



Lycée François-René DE CHATEAUBRIAND  
136 BOULEVARD DE VITRÉ, CS 10637  
35706 RENNES CEDEX 7  
CLASSE PRÉPARATOIRE BCPST 1  
Biologie Chimie Physique Sciences de la Terre

Collection 1 : une phylogénie d'Arthropodes (3 taxons + 1 extragroupe)  
D'après SEGARRA, PIÈTRE *et al.* (2023)

ENSEIGNEMENT DE SCIENCES DE LA VIE ET DE LA TERRE (SVT)  
°° SCIENCES DE LA VIE °°  
>> Travaux pratiques <<

TP SV K

## Reconstruction phylogénétique

### SUPPORT ÉTUDIANT

#### Objectifs : extraits du programme

##### Capacités exigibles

- Construire une phylogénie par parcimonie (cladogramme) à partir d'un jeu limité de taxons et de caractères fournis (chez les Métazoaires ou les Embryophytes)
- Réaliser et exploiter des alignements de séquences afin de construire un arbre phylogénétique à partir d'une méthode fournie avec l'aide ou non d'un logiciel dédié.

#### Introduction

Dans ce TP, on se propose de :

- Construire des **arbres phylogénétiques** par **analyse cladistique** sur des **exemples simples** ;
- Construire un **arbre phylogénétique** par une **méthode moléculaire simple**.

Comment construit-on des arbres phylogénétiques ?

#### Activité 1. Analyse cladistique

À partir des collections d'organismes proposés et des choix de caractères, **remplissez** les matrices de caractères (d'après vos connaissances de TP ou de cours ; le professeur pourra compléter), **produisez** tous les arbres possibles (sur la feuille annexe) et **recopiez** proprement l'arbre le plus parcimonieux.

##### Remarques :

- En cas d'**équiprobabilité** du scénario « convergence » et du scénario « réversion », vous suivrez le scénario proposé. En l'absence d'indication, les **deux options** sont possibles.
- Vous **polariserez** les caractères uniquement par la **méthode de l'extragroupe**. La 1<sup>re</sup> ligne de chaque matrice correspond à l'extragroupe.
- En cas d'arbres équiprobables, vous **résoudrez** le conflit phylogénétique par consensus.
- Vous **noterez** les **acquisitions d'états dérivés** de caractères en rouge et les **pertes par réversion** en vert.

(1) En cas d'équiprobabilité du scénario « convergence » et du scénario « réversion », **envisagez** les **deux scénarios**.

Caractères	[1] Symétrie bilatérale	[2] Appendices articulés	[3] Squelette interne / externe	[4] Respiration branchiale / aérienne	[5] Antennes	[6] Mandibules	[7] Chélicères
Taxons							
Maquereau							
Criquet							
Écrevisse							
Araignée							

^ Matrice de caractères.

Polarisation des caractères :

- [1]
- [2]
- [3]
- [4]
- [5]
- [6]
- [7]

^ Arbre le plus parcimonieux.

**Collection 2 : la paraphylie des 'reptiles' (3 taxons + 1 extragroupe)**  
D'après LECOINTRE & LE GUYADER (2009)

(!) En cas d'équiprobabilité du scénario « convergence » et du scénario « réversion », **envisagez le scénario « convergence ».**

Caractères	[1] Amnios	[2] Gésier	[3] Mandibule pleine / fenestrée	[4] Bec corné	[5] Ailes	[6] Colonne vertébrale	[7] Écailles
<b>Taxons</b>							
Grenouille							
Cistude							
Crocodile							
Pigeon							

^ **Matrice de caractères.**





**Polarisation des caractères :**

- [1]
- [2]
- [3]
- [4]
- [5]
- [6]
- [7]

^ **Arbre le plus parcimonieux.**

**Collection 3 : la paraphylie des 'homéothermes' (3 taxons + 1 extragroupe)**  
D'après LECOINTRE *et al.* (2004)

(!) En cas d'équiprobabilité du scénario « convergence » et du scénario « réversion », **envisagez le scénario « convergence ».**

	Gésier [1]	Mandibule [2]	Régulation thermique [3]	Amnios [4]
EG: grenouille verte 				
Renard roux 				
Crocodile du Nil 				
Gris du Gabon 				

^ **Matrice de caractères.**

**Polarisation des caractères :**





- [1]
- [2]
- [3]
- [4]

^ **Arbre le plus parcimonieux.**

**Collection 4 : la paraphylie des organismes métamérisés et des 'vers'**  
(3 taxons + 1 extragroupe) – D'après LECOINTRE *et al.* (2004)

(!) En cas d'équiprobabilité du scénario « convergence » et du scénario « réversion », **envisagez le scénario « convergence ».**

(!) La matrice est déjà complétée.

	Cuticule [1]	Ecdysone [2]	Segmentation du corps [3]	Forme de type « ver » [4]	Symétrie bilatérale [5]
EG: <i>Pelagia noctiluca</i> (méduse) 	0	0	0	0	0
Ver de terre 	0	0	1	1	1
Criquet pèlerin 	1	1	1	0	1
<i>Chondronema passali</i> (nématode) 	1	1	0	1	1

^ **Matrice de caractères.**

**Polarisation des caractères :**

Les caractères sont codés comme suit: absence (0) ou présence (1) de cuticule; absence (0) ou présence (1) des hormones de la famille de l'ecdysone (provoquant des mues); le corps est (1) ou non (0) segmenté; l'animal n'a pas de forme allongée (0) ou l'animal est vermiforme (1); l'animal possède (1) ou non (0) une symétrie bilatérale.

^ **Arbre le plus parcimonieux.**

**Collection 4 : une phylogénie des Embryophytes (4 taxons + 1 extragroupe)**  
Original (2017)

(!) En cas d'équiprobabilité du scénario « convergence » et du scénario « réversion », **envisagez le scénario « convergence ».**

Caractère	Thalle (1)	Embryon lié à la plante mère (2)	Cuticule et sporopolléine (3)	Archégonés et anthéridies (4)	Trachéides >> Plantes vasculaires (5)	Lignine (6)	Ovule (7)	Cambium bifacial (8)	Fleur (9)	Vaisseaux vrais (10)
Ulve	0	0	0	0	0	0	0	0	0	0
Polytric	1	1	1	1	0	0	0	0	0	0
Polypode	1	1	1	1	1	1	0	0	0	0
Pin	1	1	1	0	1	1	1	1	0	0
Coquelicot	1	1	1	0	1	1	1	1	1	1

^ **Matrice de caractères.**

**Polarisation des caractères :**

- (1) :
- (2) :
- (3) :
- (4) :
- (5) :
- (6) :
- (7) :
- (8) :
- (9) :
- (10) :

^ **Arbre le plus parcimonieux.**

## Activité 2. Initiation à une méthode moléculaire (WPGMA)

### Principales méthodes utilisables en phylogénie moléculaire

On peut les classer en trois catégories principales.

#### a) Les méthodes phénétiques ou méthodes de distance

(!) Ne pas confondre **systematique phénétique** et **méthodes phénétiques** : si les premières **méthodes phénétiques** ont effectivement été inventées dans le cadre de la **systematique phénétique** (en les appliquant alors à des caractères morpho-anatomiques), elles sont **aujourd'hui appliquées à des données moléculaires dans un cadre conceptuel phylogénétique**.

#### Principe

Elles sont basées sur le **pourcentage de divergence** (ou de **similitude**) entre **séquences comparées deux à deux**. On construit pour cela des **tableaux à double entrée taxon-taxon** où l'on reporte ces **pourcentages** : ce sont des **matrices de distances**.

#### Méthodes principales

– L'**UPGMA** (*Unweighted Pair Group Method with Arithmetic mean*, méthode non pondérée de regroupement par paires par moyenne arithmétique) ou le **WPGMA** (*Weighted Pair Group Method with Arithmetic mean*, méthode non pondérée de regroupement par paires par moyenne arithmétique) sont des **méthodes simples** qui reposent sur une application des **principes phénétiques fondamentaux** : produire des **arbres basés sur la divergence quantifiée** entre **molécules**. Pour admettre la **pertinence phylogénétique** de ces méthodes, il faut adhérer sans réserve à la **théorie neutraliste de l'évolution moléculaire**. Selon cette théorie aujourd'hui bien admise (mais nuancée !), la **plupart des mutations** (et non toutes, contrairement à la caricature de certains détracteurs) sont **neutres à l'échelle moléculaire**. Dès lors, on peut **postuler** que, pour un **gène donné** et au sein d'un **ensemble de taxons donnés**, **les mutations s'accumulent de manière régulière au cours du temps** : c'est ce qu'on appelle l'**horloge moléculaire**. Aussi, l'UPGMA et le WPGMA postulent que, **plus les séquences se ressemblent, plus les taxons étudiés sont phylogénétiquement proches**. La **racine** sera placée sur la **séquence la plus différente** des autres.

– Le **NJ** (*Neighbour Joining*, rapprochement de voisinage) produit des **arbres non enracinés** sans se baser sur une stricte observance de l'horloge moléculaire, ce qui autorise la **prise en compte de potentielles variations des vitesses d'évolution des séquences** : c'est la raison de son invention en 1987 par SAITOU et NEI qui trouvaient l'UPGMA trop artificielle. Tous les taxons sont initialement **placés sur un arbre en étoile** et les taxons sont **regroupés petit à petit les uns après les autres** (on regroupe deux taxons, puis un troisième taxon avec les deux autres, puis un quatrième taxon avec l'ensemble des trois autres...) en construisant **tous les arbres possibles à chaque nouvelle étape** et en **choisissant celui qui présente une taille minimale** (notons qu'une **nouvelle matrice** est construite à chaque étape en appliquant une formule particulière). L'objectif est de **minimiser au maximum la taille de l'arbre final**.

#### Intérêts et limites

– L'**UPGMA** et le **WPGMA** reposent sur l'**horloge moléculaire** de manière **rigide** ; elle n'est donc quasiment plus employée, l'**hypothèse de non-variation des vitesses d'évolution** étant trop artificielle.

– Le **NJ** est une **méthode très populaire** depuis son invention : elle est rapide, **facile à mettre en œuvre** et peut s'appliquer à un **grand nombre de séquences**. Néanmoins, le mode de construction ne produit **pas nécessairement l'arbre le plus probable**.

– Ces méthodes **ne permettent pas de voir si un site nucléotidique a subi plus d'une mutation** (mais des **méthodes de correction** peuvent **pondérer** les résultats et **intégrer** cette **possibilité**).

#### b) Les méthodes de parcimonie

#### Principe

Il s'agit non plus de quantifier la divergence des séquences mais de **considérer chaque site (nucléotidique ou peptidique) muté** comme un **caractère**. La **parcimonie** consiste à **minimiser le nombre de « pas » (mutations)** nécessaires pour **passer d'une séquence à une autre** dans un arbre.

Cette méthode s'appuie sur les **hypothèses suivantes** :

- Les **sites** évoluent **indépendamment** les uns des autres (la séquence peut être considérée comme une **suite de caractères** non ordonnés)
- La **vitesse d'évolution** est **lente et constante** au cours du temps.

La méthode de **maximum de parcimonie** examine toutes les **topologies possibles** afin de trouver l'**arbre optimal** (le plus **parcimonieux**).

#### Intérêts et limites

- Cette méthode impose de **réaliser tous les arbres possibles**, ce qui **augmente rapidement** avec le nombre de séquences et rend le **travail très long**, même informatiquement.
- Cette méthode est **cladistique** plutôt que **phénétique** et elle intègre des **états initiaux et dérivés** ; on peut donc se tromper lors de la **polarisation**.
- La méthode ne permet toutefois **pas** de voir si un site a subi des **mutations multiples**, n'utilise **pas tous les sites existants** (mais juste les sites dits « informatifs ») et ne donne **aucune information sur la longueur des branches** (c'est-à-dire la divergence).

#### c) Les méthodes probabilistes

#### Principe

Les **méthodes probabilistes** utilisent des **formules de probabilité** en intégrant d'éventuels **modèles évolutifs** dans leurs algorithmes. Il ne s'agit plus seulement de comparer la divergence des séquences mais la **probabilité pour que les séquences aient muté sur tel ou tel site nucléotidique ou peptidique** (par exemple, on intègre la probabilité de chaque type de mutation). Elles reposent sur des **modèles d'évolution**. Lorsque le **modèle est bien choisi**, ces méthodes sont celles qui **s'approchent le mieux de la réalité**.

#### Méthodes principales

On peut citer deux grandes catégories d'approches :

- La **méthode du maximum de vraisemblance (ML, maximum likelihood)** : cette méthode travaille sur les **sites considérés comme des caractères** mais intègre la **probabilité que ceux-ci subissent une (ou plusieurs) mutations** en utilisant des modèles évolutifs choisis par le phylogénéticien.
- Les **méthodes bayésiennes** fondées sur le **cadre probabiliste de BAYES** qui sont une méthode calculatoire encore plus **complexe et lourde** (donc parfois difficile à mettre en œuvre) mais donnant d'excellents résultats.

#### Intérêts et limites

- Ces méthodes permettent d'obtenir des **résultats jugés très fiables**, en **minimisant les biais** des phylogénies moléculaires classiques. Ils sont considérés comme **les plus prometteurs** aujourd'hui.
- **Aucune référence absolue** n'existe pour **choisir le modèle évolutif pertinent** face à une séquence (et le choix d'un modèle erroné est le biais le plus préjudiciable à ces approches).
- Ces méthodes sont **extrêmement lourdes** d'un point de vue **calculatoire et algorithmique** (surtout BAYES) d'où l'emploi d'heuristiques pour limiter le nombre de configurations étudiées, ce qui implique que des possibilités ne soient pas examinées.

Appliquez l'algorithme proposé pour produire des arbres racinés selon la méthode WPGMA (parfois à tort appelée UPGMA dans quelques manuels francophones).

**Exercice 1 : une phylogénie des globines alpha de quelques Vertébrés**  
D'après LECOINTRE & LE GUYADER (2009)

On se propose de construire, par le biais de la méthode WPGMA, une phylogénie simple de quelques Vertébrés à partir des séquences peptidiques de leurs globines alpha (document a).

```

ESPÈCES          Séquences alignées
-----
Carpe commune    SLSDKDKAAV KIAWAKISPK ADDIGAEALG RMLTVYPQTK TYFAHWADLS PGSGPVRKHGK
Homme             V--PA--TN-  -A--G-VGAH -GEY-----E --FLSP-T--  ---P-F*--- H--AQ--GHG
Roussette        V--SA--TNI  -A--D-VGGN -GEY-----E --FLSP-T--  ---P-F*--- H--AQ--GHG
Coq domestique  V--AA--NN-  -GIFT--AGH -EEY---T-E  --F-T--P--  ---P-F*--- H--AQI-GHG

          Séquences alignées
          -----
KVIMGAVGDA VSKIDDLVGG LASLSELHAS KLRVDPANFK ILANHIVVGI MFYLPGFDFPP
-KVAD-LTN- -AHV--MPNA -SA--D---H  -----V--- L-SHCLL-TL AAH--AE-T-
-KVGD-LTN- -GHL---P-A  -SA--D---Y  -----V--- L-SHCLL-TL ANH--S--T-
-KVVA-LIE- ANH---IA-T -SK--D---H  -----V--- L-GQCFL-VV AIHH-AALT-

- Séquences alignées -
EVHMSVDKFF QNLALALSEK YR
A--A-L---L ASVSTV-TS- --
A--A-L---L ASVSTV-TS- --
---A-L---L CAVGTV-TA- --
    
```

Séquences alignées d'une portion de l'hémoglobine  $\alpha$ . Les lettres correspondent au code des acides aminés. Un tiret signifie identité par rapport à l'acide aminé de la première ligne; une étoile l'absence de l'acide aminé.

▲ DOCUMENT a. Matrice de position comparant les séquences alignées de globines  $\alpha$  de quatre Vertébrés.

1. On appelle **distance** ou **divergence** le pourcentage de nucléotides qui diffèrent entre les gènes de deux espèces étudiées (notons qu'un **nucléotide absent** est compté comme une **différence** dans la séquence). Recopiez et complétez la matrice de distance (document b) en utilisant les séquences proposées.

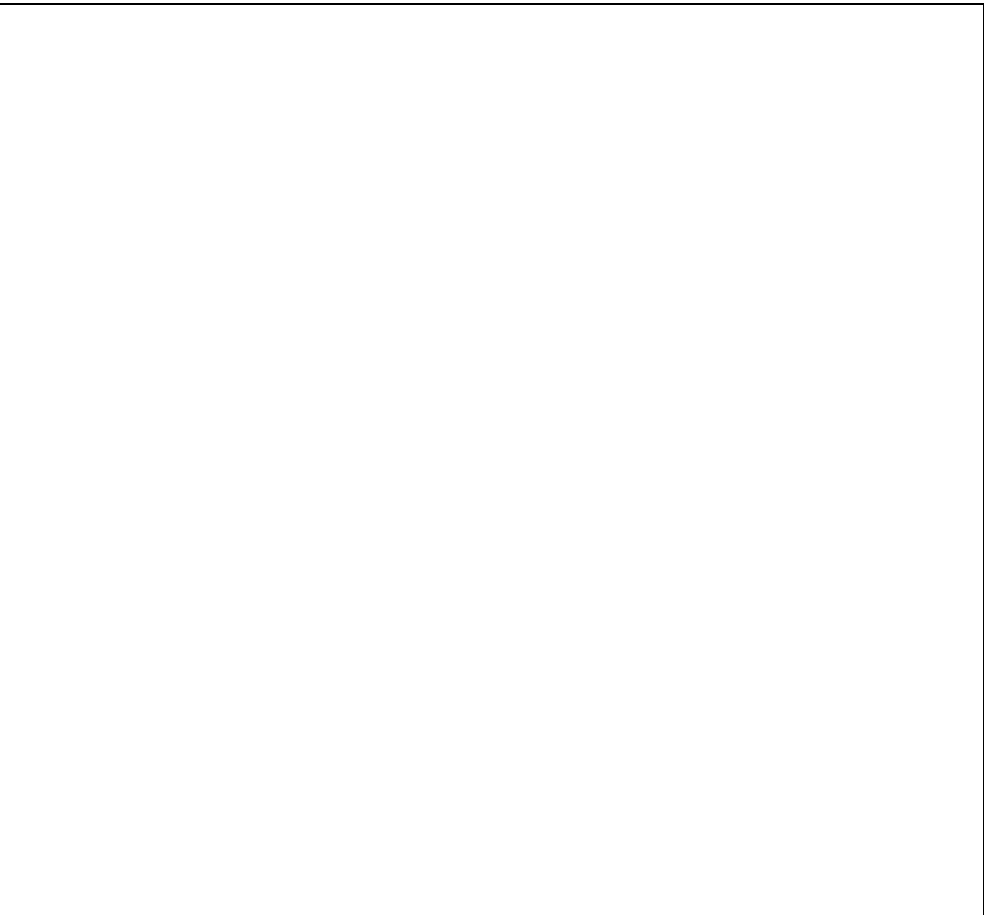
Taxons / Taxons	Carpe	Homme	Roussette	Coq
Carpe	0,00	-	-	-
Homme		0,00	-	-
Roussette			0,00	-
Coq				0,00

▲ DOCUMENT b. Matrice de distance à compléter.

2. On admettra que **toutes les séquences évoluent à la même vitesse dans toutes les branches de l'arbre** (ce qui constitue un postulat de la méthode WPGMA) et que les pourcentages calculés reflètent les différences pour l'extrait de **séquence peptidique de l'hémoglobine**. En utilisant la matrice produite à la question précédente, **construisez un phylogramme WPGMA** des taxons proposés en employant la méthode expliquée :

- Dans la matrice des distances, trouver les taxons  $i$  et  $j$  pour lesquels la distance  $d_{ij}$  est la plus petite.
- Mettre la racine à égale distance de  $i$  et de  $j$ , soit à  $d_{ij}/2$ .
- Créer un nouvel ensemble qu'on peut nommer  $X$  incluant  $i$  et  $j$ . Si  $i$  et  $j$  sont les deux derniers ensembles, l'arbre est complet.
- Recalculer une matrice de distance en prenant comme distance de  $X$  à chaque taxon  $k$ , la moyenne des distances  $d_{ki}$  et  $d_{kj}$  :  $d_{Uk} = (d_{ki} + d_{kj}) / 2$
- À partir de cette nouvelle matrice (dans laquelle il y a une entrée de moins car  $i$  et  $j$  ont été agglomérés en un ensemble  $U$ ), retourner à la première étape.

Notez qu'on vous demande un **phylogramme**, c'est-à-dire un arbre phylogénétique dont la longueur est proportionnelle à la divergence des molécules : vous devrez donc faire des branches dont la taille sera égale à chaque fois à  $d_{ij}/2$ .



**Exercice 2 : une phylogénie des globines alpha de quelques Vertébrés**  
D'après Agro-véto A BCPST 2009, remis en forme dans SEGARRA *et al.* (2015)

On se propose d'essayer de **construire**, par le biais de la **méthode WPGMA\***, une **phylogénie simple de plastes d'organismes** en comparant des **gènes d'ARN 16S** du **stroma de plastes** de 3 organismes la **Lignée verte** et du **cytoplasme de cyanobactéries**. Les séquences sont proposées dans le **tableau page suivante** où elles ont été **alignées**, la séquence d'*Arabidopsis thaliana* (Angiospermes) servant de référence.

1. On appelle **distance** ou **divergence** le pourcentage de nucléotides qui diffèrent entre les gènes de deux espèces étudiées (notons qu'un nucléotide absent est compté comme une différence dans la séquence). **Recopiez et complétez** la **matrice de distance** (**ci-dessous**) en utilisant les séquences proposées.

Taxons \ Taxons	<i>Arabidopsis</i>	<i>Anabaena</i>	<i>Cyanophora</i>	<i>Porphyra</i>
<i>Arabidopsis</i>	0,00	-	-	-
<i>Anabaena</i>		0,00	-	-
<i>Cyanophora</i>			0,00	-
<i>Porphyra</i>				0,00

2. On admettra que toutes les séquences évoluent à la même vitesse dans toutes les branches de l'arbre (ce qui constitue un postulat de la méthode WPGMA) et que les pourcentages calculés reflètent les différences pour la séquence totale de l'ARN 16S. En utilisant la matrice produite à la question précédente, **construisez** un **phylogramme** WPGMA des taxons proposés en employant la méthode expliquée :
- dans la matrice des distances, trouver les taxons *i* et *j* pour lesquels la distance  $d_{ij}$  est la plus petite ;
  - mettre la racine à égale distance de *i* et de *j*, soit à  $d_{ij}/2$  ;
  - créer un nouvel ensemble *U* incluant *i* et *j*. Si *i* et *j* sont les deux derniers ensembles, l'arbre est complet ;
  - recalculer une matrice de distance en prenant comme distance de *U* à chaque taxon *k*, la moyenne des distances  $d_{ki}$  et  $d_{kj}$ :  $d_{Uk} = (d_{ki} + d_{kj})/2$  ;
  - à partir de cette nouvelle matrice (dans laquelle il y a une entrée de moins car *i* et *j* ont été agglomérés en un ensemble *U*), retourner à la première étape.

Notez qu'on vous demande un **phylogramme**, c'est-à-dire un arbre phylogénétique dont la longueur est proportionnelle à la divergence des molécules : vous devrez donc faire des branches dont la taille sera égale à chaque fois à  $d_{ij}/2$ .



### Activité 3. Comparaison et alignement de séquences pour la réalisation d'arbres phylogénétiques

D'après SEGARRA, PIÈTRE et al. (2023)

Le séquençage massif des génomes a produit de très nombreuses séquences disponibles dans des bases de données accessibles à tous, par exemple :

- le site *National Center for Biotechnology Information* (<https://www.ncbi.nlm.nih.gov/>) qui met à disposition de très nombreuses séquences d'acides nucléiques (ADN et ARN) et peptidiques provenant de très nombreux organismes ainsi que des outils d'analyse.
- le site *The Protein Data Bank* (<https://www.rcsb.org>) qui met à disposition de nombreuses séquences peptidiques.

Ces sites proposent aussi des outils de visualisation moléculaire (affichage de séquences, modèles tridimensionnels de protéines...) et d'analyse des séquences (recherche dans les bases de données, alignement, comparaison, recherche de phases ouvertes de lecture, transcription/traduction...).

La comparaison et l'alignement de ces séquences est exploitée afin de construire des arbres phylogénétiques montrant les relations de parenté entre les séquences et, si les séquences sont orthologues, entre les organismes dont elles sont issues.

Dans cette fiche TP, on utilisera les séquences de différents transporteurs au glucose de la famille GluT.

<https://www.ncbi.nlm.nih.gov/>  
<https://www.rcsb.org/>

#### Alignement des séquences nucléotidiques et des séquences peptidiques codant GluT4

Le gène *Solute carrier family 2 (facilitated glucose transporter), member 4* est le gène codant la protéine couramment désigné GluT4 pour Glucose Transporter 4.

Les séquences codantes de ce gène ainsi que les séquences peptidiques pour différentes espèces sont disponibles dans les bases de données.

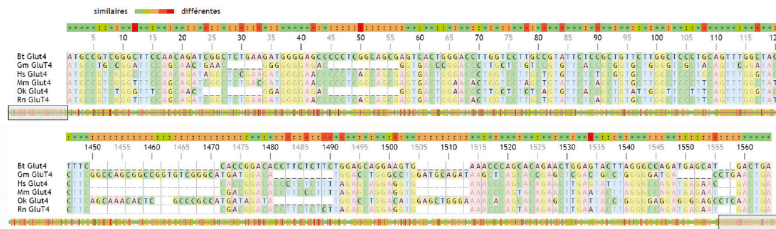
Le tableau ci-dessous récapitule quelques caractéristiques de ces séquences.

Espèce concernée	Taille de la séquence codante (nucléotides)	Taille de la séquence peptidique (acides aminés)	Nom de la séquence sur les documents
Vache domestique <i>Bos taurus</i>	1530	509	Bt GluT4
Morue de l'Atlantique <i>Gadus morhua</i>	1512	503	Gm GluT4
Homme moderne <i>Homo sapiens</i>	1530	509	Hs GluT4
Souris domestique <i>Mus musculus</i>	1530	509	Mm GluT4
Saumon pacifique <i>Oncorhynchus kisutch</i>	1518	505	Ok GluT4
Rat brun ou rat d'égout <i>Rattus norvegicus</i>	1530	509	Rn GluT4

Figure TP 25.3a

Données sur les séquences codantes et peptidiques de GluT4 utilisées dans cette étude

L'alignement des séquences nucléotidiques avec le logiciel *GenieGen2* en autorisant les discontinuités permet une comparaison des séquences conduisant à établir un pourcentage de similarité entre séquences. La figure TP 25.3b présente le résultat de l'alignement avec discontinuité des six séquences utilisées dans cette étude. L'introduction d'espaces correspondant à ces discontinuités au sein d'une séquence permet d'optimiser l'alignement des portions de séquences similaires.



**Figure TP 25.3b**  
Alignements des six séquences avec discontinuités: début et fin de l'alignement  
Chaque nucléotide est associé à une couleur (A rose, T bleu, etc.). Des tirets bas représentent un espacement introduit pour produire l'alignement le plus optimal pour l'ensemble de la séquence.  
La ligne au-dessus de l'alignement représente la similarité entre les six séquences: du vert (tous les nucléotides sont identiques) au rouge (les nucléotides sont tous différents d'une séquence à une autre).  
D'après une copie d'écran des résultats fournis par le logiciel *GenieGen2*.

Les figures TP 25.3c et TP 25.3d présentent les matrices de similarité entre les séquences nucléotidiques et peptidiques de différentes espèces. La valeur indiquée dans chaque case est un pourcentage de similitude entre les deux séquences comparées.

	Bt GluT4	Gm GluT4	Hs GLuT4	Mm GluT4	Ok GluT4	Rn GluT4
Bt GluT4	100					
Gm GluT4	67,58	100				
Hs GluT4	90,28	66,82	100			
Mm GluT4	87,4	65,66	88,17	100		
Ok GluT4	67,97	81,84	67,39	66,75	100	
Rn GluT4	87,28	65,86	88,17	95,72	66,37	100

**Figure TP 25.3c**  
Matrice de distances des séquences nucléotidiques de la portion codante du gène *GluT4*: résultats en pourcentage de similitude entre les six séquences nucléotidiques comparées.  
D'après des résultats fournis par le logiciel *GenieGen2*.

	Bt GluT4 p	Gm GluT4 p	Hs GLuT4 p	Mm GluT4 p	Ok GluT4 p	Rn GluT4 p
Bt GluT4 p	100					
Gm GluT4 p	66,02	100				
Hs GluT4 p	94,4	65,25	100			
Mm GluT4 p	93,63	64,67	95,37	100		
Ok GluT4 p	68,15	85,14	67,95	67,18	100	
Rn GluT4 p	93,63	64,86	95,37	98,07	67,57	100

**Figure TP 25.2d**  
Matrice de distances des séquences peptidiques de la protéine GluT4: résultats en pourcentage de similitude entre les six séquences peptidiques comparées. p: séquence peptidique.  
D'après des résultats fournis par le logiciel *GenieGen2*.

1. Comparez les pourcentages de similitude obtenus pour les deux types de comparaisons (nucléotidiques et polypeptidiques). Comment peut-on expliquer les différences éventuellement observées entre les deux types de séquences ?

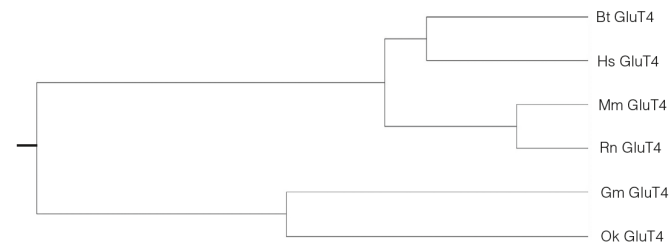
## Arbres phylogénétiques obtenus à partir des données de la comparaison de séquences codant GluT4 chez différentes espèces

À partir de matrices de similarités, il est possible de réaliser des arbres de parenté en utilisant différentes méthodes.

### Arbre à partir de la méthode UPGMA

La méthode **UPGMA** pour *Unweighted Pair Group Method with Arithmetic mean* (« méthode non pondérée de regroupement par paires basée sur la moyenne arithmétique ») est fondée sur le concept d'horloge moléculaire qui postule que pour un gène donné, **les mutations s'accumulent régulièrement** au cours du temps. Les séquences, puis les groupes, les plus proches, estimés à partir des pourcentages de divergence, sont regroupés deux à deux.

Le logiciel *GenieGen2* produit un arbre fondé sur cette méthode: figure TP 25.3e.

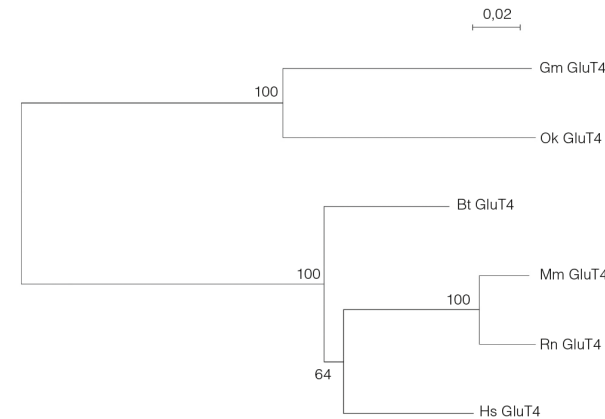


**Figure TP 25.3e**  
Arbre obtenu par la méthode UPGMA  
D'après une copie d'écran des résultats fournis par le logiciel *GenieGen2*.

### Arbre obtenu à partir de la méthode de Neighbour-Joining (NJ)

La méthode NJ (*Neighbour Joining*) est aussi une méthode de distance exploitant la divergence entre séquences en limitant les biais de l'horloge moléculaire. Elle aboutit à la production d'un arbre unique minimisant la taille finale de l'arbre.

Cet arbre (figure TP 25.3f) a été obtenu avec le logiciel *Seaview* qui compare les séquences et produit des arbres selon différentes méthodes.



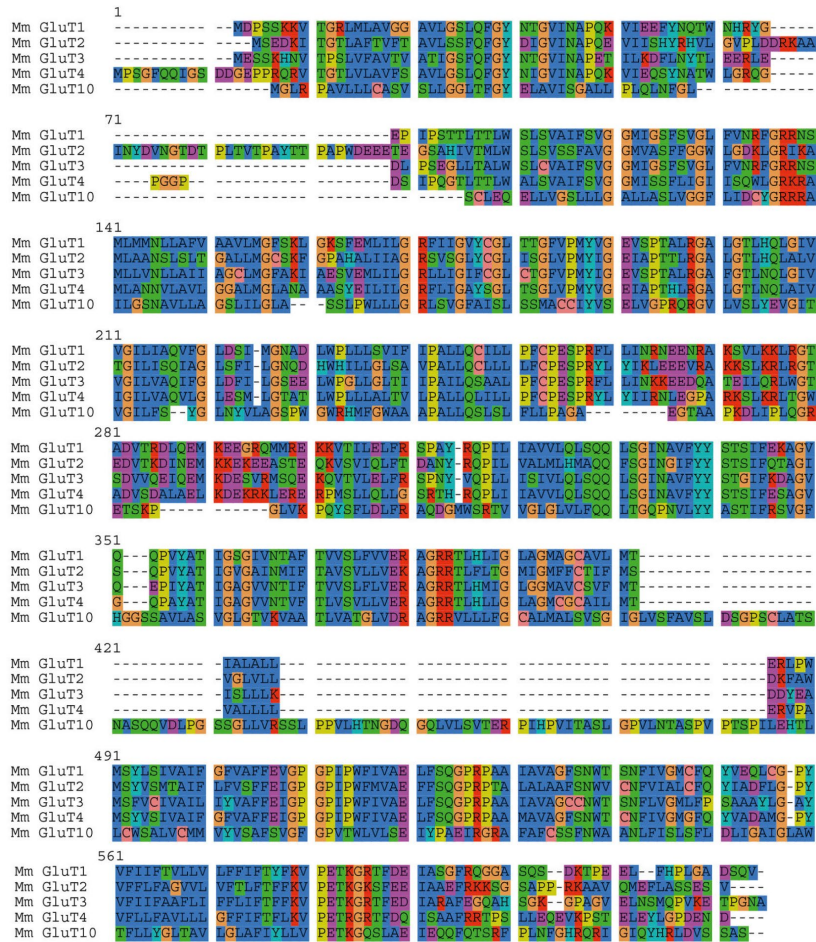
**Figure TP 25.3f**  
Arbre obtenu par la méthode NJ (méthode HKY)  
Les longueurs de branches sont proportionnelles au nombre de substitutions par site pour chaque branche. La barre en haut à gauche représente l'échelle des branches.  
Les valeurs à gauche des nœuds sont les valeurs de *bootstrap*, un test statistique par rééchantillonnage des données évaluant la robustesse des branches.

2. Comparez les deux arbres et expliquez sommairement les différences constatées.

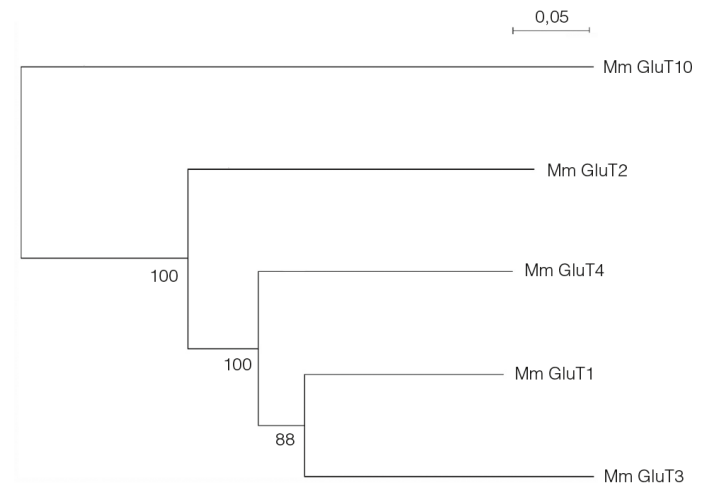
## Comparaison des séquences de différents membres de la famille des GluT

Dans cette partie, on compare les séquences de différents transporteurs GluT au sein d'une même espèce. Ce sont donc des **séquences paralogues** constituant ce qu'il est courant d'appeler une famille multigénique.

Les séquences peptidiques des transporteurs GluT1, GluT2, GluT3, GluT4 et GluT10 de la souris domestique (*Mus musculus*) sont comparées et un alignement avec des discontinuités est réalisé (figure TP 25.3g). Ces données sont ensuite traitées par le logiciel Seaview pour produire un arbre phylogénétique en utilisant la méthode de Neighbour-Joining (figure TP 25.3h).



**Figure TP 25.3g**  
Alignements avec discontinuités des cinq séquences peptidiques. Chaque couleur représente une famille d'acides aminés caractérisée par ses propriétés (polarité, charge, groupement fonctionnel). D'après une copie d'écran des résultats fournis par le logiciel Seaview.



**Figure TP 25.3h**  
Arbre obtenu par la méthode NJ. Les longueurs de branches sont proportionnelles au nombre de substitutions par site pour chaque branche. La barre en haut à gauche représente l'échelle des branches. Les valeurs à gauche des nœuds sont les valeurs de bootstrap.

### 3. Commentez l'arbre obtenu. Quelle interprétation peut-on en faire ?

#### Sources

Logiciel *GenieGen2* (Version utilisée : 2.6). Auteurs : P. Cosentino/P. Pillot

<https://www.pedagogie.ac-nice.fr/svt/productions/geniegen2/>, dernière consultation janvier 2022

Logiciel *Seaview* (Version 5.0.4). Auteurs : © 1996-2020 Manolo Gouy, Laboratoire de Biométrie et Biologie Evolutive, CNRS/Université de Lyon

<http://doua.prabi.fr/software/seaview>, dernière consultation janvier 2022

<https://www.pedagogie.ac-nice.fr/svt/productions/geniegen2/>  
<https://doua.prabi.fr/software/seaview>

s

## Références

- ALBERTS, B., A. JOHNSON, J. LEWIS, M. RAFF, K. ROBERTS & P. WALTER (2004). *Biologie moléculaire de la cellule. Quatrième édition*. Traduction de la quatrième édition américaine (2002) par F. LE SUEUR-ALMOSNI. Flammarion, Paris. Première édition américaine 1983 (1986 1<sup>re</sup> édition française).
- ALLANO, L. & A. CLAMENS (2010). *Faits et mécanismes de l'évolution biologique*. Ellipses, Paris.
- BABIN, C. (1991). *Principes de paléontologie*. Armand Colin, Paris.
- BAUTZ, A.-M. & A. BAUTZ (coll. D. CHARDARD) (2010). *Mini-manuel de Biologie animale*. Dunod, Paris, 2e édition (1e édition 2007).
- BERTHET, J. (2006). *Dictionnaire de Biologie*. De Boeck Université, Bruxelles (Belgique).
- BREUIL, M. (2007). *Biologie 1<sup>re</sup> année BCPST-véto*. Tec & Doc, Lavoisier, Paris.
- BREUIL, M. (2009). *Biologie 2<sup>e</sup> année BCPST-véto*. Tec & Doc, Lavoisier, Paris.
- CAMPBELL, N. A. & J. B. REECE (2004). *Biologie*. De Boeck Université, Bruxelles, 2<sup>e</sup> édition (1<sup>re</sup> édition 1995).
- [CAMPBELL, N. A.], J. B. REECE, L. A. URY, M. L. CAIN, S. A. WASSERMAN, P. V. MINORSKY, R. B. JACKSON (2012). *Campbell Biologie*. Adaptation française J. FAUCHER & R. LACHAÎNE. Pearson, Paris (4e édition).
- CHASSANY, V., M. POTAGE & M. RICOU (2012). *Mini manuel de Biologie végétale*. Dunod, Paris.
- DE WEVER, P., B. DAVID & D. NÉRAUDEAU (Coll. J. BROUTIN) (2010). *Paléobiosphère. Regards croisés des Sciences de la Vie et de la Terre*. Vuibert, Paris.
- DELÉAGE, G. & M. GOUY (2013). *Bioinformatique. Cours et cas pratiques*. Dunod, Paris.
- DENÉUD, J., T. FERROIR, O. GUIPPONI, H. MOREAU, M. PAULHIAC-PISON, M.-L. PONS & F. TEJEDOR (2011). *Biologie-Géologie BCPST-véto 2<sup>e</sup> année*. Tec & Doc, Lavoisier, Paris.
- DENÉUD, J., C. GODINOT, O. GUIPPONI, H. MOREAU, M. PAULHIAC-PISON & F. TEJEDOR (2013). *Biologie-Géologie BCPST-véto 1<sup>re</sup> année*. Tec & Doc, Lavoisier, Paris.
- DENÉUD, J., C. GODINOT, O. GUIPPONI, H. MOREAU, M. PAULHIAC-PISON, M.-L. PONS & F. TEJEDOR (2014). *Biologie-Géologie BCPST-véto 2<sup>e</sup> année*. Tec & Doc, Lavoisier, Paris.
- GODINOT, C., H. MOREAU, M. PAULHIAC-PISON & F. TEJEDOR (2010). *Biologie-Géologie 1<sup>re</sup> année BCPST-véto*. Tec & Doc, Lavoisier, Paris.
- HALL, B. K. & B. HALLGRIMSSON (2008). *Strickberger's evolution*. Jones and Bartlett, Sudbury, MA (USA), 4e édition (1e édition par M. W. Strickberger 1990).
- HARRY, M. (2008). *Génétique moléculaire et évolutive*. Maloine, Paris, 2e édition (1e édition 2001).
- HENNIG, W. (1966). *Phylogenetic Systematics*. University of Illinois Press, Urbana, Illinois (USA).
- LAFON, C. (2003). *La biologie autrement. 100 questions de synthèse*. Ellipses, Paris.
- LE GUYADER, H. (dir.) (1998). *L'Évolution*. Belin-Pour la Science, Paris.
- LECOINTRE, G. (dir.) (2004). *Comprendre et enseigner la classification du vivant*. Belin, Paris.
- LECOINTRE, G. (dir.), C. FORTIN, G. GUILLOT & M.-L. LE LOUARN-BONNET (2009). *Guide critique de l'évolution*. Belin, Paris.
- LECOINTRE, G. & H. LE GUYADER (2009). *Classification phylogénétique du vivant*. Illustrations D. VISETT. Belin, Paris, 3e édition (1e édition 2001).
- LECOINTRE, G. & H. LE GUYADER (2013). *Classification phylogénétique du vivant Plantes à fleurs. Cnidaires. Insectes. Squamates. Oiseaux. Téléostéens*. Belin, Paris.
- MEYER, S., C. REEB & R. BOSDEVEIX (2008). *Botanique. Biologie et physiologie végétales*. Maloine, Paris, 2e édition (1e édition 2004).
- MORÈRE, J.-L., R. PUJOL (coord.), J.-C. CALLEN, L. CHESNOY, J.-P. DUPONT, A.-M. GIBERT-TANGAPREGASSOM, G. RICOU, N. TOUZET (dir.) et collaborateurs (2003). *Dictionnaire raisonné de Biologie*. Frison-Roche, Paris.
- PERRIÈRE, G. & C. BROCHIER-ARMANET (2010). *Concepts et méthodes en phylogénie moléculaire*. Springer, Berlin (D).
- PEYCRU, P. (dir.), J.-F. FOGELGESANG, D. GRANDPERRIN, B. AUGÈRE, J.-C. BAEHR, C. PERRIER, J.-M. DUPIN & C. VAN DER REST (2010a). *Biologie tout-en-un BCPST 1<sup>re</sup> année*. Dunod, Paris, 2<sup>e</sup> édition (2009), réimpression corrigée (2010) (1<sup>re</sup> édition 2006).
- PEYCRU, P. (dir.), J.-C. BAEHR, F. CARIU, D. GRANDPERRIN, C. PERRIER, J.-F. FOGELGESANG & J.-M. DUPIN (2010b). *Biologie tout-en-un BCPST 2<sup>e</sup> année*. Dunod, Paris, 2<sup>e</sup> édition (1<sup>re</sup> édition 2007).
- PEYCRU, P., D. GRANDPERRIN, C. PERRIER (dir.), B. AUGÈRE, T. DARRIBÈRE, J.-M. DUPIN, C. ESCUYER J.-F. FOGELGESANG, & C. VAN DER REST (2013). *Biologie tout-en-un BCPST 1<sup>re</sup> année*. Dunod, Paris, 3<sup>e</sup> édition (1<sup>re</sup> édition 2006).
- PEYCRU, P., D. GRANDPERRIN, C. PERRIER (dir.), B. AUGÈRE, J.-F. BEAUX, F. CARIU, P. CARRÈRE, T. DARRIBÈRE, J.-M. DUPIN, C. ESCUYER, J.-F. FOGELGESANG, S. MAURY, É. QUÉINNEC, E. SALGUEIRO & C. VAN DER REST (2014). *Biologie tout-en-un BCPST 2<sup>e</sup> année*. Dunod, Paris, 3<sup>e</sup> édition (1<sup>re</sup> édition 2007).
- PRAT, D., A. RAYNAL-ROQUES & A. ROQUENANT (dir.) (2008). *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*. Belin, Paris.
- RAVEN, P. H., G. B. JOHNSON, J. B. LOSOS, S. S. SINGER (2007). *Biologie*. De Boeck, Bruxelles.
- REVIERS, B. DE (2002, 2003). *Biologie et phylogénie des algues. Tome 1. Tome 2*. Belin, Paris.

- RICHARD, D. (dir.), P. CHEVALET, S. FOURNEL, N. GIRAUD, F. GROS, P. LAURENTI, F. PRADÈRE & T. SOUBAYA (2012). *Biologie. Tout le cours en fiches. Licence. CAPES. Prépas*. Dunod, Paris, 2<sup>e</sup> édition (1<sup>re</sup> édition 2010).
- RIDLEY, M. (1997). *Évolution*. De Boeck, Bruxelles, traduction de la 2e édition américaine (1996).
- RIDLEY, M. (2004). *Évolution*. Blackwell, Malden, MA, USA, 3e édition (1e édition 1993).
- SCHUH, R. T. & A. W. BROWNER (2009). *Biological systematics. Principles and applications. Second edition*. Cornell University Press, New York, USA (1e édition 2000).
- SEGARRA, J. (dir.), É. CHAUVET, C. COLSON-PROCH, M. HUILLE, M. LABROUSSE, F. LOUET, F. METZ & E. PIÈTRE (2014). *Biologie BCPST 1<sup>re</sup> année*. Ellipses, Paris.
- SEGARRA, J. (dir.), G. BAILLY, O. CHASSAING, D. FAVRE, T. JEAN, F. METZ, C. MEUNIER & E. PIÈTRE (à paraître). *Biologie BCPST 2<sup>e</sup> année*. Ellipses, Paris.
- SERRE, J.-L. (2006a). *Génétique*. Dunod, Paris, 3e édition (1e édition 2001).
- SOLIGNAC, M., G. PERIQUET, D. ANXOLABÈRE & C. PETIT (1995a). *Génétique et évolution. Tome I. La variation, les gènes dans les populations*. Hermann, Paris.
- SOLIGNAC, M., G. PERIQUET, D. ANXOLABÈRE & C. PETIT (1995b). *Génétique et évolution. Tome II. L'espèce, l'évolution moléculaire*. Hermann, Paris.
- STEARNS, S. C. & R. F. HOEKSTRA (2005). *Evolution : an introduction*. Oxford University Press, Oxford (UK), 2e édition (1e édition 2000).
- THOMAS, F., T. LEFÈVRE & M. RAYMOND (dir.) (2010). *Biologie évolutive*. De Boeck, Bruxelles.

## Plan du TP

<b>Objectifs : extraits du programme</b>	<b>1</b>
<b>Introduction</b>	<b>1</b>
<b>Activité 1. Analyse cladistique</b>	<b>1</b>
<b>Activité 2. Initiation à une méthode moléculaire (WPGMA)</b>	<b>4</b>
<b>Activité 3. Comparaison et alignement de séquences pour la réalisation d'arbres phylogénétiques</b>	<b>8</b>
<b>Références</b>	<b>12</b>
<b>Plan du TP</b>	<b>12</b>

© Tanguy JEAN. Les textes et les figures originales sont la propriété de l'auteur. Les figures extraites d'autres sources restent évidemment la propriété des auteurs ou éditeurs originaux.  
Inspiré de documents variés antérieurs. Adaptation : juin 2023. Dernière relecture : mai 2024.  
Contact : [Tanguy.Jean4@gmail.com](mailto:Tanguy.Jean4@gmail.com)  
Adresse de téléchargement : <https://www.tanguy-jean-svt.com/>



Ces données sont placées sous licence *Creative Commons Attribution – Pas d'Utilisation commerciale 4.0 CC BY NC* qui autorise la reproduction et la diffusion du document, à condition d'en citer explicitement la source et de ne pas en faire d'utilisation commerciale.