

**M1 MABS**  
**Examen terminal d'Evolution Moléculaire (durée 2h)**  
**(EM7BMAAME2) – Janvier 2012**

**Questions de cours**

- Donner la définition d'un profil génétique. Comment l'établit-on ? Pourquoi établir la similarité entre les profils génétiques de plusieurs espèces?
- Comment en pratique identifie-t-on les gènes orthologues à partir des séquences ?

**Problème 1**

Une partie de l'alignement de la séquence nucléotidique de l'exon 2 d'un gène de 8 individus de *Phytophthora infestans* (pathogène de la pomme de terre) est présenté en Figure 1a. La séquence ancestrale la plus proche a été inférée par une approche phylogénétique, incluant d'autres espèces de *Phytophthora* (Figure 1b). Les sites variables sont indiqués en gras et soulignés. Justifiez à chaque fois vos réponses:

- expliquez quelle(s) force(s) évolutive(s) est/sont à l'origine du profil de polymorphisme observé sur cette séquence.

Par des approches *in silico*, il a été mis en évidence que cette séquence contient un motif moléculaire pouvant être impliqué dans des interactions protéines-protéines. Une analyse cytologique de plantes infectées par les différents individus de *Phytophthora infestans* indique que (i) la protéine codée par l'individu 8, interagit avec une protéine cytosolique végétale ii) les protéines codées par les individus 1 à 7 sont adressées au noyau de la cellule végétale. De plus, les individus 1 à 7 causent des symptômes nettement plus importants que l'individu 8 sur des plants de pomme de terre.

- donnez une explication biologique évolutive à l'ensemble de ces données.

**Figure 1a:** séquence de l'exon 2, chez 8 individus de *Phytophthora infestans*:

1 -	<u>ACT</u>	<u>GCG</u>	CTG	TCC	<u>CCG</u>	CGC	TCA	GGC	<u>AGC</u>
2 -	<u>ACT</u>	<u>GCG</u>	CTG	TCC	<u>CCA</u>	CGC	TCA	GGC	<u>AGT</u>
3 -	<u>ACT</u>	<u>GCG</u>	CTG	TCC	<u>CCG</u>	CGC	TCA	GGC	<u>AGT</u>
4 -	<u>ACT</u>	<u>GCG</u>	CTG	TCC	<u>CCA</u>	CGC	TCA	GGC	<u>AGC</u>
5 -	<u>ACT</u>	<u>GCG</u>	CTG	TCC	<u>CCA</u>	CGC	TCA	GGC	<u>AGC</u>
6 -	<u>ACT</u>	<u>GCG</u>	CTG	TCC	<u>CCA</u>	CGC	TCA	GGC	<u>AGC</u>
7 -	<u>ACT</u>	<u>GCT</u>	CTG	TCC	<u>CCA</u>	CGC	TCA	GGC	<u>AGC</u>
8 -	<u>TCT</u>	<u>GCT</u>	CTG	TCC	<u>CCA</u>	CGC	TCA	GGC	<u>AGC</u>

**Figure 1b:** séquence de l'exon 2, chez l'ancêtre le plus proche de *Phytophthora infestans*:

TCT GCT CTG TCC CCA CGC TCA GGC AGC

NB :

ACT = Thr, TCT = Ser,

GCG = Ala, GCT = Ala,

CCA = Pro, CCG = Pro

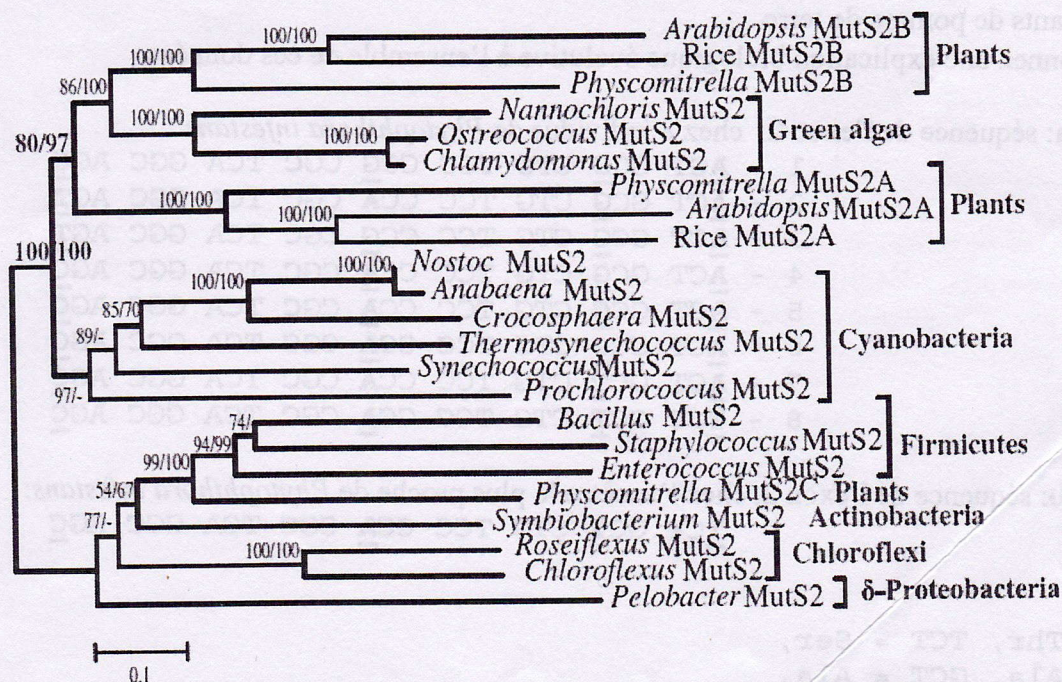
AGC = Ser, AGT = Ser,



## Problème 2

Zhenguo et collaborateurs (Nucleic Acids Res., 2007, 1-13) ont conduit une analyse phylogénétique des composants clés du système de réparations des mésappariements au niveau de l'ADN. Ces mésappariements sont introduits par l'ADN polymérase lors de la division cellulaire et leur non correction entraînerait des mutations. Le gène codant pour la protéine MutS est un acteur de ce système de réparation. Chez *Escherichia coli* il a été démontré qu'un homodimère de MutS se lie sur les nucléotides mal appariés et forme un complexe MutS-ADN. Cependant plusieurs gènes homologues à *mutS* existent dans *E. coli* et une recherche de similarité dans 461 génomes bactériens a montré qu'on pouvait distinguer 4 sous-familles de protéines MutS, MutS1 (correspondant à la protéine MutS dont la fonction est décrite ci-dessus) à MutS4. Chez les eucaryotes, 7 gènes homologues aux gènes MutS bactériens ont été identifiés. Les arbres phylogénétiques ont montré qu'ils appartenaient aux sous-familles MutS1 et MutS2. Nous allons nous intéresser ici à l'évolution des séquences protéiques MutS2. Ces dernières n'ont été identifiées que dans les génomes nucléaires des espèces possédant des chloroplastes, à savoir plantes à fleurs et algues vertes. L'analyse évolutive de cette sous-famille de protéines a été réalisée sur un ensemble d'espèces représentatives et a conduit à l'arbre de la figure 2. La reconstruction phylogénétique a été réalisée à l'aide de deux méthodes la Neighbor Joining (NJ) et maximum de vraisemblance (PhyML). Les deux topologies étant congruentes, *i.e.* les mêmes, seul l'arbre obtenu par la NJ est présenté mais les valeurs de bootstrap obtenues avec chacune des méthodes sont reportées (NJ et PhyML).

- Analyser l'arbre de la figure 2 pour émettre des hypothèses quant à l'origine et l'évolution de ce gène : duplication, transferts horizontaux, endosymbiose (il est actuellement accepté que les mitochondries proviendraient d'endosymbiontes apparentées aux  $\alpha$ -protéobactéries et les chloroplastes d'endosymbiontes apparentées aux cyanobactéries), etc.



**Figure 2:** Arbre phylogénétique obtenu sur un ensemble de séquences protéiques de la sous-famille MutS2 (extrait de Zhenguo *et al.*, 2007, Nucleic Acids Res., 1-13).



### Problème 3

Afin de rechercher des gènes fonctionnellement importants dans le génome de la plante modèle *Medicago truncatula*, des chercheurs ont recherché des traces de sélection naturelle chez deux lignées vivant dans des environnements très distincts. Un test de neutralité de Tajima a été effectué sur 30 gènes dans chacune des deux lignées, basé sur l'alignement des séquences de plusieurs individus pour chaque lignée. La formule de la statistique  $D$  du test de Tajima est :

$$D = \frac{\hat{\theta}_{\pi} - \hat{\theta}_s}{SE(\hat{\theta}_{\pi} - \hat{\theta}_s)}$$

où  $\hat{\theta}_{\pi}$  et  $\hat{\theta}_s$  sont deux estimateurs de la diversité au niveau nucléotidique sur une séquence (SE = écart-type).

- Lequel de ces deux estimateurs évalue la probabilité qu'un nucléotide soit polymorphe ? Expliquez comment vous feriez pour le calculer.
- Lequel de ces deux estimateurs évalue la probabilité d'hétérozygotie par nucléotide ? Expliquez comment vous feriez pour le calculer.

Le tableau ci-dessous résume les résultats du test de Tajima. Les 30 gènes sont classés selon les différents types de valeurs prises par  $D$  et selon les mécanismes moléculaires auxquels ils sont associés.

Lignée	Mécanismes moléculaires	$D > 0$	$D < 0$	$D = 0$
1	stress abiotique prolongé	0	3	10
	résistance aux pathogènes (gènes paralogues)	12	0	5
2	stress abiotique prolongé	0	10	3
	résistance aux pathogènes (gènes paralogues)	5		12

- Que signifient des valeurs de  $D = 0$ ,  $< 0$  et  $> 0$ .
- Interprétez les différences entre lignées en termes de pressions de sélections exercées par leur environnement et de conséquence sur le polymorphisme moléculaire de ces gènes (justifiez qualitativement).

NB : le test est significatif au seuil 0.001 pour tous les gènes avec  $D > 0$  et  $D < 0$