



وزارة التعليم العالي و البحث العلمي  
جامعة عباس لغرور - خنشلة  
كلية علوم الطبيعة و الحياة

*Ministère de l'enseignement supérieur et de la recherche scientifique*

*Université Abbes LAGHROUR*

*Faculté des sciences de la vie et de la nature*

*Département de génétique et biologie moléculaire.*



Matière :

Génétique évolutive L3 Génétique

Enseignant. BENSADA Mostefa MCB

# Sommaire du cours génétique évolutive

---

I.	L'ADN est le lieu des changements génétique.....	5
I.1	La double hélice :.....	5
I.2	Structure des pyrimidines .....	5
I.3	La réplication de l'ADN .....	6
I.4	Les lésions des nucléotides :.....	7
I.5	Les mécanismes de réparation de l'ADN sont sources de mutations:.....	8
I.6	La recombinaison génétique participe à la diversité génétique :.....	9
I.7	<i>La variation du génome</i> est assurée par d'autres mécanismes :.....	9
I.8	Les Mutations comme sources de variabilité :.....	9
I.9	Conséquences des mutations. ....	10
II.	Les théories de l'évolution : .....	11
II.1	Lamarckisme.....	11
II.2	Le darwinisme.....	11
II.3	Théorie synthétique de l'évolution : .....	11
II.4	La théorie neutraliste de Kimura (1968) :.....	11
II.5	L'horloge moléculaire : .....	11
II.6	Quelques définitions .....	13
II.7	l'arbre du vivant :.....	13
III.	La biodiversité :.....	14
III.1	Origine de La biodiversité :.....	14
III.2	Niveaux de biodiversités :.....	14
IV.	Sélection naturelle et dérive génétique.....	15
IV.1	La génétique et l'évolution :.....	15
IV.2	La génétique et la sélection naturelle :.....	15
IV.4	Hierarchie de la sélection naturelle :.....	15

V.	Les origines de la biodiversité « Spéciation ».....	16
V.1	Mécanismes de la spéciation .....	16
V.2	Type de spéciation :.....	16
V.3	La génétique et l'adaptation : .....	17
V.4	Dérive génétique :.....	17
V.5	Effet fondateur et génétique :.....	17
VI.	Evolution des génomes :.....	18
VI.1	Impact des éléments transposables sur les génomes.....	18
VI.2	Les éléments transposables dans les génomes : .....	19
VI.3	Evolution de la structure des gènes par brassage d'exons. ....	19
VI.4	Pseudogènes et rétropseudogènes : .....	19
VII.	Introduction à la phylogénie moléculaire .....	21
VII.1	Définition d'un caractère moléculaire : .....	21
VII.2	La phylogénétique.....	21
VII.2.1	La classification phylogénétique : .....	21
VII.2.2	Phylogénie : .....	22
VII.2.3	éléments de la construction phylogénétique :.....	22
VIII.	Caractères et états de caractères.....	25
IX.	Les méthodes phénétiques :.....	26
IX.1	Cladisme :.....	26
IX.2.	Approches cladistiques :.....	27
IX.3	Distance évolutive degré de dissemblance.....	27
X.	Différentes représentations graphiques des arbres .....	27
X.1	La parcimonie et le Maximum de Parcimonie.....	28
X.2	Le maximum de parcimonie .....	29
X.3	Principes de calculs .....	30
X.4	La mesure de l'homoplasie.....	31
X.5	Les arbres de consensus.....	31

X.6 Le format NEWICK ou Les arbres en notation parenthésée : .....	31
X.7 La méthode UPGMA.....	32

## Références bibliographiques

## *Préambule :*

---

S'il y a des matières enseignées dans nos universités émanant de connaissances et de principes qui ne sont pas évident à prodiguer à nos étudiants, *la génétique évolutive* en ait l'illustration. Cette Matière fait appel à des connaissances essentielles acquises aux premiers stades du processus LMD. Pour appréhender les enseignements de cette matière, des connaissances de la génétique mendélienne et des populations, la biologie moléculaire et l'étude des structures des gènes et des génomes, qui viennent se combiner à la botanique, l'écologie et à la bioinformatique sont essentiels. De plus, les connaissances acquises lors de l'enseignement de cette matière se prolongent dans les domaines de la recherche sur les maladies complexes ; maladies métaboliques, le diabète, le cancer, l'Alzheimer... En effet, les modèles d'analyse en phylogénétique moléculaires sont les mêmes utilisés dans les programmes de modélisation à la recherche de la signature moléculaire de ces pathologies.

Si les études des théories de l'évolution des espèces a permis la classification du vivant en espèces, familles, ordres.... La phylogénétique a permis de faire de ces derniers des entités interconnectés et permet donc de mieux comprendre l'évolution des caractères. Afin de faciliter l'enseignement de cette matières 2 chapitres sont consacrés aux rappels de bases sur les natures chimiques de modifications qui affectent l'ADN et aux diverses théories de l'évolution des espèces.

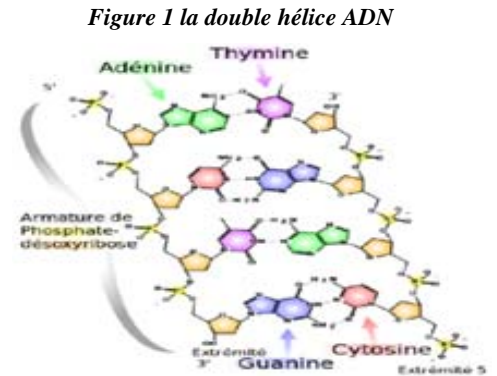
Un 3<sup>ième</sup> chapitre aborde à son tour l'évolution d'un point de vue écologique, par l'étude de la biodiversité des écosystèmes et de la spéciation. Le 4<sup>ième</sup> chapitre se décline sous la forme d'une jonction entre la génétique et l'évolution, par l'analyse des moteurs moléculaires de la biodiversité, l'effet fondateur et la dérive génétique. Enfin les chapitres 5 et 6 introduisent et expliquent d'une manière certes, succincte, les apports de la génétique dans l'analyse de l'évolution : les bases de calcul et de la construction de l'analyse phylogénétique.

# Chapitre I : L'ADN et les modifications génétiques

## I. L'ADN EST LE LIEU DES CHANGEMENTS GENETIQUE

### I.1 La double hélice :

L'ADN forme souvent des doubles brins (double hélice) par complémentarité de séquences et elle est le support de l'information génétique. Les 4 types de bases qui composent l'ADN sont: Adénine (A), Cytosine (C), Guanine (G) et Thymine (T). Un nucléotide est formé d'un sucre (ribose dans ARN et désoxyribose dans ADN), d'un groupe phosphate et d'une base.

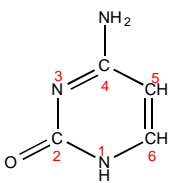


Les bases pyrimidines C et T : ce sont des précurseurs monomériques de l'ADN (maintien de l'information) et de l'ARN (expression de l'information). Mais aussi des précurseurs énergétiques.

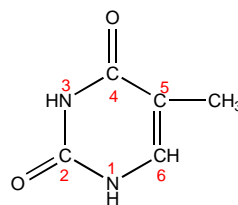
### I.2 Structure des pyrimidines

Elles se définissent comme étant un noyau hétérocyclique à 6 atomes : 4 C et 2 N. Dans l'ADN : cytosine et thymine, dans l'ARN : cytosine et uracile.

- Bases pyrimidiques de l'ADN

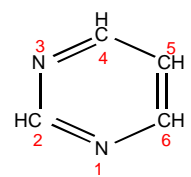


Cytosine :

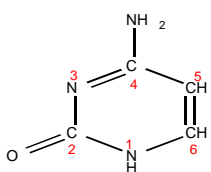


Thymine :

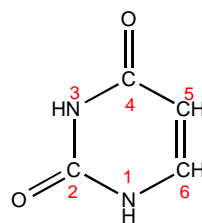
Noyau pyrimidine



Bases pyrimidiques de l'ARN

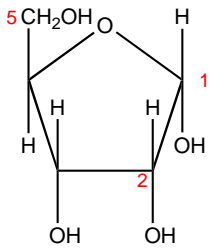


Cytosine

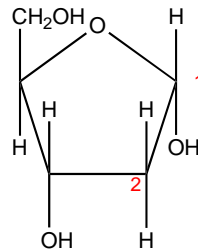


Uracile

Ces bases sont associées à un sucre à 5 carbones formant un nucléoside

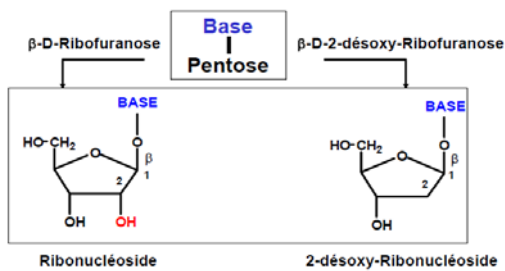


le ribose dans l'ARN



le désoxyribose dans l'ADN

L'association base + sucre + Ester de phosphate (sur C<sub>5</sub> du ribose) donne un **nucléotide**.



### I.3 La réplication de l'ADN

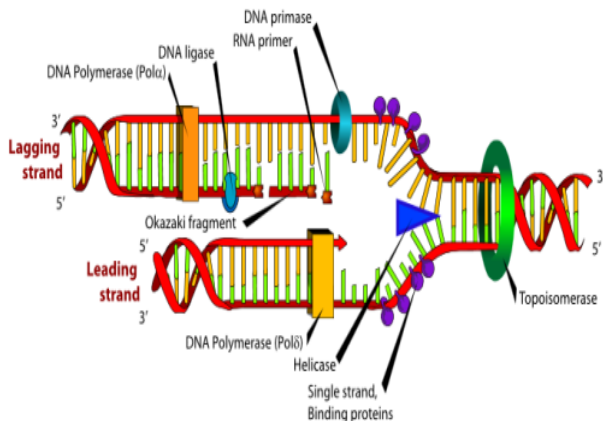


Figure 2 : la processus de réplication chez les eucaryotes.

L'ADN ne peut servir de lien génétique que si la séquence nucléotidique est strictement reproduite avant d'être transmise à la descendance. L'ADN se réplique de façon semi-conservative, et selon une orientation définie et par des mécanismes qui diffèrent en fonction de l'organisme, virus, procaryotes ou eucaryotes. Sous le contrôle d'enzymes : hélicases, polymérase III....

#### I.3 Les modifications des bases :

Les mutations sont des modifications du matériel génétique. Elles peuvent avoir des conséquences délétères, donc rarement avantageuses pour l'organisme et elles sont à l'origine des modifications évolutives. On doit distinguer entre les mutations germinales qui affectent les gamètes et sont transmissibles et les mutations somatiques qui affectent les autres cellules (non transmissibles).

Une mutation est rarement réversible : le plus souvent le gène muté est réparé ou détruit. La fréquence des mutations est liée, d'une part à la dimension du gène (plus le gène est "volumineux" ou plus il renferme de régions codantes et plus la probabilité de mutation est forte), d'autre part à l'existence dans le gène de séquences fortement mutagènes ou "points chauds de mutations". Elles peuvent être silencieuses c'est à dire ne pas

engendrer de modifications d'acides aminés dans la protéine (ce qui est le plus souvent pour les modifications affectant la troisième base du triplet).

Aberrations	Facteurs	Réparation
Perte d'une base : Apyrimidique ou Apurinique	1100/ jour	Excision réparation
Désamination (NH <sub>2</sub> )	Instabilité	Excision réparation
Hydroxylation (+OH)	Radiations, Produits chimiques.	Excision réparation
Alkylation, méthylation	Donneur cellulaire	Excision réparation
Dimères de TT ou CC	Rayons UV	Poste réplivative
Mauvaise séquence	Erreurs de réplication	Correction d'appariements
Coupures de brins	Radiations Ionisantes	Excision réparation

#### La dégradation spontanée d'une base

Un autre processus de mutagenèse est la dégradation spontanée d'une base. La désamination de C en U est fréquente. Elle peut être réparée par un processus qui détecte l'uracile. Dans le cas contraire, le U engendre en vis à vis l'insertion d'un A et entraîne une transition de GC en TA lors de la réplication. La désamination de la méthylcytosine en T peut aussi intervenir. La méthylcytosine est présente dans le génome humain dans les séquences 5'-CpG-3' qui n'existe normalement pas dans les séquences codantes des gènes. Cependant, si elle existe et si elle est désaminée en T, aucun système de réparation ne peut le reconnaître (car T est une base normale). Le doublet CpG est donc un « point chaud » de mutations.

**Tableau 1 : Les modifications chimiques des nucléotides.**

Une mutation spontanée résulte d'un processus naturel, que l'on doit la distinguer des mutations induites qui résultent d'une interaction entre l'ADN et un agent extérieur ou un mutagène (des composés génotoxiques ; comme la chaleur, les radiations ionisantes ou les UV). Toutefois les mutations qui résultent d'erreurs de réplication sont spontanées.

#### **I.4 Les lésions des nucléotides :**

Les lésions de l'ADN ont plusieurs origines dont certaines sont liées à l'action d'agents chimiques, alors que d'autres sont dues à la chaleur ou aux radiations ionisantes et les UV. Les conséquences sont aussi multiples, qui ont des cassures simples ou doubles brins, à la mutation par perte de base, altération de la base ou sa substitution.



Les lésions de l'ADN	Exemple/ cause
Perte de base	Suppression de purines par les acides et la chaleur, élimination de bases altérées par les ADN glycosylases.
Altération de base	Radiations ionisantes et les agents alkylants.
Base incorrecte	Les mutations affectant la fonction d'édition exonucléasique 3' 5'; incorporation de base incorrecte.
Insertion ou délétion de nucléotide	Agents intercalant (ex: acridine) qui induisent une addition ou une délétion de nucléotide.
Cassures simple ou double brin	Cassures de liaison phosphodiester par des radiations ionisantes et des agents chimiques.

Les mécanismes de réparation de l'ADN sont eux aussi une source de mutations et donc sources de variabilité génétique. Les systèmes de contrôles et la réparation, qui permet la correction des erreurs d'appariement. En particulier les mécanismes d'excision-réparation de bases ou d'excision-réparation de nucléotides, sont des systèmes de réparations post-répliquatives.

Tableau 2 : les principales lésions de l'ADN

### I.5 Les mécanismes de réparation de l'ADN sont sources de mutations:

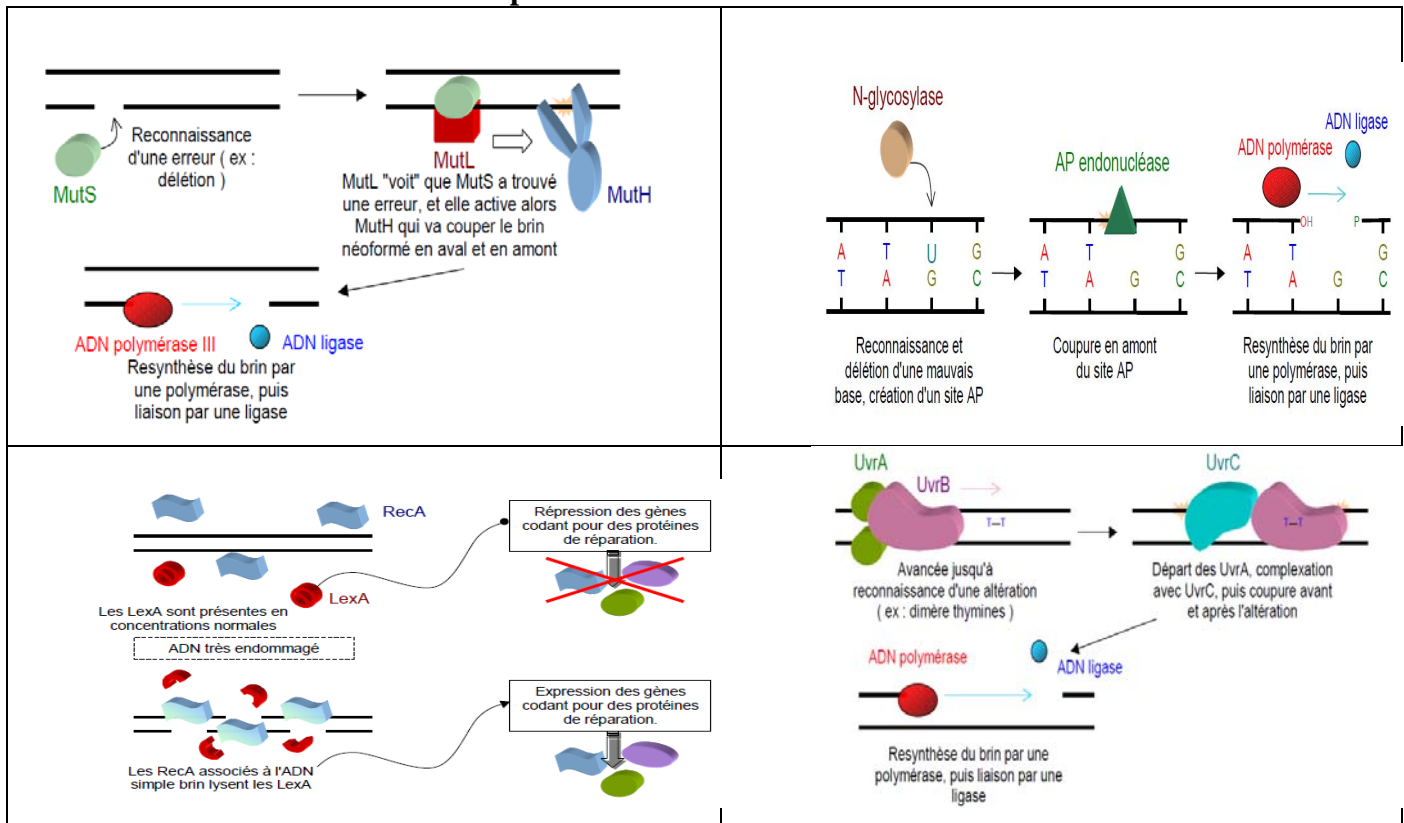


Figure 3 : mécanismes de réparation de l'ADN et sources de mutations

Par exemple Le système SOS. Ces systèmes si ils sont défaillant permettent l'apparition de mutations qui seront introduites dans l'ADN lors de le mécanisme de répllication.

## I.6 La recombinaison génétique participe à la diversité génétique :

La recombinaison génétique intervient dans de multiples événements cellulaires et participe aussi bien à la constance qu'à la variabilité du génome.

La recombinaison génétique permet *la réorganisation du génome* :

- soit par réorganisation au hasard des chromosomes à la méiose,

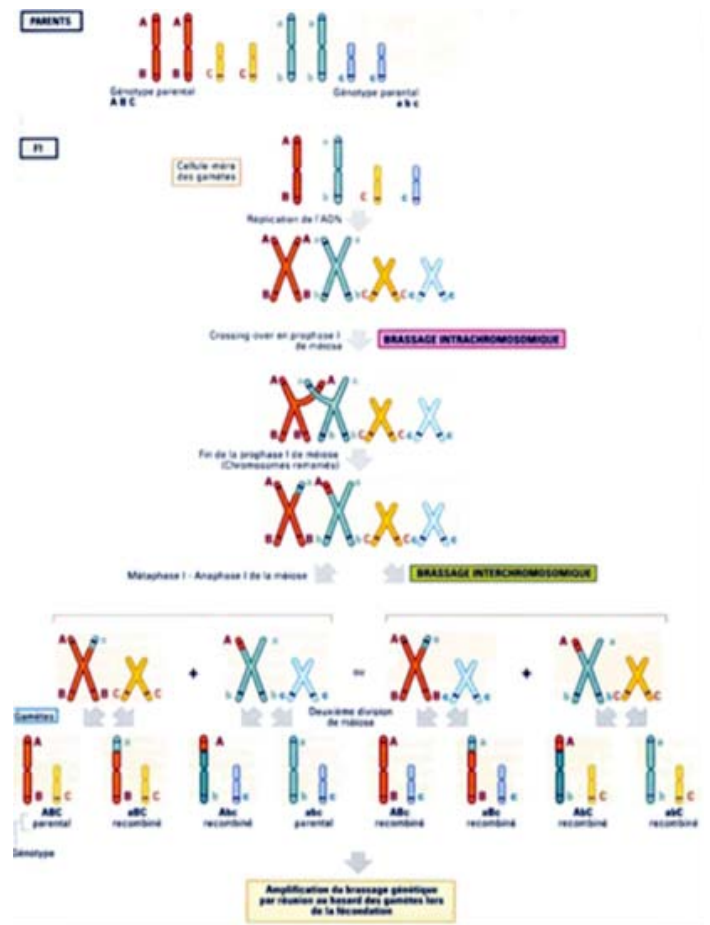
- soit plus spécifiquement par cassure et recollage des brins d'ADN. Alors que **La stabilité du génome** assurée par la :

- *Fidélité de la réplication de l'ADN*,

- *Précision de la ségrégation* en mitose et en méiose des chromosomes,

- *la Réparation des altérations* de l'ADN au cours de la vie cellulaire.

**Figure 4 : mécanisme de recombinaison est le moteur de la variabilité génétique.**



## I.7 La variation du génome est assurée par d'autres mécanismes :

Les réarrangements d'allèles à la méiose par : le Crossing-over, les Mutations, les Réarrangements, les Transpositions et enfin les Conversions géniques.

Par exemple, chez l'homme : En considérant uniquement le brassage interchromosomique, le spermatozoïde et l'ovule possèdent une garniture allélique parmi les  $2^{23}$  possibles (23 nombre haploïde des chromosomes). Le nombre de cellules-œuf que la fécondation peut générer est de  $2^{46}$  ( $2^{23} \times 2^{23}$ ) soit plus de 70 000 milliards. *La Méiose et la fécondation sont donc sources de variabilités génétiques.*

## I.8 Les Mutations comme sources de variabilité :

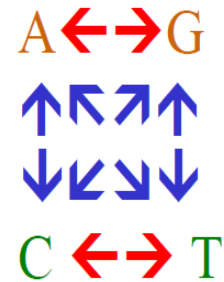
Lors de la réplication de l'ADN, la polymérase en charge de la synthèse des nucléotides est susceptible de faire des erreurs qui sont en principe réparées par des systèmes de contrôles enzymatiques. Mais, des erreurs dans la séquence nouvellement synthétisée peuvent survenir et qui échappent aux systèmes de contrôle et de réparation. Ces mutations sont alors héritées par la cellule fille. Si ces mutations sont silencieuses elles seront

considérées comme un polymorphisme et leur multiplication d'une division cellulaire à l'autre finissent par créer une différence génétique notable avec la cellule mère.

Types de mutations :

- Les **substitutions**, où la base est remplacée par une autre.
- Les **additions**, où une base est ajoutée dans la séquence d'ADN.
- Les **délétions** où l'on observe la disparition d'une base.

Il y a 12 possibilités de changements : 4 transitions possibles et 8 transversions.



### **I.9 Conséquences des mutations.**

L'ADN n'est pas un élément statique ; une mutation ponctuelle peut se produire dans un gène (dans un exon ou dans un intron) ou dans une séquence non codante (séquence répétée). Si elle se produit dans une séquence codante, elle sera à l'origine des différentes versions possibles d'un gène, c'est-à-dire à l'origine des *allèles* d'un gène ou poly-allélisme si le cas de plusieurs gènes (fréquence inférieure à 1%). L'ADN est sujet à des modifications silencieuses et transmissibles. Ces variations de séquence sont appelées *polymorphismes* lorsqu'elles surviennent à une fréquence  $\geq 1\%$ .

# Chapitre II : théories de l'évolution

---

## II. LES THEORIES DE L'EVOLUTION :

L'évolution désigne la transformation des espèces vivantes qui se manifeste par des changements de leurs caractères génétiques et morphologiques aux cours des générations. En génétique évolutive la sélection s'exerce sur le **phénotype** et non sur le **génotype**. De même que la sélection naturelle influence globalement l'ensemble des caractères et non caractère par caractère. Il n'y a donc pas de sélection naturelle efficace, car il n'y a pas de conséquences sur la descendance si la sélection n'affecte que le phénotype.

### II.1 Lamarckisme (1744 – 1829) :

La diversification des êtres vivants sous l'effet des circonstances variées. La complexification croissante de l'organisation sous l'effet de la dynamique interne propre à leur métabolisme.

### II.2 Le darwinisme (1809-1882) :

Une vision de la transformation des espèces sous la forme d'un processus de transformation et de ramification. *La sélection naturelle*.

### II.3 Théorie synthétique de l'évolution :

Les variations continues peuvent être expliquées par l'action combinée d'un grand nombre de gènes (Fisher 1918). Apparue après la découverte des lois de la génétique mendélienne.

### II.4 La théorie neutraliste de Kimura (1968) :

Cette théorie explique que le hasard ou la dérive génétique soit presque l'unique clé de l'évolution. Cette théorie est à la base de la génétique évolutive. Il y a fixation aléatoire des mutations et les mutations sont sélectivement neutres. C'est-à-dire que les mutations sont dues au hasard, la sélection existe mais la plus grande partie de la diversité est due au hasard par des mutations spontanées.

### II.5 L'horloge moléculaire :

Les théories modernes de l'évolution se portent sur le rôle que joue le "hasard" ou plus spécifiquement la dérive génétique. Pour la théorie neutraliste de Kimura, la plupart des mutations ne sont pas adaptatives et la majeure partie des mutations est neutre ne procurant ni avantage, ni handicap, alors qu'une autre partie importante est handicapante. Donc l'évolution se fait par petites étapes et impliquant les séquences non codantes plus que les séquences géniques.

De là est venue la théorie de l'horloge moléculaire Zuckerkandl et Pauling : la différence de protéines homologues chez différents organismes est fonction du temps depuis lequel ces organismes ont évolué indépendamment depuis leur dernier ancêtre commun. La ressemblance des gènes ou des protéines est en termes de séquences en nucléotides ou en acides aminés (AA). Cette ressemblance peut être quantifiée par le

pourcentage d'identité en nucléotides ou acides aminés que l'on va retrouver entre les gènes ou les protéines comparés. Si l'identité est importante sur une partie significative des gènes ou protéines, on considérera généralement que cette identité n'est pas survenue par hasard mais qu'elle est le signe d'une ascendance commune (homologie évolutive). Pour une classe donnée de protéines, la vitesse de remplacement des AA est à peu près constante. L'évolution adaptative des protéines peut ne pas requérir beaucoup de changements. La plupart des changements d'AA observables sont sélectivement neutres. Le rôle d'un AA particulier dans une protéine peut être rempli par d'autres AA (dégénérescence du code génétique). Donc, Une séquence d'ADN partagée par plusieurs génomes peut être considérée comme ayant une origine commune.

Exemples : Les immunoglobulines avec des structures secondaires variables, la superposition des parties centrales des protéines permanence des feuillets  $\beta$  dans les feuillets C et V.

Cette théorie souffre de plusieurs approximations :

- donne une probabilité de mutation.
- L'horloge favorise une sélection naturelle, dite positive, qui favorise les mutations favorable, alors que dans la majorité des cas toute mutation est négative.
- Elle ne tient pas compte des mutations qui s'annulent exemple T vers C vers T.
- Dans une population, le nombre de mutations par génération (taux de mutation) peut être élevé, mais on ne peut étudier que les mutations fixées.

# Chapitre III : biodiversité et spéciation

## II.6 Quelques définitions

**Une espèce :** Anciennement, on définissait une espèce comme étant un ensemble d'individus se ressemblant, *interféconds*, ainsi que leur descendance. L'interfécondité est une condition primordiale. Actuellement, cette notion d'espèce a été enrichie par l'idée que ces individus *évoluent conjointement sur le plan héréditaire*.

**Une population :** Une population est un ensemble d'individus se reproduisant ensemble (à l'inverse, les espèces sont des ensembles d'individus potentiellement interféconds). Ce sont donc des individus de la même espèce qui ont la possibilité d'interagir entre eux au moment de la reproduction.

**Taxon (Lam, 1950 ; Mayr, 1953) :** Un taxon est groupe d'organismes reconnu en tant qu'unité formelle à chacun des niveaux de la classification biologique hiérarchisée. *Une autre définition pour un Taxon :* Groupe d'organismes formant une unité bien délimitée à chaque niveau hiérarchique de la classification. Le pluriel de taxon ce dit taxa.

**Exemple :** taxon (la sous-espèce) désigné(e) sous le nom *Canis lupus familiaris* qui englobe la totalité des chiens domestiques. A ne pas confondre avec Taxonomie ou Taxinomie qui est la théorie et pratique de la classification des organismes. Ou plus simplement c'est l'étude des relations entre des groupes d'organismes, ou l'art de classer des entités dans des groupes.

## II.7 l'arbre du vivant :

The Last Universal Common Ancestor ou « LUCA »: Cet être vivant hypothétique serait à l'origine du monde vivant. Il serait à l'origine ARN, puis avec le temps il est passé à un génome ADN. Que LUCA a donné naissance à la branche Bactérie de la vie et qu'à partir de là, les 2 autres branches les Archées et les Eucaryotes. LUCA n'a aucune réalité biologique, mais il est une théorie qui tente d'expliquer.

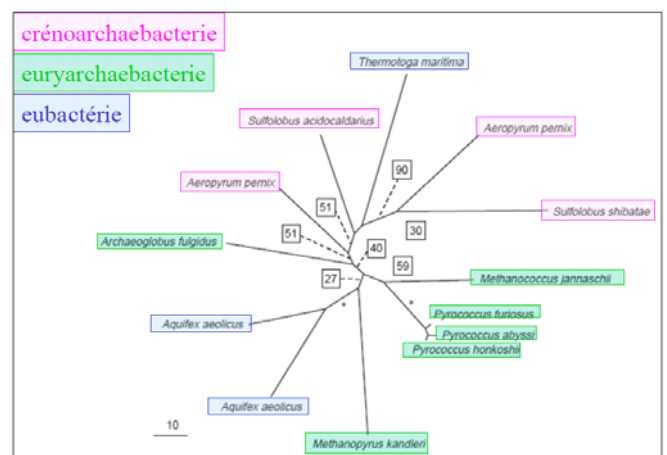


Figure : arbre du vivant avec ces 4 formes Eucaryotes, procaryotes, virus et Archeas.

**Notions à retenir :**

**Parenté** : Tous les êtres vivants sont apparentés, car ils ont un ancêtre commun unique LUCA.

**Similarité** : Notion empirique, la ressemblance peut avoir été acquise par homologie, par convergence ou par conversion de gène.

**Homologie** : ce terme sous-entend un ancêtre commun et pas uniquement une ressemblance remarquée entre deux états de caractère. C'est la relation de deux caractères qui sont descendus, généralement avec divergence, d'un caractère ancestral commun.

**Homologie et similarité** : L'homologie est plutôt une propriété non mesurable contrairement à la similarité. Elle a une connotation génétique. Si le taux de similarité est important (significatif), les séquences peuvent alors être considérées comme homologues, donc supposées descendre d'un même ancêtre.

**Orthologue** : un cas particulier d'homologie, la divergence s'est faite avant la spéciation. Dans le cas de 2 gènes ; ce sont 2 gènes homologues chez différentes espèces ayant des fonctions analogues, ils sont issus d'un ancêtre commun et non la conséquence d'une duplication d'un premier gène.

**Paralogue** : cas particulier d'homologie où dans quelques taxons la divergence s'est faite par duplication de gènes.

**Xénologue** : le gène a été hérité par transfert latéral et n'était pas représenté dans l'ancêtre commun.

### III. LA BIODIVERSITE :

La diversité biologique est la variabilité des organismes vivants de toute origine y compris, entre autres, les écosystèmes terrestres, marins et autres **écosystèmes** aquatiques et les complexes écologiques dont ils font partie; cela comprend la diversité au sein des espèces et entre espèces ainsi que celle des écosystèmes.

#### III.1 Origine de La biodiversité :

La biodiversité trouve son origine dans plusieurs facteurs, comme le niveau de diversité à l'intérieur des populations et l'organisation de la diversité sur un ensemble de populations au sein d'une espèce donnée. Mais quelle diversité utiliser ? La variabilité intraspécifique de caractères quantitatifs? Ou l'estimation de la phylogénie? Ou encore la génétique? Sachant que plus de 99% des espèces ayant existé ont maintenant disparu, donc apparition d'espèces récentes. De même que l'activité humaine a conduit à la disparition d'espèces sauvages et l'apparition des espèces domestiquées.

#### III.2 Niveaux de biodiversités :

- **Biodiversité génétique** : La variété qui existe au niveau des allèles, celui des gènes entiers ou celui de la structure chromosomique à l'intérieur des espèces. Exemple la formule chromosomique (nombre de chromosome) est spécifique à chaque espèce : 46 pour l'homme 38 pour le chat et 4 chez la drosophile.

- Biodiversité spécifique : la variété qui existe au niveau des différentes espèces trouvées dans une aire donnée : une forêt méditerranéenne renferme une variété d'espèces spécifique.

- Biodiversité des écosystèmes : la variété qui existe au niveau des environnements physiques et des communautés biotiques dans un paysage. Exemple au sein même d'une forêt il y a une variété d'espèces qui la peuple.

- Dimension de la biodiversité : Composition : ce qui est présent. Structure : comment les éléments présents sont organisés les uns par rapport aux autres. Fonction : les processus qui génèrent la biodiversité et qui affectent la structure et la composition.

#### IV. SELECTION NATURELLE ET DERIVE GENETIQUE

##### IV.1 La génétique et l'évolution :

La plupart des mutations ne sont pas adaptatives. La majeure partie des mutations est neutre ne procurant ni avantage, ni handicap et une autre partie importante est handicapante. La théorie neutraliste qui veut que le hasard ou la dérive génétique soit presque l'unique moteur de l'évolution. Donc ce qui fait évoluer les génomes sont les événements qui n'affectent pas la survie de l'espèce.

##### IV.2 La génétique et la sélection naturelle :

La fréquence des allèles dans une population est modifiée en fonction d'un paramètre environnemental qui exerce une pression de sélection. Donc il y a l'existence d'une variation phénotypique au sein de la population, d'un déterminisme génétique pour cette variation (un phénotype est lié à un génotype) et enfin d'une relation entre la variation phénotypique et la survie et/ou l'aptitude à la reproduction des divers phénotypes.

Exemple : drépanocytose ou le paludisme et l'allèle  $\beta S$  de l'hémoglobine la fréquence de l'allèle  $\beta S$  est très élevée en Afrique de l'Ouest, de l'ordre de 10 à 15 %. Les individus hétérozygotes  $\beta A // \beta S$  (génotype) présentent une résistance au Paludisme, (phénotype) ce qui leur offre un avantage sélectif (L'aptitude).

##### IV.4 Hiérarchie de la sélection naturelle :

Quel est le niveau le plus influent entre ?

- gène → génome entier
- organites/noyau → cellule eucaryote
- cellule → organisme multicellulaire
- individu → groupe

S'il existe un conflit entre deux niveaux, il sera en principe gagné par le niveau le plus bas, **gène l'emportera contre le génome entier** par exemple. L'arrangement en chromosomes impose une réplication simultanée des différents gènes ce qui garantit une uniformité de réplication entre les gènes. S'il y a un désordre dans l'agencement des gènes il n'y aurait pas de génome du tout.



## V. LES ORIGINES DE LA BIODIVERSITE « SPECIATION ».

La spéciation est, en **biologie**, le processus **évolutif** par lequel de nouvelles **espèces** vivantes apparaissent. Une espèce n'apparaît pas instantanément par une mutation conduisant à l'apparition d'un individu d'un type nouveau. La spéciation résulte de la **sélection naturelle** et/ou de la **dérive génétique**, qui sont les deux moteurs de l'évolution.

### V.1 Mécanismes de la spéciation

La disparition d'interfécondité signe l'existence d'espèces qui seront chacune une espèce à part entière.

Cette disparition d'interfécondité peut être due à :

1. *l'isolement écologique* soit par coupure des aires de répartition, soit par fréquentation d'habitats différents dans une même aire de répartition ;
2. *l'isolement éthologique* par non perception des signaux de parade visuels, olfactifs ou auditifs les chants;
3. *l'isolement mécanique et l'isolement gamétique* quand la copulation ne peut avoir lieu ou s'il y a copulation qu'elle ne donne pas de fécondation.

### V.2 Type de spéciation :

L'isolement géographique entre 2 espèces conduit à leur séparation.

- **Spéciation vicariante:** Chaque population évolue indépendamment de l'autre à cause de son isolement géographique (l'apparition d'une rivière...).
- **Spéciation péripatrique:** Par la colonisation d'une île par une partie de la population qui finit de s'isoler de la population générale d'origine.
- **Spéciation parapatrique:** L'isolement n'est pas total, il existe une zone de contact, même il y a une possibilité d'interfécondité, mais avec des hybrides peu adaptés.
- **Spéciation sympatrique :** Des populations non isolées géographiquement peuvent évoluer en espèces distinctes.
- **Spéciation par la distance:** La possibilité de trouver des populations intermédiaires qui sont proches et interfécondes qui séparent 2 populations non fécondes.

# Chapitre IV : génétique et évolution

---

## V.3 La génétique et l'adaptation :

Les adaptations phénotypiques ne sont en général pas transmises aux générations suivantes, ce qui est transmis, sur la base de gènes, c'est la possibilité de réaliser ces adaptations phénotypiques. L'adaptation locale tend à s'améliorer par la sélection naturelle, si les conditions environnementales **ne changeaient pas**. Ce qui signifie, qu'un génotype va se fixer dans une population si ce génotype permet l'apparition d'un phénotype favorable. La sélection s'exerce sur le **phénotype** et non sur le **génotype**. Il n'y a pas de sélection naturelle efficace, car il n'y a pas de conséquences sur la descendance. La sélection naturelle influence globalement l'ensemble des caractères et non caractère par caractère.

## IV.5 Dérive génétique :

La variation aléatoire des fréquences alléliques (allèles neutres) d'une génération à la suivante au sein d'une petite population. Un allèle neutre est un allèle ne conférant au phénotype codé ni avantage, ni désavantage. Au contraire, si l'un des allèles confère un avantage à celui qui le porte alors sa fréquence va augmenter : nous sommes dans le cas de la sélection naturelle.

Dans les conditions naturelles la rencontre des gamètes se fait au hasard, si l'effectif des individus qui se reproduisent entre eux est grand et si la répartition des allèles est homogène, nous sommes donc en condition de Panmixie. Toute situation qui perturbera ces deux dernières conditions modifiera les proportions des gamètes à la génération n+1 sans que la sélection naturelle intervienne : c'est la dérive génétique.

**Exemple :** Dans un groupe d'élèves dans laquelle 8 ont les yeux bleus, 7 filles et 1 garçon. Par manque de place il nous faut renvoyer huit élèves tirés au sort. L'un des tirages au sort pourra nous donner 8 élèves aux yeux bleus. La fréquence de l'allèle récessif yeux bleus est de 100 % dans ce groupe, alors qu'il ne l'était pas dans la classe entière. Si nous avons abandonné ces 8 élèves sur une île déserte quelques dizaine d'années plus tard tous les habitants de l'île auraient eu les yeux bleus.

## V.3 Effet fondateur et génétique :

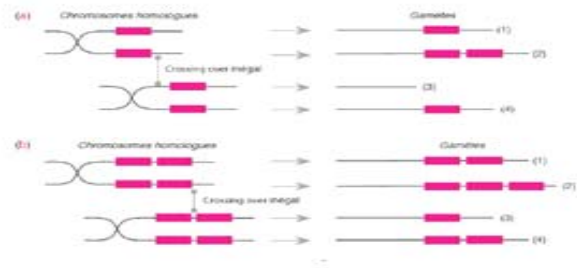
On considère ici la diffusion d'une mutation ou d'un trait génétique qui va ensuite caractériser un groupe d'individus, exemple une maladie génétique. Si la fréquence de la est de 1 sur 100 000 dans la population générale. Cette fréquence sera plus élevée dans un isolat (groupe d'individus apparenté) de migrants. La raison est qu'un individu dans l'isolat de départ avait cette maladie. Son effet est alors plus marqué quelques générations plus tard. Le plus souvent il ne s'agit pas d'une mutation génétique unique, mais d'une divergence progressive de l'ensemble du génome entre deux populations d'origine et celle de l'isolat.

# Chapitre V : Evolution des génomes

## VI. EVOLUTION DES GENOMES :

La plus part des génomes évoluent par addition, duplication ou substitution de bases, comme l'addition, duplication ou substitution de séquences entières. La génétique évolutive prend en compte ces modifications du génome qui peuvent affectées la séquence d'un gène ou d'un chromosome en entier cas de ploïdies.

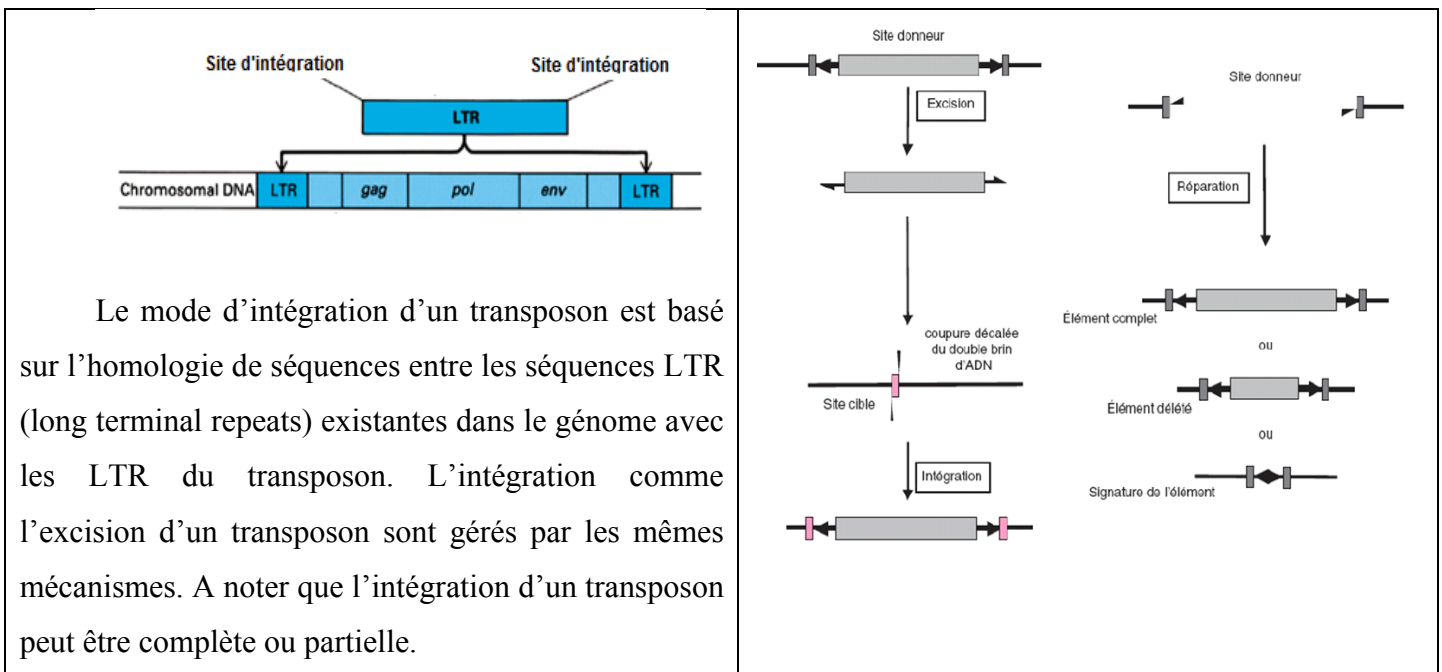
Après duplication, les deux copies du gène sont identiques l'une à l'autre. L'augmentation de la quantité de protéines qui s'en suit peut même être néfaste à la cellule, entrainant la mort cellulaire.



Les 2 gènes sont redondants, donc une seule des deux copies étant suffisante pour remplir la fonction en question, *l'autre copie peut évoluer librement en accumulant des mutations sans être sujet à la sélection.*

### VI.1 Impact des éléments transposables sur les génomes.

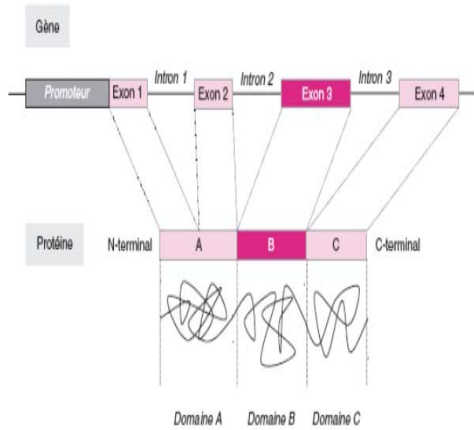
Un élément transposable peut être défini comme "une séquence d'ADN capable de se déplacer de façon autonome et de se multiplier dans le génome".



## VI.2 Les éléments transposables dans les génomes :

Il existe plusieurs copies d'éléments transposons ; de classe I, avec LTR ou non et des transposons de classe 2, dans le génome des eucaryotes ce qui démontre la dynamique de ce phénomène au cours de l'évolution.

## VI.3 Evolution de la structure des gènes par brassage d'exons.



Une protéine est constituée d'un ensemble de régions appelées **domaines**. Un domaine est défini comme une partie de la protéine qui possède une « autonomie » fonctionnelle et structurale.

Par exemple : Le gène de la globine qui est composé de 4 sous unités indépendantes dans leur fonctionnement : 2 sous unités  $\alpha$  et 2 autres sous unités  $\beta$ .

## VI.4 Pseudogènes et rétropseudogènes :

Un **pseudogène** résulte de la duplication d'un gène suivie de l'inactivation d'une des 2 copies, soit par mutation au niveau du promoteur inhibition de l'activité, soit au niveau des exons avec une possibilité d'avoir une protéine non fonctionnelle.

*Tableau ci-contre la distribution des Pseudogènes dans divers organismes.*

Organisme	Taille du génome haploïde (x 10 <sup>9</sup> pb)	Nombre de gènes	Nombre de pseudogènes	Dont pseudogènes dupliqués	Dont rétro-pseudogènes
<i>Escherichia coli</i> K12 (bactérie)	4,6	1 100	95	95	0
<i>Saccharomyces cerevisiae</i> (levure)	12,1	6 340	241	241	0
<i>Caenorabditis elegans</i> (nématode)	102,9	20 009	2 168	1 960	208
<i>Drosophila melanogater</i> (drosophile)	128,3	14 332	110	76	34
<i>Arabidopsis thaliana</i> (arabette)	115,4	25 464	> 700	?	?
<i>Mus musculus</i> (souris)	2 581	25 000	9 173	4 326	4 847
<i>Homo sapiens</i> (homme)	3 040	22 000-39 000	19 293	9 546	9 747

**Le Rétropseudogène** : est un ARNm qui a été rétrotranscrit en ADN complémentaire (ADNc), par la suite il s'est intégré au génome. Il est composé que d'exons et il n'a pas de promoteur, présence de séquences répétées directes et il a une séquence polyAAAA.

**Exemple** du gène BC1 RNA chez les rongeurs et le gène BC200 RNA chez les primates. L'ARNt-ala a subi une rétrotransposition, passage par un ARN, puis un ADNc, il y a 55 millions d'années. Cette copie s'est intégrée dans le génome et elle s'exprime dans les neurones sous forme d'un ARN non messager, **BC1**. Cet

élément à lui aussi subit une autre rétrotransposition pour donner les éléments ID. Les éléments ID appartiennent aux séquences répétées et non codantes SINES.

Chez les primates le gène BC200 RNA a subi lui aussi le même processus, il donne un ARN non messenger, qui s'exprime dans le neurone et il est à l'origine de séquences répétées de type SINES.

# Chapitre VI : phylogénétique moléculaire

---

## VII. INTRODUCTION A LA PHYLOGENIE MOLECULAIRE

### VII.1 Définition d'un caractère moléculaire :

Par définition, un caractère correspond à toute qualité observable d'un organisme, la couleur des yeux par exemple. Récemment, biochimie et la biologie moléculaire, principalement sous la forme de l'analyse des protéines et des acides nucléiques, constituent une autre source d'information en génétique évolutive. Un marqueur moléculaire correspond donc à un gène, une protéine, les diverses formes d'un même gène dans différentes espèces, les diverses formes de microsatellites, de SNP, la carte d'un génome....

Il est donc possible de comparer les séquences chez plusieurs êtres vivants et de quantifier leur ressemblance par un simple pourcentage que l'on assimile à la distance génétique entre les deux taxons auxquels appartiennent les deux êtres vivants. Ces marqueurs moléculaires se caractérisent par divers aspects principalement:

Polymorphe	- Variable entre individus
Discriminant	- Différencie des individus très proches
Multiallélique	- Possède plusieurs allèles à un même locus
Codominant	- L'hétérozygote présente les caractéristiques des 2 homozygotes
Non épistatique	- La lecture d'un marqueur est indépendante des autres marqueurs
Indépendant du milieu	- Le génotype du marqueur est déterminé à partir de son phénotype uniquement

### VII.2 La phylogénétique

« *En remontant le temps*, nous pouvons espérer *relier les 2 espèces* qui dérivent de *la même espèce ancestrale* ; en remontant plus loin encore, nous rencontrons l'espèce *ancestrale* des 3 espèces ».

#### VII.2.1 La classification phylogénétique :

Système de **classification** des êtres vivants qui a pour objectif de rendre compte *des degrés de parenté* entre les espèces et qui permet donc de comprendre leur histoire évolutive (ou phylogénie). Elle ne reconnaît pas certains groupes comme les **reptiles** ou les **poissons**, contrairement à la **classification classique**. Cette dernière classification, qui se base sur les ressemblances les plus visibles entre les espèces, est facilement utilisable par le grand public, mais elle ne reflète pas correctement les proximités évolutives entre espèces. La classification phylogénétique a remplacé la **classification traditionnelle** dans la plupart des milieux

scientifiques. L'analyse cladistique qui sert de base à l'établissement de cette classification considère les caractères à toutes les échelles à valeur égale : les caractères macroscopiques et microscopiques issus de l'anatomie comparée et de l'embryologie, les caractères moléculaires issus de la biochimie et de la biologie moléculaire, ainsi que les données apportées par la paléontologie.

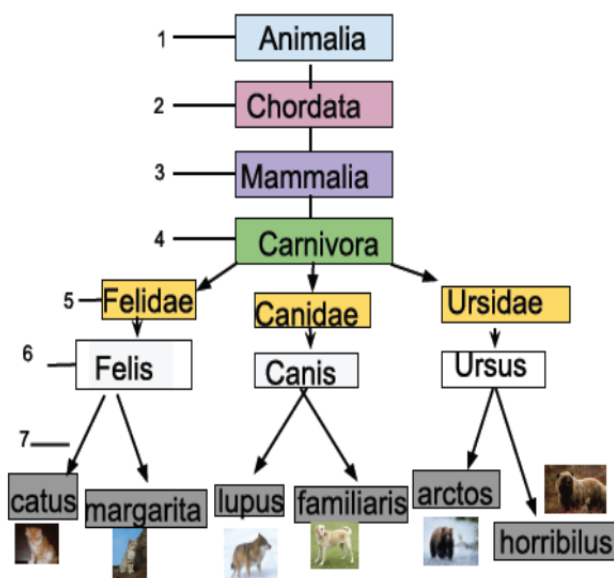
### VII.2.2 Phylogénie :

La phylogénie est le cours historique de la descendance des êtres organisés. La phylogénie permet d'étudier les espèces végétales, animales et microbiennes, sur les deux plans **phénotypique** et **génotypique**, afin de les classer en fonction *de leurs ressemblances* et *en fonction de leurs structures géniques* (liens de parenté). La phylogénie étudie, en fait, *les relations de parenté entre les individus* et représente sous forme d'arbre le résultat de ces relations.

### VII.2.3 éléments de la construction phylogénétique :

L'unité de base de la construction phylogénétique est très souvent l'espèce, puisqu'elle est un groupe génétique fermé : l'interfertilité existe uniquement entre ses membres.

- **Taxon** : est une entité conceptuelle qui est censée regrouper tous les organismes vivants possédant en commun certains caractères. Donc un taxon est un ensemble d'individus réels, et non un regroupement d'entités purement théoriques.
- Exemple : le taxon (la sous-espèce) désigné(e) sous le nom *Canis lupus familiaris* englobe la totalité des chiens domestiques. Dans le contexte phylogénétique, ces derniers, pour lesquels on dispose de données observées, sont les extrémités ou taxons terminaux, ou unités évolutives.

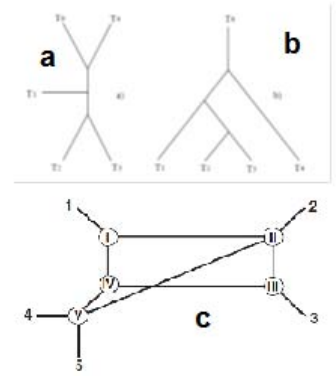


- **Unité Evolutive** (UE) ou **Taxon terminal** : taxon dont les données sont observées *a priori*,
- **Unité Evolutive Hypothétique** (UEH) : résultant *a posteriori* du processus de reconstruction,
- **Branche** : relation entre deux taxons (UEH et/ou UE),
- **Arbre non enraciné** : graphe connexe non cyclique. Il n'y a qu'un seul et unique chemin pour passer d'un sommet à l'autre.
- **Arbre enraciné** : possède une contrainte supplémentaire par rapport au précédent. Présence de liens orientés depuis une origine ou un ancêtre.

- Figure ci-contre : L'arbre (a) est non enraciné tandis que l'arbre (b) possède une racine  $T_0$ .

- **Arbre en Réseau** : arbre non raciné dans lequel on aura différents chemins évolutifs. C'est une phylogénèse dans laquelle il y a encore des inconnus (C).

- **Dendrogramme** : Le dendrogramme est une figure arborescente. C'est un arbre exprimant les liens entre taxons sous la forme d'une succession de branchements, est une terminologie globale qui s'applique à tous les types de graphiques arbores, c'est-à-dire les cladogrammes, les phénogrammes, les phylogrammes.

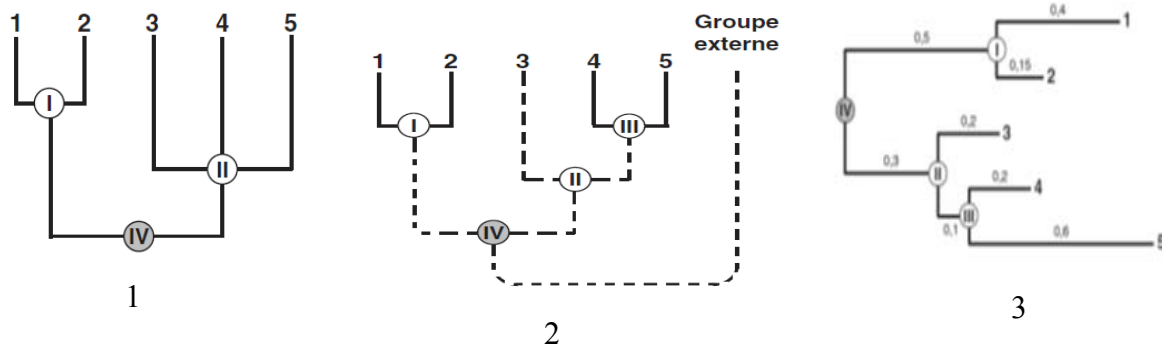


Si, dans sa construction, l'on introduit l'hypothèse que les ressemblances sont le reflet d'une relation de parenté, le dendrogramme est *généalogique* ; si l'on introduit celle que les ressemblances évoluent au cours du temps, le dendrogramme est *phylogénétique*.

- **Cladogramme** : dendrogramme exprimant les relations phylogénétiques (de parenté) entre taxons et construit à partir d'une analyse cladistique.

- **Phénogramme** : dendrogramme produit par la taxinomie numérique où les relations entre taxons expriment les degrés de similitude globale.

- **Phylogramme** : dendrogramme exprimant les branchements cladistiques et le degré de divergence adaptative subséquente aux branchements.

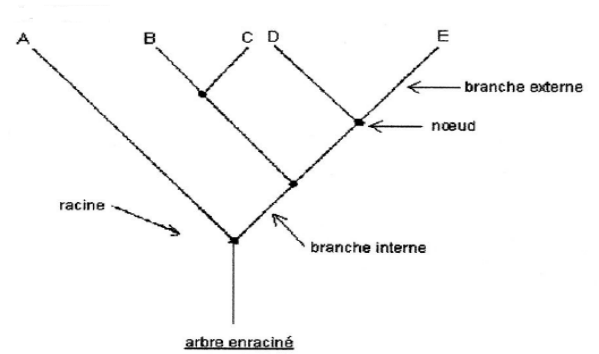


Représentations graphiques d'un cladogramme (1), d'un phénogramme (2) et Phylogramme (3)

**Les branches ou liens** : La relation entre deux sommets constitue un lien appelé souvent segment ou branche. Deux nœuds internes (I) ou (II) sont reliés par un lien interne (en gras), tandis qu'un nœud et une feuille sont reliés par un lien externe, désigné généralement comme une branche terminale ou périphérique (en pointillée). Ce sont les branches qui indiquent le degré de parenté des différents taxons. Il n'y a qu'un seul et unique chemin pour passer d'un sommet à l'autre. La longueur des branches est proportionnelle au temps ou bien aux différences de caractères entre taxons.

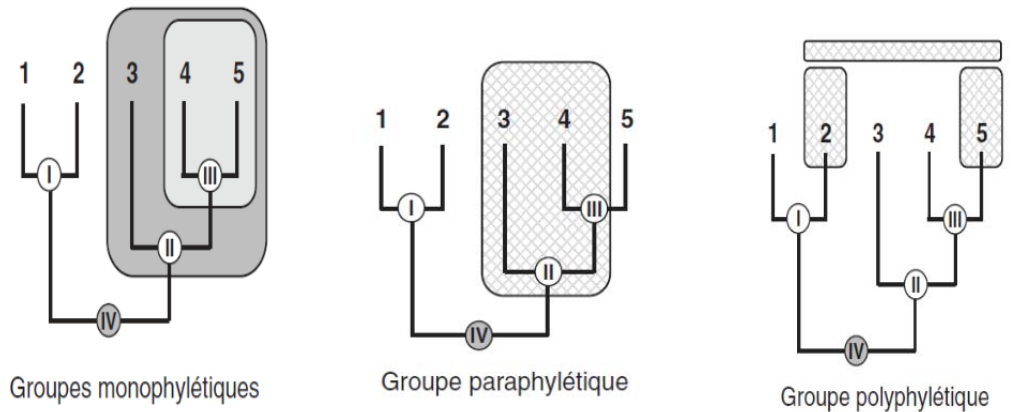


- **Le nœud** : Les arbres sont composés de deux régions : les nœuds où sont placés les taxons qui sont souvent des UEH et les branches qui indiquent le degré de parenté des différents taxons. La longueur des branches est proportionnelle au temps ou bien aux différences entre taxons.



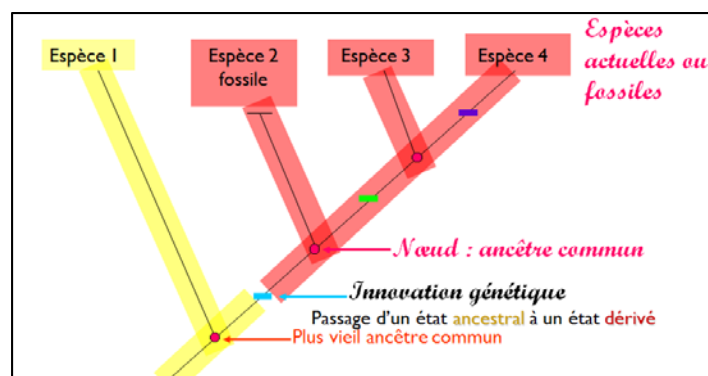
En phylogénétique moléculaire et pour éviter l'emploi du mot « espèce » en phylogénétique on parle d'Unités Evolutive (UE) ou Unités évolutives hypothétiques (UEH).

Un CLADE= Construction de groupes **monophylétiques** qui incluent un ancêtre commun et l'ensemble de sa descendance. Les taxons 4 + 5 descendent du nœud III et constituent un groupe monophylétique.



Les taxons 3 + 4 + 5 descendent du nœud II et constituent un autre groupe monophylétique. Alors que 3 et 4 forment un groupe paraphylétique. Un groupe polyphylétique, ou encore groupe artificiel, contient les descendants de plusieurs ancêtres distincts. Les taxons 2 + 5 constituent un groupe polyphylétique : le taxon 2 descend du nœud I tandis que le taxon 5 descend du nœud III.

- **Cladogramme** : diagramme en branches présentant les liens de parentés entre différents groupes basés sur la distribution des caractères dits « dérivés ». **Caractères dérivés** ou état dérivé ou évolué état apparu plus récemment par modification de l'état primitif. L'innovation génétique sépare donc 2 états passage d'un état ancestral à un état dérivé.



### VIII. CARACTERES ET ETATS DE CARACTERES

Un état de caractère est donc un attribut observable et mesurable sur un individu et qui peut être d'une quelconque nature (moléculaire, physiologique, ...) :

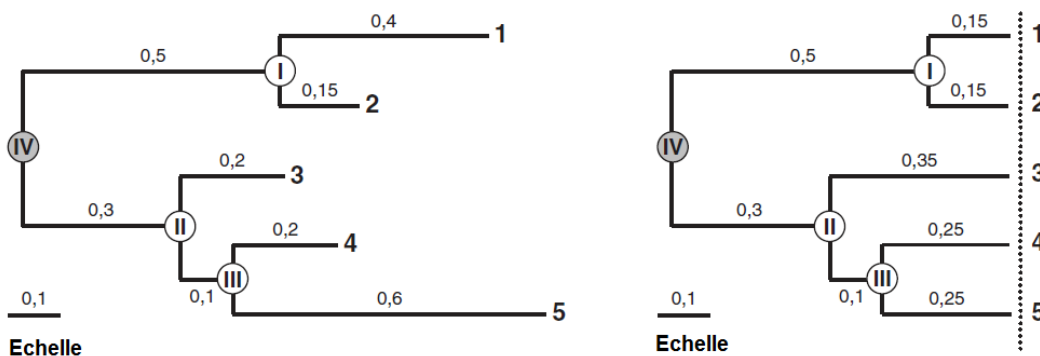
Caractères	Etats du caractère				
	Croissance à 5°C		Croissance à pH acide		
physiologique	1	0	1	0	
Couleur de la colonie	Blanche	Crème	Jaune	Orange	
Motif ATATA	1		0		
Ag O	1		0		
Nageoires	Oui		Non		
Séquence gène	A	C	G	T	Gap
Séquence protéine	20 Acides aminés ou Gap				

La phylogénie se base sur le principe de la comparaison de caractères spécifiques pour un ensemble d'individus. Ces caractères sont en général homologues et appartiennent à des organismes contemporains. Sauf que leur comparaison, par le biais des méthodes phylogénétiques, va permettre de postuler des hypothèses quant à l'éventuelle histoire commune ou non entre ces individus du point de vue moléculaire et phénotypique. Il est donc important de préciser la nature du caractère, en réalité les méthodes phénétiques ne s'intéressent pas au concept de caractère et d'état de caractère pour ne plus s'intéresser qu'à celui d'UE. Ceci est le résultat de la transformation de la matrice de caractères en matrice de distances.

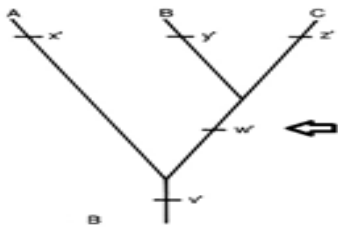
Autrement dit, le phénogramme n'est pas un cladogramme puisque les nœuds n'y représentent pas les états ancestraux des caractères mais seulement les degrés de similitude entre les UE qui en dérivent. De plus tous les caractères utilisés dans la constitution de la distance sont généralement considérés comme ayant le même poids, c'est-à-dire qu'ils participent de manière égale à la ressemblance globale.

Exemple 1 : sur le phénogramme de gauche la distance génétique entre les taxons 1 et 2 par rapport au nœud est différente : 0.4 pour le taxon 1 et 0.15 pour le taxon 2. Donc les longueurs des branches est différente.

Alors que sur celui de droite les distances sont égales : 0.15 nœud-taxon 1 et de 0.15 nœud-taxon 2.



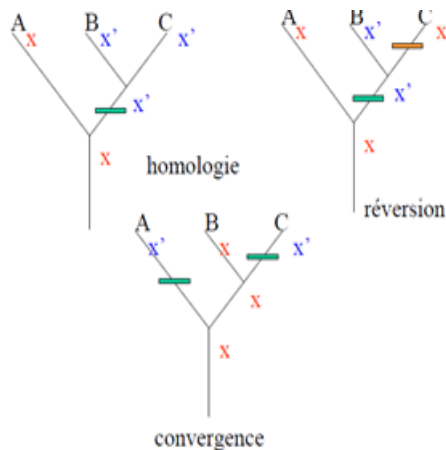
Exemple 2 : un cladogramme rapporte les relations phylogénétiques entre taxons et construit à partir de l'analyse cladistique où les points de branchements (les nœuds) sont définis par des synapomorphies, c'est-à-dire que les parentés entre les taxons étudiés sont fondées sur les seuls états dérivés partagés, hérités d'un ancêtre commun.



Pour faire simple : dans un cladogramme ce qui est important c'est les caractères commun à B et C hérités de l'ancêtre commun ⇔

### IX. LES METHODES PHENETIQUES :

**L'homoplasie** est divisée en **convergence** (apparition indépendante d'un caractère chez deux espèces) et **réversion** (apparition d'un caractère **ayant l'apparence de** la morphologie ancestrale). L'Homoplasie est la similitude non héritée et la convergences lorsque ces similitudes sont acquises indépendamment à partir de morphologies différentes.



Exemple : similitude ou ressemblance due à des ancêtres communs: baleines et mammifères.

Alors que la convergence, est la ressemblance des baleines et des poissons.

Une similitude peut représenter une homologie ou une convergence. L'**homologie** est une similitude héritée d'un ancêtre commun, contrairement à l'**homoplasie**.

La réversion est un retour apparent à l'état initial et la convergence c'est acquérir en parallèle un même caractère.

#### IX.1 Cladisme :

Le cladisme qui regroupe dans un ; même ensemble le plus proche ancêtre commun et tous ses descendants. La phénétique numérique objective qui se donne pour but de quantifier le degré de similitude entre taxa. Donc, Le cladisme est une méthode de classification des individus et des taxons en fonction de leur degré de parenté.

**Attention** : Le cladisme (ou la cladistique) est la systématique phylogénétique, le cladogramme est un « schéma d'argumentation phylogénétique », le clade est un « groupe monophylétique ».

## IX.2. Approches cladistiques :

Les principes de l'analyse cladistique ont été énoncés par Hennig (1960), sous l'expression « systématique phylogénétique ». La phylogénie est reconstruite à partir d'une analyse de caractères qui vise à identifier les états plésiomorphe (primitif) et apomorphe (dérivé). Les caractères primitifs partagés sont les symplesiomorphies. Les caractères dérivés présents chez un seul taxon sont les autapomorphies.

Les parentés entre les taxons étudiés sont fondées sur les seuls états dérivés partagés (synapomorphies), hérités d'un ancêtre commun. Les groupes définis par des synapomorphies, ou clades, sont monophylétiques (renfermant la totalité de la descendance à partir d'un ancêtre commun exclusif), paraphylétiques (ne renfermant pas la totalité de la descendance d'un ancêtre commun exclusif) ou polyphylétiques (ne renfermant pas d'ancêtre commun exclusif).

## IX.3 Distance évolutive degré de dissemblance

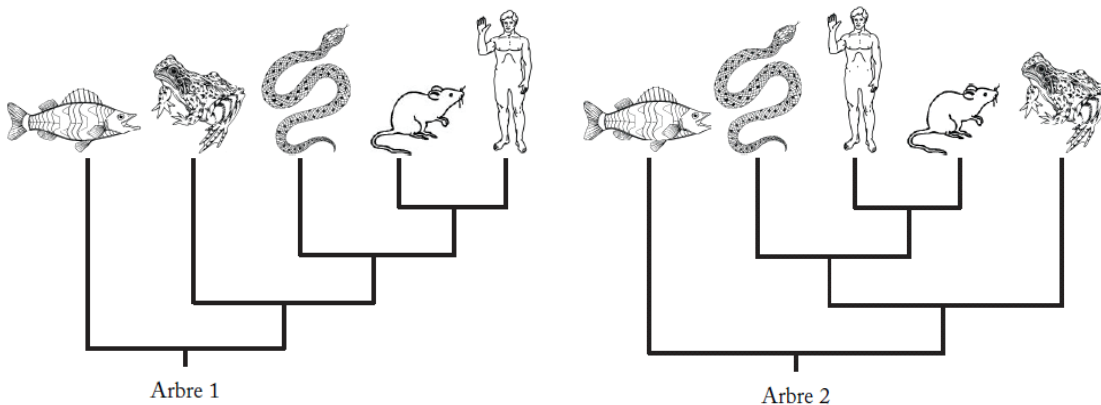
Le concept de base des méthodes phénétiques est celui de similitude globale : plus la ressemblance entre deux UE est importante, plus la parenté entre elles a des chances d'être proche.

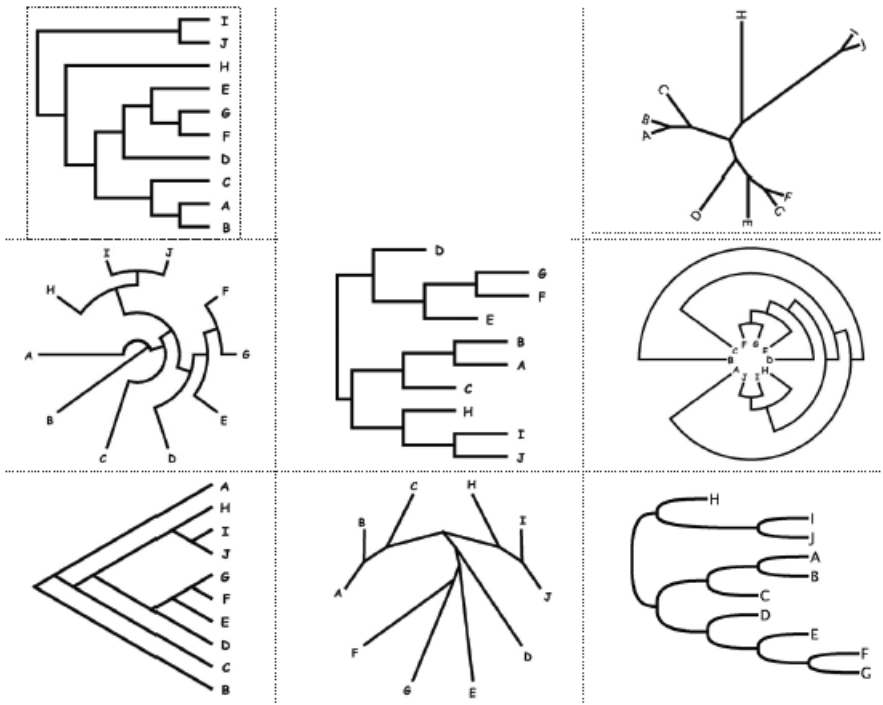
La ressemblance s'établit à partir d'informations biologiques de nature très variée :

- **alternative** : présence ou absence de particularités morphologiques ou génétiques ;
- **qualitative** : séquences d'acides aminés ou d'acides nucléiques, états multiples de caractères morphologiques ; il est parfois possible de transformer ce type de données qualitatives en données alternatives ;
- **quantitative** : fréquences géniques, génotypiques ou phénotypiques, mesures morphométriques, etc.

## X. DIFFERENTES REPRESENTATIONS GRAPHIQUES DES ARBRES

Les arbres phylogénétiques sont construits autour d'un axe qui permet leur articulation sans pour autant changer les liens de parentés.





Les différentes représentations de ces arbres phylogénétiques possèdent la même topologie et leur dessin retrace les liens des relations de parenté, il est donc composé d'une succession de branchements, pour cette raison que les relations entre les UE restent identiques, même si la forme de l'arbre importe peu. Avec l'arbre phylogénétique, l'itinéraire nous est donné par le point de départ obligé : **la racine de l'arbre**.

### X.1 La parcimonie et le Maximum de Parcimonie

Les procédures de parcimonie : à la recherche de l'arbre le plus court. Pour déterminer l'arbre le plus parcimonieux, c'est-à-dire celui qui offre le moins de pas évolutifs (mutations) par rapport à la séquence ancestrale théorique. La phylogénie des taxons ne se conçoit qu'au travers des phylogénies de caractères. Plus les caractères sont nombreux, plus le résultat sera informatif. La superposition de tous les cladogrammes de caractères permet de définir le cladogramme de congruence optimale, c'est-à-dire l'arbre le plus court.

Il existe plusieurs modèles de parcimonie.

La parcimonie de Wagner (Kluge et Farris, 1969 ; Farris, 1970) autorise *a priori* convergences et réversions ( $0 \rightarrow 1$  et  $1 \rightarrow 0$ ). Aucune contrainte n'est imposée au mode de transformation des caractères.

La parcimonie de Camin-Sokal (Camin et Sokal, 1965) n'autorise que les convergences.

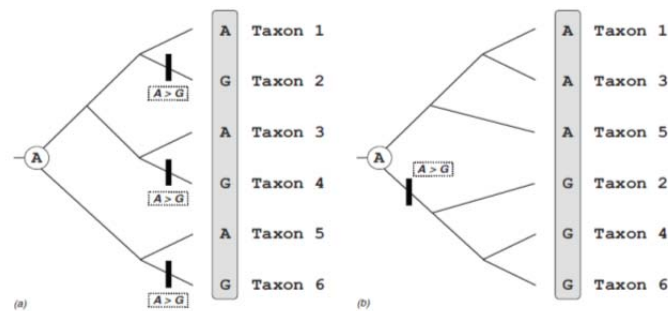
La parcimonie de Dollo (Le Quesne, 1972) n'accepte que les réversions.

Ceci montre la difficulté par l'absence de consensus dans la manière de calculer les chemins entre les UE. Le mode de parcimonie de Wagner, qui est le moins restrictif, est également le plus en adéquation avec le principe

même de parcimonie. Heureusement qu'il existe plusieurs logiciels qui peuvent faire ces calculs en fonction de l'hypothèse de départ. Nous citons ici quelques-uns.

Algorithme exact : garantit la découverte de l'arbre le plus court. Temps de calcul augmentant exponentiellement avec le nombre de taxons. Technique du *branch and bound* (Hendy et Penny, 1982) : algorithme exact qui ne nécessite pas une recherche exhaustive. Un arbre est pris comme référence (longueur L). On change de topologie. Dès qu'un arbre est aussi court que L ou plus court, il devient la nouvelle référence, etc.

Algorithme heuristique : Temps de calcul court, mais l'arbre le plus court n'est pas toujours découvert. Technique d'addition pas-à-pas (*stepwise addition*) sensible à l'ordre d'introduction des taxons ; technique de « réarrangement des branches » (*branch swapping*) : recherche d'un réarrangement plus court, qui devient la référence, etc. Ne donnent pas forcément **tous** les arbres de longueur minimale.



A : Un premier exemple de relations phylogénétiques entre 6 taxons. Les nucléotides A et G rencontrés à un même site dans leur génome sont indiqués dans le rectangle grise. Le cercle indique que le nucléotide ancestral supposé est A. Trois changements de A vers G sont indiqués pour expliquer le profil observé pour ce site.

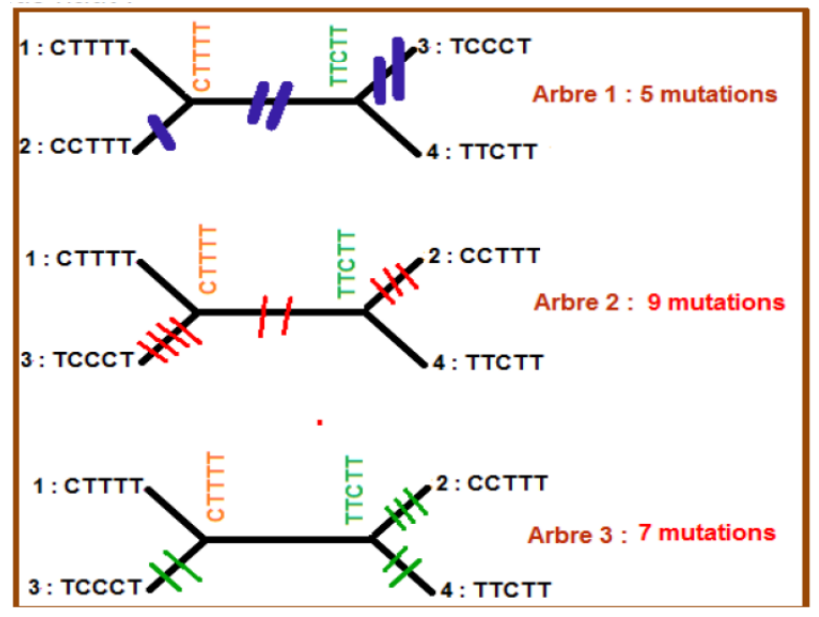
B : Un second exemple de relations phylogénétiques entre ces 6 taxons. Le changement de relations de parenté implique qu'un unique changement de A vers G suffit à expliquer le profil observé pour le site. D'autres scénarios évolutifs sont bien entendu possibles, mais celui-ci est le plus économe en termes de changement évolutif.

## X.2 Le maximum de parcimonie

Le principe fondamental des méthodes de parcimonie consiste à diminuer au maximum le nombre d'évènements génétiques (mutations, substitutions) pour comparer deux séquences ou deux individus. L'arbre qui en découle compte le moins de pas évolutifs, c'est-à-dire le moins de mutations possibles ; c'est l'arbre optimal.

Dans l'exemple suivant, nous allons rechercher l'arbre le plus parcimonieux parmi un ensemble d'arbres. Pour cela, on suppose que les quatre individus possèdent les séquences suivantes :

Ind1	C	C	T	C	C
Ind2	C	C	C	T	C
Ind3	C	C	C	C	T
Ind4	C	C	T	T	T



Le nombre d'arbres non racinés théorique entre ces quatre individus est égal à trois selon la formule donnée plus haut :

Pour déterminer l'arbre le plus parcimonieux, c'est-à-dire celui qui offre le moins de pas évolutifs (mutations) par rapport à la séquence ancestrale théorique.

Dans notre cas, nous allons supposer les deux séquences CCTCC et CCTTT comme référence ancestrale avec lesquelles nous allons calculer le nombre de mutations par comparaison aux quatre individus :

Dans le cas de l'arbre 1, le nombre de mutations entre l'individu 2 et la séquence repère est égal à 1. Entre les deux séquences repères CTTTT et TTCTT, il y a deux mutations. Entre la séquence repère TTCTT et l'individu 3, il y a deux mutations.

Au total, il y a :  $1 + 2 + 2 = 5$  mutations.

En refaisant les calculs pour les cas des arbres 2 et 3, le nombre de mutations est neuf pour l'arbre 2 et sept pour l'arbre 3. Donc l'arbre le plus parcimonieux, parmi les trois, est donc l'arbre 1. Il considère qu'il y a eu moins de mutations entre les individus ayant contribué à la topologie de cet arbre.

### X.3 Principes de calculs

Caractères binaires :  $0 - 1$  ;  $a - b$ ,

Caractères à états multiples (présentent déjà une hypothèse phylogénétique) :  $0 - 1 - 2$  ;  $a - b - c$ ,

- Relations non ordonnées (non additives) : chaque transformation compte pour un pas (A, G, C, T),
- Relations ordonnées (additives) : caractères exprimés sous forme de séries (linéaire ou non ; orientée ou non).

On peut transformer une série de transformations en plusieurs caractères binaires (factorisation).

Polymorphisme : les états plésiomorphe et apomorphe du caractère sont présents dans un même taxon (plusieurs traitements possibles : états multiples ; fréquences).

Optimisation : procédé permettant de traiter des données manquantes (?). L'état de caractère est alors inféré grâce à l'environnement du taxon considéré dans l'arbre (taxons voisins).

Pondération : peut être utilisée pour les caractères et pour les transformations.

#### **X.4 La mesure de l'homoplasie**

L'indice de cohérence, IC (*consistency index* ; Kluge et Farris, 1969) est égal au rapport entre le nombre de transformations nécessaires pour expliquer les états de tous les caractères et le nombre de transformations effectivement observées ( $I = R/L$ ). IC est inférieur ou égal à 1 ( $\cong 0,2$  en *random*). Il doit être le plus proche possible de 1 (pas d'homoplasie). Problème : il ne permet pas de distinguer autapomorphies et synapomorphies.

L'indice de rétention, IR (*retention index* ; Farris, 1989) représente le rapport entre le nombre d'homoplasies observables et le nombre d'homoplasies observées. Il doit être le plus élevé possible.

#### **X.5 Les arbres de consensus**

L'arbre de consensus strict (*strict consensus tree* ; Sokal et Rohlf, 1962, 1981) est construit en ne retenant des arbres comparés (et de longueur minimale) que les dichotomies présentes dans tous les arbres. Les ambiguïtés sont représentées par des polytomies/multifurcations.

L'arbre de consensus majoritaire (*majority rule consensus tree*) est établi en conservant les clades les plus fréquemment rencontrés dans les topologies parcimonieuses alternatives.

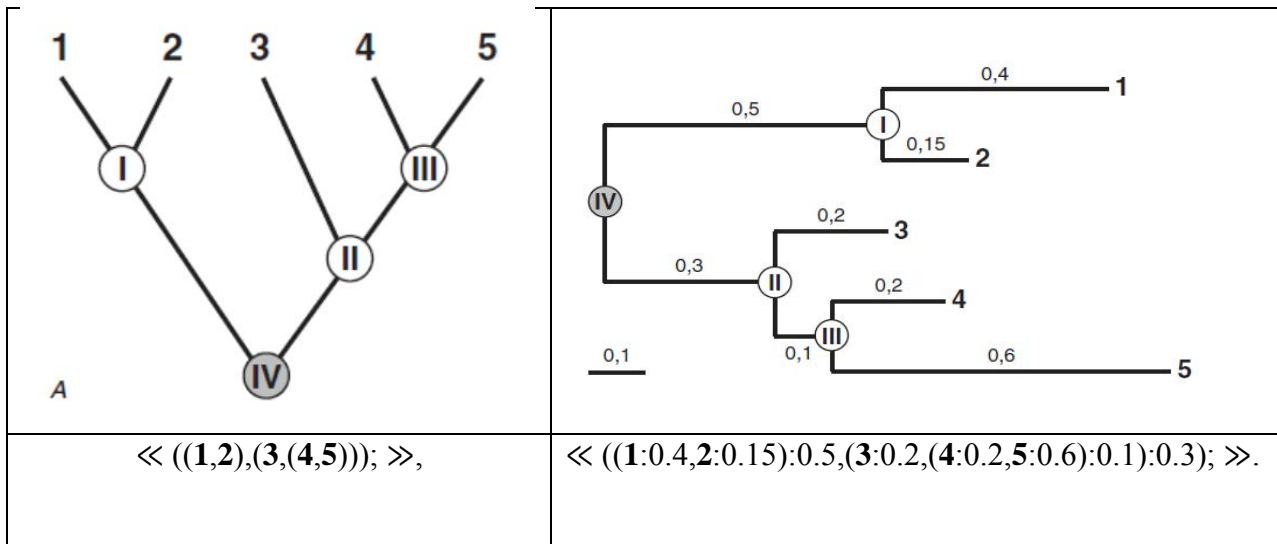
- Le Bootstrap (Efron, 1979 ; Felsenstein, 1985) consiste à tirer au hasard **avec remise** un ensemble de K caractères parmi les K caractères d'une analyse. La nouvelle matrice contient K caractères, parmi lesquels certains caractères seront présents plusieurs fois et d'autres jamais. C'est une pondération aléatoire. Cette matrice fait l'objet d'une analyse phylogénétique. La procédure peut être répétée N fois. Il y aura de 1 à N arbres différents. Il suffit ensuite, pour chaque nœud présent, dans l'arbre de consensus strict de définir la fréquence (%) de présence.

- Le Bremer Support est un indice (dit de Bremer / « *decay indices* ») qui permet de contrôler la robustesse d'un nœud. Pour chaque dichotomie de l'arbre le plus parcimonieux, il correspond au nombre de pas supplémentaires nécessaires à la disparition de cette dichotomie. Plus l'indice est élevé ( $n \geq 1$ ), plus le nœud est robuste.

#### **X.6 Le format NEWICK ou Les arbres en notation parenthésée :**

Dans certaines situations, par exemple pour communiquer avec un ordinateur, il peut être utile de disposer d'une représentation non graphique des arbres. Le format dit « Newick 8:45 » permet par un jeu de parenthèses de s'affranchir du graphisme tout en conservant les différentes informations.





Il s'agira d'encadrer par une parenthèse ouvrante et une parenthèse fermante l'ensemble des taxons descendant d'un même nœud. Ces taxons (en gras ci-dessous) seront séparés par une virgule et la ligne sera terminée par un point-virgule. Par suite, l'arbre deviendra  $\ll ((1,2),(3,(4,5))); \gg$ . Il est également possible de rendre compte du degré de divergence évolutive entre les taxons en indiquant la longueur du lien après chaque taxon et chaque nœud sépare par deux points. Attention, suivant la notation anglo-saxonne et pour éviter les confusions, les décimales sont séparées par des points et non par des virgules. Le Phylogramme de la deviendra donc :  $\ll ((1:0.4,2:0.15):0.5,(3:0.2,(4:0.2,5:0.6):0.1):0.3); \gg$ .

### X.6 La méthode UPGMA (Unweight Pair Group Method with Arithmetic mean).

Cette méthode est utilisée pour construire des arbres phylogénétiques si les séquences ne sont pas trop divergentes. Cela veut dire que cette méthode repose sur la théorie de l'horloge moléculaire dans laquelle on suppose que tous les individus issus de parents communs subissent le même taux de mutation (substitutions, délétions et insertions) au cours de leur évolution. Par exemple, des protéines ayant la même fonction auraient subi les mêmes mutations et avec les mêmes taux. Elle utilise un algorithme de clustérisation séquentielle car elle construit l'arbre pas à pas, au fur et à mesure que les clades sont définis. Un clade est une paire d'individus très similaires regroupés ensemble. Il y a identification du premier couple d'individus (1er clade) les plus proches qui seront considéré comme un seul individu pour la suite de la clustérisation.

**Comprendre UPGMA par l'exemple :** Le tableau suivant résume une portion d'un alignement multiple entre cinq séquences.

Références bibliographiques :

- **Livres :**

- C.W. FOX and J.B. WOLF: Evolutionary genetic concepts and cases studies oxford university press (2006).
- E.J.P. DOUZERY et al., Phylogénétique moléculaire (2006)
- Pierre Darlu et Pascal Tassy : *La Reconstruction phylogénétique. Concepts et Méthodes. Edition électronique (Masson Avril 2004).*
- M. A. HAMIDECHI et A. DJEKOUN Cours de phylogénie moléculaire Distances et constructions phylogénétiques. 2011.

Articles scientifiques :

- A. CRISCUOLO, et al. A Fast Distance-Based Approach for (Super)Tree Building in Phylogenomics - 2006). *Syst. Biol.* 55(5):740-755.

Sites Web :

Stanford Encyclopedia of philosophy ; <http://plato.stanford.edu/> Evolutionary Genetics.