Palma de Mallorca, February 2008

From genes to species: Universal Scaling?

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

alejandro@ifisc.uib.es

[1]IFISC, Institut de Física Interdisciplinària i Sistemes Complexos (CSIC-UIB),
E-07122 Palma de Mallorca, Spain
[2]ETH Zürich, KPL F 31.1, Kreuzplatz 5, CH-8032 Zürich, Switzerland
[3]Bioinformatics Group, Department of Computer Science,
University of Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany
[4]IMEDEA, Institut Mediterrani d'Estudis Avançats (CSIC-UIB),
C/ Miquel Marqués 21, E-07190 Esporles, Spain

*IFISC

From genes to species: Universal Scaling?

Review TRENDS in Ecology and Evolution Vol.22 No.11

ELSEVIER

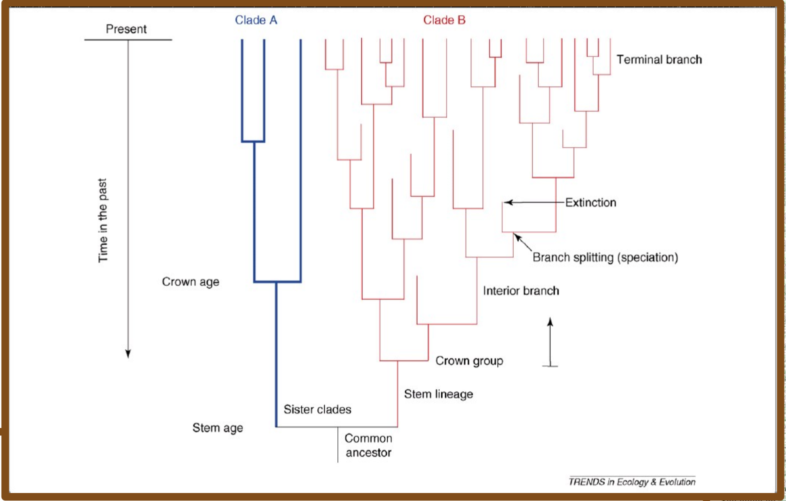
Full text provided by www.sciencedirect.com ScienceDirect

Estimating diversification rates from phylogenetic information

Robert E. Ricklefs

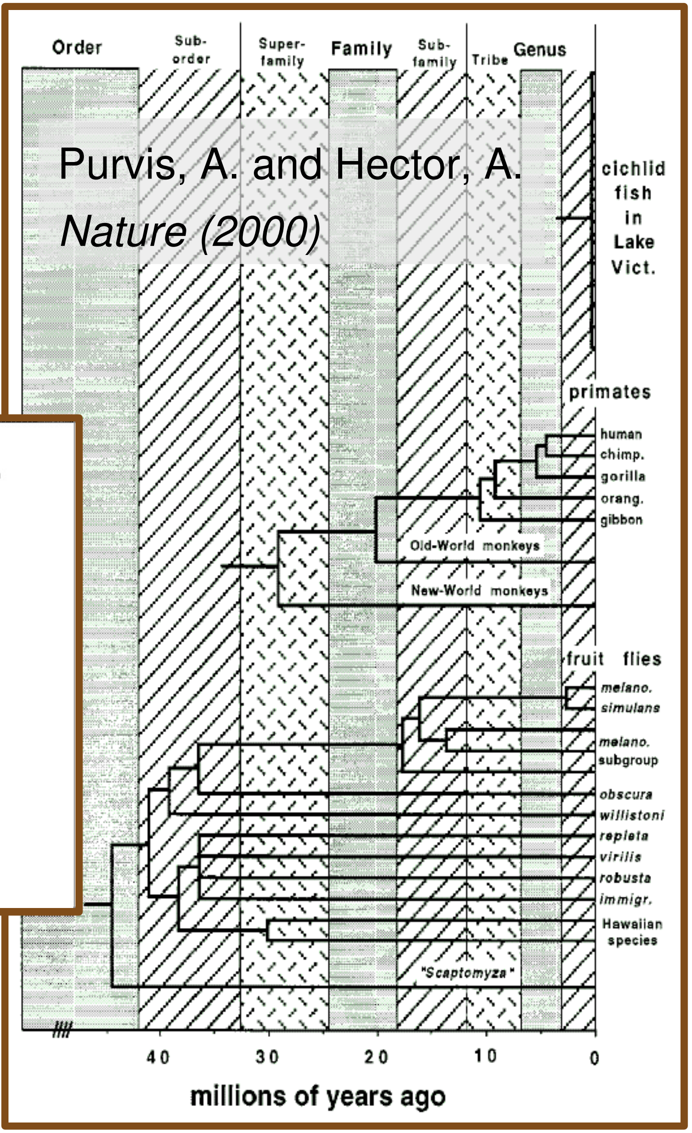
Department of Biology, University of Missouri-St Louis, MO 63121-4499, USA

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, conditions of the environment, and attributes of individuals and species. Diversity within clades also depends on age and thus the time available for accumulating species. Estimating rates of diversification is key to understanding how these factors have shaped patterns of species richness. Several approaches to calculating both relative and absolute rates of speciation and extinction within clades are based on phylogenetic reconstructions of evolutionary relationships. As the size and quality of phylogenies increases, these approaches will find broader application. However, phylogeny reconstruction fosters a perceptual bias of continual increase in species richness, and the analysis of primarily large clades produces a data selection bias. Recognizing these biases will encourage the development of more realistic models of diversification and the regulation of species richness.



Ricklefs, R. E.

Trends Ecol. Evol. (2007)



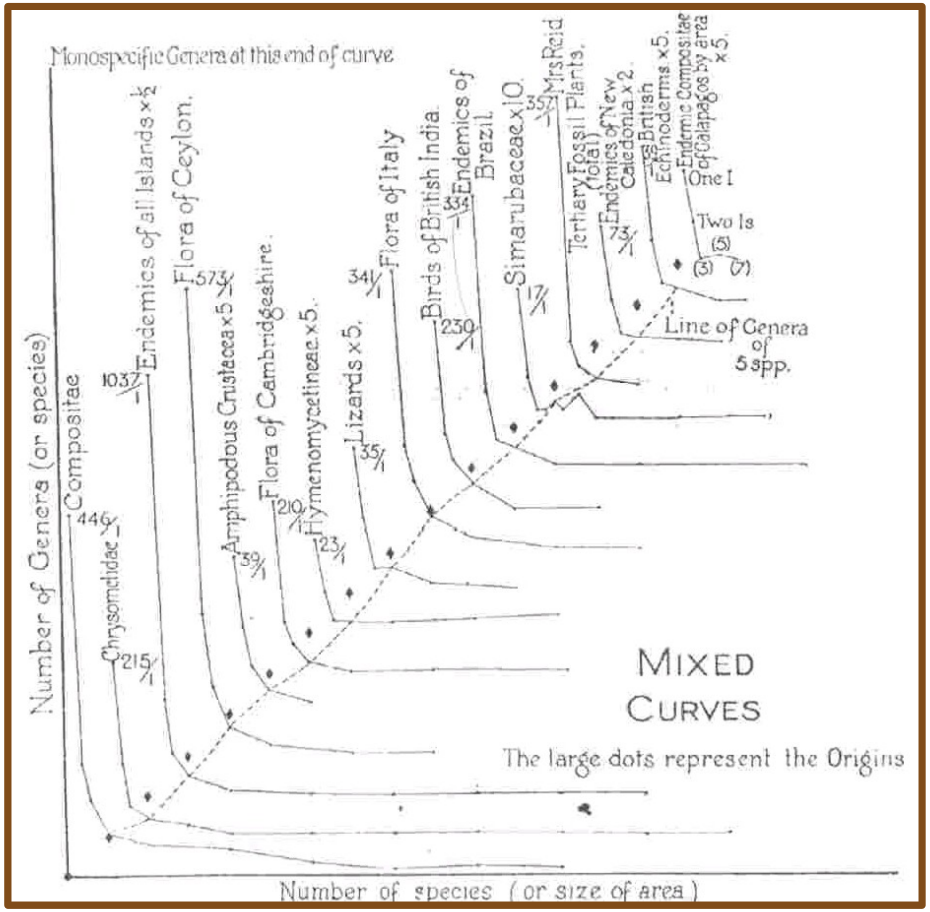
Purvis, A. and Hector, A.

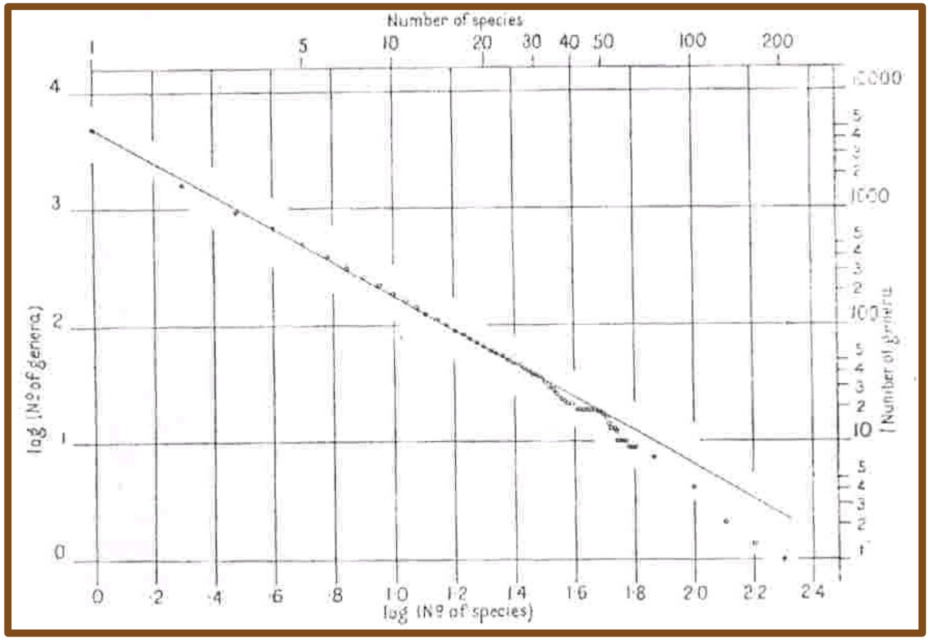
Nature (2000)

Size (Frequency distribution) ...

Numbers of species in biological taxa

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





***Darwin* (1859)**

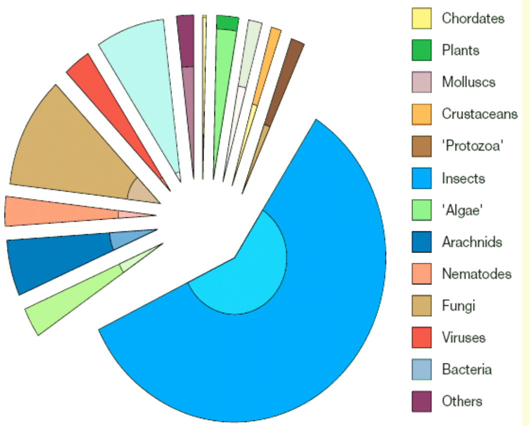
“Species in species-rich genera had more subspecific varieties.”

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

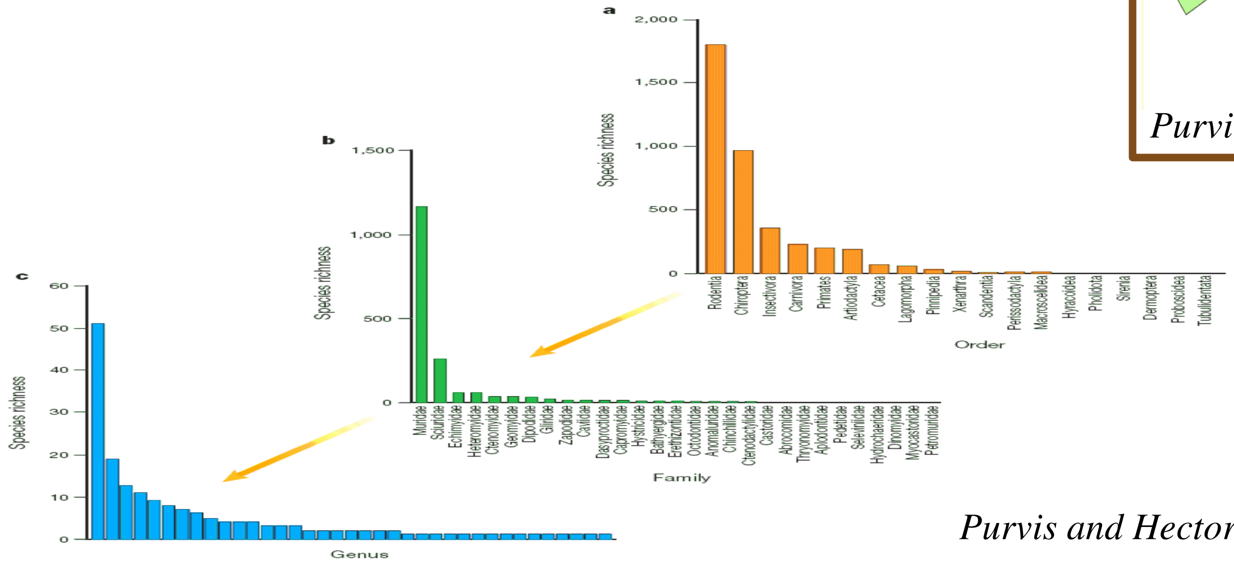
Subtaxa within taxa are often distributed unevenly.

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

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



Purvis and Hector Nature (2000)

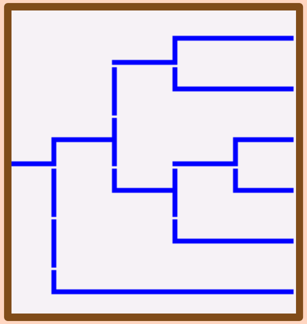


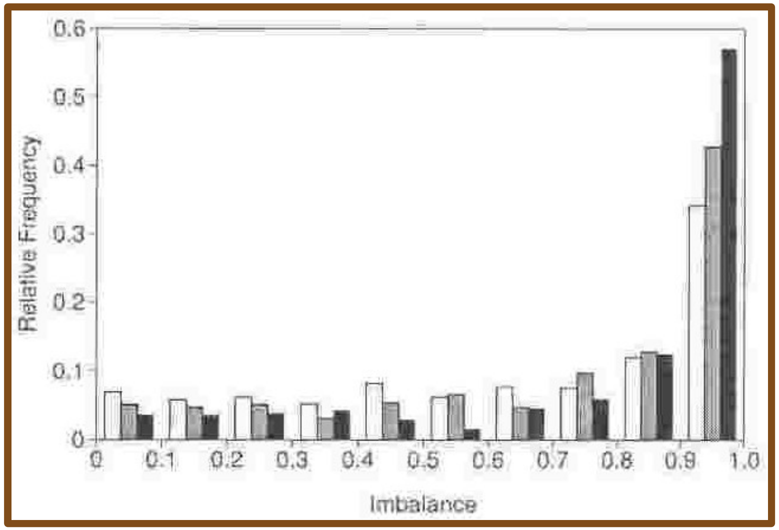
Purvis and Hector Nature (2000)

Shape (Frequency distribution) ...

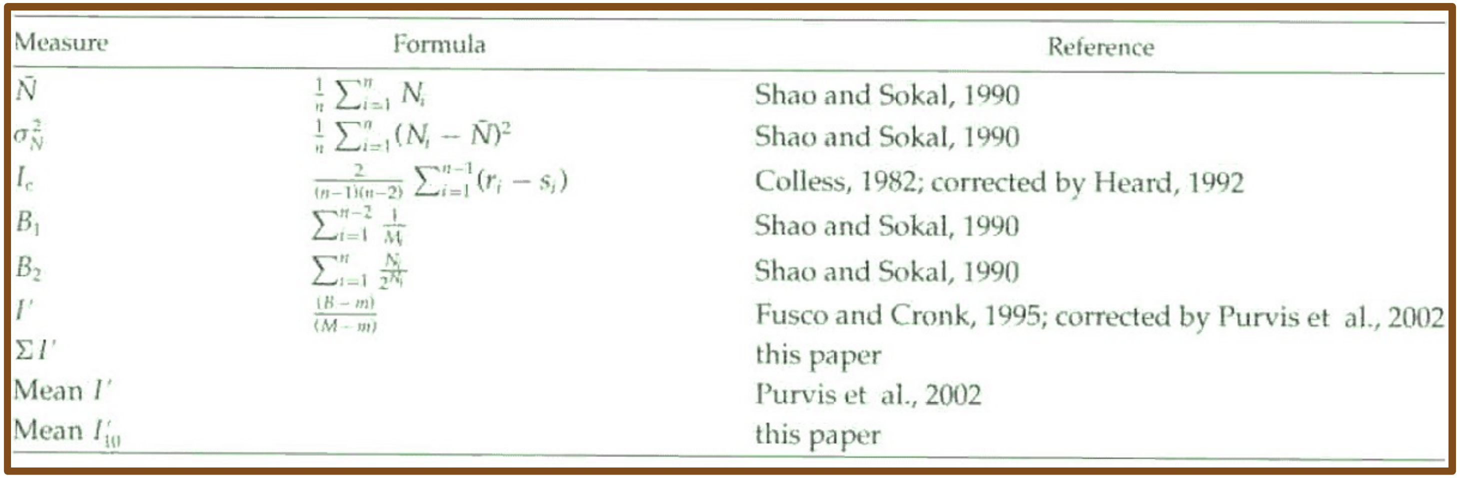
Colless (1982). Colless index of imbalance Ic.

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





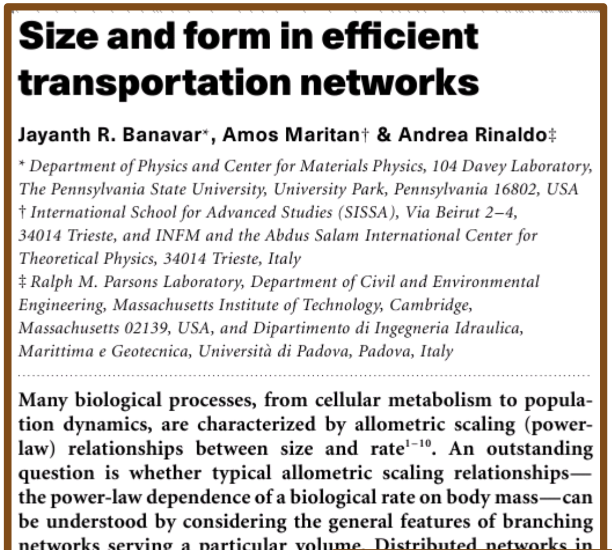
Holman, E. W. *Sys. Biol. (2005)*



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)

From genes to species: Universal Scaling?

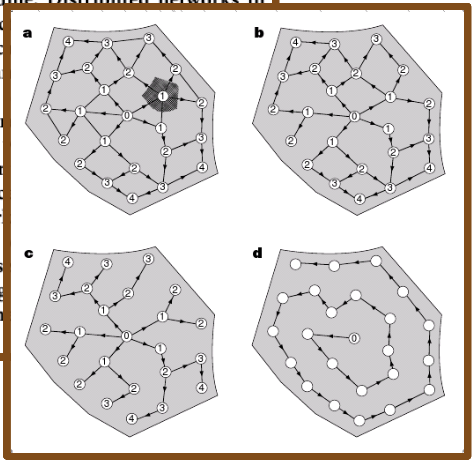


Size and form in efficient transportation networks

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

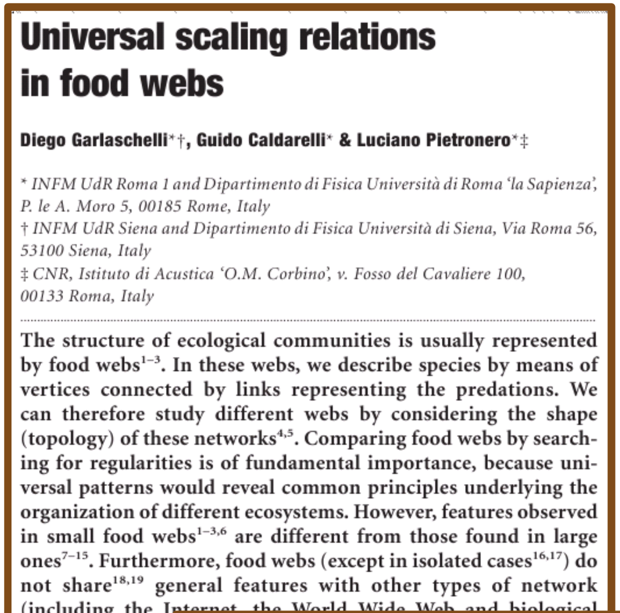
** Department of Physics and Center for Materials Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
† International School for Advanced Studies (SISSA), Via Beirut 2–4, 34014 Trieste, and INFM and the Abdus Salam International Center for Theoretical 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 r
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)

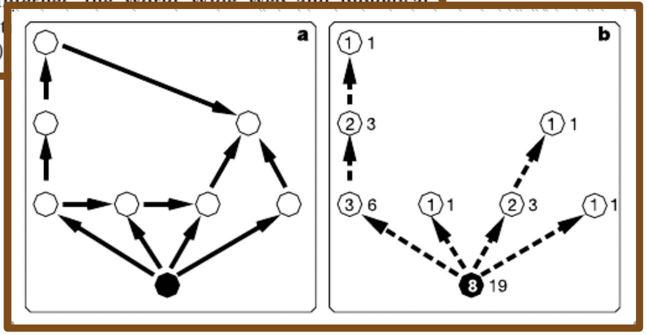


Universal scaling relations in food webs

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)

Allometric Scaling

in Phylogenies...

$$Y \sim M^b$$

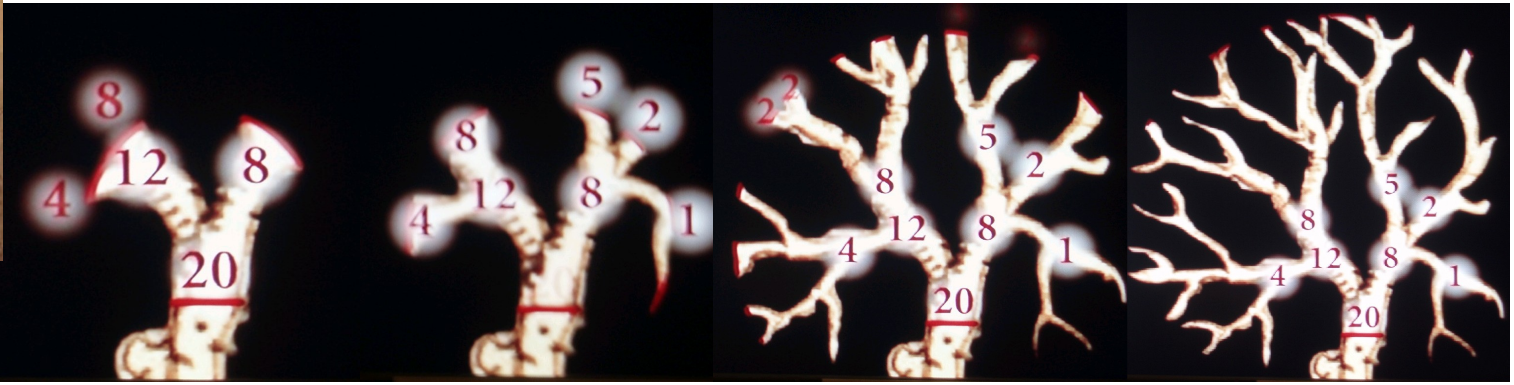
Y: Biological property

M: Mass of the organism

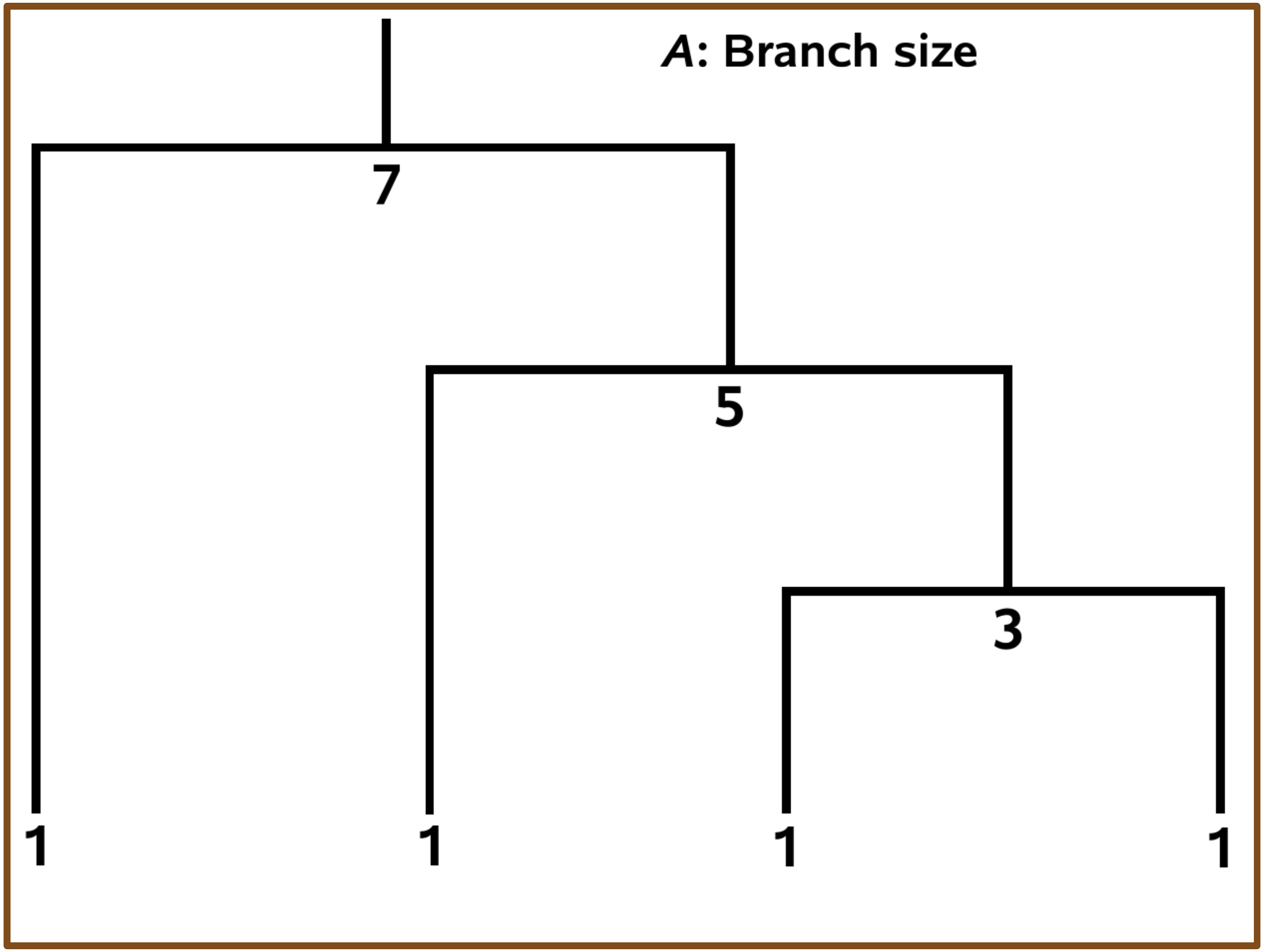
$$C \sim A^\eta$$

C: Phylogenetic tree ***shape***

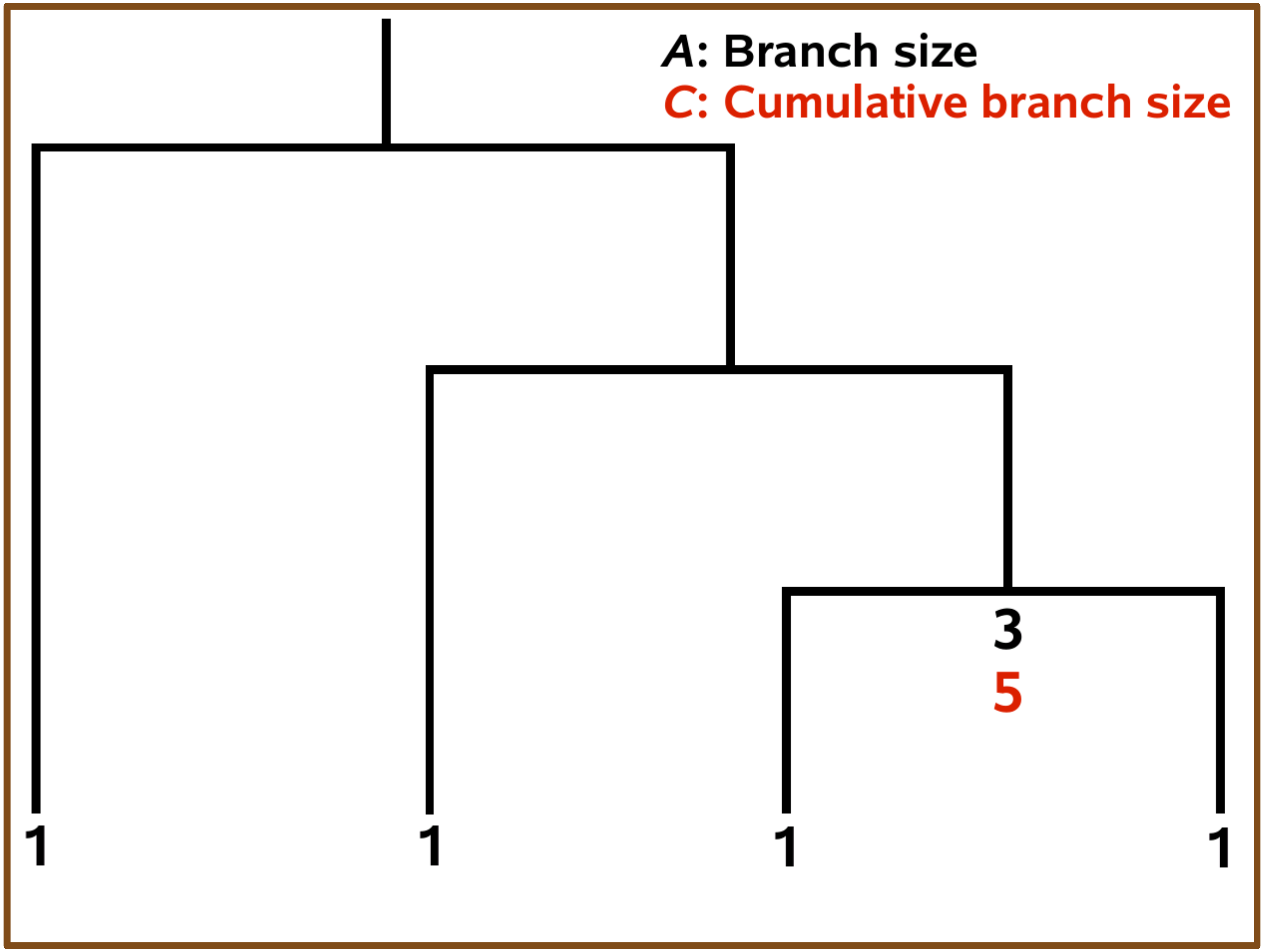
A: Phylogenetic tree ***size***



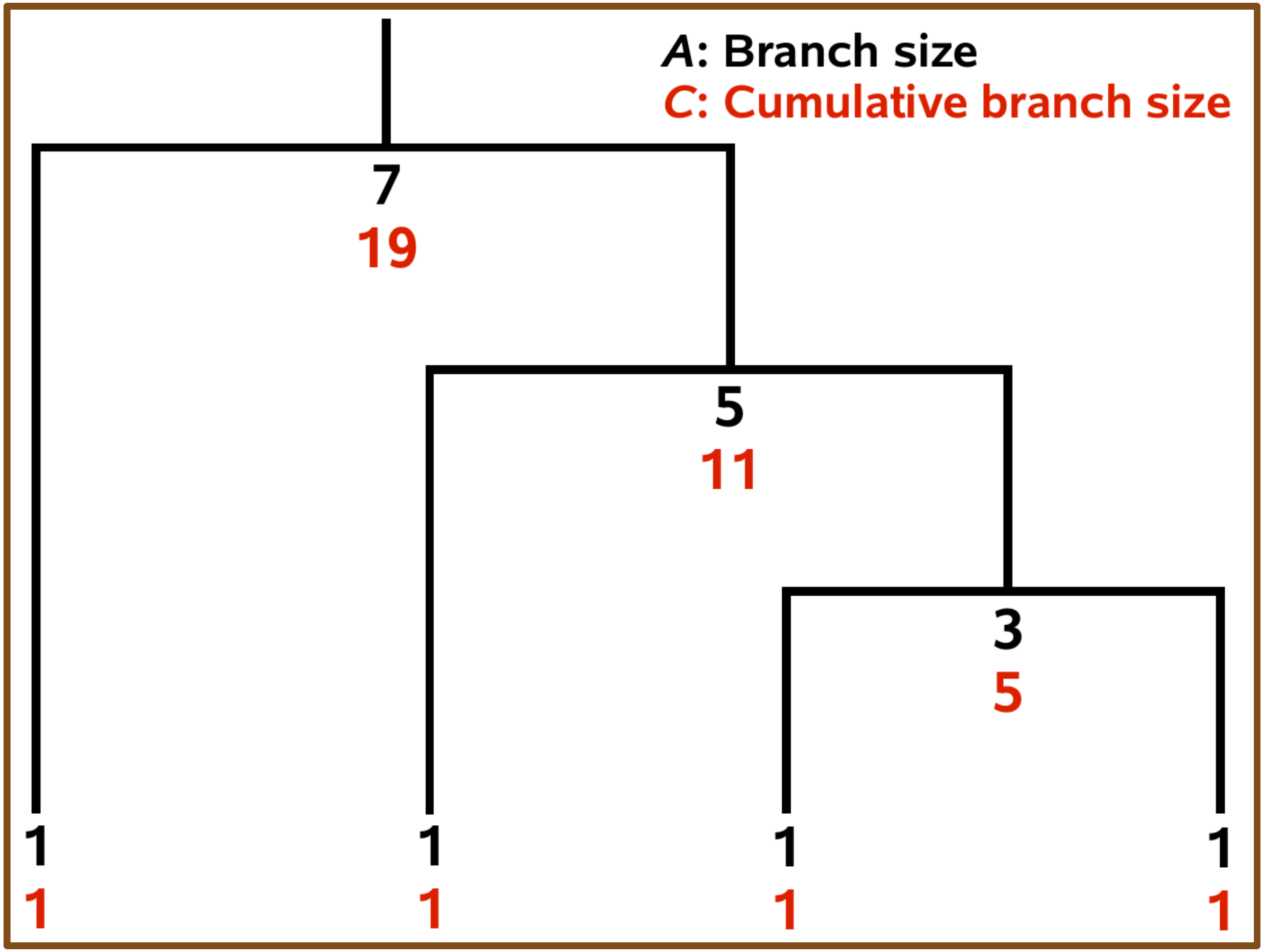
Branch size (A)



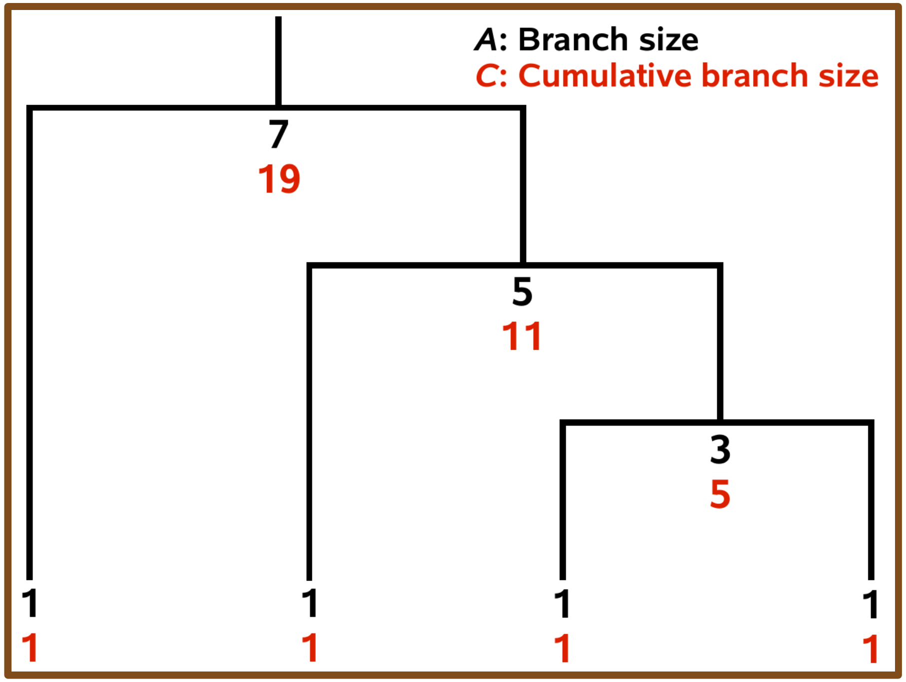
Cumulative branch size (C)

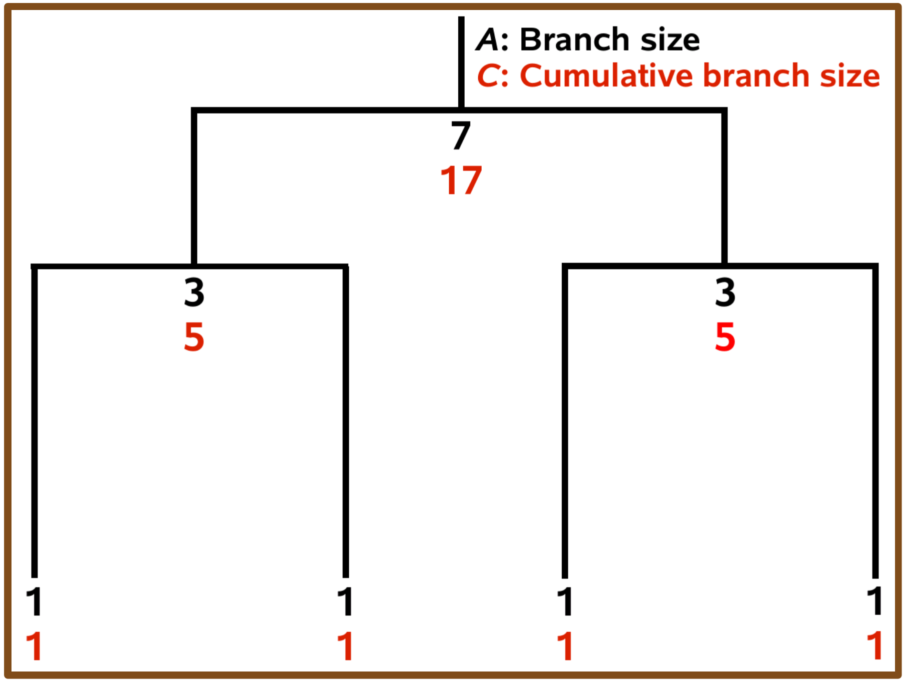


Cumulative branch size (C)

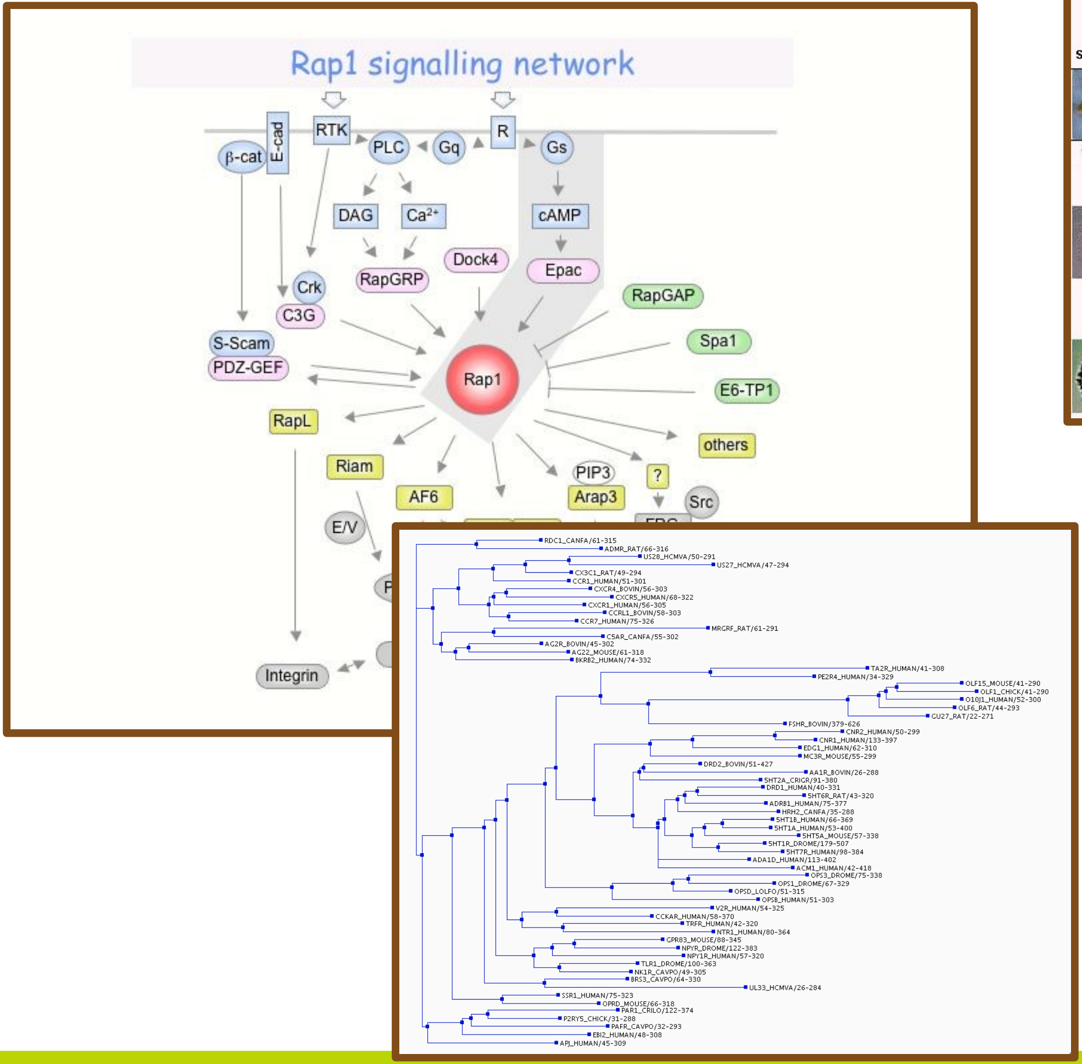


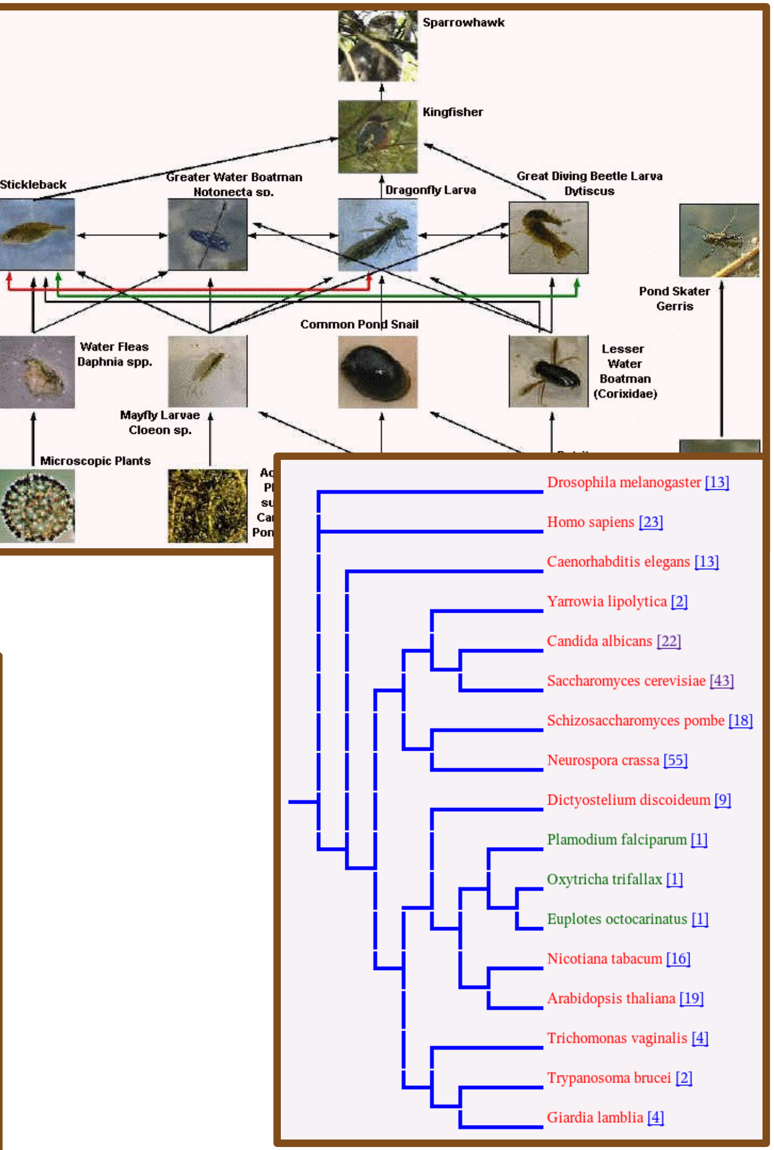
Allometric scaling ($C \approx A^{\eta}$)



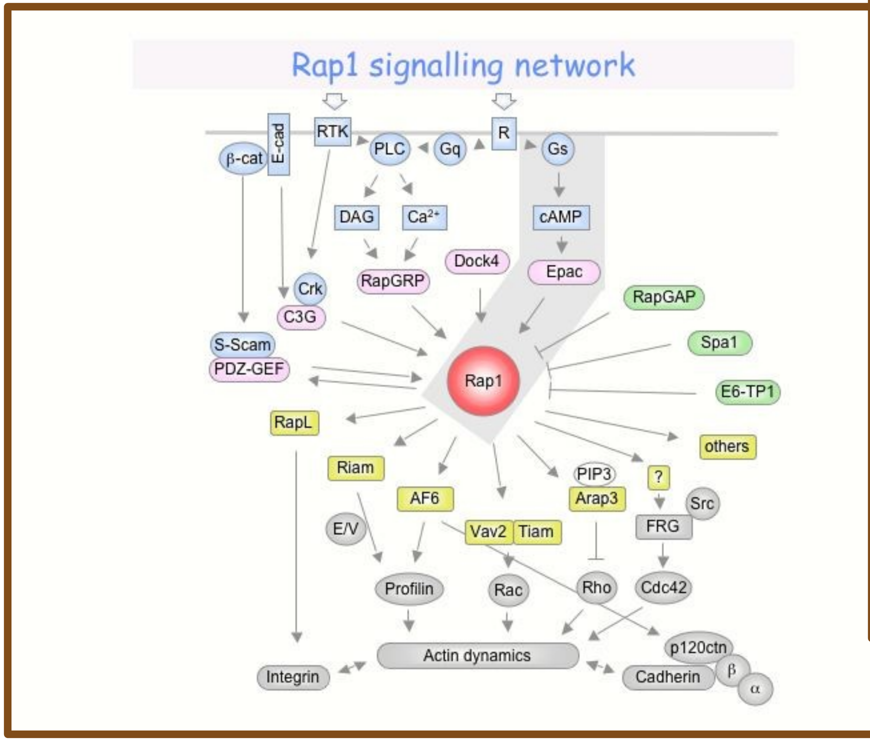


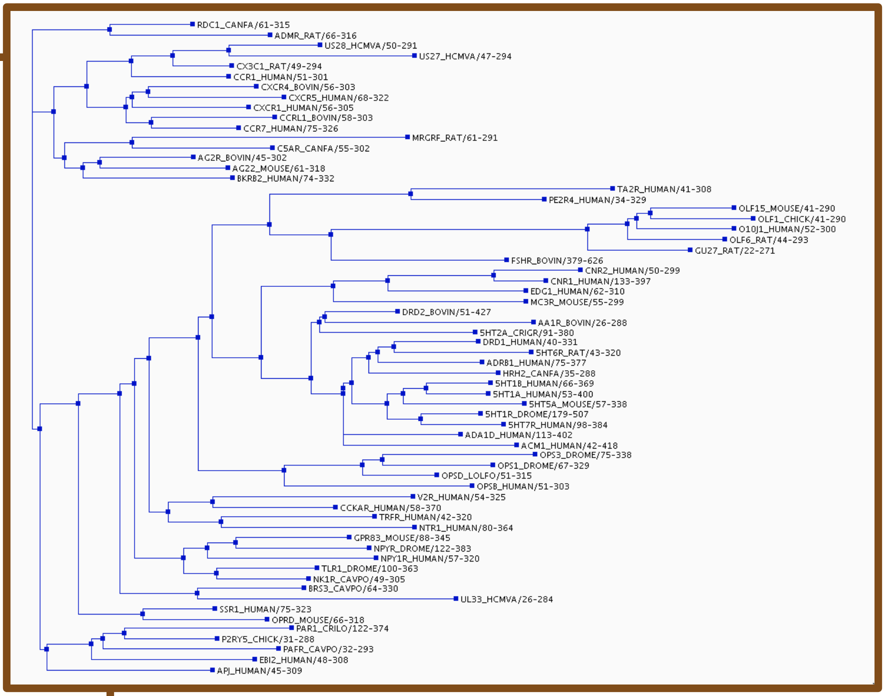
Phylogenies: Proteins and species



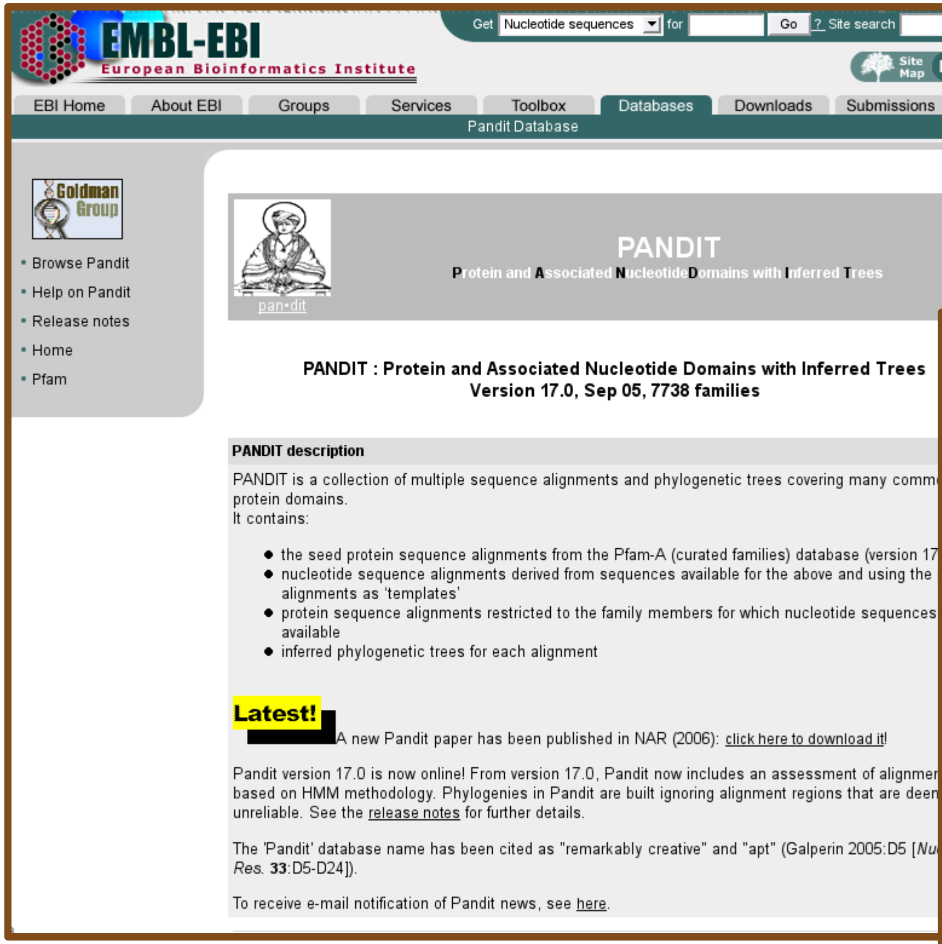


Protein phylogenies





IFISC



PANDIT : Protein and Associated Nucleotide Domains with Inferred Trees
Version 17.0, Sep 05, 7738 families

PANDIT description

PANDIT is a collection of multiple sequence alignments and phylogenetic trees covering many comm
protein domains.
It contains:

- the seed protein sequence alignments from the Pfam-A (curated families) database (version 17
- nucleotide sequence alignments derived from sequences available for the above and using the alignments as 'templates'
- protein sequence alignments restricted to the family members for which nucleotide sequences available
- inferred phylogenetic trees for each alignment

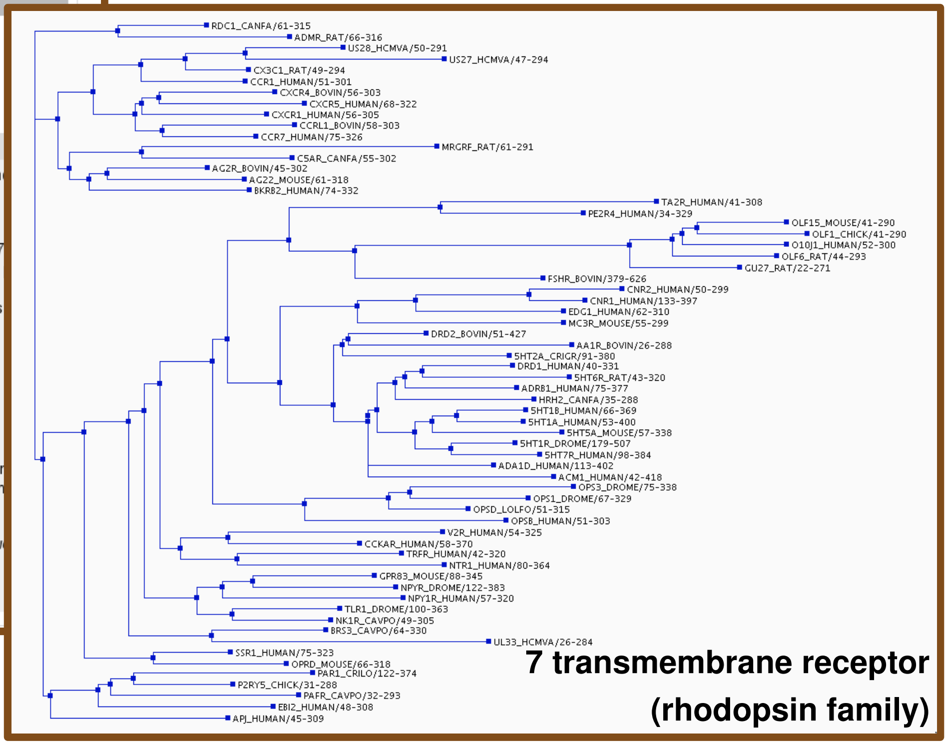
Latest! A new Pandit paper has been published in NAR (2006): click here to download it!

Pandit version 17.0 is now online! From version 17.0, Pandit now includes an assessment of alignmen
based on HMM methodology. Phylogenies in Pandit are built ignoring alignment regions that are deen
unreliable. See the release notes for further details.

The 'Pandit' database name has been cited as "remarkably creative" and "apt" (Galperin 2005:D5 [*Nu*
Res. **33**:D5-D24]).

To receive e-mail notification of Pandit news, see here.

Pandit: 7738 protein families

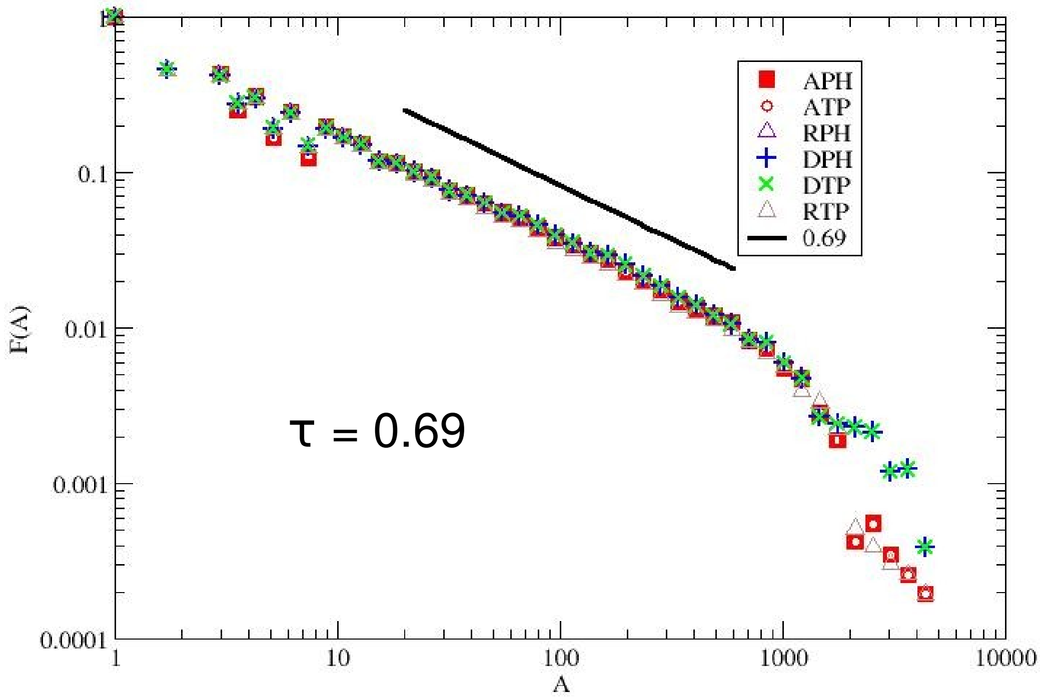


7 transmembrane receptor
(rhodopsin family)

http://www.ebi.ac.uk/goldman-srv/pandit/

Branch Size distribution ($F(A) = A^{-\tau}$)

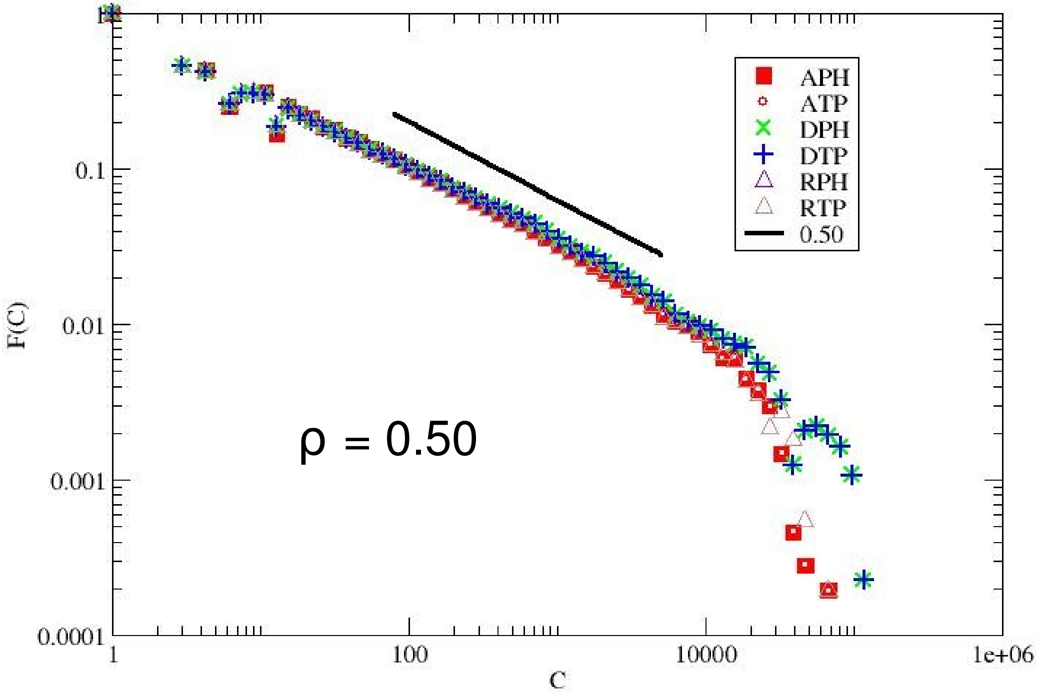
$F(A) = \text{prob}\ (A_i \geq A)$



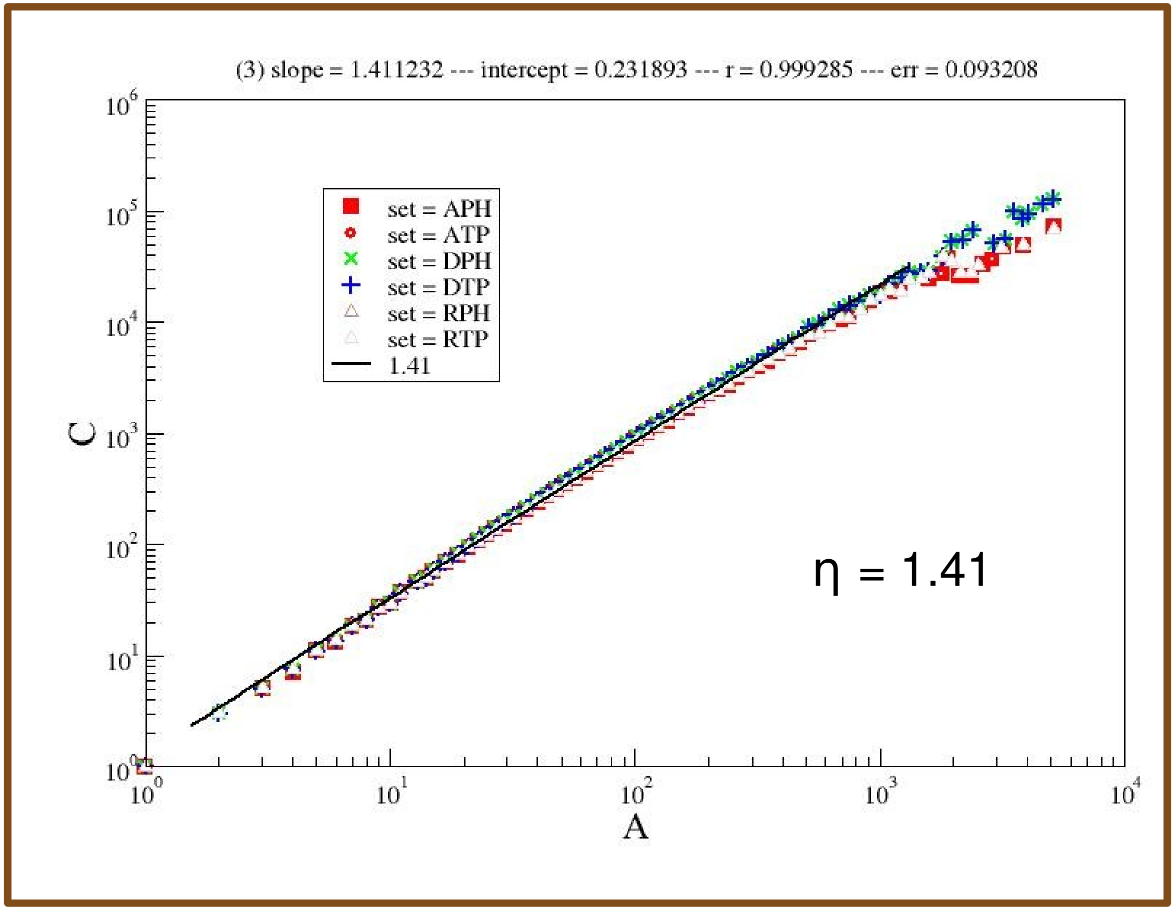
$\tau = 0.69$

Cumulative branch Size distribution ($F(C) = C^{-\rho}$)

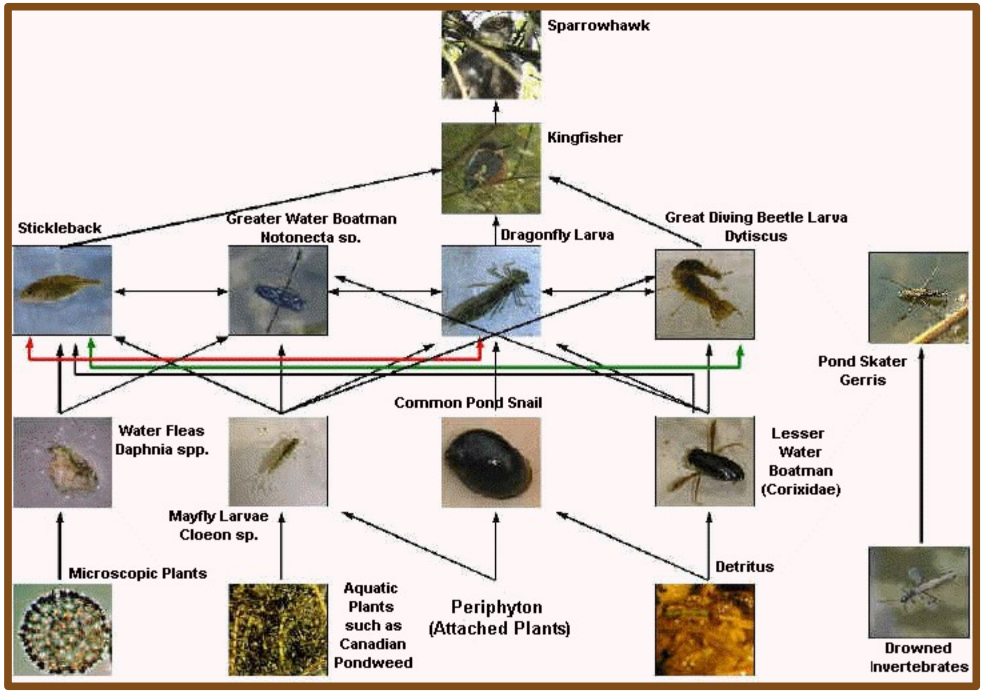
$F(C) = \text{prob}\,(C_i \geq C)$

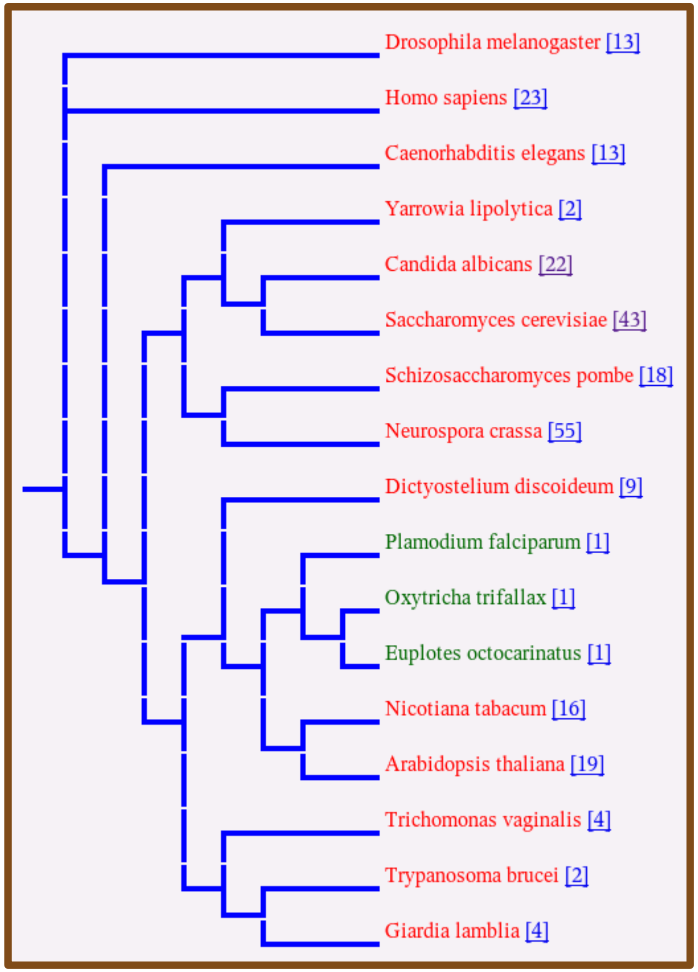


Allometric scaling ($C \approx A^{\eta}$)



Species phylogenies





TreeBASE

A Database of Phylogenetic Knowledge

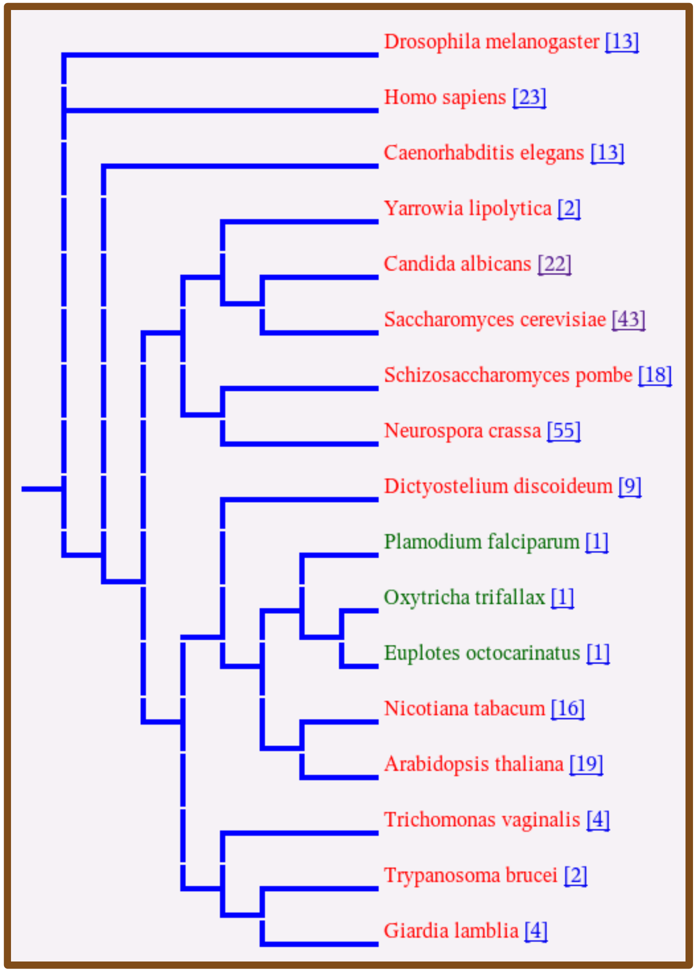
TreeBASE is a relational database of phylogenetic information hosted by the University at Buffalo. In previous years the database has been hosted by Harvard University Herbaria, Leiden University EEW, and the University of California, Davis. TreeBASE stores phylogenetic trees and the data matrices used 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.

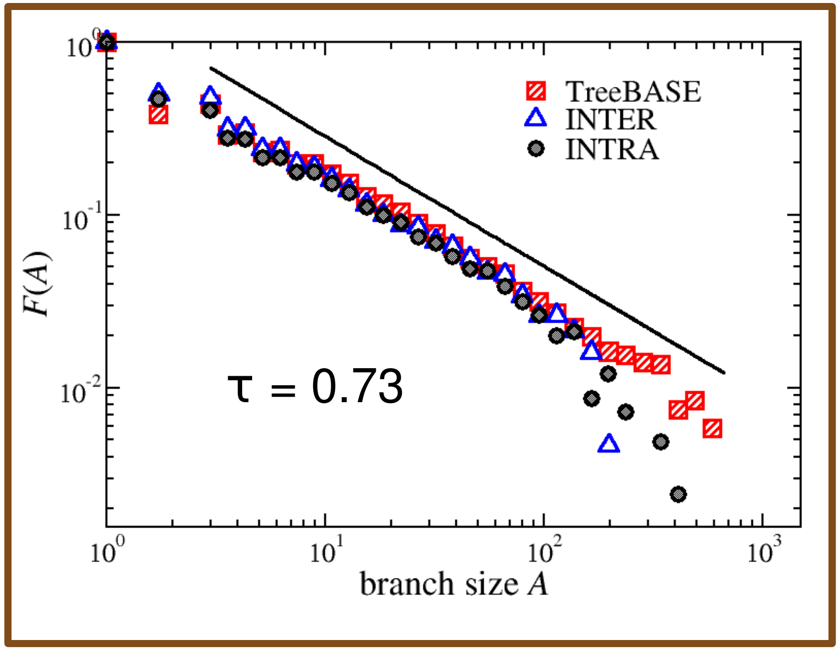
http://www.treebase.org

TreeBASE: 4,020 phylogenetic trees

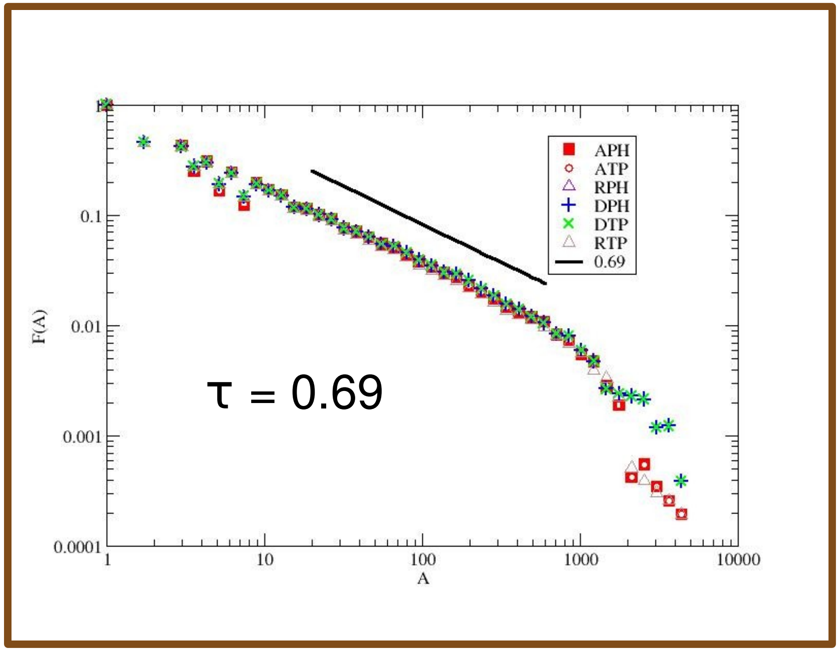


Branch Size distribution ($F(A) = A^{-\tau}$)

$F(A) = \text{prob}\ (A_i \geq A)$



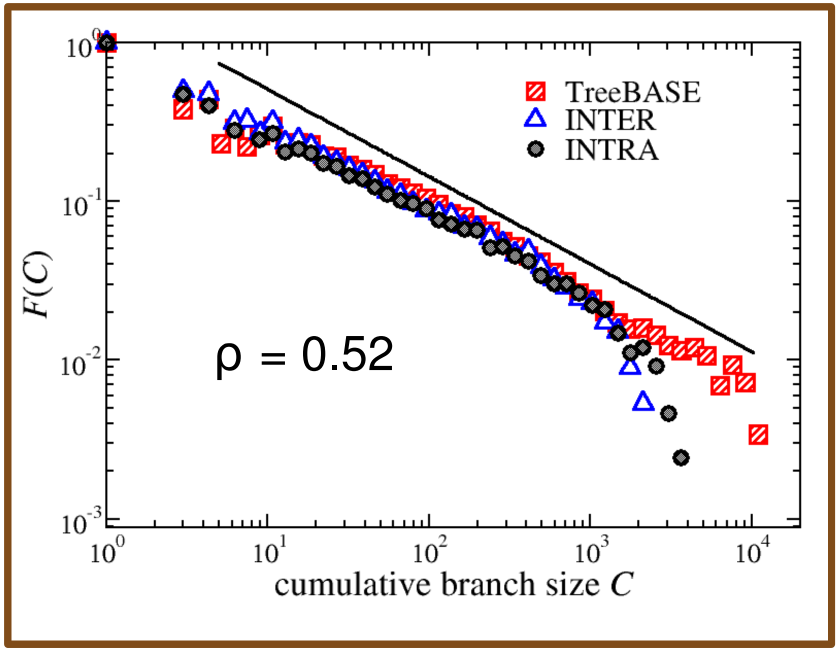
$\tau = 0.73$

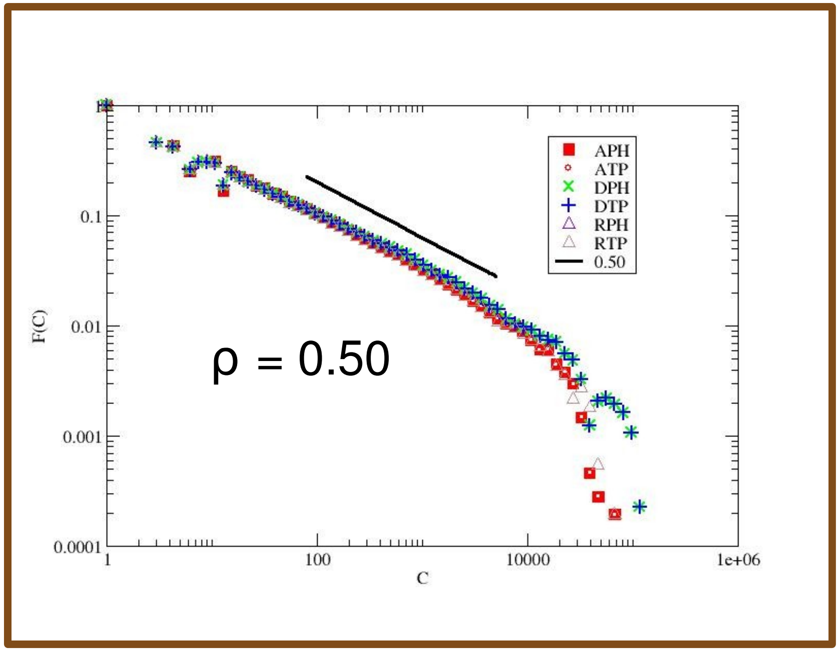


$\tau = 0.69$

Cumulative branch Size distribution ($F(C) = C^{-\rho}$)

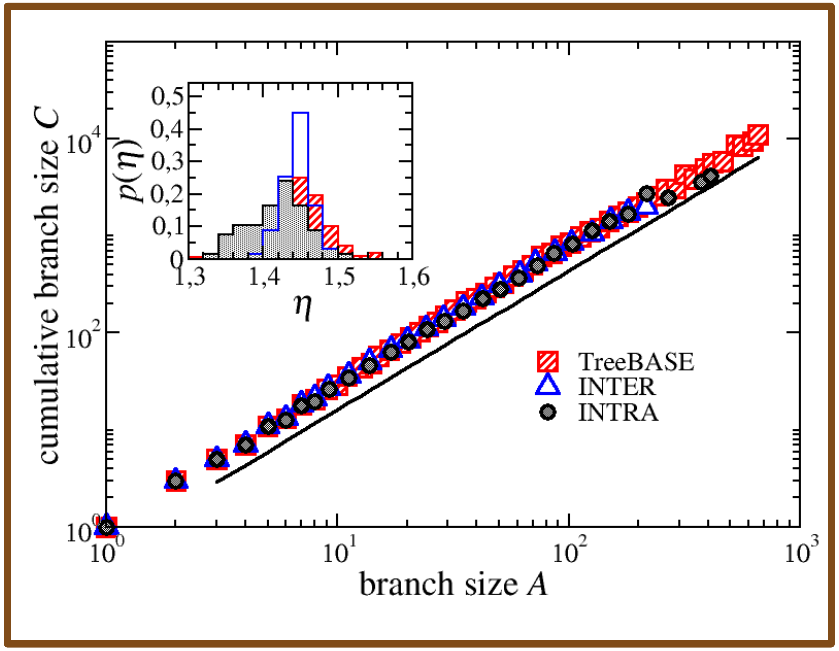
$F(C) = \text{prob}\,(C_i \geq C)$



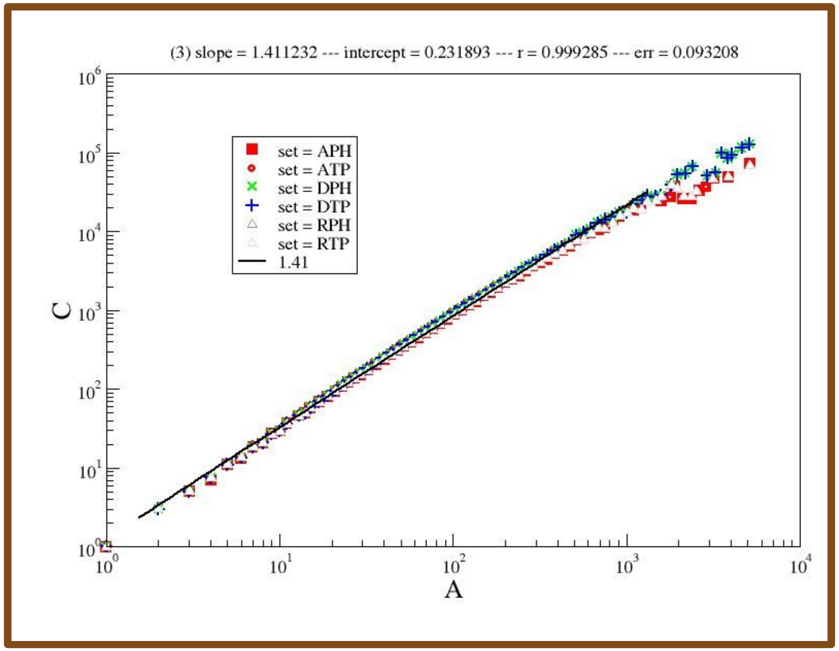


Allometric scaling ($C \approx A^{\eta}$)

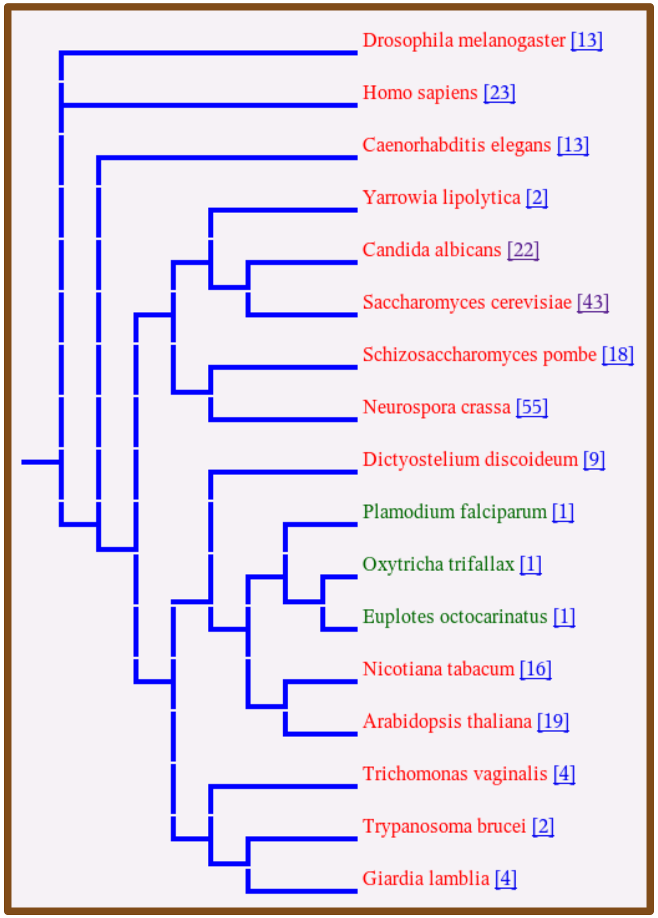
$\eta = 1.44$



$\eta = 1.41$



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

Catalogue of Life

Catalogue of Life: 2007 Annual Checklist

indexing the world's known species

Search the Annual Checklist - fixed edition each year

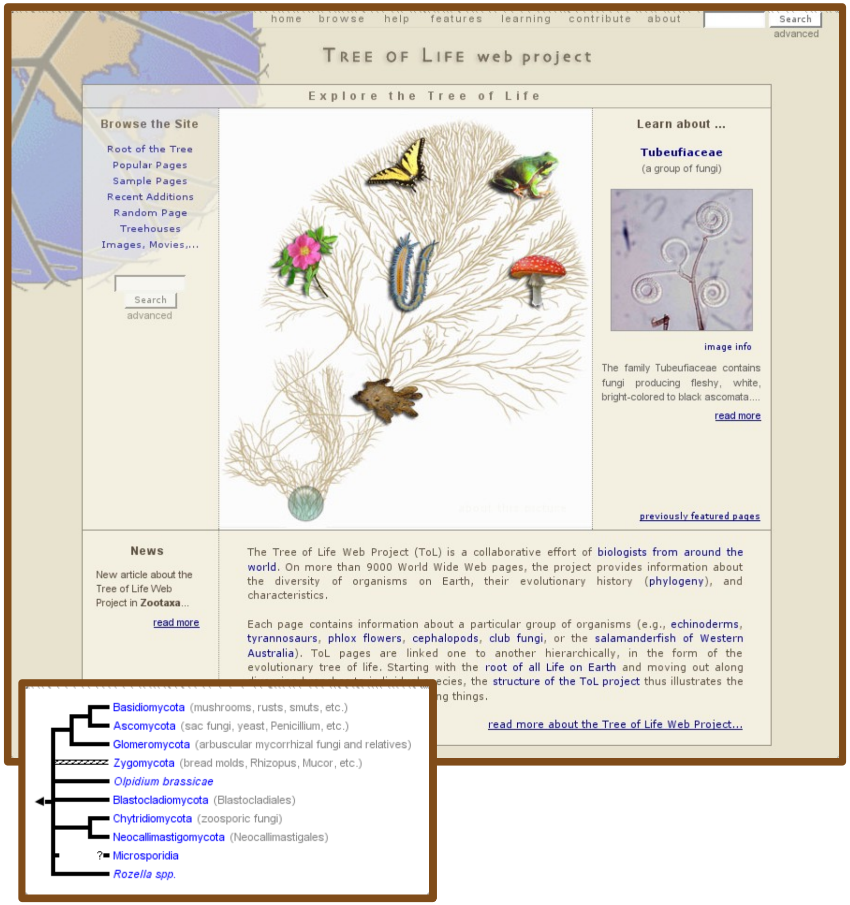
Search for:

☑ 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

Root of the Tree
Popular Pages
Sample Pages
Recent Additions
Random Page
Treehouses
Images, Movies,...

Learn about ...

Tubeufiaceae
(a group of fungi)

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

read more

previously featured pages

News

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

read more

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.

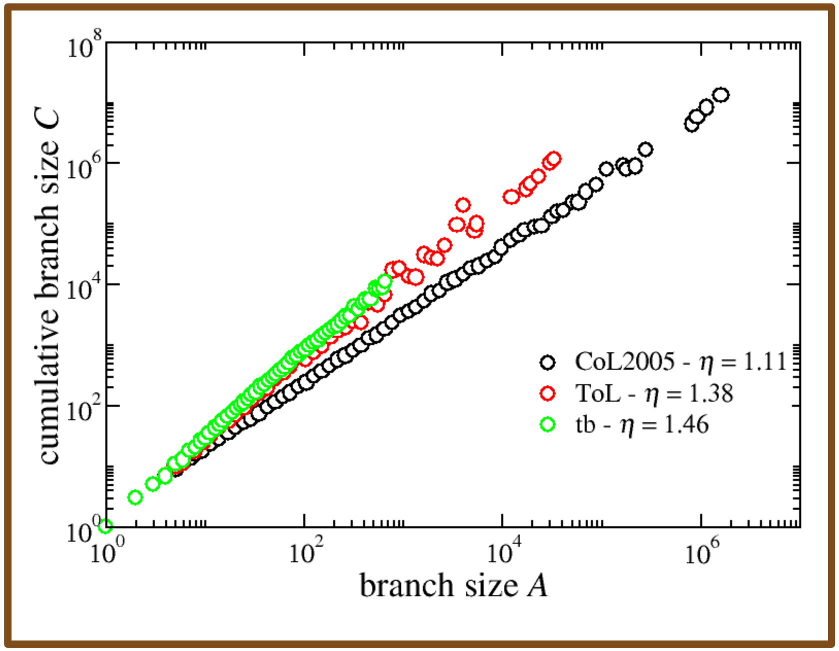
Each page contains information about a particular group of organisms (e.g., echinoderms, tyrannosaurs, phlox flowers, cephalopods, club fungi, or the salamanderfish of Western Australia). ToL pages are linked one to another hierarchically, in the form of the evolutionary tree of life. Starting with the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

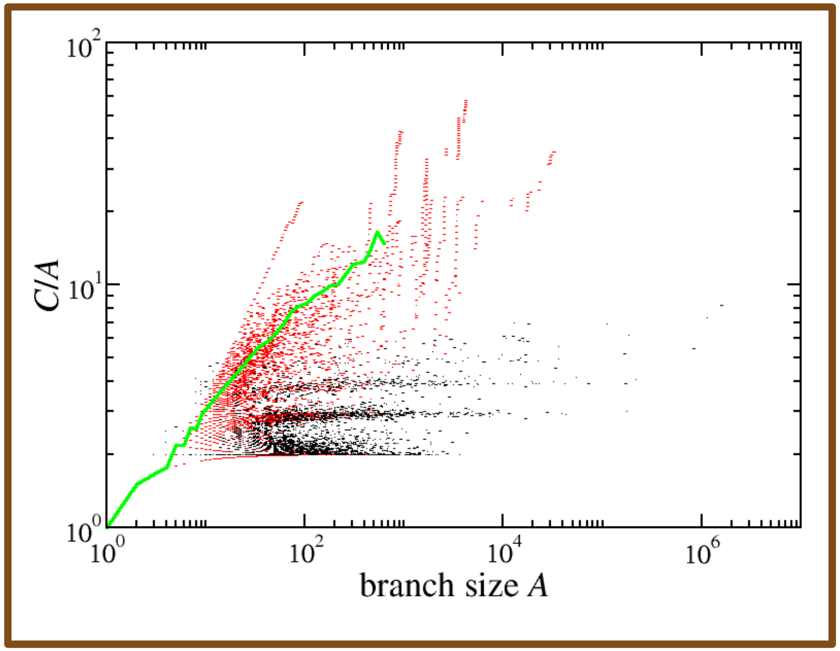
read more about the Tree of Life Web Project...

Basidiomycota (mushrooms, rusts, smuts, etc.)
Ascomycota (sac fungi, yeast, Penicillium, etc.)
Glomeromycota (arbuscular mycorrhizal fungi and relatives)
Zygomycota (bread molds, Rhizopus, Mucor, etc.)
Olpidium brassicae
Blastocladiomycota (Blastocladiales)
Chytridiomycota (zoosporic fungi)
Neocallimastigomycota (Neocallimastigales)
? Microsporidia
Rozella spp.

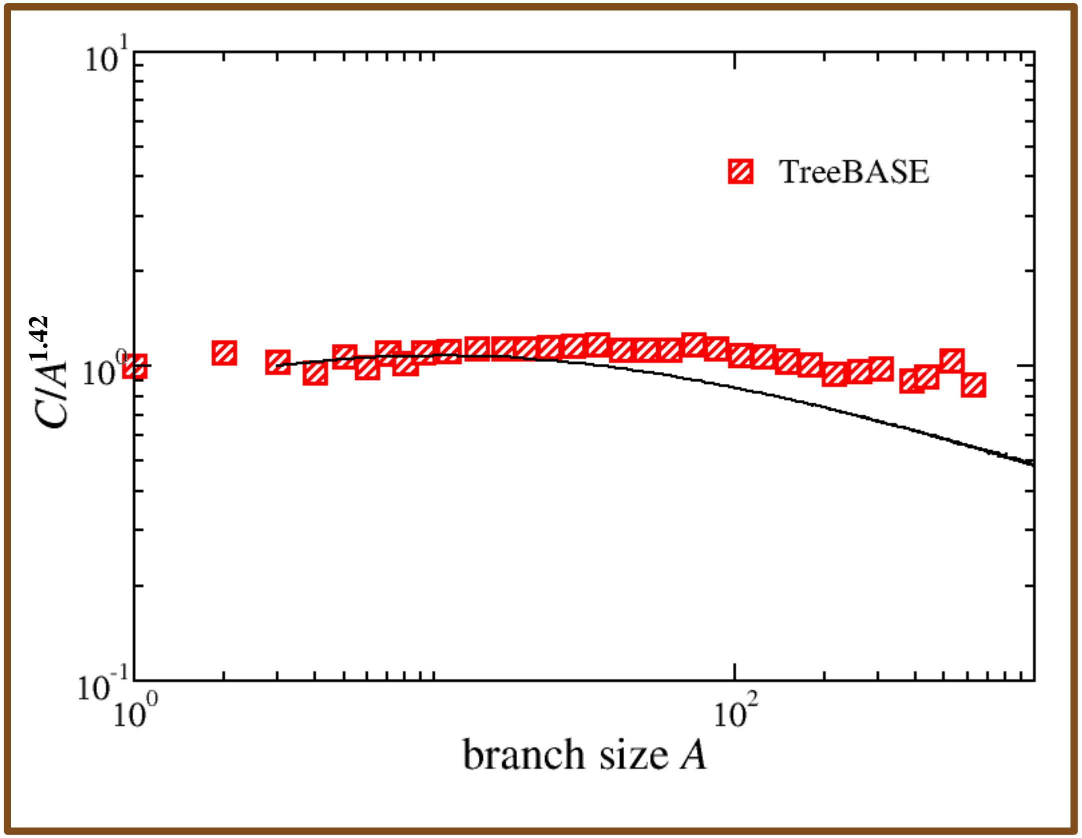
http://www.tolweb.org

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





Equal-rates Markov (ERM) model



Conclusions...

- ➔ We have found a **power-law distribution** for the branch size (0.70-0.75), for the cumulative branch size (0.52-056), and for the C vs A scaling (1.41-1.44).

- ➔ We have found **striking similarities** between the power-law exponents for the protein families and for the species phylogenies, both in the branch size (A) and the cumulative branch size (C) distribution and in the C vs A scaling.

- ➔ The **ERM model does not fully explain** the C vs A scaling. The phylogenies display correlations not contained in the ERM model.

- ➔ The inclusion of a **memory-dependent process** in the null model could help us to explain the dynamics of the evolutionary events.