

**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

Plan du TP :**1. Séquences et arbres phylogénétiques**

1.1. Logiciels utilisables ←

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

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

2. Réalisation de cladogrammes

2.1. Principe de réalisation d'un cladogramme ←

2.2. Logiciels utilisables ←

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

2.4. Réalisation d'un cladogramme avec Phylogène ←

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

3. Quelques exercices supplémentaires... en fonction du temps disponible...**Travail préparatoire :****1. Séquences et arbres phylogénétiques**

1.1. Logiciels utilisables

La comparaison de séquences nécessite l'utilisation de logiciels afin de réaliser les alignements nécessaires. Le rapport de l'épreuve de TP du concours agro-véto (version 2019) indique comme logiciel utilisable Anagène. Il est donc nécessaire de savoir manipuler ce logiciel.

Il est aussi possible de s'entraîner avec le logiciel en ligne Genieen 2, accessible gratuitement, et qui pourrait faire son apparition au concours agro-véto.

Ces logiciels sont basés sur l'algorithme BLAST (*Basic Local Alignment Sequence Tool*) ; il est possible d'accéder directement aux différents outils BLAST en ligne sur le site du NCBI (*National Center for Biotechnology Information*)(en anglais).

Logiciel	Type	Licence	Page Internet
Anagène 2	Exécutable PC	Gratuit	http://acces.ens-lyon.fr/acces/thematiques/evolution/logiciels/anagene/telechargement-eleves-maj-2018
Anagène 2 en ligne	En ligne	Payant avec abonnement	https://anagene.reseau-canope.fr/#/
Genieen 2	En ligne et Exécutable et Android	Gratuit	https://cosphilog.fr/genieen2/ --- https://play.google.com/store/apps/details?id=fr.cosphilog.genieen2
Blast	En ligne	Gratuit	https://blast.ncbi.nlm.nih.gov/Blast.cgi

➤ Prise (ou reprise) en main d'un logiciel de traitement de séquences : Genieen 2

- **Ouvrir** le logiciel en ligne Genieen 2 (*voir lien ci-dessus*).
 - Un **mode d'emploi simplifié** est à la fin de ce polycopié.
 - (1) Choisir « ouvrir la banque de séquences », puis cliquer sur « Syncytine 2 chez différents primates » ; valider en cliquant sur « Charger ces séquences ».
 - (2) Observer les séquences : Ce sont des séquences protéiques (= de protéines), donc des successions d'acides aminés. Chaque acide aminé est identifié par un code « 3 lettres » (*voir cours de Biochimie pour la diversité des acides aminés*) ; par exemple : Met = méthionine, Gly = glycine, Pro = proline...
 - (3) En passant par Fichier / Ouvrir la banque de séquences, charger le pack de séquences « Prédisposition au cancer du sein ».
 - (4) Observer les séquences : ce sont des séquences nucléotidiques, c'est-à-dire des successions de nucléotides, identifiés par leurs bases azotées. A = adénine, C = cytosine, G = guanine, T = thymine (dans l'ADN), U = uracile (dans l'ARN). *Les noms des 5 bases azotées sont à **apprendre** et à savoir orthographier !*
 - (5) Justifier qu'il s'agit ici de séquences d'ADN.
 - (6) A l'aide de la fiche technique (ou en fonctionnement de manière intuitive), convertir une de ces séquences ADN en une séquence d'ARNm (ARN messenger ; il s'agit donc de faire une **transcription**). Convertir une de ces séquences ADN en séquence protéique (il s'agit donc de faire une **traduction**).
 - (7) S'entraîner à manipuler les séquences, en suivant les catégories du mode d'emploi simplifié :
- Charger/ouvrir des séquences – aligner des séquences – comparer des séquences – afficher le tableau de comparaison - afficher l'arbre matérialisant le degré de similarité (phénogramme) – traiter des séquences – se déplacer et se repérer au sein d'une séquence

L'action des enzymes de restriction n'est pas utile pour ce TP.

2. Réalisation de cladogrammes

2.1. Principe de réalisation d'un cladogramme

La phylogénie vise à regrouper les espèces ou groupes en fonction de leurs **relations de parentés** : 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 ainsi 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.

Un cladogramme est un arbre phylogénétique construit sur la base de la comparaison de **caractères homologues** ; on ne peut ainsi comparer entre eux que des caractères dérivant d'un même caractère ancestral (homologie, et non analogie). Afin d'obtenir des groupes monophylétiques (*voir le cours*), on regroupe les espèces ou groupes partageant des caractères homologues à l'état **dérivé** (synapomorphie).

Cette démarche nécessite de procéder en plusieurs temps :

- 1- Choisir les caractères pris en compte.
- 2- Identifier l'état descriptif de chaque caractère, pour chaque espèce : construction de la **matrice de caractères** (tableau à double entrée : espèces / caractères).
- 3- **Polariser** la matrice de caractères : indication de l'état ancestral (codage : 0) ou dérivé (codage : 1 *et autres nombres s'il y a plusieurs états dérivés possibles*) de chaque caractère.
- 4- Construction de **l'arbre** (ou de **plusieurs** arbres possibles), en regroupant les espèces sur la base du partage de caractères à l'état dérivé (en évitant les symplésiomorphies : regroupements par partages de caractères à l'état ancestral).
- 5- Dans le cas général (construction de plusieurs arbres), **choix** de l'arbre le plus vraisemblable, en général par principe de parcimonie (*voir cours*).

Polarisation des caractères :

Nous allons utiliser ici la démarche de **l'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 (*ou plus si plusieurs états dérivés différents*).

2.2. Logiciels utilisables

La réalisation d'une phylogénie par parcimonie, ou cladogramme, nécessite de disposer de données sur les caractères observables des espèces à classer. Ceci peut se faire à partir de documents, ou en utilisant un logiciel. Il est possible d'utiliser les logiciels de comparaisons de séquences, dans le cas d'analyses prenant les nucléotides comme caractères.

Le logiciel **Phylogène** permet l'affichage de caractères et la construction d'arbres. Ce logiciel, gratuit, est téléchargeable à l'adresse suivante :

<http://accres.ens-lyon.fr/accres/thematiques/evolution/logiciels/phylogene/telechargement-eleves/telechargement-eleves>

Voir en annexe la fiche-méthode sur Phylogène.

2.4 Réalisation d'un cladogramme avec Phylogène

Vous pouvez réaliser ce travail par groupes de 2 ou 3, en fonction des étudiants disposant d'un ordinateur.

➤ Télécharger et installer Phylogène, puis réaliser un arbre des Vertébrés à partir de la démarche indiquée.

Sur Mac, une page accessible à partir de la page de téléchargement indique une démarche permettant l'installation et l'utilisation de Phylogène. Pas de version « tablette – téléphone » pour Phylogène...

➤ (1) **Ouvrir** la collection « Vertébrés-lycées ».

➤ (2) **Sélectionner** les espèces suivantes :

Homme	Chat sauvage	Dauphine	Coelacanthé	Eléphant
Ornythorinque	Aigle	Oreillard	Pigeon	Babouin
Manchot	Grenouille	Sardine	Cheval	

➤ (3) **Construire** une matrice de caractères, en choisissant les caractères suivants :

Ailes	Allaitement	Appendices pairs	Branchies	Gésier
Deux fenêtres temporales		Placenta	Plumes	Poils
Poumons fonctionnels		Amnios	Pièces basales	Ongles plats
Crâne et Vertèbres		Phalanges spatulées		

➤ (4) **Polariser** la matrice (ce qui implique de choisir un extragroupe). Faire une capture d'écran.

➤ (5) **Construire** le cladogramme correspondant. Faire une capture d'écran.

ANNEXE 1 : MODE D'EMPLOI SIMPLIFIÉ DE GENIEGEN 2

Traiter des séquences avec le logiciel Geniegen2

Charger / ouvrir des séquences	Traiter des séquences
- Si les séquences sont fournies sous la forme d'un fichier, cliquer sur « charger des séquences (.edl) » sur le panneau d'accueil ou dans le menu « Fichier » - Si les séquences font partie de la banque de Geniegen2 : - o Cliquer sur « Ouvrir la banque de séquences » - o Rechercher les séquences en saisissant des mots clés - o Cliquer sur le nom des séquences ou packs de séquences souhaitées - o Cliquer sur « Charger ces séquences »	- Sélectionner les séquences à traiter en cochant la case située à gauche de leur nom - Choisir le traitement à effectuer dans le menu « Actions » : il est possible de transcrire une séquence d'ADN en ARN, de traduire une séquence d'ADN ou d'ARN en protéine, ou encore d'obtenir la séquence complémentaire d'une séquence d'ADN ou d'ARN Dans le cas où l'on souhaite ne traiter qu'une seule séquence, il est possible de le faire par un clic droit sur le nom de cette séquence. Il est également possible de supprimer des séquences en procédant ainsi.
Aligner des séquences	Se déplacer et se repérer au sein d'une séquence
- Si nécessaire, sélectionner les séquences à aligner en cochant la case située à gauche de leur nom (remarque : on ne peut aligner que des séquences de même nature : nucléotides ou acides aminés) - Cliquer sur « Aligner les séquences sélectionnées » dans le menu « Actions » ⇒ Les séquences alignées sont affichées dans la moitié inférieure de l'écran ; les nucléotides ou acides aminés manquants sont représentés par des tirets. L'alignement permet de prendre en compte d'éventuelles délétions ou insertions (discontinuités) qui sinon décaleraient les séquences lors de leur comparaison.	- Pour faire défiler rapidement les séquences, déplacer le curseur mobile situé sous les séquences (des marques colorées indiquent les zones similaires ou différentes) - Pour un défilement plus précis et plus lent, cliquer directement sur les séquences et les faire défiler en bougeant la souris tout en maintenant le bouton gauche enfoncé - Pour connaître précisément la position d'un nucléotide, d'un acide aminé, ou d'un codon, survoler la séquence à l'aide de la souris, sans cliquer Une règle graduée, située au-dessus des séquences, permet également de se repérer. Il est possible de changer le mode de numérotation (nucléotide, codon) via le menu « Options »
Comparer des séquences	Action des enzymes de restriction
Une ligne de comparaison colorée met en évidence les positions où les nucléotides (ou les acides aminés) sont identiques (étoiles sur fond vert) ou différents (points sur fond orange ou rouge selon le degré de différence). Ces couleurs se retrouvent également au niveau du curseur mobile, que vous pouvez déplacer à l'aide de la souris pour faire défiler les séquences sur toute leur étendue. Remarque : un mode adapté aux docteurs est disponible dans les options	- Sélectionner les séquences à traiter (uniquement des séquences d'ADN) en cochant la case située à gauche de leur nom - Cliquer sur « Enzymes de restriction » dans le menu « Actions » ⇒ Les séquences sélectionnées apparaissent alors la moitié inférieure de l'écran, ainsi que leurs séquences complémentaires (double brin d'ADN) - Cliquer sur « Ouvrir la banque d'enzymes », rechercher les enzymes souhaitées et les sélectionner en cliquant sur leurs noms, puis sur « Fermer la banque » ⇒ Les enzymes choisies sont maintenant disponibles dans le menu déroulant intitulé « Choisir une enzyme » - Choisir une enzyme dans le menu déroulant
Afficher le tableau de comparaison - Une fois les séquences alignées, cliquer sur « Afficher le tableau de comparaison » dans le menu « Affichage » - Si nécessaire, décocher « Similitudes » pour visualiser les différences, et décocher « en % » pour avoir des nombres plutôt que des pourcentages	Le résultat de l'action enzymatique apparaît sous la forme de traits rouges matérialisant les sites de coupure de l'enzyme
Afficher l'arbre matérialisant le degré de similarité (phénoگرامme) - Une fois les séquences (au moins 3) alignées, cliquer sur « Afficher le phénoگرامme » dans le menu « Affichage »	

ANNEXE 2 : MODE D'EMPLOI SIMPLIFIE DE PHYLOGENE

UTILISATION DE PHYLOGENE 2009

La barre d'outils

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16								
1 – Choisir une collection. 2 – Observer. 3 – Comparer. 4 – Construire une matrice de caractères. 5 – Afficher le tableau de référence de la collection choisie. 6 – Polariser et coder les états de caractères de la matrice. 7 – Classer dans des groupes emboîtés.							8 – Établir des parentés. 9 – Ouvrir un fichier de séquences. 10 – Ouvrir un fichier de groupes emboîtés. 11 – Imprimer. 12 – Copier. 13 – Enregistrer. 14 – Ouvrir un fichier 15 – Choix possibles 16 – Choisir une sous collection							Construire une matrice de caractères Sélectionner une collection : fichier/Sélectionner collection OK. Cliquer sur «construire». Sélectionner les espèces en cliquant sur chacune d'elles. Un nouveau clic sur la figure enlève l'espèce de la sélection. Choisir les caractères dans le menu déroulant. Remplir le tableau en cliquant dans chacune des cases et en utilisant les informations qui apparaissent en bas, à droite de l'écran. Vérifier le tableau et corriger si nécessaire.		Polariser et coder les états des caractères Après avoir construit et validé une matrice de caractères, cliquer sur l'icône «Polariser». Choisir un taxon extragroupe pour afficher les états primitifs (menu déroulant). Les fossiles apparaissent en rouge et l'extragroupe choisi en bleu. Colorer les états primitifs comme ceux de l'extragroupe, puis changer de couleur pour colorer les états dérivés. Cliquer sur «vérifier» à la fin.		Construire un arbre phylogénétique Cliquer sur le menu «établir des parentés». Les colonnes et les lignes peuvent être déplacées pour regrouper les caractères de mêmes états : cliquer sur «Organiser le tableau» puis sur le nom de l'espèce ou sur le caractère et déplacer. «Choisir les taxons» puis cliquer sur le nom de l'espèce. En cliquant sur un caractère dans la matrice, un code couleur apparaît. L'hypothèse de départ est que toutes les espèces choisies ont une origine commune (un même ancêtre commun). Toutes les branches sont réunies au même nœud. Cliquer sur la branche à déplacer avec le bouton gauche de la souris. Sans lâcher ce bouton, déplacer la souris de façon à venir faire le branchement là où on le souhaite. Si le branchement est possible, une croix apparaît.		Les outils pour construire un arbre Annuler la dernière opération Déplacer les branches Relier toutes les branches au même nœud Permuter des branches autour d'un nœud. Cliquer sur cette icône puis se placer sur un nœud et cliquer. Thème Inter Accéder au menu permettant de choisir le mode de représentation de l'arbre (avec ou sans boîtes), de donner un nom aux boîtes et de choisir le mode de travail (exploration ou édition) «copier l'état actuel» permet de comparer des arbres		Vérification de l'arbre En mode exploration. («choix») Lorsque l'on clique sur un nœud représentant un ancêtre, la liste de ses états de caractères apparaît (telle qu'elle peut être déduite de l'arbre dessiné). Si le groupe des descendants partage un état dérivé commun exclusif, les branches sont mises en vert. Enjaune : états dérivés communs non exclusifs. Sinon, une mise en garde est affichée et l'arbre est à modifier Attention : le nœud n'est pas justifié par un état dérivé commun exclusif Les dates sont celles du premier fossile avant cet état.	

Ouvrir un tableau

Phylogène a besoin de données pour pouvoir être utilisé.

Il est possible **d'ouvrir une collection**, qui comporte toutes les données nécessaires :

Fichier > Sélectionner une collection

Il est aussi possible **d'ouvrir un tableau de molécules** (.aln) ou un **tableau de caractères** (.tab) :

Fichier > Ouvrir > Tableau de caractère *ou* Tableau de molécules

e

Observer une espèce

Choisir : Observer

A partir d'un tableau de caractères, on peut observer une espèce soit en la sélectionnant dans la liste en bas à gauche, soit en cliquant sur l'icône ci-contre pour aboutir à une page présentant toutes les espèces du tableau.

Les menus défilants situés en bas permettent de choisir les données visualisées : image de l'espèce, données sur son mode de vie, de reproduction, anatomie de ses membres, de ses téguments, de son cœur, etc.

Les espèces présentant la lettre F à côté de leur image sont des espèces fossiles.



Comparer des espèces

Choisir : Comparer

On peut alors sélectionner (comme pour « observer une espèce ») de une à quatre espèces différentes. Attention aux espèces fossiles : certaines données ne sont pas disponibles.

Réaliser des matrices de caractères personnalisés

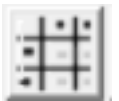
Choisir : construire une matrice

Puis **sélectionner les espèces désirées** (à partir de leurs images) et les **caractères voulus** (à partir de la liste présentée).

Remplir alors le tableau :

- Aller sur chaque case, et choisir la valeur du caractère, parmi celles proposées ;
- Une aide est fournie à chaque fois, en bas à droite ;
- Cliquer sur vérifier, et corriger si besoin (les cases erronées sont entourées de rouge).

Puis il est nécessaire de **polariser** la matrice : choisir un extragroupe, puis colorer les états ancestraux (primitifs), et cliquer sur les états dérivés pour les colorer d'une autre couleur.



Construire un arbre à partir de données anatomiques

Il faut d'abord avoir une **matrice de caractère** (soit fourni, soit personnalisé) sans erreur.

Choisir : Etablir des parentés

Un arbre apparaît. On peut alors :

- déplacer des espèces en maintenant le bouton gauche de la souris enfoncé sur la branche de l'espèce, pour l'amener à l'endroit qui semble le plus pertinent ;
- faire apparaître les états d'un caractère, en cliquant sur la colonne du caractère dans la matrice de caractères

Deux boutons utiles :



pour annuler la dernière action



pour procéder à une « rotation » autour d'un noeud



Construire un arbre à partir de données moléculaires

Il faut d'abord ouvrir un tableau de molécules.

Sélectionner les molécules voulues en cliquant sur le nom des espèces, ou par le bouton « tout sélectionner ».

Si besoin (mais cela vous sera alors précisé), aligner les molécules, en cliquant sur le bouton Aligner.

On peut alors :

- Afficher automatiquement l'arbre calculé par le logiciel : Arbre
- Afficher la matrice des distances : Matrice des distances

La matrice des distances donne les valeurs utilisées par le logiciel pour construire l'arbre : Entre deux espèces, une valeur de distance est calculée, à partir de la comparaison des séquences de la molécule chez les deux espèces. Plus les séquences diffèrent, plus la valeur est importante.