

SV-K-2.1 poly 02

Texte du cours

Sciences de la Vie K – évolution et phylogénie
Chapitre 2 : une approche phylogénétique de la biodiversité

(1) Classer la biodiversité

La **biodiversité** correspond à la diversité des organismes vivants, aussi bien à une échelle globale (Terre entière) qu'à une échelle plus localisée, comme au niveau d'un écosystème par exemple. Elle est donc variable d'un point de vue géographique (écosystèmes différents), tout comme d'un point de vue temporel, du fait de l'apparition, disparition et évolution des espèces.

Selon les études, on évalue la biodiversité mondiale entre 3 et 100 millions d'espèces, le chiffre de 10 millions d'espèces étant retenu comme le plus probable.

La **systématique** a pour objet l'établissement de l'inventaire des êtres vivants, en se basant sur leur description, leur nomenclature et leur taxonomie.

La **nomenclature** utilisée a été mise en place par le suédois Carl Von Linné (1707-1778) : elle est binomiale, composée d'un genre (débutant par une majuscule) et d'une espèce (écrite en minuscules) : *Homo sapiens* (espèce humaine), *Mus musculus* (la souris), *Ovis aries* (le mouton), *Capra hircus* (la chèvre), *Capra ibex* (le bouquetin des Alpes), *Zea mays* (le maïs), *Saccharomyces cerevisiae* (la levure du boulanger), etc.

La **taxonomie** (=science de la classification) a pour objet de regrouper les êtres vivants au sein de **taxons**, et d'ordonner ces taxons les uns par rapport aux autres.

Comment organiser la diversité du vivant ? Comment concilier classification et évolution ?

Rappel :

On définit comme une **espèce** l'ensemble des organismes interféconds, c'est-à-dire capables de se reproduire entre eux, dans les conditions naturelles, en donnant une descendance fertile. Pour mémoire, cette notion peut être discutée (inter-fécondabilité au sein du genre *Canis*, « espèces » bactériennes définies en absence de processus de reproduction sexuée...).

I. Une diversité de classifications

Le type de classification produit dépend de l'objectif de cette classification. Ainsi, il n'est pas étonnant de trouver des classifications différentes selon l'objet de cette classification ou le public l'utilisant.

A. Classer et trier pour ordonner le vivant

On peut distinguer classer et trier :

- Une **classification** a pour objectif la définition de taxons, et fait donc partie d'un processus de systématique.
- Un **tri** correspond uniquement en un regroupement d'organismes selon divers critères, afin d'aboutir, par exemple, à leur détermination. Ce tri n'a pas d'obligation de suivre les règles d'une classification. *Exemple : nombre de flores permettent de trier les fleurs en fonction de leur couleur, du nombre de pétales, etc., afin de déterminer le nom de l'espèce.*

1. Historiquement : différentes approches scientifiques

Les premières classifications, effectuée pour les plantes, sont des classifications utilitaires : plantes aromatiques, alimentaires, médicinales, vénéneuses, venimeuses.

Vers 1550, classification par ordre alphabétique ! Aucune utilité...

Par la suite, de nombreux essais de classification avec tous les critères possibles selon 2 méthodes :

- **Partage** selon des **critères prédéfinis** (a priori), classification divisive, par dichotomie, avec souvent des définitions négative (absence de ...).
- **Agglomération** sur **critères de similarité**, basés sur l'observation.

Carl von Linné (1707-1778) utilise un mélange des deux méthodes et codifie la hiérarchisation des espèces en 7 niveaux taxonomiques :

Règne, Embranchement, Classe, Ordre, Famille, Genre, Espèce.

Toutefois ce système posait plusieurs types de problèmes :

- On ne retombe pas forcément sur le même nombre de niveaux, d'où : sous-ordre, super-classe, etc.
- Critères choisis arbitrairement.
- L'ordre d'utilisation des critères change la classification...

Plusieurs systèmes ont été proposés à la suite de Linné, avec l'idée d'aboutir à une classification « naturelle », qui ne serait pas qu'un simple tri des espèces.

2. Une diversité d'objectifs

On peut réaliser plusieurs classifications différentes, en fonction de l'objectif recherché.

a. Classification utilitaire

Classement des espèces en fonction d'une **utilisation** par l'Homme.

Par exemple, les classifications des plantes à l'antiquité et au moyen-âge avaient pour but de regrouper les plantes par type d'utilisation médicinale.

b. Classification fonctionnelle

On utilise des **caractéristiques fonctionnelles** des organismes pour les regrouper. Ceci peut concerner :

- Des traits physiologiques : type de nutrition par exemple.
- Des traits écologiques (correspondant aux interactions entre l'organisme et son écosystème).

3. Une diversité de caractères utilisables

Une classification se construit sur la base de la comparaison de caractères présentés par diverses espèces. Les caractères en question peuvent être de diverses natures, une classification pouvant utiliser différents types de caractères.

Exemples de caractères utilisables :

- Caractères **fonctionnels** : type de respiration, mode déplacement, etc.
- Caractères **morphologiques** : nombre de membres et structure générale, axes de polarités, etc.
- Caractères **anatomiques** : organisation du système nerveux central, structure de la pompe sanguine, etc.
- Caractères **embryologiques** (chez les pluricellulaires) : modalités de la segmentation (divisions cellulaires au début du développement), de la gastrulation chez les animaux (mouvements mettant en place le tube digestif primitif), etc.
- Caractères **biochimiques** : type de molécules de réserves, voies métaboliques de détoxification des déchets azotés, présence/absence d'une molécule donnée, etc.
- Caractères **moléculaires** : séquences d'ADN, de protéines, etc.

4. Classifications et ressemblances

Une classification repose donc sur le regroupement d'organismes possédant un même caractère, c'est-à-dire un attribut observable présentant le même état. Pour être utilisable dans une classification, un caractère doit donc exister sous plusieurs états.

Lorsque deux caractères sont identiques (= même état) entre deux espèces, on parle **d'analogie**.

L'analogie correspond donc à une ressemblance entre deux individus. Toutefois, cette ressemblance peut avoir deux origines d'un point de vue évolutif.

a. Homologies

Elle peut être due au fait que le caractère étudié a été **hérité** d'un ancêtre commun. Un même caractère, présent chez deux espèces car hérité d'un caractère présent chez un ancêtre commun à ces deux espèces, est dit **homologue**.

Un caractère homologue peut se présenter sous deux états :

- Un **état ancestral** = **plésiomorphe**, qui correspond à celui présent chez l'ancêtre commun.
- Un **état dérivé** = **apomorphe**, différent de celui présent chez l'ancêtre commun, et qui correspond donc à l'acquisition d'une innovation évolutive.

Un caractère homologue, à l'état **dérivé**, partagé entre deux espèces est dit **synapomorphique**.

Un caractère homologue, à l'état **ancestral**, partagé entre deux espèces est dit **symplesiomorphique**.

b. Homoplasies

Deux structures similaires, analogues, peuvent ne pas dériver d'une même structure ancestrale. On parle alors d'**homoplasie**.

Cette homoplasie peut avoir différentes origines :

- Une **convergence** évolutive : deux structures non homologues ont évolué vers des formes et fonctions similaires, permettant une réponse similaire à une même contrainte évolutive. Ex : l'aile des chiroptères vs. l'aile des oiseaux.
- Une **réversion** : retour à l'état ancestral (ancestral → dérivé → ancestral). *Ceci concerne essentiellement des caractères moléculaires ou biochimiques.*
- Un **parallélisme** : évolution similaire et contemporaine de deux espèces proches, suite à une même pression évolutive. Ex : taupe vs. rat-taupe.

B. Classifications phénétiques et probabilistes

1. Mesure de la ressemblance : classifications phénétiques

Les classifications phénétiques se basent sur une ressemblance globale : On calcule un indice de similitude globale entre les taxons, c'est-à-dire une distance pour chaque couple de taxons.

Par exemple, dans le cas de l'utilisation de séquences de gènes, on procède à un alignement de ces séquences, puis on détermine le pourcentage de différences entre les espèces prises deux à deux. Ceci permet de construire une matrice des distances, qui est ensuite traitée pour produire un arbre phénétique. La longueur des traits reliant les espèces est proportionnelle à la distance génétique calculée. *Il est possible de corriger les distances calculées pour prendre en compte des différences de vitesse d'évolution.*

Ex : la mise en évidence de trois domaines du vivant par comparaison de la séquence des ARNr 18S par Woese.

2. Méthodes probabilistes : maximum de vraisemblance

Ces méthodes établissent un modèle constitué d'un ensemble de paramètres dont le réglage formule différentes hypothèses. Ces hypothèses concernent surtout l'évolution des états de caractères et sont exprimées en termes de probabilités ;

ex : probabilités lors d'une mutation d'une Adénine

$$A \rightarrow C \quad p = 0,3 \cdot 10^{-9}$$

$$A \rightarrow G \quad p = 0,5 \cdot 10^{-9}$$

$$A \rightarrow T \quad p = 0,2 \cdot 10^{-9}$$

Toutes les probabilités associées aux transformations impliquées par un arbre donné vont se multiplier et fournir une valeur globale de **vraisemblance** des données associées à cet arbre.

Parmi les arbres possibles, l'arbre choisi est celui dont la vraisemblance des données au vu du modèle est maximale.

Cette méthode fonctionne sur les caractères moléculaires, pour lesquels on peut établir des modèles d'évolution des protéines ou des acides nucléiques.

Pour les caractères morphologiques, cette méthode est très difficile à appliquer, car, compte tenu de leur nature hautement intégrée, on n'a aucune idée des probabilités de transformation d'un état à un autre, ou plus généralement des modèles à construire.

C. Classifications post-Darwiniennes

Suite aux publications de Darwin, il est apparu qu'une classification « naturelle » est une classification reflétant les parentés évolutives, et dont l'interprétation permet donc de mener une réflexion sur l'évolution des espèces.

1. Classifications évolutionnistes

Les classifications évolutionnistes se basent sur un regroupement des espèces partageant les mêmes caractères homologues.

Ce type de classification **n'est plus utilisé**, car il acceptait un regroupement sur la base de symplésiomorphies, ce qui conduit à des groupes paraphylétiques (cf. 3.).

2. Classifications phylogénétiques

En dehors des situations où une classification utilitaire ou fonctionnelle est d'intérêt, les classifications actuelles sont des classifications phylogénétiques = **cladistiques**. Les taxons déterminés dans une classification phylogénétique sont des **clades**.

Afin de rechercher les parentés entre espèces, on les regroupe sur la base de synapomorphies, ce qui signifie donc que les caractères utiles sont :

- Homologues
- A l'état dérivé (apomorphes)

Les fondements de la cladistique ont été posés par Willi Hennig en 1966.

3. Différents types de groupes selon les classifications réalisées

Une classification phylogénétique cherche à produire des groupes **monophylétiques**. Un groupe monophylétique est construit en utilisant une **synapomorphie** : il regroupe un ancêtre commun et l'intégralité de ses descendants, qui partagent tous un même caractère homologue à l'état dérivé.

D'autres classifications, ainsi que les habitudes de langage, utilisent des groupes qui ne sont pas monophylétiques : ces regroupements ne peuvent être concernés comme valides d'un point de vue phylogénétique, et les taxons en question ne sont donc pas des clades.

Il peut s'agir :

- De groupe **paraphylétiques** : construits en regroupant des espèces sur la base d'une symplésiomorphie, donc en utilisant l'état **ancestral** d'un caractère homologue, et donc une **symplésiomorphie**. Un tel groupe n'est pas valide phylogénétiquement parlant, car excluant une partie des descendants de l'ancêtre commun au groupe. Ex : *poissons*, *reptiles*.

- De groupes **polyphylétiques** : construits en regroupant des espèces sur la base de caractères analogues et non homologues. Un tel groupe n'est pas valide phylogénétiquement parlant, car sans relation avec les parentés évolutives, mais basé sur une **homoplasie**. Ex : animaux « à sang chaud », pachydermes, mammifères marins.

II. L'élaboration de phylogénies

L'analyse cladistique (= systématique phylogénétique) vise à reconstruire la phylogénie d'un taxon par distinction, au sein d'un caractère, de l'état primitif (= plésiomorphe) de l'état dérivé (= apomorphe).

Chercher la parenté c'est chercher le groupe frère et non l'ancêtre. Il s'agit de chercher, pour une espèce ou un groupe donné, le groupe avec lequel il partage un ancêtre exclusif, qu'il ne partage avec aucun autre groupe de l'échantillon.

Les nœuds de l'arbre obtenu correspondent à des ancêtres communs, qui ont une valeur théorique : Ils ne correspondent pas à un organisme actuel et il n'est pas possible de lui attribuer une espèce fossile, faute de données suffisantes.

La construction de l'arbre conduit toutefois à déduire certains des caractères que devait posséder cet ancêtre à partir des caractères dérivés qui définissent le groupe (synapomorphies).

A. Réalisation d'une matrice de caractères

La reconstitution des liens de parenté ne résulte pas d'une découverte fortuite ou révélée mais de décisions préalables réfléchies.

La première étape est le choix du groupe à étudier, dans lequel on choisit un ou plusieurs échantillons. La phylogénie construite dépend donc, dans une certaine mesure, des choix d'échantillonnage qui sont réalisés...

Une fois l'échantillon choisi, on procède au choix de caractères, qui vont permettre la comparaison. On note dans un tableau (la matrice de caractères) l'état des différents caractères.

Les caractères doivent être homologues, hérités d'un ancêtre commun. On fait une hypothèse d'homologie primaire soutenue par des indices de similitude de plan d'organisation, de même origine embryologique ou des études paléontologiques. L'analyse cladistique permettra de confirmer cette hypothèse (homologie secondaire) ou de la réfuter (homoplasie).

A ce stade, les caractères ne sont pas forcément polarisés, c'est-à-dire que l'on ne sait pas *a priori* l'état ancestral (plésiomorphe) ou dérivé (apomorphe) des caractères.

B. L'identification des états plésiomorphes et apomorphes

Plusieurs méthodes existent, la principale étant l'utilisation d'un groupe externe, ou extragroupe.

Le caractère ancestral ou dérivé d'un caractère peut dépendre de l'échantillonnage réalisé.

1. Grâce à un extragroupe

La méthode consiste à trouver une espèce (ou un groupe d'espèces) dont on soit **sûr** qu'elle est **extérieure** au groupe d'étude. La définition de l'extragroupe fait partie des **hypothèses de départ**.

Par définition, l'extragroupe possède l'état **ancestral** pour tous les caractères étudiés (noté 0). Lorsqu'une espèce du groupe d'étude possède le même état de caractère que l'extragroupe, l'état de ce caractère sera considéré comme ancestral et codé 0, lorsqu'il possède un état de caractère différent il sera considéré comme dérivé, il sera noté 1.

2. Grâce à d'autres approches

Le critère ontogénétique : il est fondé sur la « loi » de Haeckel (1834-1919) selon laquelle « l'ontogénie récapitule la phylogénie » dans la mesure où, lors d'un développement embryonnaire, les caractères généraux apparaissent avant les caractères plus particuliers à l'espèce. S'il est possible d'observer, chez une espèce du groupe d'étude, le développement du caractère que l'on veut polariser, on cherchera à déterminer lequel des deux états apparaît après l'autre. Celui qui apparaît le premier sera considéré comme l'état ancestral, puis, si ce dernier se modifie, l'état modifié sera l'état dérivé.

Par exemple, les Mammifères développent des arcs branchiaux à l'état embryonnaire : ces arcs sont les homologues des arcs branchiaux des Téléostéens (« poissons »), qui soutiennent leurs branchies. Chez les Mammifères, ces arcs donnent d'autres structures au cours du développement, par exemple au niveau de la face et du cou. Les arcs « réellement branchiaux » des Téléostéens correspondent donc à l'état ancestral, alors que les structures observées chez les Mammifères sont des états dérivés.

Il est à noter que ce critère n'est pas toujours utilisable, car la « loi » de Haeckel n'est pas observée de manière systématique.

L'utilisation de **fossiles** permet aussi d'émettre des hypothèses sur les caractères ancestraux, en mettant en évidence l'apparition de formes dérivées, absentes chez les fossiles plus anciens.

3. Détermination a posteriori

Dans la majorité des situations où un extragroupe n'est pas utilisé (souvent car cet extragroupe n'est pas définissable, ou trop différent des espèces étudiées pour être utile), un arbre peut être réalisé sur la base du maximum de probabilité, en partant d'un modèle évolutif postulé (par exemple grâce aux études de fossiles).

Dans ce cas de figure, c'est l'arbre réalisé qui permet de polariser les caractères et donc de déterminer leur état ancestral ou dérivé : on se base pour cela sur la notion de descendance avec modification. En effet, la polarisation des caractères signifie que l'état dérivé est le produit d'une transformation de l'état ancestral et que ce produit a été transmis de génération en génération.

C. Détermination de l'arbre le plus probable

1. Regroupement et caractères

On regroupe les espèces sur la base de caractères partagés à l'état dérivé (**synapomorphies**).

Il est à noter que cette démarche suppose initialement que tous les caractères ont la même valeur. Or, certains caractères ne sont pas utilisables, car pas **informatifs** : c'est le cas d'un caractère partagé par toutes les espèces de l'échantillon...

Les regroupements réalisés conduisent à produire **plusieurs arbres**, qui correspondent à des scénarios évolutifs différents : les caractères semblent donc se « contredire », ce qui correspond à la possibilité d'homoplasies (convergences, réversions...).

Il est donc nécessaire de choisir entre les différents arbres produits.

2. Maximum de vraisemblance

Lorsque les différentes innovations évolutives sont affectées d'une probabilité, il est possible de calculer une probabilité totale pour chaque arbre, et ainsi de déterminer un maximum de vraisemblance.

3. Principe de parcimonie

Chaque arbre produit correspond à un scénario évolutif différent. Afin de choisir entre ces différentes possibilités, on place sur chaque arbre les pas évolutifs correspondants : apparition d'un caractère dérivé (apomorphie), convergence (homoplasie), réversion(homoplasie).

On peut alors compter le nombre de pas évolutifs correspondant à chaque arbre : l'arbre considéré comme le plus vraisemblable est celui présentant le moins de pas évolutifs : il s'agit du principe de parcimonie.

Remarques :

★ Le principe de parcimonie peut être vu comme un cas particulier de la méthode de maximum de vraisemblance avec la même probabilité pour chaque pas évolutif.

★ Le principe de parcimonie est une application à la phylogénétique du rasoir d'Occam, principe posant que la solution la plus probable est celle comportant le moins d'hypothèses.

L'analyse des arbres phylogénétiques et phénétiques permet de proposer des scénarios évolutifs à l'histoire du vivant : voir BCPST 2 (SV-K-2.2).