

# Traitement informatique des données

## **Présentation du module:**

Cette matière contribue au développement des acquis de l'étudiant en ce  
Qui concerne les nouvelles bases de la bioinformatique

### **Connaissances préalables recommandées**

Génétique  
Biologie moléculaire  
Microbiologie  
Biochimie

## **Cours 02: Initiation à la phylogénie moléculaire**

Chargée de cours : ALIANE S

### 1. Introduction :

La phylogénie est l'étude de la relation d'évolution entre des groupes d'organismes (espèces, populations). Ce que nous permet de retracer les principales étapes de l'évolution des organismes depuis un ancêtre commun et ainsi de classifier (taxonomie) plus précisément les relations des parentés entre les êtres vivants.

La taxonomie est une discipline qui permet de classifier, identifier, et nommer les organismes. Ceci est basé sur les caractéristiques communes des espèces.

### 2. Les données de la phylogénie

#### 2.1. Données phylogénétiques

Ces données correspondent aux caractères observables citant les traits morphologique, biochimiques et physiologique, patterns binaires (présence/ absence d'un caractère donné).

Matrice des caractères							
	1 Sacs Aériens	2 Appendices Pairs	3 Mandibule	4 Glandes Mammaires	5 Ailes	6 Dents	7 Colonne Vertébrale
Truite	0	0	0	0	0	0	0
Chauve-souris	1	1	1	1	1	0	0
Homme	1	1	1	1	0	1	0
Pigeon	1	1	0	0	1	1	0

#### 2.2. Données moléculaires

Les données moléculaires sont des marqueurs, le plus souvent des séquences particulières de gènes (ARNm, RFLPs, Microsatellites, SNPs, ITS).

## alignement multiple de séquences homologues

goshawk	GALSKLSDLHAQKLRVDPVNFKLLGOCFLVVVAIHHP SVLTPPEVHASLDKFLCAVGNVLTAKYR
vulture	GALSKLSDLHAQKLRVDPVNFKLLGOCFLVVVAIHHP SVLTPPEVHASLDKFLCAVGNVLTAKYR
duck	GALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPAALTPEVHASLDKFMCAVGAVLTAKYR
alligator	GALCRLSELHAHSLRVDPVNFKFLAHCVLVVFAIHHP SALSPEIHASLDKFLCAVSAVLTAKYR
lesser	GALSALSDDLHAHKL RVDPVNFKLLSHCLLVTLACHHPAEFTPAVHASLDKFFSAVSTVLTAKYR
giant	GALSALSDDLHAHKL RVDPVNFKLLSHCLLVTLASHHPAEFTPAVHASLDKFFSAVSTVLTAKYR
moose	GTLSDLSDDLHAHKL RVDPVNFKLLSHTLLVTLAAHLP SDFTPAVHASLDKFLANVSTVLTAKYR
axolotl	ATLVKLSDKHAHDL MVDPAEFPR LAEDILVVLGFHLP AKFTYAVQCSIDKFLHVTMRLCI SKYR

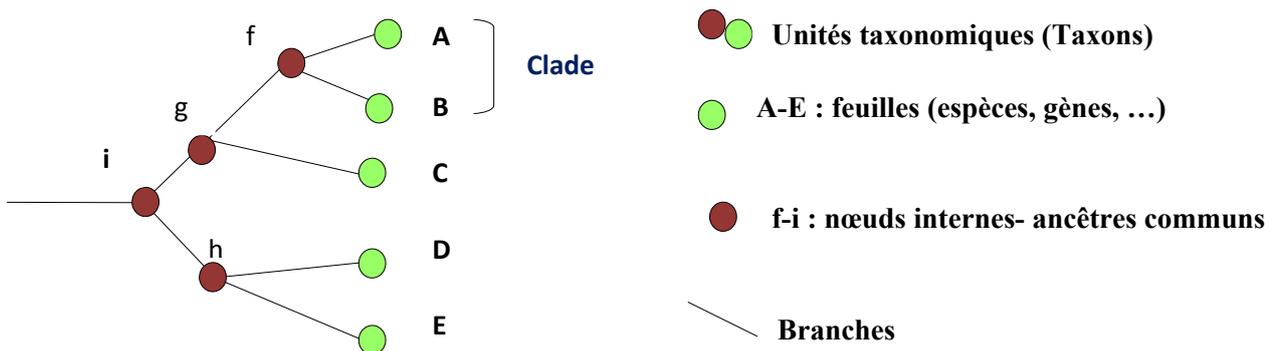
### 3- Arbres phylogénétiques

Le premier objectif de l'étude phylogénétique est reconstruction de l'arbre de vie de toutes les espèces vivantes à partir des données génétiques observées.

L'arbre phylogénétique est un graphe acyclique ; ensemble des nœuds (ou sommets) connectés par les arêtes (ou branches) de telle sorte que toute paire de nœuds est reliée par un chemin.

La phylogénie moléculaire renseigne sur les changements survenus au niveau des séquences au cours de l'évolution (modifications graduelles au cours du temps).

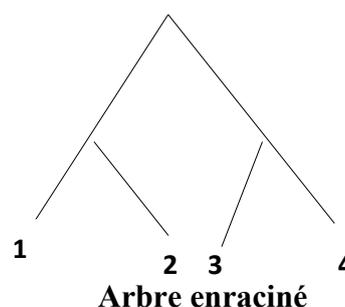
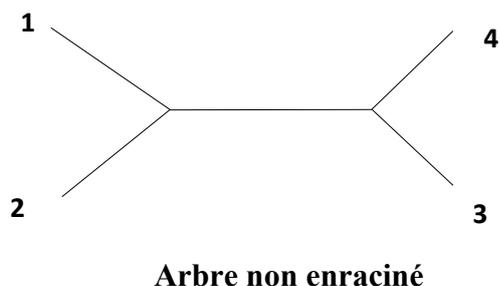
#### 3.1 Structure d'un arbre phylogénétique



### 3.2. Arbre enraciné et arbre non enraciné

Dans l'**arbre enraciné**, la racine représente l'ancêtre commun le plus récent de tous les taxons considérés. Un arbre enraciné est donc dirigé et prend un « chemin évolutif » de l'ancêtre commun aux taxons actuels.

Un **arbre non-enraciné** ne représente que les relations entre les taxons.



### 3.3. Méthodes de construction

Un arbre phylogénétique est un arbre schématique qui montre les relations de parentés entre des groupes d'êtres vivant. Il existe plusieurs techniques de construction des arbres phylogénétiques, plus ou moins rapides et plus ou moins fiables. Trois méthodes sont le plus souvent utilisées :

#### 3.3.1. Méthode des distances (Unweighted Pair Group Method with Arithmetic Mean)

C'est une méthode basé sur les similarités entre paires de séquences, fondée sur l'alignement multiple et le calcul des distances entre ces séquences. La topologie de l'arbre est construite par la méthode Neighbor-Joining (NJ), c'est aussi une méthode de distance, qui consiste à calculer la longueur des branches, de telle manière à ce que les distances déduites de l'arbre soient les plus proches des distances mesurées. Elle a l'avantage d'être vraiment rapide. En général, elle est utilisée pour faire des arbres de plusieurs milliers de séquences. C'est la méthode utilisée par CLUSTALW.

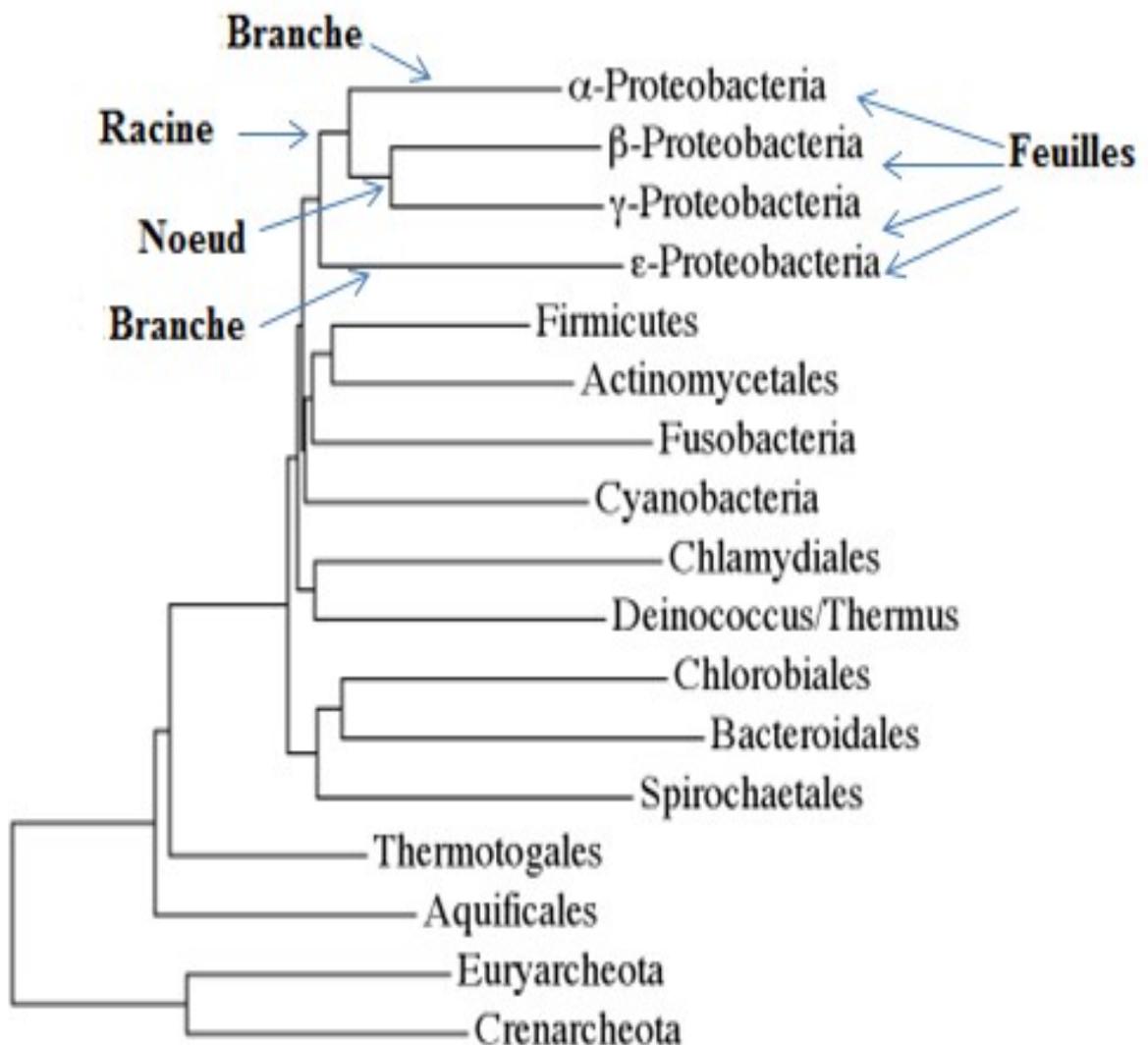
#### 3.3.2. Méthode de parcimonie (Maximum Parsimony)

Elle est fondée sur une méthode probabiliste qui recherche l'arbre le plus parcimonieux parmi tous les arbres possibles, c'est à dire celui qui nécessite le moins de changements pour expliquer les différences observées entre les séquences. C'est une méthode très lente par

rapport à la méthode des distances et pas aussi précise que la méthode de maximum de vraisemblance (ML).

### 3.3.3. Méthode du maximum de vraisemblance (Maximum Likelihood-ML)

Cette méthode est souvent décrite comme étant la meilleure méthode et la plus fiable, c'est-à-dire la plus efficace pour trouver l'arbre le plus proche de la réalité. Elle repose sur un ou plusieurs caractères à étudier. Il s'agit d'une méthode probabiliste qui nécessite un modèle d'évolution. Le choix de ce modèle est crucial pour la qualité de l'arbre. L'inconvénient de cette méthode qu'elle est longue et nécessitant une grande puissance de calcul.



**Fig.1:** Arbre phylogénétique.

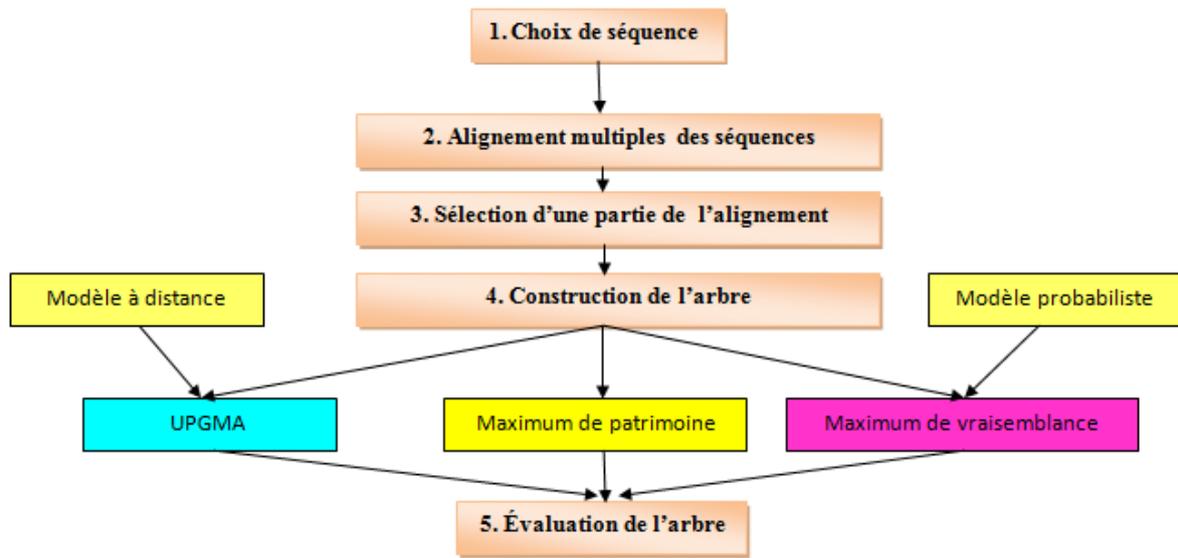


Fig.2: Construction d'un arbre phylogénétique.