

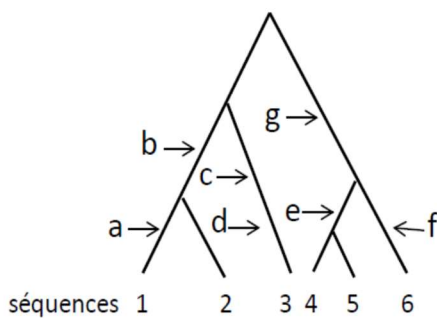
TD3 Evolution Moléculaire

Exercice 1 : Généalogie, polymorphisme et signature intraspécifique de sélection dans un gène.

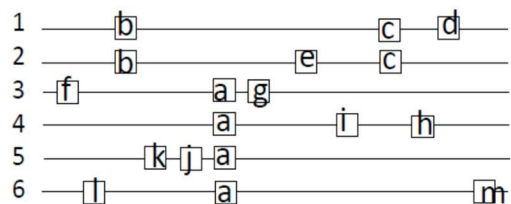
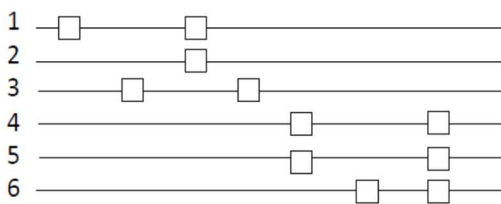
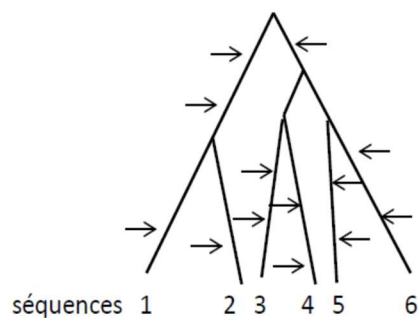
Chez la plante *Arabidopsis thaliana*, le gène *TX* codant pour une enzyme de détoxification d'un métal lourd, le cadmium, a été séquencé chez 6 individus dans 2 populations différentes : la population 1 située dans une zone peu polluée, et la population 2 située dans une zone périurbaine soumise à certains rejets de déchets industriels.

Voici 2 généalogies, ainsi qu'une schématisation de l'alignement des 300 paires de bases séquencées chez les 6 individus dans chaque population, avec les mutations aux sites polymorphes symbolisées par un carré.

généalogie pop 1



généalogie pop 2



1 – placez les mutations apparues dans la généalogie 1 sur l'alignement obtenu pour la population 1.

2 – placez sur la généalogie de la population 2, les mutations identifiées sur l'alignement.

3 – pour chacune des populations, calculez l'estimateur de polymorphisme de Watterson

$$\theta_S = S / \sum_{i=1}^{n-1} (1/i)$$

avec S = nombre sites polymorphes, n = taille de l'échantillon.
Comparez-les.

4 – On donne $\theta_\pi = 3,2$ et $4,93$ pour les populations 1 et 2, respectivement. On donne aussi

$$\sqrt{\text{Var}(\theta_\pi - \theta_S)} = 0,53 \text{ et } 0,93 \text{ pour les populations 1 et 2, respectivement.}$$

Calculez la statistique neutralité D de Tajima pour le gène TX dans chaque population.

$$D = \frac{\theta_\pi - \theta_S}{\sqrt{\text{Var}(\theta_\pi - \theta_S)}}$$

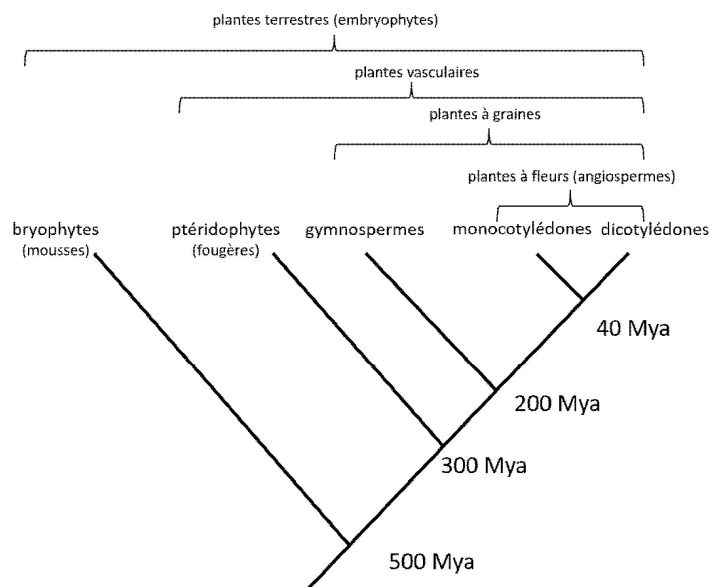
5 – On donne, sous l'hypothèse nulle que $D \sim N(0,1)$:

- $\text{Prob}(D < 0.19) = 0.61$
- $\text{Prob}(D < -0.83) = 0.02$

Interprétez ces résultats en termes de différences de signatures moléculaires de sélection sur le gène TX entre les 2 populations.

Exercice 2 : Comparaison des taux de substitutions synonymes et non synonymes à l'échelle phylogénétique

L'arbre phylogénétique ci-dessous résume l'évolution des grands groupes de plantes terrestres.



L'apparition de l'appareil reproducteur floral est une innovation majeure dans la diversification des végétaux. Le facteur de transcription APETALA3 est connu pour être un régulateur central du développement floral chez les angiospermes. L'orthologue d'APETALA3 est présent chez les grands groupes de plantes terrestres. Des chercheurs ont séquencé la séquence codante d'APETALA3 chez plusieurs espèces, dont 1 représentant par groupe. Nous nous limitons ici aux séquences présentes chez les plantes à graines (gymnospermes et angiospermes). Un alignement d'une partie de l'exon 1, dont la traduction protéique présente une signature de domaine de liaison à l'ADN de type MADS-box, est présentée ci-dessous :

Gymno	CTG	GCG	CCA	TCA
Monocot	CCG	GCT	TCA	ACA
Dicot	CCG	GCA	TCA	ACA

En vous aidant du code génétique, ci-après :

- 1 – Traduisez en séquences protéiques (donner la séquence en nomenclature internationale).
- 2 – Pour chaque séquence (chaque espèce), déterminez le nombre de sites synonymes et non-synonymes. Puis déterminez le nombre moyen de sites synonymes et non-synonymes entre chaque paire de séquence.
- 3 – Déterminez les taux de substitutions non-synonymes (d_N) et synonymes (d_S), puis calculez le ratio $\omega = d_N/d_S$ pour chaque paire de séquences. Interprétez les résultats.

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA STOP UAG STOP	UGU } Cys UGC } UGA STOP UGG Trp	U C A G	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } Ile AUC } AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } Val GUC } GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	

Key:

- Ala = Alanine (**A**)
- Arg = Arginine (**R**)
- Asn = Asparagine (**N**)
- Asp = Aspartate (**D**)
- Cys = Cysteine (**C**)
- Gln = Glutamine (**Q**)
- Glu = Glutamate (**E**)
- Gly = Glycine (**G**)
- His = Histidine (**H**)
- Ile = Isoleucine (**I**)
- Leu = Leucine (**L**)
- Lys = Lysine (**K**)
- Met = Methionine (**M**)
- Phe = Phenylalanine (**F**)
- Pro = Proline (**P**)
- Ser = Serine (**S**)
- Thr = Threonine (**T**)
- Trp = Tryptophan (**W**)
- Tyr = Tyrosine (**Y**)
- Val = Valine (**V**)