I·M·E·D·E·A
Institut Mediterrani d'Estudis Avançats

Scaling properties in the Tree of Life

E. Alejandro Herrada*,

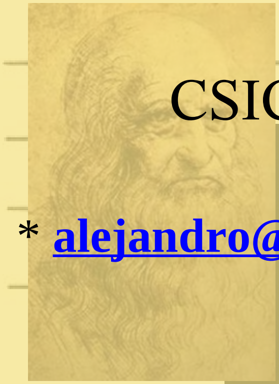
Claudio J. Tessone, Víctor M. Eguíluz, Emilio Hernández-García

and Carlos M. Duarte

Instituto Mediterráneo de Estudios Avanzados (IMEDEA)

www.imedea.uib.es/physdept/

CSIC-Universitat de les Illes Balears, Palma de Mallorca, Spain

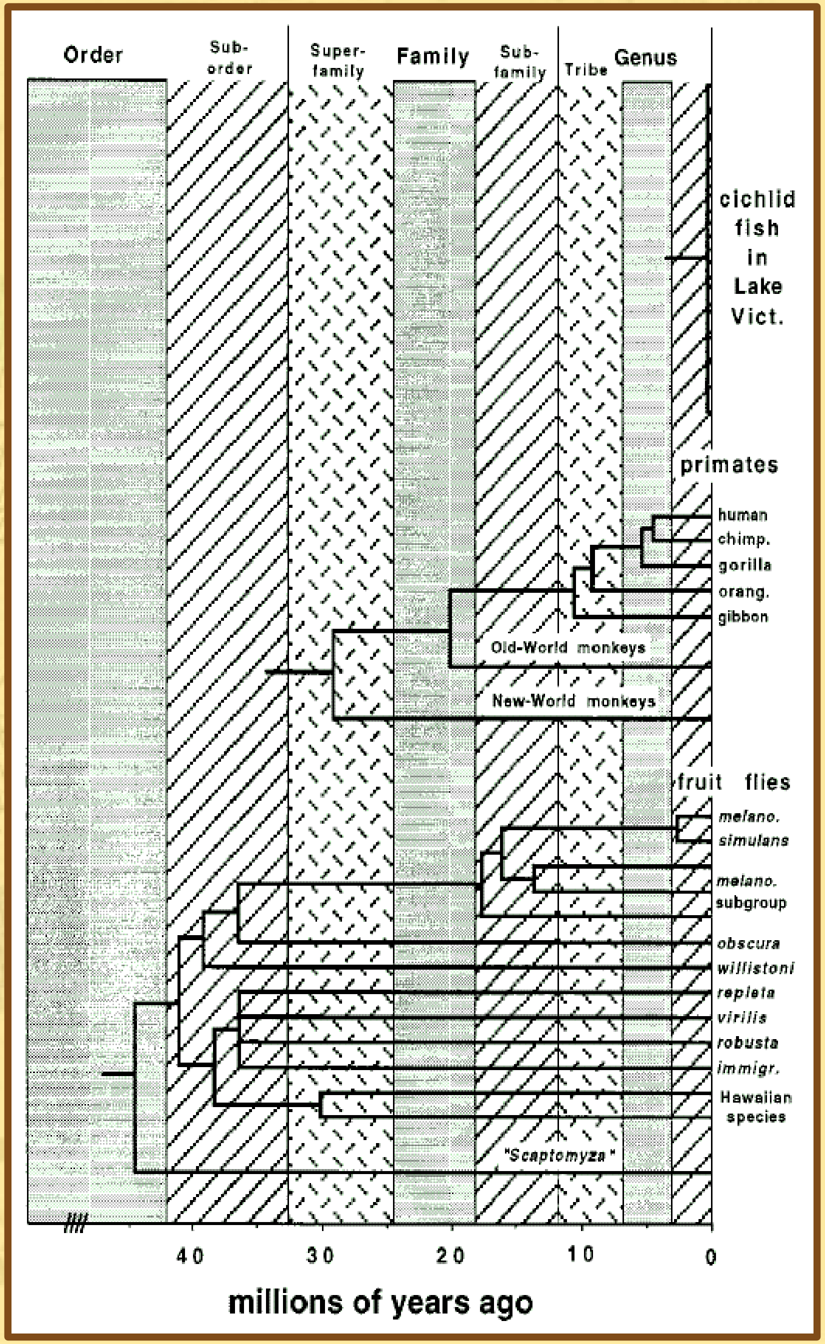


* **alejandro@imedea.uib.es**

Palma de Mallorca (Illes Balears), Janury 2007

Inferring Evolutionary process from phylogenetic tree shape...

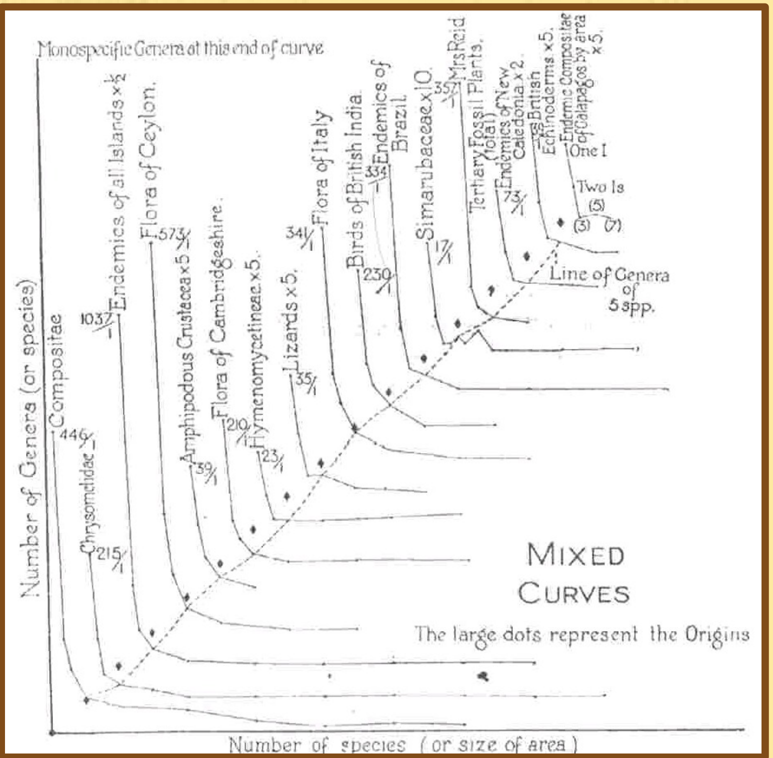
Purvis, A. and Hector, A.
Nature (2000)

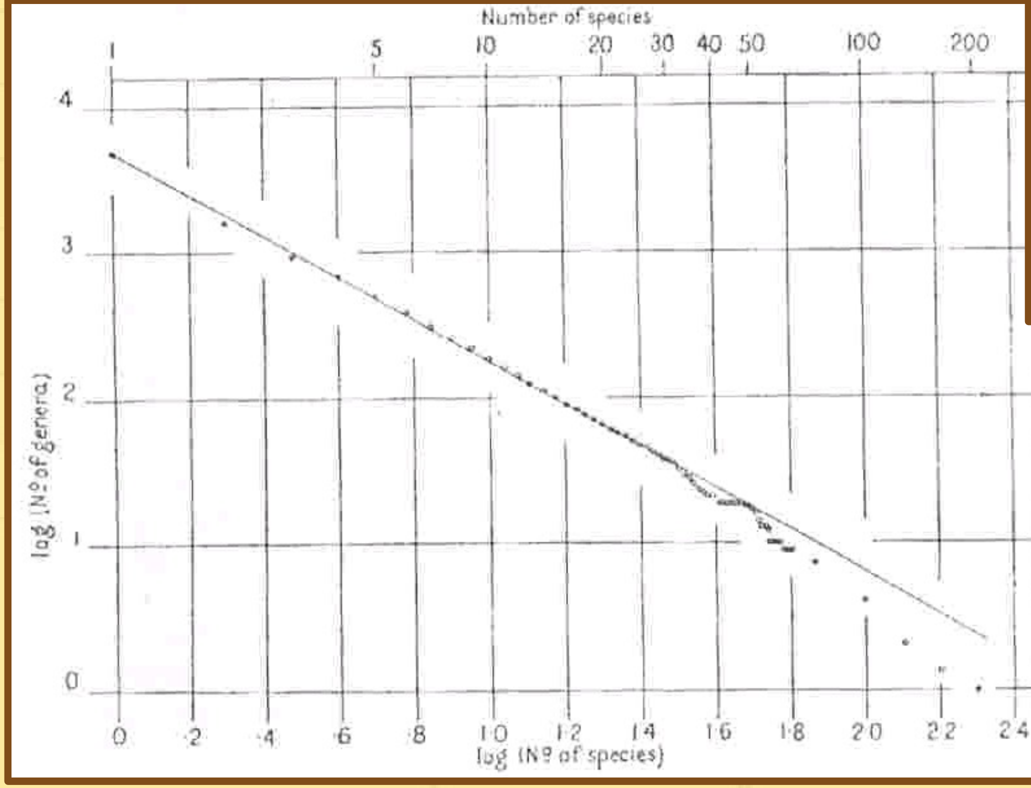


Size (Frequency distribution) ...

Numbers of species in biological taxa

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





Introduction. *Historical context and concepts*

Darwin **(1859)**

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

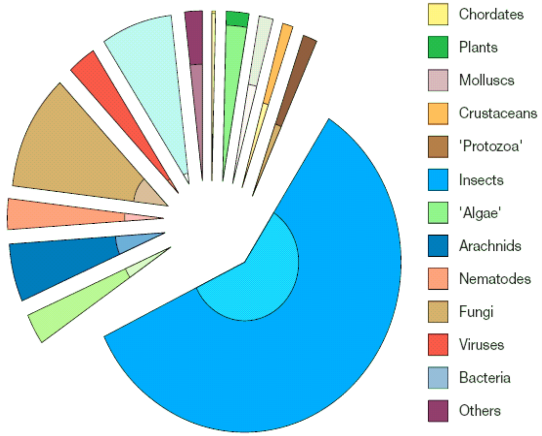
Willis **(1922),** ***Corbet*** **(1942),** ***Anderson*** **(1975).**

Subtaxa within taxa are often distributed unevenly.

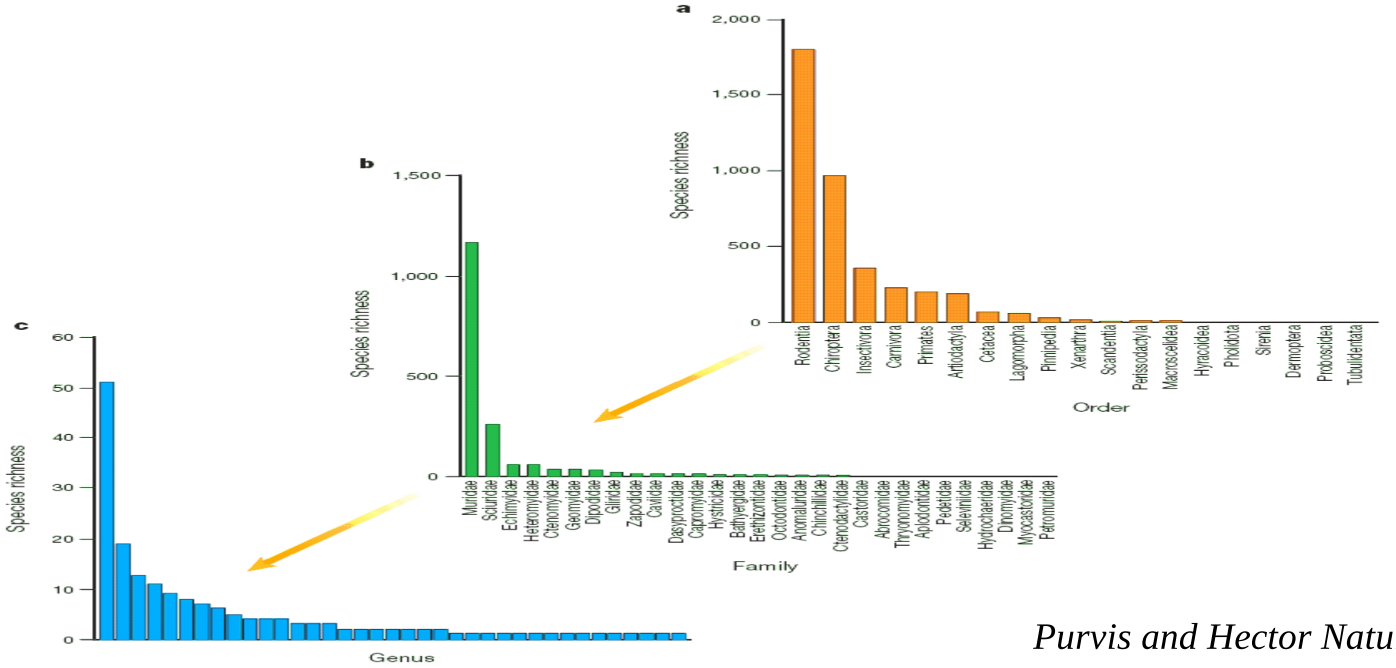
Burlando **(1990, 1993),** ***Minelli*** **(1991),** ***Nee*** **(1992).**

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

Scaling pattern in intraspecific diversity?



Purvis and Hector Nature (2000)

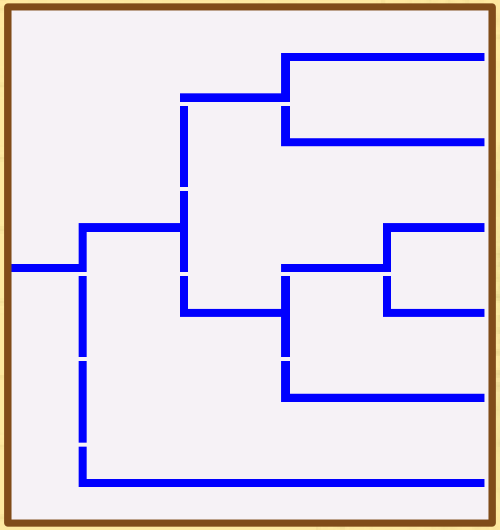


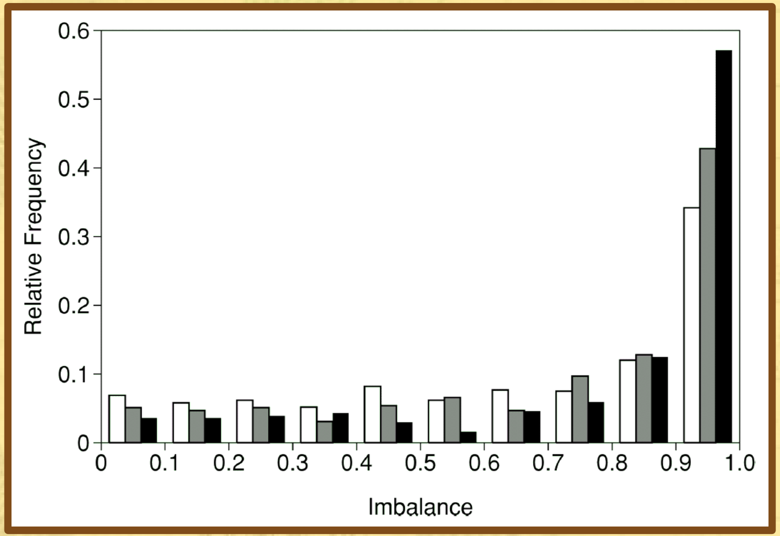
Purvis and Hector Nature (2000)

Shape **(Frequency distribution) ...**

Colless (1982). Colless index of imbalance Ic.

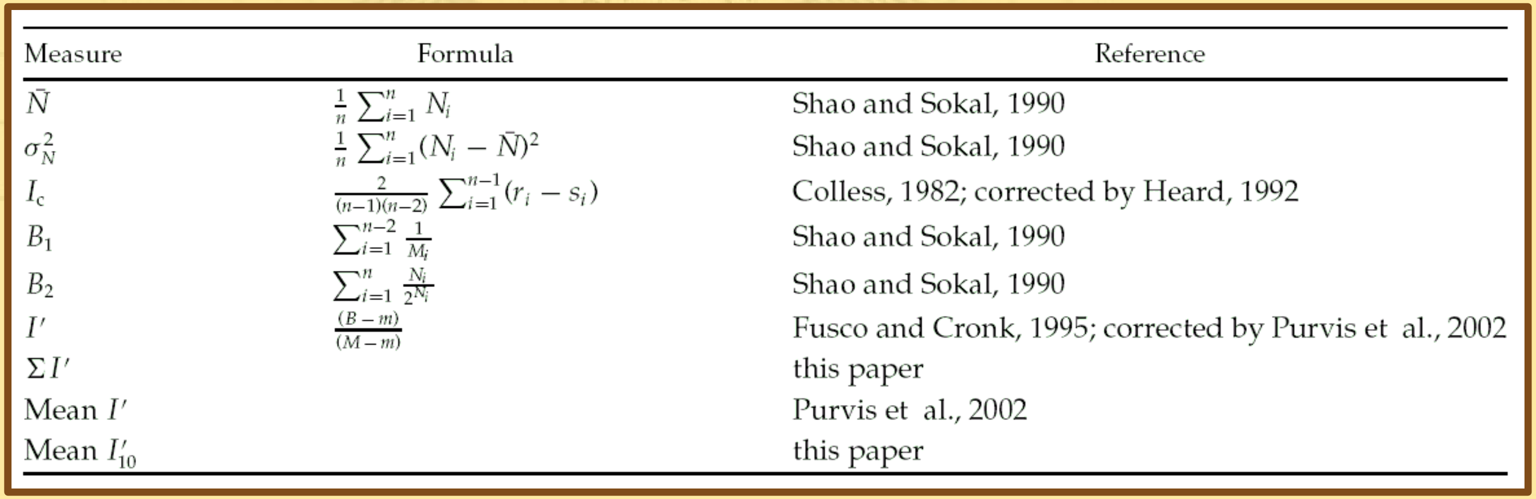
Savage (1983). Patterns in the frequencies of all possible tree topologies.





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

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



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

How changes the phylogenetic tree *shape* with its *size*...

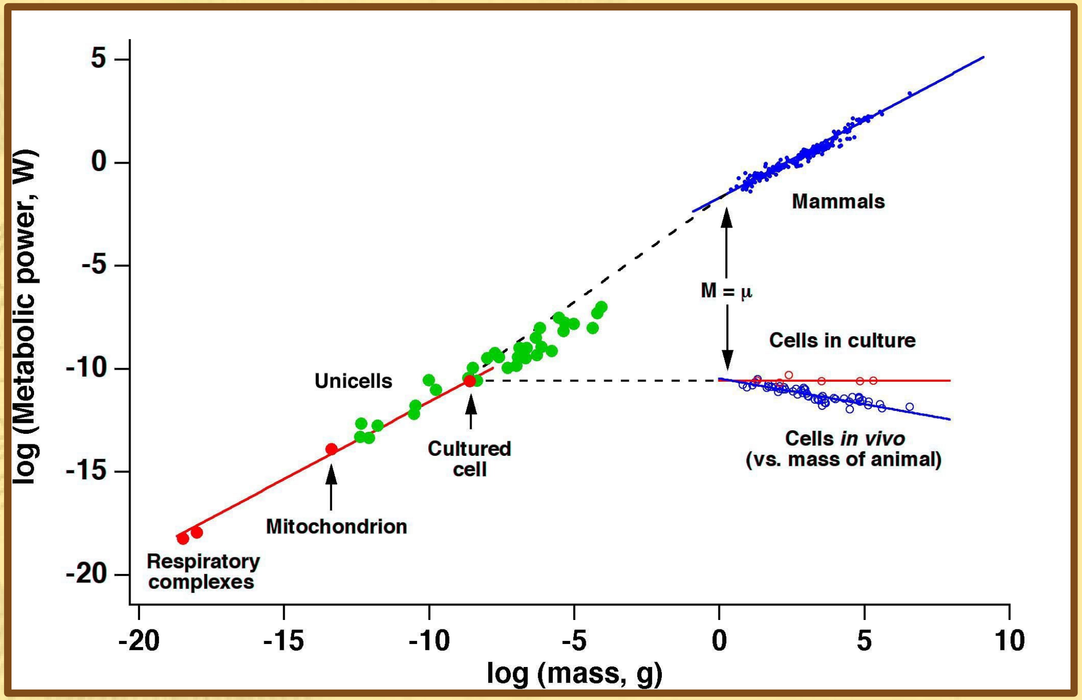
Allometric scaling

$$Y \sim M^b$$

Y: Biological property
M: Mass of the organism

Examples:

- *Metabolic rate (¾)*
- *Lifespan (¼)*
- *Densities of mitochondria, chloroplasts and ribosomes (¼)*



West, G. B. and Brown, J. H. *J. Exp. Biol. (2005)*

Allometric Scaling

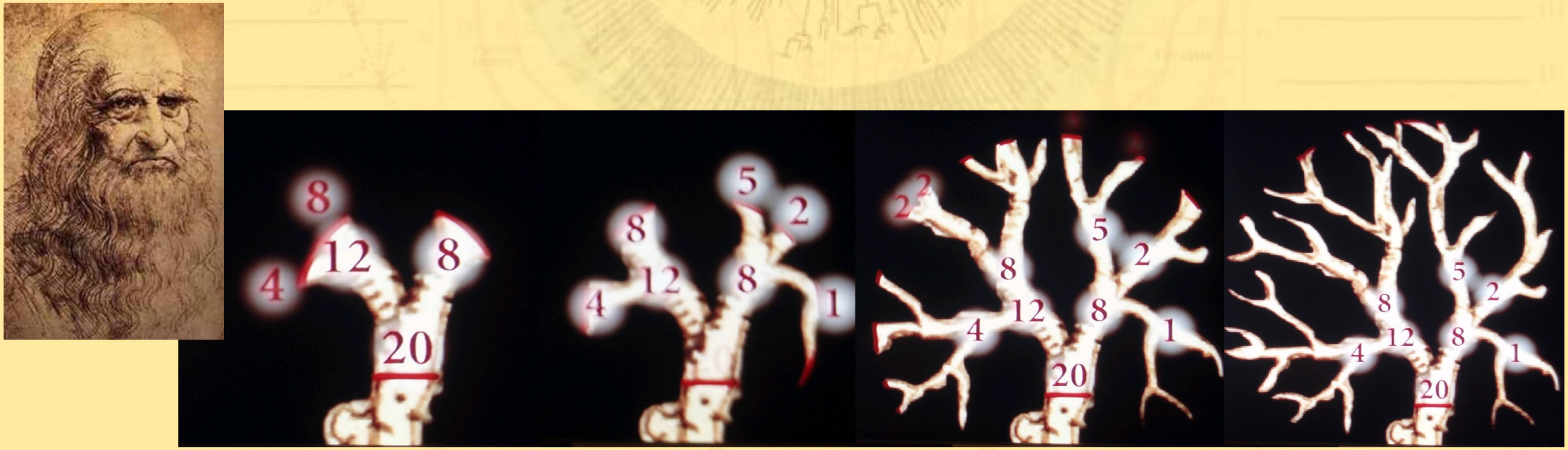
in Phylogenetic Trees...

$$Y \sim M^b$$

Y: Biological property
M: Mass of the organism

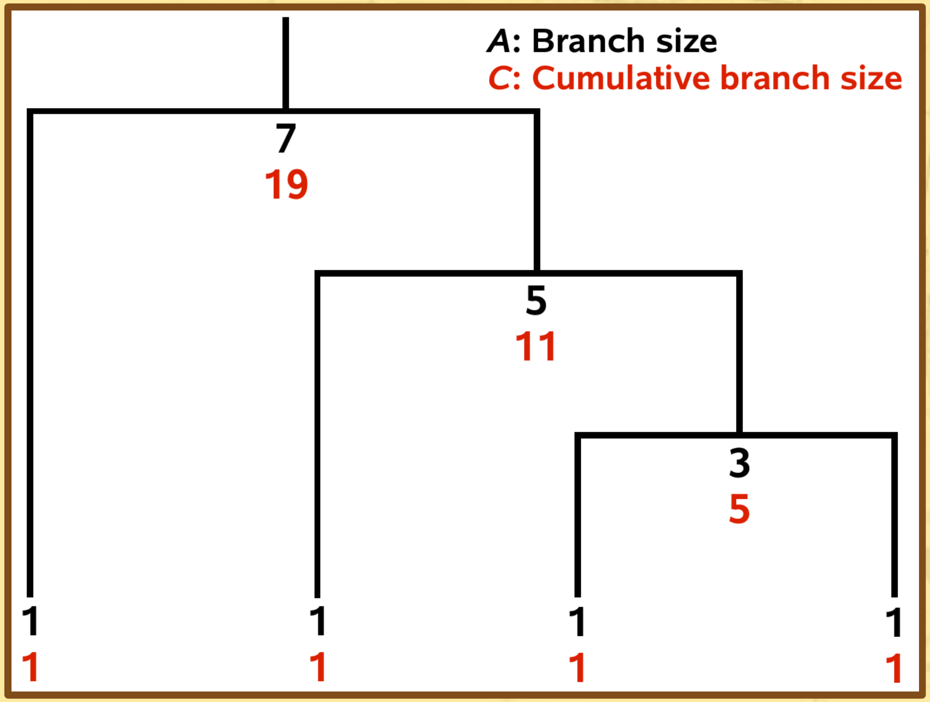
$$C \sim A^\eta$$

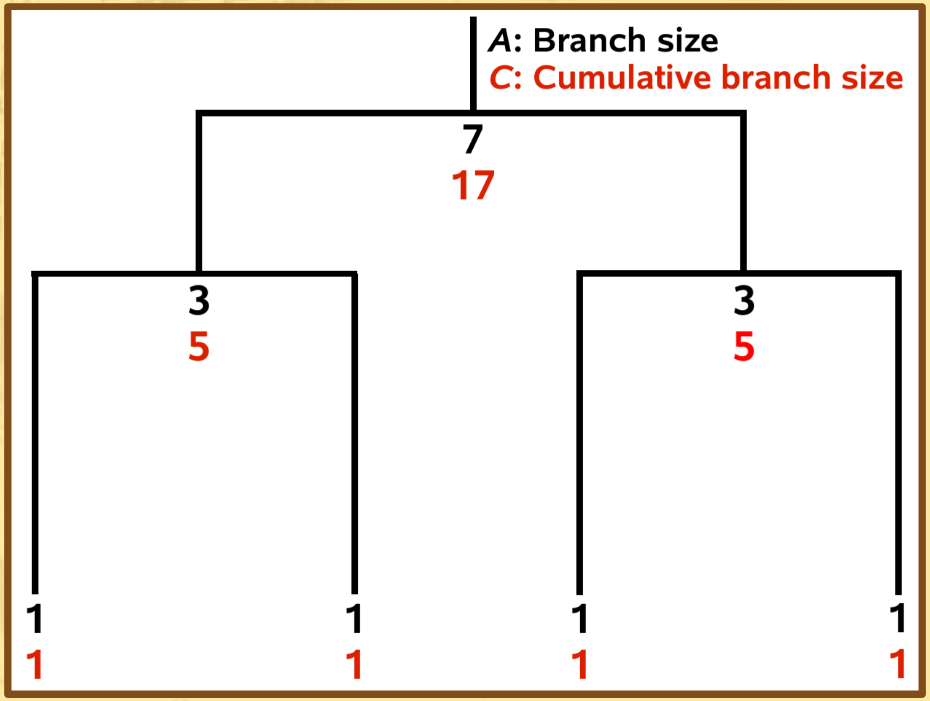
C: Phylogenetic tree ***shape***
A: Phylogenetic tree ***size***

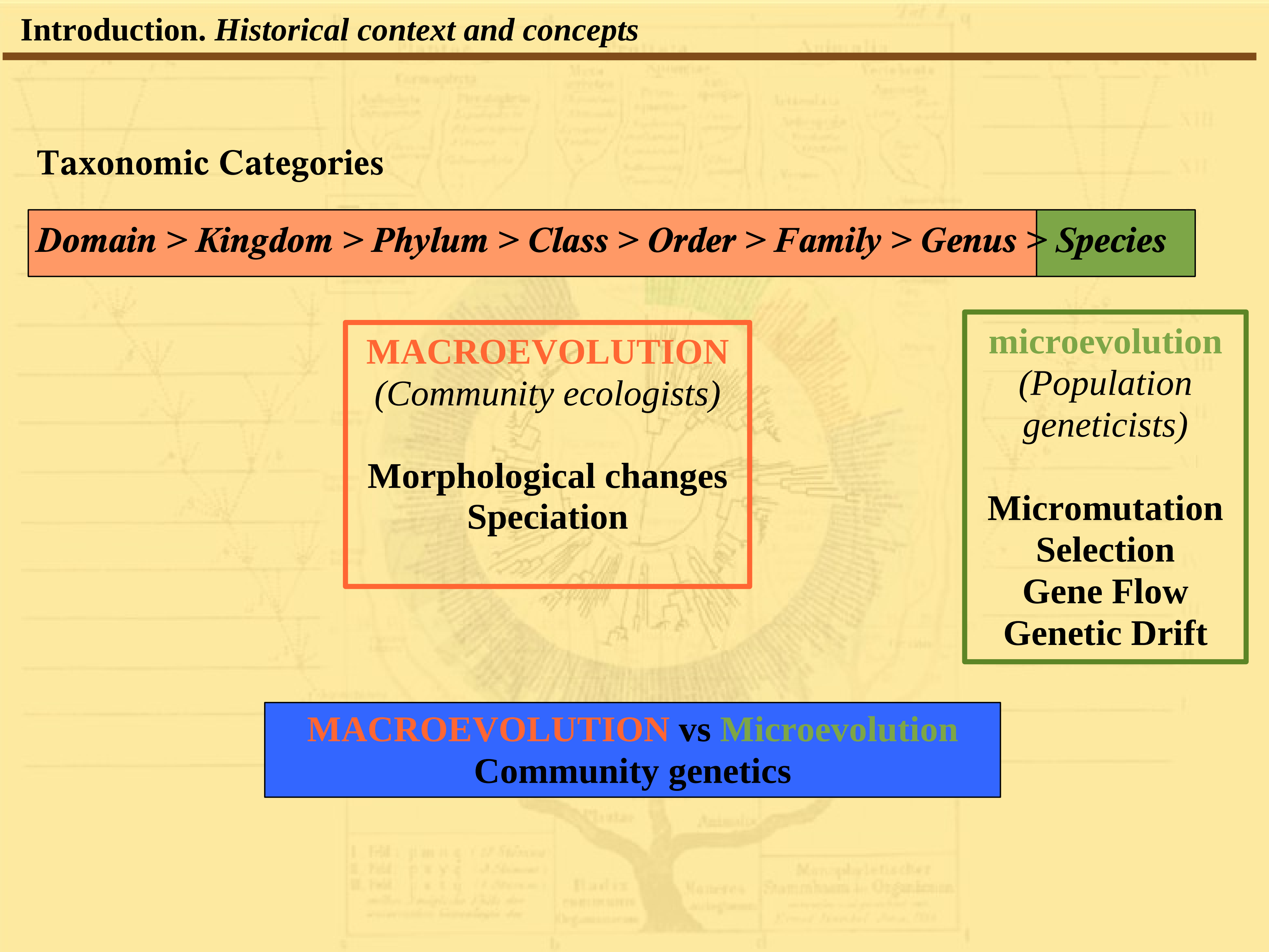


Allometric scaling analysis ($C \approx A^{\eta}$)

- ✓ For each node i of a tree, S_i is the subtree rooted at the node i containing all nodes bellow *i*.
- ✓ The number of nodes in the subtree is called **branch size**, ***A***, which is defined, as $A = |S_i|$.
- ✓ A parameter to describe the shape of a subtree is ***C***, the **cumulative branch size**.







Taxonomic Categories

Domain > Kingdom > Phylum > Class > Order > Family > Genus > Species

MACROEVOLUTION
(Community ecologists)

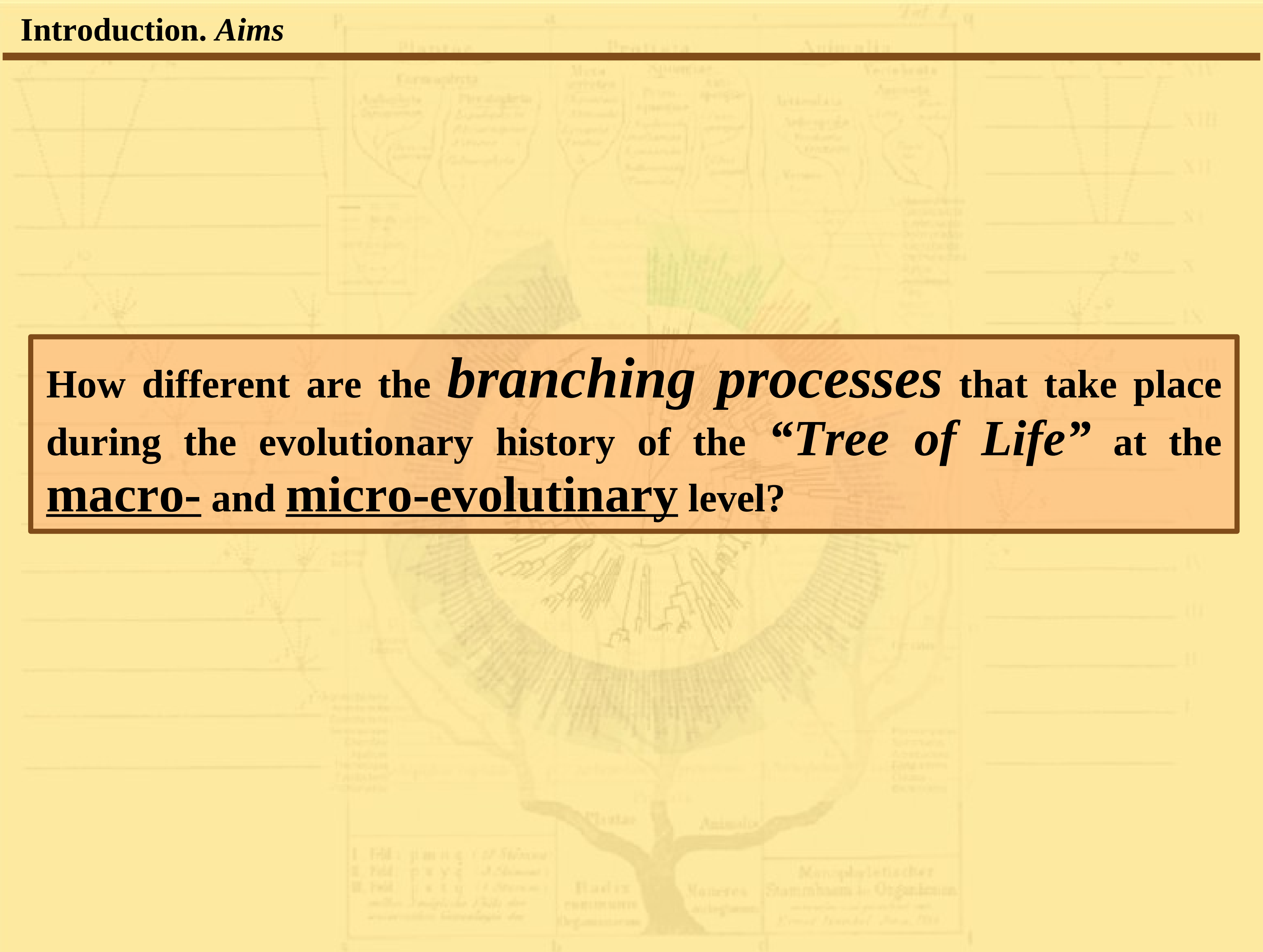
Morphological changes
Speciation

microevolution
(Population geneticists)

Micromutation
Selection
Gene Flow
Genetic Drift

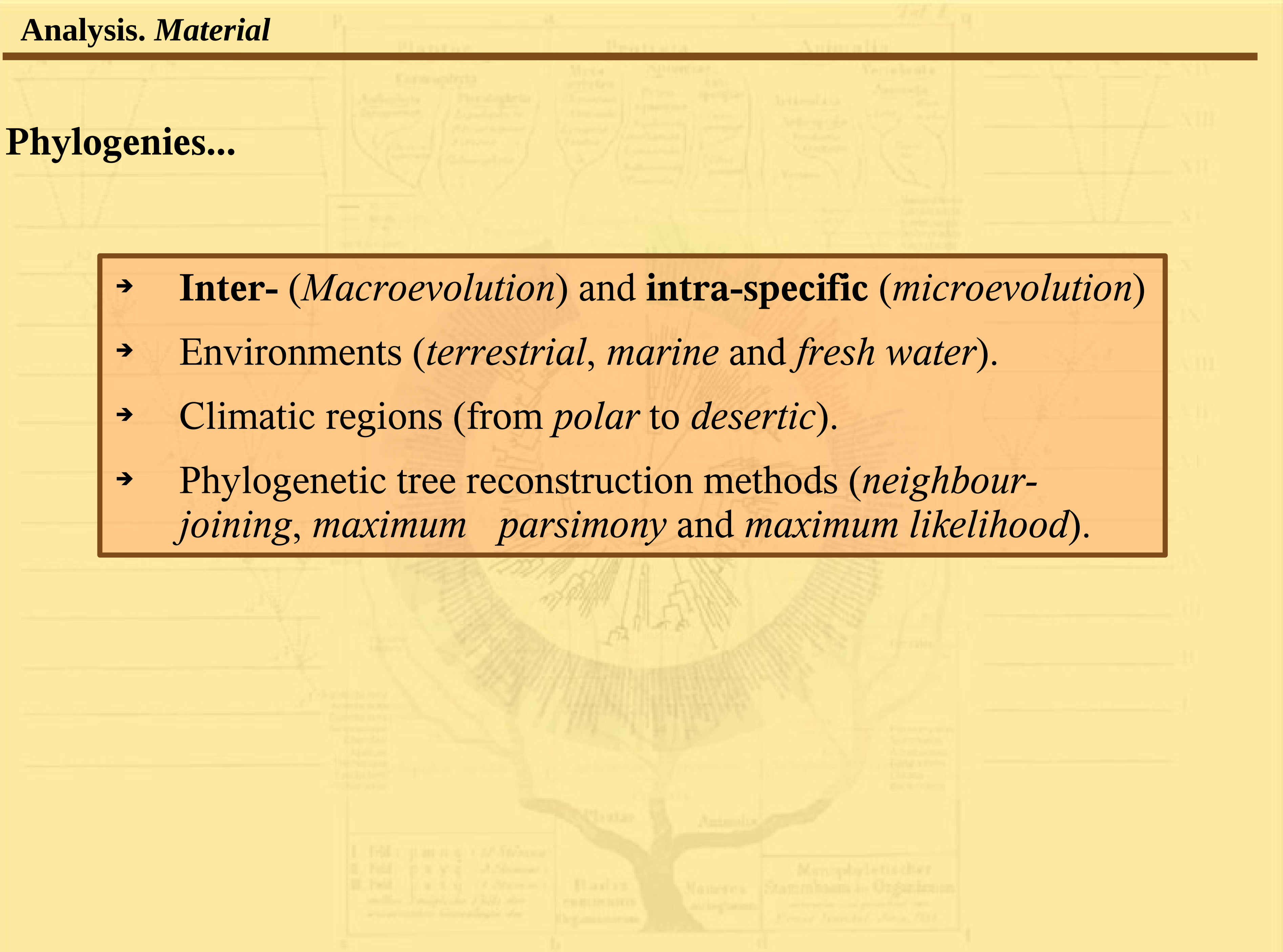
MACROEVOLUTION vs Microevolution
Community genetics

Introduction. *Aims*



How different are the *branching processes* that take place during the evolutionary history of the *"Tree of Life"* at the macro- and micro-evolutinary level?

Analysis. *Material*



Phylogenies...

- **Inter-** (*Macroevolution*) and **intra-specific** (*microevolution*)
- Environments (*terrestrial*, *marine* and *fresh water*).
- Climatic regions (from *polar* to *desertic*).
- Phylogenetic tree reconstruction methods (*neighbour-joining*, *maximum parsimony* and *maximum likelihood*).

Macroevolutionary Phylogenies...

TreeBASE: 4,020 phylogenetic trees.

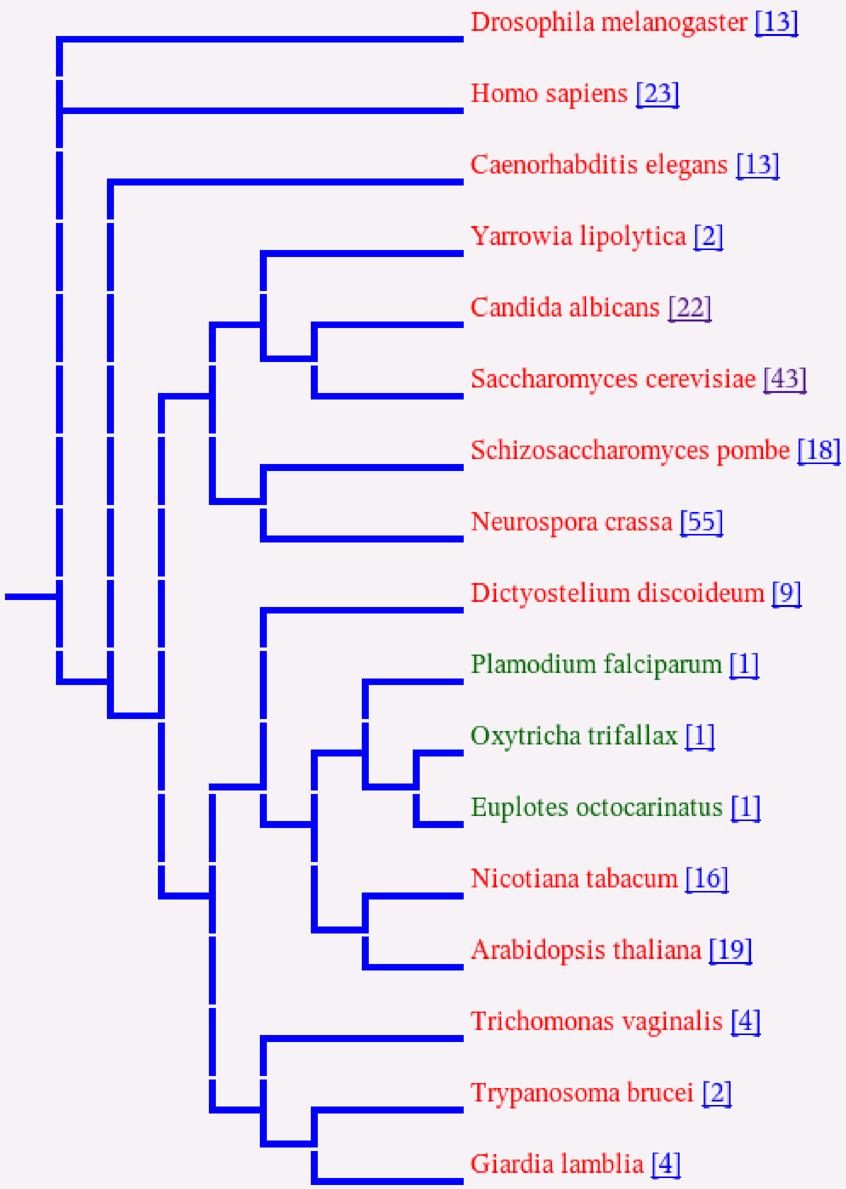
TreeBASE

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.

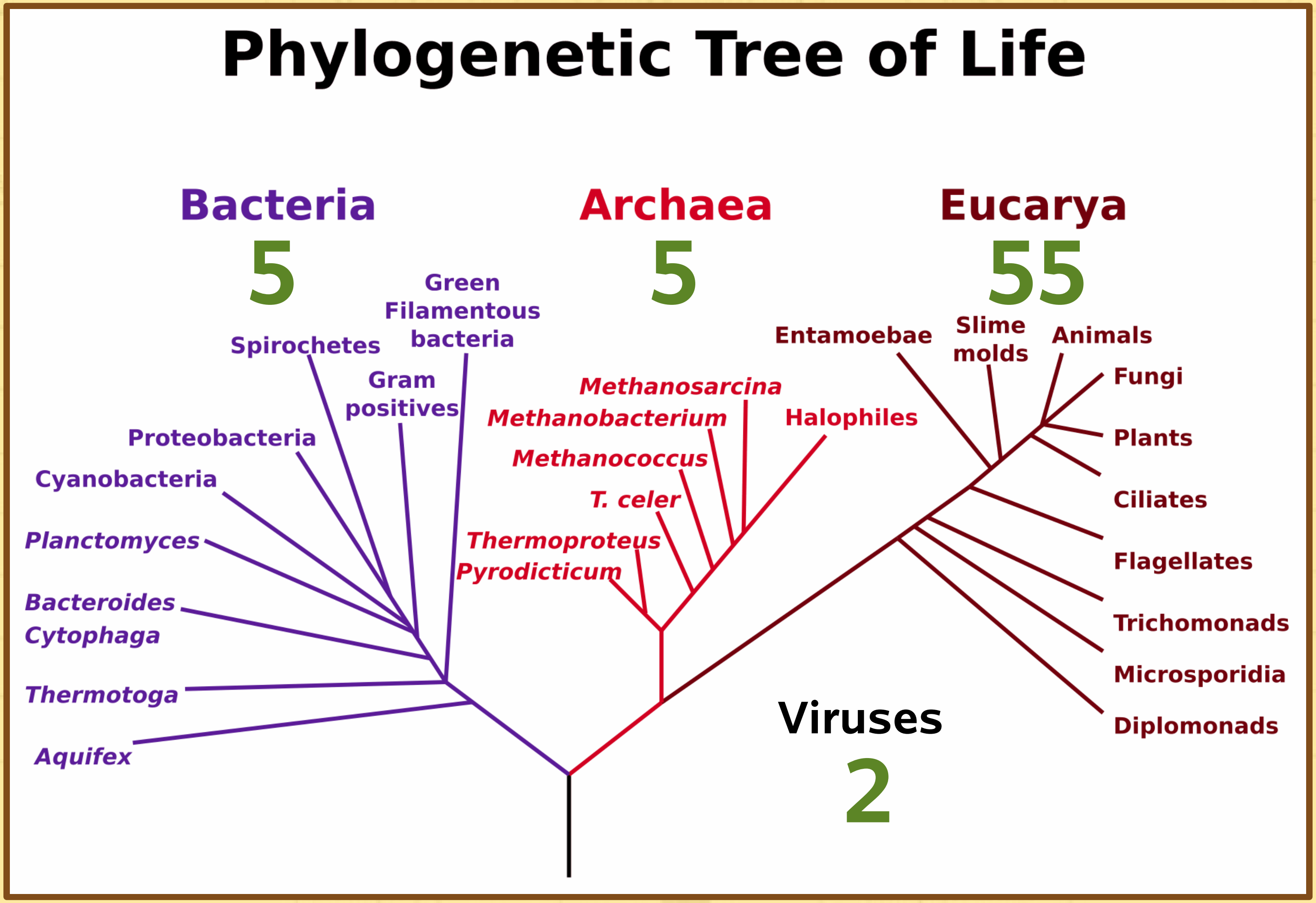
The WWW implementation of TreeBASE requires a forms-capable and frames-capable browser. We would be very grateful for any feedback on TreeBASE, including suggestions for improvement. In particular, if you encounter any errors please let us know.



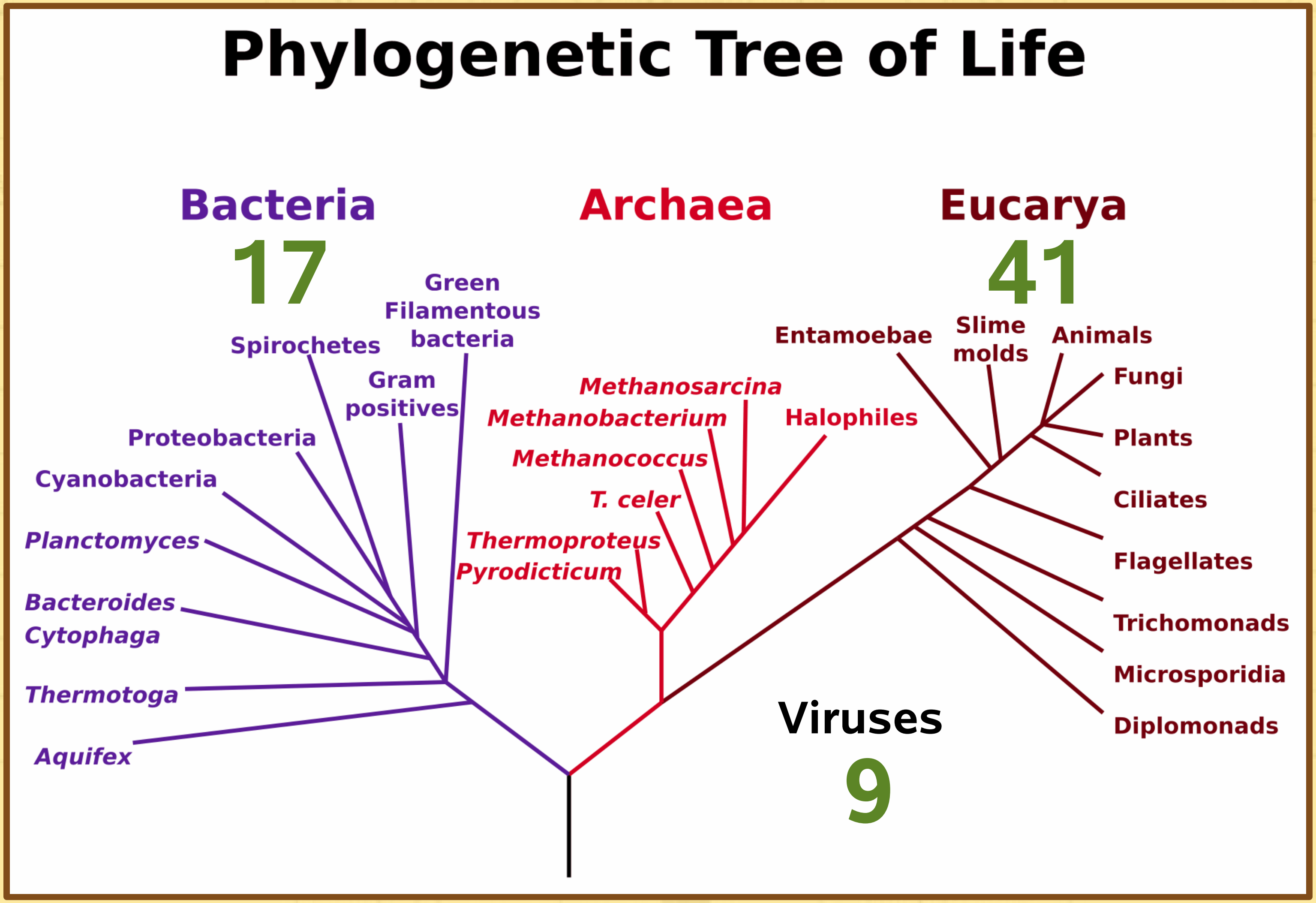
http://www.treebase.org

Analysis. *Material*

…



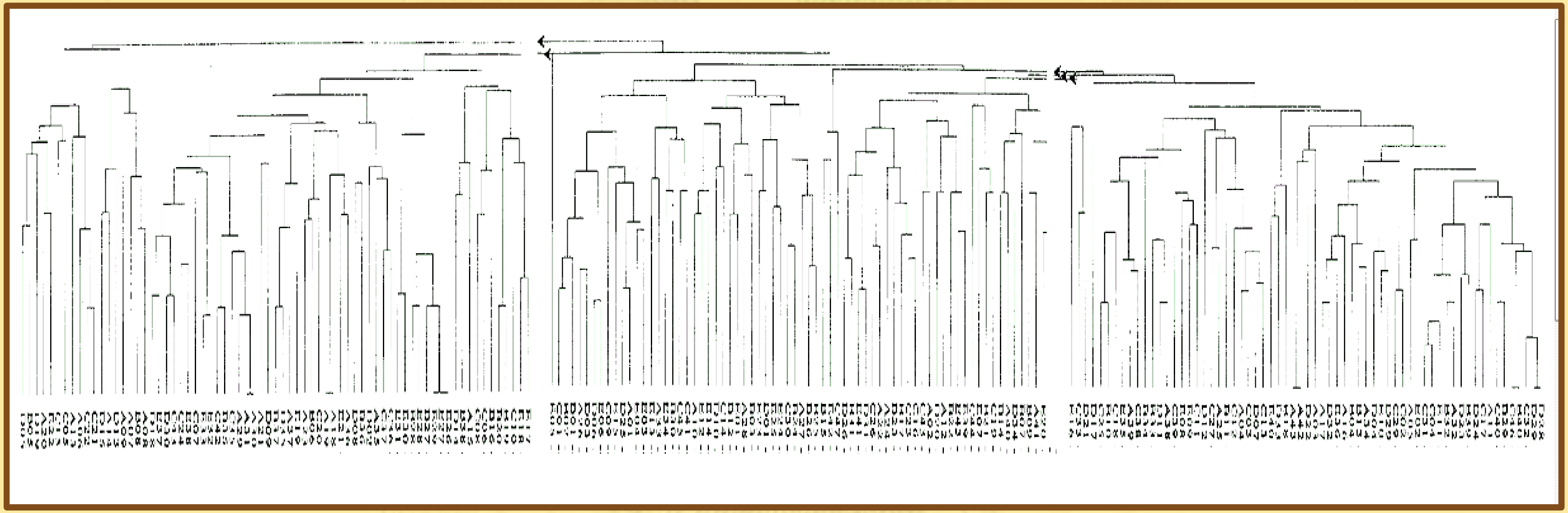
microevolutionary phylogenies...



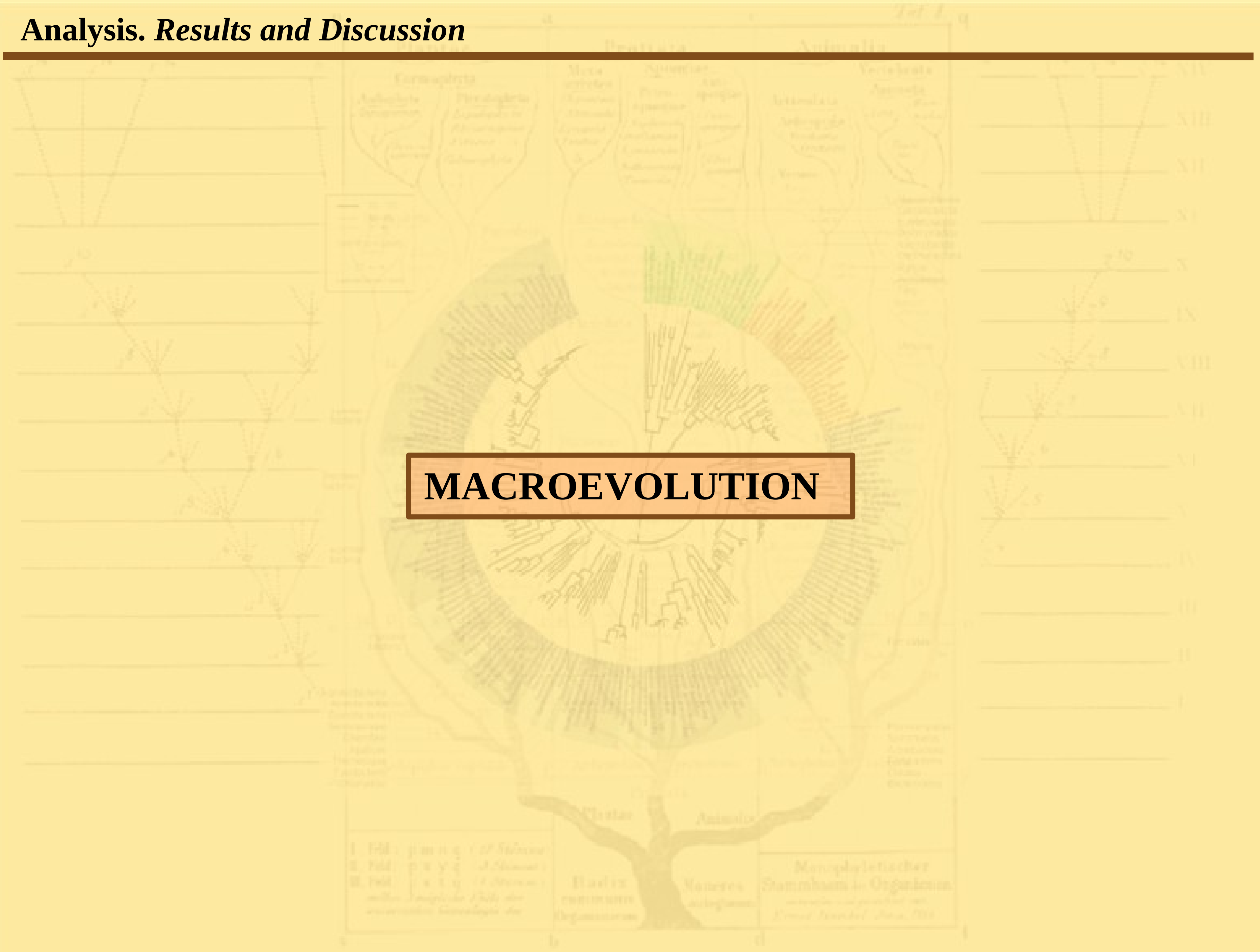
Analysis. *Material*

Vibrio vulnificus **(Bacteria)**

Lin, M. *Appl. Environ. Microbiol. (2003)*



Analysis. *Results and Discussion*

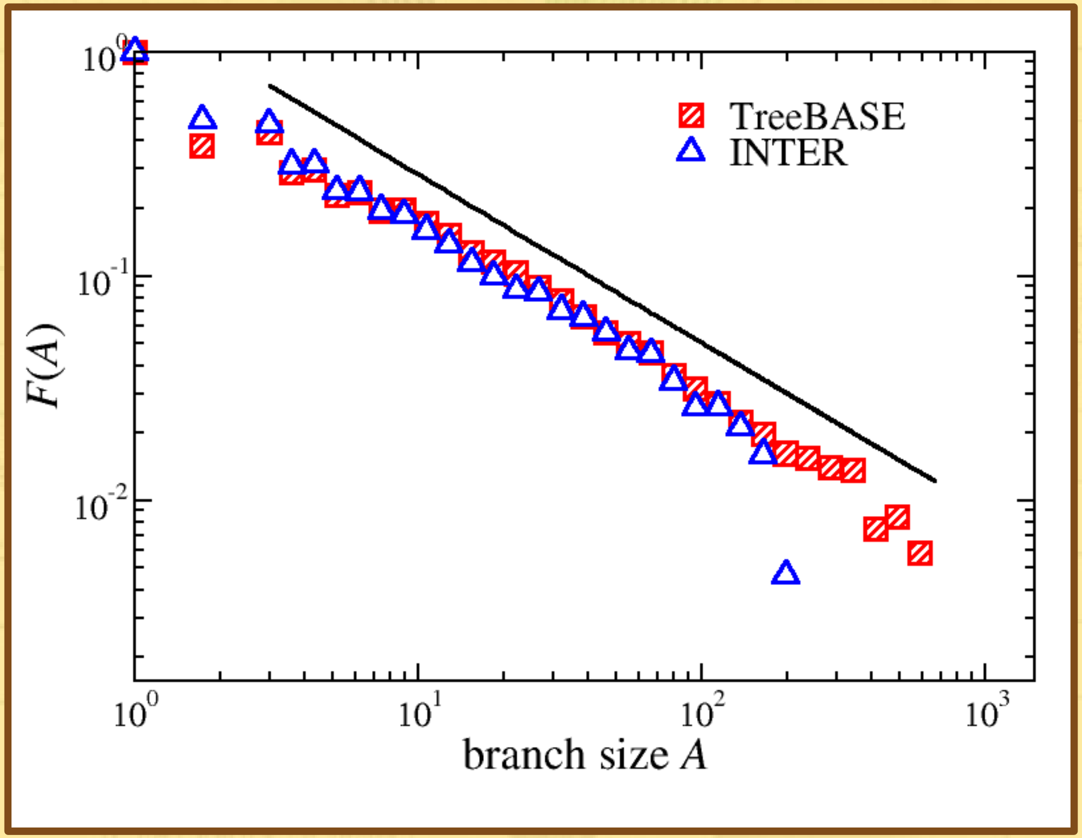


MACROEVOLUTION

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

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

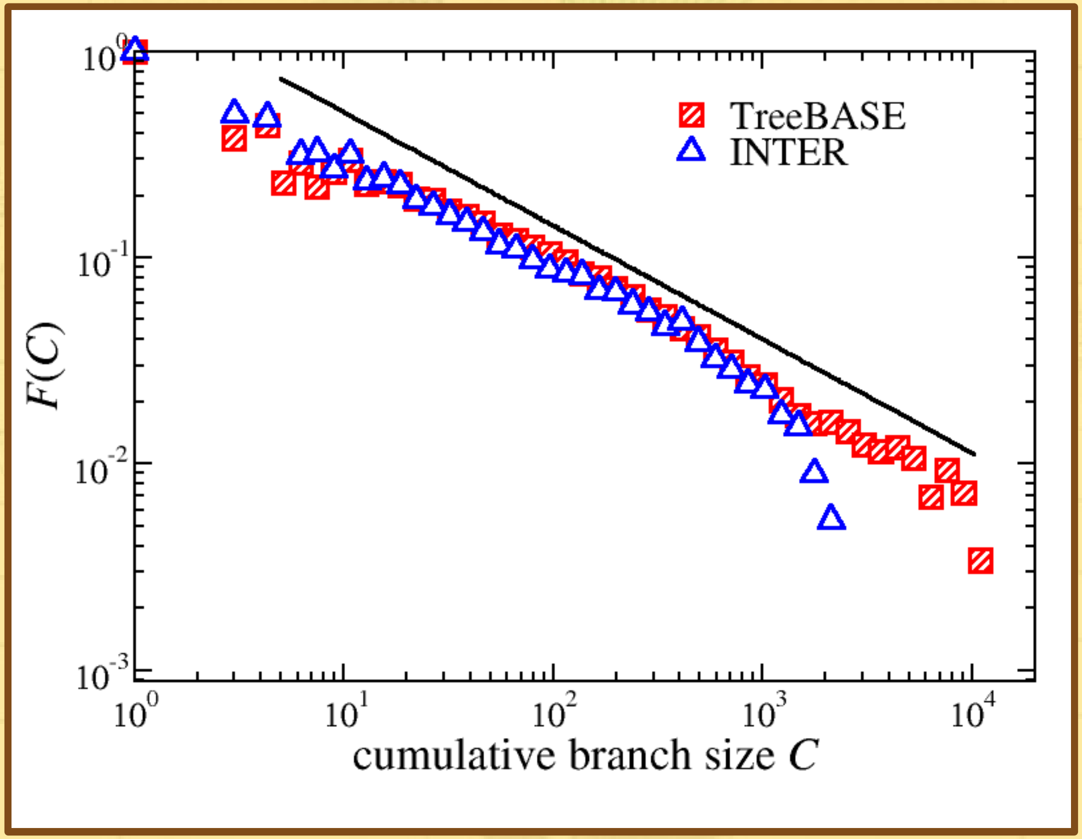
TreeBASE: 0.74
INTER: 0.73



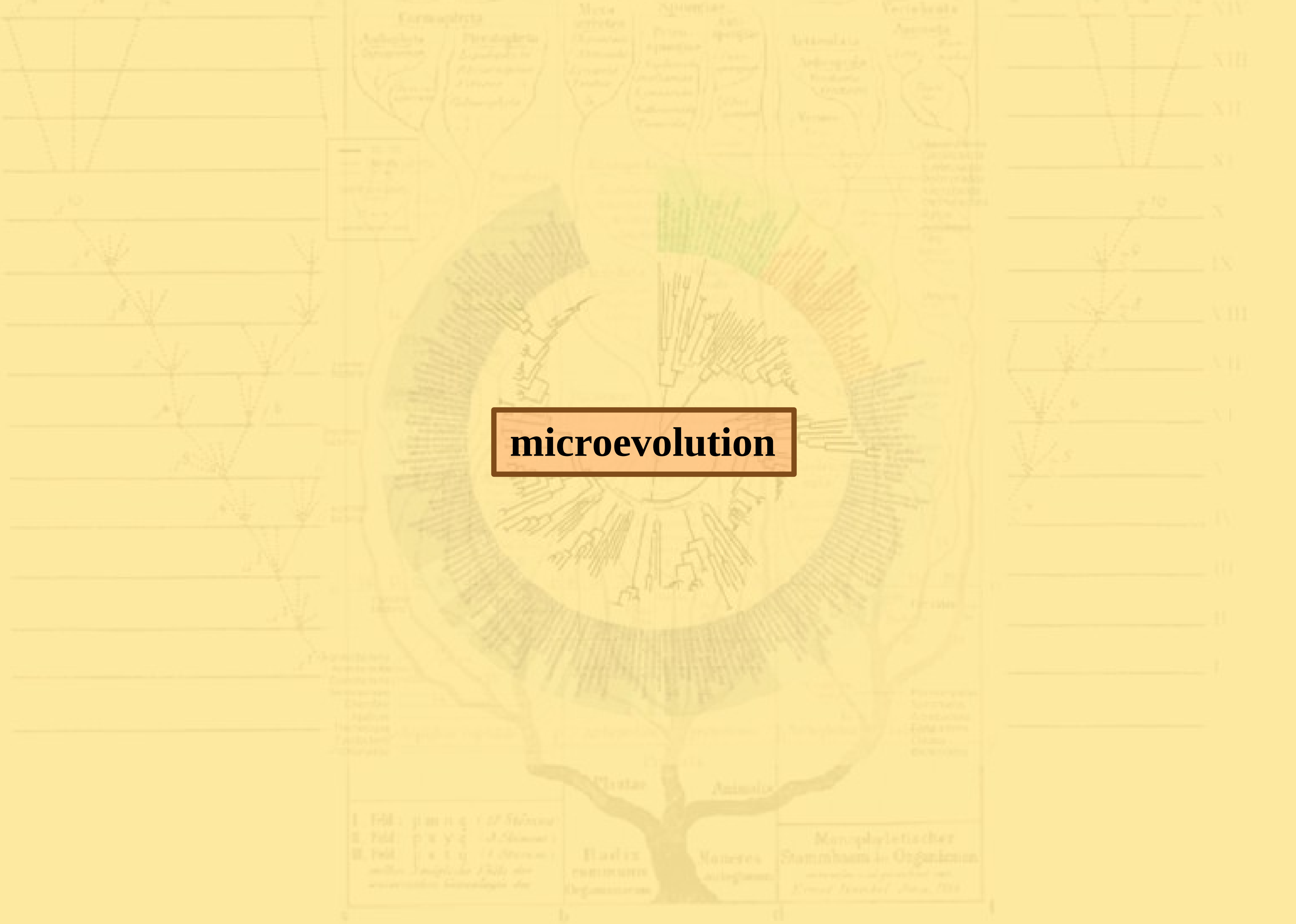
Cumulative branch Size distribution ($F(C) = C^{-\rho}$)

$F(C)$ = prob ($C_i \geq C$)

TreeBASE: 0.53
INTER: 0.52



Analysis. *Results and Discussion*

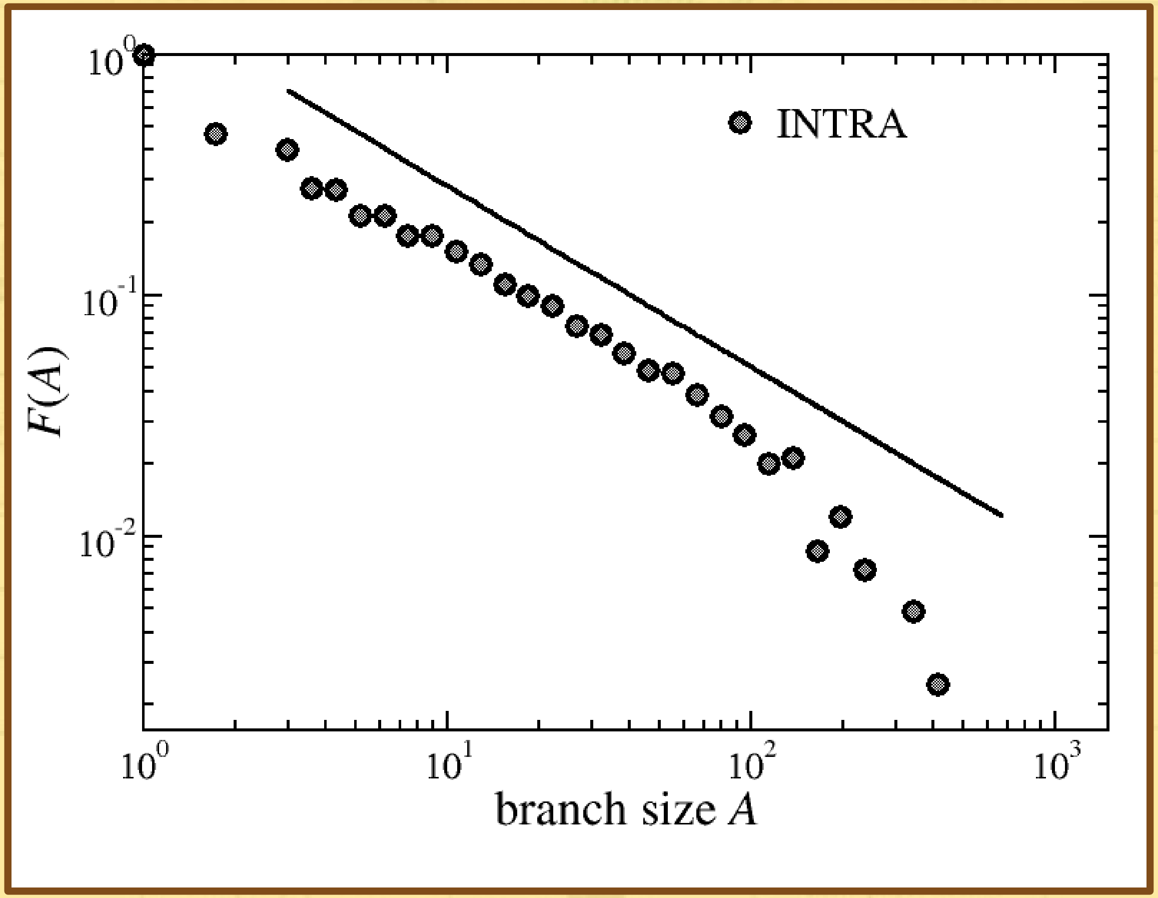


microevolution

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

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

INTRA: 0.73

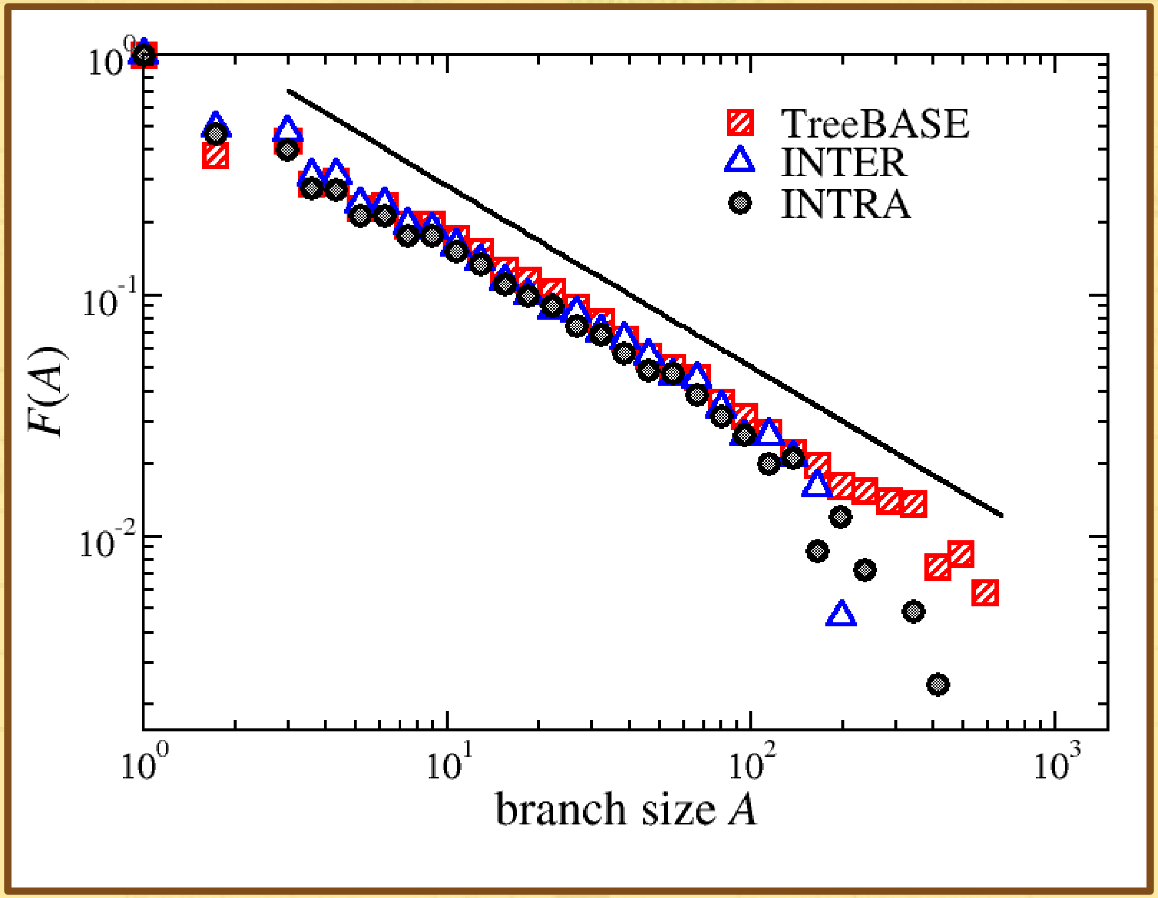


Analysis. ***Results and Discussion***

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

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

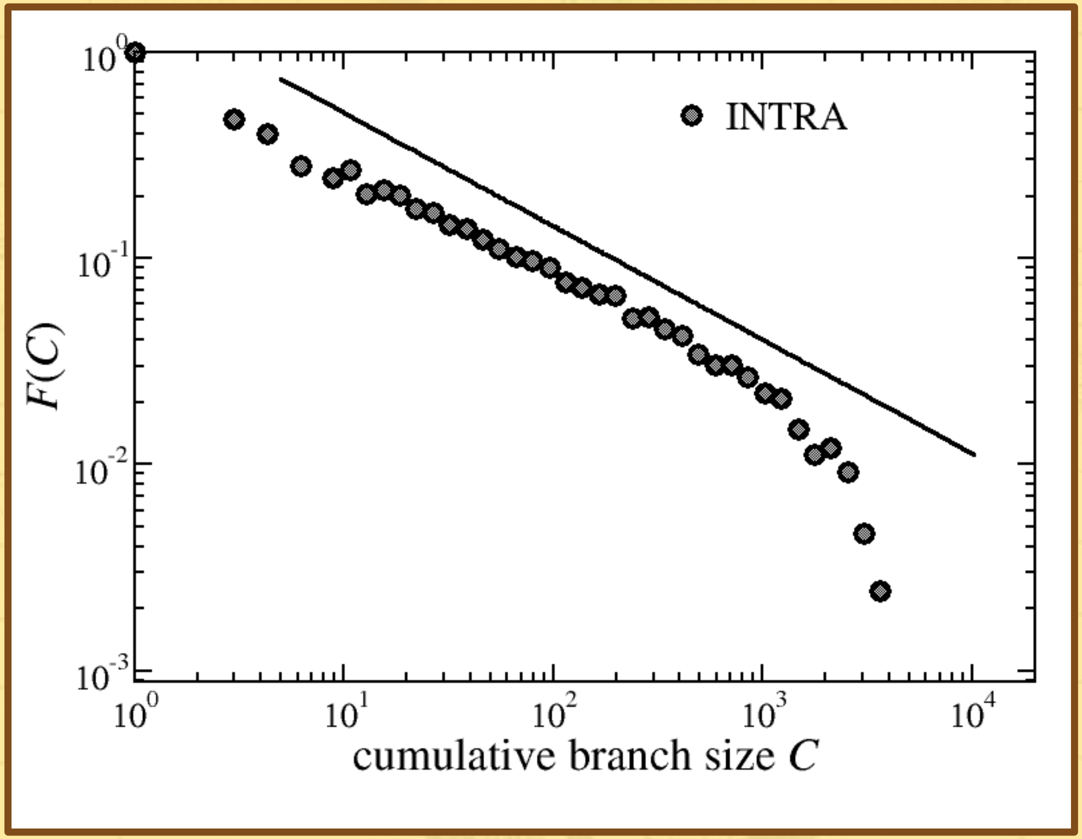
TreeBASE: 0.74
INTER: 0.73
INTRA: 0.73



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

$F(C)$ = prob ($C_i \geq C$)

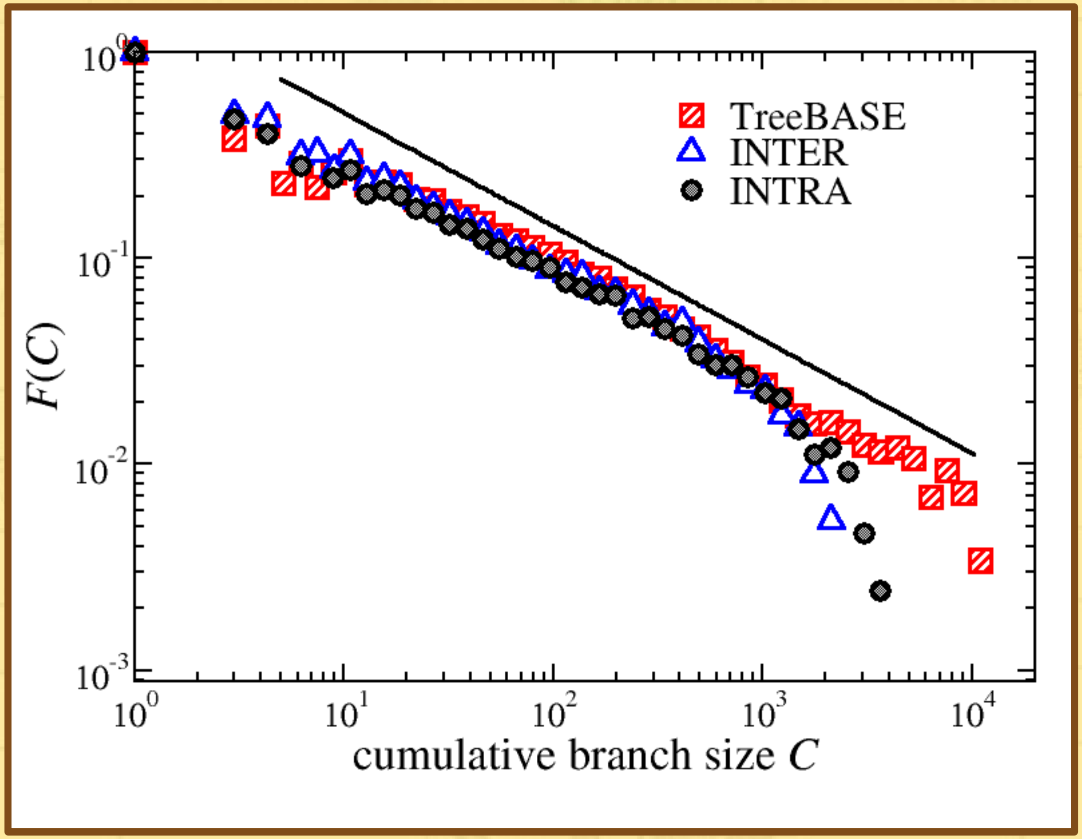
INTRA: 0.51



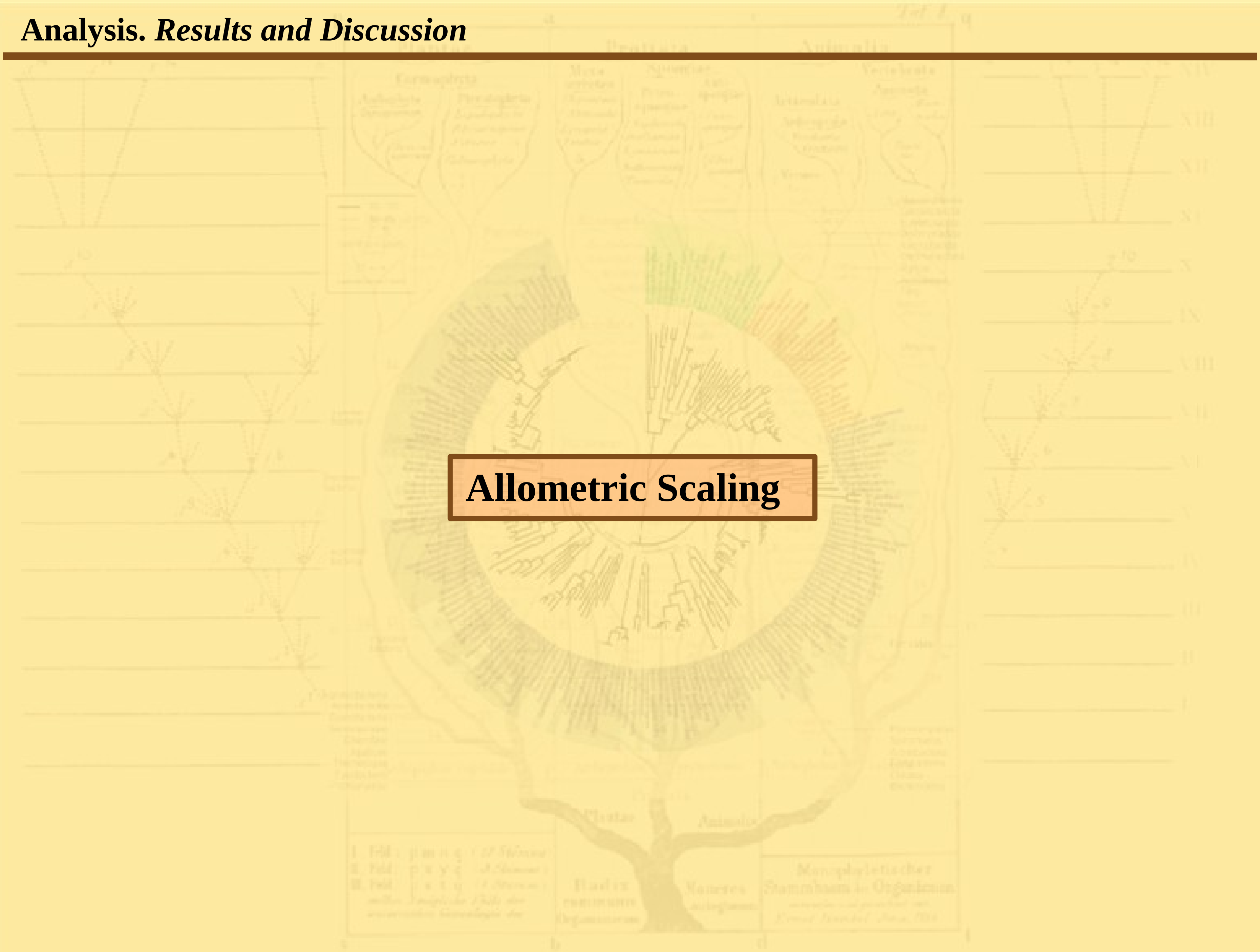
Cumulative branch Size distribution ($F(C) = C^{-\rho}$)

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

TreeBASE: 0.53
INTER: 0.52
INTRA: 0.51



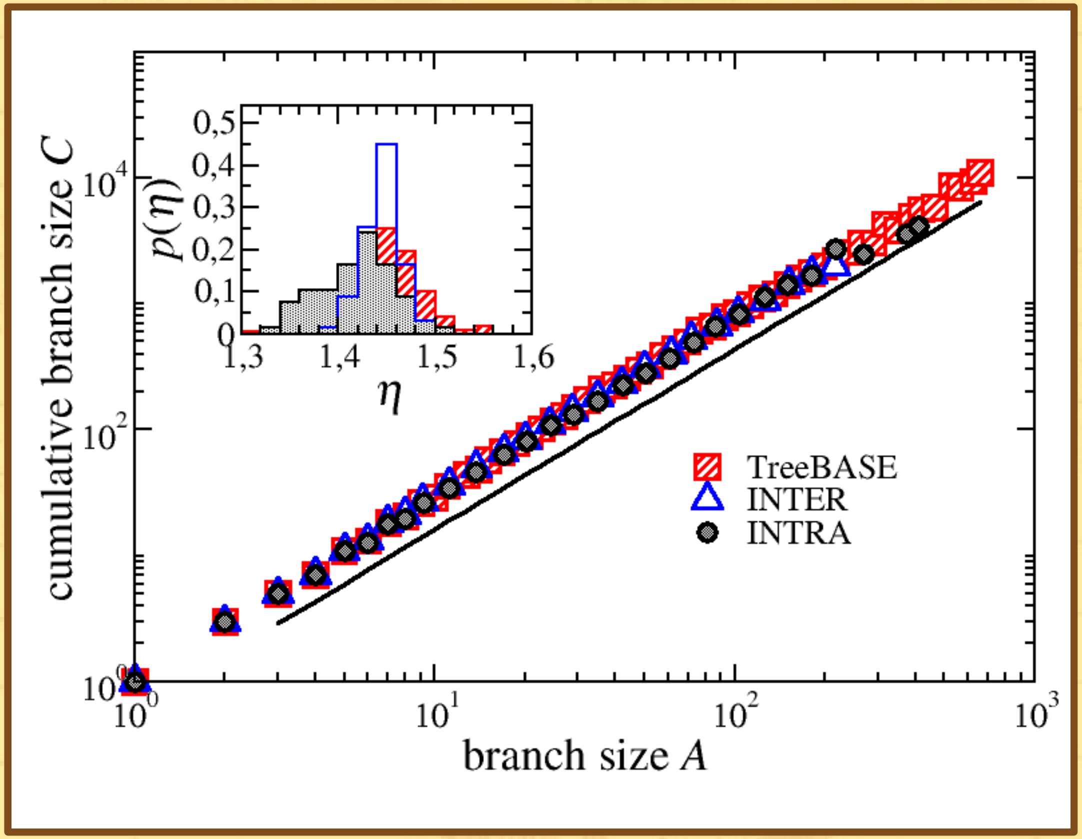
Analysis. *Results and Discussion*



Allometric Scaling

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

TreeBASE: 1.46
INTER: 1.43
INTRA: 1.43



Inferring Evolutionary process from phylogenetic tree shape...

Null models:

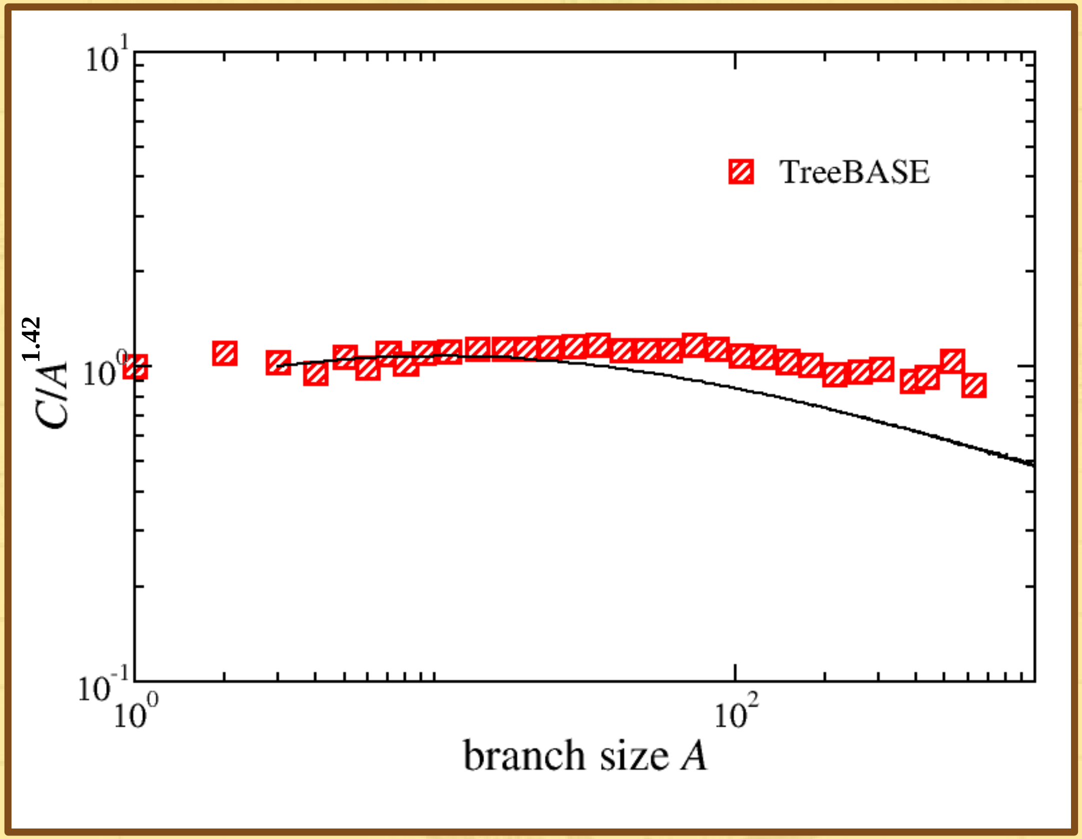
- ***Equa-Rates-Markov (ERM) model:*** Yules(1924), Cavalli-Sforza(1967) and Harding (1971)
- ***Proportional-to-Distinguishable-Arrangements (PDA) model:*** Rose (1978)
- ***Equiprobable-types (EPT) model:*** Simberloff et al. (1981)

Null model...

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.
- We generated a random tree, whose size goes to 10^6 tips, following the same evolutionary principle of the equal-rates Markov (ERM) model.

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



Conclusions

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 (1.45-1.47).
- We have found striking similarities between the power-law exponents for the intra- and for the inter-specific 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.

I·M·E·D·E·A
Institut Mediterrani d'Estudis Avançats

Topological diversity in phylogenies: microevolution vs macroevolution

E. Alejandro Herrada*,

Claudio J. Tessone, Víctor M. Eguíluz, Emilio Hernández-García

and Carlos M. Duarte

Instituto Mediterráneo de Estudios Avanzados (IMEDEA)

www.imedea.uib.es/physdept/

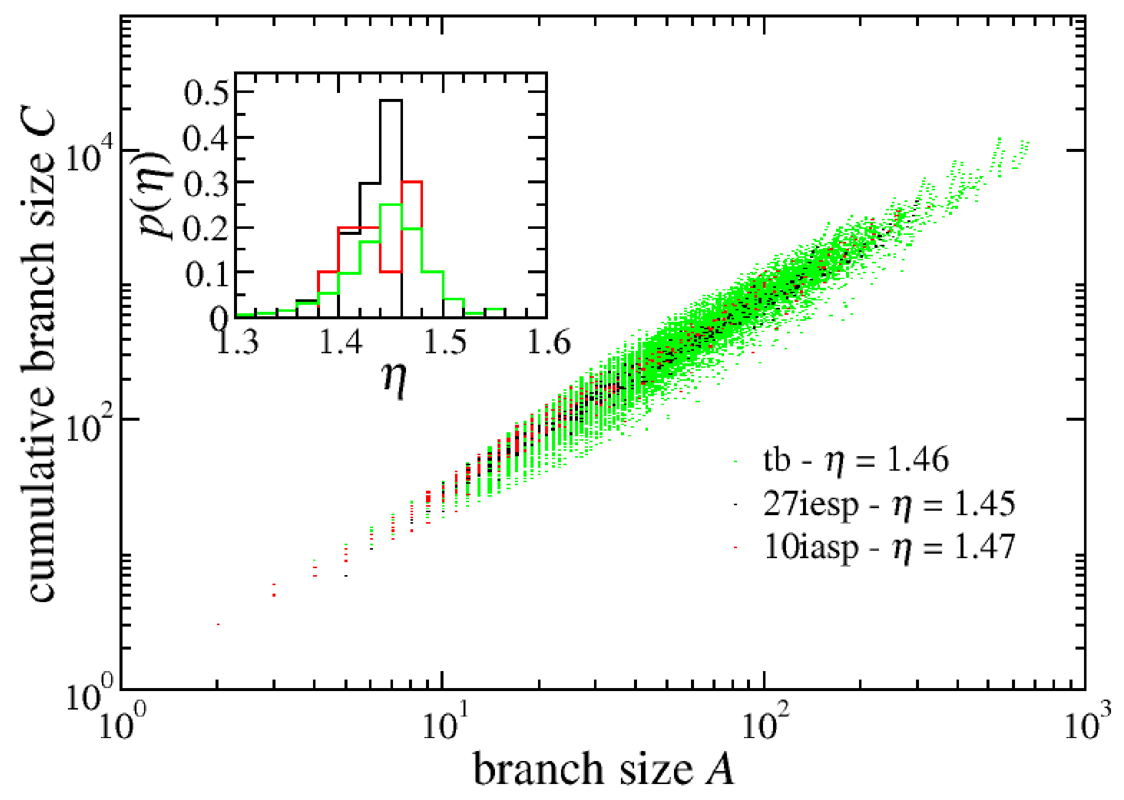
CSIC-Universitat de les Illes Balears, Palma de Mallorca, Spain

* alejandro@imedea.uib.es

Girona (Catalunya), November 2006

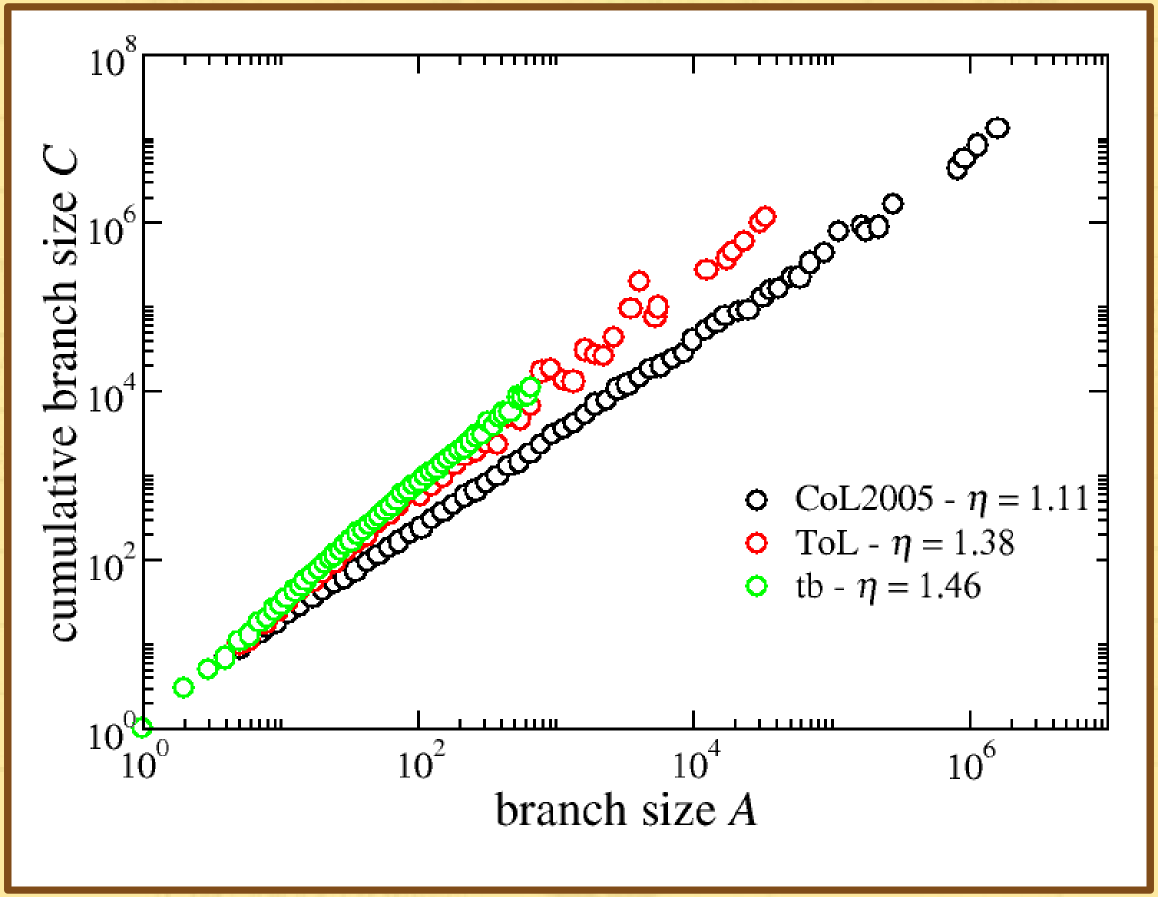
Analysis. ***Results and Discussion***

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



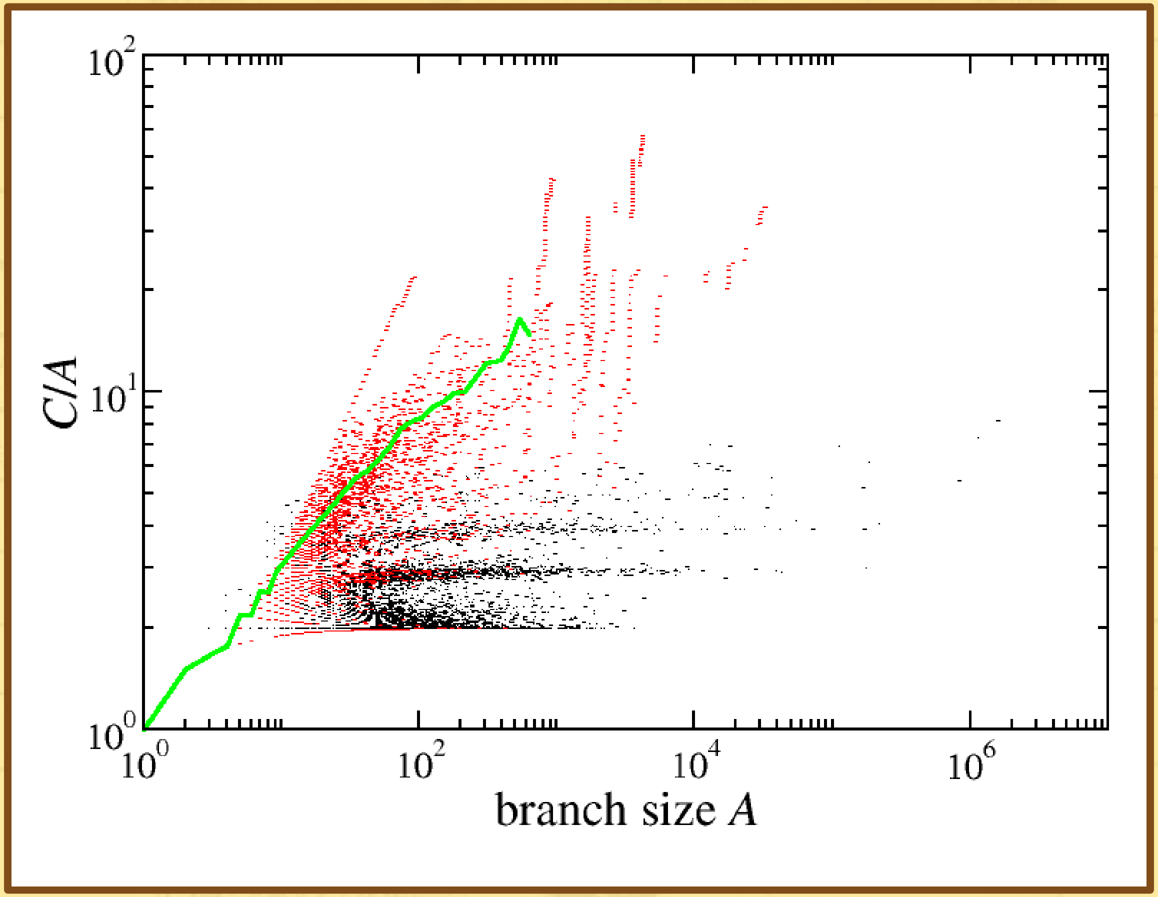
Analysis. *Results and Discussion*

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



Analysis. *Results and Discussion*

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



Analysis. *Results and Discussion*

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