

# Formation à la phylogénie moléculaire

Lyon – 19-22 mars 2013

## Introduction à la phylogénie moléculaire - aspects biologiques -

Manolo Gouy, CNRS, Lyon

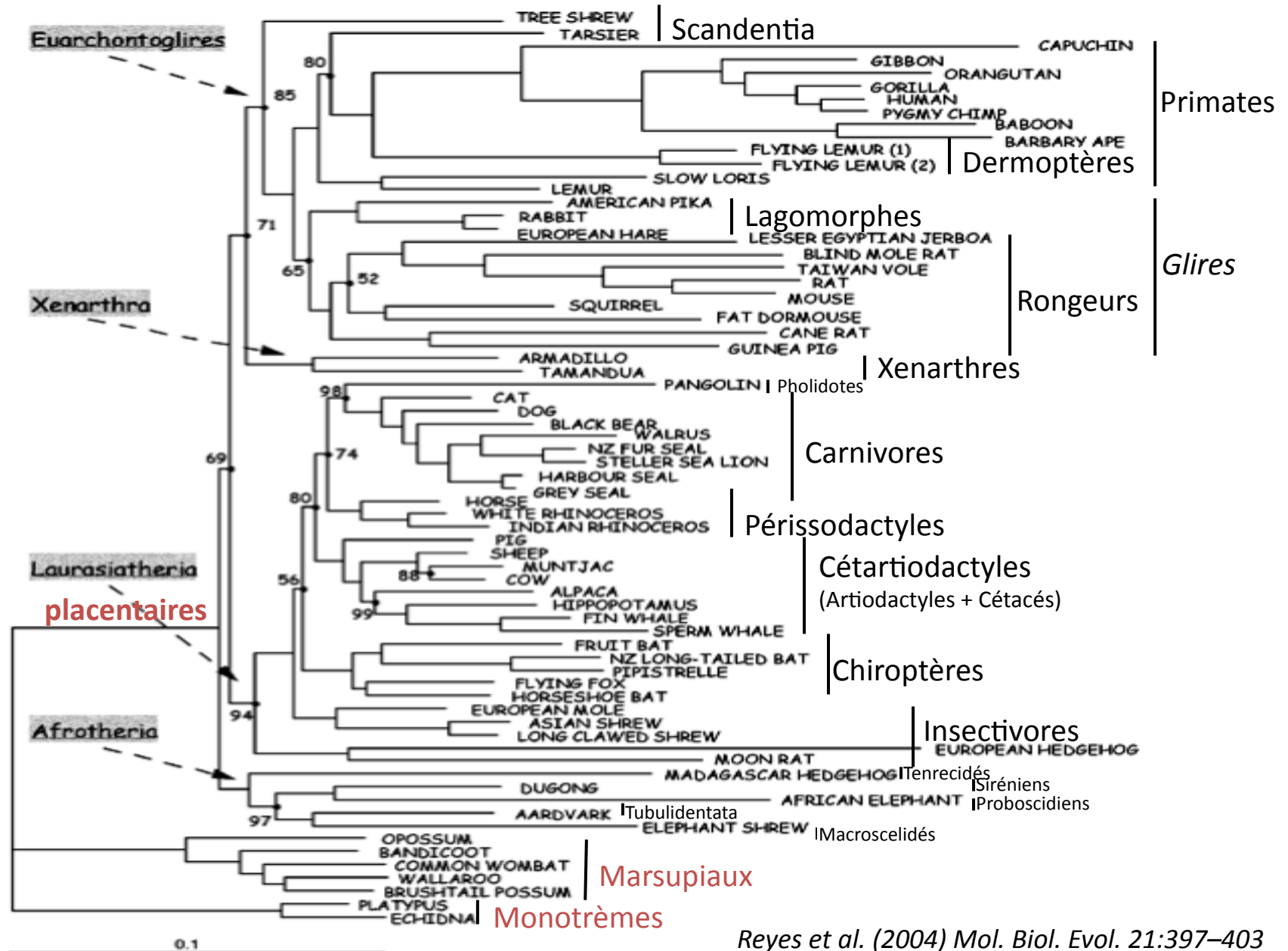
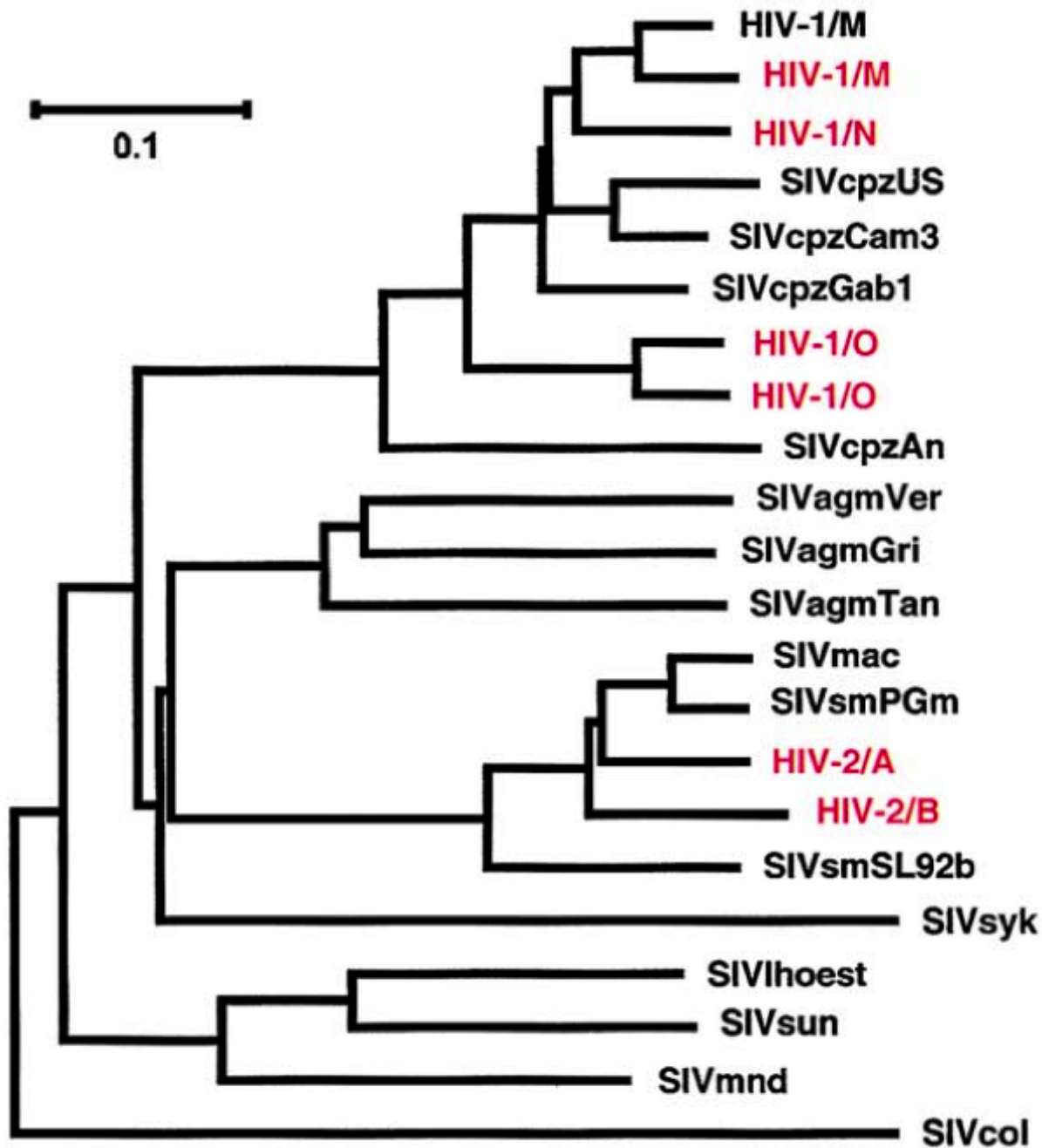


FIG. 1.—Phylogenetic tree of placental mammals reconstructed using the program MrBayes from mitochondrial H-stranded protein-coding genes using ungapped first and second codon positions with the exclusion of Leu synonymous sites. Posterior probabilities (PP) supporting the tree nodes are only reported when less than 100. Marsupialia and Monotremata were used as outgroups. The lengths of the branches are proportional to the number of nucleotide substitutions per site.

*Reyes et al. (2004) Mol. Biol. Evol. 21:397–403*



## Origine du virus du SIDA

cpz: chimpanzé --> HIV-1

agm: singe vert africain

mac: macaque

sm: *Cercocebus atys* --> HIV-2

syk: *Cercopithecus albogularis*

lhoest: *C. lhoesti*

sun: *C. solatus*

mnd: *Mandrillus sphinx*

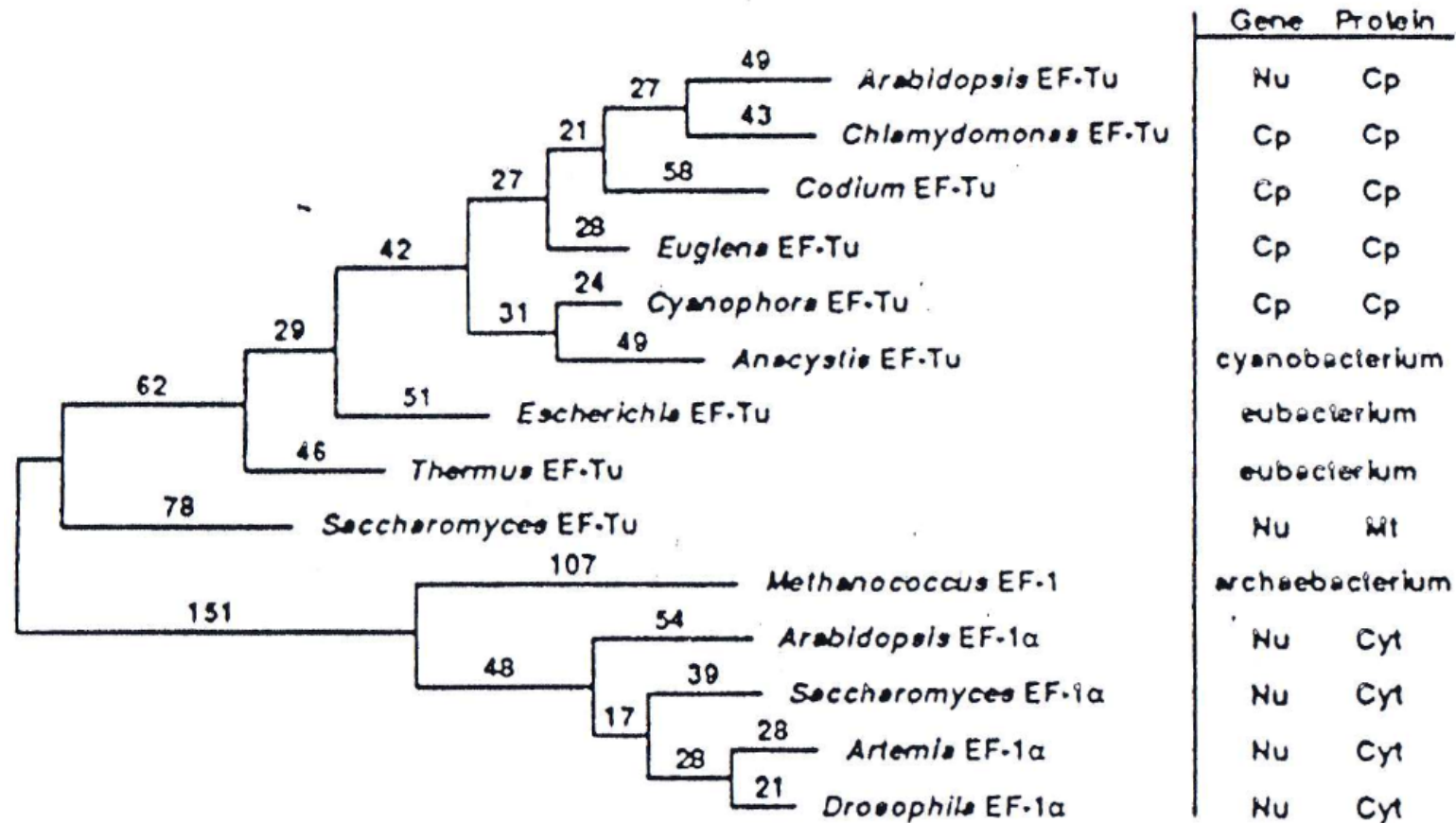
col: *Colobus guereza*

La vitesse d'évolution des molécules utilisées doit être adaptée à l'échelle temporelle du phénomène étudié.

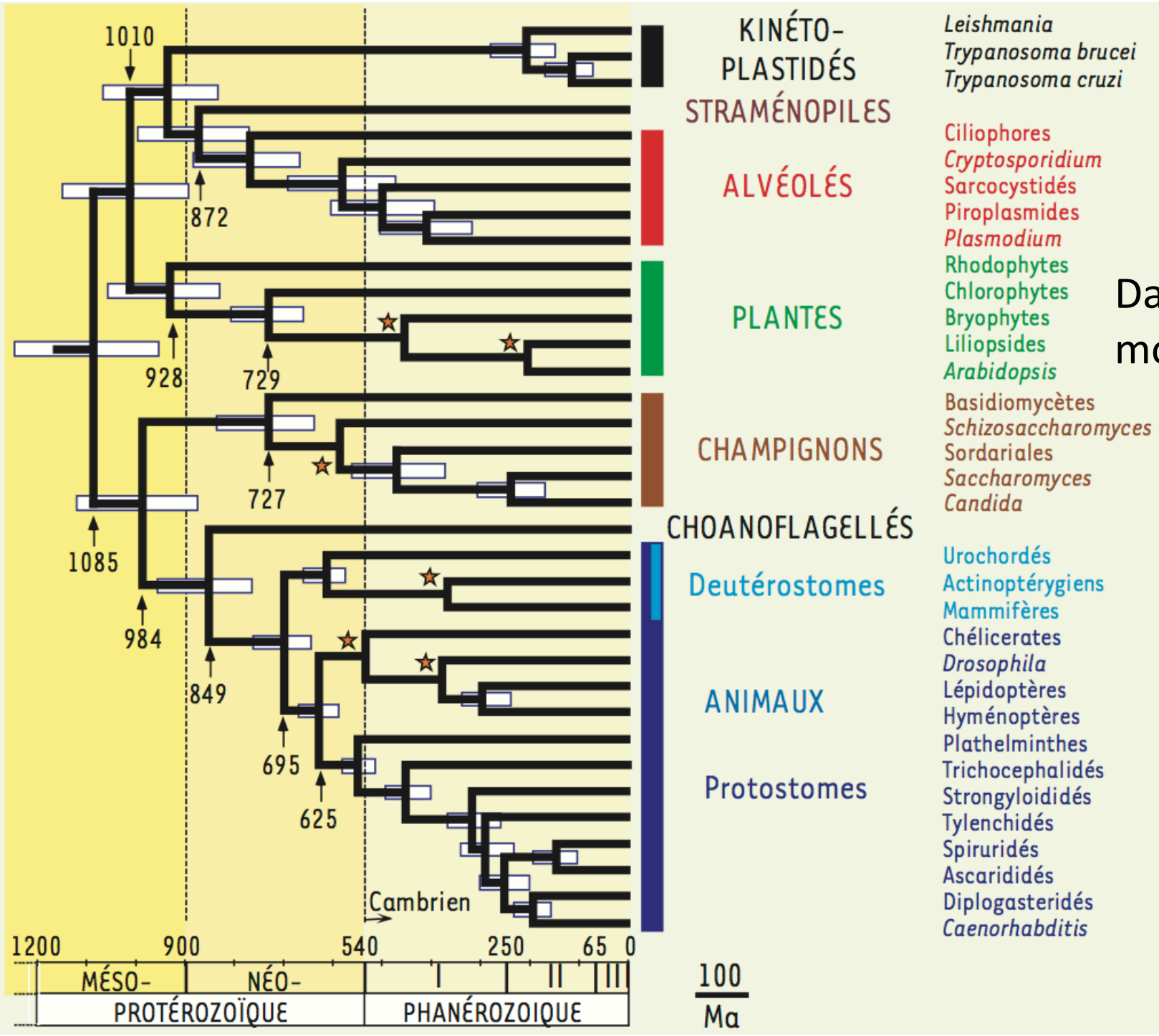
Figure 1. Evolution of AIDS Viruses

Evolutionary transfer of the chloroplast tufA gene to the nucleus.

LETTERS TO NATURE



# Datation moléculaire

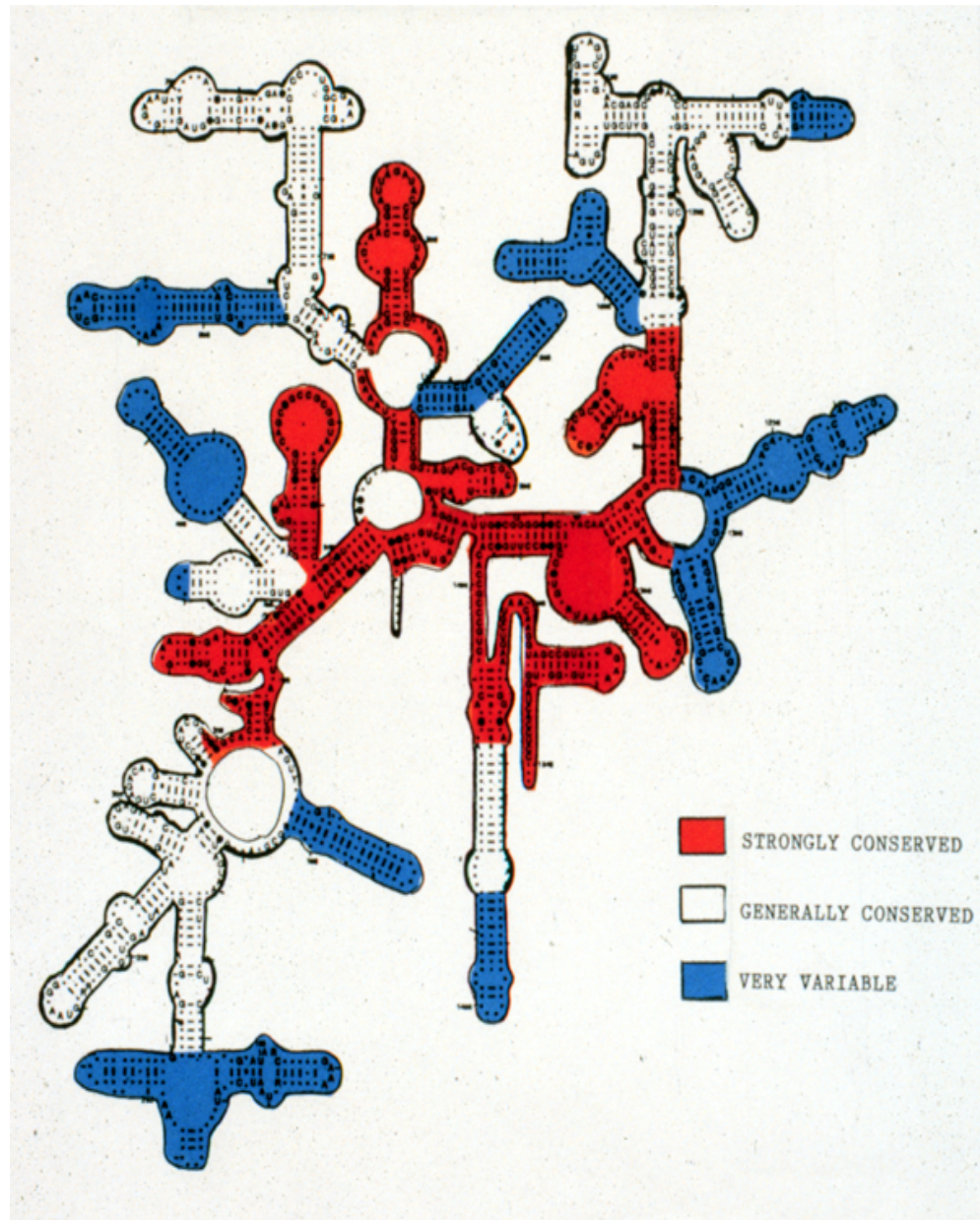


# Alignement et Gaps

- La qualité de l'alignement est essentielle : chaque colonne de l'alignement (site) est supposée contenir des résidus homologues (nucléotides, acides aminés) qui dérivent d'un ancêtre commun.  
=> Les parties non fiables de l'alignement doivent être omises du reste des analyses.
- La plupart des méthodes ne tiennent compte que des substitutions ; les gaps (événements d'insertion/délétion) ne sont pas utilisés.  
=> les sites contenant des gaps sont ignorés.



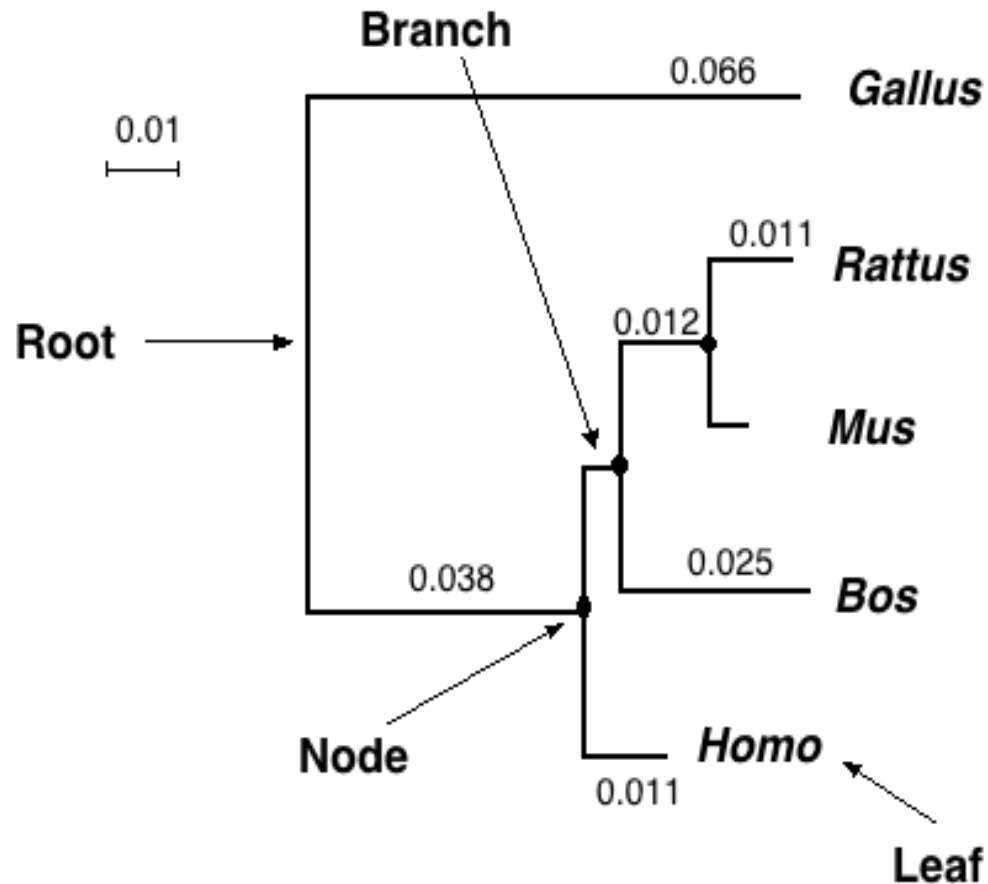
# Variation de la vitesse d'évolution entre sites



**Small subunit  
ribosomal RNA  
(18S or 16S)**

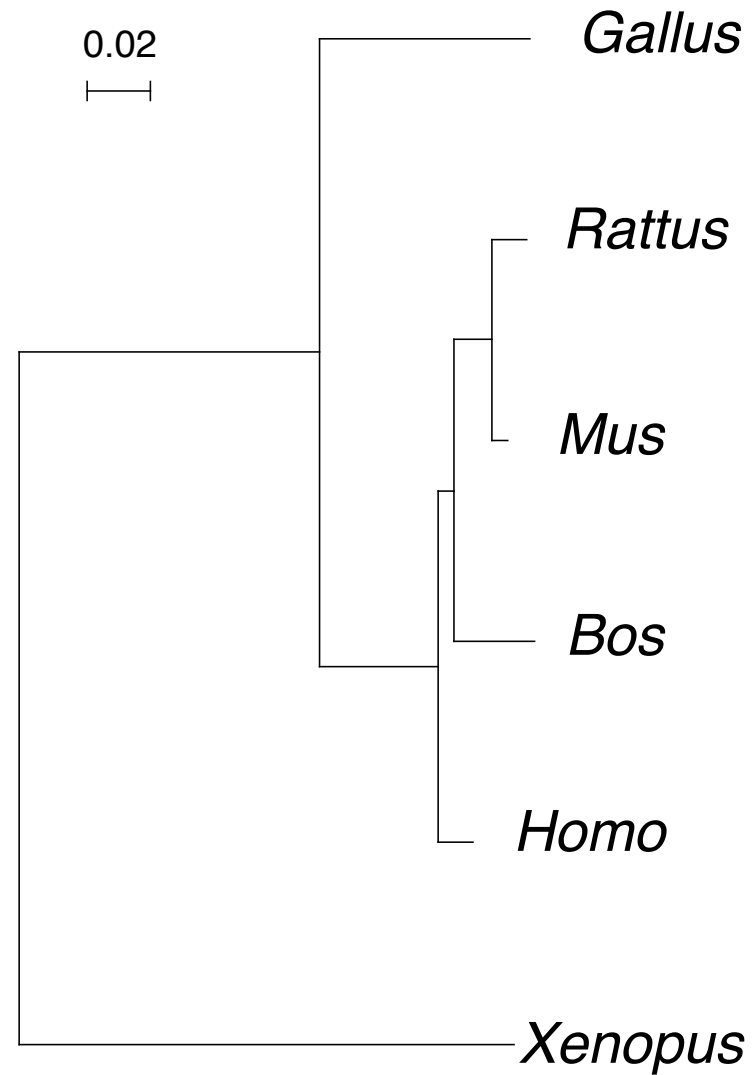
# Arbre Phylogénétique

- Branche Interne: entre 2 nœuds. Branche Externe: entre un nœud et une feuille
- Les longueurs des branches horizontales sont proportionnelles aux distances évolutives entre séquences ancestrales (unité = substitution / site).
- Topologie d'arbre = forme de l'arbre = ordre de branchement des nœuds

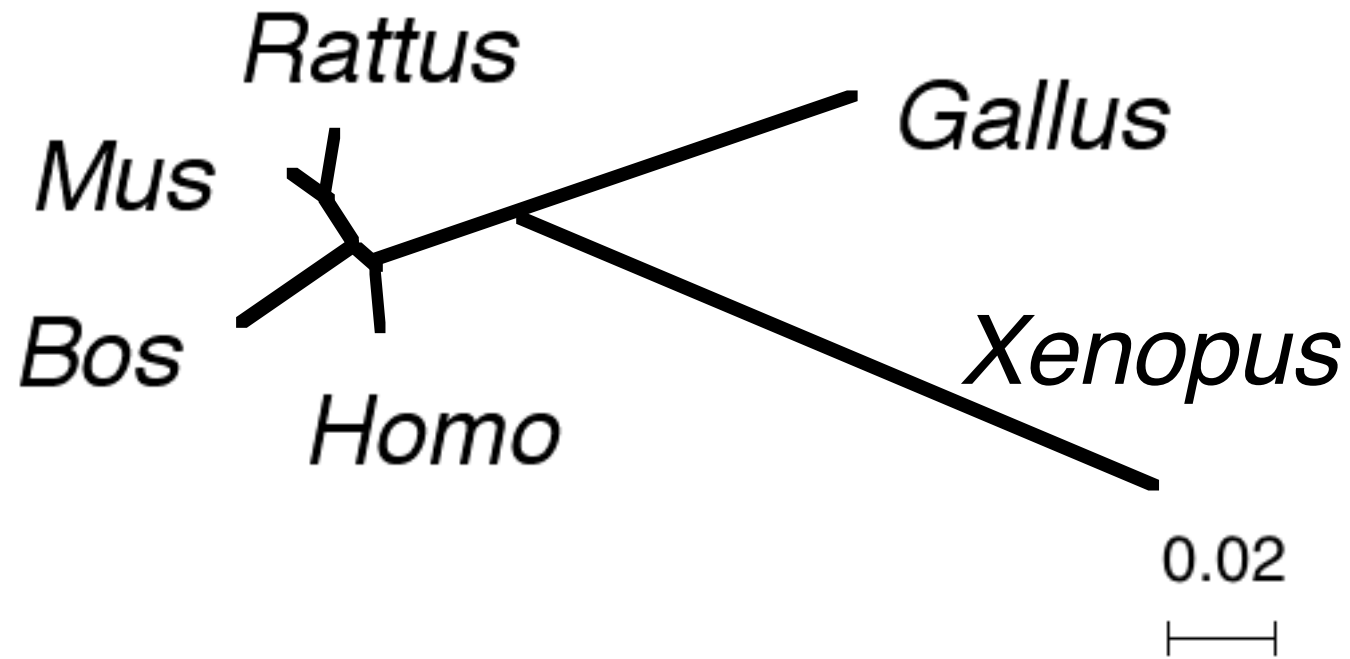




# Arbre raciné



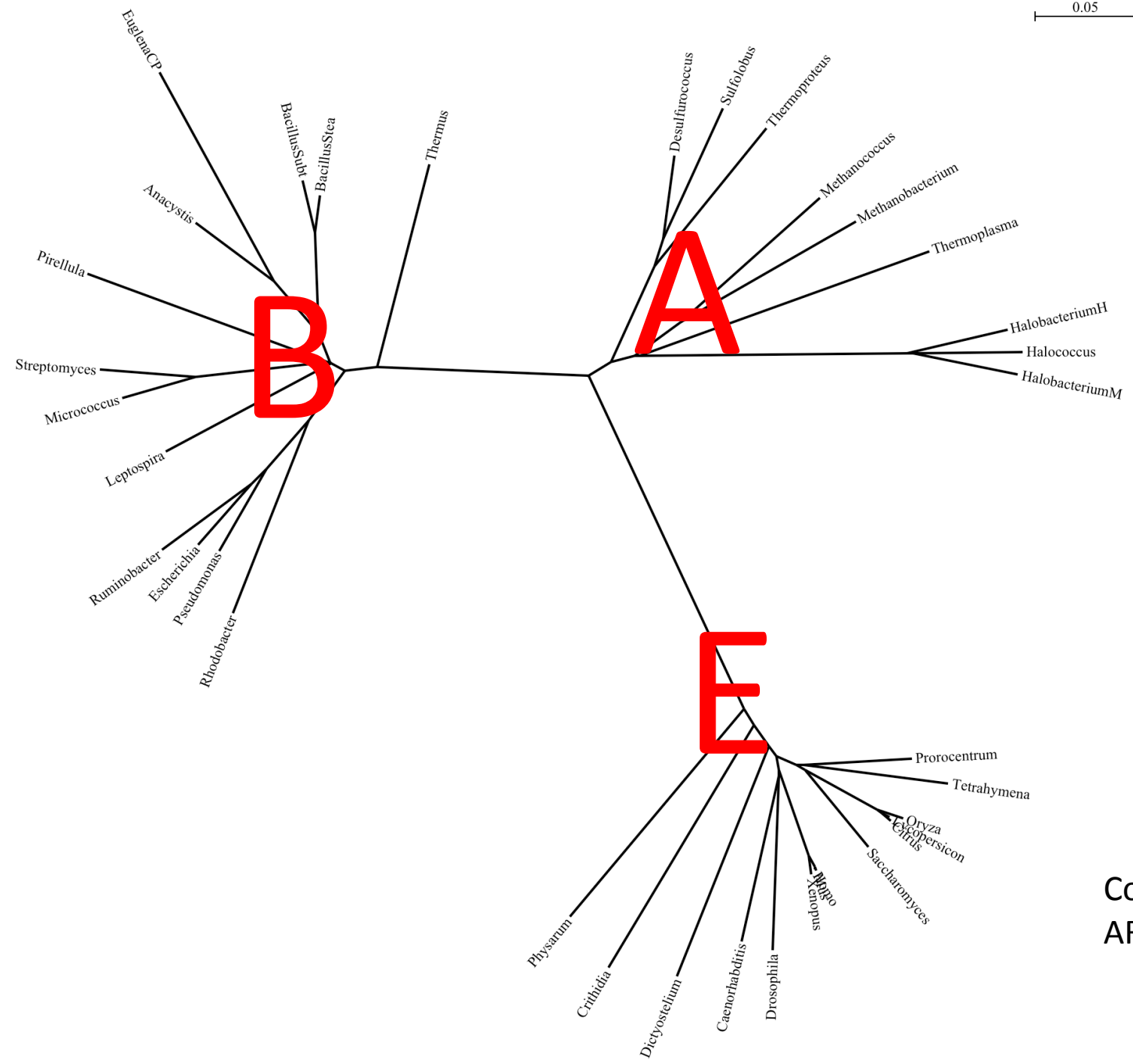
# Arbre non raciné



## Arbres racinés et non-racinés

- La plupart des méthodes phylogénétiques produisent des arbres non racinés. La raison est que les méthodes détectent des différences entre séquences, sans avoir le moyen de les orienter temporellement.
- Deux façons d'enraciner un arbre non raciné:
  - Méthode du groupe externe : inclure dans l'analyse un groupe de séquences dont on sait *a priori* qu'elles sont externes au groupe étudié; la racine est sur la branche qui relie le groupe externe aux autres séquences.
  - Faire l'hypothèse de l'horloge moléculaire : toutes les lignées sont supposées évoluer à la même vitesse depuis leur divergence; la racine est au point de l'arbre équidistant de toutes ses feuilles.

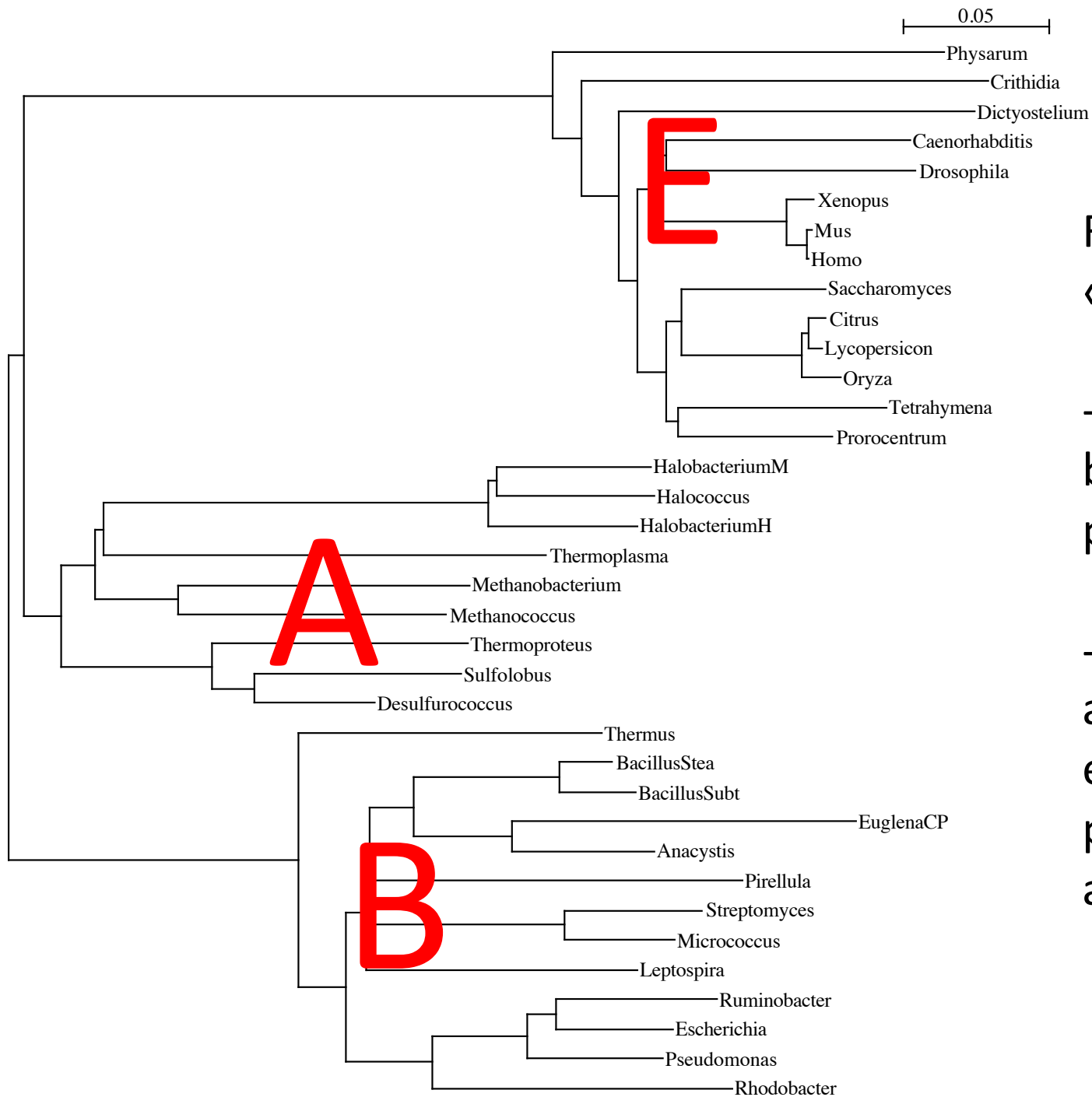
0.05



Phylogénie universelle non racinée

Trois domaines  
A Archées  
B Bactéries  
E Eucaryotes

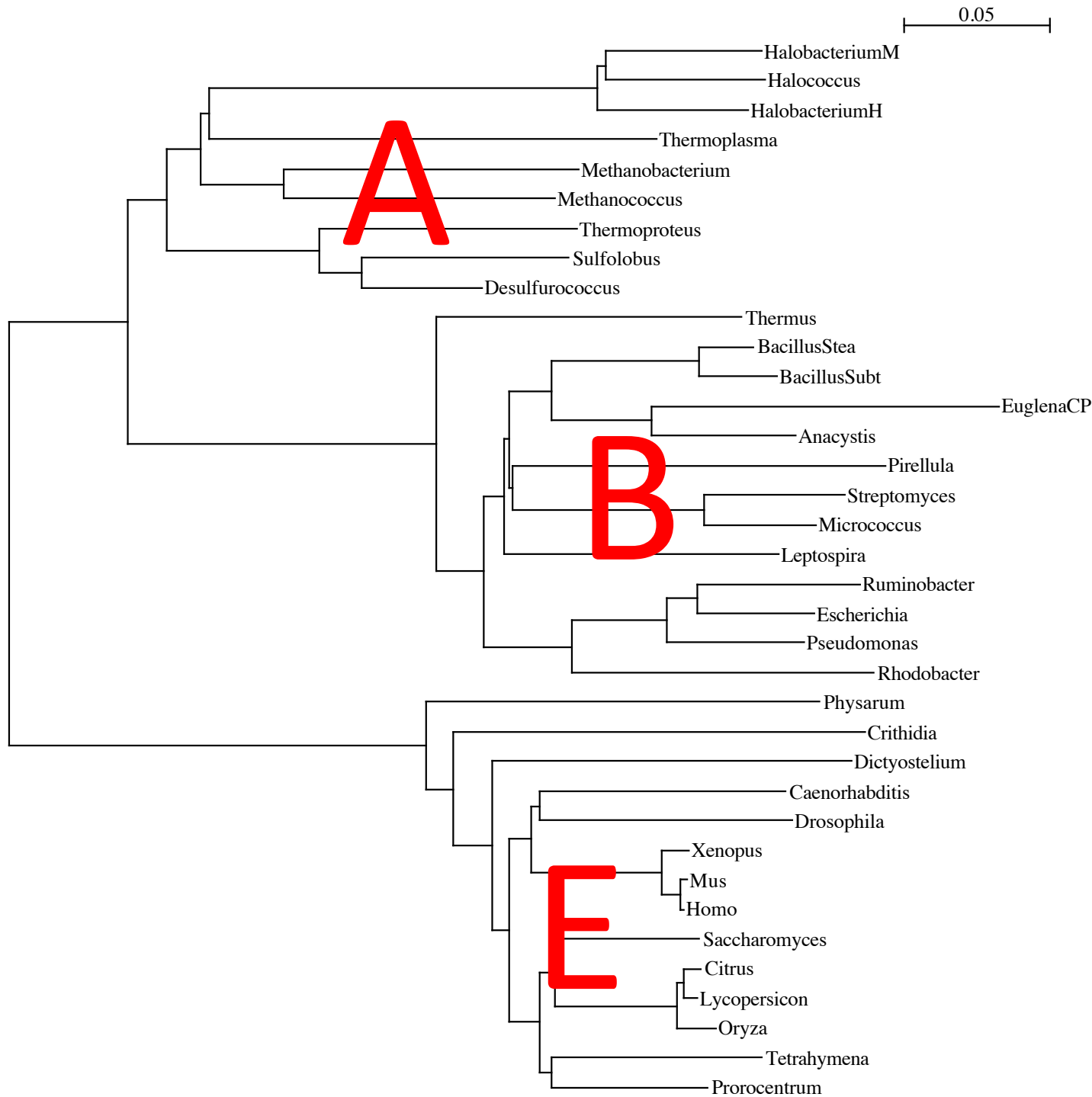
Concaténation des ARNr SSU et LSU



Racinement  
« standard »:

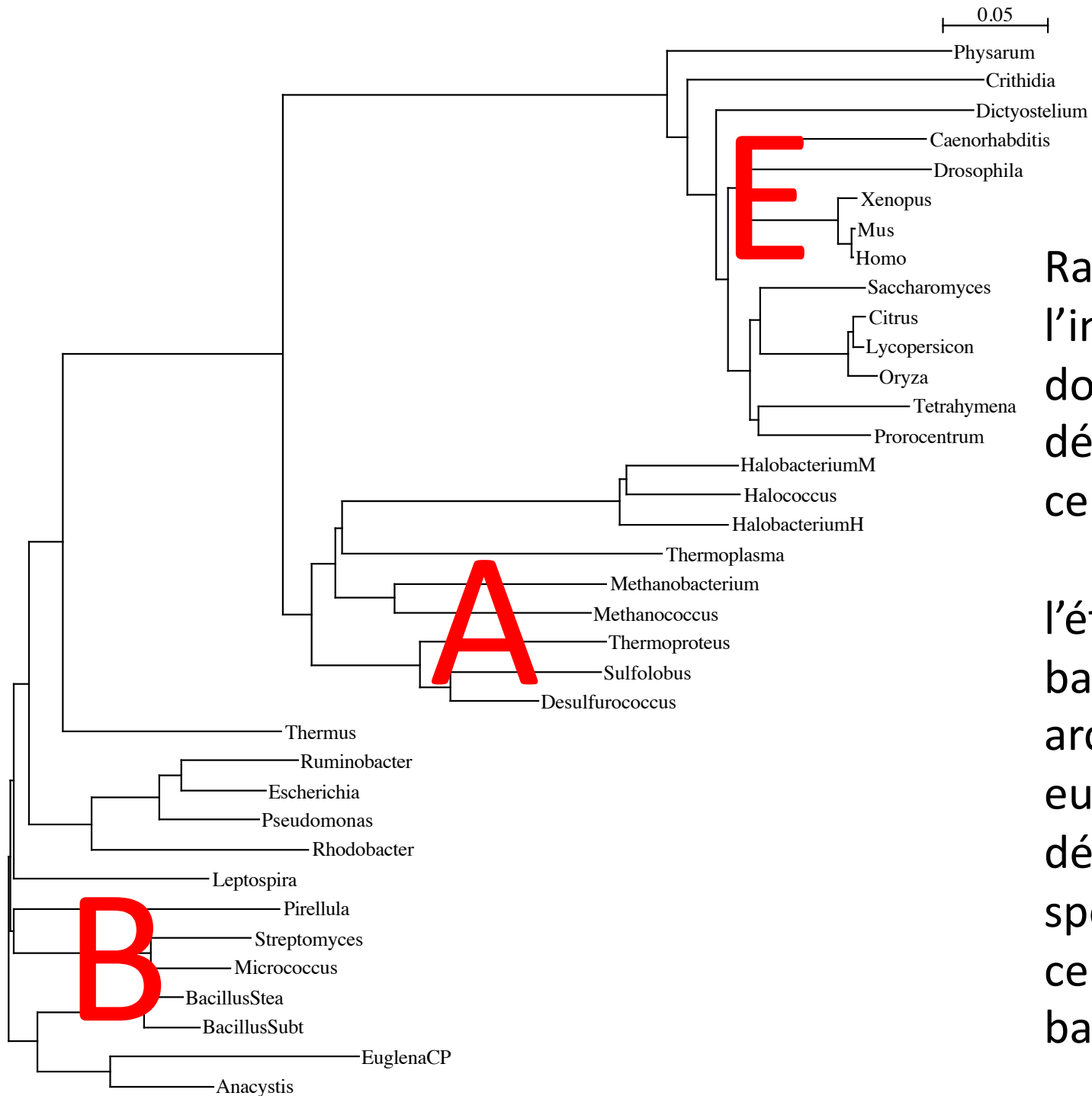
- le domaine  
bactérien est le  
premier à diverger

- les domaines  
archéen et  
eucaryote  
possèdent un  
ancêtre exclusif



Racinement par  
le centre :

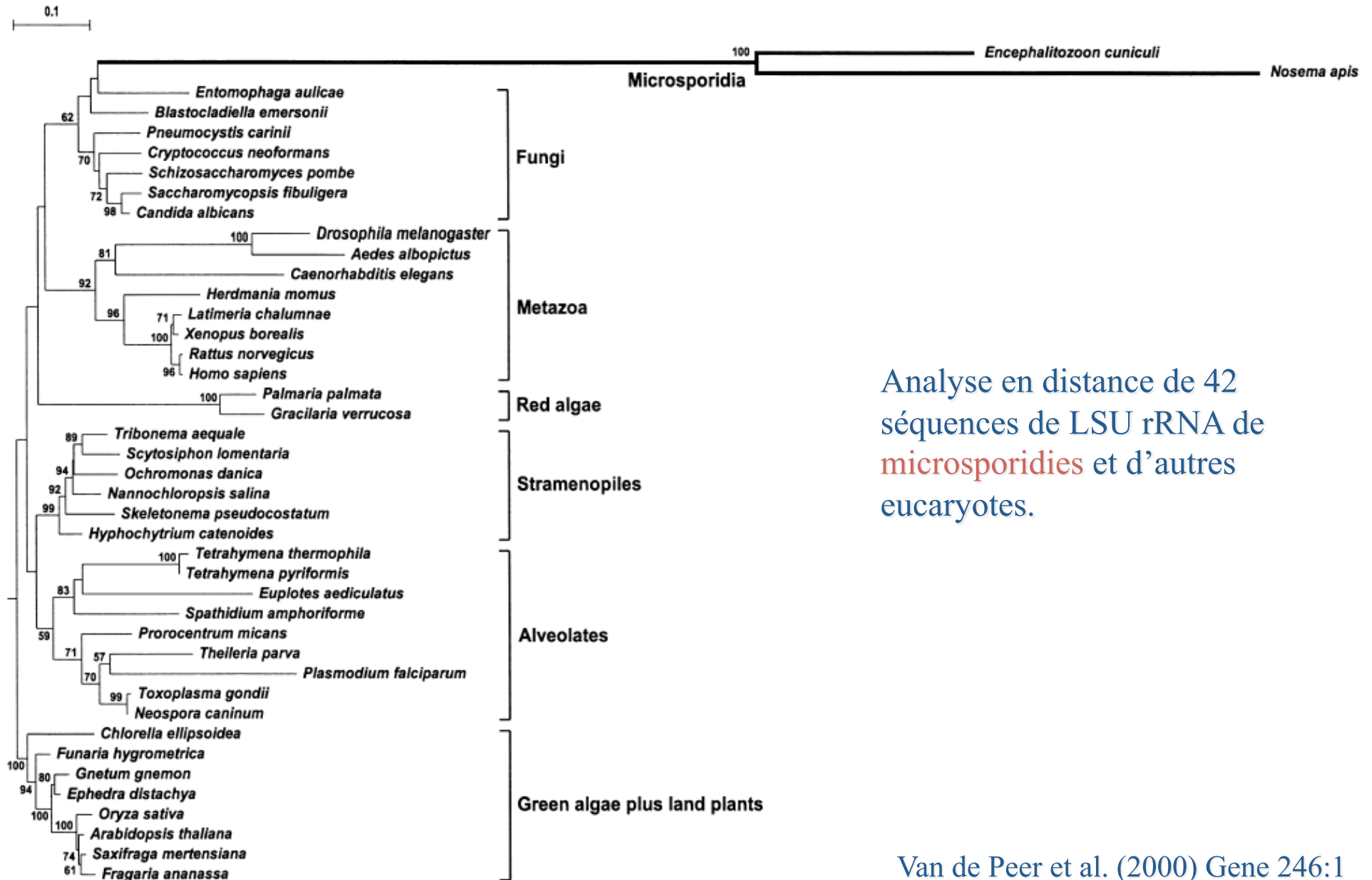
opposition  
procaryote/  
eucaryote



Racinement à l'intérieur du domaine bactérien défendu par certains auteurs :

l'état ancestral est bactérien, les archées et les eucaryotes sont dérivés spécifiquement de certains types de bactéries.

# Racinement par le centre: incorrect si fortes différences de vitesse entre lignées



Analyse en distance de 42 séquences de LSU rRNA de **microsporidies** et d'autres eucaryotes.



## Arbre des gènes vs. Arbre des espèces

- L'histoire évolutive des gènes reproduit celle des espèces qui les portent, sauf si:
  - Transfert horizontal = transfert de gène entre espèces
  - Duplication génique : orthologie/ paralogie

# Orthologie / Paralogie

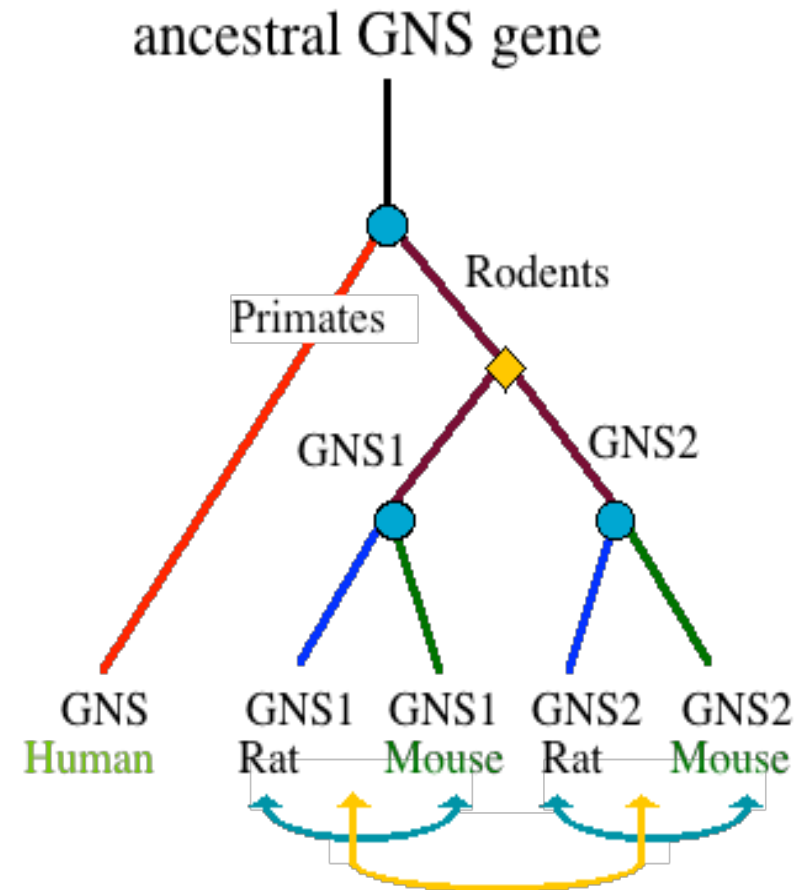
- speciation
- ◆ duplication

*Homology* : two genes are homologous iff they have a common ancestor.

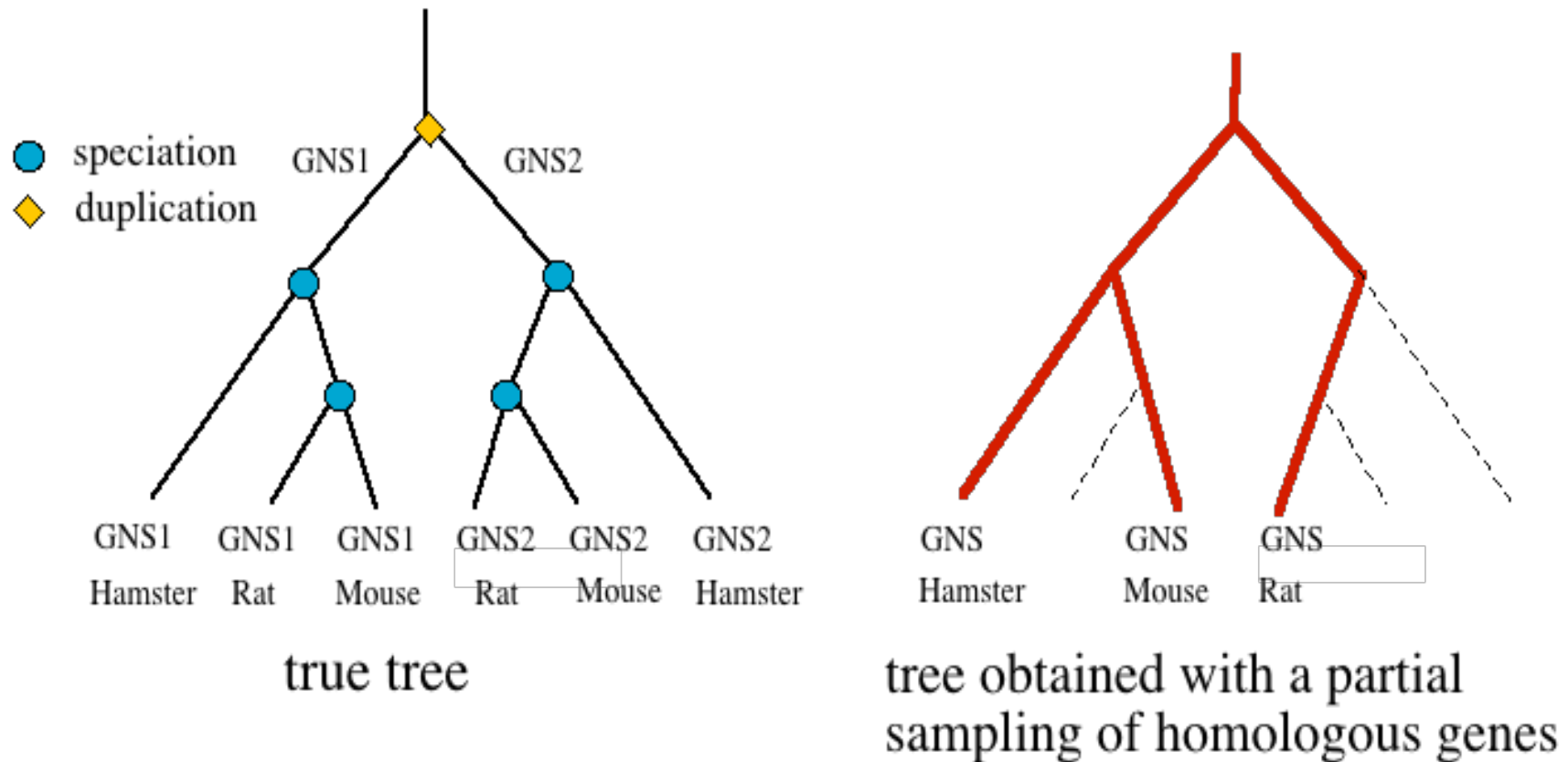
↔ *Orthology* : two genes are orthologous iff they diverged following a speciation event.

↔ *Paralogy* : two genes are paralogous iff they diverged following a duplication event.

⚠ Orthology ≠ functional equivalence

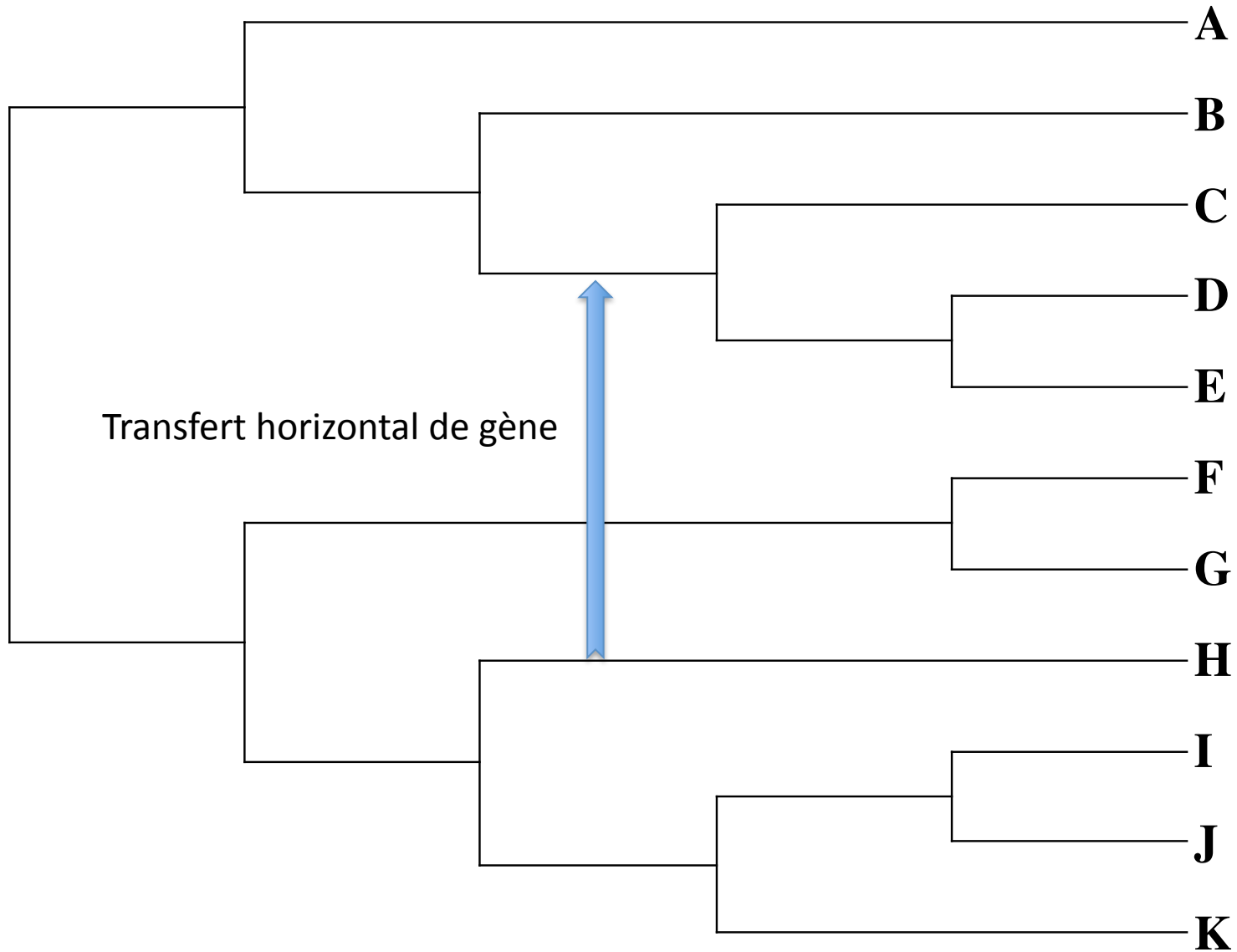


# Reconstruction de la phylogénie des espèces: artéfacts dus à la paralogie

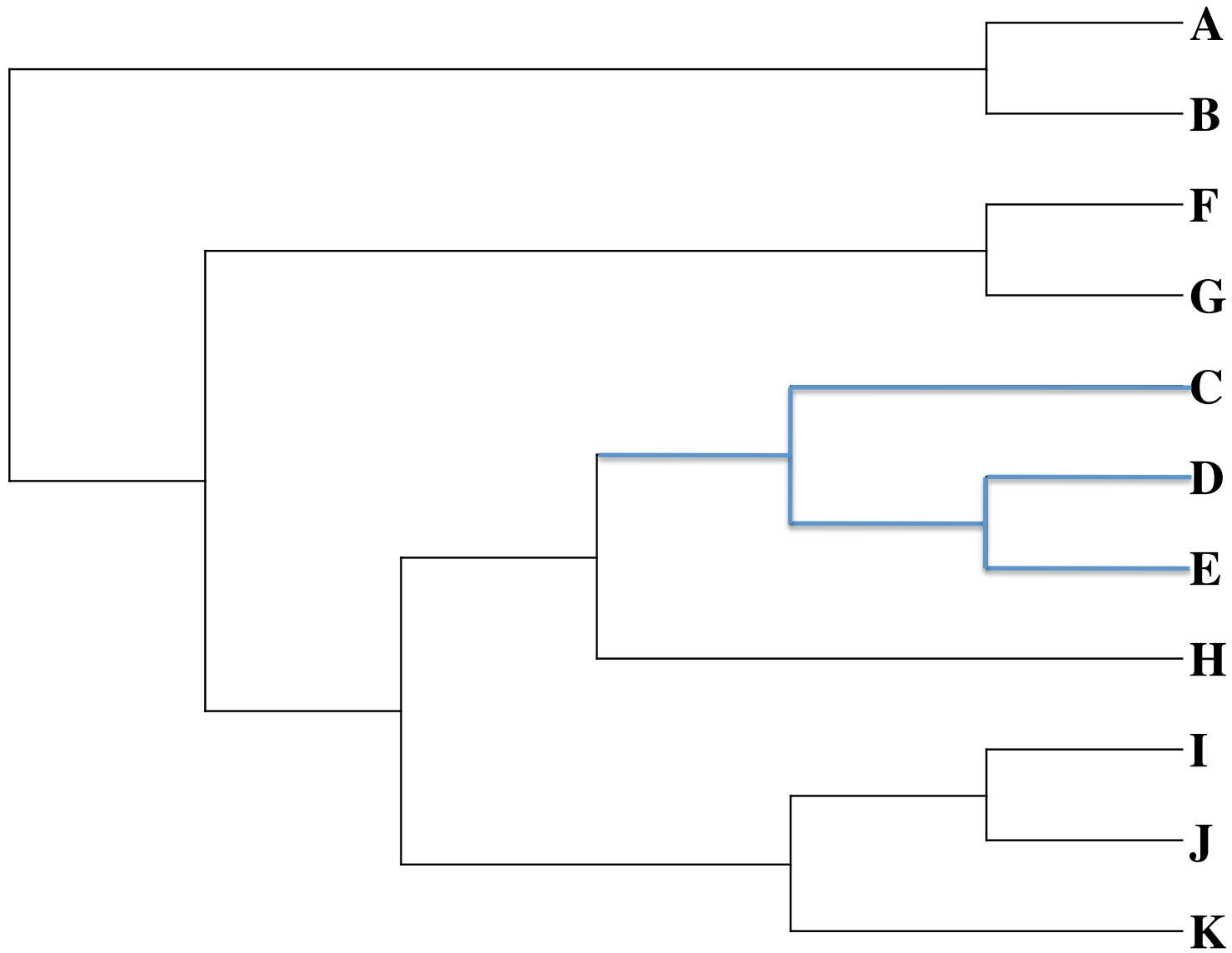


!! Des pertes de gènes peuvent se produire au cours de l'évolution : même avec des séquences génomiques complètes, il peut être difficile de détecter la paralogie !!

# Arbre des espèces

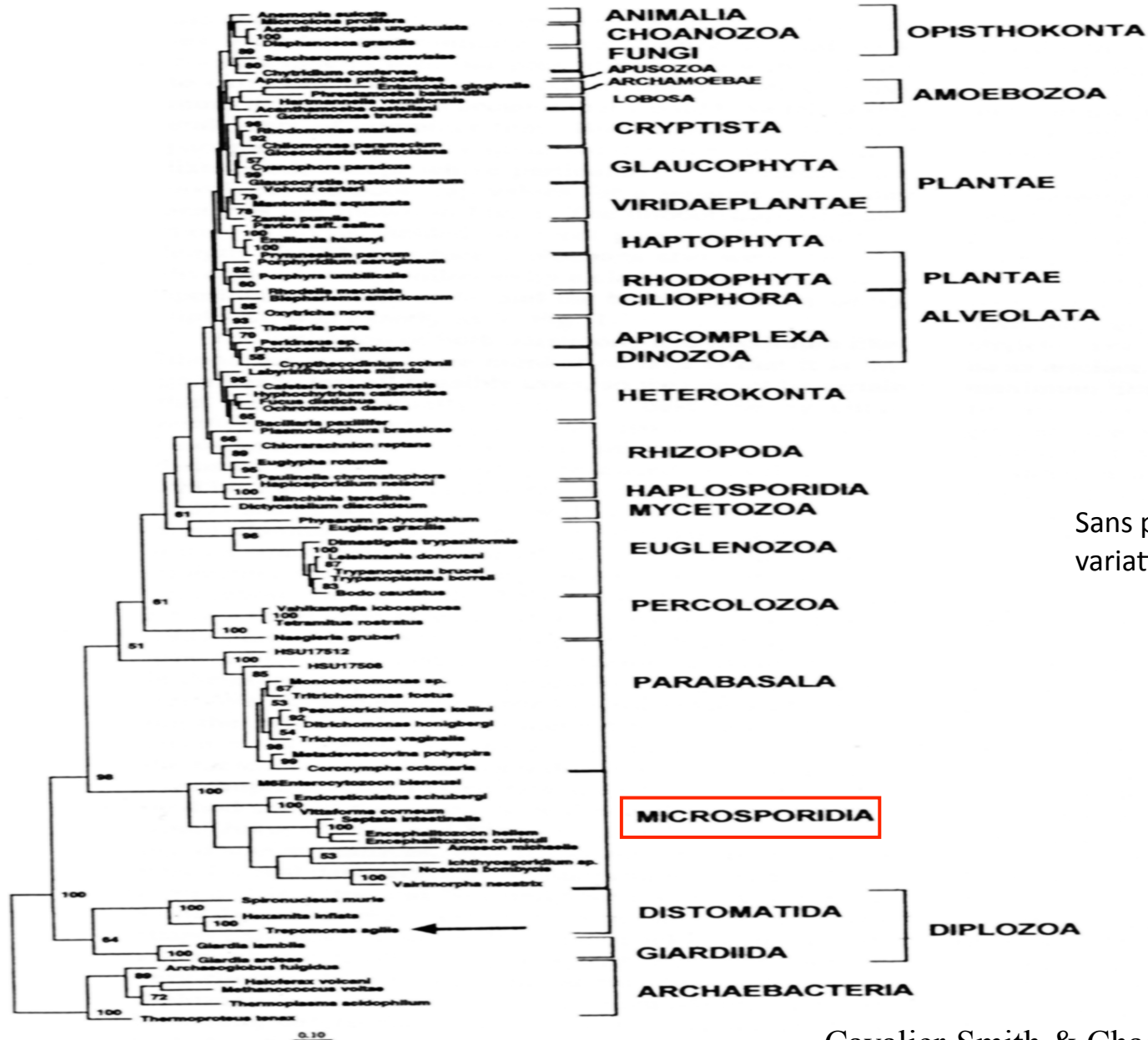


# Arbre du gène



L'artefact d'attraction des longues  
branches

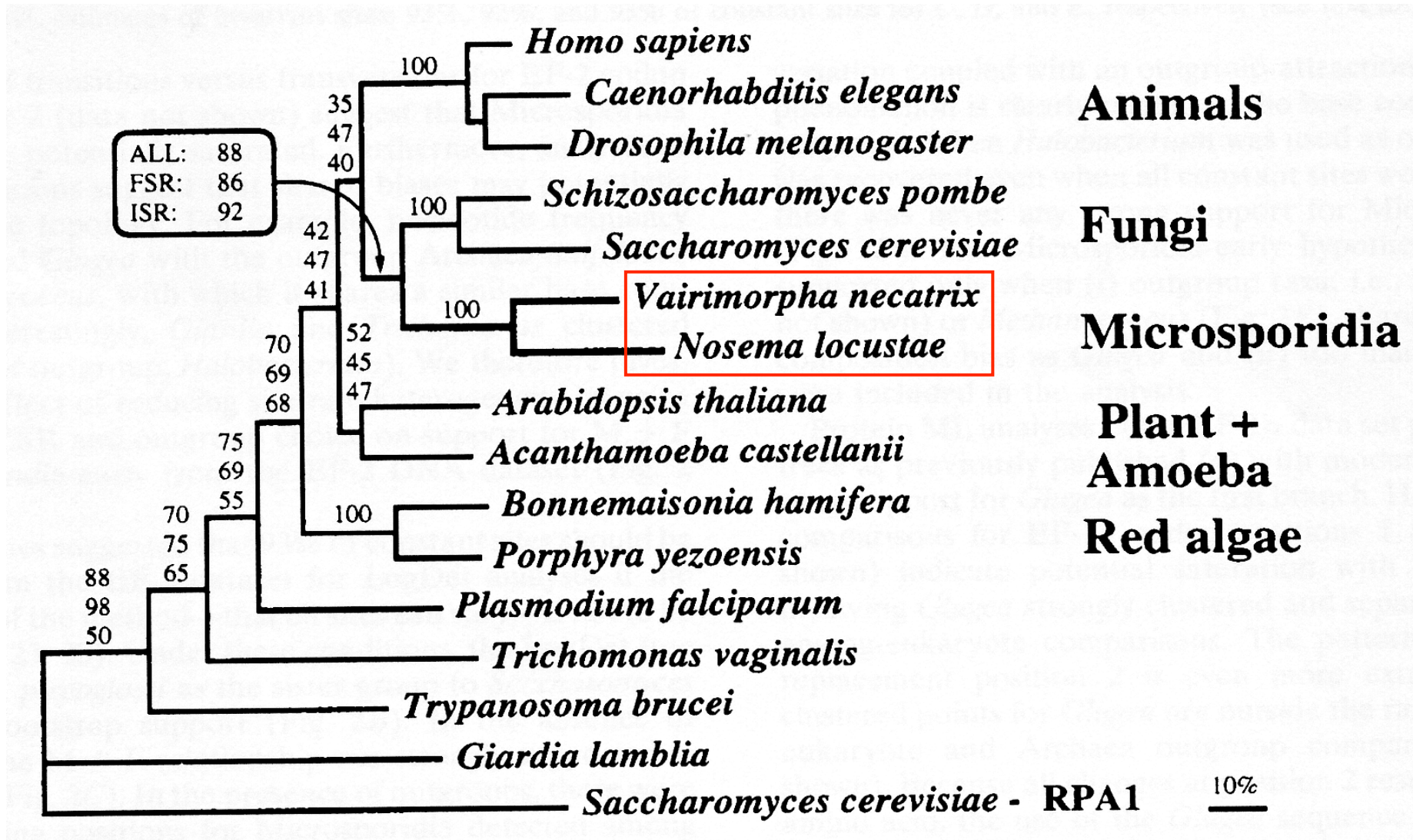
# Phylogenetic analysis of eukaryotic small subunit ribosomal RNA



Sans prise en compte de la variation des vitesses entre sites

# Phylogenetic analysis of RNA polymerase II large subunit

Hirt *et al.* (1999) Proc.Natl.Acad.Sci. USA 96:580



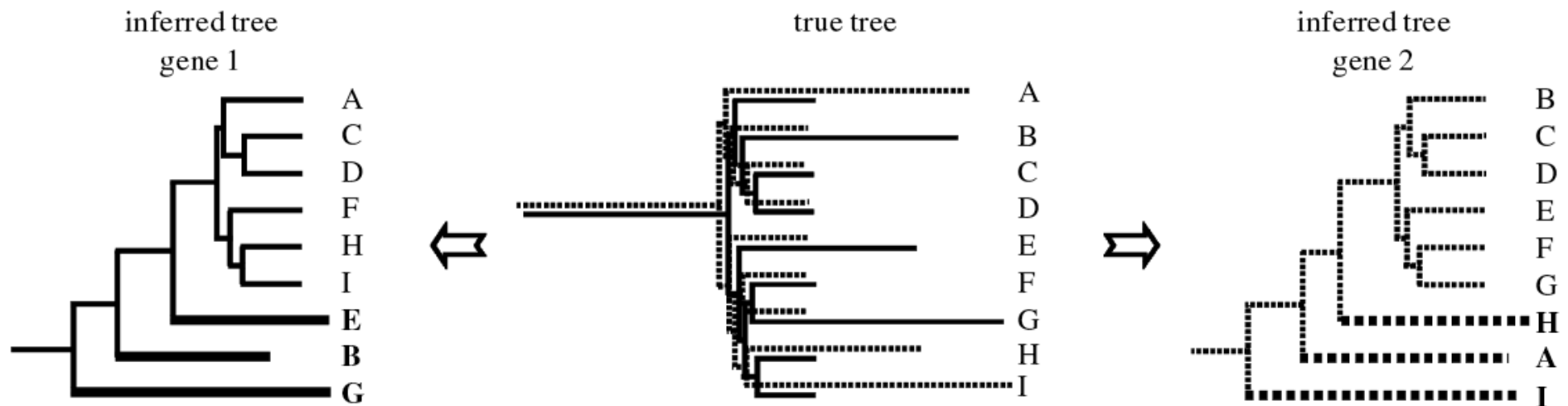
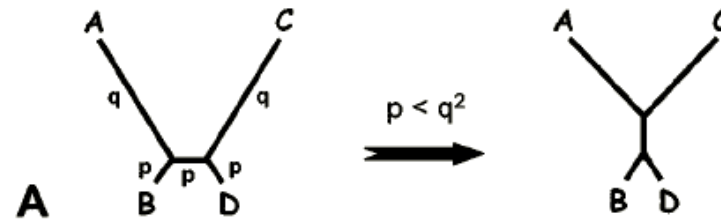


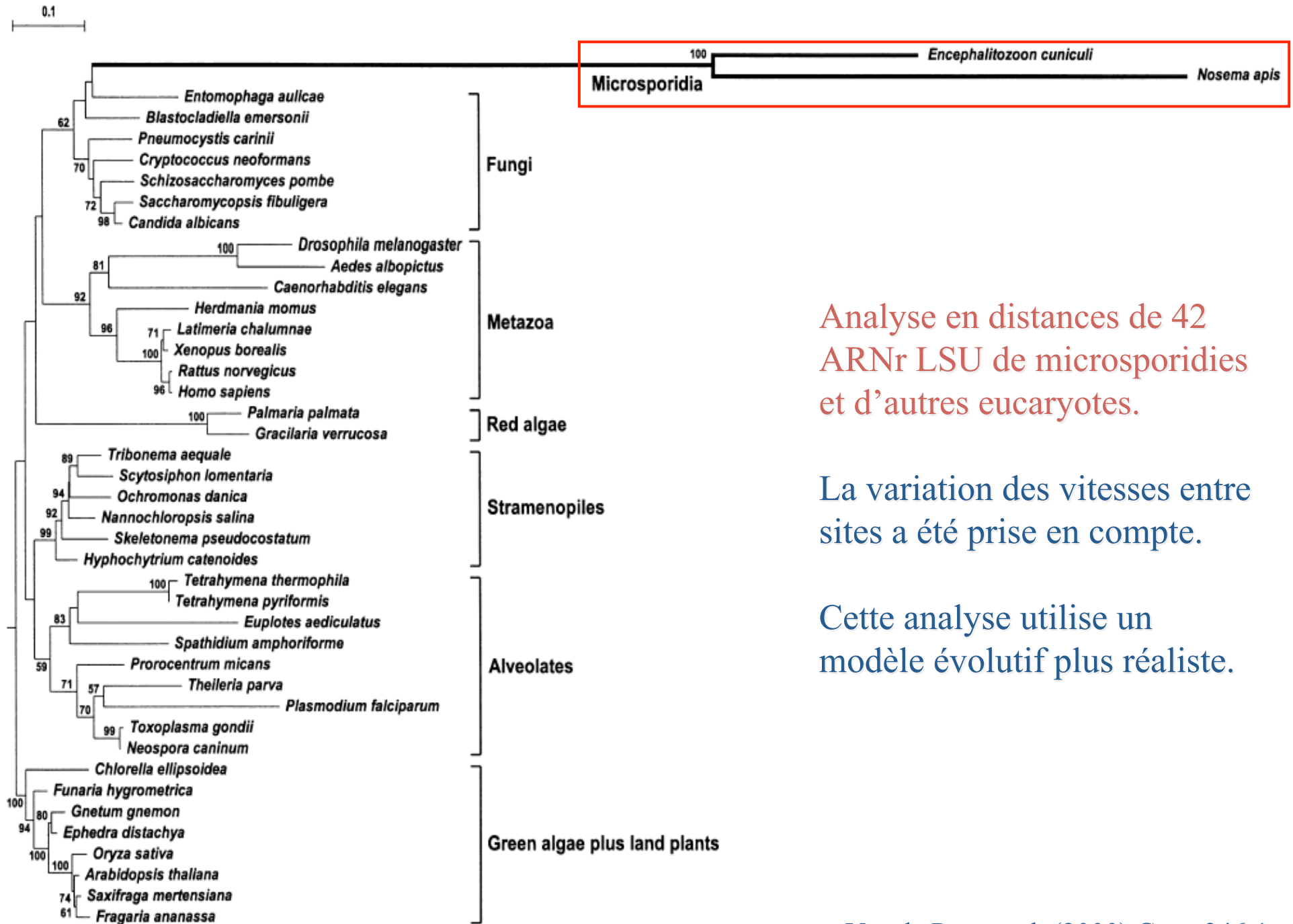
# Peut-on réconcilier les ARNr et les ARN polymérases ?



# The Long Branch Attraction artifact

[ Felsenstein (1978) *Syst Zool* 27:401 ]





Analyse en distances de 42 ARNr LSU de microsporidies et d'autres eucaryotes.

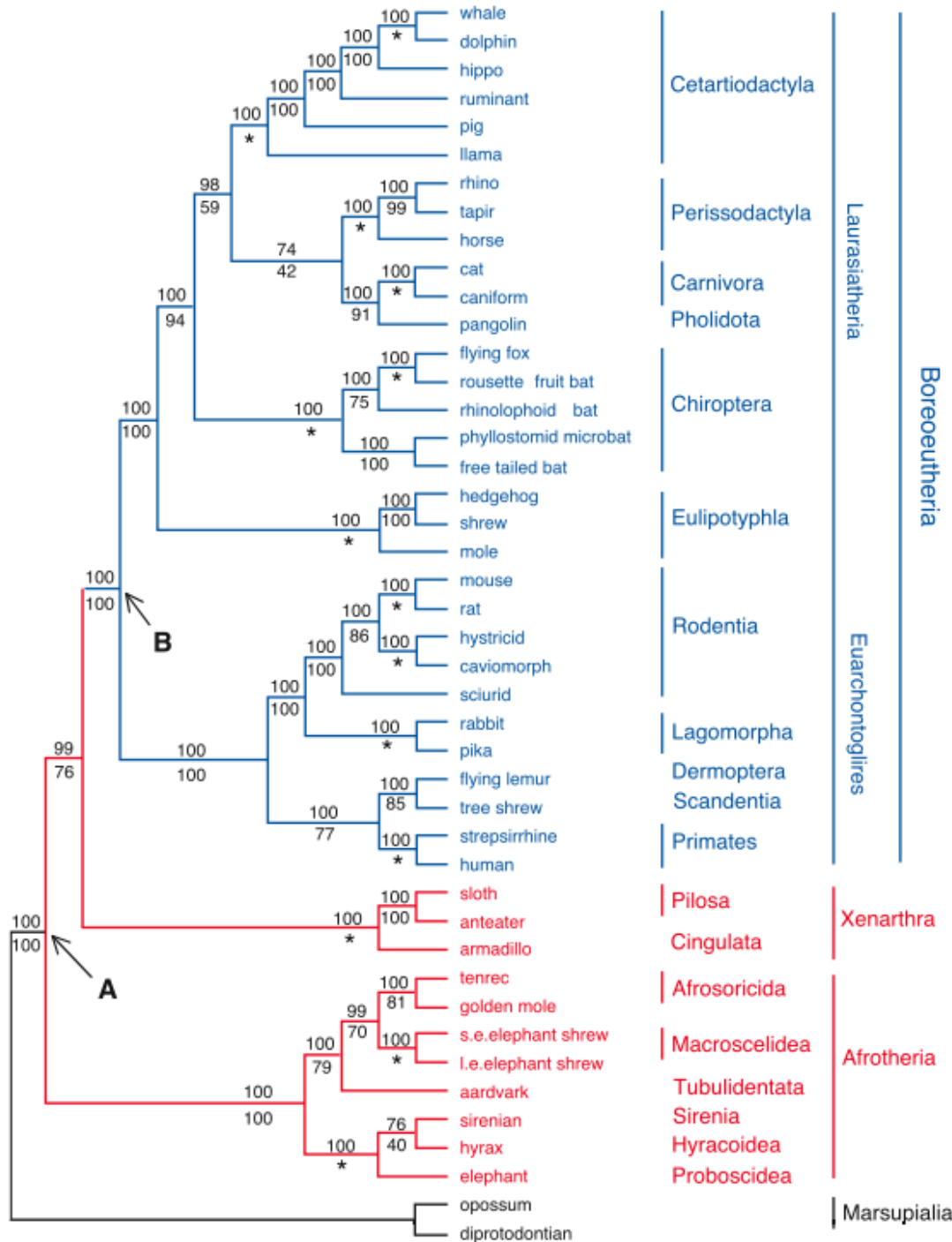
La variation des vitesses entre sites a été prise en compte.

Cette analyse utilise un modèle évolutif plus réaliste.

L'effet du modèle évolutif utilisé:

les modèles plus réalistes sont meilleurs

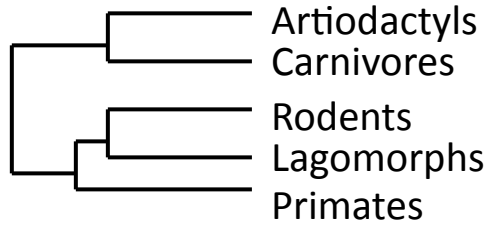
# Phylogénie des ordres de mammifères



## Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics

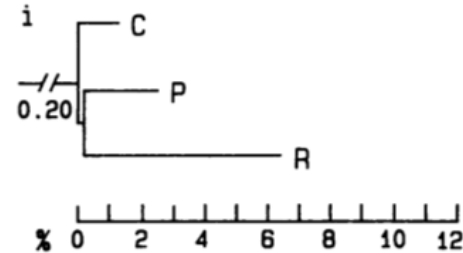
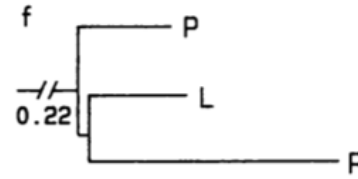
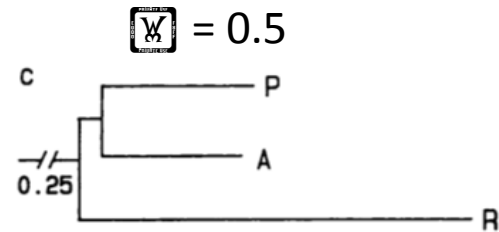
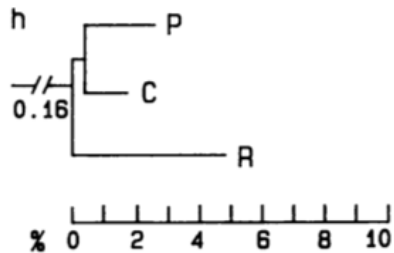
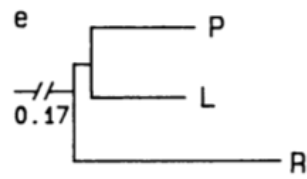
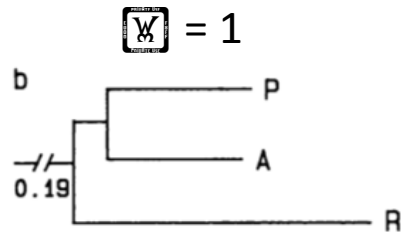
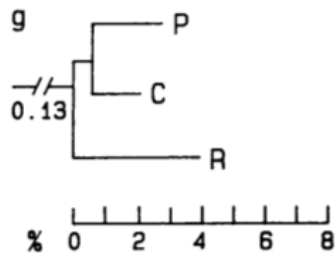
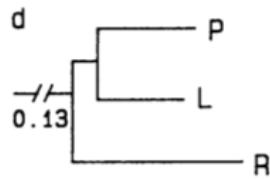
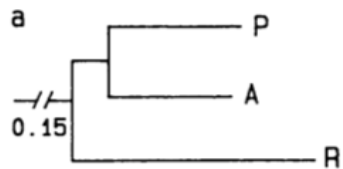
William J. Murphy,<sup>1\*</sup> Eduardo Eizirik,<sup>1,2\*</sup> Stephen J. O'Brien,<sup>1†</sup> Ole Madsen,<sup>3</sup> Mark Scally,<sup>4,5</sup> Christophe J. Douady,<sup>4,5</sup> Emma Teeling,<sup>4,5</sup> Oliver A. Ryder,<sup>6</sup> Michael J. Stanhope,<sup>5,7</sup> Wilfried W. de Jong,<sup>3,8</sup> Mark S. Springer<sup>4†</sup>

True tree :



with across-site rate variation

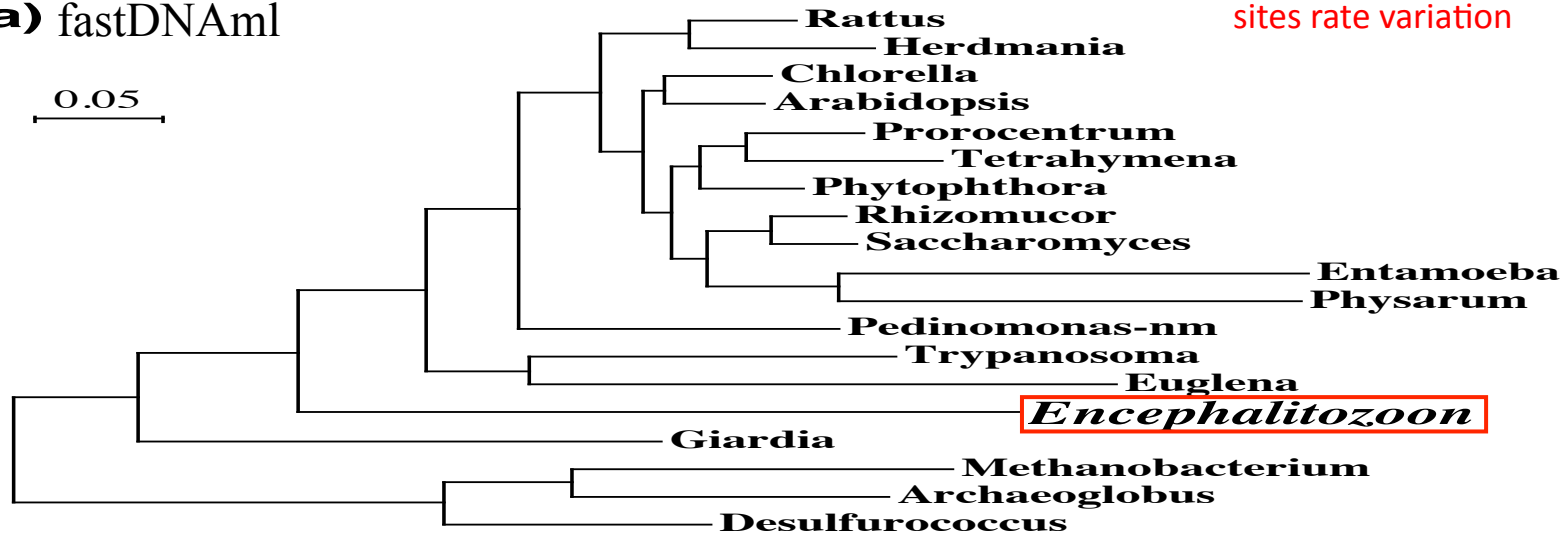
no across-site  
rate variation



# Phylogenetic analysis of LSU rRNA

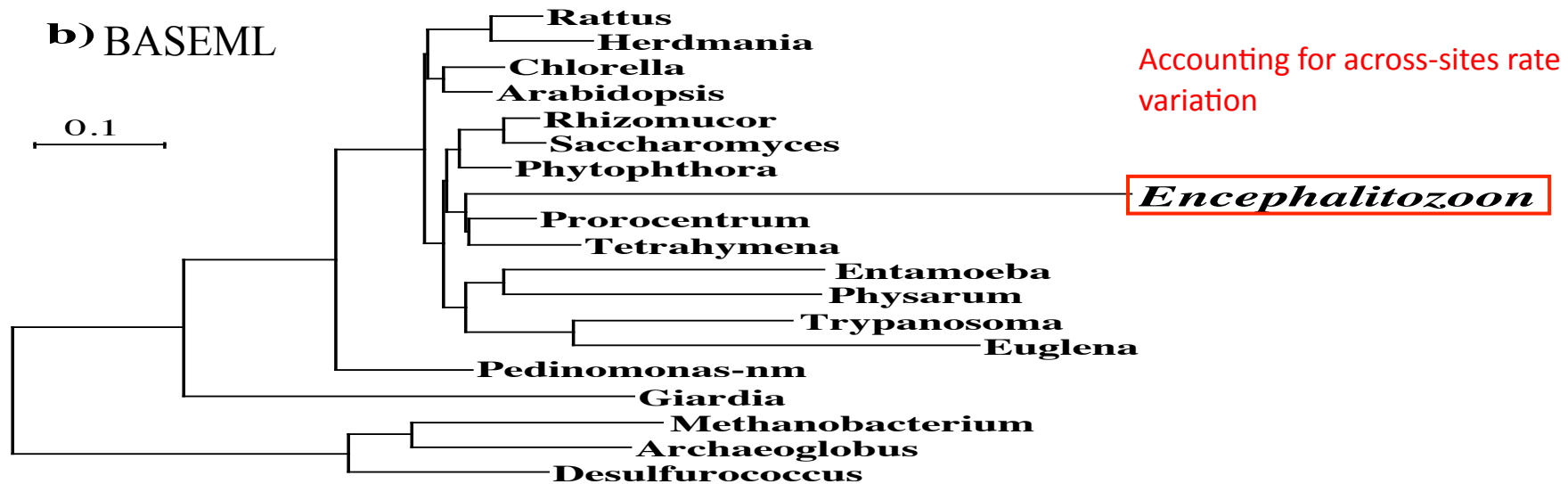
a) fastDNAmI

0.05



b) BASEML

0.1



# L'effet de l'échantillonnage taxonomique



“Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi”  
Gouy & Li (1989)

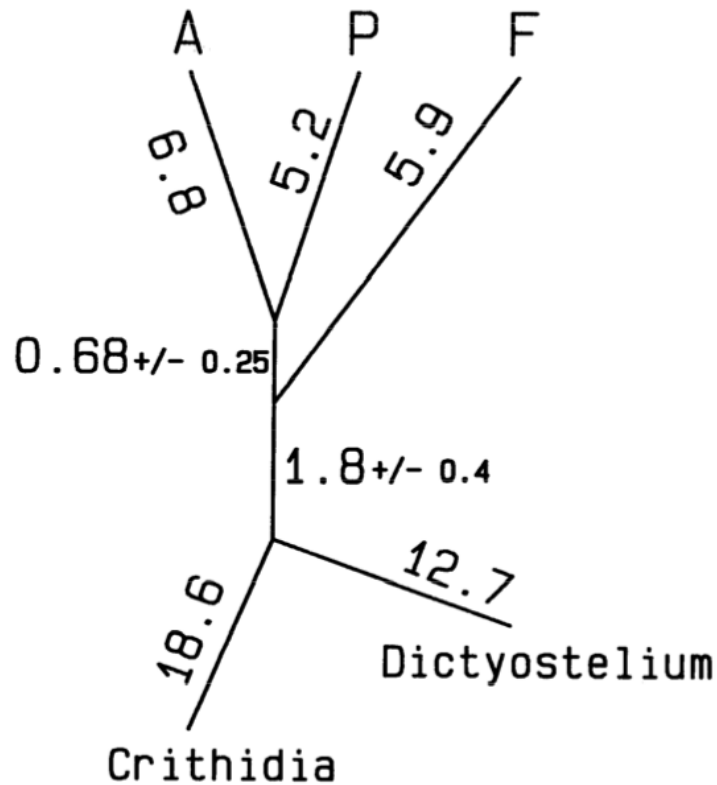


FIG. 2.—Unrooted phylogenetic tree inferred from rRNA sequences. A total of 2,971 sites were analyzed.

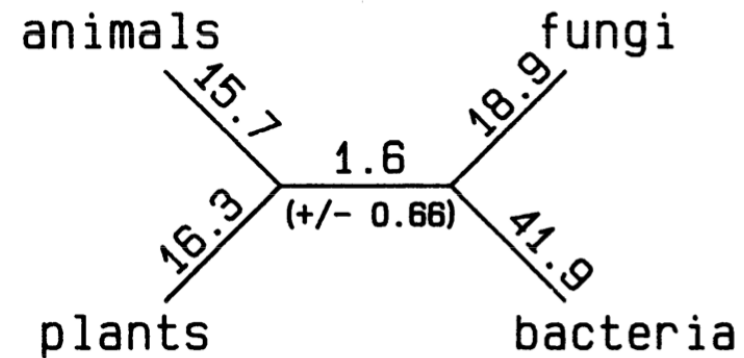
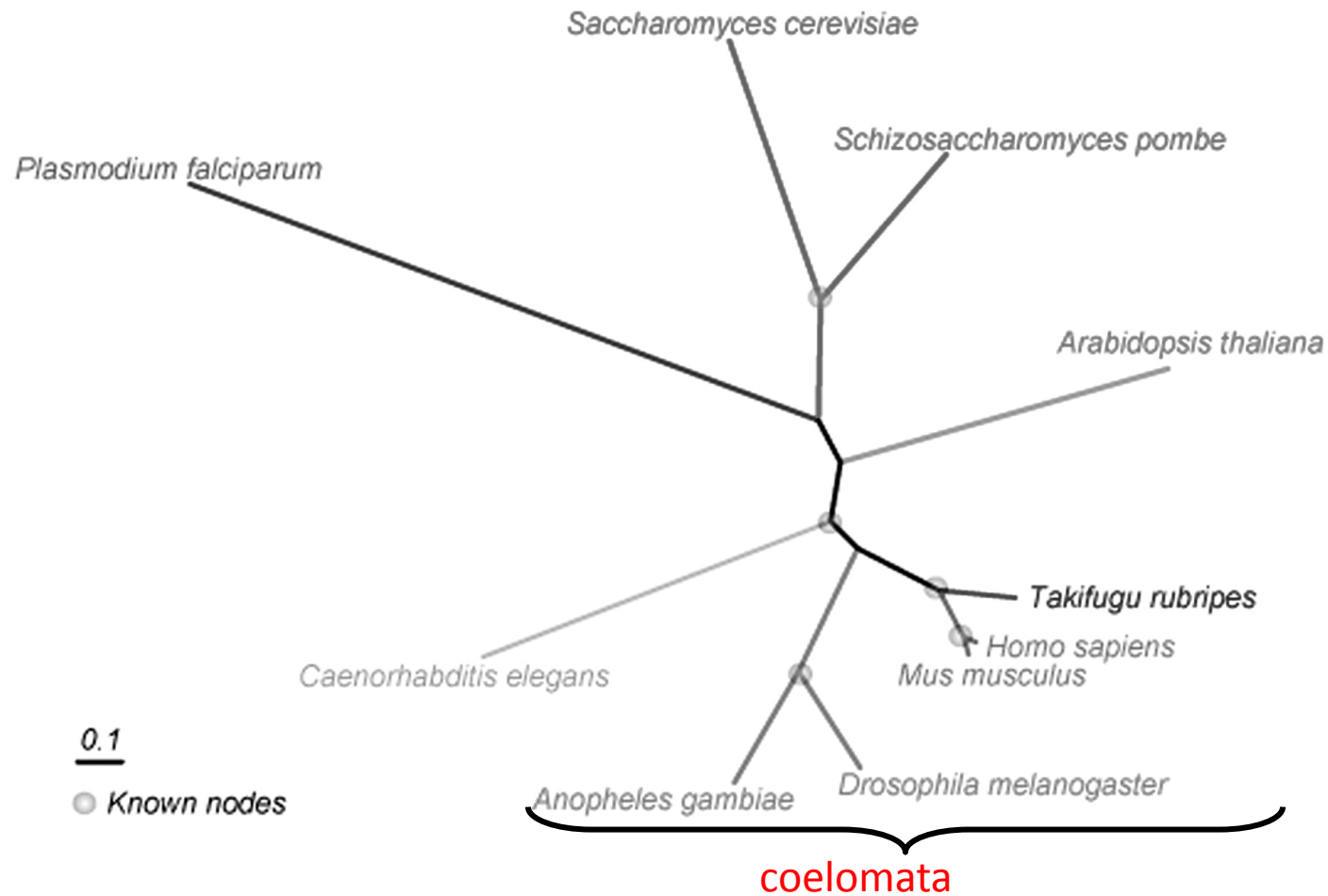


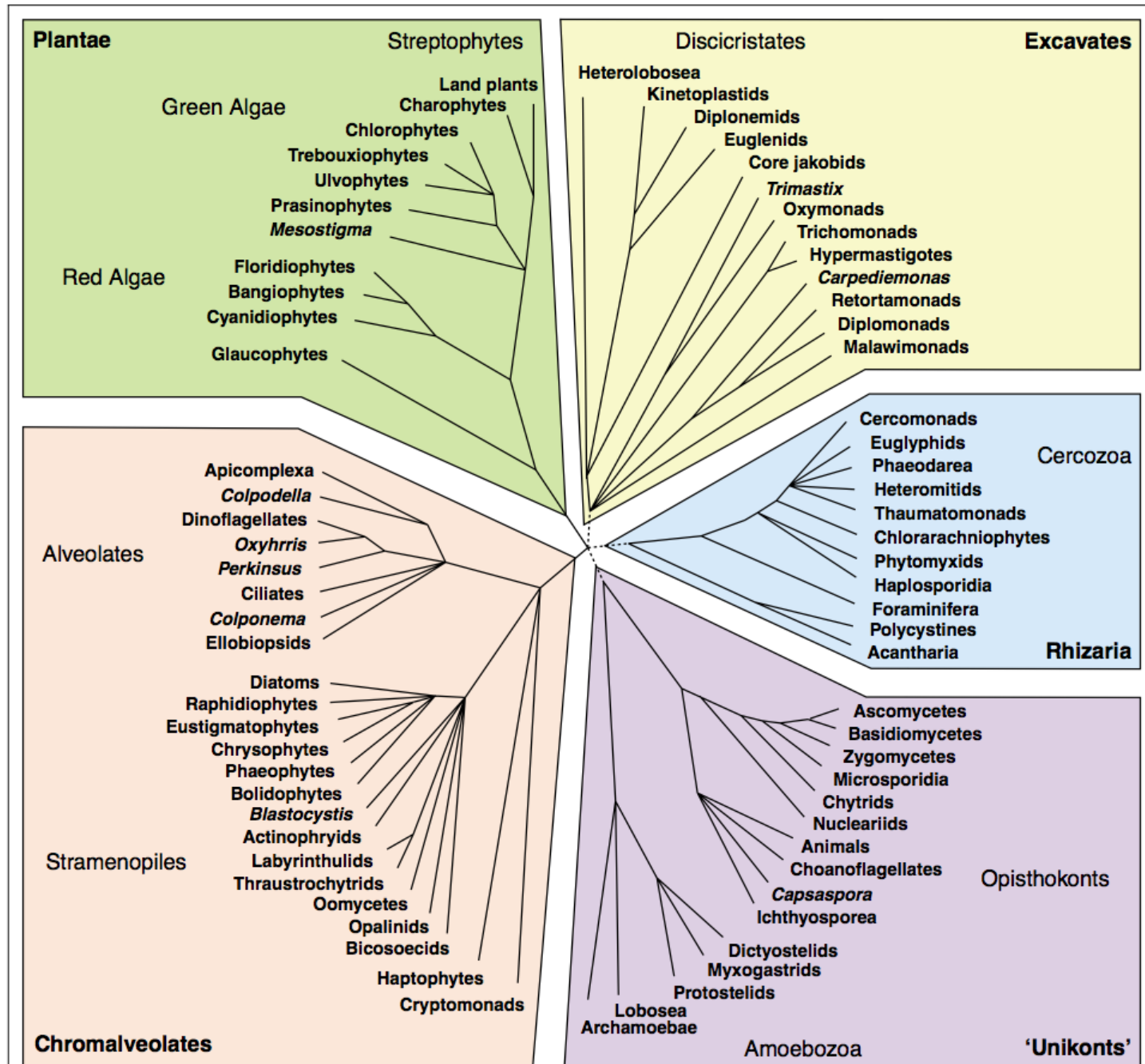
FIG. 4.—Unrooted phylogenetic tree inferred from the pooled protein data set. A total of 1,634 sites were analyzed. Branch lengths are in percent substitutions. The SE of the internal branch length estimate is shown. See table 3 for a description of the data set.

Les temps reculés: à la fois peu de gènes et peu d'espèces

“The Opisthokonta and the Ecdysozoa May Not Be Clades: Stronger Support for the Grouping of Plant and Animal than for Animal and Fungi and Stronger Support for the Coelomata than Ecdysozoa”



Arbres phylogénétiques de 780 gènes de 10 génomes complètement séquencés amalgamés en un unique superarbre.



Eukaryotic domain phylogeny.

Emerging consensus for the identification of five super- phyla.

Relationships between them remain very uncertain.

## Phylogénie des métazoaires

- Rejet du concept acoelomate, pseudocoelomate, coelomate :

la division pertinente est lophotrochozoa / ecdysozoa

- Bilateria vs. cnidaria, porifera et ctenophora

- Protostomes vs. Deutérostomes

