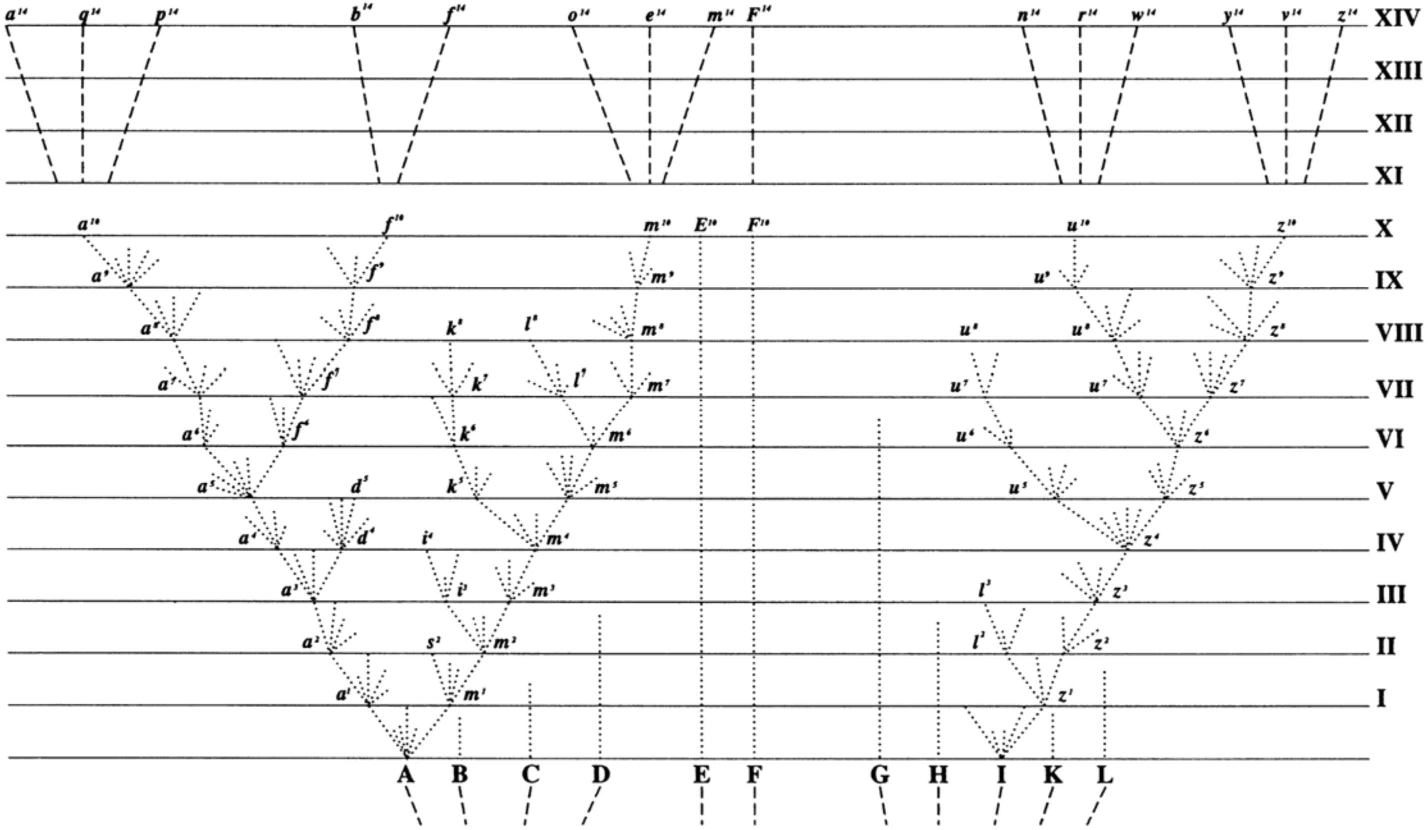
Net-Works 2008, Pamplona, June 2008

Universal branching in phylogenetic trees

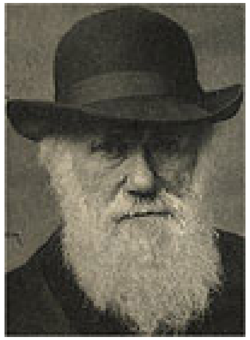
Emilio Hernández-García[1], E. Alejandro Herrada[1], Claudio J. Tessone[2], Konstantin Klemm[3], Víctor M. Eguíluz[1], Carlos M. Duarte[4]

1. **IFISC**, Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB),Palma de **Mallorca**, Spain
2. **ETH Zürich**,Systemgestaltung, Zürich, Switzerland
3. **Bioinformatics** Group, Computer Science Department, University of **Leipzig**, Germany
4. **IMEDEA**, Instituto Mediterráneo de Estudios Avanzados (CSIC-UIB), Esporles, **Mallorca**, Spain

Universal branching in phylogenetic trees

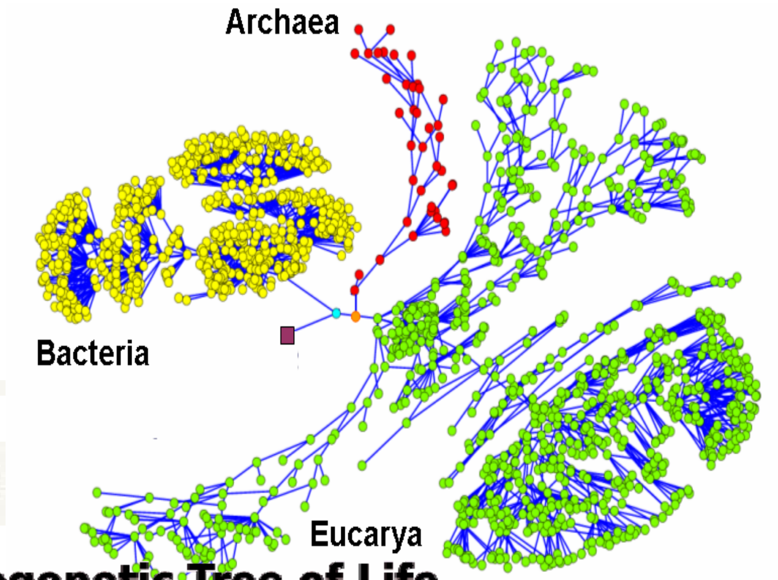


Charles Darwin in On the Origin of Species, 1859

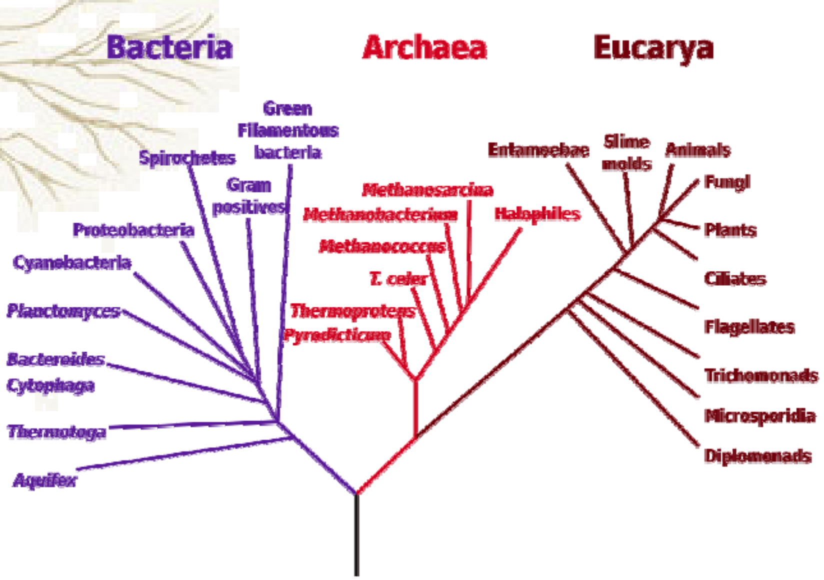


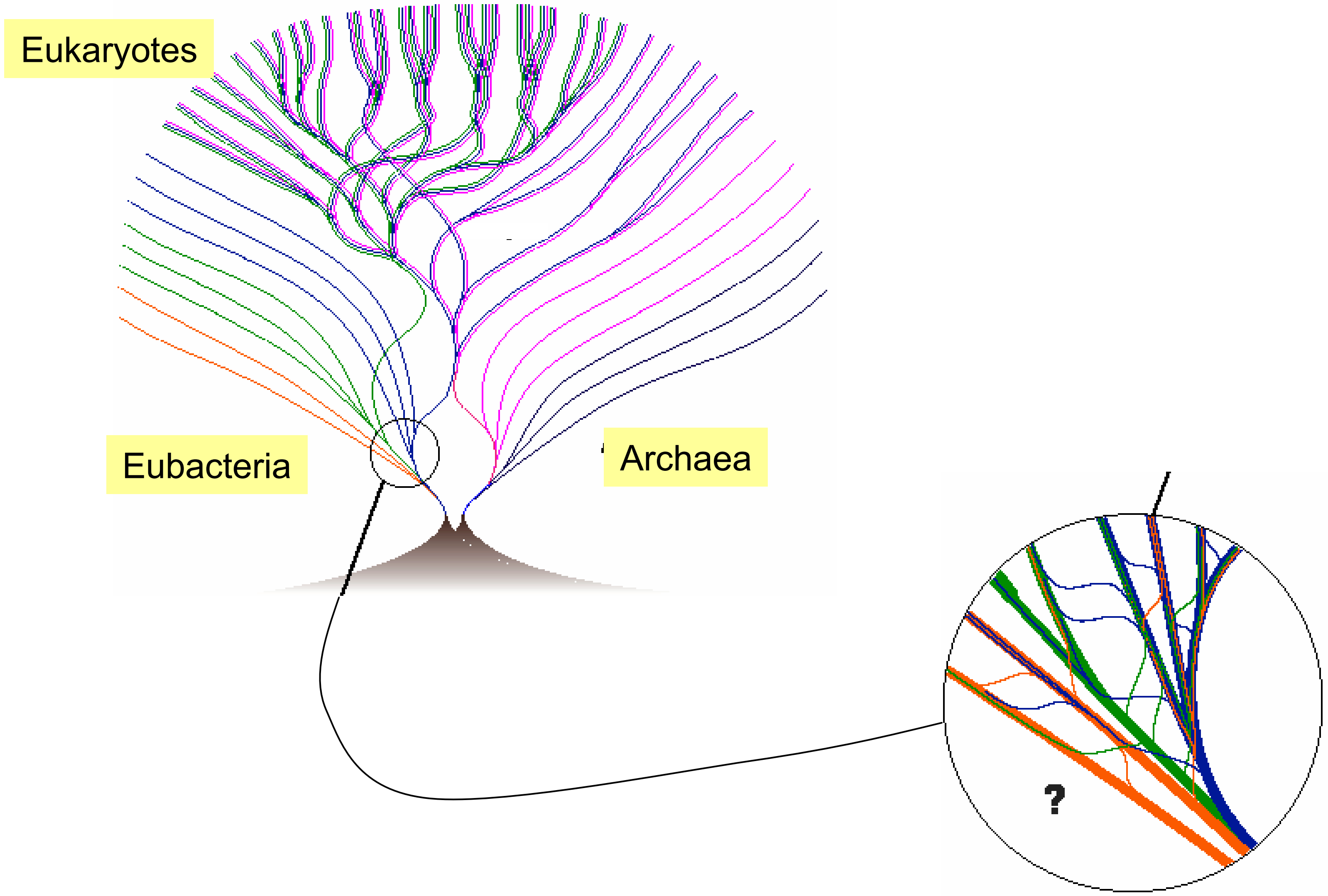
Universal branching in phylogenetic trees

Tree of Life from **tolweb.org**



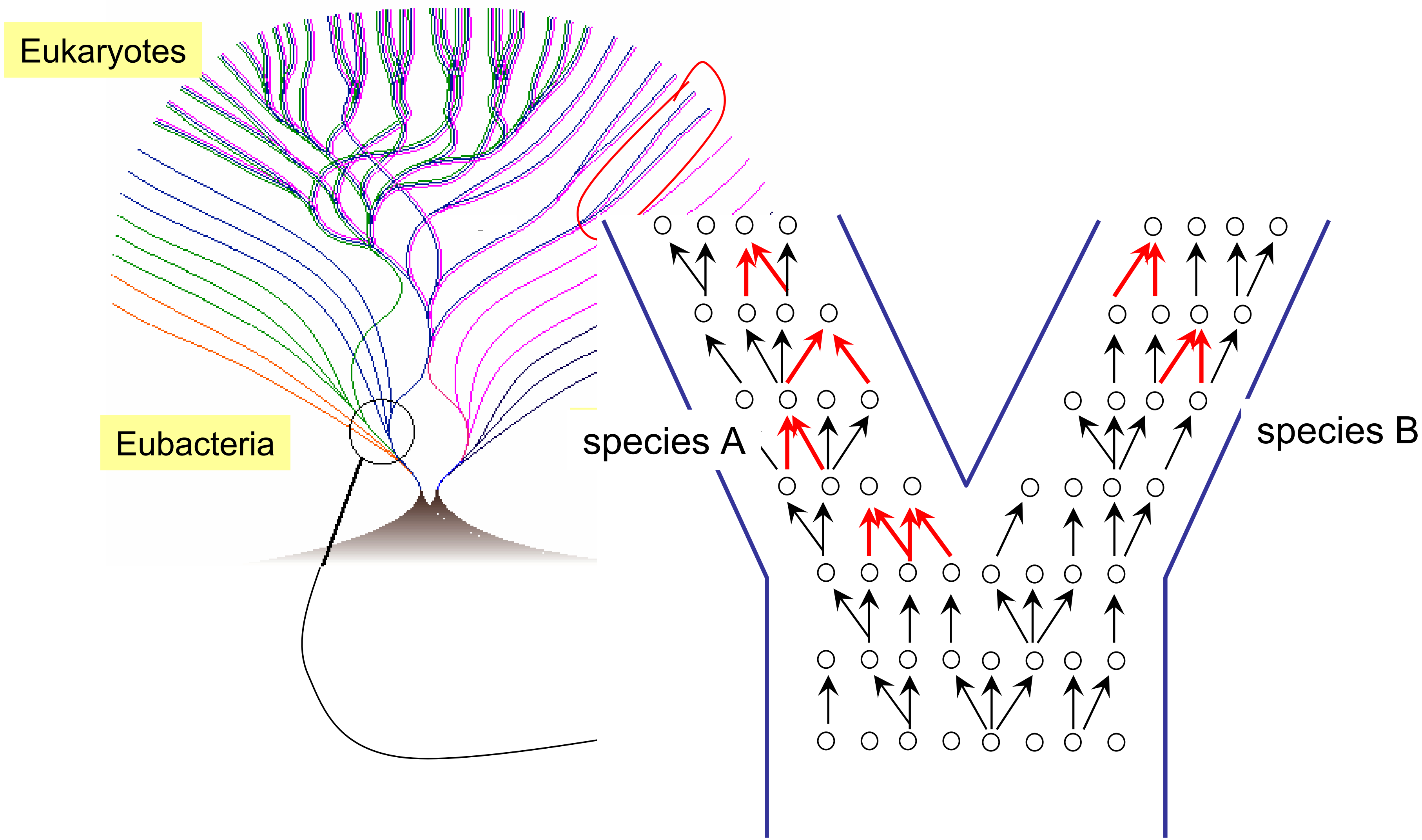
Phylogenetic Tree of Life



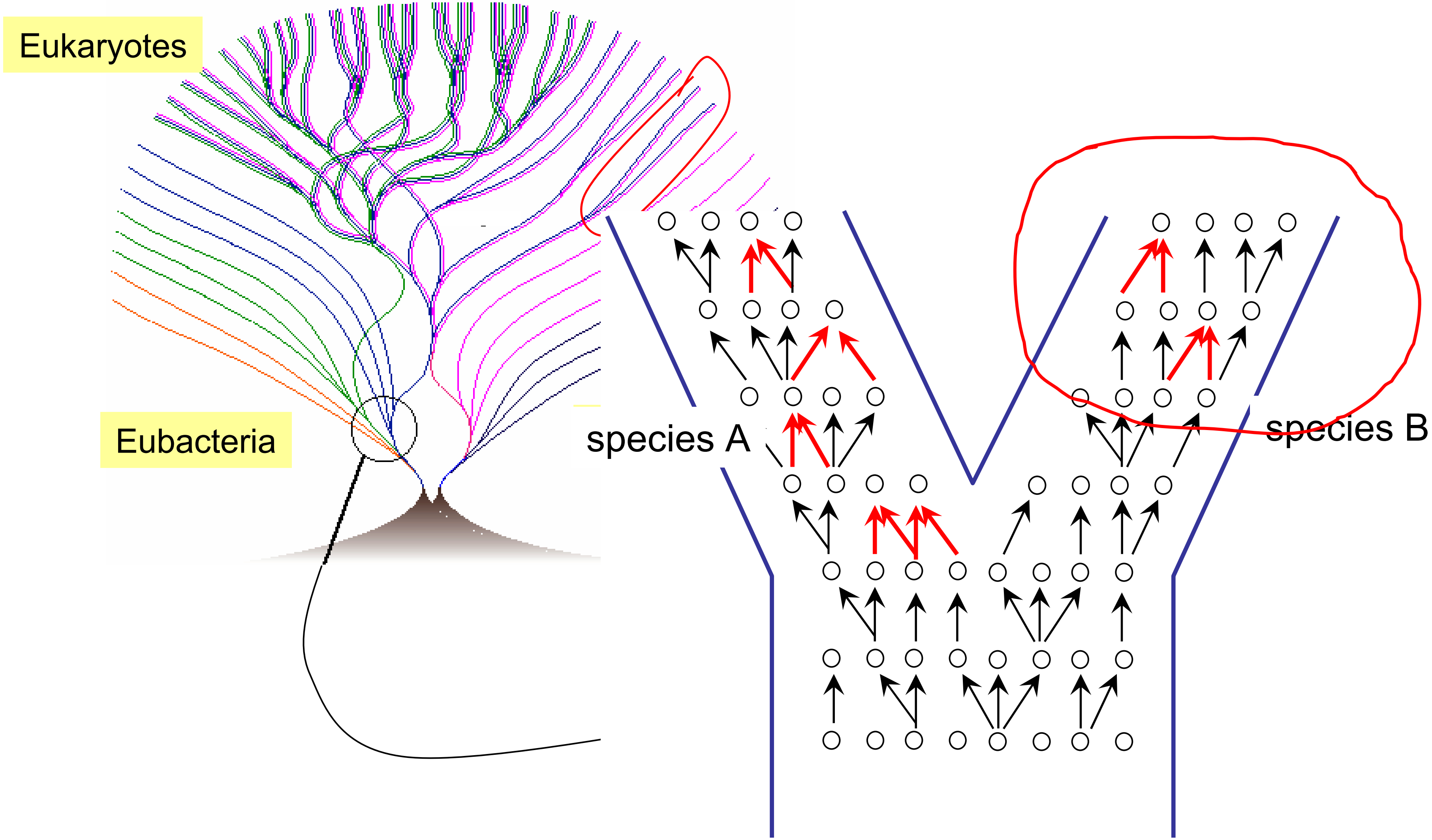


W. Martin, BioEssays 21.2, 1999

Universal branching in phylogenetic trees



Universal branching in phylogenetic trees



GENETIC SIMILARITY NETWORKS

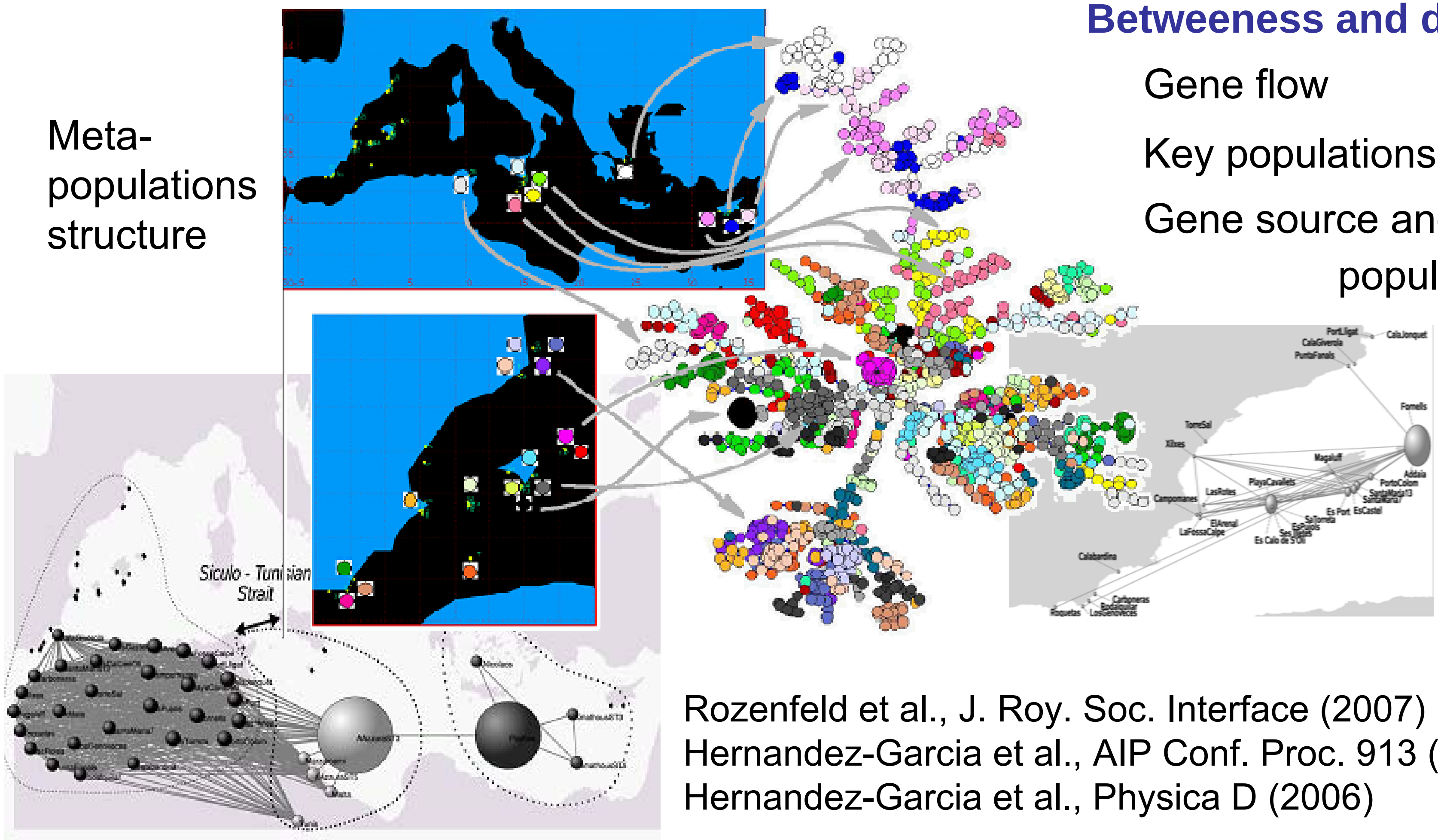
Meta-
populations
structure

Betweeness and degree:

Gene flow

Key populations

Gene source and sink
populations



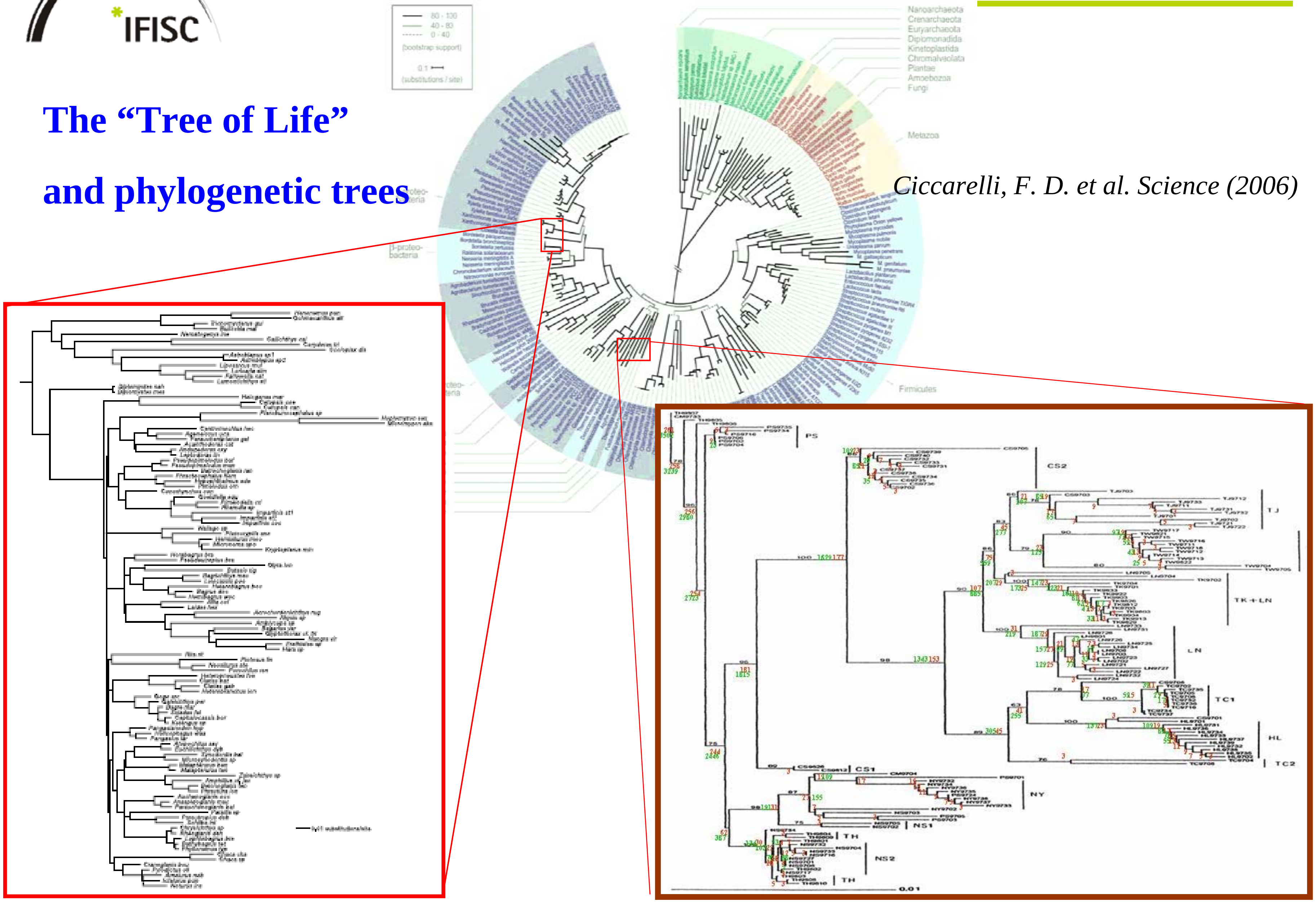
Rozenfeld et al., J. Roy. Soc. Interface (2007)
Hernandez-Garcia et al., AIP Conf. Proc. 913 (2007)
Hernandez-Garcia et al., Physica D (2006)

Ifisc.uib-csic.es/EDEN

ifisc.uib-cisc.es/publications/
ifisc.uib-csic.es/eng/lines/bio.html

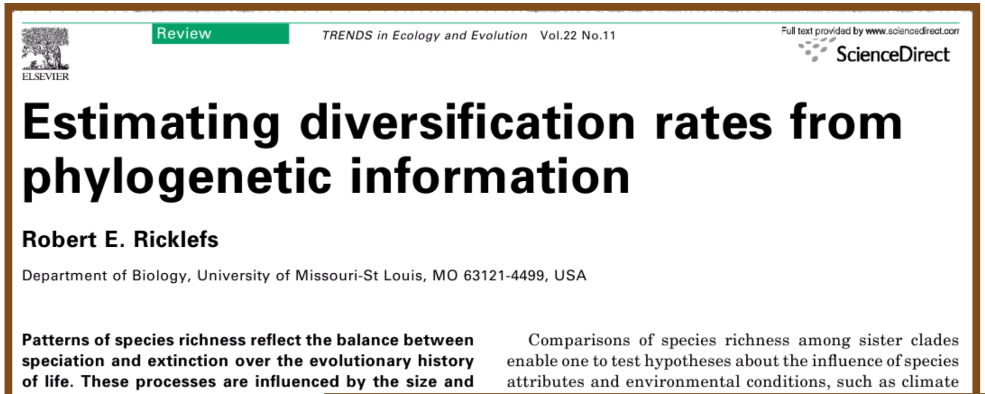
Universal branching in phylogenetic trees

The “Tree of Life” and phylogenetic trees



Ciccarelli, F. D. et al. Science (2006)

Universal branching in phylogenetic trees



Review TRENDS in Ecology and Evolution Vol.22 No.11
Full text provided by www.sciencedirect.com
ScienceDirect
ELSEVIER

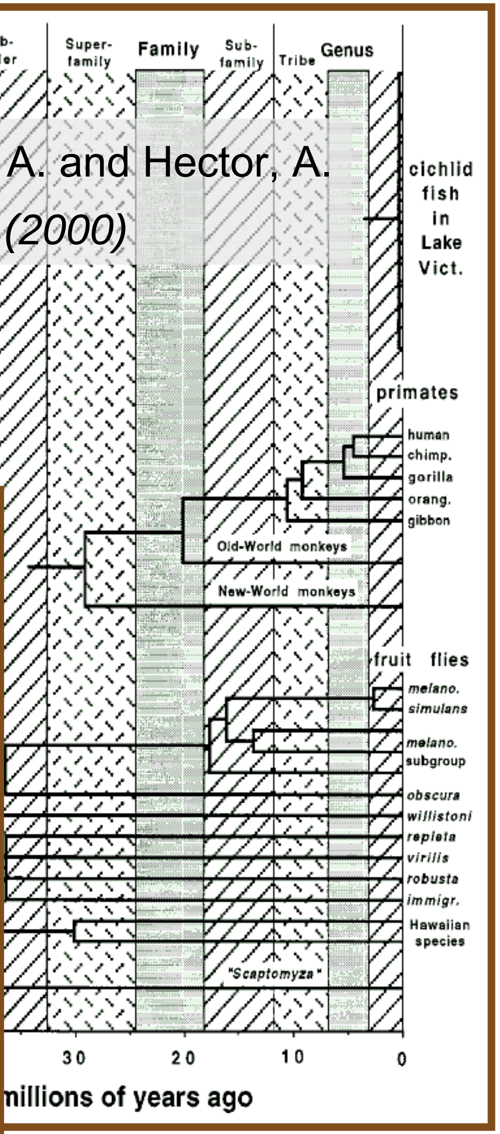
Estimating diversification rates from phylogenetic information

Robert E. Ricklefs

Department of Biology, University of Missouri-St Louis, MO 63121-4499, USA

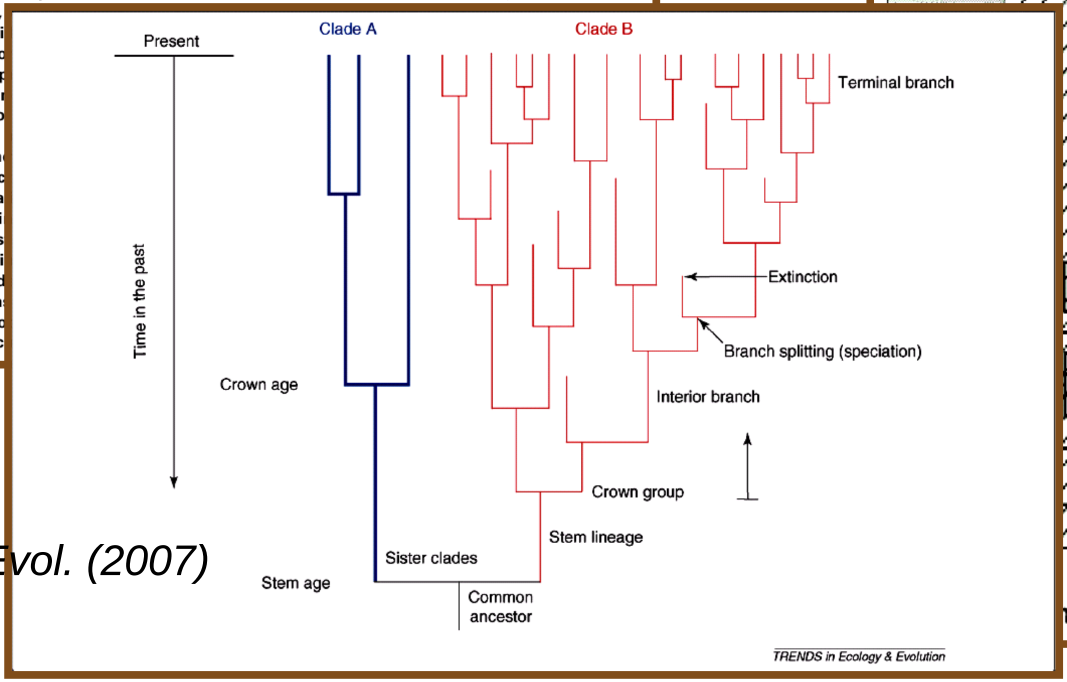
**Patterns of species richness reflect the balance between speciation and extinction over the evolutionary history of life. These processes are influenced by the size and
geographical complexity of regions,
environment, and attributes of indivi
Diversity within clades also depends o
time available for accumulating sp
rates of diversification is key to ur
these factors have shaped patterns o
Several approaches to calculating
absolute rates of speciation and extin
are based on phylogenetic reconstruc
ary relationships. As the size and qua
increases, these approaches will fi
cation. However, phylogeny recons
perceptual bias of continual increase i
and the analysis of primarily large clad
selection bias. Recognizing these bia
the development of more realistic mo
tion and the regulation of species ric**

Comparisons of species richness among sister clades enable one to test hypotheses about the influence of species attributes and environmental conditions, such as climate



Purvis, A. and Hector, A.

Nature (2000)



Ricklefs, R. E.

Trends Ecol. Evol. (2007)

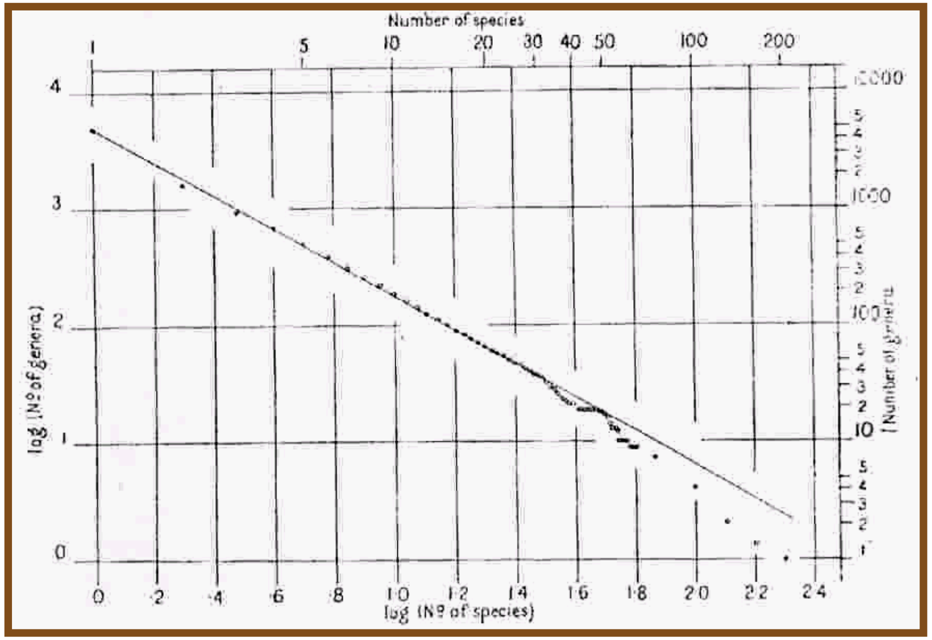
CLASSICAL STUDIES OF PHYLOGENETIC BIODIVERSITY

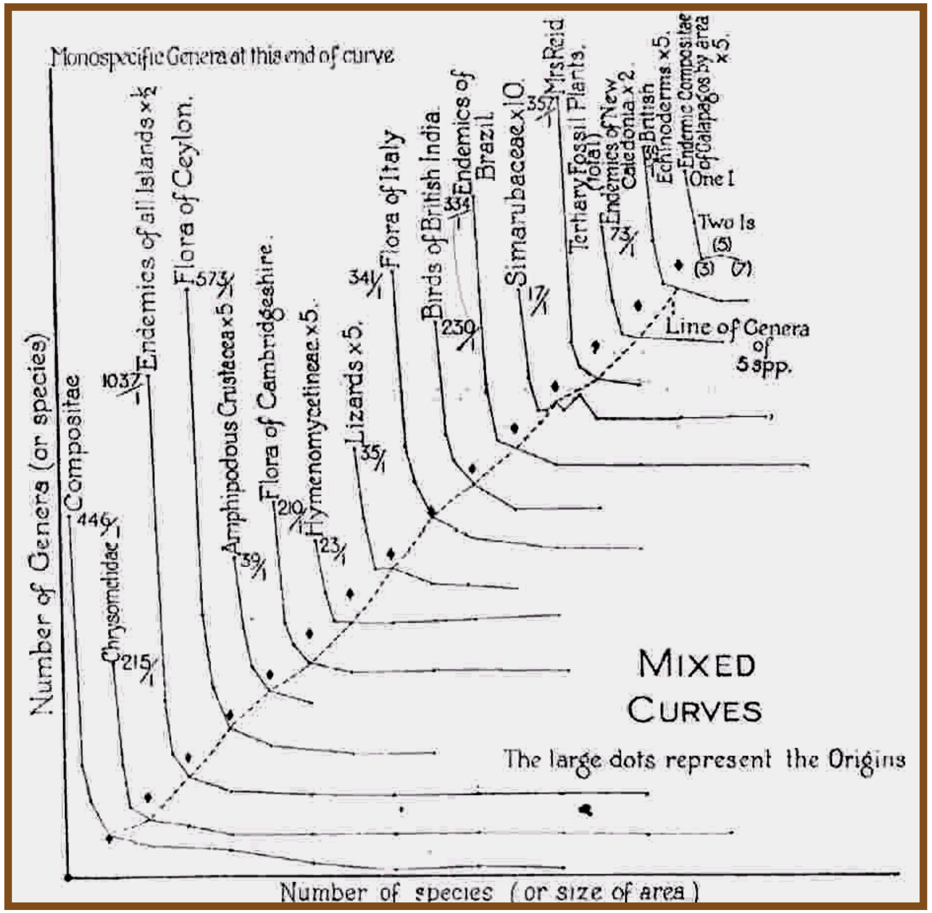
- Distributions of species within genera (or of subtaxons within taxons)
- Symmetry of the branching: balance vs unbalance

Size (Frequency distribution) ...

Numbers of species in biological taxa

Willis, J. C. and Yule, G. U. *Nature (1922)*





$$n_g(n_s) \sim n_s^{\gamma} \;, \quad \gamma \approx 1.5–2$$

Universal branching in phylogenetic trees

Darwin **(1859)**

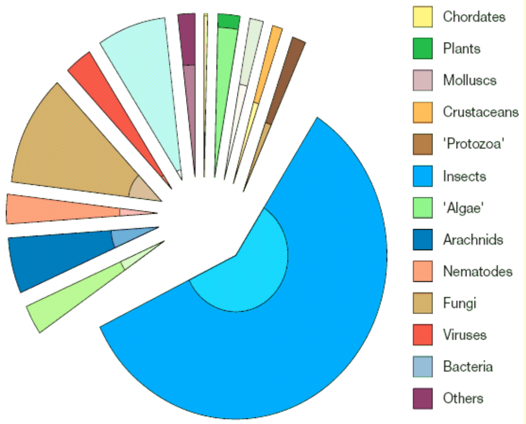
“Species in species-rich genera had more subspecific varieties.”

Willis **(1922),** ***Corbet*** **(1942),** ***Anderson*** **(1975).**

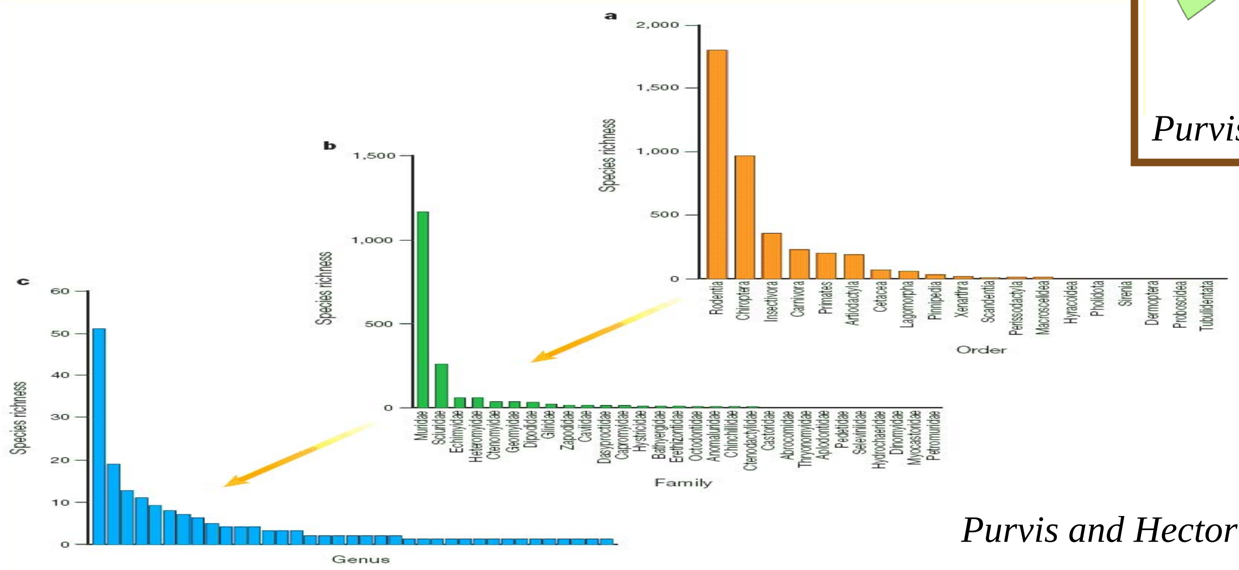
Subtaxa within taxa are often distributed unevenly.

Burlando **(1990, 1993),** ***Minelli*** **(1991),** ***Nee*** **(1992).**

Described the uneven distribution in terms of **fractal geometry**.



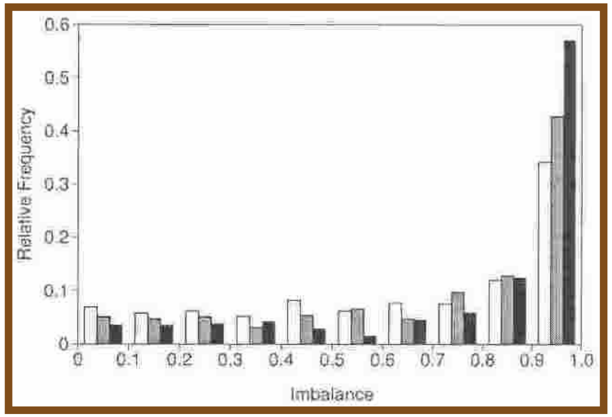
Purvis and Hector Nature (2000)



Purvis and Hector Nature (2000)

Tree shape ...

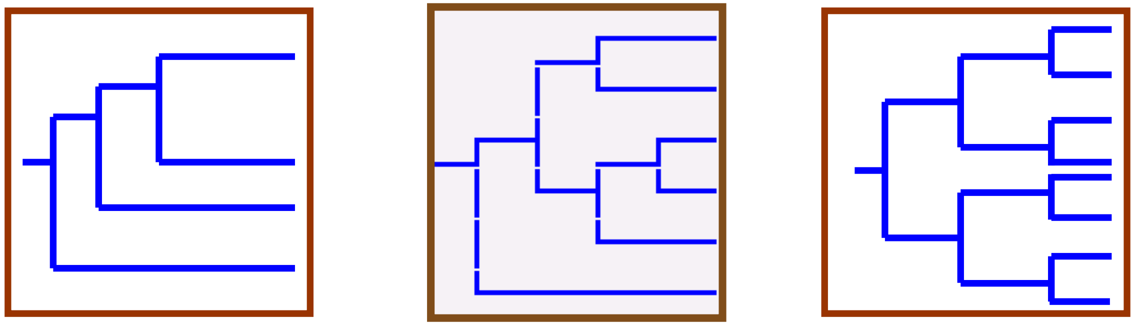
Holman, E. W. *Sys. Biol. (2005)*



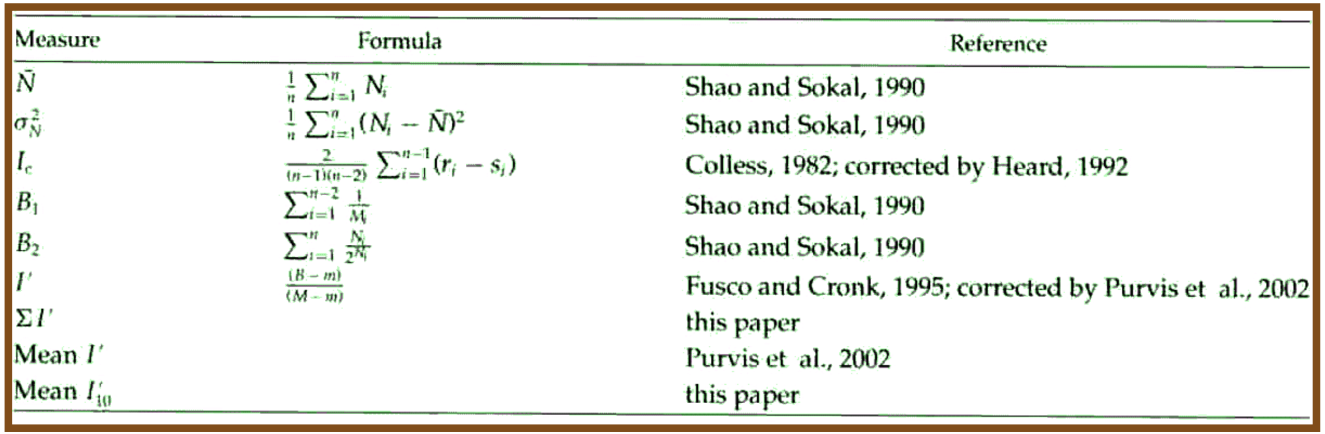
Colless (1982). Colless index of imbalance

Savage (1983). Patterns in the frequencies of all possible tree topologies.

UNBALANCED



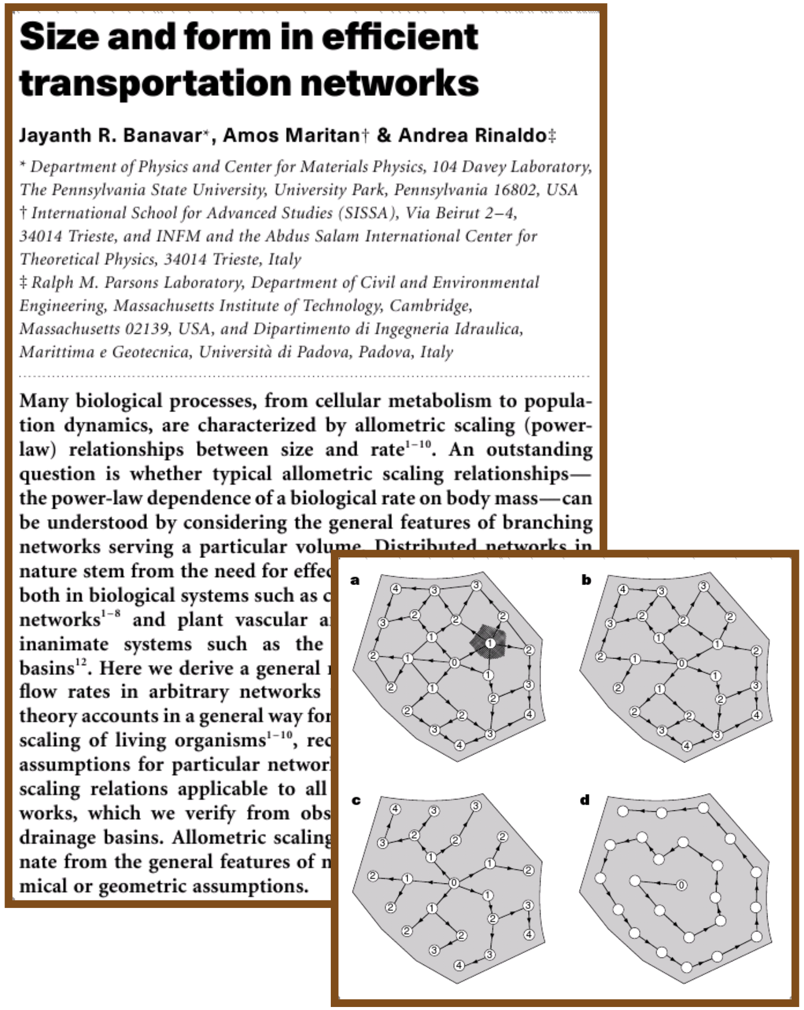
BALANCED



Measure	Formula	Reference
$\bar{N}$	$\frac{1}{n}\sum_{i=1}^{n} N_i$	Shao and Sokal, 1990
σ_N^2	$\frac{1}{n}\sum_{i=1}^{n}(N_i - \bar{N})^2$	Shao and Sokal, 1990
I_c	$\frac{2}{(n-1)(n-2)}\sum_{i=1}^{n-1}(r_i - s_i)$	Colless, 1982; corrected by Heard, 1992
B_1	$\sum_{i=1}^{n-2}\frac{1}{M_i}$	Shao and Sokal, 1990
B_2	$\sum_{i=1}^{n}\frac{N_i}{2^{N_i}}$	Shao and Sokal, 1990
I'	$\frac{(B-m)}{(M-m)}$	Fusco and Cronk, 1995; corrected by Purvis et al., 2002
$\Sigma I'$		this paper
Mean I'		Purvis et al., 2002
Mean I'_{10}		this paper

Agapow, P.-M. and Purvis, A. *Sys. Biol. (2002)*

Universal branching in phylogenetic trees



Size and form in efficient transportation networks

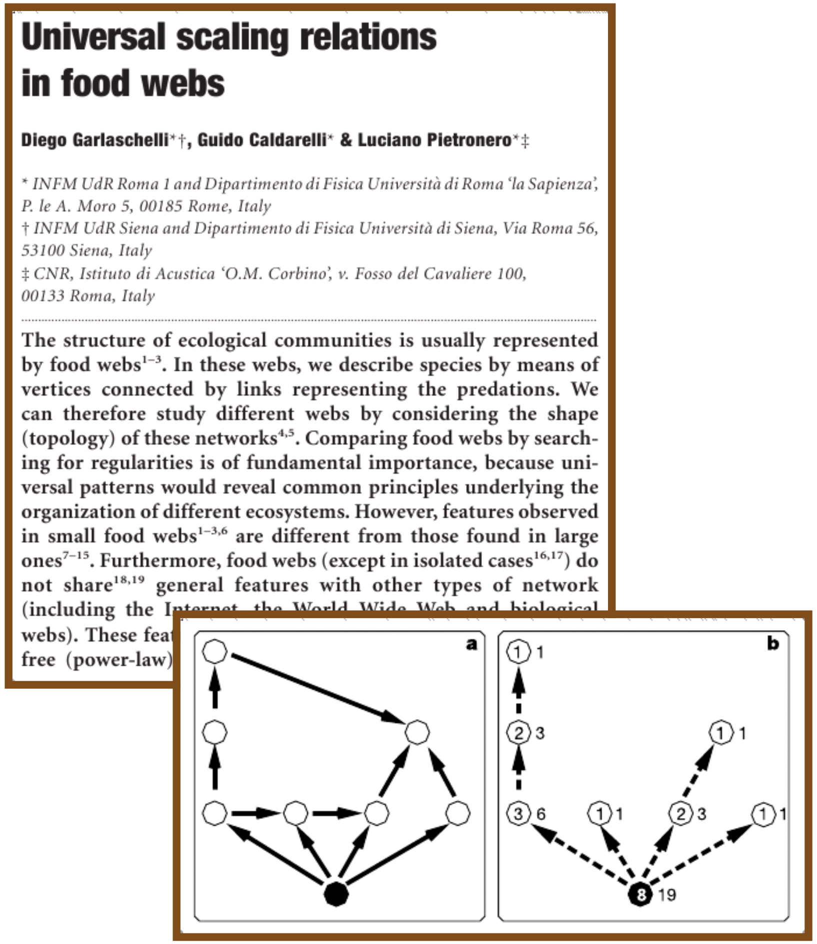
Jayanth R. Banavar*, Amos Maritan† & Andrea Rinaldo‡

* *Department of Physics and Center for Materials Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
† *International School for Advanced Studies (SISSA), Via Beirut 2–4, 34014 Trieste, and INFM and the Abdus Salam International Center for Theoretical Physics, 34014 Trieste, Italy*
‡ *Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA, and Dipartimento di Ingegneria Idraulica, Marittima e Geotecnica, Università di Padova, Padova, Italy*

Many biological processes, from cellular metabolism to population dynamics, are characterized by allometric scaling (power-law) relationships between size and rate[1–10]. An outstanding question is whether typical allometric scaling relationships—the power-law dependence of a biological rate on body mass—can be understood by considering the general features of branching networks serving a particular volume. Distributed networks in
nature stem from the need for effec
both in biological systems such as c
networks[1–8] and plant vascular a
inanimate systems such as the
basins[12]. Here we derive a general
flow rates in arbitrary networks
theory accounts in a general way for
scaling of living organisms[1–10], rec
assumptions for particular networ
scaling relations applicable to all
works, which we verify from obs
drainage basins. Allometric scaling
nate from the general features of n
mical or geometric assumptions.

Banavar, J. R., Maritan, A. and Rinaldo, A.

Nature (1999)



Universal scaling relations in food webs

Diego Garlaschelli*†, Guido Caldarelli* & Luciano Pietronero*‡

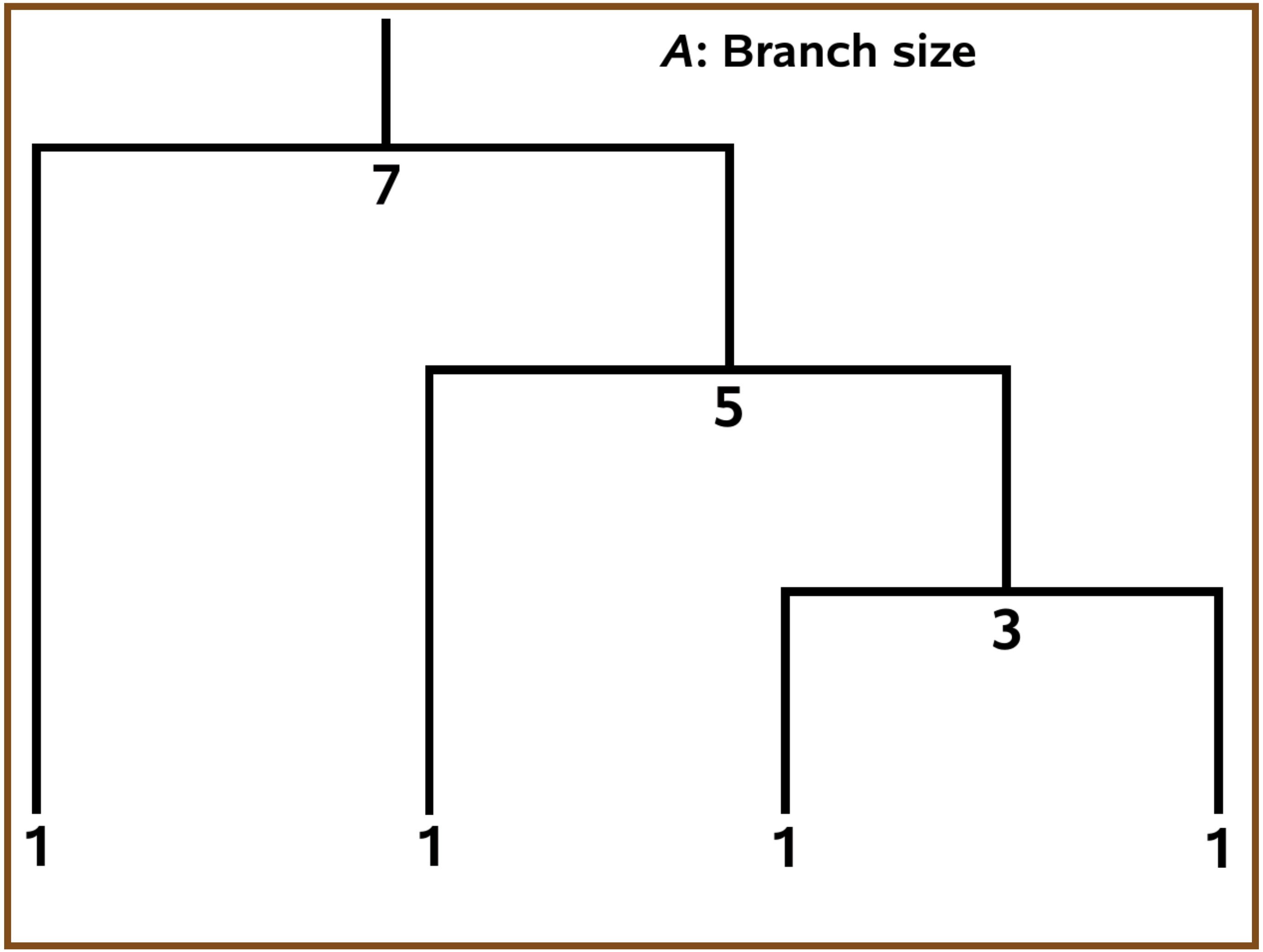
* *INFM UdR Roma 1 and Dipartimento di Fisica Università di Roma 'la Sapienza', P. le A. Moro 5, 00185 Rome, Italy*
† *INFM UdR Siena and Dipartimento di Fisica Università di Siena, Via Roma 56, 53100 Siena, Italy*
‡ *CNR, Istituto di Acustica 'O.M. Corbino', v. Fosso del Cavaliere 100, 00133 Roma, Italy*

The structure of ecological communities is usually represented by food webs[1–3]. In these webs, we describe species by means of vertices connected by links representing the predations. We can therefore study different webs by considering the shape (topology) of these networks[4,5]. Comparing food webs by searching for regularities is of fundamental importance, because universal patterns would reveal common principles underlying the organization of different ecosystems. However, features observed in small food webs[1–3,6] are different from those found in large ones[7–15]. Furthermore, food webs (except in isolated cases[16,17]) do not share[18,19] general features with other types of network (including the Internet, the World Wide Web and biological
webs). These feat
free (power-law)

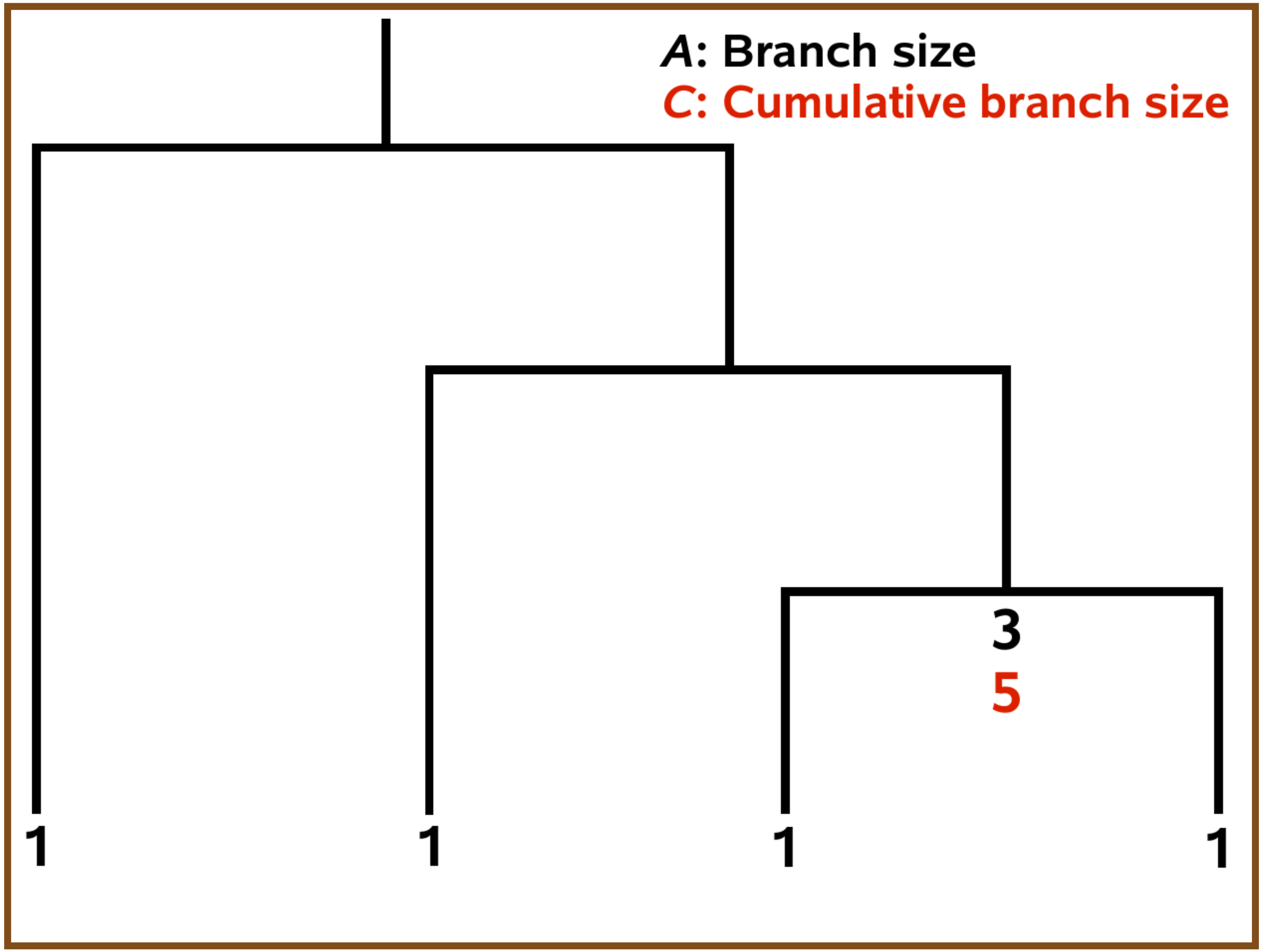
Garlaschelli, D., Caldarelli, G. and Pietronero, L.

Nature (2003)

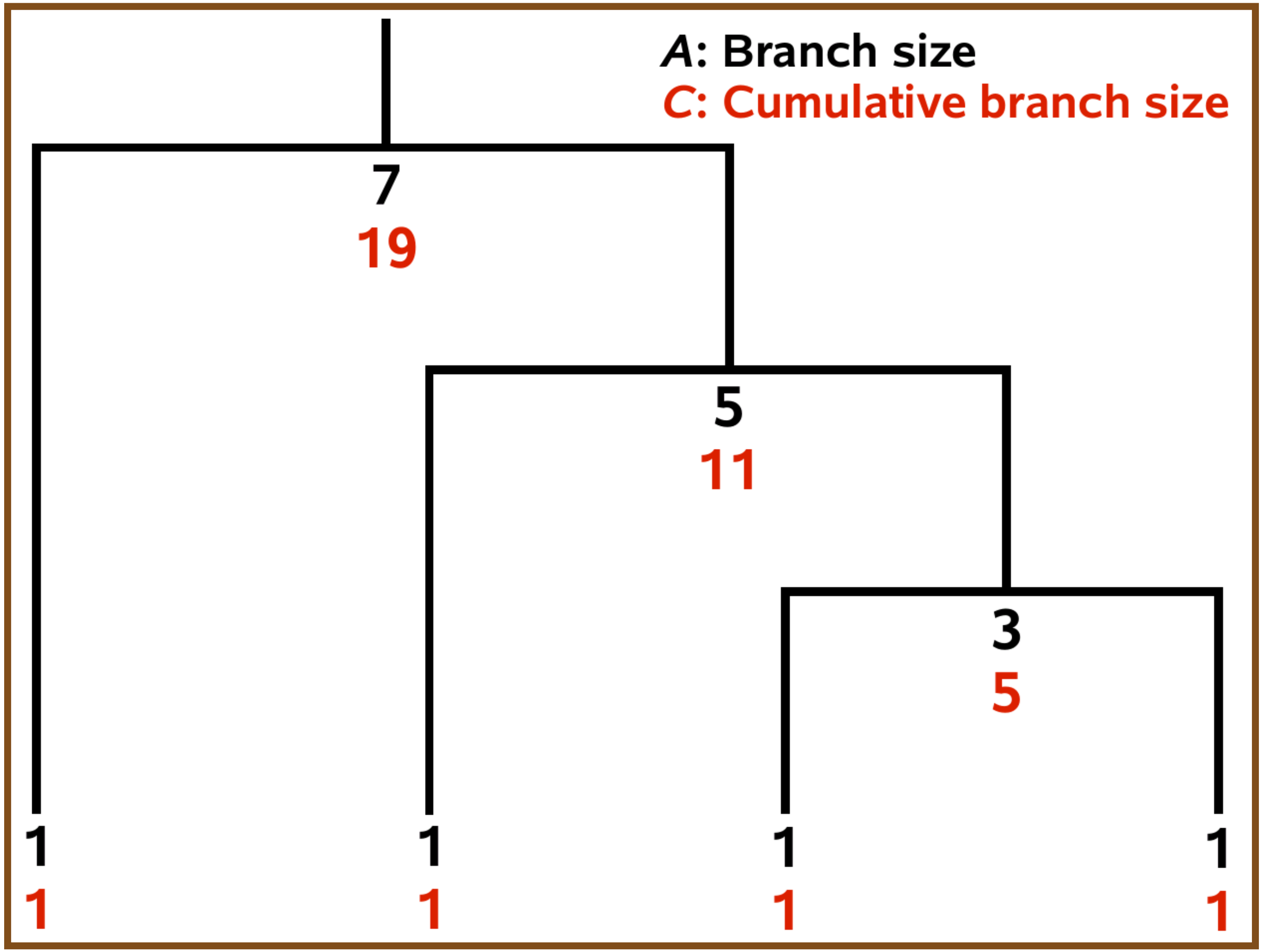
Branch size (A)



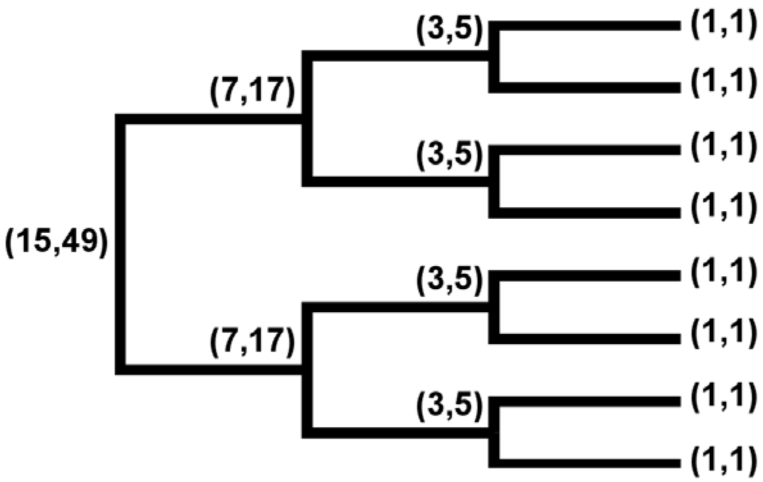
Cumulative branch size (C)

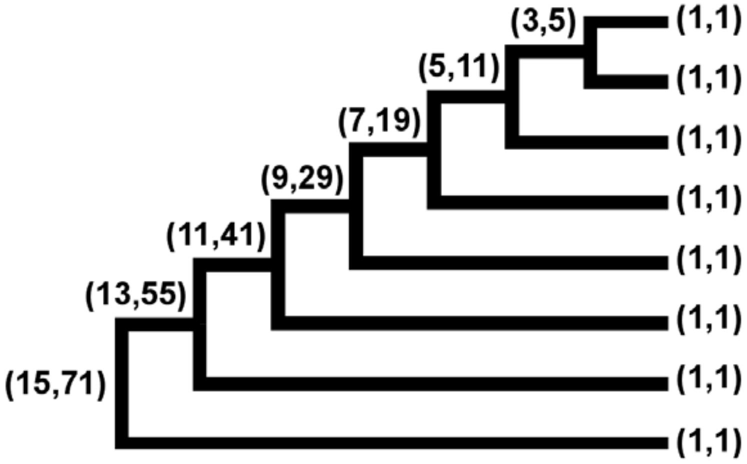


Cumulative branch size (C)

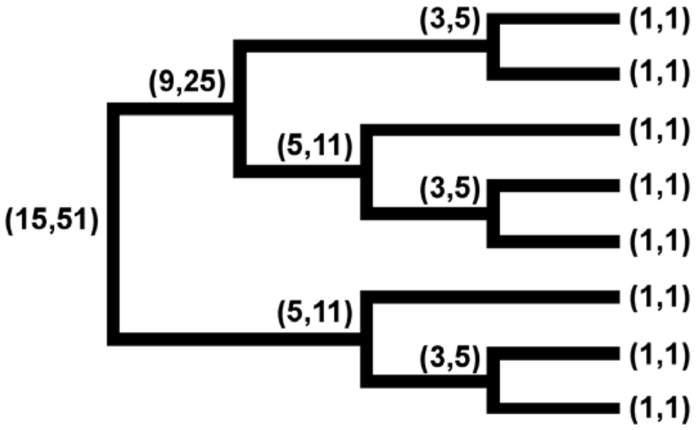


Universal branching in phylogenetic trees





(A_i , C_i)



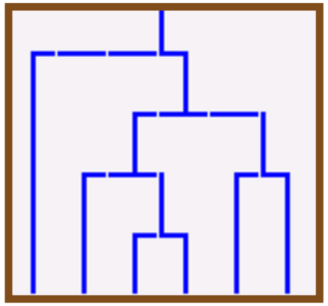
portion of a catfish filogeny

Branch size A, cumulative branch size, C, and classical quantifiers

- *P(A)* $\sim A^{-\tau_A}$ *branch size distribution: related to subtaxa within taxa .* *P(C)* $\sim A^{-\tau_C}$
- C_i/A_i mean distance of leaves to node *i*: mean depth (Sackin's index). Related to Coless index of unbalance.
- $C_i \sim A_i^{\eta}$ *allometric scaling: shape vs size. Related to efficiency measures.*

FULLY UNBALANCED: $\eta = 2$, $\tau_A = 0$, $\tau_C = 0.5$

SIMMETRIC (and random!): $\eta = 1$ (+log), $\tau_A = \tau_C = 2$



AIM:

To explore interesting objects from the point of view of network science, *but also*:

To use this topological characterization of phylogenies to compare **macroevolutionary** processes (leading to speciation) with **microevolution** (leading to population differentiation within the same species)

IFISC

Universal branching in phylogenetic trees

TreeBASE@30June2007: 5212 trees of interspecific phylogenetic relationships

TreeBASE

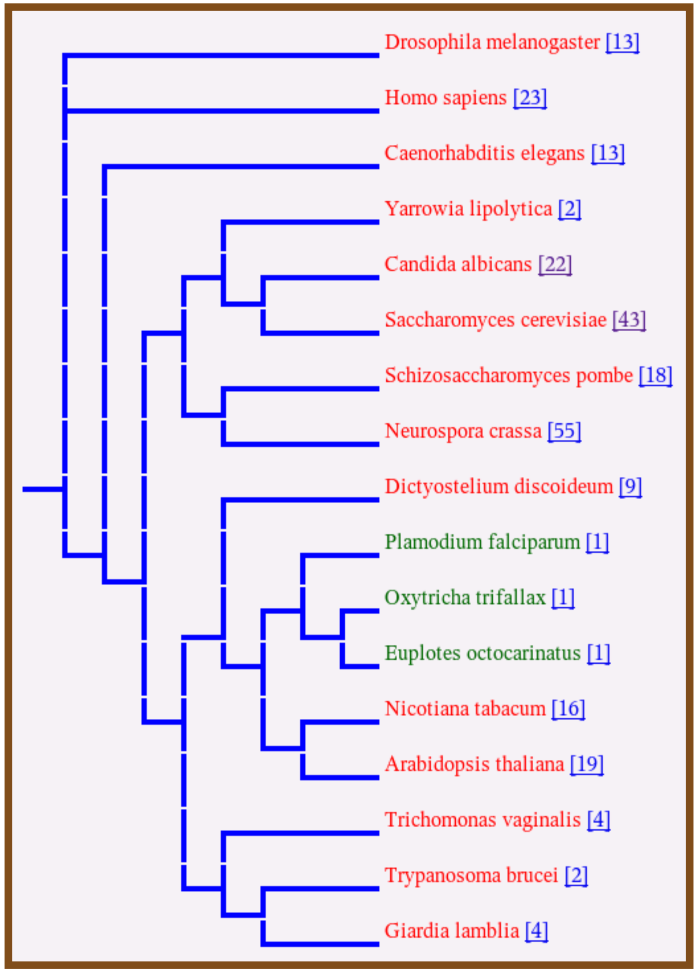
A Database of Phylogenetic Knowledge

TreeBASE is a relational database of phylogenetic information hosted by the University at Buffalo. In previous years the database has been hosted by Harvard University Herbaria, Leiden University EEW, and the University of California, Davis. TreeBASE stores phylogenetic trees and the data matrices used to generate them from published research papers. We encourage biologists to submit phylogenetic data that are either published or in press, especially if these data were not fully presented in the publication due to space limitations. TreeBASE accepts all types of phylogenetic data (e.g., trees of species, trees of populations, trees of genes) representing all biotic taxa. For more information, see an introduction to TreeBASE, information on searching, the database schema, and a graphic presentation of the web site's internal structure. Also, check out some ideas on why you might want to use TreeBASE.

TreeBASE is now a participant in CIPRes , the NSF-sponsored Cyberinfrastructure for Phylogenetic Research project. As such, it is being redesigned from the ground up through collaborative research among Computer Scientists, Biologists, and Programmers. Presently TreeBASE is being mirrored at the San Diego Supercomputer Center at UCSD. Eventually, the redesigned, new and improved CIPRes version of TreeBASE will take over. In the meantime, please send us suggestions of what kinds of features or functions would you like designed into the new database? Are there new or unusual data types , queries, and functions that are not already offered by the current version of TreeBASE? Please send your suggestions here.

The WWW implementation of TreeBASE requires a forms-capable and frames-capable browser. We would be very grateful for any feedback on TreeBASE, including suggestions for improvement. In particular, if you encounter any errors please let us know.

http://www.treebase.org



TreeBASE: 5212 interspecific phylogenies

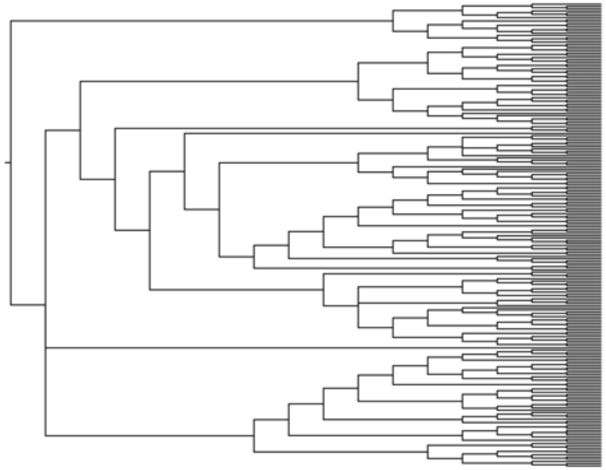
Three data sets

INTRAspecific set of 67 manually selected phylogenies.

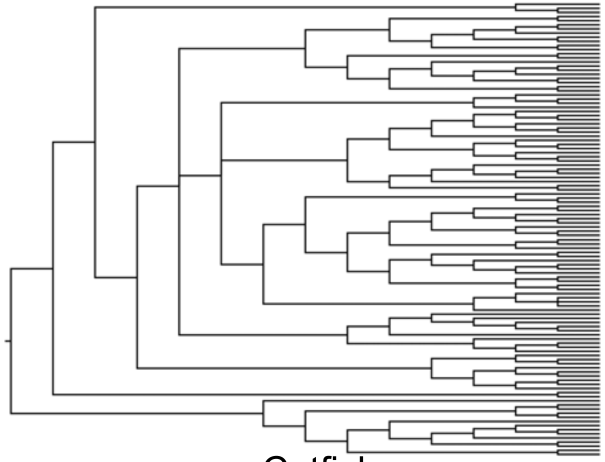
- from all kingdoms
- climatic regions (*terrestrial*, *marine, fresh water…)*
- environments (from polar to desertic)

and reconstruction methods (*neighbour joining*, *maximum parsimony* and *maximum likelihood*)

INTERspecific set of 67 manually selected phylogenies (same criteria)



Vibrio vulnificus



Catfishes

Branch Size distribution

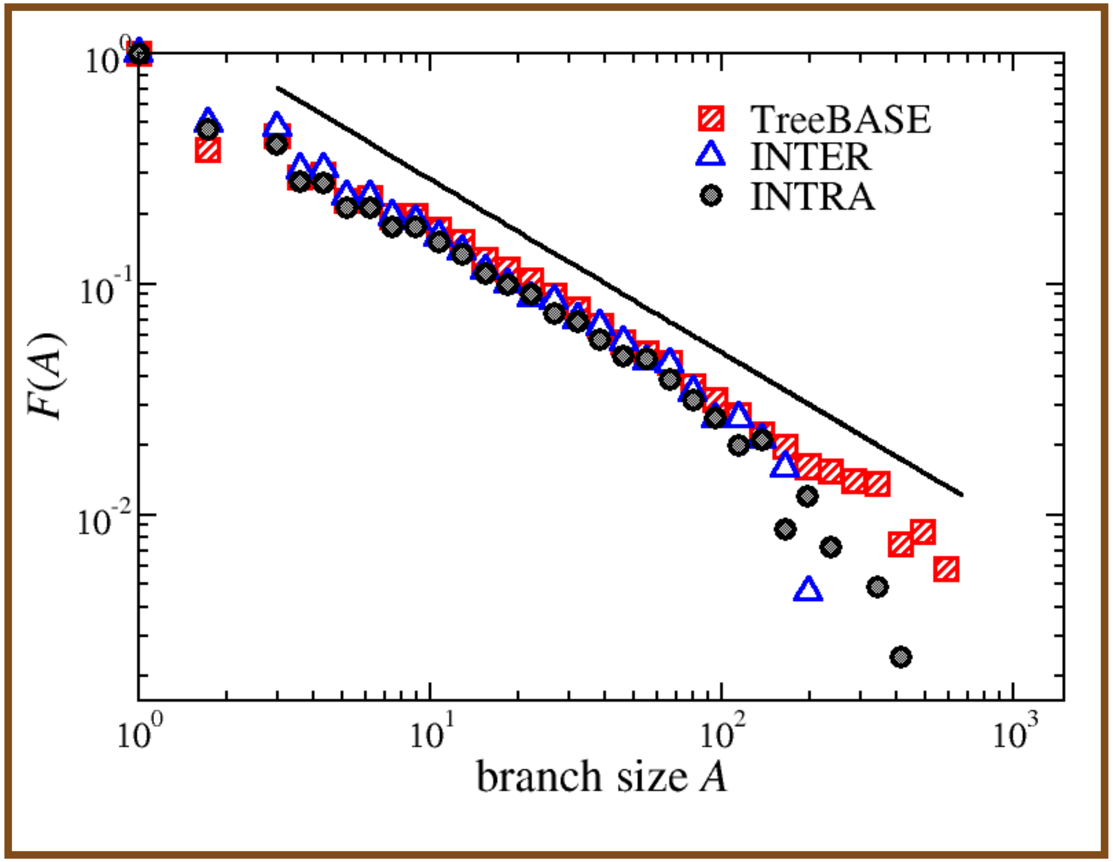
$F(A)$ = prob $(A_i \geq A)$

$$F(A) \sim A^{1-\tau_A}, \; P(A) \sim A^{-\tau_A}$$

τ_A

TreeBASE: 1.74
INTER: 1.73
INTRA: 1.73

Universality across evolutionary scales,
kingdoms,
geographic regions,
…



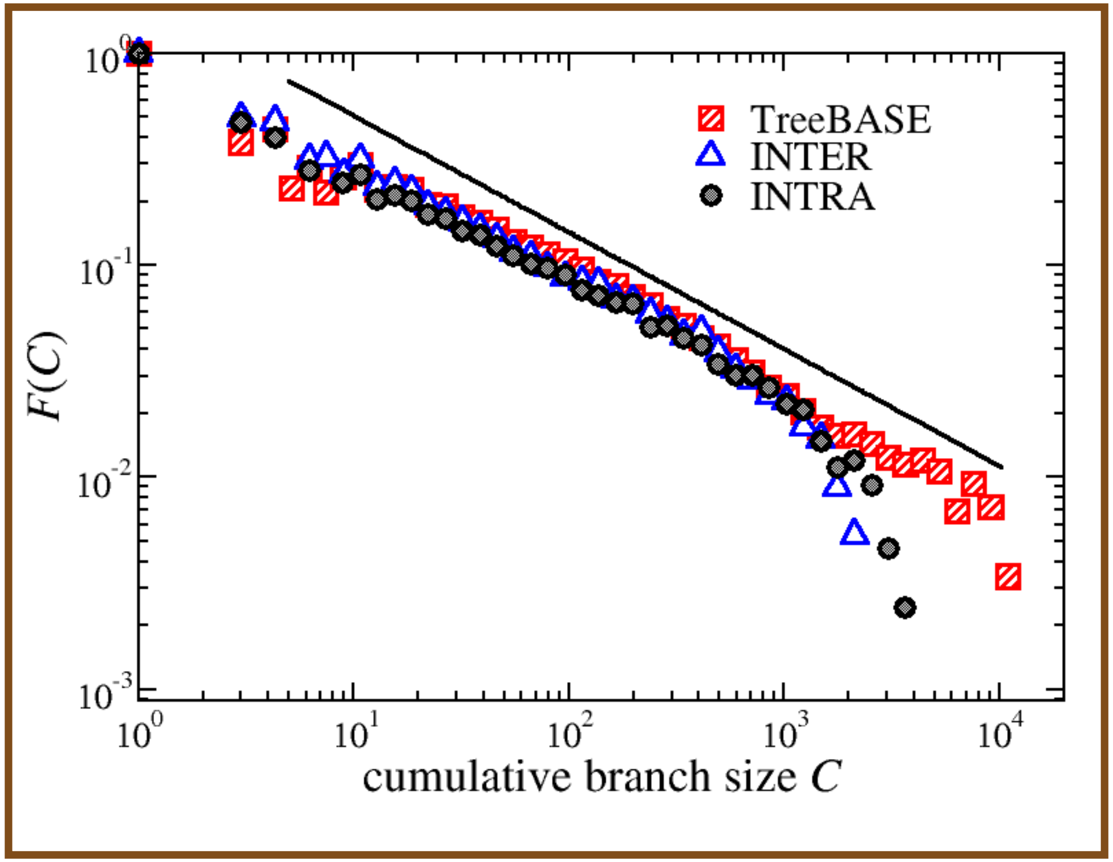
Cumulative Branch Size distribution

$F(C) = \text{prob}\,(C_i \geq C)$

$F(C) \sim A^{1-\tau_C}$, $P(C) \sim A^{-\tau_C}$

τ_C

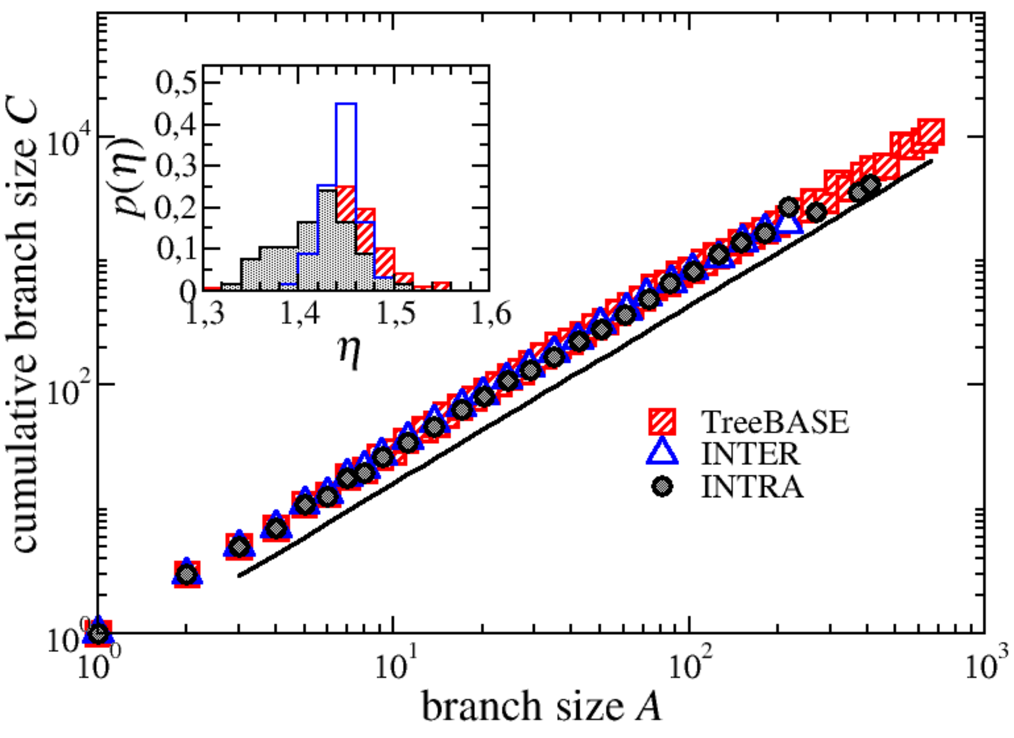
TreeBASE: 1.53
INTER: 1.52
INTRA: 1.51



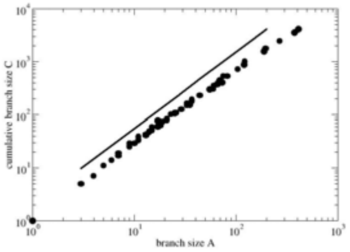
Allometric scaling ($C \approx A^{\eta}$)

η

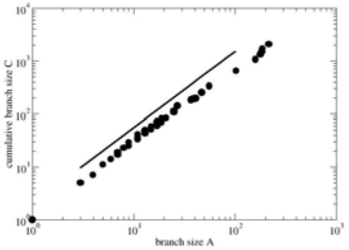
TreeBASE: 1.46
INTER: 1.43
INTRA: 1.43



vibrio



catfish

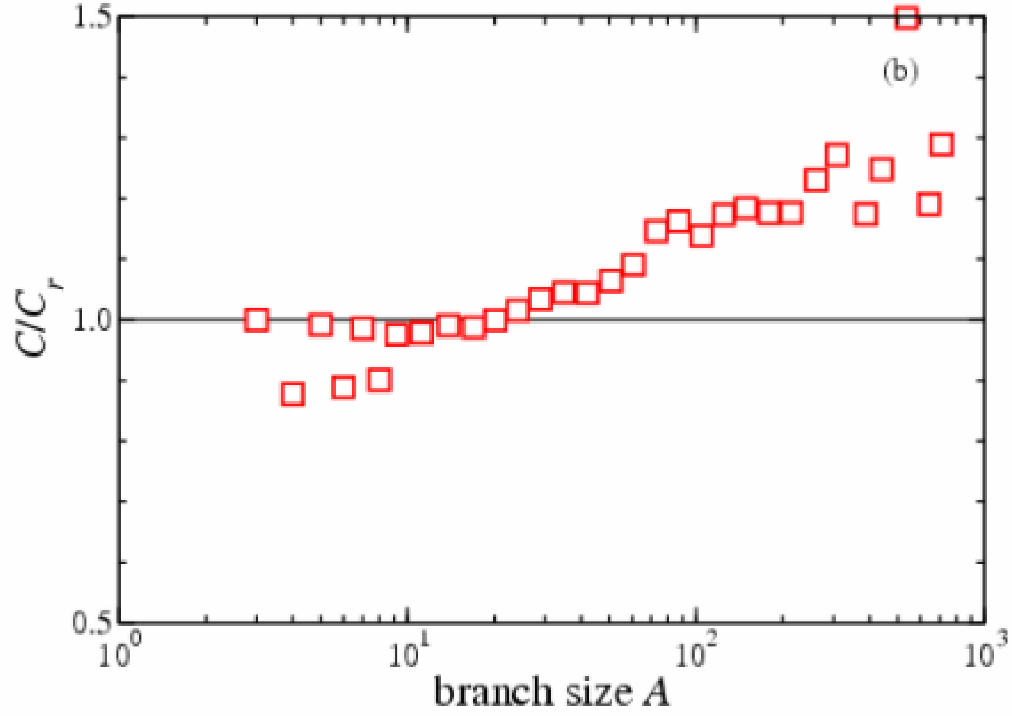


CAN THIS BE EXPLAINED BY RANDOM BRANCHING?

Equal-rates Markov (ERM) model (ERM model):

✓ The phylogenies are the product of random branching.
✓ The "effective speciation rate" (the difference between extinction and speciation rate) is equal for all species.
✓ The process is memoryless.

$C_r \sim A \log A$



TreeBASE data

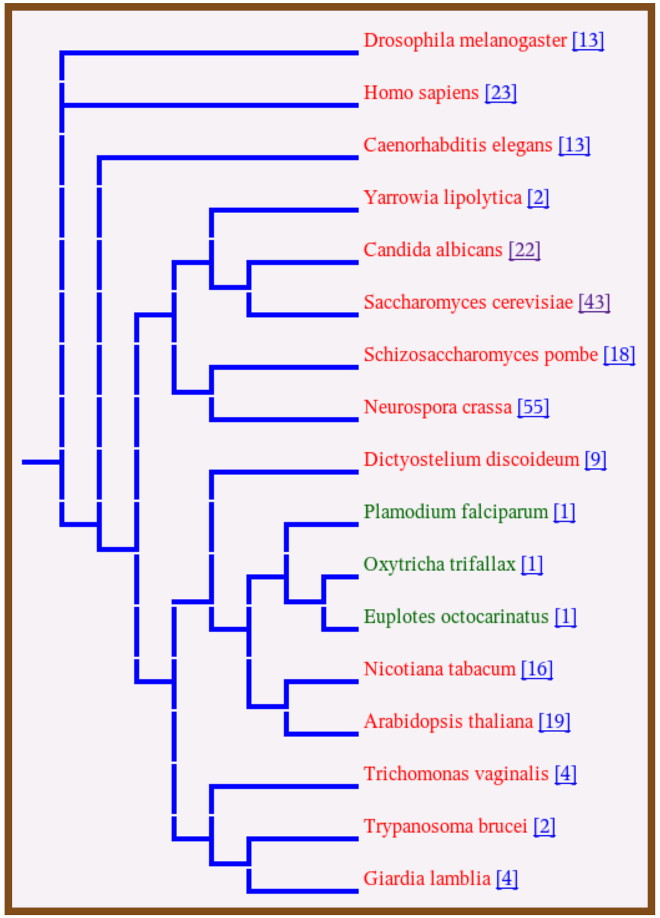
Random-like scaling (η=1, τ_A = 2) is extremely robust: ERM, Yule, Cayley, Moran, coalescent, … any model in which all nodes are equivalent at a given time

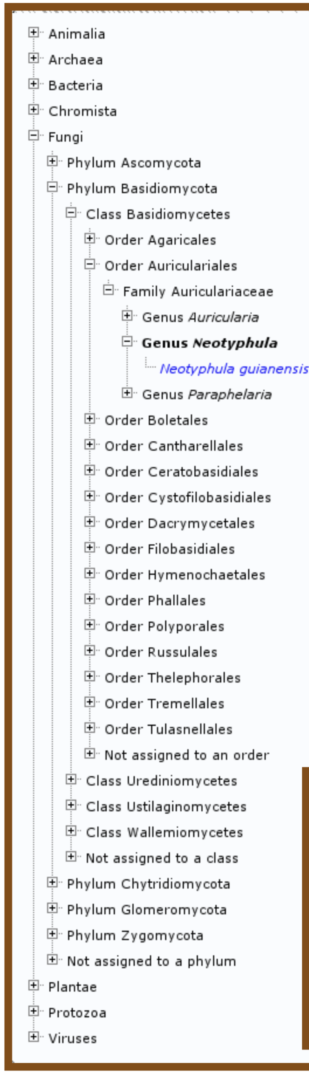
Alternatives:

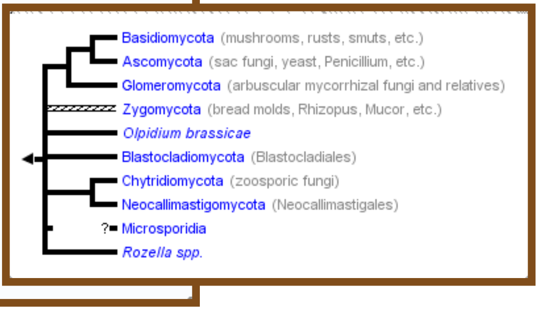
Proportional-to-Distinguishable-Arrangements (PDA) model: Rose (1978)
Aldous beta model (2001)

Universal branching in phylogenetic trees

Phylogenies vs Taxonomies




- ⊞ Animalia
- ⊞ Archaea
- ⊞ Bacteria
- ⊞ Chromista
- ⊟ Fungi
  - ⊞ Phylum Ascomycota
  - ⊟ Phylum Basidiomycota
    - ⊟ Class Basidiomycetes
      - ⊞ Order Agaricales
      - ⊟ Order Auriculariales
        - ⊟ Family Auriculariaceae
          - ⊞ Genus *Auricularia*
          - ⊟ **Genus *Neotyphula***
            - *Neotyphula guianensis*
          - ⊞ Genus *Paraphelaria*
      - ⊞ Order Boletales
      - ⊞ Order Cantharellales
      - ⊞ Order Ceratobasidiales
      - ⊞ Order Cystofilobasidiales
      - ⊞ Order Dacrymycetales
      - ⊞ Order Filobasidiales
      - ⊞ Order Hymenochaetales
      - ⊞ Order Phallales
      - ⊞ Order Polyporales
      - ⊞ Order Russulales
      - ⊞ Order Thelephorales
      - ⊞ Order Tremellales
      - ⊞ Order Tulasnellales
      - ⊞ Not assigned to an order
    - ⊞ Class Urediniomycetes
    - ⊞ Class Ustilaginomycetes
    - ⊞ Class Wallemiomycetes
    - ⊞ Not assigned to a class
  - ⊞ Phylum Chytridiomycota
  - ⊞ Phylum Glomeromycota
  - ⊞ Phylum Zygomycota
  - ⊞ Not assigned to a phylum
- ⊞ Plantae
- ⊞ Protozoa
- ⊞ Viruses



Domain

Kingdom

Phylum

Class

Order

Family

Genus

Species

IFISC

Universal branching in phylogenetic trees

Catalogue of Life

Species 2000 ITIS

Catalogue of Life: 2007 Annual Checklist

indexing the world's known species

Search the Annual Checklist - fixed edition each year

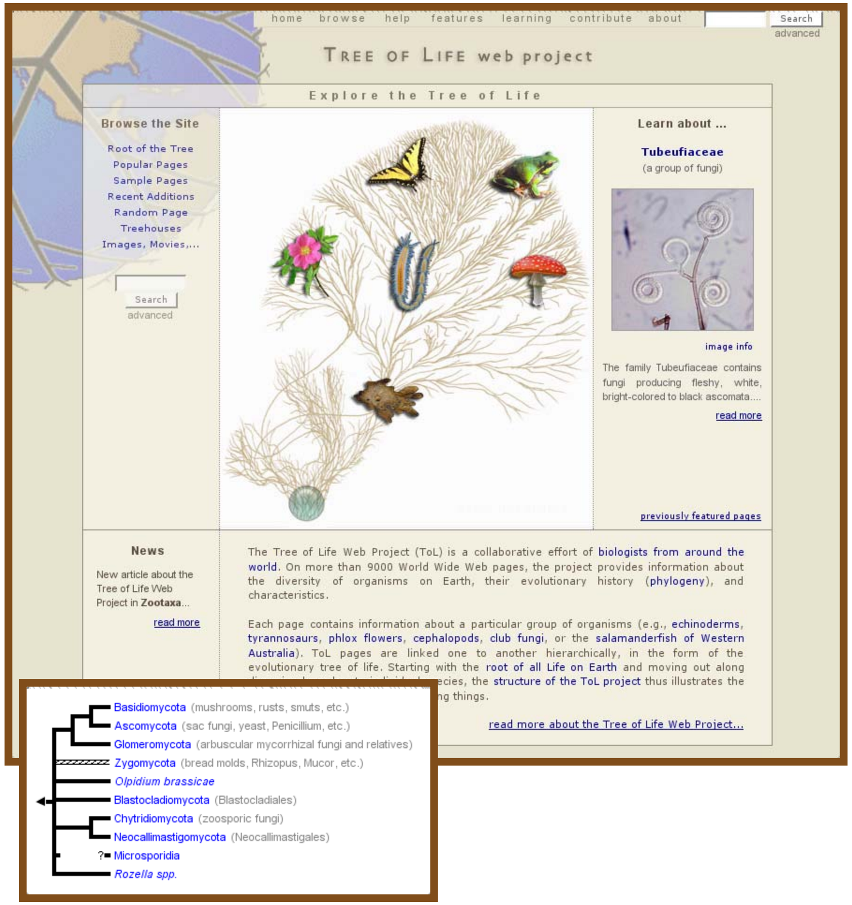
Search for:

☑ Match whole words only

- Animalia
- Archaea
- Bacteria
- Chromista
- Fungi
  - Phylum Ascomycota
  - Phylum Basidiomycota
    - Class Basidiomycetes
      - Order Agaricales
      - Order Auriculariales
        - Family Auriculariaceae
          - Genus *Auricularia*
          - **Genus *Neotyphula***
            - *Neotyphula guianensis*
          - Genus *Paraphelaria*
      - Order Boletales
      - Order Cantharellales
      - Order Ceratobasidiales
      - Order Cystofilobasidiales
      - Order Dacrymycetales
      - Order Filobasidiales
      - Order Hymenochaetales
      - Order Phallales
      - Order Polyporales
      - Order Russulales
      - Order Thelephorales
      - Order Tremellales
      - Order Tulasnellales
      - Not assigned to an order
    - Class Urediniomycetes
    - Class Ustilaginomycetes
    - Class Wallemiomycetes
    - Not assigned to a class
  - Phylum Chytridiomycota
  - Phylum Glomeromycota
  - Phylum Zygomycota
  - Not assigned to a phylum
- Plantae
- Protozoa
- Viruses

http://www.catalogueoflife.org

Tolweb



TREE OF LIFE web project

Explore the Tree of Life

Browse the Site

Learn about ...

Tubeufiaceae
(a group of fungi)

The family Tubeufiaceae contains fungi producing fleshy, white, bright-colored to black ascomata....

News

New article about the Tree of Life Web Project in **Zootaxa**...

The Tree of Life Web Project (ToL) is a collaborative effort of biologists from around the world. On more than 9000 World Wide Web pages, the project provides information about the diversity of organisms on Earth, their evolutionary history (phylogeny), and characteristics.

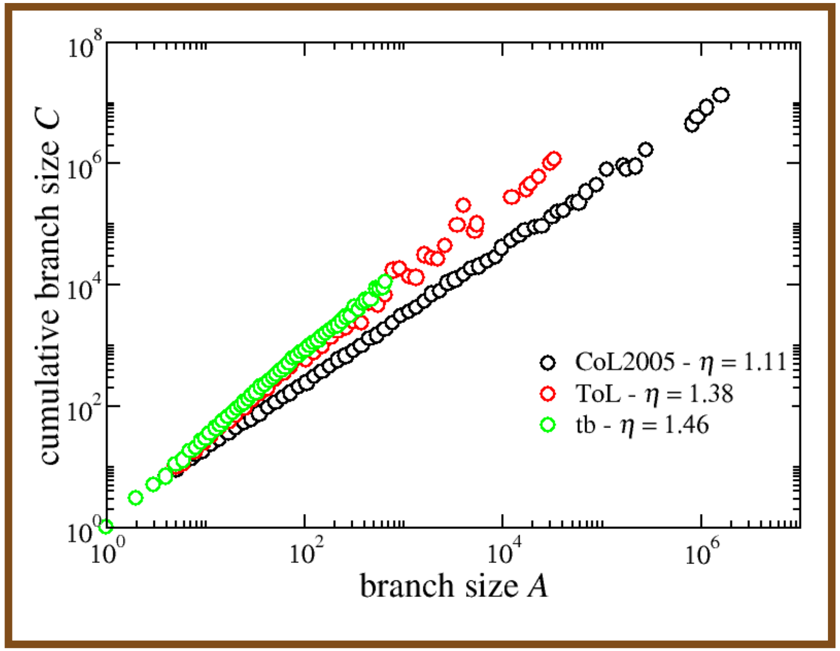
Each page contains information about a particular group of organisms (e.g., echinoderms, tyrannosaurs, phlox flowers, cephalopods, club fungi, or the salamanderfish of Western Australia). ToL pages are linked one to another hierarchically, in the form of the evolutionary tree of life. Starting with the root of all Life on Earth and moving out along

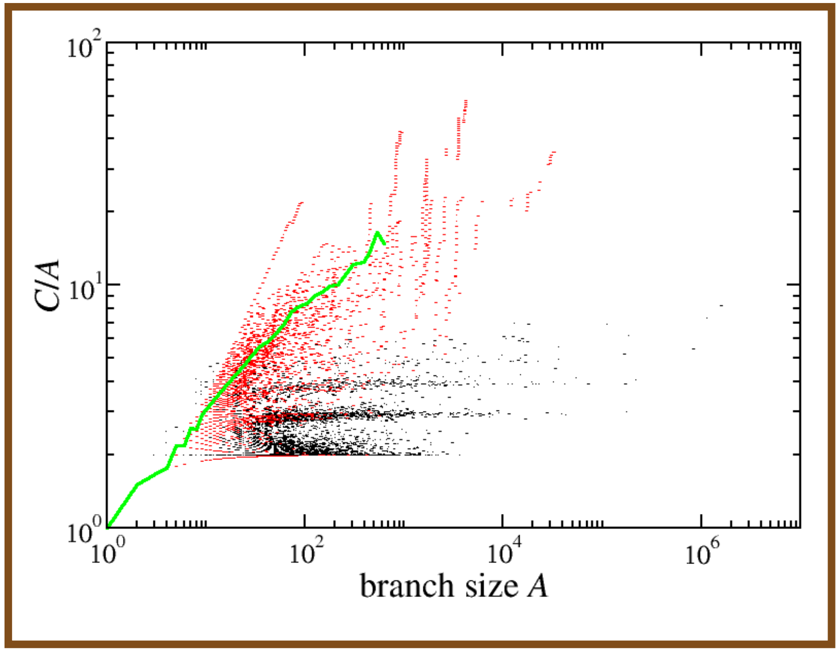
ecies, the structure of the ToL project thus illustrates the

ng things.

http://www.tolweb.org

Allometric scaling ($C \approx A^{\eta}$)





Catalog of life

ToLweb

TreeBASE

Conclusions...

➔ We have characterized large sets of phylogenies with **topological quantities** which combine previous approaches based on diversity statistics, unbalance, and efficiency of transportation networks.

➔ There is a striking **universality** across domains of life, environments, and scales of evolution of the **branching behavior** of the above topological indices. In particular we found the same branching patterns at the macroevolutionary scale (species formation) as at the microevolutionary/ecological scale (population differentiation).

➔ **Power-law distributions** and **allometric scaling** characterize the branching properties as a function of tree size (A).

➔ Simple **random branching models do not explain** the observed behavior. Memory of previous diversification history seems to be needed.

➔ **Taxonomical** classifications have rather different topological properties

ifisc.uib-cisc.es/publications/

Ifisc.uib-csic.es/EDEN ifisc.uib-csic.es/research/research_bio.php