

Phylogénie

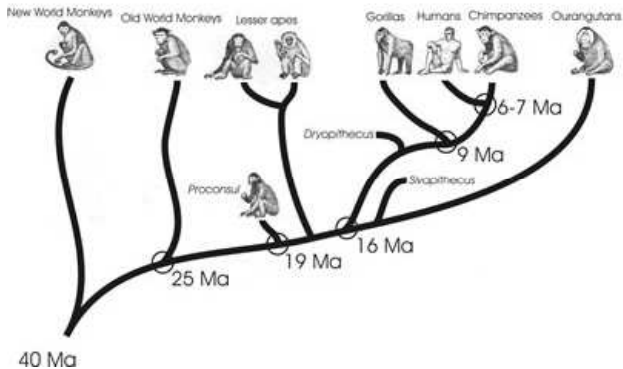
Hélène Touzet

Licence – master

www.lifl.fr/~touzet/M1/

Phylogénie

- ▶ Retracer l'historique des espèces à partir des évolutions observées
- ▶ 1859 – Darwin
"Les lignes généalogiques de tous les êtres organisés".
- ▶ 1866 – Haeckel
"L'enchaînement des espèces animales et végétales au cours du temps"



Les buts de la reconstruction phylogénétique

- ▶ **C**omprendre l'origine de la vie
- ▶ **E**tudier la biodiversité.
- ▶ **D**éterminer l'origine géographique des espèces
- ▶ **C**omprendre les mécanismes moléculaires
- ▶ etc.

Caractères morphologiques

- ▶ Exemple : phylogénie des ongulés
 - ▶ dentition : présence d'incisives, de canines
 - ▶ nombre d'ongles
 - ▶ appareil digestif : ruminant ou pas



Espèce	ongles	dentition	appareil digestif			
bison	2	ni incisives, ni canines	ruminant	0100	00	1
hippopotame	4	incisives et canines	estomac simple	0001	11	0
rhinocéros	3	incisives	estomac simple	0010	10	0
vache	2	ni incisives, ni canines	ruminant	0100	00	1
zèbre	1	incisives	estomac simple	1000	10	0



Limites des caractères morphologiques



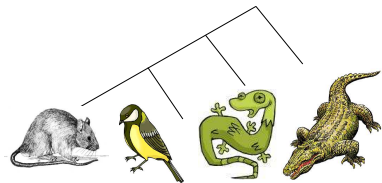
dauphin



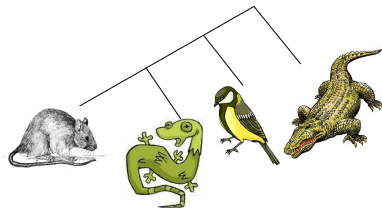
thon

Limites des caractères morphologiques

	rat	oiseau	lézard	croco
écailles	0	0	1	1
ovipare	0	1	1	1



	rat	oiseau	lézard	croco
écailles	0	0	1	1
ovipare	0	1	1	1
oeil	0	1	0	1
gésier	0	1	0	1





Archaeopteryx - oiseau dinosaure

Phylogénie moléculaire

- ▶ **Données** : séquences
séquences ARN ribosomiques, protéines
- ▶ **Evolution** : mutations
- ▶ **En pratique**
 - ▶ alignement multiple
 - ▶ colonne = caractère

éléphant	FVNQHLCGSHLVEALYLVCGERGFFYTPKTGIVEQCCTGVCSLYQLENYCN
hamster	FVNQHLCGSHLVEALYLVCGERGFFYTPKSGIVDQCCTSICSLYQLENYCN
baleine	FVNQHLCGSHLVEALYLVCGERGFFYTPKAGIVEQCCASTCSLYQLENYCN
alligator	AANQRLCGSHLVDALYLVCGERGFFYSPKGGIVEQCCHNTCSLYQLENYCN

Méthodes de parcimonie

Rasoir d'Occam

*"Pluralitas non est
ponenda sine neccesitate"*

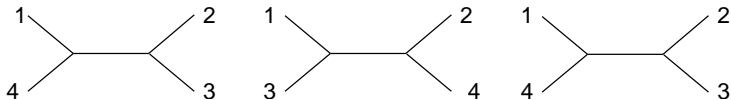


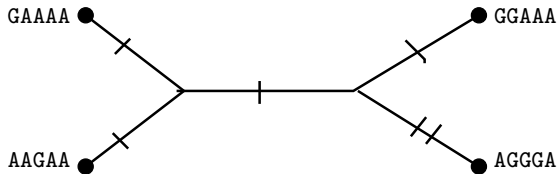
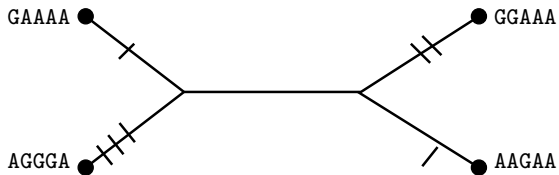
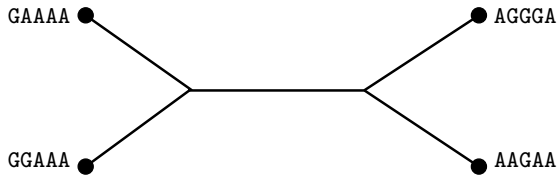
- ▶ **P**rivilégier l'arbre qui minimise le nombre de mutations
- ▶ **L**e nombre global de mutations est obtenu en faisant la somme des mutations le long de chaque branche

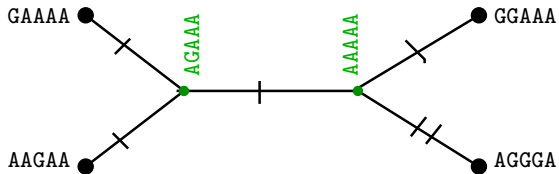
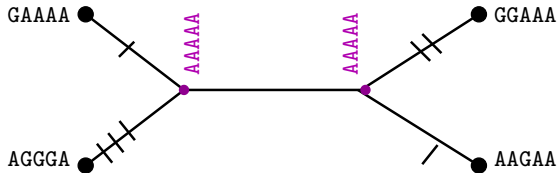
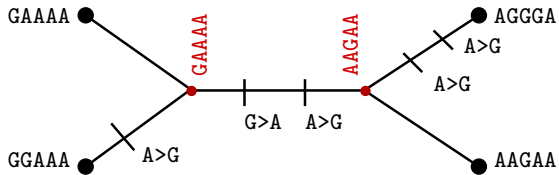
► Exemple

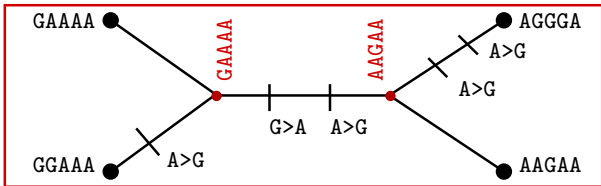
G A A A A	orang-outan
G G A A A	gorille
A G G G A	chimpanzee
A A G A A	humain

► 4 espèces : 3 arbres non enracinés possibles

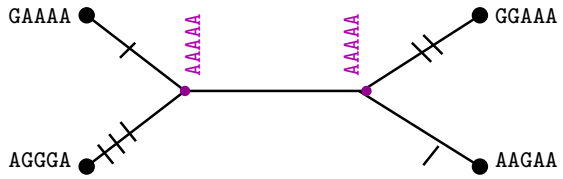




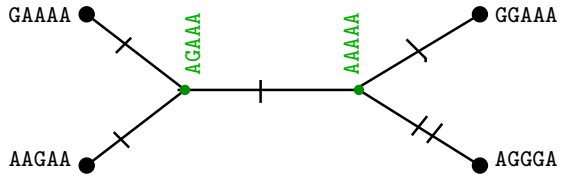




coût: 5 mutations

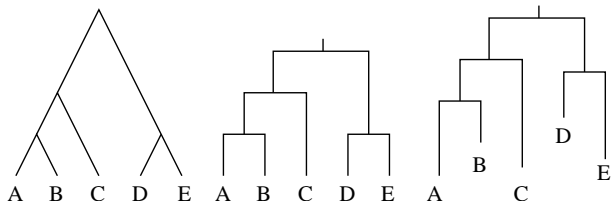


coût: 7 mutations

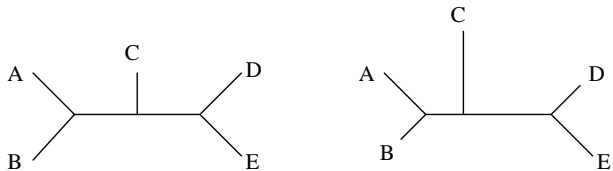


coût: 6 mutations

► **Arbre enraciné**



► **Arbre non-enraciné**



Solution algorithmique

- ▶ **Problème de décision associé NP-complet**
- ▶ **Algorithme exact de **branch and bound****
 - ▶ coût initial : L
 - ▶ arbre initial : 3 séquences
 - ▶ exploration en profondeur de l'espace des solutions en ajoutant une séquence à chaque étape
 - ▶ abandon d'une branche dès que l'on rencontre un arbre de coût supérieur à L
 - ▶ mise à jour de L quand on rencontre un arbre complet de meilleur score

Permet de traiter jusqu'à 20 à 30 données, surtout si les séquences sont proches

- ▶ **Heuristiques:** méthodes de distance, ...

Méthodes de distance

- ▶ **P**oint de départ : alignement multiple
- ▶ **M**atrice de toutes les distances deux à deux
- ▶ **C**lassification hiérarchique

On construit l'arbre à partir des feuilles en regroupant progressivement les noeuds 2 à 2 pour former des **clusters**.

Exemple : UPGMA

- ▶ **U**nweight **P**air **G**roup **M**ethod with **A**rithmetic mean
- ▶ **À** chaque étape, on regroupe les deux clusters les plus proches
- ▶ **La** distance entre le nouveau groupe et les autres éléments est calculée en faisant la moyenne arithmétique

1	G A A A A	orang-outan
2	G G A A A	gorille
3	A G G G A	chimpanzee
4	A A G A A	humain

Matrice initiale

	1	2	3	4
1	0	1	4	2
2		0	3	3
3			0	2
4				0

Après une itération

	1 + 2	3	4
1 + 2	0	3,5	2,5
3		0	2
4			0

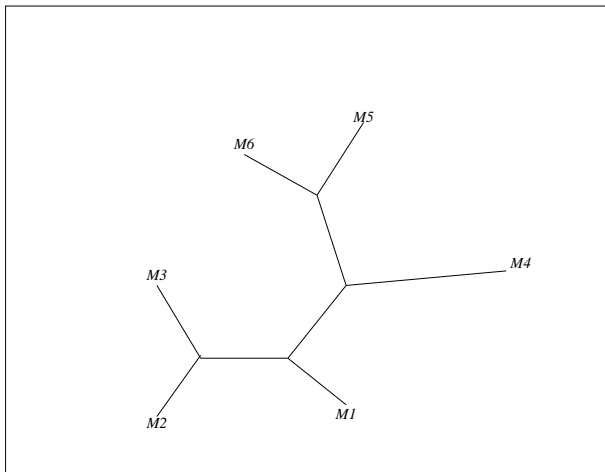
Robustesse de l'arbre : bootstrap

- ▶ **E**volution indépendante des caractères
- ▶ **P**rotocole en 3 étapes (réalisé au moins 100 fois) :
 1. réalisation d'un alignement artificiel à partir de l'alignement initial en prenant arbitrairement n colonnes (avec remise),
 2. estimation de l'arbre obtenu T' ,
 3. comparaison des arbres T et T' : ils doivent être identiques.
- ▶ **S**i on obtient toujours le même arbre, celui-ci semble pertinent

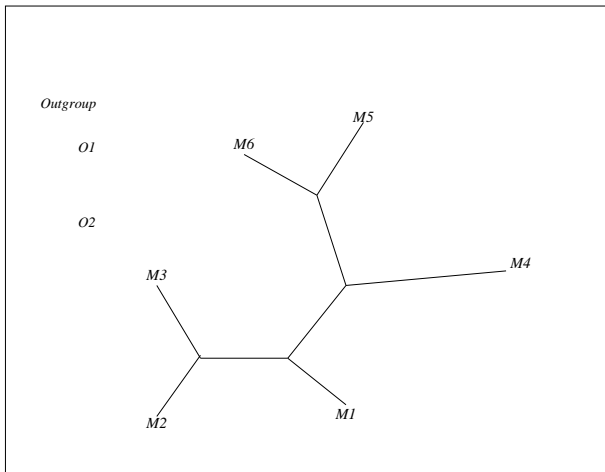
Enracinement de l'arbre phylogénétique

- ▶ **P**assage d'un arbre non-enraciné à un arbre enraciné
- ▶ **A**jout d'un groupe externe (outgroup)
Ensemble de séquences venant d'espèces assez éloignées
- ▶ **L**a racine se situe sur la branche menant à l'outgroup

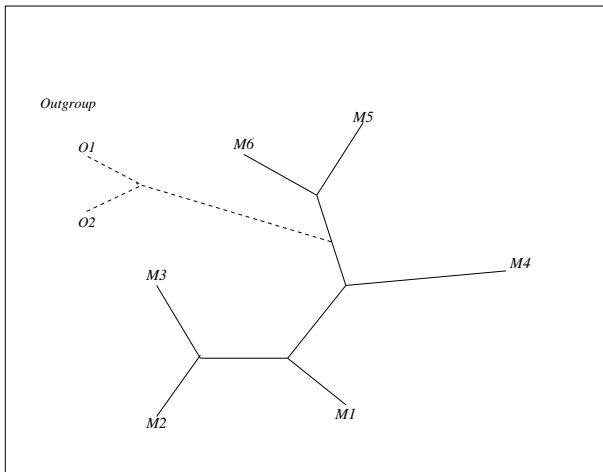
Enracinement



Enracinement



Enracinement



Enracinement

