

Funded by the Horizon 2020 Framework Programme of the European Union

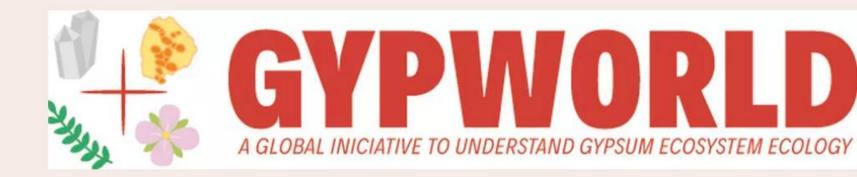




# 1st GYPSUM ECOSYSTEM RESEARCH CONFERENCE

June 3-9, 2018 Ankara/Turkey - Gypsum Ecosystems as biodiversity hotspots

Registration Deadline is May 15th 2018.>





# GYPWORLD Training Course Analysis of Phylogenies

# **Part 1: Phylogenetic study design:** How to sample appropriate study units



Helga Ochoterena

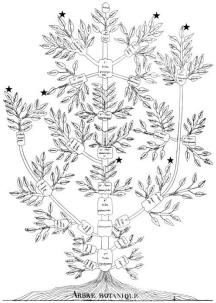
# Introduction: evolution vs phylogeny

Darwin suggested in 1859 that all living organisms are connected through evolution, sharing a single common ancestor

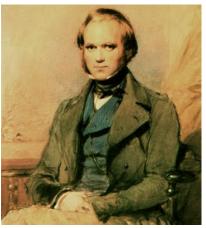
### (descendent with modification).

Even though the notion of modification through time (evolution) was present before, Darwin's contribution revolutionize science, since he suggested a mechanism to explain evolution

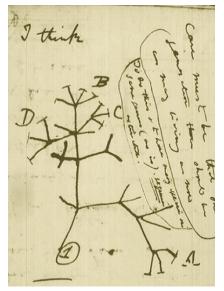
(adaptation and natural selection).



Augustin Augier 1801

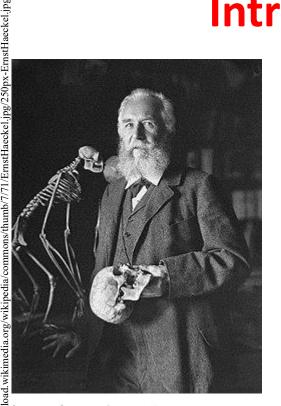


http://www.charlesdarwin.fr/dico\_esp.html

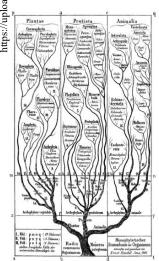


http://cmcdianatausia.blogspot.mx/2012.html

# Introduction: evolution & phylogeny



Since that moment, questions related to evolution (e.g., how to establish evolutionary hypotheses, major adaptive changes, "trends", etc.) start to become popular and intense debates about how to reflect evolution in classifications created heated discussions.



Ernst Haeckel (1834-1919)



http://es.wikipedia.org/wiki/Ernst\_Haeckel

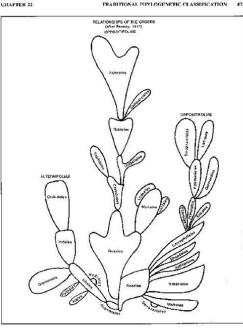


Figure 22-1. Diagram used by Bessey (1915) to illustrate putative relationships among major taxa of the Angiosperms. Used by permission of the Missouri Betanical Garden.

# Bessey's "Cactus'

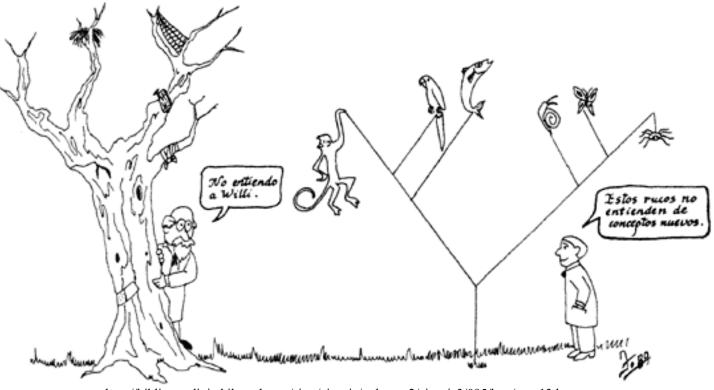
Placed plant groups with many floral parts in a basal position as the 'ancestral' forms.

Outlined 'dicta' for the construction of phylogenies using the evolutionary trends in character changes.

Polypetalous flowers, insect pollination, cycad-like ancestors

# **Cladistics (also known as phylogenetic systematics)**

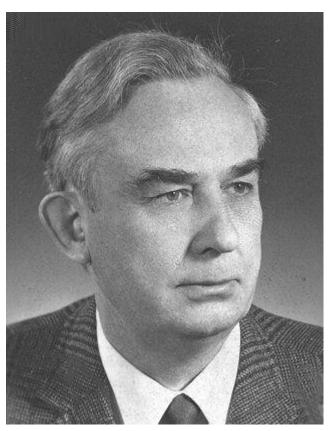
Cladistics was born after the publication of Willi Hennig's work (*Grundzüge einer Theorie der phylogenetischen Systematik* 1950), around the time that phenetics became popular.



http://bibliotecadigital.ilce.edu.mx/sites/ciencia/volumen2/ciencia3/095/htm/sec\_13.htm

# Cladistics

Hennig's work was not acknowledge until it was translated into English (1966), as a book entitled *Phylogenetic systematics*.



http://commons.wikimedia.org/wiki/File:Willi\_Hennig2.jpg

# **Cladistics: Hennig's basic distinctions**

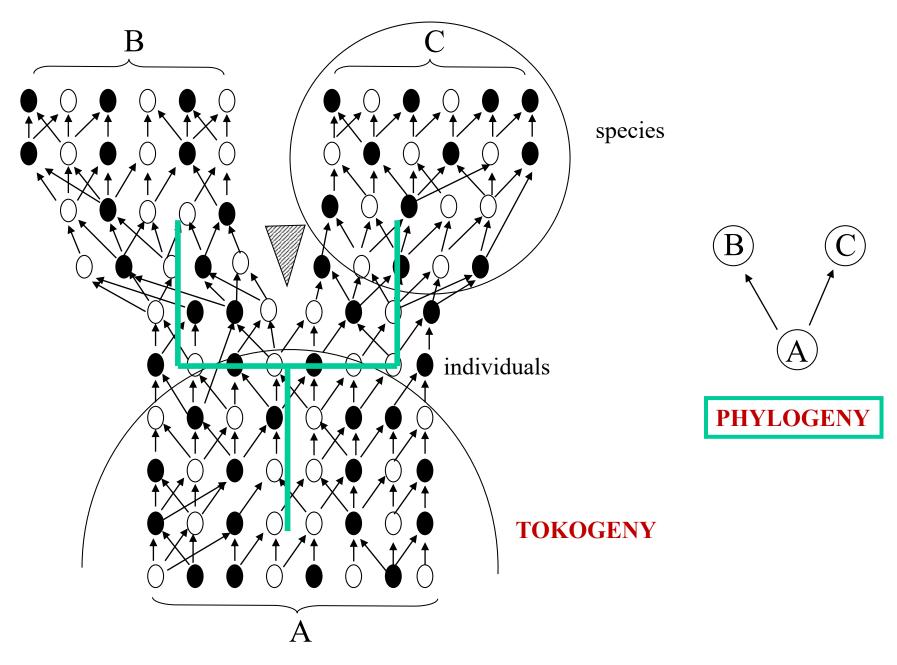
Two types of relationships ocure among organisms:

### Tokogenetic

With reticulate structure Among individuals (e.g. father-son)

### **Phylogenetic**

At the level of species (ancester-descendent) Typically with hierarchical structure Resulting from descendence with modification followed by genetic isolation

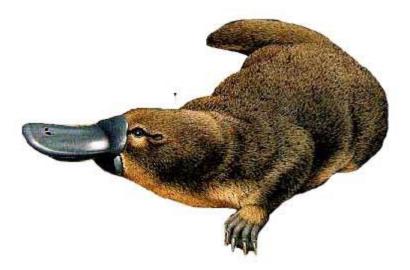


Modified from Hennig 1966

# **Cladistics: Very logical**

The main driver in evolution is **descendency with modification** 

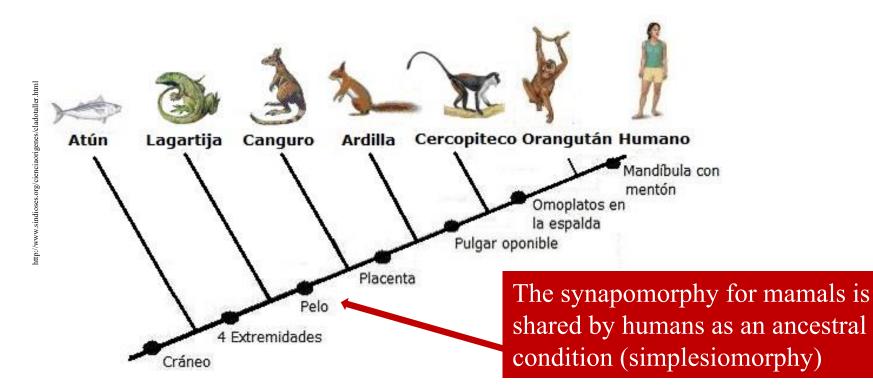
All the species that descend from the same ancestor will have the modification that gave rise to the lineage they belong in; species have shared derived character states (synapomorphies) hierarchically accumulated; they also have the **modifications that gave rise to their particular species (the more recent lineage)** 



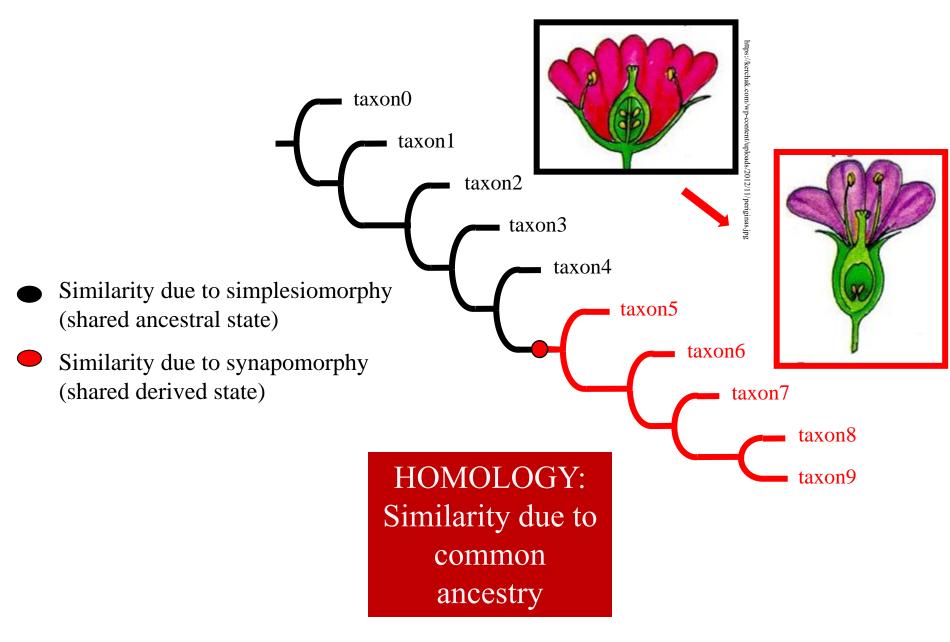
http://www.articulosweb.net/animales/ornitorrinco

### **SINAPOMORPHIES (shared derived conditions) provide evidence for different hierarchical relationships**

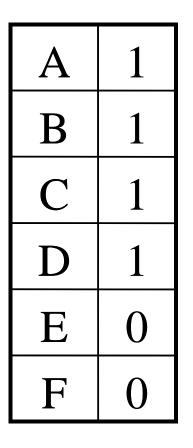
More or less nested synapomorphies help postulating hypotheses about how species are related and how relatively recent the divergency was. Different setts of characters provide evidence at different hierarchical levels depending on when they evolved and how fast the rates of evolution are for that particular character.

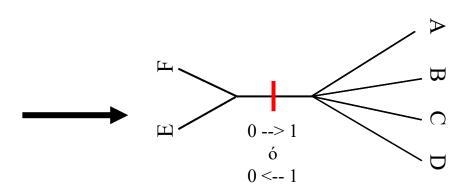


# **HOMOLOGY: NOT JUST SINAPOMORPHY**



# **HOMOLOGY vs SINAPOMORPHY**



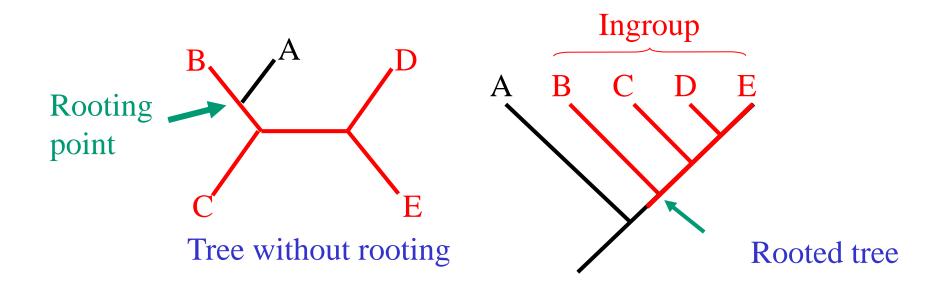


Because both, the ancestral and the derived conditions are hypothesis of homology, it is not necessary to *a priori* know which of the alternative conditions is the derived one and which one is the ancestral: it is the postulation of a change what reflects homoplogy

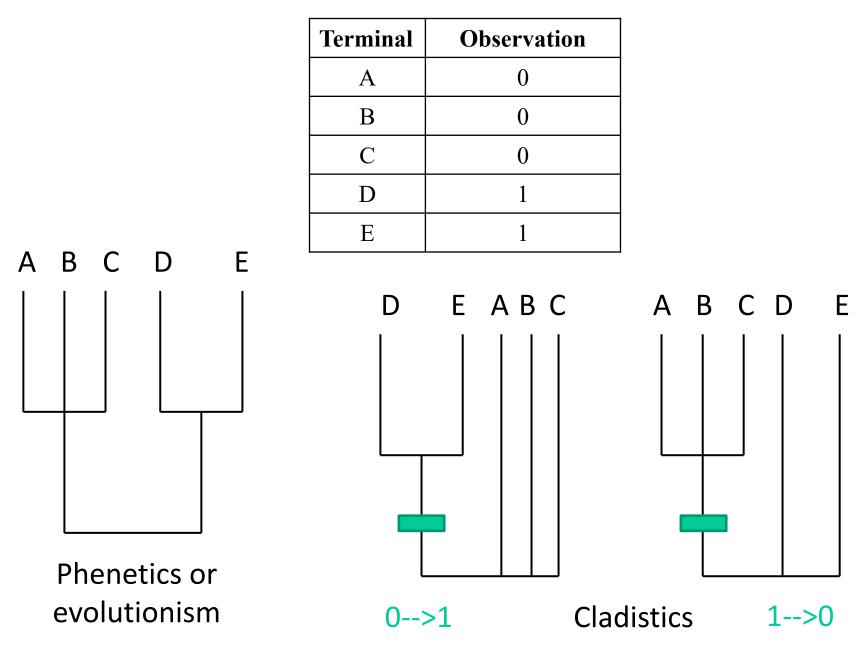
Groups are not recgnized by the synapomorphies but instead by the hypothesis of a single transformation between the alternative observations

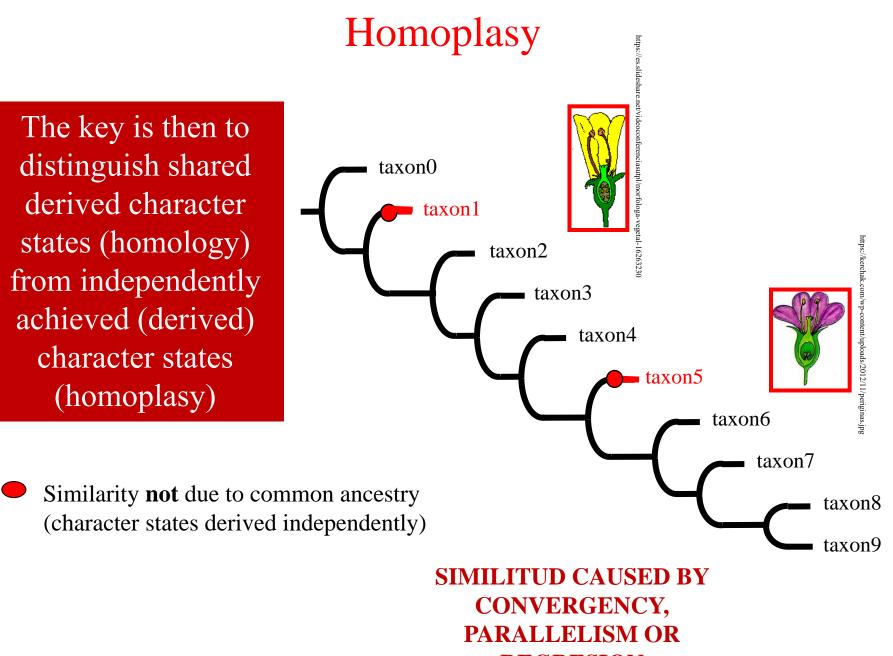
### **OUTGROUP and character polarity**

To interpret the directionality of the character change (the polarity of the character) it is necessary to orient (root) the tree. One could arbitrarily root it between any two internodes, but depending on how the tree is oriented the interpretations of relationships can be very different; therefore, one most include in the analyses taxa that do not belong to the immediate group of interest (outgroup)



# Feneticismo vs Cladismo





REGRESION

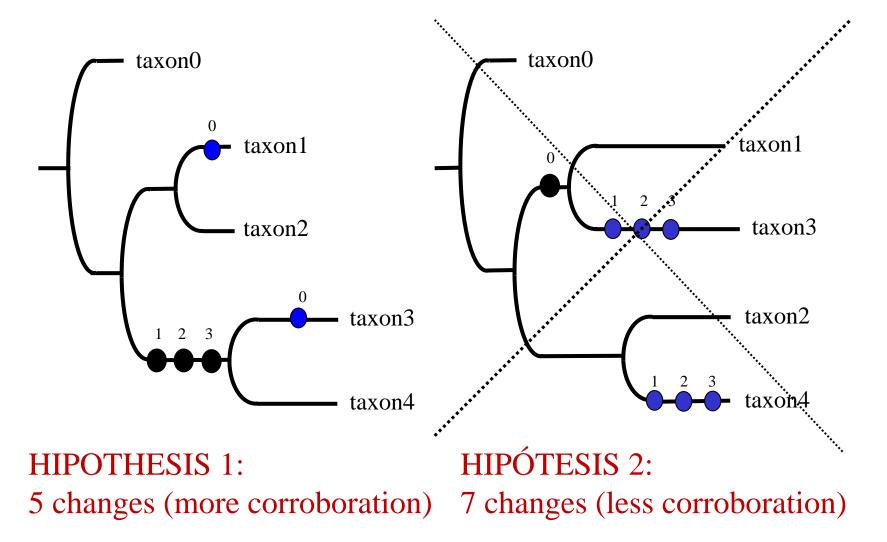
# **PARSIMONY** was proposed for this pourposes by Hennig



Every comparative observation that can be inherited can be postulated as a homology hypothesis; all hypothesis are subjected to a corroboration/refusal test (parsimony) at once; the tree that implies the least number of changes is accepted as the preferable phylogenetic hypothesis

Parsimony in a scientific context means not to assume more causes than those minimally needed to explain the observations

# PARSIMONY: Maximum corroboration of the homology hypotheses (observations)



# **PROBABILITY BASED ON MODELS**

### Likelihood of hypothesis = Probability of data given hypothesis

• Fair or unfair coin?

 $P_{\text{head}} = 0.5$  $P_{\text{head}} = 0.67$  Fair Unfair



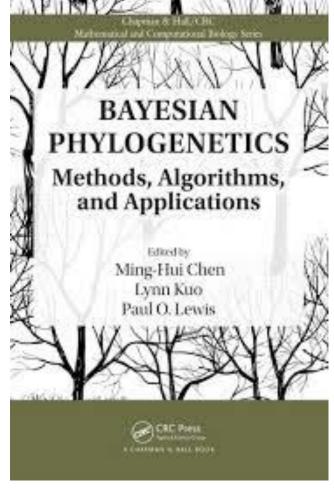
• Flip coin 4 times, get:

3 heads, 1 tail

| ,       | Fair                         | Unfair                       |
|---------|------------------------------|------------------------------|
| НхНхНхТ | 1/2 x 1/2 x 1/2 x 1/2 = 1/16 | 2/3 x 2/3 x 2/3 x 1/3 = 8/81 |
| НхНхТхН | 1/2 x 1/2 x 1/2 x 1/2 = 1/16 | 2/3 x 2/3 x 1/3 x 2/3 = 8/81 |
| НхТхНхН | 1/2 x 1/2 x 1/2 x 1/2 = 1/16 | 2/3 x 1/3 x 2/3 x 2/3 = 8/81 |
| ТхНхНхН | 1/2 x 1/2 x 1/2 x 1/2 = 1/16 | 1/3 x 2/3 x 2/3 x 2/3 = 8/81 |
| Total   | 1/4                          | 32/81                        |
|         | (0.25)                       | (0.40)                       |

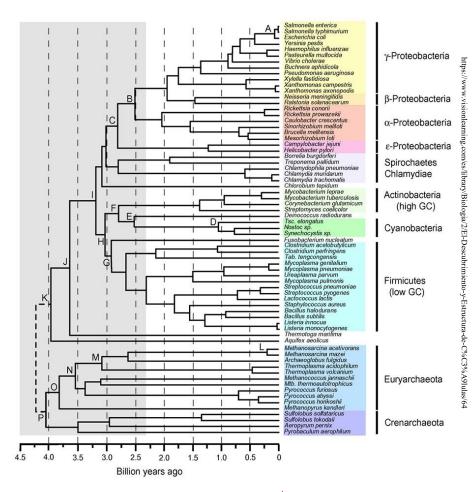
https://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/Phylogenetics/images/phylonv48.gif

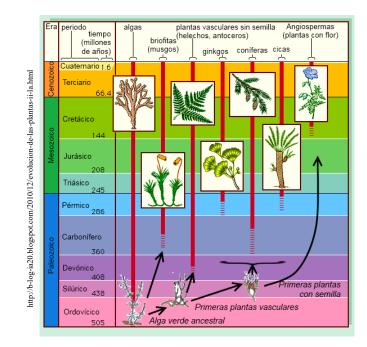
Other methods to choose upon alternative phylogenetic hypotheses are now more frequently used



 $https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRkwxc-Y2ploWQ58mb_7FZSkCrMV-i5FBFSUBBtpVh0KWTz2SMf_w$ 

# In phylogenetics the study units are considered as «TERMINALS»





### Ancestors are hypothetic

VS.

# **Terminals are discoverable (like fossils)**

# **TERMINALS in cladistic analyses**

Theoretically, any type of terminal can be included in a cladistic analysis

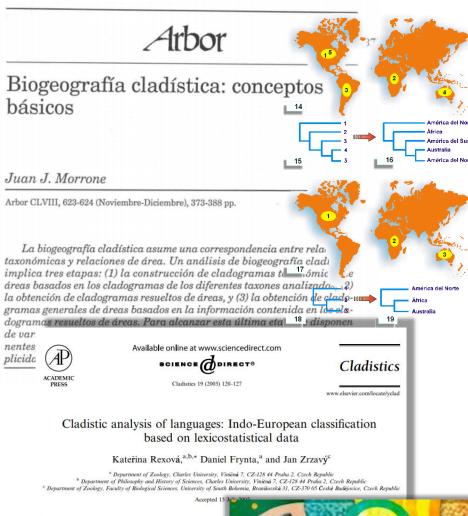
Therefore, the sampling of terminals has to be designed strictly accordingly to the objectives of the study and it will determine, both, the results and their interpretation





https://www.ecologiaverde.com/se-extinguen-150-especies-animales-por-dia-3.html

# **TYPE OF TERMINALS**



#### Abstract

The phylogeny of the Indo-European (IE) language family is relexicostatistical dataset collected by Dyen (about 200 meanings, 8 outgroup). Three different methods of character coding provide trees clade, Romano-Germano-Celtic clade, Armenian-Greek group, and (b) the unstable position of the Albanian language; (c) the unstable p existence of the Balto-Slavonic-Indo-Iranian ("satem") and the Rom compared with the phenetic approach to levicostatistical data, the re basal pattern. The results suggest a predominantly branching patte individual words. Different scenarios of IE differentiation based on 2 020 The Will Hennig Society, Published by Elsevier Science (U





#### Molecular Phylogenetics and Evolution

Contents lists available at SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/ympev

#### Review

njo/Fig1419

Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory

Asher D. Cutter\*

Department of Ecology & Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, ON M5S 3B2, Canada

#### ARTICLE INFO ABSTRACT

Article history: Received 6 April 2013 Revised 6 June 2013 Accepted 12 June 2013 Available online 22 June 2013

Keywords:

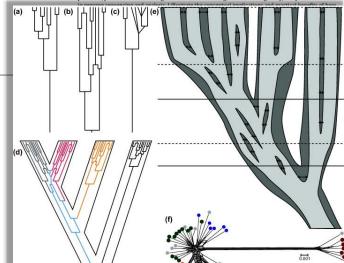
Phylogeography

Genome evolution

Coalescent theory

Species trees Speciation Evolutionary theory is primed to synthesize microevolutionary processes with macroevolutionary divergence by taking advantage of multilouus multispecies genomic data in the molecular evolutionary analysis of biodiversity. While coalescent theory bridges across timescales to facilitate this integration, it is important to appreciate the assumptions, caveats, and recent theoretical advances so as to most effectively exploit genomic analysis. Here I outline the connections between population processes and phylogeny, with special attention to how genomic features play into underlying predictions. I discuss empirical and theoretical attention to how genomic features play into underlying predictions. I discuss empirical and theoretical complications, and solutions, relating to recombination and multifurcating genealogical choiceses. predictions about how enome structure affects ener the hereoneneity, and tractical choices

CrossMark



https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcStw\_g7DRXIP1y0Rd8Ti3I7kCjc\_A5z6yS2YXJFODawsNdt11cA

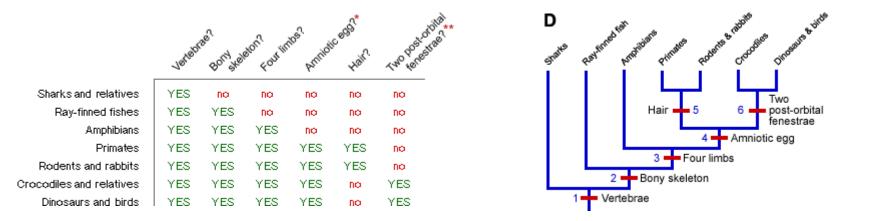
# **Phylogenetic analyses: potential objectives in the gypworld context**

- 1) To know which is the sister species of a gypsophyte in order to understand its affinity to/origin in/adaptation to gypsum
- 2) To understand how & when the species reached the gypsum habitat(s) in order to understand the origin of the flora
- 3) To identify (morphological) **adaptations** that allow the species to have a better fit to gypsum than outside it
- 4) To explore if different gypsum areas could be connected through colonisations of already adapted species followed by diversification or de-novo



# If a cladogram is to be interpreted as a phylogeny terminals need to be potential evolutionary units

To answer questions like those in the previous slide, one needs to consider terminals as potential evolutionary units and characters as inheritable traits with hypothesized evolutionary transformations: 1) they most be intrinsic; 2) the character states (variations of the trait) most be shared by more than one terminal



# **TERMINALS can cover differnt hierarchical levels**

Phylogenetic analysis can use individuals, populations, species, genera, families, etc. as sampling units

Neogene

4.0

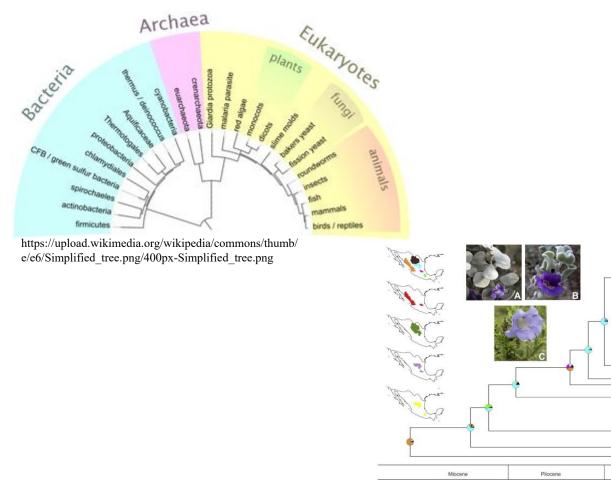
0.8 nucleotide

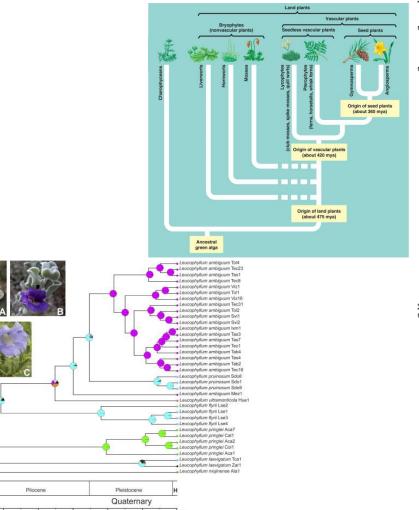
3.0

2.0

1.0

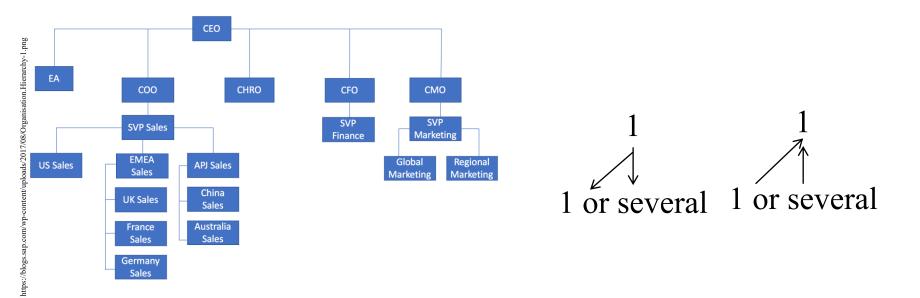
p.oMa





# **TOKOGENY & PHYLOGENY**

In general, tokogenetic relationships **do no** produce **natural** hierarchical structure, while phylogenetic relationships normally result in **natural** hierarchical structure

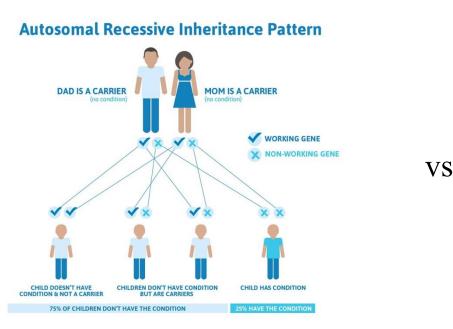


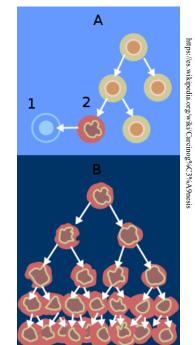
**Hierarchy**: a structure that is established with a subordination criterium

# **TYPE OF TERMINALS**

### According to Hennig's model, the appropriate terminals for phylogenetic studies are those that replicate (and hence produce hierarchical patterns)

Replication implies clonal reproduction or total inheritance

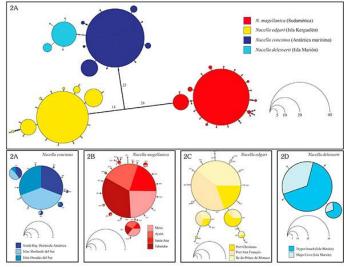




https://i1.wp.com/www.geneticsupportfoundation.org/wp-content/uploads/2014/12/autosomal recessive inheritance pattern.jpg?resize=1000%2C885

# **TOKOGENY vs PHYLOGENY**

Because **TOKOGENETIC** relationships are normally reticulated, the study of relationships among individuals of the same species is normally conducted through phylogeographic methods, rather than phylogenetic methods

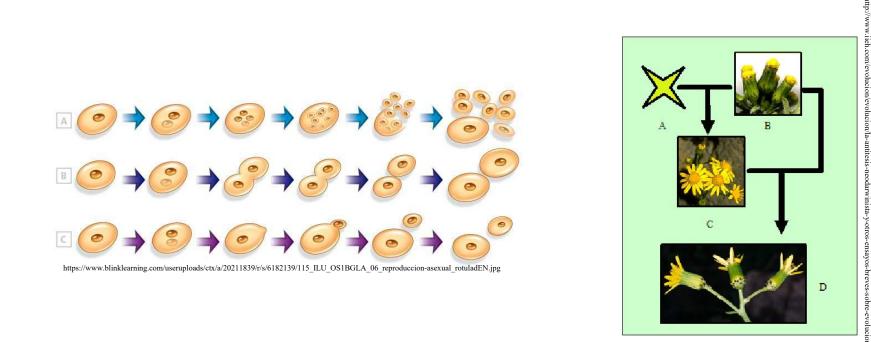


https://scielo.conicyt.cl/scielo.php?script=sci\_arttext&pid=S0718-686X2016000300005

Haplotypes of markers with fast evolving rates (such as microsatelites) using methods allowing networks reconstructions are the most commonly used in these cases

# **TOKOGENY & PHYLOGENY**

Despite the common rule of thumb, tokogenetic relationships may produce hierarchical structure (e.g., those among bacteria) and phylogenetic relationships may **not** necessarily result in hierarchical structure (e.g., if there is hybridization)

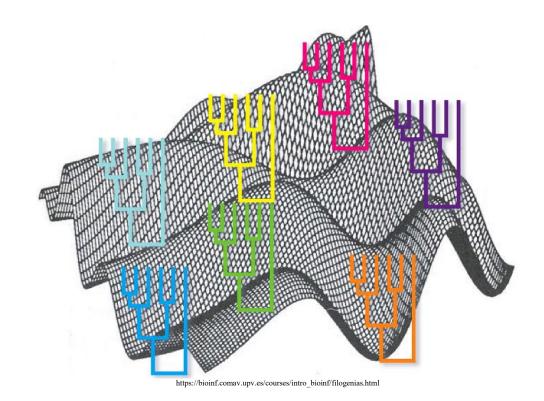


i.e., both, tokogeny and phylogeny, in both cases, the process and the pattern, may or may not be hierarchical

# Phylogenetic methods: TYPE OF TERMINALS

Phylogenetic methods group morphoclines assuming a hierarchical pattern:

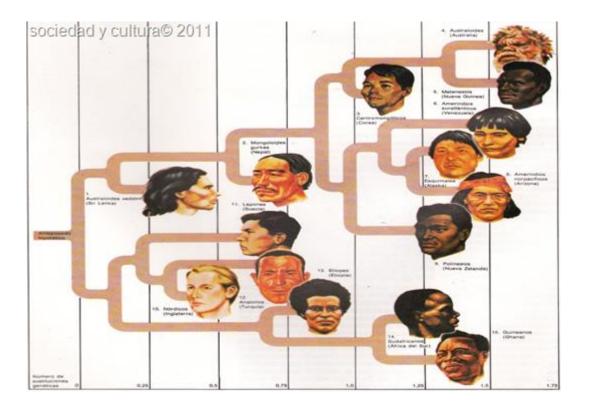
Parsimony explicitly groups morphoclines in such a way that the number of origins is minimized, but probabilistic methods (Likelihood or Bayesian), generally also result in resolution (hierarchical patterns), independently on the type of relationships that the terminals had (tokogenetic or phylogenetic)



# **TYPE OF TERMINALS**

Consequently, analyses that use units below species (intraspecific terminals, such as individuals) can also result in hierarchical patterns, even when the relationships among the terminals (the process) should not naturally produce hierarchies

How to interpret these patterns?

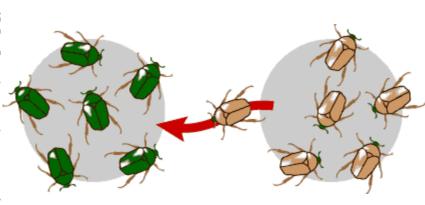


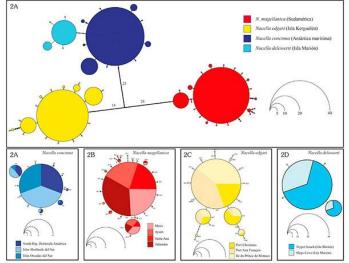
# **TYPE OF TERMINALS**

A hierarchical pattern results from the elimination of possible combinations among terminals during the evolution **or the sampling** 

# It can be caused by genetic drift, lack of recombination, small population sizes or INCOMPLET SAMPLING

Hence, the branches in a cladogram can be interpreted as total divergeny (phylogeny), genic flou, or partial divergeny





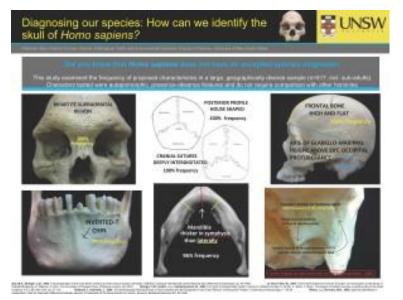
https://scielo.conicyt.cl/scielo.php?script=sci\_arttext&pid=S0718-686X2016000300005

# **SAMPLING DENSITY: How many units to include?**

### In general, sampling density afects more drastically the results when individuals are used as terminals than when species or supraspecific taxa are used (even when they are represented by individuals)

This is because species and supra-specific taxa are mainly or exclusively circumscribed using fixed characters, which therefore are present among the terminals independently of their sampling density

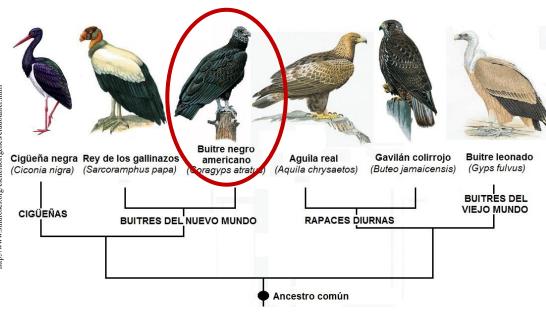




https://www.australianarchaeologicalassociation.com.au/gallery/diagnosing-our-species-how-can-we-identify-the-skull-of-homo-sapiens/

# **TYPE OF TERMINALS**

If a cladogram/tree is to be interpreted as a phylogeny, the terminal units that are included **or represented MUST be species or supraspecific taxa**, which normally have phylogenetic relationships resulting in **total diverging evolutionary units** producing hierarchical patterns



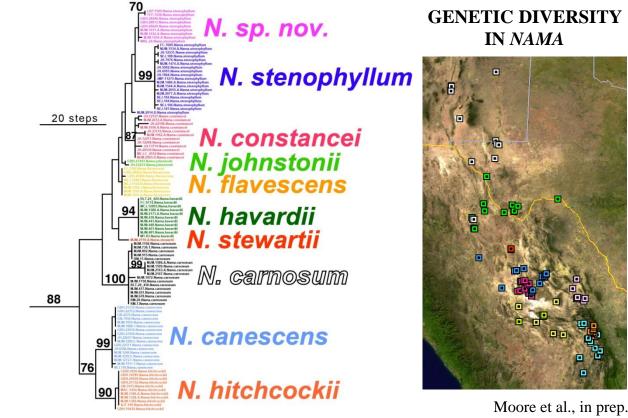


# **SAMPLING DENSITY: How many units to include?**

If the sampling of individuals and characters is dense enough, the

problem of using a phylogenetic method for non-fixed haplotypes can be overcome as synapomorphies should prevail to group individuals of the same species due to common ancestry

In this case, trees can be interpreted as phylogenies using individuals as terminal units



# **SPECIES CONCEPT**

To interpret a cladogram/tree as a phylogeny one must take into consideration (or imply) a species concept that is compatible with Hennig's principles

#### The Systematics Association Special Volume Series 54

ony whited Material



The units of biodiversity

EDITED OF M.F. Claridge, H.A. Dawah and M.R. Wilson

4

Morphologic species

### Nominalistic species

Biological species (Dobzhansky, 1935; Mayr, 1942)

# Ecological species (Van Valen, 1976): a species is a

set of organisms adapted to a particular set of resources, called a niche, in the environment. According to this concept, populations form the discrete phenetic clusters that we recognize as species because the ecological and evolutionary processes controlling how resources are divided up tend to produce those clusters.

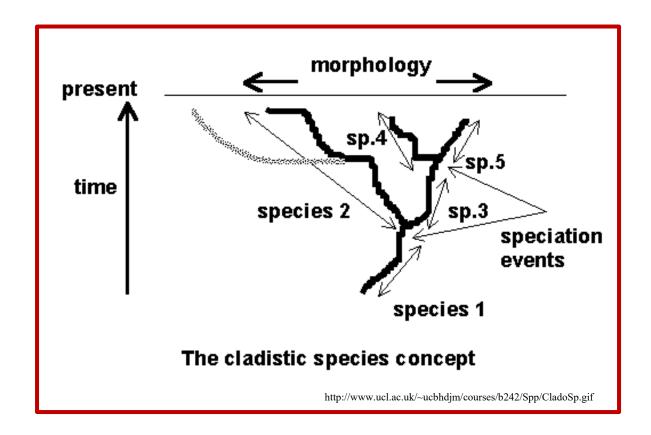
 $http://www.blackwellpublishing.com/ridley/a-z/Ecological\_species\_concept.asp$ 

# Evolutiary species (Wiley, 1978) Phylogenetic specie (Cracraft, 1989)

# **SPECIES CONCEPT**

### **GOOD NEWS: they (almost)all are**

Therefore the species concept normally has no major effect in the implementation of the phylogenetic method



#### **TYPE OF TERMINALS**

However, it must be understood that the study units (terminals) need to be defined before doing a phylogenetic analysis according to the objectives, because they will determine the interpretation of the results





https://www.ecologiaverde.com/se-extinguen-150-especies-animales-por-dia-3.html

### **TWO SETS OF TERMINALS**

After deciding which hierarchical level is appropriate for the question we want to answer (individuals, populations, species, etc.), the second important research design has to do with the what we call the ingroup and the outgroup



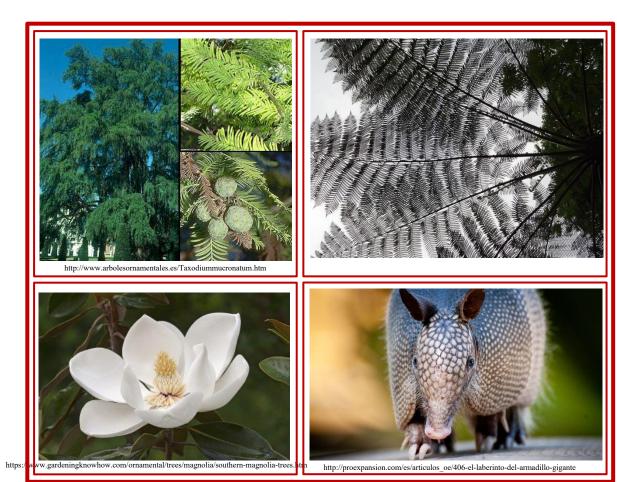


#### **OUTGROUP:** definition

## An outgroup is the taxon o taxa that does not belong to the group of immediate interes

e.g., the outgroup in an analysis of species from the genus *Pinus* could be *Taxodium*, a Magnolia species, a fern, an animal or simultaneously all of the above

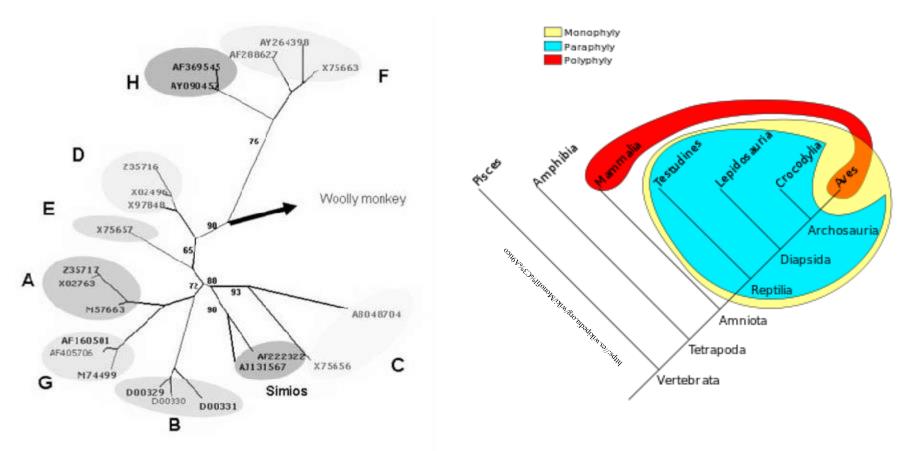




## **OUTGROUP: objective**

The outgroup in phylogenetic analysis fulfils two main objectives:

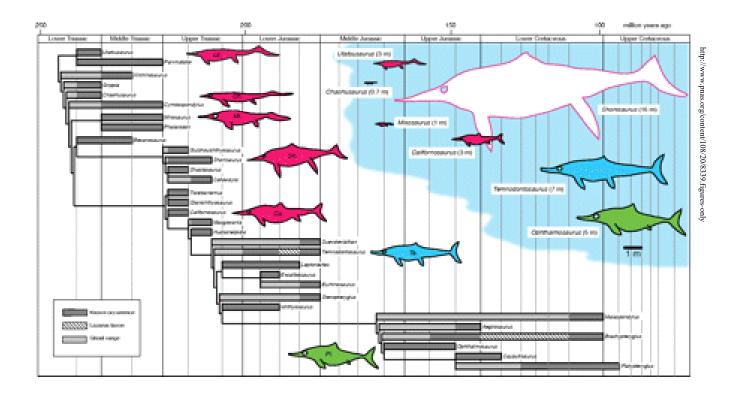
#### -- To orient (root) the tree -- To test the monophyly of the group of interest

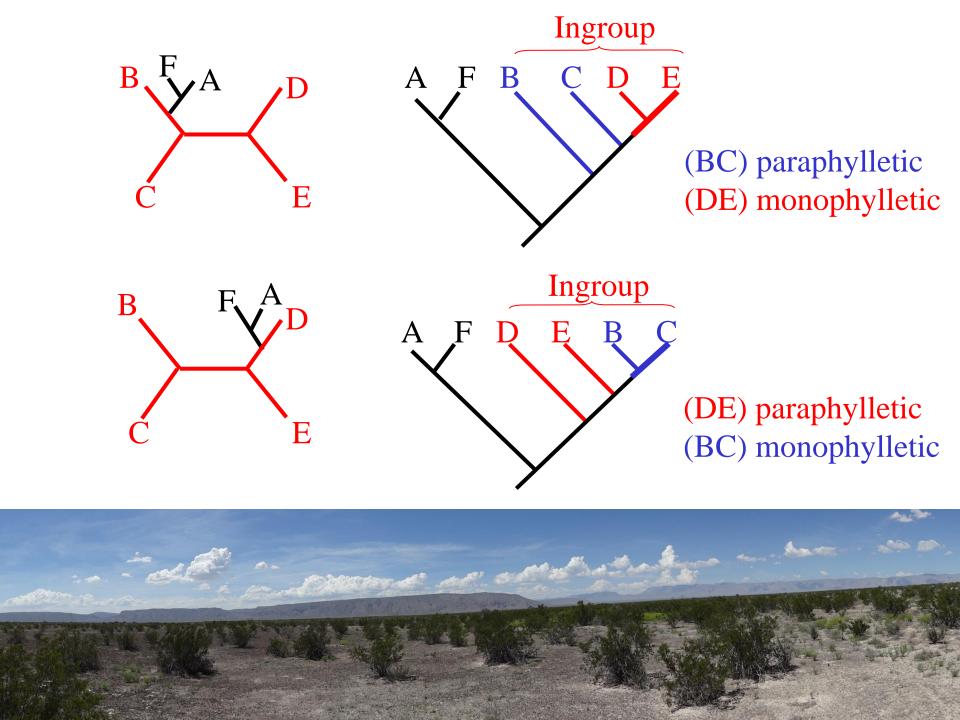


https://www.researchgate.net/publication/234138763\_Rol\_of\_Hepatitis\_B\_virus\_genotype\_ and\_precorecore\_variantson\_antiviral\_treatment\_response\_and\_clinical\_outcome/figures

## **OUTGROUP:** function

The orientation of the tree is totally dependent upon the placement of the outgroup in reference to the taxa included in the ingroup and determines the internal structure of the ingroup, defining for instance the monophylly or paraphylly of the included taxa

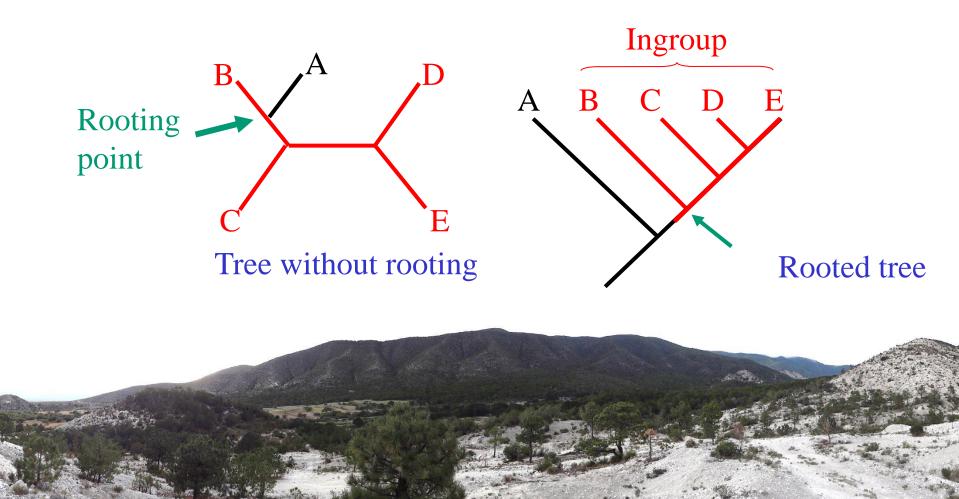




#### **OUTGROUP:** function

If the analysis includes **only a single terminal as outgroup**, the tree will be arbitrarily oriented (rooted) between it and the ingroup

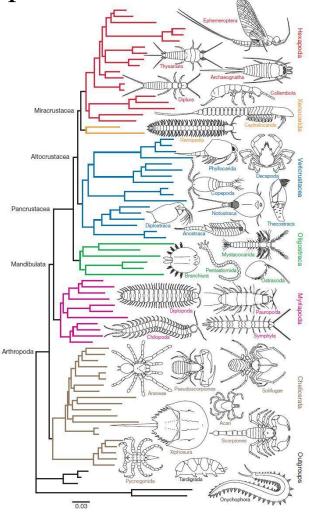
Consequently, the monophylly of the ingroup will not be tested



#### **OUTGROUPS & monophylly**

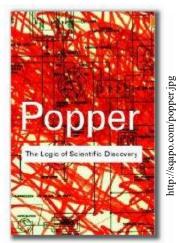
#### The monophylly hypothesis of the ingroup **CAN ONLY be tested if more than one terminal (taxon) is simultaneously included** in the analysis as outgroup

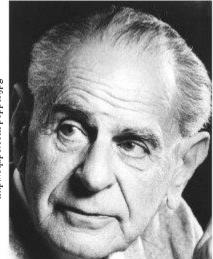
To do this, it is necessary to include ALL the members of the outgroup in the matrix together with those of the ingroup (as if there was no difference among them); characters most be gathered and codified for all terminals included in the matrix



http://brendaortunoartnoinsc.blogspot.mx/2013/

#### **OUTGROUP: the best selection seeks for rigor**





According to Popper's philosophy, our confidence in the results will increase if we try to reject the hypothesis and still it is accepted





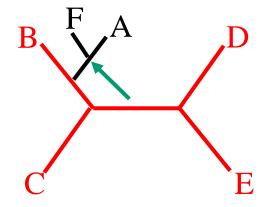


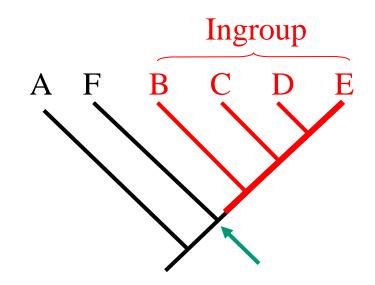




http://proexpansion.com/es/articulos\_oe/406-el-laberinto-del-armadillo-gigante https://www.gardeningknowhow.com/ornamental/trees/magnolia/southern-magnolia-trees.htm

#### Monophyletic ingroup





T. Gamble et al. • New trans-Atlantic gecko clade

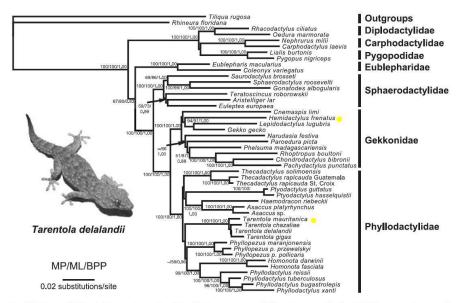
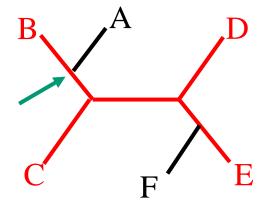
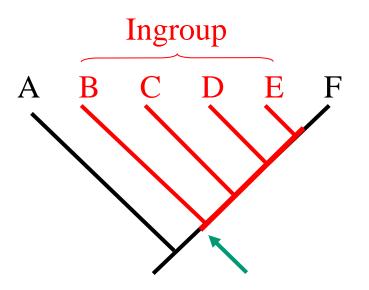


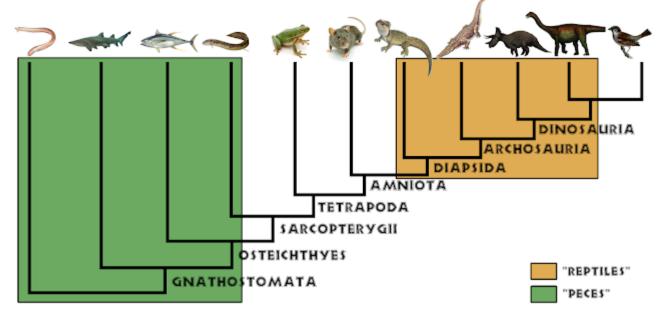
Fig. 2 Maximum likelihood phylogeny for combined data. Maximum parsimony (MP) and maximum likelihood (ML) bootstrap values as well as Bayesian posterior probabilities (BPP) are indicated. Clade names are shown on the right. Photo by T. Gamble.

http://4.bp.blogspot.com/-kVjWnCjL6YQ/TfDusGd5uyI/AAAAAAAAmg/FuXbEbxcgsE/s1600/Clado%2Bmonofiletico%2Bgecos.jpg

Paraphyletic ingroup





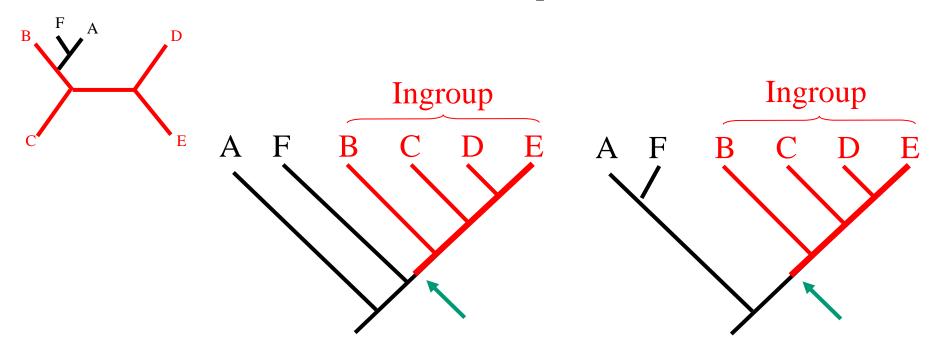


http://dinosaurrenaissance.blogspot.mx/2013/04/tutorial-2-filogenia-basica-de.html

## **OUTGROUP:** function

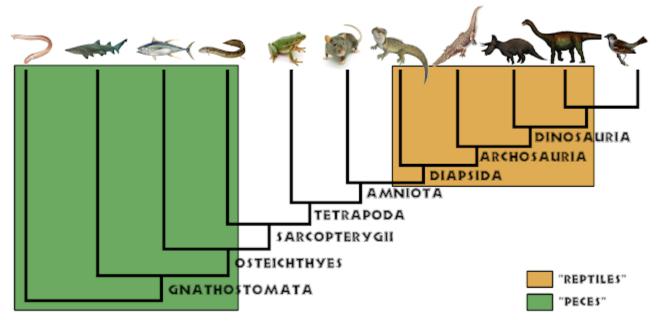
If in the analyses more than one taxon is used as outgroup, the tree will be arbitrarelly rooted between one of them and the ingroup (this is what I call the functional outgroup)

If the ingroup is corroborated as monophyletic, the rooting decision (merely a graphic decision) will not impact the conclusion of monophyly of the ingroup, but it can impact the conclusion about the sister relationships



#### **OUTGROUP**

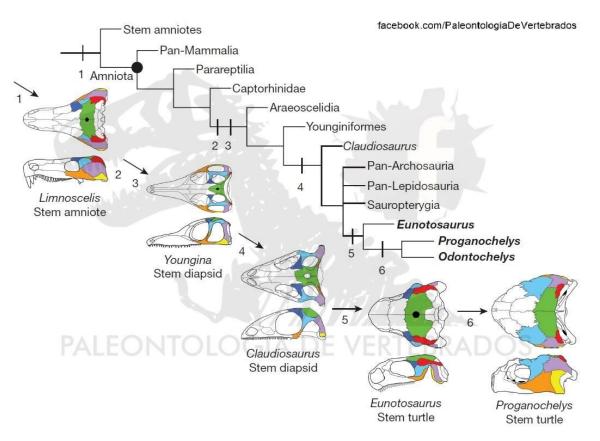
So, when more than one taxon is used as an outgroup, the one that serves as rooting reference has to be chosen arbitrarily, but not necessarily unjustifiably: it should be the theoretically most distantly related taxon



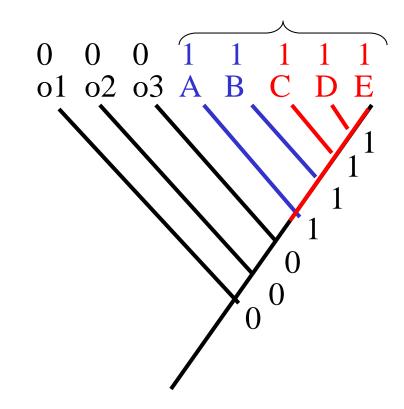
http://dinosaurrenaissance.blogspot.mx/2013/04/tutorial-2-filogenia-basica-de.html

#### **OUTGROUP: character polarity**

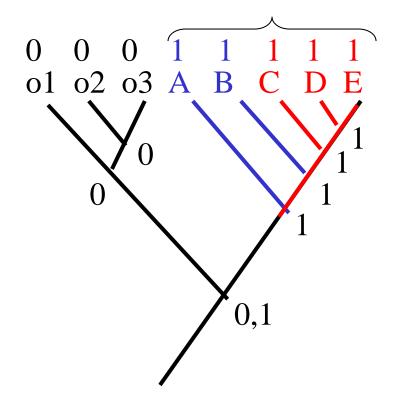
#### Also the interpretation of character evolution can change depending on the orientation of the tree (on the relative position of the out group)



## **OUTGROUPS: character polarity**



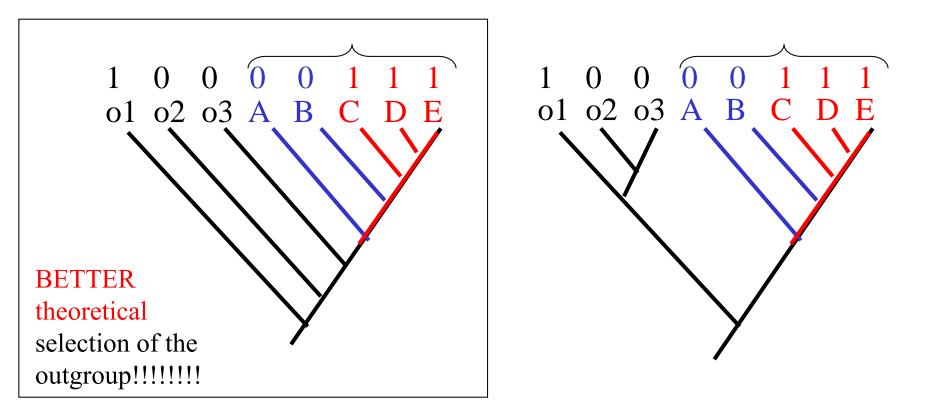
The ancestral character state for the ingroup would be interpreted to be 0



The ancestral character state for the ingroup could be either 0 or 1

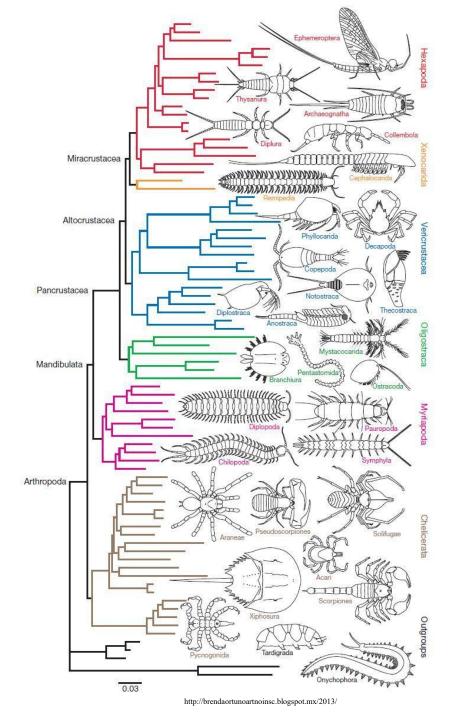
# **OUTGROUP:** selection and graphic representation

A diverse sampling of the outgroup generally results in a paraphiletic group, with a gradient from the ingroup to a more distantly related taxon the functions to roote the tree



# OUTGROUP: monophylly

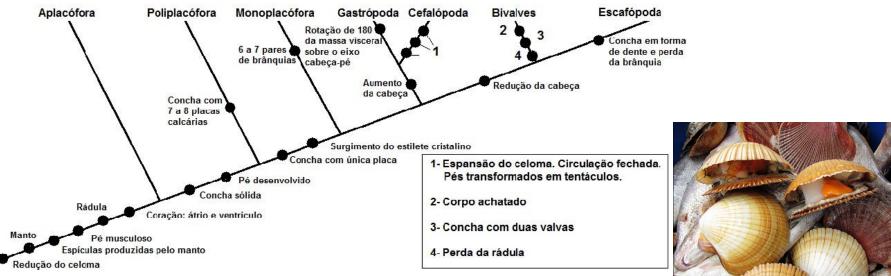
The more diverse the representation of taxa (outgroup and ingroup) the more rigor in the monophylly test for the ingroup and the groups inside it



# **OUTGROUP: selection**

There is no universal rule to appropriatelly select the outgroup because this depends upon each group, including type and quantity required to produce stable results

The best way to design the sampling of appropriate outgroups is to postulate hypothetical SINAPOMORPHIES that are share by the members of the ingroup (character states present in the ingroup and absent in the outgroup) plus hypothetical synapomorphies shared by the ingroup and some of the outgroup members, in such a way that there is a theoretical hirarchy of synapomorphies and terminals



https://zoologia-ii-ufes-turma-i.webnode.com/products/mollusca2/

# **OUTGROUP: selection**

For a good design of outgroup sampling one can use as preliminary information CLASSIFICATIONS, PREVIOUS ANALISES, or PUTATIVE SINAPOMORPHIES that can be postulated according to the hypothesis of monophyly of the ingroup

It is very important to *a priori* select one terminal as the one that unequivocally is the most distantly related and that will serve to root the tree



https://es.wikipedia.org/wiki/Parides

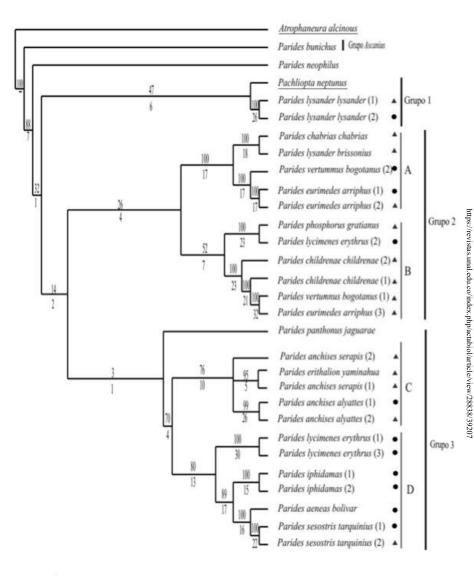


Figura 3. Árbol filogenético según análisis Bayesiano. Los valores de las ramas indican la probabilidad posterior Bayesiana. Las especies del grupo externo se muestran subrayadas.

#### **TERMINAL SAMPLING**

There is no rule to predict how many terminals fulfil the requirements for the interpretation of a tree as a phylogeny and for the correct retrieval of natural relationships

The objective is to obtain results that reflect phylogenetic relationships that we believe are natural and that the results are stable

For that, it is desirable to include in a phylogenetic analysis as much comparative information as possible depending upon availability (at the level of terminals, but also characters)

| CLADISTICS                                  |  |
|---|--|
| Study units (terminals)                     | Populations (becoming more common), species or higher hierarchical levels (genera, family, etc.)   |
| Number of characters                        | As many as possible  |
| Type of characters                          | Qualitative and/or Quantitative (now possibe)  |
| Similitud                                   | Special: reflecting common ancester (homology)   |
| Method                                      | Parsimony (not assuming more explanations than those that are strictly necessary to explain the observations) or probabilistic (applyin evolutionary models) |
| Importance of the degree if diversification | Irrelevant   |
| <b>Evolutionary interpretation</b>          | Not necesary   |
| Main component in evolution                 | Cladogenesis   |
| Ancestors                                   | Hypothetic   |
| Homology                                    | Hypothetizizd to begin (Primary) followed by a test  |
| Diagram                                     | Cladogram or phylogram (with branch lenghts)   |
| Direction of evolutionary changes           | A posteriori   |
| Groups accepted in the clasification        | Monophyletic   |
| Non inheritable features (e.g., geography)  | A posteriori   |