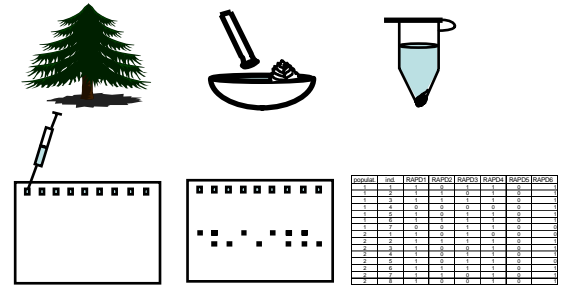


## 3. Phylogénie moléculaire

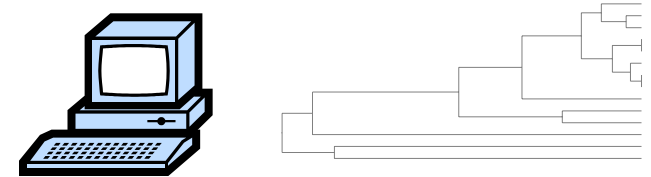
- 3.1. Introduction: l'arbre phylogénétique
- 3.2. Les méthodes des distances génétiques
- 3.3. Les méthodes de maximum de parcimonie
- 3.4. Les méthodes de maximum de vraisemblance
- 3.5. Test de robustesse des reconstructions phylogénétiques

# 3. Phylogénie moléculaire

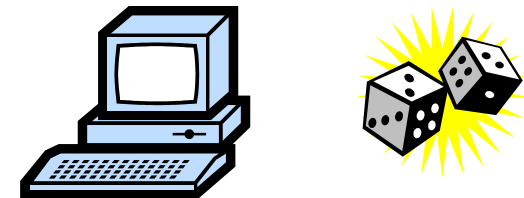
Acquisition de données moléculaires  
(séquences d'ADN)



Reconstructions phylogénétiques

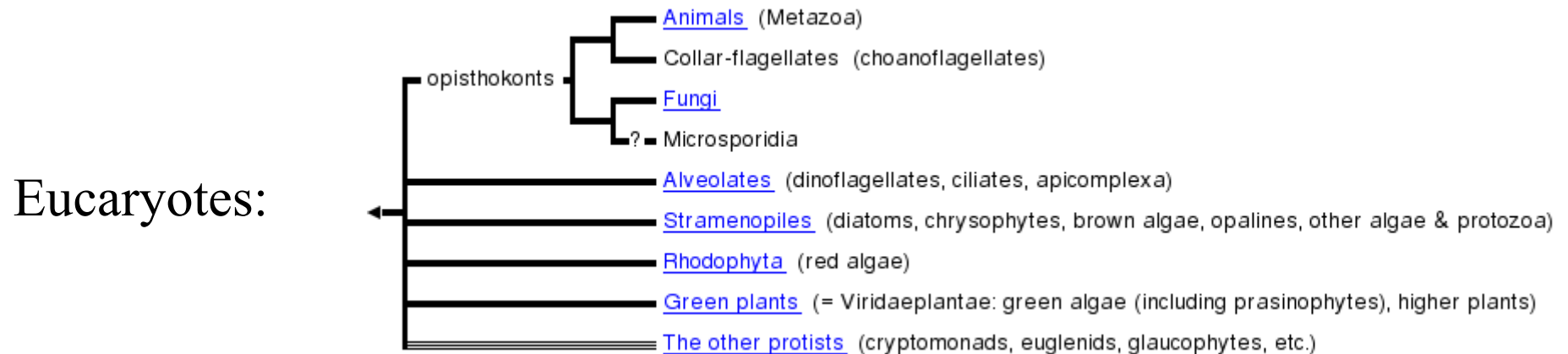
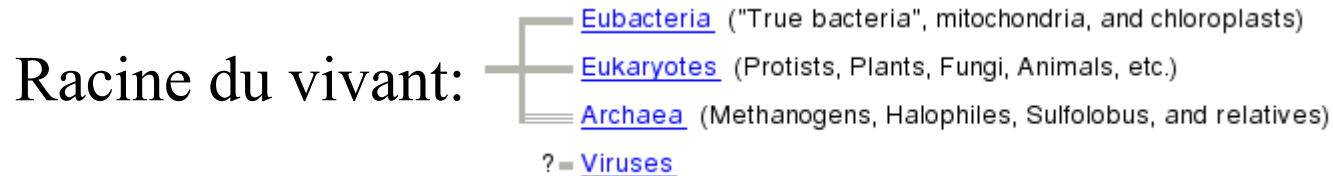


Test de robustesse des reconstructions  
phylogénétiques



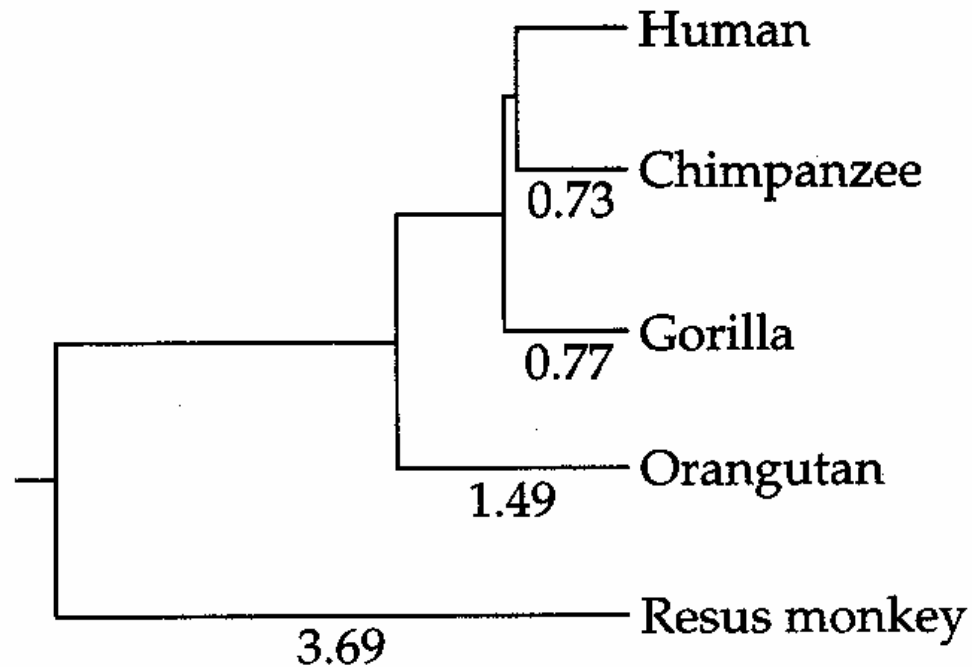
# 3.1. Introduction: l'arbre phylogénétique

Principe de base: **évolution biologique**  $\Rightarrow$  tous êtres vivants ont un ancêtre commun  $\Rightarrow$  sont reliés par des branches dans l'**arbre phylogénétique du vivant**  
(cf. "Tree of life": <http://tolweb.org/tree/phylogeny.html>)



# 3.1. Introduction: l'arbre phylogénétique

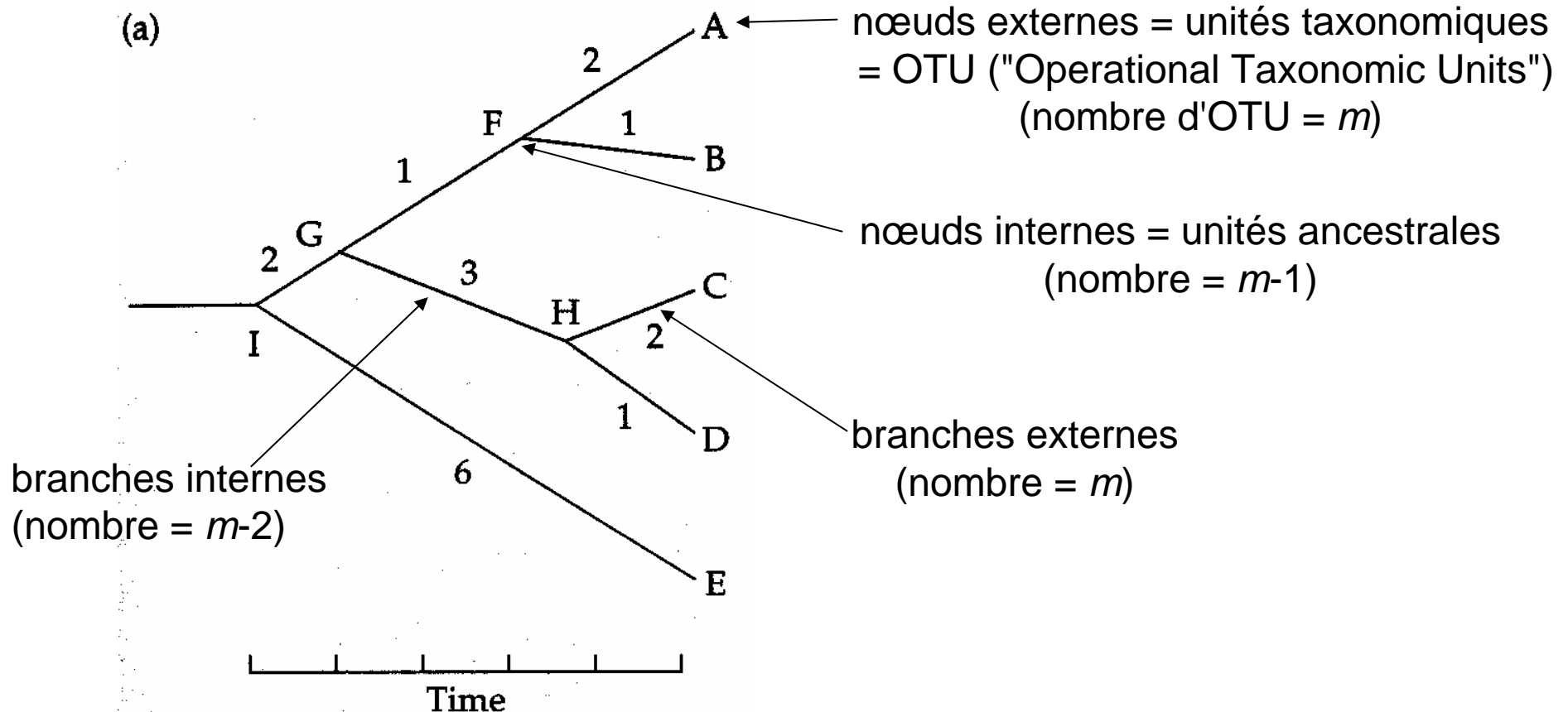
Comparaison des **séquences d'ADN homologues** au sein d'un échantillon d'espèces  $\Rightarrow$  reconstruction de la séquence évolutive  $\Rightarrow$  élaboration d'un **arbre phylogénétique inféré** (s'approchant le plus possible de la "vraie" séquence évolutive)



# 3.1. Introduction: l'arbre phylogénétique

## Représentation d'un arbre phylogénétique

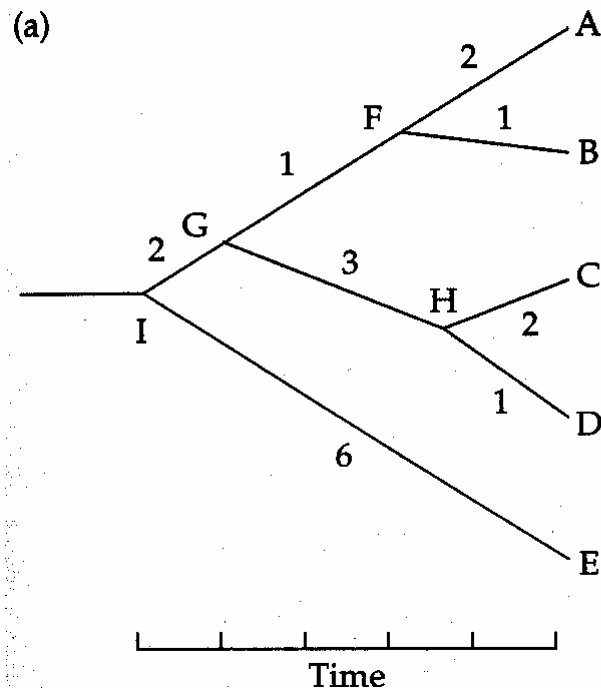
- **nœuds:** A, B, C,...I (nombre de nœuds =  $2m-1$ )
- **branches:** FA, FB, IG,... (nombre branches =  $2m-2$ )



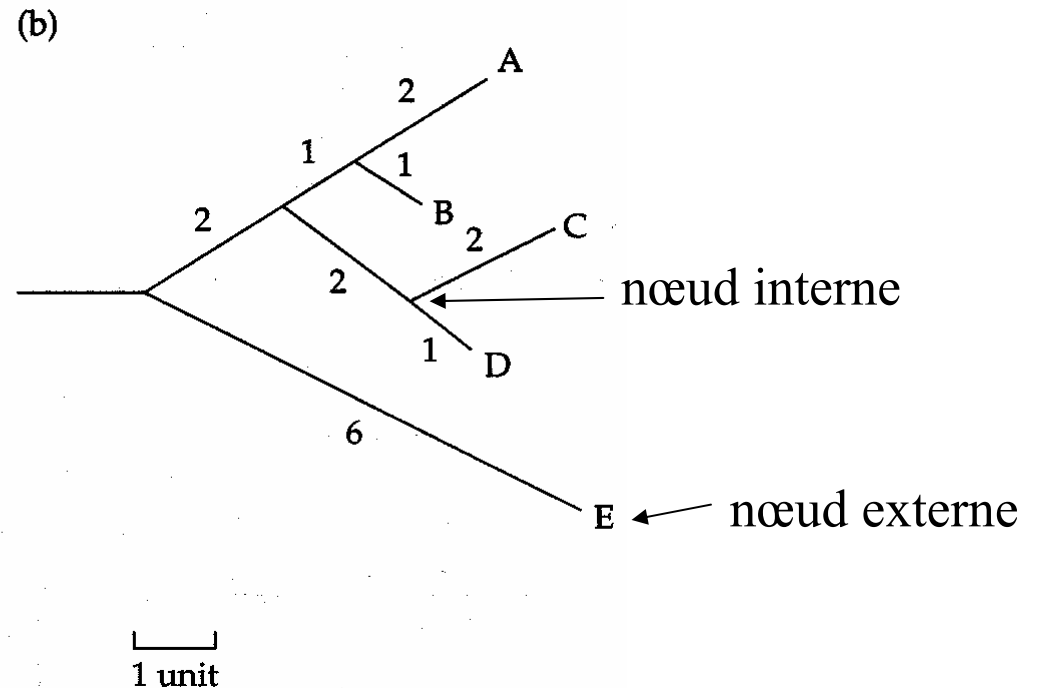
# 3.1. Introduction: l'arbre phylogénétique

## Types de représentation d'arbres phylogénétiques

branches non proportionnelles  
à **divergence moléculaire**:  
position des nœuds internes  
proportionnelle au **temps**



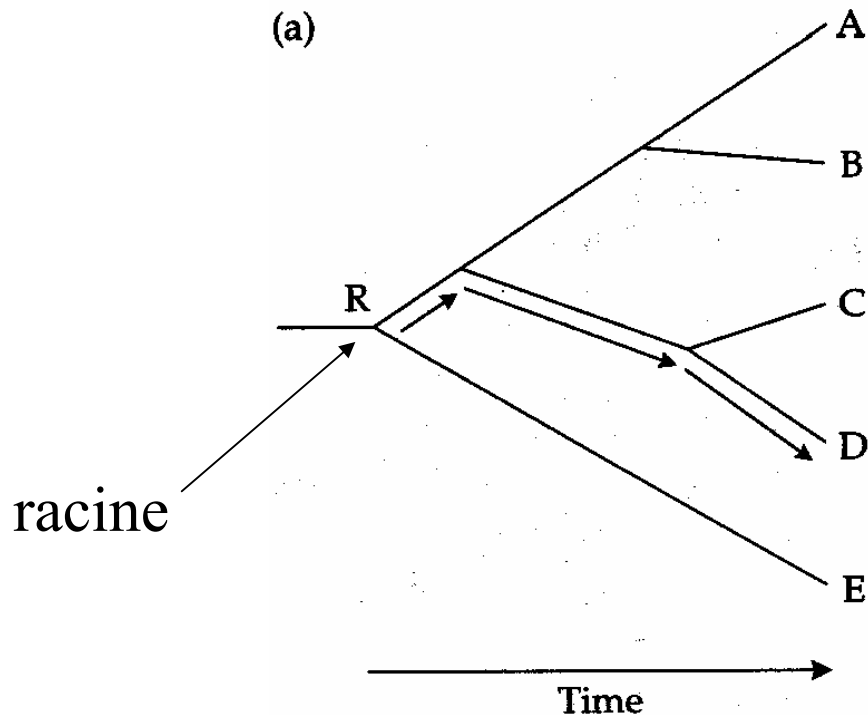
branches proportionnelles  
à **divergence moléculaire**:  
position des nœuds internes  
non proportionnelle au **temps**



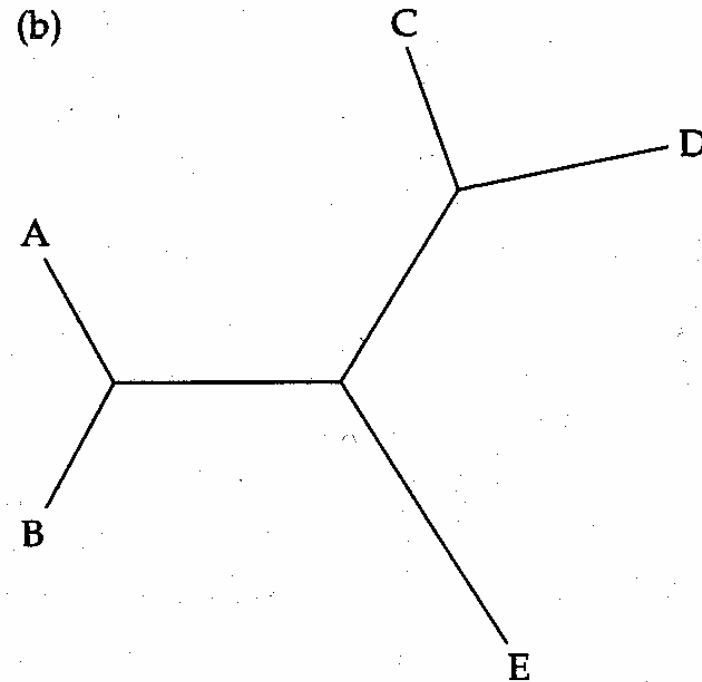
# 3.1. Introduction: l'arbre phylogénétique

## Types de représentation d'arbres phylogénétiques

arbre phylogénétique  
enraciné



arbre phylogénétique  
non enraciné



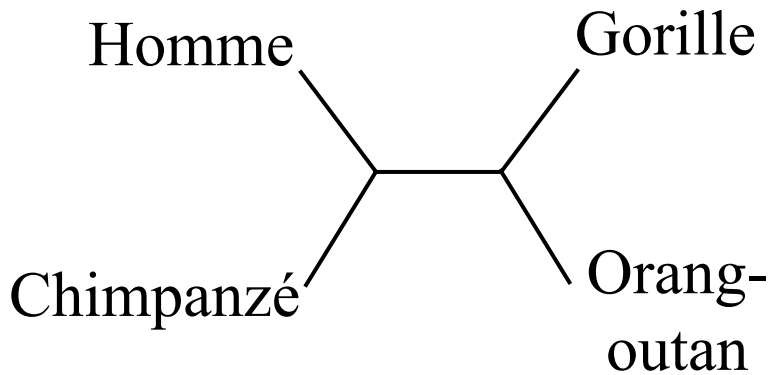
Arbre enraciné  $\Rightarrow$  direction du chemin évolutif est déterminée

# 3.1. Introduction: l'arbre phylogénétique

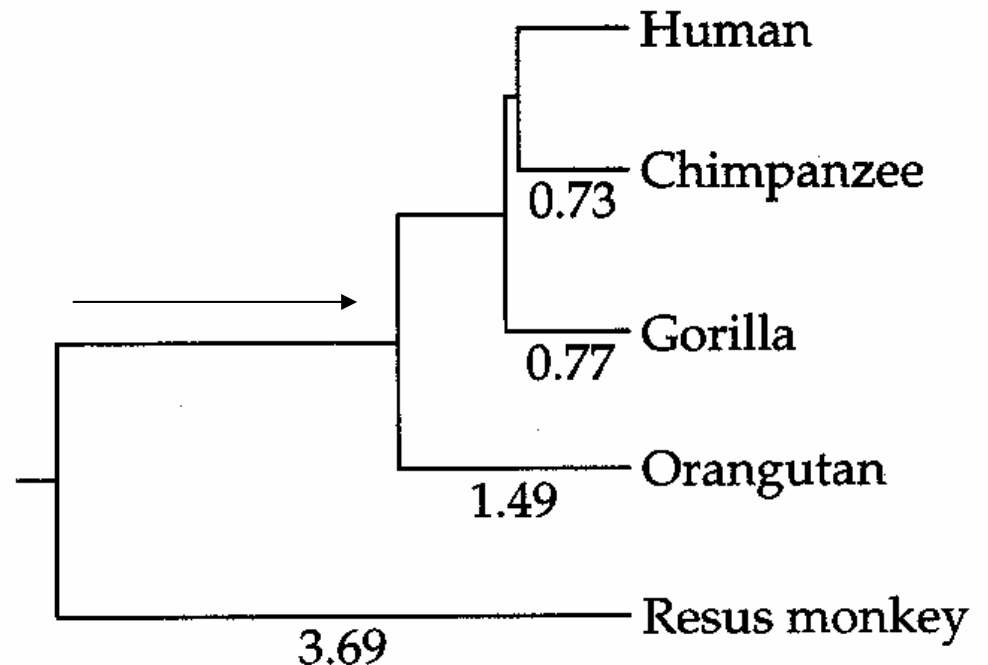
Méthode d'enracinement d'un arbre phylogénétique:

introduction d'un OTU ("Operational Taxonomic Unit") dont on sait qu'il est externe au groupe étudié = "**outgroup**" (**groupe externe**)

arbre sans outgroup  
direction évolutive  
indéterminée



arbre avec outgroup (singe rhesus n'appartenant pas aux primates)





## 3.1. Introduction: l'arbre phylogénétique

### Nombre de topologies alternatives

- arbre phylogénétique non enraciné avec  $n$  OTUs ( $n \geq 3$ )

$$N_U = \frac{(2n-5)!}{[2^{n-3} (n-3)!]}$$

- arbre phylogénétique enraciné avec  $n$  OTUs ( $n \geq 2$ )

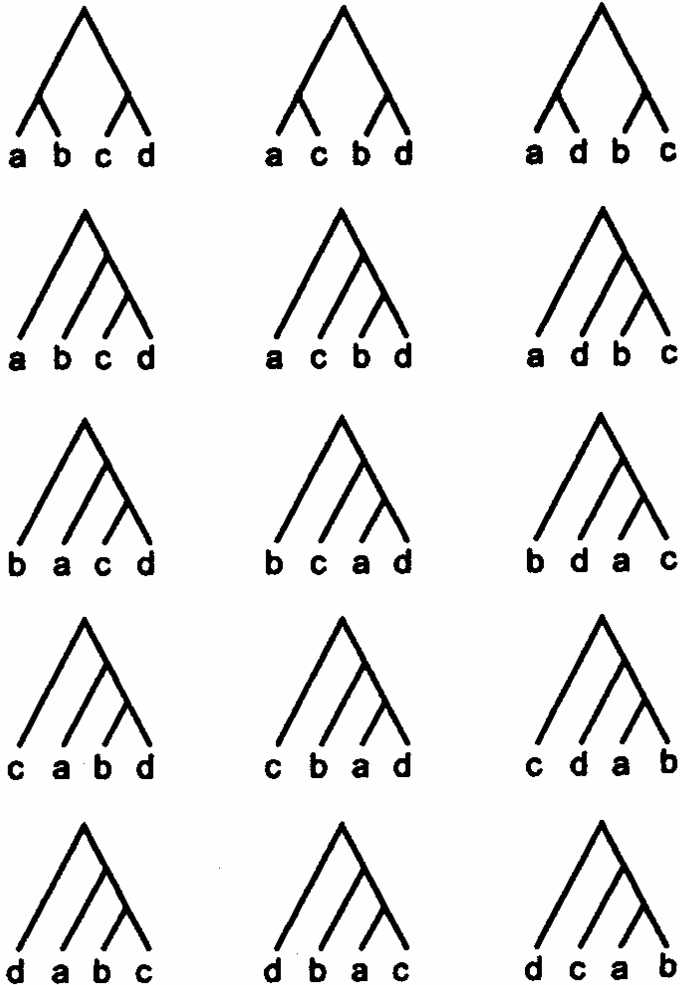
$$N_R = \frac{(2n-3)!}{[2^{n-2} (n-2)!]}$$

- si  $n = 3 \Rightarrow 3$  arbres non enraciné et **1** arbre enraciné

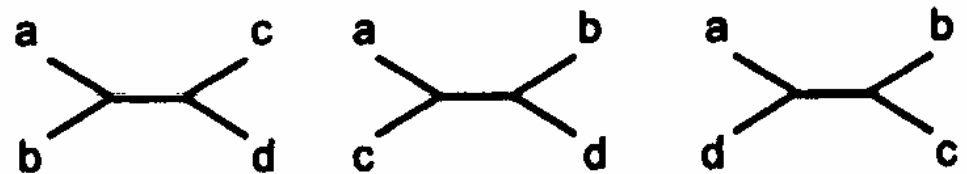
# 3.1. Introduction: l'arbre phylogénétique

Nombre de topologies alternatives:  $n = 4$

arbre phylogénétique enraciné



arbre phylogénétique non enraciné



## 3.1. Introduction: l'arbre phylogénétique

Nombre de topologies alternatives

**TABLE 5.1** Possible numbers of rooted and unrooted trees up to 10 OTUs

| <i>Number of OTUs</i> | <i>Number of rooted trees</i> | <i>Number of unrooted trees</i> |
|-----------------------|-------------------------------|---------------------------------|
| 2                     | 1                             | 1                               |
| 3                     | 3                             | 1                               |
| 4                     | 15                            | 3                               |
| 5                     | 105                           | 15                              |
| 6                     | 954                           | 105                             |
| 7                     | 10,395                        | 954                             |
| 8                     | 135,135                       | 10,395                          |
| 9                     | 2,027,025                     | 135,135                         |
| 10                    | 34,459,425                    | 2,027,025                       |

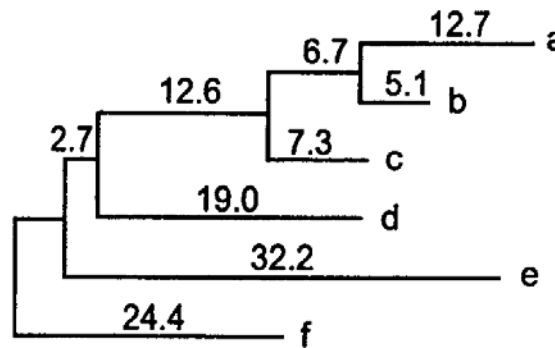
From Felsenstein (1978).

# 3.1. Introduction: l'arbre phylogénétique

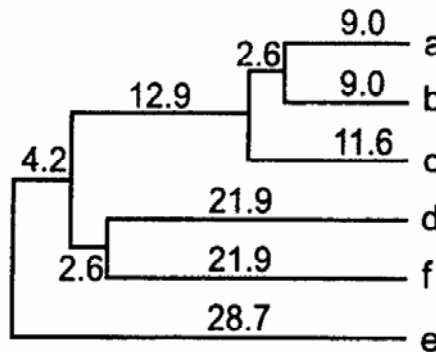
## Expression symbolique des topologies

- arbres phylogénétiques enracinés

(f(e(d(c(b,a))))))



(e((d,f)(c(b,a))))



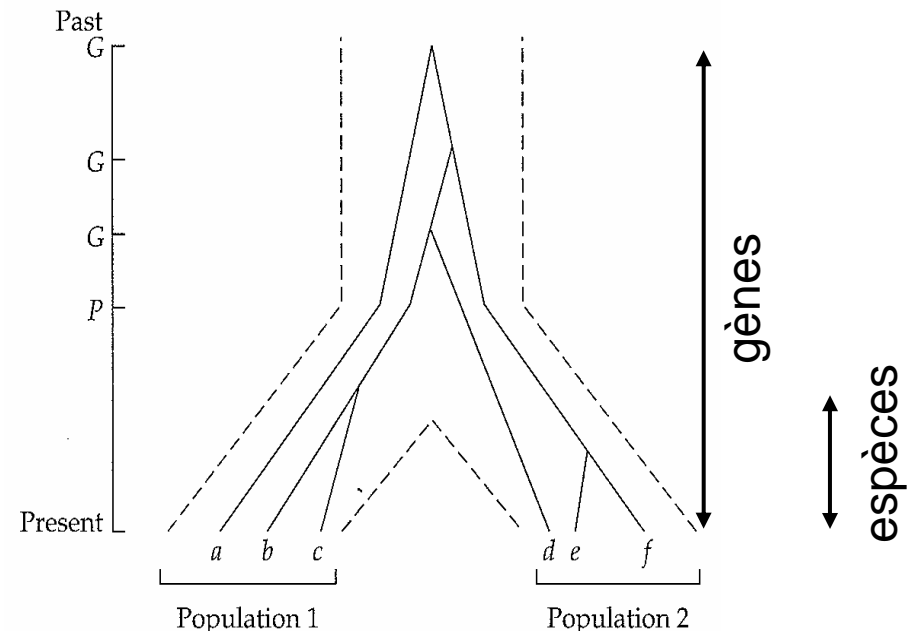
# 3.1. Introduction: l'arbre phylogénétique

## Arbre d'espèces et arbres de gènes

– **arbre d'espèces:** représente l'histoire évolutive d'un groupe d'espèces ou de populations. Le temps de divergence entre 2 espèces correspond au temps à partir duquel les 2 espèces sont devenues isolées génétiquement (barrières reproductives)

– **arbre de gènes:** l'histoire évolutive d'un échantillon de copies de gènes

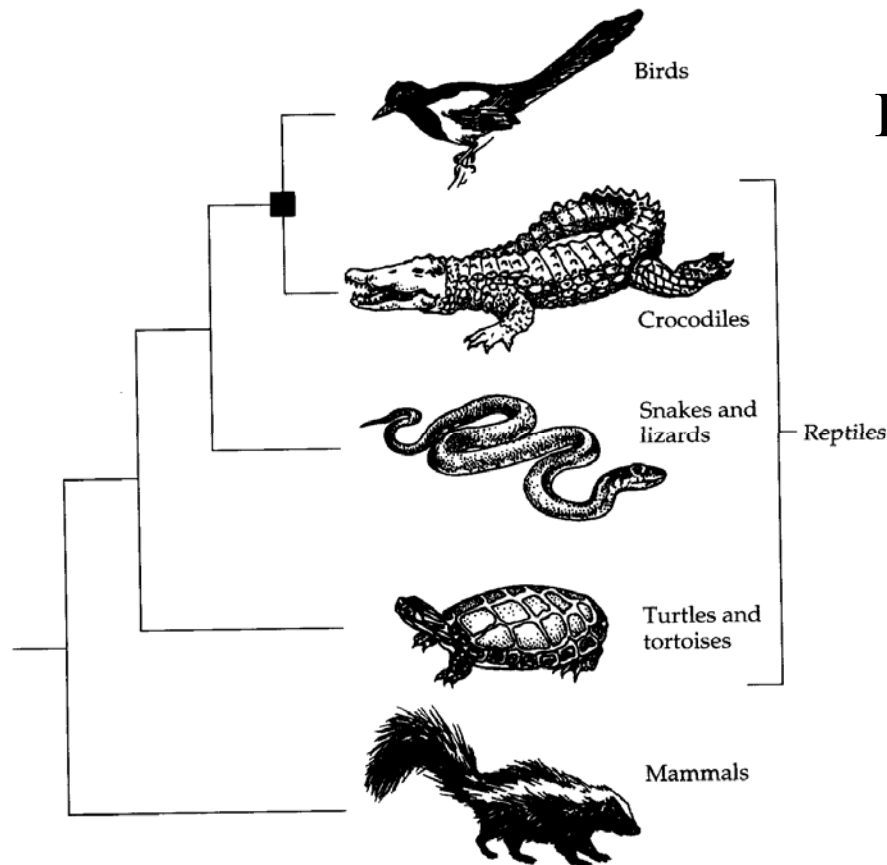
si  $\exists$  allèles polymorphes lors de la spéciation  $\Rightarrow$  temps de divergence des gènes  $>$  temps de divergence des espèces



# 3.1. Introduction: l'arbre phylogénétique

Groupes monophylétiques ou polyphylétiques:

Un groupe de taxons est **monophylétique** s'il existe un ancêtre commun exclusif de tous les taxons du groupe

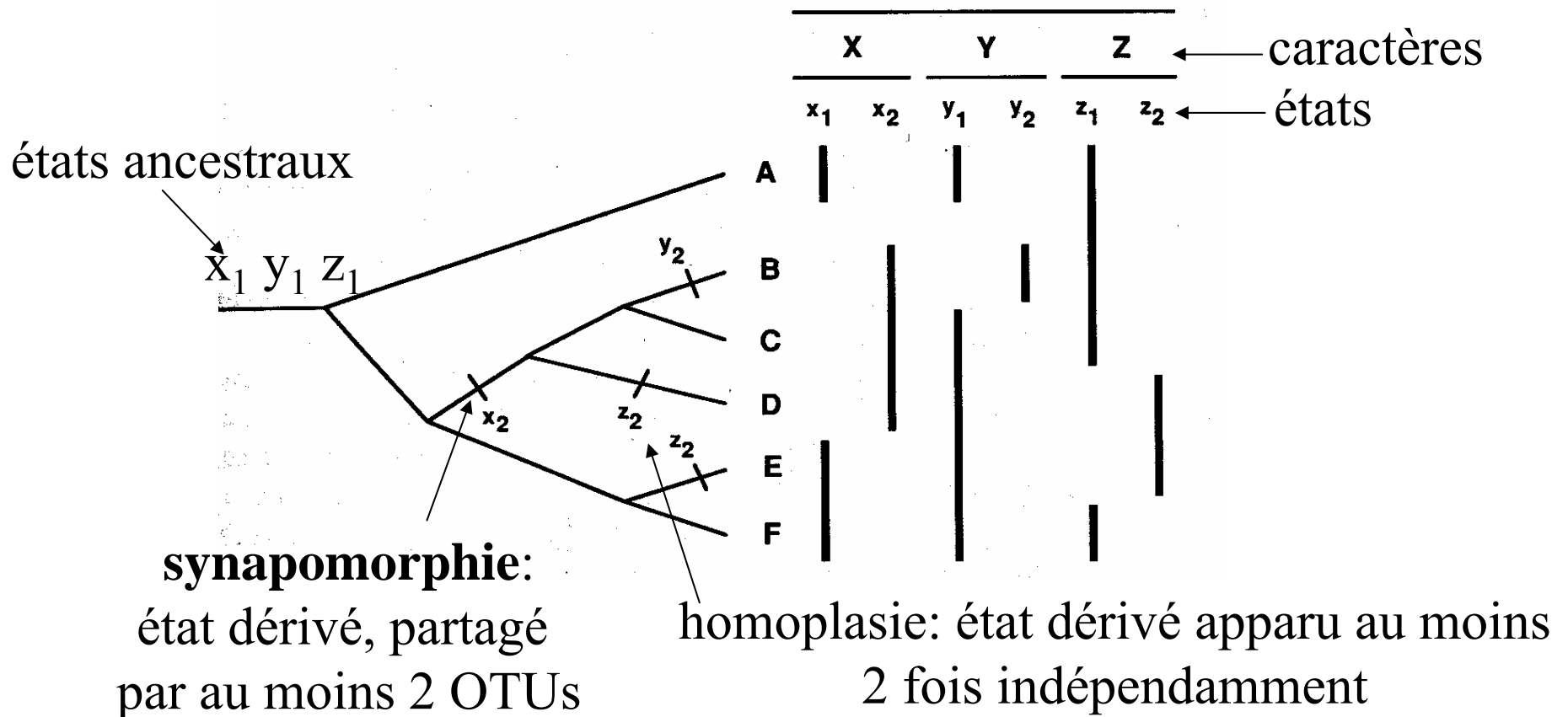


Les oiseaux et les crocodiles forment un groupe monophylétique, les Archosaures

La classe des reptiles n'est pas monophylétique

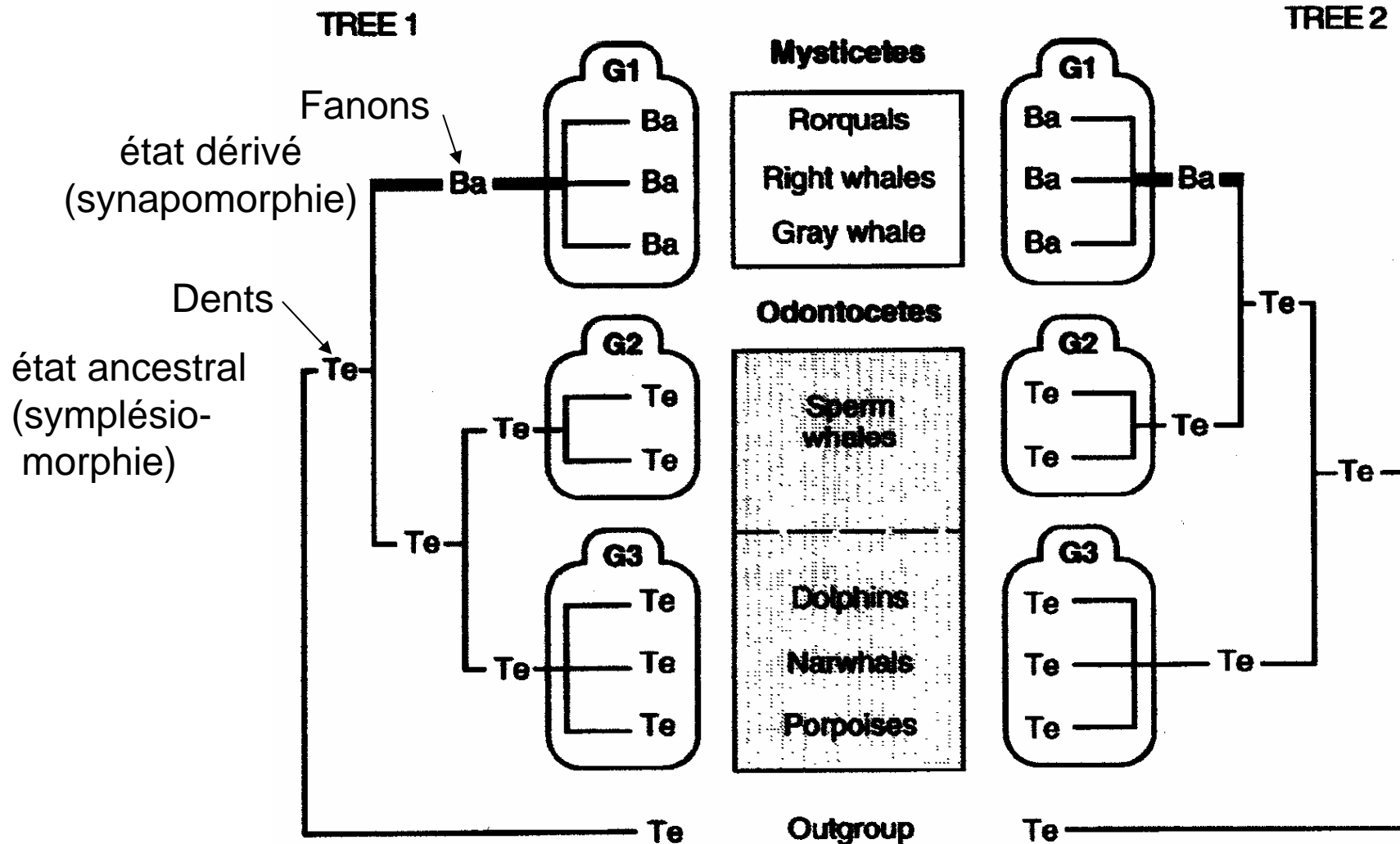
# 3.1. Introduction: l'arbre phylogénétique

## Etats dérivés et états ancestraux



# 3.1. Introduction: l'arbre phylogénétique

Etats dérivés et états ancestraux: ex. Phylogénie des Cétacés



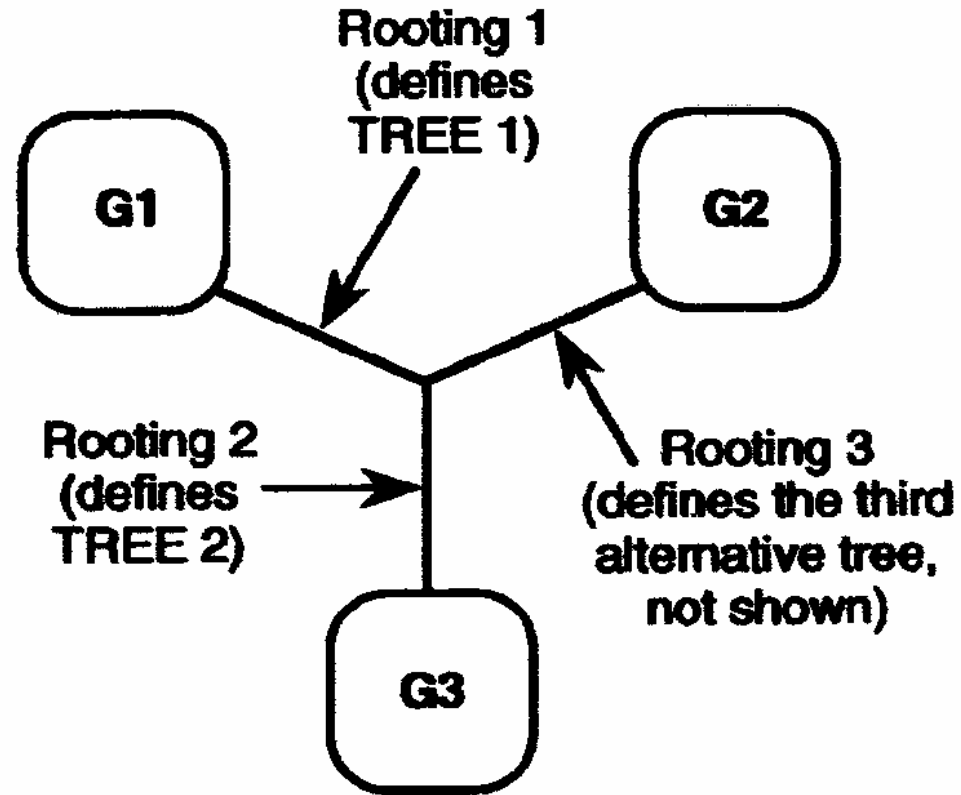


# 3.1. Introduction: l'arbre phylogénétique

Etats dérivés et états ancestraux: ex. Phylogénie des Cétacés

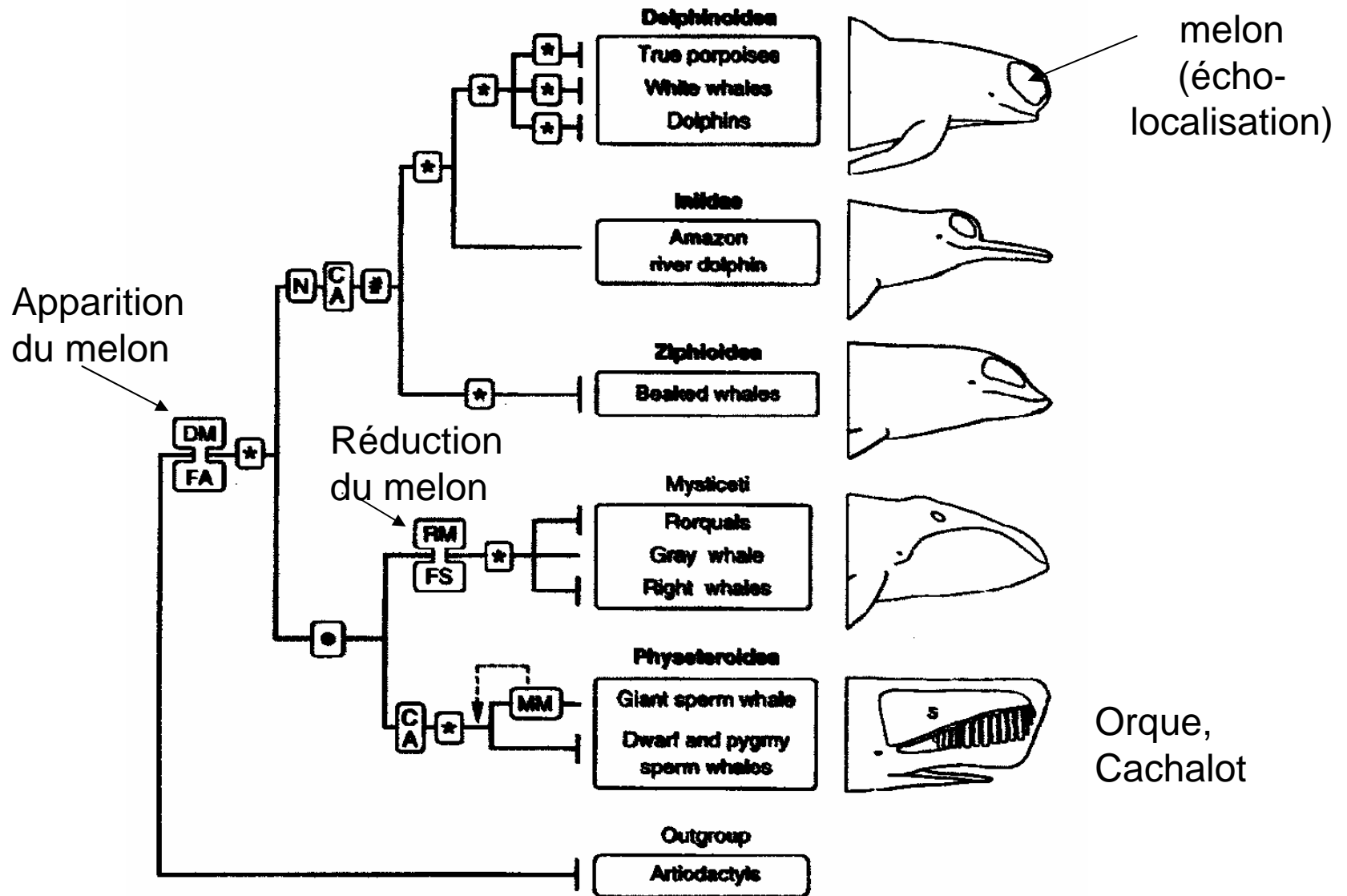
Baleines  
état dérivé  
(synapomorphie)

Dents  
état ancestral  
(symplesiomorphie)



# 3.1. Introduction: l'arbre phylogénétique

Etats dérivés et états ancestraux: ex. Phylogénie des Cétacés



# 3.1. Introduction: l'arbre phylogénétique

## Trois types de méthodes de reconstruction phylogénétique

- Méthodes des **distances génétiques**: calcul de distance génétique (nombre estimé de substitutions nucléotidiques) entre **chaque paire** d'OTU; construction progressive de l'arbre phylogénétique à partir de la matrice de distances
- Méthodes de **maximum de parcimonie**: }
- Méthodes de **maximum de vraisemblance**: }
  - Exploration d'un très grand nombre de topologies alternatives
  - Pour chaque topologie examinée  $\Rightarrow$  estimation de son score par confrontation avec les données moléculaires
    - parcimonie: nombre d'étapes évolutives (substitutions) nécessaires
    - vraisemblance: probabilité des données si topologie est la bonne
  - Choix de la (ou des) topologie(s) avec le meilleur score

## 3.2. Les méthodes des distances génétiques

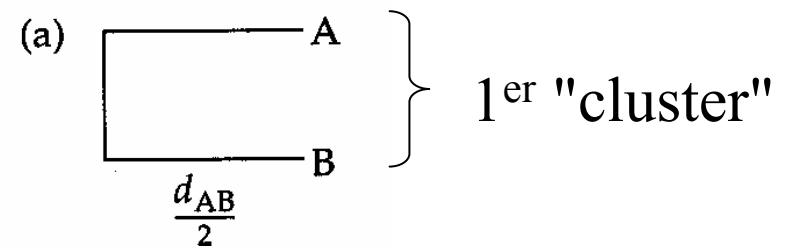
A. Méthode UPGMA ("unweighted pair-group method with arithmetic mean"): algorithme simple produisant des arbres à branches non proportionnelles à la divergence moléculaire  $\Rightarrow$  valable si les taux d'évolution varient peu entre les différentes lignées évolutives

Algorithme séquentiel: à chaque étape  $\rightarrow$  regroupement des 2 OTUs présentant la distance génétique la plus faible  $\Rightarrow$  1 OTU composite; réactualisation de matrice de distances génétiques entre OTUs (y compris OTUs composites)

min.  $\swarrow$

|     |          | <u>Matrice distances</u> |          |          |
|-----|----------|--------------------------|----------|----------|
|     |          | OTU                      |          |          |
| OTU |          | A                        | B        | C        |
| B   | $d_{AB}$ |                          |          |          |
| C   | $d_{AC}$ |                          | $d_{BC}$ |          |
| D   | $d_{AD}$ |                          | $d_{BD}$ | $d_{CD}$ |

Arbre après étape 1



## 3.2. Les méthodes des distances génétiques

### A. Méthode UPGMA:

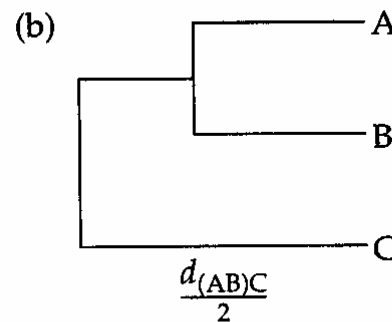
Recalcul de la matrice de distances avec l'OTU composite (AB) avec:

$$d_{(AB)C} = (d_{AC} + d_{BC})/2 \text{ et } d_{(AB)D} = (d_{AD} + d_{BD})/2$$

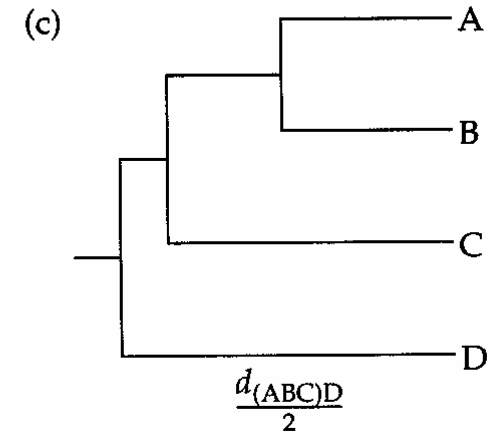
#### Nouvelle matrice distances

|     | OTU         |          |
|-----|-------------|----------|
| OTU | (AB)        | C        |
| C   | $d_{(AB)C}$ |          |
| D   | $d_{(AB)D}$ | $d_{CD}$ |

#### Arbre après étape 2



#### Arbre après étape 3



## 3.2. Les méthodes des distances génétiques

B. Méthode de "**neighbor-joining**": Saitou & Nei, 1987

Basée sur le principe d'**évolution minimale**

Si  $S$  = somme de toutes les branches de l'arbre

avec  $T$  = nombre de branches =  $2m-3$

et  $b_i$  = longueur de branche  $i$

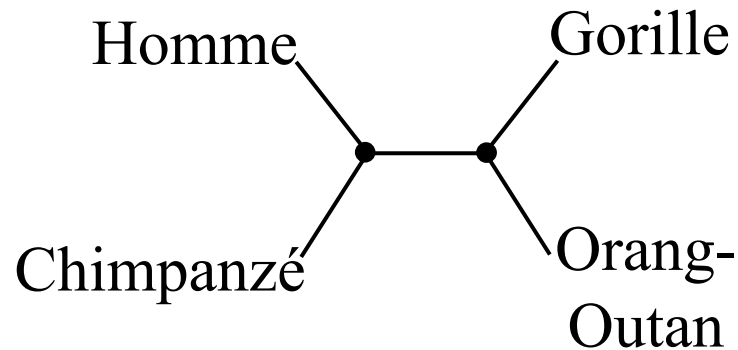
⇒ Meilleure topologie →  $S$  minimum

$$S = \sum_i^T b_i$$

## 3.2. Les méthodes des distances génétiques

### B. Méthode de "neighbor-joining":

2 OTUs sont **voisins** ("neighbors") si connectés par un seul nœud interne

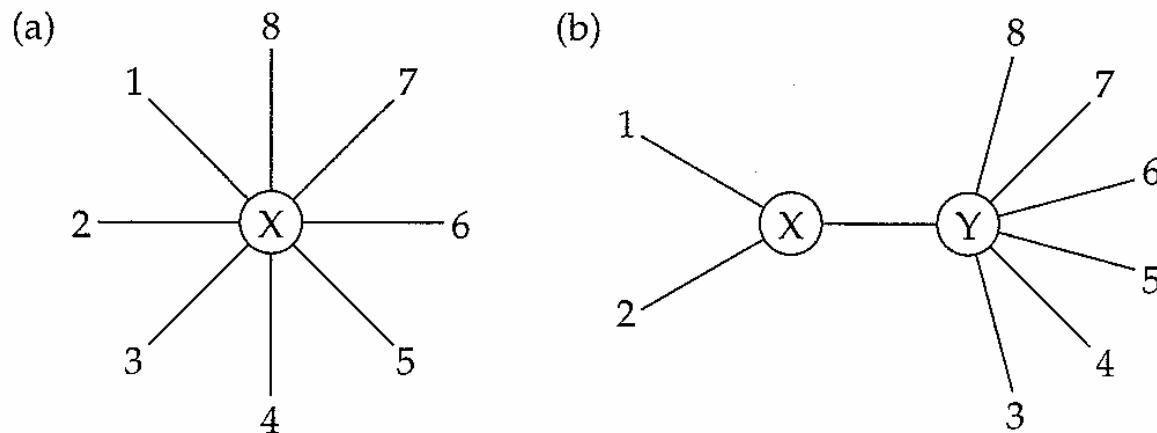


Si H et C sont voisins, et si G et O sont voisins  $\Rightarrow$   
 $d_{HC} + d_{GO} < d_{HG} + d_{CO}$  et  $d_{HC} + d_{GO} < d_{HO} + d_{CG}$

## 3.2. Les méthodes des distances génétiques

### B. Méthode de "neighbor-joining":

Algorithme séquentiel: démarrage avec une topologie étoile ("star topology");  
à chaque étape → test de toutes les paires d'OTUs choisies comme OTUs voisins  
et estimation de la longueur totale de l'arbre (somme de toutes les longueurs de  
branches) → choix de la paire produisant une longueur totale minimale  $\Rightarrow$  1  
OTU composite; réactualisation de matrice de distances génétiques entre OTUs  
(y compris OTUs composites)





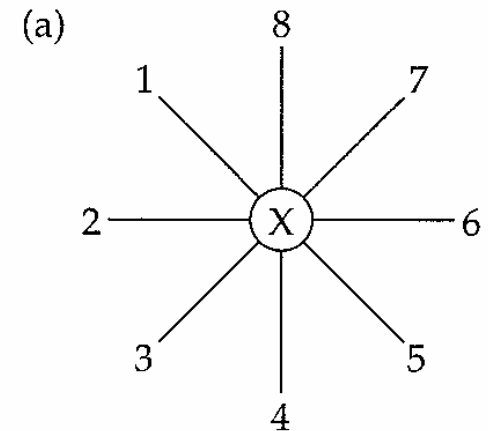
## 3.2. Les méthodes des distances génétiques

B. Méthode de "neighbor-joining":

Topologie étoile: longueur totale  $S_0$

$$S_0 = \sum_{i=1}^m L_{iX} = \frac{1}{m-1} \sum_{i<j}^m d_{ij}$$

avec  $L_{iX}$  = longueur branche reliant nœuds  $i$  et  $X$   
et  $d_{ij}$  = distance génétique entre OTUs  $i$  et  $j$

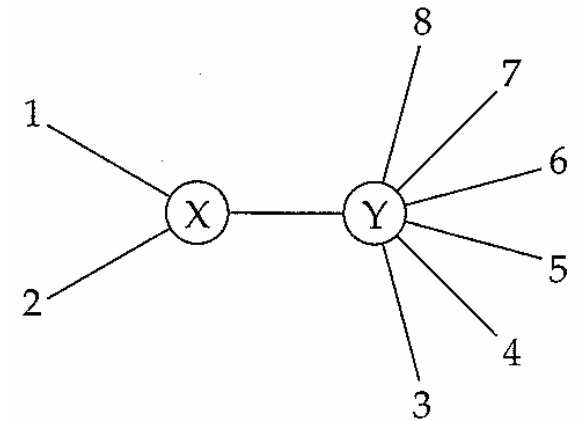


## 3.2. Les méthodes des distances génétiques

B. Méthode de "neighbor-joining":

Topologie voisins 1-2: longueur totale  $S_{12}$

$$S_{12} = L_{1X} + L_{2X} + L_{XY} + \sum_{i=3}^m L_{iY}$$



avec

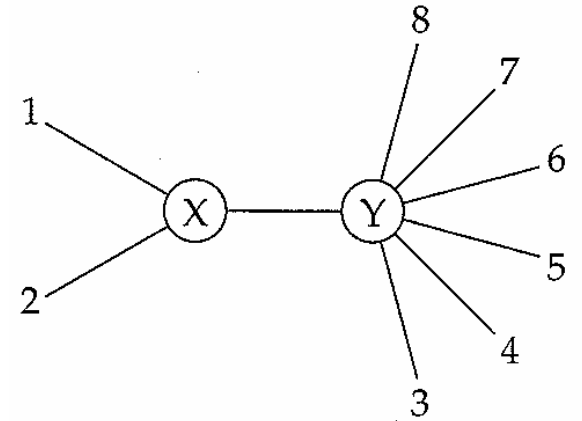
$$L_{XY} = \frac{1}{2(m-2)} \left[ \sum_{i=3}^m (d_{1i} + d_{2i}) - (m-2)(L_{1X} + L_{2X}) - 2 \sum_{i=3}^m L_{iY} \right]$$

$$\sum_{i=3}^m L_{iY} = \frac{1}{m-3} \sum_{3 \leq i < j} d_{ij}$$

## 3.2. Les méthodes des distances génétiques

B. Méthode de "neighbor-joining":

Topologie voisins 1-2: longueur totale  $S_{12}$



ensuite on remplace nœuds 1 et 2 par nœud X et on recalcule la matrice de distances génétiques:

$$d_{Xk} = \frac{(d_{ik} + d_{jk} - d_{ij})}{2}$$

## 3.3. Les méthodes de maximum de parcimonie

### Méthodes de **maximum de parcimonie** :

**Principe:** recherche de l'arbre phylogénétique dont la topologie demande le **nombre le plus faible de changements évolutifs** (substitutions nucléotidiques) nécessaires pour expliquer les **différences observées entre OTUs**

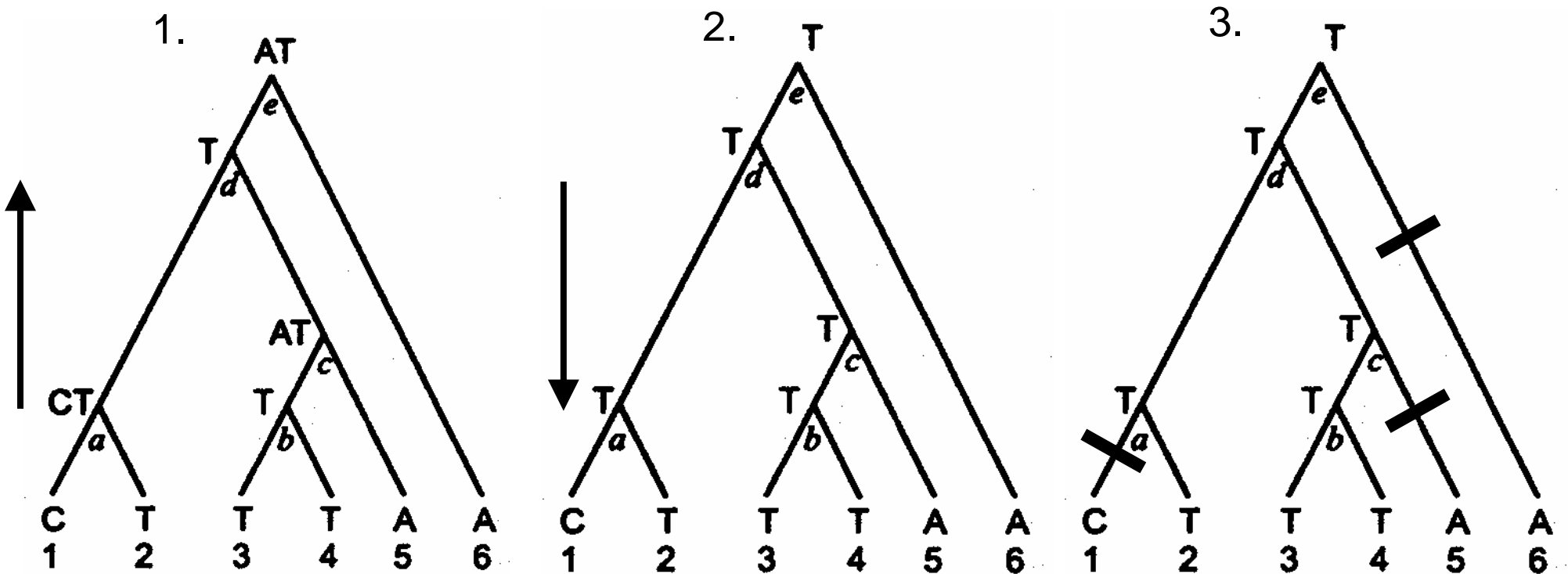
#### **Procédure:**

- algorithme passant en revue un très grand nombre de topologies différentes
- pour chaque topologie et chaque site, calcul du nombre de changements évolutifs nécessaires pour expliquer les données de l'échantillon
- somme sur tous les sites nucléotidiques  $\Rightarrow$  score de la topologie
- choix de la (ou des) topologie(s) avec le score le plus faible

## 3.3. Les méthodes de maximum de parcimonie

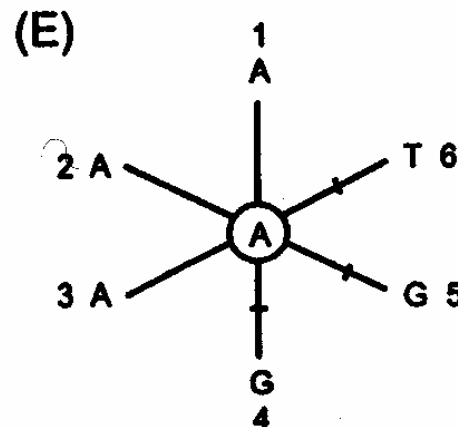
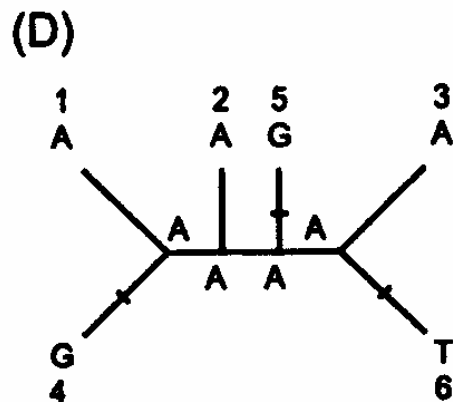
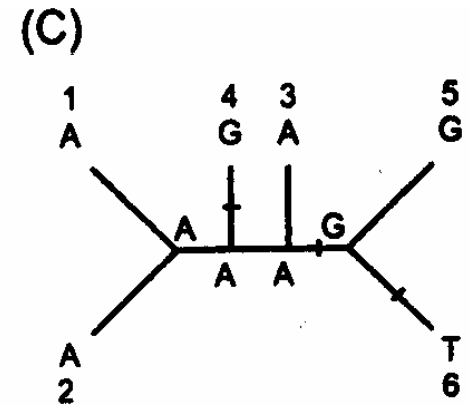
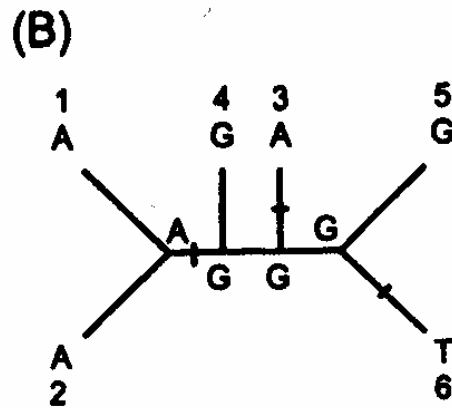
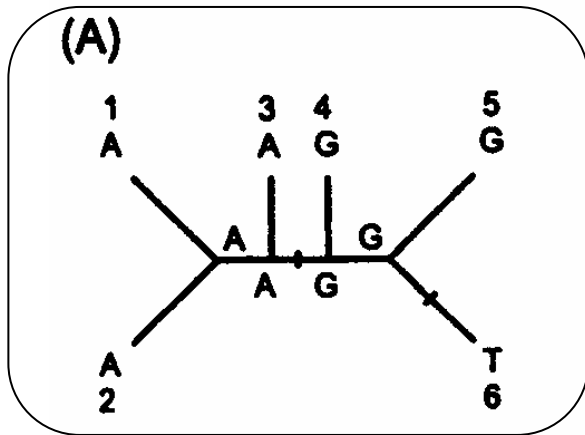
### 3.3.1. Calcul du nombre minimum de substitutions : algorithme de Fitch (1971)

1. reconstitution exhaustive des états ancestraux possibles
2. choix des états ancestraux les plus parcimonieux
3. calcul du nombre de substitutions



# 3.3. Les méthodes de maximum de parcimonie

3.3.1. Calcul du nombre minimum de substitutions :  
→ Comparaison de différentes topologies



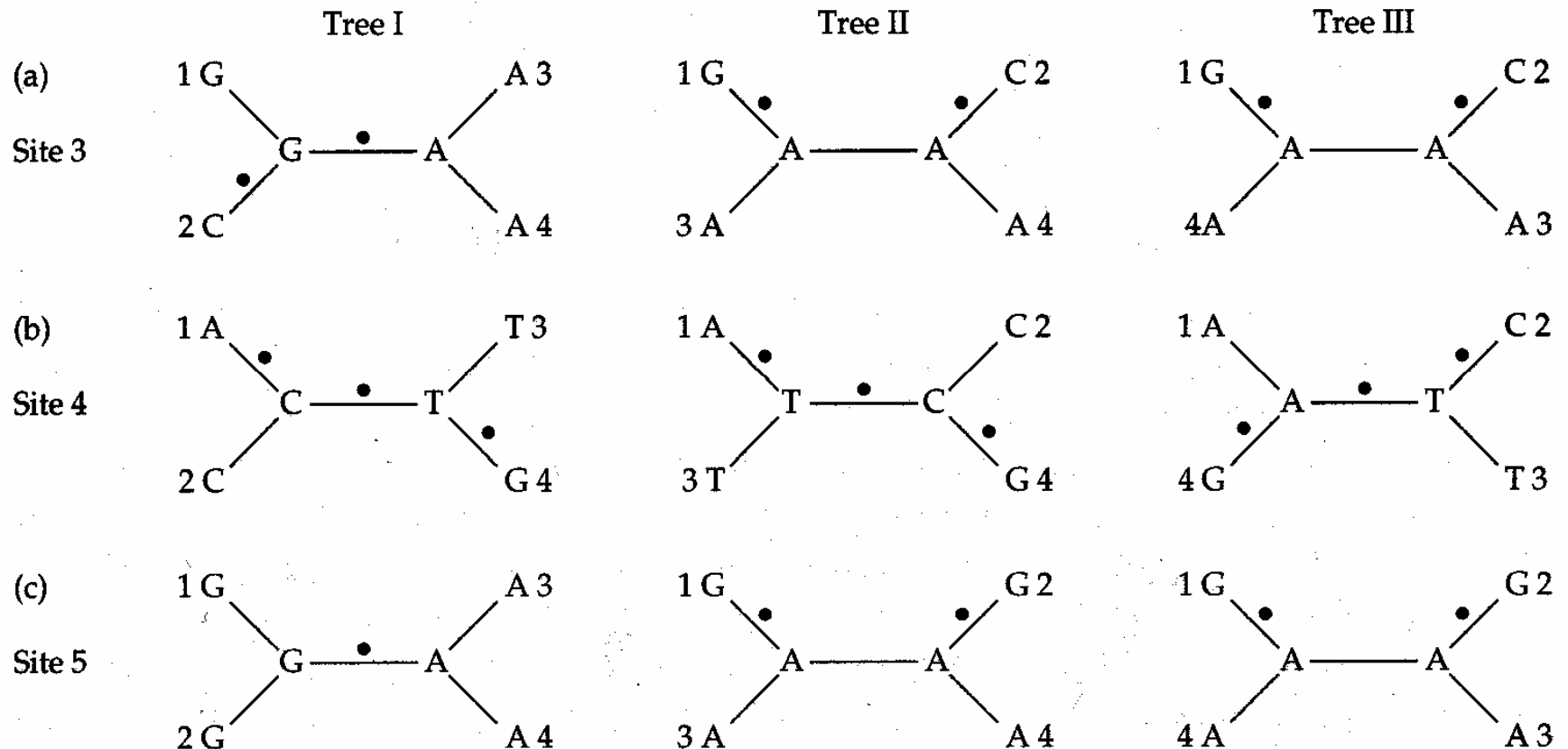
6 OTUs  $\Rightarrow$  105  
arbres non enracinés

Calcul de  $L$  = long. de l'arbre  
=  $\Sigma$  nombres min. substitutions  
sur tous sites

# 3.3. Les méthodes de maximum de parcimonie

## 3.3.1. Calcul du nombre minimum de substitutions :

Sites phylogénétiquement **informatifs** et sites non-informatifs



## 1.2. Méthodes de reconstruction phylogénétique

### 3.3.1. Calcul du nombre minimum de substitutions :

Sites phylogénétiquement **informatifs** et sites non-informatifs:

Seuls les sites informatifs permettent de discriminer entre topologies différentes: un site est phylogénétiquement informatif s'il y a au moins 2 nucléotides différents en ce site, chacun étant représenté dans au moins 2 des OTUs analysés

|          | Site |   |   |   |   |   |   |   |   |
|----------|------|---|---|---|---|---|---|---|---|
| Sequence | 1    | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1        | A    | A | G | A | G | T | G | C | A |
| 2        | A    | G | C | C | G | T | G | C | G |
| 3        | A    | G | A | T | A | T | C | C | A |
| 4        | A    | G | A | G | A | T | C | C | G |

\*                      \*                      \*

↙                      ↗                      ↘

sites informatifs



## 3.3. Les méthodes de maximum de parcimonie

### 3.3.2. Recherche de topologies alternatives :

- Recherches exhaustives :  $m < 10$
- Recherches non-exhaustives:
  - Méthodes "**branch-and-bound**" (Hendy & Penny 1982):  $m < 20$   
Élimination des familles de topologies dont  $L >$  valeur de référence
    - ⇒ recherche seulement dans l'espace pertinent de topologies
    - ⇒ certain de trouver le/les arbre(s) le(s) plus parcimonieux
  - Méthodes **heuristiques** (exploration par évaluations successives):  
 $m > 20$   
Examen d'une petite portion seulement des topologies avec  
algorithme "intelligent"
    - ⇒ pas de garantie de trouver l'arbre le plus parcimonieux

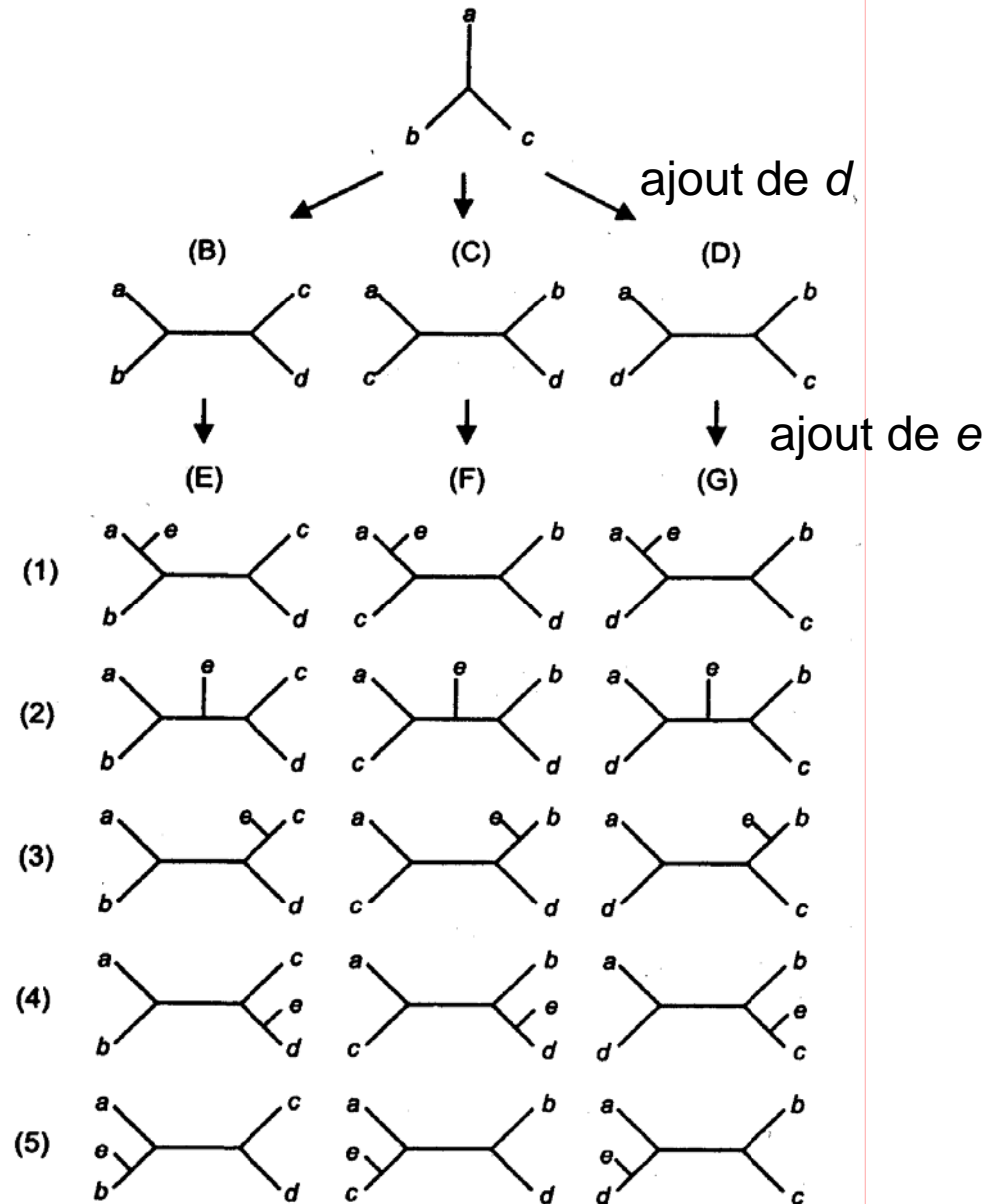
## 3.3. Les méthodes de maximum de parcimonie

### 3.3.2. Recherche de topologies alternatives :

- Méthodes "**branch-and-bound**" (Hendy & Penny 1982):  $m < 20$ 
  - on démarre avec un arbre à 3 OTUs
  - on ajoute un OTU supplémentaire dans chaque position possible et on calcule la longueur de l'arbre (incomplet) correspondant
  - le ou les arbres dont  $L > L_{\text{seuil}}$  sont éliminés ainsi que leurs "descendants"
  - on recommence l'ajout d'un OTU supplémentaire à chaque arbre non éliminé ...
  - le choix des 3 OTUs de départ et de l'ordre d'introduction des OTUs supplémentaires est déterminé par un algorithme précis

### 3.3. Les méthodes de maximum de parcimonie

Méthodes "branch-and-bound"



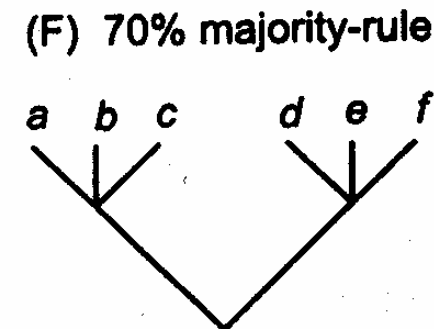
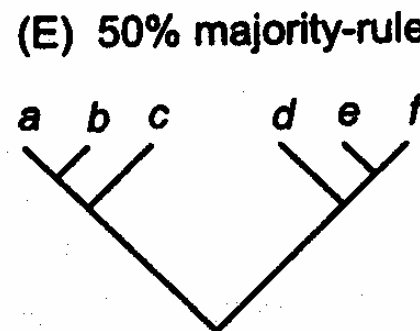
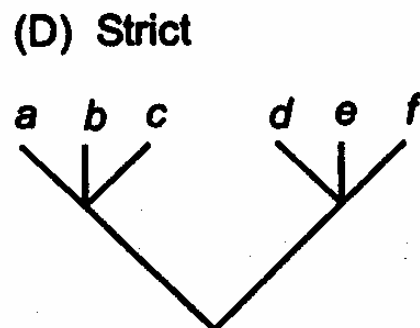
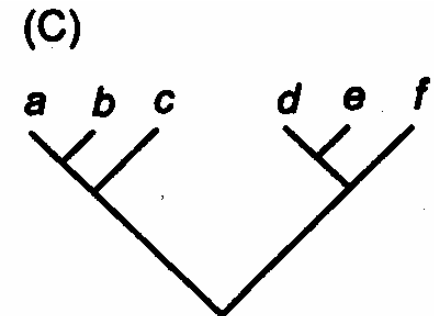
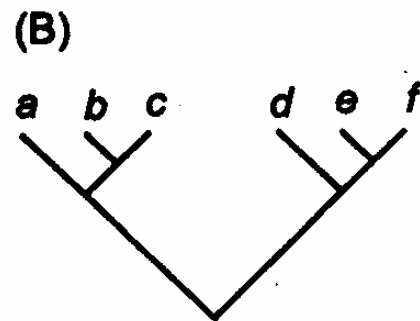
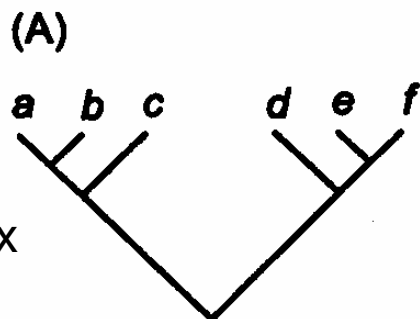
(tiré de Nei & Kumar, 2000)

## 3.3. Les méthodes de maximum de parcimonie

### 3.3.3. Arbres consensus :

- Si on obtient plusieurs arbres également parcimonieux (même  $L$ )
  - ⇒ établissement d'un **arbre consensus** (arbre composite)
- arbre à consensus strict
- arbre à consensus majoritaire

3 arbres également parcimonieux



## 3.3. Les méthodes de maximum de parcimonie

### 3.3.4. Parcimonie pondérée:

- Différentes substitutions  $\Rightarrow \neq$  probabilités d'occurrence  $\Rightarrow \neq$  poids dans le calcul de  $L$ 
  - ex.: poids des transitions  $>$  poids des transversions

#### (A) Weighted parsimony

|   | A | T | C | G |
|---|---|---|---|---|
| A |   | w | w | 1 |
| T | w |   | 1 | w |
| C | w | 1 |   | w |
| G | 1 | w | w |   |

## 3.3. Les méthodes de maximum de parcimonie

### 3.3.5. Estimation des longueurs de branches:

Parcimonie  $\Rightarrow$  choix de topologie mais longueurs de branches arbitraires (n'intervient pas dans calcul du score  $L$ )

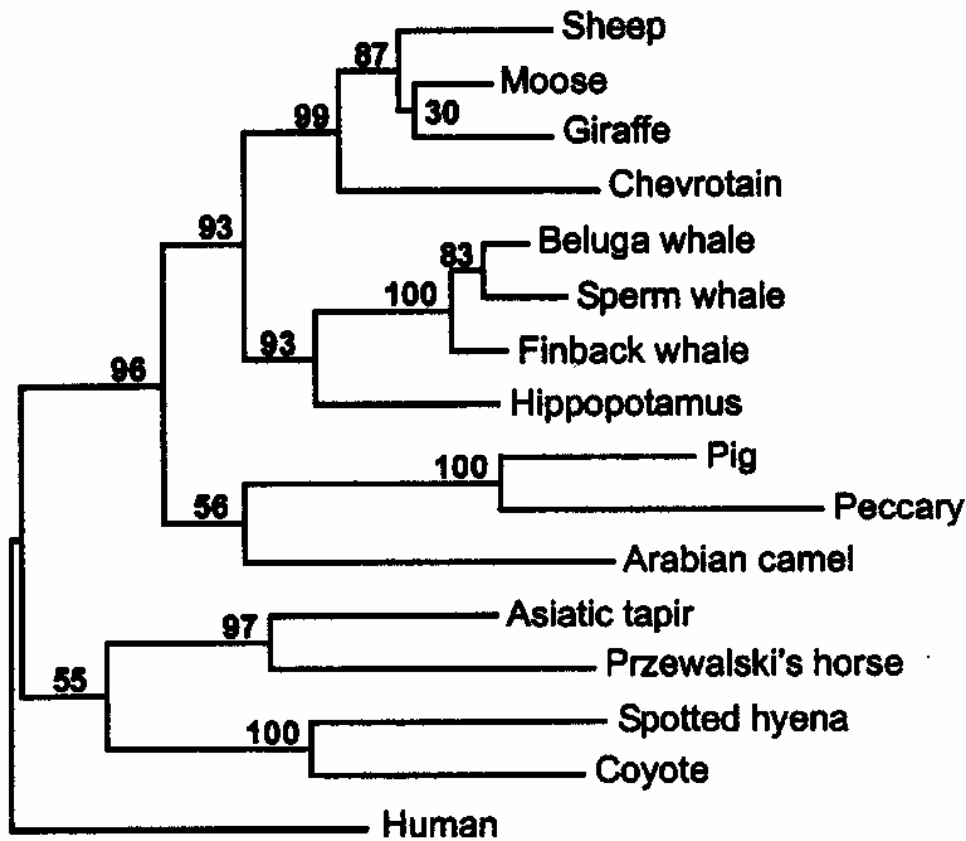
$\rightarrow$  après choix de topologie  $\Rightarrow$  calcul des longueurs des branches ( $\neq$  méthodes)

# 3.3. Les méthodes de maximum de parcimonie

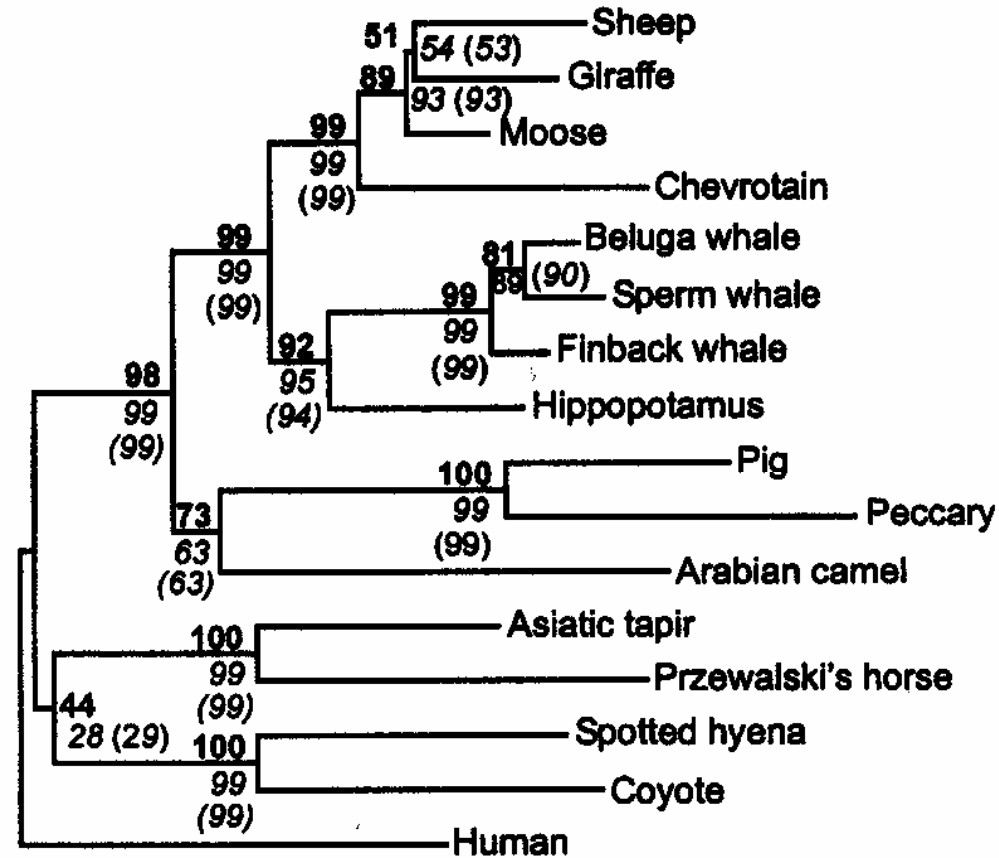
## 3.3.6. Exemple: origine des cétacés

Comparaison de reconstructions par parcimonie et neighbor-joining

(A) MP bootstrap consensus tree



(B) NJ tree (Kimura model)



0.05 substitutions/site

0.05 substitutions/site

## 3.3. Les méthodes de maximum de parcimonie

### 3.3.6. Exemple: origine des cétacés

- Artiodactyles (ordre Mammifères)
  - doigts en nombre pair
  - Suiformes
    - porcs, pécaris
    - Hippopotamidés
  - Tylopodes
    - Camélidés
  - Ruminants
    - Cervidés
    - Giraffidés
    - Antilocapridés
    - Bovidés
- Périssodactyles (Chevaux,...)
  - doigts en nombre impair

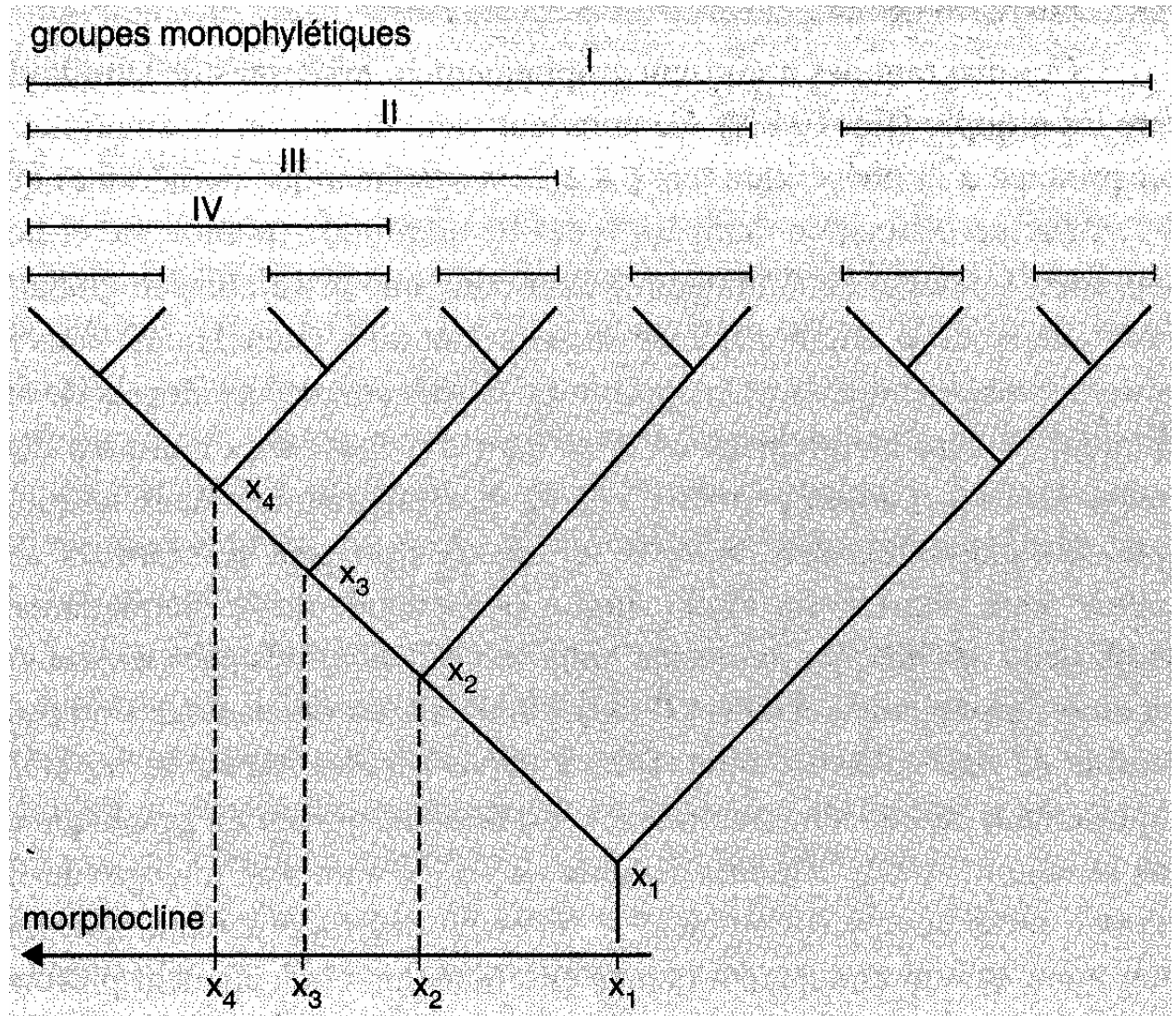
Artiodactyles ne forment pas un groupe monophylétique!  
⇒ nouvelle nomenclature:  
Artiodactyles + Cétacés  
= Cétartiodactyles



# 3.3. Les méthodes de maximum de parcimonie

## 3.3.6. Exemple: origine des cétacés

Reconnaissance  
de groupes  
monophylétiques



## 3.3. Les méthodes de maximum de vraisemblance

**Principe:** recherche de l'arbre phylogénétique dont la topologie est la plus vraisemblable étant donné les séquences nucléotidiques des OTUs (les données) et le **modèle d'évolution** des séquences nucléotidiques choisi

### **Procédure:**

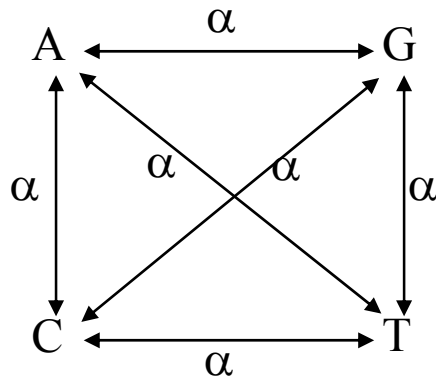
- choix d'un modèle d'évolution des séquences nucléotidiques (ex. Jukes & Cantor à 1 paramètre)
- algorithme passant en revue un très grand nombre de topologies différentes
- pour chaque topologie, estimation des longueurs de branches optimales et calcul de la vraisemblance (probabilité conditionnelle = "likelihood") étant donné les séquences nucléotidiques des OTUs (les données) et le modèle d'évolution des séquences choisi
- choix de la topologie avec la vraisemblance la plus forte

## 3.3. Les méthodes de maximum de vraisemblance

### 3.3.1. Calcul de la vraisemblance

- modèle d'évolution des séquences nucl.(Jukes & Cantor)
- topologie à tester
- séquences nucléotidiques

#### Modèle de Jukes & Cantor



même nucléotide:

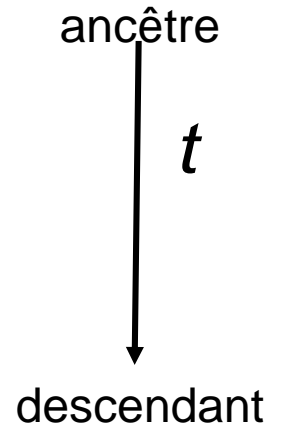
$$P_{ii(t)} = \frac{1}{4} + \frac{3}{4} e^{-4\alpha t}$$

nucléotide différent:

$$P_{ij(t)} = \frac{1}{4} - \frac{1}{4} e^{-4\alpha t}$$

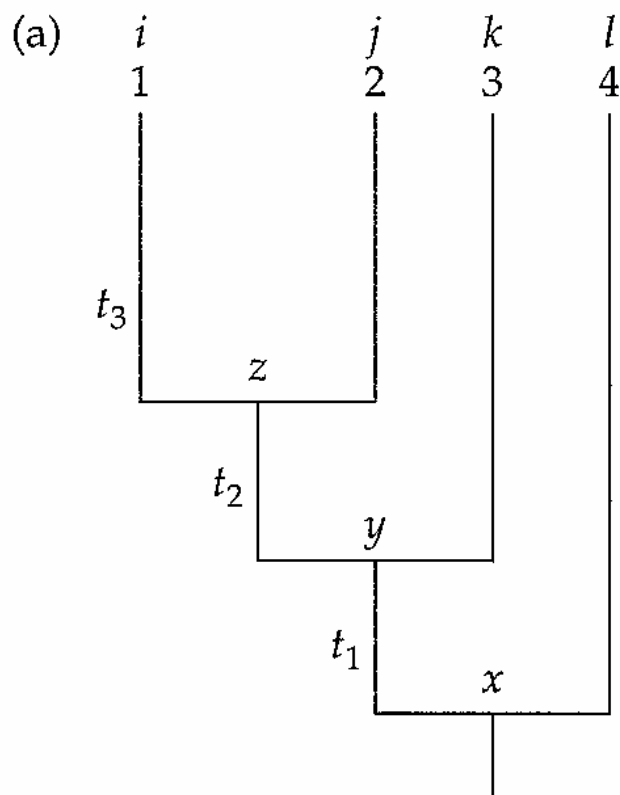
ancêtre

descendant



## 3.3. Les méthodes de maximum de vraisemblance

- 3.3.1. Calcul de la vraisemblance
  - modèle d'évolution des séquences nucl.(Jukes & Cantor)
  - topologie à tester
  - séquences nucléotidiques (site  $s$ : nucl.  $i,j,k,l$  aux OTUs 1,2,3,4)



-Si les **états ancestraux** sont connus ( $x,y,z$ )  
-Si les **longueurs de branches** sont connues ( $t_1,t_2,t_3$ )  $\Rightarrow$  calcul de vraisemblance au site  $s$ :

$$L_s = \text{Prob}(i,j,k,l) = P_{xl}(t_1+t_2+t_3) \cdot P_{xy}(t_1) \cdot P_{yk}(t_2+t_3) \cdot P_{yz}(t_2) \cdot P_{zi}(t_3) \cdot P_{zj}(t_3)$$

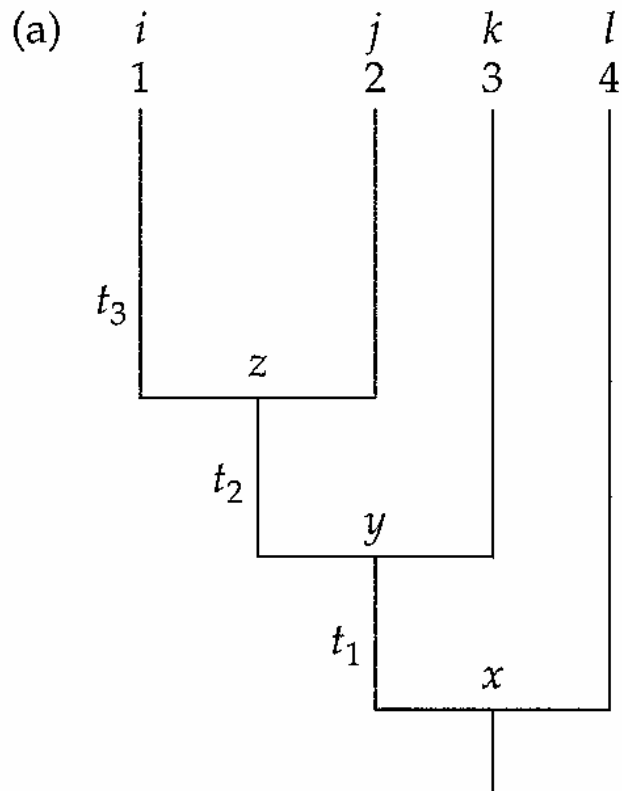
Vraisemblance totale  $L =$  produit des  $L_s$

$$\ln L = \sum_{s=1}^n \ln L_s \quad \rightarrow \text{"log-likelihood"}$$

## 3.3. Les méthodes de maximum de vraisemblance

- 3.3.1. Calcul de la vraisemblance
  - en pratique, **états ancestraux sont non connus**  $\Rightarrow$  vraisemblance au site  $s =$  somme des vraisemblances pour toutes les valeurs possibles des états ancestraux  $(x,y,z)$

$$L_s = \sum_x \sum_y \sum_z \text{Prob}(i, j, k, l)$$

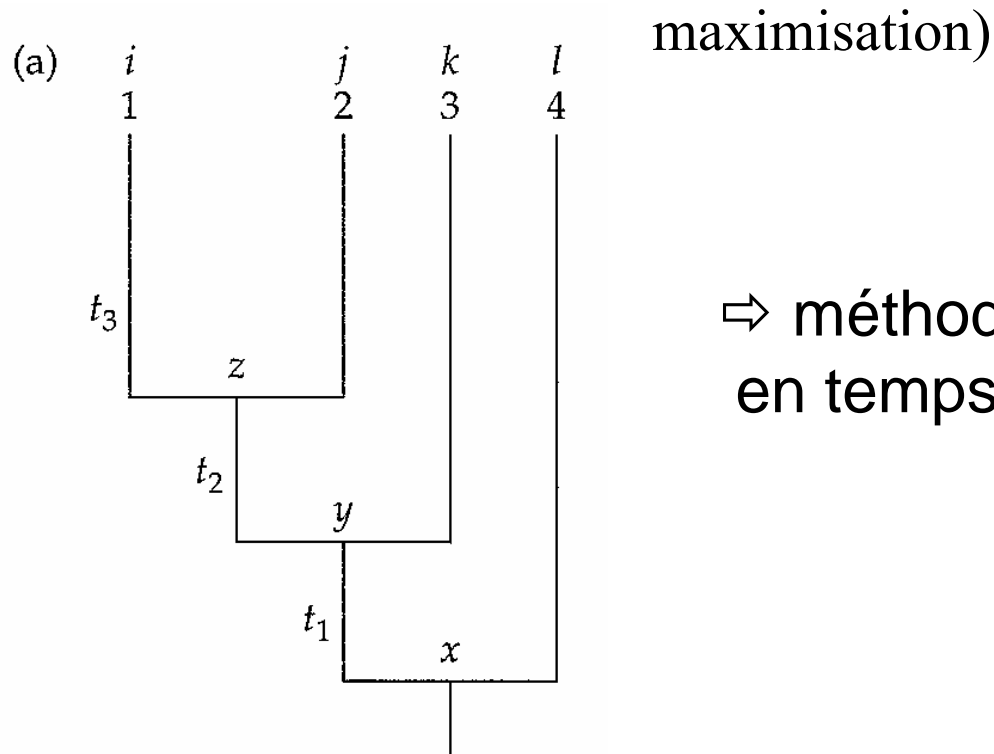


avec  $x,y,z = [A,G,C,T]$

$$\text{et Prob}(i,j,k,l) = P_{xl}(t_1+t_2+t_3) \cdot P_{xy}(t_1) \cdot P_{yk}(t_2+t_3) \cdot P_{yz}(t_2) \cdot P_{zi}(t_3) \cdot P_{zj}(t_3)$$

## 3.3. Les méthodes de maximum de vraisemblance

- 3.3.1. Calcul de la vraisemblance
  - en pratique, **longueurs des branches sont non connues**  $\Rightarrow$  (1) on choisit des valeurs de  $t_1, t_2, t_3$ ; (2) on calcule la vraisemblance totale  $\ln L$ ; et ensuite (3) on prend d'autres valeurs jusqu'à trouver celles qui maximisent la vraisemblance totale  $\ln L$  (méthodes numériques de



$\Rightarrow$  méthodes terriblement exigeantes en temps de calcul!

## 3.3. Les méthodes de maximum de vraisemblance

- 3.3.2. Modèles de substitutions nucléotidiques

- méthode applicable à tous modèles disponibles (grande flexibilité)
- permet l'élaboration de modèles très complexes:
  - ex.: variation des taux de substitutions entre sites nucléotidiques (position dans codon,...)

Longueurs des branches exprimées en nombre attendu de substitutions nucléotidiques:  $v = r.t$ ;  $t$  est constant pour tous sites mais  $r$  peut être utilisé comme une variable avec une distribution entre sites caractérisée par un ou plusieurs paramètres que l'on estime conjointement aux longueurs de branches par maximisation

- permet de tester l'horloge moléculaire:
  - calculer la vraisemblance pour deux arbres de topologie identique mais (1) avec la contrainte de branches strictement proportionnelles au temps et (2) sans cette contrainte (branches de longueurs différentes selon lignées)
  - test d'une différence significative entre les deux likelihoods.

## 3.5. Test de robustesse des reconstructions phylogénétiques

### Méthode du **bootstrap**:

**Principe:** recréer un grand nombre de jeux de données par rééchantillonnage numérique à partir du jeu de données réel; construire un arbre phylogénétique pour chaque jeu de données simulé, en utilisant l'une des méthodes de reconstruction phylogénétique; compter la proportion de ces arbres obtenus par "bootstrap" qui montrent un même regroupement phylogénétique (une même branche dans l'arbre).



# 1.3. Test de robustesse des reconstructions phylogénétiques

## Méthode du bootstrap:

Rééchantillonnage numérique par **tirage aléatoire avec remplacement** des sites nucléotidiques à introduire dans le nouveau jeu de données:

Jeu de données réel:

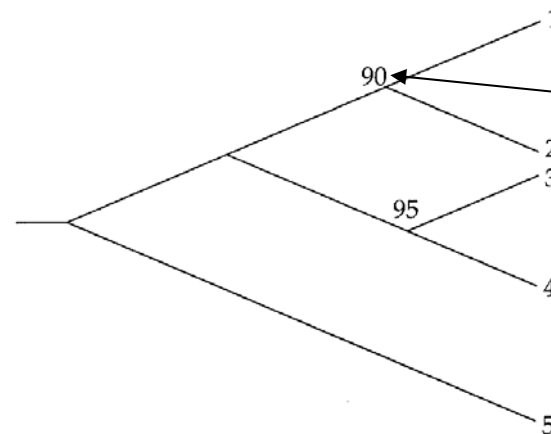
| Taxa | Sequence    |
|------|-------------|
| 1    | GCAGTACT... |
| 2    | GTAGTACT... |
| 3    | ACAATACC... |
| 4    | ACAACACT... |
| 5    | GCGGCATT... |

Jeu de données rééchantillonné:

(tirage aléatoire des sites 6,1,6,...)

|   |             |
|---|-------------|
| 1 | AGATACTC... |
| 2 | AGATGCTT... |
| 3 | AAACACTC... |
| 4 | AAACACCC... |
| 5 | AGACATCC... |

Un total de 100 jeux de données  
et donc 100 arbres sont obtenus  
dont 90 → clade (12)  
et 95 → clade (34)

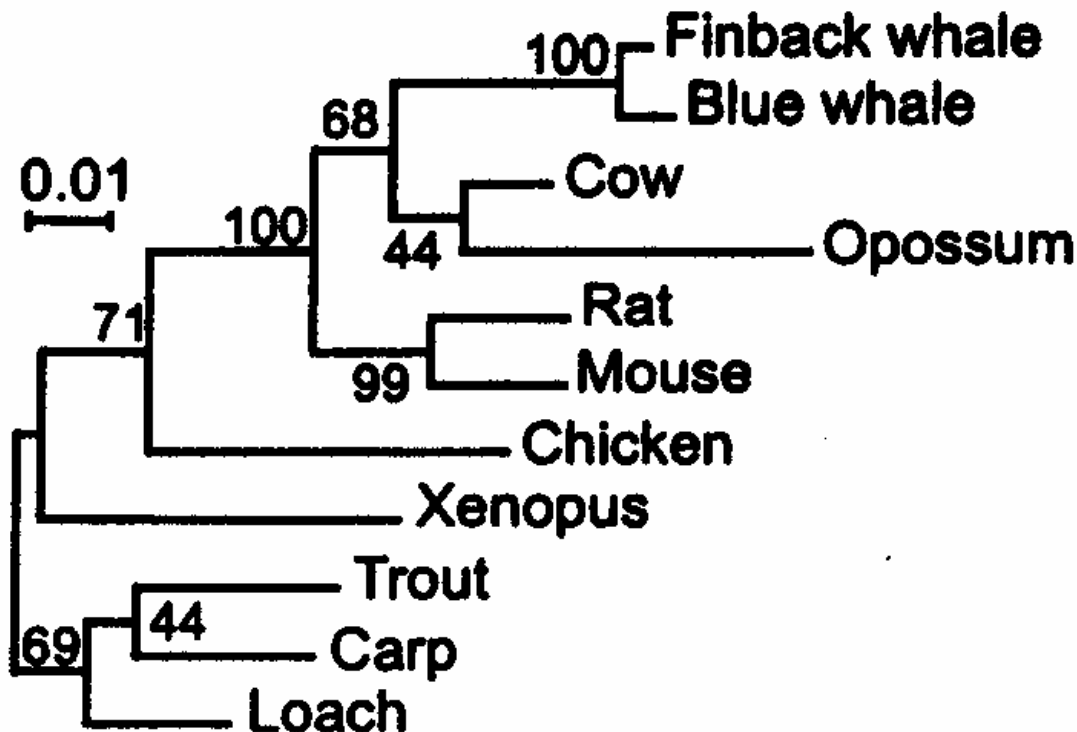


"proportion  
de bootstrap"  
= niveau de  
confiance

# 1.3. Test de robustesse des reconstructions phylogénétiques

## Méthode du bootstrap:

Interprétation des valeurs de **proportion de bootstrap** pour une branche donnée d'un arbre phylogénétique:  $\approx$  probabilité de confiance que la longueur de la branche soit supérieure à zéro (la branche est dite "significative" si  $>95\%$  ou  $99\%$ )



ex. arbre consensus sur 500 bootstraps avec méthode ML sur données de séquences de la cytochrome oxidase (gène mitochondrial)

(Nei & Kumar, 2000)