Introduction :

Historique.

De tout temps on a classé les objets de la biologie en fonction de leur ressemblance. La première classification connue vient d' Egypte 1600 ans avant Jésus Christ, elle classait les plantes selon leurs propriétés médicinales.

On s'est longtemps attaché à dégager l'essence des organismes observés donc à décrire les caractères communs a divers spécimens.

Dès l'antiquité avec Platon 428-348 BC, Aristote (384-322 BC) vers 325 avant Jésus Christ dans son traité des plantes classait les arbres, les arbustes selon leur taille.

A la renaissance, le but de ces classifications était l'identification des échantillons. Césalpin (1519-1603) dans De Plantis paru en 1583 s'inspire toujours d'Aristote. Sans système de fond, il regroupe les échantillons selon le type de croissance, la présence ou non d'épines, l'origine cultivée ou spontanée. Il produit une classification descendante par dichotomies successives en partant d'associations plus ou moins naturelles et tout à fait subjectives. Tournefort(1656-1708) introduit la notion de genre et Linnée (1707-1778) invente le binominalisme et produit une classification toujours descendante basée sur des caractères qui pour l'essentiel concernent la sexualité des espèces. Tous ces auteurs (et leurs contemporains) sont des essentialistes. L'apport de Linnée , essentialiste strict qui admet la continuité entre les espèces, est l'ordre et la simplification de la taxinomie (synonymes, diagnoses rigoureuses, etc.). Au sein de chaque règne il a introduit un système hiérarchique à 4 niveaux : classe, ordre, genre, espèce. Cette hiérarchie ne représentant qu'une commodité de classement.

Parallèlement dans cette même période on voit apparaître des méthodes un peu différentes. Un des premiers, Magnol (1638-1715) recommande dans Prodomus (1689) la combinaison de caractères qui conduit à une classification ascendante. Vers la même époque Adanson (1727-1806) publie Familles des plantes (1763) où les ordres linnéens deviennent les familles adansoniennes. Dans sa méthode, il affirme qu'il ne faut pas s'attacher à certains caractères naturels pour classer les plantes mais à tous les caractères d'une plante. Il considère que tous les caractères ne sont pas au même niveau et calcule des distances entre les différentes familles. Il est à la base de la taxonomie botanique moderne.

Antoine Laurent de Jussieu (1748-1836) reprendra la méthode d' Adanson en étudiant les affinités entre les végétaux. Il formalisera le principe de la subordination des caractères. Il publie en 1789 son Genera plantarum qui servira à la nomenclature des familles.

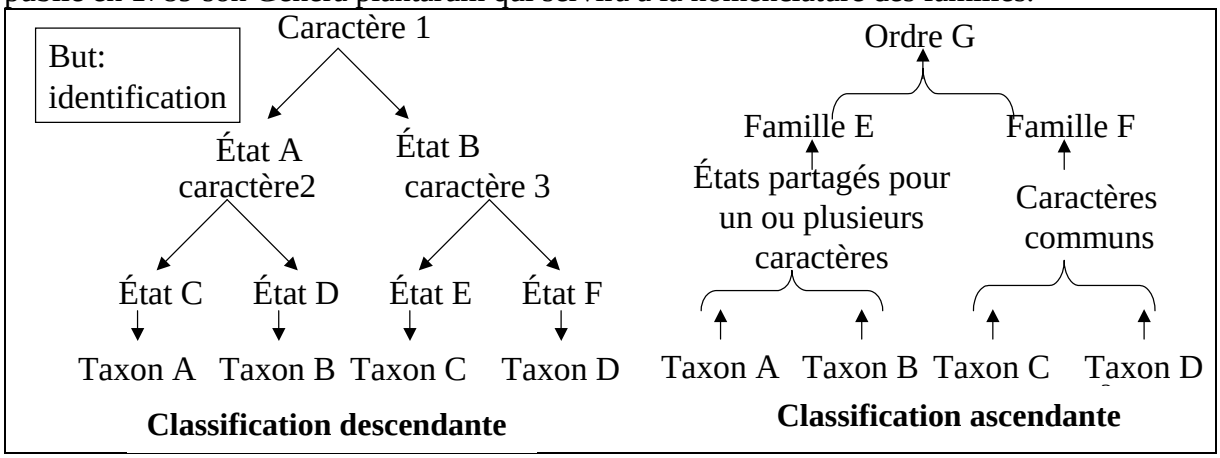


Figure I- 1. Les deux méthodes de classification.

Dans le monde animal l'évolution des idées est comparable. Buffon (1707-1788) s'attache à décrire tous les détails des échantillons qu'il a entre les mains au lieu d'en tirer l'essentiel. Ce qui en fait un nominaliste. En 1755 il reconnaît des espèces plus ou moins apparentées.

A cette époque se développe l'anatomie comparée avec Cuvier (1769-1832) et Geoffroy Saint Hilaire (1772-1844).Cuvier reprend le concept de la subordination[1] des caractères. Bien que résolument fixiste c'est le fondateur de la paléontologie. Au contraire, les travaux de Geoffroy Saint Hilaire portant essentiellement sur l'anatomie comparée, sur les embryons, les fossiles, lui permettent de trouver des liens entre les espèces anciennes ou contemporaines (d'ou leur querelle de 1830). Geoffroy Saint Hilaire (1772-1844) fonde la notion d'analogie (devenue homologie). Il est ainsi précurseur de l'évolutionnisme. Lamarck (1744-1829) croit à l'importance primordiale de l'adaptation au milieu dans l'évolution du vivant.
En 1859 la publication de Darwin (1809-1882) rend compte des affinités et restaure l'idée de continuité illustrée par Haeckel (son arbre où l'on passe d'une espèce à une autre).

Vers 1820-40, zoologistes et botanistes reconnaissent 2 types de similitudes qui posent les bases de la classification ?
affinité (=ressemblance due à des ancêtres communs, ex : baleines et mammifères)
analogie (= convergence, ex : ressemblance des baleines et des poissons)

Les classifications sont fondées sur des relations d'apparentement : la théorie de la récapitulation selon laquelle l'ontogenèse récapitule les stades adultes ancestraux
et en 1836, le triple parallélisme d'Agassiz[2] (1807-1873) fixiste convaincu.
histoire des fossiles
développement embryonnaire
rang dans la classification
A l'aube du XXe siècle la naissance de la génétique va faire évoluer cette situation. En 1883-1884 la publication d'August Weismann (1834-1914) sur la théorie de la lignée germinale et de la lignée somatique est la négation du lamarckisme. Vers 1900, les travaux de Mendel sont redécouverts et favorisent l'hypothèse saltatoire opposée au gradualisme. Deux courants évolutionnistes vont dès lors s'opposer.
Ceux qui recherchent les causes proximales, les facteurs génétiques et leur origine (Bateson 1861-1926, de Vries (1848-1935), Morgan (1866-1945)) s'intéressent aux gènes plus qu'aux taxa.

---

[1] *Subordination des caractères*. Certains caractères semblent dominer les autres, en ce que les seconds ne peuvent exister sans les premiers. On ne peut être mammifère si l'on n'est vertébré. C'est par cette subordination que s'établit la classification.
Les progrès vont être très importants quand, avant la Révolution, **Bernard de Jussieu** (1799-1876) et son neveu Antoine-Laurent (1748-1836) vont définir la notion-clé de subordination des caractères. Ils se rendent compte que, pour déterminer un taxon donné, une classe par exemple, l'idéal est d'avoir un (ou plusieurs) caractère(s) constant(s) à l'intérieur de cette classe et variable(s) dans toutes les autres. Un type de caractère est donc utile à un niveau précis de la classification, certains au niveau de l'ordre, d'autres au niveau du genre... Les caractères sont donc "subordonnés". On peut alors faire un tri parmi les caractères issus de descriptions détaillées, et faire surgir ceux qui sont pertinents d'un point de vue **taxinomique**, en les hiérarchisant. On pensait avoir trouvé la "Méthode Naturelle" menant à la Classification Naturelle. Si les Jussieu l'appliquent en botanique, dès la fin du XVIIIème siècle, **Jean-Baptiste Lamarck** (1744-1829) et **Georges Cuvier** (1769-1832) l'appliquent au monde animal, le premier essentiellement chez les mollusques et les animaux vermiformes, le second chez les vertébrés et pour l'ensemble des animaux.

[2] Néanmoins il ne faut pas se contenter d'une classification ascendante comme on peut en trouver chez Agassiz (1807-1873) qui publia un schéma associant dans un contexte géologique groupes fossiles et actuels de poissons, les groupes décrits ayant tous une dimension stratigraphique. On aurait pu y voir l'application pure et simple du modèle darwinien. Or Agassiz est créationniste. Son schéma suggère des connexions mais celles-ci ne sont pas interprétées comme des preuves de la transformation des espèces au cours du temps. (Tassy : l'arbre à remonter le temps). Bien au contraire, pour lui et les progressionnistes la séquence des fossiles reflète simplement la maturation du plan de la création dans l'esprit du créateur.

Ceux qui recherchent les causes ultimes de la biodiversité : les taxinomistes qui centrent leurs études sur les espèces et les paléontologues et anatomistes comparés qui étudient plutôt l'apparition de taxa supérieurs.

Les premiers, mendéliens, sont d'ardents typologistes pour lesquels tout changement est dû à des mutations :
l'évolution est un processus mutationnel
la sélection est peu active
il est inutile de prendre en compte les variables individuelles

Les seconds, naturalistes font erreur sur la nature de l'hérédité et de la variation. Ils étudient les populations naturelles dans leur diversité. Ils sont dans la lignée de Darwin
Les variations géographiques produisent des gradients dans les populations
le poids de l'hérédité mendélienne est minimisée.

Le conflit ne sera réconcilié (?) que vers 1970 par des généticiens agronomes qui établiront l'existence de plus d'un type de variabilités de natures différentes les grandes et les toutes petites.
Toutes ne sont pas nuisibles , il y en a aussi de bénéfiques.
Le matériel génétique est invariant.
Dans les populations, la principale source de variation est la recombinaison, l'action de facteurs multiples, les interactions épistatiques.
Un seul gène peut affecter plusieurs caractères phénotypiques.
La sélection naturelle est efficace.

Phénétique et cladisme

En 1880 c'est le déclin de la taxinomie et de la phylogénétique, faute de résultats probants. 1901, Fleischman parlant de la théorie de la descendance considère que c'est « un magnifique mythe dépourvu de base effective »[3].

Deux façons de voir les choses jusqu'en 1950

Les données taxonomiques fournissent le matériel pour la phylogénie qui peut à son tour modifier la taxonomie, cette circularisation déniant un caractère scientifique à ces matières. Les données fournies par l'étude des groupes naturels (méthodes ? ou associations subjectives) étaient traitées indépendamment par les taxinomistes et par les phylogénéticiens.

[3] "The Darwinian theory of descent has not a single fact to confirm it in the realm of nature. It is not the result of scientific research, but purely the product of imagination."—*Dr. Fleischman [Erlangen zoologist]*. la théorie de la descendance avec modification de Darwin n'a pas un seul fait qui la confirme dans le royaume de la Nature. Ce n'est pas le résultat d'une démarche scientifique mais un pur produit de l'imagination.
Dr. Albert Fleischmann, of the University of Erlangen,
"I reject evolution because I deem it obsolete; because the knowledge, hard won since 1830, of anatomy, histology, cytology, and embryology, cannot be made to accord with its basic idea. The foundationless, fantastic edifice of the evolution doctrine would long ago have met with its long-deserved fate were it not that the love of fairy tales is so deep-rotted in the hearts of man."[1]
Je rejette le fait évolutif parceque je l'estime dépassé car le savoir qui a progressé rapidement depuis 1830, n'arrive pas à accorder les nouvelles données en anatomie, cytologie, embryologie avec l'idée d'évolution. Le manque de bases, l'aspect fantastique de la doctrine de l'évolution aurait dû depuis longtemps avoir le sort qu'elle mérite, si ce n'était l'amour des contes de fées si profondément ancré dans le coeur des hommes.

Depuis 1950 on assiste à un réveil phylotaxonomiste. En 1957 au congrès « La Systématique aujourd'hui » sur 29 communications 4 seulement étaient consacrées à la macrotaxonomie. On voyait souvent un seul caractère basant les regroupements d'ordre supérieur. La macrotaxinomie était principalement qualitative et subjective.

Exemples

la façon de considérer les apétales parmi les angiospermes, en fonction de l'analogie reconnue entre le strobile des gnétales et la fleur des apétales (à sexes séparés).
Jusqu'en 1915 le strobile des apétales évoque celui des gnétales, tous deuxw sont à sexes séparés et les apétales sont considérées comme primitives (Engler et Prantl). A partir de 1915 une autre interprétation (Bessey) place les Ranales au strobile bisexué en position ancestrale, rejetant les apétales parmi les formes plus dérivées.

la position du ginkgo parmi les gymnospermes (Cycadales + Pinales) en fonction de l'aplatissement de la graine ou de la prise en considération des vaisseaux qui irriguent les enveloppes.

Souvent les ginkgopsides ont été regroupés avec les Coniférales sur la base de leur platyspermie commune alors que les graines de cycadales sont de type radial.

Meyen (1984) reconsidère ce regroupement en prenant en compte la vascularisation et les formes fossiles connues.

La lignée des ginkgoosides , où tégument et nucelle ne sont pas fusionnés présentent deux faisceaux de vaisseaux dans le plan principal de l'ovule. Ils présentent une platyspermie primaire (originelle)

Lyrasperma (Carbonifère inférieur)

Callospermarion, la cupule se referme en un micropyle (Carbonifère supérieur)

D'autre part Hydrasperma du carbonifère inférieur présente plusieurs ovules à symétrie radiale. Une première étape évolutive réduit le nombre de ces ovules (ovule unique) donnant deux formes E (hypothétique) et Lagenostoma (carbonifère supérieur). Chacune de ces deux formes va évoluer indépendamment avec la fusion tégument-nucelle. Cette nouvelle structure est vascularisée. Dans une lignée, la graine s'aplatit (platyspermie secondaire, Nucellangium, Carbonifère supérieur) Mais la vascularisation des tissus fusionnés présente encore la symétrie de type radial. Dans l'autre lignée la symétrie radiale est totalement conservée (Pachytesta, Carbonifère supérieur). Chez les Cycadopsides il peut y avoir une platyspermie secondaire avec des traces de la symétrie radiale au niveau de la vascularisation. Nucellangium (lignée des Pinopsides évolue ensuite avec une réduction de la vascularisation (Mitrospermum, Carbonifère supérieur,Pennsylvanian puis Pinus).

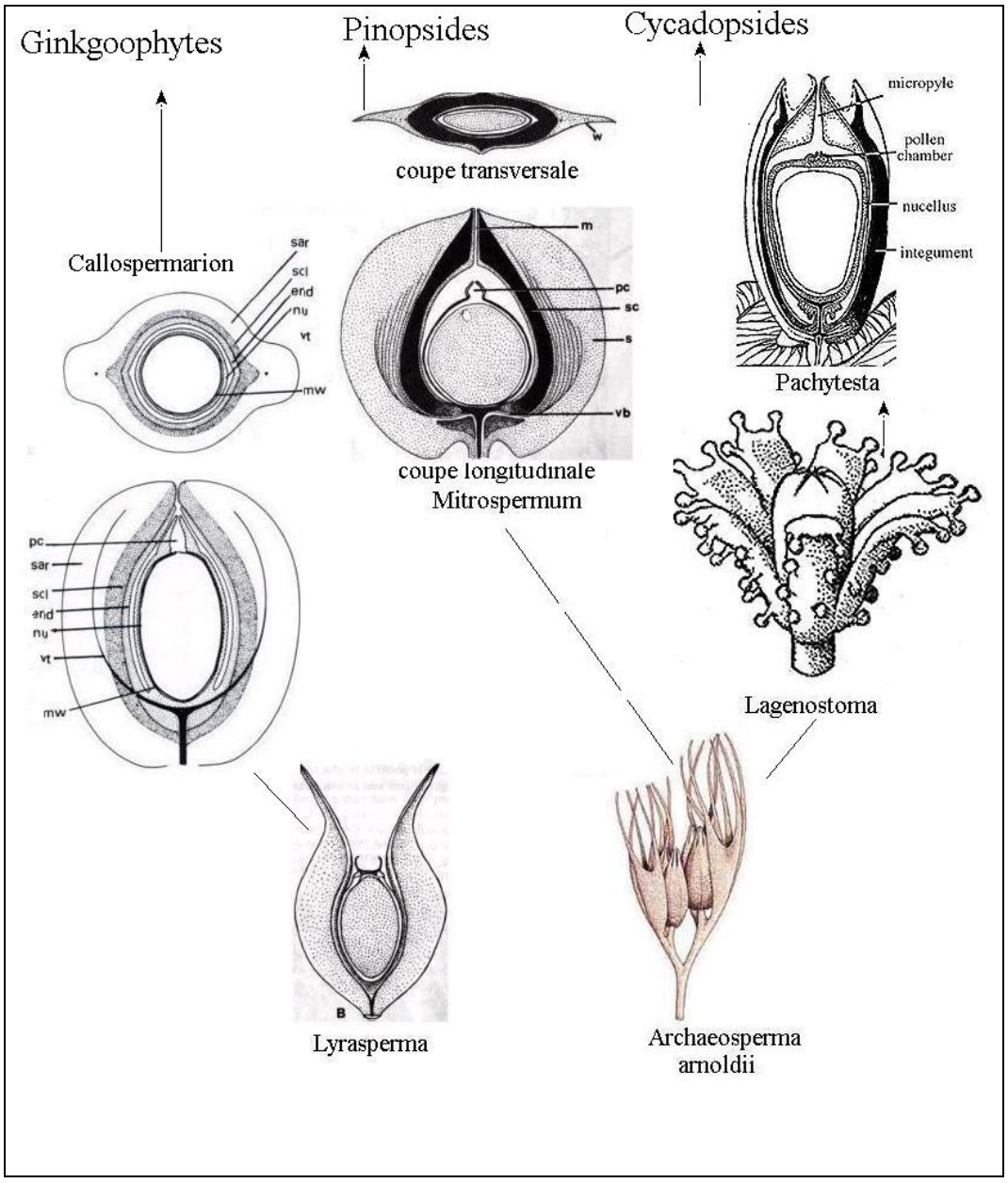


Figure I- 2. L'évolution dans la lignée gymnosperme selon Meyen.

Une nouvelle méthodologie voit le jour qui tente de prendre plusieurs caractères en compte et d'appliquer des méthodes objectives. Deux types de solutions se sont développées.

La phénétique numérique objective qui se donne pour but de quantifier le degré de similitude entre taxa
Le cladisme qui regroupe dans un ;même ensemble le plus proche ancêtre commun et tous ses descendants.

Les premières tentatives de phénétique datent des années 1850 et portaient surtout sur des différences entre variétés. En 1939 Sturtevant (1891-1970) dans le but d'exclure des caractères adaptatifs, étudiant des drosophiles définit 39 **Bons** caractères (qui sont corrélés avec d'autres) et montre que ces méthodes sont plus fiables si les groupes étudiés sont plus proches. En 1957 l'apparition des ordinateurs permet des avancées appréciables avec Michener et Sokal (1926-) aux USA et Sneath en Grande Bretagne. Ils donnent un poids égal

à tous les caractères puis en font une évaluation phylétique. 1963 première publication de Principles of Numerical Taxonomy, 1973 réédition corrigée. Ces auteurs ne prennent pas en considération les données relatives à la descendance d'ancêtres communs. Ils font des groupes et essaient ensuite de voir s'ils sont monophylétiques.

Le cladisme qui veut également éliminer la subjectivité a été fondé par Hennig (1913-1976) (publication en allemand 1950 « Phylogenetic systematics », traduction en anglais 1966). C'est un classement basé sur la généalogie qui est représenté comme une suite de dichotomies d'espèces une espèce parentale se divisant entre deux espèces sœurs (même si l'espèce ancêtre et une des deux sœurs reste identique). Ce n'est qu'à partir de 1966 date de la publication de la traduction anglaise de son ouvrage que cette théorie va prendre de l'importance. Contrairement aux autres évolutionnistes Hennig distingue les états ancestraux (plésiomorphie) des états dérivés (apomorphies). Cette méthode a pour ambition de restituer (reconstruire) la phylogenèse sans recours aux fossiles. Depuis Darwin la recherche portait essentiellement sur des taxa monophylétiques. Ici cette monophylie est basée uniquement sur des synapomorphies. On a une nouvelle façon d'évaluer les caractères. Une similitude peut représenter une homologie[4] ou une convergence. A son tour une homologie vraie peut résulter d'une synplésiomorphie ou d'une synapomorphie. Seuls ces derniers caractères partagés sont considérés comme recevables pour établir la phylogénie. Une des critiques de cette méthode est qu'elle amène à rassembler des taxa (tels qu'oiseaux et reptiles) si différents qu'ils n'ont plus que des autapomorphies. Un groupe a pu changer en passant dans une autre niche alors que l'autre n'a que peu varié.

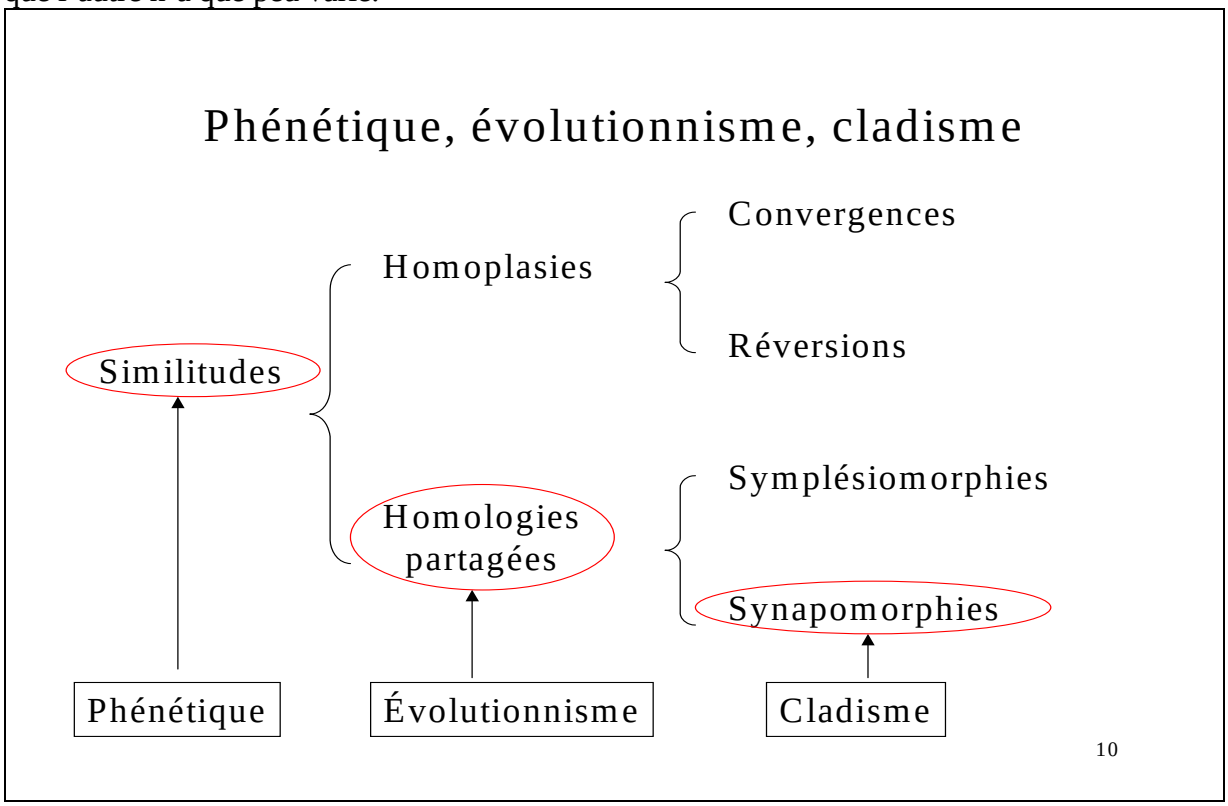


Figure I- 3.Les différents types de ressemblances et les écoles phylogénétiques.

[4] Homologie: un trait est homologue dans deux taxa si l'on peut démontrer qu'il provient du même trait chez l'ancêtre présumé de ces taxa. L'homologie de position (selon Geoffroi Saint Hilaire) ne s'applique pas aux homologies biochimiques, comportementales, etc.

L’étude cladistique produit un cladogramme qui reflète les parentés des différents taxa. Ce cladogramme peut devenir un phylogramme pour montrer également la quantité des différences, c’est la façon de voir des taxinomistes évolutionnistes ou éclectiques qui accordent du poids aux autapomorphies (Mayr 1969).

Origine de la diversité

Hypothèses fondatrices

C’est au XXe siècle que l’on voit les principales avancées dans l’estimation de la phylogénie. Jusque dans les années 60 seules les variations morphologiques et comportementales étaient prises en compte. Ensuite des propriétés immunologiques furent considérées ainsi par exemple que la présence/absence de protéines enzymatiques, les cartes de restriction et le degré d’hybridation des acides nucléiques d’espèces différentes. Depuis les données biologiques ont pris de plus en plus d’importance :
les protéines furent séquencées à partir de 1949 avec les travaux de Sanger
les acides nucléiques (ADN et ARN) à partir de 1975 avec les travaux de Maxam et Sanger. Ce domaine s’est énormément développé avec la mise au point de la Réaction de Polymérisation en Chaîne (PCR) automatisée en 1989
L’analyse des premières séquences protéiques conduisit Zuckerkandl et Pauling à formuler l’hypothèse que la différence de protéines homologues chez différents organismes est fonction du temps depuis lequel ces organismes ont évolué indépendamment depuis leur dernier ancêtre commun, hypothèse connue sous l’appellation d’horloge moléculaire. Ils ont remarqué différentes caractéristiques de l’évolution des protéines homologues dans différents organismes :
Les différents acides aminés ont des propriétés physicochimiques chevauchantes
Le rôle d’un acide aminé particulier dans une protéine peut être rempli par d’autres acides aminés
La plupart des changements observables d’acides aminés sont sélectivement neutres
L’évolution adaptative des protéines peut ne pas requérir beaucoup de changements
Pour une classe donnée de protéines, la vitesse de remplacement des acides aminés est à peu près constante

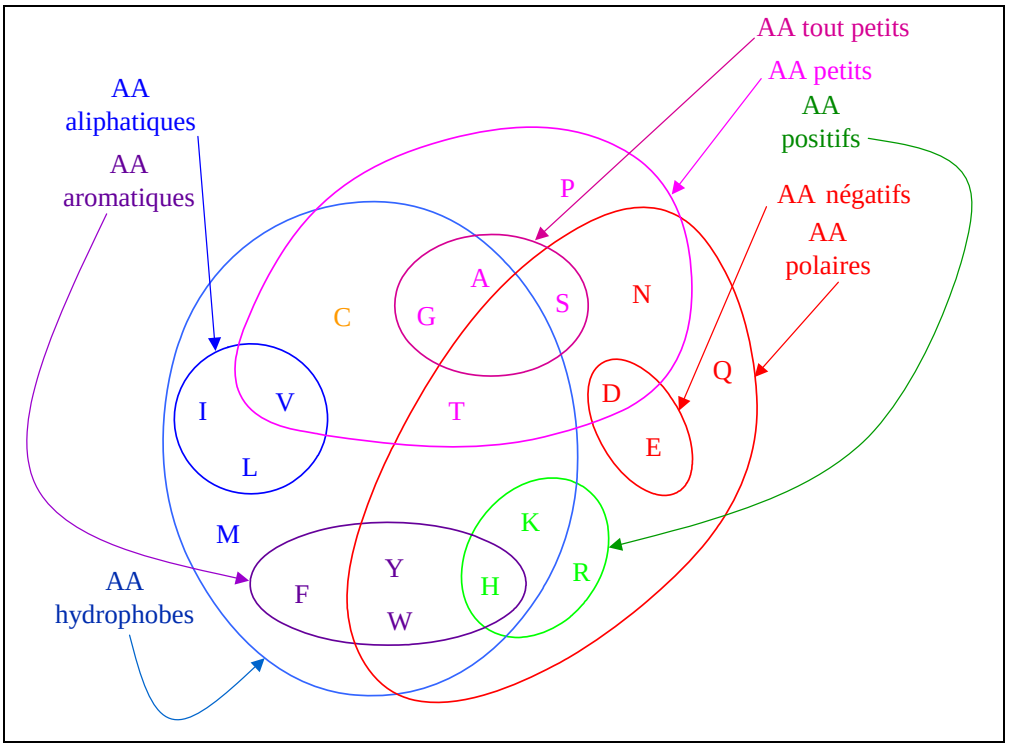


Figure I- 4. Les différent acides amines ont des propriétés physicochimiques chevauchantes.

En 1970, la théorie neutraliste de Kimura donnait un fondement théorique à ces observations.
Les hypothèses de cette théorie neutraliste sont :
polymorphisme sélectivement neutre
mutations sélectivement neutres
fixation aléatoire de ces mutations
Ce qui a pour conséquence la distribution des mutations (fr M) dans une population sur T générations = TM

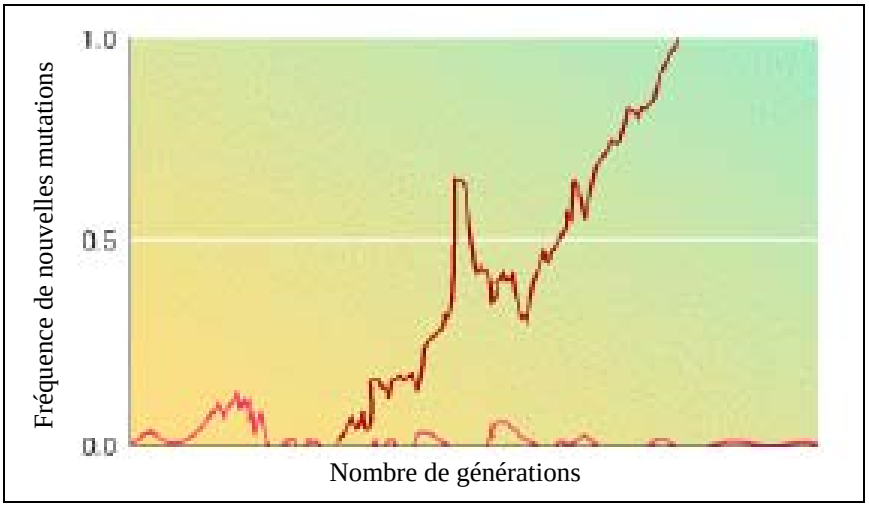


Figure I- 5. L'horloge moléculaire prévoit que la probabilité qu'une substitution dans l'ADN soit fixée par unité de temps est constante, mais pas que cette vitesse de fixation soit constante. On doit attendre des variations statistiques dans les changements de l'ADN parmi différentes lignées pour une période donnée.

Dans le cas de données moléculaires un caractère peut être

- la carte de restriction de parties équivalentes de différents génomes etc.(RFLP)
- un gène absent ou présent
- les diverses formes (séquence) d'un même gène dans différentes espèces
- Les diverses formes de microsatellites
- Présence ou absence de fragments comparables du génome (AFLP, RAPD)
- la carte d'un génome (exemple croissant des génomes mitochondriaux)

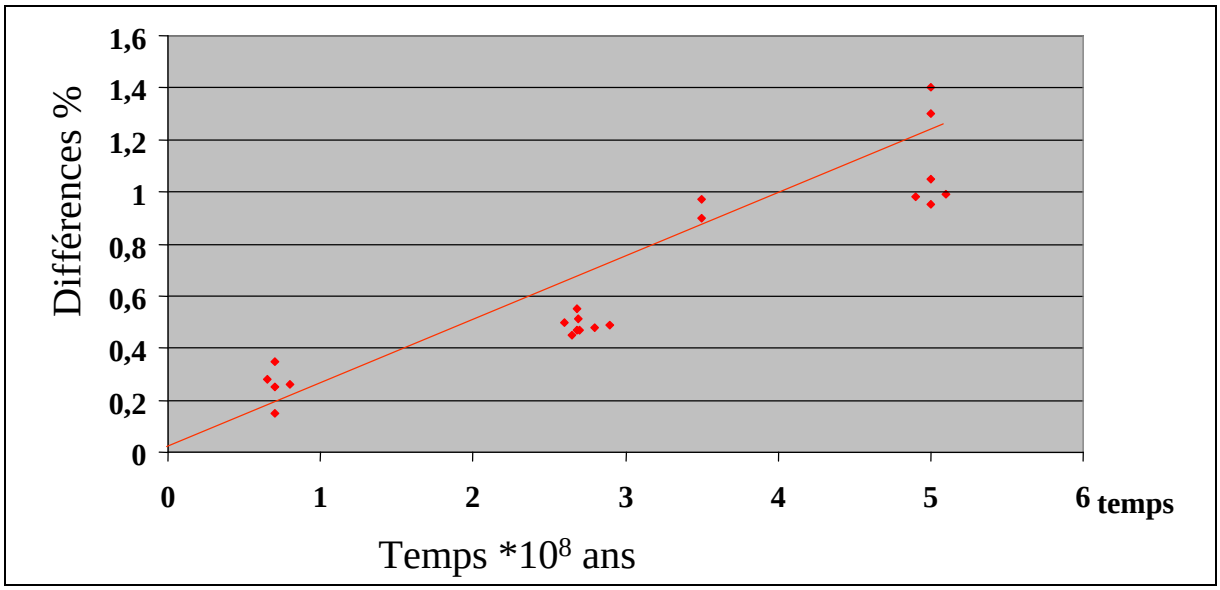


Figure I- 6. Divergence entre globines (AA) en fonction du temps supposé qui sépare 2 taxa

LUCA

Si l'on considère l'idée actuelle de la construction du monde vivant, tous les organismes actuels dérivent d'un unique ancêtre commun (Last Universal Common Ancestor) qui devait déjà posséder un génome assez complexe d'après les dernières estimations mais qui s'est complexifié au cours du temps.

Quelle était la taille du génome de LUCA ?

Familles multigéniques

Lorsqu'on hybride les clones d'une banque d'ADN génomique avec un ADNc donné, dans des conditions de moyenne stringence, il arrive assez souvent que l'on repère plus d'un gène. Les gènes ainsi obtenus présentent une certaine ressemblance dans leur séquence primaire (nucléotides ou acides aminés) que l'on interprète comme le reflet d'une origine commune.

L'examen des protéines d'un organisme révèle dans la séquence primaire ou dans la structure dans l'espace. De telles séquences sont considérées comme descendant d'un gène unique par le jeu de duplications multiples (origine modulaire des protéines)

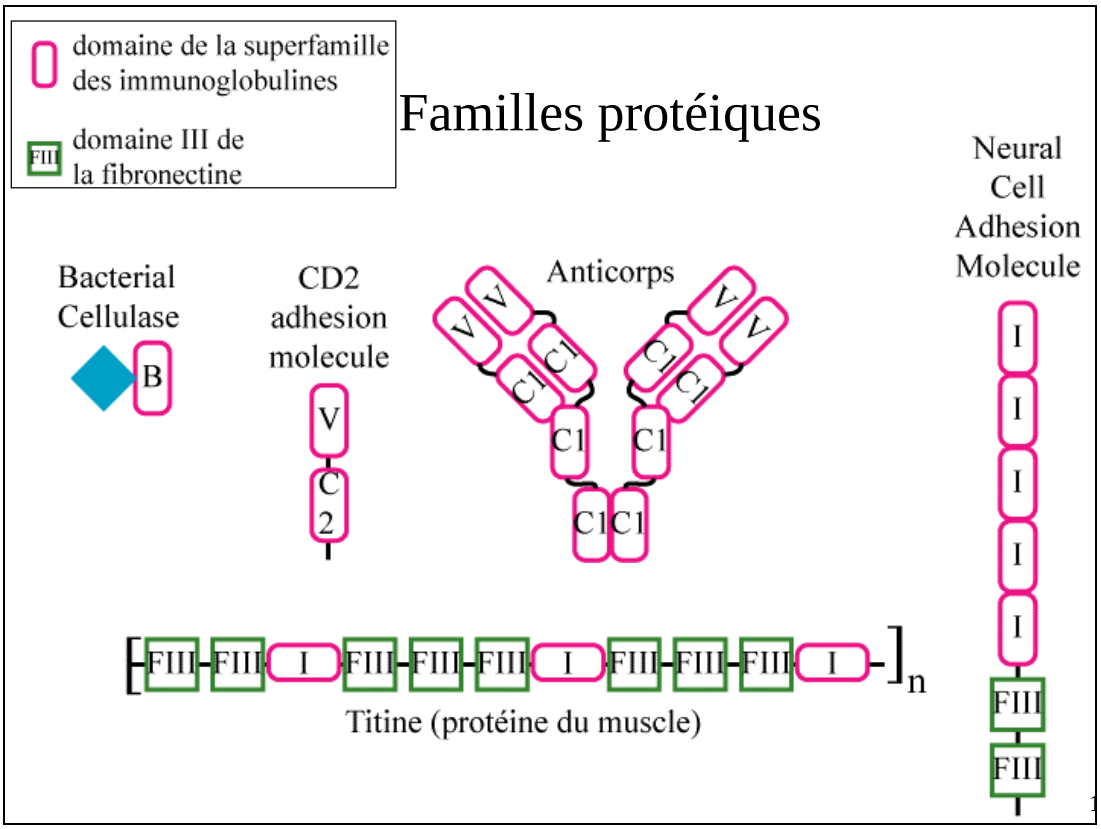


Figure I- 7.Quelques familles protéiques qui présentent des modules en commun.

Lorsque l'on ne peut pas trouver d'analogie entre les séquences primaires cela ne signifie pas qu'il n'y a pas une lointaine origine commune.

L'étude des immunoglobulines montre que des molécules différentes sont constituées de modules variables et constants dont on peut comparer la structure tridimensionnelle. Dans ces conditions on arrive à superposer des parties de molécule : plutôt les parties centrales (permanence des feuillets β). Ce qui permet de repérer des homologies indécelables autrement.

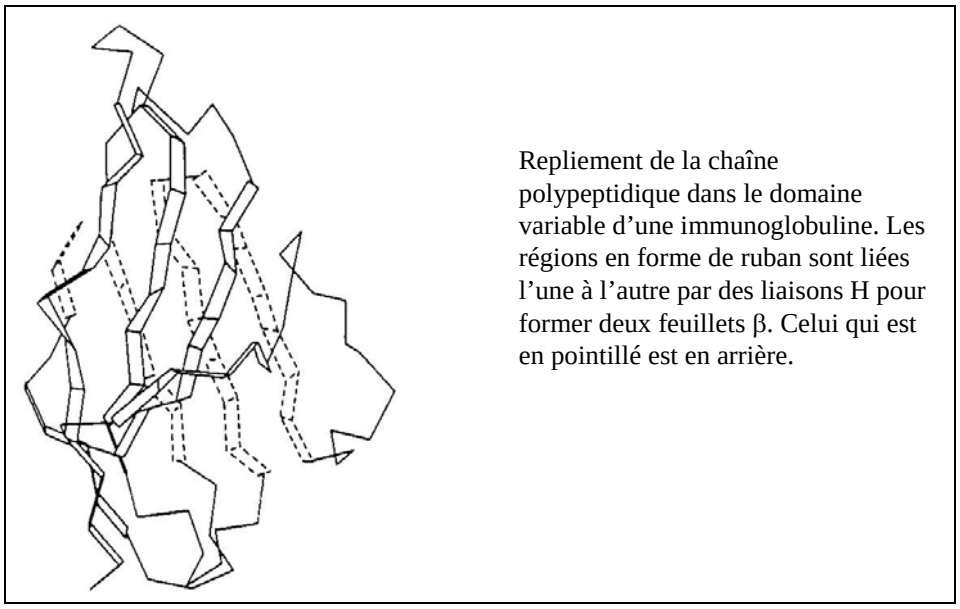
V 110 AA présentent 50% de conformation identique

C 100 AA présentent 55% de conformation identique

mais si l'on se contente de la structure primaire il n'y a plus que 15% de résidus identiques. Cela se retrouve dans d'autres protéines : lorsqu'il y a une conservation de l'ordre de 20%, cela n'est plus interprétable car une ressemblance fortuite (tirage au hasard avec la même composition) pourrait donner la même chose.

Identité en structure primaire	Homologie de conformation (structure secondaire)
40%	bien conservée ≈100%
20%	50%

Remarque : 20% d'identité dans la SI n'est pas distinguable du bruit de fond, alors que 50% d'homologie conformationnelle est encore repérable.



Repliement de la chaîne polypeptidique dans le domaine variable d'une immunoglobuline. Les régions en forme de ruban sont liées l'une à l'autre par des liaisons H pour former deux feuillets β. Celui qui est en pointillé est en arrière.

Figure I- 8. Structure des feuillets β dans les domaines V et C des immunoglobulines

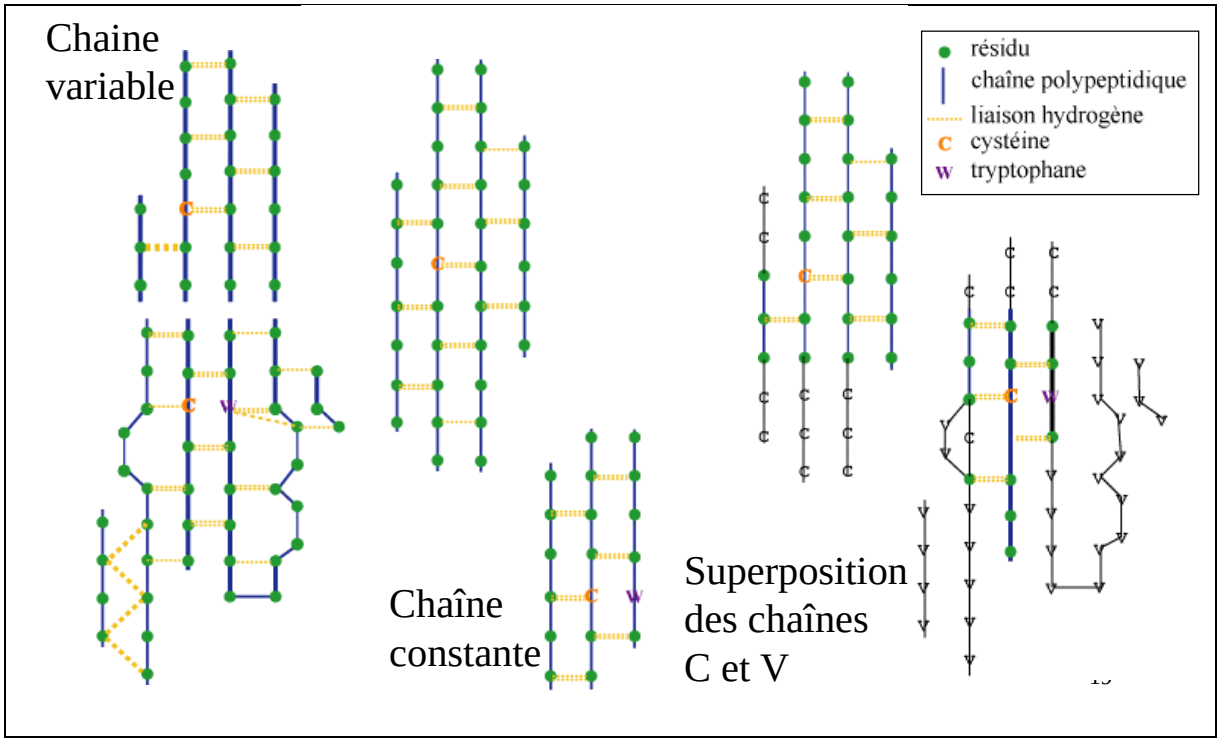


Figure I- 9. Comparaison des structure des feuillets β dans les domaines V et C des immunoglobulines

Etude des banques

Estimation du nombre de gènes de base d'après une récapitulation de 1994.

*Des séquences nouvelles vers les banques : Les banques constituent une source des gènes actuellement connus. Elles contiennent un certain nombre d'ORF. Parmi les ORF nouvellement produites on constate qu'un tiers seulement présente de l'homologie avec ce qu'il y a dans les banques. Celles-ci représentent donc 1/3 de l'ensemble des gènes.

*Des banques vers les banques : Parmi les séquences des banques 28% présentent de l'homologie avec des séquences déjà connues ; autrement dit ¼ des protéines appartiennent à une famille protéique dont on connaît déjà la structure moléculaire.

*Dans la banque Brookhaven protéique il y a un ou plusieurs représentants de 120 protéines appartenant à des familles différentes.

A partir de ces données on peut estimer grossièrement la taille du génome minimal qui a ensuite donné naissance à la diversité actuelle.

- Si l'on admet que l'on a détecté toutes les homologies existantes on peut faire le calcul suivant :

120 familles
x 4 dans les banques ¼ des protéines appartiennent à 1 famille connue
x 3 les banques contiennent 1/3 de ce qui est connu

1500 familles issues de 1500 gènes différents

- Si l'on admet que l'on n'a détecté que 80% des homologies existantes

120 familles
x 4 dans les banques ¼ des protéines appartiennent à 1 famille connue
x 0,8 car il n'y en a que 80% qui sont différentes, les autres homologies existent mais n'ont pas été détectées
x 3 les banques contiennent 1/3 de ce qui est connu
x 0,8

1000 familles issues de 1000 gènes différents

Comparaison de génomes

En 1999 la première séquence complète d'archaebactérie (*Methanococcus jannaschii*) a permis de comparer des génomes complets de chacun des trois domaines. Parmi 1756 gènes identifiés (ORF) certains (324) avaient un homologue dans chaque domaine. Les fonctions représentées étaient principalement des protéines enzymatiques du métabolisme, des transporteurs, des protéines se liant à l'ATP ou au GTP, des tyrosine-phosphatases, des protéines ribosomiques, des amino acyle ARNt synthétases, des facteurs d'initiation de la traduction, des hélicases et des ARN polymérases. Ce nouveau génome a permis de combler le fossé entre Eubactéries et Eucaryotes. Le génome de LUCA devait être composé principalement d'éléments structuraux alors que la composante régulatrice était absente. Deux interprétations possibles : soit cette régulation a tellement changé d'un domaine à l'autre qu'on ne reconnaît plus les éléments paralogues soit cette composante était absente chez LUCA. 3 conclusions principales peuvent être tirées de ces observations.

les archaebactéries semblent plutôt représenter une forme ancestrale de la vie qu'une chimère. L'étude des histones archaebactériennes a montré qu'elles étaient plus poly fonctionnelles que les histones hautement spécialisées des Eucaryotes (motifs archaébactériens communs à histones et CBF (-B :CCAAT-binding transcription factor, CBFD : Histone-like transcription factor (CBF/NF-Y) and archaeal histone) Eukaryotes
Les archaebactéries semblent partager une plus grande partie de génome avec les Eubactéries qu'avec les Eucaryotes, sans dire qu'elles en sont plus proches.
On peut maintenant préciser la nature du génome de LUCA qui doit avoir déjà acquis une complexité certaine. Il est constitué de gènes enzymatiques et de systèmes génétiques analogues à ceux des unicellulaires actuels. (Kyrpidez, Overbeek, Ouzounis ,1999 J. Mol. Evol. **49** pp413-423)

Racine de l'arbre du vivant

L'arbre universel du vivant reconnaît trois groupes
Les Archaebactéries (A)
Les Eubactéries (B)
Les Eucaryotes (E)

Par contre la position de la racine et donc de LUCA reste ambiguë. On a tenté de la placer en construisant l'arbre avec deux protéines qui sont issues d'une duplication très ancienne.
Cet arbre présente un certain nombre d'artefacts.
Les protistes dépourvus de mitochondries sont placés à la base de l'arbre. C'est un phénomène d'attraction de longues branches (vitesse d'évolution plus grande et ressemblance plus grande par hasard (plus d'espace pour les positions évolutives).
Le groupement des hyperthermophiles à la base de l'arbre est également un artéfact (attraction des branches courtes). Ces organismes ont un plus forte proportion de GC ce qui réduit l'espace évolutif.
La reconstruction des ARNr de LUCA (méthode de parcimonie indique qu'il ne serait pas hyperthermophile (% GC de type mésophile)

- Enfin, la position de la racine serait également due à une attraction des longues branches.

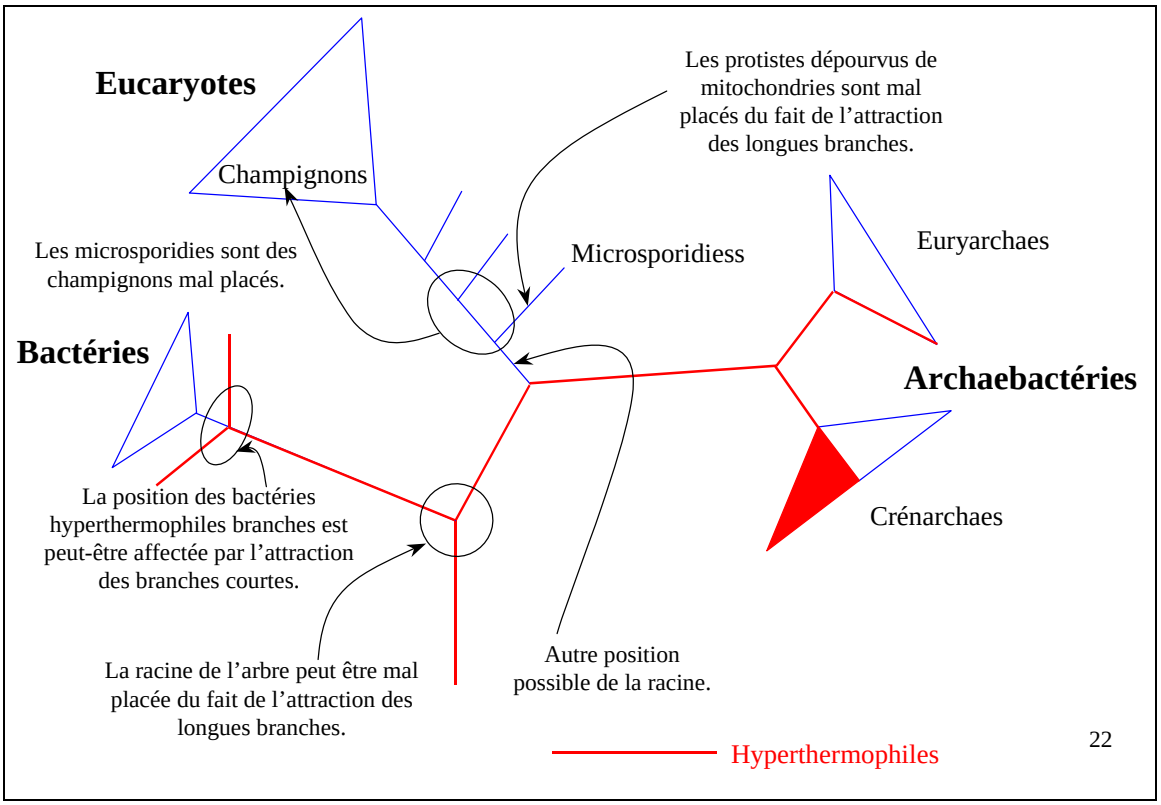


Figure I- 10. Une version de l'arbre universel du vivant.

D'autres hypothèses ont été proposées.

La ré analyse des données protéiques suggère une racine sur la branche la plus longue des Eucaryotes (H3)

Pour éviter les contradictions entre différents arbres protéiques, une autre hypothèse (H1) serait que les Eucaryotes proviendraient pour leur cytoplasme des Bactéries et pour leur noyau des Archaebactéries. Dans ce cas cependant on a du mal à expliquer la disparition au niveau des ribosomes Eucaryotes de la forme Bactérienne remplacée par Archaebactérienne

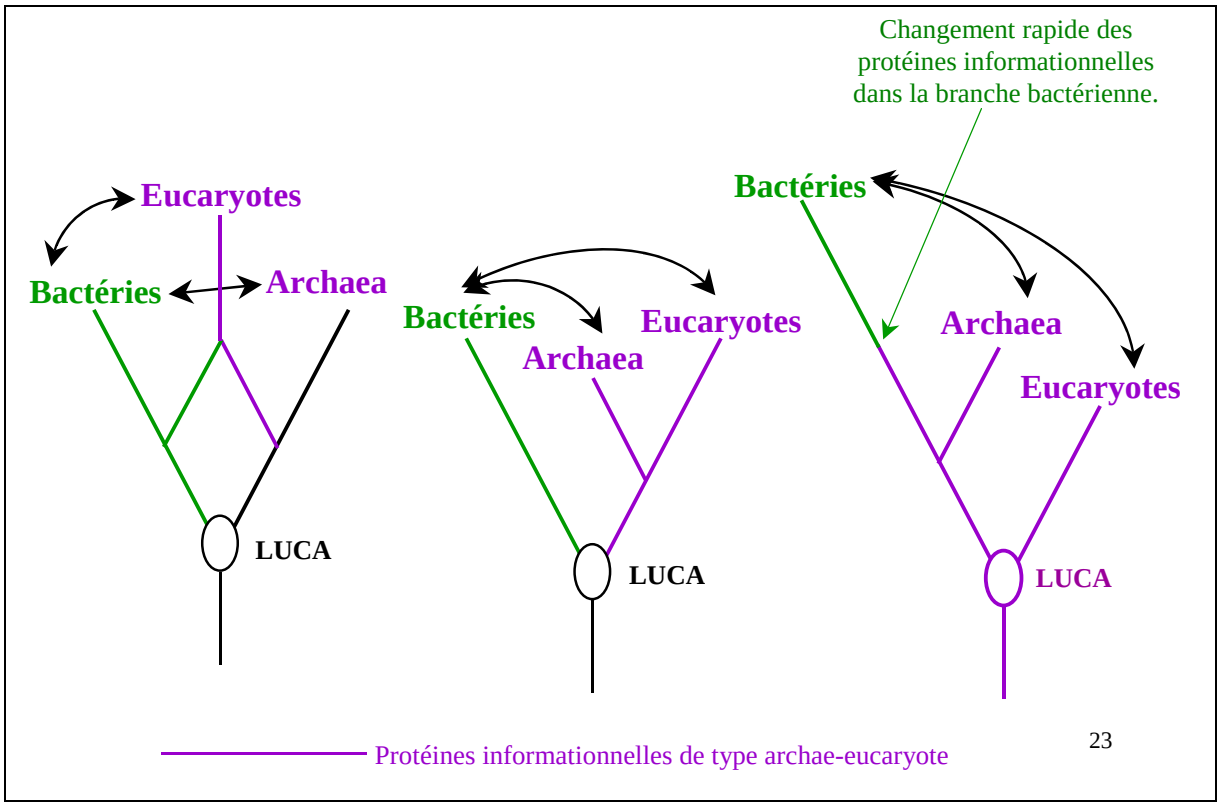


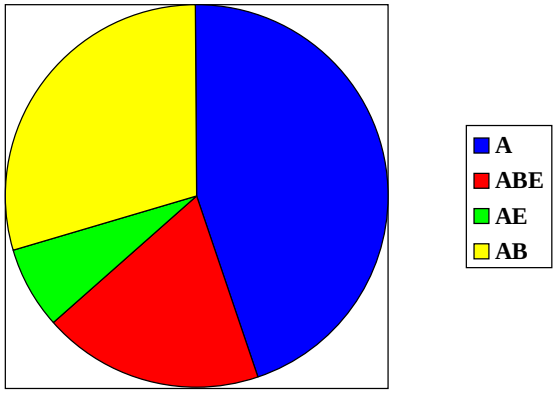
Figure I- 11.Les différentes racines possibles de l'arbre universel.

En 2001 on disposait de

- 28 génomes B complets
- 10 génomes A complets
- 5 génomes E complets.

La génomique confirme l'existence des trois domaines avec des protéines spécifiques à chacun d'eux. Tous sont monophylétiques (ce qui élimine une hypothèse selon laquelle les A auraient été divisées en 2 EuryA (plus proches de B) et CrénoA (plus proches de E).

Domaine	Classes				
	Energie	Information	Communication	Hypothétique	Total
Universel ABE	178	103	20	23	324
Non eucaryote AB	264	94	32	132	522
Non bactérien AE	10	79	7	27	123
Archéen A	80	34	11	662	787
Total	532	310	70	844	1756



21

Figure I- 12. Résultats de la comparaison des génomes complets des trois domaines

On retrouve parmi les protéines appartenant à plusieurs domaines ABE, AB, AE et EB. Il a donc dû y avoir du transfert latéral de gène sans doute plus important qu'on ne le pense. En particulier entre organismes qui ont partagé le même environnement :

- les bactéries hyperthermophiles *Aquifex* et *Thermotoga* ont plus de protéines de type archéen que de protéines de type bactérien mésophiles
- l' Archébactérie thermoacidophile *Thermoplasma acidophilum* (euryA) présente plus de protéines apparentées à *Sulfolobus solfataricus* (crénoA) qui vit dans le même milieu que de protéines de type Eurybactérien.
- L'examen de l'arbre phylogénique de la réverse gyrase qui n'existe que chez les hyperthermophiles met également en évidence des transferts latéraux

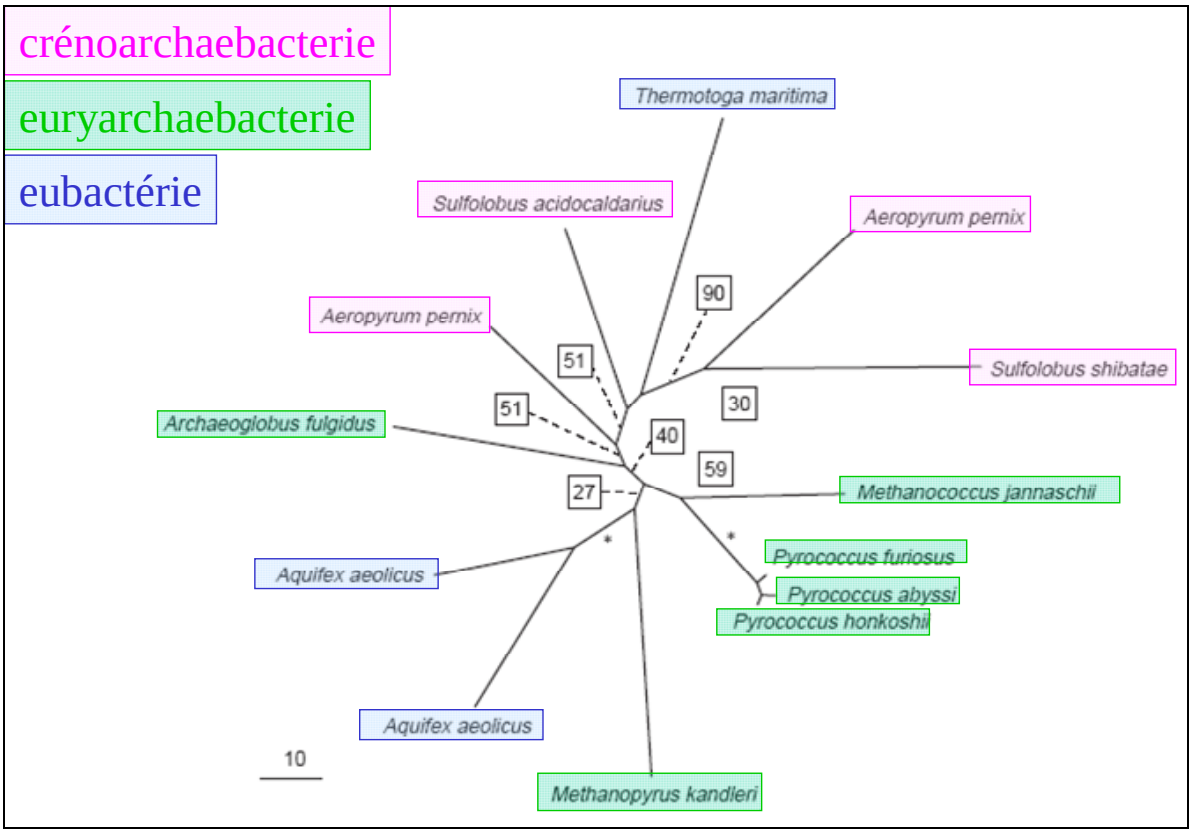


Figure I- 13. Arbre obtenu avec la séquence de la réverse gyrase spécifique des bactéries hyperthermophiles.

Parmi les gènes susceptibles de composer le génome minimum (256gènes plus peut-être quelques autres obtenu par comparaison des génomes complets d'*Hemophilus influenza* et *Mycoplasma genitalium*), on trouve

- Un équipement presque complet pour la traduction
- Un équipement presque complet pour la réplication
- Un système très rudimentaire de réparation
- 4 sous unités de transcription, aucune protéine régulatrice
- un lot étonnamment grand de chaperonines (Chaperonine: une protéine spécialisée qui associée à un polypeptide naissant en cours de synthèse l'empêche de se replier prématurément).
- Gènes pour un métabolisme intermédiaire anaérobie
- Pas de synthèse d'aminoacides ni de nucléotides ni d'acides gras.

On constate que les gènes impliqués dans la réplication de l'ADN n'ont pas d'orthologues dans l'ensemble AE. Ceci amène à deux nouvelles hypothèses..

- Soit LUCA avait déjà un génome à ADN et sur la branche (sur la longue branche B° il y a eu substitution du système AE par un autre non homologue
- Soit LUCA avait un génome à ARN et le passage à l'ADN s'est fait 2 fois indépendamment avec deux systèmes de réplication différents.

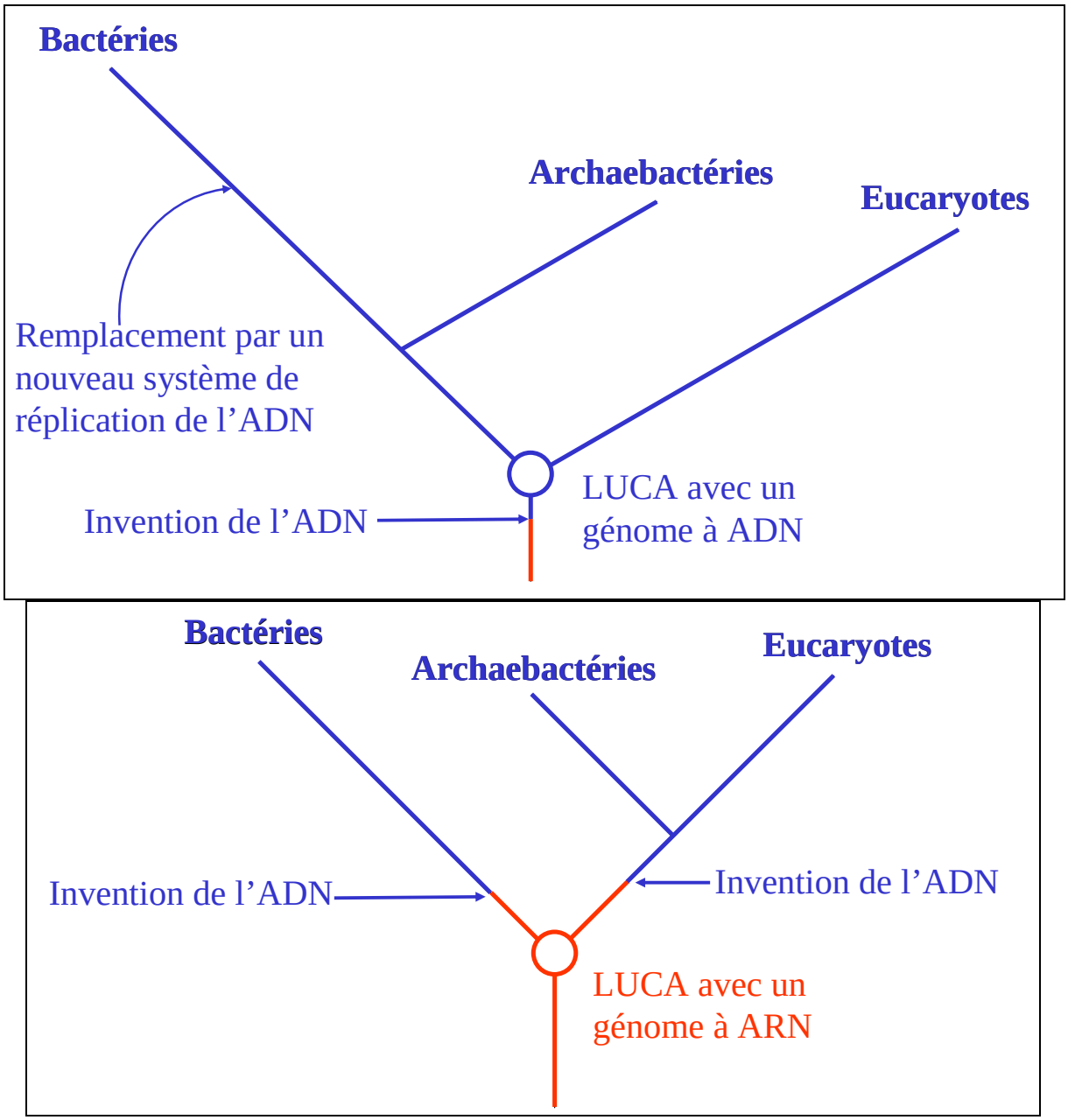


Figure I- 14. Les deux hypothèses pour l'invention de l'ADN.

Gain de complexité

Les procédés de complexification sont une (ou plusieurs) duplication d'un gène donné ou du génome entier (les deux sont possibles). Ensuite il y a donc un exemplaire surnuméraire qui peut être perdu, conservé avec la même fonction (gènes non essentiels qui assurent la même fonction et pour lesquels on ne peut pas ou difficilement, obtenir des mutants), ou évoluer vers une autre fonction (on connaît actuellement des protéines qui assurent des fonctions comparables bien que chacune spécialisée, la séquence protéique et donc la séquence génique de chacune de ces molécule présente des caractéristiques communes qui attestent une origine commune, cas des cyclines dans la division cellulaire).

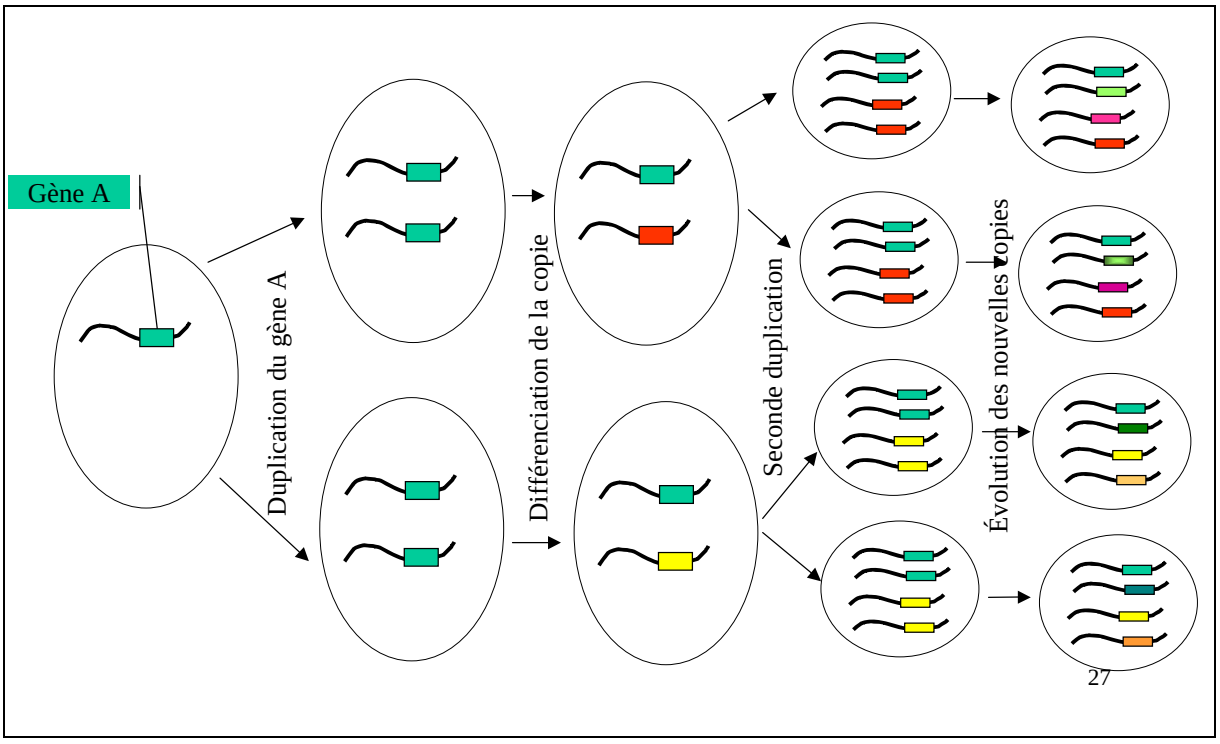


Figure I- 15. Principe de l'évolution du génome de LUCA aux taxons actuels.

C'est cette image de la complexification progressive du vivant qui a conduit à une représentation de son histoire par un arbre dichotomique.

Définitions

Parmi les gènes homologues, il y a lieu de distinguer les gènes paralogues (provenant d'un ancêtre commun par duplication dans des espèces différentes ou au sein de la même espèce) et les orthologues (provenant d'un ancêtre commun par modifications dans des espèces différentes).

Quelques définitions

- Similarité : c'est une observation empirique qui peut être quantifiée. Cette 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 0 ou 1. C'est déjà une hypothèse que de parler de caractères homologues
- Orthologue : un cas particulier d'homologie, la divergence s'est faite avant la spéciation, donc dans toutes les espèces on voit un gène qui est issu du même gène dans le dernier ancêtre commun à ces taxons.
- 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.

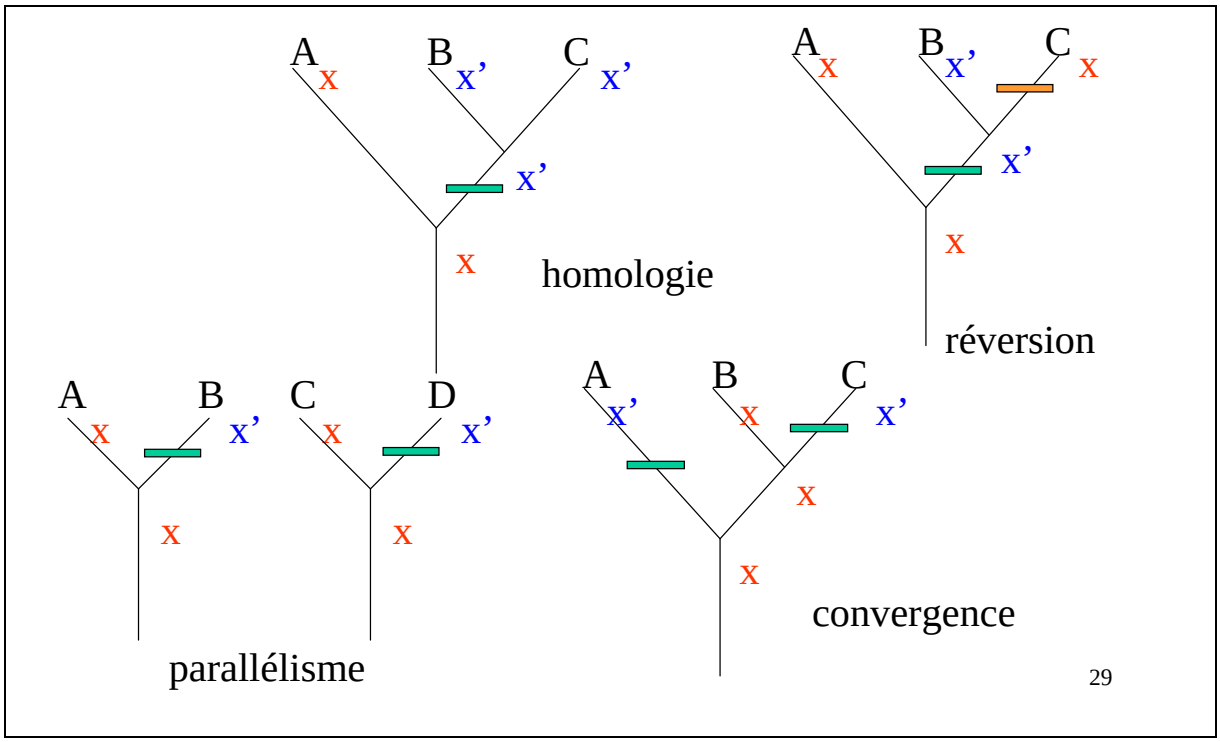


Figure I- 16. Les différents cas possibles de similitude.

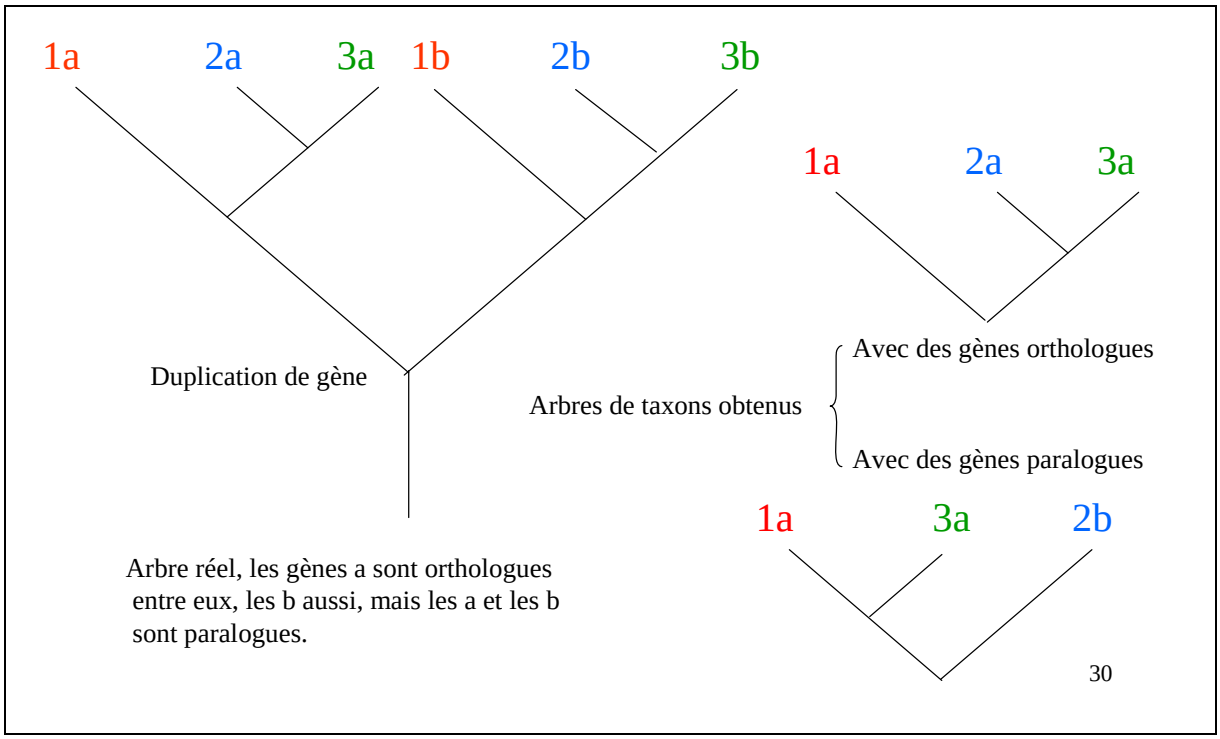


Figure I- 17. En haut à gauche, l'arbre vrai avec trois taxons et les gènes orthologues et paralogues, à droite, en haut on retrouve cet arbre des espèces avec les gènes orthologues ; par contre en bas un gène paralogue remplace un orthologue et les relations entre taxons ne correspondent plus à l'arbre vrai.

L'utilisation de gènes orthologues permet de retracer l'histoire de la spéciation. Par contre, si des gènes paralogues sont utilisés par erreur (la distinction orthologues / paralogues n'est pas toujours simple) l'histoire retrouvée par des méthodes de construction d'arbres est erronée (figure 17).

Lorsqu'on a fait cette distinction deux problèmes différents peuvent être proposés :

- reconstruire l'histoire d'un ensemble d'espèces et il faudra utiliser des gènes dont on est sûr que ce sont bien des orthologues
- reconstruire l'histoire du gène et de ses descendants (cas des gènes de globine). Dans ce cas on prendra des gènes paralogues et orthologues dans différentes espèces dont on a déjà une idée de la phylogénie de préférence. (Exemple de l'arbre des globines).

Dans l'arbre des globines on retrouve des événements de duplication qui ont généré les gènes myo, α, β, δ, γ, ζ, .. globines et pour chaque molécule, idéalement on devrait retrouver le même arbre des espèces.

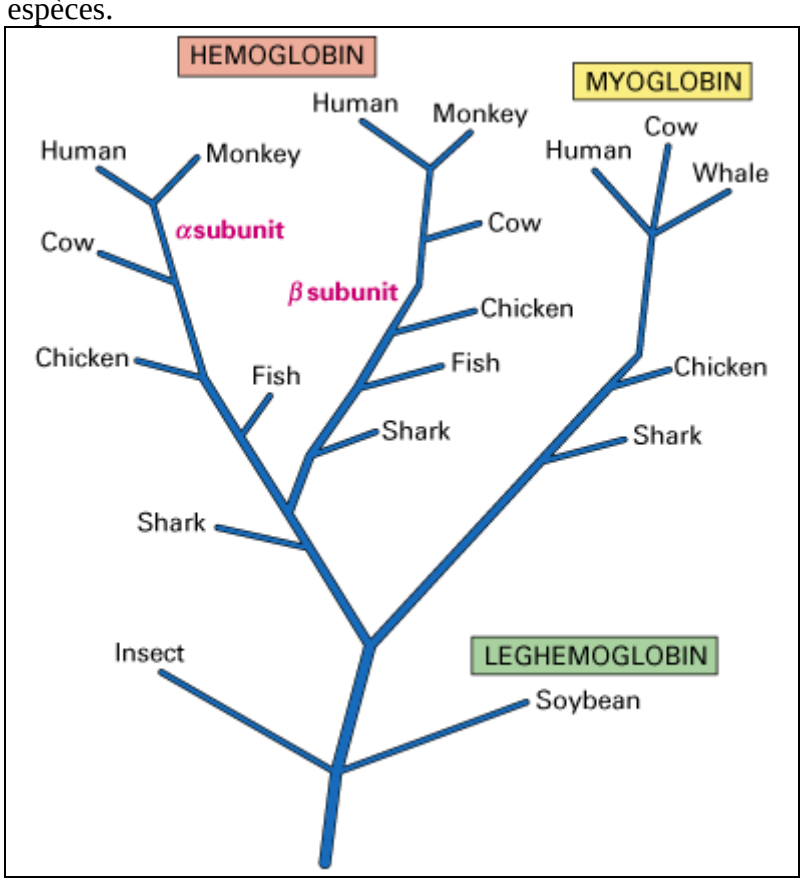


Figure I- 18.Dans cet arbre des myoglobines et globines α et β, les relations entre taxons sont les mêmes pour les trois gènes.

Outils phylogénétiques

Arbres ou réseaux

Ce concept phylogénétique de descendance avec modification peut être rendu par un arbre phylogénétique dans lequel on suit le long des branches l'évolution d'un gène donné au sein des différentes espèces étudiées. Mais la nature n'est pas si simple. Il existe des phénomènes dits de transferts horizontaux, l'exemple le plus courant étant l'acquisition d'un second génome qui devient une mitochondrie (ou un chloroplaste). Ces événements ne sont cependant pas limités à des événements ancestraux et sont encore observables de nos jours. On les observe dans le monde animal par passage d'éléments transposables d'un taxon à un autre et dans le monde végétal chez des organismes aptes à se reproduire et qui sont issus

d'hybridations naturelles. Dans de tels cas le modèle de l'arbre n'est plus adéquat pour rendre compte de l'évolution, il faudrait avoir recours à un modèle plus complexe de type réseau.

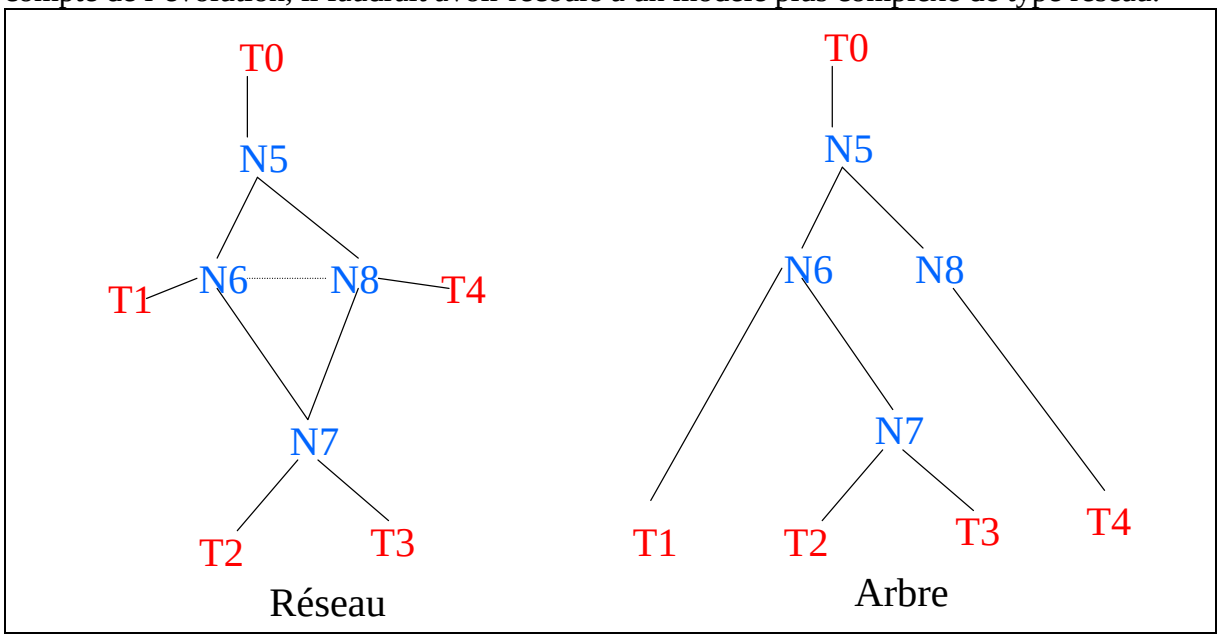


Figure I- 19. Les réseaux permettent des passages latéraux que n'autorisent pas les arbres.

Dans un réseau de relations entre divers taxons, on désigne

- les sommets externes ou feuilles constitués par les unités évolutives ou taxons terminaux
- les sommets internes ou nœuds constitués par les unités évolutives hypothétiques
- les liens auxquels on peut attacher des mesures, distance (génétique) un poids (vitesse d'évolution, qui peuvent être orientés (sens de parcours imposé).

Un réseau présente au moins un chemin entre chaque paire de sommets. Il est cyclique.

Un arbre présente au moins un chemin entre chaque paire de sommets mais il n'est pas cyclique. Actuellement on s'en tient au modèle le plus simple : celui de l'arbre qui pose déjà un certain nombre de problèmes dans sa recherche.

Parmi les diagrammes ayant une allure générale d'arbres, que l'on peut désigner d'un nom général : **les dendrogrammes**. Parmi eux, on distingue

- Les phénogrammes qui sont produits par les méthodes phénétiques où les relations entre taxons expriment les degrés de similitude globale.
- Les cladogrammes qui sont des schéma (dendrogramme) exprimant les relations de proche parenté entre taxa construit à partir de l'analyse cladistique, où les points de branchement (ou noeuds) sont définis par des synapomorphies. Contrairement au phylogramme, les points de branchement ne reflètent pas les taux de divergence
- Les phylogrammes sont des schéma de relations de parenté (dendogramme) exprimant les branchements et le degré de divergence adaptative associé à chaque branche (taxon)

Quels sont les éléments que l'on va être amené à utiliser au cours d'études phylogénétiques ?

Taxons et caractères

Taxon : groupe d'organismes reconnu en temps qu'unité formelle à chacun des niveaux de la classification. L'espèce devrait être l'unité de base, mais c'est cependant déjà un agrégat de populations. Les taxons supra spécifiques doivent être monophylétiques.

Définitions

- Taxons monophylétiques
- Taxons paraphylétiques
- Taxons polyphylétiques

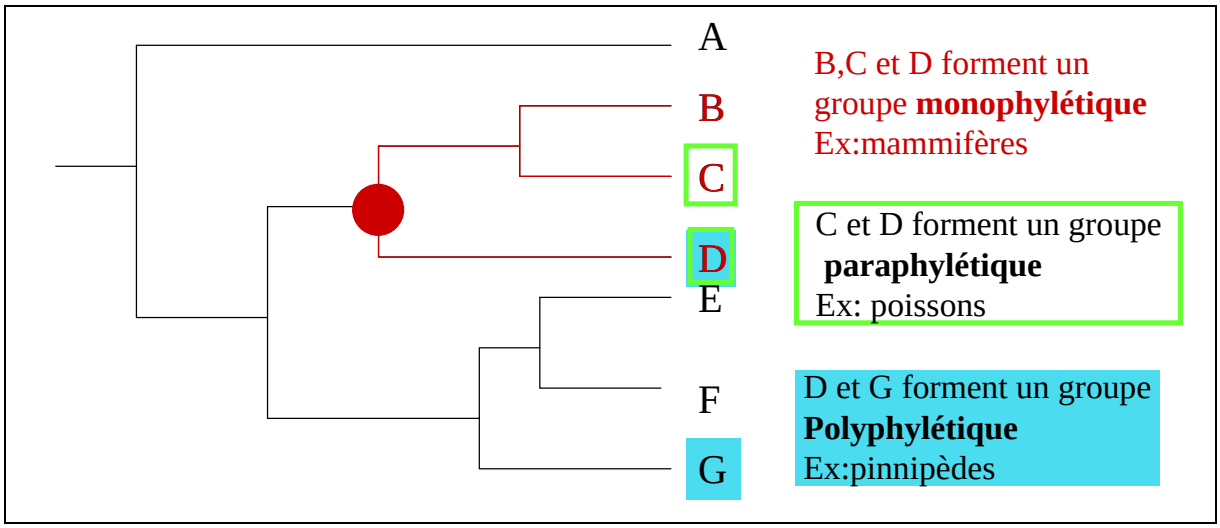


Figure I- 20. Mono, para et polyphilie.

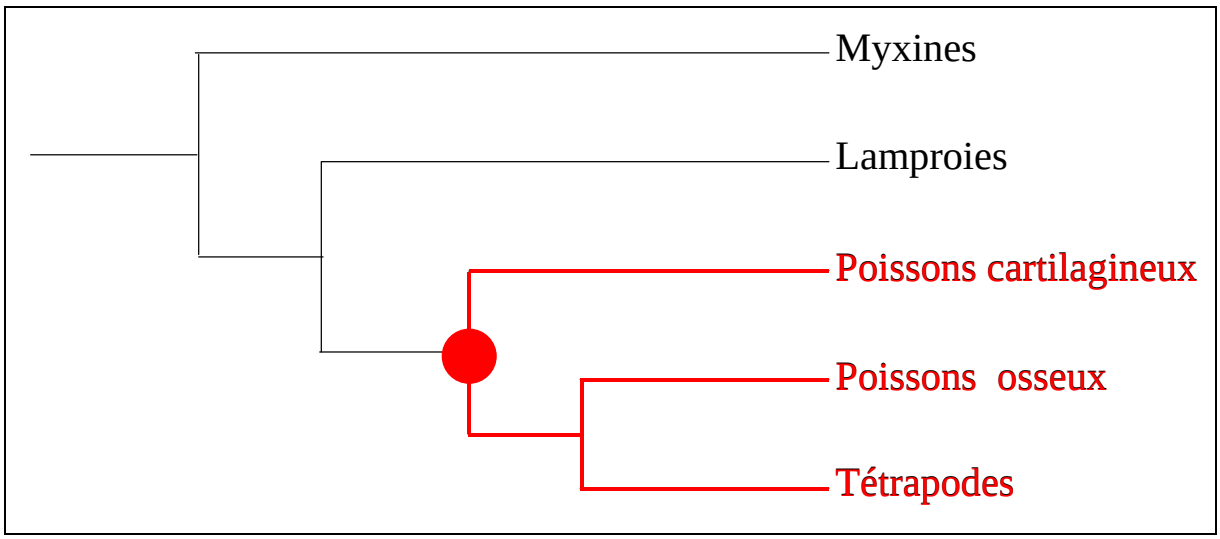


Figure I- 21. Les poissons et les tétrapodes constituent un ensemble monophylétique.

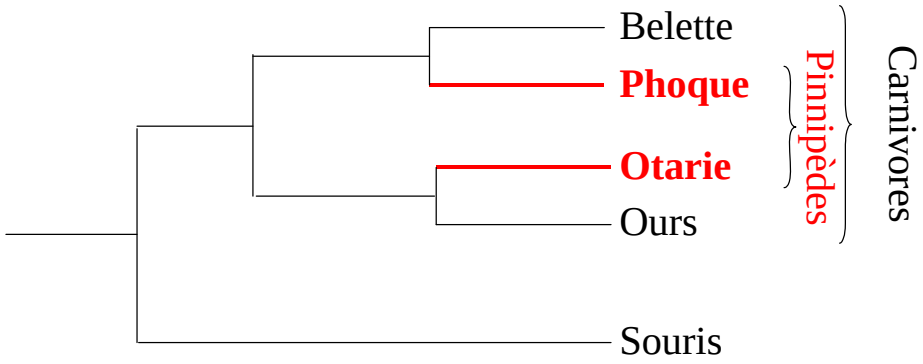


Figure I- 22.Arbre obtenu avec une séquence protéique partielle d'α globine. Les pinnipèdes forment un groupe polyphylétique.

S'agissant de caractères morphologiques on peut être tenté de choisir comme taxons les espèces ou des genres voire des familles ou des ordres, avec le risque pour certains caractères d'avoir du polymorphisme. Lorsqu'il s'agit de caractères moléculaires on prend forcément des représentants d'une espèce donnée, ce problème du polymorphisme se retrouvant éventuellement au sein des populations (par exemple cas des allèles rares d'un gène, polymorphisme tel que celui des groupes sanguins, coexistence de plusieurs allèles en proportions non négligeables).

Caractère : tout attribut observable d'un organisme. On va observer un caractère donné chez différents organisme. Un caractère donné n'aura pas la même forme chez deux organismes différents. Il faut donc définir ce que l'on entend comme caractère comparable ou homologue.

Cette notion d'homologie a été définie par Geoffroy Saint Hilaire en 1818 « *la seule généralité à appliquer dans l'espèce est donnée par la position, les relations et dépendances des parties, c'est à dire par ce que j'embrasse et ce que je désigne sous le nom de* ***connexions*** » le principe des connexions, repris par Owen en 1843, exemple : ailes des oiseaux des drosophiles des anges et bras des hommes).

Figure I- 23. Parmi ces taxons quels sont les membres antérieurs homologues?

Neutralité des caractères

Dès 1962 en observant les séquences protéiques chez divers organismes Zuckerkandl émettait l'idée de l'horloge moléculaire. A ce moment on espérait ainsi pouvoir donner le temps de divergence entre deux taxa pour lesquels on ne possédait aucun fossile. Bien qu'aujourd'hui on admette une corrélation entre la divergence moléculaire et le temps, on est loin d'accepter que la vitesse soit constante. Gillespie en 1987 notait le rapport du nombre de substitutions par position à la moyenne des substitutions qui atteignait 1/35 au sein d'une généalogie pour les acides aminés et 1/19 pour les substitutions silencieuses, ce qui trahit des fluctuations importantes.

La constance de la vitesse d'évolution est une des pierres d'achoppement de la théorie neutraliste. Théorie qui est largement admise (Kimura 1968) mais avec des modifications variées selon les cas. Les caractères moléculaires sont-ils sélectivement neutres ? Compte tenu du peu de conclusions sur les données moléculaires, on peut prendre comme hypothèse de départ cette neutralité en sachant bien que ce n'est qu'une hypothèse. L'impact de la sélection sur les études phylogénétiques dépend du nombre de cibles (nucléotides) affectées. De nombreux écarts à cette neutralité sont locus spécifique, il est largement admis que la sélection n'a qu'un impact mineur sur l'analyse lorsque de nombreux sites sont examinés.

Différents caractères doivent apporter une information originale donc être indépendants les uns des autres. On ne sait pas définir les interactions entre les différentes positions nucléotidiques, il est néanmoins évident qu'elles existent (structure secondaire de l'ARN). De la même manière on corrige cette méconnaissance en utilisant plus de sites ou l'on se sert du peu que l'on sait pour améliorer l'alignement et donc l'homologie.

Caractères morphologiques ou moléculaires ?

Les caractères sont tout ce qui s'observe que ce soit morphologique ou moléculaire. Les taxons sont un groupe d'organismes qui partagent des caractères en commun. Ils peuvent être d'ordre supérieur (genre, famille, ordre etc.), dans ce cas il peut se poser des problèmes de polymorphisme, ce qui n'est pas exclu non plus avec des taxons du type espèce car le polymorphisme existe aussi au sein des populations. Avec des caractères moléculaires on séquence obligatoirement un gène dans un individu ou une population et dans ce cas la séquence est le consensus de ce que l'on trouve dans la population.

Le débat entre la validité de ces deux sortes de caractères a été vif. Kluge en 1983 soutient que les caractères moléculaires ont un poids relativement faible tandis que d'autres soulignent que les caractères morphologiques peuvent entraîner des biais parce qu'ils ont été mal interprétés (Frelin et Vuilleumier 1979 ; Sibley et Ahlquist 1987).

Exemples tirés d'études morphologiques

Exemple des amentifères : selon que l'on considère le caractère apétale de la fleur ancestral ou dérivé l'interprétation de la position de ce groupe change.

De 1887 à 1915 Engler et Prantl mettant en homologie le strobile des gnétales et les fleurs apétales à sexe séparé des Amentifères proposèrent que ce groupe soit une forme ancestrale des Angiospermes. A partir de 1915 Bessey considéra les Ranales (order with plants like Ranunculaceae and Magnoliaceae) comme primitives; le strobilus n'étant pas unisexué mais bisexué, les fleurs apetales devenaient évoluées et non plus primitives.

Apport des caractères moléculaires

Certaines classifications ont été basées sur un seul caractère morphologique. Pour ce qui est des caractères moléculaires si le changement d'une base n'a que très peu de sens, la capacité actuelle de séquençage permet une accumulation de caractères qui doivent dans leur ensemble porter un message phylogénétique pour autant que l'on soit capable de le décoder dans sa complexité. La tendance actuelle est de reprendre l'ensemble des caractères morphologiques, éventuellement de les vérifier car certaines observations sont anciennes et peuvent être améliorées à la lumière de nouveaux travaux, d'en faire une matrice la plus complète possible et d'en tirer une phylogénie. Si dans un premier temps on s'est contenté d'étudier une partie d'un gène ou un gène unique pour un groupe de taxons dont on voulait préciser les relations de parenté, aujourd'hui avec l'extension des banques, on essaie de plus en plus de faire de grandes matrices qui contiennent souvent de nombreux taxons et des gènes qui sont de plus en plus nombreux. Les développements informatiques ont favorisé cette évolution car le traitement de grands jeux de données posait en 1993 de gros problèmes (matrice rbcL de Chase et al comportant quelques 500 taxons qu'il a fallu soumettre à plusieurs mois de traitement, en 2000 certains logiciels sont capables de faire l'équivalent en quelques jours).

Il faut également voir que les pressions de sélection sont différentes sur les caractères morphologiques et sur les caractères moléculaires. L'étude simultanée de ces deux sortes de caractères n'est pas antinomique.

Les caractères morphologiques sont les seuls à être utilisables pour les fossiles anciens (dès un âge de plus de 40 000ans) et à permettre l'usage du cadre ontogénique pour orienter l'arbre (voir plus loin les problèmes de racine).

Les caractères moléculaires n'ont comme limite que la taille du génome et offrent l'avantage d'avoir une base génétique claire.

Les résultats les plus satisfaisants sont obtenus en faisant une synthèse des deux types de caractères. Dans des cas de conflit entre différents jeux de données (matrices différentes correspondant à différents gènes ou à des gènes d'une part et des caractères morphologiques de l'autre) le retour aux caractères permet de voir si une interprétation différente des dits caractères serait plus judicieuse.

Exemple des éponges

Parmi les éponges à squelette siliceux (Démosponges) on distingue les éponges présentant des spicules de grande taille (macrosclères) ou petits (microsclères). La sous classe des Tétractinomorpha se caractérise par ses microsclères de type aster : tout microsclère dans lequel les rayons proviennent d'un centre ou d'une tige axiale qui peut être droite, en forme de C ou spiralée. On considère qu'il existe deux types mutuellement exclusifs, les euasters où les rayons proviennent d'un point central ; les streptasters dans lesquels les rayons proviennent d'un axe allongé qui est le plus souvent spiralé.

Taxon	Sous classe	Ordre	Sous ordre	Famille
Pachastrissa	Tetractinom.	Astrophorida	Euastrophor.	Calthropellidae
Erylus	Tetractinom.	Astrophorida	Euastrophor.	Geodiidae
Pachymatisma	Tetractinom.	Astrophorida	Euastrophor.	Geodiidae
Penares	Tetractinom.	Astrophorida	Euastrophor.	Ancorinidae
Stryphnus	Tetractinom.	Astrophorida	Euastrophor.	Ancorinidae
Discodermia	Tetractinom.	Astrophorida	Streptosclerop	Theonellidae
Corallistes	Tetractinom.	Astrophorida	Streptosclerop	Corallistidae
Poecillastra	Tetractinom.	Astrophorida	Streptosclerop	Pachastrellidae
Cynachyrella	Tetractinom.	Spinophorida		Tetillidae

Tableau I- 1. Classification de quelques trétraticnomorphes d'après des données morphologiques

Le séquençage de 850 bases de la partie 5' du rDNA 28S donne un arbre phylogénétique qui contredit la classification généralement admise. C. Chombard (thèse) choisit d'étudier les spicules en microscopie électronique et montre que Stryphnus possède non seulement des oxyasters (en étoile) mais aussi des sanidasters (avec un axe) de 13µm de long non identifiables en microscopie optique, ce qui lui permet de proposer une révision de la classification et une hypothèse évolutive pour les spicules dans ces familles.

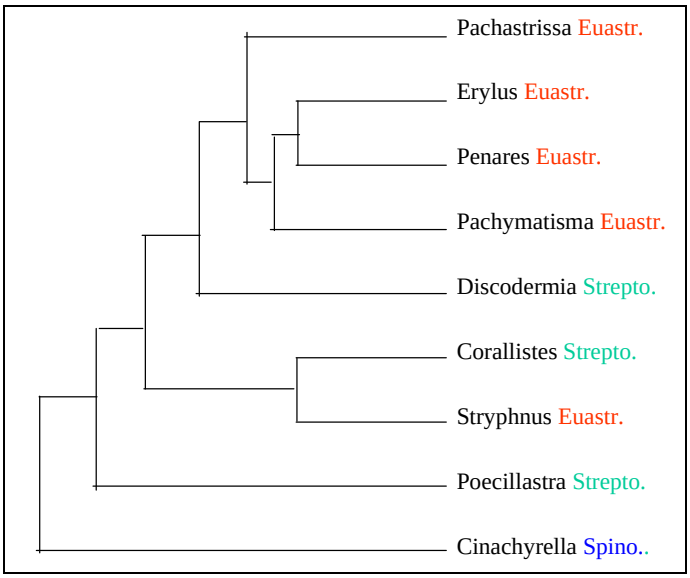


Figure I- 24. Arbre obtenu avec le domaine D2 du gène nucléaire codant l'ARN ribosomique 28S.

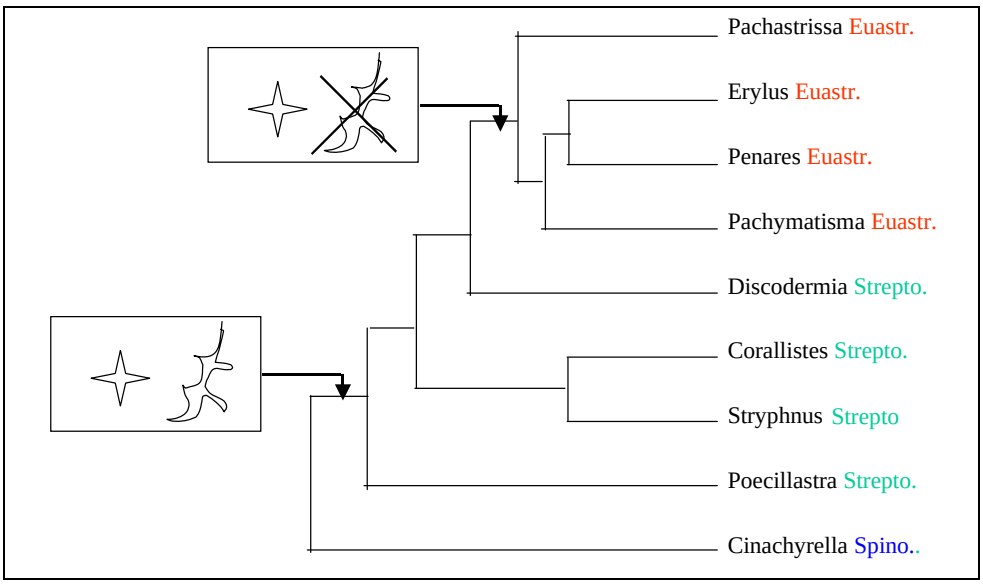


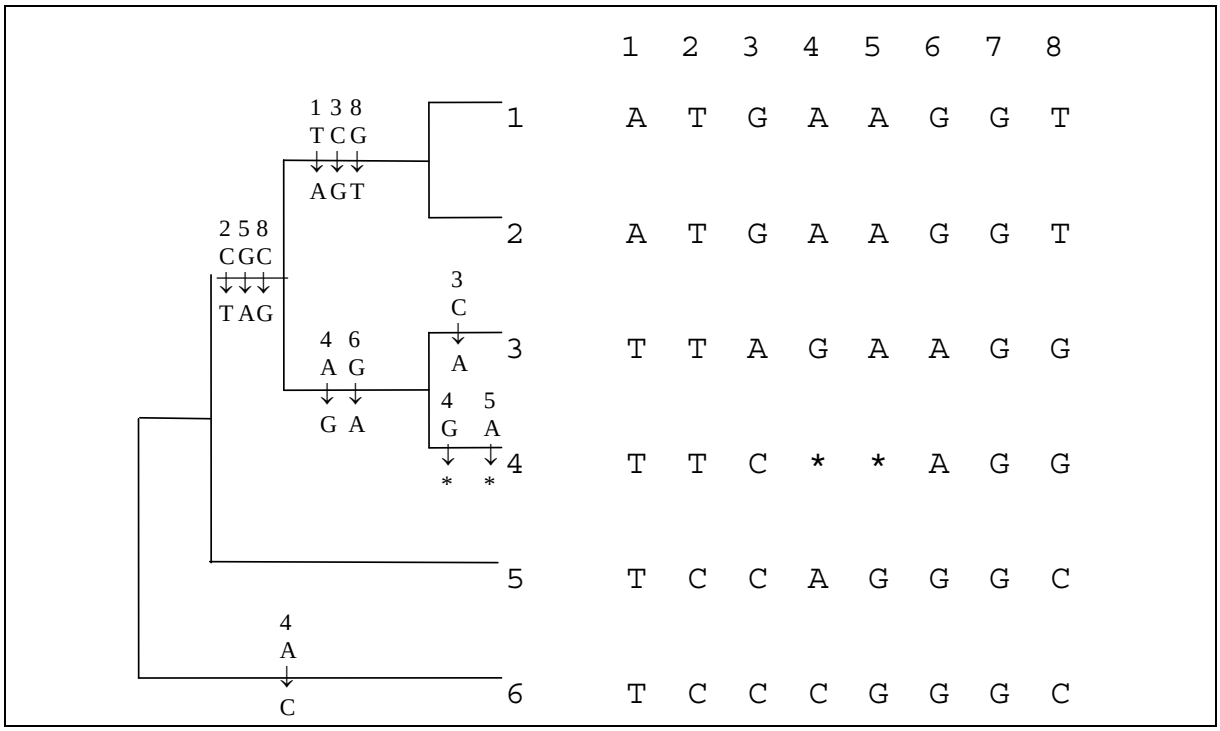
Figure I- 25.Le réexamen par microscopie électronique des spicules réconcilie les deux arbres et requalifie Stryphnus.

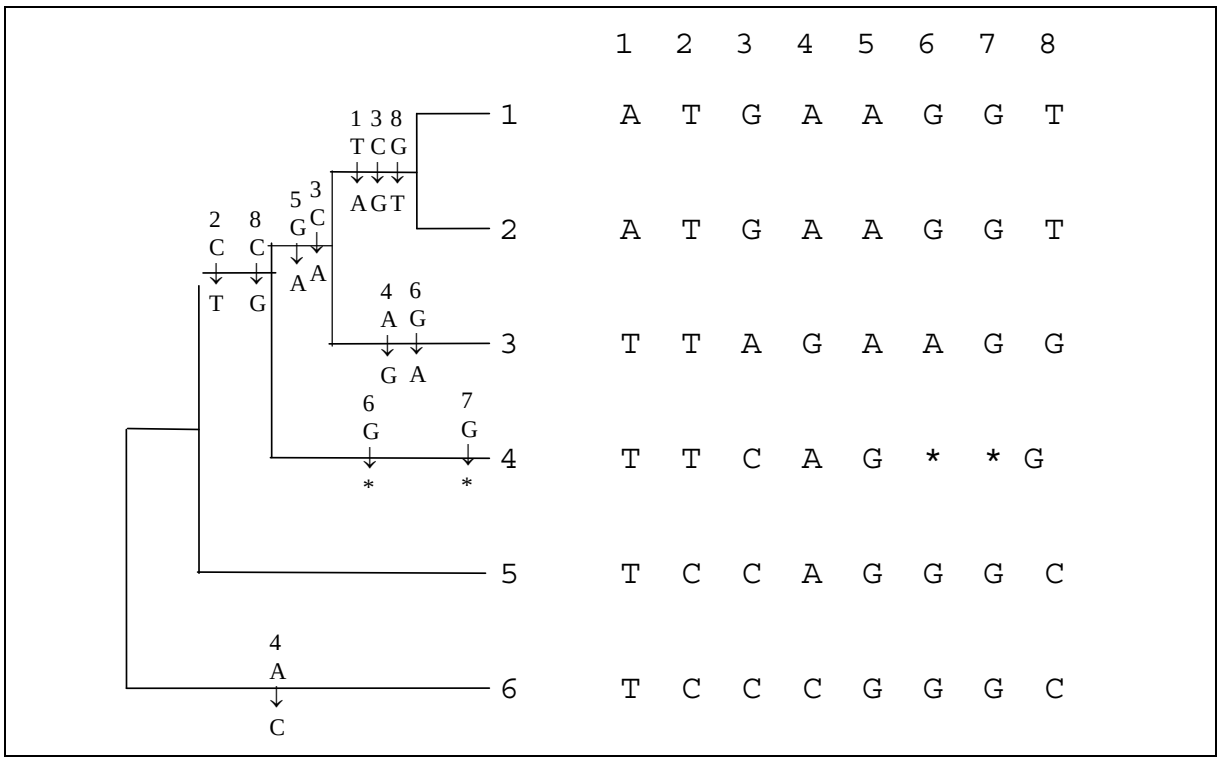
Conclusion

Si l'on sait quoi comparer (homologie) sait-on reconstruire l'histoire évolutive

- d'un ensemble de gènes
- d'un ensemble de taxons ?

Avant de chercher les logiciels qui permettront ce calcul sous diverses hypothèses évolutives il faut s'assurer que l'on comparable ce qui l'est. Avec les caractères moléculaires il faut résoudre le problème de l'alignement des séquences.


1 2 3 4 5 6 7 8
1 3 8
T C G
A G T
2 5 8
C G C
T A G
3
C
A
4 6
A G
G A
4 5
G A
* *
4
A
C
1 A T G A A G G T
2 A T G A A G G T
3 T T A G A A G G
4 T T C * * A G G
5 T C C A G G G C
6 T C C C G G G C


1 2 3 4 5 6 7 8
1 3 8
T C G
A G T
5 3
G C
A A
2
C
T
8
C
G
4 6
A G
G A
6
G
*
7
G
*
4
A
C
1 A T G A A G G T
2 A T G A A G G T
3 T T A G A A G G
4 T T C A G * * G
5 T C C A G G G C
6 T C C C G G G C

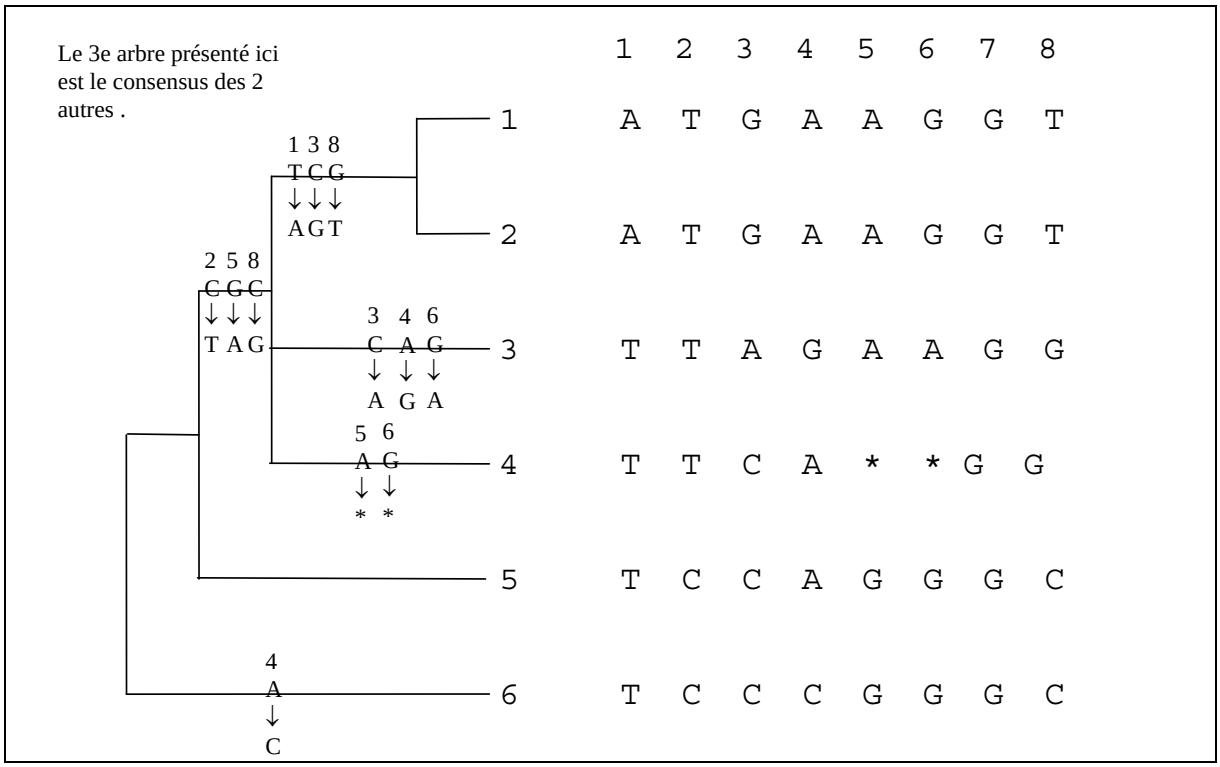


Figure I- 26. Pour ces six séquences trois alignements sont possible, générant trois arbres distinct.

Un tel exemple n'est pas rare car dans certaines séquences certaines portions s'alignent parfaitement bien mais d'autres sont beaucoup problématiques (variabilité qui diffère au long de la molécule).

Ensuite il faut choisir la méthode de reconstruction adaptée au(x) gène(s) observé(s) pour inférer les étapes ancestrales.