

TD N°1 – L'ALIGNEMENT DES SÉQUENCES : PHYLOGÉNIE DES MAMMIFÈRES ET ÉVÉNEMENTS D'INDELS.

Objectifs du TP : Aligner l'exon 11 du gène **BRCA1** (*BReast and ovarian CAncer susceptibility gene N°1*) chez 8 mammifères placentaires et un groupe externe marsupial. Comprendre la dynamique évolutive des indels et évaluer leur signal phylogénétique. Inférer et comparer des arbres phylogénétiques par méthodes de distance (**NJ**), de cladistique avec parcimonie (**MP**), et par approche probabiliste avec maximum de vraisemblance (**ML**).

Nom latin	Nom commun	Taxonomie	Accessions	Phylogénie actuellement admise
<i>Vombatus ursinus</i>	wombat	Marsupiaux	AF284031	
<i>Orycteropus afer</i>	oryctérope	Afrothériens	AF284030	
<i>Elephas maximus</i>	éléphant	Afrothériens	AF284022	
<i>Dasybus novemcinctus</i>	tatou	Xénarthres	AF283999	
<i>Bradypus tridactylus</i>	paresseux	Xénarthres	AF284002	
<i>Homo sapiens</i>	homme	Euarchontoglires	AY304547	
<i>Mus musculus</i>	souris	Euarchontoglires	U31625	
<i>Bos taurus</i>	vache	Laurasiathériens	AF284013	
<i>Erinaceus europaeus</i>	hérisson	Laurasiathériens	AF284008	

Question 1 : Rapatriez les 9 séquences correspondantes à partir de la base de données **GenBank** du **NCBI**. Veillez à utiliser le format **FASTA**.

Question 2 : Utilisez **SEAVIEW** [<http://doua.prabi.fr/software/seaview>] pour aligner les séquences à l'aide de **CLUSTAL** et à l'aide de **MUSCLE**.

Question 3 : Que remarquez vous sur la taille des *indels* ? A votre avis à quoi cela est-il dû ? Certains *indels* ne possèdent pas cette propriété. Trouvez un exemple et une explication à ce phénomène.

Question 4 : Trouvez des exemples *d'indels* qui sont en accord avec la taxonomie moléculaire et trouvez en qui illustrent l'homoplasie. Comment peut-on trancher entre les deux phénomènes ?

Question 5 : Choisissez un *indel* particulier. En supposant que la phylogénie ci-dessus est la bonne, utilisez la comme support pour représenter deux histoires possibles permettant d'expliquer l'existence de cet *indel* chez les espèces correspondantes.

Question 6 : À partir d'un alignement respectueux du cadre de lecture de **BRCA1**, les indels ont été codés proportionnellement à leur longueur comme des caractères avec deux états : présence (**I** = insertion) ou absence (**D** = délétion). Deux matrices de caractères ont été construites : soit en prenant **tous les 366 indels** [fichier **BRCA1_indels_all.fas**], soit en ne gardant que les **285 indels non-emboîtés** [fichier **BRCA1_indels_non_embedded.fas**]. Analysez les matrices binaires obtenues par maximum de parcimonie avec bootstrap. Comparez les topologies obtenues à celle habituellement admise. Qu'en pensez-vous ?

Question 7 : Comparez les arbres phylogénétiques inférés (i) par la méthode de distances **NJ** (ou **BioNJ**) appliquée aux distances génétiques calculées sous le modèle de Jukes-Cantor (**J-C**) ; (ii) par maximum de parcimonie ; et (iii) par maximum de vraisemblance sous le modèle de Jukes-Cantor (**JC69**) et avec l'opérateur SPR de recherche d'arbres.