

**TP SV
K**
PHYLOGENIES

COURS : SV-K-2.1 + COURS « EVOLUTION » EN BCPST2



L'objectif des classifications actuelles est d'aboutir à un classement des espèces reproduisant leur histoire évolutive, et ainsi de produire des groupes monophylétiques. Ces classifications peuvent être produites par une démarche cladistique, basée sur la comparaison de caractères, ou une démarche phénétique, basée par exemple sur des comparaisons de séquences.

L'objectif de ce TP est de mettre en œuvre les différentes techniques conduisant à des classifications phylogénétiques, et d'exploiter de manière critique les arbres obtenus.

Programme officiel :

Construire une phylogénie par parcimonie (cladogramme)
Réaliser et exploiter des alignements de séquences afin de construire un arbre phylogénétique

Compétences :

Exploiter des données morpho-anatomiques ou moléculaires pour positionner un organisme dans un arbre phylogénétique ou construire une phylogénie :

- Construction d'une matrice taxons-caractères ou d'une matrice de distances
- Polarisation des caractères
- Construction d'arbres phylogénétiques et application du principe de parcimonie

Réalisation et exploitation d'une comparaison de séquences à l'aide de logiciels

1. Séquences et arbres phylogénétiques

1.1. Logiciels utilisables

Voir poly 1 – préparation

1.2. Comparaison de séquences avec Anagène 2 et construction de l'arbre

Voir en annexe 1 la fiche-méthode sur Anagène 2.

➔ **Ouvrir** Anagène 2, puis ouvrir la banque de séquences « gène CDC2 » (*Terminale S-2000 > parenté entre les êtres vivants actuels et fossiles > relations de parenté entre les êtres vivants > eucaryotes*).

➔ **Sélectionner** l'intégralité des séquences d'ADN proposées.

➔ **Aligner** les séquences et observer les différences.

➔ **Construire** la matrice de distances pour 6 espèces au choix (*il faut pour cela réaliser plusieurs alignements, en prenant chaque fois une séquence différente en référence – la séquence en référence est la première dans la liste des séquences*).

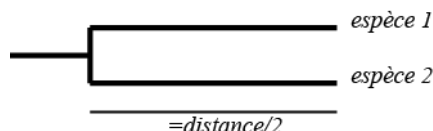
Plusieurs techniques de construction d'un arbre existent, la plus simple est celle de l'UPGMA (*Unweighted Pair Group Method using Averages*), qui trouve un seul arbre à partir d'une matrice, par agglomération successive des espèces, des plus proches aux plus éloignées. L'avantage est que l'arbre peut se réaliser à la main, mais l'inconvénient est que l'on fait l'hypothèse que les séquences évoluent à la même vitesse dans toutes les branches de l'arbre... Si ce n'est pas le cas, l'arbre est faux... Cette technique est ainsi de plus en plus rarement utilisée.

Méthode : réalisation d'un phénogramme par UPGMA

1- On part des deux espèces les plus proches (faible distance)

2- On calcule : $\frac{\text{distance entre les 2 espèces}}{2} = \text{longueur des branches horizontales reliant les 2 espèces}$

3- On trace ces deux premières branches :



4- On regroupe les deux espèces comme une seule entité, et on produit une nouvelle matrice : les nouvelles distances sont les moyennes des distances des autres espèces par rapport aux deux premières espèces rapprochées

5- On recommence le processus jusqu'à ce que toutes les espèces soient intégrés au phénogramme

➡ **Réaliser le phénogramme** des Eucaryotes (6 espèces) basé sur le gène CDC2.

➡ **Réaliser** un alignement des séquences protéiques, et comparer les valeurs obtenues par rapport à celles de l'alignement nucléotidique : comment expliquer la présence de différences ?

1.3. Comparaisons de séquences avec Genieen 2 et exploitation des phénogrammes

Le logiciel Genieen 2 permet d'obtenir directement la matrice des distances ainsi que le phénogramme à partir d'un alignement.

On travaille dans un premier temps avec le gène codant la prolactine, une hormone impliquée en particulier (mais pas seulement) dans la lactation.

➡ **Ouvrir** la banque de séquences « gène de la prolactine chez différents Vertébrés » dans Genieen 2, puis **aligner** les gènes.

➡ **Afficher** la matrice de caractères puis le phénogramme proposé : **vérifier** la concordance entre la matrice et le phénogramme.

On travaille dans un deuxième temps sur l'hexokinase :

- Cette enzyme catalyse la phosphorylation du glucose, qui correspond à la première réaction de la glycolyse
Glucose + ATP = glucose-6-phosphate + ADP
- Les séquences utilisées ici sont des séquences « brutes », directement extraites des bases de données, sans avoir été retravaillées (au contraire des banques de séquences livrées avec les différents logiciels)
- Les séquences correspondent à des séquences d'ARNm (retranscrites en ADN)

➡ **Ouvrir** le fichier **hexokinase.edi** dans Genieen 2 et **aligner** les séquences.

➡ **Afficher** la matrice de caractères puis le phénogramme : **Discuter** le phénogramme obtenu.

2. Réalisation de cladogrammes

2.1. Principe de réalisation d'un cladogramme

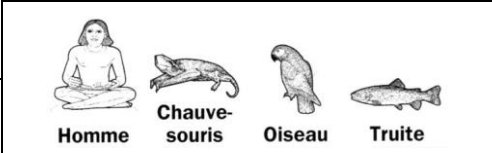
Voir poly 1 – préparation

2.2. Logiciels utilisables

Voir poly 1 – préparation

2.3. Réalisation d'arbres et choix de l'arbre le plus vraisemblable

On souhaite classer les 4 espèces suivantes :

Homme		Molosse de Cestoni	
Primate – <i>Homo sapiens</i>	 Homme Chauve-souris Oiseau Truite	Chiroptère – <i>Tadarida teniotis</i>	
Gris du Gabon ou Perroquet jaco		Truite d'Europe	
Oiseau - <i>Psittacus erithacus</i>			Téléostéen – <i>Salmo trutta</i>

➡ Quelle espèce choisir comme **extragroupe** ?

Pour comparer les espèces de notre échantillon, il faut choisir un ensemble de caractères sur lesquels on réalise la comparaison.

Pour notre exemple 6 caractères :

1. Mâchoire : deux états, présence absence
2. Type d'appendices pairs : deux états, nageoire rayonnée, membre chiridien
3. Dents : deux états, présence, absence

4. Constitution de la mandibule : deux états, un seul os, le dentaire, plusieurs os
5. Réserves vitellines de l'œuf : deux états, œuf alécithe (avec très peu de réserve vitelline) œuf télolécithe (avec de grandes quantités de réserves vitellines)
6. Appendices pairs antérieurs : deux états, transformés en ailes, absence d'aile

➡ **Identifier** l'état de chaque caractère dans le tableau suivant.

	Homme	Chauve-Souris	Oiseau	Truite
1. Mâchoire				
2. Membres				
3. Dents				
4. Constitution de la mandibule				
5. Réserves vitellines de l'œuf				
6. Ailes				

➡ A partir de l'extragroupe, **polariser** l'état de chaque caractère dans le tableau suivant.

(1 = dérivé, apomorphe /// 0 = ancestral, plésiomorphe)

	Homme	Chauve-Souris	Oiseau	Truite
1. Mâchoire				
2. Membres				
3. Dents				
4. Constitution de la mandibule				
5. Réserves vitellines de l'œuf				
6. Ailes				

Certains caractères peuvent se contredire, du fait d'homoplasies (convergences ou réversions). Concrètement, il est ici possible de construire 6 arbres : 3 arbres basés sur des convergences, et 3 arbres basés sur des réversions.

➡ **Construire** et représenter les 3 arbres possibles basés sur les convergences (*si vous avez le temps : construire les 3 arbres basés sur des réversions*).

➡ **Indiquer** sur chaque arbre :

- L'apparition d'état dérivés (apomorphies) par un rond noir ●
- Les convergences par un trait noir —
- Les réversions par un rond blanc ○
- Le nombre total de pas évolutifs ainsi mis en évidence

(ces symboles n'ont aucune valeur « générale » et sont définis ici de manière arbitraire)

➡ Sur la base du principe de parcimonie, **choisir** l'arbre le plus vraisemblable.

2.4. Réalisation d'un cladogramme avec Phylogène

Voir poly 1 – préparation

2.5. Réalisation d'un cladogramme à partir d'un alignement de séquences

Il est possible de réaliser un cladogramme à partir d'un alignement de séquences nucléotidiques : pour cela il faut considérer chaque nucléotide comme un caractère, et le nucléotide précis comme l'état du caractère.

On se propose ici d'aborder de manière simplifiée cette méthodologie, à partir de courtes séquences alignées dans Anagène.

➡ **Ouvrir** Anagène 2, puis ouvrir la banque de séquences « B7 - Ubx » (*Terminale S-2000 > parenté entre les êtres vivants actuels et fossiles > relations de parenté entre les êtres vivants > eucaryotes > gènes homéotiques*).

On étudie ici les gènes Hox B7 de Vertébrés, ainsi que leur homologue Ultrabithorax de la Drosophile. Ces gènes font partie du complexe homéotique, ensemble de gènes impliqués dans l'identité des cellules le long de l'axe antéro-postérieur de l'embryon.

➡ **Aligner** les quatre séquences d'ADN de la banque de séquences.

➡ On choisit la Drosophile comme extragroupe : pour chaque position, le nucléotide présent chez la Drosophile est considéré comme représentant l'état ancestral. **Trouver** 4 nucléotides que l'on peut considérer comme informatif (= permettant un regroupement d'espèces sur la base d'une synapomorphie), puis **construire** la matrice de caractères polarisée, en identifiant pour chaque caractère (= nucléotide) son état (0 = ancestral ; 1,2,3 = dérivé).

➡ **Construire** le cladogramme correspondant. **Discuter** l'arbre obtenu.

3. Quelques exercices supplémentaires... en fonction du temps disponible...

Exo 1 : Réaliser un **cladogramme**, dans **Phylogène**, à partir de la collection « **végétaux** – seconde », en prenant au minimum 5 espèces et 6 caractères.

Exo 2 : Réaliser un **phénogramme** à partir de l'alignement, dans Anagène, de 4 séquences protéiques de **gènes homéotiques**.

Exo 3 : Un arbre **au choix**, par méthodologie au choix ☺

ANNEXE : MODE D'EMPLOI SIMPLIFIE D'ANAGENE

Anagène est un logiciel payant, permettant l'affichage, la comparaison et le traitement de séquences moléculaires (ADN, ARN et Protéines). La présente fiche est basée sur les versions 1.3 et 1.8 du logiciel.

La barre d'outil : description



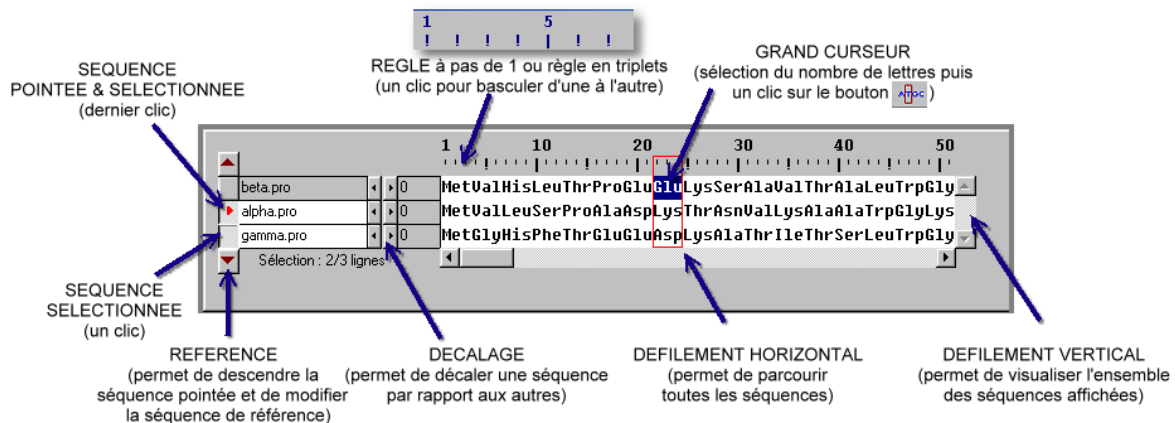
Accès aux ressources		Sauvegarde		Edition des sélections		Traitements des séquences		Autres	
	Banque de séquences		Accès au classeur		Couper		Convertir (transcription, traduction)		Information sur l'objet pointé
	Thèmes d'études		Enregistrer		Copier		Comparer		Code génétique
	Programmes & documents		Imprimer		Coller		Enzymes de restriction		Grand curseur
					Supprimer				Fermeture de toutes les fenêtres

Affichage de séquences – base de donnée du logiciel



Le logiciel dispose dans sa banque de séquence de nombreuses séquences. Il s'agit de séquences d'ADN (.adn), d'ARN (.arn), ou de protéines (.pro). Les thèmes d'étude regroupent plusieurs séquences et documents en un même fichier. Il est possible de sélectionner plusieurs séquences en même temps, en utilisant les touches CTRL et MAJ.

L'affichage



Comparaison de séquences – alignements → TP K1



En cliquant sur les carrés à gauche des séquences, il est possible de les sélectionner. La dernière séquence sélectionnée est celle « pointée ». C'est à cette séquence que seront ensuite comparées toutes les autres séquences sélectionnées.

Le bouton de comparaison permet ensuite de comparer les séquences sélectionnées. La comparaison simple est la plus rapide.

La comparaison avec discontinuité est plus longue, mais souvent plus efficace. Elle permet de gérer les séquences ne débutant pas toute au même niveau, par exemple.

Résultats : - et * : identité _ : manque (rien au niveau de la séquence comparée)

Le résultat de comparaison est affiché dans une nouvelle fenêtre.

Traiter des séquences – code génétique



Anagène permet de convertir toutes les séquences entre ADN, ARN et protéines

Traduction : Soit traduction « simple », normale, Soit « phases ouvertes de lecture », qui recherche les gènes possibles dans la séquence (recherche des codons d'initiation et des codons STOP)

Le logiciel peut aussi afficher le tableau du code génétique.

Digestion par des enzymes de restriction



Sélectionner une ou plusieurs séquences d'ADN. Puis cliquer sur le bouton, et sélectionner la ou les enzymes à utiliser. La représentation graphique indique les sites de restriction (en rouge).

IMPORTANT !



C'est LE bouton le plus important. Il permet d'obtenir des informations sur la séquence pointée. Après une comparaison de séquences, il permet d'afficher le bilan de la comparaison : nombre et pourcentage de différence, etc (penser à sélectionner toutes les séquences pour avoir un bilan complet ; attention, la comparaison est toujours faite par rapport à la première séquence).

Après digestion enzymatique, il permet d'avoir aussi les informations sur le résultat du traitement effectué.