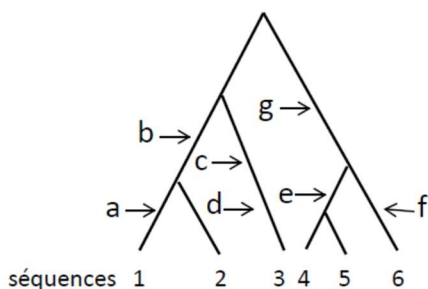


# TD3 Evolution Moléculaire **CORRECTION**

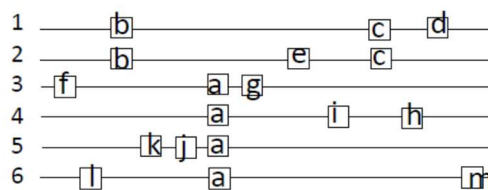
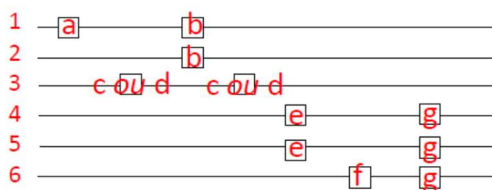
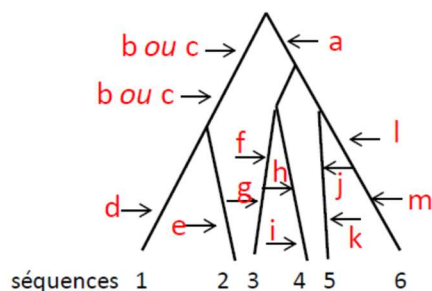
## Exercice 1 : Généalogie, polymorphisme et signature intraspécifique de sélection dans un gène.

### Questions 1 et 2

généalogie pop 1



généalogie pop 2



### Question 3:

$$\theta_s \text{ pop}_1: 7/[1+(1/2)+(1/3)+(1/4)+(1/5)] = 3.065693 = 3.1$$

$$\theta_s \text{ pop}_2: 13/[1+(1/2)+(1/3)+(1/4)+(1/5)] = 5.693431 = 5.7$$

le polymorphisme (en terme de nombre de site polymorphes) semble supérieur dans la pop2. Mais on ne sait pas si cela est significatif.

### Question 4 et 5:

- $D_{(TX\text{-pop}_1)}: (3.2-3.1)/0.53 = 0.1886792 = \mathbf{0.19}$
- $\rightarrow$  p-valeur en test unilatéral ( $H_1$  = sélection négative OU sweep) :  $\text{Prob}(D < 0.19) = 0.61 \rightarrow$  non significatif au seuil de 5%  $\rightarrow$  on ne rejette pas la neutralité.
- $\rightarrow$  p-valeur en test unilatéral ( $H_1$  = sélection balancée) :  $\text{Prob}(D > 0.19) = 0.39 \rightarrow$  non significatif au seuil de 5%  $\rightarrow$  on ne rejette pas la neutralité.
- $D_{(TX\text{-pop}_2)}: (4.93-5.7)/0.93 = -0.827657 = \mathbf{-0.83}$

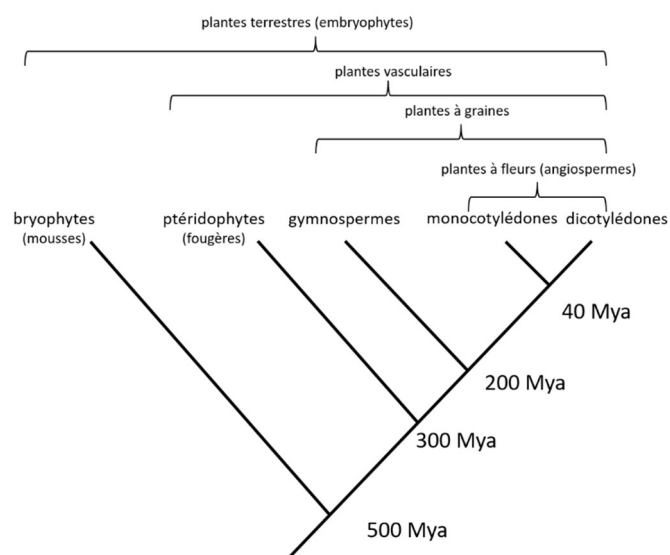
- → p-valeur en test unilatéral ( $H_1 =$  sélection négative OU sweep) :  $\text{Prob}(D < -0.83) = 0.02$  → significatif au seuil de 5% → on rejette  $H_0$  (neutralité) → **gène sous sélection négative OU balayage selectif (sweep).**

Dans la pop 1, TX évolue de manière neutre, tandis que dans la pop 2 ( $D=-0.83$  significativement  $< 0$ ) → traduit un excès d'allèles rares compatible avec de la sélection négative (conservation de séquences) ou un « selective sweep » (sélection positive forte sur une mutation avantageuse qui se fixe rapidement, et mutations neutres qui apparaissent par la suite sur l'haplotype portant cette mutation rapidement fixée). On ne peut distinguer les 2 hypothèses avec cette statistique.

En l'absence d'information supplémentaire, l'hypothèse de sélection négative (= purifiante) est suffisante (la mutation « a » est optimale, toute autre mutation sur la séquence est contre-sélectionnée).

## 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).

Gymno	Leu	Ala	Pro	Ser
Monocot	Pro	Ala	Ser	Thr
Dicot	Pro	Ala	Ser	Thr

Gymno	LAPS
Monocot	PAST
Dicot	PAST

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 pour chaque paire de séquence.

#### *Nombre de sites synonymes et non-synonymes par séquences*

Gymno	CTG	GCG	CCA	TCA	
syn	$\frac{1}{3}01$	001	001	001	$\Rightarrow 4 + \frac{1}{3} = \frac{13}{3} = 4.33$
non_syn	$\frac{2}{3}10$	110	110	110	$\Rightarrow 7 + \frac{2}{3} = \frac{23}{3} = 7.66$
Monocot	CCG	GCT	TCA	ACA	
syn	001	001	001	001	$\Rightarrow 4$
non_syn	110	110	110	110	$\Rightarrow 8$
Dicot	CCG	GCA	TCA	ACA	
syn	001	001	001	001	$\Rightarrow 4$
non_syn	110	110	110	110	$\Rightarrow 8$

#### *Nombre moyen de sites synonymes et non synonymes pour chaque paire de séquences*

Gymno vs Monocot  
 $n_{syn} = (4.33+4)/2 = 4.17$   
 $n_{non\_syn} = (7.66+8)/2 = 7.83$

Gymno vs Dicot  
 $n_{syn} = (4.33+4)/2 = 4.17$   
 $n_{non\_syn} = (7.66+8)/2 = 7.83$

Monocot vs Dicot  
 $n_{syn} = (4+4)/2 = 4$   
 $n_{non\_syn} = (8+8)/2 = 8$

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.

#### Gymno vs Monocot

Gymno	C <b>T</b> G	G <b>C</b> G	<b>C</b> CA	<b>T</b> CA
Monocot	<b>C</b> CG	G <b>C</b> T	<b>T</b> CA	<b>A</b> CA

$$d_S = 1 / 4.17$$

$$d_N = 3 / 7.83$$

$$\omega_{\text{Gymno\_vs\_Monocot}} = d_N/d_S = 1.6$$

#### Gymno vs Dicot

Gymno	C <b>T</b> G	G <b>C</b> G	<b>C</b> CA	<b>T</b> CA
Dicot	<b>C</b> CG	G <b>C</b> A	<b>T</b> CA	<b>A</b> CA

$$d_S = 1 / 4.17$$

$$d_N = 3 / 7.83$$

$$\omega_{\text{Gymno\_vs\_Dicot}} = d_N/d_S = 1.6$$

#### Monocot vs Dicot

Monocot	CCG	G <b>C</b> T	TCA	ACA
Dicot	CCG	G <b>C</b> A	TCA	ACA

$$d_S = 1 / 4$$

$$d_N = 0 / 8$$

$$\omega_{\text{Monocot\_vs\_Dicot}} = d_N/d_S = 0$$

La séquence peptidique du motif de liaison à l'ADN du facteur de transcription APETALA3 est parfaitement conservée chez les plantes à fleurs (pas de subst non-syn,  $\omega = 0$ ) ; on observe seulement 1 substitution synonyme entre monoco et dicot. Ceci signifie que les changements d'acide aminés sont contre-sélectionnés car le domaine « PAST » d'APETALA3 est primordial d'un point de vue fonctionnel chez les plantes à fleurs.

Quand l'on compare les angiospermes aux gymnospermes,  $\omega = 1.6$ , ce qui indique une accélération des changements d'acides aminés par sélection positive, correspondant ainsi à une adaptation apparue après la divergence avec les gymnospermes, chez l'ancêtre des angiospermes, avant leur diversification.

On rappelle que :

- $d_N$  = nombre de substitutions non synonymes / nombre de sites non synonymes
- $d_S$  = nombre de substitutions synonymes / nombre de sites synonymes
- $0 < \omega < 1 \rightarrow$  sélection négative (forte quand  $\omega$  tend vers 0)
- $\omega = 1 \rightarrow$  neutralité
- $\omega > 1 \rightarrow$  sélection positive, diversifiante (forte quand  $\omega$  tend vers  $+\infty$ )

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA <b>STOP</b> UAG <b>STOP</b>	UGU } Cys UGC } UGA <b>STOP</b> 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 <b>Met</b>	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**)