

Michel C. Milinkovitch
Dpt. Genetics & Evolution,
University of Geneva
Michel.Milinkovitch@unige.ch

www.lanevol.org

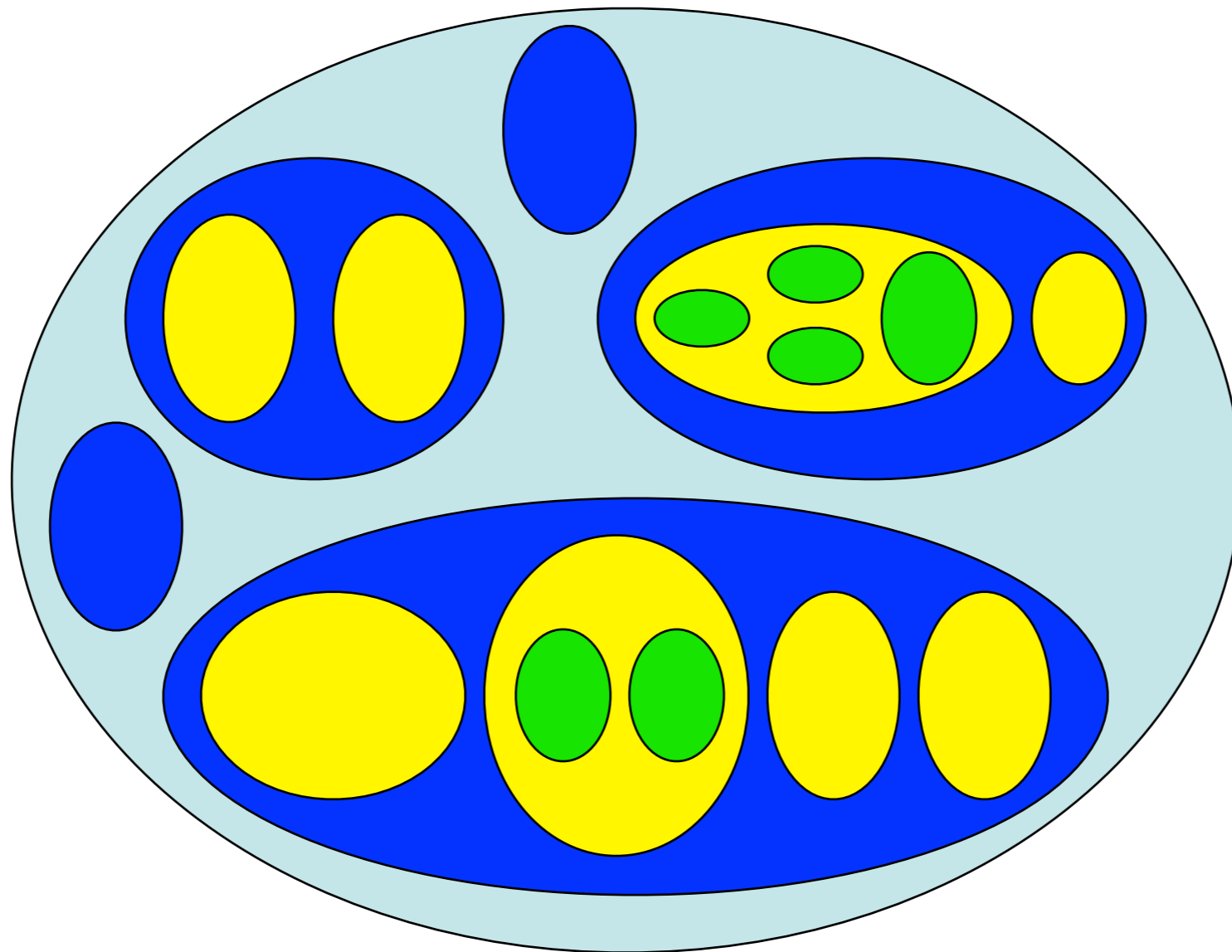


Laboratory of
Artificial & Natural
Evolution



Since Platon & Aristotile

Taxonomy (classification)



Linné (1758): binomial nomenclature



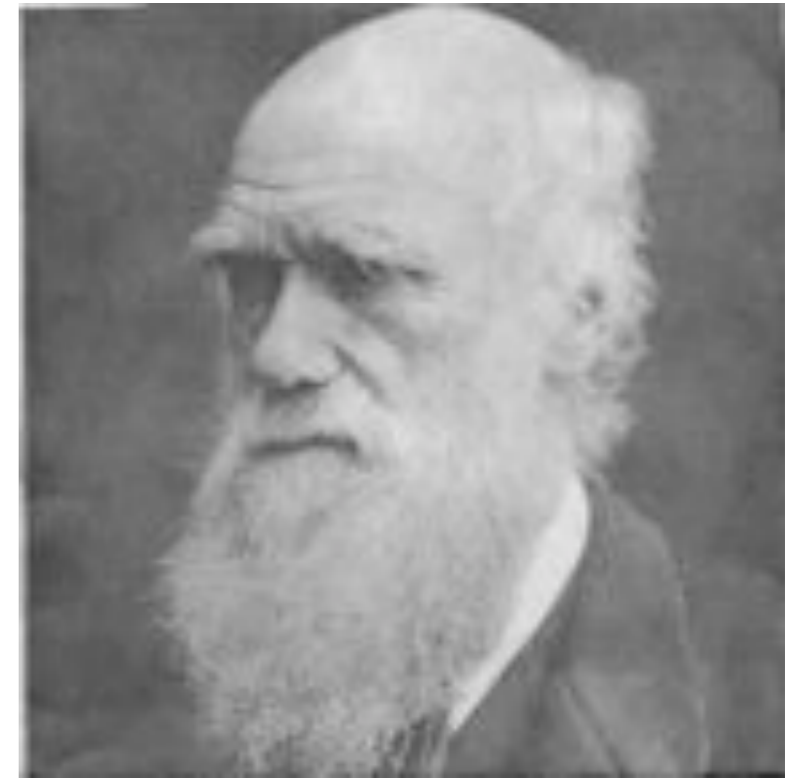
Kingdom	-	Animalia
Phylum	-	Chordata
Subphylum	-	Vertebrata
Class	-	Mammalia
Order	-	Carnivora
Family	-	Ursidae
Gender	-	Ursus
Species	-	U. maritimus

Since Darwin (1859)

1. Variation & heritability,
2. Natural selection,
3. Phylogeny.

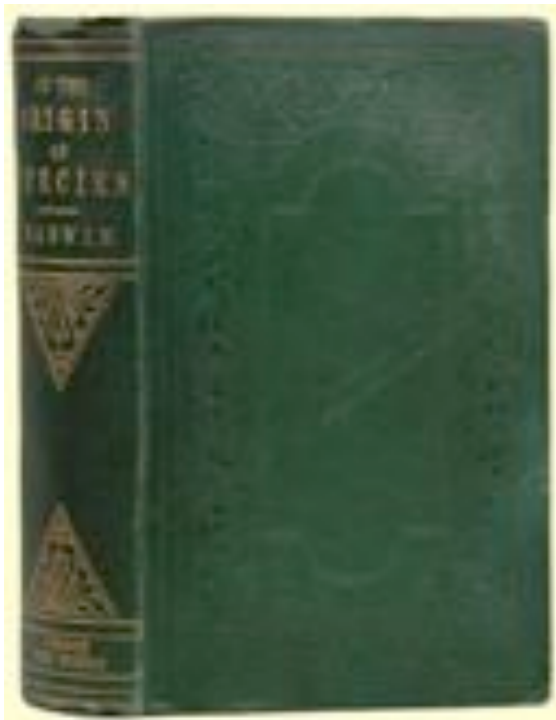
Since Darwin (1859)

1. Variation & heritability,
2. Natural selection,
3. Phylogeny.



**On the origin of species by means of
Natural Selection or the Preservation of
Favoured Races in the Struggle for Life**

(Charles Darwin, 1859)



Since Darwin (1859)

1. Variation & heritability,
2. Natural selection,
3. Phylogeny.

- ✓ There is variation among individuals within a species
- ✓ This variation is, partly, heritable
- ✓ Individuals of generation n have unequal contributions to generation $n+1$ (resources are limited; bearers of advantageous variation contribute more to the next generation).

Since Mendel (1856-1871 -> 1900)



Gregor Mendel
(1822-1884)



F1



100%



F2



25%

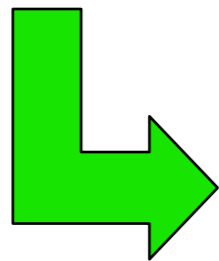
50%

25%

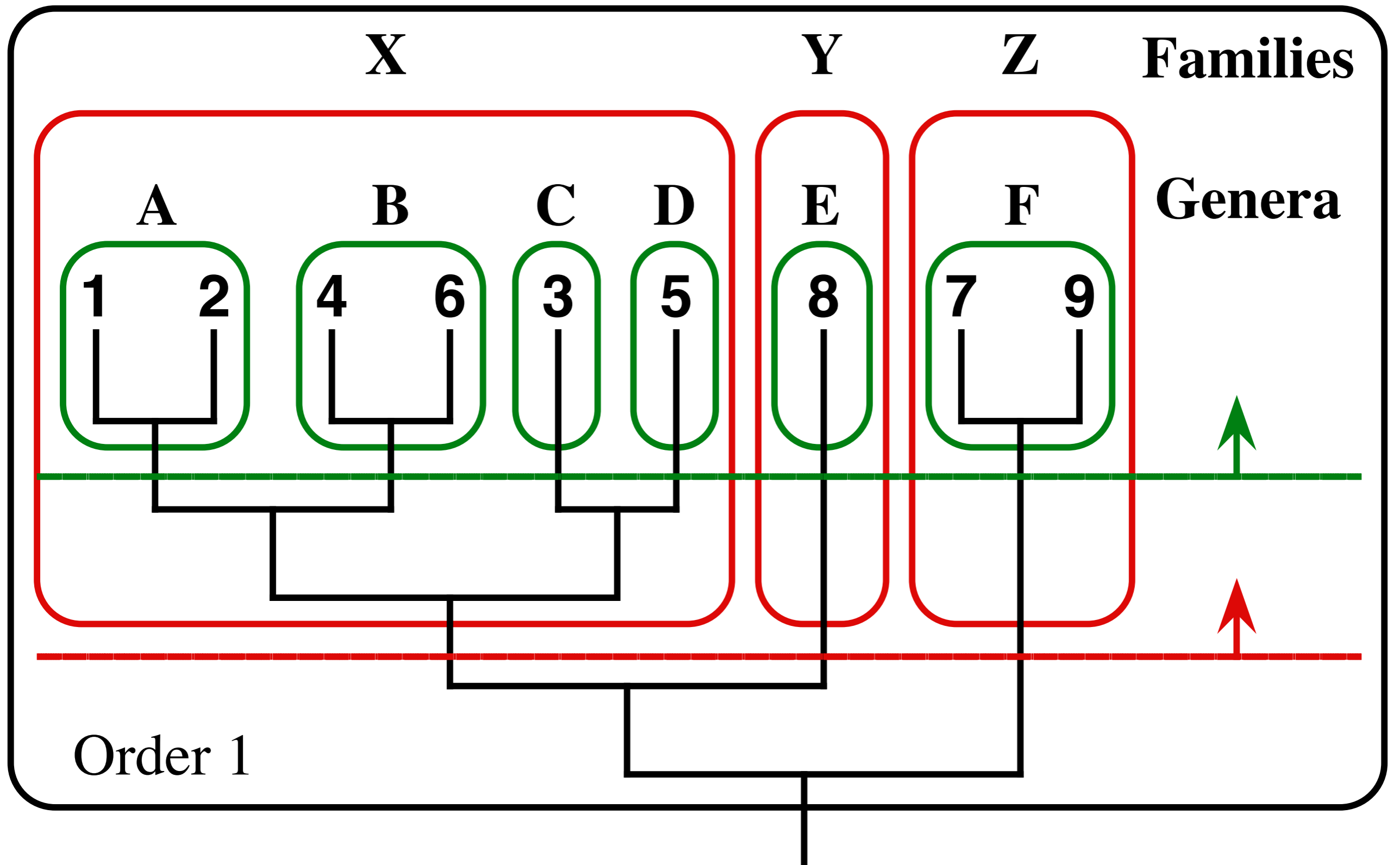
Since Darwin (1859)

1. Variation & heritability,
2. Natural selection,
3. Phylogeny.

The hierarchical organization of biological diversity results from modification with descent



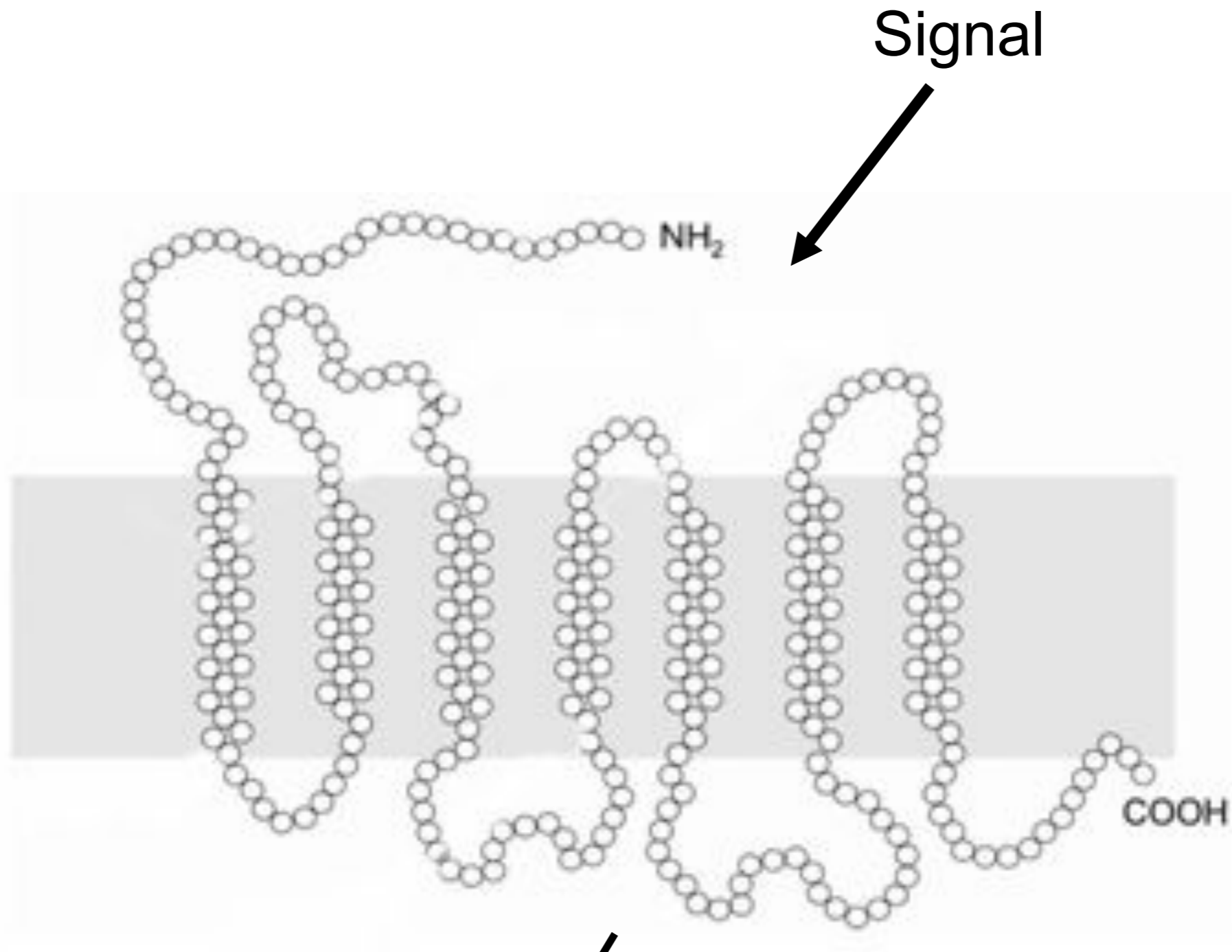
The Linnean system should be strictly genealogical



The origin of variation: mutations

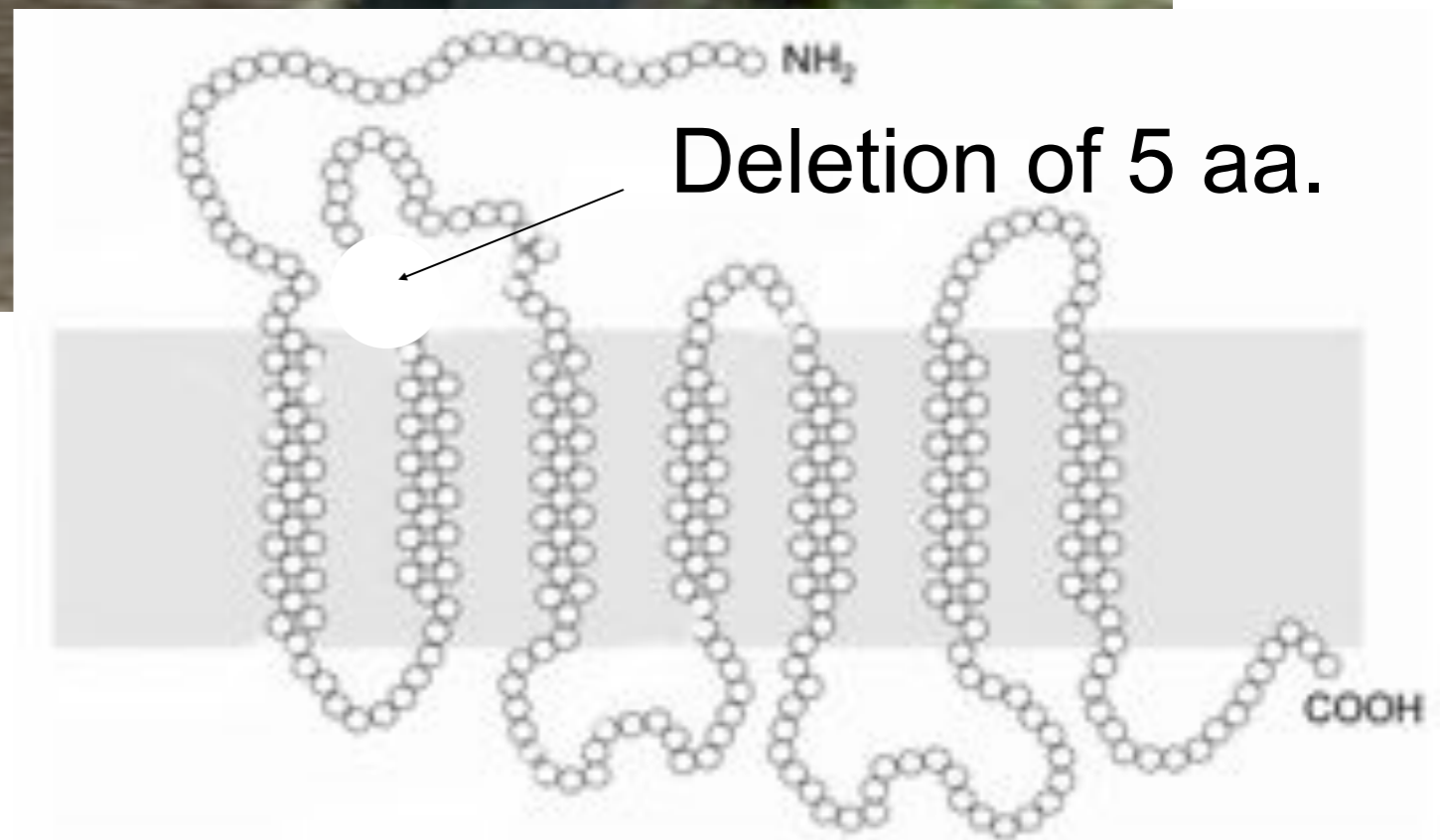


Example: the MC1R



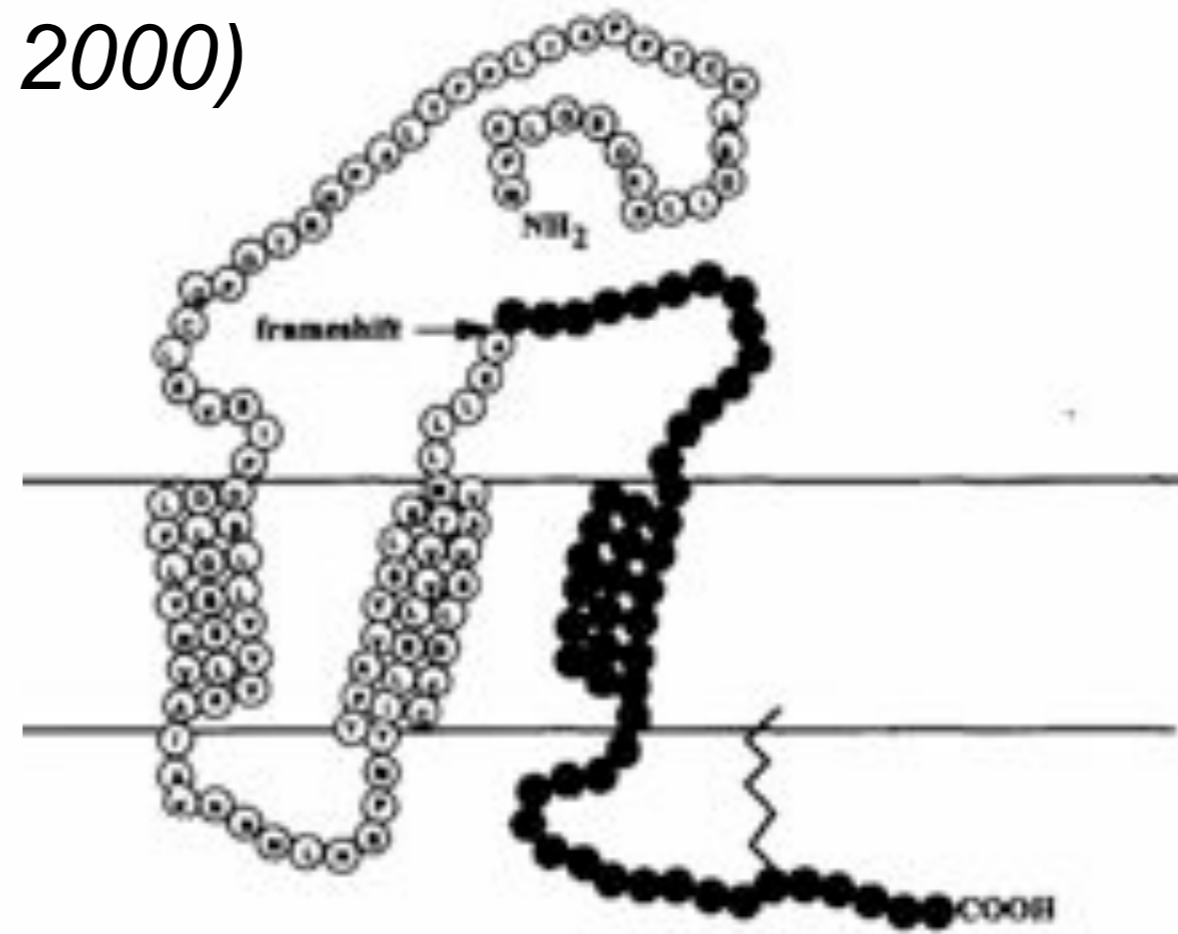
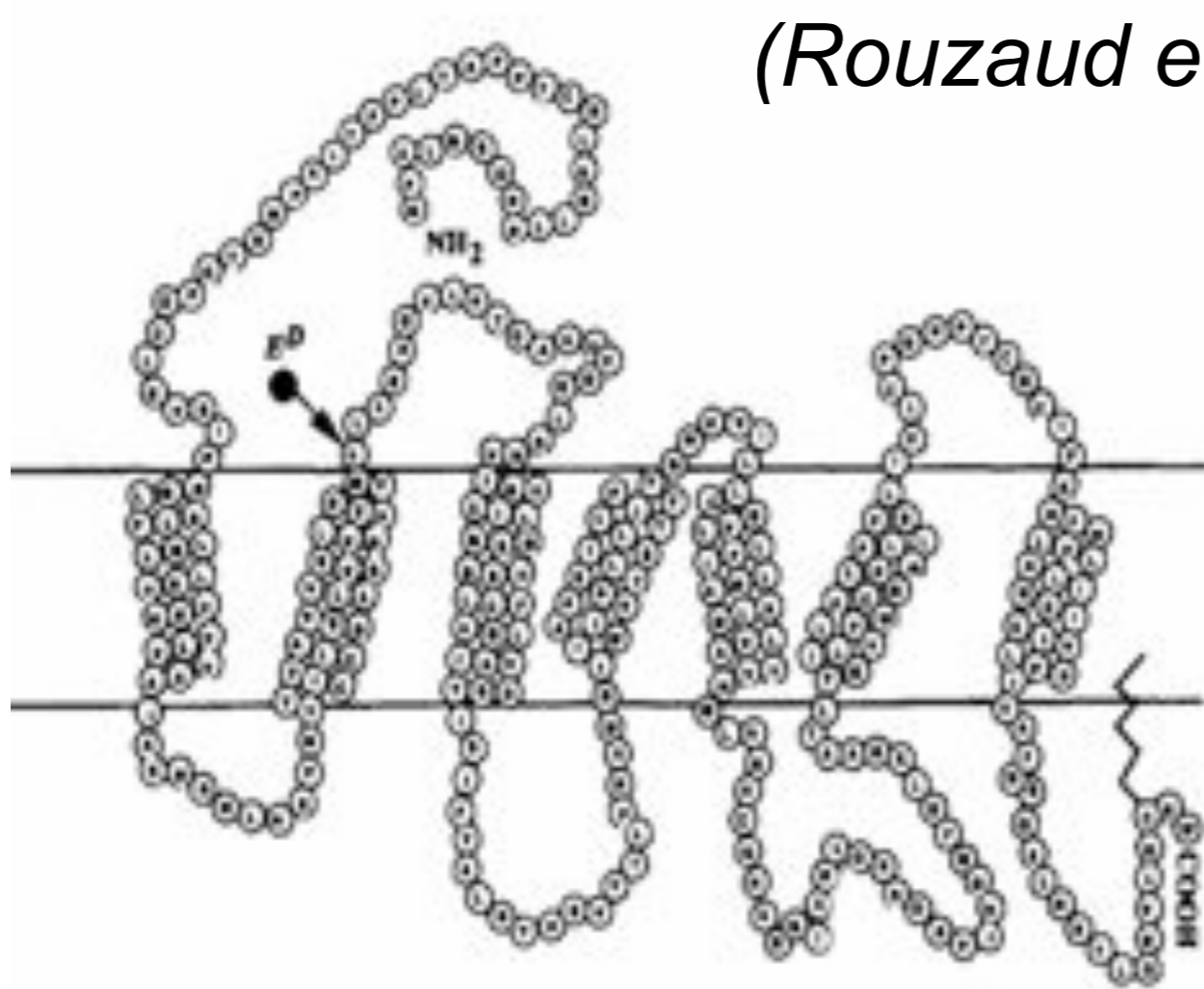
Eumelanin

Phaeomelanin

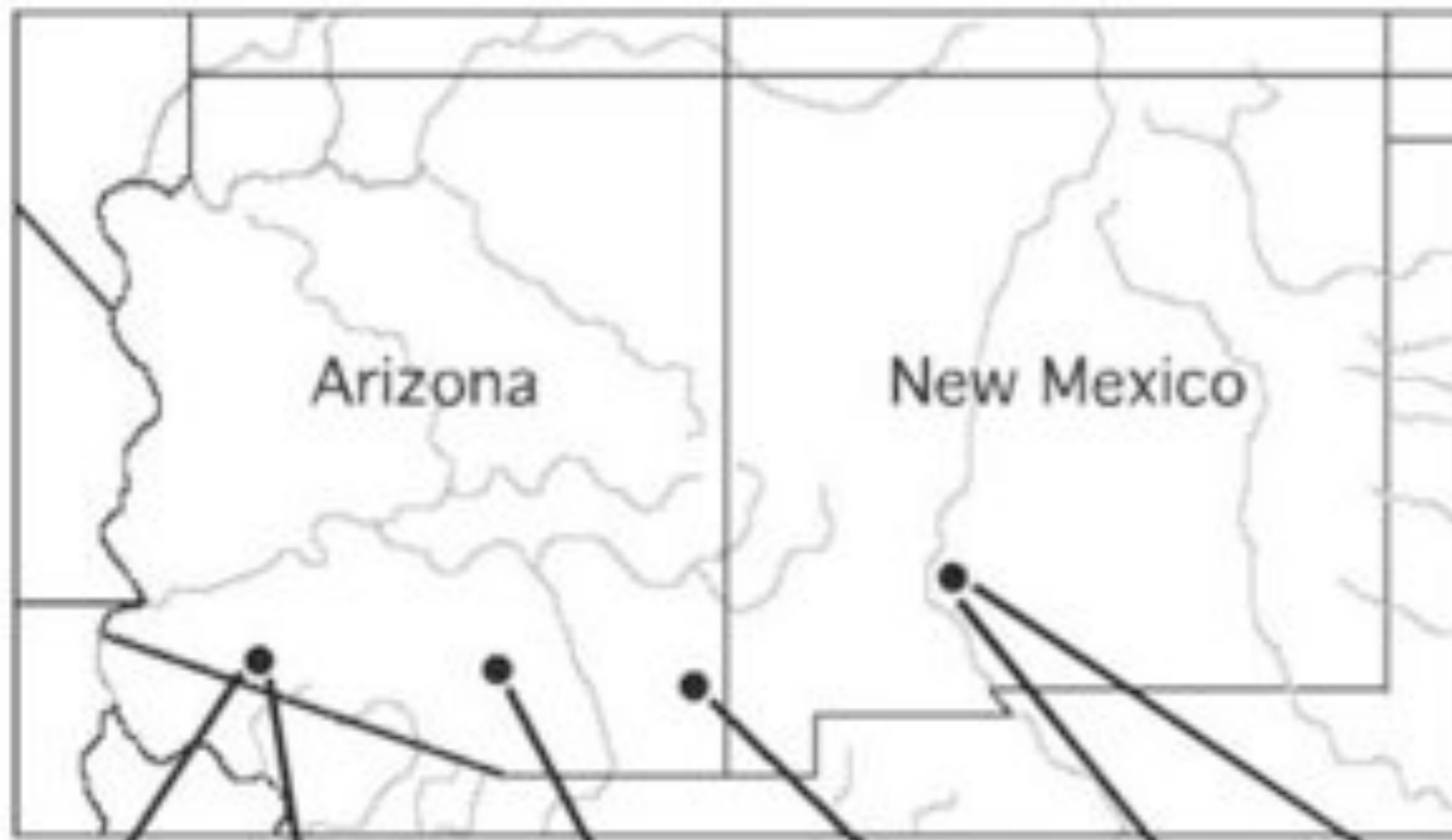


(Eizirik et al., 2003)

(Rouzaud et al., 2000)



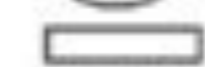
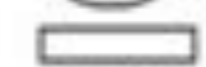
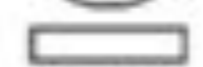
Natural selection



mouse color



rock color



N = 18

N = 11

N = 15

N = 5

N = 12

N = 8

Pinacate

Avra Valley

Portal

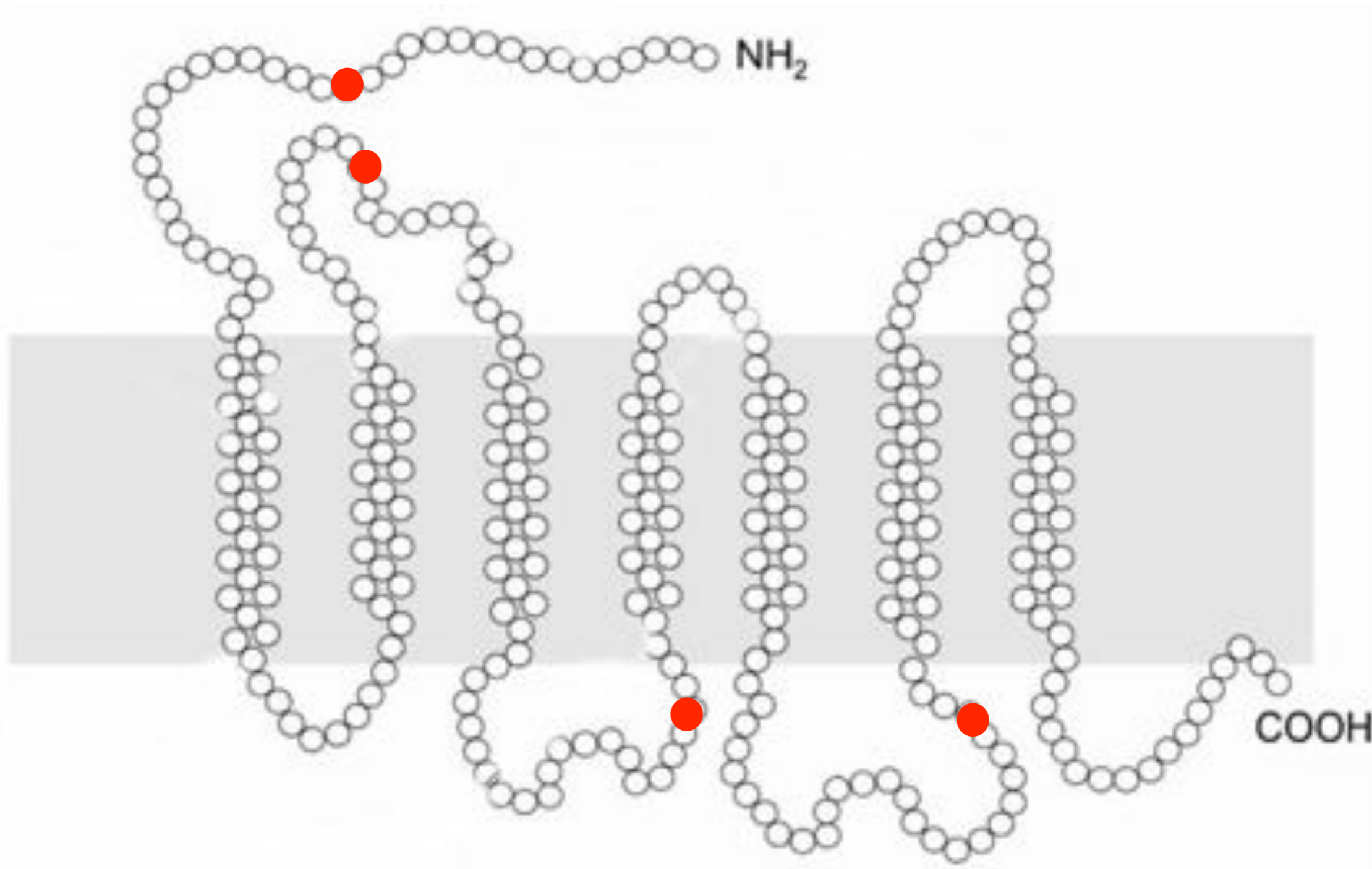
Armendaris



(Nachman et al.,
2003)



Some of the identified mutations in vertebrate MC1R



(Mundy, 2005)

Phylogeny "reconstruction" is an inference



desktop by Mark Seitz from MacDesktop.com

METRO-GOLDWYN-MAYER PRESENTS THE SPECTACULAR GEORGE PAL PRODUCTION BY H. G. WELLS'

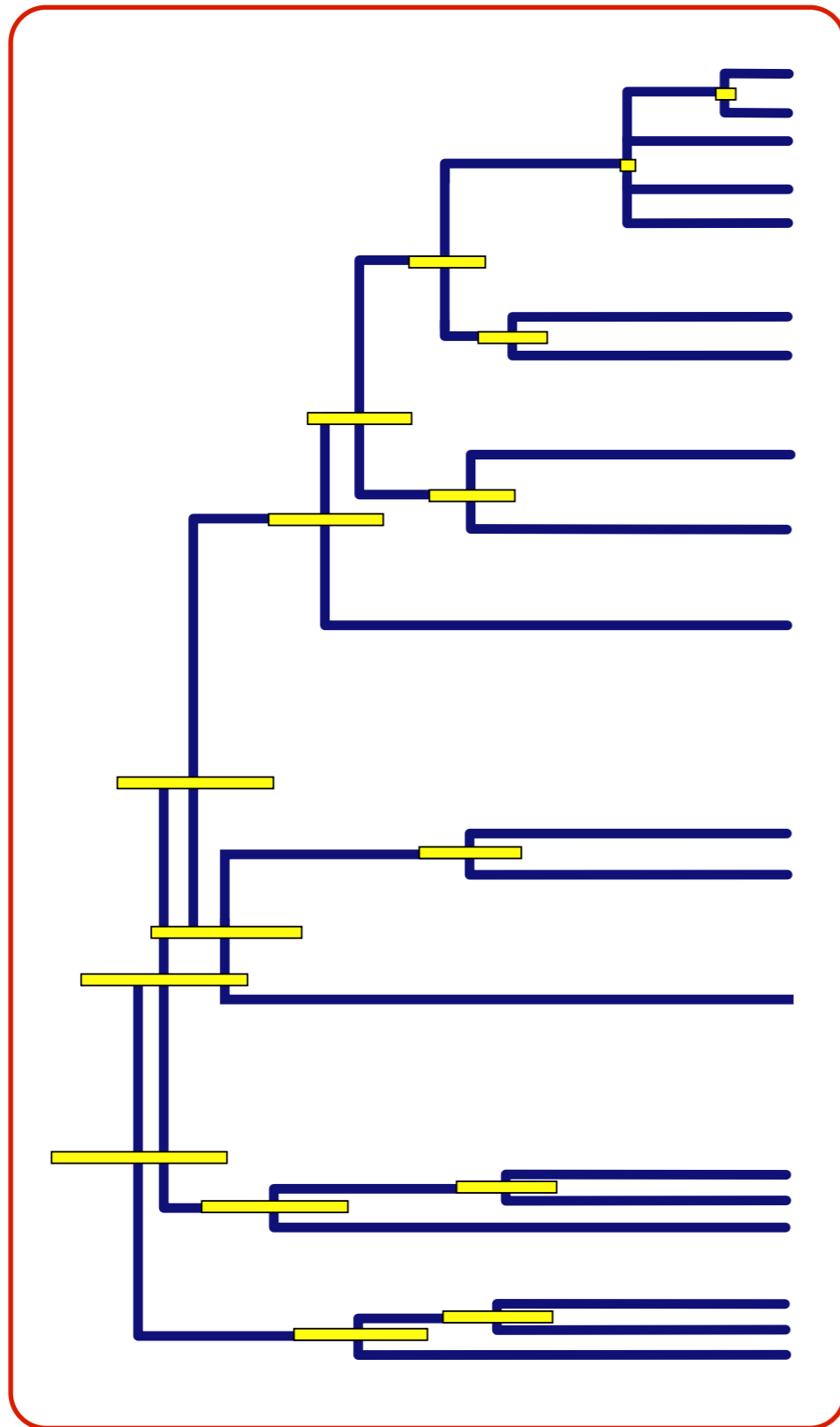


STARRING ROD TAYLOR · ALAN YOUNG · YVETTE MINIEUX · SEBASTIAN CABOT · TOM HELMORE

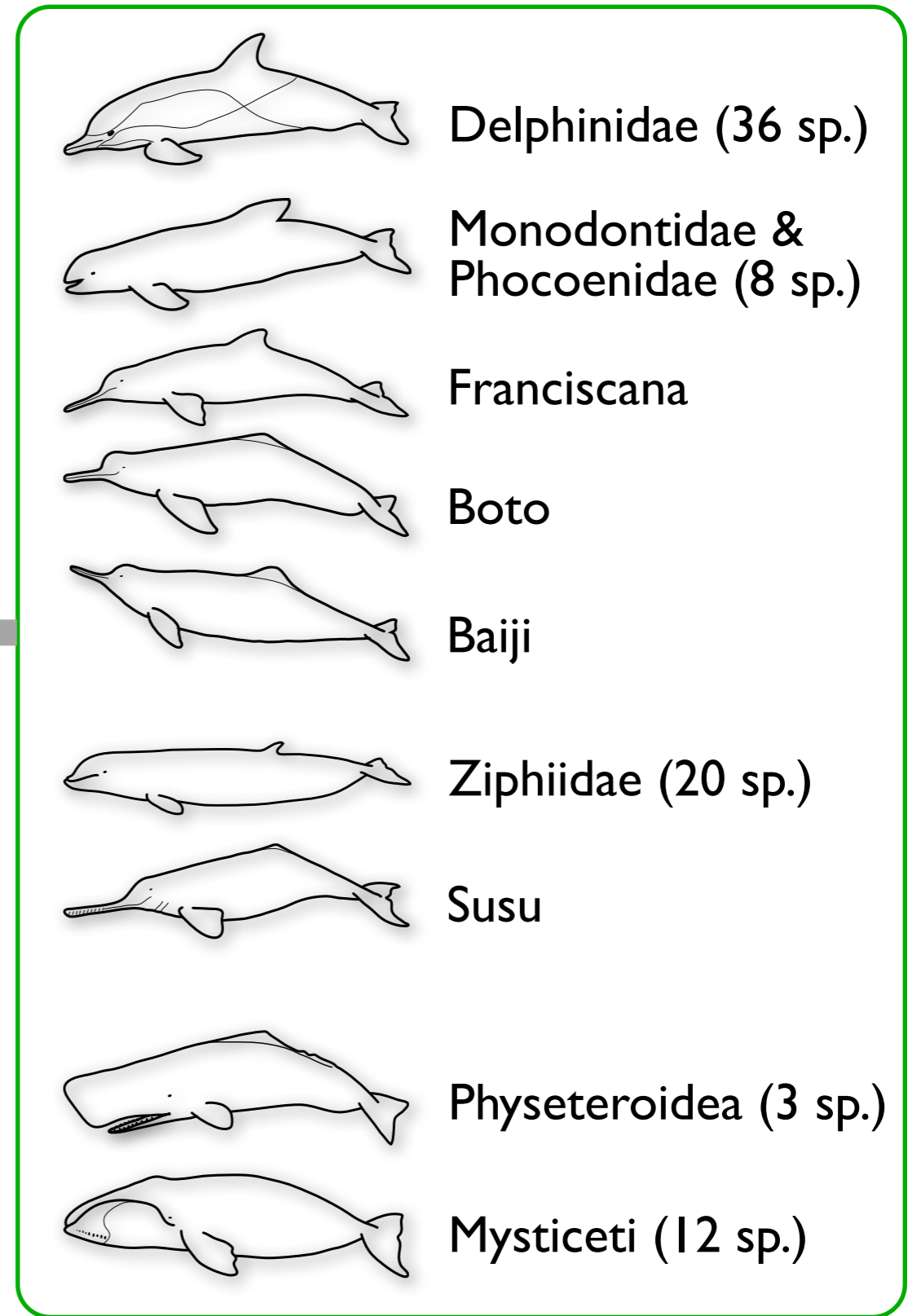
SCREENPLAY BY DAVID DONICAN BASED UPON THE NOVEL BY H. G. WELLS DIRECTED BY GEORGE PAL



50 40 30 20 10 5 0 mya

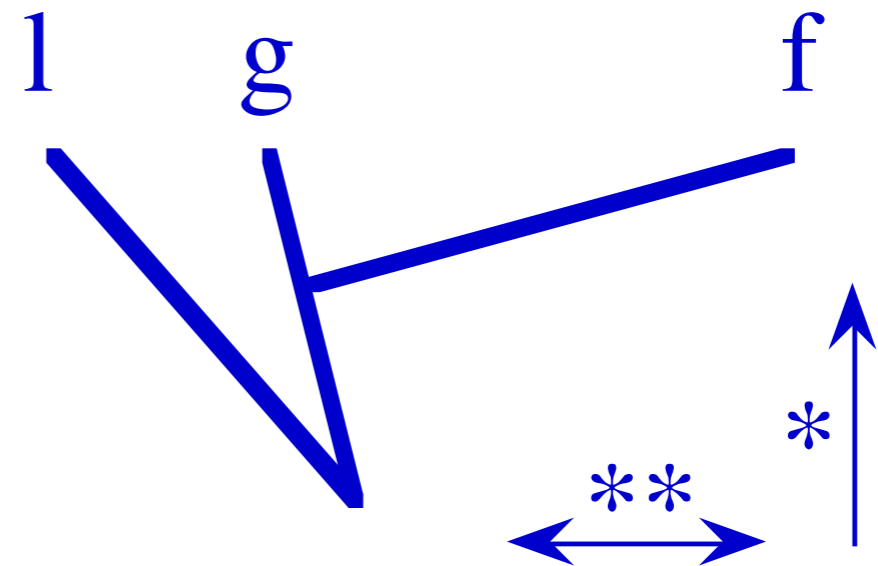
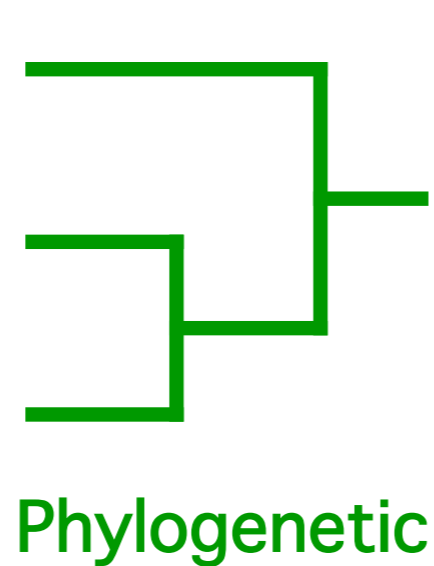
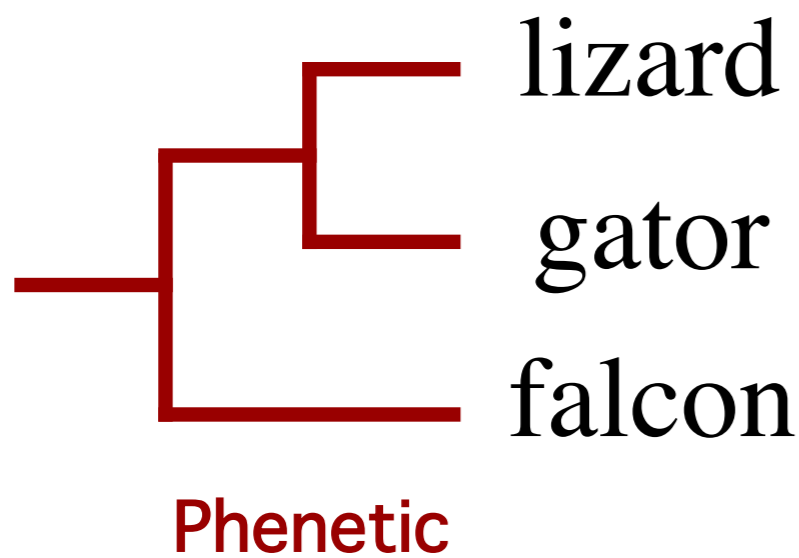


Inferred (topology,
BrL, age of nodes)



Observed
(Data)

Phylogenetic relationships among lineages can be counter-intuitive



What is Phylogeny ?

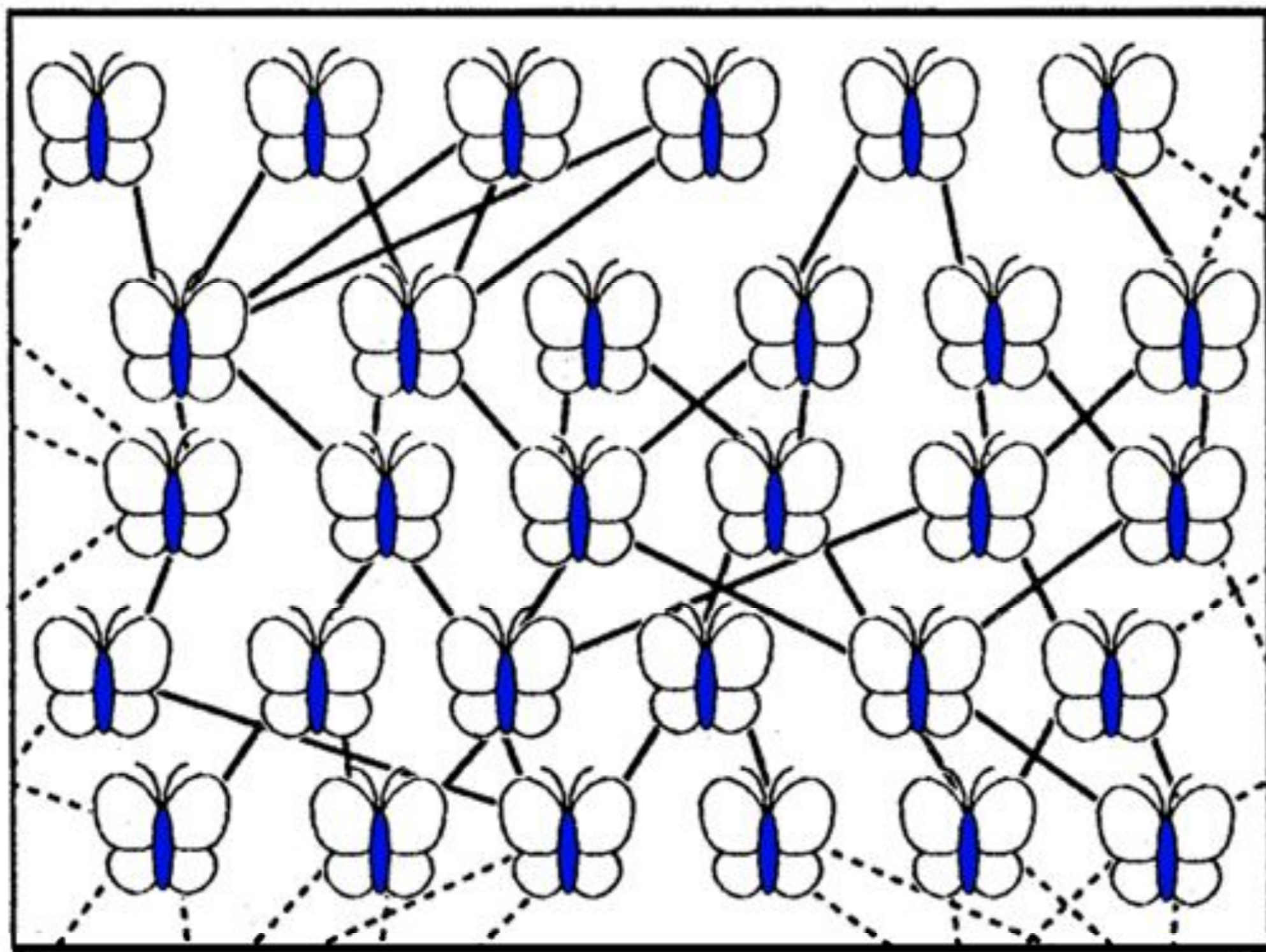
The evolutionary relationships
among organisms;

the patterns of lineage branching
produced by the true evolutionary
history of the organisms
being considered.

The Tree of Life then represents the history of all organismal lineages as they change through time ...



individus



1997

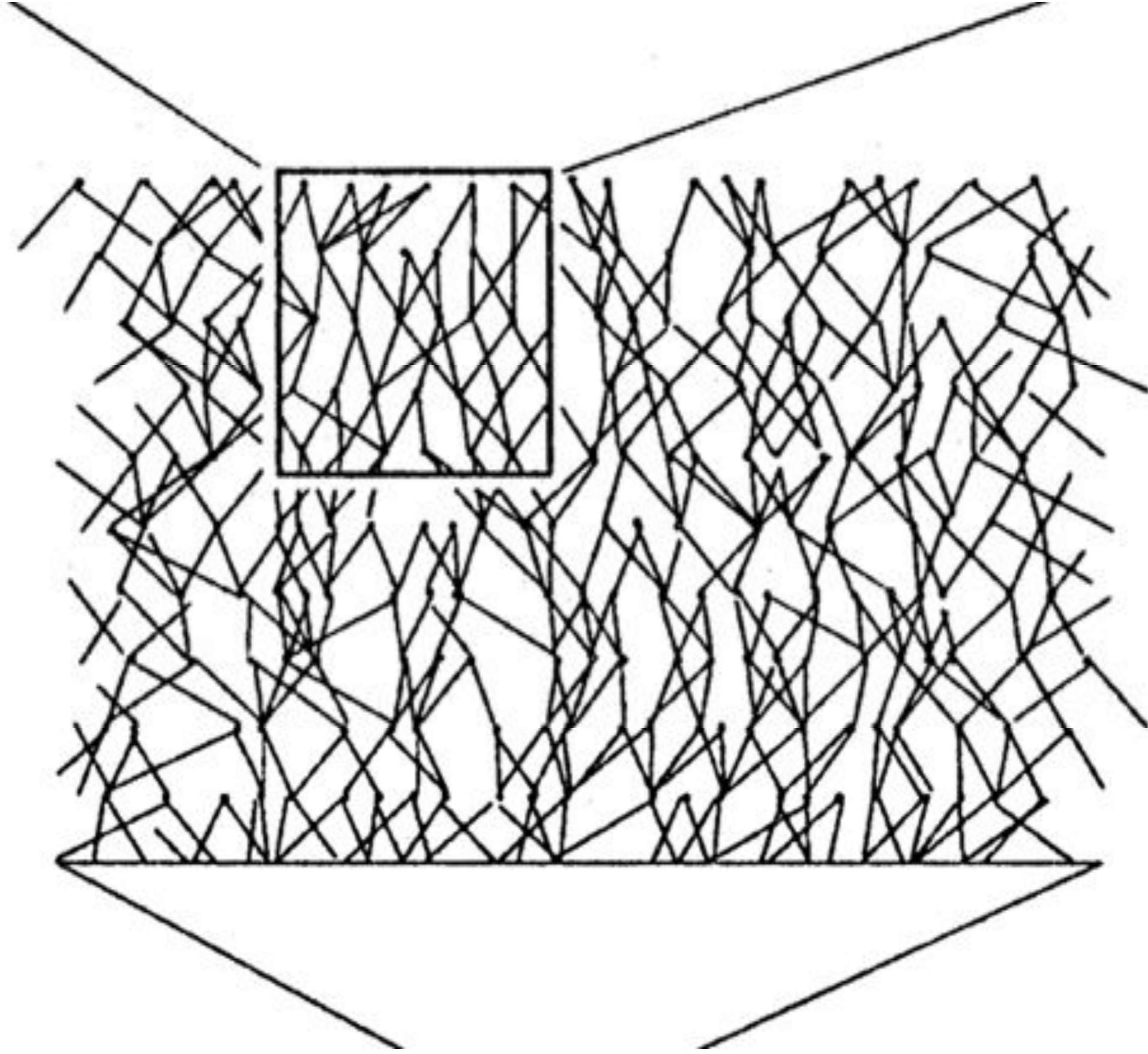
1996

1995

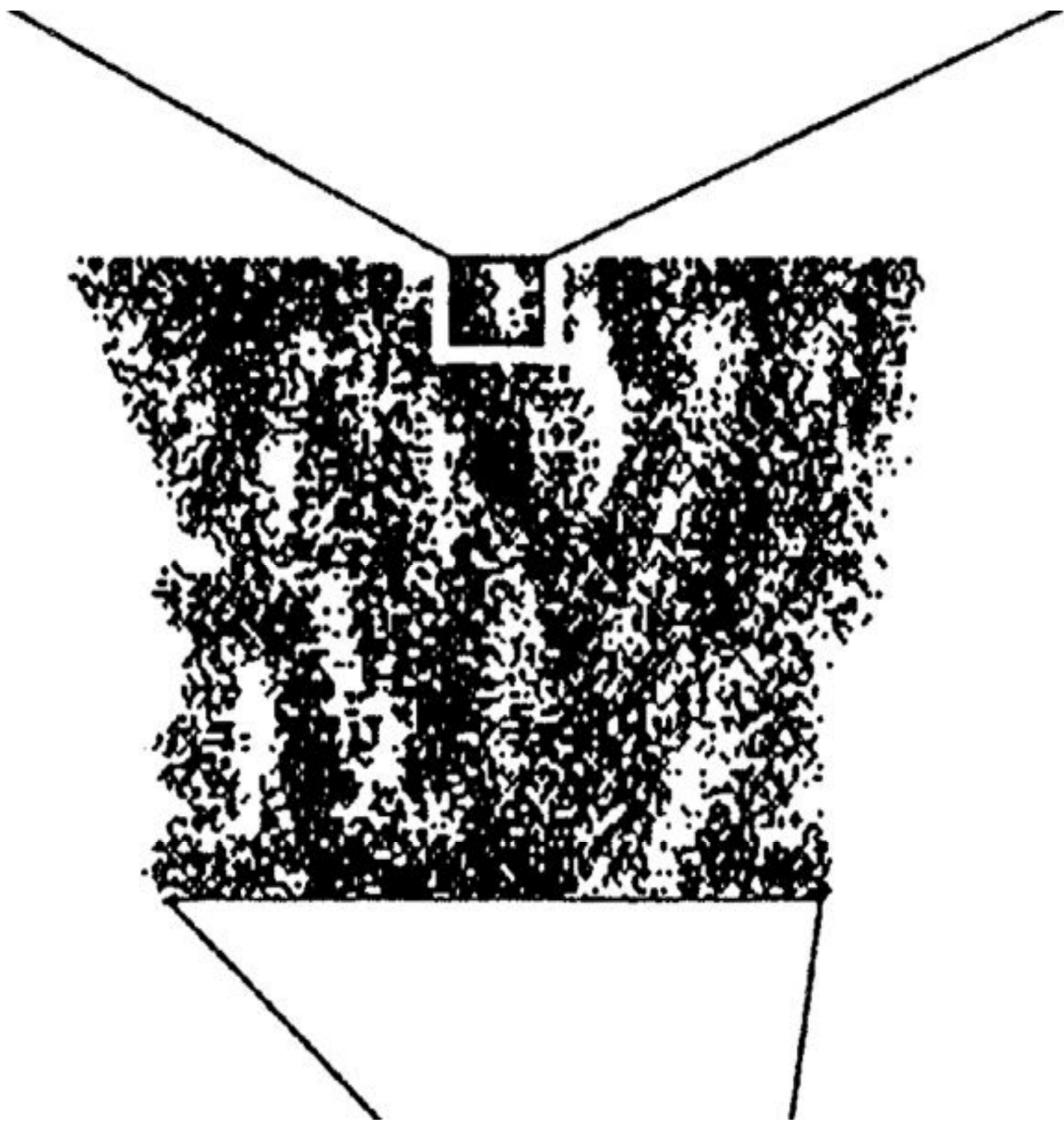
1994

1993

population



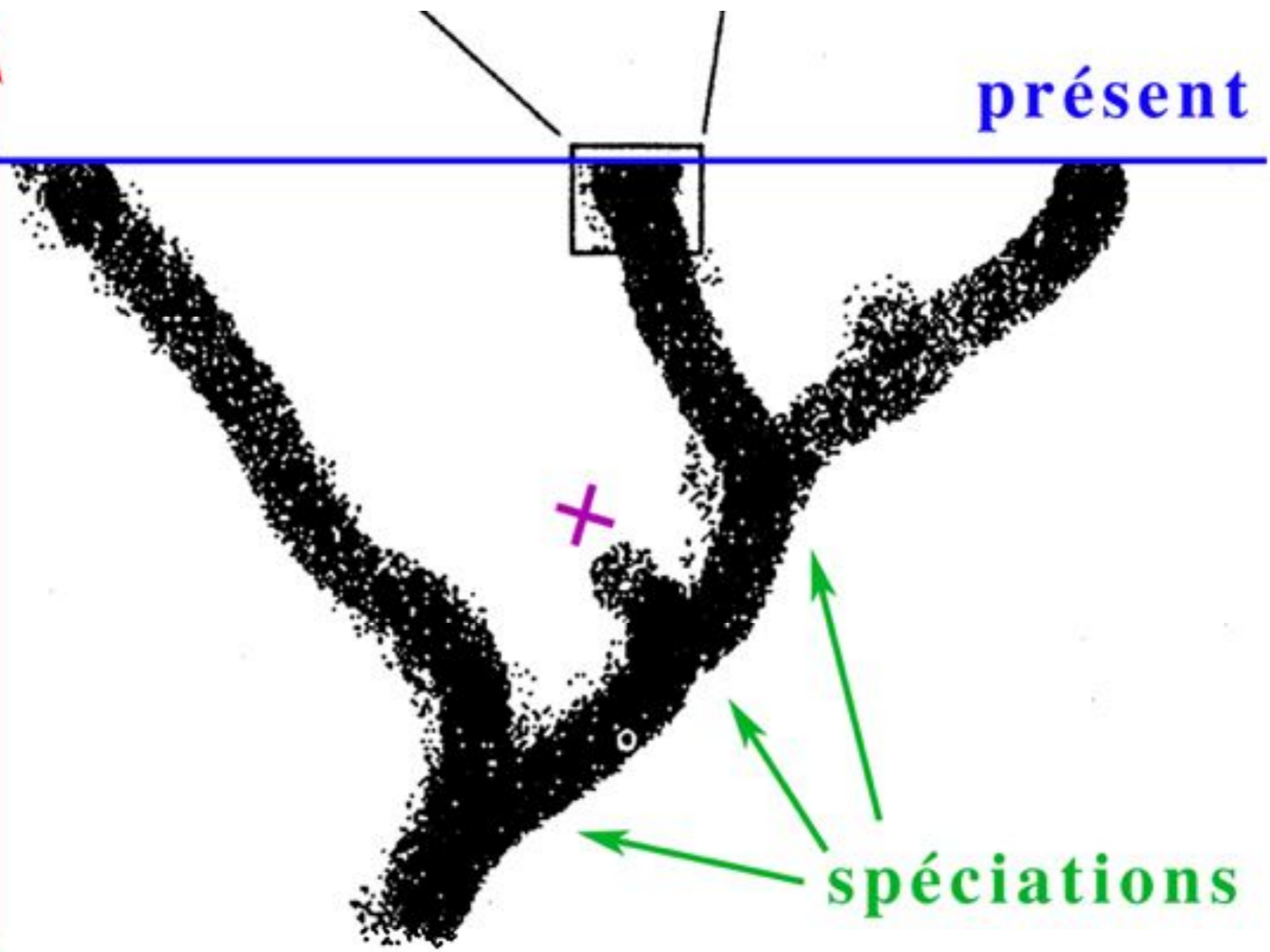
espèce



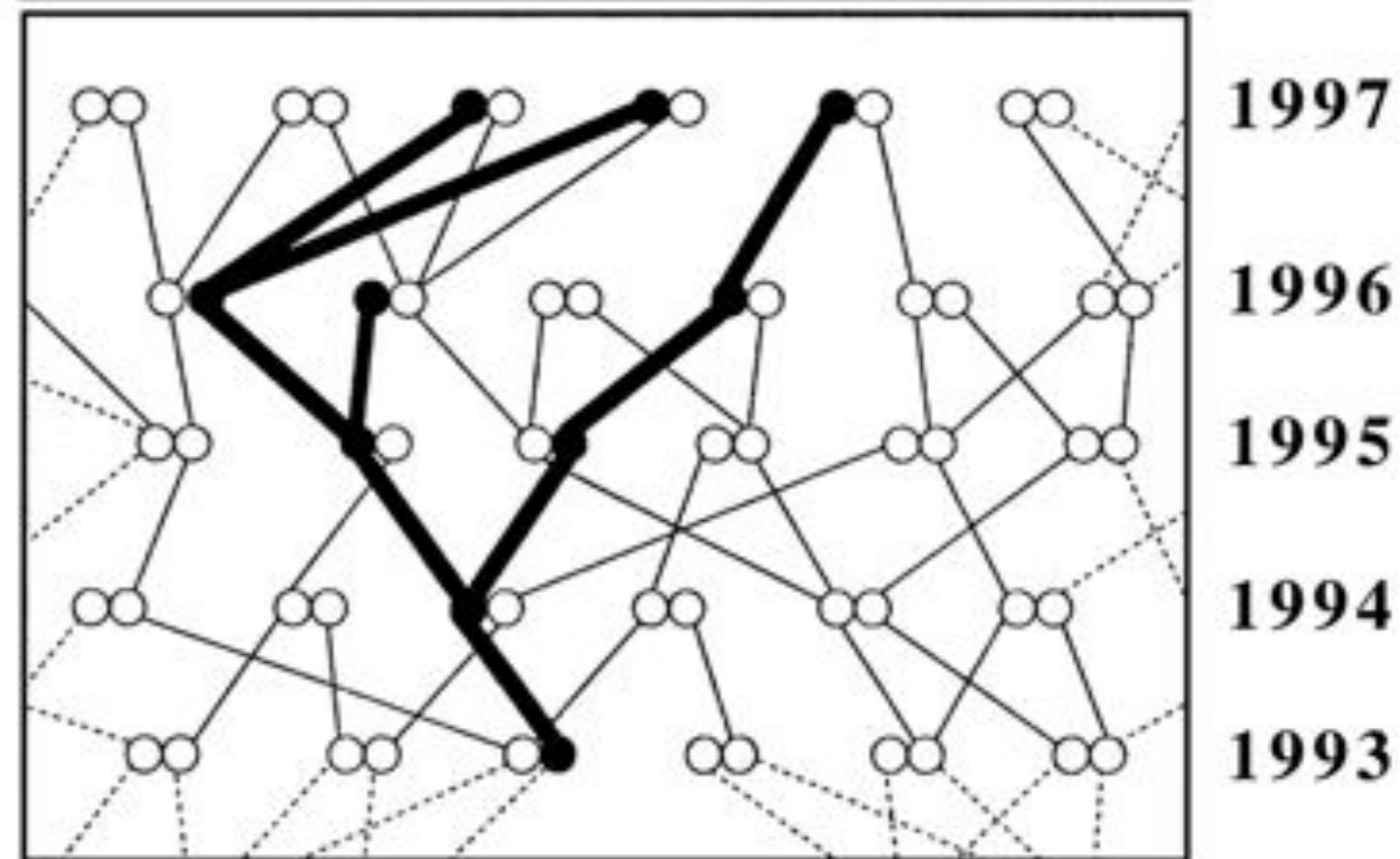
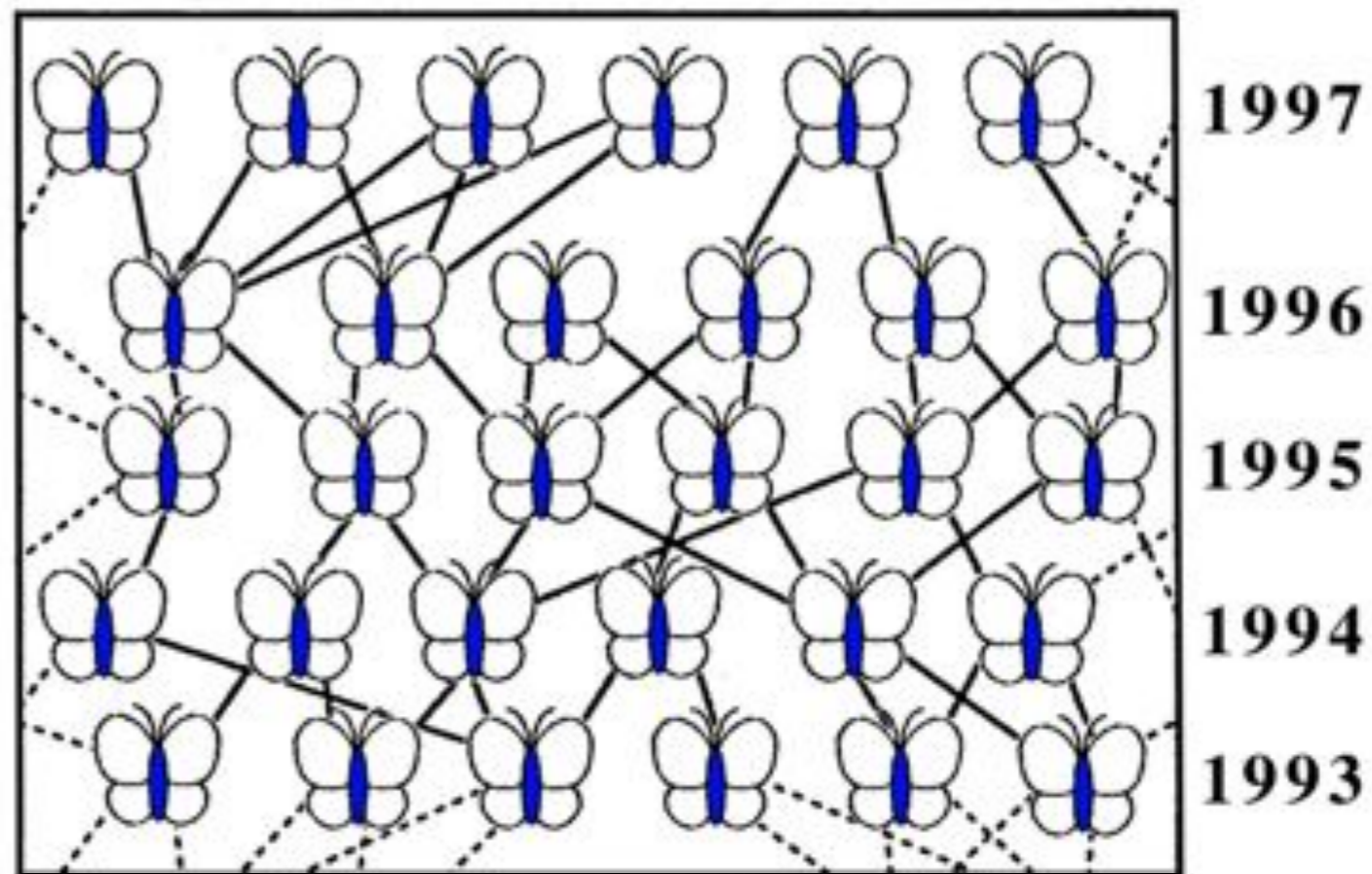
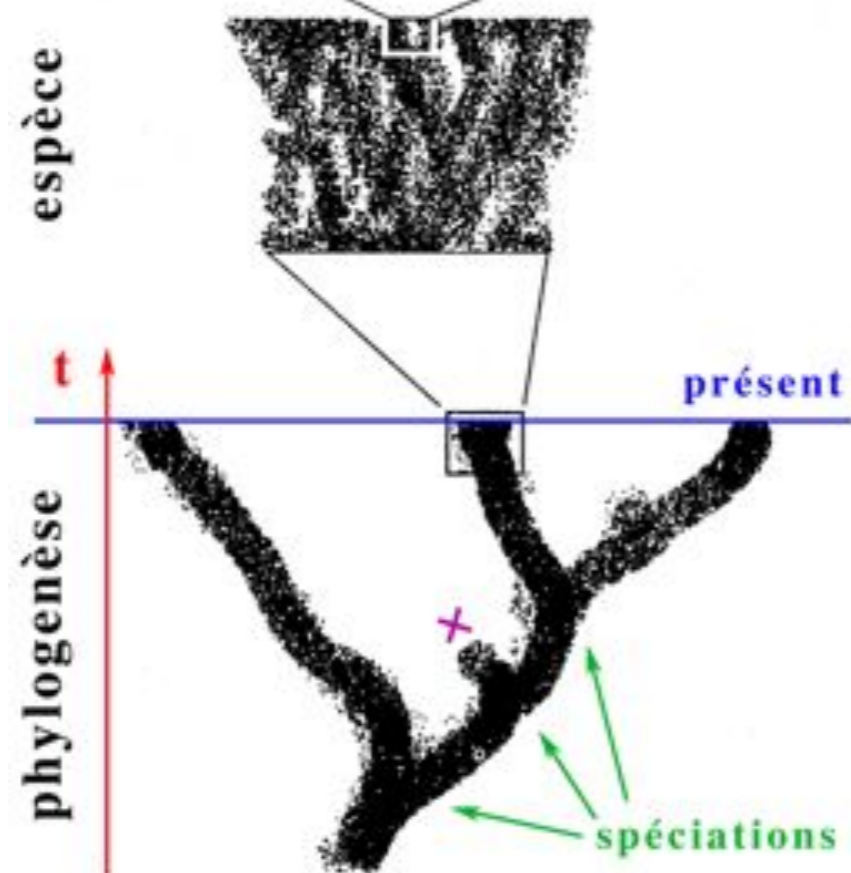
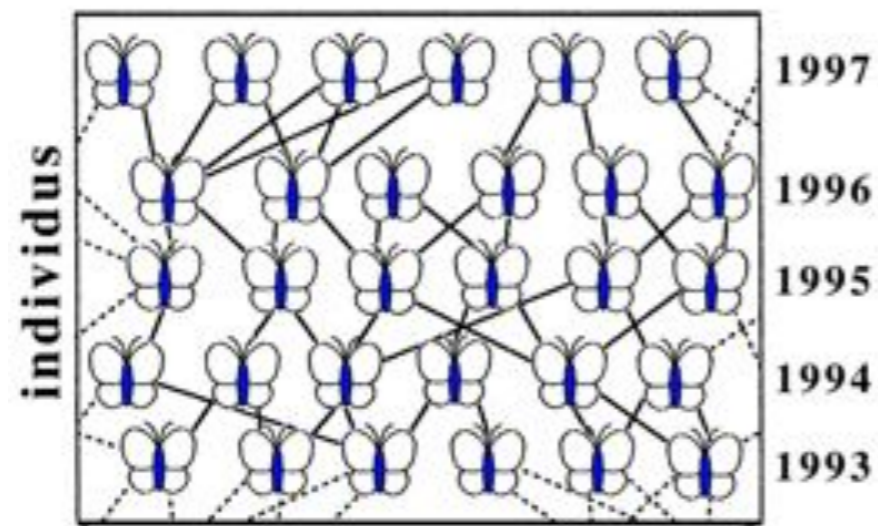
phylogenèse



présent

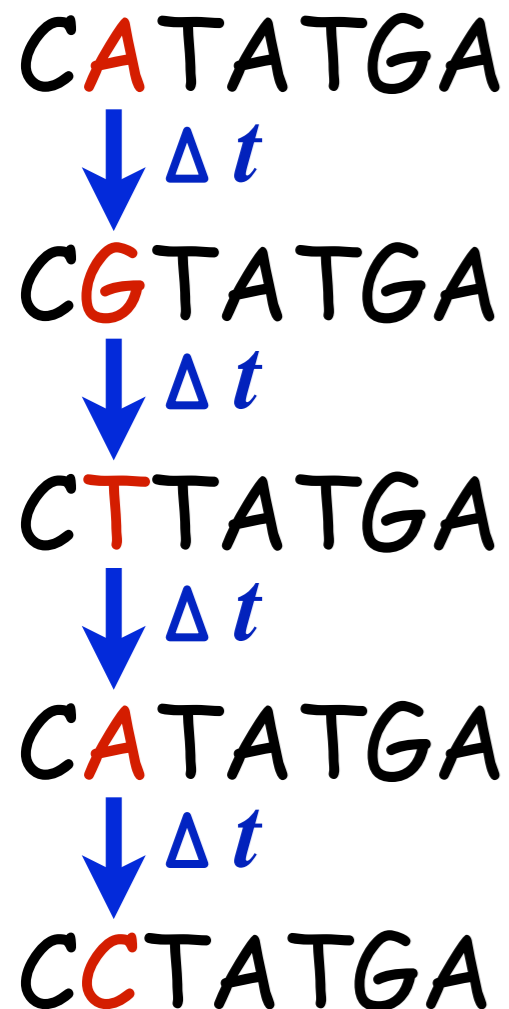


spéciations



Computing pairwise distances: the problem

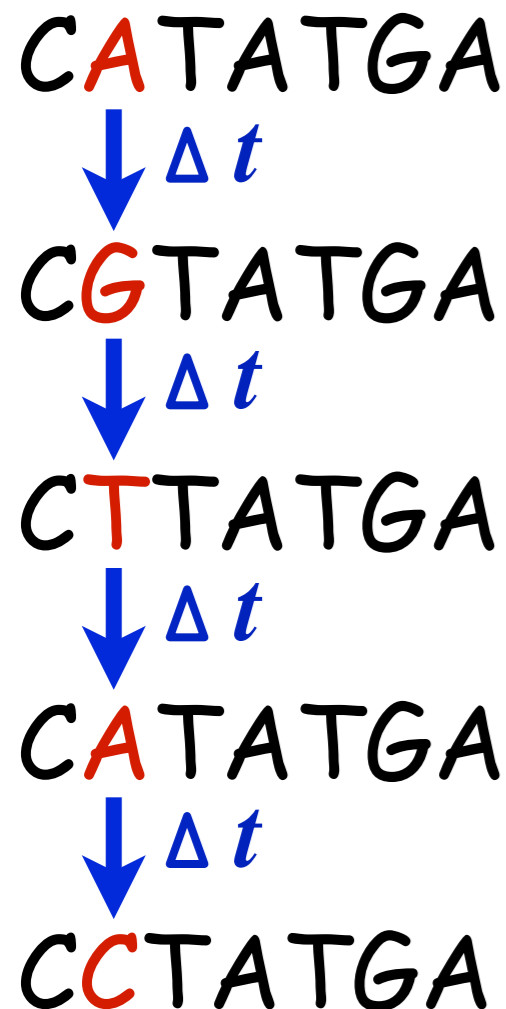
The observed number of differences between two aligned sequences separated by a time t is a poor indicator of the number of substitutions that occurred between these two sequences (unless t is small).



Divergence between two sequences will not increase linearly with time as multiple substitutions may occur at the same site. This is true even for infinite-length sequences. Back, parallel, & convergent substitutions can even INCREASE similarity, locally in time.

Computing pairwise distances: the problem

The observed number of differences between two aligned sequences separated by a time t is a poor indicator of the number of substitutions that occurred between these two sequences (unless t is small).



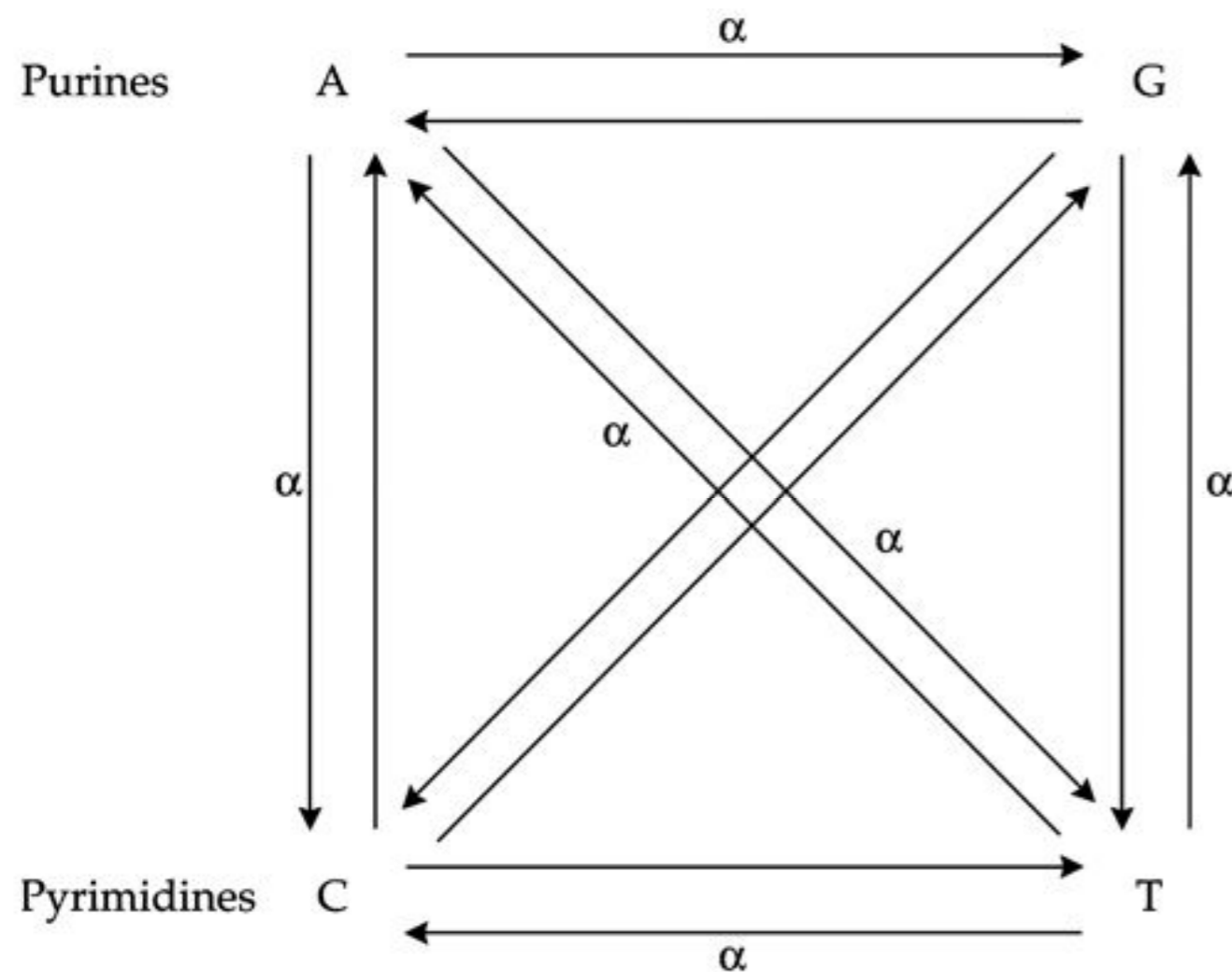
In probabilistic terms:

If the character state is "A" at time t_0 , what is the probability, $P_{A(t)}$, that the character state will be "A" at time t ?

Models have been proposed to approximate the true dynamic of nucleotide substitution.

Computing distances I: the Jukes-Cantor (JC) model

Assumption: Substitutions occur with equal rates among the four nucleotide states
--> 1 single parameter

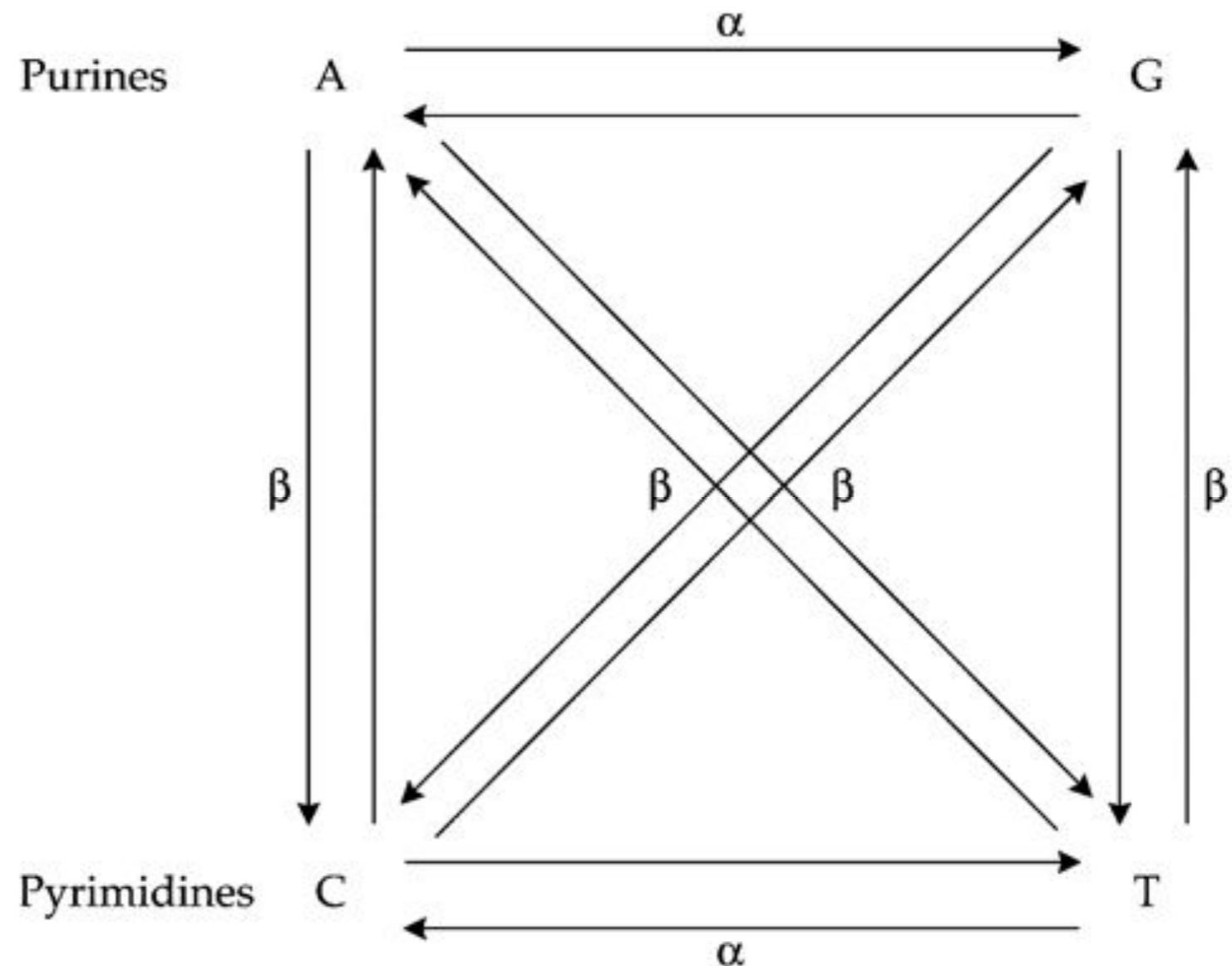


alpha is a rate,
i.e., for example a
number of substitutions
per site per year.

Computing distances II: the K2P model

Assumption: T_i and T_v and different rates among the four nucleotide states and all state equilibrium frequencies are equal (0.25).

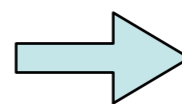
--> 2 parameters



Computing distances III: GTR

The transition probability matrix can be obtained by:

$$R = \begin{bmatrix} \cdot & \alpha\pi_T & \beta\pi_C & \gamma\pi_G \\ \alpha\pi_A & \cdot & \rho\pi_C & \sigma\pi_G \\ \beta\pi_A & \rho\pi_T & \cdot & \tau\pi_G \\ \gamma\pi_A & \sigma\pi_T & \tau\pi_C & \cdot \end{bmatrix}$$



$$\mathbf{P}(t) = \mathbf{e}^{\mathbf{R}t} \quad *$$

The GTR model assumes reversibility: the net rate from j to i is equal to the net rate from i to j .

i.e., $\pi_i r_{ij} = \pi_j r_{ji}$ **

Given * & **, Rodriguez et al (1990) demonstrated:

$$\mathbf{\Pi P}(t) = \mathbf{P}(t)^T \mathbf{\Pi}$$

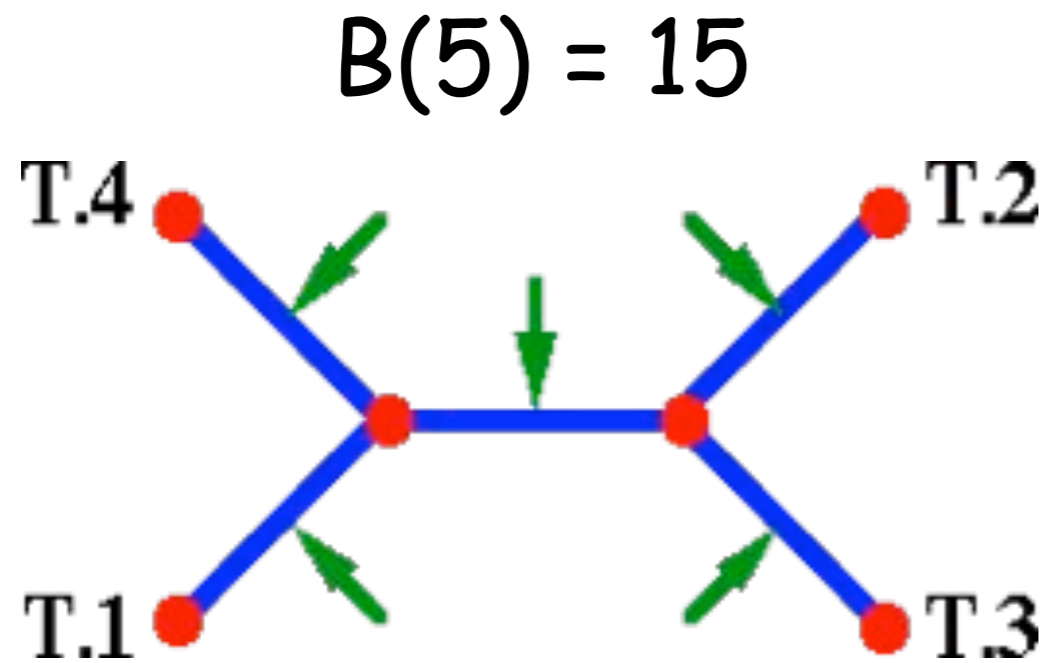
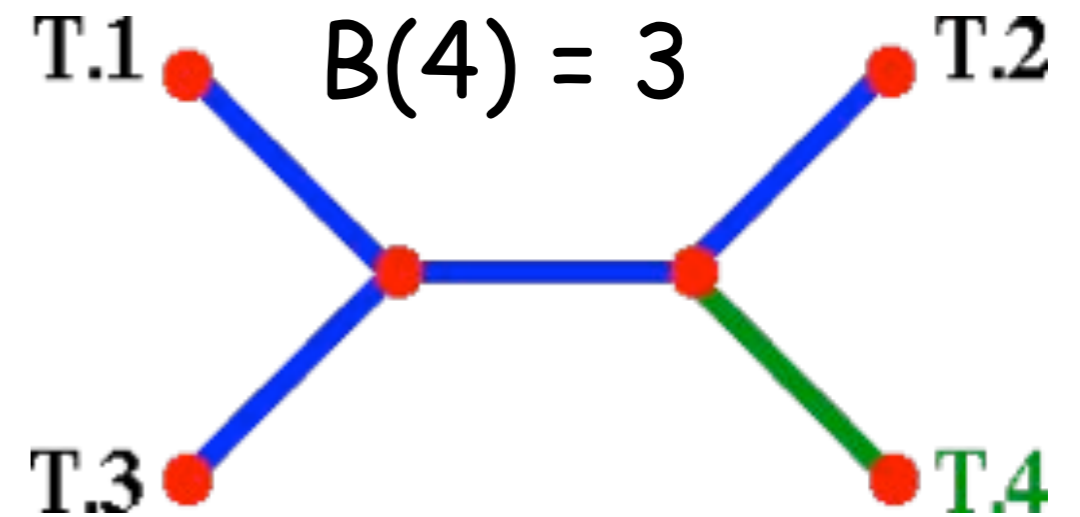
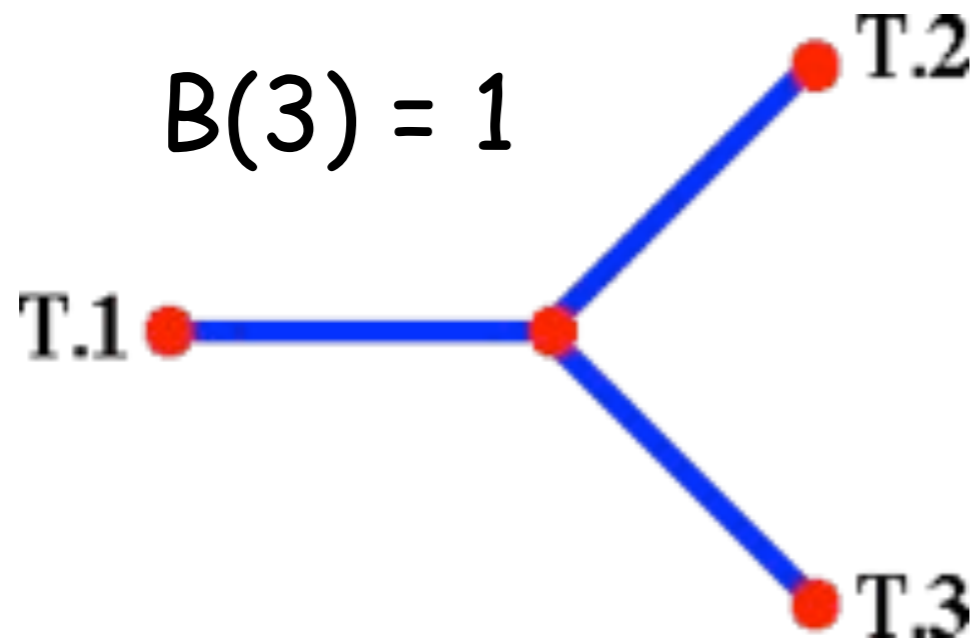
where $\mathbf{\Pi}$ is the diagonal matrix of state equilibrium frequencies.

$$\mathbf{\Pi} = \begin{bmatrix} \pi_A & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_G & 0 \\ 0 & 0 & 0 & \pi_T \end{bmatrix}$$

and

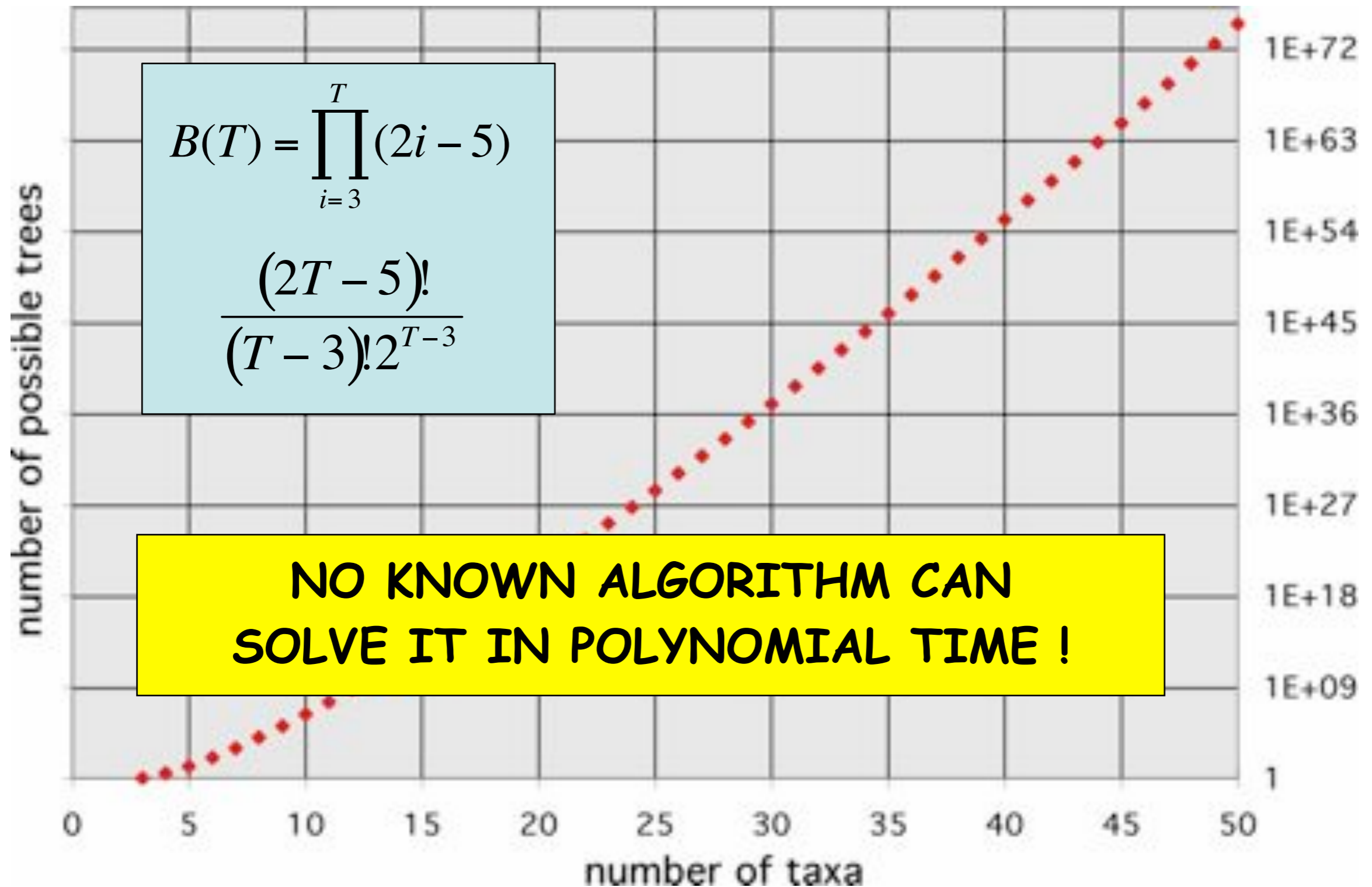
$$\mathbf{P}(\hat{t}) = \mathbf{P} = \mathbf{\Pi}^{-1} (\mathbf{P}(\hat{t})^T \mathbf{\Pi}) = \mathbf{\Pi}^{-1} \mathbf{F}^\#$$

Number of possible trees

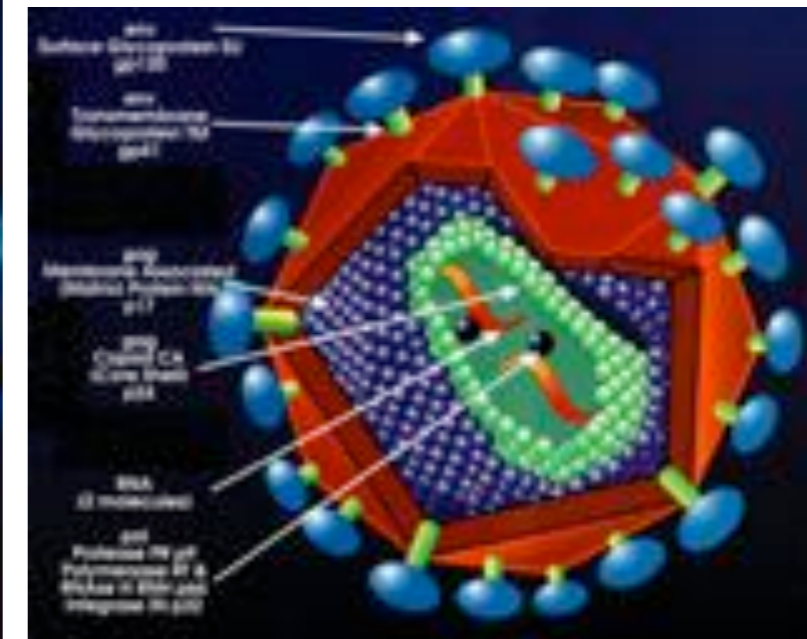
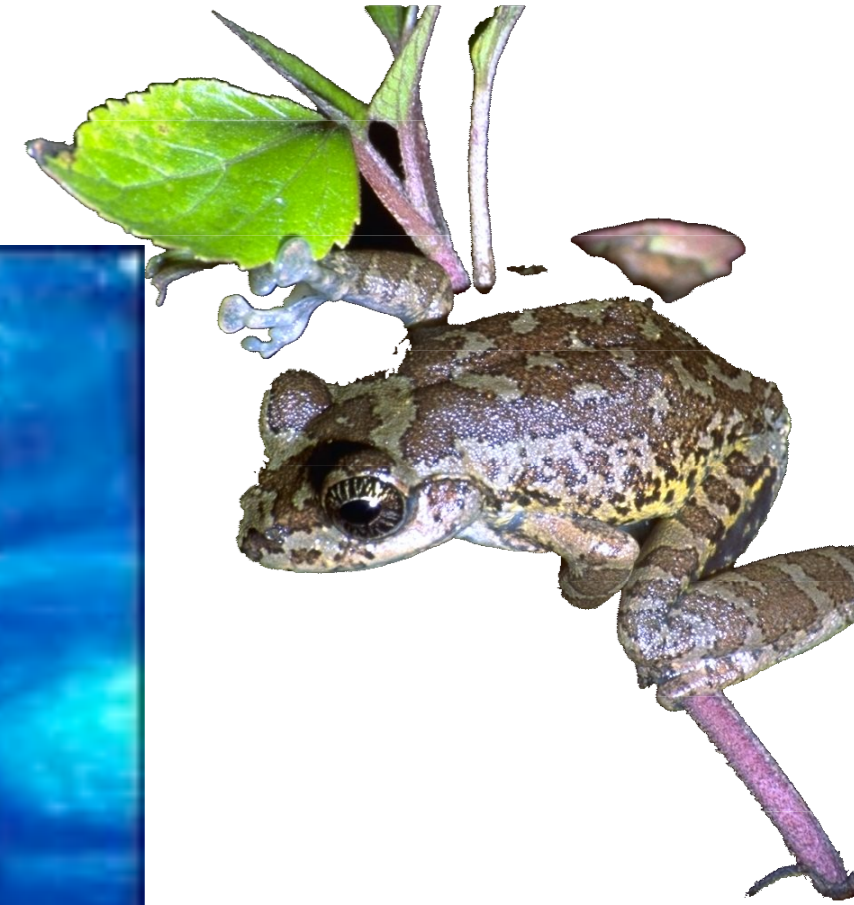


$$B(T) = \prod_{i=3}^T (2i - 5)$$
$$\frac{(2T - 5)!}{(T - 3)! 2^{T-3}}$$

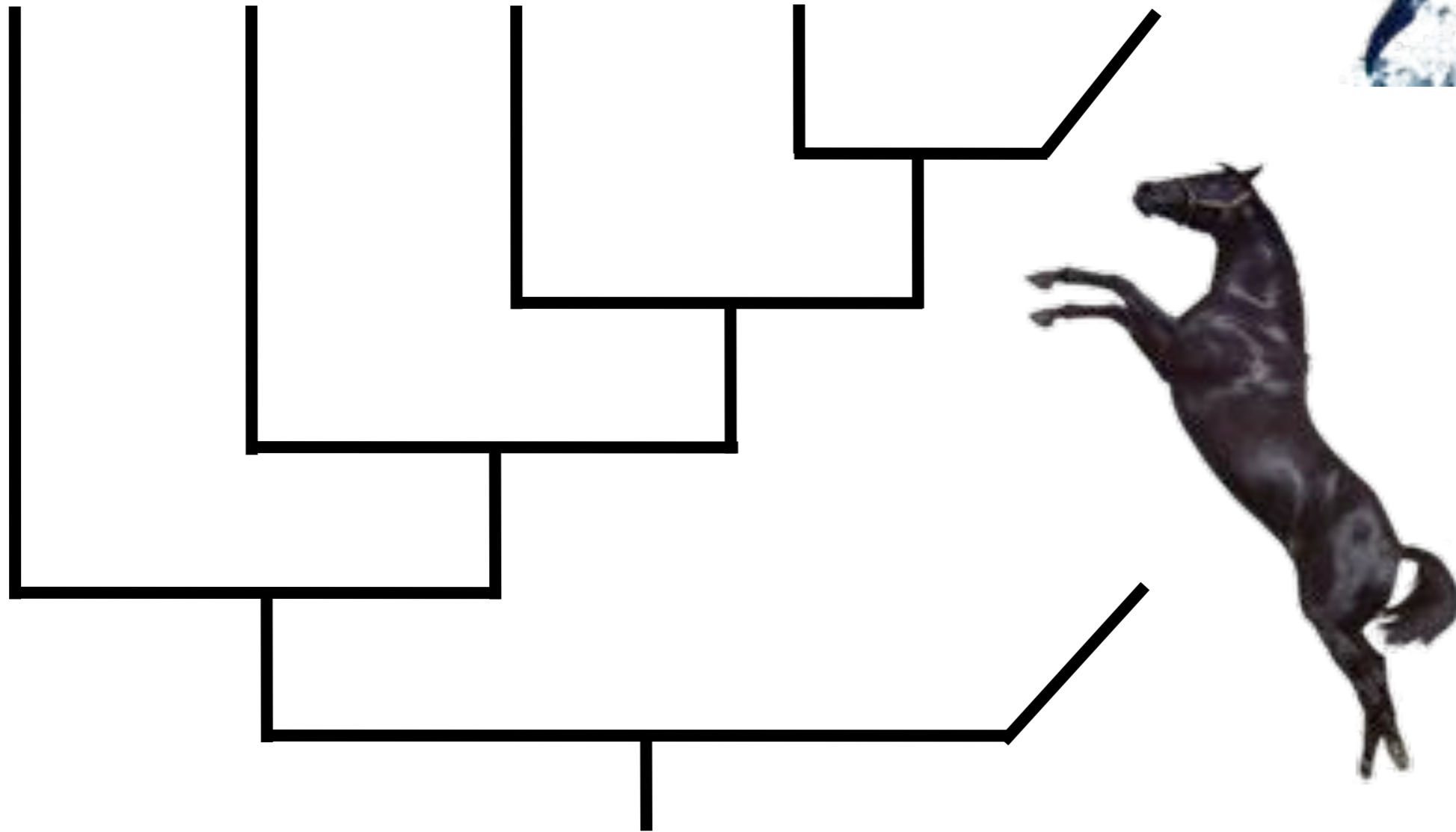
"NP-h" combinatorial optimization problem



● Molecular Phylogeny inference



derived
morphology



e.g., Milinkovitch & Thewissen. *Nature* (1997)
Gatesy et al. *Systematic Biology* (1999)



E.g.,

Bossuyt & Milinkovitch *PNAS* (2000)

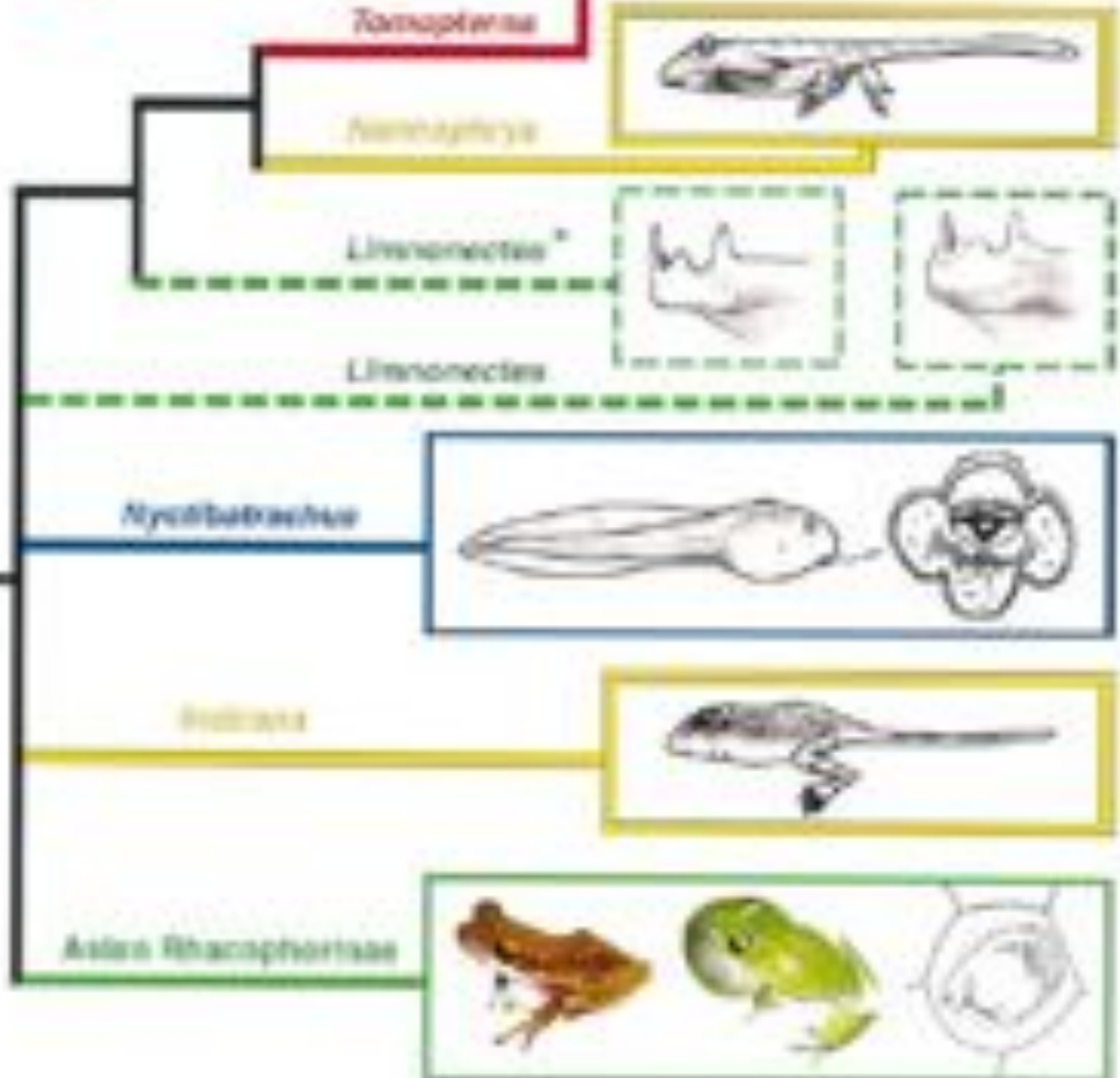
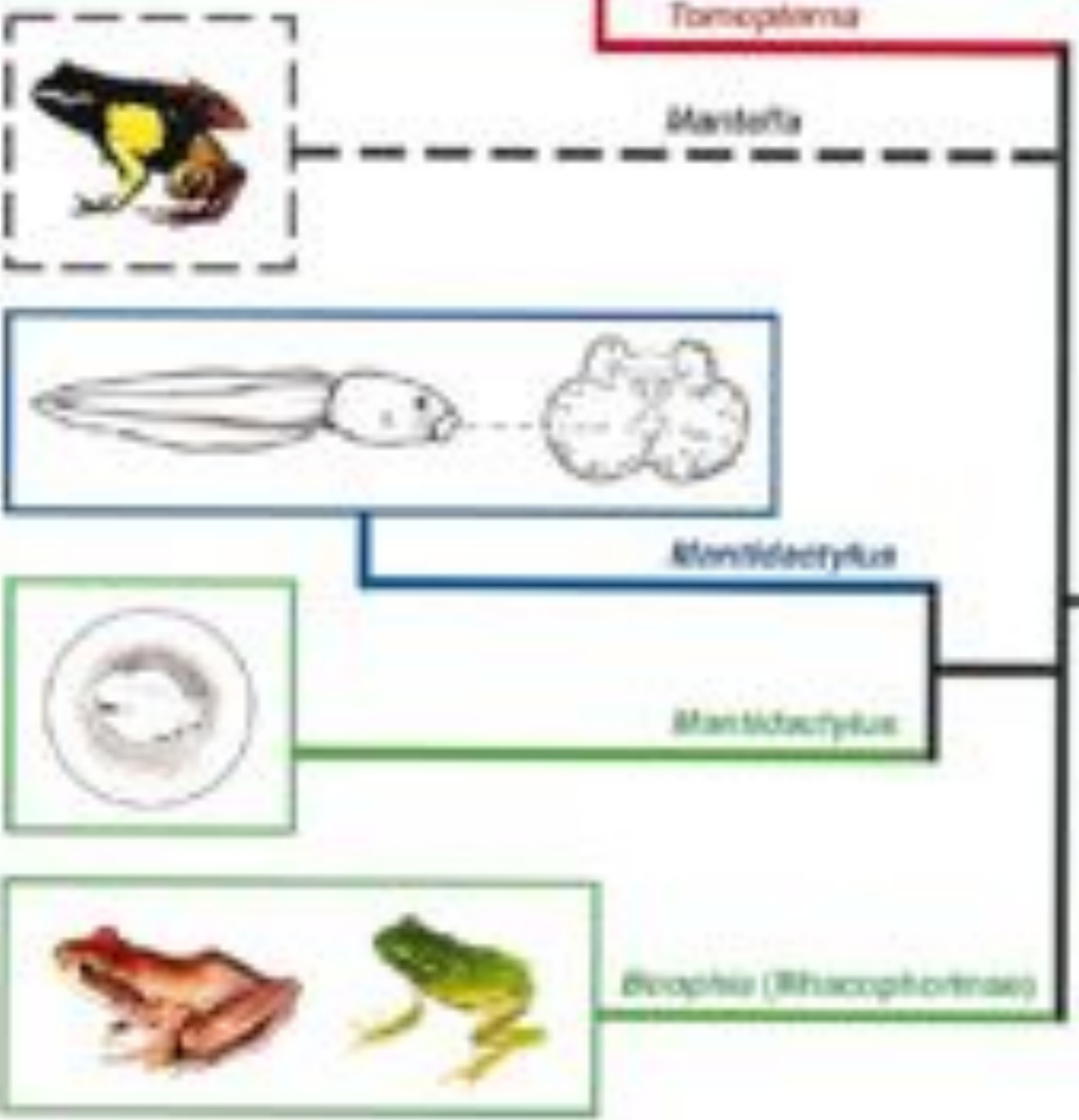
Bossuyt & Milinkovitch *Science* (2001)

Bossuyt *et al.* *Science* (2004)

Bossuyt *et al.* *Systematic Biology* (2006)

Madagascar

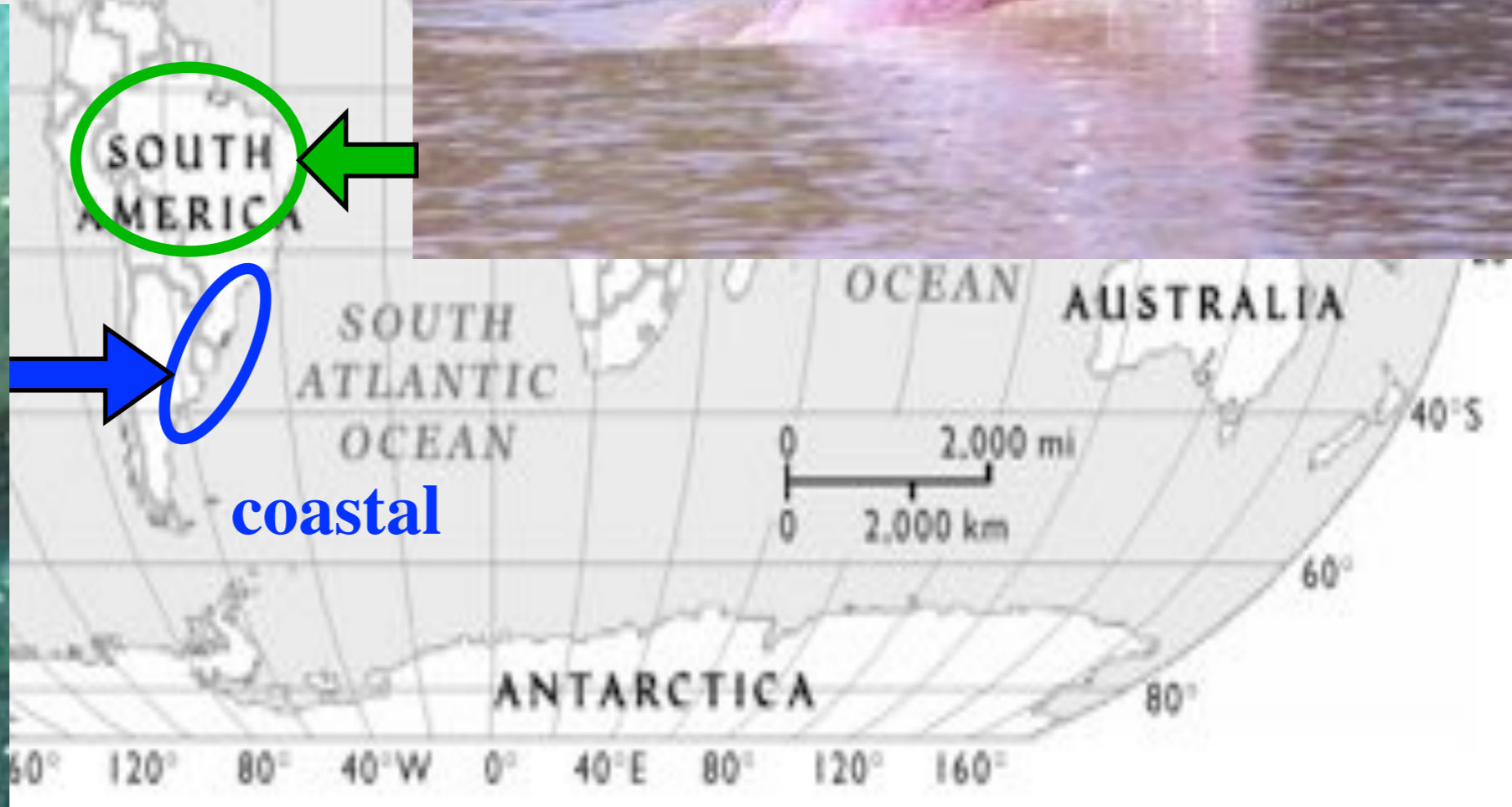
Asia



■ Burrowing
 ■ Rock-dwelling
 ■ Fanged
 ■ Poisonous
 ■ Arboreal
 ■ Terrestrial

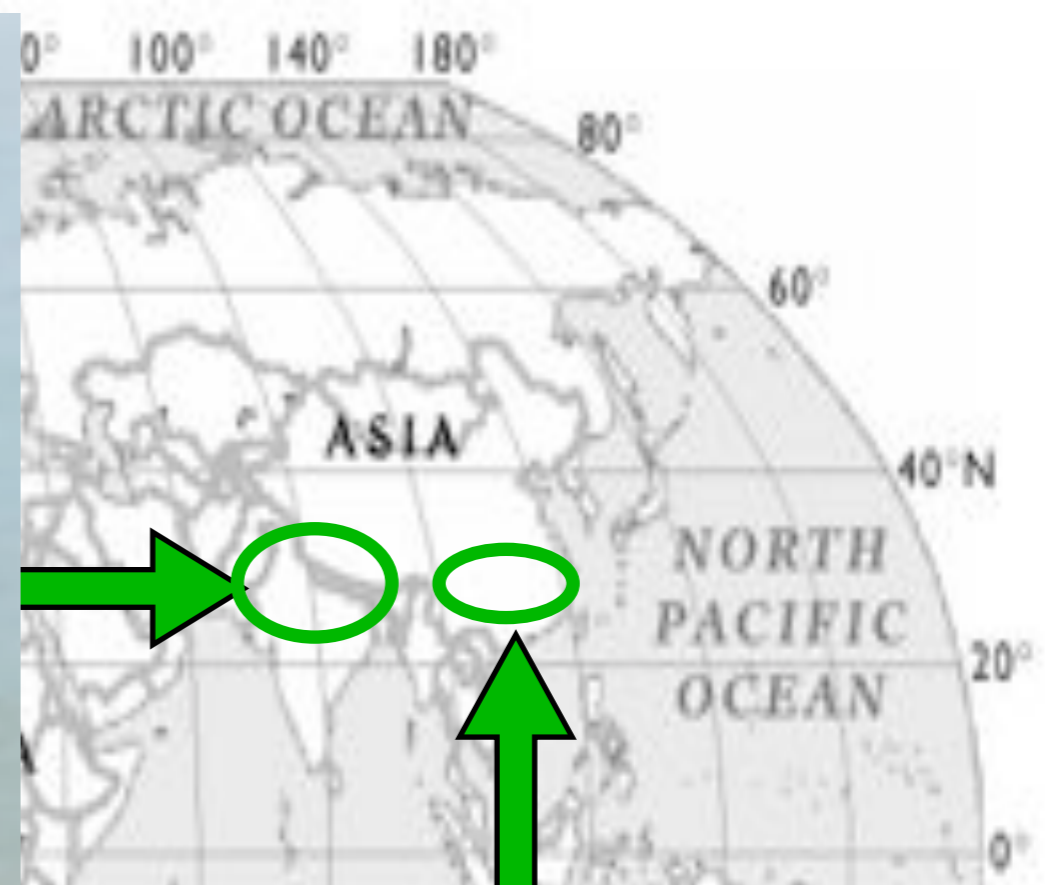


I. geoffrensis
(Boto)

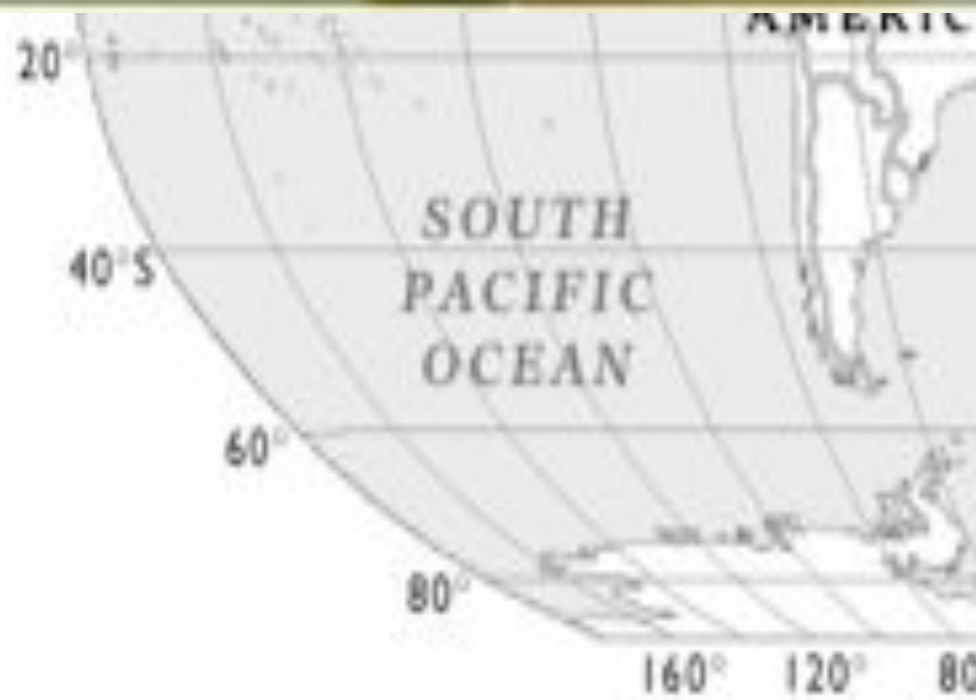


P. blainvillei
(Franciscana)

P. gangetica
(Susu)



L. vexillifer.
(Baiji)



Similar external morphology



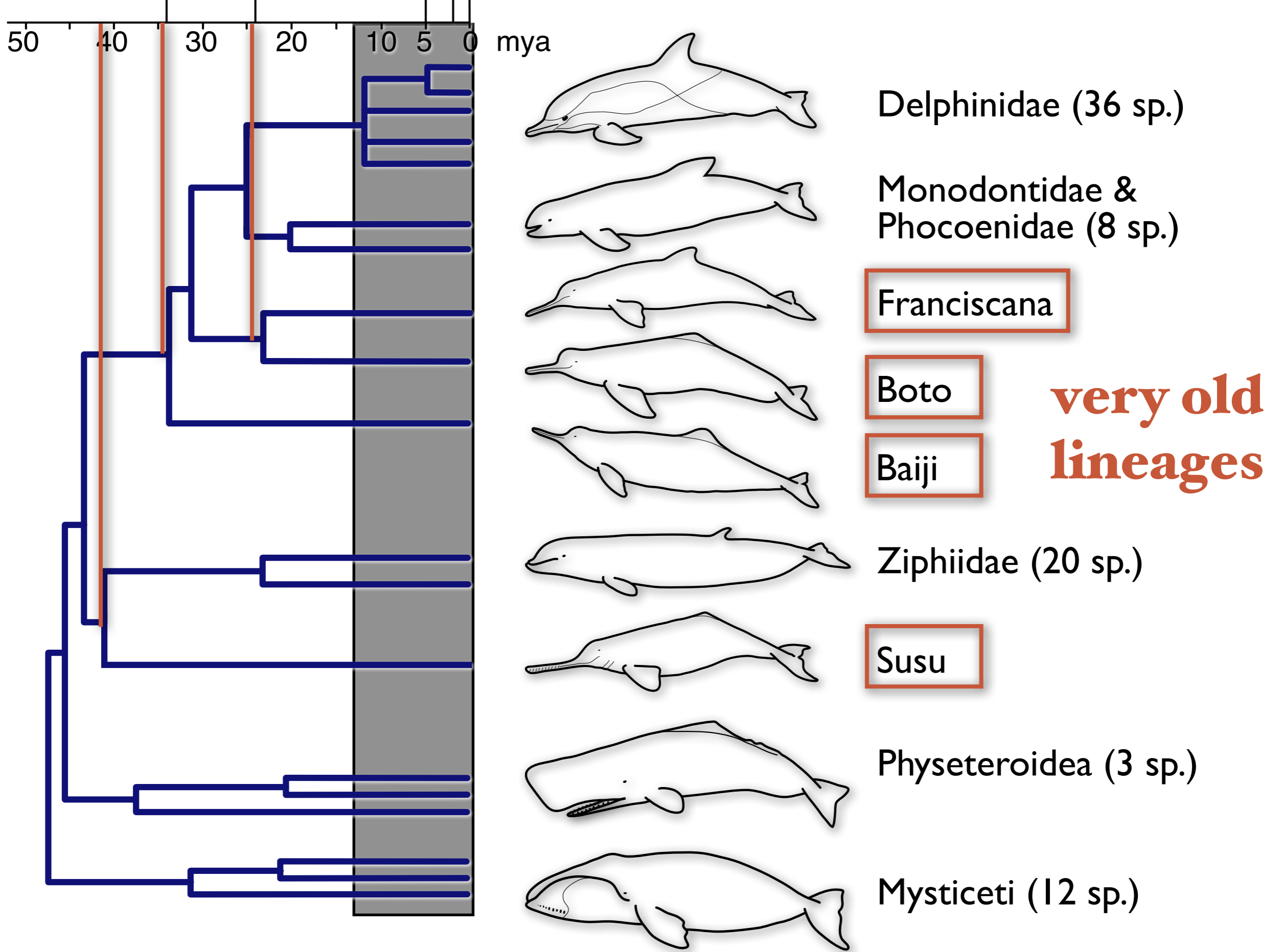
**Large,
flexible
flippers**



Long beak

**Low,
triangular
dorsal fin**





Super-Order Afrotheria

5g à 5T!!



- ✓ O. Proboscidea
- ✓ O. Macroscelidea (elephant shrews),
- ✓ O. Tubulidentata
- ✓ O. Afrosoricida (Golden Moles & tenrecs)
- ✓ O. Hyracoidea
- ✓ O. Sirenia



Springer *et al.* Nature (1997)
Stanhope *et al.* PNAS (1998)
Springer *et al.* Syst Biol (1999)
Van Dijk *et al.* PNAS (2001)

Super-Order Afrotheria



Isolated Africa (105-→ 45 mya)
→ collision with Eurasia.

**No known shared derived
morphological character**

Michel C. Milinkovitch
Dpt. Genetics & Evolution,
University of Geneva
Michel.Milinkovitch@unige.ch

www.lanevol.org



Laboratory of
Artificial & Natural
Evolution

