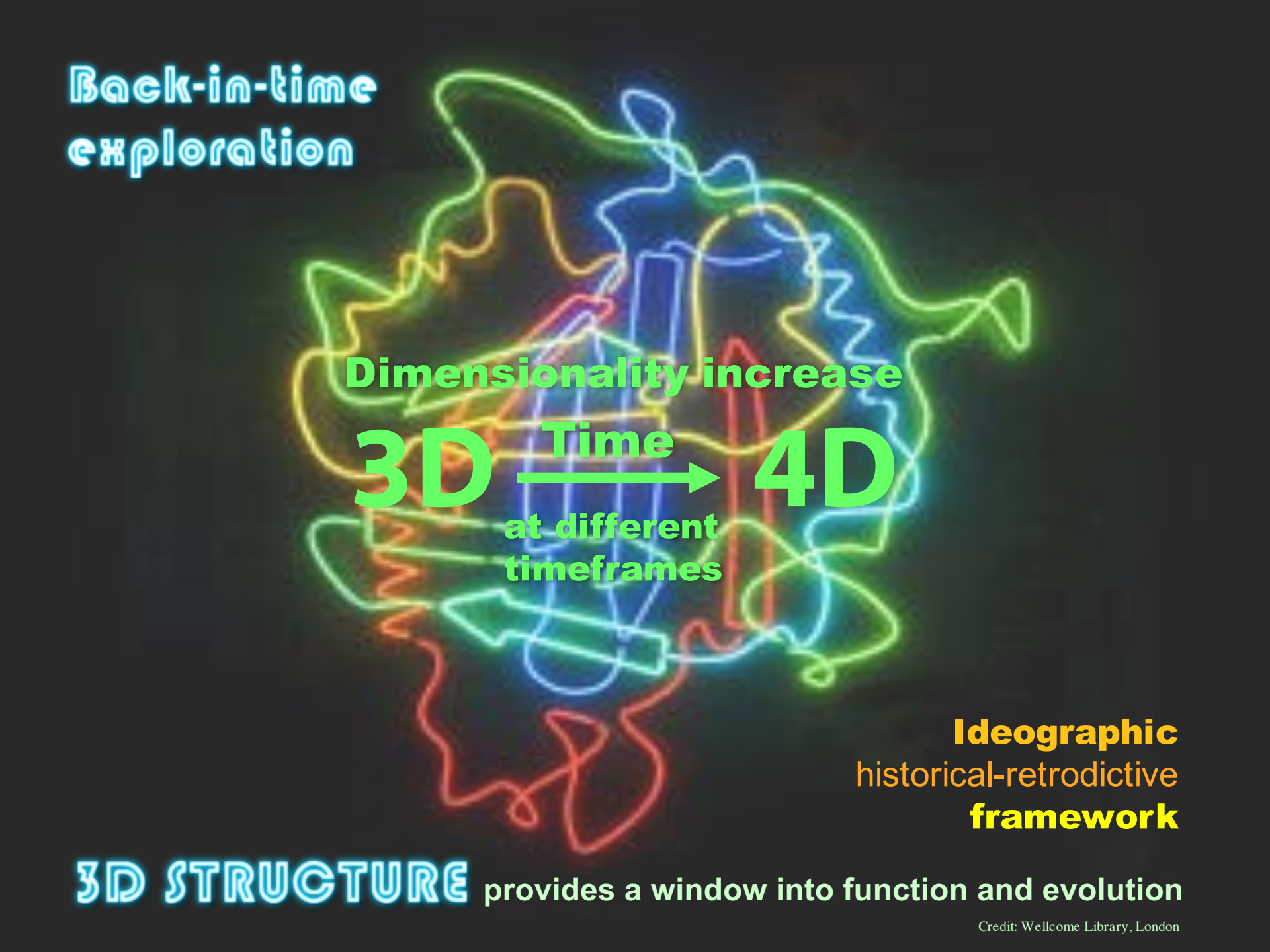
Gustavo Caetano-Anollés

University of Illinois, Urbana-Champaign, Illinois, USA

June 4-8, 2018 | Evolution – Genetic Novelty/Genomic Variations by RNA networks and Viruses, Salzburg

PUNCHLINE: A phylogenomic-based theory for the generation of structural modules explains network evolution and is congruent with Empedocles, *On Nature, P. Strasb.* Gr. Inv. 1665-66

Photo credit:
Biomedical Computation REVIEW


Back-in-time exploration
Dimensionality increase
3D
Time
4D
at different timeframes
Ideographic
historical-retrodictive
framework
3D STRUCTURE
provides a window into function and evolution
Credit: Wellcome Library, London

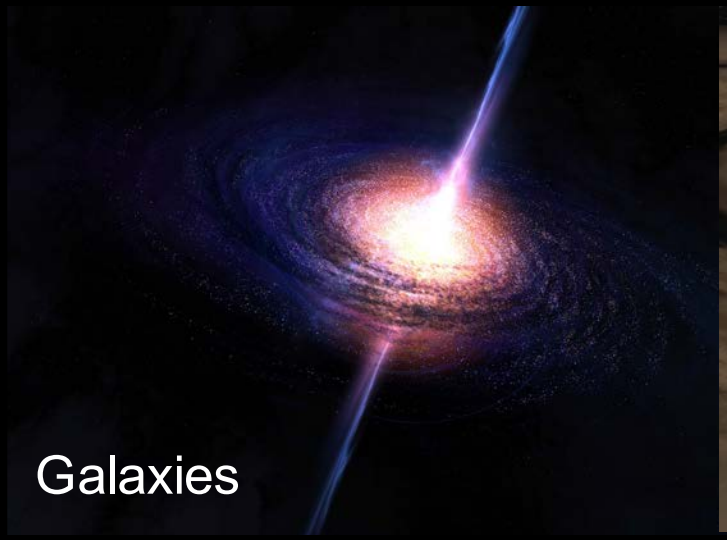
Accretion

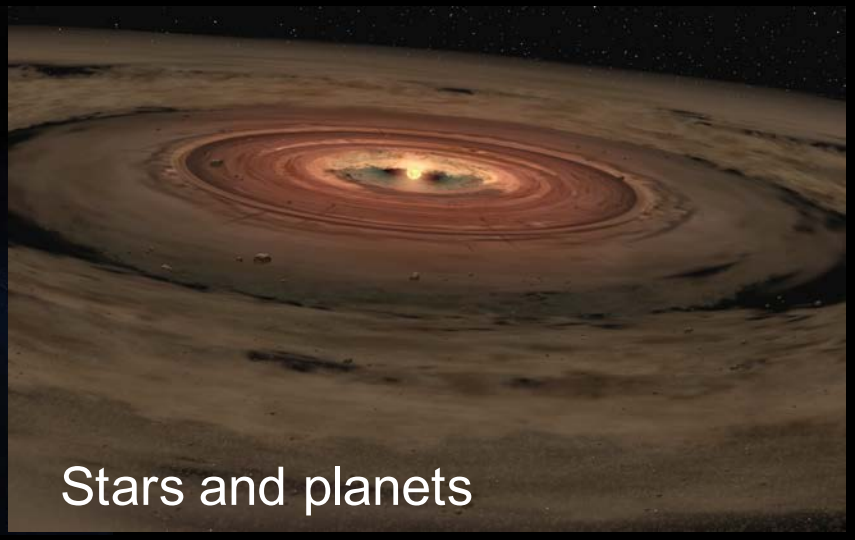
Accretion: A process of growth and increase, typically by the accumulation of entities and their interconnections

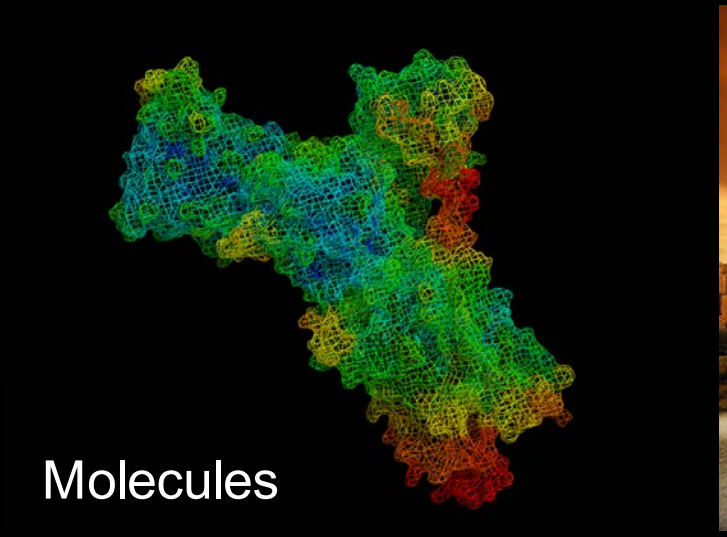
Accretion of networks: The process of growth and increase of nodes and links defining a network system, including their associated properties

accretion appears universal

EVOLUTION OF BIOLOGICAL NETWORKS







accretion appears universal

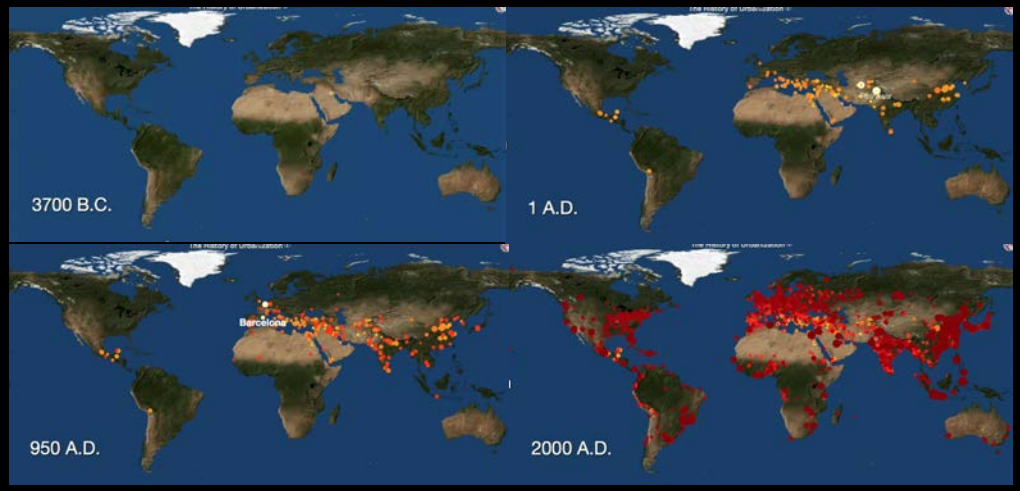
EVOLUTION OF BIOLOGICAL NETWORKS

Accretion of high-rise buildings in Lower Manhattan

CubeLease & NY City records

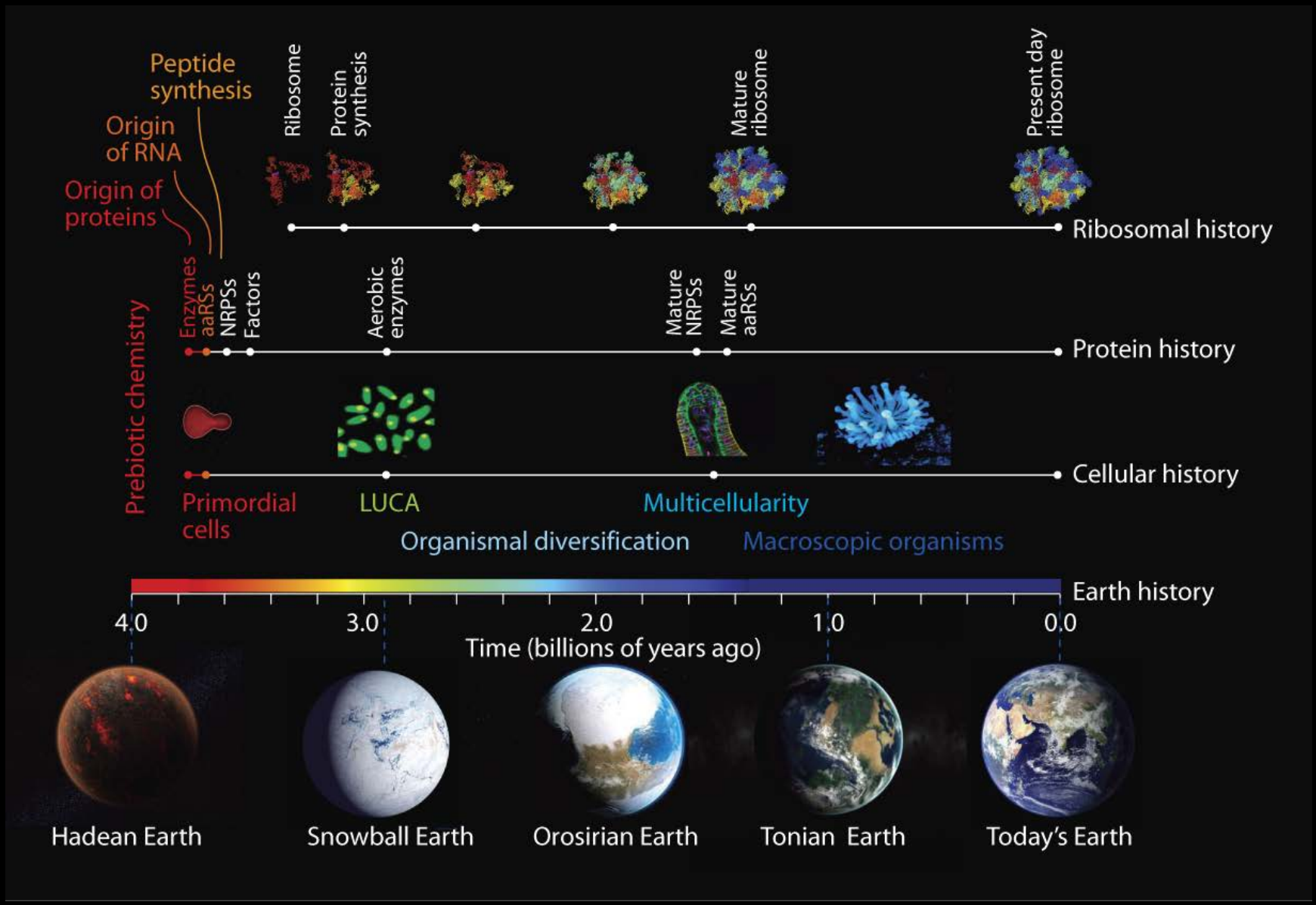
Planetary accretion of cities

Reba et al. 2016 *Sci Data* 3:160034



accretion at planetary scale

EVOLUTION OF BIOLOGICAL NETWORKS



accretion : theory

Biphasic model

Mittenthal *et al.* (2012) *Front Genet* 3: 147

The evolution of biological structure is driven by a 'double tale' of accretion and change

Accretion brings together disparate parts to form bigger wholes
Change provides opportunities for growth and innovation

In a *first phase*, parts are at first weakly linked and associate variously. As parts diversify they compete with each other and are selected for performance. The emerging interactions constrain their structure and organization. This causes parts to self-organize into modules with tight linkage.

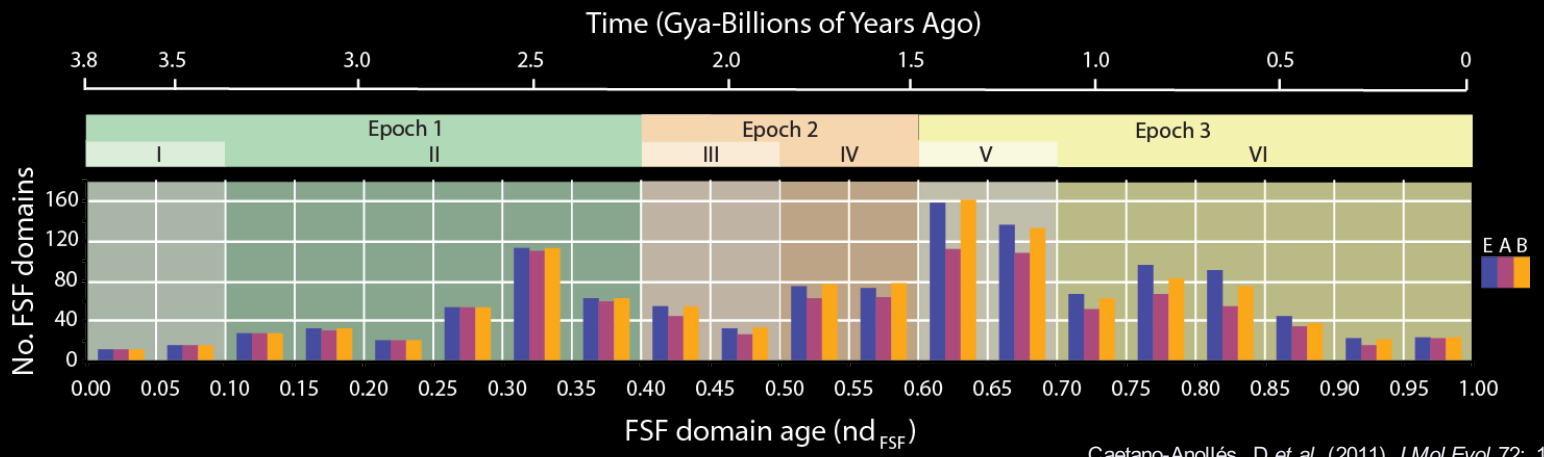
In a *second phase*, variants of the modules evolve and become new parts for a new generative cycle of higher-level organization.

accretion : theory

EVOLUTION OF BIOLOGICAL NETWORKS

Initial support for the biphasic model

Mittenthal *et al.* (2012) *Front Genet* 3: 147



Caetano-Anollés D *et al.* (2011) *J Mol Evol* 72: 14-32

- Developmental hourglass
- Diversification of macromolecules:
  - protein structures
  - shapes of RNA molecules
- Diversification of networks
  - early metabolism
  - evolution of codes
  - evolution of development and epigenetics

accretion : philosophy

EVOLUTION OF BIOLOGICAL NETWORKS

The Strasbourg papyrus (Gr. Inv. 1665-6)

Janko (2012) *Zeitschrift für papyrologie und epigraphik Bd* 150: 1-26

The ~2,000-year old papyrus of the ancient city of Panopolis in Upper Egypt describes a puzzling process of evolutionary growth:

"A double tale I'll tell. At one time one thing grew to be just one from many, at another many grew from one to be apart. Double the birth of mortal things, and double their demise."

Lines 233-235

Alain Martin in 1992 used papyrological and doxographic evidence to support a scribal transmission of Empedocles, *On Nature* (ca. 450 BC) that matches quotations from Simplicius (ca. 400 AD) and describes a 'cosmic cycle'.

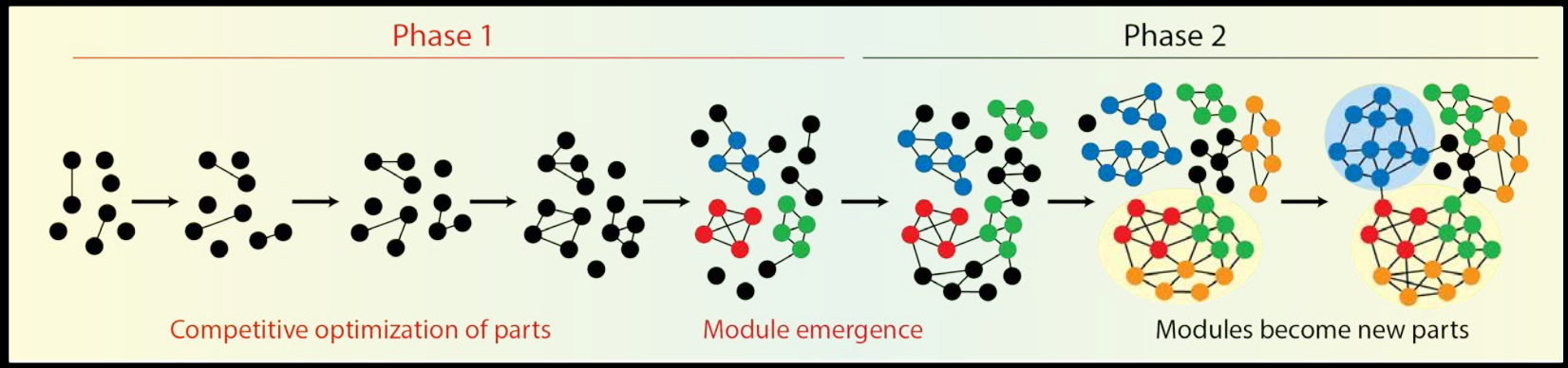
The hexameter text is a poem that embodies Darwinian and network thinking ~2,500 years ago and matches our theory (ms. in preparation)

Stichometric scribal sign Γ= stichos 300

accretion : network model

The ‘double tale’ is a biphasic model of module creation

At one time one thing grew to be just one from many… *at another many grew from one to be apart*



The ‘double tale’ explains the *emergence of network structure* in evolution

Nodes of the network are parts of a growing system
Connections represent links of interaction

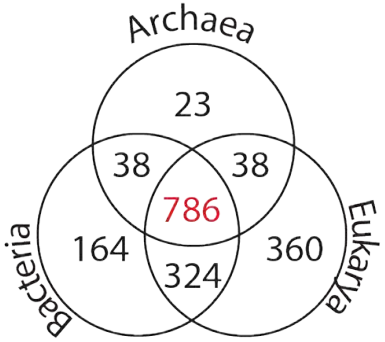
The larger the number of links the more cohesive and stable is the structure of the network

During Phase 1, the emergence of *hierarchical modularity* gives rise to small highly connected subnetworks, which can become modules.
During Phase 2, modules coalesce by combination into higher modules of network structure

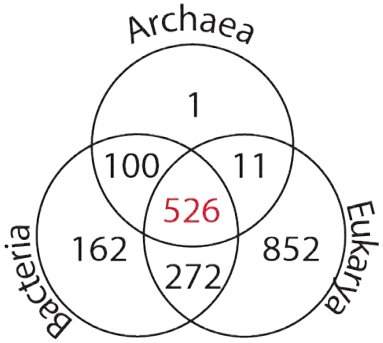
Caetano-Anollés et al. (2018) arXiv, http://arxiv.org/abs/1805.06487

accretion : molecular repertoires

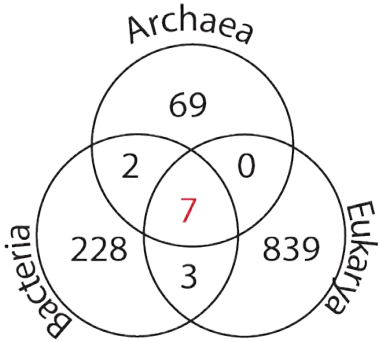
Ancient universal cores (in red) and derived peripheries support the 'double tale'



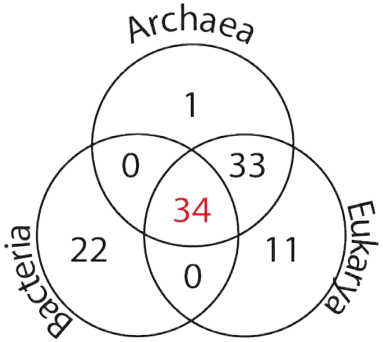
FSF domains



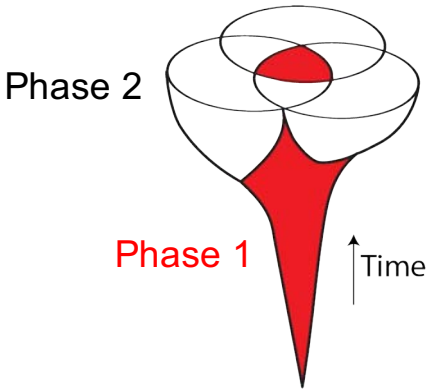
GO terms



RNA families



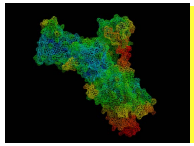
r-proteins



Venn diagrams of the distribution of fold superfamilies (FSF) of domains, Gene Ontology (GO) terms of molecular functions, RNA families, and ribosomal proteins (r-proteins) in genomes suggest *biphasic historical patterns*

accretion : reconstructing molecular history

EVOLUTION OF BIOLOGICAL NETWORKS



Mining molecular change in the 'parts' and 'wholes' of proteins and RNA

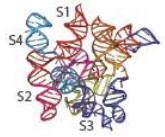
Genomes
↓ Structures, HMMs
Proteomes
↓ Structural census (~1 million domain structures), Functional annotations
Demography tables
↓ Data normalization, Character coding, Character argumentation
↓ Phylogenetic analysis

Trees of life (ToLs)
Tree of proteomes (ToP)
Parts: Protein domains
Data matrix
$g_{1,1}\ g_{1,2} \cdots g_{1,n}$
$g_{2,1}\ g_{2,2} \cdots g_{1,n}$
$g_{m,1}\ g_{m,2} \cdots g_{mn}$
Wholes: Proteomes
Characters: Protein domains in proteomes
Tree of domains (ToD)
↓ Age calculation (nd values)
Age of structures and functions
↓ Functional annotation, Structural annotation, Molecular clock analysis
Timeline of structural and functional discovery

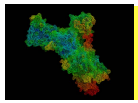
```
#NEXUS
begin data;
dimension ntax=185 nchar=776
format symbols="0~9 A~K"
transpose;
taxlabels
Homo_sapiens_22_34d
Pan_troglodytes_22_1
Mus_musculus_22_32b
Rattus_norvegicus_22_3b
Gallus_gallus_22_1
Xenopus_tropicalis_2_0
...;
matrix
a.1   EEDECDCG9BAFF57778666...
a.10  KKDK0000000DD0000000D...
a.100 JIHHHKIHFGIHFDGGGIFGH...
a.101 DAKF00000000700000000...
a.102 HHGGHHIJEGFECCGEFFBBB...
a.103 EGGD6DED9GDDD9DDBBBBB...
a.104 GGHHGGFGFFGFFAGCGG555...
a.108 AAD606606666A6A666666...
a.109 GGGGGKA00000000000000...
...;
endblock;
```

RNA structures
↓ Data normalization, Character coding, Character argumentation
↓ Phylogenetic analysis

Trees of life (ToLs)
Tree of molecules (ToM)
Parts: Molecular substructures
Data matrix
$s_{1,1}\ s_{1,2} \cdots s_{1,n}$
$s_{2,1}\ s_{2,2} \cdots s_{1,n}$
$s_{m,1}\ s_{m,2} \cdots s_{mn}$
Wholes: Molecules
Characters: Molecular substructures in molecules
Tree of molecular substructures (ToS)
↓ Age calculation (nd values), Structures, Structural annotations, Functional annotations
Evolutionary tracings and models

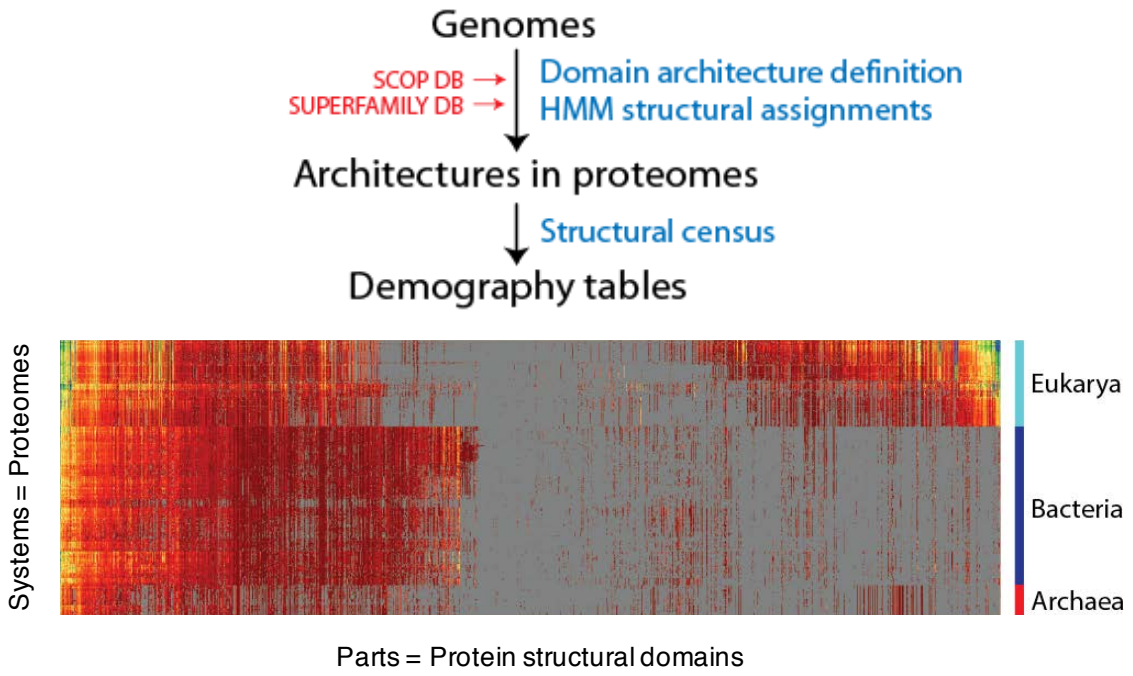


accretion : reconstructing molecular history



Computing the history of the protein world

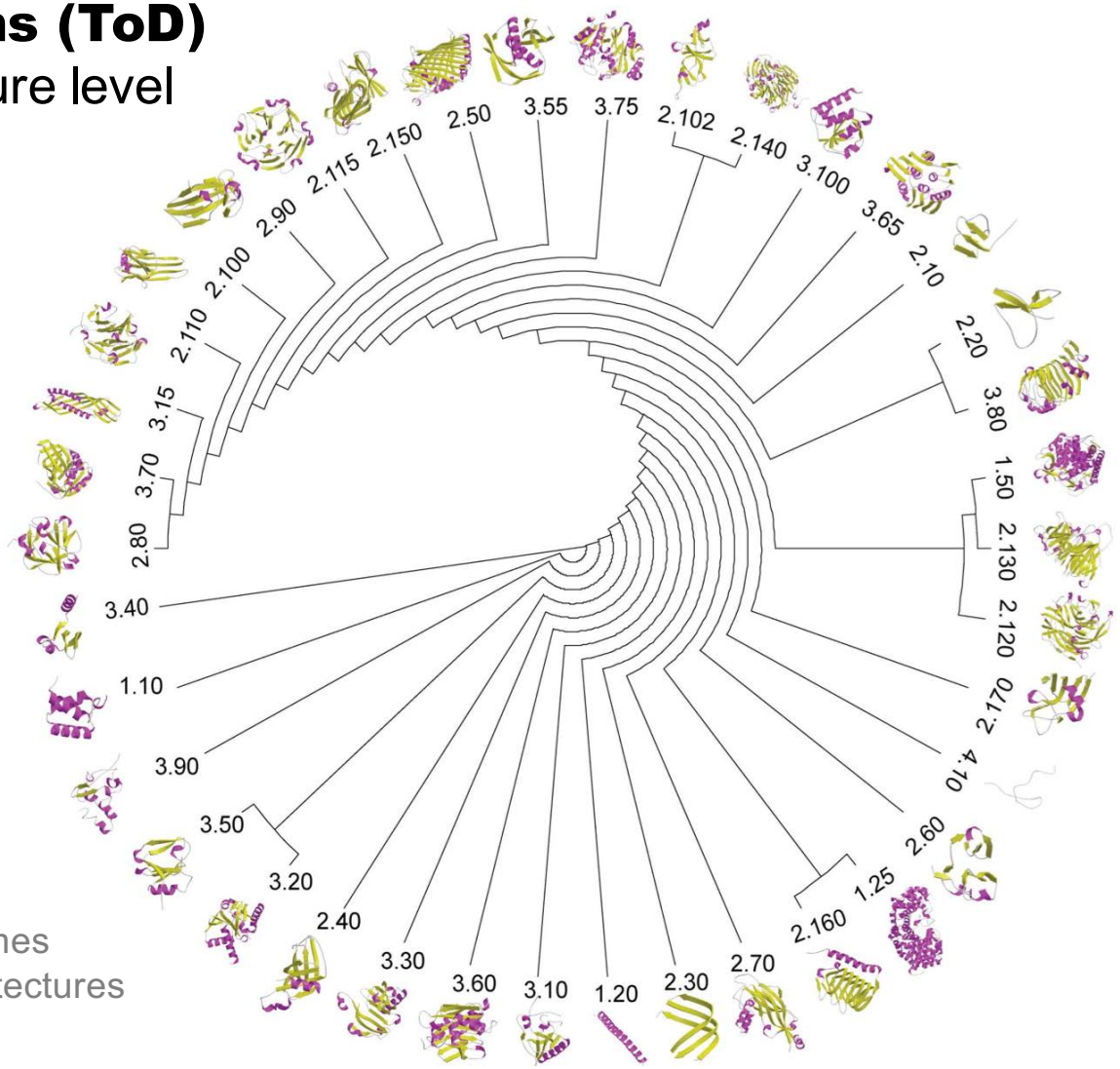
The collective of proteins (proteomes) and first decomposed into ***protein domains***, their structural, functional and evolutionary units



accretion … of fold structures

Tree of Domains (ToD)
at CATH architecture level

Bukhari and Caetano-Anollés (2013)
PLoS Comput Biol 9: e1003009

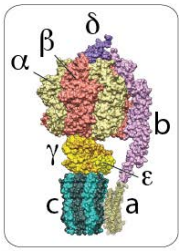


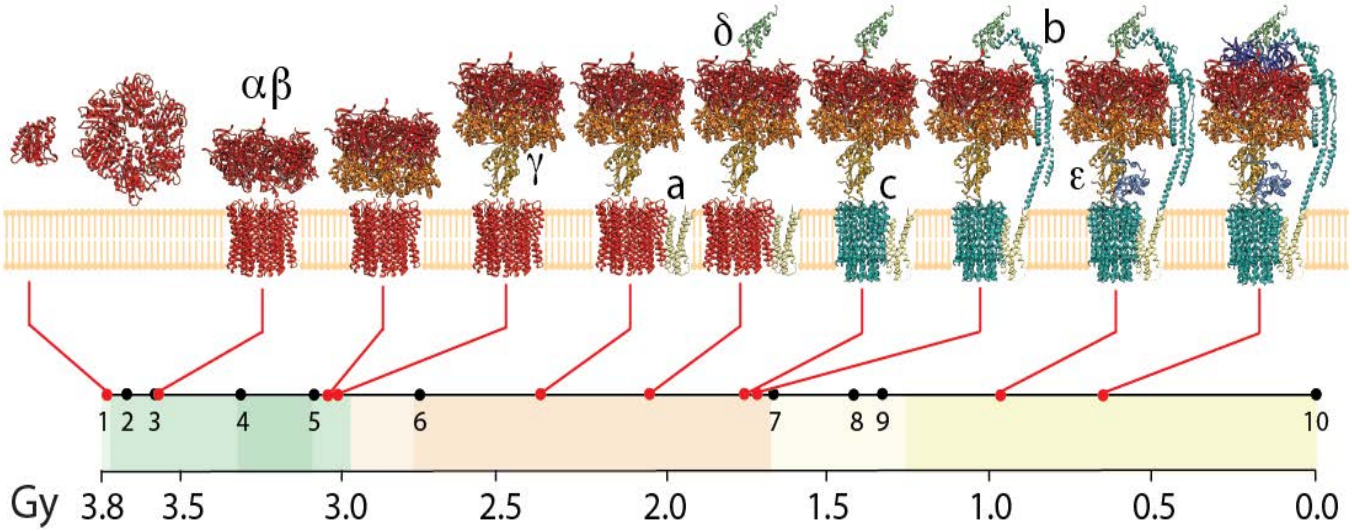
Systems: Proteomes
Parts: CATH architectures

accretion … in biological subsystems

Protracted evolution of ATP synthases: domain accretion took place during three quarters of the history of the protein world

Caetano-Anollés and Seufferheld (2013) *J Mol Microbiol Biotechnol* 23:152-177

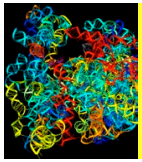




Molecular complexes are permanently optimized

EVOLUTION OF BIOLOGICAL NETWORKS

accretion … and ribosomal evolution



Phylogenomic 4D model of ribosomal evolution

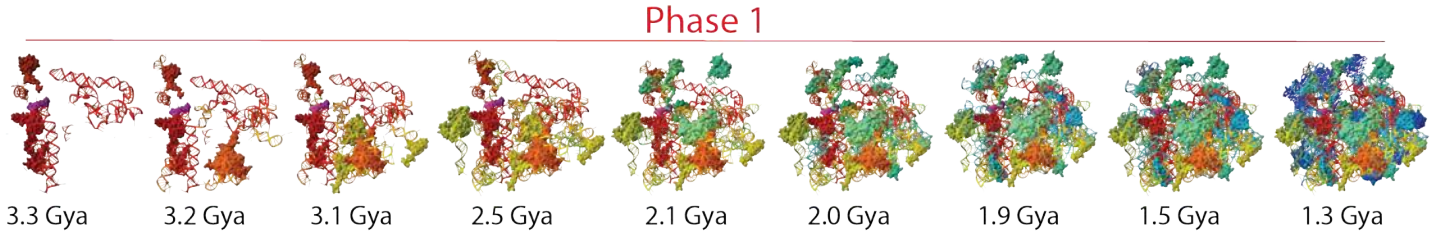
A biphasic 'double tale' is clearly evident

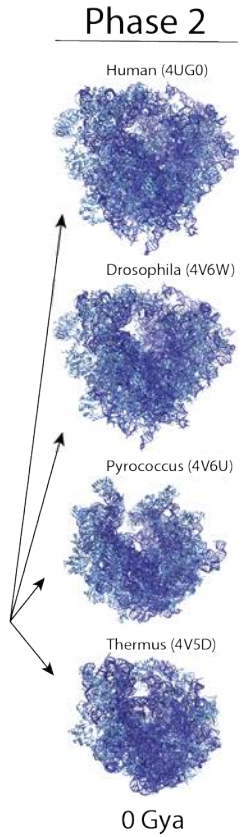
Harish and Caetano-Anollés (2012) *PLoS ONE* 7: e32776
Caetano-Anollés and Sun (2014) *Front Genet* 5: 127
Caetano-Anollés and Caetano-Anollés (2015) *CSBJ* 13: 427-447

An evolutionary timeline of ribosomal RNA and proteins inferred directly from phylogenomic data shows:

Phase 1: RNA helices interacted with proteins to form a processivity core 3.1 Gya, which served as center for coordinated ribonucleoprotein accretion.

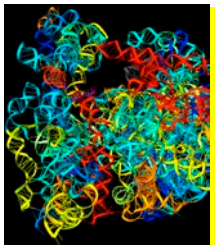
Phase 2: Core diversification started 1.3 Gya.





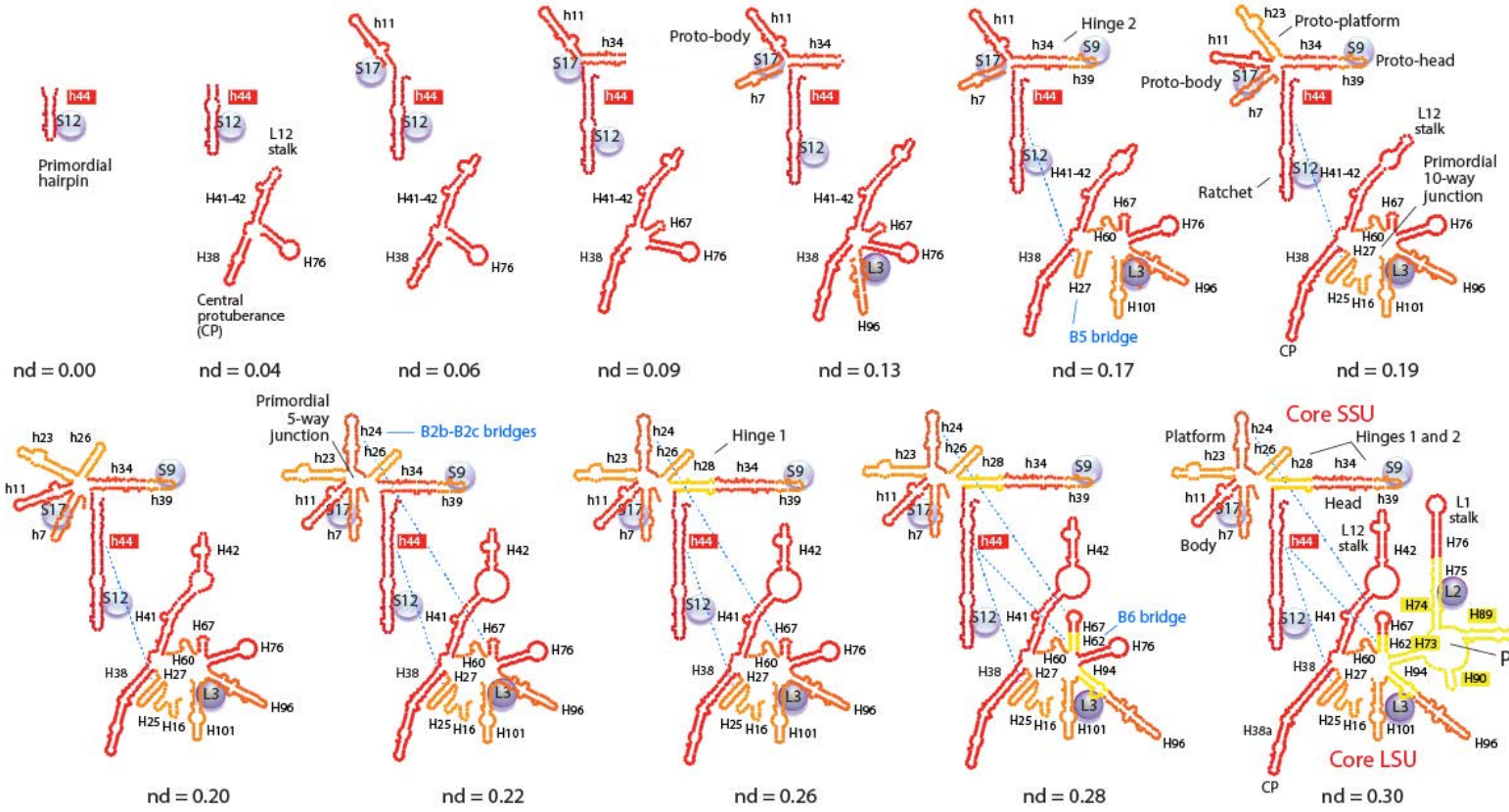
A biphasic 'double tale' pattern is clearly evident

accretion … the phylogenomic 4D model is powerful



Co-evolution of interactions

The 4D model predicts the history of important interactions (e.g. inter-subunit bridges, tRNA and ribosomal protein binding sites, hinges, A-minor components)



Caetano-Anollés D & Caetano-Anollés G (2015) *Comput Struct Biotech J* 13:427-447

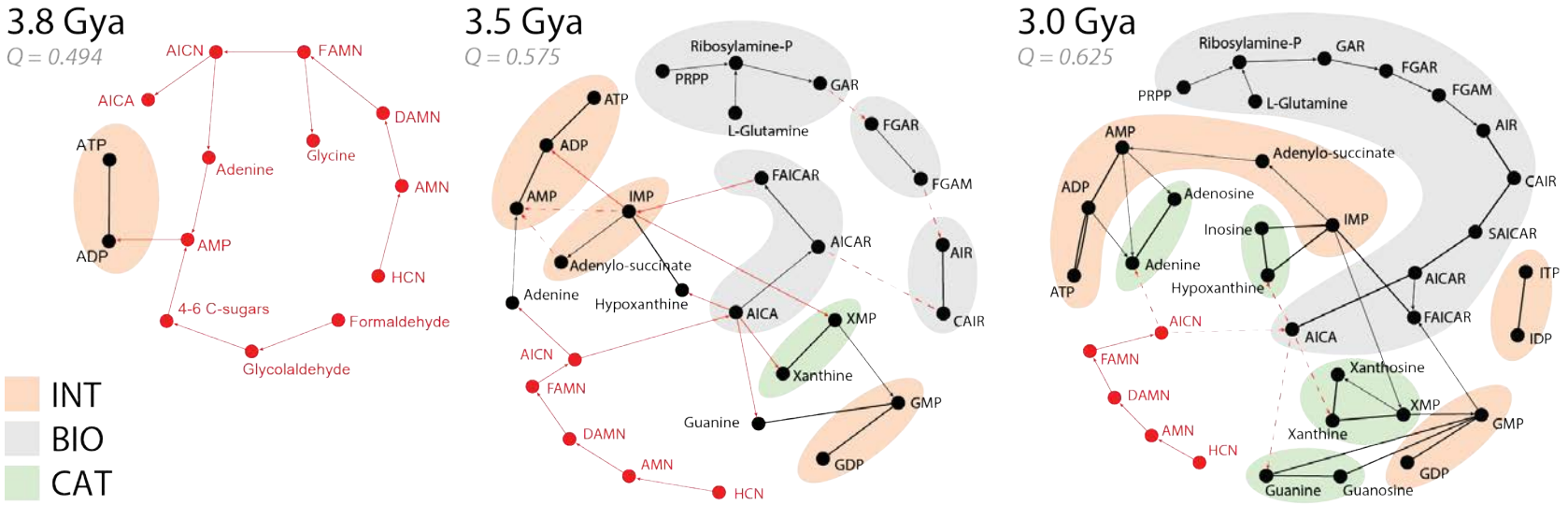
accretion … and metabolic evolution

EVOLUTION OF BIOLOGICAL NETWORKS

Phylogenomic 4D model of evolution of purine metabolism

Network modularity increases in evolution

Caetano-Anollés K & Caetano-Anollés G (2016) *PLoS One* 8:e59300



- Extant metabolites
- Plausible prebiotic metabolites
- -- Unknown reaction candidates/withering prebiotic pathways

Piecemeal recruitment of functional modules for the nucleotide interconversion (INT), catabolism and salvage (CAT) and biosynthetic (BIO) pathways.

Separate components unify into a cohesive network of modules. Modularity (Q), which measures the density of connections in node communities, increases in evolution.

accretion … and metabolic evolution

Phylogenomic 4D model of all metabolic networks

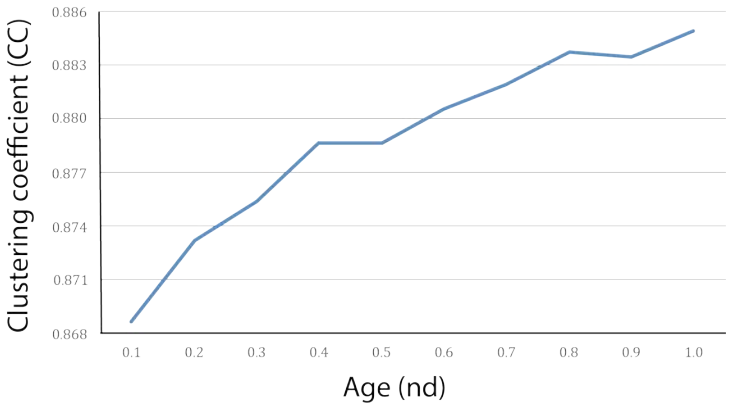
Pervasive trend of increase of network modularity in evolution

Mughal and Caetano-Anollés (2018), ms. submitted

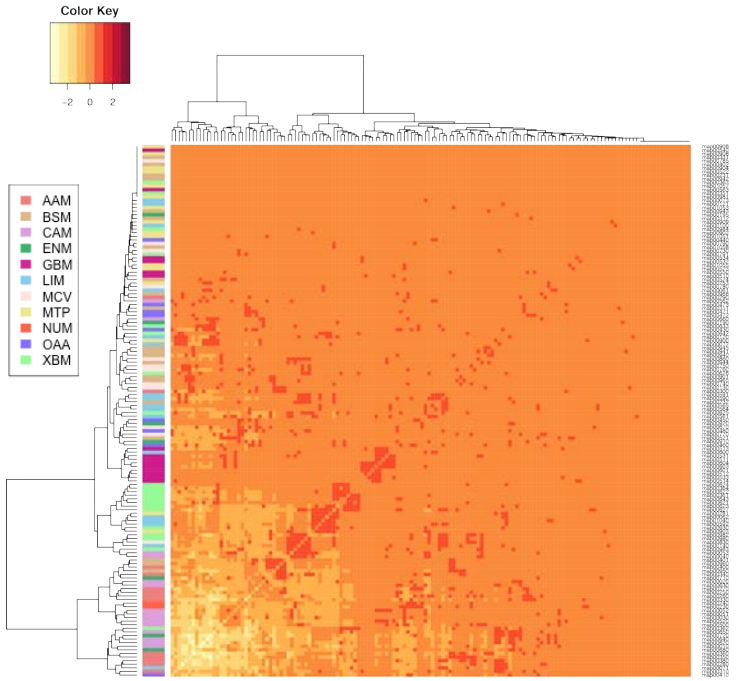
- The new MANET 3.0 database traces the ages of domains in metabolic enzymes at fold family level

Molecular ancestry networks
http://manet.illinois.edu

- The clustering coefficient and modularity metrics show an increasing trend of network modularity in evolution



- A tapestry of enzyme recruitment and modular structure is clearly visible in the adjacency matrix of metabolism



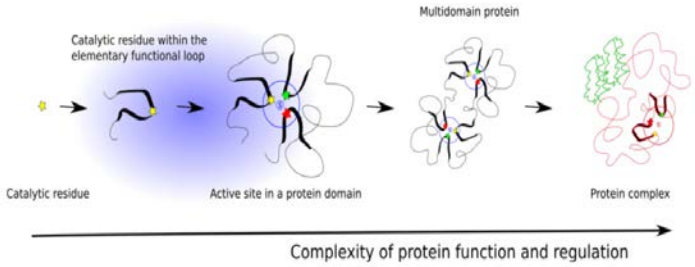
accretion … and the origin of protein domains

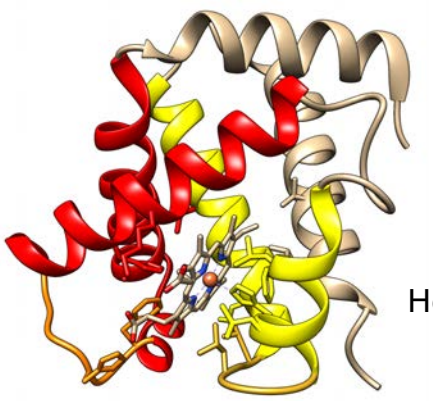
Phylogenomic 4D model of structural domain modularity

Domains emerge from the combination of 'elementary functional' loops

- Loop structures combine to make the structural domains of proteins

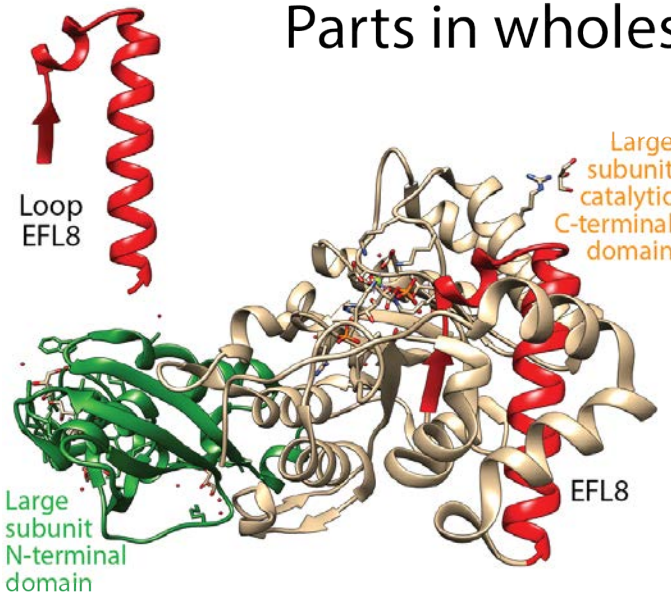
Goncearenco and Berezovsky (2015)
Physical Biol 12(4): 045002



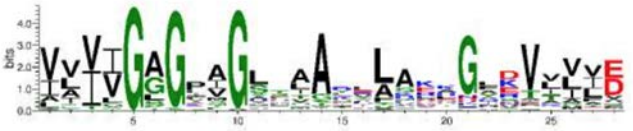


Hemoglobin

Parts in wholes



RubisCO



accretion … and the origin of protein domains

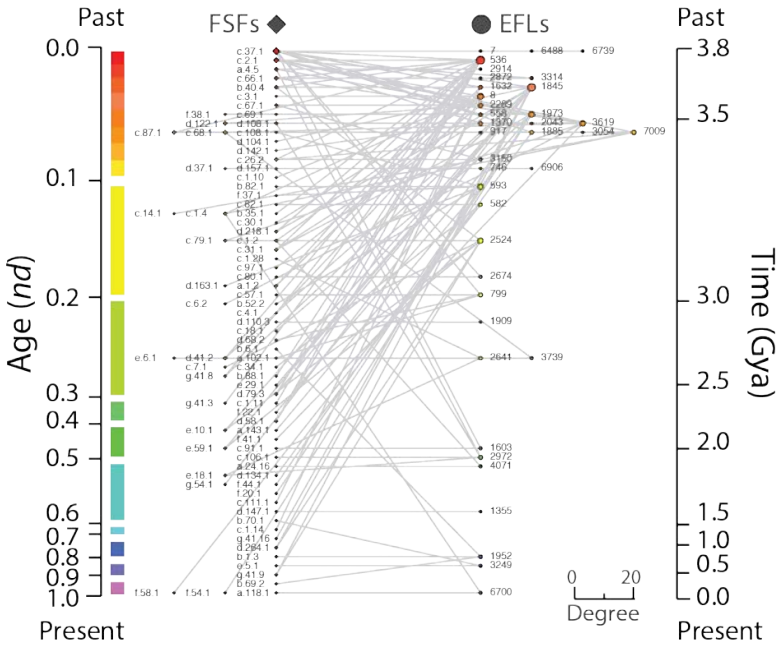
EVOLUTION OF BIOLOGICAL NETWORKS

Phylogenomic 4D model of protein domain modularity

Domains emerge from the combination of ‘elementary functional’ loops

Aziz et al. (2016) *Sci Rep* 6: 25058

- The dynamics of recruitment of loop motifs by domains structures is ongoing and highly modular
- The emergence of molecular functions show properties of hierarchical modularity and emergent power law behavior
- Modules in the bipartite network behave as integral communities of interacting structural parts defining classes of molecular functions
- The clustering coefficient shows an increasing trend of network modularity in evolution

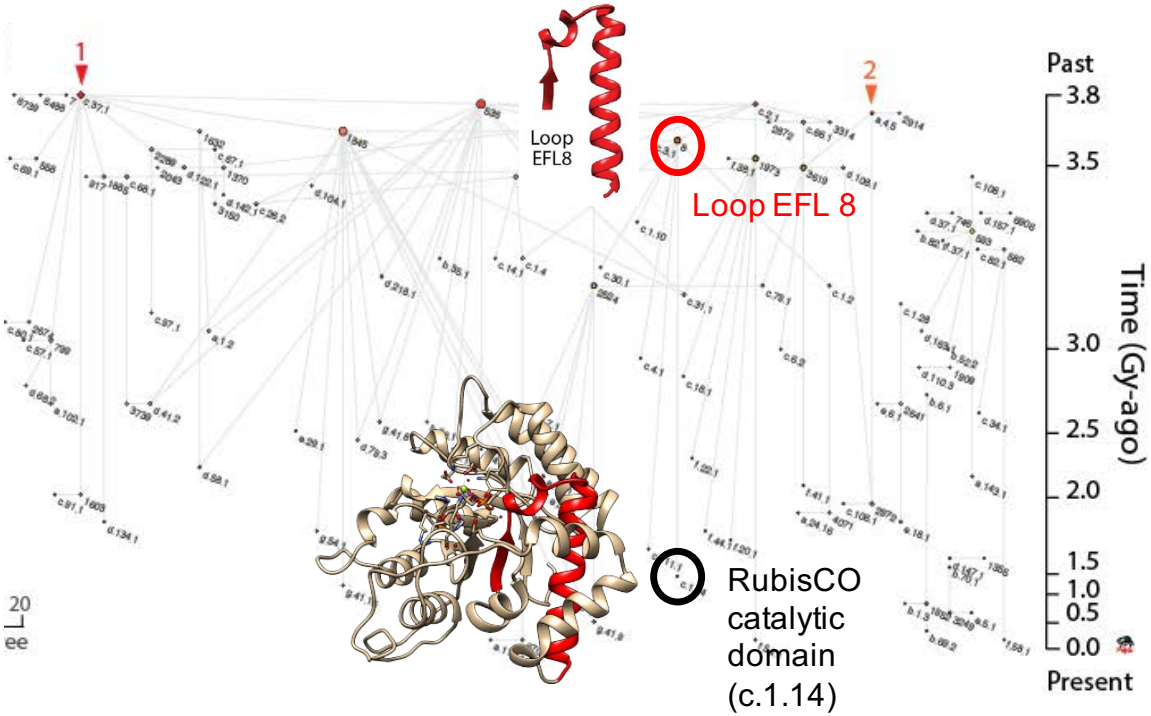


accretion … and the origin of protein domains

Phylogenomic 4D model of protein domain modularity

Domains emerge from the combination of 'elementary functional' loops

Aziz et al. (2016) *Sci Rep* 6: 25058

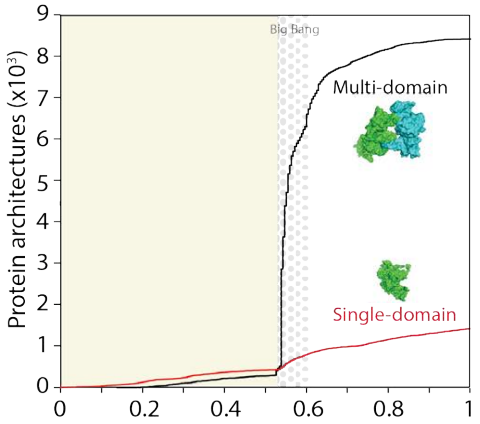


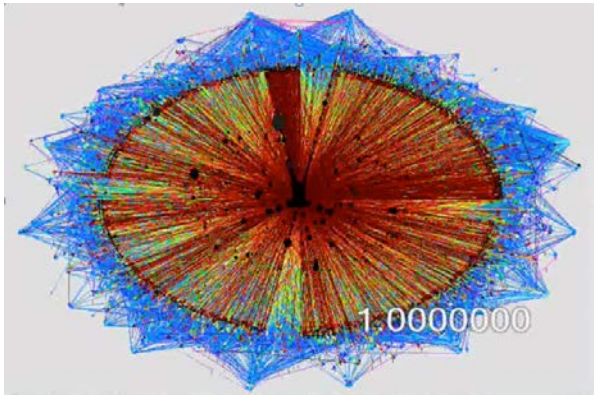
Two clear waves of functional innovation involved ancient loops and founder 'p-loop' and 'winged helix' domain structures

accretion … and the combination of domains

Domain rearrangements are pervasive in biology

A network of domains connected by their associations in proteins shows that domain combinations and rearrangements are abundant and widely distributed in the world of proteins



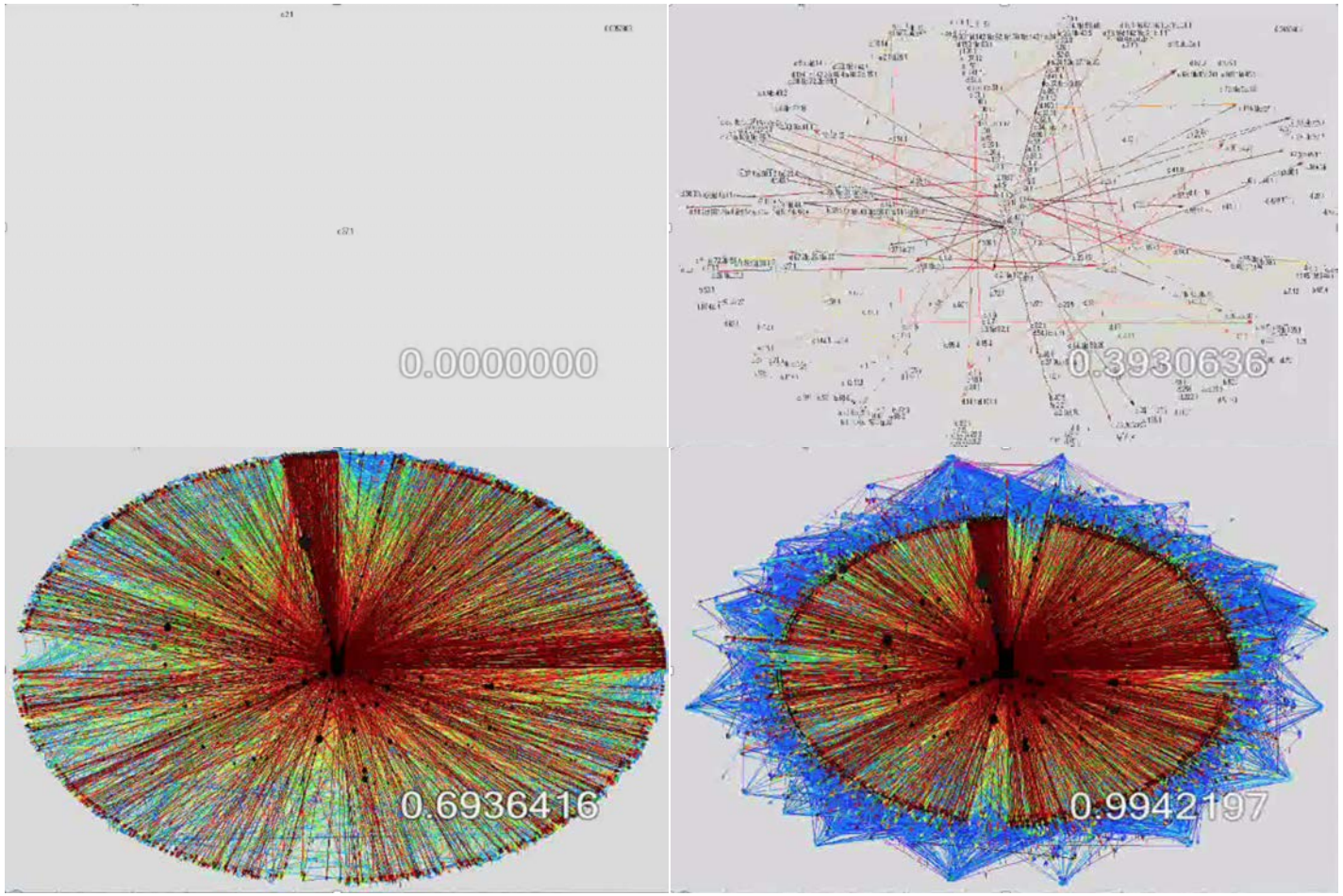


Wang & Caetano-Anollés (2009) *Structure* 17: 66-78

Aziz & Caetano-Anollés, *ms. in preparation*

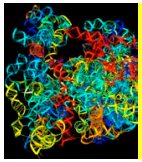
accretion … and the combination of domains

A historical and contextual network visualization approach



Aziz & Caetano-Anollés, *ms. in preparation*

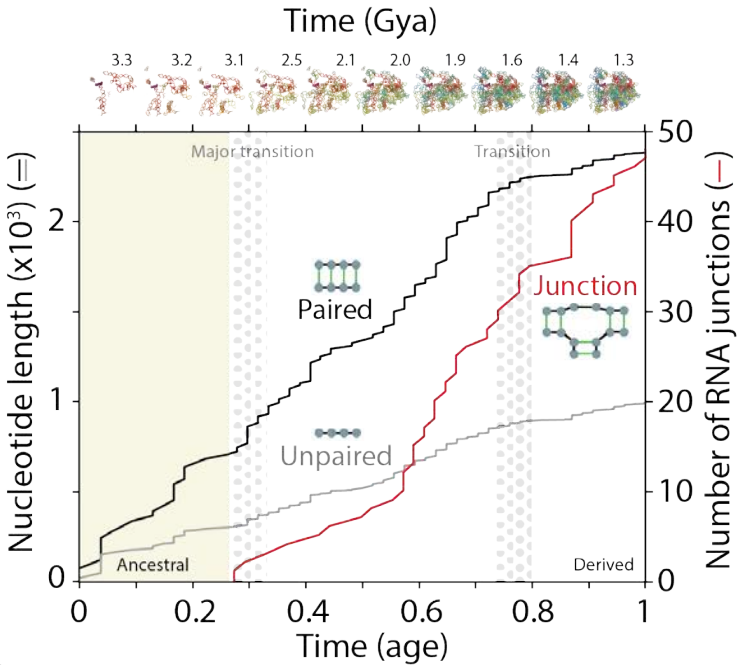
accretion … evolutionary growth of RNA

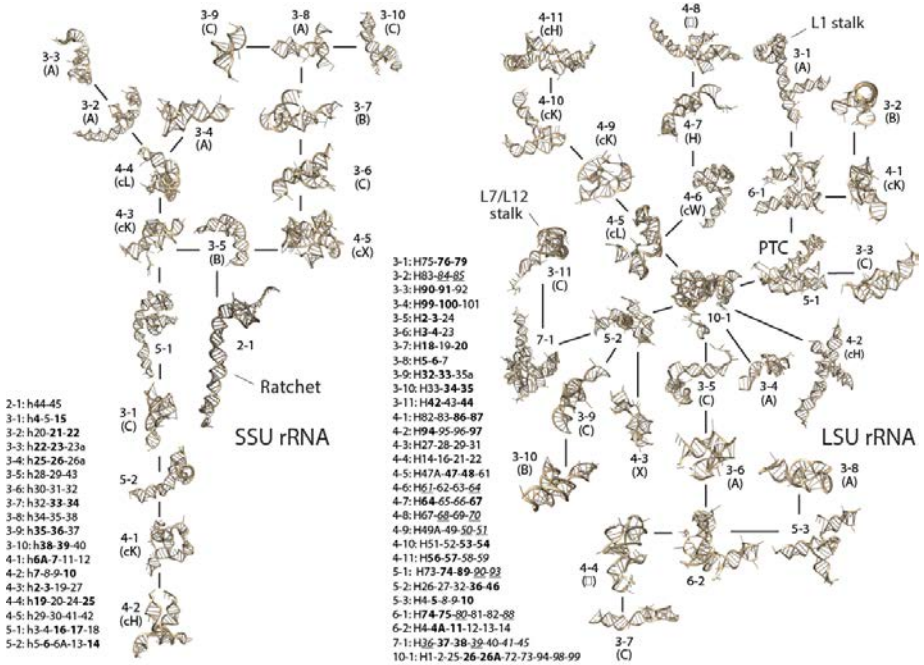


The emergence of a vocabulary of ribosomal junctions

Junctions are compact evolutionary modules of high-level hierarchy

Caetano-Anollés et al. (2018) arXiv, http://arxiv.org/abs/1805.06487





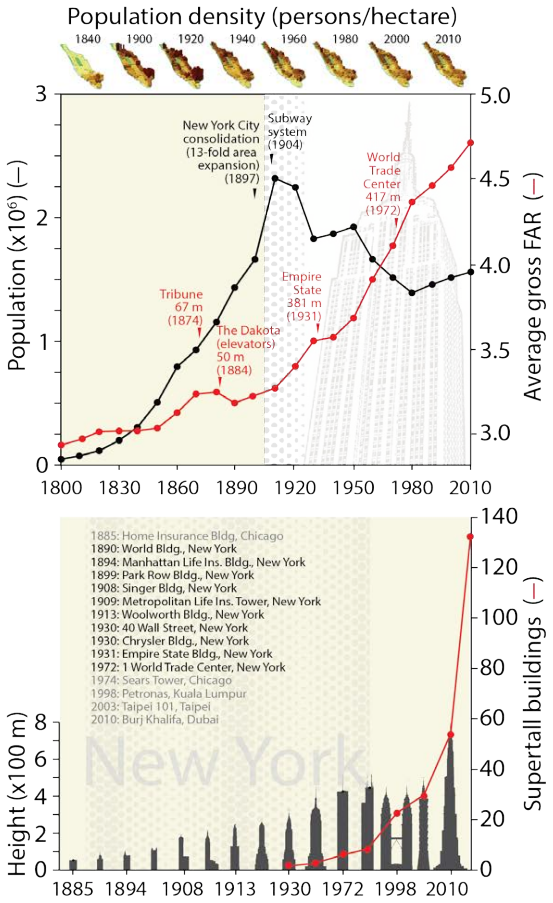
The biphasic 'double tale' pattern is again clearly evident

accretion … evolutionary growth of cities

The emergence of high-rise and supertall buildings

Buildings embed modules that accrete and diversify

Accretion



Caetano-Anollés et al. (2018) arXiv, http://arxiv.org/abs/1805.06487

Schlomo A, Lamson-Hall P (2014) Marron Inst Urban Mgmt, New York University, Working paper 14, pp. 1-48.

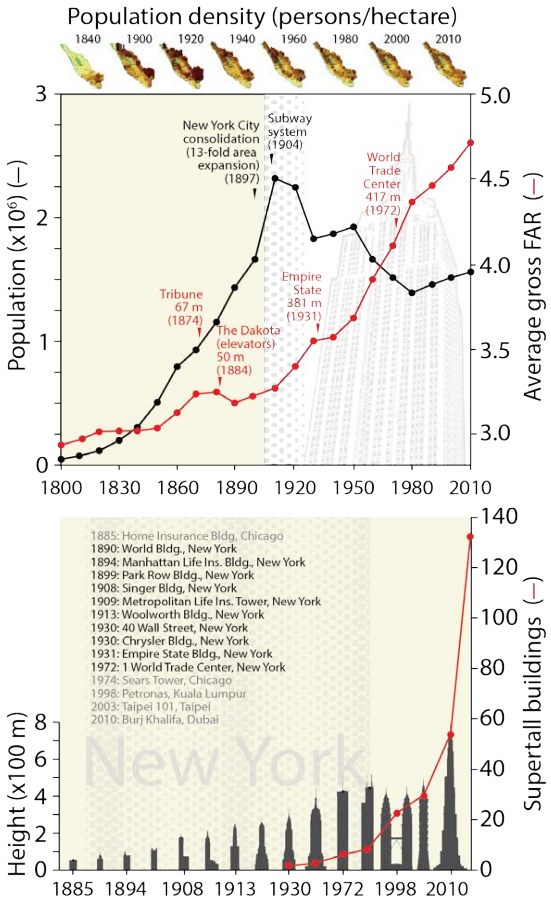
Pietrzak J (2015) *Challenges of Modern Technology* 6, 48-56 31-38.

accretion … evolutionary growth of cities

The emergence of high-rise and supertall buildings

Buildings embed modules that accrete and diversify

Accretion

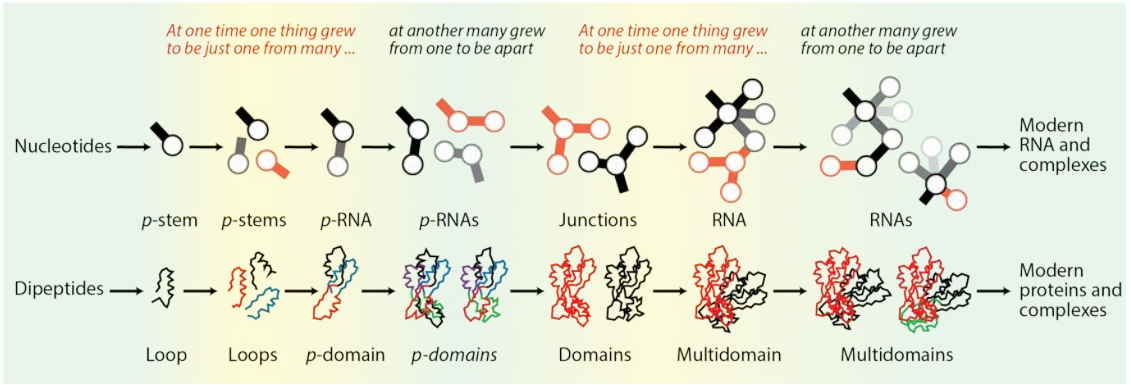


Diversification

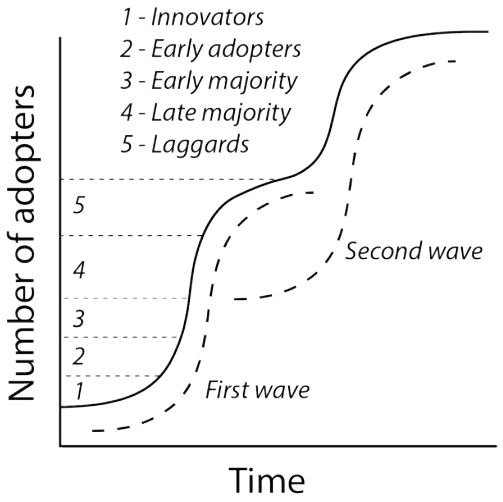
accretion … evolutionary growth cycles

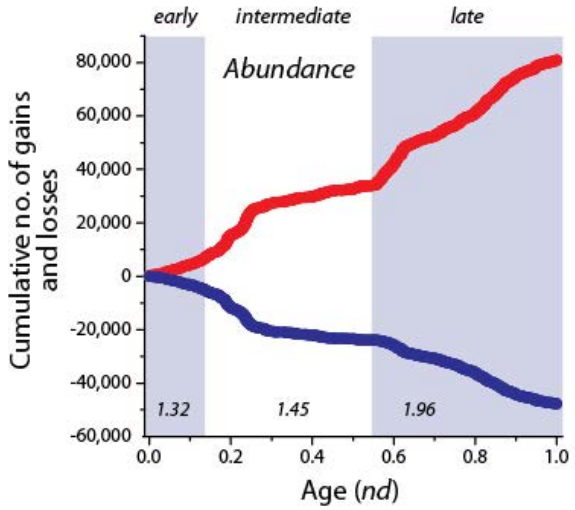
The 'double tale' explains the emergence of nucleic acids and proteins as successive cycles of module creation

Caetano-Anollés et al. (2015) arXiv, http://arxiv.org/abs/1805.06487



Logistic S-shaped wavelet analysis and the Bass equations can model accretion





Eq. 1: $\frac{d\rho(t)}{dt} = (p + q\rho(t))(1 - \rho(t))$

Eq. 2: $d\rho(t) = p + (q - p)\rho(t) - q[\rho(t)]^2$

Nasir *et al.* (2014) *PLoS Comp Biol* 10(1): e1003452

Accretion of biological networks

Jay Mittenthal
Derek Caetano-Anollés
Fayez Aziz
Kelsey Caetano-Anollés
Minglei Wang
Arshan Nasir
Abbas Bukhari
Feng Jie Sun
Fizza Mughal
Ibrahim Koç
Khuram Shahzad
Liudmila Yafremava
Bushra Minhas
Guy Tal
Heeshin Kim
Dongsheng Zhang
Ajith Harish
Manfredo Seufferheld

External collaborators:
Kyung Mo Kim (Korea)
Frauke Graeter (Germany)
Patrick Forterre (France)
Hong-Yu Zhang (China)
Chris Dupont (USA)
John Mitchell (UK)
Viviana Lepek (Argentina)
Hugo Naya (Uruguay)

accretion ‘big time’

“In the way that many arise as the one again dissolves,
in that respect they come to be an have no life eternal;
but in the way that never do they cease from change continual,
in this respect they live forever in never-ending cycle”

Emepedocles, *On Nature, P. Strasbourg*, Gr. Inv. 1665-6, lines 241-244

"I'm on my way, I'm making it, I've got to make it show…
So much larger than life, I'm going to watch it growing…

The place where I come from is a small town,
They think so small, they use small words
But not me, I'm smarter than that, I worked it out
I've been stretching my mouth, to let those big words come right out"

accretion 'big time' Peter Gabriel